Milestone 1 Report: Data Pipeline Setup

This milestone focuses on building a basic data collection and ingestion pipeline. The goal was to scrape product data from a public website and convert it into a structured dataset for further analysis.

Objectives

- Implement web scraping using Python
- Extract product details such as name, price, and ratings
- Generate sentiment labels from ratings
- Store the cleaned dataset in CSV format

Tools & Technologies Used

- Python
- Requests library
- BeautifulSoup
- CSV module

Methodology

The scraper collected product data from a public test e-commerce website. Each product card was parsed to extract the product name, price, and rating. Price values were cleaned and converted into numeric format. Based on star ratings, sentiment labels were generated as Positive, Negative, or Neutral.

Output

The final output is a structured CSV dataset containing product name, price, star rating, sentiment label, and scraped date. This dataset will be used for further analysis in future milestones.

Conclusion

This milestone successfully demonstrates the implementation of a basic data pipeline. The scraping, cleaning, and storage process has been completed and validated, forming the foundation for advanced analytics and modeling in upcoming milestones.