



12th Annual
**New England
First Amendment Institute**



October 23-25
Northeastern University

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

Maggie Mulvihill
mmulvih@bu.edu

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



Autodetect Tables											
Clear All Selections											
Preview & Export Extracted Data											
Year	Month	Revenue in Millions	Operating Income	Net Income	Operating Income	Net Income	Operating Income	Net Income	Operating Income	Net Income	Operating Income
2012	Jan	3,320,002	1,774	2,013	239	15,003	9,287	3,806	3,872	3,320,074	
	Feb	3,320,074	2,528	2,643	154	11,618	8,022	2,966	2,966	3,320,475	
	Mar	3,332,470	2,683	2,681	188	13,230	9,153	4,077	3,878	3,332,348	
	Apr	3,332,348	2,665	2,485	76	11,927	8,810	3,317	3,378	3,332,724	
	May	3,332,724	2,834	2,036	179	11,374	8,879	2,485	2,884	3,342,028	
	Jun	3,342,028	2,768	2,703	65	11,568	9,134	2,434	2,905	3,342,108	
	Jul	3,342,108	3,115	2,671	444	14,577	11,927	2,800	3,087	3,342,205	
	Aug	3,342,205	3,253	2,470	793	16,384	12,376	3,869	4,664	3,342,689	
	Sep	3,352,689	2,932	1,948	988	16,822	10,774	6,148	7,139	3,352,028	
	Oct	3,352,028	3,013	3,148	10	16,752	11,862	7,813	7,115	3,352,727	
	Nov	3,352,727	2,872	2,848	24	14,487	9,227	5,300	5,288	3,373,015	
	Dec	3,373,015	4,245	3,704	621	9,204	7,932	1,862	2,287	3,373,022	
2013	Jan	3,373,022	1,788	2,033	240	15,944	10,257	3,687	3,447	3,373,009	
	Feb	3,373,009	2,441	2,793	372	17,781	9,606	3,205	2,853	3,361,972	
	Mar	3,381,972	2,384	2,782	368	12,219	8,608	3,611	3,221	3,384,793	
	Apr	3,384,793	2,819	3,103	124	12,967	9,980	3,969	3,851	3,388,644	
	May	3,388,644	2,801	2,718	86	11,788	8,826	3,249	3,334	3,391,978	
	Jun	3,391,978	2,775	2,637	148	10,238	9,970	3,218	3,182	3,394,105	
	Jul	3,394,105	3,284	2,691	803	16,110	13,381	2,729	3,339	3,387,489	
	Aug	3,397,489	3,121	2,660	467	16,381	12,872	3,208	3,876	3,401,147	
	Sep	3,401,147	2,831	2,280	551	17,886	12,143	6,715	6,487	3,407,614	
	Oct	3,407,614	3,387	2,876	722	18,376	12,822	6,784	7,477	3,418,981	
	Nov	3,418,981	2,782	2,588	204	14,152	9,836	4,317	4,532	3,419,923	
	Dec	3,419,923	4,387	3,838	592	10,388	9,786	1,642	2,288	3,421,929	
2014	Jan	3,421,929	1,912	1,834	278	14,201	10,961	3,240	3,539	3,425,368	
	Feb	3,425,368	2,716	2,536	171	11,874	9,938	2,186	2,582	3,427,918	
	Mar	3,427,918	2,733	2,690	43	13,999	10,527	2,472	3,914	3,431,032	
	Apr	3,431,032	2,896	2,490	478	12,864	10,166	2,526	3,237	3,432,744	
	May	3,432,744	2,856	2,505	308	11,853	9,582	2,271	2,847	3,437,918	
	Jun	3,437,918	3,280	2,162	1,078	11,887	9,916	1,841	3,075	3,444,991	
	Jul	3,440,991	3,577	2,935	642	14,622	13,836	802	1,668	3,442,048	
	Aug	3,442,048	2,881	2,875	16	16,176	12,800	2,808	3,144	3,444,793	
	Sep	3,445,793	3,539	2,888	1,081	18,808	13,720	8,895	7,118	3,452,911	
	Oct	3,448,911	3,322	2,879	443	18,241	12,366	8,874	8,488	3,458,366	
	Nov	3,458,366	2,967	2,828	330	15,758	9,736	5,979	6,768	3,468,164	
	Dec	3,468,164	4,495	4,274	227	12,375	9,347	3,028	3,685	3,468,849	
2015	Jan	3,468,849	1,927	1,884	43	15,088	11,442	3,646	4,085	3,473,034	
	Feb	3,473,034	2,686	2,681	48	13,487	11,214	2,275	2,685	3,476,918	
	Mar	3,476,918	2,875	3,386	421	14,912	12,445	2,765	2,539	3,478,088	
	Apr	3,478,088	2,711	2,913	352	13,279	10,800	2,470	2,608	3,483,187	
	May	3,483,187	2,967	2,736	231	9,308	8,463	862	1,138	3,482,285	
	Jun	3,482,285	3,712	3,492	202	12,291	12,121	1,128	2,720	3,482,903	

Sometimes Government Makes It Easy To Work With Its Data

Key: ■ Yes ■ No ■ Unsure ■ No data

Sort

☐ alphabetically

☒ by score

	Transport Timetables	Government Budget	Government Spending	Election Results	Company Register	National Map	National Statistics	Legislation	Postcodes / Zipcodes	Emissions of pollutants	Total Score
United Kingdom	Yes	No	Yes	No	Yes	Yes	Yes	No	Yes	No	940
United States	Unsure	Yes	No	Yes	No	Yes	Yes	Yes	Yes	Yes	870
Denmark	No	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	No	835
Norway	No	Unsure	No	Yes	Unsure	Yes	Yes	No	No	Yes	755
Netherlands	No	Yes	No	Yes	No	Yes	No	Yes	Yes	No	740
Australia	No	No	Unsure	Yes	No	Yes	Yes	No	No	No	705
Finland	Yes	No	No	Yes	No	Yes	No	No	No	Unsure	700



Executive Department New Hires

**Data provided by the Human Resources Division*

Sometimes Government Makes It Easy to
Get Data

Just Hit the Export or Download Button on a Government Site

Commission for the Blind	\$254,426.54
Commission for the Deaf and Hard Of Hearing	\$228,090.94
Department of Agricultural Resources	\$181,676.56
Department of Children & Families	\$14,466,090.68
Department of Conservation & Recreation	\$10,641,545.17
Department of Correction	\$3,651,954.38
Department of Developmental Services	\$13,350,376.02
Department of Early Education and Care	\$995,194.44
Department of Elder Affairs	\$177,620.20
Department of Elementary and Secondary Education	\$581,488.38

Total

ler Priv



Executive Depa

for a Department, Jok

DOWNLOAD

Data

We've made the data that powers Open Payroll available to you.

Get Data

Dataset	Description	View	Download
Payroll Transactions	Data for Payroll Transactions as displayed on this website.	View	Download

CTHRU Executive Dept. New Hires Home

Mass.gov Comptroller Home CTHRU Home   

Data

We've made the data that powers Open Payroll available to you.

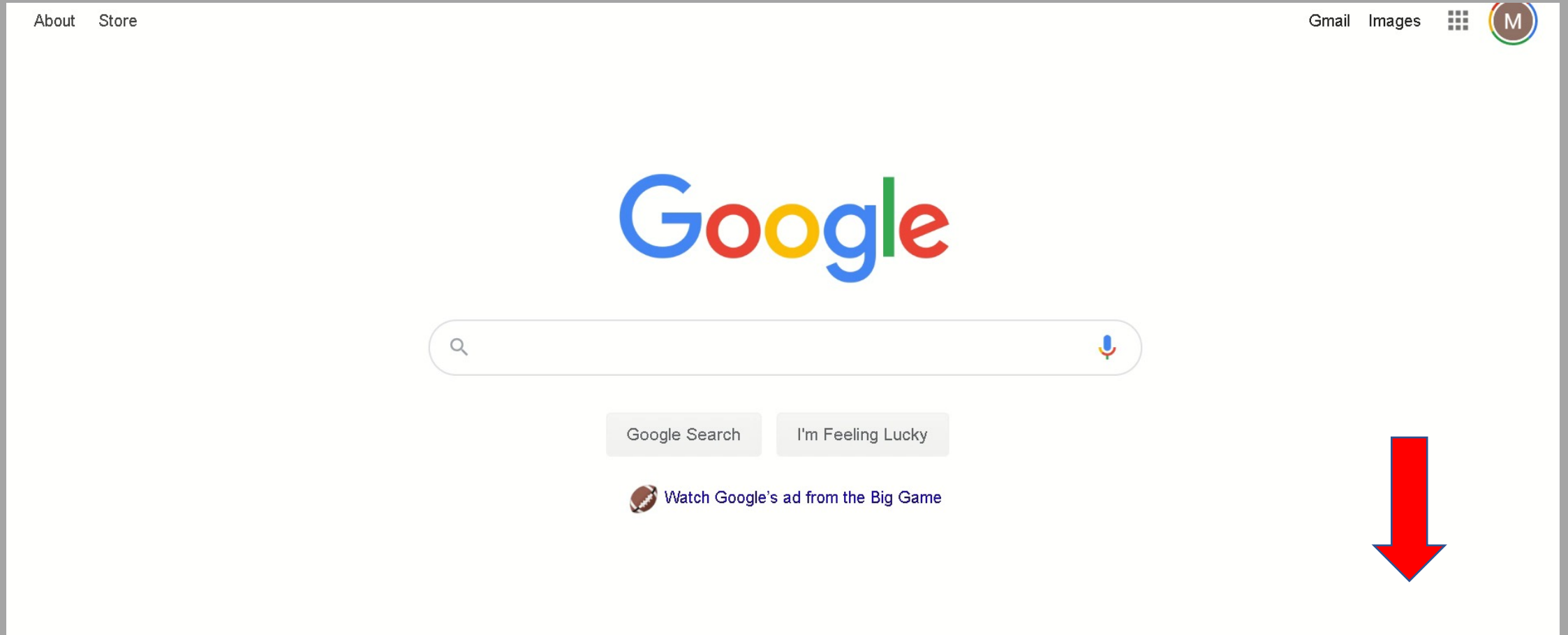
Get Data

Dataset	Description	View	Download
Payroll Transactions	Data for "Payroll Transactions" as displayed on this website.	View	Download



