**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 3/7/24 |
| Team ID | SWTID1720100721 |
| Project Title | Machine Learning Approach To Predict Price of Natural Gas |
| Maximum Marks | 6 Marks |

|  |  |
| --- | --- |
| **Section** | **Description** |
| Data Overview | Basic statistics, dimensions, and structure of the data. For example, number of rows, columns, data types, summary statistics (mean, median, standard deviation), and missing values percentages. |
| Univariate Analysis | Exploration of individual variables (mean, median, mode, etc.). For example, histograms, box plots, frequency distributions, and descriptive statistics for each variable. |
| Bivariate Analysis | Relationships between two variables (correlation, scatter plots). For example, correlation matrix, scatter plots, and pairwise relationships between variables. |
| Multivariate Analysis | Patterns and relationships involving multiple variables. For example, principal component analysis , clustering, and other dimensionality reduction techniques to visualize and understand complex relationships in the data. |
| Outliers and Anomalies | Identification and treatment of outliers. For example, box plots, z-score, or other outlier detection techniques. Imputation or removal of outliers based on the analysis. |
| **Data Preprocessing Code Screenshots** | |
| Loading Data |  |
| Handling Missing Data |  |
| Data Transformation |  |
| Feature Engineering |  |
| Save Processed Data |  |