Visualizing Factors that Make Bars, Restaurants, and Hotels Popular

Chelsey Legacy, Lindong Zhou, Evan Pete Walsh*

December 1, 2014

Abstract

Blah blah blah

*Statistics graduate students at Iowa State University of Science and Technology

Contents

1 Introduction

Lindong

2 Data

The Yelp Academic Dataset[1] provides data enthusiants with the exciting opportunity to explore an incredible collection of information regarding the characteristics and quality of hundreds of businesses across the United States, Canada, and the UK. Specifically, the data includes details and reviews on 250 of the closest businesses to 30 large universities, including the Arizona State, UNLV, the University of Edinburgh, the University of Wisconsin, and the University of Waterloo, to name a few. The raw data is in json format and contains five different types of json objects: **Business**, **Review**, **User**, **Check-in**, and **Tip**.

Each **Review** object represents an individual user-based review of a particular business. The unique encrypted business ID is given along with the date of the review, the number of stars (out of 5) that were awarded, the number and type of votes that the review received, and an optional text description provided by the user. **User** objects are unique to every person that has an active Yelp account. Each user has a name, a unique encrypted user ID, the number of votes they have cast, the average number of stars they have given, and the date they signed up for Yelp, among other things. A **Check-in** object represents the count and time of all the registered check-ins for a particular business, and **Tip** objects represent a tip given by user for a particular business. Tips include the user's ID, the business's ID, the date, and the message that the user gave. While these objects all provide a rich source of information, for the scope of this paper we will only be examing the **Business** objects.

Business objects are unique to business ID's, and include the following information:

- the name of the business,
- the name of neighborhood in which the business is located,
- the city in which the business is located,
- the full address of the business,
- the exact latidude and longitude coordinates of the business,
- the average number of stars awarded to the business,
- the number of reviews received by the business,
- whether or not the business is still open,
- the hours that the business is open,
- the categories that the business falls under,

---

[1]$https://www.yelp.com/academic\_dataset$

- and a number of different attributes which mostly concern restaurants and bars, such as whether or not smoking is allowed and the price range of the food.

To work with the data, we converted the set of all **Business** objects to a csv file in which the columns are variable names representing each aspect of a **Business** object, and each row corresponds to a unique business. Each different type of attribute was converted to its own character, numerical, or logical variable depending on what was appropriate. For example, the attribute **smoking** was converted to a logical variable where the value is "true" if the smoking is allowed at the establishment, and "false" if it is not allowed, while the attribute **price range** was converted to a numerical variable which ranges from 1 to 4.

3 Restaurants

Most of the data about restaurants in the Yelp Academic Dataset are concentrated around the following cities:

- Madison, WI,
- Las Vegas, NV,
- Pheonix, AZ,
- Waterloo, Ontario, Canada,
- and Edinburgh, Scotland.

Although the choice of cities is limited, the wealth of information coming from each city is abundant. All together, our data consists of 14,303 restaurants with 45 variables describing their attributes and the collective user sentiment towards each establishment. There is a significant number of restaurants in just about any subcategory that one can imagine. For example, Figure 1a shows all of the restaurants classified as either "hipster" or "divey" near Pheonix, AZ that have televisions, offer live music, and have more than 35 reviews on Yelp. Figure 1b shows all of the Asian and Indian food restaurants near Las Vegas, NV that offer take-out and have a full bar.

3.1 The Cost of Reservations

One question we have explored with the data is whether restaurants that take reservations are more expensive and of higher quality than those that do not. Not suprisingly, it appears that one has to pay a premium, on average, to eat out at a restaurant that accepts reservations. Figure 2 shows the average price level of restaurants by type that do or do not accept reservations. For each category, the price level is higher on average for restaurants that accept reservations.

There are a couple potential explanations for this. For one, it could just be that fancier restaurants feel they need to accept reservations to attract their target demographic. Perhaps people who are willing to spend a large amount on a meal will only do so if they know they

Figure 1: Restaurant subsets near Pheonix, AZ and Las Vegas, NV.
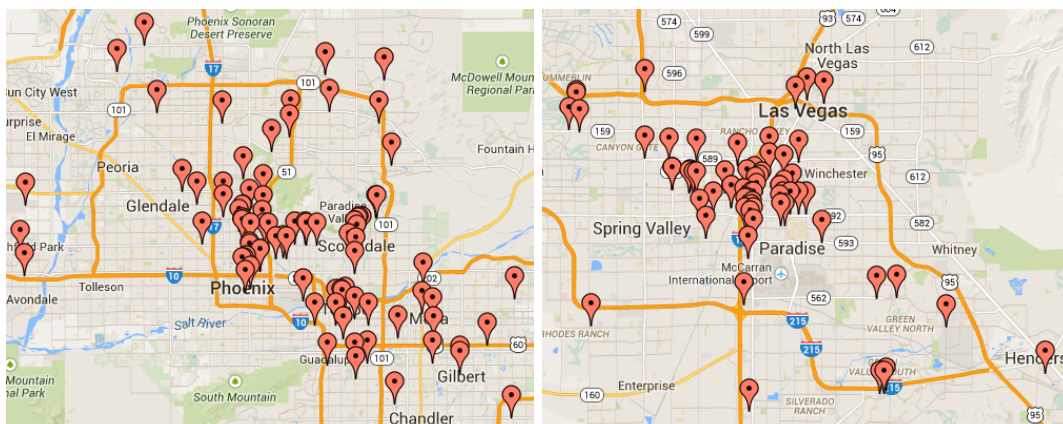
(a) All restaurants near Pheonix, AZ classified as "hipster" or "divey" that have live music, TV's, and more than 35 reviews on Yelp.

