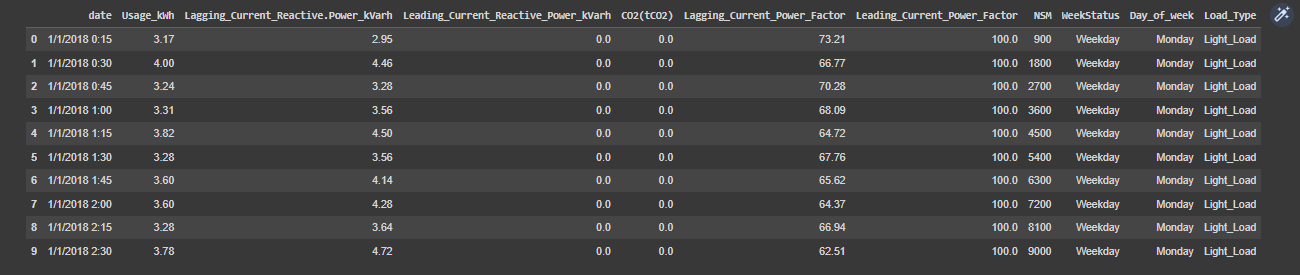
Energy Consumption Prediction

Using

Machine Learning

Name: Ali Rafeeq Amer

First Read Data

Text

Description automatically generated1 Data Preparation

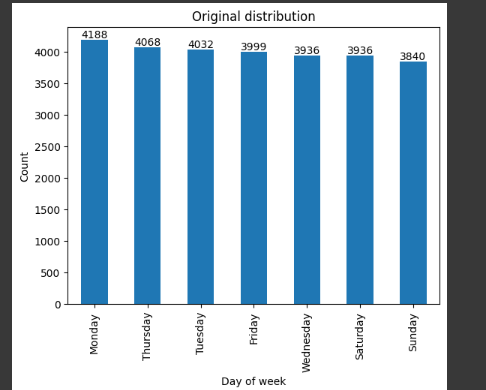
1.1 Missing Data

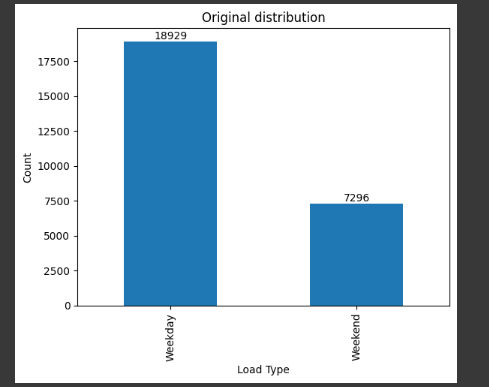
1. Leading\_Current\_Reactive\_Power\_kVarh' Null we fill it by (ffill)

The forward fill (ffill) method is a type of missing value imputation technique that propagates the last observed non-null value forward until the next non-null value is encountered. In other words, it fills the missing values with the most recently observed value.

2- Leading\_Current\_Power\_Factor' Null we fill it by (ffill)

The forward fill (ffill) method is a type of missing value imputation technique that propagates the last observed non-null value forward until the next non-null value is encountered. In other words, it fills the missing values with the most recently observed value.

Chart, bar chart

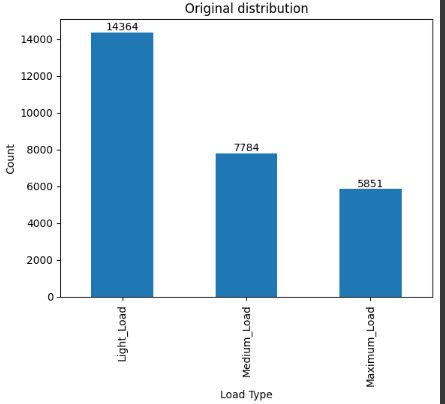
Description automatically generated

Before missing values

After missing values fill

3- By filling missing values in the 'Day\_of\_week' column using the 'date' column, the code is ensuring that the 'Day\_of\_week' column is consistent with the dates provided in the 'date' column. This can be useful in time series analysis or when analyzing trends by day of the week.

3- 'Weekday', the code is filling all missing values in the 'WeekStatus' column with the most frequent value in the column, which is 'Weekday'. This is a common approach for dealing with missing values, especially when the missing values are few compared to the total number of observations in the dataset.



3- Load\_Type , the code is filling all missing values in the Load\_Type column with the most frequent value in the column, which is 'Weekday'. This is a common approach for dealing with missing values, especially when the missing values are few compared to the total number of observations in the dataset.

1.2 Feature Engineering

Handling Datetime column

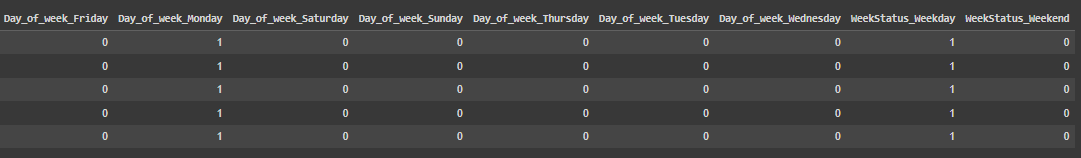
1. Converting the 'date' column to a datetime format: The first line of code uses the 'to\_datetime' method from Pandas to convert the 'date' column in 'train\_data' from a string format to a datetime format.
2. Calculating the Unix timestamp in seconds: The second line of code calculates the Unix timestamp in seconds by subtracting the datetime of January 1st, 1970 from the 'date' column, which is now in datetime format, and then taking the total number of seconds using the 'dt.total\_seconds()' method. The Unix timestamp is a widely used way to represent a point in time as the number of seconds that have elapsed since January 1st, 1970, at 00:00:00 UTC.
3. Dropping the original datetime column if desired: The third line of code drops the original 'date' column from the 'train\_data' DataFrame using the 'drop' method and specifying 'axis=1' to indicate that the column should be dropped along the columns axis. This is an optional step, and it depends on whether or not the original 'date' column is needed for further analysis or modeling.

Graphical user interface, text

Description automatically generated

Encoding the categorical columns.

in Day\_of\_week column and WeekStatus column i use One-hot encoding

n this case, it is applied twice to create binary columns for 'Day\_of\_week' and 'WeekStatus' separately. The original categorical columns are dropped from the DataFrame and replaced with the binary columns.

In Load\_Type column I use LabelEncoder

By transforming categorical variables into numerical values, the LabelEncoder makes it easier for machine learning algorithms to process the data. However, it is important to note that this transformation does not add any meaning or order to the categorical variables

Machine Learning Model

1 -linear regression

* Text

  Description automatically generatedTrain

No Hyperparameter Tuning

With Hyperparameter Tuning

Text

Description automatically generated

* Text

  Description automatically generatedTest

No Hyperparameter Tuning

With Hyperparameter Tuning

Text

Description automatically generated

2 -Ridge

* Train

Text

Description automatically generated

No Hyperparameter Tuning

With Hyperparameter Tuning

Text

Description automatically generated

* Text

  Description automatically generatedTest

No Hyperparameter Tuning

Text

Description automatically generated

With Hyperparameter Tuning



Chart, line chart

Description automatically generated

The Learning Curve in Ridge Model

3 -Lasso

* Graphical user interface, text

  Description automatically generatedTrain

No Hyperparameter Tuning

Text

Description automatically generated



With Hyperparameter Tuning

* Test

Text

Description automatically generated

No Hyperparameter Tuning

With Hyperparameter Tuning

Text

Description automatically generated

Chart, line chart

Description automatically generated

The Learning Curve in Lasso Model