

Amazon Web Scraping and Data Analysis

Project Overview

This project involves web scraping product data from Amazon's laptop listings using Python's BeautifulSoup and Requests libraries. The primary goal was to extract product details—such as titles, prices, ratings, reviews, availability, and other attributes—and perform a comprehensive analysis on the data. Through data visualization and sentiment analysis, this project aims to provide insights into pricing trends, customer preferences, and product sentiment.

Data Extraction Process

The scraping module gathers product data by navigating the Amazon search results page, extracting:

- **Title:** Scraped from the product's title section.
- **Price:** Retrieved from elements indicating regular or deal prices.
- **Rating and Reviews:** Collected from the product's rating badge and review count.
- **Availability:** Checks if the product is in stock.
- **Additional Details:** Includes brand, seller, and image URLs.

Data Processing and Cleaning

Once the data was collected:

- Missing and irregular data points were addressed.
- Prices were converted to numeric format.
- Reviews were processed to extract counts, while ratings were binned for range-based categorization.

Data Analysis and Visualization

Using Pandas, Matplotlib, and Seaborn, the data analysis focused on:

- **Price Distribution:** Histogram analysis revealed price ranges and category segmentation.
- **Sentiment Analysis:** TextBlob was applied to review text, categorizing customer sentiment as positive, negative, or neutral.
- **Price vs. Rating:** Scatter plots analyzed relationships between price and ratings.
- **Top-Rated and Cheapest Products:** Extracted and visualized top products based on rating and pricing criteria.

Key Insights and Impact

- **Customer Preferences:** Identified high-rated and popular laptops.
- **Pricing Trends:** Revealed average prices in various categories (budget, mid-range, premium).
- **Stock Analysis:** Highlighted which products are frequently available or limited in stock, offering insights into demand.