

# Coursera Capstone Project: IBM Data Science Professional Certificate

Ahmad M. Shapiro

[ahmad.shapiro@alexu.edu.eg](mailto:ahmad.shapiro@alexu.edu.eg)

Alexandria University, Egypt

## Analyzing San-Francisco's 2016 Crime Data

### **1-Introduction: -**

*“Safety and security don't just happen; they are the result of collective consensus and public investment. We owe our children, the most vulnerable citizens in our society, a life free of violence and fear.” - Nelson Mandela*

Safety is one of the most important aspects of life each person should have, and we will do our best today to assure and enhance the Police system in San-Francisco , California , USA by analyzing the Crime Data in 2016 looking for any patterns that could help us to be more safe.

### **2-Business Problem: -**

Government of San-Francisco are doing their best to make the city safer, so we will help them understand the patterns of what's actually happening and what are the most places that has crime and at exactly what time, and maybe the reason for those crimes.

According to Edwin M., SF Mayor, the budget for the City for FY 2016-17 is \$9.6 billion dollars, half of them is for the Public Health, Police, Fire, and Recreation and Parks.

So, the police department budget may be around \$2 billion dollar in 2016. Can we make a better use of this budget to attain more safety? This our problem that we will try to solve talking about an example only.

### 3-Data: -

The data for this project has been retrieved and processed through multiple sources, giving careful considerations to the accuracy of the methods used.

#### 3.1-San-Francisco 2016 Crime Data

Csv file of crime data from :- <https://www.kaggle.com/c/sf-crime/data>

IncidentNum	Category	Descript	DayOfWeek	Date	Time	PdDistrict	Resolution	Address	X	Y	Location	PdId	
0	120058272	WEAPON LAWS	POSS OF PROHIBITED WEAPON	Friday	01/29/2016 12:00:00 AM	11:00	SOUTHERN	ARREST, BOOKED	800 Block of BRYANT ST	-122.403405	37.775421	(37.775420706711, -122.403404791479)	12005827212120
1	120058272	WEAPON LAWS	FIREARM, LOADED, IN VEHICLE, POSSESSION OR USE	Friday	01/29/2016 12:00:00 AM	11:00	SOUTHERN	ARREST, BOOKED	800 Block of BRYANT ST	-122.403405	37.775421	(37.775420706711, -122.403404791479)	12005827212168
2	141059263	WARRANTS	WARRANT ARREST	Monday	04/25/2016 12:00:00 AM	14:59	BAYVIEW	ARREST, BOOKED	KEITH ST / SHAFTER AV	-122.388856	37.729981	(37.7299809672996, -122.388856204292)	14105926363010
3	160013662	NON-CRIMINAL	LOST PROPERTY	Tuesday	01/05/2016 12:00:00 AM	23:50	TENDERLOIN	NONE	JONES ST / OFARRELL ST	-122.412971	37.785788	(37.7857883766888, -122.412970537591)	16001366271000
4	160002740	NON-CRIMINAL	LOST PROPERTY	Friday	01/01/2016 12:00:00 AM	00:30	MISSION	NONE	16TH ST / MISSION ST	-122.419672	37.765050	(37.7650501214668, -122.419671780296)	16000274071000

#### 3.1-San-Francisco geojson file to visualize the map

Json file from: [https://cocl.us/sanfran\\_geojson](https://cocl.us/sanfran_geojson)

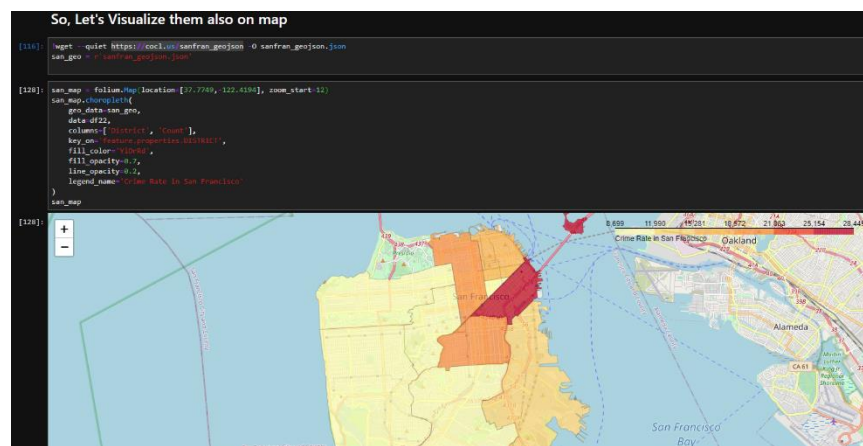
### 4-Methodology: -

Pandas Library to read the csv data file and analyzing the data frame:

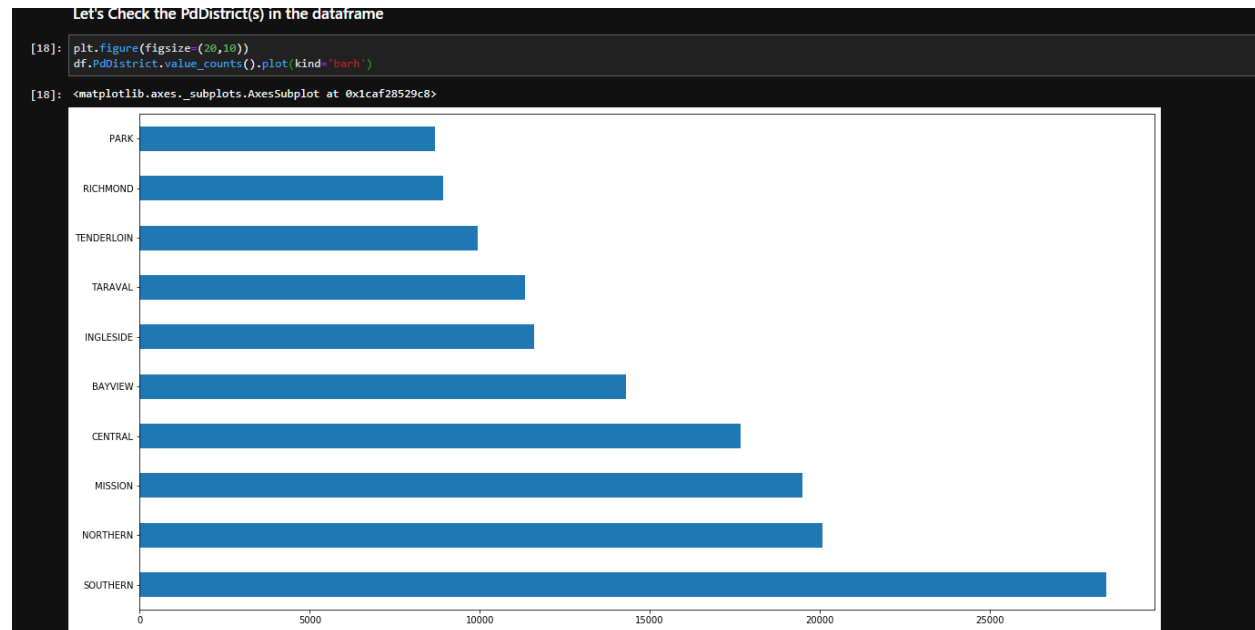
```
[120]: df=pd.read_csv('2016.csv')
[3]: df.head()
```

	IncidentNum	Category	Descript	DayOfWeek	Date	Time	PdDistrict	Resolution	Address	X	Y	Location	PdId
0	120058272	WEAPON LAWS	POSS OF PROHIBITED WEAPON	Friday	01/29/2016 12:00:00 AM	11:00	SOUTHERN	ARREST, BOOKED	800 Block of BRYANT ST	-122.403405	37.775421	(37.775420706711, -122.403404791479)	12005827212120
1	120058272	WEAPON LAWS	FIREARM, LOADED, IN VEHICLE, POSSESSION OR USE	Friday	01/29/2016 12:00:00 AM	11:00	SOUTHERN	ARREST, BOOKED	800 Block of BRYANT ST	-122.403405	37.775421	(37.775420706711, -122.403404791479)	12005827212168
2	141059263	WARRANTS	WARRANT ARREST	Monday	04/25/2016 12:00:00 AM	14:59	BAYVIEW	ARREST, BOOKED	KEITH ST / SHAFTER AV	-122.388856	37.729981	(37.7299809672996, -122.388856204292)	14105926363010
3	160013662	NON-CRIMINAL	LOST PROPERTY	Tuesday	01/05/2016 12:00:00 AM	23:50	TENDERLOIN	NONE	JONES ST / OFARRELL ST	-122.412971	37.785788	(37.7857883766888, -122.412970537591)	16001366271000
4	160002740	NON-CRIMINAL	LOST PROPERTY	Friday	01/01/2016 12:00:00 AM	00:30	MISSION	NONE	16TH ST / MISSION ST	-122.419672	37.765050	(37.7650501214668, -122.419671780296)	16000274071000

