IBM Applied Data Science

Capstone Project

Opening a Boba Tea Place in San Fernando Valley

By Scout Zhou

May 29, 2020

Introduction - Boba Milk Tea

- Milk tea + Chewy tapioca balls
- Boba milk tea or bubble milk tea has been the most trending food in Eastern Asian since early 2000
- Originated from Taipei
- Increasingly getting more attention globally in recent years.
- CoCo Fresh Tea&Juice, Gong Cha, Yi Fang Taiwan Fruit Tea, and etc.

Business Problem

It will be beneficial for people to consider taking advantage of this trend and open a Boba milk tea shop in San Fernando Valley. However it brought out a critical and essential question. Where should we invest to open a new Boba milk tea shop.

An ideal location will satisfy two conditions:

- Higher Asian population: Boba milk tea is mostly approved by the Eastern Asian group
- More Crowded Region: Better business opportunity in general
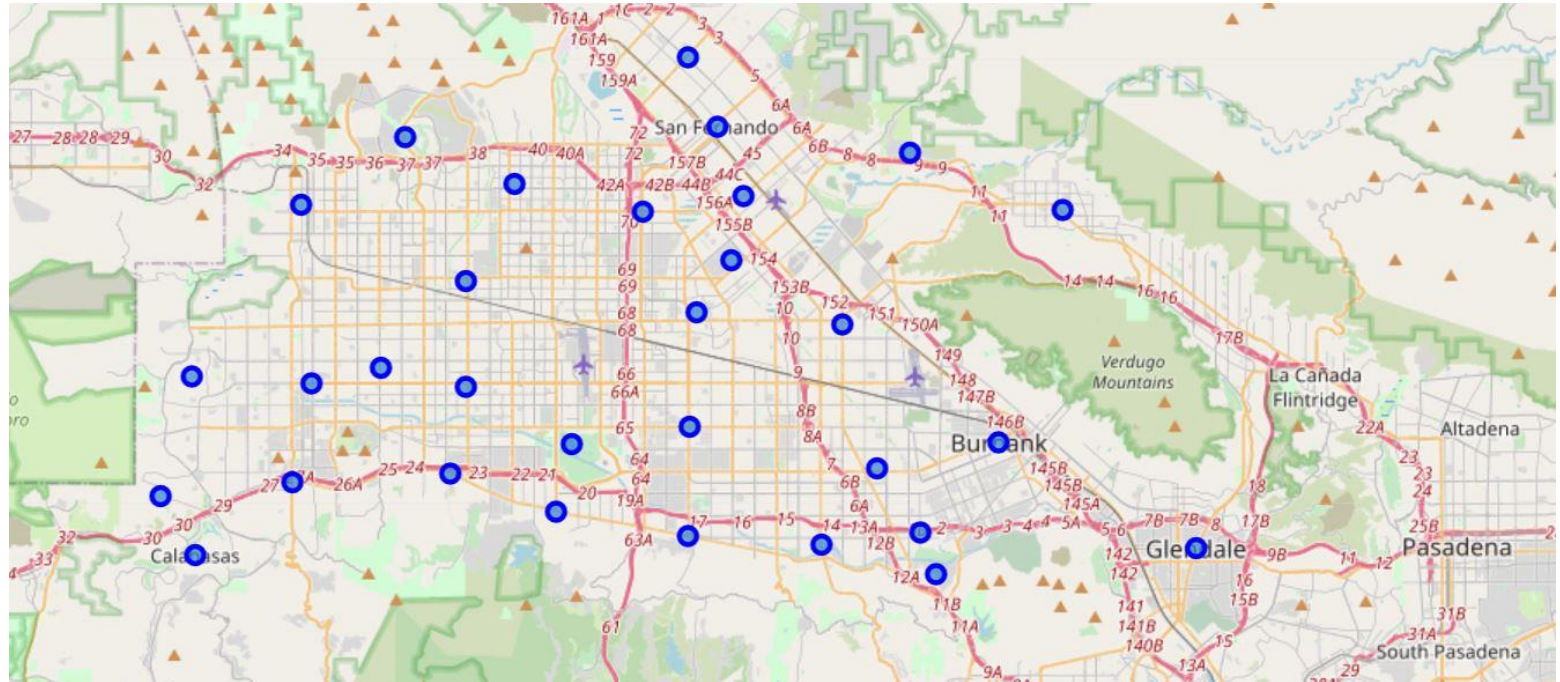
Data Acquisition

The data we needed for this project can be generally sorted into three categories, including

