**Assignment-4 :**

**Prediction of Electricity Price using Regression Analysis**

**Team Leader**

Lakshanya S

**Team Members**

A. Bhavya Sri

J. Sinduja

K. Nanda Kishore Reddi

Ejumalla Saikiran

Naveen Rampa

Submission Date:

24-12-2023

**1. Initial Exploration**

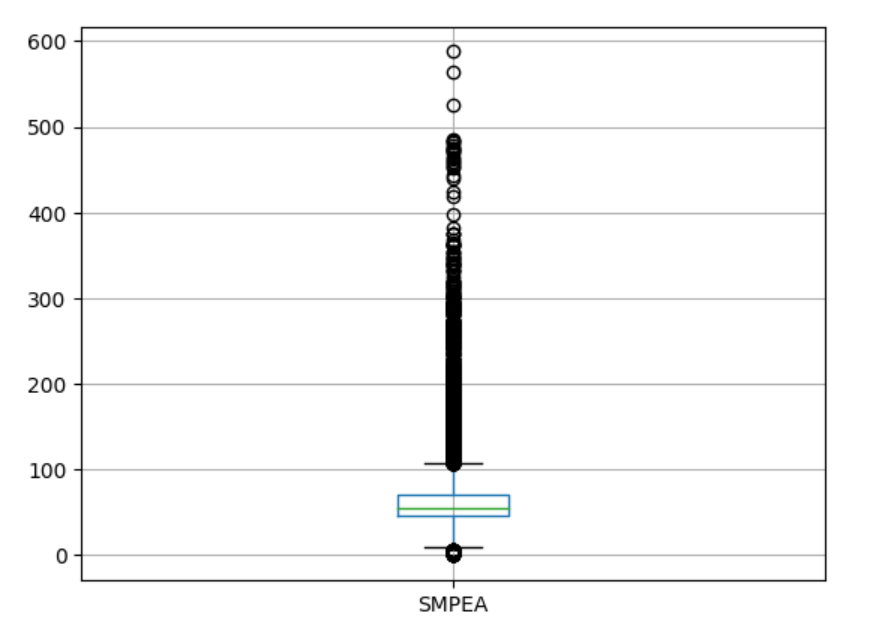
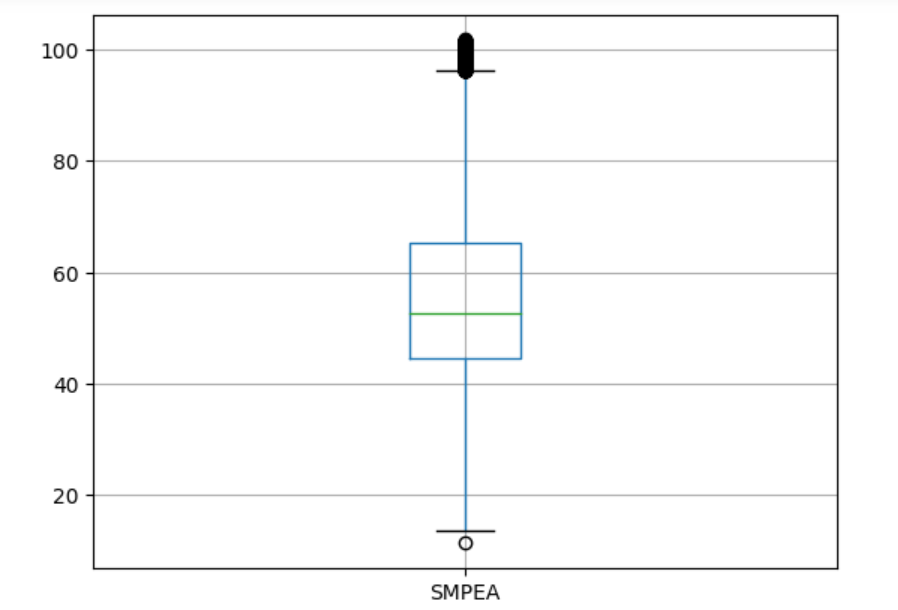
* Loaded the dataset and conducted an initial exploration using describe() and info().
* Missing values were identified in the dataset, and they were imputed by replacing them with the average values of their respective features.

**2. Exploratory Data Analysis**

2.1 Identifying and removing Outliers

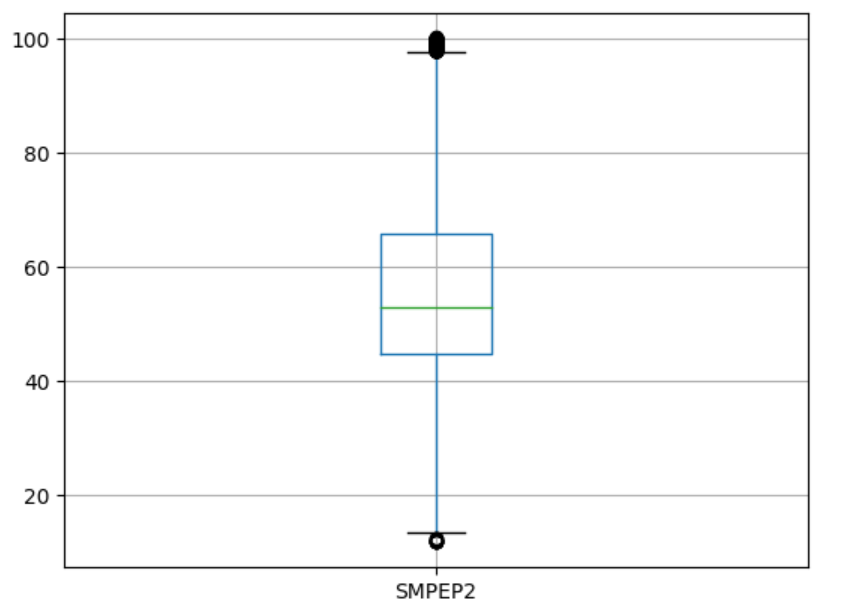
Utilizing data visualization methods, such as box plots, offered a graphical depiction of data points exhibiting notable deviations from the overall pattern. The visual representations below illustrate the presence of outliers in the dataset before and after after removing them using IQR method.

