STARTING A NEW PIZZA PLACE?

BUT WHERE IN BAY AREA?

Problem at hand

- Market competition is sky rocketing.
- Demand and competition varying from location to location.
- A deep analysis of the location need to be done before starting a new business.

Who will be interested

- People looking to open a new pizza place.
- Small pizza restaurants looking for expansion.
- Established restaurants to explore new opportunities.

Data Acquisition

- “San Francisco Neighborhoods as ZIP Codes" from San Francisco Burden of Disease & Injury Study is used to get the zip codes and population.
- “pgeocode” library is used to get the geo data of the neighborhood.
- Foursuqre api is used to obtain data related to venues of each neighborhood.

Data cleaning and preprocessing

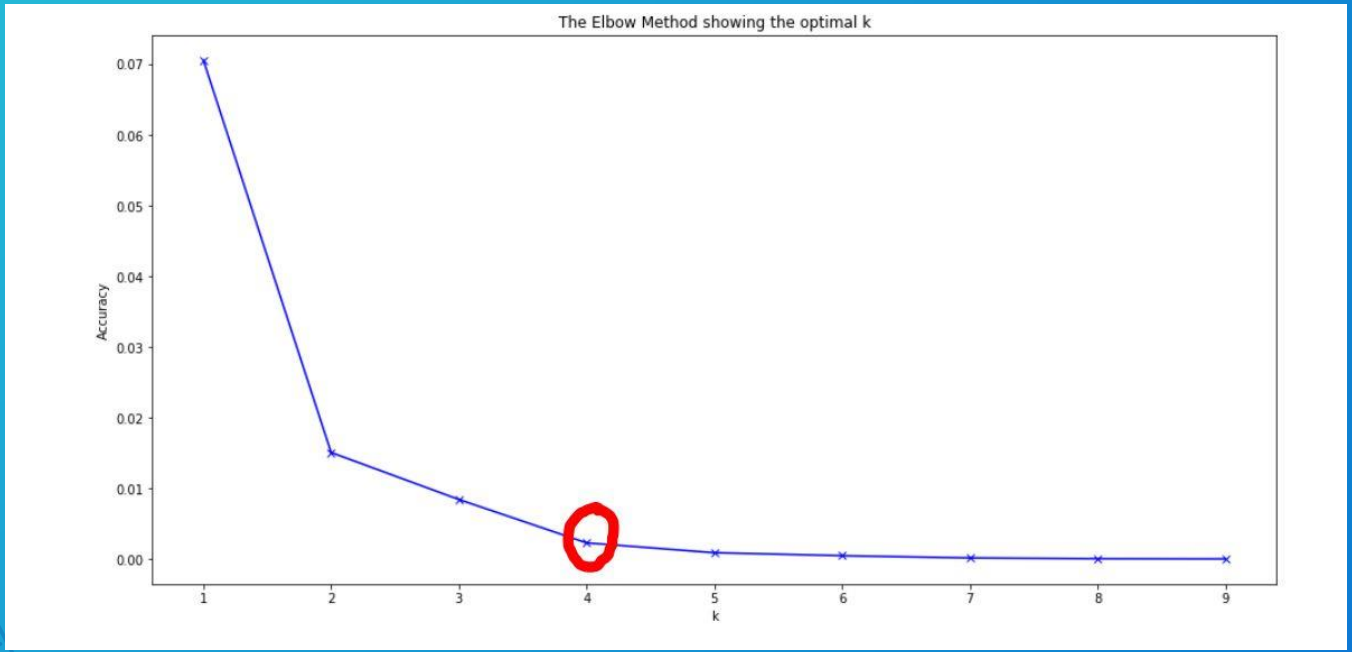
- Venues marked as “pizza place”, “bakery”, and “Italian Restaurant” are filtered and a new dataset is created as “No of pizza places” with the total number of above venues in each neighborhood.
- The ration between number of pizza places and population is calculated and added to the data frame.

