**Abstract**

For this project, the business problem I’m trying to solve is determining the ideal location and features for opening a new restaurant in Europe.

**Design**

The focus of the project is understanding which European cities and restaurant features (distinct cuisine styles and price ranges) have a strong association with high customer reviews of at least 4.0 out of 5.0

My client is a European restaurant group looking to open a new restaurant. Analyzing data about existing restaurants with high customer reviews will help them create a new restaurant that achieves the same measure of success.

**Data**

I used a Tripadvisor dataset found on [Kaggle](https://www.kaggle.com/datasets/damienbeneschi/krakow-ta-restaurans-data-raw). This dataset was scraped in 2018 and includes restaurants in 31 different European cities: Amsterdam, Athens, Barcelona, Berlin, Bratislava, Brussels, Budapest, Copenhagen, Dublin, Edinburgh, Geneva, Helsinki, Hamburg, Krakow, Lisbon, Ljubljana, London, Luxembourg, Madrid, Lyon, Milan, Munich, Porto, Oslo, Paris, Prague, Rome, Stockholm, Vienna, Warsaw, and Zurich. This data set includes all restaurants registered on Tripadvisor.com in each city, but this is not 100% comprehensive of all the restaurants that exist in each city. The dataset includes 125,433 total restaurants. Here are the additional restaurant features I focused on for my project:

* Cuisine Style – The raw data included a list of string values for each restaurant. From this data I extracted 127 different cuisine style tags.
* Rating – This is the average customer rating for each restaurant out of 5.
* Price Range – $ (Cheap Eats), $$ - $$$ (Mid-range) and $$$$ (Fine dining) are the 3 categorical values
* Reviews – The text of 2 reviews is included for each restaurant (list of lists with review text and dates)

**Algorithms**

Most of my data cleaning and transformation in Excel was focused on the Cuisine Style attribute. The draft versions of my data in Excel (v1, v2, and v3) are attached for visibility regarding the different functions I used. I SPLIT each list of strings into columns, used TRIM on the values, and found the UNIQUE values from each column. I combined my lists of unique to finally determine that there are 127 different cuisine styles present in the dataset. Then, I created a column for each cuisine style. I used ISNUMBER and SEARCH to determine if the style in the column of interest existed in the Cuisine Style list of strings column from the raw dataset. I converted the true/false outputs to binary values so that I could use the SUBTOTAL function and create value counts when filtering down the data by feature and city. From the initial 127 cuisine styles, I narrowed my focus to the 98 styles which each appeared at least 50 times in the dataset. I also used pivot tables in Excel for some initial exploratory data analysis regarding city distribution and rating distribution.

In Tableau, I connected to the 3 different sheets in my final Excel file (“TA\_restaurants\_curated - updated - binary”) and visualized the data with symbol maps, bar charts, histograms, and line charts. I brought together visualizations from different worksheets to make comprehensive dashboards.

**Tools**

I used a variety of functions in Excel and Tableau to effectively analyze the Tripadvisor dataset

**Communication**

I have attached the PowerPoint slides which includes the screenshots of my Tableau dashboards.

The interactive version of my Tableau worksheets and dashboards can be found on Tableau Public:

<https://public.tableau.com/views/EuropeanRestaurantsProject/PricebyCityDB?:language=en-US&publish=yes&:display_count=n&:origin=viz_share_link>