I.SMPEA

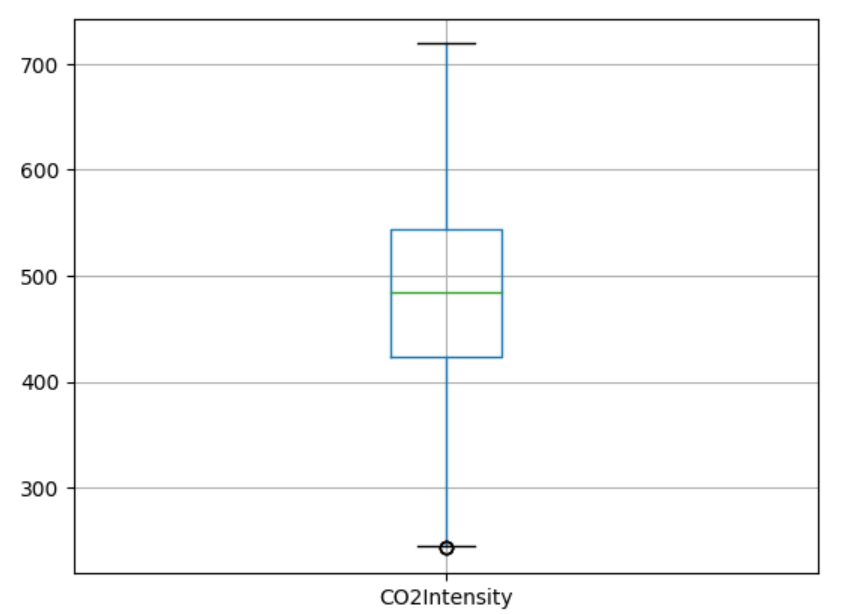
II.SMPEA2

A graph with lines and dots

Description automatically generated 

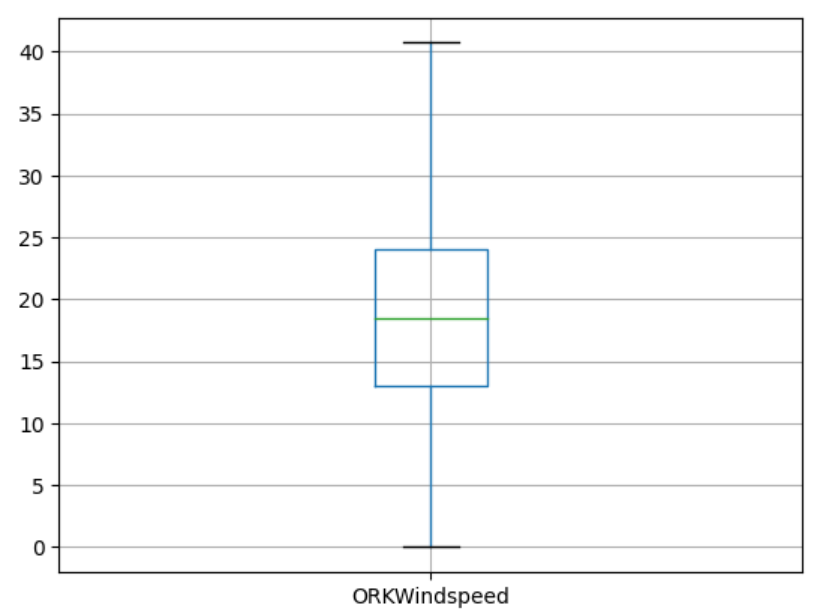
III.CO2Intensity

A graph with a blue rectangle and green rectangle

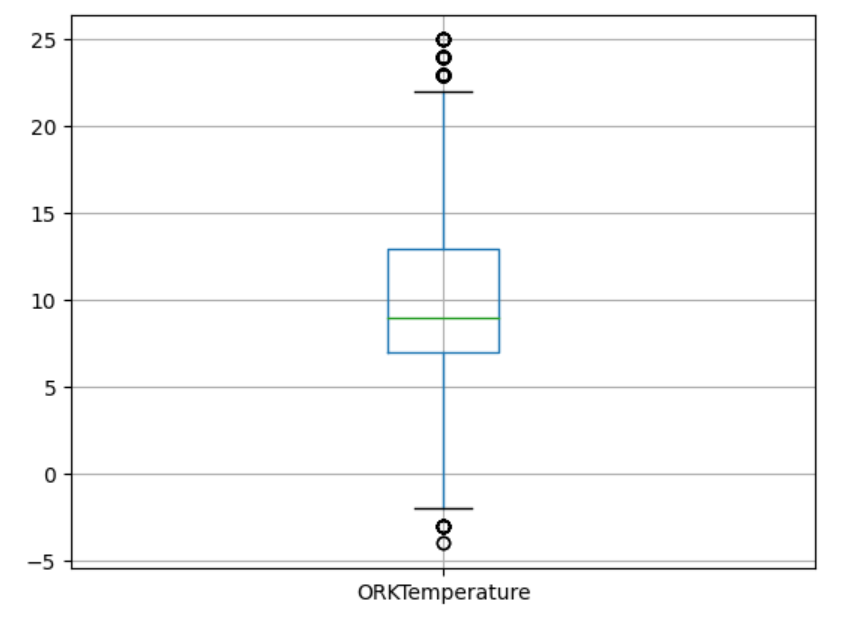
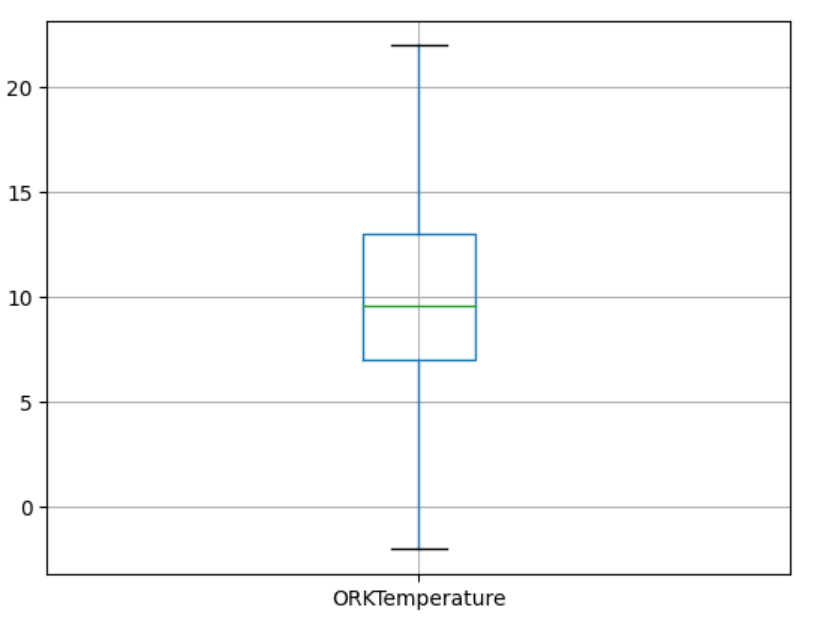
Description automatically generated 

