Report: Using Data Science to Analyze the Neighborhoods of Seattle

Fahim Tahmeed
November 11th, 2020

1. Introduction

1.1 Background

The coronavirus COVID-19 is an infectious disease that has quickly spread to over 200 countries in a matter of weeks, disrupting our daily routines and our lives as we knew them. According to Worldometers website, there has been over 24 million total cases and 882,000 total deaths in the world, as of August 25, 2020. In the United States alone, there has been almost 6 million cases and close to 200,000 deaths. Because the virus is thought to spread mainly from person to person through respiratory droplets (source), many countries, including the United States, have asked their citizens to quarantine themselves since March of 2020 and to limit travel and physical human contact outside of their household.

This has caused many disruptions to our lives, including many workplaces having their employees work from home. As more and more jobs go remote without a deadline in sight, this has given people an option to move further away from the office. We have seen a general trend where people are moving away from expensive big cities to the cheaper suburban areas or even to other states where there is no income tax, such as Washington and Texas (source).

1.2 Problem

I am currently living in Pleasanton, California, which is to the east of San Francisco Bay Area. I enjoy living here because of the proximity to a variety of local parks, international cuisine, and grocery stores. It is also safe to live here because of the low crime rate. However, the cost of living is very high. So, I'd like to move to Seattle, Washington, but I'd like to live in a comparable neighborhood to where I am now.

The aim of the project is to find a safe neighborhood in Seattle, that is surrounded by amenities like the ones in my current location. This will be determined by analyzing crime data per neighborhood, clustering neighborhoods using K-means clustering, and exploring the top common venues in the safest neighborhoods.

1.3 Interest

This exercise may be of interest to anyone also living in an expensive city looking to move to Seattle, Washington. By segmenting and clustering neighborhoods in Seattle, we can determine the most suitable neighborhood we most want to live in.

2. Data acquisition and cleaning

2.1 Data sources

The crime data in Seattle was open source and available for download directly from the Seattle Police Department's website. It contained the crime offense, the crime offense category, the district where the crime took place, and the longitude and latitude of the crime location. A screenshot of the crime data:

	Report Number	Offense ID	Offense Start Date Time	Offense End Date Time	Report Date Time	Group A B	Crime Against Category	Offense Parent Group	Offense	Offense Code	Precinct	Sector	Beat	MCPP	100 Block Address	Longitude	Latitude
0	2020-232722	14749627332	8/5/2020 22:03	8/5/2020 22:03	8/6/2020 1:10	A	PROPERTY	BURGLARY/BREAKING&ENTERING	Burglary/Breaking & Entering	220	E	E	E1	CAPITOL HILL	6XX BLOCK OF BELMONT AVE E	-122.324039	47.624705
1	2020-232740	14746508904	8/5/2020 22:45	8/5/2020 23:30	8/5/2020 23:31	A	PERSON	ASSAULT OFFENSES	Simple Assault	13B	E	C	C3	CENTRAL AREA/SQUIRE PARK	9XX BLOCK OF 25TH AVE	-122.300138	47.610795
2	2020-232748	14746068011	8/5/2020 21:00	8/5/2020 23:00	8/5/2020 23:26	A	PROPERTY	DESTRUCTION/DAMAGE/VANDALISM OF PROPERTY	Destruction/Damage/Vandalism of Property	290	E	E	E2	CAPITOL HILL	15XX BLOCK OF 12TH AVE	-122.316845	47.614684
3	2020-232689	14745070199	8/5/2020 21:10	NaN	8/5/2020 22:59	A	PERSON	ASSAULT OFFENSES	Simple Assault	13B	SW	F	F1	NORTH DELRIDGE	54XX BLOCK OF DELRIDGE WAY SW	-122.362975	47.552820

I also obtained a list of Seattle's districts and neighborhoods by web scraping a Wikipedia page using BeautifulSoup. However, this dataset lacked the geographical coordinates. So, I used geocoder to obtain the latitude and longitude coordinates for each Seattle neighborhood. A screenshot of this dataset:

	Neighborhood	District	Latitude	Longitude
0	Broadview	North Seattle	47.72238	-122.36498
1	Bitter Lake	North Seattle	47.71868	-122.35030
2	North Beach / Blue Ridge	North Seattle	47.70044	-122.38418
3	Crown Hill	North Seattle	47.69520	-122.37410
4	Greenwood	North Seattle	47.69082	-122.35529

To get venue information in each neighborhood, I called the Foursquare API. This gave me a dataset containing the venue name, latitude and longitude coordinates of the venue location, and the venue category, such as the screenshot below:

Venue	Venue Latitude	Venue Longitude	Venue Category
Warren G. Magnuson Park	47.680999	-122.258483	Park
Tennis Center Sand Point	47.681581	-122.260373	Tennis Court
Magnuson Small Dog Area	47.682112	-122.256849	Dog Run
Magnuson Park Off-Leash Dog Park	47.686004	-122.254264	Dog Run
Magnuson Cafe & Brewery	47.688135	-122.264808	New American Restaurant

2.2 Data cleaning

For the Seattle crime dataset, it had nearly 850,000 rows of crime data from 2008 to the present, which was too large to import into Jupyter notebook. So, I manually opened the csv file, filtered by the column "Offense Start Date Time" for only the dates from August 2019 to August 2020, and copied and pasted this filtered data into a new csv file. The dataset now had 76,931 rows and 17 columns. The csv file was imported into a pandas dataframe. I was really only interested in a few of the columns so I created a new dataframe consisting of just the columns I wanted and also renamed the columns to more relevant titles. I also checked if there were any missing values in any of the records by using value_counts and isnull to double check that each column printed out exactly 76,931 records. No missing values were found.

For the neighborhood and district data that was scraped, I cleaned up the dataset by deleting anything that was written in brackets for each item, including the brackets themselves. The list of Seattle

neighborhoods and districts was then put into a pandas dataframe and the column names added. Then, after geocoder was used to get the latitude and longitude coordinates for each neighborhood, I dropped any rows that didn't contain any latitude information and the dataframe index was reset. This list was exported into a csv file for easier future importing.

