



Applied Data Science - Coursera Capstone Project

Battle of the Neighbourhoods -
"Nana's Pizza"

Background

- Vancouver is a coastal seaport city in western Canada, located in the Lower Mainland region of British Columbia. The Greater Vancouver area had a population of 2,463,431 as in 2016, making it the third-largest metropolitan area in Canada. Crime in different forms is a prevalent distress to the people in Metropolitan cities and Vancouver is no exception. 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 neighbourhood before finalizing a location. We look to address this issue by analysing the crime data of Vancouver City and finding the safest borough and a neighbourhood with in the borough which best suits the requirements of our business problem.

Business problem

- The family owned restaurant “Nana’s Pizza” wants to expand their business and decided to take a large step and open another Restaurant in Vancouver. They already own two successful restaurants in Montreal and Quebec. The goal is to find a safe and secure location.
- “Nana’s Pizza” is a family friendly restaurant. To ensure a safe area, the task would be to select the safest neighborhood by analyzing crime data for opening a restaurant and shortlisting a neighborhood where restaurants are not among the most frequented locations. On the other hand they should be as close as possible to the city center.
- Equipped with the skills and the tools to use location data to explore a geographical location, an analysis with focus on the safest neighborhood and examination of its neighborhoods will be done. A proposal of the ten most frequent venues in each neighborhood will be made, so that the best neighborhood to serve “Nana’s Pizza” can be selected.

Data selection

- Key Indicators for area selection:
- Safety of district based on Vanvouver crime statistics
- Popularity of Venues
- Selection of the right neighbourhood within the district

The geographic coordinates of Vancouver will be used to generate an area map of neighborhoods that are considered safe. Those neighborhoods will be grouped and presented in the final result.

Data sources

1. Reliable data set from Kaggle. The data set includes Vancouver crimes from 2003 to 2019: A data set consisting of the crime statistics of each neighborhood court in Vancouver along with the characteristics. Type of crime, recorded in year, month and hour.
2. List of categorized boroughs from Wikipedia. The information on the boroughs will be used to map the existing data, with each borough being assigned to its respective district.
3. Consolidation of new data set. The data set consists of neighbourhoods with districts, crime data and the coordinates. This latitude and longitude data will be retrieved using the OpenCage Geocoder to explore the neighborhood by drawing with Folium and performing exploratory data analysis.
4. Creation of consolidated data set. The data set consists of neighborhoods, districts, and most common venues and neighborhoods along with coordinates. This data will be retrieved using the Four-Square API. With the application of machine learning algorithms, a cluster of the neighborhoods will be created, including maps with a detailed result.

Data Cleaning

- The data provided from Kaggle data source was a large file. The Vancouver crime report included more than 600.000 rows of input. A reduction to current crimes has been used to reduce the number of rows in the dataset. The data set has been provided in github in form of csv file and then uploaded into pandas.
- Furthermore, any information that was not relevant for the analysis of our business problem have been dropped. The resulting information have been used for conclusion.

