

Peer-graded Assignment: Capstone Project - The Battle of Neighborhoods

Moving to Houston, Texas, USA

Amir Kassenev

Contents

1.1	Introduction: Business Problem	2
2.1	Data	3
3.1	Methodology	3
4.1	Results	4
5.1	Discussion.....	9
6.1	Conclusion	9

1.1 Introduction: Business Problem

In this project, I will be using my persona real live case.

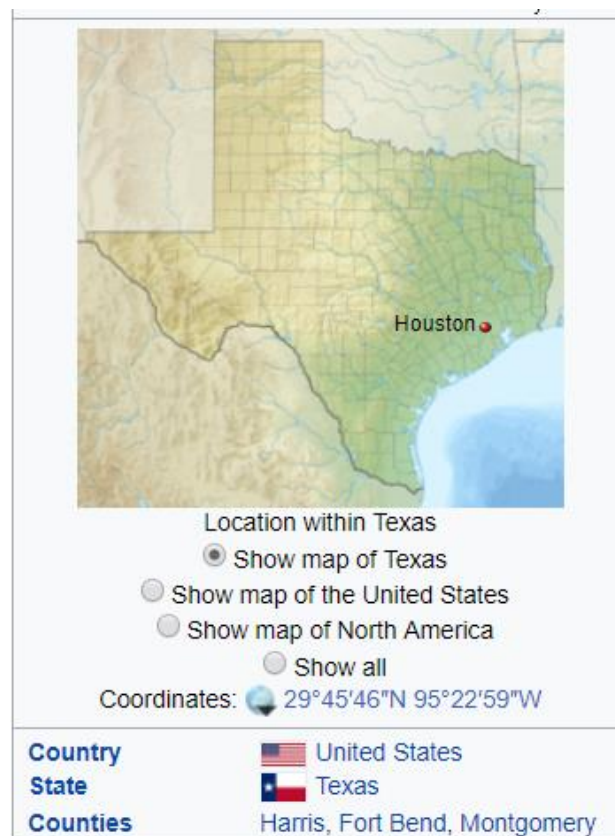
I am planning to move to Houston, Texas, U.S. in next few months and looking for an area to buy a house or apartment. This case will be useful for anyone planning to move to a new area/city and looking for an area to buy a house.

Since Houston is the most populous city in the U.S. state of Texas and the fourth most populous city in the United States with the area of 627 sq mi, it is a quite challenging task to find the right place to settle.

We will be looking for an area/neighborhood with the following requirements:

1. Neighborhood with the best schools *
2. Neighborhood with low crime rate
3. As close to the city center as possible

We will use our data science powers to generate a few most promising neighborhoods based on these criteria. Advantages of each area will then be clearly expressed so that the best possible final location can be chosen by Family.



2.1 Data

Based on definition of our task, factors that will influence our decision are:

number of Schools in the neighborhood (any type of schools) distance

neighborhood crime rate

distance of neighborhood from city center

We decided to use regularly spaced grid of locations, centered around city center, to define our neighborhoods.

Following data sources will be needed to extract/generate the required information:

Wikipedia - list of Houston neighborhoods (https://en.wikipedia.org/wiki/List_of_Houston_neighborhoods)

Crime Statistics in Houston ranking - <https://www.neighborhoodscout.com/tx/houston/crime>

Best Schools in Houston ranking - <https://www.neighborhoodscout.com/tx/houston/schools>

3.1 Methodology

The key library that was used in this project is Beautiful Soup – to scrap data from the web sites

Advanced table data manipulation was used to combine results from several requests (each to meet specific criteria) and results in final list of neighborhood

Foursquare was used to get Chinese restaurants close to the neighborhood

Folium was used to demonstrate results on the map

4.1 Results

List of neighborhood was taken first from Wikipedia

URL = "https://en.wikipedia.org/wiki/List_of_Houston_neighborhoods" #reading Houston Neighborhoods from Wikipedia

```

                                Name
0          Willowbrook\n
1      Greater Greenspoint\n
2          Carverdale\n
3 Fairbanks / Northwest Crossing\n
4          Greater Inwood\n
5          Acres Home\n
6          Hidden Valley\n
7          Westbranch\n
8      Addicks / Park Ten\n
9      Spring Branch West\n
(88, 1)
```

