NIGHT CLUB IN SAN FRANCISCO

Aruzhan Boken

Introduction

- **Business Problem**
- San Francisco became the World capital of IT innovations because of the Silicon Valley and millions of young people are craving to
- come here to build their careers and thousands of people are already came. Moreover, Stanford University is one of the best universities in the world and thousands of people come to study at Stanford each year. Thus, this project will focus on opening a night club for young people in San Francisco. We will not only focus on the most popular place but also focus on the safety since many people become drunk at night clubs and it's our duty to choose a safe place for them.
- **Target Audience**
- 1. A business entrepreneur that wants open business in San Francisco.
- 2. Business Analyst or Data Scientists, who wish to analyze the neighborhoods of San Francisco using python and jupyter notebook.

- **Data**
- In order to make a decision we will consider three types of datasets.
- 1. San Francisco Registered Business Data (https://data.sfgov.org/Economy-and-Community/Registered-Business-Locations-San-Francisco/g8m3-pdis)
- - It will help to know density of business nearby
- 2. San Francisco Crime Data(https://data.sfgov.org/Public-Safety/Police-Department-Incident-Reports-2018-to-Present/wg3w-h783)
- - It will help us to analyze the safety in different places in San Francisco
- 3. Foursquare Data
- - Foursquare data is robust and provides location data for Apple and Uber. Foursquare API allows you to retrieve information about the most popular spots in each neighborhood in San Francisco. This will be another good indication of foot traffic for particular venue types. Calling the Foursquare API returns a JSON file, which can be turned into a dataframe for analysis in python.

Crime Data

```
crime = pd.read_csv('https://data.sfgov.org/api/views/wg3w-h783/rows.csv?accessType=DOWNLOAD')
print(crime.shape)
crime.head()
```

```
(294865, 35)
```

	Incident Datetime	Incident Date	Incident Time	Incident Year	Incident Day of Week	Report Datetime	Row ID	Incident ID	Incident Number	CAD Number	Report Type Code	Report Type Description	Filed Online	Incident Code
0	2019/08/15 11:41:00 AM	2019/08/15	11:41	2019	Thursday	2019/10/01 02:06:00 PM	85424006374	854240	196208089	NaN	II	Coplogic Initial	True	6374
1	2019/09/17 10:00:00 PM	2019/09/17	22:00	2019	Tuesday	2019/10/02 10:01:00 PM	85426606374	854266	196208205	NaN	II	Coplogic Initial	True	6374
2	2019/10/04 02:25:00 PM	2019/10/04	14:25	2019	Friday	2019/10/04 04:13:00 PM	85442603474	854426	190746203	192772728.0	II	Initial	NaN	3474
3	2019/10/03 07:30:00 PM	2019/10/03	19:30	2019	Thursday	2019/10/03 11:25:00 PM	85419706244	854197	190744514	192764437.0	II	Initial	NaN	6244
4	2019/10/04 04:53:00 PM	2019/10/04	16:53	2019	Friday	2019/10/04 04:53:00 PM	85446351040	854463	190746532	192772932.0	II	Initial	NaN	51040

