

El uso de la ciencia de los datos para analizar los Barrios de Seattle

Diego Silvera
20 de Noviembre 2022

1. Introducción

1.1 Antecedentes

En los últimos dos años Estados Unidos ha sufrido un cambio en la política estatal, y esto ha generado fuertes cambios en las diferentes áreas del gobierno una de ellas la económica. Se han tenido que tomar medidas socioeconómicas que generan cambios en las diferentes industrias, un aumento en el costo de vida promedio de la sociedad incrementando la brecha que existe entre las diferentes clases sociales. Esta brecha puede haber estimulado o afectado el nivel de delincuencia en los últimos tiempos haciendo que la sociedad se vea afectada por mayor número de delitos. Se entiende que pueden ser efectos de la pandemia que acabamos de pasar muchos países, incluido Estados Unidos, solicitaron cuarentena, limitando el tránsito impactando en todo lo que es el área comercial directamente al rubro gastronómico teniendo que en muchos casos cierres de locales. La tendencia general es que la gente se aleje de las grandes y caras ciudades y se traslade a las más baratas o incluso a otros estados donde no hay impuesto sobre la renta, como Washington y Texas (fuente).

1.2 Problema

California, que está al este de la bahía de San Francisco es una de las ciudades a mi entender cuenta con un costo de vida promedio elevado pero con un nivel de seguridad alto para el promedio.

Por lo tanto, me gustaría me gustaría mudarme a Washington específicamente en Seattle, pero me gustaría vivir en un barrio con características similares a California.

El objetivo del proyecto es encontrar un barrio seguro, que esté rodeado de servicios como los que ofrece California. Esto se determinará

analizando los datos de delincuencia por barrios, agrupación de los barrios utilizando K-means clustering, y la exploración de la parte superior lugares comunes en el más seguro más seguros.

1.3 Interés

Este ejercicio puede ser de interés para cualquier persona que también viva en una ciudad cara y quiera mudarse a, Washington con características similares a Seattle. Al segmentar y agrupar los barrios de Seattle, podemos determinar el barrio más adecuado en el que queremos vivir.

2. Adquisición y limpieza de datos

2.1 Fuentes de datos

Los datos sobre la delincuencia en Seattle eran de código abierto y estaban disponibles para su descarga directamente desde el sitio web del Departamento de Policía de Seattle que solo contaban con la base hasta el año 2020.

Departamento de Policía de Seattle. Contenían como variables a tomar en cuenta son, el delito, la categoría del delito, el distrito en el que tuvo lugar el delito y la longitud y latitud del lugar del delito.

El delito se produjo, y la longitud y latitud del lugar donde se cometió, esto nos permite ubicar el mismo en zona. Una captura de pantalla de los datos sobre delitos:

	Report Number	Offense ID	Offense Start DateTime	Offense End DateTime	Report DateTime	Group A B	Crime Against Category	Offense Parent Group	Offense	Offense Code	Pr
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	
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	
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	
3	2020-232689	14745070199	8/5/2020 21:10	NaN	8/5/2020 22:59	A	PERSON	ASSAULT OFFENSES	Simple Assault	13B	
4	2020-232654	14742137960	8/4/2020 0:00	8/4/2020 23:59	8/5/2020 22:18	A	PROPERTY	LARCENY-THEFT	Theft of Motor Vehicle Parts or Accessories	23G	

También obtuve una lista de los distritos y vecindarios de Seattle mediante el raspado de una página de Wikipedia utilizando BeautifulSoup.

Sin embargo, este conjunto de datos carecía de coordenadas geográficas. Así que utilicé el geocodificador para obtener las coordenadas de latitud y longitud de cada barrio de Seattle. Una captura de pantalla de este conjunto de datos:

	Barrio	Distrito	Latitud	Longitud
0	North Seattle	Seattle	47.643724	-122.302965
1	Broadview	North Seattle	47.722380	-122.364980
2	Bitter Lake	North Seattle	47.718680	-122.350300
3	North Beach / Blue Ridge	North Seattle	47.700440	-122.384180
4	Crown Hill	North Seattle	47.695200	-122.374100

Para obtener información sobre los locales de cada barrio, llamé a la API de Foursquare.

Esto me dio un conjunto de datos que contiene el nombre del local, las coordenadas de latitud y longitud de la ubicación del local y la categoría del mismo, se muestra a continuación:

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 Limpieza de datos

El conjunto de datos sobre la delincuencia en Seattle tenía casi 850.000 filas de datos sobre la delincuencia desde 2008 hasta la actualidad, lo que era demasiado grande para importarlo al cuaderno Jupyter. Así que abrí manualmente el archivo csv, filtrado por la columna

"Offense Start Date Time" para sólo las fechas de agosto de 2019 a agosto de 2020, y copié y pegué estos datos filtrados en un nuevo archivo csv. El conjunto de datos tenía ahora 76.931 filas y 17 columnas. El archivo csv fue importado en un marco de datos de pandas. En realidad, sólo me interesaban algunas de las columnas, así que creé un nuevo marco de datos con sólo las columnas que quería y también cambié el nombre de las columnas por títulos más relevantes.

También comprobé si faltaban valores en alguno de los registros utilizando `value_counts` y `isnull` para comprobar que cada columna imprimía exactamente 76.931 registros.

No se encontró ningún valor que faltara valores perdidos.

