

# The Washingtonian Top 100 Restaurants

DNSC 6211: Programming for Analytics

Lawrence Gadsden

Eugene Hwang

Steffanie Ohms

Ramsin Toma

## Abstract

For this analysis, we decided to take a deeper look into the Washingtonian's list of the Top 100 Restaurants and see how they rate on Yelp. As the Washingtonian ratings are based on professional reviewers, we thought it would be interesting to see if the ratings from the general public agreed. Besides ratings, we also analyzed the price levels and locations of these top 100 restaurants to get a better sense of their composition. Through the use of visuals such as histograms and maps, we found that most of the restaurants had top ratings on Yelp, were priced in the lowest price tier, and were located in DC throughout the many popular dining neighborhoods.

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Background</b>	<b>3</b>
<b>3</b>	<b>Method</b>	<b>3</b>
<b>4</b>	<b>Organization</b>	<b>4</b>
4.1	Workflow . . . . .	4
4.2	Project structure . . . . .	5
4.3	Figures and Tables . . . . .	5
<b>5</b>	<b>Discussion</b>	<b>7</b>
5.1	Learnings . . . . .	8
5.2	Challenges . . . . .	8
<b>6</b>	<b>Conclusion</b>	<b>8</b>

# 1 Introduction

The topic that we investigated in our study was the relationship between the list of restaurants in Washingtonian Magazine’s “Top 100 Very Best Restaurants”, and the relevant Yelp information for that same list. The reason we chose this topic is two-fold. First, we wanted to compare the opinions of the Washingtonian “experts” with those of the “common people” from Yelp. Second, we chose this topic for slightly personal reasons as well; everyone loves food, and we all want the best value and experience when we go out to eat.

The concept of comparing and contrasting the opinions of experts and regular reviewers is not limited solely to food or restaurants. Whether discussing restaurants, movies, music, etc the expert opinion does not always match public opinion. Nickelback has been roundly panned by music experts, but they continue to sell millions of records and sell out shows around the world. Birdman won the Best Picture Oscar in 2015, but it receives only a 78% general audience score on Rotten Tomatoes, which seems far from a “best movie” score.

On a more personal level, we want to get the best value for our money when we go out to eat with family, friends, or loved ones. It could be argued that this is the very reason lists like the one in the Washingtonian even exist. We want the best experience, but should we trust the opinions of total strangers? Ultimately, that is the question we want to answer with our study.

# 2 Background

The Washingtonian publishes its 100 Very Best Restaurants, yearly. The list is extremely popular. On the other hand, Yelp is the leader in crowdsourced reviews about local businesses. We chose this data because we thought that it would be interesting to see how normal customers felt about the Washingtonian's picks, while also finding location and pricing patterns.

Yelp has ratings for more than 12,000 restaurants in DC. The Washingtonian's list allowed us to focus on a more select group. Some questions that we started with initially were, what categories do most of these restaurants belong to? What are the most popular neighborhoods? Will there be clusters of restaurants or will they mostly be spread out? Do the restaurants have similar price ranges? And, do Yelp reviewers rate these restaurants highly?

We finally got to the questions that we answered by visualizing them through plots and figures. We were able to construct plots for pricing, ratings, and locations, and we were able to build a map so that we can see their actual locations.

# 3 Method

The overall question that we set out to answer was “What do the Washingtonian Top 100 Restaurants look like?” Given the data that we had available to us, we were able to answer this question by analyzing the restaurant distribution by Yelp rating, price level, and location. We were interested in seeing how the restaurants varied by these attributes, if at all. Were these restaurants all highly rated on Yelp or did some poorly rated restaurants on Yelp somehow make the list? Were the top restaurants all expensive ones that most people would not be able to afford? Lastly, where are these restaurants located and do they tend to be located in one area more than another?

