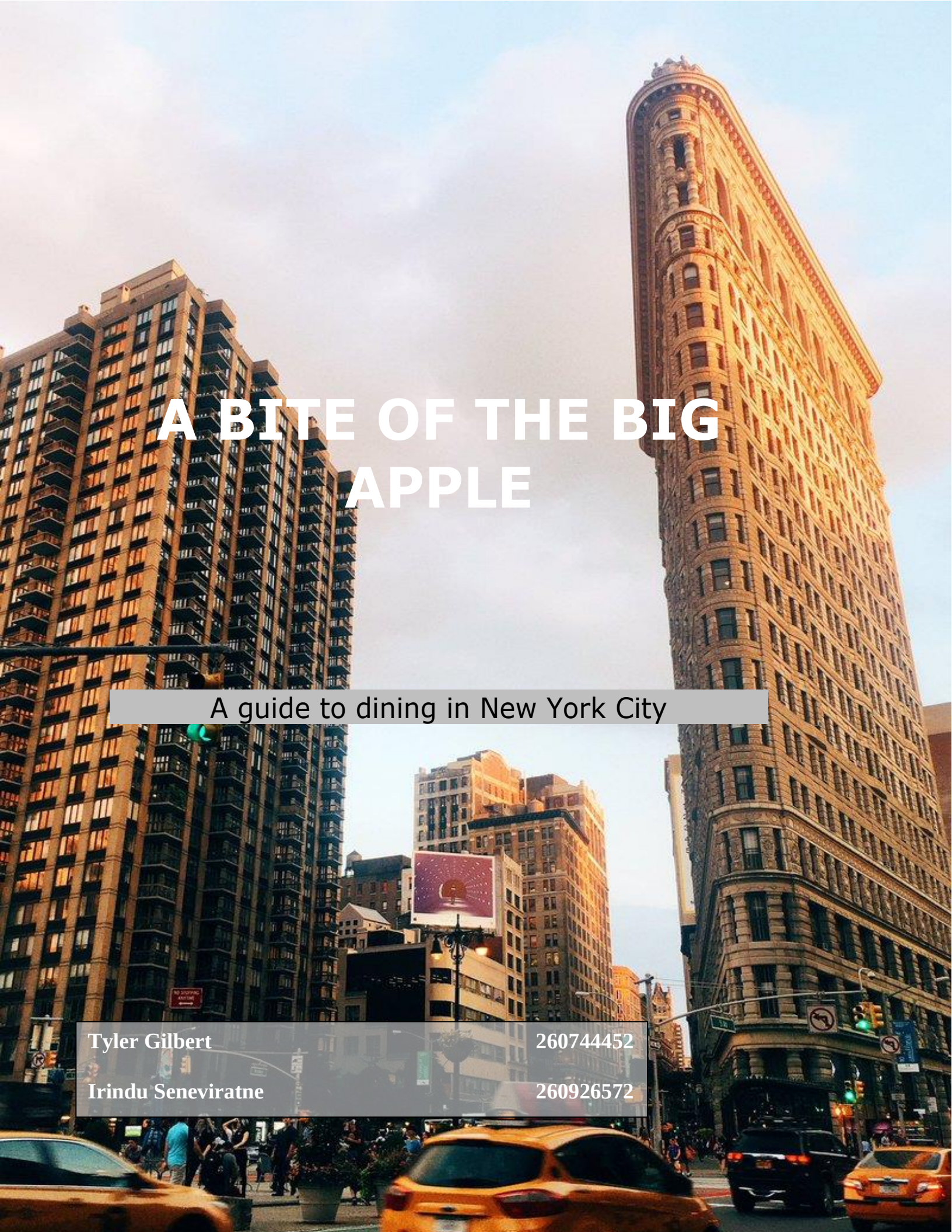


A BITE OF THE BIG APPLE

A guide to dining in New York City

Tyler Gilbert **260744452**

Irindu Seneviratne **260926572**

Contents

1. Introduction

1.1 Overview

NYC Restaurant Week is a biannual culinary festival that offers food enthusiasts with prix-fixe lunch and dinner menus in the city. This gives diners the opportunity to explore some of the top eateries in the city as there is a large selection of top restaurants that participate, and this will also allow customers to try new cuisine and be part of the growing culinary movement in NYC. This report is aimed at providing a unique dining experience for diners in Manhattan, New York during this event. The study relies on information gathered from ‘NYC Restaurant Week’ in 2018, with the hopes of providing valuable insight for selective diners looking to have an ideal night out in the city based on; the ambiance of the restaurant, value for money and the neighborhood that they choose to dine to name a few aspects that are considered.

