**PROJECT REPORT**

**ON**

“Understanding Women’s

Shopping Preferences

By

22071A6751-SRIRAM NAIK

22071A6723-KSN VARSHIN

22071A6730-MAYANK SHAMBHU

23075A6707-MD NEHAL



**WEBSITE:** [**https://www.nykaa.com**](https://www.nykaa.com)

**INDEX**

|  |  |  |
| --- | --- | --- |
| **S.NO.** | **Content** | **Page No.** |
| **1** | **Introduction** | **3** |
| **2** | **Methodology** | **4** |
| **3** | **Challenges and Issues** | **7** |
| **4** | **Result** | **8** |
| **5** | **Conclusion** | **10** |

**INTRODUCTION:**

Web scraping is the process of extracting data from websites, a practice that has become increasingly prevalent in the digital age. One of the many websites that have garnered significant attention from web scrapers is Nykaa, a popular e-commerce platform specializing in beauty and cosmetics products. Founded in 2012 by Falguni Nayar, Nykaa has rapidly risen to prominence in India, offering a wide range of beauty and skincare products. This introduction sets the stage for understanding the significance of scraping the Nykaa website.

The beauty and cosmetics industry is one of the fastest-growing sectors worldwide. In India, it has witnessed remarkable growth driven by increased beauty awareness, shifting beauty standards, and the proliferation of the e-commerce market. Nykaa, as a frontrunner in this space, has played a pivotal role in shaping consumer choices and beauty trends.

Web scraping of the Nykaa website is an endeavor aimed at unlocking the valuable data contained within its virtual aisles. By extracting information on the extensive range of beauty products, brands, customer reviews, and pricing details, web scraping enables businesses, researchers, and consumers to gain deep insights into the beauty and cosmetics industry. Whether it is tracking the popularity of specific products, understanding market trends, or comparing prices, scraping Nykaa provides a valuable repository of data for various stakeholders.

In this report, we will delve into the nuances of scraping Nykaa, including the methodologies employed, ethical considerations, challenges faced, and the ultimate impact of this process on the beauty industry. By examining this case study, we can uncover the multifaceted nature of web scraping and its crucial role in shaping the contemporary landscape of the beauty and cosmetics sector, with Nykaa as a focal point.

**METHODOLOGY:**

Web scraping the Nykaa website involves a systematic approach to extracting data from its web pages. The methodology encompasses several steps and considerations to ensure the successful collection of information while adhering to ethical and legal guidelines.

1. Selecting the Tools and Technologies:

- We have chosen Python as it is a popular choice due to its rich libraries, such as BeautifulSoup and requests, which facilitate web scraping tasks.

2. Understanding the Website Structure:

- Begin by conducting an in-depth analysis of the Nykaa website. Use browser developer tools to inspect the website's HTML structure and identify the elements that contain the desired data. This includes understanding the URL structure, pagination, and the structure of product listings.

3. Defining the Data to be Scraped:

- Clearly define the data you intend to scrape from the website. This can include product details, pricing information, customer reviews, product ratings, and more. Define the data fields and the format in which the data should be collected.

4. Sending HTTP Requests:

- Use the `requests` library in Python to send HTTP GET requests to the Nykaa website. This involves constructing URLs for the pages to be scraped, including any necessary query parameters like page numbers, sorting options, and product categories.

5. Parsing HTML Content:

- Utilize a parsing library like BeautifulSoup to extract data from the HTML content retrieved through the HTTP requests. BeautifulSoup allows you to traverse the HTML structure and select specific elements using CSS selectors or XPath.

6. Data Extraction and Cleaning:

- Extract the relevant data from the parsed HTML content. Clean and preprocess the data to ensure it is in a usable format. This may include handling missing values, removing unnecessary whitespace, and converting data types.

7. Handling Pagination:

- If the data you are interested in spans multiple pages, implement a mechanism to iterate through the pages. This often involves finding and clicking on "Next" or "Load More" buttons or updating URL parameters accordingly.

