Fast Food Restaurant

Kavin Shah

July 2020

Introduction

Description of the Problem and Context:

A group of investors started a small Fast food restaurant chain in Alabama. One of the decisions made during last business strategy company meeting was to confirm a plan to establish and deploy new stores on the west coast via launching stores in Los Angeles. As a consequence, a first pilot project is launched, and Raymond, project leader was nominated with first objective to launch 10 new stores in Los Angeles, CA. The success of this initiative is crucial for the next phases of the project and for the development of additional stores. Among Raymond's key tasks, he needs to work with a real estate consultant to identify available venues in Los Angeles and close relevant deals as soon as possible to deploy the stores. After a first discussion with Raymond's preferred real estate broker in LA, he realized that the criteria he defined for finding the stores are not accurate enough and his broker shared additional information. Here are a few problems shared: A) He may spend too much time finding the right places in such a large city like LA, so he needs to identify some preferred zones within Los Angeles to focus his search. B) He may not be the best person to define priorities and most relevant areas for the Fast food restaurant chain criteria, so he needs more views with targeted areas. As a consequence, Raymond contacted me and asked for some help on where I would recommend they should open the first stores in LA. Raymond and I discussed and we came to a conclusion that the problem could be solved with defining a list of preferred areas in Los Angeles issued from classifying neighborhoods based on exploring existing venues and most frequent categories of venues in each candidate zone. This way, we can identify similar neighborhoods, gather them within several clusters and choose the right cluster of areas within Los Angeles to focus on. Such output will serve as a view for Raymond and a list of target zones for Raymond's real estate broker.

We will use the Foursquare location data to explore neighborhoods of LA, specifically categories of venues, in a similar way we did with Toronto area in the previous lab and assignment. We plan to use unsupervised machine learning method for classification, like k-means algorithm. This topic will be developed further in the next Methodology section of this report. So, we focus for now on unsupervised clustering method based on venues around candidate locations, and within each zip code in Los Angeles, CA. Additional tools for solving the problem: In addition to Foursquare API we already mentioned, we will use Jupyter notebook for all the coding and explanations of our method, process and computations. Coding will be done in python 3, and leveraging usual libraries: NumPy and SciPy for scientific computing, Pandas for data extracting, cleaning and analysis, Matplotlib and Folium for figures, plots, maps and visualization, Scikit-learn for machine learning, in particular we plan to use clustering kmeans algorithm.

In this report, we mainly present explanations, notes and comments as described above, but we don't insert actual code. Another note book is provided as part of the assignment and gathers all the python code, dataframes, plots and maps involved in the capstone project. The objective is to build clusters to partitions Los Angeles, CA in similar areas and identify the most suitable areas for launching new stores for Fast food restaurants. Once clusters are created, we will review the clusters and identify similarities within a given cluster and similarities between two different clusters.

Context

This is a list of over 10,000 fast food restaurants provided by Datafiniti's Business Database. The dataset includes the restaurant's address, city, latitude and longitude coordinates, name, and more.

Dataset acquisition and Cleaning.

Data was acquired from Datafiniti's Business Database. Apart from that I also took dataset for state codes and the population dataset of the states to get further insights in the relation of location and the restaurant.

From the perspective of Null values there weren't any row with null values. The main problem the dataset had was mention of same restaurant in different forms. For ex. Mc Donalds was also written had mcdonalds or McDonald's. So, I used NLTK's edit distance function to find the edit distance between those names Edit Distance: The distance between the source string and the target string is the minimum number of edit operations (deletions, insertions, or substitutions) required to transform the source into the target. The lower the distance, the more similar the two strings.

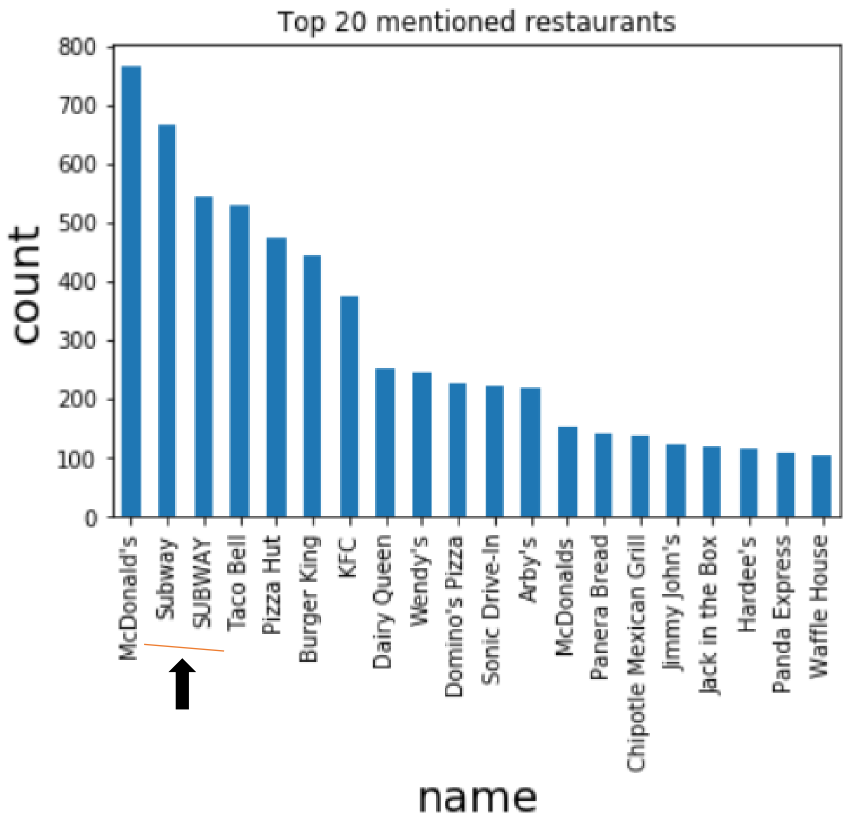
This helped correct one big analysis. Before McD was the largest food chain, after applying nltk Subway became the largest chain.

