**Opportunity of Opening a Restaurant in Salt Lake City**

# Kan Zhang

## June 11, 2021

### **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.

### **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).

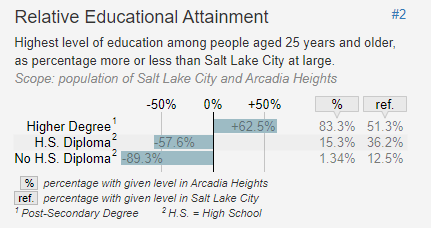


Figure 1. Example of figure on the website where demographic data is retrieved. In this case, the percentage data (83.3%, 15.3%, 1.34%) were retrieved by locating this figure and the values in the html script.

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

### **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).

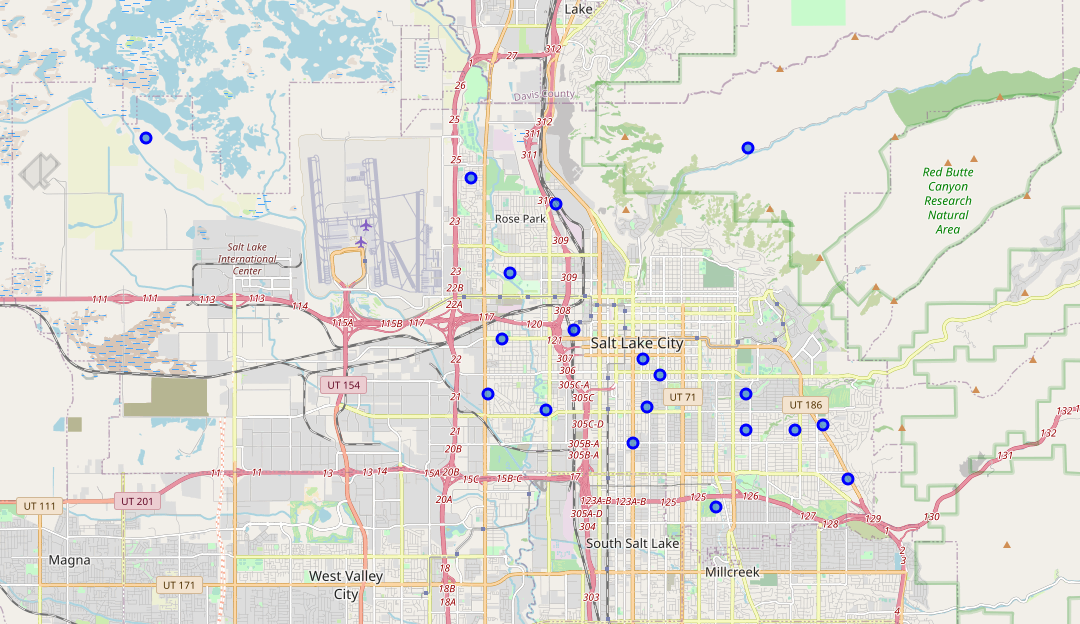


Figure 2. Salt Lake City map with markers for each neighborhood.



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.

