Relationships between Toronto apartment's

rental price and safety issue

Introduction

1.1 Background

Being one of the biggest cities in Canada, Toronto is welcoming a great amount of people from all over the world to visit, work and study, and its population is expected to grown to 3,560,000 by 2031, with an annual average growth of 41,000 [1]. Thus lead Toronto's rental market become quite competitive. For instance, a 1 bedroom average apartment rent is around $1,270 in 2019, which has increased 23% comparing with 2013 [2]. For those who is about to settle down in Toronto for the first time, renting a solid apartment sounds like the first thing to do. Though not knowing much about this city, new lessee would still love to find a safe district under their budget.

1.2 Business Problem

In this project, rental price, regional safety and house rental price will be analyzed for newcomers to find an ideal place efficiently in Toronto. Previous to this capstone project, we have already clustered Toronto's neighborhoods. For further steps, We will explore: Which part of the Toronto has less criminal risks, and what are their expecting rental price? Does higher rental price guarantee to be safer and vice versa? Finally, what are the recommendations regarding to a student, a middle class and a retired man who is trying to find an apartment to rent in Toronto.

1.3 Expecting Purposes

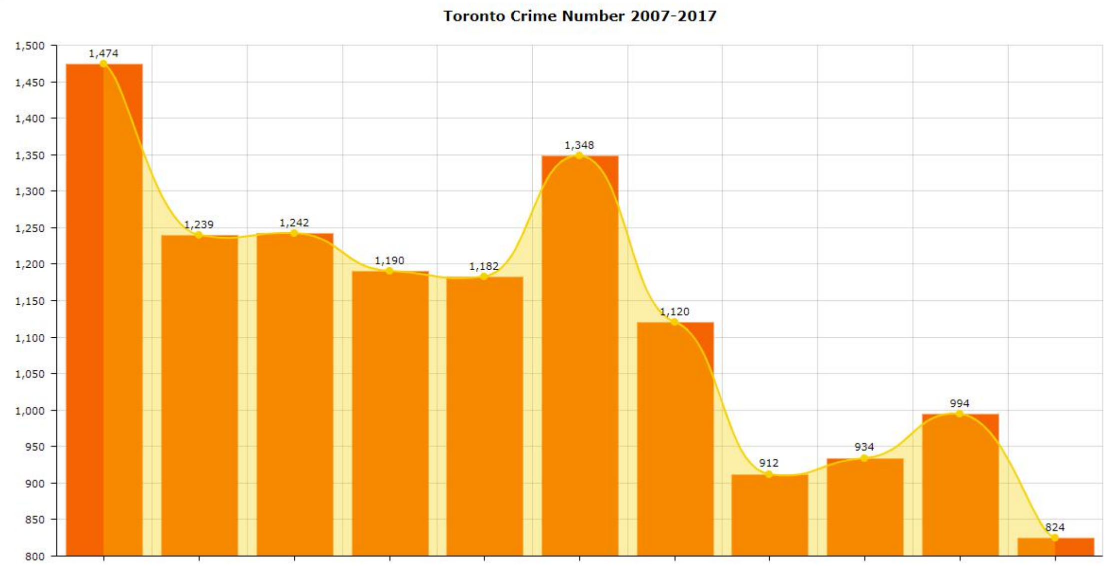
The result of the project is expected to be helpful on multiple purposes: Mostly for people who need to rent an apartment and can have a roughly understanding about Toronto house rental market. Meanwhile, real estate agent can also use the result to have a clearer vision of the advantages and disadvantages of their properties.