1.2 Objective

By dissecting the gathered data, we hope to provide a guide to dining during future Restaurant Weeks that are sure to take place, and direct customers to some of the best the city has to offer while ensuring that the customers needs are met throughout the dining experience. The idea behind this project was fueled by the frustration created by the overwheling number of restaurants taking part during this event, that often affects decision making when it comes to dining out, and eventually affects the overall experience. By providing data driven insight, diners can be more adventurous and be more selective of where they would like to dine, not limited by price or cuisine, but also by the level of noise and activity in each neighborhood in Manhattan. By using data on noise complaints in NYC, the study was able to extend the scope to how certain parts of the city will cater to different diners, especially when it comes to making plans after finishing their meals. Moreover, by analyzing noise activity in different neighborhoods it is easy to direct restaurant goers to specific areas based on their plans following dinner. It has become evident that most of the loudest neighborhoods in Manhattan correspond to areas that have a bustling and vibrant nightlife. Therefore, directing a diner looking to have a few drinks or experience a night out in Manhattan could be directed to a restaurant either in or within walking distance to one of the louder neighborhoods the city has to offer.

2. Data Description

2.1 Data-Sourcing

The data gathered for this project was obtained from kaggle.com. Using two datasets:

- NYC Parties
- NYC Restaurant Week

2.2 Pre-Processing

The first step to pre-processing the data was to filter out all the noise complaints in the NYC Parties dataset to only those which occurred in Manhattan. We did so because the restaurants in the restaurant week dataset only contain restaurants located in Manhattan, and this was chosen as the area to focus on. We then added a new column which contains the noise-count and gave every row (occurrence of a noise incident) a count of 1. Next, we dropped irrelevant columns such as ‘date created’, ‘date closed’, ‘location type’, ‘city’,

and ‘borough’. This left us with the 4 columns needed to do the analysis; ‘zip code’, ‘latitude’, ‘longitude’, and ‘noise count’. We kept a copy of this dataset with the individual occurrences of noise complaints to be used for our interactive map coming later.

We then summarized this dataset by grouping the occurrences by zip code, and summing the total number noise complaints by using the mutate() function. This gave us the total number of noise complaints by zip code which gave us a good picture of which neighborhoods are noisy and which are quieter (See Appendix 1). We then removed the duplicates to have the unique postal codes and their respective noise complaints.

The next step was to merge the two datasets. We merged the datasets on the zip code columns using a left-outer join. This kept all the rows in the Restaurant Week dataset and kept only those in the Parties NYC dataset that matched. We had some outliers in this merged dataset, so we removed the observation of a restaurant without a zip code. Some zip codes had no noise complaints therefore they showed up as NA for the count for some restaurants. To solve this, for the NA values, we replaced them by 0 because essentially, they received 0 noise complaints.

3. Model Selection & Methodology

As this study primarily focuses on providing further insight to diners and helping them narrow down their decisions for selecting restaurants, there was a heavy focus on three aspects for the decision-making process;

1. Classification trees
2. Principal Component Analysis
3. Clustering

3.1 Classification trees

The thought process behind classification trees was to create a map of how restaurants can be classified, and the Random Forest algorithm provided the best visualization to determine the type of restaurants that are available based on factors such as;

- food
- ambiance

- value for money
- price range

This provided a strong statistical argument as to what a diner can expect based on their selection of cuisine. For example, you can assume from the start that if you decide to go to a New York Steakhouse you should expect to pay approximately $50 or more per person. However, if someone is looking for a dining experience where you get a higher value for money, they would be better off looking choosing a restaurant oriented towards American cuisine. (See Appendix 2)

3.2 Principal Component Analysis (PCA)

PCA provided clarity on the relationship between variables, especially when it came to factors that could affect the Average Rating of a restaurant. At an initial glance some of the relationships made sense intuitively, for instance the quality of food had a very strong relationship with how the Average review would be, while factors such as a 1-star, or 2-star rating negatively impacted the average review. However, factors such as value for money, and service rating had a strong collinearity which shows that there is a possibility that cheaper restaurants may have a lower service rating, but they both had a positive impact on the Average rating of a restaurant.

Moreover, the PCA plot showed that noise levels in the corresponding neighborhoods did not provide much of an effect on the rating of a restaurant, as it shows a perpendicular relationship with the Average Reviews. This makes it clear that dining options can be selected irrespective of the noise level of the neighborhood, but the noisy neighborhoods will provide more options for diners looking to enjoy a drink or two after their meal at another establishment such as a bar or nightclub. (See Appendix 3)

3.3 Clustering

For clustering we used multiple methods to analyze various aspects of the restaurants in Manhattan, and one of the most striking results was based on clustering restaurants based on the zip code and average rating. The K-Means algorithm indicated that while most of the restaurants active during Restaurant Week in Manhattan are closely rated, there were more restaurants that were rated higher than zip codes from 10010 to 10020.

Conversely, as the zip codes increased or the more diners looked at options further uptown, the ratings for restaurants were lower, while restaurants in zip codes closer to 10050 had particularly lower ratings. Therefore, based on this clustering method we can advise diners to stay in the Lower Manhattan to Midtown areas during restaurant week if they were more concerned with the ratings of a restaurant. (See Appendix 4)

4. Results

4.1 Our Model

By observing the results from the Principal Component Analysis, it is evident that the noise count has little to no effect on the overall rating of the restaurant. This means that the noisiness is purely preferential to diners. They may want to be in a noisy neighborhood and have a "good time" or may want to be in a quiet neighborhood for a relaxing evening. Therefore, noise is not a factor in having a good dining experience.