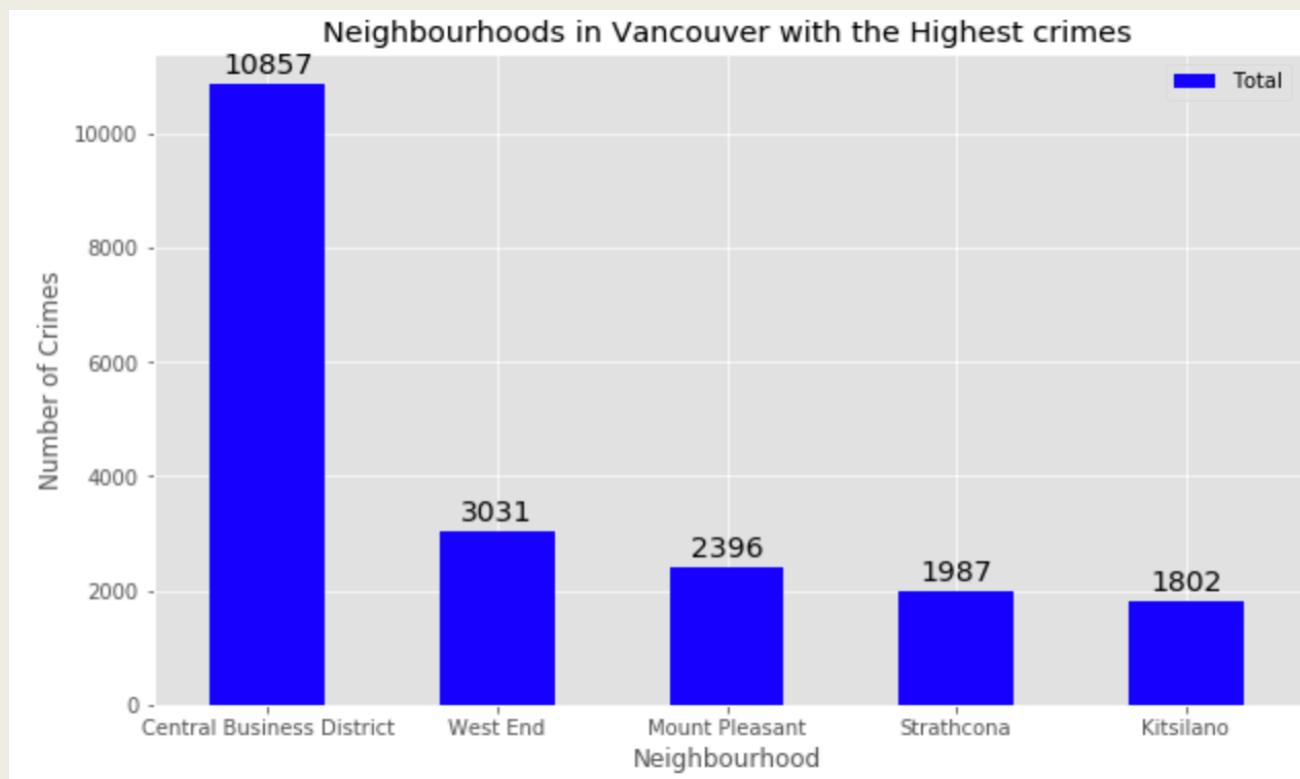
Methodology

- Statistical summary of number of Crimes committed

	YearBreak and Enter Commercial	YearBreak and Enter Residential/Other	YearMischief	YearOther Theft	YearTheft from Vehicle	YearTheft of Bicycle	YearTheft of Vehicle	YearVehicle Collision or Pedestrian Struck (with Fatality)	YearVehicle Collision or Pedestrian Struck (with Injury)	Total
count	4.000000	4.000000	4.000000	4.000000	4.000000	4.000000	4.000000	4.000000	4.000000	4.000000
mean	506.250000	599.250000	1430.250000	1236.750000	3736.500000	539.750000	286.500000	3.250000	368.500000	8707.000000
std	354.409721	488.189427	997.26572	1060.087221	2723.536977	353.955153	226.117226	3.304038	227.060198	5801.870618
min	49.000000	156.000000	187.000000	88.000000	483.000000	36.000000	71.000000	1.000000	111.000000	1182.000000
25%	314.500000	187.500000	843.25000	544.000000	2249.250000	450.000000	186.500000	1.000000	263.250000	5698.500000
50%	594.500000	599.000000	1627.000000	1185.000000	3796.000000	633.000000	235.000000	2.000000	351.500000	9802.000000
75%	786.250000	1010.750000	2214.000000	1877.750000	5283.250000	722.750000	335.000000	4.250000	456.750000	12810.500000
max	787.000000	1043.000000	2280.000000	2489.000000	6871.000000	857.000000	605.000000	8.000000	660.000000	14042.000000

