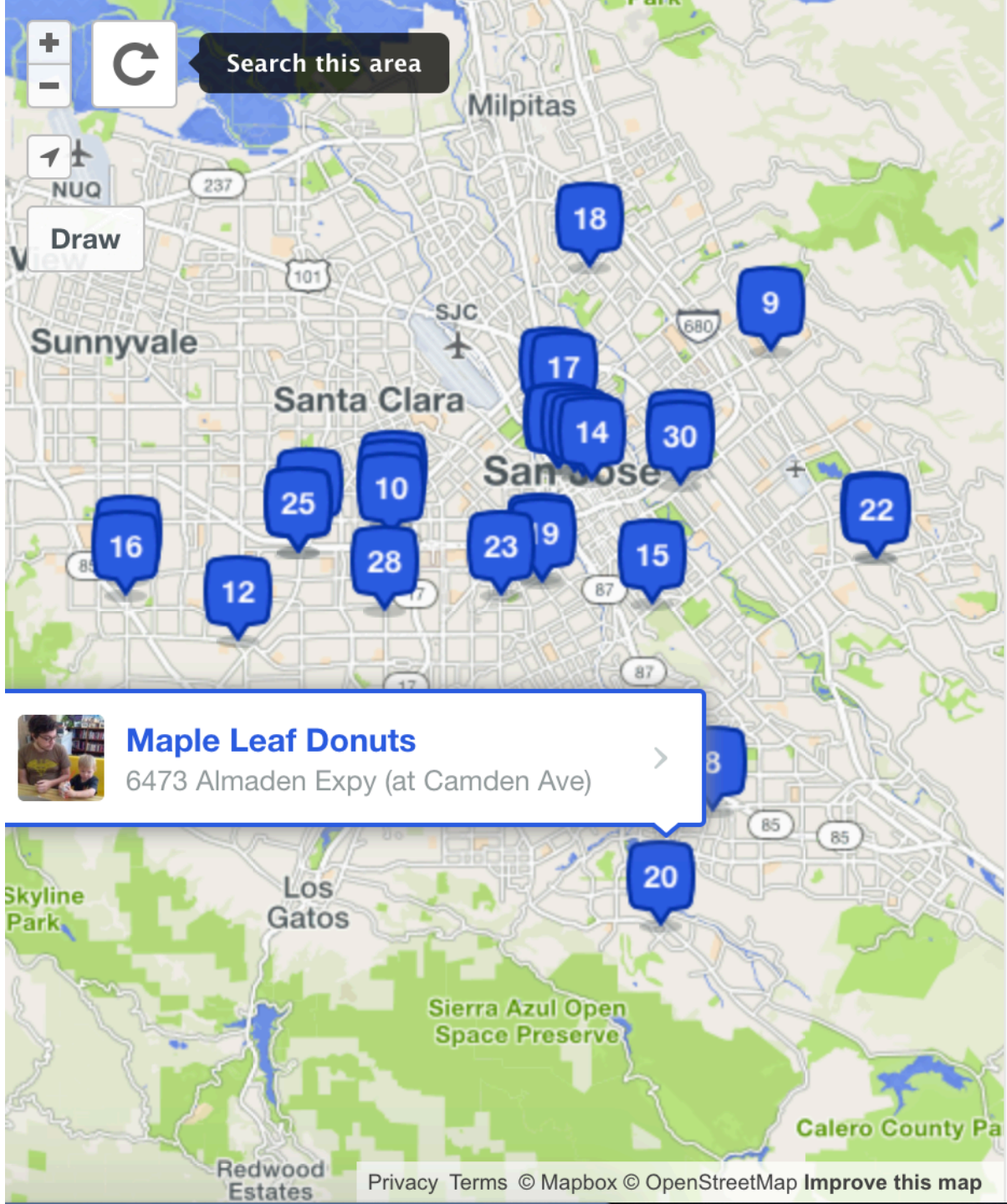
Neighborhood Analysis of Dessert Venues in SF South Bay Area

Borirak Opasanont

08Apr20

This presentation is a part of a submission for IBM Data Science Specialization capstone project on Coursera along with a Jupyter Notebook and a PDF Report, available on my Github: *https://github.com/boriopas/ds-capstone-project*



Screenshot of foursquare.com with dessert in San Jose as search term

Motivation

- Dessert scene could be an important factor to someone looking to select a neighborhood to move into
- Investors could use a high-level analysis to down-select neighborhood and dessert venue type to pursue further business plan
- This data can be used for comparing with other cities to better understand demographic preference of dessert venues

