Amazon Product Scraper and Analyzer
A Python-Based Web Scraping and Data Analysis Tool

Karthick S
Team 1 (Team
Member) Cybernaut
intern

Abstract:

With the rapid growth of e-commerce, consumers and businesses increasingly rely on timely product information for decision-making. The **Amazon Product Scraper and Analyzer** is a Python-based tool that automatically scrapes product details from Amazon India using the ZenRows API and BeautifulSoup.

The system collects essential attributes such as product title, price, rating, description, ASIN, and image URLs from multiple search result pages. Extracted data is saved in CSV, Excel, and JSON formats, enabling historical tracking, comparative analysis, and visual insights. Optional filters allow focusing on products within specific price ranges or minimum ratings.

Additionally, the tool generates visualizations such as rating distribution charts and word clouds from product descriptions, supporting market research, competitive analysis, and academic study. This system is designed for analysts, researchers, and developers who need an automated, scalable solution for product monitoring.

Introduction:

E-commerce platforms like Amazon host thousands of products across numerous categories, making manual monitoring time-consuming and error-prone. Businesses, analysts, and consumers require automated solutions to extract accurate product data, compare offerings, and identify trends in pricing and customer feedback.

Manual scraping is not scalable, while APIs often have rate limits or incomplete coverage. Web scraping enables direct access to live product information for flexible analysis.

The **Amazon Product Scraper and Analyzer** leverages Python to automate data extraction, clean and structure the data, and provide visual insights. It integrates multi-threaded scraping, filtering, and visualization to offer a complete solution for product data analysis.

Existing Methods:

Currently, e-commerce analysts and consumers track Amazon products using three main approaches:

1. **Manual Browsing**
   o Users check product listings manually.
   o **Limitations:** Slow, error-prone, hard to track multiple products or trends.
2. **API-Based Retrieval**
   o Some services or platforms provide structured product data via APIs.
   o **Limitations:** Rate limits, subscription requirements, incomplete metadata, and endpoint changes.
3. **Third-Party Tools**
   o Tools like Keepa or CamelCamelCamel provide dashboards and historical data.
   o **Limitations:** Limited customization, export restrictions, no direct access to full product pages.

Limitations of Existing Methods

- No fully automated, large-scale logging of product data.
- Heavy dependency on APIs or third-party platforms with restrictions.
- Limited filtering, alerting, or comparative analysis.

Overall Gaps

- Manual methods cannot track large numbers of products efficiently.
- API-based solutions often restrict flexibility or access to detailed product content.
- Third-party tools may not allow raw data export for custom analysis.

Our solution — the Amazon Product Scraper and Analyzer — uses a reliable API (ZenRows) to access live pages but adds **automated multi-page scraping, filtering, data cleaning, and visualizations**, providing a flexible, self-contained system.

Proposed Solution:

Python-Based Web Scraping System

The **Amazon Product Scraper and Analyzer** addresses the limitations above with a multi-threaded, automated scraping system that integrates data cleaning, filtering, and visualization.

Key Features:

1. **Automated Multi-Page Scraping**
   o Extracts title, price, rating, description, ASIN, product link, and image URL.
2. **Dynamic Filtering**
   o Filter by maximum price or minimum rating.
3. **Multi-Format Storage**
   o Save data in CSV, Excel, and JSON formats.
4. **Data Visualization**
   o Rating distribution histograms and word clouds from product descriptions.

5. **Concurrent Product Page Fetching**
    o Multi-threaded fetching to improve efficiency.
6. **Cleaned and Ready-to-Use Data**
    o Removes unwanted characters, truncates long text, standardizes attributes.

Advantages:

- Fully automated, no dependency on restrictive APIs.
- Flexible filtering and visualization options.
- Scalable for multiple products and pages.
- Supports comparative analysis and historical tracking.

Technologies To Be Used:

Programming Language: Python

Libraries:

- `requests` – HTTP requests via ZenRows API.
- `BeautifulSoup` – HTML parsing.
- `pandas` – Data handling and exporting.
- `matplotlib` – Visualizations (histograms).
- `wordcloud` – Word cloud generation.
- `concurrent.futures` – Multi-threading.
- `re` – Text cleaning and regex.
- `dotenv` – API key management.

