Profiling the Marketable Audience of Japanese Restaurant-Goers in Manhattan

Afnan Rehman

June 16, 2019

1. Introduction

The problem I am trying to solve is one of correlating user data to venue data. Which types of users frequent which kinds of restaurants? To narrow the scope and keep it relevant to the audience I will specify it to be of the category "Asian Restaurant". This type of analysis could aid in marketing efforts of a variety of locations in the area of focus.

For this analysis I will focus on the area of Manhattan, New York. Specifically, in the areas around the Empire State Building, where foot traffic and tourism would be most densely concentrated.

My audience of stakeholders would be owners of Japanese cuisine restaurants in the Manhattan area, both local and chain restaurants that could benefit from this data to improve their targeted marketing.

2. Data and Cleanup

Data will be taken from the Foursquare database and primarily use the Venues and Users endpoints for the requests. Data will be matched based on the Check-in data to get details for the restaurant and the user in order to establish a connection. We will then establish counts for things like male/female patrons, home cities, friend counts, and so on.

Based on our definition of the problem, the following data sources will be needed:

- Candidate areas based on a radius surrounding a central location (in our case the Empire State Building)
- Number of restaurants and their locations based on the Foursquare API calls
- User data matched from check-ins to those restaurants, matched by unique ids provided by Foursquare

Data was also sourced from Google location APIs that allowed for precise tracking of the longitude and latitude of the center point and radius of our analysis. Using this geographic box, we were able to feed the Foursquare API with enough location data to provide an accurate picture of which restaurants we should be concerned with.

Data was originally formatted in a versatile, but rather unwieldy JSON format. Converting this into a nice and neat tabular format is luckily fairly straightforward using the python pandas data frames. Missing data was left blank, or, if the value was necessary to the analysis, it was replaced with the mode of the data.

Some cases involved restaurants with very few to no results for user check-ins. These restaurants would still be included but would instead use the average profile of the restaurants as their recommendation. This would ensure that actionable results could be given even with lack of data for that restaurant.

3. Exploratory Data Analysis

During the exploratory data analysis phase, 30 establishments of interest were identified in the defined search radius around the Empire State Building. Of these, 21 were identified as restaurants and 9 were of various other categories. I decided to focus the analysis on those 21 restaurants.

Sampling the data for users checking into the closest restaurant to the tower, we found that we can sample several features in the data to use in our analysis. Namely: Gender, Friend Count, Home City, and Number of Tips. These allow us to track the user profile to an extent and even pursue some analytics relating to their engagement on the foursquare platform.

Other features were dropped, such as Bio, Lists, Contact info, name, and photo. Bio, lists, and contact info were often left blank in the fields and offered no useful insight due to this fact. If the Bio field were consistently used, there would be a chance of using keyword searches and possibly natural language processing to create a fuller profile of the typical Japanese restaurant-goer.

The next step was to mine all of this information from the check-in data and get a picture of where we stood with these metrics and count for any frequent occurrences.

