



## **Data Collection and Preprocessing Phase**

Date	14 <sup>th</sup> July 2024
Team ID	740083
Project Title	Sentiment Analysis of Commodity News(Gold)
Maximum Marks	6 Marks

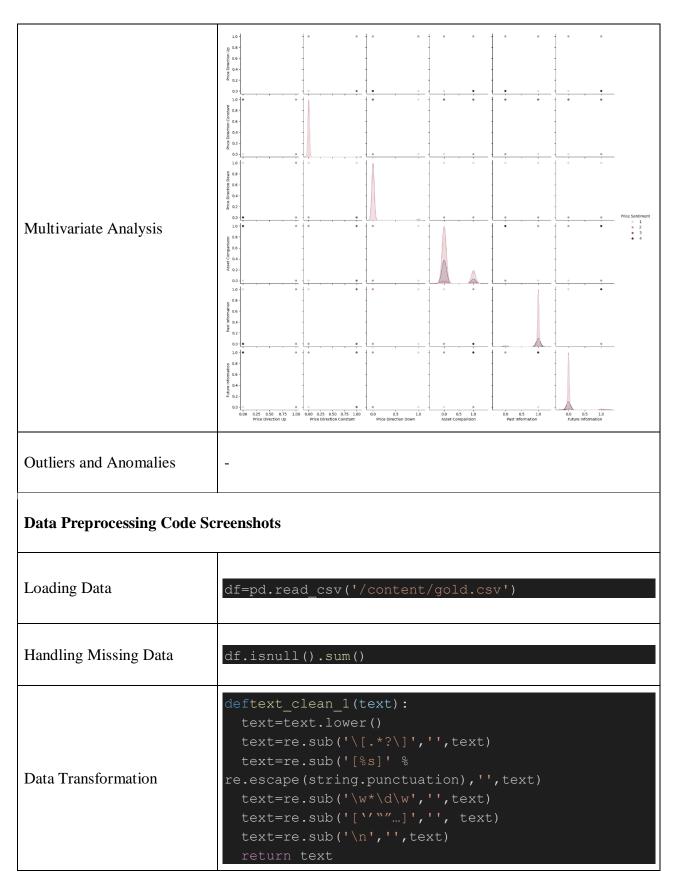
## **Data Exploration and Preprocessing Template**

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	Dimensions:(10570, 10)										
	±*	Dates	URL	News		Price Direction	Price	Asset	Past	Future	Price
	cour		10570	10570	Direction Up 10570.000000	Constant 10570.000000	10570.000000	Comparision 10570.000000	Information 10570.000000	Information 10570.00000	Sentiment 10570.000000
	uniqu	e 3761	10570	10570	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Univariate Analysis	top	30-08- 2017	http://www.marketwatch.com/story/april- gold-do	april gold down 20 cents to settle at \$1,116.1	NaN	NaN	NaN	NaN	NaN	NaN	NaN
	freq	18	1	1	NaN	NaN	NaN	NaN	NaN	NaN	NaN
	mea		NaN	NaN	0.417408	0.042006	0.370104	0.189309	0.969915	0.03018	2.047114
	std	NaN NaN	NaN NaN	NaN NaN	0.493155	0.200612	0.482855	0.391773	0.170830	0.17109	1.067259
	25%		NaN	NaN	0.000000	0.000000	0.000000	0.000000	1.000000	0.00000	1.000000
	50%	NaN	NaN	NaN	0.000000	0.000000	0.000000	0.000000	1.000000	0.00000	2.000000
	75%	NaN	NaN	NaN	1.000000	0.000000	1.000000	0.000000	1.000000	0.00000	2.000000
	max	NaN	NaN	NaN	1.000000	1.000000	1.000000	1.000000	1.000000	1.00000	4.000000
									1	<b>↓</b> ⊕ ■ :	<b>¢</b> 🖟 🖟 🗎 :
Bivariate Analysis	4000 - 3000 - 2000 - 1000 - 0	1 Pr	Price Direction Up 0 4000 - 30	Past Informa 0 1 1 2 3 Price Sentiment							











	<pre>Cleaned_News= lambda x: text_clean_1(x)</pre>
Feature Engineering	NA
Save Processed Data	<pre>import pickle pickle.dump(model,open('model.pkl','wb'))</pre>