

## Data Collection and Preprocessing Phase

Date	15 july 2024
Team ID	740032
Project Title	Price prediction of natural gas using machine learning approach.
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	Basic statistics, dimensions, and structure of the data.
Univariate Analysis	Exploration of individual variables (missing values).
Bivariate Analysis	Relationships between two variables (boxplot, scatter plots).
Multivariate Analysis	Patterns and relationships involving multiple variables.

<p>Outliers and Anomalies</p> <p><b>Data Preprocessing</b></p> <p><b>Code Screenshots</b></p>	<p>Identification and treatment of outliers.</p>
<p>Loading Data</p>	<p>✓ loading the dataset</p> <pre>[ ] data=pd.read_csv('/content/daily_csv.csv')</pre>
<p>Handling Missing Data</p>	<p>✓ checking null values and filling missing values</p> <pre>[ ] data.isnull().any()</pre> <pre>↔ Date      False    Price     True    dtype: bool</pre>
<p>Data Transformation</p>	<p>finding outliers</p> <p><math>IQR=q3-q1, upperbound=q3+1.5IQR, lowerbound=q1-1.5IQR</math></p> <pre>[ ] IQR=data['Price'].quantile(0.75)-data['Price'].quantile(0.25)    upperbound=data['Price'].quantile(0.75)+1.5*IQR</pre> <pre>[ ] IQR ↔ 2.58</pre> <pre>[ ] Lowerbound=data['Price'].quantile(0.25)-1.5*IQR    Lowerbound</pre> <pre>↔ -1.21</pre> <pre>[ ] upperbound=data['Price'].quantile(0.75)+1.5*IQR    upperbound</pre> <pre>↔ 9.11</pre>
<p>Feature Engineering</p>	<p>✓ Splitting the data</p> <pre>[ ] from sklearn.model_selection import train_test_split    x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2,random_state=0)</pre>
<p>Save Processed Data</p>	<p>✓ training the model with decision tree</p> <pre>[ ] import numpy as np    import pandas as pd # Use pandas as pd, not np    import matplotlib.pyplot as plt    import seaborn as sns    from sklearn.model_selection import train_test_split    from sklearn.tree import DecisionTreeRegressor</pre>

