**Project Overview:**

This project is a comprehensive data analytics study of the dataset of restaurants listed in Zomato from Bangalore by year 2019. The primary objective was to uncover key trends and business insights from raw data to understand better the restaurant industry, market dynamics, and consumer behavior. The analysis involved a full-cycle data workflow from data cleaning and transformation to the generation of actionable insights and making report using visualizations

**Data Exploration:**

**1. Initial Data Overview & Scope**

* The project was initiated by analyzing a dataset of 60,783 rows and 17 columns, providing comprehensive data on restaurants in Bangalore.
* The dataset was initially found to have a high percentage of missing values and duplicate rows, which required significant data cleaning before analysis.

**2. Key Data Quality Challenges & Solutions**

* Irrelevant Columns: Columns like URL and address were identified as non-essential for the core analysis and were removed to optimize the dataset's size and improve query performance.
* Data Consistency: Addressed major data inconsistencies, including improper character encoding (e.g., "Café") and mixed data types, to ensure uniformity.
* Handling Nulls: Replaced null values in the critical Ratings column with an appropriate strategy (e.g., average ratings for known restaurants and 0 for unrated ones) to prevent data loss and allow for comprehensive analysis.
* Identifying "Noise" Data: Eliminated "noise" data, such as irrelevant reviews and ratings, that were incorrectly placed in columns like Menu Item and online\_order.

**3. Data Refinement & Feature Engineering**

* De-duplication: Successfully removed duplicate rows to ensure each observation represented a unique record, leading to an accurate count of restaurants.
* Complex Cleaning: Handled intricate data cleaning tasks like separating multiple phone numbers within a single cell and resolving inconsistencies in restaurant names.

**Data & Methodology:**

* **Data Source:** The project utilized a multi-table dataset provided in CSV format, containing information on restaurants, cuisines, and customer-liked food items.
* **Data Cleaning and Preparation:**
  + Addressed data inconsistencies by handling missing values and ensuring uniform data types.
  + Cleaned and standardized textual data, including correcting duplicate restaurant names and generalizing specific food items (e.g., "Mutton Biryani," "Dum Biryani") into a single "Biryani" category for accurate popularity analysis.
  + Enriched the dataset by integrating data from multiple tables to create new key columns, such as adding Approx#Cost\_for\_two to the Liked\_Foods table.
* **Methodology:** All data manipulation and analysis were performed using **SQL Server**, leveraging advanced functionalities such as Common Table Expressions (CTEs), views, and window functions (e.g., ROW\_NUMBER(), RANK()) to segment and rank data.

**Key Findings & Insights:**

* **Market Segmentation:** Restaurants were segmented into key categories, revealing a breakdown of services offered (online order vs. table booking) and budget tiers (Costly, Budget-Friendly, Low-Budget).
* **Geographic Analysis:** The study identified the top 10 locations with the highest concentration of restaurants and correlated these areas with average ratings and cost, providing insights into a location's overall dining quality and affordability.
* **Customer Preferences:**
  + The most popular cuisines were identified based on the number of restaurants serving them.
  + A list of the most-liked food items was generated, indicating strong consumer preferences for specific dishes.
* **Relationship Analysis:** An analysis was conducted to determine the relationship between a restaurant's average rating and its cost, revealing whether higher ratings are correlated with higher prices.

**Tools & Technologies:**

* **Data Cleaning:** Power query,Excel
* **Data Analysis:** SQL (Views, CTEs, Window Functions)
* **Data Visualization:** Tableau