

## Best Restaurants in the World in 2018

### **Abstract**

The restaurant industry is one of enormous competition. A combination of a large number of restaurants placed in close proximity to each other, and the variety in cuisine options creates a difficulty in entering and succeeding in this industry. For restaurant owners, it is important to determine which restaurants are more likely to succeed in their country or city. This project utilizes visualizations to display the geographical distribution of cuisine types and describe the differences in ratings for the best restaurants in the world. The aspects of this analysis that best determines where restaurants are most likely to succeed are heavily focused on restaurant distribution and cuisine type. This correlation is fairly obvious when looking at the percentages of restaurants in each country and the percentages of cuisine types from top-rated restaurants. This report will discuss the various objectives, requirements, and conclusions explored in this project.

### **Objectives**

The goal of this project is to display the geographical distribution of cuisine types and describe the differences in ratings for approximately 9500 global restaurants. For that purpose, this project will utilize factors including city and country locations, restaurant ratings, price ranges, and the types of cuisine served. For example, a higher or lower density of restaurants in a certain country can be an indicator of low opportunities for a restaurant to succeed. Additionally, taking a look at percentages of top-rated restaurants based on factors of price, location, and cuisine type can provide a better understanding of the most desired cuisines and determine where certain restaurant types may flourish. For further understanding, this analysis will also look more carefully at restaurant distribution on a city level, especially for India, which has a much larger number of restaurants in comparison with the other countries.

### **Functional Requirements**

By the conclusion of this project, the restaurant data and relevant factors will be stored and summarized into four types of datasets. There will be a general dataset by country, a dataset on percentages by top-rated restaurants, a general dataset by cities, and a final dataset of restaurant rating scores relating to price ranges. The summarized datasets will be formatted such that effective visualizations can be created, and conclusions can be drawn. The user will be able to interface with a web page and navigate using various tabs and links. They will be able to view data and interactive visualizations by selecting relevant topics. These interactions include drop-down menus and graphs with mouse over effects.

### **System Architecture and Description**

The data preprocessing, consolidation, and summarization processes were completed using a combination of Microsoft Excel and Python in Jupyter Notebook. The raw data used for this project was taken directly from Kaggle and was updated with the most up-to-date information in 2018. For this project, multiple two-dimensional data frames were created. For each visualization, there was an associated Python data frame that was then exported into separate CSV files as data tables to be loaded into D3 JavaScript library. All of the data, including the raw data file, consolidated structured data, and data tables were stored locally.

### **Development Platforms**

The user-facing front end of this analysis is an HTML based web-page acting as a homepage with descriptions of the project, data source, references, and a link to another HTML based web-page with various tabs organized by the data content. For example, Best Rated Cuisine and Restaurants by Country are tabs on this web-page. The web-pages are interactive, with links, drop down menus, and mouse over visualizations. Hovering over the visualizations will allow the user to see call-out boxes with labels and specific information on percentages and values. Cascading Style Sheets (CSS) will be used to control the layout and presentation of the web-pages. And the visualizations will be created using the D3.js library.

### **Proposed Visualizations**

The visualizations in this analysis will be created using the D3.js library. Choropleth world maps will be used to visualize the density of restaurants for 15 specified countries using a color scale. Bar graphs will also be

used as clearer visualizations of differences between number of top-rated restaurants and cuisine types in each country. Then for percentages of top-rated restaurants, pie charts will be used to compare between countries and cuisine type. As an extension of the restaurant density visualizations, bar graphs and geographical maps will also be used to look at the data on a city basis, for countries of the highest restaurant density. Lastly, scatter plots and bar charts of price ranges against restaurant rating scores will be used to determine any correlation for cities with the largest restaurant densities.

### **Experimental Analyses and Conclusions**

The data first revealed, that by a large extent, there is the greatest number of restaurants located in India. With a total of 8652 restaurants in India, the country with the second highest density (the United States) has only 425 restaurants. But looking at the restaurant ratings, India has the lowest percentage of restaurants with a rating of Excellent (rating of 4.5 and higher), at approximately 1.3%. Taking a closer look at the percentages, the Philippines has the largest percentage of Excellent rated restaurants, at approximately 54.5%. Though there are only 22 restaurants from the Philippines in this dataset, 12 of those were Excellent rated, as opposed to India where 116 out of 8652 were Excellent rated. This is a very wide difference in percentages but may simply be a result of not enough data points from the Philippines. But when looking at the percentages for all 15 countries, there does not seem to be a clear correlation between restaurant density and rating. Rather, it seems more likely that certain countries have higher standards for restaurants than others.

The data also revealed that there is not any correlation between cuisine type or price ranges and restaurant ratings, regardless of locations and restaurant density. This indicates that there are other hidden factors effecting the ratings, since there is no other reason for pricing or cuisine type to have such minimal effect on ratings. Therefore, for a clearer conclusion to be reached, this analysis requires more relevant factors to be included.