

Steps To reproduce the results

DATA DEPENDENCIES

Monthly Data

You'll be downloading the necessary data for the month from the MA court website. The data folder of this file already contains some downloaded files. For the month of the data you're adding, you will need to create a new folder within this existing file structure (see below). Make sure the names of these folders have NO SPACES:

data

| -> 2017

| -> Jan

| -> Feb

| -> March

...

| -> 2016

| -> Jan

| -> Feb

...

Click this link to access the MA State Court website and enter the captcha to gain access:

<https://www.masscourts.org/eservices/home.page.2>

To search the website, you'll want to use the following options for Cambridge Courts:

Court Department -> District Court

Court Division -> Cambridge District Court

Court Location -> Cambridge District Court

Choose the "Case Type Tab":

Case Type -> Summary Process

City/Town -> Cambridge

Case Status -> All Statuses

Party Type -> Plaintiff

The screenshot shows the 'Search' page of the MA State Court website. The search criteria are as follows:

- Court Department:** District Court
- Court Division:** Cambridge District Court
- Court Location:** Cambridge District Court
- Number of Results:** 25
- Begin Date:** 04/01/2017
- End Date:** 04/30/2017
- Case Type:** Civil, Small Claims, Summary Process, Supplementary Process
- City/Town:** Brookline, Buckland, Burlington, Cambridge, Canton, Carlisle, Carver
- Case Status:** All Statuses, Activity Suspended, Closed, Closed Case Transferred, Disposed for Statistical Purposes, Filed, Open, Reopen
- Party Type:** All Party Types, Defendant, Plaintiff

Be sure to specify the begin and end date of the month you are adding. Only one month of data can be searched at a time.

Once you click Search, the results page should look like this:

Home Search Results Log

Search Results

Displaying all 31 matches.

Case Number	Case Type	File Date	Initiating Action	Party/Company	Party Type	Date of Birth	Case Status	Court
1752SU000167	Summary Process	04/03/2017	Summary Process - Residential (c239)	Federal Management Co., Inc., d/b/a Schochet Company, Inc., managing agent for Rindge Associates d/b/a Fresh Pond Apartments	Plaintiff		Disposed for Statistical Purposes	Cambridge District Court
1752SU000168	Summary Process	04/03/2017	Summary Process - Residential (c239)	Memorial Drive Housing LP	Plaintiff		Disposed for Statistical Purposes	Cambridge District Court
1752SU000171	Summary Process	04/10/2017	Summary Process - Residential (c239)	Fort Point Investments, LLC	Plaintiff		Disposed for Statistical Purposes	Cambridge District Court
1752SU000172	Summary Process	04/10/2017	Summary Process - Residential (c239)	Cascap, Inc.	Plaintiff		Disposed for Statistical Purposes	Cambridge District Court
1752SU000173	Summary Process	04/10/2017	Summary Process - Residential (c239)	Avalon North Point ,by and through its Managing agent,Avalonbay Communities,Inc.	Plaintiff		Disposed for Statistical Purposes	Cambridge District Court
1752SU000174	Summary Process	04/10/2017	Summary Process - Residential (c239)	Avalon North Point ,by and through its Managing agent,Avalonbay Communities,Inc.	Plaintiff		Disposed for Statistical Purposes	Cambridge District Court
1752SU000175	Summary Process	04/10/2017	Summary Process - Residential (c239)	Avalon North Point ,by and through its Managing agent,Avalonbay Communities,Inc.	Plaintiff		Disposed for Statistical Purposes	Cambridge District Court
1752SU000176	Summary Process	04/10/2017	Summary Process - Residential (c239)	Avalon North Point ,by and through its Managing agent,Avalonbay Communities,Inc.	Plaintiff		Disposed for Statistical Purposes	Cambridge District Court
1752SU000177	Summary Process	04/10/2017	Summary Process - Residential (c239)	Avalon North Point ,by and through its Managing agent,Avalonbay Communities,Inc.	Plaintiff		Disposed for Statistical Purposes	Cambridge District Court
1752SU000178	Summary Process	04/10/2017	Summary Process - Residential (c239)	Avalon North Point ,by and through its Managing agent,Avalonbay Communities,Inc.	Plaintiff		Disposed for Statistical Purposes	Cambridge District Court

Click on each case number to open the file and save each individual case to the month folder you've created (File->Save Page As, or Ctrl + S), choosing the format "Webpage, HTML only". The file name doesn't matter so long as it is unique to the folder and doesn't overwrite an existing file. If you accidentally save with the other file format, simply delete the folder it creates and resave with the correct format.