We Can Just Download Data

Let's Try It: Finding Data By File Type



Finding Data

Let's look for the most recent spreadsheet of all individuals or entities who have been "debarred" or banned from doing business in

Advanced Search

Find pages with...

all these words: Type the important words: tricolor rat terrier

this exact word or phrase: Put exact words in quotes: "rat terrier"

any of these words: Type OR between all the words you want: miniature OR standard

none of these words: Put a minus sign just before words you don't want: -rodent, -"Jack Russell"

numbers ranging from: to Put 2 periods between the numbers and add a unit of measure: 10..35 lb, \$300..\$500, 2010..2011

Then narrow your results by...

language: Find pages in the language you select.

region: Find pages published in a particular region.

last update: Find pages updated within the time you specify.

site or domain: Search one site (like wikipedia.org) or limit your results to a domain like .edu, .org or .gov

terms appearing: Search for terms in the whole page, page title, or web address, or links to the page you're looking for.

[SafeSearch:](#) Tell [SafeSearch](#) whether to filter sexually explicit content.

file type: Find pages in the format you prefer.

debarment-list [Compatibility Mode] - Excel

Mulvihill, Maggie

FILE HOME INSERT PAGE LAYOUT FORMULAS DATA REVIEW VIEW OFFICE TAB KUTOOLS ENTERPRISE

Paste Cut Copy Format Painter

MS Sans Serif 8.5 A A

B I U

Wrap Text

General

\$ % , .00 .00

Conditional Formatting Format as Table Cell Styles

Insert Delete Format

AutoSum Fill Clear

Sort & Find & Filter Select

Clipboard Font Alignment Number Styles Cells Editing

A2 16-28411

debarment-list [Compatibility Mode]

	A	B	C	D	E	F	G	H	I	J
1	SWO ID #	Company	Street	City	State	SWO Date	Industry	NAICS		
2	16-28411	1 800 Busy Dog Basement Technologies Inc	35 North Street	Canton	MA	1/14/2016		Not Enough Information/Unclassifiable		
3	14-24080	1 American Construction, Inc.	12 Wall St.	Brockton	MA	7/22/2014	Construction	238140		
4	14-24261	1 Call Does It All And Then Some, Inc.	347 Bridge Road	Florence	MA	8/11/2014		Not Enough Information/Unclassifiable		
5	15-28032	1 Up Inc DBA Dollar Deal Store (2)	19 Oxford St	Lawrence	MA	11/23/2015	Family Clothing Stores	448140		
6	15-26503	1264 Grafton Street LLC	1264 Grafton Street	Worcester	MA	5/26/2015		Not Enough Information/Unclassifiable		
7	14-22506	191 Union LLC	193 Union Street	Lawrence	MA	1/22/2014		Not Enough Information/Unclassifiable		
8	15-27329	194B Park Ave LLC	194 B Park Ave	Worcester	MA	8/27/2015		Not Enough Information/Unclassifiable		
9	16-28558	1ST Stop Cafe, Inc.	365 Walnut St. Extension	Agawan	MA	2/2/2016	Full-Service Restaurants	722511		
10	15-26573	1-Stop Property Shop Inc	35 East Main St	Webster	MA	6/1/2015		Not Enough Information/Unclassifiable		
11	16-28311	311 Centre, LLC dba Clean & Crisp Laundromat	311 Centre Ave	Rockland	MA	1/4/2016	Coin-Operated Laundries and Drycleaners	812310		
12	16-29215	360 Motorsports	10 Aborn Street	Salem	MA	4/26/2016	General Automotive Repair	811111		
13	15-26831	3B Auto Inc.	18 Canterbury St	Worcester	MA	6/30/2015	General Automotive Repair	811111		
14	16-30740	3B Services Inc	221 Massasoit Road Suite	Worcester	MA	11/7/2016		Not Enough Information/Unclassifiable		
15	15-27607	3R Recycling Solutions Inc	13 Robbie Rd	Avon	MA	10/2/2015		Not Enough Information/Unclassifiable		
16	15-28245	444 Atlantic Inc	444 Humphrey Street	Swampscott	MA	12/17/2015		Not Enough Information/Unclassifiable		
17	14-23725	4AS Corporation dba Grandes Kennedy Fried Chick	144 South Union Street	Lawrence	MA	6/9/2014	Snack and Nonalcoholic Beverage Bars	722515		
18	16-29190	5 Corner Auto Sales Inc	369 Court Street	Brockton	MA	4/21/2016	Automobile and Other Motor Vehicle Merchant Whole	423110		
19	14-24434	606 Andover Street Market, LLC (2)	606 Andover St	Lawrence	MA	8/29/2014	Convenience Stores	445120		
20	14-23606	7 Days Food Mart Inc.	1220 Wilbur Avenue	Somerset	MA	5/27/2014	Convenience Stores	445120		
21	16-30811	70 Milford Trust	33 Mulberry St.	Springfield	MA	11/16/2016		Not Enough Information/Unclassifiable		
22	16-30109	730 Recovery	506 Main St	Wakefield	MA	8/15/2016		Not Enough Information/Unclassifiable		
23	15-27920	777 Convenience	269 Main St	Charlestown	MA	11/9/2015	Convenience Stores	445120		
24	16-30571	81 Clinton Pub LLC	81 Clinton St	Shrewsbury	MA	10/12/2016	Full-Service Restaurants	722511		
25	16-28714	881 Carew Street LLC DBA Blamey Stone Tavern	885 Carew Street	Spngfield	MA	2/23/2016	Drinking Places (Alcoholic Beverages)	722410		
26	16-29014	A & B Lawn Care, LLC	12 Dogwood Dr. Belchert	South Deerfield	MA	3/29/2016	Landscaping Services	561730		
27	15-25719	A & D Associates, Inc. dba Artful Healing Massage &	898A Washington Street	Norwood	MA	2/23/2015	Other Personal Care Services	812199		
28	15-25504	A & Fren Inc dba Sunny's Auto Center	644 Belmont Street	Brockton	MA	1/15/2015	General Automotive Repair	811111		

debarment-list [Compatibility Mode] - Excel

Mulvihill, Maggie

FILE HOME INSERT PAGE LAYOUT FORMULAS DATA REVIEW VIEW OFFICE TAB KUTOOLS ENTERPRISE

Paste Cut Copy Format Painter

MS Sans Serif 8.5 A A

B I U

Wrap Text

General

\$ % , .00 .00

Conditional Formatting Format as Table Cell Styles

Insert Delete Format

AutoSum Fill Clear

Sort & Find & Filter Select

Clipboard Font Alignment Number Styles Cells Editing

A2 X ✓ fx 16-28411

debarment-list [Compatibility Mode]

