The Battle of Neighborhoods

Finding A Better Place in West Vancouver

Introduction

- The purpose of this Project is to help people like my friends, Adam and Eve and their 5-yr old, in exploring better facilities in West Vancouver. It will help them making smart and efficient decision on selecting the most suitable neighborhood out of numbers of other neighborhoods in West Vancouver, BC.

- Lots of people are migrating to West Vancouver in Greater Vancouver region of Canada and need lots of research for good housing prices and reputated schools for their children. This project is for those people who are looking for better neighborhoods. For ease of accessing to Cafe, School, Super market, medical shops, grocery shops, mall, theatre, hospital, like minded people, etc.

- This Project aims to create an analysis of features for anybody migrating to West Vancouver to search a suitable neighborhood as a comparative analysis between neighborhoods. The features include median housing price and better school according to ratings, crime rates of that particular area, road connectivity, weather conditions, good management for emergency, water resources both freash and waste water and excrement conveyed in sewers and recreational facilities.

- It will help people to get awareness of the area and neighborhood before moving to West Vancouver for their work or to start a new fresh life.
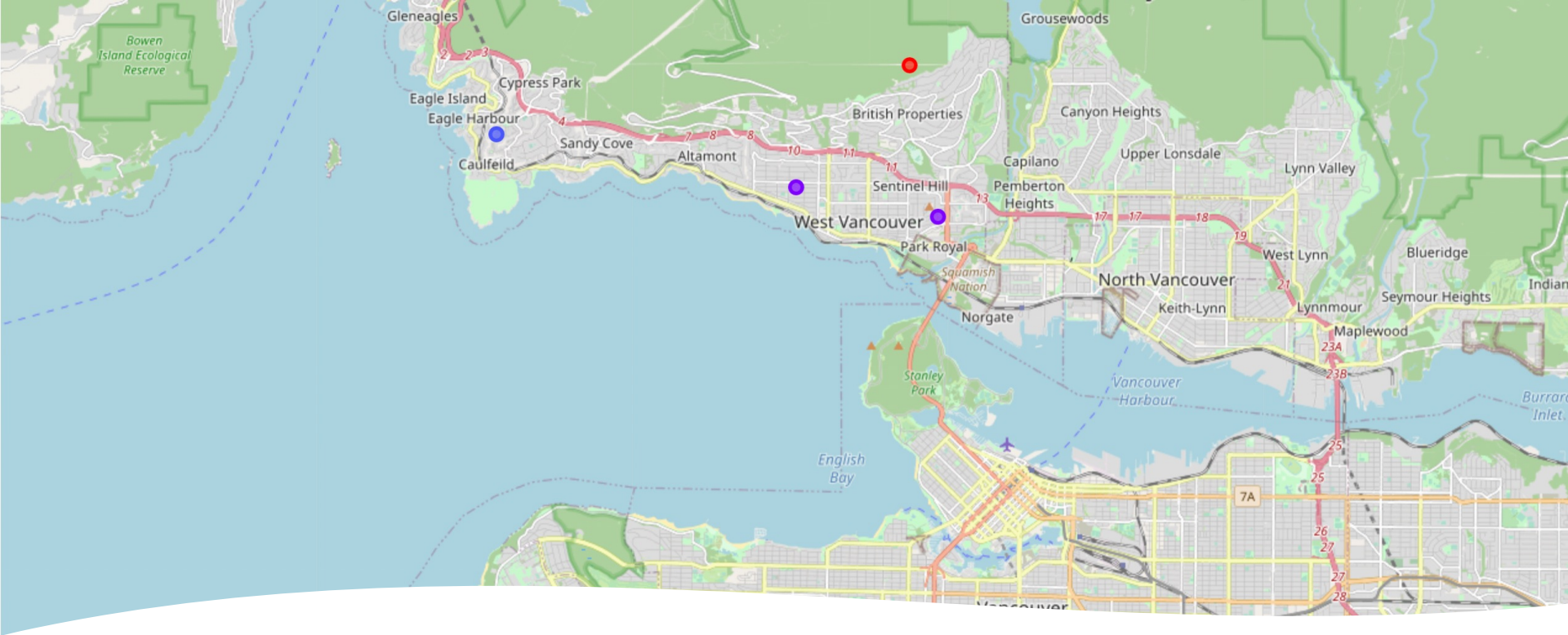
Data Extraction & Cleaning

I have used West Vancouver dataset which was scrapped from wikipedia. Dataset consisting of latitude and longitude, zip codes.

