Below is a list of the Python libraries that I utilized:

- numpy
- pandas
- matplotlib
- seaborn
- datetime
- statsmodels
- scipy
- Scikit -Learn

This report summarizes the ten questions from the third and final data analytics course activity. It includes the methods used, the results obtained, and my subjective assessment of each task.

Task 1:

I was given a link to access data on the amount of energy needed by the EirGrid system in 2014. My task was to load this data onto my computer and create a graph showing the changes in energy demand over time. I made sure to label the graph accurately and account for any missing data in the spreadsheet thanks to the following steps.

- I downloaded the provided csv file that consisted of a timeseries demand data and load it into my python working space as a data frame
- I checked if there was missing value and handled them using a linear interpolation method that estimate missing values between known data points
- I converted dates and joined it with the time as a timestamp using a pandas python library and the plotted the demand versus time

Results: The graph displayed in Figure 1 demonstrates the successful completion of the task, depicting the load demand throughout 2014 in fifteen-minute increments. The data reveals that from October to March, there was a higher demand for energy, while from April to September, the demand was lower.

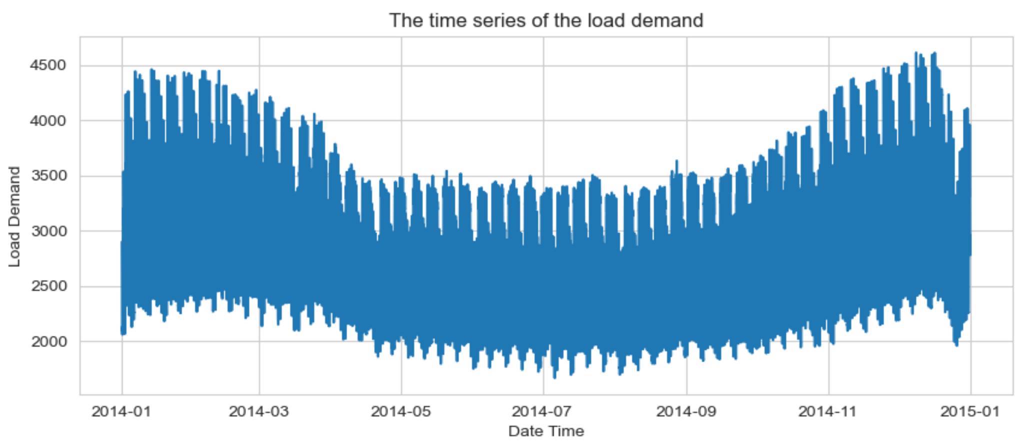

Figure 1: The energy consumers required in 2014, this energy was needed from EIR Grid system

Insight: EirGrid is a company that manages the electricity supply in Ireland [1]. I think that during the summer months, people use less electricity because it is warmer outside, and they don't need to use heaters and lights as much. Also, many people go on vacation during the summer, so they use less electricity overall.

Task 2

The task to be performed involved calculating the autocorrelation coefficients for a period of 10 days and plotting the autocorrelation against the lag. The x-axis of the plot should be labelled in days, and it is also important to provide a clear labelling for the y-axis.

- I calculated the autocorrelation of each 15-minute electricity demand up to 10 days using 'autocorr()' python function
- I represented the autocorrelation of the 10 days electrical demand using a statsmodels python library

Results: The chart presented below displays the autocorrelation with a 10-day lag that was calculated for this task. Furthermore, the autocorrelation computed was found to be 0.8125, indicating a robust positive correlation between the demand values for these 10 consecutive days in the time series.