To display the results, the 'leaflet' package provided a platform to plot an interactive map of Manhattan (See Appendix 5) with the Longitude and Latitude from the noise complaints to plot individual complaints. Since there were so many noise complaints, it made sense to cluster these using the 'markerClusterOptions' method. This displayed colored clusters based on the noise complain count as well as the actual instance count itself.

The result of the model is a highly interactive map of Manhattan containing partitioning clusters of noise complaints based on the "zoom" of the user (See Appendix 6). By looking

at this map, users can choose which area they may want to avoid or go towards for their perfect night out. In the map, there are markers which indicate the number of noise complaints within a clustered area. When a user hovers the mouse over the marker, it will show the border of area in which the noise complaints will be heard (See Appendix 7). When you zoom in or select a specific cluster, the clusters will then partition into more clusters giving a more detailed view. (See Appendix 8)

Another important feature we wanted to include in the model was to be able to select restaurants from Restaurant Week based on user preferences. (See Appendix 9)

Another interactive map was added that lets users choose the desired price range, the restaurant type and a noisy neighborhood as filter. This map will then display all the restaurants in the NYC restaurant week onto the map that the user desires. This map is interactive in the sense where you can zoom in and out to see the precise locations and names of the restaurants that the user has picked according to their criteria. (See Appendix 10)

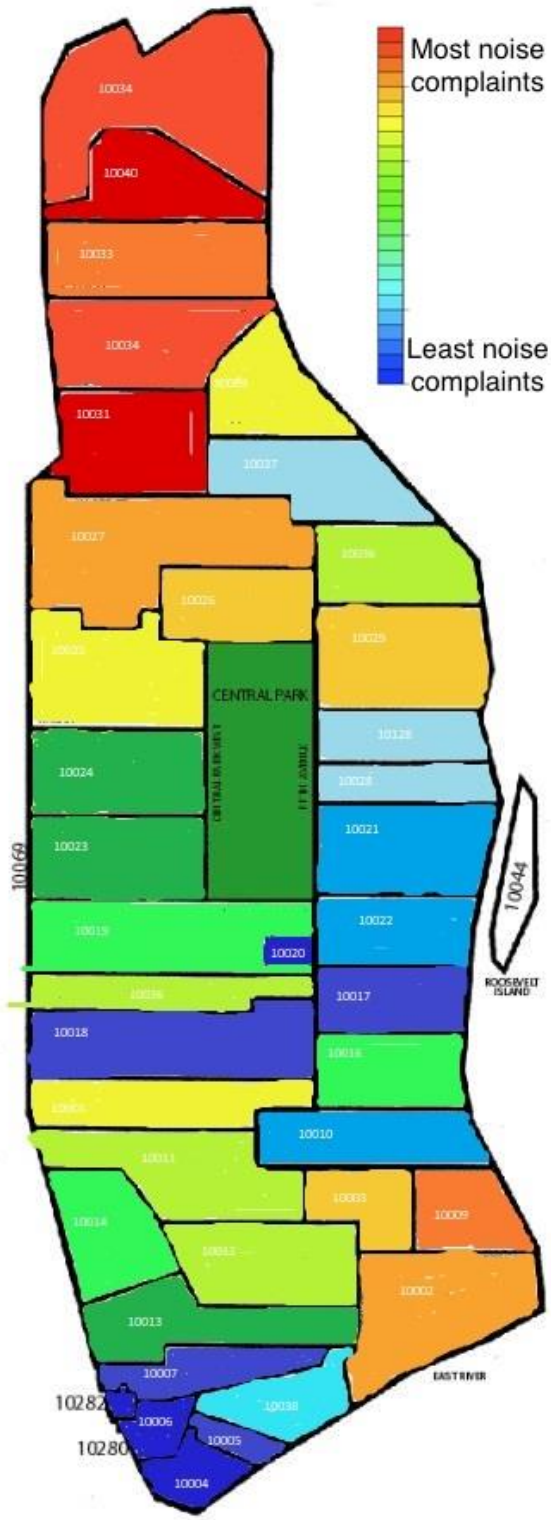
4.2 Conclusion

While NYC's restaurant week allows diners to eat at some of the best eateries in the city, most people are often just intimated by the vast amount of choices. While the options for dining are often considered good, given the timeframe this event runs for it would be ideal to optimize the experience for those hoping to make the most of this week. Therefore, we hope this will help diners find the best places to eat, based on their preferences from cuisine, price, location and their post dinner plans.

