

Web Scraping – Flipkart Mobiles – Android

Problem Statement: Scrape Flipkart website for Android Mobiles Product Data and Create Interface to Showcase Mobiles Based on Feature Requirement

Tools & Technologies Used:

- **Programming Language:** Python
- **Libraries & Packages:** 1. Requests
2. BeautifulSoup from bs4
3. Pandas
4. Matplotlib
5. Seaborn
- **Python IDE:** Google Colab
- **Data Cleaning:** 1. Excel - Power Query Editor
2. MS Power BI – Power Query Editor
- **Business Intelligence:** Microsoft Power BI Desktop

Approach:

Step1: Open Flipkart.com website and search for mobile category. Once result page arrives, filter the data for android mobiles by selection the Operating System as “Android” in filters available and copy the link.

Link : [Mobiles- Buy Products Online at Best Price in India - All Categories | Flipkart.com](https://www.flipkart.com/mobiles-buy-products-online-at-best-price-in-india-all-categories)

Step2: Open Google Colab (IDE) and import required libraries (requests, BeautifulSoup) to scrape data from website(html) and pandas for Data Manipulation.

Step3: As there are 223 pages in flipkart website, we must loop the program to parse the data and append the appropriate features to pre-created empty column lists using ‘html parser’.

Step4: The Data extracted was converted into Data Frames for data manipulation and visualization purposes.

Step5: Data Cleaning / Transformation – Data Can be transformed and cleaned using Pandas, for easy process, the Data Frame data was exported to csv file and loaded in Excel - Power Query Editor. Below are brief of data transformation steps performed:

1. Split and merging columns
2. Trimming
3. Adding Conditional Columns
4. Changing data types
5. Replacing and filling up values.

Step6: The Transformed data was again imported into Colab for Exploratory Data Analysis. Histograms and Bar plots were used to analyze the distribution of features.

Step7: To get complete insights on data, the data from csv was Extracted, Transformed and Loaded (ETL) in Microsoft Power BI Desktop tool.

The data was for further transformed by creating additional columns, groups, measures using Power Query Editor in Power BI.

Step8: The report was built in a way to understand the availability of android mobiles in Market and the average prices for various features available and finally interface to showcase the available models for specific requirements using bar chart, line chart, slicers and visual filters.

Example: 1. Mid-Range(15K-35K) Mobile with 8Gb Ram and 128Gb internal storage.

2. Premium(>35K) Mobiles with Best ratings

3. Mobile within Budget(<15K) with Full-HD Display and long-lasting battery etc.,

Step9: The report was then published to PowerBi.Com for additional purposes like dashboarding, sharing to colleagues with access or embedding the report into a scanner code (Need Access).