Siting a Pizza Place in San Francisco

Oluwagbemiga Ajayi

June 2020

Introduction

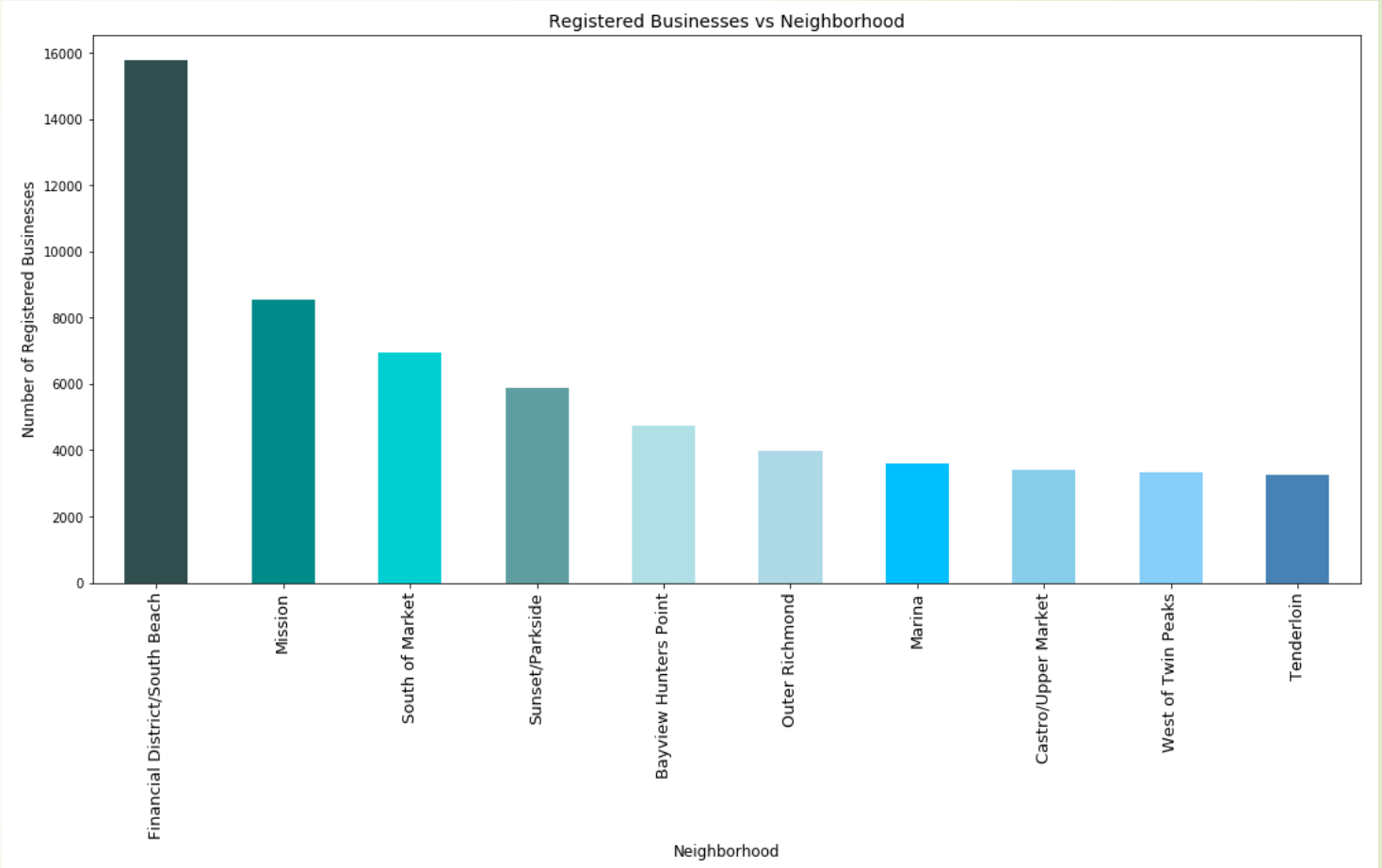
- The decision on where to site a business can go a long way to impact the success or failure of such business
- Maxwell's Pizza has been very creative in bringing something new to Pizza lovers in Sacramento and would like to get into San Francisco market
- Having learned from the bad experience from siting businesses in areas with high crime rates, Maxwell's is trying to be careful this time around
- While it's important to avoid areas with crime rates, a business should not be too far from where the patronage is.
- Maxwell's is confident of the market their pizza place is going to attract but they want to run their businesses in safe environments