	A	B	C	D	E	F	G	H	I	J
1	SWO ID #	Company	Street	City	State	SWO Date	Industry	NAICS		
2	16-28411	1 800 Busy Dog Basement Technologies Inc	35 North Street	Canton	MA	1/14/2016		Not Enough Information/Unclassifiable		
3	14-24080	1 American Construction, Inc.	12 Wall St.	Brockton	MA	7/22/2014	Construction	238140		
4	14-24261	1 Call Does It All And Then Some, Inc.	347 Bridge Road	Florence	MA	8/11/2014		Not Enough Information/Unclassifiable		
5	15-28032	1 Up Inc DBA Dollar Deal Store (2)	19 Oxford St	Lawrence	MA	11/23/2015	Family Clothing Stores	448140		
6	15-26503	1264 Grafton Street LLC	1264 Grafton Street	Worcester	MA	5/26/2015		Not Enough Information/Unclassifiable		
7	14-22506	191 Union LLC	193 Union Street	Lawrence	MA	1/22/2014		Not Enough Information/Unclassifiable		
8	15-27329	194B Park Ave LLC	194 B Park Ave	Worcester	MA	8/27/2015		Not Enough Information/Unclassifiable		
9	16-28558	1ST Stop Cafe, Inc.	365 Walnut St. Extension	Agawan	MA	2/2/2016	Full-Service Restaurants	722511		
10	15-26573	1-Stop Property Shop Inc	35 East Main St	Webster	MA	6/1/2015		Not Enough Information/Unclassifiable		
11	16-28311	311 Centre, LLC dba Clean & Crisp Laundromat	311 Centre Ave	Rockland	MA	1/4/2016	Coin-Operated Laundries and Drycleaners	812310		
12	16-29215	360 Motorsports	10 Aborn Street	Salem	MA	4/26/2016	General Automotive Repair	811111		
13	15-26831	3B Auto Inc.	18 Canterbury St	Worcester	MA	6/30/2015	General Automotive Repair	811111		
14	16-30740	3B Services Inc	221 Massasoit Road Suite	Worcester	MA	11/7/2016		Not Enough Information/Unclassifiable		
15	15-27607	3R Recycling Solutions Inc	13 Robbie Rd	Avon	MA	10/2/2015		Not Enough Information/Unclassifiable		
16	15-28245	444 Atlantic Inc	444 Humphrey Street	Swampscott	MA	12/17/2015		Not Enough Information/Unclassifiable		
17	14-23725	4AS Corporation dba Grandes Kennedy Fried Chick	144 South Union Street	Lawrence	MA	6/9/2014	Snack and Nonalcoholic Beverage Bars	722515		
18	16-29190	5 Corner Auto Sales Inc	369 Court Street	Brockton	MA	4/21/2016	Automobile and Other Motor Vehicle Merchant Whole	423110		
19	14-24434	606 Andover Street Market, LLC (2)	606 Andover St	Lawrence	MA	8/29/2014	Convenience Stores	445120		
20	14-23606	7 Days Food Mart Inc.	1220 Wilbur Avenue	Somerset	MA	5/27/2014	Convenience Stores	445120		
21	16-30811	70 Milford Trust	33 Mulberry St.	Springfield	MA	11/16/2016		Not Enough Information/Unclassifiable		
22	16-30109	730 Recovery	506 Main St	Wakefield	MA	8/15/2016		Not Enough Information/Unclassifiable		
23	15-27920	777 Convenience	269 Main St	Charlestown	MA	11/9/2015	Convenience Stores	445120		
24	16-30571	81 Clinton Pub LLC	81 Clinton St	Shrewsbury	MA	10/12/2016	Full-Service Restaurants	722511		
25	16-28714	881 Carew Street LLC DBA Blamey Stone Tavern	885 Carew Street	Spngfield	MA	2/23/2016	Drinking Places (Alcoholic Beverages)	722410		
26	16-29014	A & B Lawn Care, LLC	12 Dogwood Dr. Belchert	South Deerfield	MA	3/29/2016	Landscaping Services	561730		
27	15-25719	A & D Associates, Inc. dba Artful Healing Massage &	898A Washington Street	Norwood	MA	2/23/2015	Other Personal Care Services	812199		
28	15-25504	A & Fren Inc dba Sunny's Auto Center	644 Belmont Street	Brockton	MA	1/15/2015	General Automotive Repair	811111		

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

Massachusetts Bay Transportation Authority

Getting Around ▾Fares ▾Contact Us ▾More ▾EN ▾Search

Home > MBTA Budgets and Financials > MBTA Budget

Financials

MBTA Budgets

Capital Investment Plan

Audited Financials

Major Bonds Programs

Fiscal and Management Control Board

MBTA Back on Track

Investor Information Disclaimer

MBTA Budget

Every year, the MBTA determines which activities and actions can be taken during the next fiscal year to continue to fund programs that provide safe, reliable, and accessible transportation services. These annual budgets are produced based on statutory guidelines, and can be downloaded below.

For MBTA financial statements produced in accordance with Generally Accepted Accounting Principles (GAAP), please visit our [audited financial statements](#).

Note: The information in this section is presented subject to readers' acknowledgment of [this disclaimer](#).

MBTA Payroll

2017

2015

2014

2013

Employee ID	Last Name	First Name	Title	Hire Date	Base Pay	Total 2017 Earnings
452912	Aalto	Joanna	AGM Capital Programs	03-Apr-16	\$185,000.00	\$28,461.55
72929	Aaron	Kevar	Laborer, Track	21-Sep-15	\$57,803.20	\$72,800.63
72021	Aaron	Keyena	Operator, P-T Surface	05-May-14	\$47,548.80	\$48,448.81
68101	Abany	Samuel	Sergeant	10-Apr-06	\$95,518.40	\$202,214.19
68415	Abbit	Cynthia	Inspector	05-Sep-06	\$83,304.00	\$98,700.28
9749	Abbott	Kenneth	Operator, Surface	03-Feb-92	\$74,588.80	\$85,631.68
72103	Abbott	Charles	Operator, Surface	23-Jun-14	\$63,398.40	\$74,230.46
73625	Abbott	Corey	Operator, P-T Surface	09-Jan-17	\$30,763.20	\$29,608.46
451427	Abdelkarim	Loay	Mgr Design	27-Feb-12	\$117,279.99	\$116,354.38
71810	Abdella	Nasir	Operator, Surface	06-Jan-14	\$67,121.60	\$73,542.99
72734	Abdelouahhab	Ali	Operator, Surface	22-Jun-15	\$63,398.40	\$63,624.21
67965	Abdelrahman	Ashraf	Supv Trans - Bus	11-Oct-05	\$91,312.90	\$89,721.43
72691	Abderazzak	Sammy	Auditor, Field	01-Jun-15	\$65,880.52	\$64,723.80
466160	Abdulakhatov	Akmal	Supv Transm&Dist Eng&Sfty	12-Feb-17	\$111,152.29	\$94,330.77
68620	Abdullahi	Abdul-Hakim	Trackperson	10-Oct-06	\$71,552.00	\$93,345.76
69697	Abell	Paul	Worker, Sheet Metal	12-Aug-13	\$96,553.60	\$108,503.17
67679	Abernathy	Sandra	Motorperson, Rapid Transit	04-Apr-05	\$77,563.20	\$55,099.48
69204	Abibi	Abdelmalek	Operator, Surface	10-Sep-07	\$74,588.80	\$92,339.26
66635	Abkarian	Ropen	Operator, Surface	15-Oct-02	\$74,588.80	\$15,562.41
67420	Abraham	Bernard	Operator, Surface	07-Sep-04	\$74,588.80	\$110,055.70
369115	Abrahamson	Daniel	Spec, Energy Conservation	06-Jun-16	\$71,370.43	\$70,117.26
73316	Abramo	Michael	Chief Financial Officer	06-Jun-16	\$210,000.00	\$186,635.50
67817	Abreu	Estarlo	Operator, Surface	11-Jul-05	\$74,588.80	\$87,679.32
71537	Abreu	Yris	Operator, Surface	20-May-13	\$74,588.80	\$58,391.87
73503	Abujade	Olalekan	Operator, P-T Surface	06-Sep-16	\$41,964.00	\$41,757.36
477676	Acchat	Kerby	Operator, P-T Surface	19-Nov-17	\$30,763.20	\$2,200.00
71941	Accime	Susie	Operator, Surface	24-Feb-14	\$67,121.60	\$97,051.46
67909	Acevedo	Tyrone	Repairer, Rail	03-Oct-05	\$86,715.20	\$129,078.87
72828	Acharya	Horn	Operator, Surface	20-Jul-15	\$55,952.00	\$44,999.50
92047	Adams	Lee	Worker, Sheet Metal	22-Aug-00	\$96,553.60	\$144,456.73
7796	Adams	Edward	Inspector, Chief - SL	13-Feb-84	\$87,214.40	\$130,845.02
67527	Adams	Jeffrey	Repairer, Rail	13-Dec-04	\$86,715.20	\$119,303.28
66003	Adams	Shella	Dispatcher Bus	09-Jul-17	\$92,934.40	\$117,796.60

- What type of “file” is a PDF?
- Adobe created PDF more than 20 years ago
- Primarily used for viewing, not working with documents
- **Portable Document Format** – can look at it no matter what operating system you are using
- Create the document one time and use Adobe to look at it