Data Description and Pre-processing

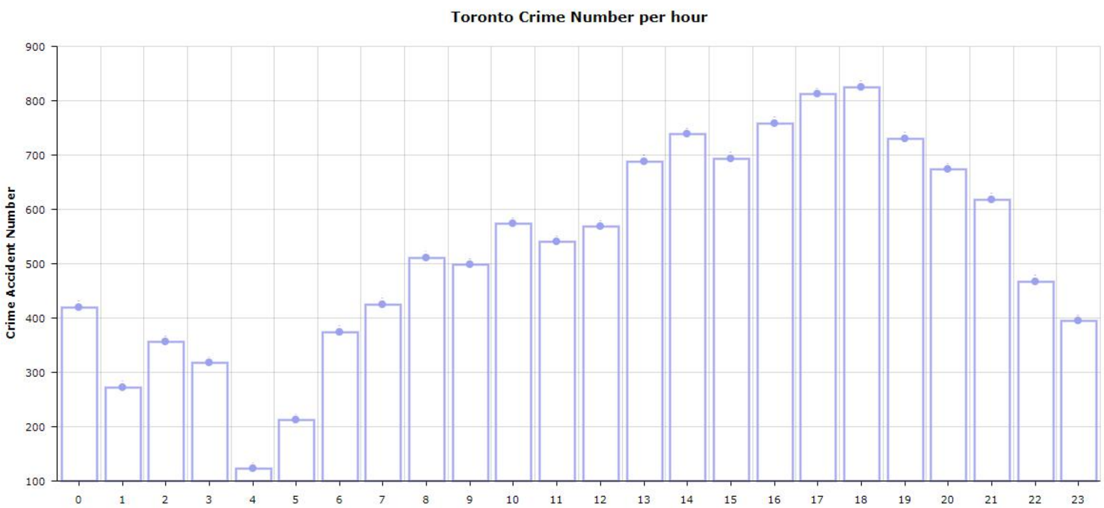
Two datasets are used in this project:
Killed or Seriously Injured (KSI) Toronto Clean:
https://www.kaggle.com/jrmistry/killed-or-seriously-injured-ksi-toronto-clean/kernels
and ***Toronto Apartment Rental Price***:
https://www.kaggle.com/rajacsp/toronto-apartment-price all from kaggle.com.

2.1 KSI Dataset

This dataset has attributes: accident number, accident year/month/day, latitude, longitude, type of the crime, result and the scene description, with length of 12557.



X-axis is year from 2007-2017, Y-axis is numbers of accidents per year. The chart is showing a decreasing trend of crime frequency in Toronto from 2007-2017, which is an overall positive public security sign.

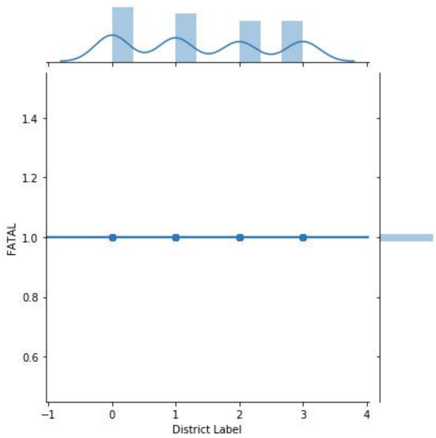


X-axis is 24 hours, Y-axis is numbers of accidents per hour, regardless of year/month.Crime accident peak hours last from 16:00 to 19:00 per day, and gets to lowest point at 4:00 in the morning. This could probably because of daily traffic. Since this project isn't going to study on hourly accidents, this plot is just for dataset visualization.

As for fatal analysis, we transfer district name into numbers from 0-3, we take a look at each district's number, fatal accidents happens within each:

District	District Code	Number	Proportion	Fatal Number	Fatal Ratio
Scarborough	0	2940	0.23	503	0.17
Toronto East York	1	4200	0.33	451	0.11
Etobicoke York	2	3004	0.23	379	0.13
North York	3	2399	0.19	382	0.16

Among four districts Toronto East York has the largest accidents number, while fatal injuries happens most likely in district Scarborough, with 17% possibilities dead. Here's the correlation between district and fatal accidents, the plot proves the observation, that Scarborough has the largest proportion of fatal accidents.



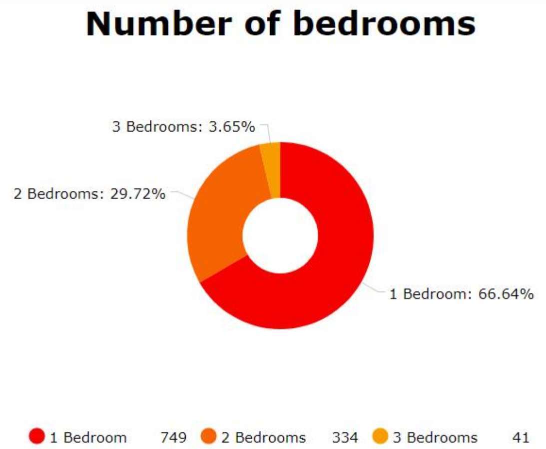
In this dataset, we extract accident number(ACCNUM), Latitude (LATITUDE) and Longitude(LONGITUDE), District(District), District Label(District Label), Fatal (FATAL) attributes for further use.

```
criminal_record.head()
```

