San Francisco Bay Area New Bubble Tea Shop Classification

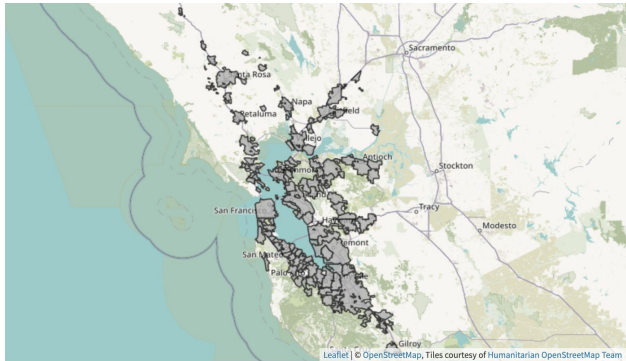
Rushil Jayant

Purpose Of Project

- Bubble Tea, known as Boba for short is a Taiwanese tea-based drink invented in Taiwan in the 1980s
- An **Interested Entrepreneur** wants to open a new Bubble Tea Shop in the San Francisco Bay Area because it is a booming trend.
- The Entrepreneur wants to **Maximize Market Share** by finding out the ideal city to open his Bubble Tea Shop
    - The City Must already have at least one existing Bubble Tea Shop
- He believes that using the tools of Unsupervised ML he can find the best City in the Bay Area

Sources of Data

- US CITIES DATA https://simplemaps.com/data/us-cities
- Stanford GeoJson file of Bay Area Cities
    - Used in Visualizations of the population, Bubble Tea shops, and classifications of Cities for this project
- List of Towns and Cities: https://en.wikipedia.org/wiki/List_of_cities_and_towns_in_the_San_Francisco_Bay_Area
- Foursquare API for Venues

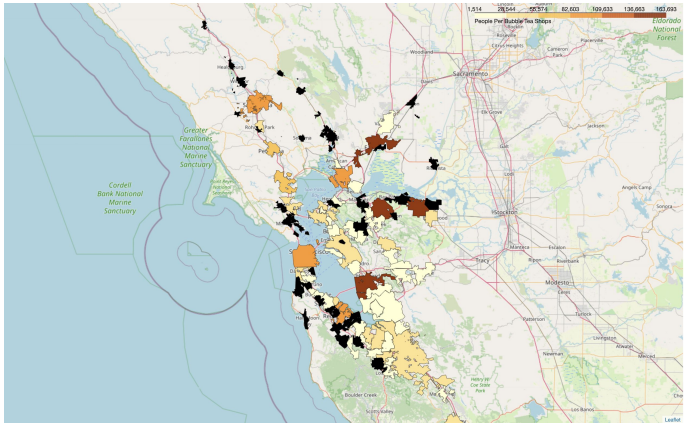


Cleaned and Prepared Data

- A DataFrame called *USCities* was used to
  - Contain all the cities in the Bay Area and their properties
  - Used to look up Bubble Tea Shops in the Foursquare API
- A DataFrame called *BobaShops* contained
  - Boba Shops in the Bay Area
  - Coordinates to different Bubble Tea Shops
  - Addresses to different Bubble Tea Shops
- A DataFrame called *BobaCityCounts* contained
  - Cities which had Bubble Tea Shops
  - Number of Bubble Tea Shops in each city
  - Population, Density, and People Per Bubble Tea Shop in each city

Preprocessing Using Geospatial Visualization

- Used four Geospatial Visualizations (in each city of the Bay Area):
  - Population
  - Density
  - Number of Bubble Tea Shops
  - People per Bubble Tea Shop
- Most Important: **People per Bubble Tea Shop**
  - Shows the cities that do not have Shops in Black
  - Darker Cities (non black) have more people per shop
  - More people per shop – More Potential Customers
  - Important to look at clusters which maximize this



Modeling using Unsupervised Learning

- Used KMeans Clustering to cluster the different cities based on (in each city):
  - Population
  - Density
  - People Per Bubble Tea Shop
- Total of 8 Clusters were produced and were manually classified into
  - Best
  - OK
  - Bad
- These ratings were the results of the Modeling

Results

- Cities with **Best** classification
  - **San Francisco**
  - **Concord**
  - **Fremont**
  - **Santa Rosa**
  - **Antioch**
- If for some reason the entrepreneur does not want the cities with the Best classification, there are several OK cities
- There are some cities that are definitely to be avoided (see Notebook)

Discussion/Ways to Expand Project

- Can Use this same process to other cities and other investors
- Use More Data like:
  - Traffic Data
  - Youth Population
  - Social Media
  - Individual Ratings
  - Etcetera
- Could use the different data and use different ML models like:
  - Logistic Regression
  - Classification

Conclusion

- Used Unsupervised Machine Learning models to predict the cities that are the best to open a new Bubble Tea Shop
- Definitely Made **Better, more Accurate, and Reliable** results than traditional processes.
- Might not be perfect, but provided better insight for the investor.