

Week 5 FINAL

August 26, 2020

1 COURSERA IBM DATA SCIENCE CAPSTONE

1.1 Buying a Home in Edgewater, NJ

By: Anna Rivera

August 2020

```
[4]: from PIL import Image  
img = Image.open('edgewater.jpg')  
img
```

[4]:



INTRODUCTION

Just across the Hudson River and facing New York City is a small yet diverse and thriving town named Edgewater, NJ. Due to its close proximity to New York, it has the advantage of attracting New Yorkers to hop across the river for more bang for their buck. The town also has continuous new property developments which is another great reason to buy property in Edgewater, NJ.

BUSINESS CHALLENGE

The purpose of this project is to review data & analyze information to determine property investment in Edgewater, NJ. Methods learned throughout this course will be used to provide a convincing argument as to why this location is ideal to invest in real estate.

TARGET AUDIENCE

This information is geared towards those wanting to purchase a residence or those seeking investment properties.

DATA

To have a compelling argument, the following information is provided:

- Listings of properties showcasing their value
- Coordinates of the location
- Map of the location

DATA EXTRACTION/DATA SOURCE

- Extract map image from Wikipedia
- Using Geocoder Package to attain coordinates data
- Data scraping DataUSA to gather more in-depth details about Edgewater, NJ

1. IMPORT NECESSARY LIBRARIES

```
[79]: import pandas as pd
import requests
pd.set_option("display.max_columns", None)
pd.set_option("display.max_rows", None)
import numpy as np
import folium
import json
from pandas.io.json import json_normalize
import matplotlib.cm as cm
import matplotlib.colors as colors
from sklearn.cluster import KMeans
! pip install geopy
from geopy.geocoders import Nominatim
! pip install geocoder
import geocoder
```

Collecting geopy

Downloading <https://files.pythonhosted.org/packages/07/e1/9c72de674d5c2b8fcb0738a5ceeb5424941fefa080bfe4e240d0bacb5a38/geopy-2.0.0-py3-none-any.whl>

```

(111kB)
|                               | 112kB 6.9MB/s eta 0:00:01
Collecting geographiclib<2,>=1.49 (from geopy)
  Downloading https://files.pythonhosted.org/packages/8b/62/26ec95a98ba642991631
99e95ad1b0e34ad3f4e176e221c40245f211e425/geographiclib-1.50-py3-none-any.whl
Installing collected packages: geographiclib, geopy
Successfully installed geographiclib-1.50 geopy-2.0.0
Collecting geocoder
  Downloading https://files.pythonhosted.org/packages/4f/6b/13166c909ad2f2
d76b929a4227c952630ebaf0d729f6317eb09cbceccbab/geocoder-1.38.1-py2.py3-none-
any.whl (98kB)
|                               | 102kB 5.3MB/s ta 0:00:011
Collecting ratelim (from geocoder)
  Downloading https://files.pythonhosted.org/packages/f2/98/7e6d147fd16a10a5f821
db6e25f192265d6ecca3d82957a4fdd592cad49c/ratelim-0.1.6-py2.py3-none-any.whl
Requirement already satisfied: six in
/home/jupyterlab/conda/envs/python/lib/python3.6/site-packages (from geocoder)
(1.15.0)
Collecting future (from geocoder)
  Downloading https://files.pythonhosted.org/packages/45/0b/38b06fd9b92dc2
b68d58b75f900e97884c45bedd2ff83203d933cf5851c9/future-0.18.2.tar.gz (829kB)
|                               | 829kB 13.1MB/s eta 0:00:01
Collecting click (from geocoder)
  Using cached https://files.pythonhosted.org/packages/d2/3d/fa76db83bf75c4f8d33
8c2fd15c8d33fdd7ad23a9b5e57eb6c5de26b430e/click-7.1.2-py2.py3-none-any.whl
Requirement already satisfied: requests in
/home/jupyterlab/conda/envs/python/lib/python3.6/site-packages (from geocoder)
(2.24.0)
Requirement already satisfied: decorator in
/home/jupyterlab/conda/envs/python/lib/python3.6/site-packages (from
ratelim->geocoder) (4.4.2)
Requirement already satisfied: idna<3,>=2.5 in
/home/jupyterlab/conda/envs/python/lib/python3.6/site-packages (from
requests->geocoder) (2.10)
Requirement already satisfied: certifi>=2017.4.17 in
/home/jupyterlab/conda/envs/python/lib/python3.6/site-packages (from
requests->geocoder) (2020.6.20)
Requirement already satisfied: urllib3!=1.25.0,!1.25.1,<1.26,>=1.21.1 in
/home/jupyterlab/conda/envs/python/lib/python3.6/site-packages (from
requests->geocoder) (1.25.10)
Requirement already satisfied: chardet<4,>=3.0.2 in
/home/jupyterlab/conda/envs/python/lib/python3.6/site-packages (from
requests->geocoder) (3.0.4)
Building wheels for collected packages: future
  Building wheel for future (setup.py) ... done
  Stored in directory: /home/jupyterlab/.cache/pip/wheels/8b/99/a0/81daf51
dcd359a9377b110a8a886b3895921802d2fc1b2397e
Successfully built future

