

Capstone project- The battle of Neighborhoods

Introduction

People in Canada move around all the time some in hunt for better life better opportunities some for better lifestyle others for various other reasons. The path towards finding the right place what we can call home is not always the case, moving to the new unknown place with myriad of resources available from internet makes its complicated to settle down to finding the right place, and people often don't feel satisfied affecting their peace of mind and health.

To refine this process, when my friend asked me that after finishing his studies he is wanting to move to Vancouver in order to look for better carrier opportunities, this agitated me use the skills of data science to find just about the right neighborhood to move to in the city based on the various factors and criteria crime statistics of the locality and the distance from grocery stores, types of cuisine available in the close vicinity and places to have fun.

The places and selection of nearby available top venues can changed based on personal choice, the main aim of this of project is to explore the safest neighborhoods with available basic amenities when moving to a new city, and live a life in a peaceful and mindful society.

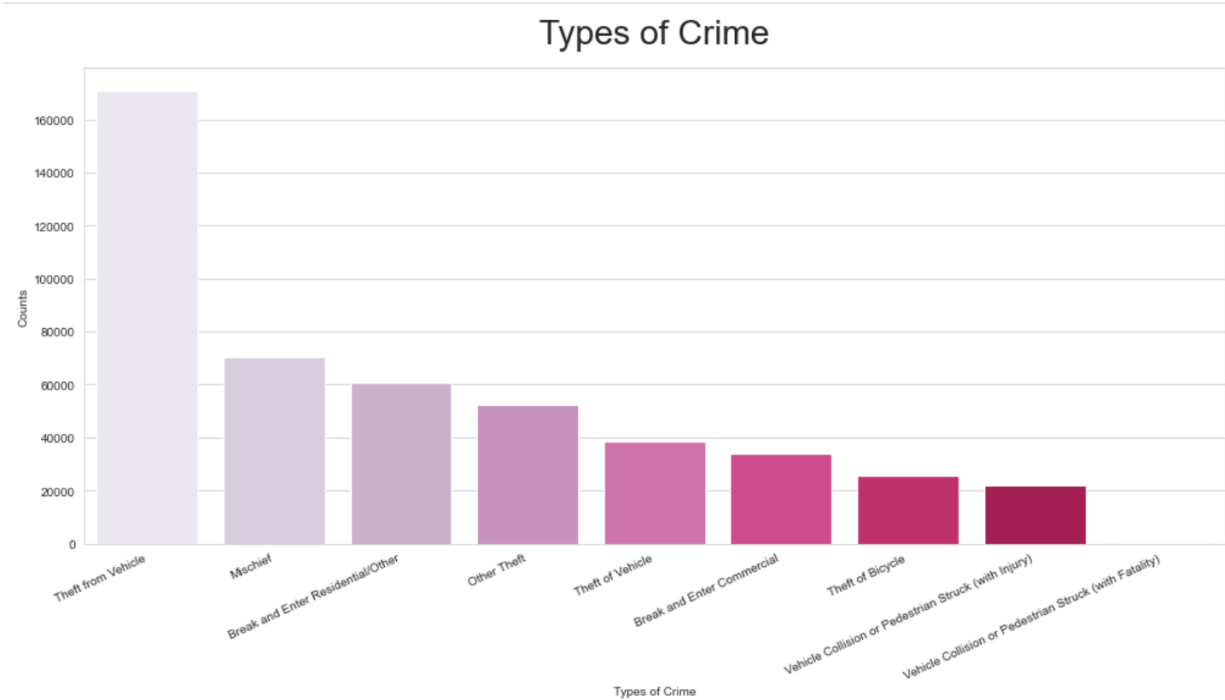
The target audience of this project is mostly people new to the country and culture looking for a safe neighborhood to settle down and live a life one desire and to find the place that one can call home.

