

# Using Data Science to Analyze the Neighborhoods of Seattle



Sunatullo Abdullaev



Since COVID-19 lockdown in the United States, trend of remote working and moving away from big, expensive cities

Interest in moving to Seattle, Washington

What is the safest neighborhood in Seattle?

What are the top common venues in the neighborhood?

# Introduction

# Methodology



Gather datasets

Data cleaning

Exploratory  
data

analysis of  
crime  
data

K-means  
clustering  
of  
crime data

Pick a cluster  
to  
explore

Explore venues in  
the cluster using  
Foursquare API

Pick the best  
neighborhood  
in  
the cluster

Seattle crime dataset: available for public download from Seattle PD website

Seattle districts & neighborhoods: web scrape Wikipedia page

Venue information: Foursquare API

## Data Sources



# Methodology



# Data Cleaning

# Data Cleaning

(Aug 2019 to Aug 2020) = 76,931 records

Used geocoder to gather latitude and longitude coordinates

Map coordinates and manually change any outside of Seattle area using Google Maps



# Methodology

Gather datasets

Data cleaning

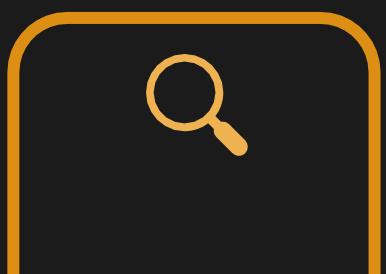
Exploratory  
data  
analysis of  
crime  
data

