Primary Finding

As a cop in San Francisco, you would have the best chance of catching some robbers on a Friday/Saturday, between 7-8pm, pretty much during the whole of the year (with possibly an exception in December)

Below are three tables to visualize this trend.

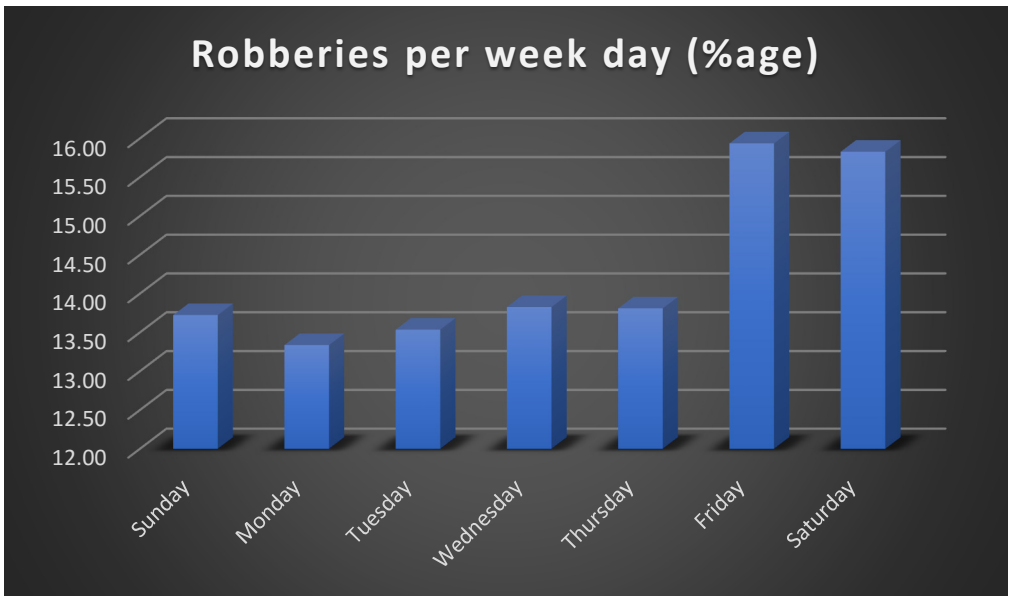
Robberies per week day		
Day	Percentage	Incident Cnt
Sun	13.72	15456
Mon	13.34	15021
Tue	13.54	15245
Wed	13.83	15573
Thu	13.81	15553
Fri	**15.94**	**17947**
Sat	**15.83**	**17826**

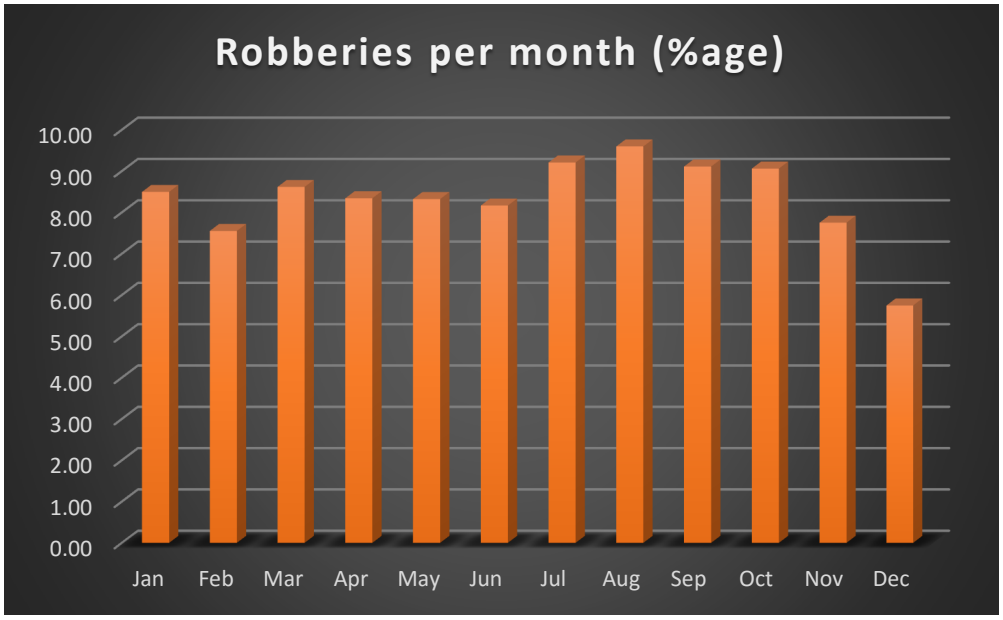
Robberies Per Month		
Month	Percentage	Incident Cnt
Jan	8.50	9571
Feb	7.55	8507
Mar	8.62	9707
Apr	8.34	9398
May	8.33	9376
Jun	8.17	9199
Jul	9.21	10373
Aug	9.60	10812
Sep	9.11	10263
Oct	9.06	10207
Nov	7.75	8732
Dec	**5.75**	**6476**