Running Scrapper to build initial csv from court dockets

Run scrapper.py

It will create initial data file in the form of csv

/csv/csvdata.csv

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
Case Number	Status	File Date	Plaintiff	P-Attorney	Defendant	D-Attorney	Property Add	Docket	Judgement C	Judgement T	JudgeMent N	Judgement T	Execution Total	
1752SU0001	Disposed for	3/20/17	Cambridge A Turk, Esq., Je Emile, Michel				30 Union Str	03/20/2017	3/31/17	Judgment fo	after defend	1,853.84	1,859.94,	
1752SU0001	Disposed for	3/17/17	Giller, David	Hegel, Esq., 'Esielomis, Ki	Hegel, Esq., '151 Auburn S		03/17/2017		3/17/17	Judgment fo	by agreement of the parties			
1752SU0001	Disposed for	3/13/17	Acorn Holdin	Turk, Esq., Je Patel, Pratick			7 Cameron A	03/13/2017	4/6/17	Not Held But	NA			
1752SU0001	Disposed for	3/20/17	Cambridge H Turk, Esq., Je Duverneau, Mederic				237 Franklin	03/20/2017	3/31/17	Judgment fo	after defend	235	235.77,	
1752SU0001	Disposed for	3/20/17	Cambridge H Turk, Esq., Je Regal, Mary				50 Churchill	03/20/2017	3/31/17	Judgment fo	after defend	1,001.76	1,005.06,	
1752SU0001	Disposed for	3/27/17	Cambridge C Krems, Esq., Rivera, Ralynda				90 Berkshire	03/27/2017	4/6/17	Notice of dis	NA			
1752SU0001	Disposed for	3/20/17	Watermark I Scolnick, Esq Voronovich, Alexander				350 Third Str	03/20/2017	3/30/17	Notice of dis	NA			
1752SU0001	Disposed for	3/27/17	54 Berkshire Turk, Esq., Je Cayman, Robert				54 Berkshire	03/27/2017	4/6/17	Judgment fo	by agreemer	677		
1752SU0001	Disposed for	3/13/17	Agor Market Turk, Esq., Je Fields, Lateisha C				25 Market St	03/13/2017	3/31/17	Judgment fo	after defendant(s) failed t	1.00,		
1752SU0001	Disposed for	3/13/17	Avalon North Quijano, Esq Aldossary, Saad				One Leightor	03/13/2017	3/27/17	Notice of dis	NA			
1752SU0001	Disposed for	3/20/17	Cambridge C Krems, Esq., Montoya, Sai Shachter, Es				20 Carlisle S	03/20/2017	5/10/17	Not Held But	NA			
1752SU0001	Disposed for	3/20/17	Cambridge H Turk, Esq., Je Maxwell, Richard				14 Roosevelt	03/20/2017	4/19/18	Judgment fo	by agreemer	75		
1752SU0001	Disposed for	3/20/17	Cambridge C Krems, Esq., Tessema, Getnet				1 Marcella S	03/20/2017	3/31/17	Judgment fo	after defend	631.89	633.97,	
1752SU0001	Disposed for	3/20/17	Cambridge H Turk, Esq., Je Horner, Benjamin				150 Erie Stre	03/20/2017	3/31/17	Judgment fo	after defend	987.71	990.96,	
1752SU0001	Disposed for	3/20/17	Scouting Wai Quijano, Esq Bartalini, Dina				10 Scouting	03/20/2017	3/30/17	Notice of dis	NA			

Clean up and adding additional information

1. Cleanup and removing duplicate

Drop duplicate case number by using pandas

```
df = df.drop_duplicates(subset='Case Number', keep="first")
```

Change the data type in judgement total and execution total to float by

- Replace coma(,) with "
- Change data type from string to Float

2. Adding Longitude & Latitude

From GeoMaps.py, call function addGeo(data)

The function will find the longitude and latitude based on address. It will append new 2 columns (longitude and latitude)

For address where the coordinates cannot be located, use google maps to find the coordinate

	Judgement Total	Execution Total	latitude	longitude
d to appear	1853.84	1859.94	42.3699653	-71.0971379
arties			42.36347410000001	-71.10460936
			42.3982092	-71.13025349
d to appear	235	235.77	42.36399119	-71.10409226
d to appear	1001.76	1005.06	42.40057355	-71.13163974
			42.3705102	-71.09054862
			42.36393082	-71.08354681
arties	677		42.3696425	-71.09104734
d to appear		0	42.3675306	-71.09468187
			42.3704771	-71.07341120000002

