



NEW BUBBLE TEA SHOP LOCATION PREDICTION IN SANTA CLARA COUNTY

ZIJIAN ZHU

BUBBLE TEA

Bubble Tea is one the most famous and well known Taiwanese drinks in Asia and even the worldwide. Along with the economic booming for the past 10 years, San Francisco Bay Area, as the central of the Technology and Innovation of the world, attracted millions of immigrants moved in and became one of the highest density areas with Asian population in the US. Hundreds of Bubble Tea Shops opened in recent decade around the Bay Area.



PROBLEM

This project targets to look for some ideal locations to open Bubble Tea shop in Santa Clara County (the heart of Silicon Valley) which might potentially have a higher than average profit and less competition.

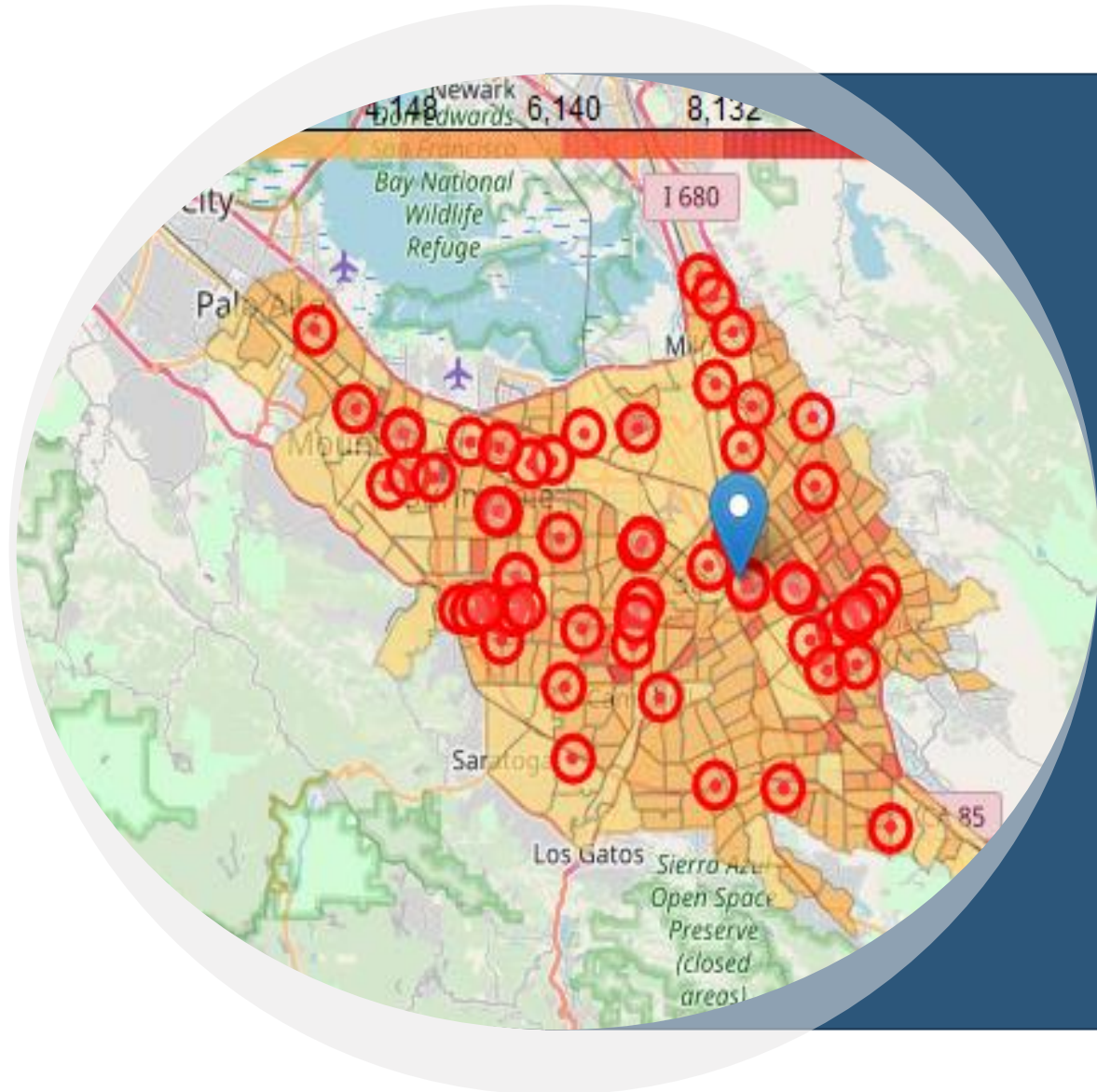
INTERESTS

The stakeholders who are looking to start Bubble Tea business or franchise expansion would be interested in searching for such kind of candidate locations.



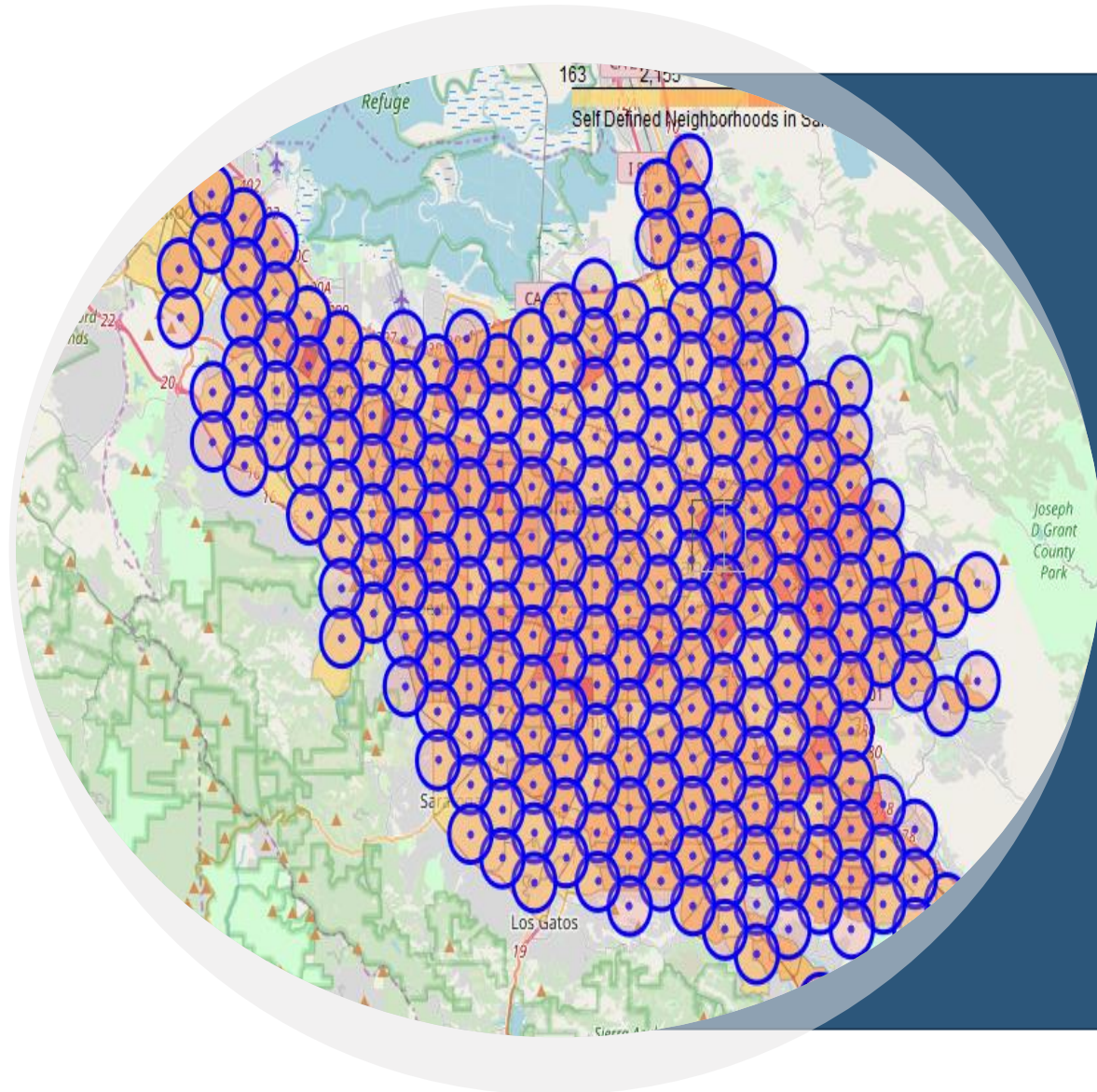
DATA

- Bubble Tea Shop List in Santa Clara County
- Bubble Tea Shop Nearby Venue Attributes Data