Feature Selection

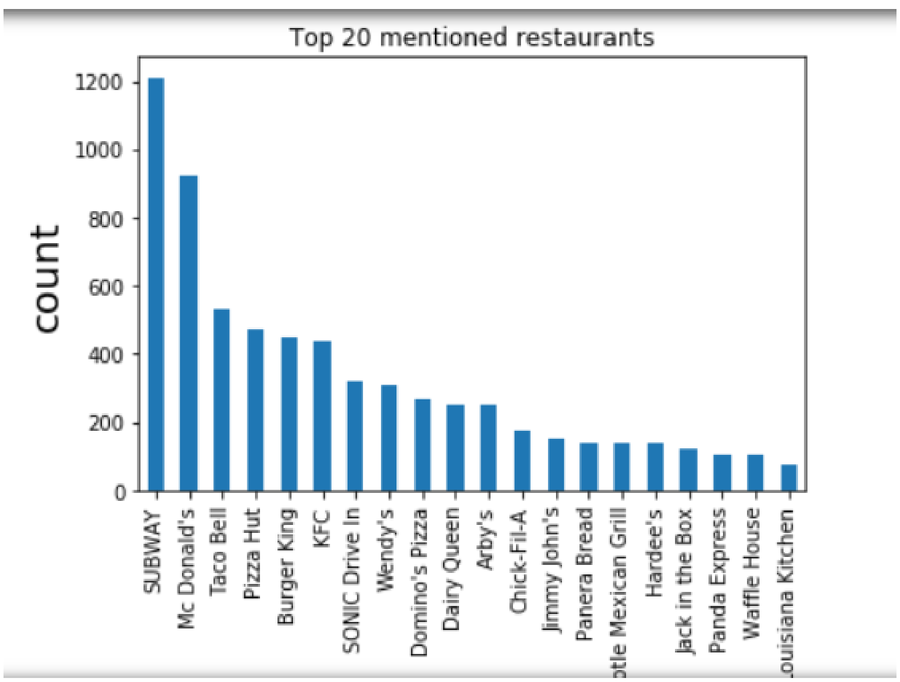
After cleaning the shape of the dataset was 10000 by 16. Upon further examining, it was clear that some of the features were not needed such as id, keys, sourceURLs, and websites which are unique identifier and would not be required for visualization or ML purposes.

EDA

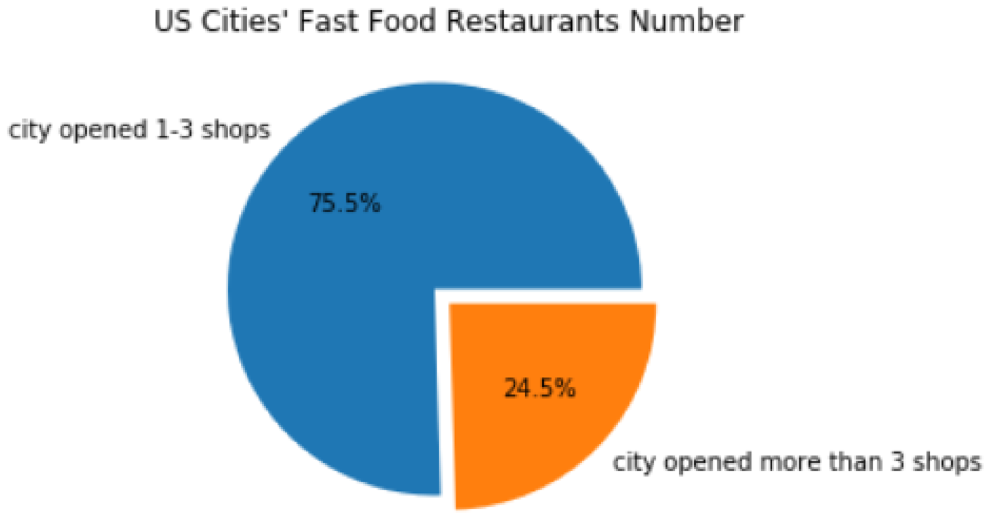
top 20 restaurants recorded by count total.