In addition the questions of interest mentioned above, we also wondered about the correlation amongst these attributes. Did higher prices usually lead to higher ratings? Did location have an effect on higher ratings? However, being that we are only analyzing the top 100 restaurants from the Washingtonian, we felt that answering these questions would not produce a fair

analysis. This subset of restaurants is highly biased since they are already the best by the Washingtonians standards. To perform such a correlation analysis, it would be better to have a sample of restaurants with more variation in the data to determine any true relationships (that is, restaurants with both high and low ratings). For this reason, we decided not to pursue the correlation aspect of our analysis any further.

## 4 Organization

To complete this analysis, we divided up the tasks within our group as follows:

- Lawrence and Ramsin collaborated to scrape the list of top restaurants from the Washingtonian website and retrieve the corresponding Yelp data for each of those restaurants from the Yelp API.
- Eugene developed the map of the restaurants and produced the display of results in an RShiny application.
- Steffanie created the histogram plots of the data and took the lead role in writing the report.

Despite these designations of tasks, all members of the group worked together and helped each other when assistance was needed.

### 4.1 Workflow

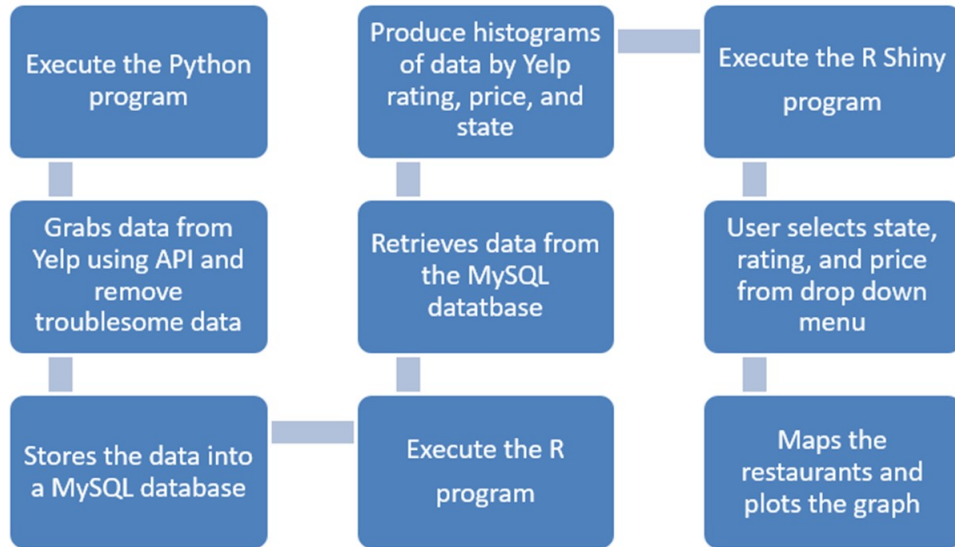


Figure 1: The project workflow

To begin this analysis, we started in Python to first obtain the list of the Washingtonian Top 100 Restaurants. We did so through web scraping of the Washingtonian site, and then using that list, we utilized the Yelp API to find the data corresponding to those restaurants.

We took the restaurant name and price level data from the Washingtonian website, and took the rating and location data from the Yelp API. Once we had all necessary data, we stored it in a MySQL database for later reference.

Moving out of Python, we performed the analytical portion of our project in R. In the R file titled “GroupProjectPlots.R”, we produced multiple histograms to understand the distribution of top restaurants by Yelp rating, price level, and state of location. To better visualize our restaurant data by location, we developed a map and placed it within an RShiny application to analyze location with more flexibility. Through the use of histograms and the map, we were able to develop the findings of our analysis.

## 4.2 Project structure

We used the Washingtonian's 100 Very Best Restaurants list and Yelp for our data sources. They are related to each other because both feature ratings and reviews for restaurants in the DC area. However, while the Washingtonian staff ranks the best restaurants, Yelp allows ordinary customers to submit ratings. Furthermore, Yelp has ratings for most of the restaurants in DC while, the Washingtonian rates 100. We used the Washingtonian to submit the Yelp restaurants to only those restaurants recognized on the Washingtonians list.

The ratings, location information, and category information from Yelp allow us to explore the restaurants recognized by the Washingtonian so that we can find similarities and patterns.