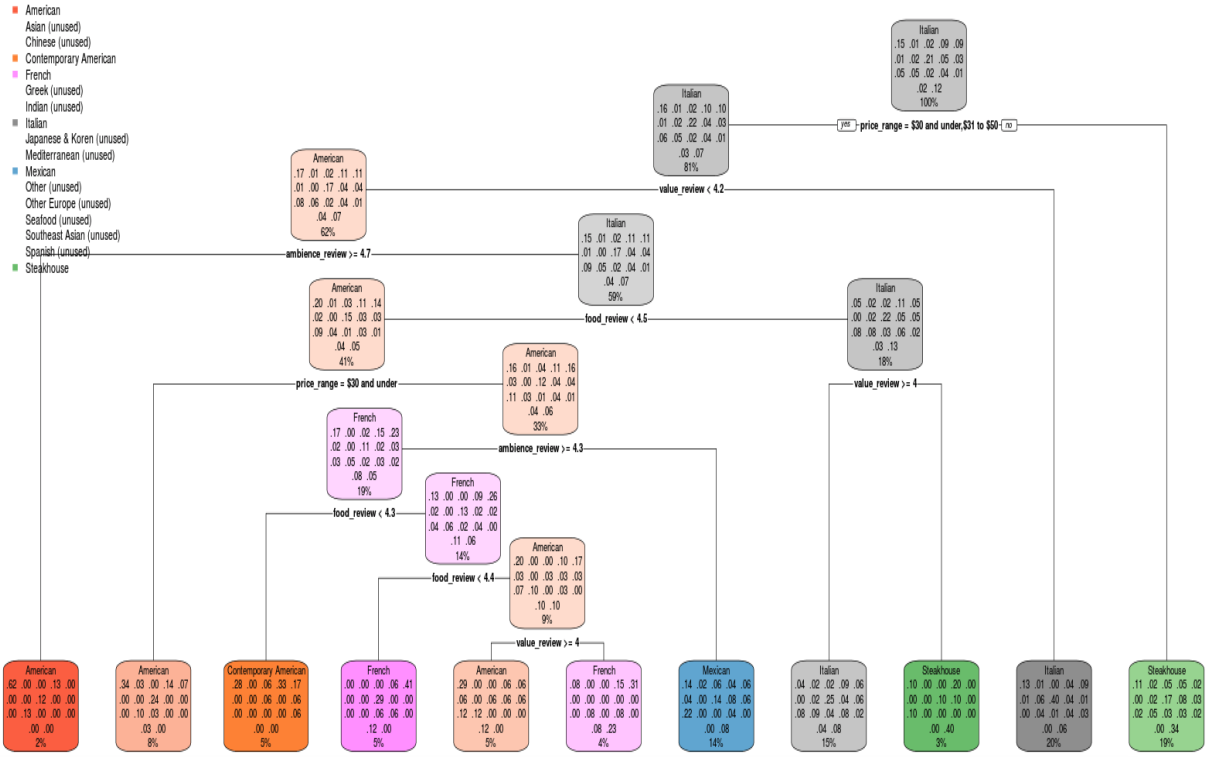
As an extension to this project, users will be able to access an interactive map of Manhattan which contains multiple layers in-built to help make informed decisions. New Yorkers and visitors alike can feel at ease stepping out their; Ubers, Yellow Cabs or Subway line, knowing they will have one of the best curated dining experiences for NYC Restaurant week. Bon Appetite!

