Amazon Products Data Analysis Project

Overview

This project is designed to scrape product data from Amazon, preprocess it, and visualize valuable insights using Python. It consists of two main scripts:

1. `main.py` : Handles web scraping to collect product details (e.g., title, price, rating, reviews, availability) from Amazon.
2. `visualisation.py`: Cleans and preprocesses the scraped data, then generates various visualizations to uncover insights.

Features

Web Scraping (`main.py`)

The scraper leverages `BeautifulSoup` and `requests` libraries to gather key product information, such as:

- **Product Title**
- **Price**
- **Rating**
- **Number of Reviews**
- **Availability Status**

The collected data is stored in a CSV file (`amazon_data.csv`) for further visualization and analysis.

Data Cleaning & Processing (`visualisation.py`)

Once the data is scraped, it undergoes several preprocessing steps:

1. Data type conversion (e.g., converting ratings and reviews to numeric).
2. Cleaning messy text data (e.g., simplifying availability statuses).
3. Handling missing or inconsistent data by filling or converting it properly.

Data Visualizations

The project generates visualizations using `Matplotlib` and `Seaborn` to uncover trends and insights based on the data.

Rating Distribution

A histogram displaying how product ratings are distributed across the dataset.

Top 5 Products by Reviews

A bar chart showing the products with the highest number of reviews.

Stock Availability Distribution

A pie chart summarizing the availability of products (e.g., "In Stock" vs. "Limited Stock").

Reviews vs. Ratings

A scatter plot showing the relationship between product reviews and ratings, categorized by availability.