3. Adding district

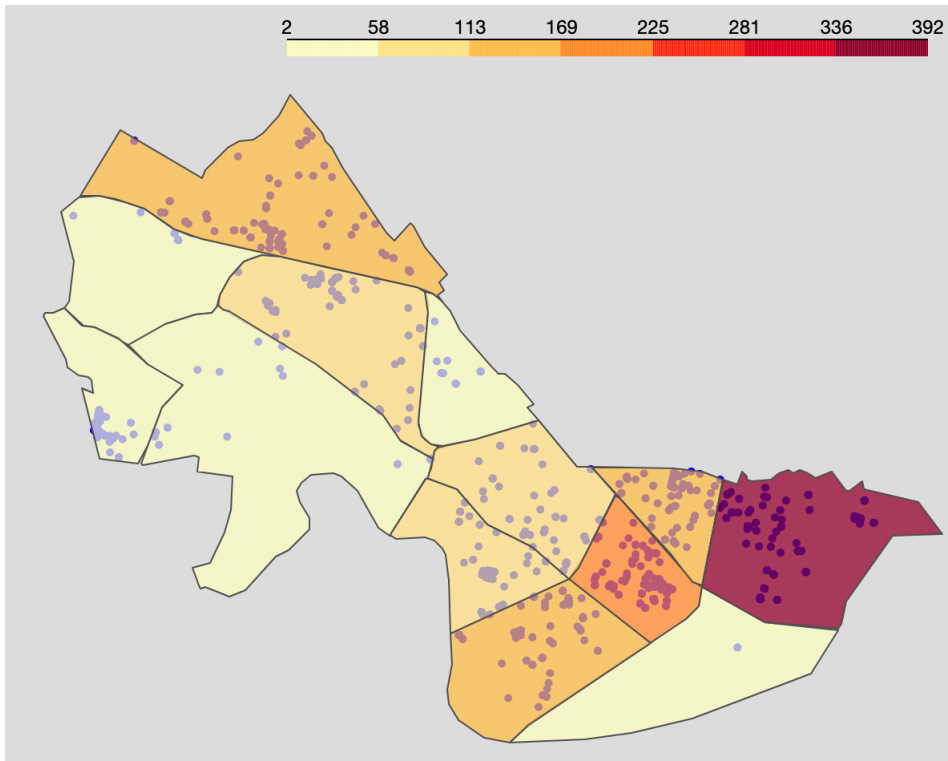
By combining with data from Neighbourhood polygon

From GeoMaps.py run addDistrict(data, districts)

DISTRICT
The Port
Cambridgeport
North Cambridge
Cambridgeport
North Cambridge
Wellington-Harrington
East Cambridge
Wellington-Harrington
The Port
East Cambridge

4. Creating Map of Cambridge with number of cases

From GeoMaps.py run createMapPlot(data, district)



5. Adding unit

The unit is the number of units that is in the building.

By using the data we have from previous step together with housing stat.csv data run findUnit.py

It will add another column Address (standardize address)and Units

Address	units
30 Union St	
151 Auburn St	
7 Cameron St	37
237 Franklin St	
50 Churchill St	
90 Berkshire St	

