



Hi! Welcome to CityMate! This is the README file for the instructions of installing and running the project.

【Youtube link of source code video】

<https://www.youtube.com/watch?v=hKihzSk-4PI>

【Github link of the project】

<https://github.com/CMU-DataFocusedPython-Group2/CityMate>

【Group Members】

Yixuan Guo yixuangu@andrew.cmu.edu
Yue Jia yjia2@andrew.cmu.edu
Yunxuan Yu yunxuany@andrew.cmu.edu
Zixin Yin zixinyin@andrew.cmu.edu
Jingwen Ma jingwenm@andrew.cmu.edu

【How to install manually】

1. pip install sodapy

Before you start running, please press Windows+R and type in “cmd” to open the command line window. And type “pip install sodapy” in the command line to install the sodapy module. Like this:

```
C:\Users\MAJIUWANG>pip install sodapy
WARNING: Retrying (Retry(total=4, connect=None, read=None, redirect=None, status=None)) after connection broken by 'ReadTimeoutError("HTTPConnectionPool(host='pypi.org', port=443): Read timed out. (read timeout=15)')': /simple/sodapy/
Collecting sodapy
  Using cached sodapy-2.1.0-py2.py3-none-any.whl (14 kB)
Requirement already satisfied: requests>=2.20.0 in d:\anaconda2020.7\lib\site-packages (from sodapy) (2.24.0)
Requirement already satisfied: certifi>=2017.4.17 in d:\anaconda2020.7\lib\site-packages (from requests>=2.20.0->sodapy) (2020.6.20)
Requirement already satisfied: chardet<4,>=3.0.2 in d:\anaconda2020.7\lib\site-packages (from requests>=2.20.0->sodapy) (3.0.4)
Requirement already satisfied: urllib3<1.25.0,!1.25.1,<1.26,>=1.21.1 in d:\anaconda2020.7\lib\site-packages (from requests>=2.20.0->sodapy) (1.25.9)
Requirement already satisfied: idna<3,>=2.5 in d:\anaconda2020.7\lib\site-packages (from requests>=2.20.0->sodapy) (2.10)
Installing collected packages: sodapy
Successfully installed sodapy-2.1.0
```

2. pip install shapely

Please type “pip install shapely” in the command line to install the shapely module. Like this:

```
C:\Users\MAJIUWANG>pip install shapely
Collecting shapely
  Using cached Shapely-1.7.1-cp38-cp38-win_amd64.whl (1.0 MB)
Installing collected packages: shapely
Successfully installed shapely-1.7.1
```

3. pip install wordcloud

Also, please type “pip install wordcloud” in the command line to install the wordcloud module. Like this:

```
C:\Users\MAJIUWANG>pip install wordcloud
Collecting wordcloud
  Using cached wordcloud-1.8.0-cp38-cp38-win_amd64.whl (159 kB)
Requirement already satisfied: pillow in d:\anaconda2020.7\lib\site-packages (from wordcloud) (7.2.0)
Requirement already satisfied: matplotlib in d:\anaconda2020.7\lib\site-packages (from wordcloud) (3.2.2)
Requirement already satisfied: numpy>=1.6.1 in d:\anaconda2020.7\lib\site-packages (from wordcloud) (1.18.5)
Requirement already satisfied: python-dateutil>=2.1 in d:\anaconda2020.7\lib\site-packages (from matplotlib->wordcloud) (2.8.1)
Requirement already satisfied: cycler>=0.10 in d:\anaconda2020.7\lib\site-packages (from matplotlib->wordcloud) (0.10.0)
Requirement already satisfied: pyparsing!=2.0.4,!2.1.2,!2.1.6,>=2.0.1 in d:\anaconda2020.7\lib\site-packages (from matplotlib->wordcloud) (2.4.7)
Requirement already satisfied: kiwisolver>=1.0.1 in d:\anaconda2020.7\lib\site-packages (from matplotlib->wordcloud) (1.2.0)
Requirement already satisfied: six>=1.5 in d:\anaconda2020.7\lib\site-packages (from python-dateutil->matplotlib->wordcloud) (1.15.0)
Installing collected packages: wordcloud
Successfully installed wordcloud-1.8.0
```