En el caso de los datos de los barrios y distritos extraídos, elimine los datos escritos entre paréntesis para cada caso limpiando así en marco, que estuviera escrito entre paréntesis para cada elemento, incluidos los propios paréntesis. La lista de barrios y distritos de Seattle en un marco de datos de pandas y se añadieron los nombres de las columnas.

A continuación, después de usar el geocodificador para obtener las coordenadas de latitud y longitud de cada barrio, eliminé las filas que sin información de latitud y se reinició el índice del marco de datos.

Exporte a un archivo csv permitiendo importar este documento a futuro.

En cuanto a las coordenadas geográficas, Geocoder generaba inexactitud en la ubicación, así que creé un mapa para comprobar si todas las coordenadas estaban dentro del área de Seattle. Si una coordenada parecía estar fuera de la zona de Seattle, utilicé Google Maps para obtener coordenadas aproximadas y cambiaba las coordenadas manualmente en el archivo csv en consecuencia.

3. Metodología

3.1 Análisis exploratorio de los datos

Como primer paso genere una exploración general en el conjunto de datos sobre delincuencia de Seattle echando primero un vistazo al número total de veces que se produjo un delito en Seattle. La mayoría de los delitos en Seattle son "robo de vehículo motorizado", "usurpación de identidad", y "allanamiento de morada", el paso siguiente fue agrupar el recuento de

cada delito por distrito. Esto me permitió generar un marco de datos, pude crear una tabla dinámica con las columnas de los delitos y las filas de los distritos. Los valores de cada celda son el recuento de delitos.

También añadí una columna "Totales" al final de la tabla dinámica que tabulaba el recuento total de delitos para cada distrito.

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
Peeping Tom	19
Welfare Fraud	19
Human Trafficking, Commercial Sex Acts	16
Drunkenness	4
Statutory Rape	3
Justifiable Homicide	2
Assisting or Promoting Prostitution	2
Negligent Manslaughter	1
Bribery	1
Human Trafficking, Involuntary Servitude	1
Name: Offense, dtype: int64	

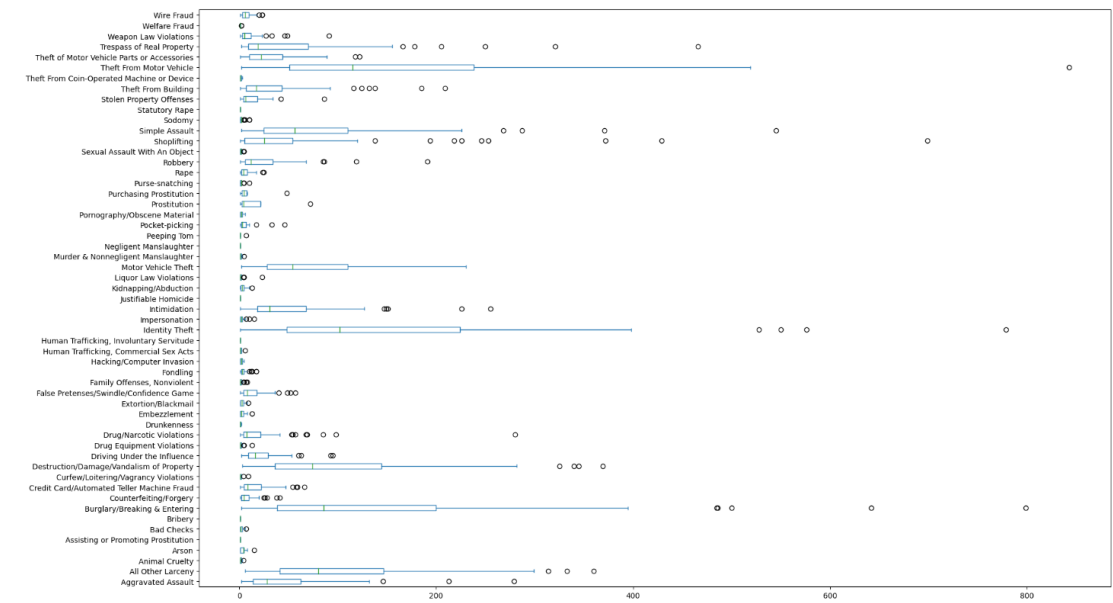
A continuación se muestra un ejemplo de esto para los primeros 5 distritos:

Offense	Aggravated Assault	All Other Larceny	Animal Cruelty	Arson	Assisting or Promoting Prostitution	Bad Checks	Bribery	Burglary/Breaking & Entering	Counterfeiting/Forgery	Card/Automated Teller Machine Fraud	...	Stolen Property Offenses
District												
ALKI	9.0	44.0	NaN	NaN	NaN	NaN	NaN	38.0	1.0	4.0	...	NaN
BALLARD NORTH	23.0	151.0	NaN	5.0	NaN	4.0	NaN	165.0	9.0	22.0	...	3.0
BALLARD SOUTH	76.0	314.0	NaN	5.0	NaN	4.0	NaN	485.0	26.0	45.0	...	22.0
BELLTOWN	99.0	114.0	NaN	1.0	NaN	1.0	NaN	259.0	9.0	19.0	...	11.0
BITTERLAKE	61.0	125.0	NaN	NaN	NaN	1.0	1.0	139.0	5.0	27.0	...	8.0

A continuación, se pueden encontrar las estadísticas descriptivas de la tabla dinámica de delitos.

