Walmart Sales Prediction Machine Learning Project

**Problem Statement**

Predict the sales of Walmart stores.

**Dataset**

Kindly download the data from [here](https://daxg39y63pxwu.cloudfront.net/hands-on-labs/125/walmart-sales-prediction-machine-learning-project/datasets/Walmart.csv).

**Tasks**

1. Hypothesis based EDA:
   * Share of sales and profit by product category (top 5)
   * Share of profit by region
   * Share of sales and profit by product name (top 5)
2. Plot the trend of sales and profit over time as line chart
3. Create new features:
   * Create average sales per category and average sales by region features and drop the category and region column
   * One-hot encode all the other categorical variables.
   * Standardize the numerical variables
4. Build Models and compare the results:
   * Build a Linear Regression and Random Forest on the above data
   * Set the date as index and use only sales column to build ARIMA and SARIMA model
   * Use facebook prophet and build the model

**FAQs**

**Q1. How can I analyze the sales and profit distribution by product category?**

Perform hypothesis-based EDA to identify the top 5 product categories' sales and profit shares.

**Q2. How can I determine the profit distribution by region?**

Conduct an analysis to uncover the share of profit attributed to different regions.

**Q3. Which models can I build for sales prediction?**

Use Linear Regression, Random Forest, ARIMA, SARIMA, and Facebook Prophet models to predict Walmart store sales.