**Research Report: WEB SCRAPING PYTHON**

**---**By Vanshika Srivastava

Jims sector-3

**Introduction**

**Purpose of Python Web Scraping:**

Web scraping is the process of automatically extracting data from websites using a program or script. In Python, web scraping is a powerful tool for collecting large amounts of data from various online sources, which can be crucial for research purposes. It allows researchers to gather information that is not readily available in structured formats like APIs or databases.

**Libraries Used:**

Several Python libraries are popular for web scraping due to their ease of use and functionality:

* BeautifulSoup: A library used to parse HTML and XML documents. It helps in navigating, searching, and modifying the parse tree, making it easier to extract data.
* Requests: A simple and elegant HTTP library for Python, used to send HTTP requests to a website and receive the response, which can then be processed.
* Scrapy: A more advanced framework for web scraping that provides tools for large-scale scraping projects. It allows you to define custom spiders for crawling and scraping websites efficiently.
* Selenium: A tool that automates web browsers, enabling scraping from dynamic websites that require interaction (like clicking buttons or filling out forms).

**System Overview**

**System Architecture:**

The system is composed of several key components that work together to achieve the goal of extracting and processing data from Flipkart.

**Functional Components:**

* Web Scraper (Core Component)

 **HTML Request Handler**: Uses the Requests library to send HTTP requests to the target URLs on Flipkart and retrieve HTML content.

 **HTML Parser**: Utilizes BeautifulSoup to parse the HTML content, extract relevant data fields (product details, prices, ratings, reviews), and navigate through the DOM tree.