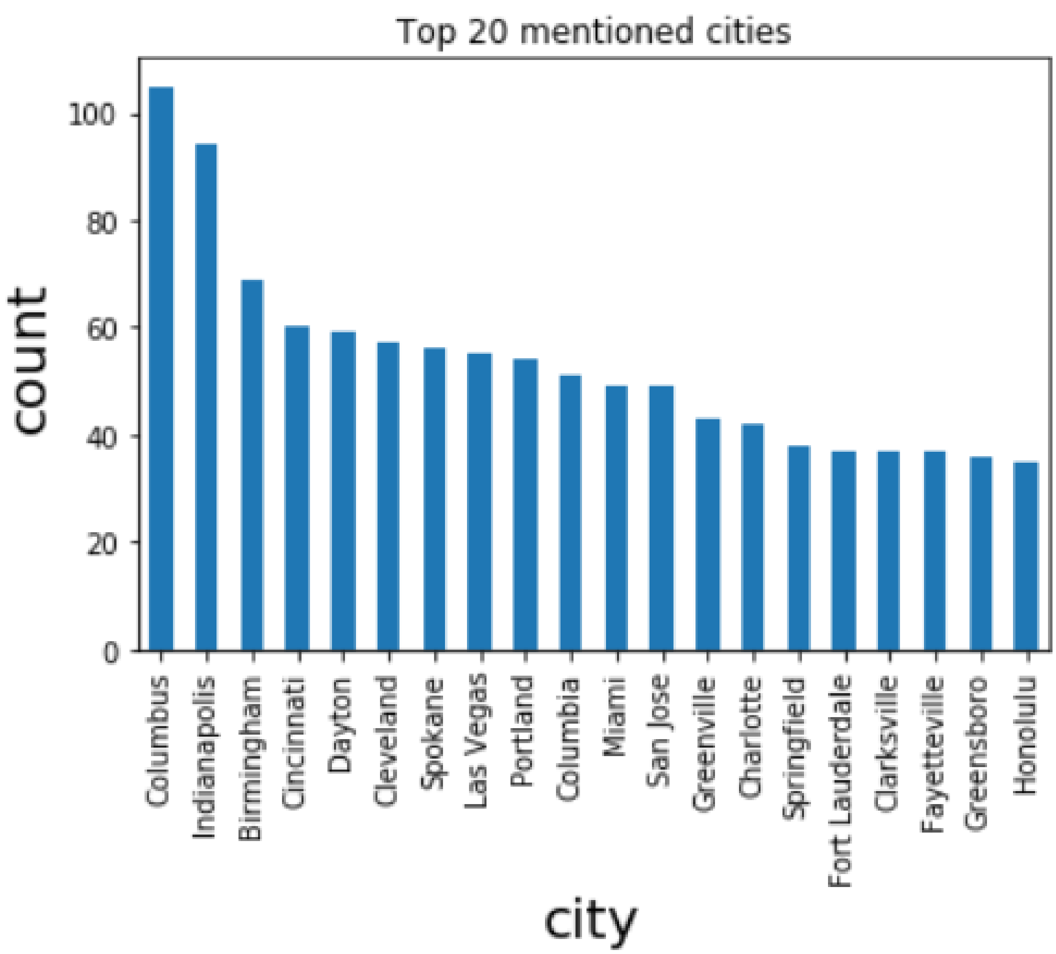
Corrected Top 20



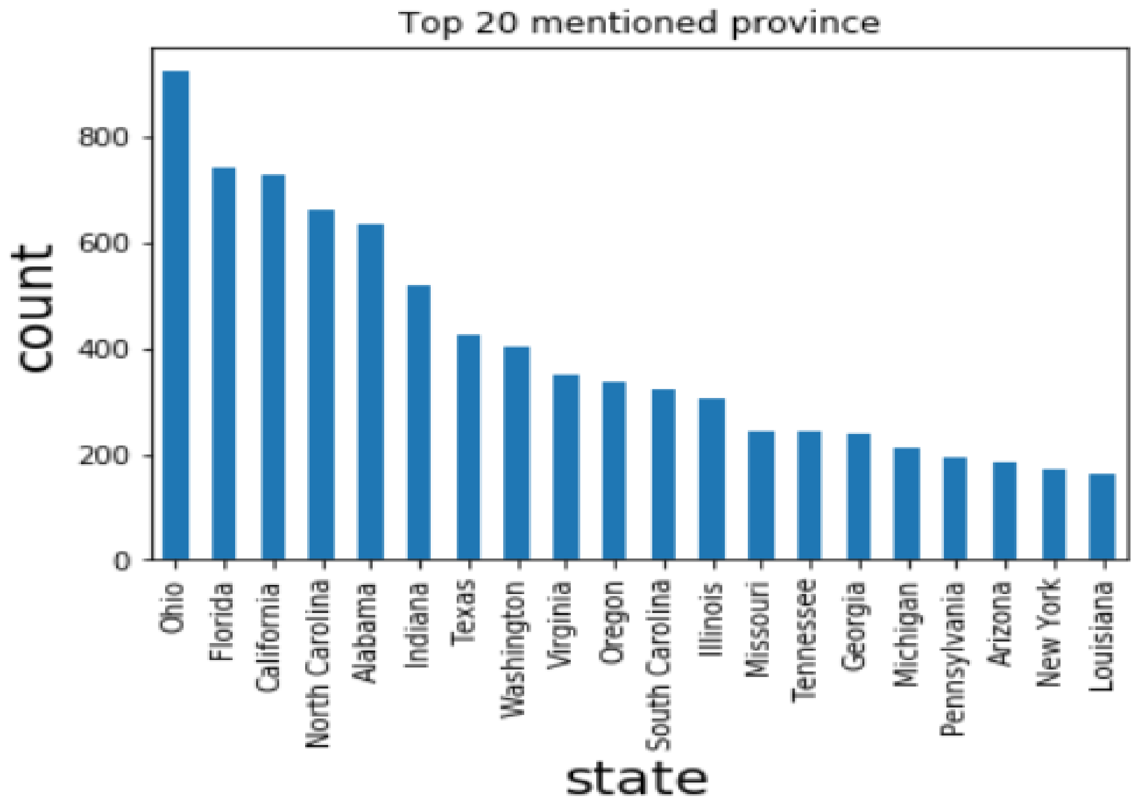
Here, we can see the % for restaurants where branches of restaurant is between 1-3 and more than 3.