```

```
Installing collected packages: ratelim, future, click, geocoder
Successfully installed click-7.1.2 future-0.18.2 geocoder-1.38.1 ratelim-0.1.6
```

```
[80]: conda install -c anaconda xlrd --yes
```

```
Collecting package metadata (current_repodata.json): done
Solving environment: done
```

```
==> WARNING: A newer version of conda exists. <==
current version: 4.8.3
latest version: 4.8.4
```

```
Please update conda by running
```

```
$ conda update -n base -c defaults conda
```

```
## Package Plan ##
```

```
environment location: /home/jupyterlab/conda/envs/python
```

```
added / updated specs:
- xlrd
```

```
The following packages will be downloaded:
```

package	build		
ca-certificates-2020.6.24	0	133 KB	anaconda
certifi-2020.6.20	py36_0	160 KB	anaconda
openssl-1.1.1g	h7b6447c_0	3.8 MB	anaconda
xlrd-1.2.0	py36_0	188 KB	anaconda
Total:		4.3 MB	

```
The following packages will be UPDATED:
```

```
ca-certificates conda-forge::ca-certificates-2020.6.2~ --> anaconda::ca-  
certificates-2020.6.24-0
```

```
The following packages will be SUPERSEDED by a higher-priority channel:
```

```
certifi conda-forge::certifi-2020.6.20-py36h9~ -->  
anaconda::certifi-2020.6.20-py36_0  
openssl conda-forge::openssl-1.1.1g-h516909a_0 -->
```

```

anaconda::openssl-1.1.1g-h7b6447c_0
  xlrld                      conda-forge/noarch::xlrld-1.2.0-py_0 -->
anaconda/linux-64::xlrld-1.2.0-py36_0

```

Downloading and Extracting Packages

```

ca-certificates-2020 | 133 KB | ##### | 100%
xlrld-1.2.0          | 188 KB | ##### | 100%
certifi-2020.6.20    | 160 KB | ##### | 100%
openssl-1.1.1g       | 3.8 MB | ##### | 100%
Preparing transaction: done
Verifying transaction: done
Executing transaction: done

```

Note: you may need to restart the kernel to use updated packages.

```
[81]: conda install bs4
```

```

Collecting package metadata (current_repodata.json): done
Solving environment: done

```

```

==> WARNING: A newer version of conda exists. <==
  current version: 4.8.3
  latest version: 4.8.4

```

Please update conda by running

```
$ conda update -n base -c defaults conda
```

Package Plan

```
environment location: /home/jupyterlab/conda/envs/python
```

```

added / updated specs:
- bs4

```

The following packages will be downloaded:

package	build	
beautifulsoup4-4.9.1	py36_0	172 KB
bs4-4.9.1	0	4 KB
ca-certificates-2020.6.24	0	125 KB

certifi-2020.6.20		py36_0	156 KB
openssl-1.1.1g		h7b6447c_0	2.5 MB
soupsieve-2.0.1		py_0	33 KB

Total:			3.0 MB

The following NEW packages will be INSTALLED:

beautifulsoup4	pkgs/main/linux-64::beautifulsoup4-4.9.1-py36_0
bs4	pkgs/main/noarch::bs4-4.9.1-0
soupsieve	pkgs/main/noarch::soupsieve-2.0.1-py_0

