

Aplicación de Data Science para el análisis de los barrios de Seattle

Diego Silvera





Introducción

- La economía actual de los últimos años, ha hecho que la gente busque nuevos lugares para vivir
- Esto hace que la gente busque lugares más económicos y seguros para vivir
- Se analiza la ciudad de Seattle como posible ciudad a mudarse
 - Cuál es el barrio más seguro?
 - Qué servicios y lugares ofrecen los barrios de Seattle



Fuente de los datos

- Conjunto de datos de los crímenes de Seattle: obtenidos de los registros públicos de la policía de Seattle.
- Conjunto de Barrios y Distritos de Seattle obtenidos de la pagina web de Wikipedia
- Información de los lugares disponibles: Foursquare



Limpieza de los datos

- Se limitó la cantidad de crímenes a los del último año
- Se eliminó cualquier información que estuviera con paréntesis del conjunto de datos
- Se utilizó geocoder para obtener la longitud y latitud de cada barrio
- Se borró cualquier columna que no tuviera coordenadas
- Se borró cualquier dato que no fuera de la ciudad de Seattle usando Google Maps



Análisis de los conjuntos de datos de crímenes

10 Crímenes mas comunes

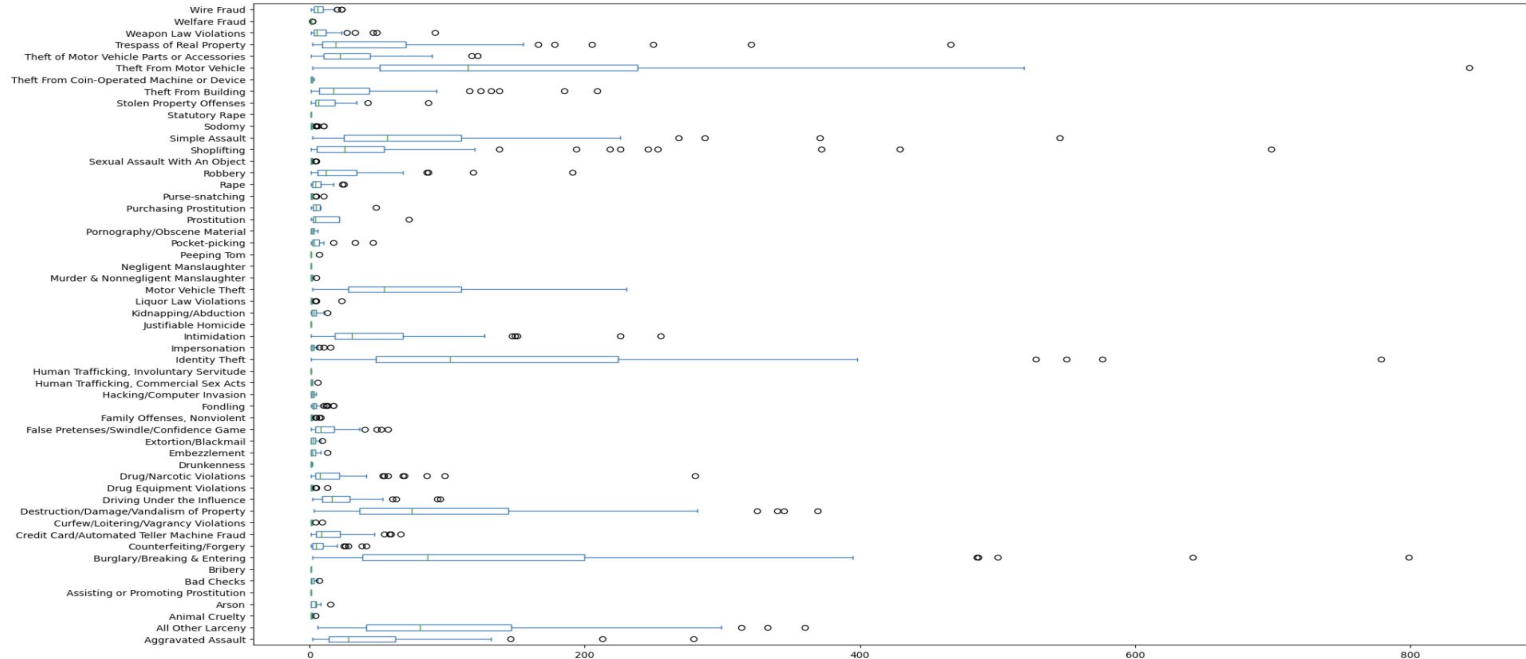
Theft From Motor Vehicle	10035
Identity Theft	9442
Burglary/Breaking & Entering	8981
All Other Larceny	6261
Destruction/Damage/Vandalism of Property	6111
Simple Assault	5164
Motor Vehicle Theft	4387
Shoplifting	3881
Trespass of Real Property	3224
Intimidation	3011



