**Elyse Hallstrom**

**Applied Data Science Capstone**

**May 2020**

**Restaurant Location Selection**

*Introduction*

Our client is wanting to open a new American Restaurant in Phoenix, AZ. They are a successful national restauranteur, but they have never opened a restaurant in Arizona, so they are unfamiliar with the region. They would like for us to recommend the best location for them to open their restaurant. The restaurant is upscale American, so their primary concern is that they are in an area that can afford to pay for the higher prices. They have had restaurant failures in the past where they did not align the price point for their food, the general interest in the surrounding area for that type of food, and the neighborhood’s ability to afford it.

We believe that this type of analysis can be recreated for any city or restaurant type. It could even be extended for other types of businesses as well.

*Data*

There are three datasets that we will use for this analysis: Foursquare API, Open Datasoft zip code longitude and latitude data, and Zip Atlas average household income by zip code.

Foursquare is a social media database that contains geolocation data for all kind of businesses and restaurants. There are many features to this data, but we utilize the Venue Category, the longitude, and the latitude for each location. See Figure 1 for an example of the data that is available via Foursquare.

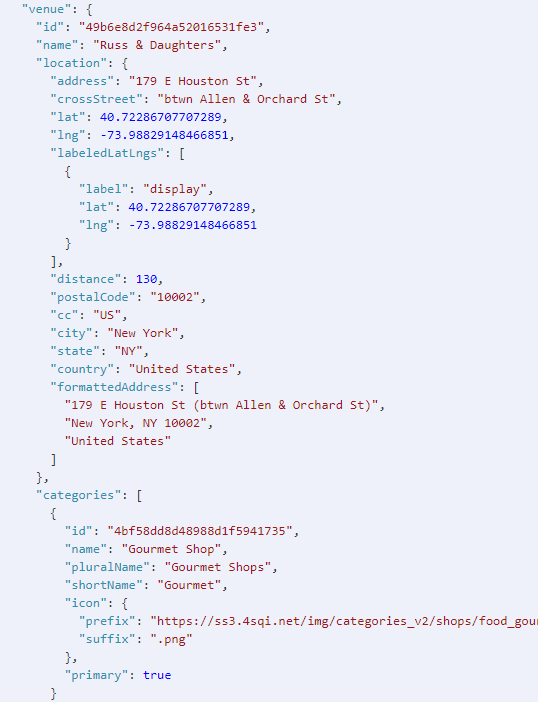


Figure 1: Foursquare API Example

Open Datasoft has free datasets that contain zip code information. This data contains all of the United States zip codes, their city, state, longitude, and latitude. We will use this data to get the list of Phoenix zip codes.

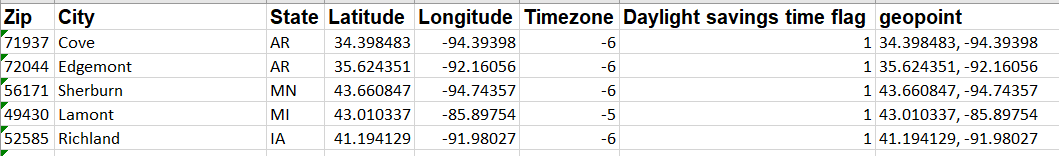


Table 1: Open Datasoft Zip Code Dataset

Zip Atlas publishes interesting datasets, one of which is the average household income by zip code. We will use this data as a way to determine whether people have enough disposable income to indulge in a more expensive meal.

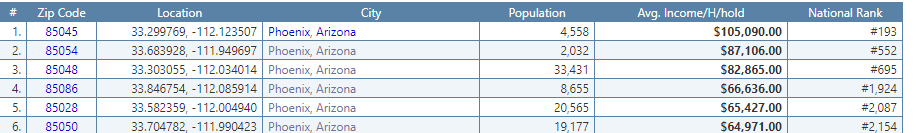


Table 2: Zip Atlas Average Household Income by Zip Code

*Methodology – exploration, statistics, what methods and why*

To join all three datasets together, we first had to clean them. Zip code data needed to be filtered down to Arizona and Phoenix for state and city, respectively. After that was completed, we only kept the zip code, longitude, and latitude. For the income dataset, we dropped everything but zip code, city (which contained city and state), and average household income. We split city into city and state as separate columns in our dataframe. We renamed the income column to ‘Avg Income’ from ‘Avg. Income/H/hold’ for ease of use. We then merged this dataset with the zipcode data for Arizona, see Table 3.

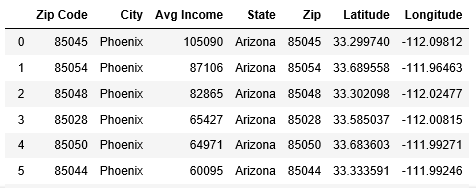


Table 3: Merged Income and Zip Code Dataframe

Once we had this dataframe, we ran through calling the Foursquare API for each zip code, using the getNearbyVenues function from the capstone labs. This resulted in 440 venues in the city of Phoenix. This dataset includes more than restaurants, as you can see from Table 4.

****

Table 4: Results from Foursquare API for Phoenix Zip Codes

Once we had the Foursquare data, we explored the data grouping it by neighborhood to see how many venues were in each zip code and by venue category just so we could see what the most common places were, see Table 5. It looks like American Restaurant is the 3rd most common type of venue in Phoenix, so American Restaurant seems like a good option for the type of restaurant to open here since they appear to be successful.