About the geographical coordinates, geocoder didn't always give me the correct coordinates for a location. So, I created a map to check if all the coordinates were within the Seattle area. If a coordinate seemed to be outside of the Seattle area, I used Google Maps to get approximate coordinates and changed the coordinates manually in the csv file accordingly.

3. Methodology

3.1 Exploratory data analysis

I explored the Seattle crime dataset by first taking a look at the total number of times an offense occurred in Seattle in the past year. Most of the offenses in Seattle are "theft from motor vehicle", "identity theft", and "burglary/ breaking and entering."

Theft From Motor Vehicle	10035
Identity Theft	9442
Burglary/Breaking & Entering	8981
All Other Larceny	6261
Destruction/Damage/Vandalism of Property	6111
Simple Assault	5164
Motor Vehicle Theft	4387
Shoplifting	3881
Trespass of Real Property	3224
Intimidation	3011
Aggravated Assault	2769
Theft From Building	2048
Theft of Motor Vehicle Parts or Accessories	1720
Robbery	1485
Driving Under the Influence	1269
Drug/Narcotic Violations	1217
Credit Card/Automated Teller Machine Fraud	923
False Pretenses/Swindle/Confidence Game	758
Stolen Property Offenses	623
Weapon Law Violations	600
Counterfeiting/Forgery	431
Wire Fraud	402
Rape	291
Fondling	218
Pocket-picking	183
Kidnapping/Abduction	154
Arson	146
Impersonation	128
Extortion/Blackmail	105
Embezzlement	87
Sodomy	82
Prostitution	81
Hacking/Computer Invasion	76
Bad Checks	74
Purchasing Prostitution	69
Family Offenses, Nonviolent	66
Drug Equipment Violations	61
Liquor Law Violations	57
Sexual Assault With An Object	45
Pornography/Obscene Material	44
Murder & Nonnegligent Manslaughter	37
Purse-snatching	37
Curfew/Loitering/Vagrancy Violations	29
Theft From Coin-Operated Machine or Device	27
Animal Cruelty	24
Welfare Fraud	19
Peeping Tom	19
Human Trafficking, Commercial Sex Acts	16
Drunkenness	4
Statutory Rape	3
Justifiable Homicide	2
Assisting or Promoting Prostitution	2
Bribery	1
Human Trafficking, Involuntary Servitude	1
Negligent Manslaughter	1

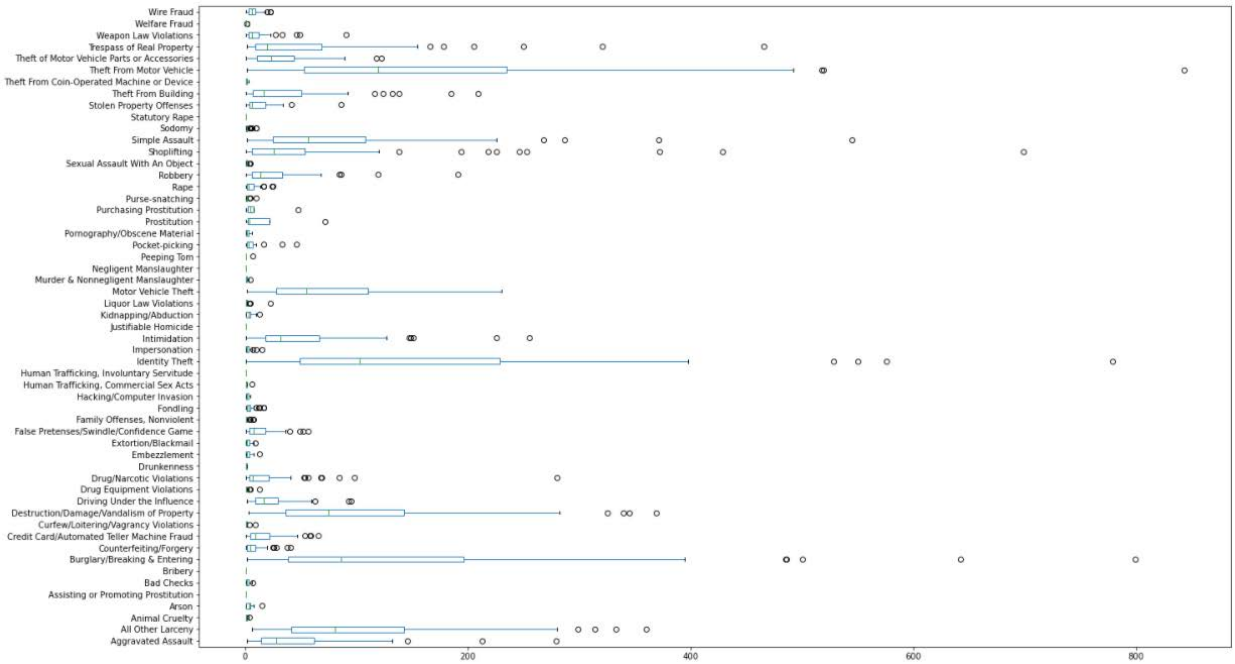
I then grouped the count of each offense by district. With this dataframe, I could then create a pivot table with the columns being the offenses, and rows as the districts. The values in each cell is the count of offense. I also added a “Totals” column at the end of the pivot table that tabulated the total count of offenses for each district. An example of this for the first 5 districts is shown below:

Offense	Aggravated Assault	All Other Larceny	Animal Cruelty	Arson	Assisting or Promoting Prostitution	Bad Checks	Bribery	Burglary/Breaking & Entering	Counterfeiting/Forgery	...	Welfare Fraud	Wire Fraud	Total
District													
ALASKA JUNCTION	35.0	99.0	NaN	NaN	NaN	2.0	NaN	166.0	7.0		NaN	8.0	1252.0
ALKI	9.0	44.0	NaN	NaN	NaN	NaN	NaN	38.0	1.0		1.0	6.0	373.0
BALLARD NORTH	23.0	151.0	NaN	5.0	NaN	4.0	NaN	165.0	9.0		1.0	18.0	1752.0
BALLARD SOUTH	76.0	314.0	NaN	5.0	NaN	4.0	NaN	485.0	26.0		1.0	9.0	3004.0
BELLTOWN	99.0	114.0	NaN	1.0	NaN	1.0	NaN	259.0	9.0		NaN	11.0	1762.0