- The names of neighborhoods of the San Fernando Valley will be used to determine the scope of this project. The list of neighborhoods in San Fernando Valley is provided by Wikipedia.
- The location of neighborhoods within the scope, including the Longitude and Latitude, will be used for visualization purposes.
- The environment of each neighborhood, especially the venue's information in each region obtained by Foursquare.
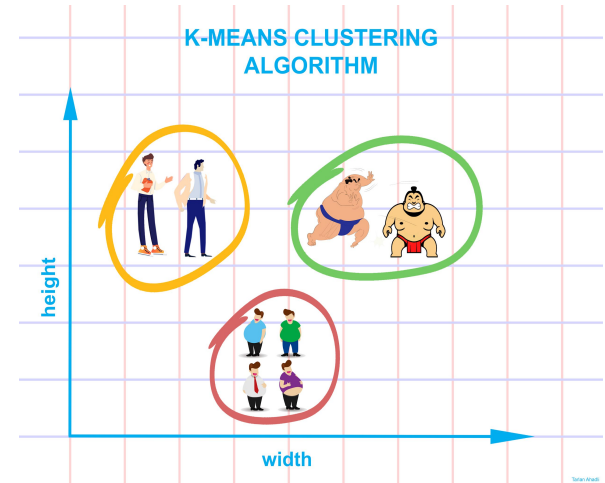
Obtained Data

Data Analysis - size of Asian population

For this project, the size of Asian population in each region is evaluated by the number of Asian Restaurants. The highest-ranked neighborhoods are listed in the provided figure.

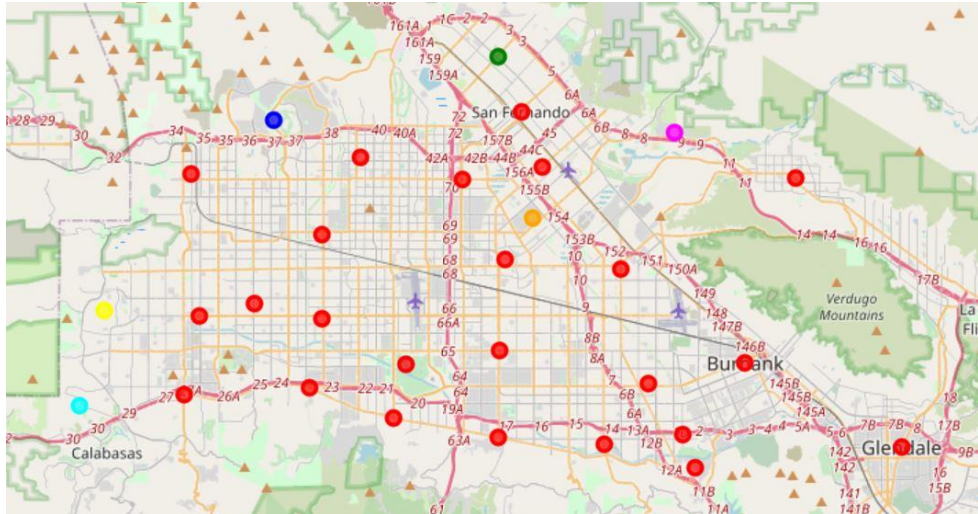|	Neighborhoods	Asian Restaurant
182	North Hollywood, Los Angeles	1
234	Northridge, Los Angeles	1
577	Woodland Hills, Los Angeles	1
20	Canoga Park, Los Angeles	1
74	Glendale, California	1
220	Northridge, Los Angeles	1
245	Panorama City, Los Angeles	1
478	Toluca Lake, Los Angeles	1
15	Canoga Park, Los Angeles	1
413	Sunland-Tujunga, Los Angeles	0
409	Sunland-Tujunga, Los Angeles	0
410	Sunland-Tujunga, Los Angeles	0
411	Sunland-Tujunga, Los Angeles	0
412	Sunland-Tujunga, Los Angeles	0
0	Arleta, Los Angeles	0

Data Analysis - business active regions

- The neighborhoods are clustered by k-mean clustering based on the venue's information
- K-mean clustering: "an iterative algorithm that tries to partition the dataset into K pre-defined distinct non-overlapping subgroups (clusters) where each data point belongs to only one group."
- The neighborhoods are divided into 7 clusters
- Based on their most popular venues

K-MEANS CLUSTERING ALGORITHM

Continue

❖ map demonstrates the distribution of each cluster in San Fernando Valley.

❖ The crowded region are marked with red or orange dot

Discussion and Conclusion

- The number of clustering can be reduced to simplify the analysis process
- The size of Asian population has more significance
- opening a Boba milk tea shop in any region with a higher Asian

  population will be an excellent choice.

- North Hollywood, Northridge, Woodland Hills, and etc.

Reference

A brief history of Boba Milk tea

https://www.foodandwine.com/tea/bubble-tea-taiwanese-street-drink-turned-american-addiction

Lab: Segmenting and CLustering Neighborhoods in New York City

https://labs.cognitiveclass.ai/tools/jupyterlab/lab/tree/labs/DP0701EN/DP0701EN-3-3-2-Neighborhoods-New-York-py-v1.0.ipynb

Wikipedia page-San Fernando Valley

https://en.wikipedia.org/wiki/Category:Communities_in_the_San_Fernando_Valley