The following packages will be SUPERSEDED by a higher-priority channel:

ca-certificates	anaconda --> pkgs/main
certifi	anaconda --> pkgs/main
openssl	anaconda --> pkgs/main

Downloading and Extracting Packages

beautifulsoup4-4.9.1	172 KB	#####	100%
openssl-1.1.1g	2.5 MB	#####	100%
ca-certificates-2020	125 KB	#####	100%
certifi-2020.6.20	156 KB	#####	100%
bs4-4.9.1	4 KB	#####	100%
soupsieve-2.0.1	33 KB	#####	100%

Preparing transaction: done
Verifying transaction: done
Executing transaction: done

Note: you may need to restart the kernel to use updated packages.

2. LOAD SOURCE WEBPAGE & ASSIGN VARIABLE SOURCE

```
[82]: import requests
import lxml
! pip install bs4
from bs4 import BeautifulSoup
data = requests.get("http://datausa.io/api/data?measure=Property Value by_
↳Bucket,Property Value by_
↳Bucket%20Moe&geo=16000US3420020,01000US&drilldowns=Value%20Bucket").text
soup = BeautifulSoup(data, 'html.parser')
```

Requirement already satisfied: bs4 in
/home/jupyterlab/conda/envs/python/lib/python3.6/site-packages (0.0.1)
Requirement already satisfied: beautifulsoup4 in
/home/jupyterlab/conda/envs/python/lib/python3.6/site-packages (from bs4)

(4.9.1)

Requirement already satisfied: soupsieve>1.2 in
/home/jupyterlab/conda/envs/python/lib/python3.6/site-packages (from
beautifulsoup4->bs4) (2.0.1)

3. LOAD CSV WRITER AND ADD COLUMNS NAMES ON THE 1ST ROW

```
[91]: import csv
csv_file = open('Property Value.csv', 'w')
csv_writer = csv.writer(csv_file)
csv_writer.writerow(['Value Bucket', 'Value Property by Bucket', 'Share'])
```

[91]: 45

4. SCRAPE THE WEBPAGE

```
[8]: ! pip install lxml
import lxml
data = requests.get("http://datausa.io/api/data?measure=Property Value by_
↳Bucket,Property Value by_
↳Bucket%20Moe&geo=16000US3420020,01000US&drilldowns=Value%20Bucket")
soup = BeautifulSoup(data.content)
table = soup.find('table', class_ = 'wikitable sortable')

ValueBuckets = []
PropertyValuebyBuckets = []
Shares = []
Geographies = []

if Geographies == 'Not assigned':
    Geographies = Edgewater, NJ
```

Collecting lxml

Downloading https://files.pythonhosted.org/packages/79/37/d420b7fdc9a550bd29b8cfeacff3b38502d9600b09d7dfae9a69e623b891/lxml-4.5.2-cp36-cp36m-manylinux1_x86_64.whl (5.5MB)

| 5.5MB 23.3MB/s eta 0:00:01

Installing collected packages: lxml

Successfully installed lxml-4.5.2

4. PLACED THE SCRAPED DATA INTO A DATAFRAME

```
[17]: ValueBuckets = []
PropertyValuebyBuckets = []
Shares = []
Geographies = []

Edgewater = {'ValueBucket' : ValueBuckets , 'PropertyValuebyBucket' :
↳PropertyValuebyBuckets, 'Share': Shares}
Edgewater_DF = pd.DataFrame.from_dict(Edgewater, orient='index')
Edgewater_DF.transpose()
```

```
[17]: Empty DataFrame
Columns: [ValueBucket, PropertyValuebyBucket, Share]
Index: []
```

```
[16]: import pandas as pd
Edgewater_DF = pd.read_csv('Property Value.csv')
Edgewater_DF.head(10)
```

```
[16]: ID Value Bucket      Value Bucket ID Year  Year \
0      0  Less Than $10,000  2017  2017
1     25  $2,000,000 or More  2017  2017
2      2  $15,000 to $19,999  2017  2017
3      3  $20,000 to $24,999  2017  2017
4      4  $25,000 to $29,999  2017  2017
5      5  $30,000 to $34,999  2017  2017
6      6  $35,000 to $39,999  2017  2017
7      7  $40,000 to $49,999  2017  2017
8      8  $50,000 to $59,999  2017  2017
9      9  $60,000 to $69,999  2017  2017
```

```
Property Value by Bucket  Property Value by Bucket Moe      Geography \
0      1476974      13039.0  United States
1           42        56.0  Edgewater, NJ
2      562022      5452.0  United States
3      603519      5001.0  United States
4      552538      4783.0  United States
5      674950      6727.0  United States
6      506636      4012.0  United States
7     1314549      8040.0  United States
8     1633945     10334.0  United States
9     1907605     11773.0  United States
```

```
ID Geography Slug Geography      share
0      01000US  united-states  0.019477
1  16000US3420020  edgewater-nj  0.022838
2      01000US  united-states  0.007411
```