DATA CON'T

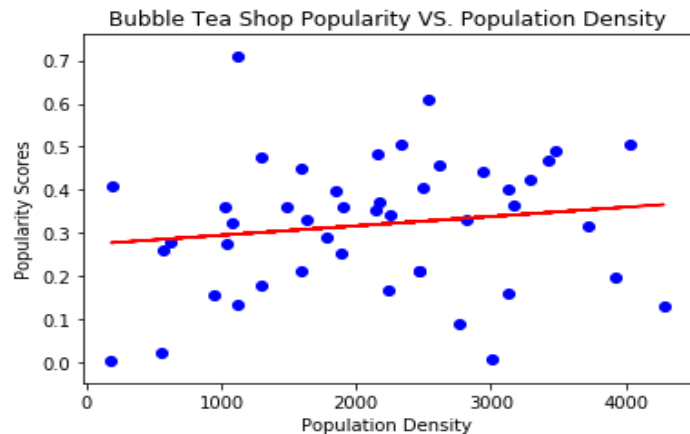
- County Neighborhoods Define and Data Generation
- Neighborhoods Nearby Venue Attributes Data



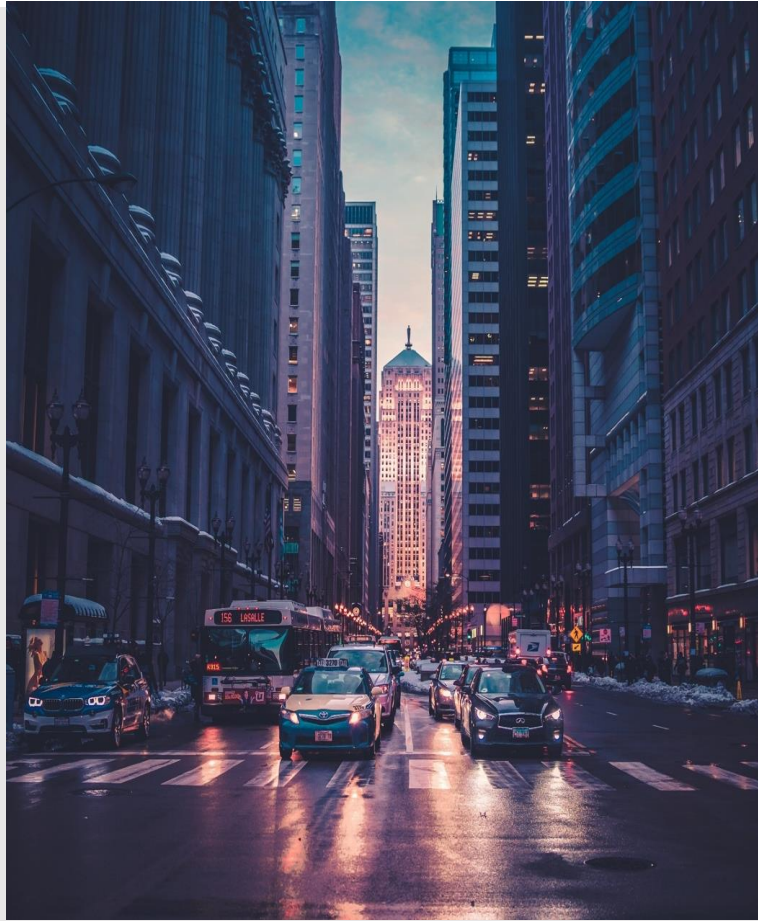
Exploratory Data Analysis



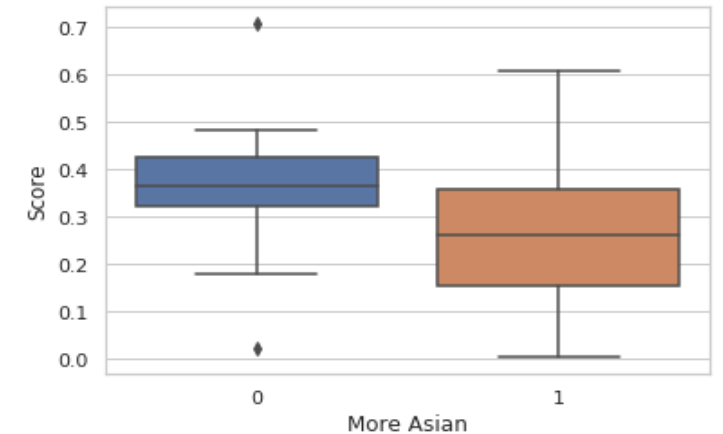
BUBBLE TEA SHOP'S SCORE VS POPULATION DENSITY



(F-score: 0.2476, P-value: 0.6212)



ASIAN PERCENTAGE VS BUBBLE TEA SHOP SCORE



(F-score: 4.3107, P-value: 0.0437)

0: below 42% , 1: above 42%

