

Opportunity of Opening a Restaurant in Salt Lake City

Capstone Project of Applied Data Science

Kan Zhang

June 11, 2021

1. Introduction

1.1 Problem Statement

We would like to understand which neighborhoods of Salt Lake City area are better for opening a restaurant and which type of restaurant (Chinese, Japanese, American, etc.) it should be.

1.2 Discussion of Background

The Salt Lake City area is drawing a lot of people moving into the area which gives a profit opportunity of opening a restaurant. We would like to first understand the current status of the neighborhoods so we know which areas have more restaurants already. We would also like to understand if demographic factors such as population, household, income, etc. can explain the difference among neighborhoods. Such understanding may give us insights on which neighborhood we should open the restaurant.

2. Data acquisition and cleaning

2.1 Data sources

2.2.1: Neighborhood data for the Salt Lake City area: Neighborhood data is taken from this website: <https://statisticalatlas.com/place/Utah/Salt-Lake-City/Overview>.

2.2.2: Venue data for each neighborhood: This data is taken from foursquare using their API.

2.2.3: Demographic data for each neighborhood: This data is taken from a series of websites based on the list of neighborhoods and factors that we are interested in. Examples of factors can be: population, number of household, age and sex, race and ethnicity, household type (married or not), income, etc. We will use these neighborhoods and factors to scan through a series of websites. One example is: <https://statisticalatlas.com/neighborhood/Utah/Salt-Lake-City/Yalecrest/Race-and-Ethnicity>

2.2 Data retrieval

For the data of neighborhood and demographic factors, the data was retrieved by scraping the corresponding website. Note that the data is not in clean format like a table on website, so it is retrieved by locating the corresponding figure and finding the text in the fixed location. Example of such figure is listed below (Fig. 1).

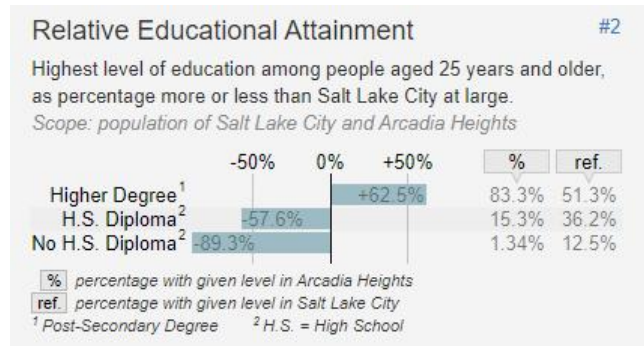


Figure 1. Example of figure on the website where demographic data is retrieved. In this case, the percentage data (83.3%, 15.3%, 1.34%) were retrieved by locating this figure and the values in the html script.

For the data of venues, the retrieval procedure is similar with the class assignment by calling foursquare APIs.

3. Analysis Method and Discussion

3.1 Visualize the neighborhood

This is done by calling Folium map function. See the generated map below (Fig. 2) and corresponding neighborhood list (Fig. 3).

