Amazon Laptop Data Analysis Report

1. Introduction

This project focuses on collecting and analyzing laptop data from Amazon.

The main goal of this project is to:

- Scrape laptop data from Amazon
- Clean the collected data
- Analyze the dataset
- Generate meaningful insights

This project helps in understanding real-world data analysis using Python.

---

2. Data Collection

The data was collected using web scraping techniques.

Tools Used:

- Python
- Requests
- BeautifulSoup

Information collected includes:

- Laptop Name
- Price
- Ratings
- Number of Reviews
- Other basic details

The raw data was saved as:

- `amazon_laptops_raw.csv`

---

3. Data Cleaning

Raw data often contains errors and missing values. So, data cleaning was performed.

Cleaning Steps:

- Removed missing values
- Removed special symbols like ₹ and commas from price
- Converted price column into numeric format
- Removed duplicate records
- Standardized column names

---

4. Data Analysis

After cleaning, the dataset was analyzed using Pandas.

Analysis Performed:

- Calculated average laptop price
- Found highest and lowest priced laptops
- Compared ratings of different laptops
- Analyzed price distribution
- Sorted laptops by price

Example Insights:

- Some laptops have high price but low ratings.
- Mid-range laptops tend to have better ratings.
- There is a wide price variation in the dataset.

---

5. Data Visualization

Matplotlib was used to create visual charts.

Charts Created:

- Bar chart for laptop prices
- Rating comparison chart
- Price distribution graph

Visualizations helped in understanding trends clearly and easily.

---

6. Key Findings

- The laptop price range varies significantly.
- Higher price does not always mean better rating.
- Popular laptops usually have more reviews.
- Most laptops fall in the mid-price category.

---

7. Conclusion

This project demonstrates:

- Web scraping techniques
- Data cleaning process
- Data analysis using Pandas
- Data visualization using Matplotlib

It provides practical experience in handling real-world data.

---

8. Future Improvements

- Add more pages for scraping
- Perform brand-wise analysis
- Create interactive dashboards
- Use advanced visualization tools like Seaborn or Plotly