Notebook

May 4, 2019

Contents

1 Introduction

This notebook is intended to help decide where to have dinner in San Francisco, safely. The data used for this are: 1. Foursquare API 2. SF Police Department Incident Reports: 2018 to Present

The final product of the notebook is an interactive map of SF, with districts colored by crime rate and markers showing restaurants, as well as sample crimes in the vicinity.

1.1 Crime dataset exploration

The dataset published by the SFPD includes important information regarding the incident (e.g. category) as well as its location. Most of the data in the 34 columns of this dataset (shown in table 1.1) will be useless for the purpose of this notebook, so only `Police District`, `Latitude`, `Longitude` and `Incident Category` will be retained, as illustrated in table 1.2.

Table 1.1: Top 5 rows of the original SFPD crimes dataset

Table 1.2: Top 5 rows, only relevant columns

	Police District	Latitude	Longitude	Incident Category
0	Southern	37.784908	-122.404795	Other Miscellaneous
1	Central	37.786410	-122.408036	Lost Property
2	Mission	37.756834	-122.406699	Assault
3	Southern	37.784007	-122.404865	Larceny Theft
4	Central	NaN	NaN	Larceny Theft

Some of the incidents reported are not relevant or to vague to include in the analysis, so they will be dropped. Incidents from categories `Gambling`, `Liquor Laws`, `Vehicle Impounded`, `Vehicle Misplaced`, `Case Closure`, `Other`, `Miscellaneous Investigation`, `Recovered Vehicle`, `Non-Criminal`, `Other Miscellaneous` amount to 36124 and after filtering the dataset is left with 158729 incidents. Table **??** shows number of incidents per category, also visualized in figure 1.2. Figure 1.1 shows the incident numbers per district and figure 1.3 as well as table **??** show that information combined.

```
                            Incident Category   Count
0       Human Trafficking, Commercial Sex Acts      5
1                              Weapons Offence      6
2                                     Homicide     30
3                                   Suspicious     32
4                          Motor Vehicle Theft?     37
5                                         Rape     56
6                               Drug Violation     61
7                                      Suicide     81
8    Human Trafficking (A), Commercial Sex Acts     85
9                               Courtesy Report    131
10                                 Fire Report    189
11                                   Vandalism    235
12                                 Embezzlement    236
13                               Family Offense    255
14                             Juvenile Offenses    263
15                             Traffic Collision    354
16                                  Sex Offense    402
17                                        Arson    453
```

```
18                                    Prostitution      569
19                                 Civil Sidewalks      640
20                     Forgery And Counterfeiting      882
21                                 Stolen Property     1037
22                                 Weapons Offense     1047
23                             Weapons Carrying Etc     1082
24                                 Other Offenses     1360
25                         Traffic Violation Arrest     2721
26                             Disorderly Conduct     3430
27                                 Suspicious Occ     3578
28     Offences Against The Family And Children     3886
29                                 Missing Person     4648
30                                       Robbery     4718
31                                 Drug Offense     5361
32                                         Fraud     5628
33                                 Lost Property     6590
34                             Motor Vehicle Theft     6820
35                                       Warrant     7423
36                                     Burglary     9077
37                           Malicious Mischief    11589
38                                     Assault    11813
39                               Larceny Theft    61902
```
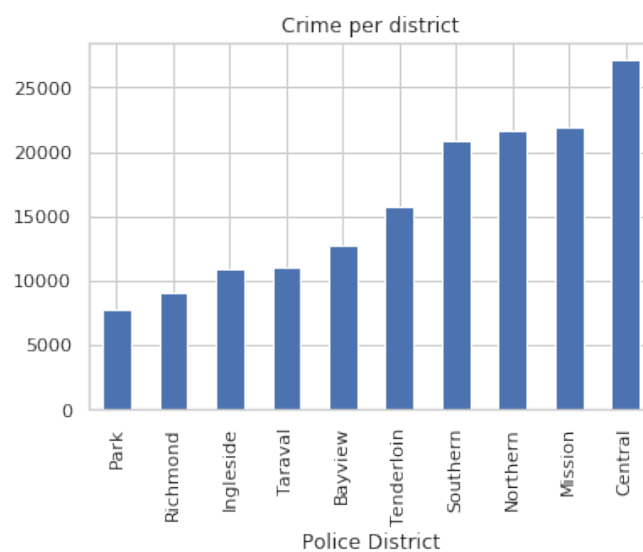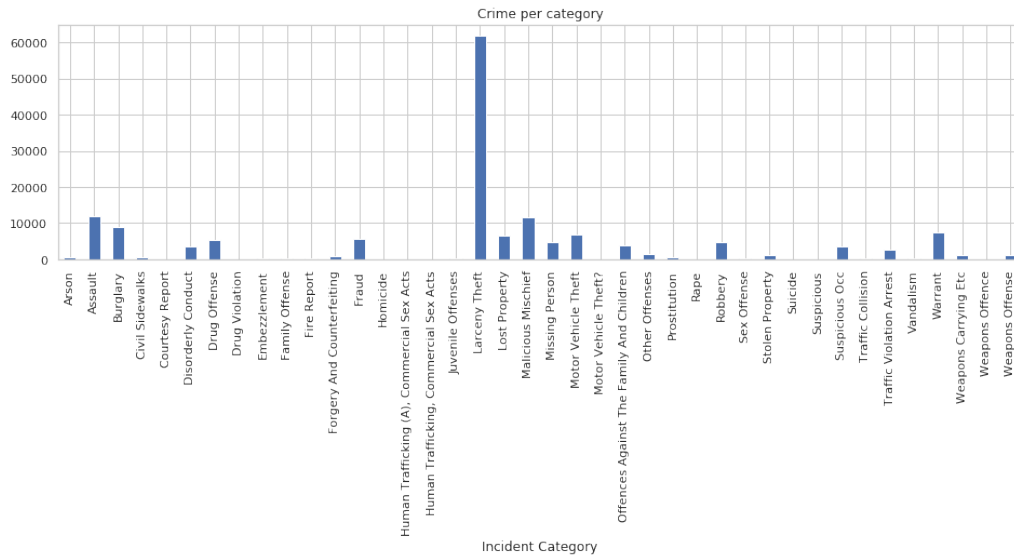

