# Restaurant Trend Analysis and Success Prediction

Team 17

NUID: Khushbu Singh 002222644

NUID: Lakshmi Kantha Monisha Pulavarthy 002291093

**Introduction**

**Background**

The dining industry has evolved significantly with the advent of digital platforms, changing consumer preferences, and the global expansion of culinary cultures. Zomato, as a leading restaurant discovery and review platform, offers a rich dataset capturing a wide array of variables related to dining establishments worldwide. The analysis of this data set presents an opportunity to uncover patterns in customer preferences, restaurant success factors, and emerging culinary trends. Traditionally, restaurant success was gauged by local word-of-mouth and critical reviews in print media. However, in the digital age, data analytics offers a nuanced understanding by correlating various factors such as location, cuisine, ratings, and reviews with restaurant popularity and success.

**Motivation**

Our motivation is driven by the potential to leverage data analytics in enhancing the dining experience, supporting restaurant owners in making informed decisions, and predicting future dining trends. By analyzing the Zomato dataset, we aim to identify key factors that contribute to a restaurant's success, understand global dining trends, and provide insights into customer preferences. This project is about harnessing data to benefit stakeholders in the culinary industry, including customers, restaurant owners, and culinary entrepreneurs.

**Goal**

Our goal is to conduct a thorough analysis of the Zomato dataset to:

1.Identify patterns and trends in customer dining preferences.

2.Determine key factors that influence restaurant ratings and success.

3.Forecast emerging culinary trends on a global and regional scale.

4.Provide actionable insights to restaurant owners for strategic decision-making.

**Methodology**

1.**Data preprocessing and cleaning**: We will clean the Zomato dataset for accuracy and usability, like removing irrelevant or duplicate data, addressing missing values, standardizing formats, and ensuring the data is structured for analysis.

2.**Exploratory data analysis (EDA)**: We will use visualization and statistical techniques such as histogram, box plot, bar charts etc., to explore the dataset, identifying trends in restaurant popularity, cuisine preferences, and the impact of location on success.

3.**Feature engineering**: We plan to select and engineer features that are indicative of restaurant success, informed by domain knowledge and statistical analysis. This may include creating composite indicators from existing variables.

4.**Model Selection and Evaluation**: Our approach includes applying machine learning models (e.g., linear regression, Random Forest, Decision Tree) to predict restaurant success and customer preferences. Model performance will be rigorously evaluated to ensure reliability and accuracy.

Decision Trees: This approach is beneficial for both classification and continuous variable prediction, making it ideal for dissecting the complex relationships within the Zomato dataset, such as predicting restaurant ratings based on various attributes.

Random Forest Classifier: Random Forest aggregates the insights from multiple decision trees to form a more accurate and stable prediction model. Its capability to manage incomplete data sets it apart as a versatile tool for analyzing the Zomato dataset, particularly when predicting success metrics for restaurants based on diverse and potentially incomplete information.

Linear Regression: For the Zomato dataset, linear regression can be used to predict quantitative outcomes such as the average cost for two or rating based on various predictors like location, number of reviews, or cuisine type.

5.**Trend analysis**: We will conduct trend analysis to identify emerging dining trends, preferences shifts, and potential market gaps in the culinary industry.

**Description of the dataset:**

The Zomato dataset is an extensive compilation of data related to restaurants listed on the Zomato platform, covering various aspects such as location, cuisine, customer ratings, and reviews. The dataset contains entries for thousands of restaurants across multiple countries, offering a comprehensive view of the global dining landscape. Key columns include Restaurant ID, Name, City, Address, Cuisine, Average Cost for two, Rating, Votes, and Reviews.

The dataset exceeds 500 MB of data containing 17 columns and 51717 entries.

**Data source:**

https://www.kaggle.com/datasets/rishikeshkonapure/zomato/data