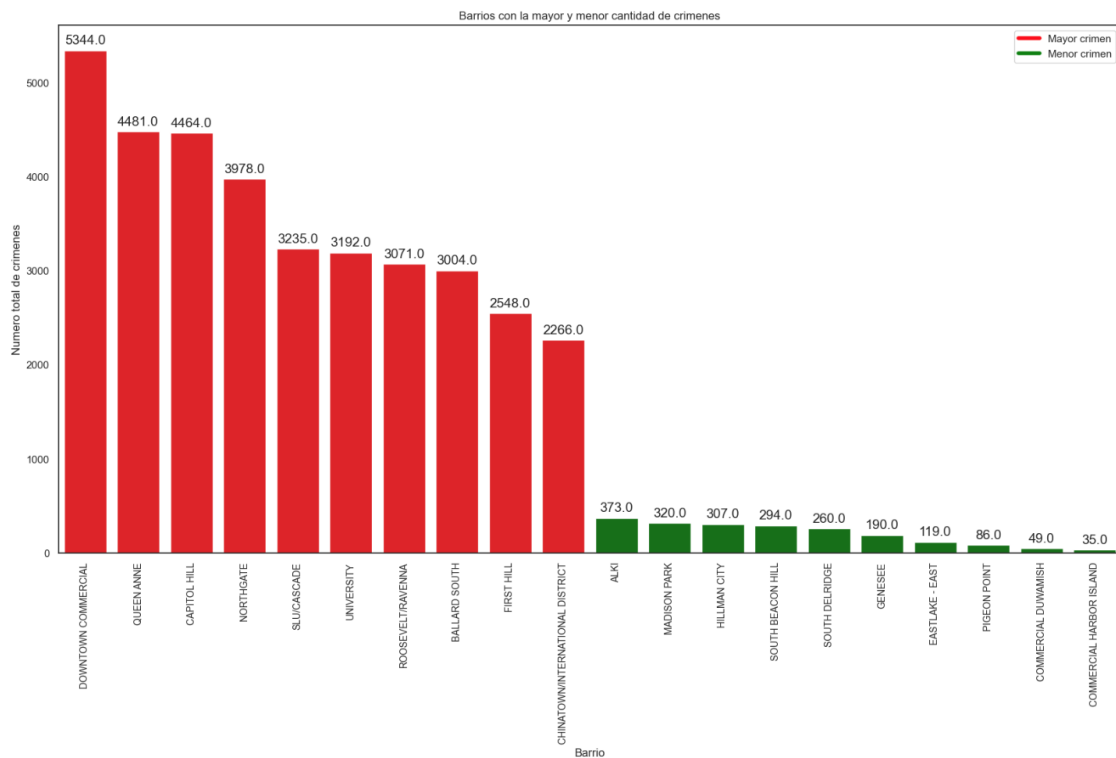
Offense	Aggravated Assault	All Other Larceny	Animal Cruelty	Arson	Assisting or Promoting Prostitution	Bad Checks	Bribery	Burglary/Breaking & Entering	Counterfeiting/Forgery	Card/Automated Teller Machine Fraud	...	Stolen Property Offenses
count	58.000000	58.000000	15.000000	40.000000	2.0	32.000000	1.0	58.000000	51.000000	54.000000	...	50.0
mean	47.137931	106.241379	1.600000	3.650000	1.0	2.250000	1.0	151.982759	8.313725	16.777778	...	12.4
std	51.929253	90.083258	0.910259	2.787794	0.0	1.741338	NaN	170.363851	9.120286	17.195509	...	14.7
min	2.000000	6.000000	1.000000	1.000000	1.0	1.000000	1.0	2.000000	1.000000	1.000000	...	1.0
25%	14.250000	41.000000	1.000000	1.000000	1.0	1.000000	1.0	38.250000	2.000000	5.000000	...	4.0
50%	28.000000	80.000000	1.000000	4.000000	1.0	1.500000	1.0	85.500000	5.000000	8.500000	...	6.5
75%	62.500000	146.750000	2.000000	5.000000	1.0	3.000000	1.0	199.750000	9.500000	22.000000	...	18.5
max	279.000000	360.000000	4.000000	15.000000	1.0	7.000000	1.0	799.000000	41.000000	66.000000	...	86.0

Para visualizar mejor las estadísticas descriptivas de cada delito, se realizó un boxplot:



De acuerdo con el boxplot, existe una dispersión significativa de la sustracción de vehículos, la usurpación de identidad, el robo, y destrucción de la propiedad, lo que se debe a que algunos distritos tienen más de estos delitos que otros distritos.

También encontré los 10 distritos con el mayor número total de delitos y el menor número total de delitos. Luego concatené estas dos listas y creé un gráfico de barras para visualizar mejor la disparidad.



Los 10 distritos con mayor número de delitos totales son:
Downtown Seattle, Queen Anne, Capitol Hill, Northgate, SLU/Cascade, University, Roosevelt/Ravenna, South Ballard, First Hill, and Chinatown/International District.

Los 10 últimos distritos son: Alki, Madison Park, Hillman City, South Beacon Hill, South Delridge, Genesee, Eastlake, Pigeon Point, Commercial Duwamish y Commercial Harbor Island.

3.2 Agrupación de K-means

Me motivo saber como podría a travez de la minería de datos con un conjunto de datos sobre delincuencia.