IV.ORKWindspeed

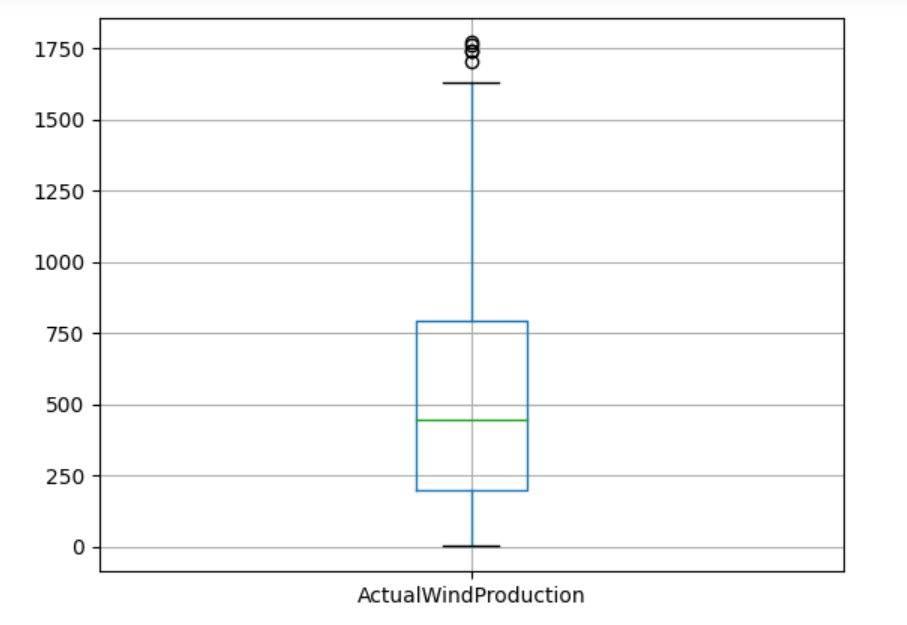
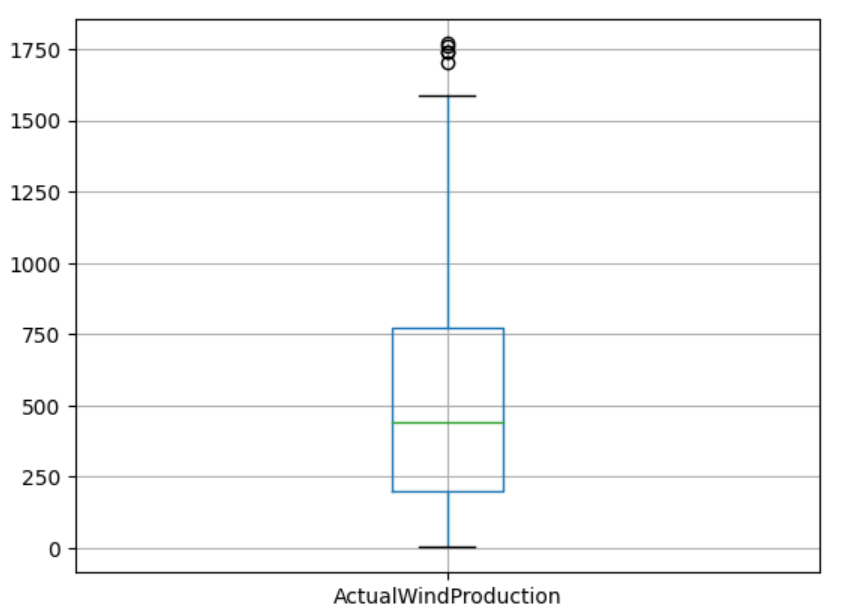
A graph of a wind speed

