# Analyzing Swiggy: Bangalore delivery outlet data

Objective:

This project's purpose is to extract, convert, and load (ETL) data from Swiggy's Bangalore delivery outlet in order to do data mining, discover critical indicators, and uncover significant links between various parameters. The project requires the use of analysis and visualization technologies such as Python, and Power BI.

Benefits:

➢ Improved decision-making for corporate strategy.

➢ Increased understanding of customer behavior and preferences.

➢ Optimization of delivery processes and logistics.

➢ Identifying trends and patterns in the online meal ordering sector.

Power BI Architecture Overview
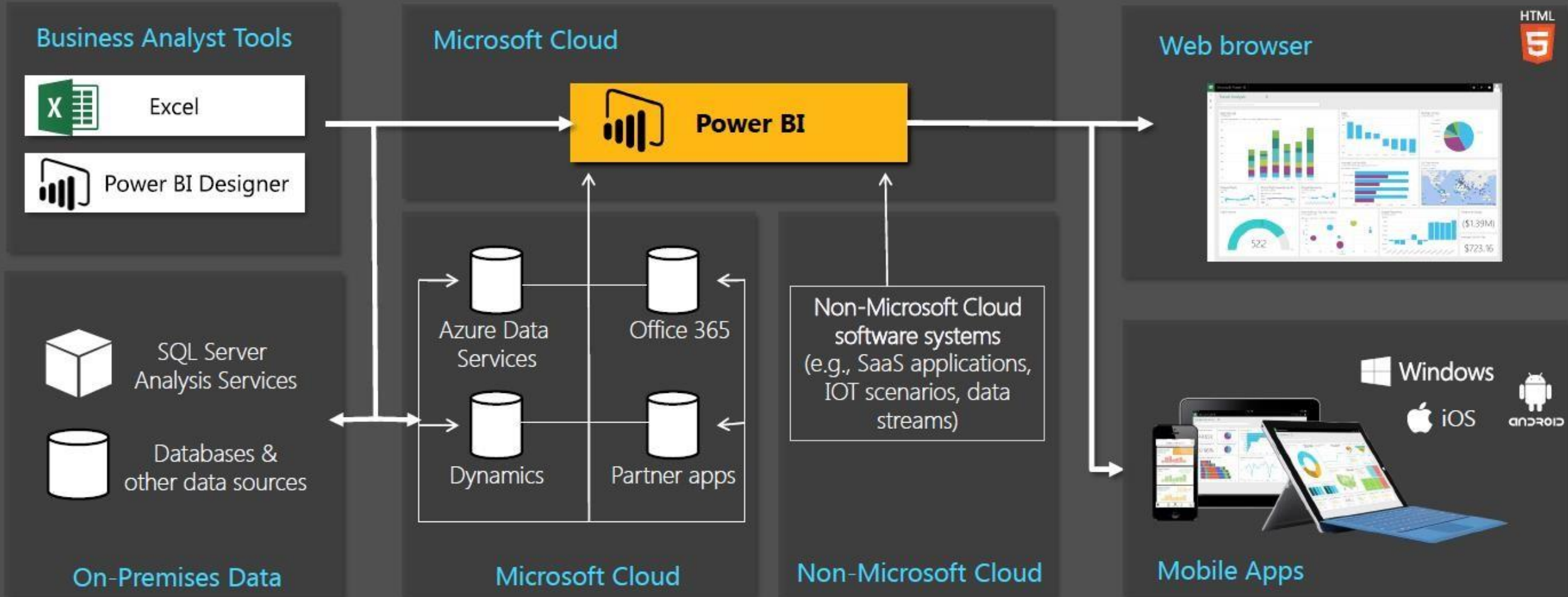
Power BI is a powerful business analytics tool that allows users to visualize data and share insights across their organizations. It can connect to hundreds of data sources, simplify data prep, and drive ad hoc analysis. Power BI provides a suite of tools including Power BI Desktop, Power BI Service, and Power BI Mobile, enabling users to create interactive reports and dashboards.

Data Validation:

Data validation assures the correctness and quality of data prior to processing. Steps involved:

Integrity checks: Ensure that there are no missing, duplicate, or incorrect data entries.

Format validation ensures that data follows the specified format (e.g., date formats and numerical values).

Range checks: Ensuring that numerical values are inside a reasonable range.

Referential Integrity: Ensuring that tables' associations are consistent (for example, valid foreign keys).

Data Transformation:

Data transformation include cleaning, organising, and enriching data to make it ready for analysis. Key steps include:

Data Cleaning: Dealing with missing values, outliers, and duplicates.

Normalisation is the process of scaling numerical data to provide uniform analysis.

Feature Engineering: Creating new features from existing data to improve the model's prediction power.

Aggregation is the process of summarising data to the necessary level of granularity.

