Project Documentation: Exploratory Data Analysis using Python

# Introduction

This project focuses on analyzing a laptop dataset comprising various specifications such as processor type, RAM, storage, brand, screen size, and price. The objective is to explore the factors influencing laptop pricing and identify trends or correlations that can assist buyers or sellers in making informed decisions.

# Aim

To perform a comprehensive exploratory data analysis on laptop specifications and prices to uncover patterns, trends, and correlations that affect pricing using visualizations and statistical tools.

# Problem Statement

The electronics market is saturated with a wide range of laptops with varying specifications and price points. Consumers and retailers face difficulty in assessing the fair price of a laptop based on its features. This project aims to identify the key determinants of laptop pricing and provide actionable insights into which features drive cost.

# Project Workflow

The project workflow includes data loading, cleaning, feature engineering, univariate, bivariate, and multivariate analysis using visualizations and statistical summaries.

# Data Understanding

The dataset includes columns such as brand, processor type, RAM, storage type and size, screen size, operating system, and price. Each attribute gives critical information for analysis.

# Data Cleaning

Missing values were handled, data types were standardized, and inconsistencies were resolved. Outliers in the price column were examined and treated accordingly.

# Obtaining Derived Metrics

Derived metrics include combined storage size, categorizing RAM size (e.g., low, medium, high), and grouping processor generations for more meaningful analysis.

# Filtering Data for Analysis

Filtering was done to isolate specific laptop segments, such as those with SSD storage, certain brands, or price ranges. This helped in focused comparison and clearer insights.

# Statistical Analysis

Descriptive statistics were computed and price comparisons were made using grouped features. Correlation analysis showed strong associations between price and features like RAM, storage, and processor.

# Exploratory Data Analysis (EDA) - Univariate Analysis

Variables like price, RAM, and screen size were explored using histograms and boxplots to understand distributions and detect outliers.

# Bivariate Analysis

Relationships such as price vs RAM, price vs brand, and price vs processor type were analyzed using scatter plots, boxplots, and bar charts.

# Multivariate Analysis

Correlation heatmaps and pair plots helped reveal multicollinearity and combined effects. Grouped visualizations showed patterns across multiple categorical and continuous variables.

# Overall Insights from Analysis

High RAM and SSD storage significantly influence laptop prices. Brands and processor types also show a considerable impact. Larger screens and newer OS versions correspond with higher prices.

# Conclusion

The analysis highlights critical factors driving laptop prices. These findings can assist consumers in evaluating their purchases and help manufacturers understand competitive pricing strategies.