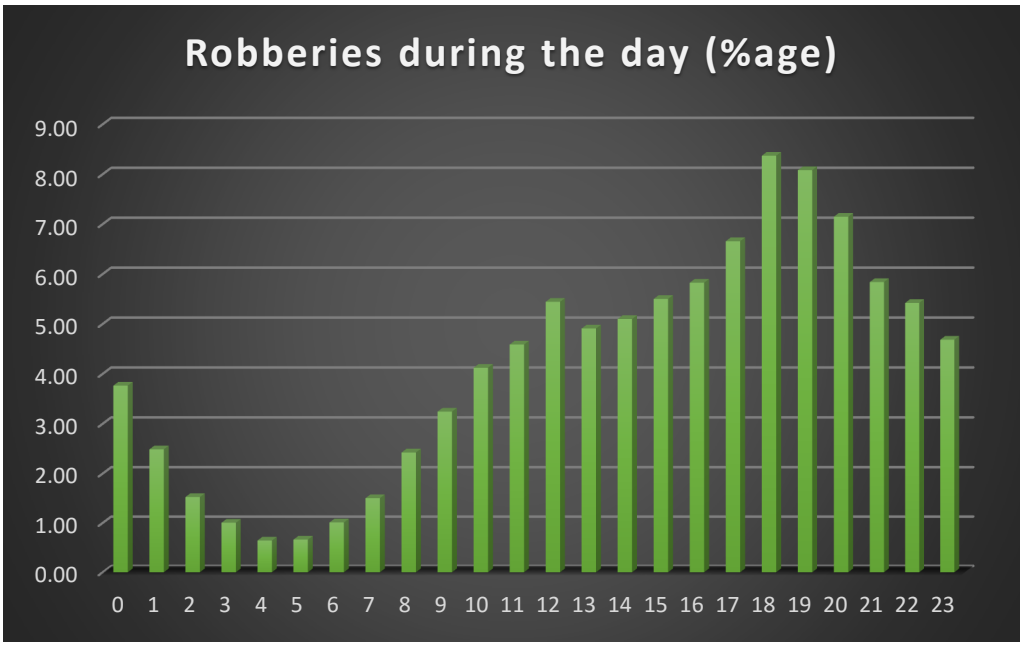
Robberies Per Hour (24 hr)		
Hour	Percentage	Incident Cnt
0	3.77	4242
1	2.49	2803
2	1.53	1719
3	1.01	1132
4	0.65	727
5	0.67	749
6	1.01	1140
7	1.51	1695
8	2.43	2733
9	3.25	3655
10	4.12	4644
11	4.59	5168
12	5.44	6132
13	4.91	5531
14	5.10	5744
15	5.50	6198
16	5.83	6564
17	**6.66**	**7500**
18	**8.37**	**9429**
19	**8.08**	**9102**
20	**7.15**	**8052**
21	5.84	6579
22	5.42	6106
23	4.69	5277