K-means  
clustering  
of  
crime data

Pick a cluster  
to  
explore

Explore venues in  
the cluster using  
Foursquare API

Pick the best  
neighborhood  
in  
the cluster

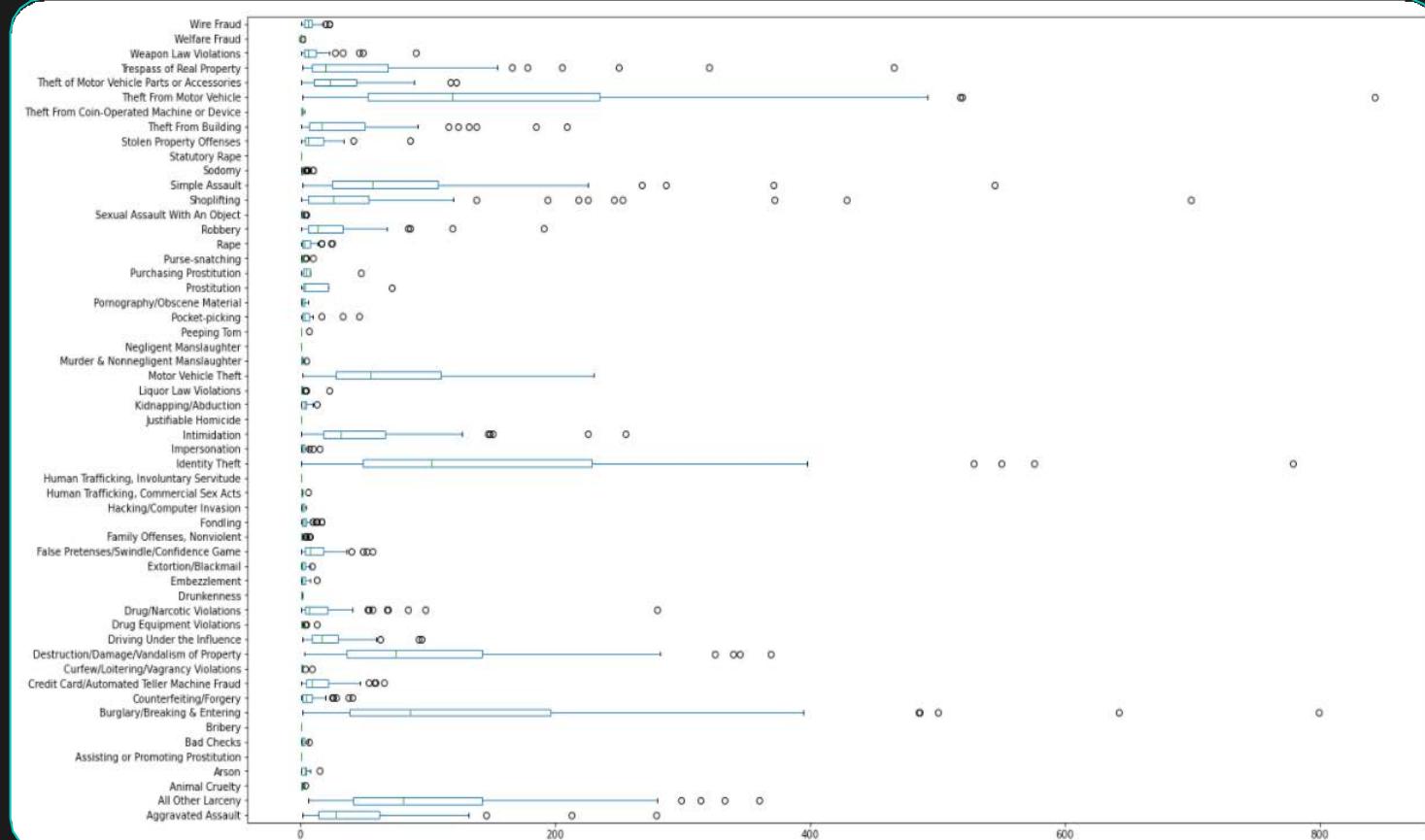


# Exploratory data analysis of crime dataset

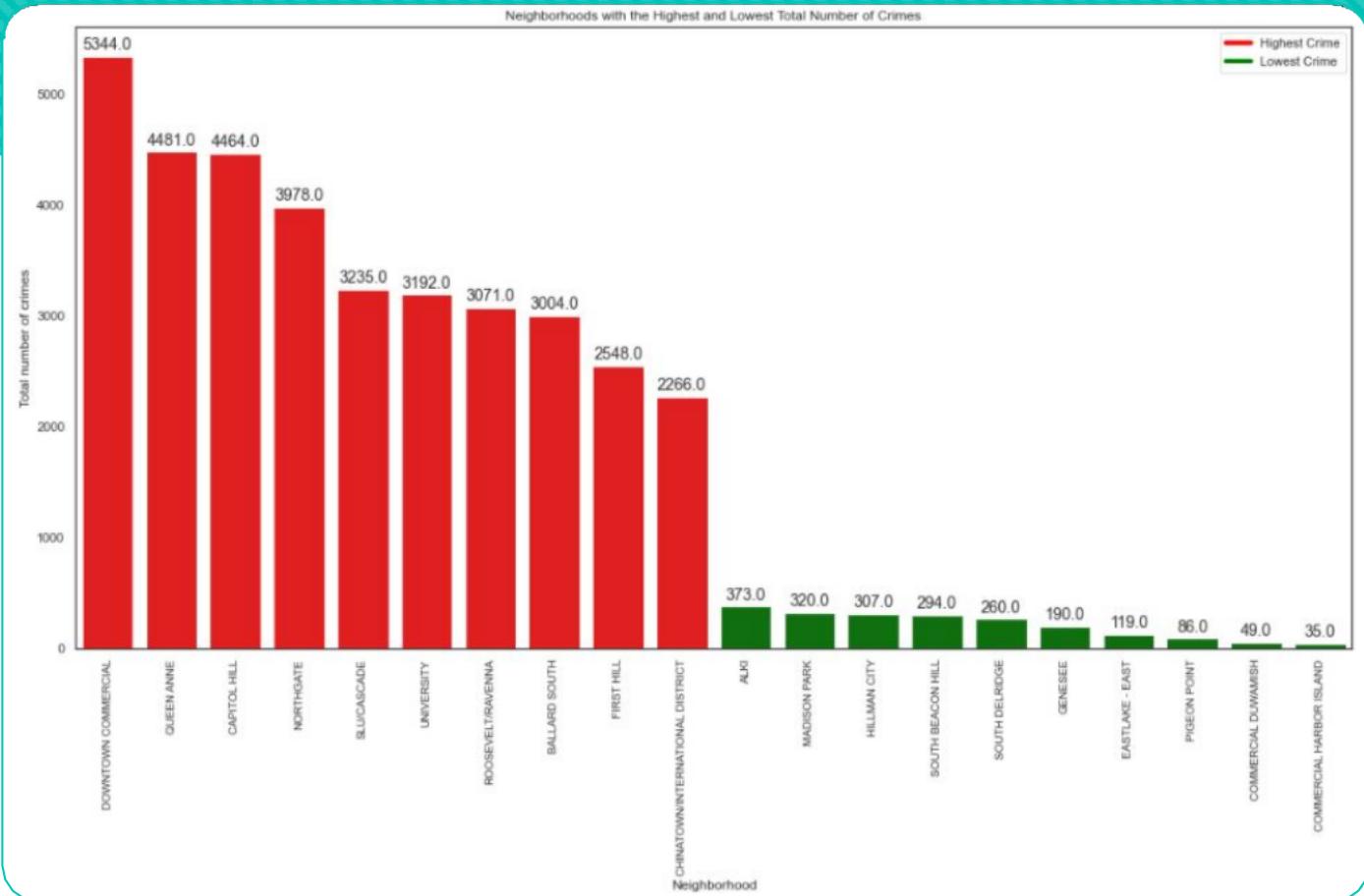
Theft From Motor Vehicle	10035
Identity Theft	9442
Burglary/Breaking & Entering	8981
All Other Larceny	6261
Destruction/Damage/Vandalism of Property	6111
Simple Assault	5164
Motor Vehicle Theft	4387
Shoplifting	3881
Trespass of Real Property	3224
Intimidation	3011