Description automatically generated 

V.ORKTemperature

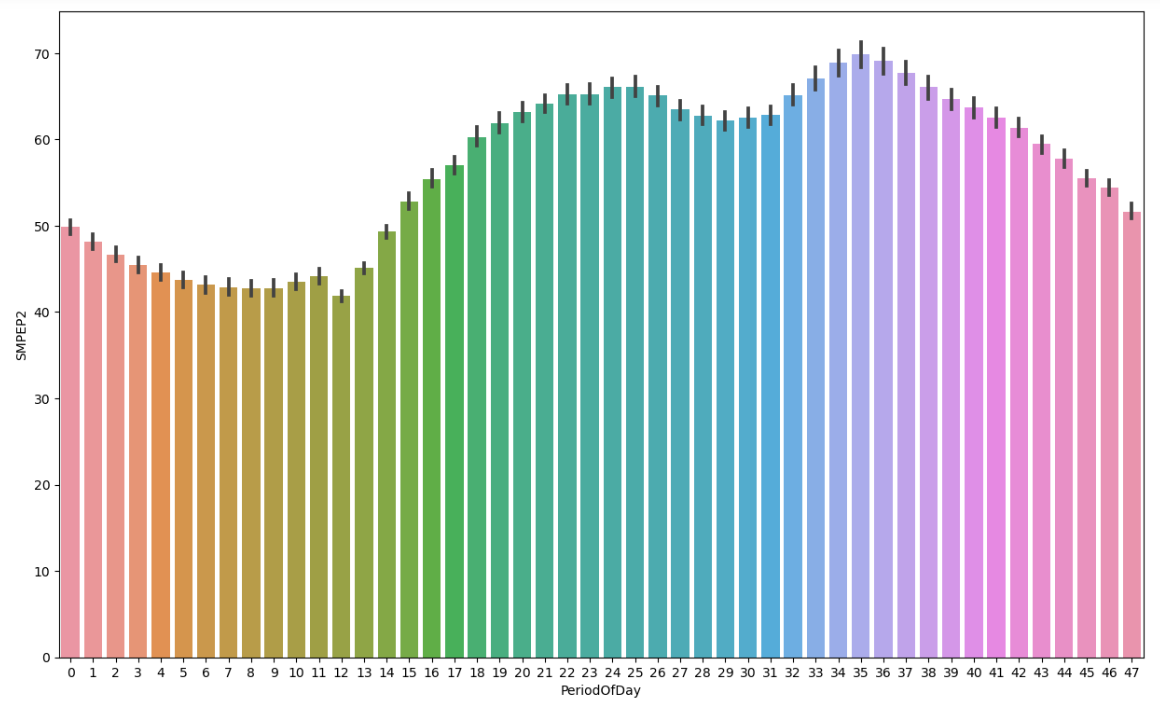
 

VI.ActualWindProduction

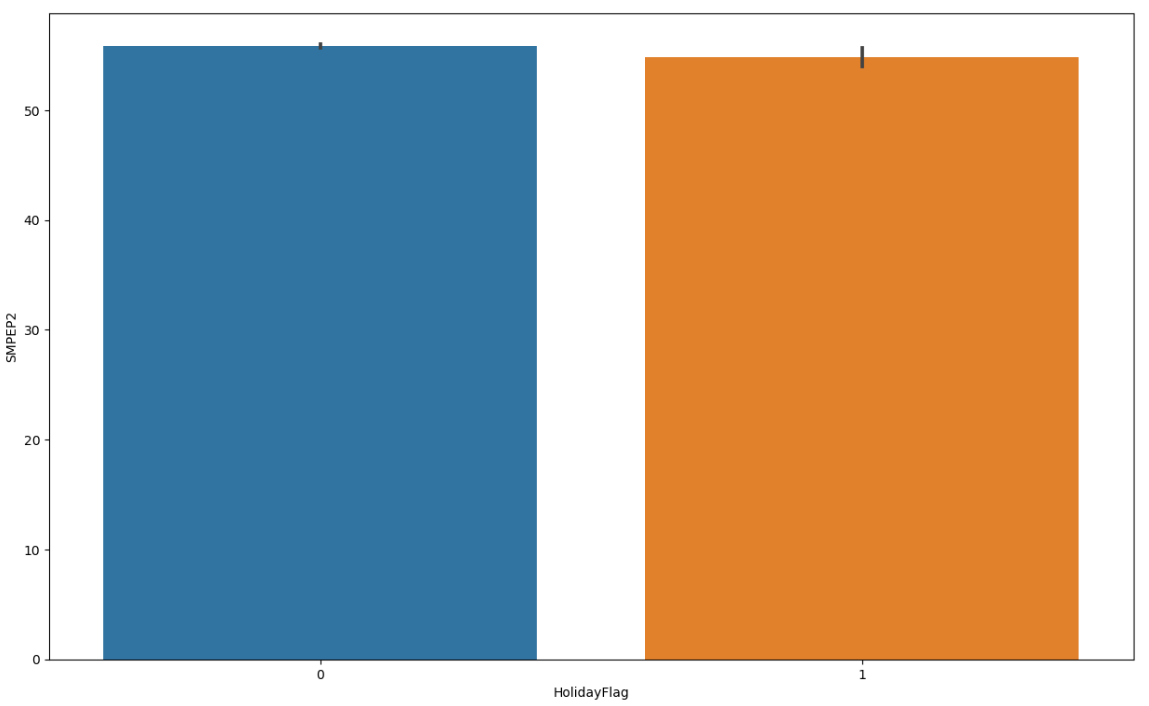
2.2 A Time-of-Day Analysis

In this visualization, we examine the fluctuation in electricity prices over different periods of the day, aiming to gain insights into the average pricing trends for every 30-minute interval. The x-axis represents distinct time intervals, segmented throughout a 24-hour day, while the y-axis illustrates the mean electricity prices during these intervals.



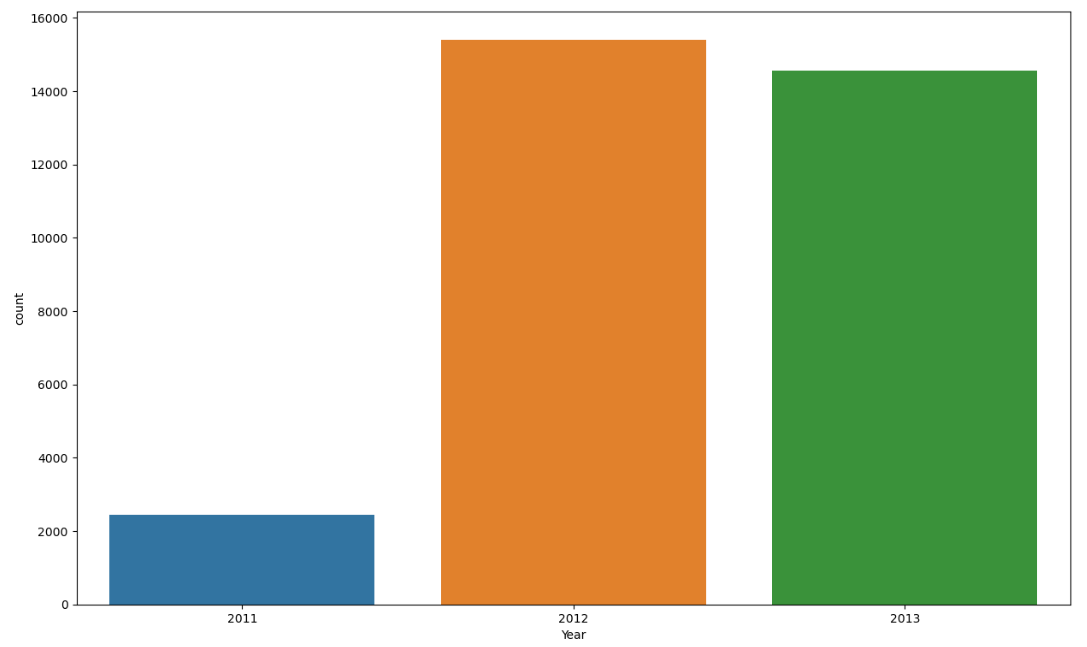
2.3 Bank Holiday vs. Non-Holiday Electricity Pricing

The data, derived for our electricity pricing analysis, reveals that there is minimal variation in electricity prices between these two categories. This observation suggests that the presence or absence of a bank holiday does not significantly impact electricity prices.

