



In this week, you will continue working on your capstone project. Please remember by the end of this week, you will need to submit the following:

**A full report consisting of all of the following components (15 marks):**

- Introduction where you discuss the business problem and who would be interested in this project.
- Data where you describe the data that will be used to solve the problem and the source of the data.
- Methodology section which represents the main component of the report where you discuss and describe any exploratory data analysis that you did, any inferential statistical testing that you performed, if any, and what machine learnings were used and why.
- Results section where you discuss the results.
- Discussion section where you discuss any observations you noted and any recommendations you can make based on the results.
- Conclusion section where you conclude the report.

**A link to your Notebook on your Github repository pushed showing your code. (15 marks)**

**Your choice of a presentation or blogpost. (10 marks)**

## **2. INTRODUCTION / BUSINESS PROBLEM:**

Everybody wants to live in New York City, however many cannot afford to. A two-bedroom apartment in a sketchy neighborhood can easily cost more than the price of an actual house outside of New York City.

So what is the solution for up-and-coming trendsetters who would like to live and work in New York City, but cannot afford to? One solution is to find a house or condo in nearby New Jersey towns and commute into the city via one of many mass-transportation services. They can build equity in their home and when they are ready they can sell and move across the Hudson River. If they still want to, that is. Many New Jerseyans have decided to stay in New Jersey after having been able to escape the craziness of New York after work. New Jersey has many of the same benefits of New York City, but without the out-of-reach housing costs and many of the urban problems (like parking/having a car).

There are many towns within New Jersey that are close to New York City, and have a thriving community and nightlife, but have much more affordable apartments for rent.

Union County is very close to New York City and has towns that can fit any budget and lifestyle. For all the city comforts of New York with in walking distance, Elizabeth is the town. For a balance of city and country, the sleepy commuter towns of Westfield, Scotch Plains, and Rahway might be choices.

## The Towns of Union County:

1. Berkeley Heights
2. Clark
3. Cranford
4. Elizabeth
5. Fanwood
6. Garwood
7. Hillside
8. Kenilworth
9. Linden
10. Mountainside
11. New Providence
12. Plainfield
13. Rahway
14. Roselle
15. Roselle Park
16. Scotch Plains
17. Springfield
18. Summit
19. Union (township)
20. Westfield
21. Winfield Township

This project and attached reports will help the potential homeowner find an area in Union County where they could find accommodations close to New York within their budget which have their own share of parks, trendy local restaurants, schools and other places of interest.

Once this report becomes used by potential homeowners for Union County, it can easily be modified to include other areas of New Jersey and other states for a larger audience.

## 3. DATA SOURCES

For this project, I will use these publicly available data sources:

### US Zip Code Latitude and Longitude:

*This database was composed using ZIP code gazetteers from the US Census Bureau from 1999 and 2000, augmented with additional ZIP code information. The database is believed to contain over 98% of the ZIP Codes in current use in the United States. Last modified: February 9, 2018*

- **Original Data Source:** <https://public.opendatasoft.com/explore/dataset/us-zip-code-latitude-and-longitude/export/>
- **Stored at:** [https://github.com/JamesH-Lantern/Coursera\\_Capstone/blob/master/us-zip-code-latitude-and-longitude.csv?raw=true](https://github.com/JamesH-Lantern/Coursera_Capstone/blob/master/us-zip-code-latitude-and-longitude.csv?raw=true)

## Data Methodology:

*The source data contains all the latitude and longitude information for all zipcodes in the United States as of February, 2018. I will need to extract just the locations from the State of New Jersey. Special note that this data has semicolon separated values instead of commas*

**Fields used:** Zipcode, City, State, Latitude, Longitude

## Union County, NJ Average Home Values

*Zillow Home Value Index (ZHVI): A smoothed, seasonally adjusted measure of the typical home value and market changes across a given region and housing type. Zillow publishes top-tier ZHVI and bottom-tier ZHVI. Zillow also publishes ZHVI for all single-family residences, condo/coops, all homes with 1, 2, 3, 4 and 5+ bedrooms, and the ZHVI per square foot.*