## 4.3 Figures and Tables

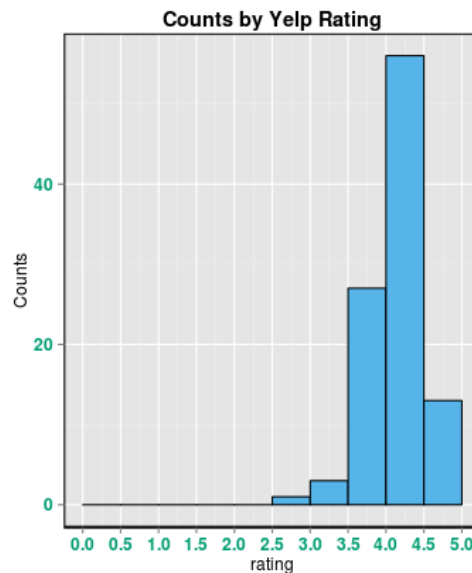


Figure 2: Restaurant Distribution by Yelp Rating

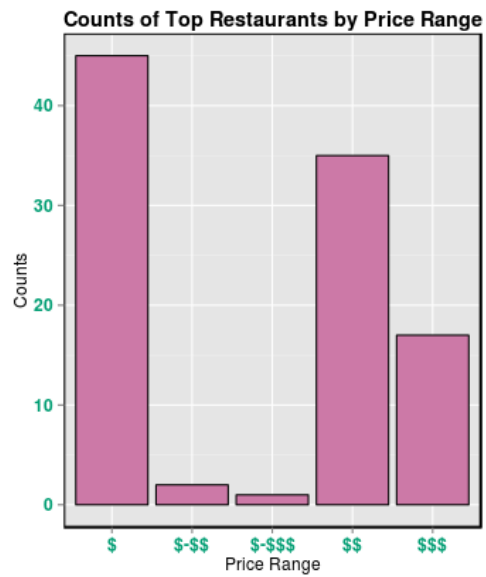


Figure 3: Restaurant Distribution by Price Range

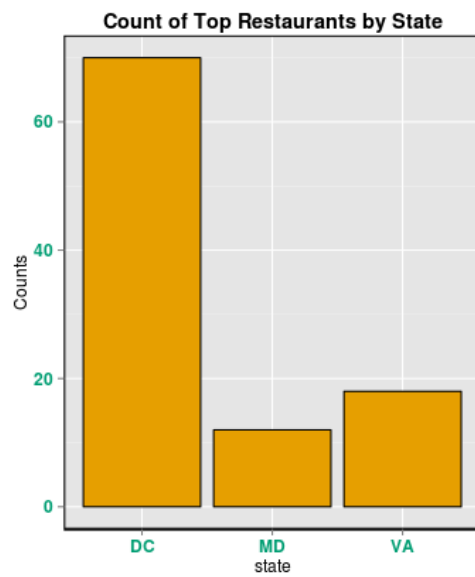


Figure 4: Restaurant Distribution by State

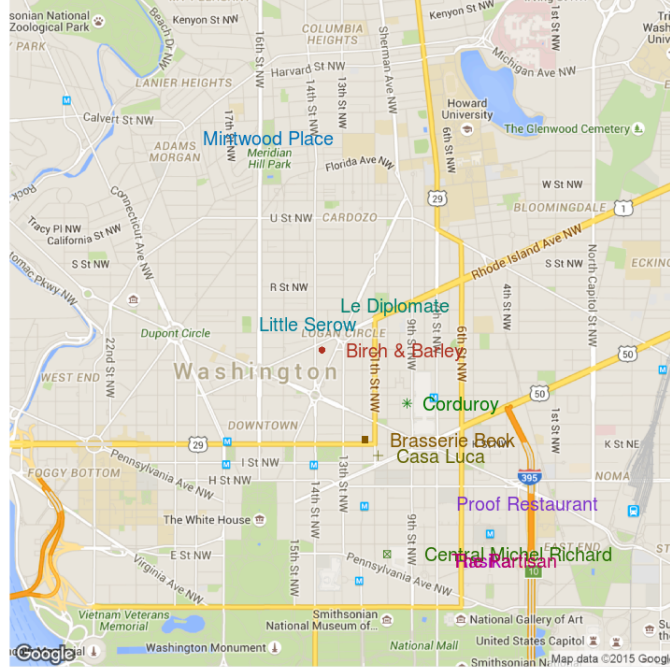


Figure 5: Map of DC Restaurants

Since the goal of our analysis was to see what the Washington Top 100 Restaurants look like by key attributes, we used histograms and maps for their ease of summarizing and visualizing the data. Figures 2, 3, and 4 are histograms that show the distribution of the restaurants by Yelp rating, price range, and state of location, respectively. These histograms show the makeup of the Washingtonians list and reveal any tendencies that it may have in relation to these three attributes. By looking at these histograms, we are able to answer most of the main exploratory questions that we had for this analysis.

The one question that the histograms are unable to answer is whether or not there are location patterns for the top restaurants. The histogram by state shows location only at a very broad level. By seeing the restaurants displayed on a map as in Figure 5, we are able to analyze the data by neighborhood and detect if there is an existence of clustering patterns. As a result, this part of the analysis explains if there is any relationship between the “top restaurant designation and where exactly the restaurant is located.

## 5 Discussion

This analysis is of value for its ability to provide further validation and details about the restaurants that are critically-acclaimed by the Washingtonian. As the Washingtonian is only one opinion (and a professional one in that), it is beneficial to have the added layer of credibility from the reviewers on Yelp. The Washingtonian list combined with the ratings from Yelp can offer a person a better idea of the overall quality of a given restaurant.

Another benefit of this analysis is the factual information about the restaurant list that it produces. As the Washingtonian Top 100 Restaurants list is considered an elite group, one

may have potentially false preconceived notions about those that make the list. The results of this analysis may shed new light on a restaurant or the list as a whole, and alter a persons consideration in both possible directions.

## 5.1 Learnings

The main learnings from our analysis of the Washingtonian Top 100 Restaurants are as follows:

- **Rating:** Most restaurants do indeed have a top rating on Yelp but this is not true for all. About 70 of the restaurants have a Yelp rating of 4 or above, leaving the remaining 30 restaurants to have mediocre ratings of 2.5-3.5.
- **Price:** Many of the restaurants are classified in the more affordable price ranges by the Washingtonian. Nearly half of the restaurants are in the lowest price tier, while only 17 restaurants on the list are in the highest price tier.
- **Location by State:** Majority of the top restaurants are located in DC, with only 30 of the restaurants located in MD and VA combined.
- **Location by Neighborhood:** The top restaurants are located throughout DC, MD, and VA without any strong clustering patterns in one neighborhood more than another. Rather, they seem to be distributed similarly to the distribution of restaurants by neighborhood overall.

## 5.2 Challenges

As mentioned earlier in the “Methods” section, a challenge with this analysis was making it more robust. Given that we were limited to a subset of already well-rated restaurants, we were unable to gleam more insights about the relationships between the restaurant attributes themselves.

On a more technical note, we faced some execution challenges throughout our analysis. The first was regarding the restaurant names and obtaining the necessary data from the Washingtonian and Yelp. There were issues with some of the special accent characters of the restaurant names in Unicode and also issues with some inconsistencies between the naming conventions for the same restaurant between the two sources. It was also difficult when a restaurant of the same title had multiple locations in the DC/MD/VA area.

## 6 Conclusion

Over the course of our project, our team worked to compare the list of The Washingtonian's “Very Best 100 Restaurants” with the data for those restaurants on Yelp. The data we collected included comparison metrics, such as price ranges and ratings, and identification metrics, such as address, neighborhood, or phone number. Over the course of our project, we used various tools in Python, MySQL, and R to get our data, clean it, store it, and describe it. What we found out about the “Top 100 Restaurants was a little surprising to us. While very best might seem synonymous with “most expensive, that is not the case as the majority of the restaurants were actually the least expensive. In addition, we found that the Yelp ratings were not quite as favorable as the expert opinions. This seems to be another case where the experts and the general public seem to have different tastes.