# Store Sales Data Forecasting using AI - Prophet Method

## **Project Overview**

This project focuses on analyzing store sales data to uncover key insights and forecast future sales. The primary objective is to use data visualization techniques to understand sales patterns across different product categories, shopping malls, and customer demographics. Subsequently, the project employs the Prophet AI model, a time series forecasting method, to predict sales for the upcoming month based on historical trends.

## **Raw Dataset**

The analysis is based on the Different\_stores\_data\_V2.csv dataset.

Link: https://www.kaggle.com/datasets/kzmontage/sales-from-different-stores

## **Insights & Findings**

Through exploratory data analysis and visualization, several key findings were identified:

* **Top Selling Categories**: **Clothing** is the highest-selling category by a significant margin, followed by Cosmetics and Food & Beverage.
* **Top Performing Malls**: The shopping malls with the highest sales volumes are **Kanyon**, **Mall of Istanbul**, and **Metrocity**. Clothing consistently emerges as the top-selling category across these leading malls.
* **Sales by Gender**: While **females** account for a higher quantity of items sold overall, the top 5 preferred categories are nearly identical for both males and females, with Clothing leading for both groups.
* **Profitability Insights**: In the top-performing malls, the most profitable categories are **Clothing**, **Technology**, and **Shoes**.

## **AI Support Explanation**

The **Prophet model** was selected for time series forecasting due to its robustness in handling datasets with seasonal patterns and trends, even with a relatively short time frame of historical data.

* **Reasoning for Choice**: Prophet is well-suited for this task as it effectively models daily and weekly seasonality, which was observed in the sales data trends. It is also resilient to missing data points and outliers.
* **Model Performance**: The model was trained on the daily sales data. The evaluation metrics below represent the model's in-sample fit error:
  + **Mean Absolute Error (MAE)**: 1339.22
  + **Mean Squared Error (MSE)**: 5218145.28
  + **Root Mean Squared Error (RMSE)**: 2284.33

The model successfully generated a forecast for the next 30 days, visualizing both the overall trend and the weekly seasonality components.