Prediction Modeling

CLASSIFICATION MODELS TEST SCORES



Test Score	Logistic Regression	K-Nearest Neighbor	SVM	Decision Trees
Jaccard Score	0.75	0.6875	0.7500	0.7500
F1 Score	0.75	0.5602	0.6859	0.7292

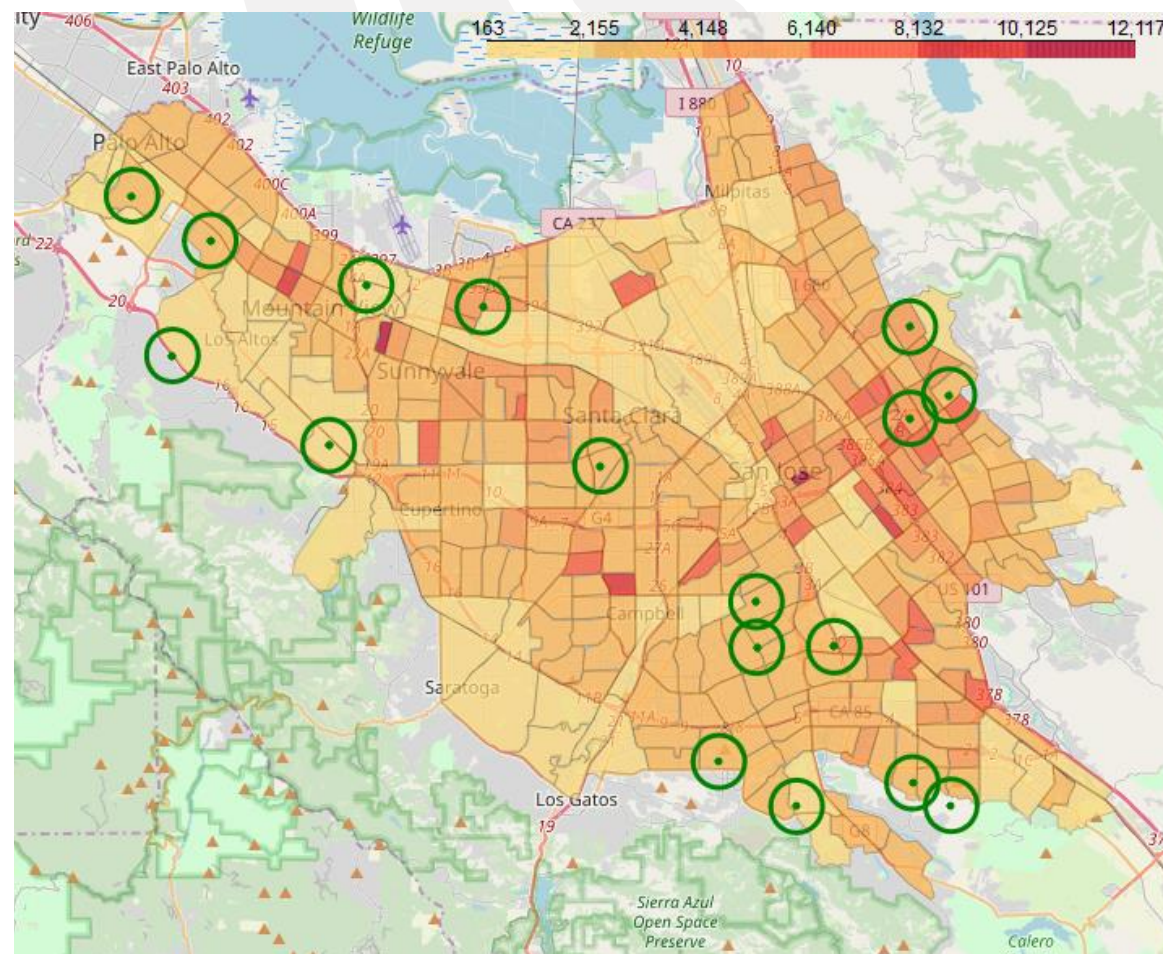
Logistic Regression model turn out the best test results (~75% accuracy and F1 score). What's more, the logistic regression model would provide probability information for every target location.



