

# Relationship between Neighbourhoods and Crime Rates in Vancouver City

Capstone Project for IBM Data Science on Coursera

by Huabei You

January 2020

- Introduction
- Materials and Methods
- Results and Discussion
- Conclusion and Future Work

# Introduction

## Background

Recent studies in social science suggest that the local crime rate may be correlated with the characteristics of the neighbourhood. On the other hand, neighbourhoods can be categorized by the types of venues located within them. Since high crime rates can have a negative impact on the property value, it is in stakeholders' best interests to invest in areas with low crime rate not only at the moment but also in the future.

## Problem

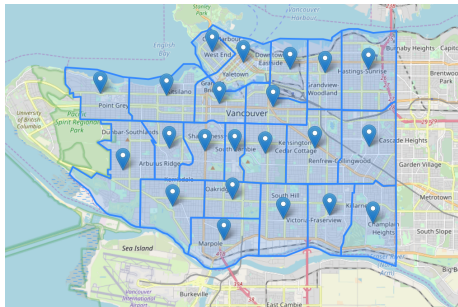
Is there any relationship between crime rates and common venue types in the neighbourhood? Can we identify the best neighbourhood for real estate investment?

Vancouver city was selected as the real life example in this study.

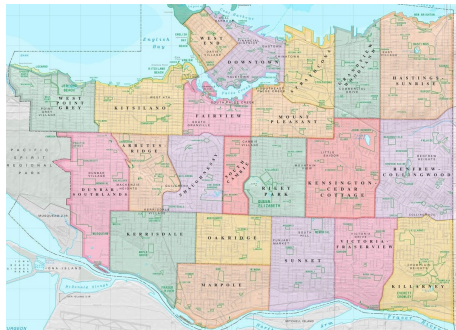
- Geographic data of the city: Vancouver city open data portal
- Crime data: Vancouver police department website
- Venue data: Foursquare Places API

# Materials and Methods

## Pre-processing Data



**Figure:** Neighbourhood divisions of Vancouver City based on the geographic data set.



**Figure:** Official Neighbourhood Map of Vancouver City.

The geographic data set contains the boundaries of 22 neighbourhoods in Vancouver city. The neighbourhood boundaries are visualized in Fig. 1 which is identical to the official map of Vancouver city (Fig. 2).

# Materials and Methods

## Pre-processing Data

The crime date set consists of 627174 records from year 2003 to 2019 with information about the type, date, time, block name, neighbourhood name, and coordinates of the crime.

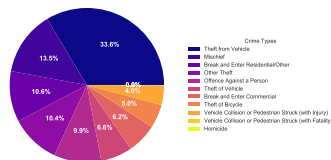
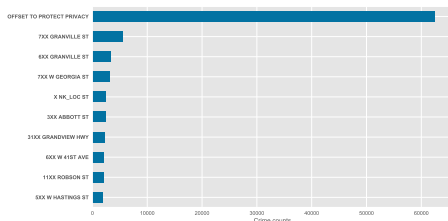


Figure: Top 10 Dangerous Blocks.

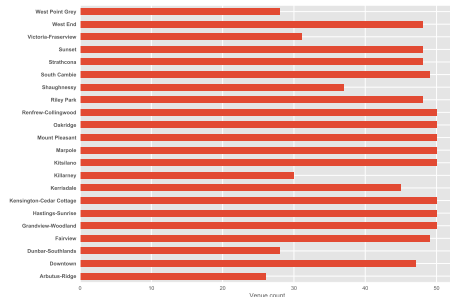
Figure: Crime Type Distributions.

- Fig. 3 shows that "OFFSET TO PROTECT PRIVACY" is the most dangerous block. Further inspection shows that crime records with this block name all have missing data of neighbourhood and coordinates.
- This is because Vancouver police removed location information for all crimes labeled "Offence Against a Person" or "Homicide" to protect privacy.
- As a result, crime types analysed in this study do not include "Offence Against a Person" and "Homicide" for the lack of location data.

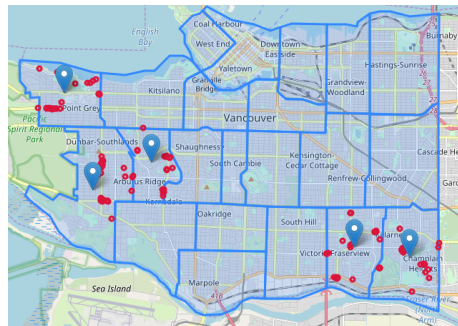
# Materials and Methods

## Pre-processing Data

The venue data set is obtained from Foursquare Place API.