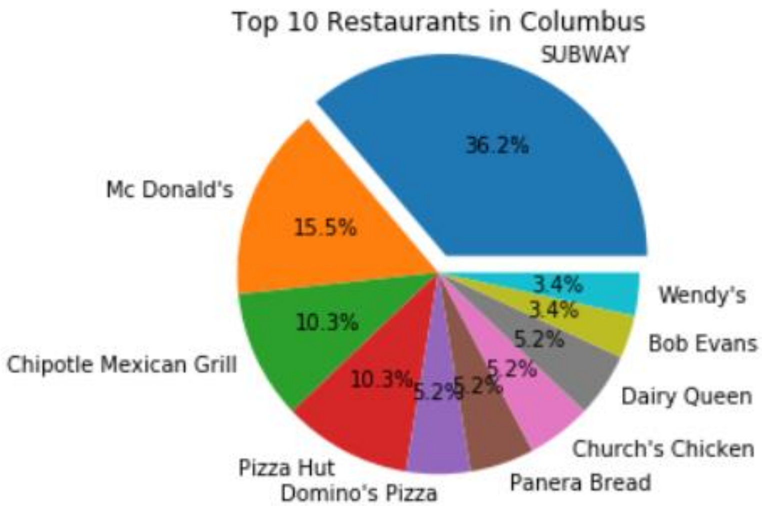


top 20 fast food populated cities recorded by count total. Cities with the most number of fast-food restaurants.

