WHERE TO OPEN A BAR IN THE BAY AREA

Matt Vestal

IBM data science capstone

RATIONALE

- The Bay Area is experiencing population growth due to the tech boom

- More people coming in means more demand for local businesses

  - One profitable business is a bar, but where is the best place to open one with minimal competition?

PROBLEM

- How do we determine which places have the least amount of bars in their top 50 venues?
 - Location data through web scraping and Foursquare API calls

DATA REQUIRED

- Location data

- Coordinates

- Venue data

- Data will be sourced from Wikipedia and Foursquare API

METHODOLOGY

- Use web scraping to obtain list of cities and population of cities in the Bay Area
  - This is done with beautifulsoup
- Use Foursquare API call to obtain venue data for top 50 venues
  - Foursquare limit is 50

- Apply K-means clustering to determine best place to open a bar

RESULTS
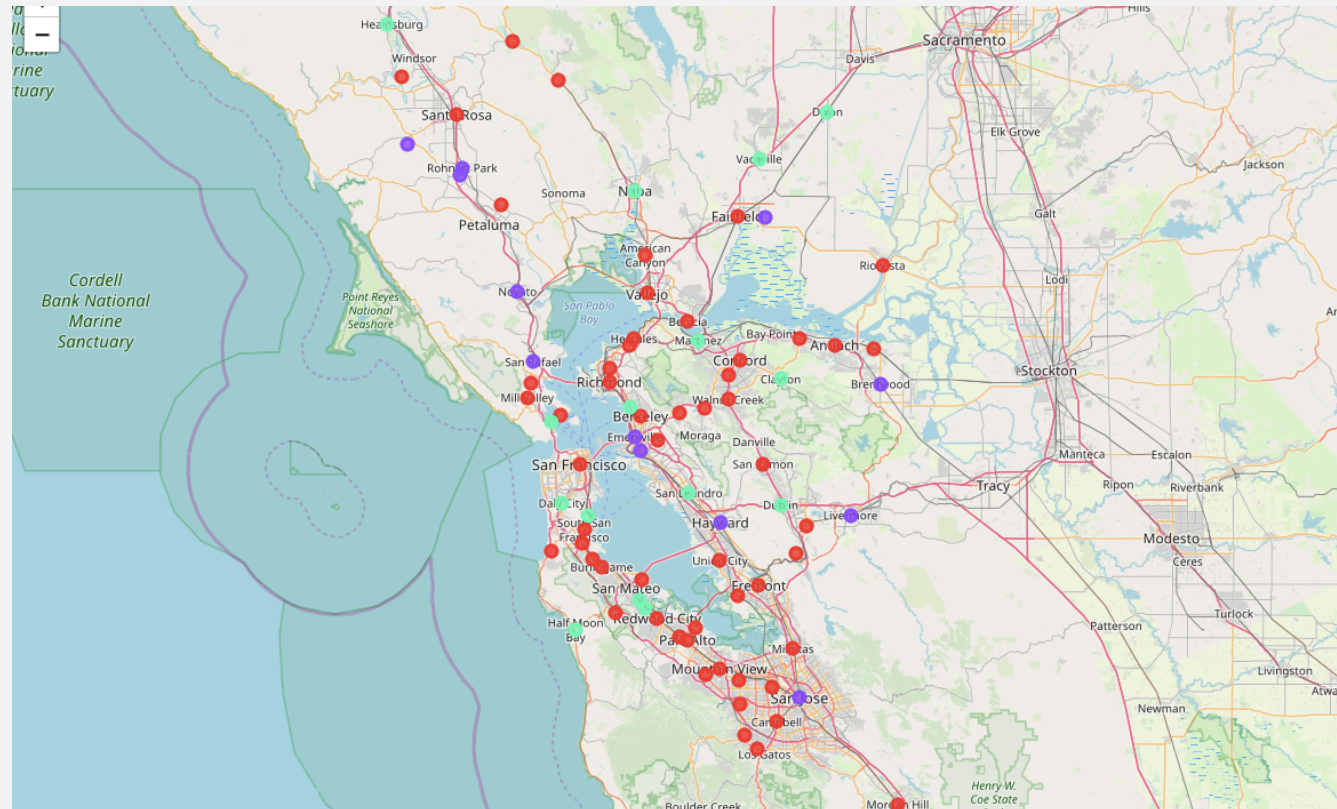
RESULTS

- The clusters were grouped as such:

-

- Cluster 0: the cluster with the least bars (red)
- Cluster 1: the cluster with the most bars (purple)
- Cluster 2: the cluster with a moderate amount of bars (green)

- The mean populations in clusters 0, 1, and 2 were 67501, 151,393, and 38,215 respectively.
- Cluster 0 had 11 major cities, cluster 1 had three major cities, and cluster 2 had just one major city.

DISCUSSION

- It appears that cluster 0 is the best place to open a bar due to the low frequency. However there is a large number of major cities here

  - Probably due to the low limit of the API call for venues

  - Despite this, its probably okay to open a bar in a cluster 0 city as long as it isn't a major city

- Cluster 2 has a moderate amount of bars but has a very low mean population. This suggests a high bar to person ratio, and it is not advisable to open a bar here

CONCLUSION

- Cluster 0 cities are the best cities to open a bar, provided they are not major cities (population < 100,000).

- It is not recommended to open a bar in cluster 1 cities due to the high number of bars and cluster 2 cities due to a high bar to person ratio