- **Original Source:** <https://www.zillow.com/research/data/>
- **Stored at:** [https://github.com/JamesH-Lantern/Coursera\\_Capstone/raw/master/NJ\\_MedianPrice\\_20200331.csv](https://github.com/JamesH-Lantern/Coursera_Capstone/raw/master/NJ_MedianPrice_20200331.csv)

**Data Methodology:** *The source data contains average home values for Union County zipcodes as of April, 2020. As there are multiple Union Counties in the United States, I will need to extract just the locations from the State of New Jersey. Special note that this data has semicolon separated values instead of commas*

**Fields used:** Zipcode, City, State, CountyName, AvgHomeValue

## Top Venue Recommendations from FourSquare API

I will be using the FourSquare API to pull the “top places of interest” in selected cities of Union, New Jersey, as well as various information

Website: <http://www.FourSquare.com>

**Data Methodology:** Through a series of API queries, I will be pulling off various reports of popular venues, schools, and restaurants.

## Importing Python Libraries

Importing various Python libraries for scraping and cleaning data, visualizations, etc.

```
***Just a sample. May change in final product***
#!pip install -c conda-forge geopy --yes
#!pip install folium
import requests # library to handle requests
import pandas as pd # library for data analysis
import numpy as np # library to handle data in a vectorized manner
import random # library for random number generation
```

```

import folium
import matplotlib.cm as cm
import matplotlib.colors as colors
from bs4 import BeautifulSoup
from geopy.geocoders import Nominatim # module to convert an address into
latitude and longitude values
#transforming json file into a pandas dataframe library
from pandas.io.json import json_normalize
from sklearn.cluster import KMeans
#libraries for displaying images
from IPython.display import Image
from IPython.core.display import HTML

```

## 4. How will we solve the Business Problem?

Using the data above, we will be doing a exploratory dive into the individual towns of Union County using the Four SquareAPI, and finding out the most popular spots in the vicinity. Which I will then group the towns into clusters.

We will be looking into spots such as:

