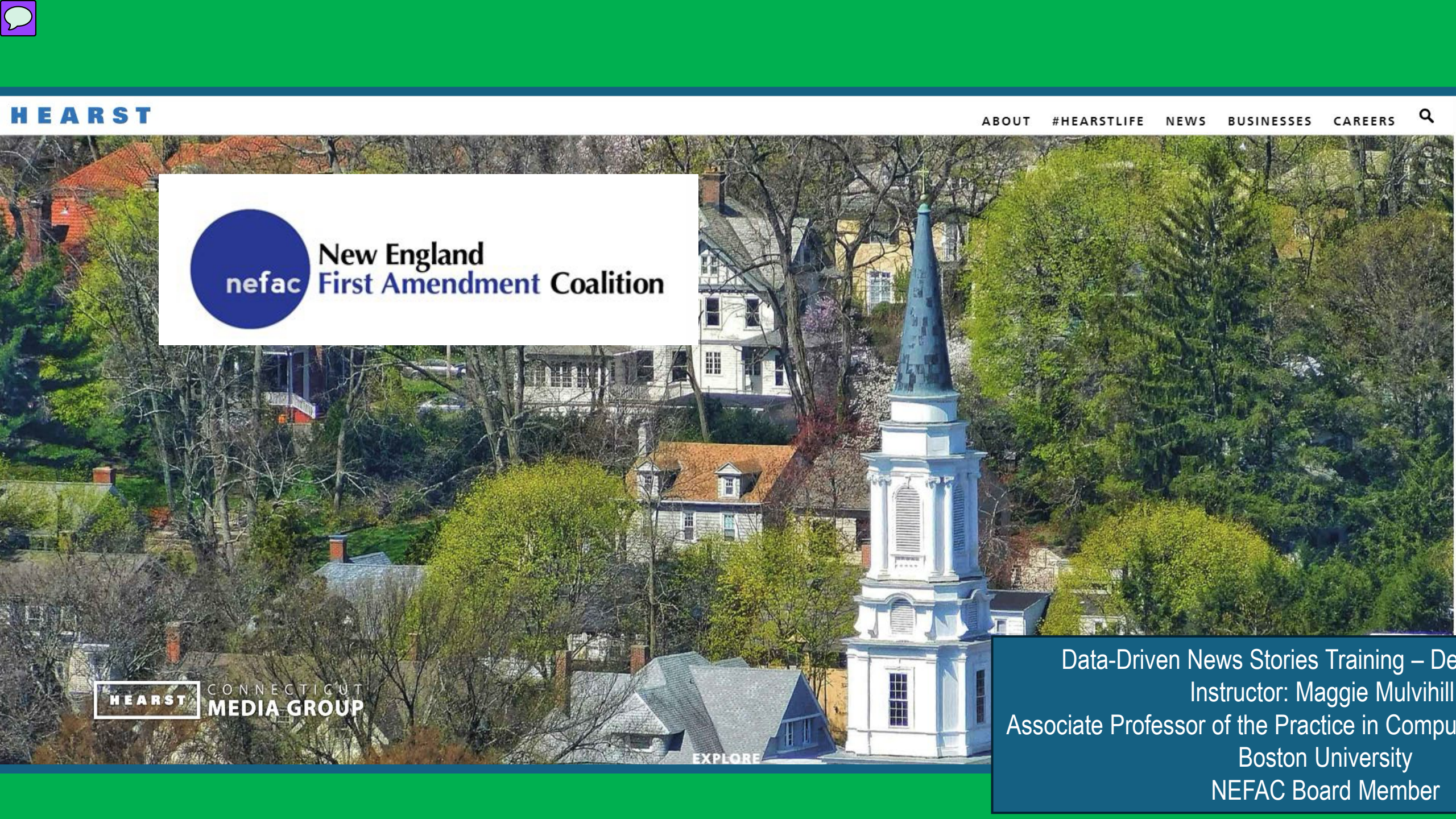
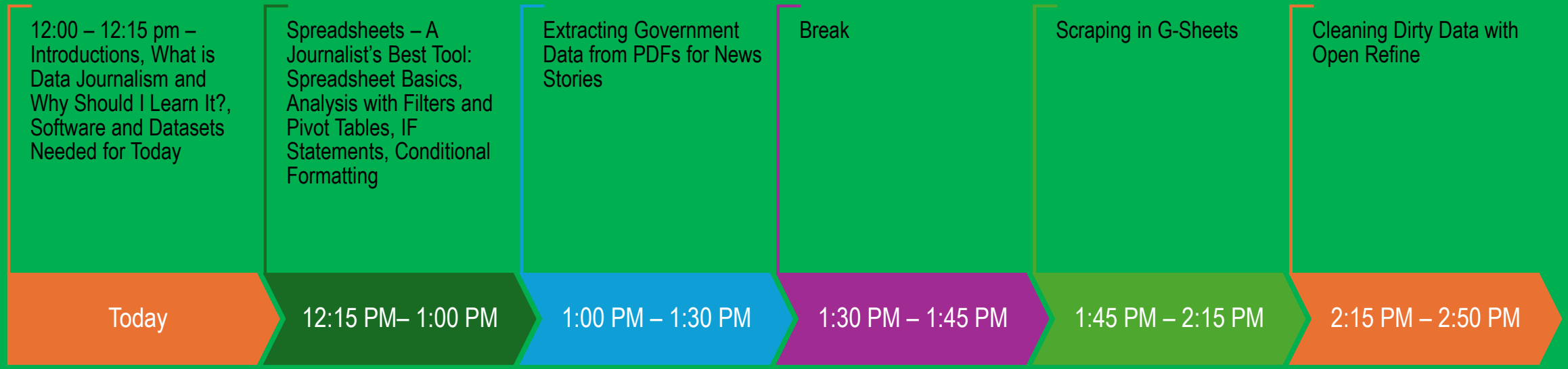


Data-Driven News Stories Training – De
Instructor: Maggie Mulvihill
Associate Professor of the Practice in Compu
Boston University
NEFAC Board Member

Today's Lesson



DPH - Regulatory Action Report

10/01/2021 to 12/31/2021

Divisio	Profession	Respondent	Credential Number	City	File Number	Other States	Date Closed	Case Resolution	Case Action	Case
OEMS	Emergency Medical Service	BUCKLEY, CHERYLANN	1466	ROCKY HILL	2020-41	None Reported	10/4/2021		Probation Completed	Business F
OEMS	Emergency Medical Technician	RING, SANDRA E	5605	SHELTON	2020-43	None Reported	10/22/2021	Consent Order	Probation/ Restriction/ Therapy	Substance Abuse/Drug R
OEMS	Paramedic	SAWICKI, ZACHARY J	2768	Torrington	2018-50	MA	12/16/2021	Consent Order	Probation/ Remedial Education/	Negli
PLIS	Advanced Practice Registered Nurse	BUCCI, KELLY M	1809	FAIRFIELD	2020-1067	None Reported	11/17/2021	Consent Order	Fine/ Reprimand/ Restriction	Drug Related, Inappropriate Pres
PLIS	Advanced Practice Registered Nurse	DUNIGAN, SUSAN M	E49878	SHELTON	2020-244	CT	12/15/2021	Consent Order	Employer Reports/ Probation/ Remedial	Incompetence/Negligence/ Mai
PLIS	Advanced Practice Registered Nurse	PYRAM, MARTINE NOEMIE	9206	Burlington	2021-71	NY/ OH	11/17/2021		Suspension	Mandatory Report/ Mental
PLIS	Barber	COPLAND, WILLIE K	3672	WEST HAVEN	2020-363	None Reported	10/7/2021	Reinstatement Consent	Fine/ Reprimand	Reinsta
PLIS	Barber	JORDAN, BLAKE O	4916	BRIDGEPORT	2021-1009	None Reported	12/28/2021	Reinstatement Consent	Fine	Reinsta
PLIS	Barber	RIVERA, JOSE A	3981	BRIDGEPORT	2021-349	None Reported	11/4/2021	Reinstatement Consent	Fine/ Reprimand	Reinsta
PLIS	Barber	RUCKER, SHAWN D	4942	NEW LONDON	2021-477	None Reported	12/6/2021	Reinstatement Consent	Fine/ Reprimand	Reinsta
PLIS	Dentist	FLANAGAN, DENNIS F	4378	WILLIMANTIC	2018-1193	None Reported	10/13/2021	Consent Order	Probation/ Remedial Education/	Incompetence/Neg
PLIS	Hairdresser/Cosmetician	AVELLANET, ERIN M	60731	Milford	2021-932	None Reported	11/19/2021	Reinstatement Consent	Fine/ Reprimand	Reinsta
PLIS	Hairdresser/Cosmetician	BURNS, JENNIFER S	49287	BRISTOL	2021-938	None Reported	11/9/2021	Reinstatement Consent	Fine/ Reprimand	Reinsta
PLIS	Hairdresser/Cosmetician	CAPUTO, LAURA	38531	GUILFORD	2021-942	None Reported	11/24/2021	Reinstatement Consent	Fine/ Reprimand	Reinsta
PLIS	Hairdresser/Cosmetician	HARDER, LAURIE J	53400	CLINTON	2021-939	None Reported	11/16/2021	Reinstatement Consent	Fine/ Reprimand	Reinsta
PLIS	Hairdresser/Cosmetician	JOINER, JOYCE B	57431	BETHEL	2021-787	CO/ NY	11/23/2021	Reinstatement Consent	Fine/ Reprimand	Reinsta
PLIS	Hairdresser/Cosmetician	LIGHT, LASETT A	64676	WEST PALM	2021-616	None Reported	10/28/2021	Reinstatement Consent	Fine/ Reprimand	Reinsta
PLIS	Hairdresser/Cosmetician	LIGUORI, ANTONIO	60090	BRANFORD	2020-594	CT	12/21/2021	Reinstatement Consent	Fine/ Reprimand	Reinsta
PLIS	Licensed Alcohol and Drug	MARTUSIS, JOSEPH M	1155	BRISTOL	2020-1254	None Reported	10/13/2021	Consent Order	Probation/ Random Urine Screens/	Mandatory Report/ Sub
PLIS	Licensed Clinical Social Worker	CALDWELL, ELIJAH	8201	WATERBURY	2019-783	FL	11/9/2021		Probation Completed	Reinsta
PLIS	Licensed Practical Nurse	BARD, LINDA C	27663	HAMPTON	2014-875	None Reported	11/30/2021		Probation Completed	Scope of F
PLIS	Licensed Practical Nurse	BRADSHAW, ROBIN E	23473	WATERBURY	2021-732	None Reported	11/1/2021	Voluntary Surrender		Incompetence/Neg
PLIS	Licensed Practical Nurse	BRICE, KRISTIN A	38597	FAIRFIELD	2020-1239	SC	10/20/2021	Memorandum of Decision	Employer Reports/ Narcotics Key	Violation of Pr
PLIS	Licensed Practical Nurse	CROTEAU, SARAH J	25849	SANDY HOOK	2021-739	None Reported	11/17/2021		Suspension	Substance Abuse/Drug F
PLIS	Licensed Practical Nurse	FITZPATRICK, ERIN K	37560	LISBON	2020-1015	None Reported	10/20/2021	Memorandum of Decision	Employer Reports/ Narcotics Key	Illegal Conduct/ Mandatory
PLIS	Licensed Practical Nurse	FITZPATRICK, ERIN K	37560	LISBON	2020-330	None Reported	10/20/2021	Memorandum of Decision	Employer Reports/ Narcotics Key	Substance Abuse/Drug F
PLIS	Licensed Practical Nurse	FREDERICK, MEGHAN H	37253	DAYVILLE	2021-206	MA	12/15/2021	Memorandum of Decision	Revocation/ Suspension	Mandatory
PLIS	Licensed Practical Nurse	HOWLEY, DANIELLE P	30383	BETHEL	2021-557	None Reported	10/20/2021		Suspension	Professional Boundaries/ Profe
PLIS	Licensed Practical Nurse	LAMBERT, ASHLEY	40682	THOMPSON	2021-931	None Reported	12/1/2021		Suspension	Violation of Pr
PLIS	Licensed Practical Nurse	MERCADO, LOURDES J	36248	WILLIMANTIC	2020-1131	None Reported	10/20/2021	Memorandum of Decision	Probation/ Remedial Education/	Violation of Pr
PLIS	Licensed Practical Nurse	RING, SANDRA E	34369	SHELTON	2020-1074	None Reported	10/20/2021	Consent Order	Employer Reports/ Narcotics Key	Substance Abuse/Drug F
PLIS	Licensed Practical Nurse	STEWART, TIONA A	38404	STRATFORD	2017-1275	None Reported	12/15/2021	Consent Order	Employer Reports/ Probation/ Remedial	Illegal C
PLIS	Licensed Practical Nurse	TESTA, TIMOTHY	36189	BRIDGEPORT	2021-443	None Reported	10/20/2021	Consent Order	Employer Reports/ Probation/ Random	Mandatory Report/ Sub

IRE
INVESTIGATIVE REPORTERS & EDITORS

REPORTERS COMMITTEE
FOR FREEDOM OF THE PRESS

BU College of Communication

For Current Students For Alumni For Parents GI

ACADEMICS ADMISSIONS RESEARCH CAMPUS LIFE ABOUT COM NEWS EVENTS APPLY



Maggie Mulvihill

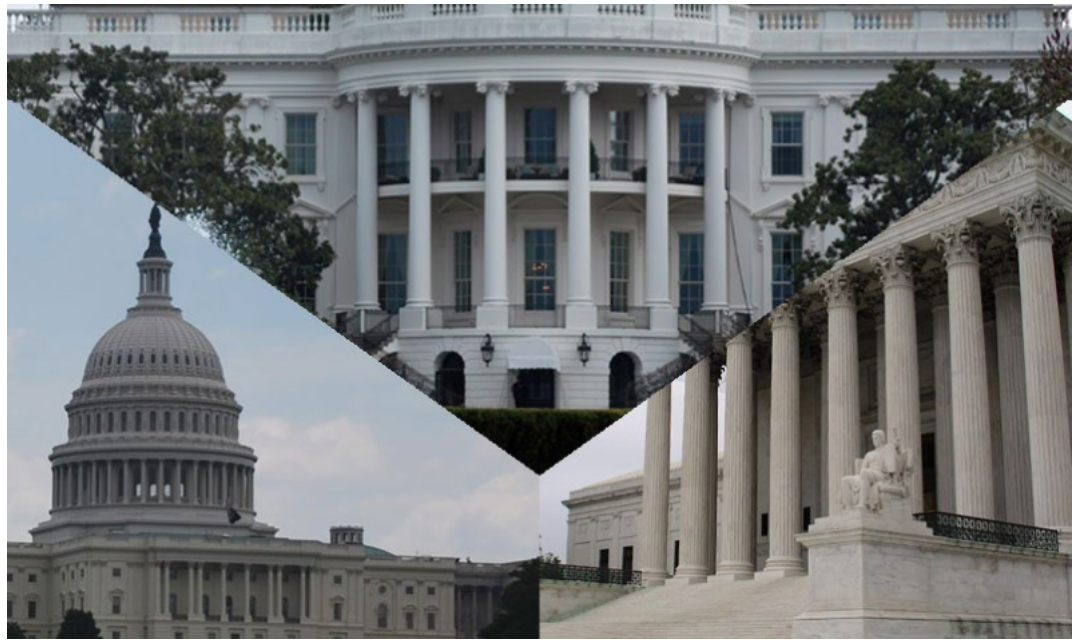
Associate Professor of the Practice of Computational Journalism

Email: mmulvih@bu.edu **Phone:** 617-353-4978



Maggie Mulvihill is a veteran reporter, data journalism trainer, news entrepreneur, First Amendment advocate and attorney. She has been a member of the BU Journalism faculty since 2009, where she developed the Journalism Department's first data journalism course and teaches the Law and Ethics of Media and reporting classes.





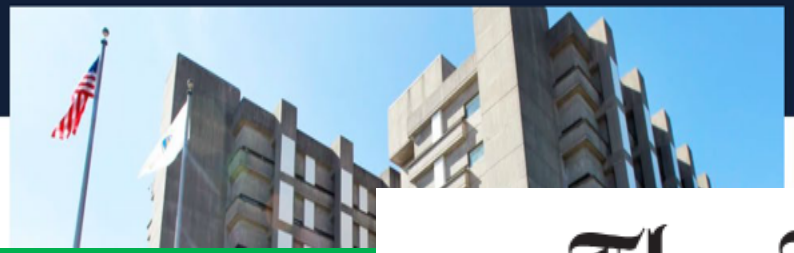
The Government Works for You, Not The Other Way Around

NBC NEWS MEET THE PRESS U.S. NEWS LOCAL POLITICS WORLD BUSINESS EDITORS' PICKS SHOPPING TIPLINE WATCH LIVE

U.S. NEWS

Epic Massachusetts crime lab scandal may involve even broader wrongdoing, judge says

State authorities said for years that one rogue chemist was behind the compromised drug tests that led to tens of thousands of overturned cases. But a judge says there may be more.

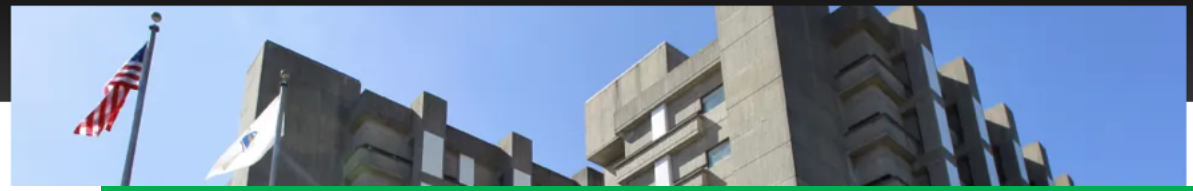


AP WORLD U.S. POLITICS SPORTS ENTERTAINMENT BUSINESS SCIENCE FACT CHECK ODDITIES BE WELL NEWSLETTERS PHOTOGRAPHY AP

Kerch strait New Jersey drones Timothée Chalamet 100 years in prison Celtic vs. Rangers

AP SETS THE STANDARD FOR P
SUPPORT INDEPENDENT, FACT-

Judge declines mass dismissal of cases from tainted drug lab



The Boston Globe

Five more linked to burgeoning Boston polic fraud scandal, including former union presi

By Dugan Arnett and Maggie Mulvihill Globe Staff and Globe correspondent,
Updated June 8, 2021, 7:41 p.m.



Another Boston police scandal is quietly unfolding

It may surpass the high-profile State Police overtime fraud debacle

By Dugan Arnett and Maggie Mulvihill Globe Staff and Globe correspondent,
Updated June 5, 2021, 4:30 p.m.

167



More cases tied to scandal-plagued drug lab could be tossed in 'nightmare scenario' for state's criminal justice system

The Middlesex DA wants the Supreme Judicial Court to investigate and potentially invalidate a decade's worth of drug tests

By Maggie Mulvihill and Dugan Arnett Globe correspondent and Globe Staff.
Updated March 26, 2021, 9:31 a.m.

Suffolk District Attorney Rachael Rollins looking into third chemist at state drug lab

Investigators are looking into longtime state chemist Della Saunders, who analyzed tens of thousands of drug samples while working alongside disgraced chemist Annie Dookhan for about a decade

By Maggie Mulvihill and Dugan Arnett Globe correspondent and Globe Staff,
Updated December 17, 2020, 10:29 a.m.

AP WORLD U.S. POLITICS SPORTS ENTERTAINMENT BUSINESS SCIENCE FACT CHECK ODDITIES BE WELL NEWSLET

Kerch strait New Jersey drones Timothée Chalamet 100 years in prison Celtic vs. Rangers

Mass. health boards overwhelmed, unready to protect workers

BY SHANNON IRIARTE, SHWETHA SURENDRAN AND MAGGIE MULVIHILL/THE HOWARD CENTER FOR INVESTIGATIVE JOURNALISM
Published 12:25 PM EST, May 12, 2021

Share

BOSTON (AP) — With federal regulators missing from the field and state leaders scrambling to manage the COVID-19 crisis, Massachusetts' 351 overtaxed local boards of health were unwittingly thrust into a new role last year — overseers of workplace safety.

Suffolk DA kicks off effort to drop tens of thousands of convictions tied to tainted state drug lab

By Maggie Mulvihill Globe correspondent, Updated April 29, 2021, 6:24 p.m.



AP

WORLD U.S. POLITICS SPORTS ENTERTAINMENT BUSINESS SCIENCE FACT CHECK ODDITIES BE WELL NEWSLETTERS

Kerch strait New Jersey drones Timothée Chalamet 100 years in prison Celtic vs. Rangers

U.S. NEWS

When House members travel the globe on private dime, families often go too



1 of 3 | FILE - Rep. Jan Schakowsky, D-Ill., with her husband Robert Creamer, center, stands for a ceremonial photo with Speaker of the House John Boehner, R-Ohio, left, in the Rayburn Room of the Capitol after the 113th Congress convened, Jan. 3, 2013, in Washington. (AP Photo/J. Scott Applewhite, File)

BY PHILIP RANDAZZO, HAIYI BI, AKANKSHA GOYAL / BOSTON UNIVERSITY, THE HOWARD CENTER FOR INVESTIGATIVE JOURNALISM AT THE UNIVERSITY OF MARYLAND, BOSTON UNIVERSITY AND THE HOWARD CENTER FOR INVESTIGATIVE JOURNALISM AT THE UNIVERSITY OF MARYLAND

Share

Maggie Mulvihill
April 10 ·

A huge salute to the hardest working reporting team around who earned first place in the in-depth category, beating out competition from colleges throughout New England, New York and Pennsylvania. congratulations to College of Communication, Boston University journalists Nicholas McCool Anoushka Dalmia Sophia Brown Noemi Arellano-Summer for keeping your shoulders firmly on the wheel to produce these prize-winning stories on housing and homelessness for @HowardCenterUMD @bost... See More

MOE 2020
MARK OF EXCELLENCE AWARDS

SPJ Region 1 Conference
April 10 ·

Congratulations to all 2020 Mark of Excellence winners and finalists! We announced the results of the judging during a virtual ceremony on Saturday, April 10. C... See More

---

enSecrets.org
Center for Responsive Politics

Search

e / News & Analysis / OpenSecrets Blog

ky-high price of hepatitis C drug produces profits or lawmakers

by Alex Rozier on May 4, 2016

hile patients in Massachusetts wait affordable doses of the ockbuster hepatitis C drug Sovaldi, ep. Joseph Kennedy III has ancially benefited from the uccess of the drug's manufacturer, ilead Sciences, according to public ancial disclosures and campaign ance records.

nd he's not the only lawmaker



s.) holds the largest Congress in Gilead

This story was produced as part of Professor Maggie Mulvihill's Journalism Research course at Boston University.

DE ISLAND | POLITICS | EDUCATION | LIFESTYLE | ARTS

For the homeless in Boston and beyond, laws can criminalize life itself

By Nick McCool, Noemi Arellano-Summer and Maggie Mulvihill Globe correspondents,
Updated June 28, 2020, 5:32 p.m.

34



The Center for Public Integrity
Winner of the 2014 and 2017 Pulitzer Prize

Home About the Center ICIJ Donate Email newsletters

Politics National Security Business Environment Workers' Rights Health

Search the Center Search

Poisoned Places

States slow to act against New England polluters

By Maggie Mulvihill Alex Burris James Robinson 6:00 am, January 12, 2012 Updated: 12:19 pm, May 19, 2014

Student Bylines: The AP, The Boston Globe, The Washington Post and many other outlets

The Washington Post

Democracy Dies in Darkness

State lotteries transfer wealth out of needy communities

By Aadit Tambe, Stephen Neukam, Kate Seltzer, Rachel Logan, Henry Kuczynski, Emmett Gartner, Trisha Ahmed, Abby Zimmardi, Michael Purdie, Shreya Vuttaluru, Jamie Lin Pinzon / Howard Center for Investigative Journalism | AP

July 11, 2022 at 3:21 p.m. EDT

WARREN, Mich. — While the growing expansion of casinos and state-sanctioned sports betting steal the spotlight, ealth transfer

e University of
s in nearly e
ose stor

AP AP NEWS

U.S. News | World News | Politics | Sports | Entertainment | Business | Technology | Health | Science | Cli

Trending News | Jan. 6 hearings | Abortion | Astronomy | Boris Johnson | Sri Lanka | Russia-Ukraine war | C

RELATED TOPICS

Stores cash customers' checks, then sell them lotto tickets

By HENRY KUCZYNSKI, ETHAN BIDDLE, ISABEL TEHAN and MELISSA ELLIN / Boston University

July 11, 2022

Sections | The Washington Post *Democracy Dies in Darkness* | Get one year for $40 | Sign in

Confusion over eviction ban led to selective enforcement



By Brenda Wintrode, Amy Dipierro, Ryan Little, Luciana Perez Uribe, Aneurin Canham-Clyne, Trisha Ahmed, Sean Mcgoey, Sophia Brown and Maya Pottiger / The Howard Center for Investigative Journalism | AP

September 2, 2020

TAMPA, Fla. — Yochebed Israel was the kind of person Congress had in mind when it voted in March to impose a temporary moratorium on many evictions as the coronavirus spread.

First, the furnace in her Tampa, Florida, apartment broke and her electric bill rose above $460 a month for five months. She fell behind on rent, forced to choose between keeping the lights on or paying the landlord.

Mass. violates juvenile justice law

The violation has cost the state more than $600,000 in federal funding for at-risk youth programs since 2014.

By **Mia Cathell and Sophia Brown** Globe Correspondents, Updated February 5, 2020, 4:01 p.m.



For teens guilty of murder, penalties can vary widely

Review finds no pattern in young killers' terms since Mass. law change



Antonio Fernandez is serving life without parole for a slayi
Massachusetts youths have received the maximum penalty

By Sarah Favot, Kirsten Berg and Jenna Ebersol
INVESTIGATIVE REPORTING DECEMBER 27, 2

Top 10 Tr
Most View
Celtics' Gord season, spec
Dan Shaugh Celtics' seas
A nurse's aid veteran dies
New tests br
Gordon Hay broken and
At least thre park
Olympian M abused for y

Our Youngest Killers

December 27, 2011

By Sarah Favot, Kirsten Berg and Jenna Ebersole

Tweet 0 Like 0 Print Email



Livingston Awards Nominees: The Nominated Work

May 2, 2012 // 0

The Center for
Public Integrity
Winner of the 2014 and 2017 Pulitzer Prize

Our youngest killers

By Sarah Favot, Kirsten Berg and Jenna Ebersole January 4, 2012

The Boston Globe

THE NEWSGUILD - CWA
THE NEWSGUILD-COMMUNICATIONS WORKERS OF AMERICA
NEWS WHO WE ARE ORGANIZING OUR UNION

DAVID S. BARR AWARD FOR STUDENT JOURNALISM

The Maine Monitor

Examples of Published Work

MONEY & BUSINESS, SPORTS · 5 minute read

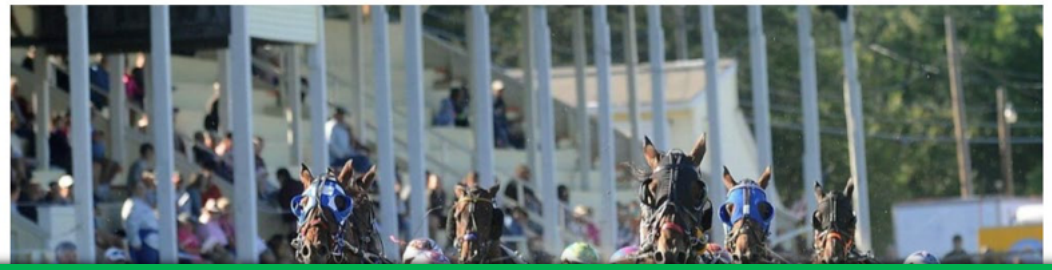
There's an ugly side to Maine harness racing



By Cici Yu · Emma Sánchez · Kendall Richards
January 29, 2023

Share

Horses have been subjected to drugging and abuse, with more than 600 violations issued by the state governing body for the sport since 2016.



EY & BUSINESS, PUBLIC SAFETY · 5 minute read

Gun manufacturers have made a dramatic exodus from Maine



By Seamus Webster · Mia McCarthy · Maya Bhat · Liam Grogan
April 16, 2023

Share

Tax breaks and a sunny political climate have convinced companies to find welcoming homes in the South.



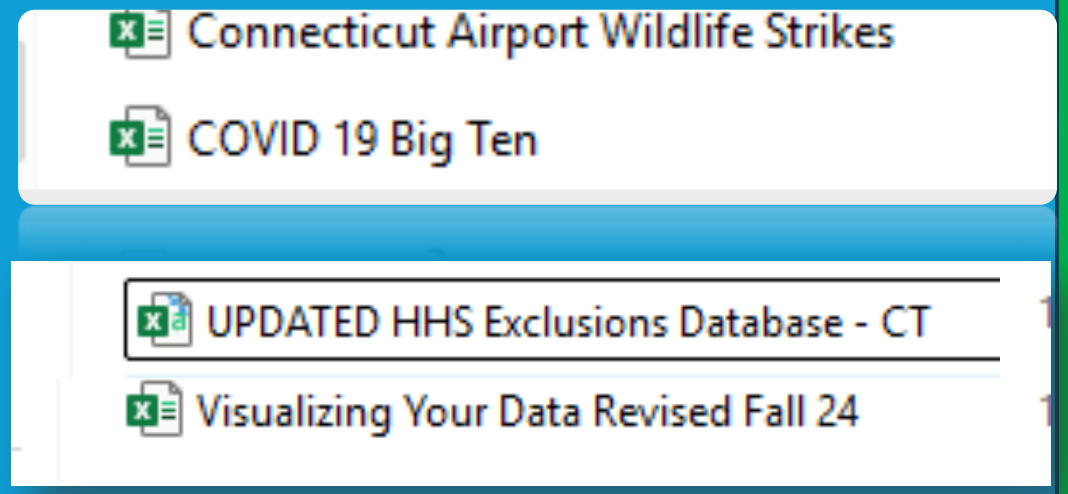
Today's Lesson – Software, Data and Documents

- Microsoft Excel
- Google Sheets
  - Tabula
- PDF to Excel
- Open Refine

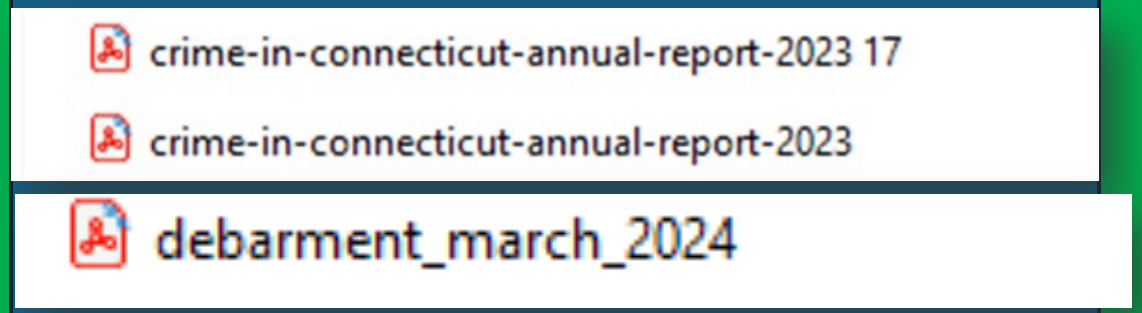


Files You Need Downloaded to Your Computer

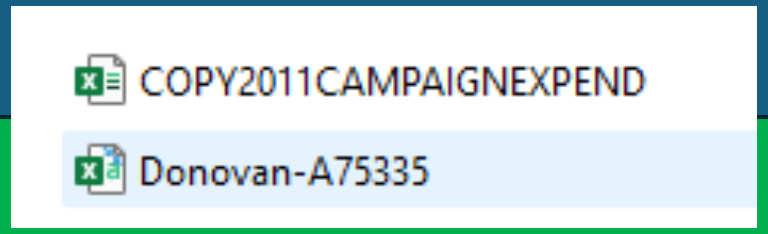
Spreadsheets



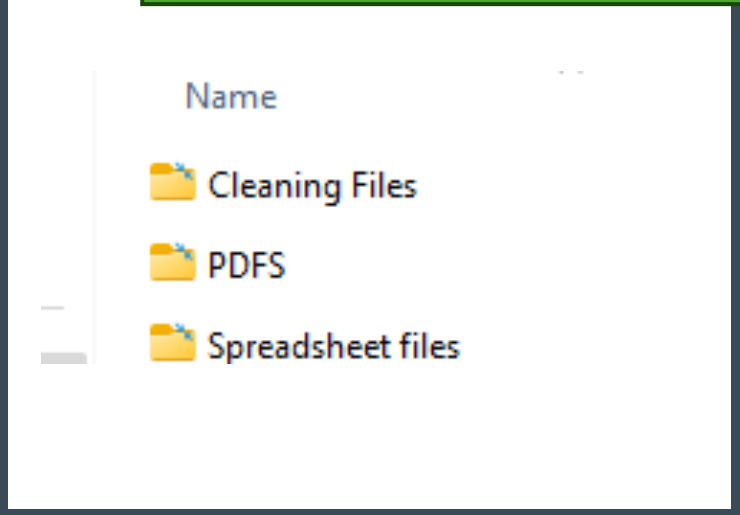
PDF Lesson

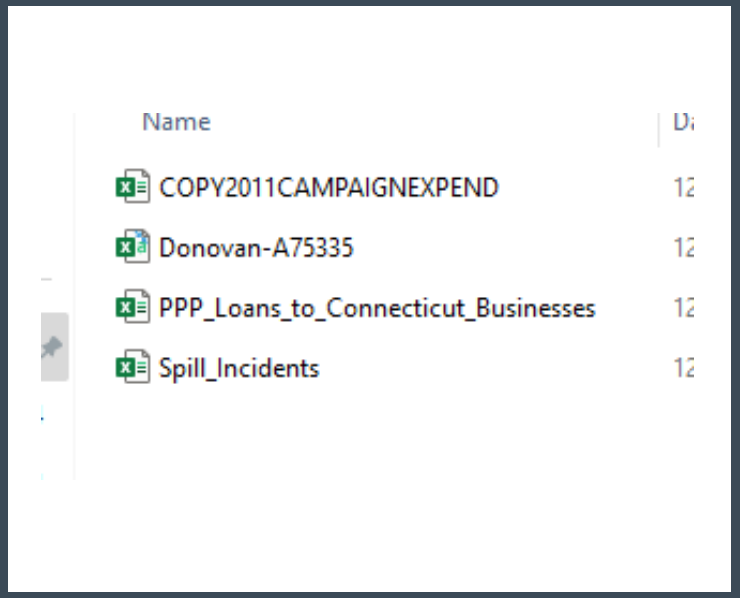


Cleaning with Open Refine

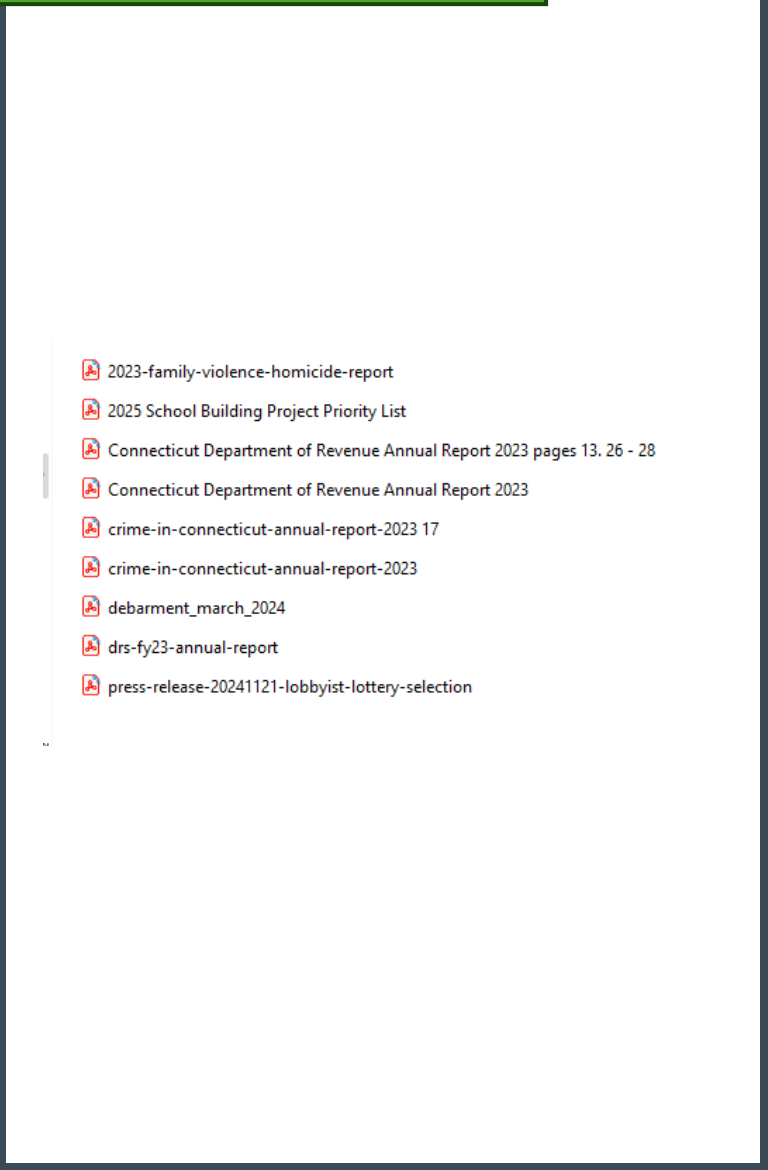


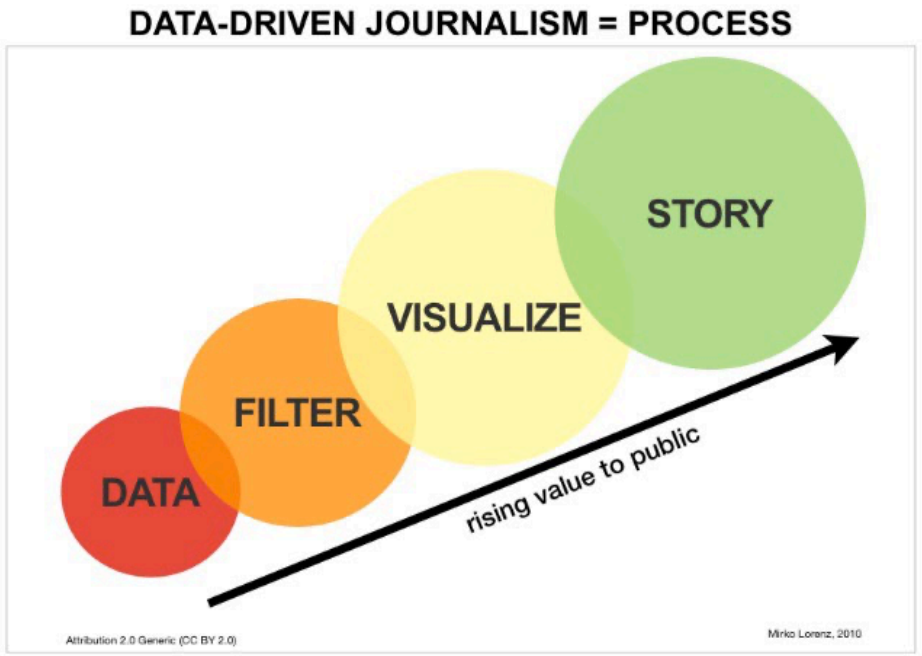
Lots of Connecticut Practice Data and Documents





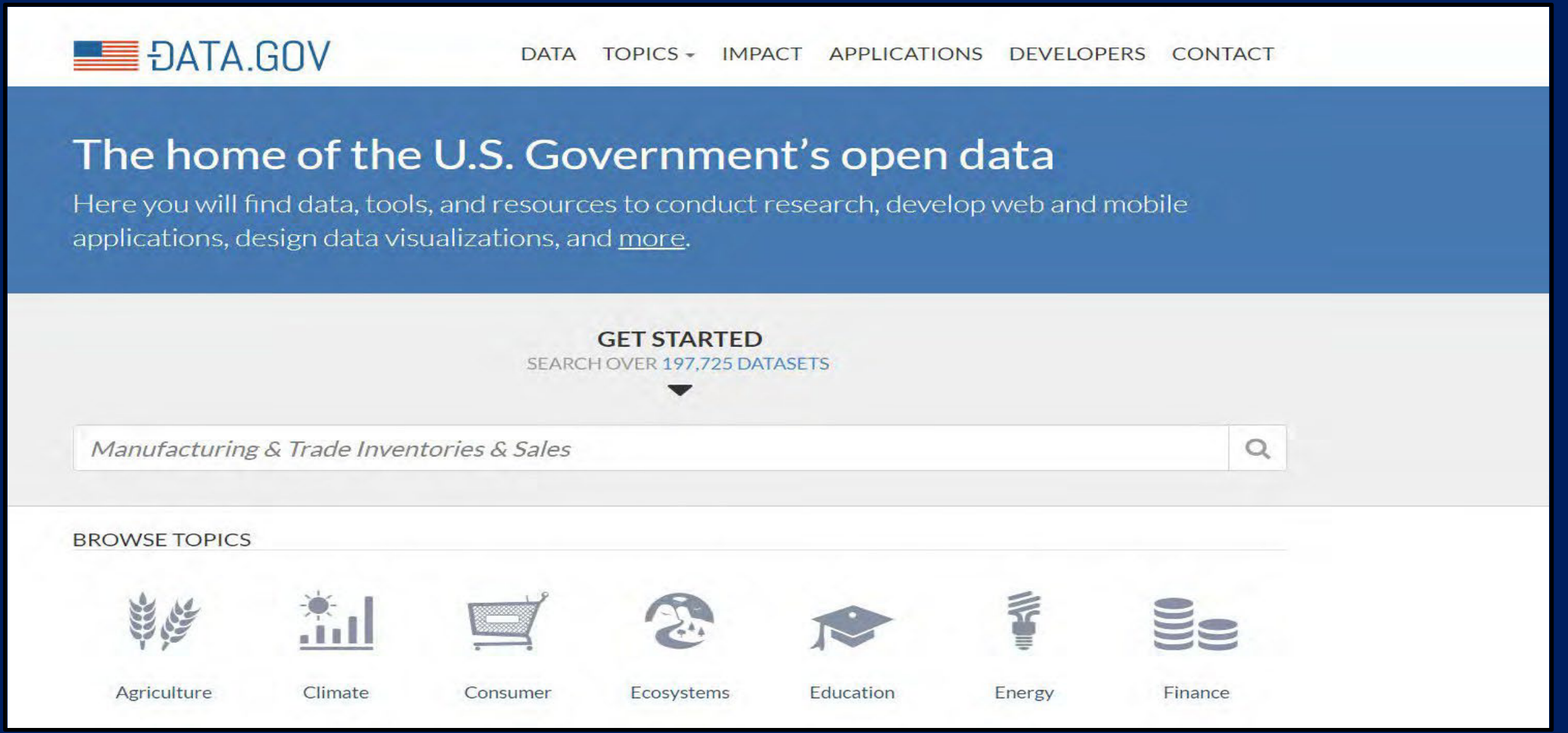








What Is Data Journalism?


DATA.GOV
DATA
TOPICS
IMPACT
APPLICATIONS
DEVELOPERS
CONTACT
The home of the U.S. Government's open data
Here you will find data, tools, and resources to conduct research, develop web and mobile applications, design data visualizations, and more.
GET STARTED
SEARCH OVER 197,725 DATASETS
Manufacturing & Trade Inventories & Sales
BROWSE TOPICS
Agriculture
Climate
Consumer
Ecosystems
Education
Energy
Finance

CT DATA

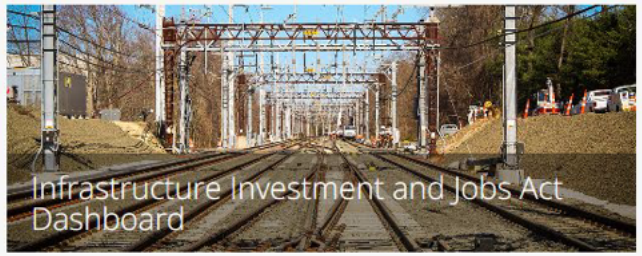
Developers Suggest a Dataset Help About Sign In

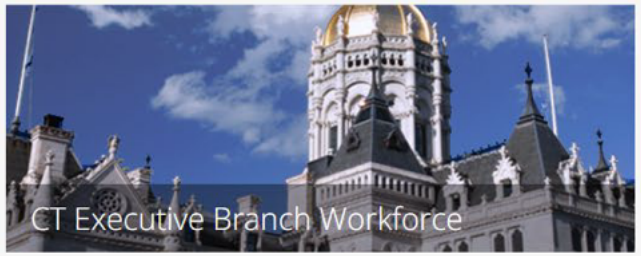
Welcome to Connecticut Open Data

Search or browse public data provided by Connecticut State Agencies

Search for Data

Featured Data Stories







Welcome to the new DocumentCloud! Questions or feedback? Use the "Feedback" button in the top right. If you need the old DocumentCloud, go to legacy.documentcloud.org -- all changes synced between sites so you can switch back and forth freely.
We'd also love your feedback at an upcoming onboarding session or stop by the #proj-documentcloud channel on News Nerdery.

Search

View

Use filters like user:, project: or organization: to refine searches. Use sort: to order results. Learn more

Documents Explore
Your Documents
Your Private Documents
Your Public Documents
Boston University's Documents

Projects Explore
Your Projects
Shared with you

Search and browse documents

Documents now load in as you scroll down. You can search by keyword or filters. Select documents to take bulk actions, like editing metadata or running an add-on.

4 of 7 ← Previous Next →

Transcribe Audio
Klaxon Cloud: Site Monitor and ...
Bad Redactions



criminal%20Calendar%20Summary Public
154 pages - Sanjin Ibrahimovic (MuckRock Staff) - Mon Dec 16 2024
Project test3



Contrato Claro PR y CEE Public
15 pages - Vanessa Colón-Almenas (Center for Investigative Journalism) - Mon Dec 16 2024



FOIA5886 A Gelzin FOIA Log (Aug-Nov 2024) - Part 2 Public
3 pages - Muckrock Staff (MuckRock Staff) - Mon Dec 16 2024
_mr_agency 3072 _mr_request 176110 _mr_jurisdiction 169



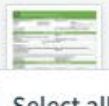
Goldsmith-NZ On Air Appointments-response letter Public
3 pages - Idiot Savant (PublicAddress.net) - Mon Dec 16 2024



Goldsmith-NZ On Air Appointments-Documents Public
89 pages - Idiot Savant (PublicAddress.net) - Mon Dec 16 2024



FOIA 5884 Patron Incident Reports - October 2024 - Part 2 Public
1 page - Muckrock Staff (MuckRock Staff) - Mon Dec 16 2024
_mr_agency 3072 _mr_request 176064 _mr_jurisdiction 169



R - 5884 - Part 2 Public
14 pages - Muckrock Staff (MuckRock Staff) - Mon Dec 16 2024

Select all Actions
Showing 25 of 5,893,096 results

Upload Documen
Share & Embed
Edit Metadata
Edit Tags & Data
Move to Project
Reprocess
Delete

2023 Commission Final Case Decisions

File No. 1 - 60 | **File No. 61 - 120** | **File No. 121 - 180**

File No. FD2023-001 - Complaint by Fernando Betancourt, Branford

File No. FD2023-002 - Referral by Registrar of Voters, Keri Rowland & Debra Fellenbaum, Milford

File No. FD2023-003 - Complaint by Linda Dalessio, Wolcott

File No. FD2023-004 - Complaint by Patricia Imperatore, Branford

File No. FD2023-005 - Complaint by Joseph A. DeLucia, Norwich

File No. FD2023-006 - Complaint by Kevin Saythany, Norwich

File No. FD2023-007 - Complaint by Eduardo Terrelba, East Haven

File No. FD2023-008 - Referral by the Audit and Disclosure Unit of the SEEC re Rogers 2018

File No. FD2023-010 - Complaint by Victor Cuevas, Waterbury

File No. FD2023-013 - Complaint by Robert Natt, Easton

File No. FD2023-015 - Complaint by Donald A. Brown Thompson & William Warner, North Grosvenordale

File No. FD2023-028 - Complaint by Mathew Hennessy, Hartford

File No. FD2023-029 - Complaint by Alan Lazowski, Hartford

File No. FD2023-030A & 2023_031B - Complaint by Silvia Ramos, Bridgeport

File No. FD2023-032 - Complaint by Debra M. Gaudette, South Windsor

File No. FD2023-033A & 2023_033B - Complaint by David Agosta, New Haven

File No. FD2023-034 - Complaint by Stephen Herbet, Thompson

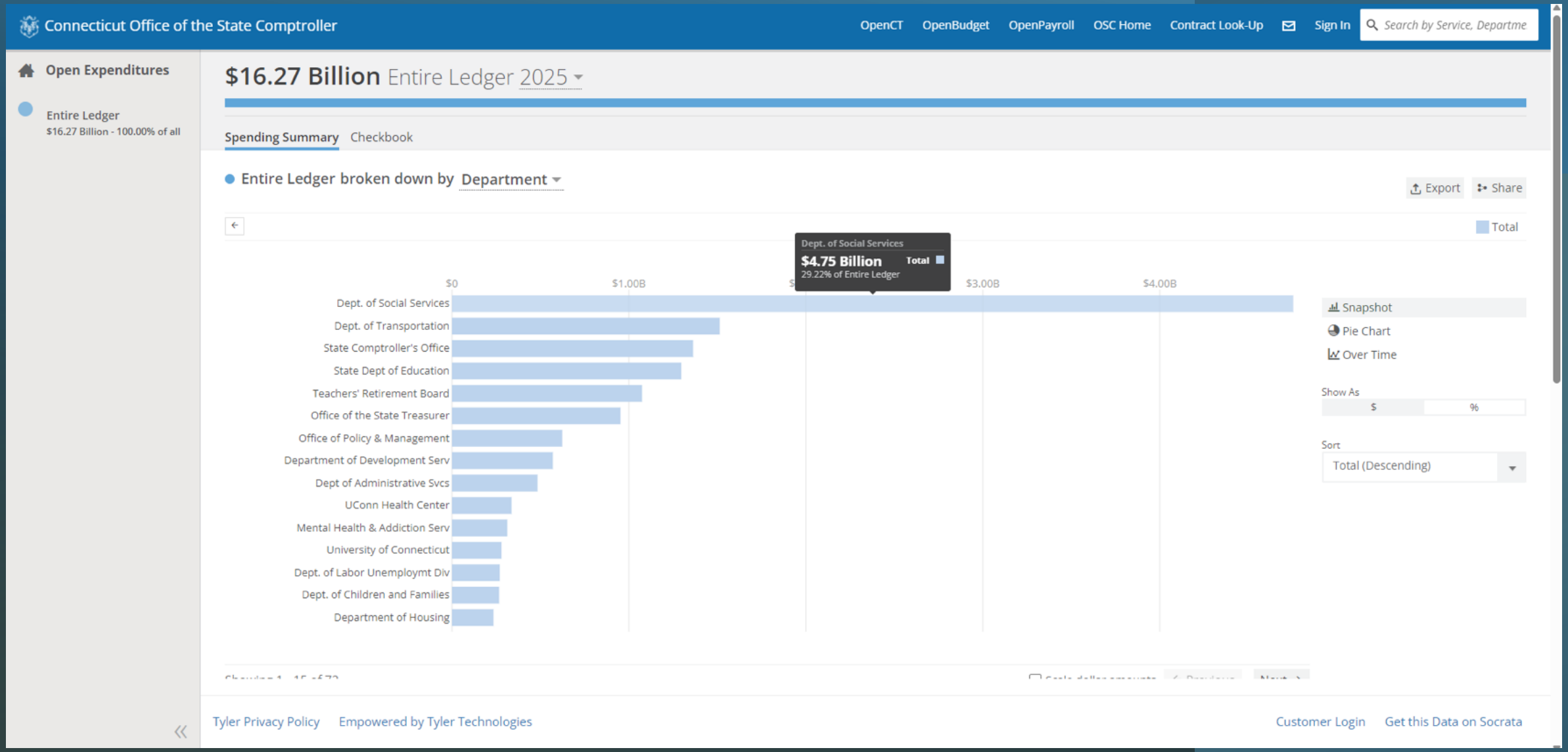
File No. FD2023-035 - Complaint by Nelson B. Stalker and Tammy A. Stalker, Sterling

File No. FD2023-036 - Complaint by Joseph D. Thompson, Bridgeport

File No. FD2023-037 - Complaint by Crystal Dailey and Walter Morton, Hamden

File No. FD2023-044 - Complaint by David G. Morin, Hartford

File No. FD2023-045 - Complaint by Michael Fallon, Middletown


Connecticut Office of the State Comptroller
OpenCT
OpenBudget
OpenPayroll
OSC Home
Contract Look-Up
Sign In
Search by Service, Departme
Open Expenditures
Entire Ledger
$16.27 Billion - 100.00% of all
$16.27 Billion Entire Ledger 2025
Spending Summary
Checkbook
Entire Ledger broken down by Department
Export
Share
Total
Dept. of Social Services
$4.75 Billion
Total
29.22% of Entire Ledger
$0
$1.00B
$3.00B
$4.00B
Dept. of Social Services
Dept. of Transportation
State Comptroller's Office
State Dept of Education
Teachers' Retirement Board
Office of the State Treasurer
Office of Policy & Management
Department of Development Serv
Dept of Administrative Svcs
UConn Health Center
Mental Health & Addiction Serv
University of Connecticut
Dept. of Labor Unemploymt Div
Dept. of Children and Families
Department of Housing
Snapshot
Pie Chart
Over Time
Show As
$
%
Sort
Total (Descending)
Tyler Privacy Policy
Empowered by Tyler Technologies
Customer Login
Get this Data on Socrata



Connecticut State
Department of Revenue Services

Upcoming CT DRS webinar: **Select** to register for the upcoming **Withholding Forms W-2 and 1099 Annual Filing Webinar** on Wednesday, January 15, 2025, at 10:00 a.m.

- Featured Items
- For Businesses
- For Individuals
- Practitioners
- Forms
- Publications
- Research Library

Search Department of Revenue Services

by Keyword

90 Day Delinquency List

Connecticut General Statutes Section 12-7a requires the Commissioner of Revenue Services to prepare and maintain a list of delinquent state taxes unpaid for a period greater than 90 days after all appeal rights have expired. That list, including the names, addresses, tax type and amount due (tax, interest and penalties), must be available to the public for inspection by any person. The complete list of these delinquent taxpayers is available by sending an email to **DRS@ct.gov** requesting the 90 Day Delinquency List. DRS will reply with an electronic version of the list.

About

I'm an editor overseeing data and investigative projects for Hearst Connecticut Media and the Times Union in Albany. I'm also an adjunct professor teaching investigative reporting techniques at Columbia University's Graduate School of Journalism.

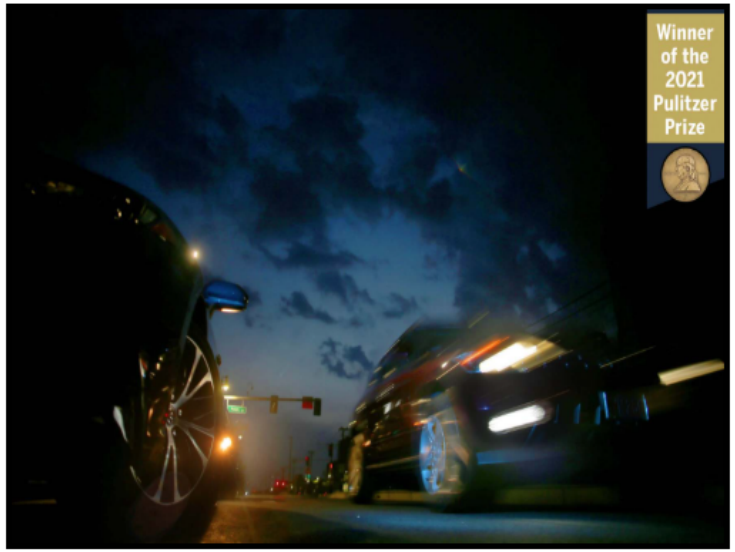
I previously worked as a Spotlight reporter for the Boston Globe, where I won the 2021 Pulitzer Prize in Investigative Reporting.



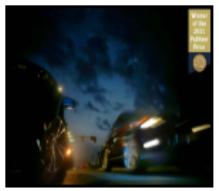
At the Globe, I specialized in using data and public records for both long- and short-term investigations. I covered beats on criminal justice, "things that work," higher education, Boston neighborhoods, business, and

| A SELECTION OF RECENT STORIES |

America's blind spot



Across the country, dangerous drivers should have lost their licenses. They didn't — and others died. Full Story (Photo: Lane Turner/Globe Staff)







Search

Bookmark and Share

Posts from @mrochele

Nothing to see here - yet

When they post, their posts will show up here.

View on X

Amusing articles



A Data Journalist's Best Tool: The Spreadsheet: Filtering, Pivot Tables, IF Statements and Conditional Formatting

Spreadsheets= Structured Data

The vertical sections are "**columns**," and they're referenced with letters.

The horizontal sections are "**rows**," and they're referenced with numbers.

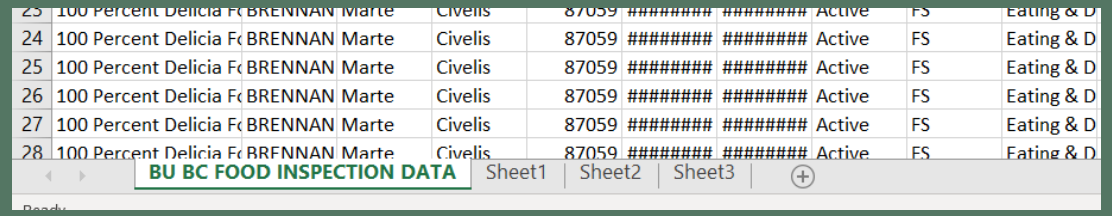
Copy of ADMIN LEAVE DATA 2014 - PRESENT (1) - Compatibility Mode | ADMIN LEAVE DATA 2014 - PRESENT (1) - Read-Only -

	A	B	C	D	E	F	G
1	ID	Name	Dept ID	Union Code	Rpt Dt	Descr	Hours
2	010608	Wright,Lysander E	211000	PDB	01/01/14	Administrative Leave Paid	8.000000
3	011047	Parker,Henderson O	211000	BPP	01/01/14	Administrative Leave Paid	8.000000
4	012276	Pierre-Thomas,Joslin A	211000	BPP	01/01/14	Administrative Leave Paid	8.000000
5	097067	Hendricks,Sharon A.	211000	SU4	01/01/14	Administrative Leave Paid	8.000000
6	103529	Ricci,Helen	211000	BPP	01/01/14	Administrative Leave Paid	8.000000
7	106710	Fonseca,Sandro	211000	BPP	01/01/14	Administrative Leave Paid	8.000000
8	108900	Li,Wei	211000	SU4	01/01/14	Administrative Leave Paid	8.000000
9	009443	Cohen,Mark A	211000	EXM	01/02/14	Administrative Leave Paid	7.000000
10	010139	Brown,Leonard F	211000	BPP	01/02/14	Administrative Leave Paid	8.000000
11	010286	Marzano,Frederick G	211000	BPP	01/02/14	Administrative Leave Paid	8.000000
12	086147	Donovan,Patrick J	211000	BPP	01/02/14	Administrative Leave Paid	8.000000
13	097067	Hendricks,Sharon A.	211000	SU4	01/02/14	Administrative Leave Paid	8.000000
14	100375	Harty,Janice M	211000	SU4	01/02/14	Administrative Leave Paid	8.000000
15	103529	Ricci,Helen	211000	BPP	01/02/14	Administrative Leave Paid	8.000000
16	106710	Fonseca,Sandro	211000	BPP	01/02/14	Administrative Leave Paid	8.000000
17	108900	Li,Wei	211000	SU4	01/02/14	Administrative Leave Paid	8.000000
18	009443	Cohen,Mark A	211000	EXM	01/03/14	Administrative Leave Paid	7.000000
19	010139	Brown,Leonard F	211000	BPP	01/03/14	Administrative Leave Paid	8.000000
20	010286	Marzano,Frederick G	211000	BPP	01/03/14	Administrative Leave Paid	8.000000
21	086147	Donovan,Patrick J	211000	BPP	01/03/14	Administrative Leave Paid	8.000000
22	097067	Hendricks,Sharon A.	211000	SU4	01/03/14	Administrative Leave Paid	8.000000
23	100375	Harty,Janice M	211000	SU4	01/03/14	Administrative Leave Paid	8.000000
24	103529	Ricci,Helen	211000	BPP	01/03/14	Administrative Leave Paid	8.000000
25	106710	Fonseca,Sandro	211000	BPP	01/03/14	Administrative Leave Paid	8.000000
26	108900	Li,Wei	211000	SU4	01/03/14	Administrative Leave Paid	8.000000
27	010139	Brown,Leonard F	211000	BPP	01/04/14	Administrative Leave Paid	8.000000
28	010286	Marzano,Frederick G	211000	BPP	01/04/14	Administrative Leave Paid	8.000000
29	010608	Wright,Lysander E	211000	PDB	01/04/14	Administrative Leave Paid	8.000000
30	011047	Parker,Henderson O	211000	BPP	01/04/14	Administrative Leave Paid	8.000000

01-01-14 to 06-10-20 | 06-11-20 to present

What Is A Spreadsheet?

Columns and Rows Together = Spreadsheet



24	100 Percent Delicia Fc	BRENNAN	Marte	Civelis	87059	########	########	Active	FS	Eating & D
25	100 Percent Delicia Fc	BRENNAN	Marte	Civelis	87059	########	########	Active	FS	Eating & D
26	100 Percent Delicia Fc	BRENNAN	Marte	Civelis	87059	########	########	Active	FS	Eating & D
27	100 Percent Delicia Fc	BRENNAN	Marte	Civelis	87059	########	########	Active	FS	Eating & D

BU BC FOOD INSPECTION DATA | Sheet1 | Sheet2 | Sheet3

Multiple Sheets Can Make Up A Workbook



	violstatus	statusdate	comments
2	Fail		Provide glass storage rack.
3	Fail		Provide dish rack over 3 bay sink to replace missing drainboards at each end of 3 bay sink.
4	Fail		Provide dish washer at ware washing area.
5	Fail		Provide mop sink at basement.
6	Fail		No hand sink at Bar Area provide.
7	Fail		Provide door and finish miscellaneous storage room.
8	Fail		Provide dish washer at ware washing area.
9	Fail		Provide chefs thermometer for monitoring food temperatures.
10	Fail		Replace heavily scorded cutting board.
11	Fail		Dishmachine not operating repair.
12	Fail		Provide sink plugs for sink behind service bar.
13	Fail		Provide dispenser for paper towels.
14	Fail		Exhausthood sticker not visible provide current documentation on exhausthood cleaning.
15	Fail		Repair loose pipe below hand sink behind beverage bar.
16	Fail		Replace missing floor tiles at entrance to kitchen.
17	Fail		Paint wall for storage.
18	Fail		Observed raw chicken stored over ready to eat product.Provide proper storage of raw and ready to eat products.
19	Fail		Closet next to microwave with pots stored on the floor.Discontinue.Store clean pots at least 6 inches off the floor
20	Fail		Observed buckets of food products stored directly on the floor.Maintain storage at least 6 inches of the floor.
21	Fail		Provide working internal thermometers for all refrigeration.
22	Fail		Provide chlorine test kit for 3 bay sink
23	Fail		Provide proper sanitizer solution to store in use wiping cloths.
24	Fail		
25	Fail		Hand sink leaking.PIC states it is in the process of being fixed at the time of inspection.
26	Fail		Provide new battery for faucet water trickling out.
27	Fail		Provide all handsinks with soap and paper towels at all times for proper handwashing.
28	Fail		Provide covered trash receptacle

BU BC FOOD INSPECTION DATA | Sheet1 | Sheet2 | Sheet3

Ready

Why Spreadsheets?



- A spreadsheet can be a calculator or database
- Can import, analyze and visualize large amounts of data
  - Statistical data
  - Text
  - Dates and times
- Do calculations in seconds on vast amounts of financial information
- Allows you to drill down deep into the data to look for story leads and angles

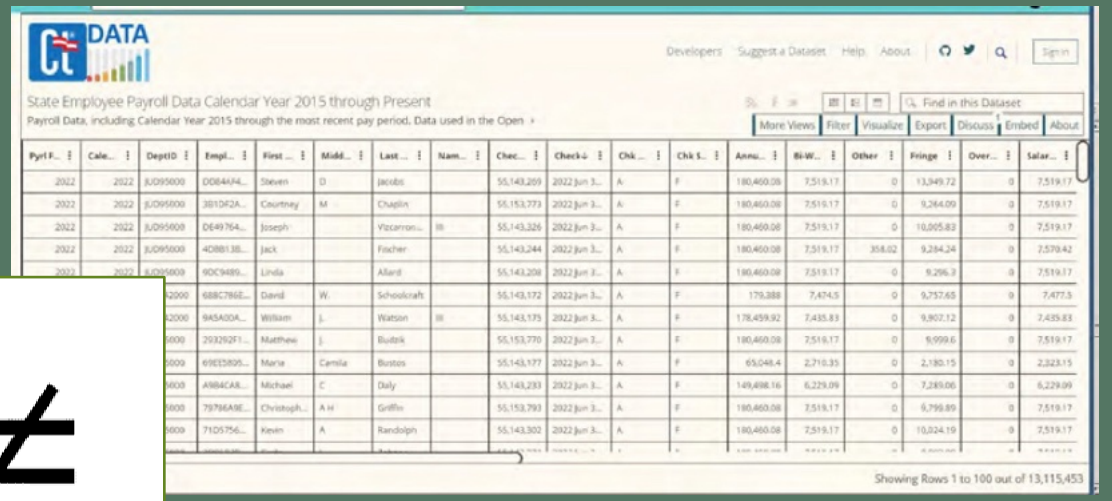
Spreadsheet Limitations

Missing People in Connecticut



≠

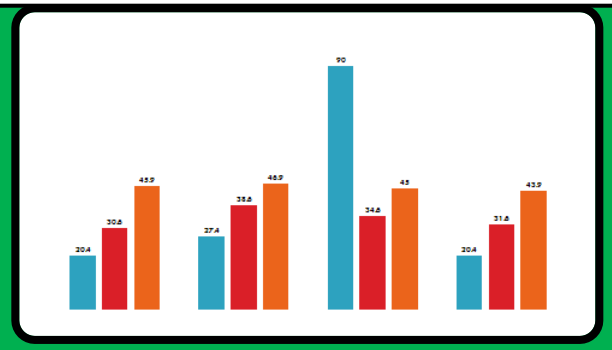
Connecticut State Payroll

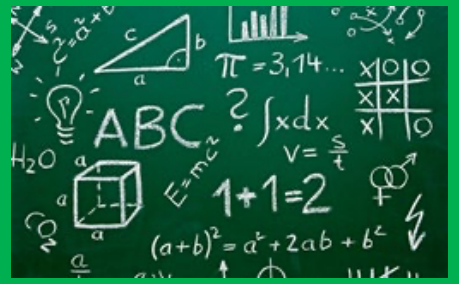


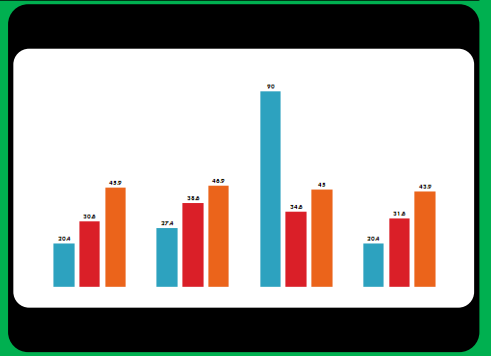
Why Numbers Matter

- Context for stories
- Must always ask compared to what?

- A number standing alone means nothing
- Is this a lot or a little or average?







Always Ask – What Does This Number Mean?



Demonstrators at a July rally for criminal justice reform. BANNER PHOTO

Boston Police officers have contributed more than $1 million to Boston elected officials over the last 10 years, with the lion's share of the dollars — $456,548 — going to Mayor Martin Walsh, according to an analysis released by the political organizing group Mijente and ACLU of Massachusetts this week.

The data, analyzed by Boston University's BU Spark! data project, show the majority of campaign contributions to city councilors going to a handful members with at-large Councilor Michael Flaherty receiving $38,700 since his re-election to the council in 2013 and at-large Councilor Annissa Essaibi-George receiving $26,700.

Councilors Frank Baker, Matt O'Malley and Ed Flynn all received more than any of the black councilors, including Council President Kim Janey, who netted $6,100 from police officers.

The donations from police represent a small sliver of overall donations to candidates.

Donors associated with the real estate industry, including real estate brokers, developers, architects, attorneys and building trades members have long dominated political donations in cities across the United States.

Some Key Numbers to Pay Attention To

Difference

Percent difference

Percent of total

Sum

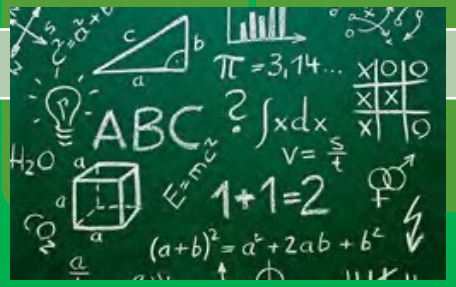
Average

Rate

Median

Mean

Ratio



An official website of the United States government Here's how you know

Overview

Wildlife Strike Database

Wildlife Strike FAQs

Wildlife Strike Gallery

Wildlife Management

Wildlife Strike Research & Development

Wildlife Regulations, Guidance, & Resources

Feather Identification Lab

Related Links

Wildlife Hazard Mitigation

During the past century, wildlife-aircraft strikes have resulted in the loss of hundreds of lives worldwide, as well as billions of dollars in aircraft damage. The FAA maintains a comprehensive program to address wildlife hazards. Through policy and guidance, research and outreach, we strive to stay ahead of the issue.

Hot Items

- **August 19, 2024:** Advisory Circular 5200-32C, Reporting Wildlife Aircraft Strikes
- **June 5, 2024:** Wildlife Strikes to Civil Aircraft in the United States, 1990 - 2023
- **August 20-22, 2024:** Aviation Wildlife Management Conference
- Receive email alerts when this page is updated.

As of November 27, 2024 current strike validation progress is as follows:

- 5200-7 processed through October 15, 2024
- 5200-7E processed through August 15, 2024
- Mandatory Occurrence Reports (MOR) processed through May 6, 2024

If you have any issues submitting strikes via this new interface, please contact matthew.nelson@ees-net.com. For those using the automated 'One-Stop' Reporting function (i.e. Web Service), please contact wesley.major@faa.gov to ensure that your system is compatible with this new wildlife strike reporting software or to inquire if this service is right for your organization.

FAA Wildlife Strike Database

AD

The FAA Wildlife Strike Database contains records of reported wildlife strikes since 1990. Strike reporting is voluntary. Therefore, this database only represents the information we have received from airlines, airports, pilots, Mandatory Occurrence Reports (MOR), incident/accident information, and other sources.

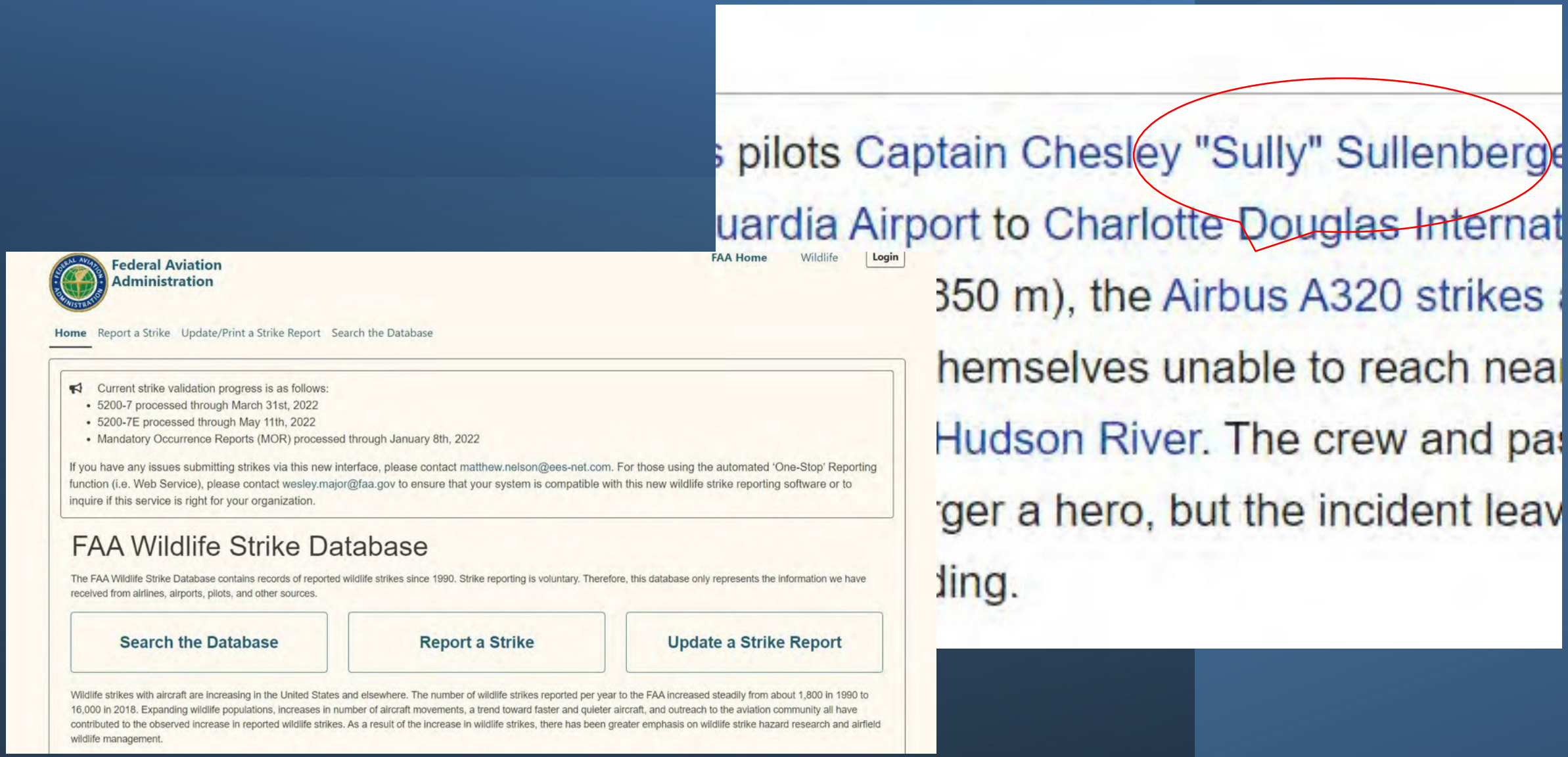
Search the Database **Report a Strike** **Update a Strike Report**

Wildlife strikes with aircraft are increasing in the United States and elsewhere. About 291,600 wildlife strikes with civil aircraft were reported in USA between 1990 and 2023 (about 19,400 strikes at 713 U.S. airports in 2023). An additional 5,100 strikes were reported by U.S. Air Carriers at foreign airports, 1990–2023 (about 236 strikes at 92 airports in 55 countries in 2023). Expanding wildlife populations, increases in number of aircraft movements, a trend toward faster and quieter aircraft, and outreach to the aviation community all have contributed to the observed increase in reported wildlife strikes. As a result of the increase in wildlife strikes, there has been greater emphasis on wildlife strike hazard research and airfield wildlife management.

For over three decades, the FAA and USDA have conducted a concerted effort to collect accurate data on wildlife strikes to better understand the scope and nature of the problem. These strike data provide a scientific foundation for management programs to mitigate risk. A major part of collecting the data is providing the general public with an easy way to submit strike reports in a consistent format. This continues to be accomplished through the use of the FAA's Wildlife Hazard Mitigation website (https://wildlife.faa.gov/) and the FAA's Wildlife Strike Database.

The FAA intends to expand its research and development activities in the area of wildlife hazard and risk mitigation on and near airports. Wildlife hazard mitigation is a broad discipline incorporating aspects of one or more of four research areas:

1. Habitat management



s pilots Captain Chesley "Sully" Sullenberge
uardia Airport to Charlotte Douglas Internat
350 m), the Airbus A320 strikes
hemselves unable to reach nea
Hudson River. The crew and pa
ger a hero, but the incident leav
ding.

Federal Aviation Administration

FAA Home Wildlife Login

Home Report a Strike Update/Print a Strike Report Search the Database

Current strike validation progress is as follows:
- 5200-7 processed through March 31st, 2022
- 5200-7E processed through May 11th, 2022
- Mandatory Occurrence Reports (MOR) processed through January 8th, 2022

If you have any issues submitting strikes via this new interface, please contact matthew.nelson@ees-net.com. For those using the automated 'One-Stop' Reporting function (i.e. Web Service), please contact wesley.major@faa.gov to ensure that your system is compatible with this new wildlife strike reporting software or to inquire if this service is right for your organization.

FAA Wildlife Strike Database

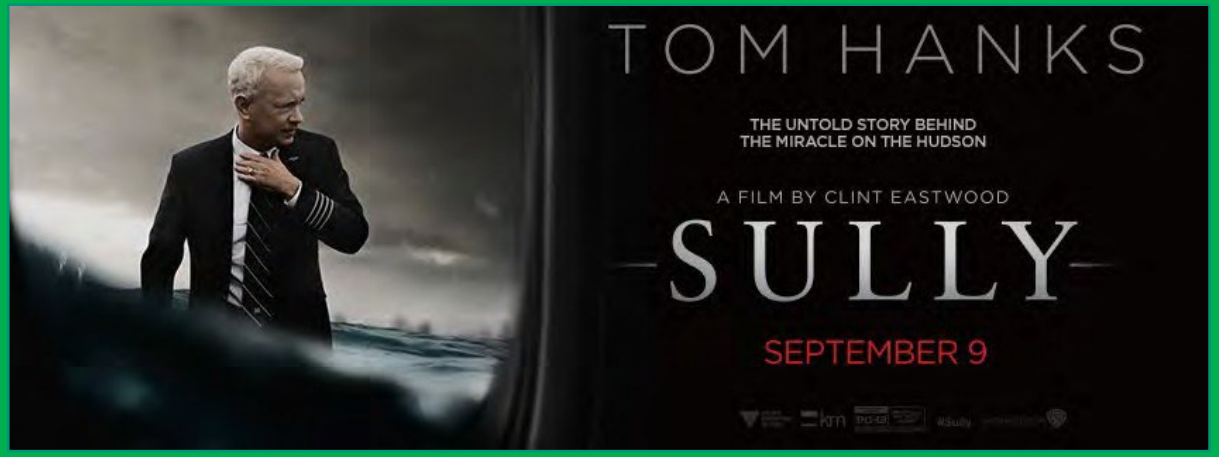
The FAA Wildlife Strike Database contains records of reported wildlife strikes since 1990. Strike reporting is voluntary. Therefore, this database only represents the information we have received from airlines, airports, pilots, and other sources.

Search the Database | Report a Strike | Update a Strike Report

Wildlife strikes with aircraft are increasing in the United States and elsewhere. The number of wildlife strikes reported per year to the FAA increased steadily from about 1,800 in 1990 to 16,000 in 2018. Expanding wildlife populations, increases in number of aircraft movements, a trend toward faster and quieter aircraft, and outreach to the aviation community all have contributed to the observed increase in reported wildlife strikes. As a result of the increase in wildlife strikes, there has been greater emphasis on wildlife strike hazard research and airfield wildlife management.


SULLY




TOM HANKS
THE UNTOLD STORY BEHIND
THE MIRACLE ON THE HUDSON
A FILM BY CLINT EASTWOOD
SULLY
SEPTEMBER 9



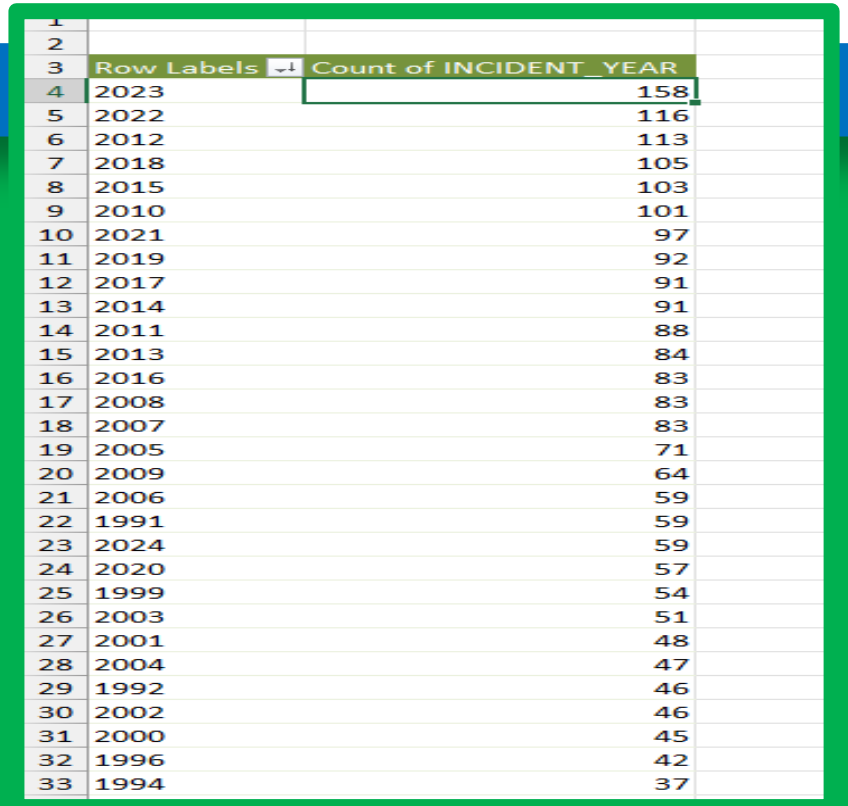
NBC New York + Follow 17.3K Followers

Caught on camera: Bird strike at LaGuardia forces emergency landing

Story by NBC New York Staff • 2h • 1 min read



Wildlife strikes with aircraft are increasing in the United States and elsewhere. About 272,000 wildlife strikes with civil aircraft were reported in USA between 1990 and 2022 (about 17,000 strikes at 693 U.S. airports in 2022). An additional 4,800 strikes were reported by U.S. Air Carriers at foreign airports, 1990–2022 (about 230 strikes at 91 airports in 53 countries in 2022). Expanding wildlife populations, increases in number of aircraft movements, a trend toward faster and quieter aircraft, and outreach to the aviation community all have contributed to the observed increase in reported wildlife strikes. As a result of the increase in wildlife strikes, there has been greater emphasis on wildlife strike hazard research and airfield wildlife management.



Row Labels	Count of INCIDENT_YEAR
2023	158
2022	116
2012	113
2018	105
2015	103
2010	101
2021	97
2019	92
2017	91
2014	91
2011	88
2013	84
2016	83
2008	83
2007	83
2005	71
2009	64
2006	59
1991	59
2024	59
2020	57
1999	54
2003	51
2001	48
2004	47
1992	46
2002	46
2000	45
1996	42
1994	37

Hear how JFK air traffic controllers guided a Chinese cargo plane back to safety after an engine failed due to a bird strike

Story by psyme@insider.com (Pete Syme) • 9mo • 2 min read





Birds and Wildlife Pests Pose Numerous Aircraft Hazards

An engineer's perspective on how birds and wildlife pests can damage aircraft.



Subscribe to our Newsletters

Enter your email for an upgraded inbox!

Email Submit

Easy fleet management with FieldRoutes

CNN travel Destinations Food & Drink News Stay Video

This is what happens when a plane collides with a bird

THE CONVERSATION By Doug Drury
3 minute read · Updated 10:34 AM EDT, Thu June 27, 2024
9 comments



A flock of birds flies near a plane as it lands at Reagan National Airport in Washington, DC, in a 2015 file photo. Andrew Caballero-Reynolds/AFP/Getty Images/FILE

SCIENCE AND TECH › ANIMALS AND INSECTS

4-legged robot to help deter wildlife strikes at Alaska airport

In 2023, 93 wildlife-aircraft strikes at Alaska airports were reported to the Federal Aviation Administration.



SCRIPPS NEWS Life

Covering the stories that inspire and impact everyday lives.

LOCAL

Blue Angels jet makes safe landing in Traverse City after bird strike

ADVERTISEMENT

Megan Huiskens



BLUE ANGELS BIRD INCIDENT

TRAVERSE CITY — Fat Albert, the aircraft that is responsible for carrying the team's equipment and

NEWSLETTER SIGNUP

Stay up to date on important stories from 9&10 News delivered free to your inbox.

Email Address
First Name
Last Name
Zip Code
SUBSCRIBE

Multiple bird strike in flair caused rough landing.

ONIAN, FAA 2630.

PILOT MADE A GO AROUND.

Aborted Take-off

Aborted Take-off

IT APPEARED OUT OF NOWHEERE AS IF IT WAS DIVING DOWN

INSTRUCTED TO LAND ON ALTERNATE RWY. A BLACK BEAR CLIMBED OVER THE PERIMETER FENCE OF THE AIRPORT

While entering the runway, the crew reported a coyote on the runway in front of them.

NEGATIVE EFFECT ON FLIGHT. Aircraft was on final approach and had to abort landing due to coyote on the runwa

EFFECT ON FLIGHT. DEER WAS RUNNING DOWN RWY AND AIRCRAFT HAD TO GO AROUND.

AD A NEGATIVE EFFECT ON FLIGHT. Aircraft was on final approach and had to abort landing due to coyote on the runwa

. Animal was runover by aircraft and remains were scattered around 50 foot area.

NEGATIVE EFFECT ON FLIGHT. Pilot decided to go-around on final approach after he was warned that a Fox was heading out towards the ru

NEGATIVE EFFECT ON FLIGHT. Aircraft performed a go around during the left downwind as per ATC instruction and preca

BIRD STRIKE. CREW ADVISED Y90 APPROACH THAT THEY HAD NOSE CONE DAMAGE

BIRD HIT WINDSHLD. NOT MUCH LEFT OF BIRD.

C DID A GO AROUND AND UPON LANDING LOST DIRECTIONAL CONTROL UPON EXITING

ABORTED LANDING AND LOST DIRECTIONAL CONTROL ON TAXI

Do Not Analyze Data Until You . . . .