Data Storage: CSV, Excel, JSON

Execution: Console-based Python script; can be scheduled for repeated runs.

Methods:

1. **Initialization**
    o Load API key and user inputs (product search term, number of pages, optional filters).
2. **Page Access and Scraping**
    o Construct Amazon search URLs.
    o Fetch pages using ZenRows API to bypass anti-scraping protections.
    o Parse product cards with BeautifulSoup.
3. **Product Details Extraction**
    o Extract title, price, rating, ASIN, product link, image URL, and description.
    o Fetch individual product pages concurrently for detailed descriptions.
4. **Data Cleaning and Processing**
    o Remove unwanted characters from descriptions.
    o Convert prices and ratings to numeric formats.
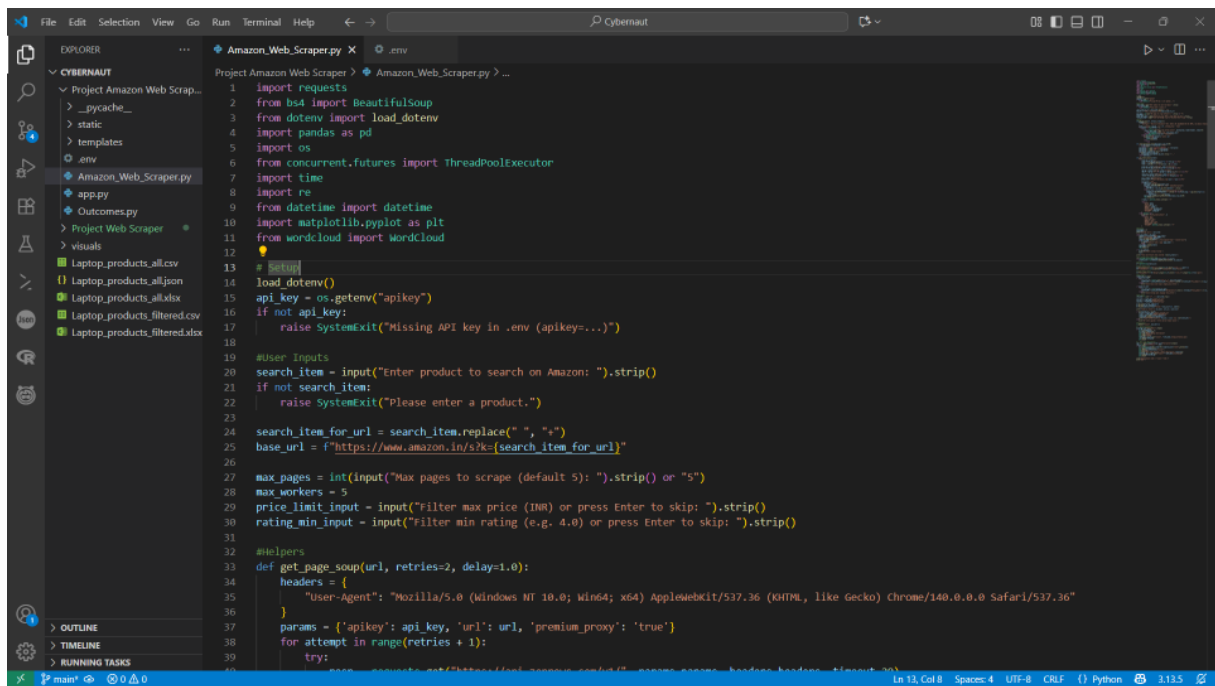   o Apply optional filters for price and rating.