Neighborhood	Population	Pizza Place Count	num/pop	Postal Code	latitude	longitude
South of Market	23016	1.0	0.043448	94103	37.7725	-122.4147
Potrero Hill	17368	1.0	0.057577	94107	37.7621	-122.3971
Ingelside-Excelsior/Crocker-Amazon	73104	3.0	0.041037	94112	37.7195	-122.4411
Parkside/Forest Hill	42958	2.0	0.046557	94116	37.7441	-122.4863
Haight-Ashbury	38738	2.0	0.051629	94117	37.7712	-122.4413
Outer Richmond	42473	2.0	0.047089	94121	37.7786	-122.4892
Sunset	55492	2.0	0.036041	94122	37.7593	-122.4836
Bayview-Hunters Point	33170	1.0	0.030148	94124	37.7309	-122.3886
Lake Merced	26291	1.0	0.038036	94132	37.7211	-122.4754

Clustering

- K-Means clustering is used.
- Neighborhoods are clustered according to the ratio between number of pizza places and population.
- The optimal number of clusters is found using “elbow joint" method.

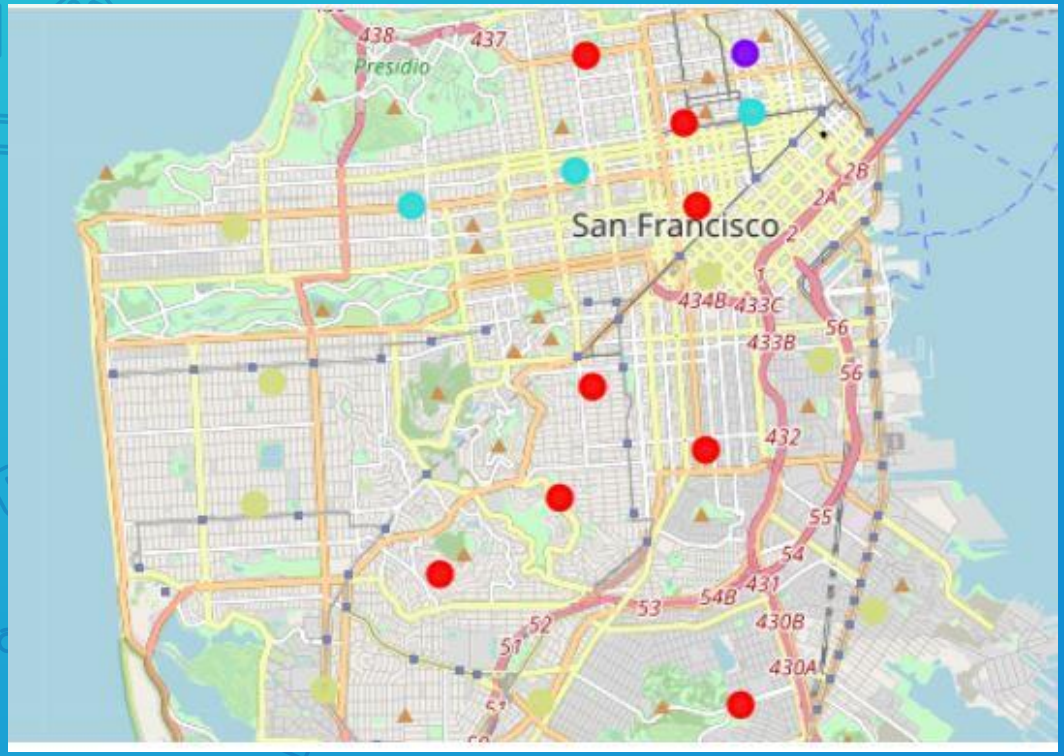
Elbow joint method



Results

- 4 Clusters have been identified.
  - Cluster 0 : Little or no competition
  - Cluster 2 : High competiton
  - Cluster 2 : Medium competition
  - Cluster 3 : Small competition

Neighborhoods after clustering



- Cluster 0
- Cluster 3
- Cluster 2
- Cluster 1

Discussion

- Cluster 0 is identified as the best option for a new pizza place to be stared as it has the lowest competition.
- Cluster 1 has a heavy competition thus not suitable for a new comer.

Conclusion

- Only population and number of existing pizza places are considered.
- More factors affect the success of a business.
  - GDP
  - Cost of living
  - Taxes
  - People's interests
- More updated data is necessary.

Thanks!

Any questions?

You can find me at herath.charithjay@gmail.com