3	01000US	united-states	0.007959
4	01000US	united-states	0.007286
5	01000US	united-states	0.008900
6	01000US	united-states	0.006681
7	01000US	united-states	0.017335
8	01000US	united-states	0.021547
9	01000US	united-states	0.025155

4A. TRI STATE AREA COMPARABLE PROPERTIES

```
[2]: import pandas as pd
TriState_DF = pd.read_csv('Tri-State Property Value.csv')
TriState_DF.head(15)
```

[2]:	ID	Value Bucket	Value Bucket	ID	Year	Year \
	0	16	\$175,000 to \$199,999	2018	2018	
	1	25	\$2,000,000 or More	2018	2018	
	2	24	\$1,500,000 to \$1,999,999	2018	2018	
	3	1	\$10,000 to \$14,999	2018	2018	
	4	2	\$15,000 to \$19,999	2018	2018	
	5	23	\$1,000,000 to \$1,499,999	2018	2018	
	6	3	\$20,000 to \$24,999	2018	2018	
	7	4	\$25,000 to \$29,999	2018	2018	
	8	22	\$750,000 to \$999,999	2018	2018	
	9	5	\$30,000 to \$34,999	2018	2018	
	10	6	\$35,000 to \$39,999	2018	2018	
	11	21	\$500,000 to \$749,999	2018	2018	
	12	7	\$40,000 to \$49,999	2018	2018	
	13	8	\$50,000 to \$59,999	2018	2018	
	14	20	\$400,000 to \$499,999	2018	2018	

	Property Value by Bucket	Property Value by Bucket Moe \
0	66214	3278.0
1	114348	4809.0
2	81294	4190.0
3	561871	12432.0
4	480004	10818.0
5	197870	6265.0
6	519620	11799.0
7	465467	11618.0
8	379279	8815.0
9	581079	12322.0
10	413843	9243.0
11	854065	12661.0
12	1097775	16811.0
13	1363891	16708.0
14	600830	9791.0

	Geography	ID Geography \
0	New York-Newark-Jersey City, NY-NJ-PA	31000US35620
1	New York-Newark-Jersey City, NY-NJ-PA	31000US35620
2	New York-Newark-Jersey City, NY-NJ-PA	31000US35620
3	United States	01000US
4	United States	01000US
5	New York-Newark-Jersey City, NY-NJ-PA	31000US35620
6	United States	01000US
7	United States	01000US
8	New York-Newark-Jersey City, NY-NJ-PA	31000US35620
9	United States	01000US
10	United States	01000US
11	New York-Newark-Jersey City, NY-NJ-PA	31000US35620
12	United States	01000US
13	United States	01000US
14	New York-Newark-Jersey City, NY-NJ-PA	31000US35620

	Slug Geography	share
0	new-york-northern-new-jersey-long-island-ny-nj...	0.017566
1	new-york-northern-new-jersey-long-island-ny-nj...	0.030335
2	new-york-northern-new-jersey-long-island-ny-nj...	0.021566
3	united-states	0.007231
4	united-states	0.006177
5	new-york-northern-new-jersey-long-island-ny-nj...	0.052492
6	united-states	0.006687
7	united-states	0.005990
8	new-york-northern-new-jersey-long-island-ny-nj...	0.100618
9	united-states	0.007478
10	united-states	0.005326
11	new-york-northern-new-jersey-long-island-ny-nj...	0.226572
12	united-states	0.014127
13	united-states	0.017551
14	new-york-northern-new-jersey-long-island-ny-nj...	0.159392