5. Appendices

Appendix 1: Number of noise complaints by zip-code

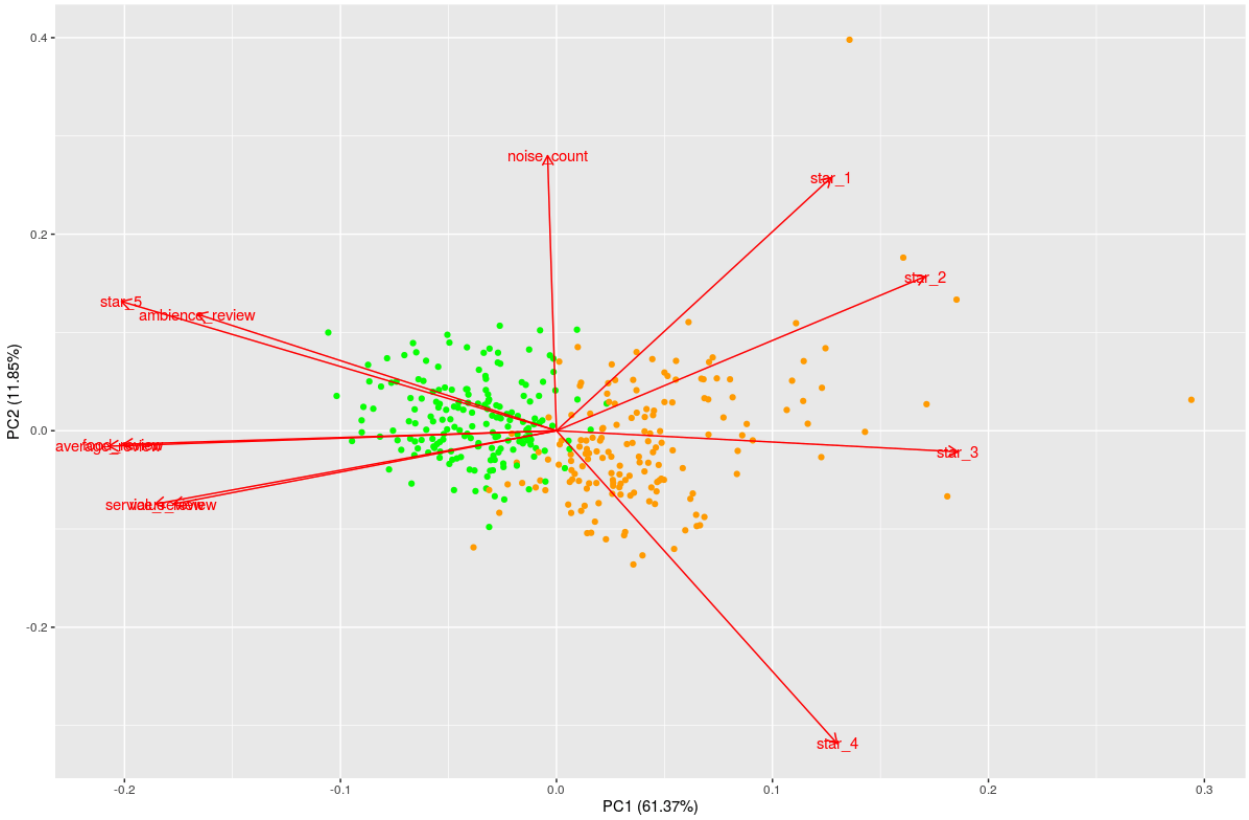


Appendix 2: Classification Tree



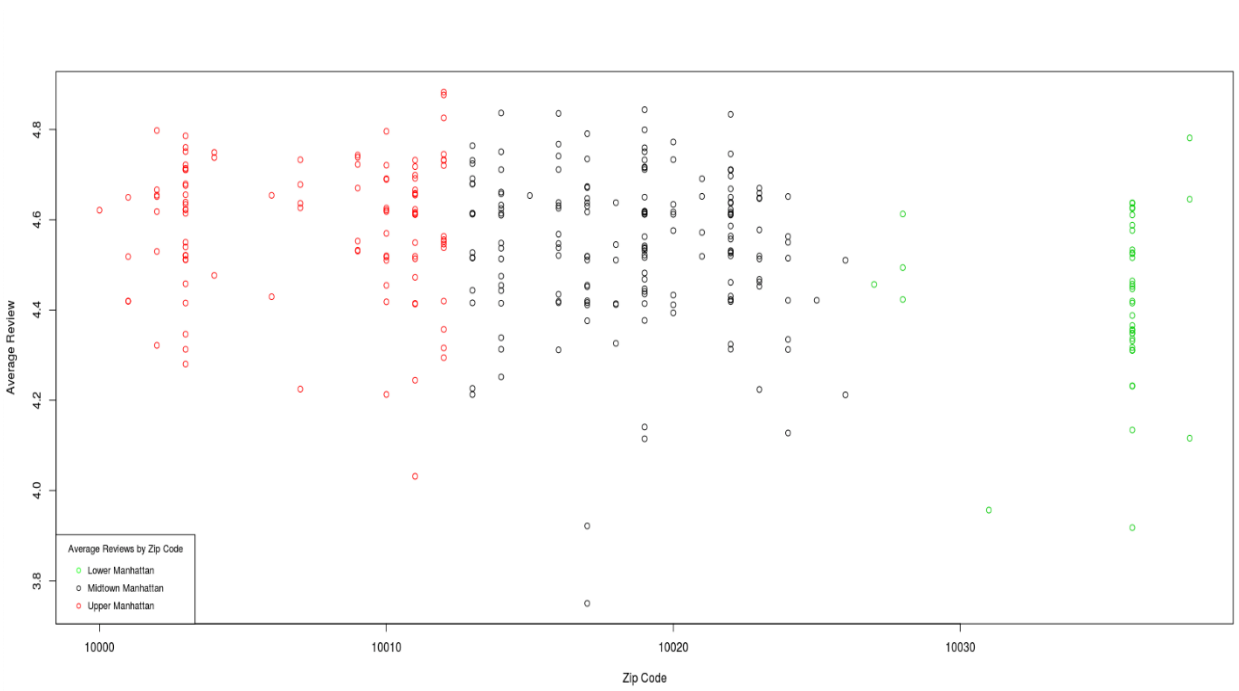
Appendix 3: Principal Component Analysis

Green points signify restaurants that have a rating higher than average, orange points signify restaurants that have a rating lower than average.