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	

Descriptive stats of each offense



# Seattle districts with highest and lowest crime



# Methodology

Gather datasets

Data cleaning

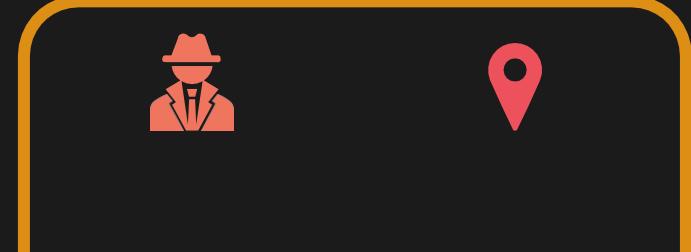
Exploratory  
data  
analysis of  
crime  
data

K-means  
clustering  
of  
crime data

Pick a cluster  
to  
explore

Explore venues in  
the cluster using  
Foursquare API

Pick the best  
neighborhood  
in  
the cluster



# K-means clustering of crime data

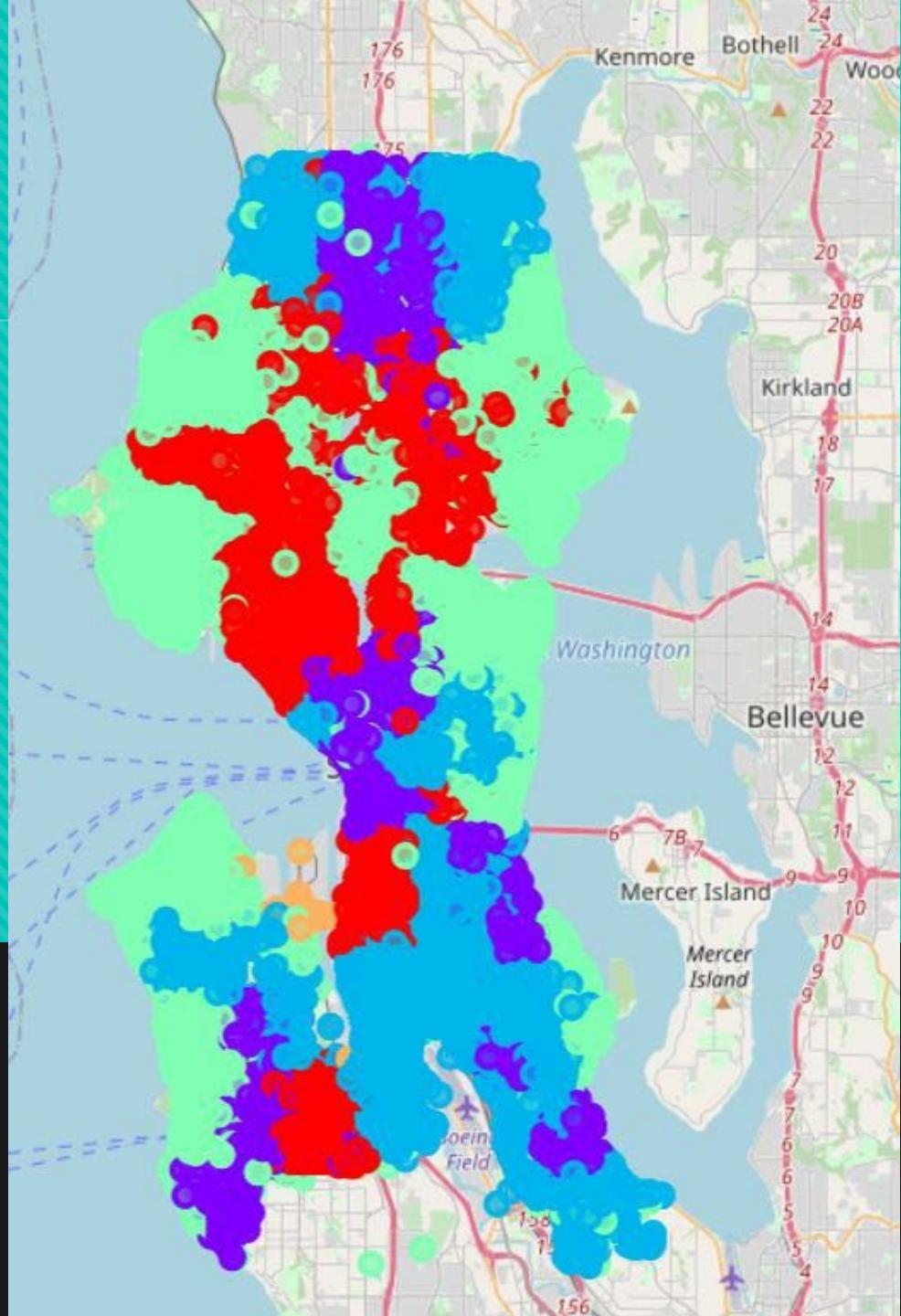
**What is it and why use it?** Unsupervised learning algorithm that will be used to group or classify large crime dataset to identify districts based on safety