**Figure:** Number of venues returned by foursquare API.

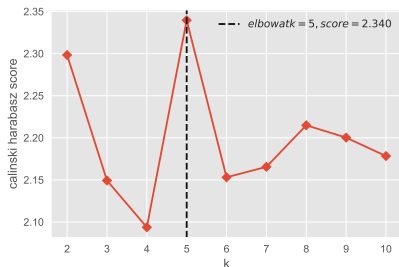


**Figure:** Venue distributions in neighbourhoods with less data.

- Not all queries have received the required amount of information (Fig. 5).
- In order to eliminate the influence of searching radius, venue locations in five neighbourhoods with less data are plotted in Fig. 6.
- The visualization shows that the radius used here is enough to cover the entire area without data overlap.

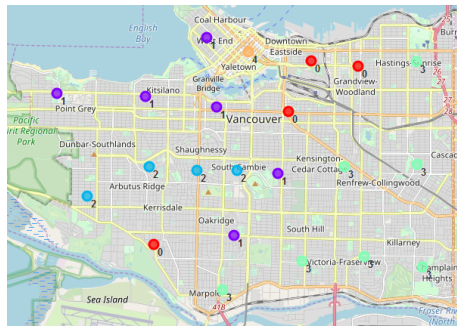
# Results

## Clustering Result



**Figure:** The Calinski Harabasz Score Based Elbow for K-means Algorithm.

- K-means algorithm is used to partition neighbourhoods into k clusters in which a data point belongs to the cluster where the sum of squares distance between within-cluster points to the centroid of the cluster is minimized.
- The optimal number of clusters is 5 based on the Calinski-Harabasz score (Fig. 5).



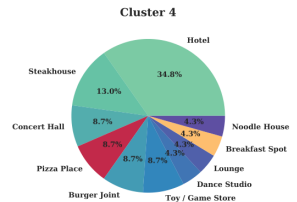
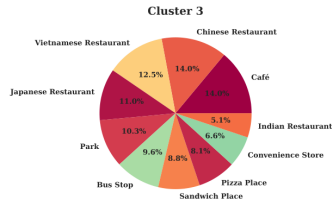
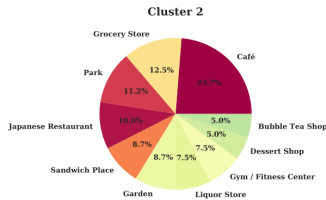
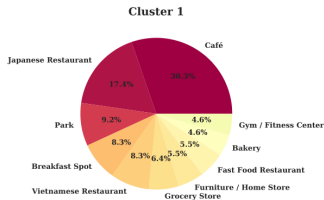
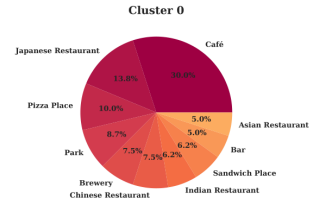
**Figure:** Clustering Results.



# Results and Discussion

## Clustering Result

Top 10 common venues in each cluster.

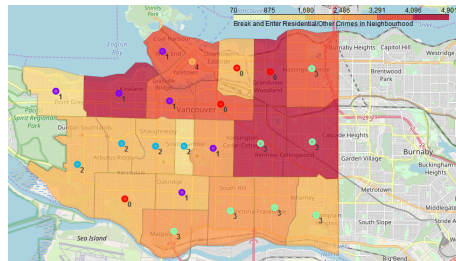
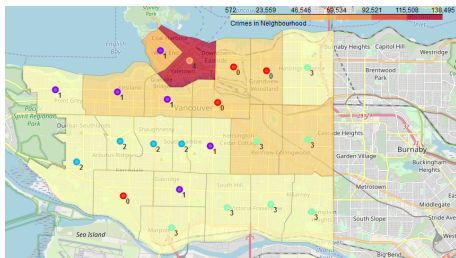


All clusters have different venue compositions, especially Cluster 4 which corresponds to the Downtown area.

# Results and Discussion

## Crime Choropleth Map

A choropleth map that colors the neighbourhoods in different shade based on the total number of crimes in that respective area is on the left. The choropleth map based on the number of occurrence of Break and Enter Residential/Others is on the right.