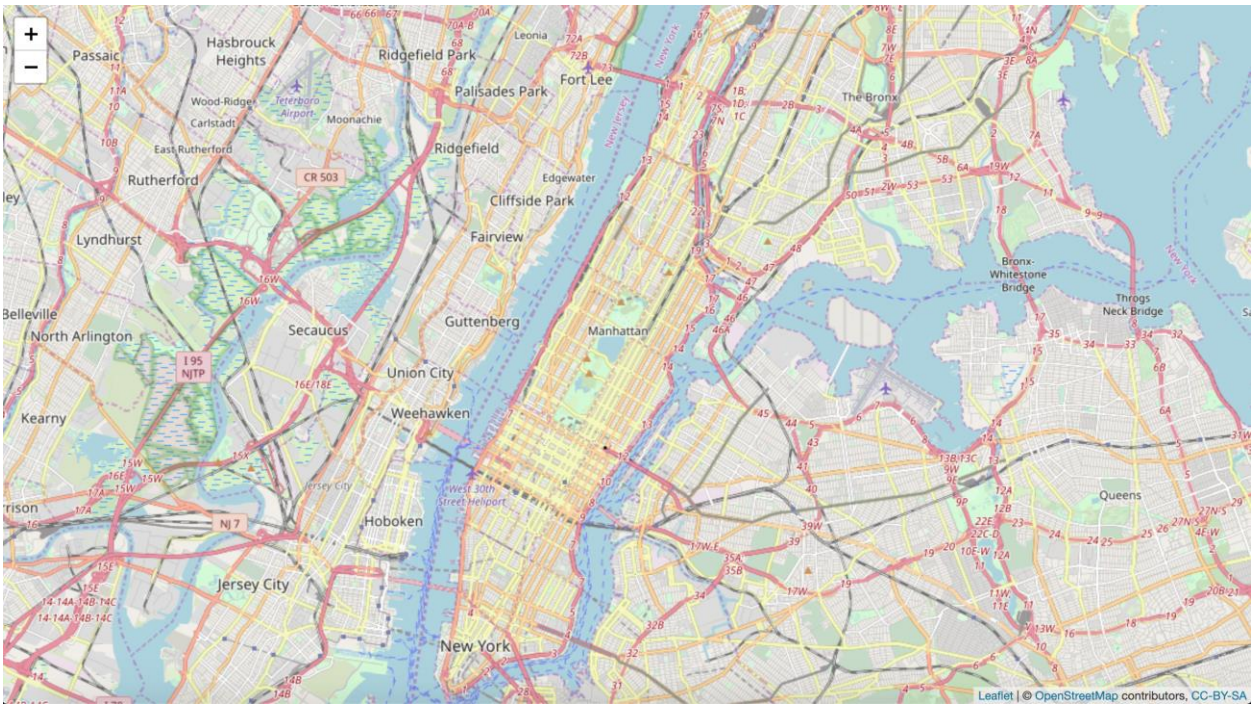


	PC1
average_review	-0.369770802
food_review	-0.357892938
service_review	-0.331532592
ambience_review	-0.296426751
value_review	-0.316341004
star_1	0.227031194
star_2	0.304650109
star_3	0.331688790
star_4	0.231834632
star_5	-0.359380832
noise_count	-0.007328349