The descriptive statistics of the crime pivot table could then be found. An example below:

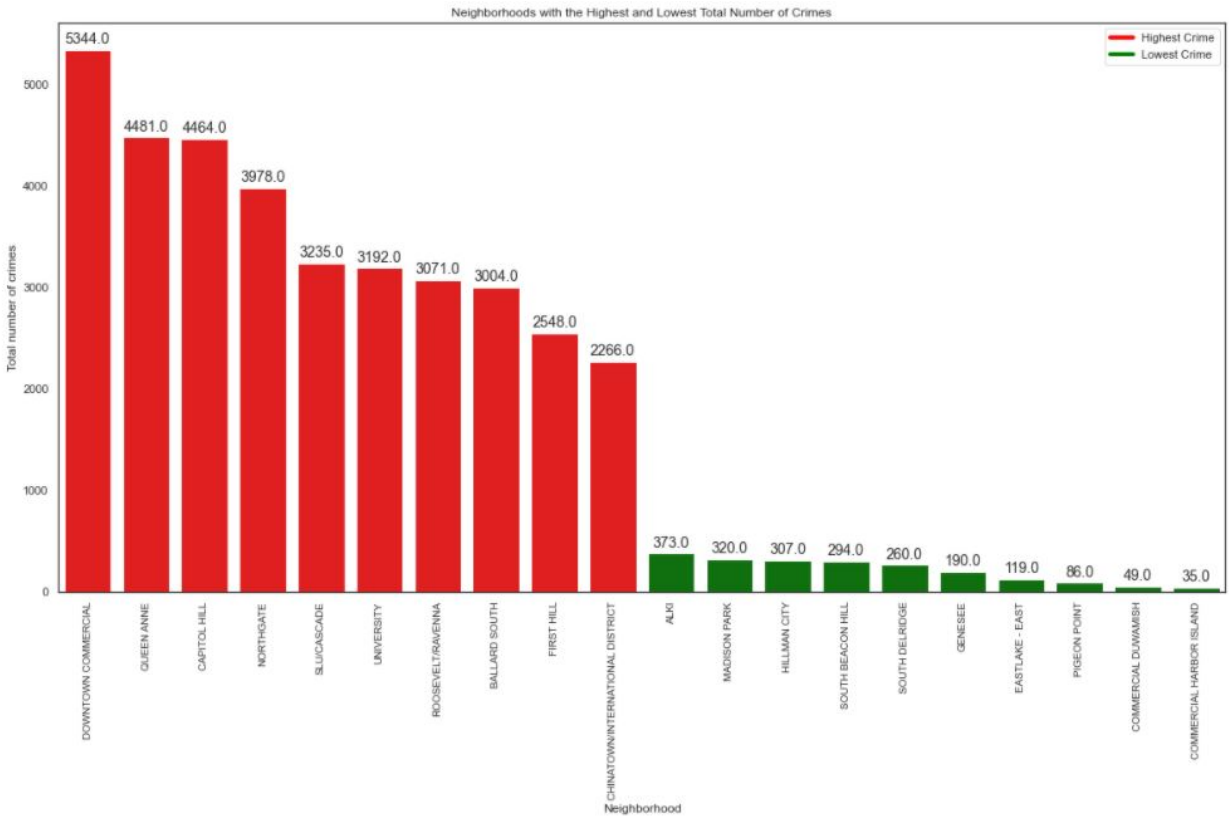
Offense	Aggravated Assault	All Other Larceny	Animal Cruelty	Arson	Assisting or Promoting Prostitution	Bad Checks	Bribery	Burglary/Breaking & Entering	Counterfeiting/Forgery	...	Welfare Fraud	Wire Fraud	Total
count	59.000000	59.000000	15.000000	40.000000	2.0	33.000000	1.0	59.000000	52.000000		17.000000	53.000000	59.000000
mean	46.932203	106.118644	1.600000	3.650000	1.0	2.242424	1.0	152.220339	8.288462		1.117647	7.584906	1303.915254
std	51.503888	89.308278	0.910259	2.787794	0.0	1.714466	NaN	168.898671	9.032266		0.332106	5.641310	1212.340616
min	2.000000	6.000000	1.000000	1.000000	1.0	1.000000	1.0	2.000000	1.000000		1.000000	1.000000	35.000000
25%	14.500000	42.000000	1.000000	1.000000	1.0	1.000000	1.0	38.500000	2.000000		1.000000	3.000000	491.000000
50%	28.000000	81.000000	1.000000	4.000000	1.0	2.000000	1.0	86.000000	5.000000		1.000000	6.000000	894.000000
75%	62.000000	142.500000	2.000000	5.000000	1.0	3.000000	1.0	196.500000	9.250000		1.000000	9.000000	1757.000000
max	279.000000	360.000000	4.000000	15.000000	1.0	7.000000	1.0	799.000000	41.000000		2.000000	23.000000	5344.000000

To better visualize the descriptive statistics for each crime, a boxplot was made:



According to the box plot, there is significant dispersion of motor vehicle theft, identity theft, burglary, and destruction of property, which is to be because some districts have more of these crimes than other districts.

I also found the 10 districts with the highest total number of offenses and lowest total number of offenses. I then concatenated these 2 lists and created a bar chart to better visualize the disparity.



The top 10 districts for total crime count include: Downtown Seattle, Queen Anne, Capitol Hill, Northgate, SLU/Cascade, University, Roosevelt/Ravenna, South Ballard, First Hill, and Chinatown/International District. The bottom 10 districts include: Alki, Madison Park, Hillman City, South Beacon Hill, South Delridge, Genesee, East Eastlake, Pigeon Point, Commercial Duwamish, and Commercial Harbor Island.

3.2 K-means clustering

I now want to see how data mining would perform with a crime dataset. K-means clustering separates a large dataset into distinct subgroups based on a feature. The unsupervised learning algorithm of k-means clustering was chosen in order to group or classify the giant crime dataset to identify districts based on safety.