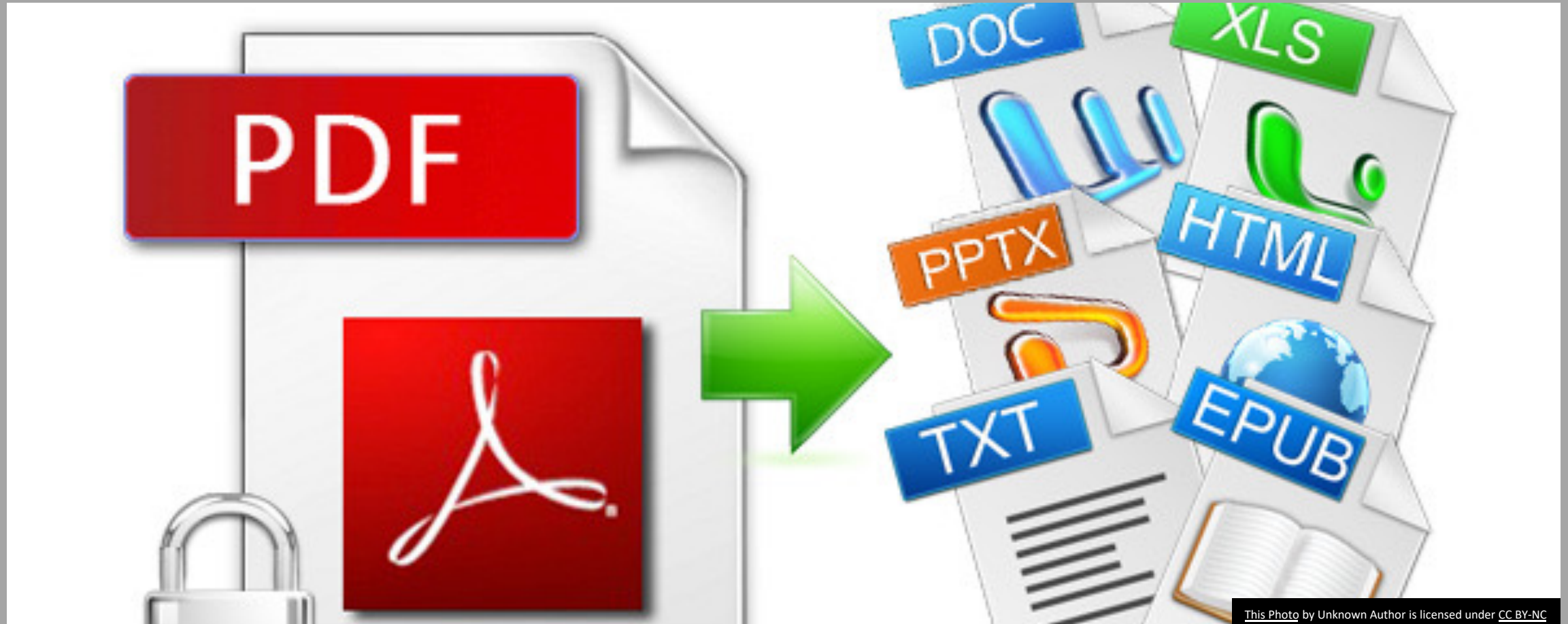
Obtaining Data: Converting Files

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

The logo for Cometdocs, featuring the word "cometdocs" in a lowercase, sans-serif font. The "comet" part is in grey, and the "docs" part is in a lighter grey. A stylized orange and yellow comet tail graphic is positioned above the "o" in "comet".

Use a PDF CONVERTER to extract the data from PDF documents so you can work with it



This Photo by Unknown Author is licensed under [CC BY-NC](#)

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

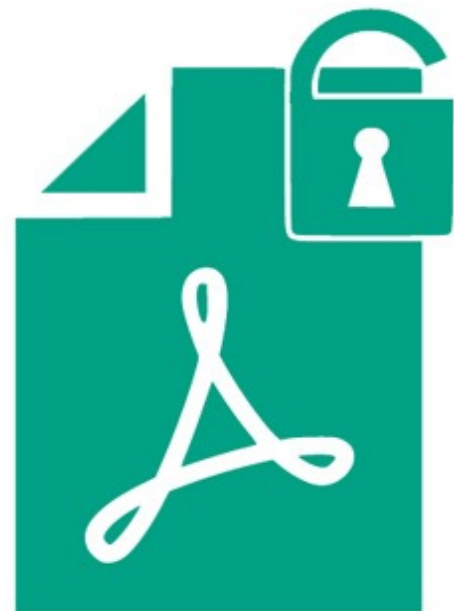
Extracting Data from PDF's with Tabula



TABULA

liberate data tables trapped inside PDFs

A free tool like
Tabula comes in
handy for PDF's
with multiple
tables



TABULA

liberate data tables trapped inside PDFs

What Is TABULA?

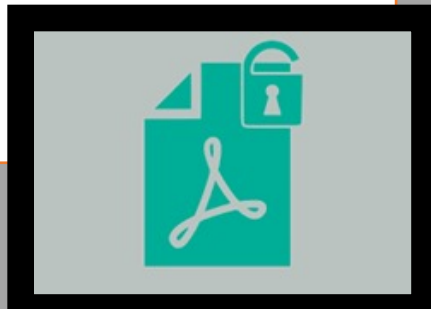


- Free extraction tool to turn tabular data locked in PDF'S into exportable CSV or spreadsheet files
- Works on Windows, MAC or Linux systems
- Works on text-based PDF's, where you can select text
- Not scanned documents

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 PDF's are a collection of images
- Can't extract the characters
- Could use OCR but accuracy can be problem



Scanned Image vs. Text

The image shows two side-by-side Adobe Reader windows. The left window, titled '1965AMLC', displays a scanned document page. It features a table of contents with a blue selection box around the text: 'President', '1st Vice-President', '2nd Vice-President', 'Secretary-Treasurer', and 'Members-at-large'. A 'Copy Image' tooltip is visible over the selection. The right window, titled '1218205495', displays a normal text document. It has a blue selection box around the 'INTRODUCTION' section, which contains two numbered list items. A 'Copy' tooltip is visible over the selection. Below the windows, the word 'Scanned' is centered under the left window and 'Normal' is centered under the right window. A central icon of a document with a lock and a checkmark is positioned between the two labels.

Adobe Reader File Edit View Window Help 1965AMLC 3 / 24 10

Adobe Reader File Edit View Window Help 1218205495 1 / 36 83

BOARD OF D (newly e)

President
1st Vice-President
2nd Vice-President
Secretary-Treasurer
Members-at-large

Copy Image

INTRODUCTION

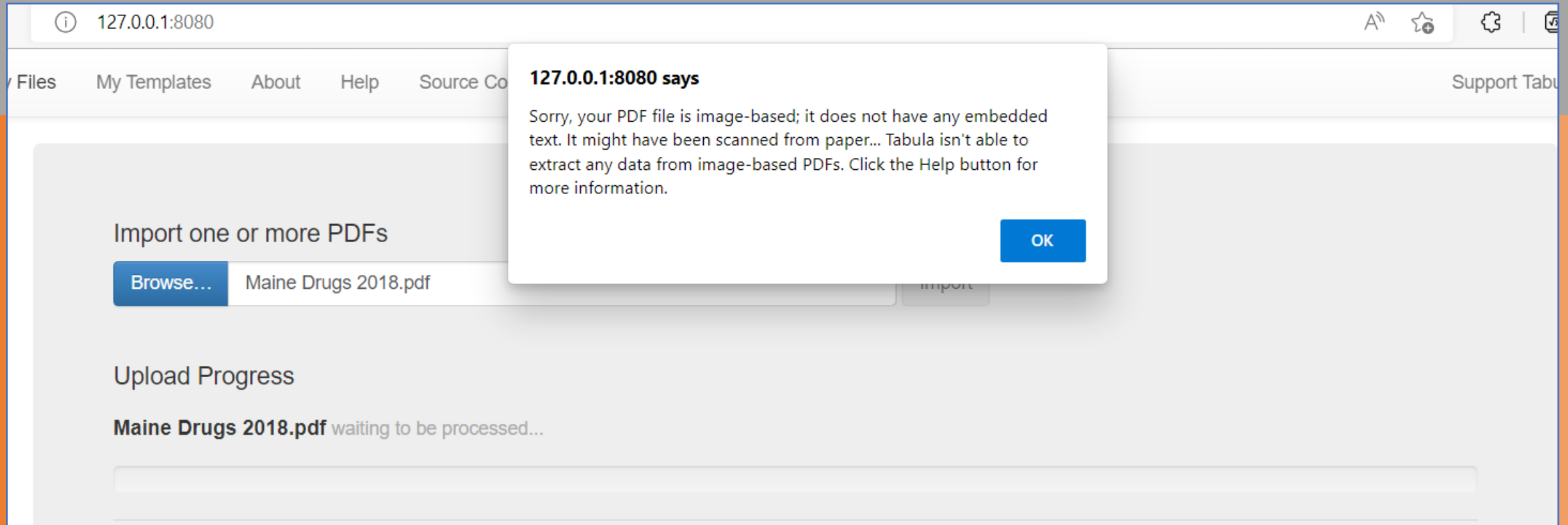
1. These notes relate to the receipt of Assent on 8th November 2005 by the Department for Education and Skills in relation to the Act. They do not form part of the Act.
2. The notes need to be read in conjunction with the Act. To be, a comprehensive description of the Act does not seem to require any explanation.

SUMMARY AND BACKGROUND

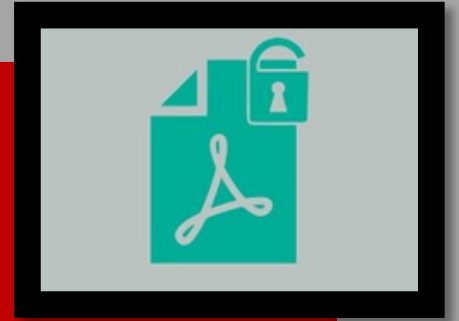
3. The current system for vetting people

Scanned Normal

TABULA CANNOT IMPORT IMAGE-BASED DOCUMENTS



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 elephant1
Database changed
mysql> SOURCE C:\xampp\htdocs\elephant12\db\mas_gis_city_ip.sql;
ERROR:
Unknown command '\x'.