Data Description

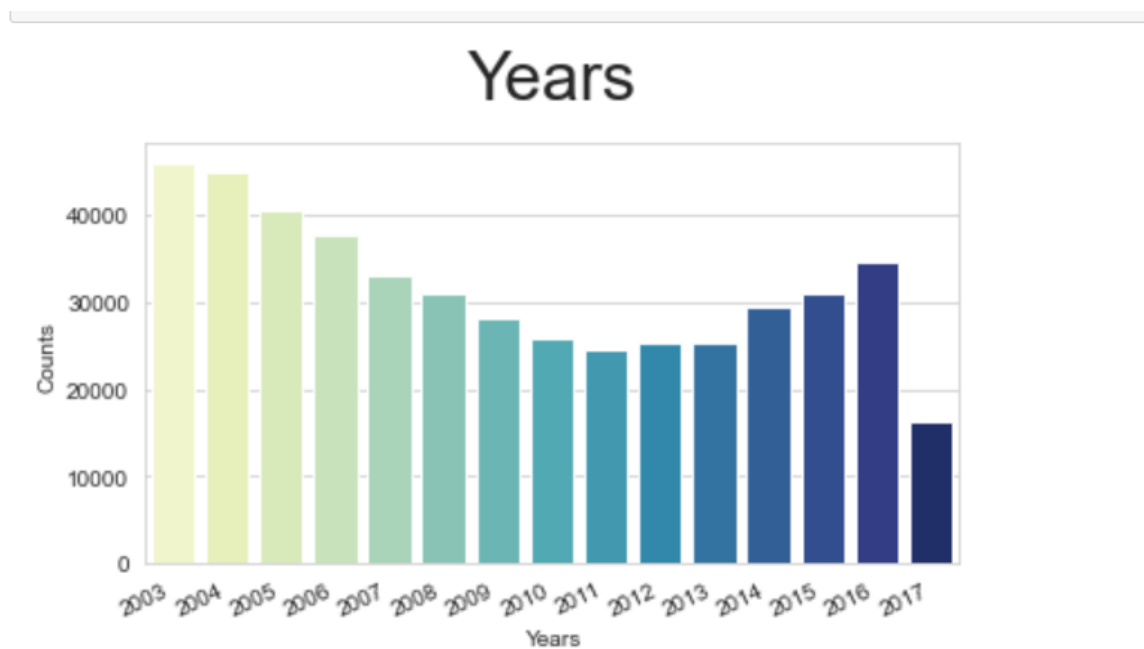
For this project I have used crime statistics data of Vancouver available on kaggle.com and did exploratory data analysis to find the top safest neighborhoods in the city of Vancouver along with their latitude and longitude coordinates and then I employed foursquare location API to get the detailed overview of the best venues the neighborhood has to offer and did clustering of neighborhoods based on their similarity.

Methodology

Based on the data gathered on the crime statistics of the Vancouver area, neighborhoods with highest and lowest crime rate were visualized based on the type of crimes committed.



Majority of the crimes in the city were related to the vehicle theft. After getting the idea of type of crimes that is most likely to happen visual representation of crime along the years was done to understand the trend.



After getting the type and rate at which crime were committed I started filtering neighborhoods based on the frequency of the criminal incidents.

Based on that I was able to determine the top safe and top unsafe neighborhoods based on their crime count.

```
: NEIGHBOURHOOD
Central Business District    110945
West End                    41352
Fairview                    32161
Mount Pleasant              30534
Grandview-Woodland          27180
Renfrew-Collingwood          26761
Kitsilano                   26698
Kensington-Cedar Cottage    24941
Strathcona                  20917
Hastings-Sunrise            18126
Name: TYPE, dtype: int64

: NEIGHBOURHOOD
Musqueam                    532
Stanley Park                3775
South Cambie                5212
Shaughnessy                 5426
West Point Grey             5870
Arbutus Ridge               6066
Kerrisdale                  7447
Dunbar-Southlands           7746
Oakridge                    8035
Killarney                   10475
Name: TYPE, dtype: int64
```

The above figure shows the frequency of crime committed in the different neighborhoods of the city, it can be seen business district, west end and Fairview are areas with highest crime intensity and Musqueam, Stanley park and South Cambie are the ones with lest crime rate. So we steer clear of the areas with high crime rate and focus our neighborhoods search based on these low crime regions.