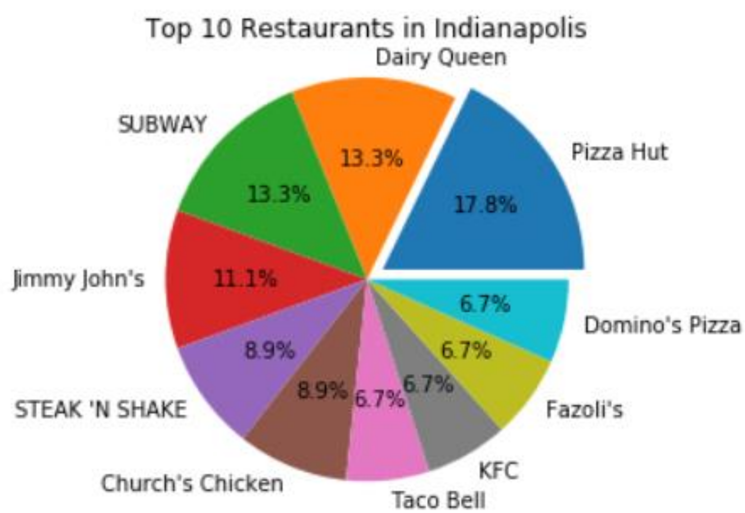


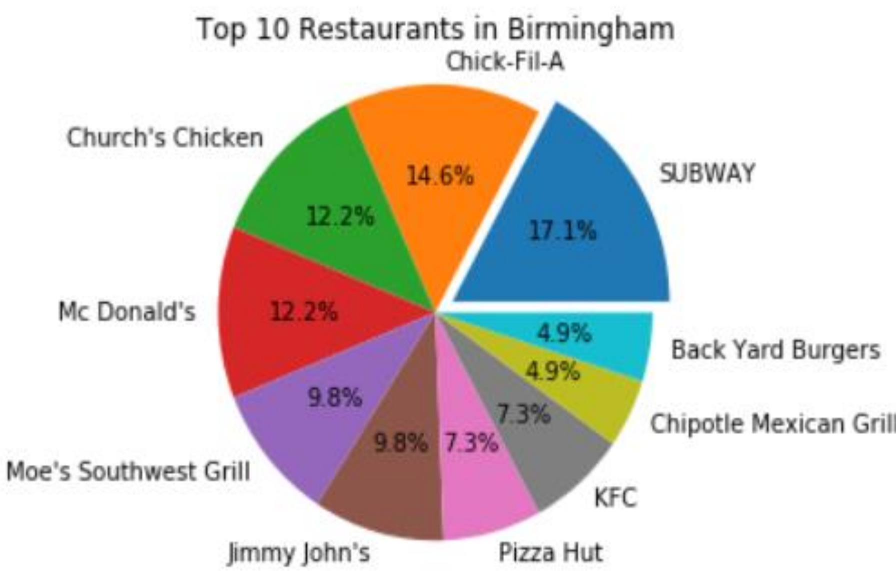
Top 20 fast food populated states recorded by count total.





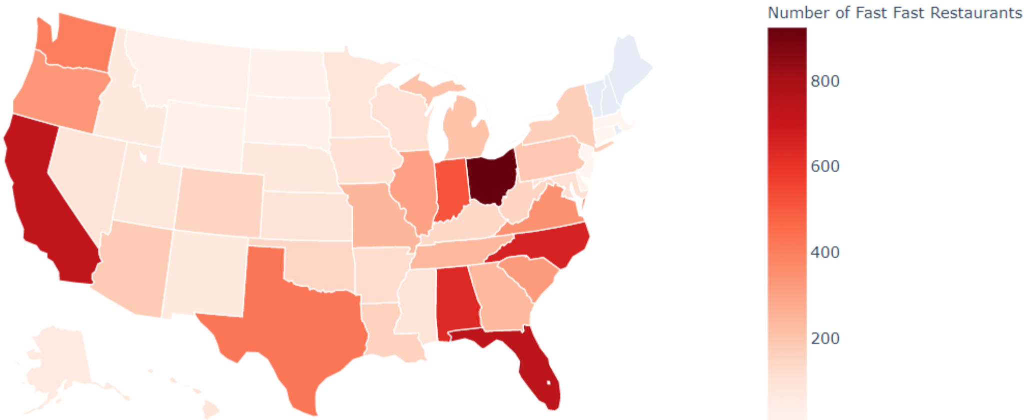
This is the breakdown of fast food restaurants for the 3 top cities with the highest number of fast food restaurants.





We can see that CA, OH, FL have darker red color indicating more restaurants.

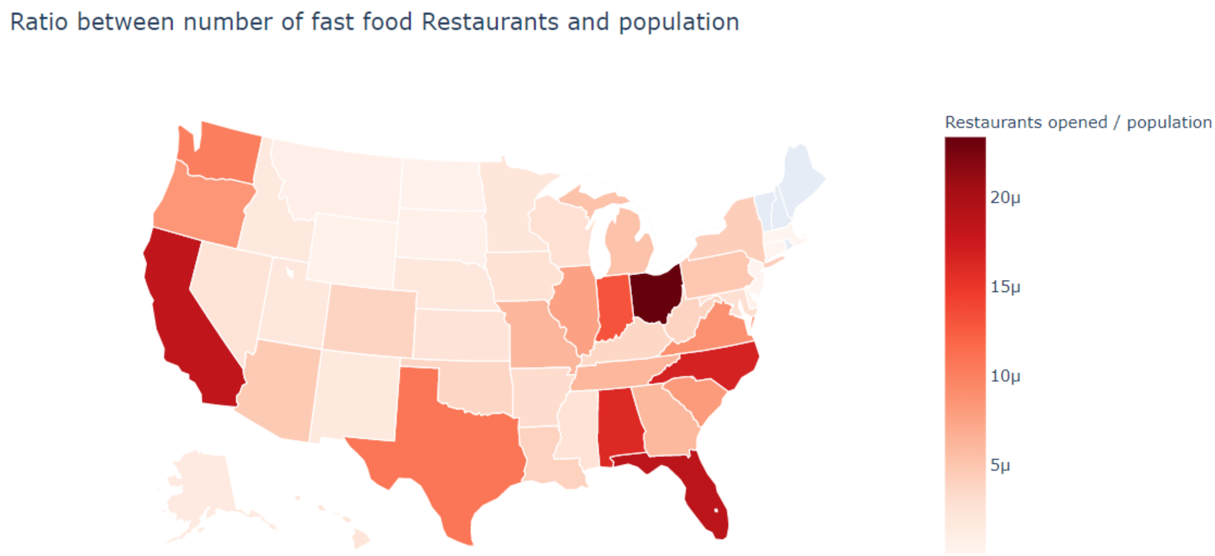
2019 US Fast Food Restaurants by State



Here, we see that Coastal parts of the country have more fast food restaurants than middle part.

