

The Dining Out Habits of Restaurant Reviewers in Madison, Wisconsin

By: Lina Sheremet, Ernesto Oropeza, Jaclyn Andrews

Repository Name: Project2_Sheremet_Oropeza_Andrews

Context:

The rise of restaurant reviewing interfaces and apps has allowed foodies to identify the best restaurants in their area, as well as other important aspects about potential restaurants that they plan to visit. Our project seeks to understand the dining behaviors of users of the app Yelp in Madison, Wisconsin.

Questions:

1. How has the Madison, Wisconsin restaurant scene changed over time?
2. Does weather influence the restaurant business in Madison?
3. Are there any restaurants in Madison that are outliers in terms of reviews?

Source Data and Scope:

The primary dataset we are analyzing was obtained from Kaggle and contains data on 192,609 businesses with 6,685,900 reviews from Yelp users spanning from 2005 to 2017. In order to keep our analyses more focused, we have decided to only analyze restaurants in Madison, Wisconsin. After this filtering, our dataset now contains information on 1,026 businesses with 59,930 total reviews. The dataset is split into a few groups of differing data, but we chose to focus on 2: businesses and reviews. Business data contains columns like the business id (unique for each restaurant), the location of the business (lat, lon), the average number of stars that business got, and some of its characteristics/categories. Review data is row-by-row reviews of each business on any given data by any given user. There is a rating, as well as a free text field. There were a couple other data sources such as check-in and business attributes, but we did not find them to be useful for our purposes, so we chose to focus on the two mentioned above. The data we used can be found [here](#).

In order to compliment our analyses of the businesses, we have decided to supplement with data on the weather conditions of Madison during the time period of these reviews. We obtained the weather data from the National Oceanic and Atmospheric Administration, and it has the weather conditions, as well as the minimum and maximum temperature for any given date. We are interested in whether certain weather conditions are correlated with what kind of food people eat and how they rate the restaurant, so we joined the weather data to our yelp data on date. We now have the exact weather conditions on any given day of a restaurant review. The data used can be found [here](#).

Data Cleaning:

In the Yelp business data set, the “categories” field contained at least one but usually many tags that indicate the type of restaurant. In order to make the category field more digestible, we created a cuisine grouping. If a business was tagged outright as a certain cuisine type like “Italian”, “Thai”, “Latin American”, etc. we used that tag to determine the cuisine. If there was no direct cuisine tagging, we used assumptions to group the cuisines. For example, we categorized restaurants tagged with “burgers” as American.

The Temperature data we used from the National Oceanic and Atmospheric Administration corresponds to the temperature forecast of several stations across Dane county in Wisconsin with several different temperature measurements. The most useful information for our purposes was average temperature, but the average was not reported through the complete period of the study. To overcome this lack of information we used the maximum and minimum temperature to calculate the average for dates that did not have an average. In the end, we only ended up using date and average temperature from this dataset for our analysis.

Severe weather conditions are reported in a separate database by the same agency. This data consists of the reported event (e.g. Hail, Strong Winds, Tornado) with some details such as starting time, initial location, final location and final time along with date and event name. Multiple events could have occurred on the same date. To build a unique database we grouped by date and counted the number of “severe weather” events for each date. When merging this data with the temperature data, we flagged “severe weather” as “True” if at least one event occurred on that date and “Normal” if no events occurred.

Exploratory Data Analysis:

There were 1,026 restaurants total in the Yelp dataset. The distribution of the average ratings for these restaurants can be found in *Figure 1* below. Most restaurants had 3.5 stars (263 restaurants) or 4 stars (302 restaurants). The distribution is skewed slightly right. Only 21 restaurants had an average of 5 stars, which makes sense, since even one bad review can skew your rating down. On the other hand, only 11 restaurants had 1 star.