4. Results

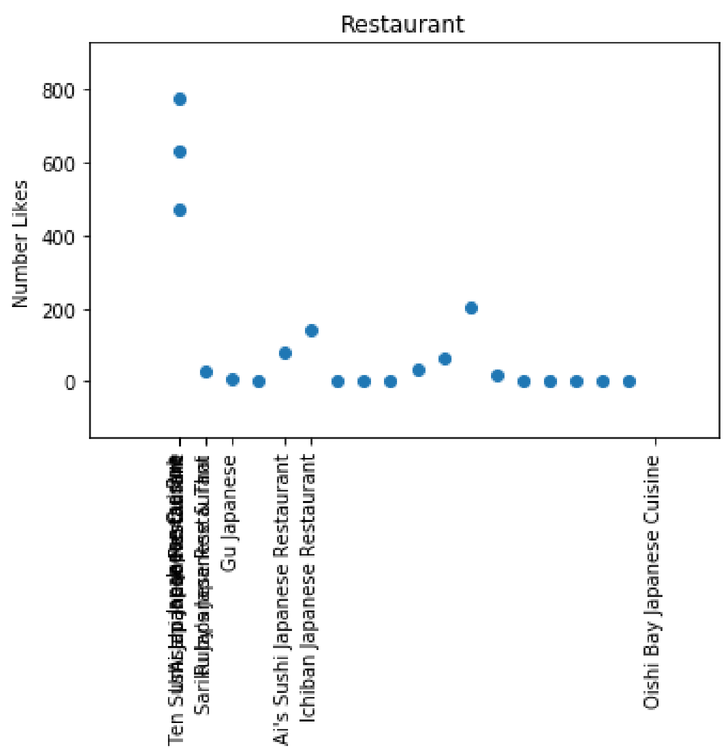

Figure 1 Likes per restaurant

Users predominantly fall into two categories when it comes to number of friends they have. Either there are a lot of friends or very few if any.

This tells us there is perhaps a value in targeting advertising to users who have a lot of friends on the app. This would narrow down the number of recipients as well as ensure reaching a wider audience through the recommendation of those individuals.

Looking at gender next, we found there to be no clear advantage to either gender as there was 17 female, 14 male, and 27 not identified. It seems that in this case there might be a slight preference towards females, but we cannot tell with much of the information obscured by the fact that many of the users choose not to identify their gender.

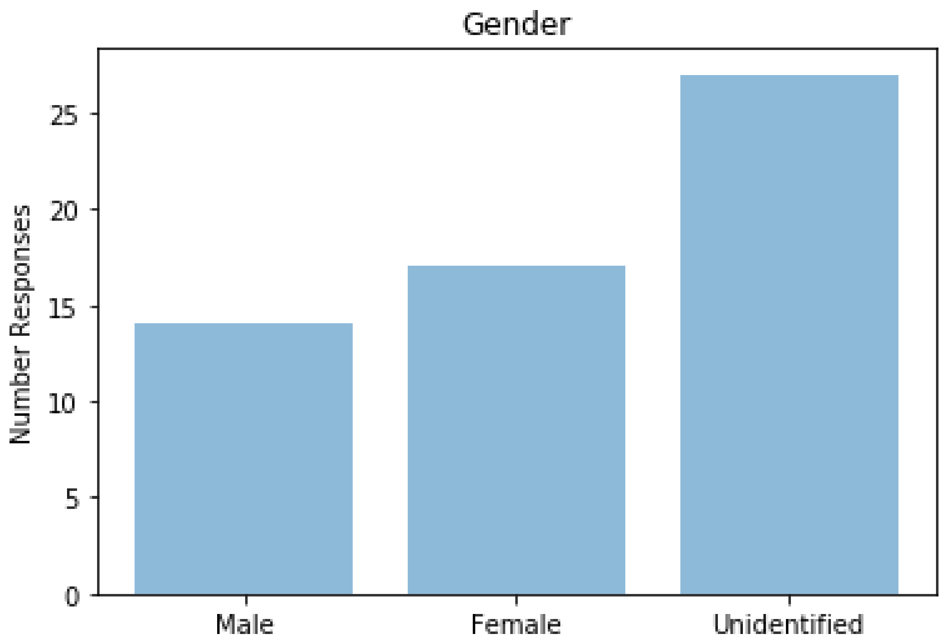

Figure 2 Gender Distribution amongst all restaurants

We should avoid targeting one gender or another in marketing in order to ensure that we reach the current audience in keeping with the current statistics.

5. Conclusion

Overall, this analysis has shown that it is best to target high-usage and high-engagement users of the platform, and then cast as wide a net as possible after that. With the analysis being centered around a more traditional cuisine, there seem to be no obvious bias in gender to be had that would greatly affect marketing strategy.

In the future, more detailed user profiles or even surveys could be used to mine more data and draw better conclusions. Given the user information of the managers of these venues, I would have access to more and cleaner statistics, which would also improve these analyses greatly.

Overall, these analyses can help restaurants in the area either discover new information, or more likely, confirm their preconceived notions based on their actual experience. Providing data backing to their experience will assist them in making decisions with confidence.