# Exploratory Data Analysis (EDA) Report

## Introduction

This report performs an exploratory data analysis on the backpack dataset which contains details about various backpacks including their brand, material, size, number of compartments, laptop compartments, waterproofness, the style, color, weight capacity and their price. The aim of the analysis is to uncover patterns and relationships between these variables.

## Data Overview

There are two dataset, train dataset is for training the model and the test for making predictions.

### 2.1. Dataset Shape

The train dataset has **300,000** **rows** and **11 columns** and the test dataset has **200,000 rows** and 10 columns.

### 2.2 Column Descriptions

**- Brand**: The brand of the bag.  
**- Material**: The primary material used.  
**- Size**: The size of the bag.  
**- Compartments**: The number of compartments in the bag (1 to 10).  
**- Laptop Compartment**: Whether the bag has a laptop compartment(Yes/No).  
**- Waterproof**: Whether the bag is waterproof (Yes/No).  
**- Style**: The style of the bag.  
**- Color**: The color of the bag.  
**- Weight Capacity (kg)**: The maximum weight the bag can hold.  
**- Price**: The price of the bag in USD.

### 2.3 Missing Values

Before starting the analysis, I checked for any missing values:

The dataset had **no missing values** in the Compartment and Price columns.

The Brand, Material, Size, Laptop Compartments, Waterproof, Style, Color, and Weight Capacity (kg) columns contained a small number of missing entries, which were filled using different imputation methods.

## Univariate Analysis

### 3.1. Distribution of Brand

Adidas is the most common bags brand in the dataset, followed by Under Armour and Nike.

### 3.2. Distribution of Material

Polyester is the most used material for bags in this dataset, followed by leather.

### 3.3. Distribution of Size

Medium-sized bags are more popular than large and small ones. This shows that people prefer a balance between space and ease of carrying.

### 3.4. Distribution of Compartments

Bags with fewer compartments (1 to 4) are the most common with single-compartment backpacks being the most preferred. This suggests that users generally opt for simpler designs over multi-compartment options.

### 3.5. Distribution of Laptop Compartment

Bags with a laptop compartment are slightly more common than those without but the difference is not significant. This shows that having a laptop compartment is a common feature but not necessarily a strong differentiator among bags in the dataset.

### 3.6. Distribution of Waterproof

Bags that are waterproof are slightly more common than those that are not waterproof, but the difference is not significant.

### 3.7. Distribution of Style

Messenger bags are the most preferred type of backpack, followed by tote bags, regular backpacks, and other designs.

### 3.8. Distribution of Color

Pink is the most preferred color of bag compared to the other colors.

### 3.9. Distribution of Weight Capacity (kg)

The Weight Capacity distribution is platykurtic (flat peak and light tails), showing that most bags weigh in the mid-range, with fewer extreme values at both the low and high ends. This suggests a more uniform spread of weight with some outliers but not an excessive concentration of data around a central value. The average weight capacity of a bag is 18.03kg.

### 3.10. Distribution of Price

The price distribution is platykurtic (flat peak and light tails), showing that most bags are priced in the mid-range, with fewer extreme values at both the low and high ends. This suggests a more uniform spread of prices with some outliers but not an excessive concentration of data around a central value. The average price of a bag is 81.41 USD.

## Bivariate Analysis