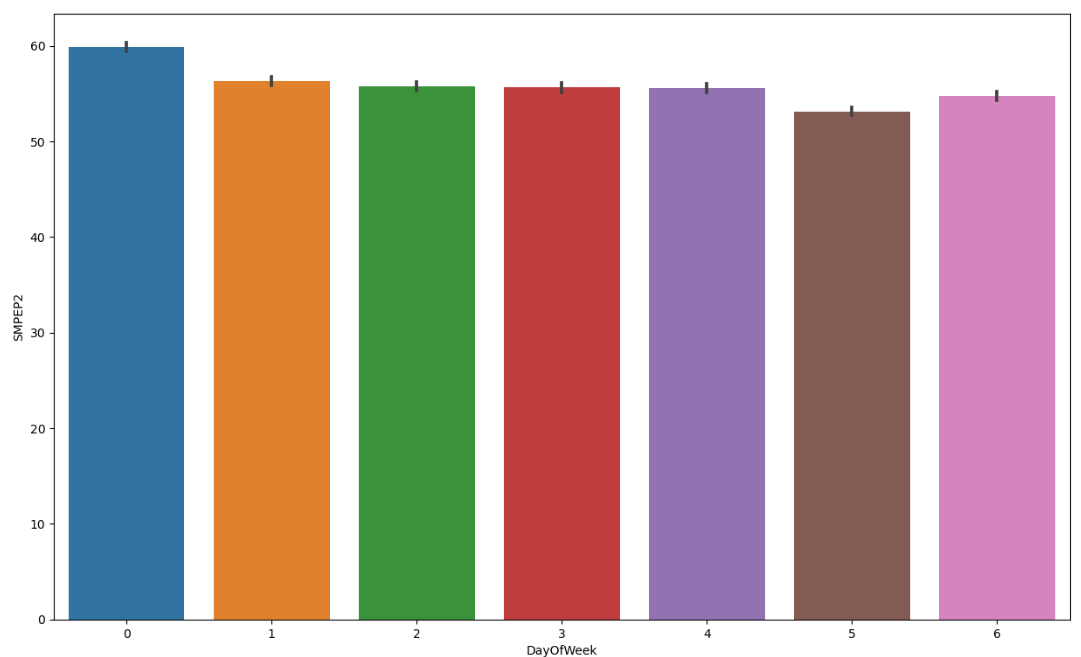


2.4 Distribution of Data Across Years

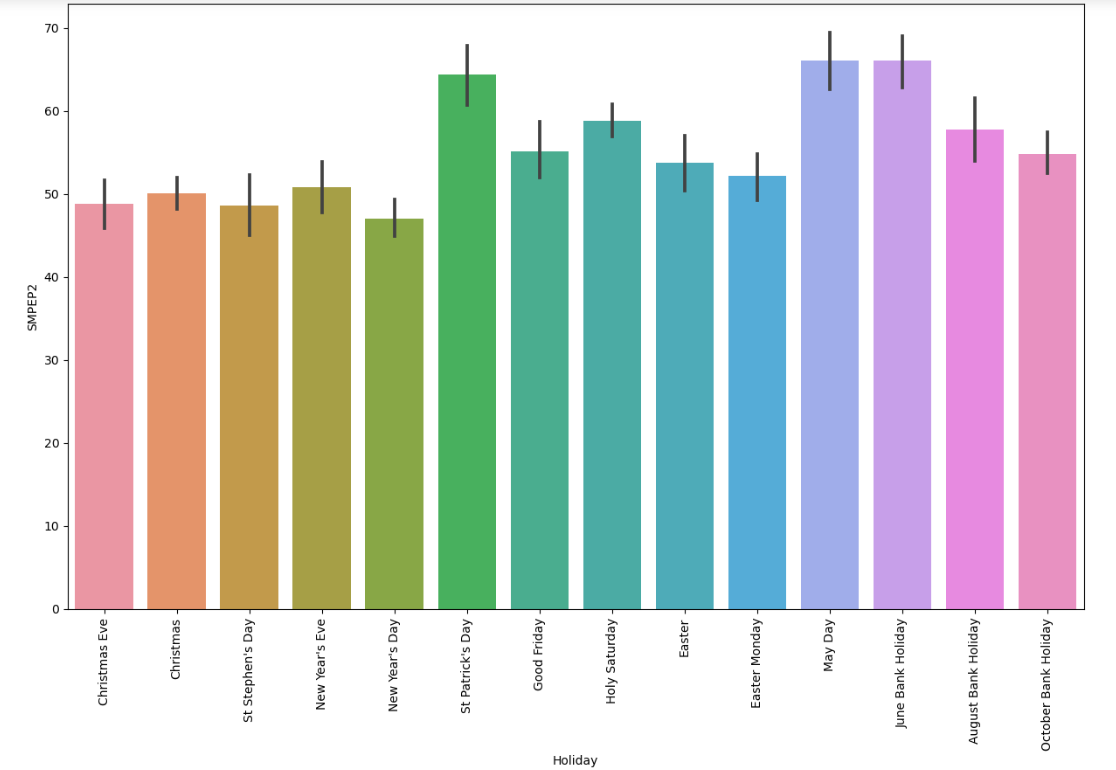
This count plot illustrates the frequency distribution of the dataset across three years: 2011, 2012, and 2013. Strikingly, the data is most prevalent for the year 2012, followed by 2013, with 2011 trailing behind. This visual representation highlights the concentration of observations in 2012, underscoring its significance in the dataset compared to the other two years.



