Seattle

Crime Analysis 2008-2020

By Ziyi Zhao

Table of Content

01 **Introduction**

02 **Data Overview & Pre-processing**

03 **Data Exploration & Insights**

1 Crime Categories

2 Crime vs. Districts

3 Crime in Downtown

6 Downtown vs. Category & Hour

5 Hourly Crime

4 Monthly Crime

04 **Unsupervised Learning**

05 **Conclusion**

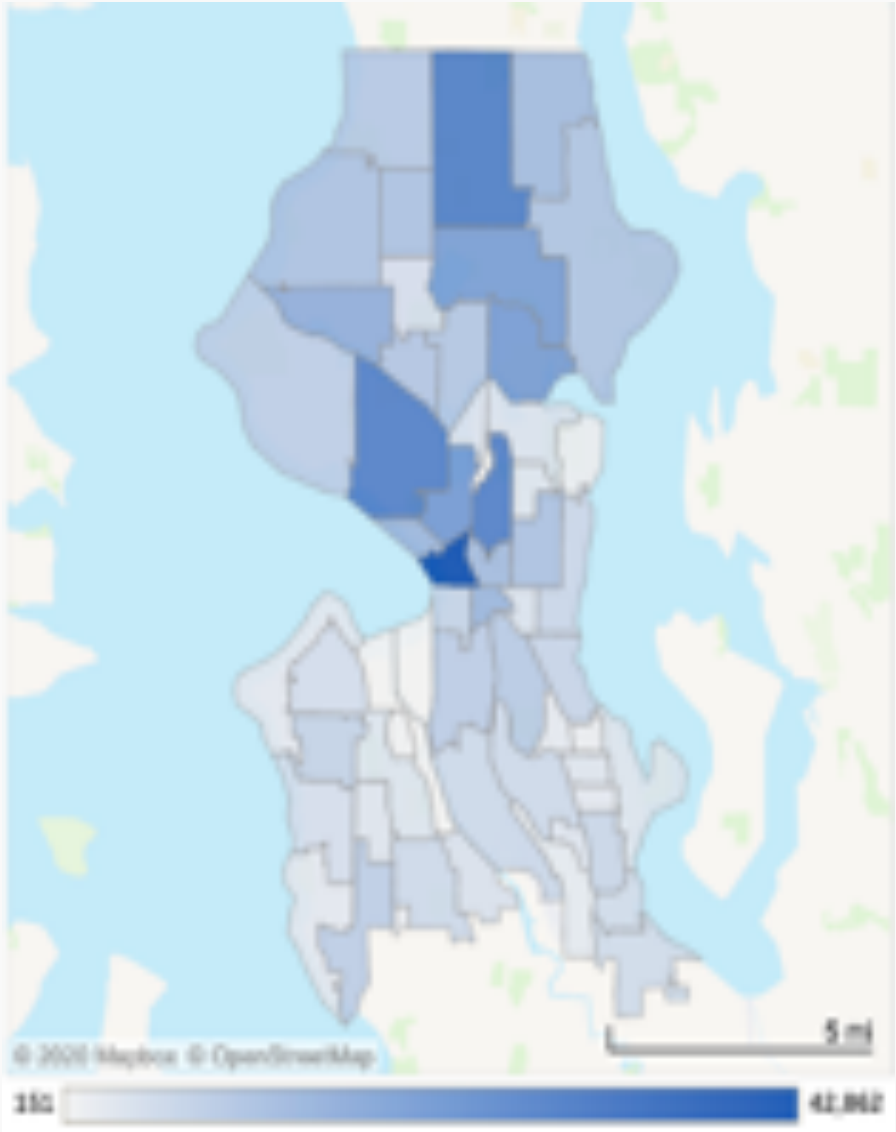
01 Introduction

Vision

Analyze the crime data of Seattle since 2008. Explore the types of crime, time, and region, and attempt to find potential insights to provide safety recommendations for residents

Specific Topics

- Explore the crimes of the three major categories and 30 parent categories
- Rank the crimes in different regions
- Compares the monthly and hourly patterns
- Crime in the downtown area vs categories and start time



02 Data Overview & Pre-processing

01

Dataset Introduction

Crime record for the city of Seattle from January 1st, 2008 to the end of April 2020 with 824220 rows and 17 columns

02

Data Dictionary

The data dictionary explained each features in this dataset in details, which can be found in Appendix.

03

Data Sampling

Randomly sampled 5% (41211 rows) from the original dataset, dropped 4 columns: Precinct, Sector, Beat and _c0. Rename the columns.

04

Missing Data

Dropped Offense End Datetime (19866 missing and useless for analysis), ignored 3.6% missing data of address.

03 Data Exploration & Insights

1 Crime Categories

2 Crime vs. Districts

3 Crime in Downtown

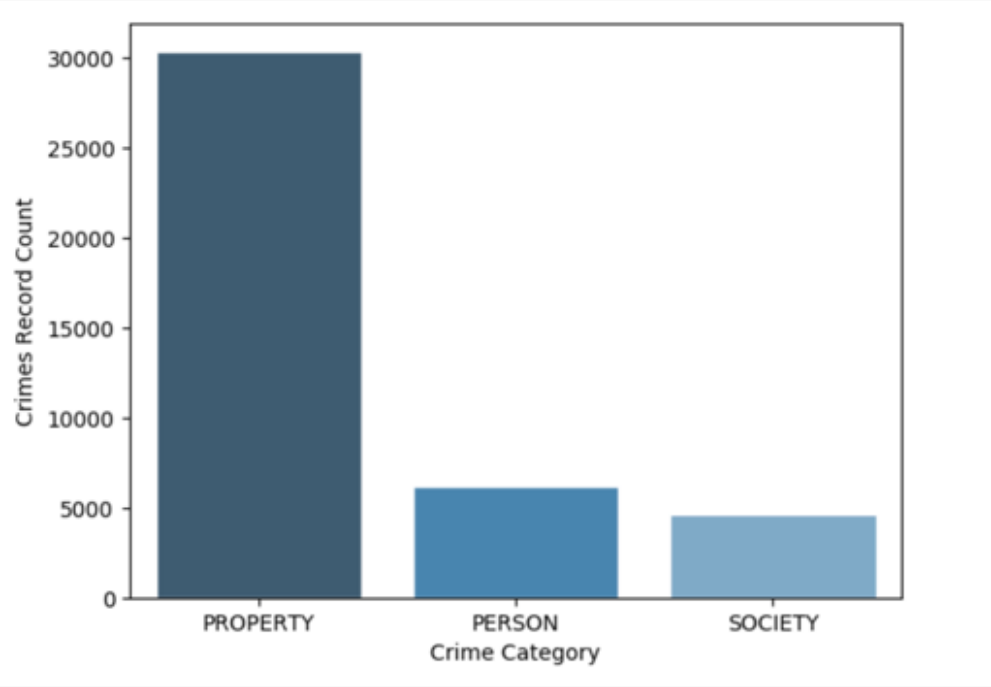
4 Monthly Crime

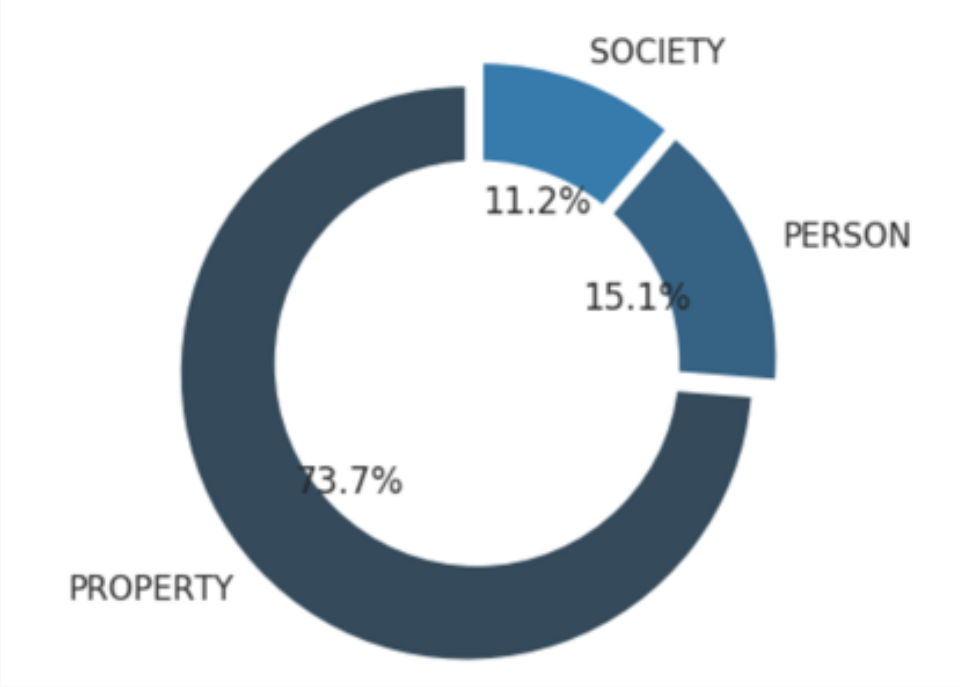
5 Hourly Crime

6 Downtown vs. Category & Hour

1-1 Crime Main Category

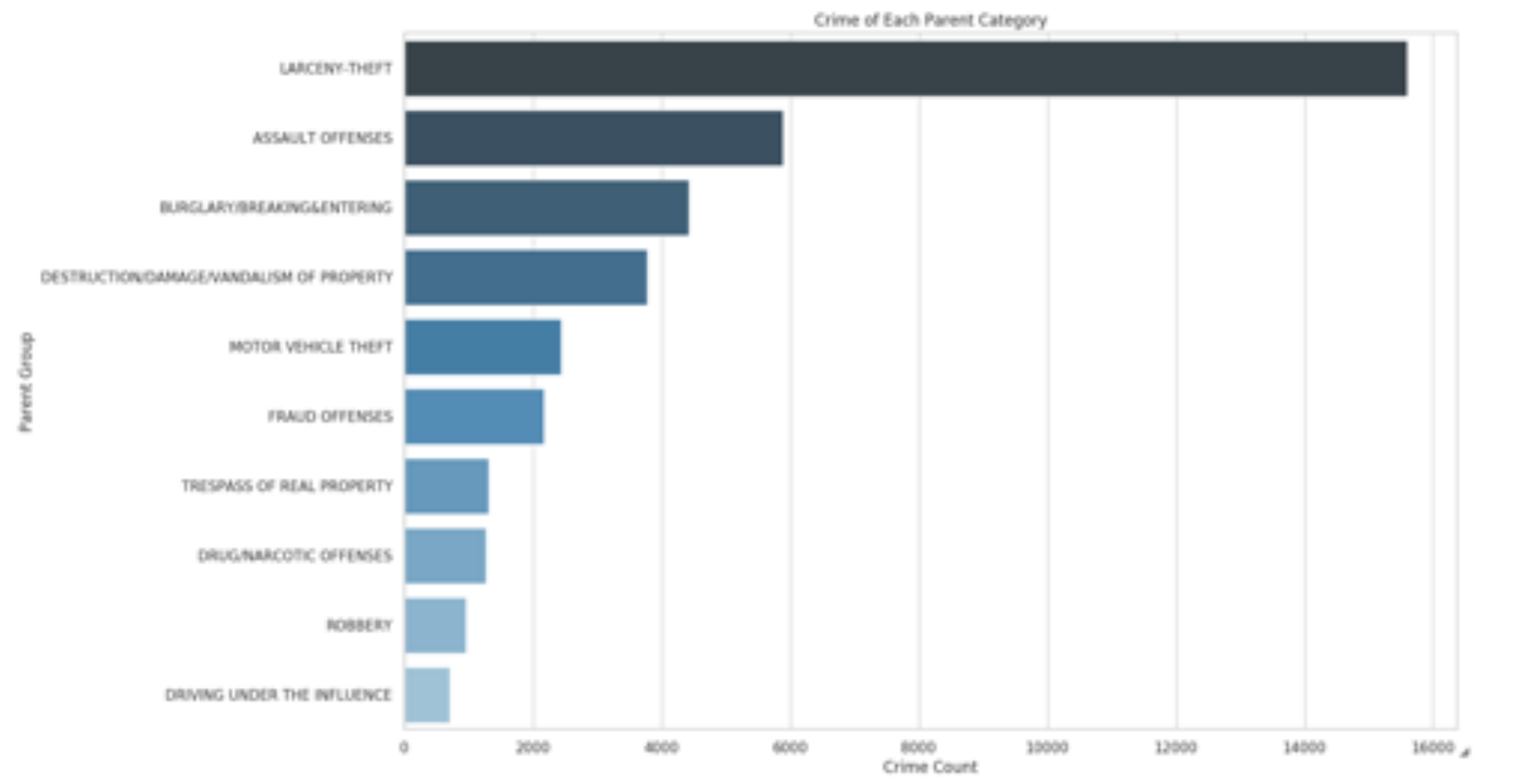
Category	count
PROPERTY	30384
PERSON	6228
SOCIETY	4599





