Analyzing house prices in Vancouver

Omar Altrad

Introduction

- **Project Idea**
- My idea for the Capstone Project is to show that when driven by venue and location data from FourSquare, backed up with open source crime data in Vancouver, that it is possible to predict the price of a home based on your own criteria.

Introduction

- Utilizing the Foursquare location data, crime statistics and restaurants availability, schools and other convenient shopping stores in Vancouver for homebuyer's decisions and house price estimation.
- FourSquare is one of the best sites that will tell you all about places to go, things to see, restaurants to eat at, bars to drink in, nightclubs to part the night away in and then where to go in the morning to get breakfast and a strong coffee.
- Foursquare alone can't give you hints of which areas are perfect to settle or live.
- If we combine data from several sources we might be able to build a model that will help homebuyers to take such a life decision with accurate if not close to accurate about the real price.

Introduction

- You like to plan ahead and always review your options and make your choices about where you will settle and live.
- You are in a city and you plan to move to a different one (example Vancouver) to settle after you got a job offer.
- You start exploring the chances to buy a home in an area in Vancouver.
- But you know no one in Vancouver to show you around to all the best areas and even if so they don’t have much of information about the house prices or they don’t know what you really looking for.
- Also you want to avoid directly asking a real state agent and even if so you need a full and accurate estimate of the prices based on your own criteria.

Data

- **Data sources**
  - Vancouver Crime Data to determine the crime rate at different Vancouver areas

    This dataset reflects reported incidents of crime (with the exception of murders where data exists for each victim) that occurred in the City of Vancouver from 2005 to present, minus the most recent seven days (Source:https://data.vancouver.ca/datacatalogue/crime-data-details.htm?)
  - Top attractions and restaurants acquired from FourSqaure website within the top attractions, find the schools, universities, transportations and parks and their geographical locations

Data content and cleaning

- Crime Data:

[17]:

	TYPE	YEAR	MONTH	DAY	HOUR	MINUTE	HUNDRED_BLOCK	NEIGHBOURHOOD	X	Y	Latitude	Longitude
0	Other Theft	2003	5	12	16.0	15.0	9XX TERMINAL AVE	Strathcona	493906.5	5457452.47	49.269802	-123.083763
1	Other Theft	2003	5	7	15.0	20.0	9XX TERMINAL AVE	Strathcona	493906.5	5457452.47	49.269802	-123.083763
2	Other Theft	2003	4	23	16.0	40.0	9XX TERMINAL AVE	Strathcona	493906.5	5457452.47	49.269802	-123.083763
3	Other Theft	2003	4	20	11.0	15.0	9XX TERMINAL AVE	Strathcona	493906.5	5457452.47	49.269802	-123.083763
4	Other Theft	2003	4	12	17.0	45.0	9XX TERMINAL AVE	Strathcona	493906.5	5457452.47	49.269802	-123.083763

Data content and cleaning

Crime data

- Keep only the related columns of the data as no need for any other column in feature analysis.

[19]:

	TYPE	YEAR	NEIGHBOURHOOD	Latitude	Longitude
99207	Break and Enter Commercial	2005	Strathcona	49.270524	-123.087369
99208	Theft from Vehicle	2005	Oakridge	49.233524	-123.118504
99209	Break and Enter Commercial	2005	West End	49.285132	-123.123461
99210	Theft from Vehicle	2005	Oakridge	49.233524	-123.118504
99211	Theft of Vehicle	2005	Renfrew-Collingwood	49.258095	-123.028596

Data content and cleaning

- Crime data
- The number of unique Crimes categories based on the Primary type is extracted and then the most common crime is written to a new dataframe.

[25]:

	TYPE	Count
6	Theft from Vehicle	136752
3	Mischief	58421
1	Break and Enter Residential/Other	47441
5	Other Theft	46980
4	Offence Against a Person	46831
0	Break and Enter Commercial	27365
8	Theft of Vehicle	25955
7	Theft of Bicycle	23082
10	Vehicle Collision or Pedestrian Struck (with I...	18230
9	Vehicle Collision or Pedestrian Struck (with F...	208
2	Homicide	180

Data content and cleaning

- Foursquare data (Features)
- The generated data from each URL request is converted into a Jason file. Then only the related data fields are extracted and stored into a separate data frame. See below an example:

]:

	venue.name	venue.categories	venue.location.lat	venue.location.lng
0	CF Pacific Centre	[{'id': '4bf58dd8d48988d1fd941735', 'name': 'S...	49.282980	-123.118361
1	City Square Shopping Centre	[{'id': '4bf58dd8d48988d1fd941735', 'name': 'S...	49.260947	-123.116535
2	Centrepoint Shopping Mall	[{'id': '4bf58dd8d48988d1fd941735', 'name': 'S...	49.259030	-123.101381
3	Harbour Centre Mall	[{'id': '4bf58dd8d48988d1fd941735', 'name': 'S...	49.284562	-123.112296
4	Granville Mall	[{'id': '4bf58dd8d48988d1fd941735', 'name': 'S...	49.281772	-123.120302
5	King Edward Mall	[{'id': '4bf58dd8d48988d1fd941735', 'name': 'S...	49.248629	-123.125933
6	Hudson's Bay	[{'id': '4bf58dd8d48988d1f6941735', 'name': 'D...	49.282483	-123.117514
7	Kingsgate Mall	[{'id': '4bf58dd8d48988d1fd941735', 'name': 'S...	49.262375	-123.097736
8	Apple Pacific Centre	[{'id': '4bf58dd8d48988d122951735', 'name': 'E...	49.283404	-123.117292

Data content and cleaning

- Each dataframe is filtered based on the category field, example shown below

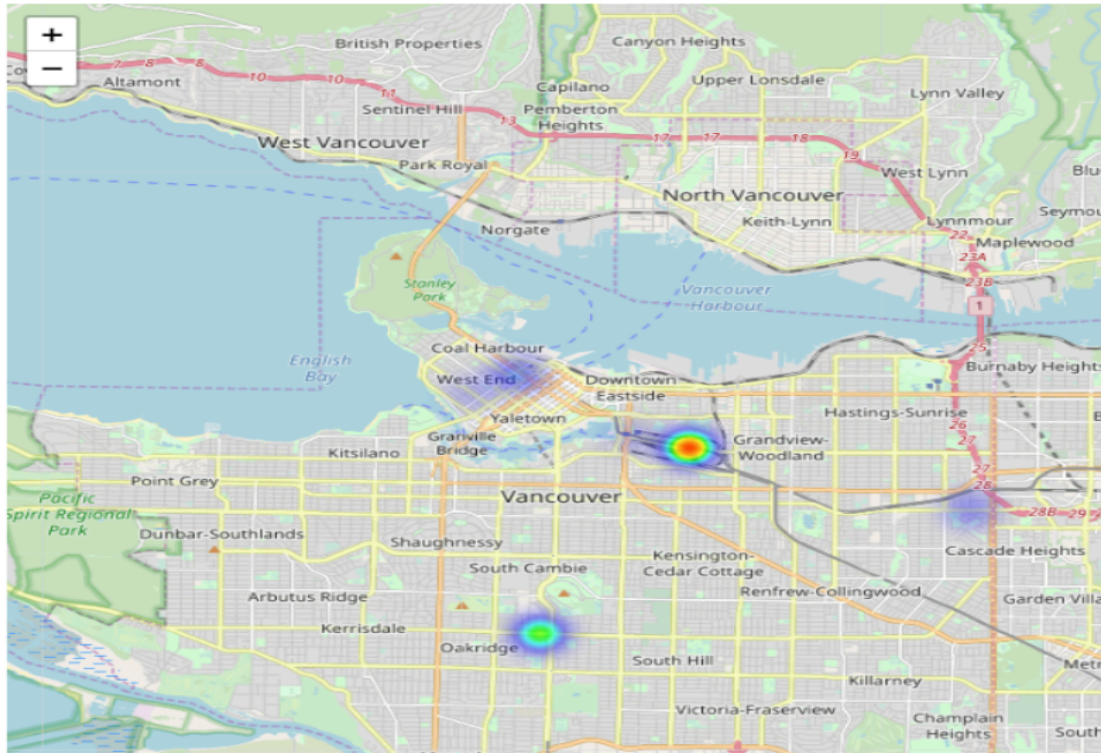
[61]:

	venue.name	venue.categories	venue.location.lat	venue.location.lng
0	CF Pacific Centre	Shopping Mall	49.282980	-123.118361
1	City Square Shopping Centre	Shopping Mall	49.260947	-123.116535
2	Centrepoint Shopping Mall	Shopping Mall	49.259030	-123.101381
3	Harbour Centre Mall	Shopping Mall	49.284562	-123.112296
4	Granville Mall	Shopping Mall	49.281772	-123.120302
5	King Edward Mall	Shopping Mall	49.248629	-123.125933
6	Hudson's Bay	Department Store	49.282483	-123.117514
7	Kingsgate Mall	Shopping Mall	49.262375	-123.097736