(b) All Asian and Indian food restaurants near Las Vegas, NV that offer take-out and have a full bar.

will be seated right when they arrive. Or maybe restaurants that take reservations lose some money in transition time when open tables are kept unoccupied until the party that made the reservation arrives. The higher average price of restaurants that take reservations could just be making up for this inefficiency.

Figure 2: Price scale of restaurants by type and whether they accept reservations.
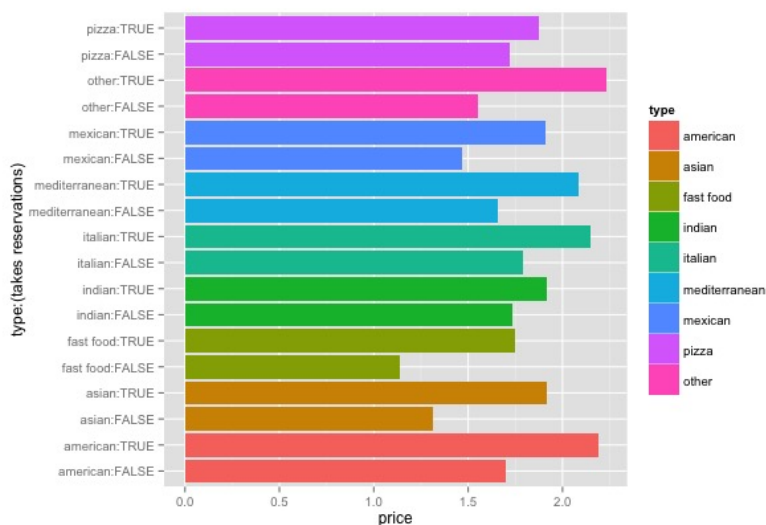
Although a premium is payed to eat at restaurants that accept reservations, we find that the quality of the experience does not necessarily increase, as measured by the average stars awarded by Yelp users. Figure 3 compares the average stars received by each category by whether or not reservations are accepted. While the average number of stars is slightly higher for pizza, fast food, Asian, American, and "other" style restaurants that take reservations,

the average star count is actually lower for Mexican, Mediterranean, Italian, and Indian style restaurants.

Figure 3: Rating of restaurants by type and whether they accept reservations.
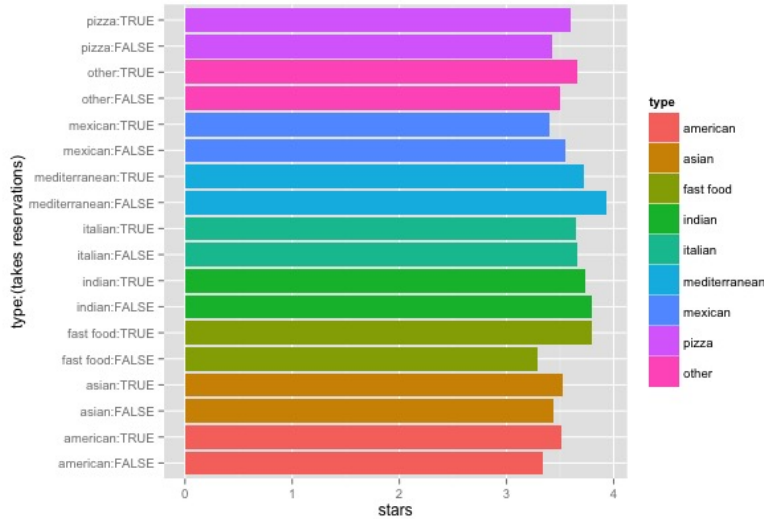
The reason for this discrepancy is unclear. Maybe popular restaurants in certain categories just don't accept reservation because it would be too inefficient, and customers are fine with that. Price may also have something to do with it. Figure 4 shows the rating of restaurants by type, whether they accept reservations, and price level. For Mexican restaurants, the average rating of low priced establishments is indistinguishable between those that accept reservations and those that do not. Yet for those in the third price level, the rating is clearly higher for those that do not take reservations. The same is true for Asian style restaurants while the opposite holds for Italian and American style restaurants. This could be because people who generally like to eat at pricier Italian and American food restaurants prefer to be able to make reservations, so these restaurants that allow reservations end up with higher ratings.

3.2 Loud can be Good, but not too Loud

Another question we sought to answer was how noise pollution affects the quality of a dining experience, especially when we consider whether or not the establishment offers live music. One can imagine that people deliberately going out to eat at a restaurant with a band playing might bear a higher noise level, and thus not be as quick to penalize the restaurant for being too noisy in their review. We find that the data is consistent with this hypothesis.