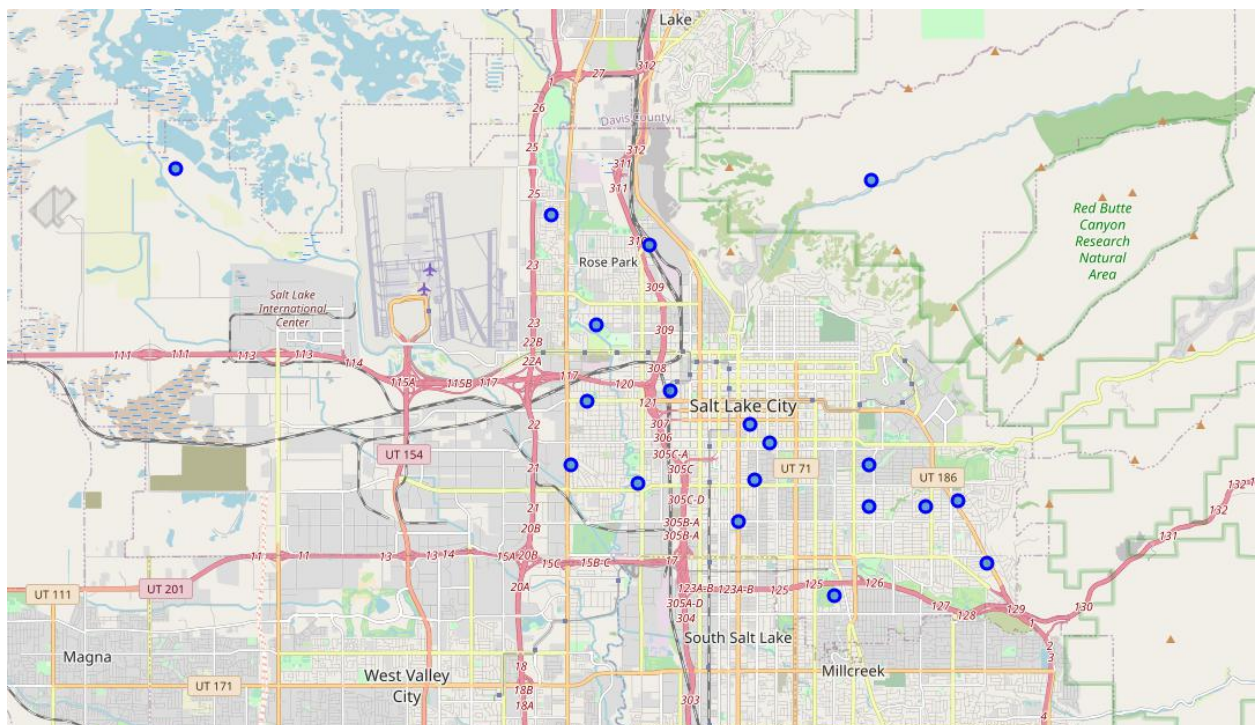


Figure 2. Salt Lake City map with markers for each neighborhood.

	Neighborhood	Latitude	Longitude
0	Arcadia Heights	40.724061	-111.814266
1	Bonneville Hills	40.736822	-111.832721
2	Capitol Hill	40.795715	-111.914489
3	Central City	40.755205	-111.884654
4	Central City-Liberty Wells	40.742818	-111.883353
5	Downtown	40.762823	-111.908463
6	East Bench	40.738281	-111.823032
7	East Central	40.751146	-111.879096
8	Fairpark	40.777661	-111.930275
9	Glendale	40.746247	-111.938072
10	Greater Avenues	40.810052	-111.848884
11	Jordan Meadows	40.609978	-112.020077
12	Liberty Wells	40.733520	-111.888309
13	People's Freeway	40.741981	-111.918079
14	Poplar Grove	40.760501	-111.933048
15	Rose Park	40.802324	-111.943955
16	Sugar House	40.716698	-111.859792
17	Wasatch Hollow	40.736946	-111.849545
18	Westpointe	40.812862	-112.055311
19	Yalecrest	40.746304	-111.849564

Figure 3. List of SLC neighborhoods

3.2 Cluster neighborhoods and visualize the clusters

This is done by calling foursquare APIs and following similar procedure in the class assignment. The visualization in a map is shown in Fig. 4 and the list of output cluster and top venues for each neighborhood is shown in Fig. 5.