To prepare the dataset for machine learning, one hot encoding was first used to create a table that turned our crime categorical variables into binary vectors. The dataset was then grouped by district by taking the mean.

I chose to cluster the dataset into 5 groups. With the cluster information, I merged it with the one hot encoding table, as well as with the latitude and longitude coordinates. A color-coded map of the clusters was then created for better visualization.

3.3 Exploring venues in a neighborhood

After a cluster was chosen for further exploration, I explored the venues within the neighborhoods of the cluster by converting the District and Neighborhood name in the Seattle pandas dataframe into all upper case. I then gathered only the rows of the Seattle pandas dataframe in which the Neighborhood name contained a word in the list of districts of the cluster. I also did the same for District names that matched a district in the cluster. Then, I concatenated the 2 dataframes into a new pandas dataframe, dropped the duplicate rows, and reset the index. This gave me a dataframe consisting of the cluster's neighborhoods, districts, and their latitude and longitude coordinates.

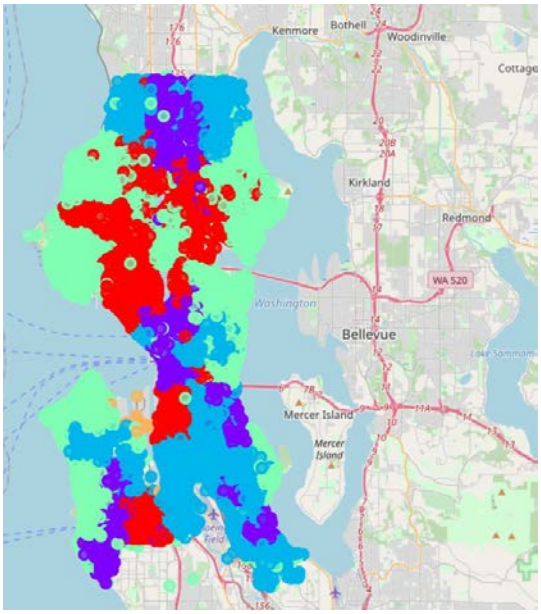
I then defined my Foursquare credentials and API version. The getNearbyVenues function uses the Foursquare API to get the top 100 venues within a radius of 1609 meters (or approximately 1 miles) from a given location. I called the getNearbyVenues function on each neighborhood's latitude and longitude coordinates to get a list of the top 100 venues per neighborhood of the cluster.

With this list, I once again used one hot encoding in order to groups the rows by neighborhood and take the mean of the frequency of the occurrence of each venue category. We can now print the top 5 most common venue categories per neighborhood with this frequency table. In order to make this easier to read, I put this into a pandas dataframe in descending order by venue category. To do this, I defined a function called return_most_common_venues that took each neighborhood row and the number of venues and returned the venues for that neighborhood row sorted in descending order.

4. Results

4.1 Results from k-means clustering

A screen shot of the color-coded map of the clusters can be seen below. It is also available for viewing in this directory as an html file. The red dots are Cluster 0, the purple dots are Cluster 1, the blue dots are Cluster 2, the green dots are Cluster 3, and orange dots are Cluster 4.



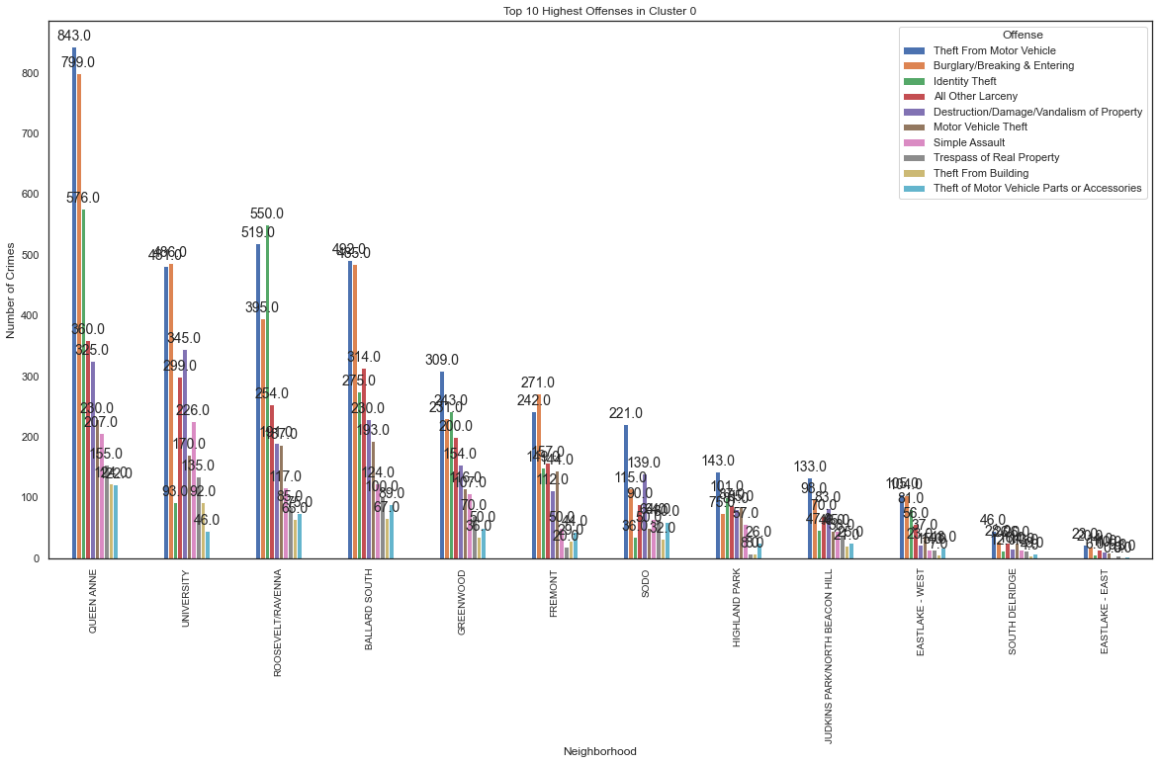
In the table below, you can see the number of crime records in each cluster.

```
Cluster
0    20853
1    24579
2    17749
3    13666
4       84
Name: District, dtype: int64
```