1. Do a “four corners” check of your dataset to see what you should get

2. Make sure you have a data dictionary or record layout explaining the data

3. Always make a copy of your data

Take 3 Important Steps

Wildlife Database Reference Tables

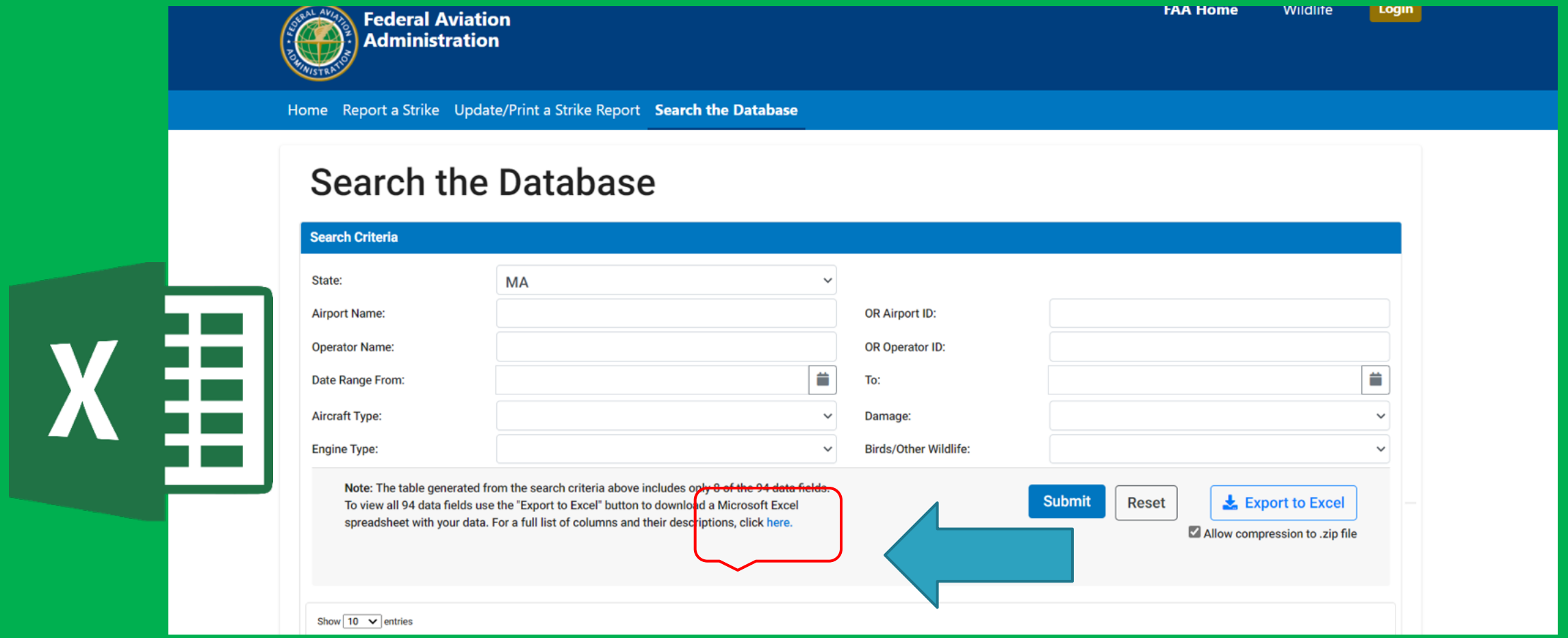
Engine Type	
A	Reciprocating
B	Turbojet
C	Turboprop
D	Turbofan
E	None (glider)
F	Turboshaft (helicopter)
Y	Other

Damage (Civil)	
N	None
M	Minor
M?	Uncertain
S	Substantial
D	Destroyed

Damage (Military)	
Class A	Over $2,000,000
Class B	$500,000 - $2,000,000
Class C	$50,000 - $499,999
Class D	$20,000 - $49,999
Class E	Less than $20,000



Make Sure To Get The Data Dictionary



Column Name	Explanation of Column Name and Codes
INDEX NR	Individual record number
OPID	Airline operator code
OPERATOR	A three letter International Civil Aviation Organization code for aircraft operators. (BUS = business, PVT = private aircraft other than business, GOV = government aircraft, MIL - military aircraft.)
ATYPE	Aircraft
AMA	International Civil Aviation Organization code for Aircraft Make
AMO	International Civil Aviation Organization code for Aircraft Model
EMA	Engine Make Code (see Engine Codes tab below)
EMO	Engine Model Code (see Engine Codes tab below)
AC_CLASS	Type of aircraft (see Aircraft Type tab below)
AC_MASS	1 = 2,250 kg or less: 2 = ,2251-5700 kg: 3 = 5,701-27,000 kg: 4 = 27,001-272,000 kg: 5 = above 272,000 kg
NUM_ENGS	Number of engines
TYPE_ENG	Turboshaft (helicopter): Y = Other
ENG_1_POS	Where engine # 1 is mounted on aircraft (see Engine Position tab below)
ENG_2_POS	Where engine # 2 is mounted on aircraft (see Engine Position tab below)
ENG_3_POS	Where engine # 3 is mounted on aircraft (see Engine Position tab below)
ENG_4_POS	Where engine # 4 is mounted on aircraft (see Engine Position tab below)
REG	Aircraft registration
FLT	Flight number
REMAINS_COLLECTED	Indicates if bird or wildlife remains were found and collected
REMAINS_SENT	Indicates if remains were sent to the Smithsonian Institution for identifcation
INCIDENT_DATE	Date strike occurred
INCIDENT_MONTH	Month strike occurred
INCIDENT_YEAR	Year strike occurred
TIME_OF_DAY	Light conditions
TIME	Hour and minute in local time
AIRPORT_ID	International Civil Aviation Organization airport identifier for location of strike whether it was on or off airport
AIRPORT	Name of airport
STATE	State
FAAREGION	FAA Region where airport is located
ENROUTE	If strike did not occur on approach, climb, landing roll, taxi or take-off, aircraft was enroute. This shows location.
RUNWAY	Runway
LOCATION	Various information about aircraft location if enroute or airport where strike evidence was found. Some locations show the two airports for the flight departure and arrival if pilot was unaware of the strike.
HEIGHT	Feet Above Ground Level
SPEED	Knots (indicated air speed)
DISTANCE	Nautical miles from airport
PHASE_OF_FLT	Phase of flight during which strike occurred
DAMAGE	
Blank	Unknown
N = none	No damage was reported.
M = minor	When the aircraft can be rendered airworthy by simple repairs or replacements and an extensive inspection is not necessary.
M? = uncertain level	The aircraft was damaged, but details as to the extent of the damage are lacking.
S = substantial	When the aircraft incurs damage or structural failure which adversely affects the structure strength, performance or flight characteristics of the aircraft and which would normally require major repair or replacement of the affected component

Open The **Connecticut Airport Wildlife Strikes** Data

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y
1	INDX_NR	INCIDENT_	INCIDENT_	INCIDENT_	TIME	TIME_OF_I	AIRPORT_I	AIRPORT	AIRPORT_I	AIRPORT_I	RUNWAY	STATE	FAAREGIO	LOCATION	ENROUTE_	OPID	OPERATOF	REG	FLT	AIRCRAFT	AMA	AMO	EMA	EMO	AC_CLASS
2	1574725	2024-08-2	8	2024	10:40	Day	KBDR	IGOR I SIK	41.16348	-73.12618		CT	ANE			BUS	BUSINESS	N7667S	75332	SIKORSKY	813	14	43	01	B
3	1577910	2024-08-2	8	2024	11:00	Day	KGON	GROTON-N	41.33006	-72.04514	23	CT	ANE			BUS	BUSINESS	N757AD		C-152	226	36	07	07	A
4	1574326	2024-08-2	8	2024	23:32	Night	KBDL	BRADLEY I	41.93887	-72.68323	24	CT	ANE			UAL	UNITED AI	N47570	2254	B-737-9	148		10		A
5	1576858	2024-08-2	8	2024	21:12	Night	ZZZZ	UNKNOWN						14 MILES K	CT	PVT	PRIVATELY	N21966		C-150	226	05	13	01	A
6	1576072	2024-08-1	8	2024	13:30	Day	KBDR	IGOR I SIK	41.16348	-73.12618	6	CT	ANE			BUS	BUSINESS	N6JS		CIRRUS SR		01	13	20	A
7	1575118	2024-08-1	8	2024	18:55	Day	KBDL	BRADLEY I	41.93887	-72.68323	24	CT	ANE			JIA	PSA AIRLIN	N537EA	5310	CRJ700	188	16	22	04	A
8	1575035	2024-08-1	8	2024	08:01	Dawn	KHVN	TWEED-NE	41.2639	-72.8871	20	CT	ANE			VXP	AVELO AIR	N707VL	401	B-737-7	148		10	01	A
9	1574928	2024-08-0	8	2024	10:35	Day	KBDL	BRADLEY I	41.93887	-72.68323	6	CT	ANE			JBU	JETBLUE A	N554JB	1142	A-320	04A	03	23	01	A
10	1574268	2024-08-0	8	2024	14:40	Day	KHFD	HARTFORI	41.73627	-72.65021	2	CT	ANE			BUS	BUSINESS	N201EM		MOONEY-	632	10	07	19	A
11	1574463	2024-08-0	8	2024	17:41	Day	KGON	GROTON-N	41.33006	-72.04514	05	CT	ANE			BUS	BUSINESS	N599HA		CL-300	188		02		A
12	1574338	2024-08-0	8	2024	09:27	Day	KHVN	TWEED-NE	41.2639	-72.8871	2	CT	ANE			VXP	AVELO AIR	N802XT	678	B-737-8	148		10	01	A
13	1574967	2024-08-0	8	2024			KBDL	BRADLEY I	41.93887	-72.68323	15/33	CT	ANE			UNK	UNKNOWN			UNKNOWN					
14	1574925	2024-08-0	8	2024			KBDL	BRADLEY I	41.93887	-72.68323	15/33	CT	ANE			UNK	UNKNOWN			UNKNOWN					
15	1573863	2024-08-0	8	2024	08:45	Day	KHVN	TWEED-NE	41.2639	-72.8871	20	CT	ANE			VXP	AVELO AIR	N705VL	404	B-737-700	148	42	10	01	A
16	1574422	2024-08-0	8	2024	22:57	Night	KBDL	BRADLEY I	41.93887	-72.68323	24	CT	ANE			JBU	JETBLUE A	N598JB	460	A-320	04A	03	23	01	A
17	1571678	2024-07-2	7	2024	14:05	Day	KHVN	TWEED-NE	41.2639	-72.8871	20	CT	ANE			VXP	AVELO AIR	N706VL	402	B-737-700	148	42	10	01	A
18	1572292	2024-07-2	7	2024			KGON	GROTON-N	41.33006	-72.04514	23	CT	ANE			UNK	UNKNOWN			UNKNOWN					
19	1572792	2024-07-2	7	2024			KBDL	BRADLEY I	41.93887	-72.68323	24	CT	ANE			UNK	UNKNOWN			UNKNOWN					
20	1571685	2024-07-2	7	2024	15:45	Day	KBDL	BRADLEY I	41.93887	-72.68323	24	CT	ANE			PDT	PIEDMONT	N933JN	5859	EMB-145	332	14	01	10	A
21	1573082	2024-07-2	7	2024			KBDL	BRADLEY I	41.93887	-72.68323		CT	ANE			MXY	BREEZE AI	N218BZ	541	A-220	04B		34		A
22	1569909	2024-07-2	7	2024	07:27	Day	KBDL	BRADLEY I	41.93887	-72.68323	24	CT	ANE			AAL	AMERICAN	N939NN	2250	B-737-800	148	43	10	01	A
23	1569891	2024-07-2	7	2024	09:42	Day	KHFD	HARTFORI	41.73627	-72.65021	02	CT	ANE			BUS	BUSINESS	N865G		PA-28	729	23	07	19	A
24	1572628	2024-07-2	7	2024	07:30	Dawn	KHVN	TWEED-NE	41.2639	-72.8871	20	CT	ANE			VXP	AVELO AIR	N705VL	401	B-737-700	148	42	10	01	A
25	1572635	2024-07-2	7	2024			KHVN	TWEED-NE	41.2639	-72.8871	2/20	CT	ANE			UNK	UNKNOWN			UNKNOWN					
26	1572170	2024-07-1	7	2024	07:38	Day	KBDL	BRADLEY I	41.93887	-72.68323	6/24	CT	ANE			AAL	AMERICAN	N358PW	2250	B-737-800	148	43	10	01	A
27	1571450	2024-07-1	7	2024	08:18	Day	KBDL	BRADLEY I	41.93887	-72.68323	24	CT	ANE			UPS	UPS AIRLIN	N121UP	2061	A-300	04A	01	34	46	A
28	1570337	2024-07-1	7	2024			KBDL	BRADLEY I	41.93887	-72.68323	24	CT	ANE			UNK	UNKNOWN			UNKNOWN					
29	1568438	2024-07-1	7	2024			KHVN	TWEED-NE	41.2639	-72.8871	20	CT	ANE			UNK	UNKNOWN			UNKNOWN					
30	1570019	2024-07-1	7	2024			KBDL	BRADLEY I	41.93887	-72.68323	24	CT	ANE			UNK	UNKNOWN			UNKNOWN					
31	1567873	2024-07-0	7	2024			KHVN	TWEED-NE	41.2639	-72.8871		CT	ANE			UNK	UNKNOWN			UNKNOWN					
32	1567483	2024-07-0	7	2024	11:14	Day	KBDL	BRADLEY I	41.93887	-72.68323	24	CT	ANE			JBU	JETBLUE A	N760JB	1142	A-320	04A	03	23	01	A
33	1568067	2024-07-0	7	2024	23:26	Night	KBDL	BRADLEY I	41.93887	-72.68323	24	CT	ANE			JBU	JETBLUE A	N712JB	460	A-320	04A	03	23	01	A
34	1564295	2024-06-3	6	2024			KHVN	TWEED-NE	41.2639	-72.8871		CT	ANE			UNK	UNKNOWN			UNKNOWN					

Basic Spreadsheet Navigation: Reviewing Data





Your cursor needs to be somewhere in your data

Move to last cell in worksheet	Ctrl End
Move to first cell in worksheet	Ctrl Home



Basic Spreadsheet Navigation: Reviewing Data







Move to last column by hitting CTRL on a PC and your **right arrow key** on your keyboard

Click the **left arrow key** to go back

Reviewing Data

How many records or rows do we have?

What data do we have?

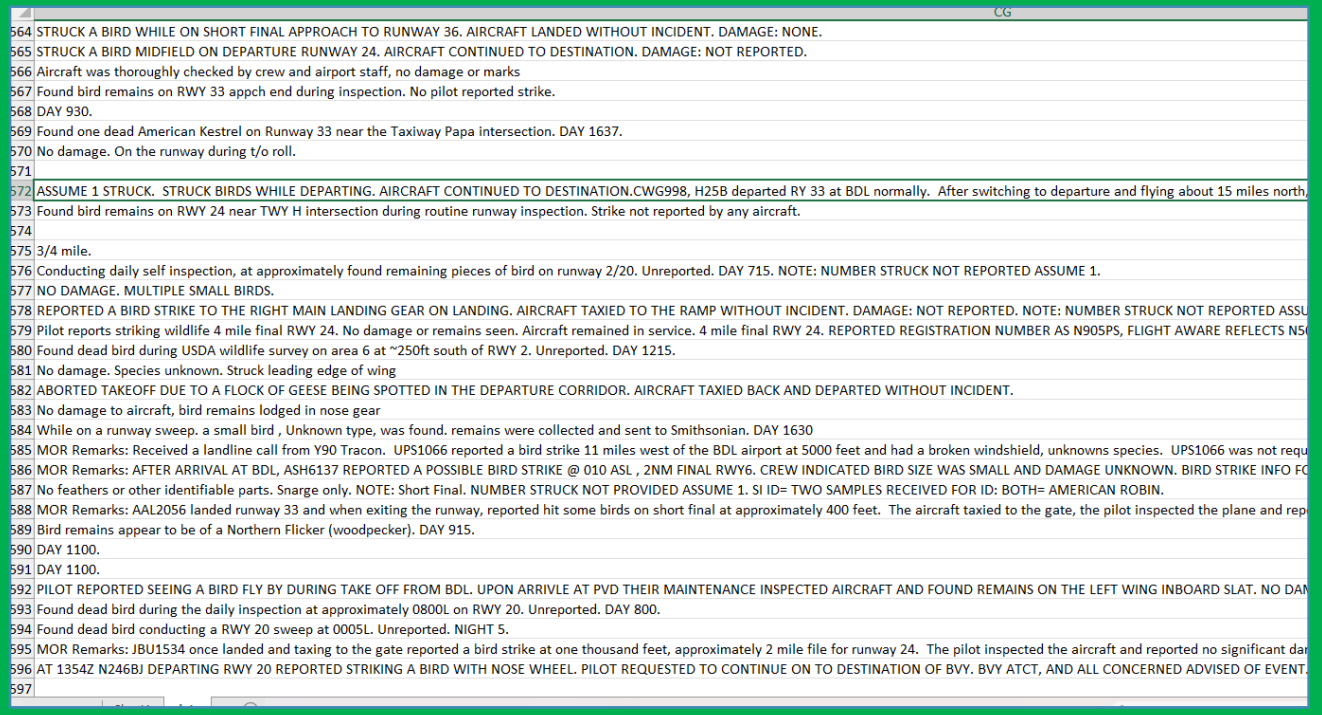
How many columns do we have?

Is the data dirty?

What is missing?

What time period does our data cover?

What questions do we want to ask?



	CG
564	STRUCK A BIRD WHILE ON SHORT FINAL APPROACH TO RUNWAY 36. AIRCRAFT LANDED WITHOUT INCIDENT. DAMAGE: NONE.
565	STRUCK A BIRD MIDFIELD ON DEPARTURE RUNWAY 24. AIRCRAFT CONTINUED TO DESTINATION. DAMAGE: NOT REPORTED.
566	Aircraft was thoroughly checked by crew and airport staff, no damage or marks
567	Found bird remains on RWY 33 appch end during inspection. No pilot reported strike.
568	DAY 930.
569	Found one dead American Kestrel on Runway 33 near the Taxiway Papa intersection. DAY 1637.
570	No damage. On the runway during t/o roll.
571	
572	ASSUME 1 STRUCK. STRUCK BIRDS WHILE DEPARTING. AIRCRAFT CONTINUED TO DESTINATION.CWG998, H25B departed RY 33 at BDL normally. After switching to departure and flying about 15 miles north,
573	Found bird remains on RWY 24 near TWY H intersection during routine runway inspection. Strike not reported by any aircraft.
574	
575	3/4 mile.
576	Conducting daily self inspection, at approximately found remaining pieces of bird on runway 2/20. Unreported. DAY 715. NOTE: NUMBER STRUCK NOT REPORTED ASSUME 1.
577	NO DAMAGE. MULTIPLE SMALL BIRDS.
578	REPORTED A BIRD STRIKE TO THE RIGHT MAIN LANDING GEAR ON LANDING. AIRCRAFT TAXIED TO THE RAMP WITHOUT INCIDENT. DAMAGE: NOT REPORTED. NOTE: NUMBER STRUCK NOT REPORTED ASSU
579	Pilot reports striking wildlife 4 mile final RWY 24. No damage or remains seen. Aircraft remained in service. 4 mile final RWY 24. REPORTED REGISTRATION NUMBER AS N905PS, FLIGHT AWARE REFLECTS N5
580	Found dead bird during USDA wildlife survey on area 6 at ~250ft south of RWY 2. Unreported. DAY 1215.
581	No damage. Species unknown. Struck leading edge of wing
582	ABORTED TAKEOFF DUE TO A FLOCK OF GEESE BEING SPOTTED IN THE DEPARTURE CORRIDOR. AIRCRAFT TAXIED BACK AND DEPARTED WITHOUT INCIDENT.
583	No damage to aircraft, bird remains lodged in nose gear
584	While on a runway sweep. a small bird , Unknown type, was found. remains were collected and sent to Smithsonian. DAY 1630
585	MOR Remarks: Received a landline call from Y90 Tracon. UPS1066 reported a bird strike 11 miles west of the BDL airport at 5000 feet and had a broken windshield, unknowns species. UPS1066 was not requ
586	MOR Remarks: AFTER ARRIVAL AT BDL, ASH6137 REPORTED A POSSIBLE BIRD STRIKE @ 010 ASL , 2NM FINAL RWY6. CREW INDICATED BIRD SIZE WAS SMALL AND DAMAGE UNKNOWN. BIRD STRIKE INFO FO
587	No feathers or other identifiable parts. Snarge only. NOTE: Short Final. NUMBER STRUCK NOT PROVIDED ASSUME 1. SI ID= TWO SAMPLES RECEIVED FOR ID: BOTH= AMERICAN ROBIN.
588	MOR Remarks: AAL2056 landed runway 33 and when exiting the runway, reported hit some birds on short final at approximately 400 feet. The aircraft taxied to the gate, the pilot inspected the plane and rep
589	Bird remains appear to be of a Northern Flicker (woodpecker). DAY 915.
590	DAY 1100.
591	DAY 1100.
592	PILOT REPORTED SEEING A BIRD FLY BY DURING TAKE OFF FROM BDL. UPON ARRIVLE AT PVD THEIR MAINTENANCE INSPECTED AIRCRAFT AND FOUND REMAINS ON THE LEFT WING INBOARD SLAT. NO DAM
593	Found dead bird during the daily inspection at approximately 0800L on RWY 20. Unreported. DAY 800.
594	Found dead bird conducting a RWY 20 sweep at 0005L. Unreported. NIGHT 5.
595	MOR Remarks: JBU1534 once landed and taxing to the gate reported a bird strike at one thousand feet, approximately 2 mile file for runway 24. The pilot inspected the aircraft and reported no significant dar
596	AT 1354Z N246BJ DEPARTING RWY 20 REPORTED STRIKING A BIRD WITH NOSE WHEEL. PILOT REQUESTED TO CONTINUE ON TO DESTINATION OF BVY. BVY ATCT, AND ALL CONCERNED ADVISED OF EVENT.
597	

Basic Spreadsheet Navigation: Enlarging Columns

Resizing Columns: Another Option

Resize columns

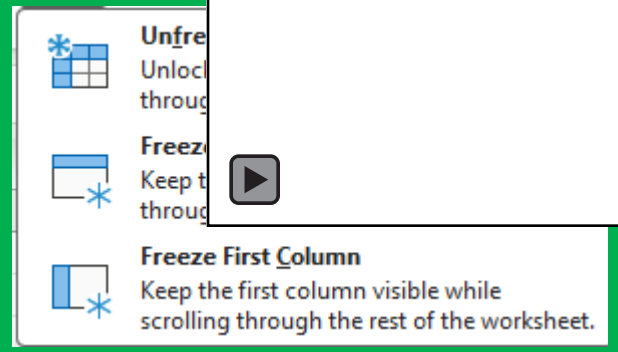
1. Select a column or a range of columns.

2. Place the pointer on the boundary between the column headers.

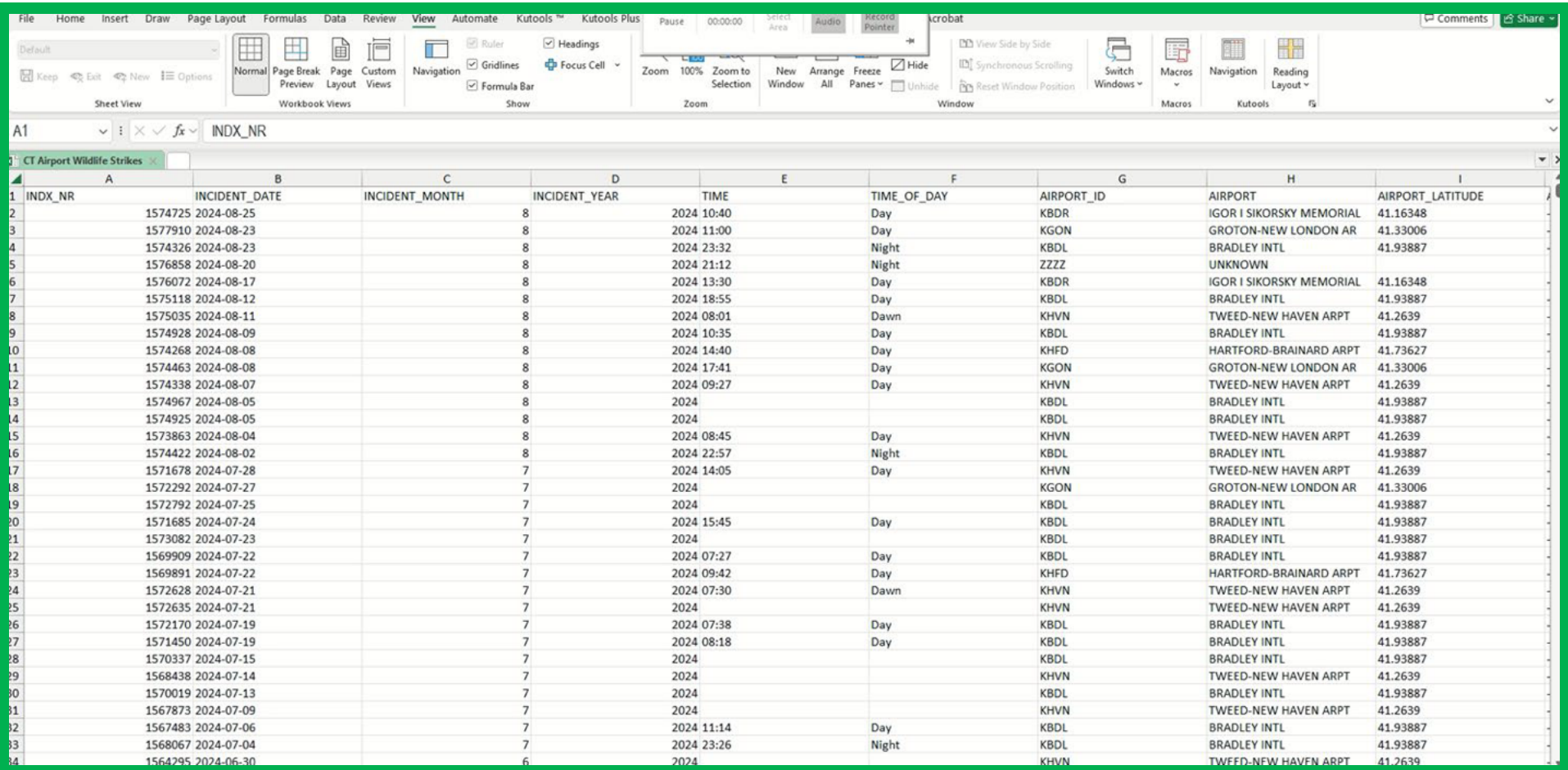


3. Drag the boundary to change the width. This resizes all of the selected columns to the same width.

Basic Spreadsheet Navigation: Freeze Panes



Basic Spreadsheet Navigation: Freeze Panes



	A	B	C	D	E	F	G	H	I
1	INDX_NR	INCIDENT_DATE	INCIDENT_MONTH	INCIDENT_YEAR	TIME	TIME_OF_DAY	AIRPORT_ID	AIRPORT	AIRPORT_LATITUDE
2	1574725	2024-08-25	8	2024	10:40	Day	KBDR	IGOR I SIKORSKY MEMORIAL	41.16348
3	1577910	2024-08-23	8	2024	11:00	Day	KGON	GROTON-NEW LONDON AR	41.33006
4	1574326	2024-08-23	8	2024	23:32	Night	KBDL	BRADLEY INTL	41.93887
5	1576858	2024-08-20	8	2024	21:12	Night	ZZZZ	UNKNOWN	
6	1576072	2024-08-17	8	2024	13:30	Day	KBDR	IGOR I SIKORSKY MEMORIAL	41.16348
7	1575118	2024-08-12	8	2024	18:55	Day	KBDL	BRADLEY INTL	41.93887
8	1575035	2024-08-11	8	2024	08:01	Dawn	KHVN	TWEED-NEW HAVEN ARPT	41.2639
9	1574928	2024-08-09	8	2024	10:35	Day	KBDL	BRADLEY INTL	41.93887
10	1574268	2024-08-08	8	2024	14:40	Day	KHFD	HARTFORD-BRAINARD ARPT	41.73627
11	1574463	2024-08-08	8	2024	17:41	Day	KGON	GROTON-NEW LONDON AR	41.33006
12	1574338	2024-08-07	8	2024	09:27	Day	KHVN	TWEED-NEW HAVEN ARPT	41.2639
13	1574967	2024-08-05	8	2024			KBDL	BRADLEY INTL	41.93887
14	1574925	2024-08-05	8	2024			KBDL	BRADLEY INTL	41.93887
15	1573863	2024-08-04	8	2024	08:45	Day	KHVN	TWEED-NEW HAVEN ARPT	41.2639
16	1574422	2024-08-02	8	2024	22:57	Night	KBDL	BRADLEY INTL	41.93887
17	1571678	2024-07-28	7	2024	14:05	Day	KHVN	TWEED-NEW HAVEN ARPT	41.2639
18	1572292	2024-07-27	7	2024			KGON	GROTON-NEW LONDON AR	41.33006
19	1572792	2024-07-25	7	2024			KBDL	BRADLEY INTL	41.93887
20	1571685	2024-07-24	7	2024	15:45	Day	KBDL	BRADLEY INTL	41.93887
21	1573082	2024-07-23	7	2024			KBDL	BRADLEY INTL	41.93887
22	1569909	2024-07-22	7	2024	07:27	Day	KBDL	BRADLEY INTL	41.93887
23	1569891	2024-07-22	7	2024	09:42	Day	KHFD	HARTFORD-BRAINARD ARPT	41.73627
24	1572628	2024-07-21	7	2024	07:30	Dawn	KHVN	TWEED-NEW HAVEN ARPT	41.2639
25	1572635	2024-07-21	7	2024			KHVN	TWEED-NEW HAVEN ARPT	41.2639
26	1572170	2024-07-19	7	2024	07:38	Day	KBDL	BRADLEY INTL	41.93887
27	1571450	2024-07-19	7	2024	08:18	Day	KBDL	BRADLEY INTL	41.93887
28	1570337	2024-07-15	7	2024			KBDL	BRADLEY INTL	41.93887
29	1568438	2024-07-14	7	2024			KHVN	TWEED-NEW HAVEN ARPT	41.2639
30	1570019	2024-07-13	7	2024			KBDL	BRADLEY INTL	41.93887
31	1567873	2024-07-09	7	2024			KHVN	TWEED-NEW HAVEN ARPT	41.2639
32	1567483	2024-07-06	7	2024	11:14	Day	KBDL	BRADLEY INTL	41.93887
33	1568067	2024-07-04	7	2024	23:26	Night	KBDL	BRADLEY INTL	41.93887
34	1564295	2024-06-30	6	2024			KHVN	TWEED-NEW HAVEN ARPT	41.2639

Freezing the panes let's us freeze rows or columns

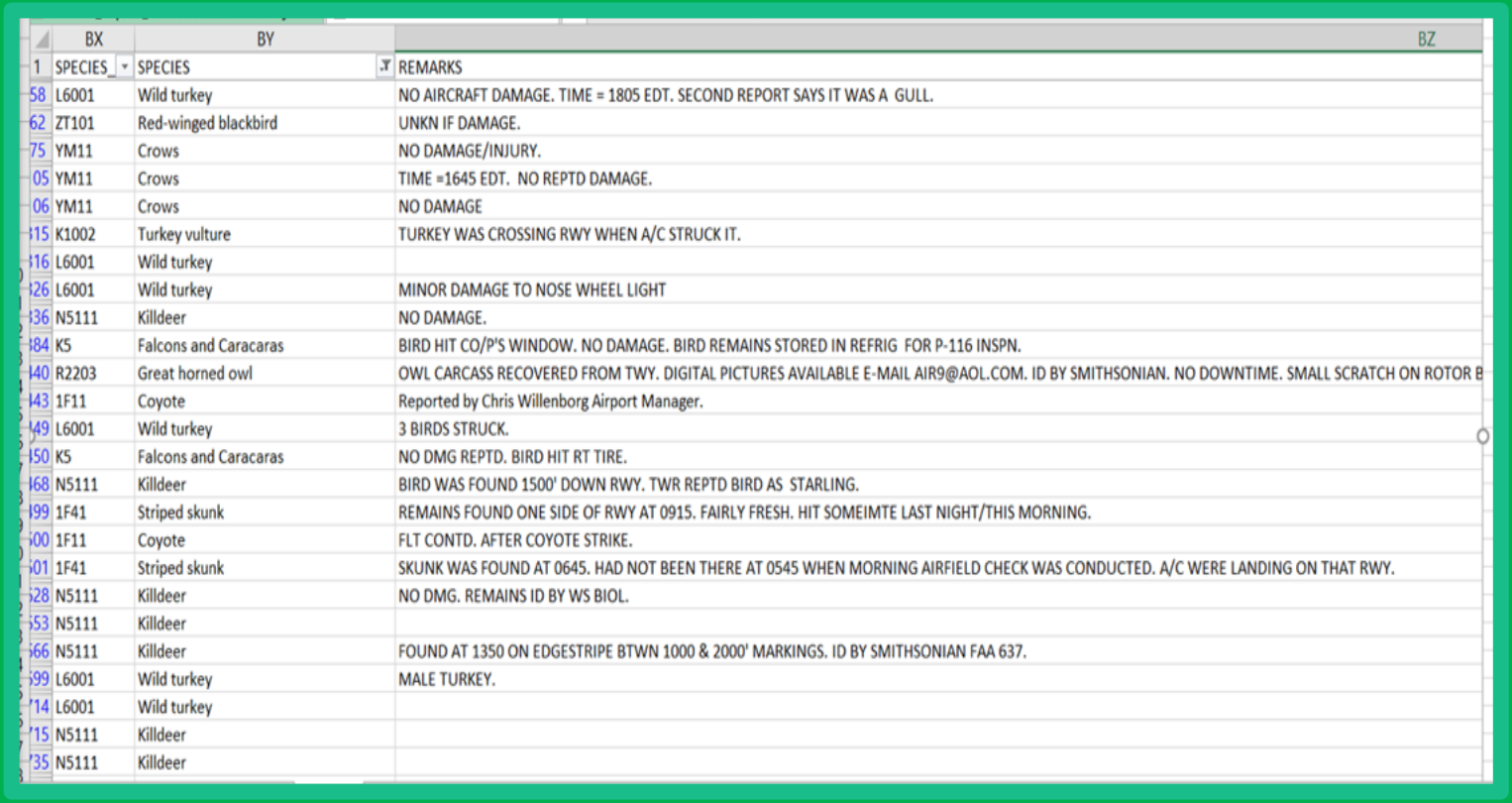
Then we can scroll through our data and still see the headers at the top

Basic Spreadsheet Navigation – Inserting Columns

- Select a column heading
  - Right click on it
- Excel will insert a column to the left



Inserting Sheets



	BX	BY	BZ
1	SPECIES_	SPECIES	REMARKS
58	L6001	Wild turkey	NO AIRCRAFT DAMAGE. TIME = 1805 EDT. SECOND REPORT SAYS IT WAS A GULL.
62	ZT101	Red-winged blackbird	UNKN IF DAMAGE.
75	YM11	Crows	NO DAMAGE/INJURY.
05	YM11	Crows	TIME =1645 EDT. NO REPTD DAMAGE.
06	YM11	Crows	NO DAMAGE
15	K1002	Turkey vulture	TURKEY WAS CROSSING RWY WHEN A/C STRUCK IT.
16	L6001	Wild turkey	
26	L6001	Wild turkey	MINOR DAMAGE TO NOSE WHEEL LIGHT
36	N5111	Killdeer	NO DAMAGE.
84	K5	Falcons and Caracaras	BIRD HIT CO/P'S WINDOW. NO DAMAGE. BIRD REMAINS STORED IN REFRIG FOR P-116 INSPN.
40	R2203	Great horned owl	OWL CARCASS RECOVERED FROM TWY. DIGITAL PICTURES AVAILABLE E-MAIL AIR9@AOL.COM. ID BY SMITHSONIAN. NO DOWNTIME. SMALL SCRATCH ON ROTOR B
43	1F11	Coyote	Reported by Chris Willenborg Airport Manager.
49	L6001	Wild turkey	3 BIRDS STRUCK.
50	K5	Falcons and Caracaras	NO DMG REPTD. BIRD HIT RT TIRE.
68	N5111	Killdeer	BIRD WAS FOUND 1500' DOWN RWY. TWR REPTD BIRD AS STARLING.
99	1F41	Striped skunk	REMAINS FOUND ONE SIDE OF RWY AT 0915. FAIRLY FRESH. HIT SOMEIMTE LAST NIGHT/THIS MORNING.
00	1F11	Coyote	FLT CONTD. AFTER COYOTE STRIKE.
01	1F41	Striped skunk	SKUNK WAS FOUND AT 0645. HAD NOT BEEN THERE AT 0545 WHEN MORNING AIRFIELD CHECK WAS CONDUCTED. A/C WERE LANDING ON THAT RWY.
28	N5111	Killdeer	NO DMG. REMAINS ID BY WS BIOL.
53	N5111	Killdeer	
66	N5111	Killdeer	FOUND AT 1350 ON EDGESTRIPE BTWN 1000 & 2000' MARKINGS. ID BY SMITHSONIAN FAA 637.
99	L6001	Wild turkey	MALE TURKEY.
14	L6001	Wild turkey	
15	N5111	Killdeer	
35	N5111	Killdeer	

We can add new sheets to the bottom of our spreadsheet and rename the sheets as we do different analyses on the data

Basic Spreadsheet Navigation – Inserting Rows

- Select the left-hand side of a row
  - Right click on it
  - Excel will insert a row above it

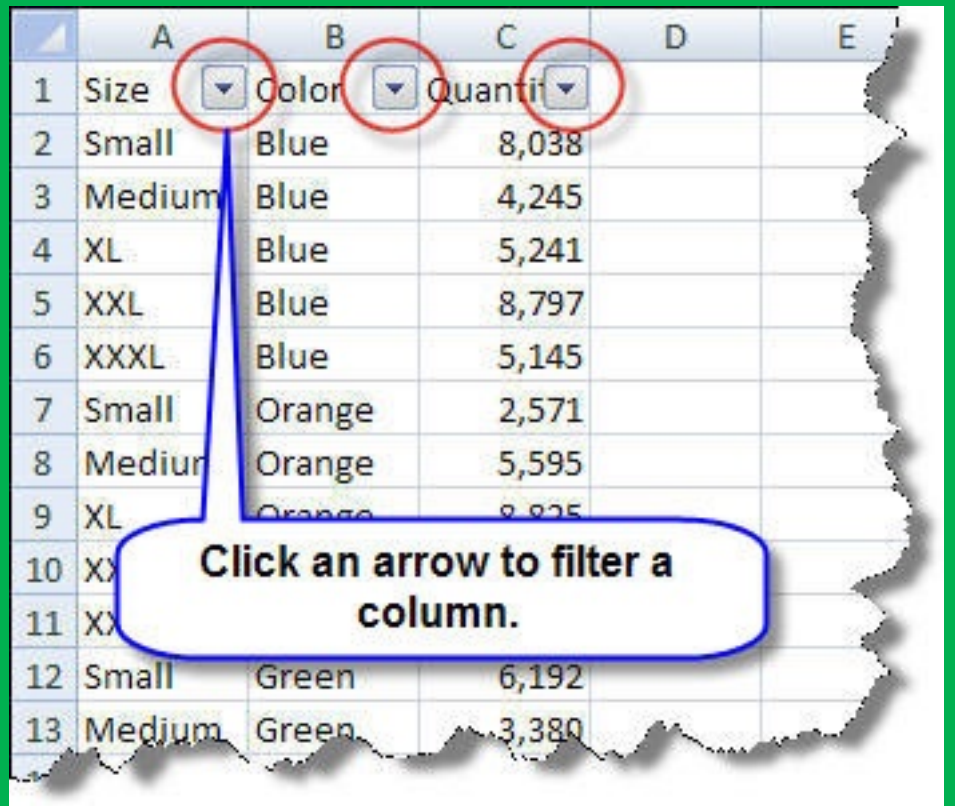


Basic Spreadsheet Navigation – Hiding Columns and Rows



- Select the column
- Right click
- Choose delete
- Same for deleting rows

Filtering

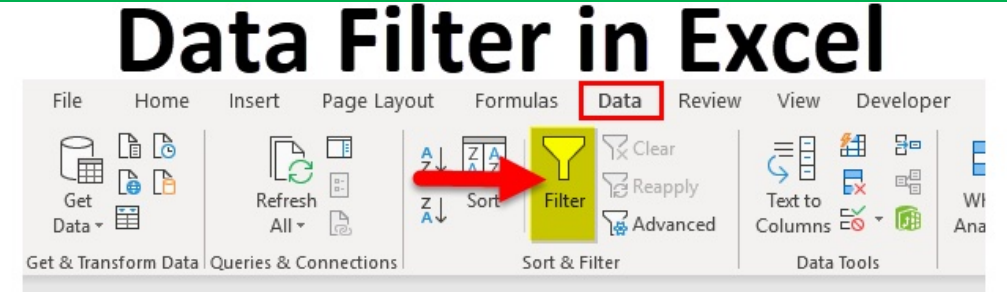

A
B
C
D
E
1 Size Color Quanti
2 Small Blue 8,038
3 Medium Blue 4,245
4 XL Blue 5,241
5 XXL Blue 8,797
6 XXXL Blue 5,145
7 Small Orange 2,571
8 Mediu Orange 5,595
9 XL
10 XX
11 XX
12 Small Green 6,192
13 Medium Green 3,380
Click an arrow to filter a column.

Basic Spreadsheet Navigation: Working with Columns

- Scroll over to column D – “Incident Year”

C	D	E	F
NT_MONTH	INCIDENT_YEAR	TIME	TIME_OF_DA
8	2024	10:40	Day
8	2024	11:00	Day
8	2024	23:32	Night
8	2024	21:12	Night
8	2024	13:30	Day
8	2024	18:55	Day
8	2024	08:01	Dawn
8	2024	10:35	Day
8	2024	14:40	Day
8	2024	17:41	Day
8	2024	09:27	Day
8	2024		
8	2024		
8	2024	08:45	Day

Basic Spreadsheet Navigation: Filtering



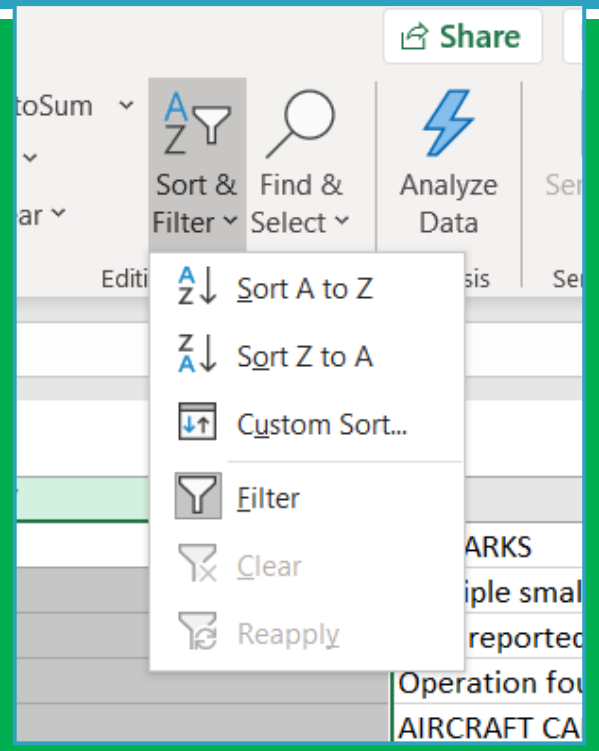
- We can filter our data to look more granularly at detail to find stories
- Let's filter on column **D**
- Scroll to it
- Select your data with CMND A or CTRL A (for PC)
- Go the ribbon under "Home" and click to "turn on" your filter

Remember: This is Dirty Data!

	A	B	C	D	E	F	G	H	I
1	INDX_NR	INCIDENT_DATE	INCIDENT_MONTH	INCIDENT_YEAR	TIME	TIME_OF_DAY	AIRPORT_ID	AIRPORT	AIRPORT_LATITUDE
2	1574725	2024-08-25	8	2024	10:40	Day	KBDR	IGOR I SIKORSKY MEMORIAL	41.16348
3	1577910	2024-08-23	8	2024	11:00	Day	KGON	GROTON-NEW LONDON AR	41.33006
4	1574326	2024-08-23	8	2024	23:32	Night	KBDL	BRADLEY INTL	41.93887
5	1576858	2024-08-20	8	2024	21:12	Night	ZZZZ	UNKNOWN	
6	1576072	2024-08-17	8	2024	13:30	Day	KBDR	IGOR I SIKORSKY MEMORIAL	41.16348
7	1575118	2024-08-12	8	2024	18:55	Day	KBDL	BRADLEY INTL	41.93887
8	1575035	2024-08-11	8	2024	08:01	Dawn	KHVN	TWEED-NEW HAVEN ARPT	41.2639
9	1574928	2024-08-09	8	2024	10:35	Day	KBDL	BRADLEY INTL	41.93887
10	1574268	2024-08-08	8	2024	14:40	Day	KHFD	HARTFORD-BRAINARD ARPT	41.73627
11	1574463	2024-08-08	8	2024	17:41	Day	KGON	GROTON-NEW LONDON AR	41.33006
12	1574338	2024-08-07	8	2024	09:27	Day	KHVN	TWEED-NEW HAVEN ARPT	41.2639
13	1574967	2024-08-05	8	2024			KBDL	BRADLEY INTL	41.93887
14	1574925	2024-08-05	8	2024			KBDL	BRADLEY INTL	41.93887
15	1573863	2024-08-04	8	2024	08:45	Day	KHVN	TWEED-NEW HAVEN ARPT	41.2639
16	1574422	2024-08-02	8	2024	22:57	Night	KBDL	BRADLEY INTL	41.93887
17	1571678	2024-07-28	7	2024	14:05	Day	KHVN	TWEED-NEW HAVEN ARPT	41.2639
18	1572292	2024-07-27	7	2024			KGON	GROTON-NEW LONDON AR	41.33006
19	1572792	2024-07-25	7	2024			KBDL	BRADLEY INTL	41.93887
20	1571685	2024-07-24	7	2024	15:45	Day	KBDL	BRADLEY INTL	41.93887
21	1573082	2024-07-23	7	2024			KBDL	BRADLEY INTL	41.93887
22	1569909	2024-07-22	7	2024	07:27	Day	KBDL	BRADLEY INTL	41.93887
23	1569891	2024-07-22	7	2024	09:42	Day	KHFD	HARTFORD-BRAINARD ARPT	41.73627
24	1572628	2024-07-21	7	2024	07:30	Dawn	KHVN	TWEED-NEW HAVEN ARPT	41.2639
25	1572635	2024-07-21	7	2024			KHVN	TWEED-NEW HAVEN ARPT	41.2639
26	1572170	2024-07-19	7	2024	07:38	Day	KBDL	BRADLEY INTL	41.93887
27	1571450	2024-07-19	7	2024	08:18	Day	KBDL	BRADLEY INTL	41.93887
28	1570337	2024-07-15	7	2024			KBDL	BRADLEY INTL	41.93887
29	1568438	2024-07-14	7	2024			KHVN	TWEED-NEW HAVEN ARPT	41.2639
30	1570019	2024-07-13	7	2024			KBDL	BRADLEY INTL	41.93887
31	1567873	2024-07-09	7	2024			KHVN	TWEED-NEW HAVEN ARPT	41.2639
32	1567483	2024-07-06	7	2024	11:14	Day	KBDL	BRADLEY INTL	41.93887
33	1568067	2024-07-04	7	2024	23:26	Night	KBDL	BRADLEY INTL	41.93887
34	1564295	2024-06-30	6	2024			KHVN	TWEED-NEW HAVEN ARPT	41.2639

Basic Spreadsheet Navigation: Filtering

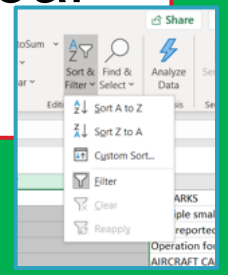
	A	B	C	D	E	F	G	
1	INDX_NR	INCIDENT_DATE	INCIDENT_MONTH	INCIDENT_YEAR	TIME	TIME_OF_DAY	AIRPORT_ID	AIRPORT
2	1574725	2024-08-25	8	2024	10:40	Day	KBDR	IGOR I SIKORSK
3	1577910	2024-08-23	8	2024	11:00	Day	KGON	GROTON-NEW
4	1574326	2024-08-23	8	2024	23:32	Night	KBDL	BRADLEY INTL
5	1576858	2024-08-20	8	2024	21:12	Night	ZZZZ	UNKNOWN
6	1576072	2024-08-17	8	2024	13:30	Day	KBDR	IGOR I SIKORSK
7	1575118	2024-08-12	8	2024	18:55	Day	KBDL	BRADLEY INTL
8	1575035	2024-08-11	8	2024	08:01	Dawn	KHVN	TWEED-NEW H
9	1574928	2024-08-09	8	2024	10:35	Day	KBDL	BRADLEY INTL
10	1574268	2024-08-08	8	2024	14:40	Day	KHFD	HARTFORD-BRA
11	1574463	2024-08-08	8	2024	17:41	Day	KGON	GROTON-NEW
12	1574338	2024-08-07	8	2024	09:27	Day	KHVN	TWEED-NEW H
13	1574967	2024-08-05	8	2024			KBDL	BRADLEY INTL
14	1574925	2024-08-05	8	2024			KBDL	BRADLEY INTL
15	1573863	2024-08-04	8	2024	08:45	Day	KHVN	TWEED-NEW H
16	1574422	2024-08-02	8	2024	22:57	Night	KBDL	BRADLEY INTL

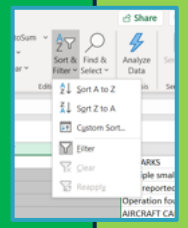


There Are Three Different Ways To Filter

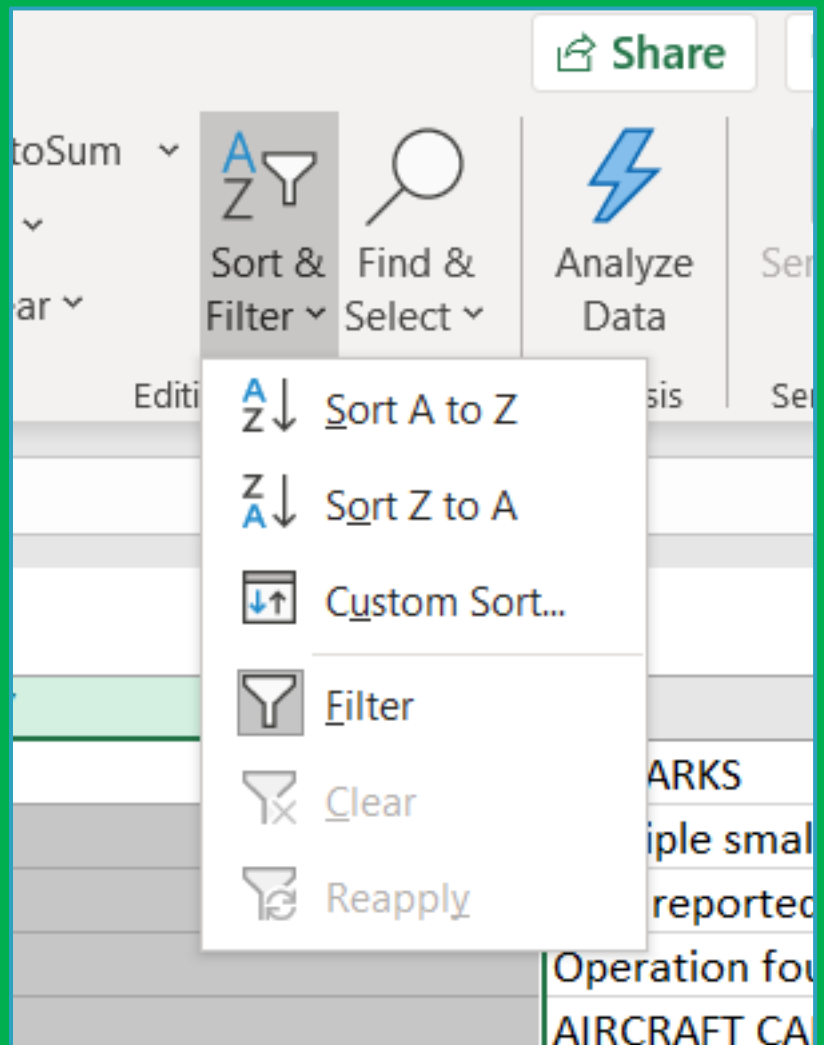
- Put your cursor in the column heading or right on the letter of the column
- Go to your ribbon and find the filter icon

- It looks like a funnel
- On a PC, it will be under the HOME tab on the ribbon
- On a MAC it should also be under the HOME tab
- Once you click the funnel, a small down arrow will appear in your column





Filtering

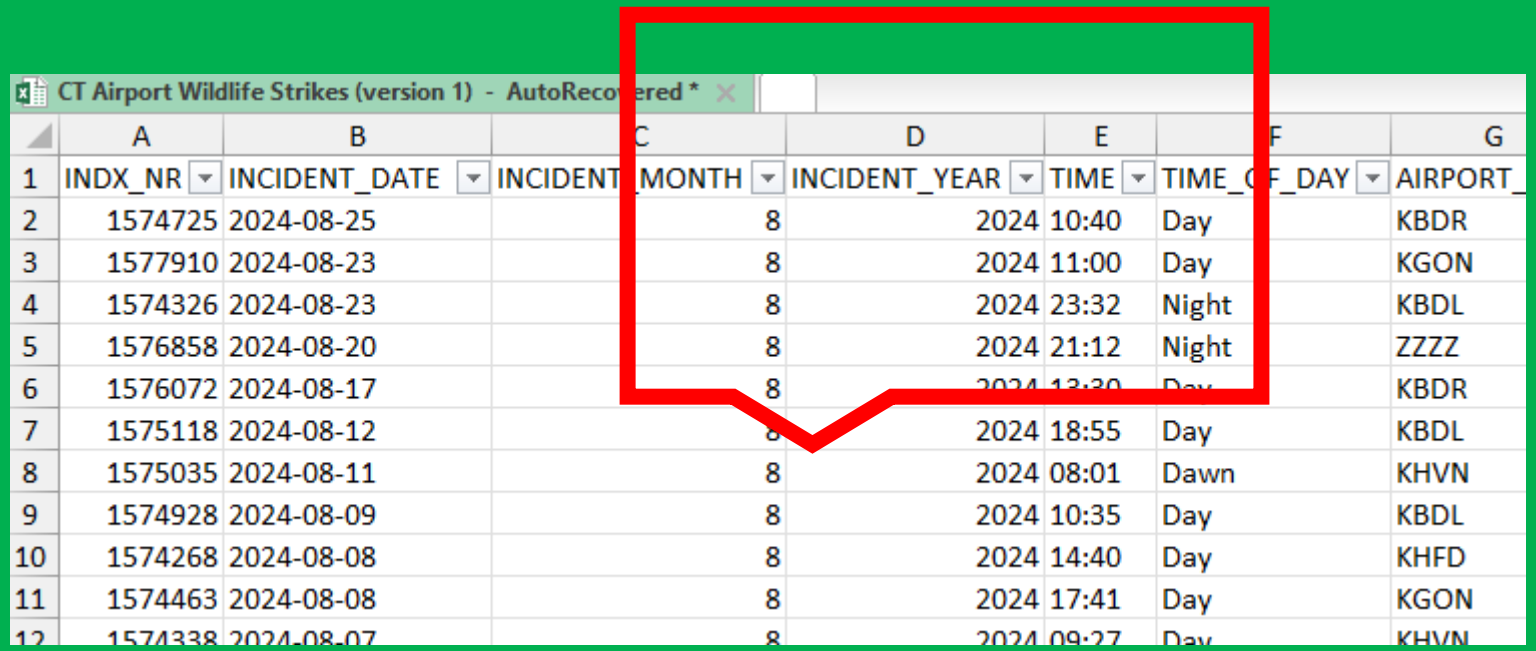


CT Airport Wildlife Strikes (version 1) - AutoRecovered *

	A	B	C	D	E	F	G	H
1	INDX_NR	INCIDENT_DATE	INCIDENT_MONTH	INCIDENT_YEAR	TIME	TIME_OF_DAY	AIRPORT_ID	AIRPORT
2	1574725	2024-08-25	8	2024	10:40	Day	KBDR	IGOR I SIKORSKY MEMORIAL
3	1577910	2024-08-23	8	2024	11:00	Day	KGON	GROTON-NEW LONDON AR
4	1574326	2024-08-23	8	2024	23:32	Night	KBDL	BRADLEY INTL
5	1576858	2024-08-20	8	2024	21:12	Night	ZZZZ	UNKNOWN
6	1576072	2024-08-17	8	2024	13:30	Day	KBDR	IGOR I SIKORSKY MEMORIAL
7	1575118	2024-08-12	8	2024	18:55	Day	KBDL	BRADLEY INTL
8	1575035	2024-08-11	8	2024	08:01	Dawn	KHVN	TWEED-NEW HAVEN ARPT
9	1574928	2024-08-09	8	2024	10:35	Day	KBDL	BRADLEY INTL
10	1574268	2024-08-08	8	2024	14:40	Day	KHFD	HARTFORD-BRAINARD ARPT
11	1574463	2024-08-08	8	2024	17:41	Day	KGON	GROTON-NEW LONDON AR
12	1574338	2024-08-07	8	2024	09:27	Day	KHVN	TWEED-NEW HAVEN ARPT
13	1574967	2024-08-05	8	2024			KBDL	BRADLEY INTL
14	1574925	2024-08-05	8	2024			KBDL	BRADLEY INTL
15	1573863	2024-08-04	8	2024	08:45	Day	KHVN	TWEED-NEW HAVEN ARPT
16	1574422	2024-08-02	8	2024	22:57	Night	KBDL	BRADLEY INTL
17	1571678	2024-07-28	7	2024	14:05	Day	KHVN	TWEED-NEW HAVEN ARPT
18	1572292	2024-07-27	7	2024			KGON	GROTON-NEW LONDON AR
19	1572792	2024-07-25	7	2024			KBDL	BRADLEY INTL
20	1571685	2024-07-24	7	2024	15:45	Day	KBDL	BRADLEY INTL
21	1573082	2024-07-23	7	2024			KBDL	BRADLEY INTL
22	1569909	2024-07-22	7	2024	07:27	Day	KBDL	BRADLEY INTL
23	1569891	2024-07-22	7	2024	09:42	Day	KHFD	HARTFORD-BRAINARD ARPT
24	1572628	2024-07-21	7	2024	07:30	Dawn	KHVN	TWEED-NEW HAVEN ARPT
25	1572635	2024-07-21	7	2024			KHVN	TWEED-NEW HAVEN ARPT
26	1572170	2024-07-19	7	2024	07:38	Day	KBDL	BRADLEY INTL
27	1571450	2024-07-19	7	2024	08:18	Day	KBDL	BRADLEY INTL
28	1570337	2024-07-15	7	2024			KBDL	BRADLEY INTL
29	1568438	2024-07-14	7	2024			KHVN	TWEED-NEW HAVEN ARPT
30	1570019	2024-07-13	7	2024			KBDL	BRADLEY INTL
31	1567873	2024-07-09	7	2024			KHVN	TWEED-NEW HAVEN ARPT
32	1567483	2024-07-06	7	2024	11:14	Day	KBDL	BRADLEY INTL
33	1568067	2024-07-04	7	2024	23:26	Night	KBDL	BRADLEY INTL
34	1564295	2024-06-30	6	2024			KHVN	TWEED-NEW HAVEN ARPT

We Will Filter By Year To Determine Which Year Experienced the Most Incidents

- Once you have selected “filter” you will see this screen with small down arrows to the right of every column
- Click on the down arrow to show the data you have

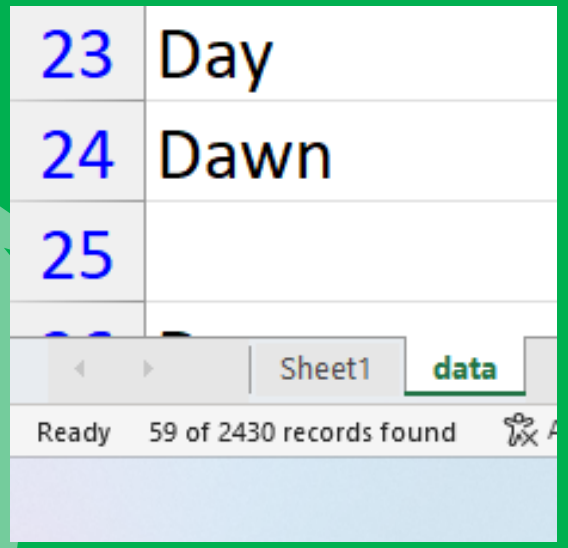


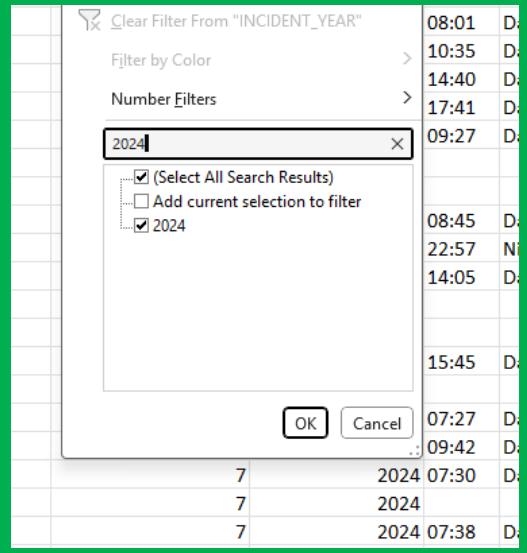
CT Airport Wildlife Strikes (version 1) - AutoRecovered *

	A	B	C	D	E	F	G
1	INDX_NR	INCIDENT_DATE	INCIDENT_MONTH	INCIDENT_YEAR	TIME	TIME_OF_DAY	AIRPORT_
2	1574725	2024-08-25	8	2024	10:40	Day	KBDR
3	1577910	2024-08-23	8	2024	11:00	Day	KGON
4	1574326	2024-08-23	8	2024	23:32	Night	KBDL
5	1576858	2024-08-20	8	2024	21:12	Night	ZZZZ
6	1576072	2024-08-17	8	2024	13:30	Day	KBDR
7	1575118	2024-08-12	8	2024	18:55	Day	KBDL
8	1575035	2024-08-11	8	2024	08:01	Dawn	KHVN
9	1574928	2024-08-09	8	2024	10:35	Day	KBDL
10	1574268	2024-08-08	8	2024	14:40	Day	KHFD
11	1574463	2024-08-08	8	2024	17:41	Day	KGON
12	1574338	2024-08-07	8	2024	09:27	Day	KHVN

Type “2024” Into the Text Box

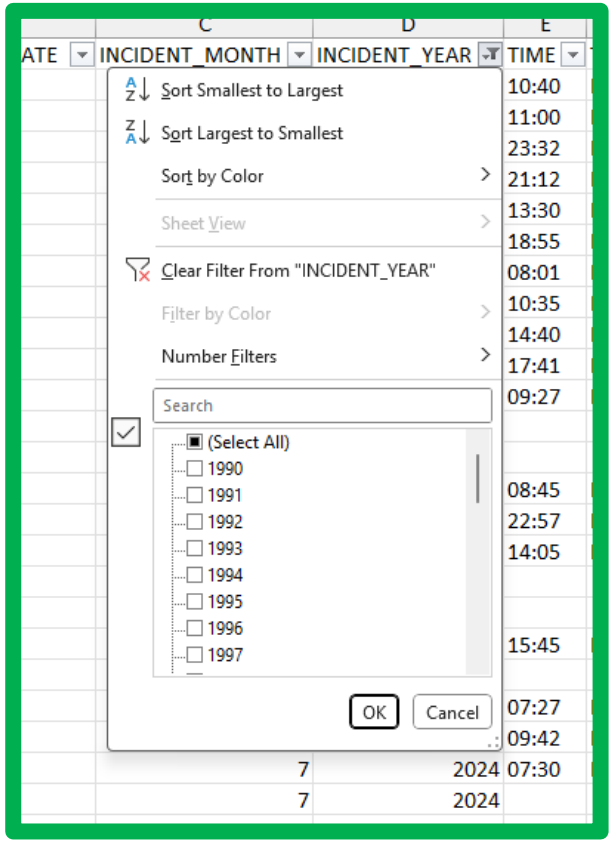
- Remember to check which month we have data for in 2024; it may not be the full year
- Then click OK
- If you look at the bottom of your screen
- You see you have 59 incidents
- You can then compare those with other years
- To determine if incidents are rising





Always Remember to Clear Your Filter

- You do want to filter on top of other filters and get an error
- Clear that filter and select with your cursor multiple years – do you see any trends?



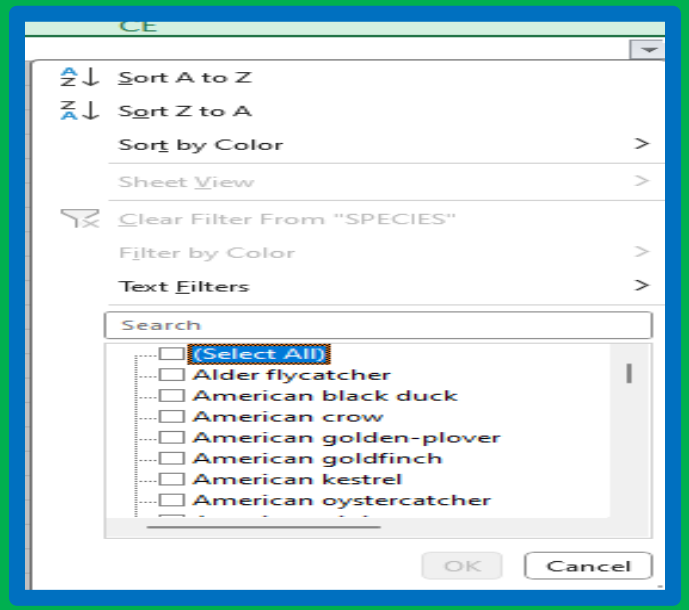
Let's Try That Again with "Species"



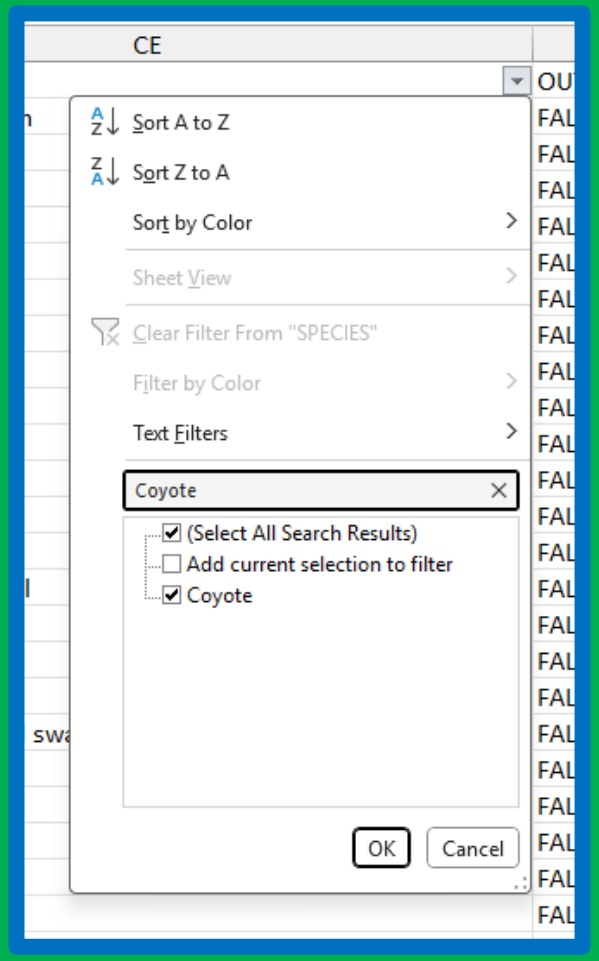
- Scroll over to column CE which is "species"
- EXCEL ran out of letters to name the columns so is doubling up
- We can delete columns we don't need
- Or we can hide them

Let's Filter for Just Coyotes

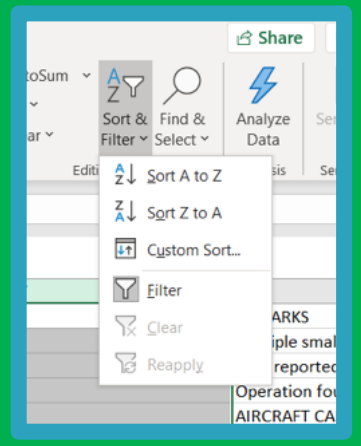
- Uncheck SELECT ALL
- You can either type in the search terms you want, or you can scroll through the choices and check them

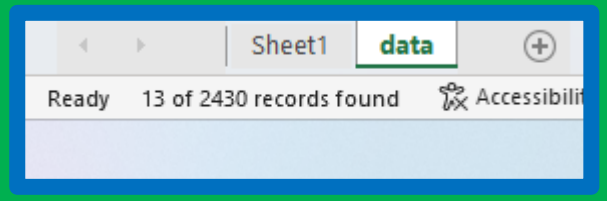


I Typed “Coyote” To Find All Species With “Coyote” or “coyote” In The Name, Then Clicked OK

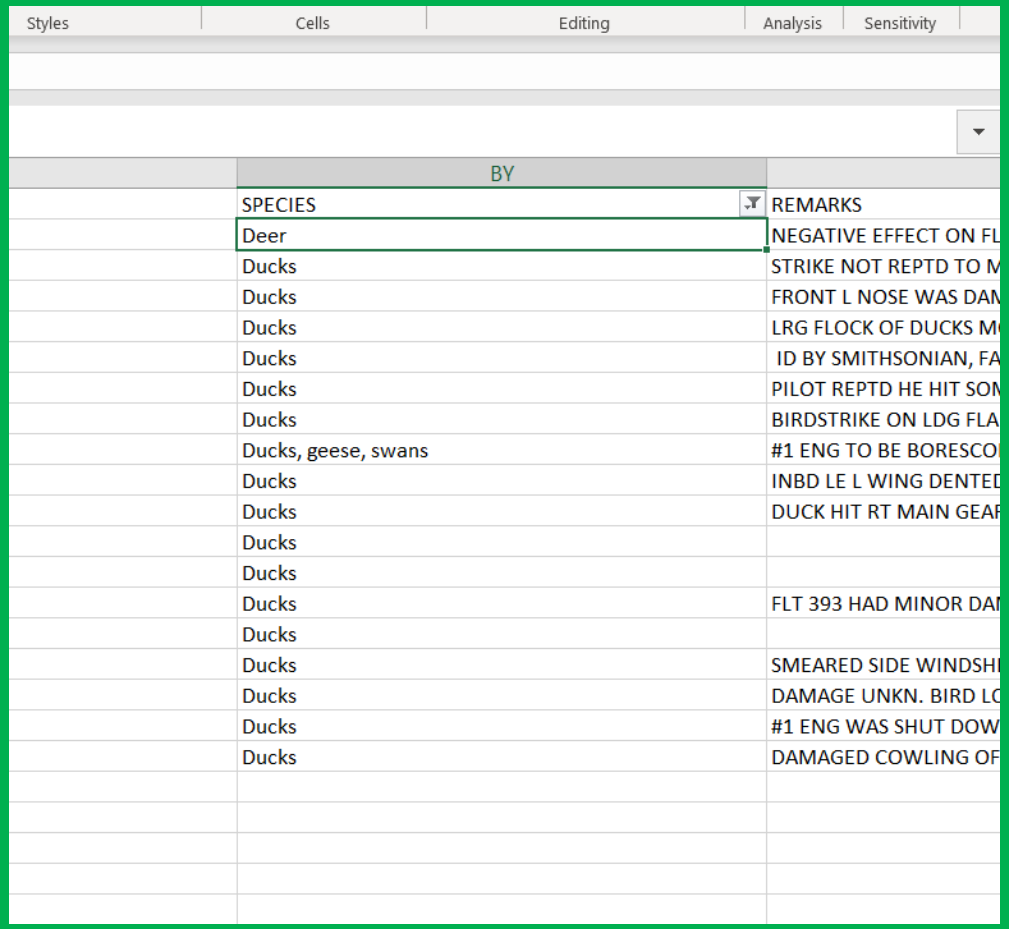


	CB	CC	CD	CE	CF	
1		EFFECT_OTHER	SPECIES_ID	SPECIES	OUT_OF_RANGE_SPECIES	REMARKS
4			1F11	Coyote	FALSE	737-9, Engine Manufacturer CFM INTL Classification Standard Engine Mod
69		GO-AROUND	1F11	Coyote	FALSE	NOT A STRIKE BUT HAD A NEGATIVE EFFECT ON FLIGHT. While UPS 1120
160			1F11	Coyote	FALSE	Struck at 4000' remaining for Rwy 6. Coyote completely intact. No known c
298		Go-around	1F11	Coyote	FALSE	NOT A STRIKE BUT HAD A NEGATVE EFFECT ON FLIGHT. N56582 went aro
404			1F11	Coyote	FALSE	Dead Coyote found in RWY 2/20 safety area, 5 feet from the pavement ed
412		ABORTED LANDING	1F11	Coyote	FALSE	NOT A STRIKE BUT HAD A NEGATIVE EFFECT ON FLIGHT. Aircraft was on f
452		ABORTED LANDING	1F11	Coyote	FALSE	NOT A STRIKE BUT HAD A NEGATIVE EFFECT ON FLIGHT. Aircraft was on
655		Go Around	1F11	Coyote	FALSE	NOT A STRIKE, BUT NEGATIVE EFFECT ON FLIGHT. Aircraft performed a go
679		Go around	1F11	Coyote	FALSE	N430EP in the pattern, 1 Coyote crossed runway 2/20, tower had aircraft g
782			1F11	Coyote	FALSE	GGN7305 was cleared for take-off RWY 33. While entering the runway, th
1017			1F11	Coyote	FALSE	ID BY SMITHSONIAN. A COYOTE WAS FOUND ON RWY 2/20. 5:30 AM DA
1554		GO-AROUND	1F11	Coyote	FALSE	NOT A STRIKE, PILOT MADE A GO AROUND.
1792			1F11	Coyote	FALSE	DEAD COYOTE PUP FOUND ON RWY 2/20 MIDFIELD.
2432						



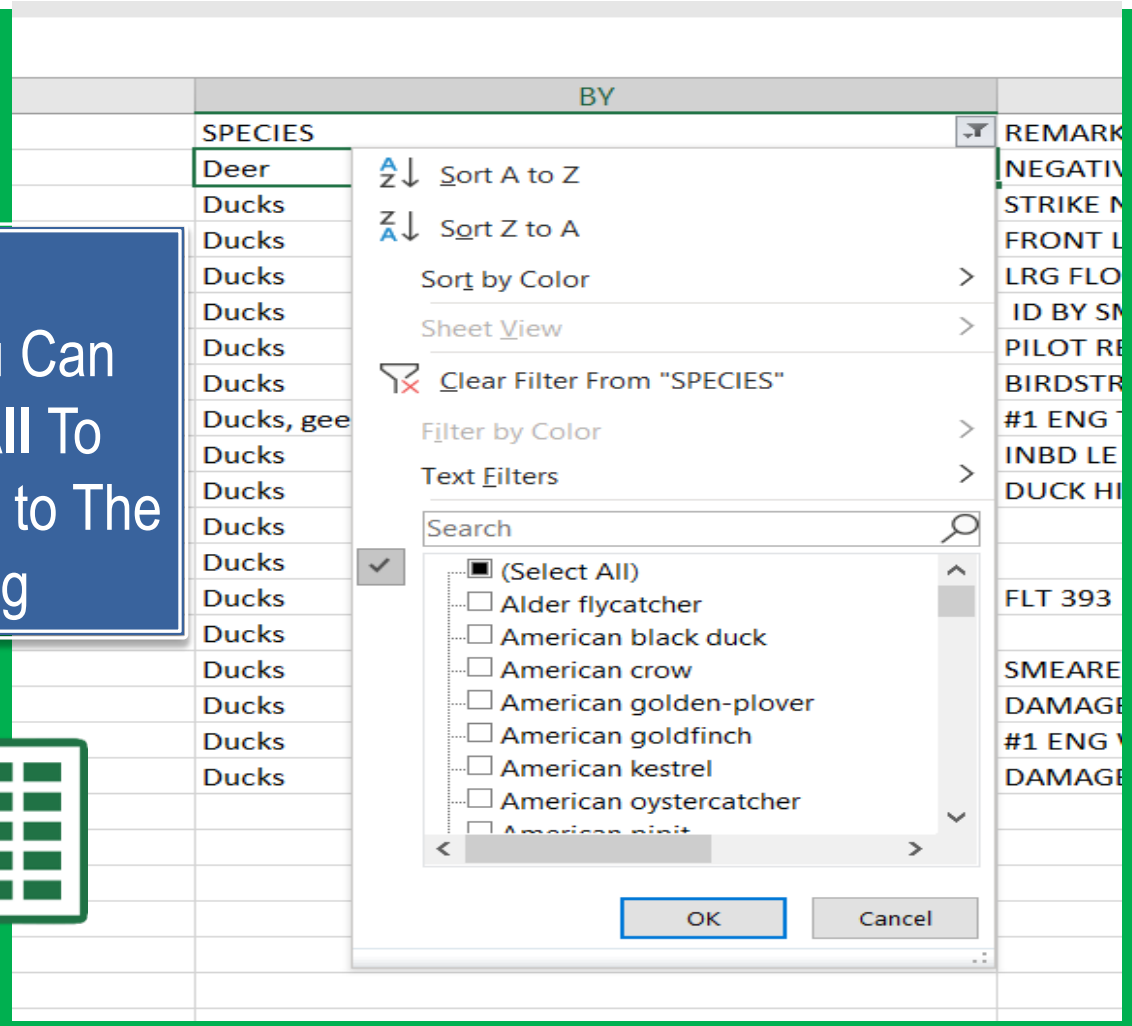


To Start A New Filter, Click On The Funnel Icon

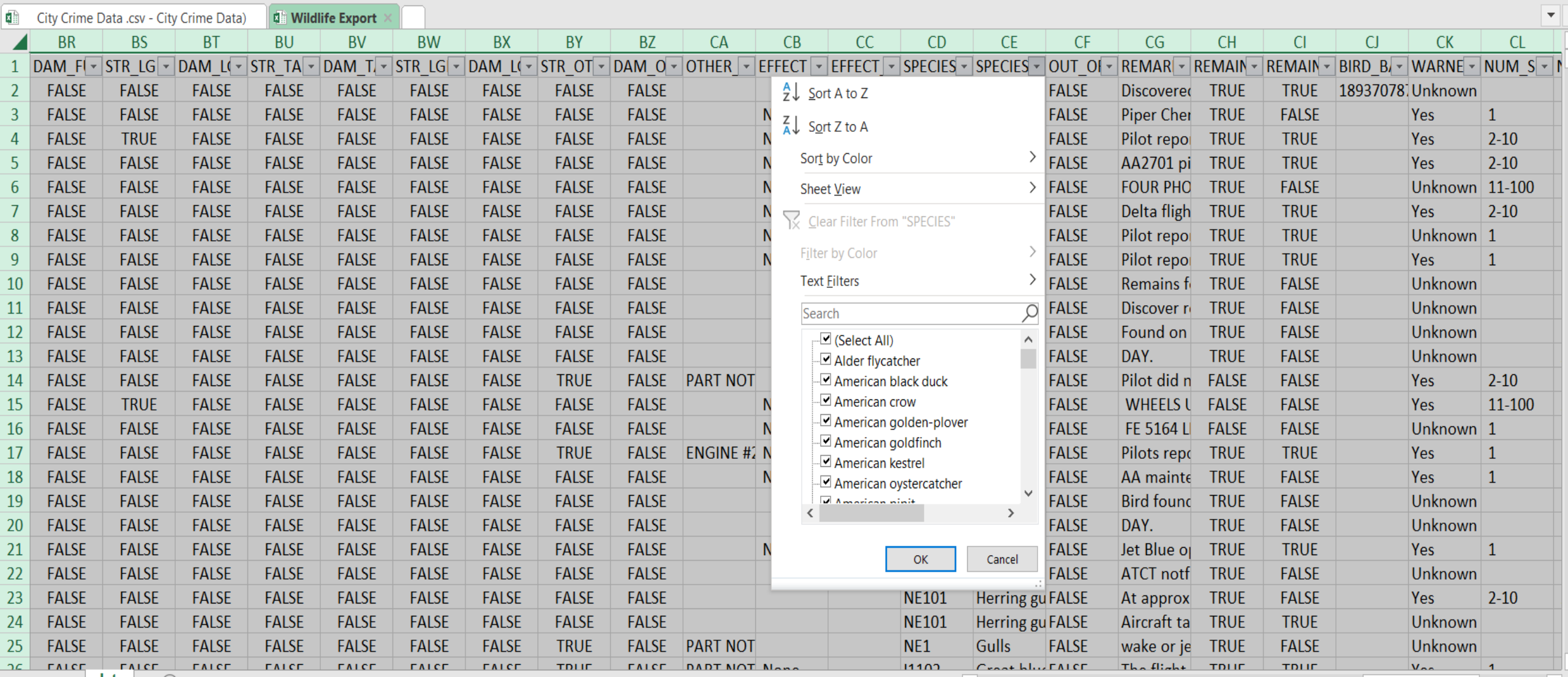


Now You Can **Select All** To Go Back to The Beginning



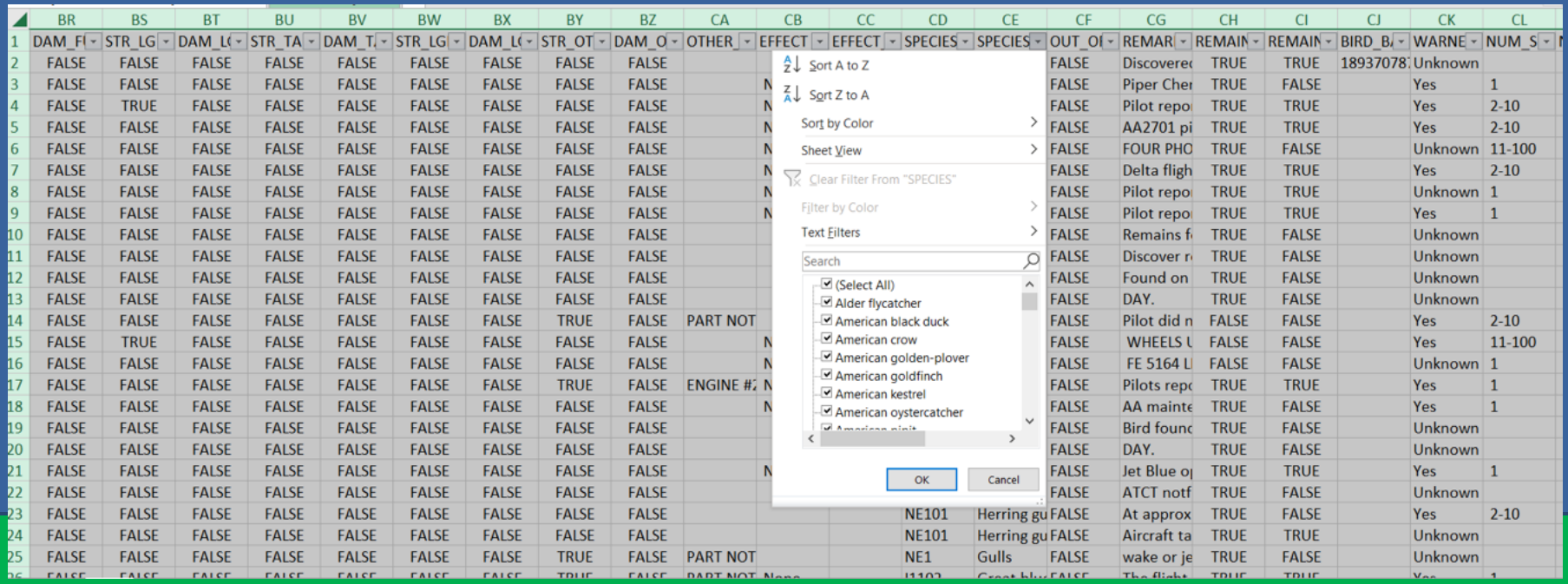


You Can Practice By Filtering to A Specific Species You Are Interested In



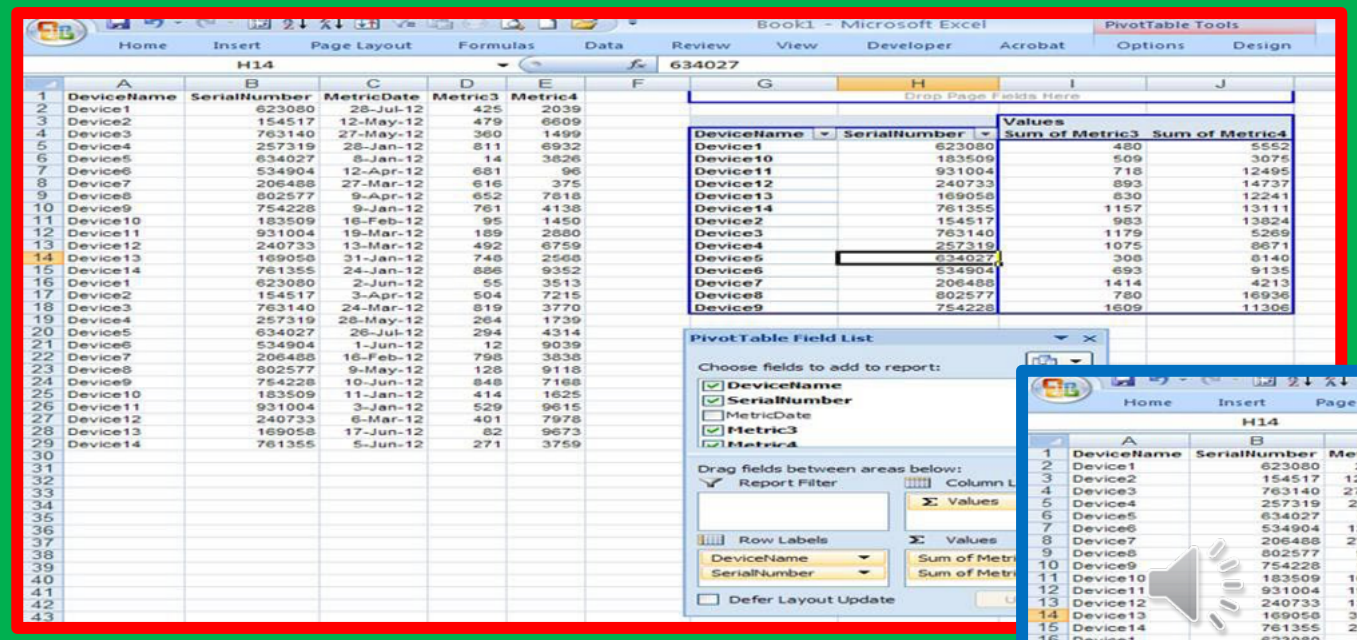
If You Move On To Another Column . . . .

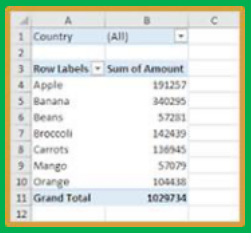
- Remember to click “select all” to choose all the entries in your column so you aren’t filtering incorrectly



Pivot Tables







	A	B
1	Country	(All)
2		
3	Row Labels	Sum of Amount
4	Apple	191257
5	Banana	340295
6	Beans	57281
7	Broccoli	142439
8	Carrots	136945
9	Mango	57079
10	Orange	104438
11	Grand Total	1029734
12		



Pivoting Your Data

What can a pivot table do?