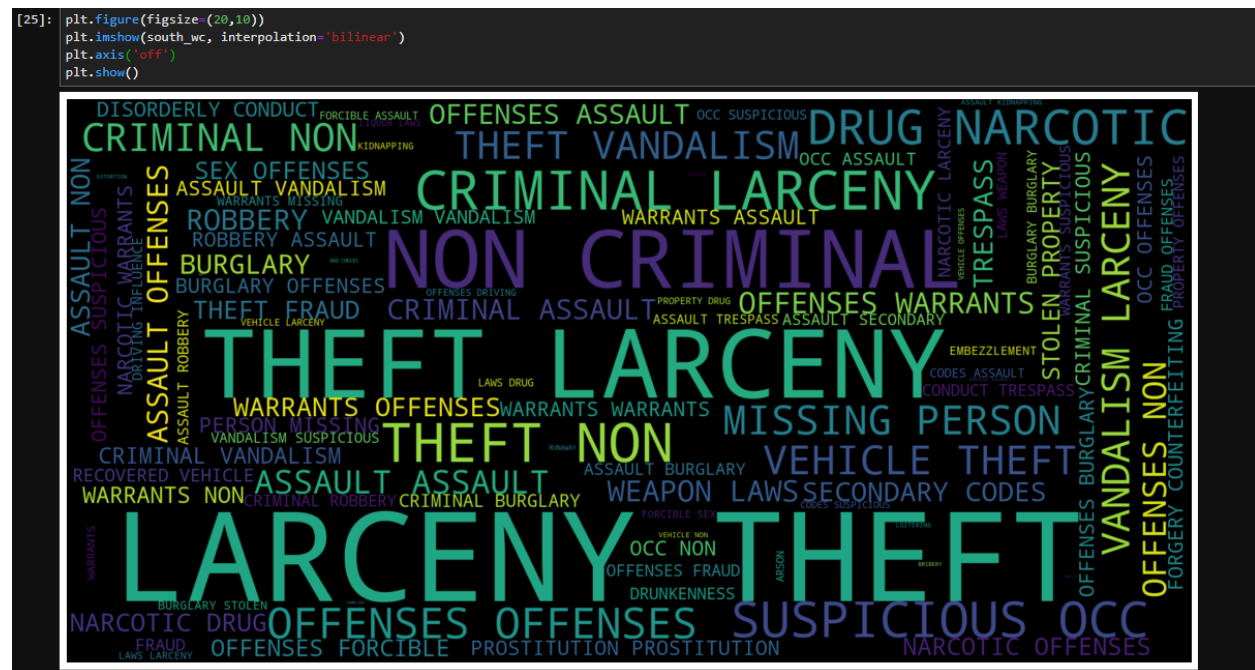
Folium Library to map the number of crimes on the Map of SF using the crime data set and SF geojson file:



Matplot Library to plot some statistics about the data:-



Wordcloud library to see the most common crime categories In the Southern District:



## Foursquare API service to know the most close venues to the most common locations of crimes

### Function for Returning The Nearest Venue According to the Foursquare API

```
[61]: def get_info(lat,lng):
# create the API request URL

url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={}&radius={}&limit={}'.format(
    CLIENT_ID,
    CLIENT_SECRET,
    VERSION,
    lat,
    lng,
    100,
    5)

# make the GET request
results = requests.get(url).json()['response']['groups'][0]['items']

venues_list=[]
# return only relevant information for each nearby venue
venues_list.append([
    lat,
    lng,
    v['venue']['name'],
    v['venue']['categories'][0]['name'],
    v['venue']['location']['distance'],
    1-v['venue']['location']['distance']/100 for v in results])

nearby_venues = pd.DataFrame([item for venue_list in venues_list for item in venue_list])
nearby_venues.columns = ['Latitude',
    'Longitude',
    'Venue',
    'Venue Category',
    'Distance From Location',
    'Accuracy']
nearby_venues.sort_values('Accuracy',ascending=False,inplace=True,)
return nearby_venues.iloc[:,:]

[115]: southernloc
```

