

Coursera Capstone Week 5
IBM Applied Data Science Capstone

***Recommendation for best location
to build new housing complex in
Hudson County, New Jersey***

Saheli Ghosh
July, 2020

Business Problem

- ▶ Hudson County is one of the fastest growing county in New Jersey according to census data.
- ▶ Due to its proximity to New York City many people would like to buy home in Hudson County.
- ▶ What location in Hudson county is most suitable for construction of a new housing complex?

Description of data and its source

- ▶ List of municipalities in Hudson County from Wikipedia
https://en.wikipedia.org/wiki/Hudson_County,_New_Jersey
- ▶ Latitude and longitude coordinates of the municipalities obtained by Python Geocoder Library.
- ▶ Venue data for each of the municipalities from Foursquare API.

Methodology:

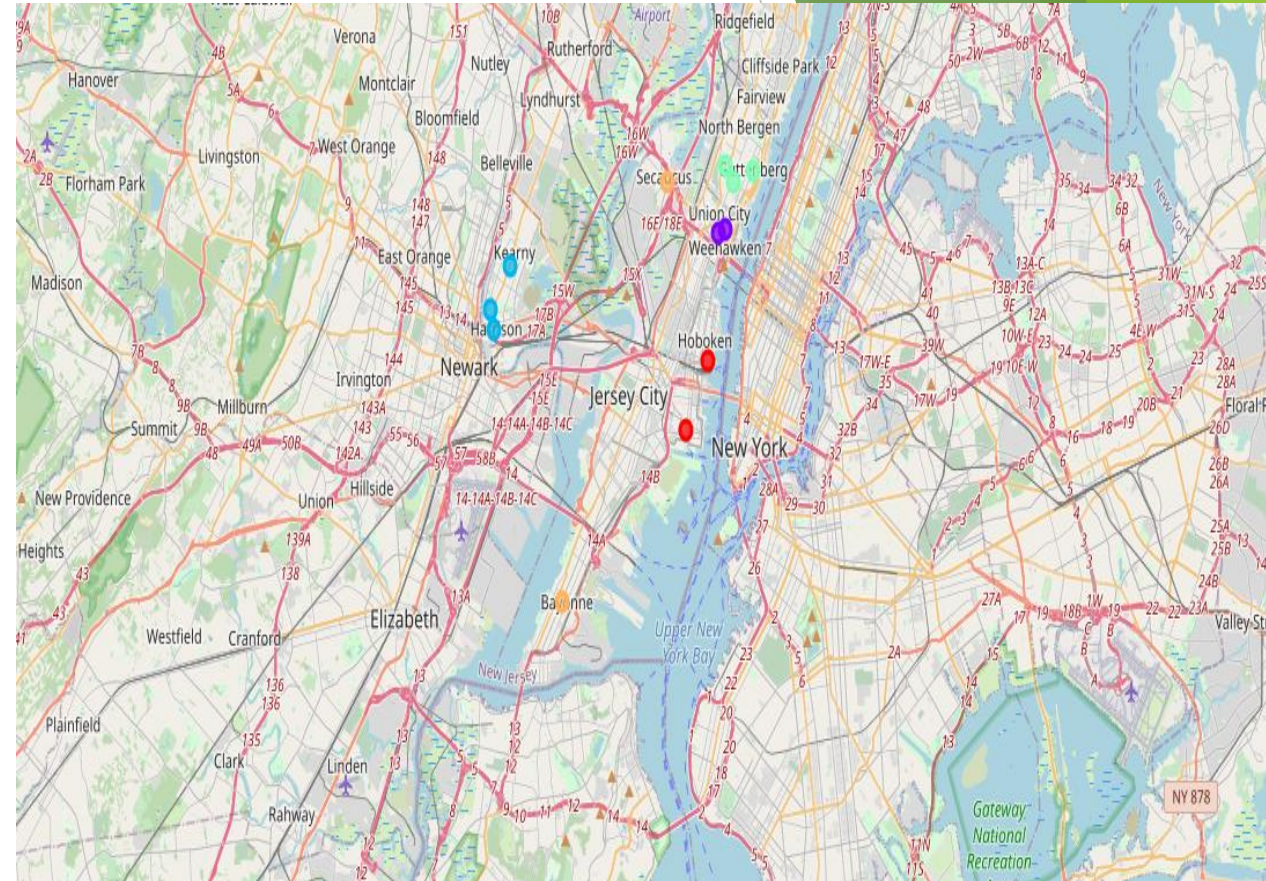
- ▶ Web scraping techniques using Python libraries requests and Beautiful Soup to extract the name of the municipalities from the Wikipedia page.
- ▶ Python geocoder library to get the geographical coordinates for each municipality of Hudson County.
- ▶ Foursquare API to get venue data

Methodology (cont.):

- ▶ Calculate the mean of the frequency of occurrence of each venue category.
- ▶ K-means clustering to group similar data municipalities based on the ranking of venue data.
- ▶ Use python package folium to visualize the resulting clusters

Results

- ▶ Cluster 0 - mostly parks, gyms and some specialty eateries.
- ▶ Cluster 1 - mostly theaters, concert halls, art galleries, parks and few restaurants.
- ▶ Cluster 2 - mostly restaurants.
- ▶ Cluster 3 - mostly parks, theaters, gym and some restaurants
- ▶ Cluster 4 - mostly restaurants, parks, grocery store and scenic lookout



Discussion and observation

- ▶ Based on the clustering result municipalities in cluster 4 (Bayonne and Secaucus) most suitable for a new housing complex.
- ▶ The most common venues are restaurant, park and grocery store which might be lucrative features for the prospective home buyers.

Conclusion:

- ▶ This project makes successful use of various technical, statistical and data science skills like web scraping, working with API interface, data wrangling, machine learning algorithm and visualization technique.
- ▶ The findings of this project will help relevant stakeholders to take informed decision regarding the best location for construction of a housing complex in Hudson County, New Jersey.