Quickly summarize data to show you the most and least (and a lot more)

Quick way to find anomalies for stories

Also helps check data integrity – how clean is your data

Pivoting Your Data

Pivoting Your Data

- With a pivot table, you set the criteria for analyzing the data
- The spreadsheet program then produces a table that displays the results

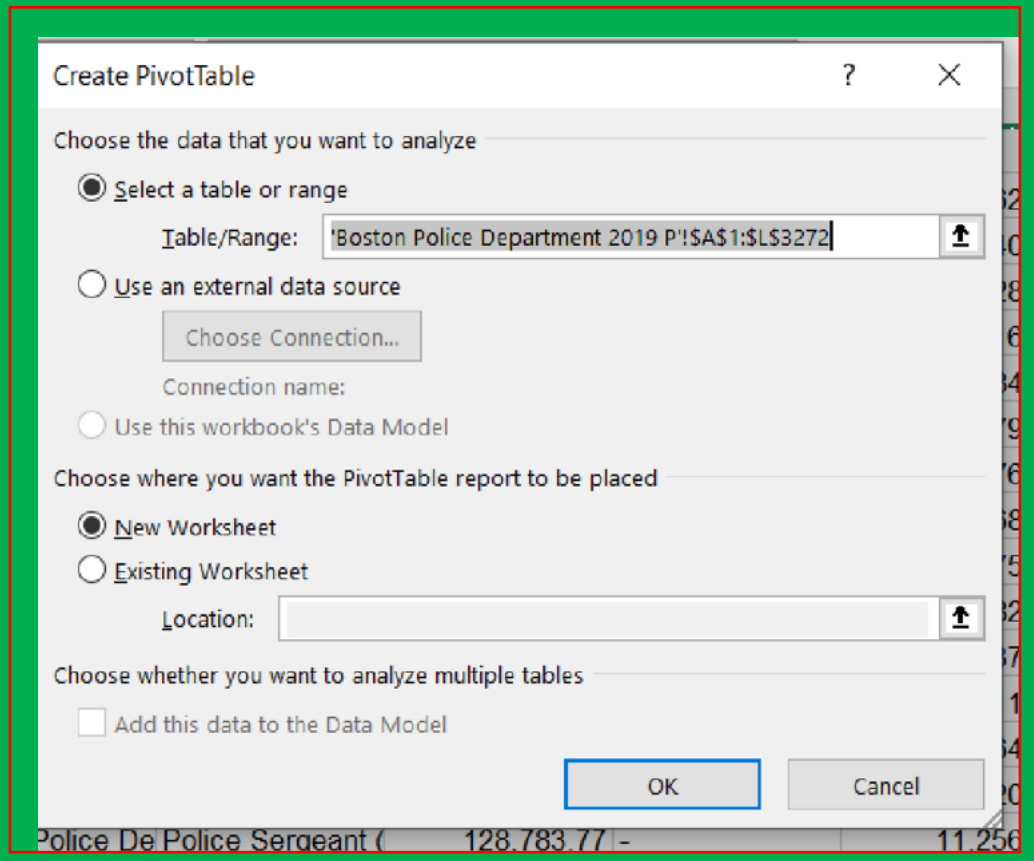
CDC CRUISE SHIP DATA 2000-2010

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	Ship Nam	Submitted	Illness Typ	Voyage Nu	Cruise Len	NextUSPo	Next US Pc	Next US Ai	Embarkati	Embarkati	Embarkati	Disembarl	Disembarl	Disembarl	Total Pass	Total Crew	Passenger	Crew Cases	
2	Disney Ma	########	24HR		7	PCF	Port Cana	########	PCF	Port Cana	########	PCF	Port Cana	########	2406	998	10	4	
3	Celebrity (	########	24HR		7	KWE	Key West,	########	PEV	Port Everg	########	PEV	Port Everg	########	1809	861	5	3	
4	Fantasy	########	24HR		4	PCF	Port Cana	########	PCF	Port Cana	########	PCF	Port Cana	########	2362	929	1	0	
5	Celebrity (	########	24HR		7	MIA	Port of Mi	########	MIA	Port of Mi	########	MIA	Port of Mi	########	1809	864	6	3	
6	Paradise	########	24HR		7	MIA	Port of Mi	########	MIA	Port of Mi	########	MIA	Port of Mi	########	2338	891	2	1	
7	Disney W	########	24HR		4	PCF	Port Cana	########	PCF	Port Cana	########	PCF	Port					1	
8	Fascinatio	########	24HR		7	SJU	San Juan,	########	SJU	San Juan,	########	SJU	San					4	
9	Holiday	########	4HR		4	LAX	Los Angele	########	LAX	Los Angele	########	LAX	Los					2	
10	Jubilee - I	########	24HR		5	TAM	Tampa, FL	########	TAM	Tampa, FL	########	TAM	Tar					2	
11	Sensation	########	24HR		7	TAM	Tampa, FL	########	TAM	Tampa, FL	########	TAM	Tar					0	
12	Fascinatio	########	4HR		7	SJU	San Juan,	########	SJU	San Juan,	########	SJU	San					4	
13	Disney Ma	########	4HR		7	PCF	Port Cana	########	PCF	Port Cana	########	PCF	Por					7	
14	Fantasy	########	24HR		3	PCF	Port Cana	########	PCF	Port Cana	########	PCF	Port Ca					0	
15	Paradise	########	24HR		7	MIA	Port of Mi	########	MIA	Port of Mi	########	MIA	Port of Mi	########	2336	884	3	1	
16	Carnival D	########	24HR		7	SJU	San Juan,	########	SJU	San Juan,	########	SJU	San Juan,	########	3138	1040	8	2	
17	Carnival V	########	24HR		8	MIA	Port of Mi	########	MIA	Port of Mi	########	MIA	Port of Mi	########	2990	1069	3	3	
18	Elation	########	24HR		7	LAX	Los Angele	########	LAX	Los Angele	########	LAX	Los Angele	########	2493	908	3	3	
19	Ecstasy	########	24HR		3	MIA	Port of Mi	########	MIA	Port of Mi	########	MIA	Port of Mi	########	2262	932	0	0	
20	Elation	########	4HR		7	LAX	Los Angele	########	LAX	Los Angele	########	LAX	Los Angele	########	2493	908	6	3	

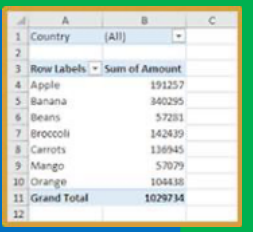
Here is a spreadsheet with many columns and rows

Pivoting Your Data

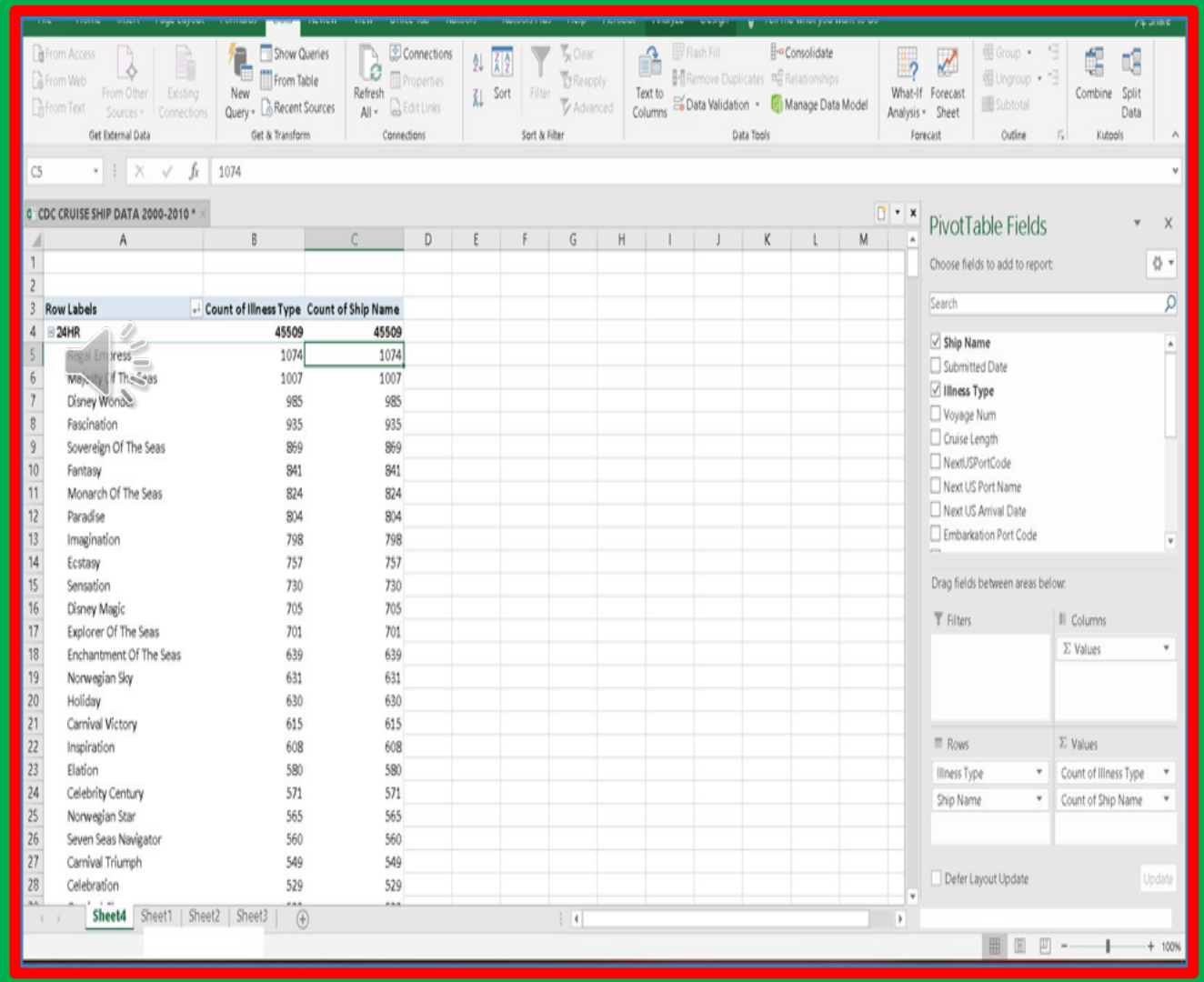




How To Build A Pivot Table

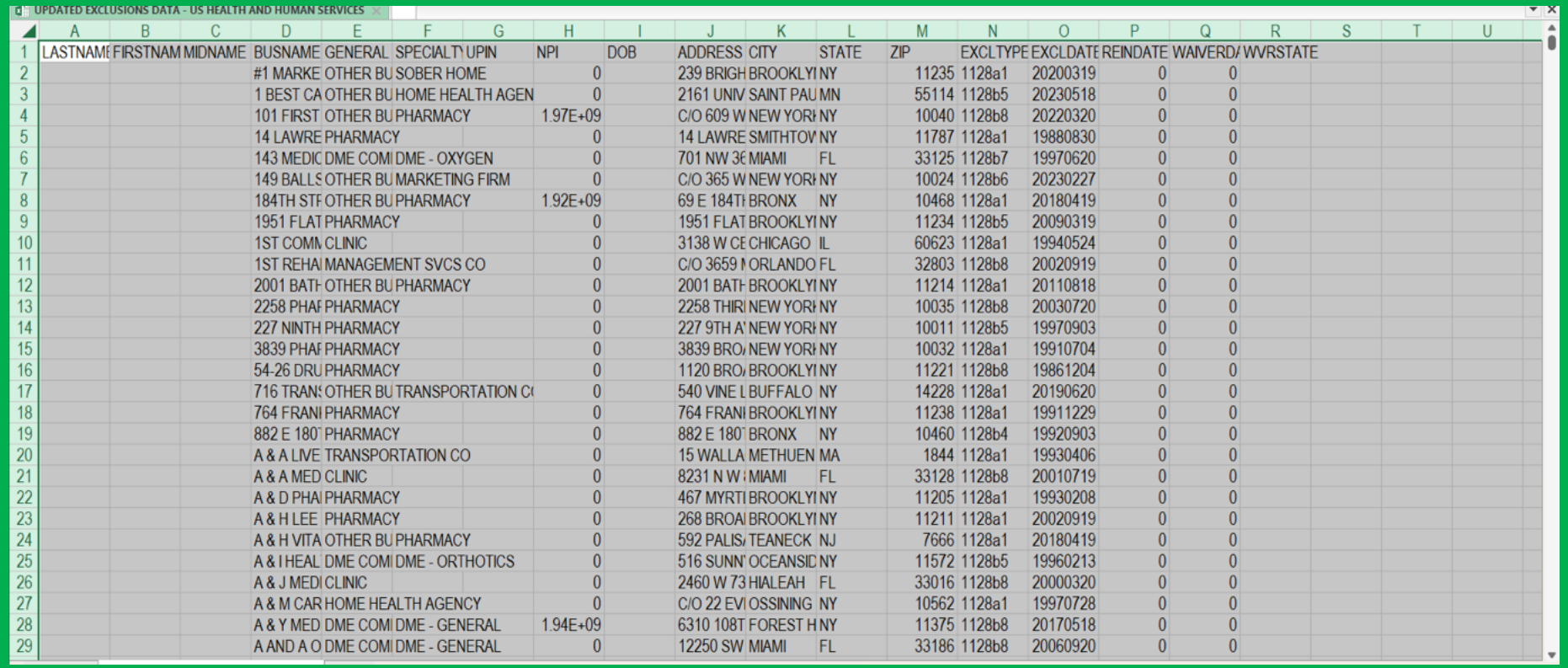


- You build PIVOT TABLES by dragging and dropping your criteria from the FIELD LIST into the four boxes in your grid
- You can also right click and select your criteria instead of dragging and dropping
- Learn how to make simple PIVOT TABLES first
- Then you can layer additional criteria into your table from the FIELD LIST to create more complex PIVOT TABLES



Open the File “Updated HHS Exclusions Database - CT”

Do a four-corners check to determine how many records you have and what data is available



UPDATED EXCLUSIONS DATA - US HEALTH AND HUMAN SERVICES

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	LASTNAME	FIRSTNAM	MIDNAME	BUSNAME	GENERAL	SPECIALTY	UPIN	NPI	DOB	ADDRESS	CITY	STATE	ZIP	EXCLTYPE	EXCLDATE	REINDATE	WAIVERD/	WVRSTATE			
2				#1 MARKE	OTHER BU	SOBER HOME		0		239 BRIGH	BROOKLYI	NY	11235	1128a1	20200319	0	0				
3				1 BEST CA	OTHER BU	HOME HEALTH AGEN		0		2161 UNIV	SAINT PAU	MN	55114	1128b5	20230518	0	0				
4				101 FIRST	OTHER BU	PHARMACY		1.97E+09		C/O 609 W	NEW YORI	NY	10040	1128b8	20220320	0	0				
5				14 LAWRE	PHARMACY			0		14 LAWRE	SMITHTOV	NY	11787	1128a1	19880830	0	0				
6				143 MEDIC	DME COMI	DME - OXYGEN		0		701 NW 3(	MIAMI	FL	33125	1128b7	19970620	0	0				
7				149 BALLS	OTHER BU	MARKETING FIRM		0		C/O 365 W	NEW YORI	NY	10024	1128b6	20230227	0	0				
8				184TH STF	OTHER BU	PHARMACY		1.92E+09		69 E 184TI	BRONX	NY	10468	1128a1	20180419	0	0				
9				1951 FLAT	PHARMACY			0		1951 FLAT	BROOKLYI	NY	11234	1128b5	20090319	0	0				
10				1ST COMM	CLINIC			0		3138 W CE	CHICAGO	IL	60623	1128a1	19940524	0	0				
11				1ST REHA	MANAGEMENT SVCS CO			0		C/O 3659 I	ORLANDO	FL	32803	1128b8	20020919	0	0				
12				2001 BATH	OTHER BU	PHARMACY		0		2001 BATH	BROOKLYI	NY	11214	1128a1	20110818	0	0				
13				2258 PHAI	PHARMACY			0		2258 THIRI	NEW YORI	NY	10035	1128b8	20030720	0	0				
14				227 NINTH	PHARMACY			0		227 9TH A	NEW YORI	NY	10011	1128b5	19970903	0	0				
15				3839 PHAI	PHARMACY			0		3839 BRO/	NEW YORI	NY	10032	1128a1	19910704	0	0				
16				54-26 DRU	PHARMACY			0		1120 BRO/	BROOKLYI	NY	11221	1128b8	19861204	0	0				
17				716 TRANS	OTHER BU	TRANSPORTATION C(		0		540 VINE L	BUFFALO	NY	14228	1128a1	20190620	0	0				
18				764 FRANI	PHARMACY			0		764 FRANI	BROOKLYI	NY	11238	1128a1	19911229	0	0				
19				882 E 180	PHARMACY			0		882 E 180	BRONX	NY	10460	1128b4	19920903	0	0				
20				A & A LIVE	TRANSPORTATION CO			0		15 WALLA	METHUEN	MA	1844	1128a1	19930406	0	0				
21				A & A MED	CLINIC			0		8231 N W	MIAMI	FL	33128	1128b8	20010719	0	0				
22				A & D PHAI	PHARMACY			0		467 MYRTI	BROOKLYI	NY	11205	1128a1	19930208	0	0				
23				A & H LEE	PHARMACY			0		268 BROAI	BROOKLYI	NY	11211	1128a1	20020919	0	0				
24				A & H VITA	OTHER BU	PHARMACY		0		592 PALIS/	TEANECK	NJ	7666	1128a1	20180419	0	0				
25				A & I HEAL	DME COMI	DME - ORTHOTICS		0		516 SUNN	OCEANSID	NY	11572	1128b5	19960213	0	0				
26				A & J MEDI	CLINIC			0		2460 W 73	HIALEAH	FL	33016	1128b8	20000320	0	0				
27				A & M CAR	HOME HEALTH AGENCY			0		C/O 22 EVI	OSSINING	NY	10562	1128a1	19970728	0	0				
28				A & Y MED	DME COMI	DME - GENERAL		1.94E+09		6310 108T	FOREST H	NY	11375	1128b8	20170518	0	0				
29				A AND A O	DME COMI	DME - GENERAL		0		12250 SW	MIAMI	FL	33186	1128b8	20060920	0	0				

Exclusions Program

OIG has the authority to exclude individuals and entities from Federally funded health care programs.

Exclusions Program
Online Searchable Database
LEIE Downloadable Databases
Monthly Supplement Downloads
Quick Tips
Waivers
Background Information
Applying for Reinstatement
Contact the Exclusions Program

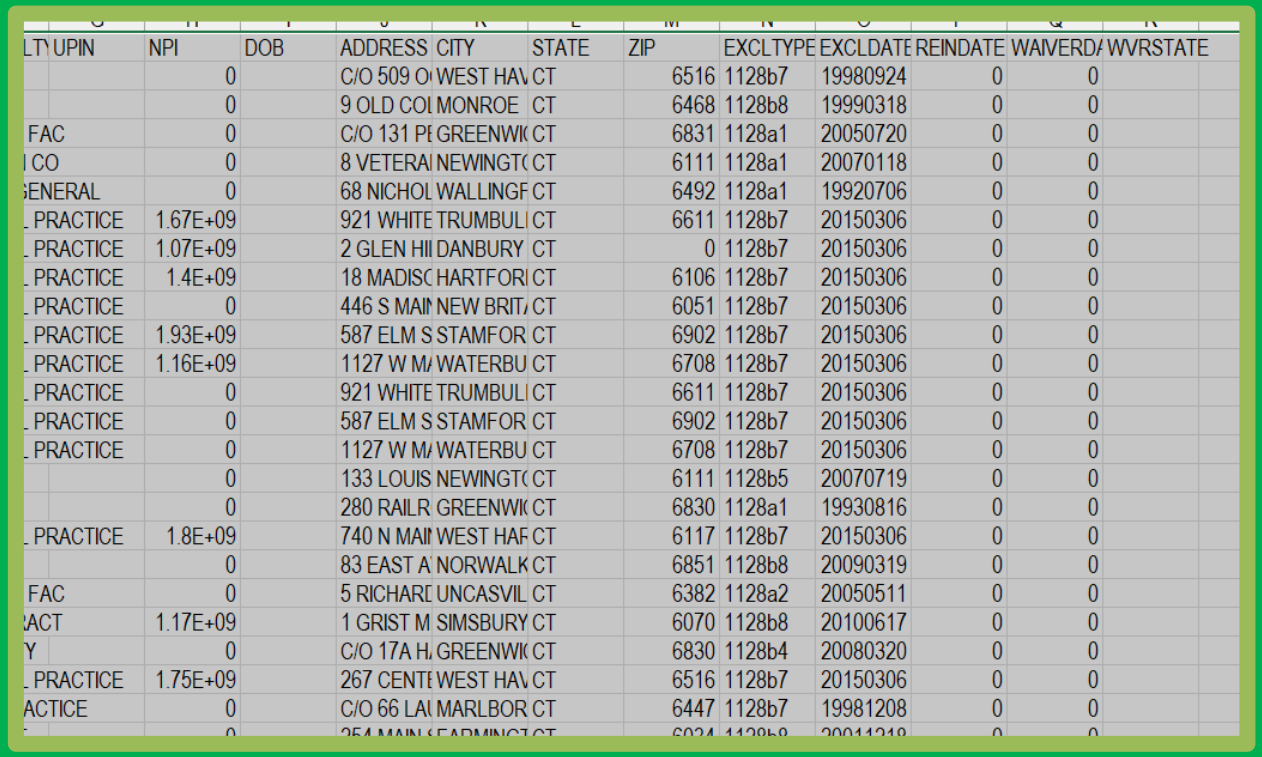
This webpage provides information about OIG's exclusion authority and activities. OIG has the authority to exclude individuals and entities from Federally funded health care programs for a variety of reasons, including a conviction for Medicare or Medicaid fraud. Those that are excluded can receive no payment from Federal health care programs for any items or services they furnish, order, or prescribe. This includes those that provide health benefits funded directly or indirectly by the United States (other than the Federal Employees Health Benefits Plan).

OIG maintains a list of all currently excluded individuals and entities called the List of Excluded Individuals/Entities (LEIE). Anyone who hires an individual or entity on the LEIE may be subject to civil monetary penalties (CMP). To avoid CMP liability, health care entities should routinely check the list to ensure that new hires and current employees are not on it.

Help imp

How To Build A Pivot Table

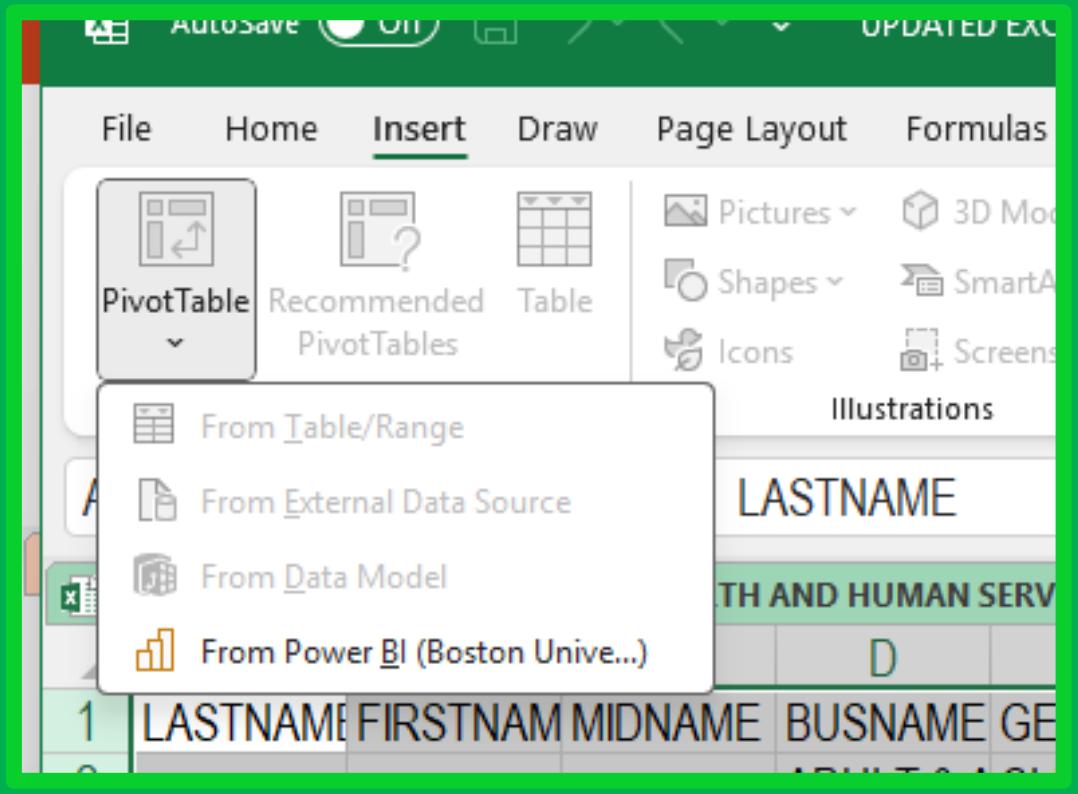
- In Your CT Sheet:
- **SELECT** Your Data
- On A PC, Put Your Cursor Somewhere In Your Data And Select CTRL A
- On A MAC, Use CMND A



LT	UPIN	NPI	DOB	ADDRESS	CITY	STATE	ZIP	EXCLTYPE	EXCLDATE	REINDATE	WAIVERDA	WVRSTATE
		0		C/O 509 O	WEST HAV	CT	6516	1128b7	19980924	0	0	
		0		9 OLD COI	MONROE	CT	6468	1128b8	19990318	0	0	
	FAC	0		C/O 131 PI	GREENWI	CT	6831	1128a1	20050720	0	0	
	CO	0		8 VETERAI	NEWINGT	CT	6111	1128a1	20070118	0	0	
	ENERAL	0		68 NICHOL	WALLINGF	CT	6492	1128a1	19920706	0	0	
	PRACTICE	1.67E+09		921 WHITE	TRUMBUL	CT	6611	1128b7	20150306	0	0	
	PRACTICE	1.07E+09		2 GLEN HI	DANBURY	CT	0	1128b7	20150306	0	0	
	PRACTICE	1.4E+09		18 MADISC	HARTFOR	CT	6106	1128b7	20150306	0	0	
	PRACTICE	0		446 S MAI	NEW BRIT	CT	6051	1128b7	20150306	0	0	
	PRACTICE	1.93E+09		587 ELM S	STAMFOR	CT	6902	1128b7	20150306	0	0	
	PRACTICE	1.16E+09		1127 W M	WATERBU	CT	6708	1128b7	20150306	0	0	
	PRACTICE	0		921 WHITE	TRUMBUL	CT	6611	1128b7	20150306	0	0	
	PRACTICE	0		587 ELM S	STAMFOR	CT	6902	1128b7	20150306	0	0	
	PRACTICE	0		1127 W M	WATERBU	CT	6708	1128b7	20150306	0	0	
		0		133 LOUIS	NEWINGT	CT	6111	1128b5	20070719	0	0	
		0		280 RAILR	GREENWI	CT	6830	1128a1	19930816	0	0	
	PRACTICE	1.8E+09		740 N MAI	WEST HAF	CT	6117	1128b7	20150306	0	0	
		0		83 EAST A	NORWALK	CT	6851	1128b8	20090319	0	0	
	FAC	0		5 RICHARI	UNCASVIL	CT	6382	1128a2	20050511	0	0	
	ACT	1.17E+09		1 GRIST M	SIMSBURY	CT	6070	1128b8	20100617	0	0	
	Y	0		C/O 17A H	GREENWI	CT	6830	1128b4	20080320	0	0	
	PRACTICE	1.75E+09		267 CENTI	WEST HAV	CT	6516	1128b7	20150306	0	0	
	ACTICE	0		C/O 66 LAI	MARLBOR	CT	6447	1128b7	19981208	0	0	

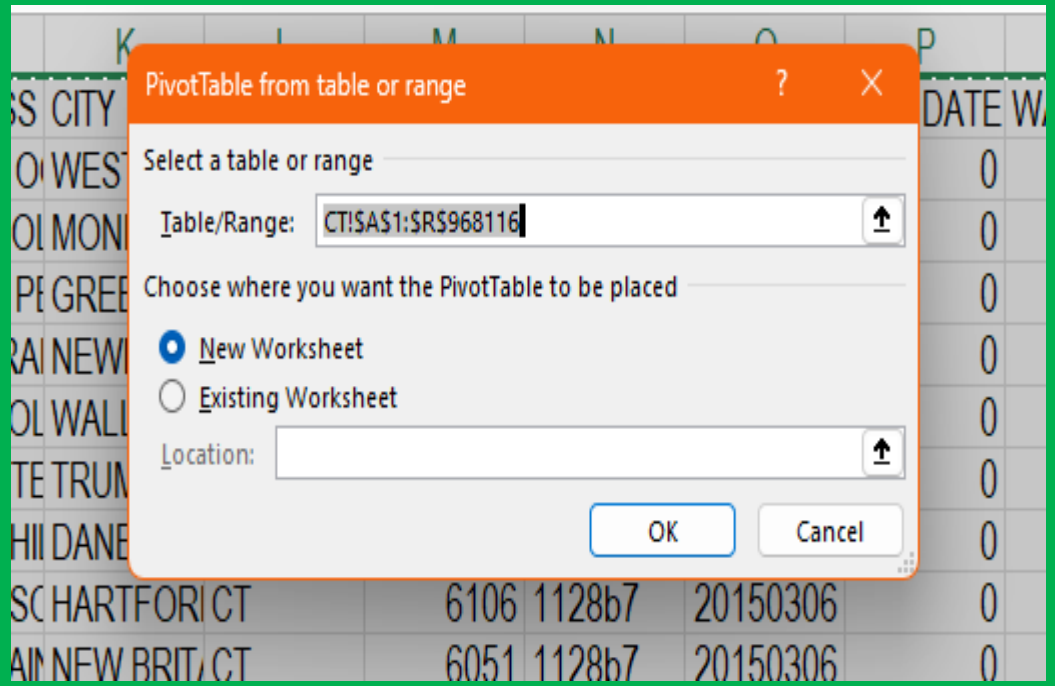
How To Build A Pivot Table

- **INSERT A PIVOT TABLE**
- On a PC, go to **INSERT** and click on **PIVOT TABLE**



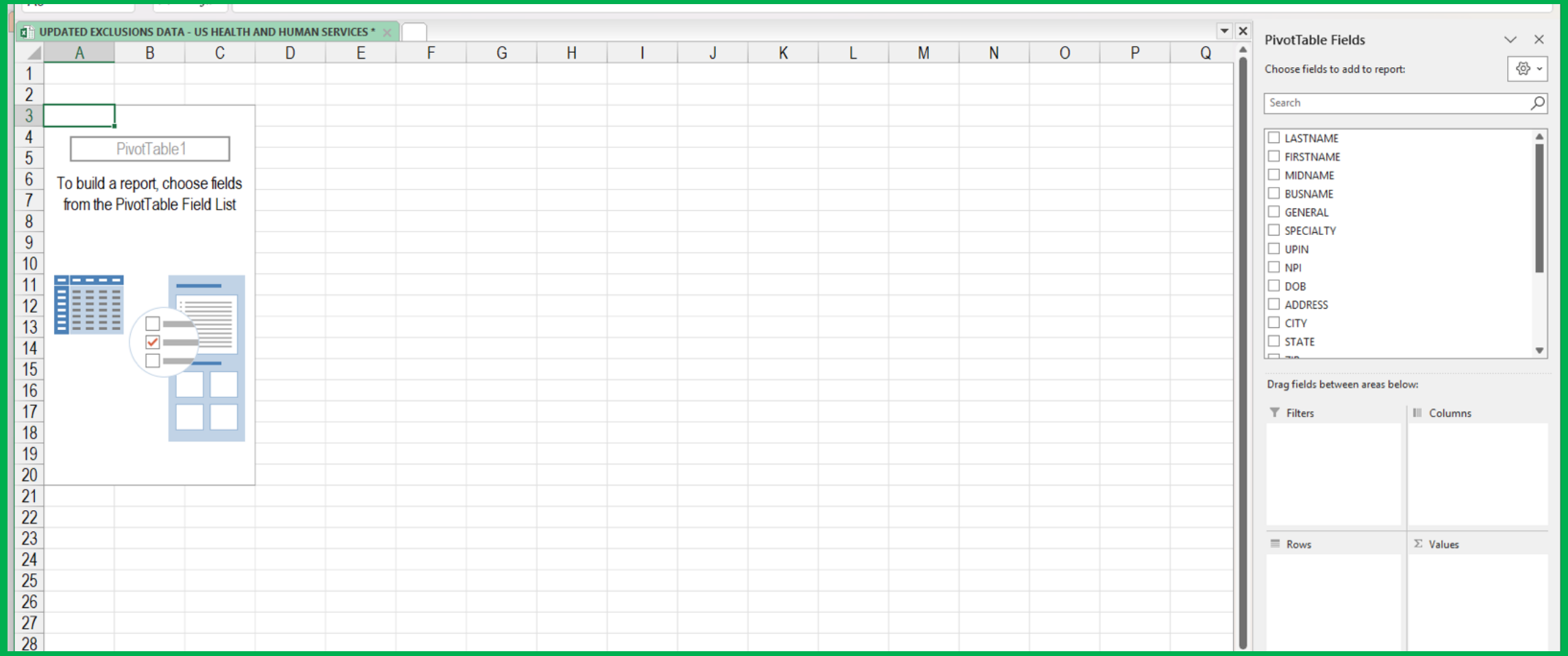
How To Build A Pivot Table

- Double check that your full range of data is selected
- You'll get an option for where to place your **PIVOT TABLE**
- Select **NEW WORKSHEET**
- CLICK **OK**



How To Build A Pivot Table

- All the fields appear as criteria for your analysis
- You can analyze by rows, values, filters or columns
- Let's start with rows

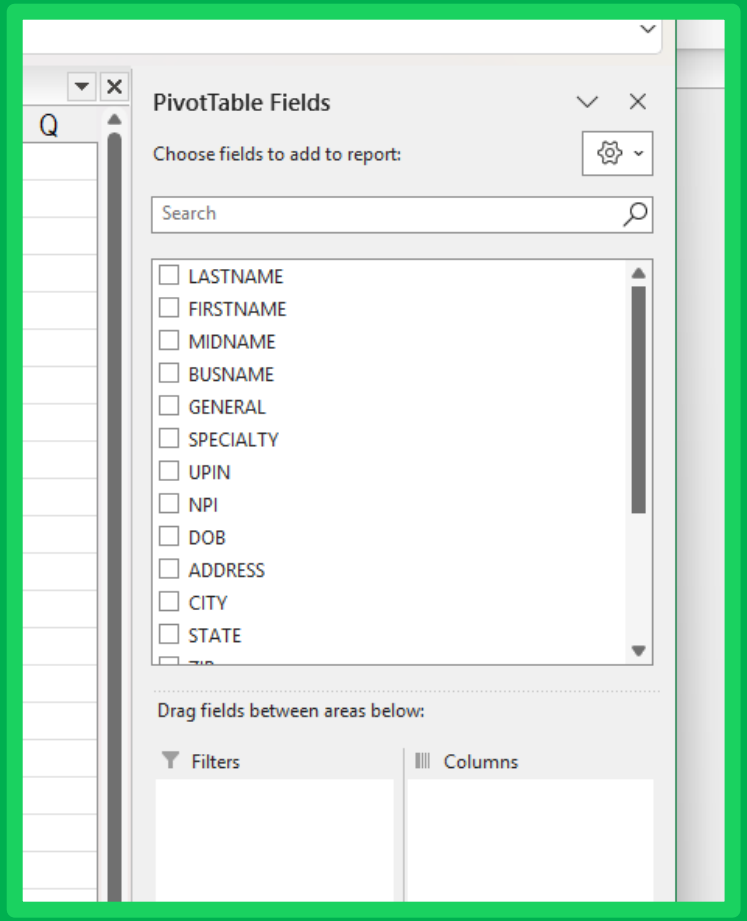


How To Build A Pivot Table

You'll see a PIVOT TABLE PANEL

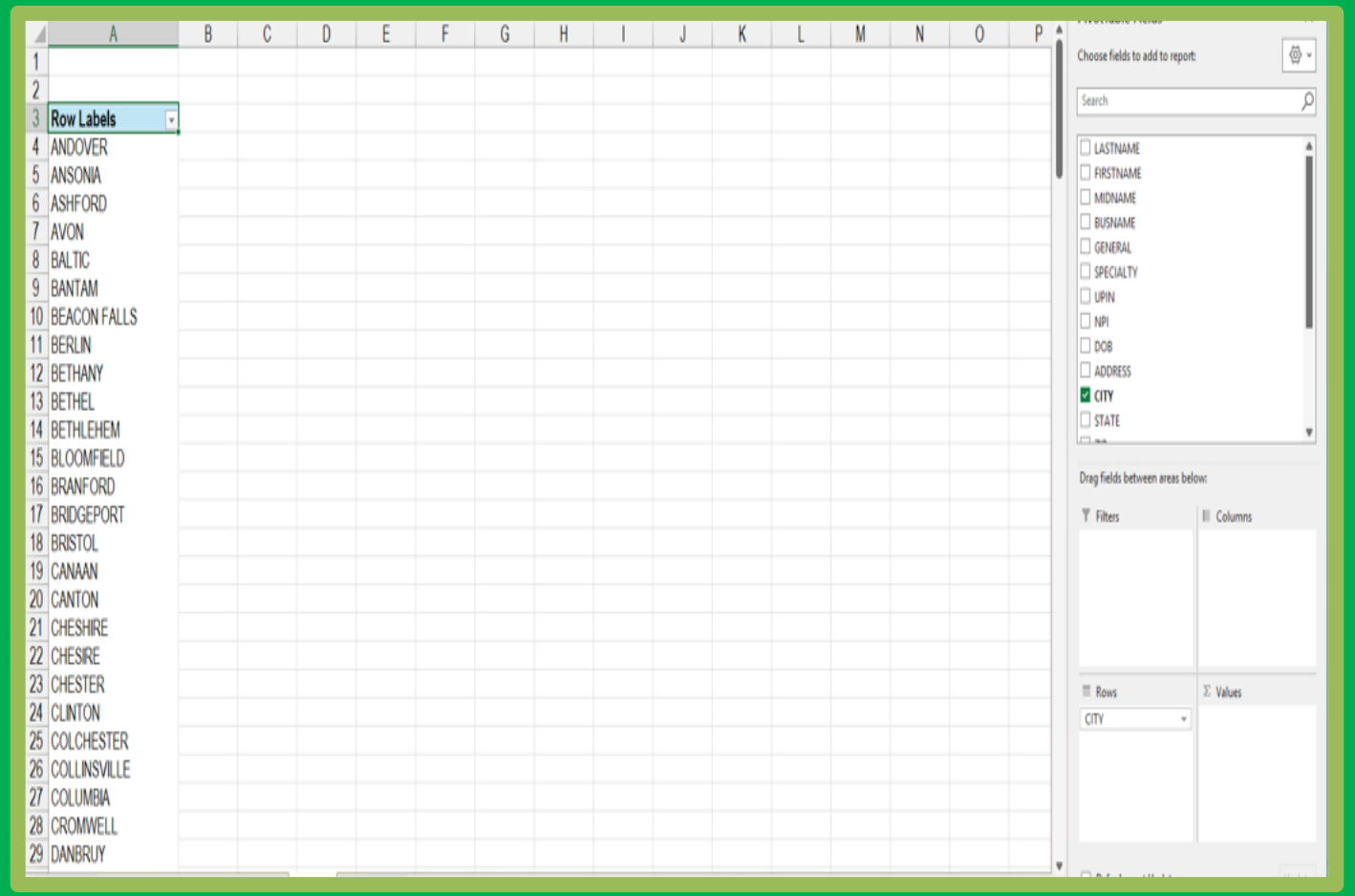
There will be a grid that will allow you to select criteria

This is called the PIVOT TABLE FIELD LIST



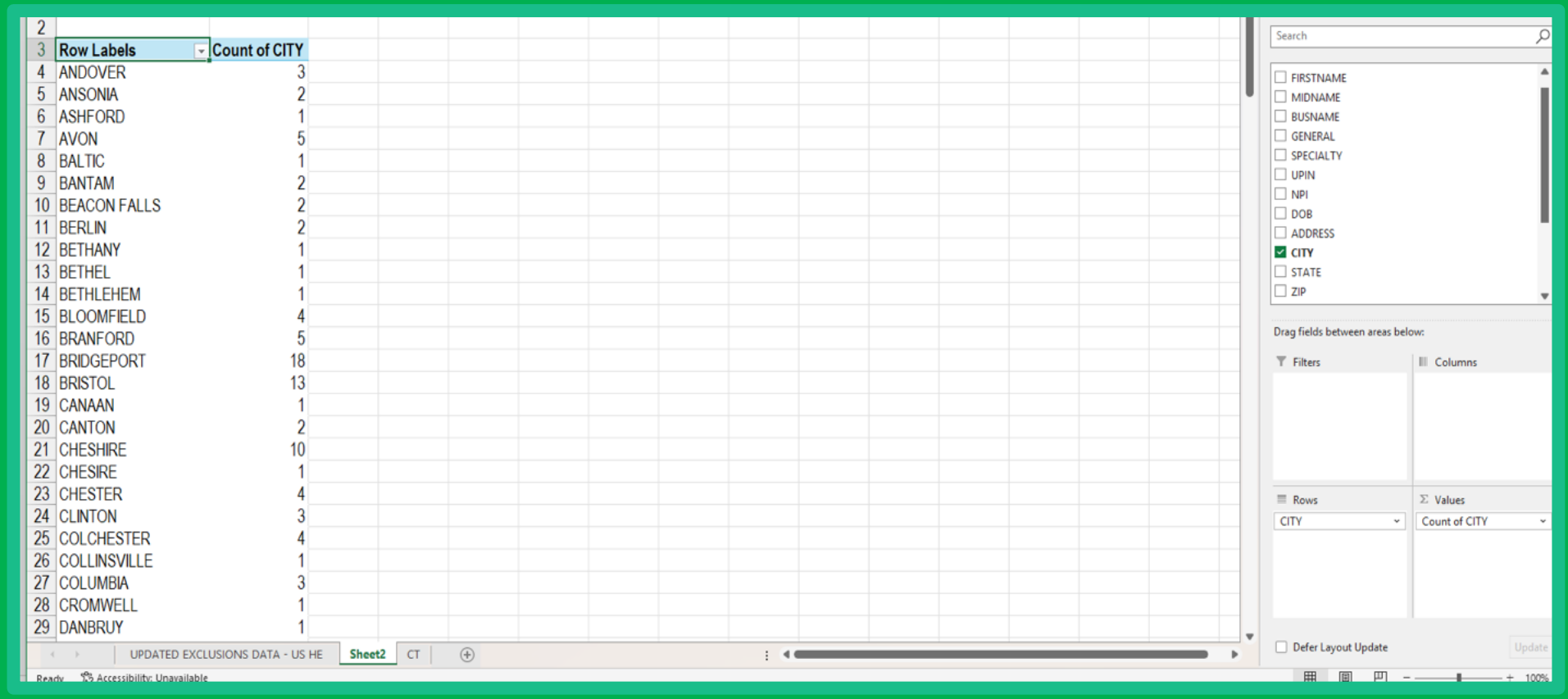
How To Build A Pivot Table

- You'll drag and drop your criteria into the grid to pivot your data
- You can also right click on a field name and add it to rows, values, columns or filters
- We'll start by dragging city to rows



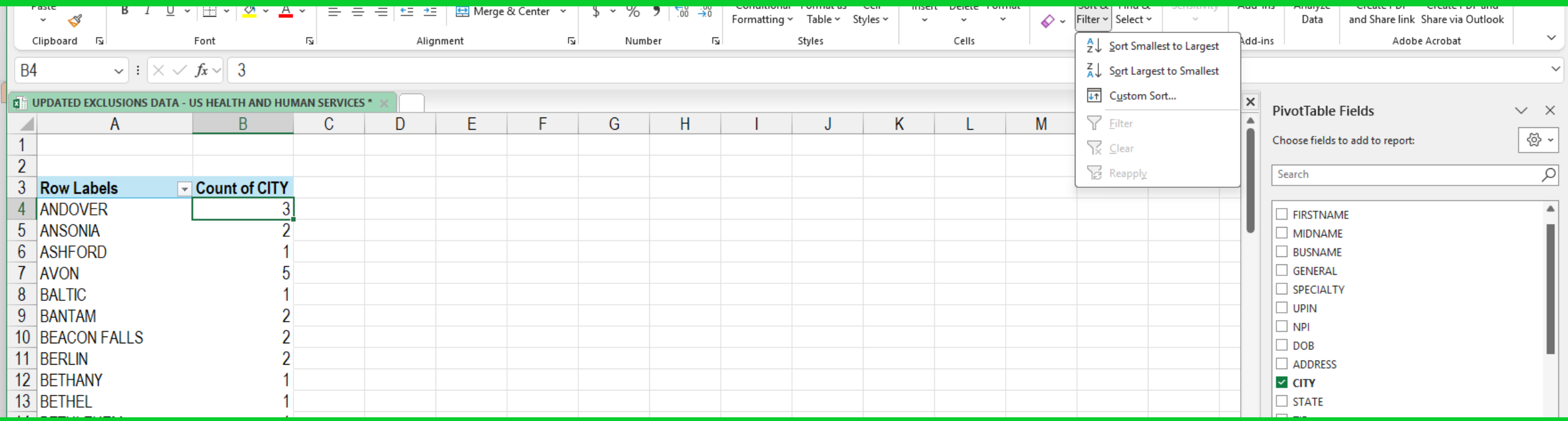
Let's Look At Which City Had The Most Businesses, Organizations or Individuals Excluded The Most Times

- Drag City to Rows
- Then Drag it Again to Values
- Excel Will Default to Count
- You are Counting the Number of Times City Appears In The Data



Row Labels	Count of CITY
ANDOVER	3
ANSONIA	2
ASHFORD	1
AVON	5
BALTIC	1
BANTAM	2
BEACON FALLS	2
BERLIN	2
BETHANY	1
BETHEL	1
BETHLEHEM	1
BLOOMFIELD	4
BRANFORD	5
BRIDGEPORT	18
BRISTOL	13
CANAAN	1
CANTON	2
CHESHIRE	10
CHESIRE	1
CHESTER	4
CLINTON	3
COLCHESTER	4
COLLINSVILLE	1
COLUMBIA	3
CROMWELL	1
DANBRUY	1

We Want to Sort This in Descending Order
Put Your Cursor on the First Number – 3
Then Go to Sort on the Ribbon and Sort This In Descending Order



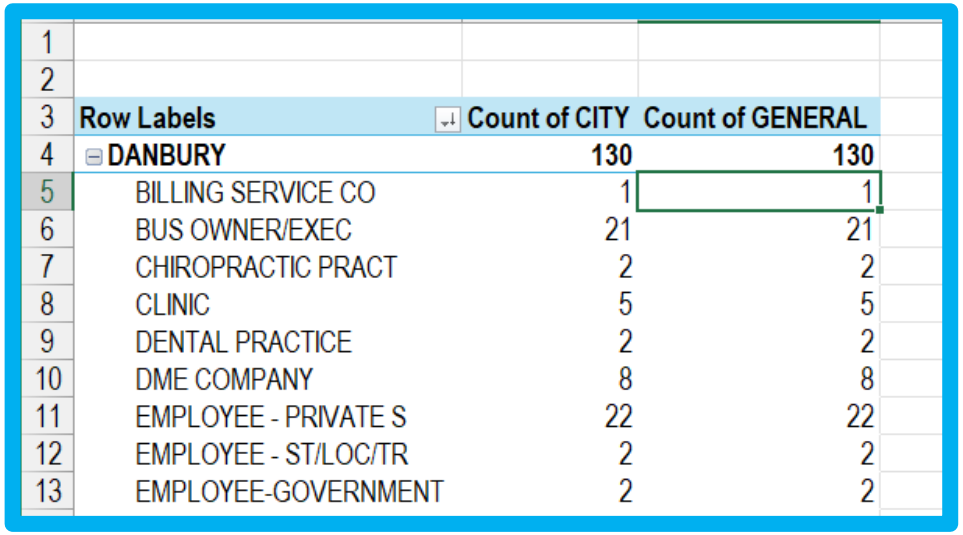
Row Labels	Count of CITY
ANDOVER	3
ANSONIA	2
ASHFORD	1
AVON	5
BALTIC	1
BANTAM	2
BEACON FALLS	2
BERLIN	2
BETHANY	1
BETHEL	1

What Do You Have To Be Careful Of?
What Could a Lede Be?

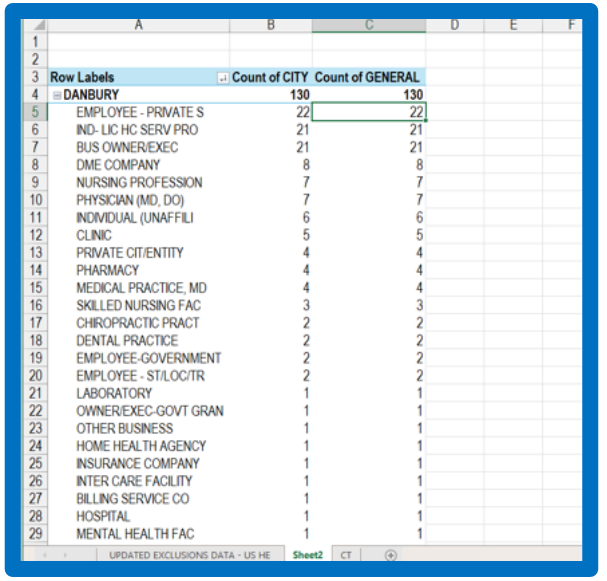
Row Labels	Count of CITY
DANBURY	130
STAMFORD	20
WATERBURY	19
BRIDGEPORT	18
HARTFORD	17
WEST HAVEN	16
NEW BRITAIN	14
NEW HAVEN	14
STRATFORD	13
BRISTOL	13
WALLINGFORD	12
TORRINGTON	11
WEST HARTFORD	11
HAMDEN	11
MERIDEN	10
ENFIELD	10
CHESHIRE	10
MANCHESTER	9
EAST HARTFORD	9
GREENWICH	9
NEWINGTON	9
NEW MILFORD	9

Adding Multiple Values

- Let's Look at What Types of Businesses Have Been Sanctioned In These Cities
- Drag General to Our Rows
- Drag It Again to Our Values
- Now Put Your Cursor on the First Number Showing a Type of Business
- Sort That in Descending Order



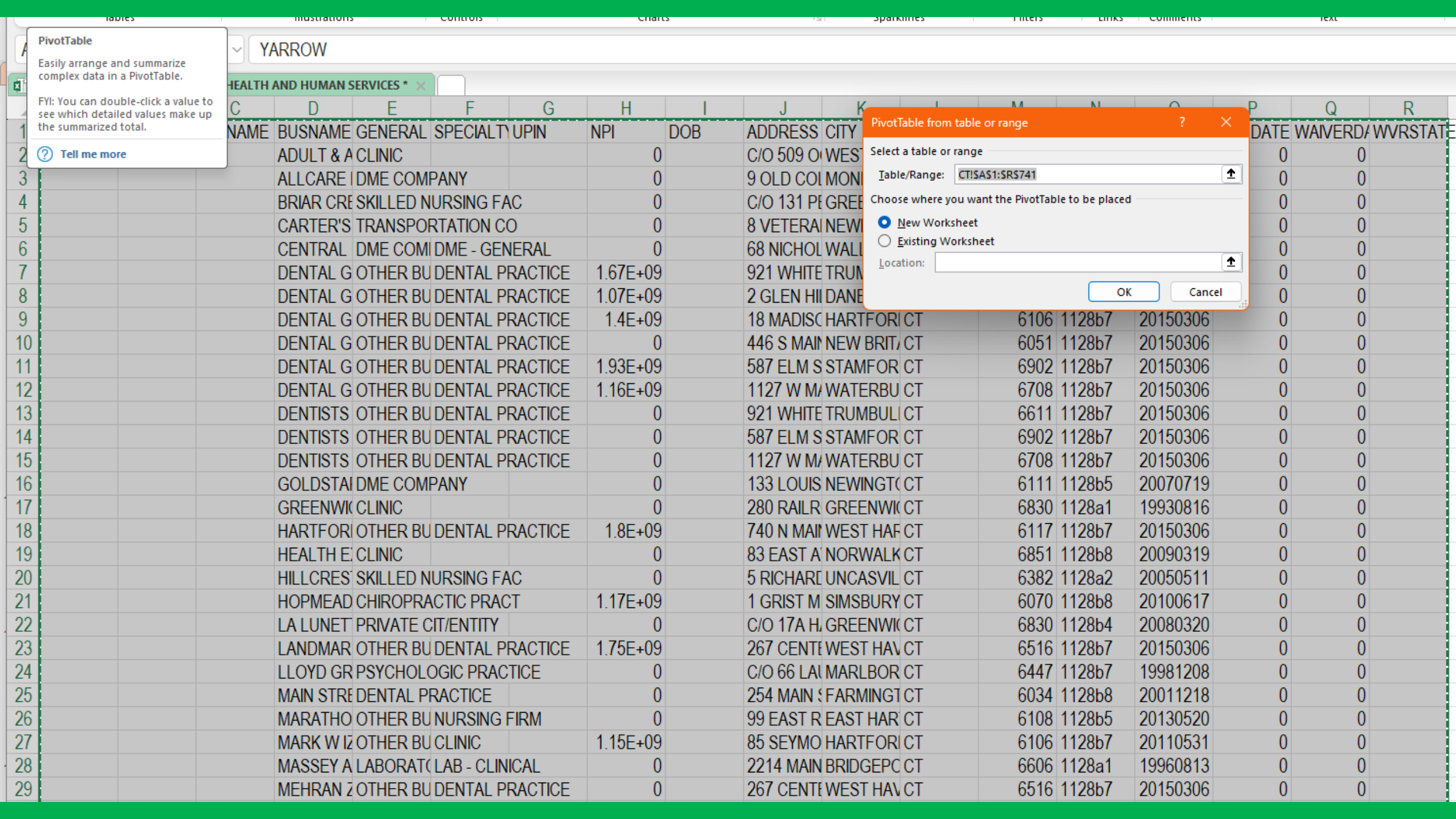
Row Labels	Count of CITY	Count of GENERAL
DANBURY	130	130
BILLING SERVICE CO	1	1
BUS OWNER/EXEC	21	21
CHIROPRACTIC PRACT	2	2
CLINIC	5	5
DENTAL PRACTICE	2	2
DME COMPANY	8	8
EMPLOYEE - PRIVATE S	22	22
EMPLOYEE - ST/LOC/TR	2	2
EMPLOYEE-GOVERNMENT	2	2



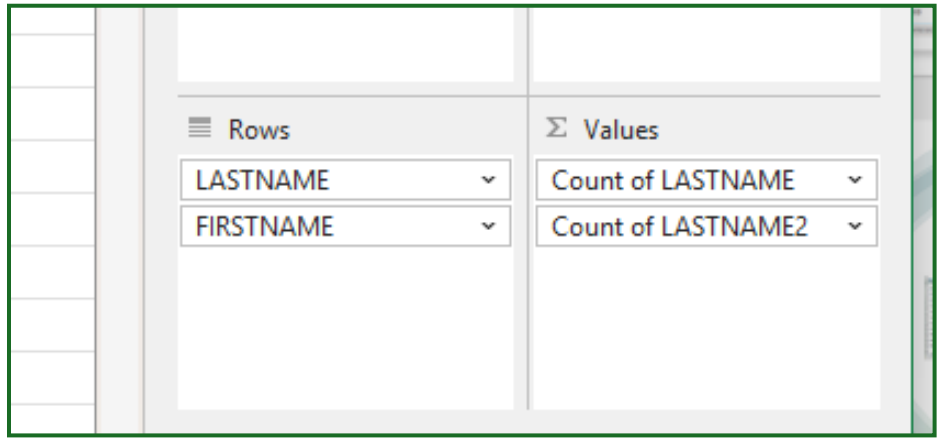
Row Labels	Count of CITY	Count of GENERAL
DANBURY	130	130
EMPLOYEE - PRIVATE S	22	22
IND- LIC HC SERV PRO	21	21
BUS OWNER/EXEC	21	21
DME COMPANY	8	8
NURSING PROFESSION	7	7
PHYSICIAN (MD, DO)	7	7
INDIVIDUAL (UNAFFILI	6	6
CLINIC	5	5
PRIVATE CIT/ENTITY	4	4
PHARMACY	4	4
MEDICAL PRACTICE, MD	4	4
SKILLED NURSING FAC	3	3
CHIROPRACTIC PRACT	2	2
DENTAL PRACTICE	2	2
EMPLOYEE-GOVERNMENT	2	2
EMPLOYEE - ST/LOC/TR	2	2
LABORATORY	1	1
OWNER/EXEC-GOVT GRAN	1	1
OTHER BUSINESS	1	1
HOME HEALTH AGENCY	1	1
INSURANCE COMPANY	1	1
INTER CARE FACILITY	1	1
BILLING SERVICE CO	1	1
HOSPITAL	1	1
MENTAL HEALTH FAC	1	1

Let's Create a New Pivot Table

- You can right click on the one you have and name it
- Then go back to the original CT data
- Select all of the data, and insert a new pivot table
- Make sure you get the right data
- Then you will put it in a new workbook



	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	NAME	BUSNAME	GENERAL	SPECIALTY	UPIN	NPI	DOB	ADDRESS	CITY					DATE	WAIVERDA	WVRSTATE
2		ADULT & A	CLINIC			0		C/O 509 O	WEST					0	0	
3		ALLCARE I	DME COMPANY			0		9 OLD COI	MONI					0	0	
4		BRIAR CRE	SKILLED NURSING FAC			0		C/O 131 PI	GREE					0	0	
5		CARTER'S	TRANSPORTATION CO			0		8 VETERAI	NEWI					0	0	
6		CENTRAL	DME COMI	DME - GENERAL		0		68 NICHOL	WALL					0	0	
7		DENTAL G	OTHER BU	DENTAL PRACTICE		1.67E+09		921 WHITE	TRUM					0	0	
8		DENTAL G	OTHER BU	DENTAL PRACTICE		1.07E+09		2 GLEN HII	DANE					0	0	
9		DENTAL G	OTHER BU	DENTAL PRACTICE		1.4E+09		18 MADISC	HARTFORI	CT	6106	1128b7	20150306	0	0	
10		DENTAL G	OTHER BU	DENTAL PRACTICE		0		446 S MAI	NEW BRIT/	CT	6051	1128b7	20150306	0	0	
11		DENTAL G	OTHER BU	DENTAL PRACTICE		1.93E+09		587 ELM S	STAMFOR	CT	6902	1128b7	20150306	0	0	
12		DENTAL G	OTHER BU	DENTAL PRACTICE		1.16E+09		1127 W M/	WATERBU	CT	6708	1128b7	20150306	0	0	
13		DENTISTS	OTHER BU	DENTAL PRACTICE		0		921 WHITE	TRUMBULI	CT	6611	1128b7	20150306	0	0	
14		DENTISTS	OTHER BU	DENTAL PRACTICE		0		587 ELM S	STAMFOR	CT	6902	1128b7	20150306	0	0	
15		DENTISTS	OTHER BU	DENTAL PRACTICE		0		1127 W M/	WATERBU	CT	6708	1128b7	20150306	0	0	
16		GOLDSTAI	DME COMPANY			0		133 LOUIS	NEWINGT(	CT	6111	1128b5	20070719	0	0	
17		GREENWI(	CLINIC			0		280 RAILR	GREENWI(	CT	6830	1128a1	19930816	0	0	
18		HARTFORI	OTHER BU	DENTAL PRACTICE		1.8E+09		740 N MAI	WEST HAF	CT	6117	1128b7	20150306	0	0	
19		HEALTH E:	CLINIC			0		83 EAST A	NORWALK	CT	6851	1128b8	20090319	0	0	
20		HILLCRES	SKILLED NURSING FAC			0		5 RICHARI	UNCASVIL	CT	6382	1128a2	20050511	0	0	
21		HOPMEAD	CHIROPRACTIC PRACT			1.17E+09		1 GRIST M	SIMSBURY	CT	6070	1128b8	20100617	0	0	
22		LA LUNET	PRIVATE CIT/ENTITY			0		C/O 17A H/	GREENWI(	CT	6830	1128b4	20080320	0	0	
23		LANDMAR	OTHER BU	DENTAL PRACTICE		1.75E+09		267 CENTI	WEST HAV	CT	6516	1128b7	20150306	0	0	
24		LLOYD GR	PSYCHOLOGIC PRACTICE			0		C/O 66 LA(	MARLBOR	CT	6447	1128b7	19981208	0	0	
25		MAIN STRI	DENTAL PRACTICE			0		254 MAIN S	FARMINGT	CT	6034	1128b8	20011218	0	0	
26		MARATHO	OTHER BU	NURSING FIRM		0		99 EAST R	EAST HAR	CT	6108	1128b5	20130520	0	0	
27		MARK W IZ	OTHER BU	CLINIC		1.15E+09		85 SEYMO	HARTFORI	CT	6106	1128b7	20110531	0	0	
28		MASSEY A	LABORAT(	LAB - CLINICAL		0		2214 MAIN	BRIDGEPO	CT	6606	1128a1	19960813	0	0	
29		MEHRAN Z	OTHER BU	DENTAL PRACTICE		0		267 CENTI	WEST HAV	CT	6516	1128b7	20150306	0	0	



Let's Look at The Individuals Who Have Been Sanctioned

Row Labels	Count of LASTNAME	Count of LASTNAME2
ABRAMS	1	1
LEE	1	1
ADAMS-PULEO	1	1
LISA	1	1
AFFRUNTI	1	1
DANIEL	1	1
AGRO	1	1
LINDA	1	1
AHMED	2	2
SHEIKH	2	2
ALEXANDER	1	1
LISA	1	1
ALLEN	1	1
COLLEEN	1	1
ALTRECHE	1	1
NAKISHA	1	1
ALVARADO	1	1
BRENDA	1	1
AMICO	1	1
RICHARD	1	1
ANASTASIO	2	2
ANDREW	2	2
ANDERSON	2	2
MEGAN	1	1
SHERWOOD	1	1
ANDRADE	1	1

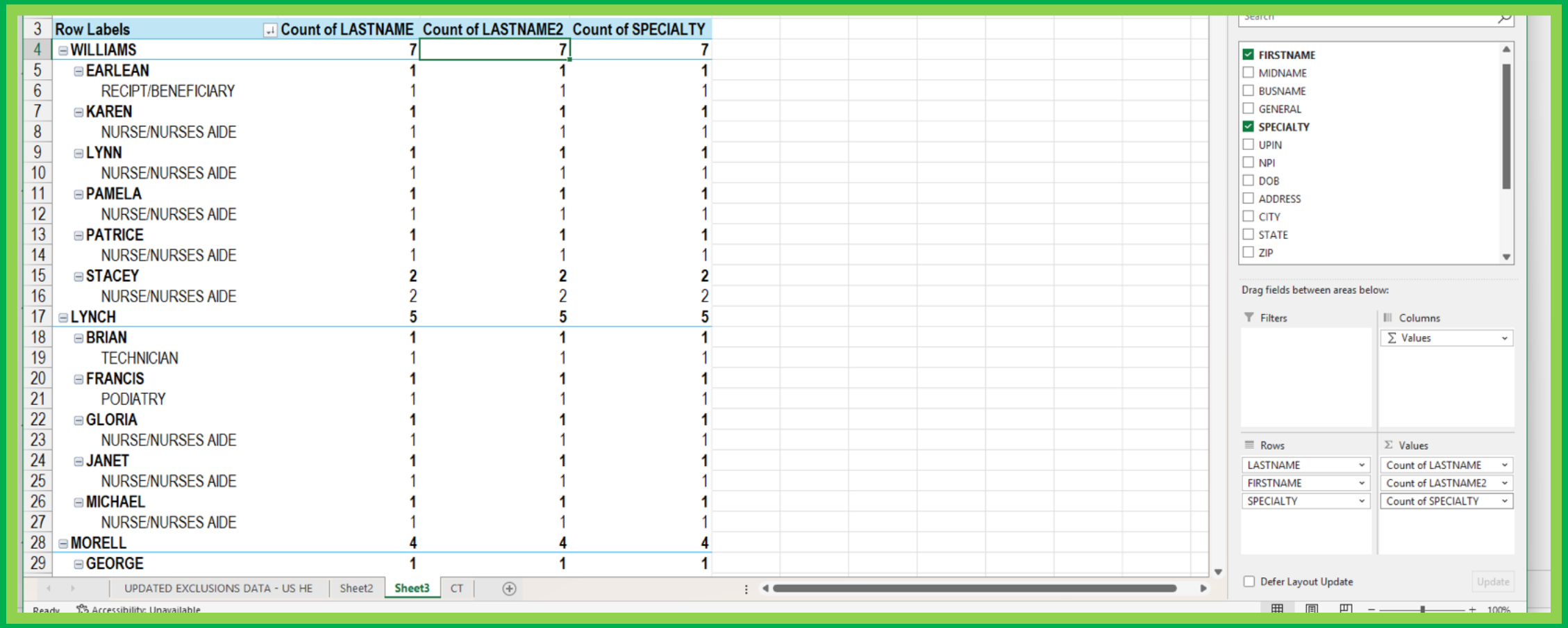
UPDATED EXCLUSIONS DATA - US HE | Sheet2 | Sheet3 | CT

- Drag Last Name to Your Rows AND Values
- Drag First Name to Your Rows AND Values

Now Sort This in Descending Order – Are There Any Repeats?

Row Labels	Count of LASTNAME	Count of LASTNAME2
WILLIAMS	**7**	**7**
EARLEAN	1	1
KAREN	1	1
LYNN	1	1
PAMELA	1	1
PATRICE	1	1
STACEY	2	2
LYNCH	**5**	**5**
BRIAN	1	1
FRANCIS	1	1
GLORIA	1	1
JANET	1	1
MICHAEL	1	1
MORELL	**4**	**4**
GEORGE	1	1
JOHN	1	1
KAREN	1	1
ROBERT	1	1
JONES	**4**	**4**
ANDREA	1	1
ANN	1	1
STEPHANIE	1	1
VINCENT	1	1
MILLER	**3**	**3**
KIMBERLY	1	1
REBECCA	1	1

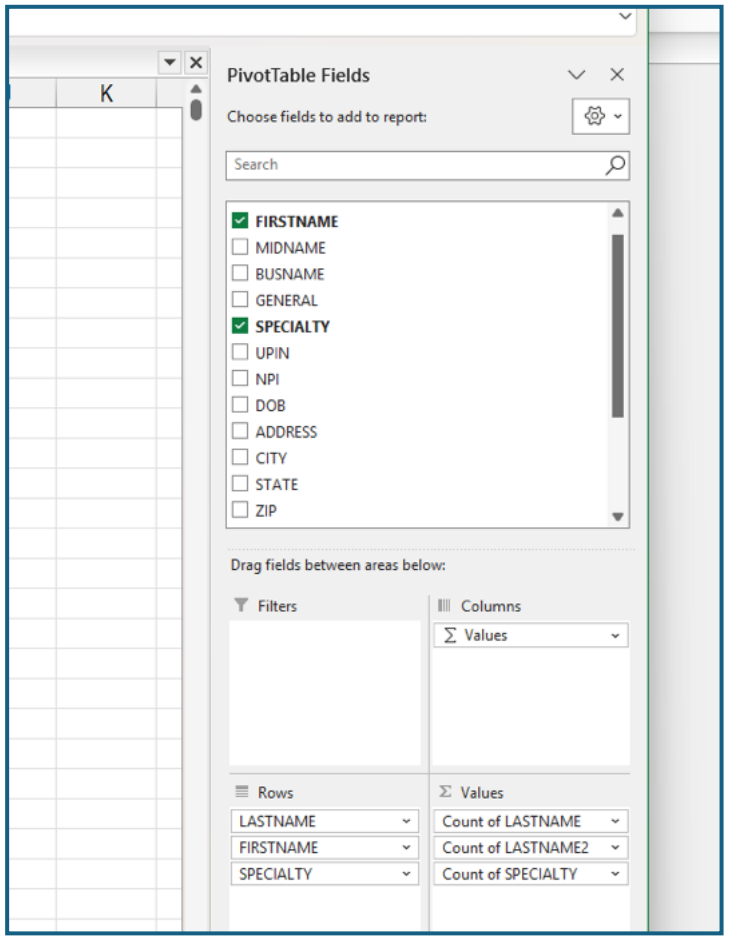
Now Add The Specialty By Dragging It To Rows and Values – You Can Also Sort This in Descending Order



Row Labels	Count of LASTNAME	Count of LASTNAME2	Count of SPECIALTY
WILLIAMS	7	7	7
EARLEAN	1	1	1
RECIPT/BENEFICIARY	1	1	1
KAREN	1	1	1
NURSE/NURSES AIDE	1	1	1
LYNN	1	1	1
NURSE/NURSES AIDE	1	1	1
PAMELA	1	1	1
NURSE/NURSES AIDE	1	1	1
PATRICE	1	1	1
NURSE/NURSES AIDE	1	1	1
STACEY	2	2	2
NURSE/NURSES AIDE	2	2	2
LYNCH	5	5	5
BRIAN	1	1	1
TECHNICIAN	1	1	1
FRANCIS	1	1	1
PODIATRY	1	1	1
GLORIA	1	1	1
NURSE/NURSES AIDE	1	1	1
JANET	1	1	1
NURSE/NURSES AIDE	1	1	1
MICHAEL	1	1	1
NURSE/NURSES AIDE	1	1	1
MORELL	4	4	4
GEORGE	1	1	1

Clear Your Pivot Table by Unchecking Your Selections

Repeat Your Steps But Look Only At Businesses



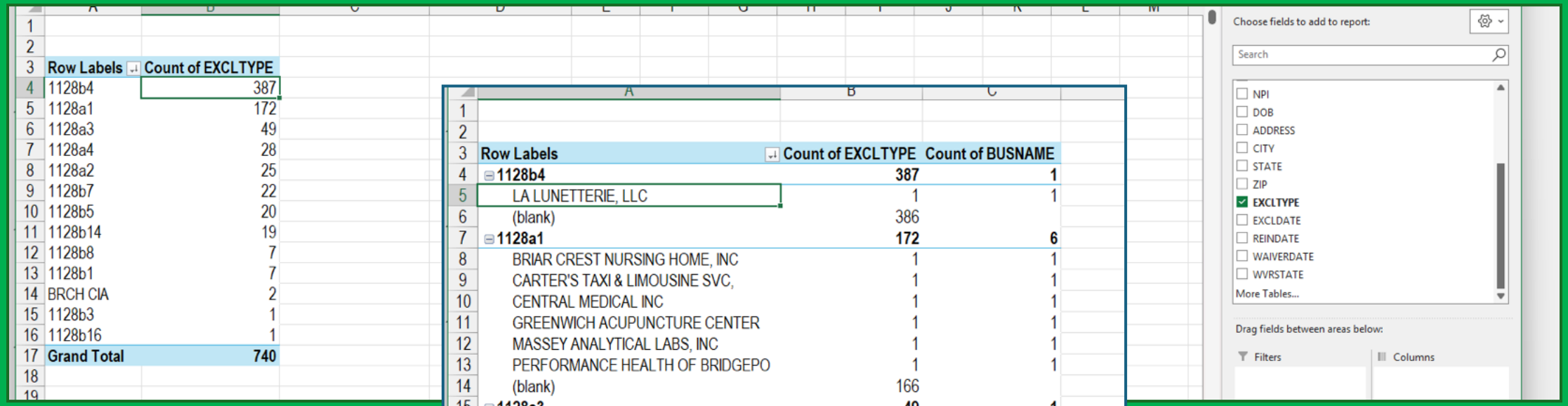
Row Labels	Count of BUSNAME
P & D SURGICAL, INC	1
MAIN STREET DENTAL ASSOCIATES	1
VAUGHAN'S PHARMACY, INC	1
ALLCARE HEALTH SERVICES, INC	1
MEHRAN ZAMANI, DDS, PC	1
BRIAR CREST NURSING HOME, INC	1
R G PHARMACY	1
CARTER'S TAXI & LIMOUSINE SVC,	1
LANDMARK DENTAL	1
CENTRAL MEDICAL INC	1
MARK W IZARD, M D, P C	1
DENTAL GROUP OF CONNECTICUT, L	1
ORANGE CENTER PHARMACY, INC	1
DENTAL GROUP OF DANBURY, PC	1
PETER J EMBRIANO, M D, P C	1
DENTAL GROUP OF HARTFORD, PC	1
THATCHER'S PHARMACY, INC	1
DENTAL GROUP OF NEW BRITIAN, P	1
ADULT & ADOLESCENT PSYCHIATRY	1
DENTAL GROUP OF STAMFORD, LLC	1
LLOYD GROUP LTD	1
DENTAL GROUP OF WATERBURY, LLC	1
MARATHON HEALTHCARE GROUP, LLC	1
DENTISTS GROUP OF CONNECTICUT,	1
MASSEY ANALYTICAL LABS, INC	1
DENTISTS GROUP OF STAMFORD, PC	1

Your Turn to Practice

- Now Try it Again
- Clear Out Your Pivot Table
- Choose Something Else to Pivot
- I Am Pivoting on **Type of Exclusion**
- I Looked Up the Codes on The HHS Website

1128(b)(1)(B)	1320a-7(b)(1)(B)	Conviction relating to fraud in non-health care programs. Baseline Period: 3 years
1128(b)(2)	1320a-7(b)(2)	Conviction relating to obstruction of an investigation or audit. Baseline Period: 3 years
1128(b)(3)	1320a-7(b)(3)	Misdemeanor conviction relating to controlled substance. Baseline Period: 3 years
1128(b)(4)	1320a-7(b)(4)	License revocation, suspension, or surrender. Minimum Period: Period imposed by the state licensing authority.
1128(b)(5)	1320a-7(b)(5)	Exclusion or suspension under federal or state health care program. Minimum Period: No less than the period imposed by federal or state health care program.
1128(b)(6)	1320a-7(b)(6)	Claims for excessive charges, unnecessary services or services which fail to meet professionally recognized standards of health care, or failure of an HMO to furnish medically necessary services. Minimum Period: 1 year
1128(b)(7)	1320a-7(b)(7)	Fraud, kickbacks, and other prohibited activities. Minimum Period: None
1128(b)(8)	1320a-7(b)(8)	Entities controlled by a sanctioned individual. Minimum Period: Same as length of individual's exclusion.
1128(b)(8)(A)	1320a-7(b)(8)(A)	Entities controlled by a family or household member of an excluded individual and where there has been a transfer of ownership/control. Minimum Period: Same as length of individual's exclusion.

I Pivoted on Column N – Exclusion Type
I Then Added Business Name



Row Labels	Count of EXCLTYPE
1128b4	387
1128a1	172
1128a3	49
1128a4	28
1128a2	25
1128b7	22
1128b5	20
1128b14	19
1128b8	7
1128b1	7
BRCH CIA	2
1128b3	1
1128b16	1
Grand Total	**740**

Row Labels	Count of EXCLTYPE	Count of BUSNAME
1128b4	**387**	**1**
LA LUNETTERIE, LLC	1	1
(blank)	386	
1128a1	**172**	**6**
BRIAR CREST NURSING HOME, INC	1	1
CARTER'S TAXI & LIMOUSINE SVC,	1	1
CENTRAL MEDICAL INC	1	1
GREENWICH ACUPUNCTURE CENTER	1	1
MASSEY ANALYTICAL LABS, INC	1	1
PERFORMANCE HEALTH OF BRIDGEPO	1	1
(blank)	166	
1128a3	**49**	**1**
PURDUE FREDERICK COMPANY, INC	1	1
(blank)	48	
1128a4	**28**	
(blank)	28	
1128a2	**25**	**1**
HILLCREST HEALTH CARE CENTER	1	1
(blank)	24	
1128b7	**22**	**19**
ADULT & ADOLESCENT PSYCHIATRY	1	1
DENTAL GROUP OF CONNECTICUT, L	1	1
DENTAL GROUP OF DANBURY, PC	1	1
DENTAL GROUP OF HARTFORD, PC	1	1
DENTAL GROUP OF NEW BRITIAN, P	1	1
DENTAL GROUP OF STAMFORD, LLC	1	1

LA LUNETTERIE LLC

Connecticut Secretary Of State Business Registration · Updated 11/28/2024

Sponsored Links

Write Review Upgrade Claim

LA LUNETTERIE LLC is a Connecticut LLC filed on November 16, 2004. The company's filing status is listed as Forfeited and its File Number is 0802116.

The Registered Agent on file for this company is Michael Baker and is located at 71 Lewis Street, Greenwich, CT 06830. The company's principal address is 71 Lewis Street, Greenwich, CT 06830 and its mailing address is 71 Lewis St., Greenwich, CT 06830.

The company has 1 contact on record. The contact is Michael Baker from Greenwich CT.

Like 33K

Company Information

Company Name:	LA LUNETTERIE LLC
Entity Type:	CONNECTICUT LLC
File Number:	0802116
Filing State:	Connecticut (CT)
Filing Status:	Forfeited
Filing Date:	November 16, 2004
Company Age:	20 Years, 1 Month
Registered Agent:	Michael Baker 71 Lewis Street Greenwich, CT 06830
Principal Address:	71 Lewis Street Greenwich, CT 06830
Mailing Address:	71 Lewis St. Greenwich, CT 06830
Governing Agency:	Connecticut Secretary of State

Sponsored Links

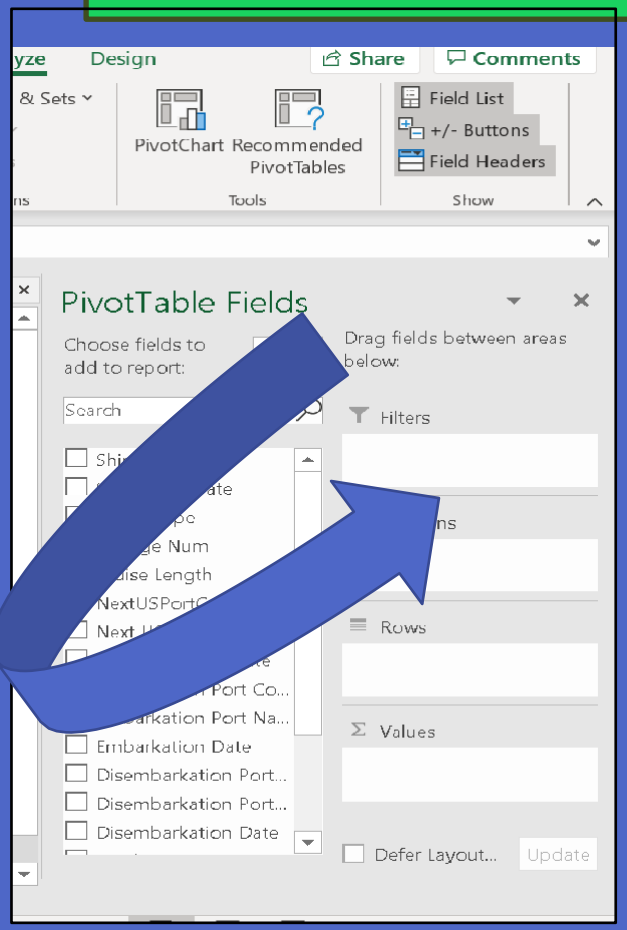
Company Contacts

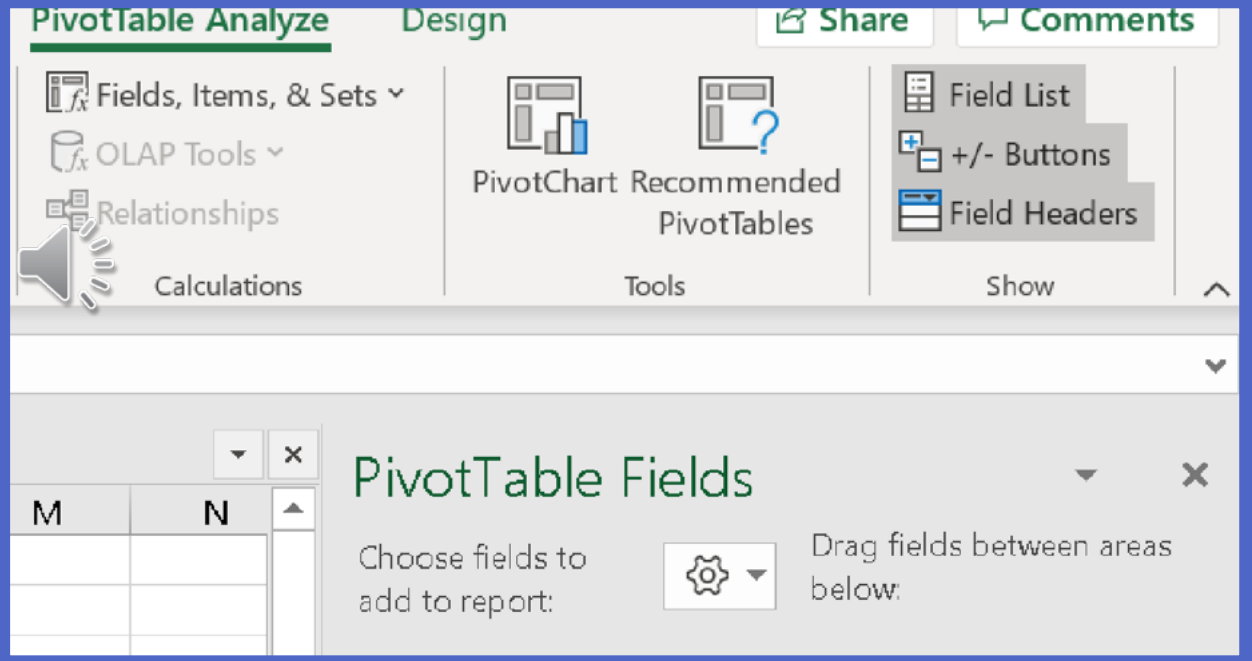
MICHAEL BAKER
Member
71 Lewis Street
Greenwich, CT 06830

Spreadsheet Basics

Pivot Table Tips

To remove the Pivot Table Field list, click **close** under the drop-down arrow in the right-hand corner





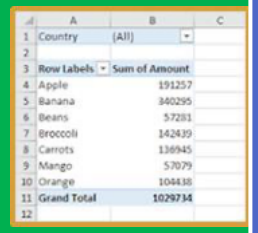
If you need it back, choose “Field List” from the Pivot Table Analysis on the top

	A	B	C	D	E	F	G	H	I	J	K	L	
1	Ship Name	Submitted	Illness Typ	Voyage Nu	Cruise Len	NextUSPo	Next US Pc	Next US A	Embarkati	Embarkati	Embarkati	Disembarl	Dis
2	Disney Ma	########	24HR		7	PCF	Port Cana	########	PCF	Port Cana	########	PCF	Po
3	Celebrity (	########	24HR		7	KWE	Key West,	########	PEV	Port Everg	########	PEV	Po
4	Fantasy	########	24HR		4	PCF	Port Cana	########	PCF	Port Cana	########	PCF	Po
5	Celebrity (	########	24HR		7	MIA	Port of Mi	########	MIA	Port of Mi	########	MIA	Po
6	Paradise	########	24HR				Port of Mi	########	MIA	Port of Mi	########	MIA	Po
7	Disney Wo	########	24HR				Port Cana	########	PCF	Port Cana	########	PCF	Po
8	Fascinatio	########	24HR				San Juan,	########	SJU	San Juan,	########	SJU	Sa
9	Holiday	########	4HR				Los Angele	########	LAX	Los Angele	########	LAX	Lo
10	Jubilee - In	########	24HR				Tampa, FL	########	TAM	Tampa, FL	########	TAM	Ta
11	Sensation	########	24HR				Tampa, FL	########	TAM	Tampa, FL	########	TAM	Ta
12	Fascinatio	########	4HR		7	SJU	San Juan,	########	SJU	San Juan,	########	SJU	Sa
13	Disney Ma	########	4HR		7	PCF	Port Cana	########	PCF	Port Cana	########	PCF	Po
14	Fantasy	########	24HR		3	PCF	Port Cana	########	PCF	Port Cana	########	PCF	Po
15	Paradise	########	24HR		7	MIA	Port of Mi	########	MIA	Port of Mi	########	MIA	Po
16	Carnival D	########	24HR		7	SJU	San Juan,	########	SJU	San Juan,	########	SJU	Sa
17	Carnival V	########	24HR		8	MIA	Port of Mi	########	MIA	Port of Mi	########	MIA	Po
18	Elation	########	24HR		7	LAX	Los Angele	########	LAX	Los Angele	########	LAX	Lo
19	Ecstasy	########	24HR		3	MIA	Port of Mi	########	MIA	Port of Mi	########	MIA	Po
20	Elation	########	4HR		7	LAX	Los Angele	########	LAX	Los Angele	########	LAX	Lo
21	Celebrity (	########	24HR		7	SJU	San Juan,	########	PEV	Port Everg	########	PEV	Po

Always name your sheets for easy organization and to avoid mistakes

Illness by Ship | Crew Cases Only | CDC CRUISE SHIP DATA 2000-2010

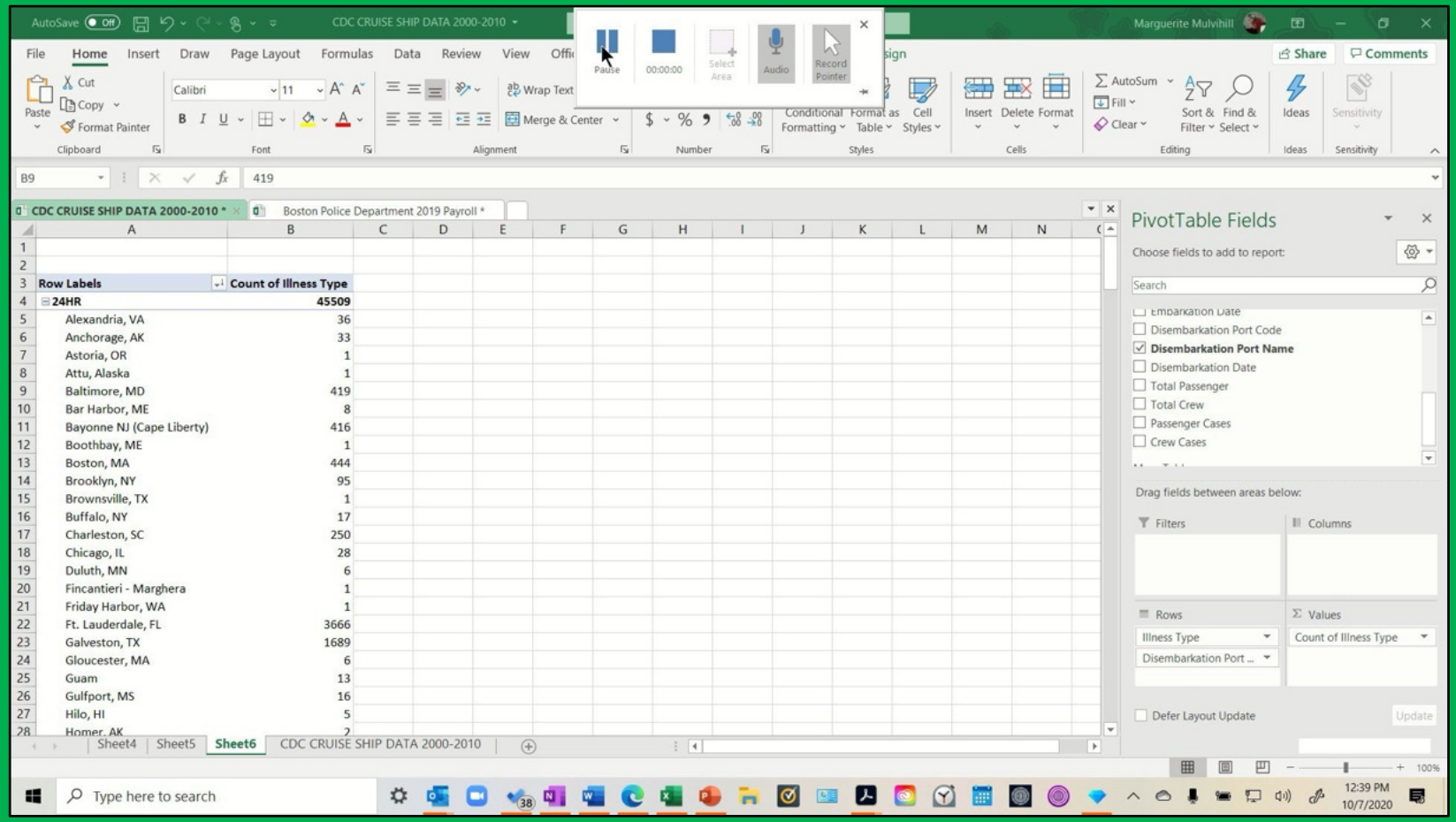
Pivot Table Tips



- Isolate and zoom in on specific data by double-clicking a tally within your pivot table
- This will open a whole new sheet for you that you can save as a new spreadsheet

3	Row Labels	Count of Illness Type
4	24HR	45509
5	Alexandria, VA	73
6	Anchorage, AK	5
7	Astoria, OR	2
8	Baltimore, MD	418
9	Bar Harbor, ME	8
10	Bayonne NJ (Cape Liberty)	409
11	Boston, MA	381
12	Brooklyn, NY	54
13	Buffalo, NY	6
14	Charleston, SC	241
15	Chicago, IL	24
16	Dutch Harbor, AK	3
17	Fincantieri - Marghera	3
18	Ft. Lauderdale, FL	3460
19	Galveston, TX	1688
20	Guam	3
21	Gulfport, MS	16

Zoom In On Your Findings



Row Labels	Count of Illness Type
24HR	45509
Alexandria, VA	36
Anchorage, AK	33
Astoria, OR	1
Attu, Alaska	1
Baltimore, MD	419
Bar Harbor, ME	8
Bayonne NJ (Cape Liberty)	416
Boothbay, ME	1
Boston, MA	444
Brooklyn, NY	95
Brownsville, TX	1
Buffalo, NY	17
Charleston, SC	250
Chicago, IL	28
Duluth, MN	6
Fincantieri - Marghera	1
Friday Harbor, WA	1
Ft. Lauderdale, FL	3666
Galveston, TX	1689
Gloucester, MA	6
Guam	13
Gulfport, MS	16
Hilo, HI	5
Homer, AK	2

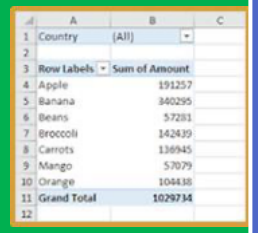
Date	Illness Type	Voyage Num	Cruise Length	NextUSPortCode	Next US Port Name	Next US Arrival Date	Embarkat
10 1:59	24HR	24100423	7	BOS	Boston, MA	4/30/2010	BOS
09 1:30	24HR	24091023	16	NOL	New Orleans, LA	11/8/2009	BOS
9 22:53	24HR	26233	10	MIA	Miami, FL	11/6/2009	BOS
09 4:31	24HR	24091023	16	STT	St. Thomas, VI	10/27/2009	BOS
9 18:59	24HR	26232	5	BOS	Boston, MA	10/27/2009	BOS
9 23:00	24HR	JW26231	5	BOS	Boston, MA	10/22/2009	BOS
9 19:42	24HR	26230	7	BOS	Boston, MA	10/17/2009	BOS
9 19:04	24HR	JW26229	7	BOS	Boston, MA	10/10/2009	BOS
9 22:29	24HR	JW26228	7	BOS	Boston, MA	10/3/2009	BOS
09 8:12	24HR	26227	7	BOS	Boston, MA	9/26/2009	BOS
9 18:49	24HR	26226	5	BOS	Boston, MA	9/19/2009	BOS
09 4:45	24HR		7	BOS	Boston, MA	9/18/2009	BOS
9 11:41	24HR		7	BOS	Boston, MA	9/11/2009	BOS
09 7:11	24HR		7	BAR	Bar Harbor, ME	9/5/2009	BOS
09 0:24	24HR		7	BOS	Boston, MA	9/4/2009	BOS
9 10:53	24HR		6	BOS	Boston, MA	8/29/2009	BOS
09 6:40	24HR		7	BOS	Boston, MA	8/28/2009	BOS
09 0:37	24HR		7	BOS	Boston, MA	8/21/2009	BOS
09 0:05	24HR		7	BOS	Boston, MA	8/14/2009	BOS
09 1:07	24HR	24090731	7	BOS	Boston, MA	8/7/2009	BOS

Make sure to rename your sheet if you want to save it

Crew Cases Only | **Sheet5** | Sheet3 | CDC CRUISE SHIP DATA 2000-2010

Average: 920340.0592 Count: 6621 Sum: 3267207210

Pivot Table Tips

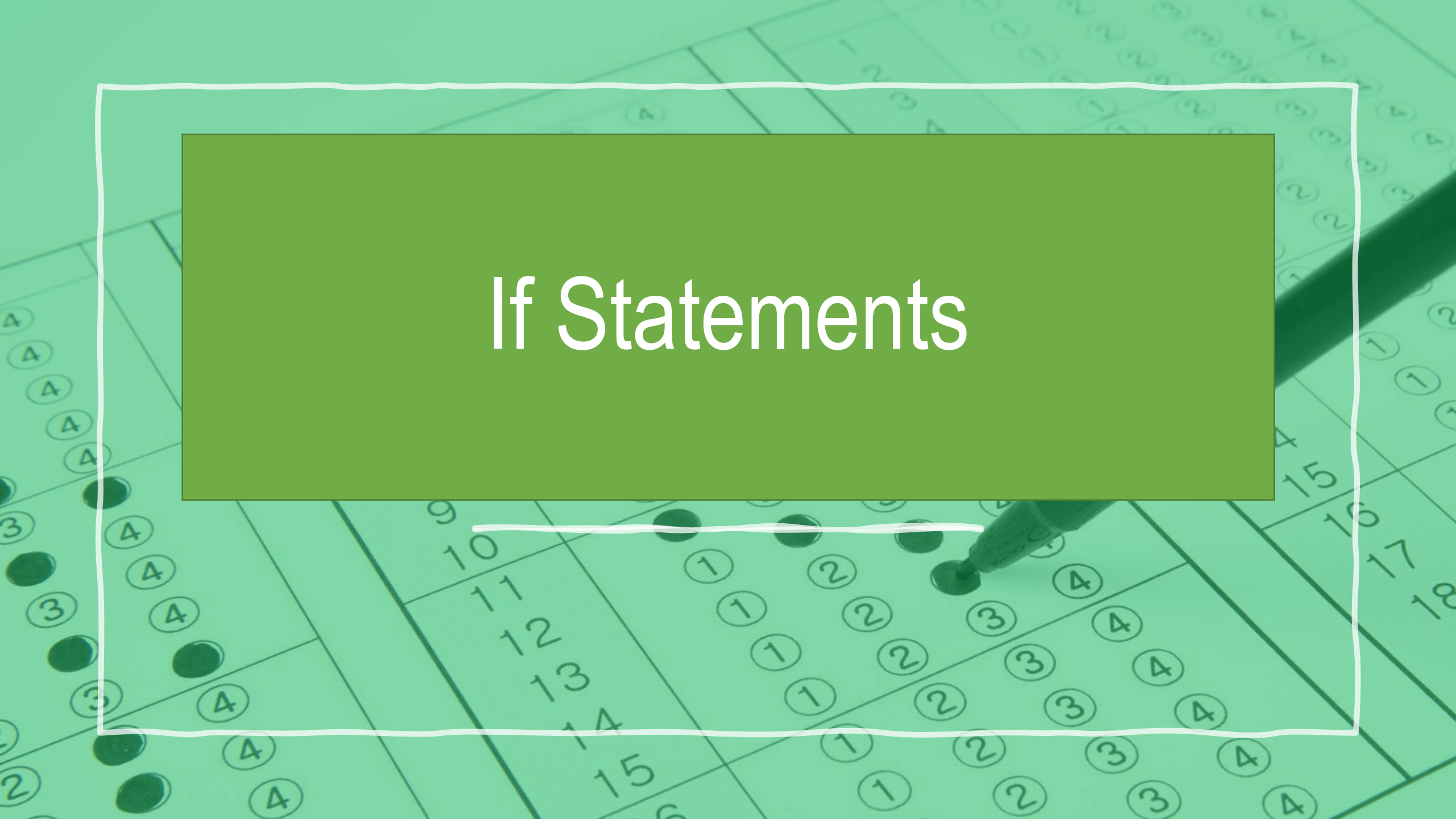


- You can also copy the field names and put them in a sheet so you can see them more quickly and just choose the ones you want to use
- To do this, copy the row of column names from your original data set

