Coursera Capstone Project

Building or Buying a House in Seattle Washington

Introduction

- This is for anyone who is interested in buying or building a new house!

Seattle is one of the “Hottest” housing markets in America

Some houses sell as high as 22% above list price

Big companies

-Amazon

-Boing

-Microsoft

Business Problem

- Find a good location to build or buy a home
- In this project we are going to analyze where is the best location to buy or build a new home
- To do so we are going to use Seattle Neighborhoods and location data

Data

Data needed:

- -List of neighborhoods
- -Latitude and longitude coordinates of the Neighborhoods
- -Venues data in and around the Neighborhoods

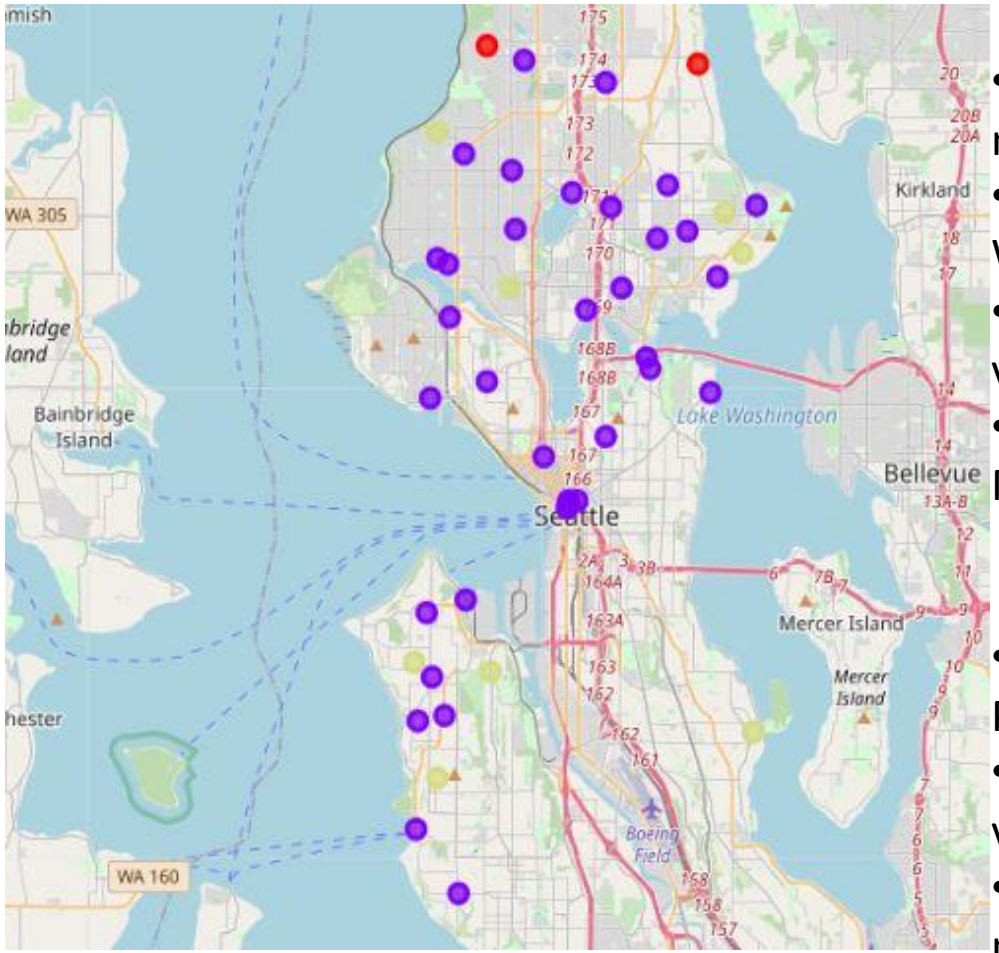
Data Sources:

- Wikipedia page: https://en.wikipedia.org/wiki/List_of_neighborhoods_in_Seattle , which has all the 127 neighborhoods in Seattle.
- Python Geocoders package for latitude and longitude
- Foursquare API for Venues Data

Methodology

- Python web scraping Wikipedia Page for Neighborhood Data
- Use Geocoder package to get Latitude and Longitude Coordinates
- Explore the neighborhoods by using Foursquare API to get the location-based venue data
- Get top 5 and or 10 venues each Neighborhood by taking the mean of the frequency of occurrence and ranking them
- Cluster by similarity using k-means clusters
- Visualize Neighborhoods in the clusters using Folium Package
- Based on deciding factors like “Parks” make a new data frame
- Finally cluster and visualize again using “Park” data and explore the clusters.