Registered Business Data

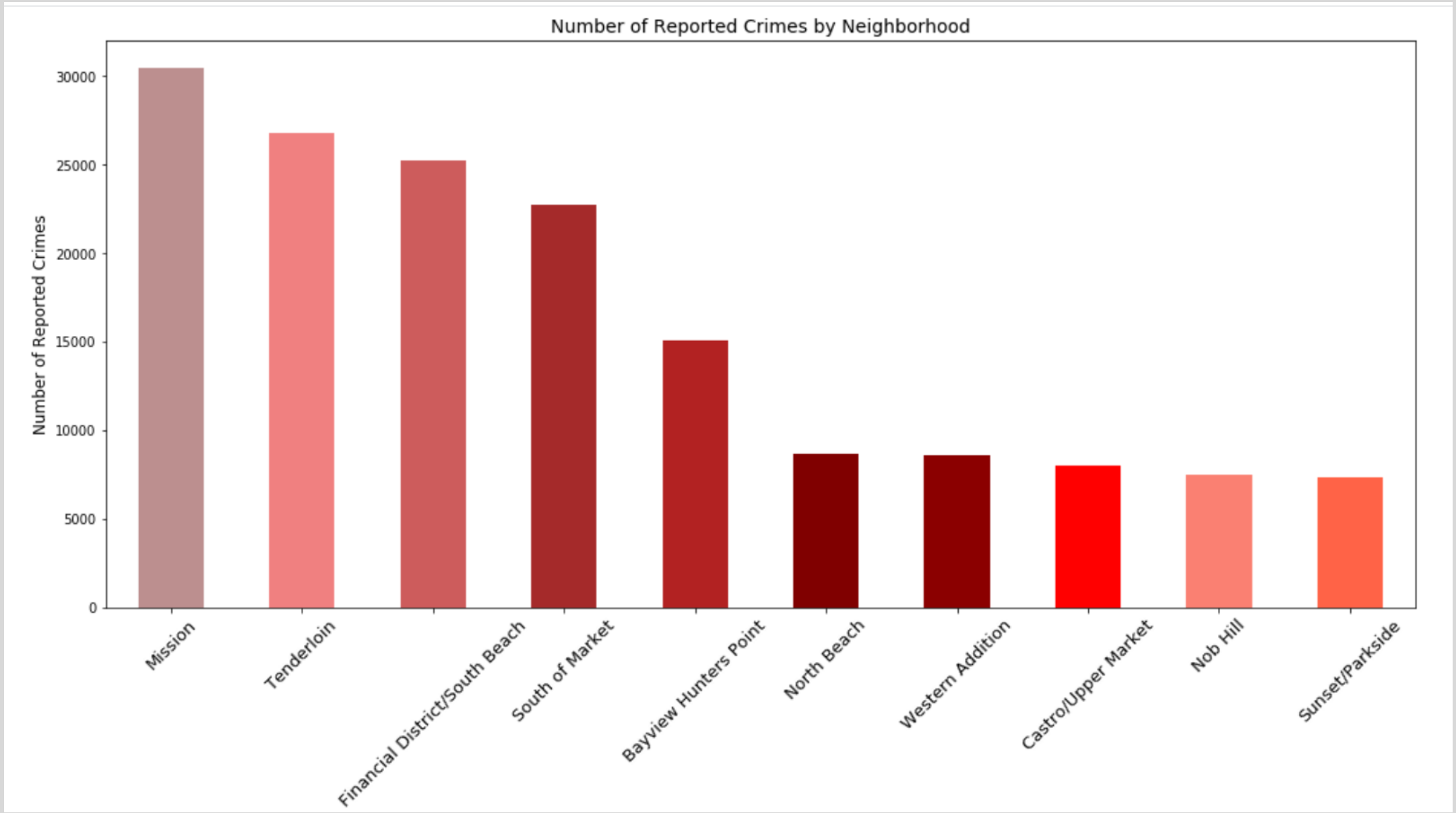
```
business = pd.read_csv('https://data.sfgov.org/api/views/g8m3-pdis/rows.csv?accessType=DOWNLOAD')
print(business.shape)
business.head()
```