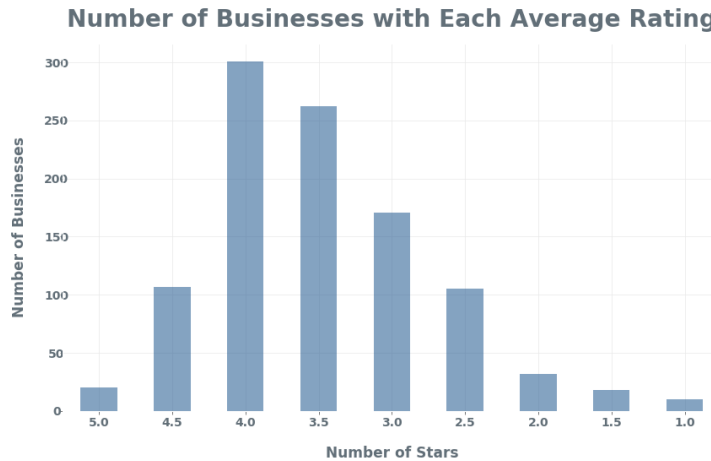


Figure 1: Distribution of average stars for restaurants in Madison

Figure 2 below displays the distribution for the number of reviews for businesses in Madison. Most restaurants in Wisconsin had more than 10 but less than 50 reviews. However, quite a large chunk of Madison restaurants had less than 10 reviews, which is surprising considering this is the total for a period of time greater than 10 years (2005 to 2017). The fewest restaurants had over 200 reviews (a little over 50). We discuss the outliers further later on this report.

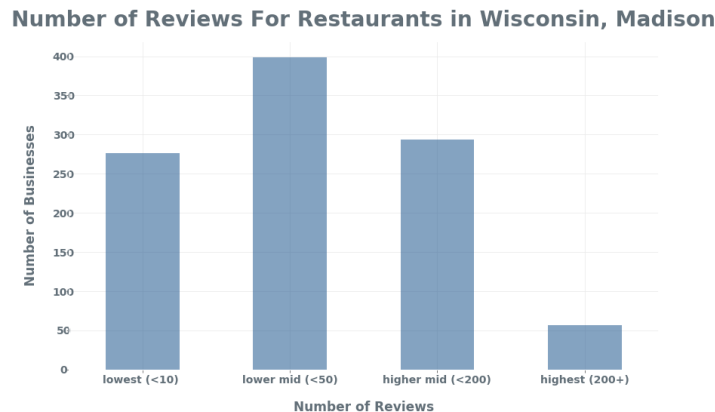


Figure 2: Distribution of number of reviews for restaurants in Madison

The heat map in Figure 3 displays where most restaurants are located in Madison, Wisconsin. Red and yellow areas have more restaurants in the area, whereas green and blue areas are less popular. The area between Lake Mendota and Lake Monona clearly has the most restaurants in Madison. This makes sense, since water would provide for good views and a more enjoyable eating experience. However, the western outskirts south of Middleton and the eastern outskirts by the airport are also popular.

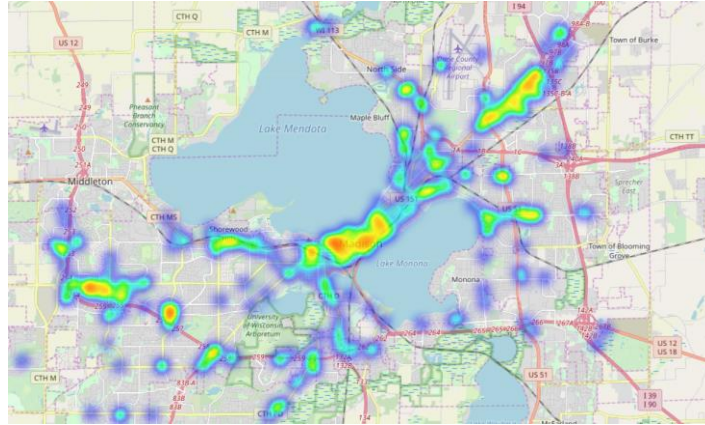


Figure 3: Heat map of restaurant locations in Madison

Table 1 shows part of the “Severe Weather Conditions” data for Dane County. It shows all reported events occurred in a single day in different parts of the County (See “BEGIN_LOCATION”). Also the data include the event type (e.g. Hail, Tornado). Some dates have a single events and others none.