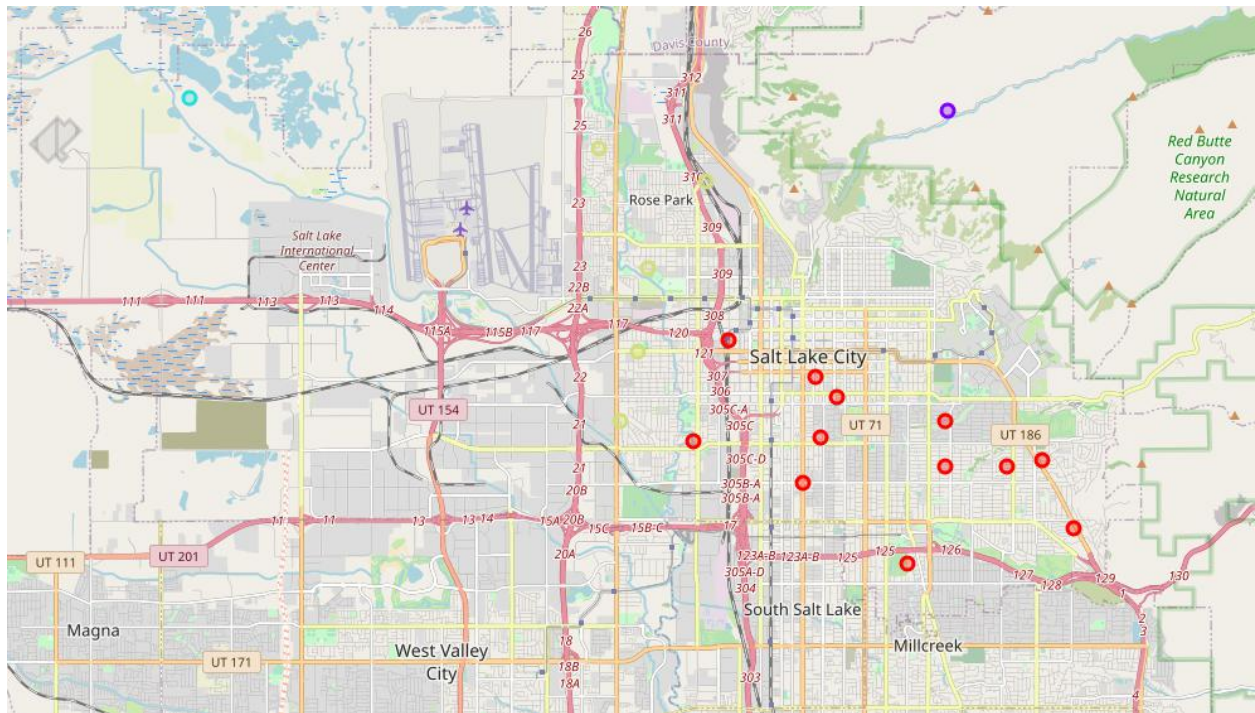


Figure 4. SLC map with clustered neighborhoods

	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Arcadia Heights	40.724061	-111.814266	0	Pizza Place	Video Store	Mexican Restaurant	Zoo Exhibit	Sporting Goods Shop	Bank	Burger Joint	Coffee Shop	Convenience Store	Fast Food Restaurant
1	Bonneville Hills	40.736822	-111.832721	0	Coffee Shop	Video Store	Pizza Place	Shipping Store	Sandwich Place	Mexican Restaurant	Ice Cream Shop	Seafood Restaurant	Salon / Barbershop	Park
2	Capitol Hill	40.795715	-111.914489	3	Mexican Restaurant	Pizza Place	Convenience Store	Park	Grocery Store	Chinese Restaurant	Discount Store	Coffee Shop	Gas Station	Train Station
3	Central City	40.755205	-111.884654	0	Coffee Shop	Bar	Bakery	New American Restaurant	Brewery	Grocery Store	Italian Restaurant	Vegetarian / Vegan Restaurant	Cocktail Bar	Burger Joint
4	Central City-Liberty Wells	40.742818	-111.883353	0	Coffee Shop	Bakery	Bar	Asian Restaurant	Cocktail Bar	Vietnamese Restaurant	Brewery	Vegetarian / Vegan Restaurant	Juice Bar	New American Restaurant
5	Downtown	40.762823	-111.908463	0	Coffee Shop	Mexican Restaurant	Hotel	Brewery	Bakery	Cocktail Bar	Hotel Bar	Bar	New American Restaurant	BBQ Joint
6	East Bench	40.738281	-111.823032	0	Gym / Fitness Center	Park	Video Store	Zoo Exhibit	Trail	American Restaurant	Bakery	Coffee Shop	Deli / Bodega	Pizza Place
7	East Central	40.751146	-111.879096	0	Coffee Shop	Bar	Bakery	New American Restaurant	Burger Joint	Brewery	Cocktail Bar	Grocery Store	Vegetarian / Vegan Restaurant	Asian Restaurant
8	Fairpark	40.777661	-111.930275	3	Mexican Restaurant	Convenience Store	Pizza Place	Gas Station	Hotel	Fast Food Restaurant	Airport Service	Airport Terminal	Burger Joint	Discount Store
9	Glendale	40.746247	-111.938072	3	Fast Food Restaurant	Convenience Store	Gas Station	Mexican Restaurant	Sandwich Place	Bank	Indian Restaurant	Discount Store	Pizza Place	Light Rail Station
10	Greater Avenues	40.810052	-111.848884	1	Trail	Mountain	Donut Shop	Food Truck	Food Stand	Fish & Chips Shop	Financial or Legal Service	Fast Food Restaurant	Exhibit	Ethiopian Restaurant
11	Jordan Meadows	40.609978	-112.020077	3	Fast Food Restaurant	Video Store	Pizza Place	Gym	Burger Joint	Convenience Store	Grocery Store	Gas Station	Big Box Store	Construction & Landscaping
12	Liberty Wells	40.733520	-111.888309	0	Coffee Shop	Bar	Brewery	Thai Restaurant	Bakery	BBQ Joint	Vietnamese Restaurant	Breakfast Spot	Pizza Place	Donut Shop
13	People's Freeway	40.741981	-111.918079	0	Mexican Restaurant	Brewery	Bar	Thai Restaurant	Sandwich Place	BBQ Joint	Coffee Shop	Convenience Store	Pizza Place	Hotel
14	Poplar Grove	40.760501	-111.933048	3	Mexican Restaurant	Fast Food Restaurant	Convenience Store	Pizza Place	Hotel	Gas Station	Bank	Sandwich Place	Grocery Store	Discount Store
15	Rose Park	40.802324	-111.843955	3	Pizza Place	Airport Service	Mexican Restaurant	Discount Store	Gas Station	Sandwich Place	Pet Store	Park	Convenience Store	Bank
16	Sugar House	40.716698	-111.859792	0	Coffee Shop	Pizza Place	Mexican Restaurant	Deli / Bodega	Gym / Fitness Center	Grocery Store	American Restaurant	Bagel Shop	Sushi Restaurant	Diner
17	Wasatch Hollow	40.736945	-111.849545	0	Coffee Shop	Video Store	ATM	American Restaurant	Grocery Store	Gym / Fitness Center	Mexican Restaurant	Bookstore	Thai Restaurant	Pizza Place
18	Westpointe	40.812862	-112.055311	2	Sports Club	Zoo Exhibit	Dog Run	Food Stand	Fish & Chips Shop	Financial or Legal Service	Fast Food Restaurant	Exhibit	Ethiopian Restaurant	Donut Shop
19	Yalecrest	40.746304	-111.849564	0	Pizza Place	Coffee Shop	Grocery Store	Yoga Studio	Bookstore	New American Restaurant	Japanese Restaurant	Video Store	Cosmetics Shop	Deli / Bodega

