



IBM Data Science Capstone

Asian Fusion Restaurant in South Bay

Introduction

An Asian American businessman is looking to open an Asian fusion restaurant in South Bay region of northern California. He wishes to open the restaurant in an area where there are good percentage of Asian Americans and where there are less competition.



Data

Percent Asian population per zip code data will be extracted from zipatlas.com, this will also include longitude and latitude values of each zip code. Foursquare data will also be gathered to analyze the popular spots per zip code. The analysis will enable the businessman to target an area where there are decent percentage of Asian population and not as many competing businesses.

Percent Population per Zip Code Source:

<http://zipatlas.com/us/ca/zip-code-comparison/percentage-asian-population.htm>

California Geojson Source:

https://github.com/OpenDataDE/State-zip-code-GeoJSON/blob/master/ca_california_zip_codes_geo.min.json



Methodology

Once all the required data were obtained, k means was utilized to cluster the zip codes based on the popularity data extracted from Foursquare. K means was used because it determines if there are any sort of pattern or similarities between population in certain zip codes. The Foursquare data was then sorted to identify top 5 popular spots per cluster. Finally the % Asian population data and the Foursquare cluster data were overlayed on map of South Bay using Folium.

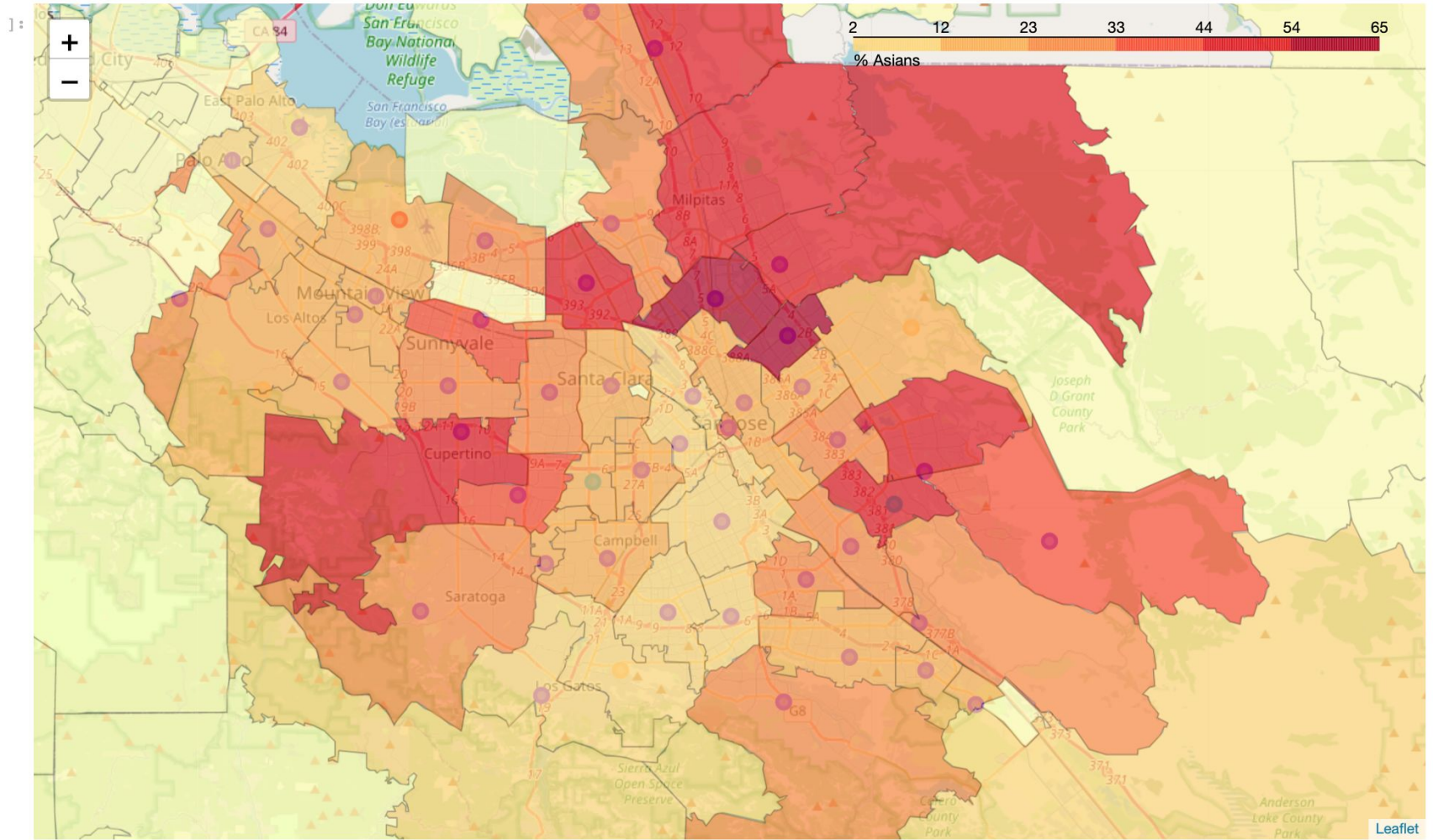


City		Zip Code	Latitude	Longitude	City	Population	% Asians	National Rank	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
San Jose	0	95131	37.387027	-121.897775	San Jose, California	26,384	64.10	#7	1	Shipping Store	Fast Food Restaurant	Mexican Restaurant	Food Court	Sandwich Place
	1	95133	37.371862	-121.860349	San Jose, California	26,018	55.01	#25	1	Park	Pool	High School	Trail	Bus Stop
	2	94555	37.572949	-122.051692	Fremont, California	33,863	53.65	#29	1	Zoo	Lounge	Park	Steakhouse	Bus Station
	17	95035	37.441931	-121.878502	Milpitas, California	62,840	51.69	#34	3	Track	Fish & Chips Shop	Furniture / Home Store	Frozen Yogurt Shop	Fried Chicken Joint
	3	95132	37.401061	-121.864386	San Jose, California	40,563	50.70	#39	1	Pet Service	Art Gallery	Park	Tennis Court	Fish & Chips Shop
	4	95148	37.315806	-121.788841	San Jose, California	44,095	50.48	#41	1	Pizza Place	Fast Food Restaurant	Indian Restaurant	Bank	Pharmacy
	5	94539	37.490497	-121.929693	Fremont, California	46,910	50.08	#43	1	Chinese Restaurant	Bubble Tea Shop	Pizza Place	Coffee Shop	Hotpot Restaurant