La agrupación de K-means separa un de datos en distintos subgrupos basados en una característica.

El algoritmo de aprendizaje no supervisado de kmeans clustering se eligió para agrupar o clasificar el gigantesco conjunto de datos sobre delincuencia para identificar los distritos en función de la seguridad.

Para preparar el conjunto de datos para el aprendizaje automático, se utilizó primero una codificación en caliente para crear una tabla que convirtió nuestras variables categóricas de delincuencia en vectores binarios.

A continuación, el conjunto de datos se agrupó por distritos mediante tomando la media.

Elegí agrupar el conjunto de datos en 5 grupos. Con la información de los grupos, la fusioné con la tabla de tabla de codificación, así como con las coordenadas de latitud y longitud. A continuación, se creó un mapa de los clusters codificado por colores para una mejor visualización.

3.3 Exploración de locales en un barrio

Una vez elegido un cluster para su exploración, comencé explorando los locales dentro de los barrios de la agrupación convirtiendo el nombre del distrito y del barrio en el marco de datos del pandas de Seattle en mayúsculas, proseguí agrupando las filas del marco de datos del pandas de Seattle en las que el nombre del barrio contenía una palabra de la lista de distritos del clúster, haciendo lo mismo con los nombres de los distritos que coincidían con un distrito del clúster.

Luego, concatene los 2 dataframes en un nuevo dataframe de pandas, eliminé las filas duplicadas y reinicié el índice.

Esto me dio un marco de datos que consistía en el clúster de barrios, distritos y sus coordenadas de latitud y longitud.

A continuación, definí mis credenciales de Foursquare y la versión de la API. La función getNearbyVenues utiliza la API de Foursquare para obtener los 100 mejores locales en un radio de 1609 metros (o aproximadamente 1 milla) de una ubicación determinada.

Llamé a la función `getNearbyVenues` con las coordenadas de latitud y longitud de cada barrio para obtener una lista de los 100 mejores locales por barrio del clúster.

Con esta lista, volví a utilizar una codificación en caliente para agrupar las filas por barrio y tomar la media de la frecuencia de aparición de cada categoría de local. Ahora podemos imprimir las 5 categorías de locales más categorías de locales más comunes por barrio con esta tabla de frecuencias.

Para que sea más fácil de leer, he puesto esto en un marco de datos de pandas en orden descendente por categoría de local. Para ello, he definido una función llamada `return_most_common_venues` que tomaba cada fila del barrio y el número de locales y devolvía los locales de esa fila del barrio ordenados de forma descendente.

4. Resultados

4.1 Resultados de la agrupación de *k-means*

En la tabla de abajo se puede ver el número de registros de delitos en cada clúster.

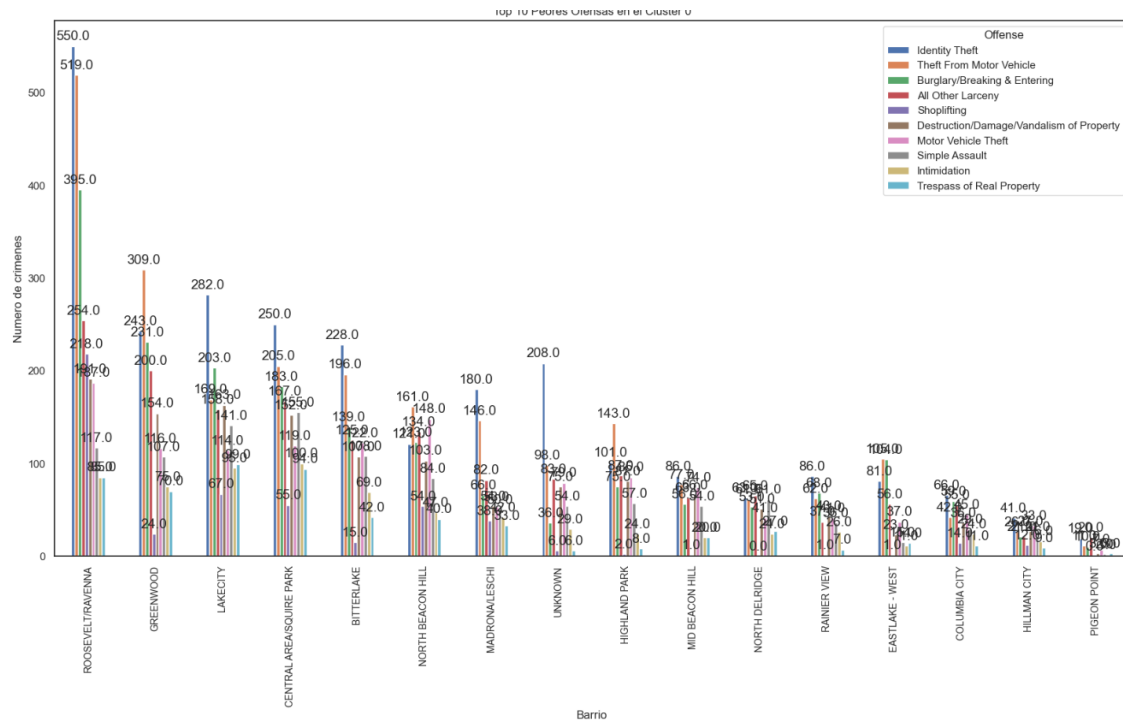
```
Cluster
0      18564
1      26488
2      17253
3      11767
4       2859
Name: District, dtype: int64
```

También agrupé la tabla para encontrar los distritos dentro de cada grupo y el número de delitos tabulados por distrito.

