

# Data Science Assignment: eCommerce Transactions Dataset

## Overview:

You are provided with an eCommerce Transactions dataset consisting of three files: Customers.csv, Products.csv, and Transactions.csv. Your task is to perform exploratory data analysis (EDA), build predictive models, and derive actionable insights. This assignment will test your data analysis, machine learning, and business insight generation skills.

## Task 1: Exploratory Data Analysis (EDA) and Business Insights

### Business Insights

#### 1. Customer Distribution by Region

Analysing the dataset reveals that South America accounts for the highest number of customers (59), followed by Europe (50), North America (46), and Asia (45). The relatively balanced distribution across regions highlights the company's global reach but also suggests untapped growth potential in Asia and North America. Business efforts could be tailored to enhance customer acquisition strategies in underperforming regions. For example, targeted marketing campaigns or localized promotions in Asia might help to boost engagement and sign-ups.

#### 2. Outliers in Daily Sign-Ups Indicate High-Impact Events

Certain dates, such as November 11, 2024, and April 16, 2022, show unusually high customer sign-ups, with 3 sign-ups each, compared to the average daily rate. These outliers may coincide with special campaigns, promotional events, or significant external factors. Investigating these dates can help identify effective strategies or events that contributed to the spike. Replicating similar conditions in future campaigns could drive higher customer acquisition. The presence of such peaks highlights the importance of targeted efforts and their measurable impact.

#### 3. Unique Customer Records Ensure Data Integrity

The dataset confirms that all 200 entries for CustomerID and Customer Name are unique, indicating no duplication or data quality issues. This ensures the reliability of subsequent analyses and insights derived from the dataset. High data integrity allows the company to confidently base decisions on the dataset, such as designing region-specific marketing strategies or conducting customer segmentation. Regular audits and data cleaning processes should be maintained to preserve this standard of data quality.

#### 4. Consistency in Customer Acquisition Over Time

The range of sign-up dates spans from January 22, 2022, to December 28, 2024, reflecting consistent efforts in customer acquisition. This continuity suggests a well-maintained strategy and operational focus on onboarding customers regularly. However, identifying seasonal or monthly variations within this period can reveal opportunities for optimizing resource allocation. For example, if certain months consistently show low sign-ups, targeted interventions during those periods could balance the acquisition trend.

## **5. Positive Growth in Sign-Up Trends**

- In 2022, the company acquired 64 customers.
- In 2023, the number decreased to 57.
- In 2024, the number increased to 79, showing significant growth from the previous year.
- Growth between years can be calculated using percentage changes, which reflect the effectiveness of strategies like marketing or customer satisfaction initiatives.