Property (73.7%) > Person (15.1%) > Society (11.2%)

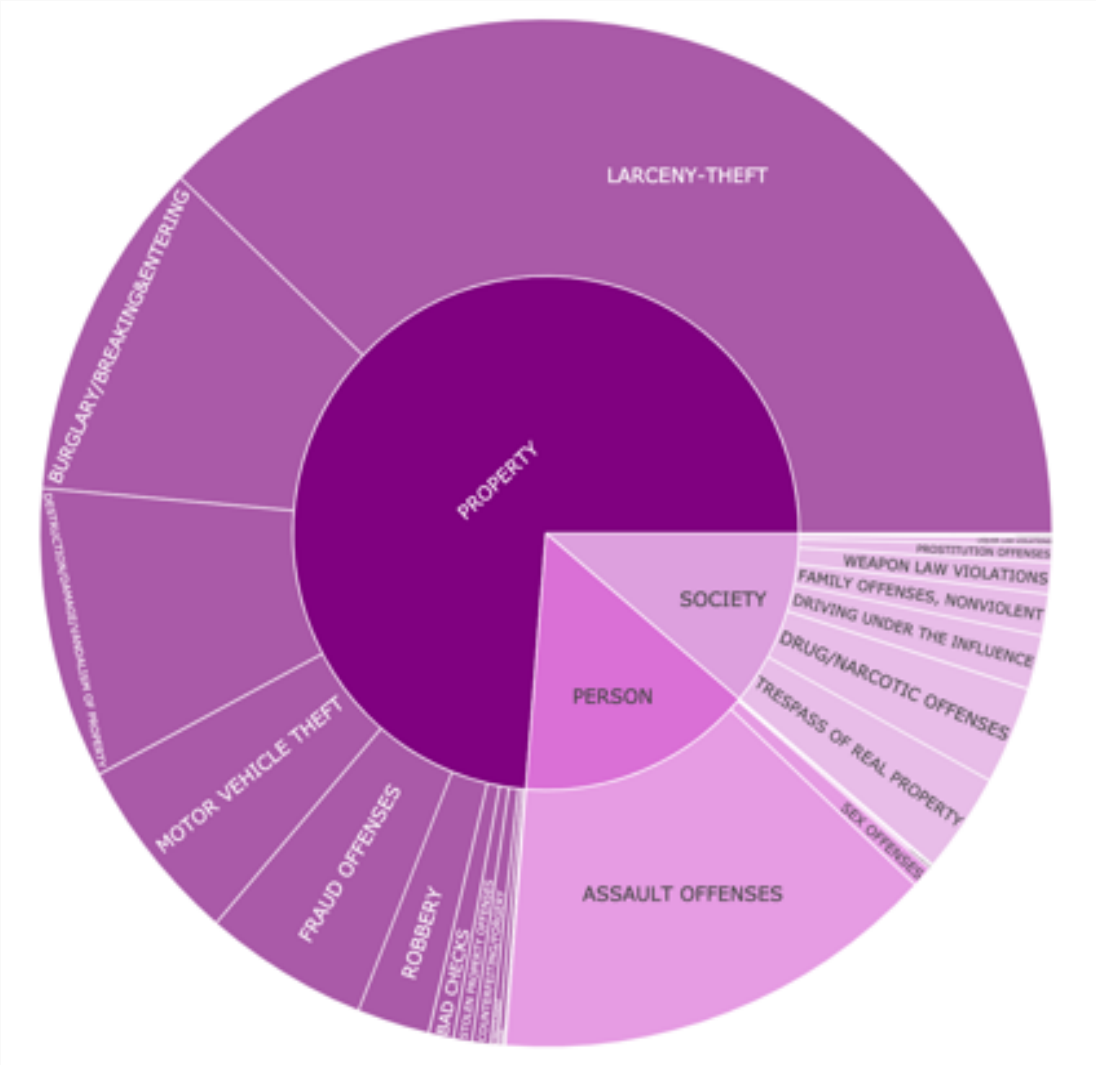
1-2 Crime Parent Group



Among 30 parent groups, Larceny Theft was the most common crime

1-3 Two Categories - Sunburst

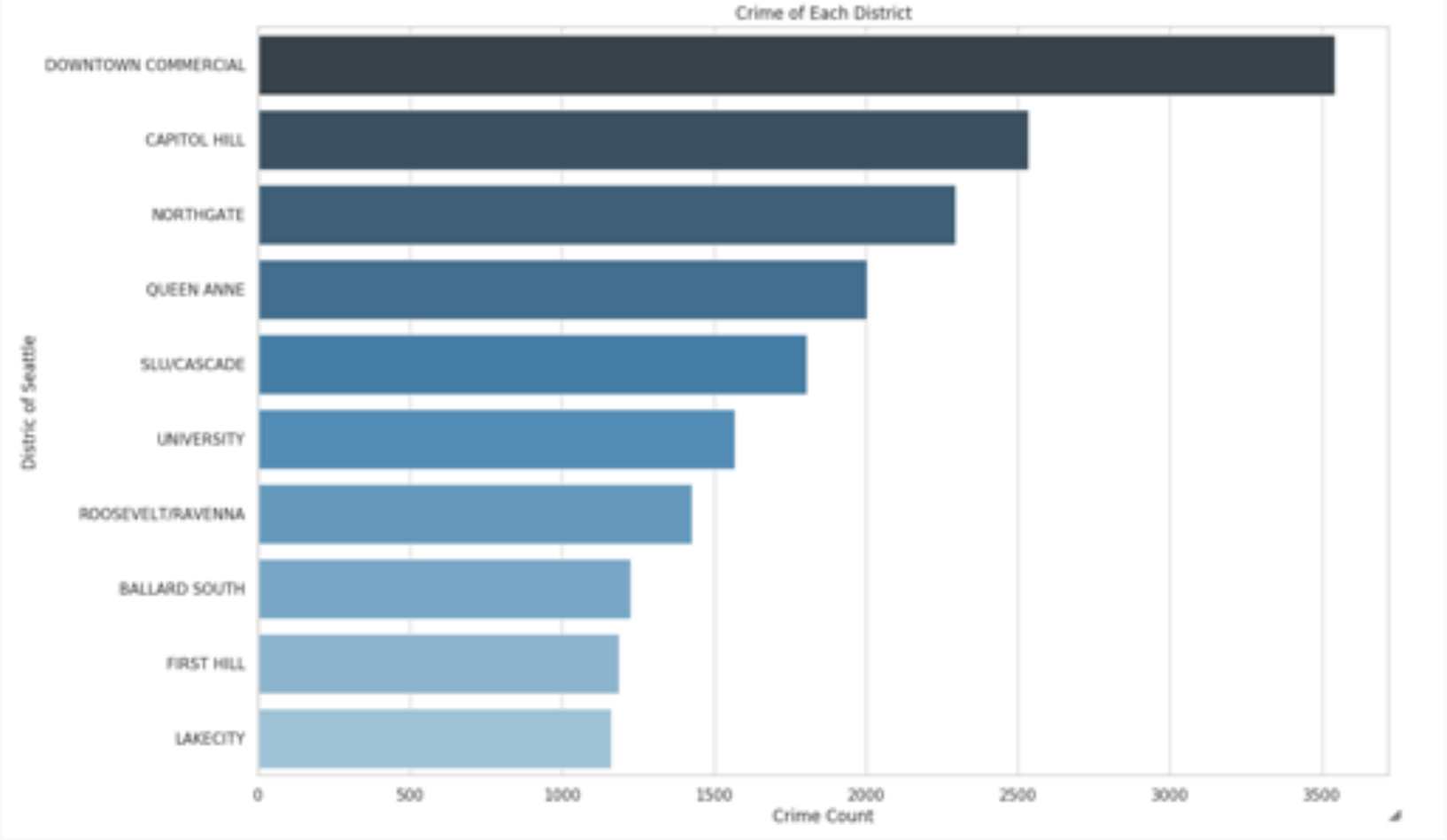
Bring two levels of categories together to see proportions of each main category and corresponding parent groups



2 Crime vs. Districts

Downtown Commercial is the area with the most cases

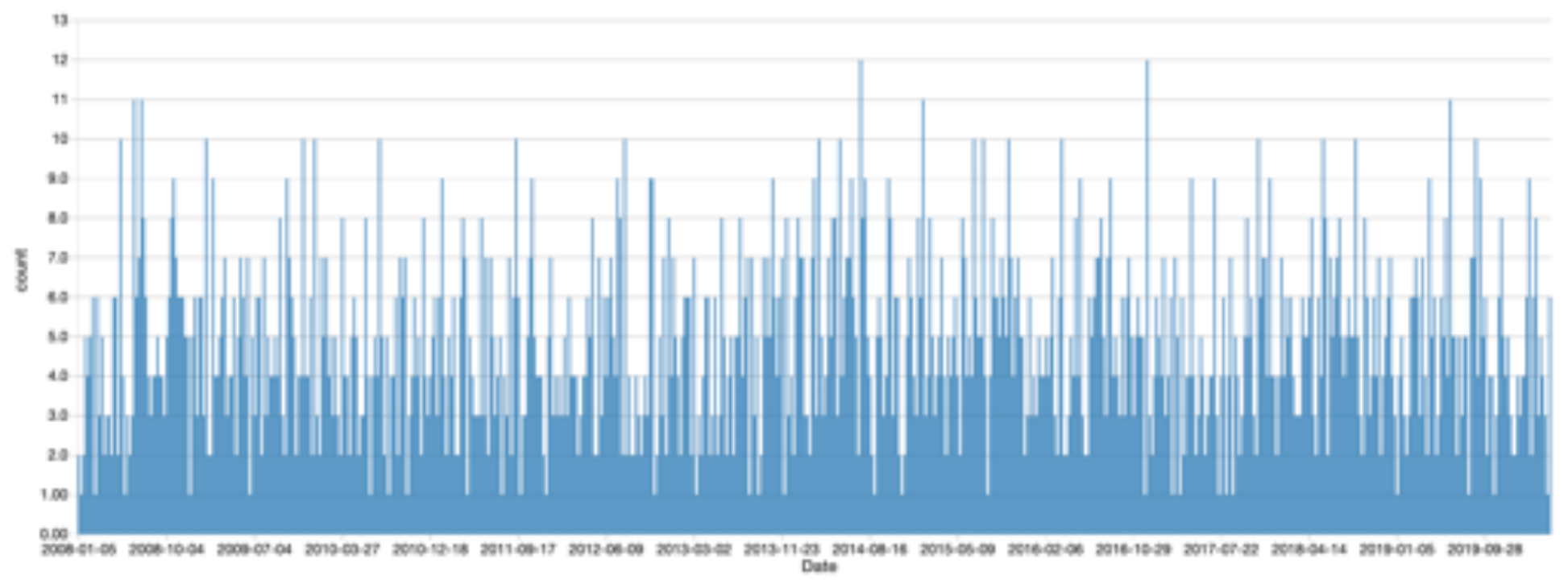
- population density in the city center is relatively large and the economy is relatively concentrated