EVENT_ID	CZ_NAME_STR	BEGIN_LOCATION	BEGIN_DATE	BEGIN_TIME	EVENT_TYPE	MAGNITUDE
5440237	DANE CO.	MAZOMANIE	03/30/2005	1331	Hail	0.75
5440238	DANE CO.	MIDDLETON	03/30/2005	1335	Hail	0.88
5440239	DANE CO.	WAUNAKEE	03/30/2005	1355	Hail	0.75
5440248	DANE CO.	SUN PRAIRIE	03/30/2005	1403	Hail	1.25
5440241	DANE CO.	MADISON	03/30/2005	1405	Hail	1.25
5440242	DANE CO.	WAUNAKEE	03/30/2005	1410	Tornado	
5440240	DANE CO.	WAUNAKEE	03/30/2005	1420	Hail	0.75
5440244	DANE CO.	DE FOREST	03/30/2005	1420	Funnel Cloud	
5440330	DANE CO.	MADISON	03/30/2005	1435	Hail	1.50
5440245	DANE CO.	SUN PRAIRIE	03/30/2005	1452	Hail	1.25
5440249	DANE CO.	WINDSOR	03/30/2005	1500	Hail	1.50
5440256	DANE CO.	MADISON	03/30/2005	1605	Thunderstorm Wind	52
5440255	DANE CO.	SUN PRAIRIE	03/30/2005	1616	Hail	0.88
5440318	DANE CO.	EAST BRISTOL	03/30/2005	1622	Thunderstorm Wind	61
5440254	DANE CO.	MT HOREB	03/30/2005	1653	Hail	0.88

Table 1. Severe weather condition original data for March 30 2005.

Analysis of Questions:

1. How has the Madison restaurant scene changed over time?

The Yelp dataset contains restaurant review data from the inception of Yelp in 2005 through the end of 2017. To learn more about how the restaurant scene has changed over time, we looked at review data each year by geography, stars and cuisine groups. Since 2005 and 2006 have limited data, we used 2007 - 2017 for the analysis.

First we looked at geography by splitting up the city into three main areas - between Lake Mendota and Lake Monona, Northeast of the lakes, and Southwest of the lakes. From *Figure 4*, we can conclude that the area with the highest number of reviews consistently is the area between the lakes. This finding is not surprising given the highest number of restaurants are located in the area between the lakes, as seen in our exploratory analysis heat map. Over time, the percent of reviews on restaurants in the Southwest and Northeast regions has increased. Our finding indicates that either more restaurants have opened in these areas or the existing restaurants there are becoming more popular.

