

OK DEAL!'s README



Course: 95-888 Data-Focused-Python
Carnegie Mellon University

Introducing the team at OK DEAL!

1. Adrian Menezes (ajmeneze@andrew.cmu.edu)
2. Gautam Naik (gnaik@andrew.cmu.edu)
3. Michael Affare (maffare@andrew.cmu.edu)
4. Ruchi Bhatia(rsbhatia@andrew.cmu.edu)
5. Shivam Patel (shpatel@andrew.cmu.edu)
6. Yao-Chang Wang (yaochanw@andrew.cmu.edu)

Link to our Application Demo

<https://youtu.be/e4uirgoezNI>

Steps to install our application

1. Download 'Data-Focused-Python.zip' from Canvas, unzip it and store it at the desired location.
2. Open Anaconda prompt or a similar terminal
3. Change your path to the **Data-Focused-Python** folder (stored into your computer) using the cd command (you should be inside the Data-Focused-Python folder, for e.g. cd Data-Focused-Python)
4. Install the required dependencies by running the following command:
pip install -r requirements.txt
Note: It will take a few minutes to install all the dependencies
5. Our folder structure looks like below:

```
— README.md
— final
  — data
    — amazon
      — amazon_products_cleaned.csv
      — amazon_products_raw.csv
      — amazon_reviews_cleaned.csv
      — amazon_reviews_cleaned.json
      — amazon_reviews_raw.xlsx
    — analysis
      — Crosstab_variant_amazon.xlsx
      — Crosstab_variant_croma.xlsx
      — Crosstab_variant_flipkart.xlsx
      — stats.xlsx
    — croma
      — croma_products_cleaned.csv
      — croma_products_raw.csv
      — croma_reviews_cleaned.csv
      — croma_reviews_cleaned.json
      — croma_reviews_raw.csv
    — currency_api
      — currency_conversion_api_data_cleaned.csv
      — currency_conversion_api_data_raw.csv
    — flipkart
      — flipkart_products_cleaned.xlsx
      — flipkart_products_raw.csv
      — flipkart_reviews_cleaned.csv
      — flipkart_reviews_cleaned.json
      — flipkart_reviews_raw.csv
    — merged
      — products
        — Merged_Products_Data_Cleaned.csv
        — Merged_Products_Data_Raw.csv
      — reviews
        — Amazon\ Reviews\ Wordcloud.png
        — Croma\ Reviews\ Wordcloud.png
        — Flipkart\ Reviews\ Wordcloud.png
        — Merged_Products_Reviews_Cleaned.csv
        — Merged_Products_Reviews_Raw.csv
      — tweets
        — negative_tweets.csv
        — neutral_tweets.csv
        — positive_tweets.csv
        — tweets_cleaned.csv
        — tweets_merged_not_cleaned.csv
```

```

├── phonearena
│   ├── phonearena_cleaned.csv
│   ├── phonearena_links.txt
│   ├── phonearena_raw.csv
│   └── phonearena_scrapped.json
├── tweets
│   ├── tweets_apple_iphone_13_mini.csv
│   ├── tweets_apple_iphone_xr.csv
│   ├── tweets_oppo_a15.csv
│   ├── tweets_oppo_a15s.csv
│   ├── tweets_oppo_a16.csv
│   ├── tweets_oppo_f17.csv
│   ├── tweets_realme_8i.csv
│   ├── tweets_samsung_galaxy_a03s.csv
│   └── tweets_samsung_galaxy_z_fold3_5g.csv
├── engines
│   ├── Product_Engine.py
│   ├── Review_Engine.py
│   └── Specs_Engine.py
├── main.py
├── phonearena_cleaned.csv
├── phonearena_links.txt
├── phonearena_raw.csv
├── phonearena_scrapped.json
├── results
│   ├── samsung\ galaxy\ f62.txt
│   └── samsung\ galaxy\ m02.txt

```

```

├── scripts
│   ├── Products_Analysis.py
│   ├── Tweet_Sentiment_Analysis.py
│   ├── __init__.py
│   └── data_gathering
│       ├── amazon
│       │   ├── clean_amazon_products.py
│       │   ├── clean_amazon_reviews.py
│       │   ├── resources
│       │   │   ├── search_results.yml
│       │   │   ├── search_results_reviews.yml
│       │   │   └── search_results_urls.txt
│       │   ├── scrape_amazon_products.py
│       │   └── scrape_amazon_reviews.py
│       ├── cromia
│       │   ├── chromedriver.exe
│       │   ├── scrape_cromia_products.py
│       │   └── scrape_cromia_reviews.py
│       ├── currency
│       │   └── currency_api.py
│       ├── flipkart
│       │   ├── scrape_flipkart_products.py
│       │   └── scrape_flipkart_reviews.py
│       ├── phonearena
│       │   ├── phonearena_cleaning.py
│       │   └── scrape_phonearena.py
│       └── twitter
│           ├── clean_tweets.py
│           └── scrape_tweets.py
├── merge_files.py
├── review_analysis.py
├── scraper.py
└── requirements.txt