2.5 Weekly Price Pattern  
In this analysis, we scrutinize the average electricity prices over a period of time, categorizing the data by days of the week.Notably, Mondays exhibit a slightly higher average price compared to other weekdays, suggesting a potential premium associated with the beginning of the workweek. Conversely, Saturdays showcase a modest dip in average prices, indicating a subtle decrease in comparison to other days of the week.

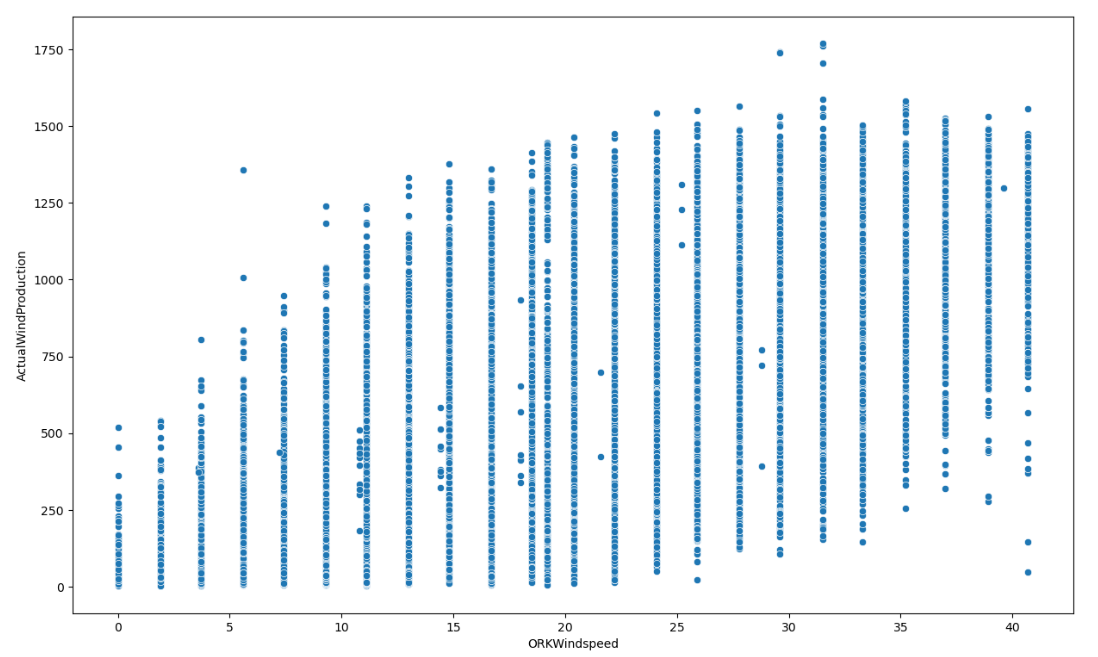


2.6 Analysis of Energy Consumption on Public Holidays  
The average energy consumption during certain holidays varies, suggesting potential impacts on energy demand during these periods.Notably, holidays like "June Bank Holiday," "May Day," and "St Patrick's Day" have relatively high average energy consumption, indicating increased energy usage or demand during these celebrations.

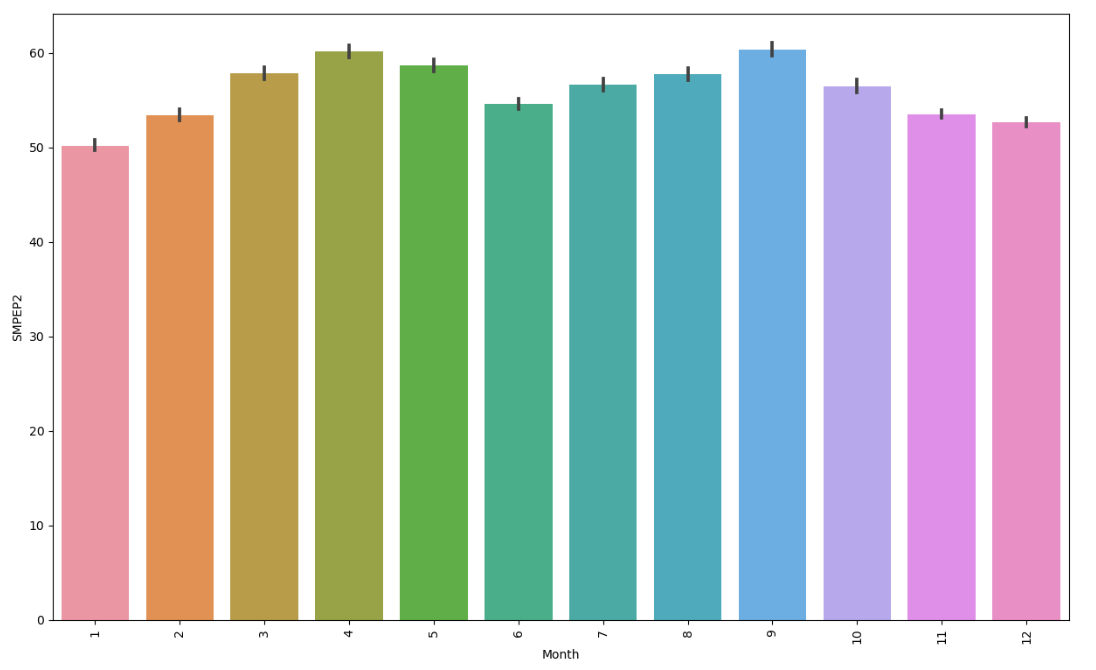


2.7 Analyzing Positive Correlation: Wind Speed and Wind Production

There is a clear positive correlation between wind speed and wind production. As the wind speed increases, the wind production also increases.There are a few outliers in the data, where the wind production is much higher or lower than what would be expected based on the wind speed. These outliers could be due to a number of factors, such as measurement errors, sudden changes in wind direction, or maintenance issues with the wind turbines.