3 Downtown Crimes on Sunday

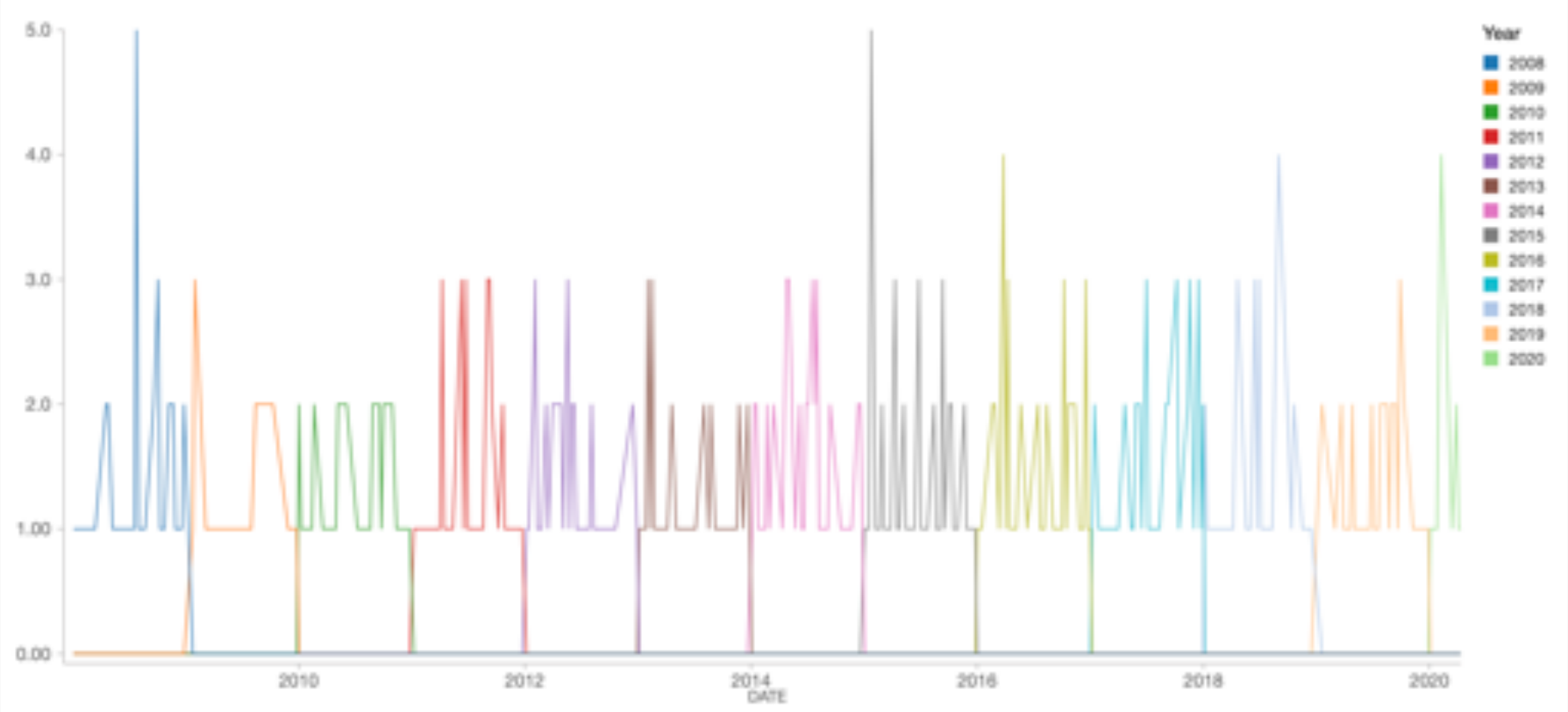
- Manually ranged downtown

avg(count)	max(count)	min(count)
4.794348508634223	12	1



Define the range of downtown within longitude (-122.285167, -122.385167) and latitude (47.558013,47.658013).

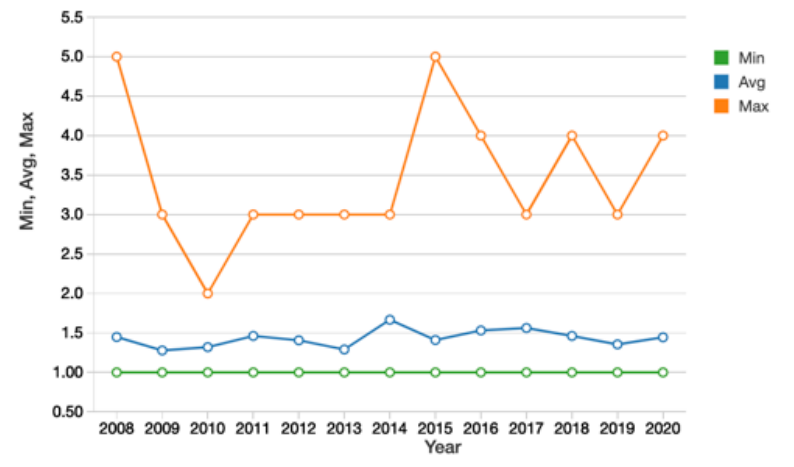
3 Downtown Crimes on Sunday



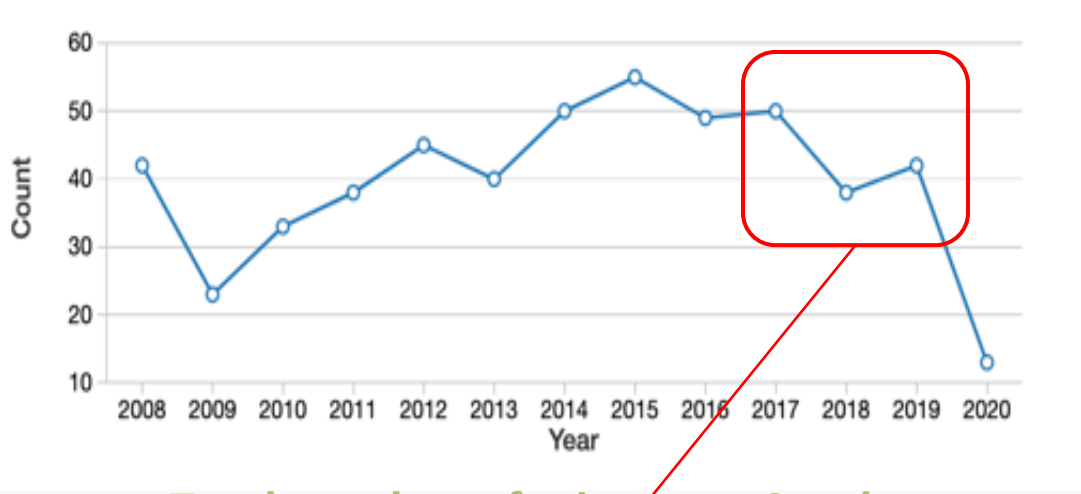
Regard the 'Downtown Commercial' District as the downtown

2008, 2015, 2016, and 2018 have more cases on Sundays.