```

Our key folders are:

- a. **data** - Contains all the scraped data
- b. **engines** - Contains Python files that link other Python files (basically acting as an engine)
- c. **results** - Contains the results of searching a product on OK Deal! It contains text files which store results (product details, reviews and specification) for the searched product
- d. **scripts** - Contains all the Python files for making OK Deal! work

6. Currency API process

For using the current currency conversion rates, we used an API called [ExchangeRates](#).

Following are the steps we need to follow to use the API:

1. Create an account and sign up for the free plan
2. After creating an account, you will receive your API key.
3. We can request the most recent exchange rate from the API by using:
http://api.exchangeratesapi.io/v1/latest?access_key=YOUR_API_KEY
4. Then we can integrate this into our application and access the data which is presented to us in a json format.

7. Since scraping all our data takes a lot of time (2-3 days), we have scraped it and put all our scraped data into **Data-Focused-Python/final/data** folder.

Note: If you scrape new data (using our interactive cell below), the current data files will get overridden.

Steps to run our application

1. After installation, open Data-Focused-Python/final/main.py in Spyder or any other Python IDLE. The main.py is our main Python file that imports the other necessary modules of OK Deal!
2. Run main.py
Note: Once you run main.py, our data preparation module will take around 1-2 minutes to merge the data for the application.
3. You will be presented with an interactive menu which looks like this:

```
WELCOME TO OK DEAL!

Please wait. We are getting the data ready for you...

Merging...
data/merged/products already exists ...
Continuing

data/merged/reviews already exists ...
Continuing

data/merged/tweets already exists ...
Continuing

Merging done !!!

Amazon Products:  (6762, 8)
Croma Products:   (348, 8)
Flipkart Products: (984, 8)
Merged Products:  (3935, 22)
Amazon Reviews:   (19836, 8)
Croma Reviews:    (2982, 8)
Flipkart Reviews: (8695, 8)
Merged Reviews:   (1368820, 22)

[nltk_data] Downloading package vader_lexicon to C:\Users\Gautam
[nltk_data]   Naik\AppData\Roaming\nltk_data...
[nltk_data]   Package vader_lexicon is already up-to-date!

MAIN MENU

1. Search by product name
2. Product analysis visualizations
3. Twitter sentiment analysis visualizations
4. Scrape data
5. Review analysis
0. Exit

Enter the Choice: |
```

4. From the options given, you can select '1' to search for products by their names. A screenshot detailing the option in this part is given below:

```
MAIN MENU
1. Search by product name
2. Product analysis visualizations
3. Twitter sentiment analysis visualizations
4. Scrape data
5. Review analysis
0. Exit

Enter the Choice: 1

Choose preferred currency
1. Indian Rupee INR
2. United States Dollar USD
3. Australian Dollar AUD
4. EURO EUR
0. Go back

Enter the Choice: 1
Rupee value 84.794795
Conversion value 84.794795
Enter a product name: samsung galaxy

Product not found... Showing you some suggestions
Getting suggestions...
1 samsung galaxy a21s
2 samsung galaxy f02s
3 samsung galaxy z fold3 5g
4 samsung galaxy a03s
5 samsung galaxy a22 5g
6 samsung galaxy a52s 5g
7 samsung galaxy a22
8 samsung galaxy a52
9 samsung galaxy m02
10 samsung galaxy a51
0. Go back

Enter the Choice: 10
Selected product: samsung galaxy a51

You can view the product's details, specifications and reviews at this path: Data-Focused-Python/final/results/samsung galaxy a51.txt
```

5. From the options given, you can select '2' to see the various visualizations after product analysis. A screenshot detailing the option in this part is given below:

Note: After selecting an option, it will take some time (a few seconds to a minute or so) for the visualizations to pop-up. Moreover, some of the options contain multiple plots, so you will need to close one, to see the next and so on until you reach the end. Upon closing the final plot, you will reach the program execution again.