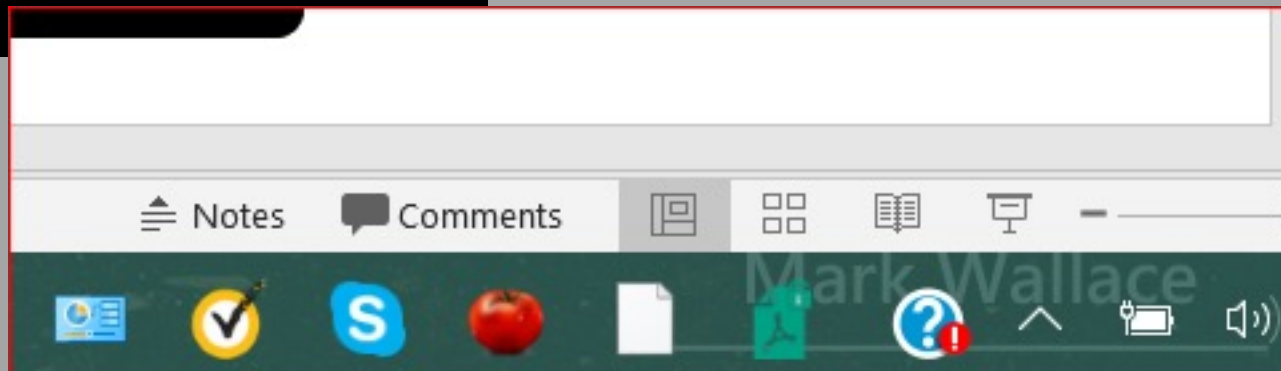
For information about MySQL products and services, visit:
  http://www.mysql.com/
For developer information, including the MySQL Reference Manual, visit:
  http://dev.mysql.com/
To buy MySQL Enterprise support, training, or other products, visit:
  https://shop.mysql.com/

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 (s) 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 PDF'S 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
- These are county by county deaths of overdoses across Massachusetts from ____ (what do I need to know?)
- Do a four-corners check – how many tables are there?
- what stories do you see?



Number of Opioid-Related Overdose Deaths, All Intent by City/Town 2015-2019

Massachusetts Department of Public Health

POSTED: NOVEMBER 2020

Note to Readers: As previously published, Table 1 contains counts of opioid-related overdose deaths in the city/town of residence for the decedents among MA residents. Table 2 contains counts of opioid-related overdose deaths in the city/town of the death occurrence. The counts in Table 2 are different from Table 1 because they include counts based on location of death versus location of the decedent's residence (Table 1). Massachusetts residents who died in Massachusetts are included in both tables. Massachusetts residents who died outside of Massachusetts are included in the "Death Occurrence" table (Table 2) as "Outside of Massachusetts". Non-Massachusetts residents who died in Massachusetts are not included in the "Residence" table (Table 1).

Table 1. Number of confirmed opioid-related overdose deaths for all intents by city/town of residence for the decedent, among MA residents, 2015-2019. For 2017 to 2019, additional cases are still being confirmed by the Office of the Chief Medical Examiner. This report will be updated quarterly, and all new confirmed cases will be included in the table below with previously confirmed cases.

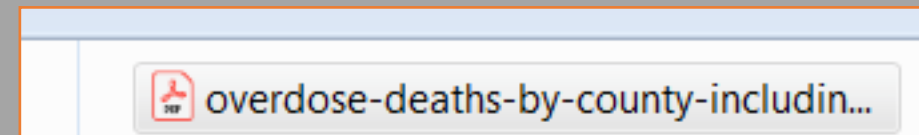
City/Town of Residence	Year of Death				
	2015	2016	2017	2018	2019
Abington	6	3	4	11	4
Acton	4	7	1	4	0
Acushnet	4	4	7	2	1
Adams	3	1	0	7	2
Agawam	5	4	7	12	13
Alford	0	0	0	0	0
Amesbury	7	8	10	9	7
Amherst	1	2	3	2	3
Andover	6	3	2	2	6
Aquinnah	1	0	0	0	0
Arlington	6	6	3	3	4
Ashburnham	1	2	0	1	2
Ashby	0	0	0	0	1
Ashfield	1	0	0	1	0
Ashland	4	4	4	2	3
Athol	3	3	6	6	7
Attleboro	10	18	26	11	21
Auburn	4	2	3	6	4
Avon	3	3	2	2	1
Ayer	1	7	1	2	4
Barnstable	13	21	13	18	18
Barre	3	0	0	3	3
Becket	0	0	0	2	0
Bedford	1	6	3	3	4

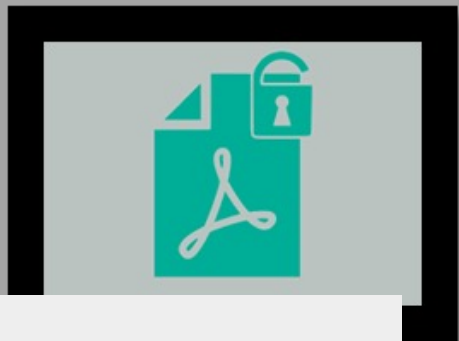


Importing to TABULA



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

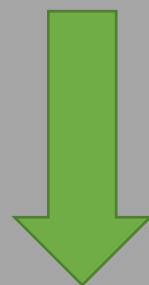




Import one or more PDFs

Browse...

Import



☒		Opioid-related-Overdose-Deaths-by-...	8/20/2021 7:00 AM	Adobe Acrobat D...	377 KB
-------------------------------------	---	---------------------------------------	-------------------	--------------------	--------



Import one or more PDFs

Browse...

Opioid-related-Overdose-Deaths-by-City-Town-November-2020.pdf

Import

Import one or more PDFs

Browse...

Opioid-related-Overdose-Deaths-by-City-Town-November-2020.pdf

Import

Upload Progress

Opioid-related-Overdose-Deaths-by-City-Town-November-2020.pdf generating page thumbnails...



Import Your PDF

Opioid-related-Overdose-Deaths-by-City-...

Templates

Clear All Selections

Autodetect Tables

Preview & Export Extracted

1.



Number of Opioid-Related Overdose Deaths, All Intents by City/Town 2015-2019

Massachusetts Department of Public Health

POSTED: NOVEMBER 2020

Note to Readers: As previously published, Table 1 contains counts of opioid-related overdose deaths in the city/town of residence for the decedents among MA residents. Table 2 contains counts of opioid-related overdose deaths in the city/town of the death location. The data for the city/town of residence for the decedents among MA residents are included in the "Death Occurred in City/Town of Residence" column. The data for the city/town of the death location are included in the "Death Occurred in City/Town of Death Location" column.

Table 1. Number of Opioid-Related Overdose Deaths, All Intents by City/Town of Residence, among MA Residents, 2015-2019

City/Town of Residence	2015	2016	2017	2018	2019
Abington	6	3	4	11	4
Acton	4	7	1	4	0
Acushnet	4	4	7	2	1
Adams	3	1	0	7	2
Agawam	5	4	7	12	13
Alford	0	0	0	0	0

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

[My Files](#)
[My Templates](#)
[About](#)
[Help](#)
[Source Code](#)

Support Tabula on OpenCollective!

Opioid-related-Overdose-Deaths-by-City-...

Templates

Clear All Selections

Autodetect Tables

Preview & Export Extracted Data

1

2

Number of Opioid-Related Overdose Deaths,
All Intents by City/Town
2015-2019

Massachusetts Department of Public Health

POSTED: NOVEMBER 2020

Note to Readers: As previously published, Table 1 contains counts of opioid-related overdose deaths in the city/town of residence for the decedents among MA residents. Table 2 contains counts of opioid-related overdose deaths in the city/town of the death occurrence. The counts in Table 2 are different from Table 1 because they include counts based on location of death versus location of the decedent's residence (Table 1). Massachusetts residents who died in Massachusetts are included in both tables. Massachusetts residents who died outside of Massachusetts are included in the "Death Occurrence" table (Table 2) as "Outside of Massachusetts". Non-Massachusetts residents who died in Massachusetts are not included in the "Residence" table (Table 1).

Table 1. Number of confirmed opioid-related overdose deaths for all intents by city/town of residence for the decedent, among MA residents, 2015-2019. For 2017 to 2019, additional cases are still being confirmed by the Office of the Chief Medical Examiner. This report will be updated quarterly, and all new confirmed cases will be included in the table below with previously confirmed cases.

City/Town of Residence	Year of Death				
	2015	2016	2017	2018	2019
Abington	6	3	4	11	4
Acton	4	7	1	4	0
Acushnet	4	4	7	2	1
Adams	3	1	0	7	2
Agawam	5	4	7	12	13
Alford	0	0	0	0	0

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



Extracting Data With Tabula

- Click “Autodetect Tables”
- TABULA will go through and highlight in pink data that it should extract from the PDF

overdose-deaths-by-county-including-ma...

⚡ Autodetect Tables

🗑 Clear All Selections

📄 Preview & Export Extracted Data

1.



Number of Opioid¹-Related Overdose Deaths, All Intent² by County, MA Residents: 2000-2016²

Massachusetts Department of Public Health

POSTED: MAY 2017

Note To Readers: This is the first quarterly report for 2017. Please note that throughout all sections of this report there has been a change in the category of deaths that are included. Death data that were previously reported as “Unintentional/Undetermined” has been updated to include deaths of “All Intent²”, which means that opioid-related deaths that have been determined by the Office of the Chief Medical Examiner as the result of suicide will now be included in the totals. This will not add significantly to the death totals, however it will allow for consistency in interpretation in the data that are presented.