4. Add the file geos_c.dll to the installation path of Anaconda, in the folder of “Library/bin/”
Please find the geos_c.dll file in our project, its location is in “CityMate-main/src/geos_c.dll”.
Like this:

```

|---README.md
|---data
|---src
|   |---data_cleaning
|   |   |---clean_covid19.py
|   |   |---clean_crime.py
|   |   |---clean_house.py
|   |   |---clean_restaurant.py
|   |   |---clean_stops.py
|   |   |---clean_theater.py
|   |   |---update_data.py
|   |---geos_c.dll
|   |---house_surroundings.py
|   |---main.py

```

Then, please copy this geos_c.dll file and paste it to the installation path of Anaconda2020.7, in the folder of “Library/bin/”. Like this:

» Anaconda2020.7 » Library » bin »

名称	修改日期	类型	大小
fax2tiff	2020/6/1 22:13	应用程序	18 KB
fixqt4headers.pl	2018/10/16 21:37	PL 文件	7 KB
freetype.dll	2020/6/19 0:26	应用程序扩展	619 KB
genbrk	2020/4/28 6:18	应用程序	21 KB
genccode	2020/4/28 6:18	应用程序	12 KB
gencfu	2020/4/28 6:18	应用程序	17 KB
gencmn	2020/4/28 6:18	应用程序	12 KB
gencnval	2020/4/28 6:18	应用程序	26 KB
gendict	2020/4/28 6:18	应用程序	28 KB
gennorm2	2020/4/28 6:18	应用程序	51 KB
genrb	2020/4/28 6:18	应用程序	135 KB
gensprep	2020/4/28 6:18	应用程序	25 KB
geos_c.dll	2019/8/12 8:14	应用程序扩展	2,700 KB
gif2h5	2018/12/20 2:37	应用程序	2,323 KB
gmp.dll	2017/11/22 4:00	应用程序扩展	500 KB

【How to run】

Now you are well set! Please follow the instructions below to run the CityMate project.

1. How to run main.py

Please find and open main.py in Anaconda, its location is in “CityMate-main/src/main.py”.

```

|---README.md
|---data
|---src
|   |---data_cleaning
|   |   |---clean_covid19.py
|   |   |---clean_crime.py
|   |   |---clean_house.py
|   |   |---clean_restaurant.py
|   |   |---clean_stops.py
|   |   |---clean_theater.py
|   |   |---update_data.py
|   |---geos_c.dll
|   |---house_surroundings.py
|   |---main.py

```

Then, click run the module to begin!

Hi, Welcome to CityMate's HouseRent Service!

Please specify your University, where do you want to live nearby?

1. Cornell University
2. Columbia University
3. New York University
4. University of Rochester
5. Rensselaer Polytechnic Institute
6. Syracuse University
7. Fordham University
8. Yeshiva University
9. Binghamton University
10. The New School
11. Clarkson University
12. Hofstra University
13. City University of New York
14. Stevens Institute of Technology
15. St.John's University

Please enter the choice number(1-15) of the university:

2

First of all, you would see the request for you to **choose a university in New York State where you want to live by**. In this case, we choose 2 (Columbia University) and press enter.

You would like to rent near Columbia University

.....Now we are preparing data for you.....

We have already updated data on 10/11/2020

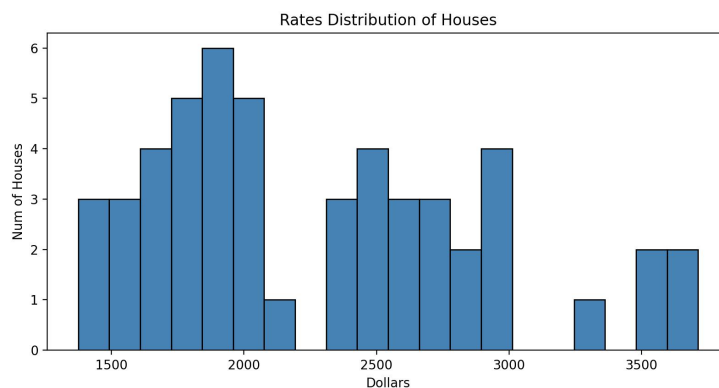
Do you need to update local data (scraping again from sites)?