Figure 1.1: Number of crimes per district

Figure 1.2: Crime per category

```
     Incident  Category  Police  District     count
0        Larceny  Theft          Central      14091
1        Larceny  Theft         Northern      11128
2        Larceny  Theft         Southern       8234
3        Larceny  Theft          Mission       7334
4        Larceny  Theft         Richmond       4533
5        Larceny  Theft          Taraval       3866
6        Larceny  Theft       Tenderloin       3734
7        Larceny  Theft          Bayview       3487
8        Larceny  Theft        Ingleside       2792
9        Larceny  Theft             Park       2703
```
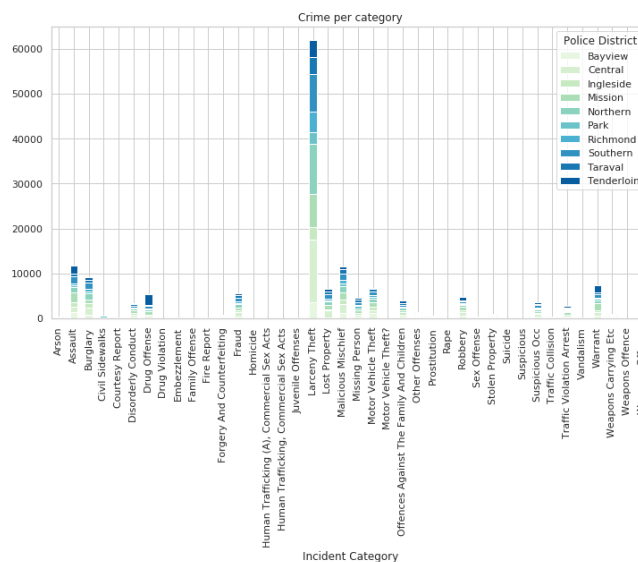

Figure 1.3: Number of crimes per category and district

Larceny Theft per district
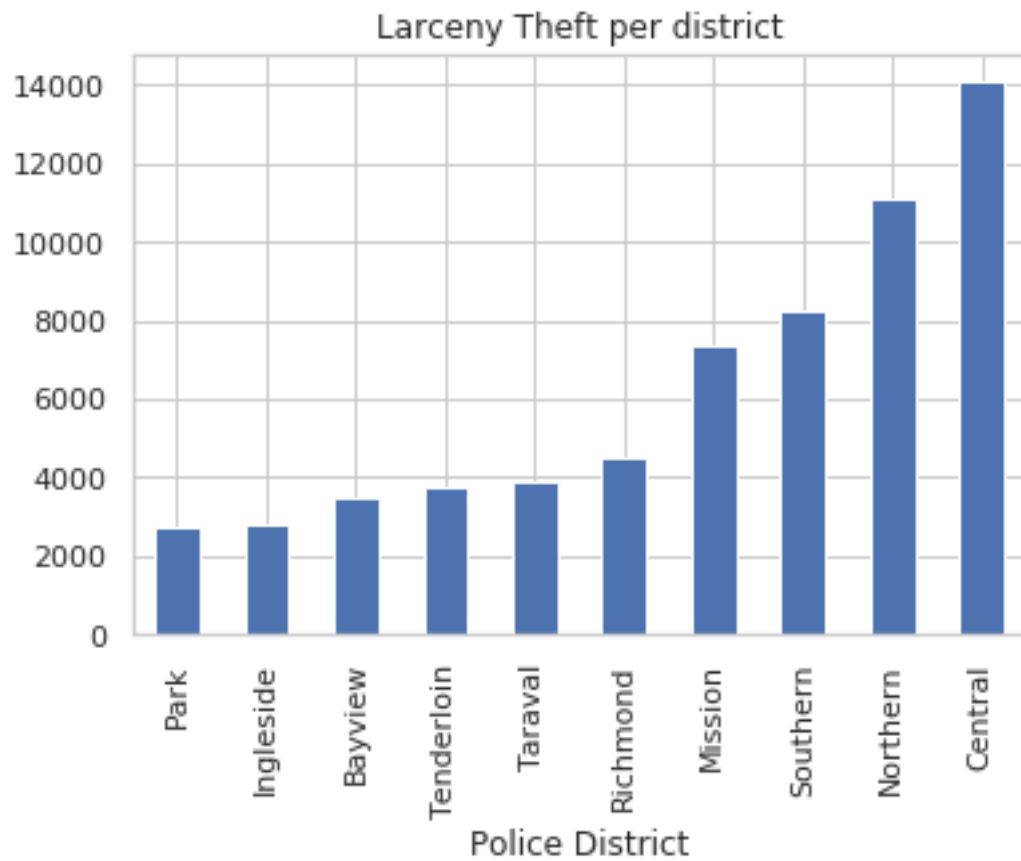
1.2 Crime dataset summary

Most important information regarding crime rate per category and district is shown in table 1.3.