From the above data set, the salient findings can be summarized as:

- Even though the number of robberies seem to be evenly distributed across the week, there's about an overall increase on Fridays and Saturdays.
- A similar trend can also be gleaned from the distribution of the incidents across the year i.e. they seem to be more or less the same per month, with the only exception of December.
- SF gets more robber friendly in the evening, from 5-8pm. I was surprised that in the early morning hours (2-6am) the robbery incidents were significantly less. I would have expected the robbers to be more alert during the sleeping hours of the city. Clearly, I would have not made for a good robber!

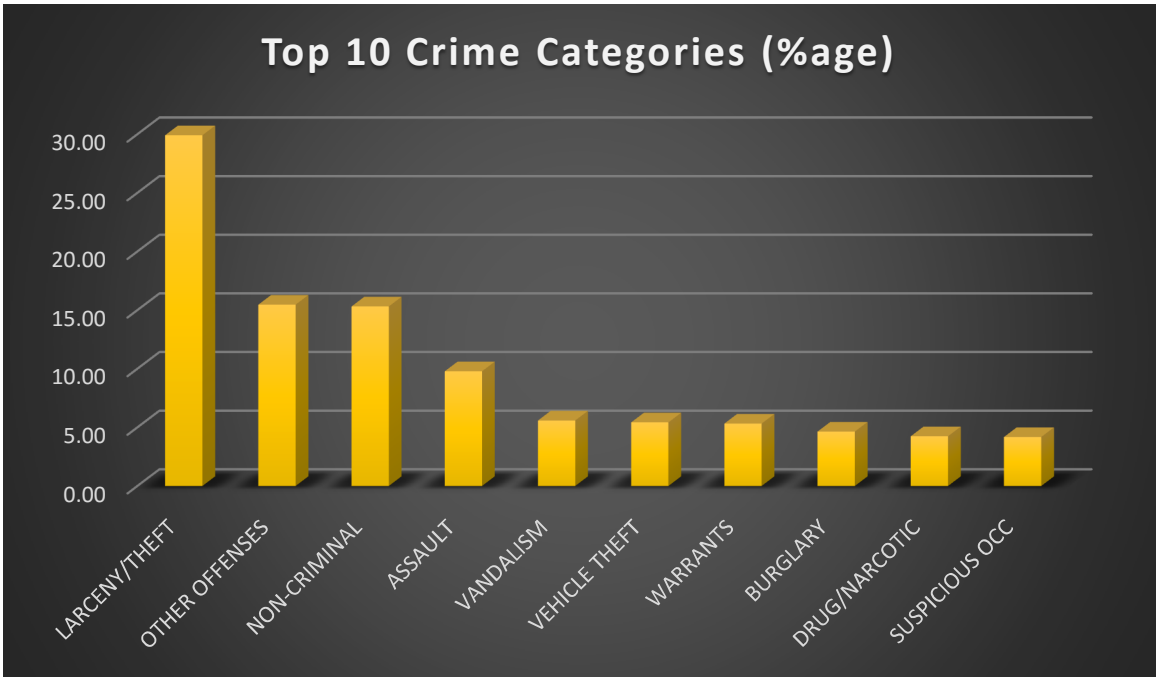

Robberies per week day (%age)
16.00
15.50
15.00
14.50
14.00
13.50
13.00
12.50
12.00
Sunday
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday


Robberies per month (%age)
10.00
9.00
8.00
7.00
6.00
5.00
4.00
3.00
2.00
1.00
0.00
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec


Robberies during the day (%age)
9.00
8.00
7.00
6.00
5.00
4.00
3.00
2.00
1.00
0.00
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23

- Of all categories of incident types in SF; ***LARCENY/THEFT*** is the most common by a significant margin. Below is the data for the top 10 categories of crime committed in SF from January 1st 2013 until November 25th 2015.