Photo of 85C Bakery by Vanessa S. on foursquare.com

Approach

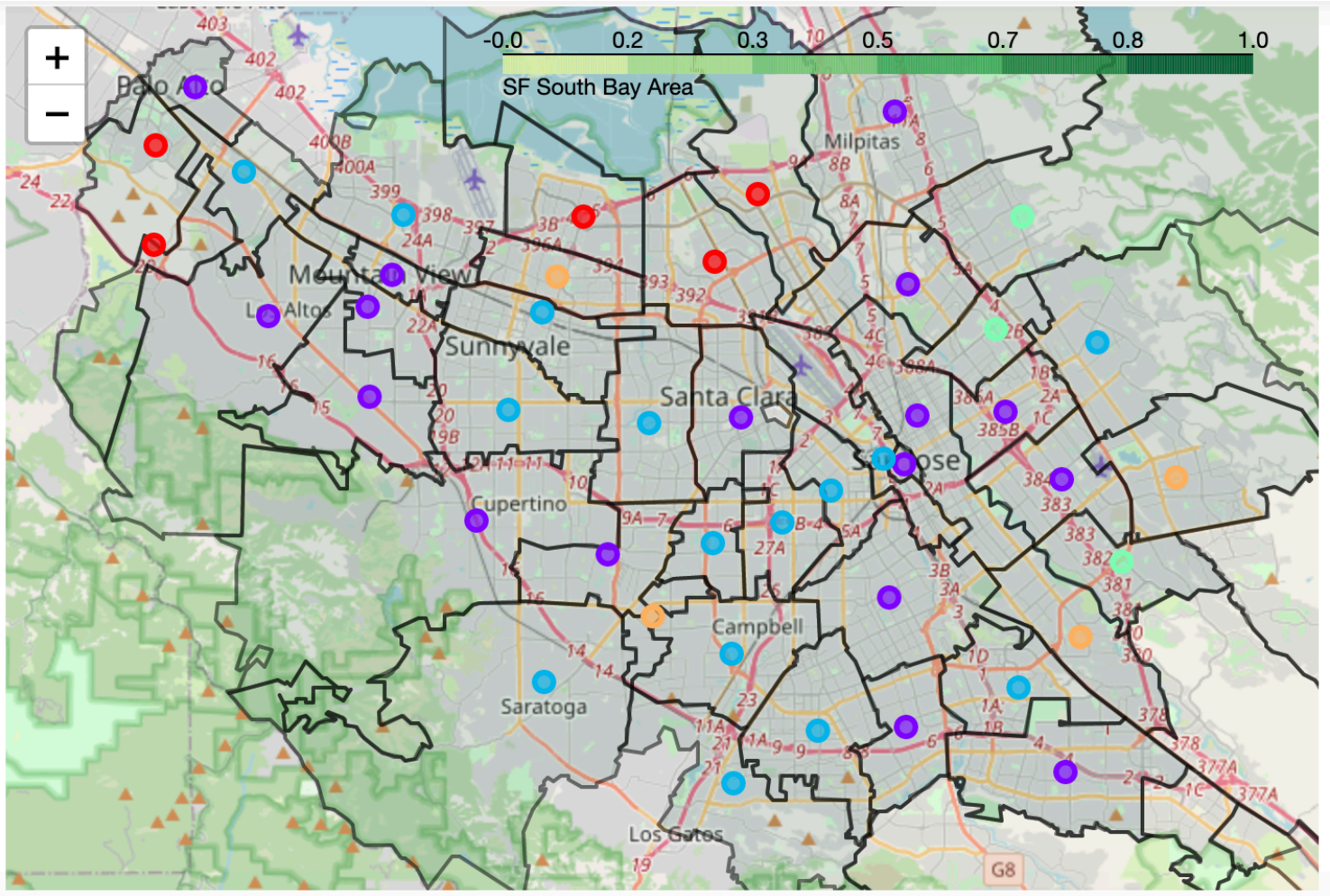
- Top 5 out of 63 dessert venue categories is used in defining clusters of neighborhood in South Bay using unsupervised k-means clustering
- Zip codes are used to represent 44 neighborhoods
- A machine learning model is developed using decision tree technique with demographic data such as age, ethnicity, language and household size to possibly explain the clustering of the neighborhoods

Photo of Campbell Creamery by Campbell Creamery on foursquare.com

Neighborhood Cluster Map

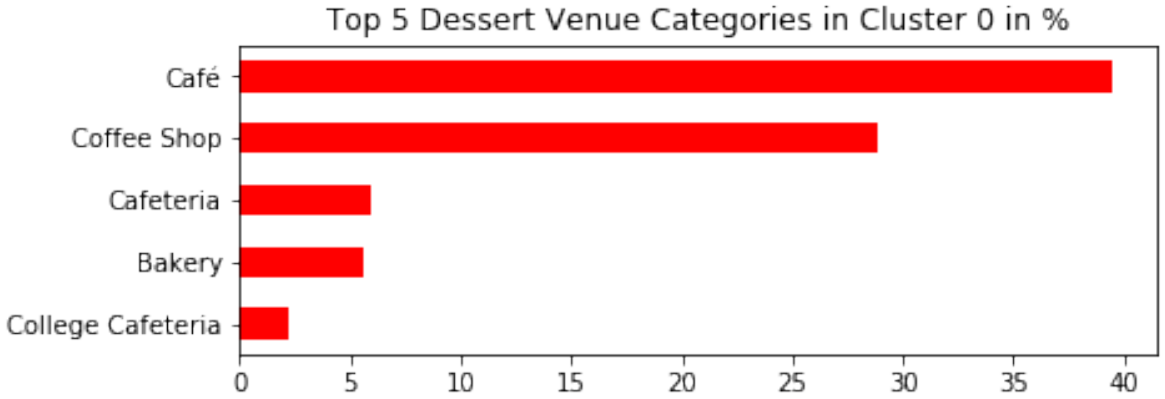
- 5 clusters obtained as a result of k-means clustering:

- Cluster 0 – Offices & College
- Cluster 1 – Mixed Use
- Cluster 2 – Offices & Residential
- Cluster 3 – Residential
- Cluster 4 – Bakery Towns