5. **Data Storage**
    o   Export full data to CSV, Excel, and JSON.
    o   Export filtered data if applicable.
6. **Visualization**
    o   Generate rating distribution histogram.
    o   Generate word cloud from product descriptions.
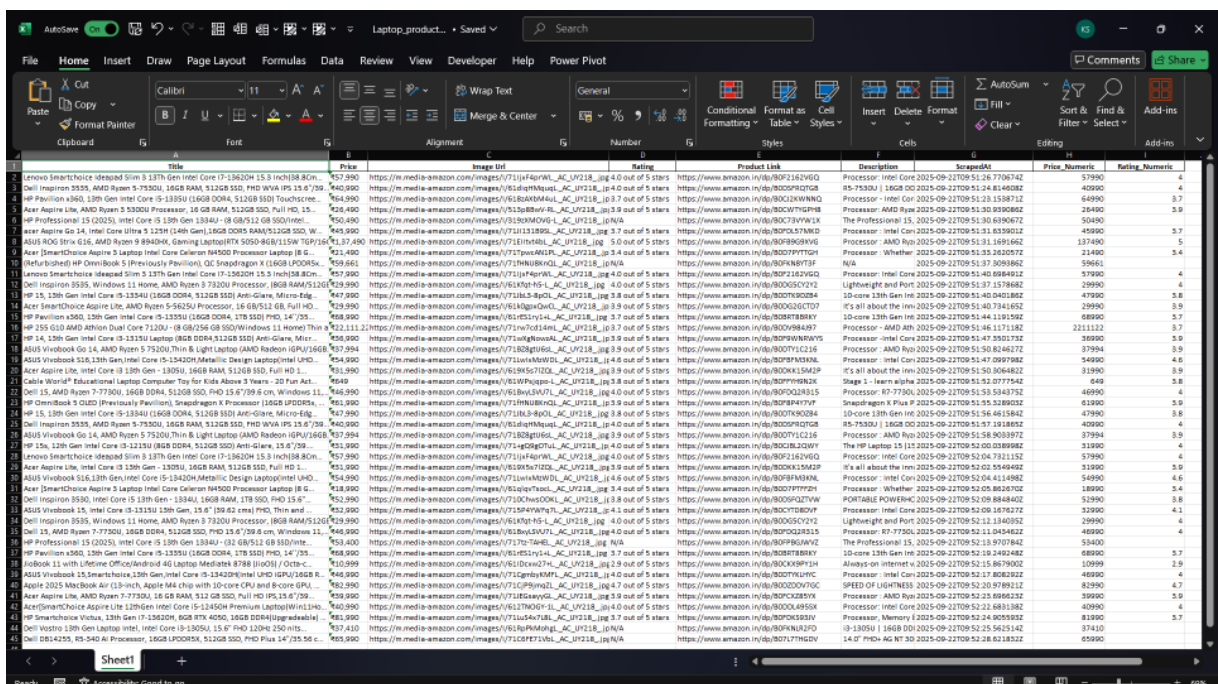   o   Save visuals in `visuals/` folder.

Screenshot Of Code Snippets:


```python
import requests
from bs4 import BeautifulSoup
from dotenv import load_dotenv
import pandas as pd
import os
from concurrent.futures import ThreadPoolExecutor
import time
import re
from datetime import datetime
import matplotlib.pyplot as plt
from wordcloud import WordCloud

# Setup
load_dotenv()
api_key = os.getenv("apikey")
if not api_key:
    raise SystemExit("Missing API key in .env (apikey=...)")

#User Inputs
search_item = input("Enter product to search on Amazon: ").strip()
if not search_item:
    raise SystemExit("Please enter a product.")

search_item_for_url = search_item.replace(" ", "+")
base_url = f"https://www.amazon.in/s?k={search_item_for_url}"

max_pages = int(input("Max pages to scrape (default 5): ").strip() or "5")
max_workers = 5
price_limit_input = input("Filter max price (INR) or press Enter to skip: ").strip()
rating_min_input = input("Filter min rating (e.g. 4.0) or press Enter to skip: ").strip()

#Helpers
def get_page_soup(url, retries=2, delay=1.0):
    headers = {
        "User-Agent": "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/140.0.0.0 Safari/537.36"
    }
    params = {'apikey': api_key, 'url': url, 'premium_proxy': 'true'}
    for attempt in range(retries + 1):
        try:
```

Output:

In xls, csv, json and
Visualization Rating Distribution and WordCloud

```json
{
    "Title":"Lenovo Smartchoice Ideapad Slim 3 13Th Gen Intel Core I7-13620H 15.3 Inch(38.8Cm...",
    "Price":"₹57,990",
    "Image Url":"https:\/\/m.media-amazon.com\/images\/I\/71IjxF4prWL._AC_UY218_.jpg",
    "Rating":"4.0 out of 5 stars",
    "Product Link":"https:\/\/www.amazon.in\/dp\/B0F2162VGQ",
    "Description":"Processor: Intel Core I7-13620H | Speed: 2.4 GHz (Base) - 4.9 GHz (Max), E-core 1.8 \/ 3.6GHz | 10 Cores | 16 Threads | 24MB ...",
    "ScrapedAt":"2025-09-22T09:51:26.7706742Z",
    "Price_Numeric":57990,
    "Rating_Numeric":4.0
},
{
    "Title":"Dell Inspiron 3535, AMD Ryzen 5-7530U, 16GB RAM, 512GB SSD, FHD WVA IPS 15.6\"\/39...",
    "Price":"₹40,990",
    "Image Url":"https:\/\/m.media-amazon.com\/images\/I\/61diqHMquqL._AC_UY218_.jpg",
    "Rating":"4.0 out of 5 stars",
    "Product Link":"https:\/\/www.amazon.in\/dp\/B0DSFRQTGB",
    "Description":"R5-7530U | 16GB DDR4 | 512GB SSD | Win 11 + Office H&S 2024 | Radeon Graphics | 15.6\" NT FHD 120Hz WVA IPS AG 250 nit Narrow...",
    "ScrapedAt":"2025-09-22T09:51:24.8146082Z",
    "Price_Numeric":40990,
    "Rating_Numeric":4.0
},
{
    "Title":"HP Pavilion x360, 13th Gen Intel Core i5-1335U (16GB DDR4, 512GB SSD) Touchscree...",
    "Price":"₹64,990",
    "Image Url":"https:\/\/m.media-amazon.com\/images\/I\/61B2AXbM4uL._AC_UY218_.jpg",
    "Rating":"3.7 out of 5 stars",
    "Product Link":"https:\/\/www.amazon.in\/dp\/B0CJ2KWNNQ",
    "Description":"Processor - Intel Core i5-1335U (up to 4.6 GHz with Intel Turbo Boost Technology(2g), 12 MB L3 cache, 10 cores, 12 threads) ;",
    "ScrapedAt":"2025-09-22T09:51:23.153871Z",
    "Price_Numeric":64990,
    "Rating_Numeric":3.7
},
{
    "Title":"Acer Aspire Lite, AMD Ryzen 3 5300U Processor, 16 GB RAM, 512GB SSD, Full HD, 15...",
    "Price":"₹26,490",
    "Image Url":"https:\/\/m.media-amazon.com\/images\/I\/513p8BwV-RL._AC_UY218_.jpg",
    "Rating":"3.9 out of 5 stars",
```