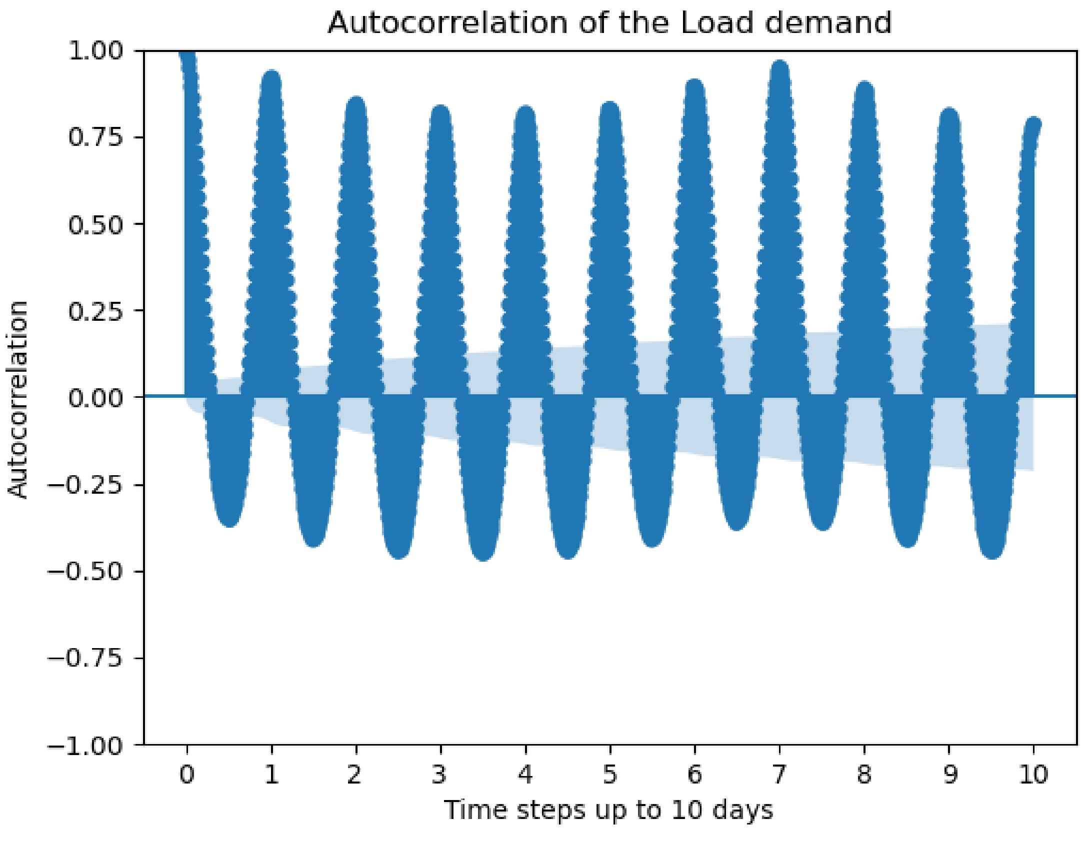

Figure 2: The autocorrelation of 2014 EIR load demand

Insight: The graph displays repeated patterns occurring at a seven-day interval, which indicates a correlation between each occurrence. This suggests that there may be weekly seasonality, meaning that there could be a daily pattern in the data that repeats itself over a seven-day period.

Task 3

For this task, you are asked to create a time of year variable that ranges from 0 to 1 and use it to plot the demand changes throughout the year. This will involve creating a graphical representation that shows how demand varies over time.

- To plot this I generated 35,040 numbers(which is the length of the data frame) between 0 and one and then I plotted the energy demand against those generated values.