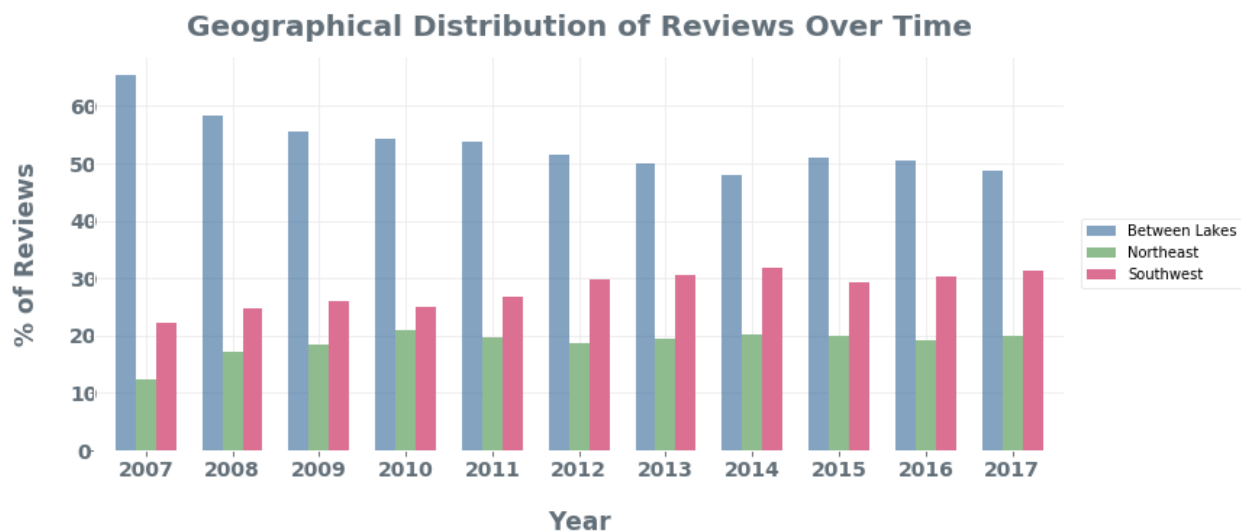


Figure 4: Restaurant Reviews over time in different locations in Madison

Next, we looked at reviewer behavior over time to see if the way Madison restaurants were rated has changed. We found in *Figure 5* that the average number of stars across all restaurant reviews remained steady around 3.7 stars throughout the decade. The average rating was highest in 2007 and 2017 at 3.8 and dipped closer to 3.6 in the years between. When looking further at how the distribution of stars shifted, we found that it became more common for 5 star ratings to be given over the years and it became less common to give out 4 stars (*Figure 6*). The percent of ratings with stars below 4 remained consistent over the decade.

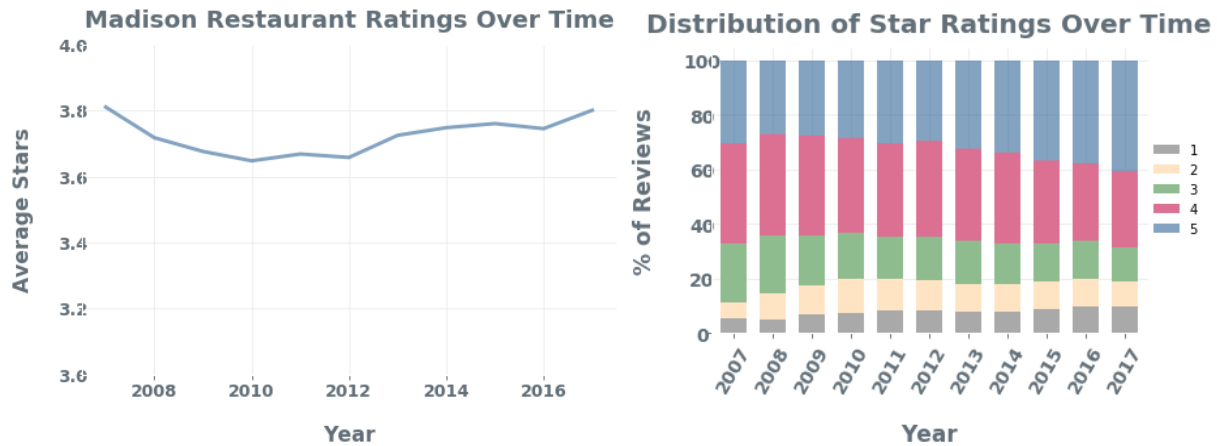


Figure 5: Average stars over time of all Madison restaurants Figure 6: Percent of reviews by star rating over time

The last trend we wanted to learn more about is if the types of cuisines Madison restaurants were serving had changed. To look at this trend, we honed in on just the years 2016 and 2017 and looked at the cuisine breakdown of restaurants with reviews in 2016 compared to the cuisine breakdown of restaurants that opened in 2017. We defined a restaurant as opening in 2017 if it had 0 reviews in 2016 and at least 10 reviews in 2017.