Table 1.3: Incident count grouped by category and district.

| Police District Incident Category	Bayview	Central	Ingleside	Mission	Northern	Park	Richmond	Southern	Taraval	Tenderloin
Arson	73.0	65.0	44.0	60.0	37.0	12.0	24.0	60.0	22.0	56.0
Assault	1311.0	1363.0	944.0	2054.0	1253.0	430.0	363.0	1586.0	694.0	1815.0
Burglary	833.0	1590.0	857.0	910.0	1358.0	491.0	449.0	1407.0	753.0	429.0
Civil Sidewalks	NaN	15.0	NaN	168.0	3.0	437.0	8.0	6.0	NaN	3.0
Courtesy Report	17.0	20.0	11.0	16.0	9.0	6.0	12.0	15.0	12.0	13.0
Disorderly Conduct	375.0	420.0	337.0	608.0	378.0	115.0	161.0	422.0	291.0	323.0
Drug Offense	188.0	375.0	216.0	791.0	347.0	181.0	85.0	664.0	112.0	2402.0
Drug Violation	8.0	3.0	9.0	8.0	2.0	1.0	NaN	8.0	NaN	22.0
Embezzlement	28.0	37.0	10.0	28.0	19.0	7.0	15.0	41.0	16.0	35.0
Family Offense	67.0	13.0	27.0	40.0	19.0	6.0	20.0	18.0	29.0	16.0
Fire Report	38.0	19.0	26.0	27.0	11.0	8.0	9.0	24.0	16.0	11.0
Forgery And Counterfeiting	57.0	156.0	69.0	143.0	113.0	34.0	49.0	136.0	58.0	67.0
Fraud	360.0	973.0	464.0	601.0	760.0	262.0	396.0	693.0	750.0	369.0
Homicide	6.0	3.0	1.0	6.0	2.0	NaN	1.0	3.0	4.0	4.0
Human Trafficking (A), Commercial Sex Acts	1.0	13.0	1.0	10.0	3.0	NaN	NaN	34.0	20.0	3.0
Human Trafficking, Commercial Sex Acts	NaN	NaN	1.0	1.0	NaN	NaN	NaN	3.0	NaN	NaN
Juvenile Offenses	12.0	1.0	17.0	10.0	11.0	26.0	15.0	7.0	158.0	6.0
Larceny Theft	3487.0	14091.0	2792.0	7334.0	11128.0	2703.0	4533.0	8234.0	3866.0	3734.0
Lost Property	285.0	1461.0	314.0	939.0	793.0	235.0	373.0	1032.0	464.0	694.0
Malicious Mischief	1275.0	1915.0	983.0	1540.0	1512.0	565.0	682.0	1570.0	828.0	719.0
Missing Person	392.0	327.0	420.0	580.0	436.0	642.0	223.0	582.0	661.0	385.0
Motor Vehicle Theft	1118.0	658.0	833.0	956.0	741.0	388.0	403.0	828.0	632.0	263.0
Motor Vehicle Theft?	4.0	5.0	7.0	7.0	3.0	1.0	1.0	3.0	4.0	2.0
Offences Against The Family And Children	466.0	364.0	472.0	539.0	456.0	141.0	200.0	412.0	380.0	456.0
Other Offenses	97.0	140.0	100.0	201.0	104.0	154.0	76.0	178.0	65.0	245.0
Prostitution	NaN	117.0	2.0	284.0	8.0	NaN	NaN	65.0	46.0	47.0
Rape	4.0	2.0	3.0	12.0	6.0	2.0	2.0	16.0	3.0	6.0
Robbery	523.0	801.0	496.0	685.0	548.0	104.0	143.0	486.0	213.0	719.0
Sex Offense	25.0	38.0	29.0	64.0	55.0	17.0	9.0	91.0	28.0	46.0
Stolen Property	97.0	207.0	75.0	141.0	150.0	28.0	49.0	125.0	52.0	113.0
Suicide	5.0	9.0	7.0	8.0	13.0	4.0	10.0	12.0	8.0	5.0
Suspicious	1.0	4.0	NaN	1.0	1.0	1.0	5.0	18.0	1.0	NaN
Suspicious Occ	413.0	501.0	315.0	450.0	275.0	146.0	292.0	529.0	249.0	408.0
Traffic Collision	55.0	39.0	46.0	53.0	41.0	16.0	29.0	31.0	26.0	18.0
Traffic Violation Arrest	160.0	258.0	342.0	693.0	186.0	115.0	162.0	317.0	181.0	307.0
Vandalism	19.0	27.0	17.0	42.0	36.0	17.0	14.0	29.0	23.0	11.0
Warrant	553.0	890.0	411.0	1462.0	651.0	349.0	184.0	943.0	251.0	1729.0
Weapons Carrying Etc	178.0	113.0	139.0	199.0	83.0	12.0	26.0	93.0	58.0	181.0
Weapons Offence	NaN	2.0	NaN	2.0	NaN	NaN	NaN	NaN	1.0	1.0
Weapons Offense	172.0	96.0	119.0	203.0	80.0	35.0	41.0	134.0	54.0	113.0