8.Rate Limiting and Ethical Considerations:

- Implement rate limiting to avoid overloading Nykaa's servers. Respect the website's terms of service, robots.txt file, and ethical guidelines throughout the scraping process. Be mindful of any legal constraints.

9. Error Handling:

- Develop error-handling procedures to address potential issues, such as connection errors, website changes, or data inconsistencies. Log errors for review and further improvement.

10. Data Storage:

- Save the scraped data into a suitable format, such as CSV, JSON, or a database. Ensure that data is organized for easy analysis and retrieval.

11. Documentation:

- Thoroughly document the entire scraping process, including code, data formats, and any specific insights or findings. This documentation will be valuable for future reference and reporting.

The methodology for scraping the Nykaa website should be systematic and well-documented to ensure the successful retrieval of data while maintaining ethical standards and legal compliance. It is essential to follow these steps to conduct a responsible and effective web scraping project on Nykaa or any other website

**CHALLENGES AND ISSUES:**

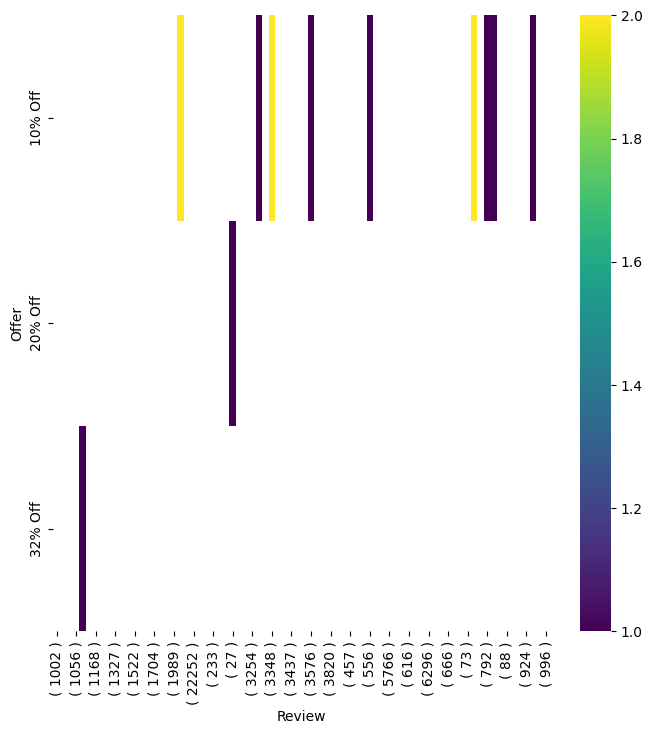
The first problem we faced are as follows:

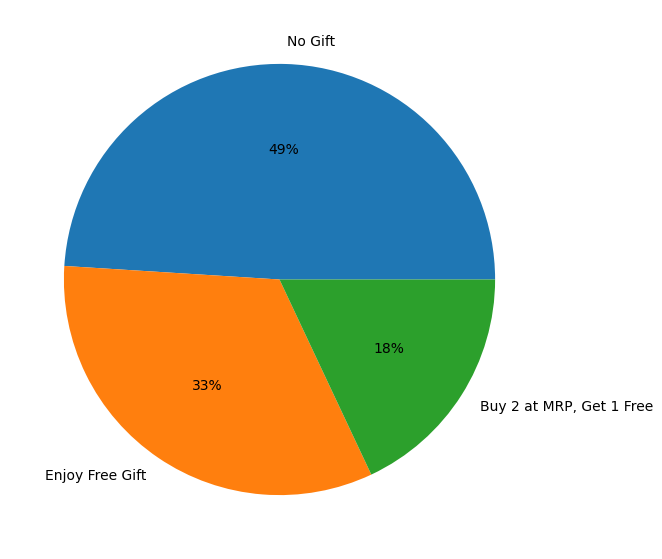
* DIFFICULTY IN UNDERSTANDING THE GIST OF THE QUESTION
* FINDING THE RESOURCES REQUIRED TO EXECUTE THE CODE
* DEBUGGING THE CODE
* UNDERSTANDING THE CODE
* EXECUTING THE DATA FOUND IN THE WEBSITE