Análisis de los conjuntos de datos de crímenes - Datos descriptivos de cada crimen

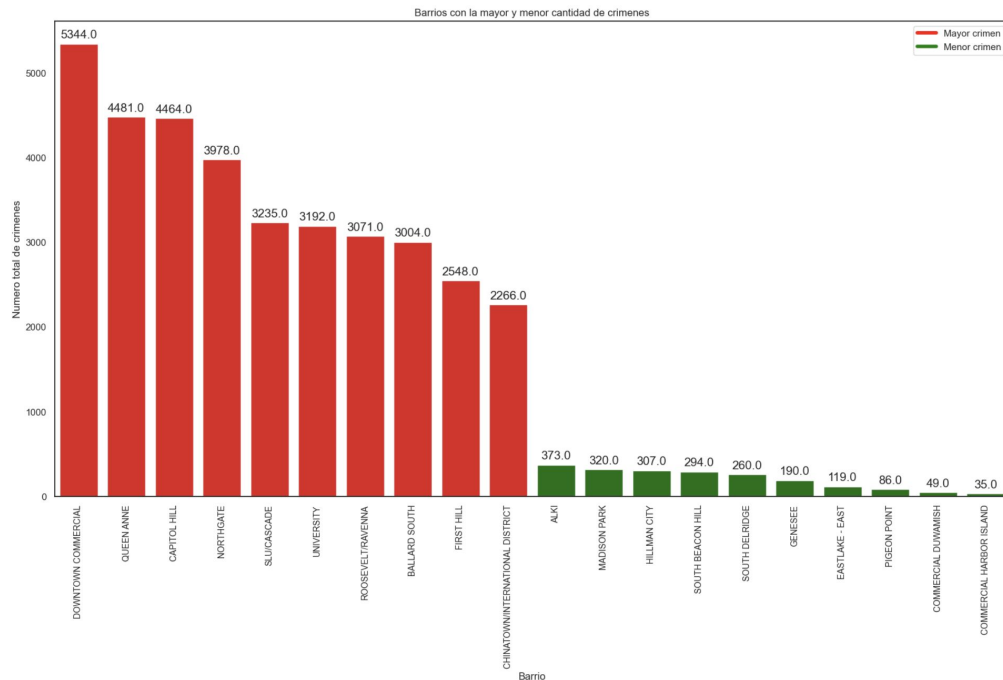
Offense	Aggravated Assault	All Other Larceny	Animal Cruelty	Arson	Assisting or Promoting Prostitution	Bad Checks	Bribery	Burglary/Breaking & Entering	Counterfeiting/Forgery	Credit Card/Automated Teller Machine Fraud	...	\$ Prec Off
count	58.000000	58.000000	15.000000	40.000000	2.0	32.000000	1.0	58.000000	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

Análisis de los conjuntos de datos de crímenes - Datos descriptivos de cada crimen