6. Adding Median Income

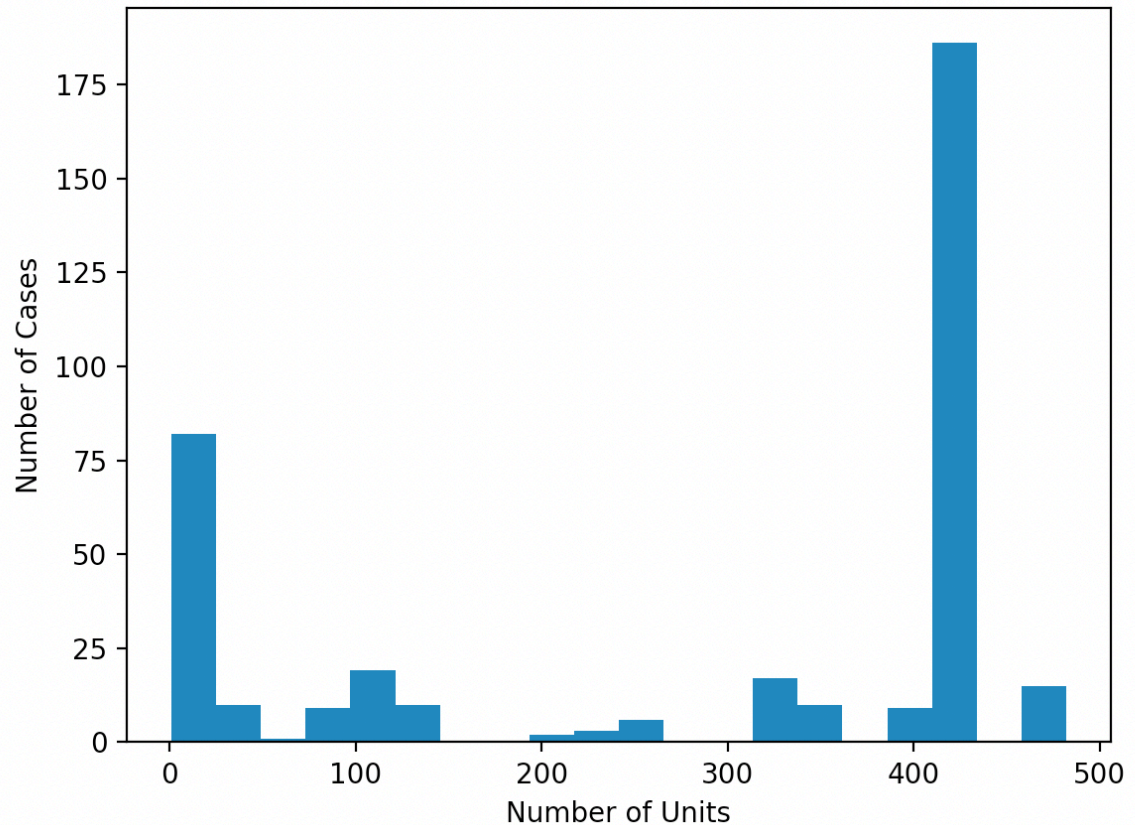
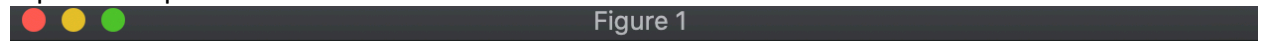
BY using previous data (csv file and American_Community_Survey_2013_-_17_Estimates_by_Neighborhood__Median_Income) run get_median_income.py

./csv/cleanData.csv is the final data we get from process 1 to 6

7. To do data science analysis (clustering)

Run clustering.py

It produces plot



8. To run data analysis by using Power BI

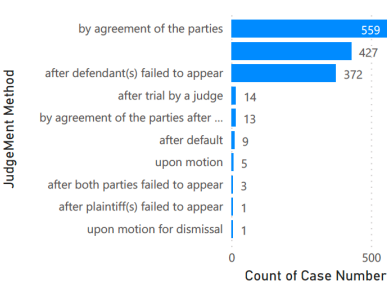
From the analysis folder, open Eviction.pbix in power BI desktop.

If there is data error, import the latest/cleanest csv file from process 1 to 6 to as the data source

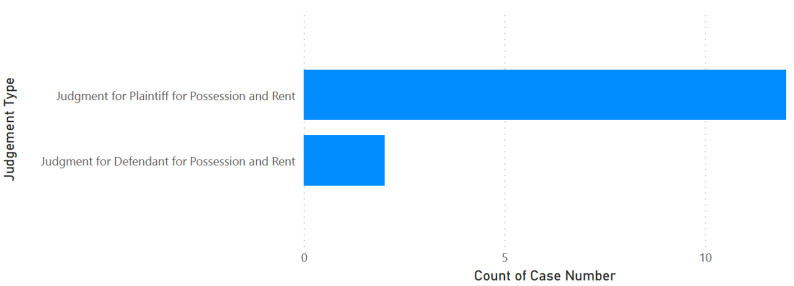
You might need to update the query in each report to reflect the data

The result of Power BI report is in analysis/data.pdf

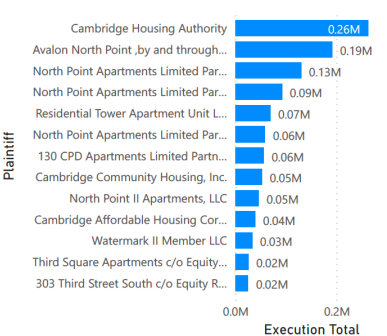
Count of Case Number by JudgeMent Method



Count of Case Number by Judgement Type



Execution Total by Plaintiff



Count of Case Number Quarterly