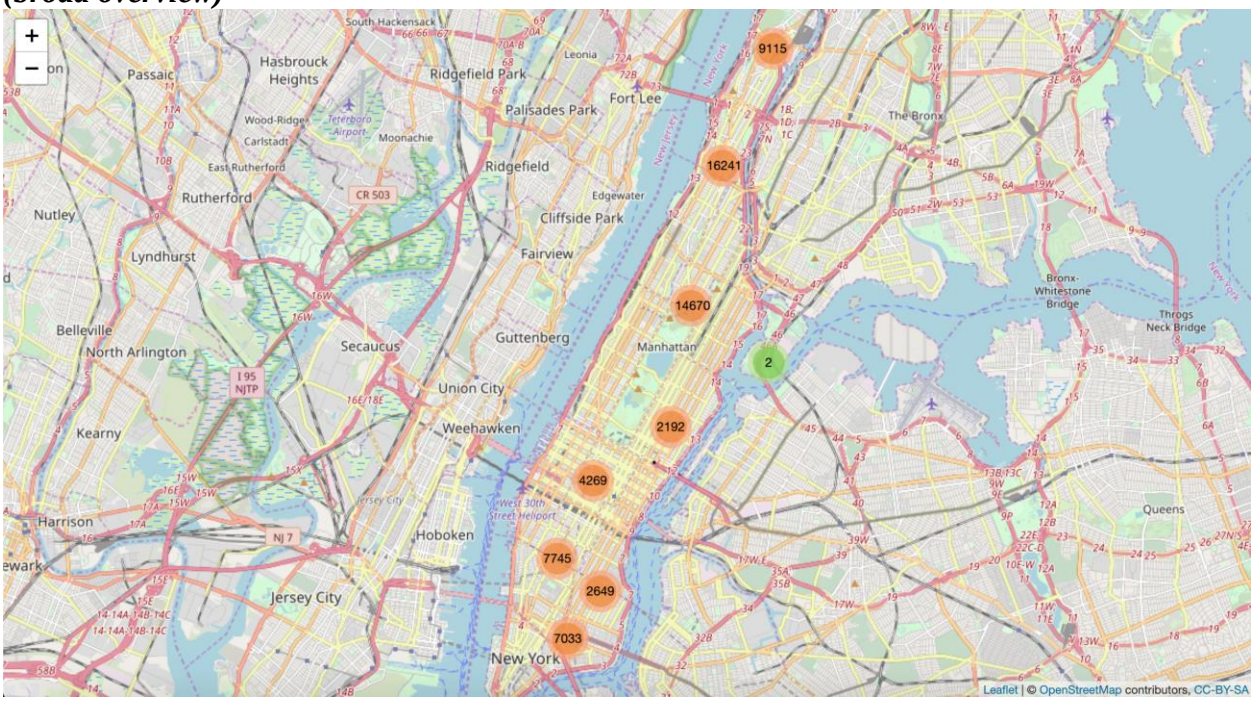
Appendix 4: Clustering



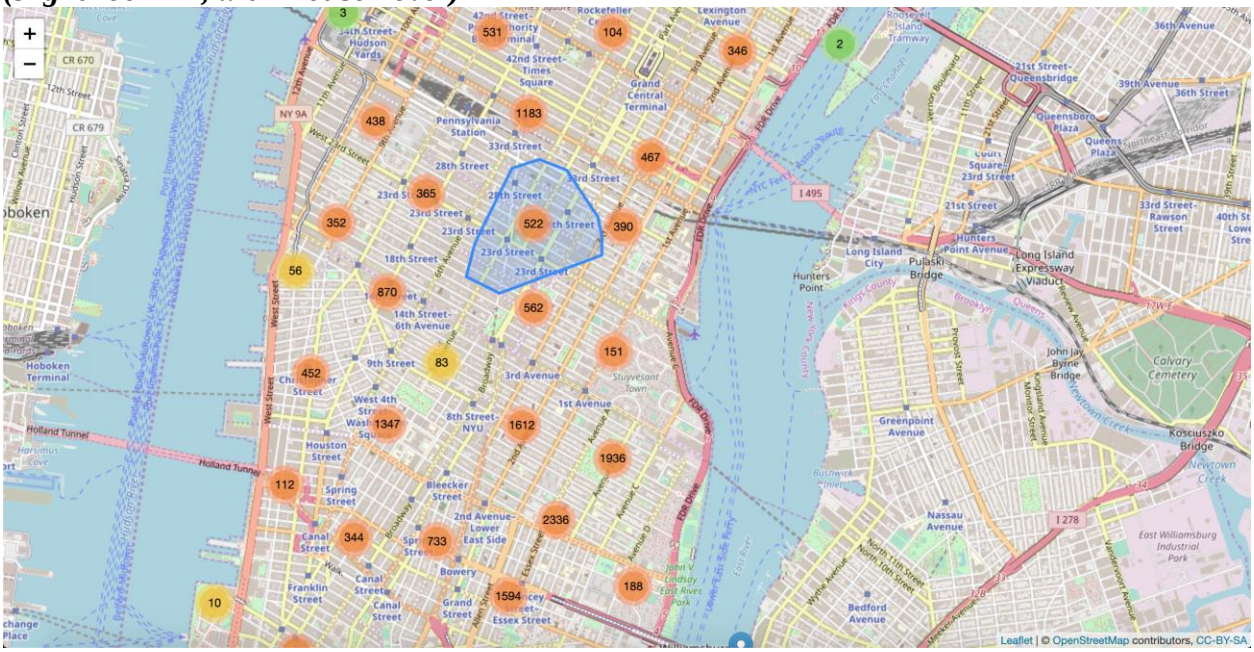
Appendix 5: Simple interactive map of Manhattan



Appendix 6: Map of Manhattan with number of noise complaints ordered by clusters (broad overview)