Distritos de Seattle con mayor y menor índice de crímenes



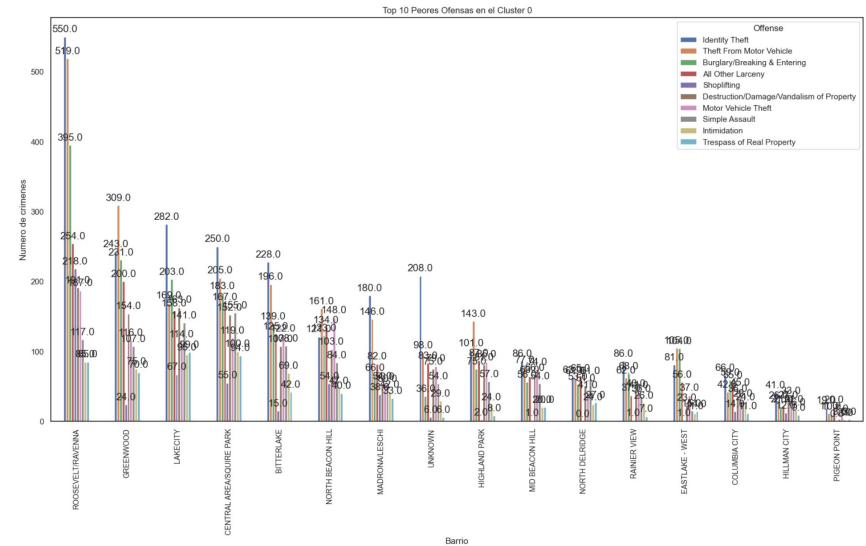


K-means para el clustering de datos por crimen

- K-means es un algoritmo que nos ayudará a categorizar el conjunto de datos de los crímenes dentro de los barrios de Seattle
- Para preparar el conjunto de datos se usa One-hot encoding que convierte las variables categóricas en vectores binarios
- Se eligió crear 5 clusters

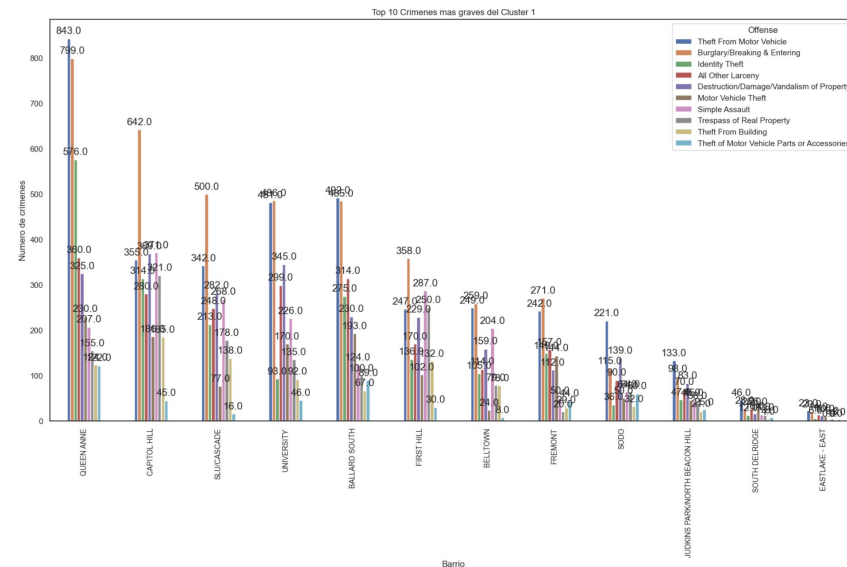
- ROOSEVELT/RAVENNA
- Greenwood
- Lakecity

est son los crímenes principales



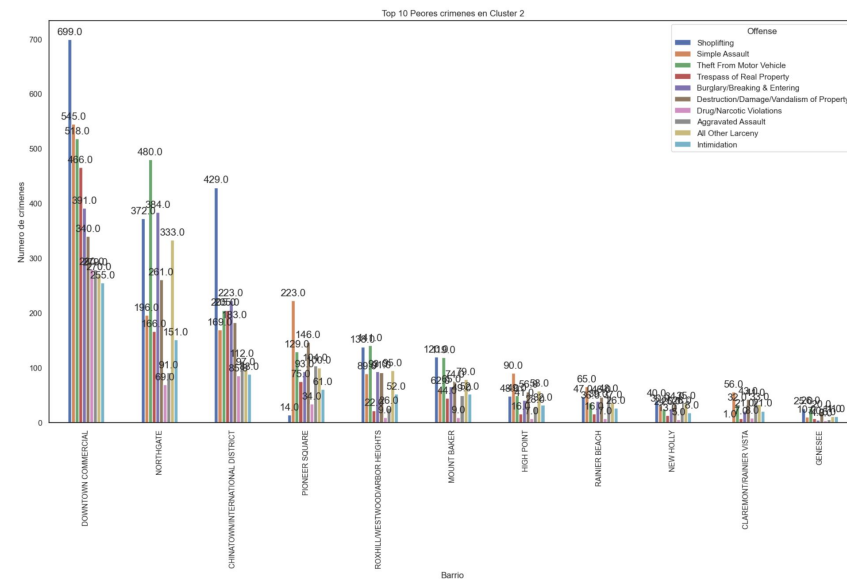
- Queen Anne
- Capitol Hill
- Slu/Cascade