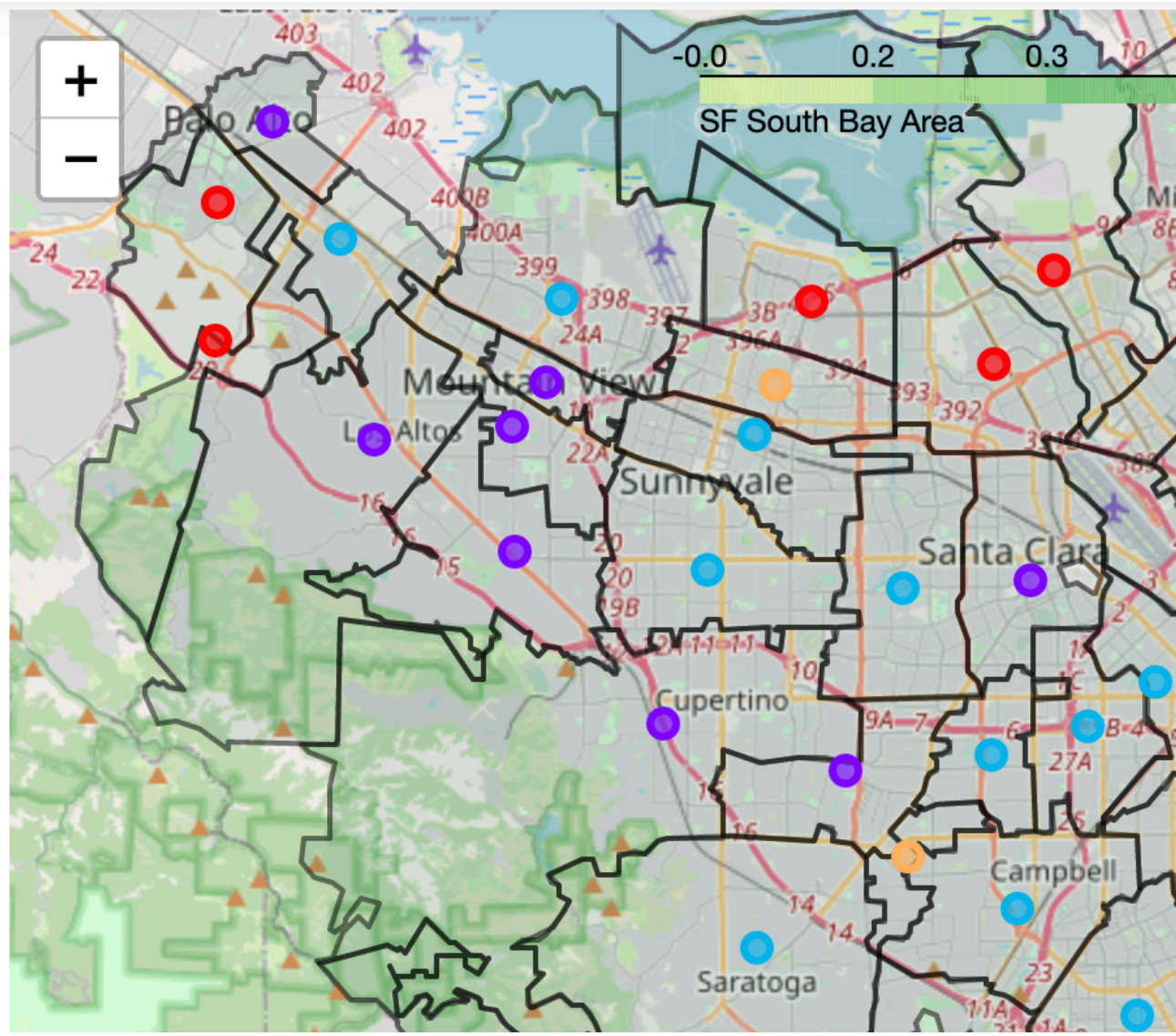


Venues category data using personal developer account from: https://developer.foursquare.com/
Demographic data and zip code boundary geojson from: https://data-sccphd.opendata.arcgis.com/datasets/demographic-statistics-zip-code
Zip code location data from: https://public.opendatasoft.com/explore/dataset/us-zip-code-latitude-and-longitude/export/?refine.state=CA

