IBM Cousera Capstone

Alex Li

Overview:

1- Introduction
2- Data Collection
3-Methodology
4-Result
5-Conclusion

Introduction

- Vancouver is a vibrant city with a constant increase in population
- It is also ranked as one of the best cities to do business.
- Vancouver is the fastest growing economy in Canada

Business Problem:

- This project tackles a hypothetical data analysis problem to determines the best neighborhood to open a business, whether it is a coffee shop or a restaurant. Vancouver has over 21 neighborhoods and each neighborhood has restaurants, coffee shops and bakeries. The project will perform basic data analysis.

Data Resources

- List of neighborhoods from Wikipedia
- API Foursquare
- Vancouver population from 2016 from website of City of Vancouver
- Python Libraries
- GitHub Vancouver Geojson file

Methodology

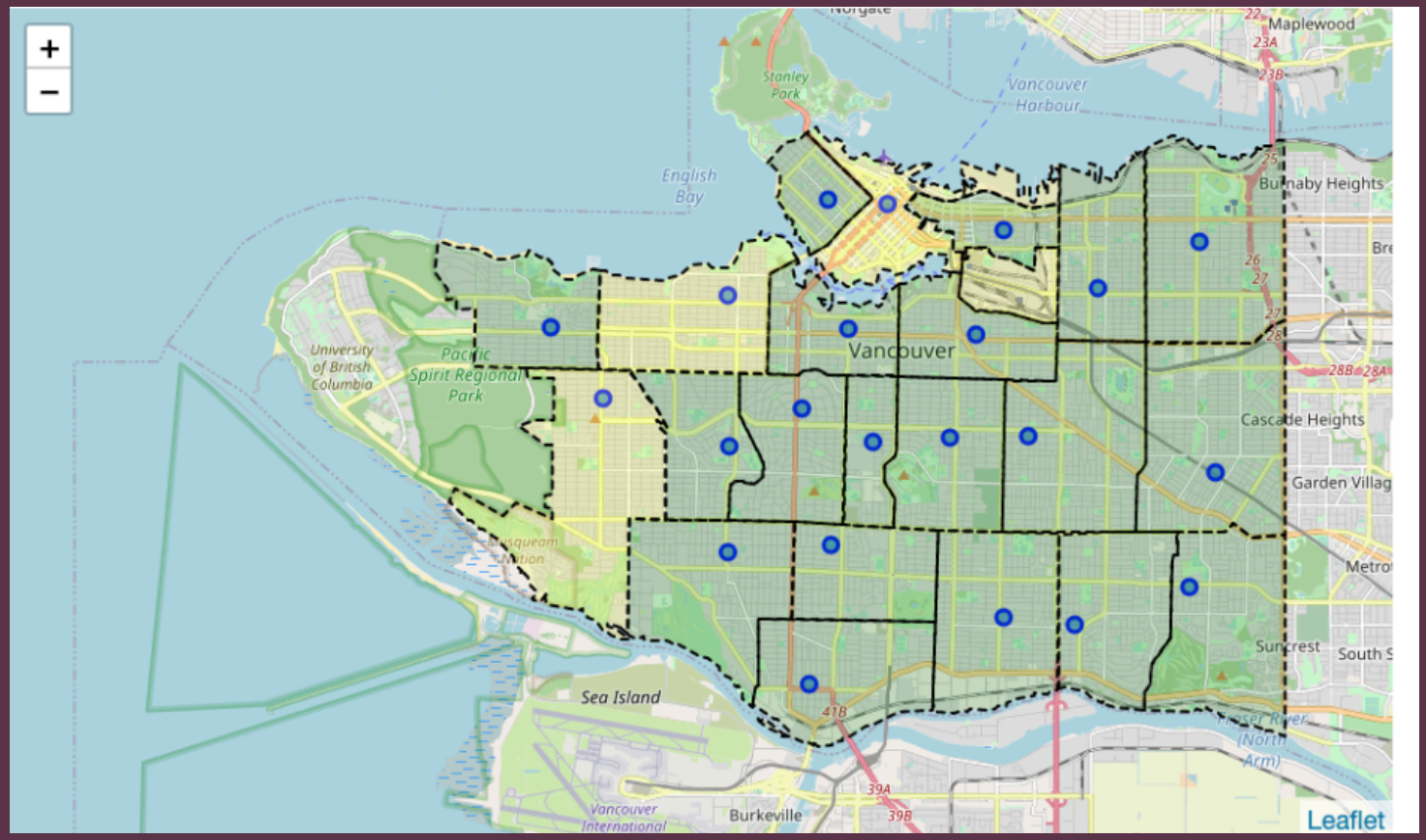
- Exploratory Data Analysis
- Data Wrangling
- Cluster Group
- K Means Clustering
- Folium Map with Clustering and Geo Coordinates

Exploratory Data Analysis:

	Neighborhood	Population	Latitude	Longitude
0	Arbutus-Ridge	15295.0	49.240968	-123.167001
1	Downtown	62030.0	49.283393	-123.117456
2	Dunbar-Southlands	21425.0	49.253460	-123.185044
3	Fairview	33620.0	49.264113	-123.126835
4	Grandview-Woodland	29175.0	49.270559	-123.067942
5	Hastings-Sunrise	34575.0	49.277594	-123.043920
6	Kensington-Cedar Cottage	49325.0	49.247632	-123.084207
7	Kerrisdale	13975.0	49.234673	-123.155389
8	Killarney	29325.0	49.224274	-123.046250
9	Kitsilano	43045.0	49.269410	-123.155267

Population and Geo Coordinates Data are extracted and grouped into a Table

Folium Map of Vancouver

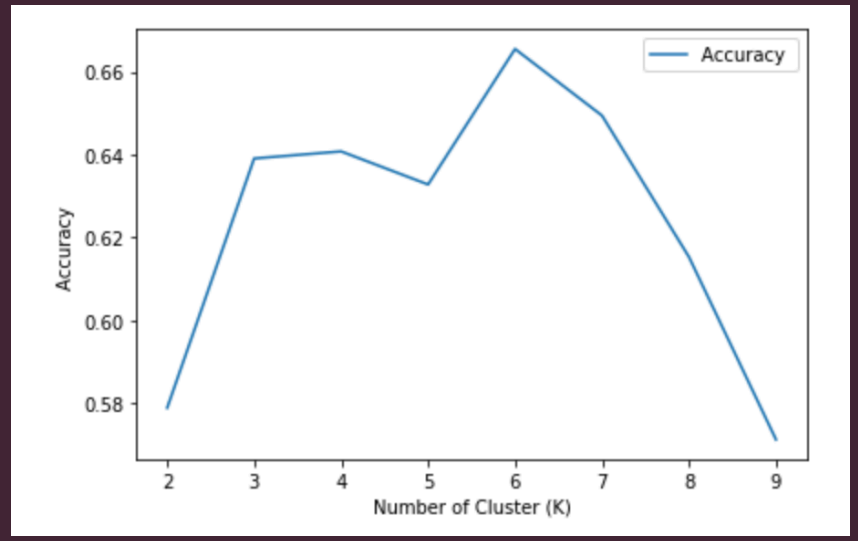


Folium Map of Vancouver with geo coordinates and Geo Json of Vancouver

K Means Clustering

- K Means Clustering requires tuning

- One group encoding is used along with clustering algorithm on the venue categories and population of each neighborhood to determine the best number of clusters .