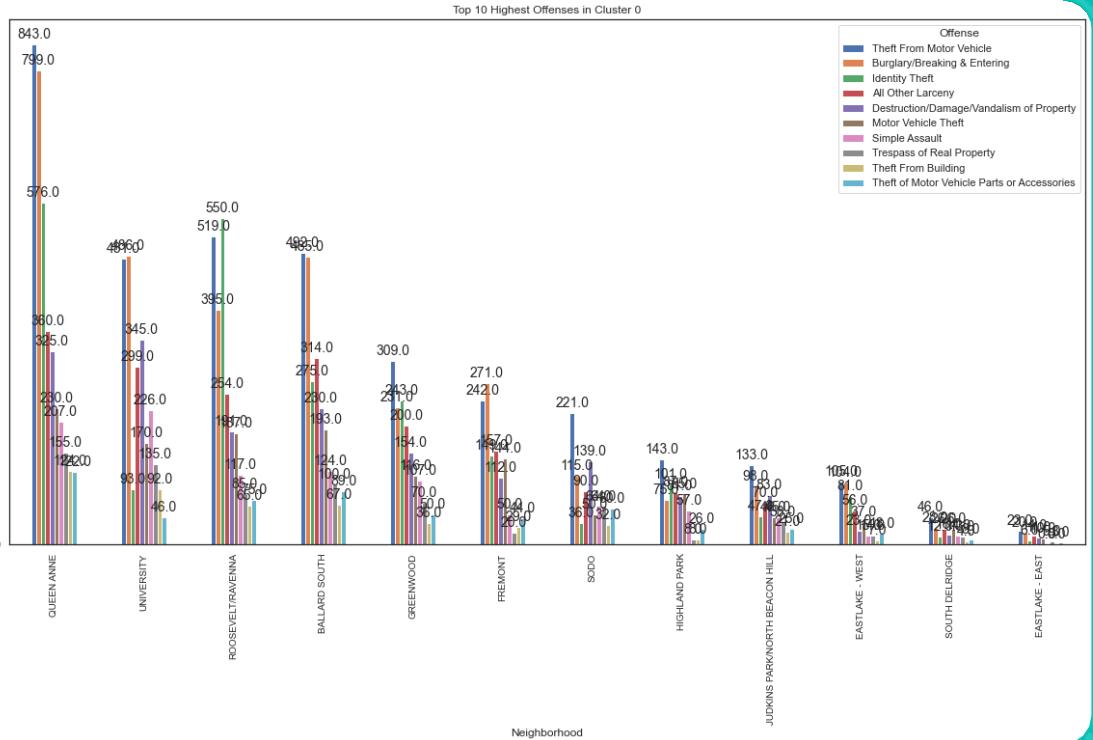
To prepare dataset: one- hot encoding turns categorical variables into binary vectors