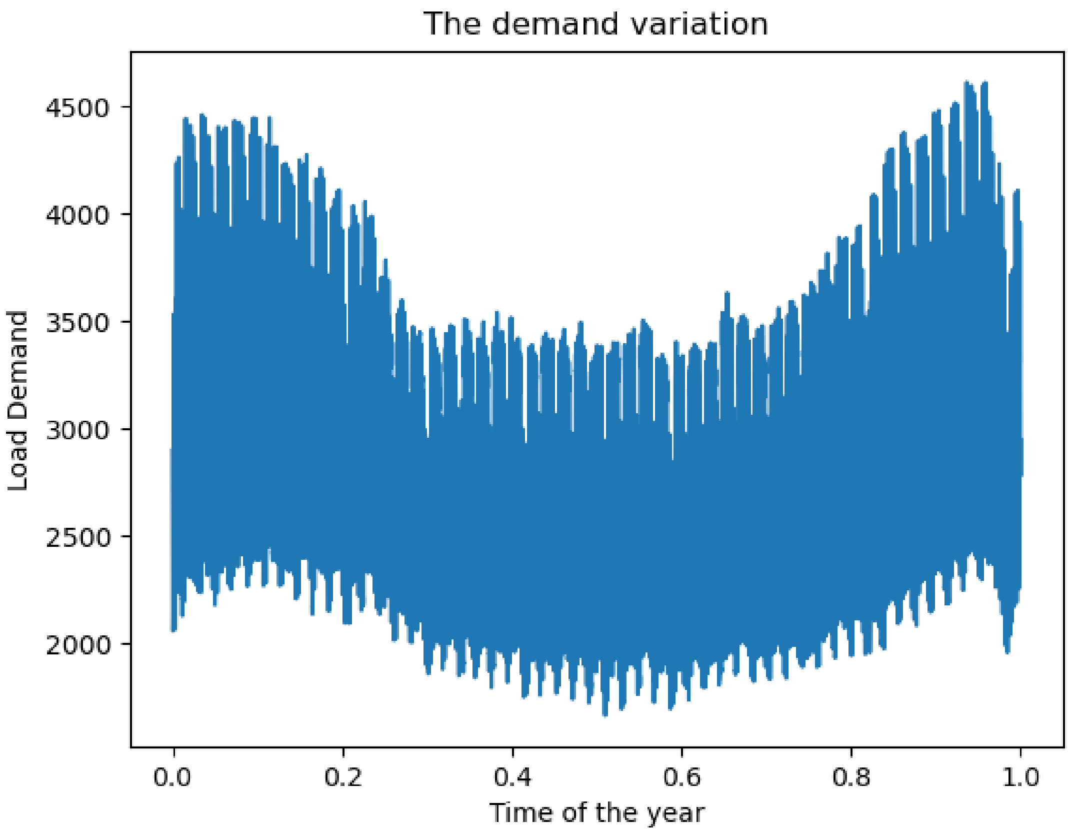

Figure 3:The EIR Load demand variation over the 2014 year

Task 4

The main aim of this task is to determine the average demand for electricity during each of the twelve months of the year. This information has been obtained presented in the form of a bar chart that is properly labelled thanks to the following steps.

- I resampled the data we had for every fifteen-minutes and aggregated them into monthly data using a panda's function "resample"
- The I plotted the monthly aggregated energy demand as a bar graph using seaborn python library

Results: The graph in Figure 4 displays the average monthly demand profile for the year 2014. It reveals that the demand for electricity during the months of June, July, and August was lower compared to the other months, but it was consistent throughout these months. On the other hand, the demand for electricity was highest during the months of December, January, and February.

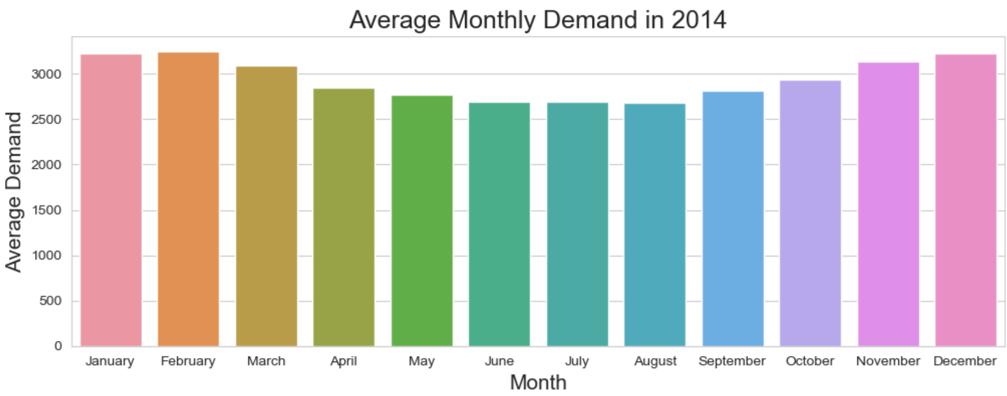

Figure 4: The monthly 2014 EIR Grid system demand profile

Insight: The data presented provides clear evidence that energy demand is lower during the summer season when compared to the winter season. One possible reason for the increased demand during winter could be that people tend to spend more time inside their homes during the colder months, which would require more power for heating and lighting purposes [2].

Task 5

To complete this task, I need to compute the average demand for each of the 24 hours of the day and represent them in a bar chart. This type of chart is known as the daily demand profile and helps to visualize how the demand for electricity varies throughout the day. Each hour of the day should be clearly labelled on the chart to aid in interpretation and the following steps contributed to the completion of this task.

- I resampled the energy demand for every fifteen-minute based on the hour and aggregated the demand with in 24 hours of a day
- I then plotted the daily average demand with respect to 24 hours of a day using seaborn python library

Result: The graph presented in Figure 5 displays the average daily demand profile for the Eir grid system in 2014. It is evident from the graph that the demand for electricity increases steadily from 6 am until 6 pm and then begins to decline gradually from 7 pm to 5 am the following day.