I've extracted the attributes 'Postalcode','Borough' and 'Neighborhood' Dataset. It was then then cleaned.

Foursquare API Data

- I need data about different venues in different neighborhoods of that specific borough. In order to gain that information I use "Foursquare" locational information. Foursquare is a location data provider with information about all manner of venues and events within an area of interest. Such information includes venue names, locations, menus and even photos. As such, the foursquare location platform will be used as the sole data source since all the stated required information can be obtained through the API.

- After finding the list of neighborhoods, I then connect to the Foursquare API to gather information about venues inside each and every neighborhood. For each neighborhood, I have chosen the radius to be 100 meter.

- The data retrieved from Foursquare contained information of venues within a specified distance of the longitude and latitude of the postcodes. The information obtained per venue as follows:

  - Neighborhood
  - Neighborhood Latitude
  - Neighborhood Longitude
  - Venue
  - Name of the venue e.g. the name of a store or restaurant
  - Venue Latitude
  - Venue Longitude
  - Venue Category

Map of West Vancouver

Geograpical Co-ordinate of Neighborhoods w.r.t address West Vancouver, BC were determined. I then fetched the nearby locations in tabular form followed by categories fetching. Some of the fetched categories include Clothing Store, Restaurant, Coffee Shop, Intersection, Sandwich Place, Gas Stations, department Store etc. This project fetched 16 categories.

Methodology

- I first encoded the categories using one hot encoding.
- To be able to explore neighborhoods, I need to cluster data which is a form of unsupervised machine learning: k-means clustering algorithm.
- I applied K-Means clustering method with value of k=3.

K-Means Clustering Approach

```
# Using K-Means to cluster neighborhood into 3 clusters
WestVancouver_grouped_clustering = WestVancouver_grouped.drop('Neighborhood', 1)
kmeans = KMeans(n_clusters=3, random_state=0).fit(WestVancouver_grouped_clustering)
kmeans.labels_
```

array([0, 1, 1, 2], dtype=int32)

```
neighborhoods_venues_sorted.insert(0, 'Cluster Labels', kmeans.labels_)

WestVancouver_merged =df_2.iloc[:16,:]

# merge WestVancouver_grouped with WestVancouver_data to add latitude/longitude for each neighborhood
WestVancouver_merged = WestVancouver_merged.join(neighborhoods_venues_sorted.set_index('Neighborhood'), on='Neighborhood')

WestVancouver_merged.head()# check the last columns!
```

	PostalCode	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue
0	V7S	West VancouverNorth	West VancouverNorth	49.35954	-123.14533	0	Trail	Wine Shop	Gastropub	Bakery	Bank	Boat or Ferry
1	V7T	West VancouverSoutheast	West VancouverSoutheast	49.33205	-123.13758	1	Bank	Department Store	Pharmacy	Mexican Restaurant	Liquor Store	Wine Shop
2	V7V	West VancouverSouth	West VancouverSouth	49.33737	-123.17716	1	Bakery	Coffee Shop	Wine Shop	Vietnamese Restaurant	Café	Gastropub
3	V7W	West VancouverWest	West VancouverWest	49.34713	-123.26034	2	Trail	Tennis Court	Playground	Boat or Ferry	Wine Shop	Department Store

