Data Collection and Preprocessing Phase

Date	11 July 2024
Team ID	SWTID1720359900
Project Title	Machine Learning Approach for Predicting The Price Of Natural Gas
Maximum Marks	6 Marks

Data Exploration and Preprocessing

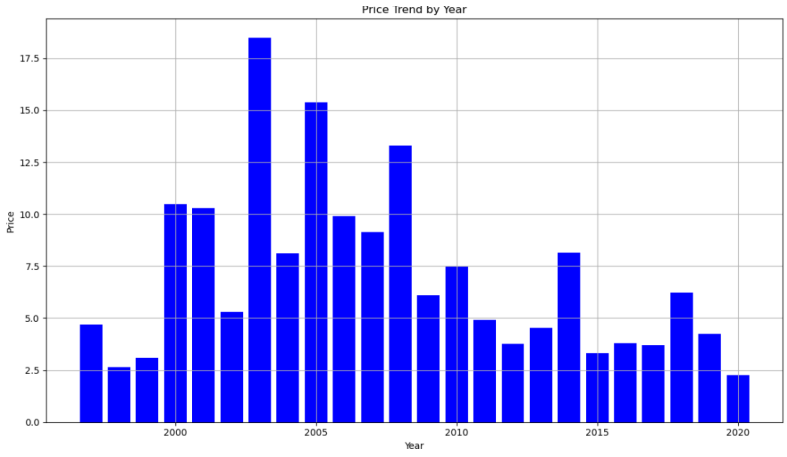
Dataset variables will be statistically analyzed to identify patterns and outliers, with Python employed for preprocessing tasks like normalization and feature engineering. Data cleaning will address missing values and outliers, ensuring quality for subsequent analysis and modeling, and forming a strong foundation for insights and predictions.

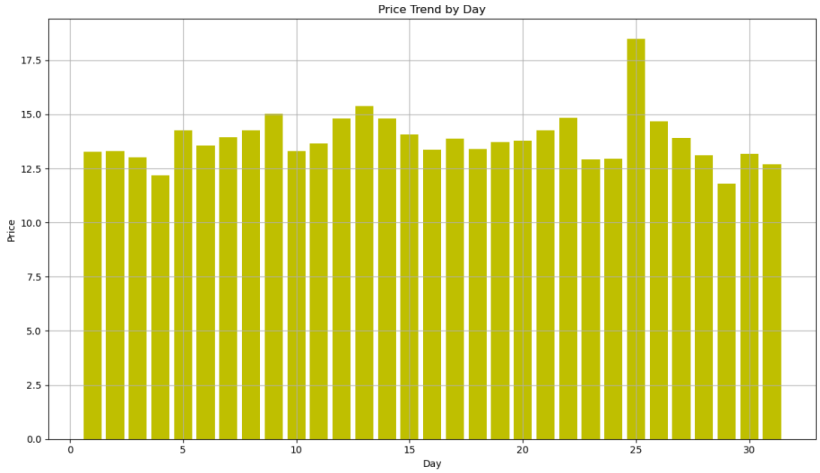
Section	Description
Data Overview	[6]: Date Price 0 1997-01-07 3.82 1 1997-01-08 3.80 2 1997-01-09 3.61 3 1997-01-10 3.92 4 1997-01-13 4.00 [5938 rows x 2 columns]>

Univariate Analysis	

	Price
count	5937.00000
mean	4.18923
std	2.19121
min	1.05000
25%	2.66000
50%	3.54000
75%	5.24000
max	18.48000

Bivariate Analysis





Data Preprocessing Code Screenshots

Loading Data

```
df.head()
```

	Date	Price
0	1997-01-07	3.82
1	1997-01-08	3.80
2	1997-01-09	3.61
3	1997-01-10	3.92
4	1997-01-13	4.00

<table>
<tr><td>Handling Missing Data</td><td>

```
df = pd.DataFrame(data)
missing_data_rows = df[df.isna().any(axis=1)]
print(missing_data_rows)
```

```
            Date  Price
5284  2018-01-05    NaN
```

```
df_2 = df.dropna()
```

```
missing_data_rows = df_2[df_2.isna().any(axis=1)]
print(missing_data_rows)
```

```
Empty DataFrame
Columns: [Date, Price]
Index: []
```

</td></tr>
<tr><td>Data Transformation</td><td>

```
[ ]: df_2['Date'] = pd.to_datetime(df_2['Date'], infer_datetime_format=True)

     # Extract year, month, and day from the 'Date' column
     df_2['year'] = df_2['Date'].dt.year
     df_2['month'] = df_2['Date'].dt.month
     df_2['day'] = df_2['Date'].dt.day

     # Drop the original 'Date' column if not needed
     df_2.drop(columns=['Date'], inplace=True)

     # Display the updated DataFrame
     print(df_2.head())
```

```
   Price  year  month  day
0   3.82  1997      1    7
1   3.80  1997      1    8
2   3.61  1997      1    9
3   3.92  1997      1   10
4   4.00  1997      1   13
```

</td></tr>
<tr><td>Feature Engineering</td><td>Attached the codes in final submission.</td></tr>
<tr><td>Save Processed Data</td><td>-</td></tr>
</table>