County	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015 ¹	2016 ¹	Total 2000-2016
Barnstable	12	19	21	15	17	20	24	31	22	21	20	19	24	43	55	71	78	512
Berkshire	2	3	1	2	6	9	1	8	3	8	4	9	18	22	29	33	35	193
Bristol	38	58	66	84	72	78	85	64	84	70	79	82	95	116	144	172	233	1620
Dukes	1	0	1	0	1	2	0	3	1	2	0	0	0	1	5	7	3 ¹	27
Essex	51	65	55	81	68	80	87	89	65	74	51	57	94	119	208	240	284	1768
Franklin	5	2	1	5	3	4	6	4	2	2	6	6	8	10	10	19	14	107
Hampden	32	37	42	46	32	36	46	38	48	46	48	45	59	69	64	102	128	918
Hampshire	5	5	5	12	8	3	10	14	11	10	12	10	11	30	26	16	34	222
Middlesex	64	86	92	115	105	122	118	110	112	124	94	130	122	155	273	354	392	2568
Nantucket	0	1	0	0	0	0	0	1	0	1	1	0	0	1	1	1 ¹	2 ¹	9 ¹
Norfolk	29	42	44	43	40	53	49	54	73	65	60	64	71	83	126	166	202	1264
Plymouth	23	25	29	47	28	38	49	52	48	53	41	67	57	86	112	178	185	1118
Suffolk	49	87	98	106	83	66	107	103	74	95	64	85	91	111	145	210	253	1827
Worcester	68	75	71	58	51	63	74	71	78	67	80	82	91	115	162	223	226	1655
TOTAL DEATHS	875	1006	1026	1214	1074	1274	1366	1427	1247	1318	1096	1242	1462	1864	2791	3464	4791	31815

¹ Opioids include heroin, opioid-based prescription painkillers, and other unspecified opioids.

² Please note that data for 2000-2014 have been updated following a review of cases that did not receive an official cause of death at the time the file was closed. Data for 2015-2016 death are preliminary and subject to updates. Case reviews of deaths are evaluated and updated on an ongoing basis. A large number of death certificates have yet to be assigned final cause-of-death codes. These counts are based on the estimates rather than confirmed cases. Data updated on 04/04/2017.

³ Numbers and calculations based on values less than 5 are suppressed for years in which the death file is not yet closed if they are based on pending cases. The deaths listed in Nantucket County for 2015 and 2016, and in Dukes County for 2016 are confirmed opioid overdose deaths.

⁴ 2001, 2005, 2008, 2012, and 2014 include 1 death of a MA resident whose city/town of residence was not known; 2006 includes 4 deaths of a MA resident whose city/town of residence was not known. Also, total may not add up due to rounding.

Please note that there is rounding of counts for 2015-2016.

Repeat this Selection

Previewing



- TABULA will generate images you can PREVIEW
- Scroll through 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



Extracting Data With Tabula

Preview & Export Extracted Data

If you like what you see, you can “Preview and Export” the PDF

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

Opioid-related-Overdose-Deaths-by-City-...

1.

2

3

4

Number of Opioid-Related Overdose Deaths, All Intents by City/Town 2015-2019

Massachusetts Department of Public Health POSTED: NOVEMBER 2020

Note to Readers: As previously published, Table 1 contains counts of opioid-related overdose deaths in the city/town of residence for the decedents among MA residents. Table 2 contains counts of opioid-related overdose deaths in the city/town of the death occurrence. The counts in Table 2 are different from Table 1 because they include counts based on location of death versus location of the decedent's residence (Table 1). Massachusetts residents who died in Massachusetts are included in both tables. Massachusetts residents who died outside of Massachusetts are included in the “Death Occurrence” table (Table 2) as “Outside of Massachusetts”. Non-Massachusetts residents who died in Massachusetts are not included in the “Residence” table (Table 1).

Table 1. Number of confirmed opioid-related overdose deaths for all intents by city/town of residence for the decedent, among MA residents, 2015-2019. For 2017 to 2019, additional cases are still being confirmed by the Office of the Chief Medical Examiner. This report will be updated quarterly, and all new confirmed cases will be included in the table below with previously confirmed cases.

City/Town of Residence	Year of Death				
	2015	2016	2017	2018	2019
Abington	6	3	4	11	4
Acton	4	7	1	4	0
Acushnet	4	4	7	2	1
Adams	3	1	0	7	2
Arawam	5	4	7	12	13

Click Preview and Export Extracted Data

 **Tabula** [My Files](#) [My Templates](#) [About](#) [Help](#) [Source Code](#) [Support Tabula on OpenCollective!](#)

Opioid-related-Overdose-Deaths-by-City-... [Templates](#) [Clear All Selections](#) [Autodetect Tables](#) [Preview & Export Extracted Data](#)

1.

2

3

4



Number of Opioid-Related Overdose Deaths, All Intents by City/Town 2015-2019

Massachusetts Department of Public Health POSTED: NOVEMBER 2020

Note to Readers: As previously published, Table 1 contains counts of opioid-related overdose deaths in the city/town of residence for the decedents among MA residents. Table 2 contains counts of opioid-related overdose deaths in the city/town of the death occurrence. The counts in Table 2 are different from Table 1 because they include counts based on location of death versus location of the decedent's residence (Table 1). Massachusetts residents who died in Massachusetts are included in both tables. Massachusetts residents who died outside of Massachusetts are included in the "Death Occurrence" table (Table 2) as "Outside of Massachusetts". Non-Massachusetts residents who died in Massachusetts are not included in the "Residence" table (Table 1).

Table 1. Number of confirmed opioid-related overdose deaths for all intents by city/town of residence for the decedent, among MA residents, 2015-2019. For 2017 to 2019, additional cases are still being confirmed by the Office of the Chief Medical Examiner. This report will be updated quarterly, and all new confirmed cases will be included in the table below with previously confirmed cases.

City/Town of Residence	Year of Death				
	2015	2016	2017	2018	2019
Abington	6	3	4	11	4
Acton	4	7	1	4	0
Acushnet	4	4	7	2	1
Adams	3	1	0	7	2
Agawam	5	4	7	12	13
Alford	0	0	0	0	0





Opioid-related-Overdose-Deaths-by-City-...

Templates

Clear All Selections

Autodetect Tables

Preview & Export Extracted Data

1.



Number of Opioid-Related Overdose Deaths, All Intents by City/Town 2015-2019

Massachusetts Department of Public Health

POSTED: NOVEMBER 2020

Note to Readers: As previously published, Table 1 contains counts of opioid-related overdose deaths in the city/town of residence for the decedents among MA residents. Table 2 contains counts of opioid-related overdose deaths in the city/town of the death occurrence. The counts in Table 2 are different from Table 1 because they include counts based on location of death versus location of the decedent's residence (Table 1). Massachusetts residents who died in Massachusetts are included in both tables. Massachusetts residents who died outside of Massachusetts are included in the "Death Occurrence" table (Table 2) as "Outside of Massachusetts". Non-Massachusetts residents who died in Massachusetts are not included in the "Residence" table (Table 1).

Table 1. Number of confirmed opioid-related overdose deaths for all intents by city/town of residence for the decedent, among MA residents, 2015-2019. For 2017 to 2019, additional cases are still being confirmed by the Office of the Chief Medical Examiner. This report will be updated quarterly, and all new confirmed cases will be included in the table below with previously confirmed cases.

City/Town of Residence	Year of Death				
	2015	2016	2017	2018	2019
Abington	6	3	4	11	4
Acton	4	7	1	4	0
Acushnet	4	4	7	2	1
Adams	3	1	0	7	2
Agawam	5	4	7	12	13
Alford	0	0	0	0	0

Preview of Extracted Tabular Data

	Year of Death				
City/Town of Residence					
	2015	2016	2017	2018	2019
Abington	6	3	4	11	4
Acton	4	7	1	4	0
Acushnet	4	4	7	2	1
Adams	3	1	0	7	2
Agawam	5	4	7	12	13
Alford	0	0	0	0	0
Amesbury	7	8	10	9	7
Amherst	1	2	3	2	3
Andover	6	3	2	2	6
Aquinnah	1	0	0	0	0
Arlington	6	6	3	3	4
Ashburnham	1	2	0	1	2
Ashby	0	0	0	0	1
Ashfield	1	0	0	1	0

Liberating Data With Tabula

Tabula Has "Freed" Your
Tables

Asks You To Double-Check Your Work

Is the extracted data incorrect?

You can revise your selected cells or try an alternate extraction method.

Revise Selected Cells

Data has been extracted from the cells you selected in the previous step. You can revise your selection(s) to add or remove cells.

[← Revise selection\(s\)](#)

Choose Alternate Extraction Method

The current preview uses the **Stream** extraction method. If the data is not mapped to the correct cells, try the **Lattice** method instead.

☒ Stream

☐ Lattice

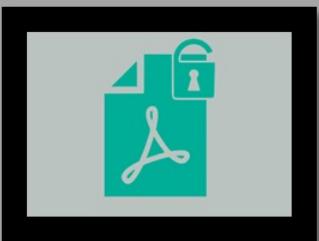
Stream looks for *whitespace* between columns, while Lattice looks for *boundary lines* between columns.