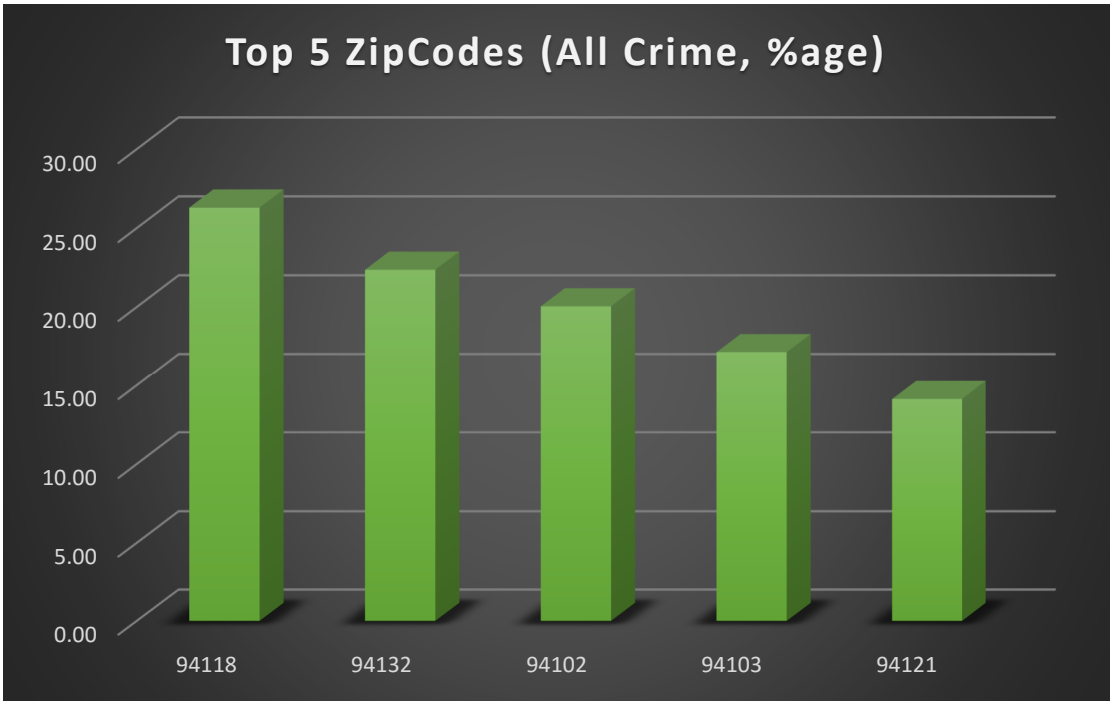
Top 10 Categories of Crime		
Category	Percentage	Incident Count
LARCENY/THEFT	29.92	112621
OTHER OFFENSES	15.48	58246
NON-CRIMINAL	15.33	57681
ASSAULT	9.80	36866
VANDALISM	5.60	21071
VEHICLE THEFT	5.44	20467
WARRANTS	5.34	20081
BURGLARY	4.65	17518
DRUG/NARCOTIC	4.26	16022
SUSPICIOUS OCC	4.19	15774

