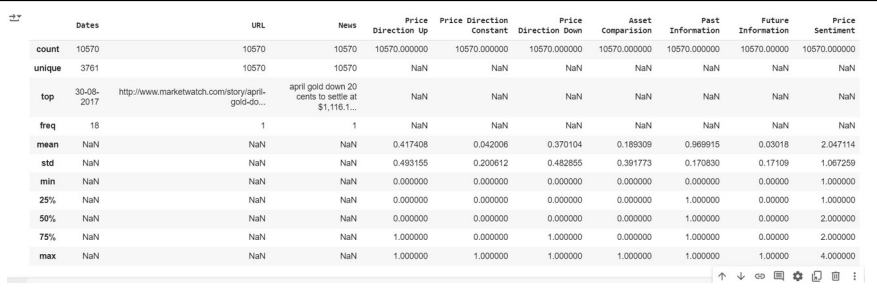
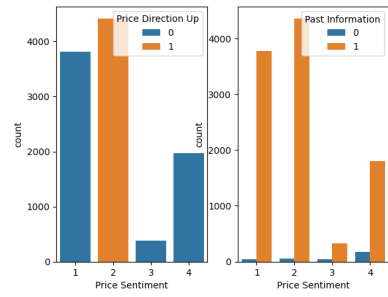


Data Collection and Preprocessing Phase

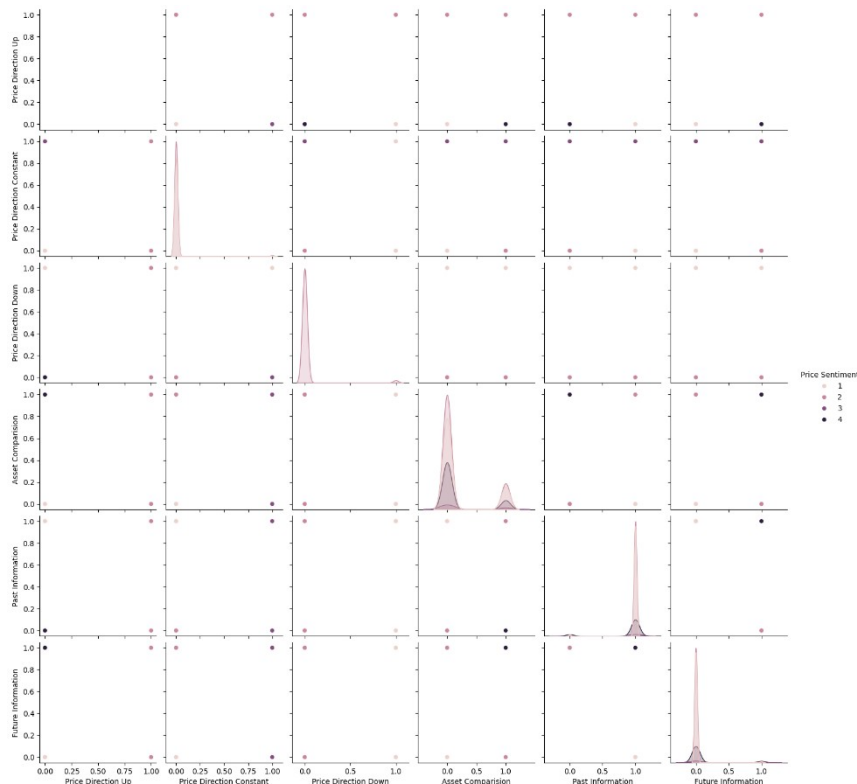
Date	14 th July 2024
Team ID	739782
Project Title	Sentiment Analysis of Commodity News(Gold)
Maximum Marks	6 Marks

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)
Univariate Analysis	
Bivariate Analysis	

Multivariate Analysis



Outliers and Anomalies

-

Data Preprocessing Code Screenshots

Loading Data

```
df=pd.read_csv('/content/gold.csv')
```

Handling Missing Data

```
df.isnull().sum()
```

Data Transformation

```
def text_clean_1(text):
    text=text.lower()
    text=re.sub('\[.*?\]', '',text)
    text=re.sub('[%s]' %
re.escape(string.punctuation), '',text)
    text=re.sub('\w*\d\w','',text)
    text=re.sub('[\'"\"...]', '', text)
    text=re.sub('\n','',text)
    return text

Cleaned_News= lambda x: text_clean_1(x)
```

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