Figure 7 shows that in 2016, approximately 60% of restaurants reviewed in Madison had American cuisine, about 20% served Asian cuisine and the remaining 20% were split between European, Latin American and other cuisine. We found that 30 restaurants opened in 2017, and the cuisine breakdown of those restaurants differed from existing restaurants. Only 40% of new restaurants were classified as American while the European and Latin American groups grew in proportion. This finding indicates a potential shift in the in-demand cuisines in Madison.



Figure 7: Cuisine types in 2016 compared to new restaurants in 2017

2. Does weather influence the restaurant business in Madison?

Weather is an important issue that may affect our daily activities. We dress according to weather. We consider weather when planning outdoor activities. In this section, we explore the influence of weather over the restaurant business perception in Madison, Wisconsin. Temperature and weather conditions were compared to the restaurant reviews in order to establish a relationship with the restaurant perception. We set temperature bands for classification:

Weather Group	Degree Range (F)
Cold	-20-32
Chilly	33-55
Warm	56-70
Hot	71-99

Table 2: The classification of weather groups

In addition, we used information on severe weather conditions to explore if reviews were related with severe weather.

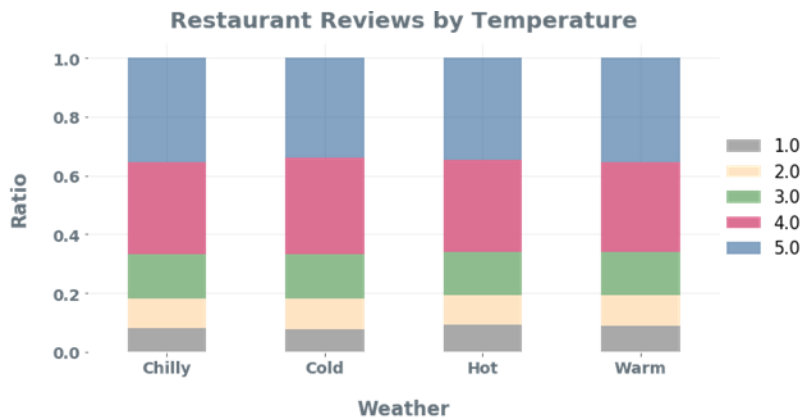


Figure 8: General reviews of the city's restaurants grouped by temperature

Figure 8 shows all reviews (stars) given by clients to restaurants during each temperature range (ranges as described above). We can observe that the ratio between reviews does not change a lot by temperature. Most evaluations (more than 60%) are good or excellent ranking restaurants 5 or 4 stars in all weather conditions. In Figure 9, the data is split for reviews done on days under severe weather conditions and reviews done on “normal” days. Results are very consistent with the data ratio observed in Figure 8: there is no obvious relationship.

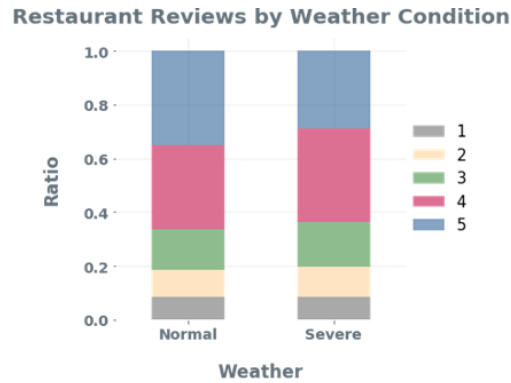


Figure 9: Restaurant reviews during reported extreme weather condition vs normal conditions

Figure 10 shows the reviews to restaurants by cuisine during the weather conditions labeled above. The distribution of reviews by weather is consistent with the general reviews count (e.g. about 60% for American restaurants). This rate does not change much for different weather conditions. Figure 11 shows the overall reviews per day to Madison's restaurants for different weather conditions. Each temperature range has a similar ratio of reviews per day. In general, weather conditions do not have a large effect on the way customers evaluate the restaurant service in Madison, Wisconsin.