Data Insertion in the Database:

Data insertion is the process of inserting cleansed and converted data into a database or data warehouse. This stage includes:

Schema Design: Defines the database's structure, including tables, fields, and relationships.
Data loading is the process of effectively inserting data into a database using ETL technologies.
Indexing refers to the process of creating indexes to increase query performance.

Model Training:

Model training is utilizing machine learning techniques to generate prediction models based on data. Steps include:

Algorithm Selection: Choosing the best algorithms for the task (for example, regression and classification).
Training Models: Data is divided into training and testing sets, the models are trained, and hyperparameters are tuned.
Model evaluation involves assessing model performance using criteria such as accuracy, precision, recall, and F1-score.

Prediction:

Using trained models to generate predictions about new data. This involves:

1. Deployment entails integrating the model into a production environment.
2. Prediction: Applying the model to fresh data to make predictions.
3. Monitoring: Constantly assessing model performance and retraining as needed.

Key Metrics and Relationships:

Some essential indicators and criteria to analyze are:

- Order volume refers to the total number of orders placed over time.
- consumer segmentation is the process of identifying various consumer categories based on their ordering behavior.
- Peak ordering times are the times of day and days of the week with the highest order volume.
- Delivery Times: Average and Distribution of Delivery Time.
- Restaurant Performance: Ratings and analysis of various establishments.
- Geographic Insights: Order dispersion in Bangalore.

1. What is the main objective of your project?

2. What tools are you using for the project?

3. What does ETL stand for and why is it important?

4. What challenges did you encounter during data cleaning?

5. How did you handle missing data?

6. What key metrics did you analyze?

7. Which machine learning algorithms did you use?

8. How did you validate the accuracy of your models?

9. What insights did geographical analysis provide?

10. How can businesses apply your findings?

1)What is the main objective of your project?

The primary goal is to analyse Swiggy's Bangalore delivery data, finding important indicators and patterns to give actionable insights that improve corporate decision-making and operational efficiency in the online food delivery industry.

2)What tools are you using for the project?

The tools used include Python for data analysis, R for statistical modeling, Tableau and Power BI for data visualization, and for efficient ETL processes and data preparation.

3)What does ETL stand for and why is it important?

ETL stands for Extract, Transform, Load. It is crucial for preparing data by extracting it from sources, transforming it into a suitable format, and loading it into a database for analysis.

4)What challenges did you encounter during data cleaning?

Challenges included handling missing values, outliers, and inconsistent data formats. Ensuring data integrity and standardization was essential for accurate analysis.

5)How did you handle missing data?

Missing data was managed through imputation, replacing missing values with mean, median, or mode, and in some cases, by removing incomplete records to ensure data quality.

6)What key metrics did you analyze?

Key metrics analyzed include order volume, peak ordering times, delivery times, customer segmentation, restaurant performance, and geographical order distribution.

7)Which machine learning algorithms did you use?

I used regression algorithms for predicting delivery times and clustering algorithms like k-means for customer segmentation to uncover distinct groups based on ordering behavior.

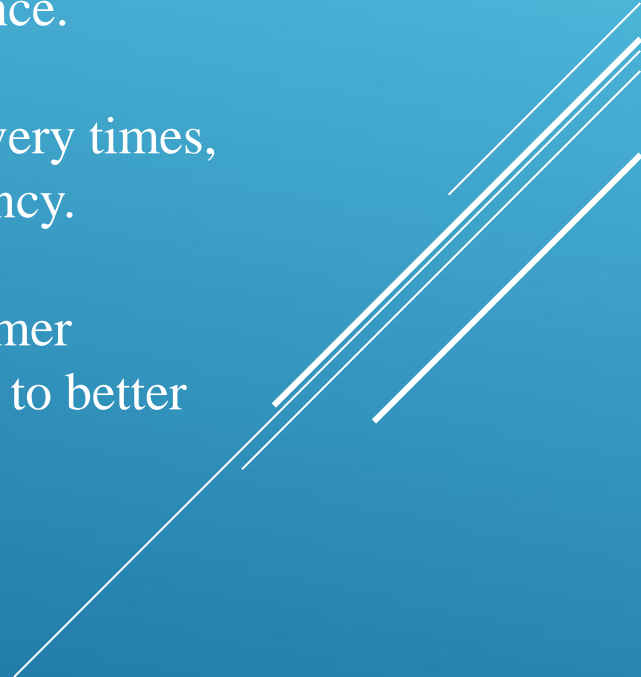**8)How did you validate the accuracy of your models?**

Model accuracy was validated through cross-validation, using metrics like RMSE for regression and silhouette score for clustering, and tuning hyperparameters for optimal performance.

9)What insights did geographical analysis provide?

The geographical analysis identified high-demand areas and regions with longer delivery times, indicating potential logistical issues that can be addressed to improve delivery efficiency.

10)How can businesses apply your findings?

Businesses can optimize delivery routes, tailor marketing strategies to different customer segments, and improve restaurant performance based on data-driven insights, leading to better customer satisfaction and operational efficiency.