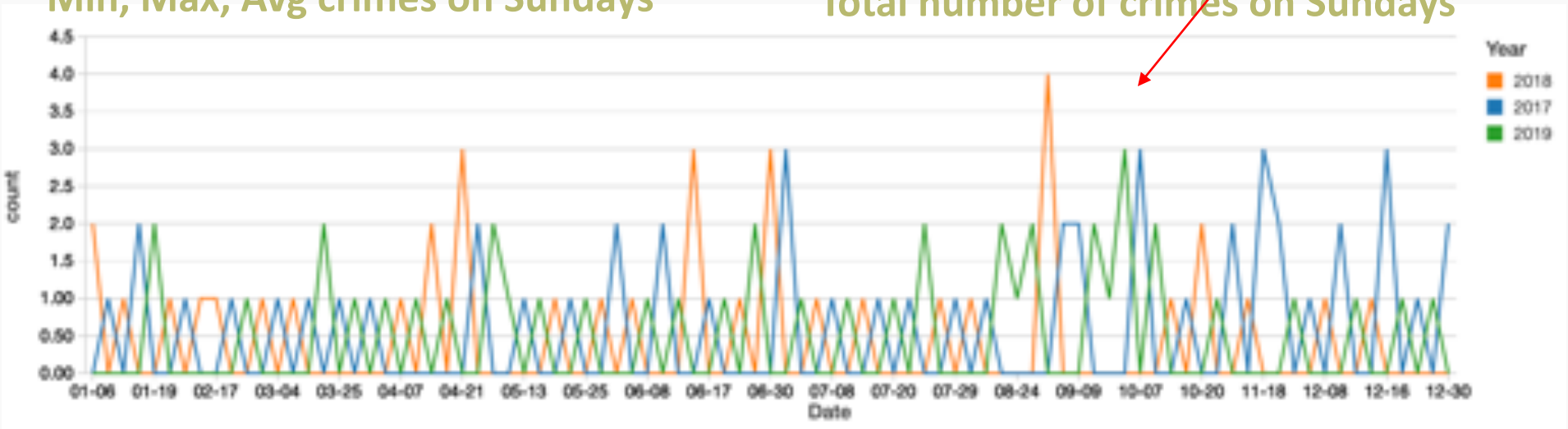
3 Downtown Crimes on Sunday



Min, Max, Avg crimes on Sundays

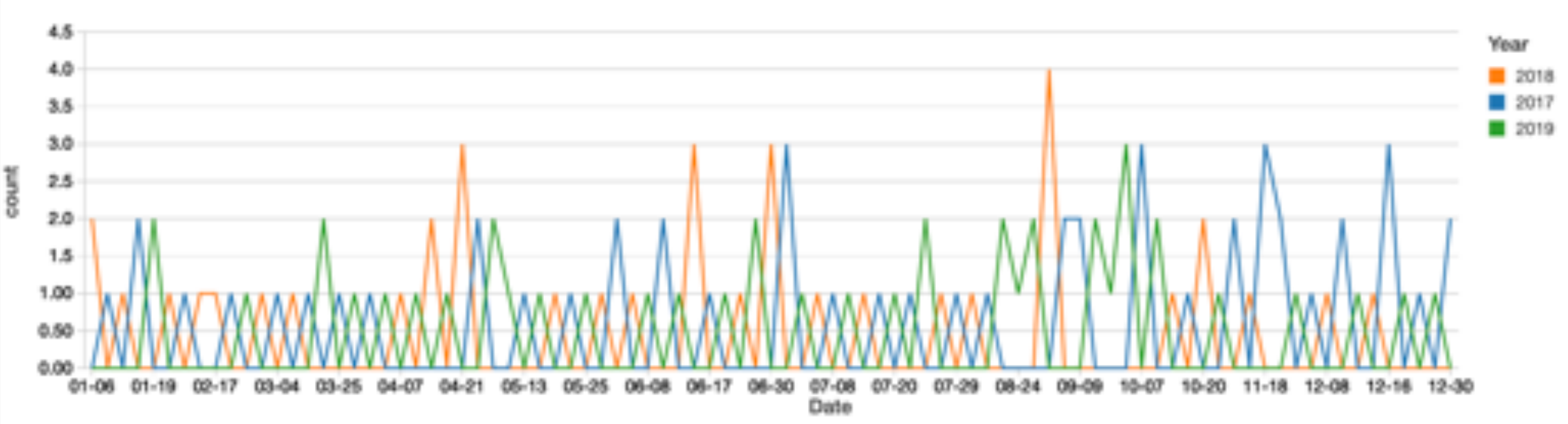


Total number of crimes on Sundays



Number of crimes on Sundays in 2017, 2018, 2019

3 Downtown Crimes on Sunday

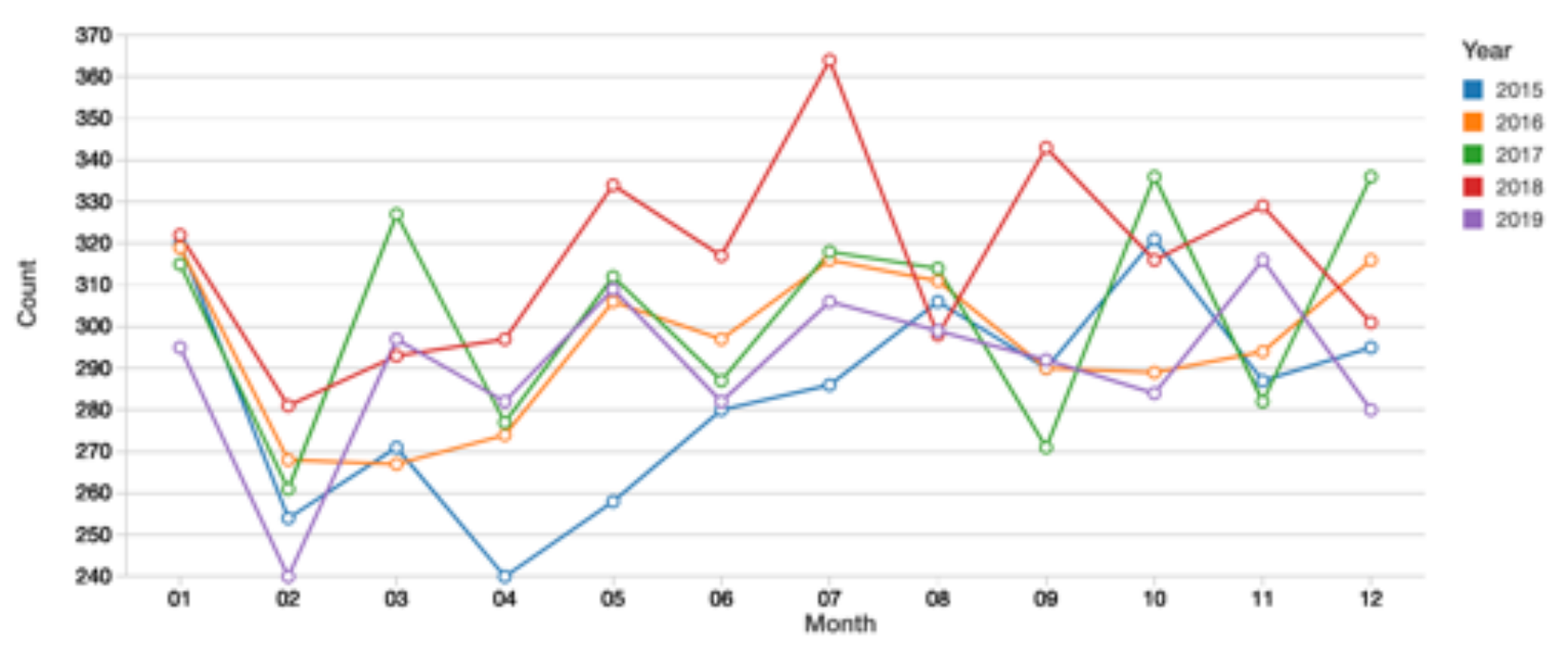


Number of crimes on Sundays in 2017, 2018, 2019

- The crime rate in the first half of the year was relatively stable
- But there may be greater fluctuations in the second half (especially in the August to October)
- In 2017, the November and December have relatively high crime cases.

4 Monthly Crime

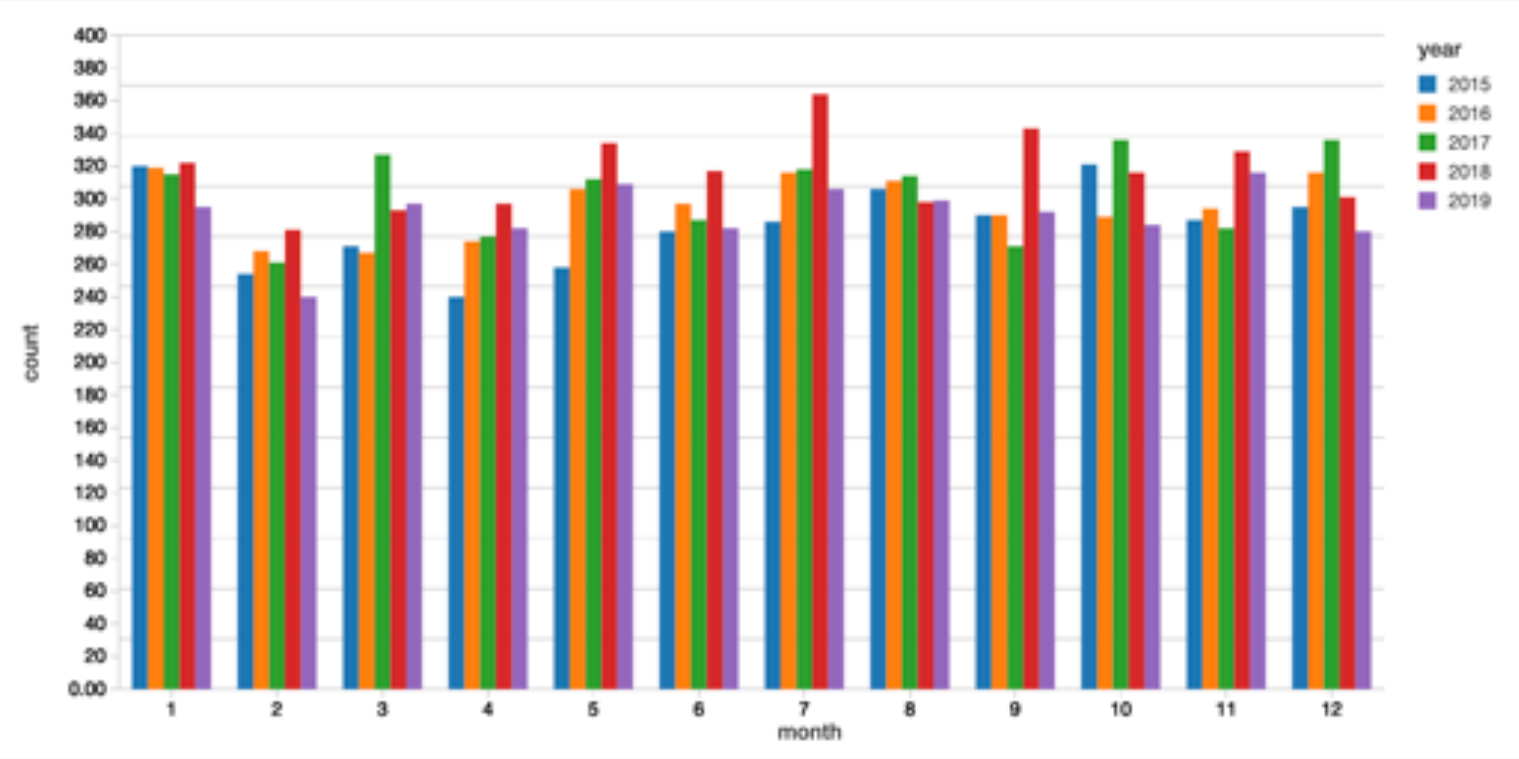
Monthly crimes 2015 - 2019



- The overall crime rate in July was relatively high, especially in 2018
- For 2018, July, May and September had higher crime rates
- For 2017, the crime rate fluctuated greatly, which was lower in January, September and November, but suddenly increased in February, October and December.

