# Description

This project is designed to analyze product data from multiple sources such as Amazon, Best Buy, Google Shopping, and Walmart. It provides insights into the top products based on various metrics including price, ratings, review counts, and sentiment analysis. The tool is aimed at helping users make informed purchasing decisions by identifying the best value products across these platforms.

# Features

- **Data Loading**: Load product data from various formats including CSV and JSON.
- **Data Processing**: Normalize data and calculate combined scores based on weighted metrics for price, rating, and reviews.
- **Sentiment Analysis**: Analyze customer reviews to assess product sentiment.
- **Top Product Identification**: Identify and display the top products based on combined scores and sentiment.

# Setup

To run this project, you will need Python installed on your machine along with several packages.

## Prerequisites

- Python 3.x
- Pandas
- NumPy
- Matplotlib
- TextBlob
- Requests
- Pathlib
- BeautifulSoup

## Installation

1. Install required Python packages:

```
pip install pandas numpy matplotlib textblob
pip install loguru
pip install scrapfly-sdk
pipx install poetry
pip install jmespath
pip install parsel
```

- You might need to install the corpora for TextBlob if you're using it for the first time:

```
python -m textblob.download_corpora
```

## Usage

To run the program in UI we designed, execute the main script from the command line in the directory you downloaded, a new window will be popped up in the back of your current window:

```
python group_5_RateMate.py
```

Follow the instruction prompted and you will proceed with our program. You can always chose to exit the program when you at the main menu.

There will be an option for you to choose to get updated scraped data. Check the box to run the data python files you want and choose a product you want to see the analysis with the dropdown box and click search butten right below it. It might take several minutes to run if you choose download all the data(You can always check back in the terminal to see the progress of downloading data).

All data files will be automatically saved in the same directory with the python files you are running with. This program eventually uses below four processed data files from the scraped data. Each file contains a combined five products' (iphone, ipad, macbook, Nintendo Switch, and PlayStation) information for each sources. (Previously extracted files already exported and exists in the directory)

1. amazon_product_reviews.csv
2. bestbuy_combined.csv
3. google_shopping_combined.csv
4. walmart_products.json

## Data

1. BestBuy’s data is scraped using scrapfly.io and Python to scrape property listing data from BestBuy.com This scraper scrapes: BestBuy product pages for product data. BestBuy search pages for product data on search pages.

2. Amazon's data is directly download from UCSD Professor Julian McAuley's Lab, collected from 1996 to 2023, https://cseweb.ucsd.edu/~jmcauley/datasets.html#amazon_reviews. Two data files is being used: "Electronics.jsonl" (product review information) and "meta_Electronics.jsonl" (product detail information). Both files is being merged based on 'parent_asin'. We uses this customer's opinion data to do sentiment analysis combined with scores of the product's pricing and rating to find which product and review combination has the best rated quality.

3. Google shopping is being scraped by using requests and Oxylabs API with the search word Electronics and set the geolocation to United States. The data provides us lists of searched products, their merchant, title, price, ratings, number of reviews. We used this dataset to do comparisons based on combination scores of price, rating, and reviews_counts.

4. Walmart data is being scraped by using BeautifulSoup and ScraperAPI. We got a list of product informations with multiple review texts for the specific product. The data provides us an option to do a deep sentiment analysis on user's reviews and combine the rating and rating_counts in parallel to generate a combined score.

## Functions Description

- load_data(): Loads data from files stored in the same directory as the script.
- process_data(data_frames, product_name): Processes data to find top products based on specified criteria.
- process_amazon_data(df): Function to get the top three Amazon reviews with product information. NOTE: Amazon data is being processed and sorted in amazon_data_process.py with a combination score of price, rating, and review sentiment.
- bestbuy_top_cheapest(df, title), google_top_cheapest(df, title): Functions that process Best Buy and Google Shopping data to find the top three products with the best combination score of ratings, prices, and review_counts/rating_counts.
- process_walmart_top_cheapest(df): Functions that process Walmart data to find the top three products according to the price, rating, and review sentiment.
- perform_sentiment_analysis(df): Analyzes sentiments of product reviews.

## License

This project is licensed under the MIT License.