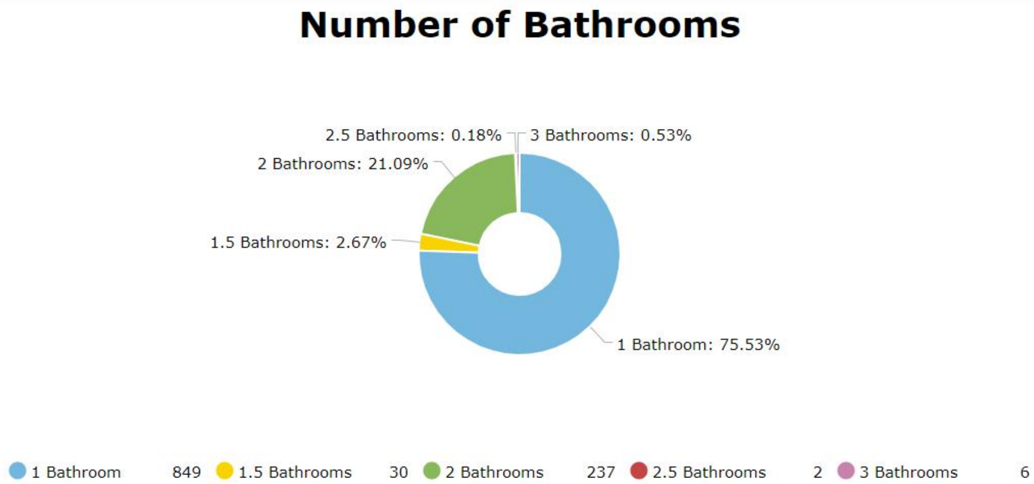
	ACCNUM	Latitude	Longitude	District	FATAL	District Label
0	1249781	43.651545	-79.383490	Toronto East York	0	1
1	1311542	43.780445	-79.300490	Scarborough	0	0
2	5002235651	43.682342	-79.328266	Toronto East York	1	1
3	1311542	43.780445	-79.300490	Scarborough	0	0
4	1311542	43.780445	-79.300490	Scarborough	0	0

2.2 Toronto Apartment Rental

This dataset contains attributes including: Bedroom number, Bathroom number, living room number, address, latitude, longitude, and rental price per month, with length of 1124 rows in total



66.64% of the whole data are 1 bedroom, left with 29.72% 2 bedrooms and only 3.65% 3 bedrooms.



1 Bathroom takes the biggest part of the whole, with proportion of 75.53%. To standardize data, data will be filted with ***1 Bedroom with 1 bathroom and 0 living room, 712 rows.*** Then it was filtered again left with rental price(Price), Latitude(Lat) and Longitude(Long) attributes for further studies.

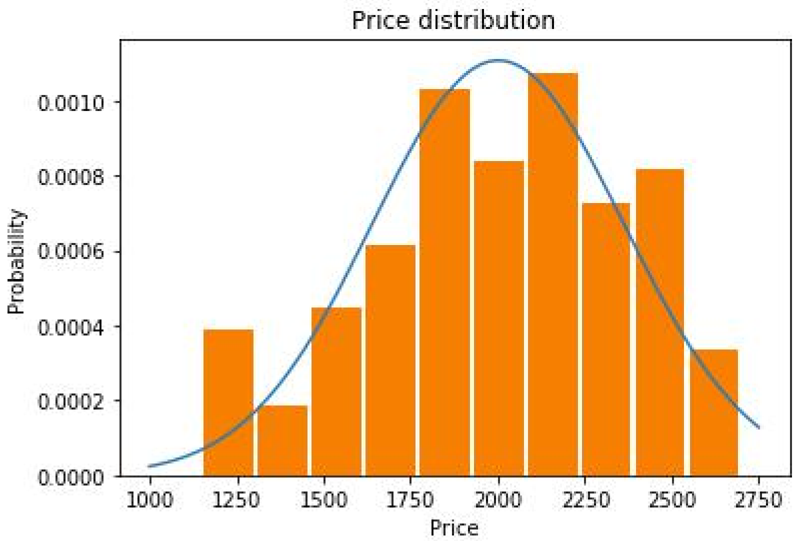
```
house_price = pd.read_csv('Toronto_apartment_rentals_2018.csv')
house_price = house_price.loc[house_price['Bedroom']==1]
house_price = house_price.loc[house_price['Bathroom']==1]
house_price = house_price[['Price','Lat','Long']]
house_price.reset_index(inplace=True)
house_price.set_index('index',inplace=True)
house_price.head()
```

