TBD 5:

Andrew Ting

Michelle Tran

Emily Wong

Emmanuel Zarate

Bao Pham

Story Description

* Our story first begins with profiling Yelp’s data where we took a look at the first 80 reviews of each business. We figured the first 80 reviews would be a good amount because Yelp’s co-founder noted that with around 75 reviews, businesses begin to enter the high review business category.
* In the profiled data, we separated each bin based on the number of reviews of each business. Bin 1 includes the first 5 reviews of each business, bin 2 includes the 6th to 10th reviews, bin 3 includes the 11th to 16th reviews, and so on. The profiled data graph we created displayed useful votes as the highest in the first 5 reviews per business.
* The data answers the main question which proves that reviews do get voted useful when there are fewer reviews for a business.
* Next, we created a box plot that illustrates the useful votes per bin. Bin 1 has 398 which was the most useful votes out of all.
* We then extended through categories to figure out the top categories. Restaurants and food were the top two categories that had the highest number of businesses. A few of the categories did not have much data so we did not combine them.
* We figured out which categories should be combined. Automotive businesses should have their own category while nightlife businesses should be combined with restaurants.
* We combined categories such as restaurants, food, and nightlife into restaurant-related categories. Home, local, and professional services were combined into service-related categories. Combining these categories will provide better visualization of the number of useful votes for each category.
* We analyze the data in combined categories and overall, restaurant-related had the highest reviews counts and useful votes out of all categories.
* We created a heat map to show that earlier reviews do receive useful votes for all categories.
* Next, we have created a boxplot to show the useful votes per bin for restaurant-related, service-related, and shopping categories. The trend of reviews is getting voted useful more often when there are fewer reviews repeating across all categories.
* Lastly, we have a graph that displays the sum of funny and cool votes. It is similar to the votes being useful when there are fewer reviews. If the question was about the funny and cool votes, it would be true as well across all bins and categories.