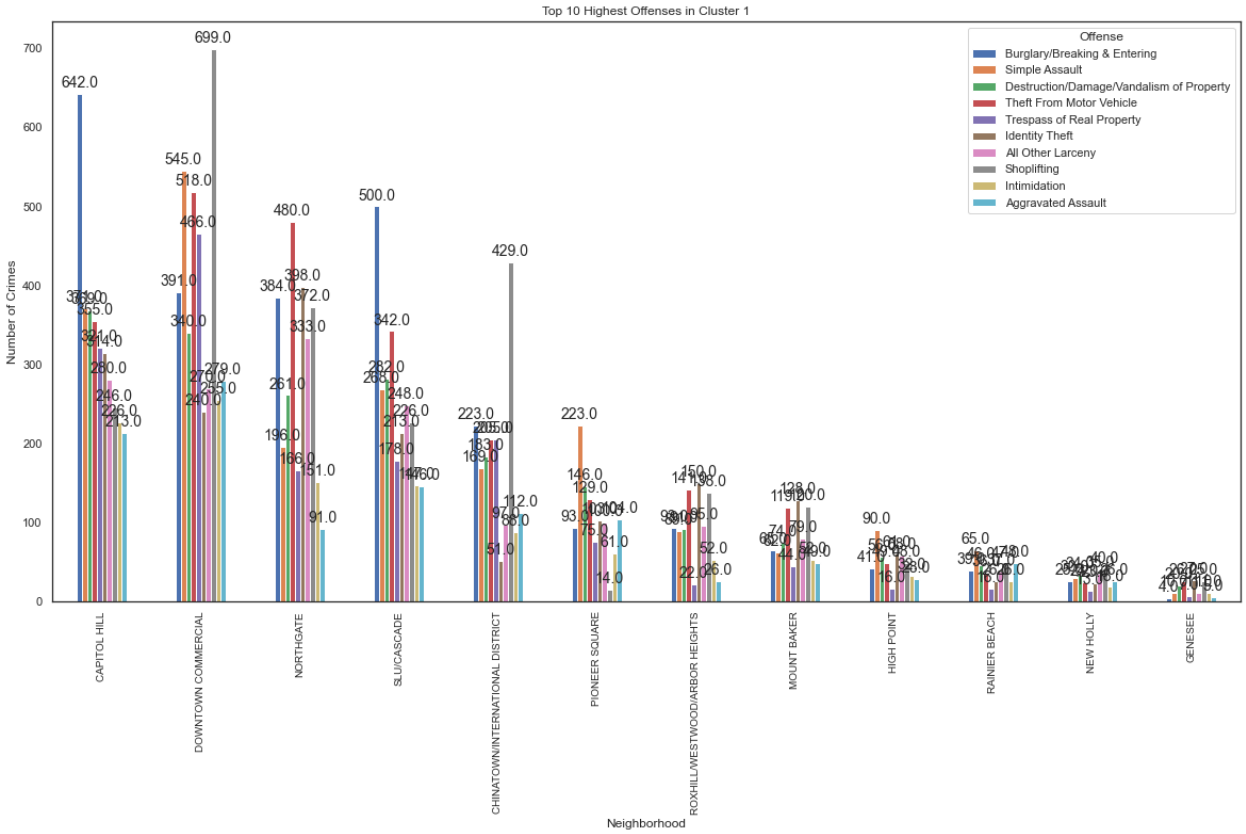
I also grouped the table to find the districts within each cluster and the number of offenses tabulated per district.

Cluster	District	
0	BALLARD SOUTH	3004
	EASTLAKE - EAST	119
	EASTLAKE - WEST	507
	FREMONT	1412
	GREENWOOD	1987
	HIGHLAND PARK	809
	JUDKINS PARK/NORTH BEACON HILL	768
	QUEEN ANNE	4481
	ROOSEVELT/RAVENNA	3071
	SODO	1243
	SOUTH DELRIDGE	260
	UNIVERSITY	3192
1	CAPITOL HILL	4464
	CHINATOWN/INTERNATIONAL DISTRICT	2266
	DOWNTOWN COMMERCIAL	5344
	GENESEE	190
	HIGH POINT	629
	MOUNT BAKER	1019
	NEW HOLLY	396
	NORTHGATE	3978
	PIONEER SQUARE	1328
	RAINIER BEACH	565
	ROXHILL/WESTWOOD/ARBOR HEIGHTS	1165
	SLU/CASCADE	3235
2	ALASKA JUNCTION	1252
	BELLTOWN	1762
	BITTERLAKE	1457
	BRIGHTON/DUNLAP	894
	CENTRAL AREA/SQUIRE PARK	1869
	CLAREMONT/RAINIER VISTA	373
	COLUMBIA CITY	475
	FIRST HILL	2548
	GEORGETOWN	1003
	HILLMAN CITY	307
	LAKECITY	1930
	MID BEACON HILL	631
	NORTH BEACON HILL	1313
	NORTH DELRIDGE	529
	RAINIER VIEW	528
	SOUTH BEACON HILL	294
	SOUTH PARK	584
3	ALKI	373
	BALLARD NORTH	1752
	FAUNTLEROY SW	395
	LAKEWOOD/SEWARD PARK	374
	MADISON PARK	320
	MADRONA/LESCHI	913
	MAGNOLIA	1489
	MILLER PARK	562
	MONTLAKE/PORTAGE BAY	532
	MORGAN	791
	NORTH ADMIRAL	792
	PHINNEY RIDGE	741
	PIGEON POINT	86
	SANDPOINT	2261
	UNKNOWN	900
	WALLINGFORD	1385
4	COMMERCIAL DUWAMISH	49
	COMMERCIAL HARBOR ISLAND	35