Figure 5. List of SLC neighborhoods with cluster labels and top venues

3.3 Examine restaurant rate by cluster

Since our goal is to find out the opportunity to open a restaurant. At this step, we would like to see if there is a clear restaurant rate difference by cluster. We derived the restaurant rate by adding all columns with "Restaurant" in column names. We do see cluster 0 and 3 have similar restaurant rate, while cluster 1 and 2 have no restaurant, as seen in Fig. 6. Looking into the top venues for cluster 1 and 2 and it was observed they are mostly outdoor activity related venues,

such as trail, zoo, etc. This information suggests we may not consider opening restaurant in cluster 1 / 2 areas.

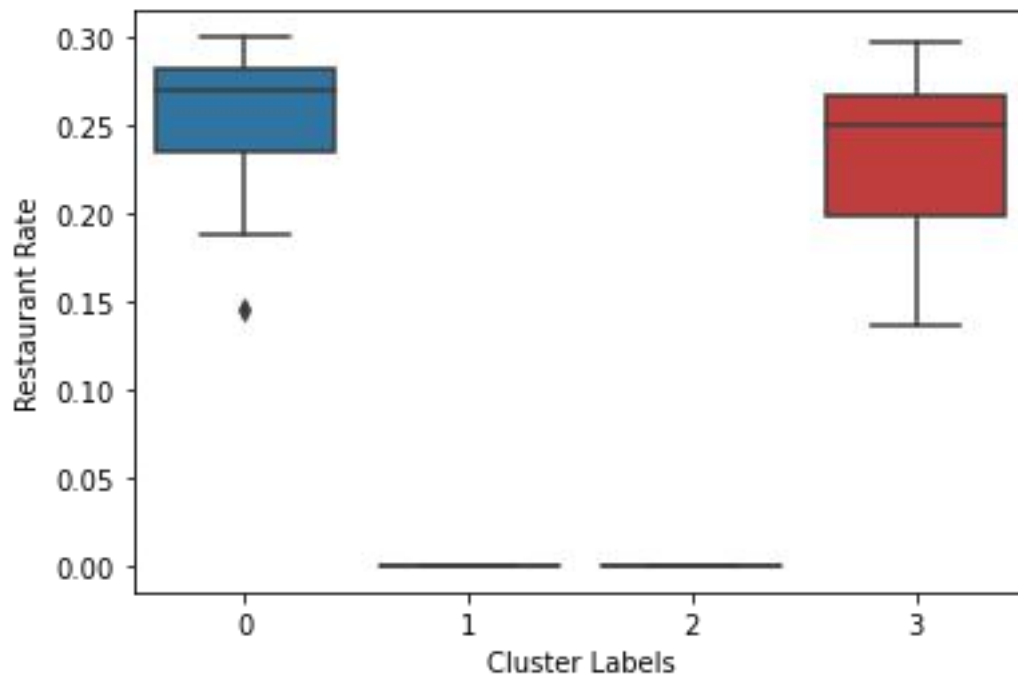


Figure 6. Restaurant rate by cluster label. Cluster 1 and 2 neighborhoods have no restaurant.

3.4 Examine the popular restaurant types

Focusing on the cluster 0 and 3 neighborhoods (since cluster 1 and 2 neighborhoods have no restaurant), we can also examine which restaurant types are more popular. As seen in Fig. 7, top restaurant types are Fast Food, Mexican, Chinese, Asian, Vegan, Thai. Such information suggests we may pick one of the popular restaurant types for better opportunity of making profit.

It was also observed the restaurant rate shows some difference by cluster for each type. It shows cluster 3 neighborhoods have high Fast Food and Mexican restaurants while cluster 0 neighborhoods have more other types of restaurant, as seen in Fig. 8.