Chose to group into 5 clusters

# Result from k-means clustering

Cluster 0 = red  
Cluster 1 = purple  
Cluster 2 = blue  
Cluster 3 = green  
Cluster 4 = orange

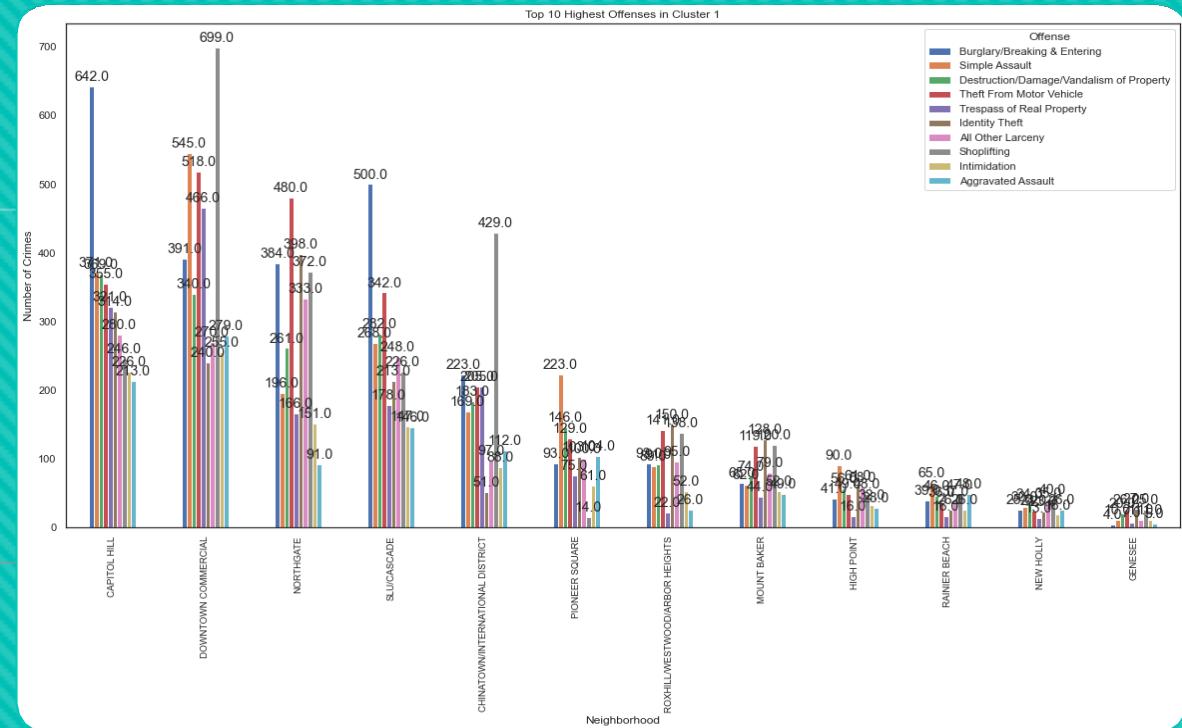




# Cluster 0

High crime districts: Queen Anne, University, Roosevelt/Ravenna, South Ballard

- Mostly motor vehicle theft & burglary



# Cluster 1

High crime districts: Capitol Hill, Downtown, Northgate, SLU/ Cascade, Chinatown/International District