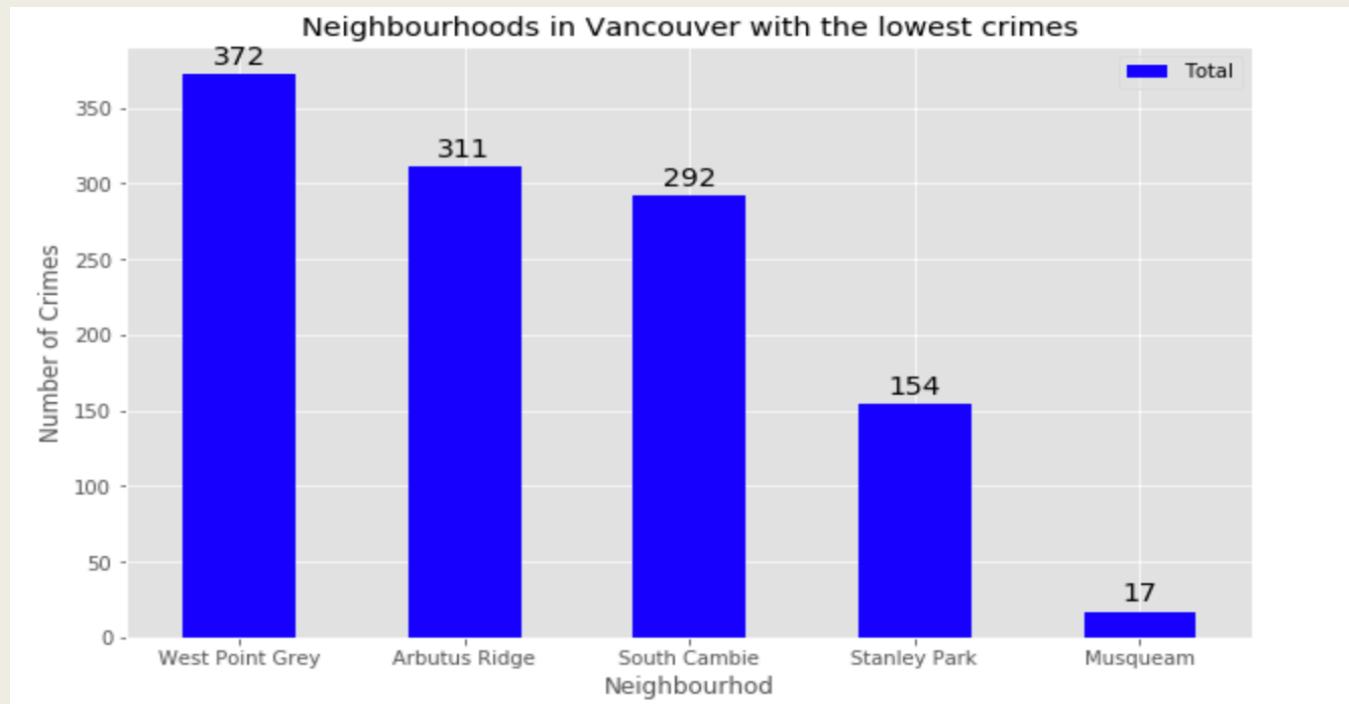
Data Vizualization

- Data Visualization of Crimes committed per neighborhood



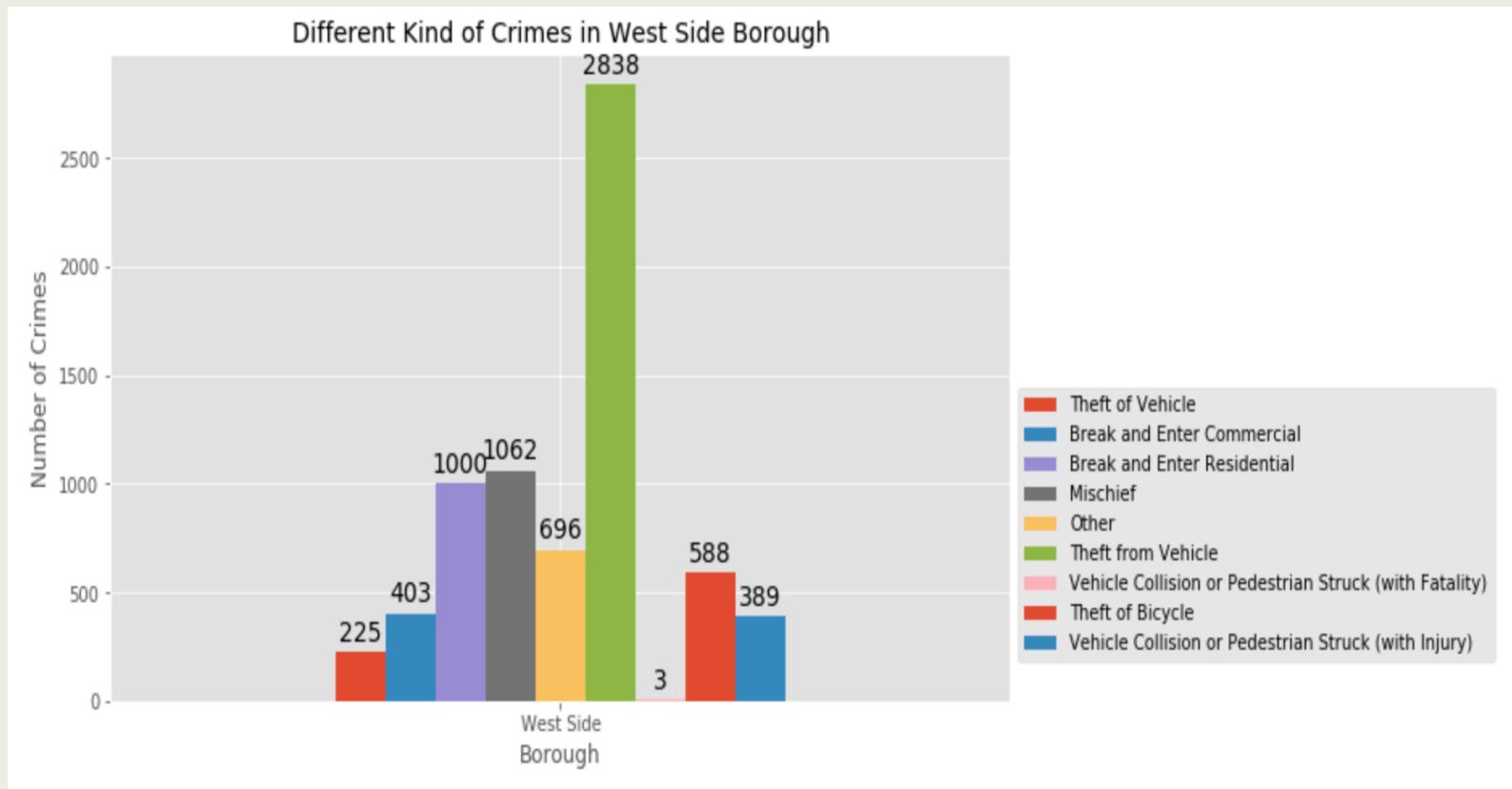
Data Vizualization

- Neighborhoods with lowest crime rates



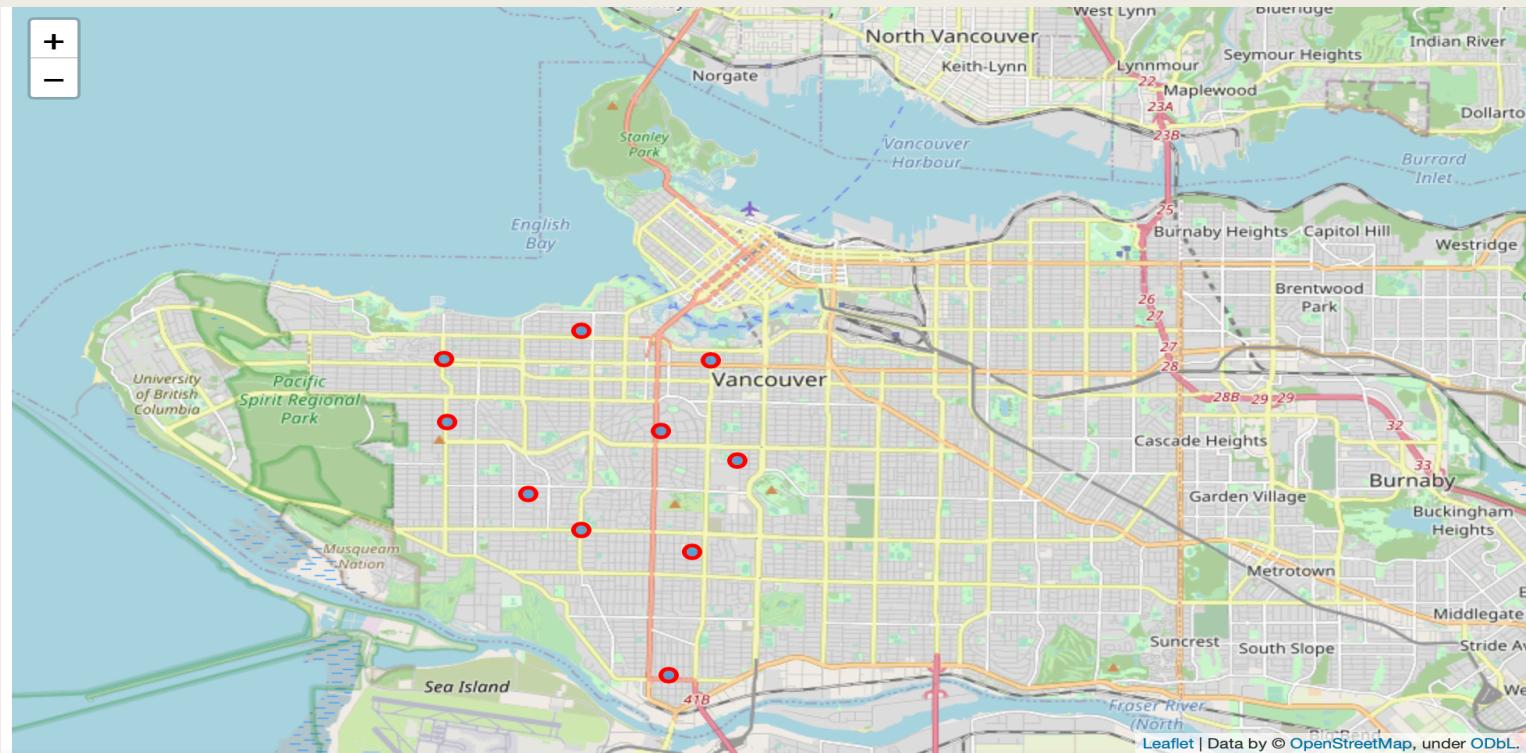