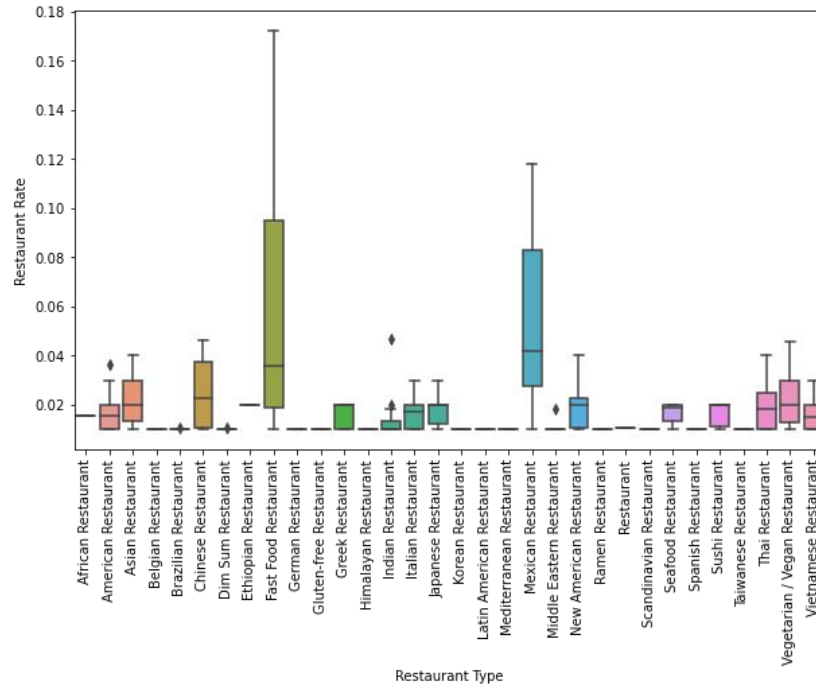


Figure 7. Restaurant rate vs. type using cluster 0 and 3 neighborhoods. Fast Food, Mexican, Chinese, Asian, Vegan, Thai are more popular than others.

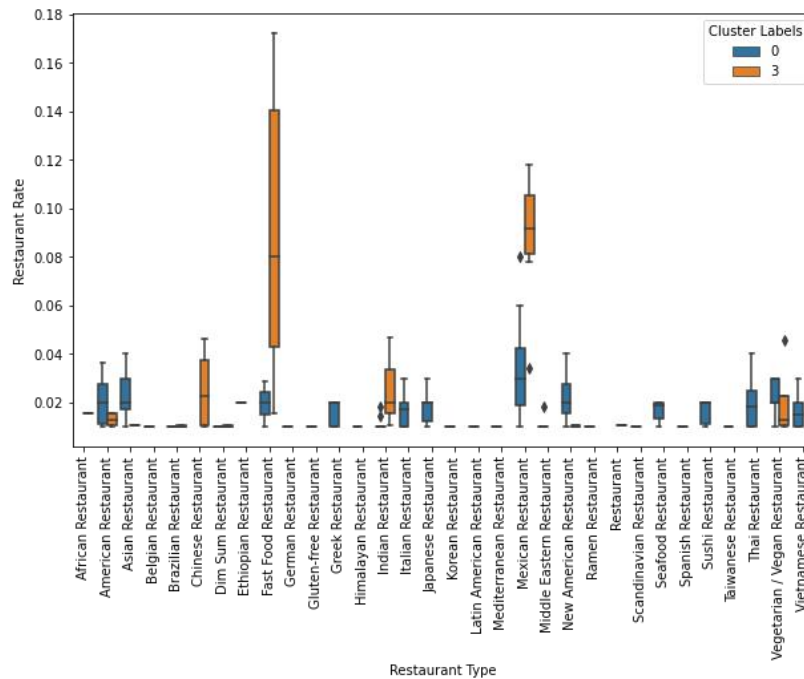


Figure 8. Further separate by cluster for restaurant rate vs. type. It shows cluster 3 neighborhoods have high Fast Food and Mexican restaurants while cluster 0 neighborhoods have more other types of restaurant.

3.5. Examine demographic factor difference by cluster

As we have seen the restaurant rate vs. type shows some cluster delta. We followed up to see if there is any cluster delta between demographic factors, so that we can use this understanding for feature selection.

Since the demographic factors' values vary significantly among factors. For example, population values are quite different than income values. We grouped the factors with similar level of values together in plots so that we can better see the difference by cluster. Fig. 9 shows the cluster difference for population and household count, but no strong cluster delta is observed.

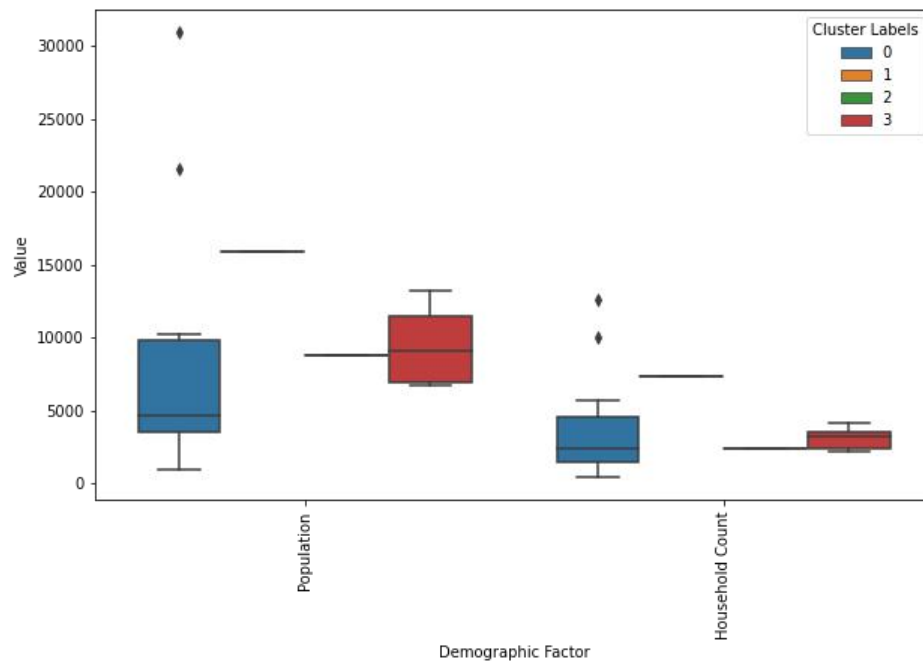


Figure 9. Population and household count by cluster labels. No strong delta is observed.