Data for the Project

- The city of San Francisco has series of useful dataset which they have generously made available to the public.
- We'll be using the registered businesses: https://data.sfgov.org/Economy-and-Community/Registered-Business-Locations-San-Francisco/g8m3-pdis/data?no_mobile=true and crimes: https://data.sfgov.org/Public-Safety/Police-Department-Incident-Reports-2018-to-Present/wg3w-h783 datasets from the city of San Francisco.
- Through the foursquare API we get location and venue information

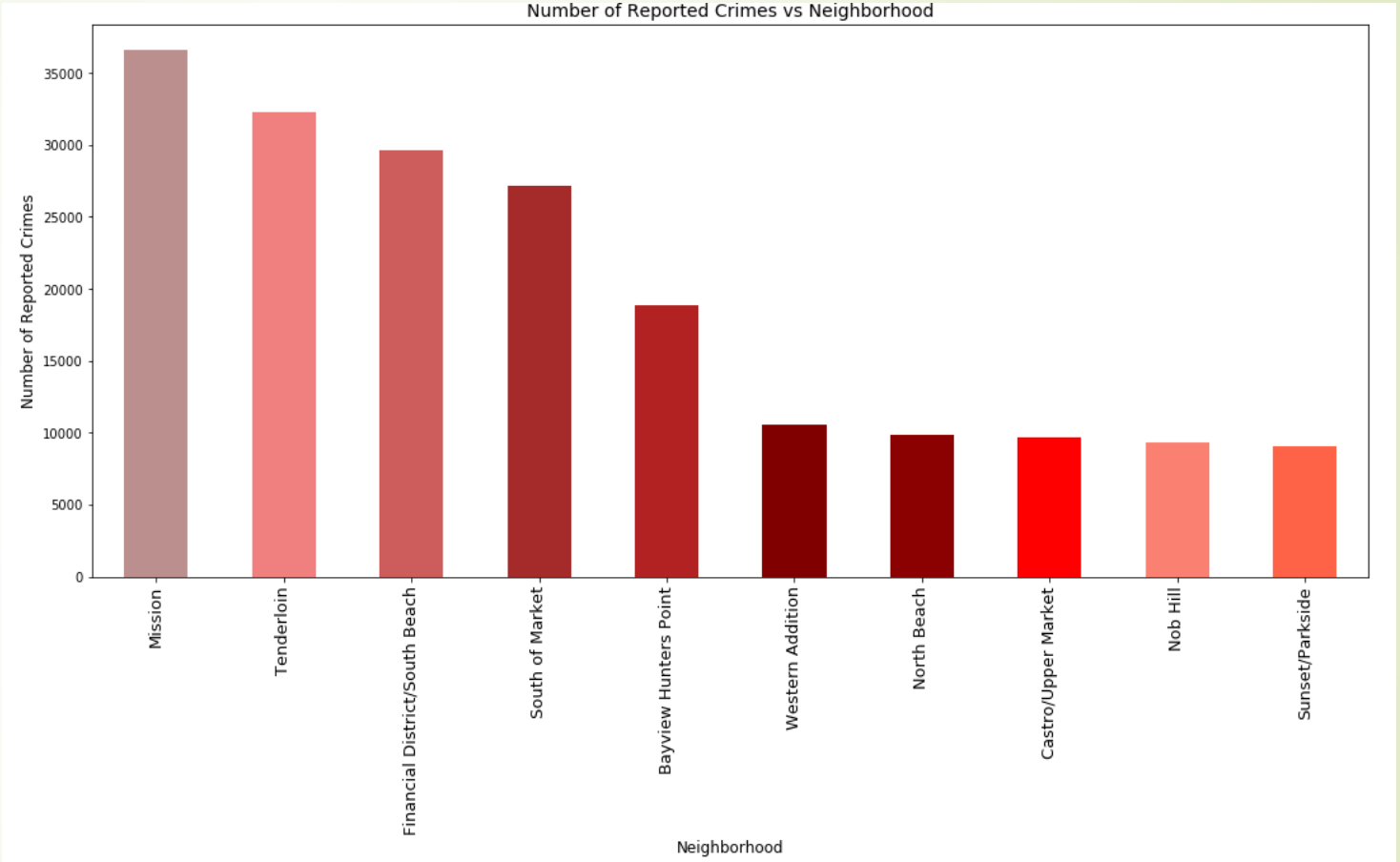
Registered Business

- We pulled data of registered businesses in San Francisco in the past 15 years (this was necessary to get enough data relevant to Pizza Places)
- Financial District neighborhood by far gets a large share of businesses in San Francisco