Houston coordinates were taking using Geocoders

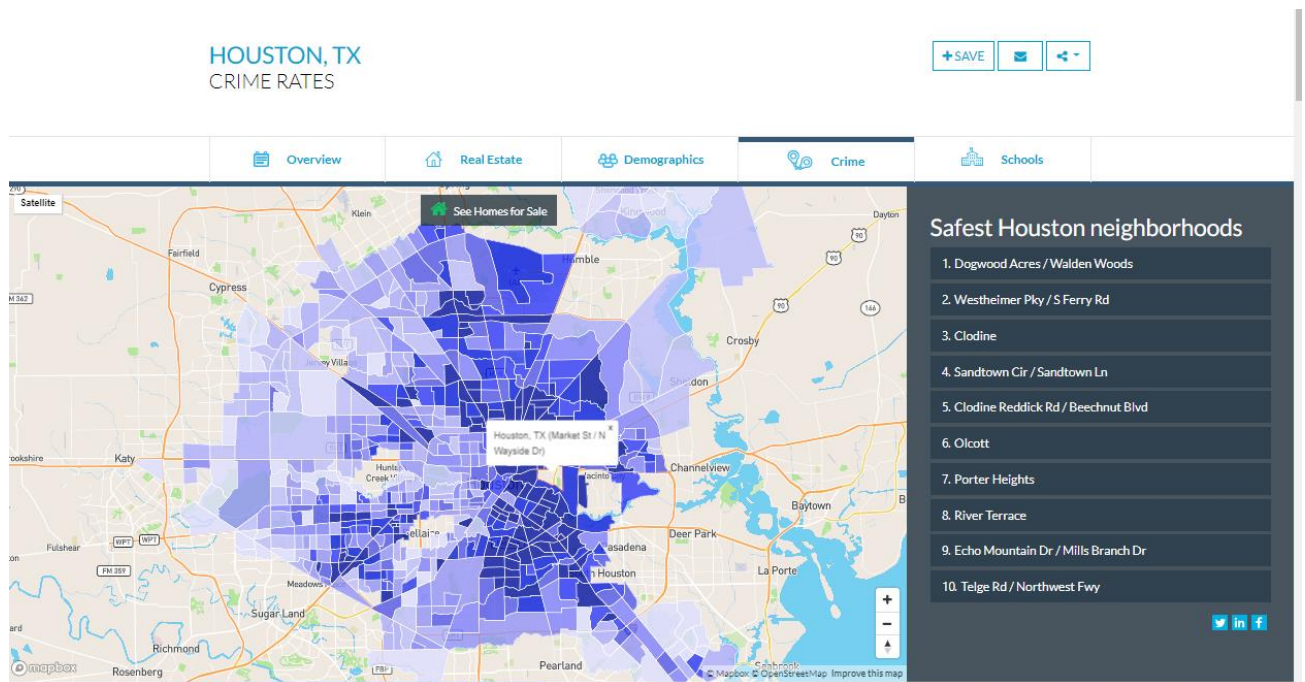
```
from geopy.geocoders import Nominatim
address = 'Houston, TX'

geolocator = Nominatim(user_agent="ny_explorer")
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude
print('The georapical coordinate of Houston are {}, {}'.format(latitude, longitude))
```

The georapical coordinate of Houston are 29.7589382. -95.3676974.

Results – safest neighborhood

- From www.neighborhoodscout.com/tx/houston we scrapped best 10 neighborhood with lowest crime rate



Python results:

