**Project Description**

Fai Abdulrahman Albadri

Abstract

The goal of this project is to use classification models to show the relationships between the elements for each restaurant, through which we can determine the preferred food for the residents of Riyadh according to the price, location and ratings, and from them we can help people who want to open a restaurant project. It worked with data provided by [Kaggle]

Design

This project is one of the T5 Data Science BootCamp requirements. Data provided by [Kaggle]has been used in this project. The [Dataset] include most of the restaurant names in Riyadh (scrapped from foursquare.com) . Classifying restaurants adresses,likes ,prices and rating using machine learning algorithms would enable us to understand if thier are relationships between restaurants addresses and prices and also comparing with the rating.

## Data

The dataset is provided in .csv format. It contains 19360 Restaurant , each Restaurant has 11 features. The most relevant feature to this project is the text which contains the Restaurant name. Some other features are extracted from other features such as the address it is extracted from location, where it contains ( Latitude,Longitude ). Other important features are the price,Rating and likes that we can extract how these features have relationship between them to effect on the customers.

This dataset contains following features:

- name: The name of the Restaurant. Some names are only available in Arabic.

- categories: The list of categories (in English) separated by a comma.

- address: The address according to foursquare. Unfortunately, it is the least useful column in the dataset because lots of restaurants lack a formal address.

- lat: Latitude.

- lng: Longitude.

- price: shows the price category (Cheap, Moderate, Expensive, Very Expensive)

- likes: The number of likes.

- photos: The number of photos. Photos themselves are not included.

- tips: The number of tips in Foursquare (e.g., Don't miss their pasta). Tips themselves are not included.

- rating: The average rating out of 10.

- ratingSignals: The number of raters.

Algorithms

Linear Regression

Shows the relation between the features in restaurant.

Plotting for all features with the price and category

Heatmap

Shows the number of likes for restaurants according to the price range and category.

Tools

- Numpy and Pandas for data manipulation

- Matplotlib and Seaborn for plotting

Communication

The slides are provided in my github