This may take a few hours.

If you want to update local data, press Y,
or press N for displaying rent information.

Please enter your input: N

Then, you would be asked whether to update all the local data sources. In this case, the data was already updated on 10/11/2020, and the updating process would cost you some time due to the web scraping and API requests process. **We highly recommend that you press "N" and then press enter here**. However, if you want to update all the data source, please follow the instruction of "How to run update_data.py" later on.



After pressing enter, you will see a picture pop up showing the histogram of house monthly rates distribution of 50 houses around the University. Please close this picture then to continue the running of main.

index	name	price	streetAddress	postcode	house_type
1	889	\$2,458	W 114th St	10026	Apartment
2	310	\$1,575	West 123 Street	10027	Apartment
3	671	\$1,925	nan	10027	Apartment
4	213	\$2,400	31 Tiemann Place,	43	10027 Apartment
5	214	\$2,800	31 Tiemann Place,	31	10027 Apartment
6	60	\$2,000	Broadway	10115	Apartment
7	210	\$3,295	219B West 116th Street	10026	Apartment
8	565	\$1,990	West 136th St.	10031	Apartment
9	209	\$2,450	140 Claremont Avenue	10115	Apartment
10	212	\$2,600	140 Claremont Avenue	10115	Apartment
11	94	\$2,450	136th street	10031	Apartment
12	623	\$1,700	118th Street	10026	Apartment
13	61	\$2,400	118th Street	10026	Apartment
14	57	\$2,400	118th Street	10026	Apartment
15	113	\$2,775	84th	10028	Apartment
16	451	\$2,770	84th	10028	Apartment
17	578	\$2,690	88	10128	Apartment
18	311	\$1,375	West 142 Street	10031	Apartment
19	961	\$1,375	West 142 Street	10031	Apartment
20	90	\$1,375	West 142 Street	10031	Apartment

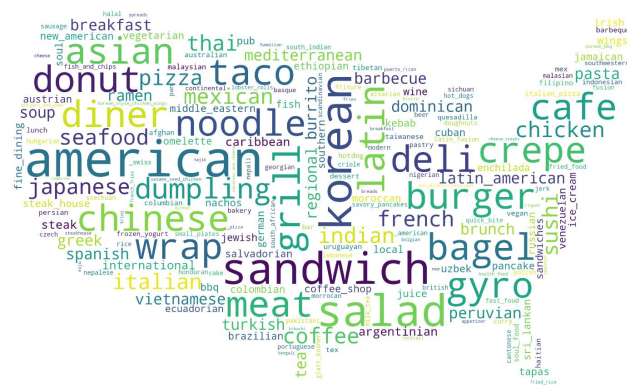
21	567	\$1,866	81 Street	10075	Apartment
22	570	\$1,733	81st Street	10075	Apartment
23	569	\$1,733	81 Street	10075	Apartment
24	130	\$1,900	nan	10031	Apartment
25	144	\$3,713	nan	10031	Apartment
26	91	\$3,529	3620 Broadway	10031	Apartment
27	709	\$1,700	nan	10031	Apartment
28	154	\$3,644	nan	10031	Apartment
29	695	\$2,000	nan	10031	Apartment
30	157	\$2,590	nan	10031	Apartment
31	687	\$1,925	nan	10031	Apartment
32	534	\$1,971	nan	10031	Apartment
33	436	\$3,500	nan	10028	Apartment
34	341	\$2600.00	nan	10128	Apartment
35	597	\$1,732	nan	10128	Apartment
36	786	\$2,500	E 83 Street	10028	Apartment
37	459	\$1,940	81st street	10028	Apartment
38	772	\$3,000	3 Ave	10128	Apartment
39	918	\$2,950	700 Riverside Drive	10031	Apartment
40	494	\$1,600	336 EAST 73 STREET	#1F 10021	Apartment
41	59	\$2,100	101st Street	10029	Apartment
42	194	\$1,675	nan	10021	Apartment
43	849	\$1,650	71st st	10021	Apartment
44	432	\$2990.00	nan	10021	Apartment
45	649	\$2000.00	nan	10128	Apartment

46	649	\$2000.00	nan	10128	Apartment
47	430	\$2,995	off West End Ave	10025	Apartment
48	763	\$1,775	316 E89th St	10128	Apartment
49	353	\$1,600	nan	10026	Apartment
50	413	\$1,950	off West End Ave	10025	Apartment

Which house do you want to know more about?