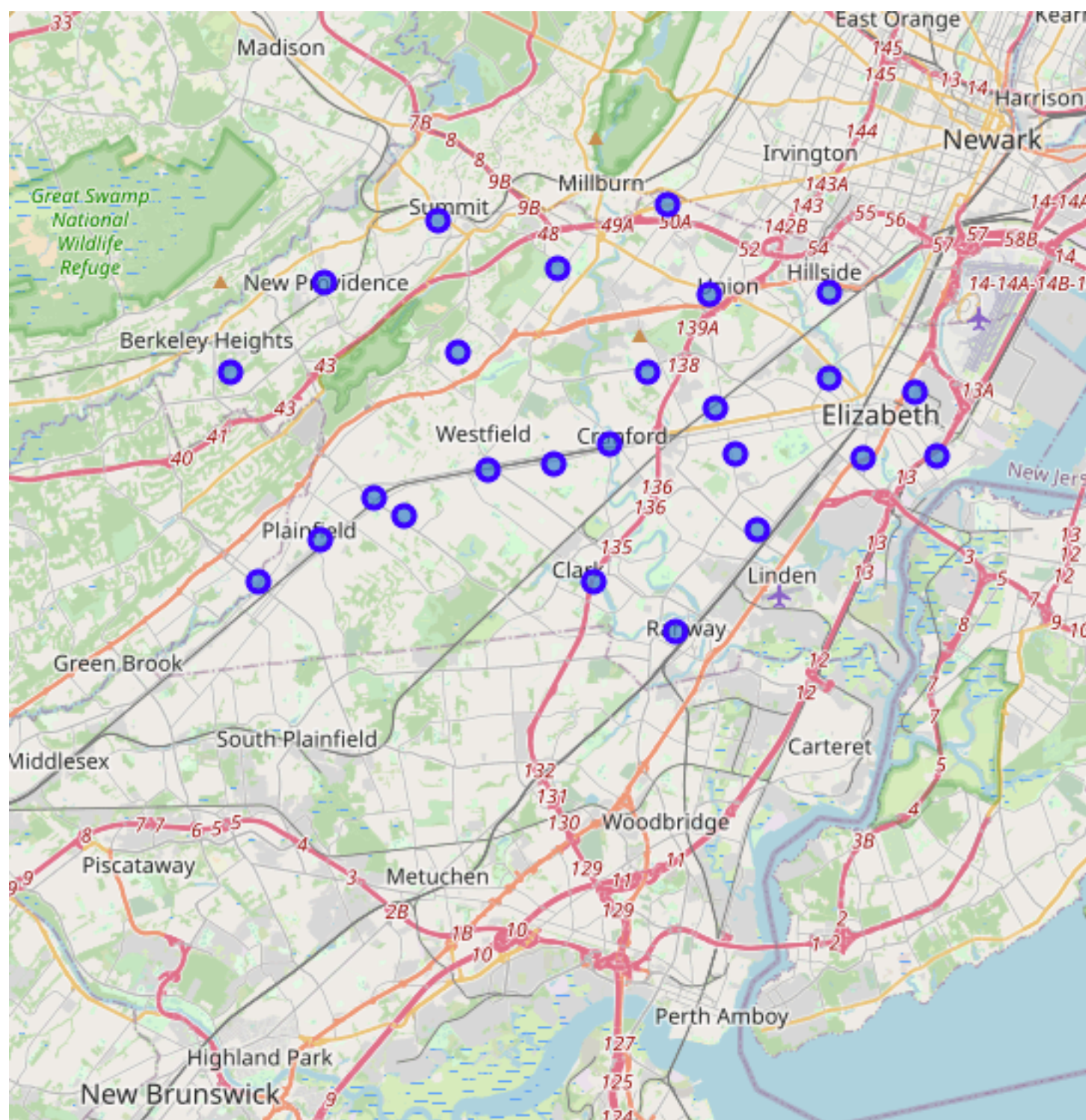
- Popular Food venues
- Outdoors and Recreation
- Nightlife
- Nearby Schools, etc.

We will also be including the average home values in the Union County towns to show the relative cost expectations between the towns, so that potential homeowners can determine whether the towns with the best variation of services is within their budget.

	zipcode	City	State	CountyName	AvgHomeValue	Latitude	Longitude
0	07083	Union	NJ	Union County	362323	40.695401	-74.26933
1	07036	Linden	NJ	Union County	311037	40.633953	-74.25339
2	07060	Plainfield	NJ	Union County	299761	40.620102	-74.42481
3	07202	Elizabeth	NJ	Union County	281318	40.652302	-74.21692
4	07065	Rahway	NJ	Union County	317059	40.607103	-74.28099
5	07208	Elizabeth	NJ	Union County	341434	40.673102	-74.22834
6	07090	Westfield	NJ	Union County	742817	40.649202	-74.34567
7	07076	Scotch Plains	NJ	Union County	513355	40.637152	-74.37493
8	07016	Cranford	NJ	Union County	517142	40.656302	-74.30371
9	07901	Summit	NJ	Union County	858275	40.714501	-74.36330