	18	95054	37.393554	-121.965114	Santa Clara, California	13,013	46.54	#53	1	Theme Park Ride / Attraction	Trail	Park	Food Truck	Liquor Store
	6	95121	37.302108	-121.805013	San Jose, California	37,126	45.58	#59	2	Intersection	Home Service	Food Truck	Track	Zoo
	19	95014	37.331686	-122.030656	Cupertino, California	54,412	44.39	#65	1	Japanese Restaurant	Electronics Store	Bagel Shop	Pool	Gym
San Jose	7	95129	37.305622	-122.000887	San Jose, California	37,565	41.93	#75	1	Ice Cream Shop	Playground	Furniture / Home Store	Dog Run	Donut Shop
	8	95135	37.286820	-121.723877	San Jose, California	15,791	35.45	#106	1	Diner	Zoo	Flea Market	Furniture / Home Store	Frozen Yogurt Shop
	9	94086	37.378541	-122.020456	Sunnyvale, California	63,997	34.38	#112	1	Indian Restaurant	Tree	Taco Place	Food Truck	Hardware Store
	10	95138	37.252697	-121.792015	San Jose, California	14,785	31.77	#128	1	Train Station	Auto Garage	Liquor Store	Convenience Store	Halal Restaurant

K-means with $k = 5$

Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	
0	0	Gym	Zoo	Flea Market	Furniture / Home Store	Frozen Yogurt Shop
1	1	Park	Mexican Restaurant	Trail	Convenience Store	Grocery Store
2	2	Intersection	Baseball Field	Home Service	Food Truck	Track
3	3	Track	Fish & Chips Shop	Furniture / Home Store	Frozen Yogurt Shop	Fried Chicken Joint
4	4	Mountain	Trail	Arts & Crafts Store	Home Service	Vegetarian / Vegan Restaurant





Conclusion

K means results indicates that most zip codes fall under cluster 1. Two of the highest % Asian zip codes, 95131 and 95133 also fall under Cluster 1. Cluster 1 however did not have any Asian restaurants in top 5 list, in fact the most popular cuisine was Mexican Restaurant. Clusters 3 had Fish and Chips as 2nd popular spot and Cluster 2 had Food Truck 4th most common venue. In terms of individual zip code data, 95131 with the highest percent Asians had Fast Food joint as the 2nd most common venue. The zip code 94539 had Chinese and hot pot restaurants as 1st and 5th most common venue.

Based on the results, it appears that there isn't a clear relationship between percent Asians and Asian restaurants being in top five most common venues. However it does not mean that there aren't Asian restaurants, it just means that those that exist are not as popular as other non-eatery venues but since they are not as popular, it could mean that they are less of competition. I would recommend to open the fusion restaurant in any of the top 10 highest % Asian zip codes, except for 95148, 94539 or 95014