	Longitude	Latitude	Address	Location	count	Neighborhood	Venue	Venue Category	Distance From Location	Accuracy
1	-122.403405	37.775421	800 Block of BRYANT ST	(37.775420706711, -122.403404791479)	3536	West SoMa	Golden Burma	Burmese Restaurant	26	0.74
2	-122.407634	37.784189	800 Block of MARKET ST	(37.7841893501425, -122.407633520742)	624	Union Square	Powell Street Cable Car Turnaround	Cable Car	51	0.49
3	-122.406521	37.785063	800 Block of MARKET ST	(37.7850629421661, -122.406520987144)	513	Union Square	The Glass Room	Art Gallery	24	0.76
4	-122.414318	37.779944	0 Block of UNITEDNATIONS PZ	(37.7799444052046, -122.414317857881)	299	San Francisco	Heart of The City Farmers Market	Farmers Market	6	0.94
5	-122.408421	37.783570	900 Block of MARKET ST	(37.7835699386918, -122.408421116922)	294	Union Square	Mikkeller Bar SF	Gastropub	70	0.30
6	-122.405395	37.783510	800 Block of MISSION ST	(37.7835100715421, -122.405395493939)	286	South of Market	Bloomingdale's	Department Store	46	0.54
7	-122.365565	37.809671	200 Block of INTERSTATE80 HY	(37.8096707013239, -122.365565425353)	244	San Francisco	Yerba Buena Island Tunnel	Tunnel	2	0.98
8	-122.406691	37.785207	800 Block of MARKET ST	(37.7852071918419, -122.406690592261)	236	Union Square	The Glass Room	Art Gallery	8	0.92
9	-122.403070	37.785269	700 Block of MISSION ST	(37.7852686309087, -122.403070000847)	234	Union Square	Yerba Buena Center for the Arts	Art Museum	59	0.41
10	-122.413046	37.773862	0 Block of DORE ST	(37.7738617771611, -122.413045971966)	195	Dore Street	Tank 18	Wine Bar	79	0.21

One-hot encoding using pandas method `get_dummies` to get the count for each crime category in each neighborhood

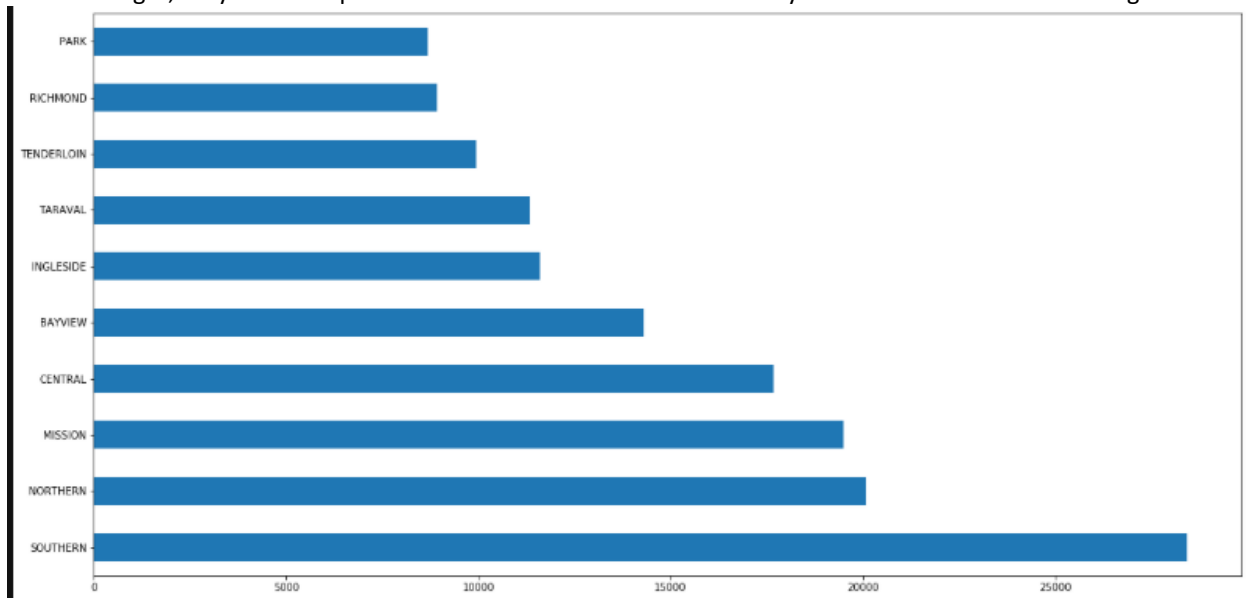
[illegible]