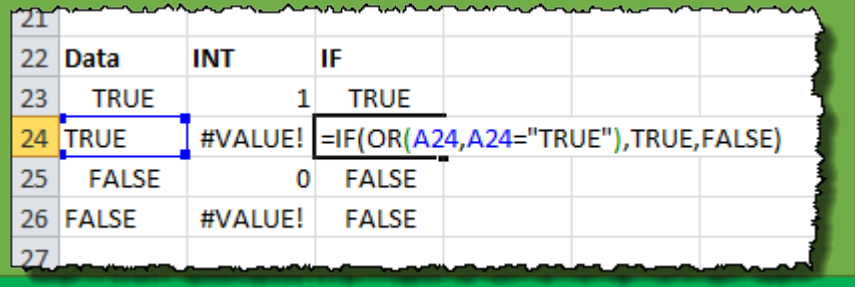
CDC CRUISE SHIP DATA 2000-2010 *

	A	B	C	D	E	F	G	H	I	J	K	L
1	Ship Name	Submitted	Illness Typ	Voyage Nu	Cruise Len	NextUSPo	Next US P	Next US A	Embarkati	Embarkati	Embarkati	Disem
2		########	24HR		7	PCF		########	PCF		########	PCF
3	Celebrity (	########	24HR		7	KWE	Key West,	########	PEV	Port Everg	########	PEV
4	Fantasy	########	24HR		4	PCF	Port Cana	########	PCF	Port Cana	########	PCF
5	Celebrity (	########	24HR		7	MIA	Port of Mi	########	MIA	Port of Mi	########	MIA
6	Paradise	########	24HR		7	MIA	Port of Mi	########	MIA	Port of Mi	########	MIA
7	Disney Wo	########	24HR		4	PCF	Port Cana	########	PCF	Port Cana	########	PCF
8	Fascinatio	########	24HR		7	SJU	San Juan,	########	SJU	San Juan,	########	SJU
9	Holiday	########	4HR		4	LAX	Los Angele	########	LAX	Los Angele	########	LAX
10	Jubilee - I	########	24HR		5	TAM	Tampa, FL	########	TAM	Tampa, FL	########	TAM
11	Sensation	########	24HR		7	TAM	Tampa, FL	########	TAM	Tampa, FL	########	TAM
12	Fascinatio	########	4HR		7	SJU	San Juan,	########	SJU	San Juan,	########	SJU
13	Disney Ma	########	4HR		7	PCF	Port Cana	########	PCF	Port Cana	########	PCF
14	Fantasy	########	24HR		3	PCF	Port Cana	########	PCF	Port Cana	########	PCF
15	Paradise	########	24HR		7	MIA	Port of Mi	########	MIA	Port of Mi	########	MIA
16	Carnival D	########	24HR		7	SJU	San Juan,	########	SJU	San Juan,	########	SJU
17	Carnival V	########	24HR		8	MIA	Port of Mi	########	MIA	Port of Mi	########	MIA
18	Elation	########	24HR		7	LAX	Los Angele	########	LAX	Los Angele	########	LAX
19	Ecstasy	########	24HR		3	MIA	Port of Mi	########	MIA	Port of Mi	########	MIA
20	Elation	########	4HR		7	LAX	Los Angele	########	LAX	Los Angele	########	LAX
21	Celebrity (	########	24HR		7	SJU	San Juan,	########	PEV	Port Everg	########	PEV

Illness by Ship | Crew Cases Only | Sheet5 | Sheet3 | CDC CRUISE SHIP DATA 2000-2010


If Statements

IF STATEMENTS



	Data	INT	IF
21			
22	Data	INT	IF
23	TRUE	1	TRUE
24	TRUE	#VALUE!	=IF(OR(A24,A24="TRUE"),TRUE,FALSE)
25	FALSE	0	FALSE
26	FALSE	#VALUE!	FALSE
27			

Here Is Your Formula:

=IF(CELL Meets CRITERIA, Then THIS, Otherwise THAT)

IF STATEMENTS

1

Open your sheet – "**Big Ten Daily COVID 19**"

These are COVID 19 cases at Big Ten Colleges and Universities

2

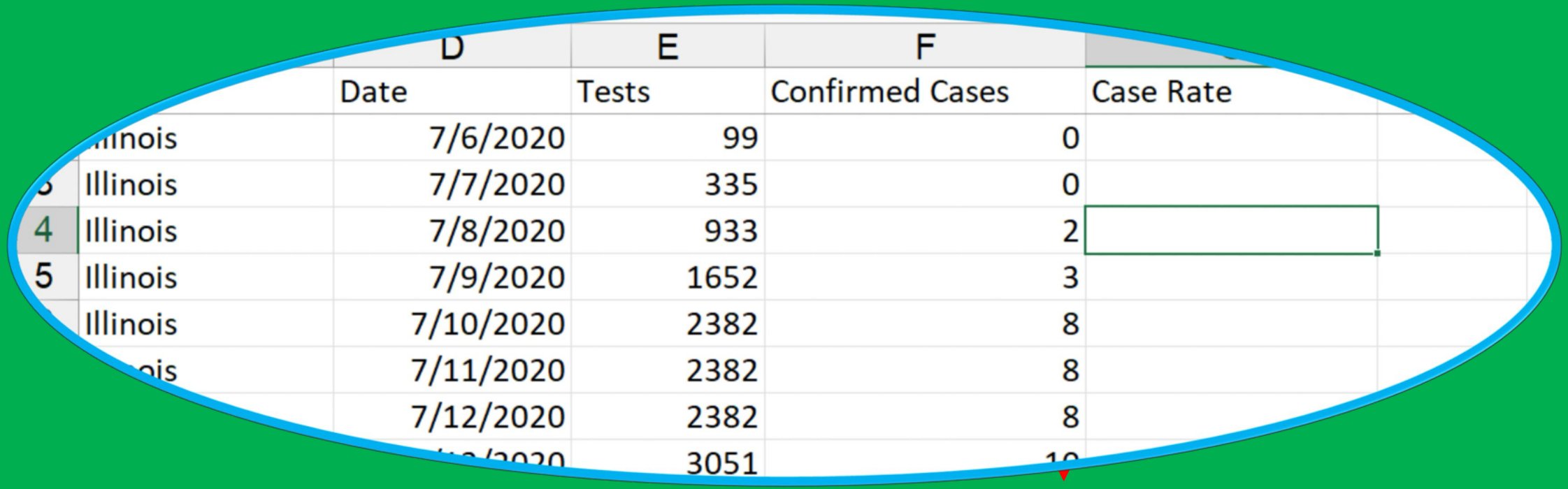
Make sure you are on the first sheet

3

Give column G the name **Case Rate**

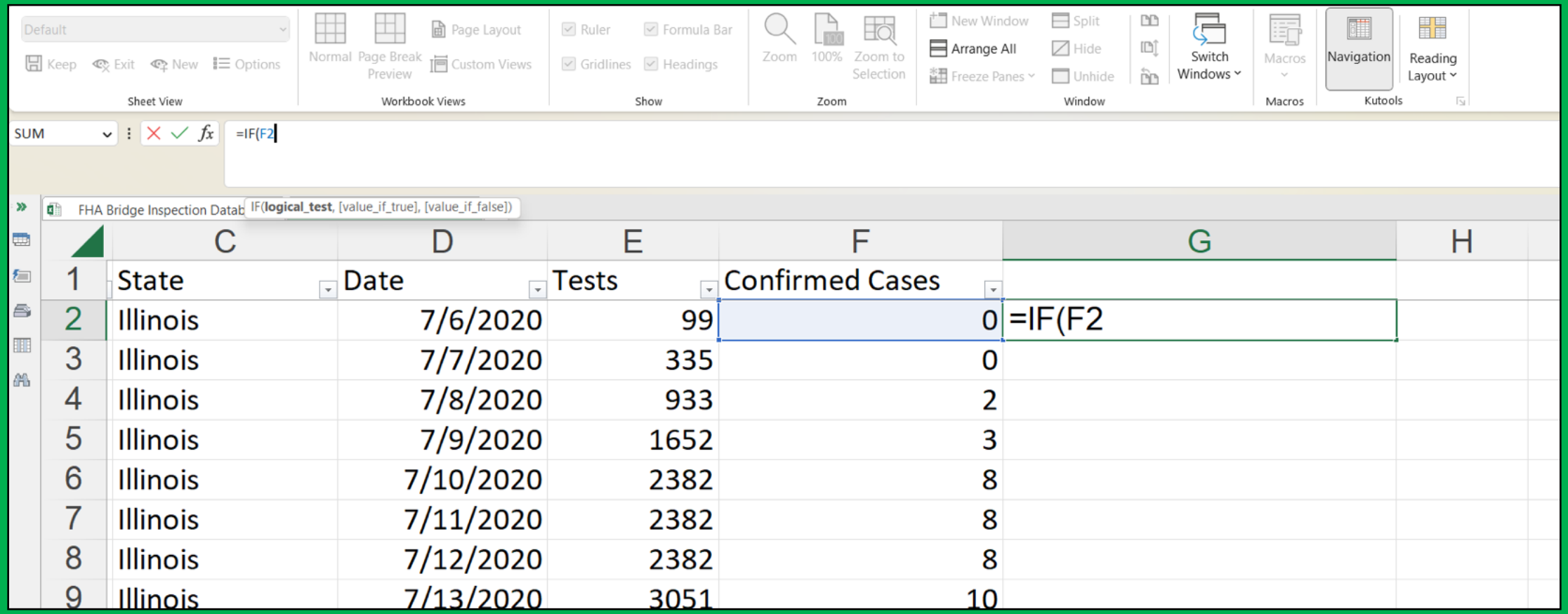
Column G

Call Your New Column
Case Rate



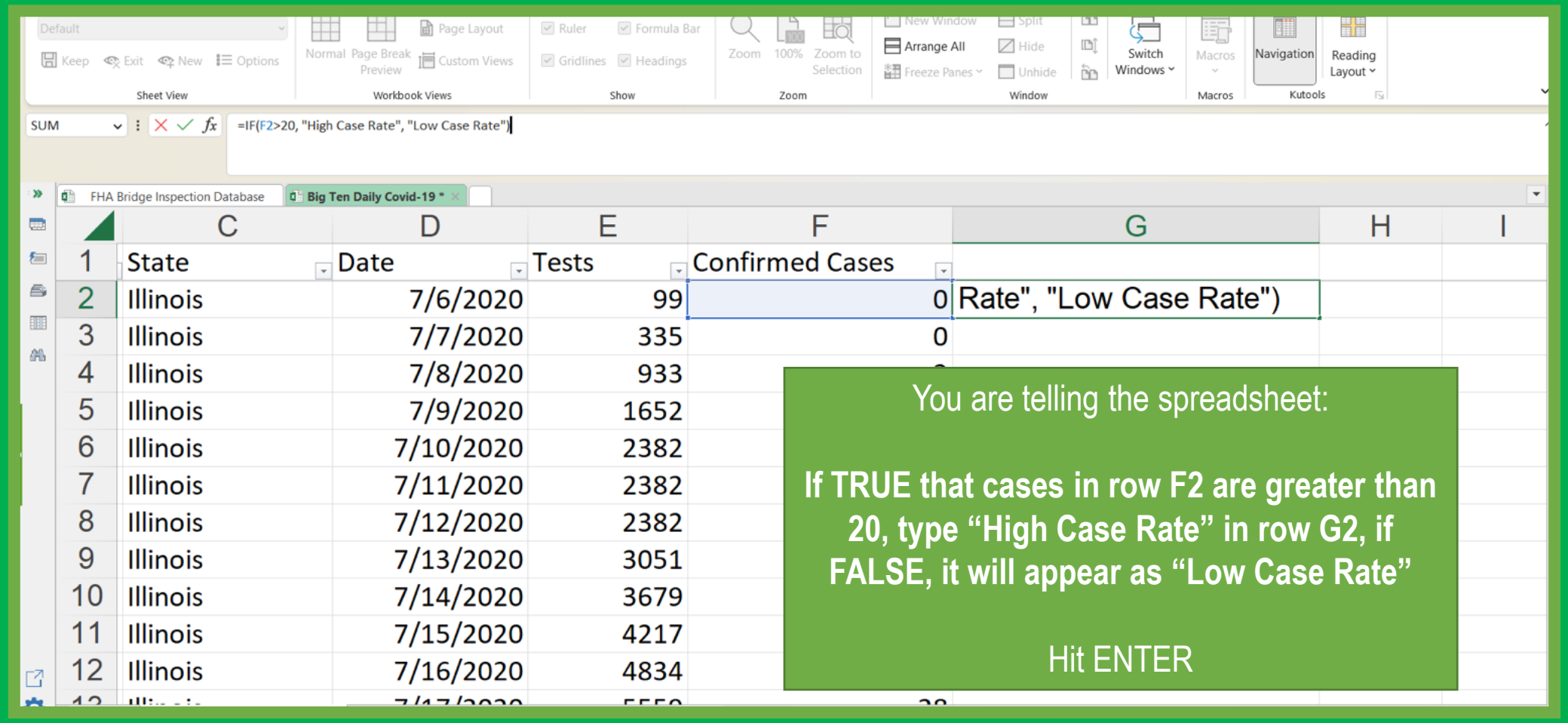
	D	E	F	
	Date	Tests	Confirmed Cases	Case Rate
	inois	7/6/2020	99	0
	Illinois	7/7/2020	335	0
4	Illinois	7/8/2020	933	2
5	Illinois	7/9/2020	1652	3
	Illinois	7/10/2020	2382	8
	ois	7/11/2020	2382	8
		7/12/2020	2382	8
		2020	3051	10

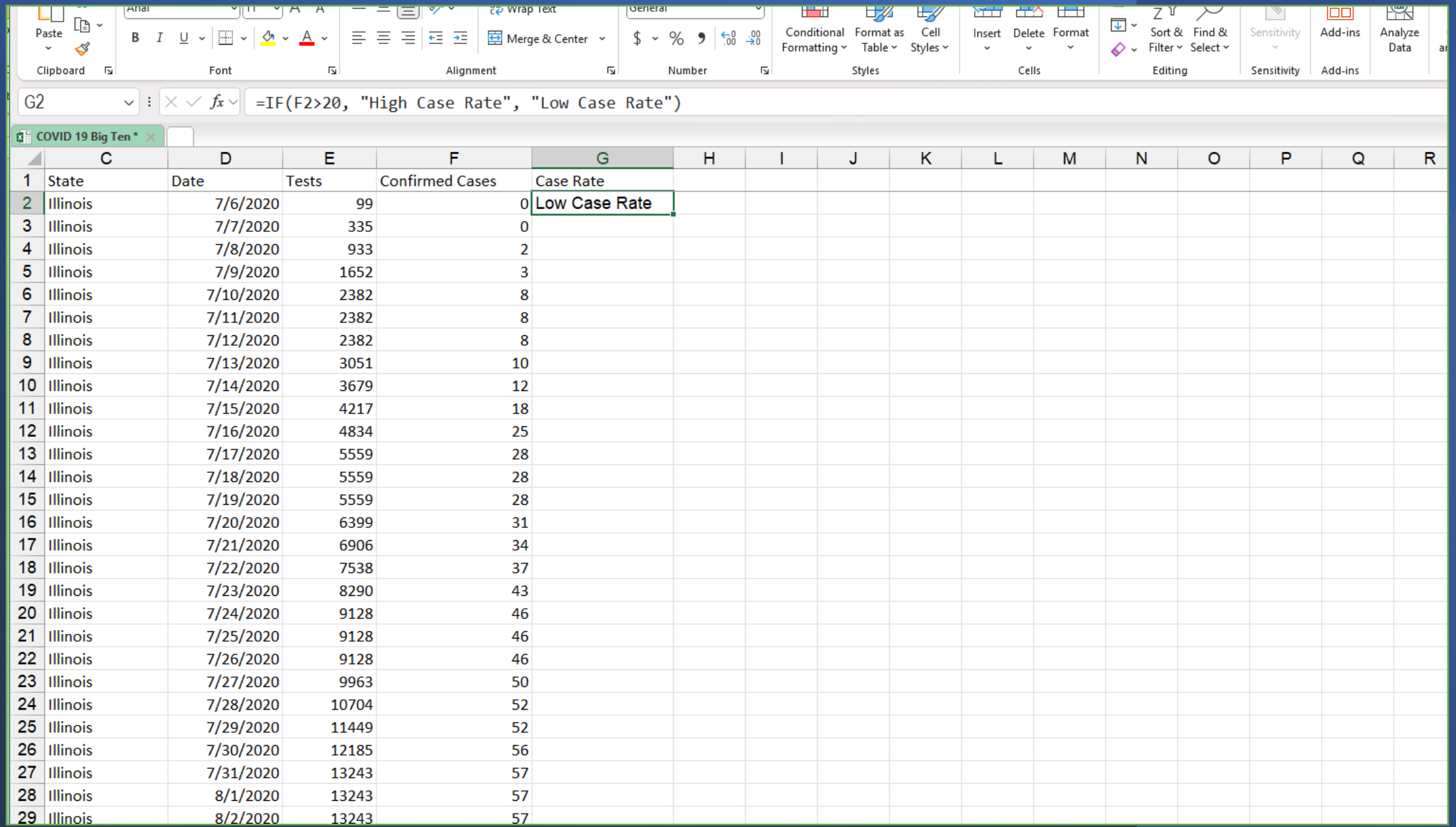
In Cell G2, We'll Start to Type Our Formula in the Formula Bar



	C	D	E	F	G	H
1	State	Date	Tests	Confirmed Cases		
2	Illinois	7/6/2020	99	0	=IF(F2	
3	Illinois	7/7/2020	335	0		
4	Illinois	7/8/2020	933	2		
5	Illinois	7/9/2020	1652	3		
6	Illinois	7/10/2020	2382	8		
7	Illinois	7/11/2020	2382	8		
8	Illinois	7/12/2020	2382	8		
9	Illinois	7/13/2020	3051	10		

In Cell G2, Type =IF(F2>20, “High Case Rate”, “Low Case Rate”)





	C	D	E	F	G
1	State	Date	Tests	Confirmed Cases	Case Rate
2	Illinois	7/6/2020	99	0	Low Case Rate
3	Illinois	7/7/2020	335	0	
4	Illinois	7/8/2020	933	2	
5	Illinois	7/9/2020	1652	3	
6	Illinois	7/10/2020	2382	8	
7	Illinois	7/11/2020	2382	8	
8	Illinois	7/12/2020	2382	8	
9	Illinois	7/13/2020	3051	10	
10	Illinois	7/14/2020	3679	12	
11	Illinois	7/15/2020	4217	18	
12	Illinois	7/16/2020	4834	25	
13	Illinois	7/17/2020	5559	28	
14	Illinois	7/18/2020	5559	28	
15	Illinois	7/19/2020	5559	28	
16	Illinois	7/20/2020	6399	31	
17	Illinois	7/21/2020	6906	34	
18	Illinois	7/22/2020	7538	37	
19	Illinois	7/23/2020	8290	43	
20	Illinois	7/24/2020	9128	46	
21	Illinois	7/25/2020	9128	46	
22	Illinois	7/26/2020	9128	46	
23	Illinois	7/27/2020	9963	50	
24	Illinois	7/28/2020	10704	52	
25	Illinois	7/29/2020	11449	52	
26	Illinois	7/30/2020	12185	56	
27	Illinois	7/31/2020	13243	57	
28	Illinois	8/1/2020	13243	57	
29	Illinois	8/2/2020	13243	57	

IF Statements – Copy Your Formula Down

Put Your Cursor On The Extreme Right-Hand Corner Of The First Cell Until It Turns Into a

Thin Black Cross

Then Double click and It Will Copy All The Way Down

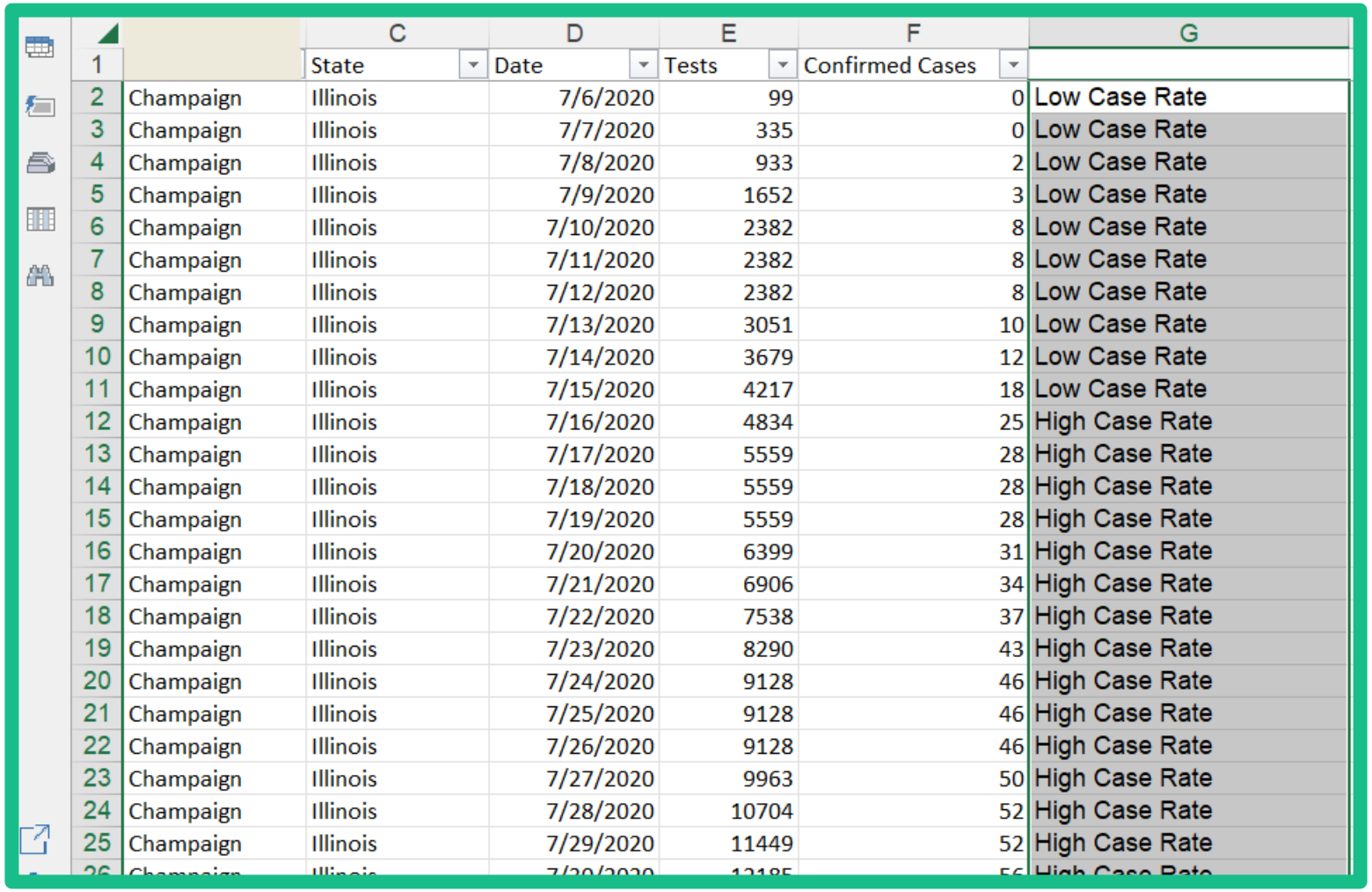


Click **ENTER** And Copy Your Formula Down

- This is also called a **FLASH FILL**
- The formula you typed will now apply to every cell below

4	Illinois	7/8/2020	933	2	Low Case Rate
5	Illinois	7/9/2020	1652	3	Low Case Rate
6	Illinois	7/10/2020	2382	8	Low Case Rate
7	Illinois	7/11/2020	2382	8	Low Case Rate
8	Illinois	7/12/2020	2382	8	Low Case Rate
9	Illinois	7/13/2020	3051	10	Low Case Rate
10	Illinois	7/14/2020	3679	12	Low Case Rate
11	Illinois	7/15/2020	4217	18	Low Case Rate
12	Illinois	7/16/2020	4834	25	High Case Rate
13	Illinois	7/17/2020	5559	28	High Case Rate
14	Illinois	7/18/2020	5559	28	High Case Rate
15	Illinois	7/19/2020	5559	28	High Case Rate
16	Illinois	7/20/2020	6399	31	High Case Rate
17	Illinois	7/21/2020	6906	34	High Case Rate
18	Illinois	7/22/2020	7538	37	High Case Rate
19	Illinois	7/23/2020	8290	43	High Case Rate

IF STATEMENTS



		C	D	E	F	G
1		State	Date	Tests	Confirmed Cases	
2	Champaign	Illinois	7/6/2020	99	0	Low Case Rate
3	Champaign	Illinois	7/7/2020	335	0	Low Case Rate
4	Champaign	Illinois	7/8/2020	933	2	Low Case Rate
5	Champaign	Illinois	7/9/2020	1652	3	Low Case Rate
6	Champaign	Illinois	7/10/2020	2382	8	Low Case Rate
7	Champaign	Illinois	7/11/2020	2382	8	Low Case Rate
8	Champaign	Illinois	7/12/2020	2382	8	Low Case Rate
9	Champaign	Illinois	7/13/2020	3051	10	Low Case Rate
10	Champaign	Illinois	7/14/2020	3679	12	Low Case Rate
11	Champaign	Illinois	7/15/2020	4217	18	Low Case Rate
12	Champaign	Illinois	7/16/2020	4834	25	High Case Rate
13	Champaign	Illinois	7/17/2020	5559	28	High Case Rate
14	Champaign	Illinois	7/18/2020	5559	28	High Case Rate
15	Champaign	Illinois	7/19/2020	5559	28	High Case Rate
16	Champaign	Illinois	7/20/2020	6399	31	High Case Rate
17	Champaign	Illinois	7/21/2020	6906	34	High Case Rate
18	Champaign	Illinois	7/22/2020	7538	37	High Case Rate
19	Champaign	Illinois	7/23/2020	8290	43	High Case Rate
20	Champaign	Illinois	7/24/2020	9128	46	High Case Rate
21	Champaign	Illinois	7/25/2020	9128	46	High Case Rate
22	Champaign	Illinois	7/26/2020	9128	46	High Case Rate
23	Champaign	Illinois	7/27/2020	9963	50	High Case Rate
24	Champaign	Illinois	7/28/2020	10704	52	High Case Rate
25	Champaign	Illinois	7/29/2020	11449	52	High Case Rate

Conditional Formatting



Conditional Formatting to Spot Trends in Data

Open Visualizing Your Data Revised Fall 2024 and Find the Sheet Called - NBA_Players_Team_Position

All 30 NBA teams are listed on the Y axis – the vertical axis

The player positions are on the X axis – the horizontal axis

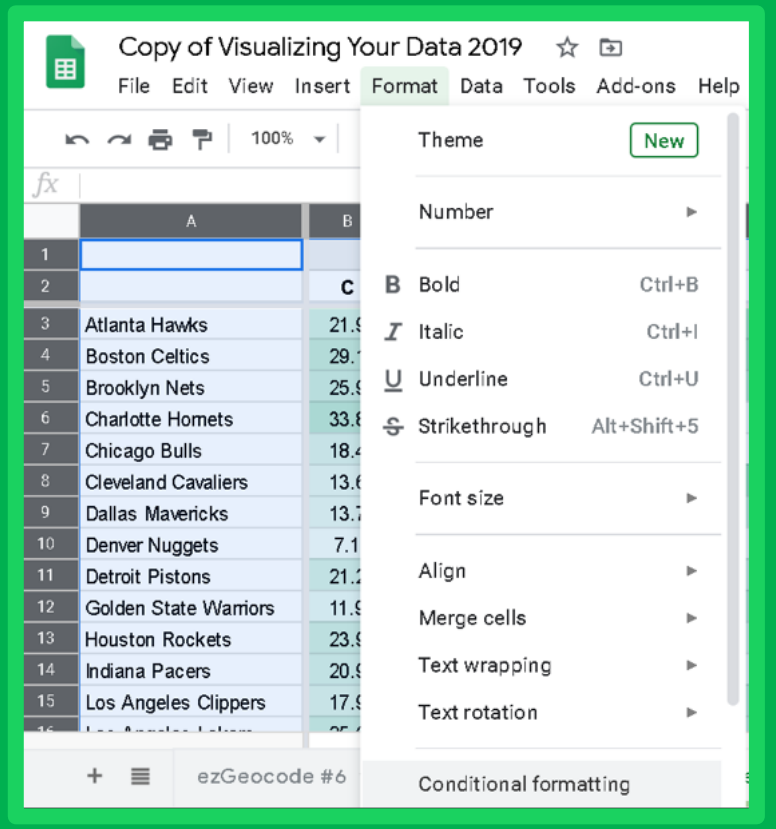
This chart shows **total points averaged by position for each team**

We are going to use a function in G-Sheets called **Conditional Formatting**

It allows you to visualize data to quickly spot trends and patterns

Conditional Formatting to Spot Trends in Data

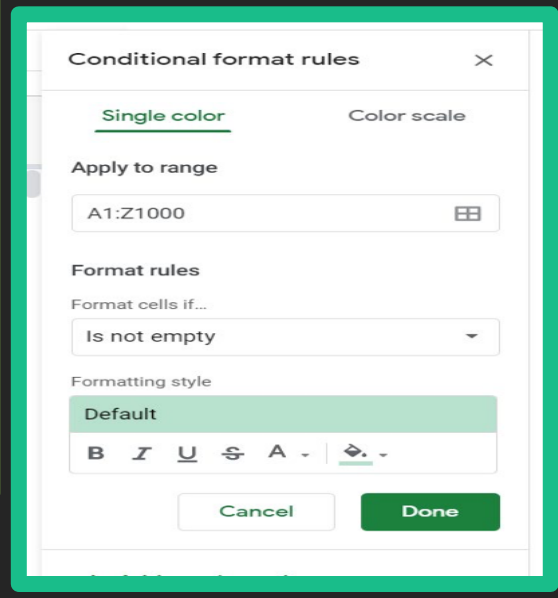
- With this data, what might we want to see?
- Which numbers are higher, and which are lower?
- Let's shade the higher number of points a darker color and the lower numbers a lighter color
- Highlight all the data
- Go to FORMAT
- Go to CONDITIONAL FORMATTING



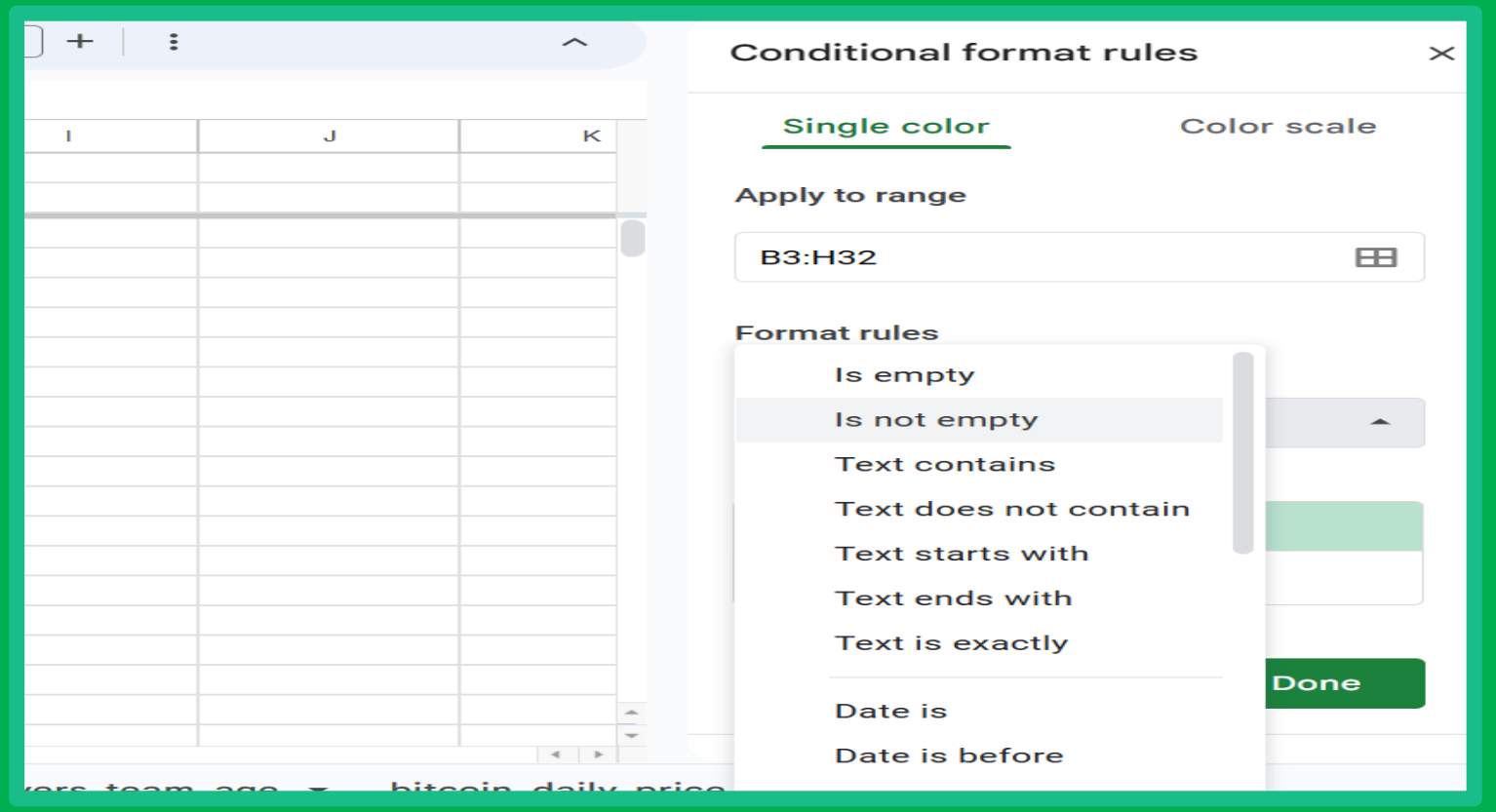
Conditional Formatting to Visualize Data

- We need to make some adjustments to our range
- Can you figure out what range we need to visualize?
- Remember your cell addresses
- We need G-sheets to visualize B3:H32
- Change the range
- We can leave the other settings as is

	A	B	C	D	E	F	G	H
1		POSITION						
2		C	F	G	PF	PG	SF	SG
3	Atlanta Hawks	21.9		4.7	26.1	34.9	44.9	32.6
4	Boston Celtics	29.1		0	10.4	58.7	47	4.1
5	Brooklyn Nets	25.9			16	55.8	37.3	46.1
6	Charlotte Hornets	33.8			23.4	29.9	9.2	38.8
7	Chicago Bulls	18.4			45.2	51.3	4	64.4
8	Cleveland Cavaliers	13.6			30.2	38.9	50.8	58.8
9	Dallas Mavericks	13.7	5.4		36.7	48.7	33.7	24.1
10	Denver Nuggets	7.1			55	20	11.5	57.3
11	Detroit Pistons	21.2			59	43.8	27.5	23.6
12	Golden State Warriors	11.9			26.4	41.4	38.1	31.3
13	Houston Rockets	23.9	8		26	54.2	25.1	36.5
14	Indiana Pacers	20.9			31.7	18.3	15.3	46
15	Los Angeles Clippers	17.9	1.8		57.8	14.3	31.9	83
16	Los Angeles Lakers	25.6		12.6	40.8	33.4	28.7	39.4
17	Memphis Grizzlies	23	8.5	9.3	28.6	39.4	52.1	42.9
18	Miami Heat	25.5			25.6	19.1	14	68.9
19	Milwaukee Bucks	23.3			61.4	30.3	28	28.8
20	Minnesota Timberwolves	23.9			24.9	27.4	46.7	11.7
21	New Orleans Pelicans	30.9			48.3	38.9	16.5	33.3
22	New York Knicks	20.1			45	38.7	27.9	36
23	Oklahoma City Thunder	15.7			6	33.3	65.8	9.8
24	Orlando Magic	34.1			4.2	30.1	41.3	50.9
25	Philadelphia 76ers	27.9			33.1	52	24.6	35.8
26	Phoenix Suns	26.3	6.4		25.5	52.3	34.2	52.5
27	Portland Trail Blazers	20.7			14.6	41	16.8	35
28	Sacramento Kings	21.6	5.7		23.2	29.8	17.6	42.6
29	San Antonio Spurs	14.2			29	29	28.7	35.7
30	Toronto Raptors	22.1	1.5	2.3	19.9	32.8	21.4	30.5
31	Utah Jazz	14.4		3.3	21.7	40.9	32	45
32	Washington Wizards	16.6			22.7	28.3	28.5	36.1

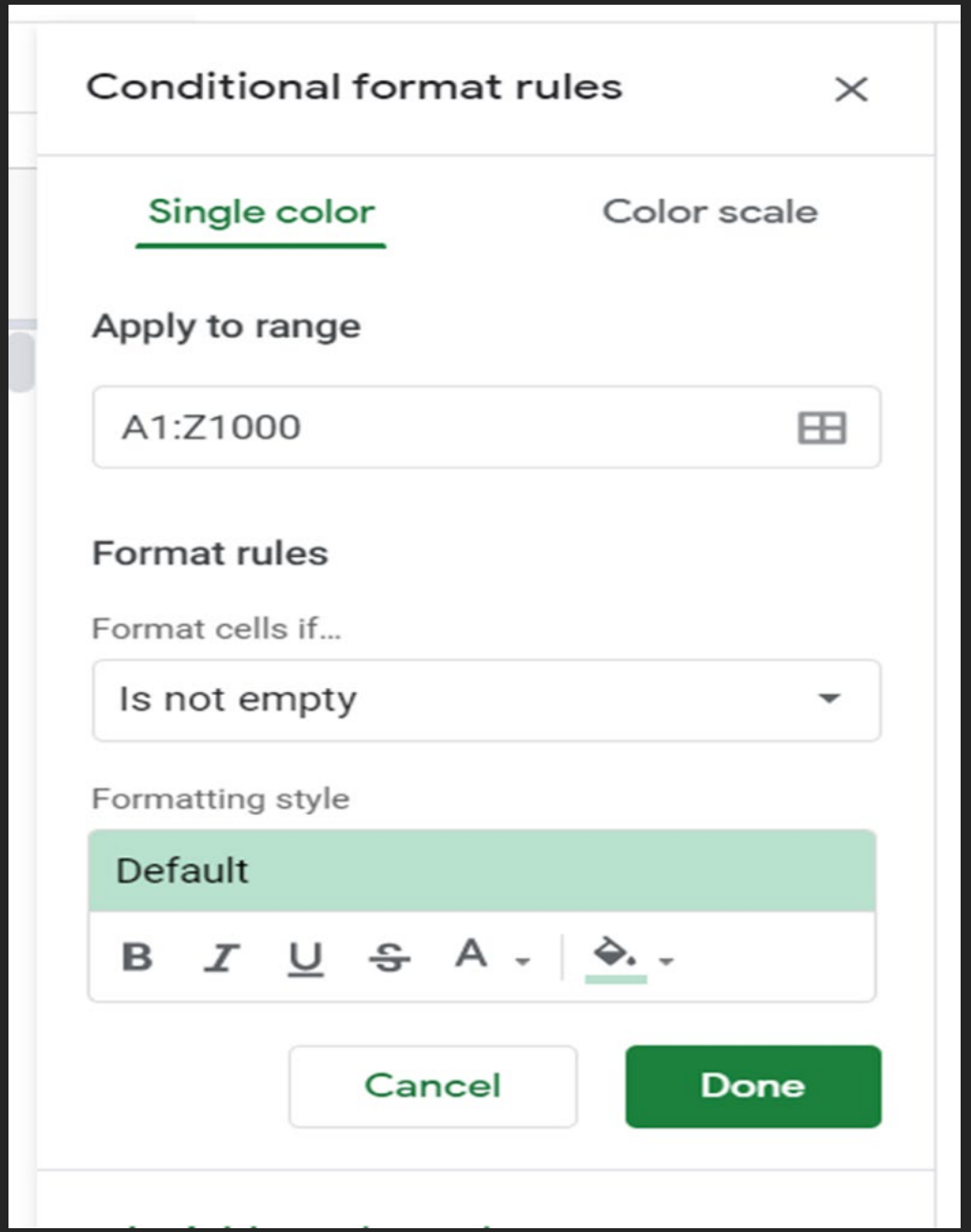


You Can Set Up Additional Formatting Rules



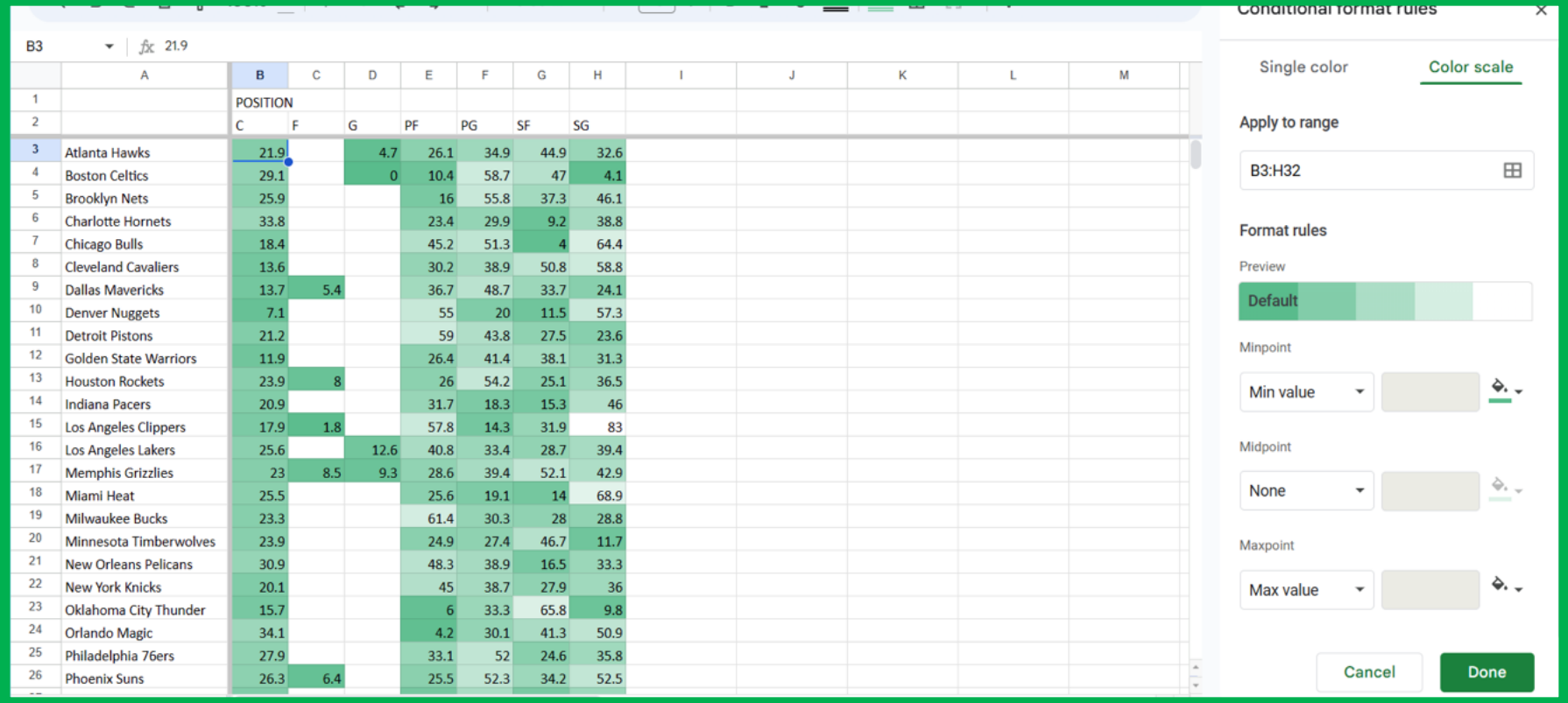
Conditional Formatting

- Now we can change the color scale to anything we want
- If you leave it at single color, or the default color range, what do you get?





Change Single Color To Color Scale To More Easily See Trends



		POSITION						
		C	F	G	PF	PG	SF	SG
3	Atlanta Hawks	21.9		4.7	26.1	34.9	44.9	32.6
4	Boston Celtics	29.1		0	10.4	58.7	47	4.1
5	Brooklyn Nets	25.9			16	55.8	37.3	46.1
6	Charlotte Hornets	33.8			23.4	29.9	9.2	38.8
7	Chicago Bulls	18.4			45.2	51.3	4	64.4
8	Cleveland Cavaliers	13.6			30.2	38.9	50.8	58.8
9	Dallas Mavericks	13.7	5.4		36.7	48.7	33.7	24.1
10	Denver Nuggets	7.1			55	20	11.5	57.3
11	Detroit Pistons	21.2			59	43.8	27.5	23.6
12	Golden State Warriors	11.9			26.4	41.4	38.1	31.3
13	Houston Rockets	23.9	8		26	54.2	25.1	36.5
14	Indiana Pacers	20.9			31.7	18.3	15.3	46
15	Los Angeles Clippers	17.9	1.8		57.8	14.3	31.9	83
16	Los Angeles Lakers	25.6		12.6	40.8	33.4	28.7	39.4
17	Memphis Grizzlies	23	8.5	9.3	28.6	39.4	52.1	42.9
18	Miami Heat	25.5			25.6	19.1	14	68.9
19	Milwaukee Bucks	23.3			61.4	30.3	28	28.8
20	Minnesota Timberwolves	23.9			24.9	27.4	46.7	11.7
21	New Orleans Pelicans	30.9			48.3	38.9	16.5	33.3
22	New York Knicks	20.1			45	38.7	27.9	36
23	Oklahoma City Thunder	15.7			6	33.3	65.8	9.8
24	Orlando Magic	34.1			4.2	30.1	41.3	50.9
25	Philadelphia 76ers	27.9			33.1	52	24.6	35.8
26	Phoenix Suns	26.3	6.4		25.5	52.3	34.2	52.5



ata Fall 2023

sert Format Data Tools Extensions Help

00% $ % .0 .00 123 Defaul... 10 B I A

B	C	D	E	F	G	H	I	J
POSITION								
C	F	G	PF	PG	SF	SG		
21.9		4.7	26.1	34.9	44.9	32.6		
29.1		0	10.4	58.7	47	4.1		
25.9			16	55.8	37.3	46.1		
33.8			23.4	29.9	9.2	38.8		
18.4			45.2	51.3	4	64.4		
13.6			30.2	38.9	50.8	58.8		
13.7	5.4		36.7	48.7	33.7	24.1		
7.1			55	20	11.5	57.3		
21.2			59	43.8	27.5	23.6		
11.9			26.4	41.4	38.1	31.3		
23.9	8		26	54.2	25.1	36.5		
20.9			31.7	18.3	15.3	46		
17.9	1.8		57.8	14.3	31.9	83		
25.6		12.6	40.8	33.4	28.7	39.4		
23	8.5	9.3	28.6	39.4	52.1	42.9		
25.5			25.6	19.1	14	68.9		
23.3			61.4	30.3	28	28.8		
23.9			24.9	27.4	46.7	11.7		
30.9			48.3	38.9	16.5	33.3		
20.1			45	38.7	27.9	36		
15.7			6	33.3	65.8	9.8		
34.1			4.2	30.1	41.3	50.9		
27.9			33.1	52	24.6	35.8		
26.3	6.4		25.5	52.3	34.2	52.5		

Click On The Bar That Says **Default** To Get More Color Options

I LIKE TO CHOOSE THE BRIGHTEST OPTION TO MORE EASILY SPOT TRENDS AND PATTERNS

File Edit View Insert Format Data Tools Extensions Help

B3 21.9

	A	B	C	D	E	F	G	H
1		POSITION						
2		C	F	G	PF	PG	SF	SG
3	Atlanta Hawks	21.9		4.7	26.1	34.9	44.9	32.6
4	Boston Celtics	29.1		0	10.4	58.7	47	4.1
5	Brooklyn Nets	25.9			16	55.8	37.3	46.1
6	Charlotte Hornets	33.8			23.4	29.9	9.2	38.8
7	Chicago Bulls	18.4			45.2	51.3	4	64.4
8	Cleveland Cavaliers	13.6			30.2	38.9	50.8	58.8
9	Dallas Mavericks	13.7	5.4		36.7	48.7	33.7	24.1
10	Denver Nuggets	7.1			55	20	11.5	57.3
11	Detroit Pistons	21.2			59	43.8	27.5	23.6
12	Golden State Warriors	11.9			26.4	41.4	38.1	31.3
13	Houston Rockets	23.9	8		26	54.2	25.1	36.5
14	Indiana Pacers	20.9			31.7	18.3	15.3	46
15	Los Angeles Clippers	17.9	1.8		57.8	14.3	31.9	83
16	Los Angeles Lakers	25.6		12.6	40.8	33.4	28.7	39.4
17	Memphis Grizzlies	23	8.5	9.3	28.6	39.4	52.1	42.9
18	Miami Heat	25.5			25.6	19.1	14	68.9
19	Milwaukee Bucks	23.3			61.4	30.3	28	28.8
20	Minnesota Timberwolves	23.9			24.9	27.4	46.7	11.7
21	New Orleans Pelicans	30.9			48.3	38.9	16.5	33.3
22	New York Knicks	20.1			45	38.7	27.9	36
23	Oklahoma City Thunder	15.7			6	33.3	65.8	9.8
24	Orlando Magic	34.1			4.2	30.1	41.3	50.9
25	Philadelphia 76ers	27.9			33.1	52	24.6	35.8
26	Phoenix Suns	26.3	6.4		25.5	52.3	34.2	52.5

What Insights Do We Have?

Remember What Our Data Is:
Average Points Per NBA Player By Team and Position?

What Position on Which Team Scores The Highest?
What Position on Which Team Scores The Lowest?

Conditional format rules

Single color | Color scale

Apply to range

B3:H32

Format rules

Preview

Default

Minpoint

Min value

Midpoint

Percentile 50

Maxpoint

Max value

Cancel Done

Conditional Formatting

Let's try that again

Open the sheet nba_players_team_age

The teams are on the Y axis – the vertical axis

The ages are on the X axis – the horizontal axis

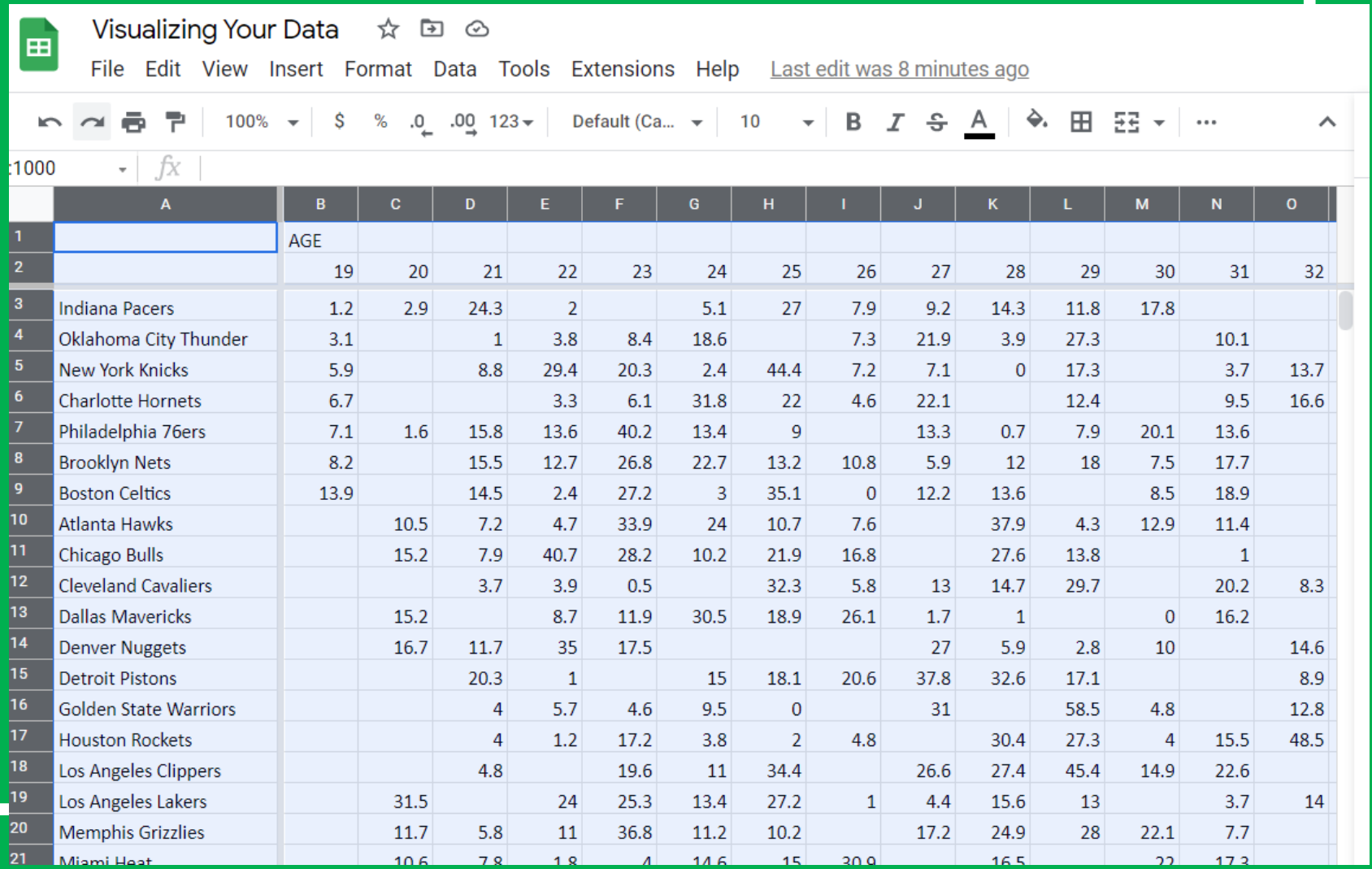
These are the average points scored by NBA players, by team and by their age

We'll repeat our steps

Select your data

Go to FORMAT

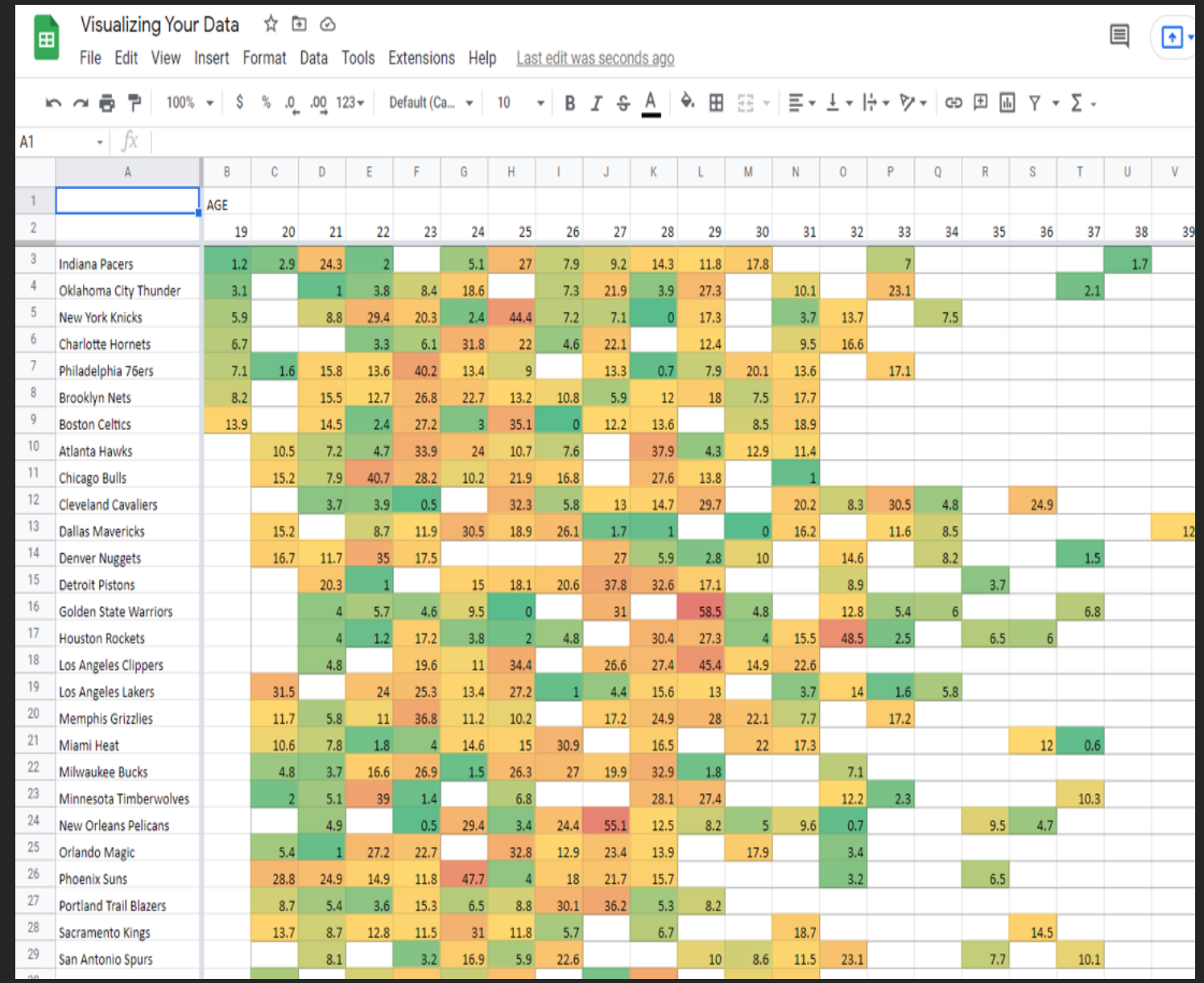
Go to CONDITIONAL FORMATTING



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1		AGE													
2		19	20	21	22	23	24	25	26	27	28	29	30	31	32
3	Indiana Pacers	1.2	2.9	24.3	2		5.1	27	7.9	9.2	14.3	11.8	17.8		
4	Oklahoma City Thunder	3.1		1	3.8	8.4	18.6		7.3	21.9	3.9	27.3		10.1	
5	New York Knicks	5.9		8.8	29.4	20.3	2.4	44.4	7.2	7.1	0	17.3		3.7	13.7
6	Charlotte Hornets	6.7			3.3	6.1	31.8	22	4.6	22.1		12.4		9.5	16.6
7	Philadelphia 76ers	7.1	1.6	15.8	13.6	40.2	13.4	9		13.3	0.7	7.9	20.1	13.6	
8	Brooklyn Nets	8.2		15.5	12.7	26.8	22.7	13.2	10.8	5.9	12	18	7.5	17.7	
9	Boston Celtics	13.9		14.5	2.4	27.2	3	35.1	0	12.2	13.6		8.5	18.9	
10	Atlanta Hawks		10.5	7.2	4.7	33.9	24	10.7	7.6		37.9	4.3	12.9	11.4	
11	Chicago Bulls		15.2	7.9	40.7	28.2	10.2	21.9	16.8		27.6	13.8		1	
12	Cleveland Cavaliers			3.7	3.9	0.5		32.3	5.8	13	14.7	29.7		20.2	8.3
13	Dallas Mavericks		15.2		8.7	11.9	30.5	18.9	26.1	1.7	1		0	16.2	
14	Denver Nuggets		16.7	11.7	35	17.5				27	5.9	2.8	10		14.6
15	Detroit Pistons			20.3	1		15	18.1	20.6	37.8	32.6	17.1			8.9
16	Golden State Warriors			4	5.7	4.6	9.5	0		31		58.5	4.8		12.8
17	Houston Rockets			4	1.2	17.2	3.8	2	4.8		30.4	27.3	4	15.5	48.5
18	Los Angeles Clippers			4.8		19.6	11	34.4		26.6	27.4	45.4	14.9	22.6	
19	Los Angeles Lakers		31.5		24	25.3	13.4	27.2	1	4.4	15.6	13		3.7	14
20	Memphis Grizzlies		11.7	5.8	11	36.8	11.2	10.2		17.2	24.9	28	22.1	7.7	
21	Miami Heat		10.6	7.8	1.8	4	14.6	15	30.9		16.5		22	17.3	

What Insights Do You See From Your Visualization?

- Very few players score points after their early 30's
- The Cleveland Cavaliers had a player at 36 who scored almost 25 points
- What is the peak age for scoring points?
- Any age score the most points?
- Any age scored the least points?



	AGE																				
	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39
Indiana Pacers	1.2	2.9	24.3	2		5.1	27	7.9	9.2	14.3	11.8	17.8			7					1.7	
Oklahoma City Thunder	3.1		1	3.8	8.4	18.6		7.3	21.9	3.9	27.3		10.1		23.1				2.1		
New York Knicks	5.9		8.8	29.4	20.3	2.4	44.4	7.2	7.1	0	17.3		3.7	13.7		7.5					
Charlotte Hornets	6.7			3.3	6.1	31.8	22	4.6	22.1		12.4		9.5	16.6							
Philadelphia 76ers	7.1	1.6	15.8	13.6	40.2	13.4	9		13.3	0.7	7.9	20.1	13.6		17.1						
Brooklyn Nets	8.2		15.5	12.7	26.8	22.7	13.2	10.8	5.9	12	18	7.5	17.7								
Boston Celtics	13.9		14.5	2.4	27.2	3	35.1	0	12.2	13.6		8.5	18.9								
Atlanta Hawks		10.5	7.2	4.7	33.9	24	10.7	7.6		37.9	4.3	12.9	11.4								
Chicago Bulls		15.2	7.9	40.7	28.2	10.2	21.9	16.8		27.6	13.8		1								
Cleveland Cavaliers			3.7	3.9	0.5		32.3	5.8	13	14.7	29.7		20.2	8.3	30.5	4.8		24.9			
Dallas Mavericks		15.2		8.7	11.9	30.5	18.9	26.1	1.7	1		0	16.2		11.6	8.5					12
Denver Nuggets		16.7	11.7	35	17.5				27	5.9	2.8	10		14.6		8.2			1.5		
Detroit Pistons			20.3	1		15	18.1	20.6	37.8	32.6	17.1			8.9			3.7				
Golden State Warriors			4	5.7	4.6	9.5	0		31		58.5	4.8		12.8	5.4	6			6.8		
Houston Rockets			4	1.2	17.2	3.8	2	4.8		30.4	27.3	4	15.5	48.5	2.5		6.5	6			
Los Angeles Clippers			4.8		19.6	11	34.4		26.6	27.4	45.4	14.9	22.6								
Los Angeles Lakers		31.5		24	25.3	13.4	27.2	1	4.4	15.6	13		3.7	14	1.6	5.8					
Memphis Grizzlies		11.7	5.8	11	36.8	11.2	10.2		17.2	24.9	28	22.1	7.7		17.2						
Miami Heat		10.6	7.8	1.8	4	14.6	15	30.9		16.5		22	17.3					12	0.6		
Milwaukee Bucks		4.8	3.7	16.6	26.9	1.5	26.3	27	19.9	32.9	1.8			7.1							
Minnesota Timberwolves		2	5.1	39	1.4		6.8			28.1	27.4			12.2	2.3				10.3		
New Orleans Pelicans			4.9		0.5	29.4	3.4	24.4	55.1	12.5	8.2	5	9.6	0.7			9.5	4.7			
Orlando Magic		5.4	1	27.2	22.7		32.8	12.9	23.4	13.9		17.9		3.4							
Phoenix Suns		28.8	24.9	14.9	11.8	47.7	4	18	21.7	15.7				3.2			6.5				
Portland Trail Blazers		8.7	5.4	3.6	15.3	6.5	8.8	30.1	36.2	5.3	8.2										
Sacramento Kings		13.7	8.7	12.8	11.5	31	11.8	5.7		6.7			18.7					14.5			
San Antonio Spurs			8.1		3.2	16.9	5.9	22.6			10	8.6	11.5	23.1			7.7		10.1		

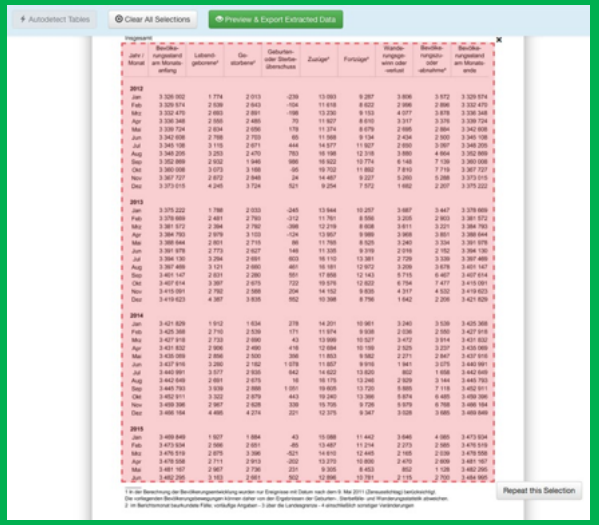


Extracting Data from PDF's: Another Way of "Scraping" Information

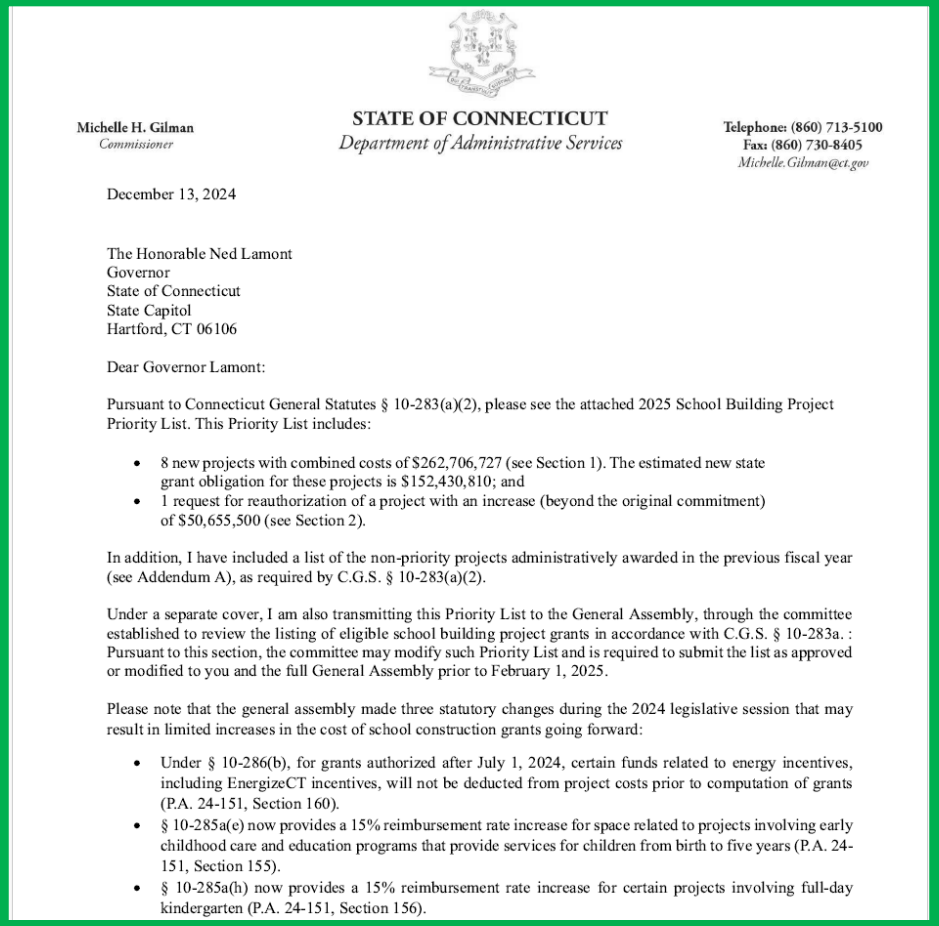
Data Extraction v. Traditional Web Scraping

- We'll be pulling valuable data out of a locked document
- We are "liberating" needed data from PDF's
- Data extraction from PDF's won't automatically update your data when the government agency updates their site
- Traditional web scraping does give you that advantage





Often Journalists Have to "Scrape" or Extract" Data from a PDF for Analysis



STATE OF CONNECTICUT
Department of Administrative Services

Michelle H. Gilman
Commissioner

Telephone: (860) 713-5100
Fax: (860) 730-8405
Michelle.Gilman@ct.gov

December 13, 2024

The Honorable Ned Lamont
Governor
State of Connecticut
State Capitol
Hartford, CT 06106

Dear Governor Lamont:

Pursuant to Connecticut General Statutes § 10-283(a)(2), please see the attached 2025 School Building Project Priority List. This Priority List includes:

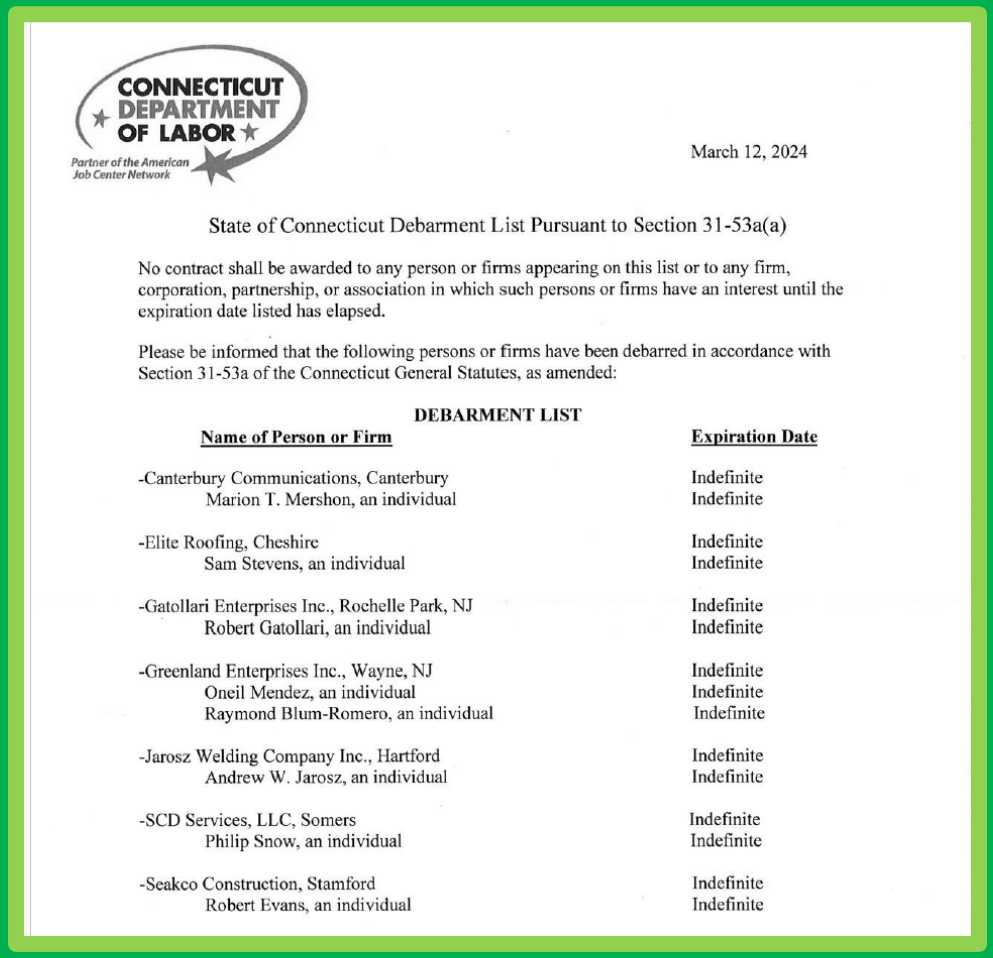
- 8 new projects with combined costs of $262,706,727 (see Section 1). The estimated new state grant obligation for these projects is $152,430,810; and
- 1 request for reauthorization of a project with an increase (beyond the original commitment) of $50,655,500 (see Section 2).

In addition, I have included a list of the non-priority projects administratively awarded in the previous fiscal year (see Addendum A), as required by C.G.S. § 10-283(a)(2).

Under a separate cover, I am also transmitting this Priority List to the General Assembly, through the committee established to review the listing of eligible school building project grants in accordance with C.G.S. § 10-283a. : Pursuant to this section, the committee may modify such Priority List and is required to submit the list as approved or modified to you and the full General Assembly prior to February 1, 2025.

Please note that the general assembly made three statutory changes during the 2024 legislative session that may result in limited increases in the cost of school construction grants going forward:

- Under § 10-286(b), for grants authorized after July 1, 2024, certain funds related to energy incentives, including EnergizeCT incentives, will not be deducted from project costs prior to computation of grants (P.A. 24-151, Section 160).
- § 10-285a(e) now provides a 15% reimbursement rate increase for space related to projects involving early childhood care and education programs that provide services for children from birth to five years (P.A. 24-151, Section 155).
- § 10-285a(h) now provides a 15% reimbursement rate increase for certain projects involving full-day kindergarten (P.A. 24-151, Section 156).



CONNECTICUT DEPARTMENT OF LABOR
Partner of the American Job Center Network

March 12, 2024

State of Connecticut Debarment List Pursuant to Section 31-53a(a)

No contract shall be awarded to any person or firms appearing on this list or to any firm, corporation, partnership, or association in which such persons or firms have an interest until the expiration date listed has elapsed.

Please be informed that the following persons or firms have been debarred in accordance with Section 31-53a of the Connecticut General Statutes, as amended:

DEBARMENT LIST

Name of Person or Firm	Expiration Date
-Canterbury Communications, Canterbury	Indefinite
Marion T. Mershon, an individual	Indefinite
-Elite Roofing, Cheshire	Indefinite
Sam Stevens, an individual	Indefinite
-Gatollari Enterprises Inc., Rochelle Park, NJ	Indefinite
Robert Gatollari, an individual	Indefinite
-Greenland Enterprises Inc., Wayne, NJ	Indefinite
Oneil Mendez, an individual	Indefinite
Raymond Blum-Romero, an individual	Indefinite
-Jarosz Welding Company Inc., Hartford	Indefinite
Andrew W. Jarosz, an individual	Indefinite
-SCD Services, LLC, Somers	Indefinite
Philip Snow, an individual	Indefinite
-Seakco Construction, Stamford	Indefinite
Robert Evans, an individual	Indefinite

Liberating Data from Documents

- Government “traps” data in PDF’s so we can’t access it easily to analyze
- We have to “free” the data so we can work with it. We use different tools to do that
- One way to do that is with PDF converters like a tool such as Cometdocs or Tabula or PDFtoEXCEL

cometdocs



PDFtoExcel.com

Liberating Data from Documents

- Know what you are working with:

  - Two types of PDF files
    - Generated by an application, like Microsoft word
    - Other is a scanned document or an image file

- Tell the difference by trying to highlight some text in a PDF. If you can drag and click the text, you have an app-generated PDF

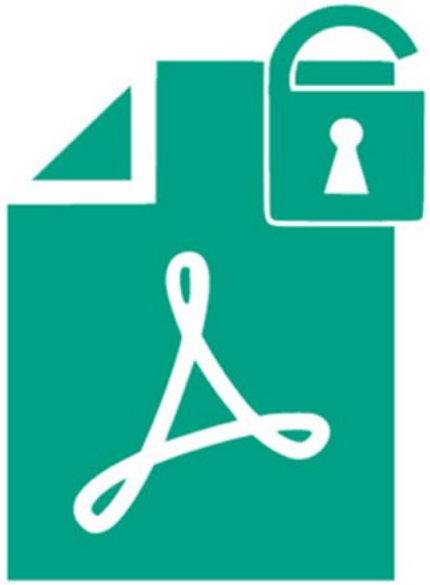
Why Government Traps Data?



*"**The crime stats are subject to being corruptible in an excel sheet.** They have been changed in the past by persons unknown and this affects the veracity of the original data posted. **If stats are posted on-line in a PDF format, this reduces the risk of contamination.** [...] Effective immediately the stats should just be posted in a PDF format."*

-Minneapolis Police Department

http://www.minnpost.com/data/2013/09/update-minneapolis-police-department-restores-accessible-data-format

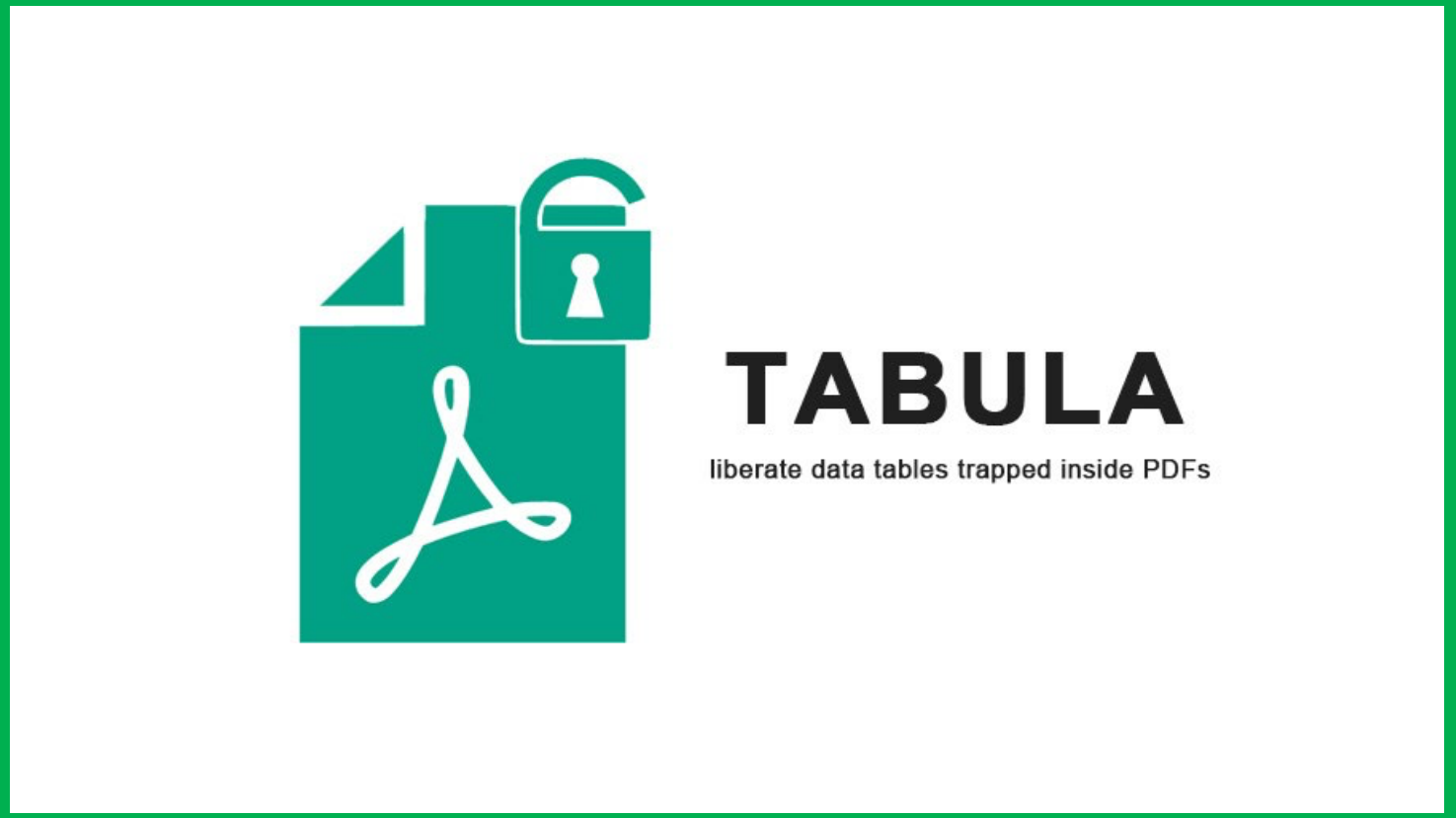


TABULA

liberate data tables trapped inside PDFs

Extracting Data from PDF's with Tabula

A Free Tool Like Tabula Comes In Handy For Pdf's With Multiple Tables



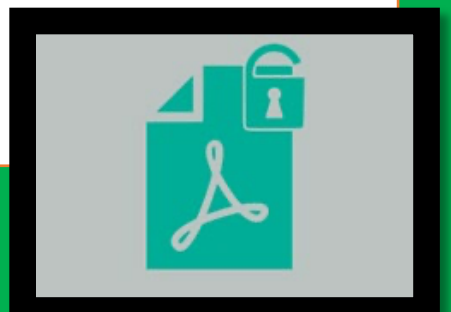
What Is TABULA?



- Free extraction tool to turn tabular data locked in a PDF into exportable CSV or spreadsheet files
- Works on Windows, MAC or Linux systems
- Works on text-based PDFs, where you can select text
- Not scanned documents

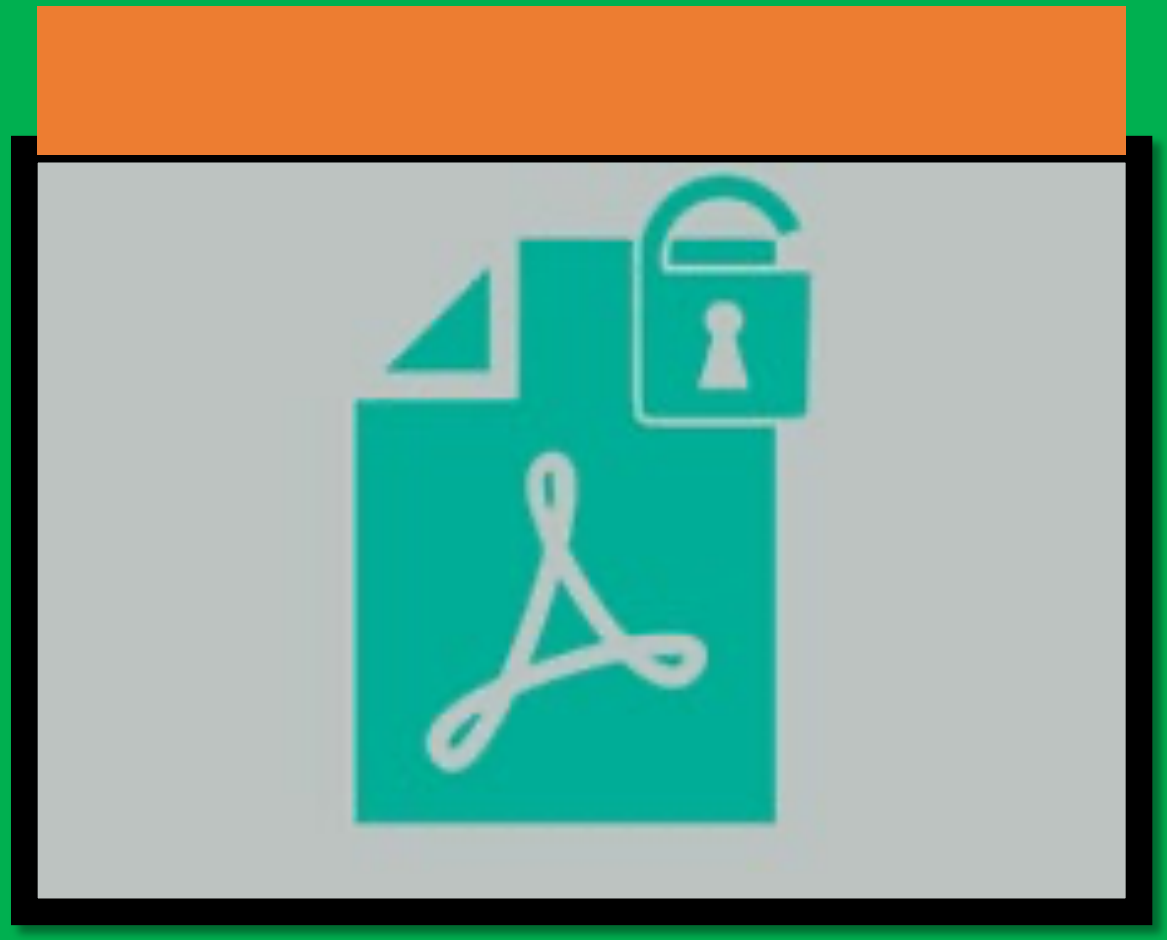
Differences: Text Based and Scanned

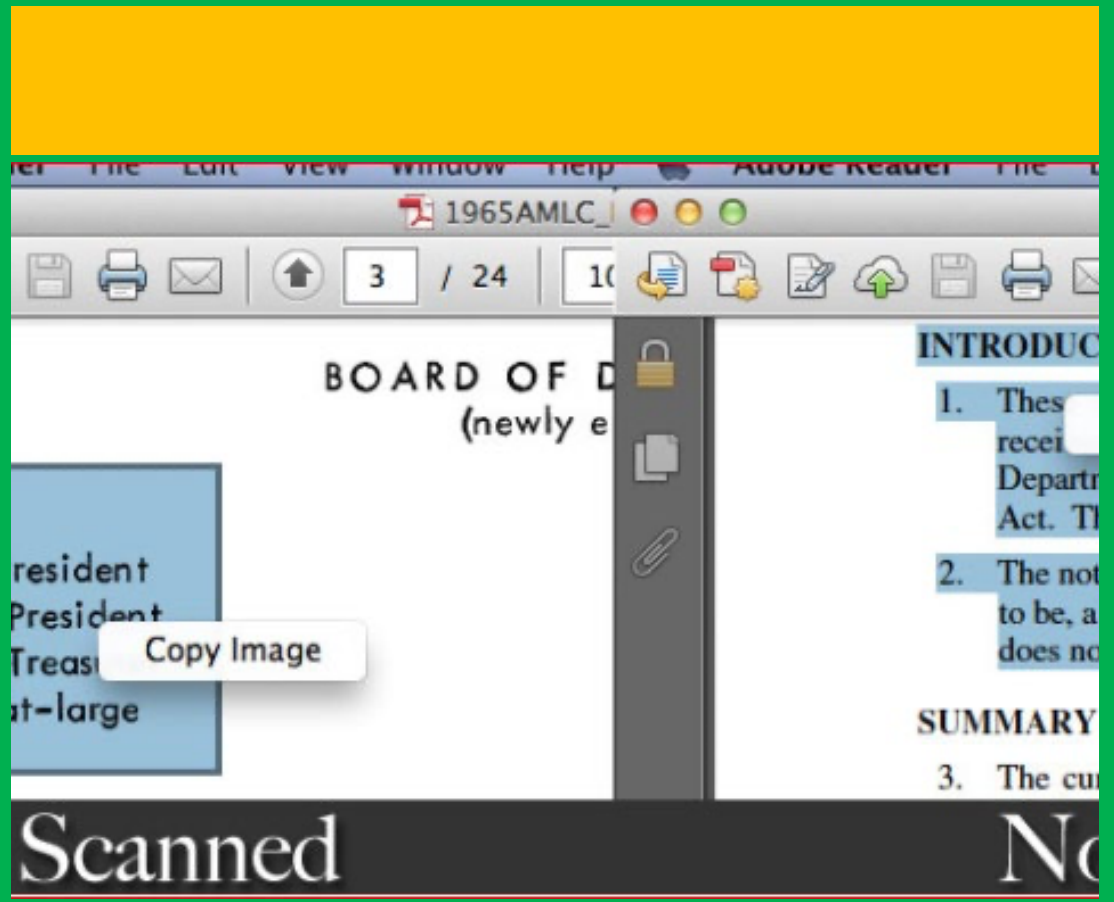
- Text-based files
- PDF stores exact positions of every character
- TABULA can take those positions and extract the characters



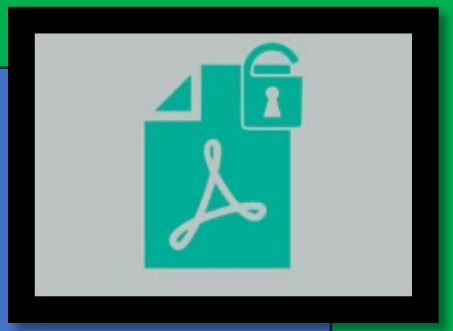
- Scanned PDFs are a collection of images
- Can’t extract the characters
- Could use OCR but accuracy can be problem

Scanned Image vs. Text





Install TABULA



- **Download & Install Tabula**
- Windows & Linux users will need a copy of Java installed. You can download Java here. (Java is included in the Mac version.)
- Download the version of Tabula for your operating system:
  - **Windows:** tabula-win.zip
  - **Mac OS X:** tabula-mac.zip
  - **Linux/Other:** tabula-jar.zip, view README.txt inside for instructions

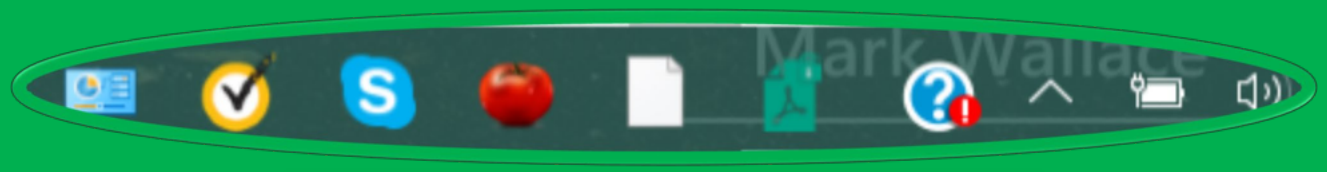
- Extract the zip file. (Instructions: Windows, Mac)
- Go into the folder you just extracted. Run the "Tabula" program inside.
- A web browser will open. If it doesn't, open your web browser, and go to http://localhost:8080. There's Tabula!