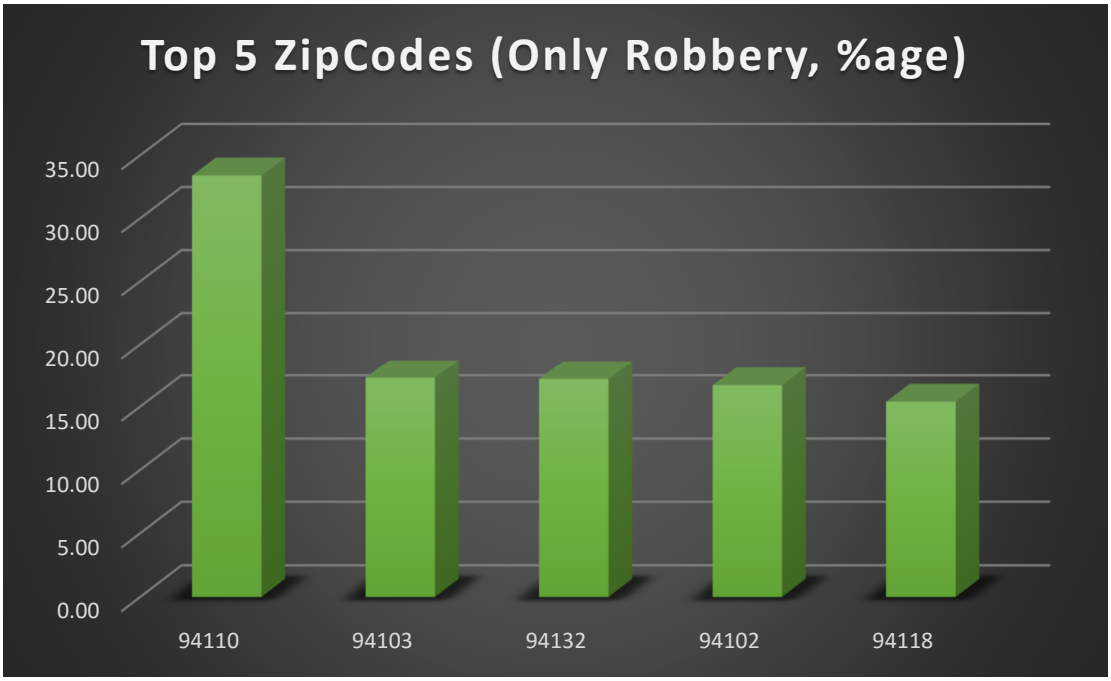


- Out of the top 5 zip codes for ALL incident types in SF and the top 5 zip codes for Robberies; 4 zip codes are the same.

Top 5 ZipCodes (ALL Crime)		
Zip Code	Percentage	Incident Cnt
94118	26.31	715
94132	22.37	608
94102	20.05	545
94103	17.14	466
94121	14.13	384

Top 5 ZipCodes (Only Robbery)		
Zip Code	Percentage	Incident Cnt
94110	33.40	2298
94103	17.30	1190
94132	17.21	1184
94102	16.70	1149
94118	15.39	1059