For restaurants that offer live music, as shown in Figure 5, loud places have a higher average rating than places with an average noise level. Yet very loud places have the lowest average rating. For restaurants that do not offer live music, the average rating decreases with the noise level.

Figure 4: Rating of restaurants by type, whether they accept reservations, and price level.
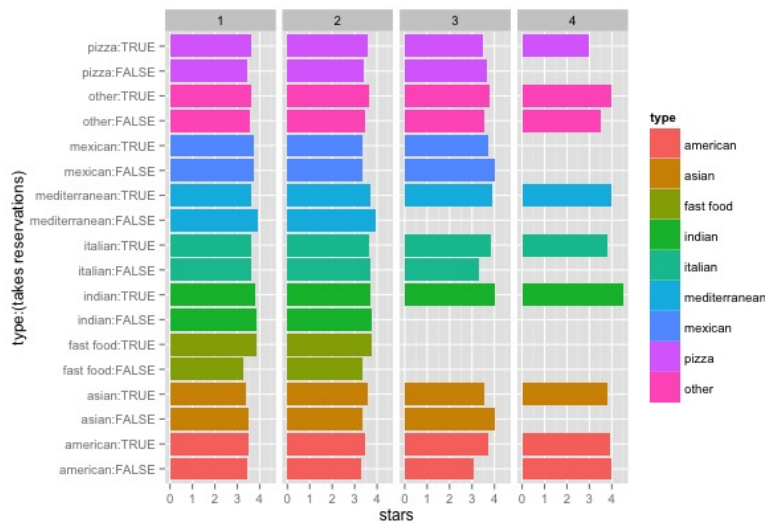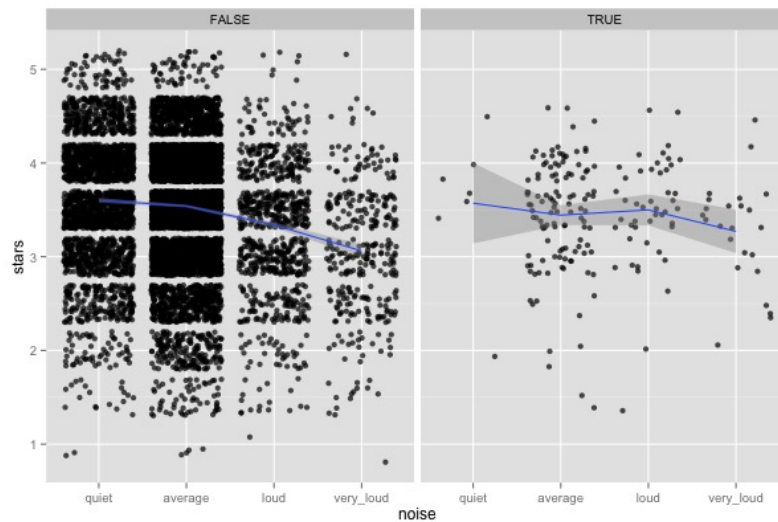

Figure 5: Rating of restaurants by their noise level and whether they host live music. The plot on the left shows the rating of restaurants by noise level that do not have live music, and the plot on the right shows the rating for restaurants with live music.

3.3   It's All About Service

Describe how the wordclouds were made: had to remove certain words

Figure 6: Wordclouds of tips for great restaurants and terrible restaurants.

(a) Tips for restaurants with atleast 4.5 stars.   (b) Tips for restaurants with less than 2 stars.
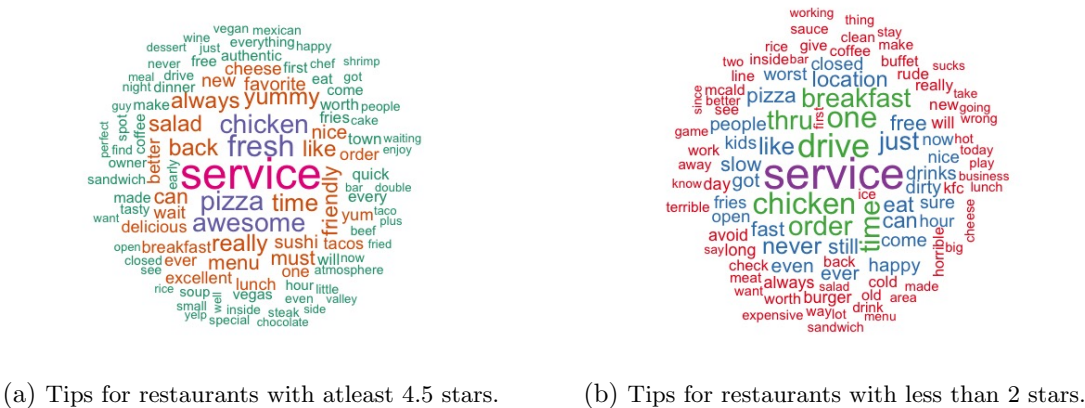
4   Bars

Chelsey

5   Hotels

Lindong

6   Fitness

Lindong

7   Conclusion

Chelsey