Exploratory data analysis

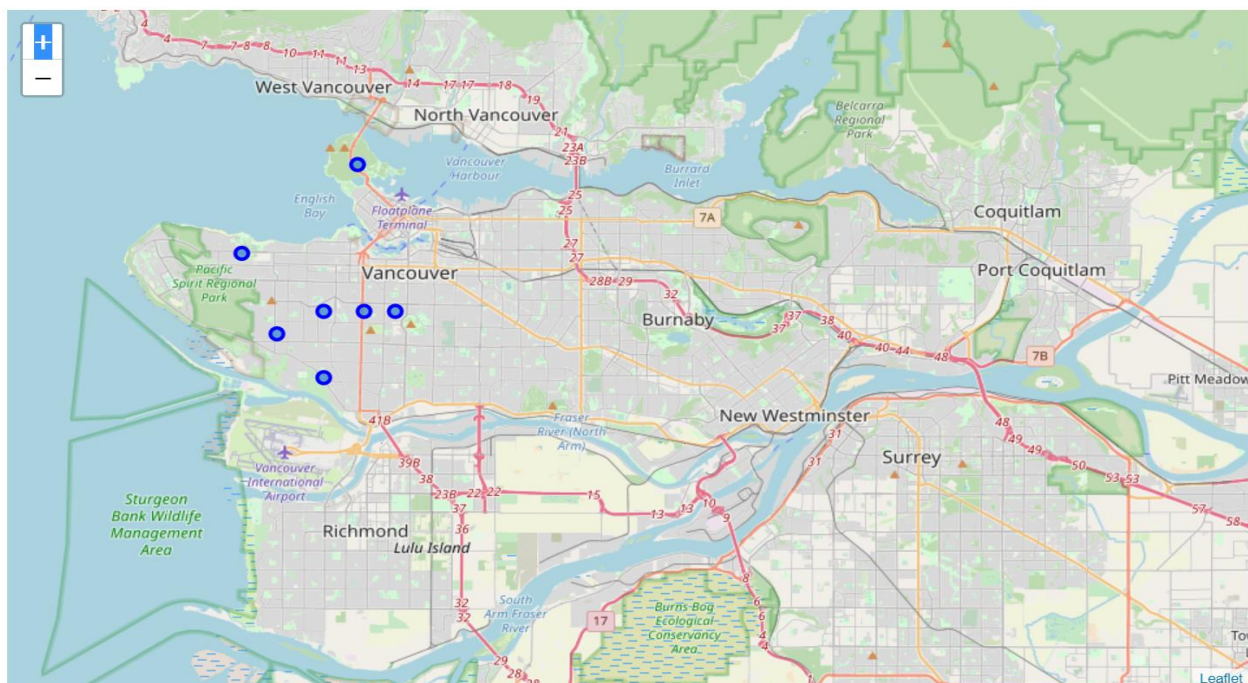
Using the describe function in python I got statistics on crimes in Vancouver data, this returns mean, standard deviation, minimum, maximum, 1st quartile (25%), 2nd Quartile(50%) and the third quartile (75%) for each of the major categories of crime. See the figure below.

TYPE	Break and Enter Commercial	Break and Enter Residential/Other	Mischief	Other Theft	Theft from Vehicle	Theft of Bicycle	Theft of Vehicle	Vehicle Collision or Pedestrian Struck (with Fatality)	Vehicle Collision or Pedestrian Struck (with Injury)
count	24.000000	24.000000	24.000000	24.000000	24.000000	24.000000	24.000000	24.000000	24.000000
mean	1410.041667	2535.666667	2923.208333	2173.333333	7120.375000	1067.500000	1597.958333	10.583333	911.958333
std	1953.128026	1253.868957	3325.900100	3959.905573	9528.086476	1606.500222	1131.298388	8.459194	676.983301
min	17.000000	65.000000	104.000000	1.000000	217.000000	7.000000	40.000000	1.000000	59.000000
25%	322.250000	1813.000000	922.750000	263.750000	2676.250000	177.250000	534.750000	5.500000	386.750000
50%	888.500000	2513.000000	1850.000000	876.000000	4210.000000	288.000000	1494.000000	9.500000	858.500000
75%	2084.500000	3486.250000	3791.500000	2621.250000	7710.500000	1112.000000	2502.500000	13.250000	1247.500000
max	9371.000000	4515.000000	16671.000000	19243.000000	48003.000000	6907.000000	4016.000000	41.000000	3188.000000

The count of each of the major category of crime returns the above values which is the number of neighborhoods of Vancouver. From above table is can concluded that vehicle theft is most common crime in the neighborhoods of Vancouver followed by other mischief and break and enter. The lowest recorded type of crime was vehicle collision or pedestrian struck (with fatality).

Modelling

After determining the safest neighborhoods the city has to offer I did visualization those parts. Visualization was done superimposing the neighborhoods on the map of Vancouver using geopy and leveraging foursquare API to get the latitudes and longitudes of the neighborhoods.