Fig. 10 shows income values by cluster. We do see cluster 0 neighborhoods generally have higher income than cluster 3 neighborhoods.

Fig. 11 shows other percentage factors (age, race, married household, education) by cluster. We can see age, race, education generally show some delta between clusters.

After all, we picked these columns as features for building a multivariate linear regression model: 'Income 80percentile', 'Income 60percentile', 'Income 50percentile', 'Age Senior Percentage', 'Age Old Adult Percentage', 'Age Young Adult Percentage', 'Race White Percentage', 'Education High Degree Percentage'.

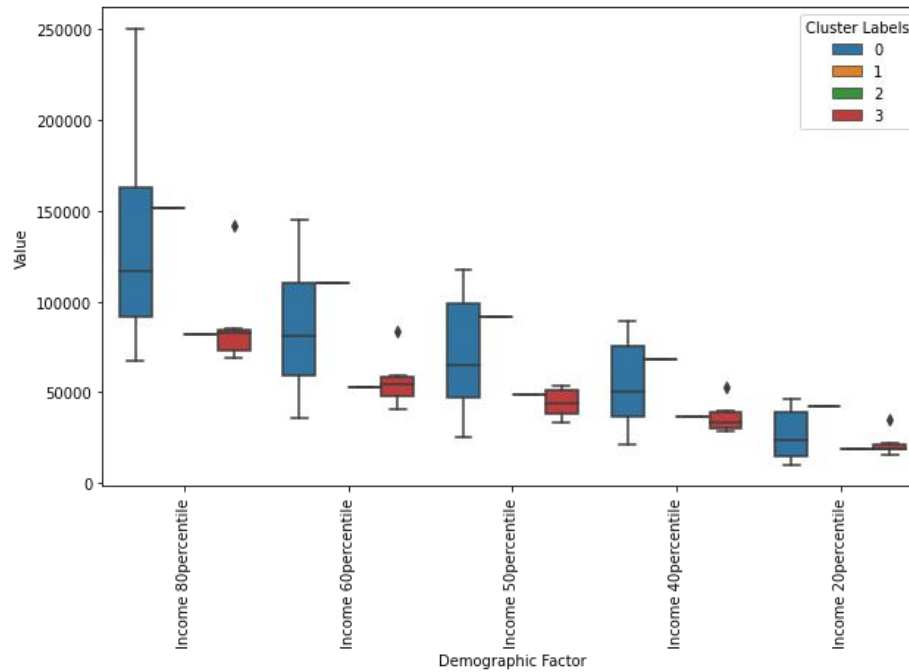


Figure 10. Income (80%tile, 60%tile, 50%tile, 40%tile, 20%tile) by cluster labels. Cluster 0 neighborhoods show higher income than cluster 3 neighborhoods.

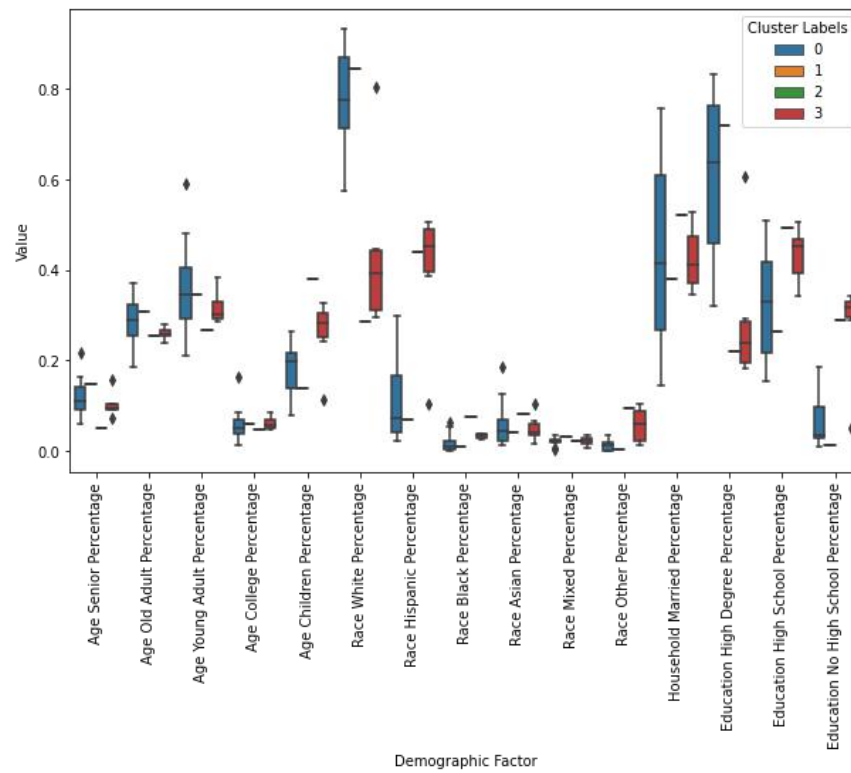


Figure 11. Age, race, married household, education factors by cluster labels.