Title,Price,Image Url,Rating,Product Link,Description,ScrapedAt,Price_Numeric,Rating_Numeric
Lenovo Smartchoice Ideapad Slim 3 13Th Gen Intel Core I7-13620H 15.3 Inch(38.8Cm...,"₹57,990",https://m.media-amazon.com/images/I/71IjxF4prWL._A...
"Dell Inspiron 3535, AMD Ryzen 5-7530U, 16GB RAM, 512GB SSD, FHD WVA IPS 15.6""/39...","₹40,990",https://m.media-amazon.com/images/I/61diqHMquqL...
"HP Pavilion x360, 13th Gen Intel Core i5-1335U (16GB DDR4, 512GB SSD) Touchscree...","₹64,990",https://m.media-amazon.com/images/I/61B2AXbM4uL...
"Acer Aspire Lite, AMD Ryzen 3 5300U Processor, 16 GB RAM, 512GB SSD, Full HD, 15...","₹26,490",https://m.media-amazon.com/images/I/513p8BwV-RL...
"HP Professional 15, 13th Gen Intel Core i3 1334U - (8 GB/512 GB SSD/Intel...","₹50,490",https://m.media-amazon.com/images/I/319zXMOVG-L...
"acer Aspire Go 14, Intel Core Ultra 5 125H (14th Gen),16GB DDR5 RAM/512GB SSD, W...","₹45,990",https://m.media-amazon.com/images/I/713JI131B95L...
"ASUS ROG Strix G16, AMD Ryzen 9 8940HX, Gaming Laptop(RTX 5050-8GB/115W TGP/16GB...","₹1,37,490",https://m.media-amazon.com/images/I/71ElIltvt4b...
Acer [SmartChoice Aspire 3 Laptop Intel C   N4500 Processor Laptop (8 G...,"₹21,490",https://m.media-amazon.com/images/I/71pwcAN1PL._A...
"(Refurbished) HP OmniBook 5 (Previously Pavilion), QC Snapdragon X (16GB LPDDR5x...","₹59,661",https://m.media-amazon.com/images/I/71fHNU8KnQL...
Lenovo Smartchoice Ideapad Slim 3 13Th Gen Intel Core I7-13620H 15.3 Inch(38.8Cm...,"₹57,990",https://m.media-amazon.com/images/I/71IjxF4prWL._A...
"Dell Inspiron 3535, Windows 11 Home, AMD Ryzen 3 7320U Processor, (8GB RAM/512GB...","₹29,990",https://m.media-amazon.com/images/I/61Kfqt-h5-L...
"HP 15, 13th Gen Intel Core i5-1334U (16GB DDR4, 512GB SSD) Anti-Glare, Micro-Edg...","₹47,990",https://m.media-amazon.com/images/I/71JbL3-8pOL...
"Acer SmartChoice Aspire Lite, AMD Ryzen 5-5625U Processor, 16 GB/512 GB, Full HD...","₹29,990",https://m.media-amazon.com/images/I/61k0gpxQwCL...
"HP Pavilion x360, 13th Gen Intel Core i5-1335U (16GB DDR4, 1TB SSD) FHD, 14''/35...","₹68,990",https://m.media-amazon.com/images/I/61rE51ry1+L...
HP 255 G10 AMD Athlon Dual Core 7120U - (8 GB/256 GB SSD/Windows 11 Home) Thin a...,"₹22,111.22",https://m.media-amazon.com/images/I/71rw7cd14mL...
"HP 14, 13th Gen Intel Core i3-1315U Laptop (8GB DDR4,512GB SSD) Anti-Glare, Micr...","₹36,990",https://m.media-amazon.com/images/I/71wXgNow2AL...
"ASUS Vivobook Go 14, AMD Ryzen 5 7520U,Thin & Light Laptop (AMD Radeon iGPU/16GB...","₹37,994",https://m.media-amazon.com/images/I/71B28gtU6sL...
"ASUS Vivobook S16,13th Gen,Intel Core i5-13420H,Metallic Design Laptop(Intel UHD...","₹54,990",https://m.media-amazon.com/images/I/71LwIxM4WDL...
"Acer Aspire Lite, Intel Core i3 13th Gen - 1305U, 16GB RAM, 512GB SSD, Full HD 1...","₹31,990",https://m.media-amazon.com/images/I/619X5s7lZQL...
Cable World® Educational Laptop Computer Toy for Kids Above 3 Years - 20 Fun Act...,₹649,https://m.media-amazon.com/images/I/61WPsjqpo-L._AC_UY2...
"Dell 15, AMD Ryzen 7-7730U, 16GB DDR4, 512GB SSD, FHD 15.6""/39.6 cm, Windows 11,...","₹46,990",https://m.media-amazon.com/images/I/618xyLSVU7L...
"HP OmniBook 5 OLED (Previously Pavilion), Snapdragon X Processor (16GB LPDDR5x, ...","₹61,990",https://m.media-amazon.com/images/I/71fHNU8KnQL...
"HP 15, 13th Gen Intel Core i5-1334U (16GB DDR4, 512GB SSD) Anti-Glare, Micro-Edg...","₹47,990",https://m.media-amazon.com/images/I/71JbL3-8pOL...
"Dell Inspiron 3535, AMD Ryzen 5-7530U, 16GB RAM, 512GB SSD, FHD WVA IPS 15.6""/39...","₹40,990",https://m.media-amazon.com/images/I/61diqHMquqL...
"ASUS Vivobook Go 14, AMD Ryzen 5 7520U,Thin & Light Laptop (AMD Radeon iGPU/16GB...","₹37,994",https://m.media-amazon.com/images/I/71B28gtU6sL...
"HP 15s, 12th Gen Intel Core i3-1215U (8GB DDR4, 512GB SSD) Anti-Glare, 15.6""/39...","₹31,990",https://m.media-amazon.com/images/I/71+gQ9gQTuL...
Lenovo Smartchoice Ideapad Slim 3 13Th Gen Intel Core I7-13620H 15.3 Inch(38.8Cm...,"₹57,990",https://m.media-amazon.com/images/I/71IjxF4prWL._A...
"Acer Aspire Lite, Intel Core i3 13th Gen - 1305U, 16GB RAM, 512GB SSD, Full HD 1...","₹31,990",https://m.media-amazon.com/images/I/619X5s7lZQL...
"ASUS Vivobook S16,13th Gen,Intel Core i5-13420H,Metallic Design Laptop(Intel UHD...","₹54,990",https://m.media-amazon.com/images/I/71LwIxM4WDL...
Acer [SmartChoice Aspire 3 Laptop Intel Core Celeron N4500 Processor Laptop (8 G...,"₹18,990",https://m.media-amazon.com/images/I/61qlqvTsocL._A...
"Dell Inspiron 3530, Intel Core 15 13th Gen - 1334U, 16GB RAM, 1TB SSD, Windows 1...","₹52,990",https://m.media-amazon.com/images/I/710ChwsOOKL...
"ASUS Vivobook 15, Intel Core i3-1315U 13th Gen, 15.6"" (39.62 cms) FHD, Thin and ...","₹32,990",https://m.media-amazon.com/images/I/715P4YWFq7L...
"Dell Inspiron 3535, Windows 11 Home, AMD Ryzen 3 7320U Processor, (8GB RAM/512GB...","₹29,990",https://m.media-amazon.com/images/I/61Kfqt-h5-L...
"Dell 15, AMD Ryzen 7-7730U, 16GB DDR4, 512GB SSD, FHD 15.6""/39.6 cm, Windows 11,...","₹46,990",https://m.media-amazon.com/images/I/618xyLSVU7L...
"HP Professional 15 (2025), Intel Core i5 13th Gen 1334U - (32 GB/512 GB SSD/Inte...","₹53,400",https://m.media-amazon.com/images/I/717tz-TAHEL...
"HP Pavilion x360, 13th Gen Intel Core i5-1335U (16GB DDR4, 1TB SSD) FHD, 14''/35...","₹68,990",https://m.media-amazon.com/images/I/61rE51ry1+L...
JioBook 11 with Lifetime Office/Android 4G Laptop Mediatek 8788 (JioOS) / Octa-c...,"₹10,999",https://m.media-amazon.com/images/I/61IDcxw27+L._A...
"ASUS Vivobook 15,Smartchoice,13th Gen,Intel Core i5-13420H(Intel UHD iGPU/16GB R...","₹46,990",https://m.media-amazon.com/images/I/71GpmbyKMFL...