	Price	Lat	Long
index			
1	$2,150.00	43.643051	-79.391643
2	$1,950.00	43.660605	-79.378635
4	$1,800.00	43.652487	-79.389622
5	$1,729.00	43.634890	-79.434654
7	$1,900.00	43.640918	-79.393982

Methodology

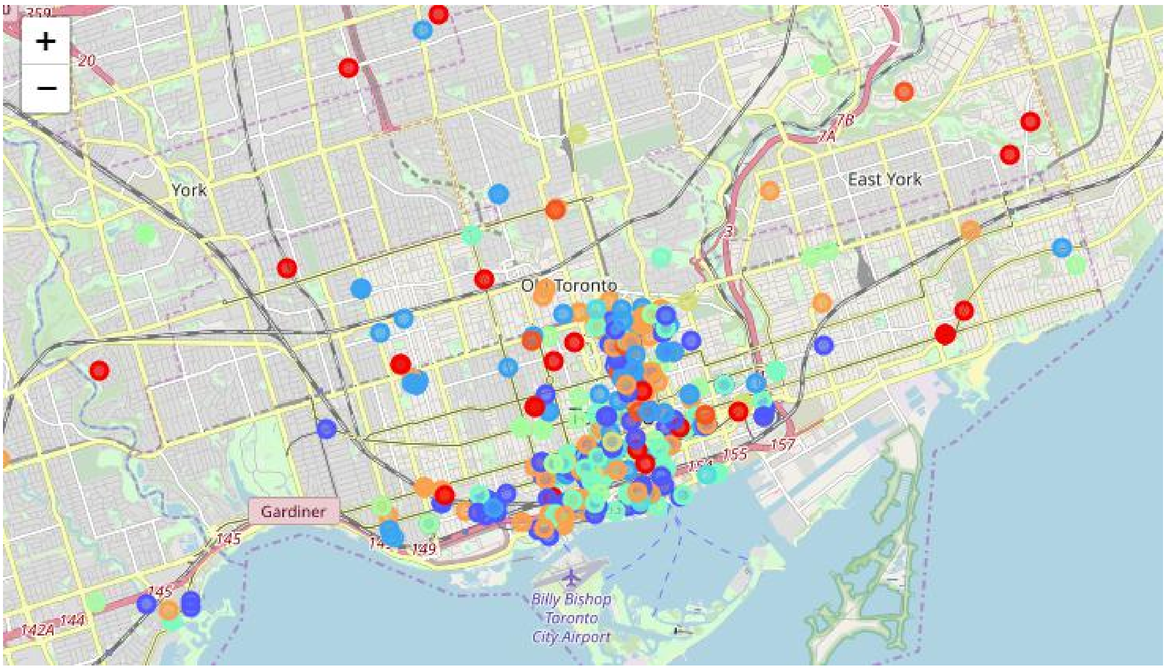
3.1 House Price Dataset Clustering

Follow graph is the distribution of house rental price:

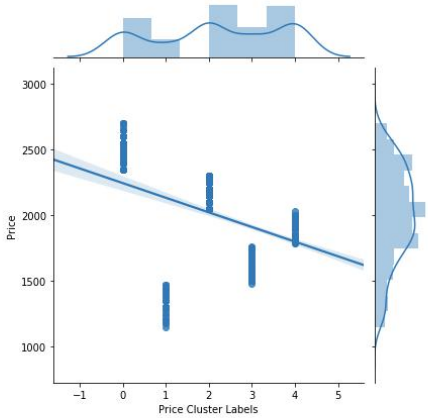


To have a clear vision of house rental price , the dataset is using k-cluster to cluster first, before that while to have a deeper look of these data, 10 clusters are chose and plotted