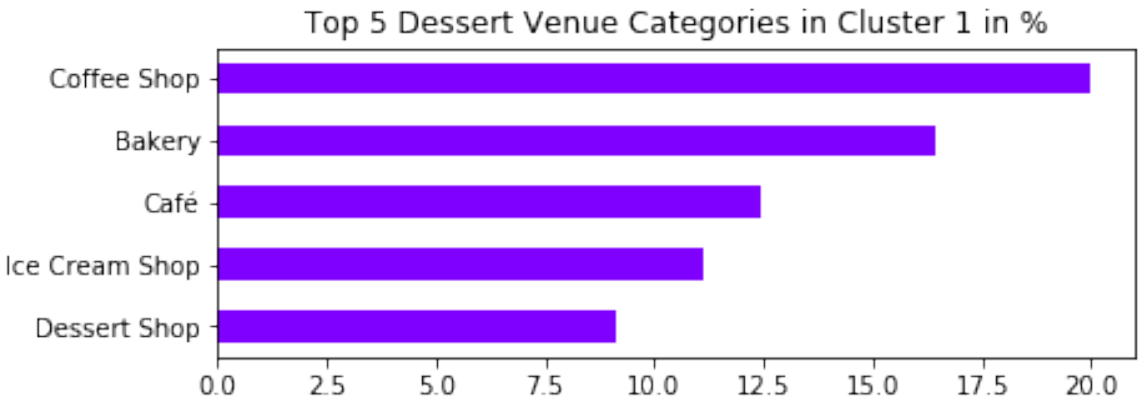
Cluster 0 – Offices & College



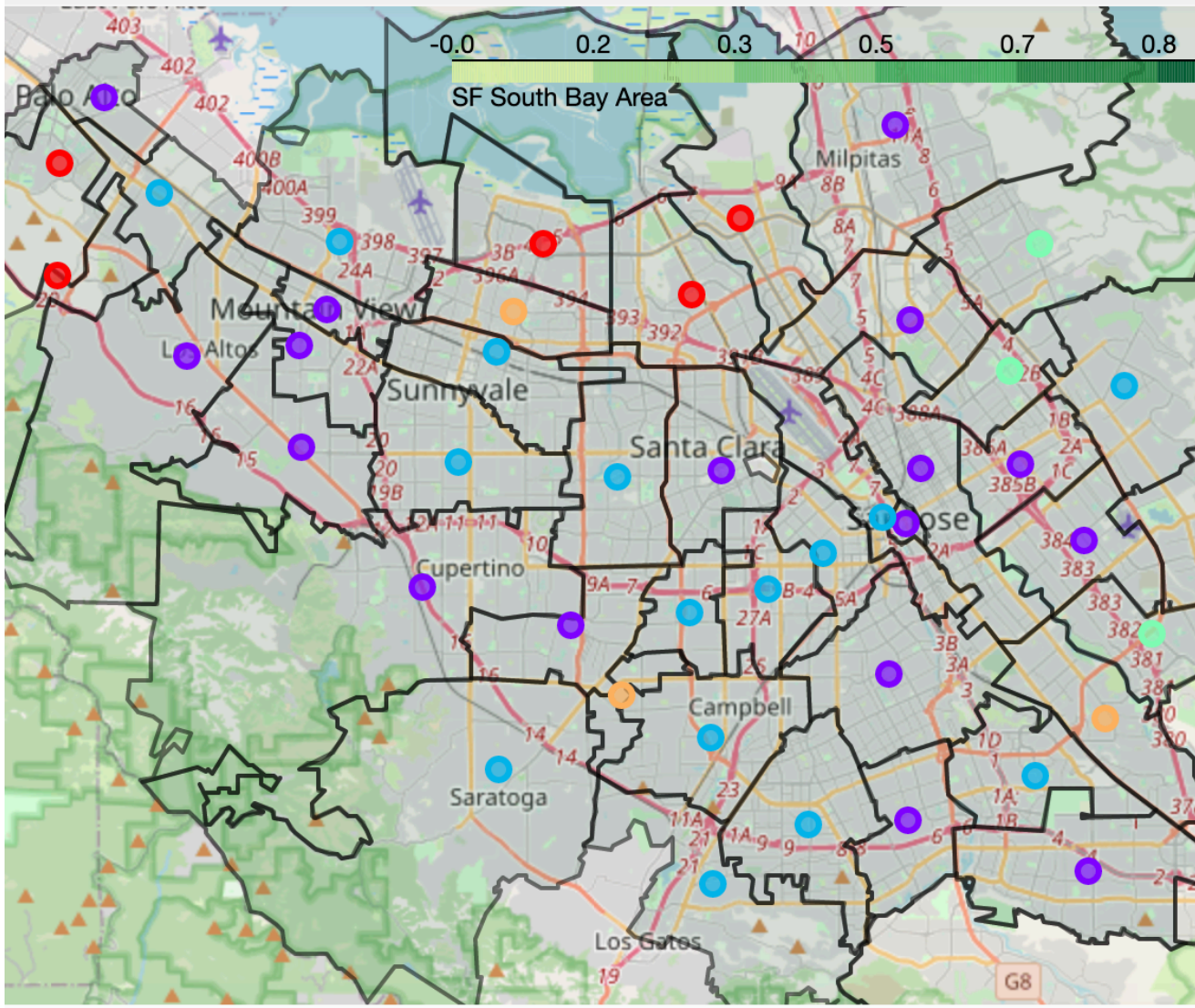
- Area around Stanford University and neighborhoods near the bay in North San Jose
- Large proportion of coffee shop and café, also cafeterias
- Busy office workers and students