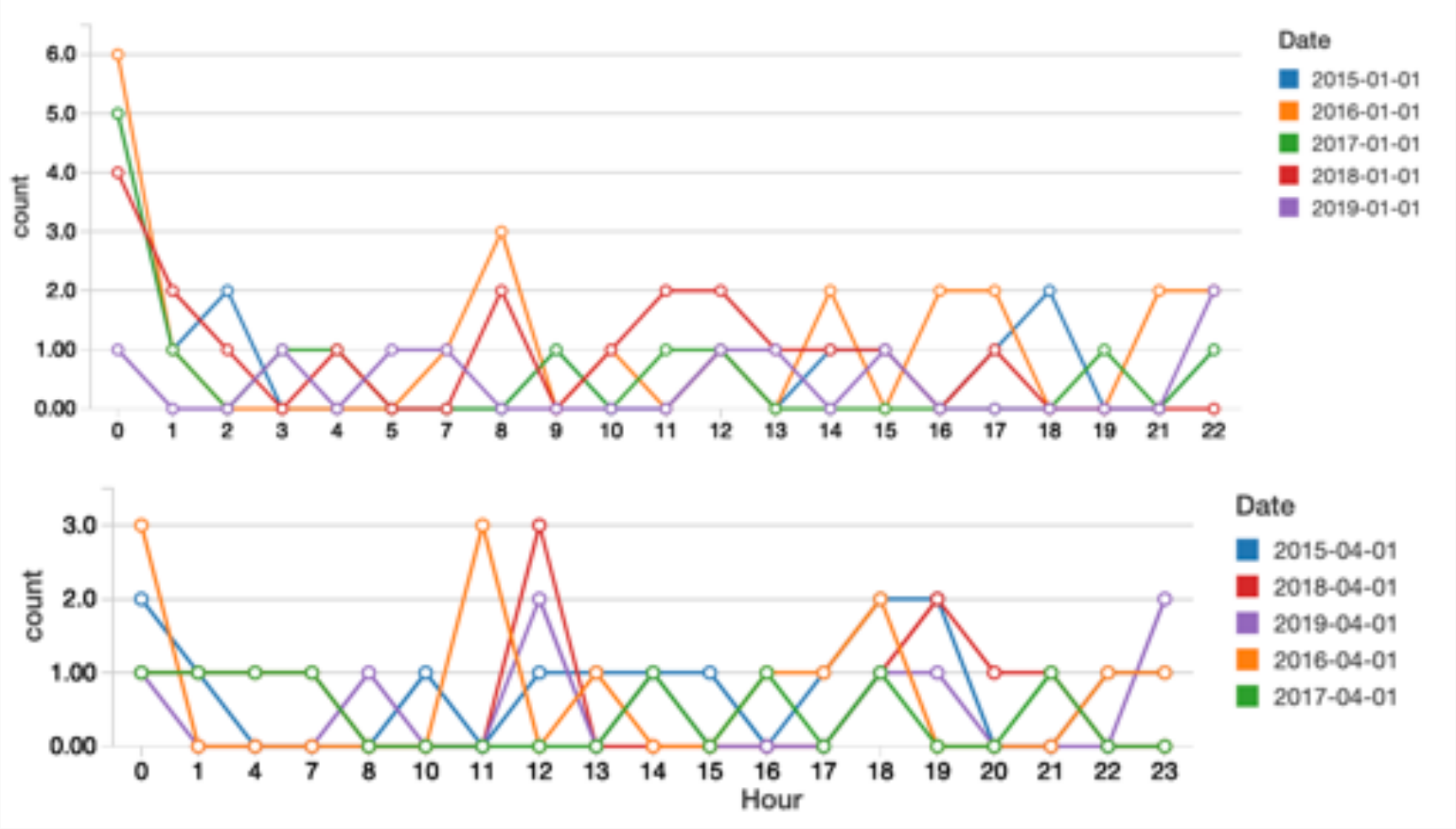
4 Monthly Crime

Monthly crimes 2015 - 2019

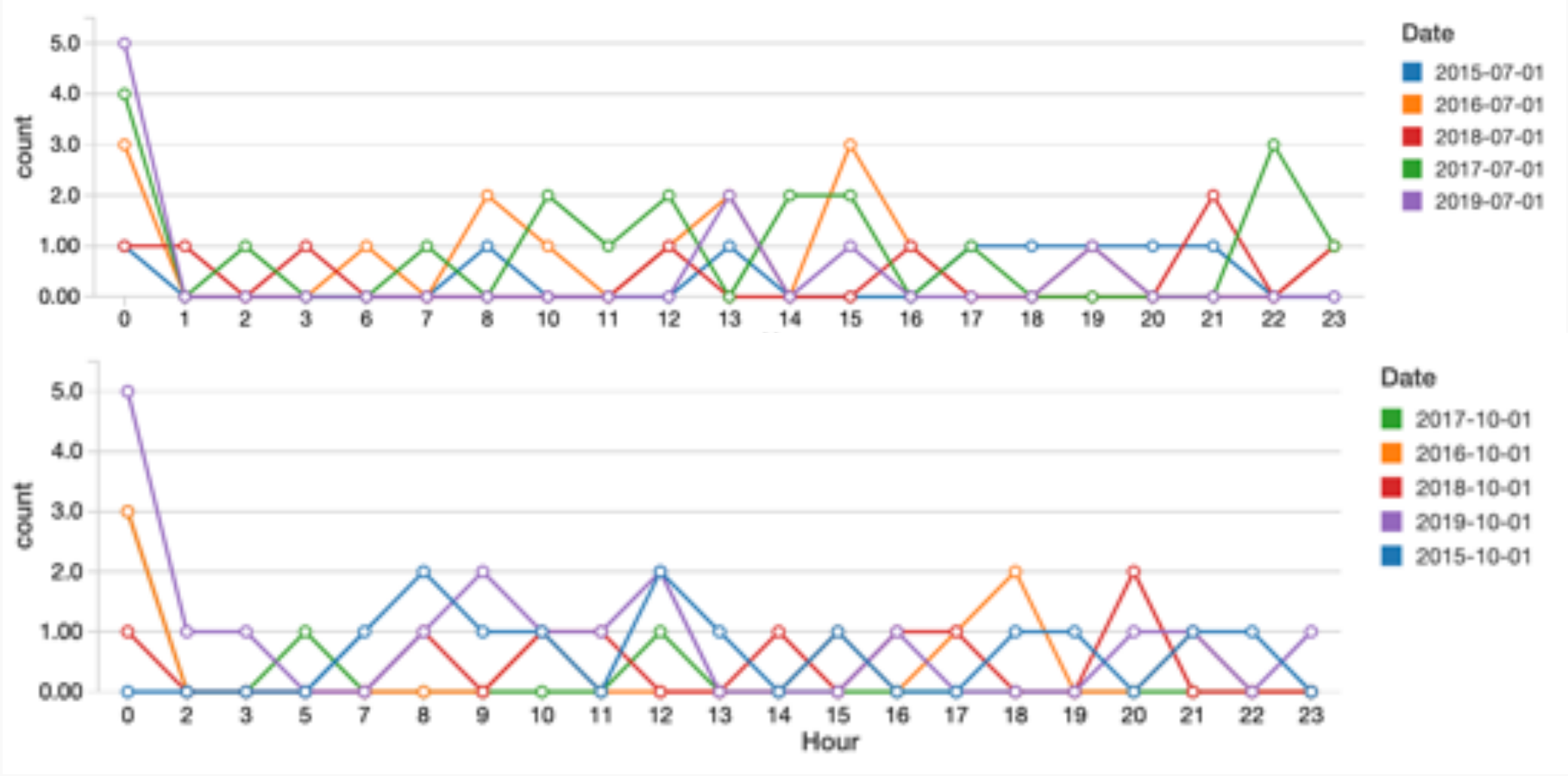


- The fluctuation of the crime rate in the second half of the year is smaller than that in the first half.
- In January and August, the crime rate in the same period in different years was not much different, but in other months it was quite different.

5 Hourly Crime - Jan 1st & April 1st



5 Hourly Crime - July 1st & October 1st



5 Hourly Crime

Findings

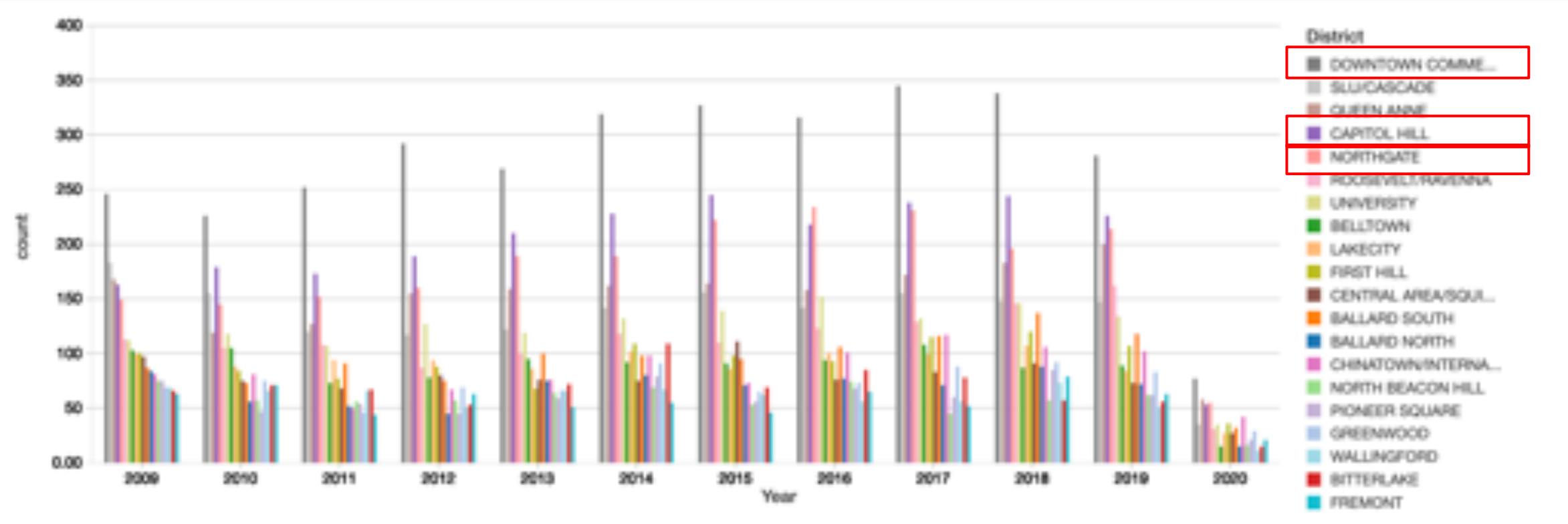
- Midnight is when Seattle has the highest crime rate
- Local theft cases are known to be the most, and property crimes including burglary and theft also frequently occur at midnight
- Afternoons in July had higher crimes, high temperatures can easily cause higher crime rates
- The average crime rate in April is less than that in January

Suggestions

- Residents should avoid go out at midnight
- Residents are advised to protect their property, such as locking doors and windows, and paying attention to anti-theft.

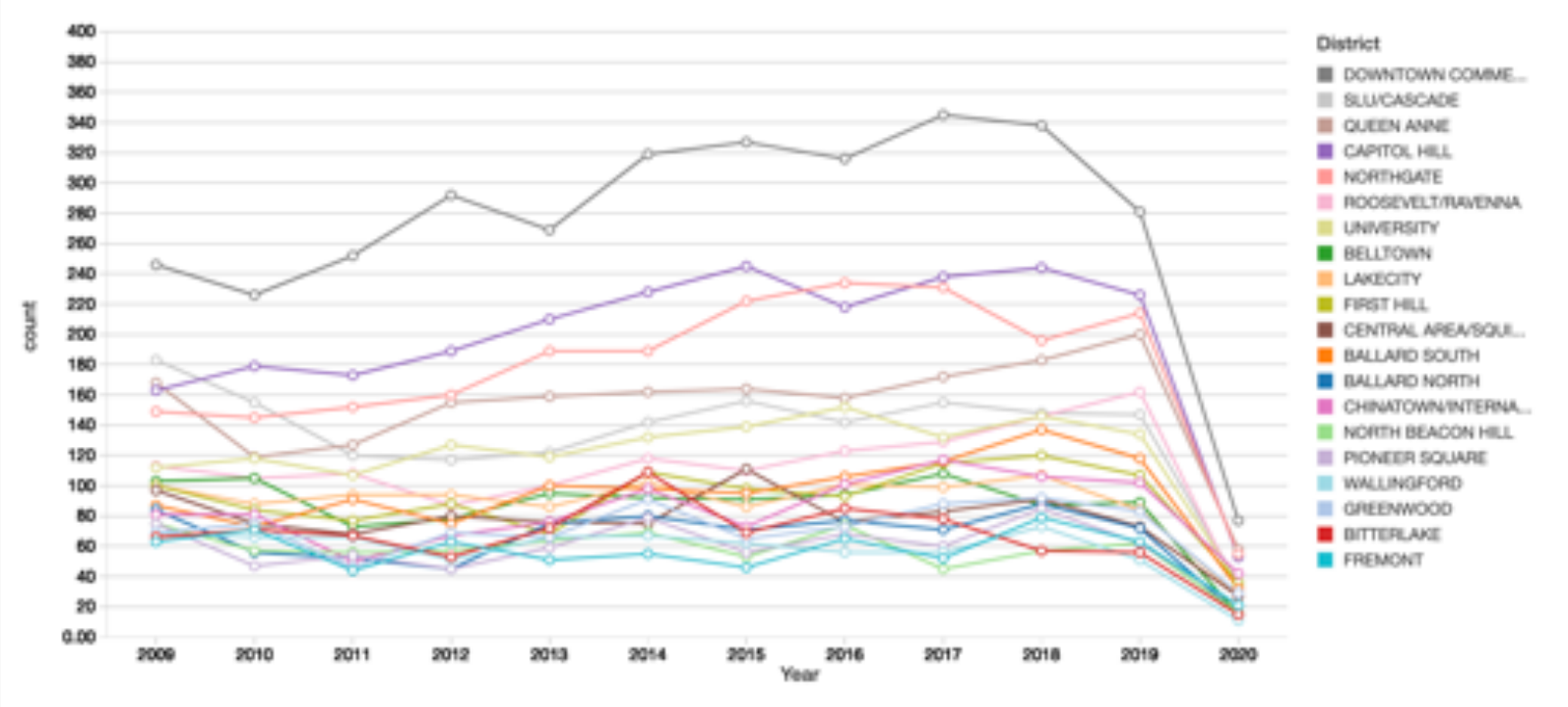
6 District vs. Time

- Three most dangerous areas



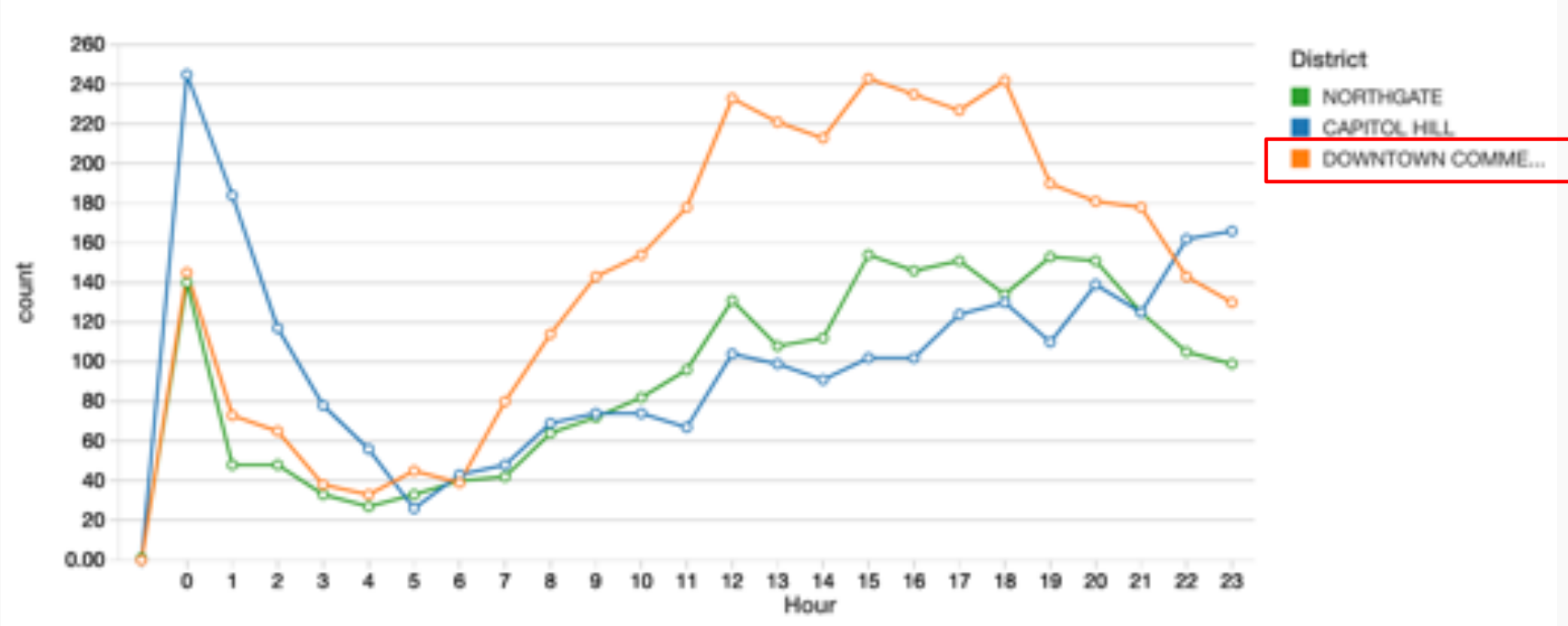
Define the areas with the highest annual crime cases as the most dangerous areas