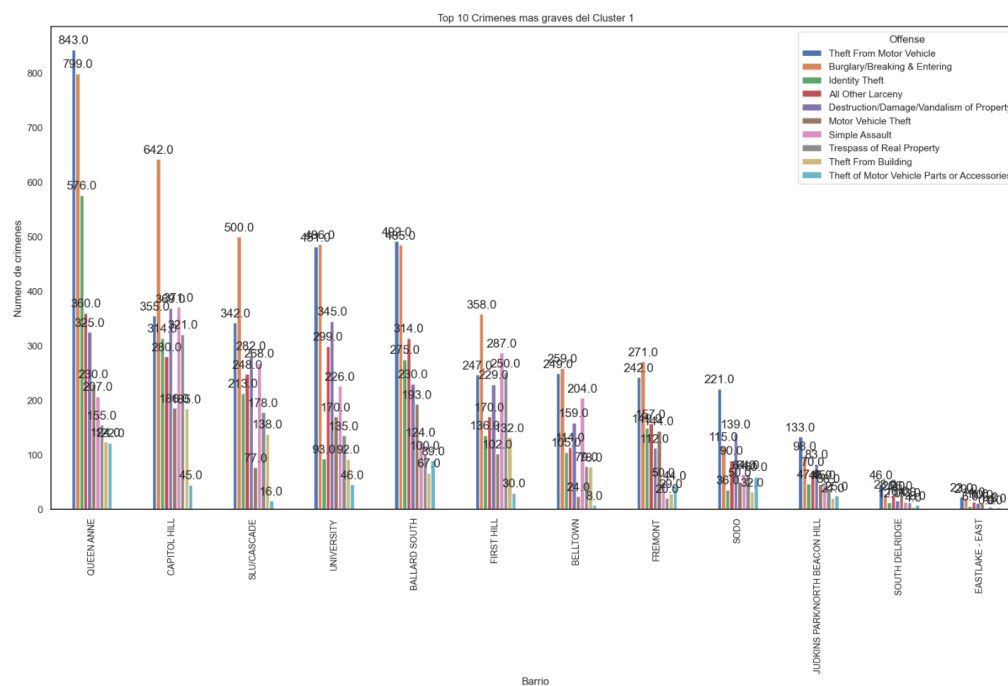
Cluster	District	
0	ALASKA JUNCTION	1252
	BITTERLAKE	1457
	CENTRAL AREA/SQUIRE PARK	1869
	COLUMBIA CITY	475
	EASTLAKE - WEST	507
	GREENWOOD	1987
	HIGHLAND PARK	809
	HILLMAN CITY	307
	LAKECITY	1930
	MADRONA/LESCHI	913
	MID BEACON HILL	631
	NORTH BEACON HILL	1313
	NORTH DELRIDGE	529
	PIGEON POINT	86
	RAINIER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

Name: Count of Offenses. dtype: int64

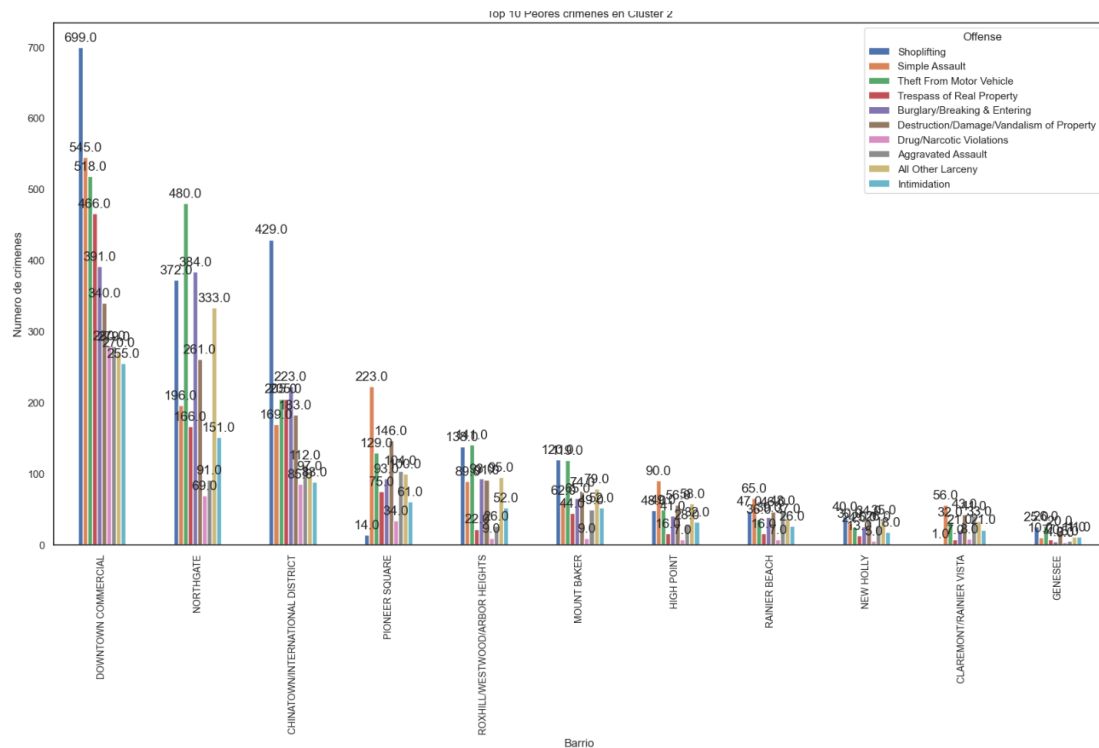
Es fácil ver que el clúster 0 contiene sobre todo robos de vehículos y hurtos.



