Myntra Review Scrapper

PW Skills

Project Overview

- This project is a Myntra review scraper that allows users to extract and analyze customer reviews from the Myntra website.
- The scraper collects valuable information, such as product ratings, reviews, and user feedback, providing insights into customer sentiments and preferences.
- Objective: Develop a web scraper to extract customer reviews and ratings from Myntra's product pages.
- Purpose: Gain insights into customer sentiments and preferences to inform business decisions. Outcome: A structured dataset containing product reviews, ratings, and user feedback.

- Data Source: Myntra's official website. Technologies Used: Python, BeautifulSoup, Requests, Pandas, Flask/Streamlit.
- Process Flow: User inputs product URL. Scraper fetches and parses the webpage. Extracted data is cleaned and stored. Data is presented in a user-friendly format

Requirements

System Requirements

- **Python Version:** 3.10
- **Package Manager:** Conda
- **Required Libraries:**
  - requests
  - beautifulsoup4
  - pandas
  - Flask/streamlit
  - gunicorn

GitHub Link: https://github.com/PWskills-DataScienceTeam/myntra-review-scrapper

Implementation Steps

- **Setup Environment:**
- Clone the repository.
- Create and activate a Conda environment.
- **Install Dependencies:**
- Run pip install -r requirements.txt.
- **Run Application:**
- Execute python app.py to start the Flask server. streamlit run app.py for starting Streamlit server
- **Access Interface:**
- Navigate to http://localhost:5000 in your browser. (Flask) but port 8501 for streamlit
- **Input URL:**
- Enter the Myntra product URL to scrape reviews.
- **View Results:**
- Scraped data is displayed and can be downloaded

Libraries for the Project

requests: Handles HTTP requests to fetch web pages.
- **beautifulsoup4:** Parses HTML content to extract specific data.
- **pandas:** Structures and manipulates data for analysis.
- **Flask/streamlit:** Creates a web interface for user interaction.
- **gunicorn:** Serves the Flask application in production environments

Data Extraction Process

- User Input: Myntra product URL.
- HTTP Request: Fetches the product page content.
- HTML Parsing: Identifies and extracts review sections.
- Data Cleaning: Removes unnecessary characters and formats data.
- Data Storage: Organizes data into a structured format (e.g., CSV)

Conclusion

Benefits and Applications

- Customer Insights: Understand customer satisfaction and areas of improvement. Product Development: Inform design and feature enhancements based on feedback.
- Marketing Strategies: Tailor campaigns to address common customer concerns. Competitive Analysis: Benchmark against competitors by analyzing similar products

Future Enhancements

- **Sentiment Analysis:** Integrate NLP techniques to gauge overall sentiment.
- **Automation:** Schedule regular scraping for continuous data updates.
- **Data Visualization:** Implement dashboards to visualize trends and patterns.
- **Scalability:** Extend scraper to other e-commerce platforms.

Recap: Developed a tool to extract and analyze customer reviews from Myntra.

Impact: Provides valuable insights for product and business strategy.