Data Vizualization

- Different kind of crimes in West Side Borough for detailed analysis



Data Vizualization

■ Neighborhoods in West side Borough



Modelling

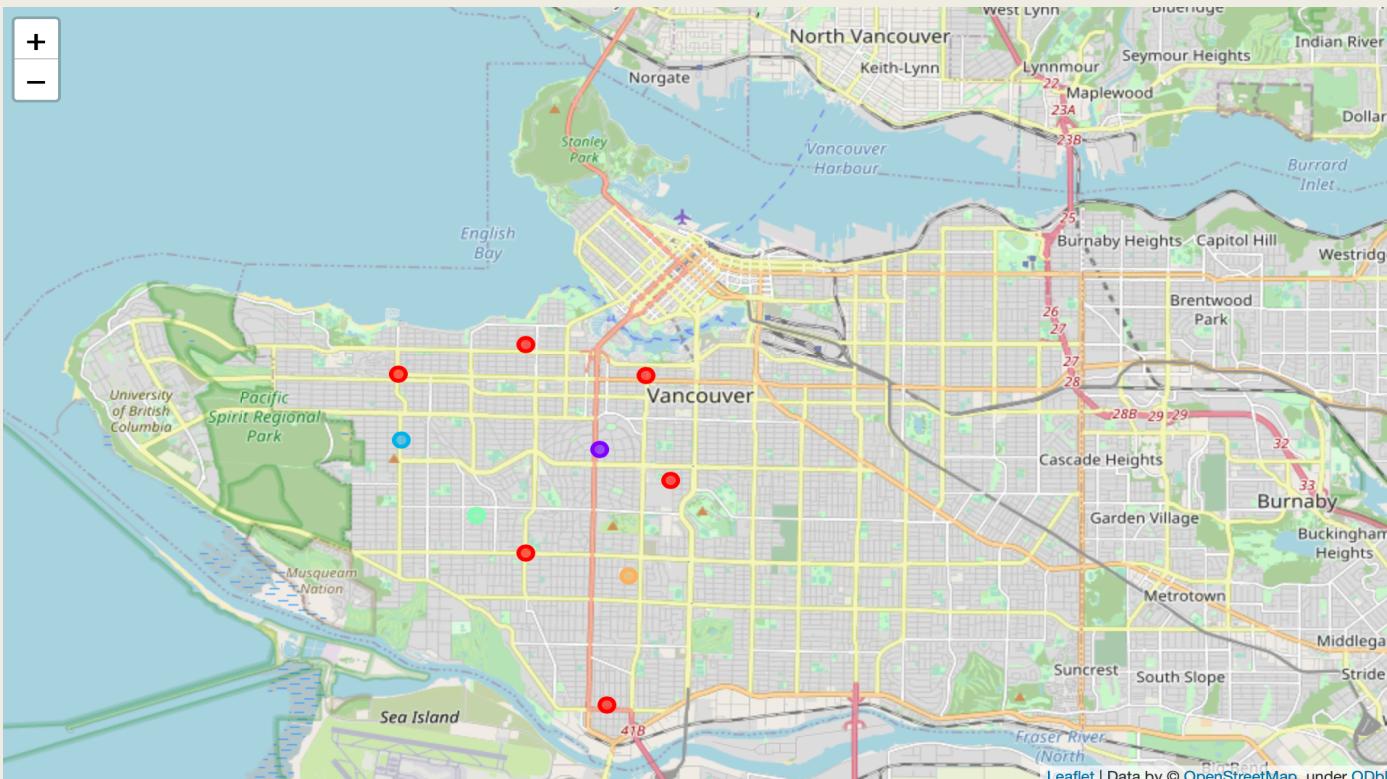
- We generated a consolidated dataset combined with the selected borough West Side. With use of longitude and latitude of West Side neighborhood, all venues within 500 meter radius of each neighborhood by connecting FourSquare API can be done. The result is a json containing all venues in each neighborhood which we convert to pandas data frame.

	Neighbourhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Category
0	Shaughnessy	49.251863	-123.138023	Angus Park	Park
1	Shaughnessy	49.251863	-123.138023	Crepe & Cafe	French Restaurant
2	Fairview	49.264113	-123.126835	Gyu-Kaku Japanese BBQ	BBQ Joint
3	Fairview	49.264113	-123.126835	CRESCENT nail and spa	Nail Salon
4	Fairview	49.264113	-123.126835	Charleson Park	Park

Results

After k-means clustering we create each cluster to evaluate which neighborhood was assigned to each of five clusters.

- Visual cluster categorization



Results

■ Cluster data frame

	Borough	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	West Side	Coffee Shop	Asian Restaurant	Park	Pharmacy	Korean Restaurant	Sandwich Place	Indian Restaurant	Japanese Restaurant	Malay Restaurant	Falafel Restaurant
3	West Side	Pizza Place	Chinese Restaurant	Sushi Restaurant	Japanese Restaurant	Café	Noodle House	Coffee Shop	Dim Sum Restaurant	Plaza	Grocery Store
4	West Side	Bakery	American Restaurant	French Restaurant	Japanese Restaurant	Thai Restaurant	Tea Room	Ice Cream Shop	Sushi Restaurant	Coffee Shop	Food Truck
5	West Side	Coffee Shop	Chinese Restaurant	Tea Room	Pharmacy	Sandwich Place	Sushi Restaurant	Fast Food Restaurant	Japanese Restaurant	Café	Liquor Store
6	West Side	Coffee Shop	Café	Japanese Restaurant	Sushi Restaurant	Bookstore	Pub	Vegetarian / Vegan Restaurant	Liquor Store	Fast Food Restaurant	Falafel Restaurant
8	West Side	Coffee Shop	Bus Stop	Park	Vietnamese Restaurant	Grocery Store	Bank	Malay Restaurant	Sushi Restaurant	Café	Cafeteria

Discussion

- The family owned restaurant “Nana’s Pizza” wants to expand their business and decided to take a large step and open another Restaurant in Vancouver. The objective was to identify one of the safest borough in Vancouver, and an appropriate neighborhood within the borough to set up a Restaurant.
- With Vancouver crime data to identify a safe borough with considerable number of neighborhood was identified. With selecting the borough it was time to choose the right neighborhood where restaurants were not among venues in a close proximity to each other. We achieved this by grouping the neighborhoods into clusters to assist the family to identify the best location to open their business.

Conclusion

- During the analysis an exploration of Vancouvers crime data has been done. After categorization into boroughs, a grouping was done and we could select the safest. Moreover an evaluation of neighbourhoods to be considered, made possible to further shortlist the neighborhoods based on the common venues, to choose a neighborhood which best suits the business problem.