Please enter the index number(1-50) for more information:
Enter 0 to exit

10



After that, you would see a list of 50 houses which are near to Columbia University, and a wordcloud picture of all kinds of cuisines nearby the house is displayed. Please then close the wordcloud picture to continue and **enter an index number of those 50 houses and press enter** to know more about other information about that house. In this case, we entered 10.

house description: Large One Bedroom / One Bath - On Sakura Park, Steps from Columbia!!
month price: \$2,450
street address: 140 Claremont Avenue
house type: Apartment
postcode: 10115

Which of the following do you want to know more about this house?

1. Distance from nearest subway
2. 5 Nearby Restaurants
3. 3 Nearby Theaters
4. COVID19 data in past 4 weeks
5. Crime Report in past 5 years

Please enter the choice number(1-5):

Press 0 for quit

1

Then, you would see a brief description of that house, and request for you to choose from the 5 fields of interest to know more about that house. Firstly, please choose "1" and press enter to see the information of "Distance from nearest subway" of that house.

The nearest subway is 116 St-Columbia University and is 327 meters away.

Please enter the choice number(1-5):

Press 0 for quit

2

The distance from nearest subway is shown and please choose "2" and press enter this time to see the information of "5 Nearby Restaurants".

The most popular cuisine is: japanese

Number restaurant within 500m: 7

The 5 nearest restaurants are:

name	cuisine	street	phone
Dinosaur BBQ	barbecue	West 125th Street	+1 212 694-1777
Shake Shack	burger	Broadway	+1 212 932-3300
Vine Sushi & Sake	japanese	Broadway	+1-212-222-3568
Go! Go! Curry!	japanese	West 125th Street	+1-646-833-7143
Ichie	japanese	West 106th Street	+1-212-865-4888

Please enter the choice number(1-5):

Press 0 for quit

3

The 5 nearby restaurants are shown and please choose "3" and press enter this time to see the information of "3 Nearby Theaters".

The 3 nearest theaters are:

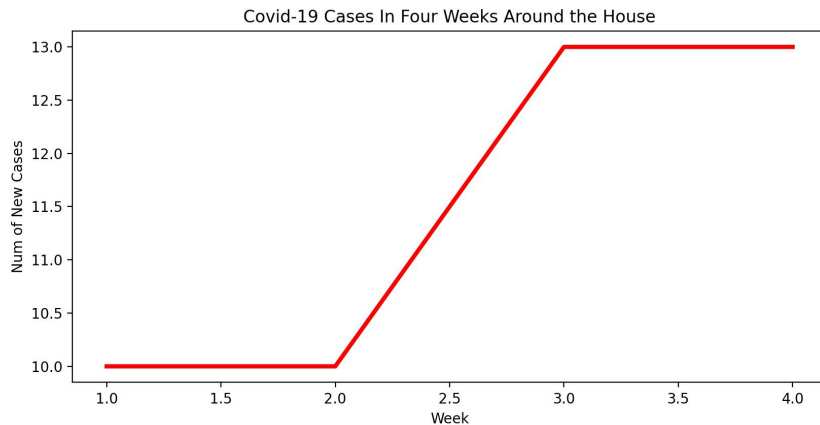
name	tel	url	address	zip
Miller Theatre			at Columbia University	(212) 854-7799
60 Broadway			10027	http://www.millertheatre.com/
Apollo Theater	(212) 531-5300	http://www.apollotheater.org/	253 West 125th Street	027
Delacorte Theater	(212) 861-7277	http://www.centralpark.com/pages/attractions/orte-theatre.html	Central Park - Mid-Park at 80th Street	0

Please enter the choice number(1-5):