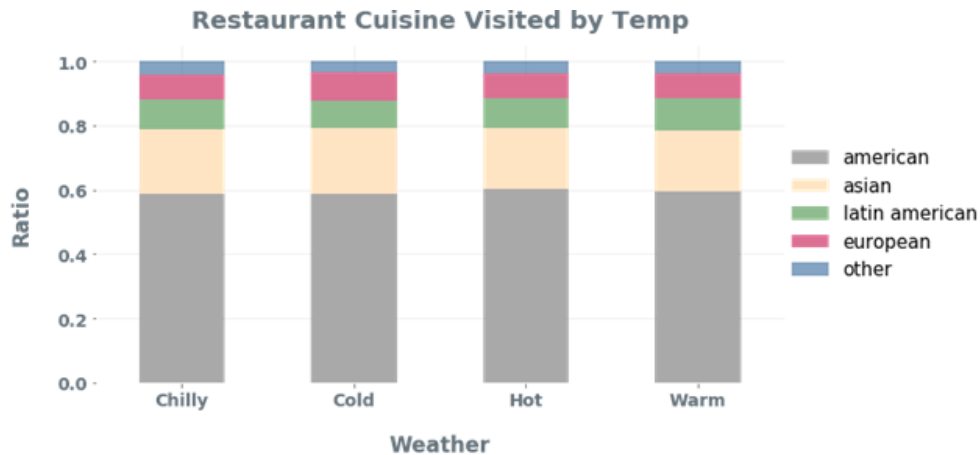


Figure 10: Top 4 cuisine visited restaurants for different temperatures

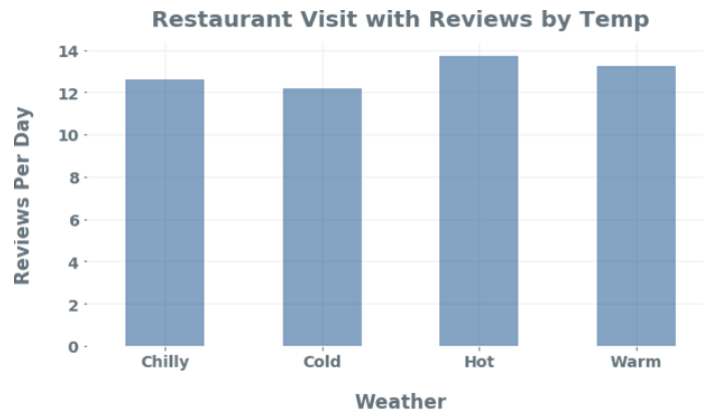


Figure 11: Reviewed restaurants for different weather (temperature) conditions.

3. Are there any restaurants in Madison that are outliers in terms of reviews?

Plotting a boxplot of the number of reviews for each restaurant in *Figure 12*, it is clear that there are some outliers with higher reviews than the rest. Specifically, there is one restaurant with over 1400 reviews (circled in red). Digging into the data, it looks like this restaurant is called "The Old Fashioned", an american restaurant that is highly recommended and regarded and is featured on many Madison travel sites (known for its fried cheese curds).

