**SABARI NATHAN S \*** [Github.com](https://github.com/Sabari-Nathan-S) **\***[sabarinathan.s.0905@gmail.com](mailto:sabarinathan.s.0905@gmail.com)

Walmart  
E-commerce Sales Dataset

\_\_\_

# Executive Summary

This project focuses on analyzing customer purchasing behavior from a Walmart dataset. The dataset includes demographic details of customers such as gender, age, occupation, and city category, alongside product details and purchase amounts. This analysis aims to explore patterns in customer purchases and identify key drivers behind purchase behavior, enabling Walmart to refine its marketing strategies and product offerings.

# Introduction

Walmart is a globally recognized retail giant, renowned for offering a wide range of products at competitive prices and providing a convenient shopping experience for millions of customers worldwide.

Analyzing this dataset offers valuable insights into Walmart's customer base and purchasing behavior. It reveals details about customer demographics, product preferences, and spending patterns. This comprehensive dataset is beneficial for understanding various aspects of Walmart's operations, such as marketing strategies, customer segmentation, and product demand. It can enhance strategic decision-making in areas like inventory management, targeted marketing, and customer relationship management.

### **Objective**

* To explore and identify key purchasing trends.
* To understand the relationship between customer demographics (e.g., age, gender, city) and their purchase behavior.
* To make recommendations based on data insights for better business decisions.

### **Stakeholders**

The insights will benefit Walmart’s marketing team, product managers, and data teams to make data-driven decisions regarding customer targeting and inventory management.

# Data Description

### **Dataset Source**

The dataset used for this analysis was sourced from [Walmart’s sales data](https://www.kaggle.com/datasets/devarajv88/walmart-sales-dataset), containing information on individual purchases made by customers.

### **Data Fields**

The Key columns in the dataset include:

* **User\_ID**: Unique identifier for each customer.
* **Product\_ID**: Unique identifier for each product.
* **Gender**: Gender of the customer.
* **Age**: Age group of the customer.
* **Occupation**: Customer’s occupation code.
* **City\_Category**: Category of the city where the customer resides.
* **Stay\_In\_Current\_City\_Years**: Number of years the customer has lived in the current city.
* **Marital\_Status**: Marital status of the customer (0 = Single, 1 = Married).
* **Product\_Category**: Category of the purchased product.
* **Purchase**: Purchase amount for the transaction.

# Data Preprocessing

### **Data Cleaning and Transformation**

* Dropped any **Duplicate Rows**
* Detect **Missing Values**
* **Outliers** - Identified and addressed potential outliers in purchase amounts.

# Exploratory Data Analysis (EDA)

### **Descriptive Statistics**

Create statistics for all numeric columns in the dataset to find the minimum, maximum, mean, median, standard deviation, etc.

### **Key Visualization and Insight**

#### **Age and Gender Insights**: Purchase Patterns by Product Category

* **Males** contribute to a larger share of purchases compared to females
* The **age group of 26-35** within **product category 1** has set a benchmark.
* **Graph:** A heat map shows the distribution of Product categories by Age and Gender.

#### **Product Category Trends:**

* Product category 5 is the most popular, with around 29% of users making purchases, indicating high demand in certain product lines.
* **Graph**: A bar chart showcasing the Counts of Users per product category.

#### **Gender Distribution in Data:**

* Males participate in purchases to a larger extent compared to females.
* The **male** population accounted for 24.7% and the **Female** population accounted for 75.3%.
* **Graph**: A bar chart analysed the Gender distribution in the Data

#### **Age Distribution in Data:**

* The age group of 26-35 tends to make more purchases compared to other age groups. They accounted for around 39.9% of all purchases across age groups.
* **Graph**: A bar chart analysed the Gender distribution in the Data

#### **Average Purchase Amount by Occupation:**

* Users occupied in the 17th made an average purchase of 5.1%, followed by users occupied in the 12th and 15th, each reaching 5.0%.
* **Graph**: A bar chart showcasing the Average purchase Amount by User’s Occupation.

#### **Relationship Analysis:**

* **Graph**: A Box plot shows the relation between **Gender, Age, Occupation, Stay in Current city years, Marital Status,** and **Product Category** how each factor affects Purchase Amount.

# Methodology

### **Descriptive Analytical Approach**

#### We used basic descriptive statistics and visualizations to explore customer demographics and their impact on purchase behavior. A correlation analysis was performed to identify relationships between variables such as occupation, marital status, and purchase amounts.

# Findings

* **Gender Impact**: Males tend to make higher-value purchases.
* **Age Group Trends**: The age group **26-35** is the most active in terms of purchases, while the younger age group **0-17** makes significantly fewer purchases.
* **Product Popularity**: Products in category **5** are the most frequently purchased followed by 1 and 8.
* **Occupation Trends:** Comparatively Users occupied 17 made more purchases.