3.6 Multivariate linear regression model to predict restaurant rate

We conducted the multivariate linear regression modelling for each type of restaurant and the average restaurant rate vs. neighborhoods. The fitting score is listed in Fig. 12.

	Restaurant	Fit Score
0	African Restaurant	0.326475
1	American Restaurant	0.637694
2	Asian Restaurant	0.426773
3	Belgian Restaurant	0.623991
4	Brazilian Restaurant	0.605012
5	Chinese Restaurant	0.266289
6	Dim Sum Restaurant	0.495331
7	Ethiopian Restaurant	0.559323
8	Fast Food Restaurant	0.625155
9	German Restaurant	0.771933
10	Gluten-free Restaurant	0.419568
11	Greek Restaurant	0.707848
12	Himalayan Restaurant	0.404605
13	Indian Restaurant	0.321377
14	Italian Restaurant	0.701652
15	Japanese Restaurant	0.780351
16	Korean Restaurant	0.396128
17	Latin American Restaurant	0.396128
18	Mediterranean Restaurant	0.573126
19	Mexican Restaurant	0.562255
20	Middle Eastern Restaurant	0.861567
21	New American Restaurant	0.513664
22	Ramen Restaurant	0.503452
23	Restaurant	0.104762
24	Scandinavian Restaurant	0.694861
25	Seafood Restaurant	0.801604
26	Spanish Restaurant	0.586915
27	Sushi Restaurant	0.646423
28	Taiwanese Restaurant	0.396128
29	Thai Restaurant	0.720832
30	Vegetarian / Vegan Restaurant	0.392854
31	Vietnamese Restaurant	0.800379
32	Restaurant Rate	0.599236

Figure 12. Fitting score list for restaurant types

Cross-referencing the popular restaurant types, we can see Fast Food and Thai are both popular and have relatively higher fitting scores (>0.6 , which is the fitting score for average restaurant rate). This information suggests we may consider opening either a fast food or a Thai restaurant since they are relatively popular and also we have a relatively good model to predict its restaurant rate.

We can then compare the predicted restaurant rate vs. its actual for fast food and Thai. If predicted value is higher than the actual value for a neighborhood, then it may suggest the demographic factors supports more such restaurant in this neighborhood and opening a new restaurant of this type in this neighborhood may make a profit. The comparison data is shown in the table in Fig. 13 where we can see “Central City” can be a good option for Thai restaurant while “Rose Park” can be a good option for Fast Food restaurant.

	Neighborhood	Thai Restaurant Actual	Thai Restaurant Predicted	Fast Food Restaurant Actual	Fast Food Restaurant Predicted	Thai Predicted - Actual	Fast Food Predicted - Actual
0	Arcadia Heights	0.000000	-0.005528	0.028986	0.038636	-0.005528	0.009651
1	Bonneville Hills	0.010000	0.008553	0.000000	0.002135	-0.001447	0.002135
2	Capitol Hill	0.000000	0.008515	0.015385	0.004326	0.008515	-0.011058
3	Central City	0.010000	0.021955	0.000000	0.002097	0.011955	0.002097
4	Central City-Liberty Wells	0.020000	0.018675	0.000000	0.011331	-0.001325	0.011331
5	Downtown	0.010000	0.000884	0.010000	0.016775	-0.009116	0.006775
6	East Bench	0.018182	0.014171	0.000000	0.002845	-0.004011	0.002845
7	East Central	0.010000	0.017578	0.000000	-0.009592	0.007578	-0.009592
8	Fairpark	0.000000	0.001715	0.043011	0.075705	0.001715	0.032694
9	Glendale	0.000000	0.000012	0.140625	0.085555	0.000012	-0.055070
11	Jordan Meadows	0.000000	0.005975	0.172414	0.102288	0.005975	-0.070126
12	Liberty Wells	0.040000	0.024856	0.000000	-0.007126	-0.015144	-0.007126
13	People's Freeway	0.040000	0.029808	0.000000	0.033383	-0.010192	0.033383
14	Poplar Grove	0.000000	-0.003922	0.080000	0.105962	-0.003922	0.025962
15	Rose Park	0.000000	0.005243	0.000000	0.064894	0.005243	0.064894
16	Sugar House	0.010000	0.014041	0.020000	-0.011964	0.004041	-0.031964
17	Wasatch Hollow	0.030000	0.030824	0.000000	-0.013418	0.000824	-0.013418
19	Yalecrest	0.020000	0.024827	0.000000	0.006587	0.004827	0.006587

Figure 12. Restaurant rate actual vs. predicted for each neighborhood for Fast Food and Thai restaurant

We can also visualize the actual vs. predicted in scatter plots. Fig. 13 shows such plot for Thai restaurant highlighting the suggested neighborhood of “Central City”. Fig. 14 shows similar plot for Fast Food restaurant highlighting “Rose Park” neighborhood.