Product Ratings Distribution

Most Common Keywords in Product Descriptions

Future Enhancements:

1. **GUI Interface**
   o Use Tkinter, PyQt, or Flask for user-friendly interaction.
2. **Real-Time Dashboard**
   o Display live product trends with Plotly Dash or Power BI.
3. **Multiple Marketplaces**
   o Extend scraping to Flipkart, Snapdeal, etc., for comparative analysis.
4. **Automated Alerts**
   o Notify users when products meet price or rating criteria.
5. **Database Storage**
   o Store data in SQL/NoSQL for scalability and historical queries.
6. **Machine Learning Insights**
   o Predict price trends, popular keywords, or product performance.
7. **Cloud Integration**
   o Deploy on AWS/Azure/GCP for continuous monitoring.

Conclusion:

The **Amazon Product Scraper and Analyzer** demonstrates a Python-based solution for automating product data extraction from Amazon India. By storing structured data in multiple formats and generating visual insights, it enables comparative analysis, market research, and academic study.

With filters, multi-threaded scraping, and visualization capabilities, this tool is suitable for analysts, researchers, and businesses. Future upgrades such as GUI, dashboards, and predictive analytics will further enhance its value as a comprehensive e-commerce analysis platform.

References:

1. Amazon India – Product Listings: https://www.amazon.in
2. ZenRows API Documentation: https://www.zenrows.com
3. BeautifulSoup Documentation: https://www.crummy.com/software/BeautifulSoup/
4. pandas Documentation: https://pandas.pydata.org/docs/
5. matplotlib Documentation: https://matplotlib.org/stable/contents.html
6. wordcloud Documentation: https://amueller.github.io/word_cloud/