- Two maps show different trends.
- Downtown area is the most dangerous area overall, but houses and apartments there are generally safer to live in than places in Kitsilano, Grandview-Woodland, Kensington-Cedar Cottage, and Renfrew-Collingwood.
- Despite having low crime rates overall, about half of the neighbourhoods in Cluster 1 and 3 have higher rates for Break and Enter Residential/Others to occur.

# Results and Discussion

## Crime-Venue Correlation

	Break and Enter Commercial	Break and Enter Residential/Other	Blasphemy	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)
Hotel	0.22	0.52	0.15	0.15	0.15	0.15	0.15	0.15	0.15
Grocery Store	-0.25	-0.25	-0.25	-0.25	-0.25	-0.25	-0.25	-0.25	-0.25
Golf Course	-0.25	-0.25	-0.25	-0.25	-0.25	-0.25	-0.25	-0.25	-0.25
Japanese Restaurant	-0.25	-0.11	-0.45	-0.25	-0.25	-0.25	-0.25	-0.25	-0.25
Breakfast Spot	0.36	0.48	0.22	0.25	0.25	0.49	0.34	0.15	0.35
Liquor Store	-0.25	-0.25	-0.25	-0.25	-0.25	-0.25	-0.25	-0.25	-0.25
Pizza Place	0.20	0.52	0.31	0.22	0.19	0.10	0.58	0.27	0.35
Bank	-0.15	-0.25	-0.25	-0.25	-0.25	-0.25	-0.25	-0.25	-0.25
Gym/Fitness Center	-0.21	-0.15	-0.25	-0.22	-0.22	-0.11	-0.25	-0.25	-0.25
Sandwich Place	-0.21	-0.15	-0.15	-0.15	-0.15	-0.25	-0.25	-0.25	-0.25
Fast Food Restaurant	-0.25	-0.25	-0.25	-0.25	-0.25	-0.25	-0.25	-0.25	-0.25
Cafe	-0.15	-0.25	-0.25	-0.25	-0.25	-0.25	-0.25	-0.25	-0.25
Park	0.25	0.25	-0.25	-0.25	-0.25	-0.25	-0.25	-0.25	-0.25
Indian Restaurant	0.10	0.25	0.25	-0.25	0.25	0.11	0.45	0.17	0.24
Convenience Store	-0.22	0.12	-0.15	-0.11	-0.15	-0.25	0.17	0.12	0.20
Chinese Restaurant	-0.21	0.25	-0.25	-0.25	-0.25	-0.25	0.11	0.27	0.25
Bakery	-0.15	0.11	-0.11	-0.15	-0.15	-0.25	0.11	0.22	0.14
Vietnamese Restaurant	-0.25	0.45	-0.25	-0.25	-0.25	-0.25	0.25	0.25	0.25
Bus Stop	-0.25	0.25	-0.25	-0.25	-0.25	-0.25	-0.25	-0.25	-0.25
Asian Restaurant	0.21	0.42	0.25	0.25	-0.25	-0.25	0.25	0.25	0.25
Pharmacy	0.15	0.25	-0.25	-0.15	-0.15	-0.25	0.25	0.15	0.25
Dessert Shop	0.20	-0.25	0.25	0.25	0.25	0.25	-0.25	0.25	0.25

- If the absolute value of correlation coefficient is below 0.6, the collinearity among the predictor variables is considered weak.
- The only venue category that has a strong relationship with crimes is Hotel which is positively correlated to the occurrence of all types of crime analysed in this study except for Break and Enter Residential/Other and Theft of Vehicle.

# Conclusion and Future Work

- Neighbourhoods in Vancouver are divided to five clusters based on the types of venues located within them.
- Neighbourhoods in Cluster 2, namely Arbutus-Ridge, Dunbar-Southlands, Shaughnessy, and South Cambie, are safer areas to live and therefore, would be recommended to real estate companies and individual house buyers.
- Little to no linear correlation was found between venue categories and crime types with one exception being Hotel.
- The results of this study are strongly dependent on the quality of venue and crime data sets which are inadequate for some neighbourhoods and some types of crimes, respectively.
- Changes in Foursquare database can affect the result of clustering.
- Future work should focus on adding more venue information to better represent the neighbourhoods.

# Thank You