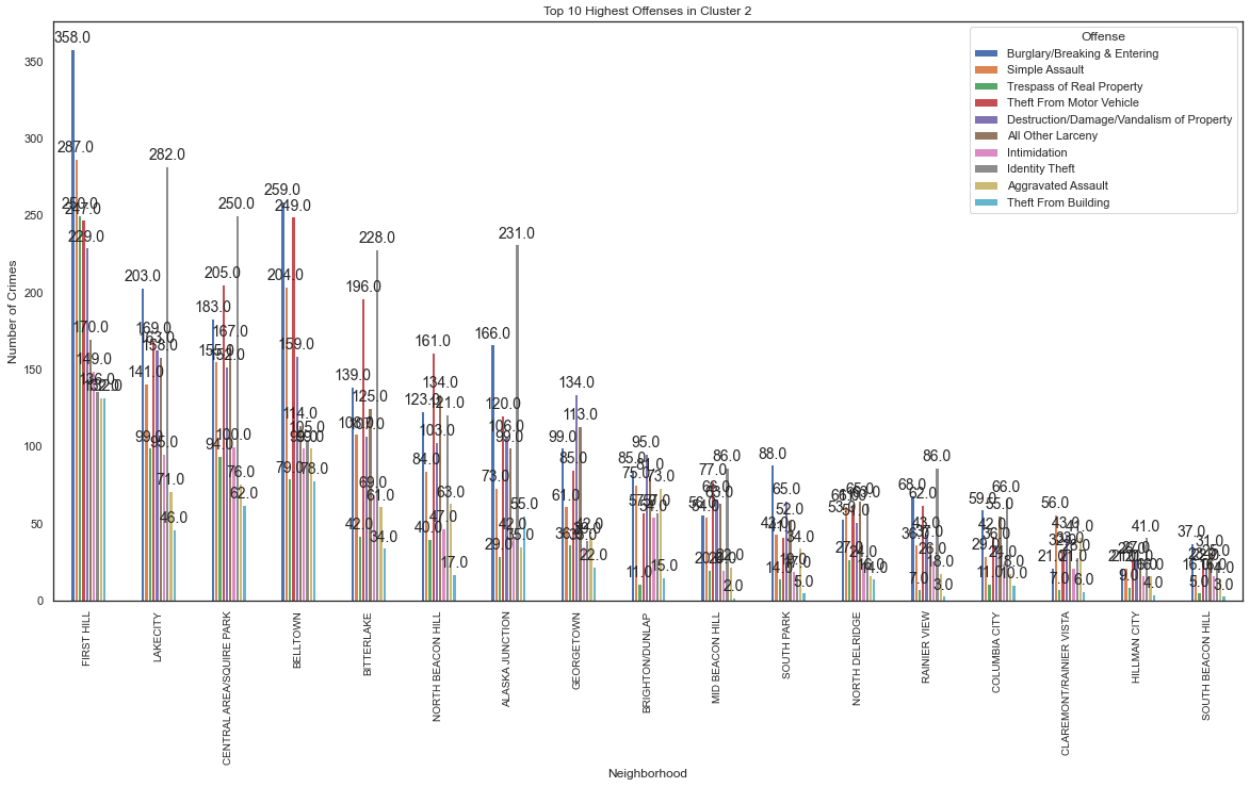
For Cluster 0, the following bar chart shows the top 10 crimes in each of the districts. It's easy to see that Cluster 0 contains mostly motor vehicle theft and burglary.



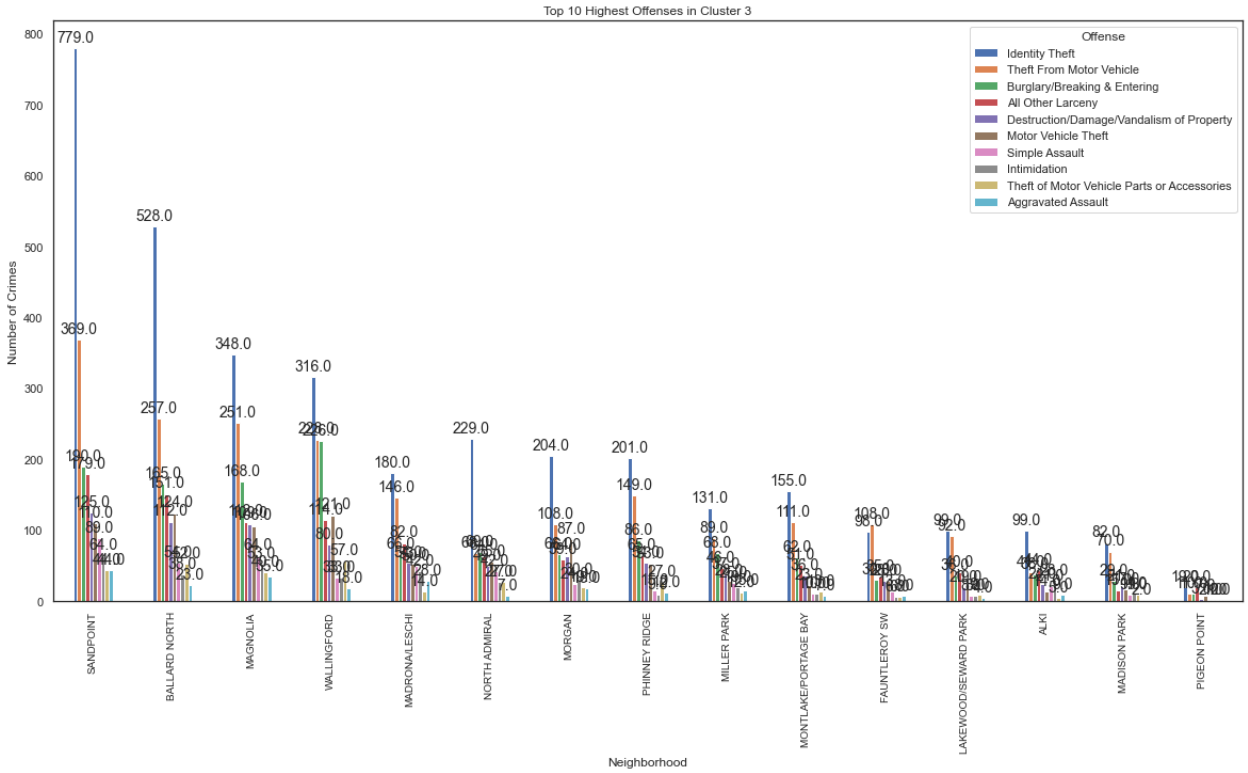
For Cluster 1, the following bar chart shows the top 10 crimes in each of the districts. Cluster 1 contains mostly burglary, shoplifting, simple assault, and motor vehicle theft.



