NAME : KATEPALLI NANDINI
ROLL NO : 23471A05DW
BRANCH : CSE
SECTION : C2- DA1

E-Commerce Data Analysis of Rechargeable Lint Removers

A data-driven approach to enhance consumer electronics purchasing decisions on e-commerce platforms.

The Problem: Navigating E-Commerce Complexity

1

Overwhelming Choices

Customers face a vast array of rechargeable lint removers on sites like Flipkart, making product selection difficult.

2

Varied Product Attributes

Significant differences in price, ratings, discounts, brand reputation, stock, and delivery times create confusion.

3

Lack of Easy Comparison

Without objective comparison tools, buyers struggle to make informed decisions, leading to potential dissatisfaction.

4

The Need for Clarity

There is a clear demand for a data-driven solution to objectively analyze and compare products.

Our Data-Driven Solution

Web Scraping

Automated extraction of product data from Flipkart using Python and BeautifulSoup.

Data Analysis

Cleaning and analysis of collected data using Pandas to uncover key insights.

Empowering Buyers

Generating actionable insights to help consumers make smarter purchasing choices.

Project Objectives: Guiding Our Analysis

Extract Product Details

Gather comprehensive information on rechargeable lint removers from Flipkart listings.

Analyze Pricing & Affordability

Understand price distribution and identify budget-friendly options.

Compare Ratings & Satisfaction

Evaluate customer feedback and satisfaction across various brands.

Study Discount Patterns

Uncover trends in promotional strategies and their impact.

Assess Availability & Delivery

Monitor stock levels and analyze shipping timelines.

Identify Value-for-Money

Pinpoint products offering the best combination of quality and price.

Methodology: Our Workflow

01

Data Collection

Collect HTML pages from Flipkart listings for rechargeable lint removers.

02

Information Extraction

Scrape product names, prices, ratings, brands, reviews, discounts, and availability.

03

Data Cleaning & Preprocessing

Standardize and refine the raw data for accurate analysis.

04

Exploratory Data Analysis (EDA)

Conduct initial analysis to discover patterns, anomalies, and test hypotheses.

05

Insight Generation & Visualizations

Develop clear insights and graphical representations of findings.

06

Reporting

Export results to CSV for easy sharing and further analysis.

Key Analyses Performed

- Price distribution of rechargeable lint removers.
- Average ratings by brand for customer satisfaction.
- Relationship between discount percentages and product ratings.
- Comparison of in-stock versus out-of-stock item availability.
- Analysis of sponsored vs. organic product listing prevalence.
- Identification of high-rated, affordably priced products.

Insights from Our Data

Budget-Friendly Focus

Most rechargeable lint removers fall within the budget to mid-range price categories.

High Satisfaction

Customer ratings are generally high, indicating strong product performance satisfaction.

Discounts & Ratings

The size of a discount does not significantly influence product ratings, suggesting inherent quality drives satisfaction.

Organic Dominance

Organic product listings are more prevalent than sponsored listings on Flipkart for this category.