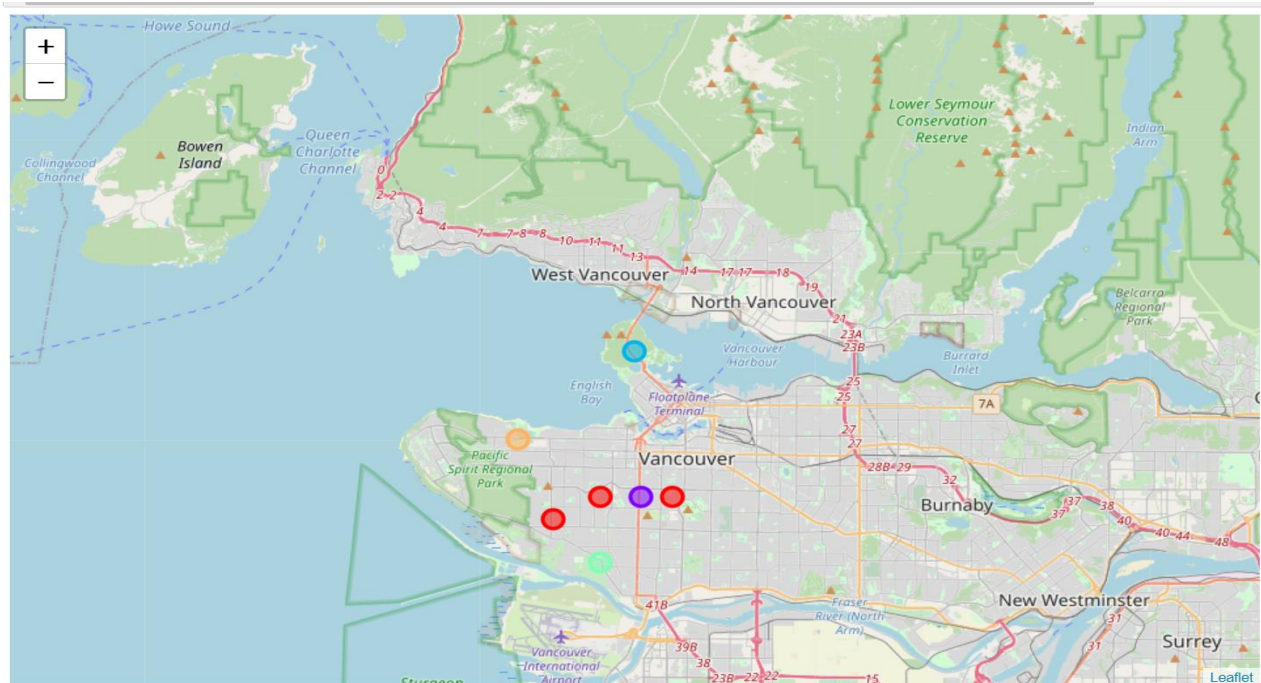
Analyzing the neighborhoods

After visualizing our targeted neighborhoods we form dataframe consisting of these neighborhoods and then leveraging foursquare API we found out the top surrounding venues within 500 meters radius. This results in Jason file displaying all the nearby venues which was further cleaned to form a dataframe showing all the top venues neighborhood has to offer with their location coordinates. Then after forming the dataframe the values were one hot encoded to deploy unsupervised machine learning model of K-means clustering to categorize neighborhoods into different clusters based on their similarity.

	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	Arbutus Ridge	Bakery	Liquor Store	Grocery Store	Shopping Mall	Seafood Restaurant	Sandwich Place	Fast Food Restaurant	Discount Store	Dance Studio	Coffee Shop
1	Dunbar-Southlands	Grocery Store	Liquor Store	Gym	Pet Store	Gym / Fitness Center	Japanese Restaurant	Coffee Shop	Café	Bus Stop	Cantonese Restaurant
2	Kerrisdale	Café	Pool	Golf Course	Park	Vietnamese Restaurant	Dance Studio	Furniture / Home Store	Fast Food Restaurant	Discount Store	Disc Golf
3	Shaughnessy	Bus Stop	Chocolate Shop	Park	Vietnamese Restaurant	Dance Studio	Gift Shop	Furniture / Home Store	Fast Food Restaurant	Discount Store	Disc Golf
4	South Cambie	Coffee Shop	Vietnamese Restaurant	Café	Grocery Store	Malay Restaurant	Gift Shop	Cantonese Restaurant	Park	Cafeteria	Bank
5	Stanley Park	Park	Trail	Lake	Golf Course	Furniture / Home Store	Fast Food Restaurant	Discount Store	Disc Golf	Dance Studio	Vietnamese Restaurant
6	West Point Grey	Gym	Performing Arts Venue	Gym / Fitness Center	Harbor / Marina	Disc Golf	Park	Sandwich Place	Café	Cantonese Restaurant	Gift Shop

Results

Based on the analysis using K-means clustering with the help of foursquare data, I was able to narrow down the hunt for safest neighborhoods in the city of Vancouver based on my preferences of the venues the locality has to offer. The different color represents the different types of neighborhoods based on their similarity of venues city has to offer.