2 Venues dataset exploration

For each of the districts in the SFPD dataset a single incident was randomly selected in order to get its latitude and longitude, which the serve to query the Foursquare API. Again the full dataset consist 31 columns, of which only several (['venue.location.lat','venue.location.lng','venue.name']) will be useful for the purpose of this analysis, other columns will be dropped, as well as locations with NaN coordinates. At the time of writing of this report - the venue dataset consisted of 884 restaurants.

```
777
```

```
478
```

|	Police District	venue.location.lat	venue.location.lng	venue.name	venue.id	rating
0	Bayview	37.734826	-122.390764	Radio Africa & Kitchen	4f56b36a6b7406c0ea010e9b	8.4
1	Bayview	37.736240	-122.390361	TATO	5b89f7f6a0215b002c7cfd73	8.4
2	Bayview	37.726908	-122.391536	Craftsman and Wolves Den	5703cab0498e52dae7c36bcf	8.2
3	Bayview	37.732747	-122.392245	Fox and Lion Bakery	5618351f498e69559d3990b7	7.8
4	Bayview	37.734287	-122.390462	Old Skool Cafe	4d0401f454d0236aa38cecd5	7.4


```
<folium.folium.Map at 0x7f507cf55c50>
```
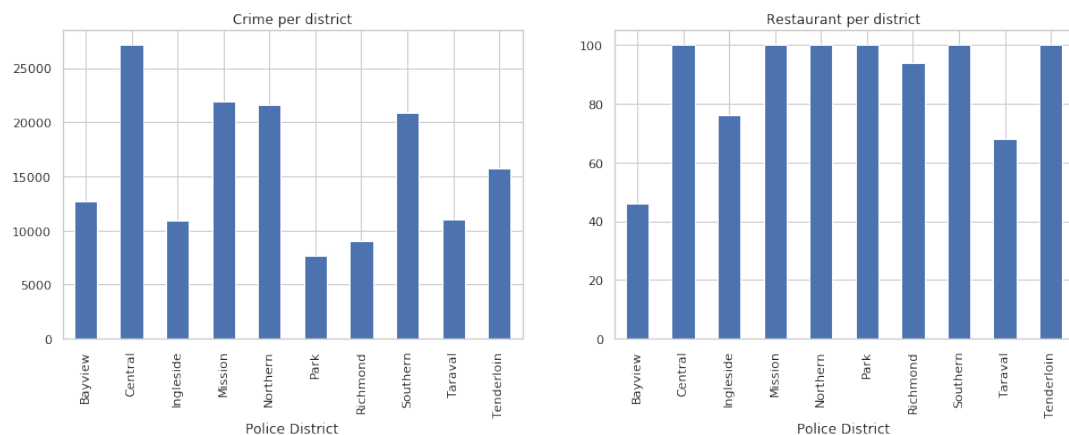
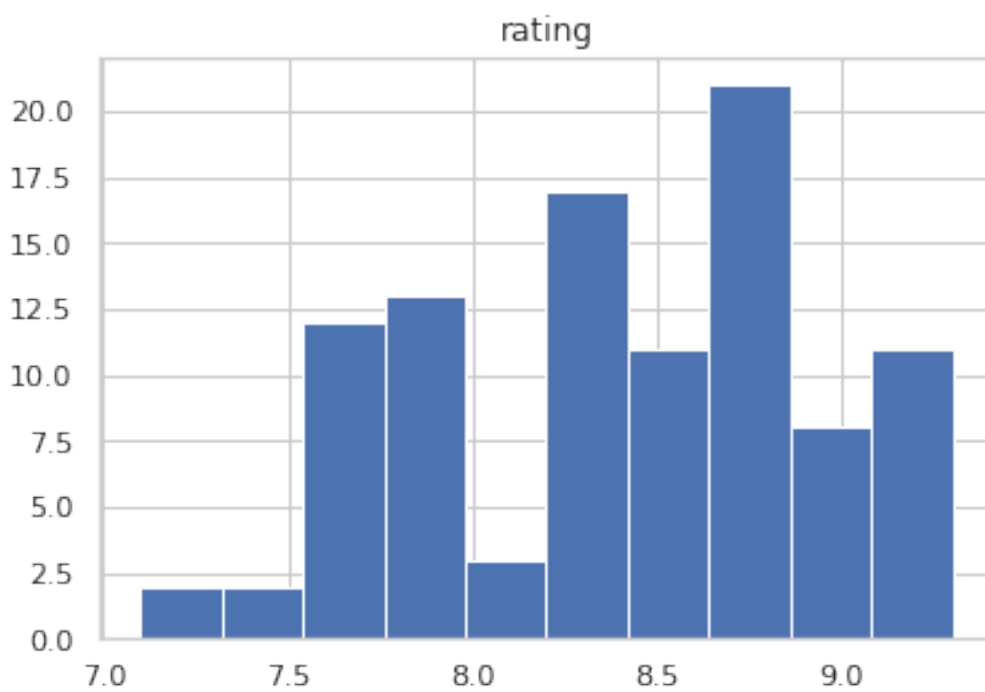```
GeoJSON file downloaded!
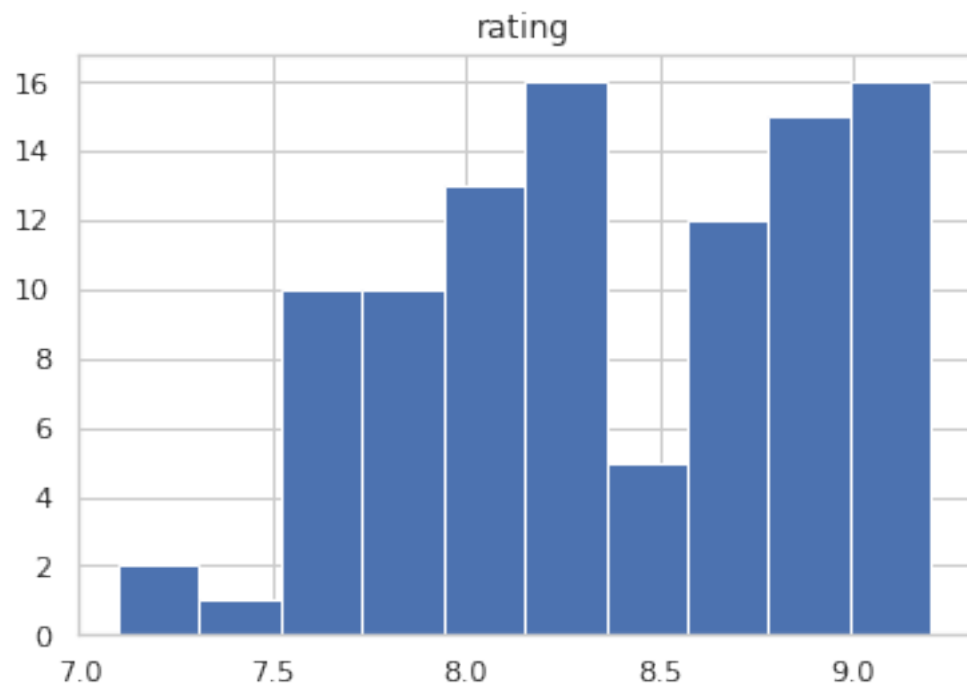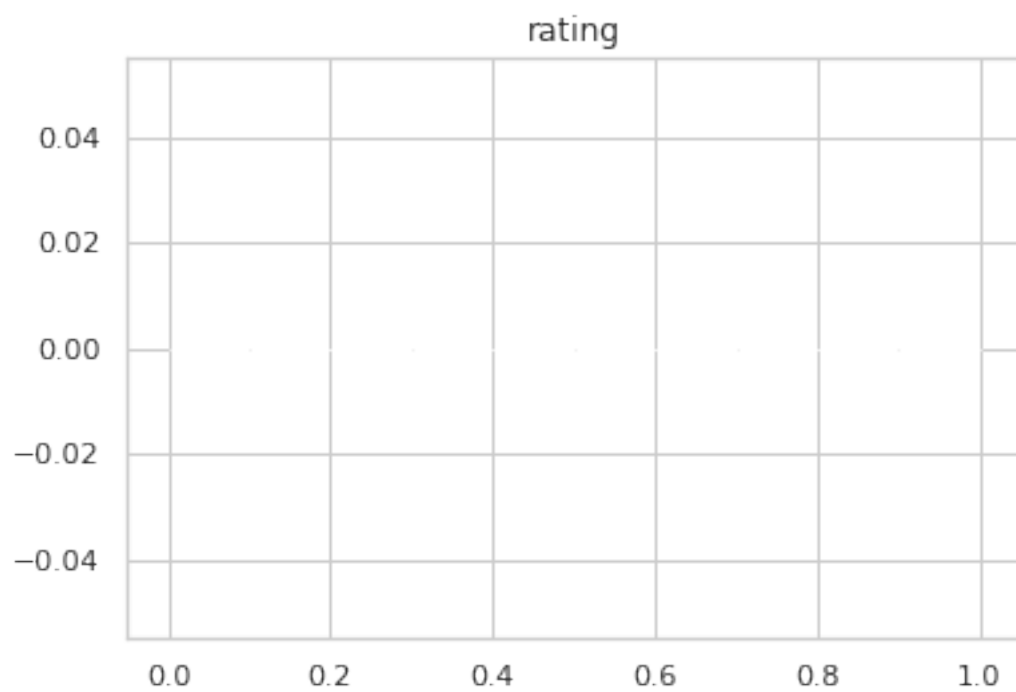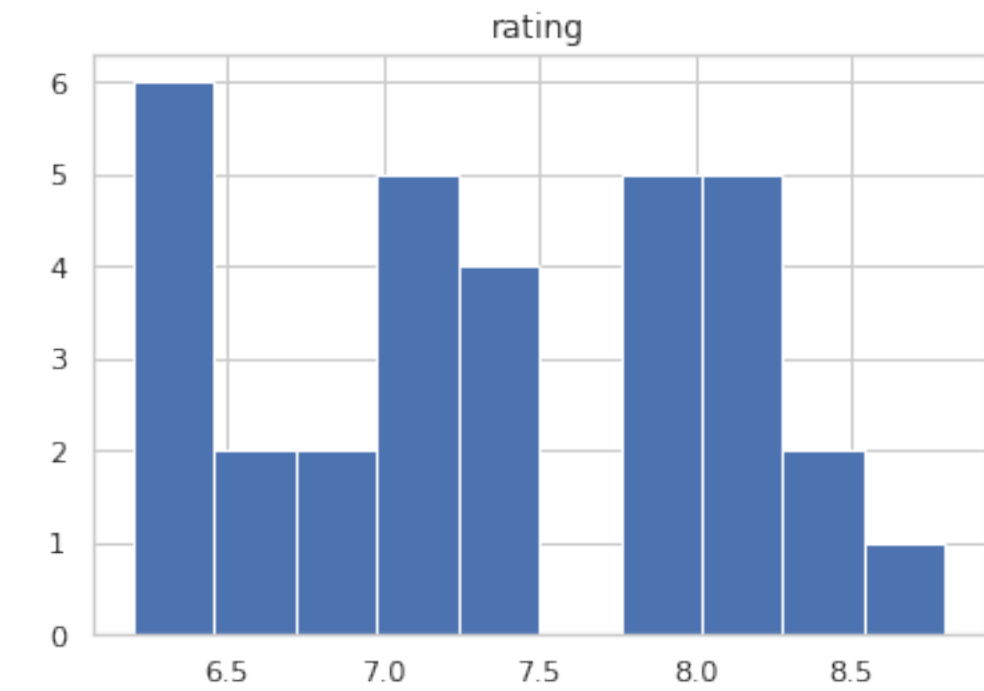```

```
<folium.folium.Map at 0x7f507f8a8d68>
```