[135]:	Longitude	Latitude	Address	Location	count	Neighborhood	Venue	Venue Category	Distance From Location	Accuracy	Category_ARSON	Category_ASSAULT	Category_BAD CHECKS	Category_BRIBERY	Category_BURGLARY
0	-122.403405	37.775421	800 Block of BRYANT ST	(37.775420706711, -122.403404791479)	3536	West SoMa	Golden Burma	Burmese Restaurant	26	0.74	3.0	196.0	1.0	3.0	63.0
1	-122.407634	37.784189	800 Block of MARKET ST	(37.7841893501425, -122.407633520742)	624	Union Square	Powell Street Cable Car Turnaround	Cable Car	51	0.49	2.0	54.0	0.0	0.0	24.0
2	-122.406521	37.785063	800 Block of MARKET ST	(37.7850629421661, -122.406520987144)	513	Union Square	The Glass Room	Art Gallery	24	0.76	1.0	35.0	0.0	0.0	21.0
3	-122.414318	37.779944	0 Block of UNITEDNATIONS PZ	(37.7799444052046, -122.414317857881)	299	San Francisco	Heart of The City Farmers Market	Farmers Market	6	0.94	0.0	21.0	0.0	0.0	0.0
4	-122.408421	37.783570	900 Block of MARKET ST	(37.7835699386918, -122.408421116922)	294	Union Square	Mikkeller Bar SF	Gastropub	70	0.30	0.0	11.0	0.0	0.0	10.0
5	-122.405395	37.783510	800 Block of MISSION ST	(37.7835100715421, -122.405395493939)	286	South of Market	Bloomingsdale's	Department Store	46	0.54	0.0	5.0	0.0	0.0	5.0
6	-122.365565	37.809671	200 Block of INTERSTATE80 HV	(37.8096707013239, -122.365565425353)	244	San Francisco	Yerba Buena Island Tunnel	Tunnel	2	0.98	1.0	22.0	0.0	0.0	11.0
7	-122.406691	37.785207	800 Block of MARKET ST	(37.7852071918419, -122.406690592261)	236	Union Square	The Glass Room	Art Gallery	8	0.92	0.0	16.0	0.0	0.0	3.0
8	-122.403070	37.785269	700 Block of MISSION ST	(37.7852686309087, -122.403070000847)	234	Union Square	Yerba Buena Center for the Arts	Art Museum	59	0.41	0.0	20.0	0.0	0.0	2.0
9	-122.413046	37.773862	0 Block of DORE ST	(37.7738617771611, -122.413045971966)	195	Dore Street	Tank 18	Wine Bar	79	0.21	0.0	3.0	0.0	0.0	1.0

## 5-Results: -

First of all, I used pandas library to transform the csv data set to a Data frame to make my analysis on,

A total of 150.5K crime record with 13. Then I thought that maybe some districts had more records than the others. And I was right, they southern pd district had much more record than any other district as shown in the figure



Exactly 28445 records out of the 150.5k record. And most of them are Theft crimes as shown in the word cloud figure in the methodology section.

Since the southern district has more than 18% of the total record of 10 other districts as shown in the figure, so I guessed it'll be better working on the southern district and base my analysis on it.

SOUTHERN	28445
NORTHERN	20100
MISSION	19503
CENTRAL	17666
BAYVIEW	14303
INGLESIDE	11594
TARAVAL	11325
TENDERLOIN	9942
RICHMOND	8922
PARK	8699

Then I had the idea that some location may have more crime rate than the others, so I grouped the data by the location, and here comes the discovery! a certain location had 3536 crime from 28445 in 2016 only!