Launching TABULA



```
mysql> use elephanti
Database changed
mysql> SOURCE C:\xampp\htdocs\elephanti2\db\mas_gis_city_ip.sql;
ERROR:
Unknown command '\x'.

For information about MySQL products and services, visit:
   http://www.mysql.com/
For developer information, including the MySQL Reference Manual, visit:
   http://dev.mysql.com/
To buy MySQL Enterprise support, training, or other products, visit:
   https://shop.mysql.com/

List of all MySQL commands:
Note that all text commands must be first on line and end with ';'
?         (\?) Synonym for `help'.
clear     (\c) Clear the current input statement.
connect   (\r) Reconnect to the server. Optional arguments are db and host.
delimiter (\d) Set statement delimiter.
ego       (\G) Send command to mysql server, display result vertically.
exit      (\q) Exit mysql. Same as quit.
go        (\g) Send command to mysql server.
help      (\h) Display this help.
notee     (\t) Don't write into outfile.
print     (\p) Print current command.
prompt    (\R) Change your mysql prompt.
quit      (\q) Quit mysql.
rehash    (\#) Rebuild completion hash.
source    (\.) Execute an SQL script file. Takes a file name as an argument.
status    (\s) Get status information from the server.
tee       (\T) Set outfile [to_outfile]. Append everything into given outfile.
use       (\u) Use another database. Takes database name as argument.
charset   (\C) Switch to another charset. Might be needed for processing binlog with multi-byte charsets.
warnings  (\W) Show warnings after every statement.
nowarning (\w) Don't show warnings after every statement.

For server side help, type 'help contents'

    ->
```

- Once you launch TABULA it will first run on the command
- Pin it to your taskbar for easy launching



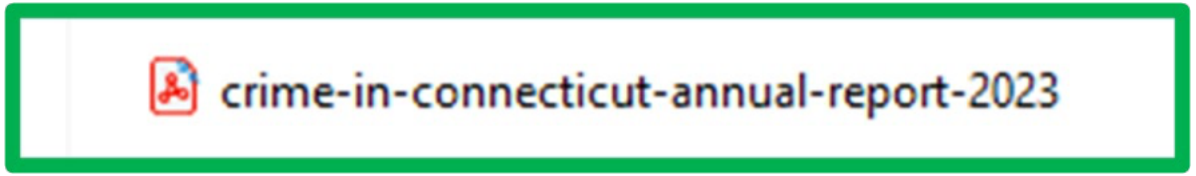
What Is TABULA?



- Free extraction tool to turn tabular data locked in PDFS into exportable CSV or spreadsheet files
- Works on Windows, MAC or Linux systems
- Works on text-based PDF's, where you can select text
- Not scanned documents

Importing Data



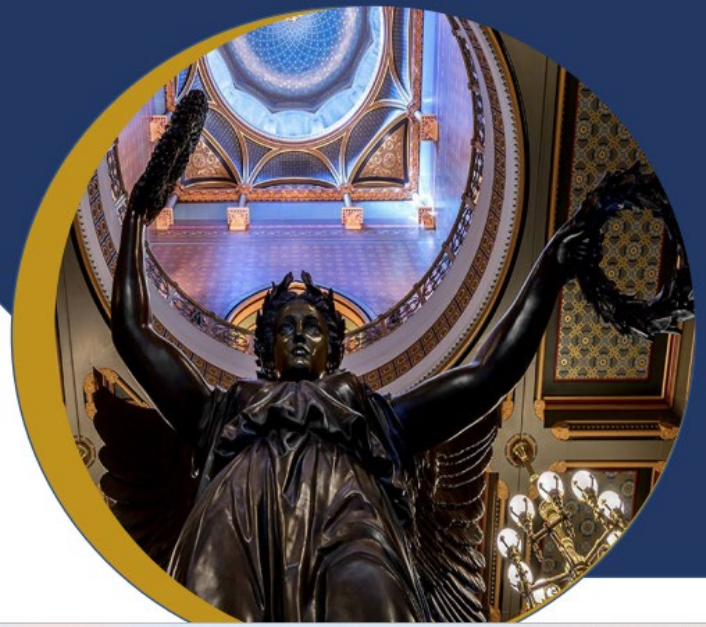


- Open The “Crime In Connecticut Annual Report 2023” PDF
- Review the report so you understand what is in it
- Do a four-corners check – how many pages are there?
- What stories do you see?

CRIME IN CONNECTICUT

2023

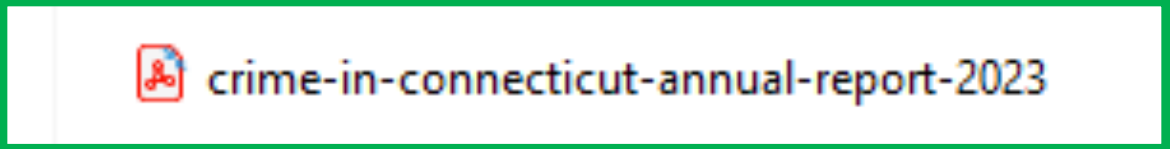
Annual Report of the Uniform Crime Reporting Program



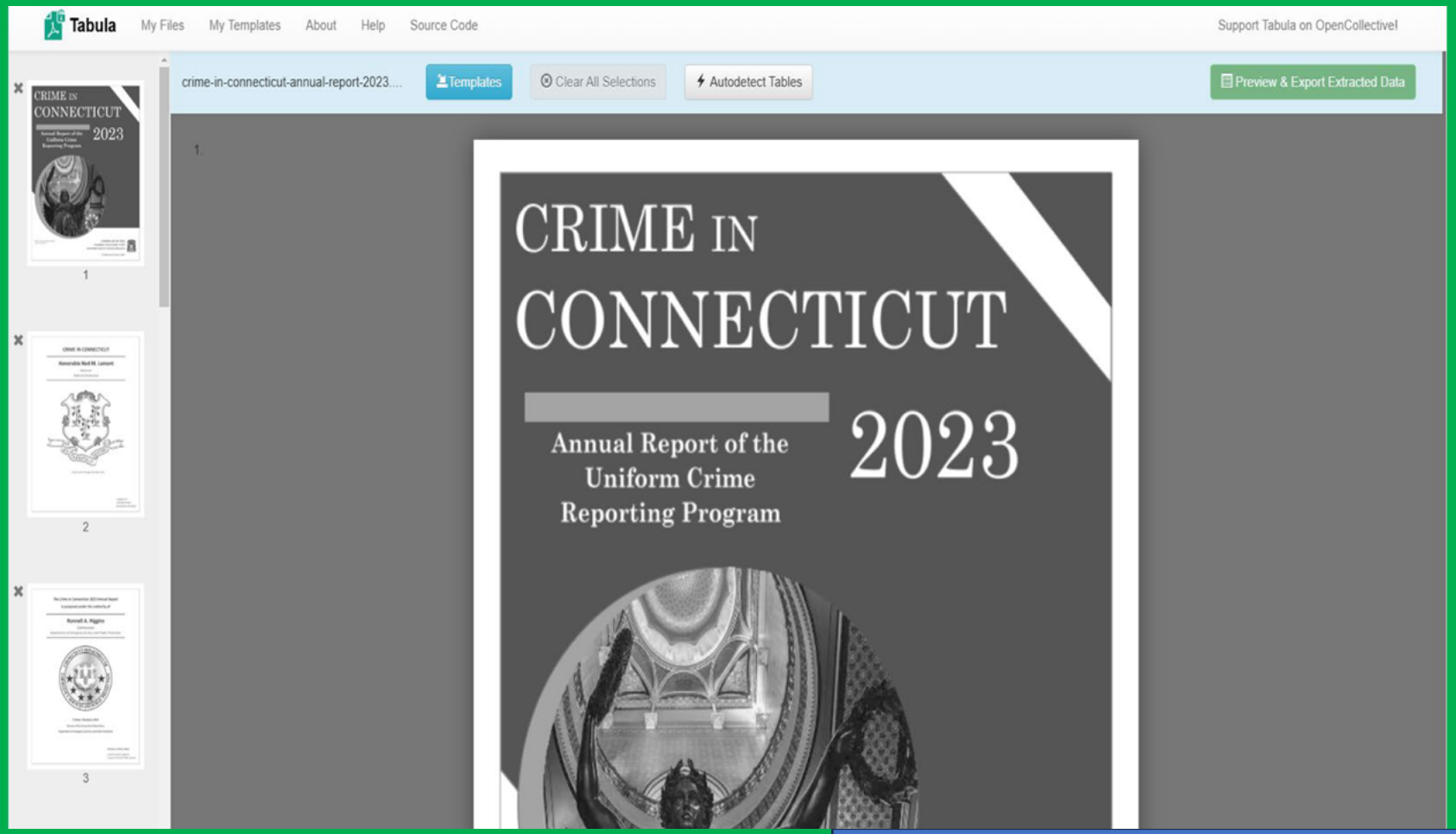
Importing to TABULA



- Open TABULA
- I save the URL to my bookmarks to easily find it
- Click on BROWSE
- Find your saved PDF
- Import it




My Files
My Templates
About
Help
Source Code
Support Tabula on OpenCollective!
crime-in-connecticut-annual-report-2023....
Templates
Clear All Selections
Autodetect Tables
Preview & Export Extracted Data
1
2
3
CRIME IN
CONNECTICUT
2023
Annual Report of the
Uniform Crime
Reporting Program



Do A Four Corners Check – Did You Get the Same Number of Pages and Records?

Previewing

- TABULA will generate images you can PREVIEW
- Scroll through the pages on the left side what it has imported to make sure it all correct
- You can see the tabular data within each page
- This is what TABULA will go in and extract



Click on “Auto Detect Tables” and Tabula Will Look for Tabular Data – and Highlight It in Pink

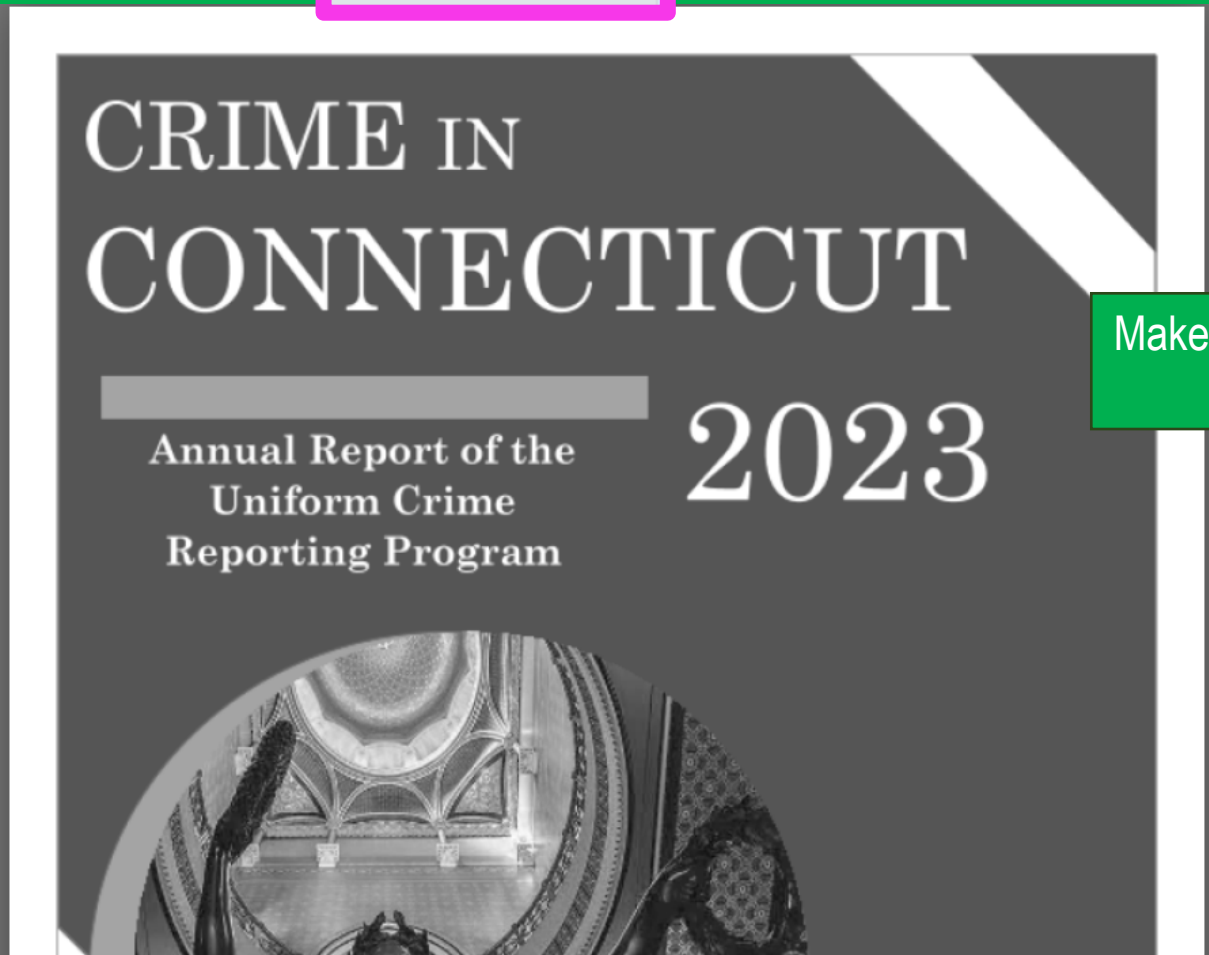
Autodetect Tables


1


2


3

1.



Make Sure To Give It Enough Time to Run

Scroll Through And You Can See It Has Identified Pages With Tabular Data

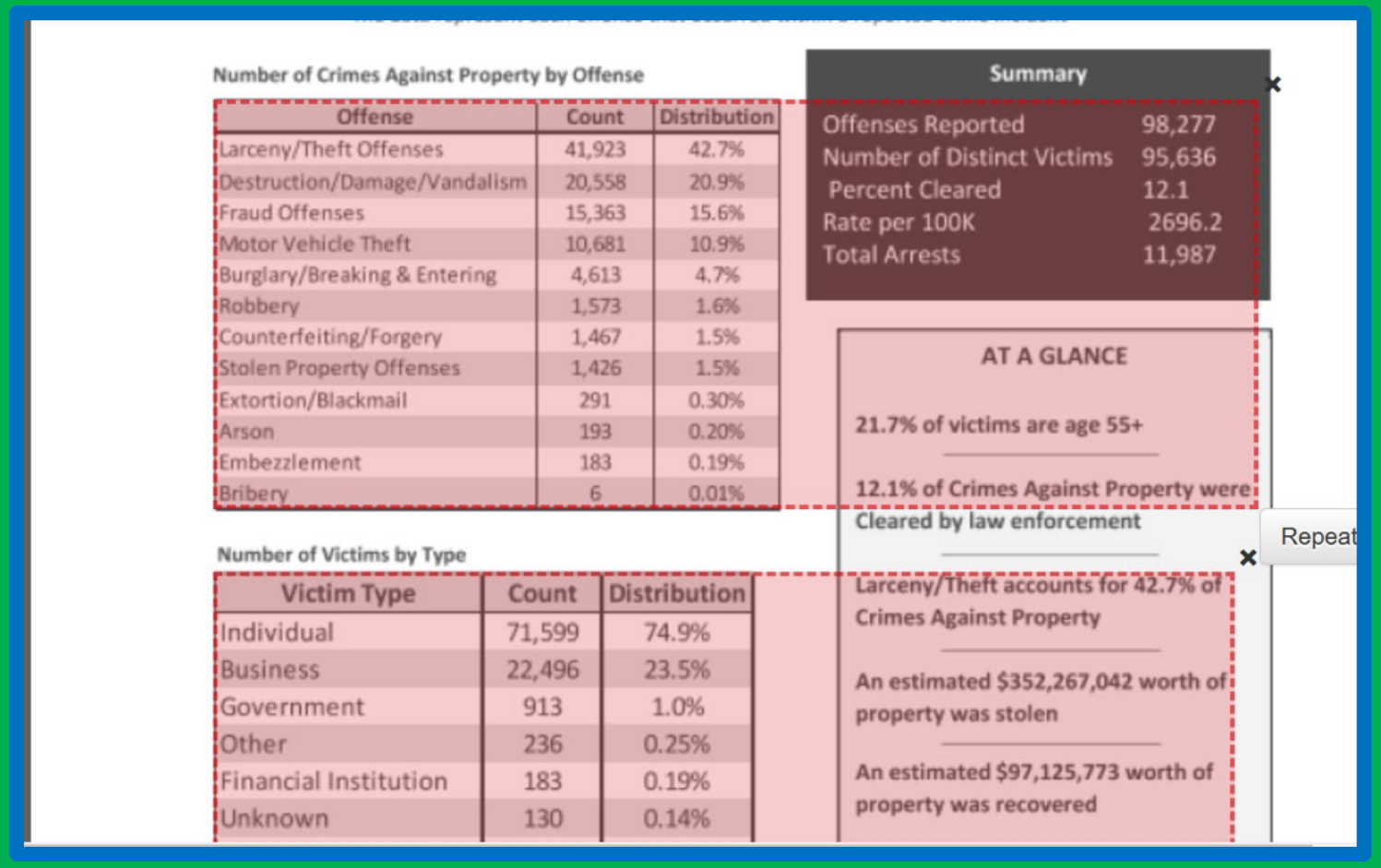


Extracting Data With Tabula

Preview & Export Extracted Data

If you like what you see, you can "**Preview and Export**" the PDF

But you will get a lot of unnecessary pages



Number of Crimes Against Property by Offense

Offense	Count	Distribution
Larceny/Theft Offenses	41,923	42.7%
Destruction/Damage/Vandalism	20,558	20.9%
Fraud Offenses	15,363	15.6%
Motor Vehicle Theft	10,681	10.9%
Burglary/Breaking & Entering	4,613	4.7%
Robbery	1,573	1.6%
Counterfeiting/Forgery	1,467	1.5%
Stolen Property Offenses	1,426	1.5%
Extortion/Blackmail	291	0.30%
Arson	193	0.20%
Embezzlement	183	0.19%
Bribery	6	0.01%

Summary

Offenses Reported	98,277
Number of Distinct Victims	95,636
Percent Cleared	12.1
Rate per 100K	2696.2
Total Arrests	11,987

AT A GLANCE

21.7% of victims are age 55+

12.1% of Crimes Against Property were Cleared by law enforcement

Larceny/Theft accounts for 42.7% of Crimes Against Property

An estimated $352,267,042 worth of property was stolen

An estimated $97,125,773 worth of property was recovered

Number of Victims by Type

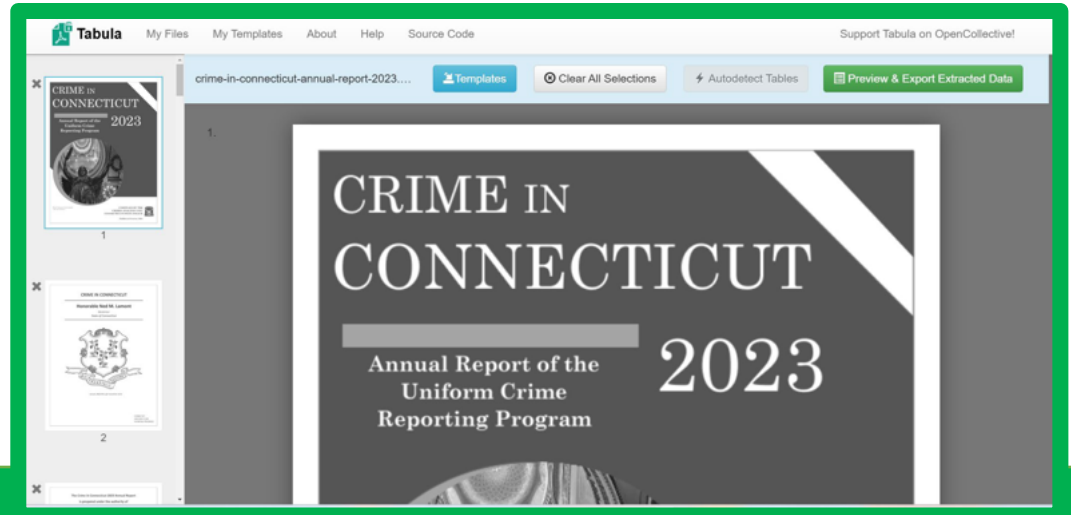
Victim Type	Count	Distribution
Individual	71,599	74.9%
Business	22,496	23.5%
Government	913	1.0%
Other	236	0.25%
Financial Institution	183	0.19%
Unknown	130	0.14%

Repeat

You Can Revise What We Want to Extract

This Is What Is So Effective About Tabula; You Can Work with Multi-Page PDF's But Isolate Out Only the High-Value Pages

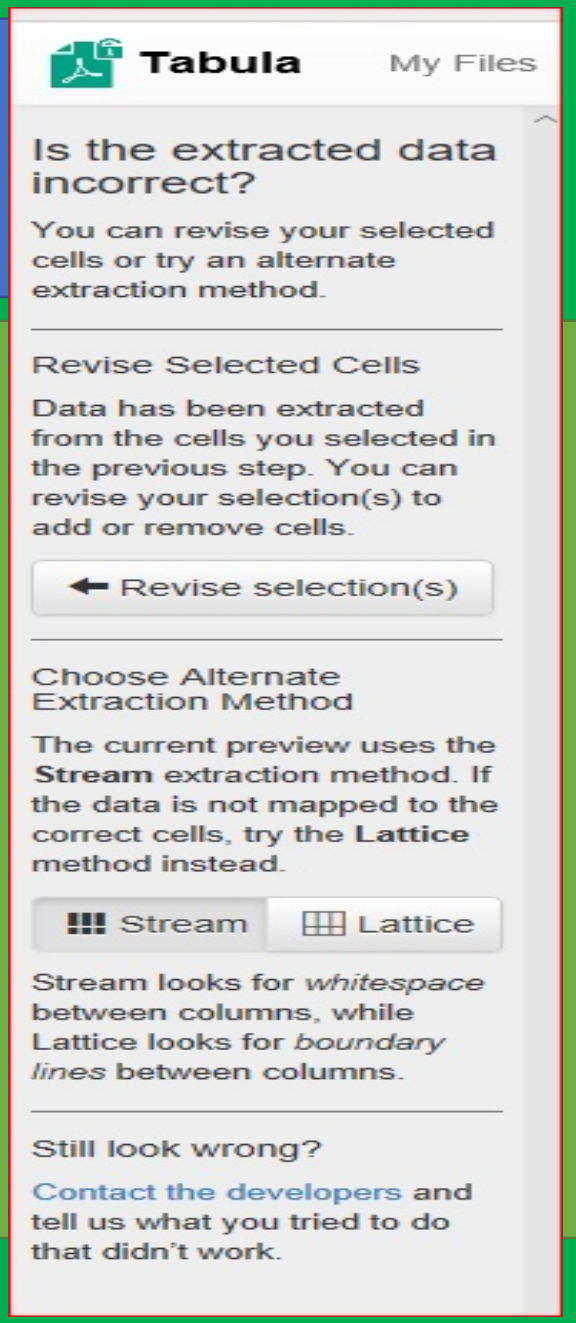
- This is a large document so it would take some time to de-select pages, but it is possible
- When you Auto-Detect Tables, you also have the option of clicking "x" to delete those pages from your extraction
- Just click on "x" and Tabula will remove those pages from your extraction



Omitting Tables

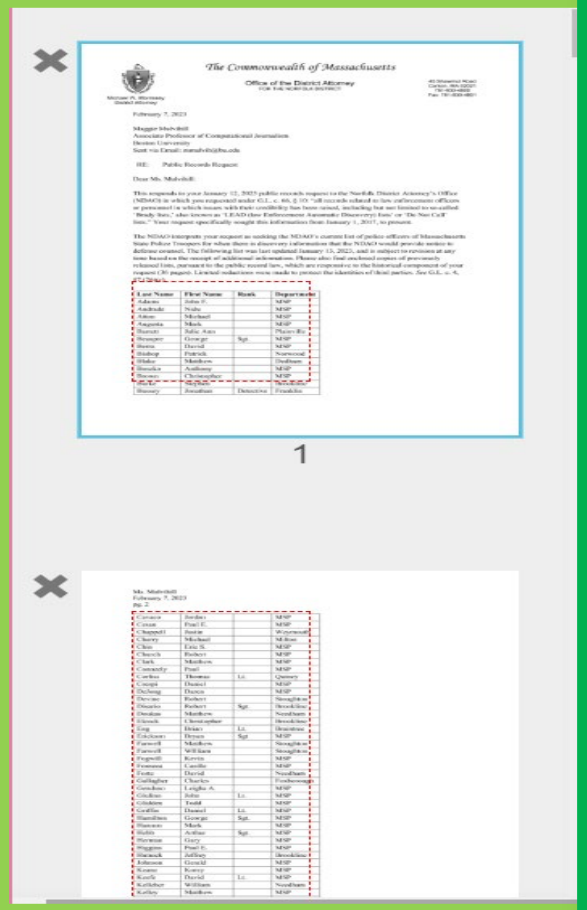
- If you want to change your selection after your extraction
- Just click on REVISE SELECTION
- Can go back to your preview page

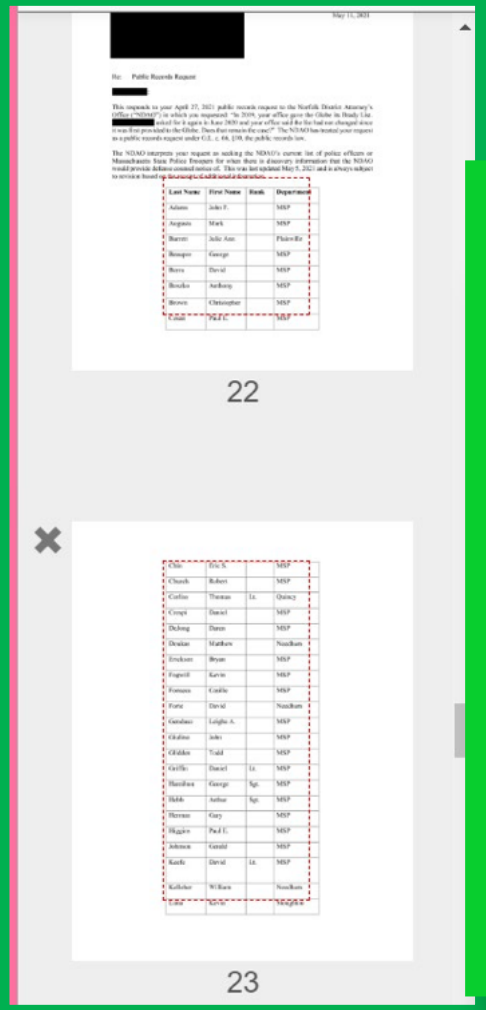


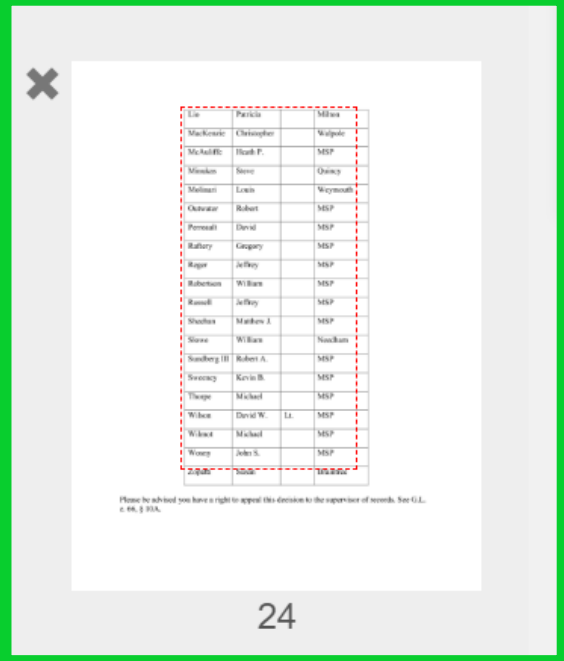


You Can "X" Out Of Pages We Do Not Need

Just Keep The X On The Tabular Data

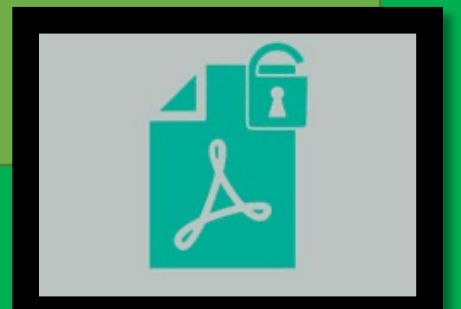






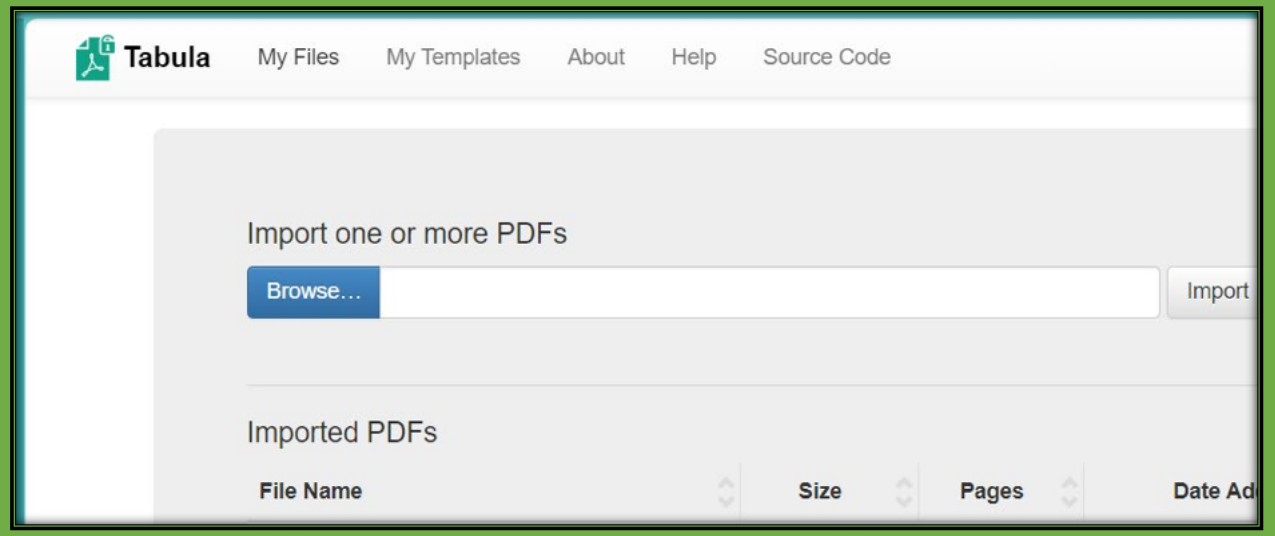
Clearing Selections

- Click CLEAR ALL SECTIONS
- You'll see the pink box around the table is cleared
- Another option: you can also click on the **X** in the upper right next to the selected table
- Close it out
- TABULA won't bring it in

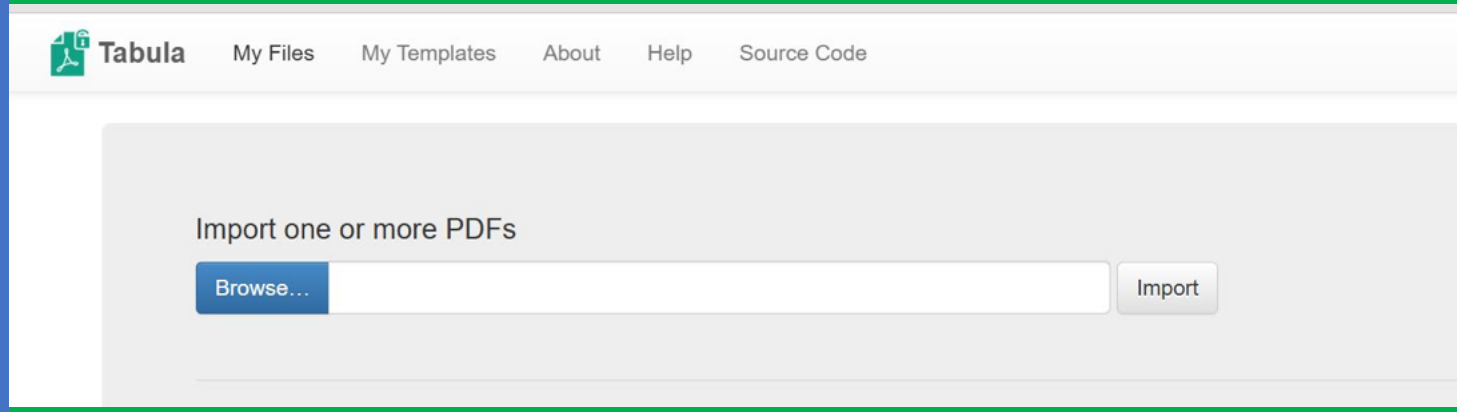


Starting a New Conversion

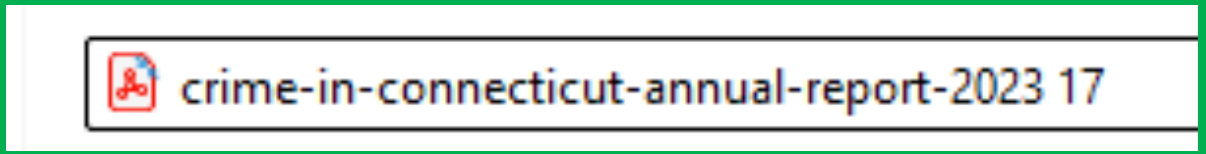
- Let's do a new conversion
- You can always click on the Tabula icon in the upper left to get to the "home" screen



To Get Back to The Home Page, Just Click on the Green Tabula Icon



Let's Extract One Page of This Report



- Download the file "Crime in Connecticut Annual Report 2023 – 17"
- This is just page 17 of this 136-page report
- Now import that to Tabula



Table 1. Distribution of Crimes by Total and Category from 2022 to 2023 and Percentage Change

Group A Offenses	2023 Crimes	2023 Total (%)	2023 Category (%)	2022 Crimes	2022 Total (%)	2022 Category (%)	Change (%)†
Crimes Against Person Total	30,354	21.8	100.0	31,546	24.1	100.0	-3.8
Simple Assault	15,172	10.9	50.0	15,624	12.0	49.5	-2.9
Intimidation	9,861	7.1	32.5	10,478	8.0	33.2	-5.9
Aggravated Assault	3,057	2.2	10.1	3,083	2.4	9.8	-0.8
Criminal Sexual Contact	912	0.7	3.0	1,045	0.8	3.3	-12.7
Rape (2023 revised)	744	0.5	2.5	670	0.5	2.1	11.0
Kidnapping/Abduction	305	0.2	1.0	337	0.3	1.1	-9.5
Murder and Nonnegligent Manslaughter	136	0.1	0.4	137	0.1	0.4	-0.7
Statutory Rape	109	0.1	0.4	115	0.1	0.4	-5.2
Negligent Manslaughter	28	0.0	0.1	26	0.0	0.1	-
Incest	19	0.0	0.1	20	0.0	0.1	-
Human Trafficking Offenses*	11	0.0	0.0	11	0.0	0.0	-
Crimes Against Property Total	98,277	70.7	100.0	88,523	67.8	100.0	11.0
Larceny/Theft Offenses*	41,923	30.2	42.7	43,109	33.0	48.7	-2.8
Destruction/Damage/Vandalism of Property	20,558	14.8	20.9	16,345	12.5	18.5	25.8
Fraud Offenses*	15,363	11.1	15.6	11,890	9.1	13.4	29.2
Motor Vehicle Theft	10,681	7.7	10.9	7,209	5.5	8.1	48.2
Burglary/Breaking & Entering	4,613	3.3	4.7	4,658	3.6	5.3	-1.0
Robbery	1,573	1.1	1.6	1,637	1.3	1.8	-3.9
Counterfeiting/Forgery	1,467	1.1	1.5	1,924	1.5	2.2	-23.8
Stolen Property Offenses	1,426	1.0	1.5	1,102	0.8	1.2	29.4
Extortion/Blackmail	291	0.2	0.3	266	0.2	0.3	9.4
Arson	193	0.1	0.2	196	0.2	0.2	-1.5
Embezzlement	183	0.1	0.2	179	0.1	0.2	2.2
Bribery	6	0.0	0.0	8	0.0	0.0	-
Crimes Against Society Total	10,328	7.4	100.0	10,565	8.1	100.0	-2.2
Drug/Narcotic Violations	4,289	3.1	41.5	4,441	3.4	42.0	-3.4
Weapon Law Violations	3,667	2.6	35.5	3,877	3.0	36.7	-5.4
Drug Equipment Violations	1,832	1.3	17.7	1,777	1.4	16.8	3.1
Pornography/Obscene Material	363	0.3	3.5	296	0.2	2.8	22.6
Animal Cruelty	119	0.1	1.2	127	0.1	1.2	-6.3
Prostitution Offenses*	51	0.0	0.5	38	0.0	0.4	-
Gambling Offenses*	7	0.0	0.1	9	0.0	0.1	-

**Sum of offenses in the category*

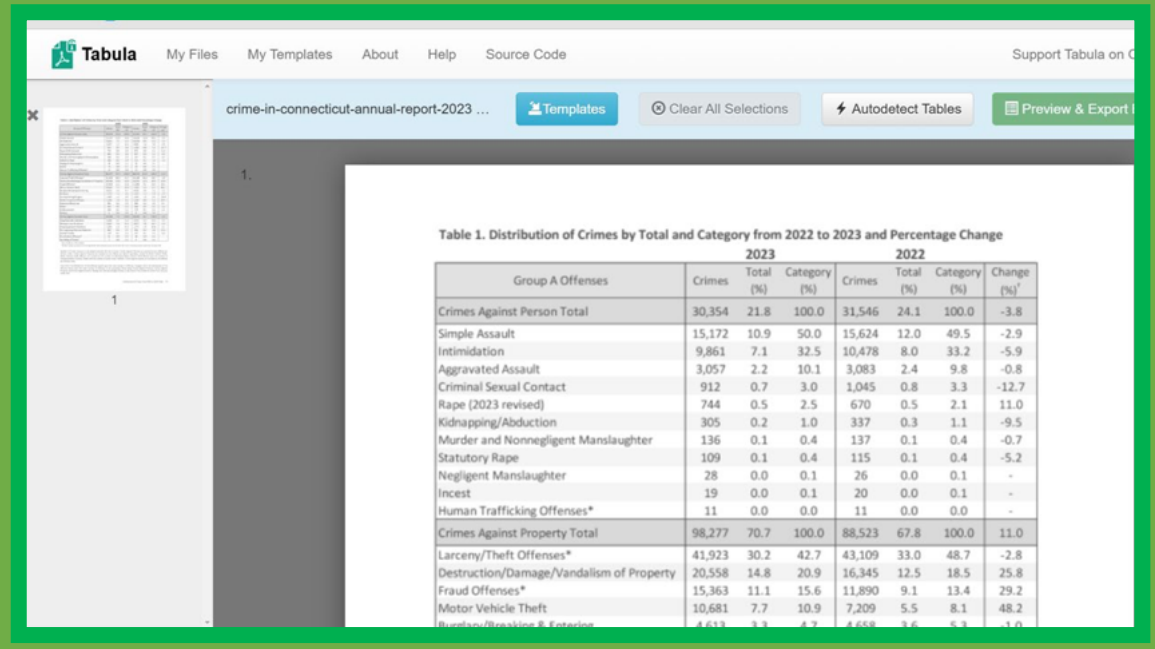
†Relative change calcuations are not appropriate when absolute values are less than 100. Due to rounding, percent totals may not equal 100

Number of Crimes measure is calculated using the FBI Unit of Count. *Crimes Against Persons* are counted as one offense per

Repeat Your Steps

crime-in-connecticut-annual-report-2023 17

- Import the file
- Select “Auto Detect Tables”



Tabula My Files My Templates About Help Source Code Support Tabula on OpenCollective!

crime-in-connecticut-annual-report-2023 … Templates Clear All Selections Autodetect Tables Preview & Export Extracted Data



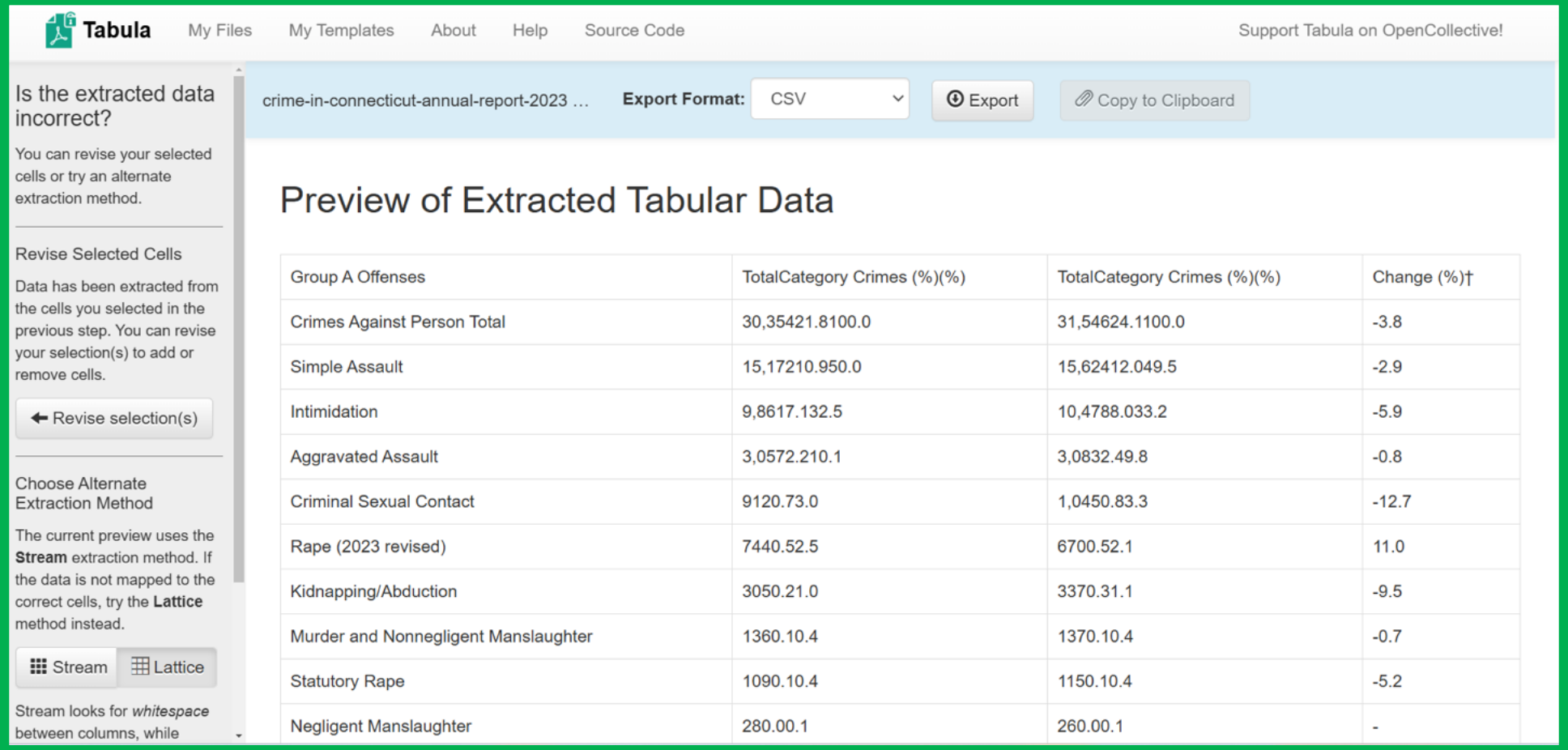
1

1.

Table 1. Distribution of Crimes by Total and Category from 2022 to 2023 and Percentage Change

Group A Offenses	2023 Crimes	2023 Total (%)	2023 Category (%)	2022 Crimes	2022 Total (%)	2022 Category (%)	Change (%)†
Crimes Against Person Total	30,354	21.8	100.0	31,546	24.1	100.0	-3.8
Simple Assault	15,172	10.9	50.0	15,624	12.0	49.5	-2.9
Intimidation	9,861	7.1	32.5	10,478	8.0	33.2	-5.9
Aggravated Assault	3,057	2.2	10.1	3,083	2.4	9.8	-0.8
Criminal Sexual Contact	912	0.7	3.0	1,045	0.8	3.3	-12.7
Rape (2023 revised)	744	0.5	2.5	670	0.5	2.1	11.0
Kidnapping/Abduction	305	0.2	1.0	337	0.3	1.1	-9.5
Murder and Nonnegligent Manslaughter	136	0.1	0.4	137	0.1	0.4	-0.7
Statutory Rape	109	0.1	0.4	115	0.1	0.4	-5.2
Negligent Manslaughter	28	0.0	0.1	26	0.0	0.1	-
Incest	19	0.0	0.1	20	0.0	0.1	-
Human Trafficking Offenses*	11	0.0	0.0	11	0.0	0.0	-
Crimes Against Property Total	98,277	70.7	100.0	88,523	67.8	100.0	11.0
Larceny/Theft Offenses*	41,923	30.2	42.7	43,109	33.0	48.7	-2.8
Destruction/Damage/Vandalism of Property	20,558	14.8	20.9	16,345	12.5	18.5	25.8
Fraud Offenses*	15,363	11.1	15.6	11,890	9.1	13.4	29.2
Motor Vehicle Theft	10,681	7.7	10.9	7,209	5.5	8.1	48.2

Check Your Extraction and Then Click “Export”



Group A Offenses	TotalCategory Crimes (%)(%)	TotalCategory Crimes (%)(%)	Change (%)†
Crimes Against Person Total	30,35421.8100.0	31,54624.1100.0	-3.8
Simple Assault	15,17210.950.0	15,62412.049.5	-2.9
Intimidation	9,8617.132.5	10,4788.033.2	-5.9
Aggravated Assault	3,0572.210.1	3,0832.49.8	-0.8
Criminal Sexual Contact	9120.73.0	1,0450.83.3	-12.7
Rape (2023 revised)	7440.52.5	6700.52.1	11.0
Kidnapping/Abduction	3050.21.0	3370.31.1	-9.5
Murder and Nonnegligent Manslaughter	1360.10.4	1370.10.4	-0.7
Statutory Rape	1090.10.4	1150.10.4	-5.2
Negligent Manslaughter	280.00.1	260.00.1	-



Let’s Export The Data to A CSV

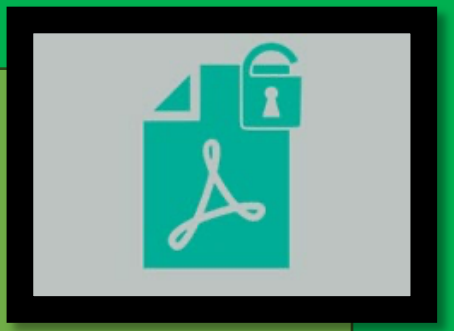


Export Your Data

- Select EXPORT
- Export to a CSV
- Open your spreadsheet
- Do a four corners check?
- What does your top row say?
- What is your bottom record?



Let TABULA Export the Data



	A	B	C	D	E	F
1	Group A Offenses	TotalCategoryCrimes(%)(%)	TotalCategoryCrimes(%)(%)	Change(%)â€		
2	Crimes Against Person Total	30,35421.8100.0	31,54624.1100.0	-3.8		
3	Simple Assault	15,17210.950.0	15,62412.049.5	-2.9		
4	Intimidation	9,8617.132.5	10,4788.033.2	-5.9		
5	Aggravated Assault	3,0572.210.1	3,0832.49.8	-0.8		
6	Criminal Sexual Contact	9120.73.0	1,0450.83.3	-12.7		
7	Rape (2023 revised)	7440.52.5	6700.52.1	11		
8	Kidnapping/Abduction	3050.21.0	3370.31.1	-9.5		
9	Murder and Nonnegligent Manslaughter	1360.10.4	1370.10.4	-0.7		
10	Statutory Rape	1090.10.4	1150.10.4	-5.2		
11	Negligent Manslaughter	280.00.1	260.00.1	-		
12	Incest	190.00.1	200.00.1	-		
13	Human Trafficking Offenses*	110.00.0	110.00.0	-		
14	Crimes Against Property Total	98,27770.7100.0	88,52367.8100.0	11		
15	Larceny/Theft Offenses*	41,92330.242.7	43,10933.048.7	-2.8		
16	Destruction/Damage/Vandalism of Property	20,55814.820.9	16,34512.518.5	25.8		
17	Fraud Offenses*	15,36311.115.6	11,8909.113.4	29.2		
18	Motor Vehicle Theft	10,6817.710.9	7,2095.58.1	48.2		
19	Burglary/Breaking & Entering	4,6133.34.7	4,6583.65.3	-1		
20	Robbery	1,5731.11.6	1,6371.31.8	-3.9		
21	Counterfeiting/Forgery	1,4671.11.5	1,9241.52.2	-23.8		
22	Stolen Property Offenses	1,4261.01.5	1,1020.81.2	29.4		
23	Extortion/Blackmail	2910.20.3	2660.20.3	9.4		
24	Arson	1930.10.2	1960.20.2	-1.5		
25	Embezzlement	1830.10.2	1790.10.2	2.2		
26	Bribery	60.00.0	80.00.0	-		
27	Crimes Against Society Total	10,3287.4100.0	10,5658.1100.0	-2.2		
28	Drug/Narcotic Violations	4,2893.141.5	4,4413.442.0	-3.4		
29	Weapon Law Violations	3,6672.635.5	3,8773.036.7	-5.4		
30	Drug Equipment Violations	1,8321.317.7	1,7771.416.8	3.1		

Check Your Data: Did It Export Correctly?

You Might Have To Do A Little Clean-Up With Your Export

What Is Your Lede?

	A	B	C	D	E
1	Group A Offenses	TotalCategoryCrimes(%)(%)	TotalCategoryCrimes(%)(%)	Change(%)â€	
2	Crimes Against Person Total	30,35421.8100.0	31,54624.1100.0	-3.8	
3	Simple Assault	15,17210.950.0	15,62412.049.5	-2.9	
4	Intimidation	9,8617.132.5	10,4788.033.2	-5.9	
5	Aggravated Assault	3,0572.210.1	3,0832.49.8	-0.8	
6	Criminal Sexual Contact	9120.73.0	1,0450.83.3	-12.7	
7	Rape (2023 revised)	7440.52.5	6700.52.1	11	
8	Kidnapping/Abduction	3050.21.0	3370.31.1	-9.5	
9	Murder and Nonnegligent Manslaughter	1360.10.4	1370.10.4	-0.7	
10	Statutory Rape	1090.10.4	1150.10.4	-5.2	
11	Negligent Manslaughter	280.00.1	260.00.1	-	
12	Incest	190.00.1	200.00.1	-	
13	Human Trafficking Offenses*	110.00.0	110.00.0	-	
14	Crimes Against Property Total	98,27770.7100.0	88,52367.8100.0	11	
15	Larceny/Theft Offenses*	41,92330.242.7	43,10933.048.7	-2.8	
16	Destruction/Damage/Vandalism of Property	20,55814.820.9	16,34512.518.5	25.8	
17	Fraud Offenses*	15,36311.115.6	11,8909.113.4	29.2	
18	Motor Vehicle Theft	10,6817.710.9	7,2095.58.1	48.2	
19	Burglary/Breaking & Entering	4,6133.34.7	4,6583.65.3	-1	
20	Robbery	1,5731.11.6	1,6371.31.8	-3.9	
21	Counterfeiting/Forgery	1,4671.11.5	1,9241.52.2	-23.8	
22	Stolen Property Offenses	1,4261.01.5	1,1020.81.2	29.4	
23	Extortion/Blackmail	2910.20.3	2660.20.3	9.4	
24	Arson	1930.10.2	1960.20.2	-1.5	
25	Embezzlement	1830.10.2	1790.10.2	2.2	
26	Bribery	60.00.0	80.00.0	-	
27	Crimes Against Society Total	10,3287.4100.0	10,5658.1100.0	-2.2	
28	Drug/Narcotic Violations	4,2893.141.5	4,4413.442.0	-3.4	
29	Weapon Law Violations	3,6672.635.5	3,8773.036.7	-5.4	
30	Drug Equipment Violations	1,8321.317.7	1,7771.416.8	3.1	

tabula-crime-in-connecticut-ann



Practice With The Connecticut Department of Revenue's Annual Report

I have given you the full 100-page report which you can import for practice

I've also given you just a few key pages to practice with



STATE OF CONNECTICUT
DEPARTMENT OF REVENUE SERVICES

ANNUAL REPORT
FISCAL YEAR 2022-2023

STATE REVENUE SOURCES

	FY 2020-2021	FY 2021-2022	FY 2022-2023
Admissions & Dues Tax *Ch. 225* and Transportation Network Company Fee *Ch. 244b*	$40,017,155	$44,945,636	$45,959,625
Alcoholic Beverages Tax *Ch. 220*	79,111,165	78,915,528	80,247,957
Ambulatory Surgical Centers *Ch. 211a (repealed)*	17,838,566	21,479,884	0
Attorney Occupational Tax *Ch. 876*	6,948,701	11,912,450	9,322,121
Beverage Container Deposits *Ch. 446d*	43,505,264	48,258,877	52,584,520
Business Entity Tax *Ch. 213a (repealed)*	6,550,175	833,643	(6,863,524)
Cable, Satellite & Video Companies Tax *Ch. 211*	45,010,159	39,072,064	49,226,215
Cannabis Tax, *Ch. 214c*	0	0	4,675,642
Cigarette Tax *Ch. 214*	324,060,357	301,655,083	268,003,557
Controlled Substances Tax *Ch. 228d*	10,511	12,008	916
Controlling Interest Transfer Tax *Ch. 228b*	10,365,593	15,604,331	9,595,329
Corporation Business Tax *Ch. 208* and Air Carriers Tax *Ch. 209*	1,144,305,750	1,397,484,273	1,519,011,011
Dry Cleaning Surcharge *Ch. 211b*	335,472	503,406	542,984
Electric Distribution Companies *Ch. 212*	138,337,220	167,554,267	149,579,820
Electronic Cigarette Products *Ch. 214b*	5,205,231	6,165,394	6,181,983
Estate and Gift Tax *Ch. 217& 228c*	303,338,908	220,117,132	218,354,438
Gas Companies *Ch. 212*	69,081,934	96,084,499	90,207,200
Gift Tax *Ch. 228c (repealed)*	39,376	154,128	41,392
Health Care Centers Tax *Ch. 207*	1,911,391	2,135,169	3,401,206
Health Care Provider Taxes and Fees *Ch. 211c*	1,019,867,122	933,602,663	900,978,299
Highway Use Fee *Ch. 222a*	0	0	29,275,901
Income Tax *Ch. 229*	10,273,322,918	12,130,517,104	11,231,163,886
Insurance Premiums Tax, Domestic *Ch. 207*	32,534,668	29,487,263	47,897,022
Insurance Premiums Tax, Foreign *Ch. 207*	158,792,653	163,866,266	192,717,482
Insurance: Non-admitted/Unauthorized *Ch. 698d*	9,025,345	9,212,643	10,288,462
Motor Carrier Road Tax - *Ch. 222*	11,455,267	8,768,694	9,893,330
Motor Vehicle Fuels Tax - Gasoline *Ch. 221*	339,398,960	267,102,572	118,263,374
Motor Vehicle Fuels Tax - Special Fuel *Ch. 221*	124,583,899	113,760,376	135,366,064
Paid Preparer Fee *Ch. 229*	93,976	69,538	92,159
Pass-Through Entity *Ch. 228z*	1,553,825,428	2,308,213,258	2,049,126,696
Petroleum Products Gross Earnings Tax *Ch. 227*	232,333,711	397,956,063	392,055,527

Practice With The Connecticut Department of Revenue's Annual Report

I have given you the full 100-page report which you can import for practice

I've also given you just a few key pages to practice with

STATE OF CONNECTICUT
DEPARTMENT OF REVENUE SERVICES

ANNUAL REPORT
FISCAL YEAR 2022-2023



STATE REVENUE SOURCES

	FY 2020-2021	FY 2021-2022	FY 2022-2023
Admissions & Dues Tax *Ch. 225* and Transportation Network Company Fee *Ch. 244b*	$40,017,155	$44,945,636	$45,959,625
Alcoholic Beverages Tax *Ch. 220*	79,111,165	78,915,528	80,247,957
Ambulatory Surgical Centers *Ch. 211a (repealed)*	17,838,566	21,479,884	0
Attorney Occupational Tax *Ch. 876*	6,948,701	11,912,450	9,322,121
Beverage Container Deposits *Ch. 446d*	43,505,264	48,258,877	52,584,520
Business Entity Tax *Ch. 213a (repealed)*	6,550,175	833,643	(6,863,524)
Cable, Satellite & Video Companies Tax *Ch. 211*	45,010,159	39,072,064	49,226,215
Cannabis Tax, *Ch. 214c*	0	0	4,675,642
Cigarette Tax *Ch. 214*	324,060,357	301,655,083	268,003,557
Controlled Substances Tax *Ch. 228d*	10,511	12,008	916
Controlling Interest Transfer Tax *Ch. 228b*	10,365,593	15,604,331	9,595,329
Corporation Business Tax *Ch. 208* and Air Carriers Tax *Ch. 209*	1,144,305,750	1,397,484,273	1,519,011,011
Dry Cleaning Surcharge *Ch. 211b*	335,472	503,406	542,984
Electric Distribution Companies *Ch. 212*	138,337,220	167,554,267	149,579,820
Electronic Cigarette Products *Ch. 214b*	5,205,231	6,165,394	6,181,983
Estate and Gift Tax *Ch. 217& 228c*	303,338,908	220,117,132	218,354,438
Gas Companies *Ch. 212*	69,081,934	96,084,499	90,207,200
Gift Tax *Ch. 228c (repealed)*	39,376	154,128	41,392
Health Care Centers Tax *Ch. 207*	1,911,391	2,135,169	3,401,206
Health Care Provider Taxes and Fees *Ch. 211c*	1,019,867,122	933,602,663	900,978,299
Highway Use Fee *Ch. 222a*	0	0	29,275,901
Income Tax *Ch. 229*	10,273,322,918	12,130,517,104	11,231,163,886
Insurance Premiums Tax, Domestic *Ch. 207*	32,534,668	29,487,263	47,897,022
Insurance Premiums Tax, Foreign *Ch. 207*	158,792,653	163,866,266	192,717,482
Insurance: Non-admitted/Unauthorized *Ch. 698d*	9,025,345	9,212,643	10,288,462
Motor Carrier Road Tax - *Ch. 222*	11,455,267	8,768,694	9,893,330
Motor Vehicle Fuels Tax - Gasoline *Ch. 221*	339,398,960	267,102,572	118,263,374
Motor Vehicle Fuels Tax - Special Fuel *Ch. 221*	124,583,899	113,760,376	135,366,064
Paid Preparer Fee *Ch. 229*	93,976	69,538	92,159
Pass-Through Entity *Ch. 228z*	1,553,825,428	2,308,213,258	2,049,126,696
Petroleum Products Gross Earnings Tax *Ch. 227*	232,333,711	397,956,063	392,055,527



Using PDFtoEXCEL – This Can Scan and Extract Image-Based PDF's

Please be informed that the following persons or firms have been debarred in accordance with Section 31-53a of the Connecticut General Statutes, as amended:

DEBARMENT LIST

Name of Person or Firm	Expiration Date
-Canterbury Communications, Canterbury	Indefinite
Marion T. Mershon, an individual	Indefinite
-Elite Roofing, Cheshire	Indefinite
Sam Stevens, an individual	Indefinite
-Gatollari Enterprises Inc., Rochelle Park, NJ	Indefinite
Robert Gatollari, an individual	Indefinite
Greenland Enterprises Inc., Wayne, NJ	Indefinite

Open The File: debarment_march_2024 file

No contract shall be awarded to any person or firms appearing on this list or to any firm, corporation, partnership, or association in which such persons or firms have an interest until the expiration date listed has elapsed.

Please be informed that the following persons or firms have been debarred in accordance with Section 31-53a of the Connecticut General Statutes, as amended:

DEBARMENT LIST

Name of Person or Firm	
-Canterbury Communications, Canterbury	Indefinite
Marion T. Mershon, an individual	Indefinite
-Elite Roofing, Cheshire	Indefinite
Sam Stevens, an individual	Indefinite
-Gatollari Enterprises Inc., Rochelle Park, NJ	Indefinite
-Greenlar	
-Jersey Welding Company Inc., Hartford	Indefinite

PDFtoExcel.com

This is a list of companies Connecticut state officials have "barred" from doing business with it

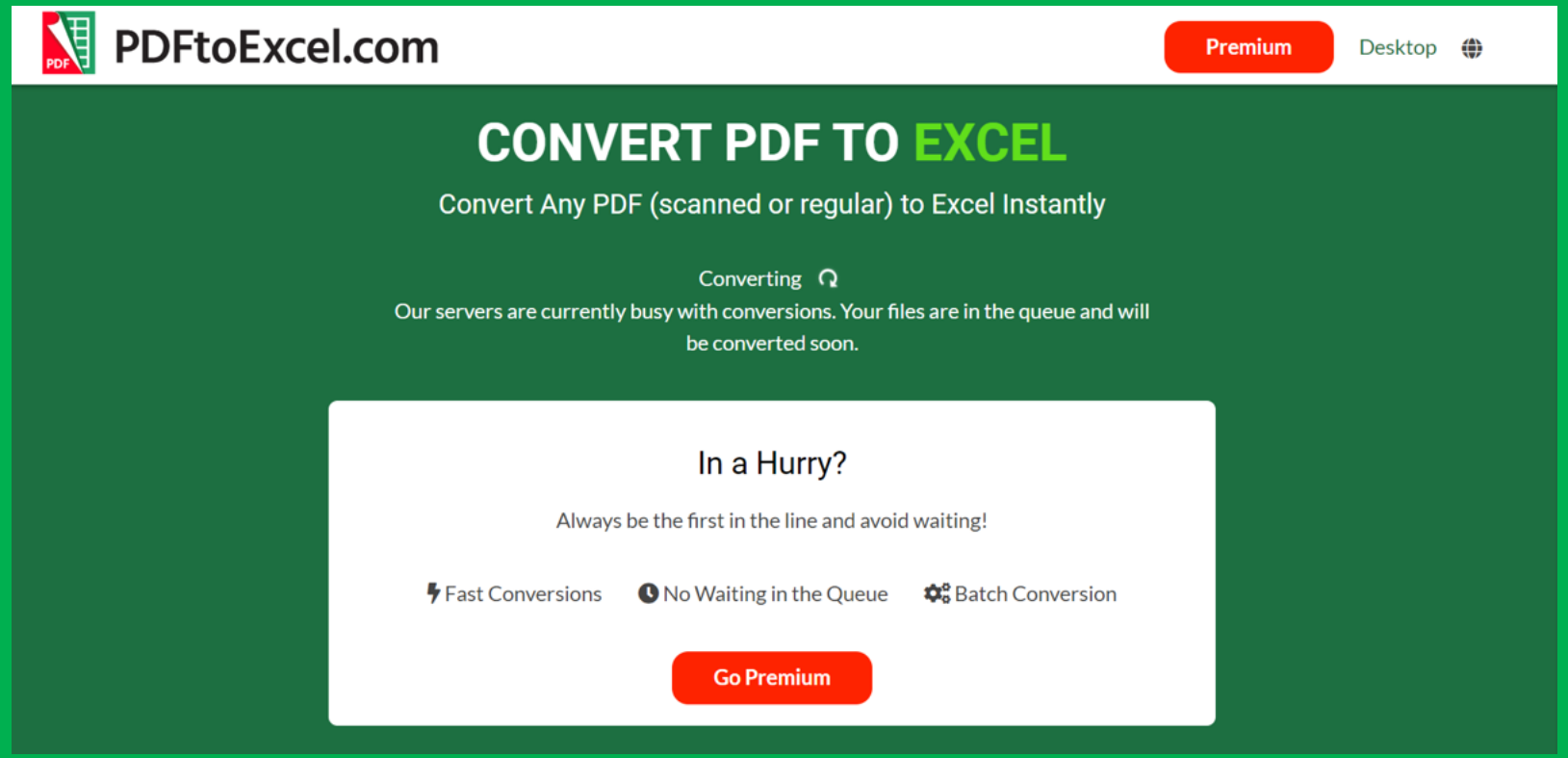
Click “Upload” And Find Your File

Do A Four Corners Check

What Is In The First Line of Records?

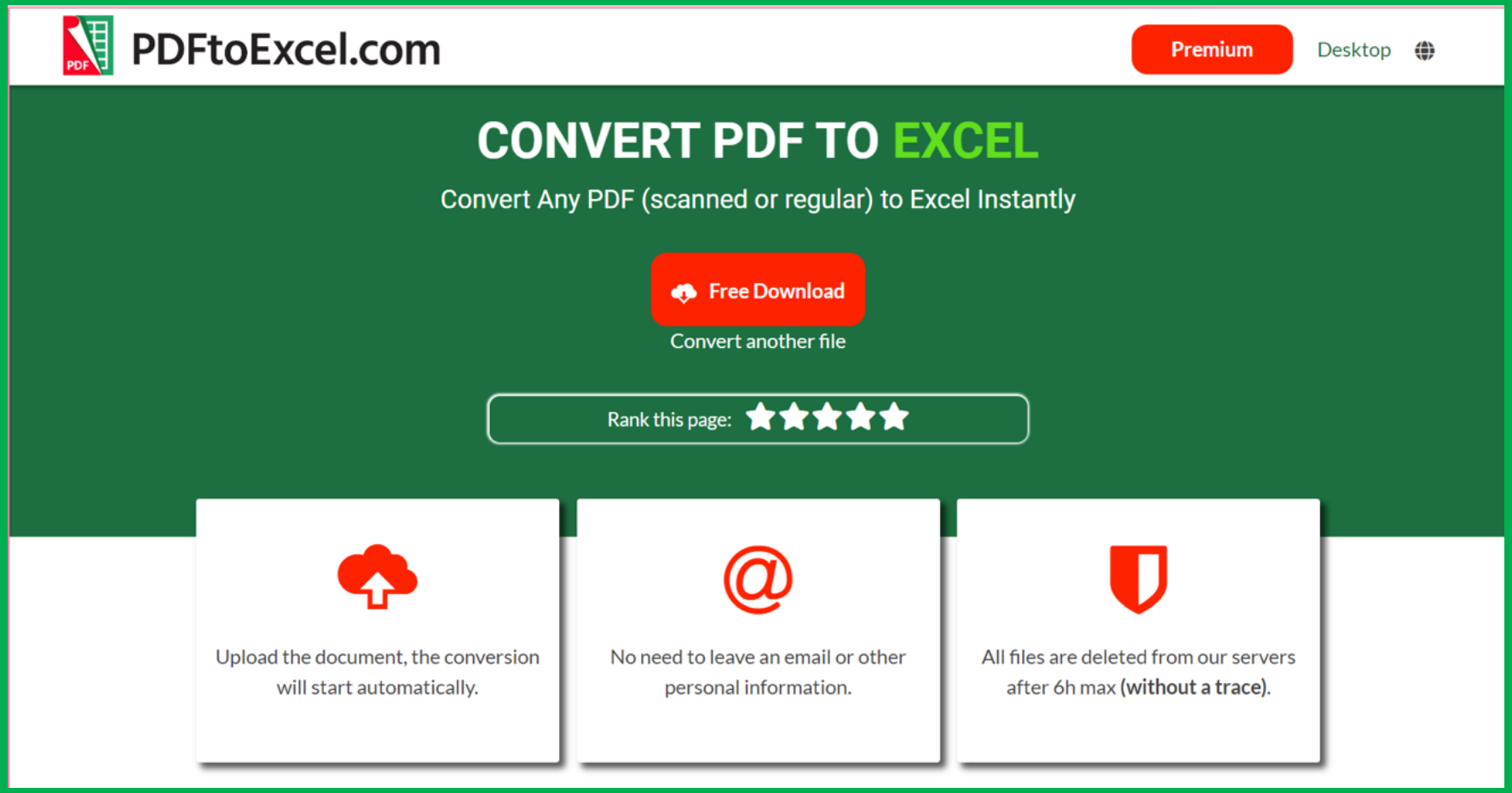
What Is In The Last Line?

Go to PDFtoEXCEL



PDFtoExcel.com

Once You Upload, Your File Will Immediately Start Converting To Excel



When It Is Finished, You'll See The Download Option Click It



CONNECTICUT

. DEPARTMENT

OF LABOR*

March 12, 2024

Fortner of the American

Job Center Network

State of Connecticut Debarment List Pursuant to Section 31-53a(a)

No contract shall be awarded to any person or firms appearing on this list or to any firm, corporation, partnership, or association in which such persons or firms have an interest until the expiration date listed has elapsed.

Please be informed that the following persons or firms have been debarred in accordance with Section 31-53a of the Connecticut General Statutes, as amended:

DEBARMENT LIST

Name of Person or Firm	Expiration Date
-Canterbury Communications, Canterbury	Indefinite
Marion T. Mershon, an individual	Indefinite
-Elite Roofing, Cheshire	Indefinite
Sam Stevens, an individual	Indefinite
-Gatollari Enterprises Inc., Rochelle Park, NJ	Indefinite
Robert Gatollari, an individual	Indefinite
-Greenland Enterprises Inc., Wayne, NJ	Indefinite
Oneil Mendez, an individual	Indefinite
Raymond Blum-Romero, an individual	Indefinite
-Jarosz Welding Company Inc., Hartford	Indefinite
Andrew W. Jarosz, an individual	Indefinite
-SCI Services, LLC, Somers	Indefinite
Philip Snow, an individual	Indefinite
-Seakco Construction, Stamford	Indefinite
Robert Evans, an individual	Indefinite
-Sun Ray Electric, Prospect	Indefinite







Web Scraping

Data Journalism

Web Scraping


edureka!
Apps
Development
XML
CSV
Webpages
Web Scraping
Structured Data



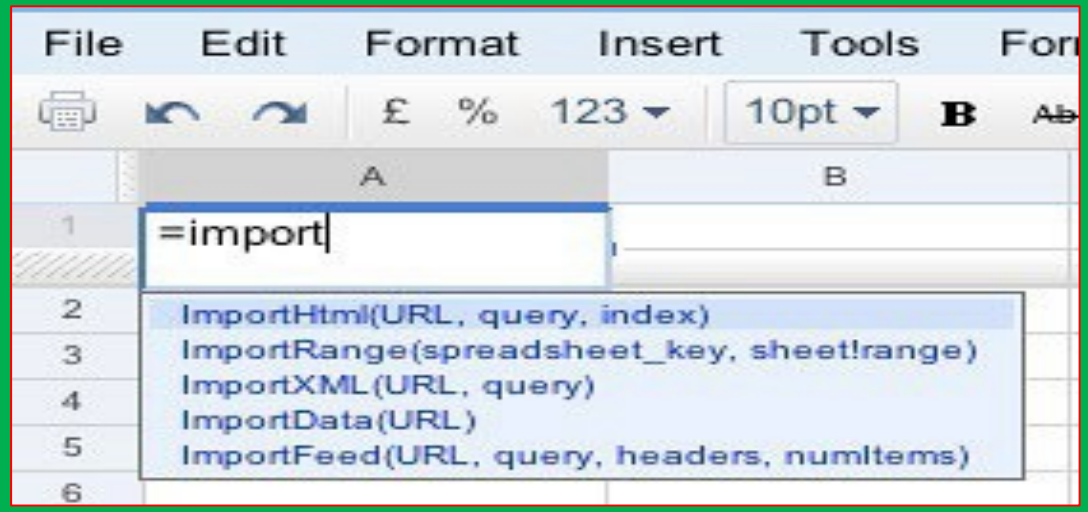
Why We Scrape Data From The Web

Why Scrape?

We can download data

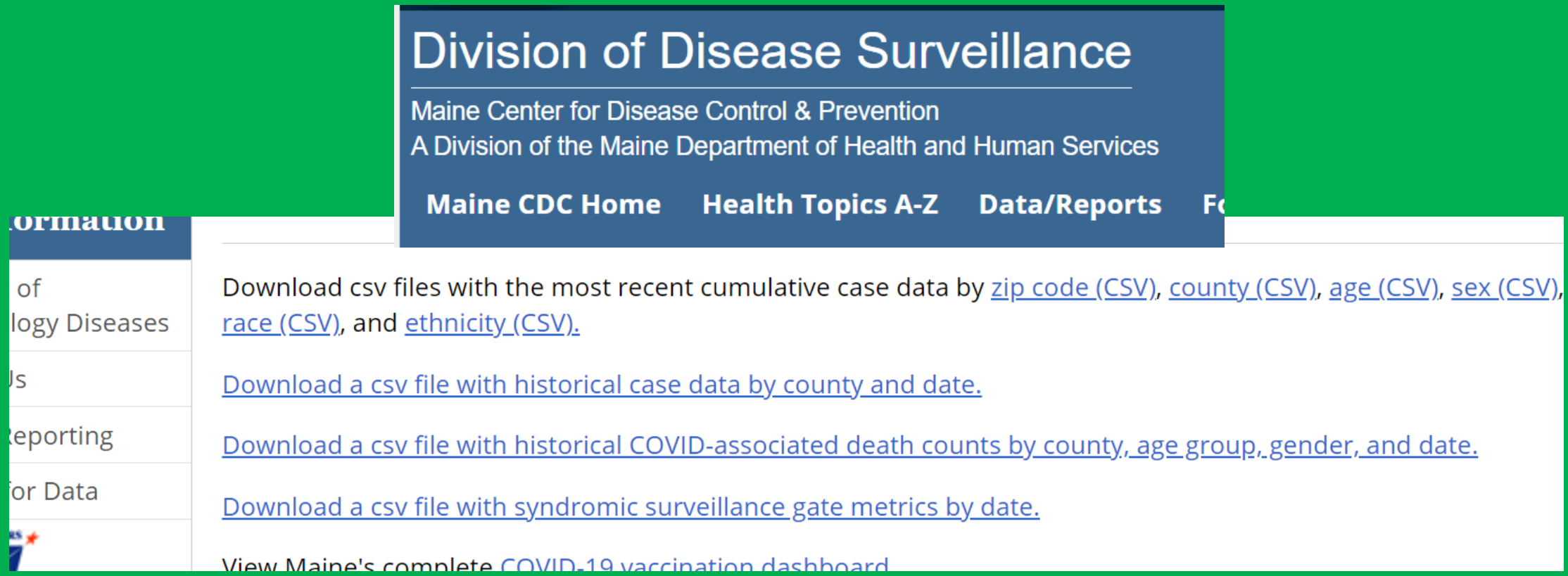
We can even copy and paste tables

What is the advantage to scraping?



Data scraping	Data Crawling
Involves extracting data from various sources including web	Refers to downloading pages from the web
Can be done at any scale	Mostly done at a large scale
Deduplication is not necessarily a part	Deduplication is an essential part
Needs crawl agent and parser	Needs only crawl agent

Sometimes We Don't Always Have A Convenient DOWNLOAD Option On A Web Site



Why Scrape? Wikipedia Page of Olympic Cities

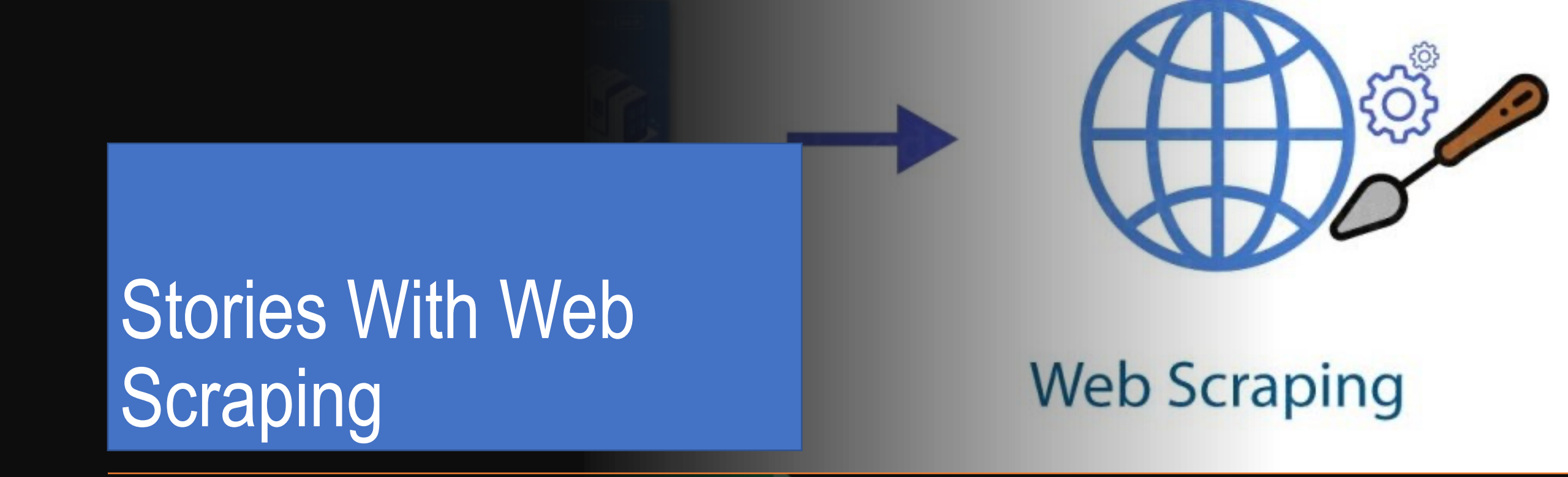
- Advantage of Scraping:
  - Each time a webpage is updated, your spreadsheet will be updated
  - Some pages have “download” buttons that allow you to export the data
  - But that download will never be automatically updated unless you go back to it and download it again
  - With scraping in G-Sheets, you always have updated data



ic Games [edit]

untry	Year	Continent	Summer	Winter	Opening ceremony
ce	1896	Europe	I		6 April 1896
nce	1900		II		14 May 1900
nited States	1904	North America	III		1 July 1904
United Kingdom	1908	Europe	IV		27 April 1908
Sweden	1912		V		6 July 1912
Germany	1916		~~VI~~		Can
Belgium	1920		VII		14 August 1920
France	1924			I	25 January 1924
France			VIII		5 July 1924
Switzerland	1928			II	11 February 1928
Netherlands			IX		28 July 1928
United States	1932	North America		III	4 February 1932
United States			X		30 July 1932
Germany	1936	Europe		IV	6 February 1936
Germany			XI		1 August 1936
Japan Germany	1940	Asia Europe		~~V~~	
Japan Finland			~~XII~~		Can
Italy	1944	Europe		~~V~~	
United Kingdom			~~XIII~~		
Switzerland	1948	Europe		V	30 January 1948
nited Kingdom			XIV		29 July 1948
way	1952			VI	14 February 1952
nd			XV		19 July 1952

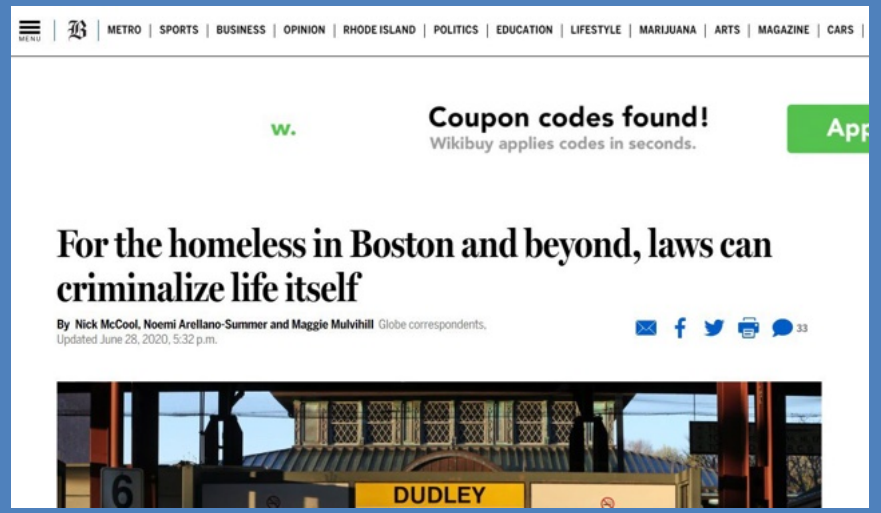
Stories With Web Scraping



Web Scraping 101



The Boston Globe



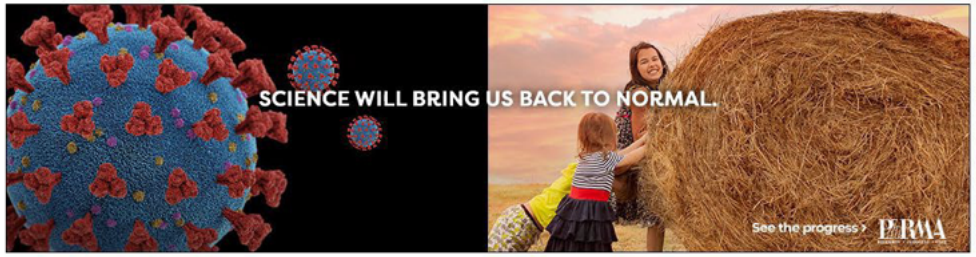
The Washington Post

Sections

The Washington Post
Democracy Dies in Darkness

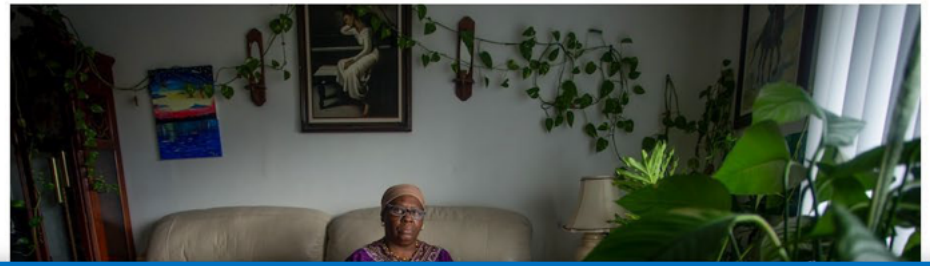
Get 1 year for $29

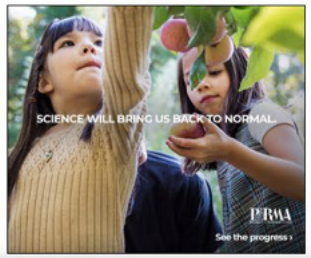
Sign in



Health

Confusion over eviction ban led to selective enforcement

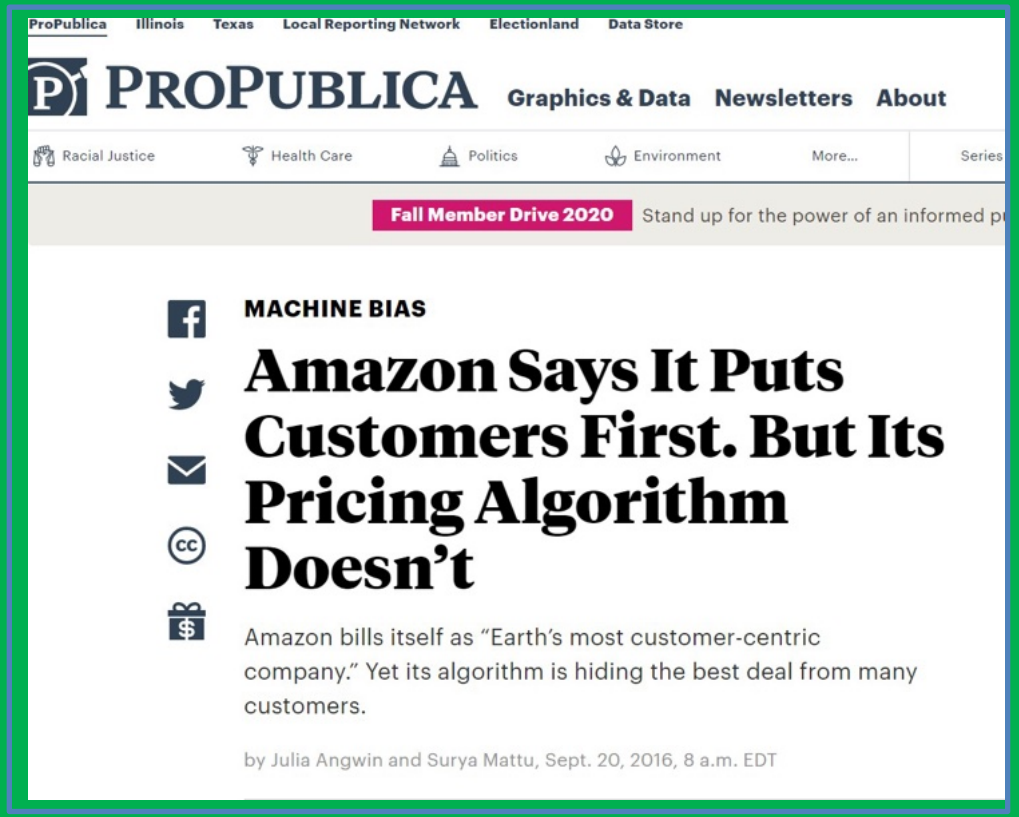




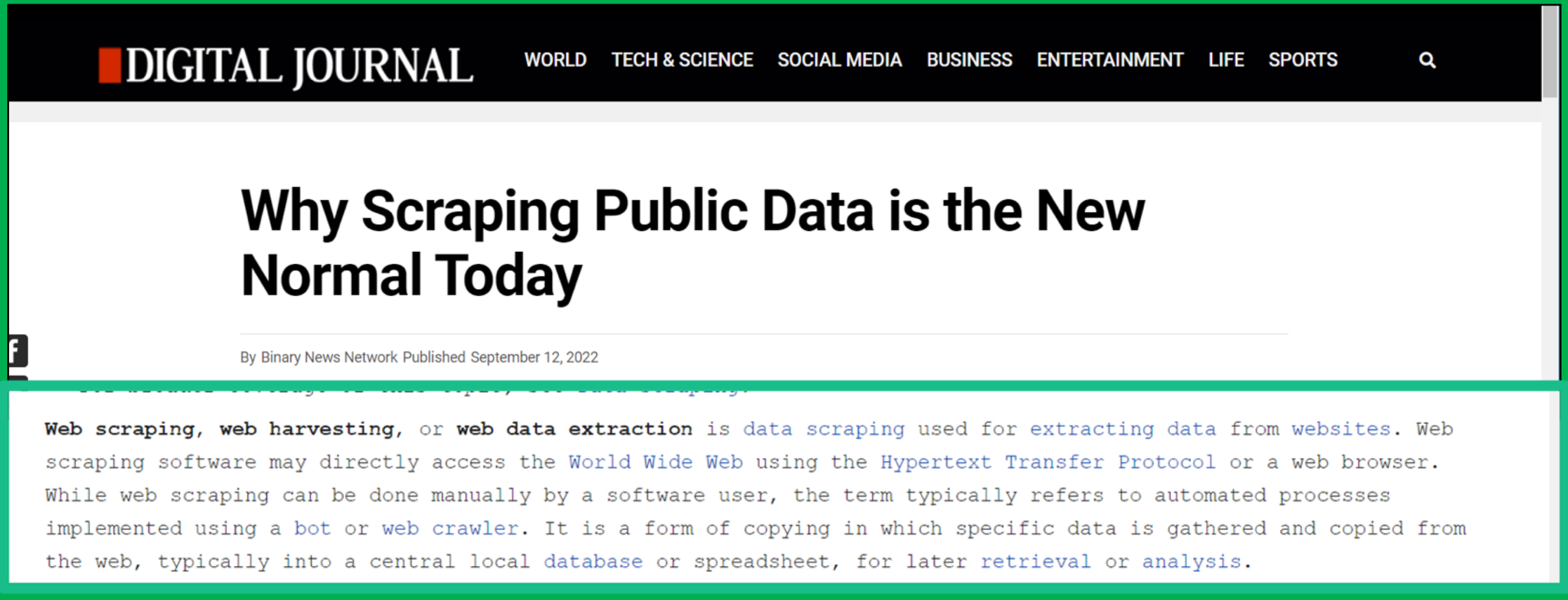
By **Brenda Wintrode, Amy Dipierro, Ryan Little, Luciana Perez Uribe, Aneurin Canham-Clyne, Trisha Ahmed, Sean Mcgoey, Sophia Brown and Maya Pottiger / The Howard Center for Investigative Journalism | AP**

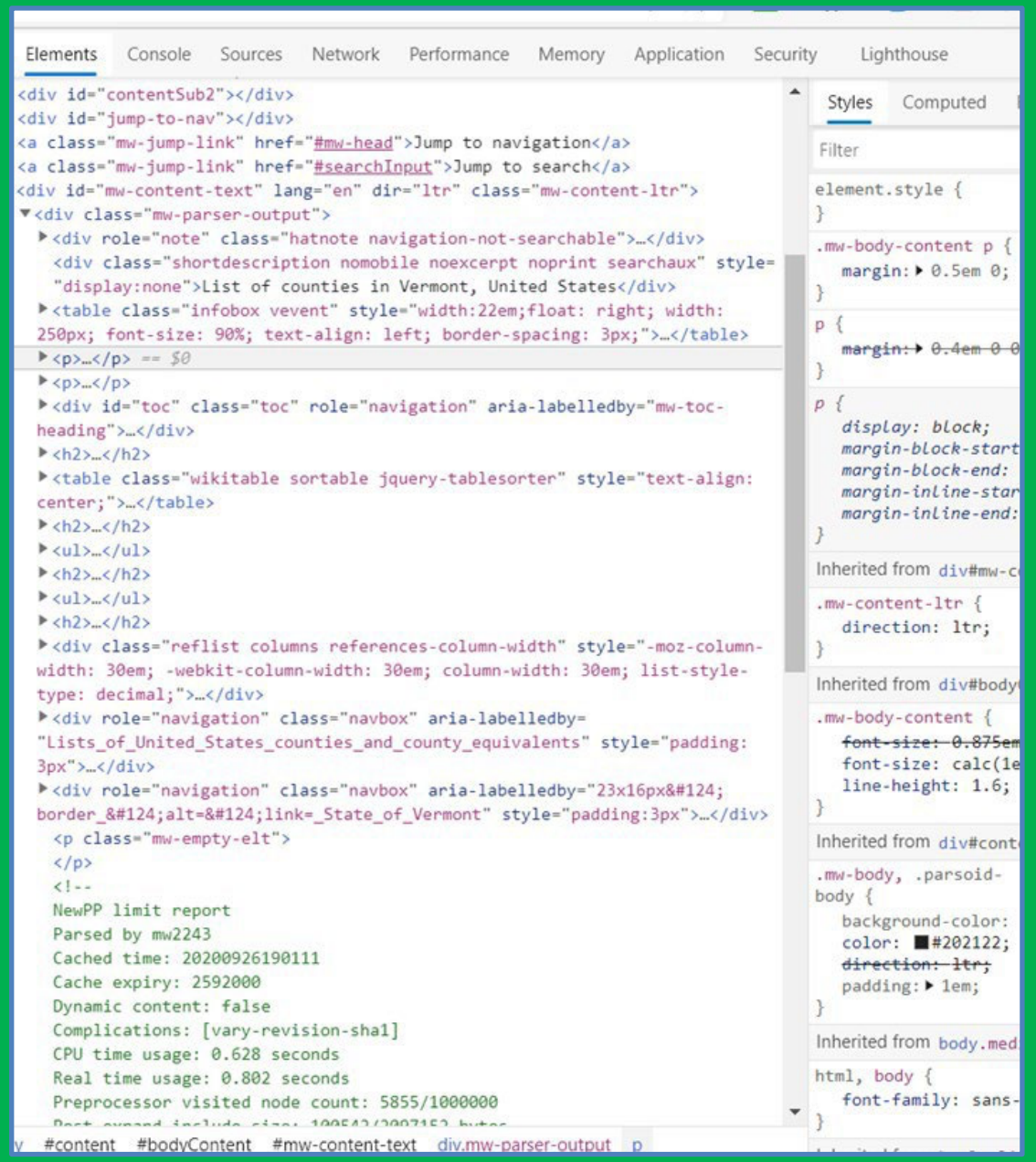
ProPublica reporters examined Amazon's shopping algorithm; scraped data from the company's website to examine listings for 250 bestselling products across a wide range of categories, from electronics to household supplies, over a period of several weeks during summer 2016.

compared pricing and shipping costs for products offered by multiple vendors, including those sold by Amazon, sellers in the "Fulfilled by Amazon" program. In total, we examined 6,973 vendor listings.