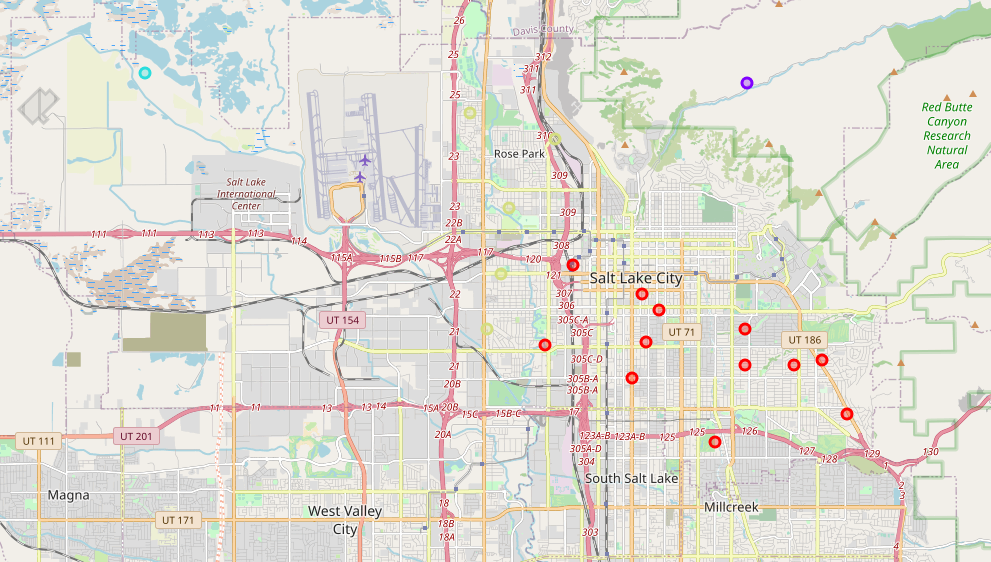


Figure 4. SLC map with clustered neighborhoods

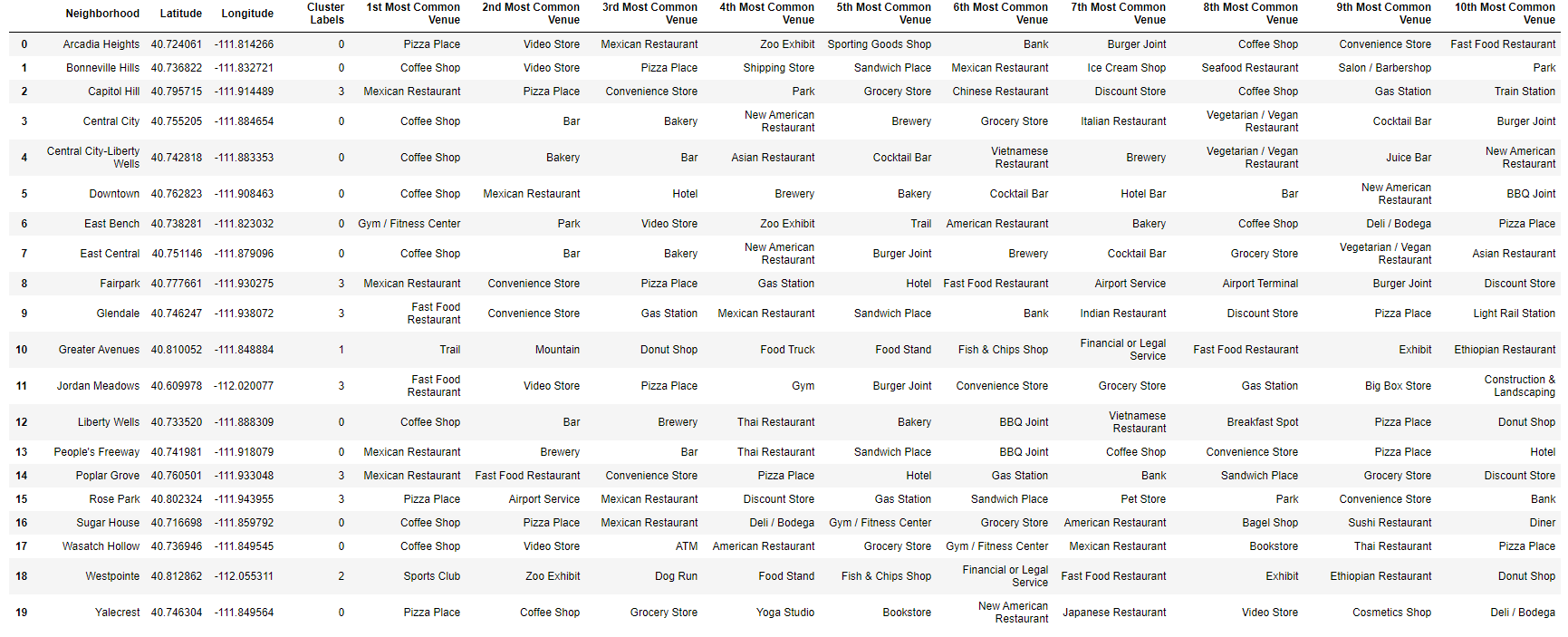


Figure 5. List of SLC neighborhoods with cluster labels and top venues

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.

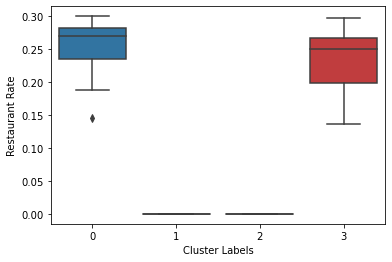


Figure 6. Restaurant rate by cluster label. Cluster 1 and 2 neighborhoods have no restaurant.

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.