[103]:	Longitude	Latitude	Address	Location	count
0	-122.403405	37.775421	800 Block of BRYANT ST (37.775420706711, -122.403404791479)		3536
1	-122.407634	37.784189	800 Block of MARKET ST (37.7841893501425, -122.407633520742)		624
2	-122.406521	37.785063	800 Block of MARKET ST (37.7850629421661, -122.406520987144)		513
3	-122.414318	37.779944	0 Block of UNITEDNATIONS PZ (37.7799444052046, -122.414317857881)		299
4	-122.408421	37.783570	900 Block of MARKET ST (37.7835699386918, -122.408421116922)		294
5	-122.405395	37.783510	800 Block of MISSION ST (37.7835100715421, -122.405395493939)		286
6	-122.365565	37.809671	200 Block of INTERSTATE80 HY (37.8096707013239, -122.365565425353)		244
7	-122.406691	37.785207	800 Block of MARKET ST (37.7852071918419, -122.406690592261)		236
8	-122.403070	37.785269	700 Block of MISSION ST (37.7852686309087, -122.403070000847)		234
9	-122.413046	37.773862	0 Block of DORE ST (37.7738617771611, -122.413045971966)		195

So I decided I should work on the top 10 locations having crimes in the Southern SF,

I used the reverse geolocating from the geopy library to get the neighborhood of each location.

[06]:	Longitude	Latitude	Address	Location	count	Neighborhood
0	-122.403405	37.775421	800 Block of BRYANT ST (37.775420706711, -122.403404791479)		3536	West SoMa
1	-122.407634	37.784189	800 Block of MARKET ST (37.7841893501425, -122.407633520742)		624	Union Square
2	-122.406521	37.785063	800 Block of MARKET ST (37.7850629421661, -122.406520987144)		513	Union Square
3	-122.414318	37.779944	0 Block of UNITEDNATIONS PZ (37.7799444052046, -122.414317857881)		299	San Francisco
4	-122.408421	37.783570	900 Block of MARKET ST (37.7835699386918, -122.408421116922)		294	Union Square
5	-122.405395	37.783510	800 Block of MISSION ST (37.7835100715421, -122.405395493939)		286	South of Market
6	-122.365565	37.809671	200 Block of INTERSTATE80 HY (37.8096707013239, -122.365565425353)		244	San Francisco
7	-122.406691	37.785207	800 Block of MARKET ST (37.7852071918419, -122.406690592261)		236	Union Square
8	-122.403070	37.785269	700 Block of MISSION ST (37.7852686309087, -122.403070000847)		234	Union Square
9	-122.413046	37.773862	0 Block of DORE ST (37.7738617771611, -122.413045971966)		195	Dore Street

Of course, Locations having this high rate of crimes through only one year, there's must be a reason for this, so I used the Foursquare API to get the most close venue in each exact location and the distance from the location.

[115]:	Longitude	Latitude	Address	Location	count	Neighborhood	Venue	Venue Category	Distance From Location	Accuracy
	-122.403405	37.775421	800 Block of BRYANT ST (37.775420706711, -122.403404791479)		3536	West SoMa	Golden Burma	Burmese Restaurant	26	0.74
	-122.407634	37.784189	800 Block of MARKET ST (37.7841893501425, -122.407633520742)		624	Union Square	Powell Street Cable Car Turnaround	Cable Car	51	0.49
	-122.406521	37.785063	800 Block of MARKET ST (37.7850629421661, -122.406520987144)		513	Union Square	The Glass Room	Art Gallery	24	0.76
	-122.414318	37.779944	0 Block of UNITEDNATIONS PZ (37.7799444052046, -122.414317857881)		299	San Francisco	Heart of The City Farmers Market	Farmers Market	6	0.94
	-122.408421	37.783570	900 Block of MARKET ST (37.7835699386918, -122.408421116922)		294	Union Square	Mikkeller Bar SF	Gastropub	70	0.30
	-122.405395	37.783510	800 Block of MISSION ST (37.7835100715421, -122.405395493939)		286	South of Market	Bloomingsdale's	Department Store	46	0.54
	-122.365565	37.809671	200 Block of INTERSTATE80 HY (37.8096707013239, -122.365565425353)		244	San Francisco	Yerba Buena Island Tunnel	Tunnel	2	0.98
	-122.406691	37.785207	800 Block of MARKET ST (37.7852071918419, -122.406690592261)		236	Union Square	The Glass Room	Art Gallery	8	0.82
	-122.403070	37.785269	700 Block of MISSION ST (37.7852686309087, -122.403070000847)		234	Union Square	Yerba Buena Center for the Arts	Art Museum	59	0.41
	-122.413046	37.773862	0 Block of DORE ST (37.7738617771611, -122.413045971966)		195	Dore Street	Tank 18	Wine Bar	79	0.21