6 District vs. Time



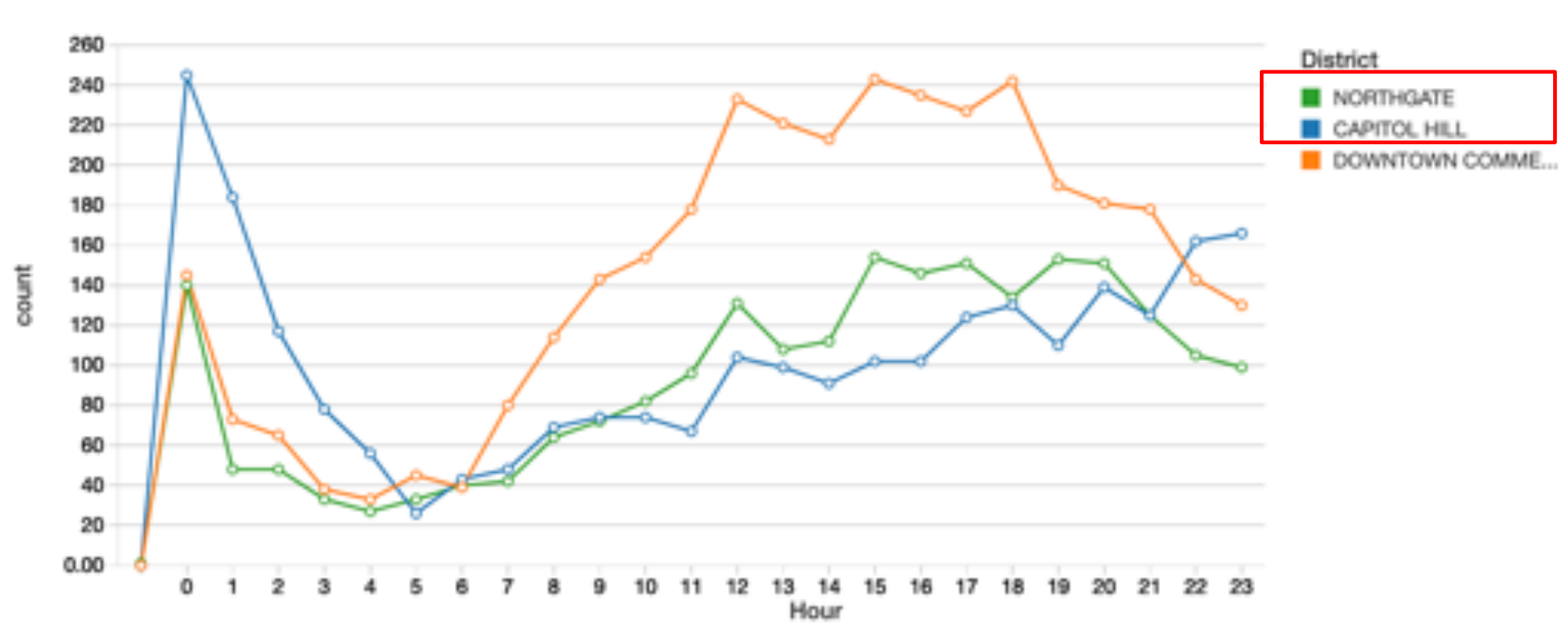
- Downtown Commercial with diverse population and business tends to provide opportunities for crimes.
- In most years except for 2016, the crime cases in Capitol hill was larger than that of Northgate.
- From 2008 to 2017, the overall crime rate showed an upward trend.
- The number of crimes in downtown increased year by year. Since 2018, this number has declined.

6 District vs. Time



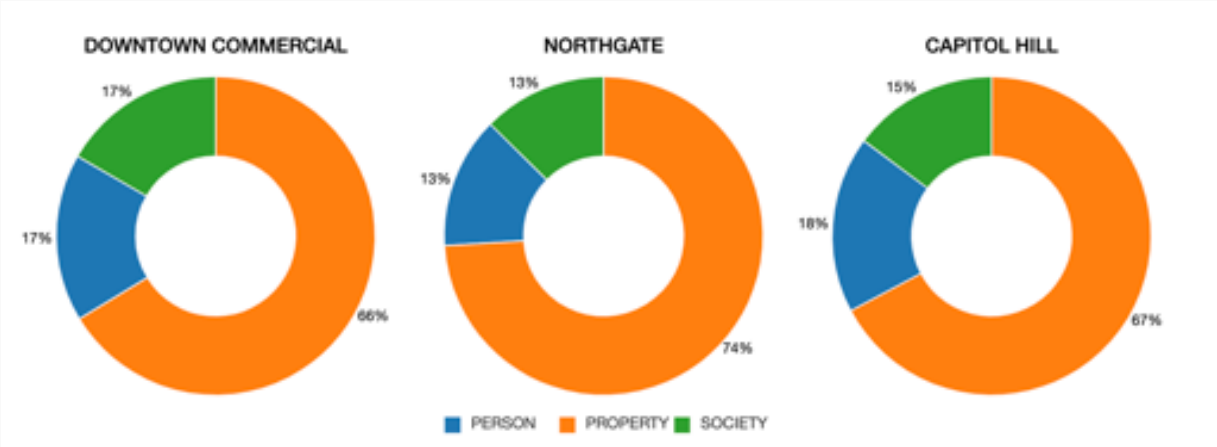
- Downtown Commercial: the crime rate continued to increase from 6 a.m. until it reached its peak around 12 p.m. and continued until about 6 p.m.
- After 6 p.m., the crime rate has declined, possibly because people in the city center come home from work and the population density has dropped rapidly.

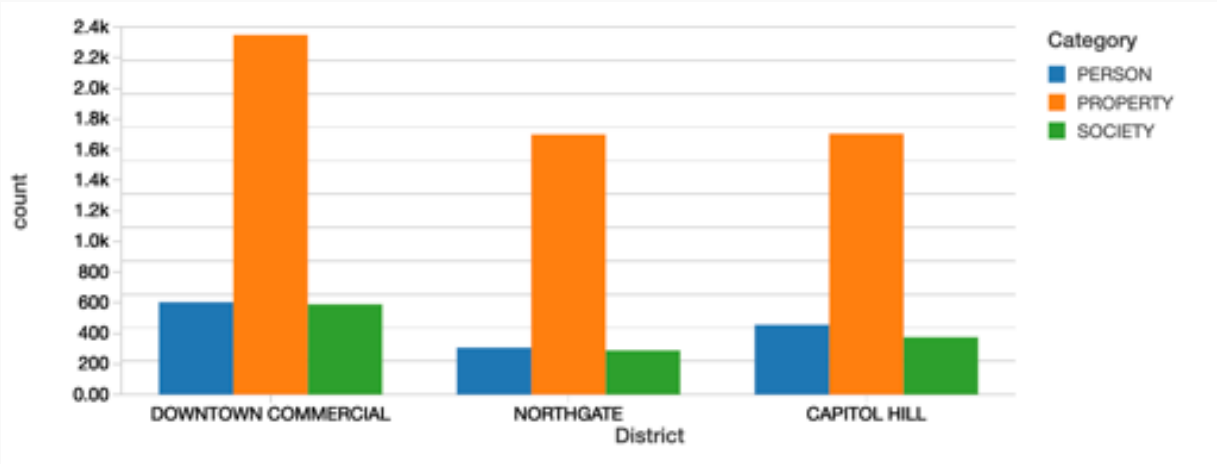
6 District vs. Time



- For Capitol Hill and Northgate, the number of crimes has gradually increased from 5am without significant decline.
- Capitol Hill is the seat of US government agencies with a very high crime rate at midnight.
- Northgate is named after and surrounds the first Northgate Mall, so Northgate has a higher crime rate as a commercial center.

6 District vs. Category





- Property crimes accounted for the largest share, with Northgate's property crime rate reaching 74%.
- In Northgate and Downtown Commercial, the number of person and society crime cases are almost the same.

04 Unsupervised Learning – K Means

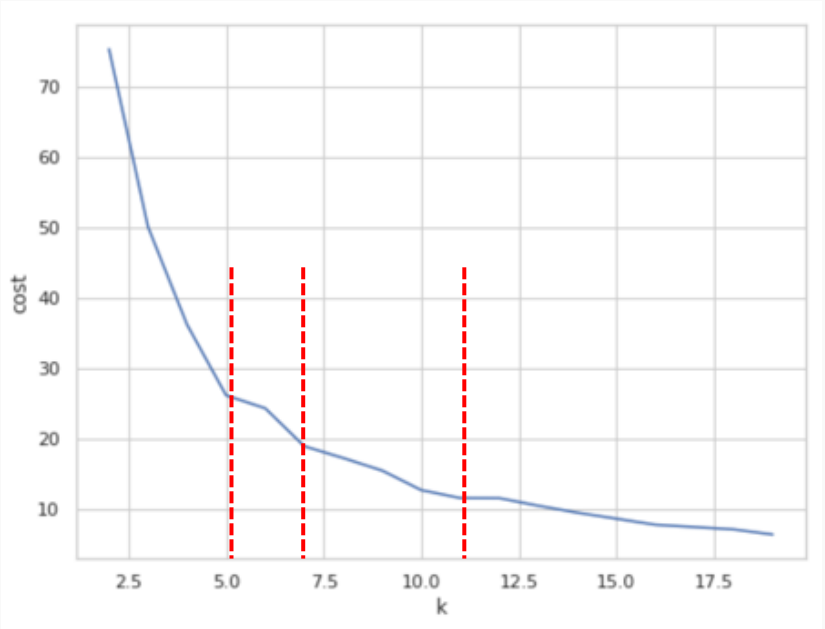
Apply Spark ML clustering algorithm to cluster the spatial data, then visualize the clustering results

