IBM – Applied Data Science Capstone

CAPSTONE Project – Final Report

The Battle of Neighbourhoods

By Lakshmi Sajja

Neighbourhood selection to start a Grocery business

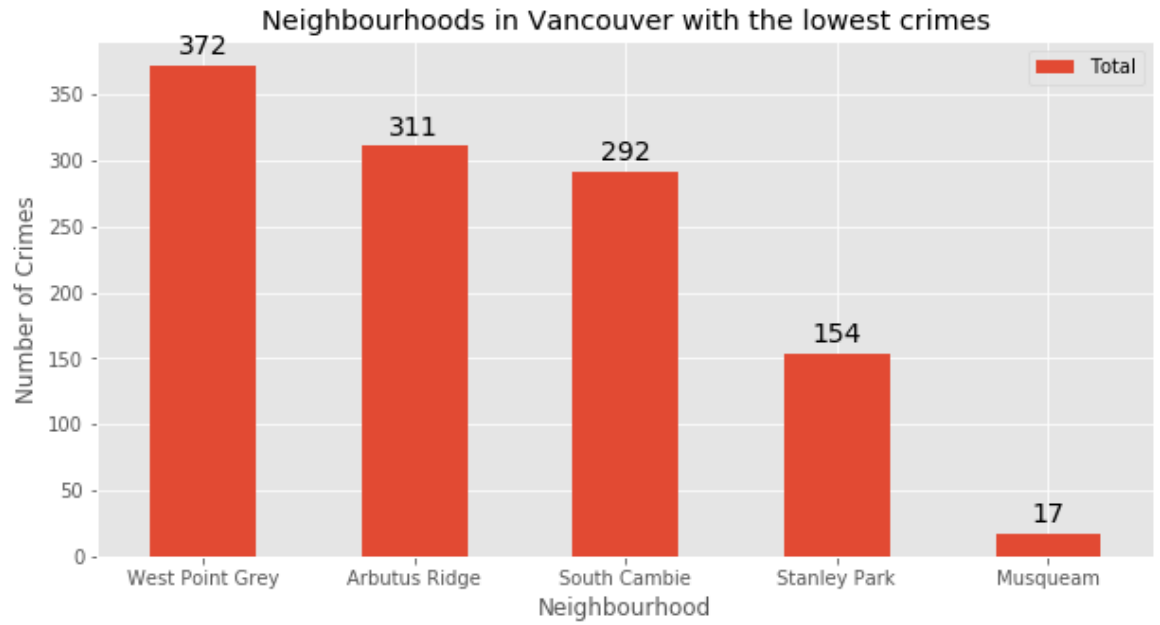
- Neighborhood plays a vital role for a successful business. Budding entrepreneurs thrive to find a safe and viable neighborhood to start a business.
- Crimes like break into commercial property to for theft are on rise and people thinking to enter into similar business should bear in mind criminal activity of the neighborhood before finalizing a location. We look to address this issue by analyzing the crime data of Vancouver City and finding the safest borough and a neighborhood with in the borough.
- Neighborhood in safest Borough with top common venue as Grocery store indicate people choice and need of grocery store

Data acquisition and cleaning

- Type of crime, recorded time and coordinates of Neighborhood in Vancouver from Kaggle data set.
- Neighborhood and Borough from Wikipedia.
- Top venues in each Neighborhood from Foursquare API.
- Raw data set from Kaggle has close to 600,000+ rows which GitHub couldn’t accommodate. Recent crime i.e. in 2018 is considered for data set.
- Fields such as month and hour in which the crime took place has been dropped because they were not in the scope of the problem.