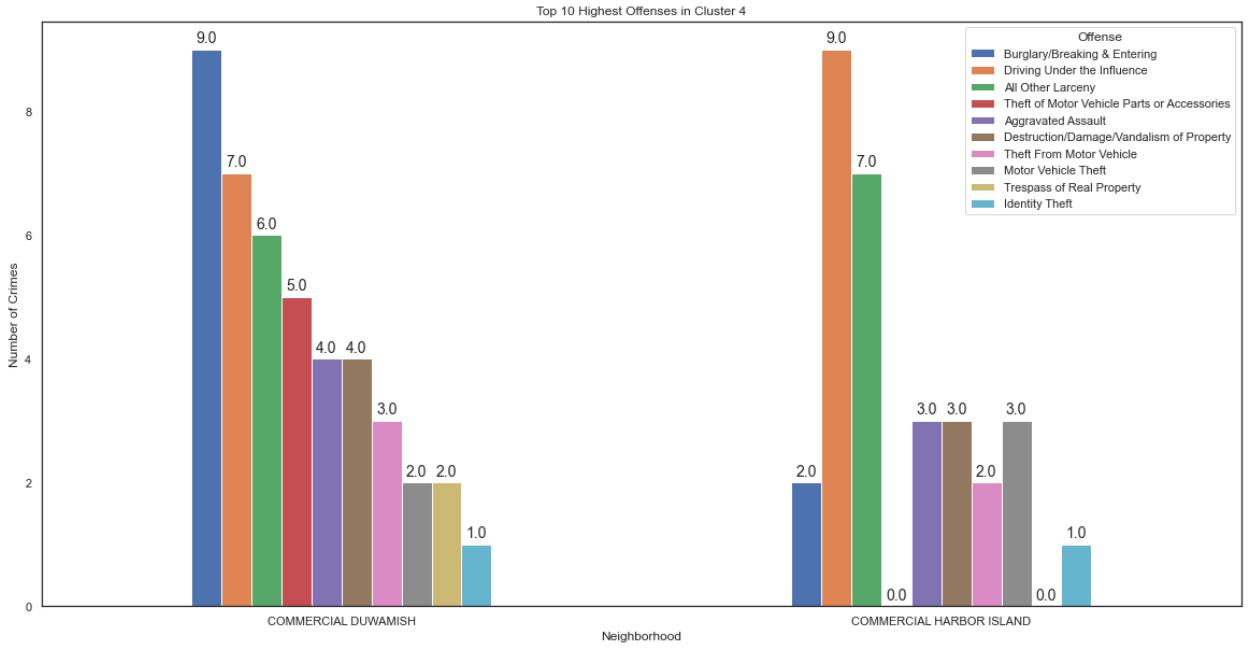
For Cluster 2, the following bar chart shows the top 10 crimes in each of the districts. Cluster 2 contains mostly burglary, identity theft, and motor vehicle theft.



For Cluster 3, the following bar chart shows the top 10 crimes in each of the districts. Cluster 3 contains mostly identity theft and motor vehicle theft.



For Cluster 4, the following bar chart shows the top 10 crimes in each of the districts. Cluster 4 contains seems to contain only the commercial areas where not a lot of crime happens.



Since Clusters 0, 1, and 2 contain districts that are in the list of top 10 districts with the most total number of crimes, and Cluster 4 only contains commercial districts (not districts for living in), this narrows down our search for a suitable, safe place to live to just Cluster 3.

4.2 Results from exploring venues in a neighborhood

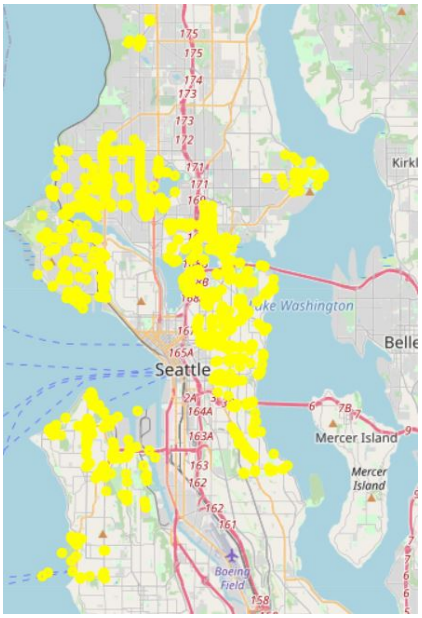
After using the Foursquare API to gather the top 100 venues within a radius of about 1 mile from each neighborhood in Cluster 3, I got a result of 1519 venues. The resulting dataframe consisted of the following columns: neighborhood name, neighborhood latitude, neighborhood longitude, venue name, venue latitude, venue longitude, and venue category. An example of this is shown below:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	SANDPOINT	47.68212	-122.26081	Warren G. Magnuson Park	47.680999	-122.258483	Park
1	SANDPOINT	47.68212	-122.26081	Tennis Center Sand Point	47.681581	-122.260373	Tennis Court
2	SANDPOINT	47.68212	-122.26081	Magnuson Small Dog Area	47.682112	-122.256849	Dog Run
3	SANDPOINT	47.68212	-122.26081	Magnuson Park Off-Leash Dog Park	47.686004	-122.254264	Dog Run
4	SANDPOINT	47.68212	-122.26081	Magnuson Cafe & Brewery	47.688135	-122.264808	New American Restaurant

Since it seemed like not all neighborhoods returned 100 venues, I wanted to group the dataframe by neighborhood to get a count of the number of venues found per neighborhood.