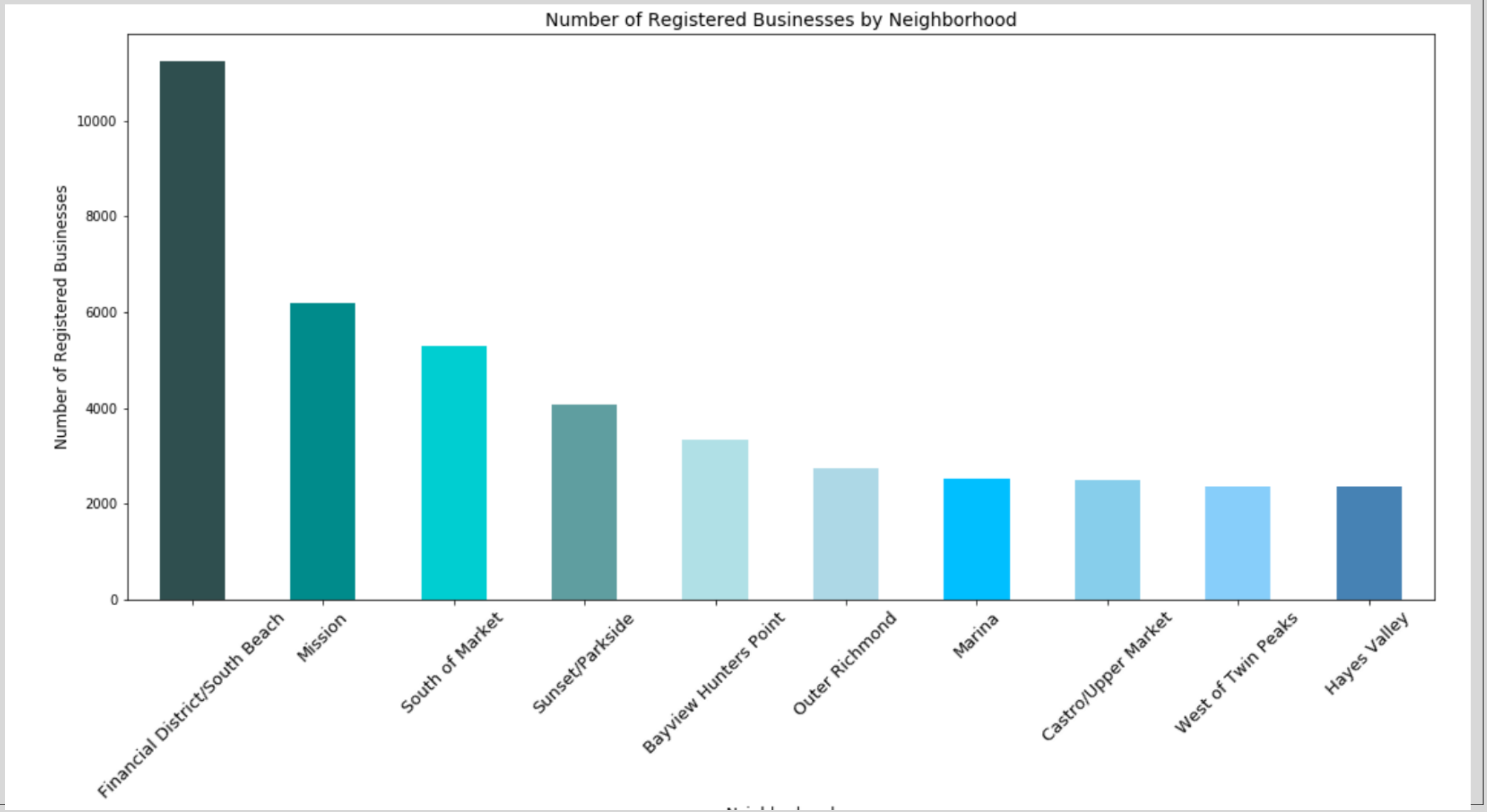
```
/opt/conda/envs/Python36/lib/python3.6/site-packages/IPython/core/interactiveshell.py:3020: DtypeWarning: Columns (12,13,14,15,21,25) h
ave mixed types. Specify dtype option on import or set low_memory=False.
  interactivity=interactivity, compiler=compiler, result=result)
```

```
(250406, 26)
```

	Location Id	Business Account Number	Ownership Name	DBA Name	Street Address	City	State	Source Zipcode	Business Start Date	Business End Date	Location Start Date	Location End Date	Mail Address	Mail City	Mail Zipcode	Ma St
0	1103593-08-161	1049564	Anjan Rajbhandari	Uber	28134 Harvey Ave	Hayward	CA	94544.0	03/24/2014	12/31/2017	03/24/2014	12/31/2017	NaN	NaN	NaN	Na
1	1218784-04-191	1100756	Luisa Alberto	High Five Sf	467 14th St	San Francisco	CA	94103.0	04/15/2019	04/15/2019	04/15/2019	04/15/2019	NaN	NaN	NaN	Na
2	1223199-05-191	1102424	Sunrun, Inc.	Sunrun, Inc	595 Market St	San Francisco	CA	94105.0	06/01/2008	06/01/2008	06/01/2008	06/01/2008	NaN	NaN	NaN	Na
3	1220748-05-191	1101579	Felix Hernandez	Tru-Tec Electric	44 Mcaker Ct	San Mateo	CA	94403.0	05/06/2019	06/18/2019	05/06/2019	06/18/2019	NaN	NaN	NaN	Na
4	1135452-02-171	1065102	Tirta Llc	Tirta	105 Hudson St 6s	New York City	NY	10013.0	06/09/2016	12/23/2018	06/09/2016	12/23/2018	NaN	NaN	NaN	Na