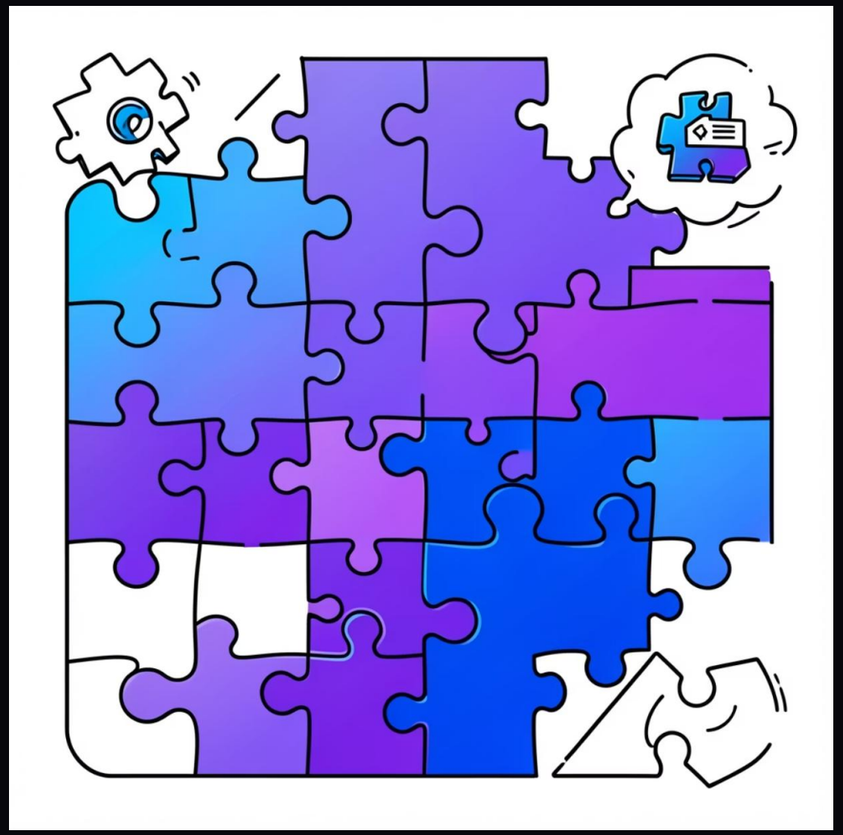
Limited High-Value Options

While high-quality, affordable lint removers exist, their availability is relatively limited.

Overcoming Challenges

Challenges Faced

- Inconsistent HTML structures across product listings.
- Missing review or discount information.
- Duplicate product entries from multiple pages.
- Limited availability of structured review data.



Solutions Implemented

- Developed robust parsing logic to handle HTML variations.
- Employed careful handling and imputation for missing values.
- Implemented deduplication algorithms to ensure data accuracy.
- Documented assumptions for unavailable data transparently.

Tools & Technologies

Python

Core programming language for scripting and data manipulation.

BeautifulSoup

Library for parsing HTML and XML documents.

Pandas

Data analysis and manipulation library.

NumPy

Fundamental package for numerical computing.

Matplotlib

Library for creating static, animated, and interactive visualizations.

Jupyter Notebook

Interactive computing environment for analysis and visualization.

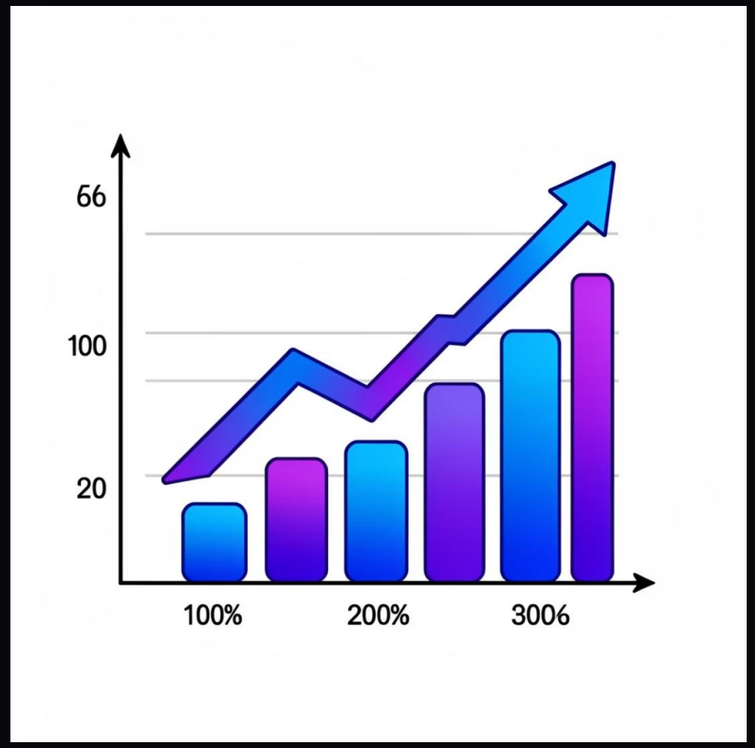
GitHub

Platform for version control and collaborative development.

Conclusion & Future Enhancements

Conclusion

This project successfully demonstrates how data analysis can resolve real consumer dilemmas. By providing clear, data-driven insights into rechargeable lint remover products, we enable users to compare options effectively and make intelligent purchasing decisions.



Future Enhancements

- Integrate **Sentiment Analysis** of customer reviews for deeper insights into product perception.
- Develop **Real-time Scraping** capabilities using e-commerce APIs for up-to-the-minute data.
- Implement a **Recommendation System** to personalize product suggestions for users.
- Create **Interactive Dashboards** for consumers to explore data intuitively.

THANK YOU