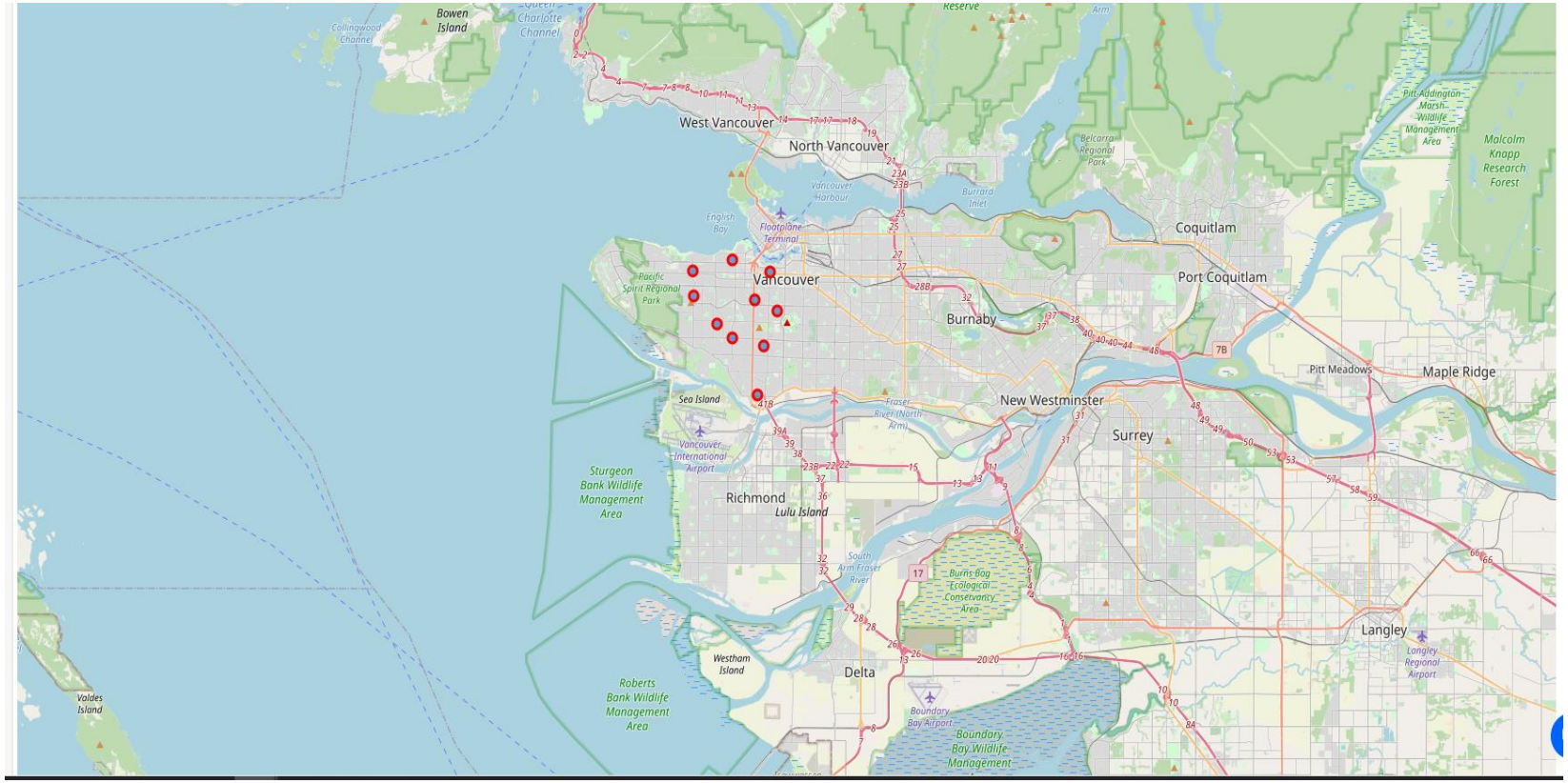
Methodology – Exploratory data analysis

- Visualise the crime repots in different Vancouver boroughs to identity the safest borough and normalise the neighbourhoods of that borough.



Methodology – Exploratory data analysis

- Neighbourhoods plot safest borough – West side



Data Modelling

- Identified top venues in West side borough using Foursquare API. Applies k-means clustering on neighbourhoods based on top venues.

Cluster 1:

```
In [70]: vancouver_merged.loc[vancouver_merged['Cluster Labels'] == 0, vancouver_merged.columns[[0] + list(range(5, vancouver_merged.shape[1]))]]
```

Out[70]:

	Neighbourhood	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
7	Arbutus Ridge	Spa	Grocery Store	Bakery	Pet Store	Nightlife Spot	Yoga Studio	Cosmetics Shop	Deli / Bodega	Dessert Shop	Dim Sum Restaurant

Cluster 2:

```
In [71]: vancouver_merged.loc[vancouver_merged['Cluster Labels'] == 1, vancouver_merged.columns[[0] + list(range(5, vancouver_merged.shape[1]))]]
```

Out[71]:

	Neighbourhood	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
1	Fairview	Coffee Shop	Park	Asian Restaurant	Korean Restaurant	Pharmacy	Chinese Restaurant	Nail Salon	Malay Restaurant	Restaurant	Diner
2	Oakridge	Sporting Goods Shop	Sushi Restaurant	Convenience Store	Park	Sandwich Place	Bubble Tea Shop	Fast Food Restaurant	Pharmacy	Vietnamese Restaurant	French Restaurant
3	Marpole	Sushi Restaurant	Japanese Restaurant	Chinese Restaurant	Pizza Place	Bus Stop	Bubble Tea Shop	Café	Sandwich Place	Gas Station	Falafel Restaurant
4	Kitsilano	Bakery	American Restaurant	Tea Room	Japanese Restaurant	Ice Cream Shop	French Restaurant	Food Truck	Sushi Restaurant	Coffee Shop	Thai Restaurant
5	Kerrisdale	Coffee Shop	Chinese Restaurant	Sushi Restaurant	Pharmacy	Sandwich Place	Boutique	Tea Room	Italian Restaurant	Pizza Place	Noodle House
6	West Point Grey	Coffee Shop	Café	Sushi Restaurant	Japanese Restaurant	Pub	Sporting Goods Shop	Pizza Place	Vegetarian / Vegan Restaurant	Bar	Falafel Restaurant
8	South Cambie	Coffee Shop	Bus Stop	Vietnamese Restaurant	Grocery Store	Light Rail Station	Bank	Gift Shop	Cantonese Restaurant	Sushi Restaurant	Malay Restaurant
9	Dunbar-Southlands	Sushi Restaurant	Italian Restaurant	Coffee Shop	Sporting Goods Shop	Ice Cream Shop	Bakery	Food Truck	Deli / Bodega	Dessert Shop	Dim Sum Restaurant

Cluster 3:

```
In [72]: vancouver_merged.loc[vancouver_merged['Cluster Labels'] == 2, vancouver_merged.columns[[0] + list(range(5, vancouver_merged.shape[1]))]]
```

Out[72]:

	Neighbourhood	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	Shaughnessy	Park	French Restaurant	Yoga Studio	Food & Drink Shop	Cosmetics Shop	Deli / Bodega	Dessert Shop	Dim Sum Restaurant	Diner	Falafel Restaurant

Conclusion and Future Directions

- Built a useful model to find safest Borough and a Neighbourhood to start a business.
- Accuracy of the model has a room for improvement.
- Future scope of the project is consider population of each neighbourhood which will be an important factor on decision making.