	<b>zipcode</b>	<b>City</b>	<b>State</b>	<b>CountyName</b>	<b>AvgHomeValue</b>	<b>Latitude</b>	<b>Longitude</b>
<b>10</b>	07201	Elizabeth	NJ	Union County	271171	40.669502	-74.19889
<b>11</b>	07203	Roselle	NJ	Union County	284227	40.653502	-74.26058
<b>12</b>	07206	Elizabeth	NJ	Union County	253298	40.653189	-74.19158
<b>13</b>	07205	Hillside	NJ	Union County	303459	40.695552	-74.22856
<b>14</b>	07081	Springfield	NJ	Union County	455190	40.701801	-74.32216
<b>15</b>	07066	Clark	NJ	Union County	459607	40.620336	-74.30934
<b>16</b>	07204	Roselle Park	NJ	Union County	326917	40.665552	-74.26768
<b>17</b>	07062	Plainfield	NJ	Union County	296946	40.631252	-74.40383
<b>18</b>	07974	New Providence	NJ	Union County	625257	40.698268	-74.40225
<b>19</b>	07922	Berkeley Heights	NJ	Union County	581746	40.675112	-74.43459
<b>20</b>	07033	Kenilworth	NJ	Union County	446286	40.674652	-74.29080
<b>21</b>	07023	Fanwood	NJ	Union County	477659	40.641852	-74.38533
<b>22</b>	07092	Mountainside	NJ	Union County	598895	40.680301	-74.35601
<b>23</b>	07027	Garwood	NJ	Union County	402907	40.650802	-74.32351
<b>24</b>	07088	Union	NJ	Union County	283429	40.718401	-74.28416





	City	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Berkeley Heights	Trail	Thai Restaurant	Gym	Italian Restaurant	Pool	Wine Shop	Coffee Shop	Fast Food Restaurant	Farmer's Market	Donut Shop
1	Clark	Thai Restaurant	BBQ Joint	Bakery	Grocery Store	Shopping Plaza	Park	Wine Shop	Dessert Shop	Food Truck	Fast Food Restaurant
2	Cranford	Gourmet Shop	Sushi Restaurant	Bar	Seafood Restaurant	Ice Cream Shop	Garden Center	Park	Coffee Shop	Dessert Shop	Food Truck
3	Garwood	Health & Beauty Service	Café	Wine Shop	Dessert Shop	Garden Center	Food Truck	Fast Food Restaurant	Farmer's Market	Donut Shop	Diner
4	Mountain Side	Bakery	Japanese Restaurant	Wine Shop	Dessert Shop	Garden Center	Food Truck	Fast Food Restaurant	Farmer's Market	Donut Shop	Diner
5	New Providence	Pool	Farmer's Market	Brewery	Diner	Dessert Shop	Coffee Shop	Garden Center	Food Truck	Fast Food Restaurant	Donut Shop
6	Plainfield	Pizza Place	Ice Cream Shop	Mexican Restaurant	Park	Golf Course	Food Truck	Fast Food Restaurant	Farmer's Market	Donut Shop	Diner
7	Rahway	Park	Wine Shop	Gourmet Shop	Garden Center	Food Truck	Fast Food Restaurant	Farmer's Market	Donut Shop	Diner	Dessert Shop
8	Scotch Plains	Sushi Restaurant	Gym	Ice Cream Shop	Breakfast Spot	Golf Course	Wine Shop	Dessert Shop	Food Truck	Fast Food Restaurant	Farmer's Market



	City	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
9	Springfield	Wine Shop	Burger Joint	Italian Restaurant	Donut Shop	Pizza Place	Golf Course	Breakfast Spot	Brewery	Bar	Café
10	Summit	Gym	Greek Restaurant	Pizza Place	Arts & Crafts Store	Bakery	Café	Coffee Shop	Farmer's Market	Ice Cream Shop	Park
11	Union	Wine Shop	Food Truck	Supermarket	Fast Food Restaurant	Coffee Shop	Garden Center	Farmer's Market	Donut Shop	Diner	Dessert Shop
12	Westfield	Park	Pizza Place	Coffee Shop	Grocery Store	Bakery	Burger Joint	Sandwich Place	Food Truck	Fast Food Restaurant	Farmer's Market

	City	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Berkeley Heights	Trail	Thai Restaurant	Gym	Italian Restaurant	Pool
1	Clark	Thai Restaurant	BBQ Joint	Bakery	Grocery Store	Shopping Plaza
2	Cranford	Gourmet Shop	Sushi Restaurant	Bar	Seafood Restaurant	Ice Cream Shop
3	Garwood	Health & Beauty Service	Café	Wine Shop	Dessert Shop	Garden Center
4	Mountainside	Bakery	Japanese Restaurant	Wine Shop	Dessert Shop	Garden Center
5	New Providence	Pool	Farmers Market	Brewery	Diner	Dessert Shop
6	Plainfield	Pizza Place	Ice Cream Shop	Mexican Restaurant	Park	Golf Course
7	Rahway	Park	Wine Shop	Gourmet Shop	Garden Center	Food Truck