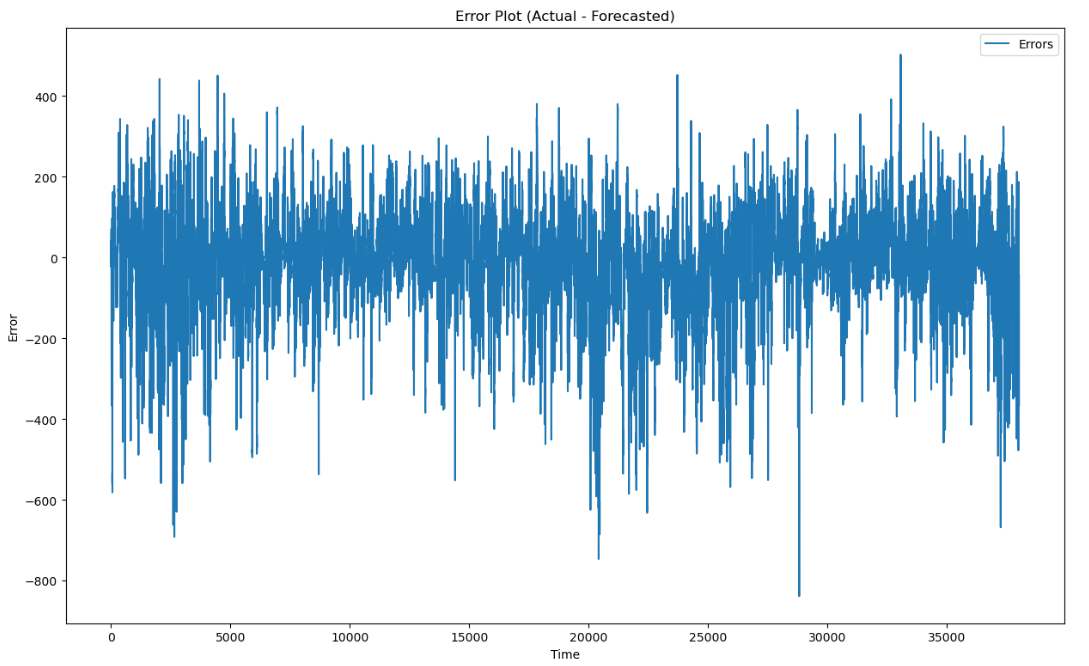


2.8 Monthly Average Energy Consumption

This analysis examines the monthly average energy consumption patterns, shedding light on variations in demand throughout the year. Key findings reveal distinct seasonal trends, with higher consumption during colder months, particularly in December and January. The moderate consumption observed in autumn and spring suggests transitional periods with milder temperatures. Notably, energy usage remains relatively stable and lower during the summer months, reflecting reduced heating demands and potential energy-saving practices.

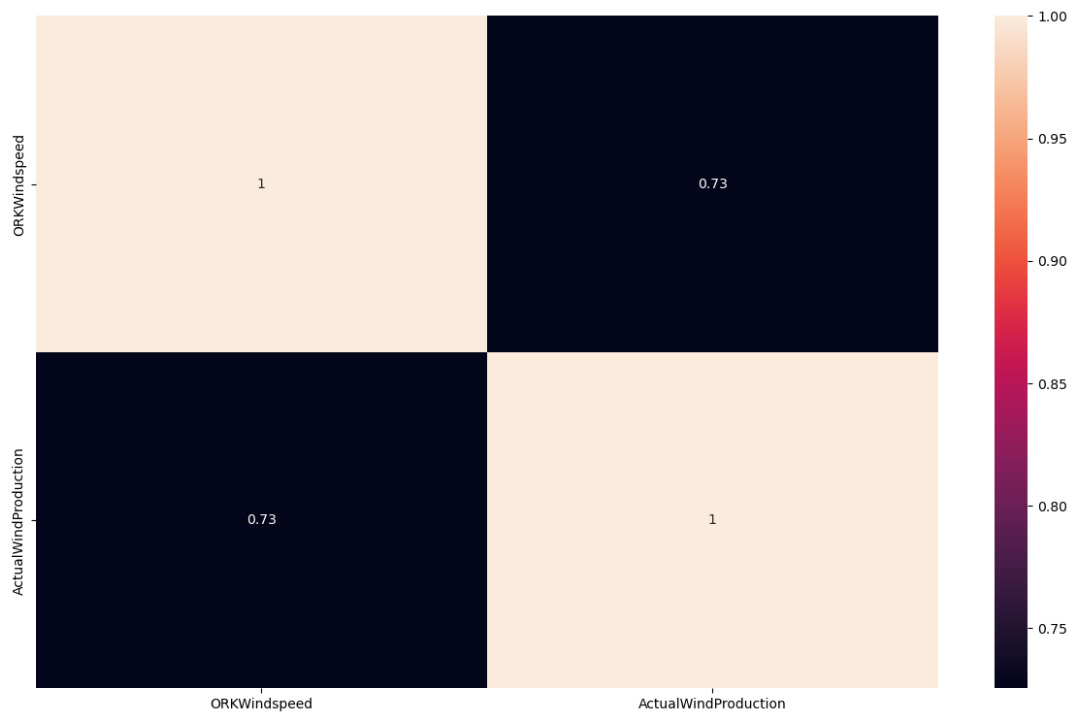
2.9 Error Analysis: Fluctuations and Outliers in Model Performance