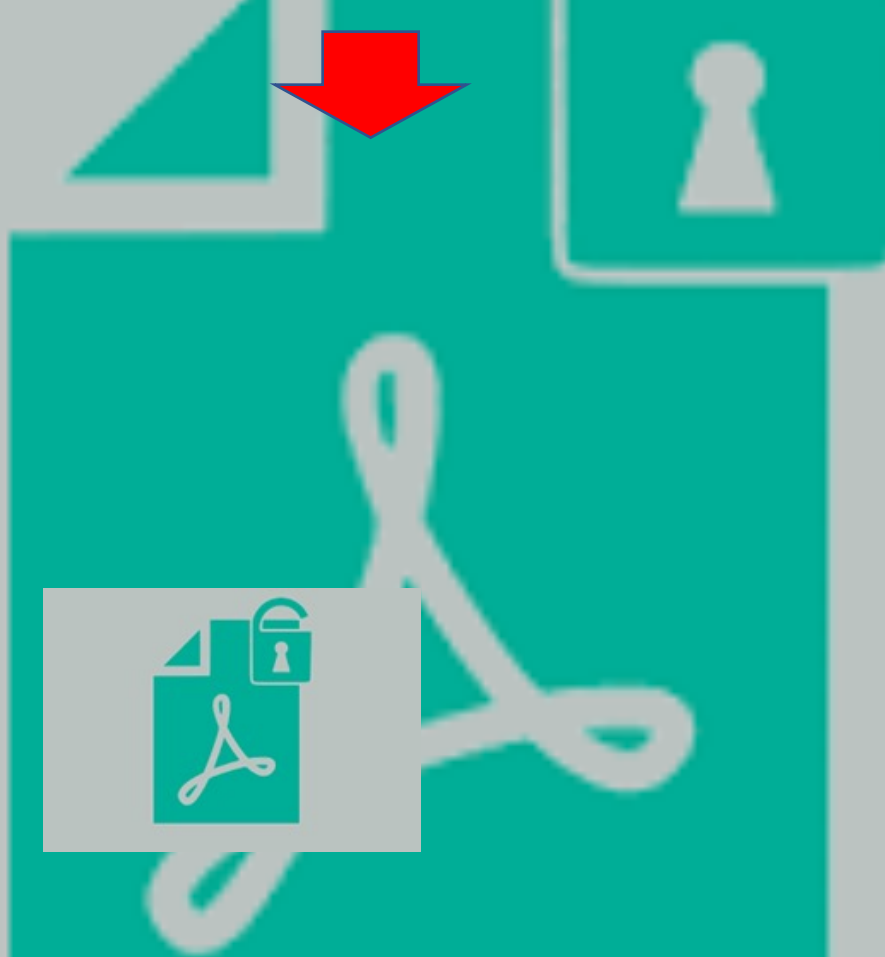
Opioid-related-Overdose-Deaths-by-City-...

Export Format: CSV

[Export](#)

[Copy to Clipboard](#)





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



My Files My Templates About Help Source Code

[Support Tabula on OpenCollective!](#)

Opioid-related-Overdose-Deaths-by-City-... Export Format: CSV Export Copy to Clipboard

Preview of Extracted Tabular Data



City/Town of Residence	Year of Death				
	2015	2016	2017	2018	2019
Abington	6	3	4	11	4
Acton	4	7	1	4	0
Acushnet	4	4	7	2	1
Adams	3	1	0	7	2
Agawam	5	4	7	12	13
Alford	0	0	0	0	0

Check Your Data: Did It Export Correctly?

Export
Your Data
to a CSV
and Open
It



Opioid-related-Overdose-Deaths-by-City-...

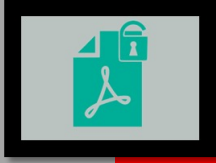
Export Format: CSV

Export

Copy to Clipboard

Preview of Extracted Tabular Data

			Year of Death		
City/Town of Residence					
	2015	2016	2017	2018	2019
Abington	6	3	4	11	4
Acton	4	7	1	4	0
Acushnet	4	4	7	2	1
Adams	3	1	0	7	2
Agawam	5	4	7	12	13
Alford	0	0	0	0	0
Amesbury	7	8	10	9	7
Amherst	1	2	3	2	3
Andover	6	3	2	2	6
Aquinnah	1	0	0	0	0
Arlington	6	6	3	3	4



Open Your Spreadsheet: Analyze

What
Stories
Do You
See?

	A	B	C	D	E	F	
1				Year of Death			
2	City/Town of Residence						
3		2015	2016	2017	2018	2019	
4	Abington	6	3	4	11	4	
5	Acton	4	7	1	4	0	
6	Acushnet	4	4	7	2	1	
7	Adams	3	1	0	7	2	
8	Agawam	5	4	7	12	13	
9	Alford	0	0	0	0	0	
10	Amesbury	7	8	10	9	7	
11	Amherst	1	2	3	2	3	
12	Andover	6	3	2	2	6	
13	Aquinnah	1	0	0	0	0	
14	Arlington	6	6	3	3	4	
15	Ashburnhar	1	2	0	1	2	
16	Ashby	0	0	0	0	1	
17	Ashfield	1	0	0	1	0	
18	Ashland	4	4	4	2	3	
19	Athol	3	3	6	6	7	
20	Attleboro	10	18	26	11	21	

Converting Data with CometDocs

	A	B	C	D	E	F	G
1				Year of Death			
2	City/Town of Residence						
3		2015	2016	2017	2018	2019	
4	Abington	6	3	4	11	4	
5	Acton	4	7	1	4	0	
6	Acushnet	4	4	7	2	1	
7	Adams	3	1	0	7	2	
8	Agawam	5	4	7	12	13	
9	Alford	0	0	0	0	0	
10	Amesbury	7	8	10	9	7	
11	Amherst	1	2	3	2	3	
12	Andover	6	3	2	2	6	
13	Aquinnah	1	0	0	0	0	
14	Arlington	6	6	3	3	4	
15	Ashburnhar	1	2	0	1	2	
16	Ashby	0	0	0	0	1	
17	Ashfield	1	0	0	1	0	
18	Ashland	4	4	4	2	3	
19	Athol	3	3	6	6	7	
20	Attleboro	10	18	26	11	21	

tabula-Opioid-related-Overdose-

- Try to “pivot” your data
 - Do you see any story ledes?
- Try to write one out and dictate it to the class

Omitting Tables

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



Tabula

My Files

Is the extracted data incorrect?

You can revise your selected cells or try an alternate extraction method.

Revise Selected Cells

Data has been extracted from the cells you selected in the previous step. You can revise your selection(s) to add or remove cells.

← Revise selection(s)

Choose Alternate Extraction Method

The current preview uses the **Stream** extraction method. If the data is not mapped to the correct cells, try the **Lattice** method instead.

Stream

Lattice

Stream looks for *whitespace* between columns, while Lattice looks for *boundary lines* between columns.

Still look wrong?

[Contact the developers](#) and tell us what you tried to do that didn't work.

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





PDFtoExcel.com

Premium

Desktop 

CONVERT PDF TO EXCEL

Convert Any PDF (scanned or regular) to Excel Instantly



UPLOAD

Drag & Drop Files



Choose from Dropbox



Choose from Google Drive



Open from OneDrive

Safe and Secure

All data uploaded to our servers will be used for conversion purposes only and deleted promptly after conversion. Your files are accessed only by automated software agents.



Excel's .xlsx Format

The new spreadsheet will be in .xlsx format ready to use in Microsoft Excel, LibreCalc, OpenCalc or any other office suite compatible with MS Office.



Quick and Easy

Full Document Conversion

The web app converts entire PDFs to Microsoft Excel. If your PDF file has multiple pages, rest assured that the full document will be converted to Excel free.



No Email

There is no need to leave an email address or other identifiable data. Upload your file and the document will be converted right on the page.



Drive, Dropbox and OneDrive

Exact Spreadsheet Recognition

All tables trapped in PDF will be extracted accurately. The rows and columns structure will be the same as in the original file, but fully editable and ready for reuse.



Extract Scanned Tables

Besides regular PDF files, the service works with scanned PDFs as well. Using the latest OCR technology, it accurately recognizes numbers and data.



