

Project Documentation: Exploratory Data Analysis using Python

Project Information

Title : Laptop Dataset

Name : Moviraj D

DA/DS : May-2025

Batch Number : RP-36

Online/Offline : Offline

Roll Number : B2024126094

Table of Contents

1. Introduction

This project explores and analyzes various laptop specifications to understand pricing trends and performance characteristics. It involves data cleaning, statistical analysis, and visualizations to uncover market insights.

2. Aim

To identify the key specifications influencing laptop prices and extract insights about performance, trends, and brand dynamics in the laptop industry.

3. Business Problem / Problem Statement

Consumers and businesses face difficulties in choosing laptops due to varying specifications and prices. This project aims to:

- Identify which features significantly affect pricing.
- Explore relationships among laptop specifications.
- Recommend strategies based on data-driven insights.

4. Project Workflow

1. Data Collection
2. Data Cleaning & Preprocessing

3. Exploratory Data Analysis (EDA)
4. Feature Engineering
5. Statistical Testing
6. Visualization & Insights
7. Conclusion

5. Data Understanding

The dataset includes laptop specifications like:

- Company
- TypeName
- Inches (Screen Size)
- Weight
- CPU, GPU, RAM
- Price (Target)

6. Data Cleaning

- **Missing Values:** Removed or imputed.
- **Outliers:** Detected using boxplots.
- **Inconsistent Values:** Unified categorical values (e.g., merging "Notebook"/"Ultrabook").

7. Obtaining Derived Metrics

- Extracted CPU and GPU brand.
- Extracted processor generation.
- Created binary flags like Touchscreen or IPS panel.

8. Filtering Data for Analysis

- Removed extreme outliers in price.
- Filtered brands/types with very low sample sizes.
- Removed duplicates and rows with unclear data.

9. Statistical Analysis

Used **ANOVA** to test if price varies significantly across companies.

```
python
CopyEdit
from scipy.stats import f_oneway
stat, p_value = f_oneway(*vl.groupby('Company')['Price'].apply(list))
```

10. Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) is a crucial step in data analysis that involves visually and statistically examining datasets to uncover patterns, spot anomalies, test hypotheses, and check assumptions. In the context of this laptop dataset, EDA helps understand the distribution of variables like price, RAM, and screen size, and how they relate to one another.

It includes **univariate analysis** (individual features), **bivariate analysis** (feature relationships), and **multivariate analysis** (interactions among multiple features). Tools like histograms, boxplots, scatterplots, and heatmaps are used to reveal trends, correlations, and outliers that guide deeper analysis and informed decision-making.

Univariate Analysis

- **Histograms:** For distribution of prices, RAM.
- **Boxplots:** To detect outliers
- **Countplots:** For categorical variables like TypeName.
- **Pie charts:** For brand/RAM proportion.

Syntax:

```
python
CopyEdit
sns.histplot(MN['Price'], bins=30, kde=True)
```

11. Bivariate Analysis

- **Scatterplots:** RAM vs Price, Inches vs Weight.
- **Boxplots:** GPU/CPU Brand vs Price.
- **Violin plots:** Price distribution by type or GPU.

- **Hexbin:** For density in numerical variables.

12. Multivariate Analysis

- **Heatmap:** To examine correlation.
- **FacetGrid:** For grouped scatterplots.
- **3D Plotting:** Price vs RAM vs Inches using Plotly or mpl_toolkits.

13. Overall Insights

- **RAM and CPU brand** are top price influencers.
- **Ultrabooks and Gaming laptops** are costlier.
- **Brand reputation** significantly impacts pricing.
- **Weight and Screen size** are moderately correlated.
- **Intel i7 and dedicated GPUs** increase price significantly.

14. Conclusion

The analysis provides actionable insights for consumers and sellers:

- Focus on key specs like RAM, CPU brand, GPU.
- Brands can adjust pricing strategies based on demand trends.
- Consumers can get value-for-money by comparing specs.