2019 US Fast Food Restaurants by GPS location

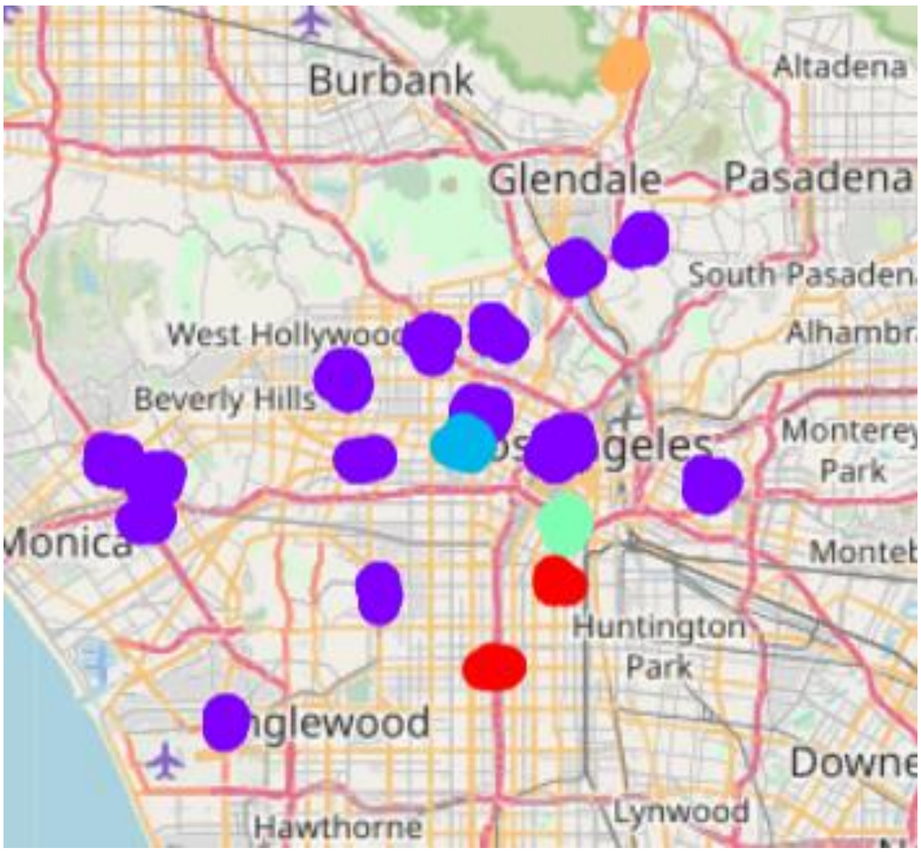
Keeping in mind the above map, we can see here that higher the population, more the restaurants, which makes sense.



Machine Learning.

Clustering.

Here, we see there are different clusters forming. We had taken the number of clusters as 5 (k=5). Most of the clusters belong to cluster 1(purple). Then next is cluster 0 (red). Cluster 2(sky blue), Cluster 3(neon green) and Cluster 4(Yellow) has only 1 Cluster each. To understand the clusters better we have also taken the top 3 most common venue returned by Foursquare to get a better idea of the neighborhood.



Results

This is a snippet of cluster 0(red). Here, we see that in this neighborhood fast food restaurants are the most visited venue along with Pizza place and Chinese Restaurant. This neighborhood is definitely a place that can be considered for opening the restaurant.

	Address	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	Cluster Labels
429	454 W Florence Ave	33.974417	-118.282092	Tom's Jr. Burgers Express	33.974439	-118.286893	Burger Joint	Fast Food Restaurant	Pizza Place	Chinese Restaurant	0
430	454 W Florence Ave	33.974417	-118.282092	SUBWAY	33.974385	-118.282363	Sandwich Place	Fast Food Restaurant	Pizza Place	Chinese Restaurant	0
431	454 W Florence Ave	33.974417	-118.282092	Little Caesars Pizza	33.974385	-118.282363	Pizza Place	Fast Food Restaurant	Pizza Place	Chinese Restaurant	0
432	454 W Florence Ave	33.974417	-118.282092	Pizza Hut	33.973828	-118.282904	Pizza Place	Fast Food Restaurant	Pizza Place	Chinese Restaurant	0
433	454 W Florence Ave	33.974417	-118.282092	Jack in the Box	33.974899	-118.278070	Fast Food Restaurant	Fast Food Restaurant	Pizza Place	Chinese Restaurant	0
434	454 W Florence Ave	33.974417	-118.282092	Shell	33.974941	-118.278692	Gas Station	Fast Food Restaurant	Pizza Place	Chinese Restaurant	0
435	454 W Florence Ave	33.974417	-118.282092	McDonald's	33.974900	-118.283005	Fast Food Restaurant	Fast Food Restaurant	Pizza Place	Chinese Restaurant	0
436	454 W Florence Ave	33.974417	-118.282092	Burger King	33.974266	-118.282955	Fast Food Restaurant	Fast Food Restaurant	Pizza Place	Chinese Restaurant	0
437	454 W Florence Ave	33.974417	-118.282092	Garage Door Springs Repair	33.976517	-118.282778	Auto Garage	Fast Food Restaurant	Pizza Place	Chinese Restaurant	0
438	454 W Florence Ave	33.974417	-118.282092	Chinatown Express	33.974396	-118.278793	Chinese Restaurant	Fast Food Restaurant	Pizza Place	Chinese Restaurant	0
	454 W										