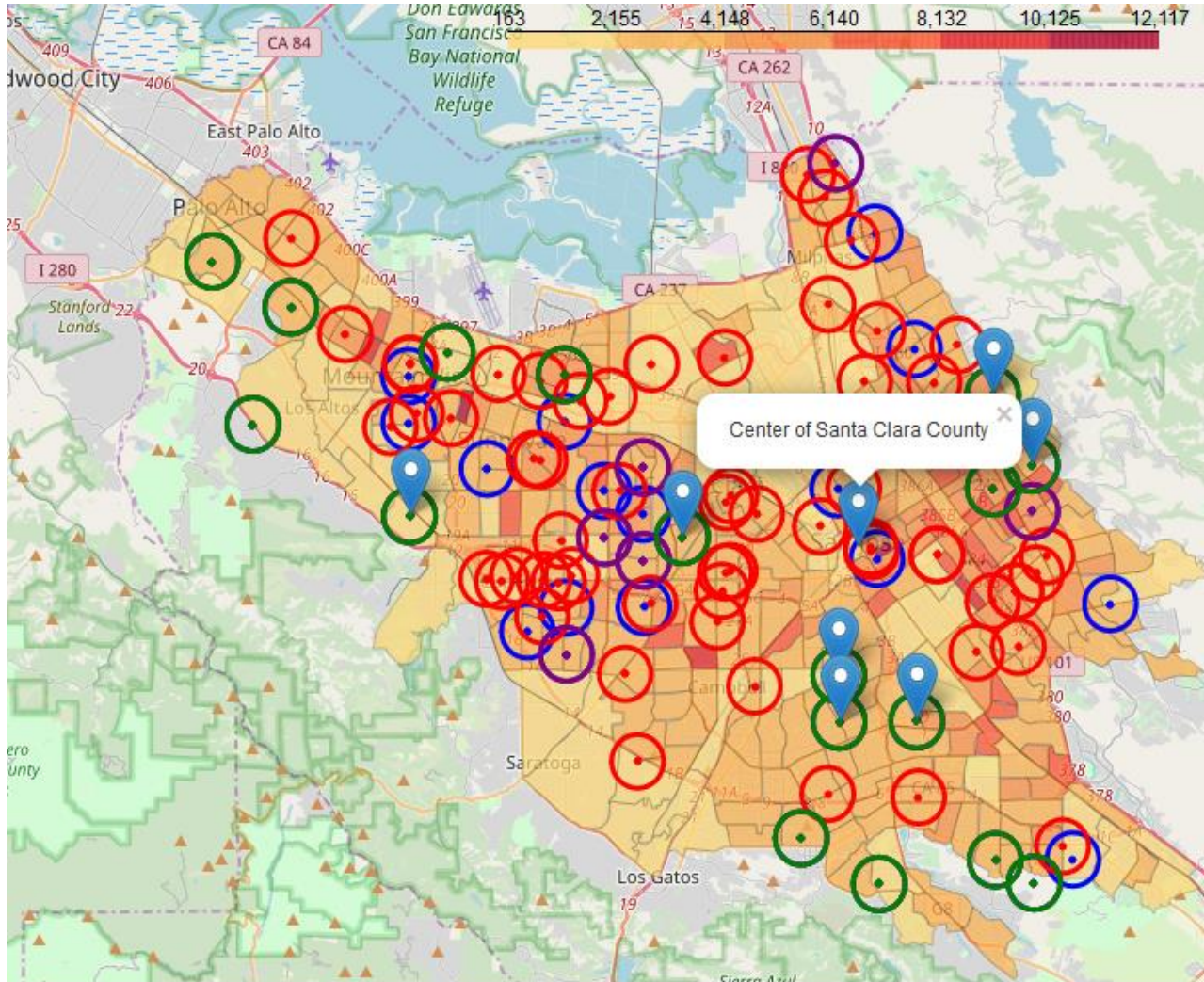
LOGISTIC REGRESSION MODEL PICKED

RESULT

	Neighborhood ID	Address	Probilities
0	303	3354, Gawain Drive, Alum Rock, San Jose, Santa...	0.510759
1	628	Saint Joseph Avenue, Creston, Los Altos, Santa...	0.509505
2	613	620, Cree Drive, Almaden Valley, San Jose, San...	0.506897
3	612	Curie Court, Almaden Valley, San Jose, Santa C...	0.505697
4	632	Junipero Serra Freeway, Los Altos Hills, Santa...	0.503966
5	528	606, Los Olivos Drive, Santa Clara, Santa Clar...	0.503062
6	365	2487, Bambi Lane, Alum Rock, San Jose, Santa C...	0.502733
7	503	146, Kittoe Drive, Whisman Station, Mountain V...	0.502403
8	540	Schwab Residential Center, 680, Serra Street, ...	0.502084
9	553	Communications Hill, Willow Glen, San Jose, Sa...	0.502000
10	586	2600, Booksin Avenue, Dry Creek, San Jose, San...	0.501976
11	675	6046, Paso Los Cerritos, Almaden Meadows, Alma...	0.501779
12	333	National Hispanic University, 14271, Story Roa...	0.501037
13	555	Willow Glen High School, Dry Creek Road, Dry C...	0.500983
14	677	Noddin Elementary School, Gilda Way, San Jose, ...	0.500915
15	538	4064, Solana Drive, Barron Park, Palo Alto, Sa...	0.500700
16	469	645, Bernal Avenue, Sunnyvale, Santa Clara Cou...	0.500285



DISCUSSION