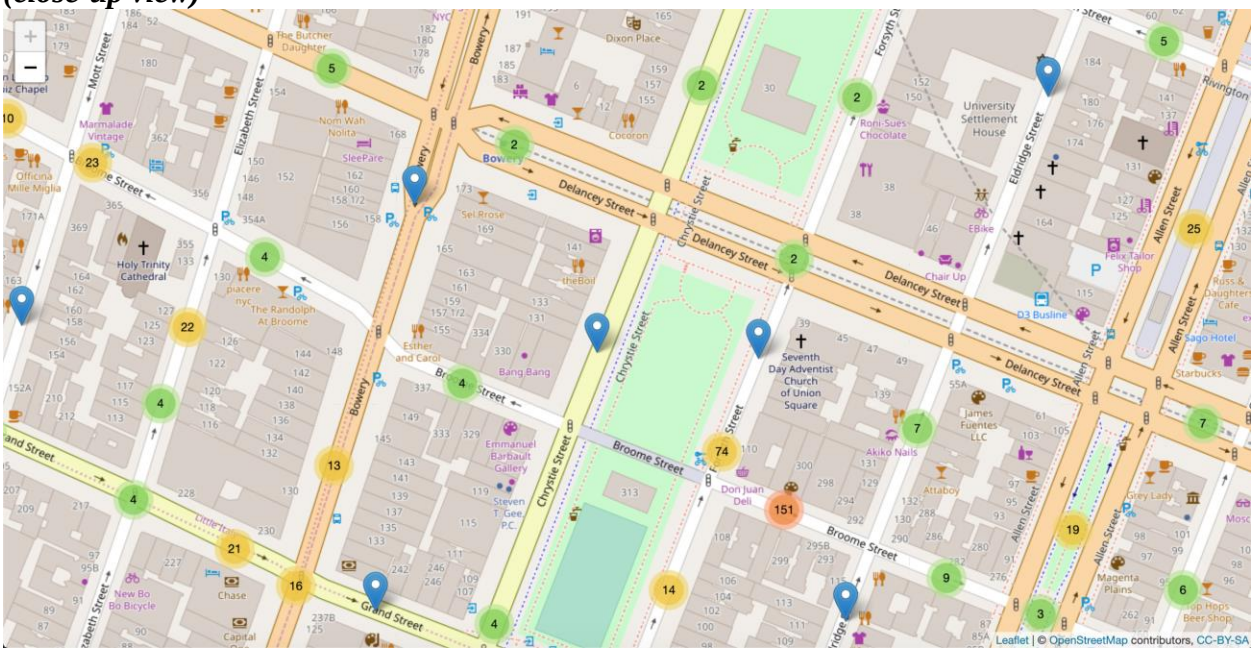


Appendix 7: Map of Manhattan with number of noise complaints ordered by clusters (slight zoom-in, with mouse hover)



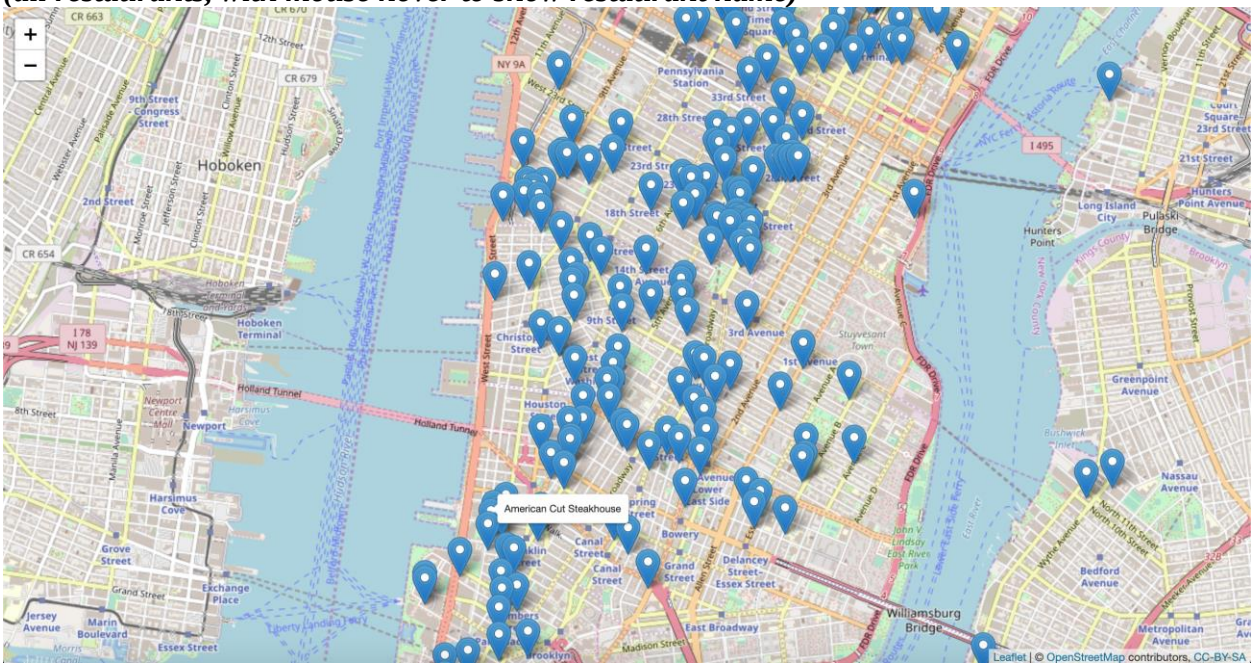
Appendix 8: Map of Manhattan with number of noise complaints ordered by clusters

(close-up view)



Appendix 9: Restaurant Week locations

(all restaurants, with mouse hover to show restaurant name)



Appendix 10: Restaurant Week locations

(Restaurants with filters applied; '$50 and over', 'noisy', 'Steakhouse')