**Project Title: Laptop E-Commerce Data Analysis and Dashboard**

**Objective:**

To scrape laptop product data from a legal e-commerce API, clean and preprocess it, perform data analysis using Python, store it in MySQL, and visualize the insights using Power BI.

**Step-by-Step Project Process:**

**Step 1: Web Scraping Using SerpAPI (Amazon)**

* Tool: Python, SerpAPI
* Target: Laptop product listings on Amazon.com
* Steps:
  1. Use SerpAPI to query laptops from Amazon (pages 1 to 50)
  2. Extract relevant fields:
     + Product Name
     + Price
     + Rating
     + Number of Reviews
     + ASIN
     + Link
  3. Add logic to parse extra info like:
     + Brand (from title using keyword matching)
     + RAM Size
     + Processor Type
  4. Store all records in a CSV file

**Step 2: Data Cleaning & Preprocessing (Python)**

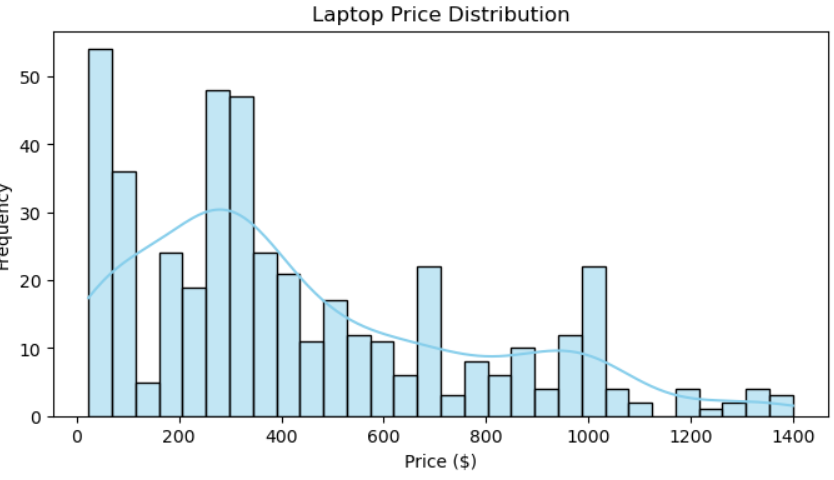
* Tool: Python (Pandas, NumPy, Matplotlib, Seaborn, SciPy)
* Tasks:
  1. Load the scraped CSV
  2. Handle missing data
     + Replace NaNs with appropriate values
     + Set unknown brands as "Unknown"
  3. Remove duplicates
  4. Convert Price and Rating columns to numeric types
  5. Extract brand using known brand list
  6. Save cleaned data to a new CSV file



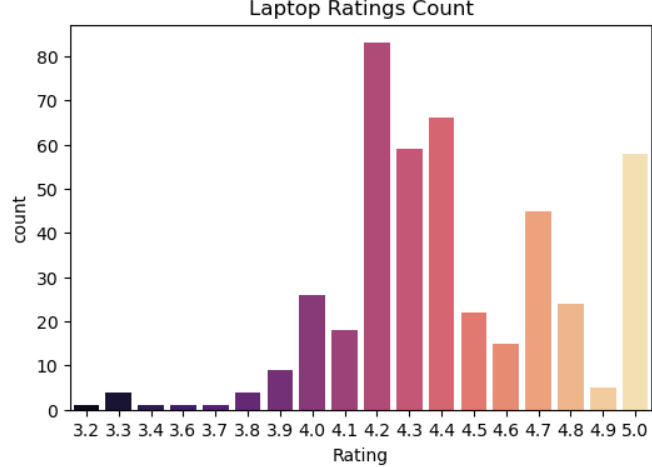
**Step 3: Exploratory Data Analysis (EDA) & Statistical Analysis**

Insights Explored:

**Price distribution**



**Rating distribution**



**Price vs Rating correlation**



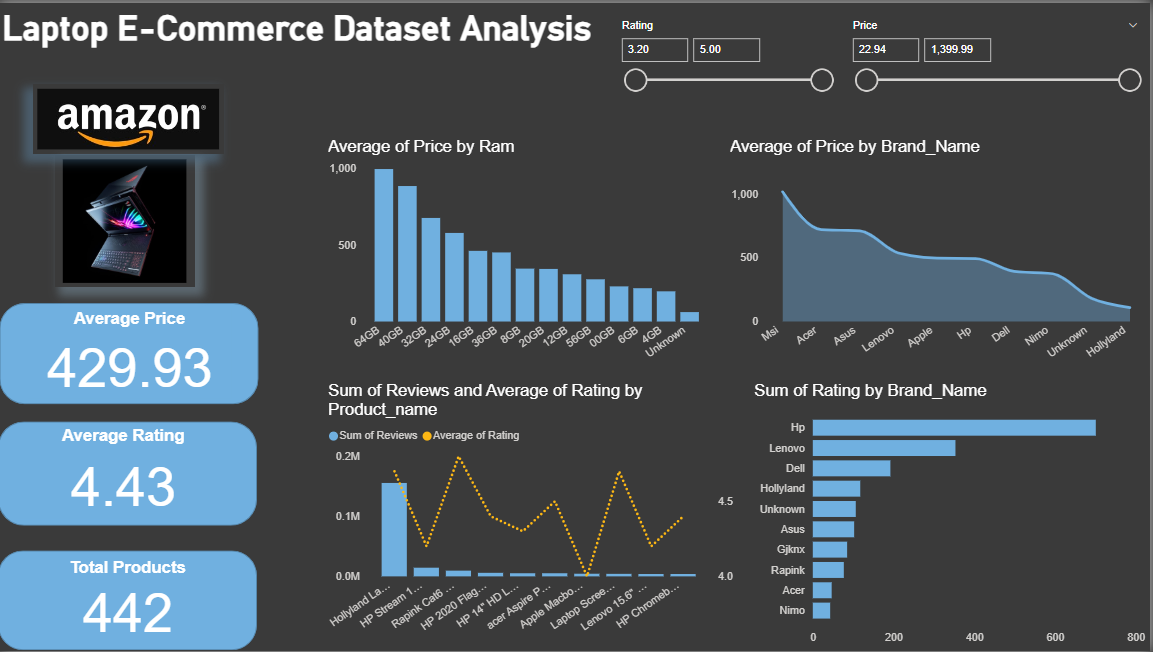
* + Hypothesis testing (e.g., do branded laptops have higher average ratings?)