- Mostly burglary, shoplifting, simple assault, & motor vehicle theft

# Cluster 2

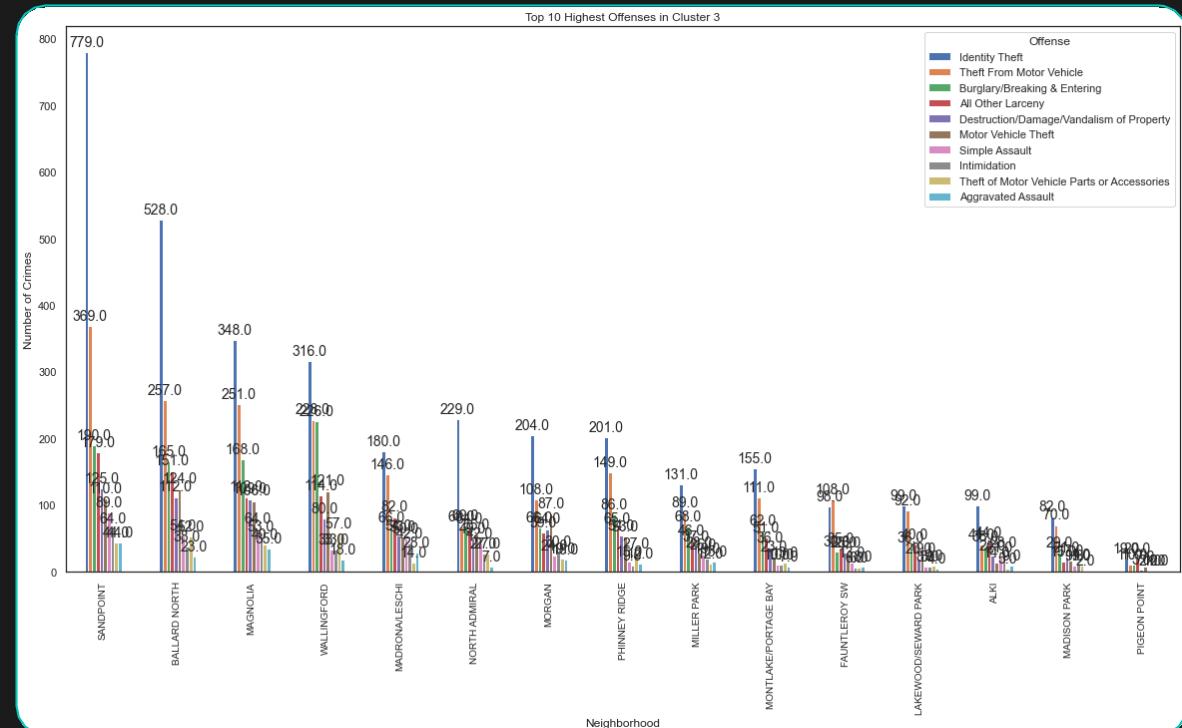
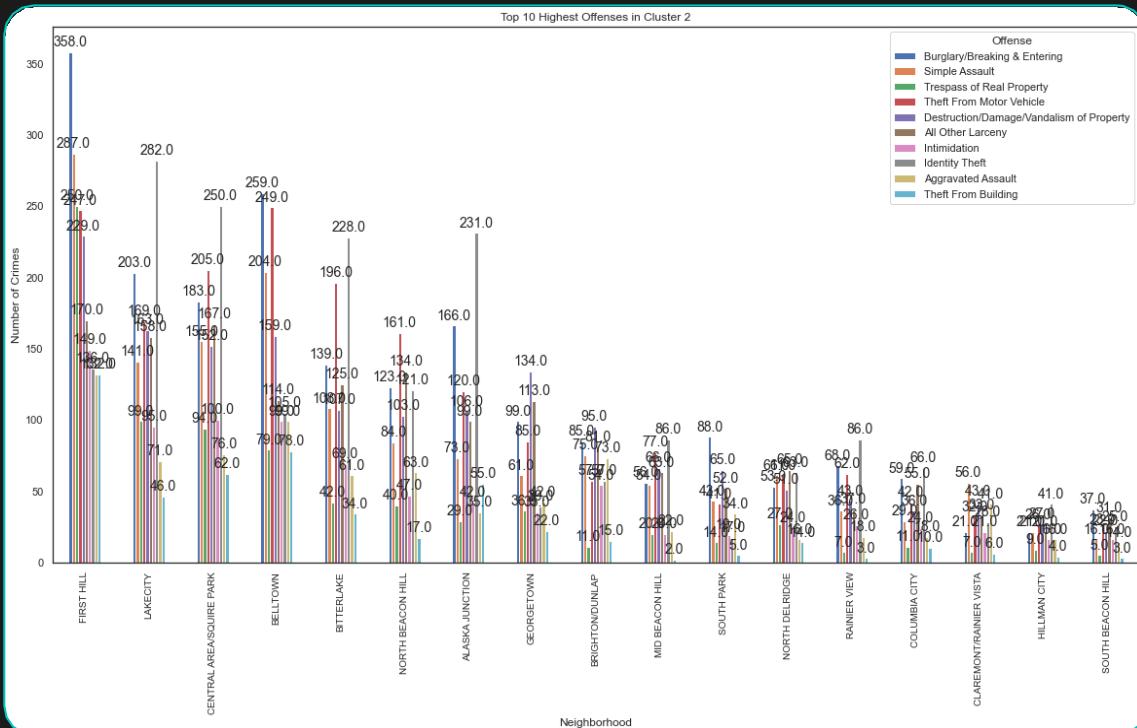
High crime districts: First Hill, Lakecity , Central Area/Square Park, Belltown, Bitterlake , Alaska Junction