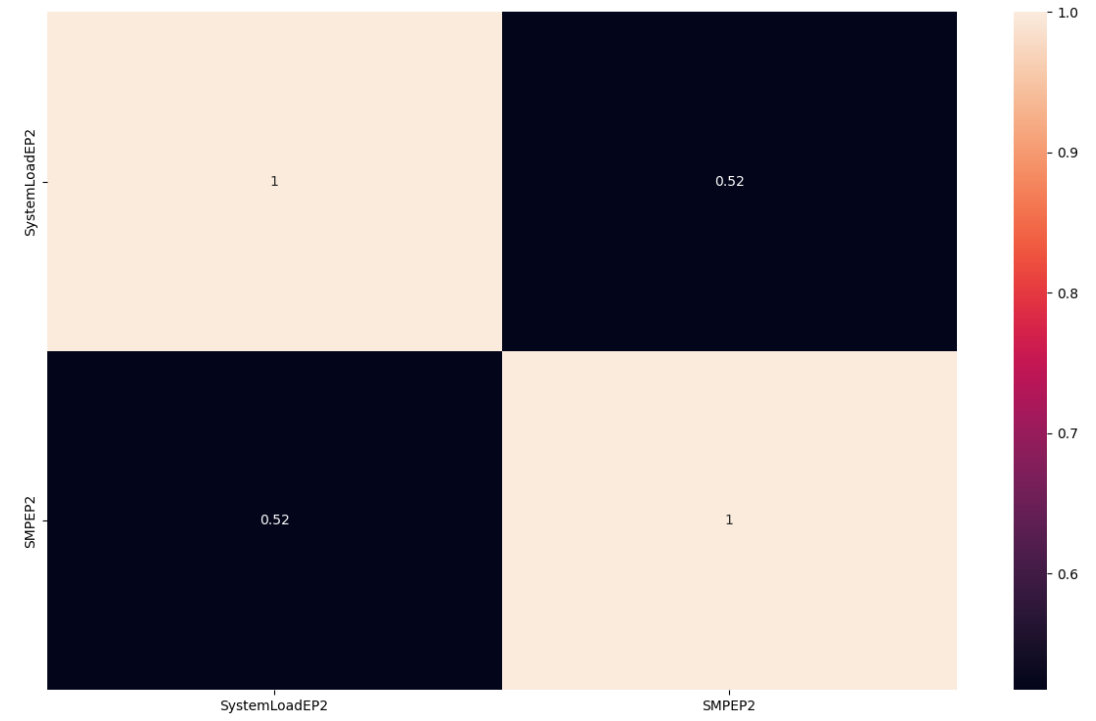
The errors fluctuate around 0 over time, indicating that the model is not consistently biased in either direction.There are large spikes in the errors, both positive and negative. These outliers could be due to a number of factors, such as inaccurate data points, unexpected events in the time series, or limitations in the forecasting model.

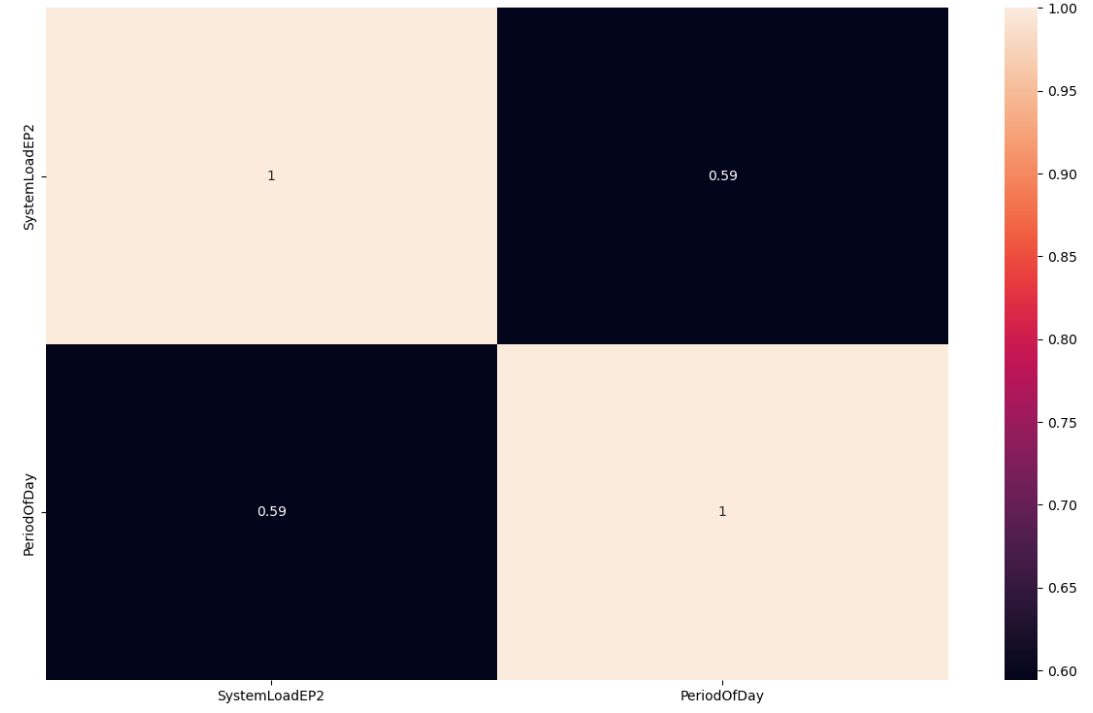


2.9 Finding Correlations

The below columns show the highest correlations indicating it could effect the price of electricity highly.







1. **Regression Model**
   1. Splitting and Standardizing

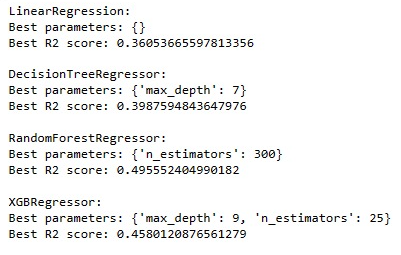
First, we segregate our features, such as consumption patterns, time of day, and other relevant factors, from the target variable we aim to predict—the electricity price. Subsequently, we divide our dataset into two segments: one dedicated to training our prediction model and the other to assess the model's performance.

* 1. Regression Model Comparison and Optimization   
     Linear Regression: Provides a reasonable baseline but may benefit from more sophisticated models.

DecisionTreeRegressor: Shows improvement over Linear Regression, emphasizing the importance of considering the depth of the tree.

RandomForestRegressor: Performs significantly better, capturing a substantial portion of price variability with 300 trees.

XGBRegressor: Strikes a balance between model complexity and accuracy, achieving a commendable R2 score. The choice among models depends on the desired trade-off between complexity and predictive performance in the context of electricity price prediction.



**4.Conclusion**  
In conclusion, when prioritizing the accuracy of predictions and the ability to capture intricate patterns in the data, the RandomForest Regressor emerges as the preferred model in this analysis. It outperforms other models and offers a more comprehensive understanding of the relationships between the input features and the target variable in the training dataset.