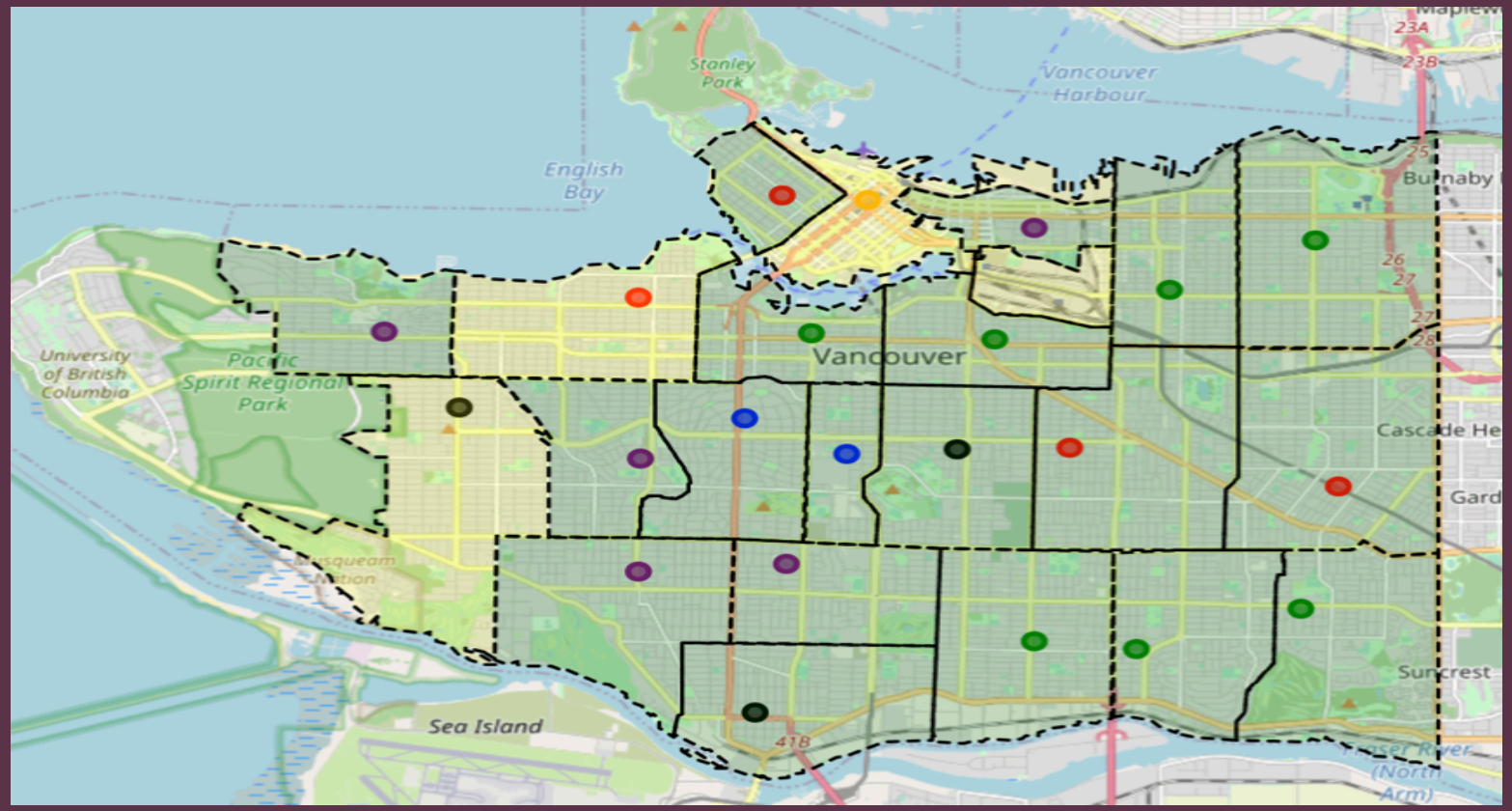
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Arbutus-Ridge	Burger Joint	Bakery	Seafood Restaurant	Dance Studio	Sandwich Place	Lounge	Liquor Store	Event Space	Discount Store	Fast Food Restaurant
1	Downtown	Hotel	Café	Coffee Shop	Sandwich Place	Seafood Restaurant	Dessert Shop	Steakhouse	Bookstore	Concert Hall	Clothing Store
2	Dunbar-Southlands	Sushi Restaurant	Ice Cream Shop	Italian Restaurant	Indian Restaurant	Coffee Shop	Ethiopian Restaurant	Filipino Restaurant	Fast Food Restaurant	Farmers Market	Falafel Restaurant
3	Fairview	Coffee Shop	Park	Asian Restaurant	Japanese Restaurant	Sandwich Place	Malay Restaurant	Salon / Barbershop	Restaurant	Camera Store	Sushi Restaurant
4	Grandview-Woodland	Coffee Shop	Italian Restaurant	Japanese Restaurant	Sushi Restaurant	Café	Park	Indian Restaurant	Pizza Place	Bakery	Record Shop
5	Hastings-Sunrise	Vietnamese Restaurant	Gas Station	Inn	Park	Coffee Shop	Food Truck	Pharmacy	Event Space	Bakery	Sushi Restaurant
6	Kensington-Cedar Cottage	Chinese Restaurant	Coffee Shop	Bus Stop	Filipino Restaurant	Vietnamese Restaurant	Sandwich Place	Convenience Store	Ice Cream Shop	Pizza Place	Café

Results and Discussion

There are total of 6 clusters in total organized based on common venues and the population of the neighborhood.

	Neighborhood	Population	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	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
3	Fairview	33620.0	3	Coffee Shop	Park	Asian Restaurant	Japanese Restaurant	Sandwich Place	Malay Restaurant	Salon / Barbershop	Restaurant	Camera Store	Sushi Restaurant
4	Grandview-Woodland	29175.0	3	Coffee Shop	Italian Restaurant	Japanese Restaurant	Sushi Restaurant	Café	Park	Indian Restaurant	Pizza Place	Bakery	Record Shop
5	Hastings-Sunrise	34575.0	3	Vietnamese Restaurant	Gas Station	Inn	Park	Coffee Shop	Food Truck	Pharmacy	Event Space	Bakery	Sushi Restaurant
8	Killarney	29325.0	3	Pool	Italian Restaurant	Track	Gym	Deli / Bodega	Dessert Shop	Dance Studio	Dim Sum Restaurant	Diner	Discount Store
11	Mount Pleasant	32955.0	3	Coffee Shop	Diner	Breakfast Spot	Sushi Restaurant	Sandwich Place	Hotel	Thrift / Vintage Store	Lounge	Brewery	Indian Restaurant
18	Sunset	36500.0	3	Cosmetics Shop	Dessert Shop	Home Service	Ethiopian Restaurant	Flower Shop	Filipino Restaurant	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Event Space
19	Victoria-Fraserview	31065.0	3	Convenience Store	Gas Station	Fast Food Restaurant	Sandwich Place	Pizza Place	Deli / Bodega	Dessert Shop	Dim Sum Restaurant	Diner	Discount Store

Results and Discussion



Map has cluster labels and geo coordinates of the neighborhoods

Conclusion

- Location and competition are two of the important factors before starting a business

- This project analysis clusters neighborhood based on population and types of businesses.

- This analysis allow investors to understand the types of business in an area and potential customers (population) as well as potential competitions.