Motor Vehicle theft y Burglary son los
crímenes principales



- Downtown Commercial
- Northgate
- Chinatown/International district

Shoplifting theft y Simple assault son los crímenes principales



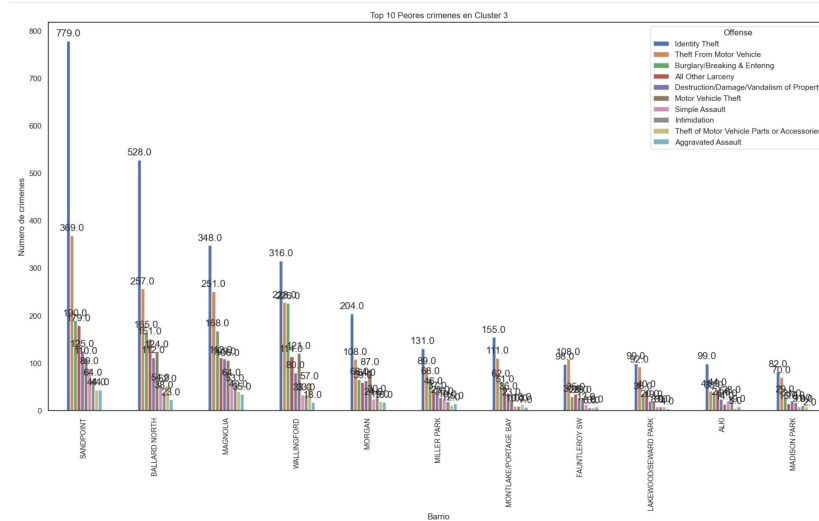


Cluster 3

Distritos con mayor índice de delito:

- Sandpoint
- Ballard North
- Magnolia

Identity theft y Theft from motor son los crímenes principales

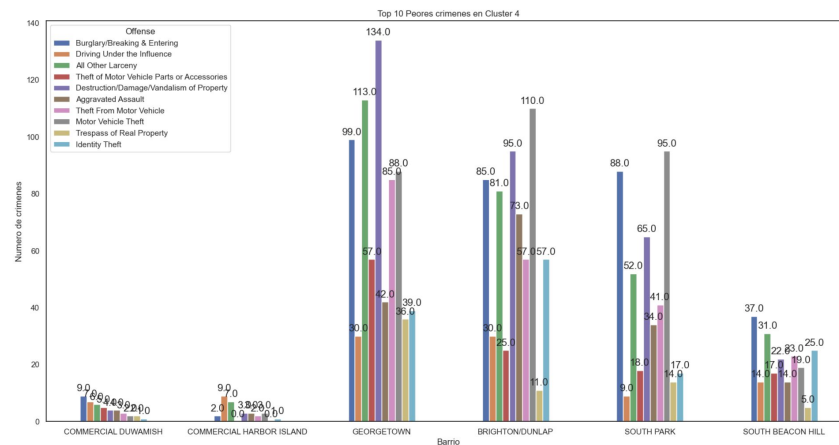




Cluster 4

Distritos con mayor índice de delito:

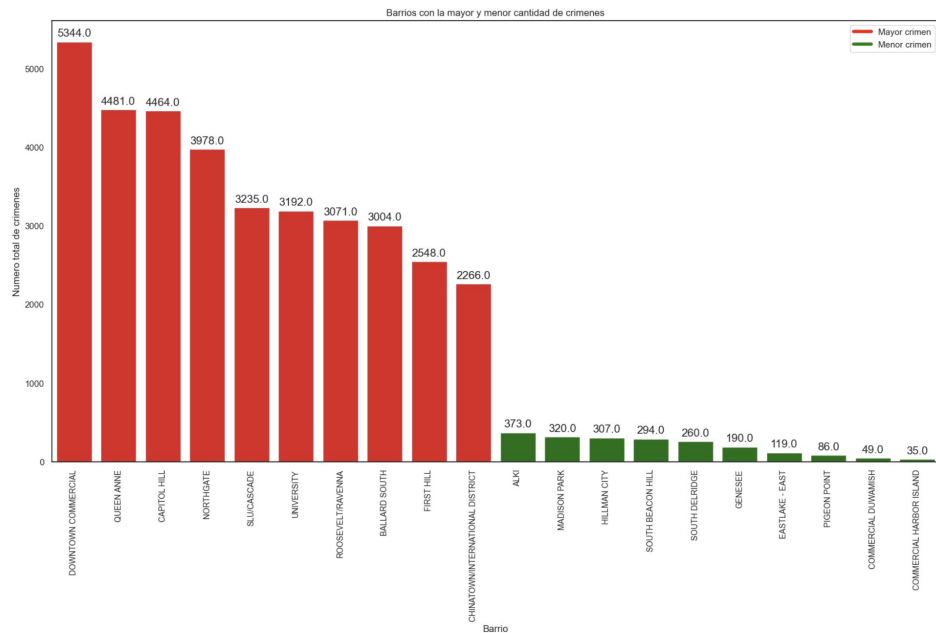
Son en su mayoría barrios comerciales y los otros barrios no tienen una cantidad significativa de crímenes registrados





Seleccionamos Cluster a analizar

Se elige el cluster 3 porque es el único cluster que no tiene barrios en el top 10 de barrios con mayor índice de crimen.