METHODOLOGY


Number of Reported Crimes by Neighborhood
Number of Reported Crimes
0
5000
10000
15000
20000
25000
30000
Mission
Tenderloin
Financial District/South Beach
South of Market
Bayview Hunters Point
North Beach
Western Addition
Castro/Upper Market
Nob Hill
Sunset/Parkside


Number of Registered Businesses by Neighborhood
Number of Registered Businesses
0
2000
4000
6000
8000
10000
Financial District/South Beach
Mission
South of Market
Sunset/Parkside
Bayview Hunters Point
Outer Richmond
Marina
Castro/Upper Market
West of Twin Peaks
Hayes Valley

The most suitable neighborhoods

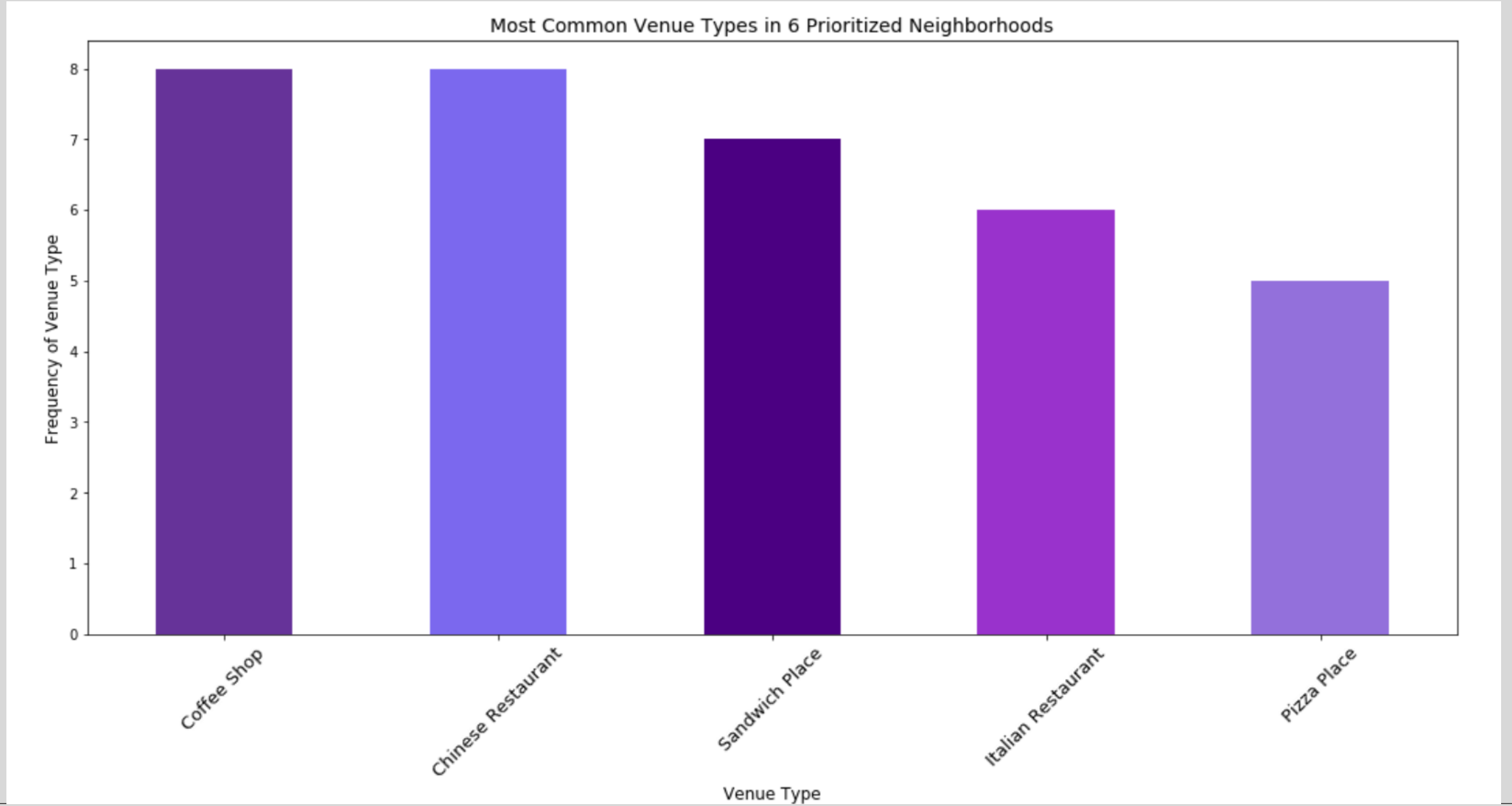
Merging Datasets

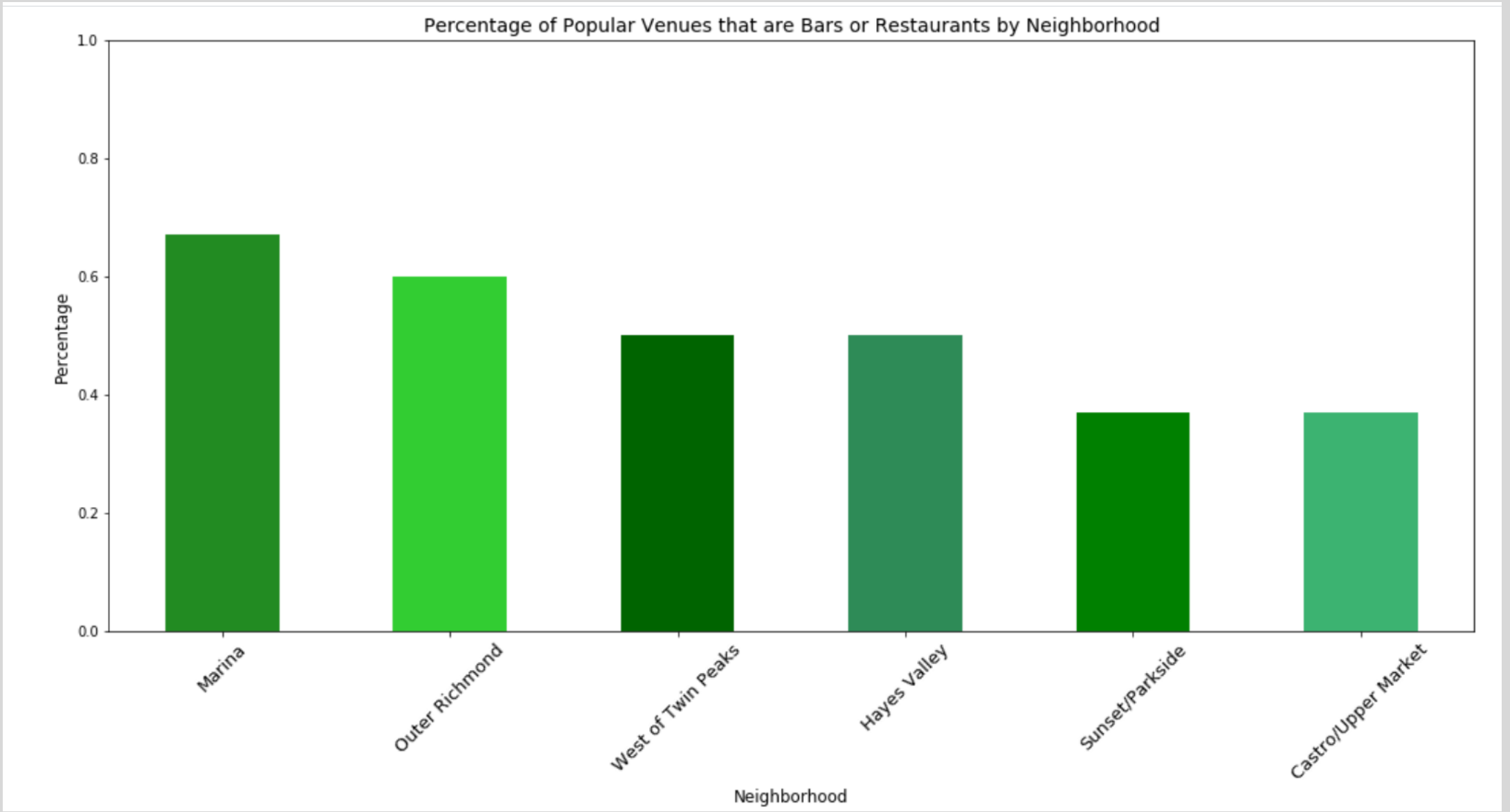
Now, we can add the Latitude and Longitude in separate columns and merge our datasets

```
SF = SF_Neighborhoods.merge(crime6, on=['Neighborhood'])
SF.rename(columns={'Incidents':'Crimes'}, inplace = True)
SF = SF[['Neighborhood', 'Businesses', 'Crimes','Coordinates']]
SF['Coordinates'].to_string()
SF[['Latitude','Longitude']] = SF['Coordinates'].str.strip('(').str.strip(')').str.split(', ', expand=True)

SF
```