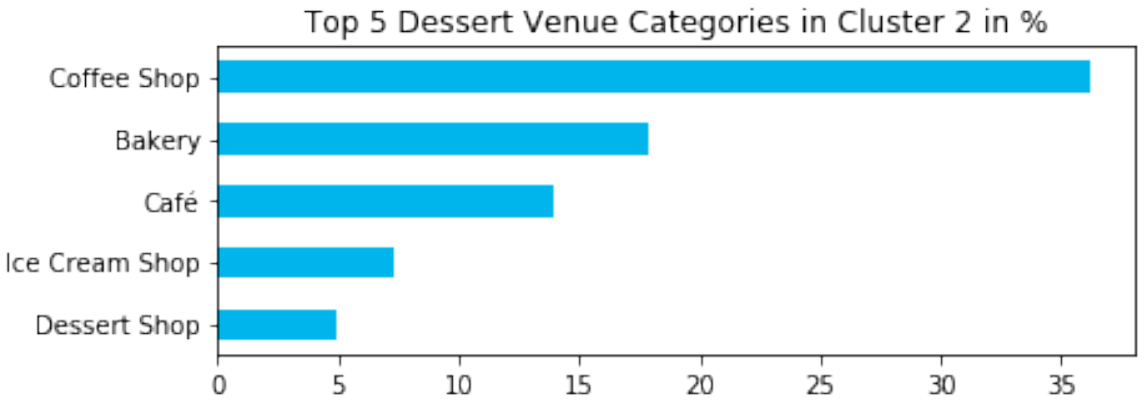
Clusters 1 – Mixed Use



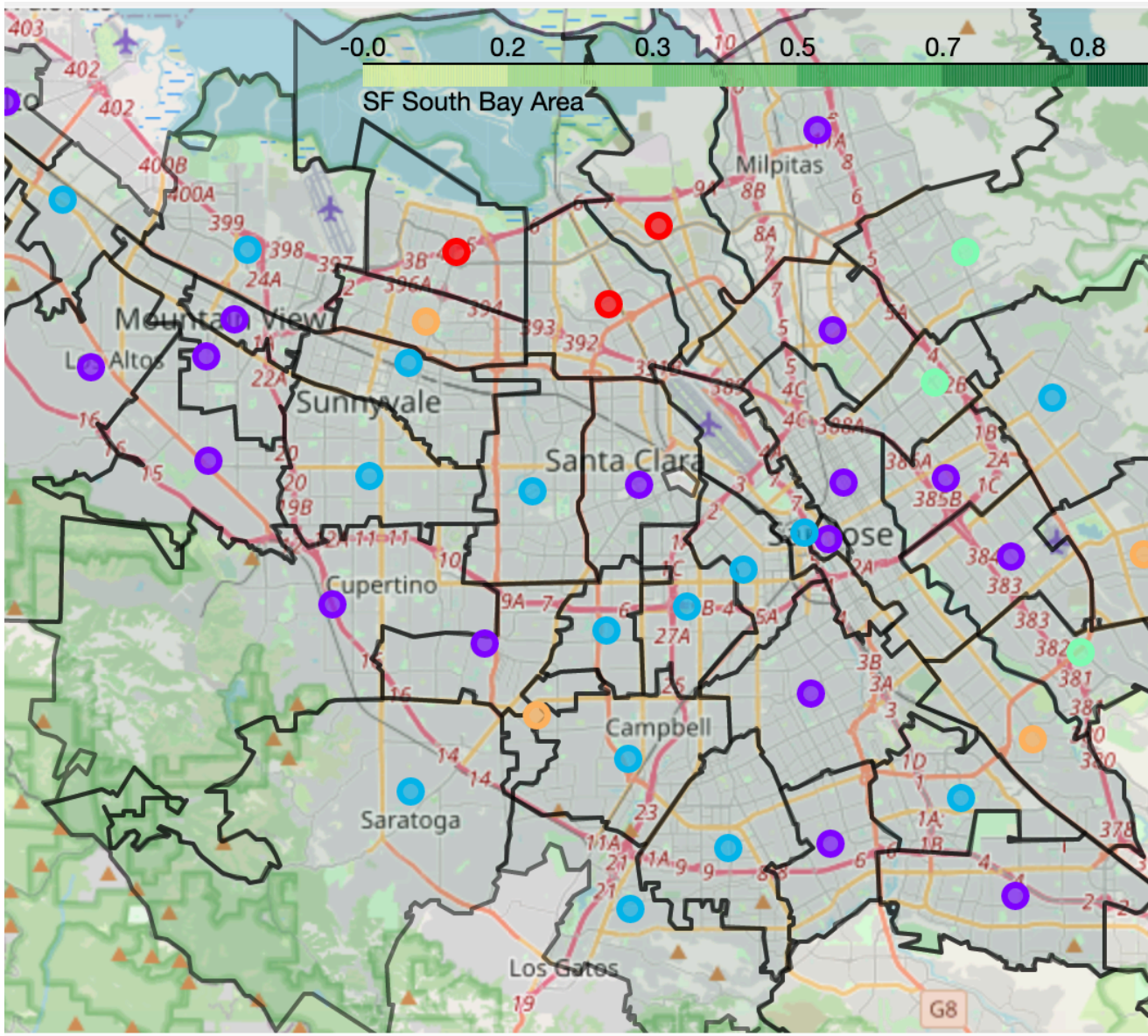
- Scattered within the South Bay Area
- Offices, industrial and residential
- Broad variety with coffee shop taking the top spot at only about 20%
- Best explained as a mix bag of dessert venues in mixed use areas



