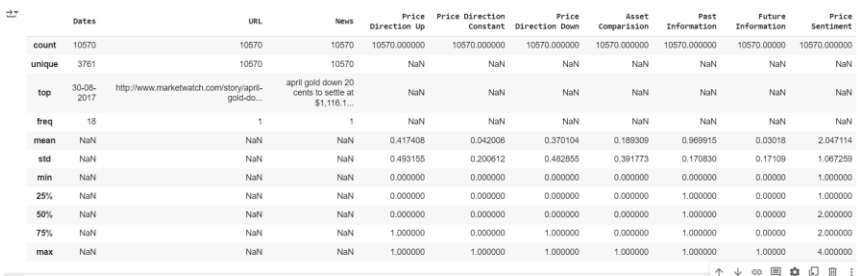
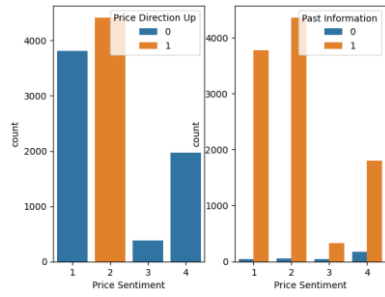


Data Collection and Preprocessing Phase

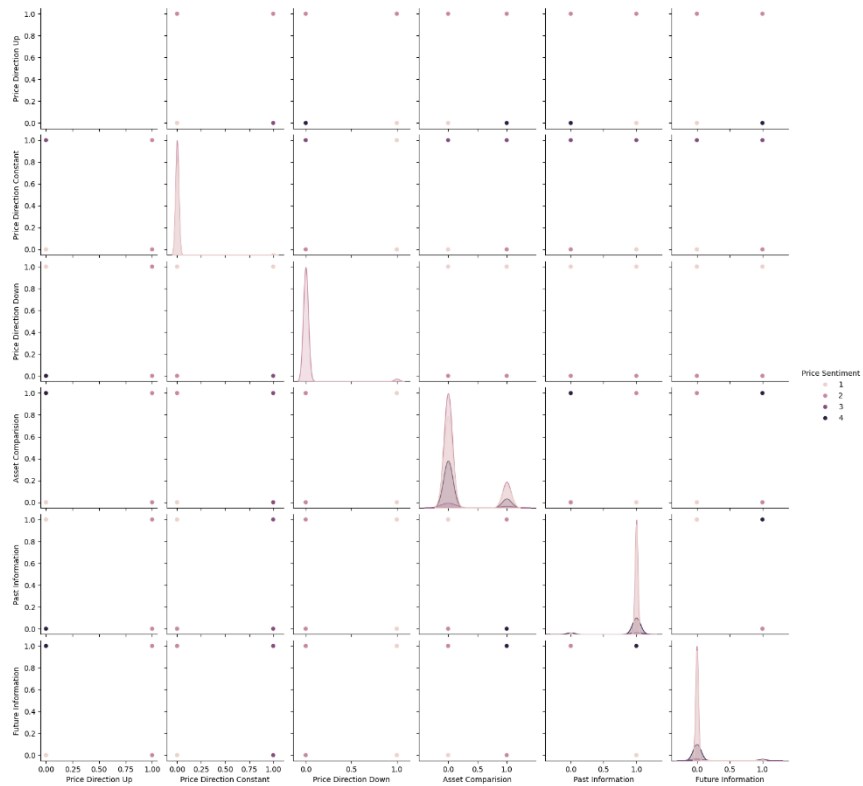
Date	14 th July 2024
Team ID	739959
Project Title	Sentiment Analysis of Commodity News(Gold)
Maximum Marks	6 Marks

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	 <p>Summary of univariate analysis for the dataset. The table shows counts, unique values, and top values for various features. The 'Price Direction Up' and 'Past Information' columns show a distribution of 0 and 1 values.</p>
Bivariate Analysis	 <p>Two bar charts showing the relationship between 'Price Sentiment' (1, 2, 3, 4) and 'Price Direction Up' (0, 1) and 'Past Information' (0, 1). The charts show that sentiment 2 is most frequent for both direction up and past information.</p>

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
```

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