**Group Project 1 (Group 2)**

**The Yelp Project**

The purpose of this analysis is to explore the best locations to visit by ratings in North America and Canada.

* These recommendations were gathered from Yelp.com

**Data limitations:**

* Yelp data is from 8 months ago
* The data does not include all 50 states in the United States
* There is a total of 29 United States and 7 Canada states
* Concluded that the data is bias since we are missing 42% of the states in the United States and missing 30% for Canada
* We suspect that Canada states abbreviations are inaccurate

**Data Transformation:**

* Original file was converted from json to csv
* Data was split into 32 unique categories for activities
* Due to the volume in the data we used the methods
  + logarithmic value = base e (ex:e^14)
  + square root of the column

**Graph Ratings by State:**

**Review Counts by State:**

* The review count state graph indicates Nevada and Arizona are the two top rated US states followed by Ontario Canada.
* Based on the Yelp reviews we can conclude Nevada and Arizona are popular states to visit.
* The least three rated states are in Canada.

**Restaurants Reviews Counts by State:**

* The restaurants reviews count by state the graph indicates Ontario, Canada has the most ratings then followed by Arizona and Nevada in US
* The least three rated states are AR, VA, TX in the US
* There are 10 states on the graph with no ratings
* Based on the Yelp reviews we can conclude Ontario, Arizona and Nevada are popular states to dine.

**Shopping Reviews Counts by State:**

* The shopping reviews count graph indicates Arizona, Nevada and the US have the most ratings followed by Ontario Canada
* Overall based on the graph individuals do not rate shopping experience

**Graph Ratings by City:**

**Restaurants Reviews by City:**

* Took a deeper dive into the restaurants ratings by city and we concluded that the top 5 cities are Toronto Canada, Las Vegas NV, Phoenix AZ, Montreal Canada, Calgary Canada
* According to this graph Canada has 3 out of the top 5 restaurant locations

**Shopping Reviews by City:**

* Took a deeper dive into the shopping ratings by city and we concluded that the top 5 cities are Las Vegas NV, Toronto Canada, Phoenix AZ, Scottsdale AZ, Charlotte NC
* According to this graph US has 4 out of the top 5 shopping ratings

**Golf Reviews by City:**

* Most of the cities we observed have less than 100 reviews meanwhile Las Vegas NV, Toronto Canada, Phoenix AZ, Charlotte NC, Scottsdale AZ, reviews are more than 1,000.00
* The golf graph data reflects that there is consistency in people sharing their reviews on their golf experiences

**Conclusion A:**

In conclusion Yelp data is incomplete most of the ratings are concentrated in Nevada, Arizona and Canada consequently causing limitations to provide objective recommendations.

The Hotel data is all-encompassing when compared to Yelp data. The most popular visited places were in the east coast and west coast, near beaches of the US.

Based on the analysis the recommended vacation place is Las Vegas Nevada.

**Stars by State:**

* The average star by state graph illustrates that Tennessee and New Jersey have the highest average of star ratings followed by West Yorkshire and London, England.
* Based on the average stars that Tennessee and New Jersey have, they would have the most recommended places to go to.
* The states with the least amount of average star ratings are Utah, Arkansas, and British Columbia.

**Stars by City:**

* When reviewing deeper into the data for the average stars by city, it seems the data may not be completely accurate and most likely biased.
* Based on the illustration, the top ten cities average star ratings are at five stars with the top city being Las Vegas and the last being Cave Creek Road.

**Conclusion B:**

Possibly many more cities have the same number of average stars while the number of reviews varies greatly in numbers. For example, some cities only have 1 review.

**Final Recap and Conclusions:**

The purpose of our project was to find interesting place to visit in the US. Our mission was to use the data from Yelp, and hotel reviews from Datafiniti. Our idea was to take the ratings of different categories from the yelp dataset and compare them with the ratings from the hotel data set and correlate their location. The idea was the places with the best reviews in dining, shopping, golf, and hotels to determine where one would likely get the biggest bang for their buck.

We found the datasets for Yelp and the hotels on Kaggle.com. Both sets were good size. The yelp set contained 192k reviews spanning many different categories. The hotel dataset contained nearly 78k reviews. The yelp data set was imported as a json file where we read it in and converted it to a csv file. This gave us tremendous difficulty. In the json file, the category field contained many commas which really skewed the data. So, when we pushed it to our project folder, it was unusable. We only discovered this when we were trying to query data for charts and such. So ultimately, we converted the file to an xlsx file that was useable and shareable. The hotel data was downloaded from Kaggle in csv format which was easy to use.

So again, our strategy was to get ratings that included shopping, golf, dining/nightlife, and hotels. We would chart all of these categories and plot their locations using the lat and lng coordinates. We found the yelp data set to be unreliable as the reviews were super concentrated in only 4 states. When compared to the hotel data, which seemed to be clean, we could not come to a conclusion that made sense based on the dataset from yelp. Everything from the yelp dataset pointed to Las Vegas.