This is a snippet of cluster 1(purple). There are many Clusters for Cluster 1. We have taken two places categorized as Cluster1. Here, we see that in one place fast food restaurants are the most visited venue and in other Coffee Shop is the most visited venue but knowing that most of the clusters are classified as Cluster 1, we can assume that there is high foot-fall in Cluster 1 places. Even this Cluster can be a suitable neighborhood to open a fast food restaurant.

	Neighborhood Latitude	Venue Longitude	Venue Category	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	Cluster Labels
0	34.088300	-118.309014	Market	Fast Food Restaurant	Grocery Store	Donut Shop	1
1	34.088300	-118.307939	Mexican Restaurant	Fast Food Restaurant	Grocery Store	Donut Shop	1
2	34.088300	-118.309499	Fast Food Restaurant	Fast Food Restaurant	Grocery Store	Donut Shop	1
3	34.088300	-118.309574	Wings Joint	Fast Food Restaurant	Grocery Store	Donut Shop	1
4	34.088300	-118.308968	Fried Chicken Joint	Fast Food Restaurant	Grocery Store	Donut Shop	1
5	34.088300	-118.308670	Korean Restaurant	Fast Food Restaurant	Grocery Store	Donut Shop	1
6	34.088300	-118.307995	Theater	Fast Food Restaurant	Grocery Store	Donut Shop	1
7	34.088300	-118.307822	Coffee Shop	Fast Food Restaurant	Grocery Store	Donut Shop	1
8	34.088300	-118.309262	Food Truck	Fast Food Restaurant	Grocery Store	Donut Shop	1
9	34.088300	-118.309316	Medical School	Fast Food Restaurant	Grocery Store	Donut Shop	1
27	34.063900	-118.288062	Korean Restaurant	Coffee Shop	Frozen Yogurt Shop	Bakery	1
28	34.063900	-118.289606	Coffee Shop	Coffee Shop	Frozen Yogurt Shop	Bakery	1
29	34.063900	-118.288899	Food Truck	Coffee Shop	Frozen Yogurt Shop	Bakery	1
30	34.063900	-118.290300	Bakery	Coffee Shop	Frozen Yogurt Shop	Bakery	1
31	34.063900	-118.291976	Food Truck	Coffee Shop	Frozen Yogurt Shop	Bakery	1
32	34.063900	-118.287369	Coffee Shop	Coffee Shop	Frozen Yogurt Shop	Bakery	1
33	34.063900	-118.289353	Dive Bar	Coffee Shop	Frozen Yogurt Shop	Bakery	1
34	34.063900	-118.290561	Video Game Store	Coffee Shop	Frozen Yogurt Shop	Bakery	1
35	34.063900	-118.290620	Breakfast Spot	Coffee Shop	Frozen Yogurt Shop	Bakery	1
36	34.063900	-118.292743	Japanese Restaurant	Coffee Shop	Frozen Yogurt Shop	Bakery	1
37	34.063900	-118.290841	Sandwich Place	Coffee Shop	Frozen Yogurt Shop	Bakery	1
38	34.063900	-118.287027	Asian Restaurant	Coffee Shop	Frozen Yogurt Shop	Bakery	1

This is the snippet of Cluster 2(sky blue). Here, we see that a very particular Restaurant is the most visited and the fact that the 3rd most visited place is the Hotel, we can say that the foot-fall might be less or there are few restaurants in the vicinity surrounding that hotel and area.

	Address	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	Cluster Labels
115	2801 West Olympic Boulevard	34.0529	-118.2952	Jeon Ju	34.052544	-118.292384	Korean Restaurant	Korean Restaurant	Coffee Shop	Hotel	2
116	2801 West Olympic Boulevard	34.0529	-118.2952	Beverly Soon Tofu	34.052982	-118.292499	Korean Restaurant	Korean Restaurant	Coffee Shop	Hotel	2
117	2801 West Olympic Boulevard	34.0529	-118.2952	Surawon Restaurant	34.052843	-118.295490	Korean Restaurant	Korean Restaurant	Coffee Shop	Hotel	2
118	2801 West Olympic Boulevard	34.0529	-118.2952	Park's BBQ	34.053920	-118.291902	Korean Restaurant	Korean Restaurant	Coffee Shop	Hotel	2
119	2801 West Olympic Boulevard	34.0529	-118.2952	World 8	34.051034	-118.291768	Video Game Store	Korean Restaurant	Coffee Shop	Hotel	2
120	2801 West Olympic Boulevard	34.0529	-118.2952	Aventura Hotel	34.052047	-118.297720	Hotel	Korean Restaurant	Coffee Shop	Hotel	2
121	2801 West Olympic Boulevard	34.0529	-118.2952	A-won Japanese Restaurant	34.055390	-118.292037	Japanese Restaurant	Korean Restaurant	Coffee Shop	Hotel	2
122	2801 West Olympic Boulevard	34.0529	-118.2952	Ong Ga Nae Korean B.B.Q	34.054748	-118.291439	Korean Restaurant	Korean Restaurant	Coffee Shop	Hotel	2
123	2801 West Olympic Boulevard	34.0529	-118.2952	I Love Boba (Vermont & Olympic)	34.052075	-118.291625	Coffee Shop	Korean Restaurant	Coffee Shop	Hotel	2
124	2801 West Olympic Boulevard	34.0529	-118.2952	Wako Donkasu	34.052539	-118.297724	Asian Restaurant	Korean Restaurant	Coffee Shop	Hotel	2