* Data Storage Module

**File Storage**: Stores the cleaned and processed data in flat files like CSV or JSON.

**System Requirements**

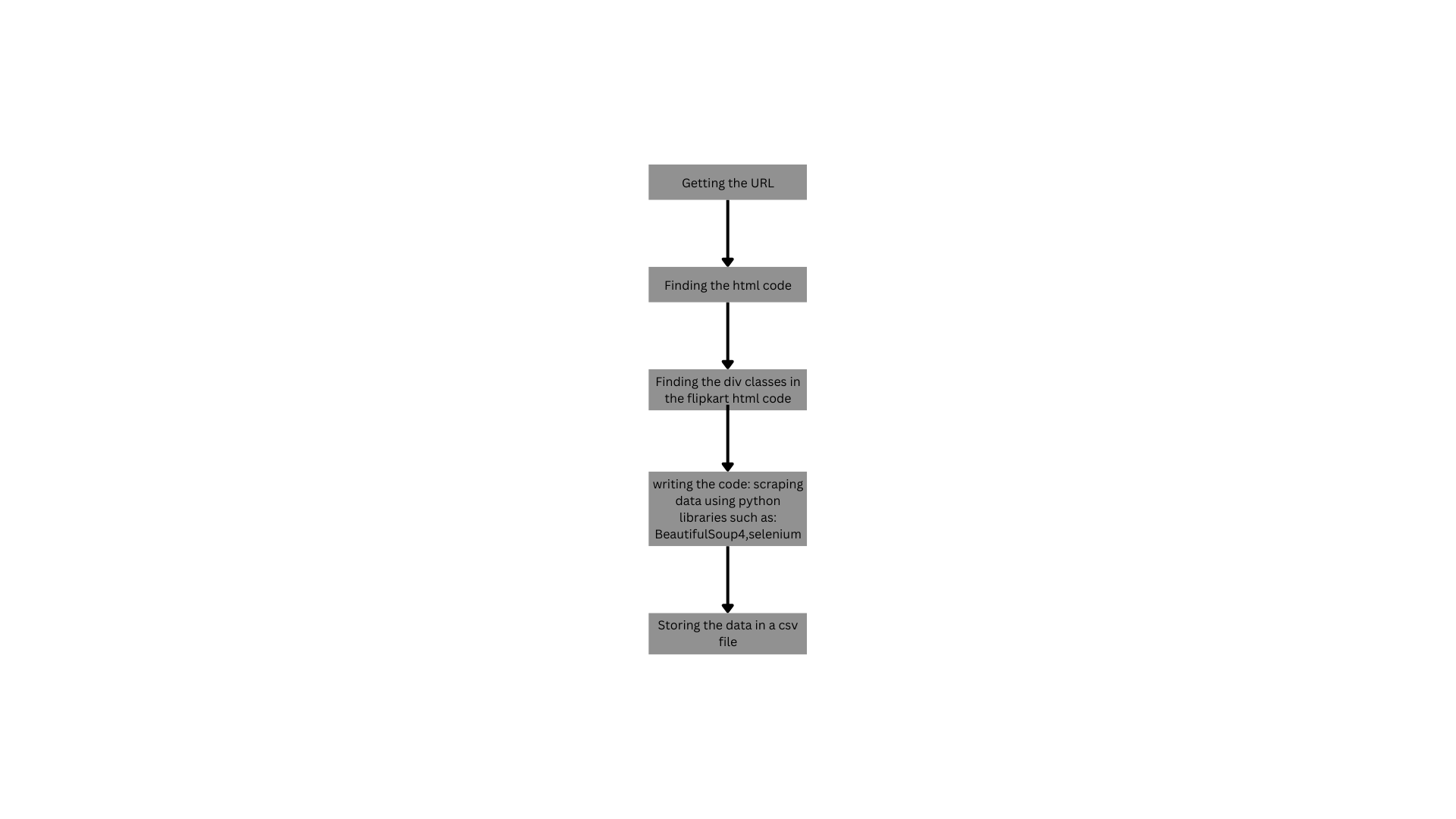
**Hardware Requirements:**

* CPU i3 7generation, Amd Ryzen series 5000

**Software Requirements:**

* Operating System (Linux/Windows)

**Flowchart**



**Design and Implementation**

**Design:**

The File will include product name, product price, description and review(rating)

**Setup:**

1. Install necessary libraries in python

* Pandas
* Requests
* beautifulSoup4

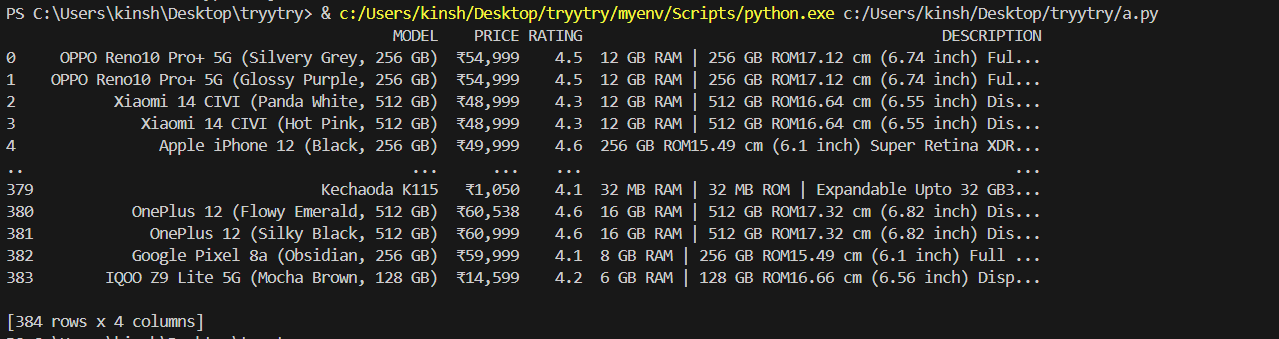
1. Taking the url of a specific product category we want to scrape data.
2. Writing the code for web Scraping.
3. Storing the data in a csv format.

**Code**





**Testing The Code**

****

**Future Enhancements**

* Scalability Enhancements
* Improved Data Management
* Multi-Platform Scraping
* Artificial Intelligence and Automation

**Conclusion**

The Python web scraping project has successfully demonstrated the ability to extract valuable data from the Flipkart website, including detailed product information, prices, ratings, and customer reviews. By leveraging powerful Python libraries such as BeautifulSoup, Requests, Selenium, and Scrapy, the project has shown how web scraping can be effectively utilized for market research, trend analysis, and competitive benchmarking.