2.1 Explore highest rated restaurants by district

As the first step to choose the best place to dine we will look into how ratings per district are distributed.
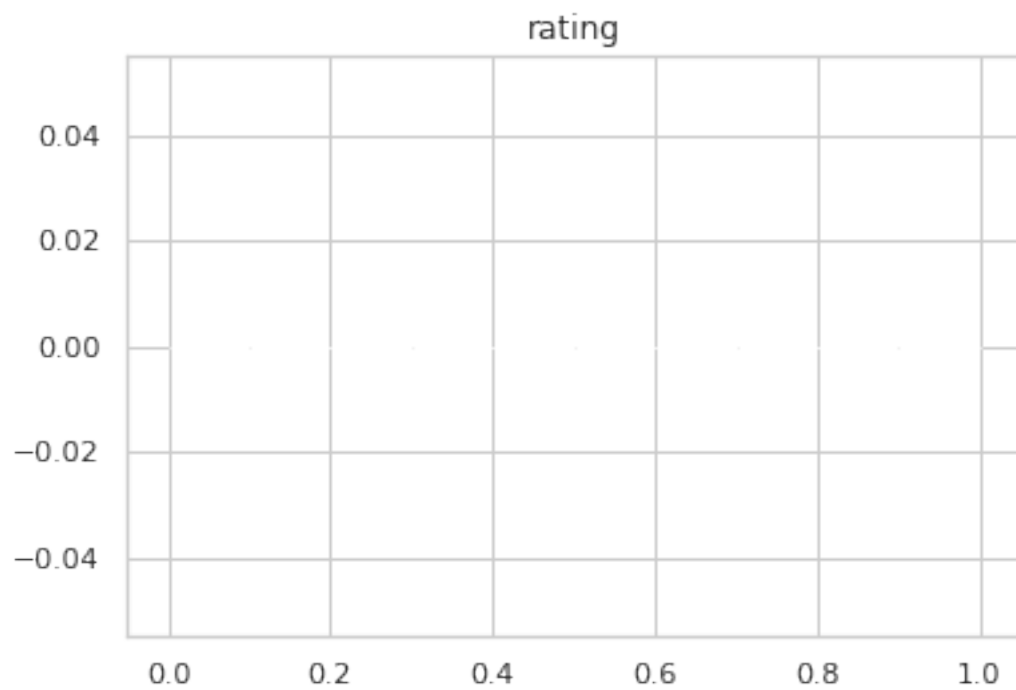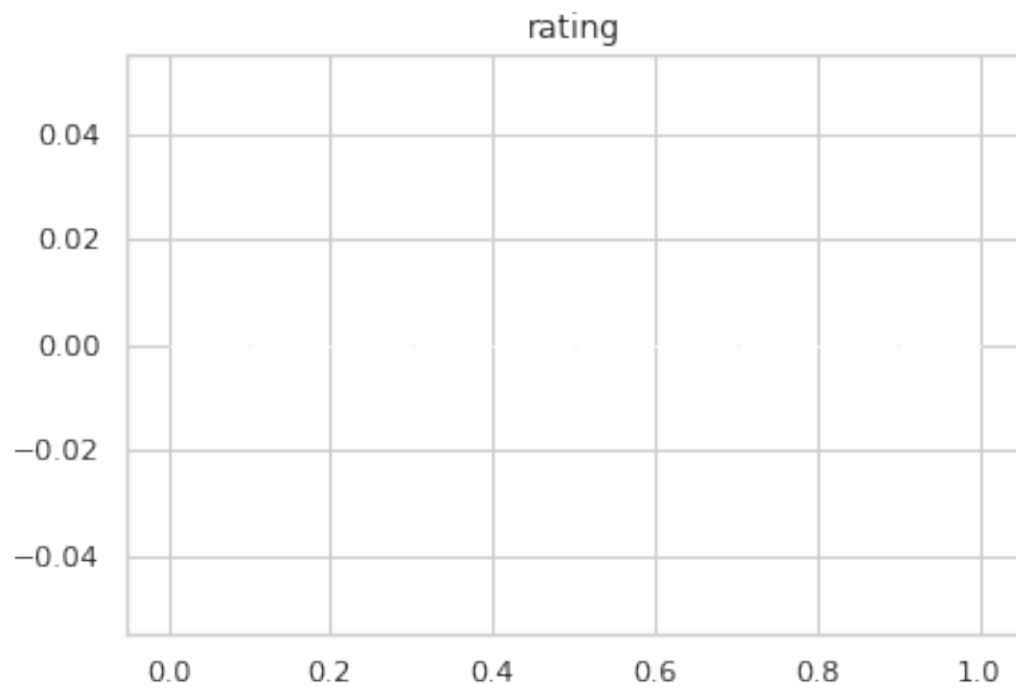
rating

