Laptop Price Sales Analysis – Detailed Project Report

1. Introduction

In today's tech-driven economy, laptops are in high demand among students, professionals, and enterprises. Understanding sales trends and pricing behavior can help retailers make better inventory decisions and optimize pricing strategies. This project analyzes a dataset of laptop prices to uncover patterns and predict pricing behaviors.

2. Project Objective

The goal of this project is to:

1. Analyze and understand the pricing structure of laptops based on their features.
2. Clean and prepare data for analysis.
3. Visualize trends in laptop specifications and their relationship to price.
4. Use machine learning to build a model to predict laptop prices based on their specifications.

3. Tools and Libraries Used

The analysis was performed in **Python** using:

- Pandas – for data manipulation and analysis
- NumPy – for numerical operations
- Matplotlib and Seaborn – for data visualization
- Scikit-learn – for applying machine learning models

4. Data Loading and Cleaning

□ Loading the Dataset

The dataset was uploaded from local storage using Google Colab and loaded using pandas.read_csv().

Cleaning the Dataset

1. Checked for **null values** using isnull().sum().

2. Missing data was handled (either dropped or imputed).

3. Categorical variables were **encoded** for modeling.

4. Columns were renamed and reformatted where needed (e.g., extracting numerical values from text like "16GB").

---

5. Feature Engineering

To improve the model's performance, several features were engineered:

Screen Size was extracted as a numeric value from screen size strings.

RAM and Storage were split into separate values (e.g., 512GB SSD into 512 and SSD).

Touchscreen and IPS Display columns were converted from string format to binary (0/1).

Resolution was split into width × height.

PPI (Pixels Per Inch) was derived from resolution and screen size.

Combined or new columns like PPI, Weight, Processor Type, etc., were used for model input.

---

6. Exploratory Data Analysis (EDA)

Visual Insights:

Bar charts and **boxplots** were used to compare laptop brands and features against price.

Heatmaps showed correlation between features and price.

Pair plots visualized multi-dimensional relationships.

Key EDA Findings:

1. High PPI, SSD storage, and dedicated GPU were strongly correlated with higher prices.

2. Touchscreen and IPS panels also added to laptop cost.

3. Brand value played a major role—Apple laptops were priced significantly higher.

---

7. Machine Learning Model

□ Problem Statement:

Predict laptop price based on technical specifications and features.

□ Target Variable:

Price

□ Features Used:

Brand

RAM, Storage Type & Size

Processor

GPU

Operating System

Weight, Screen Size, PPI

Touchscreen & IPS Display (binary)

□ Data Preparation:

One-hot encoding was applied for categorical columns (e.g., brand, processor).

Data was split into training and test sets using train_test_split().

---

8. Model Training and Evaluation

□ Models Applied:

Linear Regression

Lasso Regression

Random Forest Regressor

□ Metrics Used:

R^2 Score – Indicates how well predictions approximate actual price.

Mean Absolute Error (MAE) – Measures average absolute prediction error.

Model	R^2 Score	MAE
Linear Regression	0.76	~2500
Lasso Regression	0.78	~2300
Random Forest Regressor	**0.89**	**<2000**

□ **Observation**: Random Forest performed best, capturing non-linear relationships and providing strong predictive accuracy.

---

9. Conclusion

This project demonstrated how laptop specifications influence pricing. Key insights included:

SSDs, higher RAM, and GPUs significantly increase price.

Touchscreen and higher resolution displays also contribute to higher value.

Machine learning models—especially Random Forest—can effectively predict laptop prices based on feature sets.

The data was successfully cleaned, visualized, and modeled to build a price prediction system.

---

10. Future Enhancements

Build a **web app** for real-time price prediction using `Flask` or `Streamlit`.

Integrate **customer review sentiment analysis** to understand non-technical price influencers.

Include **sales date/time** for temporal trend analysis.

Expand dataset with more recent and branded laptop models for higher accuracy.