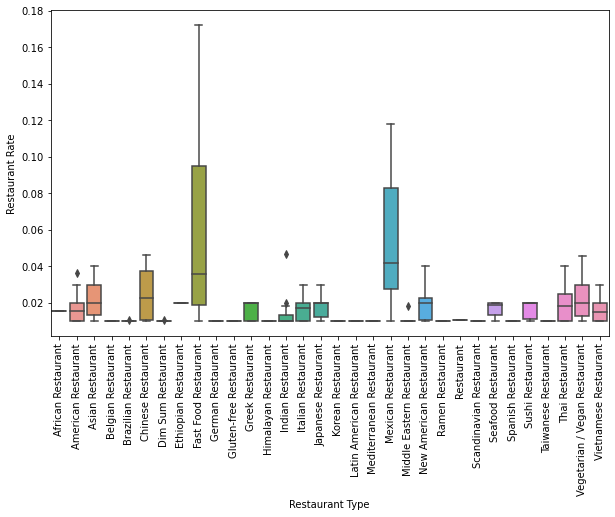


Figure 7. Restaurant rate vs. type using cluster 0 and 3 neighborhoods. Fast Food, Mexican, Chinese, Asian, Vegan, Thai are more popular than others.

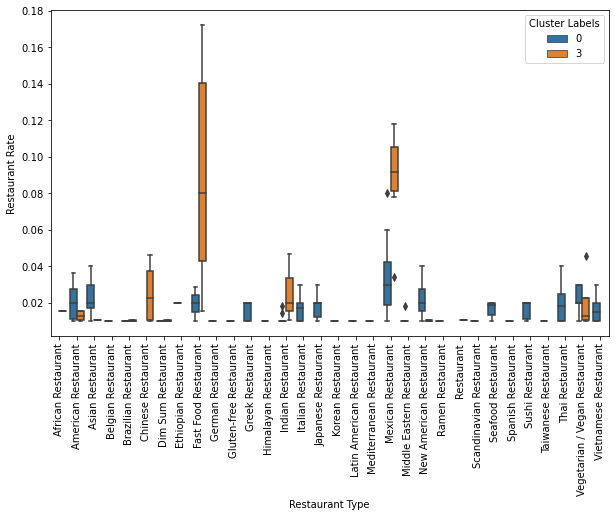


Figure 8. Further separate by cluster for restaurant rate vs. type. It shows cluster 3 neighborhoods have high Fast Food and Mexican restaurants while cluster 0 neighborhoods have more other types of restaurant.

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.

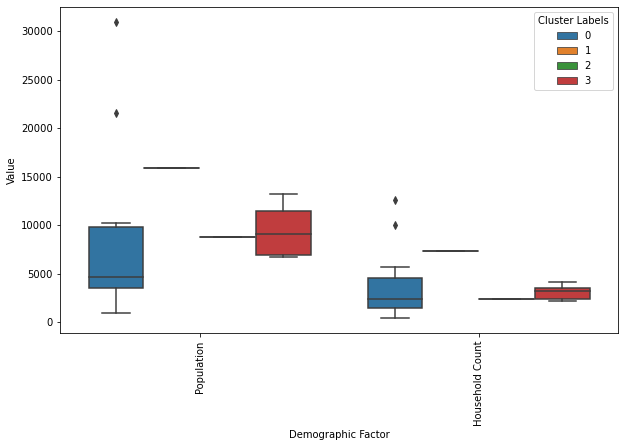


Figure 9. Population and household count by cluster labels. No strong delta is observed.

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'.

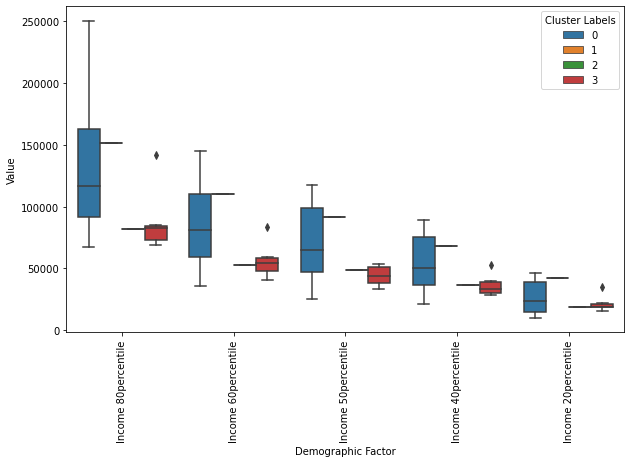


Figure 10. Income (80%tile, 60%tile, 50%tile, 40%tile, 20%tile) by cluster labels. Cluster 0 neighborhoods show higher income than cluster 3 neighborhoods.