- Mostly burglary, identity theft & motor vehicle theft

# Cluster 3

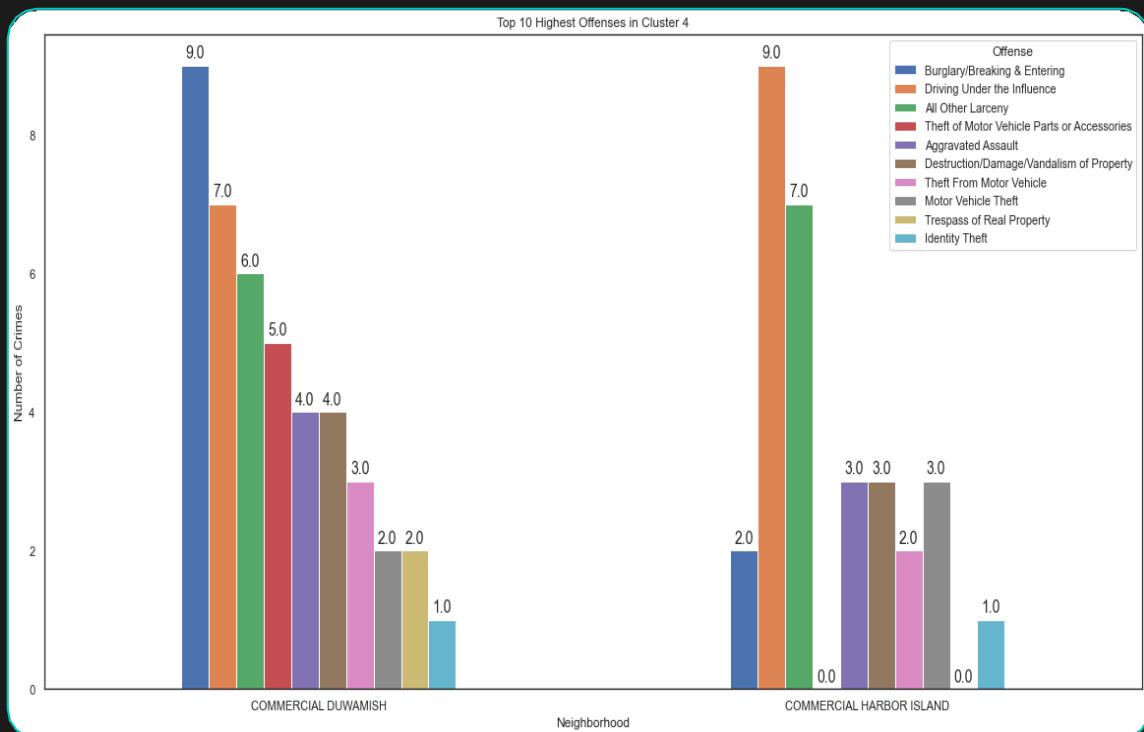
High crime districts: Sandpoint, North Ballard, Magnolia, Wallingford

