Web Data Scraping and Analysis Using Excel

Author: Aarya Khaire
Project Type: Individual
Tools Used: Microsoft Excel (Power Query, Pivot Tables, Charts)

---

1. Project Overview

This project demonstrates the use of Microsoft Excel for extracting, cleaning, analyzing, and visualizing web data. Product price data was scraped from an e-commerce practice website using Excel's Power Query feature. The collected data was transformed into a structured format and analyzed to identify pricing trends. The project highlights practical data analytics skills using Excel without the use of programming.

---

2. Objectives

- To scrape structured data from a website using Excel
- To clean and transform raw web data
- To analyze product price trends using pivot tables
- To visualize insights using charts

---

3. Data Source

- **Website:** BooksToScrape (practice e-commerce website)
- **Data Fields:** Product Name, Price, Availability

---

4. Methodology

4.1 Data Extraction

Web data was imported into Excel using the **Power Query** feature through the *Get Data → From Web* option. Relevant tables containing product information were selected and loaded into the Power Query Editor for further processing.

4.2 Data Cleaning and Transformation

The extracted data required cleaning before analysis. Currency symbols were removed from the price column, unnecessary text was extracted from availability data, and irrelevant columns were removed. Column names were standardized to improve readability and usability.

4.3 Data Analysis

A pivot table was created to calculate the **average price of products**, enabling comparison across different items and identification of pricing patterns.

4.4 Data Visualization

A column chart was generated from the pivot table to visually represent price variations and trends across products.

---

5. Results and Analysis

The analysis revealed noticeable variation in product prices across different items. While some products were priced higher than average, most products fell within a moderate price range. The visual representation using a column chart made it easier to compare prices and identify trends effectively.

---

6. Skills Demonstrated

- Web data extraction using Excel Power Query
- Data cleaning and transformation
- Pivot table analysis
- Data visualization using charts
- Analytical and problem-solving skills

---

7. Outcome

An end-to-end Excel-based data analytics workflow was successfully implemented, covering data scraping, cleaning, analysis, and visualization. This project demonstrates the practical application of Excel for real-world data analytics tasks and is suitable for entry-level data analyst roles and internships.

---

8. Future Enhancements

- Automate data refresh using Power Query
- Extend analysis to multiple product categories
- Build an interactive Excel dashboard using slicers