Methodology

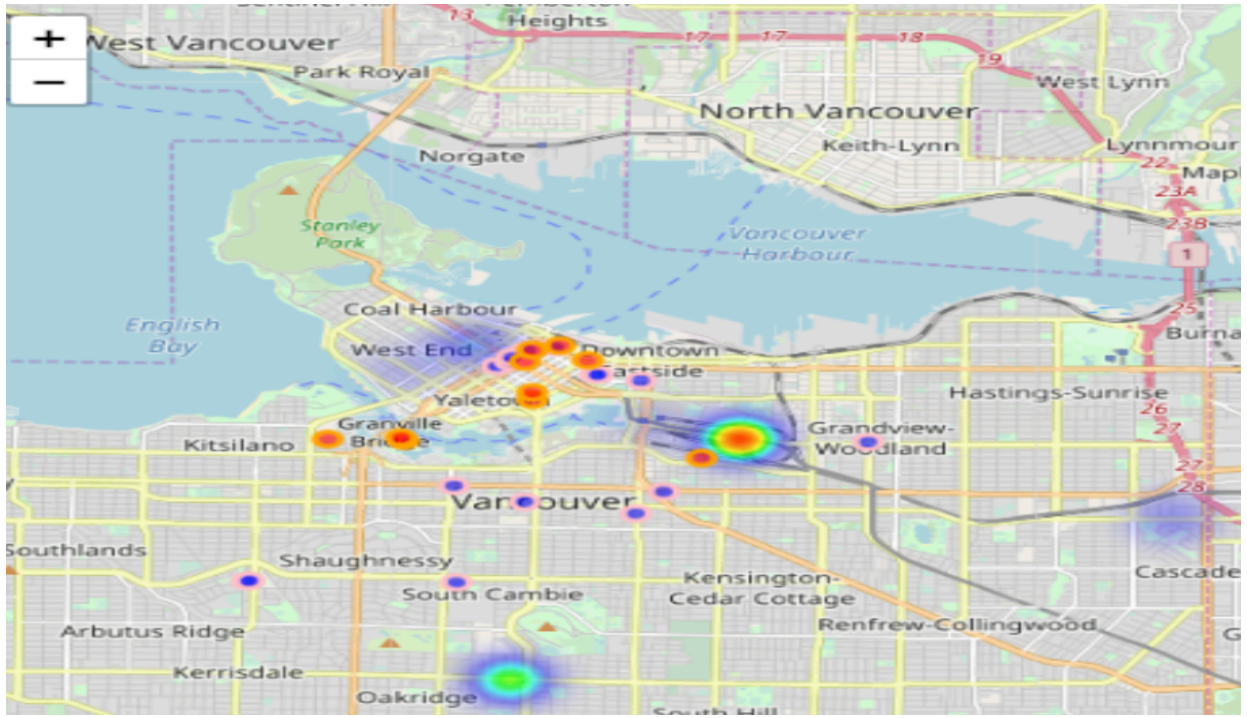
- Data Visualization (Heat Map)

	TYPE	Count
6	Theft from Vehicle	136752
3	Mischief	58421
1	Break and Enter Residential/Other	47441
5	Other Theft	46980
4	Offence Against a Person	46831
0	Break and Enter Commercial	27365
8	Theft of Vehicle	25955
7	Theft of Bicycle	23082
10	Vehicle Collision or Pedestrian Struck (with I...	18230
9	Vehicle Collision or Pedestrian Struck (with F...	208
2	Homicide	180