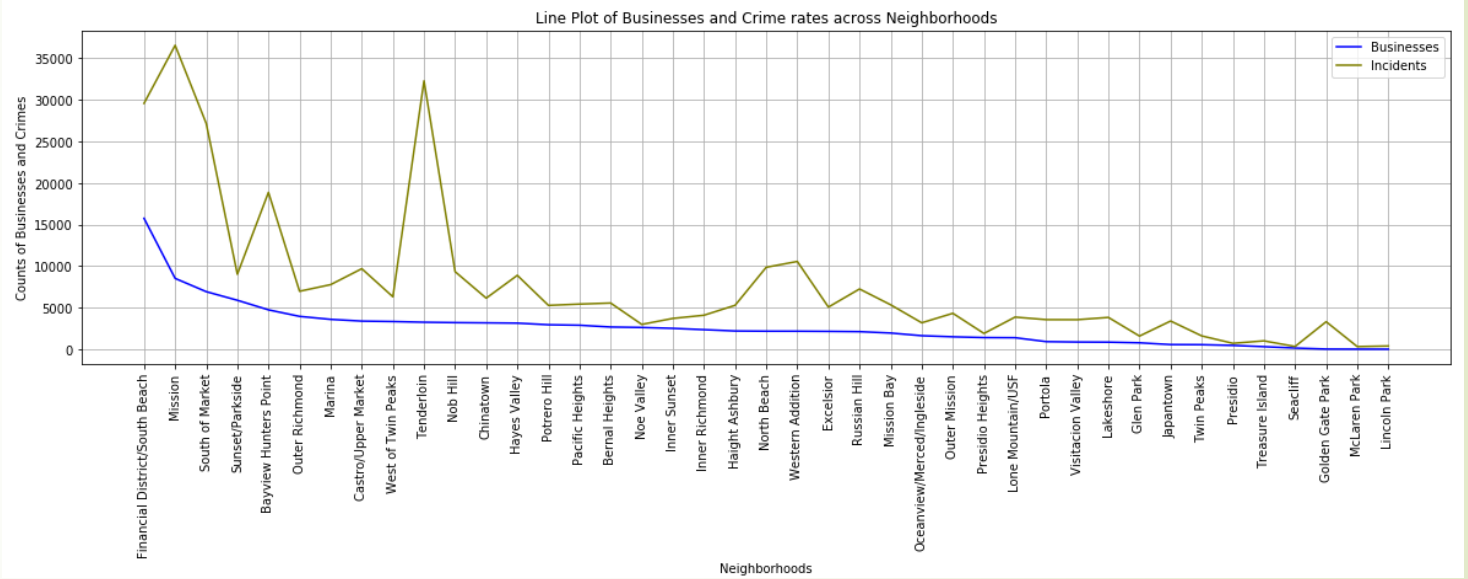
Crimes

- We pulled crimes data in the past 5 years in order to ensure we are working with crimes relevant to our time.
- Plotting number of crime incidents across neighborhood shows Mission to be a neighborhood with the highest crime rate.
- 5 neighborhoods clearly have more crime rates than the other 5



Business-Crime Combination

- We can merge the top crime and businesses data to see how far apart crimes count differ from businesses count for each neighborhood
- By far Mission neighborhood has the poorest combination of business and crime; the rate of business in Mission trails that of Financial district but crime rate in Mission clearly exceeds the rate in financial district