	City	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
8	Scotch Plains	Sushi Restaurant	Gym	Ice Cream Shop	Breakfast Spot	Golf Course
9	Springfield	Wine Shop	Burger Joint	Italian Restaurant	Donut Shop	Pizza Place
10	Summit	Gym	Greek Restaurant	Pizza Place	Arts & Crafts Store	Bakery
11	Union	Wine Shop	Food Truck	Supermarket	Fast Food Restaurant	Coffee Shop
12	Westfield	Park	Pizza Place	Coffee Shop	Grocery Store	Bakery

	City	AvgHomeValue
0	Berkeley Heights	581746.0
1	Clark	459607.0
2	Cranford	517142.0
3	Garwood	402907.0
4	Mountainside	598895.0
5	New Providence	625257.0
6	Plainfield	298353.5
7	Rahway	317059.0
8	Scotch Plains	513355.0
9	Springfield	455190.0
10	Summit	858275.0
11	Union	322876.0
12	Westfield	742817.0

	City	AvgHomeValue	ClusterLabels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	Latitude	Longitude
0	Berkeley Heights	581746.0	3	Trail	Thai Restaurant	Gym	Italian Restaurant	Pool	Wine Shop	Coffee Shop	Fast Food Restaurant	Farmers Market	Donut Shop	40.675112	-74.43459
1	Berkeley Heights	581746.0	3	Trail	Thai Restaurant	Gym	Italian Restaurant	Pool	Wine Shop	Coffee Shop	Fast Food Restaurant	Farmers Market	Donut Shop	40.675112	-74.43459
2	Berkeley Heights	581746.0	3	Trail	Thai Restaurant	Gym	Italian Restaurant	Pool	Wine Shop	Coffee Shop	Fast Food Restaurant	Farmers Market	Donut Shop	40.675112	-74.43459
3	Berkeley Heights	581746.0	3	Trail	Thai Restaurant	Gym	Italian Restaurant	Pool	Wine Shop	Coffee Shop	Fast Food Restaurant	Farmers Market	Donut Shop	40.675112	-74.43459
4	Berkeley Heights	581746.0	3	Trail	Thai Restaurant	Gym	Italian Restaurant	Pool	Wine Shop	Coffee Shop	Fast Food Restaurant	Farmers Market	Donut Shop	40.675112	-74.43459
5	Berkeley Heights	581746.0	3	Trail	Thai Restaurant	Gym	Italian Restaurant	Pool	Wine Shop	Coffee Shop	Fast Food Restaurant	Farmers Market	Donut Shop	40.675112	-74.43459
6	Clark	459607.0	1	Thai Restaurant	BBQ Joint	Bakery	Grocery Store	Shopping Plaza	Park	Wine Shop	Dessert Shop	Food Truck	Fast Food Restaurant	40.620336	-74.30934