Then I used one-hot encoding to know the number of each crime category on each exact location through the whole year of 2016, and as we knew lately from the word cloud that most of them is Theft, it was True this time!

[193]:

	Neighborhood	Venue	Venue Category	Most Crime Type	Count
0	West SoMa	Golden Burma	Burmese Restaurant	LARCENY/THEFT	1350.0
1	Union Square	Powell Street Cable Car Turnaround	Cable Car	LARCENY/THEFT	279.0
2	Union Square	The Glass Room	Art Gallery	LARCENY/THEFT	208.0
3	San Francisco	Heart of The City Farmers Market	Farmers Market	WARRANTS	86.0
4	Union Square	Mikkeller Bar SF	Gastropub	LARCENY/THEFT	158.0
5	South of Market	Bloomingdale's	Department Store	LARCENY/THEFT	188.0
6	San Francisco	Yerba Buena Island Tunnel	Tunnel	LARCENY/THEFT	64.0
7	Union Square	The Glass Room	Art Gallery	LARCENY/THEFT	154.0
8	Union Square	Yerba Buena Center for the Arts	Art Museum	LARCENY/THEFT	104.0
9	Dore Street	Tank 18	Wine Bar	MISSING PERSON	133.0

So I thought that maybe the day of the week has some influence on these high theft crime rate ! so I got the day in each location having the most crimes on it.

[297]:

	Neighborhood	Venue	Venue Category	Most Crime Type	Count	Highest Day	Highest_Day_Count	Percentage
0	West SoMa	Golden Burma	Burmese Restaurant	LARCENY/THEFT	1350.0	Friday	215	15.925926
1	Union Square	Powell Street Cable Car Turnaround	Cable Car	LARCENY/THEFT	279.0	Monday	50	17.921147
2	Union Square	The Glass Room	Art Gallery	LARCENY/THEFT	208.0	Tuesday	34	16.346154
3	San Francisco	Heart of The City Farmers Market	Farmers Market	WARRANTS	86.0	Thursday	22	25.581395
4	Union Square	Mikkeller Bar SF	Gastropub	LARCENY/THEFT	158.0	Friday	26	16.455696
5	South of Market	Bloomingdale's	Department Store	LARCENY/THEFT	188.0	Saturday	38	20.212766
6	San Francisco	Yerba Buena Island Tunnel	Tunnel	LARCENY/THEFT	64.0	Saturday	16	25.000000
7	Union Square	The Glass Room	Art Gallery	LARCENY/THEFT	154.0	Wednesday	30	19.480519
8	Union Square	Yerba Buena Center for the Arts	Art Museum	LARCENY/THEFT	104.0	Wednesday	22	21.153846
9	Dore Street	Tank 18	Wine Bar	MISSING PERSON	133.0	Monday	26	19.548872

## **6-Discussion: -**

As shown in the figures in the results sections, some neighborhoods in the southern sanfran. Has more crime rates than the other neighborhoods, and also some locations have more crime rates than the other. So, they have more active forces in this location in the days in the figure above as we discussed before, and considering the last figure, most of the theft crimes happen at the weekend, specially near Golden Burma Restaurant in west soMa neighborhood. So the sanfran PD should do more effort attaining the safety of the san Francisco citizens.

## **7-Conclusion**

Out of the 150.5K crimes in 2016, the southern district has more than 19% of them and out of this 19% there's one exact place that has more than 12% of these crimes around it "Golden Burma Restaurant in west soMa". And specially on Fridays, that's relatively high percentage regarding that san-fran has 36 neighborhoods, so the crimes are not distributed well to reflect a normal crime rate. So the pd should do more investigating regarding the places that has an up normal crime rate in it.