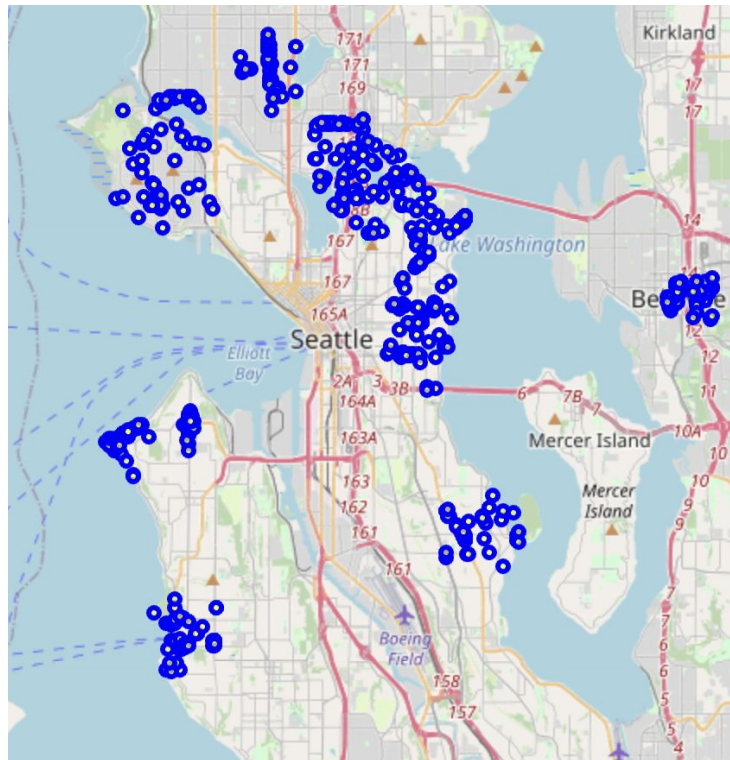


Obtenemos los lugares disponibles en el cluster 3

Se obtienen 910 lugares dentro del data set

	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

Mapa de los lugares del cluster 3





Top 10 de los categorías de lugares mas comunes de cada Barrio

	Barrio	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	Urban Park	Italian Restaurant	Coffee Shop	Bar	Fish and Chips Shop	Monument	Pizzeria	Beach	American Restaurant	Restaurant
1	BRIARCLIFF	Hiking Trail	Playground	Bank	Scenic Lookout	Pizzeria	Park	Grocery Store / Supermarket	Coffee Shop	ATM	Farmers' Market
2	BROADMOOR	Garden	Bakery	Park	Hiking Trail	Playground	Coffee Shop	Restaurant	Bank	Pizzeria	Urban Park
3	FAUNTLEROY	Park	Hair Salon	Beach	Playground	Hiking Trail	Restaurant	Attorney / Law Office	Urban Park	No Category	Coffee Shop
4	LAKEWOOD	Clothing Store	Fast Food Restaurant	Pizzeria	Department Store	Burger Joint	Bar	Furniture and Home Store	Coffee Shop	Grocery Store / Supermarket	American Restaurant
5	LAWTON PARK	Cocktail Bar	Hiking Trail	Burger Joint	Japanese Restaurant	Beer Bar	Café	Mexican Restaurant	Bakery	Playground	Brewery
6	LESCHI	Urban Park	Playground	Bar	Coffee Shop	Grocery Store / Supermarket	Café	Cocktail Bar	Bakery	Fried Chicken Joint	American Restaurant
7	MADISON PARK	Hair Salon	Spa	Park	Bank	Bakery	Furniture and Home Store	Bookstore	Hair Removal Service	Sports Bar	Grocery Store / Supermarket
8	MADRONA	Coffee Shop	Thai Restaurant	African Restaurant	Bar	Dry Cleaner	Playground	Urban Park	Bakery	Fried Chicken Joint	Café
9	MADRONA VALLEY	Industrial Equipment Supplier	ATM	Music Venue	Mediterranean Restaurant	Mental Health Clinic	Metals Supplier	Mexican Restaurant	Miscellaneous Store	Mobile Phone Store	Monument
10	MAGNOLIA	Car Dealership	Coffee Shop	Restaurant	Grocery Store / Supermarket	Furniture and Home Store	Bakery	Sushi Restaurant	American Restaurant	Spa	Sports Bar
11	MONTLAKE	Garden	Hiking Trail	Park	Bridge	Hair Salon	Urban Park	Lake	Landmarks and Outdoors	Harbor / Marina	Other Great Outdoors
12	NORTH ADMIRAL	Hair Salon	Grocery Store / Supermarket	Beer Bar	Pizzeria	Cocktail Bar	Spa	Restaurant	Bank	Miscellaneous Store	Boutique
13	NORTHLAKE	Hair Salon	Convenience Store	Restaurant	Health and Beauty Service	Coffee Shop	Playground	Deli	Department Store	Nail Salon	Financial Planner
14	PHINNEY RIDGE	Zoo	Bar	Pizzeria	Restaurant	Bakery	Beer Bar	Tattoo Parlor	Playground	Coffee Shop	Cocktail Bar
15	PIGEON POINT	Hiking Trail	Landmarks and Outdoors	Beach	American Restaurant	No Category	Seafood Restaurant	Fish and Chips Shop	Museum	Mental Health Clinic	Metals Supplier
16	PORTAGE BAY	Restaurant	American Restaurant	Playground	Urban Park	Bar	Hair Salon	Diner	Deli	Pizzeria	Landmarks and Outdoors
17	SEWARD PARK	Ethiopian Restaurant	African Restaurant	Playground	Park	Bookstore	American Restaurant	Pizzeria	Bar	Café	Hiking Trail
18	SOUTHEAST MAGNOLIA	Playground	Pizzeria	Drugstore	Coffee Shop	Sushi Restaurant	Bank	Soup Restaurant	Shoe Repair Service	Italian Restaurant	Dance Studio
19	WALLINGFORD	American Restaurant	Burger Joint	Restaurant	Bakery	Pet Supplies Store	Café	Cocktail Bar	Italian Restaurant	Music Store	Greek Restaurant



Eligiendo el mejor barrio

Se elige el barrio de Madrona Valley por su cercanía a la cocina internacional y servicios cerca disponibles.

En segundo lugar seleccionamos North Admiral por su disponibilidad de bares y pizzas.

Se destaca que todos los barrios poseen disponibilidad de cafeterías y restaurantes.



Conclusión

- Barrio mas seguros: los del cluster 3
- Mejor barrio a mi criterio para vivir: Madrona Valley
- La api de Foursquare es devuelve diferentes resultados en diferente tiempo, por lo que los servicios cambian si se llaman en otro momento
- Pueden estar faltando distritos ya que no están muy bien definidos para lo que es Seattle.