Neighborhood	Neighborhood Latitude
ALKI POINT	88
BRIARCLIFF	37
BROADMOOR	81
FAUNTLEROY	30
LAKEWOOD	45
LAWTON PARK	91
LESCHI	100
LOYAL HEIGHTS	100
MADRONA VALLEY	15
MILLER PARK	100
MONTLAKE	68
NORTH ADMIRAL	83
PHINNEY RIDGE	98
PIGEON POINT	56
PORTAGE BAY	100
SANDPOINT	34
SOUTHEAST MAGNOLIA	69
SUNSET HILL	100
WALLINGFORD	100
WHITTIER HEIGHTS	100

There were 229 unique venue categories returned and this list was printed out to verify. A map was created to visualize where all these venues are (marked in yellow).



After analyzing each neighborhood in Cluster 3 using the venues returned, I printed out a table of the top 10 most common venue categories for each neighborhood.

	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	ALKI POINT	Coffee Shop	Pizza Place	Scenic Lookout	Park	Bar	Pier	Pub	Bus Station	Café	Convenience Store
1	BRIARCLIFF	Park	Playground	Trail	Coffee Shop	Bank	Pharmacy	Pizza Place	Pub	Sandwich Place	Scenic Lookout
2	BROADMOOR	Park	Garden	Bus Stop	Bar	Bakery	Italian Restaurant	Trail	French Restaurant	Beach	Café
3	FAUNTLEROY	Park	Pool	Boat or Ferry	Convenience Store	Market	Cupcake Shop	Coffee Shop	Taco Place	Food Truck	Beach
4	LAKEWOOD	Pizza Place	Park	Vietnamese Restaurant	Video Store	Lake	Coffee Shop	ATM	Dog Run	Soccer Field	Café
5	LAWTON PARK	Park	Coffee Shop	Bar	Cocktail Bar	Trail	Mexican Restaurant	Pizza Place	Brewery	New American Restaurant	Clothing Store
6	LESCHI	Coffee Shop	Ethiopian Restaurant	Park	Playground	Bar	Bakery	Pizza Place	Grocery Store	Thai Restaurant	BBQ Joint
7	LOYAL HEIGHTS	Coffee Shop	Pizza Place	Bakery	Park	Thai Restaurant	Ice Cream Shop	Food Truck	Burger Joint	Bar	Italian Restaurant
8	MADRONA VALLEY	Chinese Restaurant	Trail	Coffee Shop	Hobby Shop	Beer Store	Dumpling Restaurant	Shopping Mall	Bank	Sushi Restaurant	Restaurant
9	MILLER PARK	Coffee Shop	Italian Restaurant	Bakery	Cocktail Bar	Café	Yoga Studio	Taco Place	Greek Restaurant	Sushi Restaurant	Thai Restaurant
10	MONTLAKE	Trail	Garden	Park	Coffee Shop	Bus Stop	Lake	Scenic Lookout	Gym	Playground	Bus Station
11	NORTH ADMIRAL	Park	Coffee Shop	Thai Restaurant	Ice Cream Shop	American Restaurant	Pizza Place	Scenic Lookout	Beach	Clothing Store	Convenience Store
12	PHINNEY RIDGE	Zoo Exhibit	Pizza Place	Ice Cream Shop	Coffee Shop	Pub	Café	Bar	Automotive Shop	Food Truck	New American Restaurant
13	PIGEON POINT	Coffee Shop	Park	BBQ Joint	Gas Station	Gym	Harbor / Marina	Pizza Place	Garden	Food Truck	Soccer Field
14	PORTAGE BAY	Coffee Shop	Park	Café	Mexican Restaurant	Italian Restaurant	Seafood Restaurant	Bar	Deli / Bodega	Bus Station	Trail
15	SANDPOINT	Park	Sculpture Garden	Dog Run	Tennis Court	Theater	Gym	Beach	Pizza Place	Rugby Pitch	New American Restaurant
16	SOUTHEAST MAGNOLIA	Bus Stop	Harbor / Marina	Pizza Place	Park	Sandwich Place	Playground	Trail	Coffee Shop	Video Store	Scenic Lookout
17	SUNSET HILL	Bar	Coffee Shop	Mexican Restaurant	Cocktail Bar	Park	Bakery	Ice Cream Shop	Italian Restaurant	Burger Joint	Farmers Market
18	WALLINGFORD	Coffee Shop	Café	Ice Cream Shop	Park	Mexican Restaurant	Seafood Restaurant	Korean Restaurant	Bar	Thai Restaurant	Bubble Tea Shop
19	WHITTIER HEIGHTS	Coffee Shop	Pizza Place	Mexican Restaurant	Bakery	Ice Cream Shop	Pub	French Restaurant	Grocery Store	Café	Breakfast Spot

Most of the first or second most common venue categories of Cluster 3 are coffee shops and parks, leading me to believe that Cluster 3 contains neighborhoods that are in the suburbs of Seattle. As a reminder, I am looking for a neighborhood similar to where I currently live, which is a safe, suburban town with lots of international cuisine. Madrona Valley has Chinese restaurants as its first most common venue category and dumpling restaurants as its sixth most common venue category, which makes it my top place to live in Seattle, because I love Chinese food. Another place I'd be interested in looking further into is Wallingford for its variety of international cuisines overall (and has bubble tea shops as its tenth most common venue category).

5. Discussion

Based on the results, I have concluded that the neighborhoods in Cluster 3 are among the safest in Seattle because they don't contain any neighborhoods from the list of top 10 highest crime neighborhoods. By investigating the venues in the neighborhoods of Cluster 3, I have further narrowed down my search for a suitable place to Broadmoor, because of its proximity to a park and italian restaurants.

An observation I noticed was that the districts in Seattle weren't well-defined, therefore the number of districts in the crime dataset didn't match the number of districts obtained from web scraping.

The venues that are given are also dependent on when the Foursquare API was called. So, doing this exercise again will yield different venue results.

6. Conclusion

In conclusion, through web scraping, data cleaning, exploratory analysis, and k-means clustering, and using the Foursquare API, I have found that the safest and most comparable place to live in Seattle, Washington is Broadmoor.