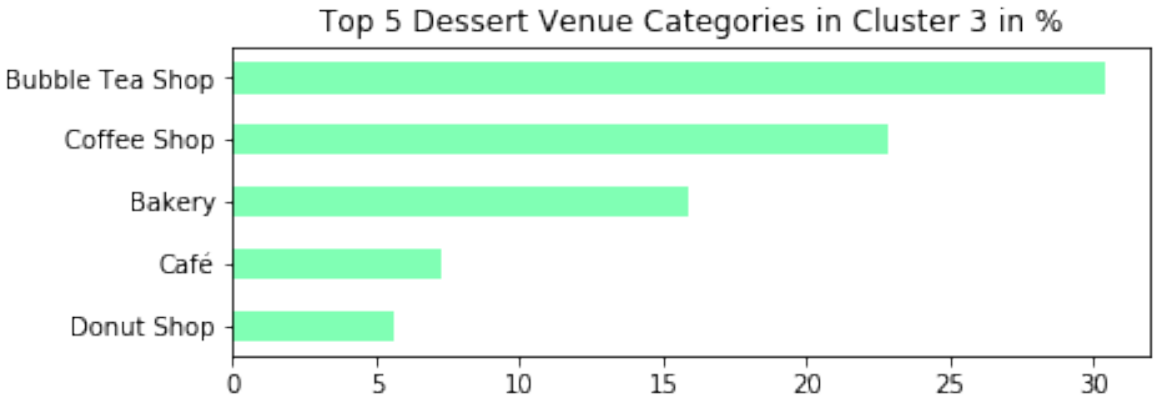
Cluster 2 – Offices & Residential



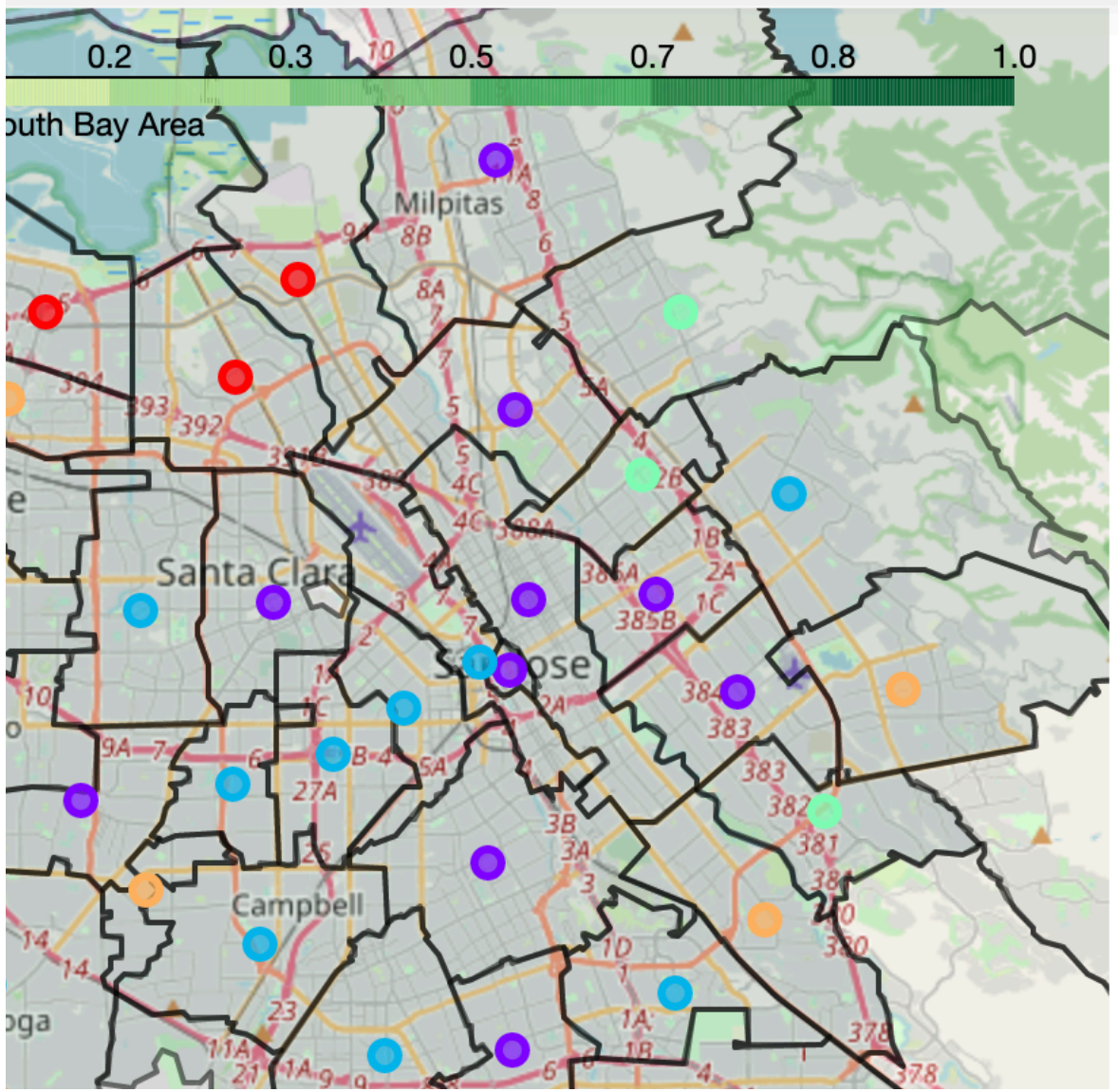
- Sunnyvale, Santa Clara and South San Jose area
- Largely residential and businesses
- Same top five as Cluster 1, but with stronger proportion of coffee shop