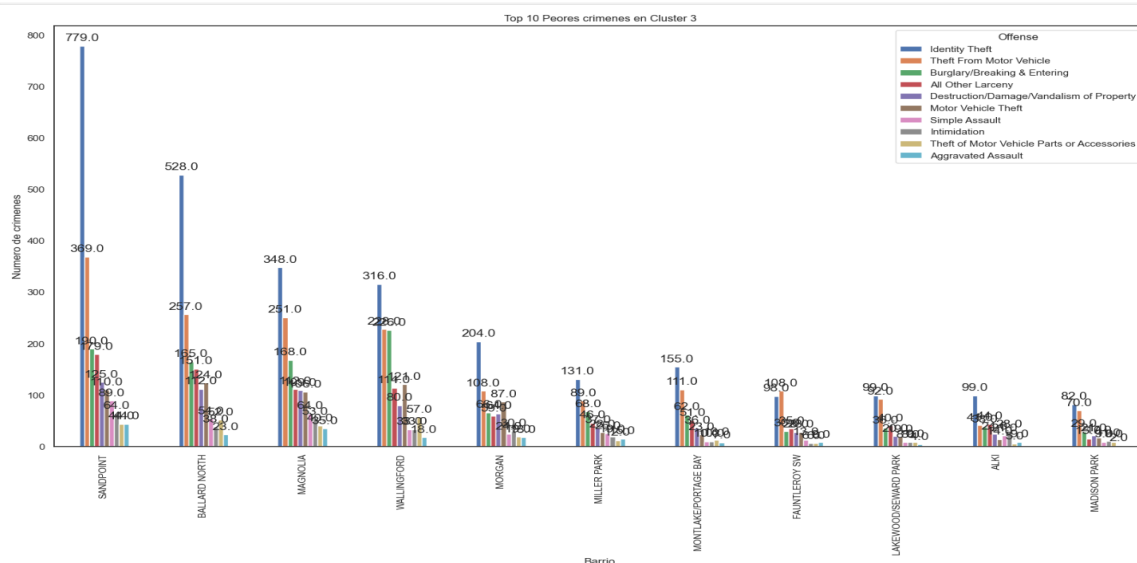
Offense	Count
Theft From Motor Vehicle	843.0
Burglary/Breaking & Entering	799.0
Identity Theft	800
All Other Larceny	800
Other	800



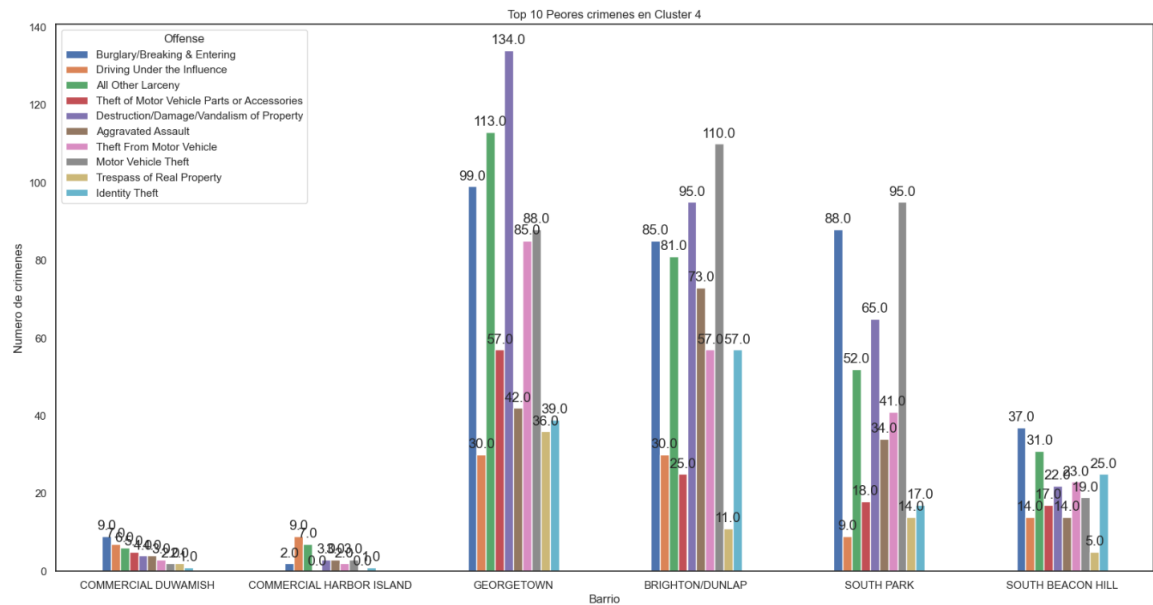
El grupo 2 contiene sobre todo robos, robos de identidad y robos de vehículos.