Free

More Options with PDFtoEXCEL

CONVERT PDF TO EXCEL

Convert Any PDF (scanned or regular) to Excel Instantly

UPLOAD

Drag & Drop Files

Choose from Dropbox

Choose from Google Drive

Open from OneDrive



PDFtoExcel.com

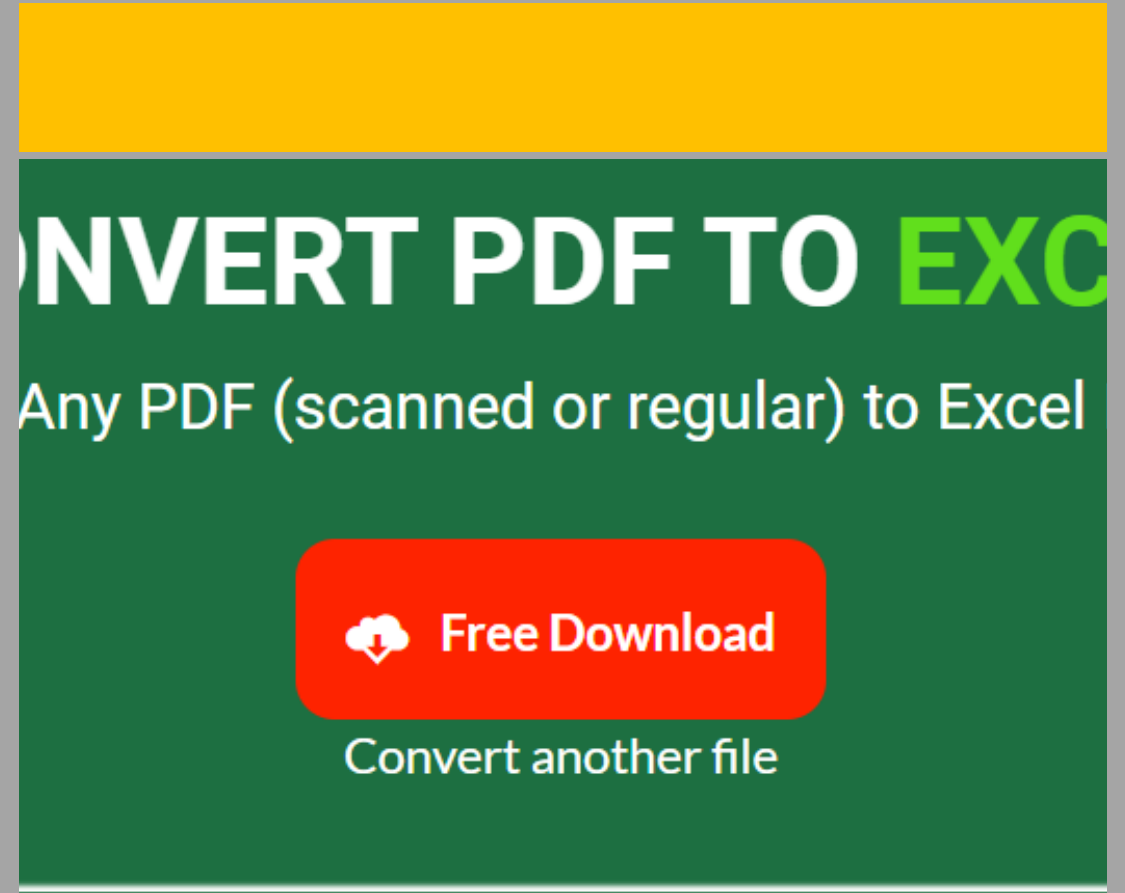
CONVERT PDF TO EXCEL

Convert Any PDF (scanned or regular) to Excel Instantly

Converting 

Our servers are currently busy with conversions. Your files are in the queue and will be converted soon.

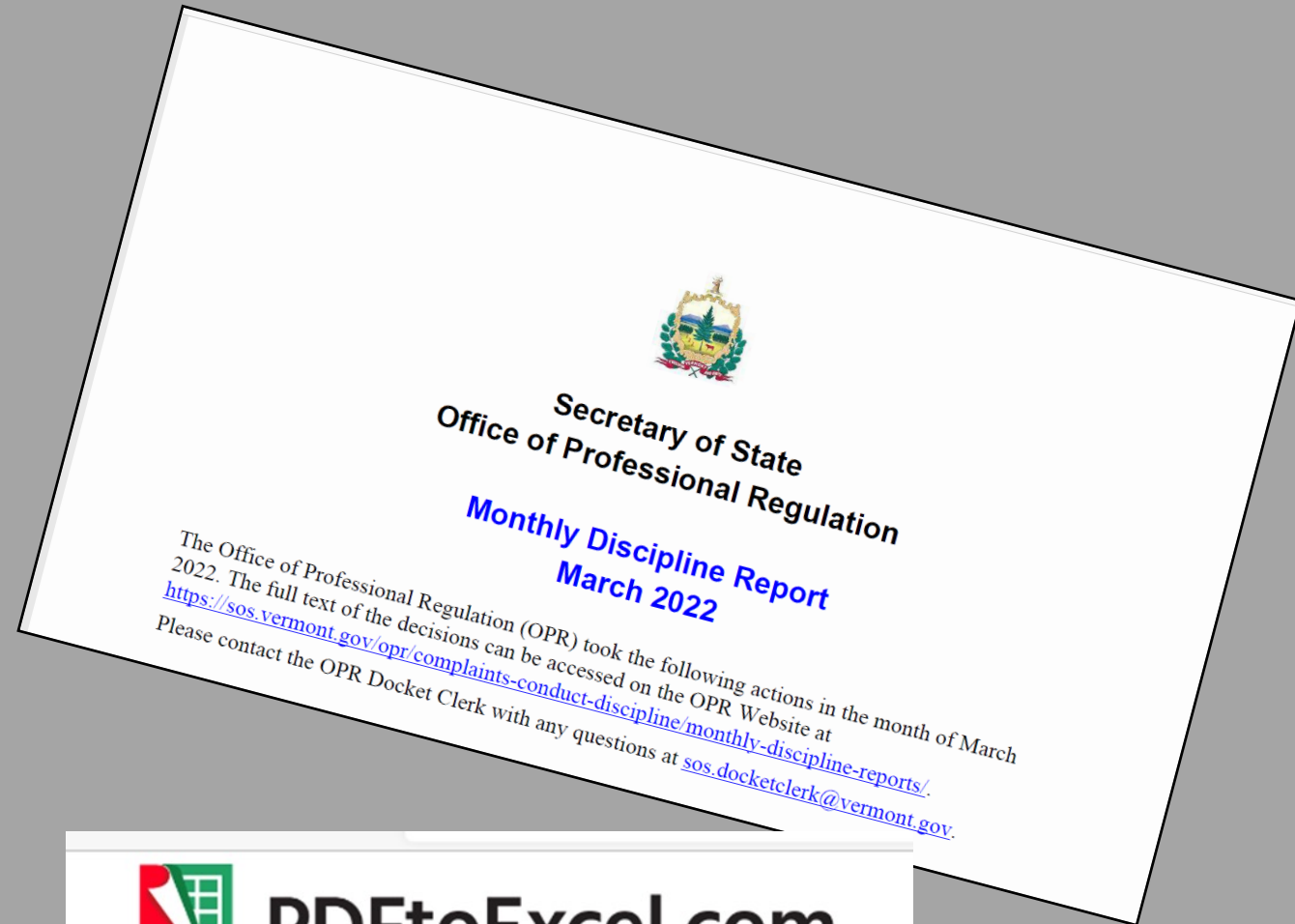
Once It's Complete, You'll See
The Option to Download Your File



Converting Data with PDFtoEXCEL

Upload the file
"March_2022_monthly
opr
discipline_report"

These are actions taken by the
Vermont Secretary of State
against licensed professionals



PDFtoExcel.com

Converting Data with PDFtoEXCEL

1									
2									
3	Secretary of State								
4	Office of Professional Regulation								
5									
6	Monthly Discipline Report								
7	March 2022								
8									
9	The Office of Professional Regulation (OPR) took the following actions in the month of March								
10	2022. The full text of the decisions can be accessed on the OPR Website at								
11	https://sos.vermont.gov/opr/complaints-conduct-discipline/monthly-discipline-reports/ .								
12									
13	Please contact the OPR Docket Clerk with any questions at sos.docketclerk@vermont.gov .								
14									
15	Tammy Larocque, St. Albans, Vermont	Dentist Assistant							
16	3/9/2022; License Summarily Suspended								
17	Pending further adjudication								
18									
19	Mark Faith, Morrisville, Vermont	Funeral Director							
20	3/10/2022; License Indefinitely Suspended	Embalmer							
21	Pending further adjudication								
22	Faith Funeral Home, Morrisville, Vermont	Funeral							
23	3/10/2022; License Indefinitely Suspended	Establishment							
24	Pending further adjudication								
25	Ivan Santiago, Portsmouth, New Hampshire	Licensed Practical							

Converting Data with PDFtoEXCEL

Now try to do
the same thing with the
file

“Maine Drugs 2018”



Maine Drug Enforcement Agency
Law Total Incident Report, by Location, Nature

Location: Androscoggin

<u>Nature of Incident</u>	<u>Total Incidents</u>
Detail - Information & Educati	2
Assist Federal Agency	1
Assist Local Police Department	45
Assist State Police	8
Assist Sheriff's Department	3
Conditional Release Violation	2
PD-Traffic Accident	2
Drug Intelligence Information	49
Criminal Intelligence Informat	23
Barbituate - Possession	1
Cocaine- Possession	16
Heroin- Possession	6



PDFtoExcel.com

Enforcement Agency

Report, by Location, Nature

Androscoggin

Nature of Incident

Total Incidents

Detail - Information & Educati

2

Assist Federal Agency

1

Assist Local Police Department

45

Assist State Police

8

Assist Sheriffs Department

3

Conditional Release Violation

2

PD-Traffic Accident

2

Drug Intelligence Information

49

Criminal Intelligence Informat

23

Barbituate - Possession

1

Cocaine- Possession

16

Heroin- Possession

6

Synthetic Narcotic- Possession

1

Public Relations - Talk or Lec

1

Cocaine- Sell

34

Heroin - Sell

15

Amphetamine - Sell

2

Synthetic Narcotic - Sell

4

Criminal Arrest Warrant

1

Total Incidents for This Location

216

Aroostook

Thank You!

Maggie Mulvihill

Associate
Professor of the
Practice in
Computational
Journalism

Boston University

mmulvih@bu.edu

MEET OUR JOURNALISM FELLOWS

12th Annual
**New England
First Amendment Institute**



October 23-25
Northeastern University