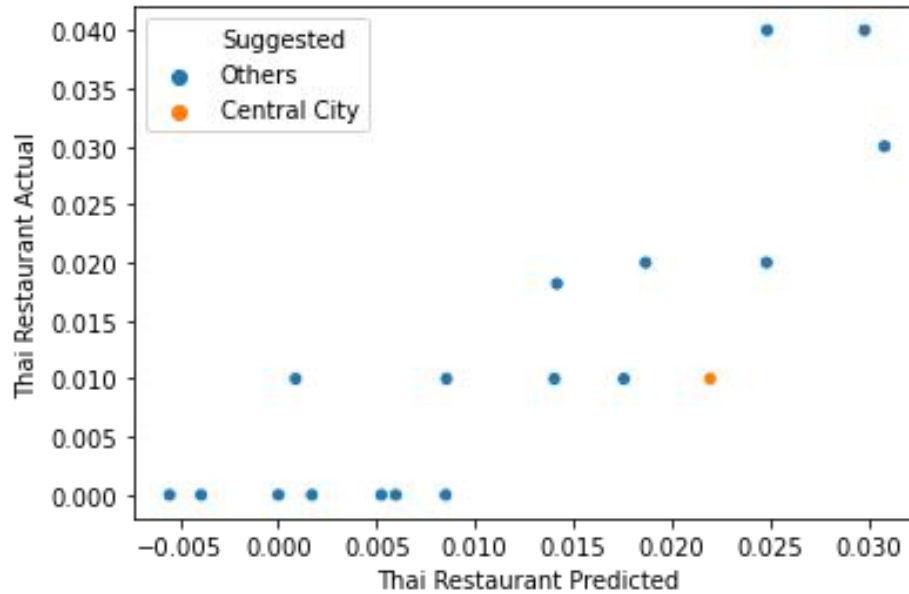


Figure 13. Thai restaurant rate actual vs. predicted, highlighting “Central City” where predicted is higher than actual, suggesting an opportunity of opening a new restaurant in this neighborhood.

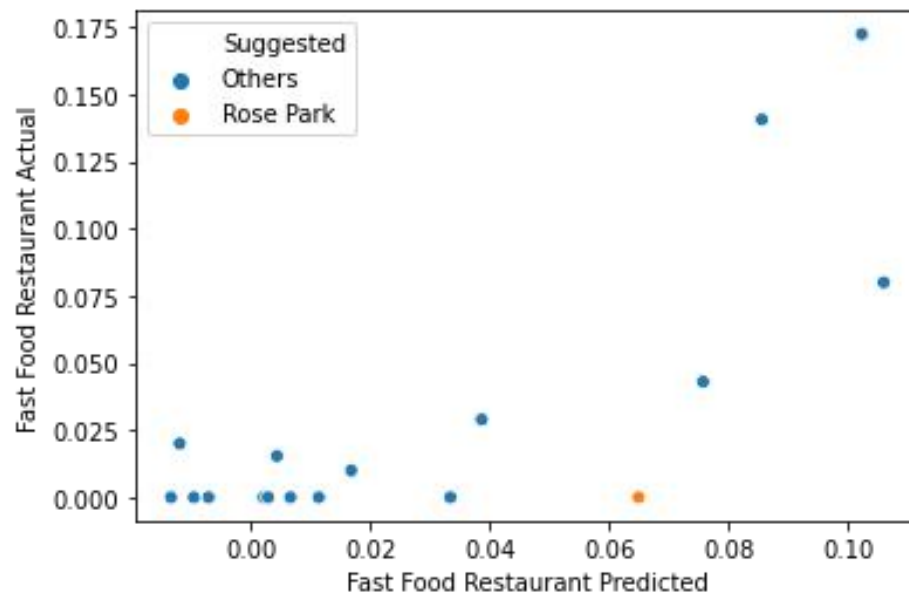


Figure 14. Thai restaurant rate actual vs. predicted, highlighting “Rose Park” where predicted is higher than actual, suggesting an opportunity of opening a new restaurant in this neighborhood.

4. Conclusion

This analysis examined the neighborhoods of Salt Lake City for their restaurant rate by type and demographic factors in order to understand the opportunity of opening a new restaurant in one of the neighborhoods. The analysis found out the most popular restaurant types and their

corresponding multivariate linear regression models. Seeking a balance between popularity and model fitting score, it is narrowed down to two types of restaurant: Thai and Fast Food. Using the fit models, the recommendation is: for opening a new Thai restaurant, "Central City" neighborhood is a good option; for Fast Food restaurant, "Rose Park" neighborhood is a good option.