MAIN MENU

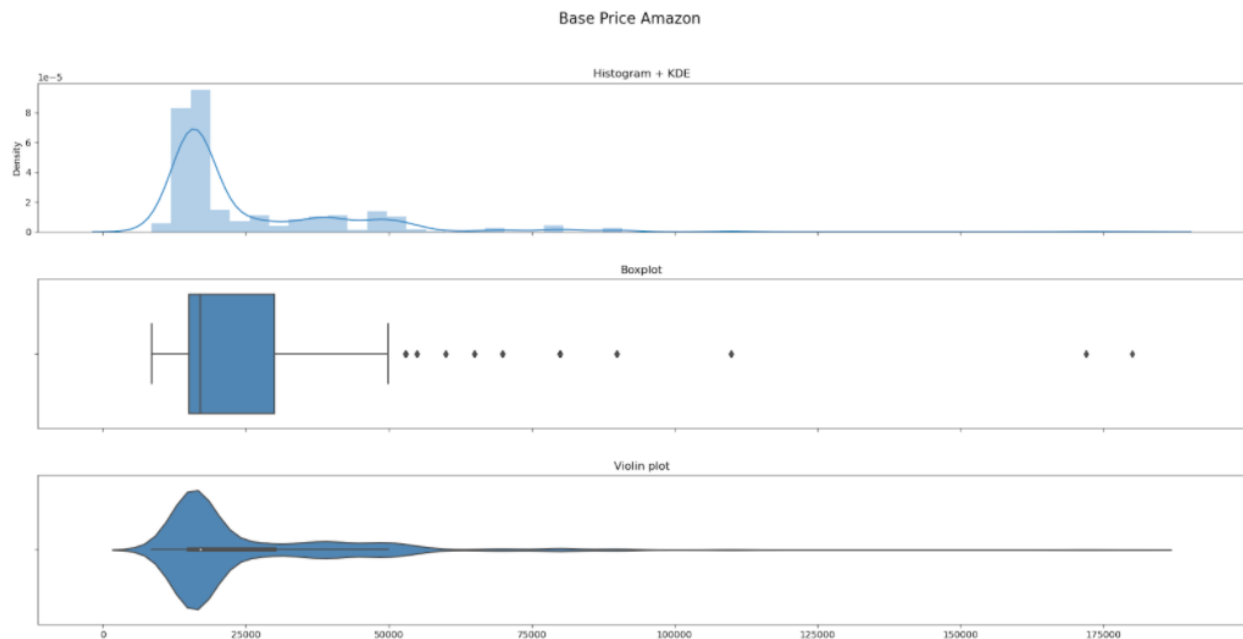
1. Search by product name
2. Product analysis visualizations
3. Twitter sentiment analysis visualizations
4. Scrape data
5. Review analysis
0. Exit

Enter the Choice: 2

Product Analysis Visualization

1. Base price distribution
2. Discount % distribution
3. Discount price distribution
4. Average rating distribution
5. Show all distribution plots
6. Base price triple plot
7. Discount percent triple plot
8. Discount price triple plot
9. Average rating triple plot
10. Descriptive statistics for Amazon, Croma and Flipkart
0. Go back

Enter the Choice: 6



6. From the options given, you can select '3' to see the various visualizations of twitter sentiments. A screenshot detailing the option in this part is given below:

Note: After selecting an option, it will take some time (a few seconds to a minute or so) for the visualizations to pop-up. Moreover, some of the options contain multiple plots, so you will need to close one, to see the next and so on until you reach the end. Upon closing the final plot, you will reach the program execution again.