Scraping Is A Key Tool for Journalists Today





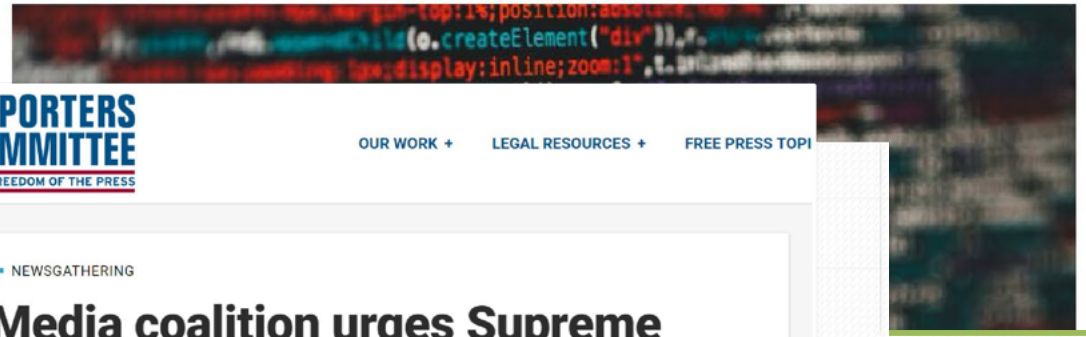
What Is Web Scraping?

REPORTERS COMMITTEE
FOR FREEDOM OF THE PRESS

OUR WORK + LEGAL RESOURCES + FREE PRESS TOPICS + WHO WE ARE + DONATE +

NEWSGATHERING

Scraping public websites likely doesn't violate the Computer Fraud and Abuse Act, court holds



Related Posts

01 Media coalition urges Supreme Court to narrow scope of federal hacking law to protect newsgathering

02 This Week in Technology + Press Freedom: Oct. 4, 2019

03 Four petitions we're watching as the US Supreme Court fills out its docket

Web scraping is legal, US appeals court reaffirms

Zack Whittaker @zackwhittaker / 3:16 PM EDT • April 18, 2022 Comment



REPORTERS COMMITTEE
FOR FREEDOM OF THE PRESS

OUR WORK + LEGAL RESOURCES + FREE PRESS TOPI

NEWSGATHERING

Media coalition urges Supreme Court to narrow scope of federal hacking law to protect newsgathering

No. 19-783

In the Supreme Court of the United States

NATHAN VAN BUREN, PETITIONER

v.

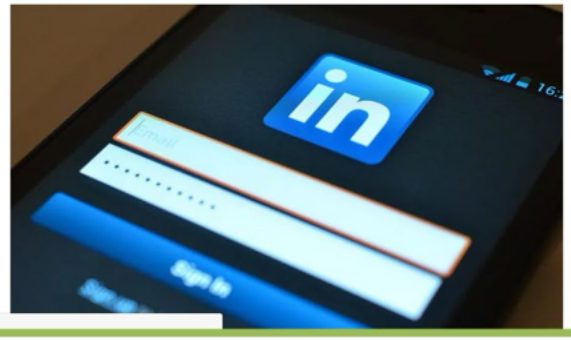
UNITED STATES

POLICY \ LAW

Scraping public data from a website probably isn't hacking, says court

By Adi Robertson | @thedextriarchy | Sep 10, 2019, 4:22pm EDT

SHARE





VERGE DEALS



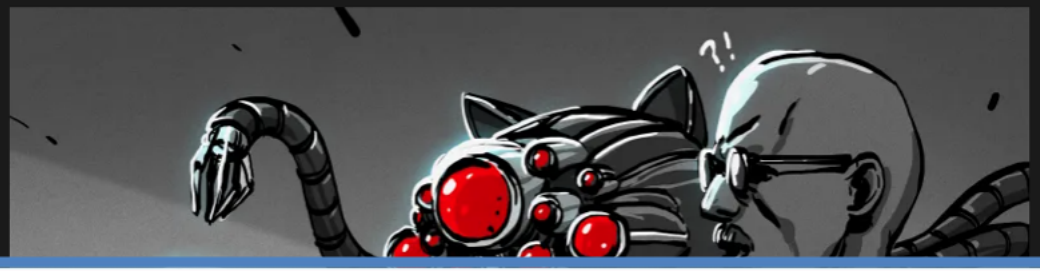
COMPONENT SEARCH ENGINE
PCB FOOTPRINTS
3D MODELS
SCHEMATIC SYMBOLS
FREE DOWNLOAD

TIRED OF WEB SCRAPING? MAKE THE AI DO IT

by: Donald Papp

19 Comments

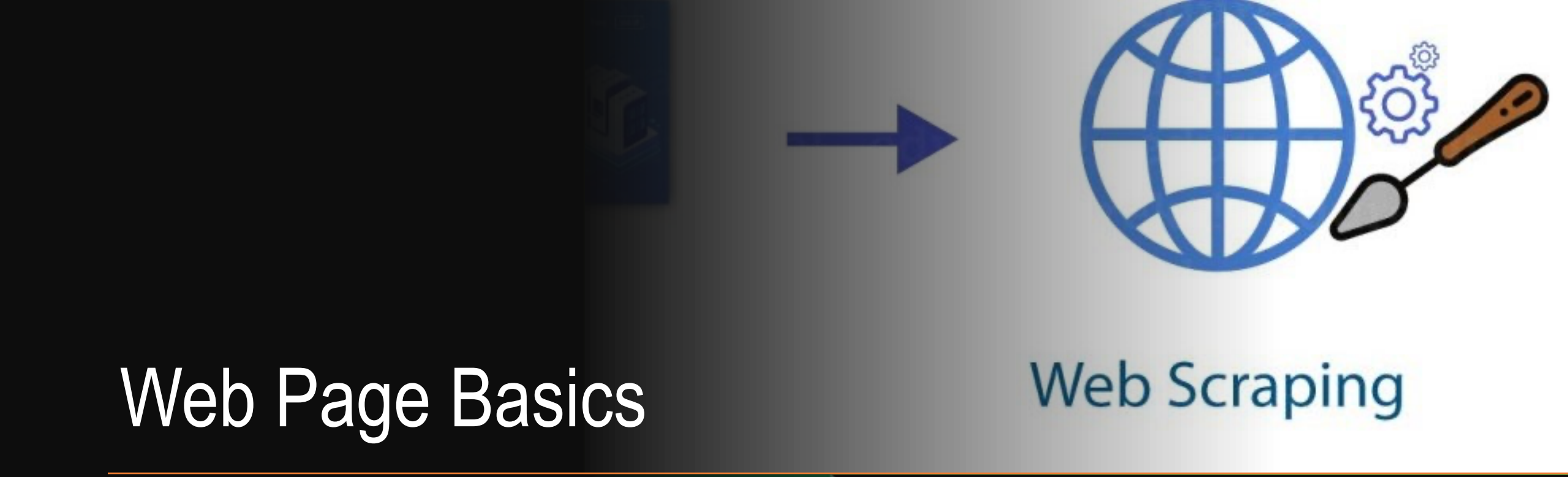
April 9, 2023



COMPONENT SEARCH ENGINE
FREE HIGH QUALITY PCB LIBRARIES FOR YOUR ECAD TOOLS
PCB FOOTPRINTS
3D MODELS
SCHEMATIC SYMBOLS
DOWNLOAD NOW

tindie

The crux of the problem lies in the Googlebot, the crawler that scours and indexes the live web to produce the results you see when you enter search terms. If publishers block Google from using their content for the AI-produced answers you now see littered at the top of many search results, they also lose the privilege of including their web pages in the standard web results.

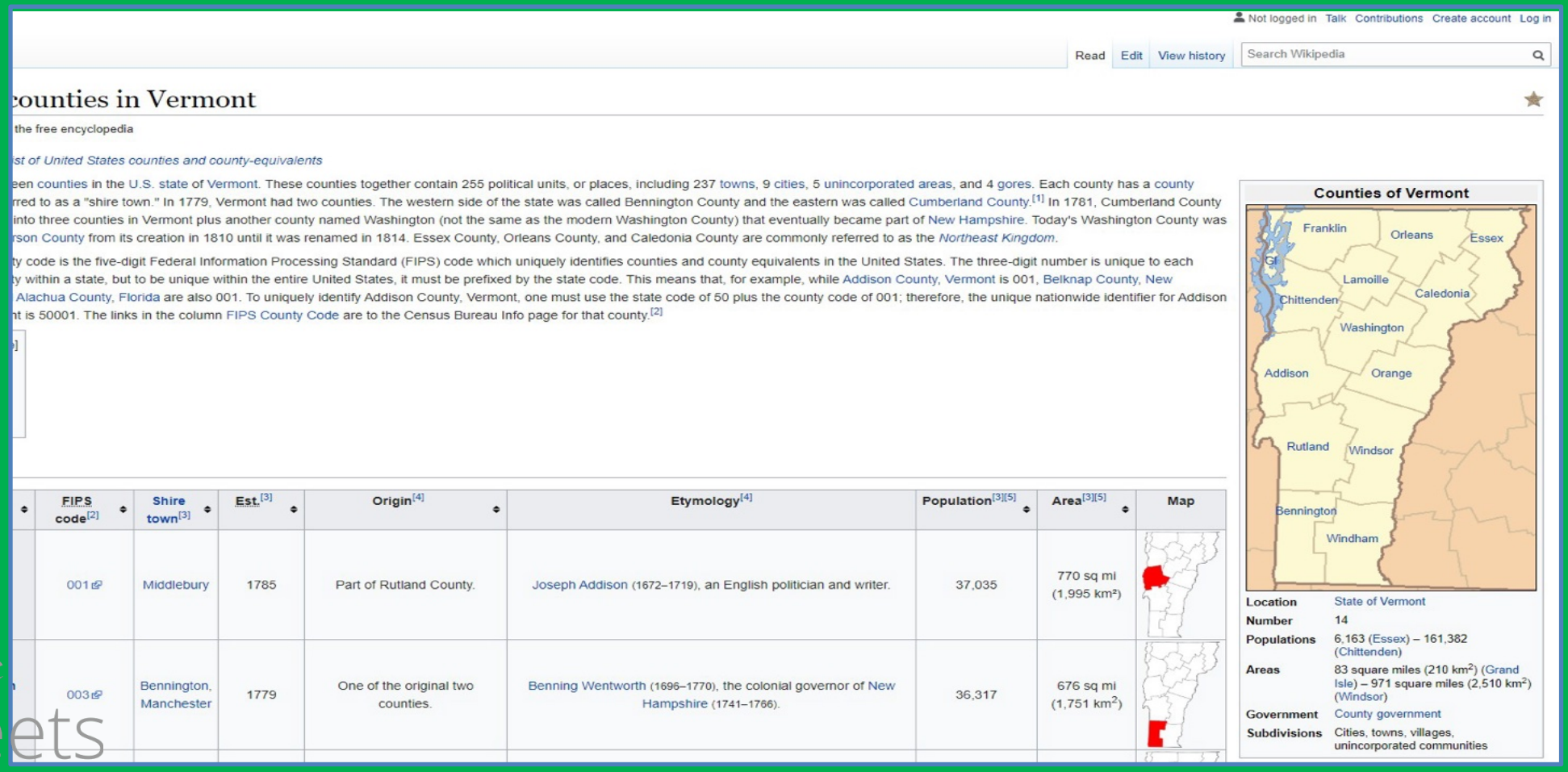


Web Page Basics

Web Scraping 101



What You See in A Web Browser



Not logged in Talk Contributions Create account Log in

Read Edit View history Search Wikipedia

counties in Vermont

the free encyclopedia

ist of United States counties and county-equivalents

een counties in the U.S. state of Vermont. These counties together contain 255 political units, or places, including 237 towns, 9 cities, 5 unincorporated areas, and 4 gores. Each county has a county rred to as a "shire town." In 1779, Vermont had two counties. The western side of the state was called Bennington County and the eastern was called Cumberland County.[1] In 1781, Cumberland County into three counties in Vermont plus another county named Washington (not the same as the modern Washington County) that eventually became part of New Hampshire. Today's Washington County was rson County from its creation in 1810 until it was renamed in 1814. Essex County, Orleans County, and Caledonia County are commonly referred to as the *Northeast Kingdom*.

ty code is the five-digit Federal Information Processing Standard (FIPS) code which uniquely identifies counties and county equivalents in the United States. The three-digit number is unique to each y within a state, but to be unique within the entire United States, it must be prefixed by the state code. This means that, for example, while Addison County, Vermont is 001, Belknap County, New Alachua County, Florida are also 001. To uniquely identify Addison County, Vermont, one must use the state code of 50 plus the county code of 001; therefore, the unique nationwide identifier for Addison nt is 50001. The links in the column FIPS County Code are to the Census Bureau Info page for that county.[2]

	FIPS code[2]	Shire town[3]	Est.[3]	Origin[4]	Etymology[4]	Population[3][5]	Area[3][5]	Map
	001	Middlebury	1785	Part of Rutland County.	Joseph Addison (1672–1719), an English politician and writer.	37,035	770 sq mi (1,995 km²)	
	003	Bennington, Manchester	1779	One of the original two counties.	Benning Wentworth (1695–1770), the colonial governor of New Hampshire (1741–1766).	36,317	676 sq mi (1,751 km²)	

Counties of Vermont

Franklin Orleans Essex Grand Isle Lamoille Caledonia Chittenden Washington Addison Orange Rutland Windsor Bennington Windham

Location	State of Vermont
Number	14
Populations	6,163 (Essex) – 161,382 (Chittenden)
Areas	83 square miles (210 km²) (Grand Isle) – 971 square miles (2,510 km²) (Windsor)
Government	County government
Subdivisions	Cities, towns, villages, unincorporated communities

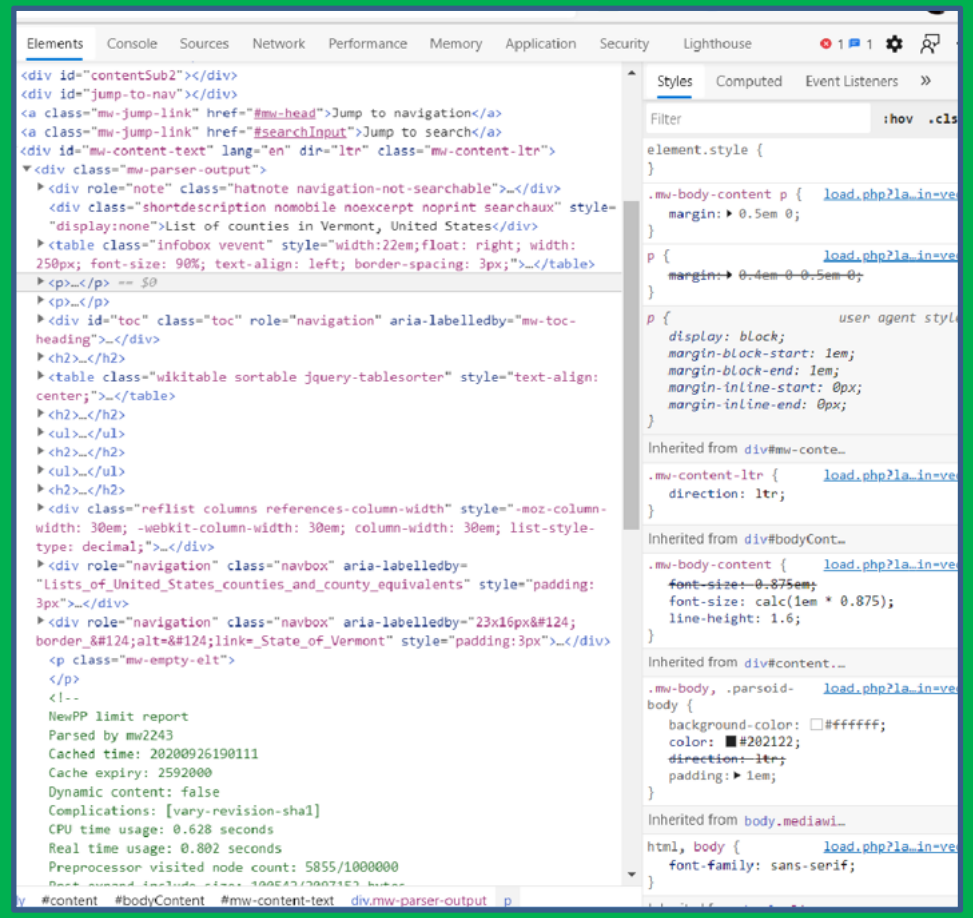
Google Sheets

The Code “Under The Hood”

On your web page, right click **“inspect”**

You’ll see the code that built the page

There is a lot of information to note

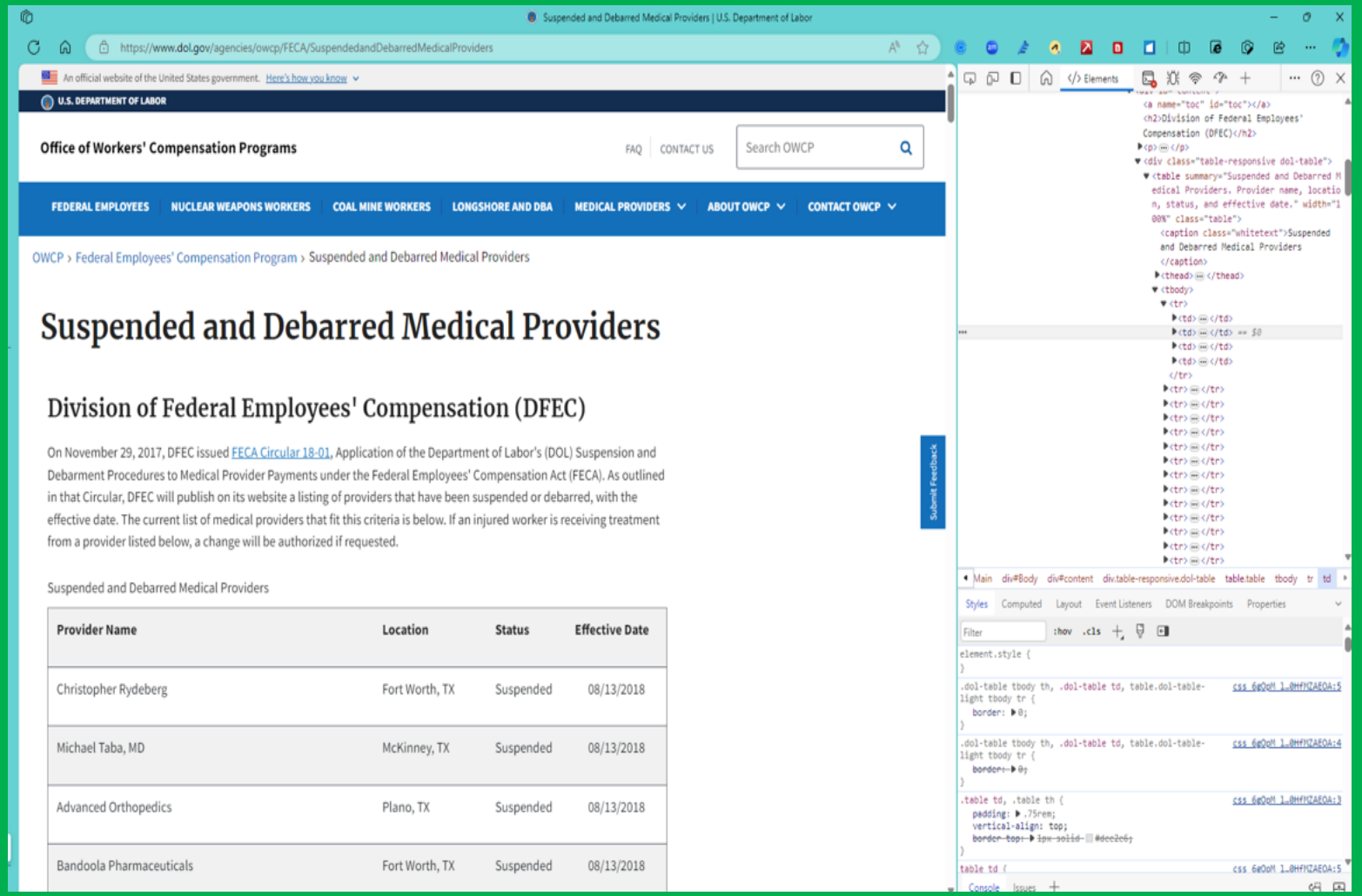


The Code “Under The Hood”

On your web page, right click **“inspect”**

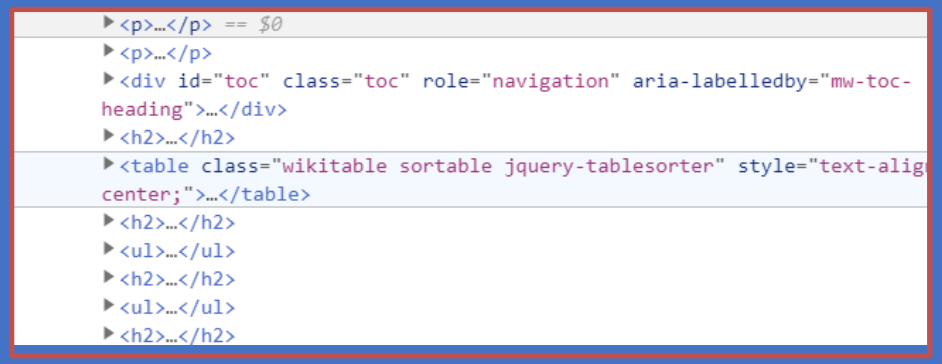
You’ll see the code that built the page

There is a lot of information to note



Web Pages Have Lots of Elements to Scrape

- The code is wrapped in “tags”
- Some of it is easy to find
- Look for “table”
- Today we’ll scrape a “table”



Google Sheets

You might want to keep an HTML dictionary handy as you learn to scrape other elements on a page

You can scrape anything from a web page once you master the basics and build on your formulas



Web Scraping

Scraping Tables With Google Sheets

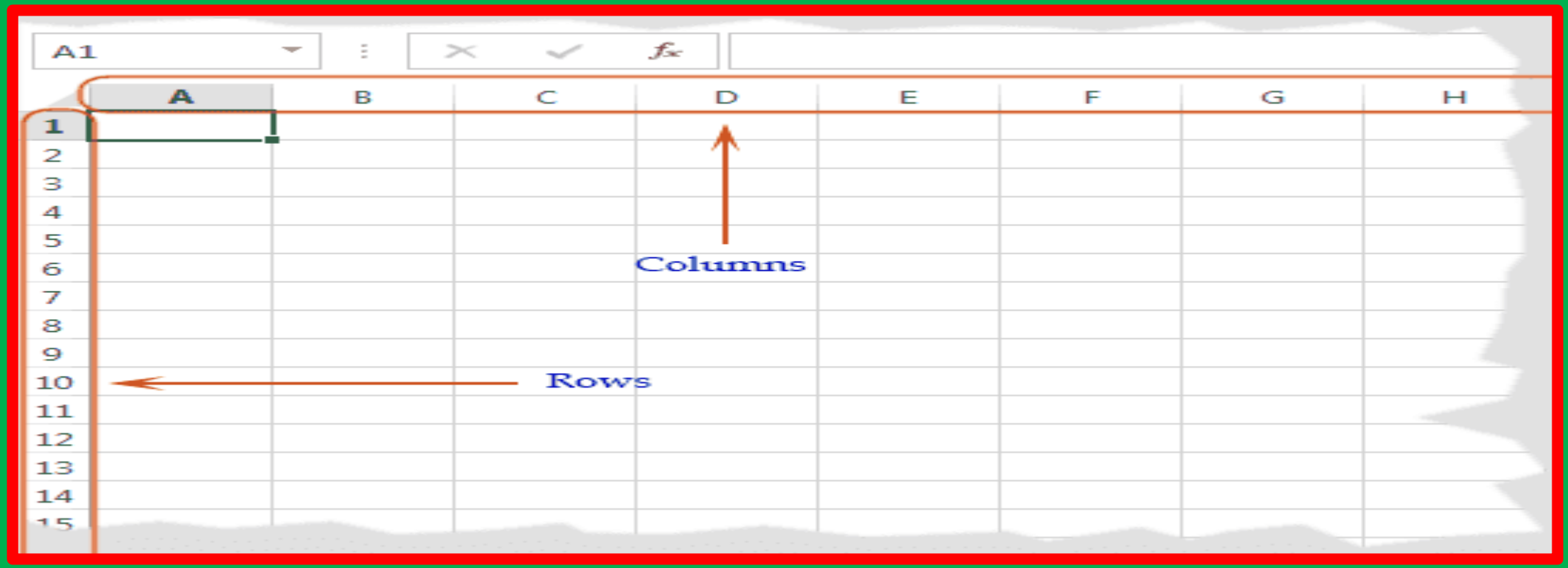
Web Scraping 101

Not All Data is Scrapeable With Google Sheets

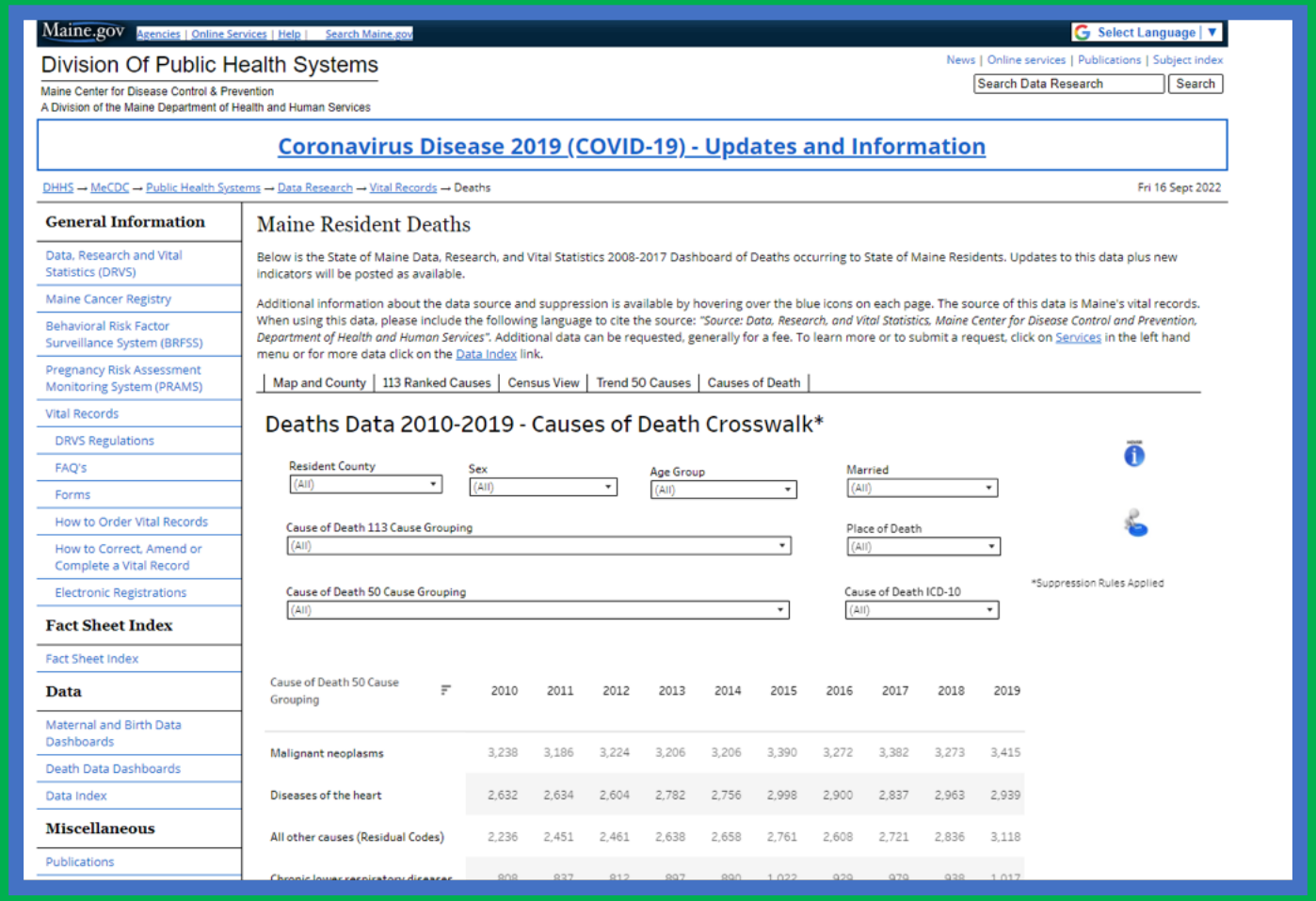
- You might need a more advanced scraper written in a programming language like R to scrape
- We are scraping data “trapped” in a website
- The first step in scraping this way is to locate a dataset can be extracted with one of G-sheets functions like =import HTML
- We want data that is “tabular” – in columns and rows – that lives on a website

Google Sheets

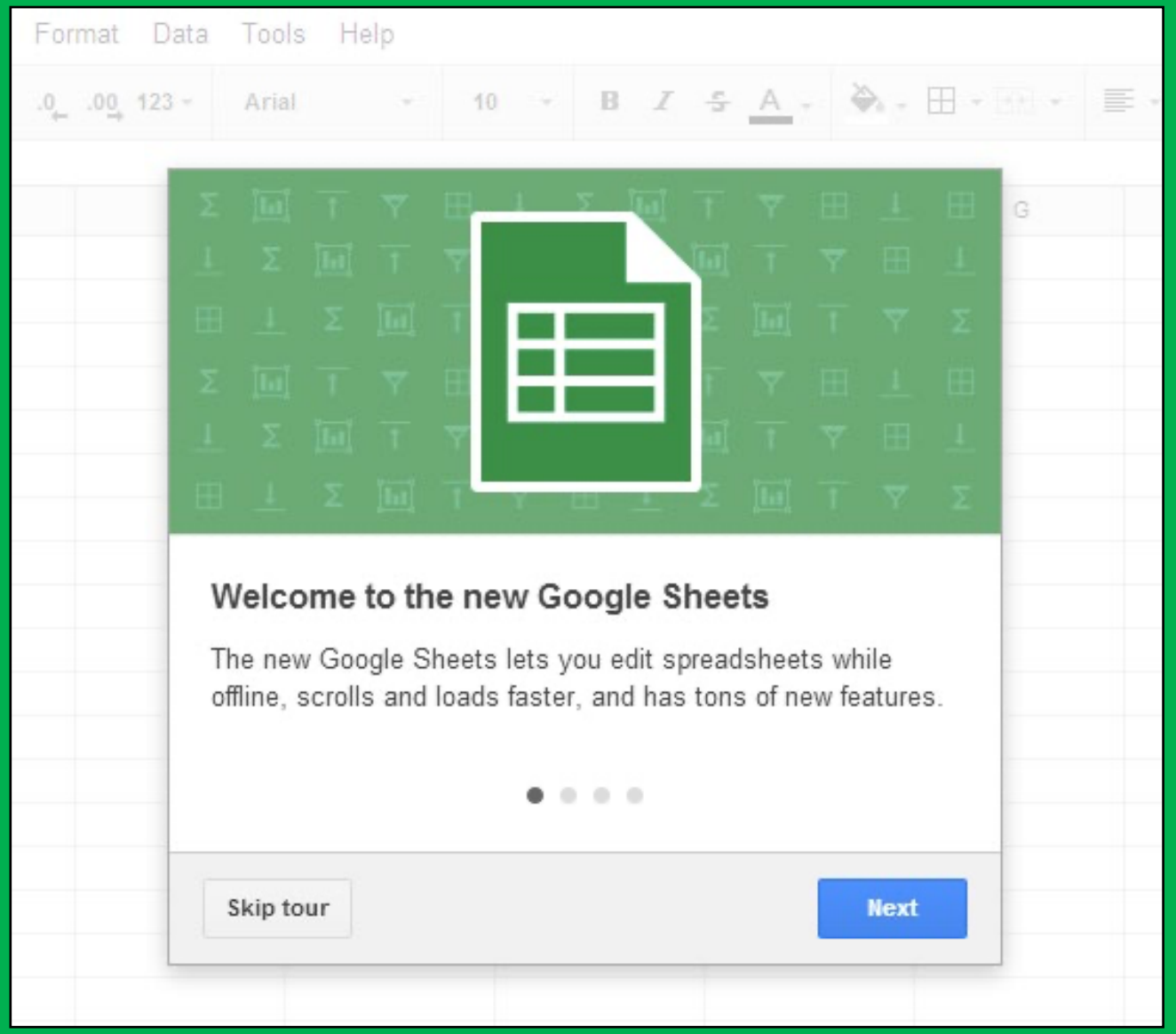
Tabular Data

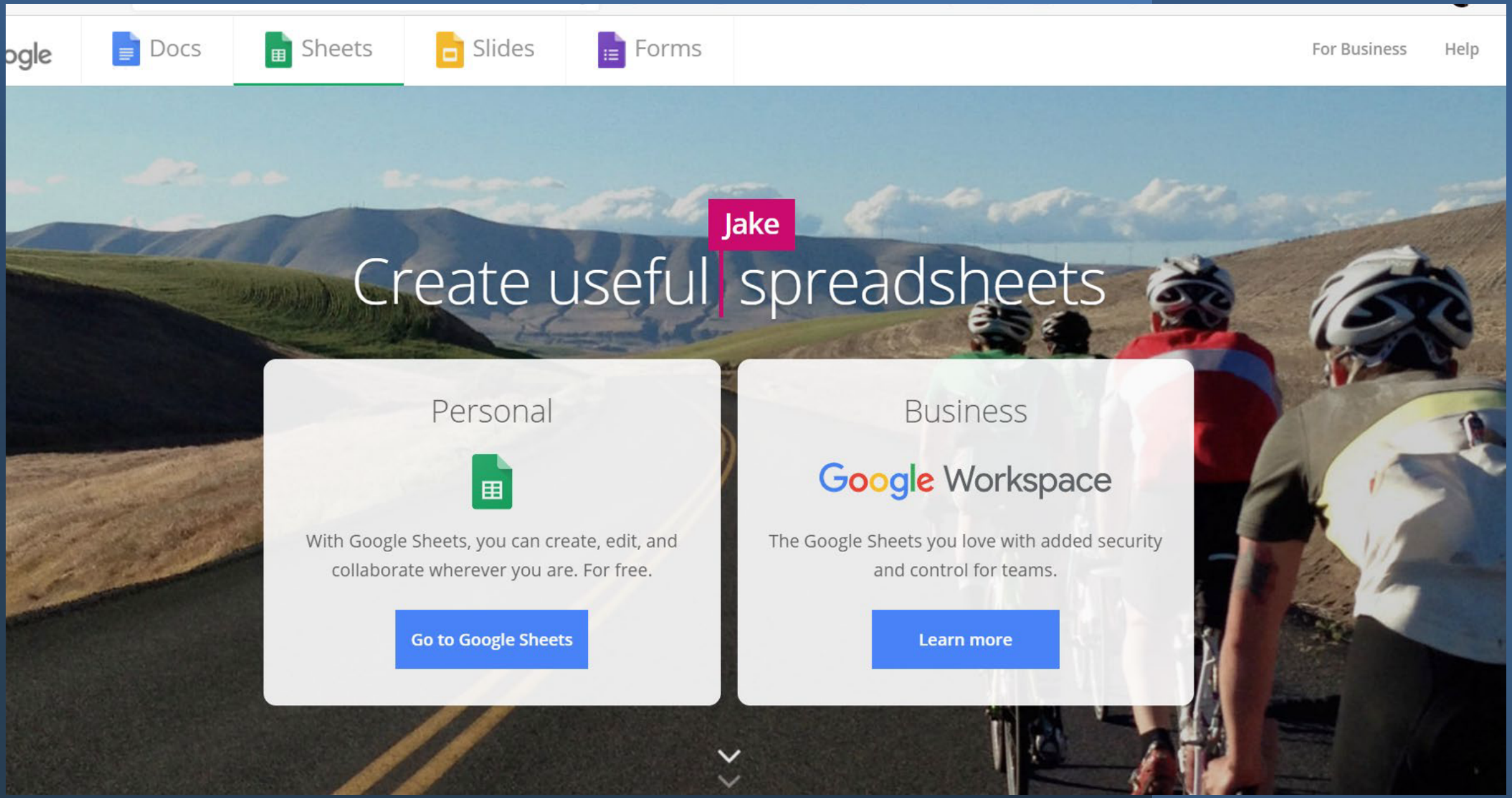


Datasets With Underlying Complex Coding Won't Be Easy To Scrape in G-Sheets



Scraping With Google Sheets




ogle
Docs
Sheets
Slides
Forms
For Business
Help
Jake
Create useful spreadsheets
Personal
With Google Sheets, you can create, edit, and collaborate wherever you are. For free.
Go to Google Sheets
Business
Google Workspace
The Google Sheets you love with added security and control for teams.
Learn more

Open Google Sheets



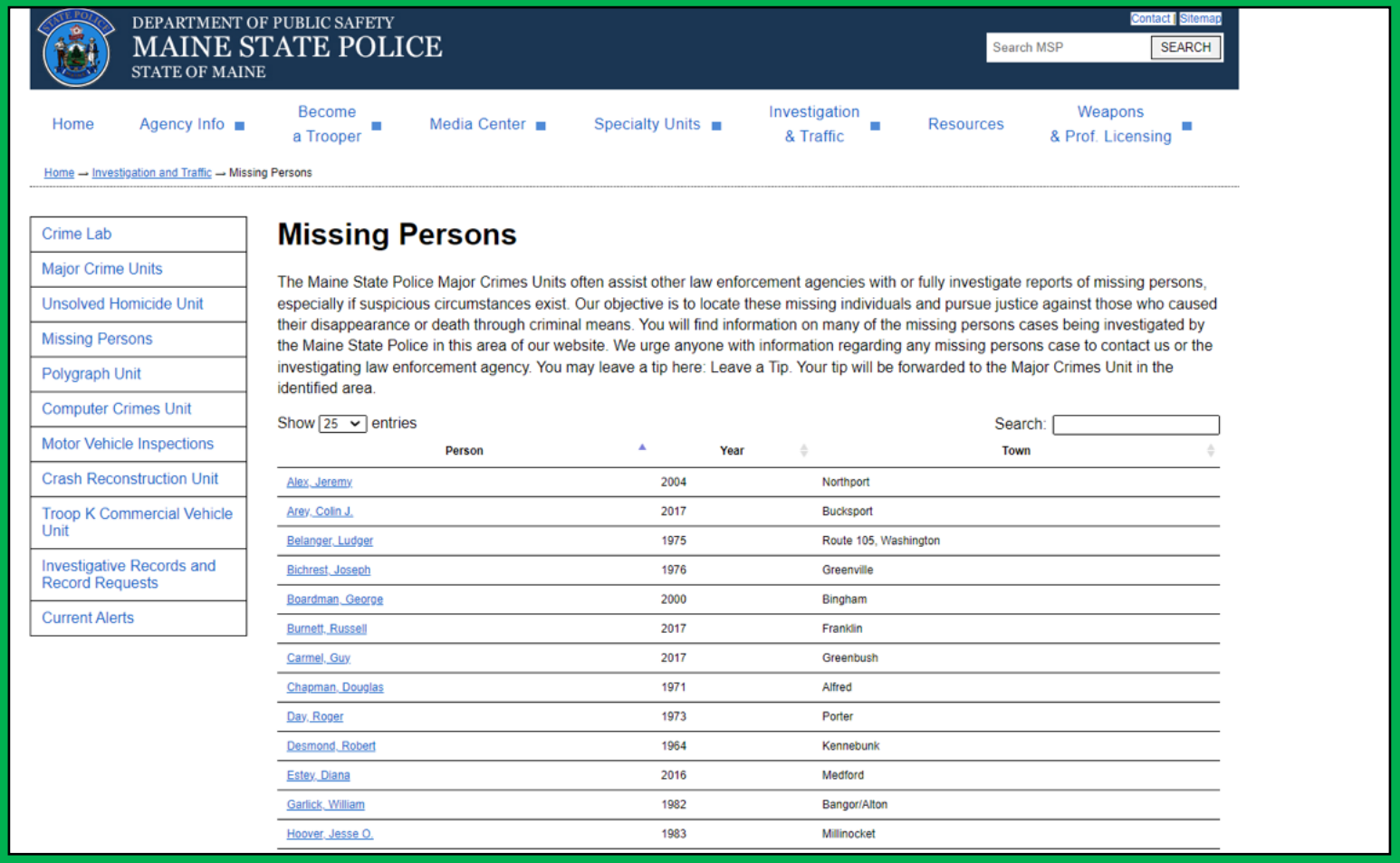
Creating a New Sheet – A Shortcut

- Once you are in your G-Drive, you can create a new sheet
- Click SHFT – S
- A new sheet will appear



Search The Web For “Maine.Gov Missing Persons”

[Missing Persons | Maine State Police]



DEPARTMENT OF PUBLIC SAFETY
MAINE STATE POLICE
STATE OF MAINE

Contact | Sitemap
Search MSP SEARCH

Home | Agency Info | Become a Trooper | Media Center | Specialty Units | Investigation & Traffic | Resources | Weapons & Prof. Licensing

Home → Investigation and Traffic → Missing Persons

- Crime Lab
- Major Crime Units
- Unsolved Homicide Unit
- Missing Persons
- Polygraph Unit
- Computer Crimes Unit
- Motor Vehicle Inspections
- Crash Reconstruction Unit
- Troop K Commercial Vehicle Unit
- Investigative Records and Record Requests
- Current Alerts

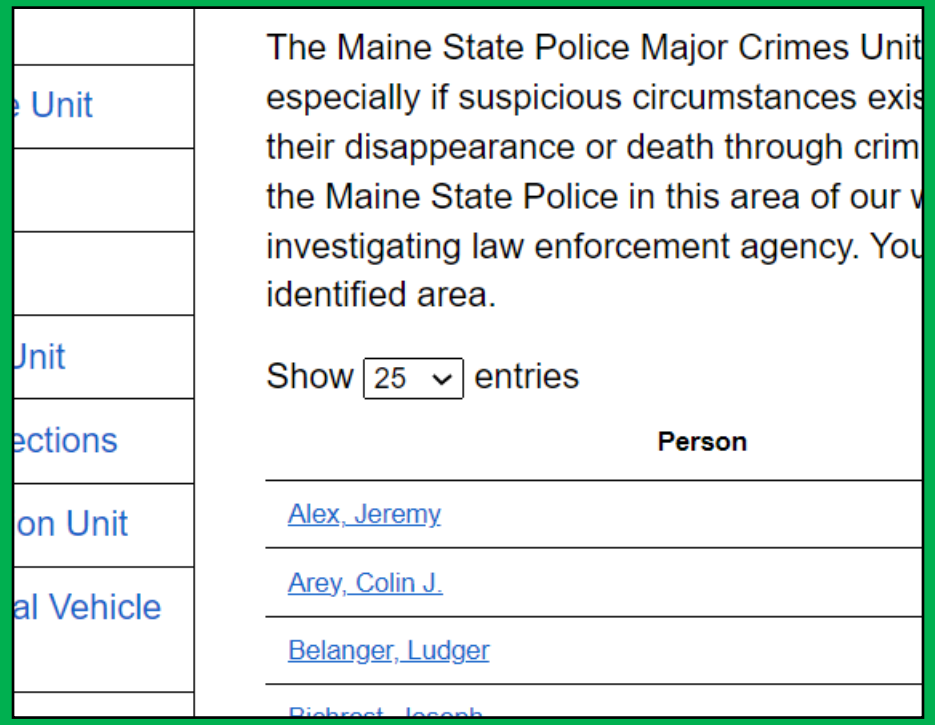
Missing Persons

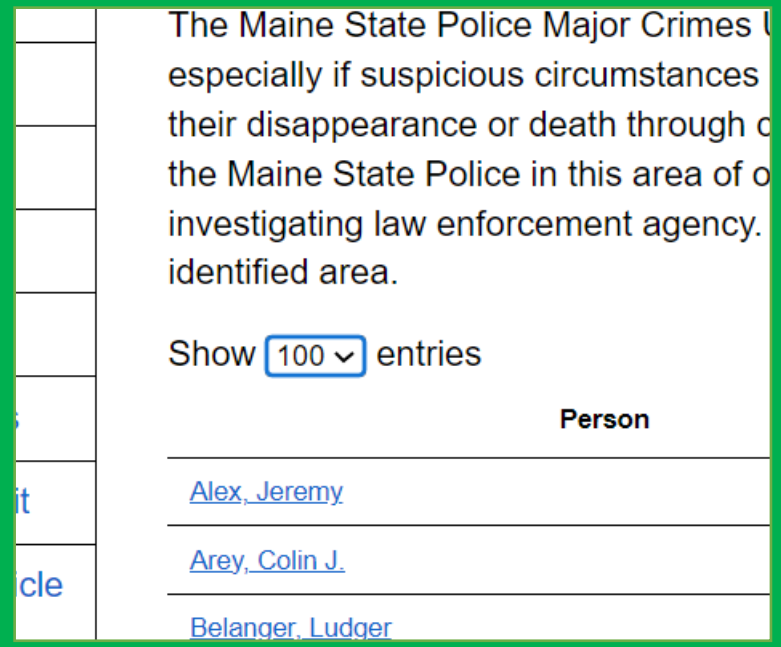
The Maine State Police Major Crimes Units often assist other law enforcement agencies with or fully investigate reports of missing persons, especially if suspicious circumstances exist. Our objective is to locate these missing individuals and pursue justice against those who caused their disappearance or death through criminal means. You will find information on many of the missing persons cases being investigated by the Maine State Police in this area of our website. We urge anyone with information regarding any missing persons case to contact us or the investigating law enforcement agency. You may leave a tip here: Leave a Tip. Your tip will be forwarded to the Major Crimes Unit in the identified area.

Show 25 entries Search:

Person	Year	Town
Alex, Jeremy	2004	Northport
Arey, Colin J.	2017	Bucksport
Belanger, Ludger	1975	Route 105, Washington
Bichrest, Joseph	1976	Greenville
Boardman, George	2000	Bingham
Burnett, Russell	2017	Franklin
Carmel, Guy	2017	Greenbush
Chapman, Douglas	1971	Alfred
Day, Roger	1973	Porter
Desmond, Robert	1964	Kennebunk
Estey, Diana	2016	Medford
Garlick, William	1982	Bangor/Alton
Hoover, Jesse O.	1983	Millinocket

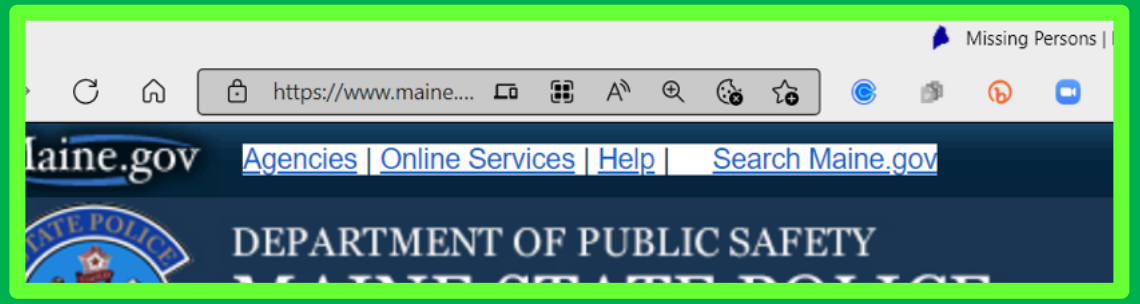
Review the Page
Change The Number 25 to 100 At The Top So You Can See All The Records





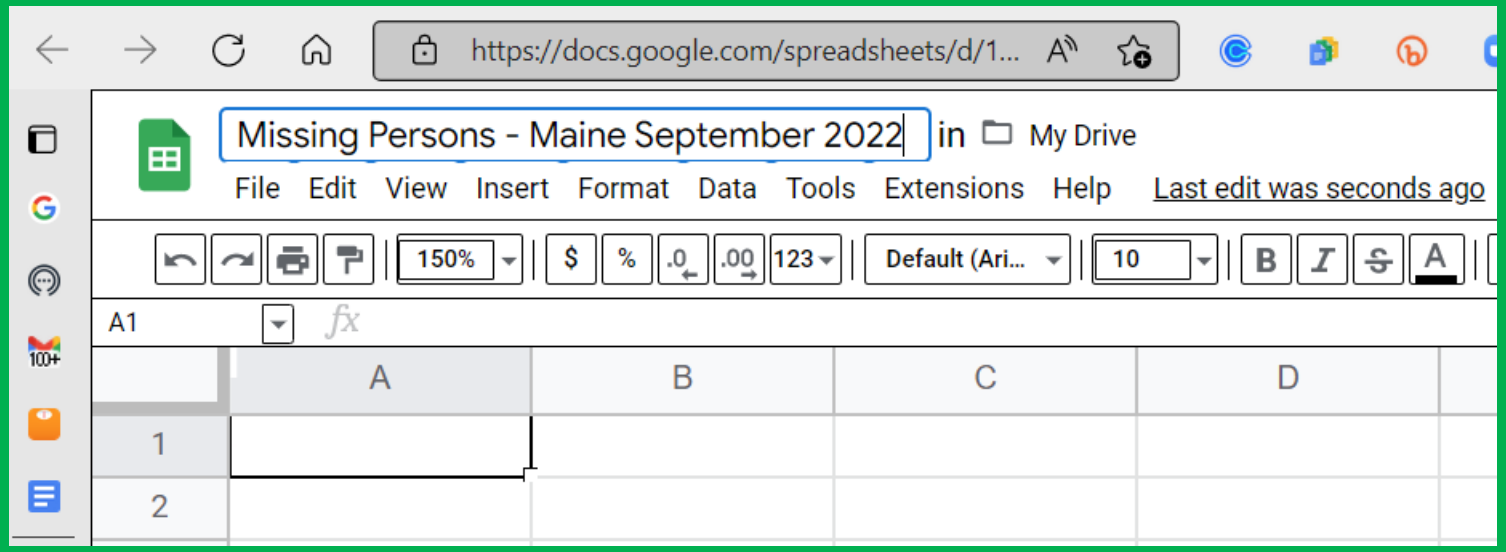
Right Click and Copy The URL

- Hover your cursor over the address bar
- Right click
- Copy or CTRL C (PC) or CMND C (Mac)



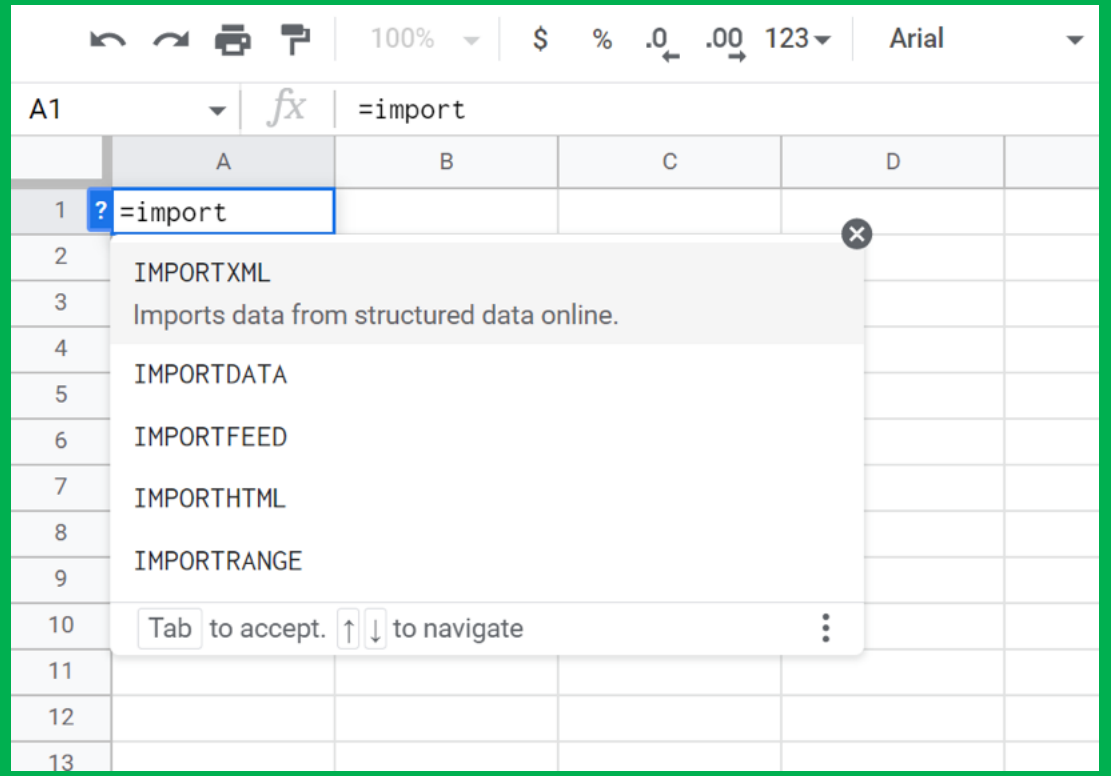
Missing Persons | Maine State Police

Give Your G-Sheet A Name



In Cell A1, type =importhtml

- Whenever you type a function in G-sheets, you need an equal sign in front of it
- The computer will start to recognize the function in G-sheets you want
- It is not case sensitive
- We are asking G-sheets to import an html table



IMPORTHTML

Imports data from a table or list within an HTML page.

Sample Usage

```
IMPORTHTML("http://en.wikipedia.org/wiki/Demographics_of_India","table",4)
IMPORTHTML(A2,B2,C2)
```

Syntax

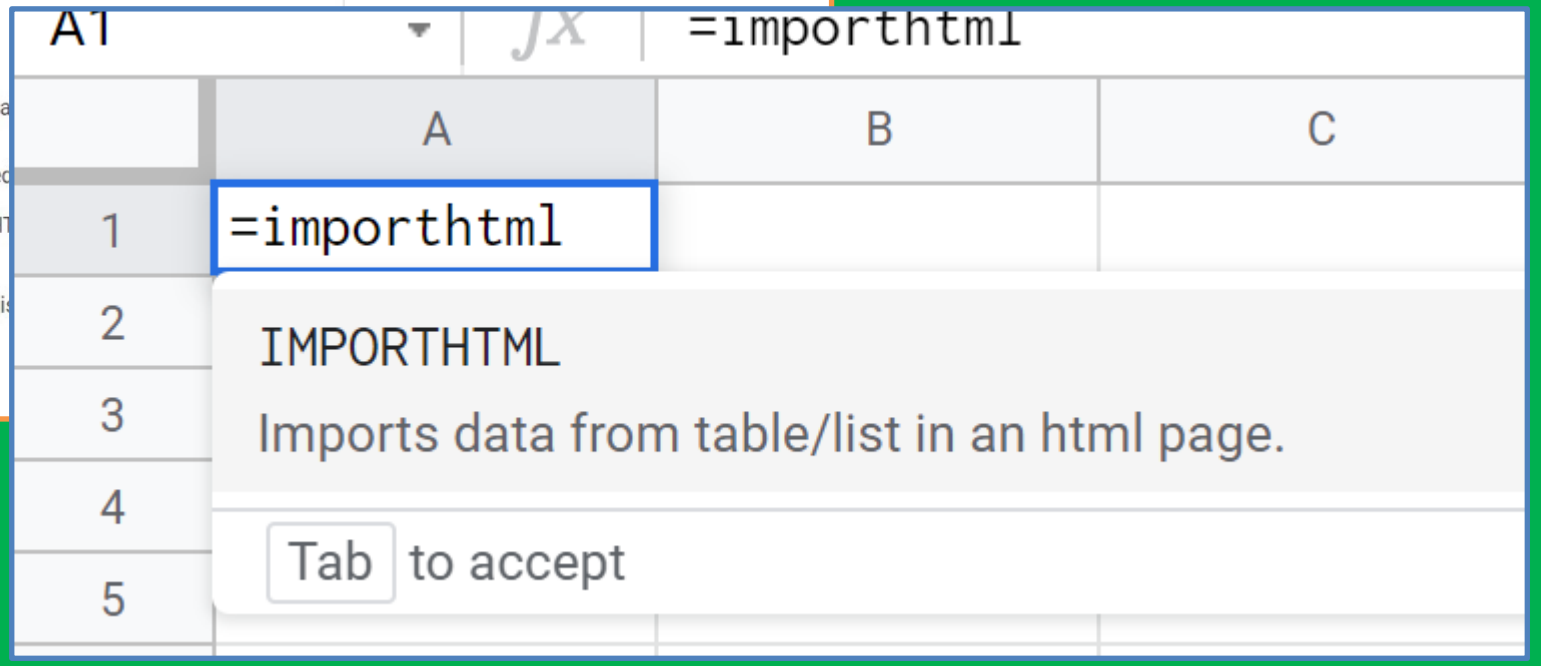
```
IMPORTHTML(url, query, index)
```

- `url` - The URL of the page to examine, including protocol (e.g. `http://`).
  - The value for `url` must either be enclosed in quotation marks or be a reference to a appropriate text.
- `query` - Either "list" or "table" depending on what type of structure contains the desired
- `index` - The index, starting at `1`, which identifies which table or list as defined in the HT returned.
  - The indices for lists and tables are maintained separately, so there may be both a lis `1` if both types of elements exist on the HTML page.

See Also

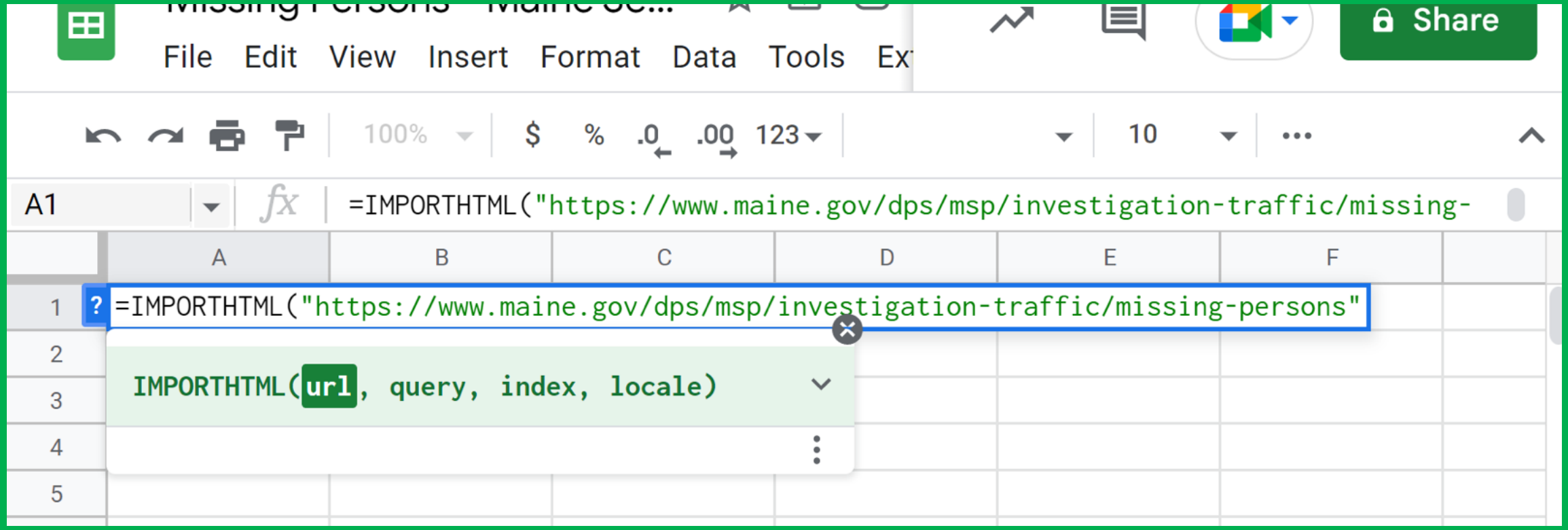
Web

- Google Sheets function list
- ENCODEURL function
- HYPERLINK
- IMPORTDATA
- IMPORTFEED
- IMPORTHTML
- IMPORTRANGE



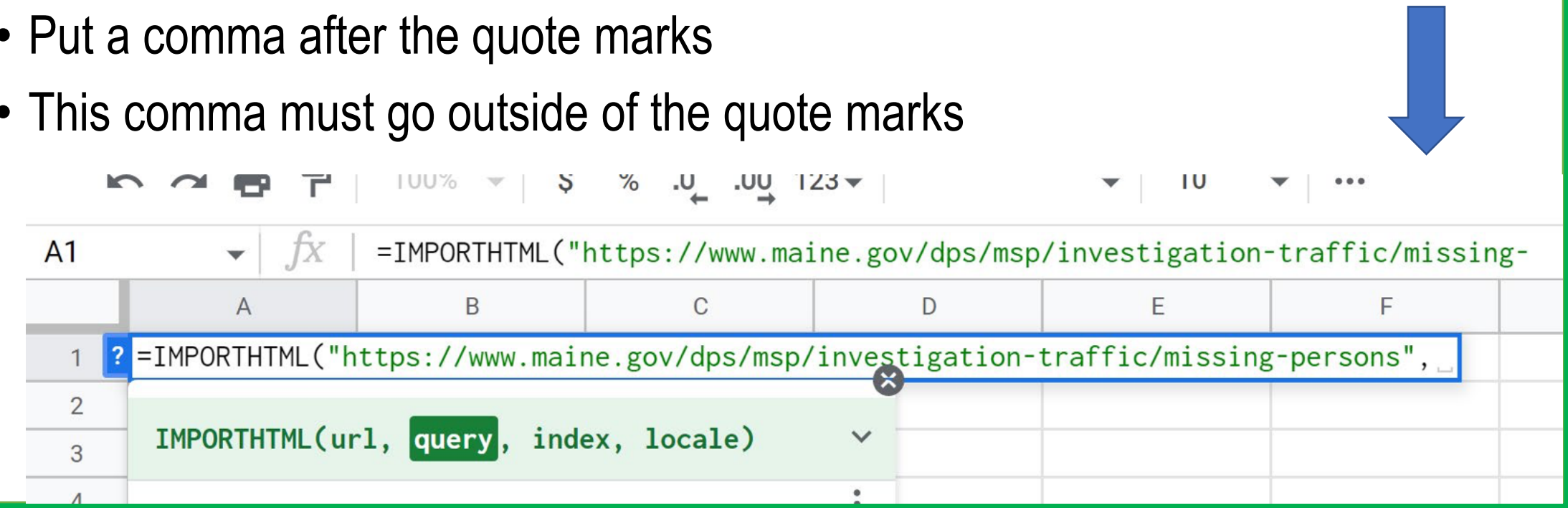
Google Sheets

In The Address Bar, Open Parentheses, Open Quotes, Paste Your URL



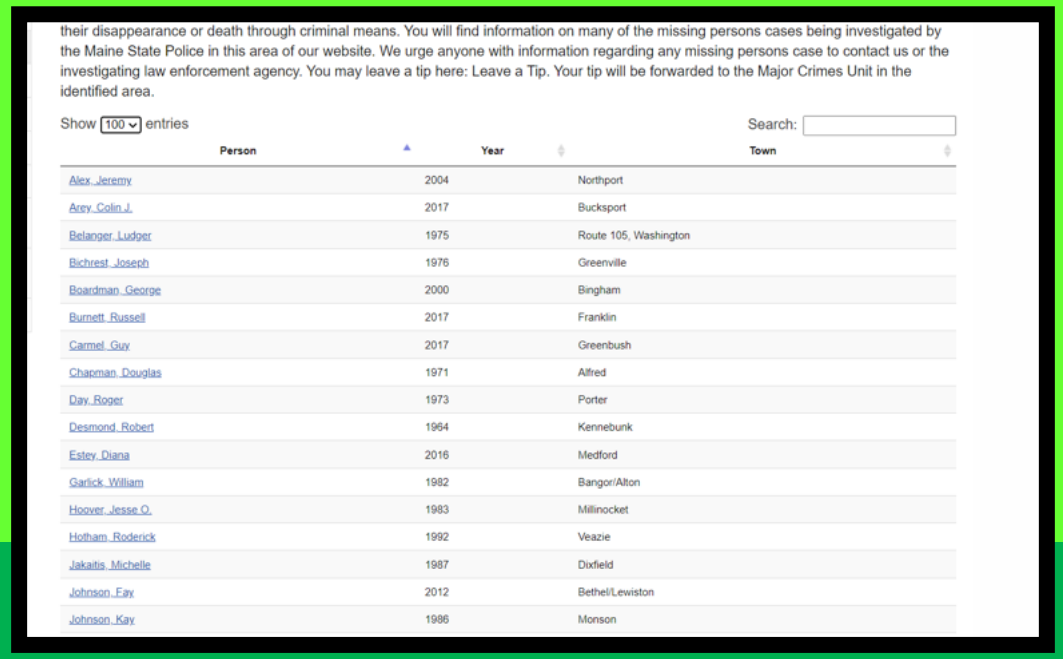
Close the URL with quote marks
Put A Comma After The Pasted URL

- Enclose the URL with quote marks
- Put a comma after the quote marks
- This comma must go outside of the quote marks



Identify The Type of Data G-Sheets Should Scrape

- Is our data a table?
- Yes – it is tabular data with columns and rows
- We need to tell G-sheets to scrape a “table

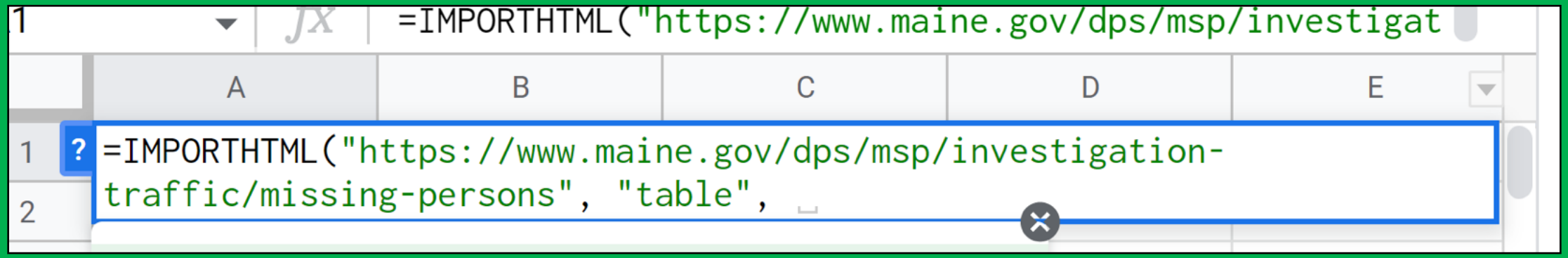


their disappearance or death through criminal means. You will find information on many of the missing persons cases being investigated by the Maine State Police in this area of our website. We urge anyone with information regarding any missing persons case to contact us or the investigating law enforcement agency. You may leave a tip here: Leave a Tip. Your tip will be forwarded to the Major Crimes Unit in the identified area.

Show 100 entries Search:

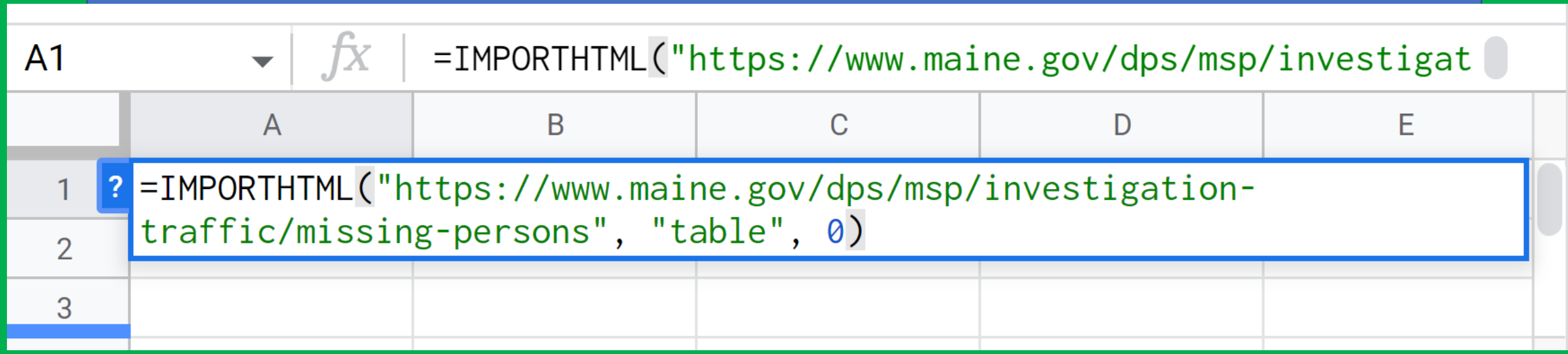
Person	Year	Town
Alex, Jeremy	2004	Northport
Arey, Colin J.	2017	Bucksport
Belanger, Ludger	1975	Route 105, Washington
Bichrest, Joseph	1976	Greenville
Boardman, George	2000	Bingham
Burnett, Russell	2017	Franklin
Carmel, Guy	2017	Greenbush
Chapman, Douglas	1971	Alfred
Day, Roger	1973	Porter
Desmond, Robert	1964	Kennebunk
Estey, Diana	2016	Medford
Garlick, William	1982	Bangor/Alton
Hoover, Jesse O.	1983	Millinocket
Hotham, Roderick	1992	Veazie
Jakaitis, Michelle	1987	Dixfield
Johnson, Fay	2012	Bethel/Lewiston
Johnson, Kay	1986	Monson

Type “Table” In Quote Marks, Add A Comma



Type a 0 – No Quotes – After The Comma

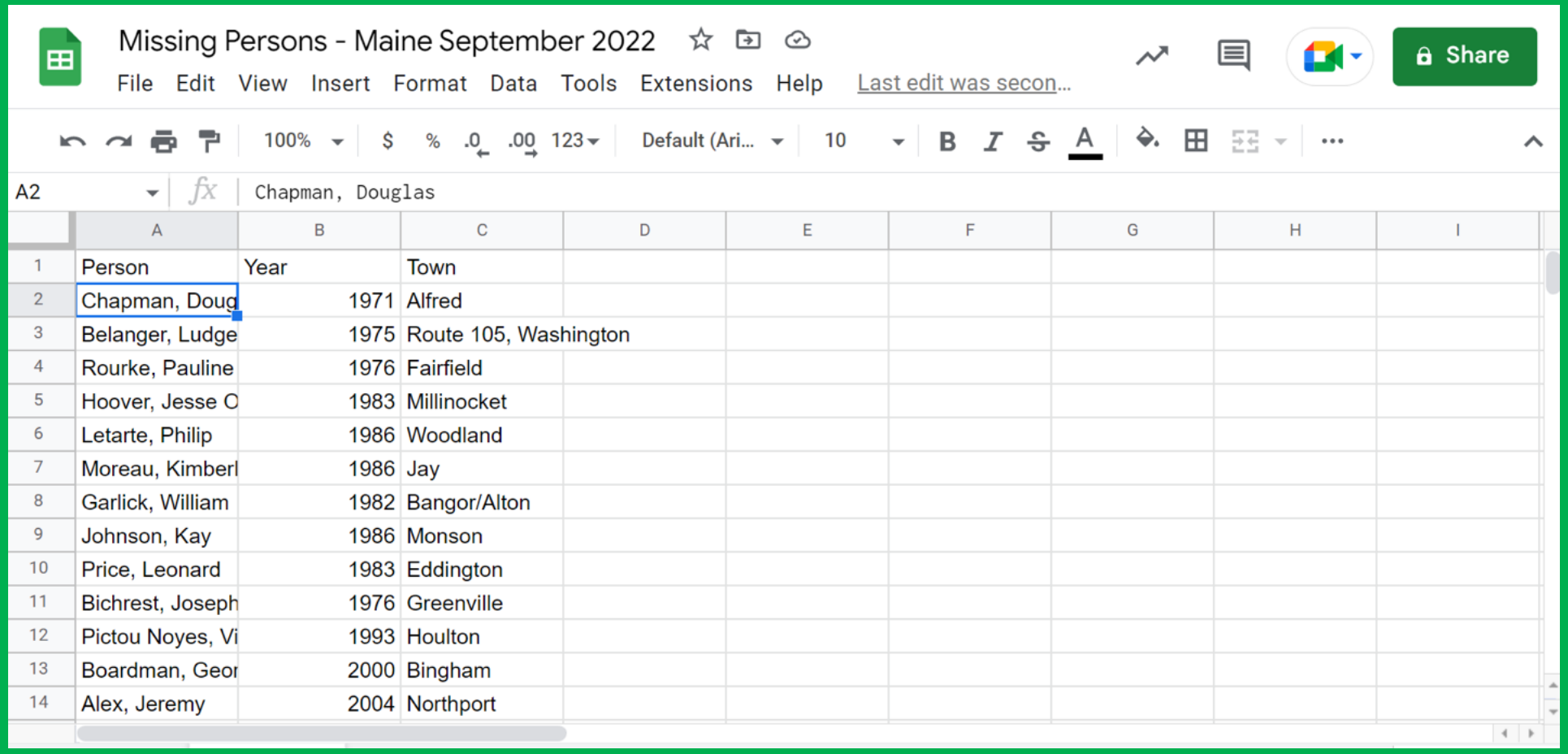
- Computers start counting elements on a page from 0
- You do not need quote marks around a number in G-sheets, only text
- Add close parentheses



Hit ENTER And Your Scrape Will Start

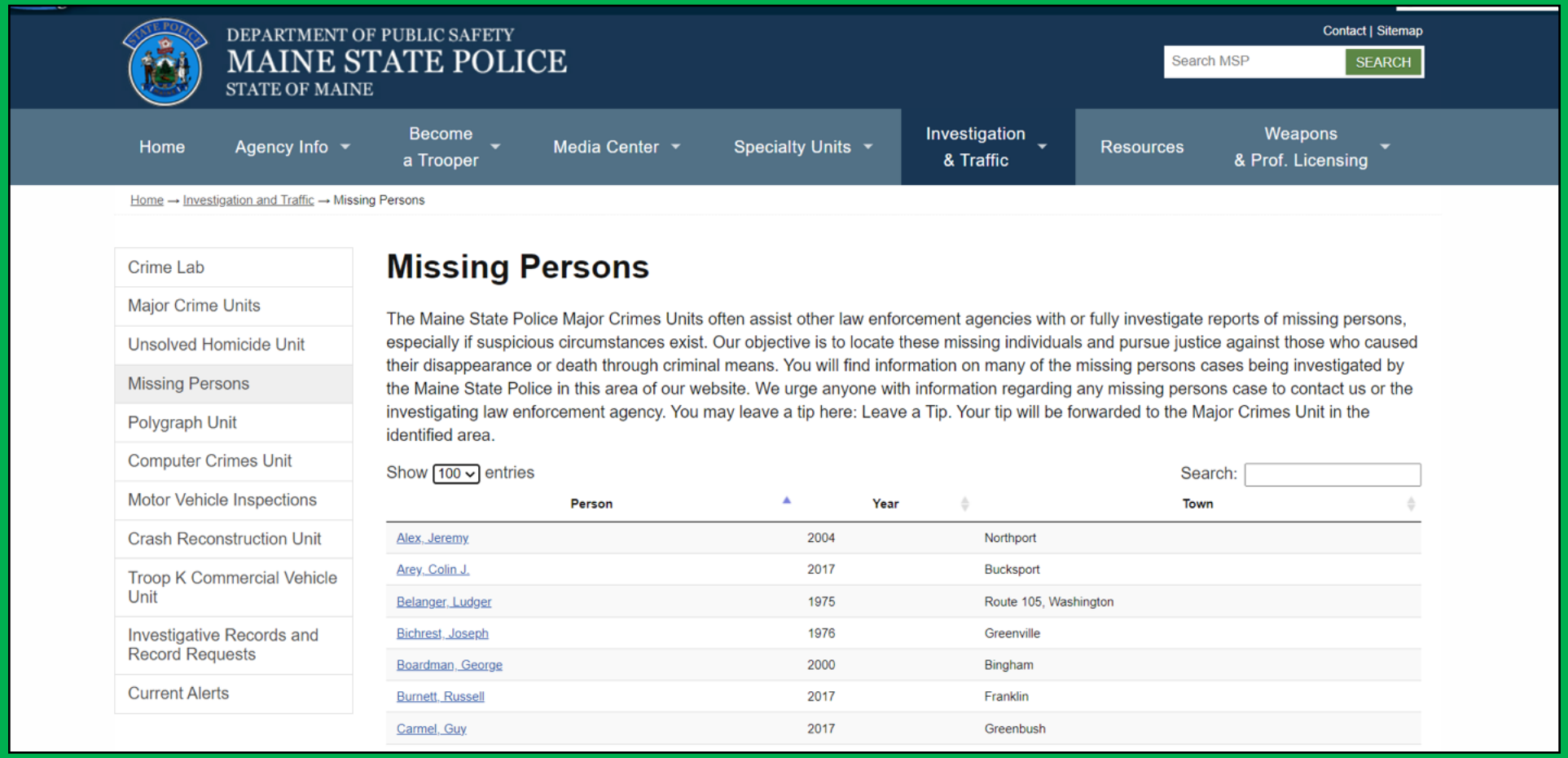


What Stories Could You Do With Your Scrape?



	A	B	C	D	E	F	G	H	I
1	Person	Year	Town						
2	Chapman, Doug	1971	Alfred						
3	Belanger, Ludge	1975	Route 105, Washington						
4	Rourke, Pauline	1976	Fairfield						
5	Hoover, Jesse C	1983	Millinocket						
6	Letarte, Philip	1986	Woodland						
7	Moreau, Kimberl	1986	Jay						
8	Garlick, William	1982	Bangor/Alton						
9	Johnson, Kay	1986	Monson						
10	Price, Leonard	1983	Eddington						
11	Bichrest, Joseph	1976	Greenville						
12	Pictou Noyes, Vi	1993	Houlton						
13	Boardman, Geor	2000	Bingham						
14	Alex, Jeremy	2004	Northport						

The Value of Scraping: Each Time The Site Is Updated, Your G-Sheet Scrape Is Updated



DEPARTMENT OF PUBLIC SAFETY
MAINE STATE POLICE
STATE OF MAINE

Contact | Sitemap

Search MSP SEARCH

Home | Agency Info | Become a Trooper | Media Center | Specialty Units | Investigation & Traffic | Resources | Weapons & Prof. Licensing

Home → Investigation and Traffic → Missing Persons

- Crime Lab
- Major Crime Units
- Unsolved Homicide Unit
- Missing Persons
- Polygraph Unit
- Computer Crimes Unit
- Motor Vehicle Inspections
- Crash Reconstruction Unit
- Troop K Commercial Vehicle Unit
- Investigative Records and Record Requests
- Current Alerts

Missing Persons

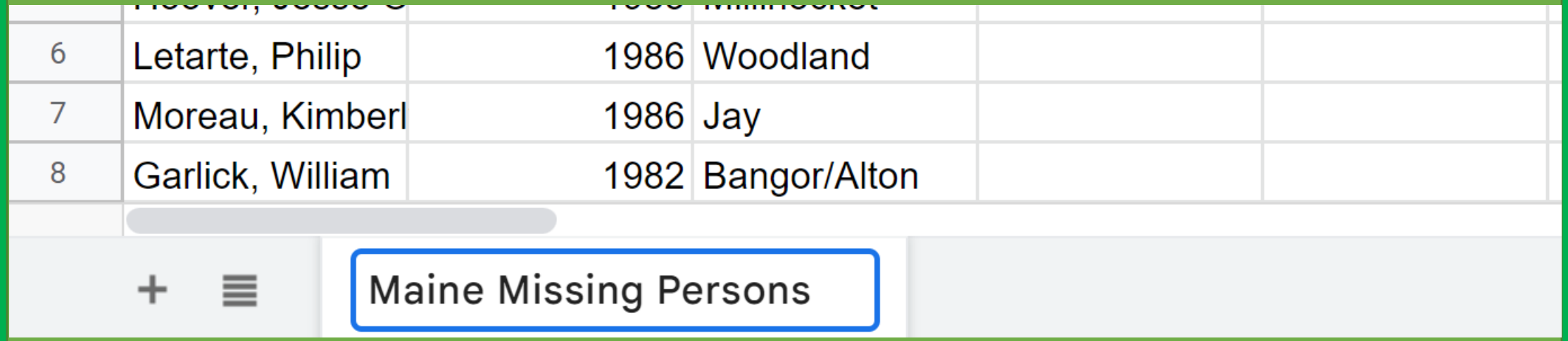
The Maine State Police Major Crimes Units often assist other law enforcement agencies with or fully investigate reports of missing persons, especially if suspicious circumstances exist. Our objective is to locate these missing individuals and pursue justice against those who caused their disappearance or death through criminal means. You will find information on many of the missing persons cases being investigated by the Maine State Police in this area of our website. We urge anyone with information regarding any missing persons case to contact us or the investigating law enforcement agency. You may leave a tip here: Leave a Tip. Your tip will be forwarded to the Major Crimes Unit in the identified area.

Show 100 entries Search:

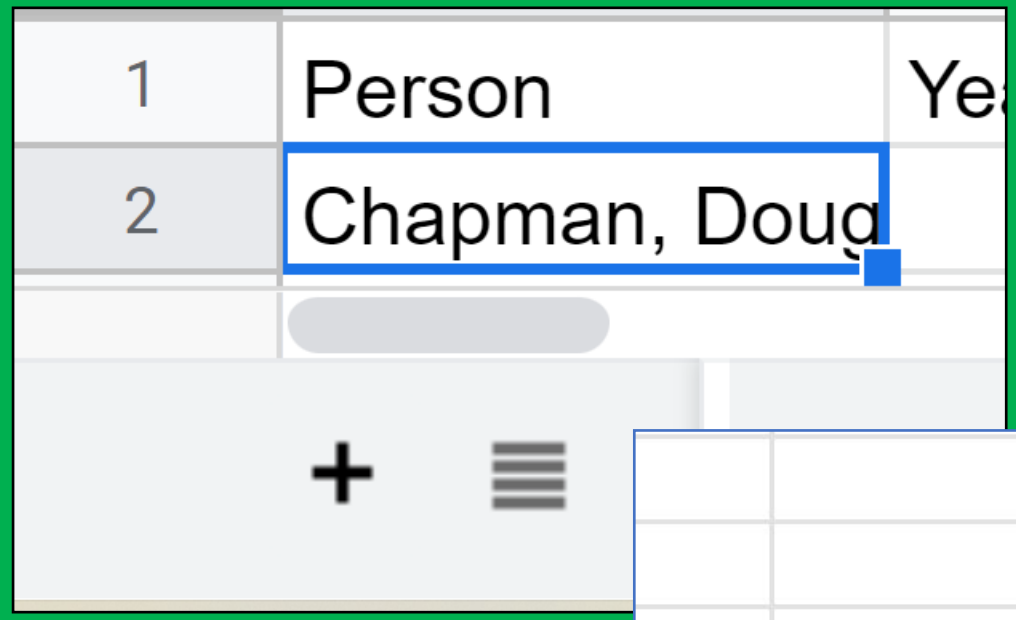
Person	Year	Town
Alex, Jeremy	2004	Northport
Arey, Colin J.	2017	Bucksport
Belanger, Ludger	1975	Route 105, Washington
Bichrest, Joseph	1976	Greenville
Boardman, George	2000	Bingham
Burnett, Russell	2017	Franklin
Carmel, Guy	2017	Greenbush

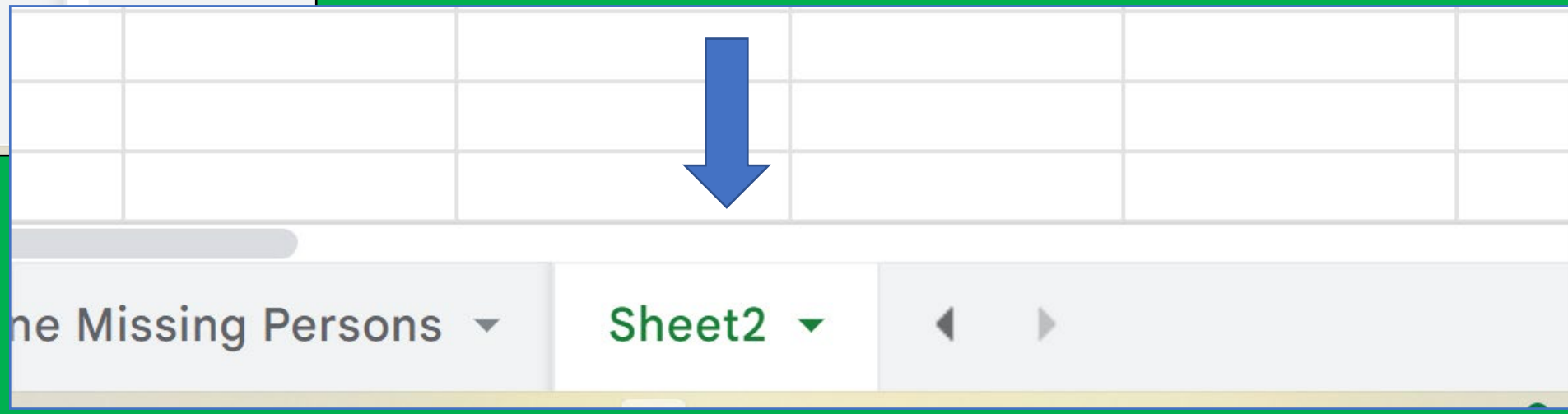
Right Click On Your Sheet At The Bottom and Name It

Good organization is key!