	Neighborhood	Businesses	Crimes	Coordinates	Latitude	Longitude
0	Sunset/Parkside	4077	7344	(37.751616, -122.490810)	37.751616	-122.490810
1	Outer Richmond	2736	5657	(37.780001, -122.490229)	37.780001	-122.490229
2	Marina	2530	6205	(37.801406, -122.439718)	37.801406	-122.439718
3	Castro/Upper Market	2483	7994	(37.762932, -122.435395)	37.762932	-122.435395
4	West of Twin Peaks	2354	5106	(37.739871, -122.460106)	37.739871	-122.460106
5	Hayes Valley	2351	7164	(37.776685, -122.422936)	37.776685	-122.422936


Most Common Venue Types in 6 Prioritized Neighborhoods
Frequency of Venue Type
0
1
2
3
4
5
6
7
8
Coffee Shop
Chinese Restaurant
Sandwich Place
Italian Restaurant
Pizza Place
Venue Type


Percentage of Popular Venues that are Bars or Restaurants by Neighborhood
Percentage
1.0
0.8
0.6
0.4
0.2
0.0
Marina
Outer Richmond
West of Twin Peaks
Hayes Valley
Sunset/Parkside
Castro/Upper Market
Neighborhood

Results and Discussion

- Based on the data about crime rates and business registrations for every neighborhood in San Francisco, the best 6 neighborhoods were identified: Sunset/Parkside, Outer Richmond, Marina, Castro/Upper Market, West of Twin Peaks, Hayes Valley. According to analysis restaurants and bars are the most common vanues in these 6 neighborhoods.
- The neighborhoods were clustered in different groups based on their most popular venues. Thus, we have 4 clusters: Hayes Valley with Castro (), Outer Richmond and Marina, West of Twin Peaks, and Sunset.
- Also, it was found that Marina, Castro, and Hayes Valley neighborhoods compared to Sunset, Outer Richmond, and West of Twin Peaks.
- According to the analysis Outer Richmond and Marina are the best neighborhoods to open bars and restaurants based on the business registrations data. Marina is an expensive option for us since we are aiming to open a nigh club mainly for students and young professionals. However, since many people in San Francisco earn quite high salary regardless their age, we still can consider Marina. In order to decide between these to neighborhoods the budget of the night club rent should be considered.

Conclusion

- Overall, based on the three data sources the best neighborhoods to open a night club in San Francisco were successfully discovered. The main work on data analysis was performed and only consideration of the financial sides is left.