Map of Clusters

```
kclusters = 10
```

```python
# create map
map_clusters = folium.Map(location=[latitude_x, longitude_y], zoom_start=11)

# set color scheme for the clusters
x = np.arange(kclusters)
colors_array = cm.rainbow(np.linspace(0, 1, kclusters))
rainbow = [colors.rgb2hex(i) for i in colors_array]
print(rainbow)
# add markers to the map

markers_colors = []
for lat, lon, nei , cluster in zip(WestVancouver_merged['Latitude'],
                                    WestVancouver_merged['Longitude'],
                                    WestVancouver_merged['Neighborhood'],
                                    WestVancouver_merged['Cluster Labels']):
    label = folium.Popup(str(nei) + ' Cluster ' + str(cluster), parse_html=True)
    folium.CircleMarker(
        [lat, lon],
        radius=5,
        popup=label,
        color=rainbow[cluster-1],
        fill=True,
        fill_color=rainbow[cluster-1],
        fill_opacity=0.7).add_to(map_clusters)

map_clusters
```
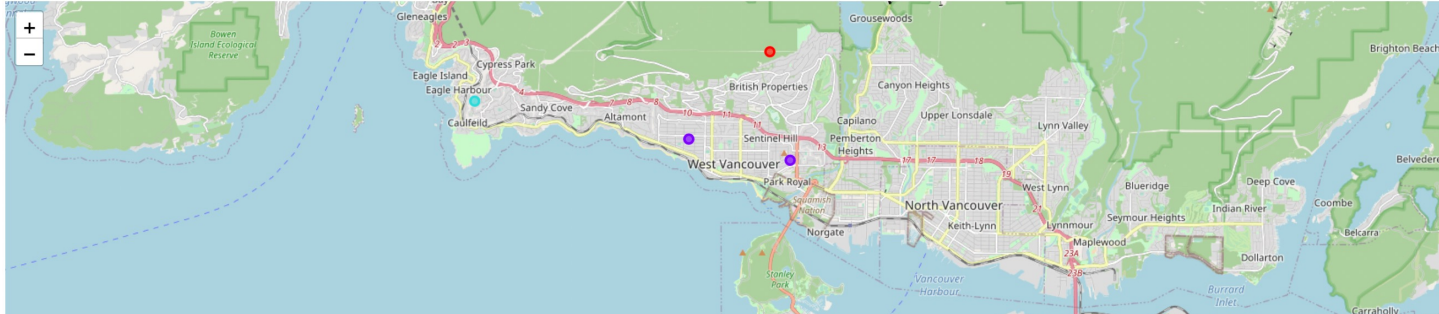
```
['#8000ff', '#2adddd', '#d4dd80', '#ff0000']
```
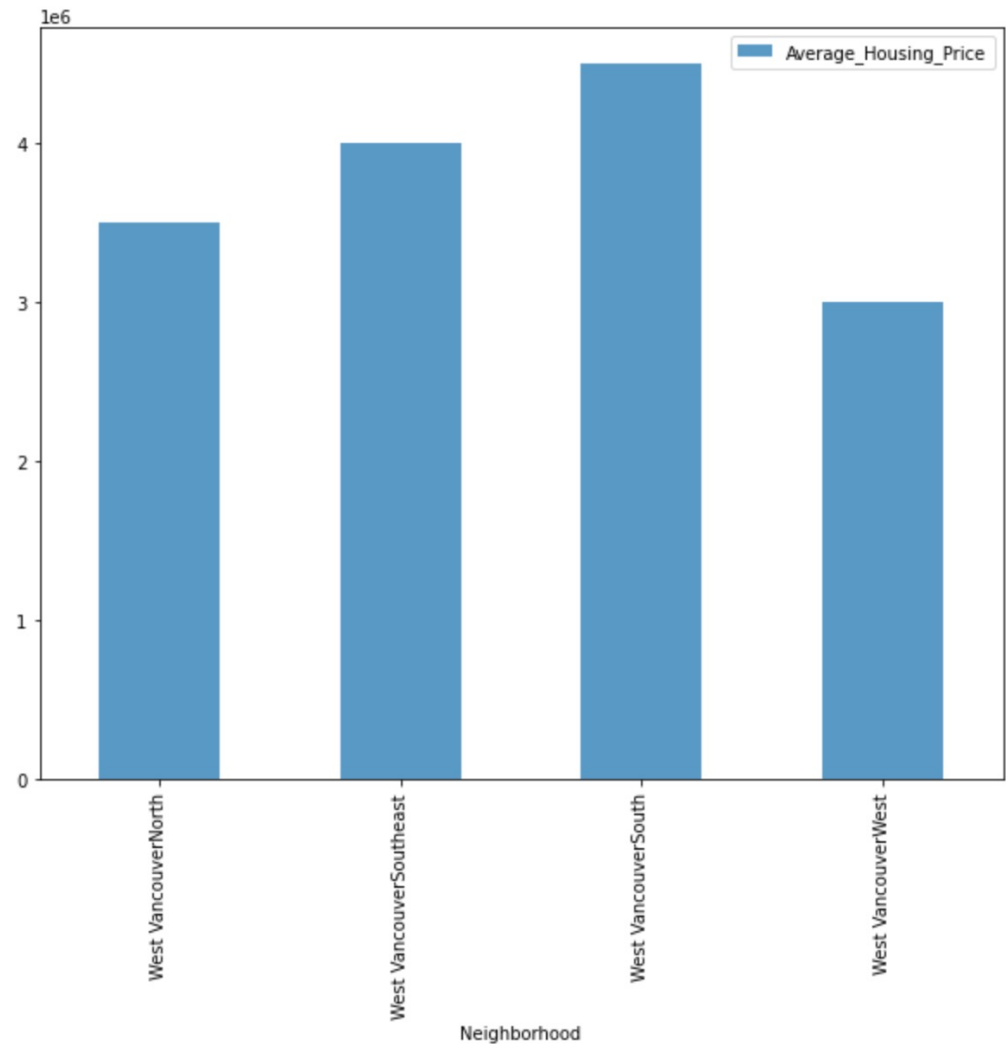