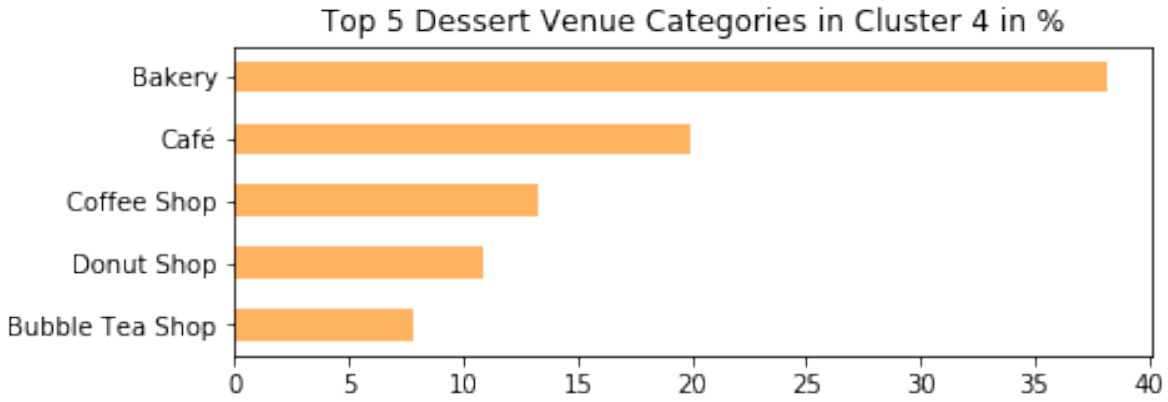
Cluster 3 - Residential



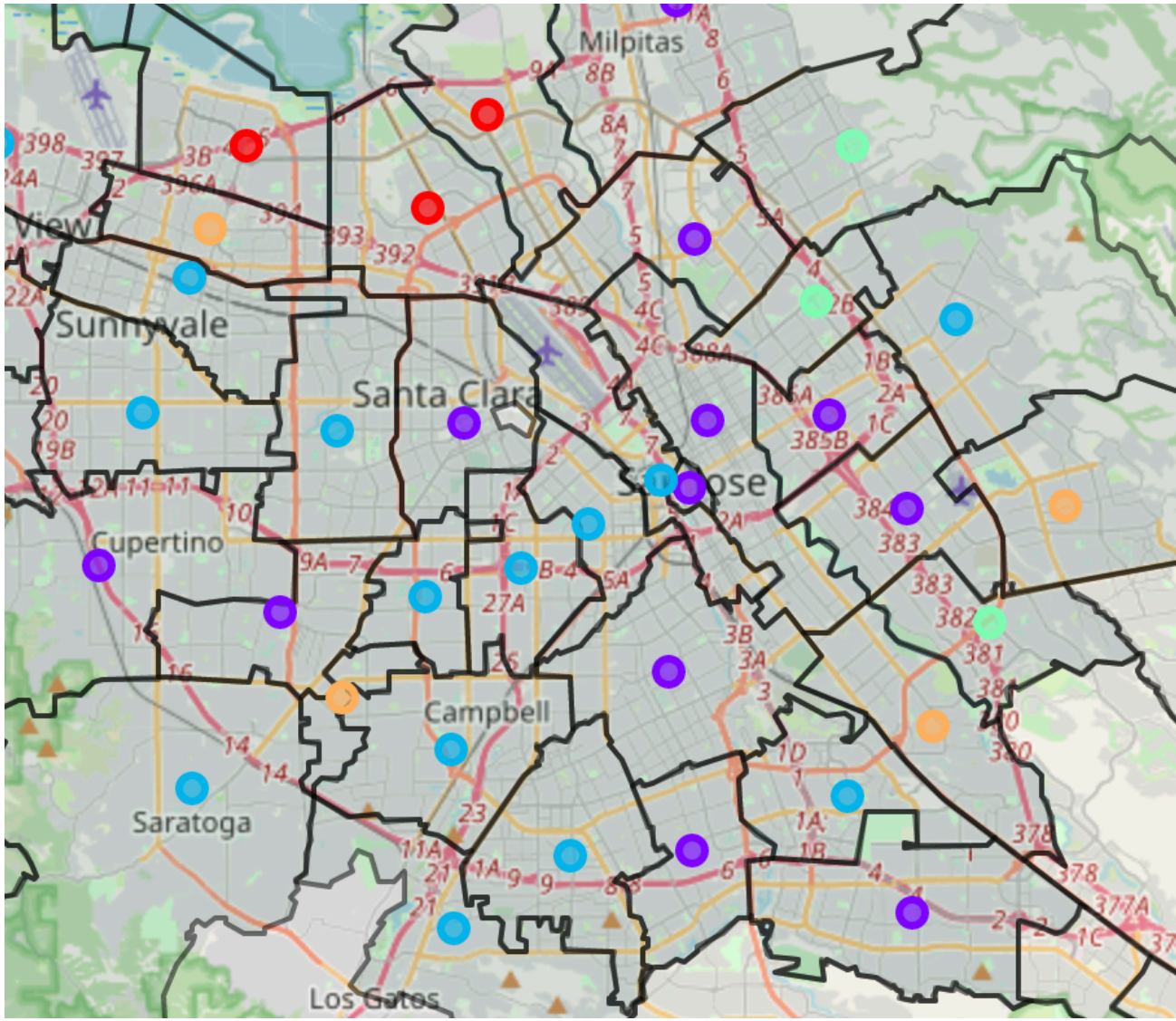
- 3 neighborhoods on East San Jose side
- Bubble tea shop takes top spot