Business-Crime Combination *cont'd*

- On the left we have the 5 neighborhoods with highest crime rates and on the right we have the other 5 neighborhoods with good business attraction and not too high crimes
- Eyeballing the previous multiple line plot we might observe that Sunset/Parkside shows a good blend decent business attraction and minimal crime rates
- This seems to agree with the table we have here as we have Sunset/Parkside on the left

	Neighborhood	Incidents		Neighborhood	Businesses
18	Mission	36566	34	Sunset/Parkside	5900
35	Tenderloin	32291	25	Outer Richmond	3963
5	Financial District/South Beach	29585	16	Marina	3602
33	South of Market	27165	2	Castro/Upper Market	3393
0	Bayview Hunters Point	18877	39	West of Twin Peaks	3342

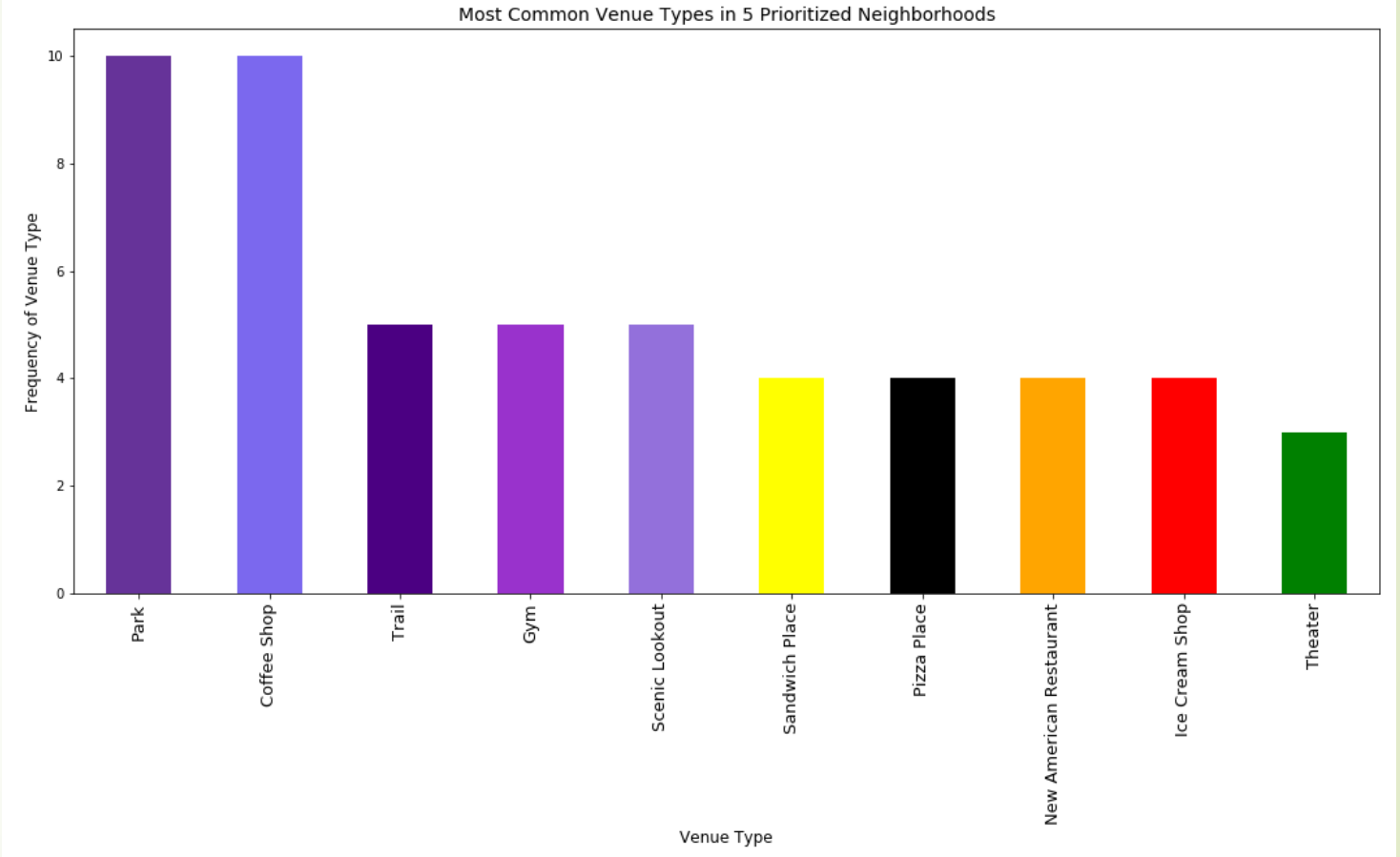
Geolocation Data

- Here we got coordinates of each of our selected neighborhoods pulling the values from https://latlong.net which provides a free service for longitude and latitude of locations
- Geocoding API of google could have been used but it attracts a cost
- We then merged our business, crime and coordinates data

	Neighborhood	Businesses	Crimes	Coordinates	Latitude	Longitude
0	Sunset/Parkside	5900	9029	(37.749690, -122.494880)	37.749690	-122.494880
1	Outer Richmond	3963	6982	(39.820580, -84.887010)	39.820580	-84.887010
2	Marina	3602	7782	(37.803850, -122.450350)	37.803850	-122.450350

Foursquare API

- Taking 1 mile (1609.34 m) radius away from our neighborhoods, we found venues around each neighborhood
- Most common venues are Parks, Coffee shop. Trail, Gyms and scenic look out follow



Common Venues

- These 2 neighborhoods have in their top common venues Pizza places, the other neighborhoods do not have Pizza places in top 20 common venues

----Outer Richmond----

	venue	freq
0	Pizza Place	0.10
1	Sandwich Place	0.07
2	American Restaurant	0.07
3	Video Store	0.07
4	Theater	0.07
5	ATM	0.03
6	Discount Store	0.03
7	Restaurant	0.03
8	Pub	0.03