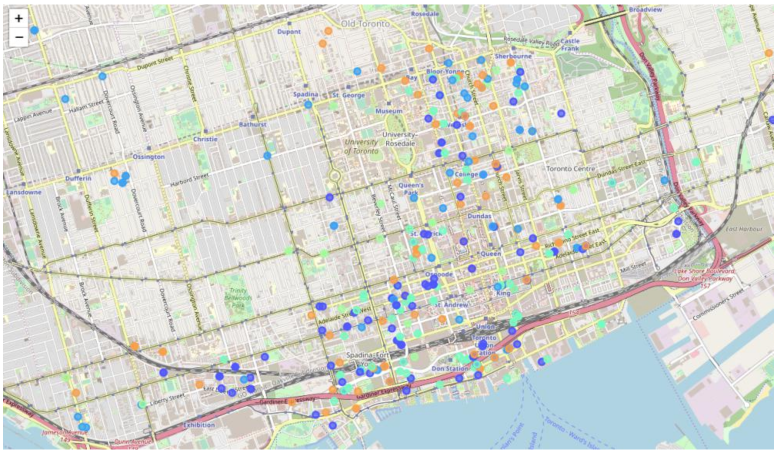
```
# set number of clusters
kclusters = 10

# toronto_house_price = house_price.drop('Price', 1)

# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(house_price)

# check cluster labels generated for each row in the dataframe
kmeans.labels_[0:10]
```



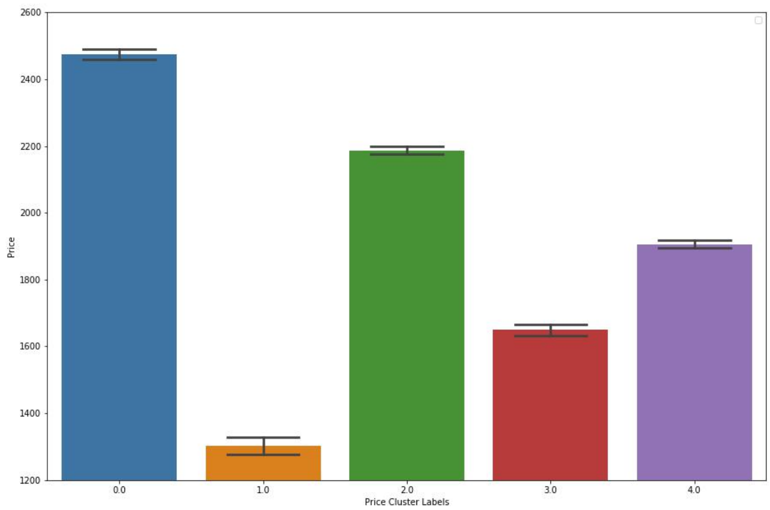
	cluster id	size	average
0	2	172	2187
1	8	171	1905
2	5	129	2474
3	3	101	1649
4	6	59	1302
5	0	37	982
6	7	21	3000
7	9	11	518
8	4	8	3927
9	1	3	535000



As shown in the graph, 10 clusters look chaotic, less organized. The table shows numbers of each cluster and its average house price, for a easier study, cluster[0,7,9,4,1] will be dropped due to small amount of people and over high/low price, which are not referential enough.



Different from KSI dataset, distribution of house price data is not located evenly on the map, while most of them are around downtown, all the way to the University of Toronto. The distribution of clusters are interrelated as well.



It's obvious that cluster 4, 2, 0 are main three price range, average value is around 1900, 2200 and 2500, where most one bedroom with 1 living room is around 2500.

3.2 District Label Classification

The purpose of this project is to explore relationship between location and rental price. To be able to connect district information with rental price, KSI dataset will be classified, to predict district code on house price data. SVM algorithm is chose classify KSI data for it's popular and efficient.