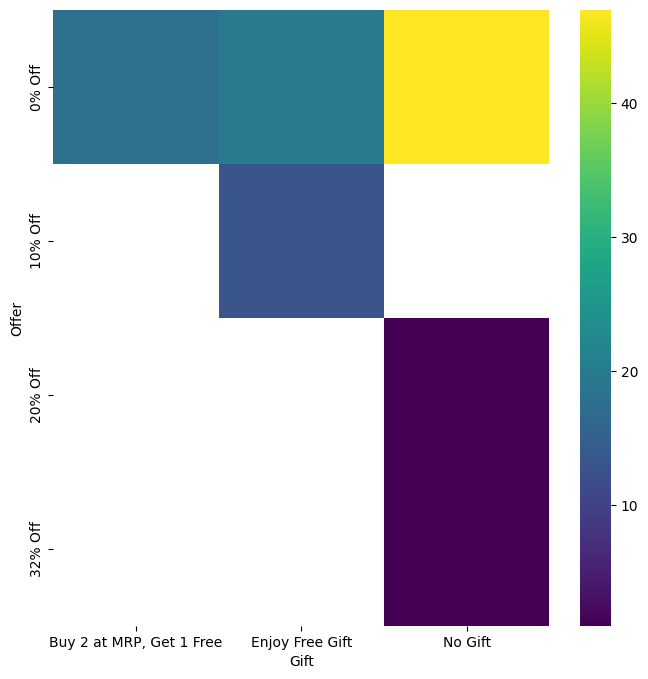
**RESULTS:**

The results we obtained are

**Offers vs reviews**



**The percentage of different Gifts given to all the product**

**offer vs free gifts given with them**

**CONCLUSION:**

The process of scraping the Nykaa website has provided invaluable insights into the beauty and cosmetics industry, underscoring the significance of web scraping in understanding market trends, consumer behavior, and product offerings. This scraping project has revealed the multifaceted nature of Nykaa as a prominent e-commerce platform, and the implications of this endeavor extend beyond the digital realm.

By systematically extracting and analyzing data related to a wide range of beauty and skincare products, pricing information, customer reviews, and product popularity, we have gained a deeper understanding of the dynamics of the beauty industry. The wealth of information obtained through web scraping has allowed us to identify trends in product demand, the influence of customer reviews on purchasing decisions, and the competitive landscape within the industry.

This project has highlighted the role of data in shaping business strategies, marketing decisions, and consumer preferences. For businesses operating in the beauty sector, the insights derived from the scraped data can inform product development, pricing strategies, and marketing campaigns. Researchers and analysts can utilize this data to conduct in-depth studies, identify emerging trends, and make data-driven recommendations.

However, it is essential to emphasize the ethical and legal considerations when conducting web scraping, particularly when scraping commercial websites like Nykaa. Respecting the terms of service, robots.txt files, and responsible scraping practices is imperative to ensure the integrity of the process and maintain a positive relationship with the website.

In conclusion, web scraping of the Nykaa website has illuminated the transformative power of data in the beauty and cosmetics industry. It has demonstrated the potential of web scraping as a tool for gaining insights, driving informed decision-making, and contributing to the broader discourse surrounding the evolving landscape of e-commerce and consumer behavior. The lessons learned from this endeavor serve as a reminder of the endless possibilities that data-driven analysis can offer to those willing to explore and harness its potential.

**REFERNCES:**

* <https://github.com/Rechel02/Web-Scraping-Analysis-of-Nykaa/blob/main/Nykaa%20project.ipynb>
* Chatgpt