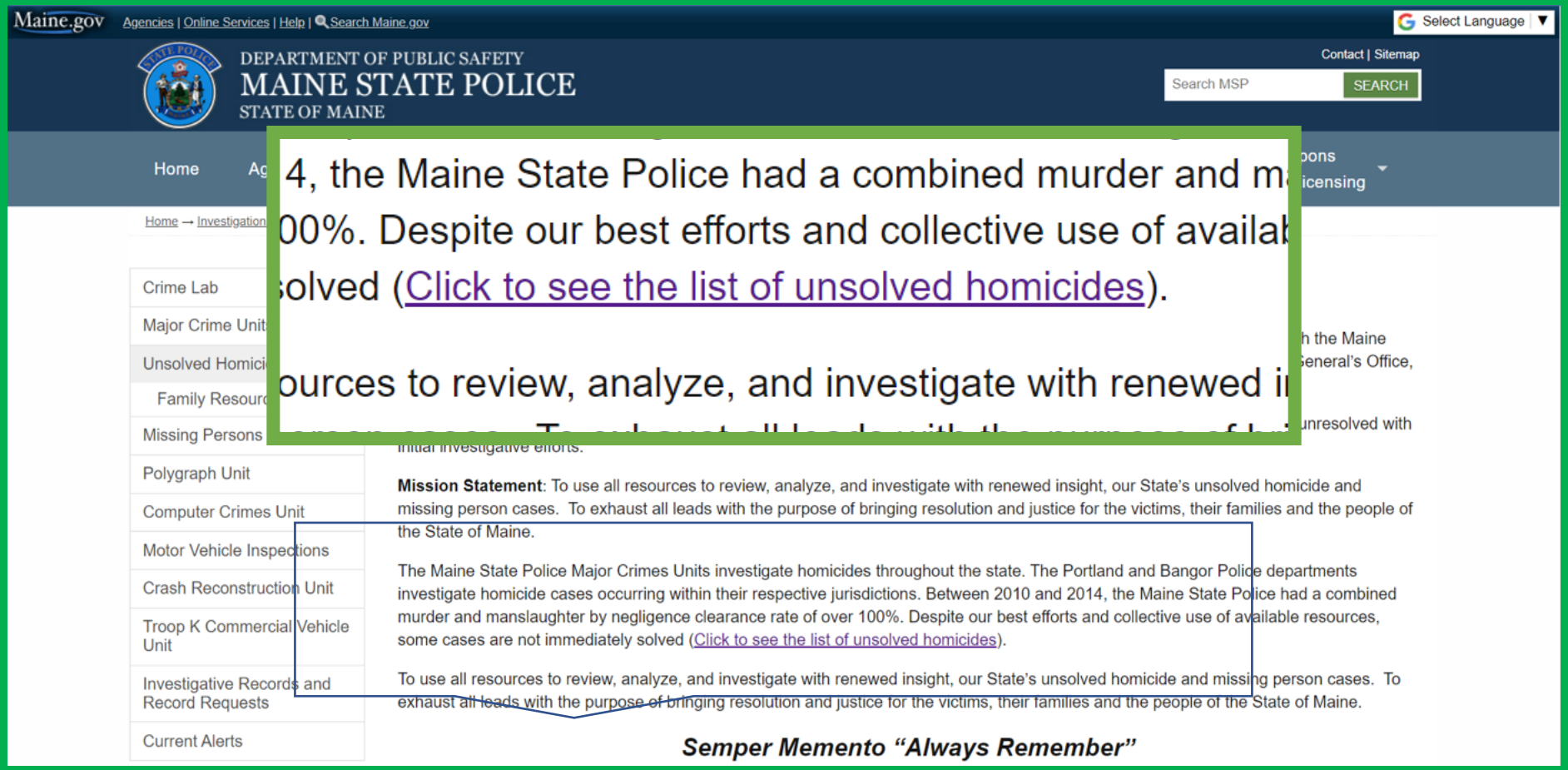


Click On the + Button On The Bottom of Your Workbook to Create a New Sheet

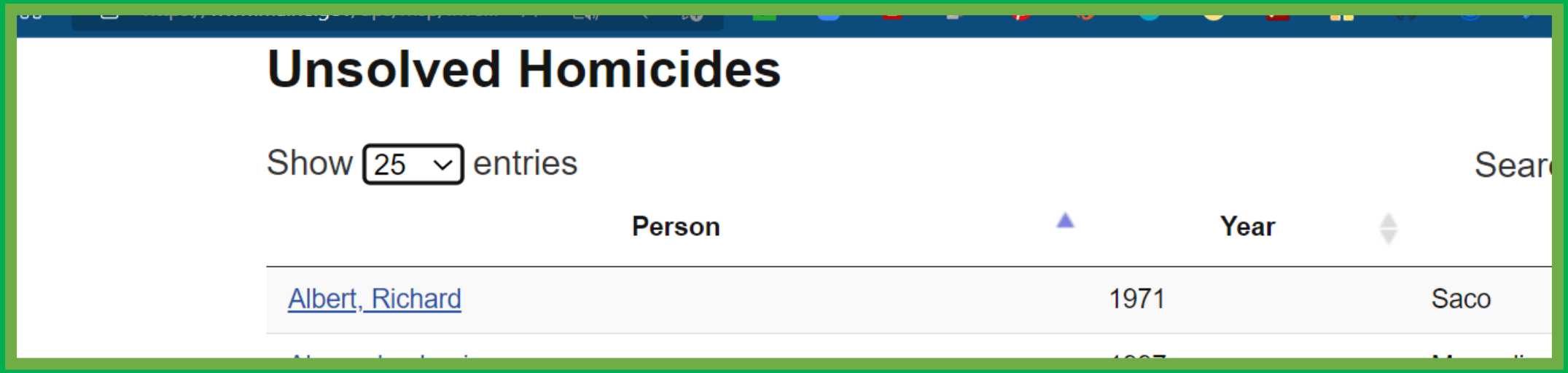




Click On The List of Unsolved Homicides On Your Web Page



Change the 25 To 100 To See The Full List



Unsolved Homicides

Show 100 entries

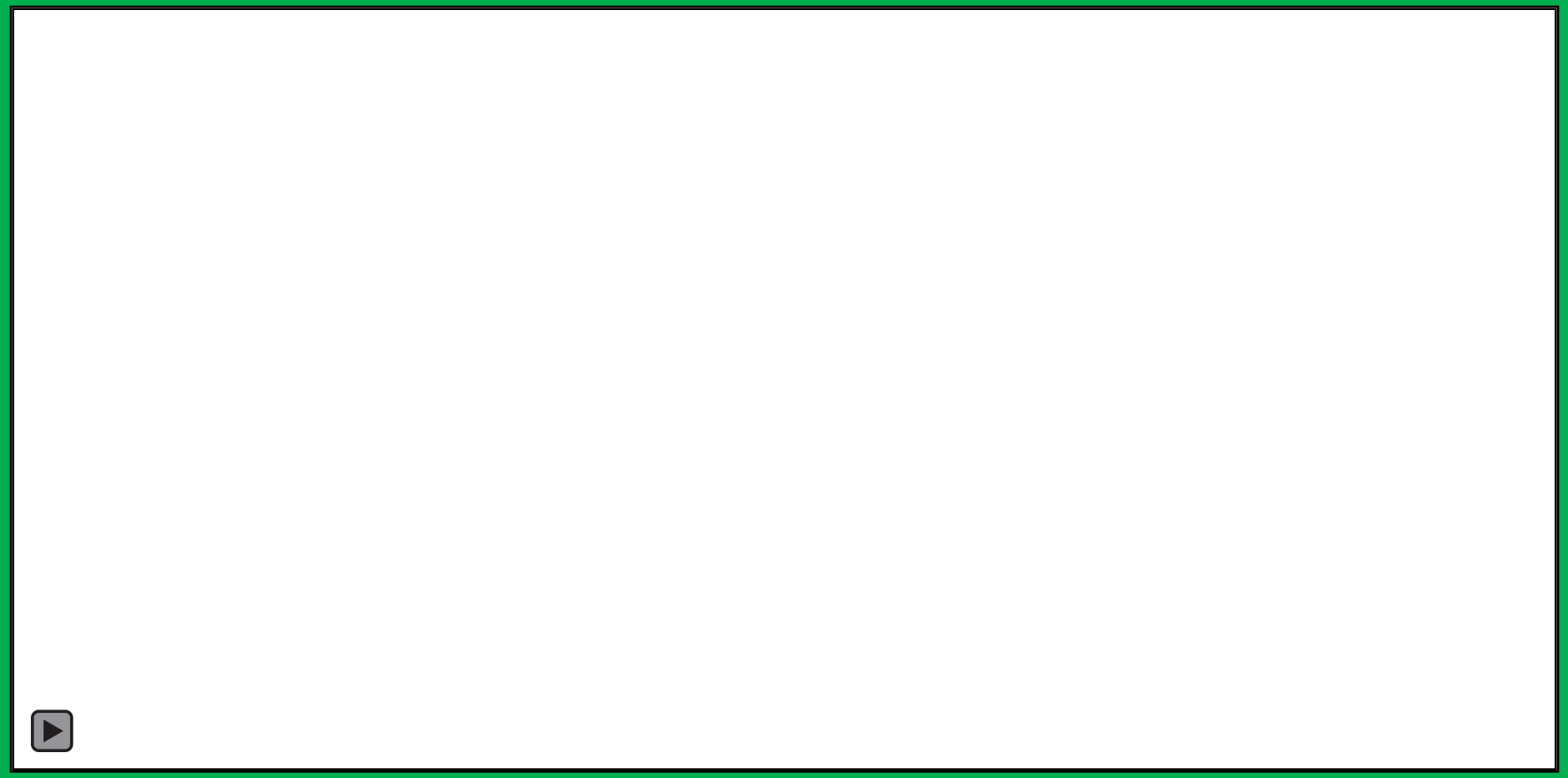
Search:

Person	Year	Town
Albert, Richard	1971	Saco
Alexander, Louis	1997	Masardis
Allen, Brian	2005	Orland
Baillargeon, Joseph	1968	Biddeford
Baldwin, Shirley	1976	Lisbon Falls
Bear, Anthony	1992	Fort Fairfield
Bitomski, Maxine T.	1993	Kittery
Blake, Alford "Buddy"	2003	Moscow
Bocuzzo, Louise	1994	Kennebunkport
Brackett, Lorna	1991	Waterville
Burke, Dorothea B.	1984	Stockton Springs
Campbell, Pamela M.	1981	Bangor/Holden

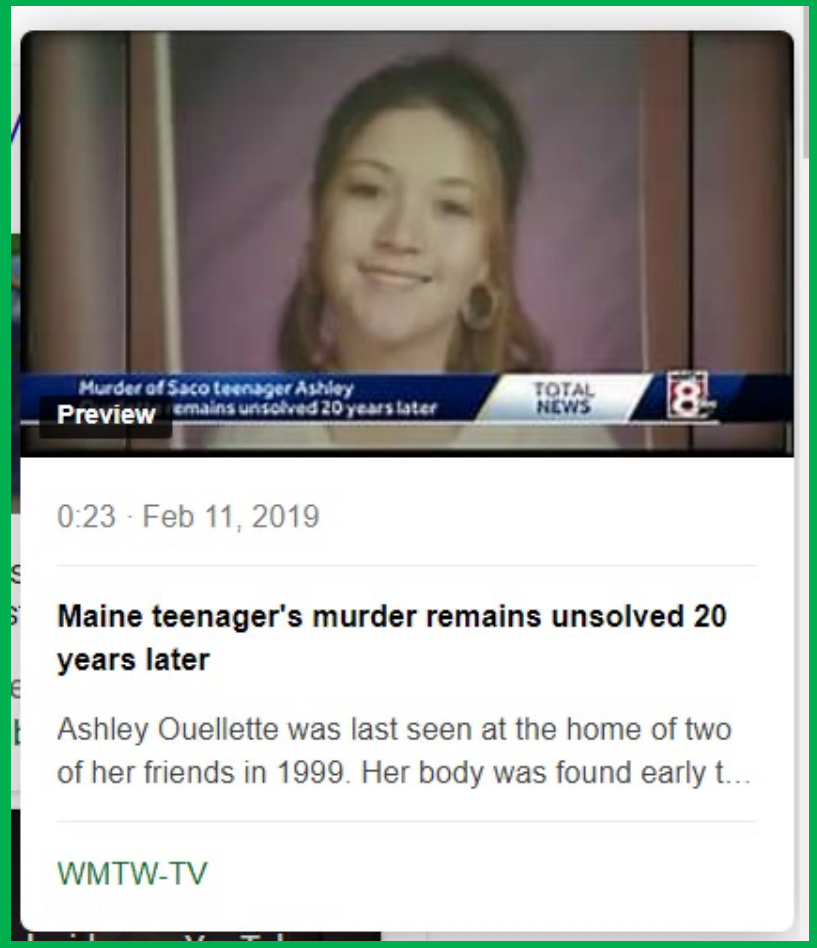
Repeat Your Scraping Steps

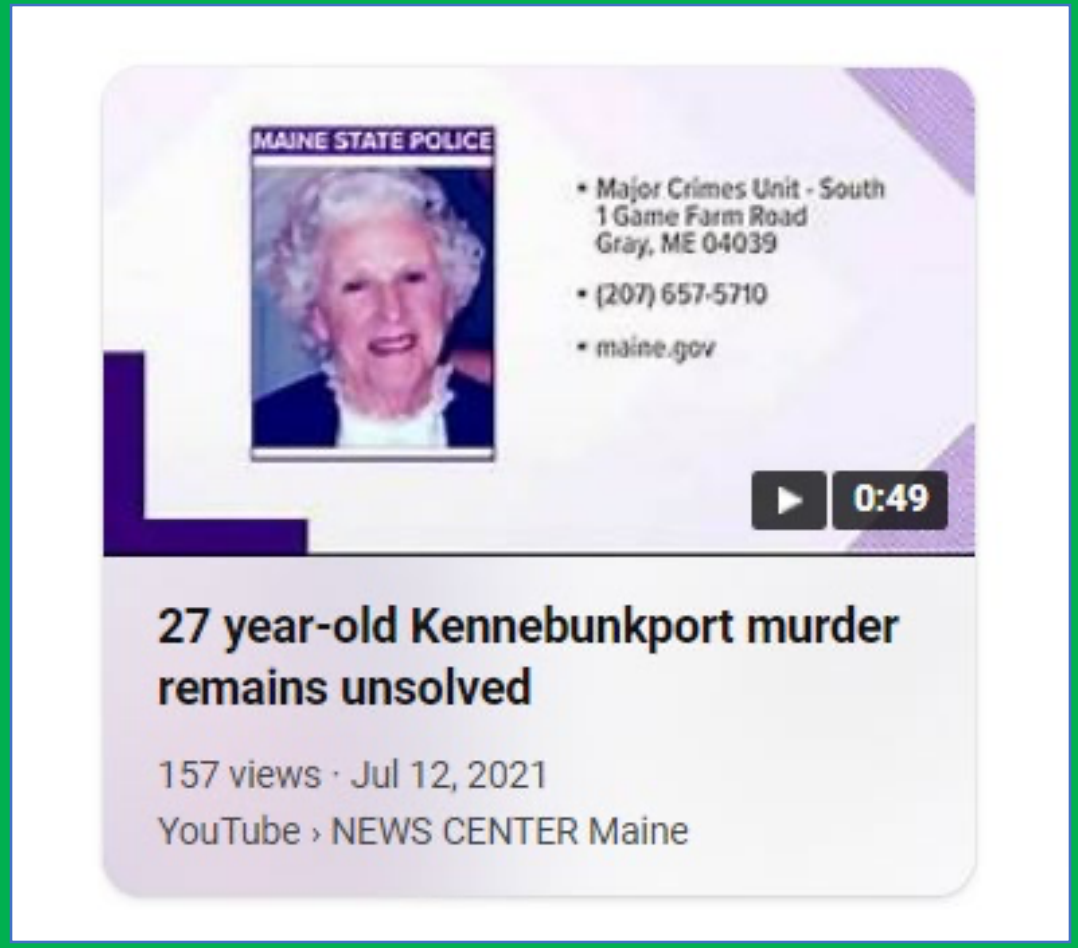
- Copy the URL
- In cell A-1 on your new sheet type =importhtml
- open parentheses, type a quote, paste in your URL, close quotes, comma
- Type “table”, add a comma after the quote mark
- Type 0 and close the parentheses
- Hit ENTER

Your Scrape Will Start



What Stories Could You Do With Your Scrape?



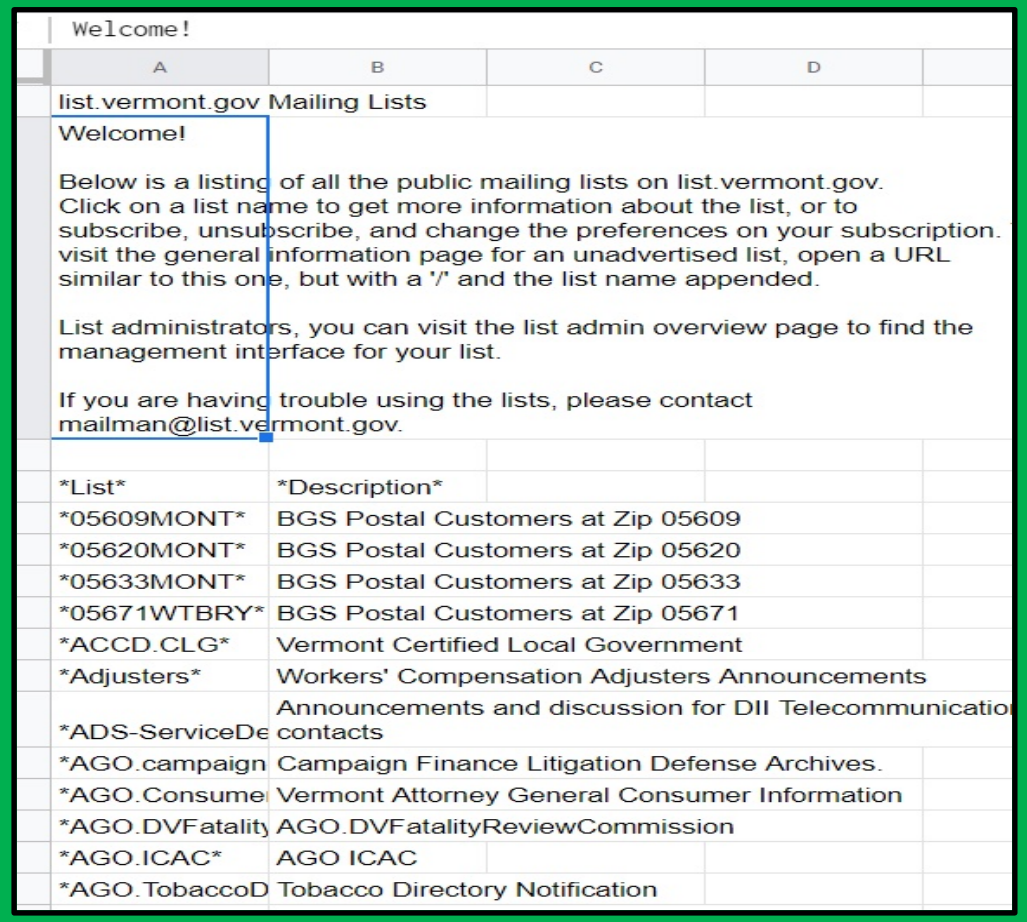


Scraping Lists With Google Sheets



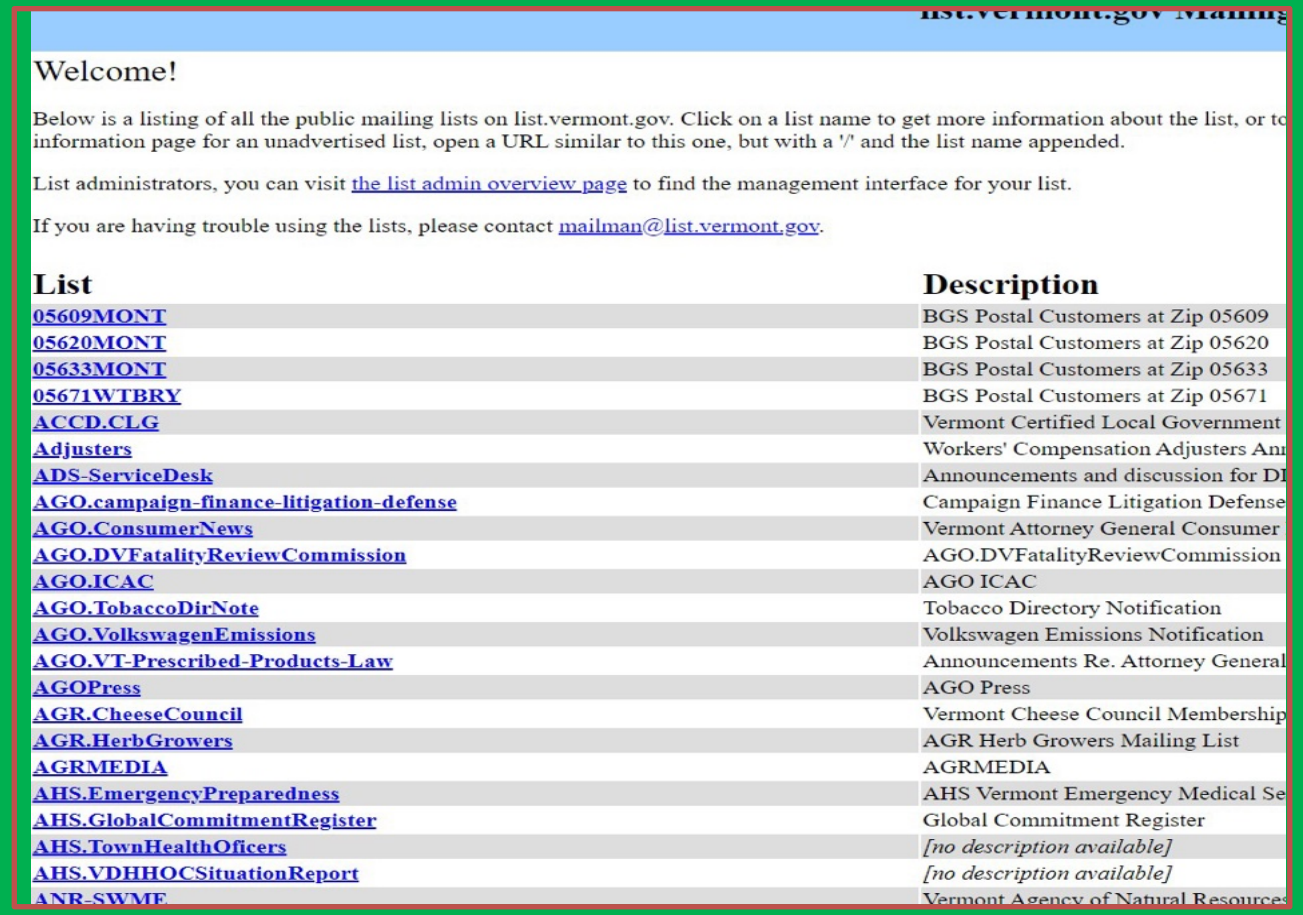
Web Scraping 101

Sometimes Tables Look Like Lists



Welcome!

	A	B	C	D
	list.vermont.gov Mailing Lists			
	Welcome! Below is a listing of all the public mailing lists on list.vermont.gov. Click on a list name to get more information about the list, or to subscribe, unsubscribe, and change the preferences on your subscription. visit the general information page for an unadvertised list, open a URL similar to this one, but with a '/' and the list name appended. List administrators, you can visit the list admin overview page to find the management interface for your list. If you are having trouble using the lists, please contact mailman@list.vermont.gov.			
	List	*Description*		
	05609MONT	BGS Postal Customers at Zip 05609		
	05620MONT	BGS Postal Customers at Zip 05620		
	05633MONT	BGS Postal Customers at Zip 05633		
	05671WTBRY	BGS Postal Customers at Zip 05671		
	ACCD.CLG	Vermont Certified Local Government		
	Adjusters	Workers' Compensation Adjusters Announcements		
	*ADS-ServiceDe	Announcements and discussion for DII Telecommunicatio contacts		
	*AGO.campaign	Campaign Finance Litigation Defense Archives.		
	*AGO.Consume	Vermont Attorney General Consumer Information		
	*AGO.DVFatality	AGO.DVFatalityReviewCommission		
	AGO.ICAC	AGO ICAC		
	*AGO.TobaccoD	Tobacco Directory Notification		



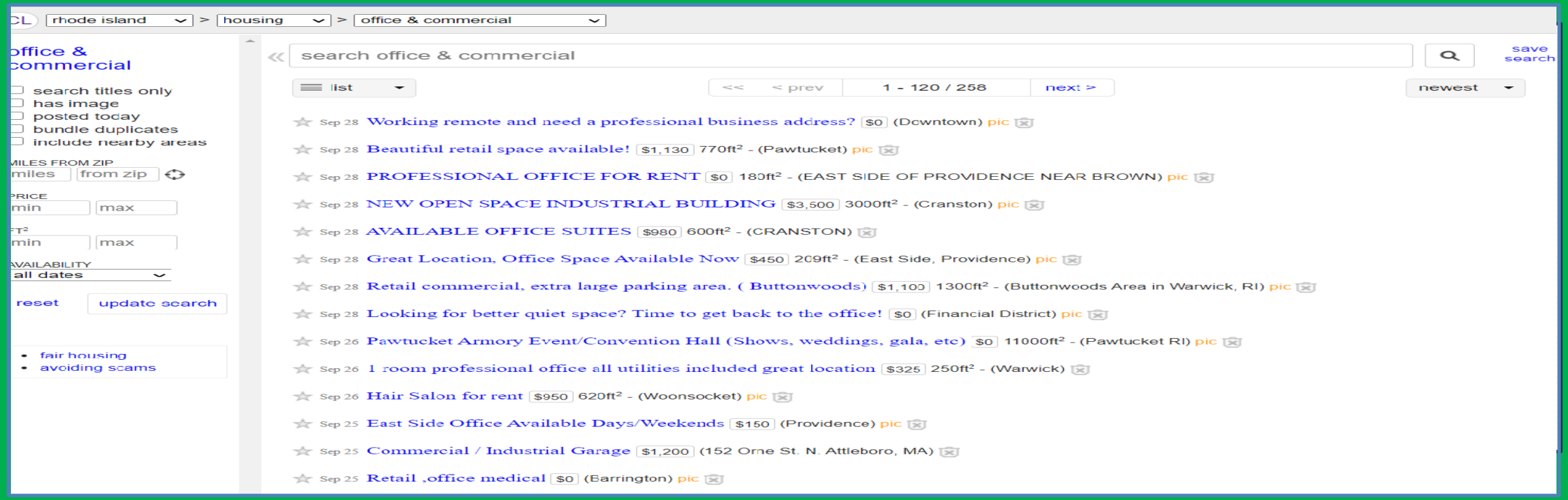
Welcome!

Below is a listing of all the public mailing lists on list.vermont.gov. Click on a list name to get more information about the list, or to information page for an unadvertised list, open a URL similar to this one, but with a '/' and the list name appended.

List administrators, you can visit the list admin overview page to find the management interface for your list.

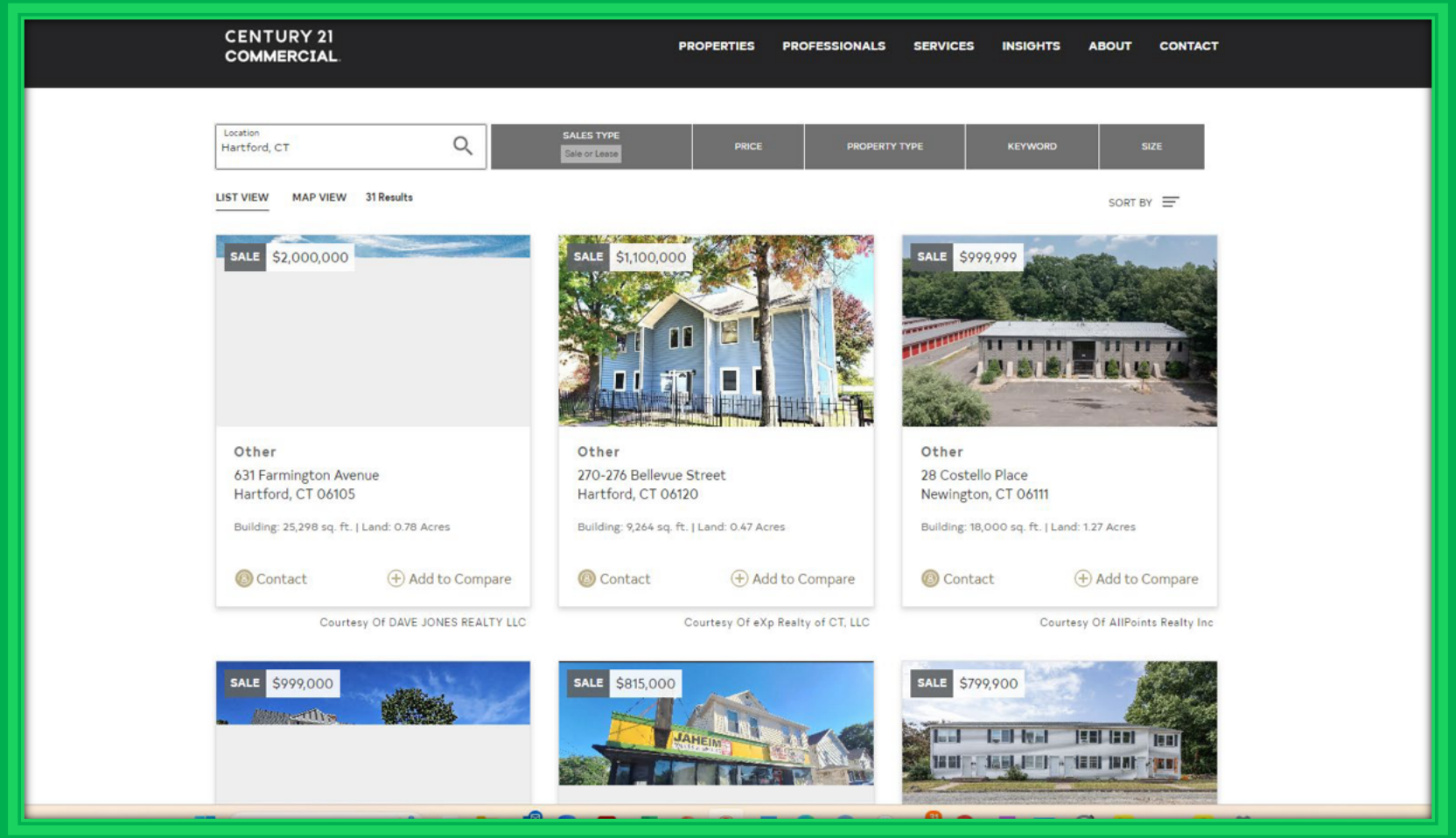
If you are having trouble using the lists, please contact mailman@list.vermont.gov.

List	Description
05609MONT	BGS Postal Customers at Zip 05609
05620MONT	BGS Postal Customers at Zip 05620
05633MONT	BGS Postal Customers at Zip 05633
05671WTBRY	BGS Postal Customers at Zip 05671
ACCD.CLG	Vermont Certified Local Government
Adjusters	Workers' Compensation Adjusters Ann
ADS-ServiceDesk	Announcements and discussion for DI
AGO.campaign-finance-litigation-defense	Campaign Finance Litigation Defense
AGO.ConsumerNews	Vermont Attorney General Consumer
AGO.DVFatalityReviewCommission	AGO.DVFatalityReviewCommission
AGO.ICAC	AGO ICAC
AGO.TobaccoDirNote	Tobacco Directory Notification
AGO.VolkswagenEmissions	Volkswagen Emissions Notification
AGO.VT-Prescribed-Products-Law	Announcements Re. Attorney General
AGOPress	AGO Press
AGR.CheeseCouncil	Vermont Cheese Council Membership
AGR.HerbGrowers	AGR Herb Growers Mailing List
AGRMEDIA	AGRMEDIA
AHS.EmergencyPreparedness	AHS Vermont Emergency Medical Se
AHS.GlobalCommitmentRegister	Global Commitment Register
AHS.TownHealthOficers	*[no description available]*
AHS.VDHHOCSituationReport	*[no description available]*
ANR-SWME	Vermont Agency of Natural Resources

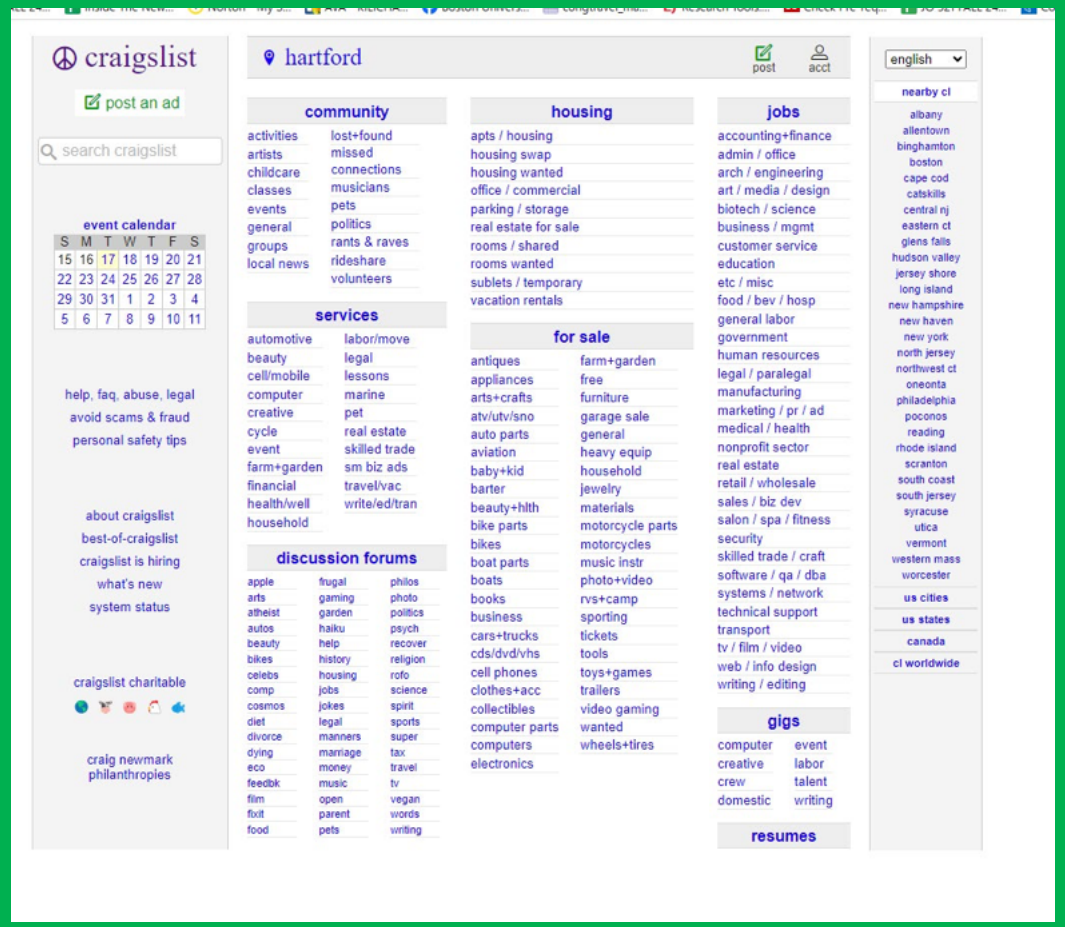


Data May Also Be In A List

Your Editor Wants You To Do A Story on the Vacant Commercial Real Estate in Hartford



We'll Scrape A List in Craig's List for Hartford To See What Properties Available and Their Costs



CL hartford housing office & commercial post acct

office & commercial

search office & commercial

price

search titles only
has image
posted today
show duplicates

MILES FROM LOCATION
miles from zip
use map...

PRICE
$ min – $ max
avg: $498
$0 $1k $2k $6.8k

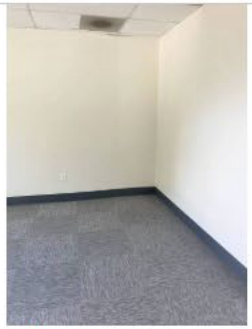
SQ FT
min - max

air conditioning

rent period
all dates

reset apply

gallery newest 1 - 69 of 69 save



Great Location - Affordable Office Space...
8h ago • Vernon, CT

$500



20% OFF Brand New Offices in New B...
8h ago • New Britain, CT

$525



Historic Downtown, Professional Neighb...
12/17 • 400ft² • Historic Downtown Dis...

$2,000



Low Rent Commercial/Light Industrial/...
12/16 • 1500ft² • New Milford



CENTER OF MANCHESTER, CT
12/16 • Manchester, CT



PROFESSIONAL OFFICES - ROCKY HI...
12/16 • Rocky Hill



SOUGHT AFTER OFFICE SPACE - BL...
12/16 • Bloomfield



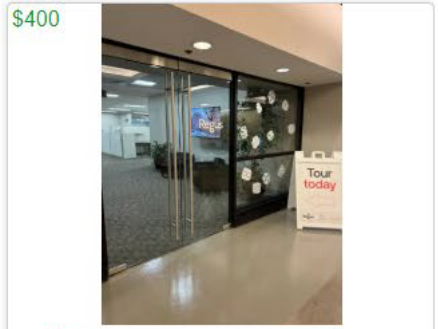
Warehouse Flex Space with Drive-in Do...
12/16 • 1000ft² • East Windsor

$400



Professional Office Space in Windsor - ri...
12/15 • Windsor/Bloomfield

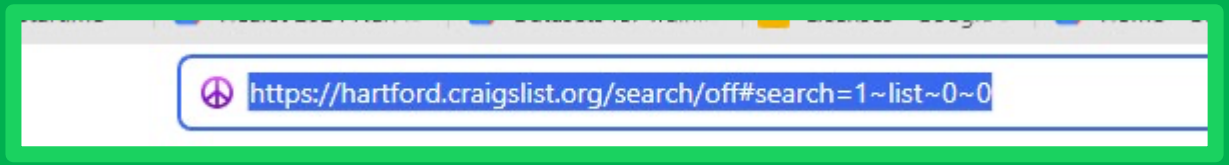
$400

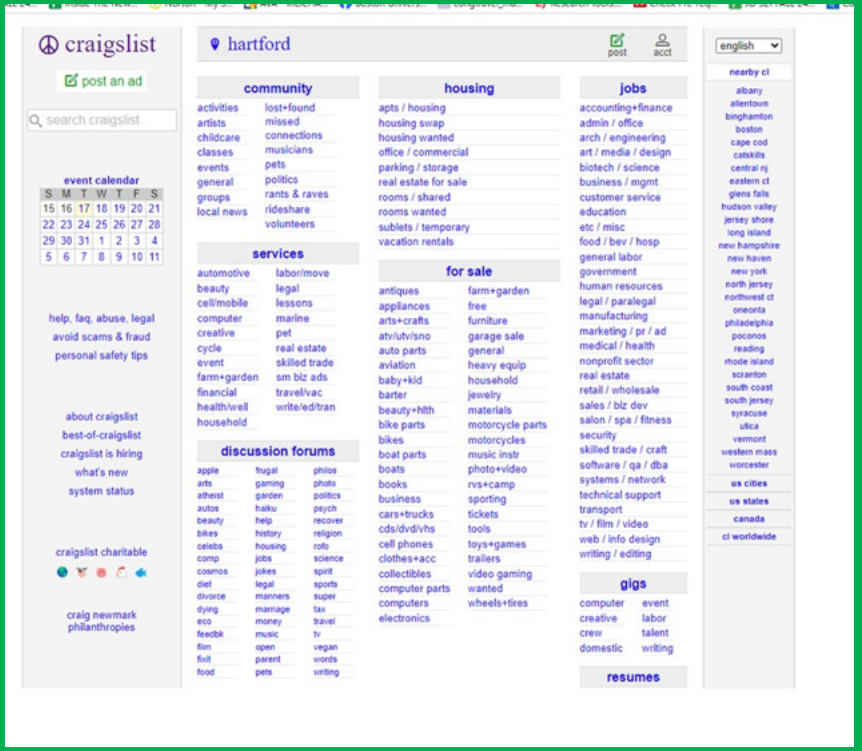


Amenity Rich Office Space on Windsor/...
12/15 • Windsor/Bloomfield

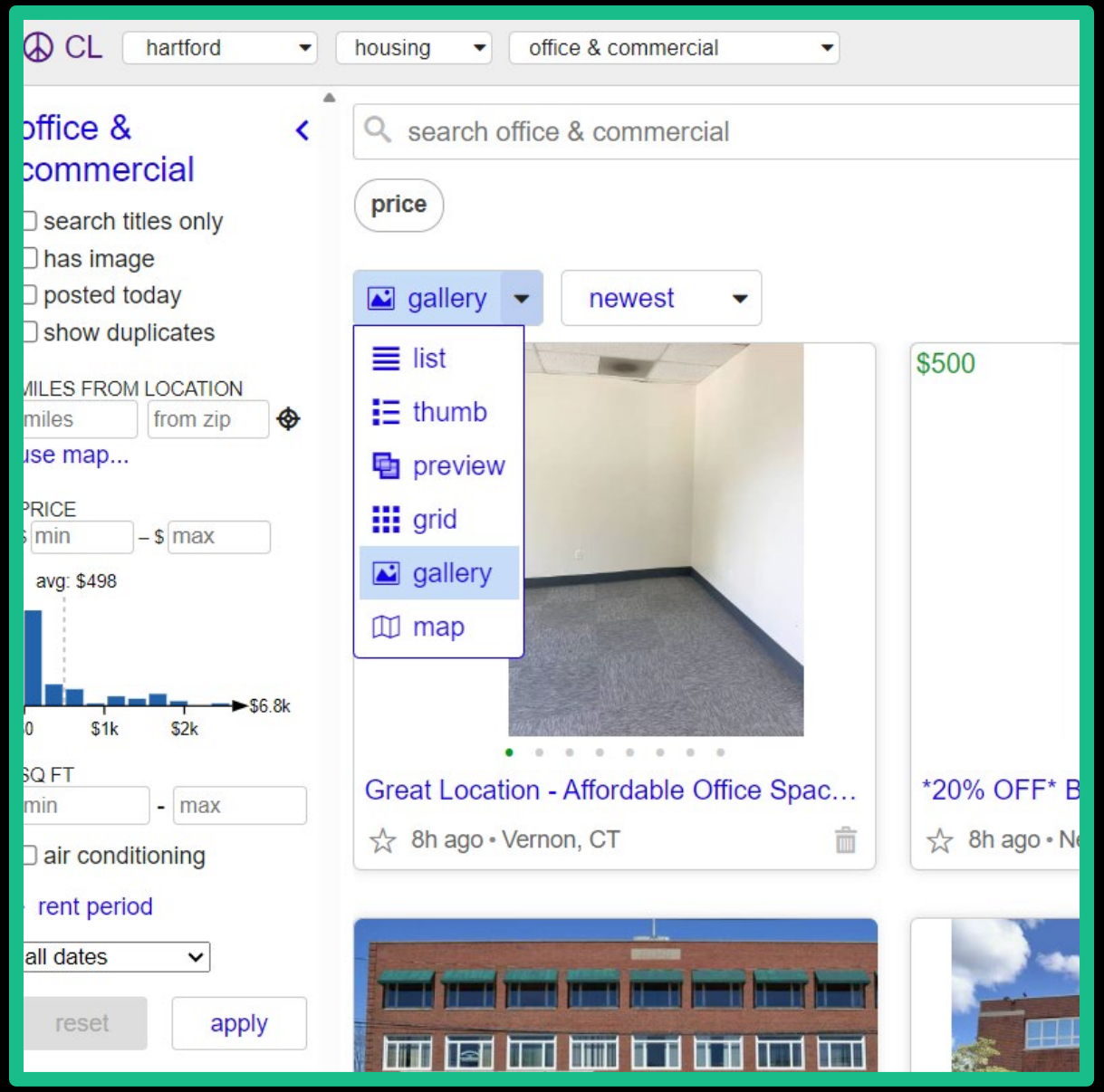
fair housing
avoiding scams

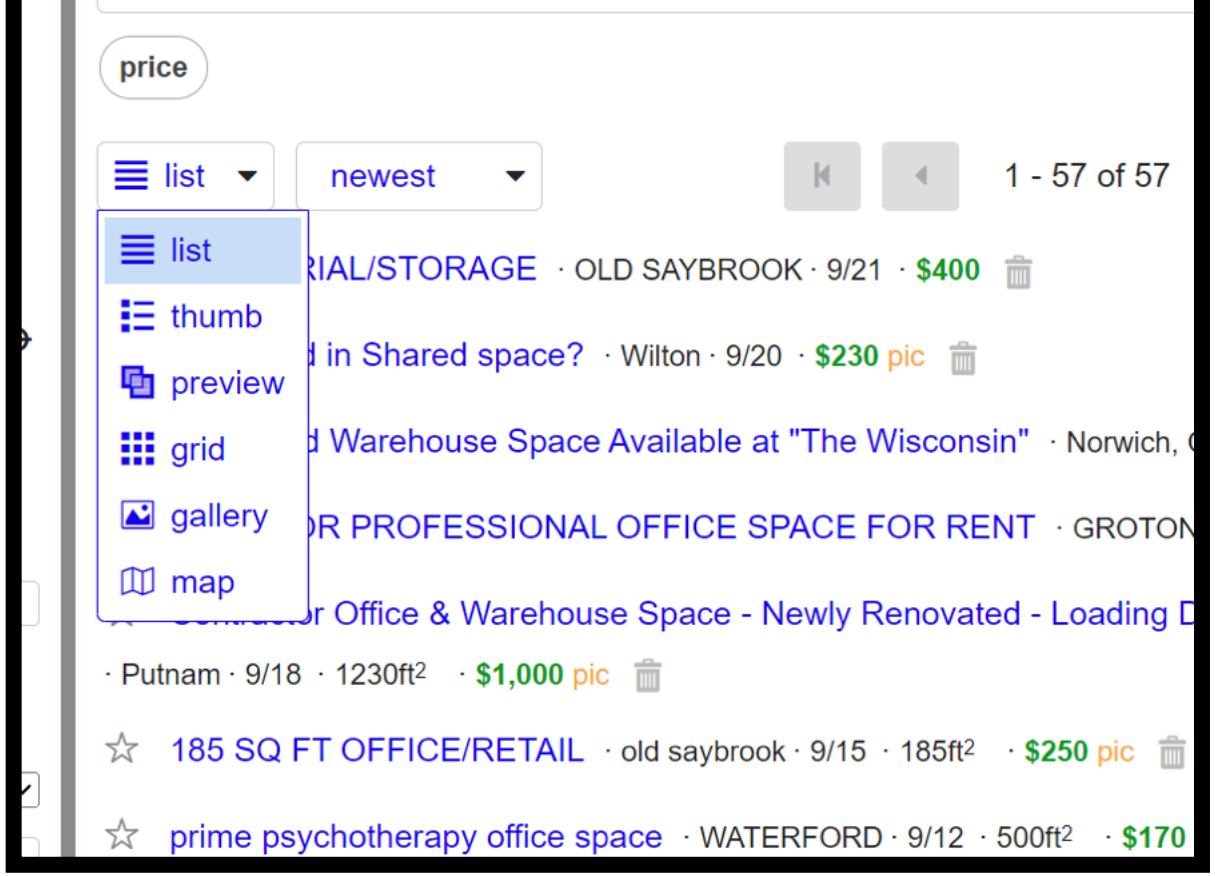
Craig's List – When You Find Hartford, Choose Housing, Then Commercial Real Estate
Right Click and Copy the URL





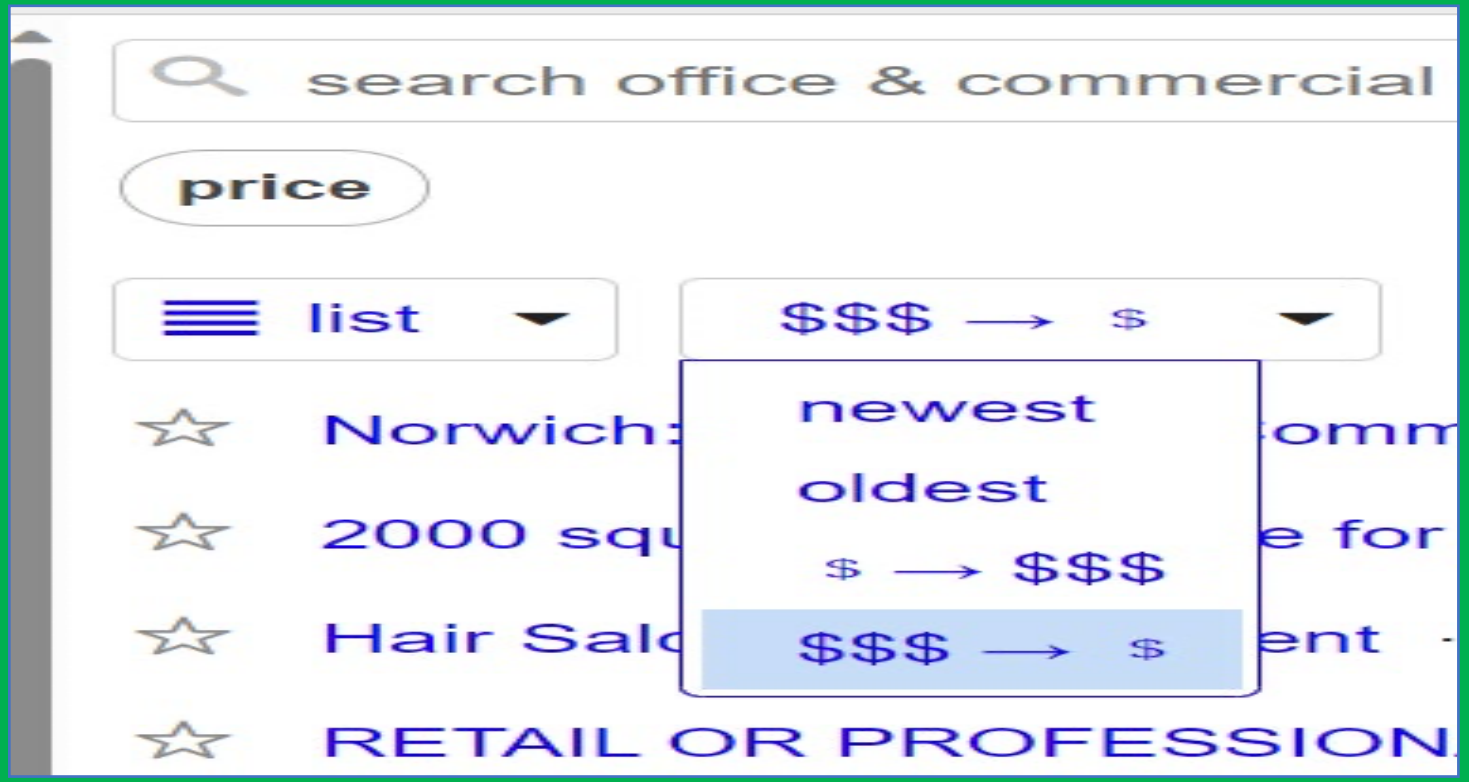
Switch to List View

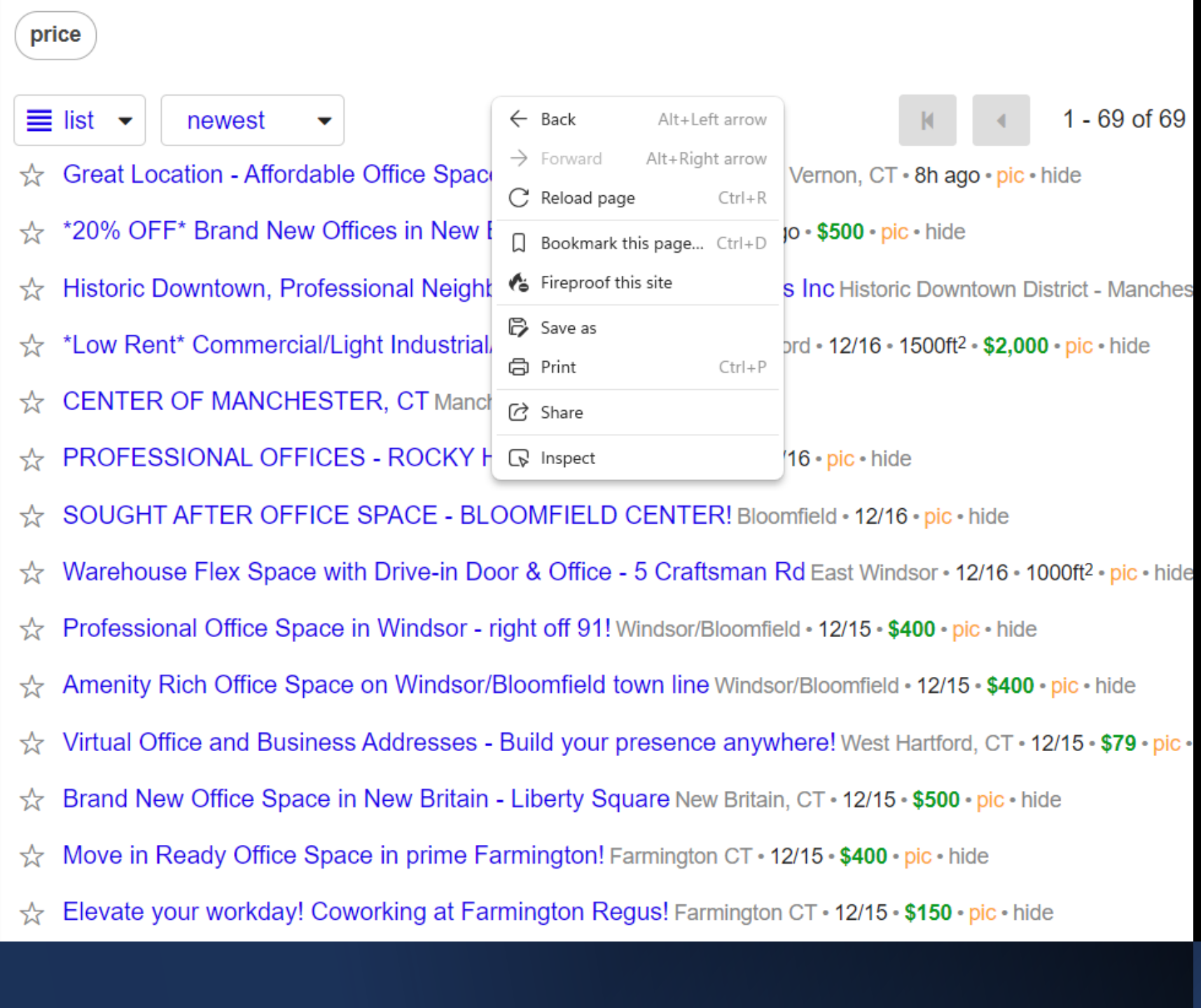




Switch From Gallery to List View

You Can Also Change the Range of Your List

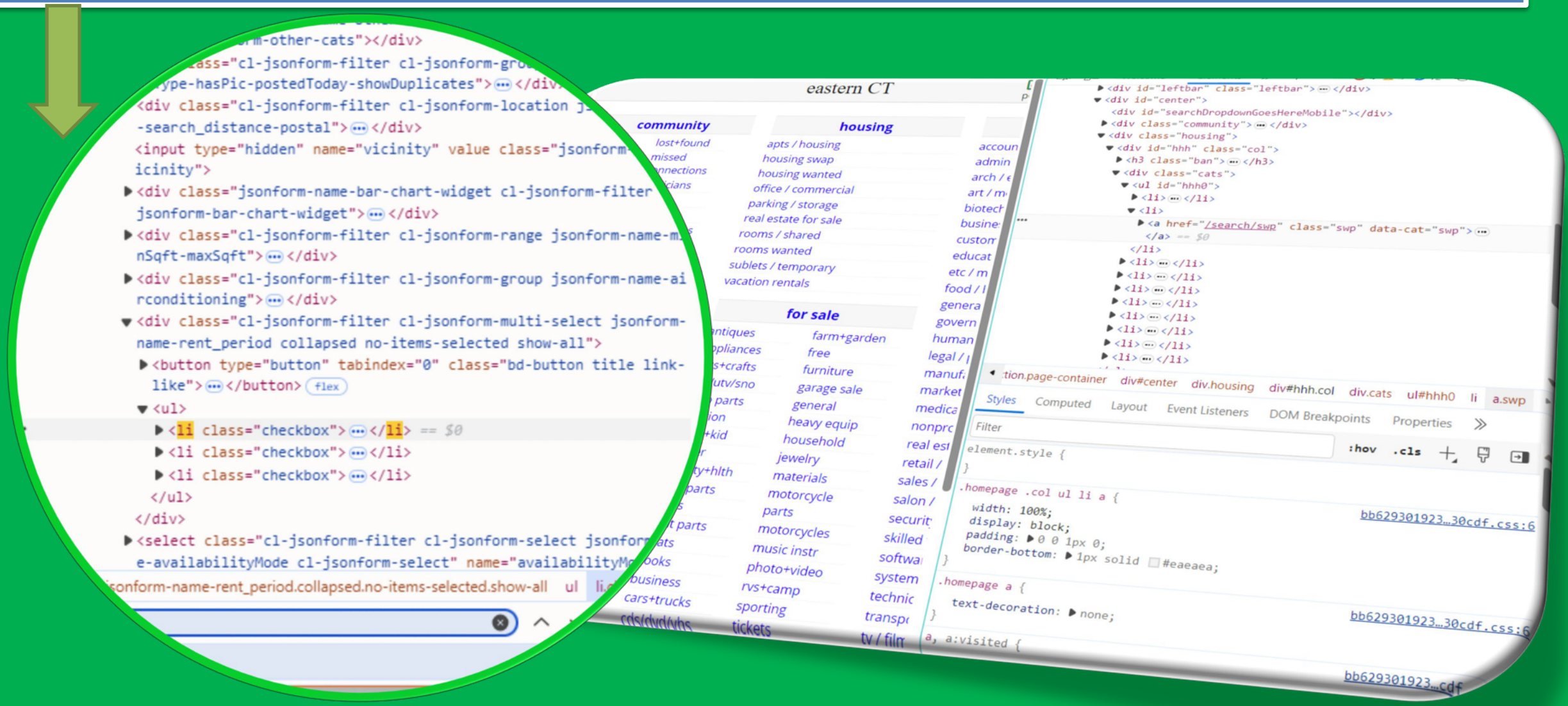


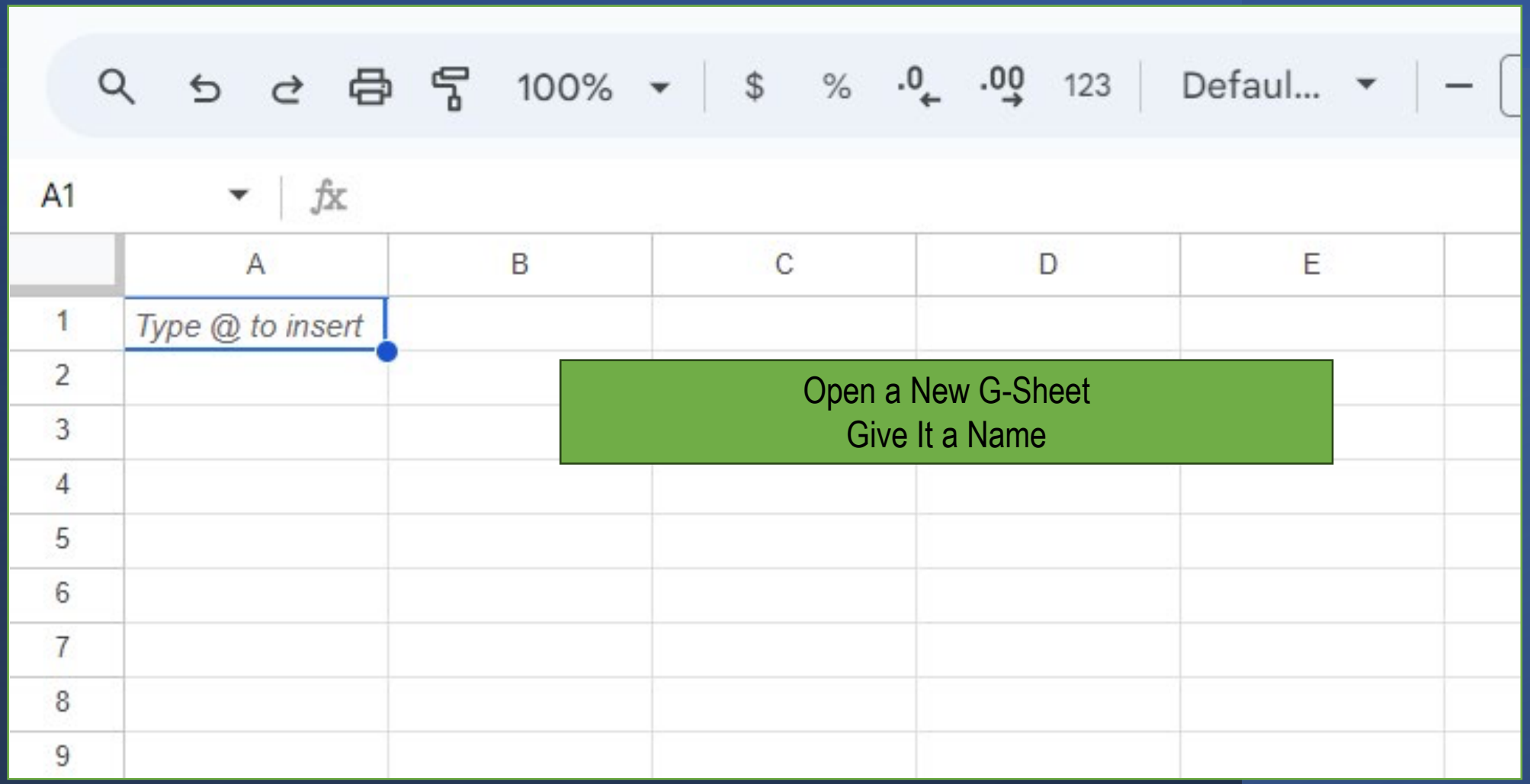


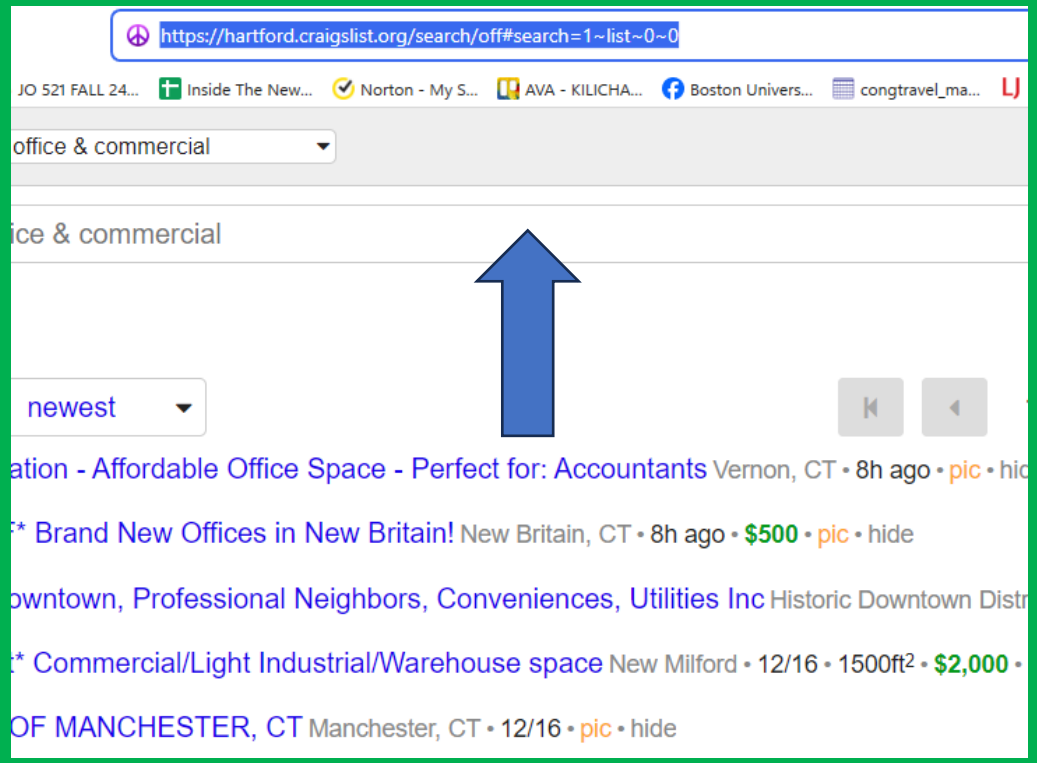
Let's Confirm This Page Is a List

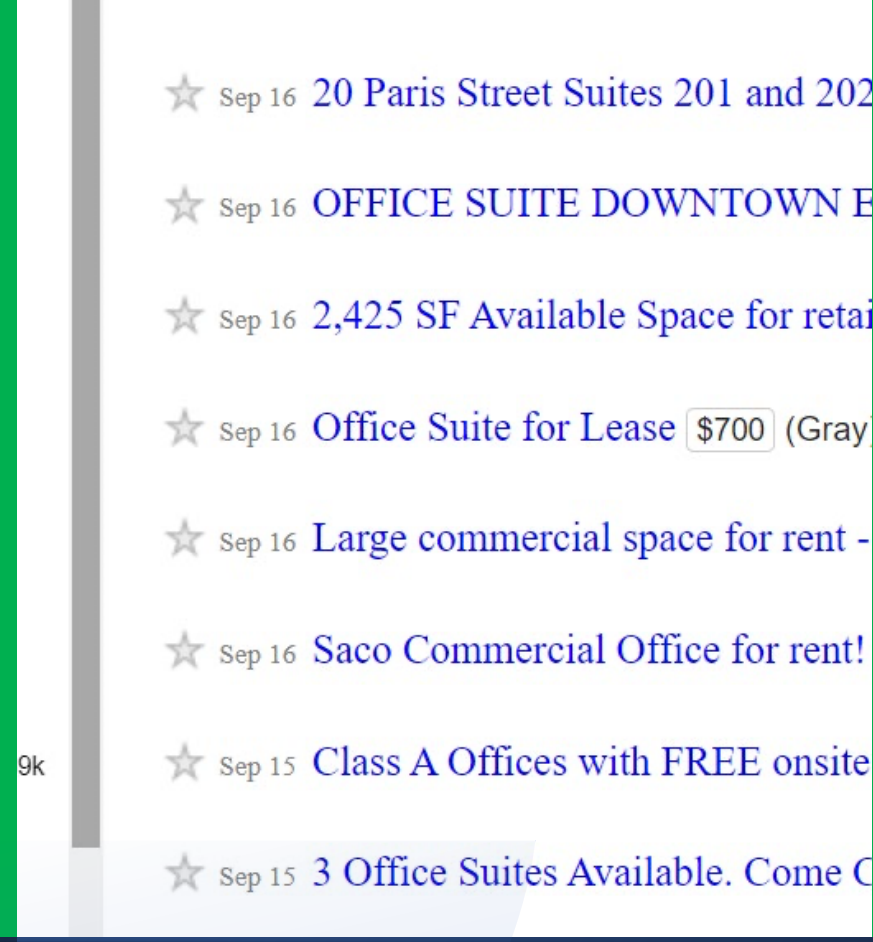
- Right click on the page
- Click "inspect"
- Do a CTRL F (PC) or CMND F (Mac)
- Search for "li" or "list"
- If you search for "table" you won't find it
- That is how we know we are going to scrape a list

You Can Doublecheck Your Element By Going “Under The Hood”




100%
$
%
.0
.00
123
Defaul...
A1
fx
A
B
C
D
E
1
Type @ to insert
2
3
4
5
6
7
8
9
Open a New G-Sheet
Give It a Name

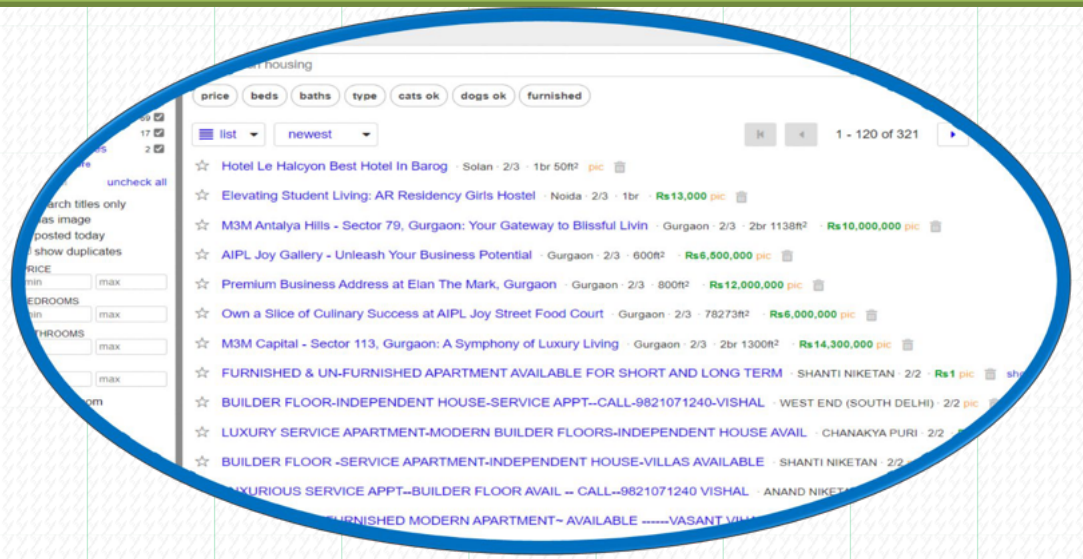




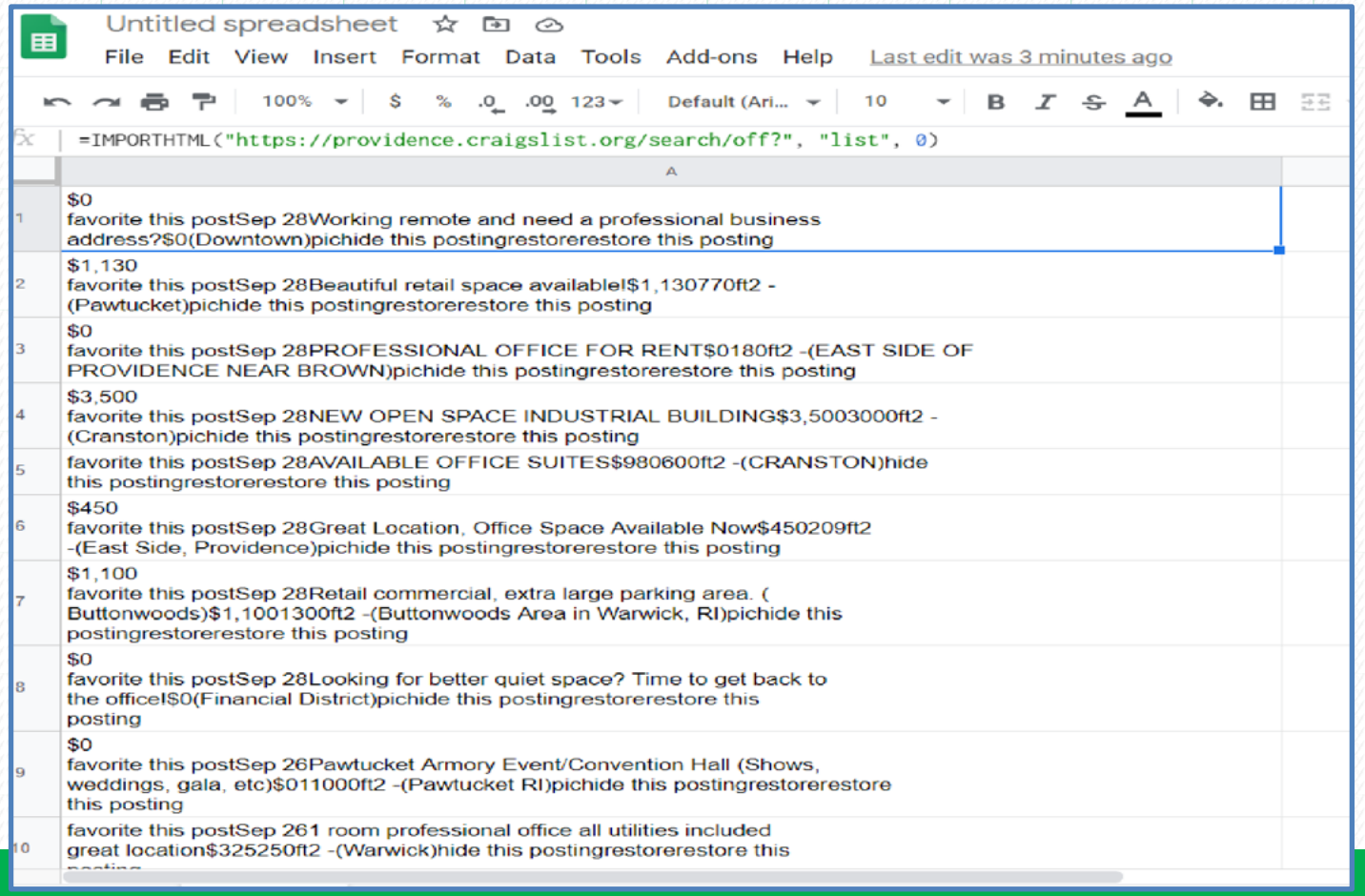
Make Sure You Have Copied Your URL

We Are Going To Tweak Our Formula

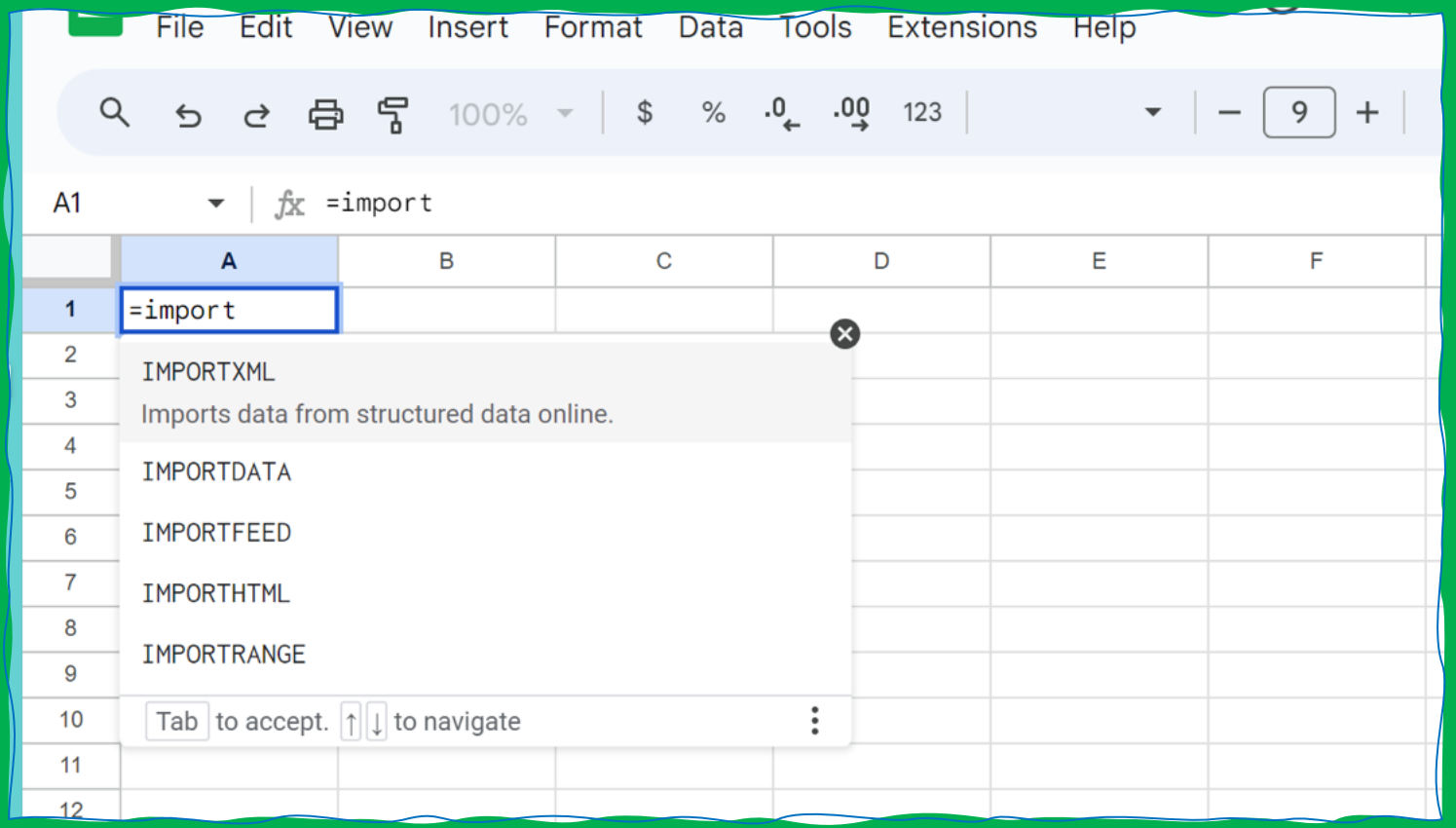
Instead of using “table” as our second element, we will use “list”



Scraping A List With G-Sheets



In Cell AI, type =importhtml



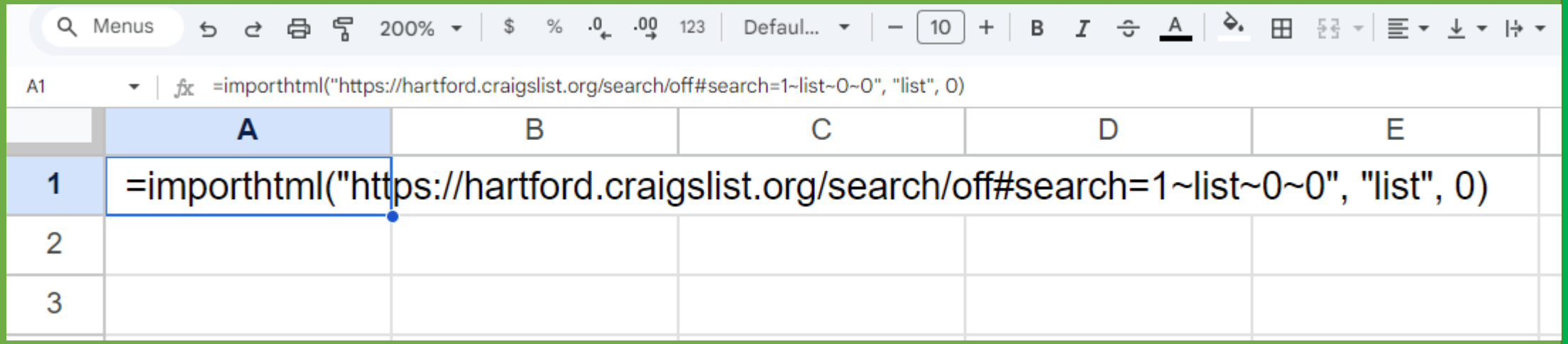
Add Your URL To Your Formula

- After import HTML, open parentheses
  - Open quotes
  - Paste your URL
  - Close quotes
  - Comma

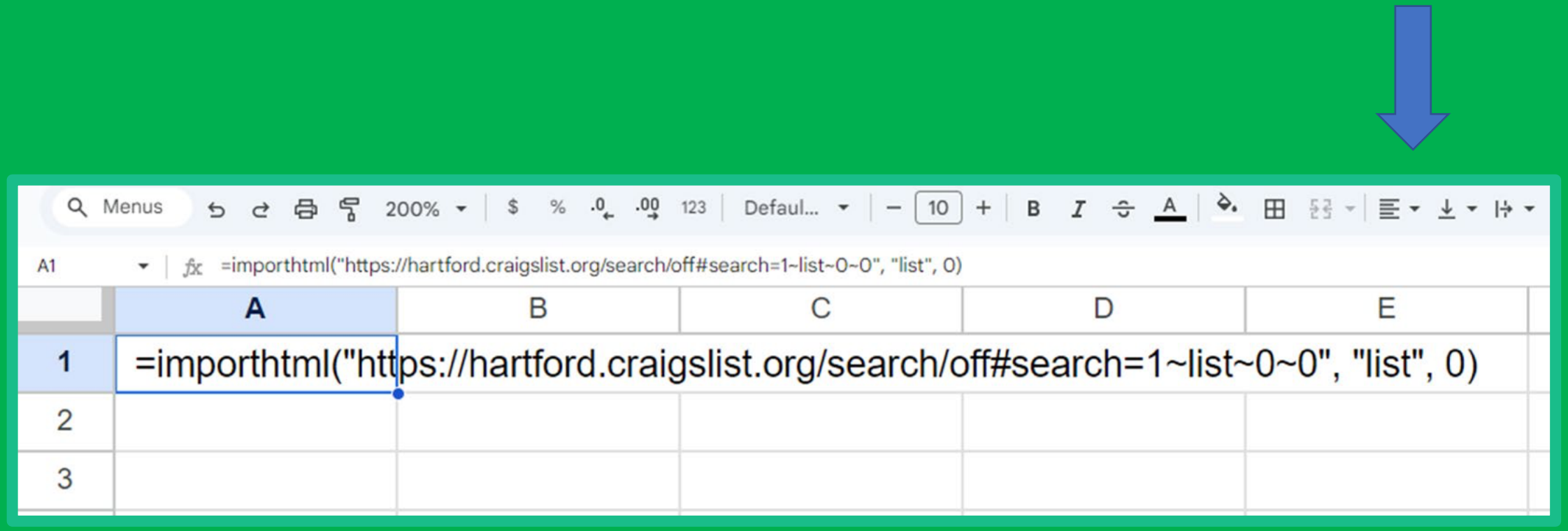


Add The Type of Element We Are Scraping To Your Formula

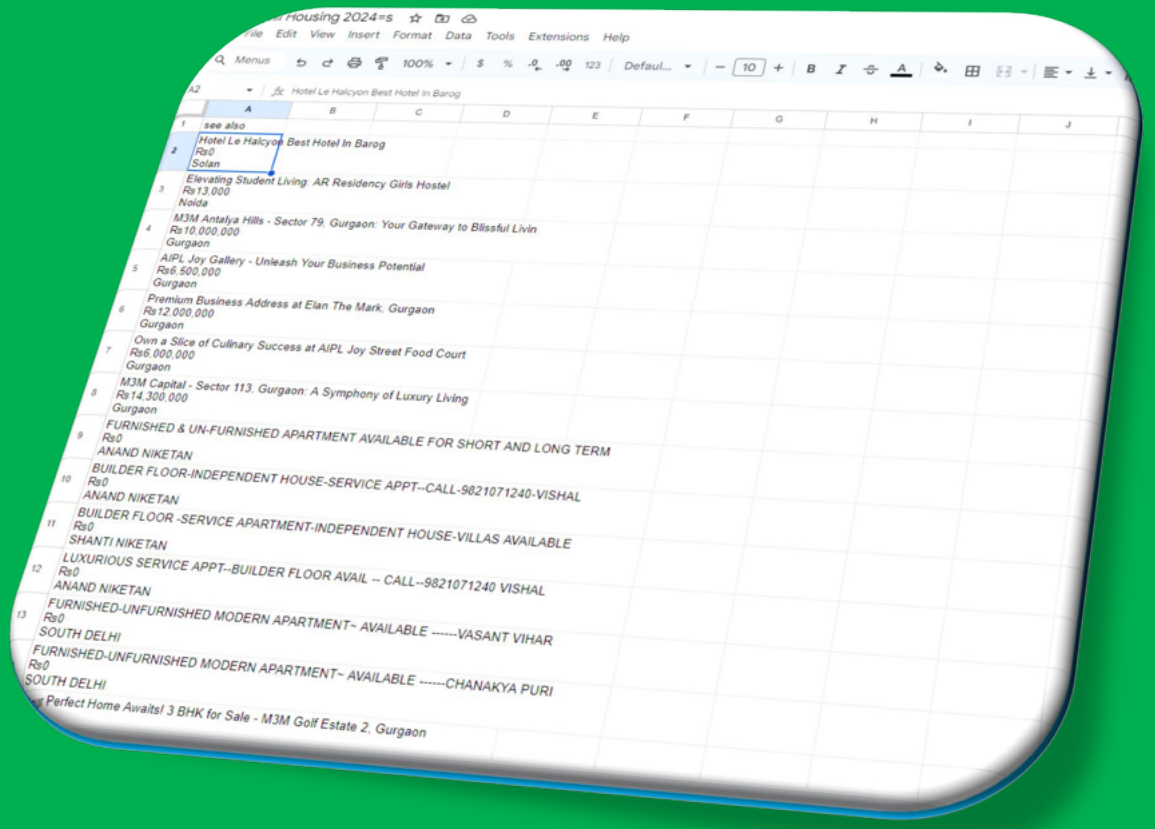
- We are scraping a list, so just replace “table” with “list”
- Don’t forget the comma after “list”



Tell G-Sheets Where On The Web Page The List Appears

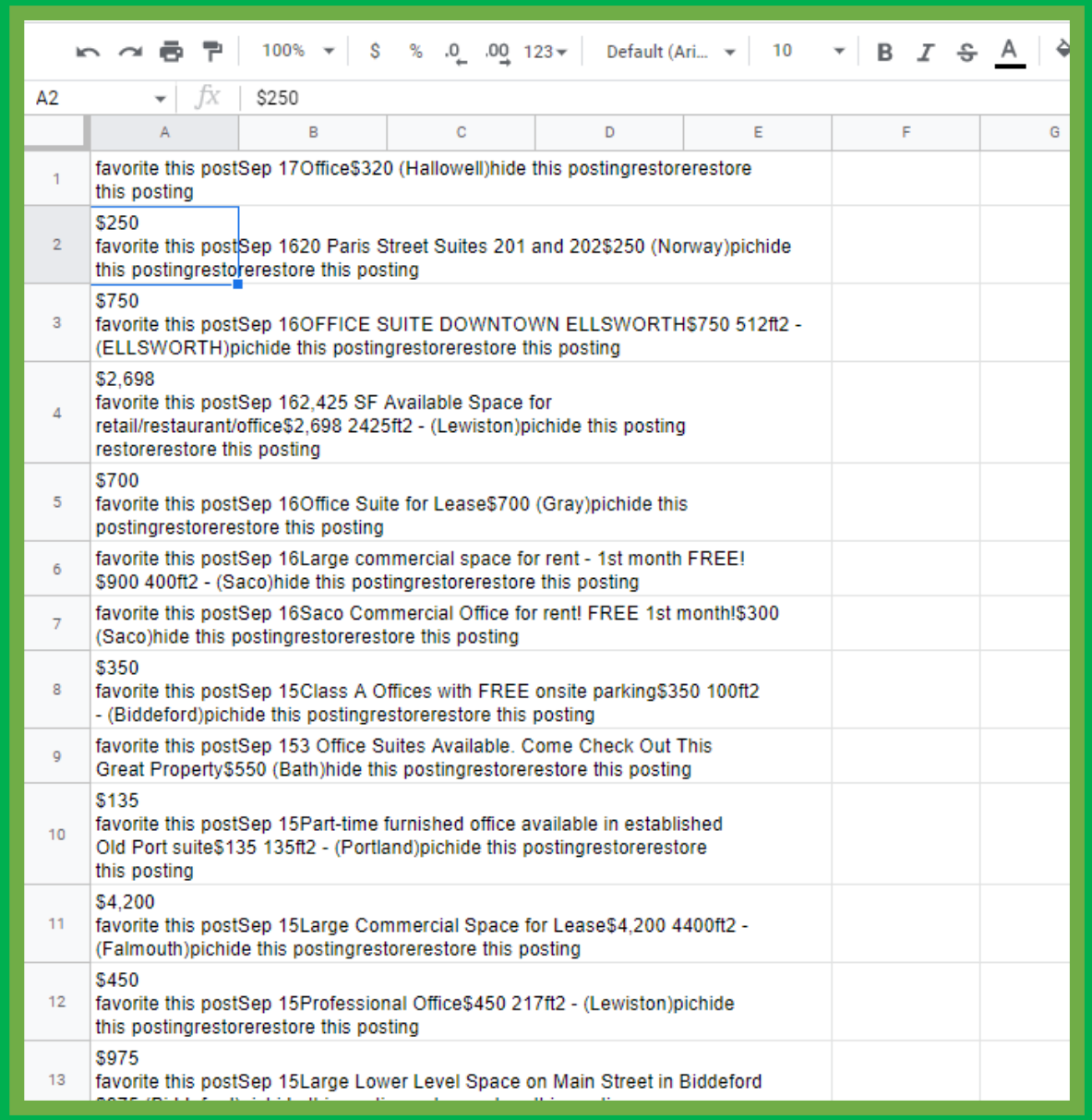


Hit Enter and Your Data Will Load



We Can Clean This Data Up By Separating Out Specific Pieces of Information

- For now, each time Craig's list updates its commercial real estate listing, your G-Sheets will update



100% $ % .0 .00 123 Default (Ari... 10

A2 $250

	A	B	C	D	E	F	G
1	favorite this postSep 17Office$320 (Hallowell)hide this postingrestorerestore this posting						
2	$250 favorite this postSep 1620 Paris Street Suites 201 and 202$250 (Norway)pichide this postingrestorerestore this posting						
3	$750 favorite this postSep 16OFFICE SUITE DOWNTOWN ELLSWORTH$750 512ft2 - (ELLSWORTH)pichide this postingrestorerestore this posting						
4	$2,698 favorite this postSep 162,425 SF Available Space for retail/restaurant/office$2,698 2425ft2 - (Lewiston)pichide this posting restorerestore this posting						
5	$700 favorite this postSep 16Office Suite for Lease$700 (Gray)pichide this postingrestorerestore this posting						
6	favorite this postSep 16Large commercial space for rent - 1st month FREE! $900 400ft2 - (Saco)hide this postingrestorerestore this posting						
7	favorite this postSep 16Saco Commercial Office for rent! FREE 1st month!$300 (Saco)hide this postingrestorerestore this posting						
8	$350 favorite this postSep 15Class A Offices with FREE onsite parking$350 100ft2 - (Biddeford)pichide this postingrestorerestore this posting						
9	favorite this postSep 153 Office Suites Available. Come Check Out This Great Property$550 (Bath)hide this postingrestorerestore this posting						
10	$135 favorite this postSep 15Part-time furnished office available in established Old Port suite$135 135ft2 - (Portland)pichide this postingrestorerestore this posting						
11	$4,200 favorite this postSep 15Large Commercial Space for Lease$4,200 4400ft2 - (Falmouth)pichide this postingrestorerestore this posting						
12	$450 favorite this postSep 15Professional Office$450 217ft2 - (Lewiston)pichide this postingrestorerestore this posting						
13	$975 favorite this postSep 15Large Lower Level Space on Main Street in Biddeford						

Introduction to Data Cleaning With Open Refine



OpenRefine

A power tool for working with messy data.

What Is Dirty Data?