Results

I used folium library of python to determine the map of clusters with k-clusters value of 10

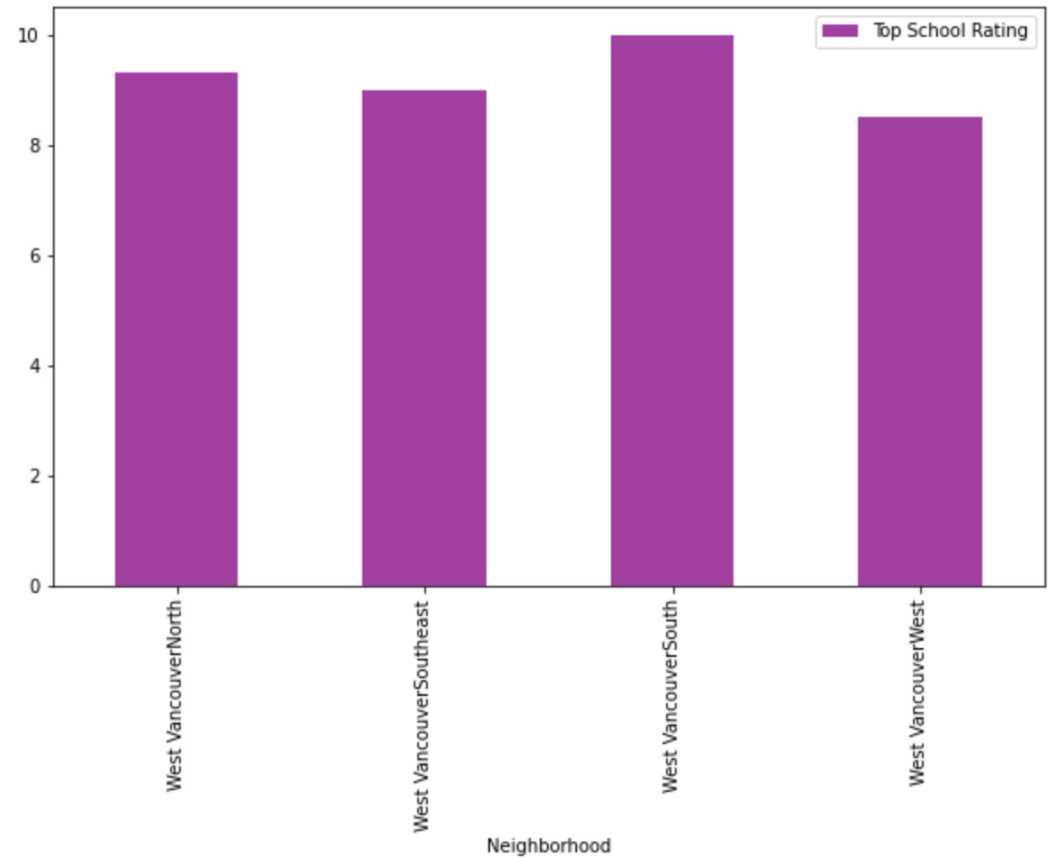**Average Housing Price by Clusters in West Vancouver**

```
WestVancouver_Avg_HousingPrice.plot(kind='bar',figsize=(10,8),alpha=0.75)
```

```
<AxesSubplot:xlabel='Neighborhood'>
```

Conclusion

- In this project, by using k-means cluster algorithm we were able to separate the neighborhood into 3 different clusters, which have very-similar neighborhoods around them. Using the charts above results presented to a particular neighborhood based on average house prices and school rating have been made.

- This project can be used as an application to resolve a real situation that has impacting personal and financial impact using Data Science tools. The mapping with Folium is a very powerful technique to consolidate information and make the analysis and decision better with confidence.

- West Vancouver South neighborhood offers the best elementary school opportunities, however, its average housing price is also the highest in West Vancouver. West Vancouver North neighborhood offers the 2nd best elementary schools, however, its average housing price is 22% lower than that of West Vancouver South. So I would like to recommend West Vancouver North to Adam and his family.

Future Works

- This project can be continued for making it more precise in terms to find best house in West Vancouver. Best means on the basis of all required things(daily needs or things we need to live a better life) around and also in terms of cost effective.