

## Group 4- Project 1 Proposal

### Topic: Michelin Star Restaurants<sup>1</sup>

Our exploratory data analysis will be over Michelin Star Restaurants in 2019, provided on Kaggle. Three datasets are given regarding one, two, and three star Michelin restaurants. The datasets provide information on restaurant location, including latitude, longitude, city, region, and zip code. The data also includes cuisine type, and price.

Inspiration regarding the data was located on Kaggle from previous analyses of the data. A previous exploratory data analysis from Luna McBride<sup>2</sup> provides ideas on dealing with null values, and incorporating map usage. Analysis from Thomas Konstantin<sup>3</sup> included ideas on how to address missing price values. Possible use of the US Census data API to the Michelin star data is also an option for analysis.

In order to analyze the data, we will merge the data sets for one, two, and three star restaurants. Null values for zip code will be addressed and the analysis will focus on latitude and longitude of the restaurants. The restaurant rating and price point are currently provided in string values and will need to be converted to integer or grouped by the string to support analysis.

Our data analysis will include multiple visual representations of the analysis. A map of the locations of restaurants will be populated using the latitude and longitudes, domestic and international. We will also color code the location based upon the Michelin Star rating. A deeper analysis on the restaurants within the United States will look at price, cuisine popularity, and region. We wish to determine the region in which certain cuisines are most prevalent. Other comparisons will include price level frequency by region or city, number of restaurants by region, and the distribution of three vs two and three versus one star restaurants by region. This will be color coded in a similar way to the map. US Census data on population by city or region can also be compared to the amount of Michelin star restaurants within said city or region.

Prediction analysis on Michelin stars based upon price or location is a possible option for our exploratory analysis. The use of a chi-square test to review data is also a possible option. The possibilities are still being explored after our initial review of the data. We plan to do a logistic regression rather than a linear regression.

---

<sup>1</sup> <https://www.kaggle.com/jackywang529/michelin-restaurants?select=one-star-michelin-restaurants.csv>

<sup>2</sup> <https://www.kaggle.com/lunamcbride24/michelin-star-exploration-classification>

<sup>3</sup> <https://www.kaggle.com/thomaskonstantin/michelin-restaurants-eda-missing-price-prediction>