



## **Data Collection and Preprocessing Phase**

| Date          | 14 <sup>th</sup> July 2024                    |
|---------------|---|
| Team ID       | 739782  |
| Project Title | Sentiment Analysis of Commodity<br>News(Gold) |
| Maximum Marks | 6 Marks                                       |

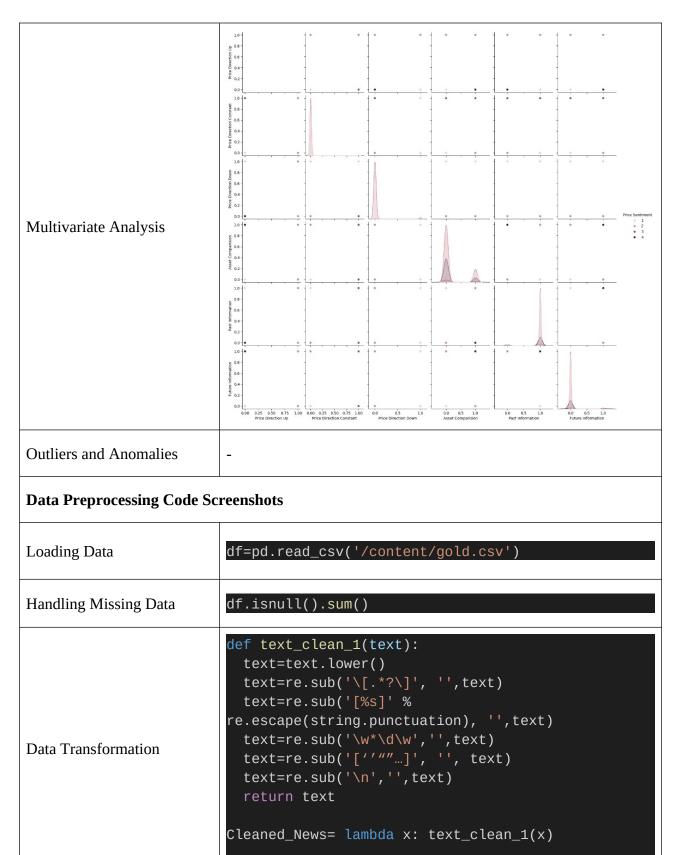
## **Data Exploration and Preprocessing Template**

Identifies data sources, assesses quality issues like missing values and duplicates, and implements resolution plans to ensure accurate and reliable analysis.

| Section             | Description                          |                        |  |   |                                  |                                  |   |                                      |                                  |                                      |                                    |
|---------------------|--------------------------------------|------------------------|--|---|----------------------------------|----------------------------------|---|--------------------------------------|----------------------------------|--------------------------------------|------------------------------------|
| Data Overview       | Dim                                  | ens                    | ions: (10570   | , 10)   |                                  |                                  |   |                                      |                                  |                                      |                                    |
| Univariate Analysis | ±v count                             | Dates                  | URL<br>10570   | News<br>10570   | Price<br>Direction Up            | Price Direction<br>Constant      | Price<br>Direction Down<br>10570.000000 | Asset<br>Comparision<br>10570.000000 | Past<br>Information              | Future<br>Information<br>10570.00000 | Price<br>Sentiment<br>10570.000000 |
|                     | unique                               | 3761<br>30-08-<br>2017 | 10570<br>http://www.marketwatch.com/story/april-<br>gold-do  | april gold down 20<br>cents to settle at<br>\$1,116.1 | NaN<br>NaN                       | NaN                              | NaN<br>NaN                              | NaN<br>NaN                           | NaN<br>NaN                       | NaN<br>NaN                           | NaN                                |
|                     | freq<br>mean                         | 18<br>NaN              | 1<br>NaN   | 1<br>NaN  | NaN<br>0.417408                  | NaN<br>0.042006                  | NaN<br>0.370104                         | NaN<br>0.189309                      | NaN<br>0.969915                  | NaN<br>0.03018                       | NaN<br>2.047114                    |
|                     | std<br>min<br>25%                    | NaN<br>NaN<br>NaN      | NaN<br>NaN   | NaN<br>NaN<br>NaN                                     | 0.493155<br>0.000000<br>0.000000 | 0.200612<br>0.000000<br>0.000000 | 0.482855<br>0.000000<br>0.000000        | 0.391773<br>0.000000<br>0.000000     | 0.170830<br>0.000000<br>1.000000 | 0.17109<br>0.00000<br>0.00000        | 1.067259<br>1.000000<br>1.000000   |
|                     | 50%<br>75%                           | NaN<br>NaN             | NaN<br>NaN   | NaN<br>NaN  | 0.000000<br>1.000000<br>1.000000 | 0.000000<br>0.000000<br>1.000000 | 0.000000<br>1.000000<br>1.000000        | 0.000000<br>0.000000<br>1.000000     | 1.000000<br>1.000000<br>1.000000 | 0.00000<br>0.00000<br>1.00000        | 2.000000<br>2.000000<br>4.000000   |
|                     | max                                  | NaN                    | NaN  | NaN   | 1.000000                         | 1.000000                         | 1.000000                                | 1.000000                             |                                  | ↓ ⇔ 🗏 ‡                              |                                    |
| Bivariate Analysis  | 4000 -<br>3000 -<br>1000 -<br>1000 - |                        | Price Direction Up 4000 - 30000 - 3000 - 3000 - 3000 - 3000 - 3000 - 3000 - 3000 - 3000 - 300 | Past Informa  0  1  1                                 | tion 4                           |                                  |   |                                      |                                  |                                      |                                    |











| Feature Engineering | NA   |
|---------------------|--|
| Save Processed Data | <pre>import pickle pickle.dump(model,open('model.pkl','wb'))</pre> |