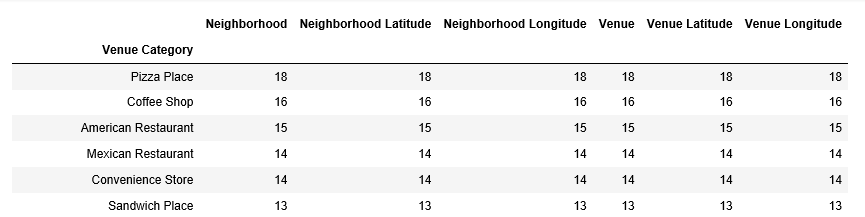


Table 5: Number of places by venue type

Using the raw Foursquare dataset, we created one hot encoding by venue type, then took the average by neighborhood as is seen in Table 6.

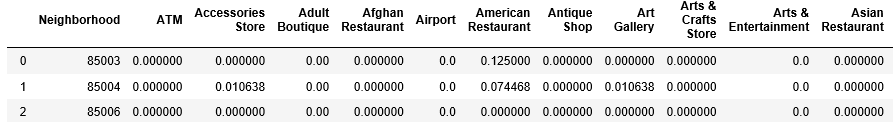


Table 6: Average by Venue by Neighborhood

We took a subset of the data from this dataframe to look at the top 10 zip codes with the highest average American Restaurants, as in Table 7. It looks like only 6 zip codes actually have American Restaurants. Table 7 is sorted from highest to lowest.



Table 7: Top 10 zip codes with highest average American Restaurants

We then merged this data with the income and zip code dataframe so that we could do easy analysis based on the American restaurant density by zip code and income.

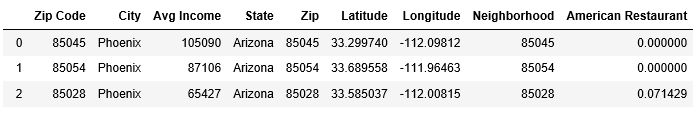


Table 8: Full dataset with income, zip code, and American Restaurant average

As is oftentimes the case with clients, they do not realize what questions they should be asking and it’s our job as data scientists to find them and ask them. One of the questions we had while putting this together is “Is average household income what we should be looking at to determine where to build our new restaurant?” We wanted to explore this further so we investigated the r squared coefficient to determine if there was indeed a correlation between American Restaurant density and Average Household income. As you can see in Figure 2, the r-squared is very poor but also is negatively correlated. Meaning that the higher income zip codes have fewer American Restaurants.

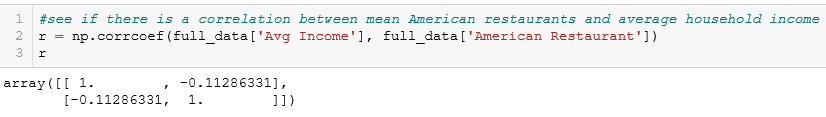


Figure 2: R-squared coefficient between American Restaurant density and Average Household income

With this finding, we wanted to increase the number of options we presented our client. We wanted to answer their original question of which zip code to look at to build their new American Restaurant that had the highest average income. But we also wanted to highlight where American Restaurants have been the most successful.

*Results*

The first proposal is to look at the highest income zip code and see what the most popular types of restaurants are there. 85045 is the most affluent zip code in Phoenix. Looking at the venue’s, however, you can see that there aren’t any restaurants at all. Since the most popular venues are trails, it makes us think that there aren’t a lot of commercial spaces in this zip code. If the client is able to find restaurant space here, it may do very well without much competition.

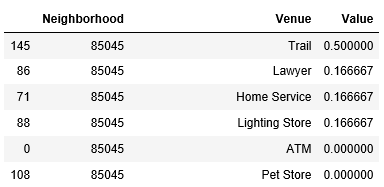


Table 9: Highest income zip’s most popular venues descending

The alternative is to look at the zip code with the highest average American Restaurants. 85021 is this zip code. You can see that it only had playground/park space and American Restaurants. So this zip code might be an alternative location to investigate since American Restaurants have been successful here.

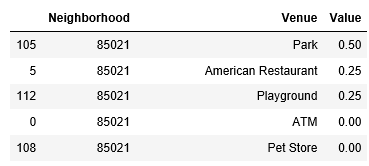


Table 10: Most American Restaurant filled zip code

*Discussion*

There are additional analyses that might be investigated after initial review with the customer, depending on their response to the analysis. One might be to find a different location dataset to group zip codes into neighborhoods. Based on the few data points per zip code, we wonder if we need to group multiple zip codes together into neighborhoods in order to have a more compelling analysis. At the beginning of this project, we looked for Phoenix neighborhood definitions by zip codes and were unsuccessful in finding anything. Living in Phoenix this makes sense to me as the neighborhood designation is something that has only developed in the last 5 years or so, so the fact that there is not a formal definition of neighborhood by zip code is unsurprising.

An additional investigation that would be good to perform would be to look at the lifespan of these restaurants and use that as a metric to determine what kinds of restaurants are the most successful. We would have to see if the Foursquare dataset contains this kind of information or if we would have to move towards something like Yelp for that kind of clarity.

*Conclusion*

Based on the request to determine where to build a new American Restaurant, we would recommend to our client to investigate zip codes 85021 and 85045. This analysis can be altered to look at different kinds of venues if the client decides they would like to look at a different type of restaurant to open up or if we get similar requests in the future.