- Mostly identity theft & motor vehicle theft



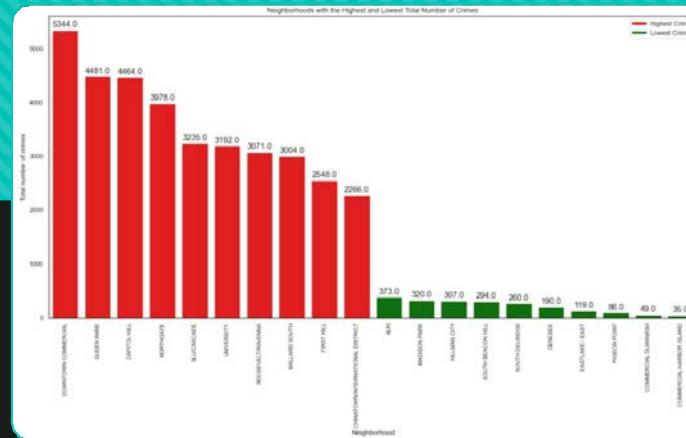
# Cluster 4

Seems to only contain commercial areas where not a lot of crime happens



Cluster 3 is the only cluster that didn't contain any districts in the top 10 highest crime district bar chart

So, chose **cluster 3** to explore further



# Now, we pick a cluster to explore

# Methodology



Gather datasets

Data cleaning

Exploratory  
data  
analysis of  
crime  
data

K-means  
clustering  
of  
crime data

Pick a cluster  
to  
explore

Explore venues in  
the cluster using  
Foursquare API

Pick the best  
neighborhood  
in  
the cluster



Combine neighborhood data  
with districts in Cluster 3

Call Foursquare API to get top  
100 venues within a radius of ~1  
mile for each neighborhood

Result = 1,519 venues

	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

## Get venue categories for Cluster 3

# Explore venue categories in Cluster 3

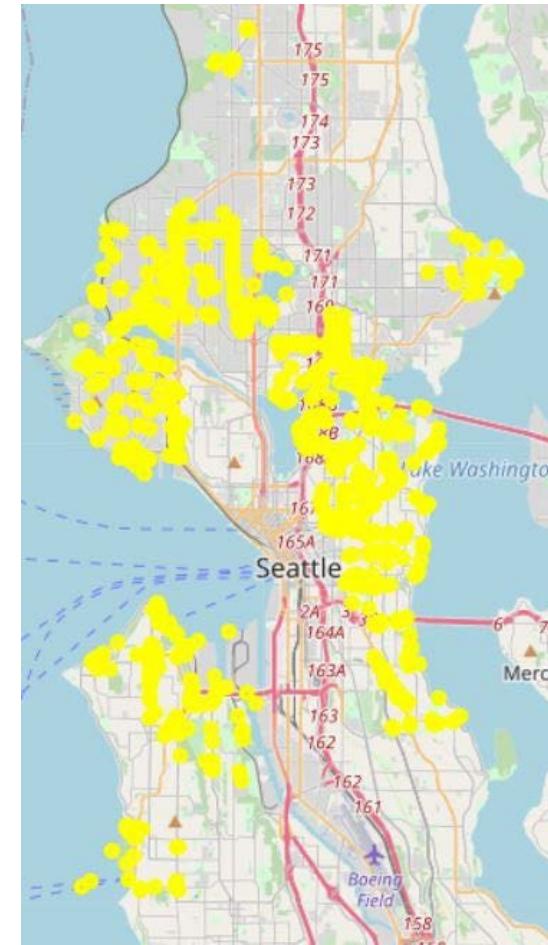
Not all neighborhoods returned 100 venues

Table that grouped by neighborhood to get count of venues

229 unique venue categories

Map of the venue locations

Neighborhood	Count of venues
ALKI POINT	88
BRIARCLIFF	37
BROADMOOR	81
FAUNTLEROY	30
LAKWOOD	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



# Top 10 most common venue categories for each Cluster 3 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

**Madrona Valley** picked as the best neighborhood for me because of its closeness to international cuisine, ie. Chinese and dumpling restaurants

Wallingford comes second (note: bubble tea shop as 10<sup>th</sup> most common venue)

Observations: coffee shops & parks as most common venue throughout = suburbs?

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	Pub	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	Cafe	Yoga Studio	Taco Place	Restaurant	Sushi Restaurant	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	Cafe	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

# Picking the best neighborhood

# Summary/Conclusion

- Safest cluster to live in: Cluster 3
- Best neighborhood in cluster 3: Aurora Valley
- Caveat: venues given by Foursquare are highly dependent on when the API was called so will yield different results each time
- Seattle districts also are not well-defined. Could be missing some districts in my analysis.

Thanks for viewing!