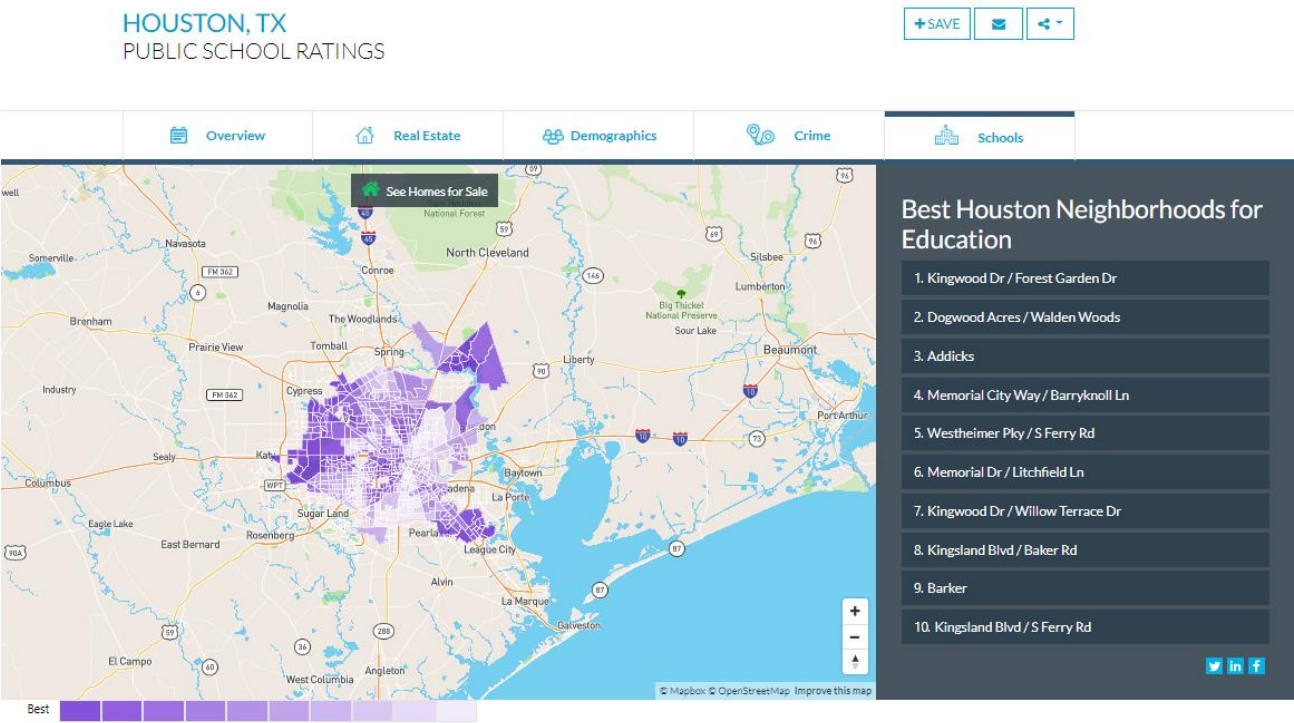
```
print(Houston_schools)
```

```

Neighborhoods
0    Kingwood Dr / Forest Garden Dr
1    Dogwood Acres / Walden Woods
2    Addicks
3    Memorial City Way / Barryknoll Ln
4    Westheimer Pky / S Ferry Rd
5    Memorial Dr / Litchfield Ln
6    Kingwood Dr / Willow Terrace Dr
7    Kingsland Blvd / Baker Rd
8    Barker
9    Kingsland Blvd / S Ferry Rd

```

Results – best schools



```
print(Houston_Neighbourhood)
```

```
Neighborhoods
0      Dogwood Acres / Walden Woods
1      Westheimer Pky / S Ferry Rd
2      Clodine
3      Sandtown Cir / Sandtown Ln
4      Clodine Reddick Rd / Beechnut Blvd
5      Olcott
6      Porter Heights
7      River Terrace
8      Echo Mountain Dr / Mills Branch Dr
9      Telge Rd / Northwest Fwy
```

Results – merge two tables

```
pd.concat([Houston_schools, Houston_Neighbourhood], axis=1)
```

```
]:
```

	Neighborhoods	Neighborhoods
0	Kingwood Dr / Forest Garden Dr	Dogwood Acres / Walden Woods
1	Dogwood Acres / Walden Woods	Westheimer Pky / S Ferry Rd
2	Addicks	Clodine
3	Memorial City Way / Barryknoll Ln	Sandtown Cir / Sandtown Ln
4	Westheimer Pky / S Ferry Rd	Clodine Reddick Rd / Beechnut Blvd
5	Memorial Dr / Litchfield Ln	Olcott
6	Kingwood Dr / Willow Terrace Dr	Porter Heights
7	Kingsland Blvd / Baker Rd	River Terrace
8	Barker	Echo Mountain Dr / Mills Branch Dr
9	Kingsland Blvd / S Ferry Rd	Telge Rd / Northwest Fwy

```
merged=pd.merge(Houston_schools, Houston_Neighbourhood,
on='Neighborhoods', how='inner')
merged
```

```
]:
```

	Neighborhoods
0	Dogwood Acres / Walden Woods
1	Westheimer Pky / S Ferry Rd

Merged:
Neighborhood with Best schools and safest!

Safest neighborhood Neighborhood with Best schools

Results – split Neighborhood name and street

- Split Neighborhood name and street into two columns as joint name will not work to find coordinates

```
merged = pd.DataFrame(merged.Neighborhoods.str.split('/').tolist(),
                        columns = ['Neighborhoods','street'])
merged
```

```
3]:
```

	Neighborhoods	street
0	Dogwood Acres	Walden Woods
1	Westheimer Pky	S Ferry Rd

Results – prepare table

- ▶ Add empty columns with coordinates (latitude and longitude)
- ▶ Add state column (Houston)
- ▶ Add new column (Neighborhood + State) that will be used in getting coordinates

```
merged ['latitude']=0
merged ['longitude']=0
merged ['state']='Houston'
merged ['get_coordinates']=merged ['Neighborhoods']+merged ['state']
merged
```

```
]:
```

	Neighborhoods	street	latitude	longitude	state	get_coordinates
0	Dogwood Acres	Walden Woods	0	0	Houston	Dogwood Acres Houston
1	Westheimer Pky	S Ferry Rd	0	0	Houston	Westheimer Pky Houston