Para el Grupo 3, el siguiente gráfico de barras muestra los 10 principales delitos en cada uno de los distritos. El grupo 3 contiene principalmente robo de identidad y robo de vehículos de motor.



Para el grupo 4, el siguiente gráfico de barras muestra los 10 principales delitos en cada uno de los distritos. El grupo 4 contiene parece contener sólo las zonas comerciales donde no se producen muchos delitos.



Dado que los conglomerados 0, 1 y 2 contienen distritos que están en la lista de los 10 distritos con mayor número total de delitos, y el conglomerado 4 contiene los distritos con mayor número de delitos. número total de delitos, y el grupo 4 sólo contiene distritos comerciales (no distritos para vivir), esto
Esto reduce nuestra búsqueda de un lugar adecuado y seguro para vivir a sólo el grupo 3.

4.2 Resultados de la exploración de lugares en un barrio

Después de utilizar la API de Foursquare para reunir los 100 mejores locales en un radio de aproximadamente 1,5 km o 1 milla, de cada de cada barrio del Cluster 3, obtuve un resultado de 1519 locales.

El marco de datos resultante constaba de las siguientes columnas: nombre del barrio, latitud del barrio, longitud del barrio, nombre del local, latitud del local, longitud del local y categoría del local.

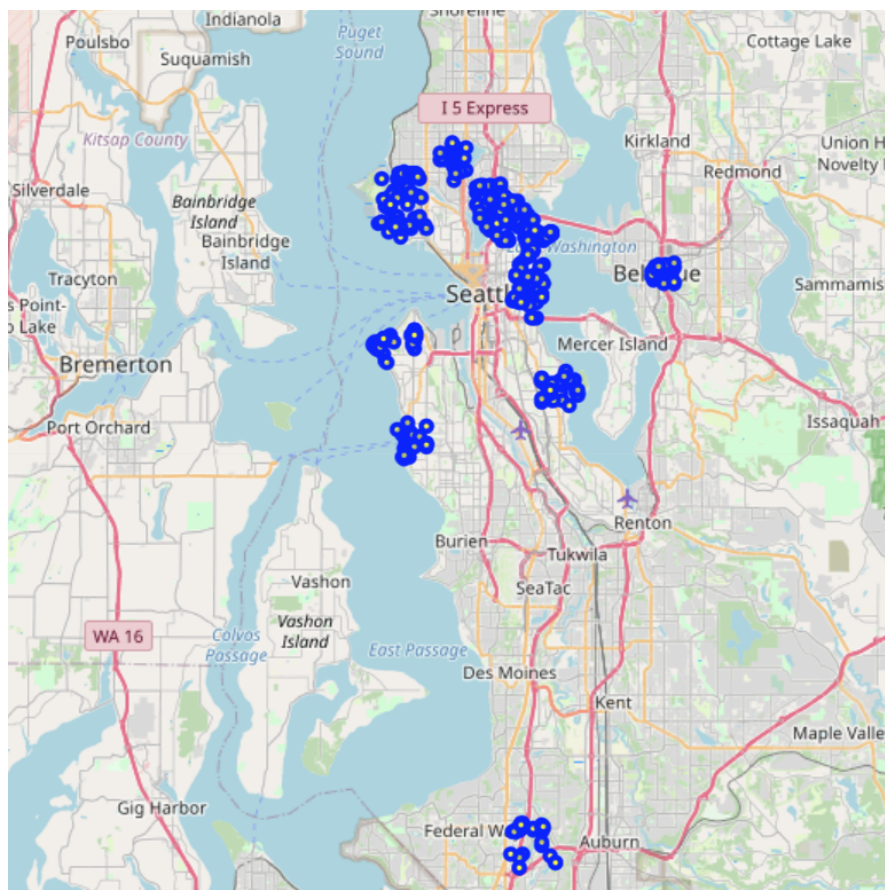
A continuación se muestra un ejemplo:

	Barrio	Barrio Latitud	Barrio Longitud	Venue	Venue Latitude	Venue Longitude	Venue Category
0	WALLINGFORD	47.65555	-122.3265	Ivar's Salmon House	47.653664	-122.324177	American Restaurant
1	WALLINGFORD	47.65555	-122.3265	Ivars Fish Bar	47.653714	-122.323941	Seafood Restaurant
2	WALLINGFORD	47.65555	-122.3265	Voula's Offshore Cafe	47.654636	-122.321453	Café
3	WALLINGFORD	47.65555	-122.3265	Westward	47.651790	-122.328831	American Restaurant
4	WALLINGFORD	47.65555	-122.3265	Whalemaker Lounge	47.653540	-122.323928	Cocktail Bar

Los resultados nos marcan que no todos los barrios tienen 100 locales, quería agrupar el marco de datos para obtener un recuento del número de locales encontrados por barrio.

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

Se obtuvieron 229 categorías únicas de locales y se imprimió esta lista para verificarla. Se creó un mapa para visualizar dónde están todos estos locales (marcados en azul).



Después de analizar cada barrio del grupo 3 con los locales devueltos, imprimí una tabla con las 10 categorías de locales más comunes para cada barrio.