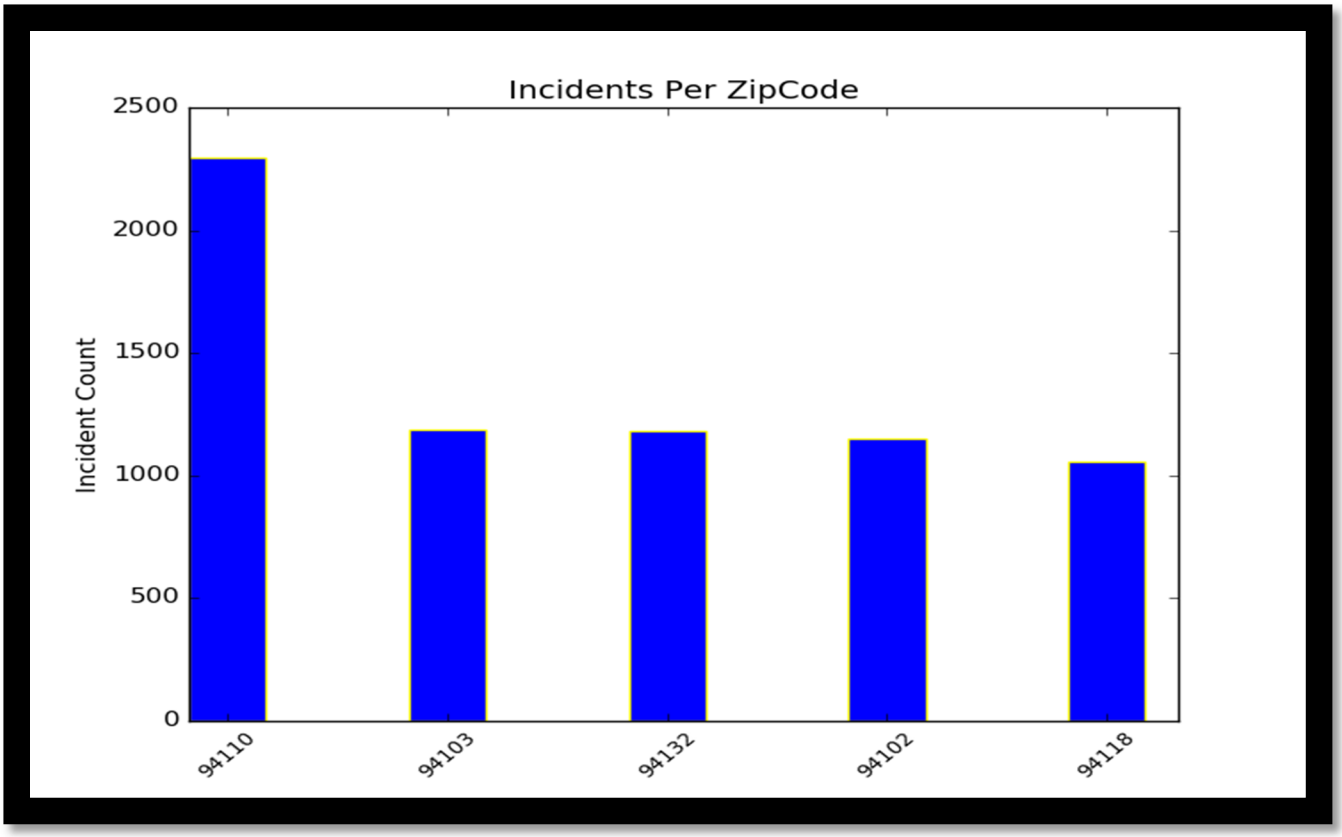


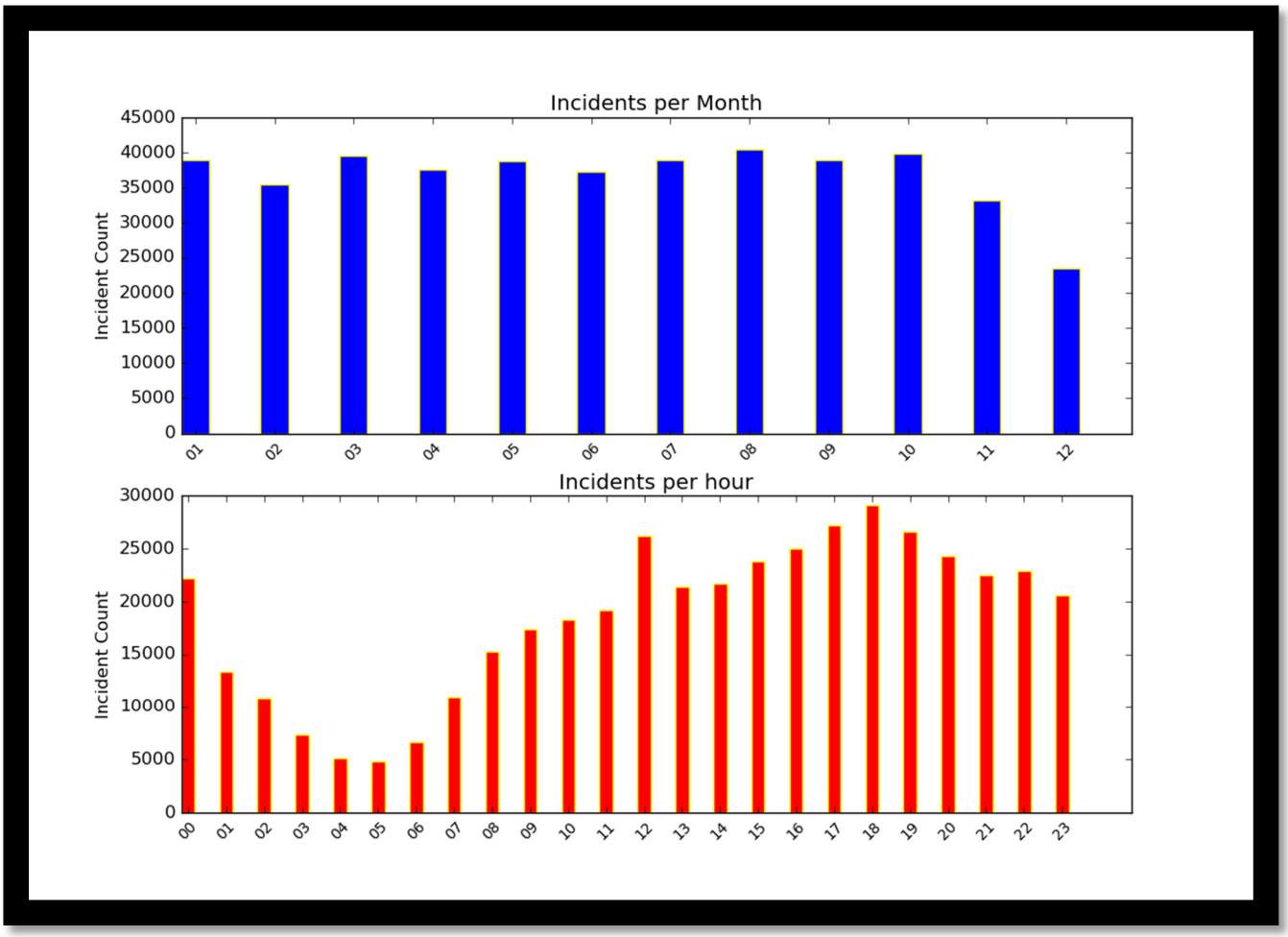


Methodology:

1. Retrieved data from 01/01/2013 – 11/25/2015 from the following link - San Francisco Data Portal
2. Wrote python scrapper, which did the following:
    a. Loaded the entire data set into a Pandas DataFrame.
    b. From the DataFrame it extracted:
        i. Top 10 categories
        ii. Incident counts per hour of the day, day of the week and month of the year
        iii. Latitude and Longitude of the top 5 zip codes; with the maximum crime rate
        iv. Steps ii and iii were repeated for the top offending category from step i
        v. The Lat/Long data extracted from the DataFrame, I wrote a python module which called the Google Maps API to GeoCode the locations i.e. for each Lat/Long tuple, I fetched its zip code
        vi. Finally, this data set was then used to plot the charts using matplotlib.

3. Some of the matplotlib charts as created from the python script are attached below. As I wasn't able to post process them in the way I best thought would benefit the reader – I ultimately used the Excel Chart Objects in word.
4. The python code for the data analysis, geo coding lat/long and plotting the bar graphs have been uploaded in the same git repo as this PDF submission. (https://github.com/ranjanabhb/uw-datascience)


Incidents Per ZipCode
Incident Count
2500
2000
1500
1000
500
0
94110
94103
94132
94102
94118


Incidents per Month
Incident Count
45000
40000
35000
30000
25000
20000
15000
10000
5000
0
01
02
03
04
05
06
07
08
09
10
11
12
Incidents per hour
Incident Count
30000
25000
20000
15000
10000
5000
0
00
01
02
03
04
05
06
07
08
09
10
11
12
13
14
15
16
17
18
19
20
21
22
23