9	Fast Food Restaurant	0.03
10	Food	0.03
11	Coffee Shop	0.03
12	Pharmacy	0.03
13	Gas Station	0.03
14	Gym	0.03
15	History Museum	0.03
16	Ice Cream Shop	0.03
17	Dance Studio	0.03
18	Rock Club	0.03
19	Bistro	0.03

----Sunset/Parkside----

	venue	freq
0	Coffee Shop	0.10
1	Chinese Restaurant	0.07
2	Ice Cream Shop	0.07
3	Playground	0.07
4	Trail	0.07
5	Dumpling Restaurant	0.07
6	Dive Bar	0.03
7	Korean Restaurant	0.03
8	Grocery Store	0.03
9	Pizza Place	0.03
10	Record Shop	0.03
11	Empanada Restaurant	0.03
12	Sandwich Place	0.03
13	Dim Sum Restaurant	0.03
14	Cocktail Bar	0.03
15	Liquor Store	0.03
16	Taco Place	0.03
17	Vietnamese Restaurant	0.03
18	Beach	0.03
19	Beer Bar	0.03

Kmeans Clustering

- It's clear already that Outer Richmond and Sunset/Parkside have in common some degree of Pizza Place popularity which suggests that there is a thriving market for Pizzas in these 2 neighborhoods
- With Pizza's Place being the most popular venue in Outer Richmond, we'd expect this neighborhood to have a good population of Pizza lovers who are going to appreciate what Maxwell's has to offer
- Lastly, we can cluster our neighborhoods using k-means to see if clustering algorithm agrees with the intuition from our analysis
- Cluster labels from Kmeans gets mapped to the neighborhood and we found Outer Richmond and Sunset/Parkside to be in the same cluster

Conclusion

- In this data science capstone project, we have used common python libraries to manipulate data sets and Foursquare API to explore the neighborhoods of San Francisco
- We used K-means clustering algorithm to verify the intuition we found through analysis which supported our observation. We would have liked to use the geocoding API but it's not available for free
- In the future we would attempt a time series analysis of businesses and crimes in neighborhoods so we can observe the trend and guard against basing our judgement on trends overall history which could have changed