	Barrio	ATM	Accounting and Bookkeeping Service	Advertising Agency	Afghan Restaurant	African Restaurant	Alternative Medicine Clinic	American Restaurant	Arcade	Art Gallery	...	Thai Restaurant	Toy / Game Store	Travel Agency	Urban Park
0	ALKI POINT	0.00	0.00	0.02	0.00	0.00	0.00	0.040000	0.00	0.02	...	0.00	0.00	0.00	0.00
1	BRIARCLIFF	0.04	0.00	0.00	0.00	0.00	0.00	0.000000	0.00	0.02	...	0.00	0.00	0.00	0.00
2	BROADMOOR	0.00	0.00	0.00	0.00	0.00	0.00	0.000000	0.00	0.00	...	0.00	0.00	0.00	0.04
3	FAUNTLEROY	0.00	0.00	0.00	0.00	0.00	0.00	0.000000	0.00	0.02	...	0.00	0.00	0.00	0.04
4	LAKEWOOD	0.00	0.00	0.00	0.00	0.00	0.00	0.040000	0.00	0.00	...	0.00	0.00	0.00	0.00
5	LAWTON PARK	0.00	0.00	0.00	0.00	0.00	0.00	0.020000	0.00	0.00	...	0.02	0.00	0.00	0.02
6	LESCHI	0.00	0.00	0.00	0.00	0.04	0.00	0.040000	0.00	0.00	...	0.02	0.00	0.00	0.00
7	MADISON PARK	0.00	0.00	0.00	0.00	0.00	0.00	0.000000	0.00	0.02	...	0.02	0.02	0.00	0.02
8	MADRONA	0.00	0.00	0.00	0.00	0.06	0.02	0.020000	0.00	0.02	...	0.06	0.00	0.00	0.04
9	MADRONA VALLEY	0.00	0.00	0.00	0.00	0.00	0.00	0.000000	0.00	0.00	...	0.00	0.00	0.00	0.00
10	MAGNOLIA	0.00	0.00	0.00	0.00	0.00	0.00	0.040000	0.02	0.00	...	0.02	0.00	0.00	0.00
11	MONTLAKE	0.00	0.00	0.00	0.00	0.00	0.00	0.000000	0.00	0.00	...	0.00	0.00	0.00	0.00
12	NORTH ADMIRAL	0.00	0.00	0.00	0.00	0.00	0.00	0.020000	0.00	0.00	...	0.02	0.00	0.00	0.00
13	NORTHLAKE	0.00	0.02	0.00	0.00	0.00	0.00	0.020000	0.00	0.00	...	0.00	0.00	0.00	0.00
14	PHINNEY RIDGE	0.00	0.00	0.00	0.00	0.00	0.00	0.020000	0.00	0.00	...	0.02	0.00	0.00	0.00
15	PIGEON POINT	0.00	0.00	0.00	0.00	0.00	0.00	0.111111	0.00	0.00	...	0.00	0.00	0.00	0.00
16	PORTAGE BAY	0.00	0.00	0.00	0.00	0.00	0.00	0.080000	0.00	0.00	...	0.02	0.00	0.00	0.00
17	SEWARD PARK	0.00	0.00	0.00	0.00	0.06	0.00	0.040000	0.02	0.00	...	0.00	0.00	0.00	0.02
18	SOUTHEAST MAGNOLIA	0.02	0.00	0.00	0.00	0.00	0.00	0.000000	0.00	0.00	...	0.00	0.00	0.02	0.00
19	WALLINGFORD	0.00	0.00	0.00	0.02	0.00	0.00	0.080000	0.00	0.00	...	0.00	0.00	0.00	0.00

La mayoría de las primeras o segundas categorías de locales más comunes del Cluster 3 son cafeterías y parques, lo que me lleva a creer que el grupo 3 contiene barrios que están en los suburbios de Seattle. Como recordatorio, estoy buscando un barrio similar a las características de California con servicios y formas de vida, que es una ciudad suburbana ciudad suburbana segura con mucha cocina internacional. Madrona Valley tiene restaurantes chinos como primera categoría de restaurantes chinos como su primera categoría más común y restaurantes de dumplings como su sexta categoría más común, lo que lo hace el lugar principal para vivir en Seattle, por la diversidad de alimentos chinos como ofrece gran parte de California. Otro lugar en el que me interesaría mirar Wallingford por su variedad de cocinas internacionales en general (y tiene tiendas de té de burbujas como su décima categoría de local más común).

5. Discusión

Tomando en cuenta los resultados, he llegado a la conclusión de que los barrios del grupo 3 se encuentran entre los más seguros de Seattle porque no contienen ningún barrio de la lista de los 10 barrios con mayor índice de criminalidad de la delincuencia.

Al investigar los locales de los barrios del Grupo 3, he reducido aún más mi búsqueda de un lugar adecuado a Madrona Valley, debido a su proximidad a los chinos y restaurantes de dumplings.

Una observación que hice fue que los distritos de Seattle no estaban bien definidos, por lo que el número de distritos en el conjunto de datos sobre la delincuencia no coincidía con el número de distritos obtenido del raspado de la web.

Los locales que se dan también dependen de cuándo se llamó a la API de Foursquare. Por lo tanto, si se repite este ejercicio de nuevo arrojará resultados de lugares diferentes.

6. Conclusión

En conclusión, mediante el raspado de la web, la limpieza de datos, el análisis exploratorio y la agrupación de k-means, y utilizando la API de Foursquare, he encontrado que el lugar más seguro y comparable para vivir en Seattle Washington es Madrona Valley.