Here's the performance of this model:

```
# Compute confusion matrix
cnf_matrix = confusion_matrix(y_test, yhat, labels=[2,4])
np.set_printoptions(precision=2)

print (classification_report(y_test, yhat))
```

	precision	recall	f1-score	support
0	0.93	1.00	0.96	289
1	0.91	0.98	0.95	414
2	0.95	0.96	0.95	303
3	0.98	0.76	0.85	249
micro avg	0.94	0.94	0.94	1255
macro avg	0.94	0.92	0.93	1255
weighted avg	0.94	0.94	0.93	1255

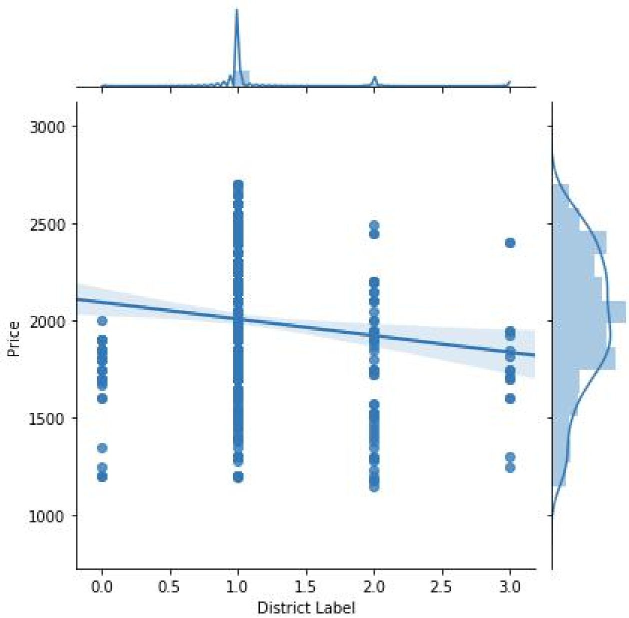
The overall classification result is around 94%, the result table shows that District 4 North York has high precision value and low recall value, means part of District 4 are classified into other classes.

The model is then used to predict on house rental price dataset, to label each row with a district label:

```python
house_geo = house_price[['Latitude', 'Longitude']]
house_price['District Label'] = clf.predict(house_geo)
house_price.sample(100)
```

	Price Cluster Labels	Price	Latitude	Longitude	District Label
405	2	2300	43.667030	-79.376386	1
643	1	1275	43.680128	-79.340131	1
188	4	1950	43.660605	-79.378635	1
575	0	1100	43.675507	-79.415001	1
181	3	1600	43.851647	-79.326656	0
72	2	2100	43.640402	-79.397147	1
619	2	2200	43.639736	-79.390417	1
438	3	1600	43.680602	-79.289333	1
114	3	1700	43.668468	-79.374834	1
480	3	1600	43.647513	-79.392702	1
239	0	2350	43.659329	-79.382675	1
451	0	2450	43.646434	-79.391869	1
568	2	2100	43.623821	-79.478728	2
148	2	2250	43.660846	-79.378756	1
678	0	900	43.741408	-79.816017	2
341	0	2400	43.829739	-79.565678	3

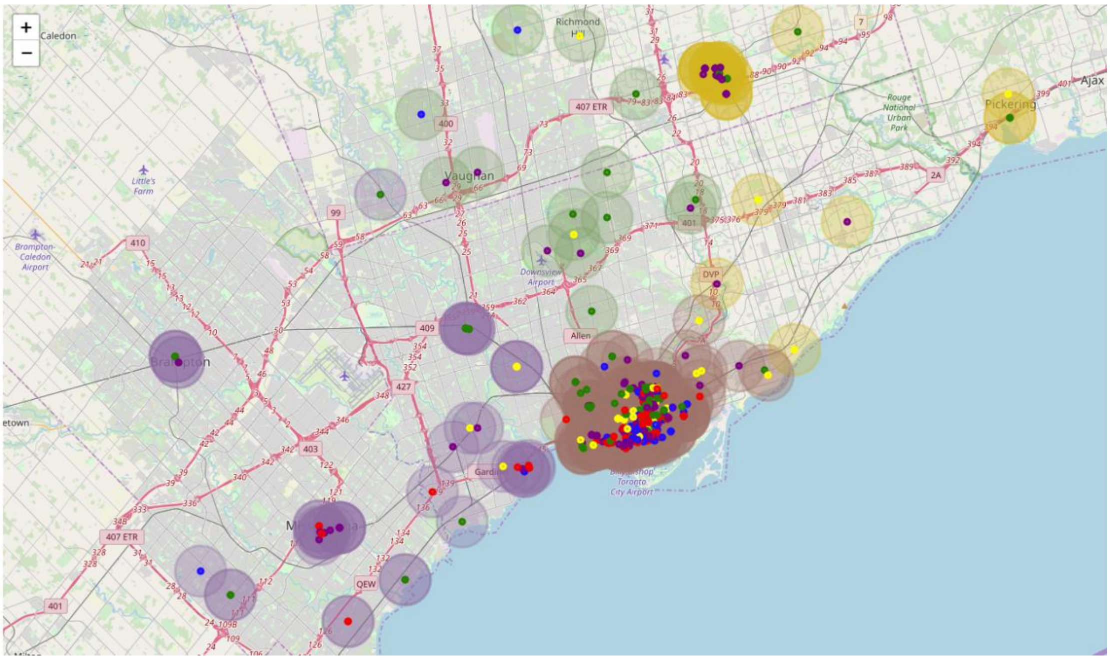
After labeling house rental price data with district, it becomes possible to explore correlations between price and district:



District 0 Scarborough has rental prices around 1300 and 1800-2000, while no more than 2000. Similar to Scarborough, District 3 North York is also around 2000 and below. District 2 Etobicoke York has more house resources, and house price ranges from 1300-2300. District 1 Toronto East York has the largest resources, also contains highest rental price up to 2500 and above.

Results

Following plot present a more intuitive observation, District labels and price labels are drawn on the same map, overlapped, to see how price varies from district to district.



As can see in the map, light color circles are district labels, while color spots are price labels. Though price labels are mixed, we could still find some patterns: blue and yellow labels(expensive ones) are mostly around light brown area, while when away from downtown, green and purple spots(relative cheaper) are spreading all over.

Here’s the safety and price table based on districts

District	District Code	Number	Proportion	atal Numb	Fatal Ratio	Safety	Price
Scarborough	0	2940	0.23	503	0.17	Low	Low
Toronto East York	1	4200	0.33	451	0.11	High	High
Etobicoke York	2	3004	0.23	379	0.13	High	Middle
North York	3	2399	0.19	382	0.16	Low	Low

Discussion

As the table listed above in the last section, it's noticeable that:

1. Lower rental price is always with less safety (see Scarborough and North York)
2. Highest price is expected to be a safer area (Toronto East York)
3. Relative safer district doesn't has to have the highest rental price (Etobicoke York).

Another purpose of this project is to give recommendations to different types of people who is looking for an apartment to rent in Toronto, such as students, people who already work, and retired people. We've made those assumptions based on common understandings:

Type	Income	Physical Condition	Social Activities
Student	Low	Good	Middle
People at work	High	Moderate	High
Retired	Middle	Relative Weak	Low

- For **students**: Considering low income and good physical condition, we recommend students considering ***Scarborough and North York*** district for their cheap rental price.
- For **people at work**: their are having a high demand of social activities, and with abundant income, we recommend ***Toronto East York*** district.
- For **retired people**: safety comes first for they are becoming weak dealing with all sorts of accidents, and they have lower frequency of social activities, ***Etobicoke York*** district is recommended for them.

Conclusion

In this project, we aim to find out relationship between Toronto rental price and safety, and try to give suggestions to different type of people who is trying to rent an apartment in Toronto. To achieve this goal, we use two datasets: KSI dataset that contains district geo information and fatal record, and House Rental Price dataset, which contains house type and rental price, latitude, longitude and lack of district information. To analyse the price data, we clustered use K-Means on House Rental Price dataset into 10 clusters, get rid of outliers, and left 5 biggest ones eventually. We also use SVM classification algorithm on KSI data, with accuracy of 94%, to predict district information on House Rental Price data. Thus we connect two dataset based on 'District'. After analyzing data, we have found that higher price comes with more safety, and vice versa. Recommendations are made based on 3 types of people: students, people at work and retired people. Scarborough and North York has cheap price and high risks, are recommended to students. Toronto East York is recommended to people at work and Etobicoke York is recommended to retired people.

Reference

[1]Toronto polulation report https://www.toronto.ca/legdocs/mmis/2019/ph/bgrd/backgroundfile-124480.pdf

[2]Toronto average rent price https://www.toronto.ca/community-people/community-partners/social-housing-providers/affordable-housing-operators/current-city-of-toronto-average-market-rents-and-utility-allowances/