MAIN MENU

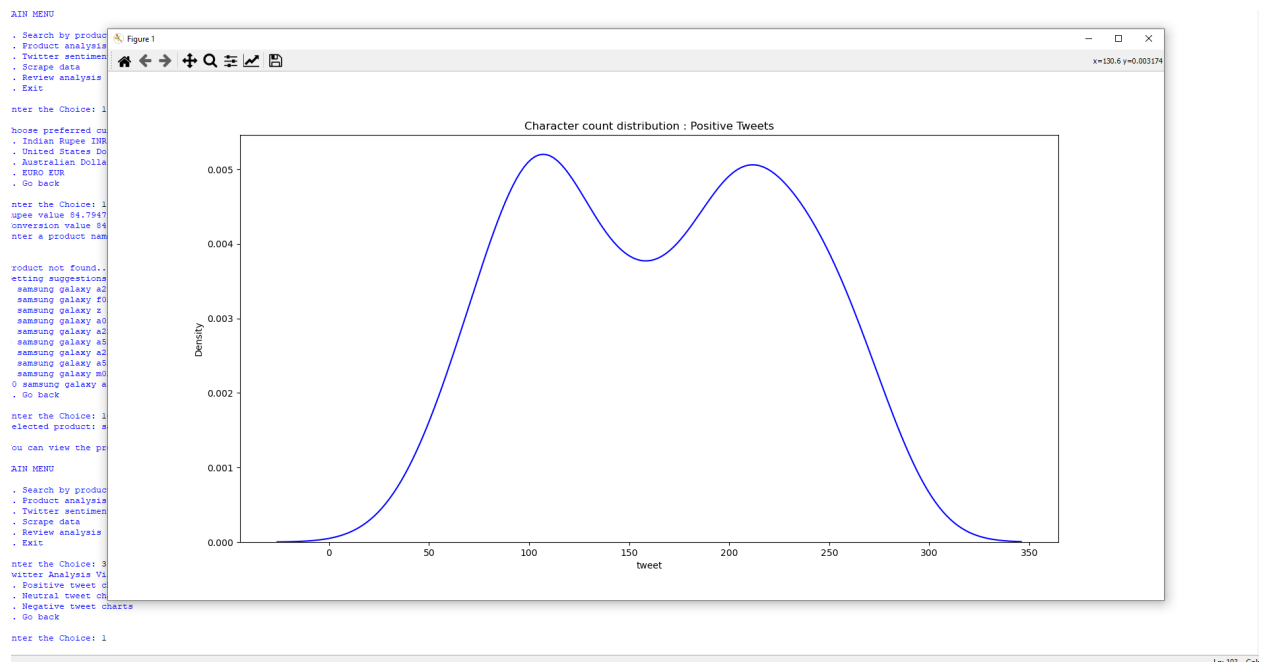
1. Search by product name
2. Product analysis visualizations
3. Twitter sentiment analysis visualizations
4. Scrape data
5. Review analysis
0. Exit

Enter the Choice: 3

Twitter Analysis Visualization

1. Positive tweet charts
2. Neutral tweet charts
3. Negative tweet charts
0. Go back

Enter the Choice: 1



7. From the options given, you can select '4' to scrape the data we have used. A screenshot detailing the option in this part is given below:

Note: It took us 2-3 days to scrape all of the data.


```

MAIN MENU

1. Search by product name
2. Product analysis visualizations
3. Twitter sentiment analysis visualizations
4. Scrape data
5. Review analysis
0. Exit

Enter the Choice: 4
Scraping data
1. Product details
2. Product reviews
3. PhoneArena specifications
4. Twitter tweets
5. Currency conversion rates
0. Go back

Enter the Choice: 1
Scraping product details

===== WebDriver manager =====
Current google-chrome version is 96.0.4664
Get LATEST chromedriver version for 96.0.4664 google-chrome
Driver [C:\Users\Gautam Naik\.wdm\drivers\chromedriver\win32\96.0.4664.45\chromedriver.exe] found in cache

Total 243 mobile specifications to scrape

===== WebDriver manager =====
Current google-chrome version is 96.0.4664
Get LATEST chromedriver version for 96.0.4664 google-chrome
Driver [C:\Users\Gautam Naik\.wdm\drivers\chromedriver\win32\96.0.4664.45\chromedriver.exe] found in cache

Scraped specifications of 10 mobiles from Amazon
Scraped specifications of 20 mobiles from Amazon
Scraped specifications of 30 mobiles from Amazon
Scraped specifications of 40 mobiles from Amazon

```

8. From the options given, you can select '5' to see the various visualizations after product review analysis. A screenshot detailing the option in this part is given below:

Note: After selecting an option, it will take some time (a few seconds to a minute or so) for the visualizations to pop-up. Moreover, some of the options contain multiple plots, so you will need to close one, to see the next and so on until you reach the end. Upon closing the final plot, you will reach the program execution again.

```

1. Search by product name
2. Product analysis visualizations
3. Twitter sentiment analysis visualizations
4. Scrape data
5. Review analysis
0. Exit

```

Review analysis

- Enter the Choice: 1



9. From the options given, you can select '0' to exit the application.

MAIN MENU

1. Search by product name
2. Product analysis visualizations
3. Twitter sentiment analysis visualizations
4. Scrape data
5. Review analysis
0. Exit

Enter the Choice: 0

Goodbye

Link to our Application Demo (again)

<https://youtu.be/e4uirgoezNI>

Thank you!