	City	AvgHomeValue	ClusterLabels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	Latitude	Longitude
7	Clark	459607.0	1	Thai Restaurant	BBQ Joint	Bakery	Grocery Store	Shopping Plaza	Park	Wine Shop	Dessert Shop	Food Truck	Fast Food Restaurant	40.620336	-74.30934
8	Clark	459607.0	1	Thai Restaurant	BBQ Joint	Bakery	Grocery Store	Shopping Plaza	Park	Wine Shop	Dessert Shop	Food Truck	Fast Food Restaurant	40.620336	-74.30934
9	Clark	459607.0	1	Thai Restaurant	BBQ Joint	Bakery	Grocery Store	Shopping Plaza	Park	Wine Shop	Dessert Shop	Food Truck	Fast Food Restaurant	40.620336	-74.30934
10	Clark	459607.0	1	Thai Restaurant	BBQ Joint	Bakery	Grocery Store	Shopping Plaza	Park	Wine Shop	Dessert Shop	Food Truck	Fast Food Restaurant	40.620336	-74.30934
11	Clark	459607.0	1	Thai Restaurant	BBQ Joint	Bakery	Grocery Store	Shopping Plaza	Park	Wine Shop	Dessert Shop	Food Truck	Fast Food Restaurant	40.620336	-74.30934
12	Cranford	517142.0	1	Gourmet Shop	Sushi Restaurant	Bar	Seafood Restaurant	Ice Cream Shop	Garden Center	Park	Coffee Shop	Dessert Shop	Food Truck	40.656302	-74.30371
13	Cranford	517142.0	1	Gourmet Shop	Sushi Restaurant	Bar	Seafood Restaurant	Ice Cream Shop	Garden Center	Park	Coffee Shop	Dessert Shop	Food Truck	40.656302	-74.30371

	City	AvgHomeValue	ClusterLabels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	Latitude	Longitude
14	Cranford	517142.0	1	Gourmet Shop	Sushi Restaurant	Bar	Seafood Restaurant	Ice Cream Shop	Garden Center	Park	Coffee Shop	Dessert Shop	Food Truck	40.656302	-74.30371
15	Cranford	517142.0	1	Gourmet Shop	Sushi Restaurant	Bar	Seafood Restaurant	Ice Cream Shop	Garden Center	Park	Coffee Shop	Dessert Shop	Food Truck	40.656302	-74.30371
16	Cranford	517142.0	1	Gourmet Shop	Sushi Restaurant	Bar	Seafood Restaurant	Ice Cream Shop	Garden Center	Park	Coffee Shop	Dessert Shop	Food Truck	40.656302	-74.30371
17	Cranford	517142.0	1	Gourmet Shop	Sushi Restaurant	Bar	Seafood Restaurant	Ice Cream Shop	Garden Center	Park	Coffee Shop	Dessert Shop	Food Truck	40.656302	-74.30371
18	Cranford	517142.0	1	Gourmet Shop	Sushi Restaurant	Bar	Seafood Restaurant	Ice Cream Shop	Garden Center	Park	Coffee Shop	Dessert Shop	Food Truck	40.656302	-74.30371
19	Cranford	517142.0	1	Gourmet Shop	Sushi Restaurant	Bar	Seafood Restaurant	Ice Cream Shop	Garden Center	Park	Coffee Shop	Dessert Shop	Food Truck	40.656302	-74.30371

## 7. Discussion and Conclusion

On this notebook, we performed an analysis of the best town venue recommendations based on the Food venue categories. Recommendations based on other user searches like available



outdoor and recreation areas are also available. Union County, NJ naturally does not have as much coverage as say New York, but there are interesting venues scattered throughout the county. The information extracted in this notebook presents the local towns and the average home values and will be a good supplement to web based recommendations for potential homeowners.

Using Foursquare API, we have collected a good amount of venue recommendations in Union County, however sourcing from a FourSquare free account does have some limitations. The list of venues is not as exhaustive as I would have liked and not all the venues found in the area have stored ratings. For this reason, the number of analyzed venues are only about 50% of all the available venues initially collected. The results therefore may significantly change, when more information are collected on those with missing data.

The generated clusters from our results shows that there are very good and interesting places located in areas where the average house values are cheaper. This may be very interesting for potential homeowners who have budget constraints. For example Plainfields average housing is about 298k while Summit, NJ is 858k. However most of the best food and outdoor venues are located in the Cranford area which has an average house value of 517k

Comparatively the average rent in New York City is

*4,208 for 703 Square Feet and the Median list price per square foot is*

1,801 and median list price: \$1,980,000. The savings on finding somewhere in Union County, NJ is definitely worth considering the Garden State

Thank you.

James Hannon email: [jameshannon@lantern-media.com](mailto:jameshannon@lantern-media.com) [www.JamesHannon.com](http://www.JamesHannon.com)