rating

rating

rating

rating

rating
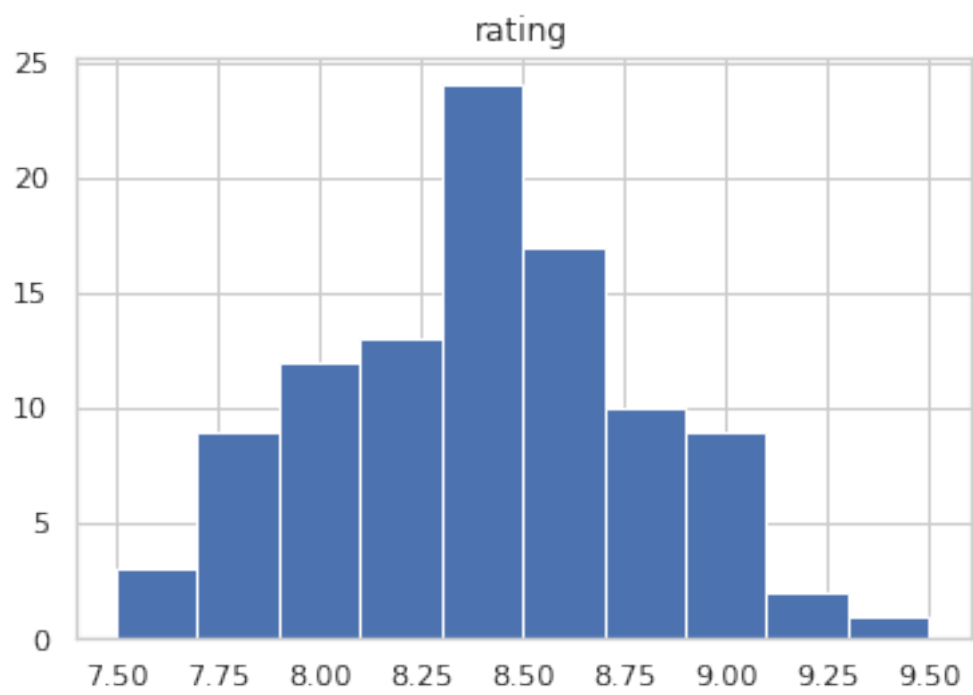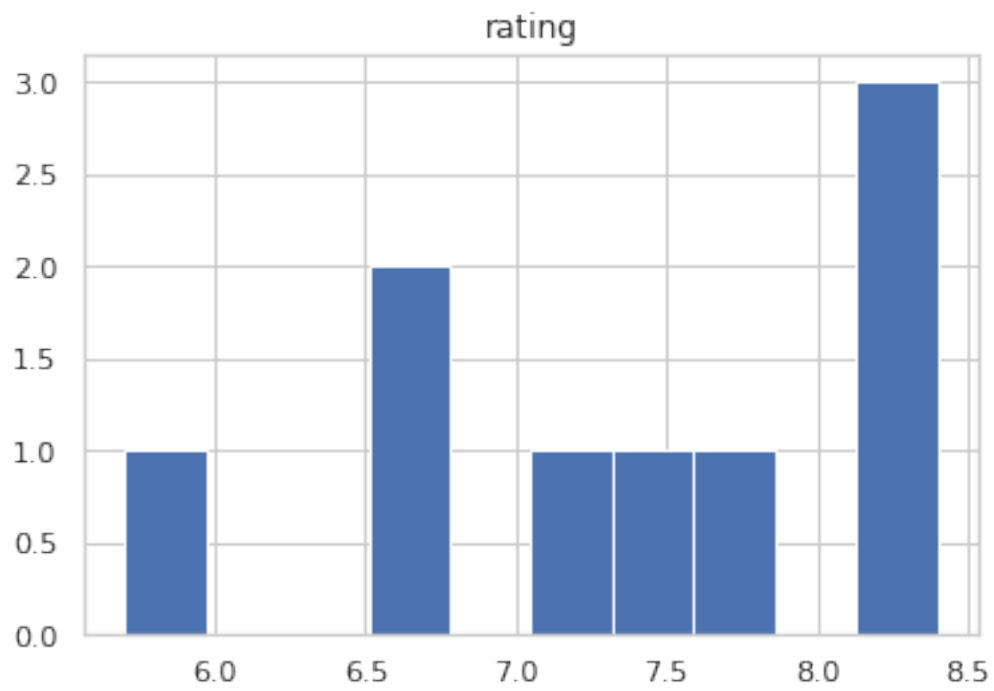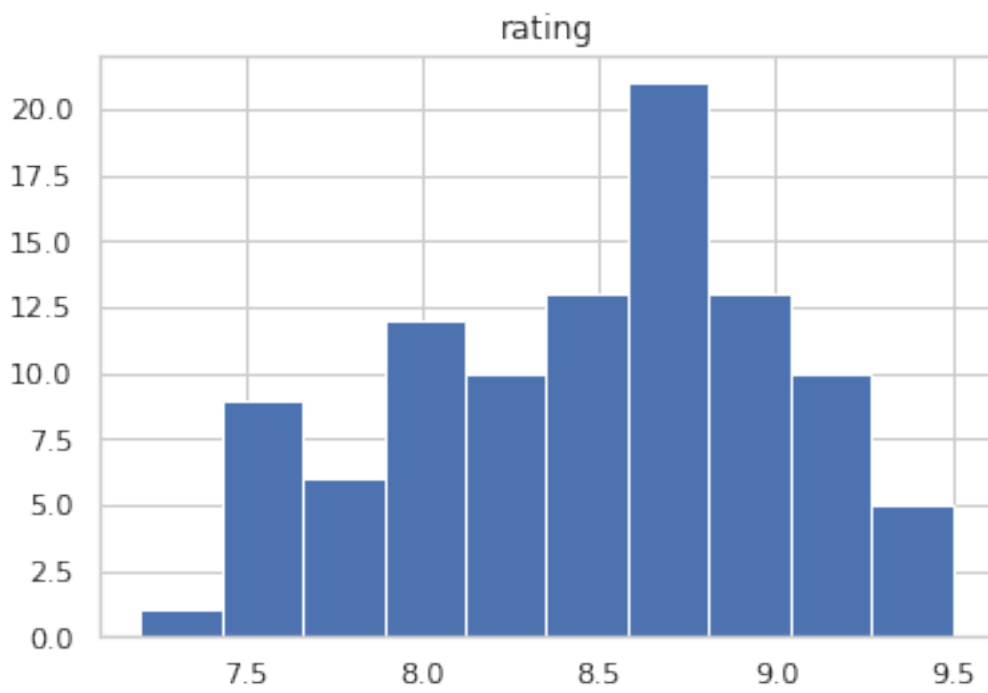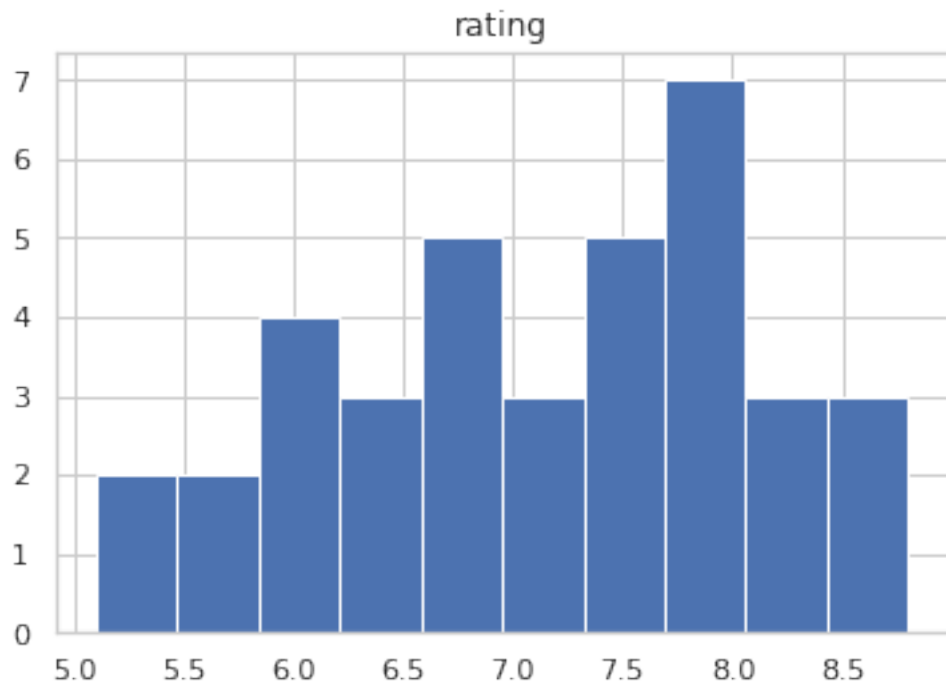
rating

rating

Table 2.1: Venues count by district and rating count

Police District rating	Bayview	Central	Ingleside	Mission	Northern	Park	Richmond
(5, 6]	1	0	4	0	0	0	0
(6, 7]	2	0	13	0	0	0	12
(7, 8]	3	24	14	24	28	31	12
(8, 9]	3	73	6	61	63	58	8
(9, 10]	0	3	0	15	9	11	0

Based on table 2.1 - the districts that have the highest rated venues are Central and Mission, however these are also heavily influenced by crime, as shown in figure 1.1. We need to select the neighborhood with best food and relatively low crime. Northern district is 2nd in terms of number of highly rated venues and 3rd in crime rate, however Park district has lowest crime yet pretty decent food, as well.

3 Conclusions

Based on analysis of the crime data and venue information provided via Foursquare API we were able to determine that the Park district is the safest (disregarding the kind of incidents, only looking at the volume) while having relatively decent restaurants.