Result 1(Top 10)



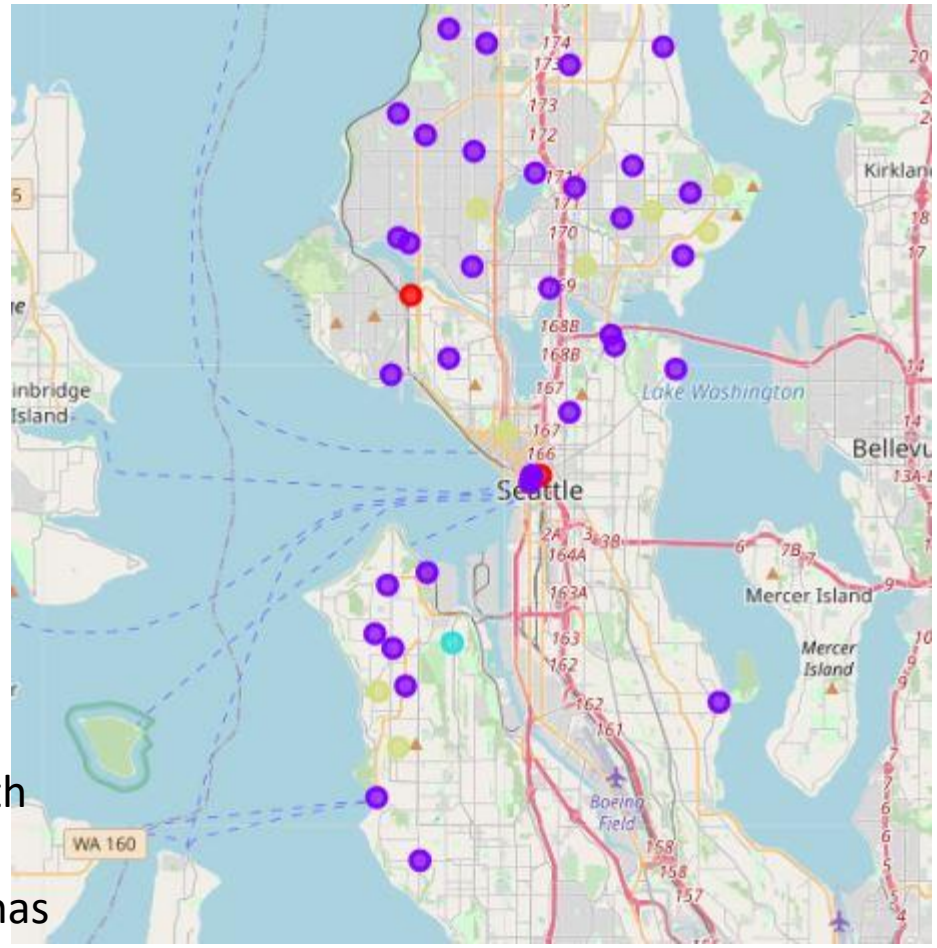
(Top 10)

- Red (Cluster 0, shows two neighborhoods),
- Purple (Cluster1,majority With less venue frequency)
- Bright-Green (Cluter2,no venues)
- Olive (Cluster 3, shows Most frequented venues)

(Parks)

- Cluster 0 and 2, no Parks total 3 Neighborhoods
- Cluster 1, still majority with very few parks
- Cluster 3 spread out and has most Parks!

Result 2 (Parks)



Discussion

Both top 10 venues and only parks were very similar in their cluster

- Cluster 1 had the most neighborhoods and cluster 2 had the least(only 1) neighborhood
- What we saw is that cluster 0 and 2 had no parks and cluster 1 had very few parks even though it had the most neighborhoods
- Cluster 3 was as well spread out as cluster 1 and yet had the most parks and most frequented venues!

Conclusion and Recommendation

- Choosing a neighborhood in cluster 3 is the best option to own a home because it’s well spread as cluster 1 and yet had the most parks!
- More work and data is needed to create comprehensive results to make a better decision as to when and where exactly to build or buy a new home.

Thank you!