# Exploratory Data Analysis Project: Online Shopping Behavior

## 1. Problem Statement

An e-commerce company wants to better understand its users' online shopping behavior to improve the overall shopping experience and optimize marketing efforts. The company has collected browsing and purchase activity data of users over the past three years. This Exploratory Data Analysis (EDA) project aims to identify data quality issues, clean the dataset, and uncover meaningful insights about user engagement, product preferences, and purchasing patterns.

## 2. Dataset Description

The dataset contains 550 records and the following variables:

|  |  |
| --- | --- |
| Variable | Description |
| User\_ID | Unique identifier for each user |
| Age | Age of the user (may include outliers and invalid values) |
| Gender | Gender of the user (may contain typos and inconsistent capitalization) |
| Country | Country of residence |
| Session\_Duration | Duration of website session in minutes (may contain negative or null values) |
| Pages\_Visited | Number of pages visited in the session |
| Product\_Category | Category of the product browsed or purchased (may include typos) |
| Purchase\_Intent | Whether the user showed intent to purchase (Yes/No with inconsistencies) |
| Purchase\_Amount | Amount spent during the session (should align logically with Purchase\_Intent) |
| Device\_Type | Device used to access the website (mobile, desktop, tablet, etc. with inconsistencies) |
| Signup\_Date | Date when the user signed up |
| Last\_Login | Most recent login date by the user |

## 3. Project Scope

As data analysts, our goal is to perform an end-to-end EDA on this dataset to meet the client’s expectations, which include:  
- Identifying and handling missing or inconsistent data entries  
- Standardizing categorical values (e.g., Gender, Product\_Category, Purchase\_Intent)  
- Identifying and correcting outliers in numerical columns like Age and Session\_Duration  
- Understanding the relationship between variables such as session duration and purchase intent  
- Analyzing purchase behavior by age group, country, and device type  
- Generating meaningful visualizations to support business decisions  
- Providing recommendations to the marketing and product teams based on user behavior patterns

## 4. Deliverables

The following deliverables will be provided upon completion of this project:  
- Cleaned and preprocessed version of the dataset  
- Visual reports and dashboards representing key user behavior insights  
- A summary report outlining the findings and actionable recommendations