▸ (1) Spark Jobs

Category	ParentGroup	OffenseName	District	Latitude	Longitude
SOCIETY	TRESPASS OF REAL ...	Trespass of Real ...	NORTHGATE	47.701792100000006	-122.344655369
PROPERTY	DESTRUCTION/DAMAG...	Destruction/Damag...	BELLTOWN	47.61497245	-122.349213984
PROPERTY	FRAUD OFFENSES	Wire Fraud	ALASKA JUNCTION	47.56897088	-122.375060686
SOCIETY	WEAPON LAW VIOLAT...	Weapon Law Violat...	DOWNTOWN COMMERCIAL	47.60241242	-122.331082199
SOCIETY	TRESPASS OF REAL ...	Trespass of Real ...	CENTRAL AREA/SQUI...	47.61702955	-122.311450327
PROPERTY	LARCENY-THEFT	Theft From Building	SLU/CASCADE	47.61421579	-122.341983607
PERSON	ASSAULT OFFENSES	Simple Assault	UNIVERSITY	47.66131587	-122.32025036799999
PROPERTY	LARCENY-THEFT	Theft From Motor ...	BELLTOWN	47.61183729	-122.34391211
PROPERTY	LARCENY-THEFT	Theft From Building	JUDKINS PARK/NORT...	47.59515308	-122.30966889
PERSON	ASSAULT OFFENSES	Simple Assault	DOWNTOWN COMMERCIAL	47.60255533	-122.33428107799999
PROPERTY	BURGLARY/BREAKING...	Burglary/Breaking...	ALASKA JUNCTION	47.56346901	-122.376189923
PROPERTY	MOTOR VEHICLE THEFT	Motor Vehicle Theft	NORTH ADMIRAL	47.5783244	-122.386659755
PROPERTY	COUNTERFEITING/FO...	Counterfeiting/Fo...	NORTHGATE	47.7086028	-122.32461515799999
PROPERTY	BURGLARY/BREAKING...	Burglary/Breaking...	FREMONT	47.66467101	-122.35001018
SOCIETY	DRUG/NARCOTIC OFF...	Drug/Narcotic Vio...	GREENWOOD	47.68875092	-122.344512557
PROPERTY	LARCENY-THEFT	Theft From Motor ...	NORTHGATE	47.70025321	-122.32578976
PROPERTY	DESTRUCTION/DAMAG...	Destruction/Damag...	CLAREMONT/RAINIER...	47.576184600000005	-122.29649406200001
PROPERTY	DESTRUCTION/DAMAG...	Destruction/Damag...	CAPITOL HILL	47.62310234	-122.313130839
PROPERTY	MOTOR VEHICLE THEFT	Motor Vehicle Theft	ROOSEVELT/RAVENNA	47.69037597	-122.318962253
PROPERTY	MOTOR VEHICLE THEFT	Motor Vehicle Theft	LAKECITY	47.70480711	-122.303604666

only showing top 20 rows

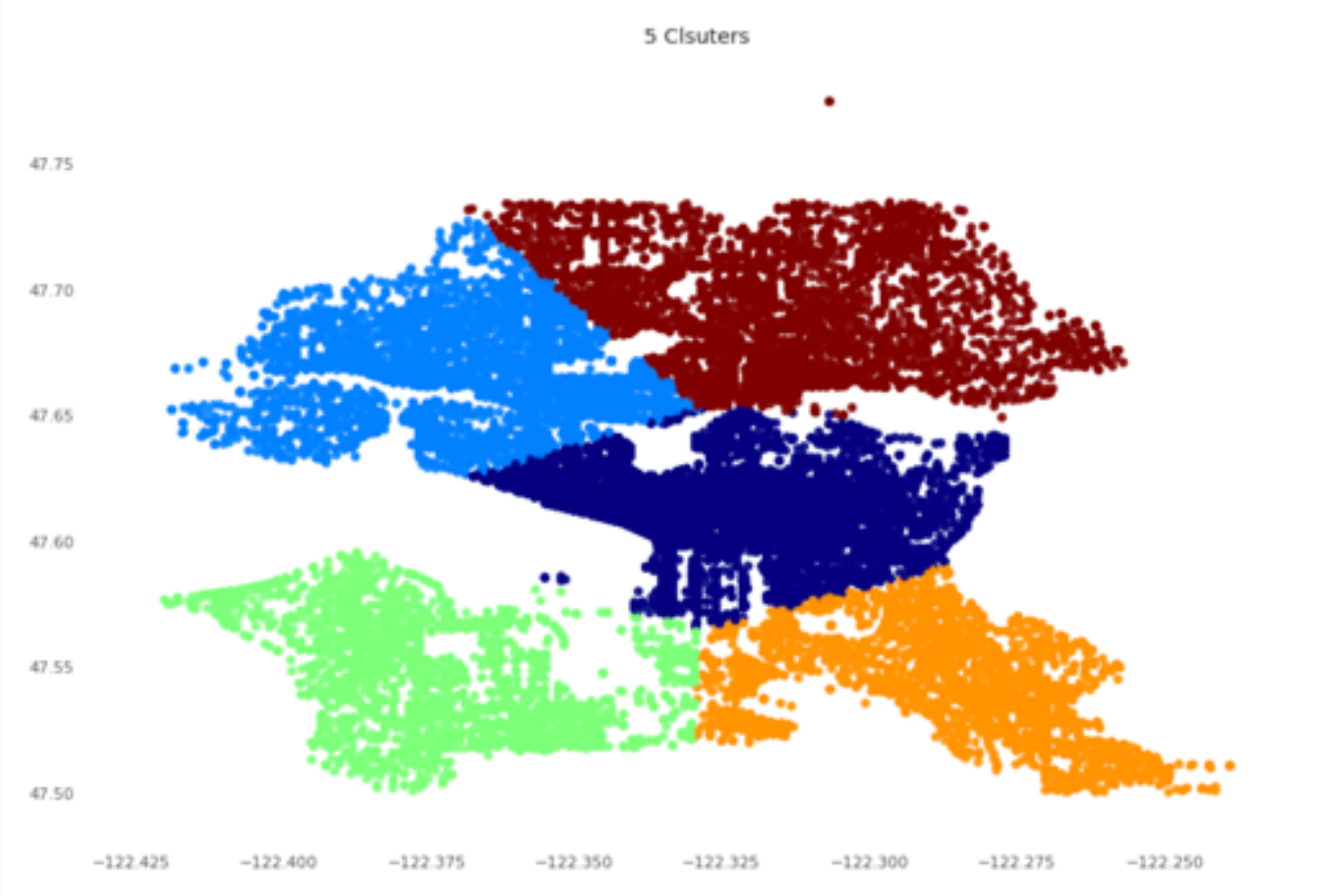
Subset the data and dropped outliers and null value



Post computational costs of different Ks to choose optimal K, try K = 5, K = 7, K = 11

4-1 Unsupervised Learning – K = 5

When K = 5, the results of K Means model:



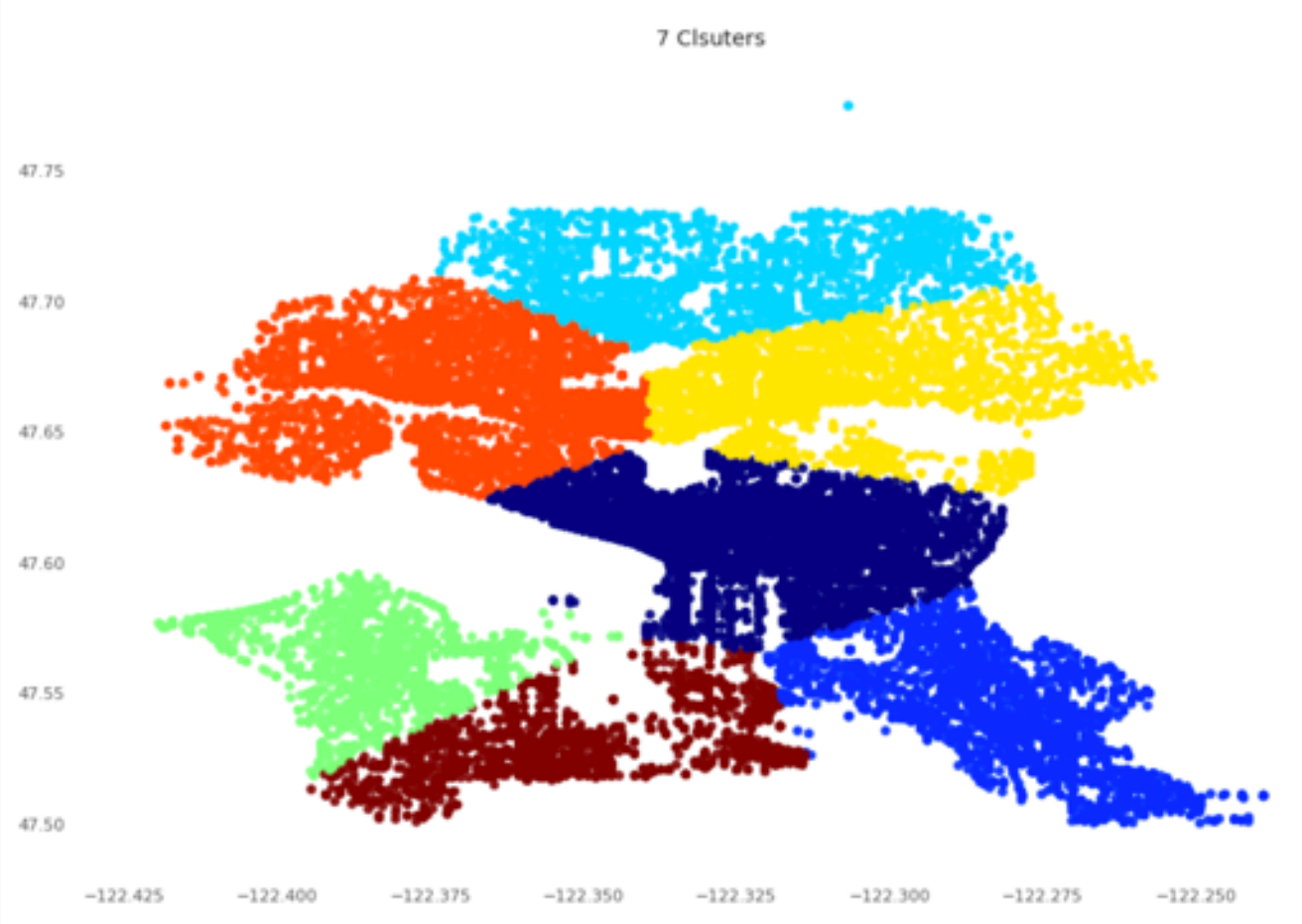

```
Cluster Centers:
[  47.61093564 -122.32778403]
[  47.6680053  -122.36781257]
[  47.54587406 -122.37121397]
[  47.54407398 -122.28748611]
[  47.6933422  -122.31773998]
```

The cluster centers when **K = 5**
Silhouette with squared Euclidean distance = **0.676685081175843**

Note: The silhouette value is a measure of how similar an object is to its own cluster (cohesion) compared to other clusters (separation). The silhouette ranges from −1 to +1, where a high value indicates that the object is well matched to its own cluster and poorly matched to neighboring clusters.

4-2 Unsupervised Learning – K = 7

When K = 7, the results of K Means model:

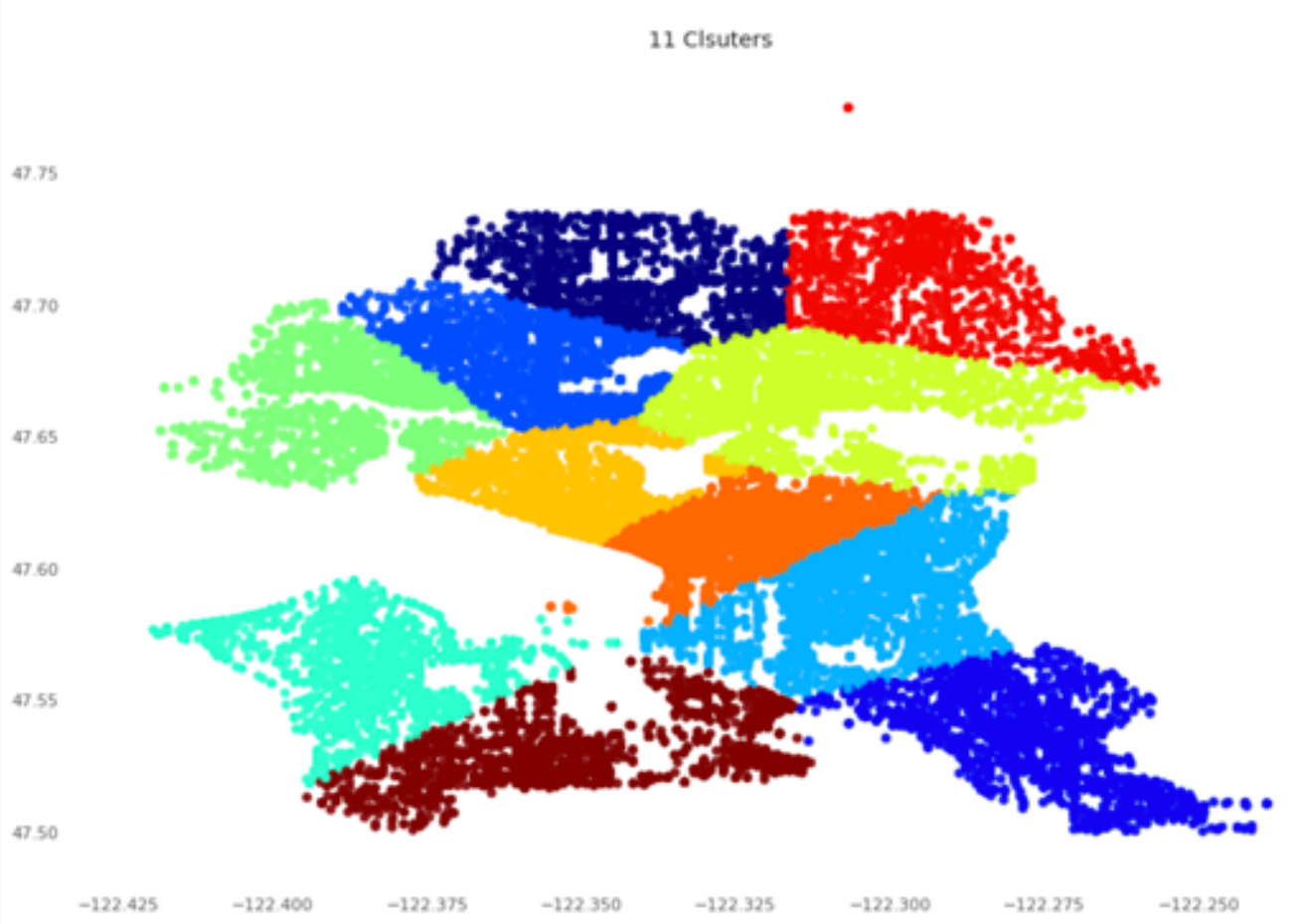



```
Cluster Centers:
[  47.60974041 -122.32821285]
[  47.54476261 -122.28224946]
[  47.7116113  -122.32847682]
[  47.56198787 -122.38382295]
[  47.66642545 -122.30928742]
[  47.6654526  -122.36980999]
[  47.53098121 -122.3499541 ]
```

The cluster centers when **K = 7**
Silhouette with squared Euclidean distance = **0.6725548253058243**

4-3 Unsupervised Learning – K = 11

When K = 11, the results of K Means model:

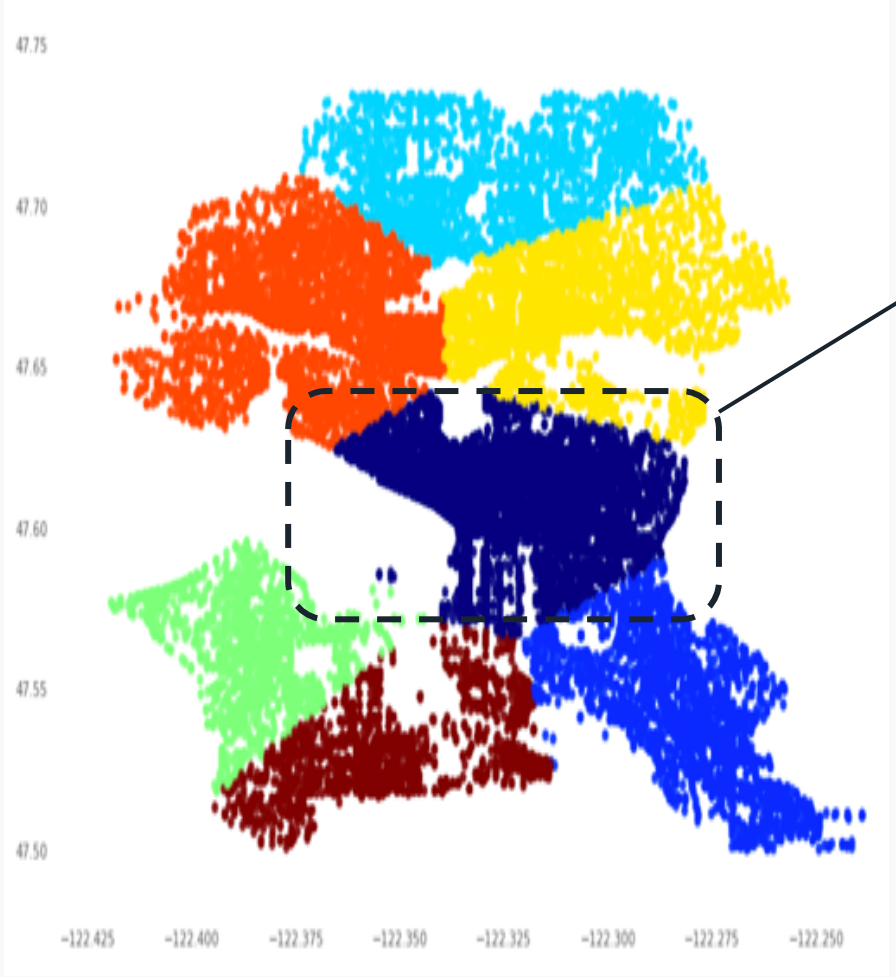



```
Cluster Centers:
[  47.71103639 -122.33895014]
[  47.53644045 -122.27666649]
[  47.67971012 -122.35872648]
[  47.58897078 -122.30601463]
[  47.56198655 -122.38380804]
[  47.66272852 -122.38605028]
[  47.66395764 -122.31265606]
[  47.62781669 -122.34913821]
[  47.61025354 -122.32868051]
[  47.70956478 -122.29431195]
[  47.53028978 -122.35014026]
```

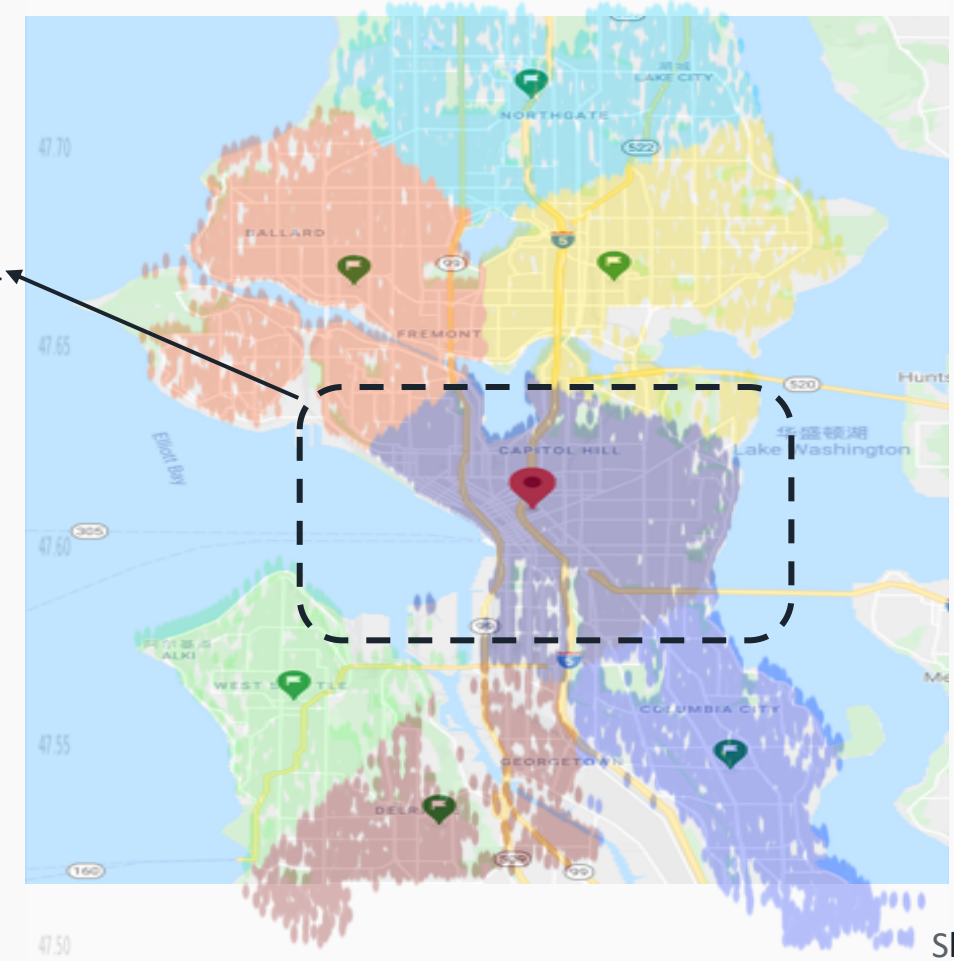
The cluster centers when **K = 11**
Silhouette with squared Euclidean distance = **0.5547070083978424**

4-4 Clustering results vs Seattle map

Finally choose model with K = 7 for after tradeoff of Silhouette and complexity



Most dangerous area



Conclusion

Objective

- Studied on the dataset of crime in Seattle from 2008 to 2020 in six aspects to seek crime patterns in Seattle and give recommendations for residents, visitor, and police on how to avoid crimes and stay safe.

Techniques

- This process is conducted by Apache Spark, using multiple tools such as Spark SQL, Dataframe, OLAP, visualization, etc. after essential data sampling and processing.

Conclusion

Findings

- The crime rate in the whole Seattle increased until 2017 and then tended to decline.
- Among three main categories of crimes, the property crime accounted for 73.7%.
- Among 30 parent groups of crimes, the larceny theft had the most cases.
- Downtown commercial, Capitol Hill, and Northgate were the most dangerous areas.
- Crime rates on Sundays were high in 2008 and 2015.
- April had relatively low crime rate and July had relatively high crime rate.
- Midnight was the most dangerous time.
- Afternoons in summer tended to have higher crime cases.

Recommendations

- Always pay attention to property safety, especially in the most dangerous areas.
- try to minimize living or staying in these areas and avoid going out at midnight.
- As crime rates tend to rise in a day, people should be more vigilant about crime in the afternoon.
- The police force coverage should pay more attention to the three most dangerous regions, help tourists identify dangerous districts and hours to visit in Seattle.

Appendix – Data Dictionary

Column Name	Description	Type
Report Number	Primary key/UID for the overall report. One report can con...	Plain Text
Offense ID	Distinct identifier to denote when there are multiple offens...	Plain Text
Offense Start DateTime	Start date and time the offense(s) occurred.	Plain Text
Offense End DateTime	End date and time the offense(s) occurred, when applicable.	Plain Text
Report DateTime	Date and time the offense(s) was reported. (Can differ fro...	Plain Text
Group A B	Corresponding offense group.	Plain Text
Crime Against Category	Corresponding offense crime against category.	Plain Text
Offense Parent Group	Offense_Parent_Group	Plain Text
Offense	Corresponding offense.	Plain Text
Offense Code	Corresponding offense code.	Plain Text
Precinct	Designated police precinct boundary where offense(s) occ...	Plain Text
Sector	Designated police sector boundary where offense(s) occurr...	Plain Text
Beat	Designated police sector boundary where offense(s) occurr...	Plain Text
MCPP	Designated Micro-Community Policing Plans (MCPP) bound...	Plain Text
100 Block Address	Offense(s) address location blurred to the one hundred blo...	Plain Text
Longitude	Offense(s) spatial coordinate blurred to the one hundred b...	Number
Latitude	Offense(s) spatial coordinate blurred to the one hundred b...	Number

References

Dataset: SPD Crime Data: 2008-Present

https://data.seattle.gov/Public-Safety/SPD-Crime-Data-2008-Present/tazs-3rd5

Jagannathan., M (2019) High temperatures can lead to more violent crime, study finds

https://nypost.com/2019/06/18/high-temperatures-can-lead-to-more-violent-crime-study-finds/

Thank You

For The Attention