Results – Coordinates are obtained, lets move forward

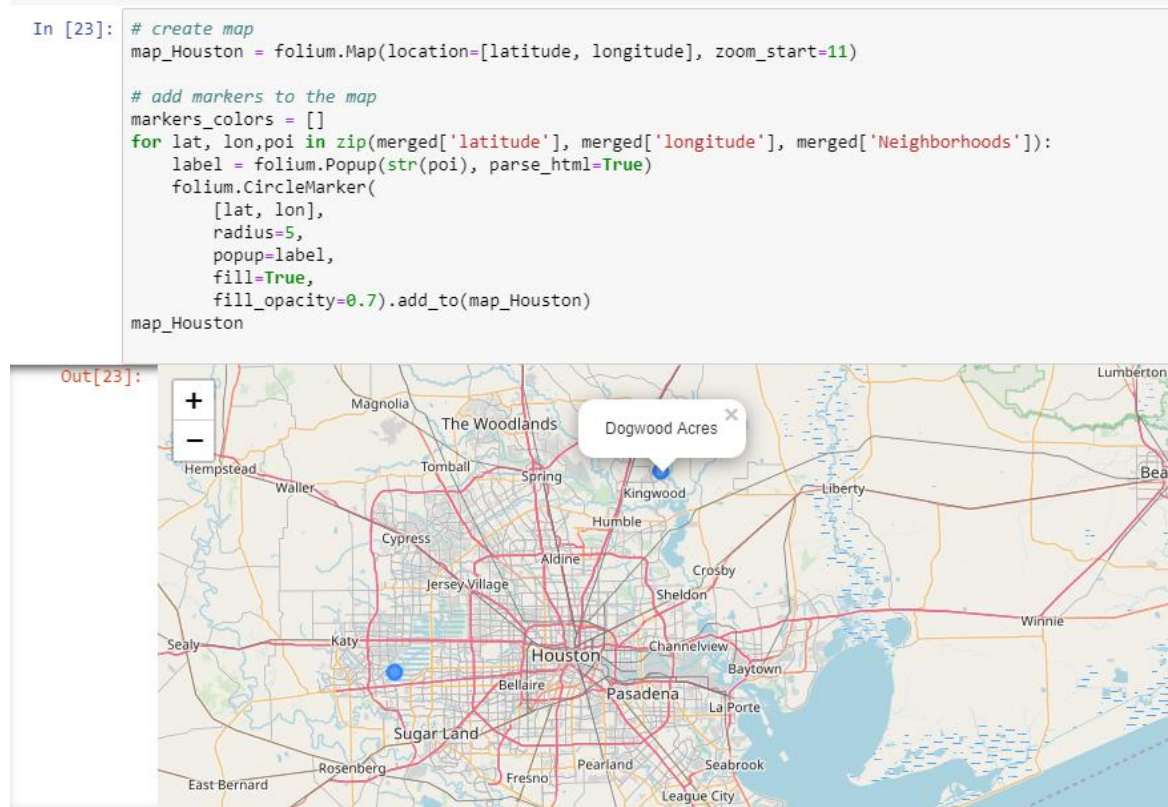
```
row=0
#or index, row in merged.iterrows():
while row<len(merged):
    #address = merged['get_coordinates']
    address=merged.loc[merged.index[row], 'get_coordinates']
    #print(address)
    #print('dine')
    geolocator = Nominatim(user_agent="ny_explorer")
    location = geolocator.geocode(address)
    latitude = location.latitude
    longitude = location.longitude
    print('The geographical coordinate of {}, {},{}'.format(address,latitude, longitude))
    merged.loc[merged.index[row], 'latitude']=latitude
    merged.loc[merged.index[row], 'longitude']=longitude
    row=row+1
merged
```

The geographical coordinate of Dogwood Acres Houston, 30.0891062,-95.1727095.
The geographical coordinate of Westheimer Pky Houston, 29.7292982,-95.7198165.

```
]:
```

	Neighborhoods	street	latitude	longitude	state	get_coordinates
0	Dogwood Acres	Walden Woods	30.089106	-95.172709	Houston	Dogwood Acres Houston
1	Westheimer Pky	S Ferry Rd	29.729298	-95.719816	Houston	Westheimer Pky Houston

Results – map is here, lets move then!



5.1 Discussion

Both neighborhood that resulted from this exercise are quite far from the city center therefore additional KPI should be considered, as home pricing, community and etc. This will be taking into account when we will visit Houston next month.

6.1 Conclusion

- ▶ In conclusion, this was very useful final project where Python skills helped me find best neighborhood as I am planning to move to Houston in reality.
- ▶ Data scrapping from the website and Fore Square API is very powerful technique
- ▶ Data visualization using mapping tools as folium is an amazing to demonstrate to my wife 😊