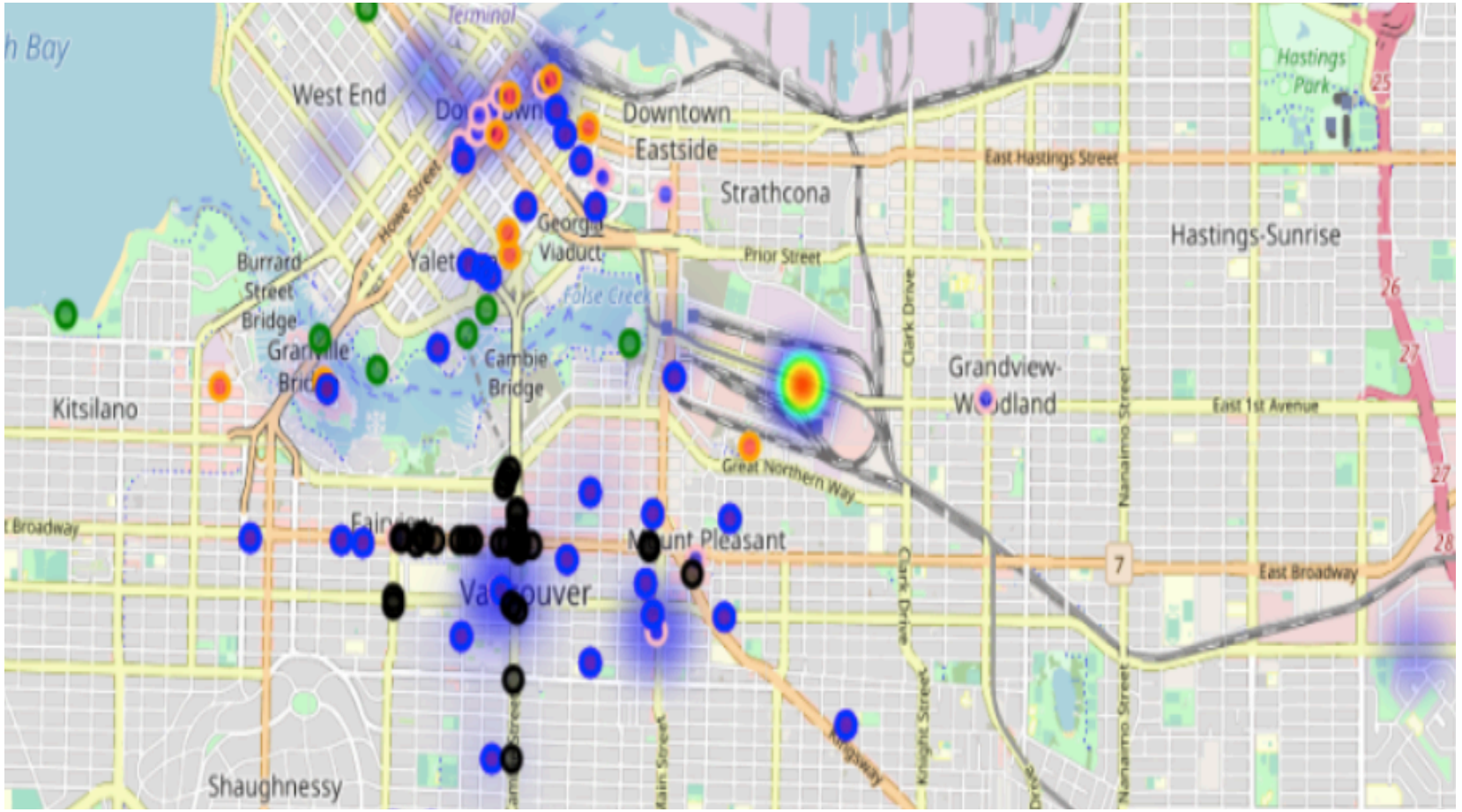


Combined Maps:

- The extracted attractions, Parks, Trails, Universities, Schools and transportation locations were added and combined with the crime occurrence data:



Results (Final Map)



Conclusions

- The use of data science can help homebuyers to find, select and have an initial estimate of a home based on the extracted features, crime data and offered prices.
- Homebuyers would like to select locations that are nearby attraction areas, safe to stay and walk in as well as near to famous restaurants and transportations.
- In this project, we analyzed and clustered Vancouver city based on tourism attractions, crime rates, public transportations and famous restaurants and parks. We also provided recommendations on the top best locations that optimizes homebuyers needs and requirements.