**Step 4: Storing Cleaned Data in MySQL**

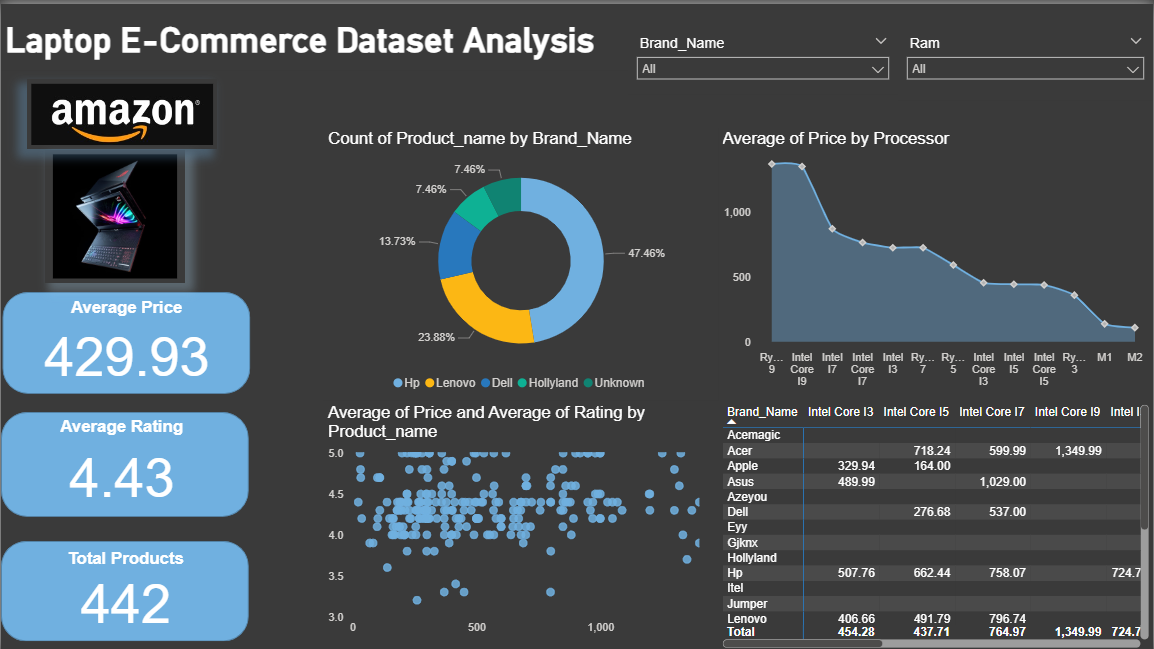
* Tool: MySQL, MySQL Connector for Python
* Tasks:
  1. Design MySQL table schema
  2. Connect to MySQL server using Python
  3. Created database and table
  4. Import cleaned CSV data into MySQL using insert queries
  5. Test successful insertion with SELECT queries

**Step 5: Power BI Dashboard Creation**

* Tool: Power BI Desktop
* Steps:
  1. Connect Power BI to MySQL database
  2. Import laptops table
  3. Transform data:
     + Replace inconsistent brand entries
     + Clean malformed text columns
  4. Create Visualizations:



* **Price vs RAM**: Higher RAM = Higher Average Price
* **Brand-wise Avg Price**: Brands like MSI, Asus lead in pricing
* **Sum of Ratings by Brand**: HP and Lenovo dominate in number of reviews and ratings
* **Top Reviewed Products**: Some individual models gain massive review counts (e.g., HP Stream)



**Visual 2**:

* **Pie Chart**: HP, Lenovo, Dell are the top 3 brands by count
* **Scatter Plot**: Shows price-rating clustering per product
* **Line Graph by Processor**: Intel Core i9 and Ryzen 9 have the highest average prices

**Skills Demonstrated:**

* Web Scraping
* Data Cleaning & Transformation
* EDA and Visualization
* Statistical Reasoning
* SQL Database Integration
* Power BI Dashboard Design
* Reporting and Presentation