Press 0 for quit

4

The 3 nearby theaters are shown and please then choose "4" and press enter this time to see the information of "COVID19 data in past 4 weeks".



```
In the past four weeks, the COVID 19 data of here
Central Harlem (South)/Morningside Heights/West Harlemlis:
Total case count:  46
Total case rate:  72.29
Total death count: 1
Total death rate: 1.57
Number of people tested: 6604
Tested positive rate: 0.7
Case counts change: [10, 10, 13, 13]
```

```
Please enter the choice number(1-5):
Press 0 for quit
```

5

A line chart of covid19 cases around the house in the last four weeks is displayed and the covid19 data in the past 4 weeks are also listed here. However, you may encounter the problem of Error 503 here, since the website for covid19 data sometimes go through a website update process. We hope it won't happen when you run this, but if it happens, we are very sorry about that and please just comment the lines for obtaining covid19 data like below in main.py for you to go through the following processes successfully.

```
# update covid19 data
print("")
print("Updating COVID19 data...")
#COVID19_DF = GET_COVID19_DF()
#COVID19_DF.to_csv('../data/updated_data/covid19_clean.csv', index=False)
print("COVID19 data updated.")
```

If there are no problems, please close the line chart to continue and choose "5" and press enter this time to see the information of "Crime report in past 5 years".

```
The house is in Precinct No. 26
Crime rate of the precinct: 0.3%
Borough of the precinct:  MANHATTAN
Address of the precinct:  520 West 126th Street
```

```
Please enter the choice number(1-5):
Press 0 for quit
```

0

```
Which house do you want to know more about?
Please enter the index number(1-50) for more information:
Enter 0 to exit
```

0

The crime data in the past 5 years of the precinct the house is in were shown. And finally, please enter "0"s for quit.

```
Thank you for using our CityMate service, if you find our service useful, please recommend it to others!:)
>>>
```

The journey of main.py ends now! Hope you had a nice time.

2. How to run update_data.py (Not recommended, because it would take some time)

After running main.py, if you want to update all the data sources, please open update_data.py file in Anaconda, its location is in "CityMate-main/src/data_cleaning/update_data.py":

```
├── README.md
├── data
└── src
    ├── data_cleaning
    │   ├── clean_covid19.py
    │   ├── clean_crime.py
    │   ├── clean_house.py
    │   ├── clean_restaurant.py
    │   ├── clean_stops.py
    │   ├── clean_theater.py
    │   └── update_data.py
    ├── geos_c.dll
    ├── house_surroundings.py
    └── main.py
```

Then, click run the module to begin!

```
Please wait a few minutes here...
```

```
Updating crime data...
```

You would need to wait a few minutes here, and when you finished, you will be shown like this:

```
Please wait a few minutes here...
```

```
Updating crime data...
```

```
Crime data updated.
```

```
Updating subway stops data...
```

```
Subway stops data updated.
```

```
Updating restaurants data...
```

```
Restaurants data updated.
```

```
Updating theaters data...
```

```
Theaters data updated.
```

```
Updating COVID19 data...
```

```
COVID19 data updated.
```

```
Merging house data... This may take 6 or 7 minutes
```

```
Merging completed.
```

```
All update process completed.
```

```
>>>
```

Afterwards, you may find all the updated and cleaned data source in the folder of

"CityMate-main/data/updated_data/":

```
|---README.md
|---data
|   |---raw_data
|   |   |---crime_raw.txt
|   |   |---restaurant_raw.csv
|   |   |---stops.txt
|   |   |---theater_raw.csv
|   |---updated_data
|   |   |---covid19_clean.csv
|   |   |---crime_clean.csv
|   |   |---house_merged.csv
|   |   |---restaurant_clean.csv
|   |   |---substops_clean.csv
|   |   |---theater_clean.csv
|   |---CityMateLogo.png
|   |---clean_house_data.xlsx
|   |---sites.txt
|---src
|   |---data_cleaning
|   |   |---clean_covid19.py
|   |   |---clean_crime.py
|   |   |---clean_house.py
|   |   |---clean_restaurant.py
|   |   |---clean_stops.py
|   |   |---clean_theater.py
|   |   |---update_data.py
|   |---geos_c.dll
|   |---house_surroundings.py
|   |---main.py
```

That's all for the installation and run process, hope you had a nice time!