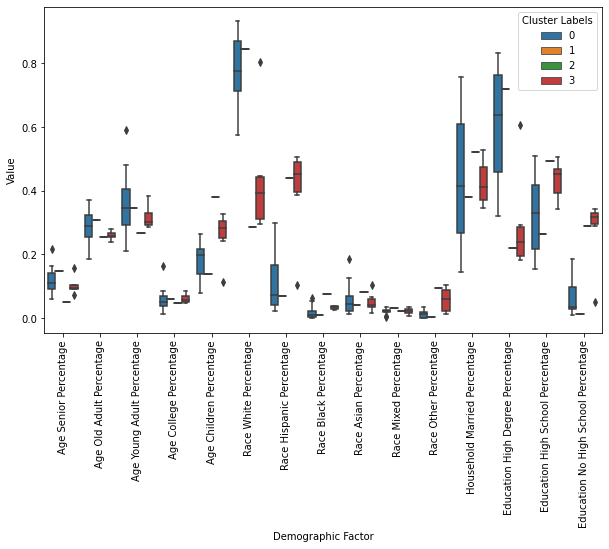


Figure 11. Age, race, married household, education factors by cluster labels.

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.



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.

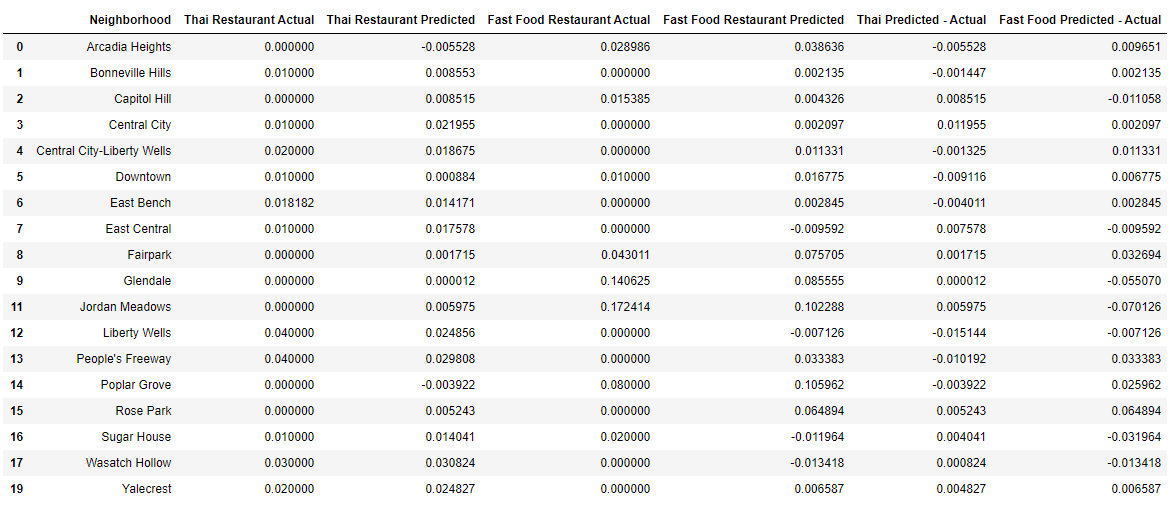


Figure 12. Restaurant rate actual vs. predicted for each neighborhood for Fast Food and Thai restaurant

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.

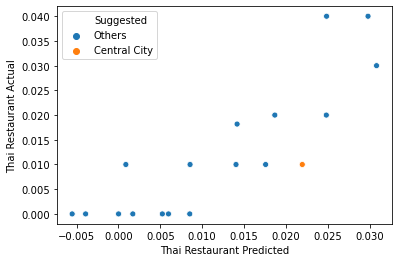


Figure 13. Thai restaurant rate actual vs. predicted, highlighting “Central City” where predicted is higher than actual, suggesting an opportunity of opening a new restaurant in this neighborhood.

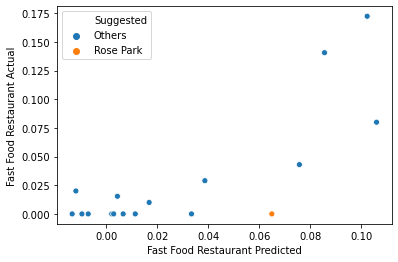


Figure 14. Thai restaurant rate actual vs. predicted, highlighting “Rose Park” where predicted is higher than actual, suggesting an opportunity of opening a new restaurant in this neighborhood.

### **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.