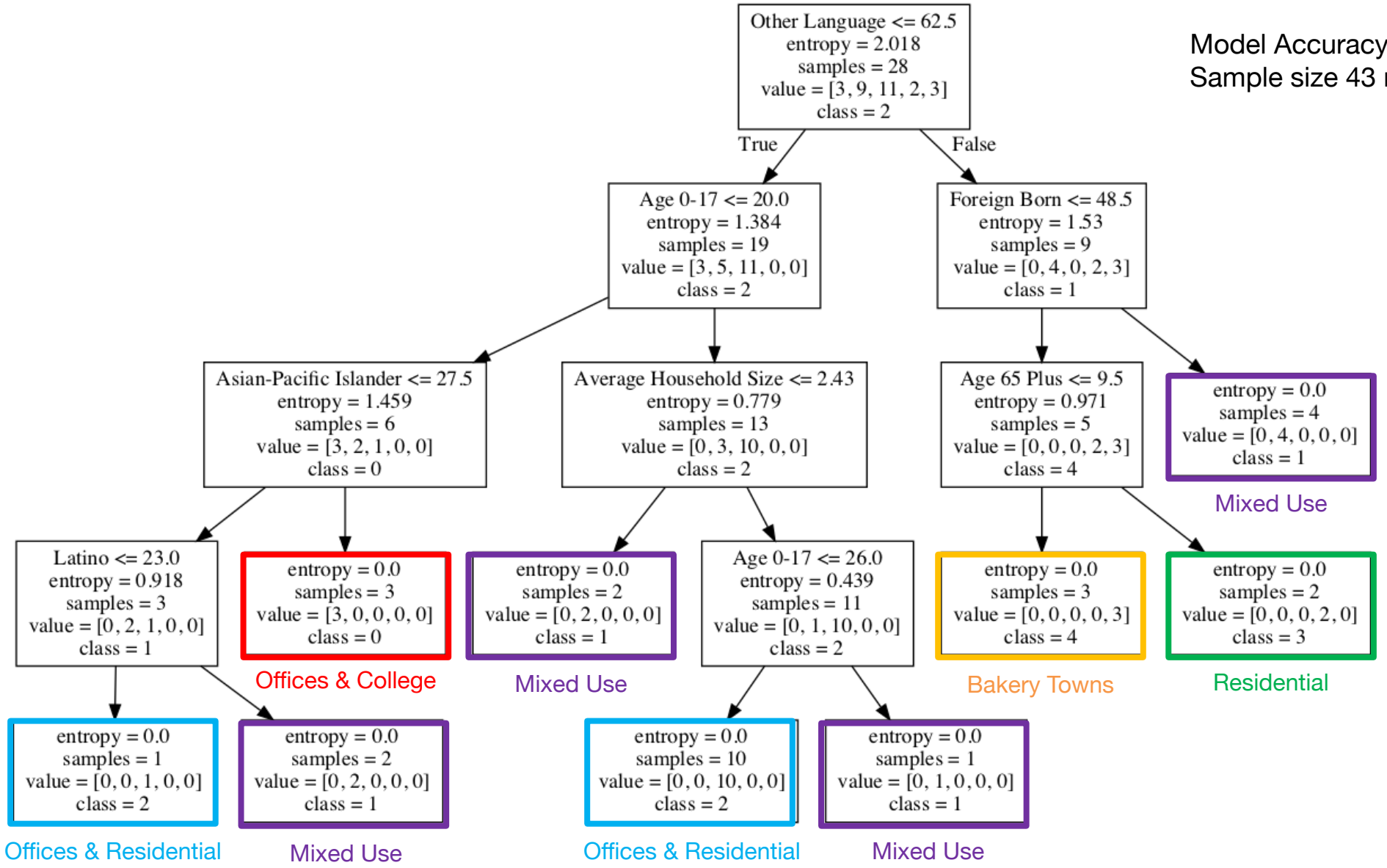
Cluster 4 – Bakery Towns



- 4 small neighborhoods, scattered spatially
- Bakery takes up 40% of dessert venues



Demographic Model of Neighborhood Clusters



Model Accuracy 46.7%
Sample size 43 neighborhoods

Summary

	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Label	Offices & College	Mixed Use	Offices & Residential	Residential	Bakery Towns
Count	5 neighborhoods	16 neighborhoods	15 neighborhoods	3 neighborhoods	4 neighborhoods
Locations	Stanford University and North San Jose area.	Offices, industrial and residential neighborhoods scattered within the Bay Area.	Residential and businesses in Sunnyvale, Santa Clara and South San Jose.	Residential neighborhoods on the East San Jose side.	Small neighborhoods scattered in the South Bay Area.
Characteristic Dessert Categories	Large proportion of coffee shop and cafe, suitable for busy office workers and students.	Broad variety of dessert venues. Best explained as a mix bag of venues and mixed use areas.	Stronger proportion of coffee shop than Cluster 1.	Bubble tea shop takes top spot.	Bakery takes up 40% of dessert venues.
Demographic Prediction By Model	Larger fraction of English as primary language, adults and of Asian/Pacific ethnicity.	Multiple demographic mix.	Larger fraction of English as primary language, smaller minor-age group, and: Either larger average household size, or smaller population of Asian/Pacific and Latino.	Larger fraction of non-English language, but smaller fraction of foreign-born population. Larger population of age > 65.	Like Cluster 3, but with smaller population of age > 65.

Conclusion

- Dessert venue categories can provide insights into SF South Bay Area neighborhood clusters
- Clusters are presented with their demographic model
- Information on five clusters can serve as a guide for someone who prioritizes dessert options in selecting neighborhood to settle in
- Model accuracy can possibly be further improved with bigger data sets like all of SF Bay Area, or California

Photo of Happy Lemon Premier by Jackie N. on foursquare.com