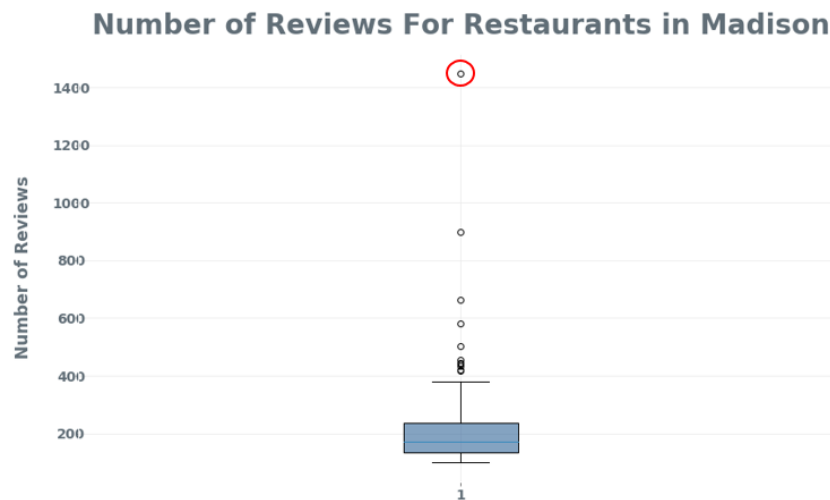


Figure 12: Boxplot of review counts for restaurants in Madison

In order to better understand the restaurant's popularity, *Figure 13* below is a plot of the number of reviews "The Old Fashioned" received over time. The restaurant has continuously increasing reviews from 2005 (likely partly due to its popularity, and partly due to the popularity of yelp) until 2015, where its number of reviews begins to slowly decline. It is difficult to conclude whether this was due to Yelp manipulating reviews (this was around the time that yelp came under controversy for manipulating reviews in order to encourage restaurants to advertise on Yelp), or

whether another restaurant started competing with "The Old Fashioned"(perhaps they had better fried cheese curds).

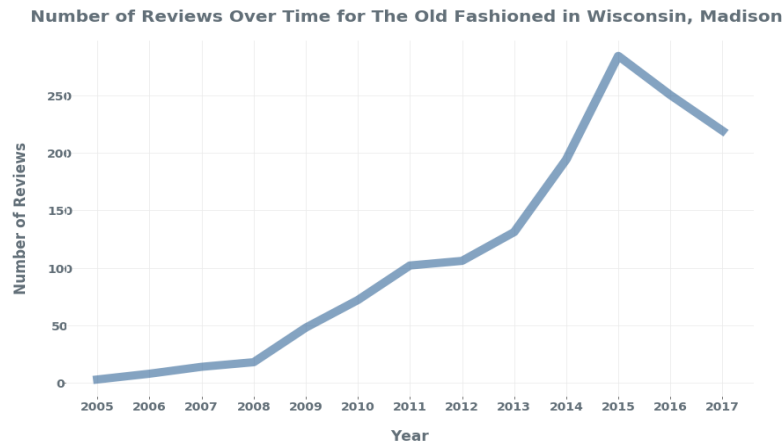


Figure 13: Reviews over time for "The Old Fashioned" Restaurant

Limitations and Future Work: (feel free to add more)

One of the limitations of our analysis was that we assumed that people reviewed a restaurant the same day they attended. Although check-in data was available in our dataset, it did not contain a specific date field, only day of the week, which would not prove useful for our purposes. Ideally we would have had check-in data by day. Another limitation of our analysis is that our data was only available through 2017, so we don't have the most up to date information through today on restaurant reviews in Madison. Our conclusions therefore can only be drawn in the context of the timeframe of our data. In the future, we hope to dive into the open text review field, instead of just focusing on the rating number. Open text can prove to be messy, but can often provide valuable insight.

Conclusions:

When looking over the course of the decade we had data available during, we found trends in geography, reviewer ratings and cuisines. We found that the percentage of restaurant reviews from the geographic region Southwest of the two main lakes in Madison increased over time from 22% to 31%. We found that while the average number of stars given to all restaurants has stayed steady around 3.7, the percent of ratings with 5 stars has increased every year since 2012. Lastly, we found that of the restaurants opened in 2017, a lower proportion of them compared to restaurants in 2016 had American cuisine and a higher proportion had European and Latin American cuisine.

In general, weather conditions do not have a large effect on the way customers evaluate the restaurant service in Madison, Wisconsin. Even extreme conditions do not seem to affect any opinion in food or service or the willingness to give a review.

Although there were a few businesses that were outliers in the Yelp dataset in terms of reviews, we chose to focus on the clear front-runner: “The Old Fashioned.” The restaurant had over 1400 reviews on Yelp, and, although it saw a steady increase throughout the years from 2005 to 2015, its review count slowly decreased after 2015. If we had data past 2017, we would be curious to see if this is still happening now.