Examining the clusters

1st cluster

	Neighbourhood	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	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
1	South Cambie	49.246464	-123.121603	0	Coffee Shop	Vietnamese Restaurant	Café	Grocery Store	Malay Restaurant	Gift Shop	Cantonese Restaurant	Park	Cafeteria	
4	Arbutus Ridge	49.246305	-123.159636	0	Bakery	Liquor Store	Grocery Store	Shopping Mall	Seafood Restaurant	Sandwich Place	Fast Food Restaurant	Discount Store	Dance Studio	
6	Dunbar-Southlands	49.237864	-123.184354	0	Grocery Store	Liquor Store	Gym	Pet Store	Gym / Fitness Center	Japanese Restaurant	Coffee Shop	Café	Bus Stop	Car Res

The above cluster is comprised of neighbourhoods of South Cambie, Arbutus ridge and Dunbar-Southlands based on their top venues offering which consists of coffee shop, bakery, grocery store, Gyms etc.,

Examining the remaining clusters

The remaining second, third, fourth and fifth cluster differ significantly from each other based on their neighborhood venue offerings. The difference arises due to their location as we can see on the Third cluster Stanley park region is comprised of chocolate Trail, park, golf course and dance studio and restaurant etc. as we can conclude this region popular for recreational

offerings. On examining the fourth and fifth cluster we can see that the popular offering are café, pool, Vietnamese restaurant, performing arts venues and sandwich place.

So based on persons personal choices and interests one can select their preferred neighborhood and then lay plans to find their dream home the respective neighborhood.

Examining second cluster

```
In [43]: van_merged[van_merged['Cluster Labels'] == 1]
```

Out[43]:

	Neighbourhood	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	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
2	Shaughnessy	49.246305	-123.138405	1	Bus Stop	Chocolate Shop	Park	Vietnamese Restaurant	Dance Studio	Gift Shop	Furniture / Home Store	Fast Food Restaurant	Discount Store	Disc Golf

examining the third cluster

```
In [44]: van_merged[van_merged['Cluster Labels'] == 2]
```

Out[44]:

	Neighbourhood	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	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Stanley Park	49.301911	-123.141541	2	Park	Trail	Lake	Golf Course	Furniture / Home Store	Fast Food Restaurant	Discount Store	Disc Golf	Dance Studio	Vietnamese Restaurant

examining the fourth cluster

```
In [45]: van_merged[van_merged['Cluster Labels'] == 3]
```

Out[45]:

	Neighbourhood	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	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
5	Kerrisdale	49.220985	-123.159548	3	Café	Pool	Golf Course	Park	Vietnamese Restaurant	Dance Studio	Furniture / Home Store	Fast Food Restaurant	Discount Store	Disc Golf

In []:

examining the fifth cluster

```
In [46]: van_merged[van_merged['Cluster Labels'] == 4]
```

Out[46]:

	Neighbourhood	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	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
3	West Point Grey	49.268102	-123.202643	4	Gym	Performing Arts Venue	Gym / Fitness Center	Harbor / Marina	Disc Golf	Park	Sandwich Place	Café	Cantonese Restaurant	Gift Shop

Discussion

The whole methodology of the project is based on the neighborhood data using foursquare API and how integrating them with criminal statistics can be employed to significantly improve the possibility of search of safest neighborhood when a person is planning to move a new unknown city. This shows the importance of neighborhood and their role in assisting the decision making process.

Using this project I tried to explore the crime statistics of neighborhoods of the city of Vancouver and then conducted exploratory data analysis on the different types of crime committed on different neighborhoods and using this I was able to narrow down the list of

safest places in city based on their crime count the severity of the crimes. This data was then used to explore the safest neighborhoods in city based on their top venue offering for example the person who likes to engage in recreational activities can look for house in areas on second cluster i.e. Stanley park, and if a person is not so outgoing and want all the basic amenities in their close vicinity i.e. grocery store, bus stops, gyms and coffee shops can look for neighborhoods in the first cluster comprising of South Cambie, Arbutus ridge and Dunbar-Southlands.

As the scope of this project was based on the battle of neighborhoods I only considered the top venues offering of particular and limited number of neighborhoods, one can dig further to a bit more of web scrapping and look for more holistic view of an neighborhood based on their housing prices and ease of commute and police surveillance and security.

Conclusion

Search for a locality that has the best to offer can be challenging and time-consuming relying on data just based on internet, as it could be vague and very vast. Through the tools I employed in this project I filtered down the safest neighborhoods the city can offer and then further analysis of each neighborhood to get idea of the top venues the locality has to offer, the search for the dream home in a city unknown to a person can be very strenuous and difficult for a person thus with application of this project that problem had been addressed. The future work can include looking for home based one's spending power and type of locality the person want to settle down to so that his/her family could safe and live the best out of everything. I'm glad to complete the search for the best neighborhood with minimum effort to enhance the very vague and time-consuming problem.