5. GET EDGEWATER GEOGRAPHICAL COORDINATES

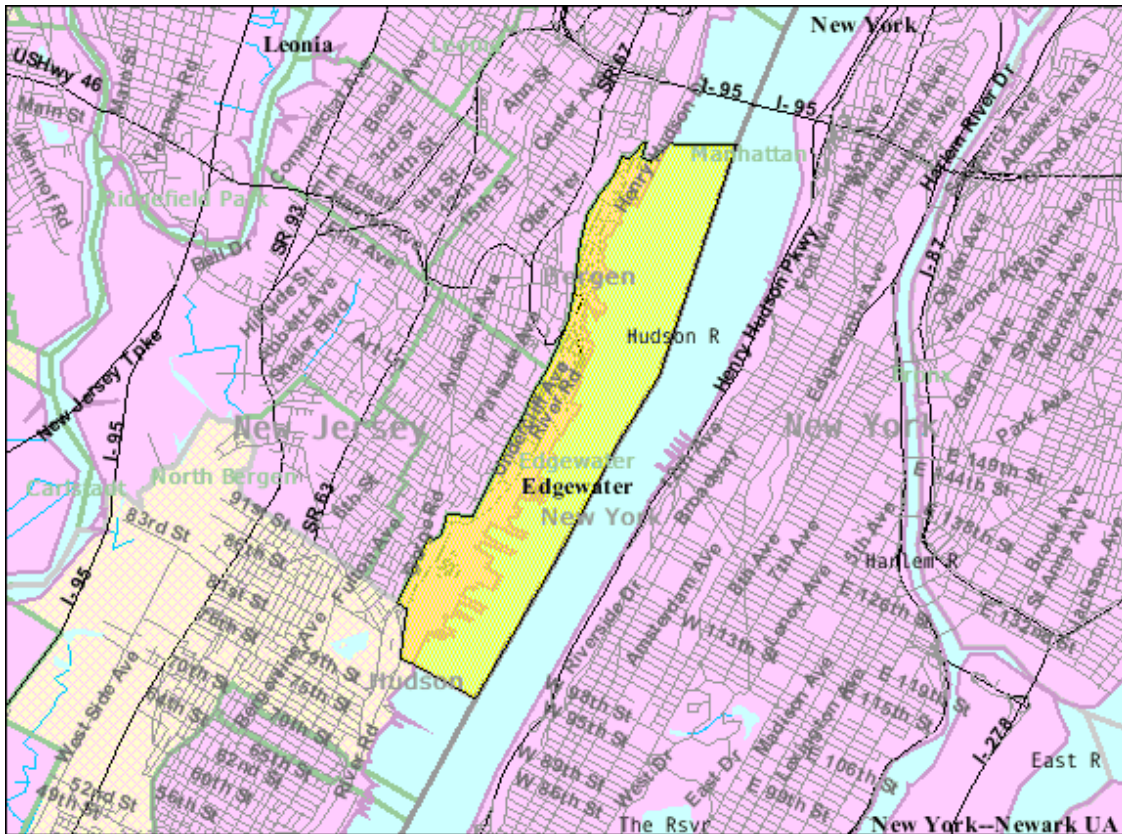
```
[3]: import pandas as pd
Coordinates = pd.read_csv("coordinates.json")
Coordinates.head()
```

```
[3]: Empty DataFrame
Columns: [{"DD":{"lat":40.82989, lng:-73.97391}, DMS:{"lat":"40°49'47.61\" N",
lng:"73°58'26.08\" W"}}, geohash:"dr72k8vg6g", UTM:"18T 586517.81862157
4520380.39565467"}]
Index: []
```

6. IMPORT A MAP OF EDGEWATER, NJ

```
[20]: from PIL import Image
img = Image.open('edgewatermap.png')
img
```

[20]:



CONCLUSION

Edgewater is a great investment opportunity due to its close proximity to NYC, bridges, tunnels, and continuous increase in property value. Considered as a suburb of NYC, it still also offers an urban feel with its array of bars, restaurants, coffee shops, shopping centers, beautiful NYC skyline view, and beautiful parks.

Thanks for your time!