This is a snippet of Cluster 3(neon green). This Cluster doesn't seem to have the attraction for the fast food restaurants and wouldn't be a recommended neighborhood.

	Neighborhood Latitude	Venue Longitude	Venue Category	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	Cluster Labels
263	34.025188	-118.251844	Print Shop	Gas Station	Mexican Restaurant	Grocery Store	3
264	34.025188	-118.252492	Sandwich Place	Gas Station	Mexican Restaurant	Grocery Store	3
265	34.025188	-118.249983	Convenience Store	Gas Station	Mexican Restaurant	Grocery Store	3
266	34.025188	-118.250811	Clothing Store	Gas Station	Mexican Restaurant	Grocery Store	3
267	34.025188	-118.252166	Grocery Store	Gas Station	Mexican Restaurant	Grocery Store	3
268	34.025188	-118.249942	Gas Station	Gas Station	Mexican Restaurant	Grocery Store	3
269	34.025188	-118.250641	Coffee Shop	Gas Station	Mexican Restaurant	Grocery Store	3
270	34.025188	-118.248962	Gas Station	Gas Station	Mexican Restaurant	Grocery Store	3
271	34.025188	-118.249800	Mexican Restaurant	Gas Station	Mexican Restaurant	Grocery Store	3
272	34.025188	-118.251602	Chinese Restaurant	Gas Station	Mexican Restaurant	Grocery Store	3
273	34.025188	-118.255846	Light Rail Station	Gas Station	Mexican Restaurant	Grocery Store	3
274	34.025188	-118.251025	Mexican Restaurant	Gas Station	Mexican Restaurant	Grocery Store	3
275	34.025188	-118.252677	Liquor Store	Gas Station	Mexican Restaurant	Grocery Store	3

This is a snippet of Cluster 4(yellow). Like, Cluster 3 even this neighborhood does not seem to be a good place for opening a fast food restaurant.

	Neighborhood Latitude	Venue Longitude	Venue Category	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	Cluster Labels
311	34.188015	-118.229316	Thai Restaurant	Poke Place	Cosmetics Shop	Liquor Store	4
312	34.188015	-118.227064	Park	Poke Place	Cosmetics Shop	Liquor Store	4
313	34.188015	-118.225854	Cosmetics Shop	Poke Place	Cosmetics Shop	Liquor Store	4
314	34.188015	-118.225454	Middle Eastern Restaurant	Poke Place	Cosmetics Shop	Liquor Store	4
315	34.188015	-118.226967	Liquor Store	Poke Place	Cosmetics Shop	Liquor Store	4
316	34.188015	-118.229080	Poke Place	Poke Place	Cosmetics Shop	Liquor Store	4

Discussion

As observations noted from the map in the Results section, most of the fast food restaurants are in the Clusters 0 and 1 and moderate number in Cluster 2. On the other hand, Clusters 3 and 4 has very low number of fast food restaurants in the neighborhoods. This can be seen in two ways, firstly, as a great opportunity and high potential areas to open new fast food restaurants as there is very little to no competition. Meanwhile, on the flip side there is a chance of missing on high footfall as suggested in Clusters 0 and 1.

Limitations and Suggestions

For Future Research In this project, we only consider one factor i.e. frequency of occurrence of fast food restaurants, there are other factors such as population and income of neighborhoods that could influence the location decision of a new fast food restaurant. However, to the best knowledge of this researcher such data are not available to the neighborhood level required by this project. Future research could devise a methodology to estimate such data to be used in the clustering algorithm to determine the preferred locations to open a new fast food restaurant.

Conclusion

In this project, we have gone through the process of identifying the business problem, specifying the data required, extracting and preparing the data, performing machine learning by clustering the data into 5 clusters based on their similarities, and lastly providing recommendations to the relevant stakeholders i.e. Raymond regarding the best locations to open a new fast food restaurant. To answer the business question that was raised in the introduction section, the answer proposed by this project is: The neighborhoods in cluster 0 and 1 are the most preferred locations to open a new fast food restaurant. The findings of this project will help the relevant stakeholders to capitalize on the opportunities on high potential locations while avoiding overcrowded areas in their decisions to open a new fast food restaurant.