If we take avoiding to overlap the 1 kilometer range with current existing Bubble Tea Shops and as closer to a high density area as possible into considerations, there are Seven better potential candidates list as following addresses:

1. 606, Los Olivos Drive, Santa Clara, Santa Clara County, California, 95050, USA
2. Willow Glen High School, Dry Creek Road, Dry Creek, San Jose, Santa Clara County, California, 95125, USA
3. 2600, Booksin Avenue, Dry Creek, San Jose, Santa Clara County, California, 95125, USA
4. Communications Hill, Willow Glen, San Jose, Santa Clara County, California, USA
5. 3354, Gawain Drive, Alum Rock, San Jose, Santa Clara County, California, 95127, USA
6. National Hispanic University, 14271, Story Road, Alum Rock, San Jose, Santa Clara County, California, 95127, USA
7. Saint Joseph Avenue, Creston, Los Altos, Santa Clara County, California, 94024-6833, USA

CONCLUSION

Final decision on optimal Bubble Tea Shop location will be made by stakeholders based on specific characteristics of neighborhoods and locations, preference of stakeholders, taking into consideration additional factors like amiabilities of plazas or shopping malls, accessibilities (major roads or freeways), rental prices, social and economic dynamics of the neighborhood etc.



THANK YOU



ZIJIAN ZHU