OPEN Refine

Environmental Spills - Connecticut

	A	B	C	D	E	F	G	H	I	J	K	L
1	Year	Case No.	Assigned to	Date Reported Time Reporte	Release date and tin	Location Of Reportec	Town of Release	State of Release		Status	Reported By	Representing
2	2009	2.01E+08	LeClerc, Ken	11/24/2009 0:00	11/24/2009 0:00	96 SUMMERS ST	SOUTHINGTON	CT		CLOSED	CL&P	Self
3	2009	2.01E+08	No Response	11/24/2009 0:00	11/24/2009 0:00	I-84 EAST EXIT 34	PLAINVILLE	CT		CLOSED	PLAINVILLE FIRE	Self
4	2009	2.01E+08	No Response	11/24/2009 9:02	11/24/2009 9:02	916 tower court	MANSFIELD	CT		CLOSED	dispatcher	uconn fire department
5	2009	2.01E+08	No Response	11/24/2009 0:00	11/24/2009 0:00	Route 12 at Sub Bas	GROTON	CT		CLOSED	Richard Conant	Groton Sub Base Environmental
6	2009	2.01E+08	Scalora, Rich	11/24/2009 0:00	11/24/2009 0:00	27 town farm rd	BERLIN	CT		CLOSED	chief simons	fd
7	2009	2.01E+08	LeClerc, Ken	11/24/2009 0:00	11/24/2009 0:00	WINSTED, 261 north	WINCHESTER	CT		CLOSED	Maryanne Graham	Lehigh Fuel
8	2009	2.01E+08	No Response	11/24/2009 11:25	11/24/2009 11:25	119 FAIRFAX STREE	WEST HAVEN	CT		CLOSED	DISPATCHER	WEST HAVEN FIRE DEPARTMENT
9	2009	2.01E+08	No Response	11/24/2009 13:14	11/24/2009 13:14	108 warren hill road	CORNWALL	CT		CLOSED	dispatcher	troop b
10	2009	2.01E+08	No Response	11/24/2009 0:00	11/24/2009 0:00	9 royal oak	WEST HARTFORD	CT		CLOSED	Charlie Freimuth	Oakland
11	2009	2.01E+08	No Response	11/24/2009 0:00	11/24/2009 0:00	15 beth drive	FAIRFIELD	CT		CLOSED	patrick neary	McGoldrick Fuel inc
12	2009	2.01E+08	No Response	11/24/2009 0:00	11/24/2009 0:00	north colony road an	BERLIN	CT		CLOSED	dave	bfd
13	2009	2.01E+08	Scalora, Rich	11/24/2009 0:00	11/24/2009 0:00	93 court st	WEST HAVEN	CT		CLOSED	fd	fd
14	2009	2.01E+08	No Response	11/24/2009 0:00	11/24/2009 0:00	5 hitchingpost drive	WALLINGFORD	CT		CLOSED	margaret bissell	Self
15	2009	2.01E+08	No Response	11/24/2009 0:00	11/24/2009 0:00	rt 9 north exit 18	CROMWELL	CT		CLOSED	dick	fd
16	2009	2.01E+08	No Response	11/24/2009 0:00	11/24/2009 0:00	911 blue hills ave	BLOOMFIELD	CT		CLOSED	anthony	ct transit
17	2009	2.01E+08	No Response	11/24/2009 0:00	11/24/2009 0:00	131 dowd av	CANTON	CT		CLOSED	fd	fd
18	2009	2.01E+08	No Response	11/24/2009 0:00	11/24/2009 0:00	I-95 SOUTH EXIT 40	MILFORD	CT		CLOSED	FD	FD
19	2009	2.01E+08	No Response	11/24/2009 0:00	11/24/2009 0:00	77 DELEO DRIVE	STAMFORD	CT		CLOSED	FD	FD
20	2009	2.01E+08	Scalora, Rich	11/25/2009 0:00	11/25/2009 0:00	IN FRONT OF 50 FIE	CROMWELL	CT		CLOSED	CROMWELL FIRE	Self
21	2009	2.01E+08	No Response	11/25/2009 0:00	11/25/2009 0:00	632 RIVERTON RD	HARTLAND	CT		CLOSED	LITCHFIELD COUNTY DISPATCH	Self
22	2009	2.01E+08	No Response	11/25/2009 0:00	11/25/2009 0:00	10 RIVER RD	LISBON	CT		CLOSED	QV DISPATCH	Self
23	2009	2.01E+08	Poynton, Dave	11/25/2009 9:08	11/25/2009 9:08	WHITE OAK AVENU	PLAINVILLE	CT		CLOSED	dispatcher	plainville fire department
24	2009	2.01E+08	No Response	11/25/2009 0:00	11/25/2009 0:00	25 washington court	STAMFORD	CT		CLOSED	Ann Dugan	Petro Oil
25	2009	2.01E+08	Aceto, John	11/25/2009 0:00	11/25/2009 0:00	9 Boxwood Road	NORWALK	CT		CLOSED	ANONYMOUS	Self
26	2009	2.01E+08	Thigpen, Donnell	11/25/2009 0:00	11/25/2009 0:00	305 main street	WALLINGFORD	CT		CLOSED	dispatch	wfd
27	2009	2.01E+08	Emanuelson, Briar	11/25/2009 0:00	11/25/2009 0:00	11 shannon drive	ENFIELD	CT		CLOSED	bob delpozzo	Cashway Oil
28	2009	2.01E+08	Stavola, Rosanne	11/25/2009 12:31	11/25/2009 12:31	408 COLONOL LED\	LEDYARD	CT		CLOSED	SALLY KROPP	KROPP ENVIRONMENTAL
29	2009	2.01E+08	Thigpen, Donnell	11/25/2009 0:00	11/25/2009 0:00	69 detroit ave	MERIDEN	CT		CLOSED	Ast. Finigan	MFD
30	2009	2.01E+08	No Response	11/25/2009 13:21	11/25/2009 13:21	600 CLEVELAND A\	BRIDGEPORT	CT		CLOSED	DICK LEWIS	ENVIROSHIELD
31	2009	2.01E+08	No Response	11/25/2009 0:00	11/25/2009 0:00	Route 44 Joe's Pizza	CANTON	CT		CLOSED	ANONYMOUS	Self
32	2009	2.01E+08	No Response	11/25/2009 14:00	11/25/2009 14:00	ROUTE 9 SOUTH E\	NEW BRITAIN	CT		CLOSED	DISPATCHER	NEW BRITAIN FIRE DEPARTMENT
33	2009	2.01E+08	No Response	11/25/2009 0:00	11/25/2009 0:00	90 post road east	WESTPORT	CT		CLOSED	dispatch	WFD
34	2009	2.01E+08	No Response	11/25/2009 0:00	11/25/2009 0:00	690 post road	WESTPORT	CT		CLOSED	Deborah Rebello	Cumberland Farms

Knowing When To Clean Data

- When Should I Clean?

- Values are missing
- Zeros replace missing values
- Data are missing you know should be there
- Rows or values are duplicated
- Spelling is inconsistent
- Name order is inconsistent
- Date formats are inconsistent
- Units are not specified
- Categories are badly chosen
- Field names are ambiguous
- Provenance is not documented
- Suspicious numbers are present
- Data are too coarse
- Totals differ from published aggregates

News

Buying Power: Massachusetts sta races cost $77 million in 2010

Updated Mar 25, 2019; Posted Feb 20, 2011

By The Republican Newsroom



Associated Press Massachusetts Republican gubernatorial candidate Charles D. Baker, left, independent Timothy Cahill, second left, moderator and political analyst Jon Keller, center, Gov. Deval Patrick, second right, and Green-Rainbow Party candidate Jill Stein, right, pose for photo prior to the start of the first televised debate at the WBZ-TV studios, Tuesday, Sept. 7, 2010, in Boston.(AP Photo/Elizuayehu Tesfaye)

By Maggie Mulvihill, Sarah Favot and Matt Porter | *New England Center for Investigative Reporting*

From political action committees to gubernatorial candidates to county prosecutors, Massachusetts campaign cash spent in 2010 topped $77 million, paying for everything from a county club membership, tuxedo rentals, expensive car leases, makeup artists, cigars, hundreds of floral arrangements and much more, newly released financial reports obtained by the

Dirty Data

Tips

- **Record counts**
  - Always be aware of how many records (rows or lines) make up your data.
    - How many should you have?
    - Does the number make sense
- **Data entry**
  - Lack of consistency or format
  - Missing information
  - Leading and trailing white spaces in data
  - Missing codes



Tips

- **Techniques to find issues**
  - Make a copy of your original data
    - Interview it
  - Get the record layout or data dictionary to make sure you understand it
    - Double check totals and calculations
  - Sort columns to look for anything that is off
    - **In EXCEL:**
      - Look at blanks – what do they mean
      - Do some pivot table “data integrity checks”
  - Look for aberrations such as one year tripling the total of all other years – could that mean a problem



As You Work

- Always work off of a backup copy.
- Keep careful notes of your work.
- Identify issues within the data and write down questions you must answer.
- Never hesitate to ask outside experts for help understanding data (inspectors, engineers, etc.).
- Talk it through with others who have worked with the data



What is Open Refine?

- A web application that runs locally on your computer
- Don't need an Internet connection to clean your data in Refine
- Can read numerous different formats of data
- Also allows for exploration of data – not just cleaning





About OpenRefine

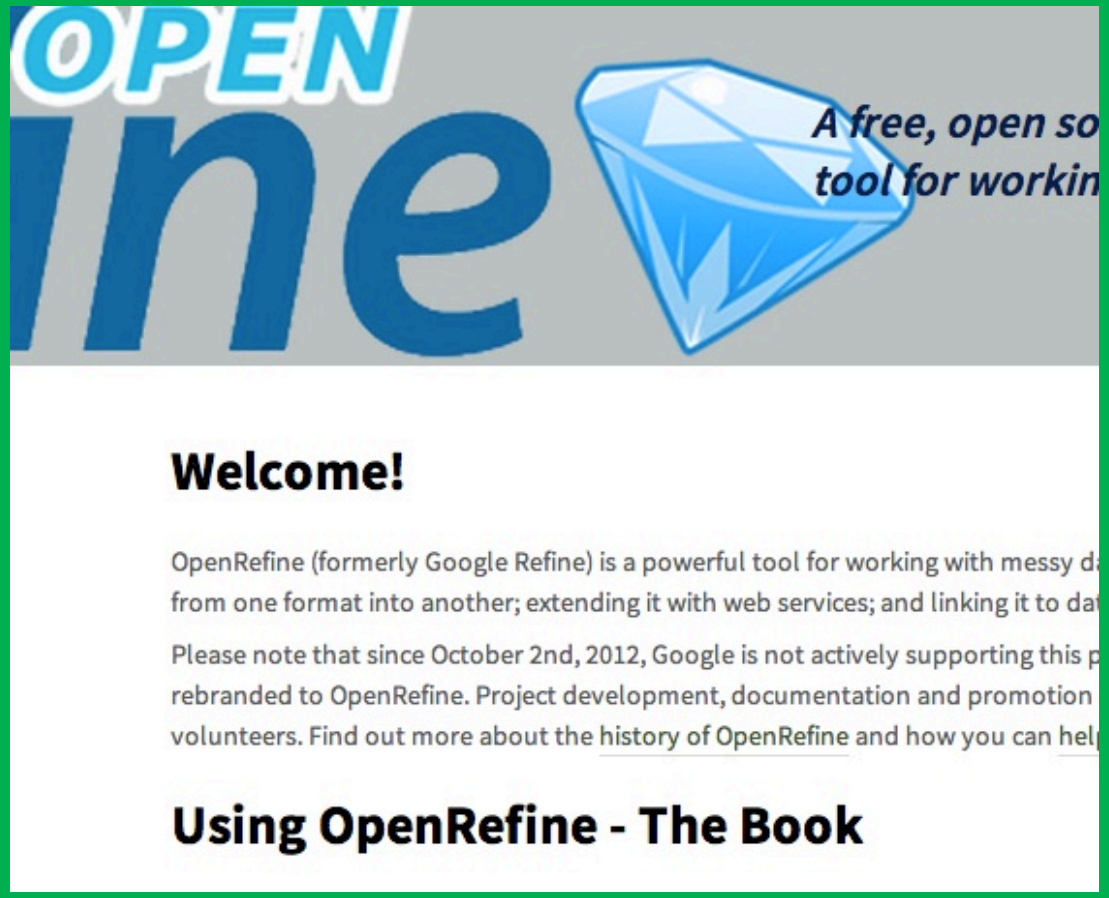
OpenRefine is a power tool for working with messy data. Use it to improve data consistency, link it to data registries like Freebase, augment it with data from other sources, transform it into different formats for other tools to consume, and contribute it to back to the original sources. OpenRefine is not a web service but a desktop app that runs on your own computer, so you can process sensitive data with privacy.

OpenRefine was originally developed as "Freebase Gridworks" by Metaweb Technologies, Inc.. Metaweb was acquired by Google in July 2010 and they renamed the product Google Refine. In October, 2012, the product was renamed OpenRefine as it transitioned to a community supported project.

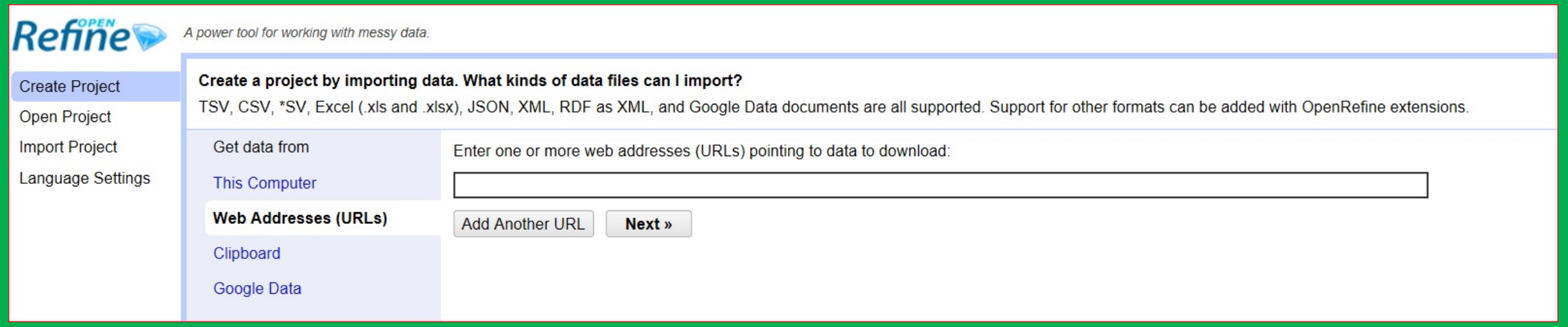
What is Open Refine?

What is Open Refine?

- Cleaning data with:
  - Cell splitting
  - Text facets
  - Clustering
  - Merging cells

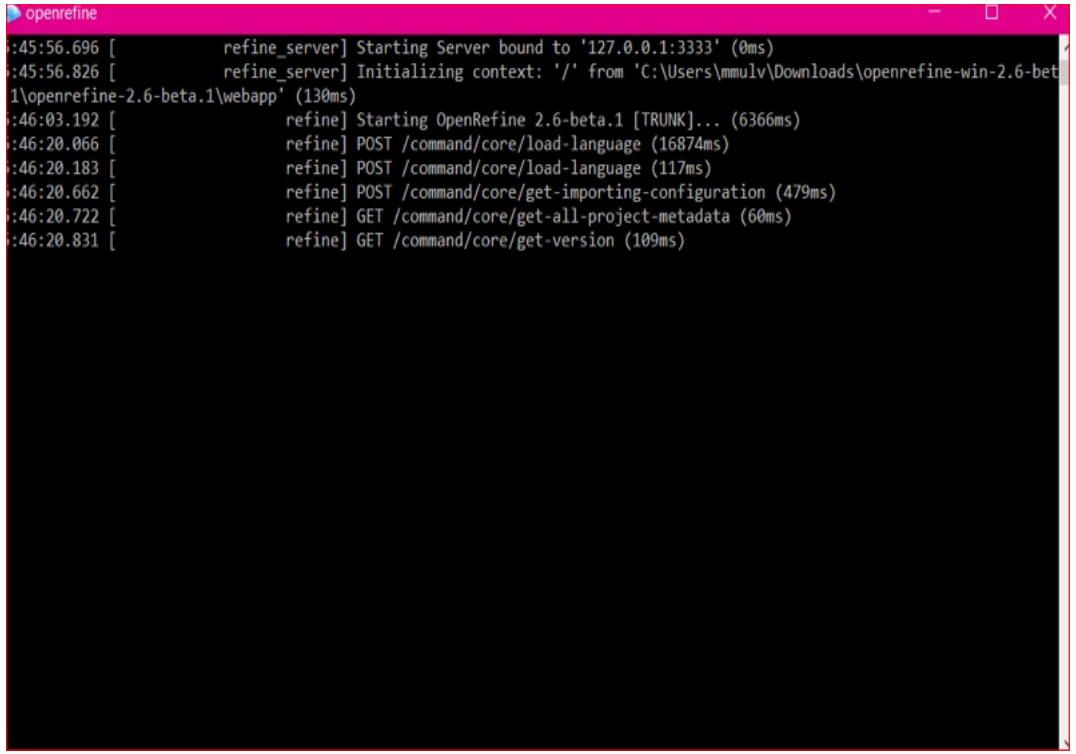


Refine Brings In Data From The Web Or Google



When You First Open the Program . . . . .





When You First Open The Program . . . . .

Projects

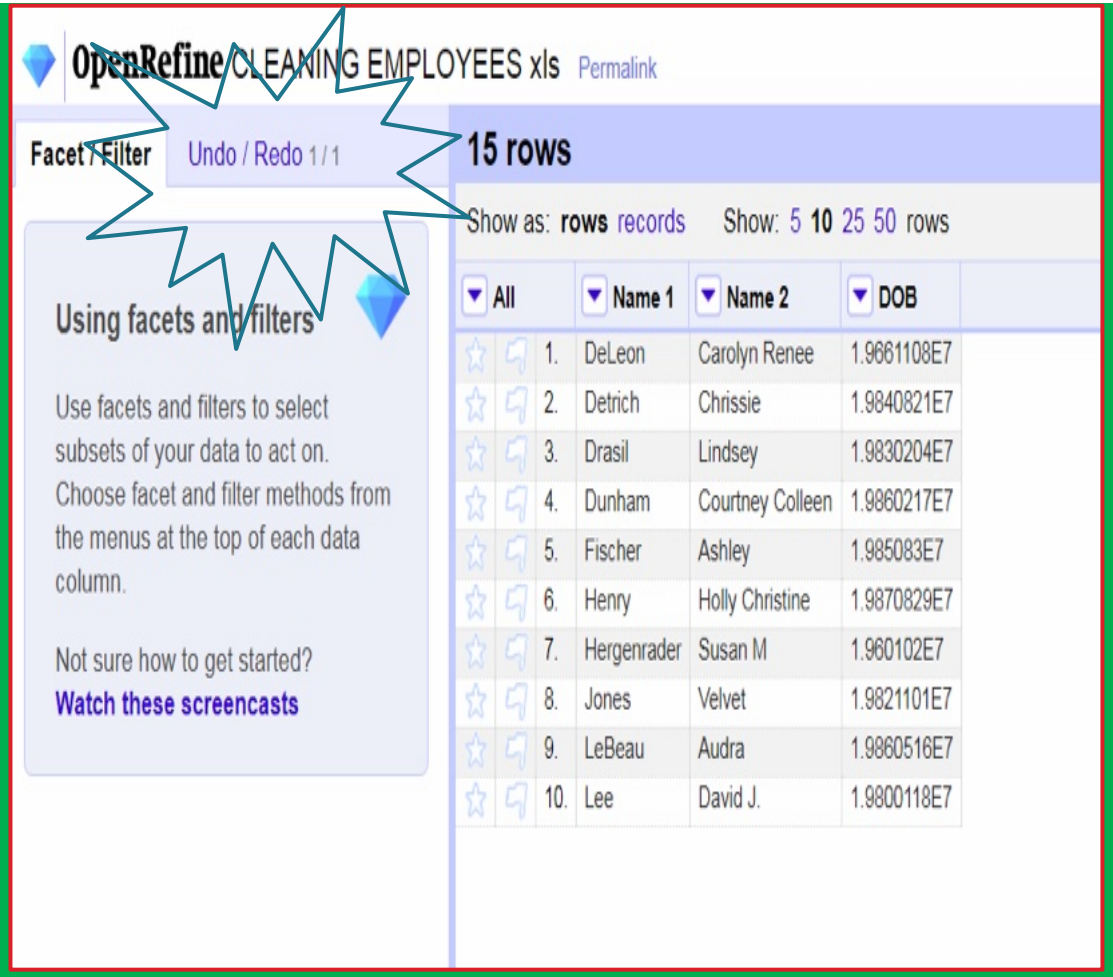
- Create
- Open
- Import
- Will save previous projects for future cleaning

Files

- Works with many different file types
- Can export projects to other formats

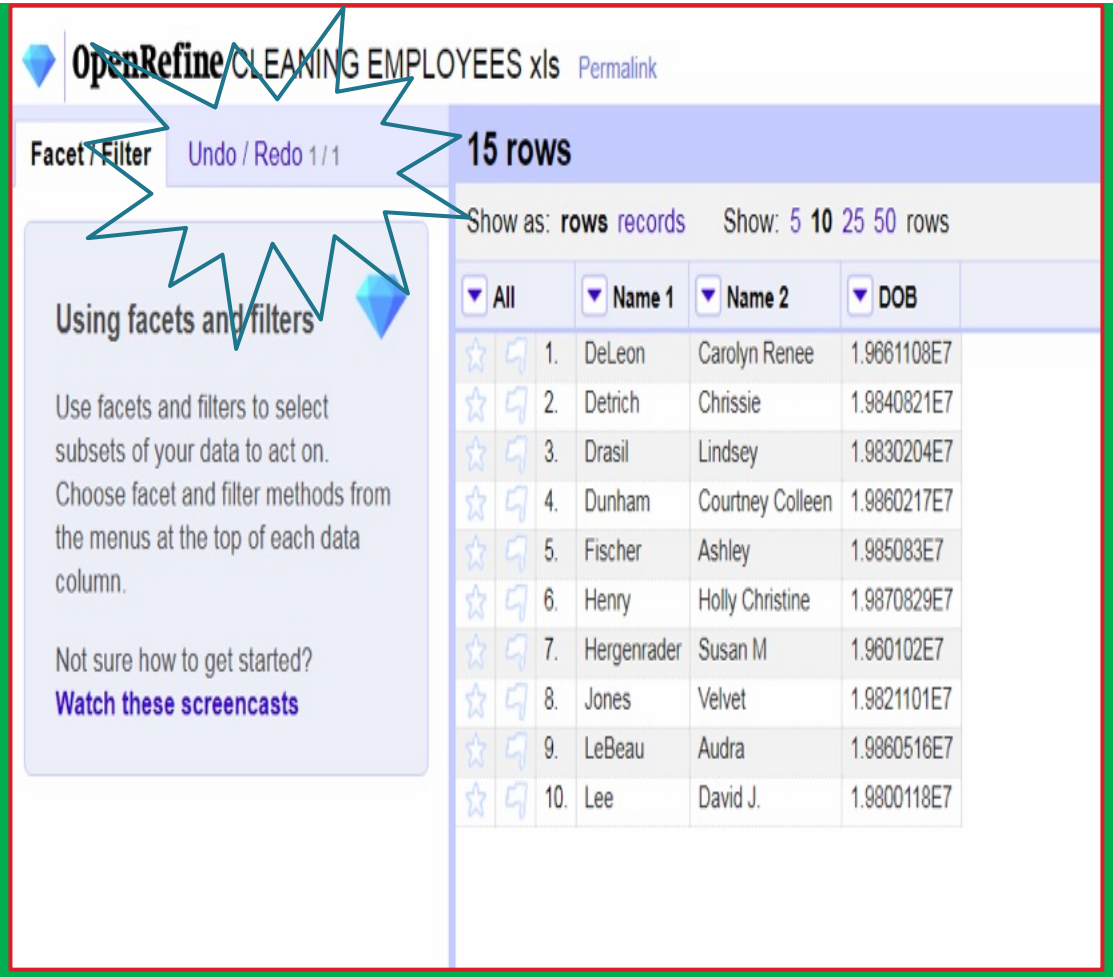
Undo and Redo

- To undo something in Refine and go backwards just click, **Undo**
- To Redo, just click **Redo**



Undo and Redo

- To undo something in Refine and go backwards just click, **Undo**
- To Redo, just click **Redo**



Creating Projects

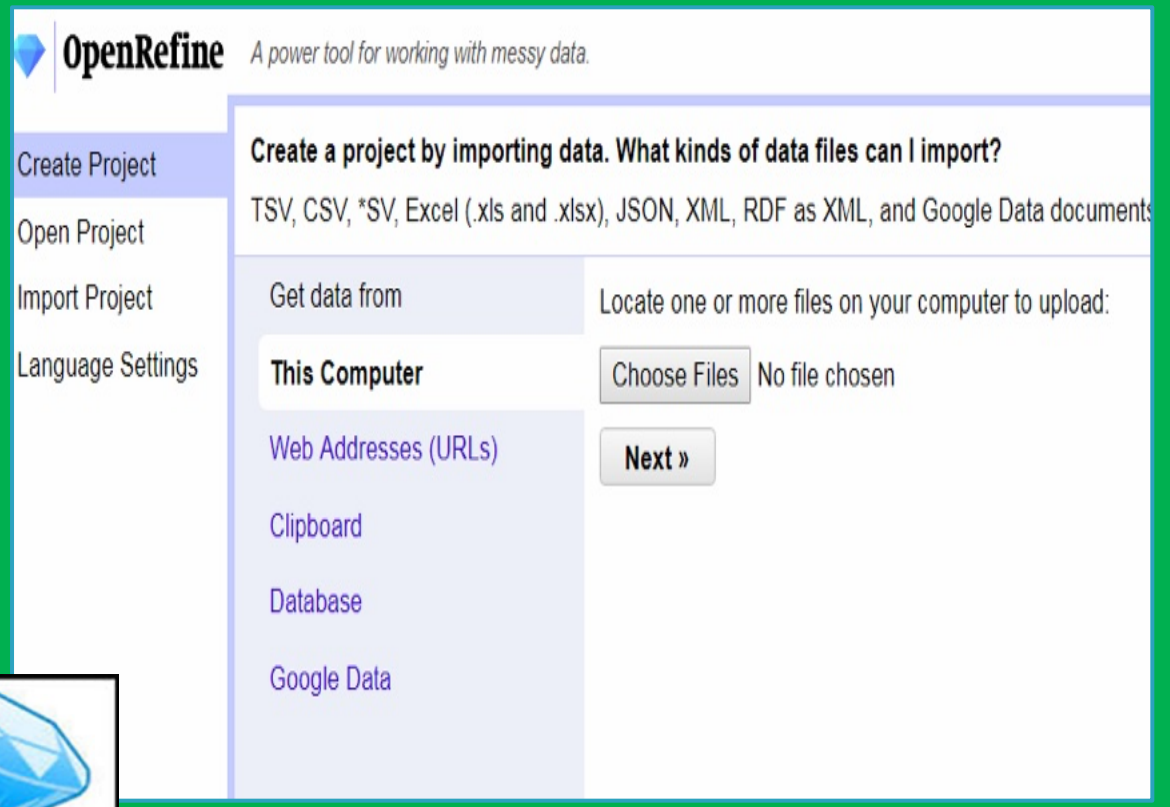


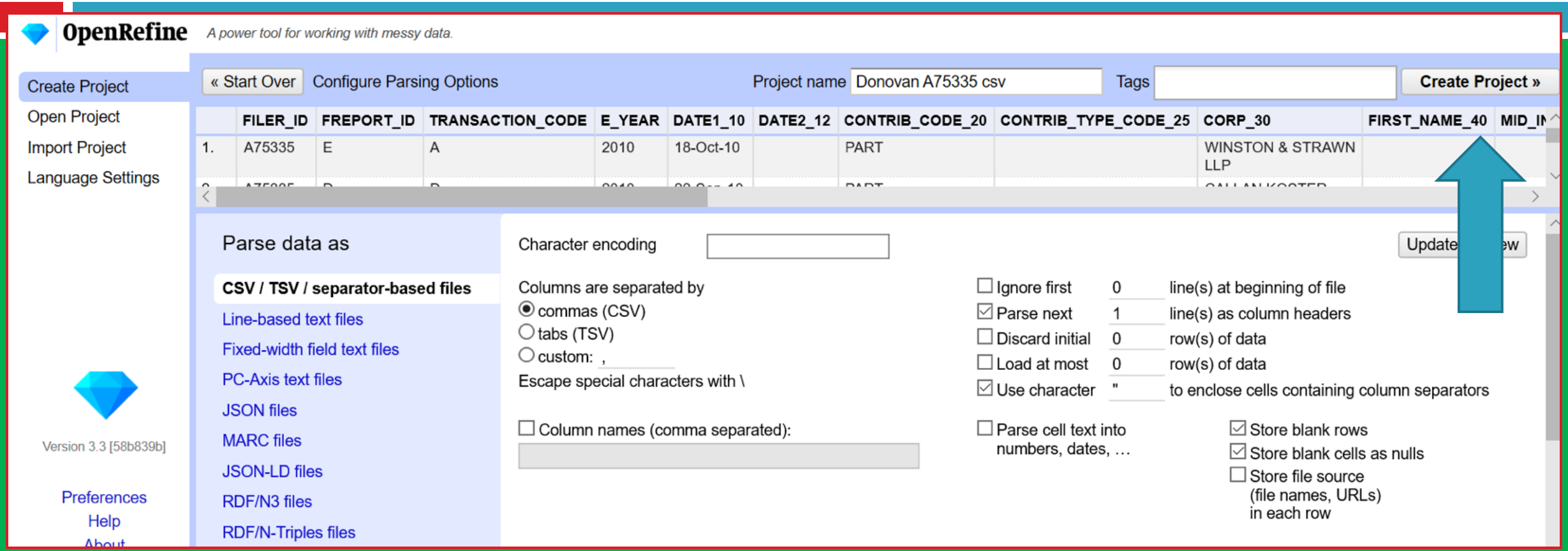
Creating Projects in Refine

- Open the spreadsheet called “Donovan-A75335”
  - Save it to your desktop
- Make a copy and do a “four corners check”
  - How many records do we have?
- Start Open Refine

Importing Data into Open Refine

- Click on **Choose Files**
- Select Donovan from wherever you saved it
- Click **Next**
- Refine will upload the data
- Select **Create Project**




OpenRefine
A power tool for working with messy data.
Create Project
Open Project
Import Project
Language Settings
« Start Over
Configure Parsing Options
Project name
Donovan A75335 csv
Tags
Create Project »
FILER_ID
FREPORT_ID
TRANSACTION_CODE
E_YEAR
DATE1_10
DATE2_12
CONTRIB_CODE_20
CONTRIB_TYPE_CODE_25
CORP_30
FIRST_NAME_40
1.
A75335
E
A
2010
18-Oct-10
PART
WINSTON & STRAWN LLP
Parse data as
CSV / TSV / separator-based files
Line-based text files
Fixed-width field text files
PC-Axis text files
JSON files
MARC files
JSON-LD files
RDF/N3 files
RDF/N-Triples files
Character encoding
Columns are separated by
commas (CSV)
tabs (TSV)
custom: ,
Escape special characters with \
Column names (comma separated):
Ignore first 0 line(s) at beginning of file
Parse next 1 line(s) as column headers
Discard initial 0 row(s) of data
Load at most 0 row(s) of data
Use character " to enclose cells containing column separators
Parse cell text into numbers, dates, …
Store blank rows
Store blank cells as nulls
Store file source (file names, URLs) in each row
Version 3.3 [58b839b]
Preferences
Help
About

Permalink

Open... Export Help

2045 rows

Extensions: Wikidata

Show as: **rows** records Show: 5 **10** 25 50 rows

« first ‹ previous **1 - 10** next › last »

All		FILER_ID	FREPORT_ID	TRANSACTION_	E_YEAR	DATE1_10	DATE2_12	CONTRIB_CODE	CONTRIB_TYPE	CORP_30	FIR
1.		A75335	E	A	2010	18-Oct-10				WINSTON & STRAWN LLP	
2.		A75335	D	D	2010	22-Sep-1				CALLAN KOSTER BRADY & BRENNAN LLP	
3.		A75335	D	A	2010	23-Sep-1				CALLAN KOSTER BRADY & BRENNAN LLP	
4.		A75335	E	A	2010	18-Oct-10				CAMACHO MAURO MULHOLLAND LLP	
5.		A75335	f	A	2010	26-Oct-10		PART		CERTILMAN BALDIN ADLER & HYMAN	
6.		A75335	F	A	2010	29-Oct-10		PART		CROTTY	

Check your data

Did the same number of records come into Refine?

Cleaning Data with FACETS



OpenRefine *A power tool for working with messy data.*

- Create Project
- Open Project
- Import Project
- Language Settings

« Start Over | Configure Parsing Options | Project name: Donovan A75335 csv | Tags: | Create Project »

	FILER_ID	FREPORT_ID	TRANSACTION_CODE	E_YEAR	DATE1_10	DATE2_12	CONTRIB_CODE_20	CONTRIB_TYPE_CODE_25	CORP_30	FIRST_NAME_40	MID_IN
1.	A75335	E	A	2010	18-Oct-10		PART		WINSTON & STRAWN LLP		
2.	A75335	D	D	2010	22-Sep-10		PART		CALLAN KOSTER BRADY & BRENNAN LLP		
3.	A75335	D	A	2010	23-Sep-10		PART		CALLAN KOSTER BRADY & BRENNAN LLP		
4.	A75335	E	A	2010	18-Oct-10		PART		CAMACHO MAURO MULHOLLAND LLP		
5.	A75335	f	A	2010	26-Oct-10		PART		CERTILMAN BALDIN ADLER & HYMAN		
6.	A75335	F	A	2010	29-Oct-10		PART		CROTTY SALAND LLP		

Version 3.2-beta [8d89a2a]

Preferences
Help
About

Parse data as

- **CSV / TSV / separator-based files**
- Line-based text files
- Fixed-width field text files
- PC-Axis text files
- JSON files
- MARC files
- JSON-LD files
- RDF/N3 files

Character encoding: | Update Preview

Columns are separated by
- (•) commas (CSV)
- ( ) tabs (TSV)
- ( ) custom: ,

Escape special characters with \

[ ] Column names (comma separated):

[ ] Ignore first 0 line(s) at beginning of file
[x] Parse next 1 line(s) as column headers
[ ] Discard initial 0 row(s) of data
[ ] Load at most 0 row(s) of data
[x] Use character " to enclose cells containing column separators

[ ] Parse cell text into numbers, dates, ...

[x] Store blank rows
[x] Store blank cells as nulls
[ ] Store file source (file names, URLs) in each row

Waiting for tools.wmflabs.org...

Refine Donovan A75335 csv Permalink

Facet / Filter | Undo / Redo 0

Using facets and filters

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started?
Watch these screencasts

2045 rows

Extensions: undefined

Show as: **rows** records Show: 5 **10** 25 50 rows « first ‹ previous **1 - 10** next › last »

All		FILER_ID	FREPORT_ID	TRANSACTION_	E_YEAR	DATE1_10	DATE2_12	CONTRIB_CODE	CONTRIB_TYPE	CORP_30	FIRST_NAME_4(	MID_INIT_42	LAST_NAM
1.	A75335	E	A	2010	18-Oct-10		PART		WINSTON & STRAWN LLP				
2.	A75335	D	D	2010	22-Sep-10		PART		CALLAN KOSTER BRADY & BRENNAN LLP				
3.	A75335	D	A	2010	23-Sep-10		PART		CALLAN KOSTER BRADY & BRENNAN LLP				
4.	A75335	E	A	2010	18-Oct-10		PART		CAMACHO MAURO MULHOLLAND LLP				
5.	A75335	f	A	2010	26-Oct-10		PART		CERTILMAN BALDIN ADLER & HYMAN				
6.	A75335	F	A	2010	29-Oct-10		PART		CROTTY SALAND LLP				
7.	A75335	D	A	2010	24-Sep-10		PART		DAVIDOFF MALITO & HUTCHER LLP				
8.	A75335	E	A	2010	18-Oct-10		PART		DELOITTE SERVICES LP				
9.	A75335	K	A	2010	22-Jun-10		PART		DR. KAREN DIBONA				
10.	A75335	D	A	2010	17-Sep-10		PART		DURKIN & DURKIN LLP				

Refine: Using FACETS

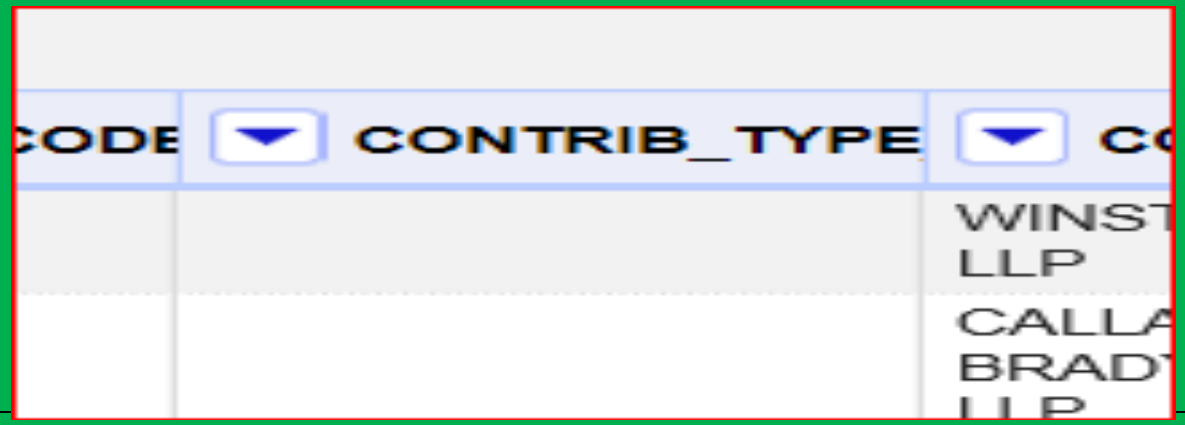


- Your row count will be at the top
- Shows up to 50 rows in your window
- Can scroll through the data by clicking NEXT
- Change your row view to 1000



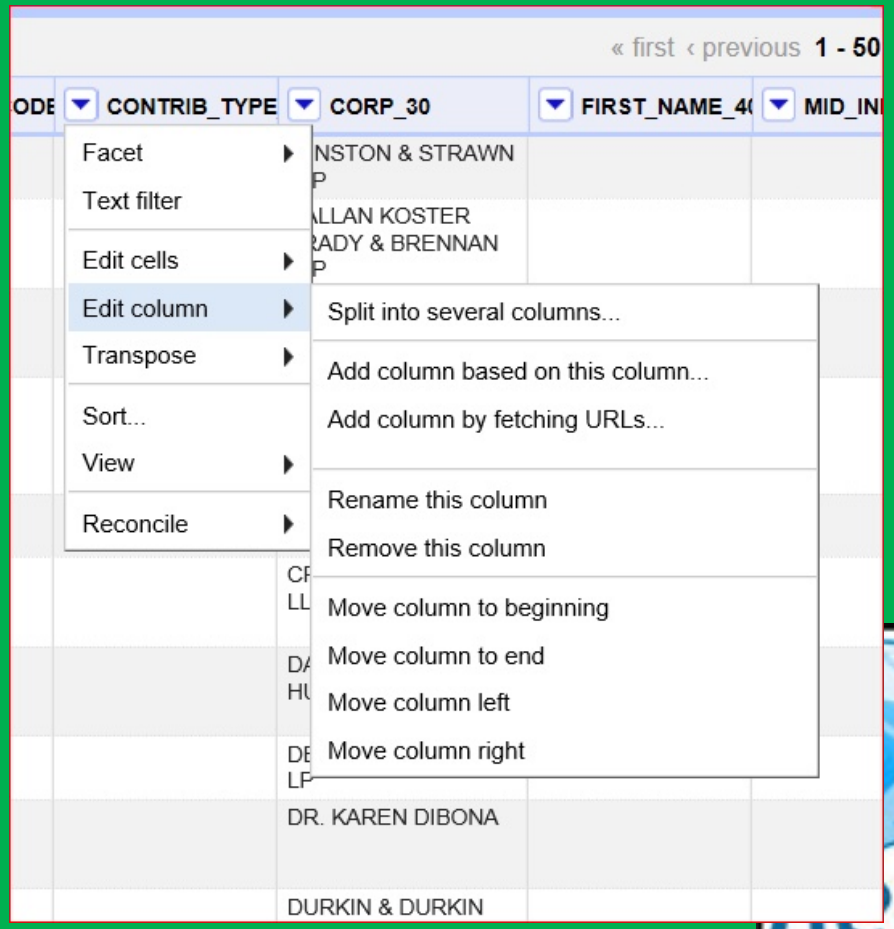
Refine: Removing Columns

- Let's remove columns we don't need
  - Go to CONTRIB_TYPE
- This is empty so let's get rid of it



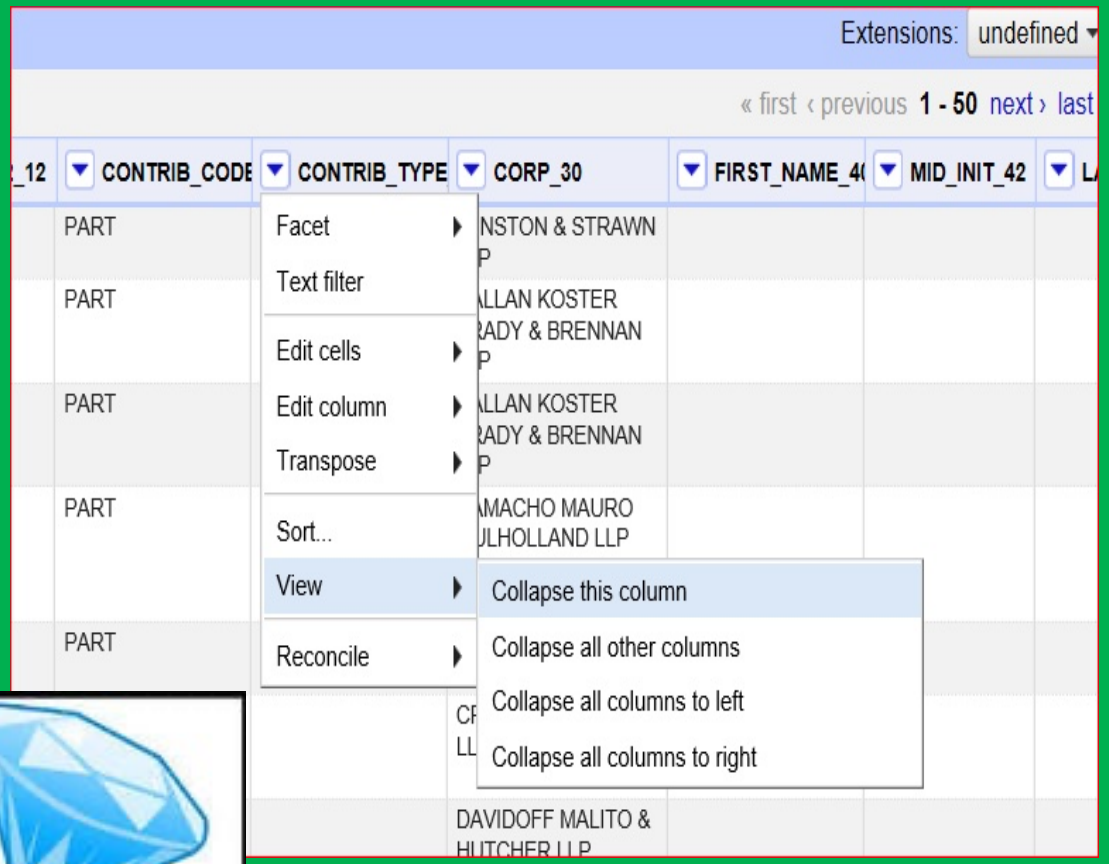
Refine: Removing Columns

- Go to the column
- Click on the down arrow
- Go to EDIT COLUMN
- Go to REMOVE THIS COLUMN



Refine: Collapsing Columns

- Collapse columns you don't need to see
- Click the down arrow on the column you want to collapse
- Click VIEW
- Click COLLAPSE COLUMN



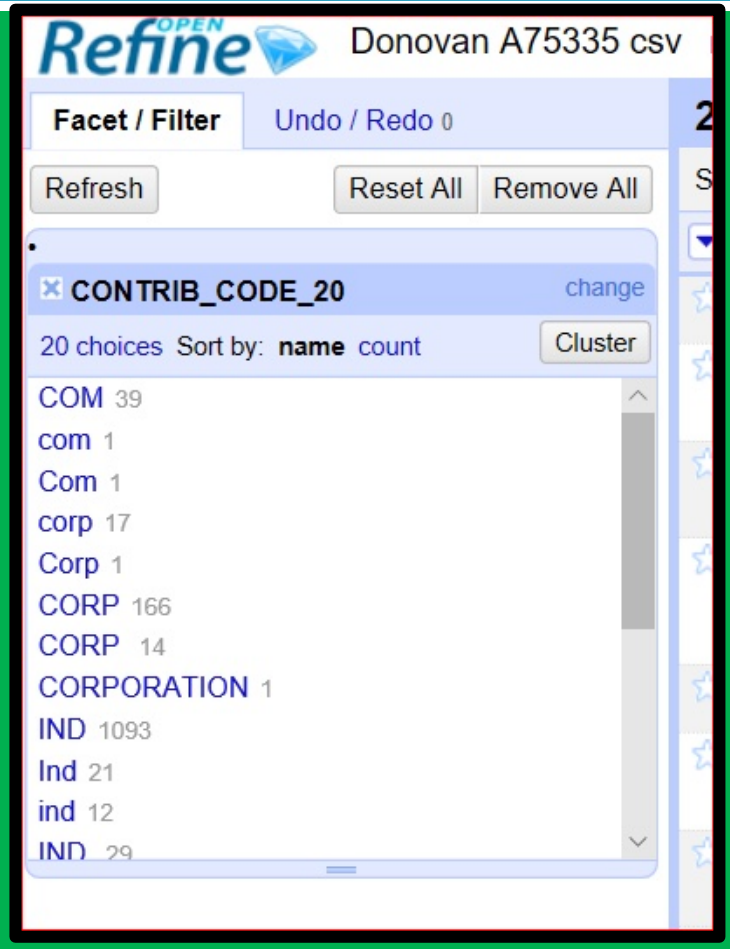
Refine: Cleaning with TEXT FACETS

- One of key tools in Refine
- Akin to a Pivot Table in EXCEL
- In Refine, a FACET is a category

- there are different types of FACETS
- Text, numbers

Refine: Cleaning with TEXT FACETS

- Click on CONTRIB_CODE
  - Click on down arrow
    - Click on FACET
  - Click on TEXT FACET



Refine: Cleaning with FACETS

- You'll see there are many different ways the text is listed for CONTRIB_CODE
- We have COM and Com
- We have LLC and llc
- Let's clean with COMMON TRANSFORMATIONS

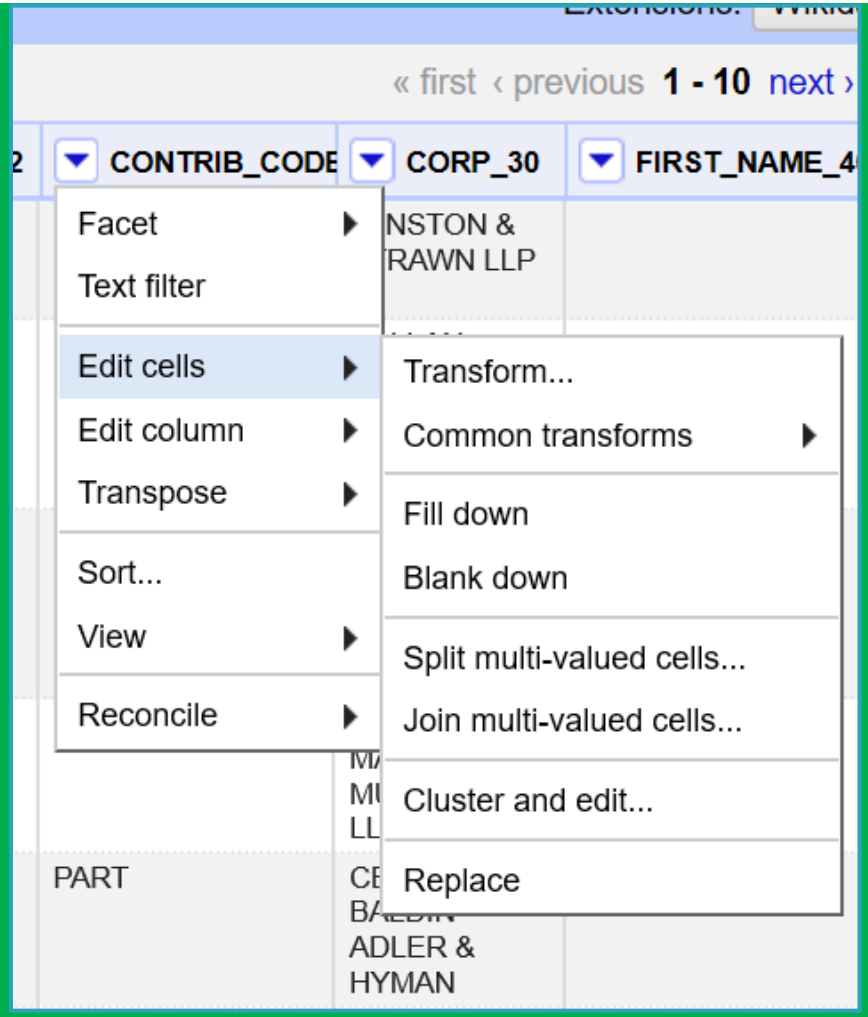
Refine: Cleaning with FACETS

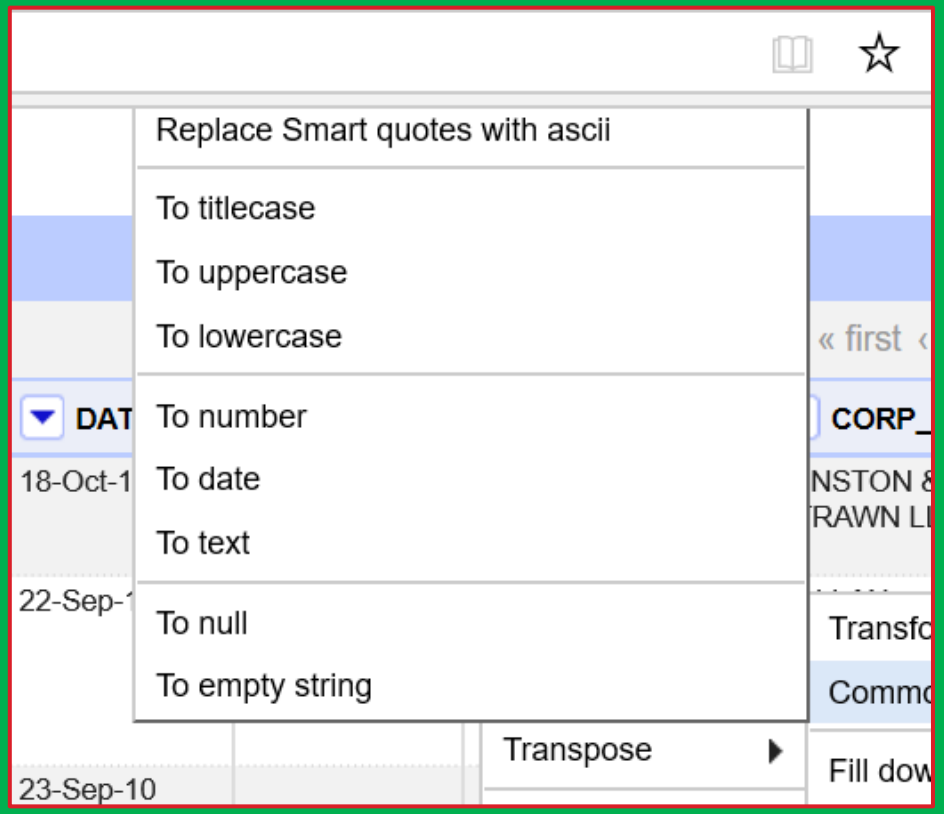


Refine: Cleaning with FACETS

- Go to *EDIT CELLS*
- Go to *COMMON TRANSFORMATIONS*
- Go to *UPPERCASE*

Refine: Cleaning with FACETS





Refine: Cleaning with COMMON TRANSFORMATIONS

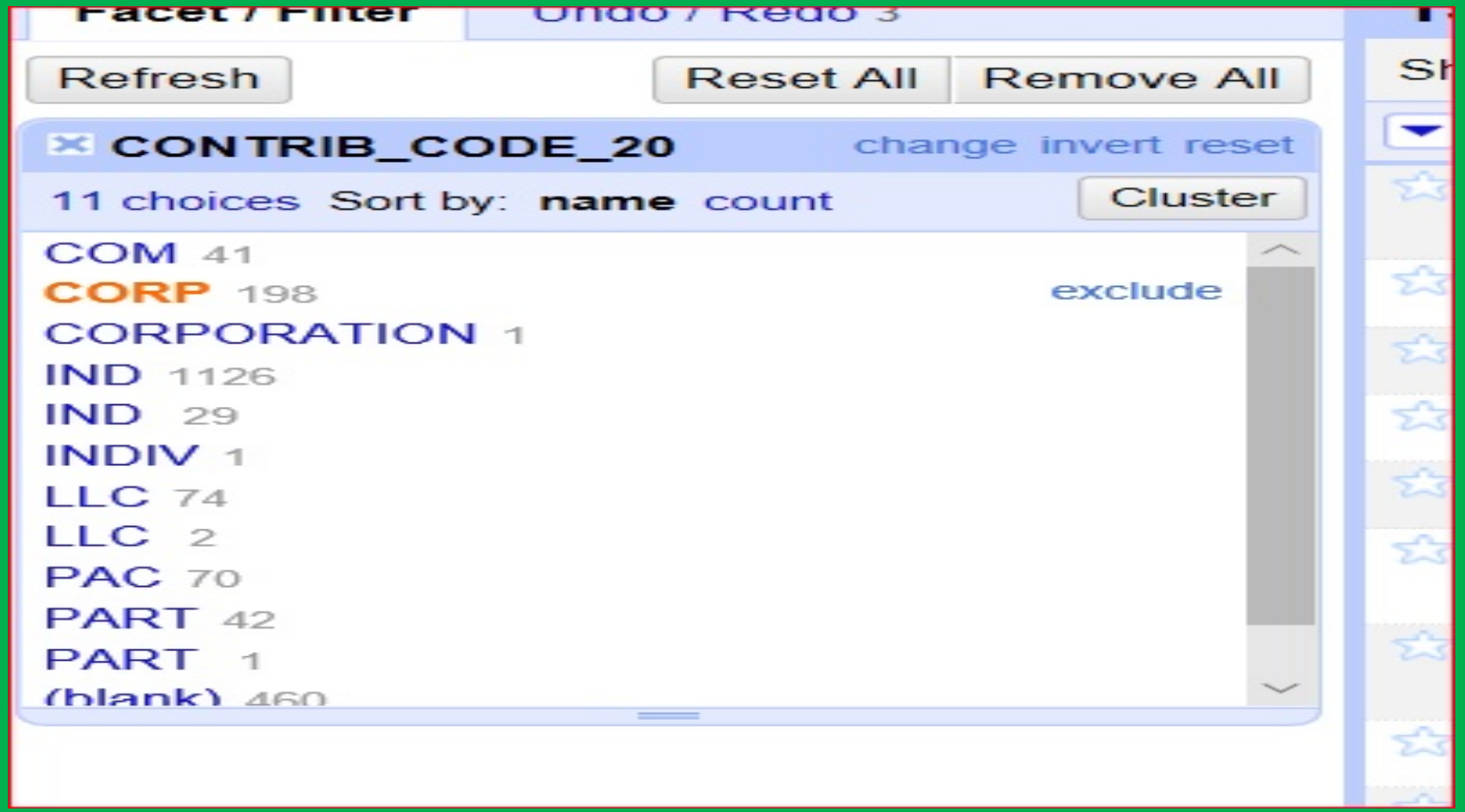
- We see Refine has cleaned up the problems with case in the text



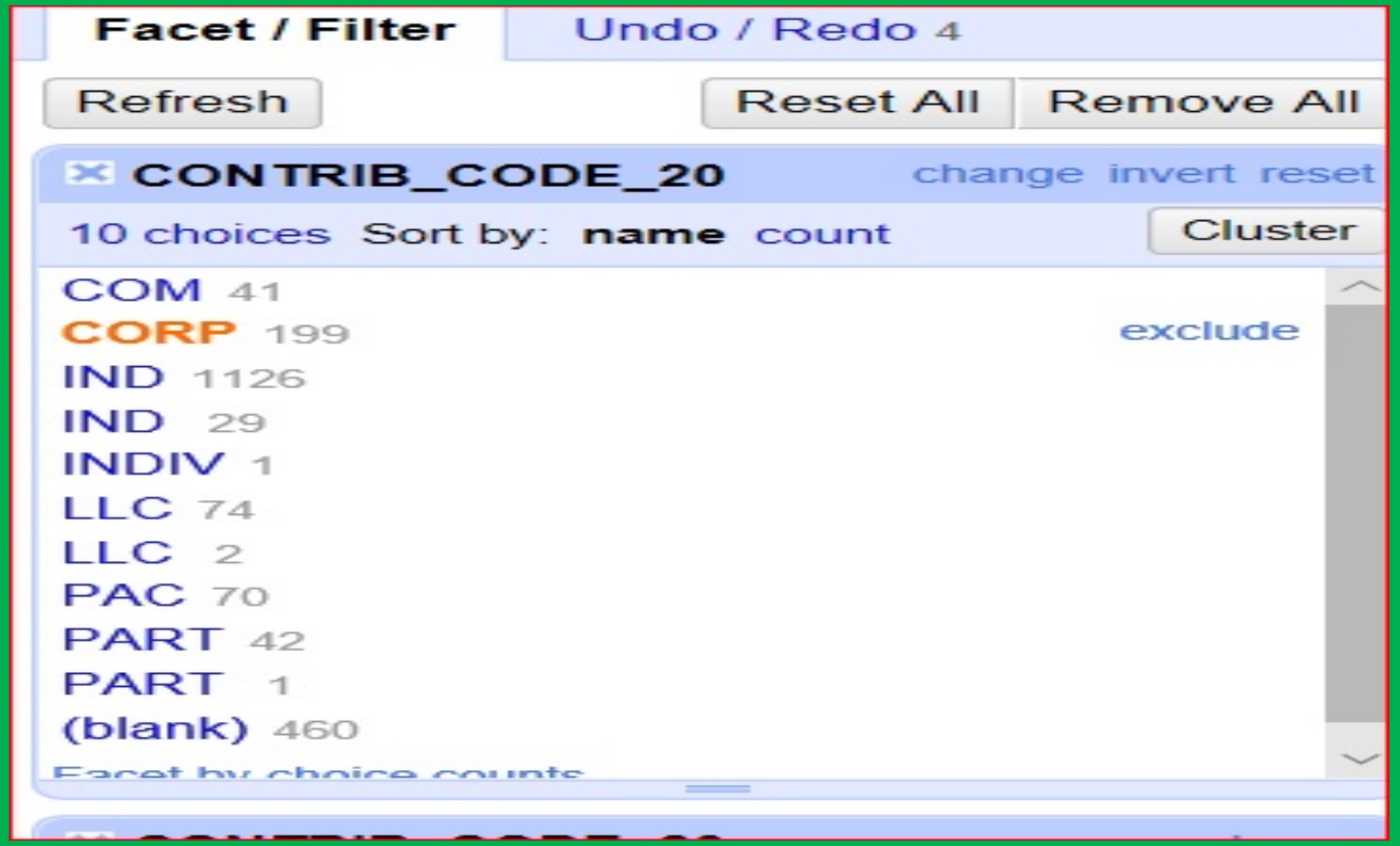
Refine: Cleaning with Common Transformations

- In the FACET/FILTER box, click on something you want to change
  - If you hover your mouse over it you'll see the option to "edit" it
    - Now change "corp" to "corporation"
      - Change "ind" to "individual"
  - Continue till you have all your names the way you want them

Refine: Cleaning with Common Transformations

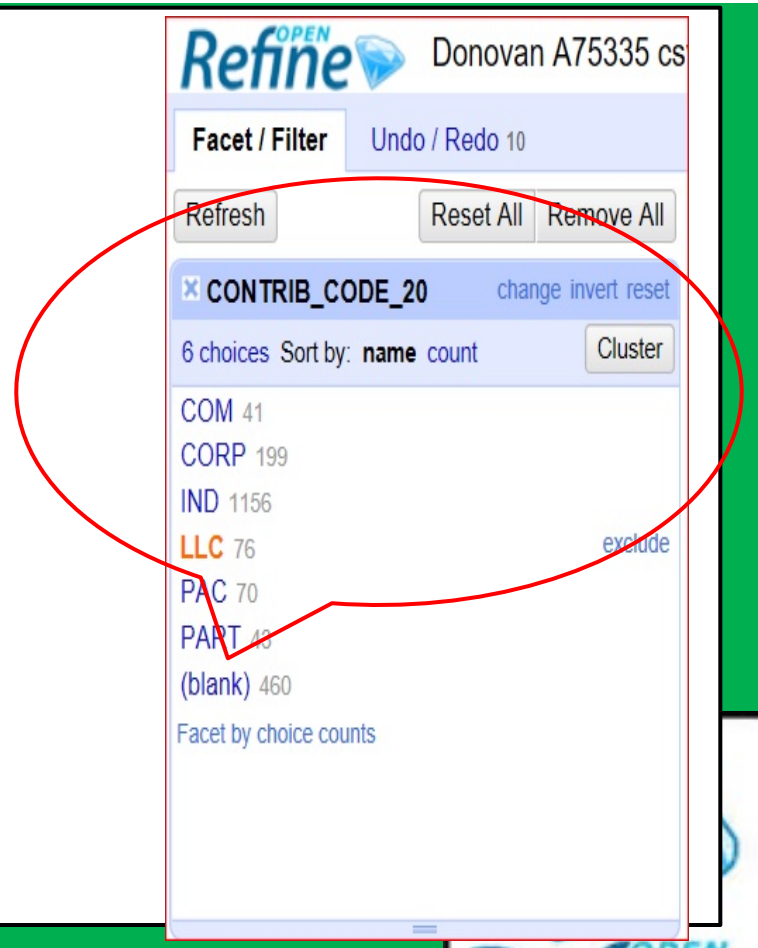


Refine: Cleaning with Common Transformations



Refine: Cleaning with FACETS

- Now let's SORT the CONTRIBUT_CODE
- SORT by COUNT
- What do you get?
- Is top CONTRIB an IND?
- You should get 1156



Refine: Duplicating Columns

- Good practice to make duplicate columns so you don't lose the original
- Refine always allows you to go back and UNDO your work

CLUSTERING

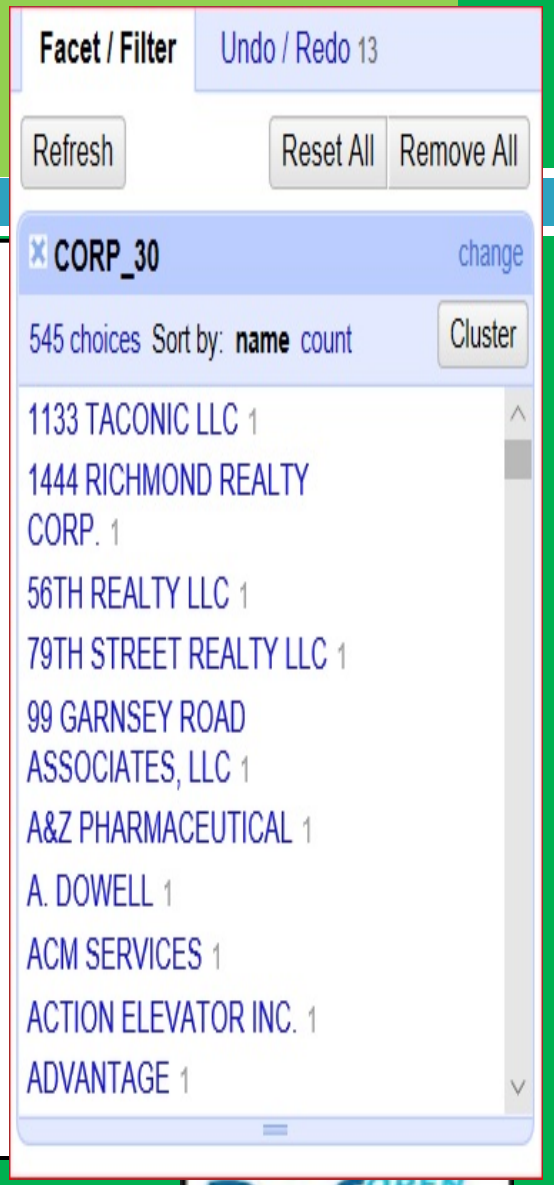


Refine: CLUSTERS

- Let's look at column CORP_30 to see how dirty it is
  - Click on the column
  - Click the down arrow
    - Click on FACET
  - Click on TEXT FACET

Refine: CLUSTERS

- A lot of different corporations gave to Donovan
- We don't want to clean this all up manually
- CLUSTERING the data allows us to do it much faster



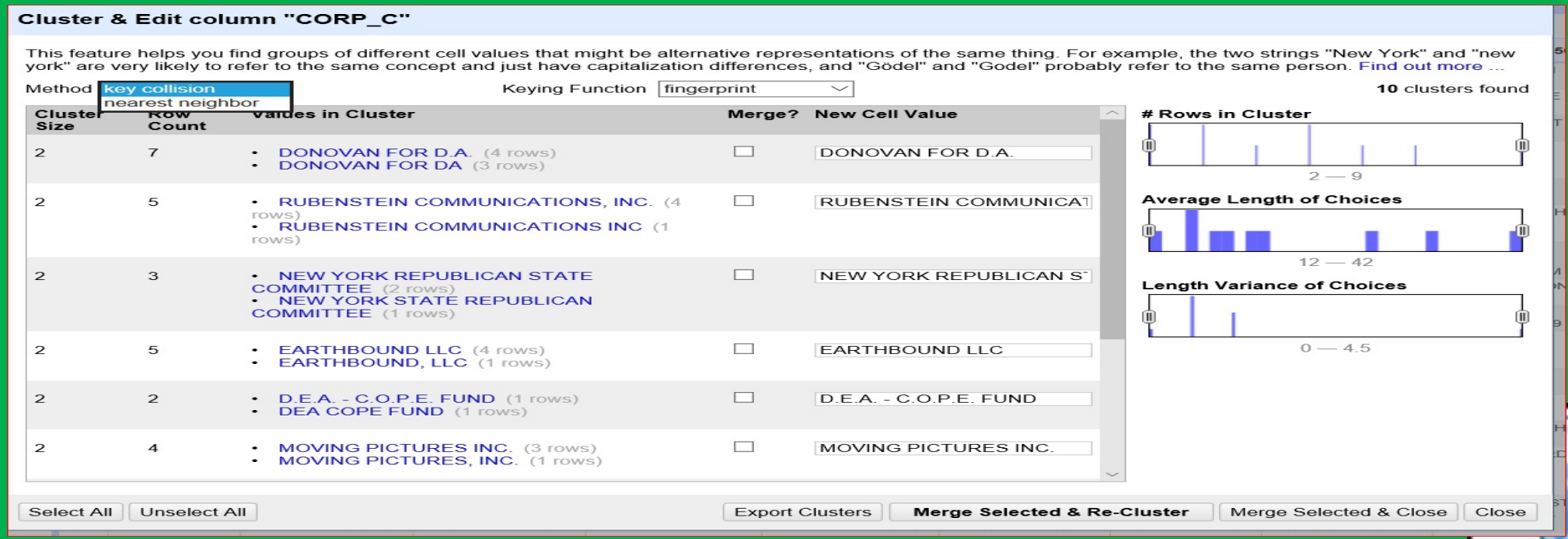
Refine: CLUSTERING

- If we do a TEXT FACET
  on our new column,
  we see we also have
  545 categories



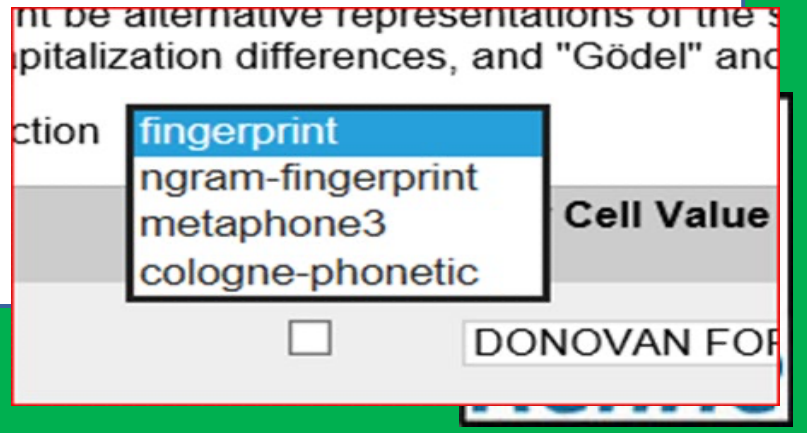
Refine: CLUSTERING

- Refine has two CLUSTERING methods



Refine: CLUSTERING

- Refine has four KEYING FUNCTIONS
- These are algorithms that try to group the data by guessing what goes together, from being very liberal in the guess to very conservative



Cluster & Edit column "CORP_C"

This feature helps you find groups of different cell values that might be alternative representations of the same thing. For example, the two strings "New York" and "new york" are very likely to refer to the same concept and just have capitalization differences, and "Gödel" and "Godel" probably refer to the same person. Find out more ...

Method key collision Keying Function fingerprint **10** clusters found

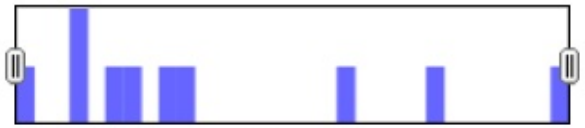
Cluster Size	Row Count	Values in Cluster	Merge?	New Cell Value
2	7	• DONOVAN FOR D.A. (4 rows) • DONOVAN FOR DA (3 rows)	☑	DONOVAN FOR DA
2	5	• RUBENSTEIN COMMUNICATIONS, INC. (4 rows) • RUBENSTEIN COMMUNICATIONS INC (1 rows)	☐	RUBENSTEIN COMMUNICAT
2	3	• NEW YORK REPUBLICAN STATE COMMITTEE (2 rows) • NEW YORK STATE REPUBLICAN COMMITTEE (1 rows)	☐	NEW YORK REPUBLICAN S
2	5	• EARTHBOUND LLC (4 rows) • EARTHBOUND, LLC (1 rows)	☐	EARTHBOUND LLC
2	2	• D.E.A. - C.O.P.E. FUND (1 rows) • DEA COPE FUND (1 rows)	☐	D.E.A. - C.O.P.E. FUND
2	4	• MOVING PICTURES INC. (3 rows) • MOVING PICTURES, INC. (1 rows)	☐	MOVING PICTURES INC.

 # Rows in Cluster



2 — 9

Average Length of Choices



12 — 42

Length Variance of Choices



0 — 4.5

Select All | Unselect All | Export Clusters | **Merge Selected & Re-Cluster** | Merge Selected & Close | Close

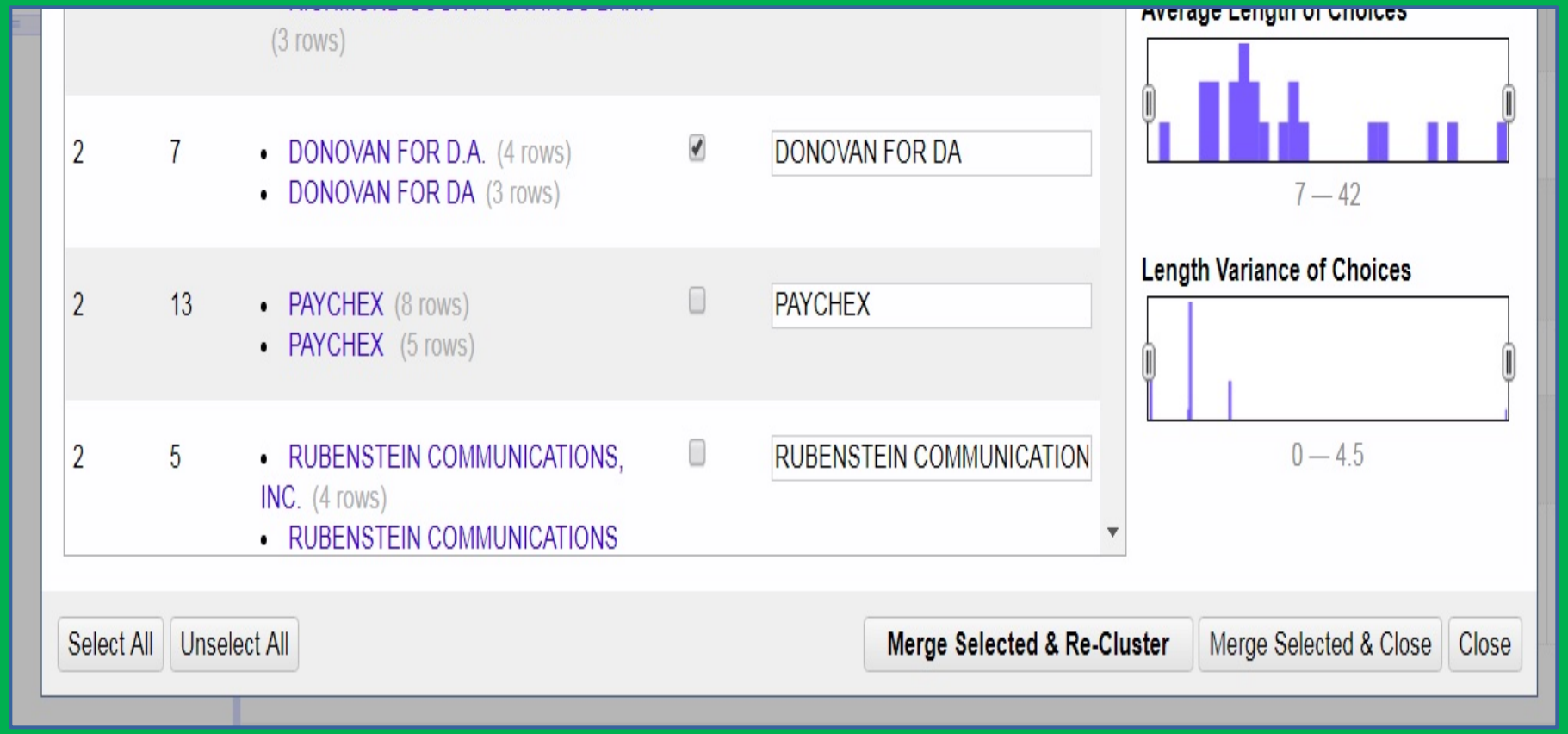
Refine: CLUSTERING

- Check the MERGE box
- Check MERGE SELECTED AND RECLUSTER
- Refine will CLUSTER the 7 rows into one name

Refine: CLUSTERING

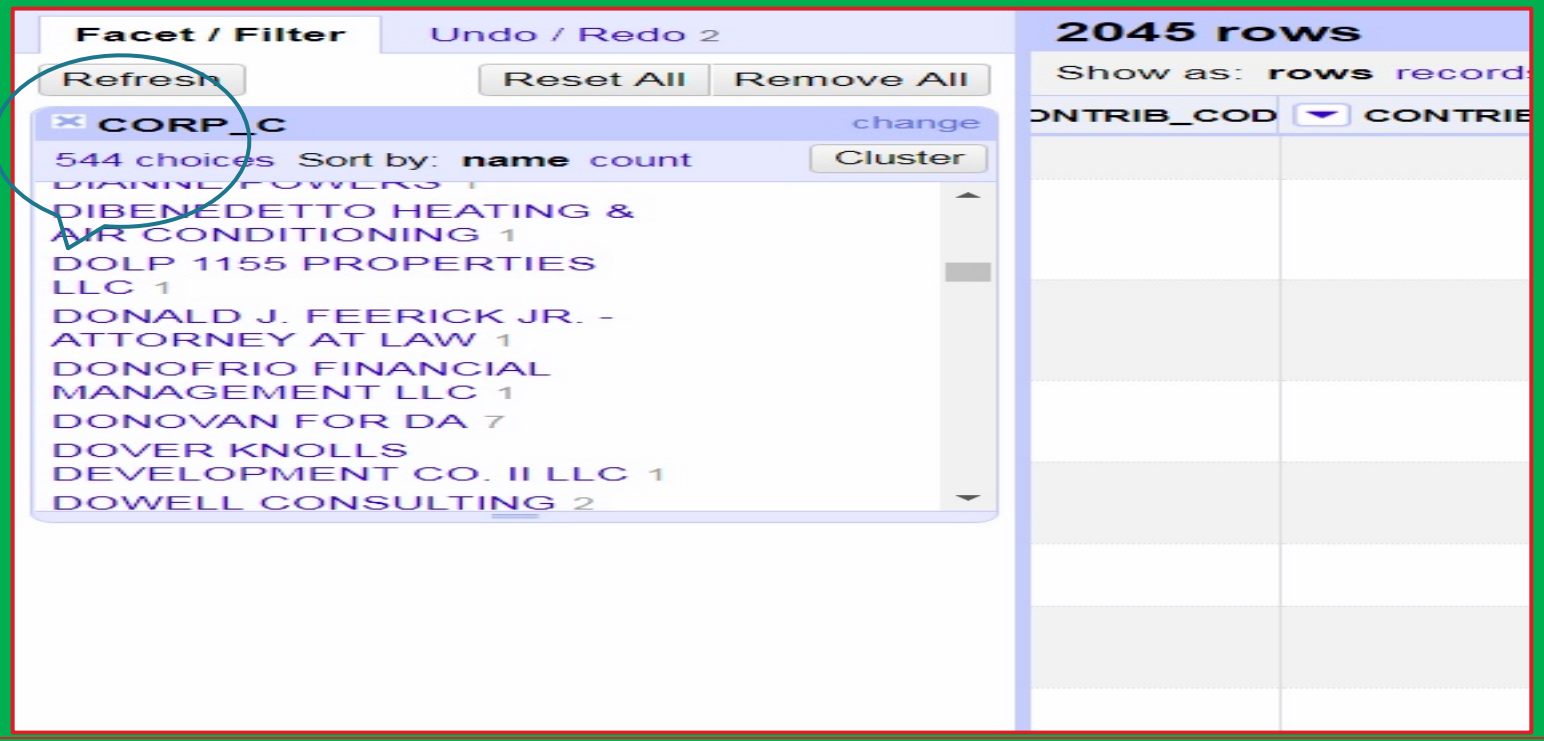
- Check the MERGE box
- Check MERGE SELECTED AND RECLUSTER
- Refine will CLUSTER the 7 rows into one name

CLUSTERING


(3 rows)
2
7
DONOVAN FOR D.A. (4 rows)
DONOVAN FOR DA (3 rows)
DONOVAN FOR DA
2
13
PAYCHEX (8 rows)
PAYCHEX (5 rows)
PAYCHEX
2
5
RUBENSTEIN COMMUNICATIONS, INC. (4 rows)
RUBENSTEIN COMMUNICATIONS
RUBENSTEIN COMMUNICATION
Average Length of Choices
7 — 42
Length Variance of Choices
0 — 4.5
Select All
Unselect All
Merge Selected & Re-Cluster
Merge Selected & Close
Close

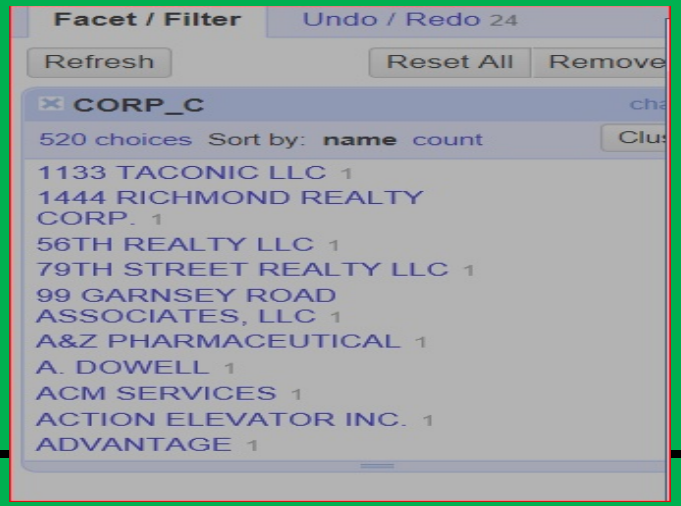
Refine: CLUSTERING with FINGERPRINT

- Your clustering will bring your choices down to 544



Refine: CLUSTERING

- Continue to go through the CLUSTERING options with the FINGERPRINT keying function
- You'll see your CLUSTERING is reducing the rows needing cleaning



Refine: CLUSTERING with N-gram

The n-gram fingerprint method is similar to the fingerprint method described above but instead of using whitespace separated tokens, it uses n-grams, where the n (or the size in chars of the token) can be specified by the user. Here is what it does (or check the code if you're curious about the details, again the order is significant):

•change all characters to their lowercase representation

Refine: CLUSTERING with N-gram

• remove all punctuation, whitespace, and control characters
•obtain all the string n-grams
•sort the n-grams and remove duplicates
•join the sorted n-grams back together
•normalize extended western characters to their ASCII representation



Cluster & Edit column "CORP_C"

This feature helps you find groups of different cell values that might be alternative representations of the same thing. For example, the two strings "New York" and "new york" are very likely to refer to the same concept and just have capitalization differences, and "Gödel" and "Godel" probably refer to the same person. Find out more ...

Method key collision | Keying Function ngram-fingerprint | Ngram Size 2 | **3** clusters found

Cluster Size	Row Count	Values in Cluster	Merge?	New Cell Value
2	7	• SUM(DESIGN) (5 rows) • SUM DESIGN (2 rows)	☐	SUM(DESIGN)
2	6	• GRASSROOTS INITIATIVE (5 rows) • GRASS ROOTS INITIATIVE (1 rows)	☐	GRASSROOTS INITIATIVE
2	6	• LEXIS NEXIS (5 rows) • LEXISNEXIS (1 rows)	☐	LEXIS NEXIS

 # Rows in Cluster



6 — 7

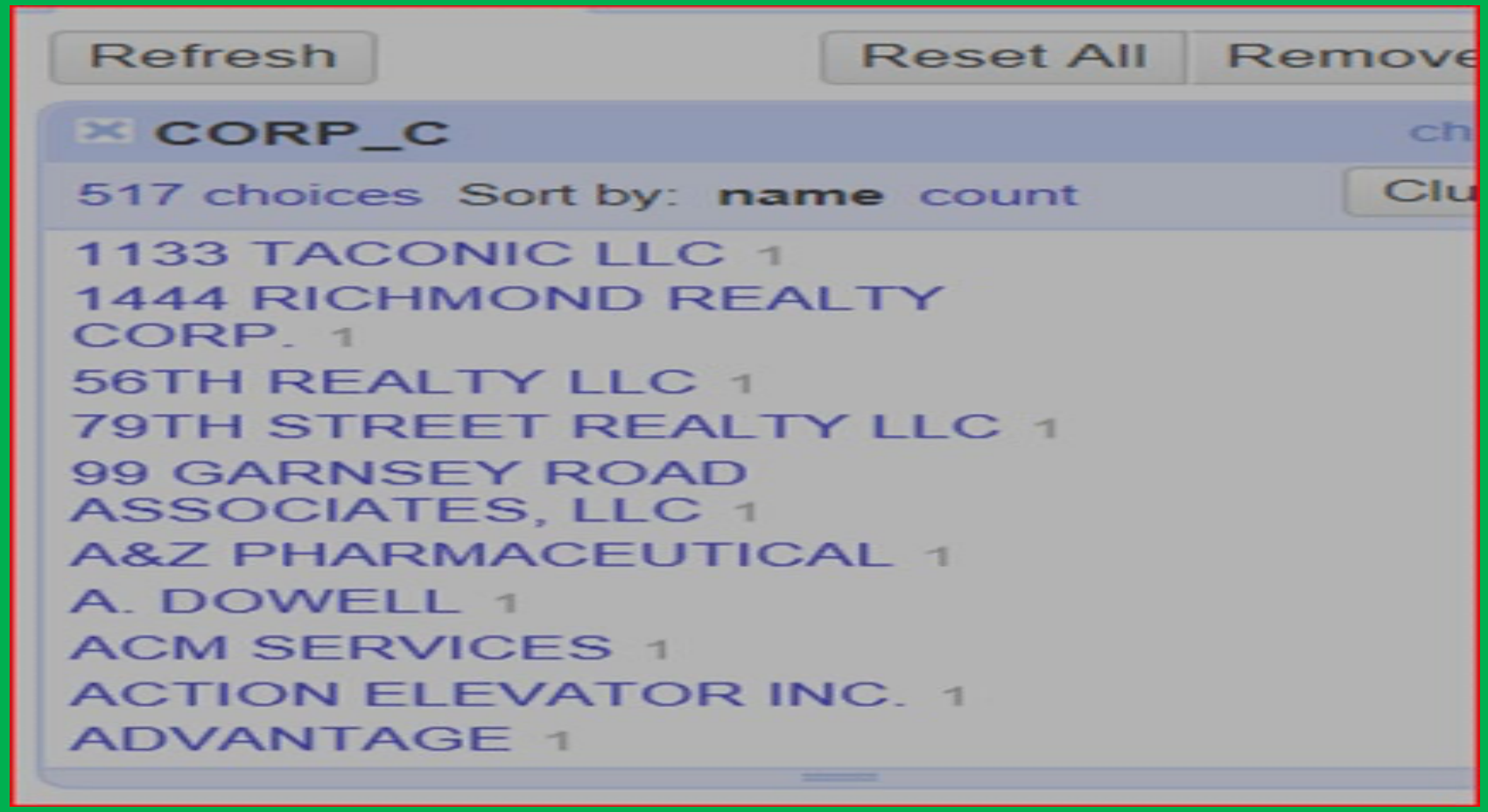
Average Length of Choices



10 — 22

Select All | Unselect All | Export Clusters | **Merge Selected & Re-Cluster** | Merge Selected & Close | Close

CLUSTERING with N-gram



Cluster & Edit column "CORP_C"

This feature helps you find groups of different cell values that might be alternative representations of the same thing. For example, the two strings "New York" and "new york" are very likely to refer to the same concept and just have capitalization differences, and "Gödel" and "Godel" probably refer to the same person. Find out more ...

Method key collision Keying Function metaphone3 **28** clusters found

Cluster Size	Row Count	Values in Cluster	Merge?	New Cell Value
6	6	• STATEN ISLAND CHIROPRACTIC ASSOCIATES PLLC (1 rows) • STATEN ISLAND HEART IMAGING PC (1 rows) • STATEN ISLAND HEART, P.C. (1 rows) • STATEN ISLAND PULMONARY ASSOCIATES (1 rows) • STATEN ISLAND PULMONARY ASSOCIATES P.C. (1 rows) • STATEN ISLAND TERMINAL LLC (1 rows)	☐	STATEN ISLAND CHIROPRA
3	10	• RICHMOND COUNTY SAVINGS BANK (8 rows) • RICHMOND COUNTY SAVINGS (1 rows) • RICHMOND COUNTY SAVINGS BANKI (1 rows)	☐	RICHMOND COUNTY SAVIN
3	3	• GREENBERG TRAURIG P.A. PAC (1 rows) • GREENBERG TRAURIG PAC (1 rows) • GREENBERG TRAURIG, LLP (1 rows)	☐	GREENBERG TRAURIG P.A.
3	4	• FIRST CLASS COMMERCIAL CLEANING SERVICES (2 rows) • FIRST CLASS COMMERCIAL (1 rows)	☐	FIRST CLASS COMMERCIAL

 # Choices in Cluster



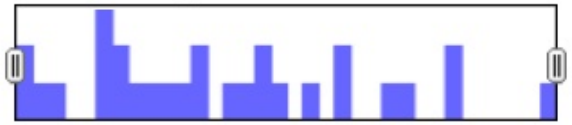
2 — 6

 # Rows in Cluster



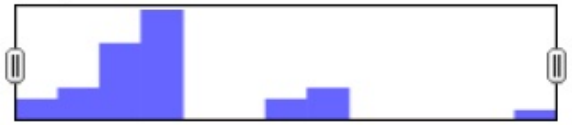
2 — 36

Average Length of Choices



13 — 46

Length Variance of Choices



0 — 12

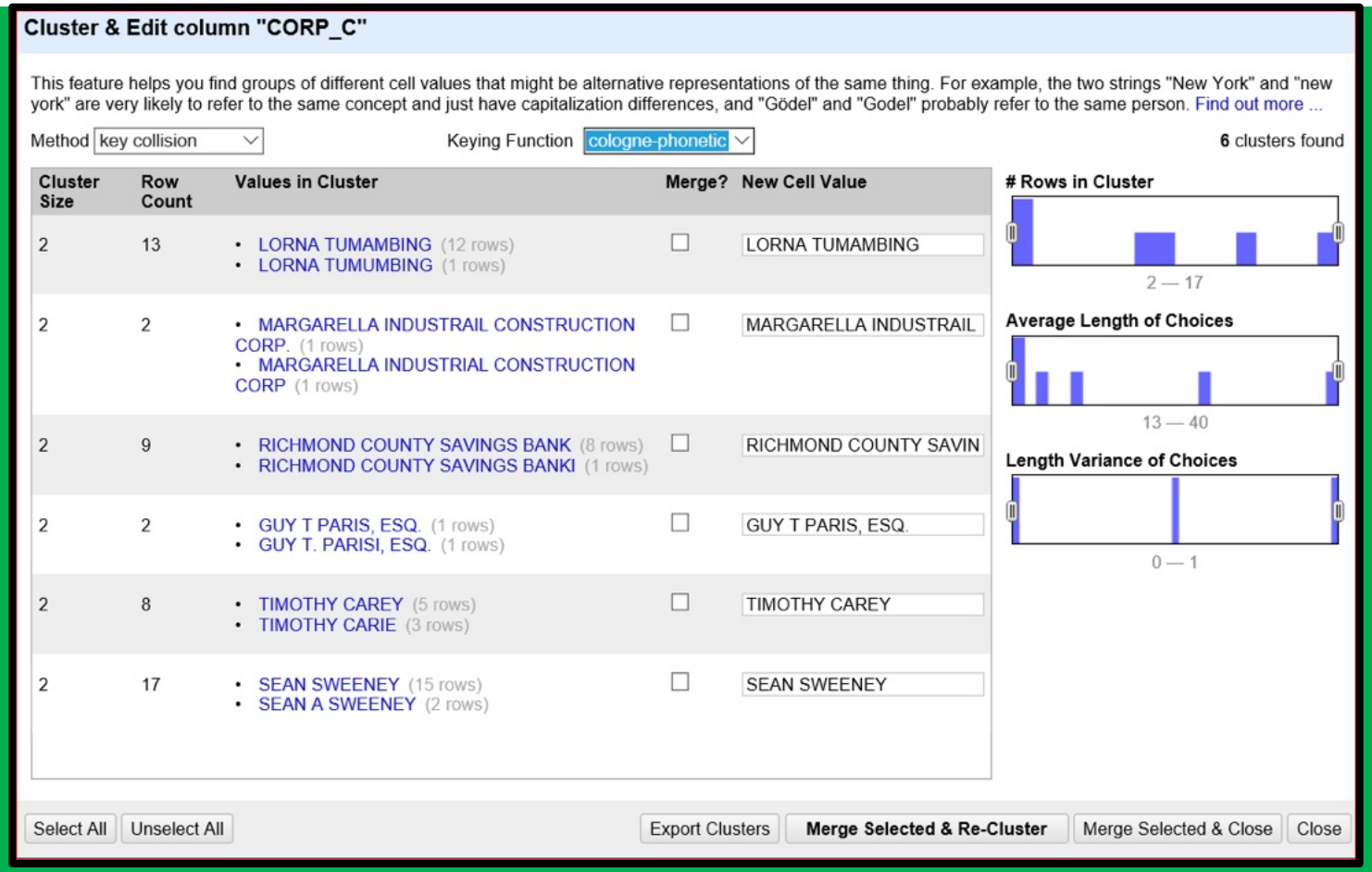
Select All | Unselect All | Export Clusters | **Merge Selected & Re-Cluster** | Merge Selected & Close | Close

CLUSTERING with metaphone3





CLUSTERING with cologne-phonetic





New England
First Amendment Coalition

Thank You!

Questions?

mmulvih@bu.edu