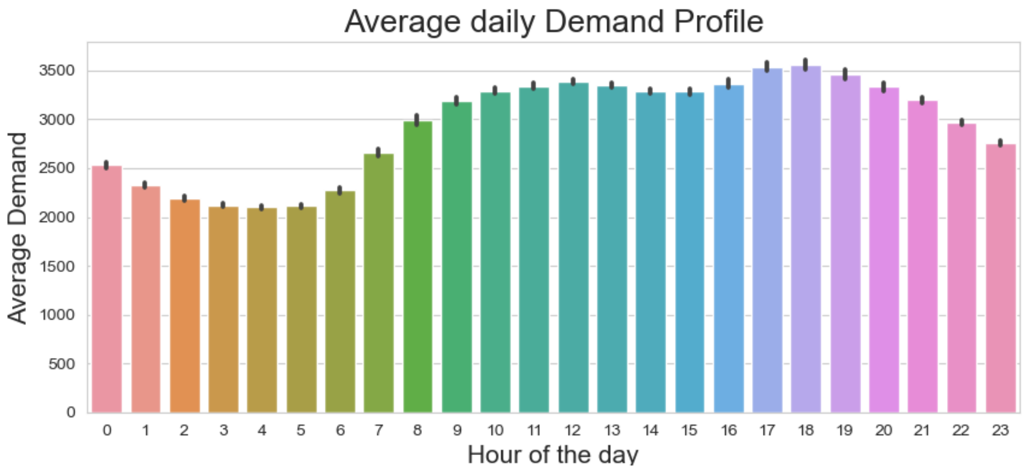

Figure 5:The average 2014 EIR Grid system demand of each of the 24 hours of the day

Insight: The data presented in the figure above illustrates that in 2014, the hours of highest electricity demand were typically between 4 pm and 6 pm. This trend may be attributed to various factors, including the heightened usage of lighting, heating, and cooking appliances in households during these hours, as well as increased energy demands from commercial and industrial sectors. By identifying these peak hours, energy providers can anticipate and manage electricity supply to meet the demands of consumers effectively [3].

Task6:

As part of the task, we need to calculate the average electricity demand for each day of the week and represent them using a bar chart. Subsequently, we are required to analyse the chart to assess whether the outcomes align with our expectations of energy usage. The following steps were used to complete this task:

- I resampled the given data as a daily average and the aggregated the energy demand for each day of the week
- I plotted the aggregated each day of the week average demand as a bar graph using seaborn python library to show the trend on each day

Result: As shown in the figure 6 below the result shows each day of the week demand profile in 2014. The demand was hire in the week days compared to the weekends as shown in the figure below.

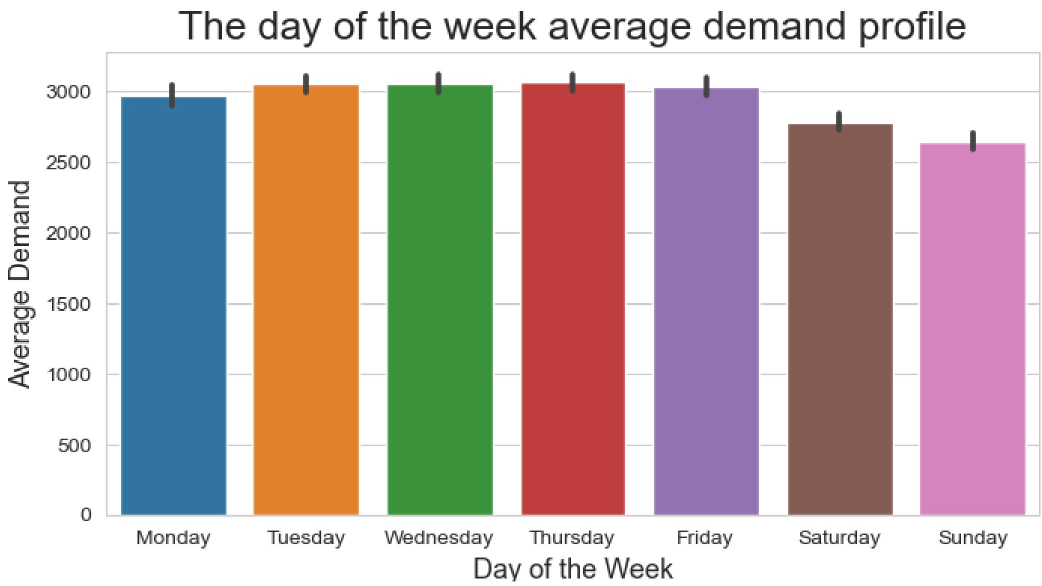

Figure 6: The average EIR Grid system demand of each day in the seven days of the week

Insight: The result does make sense because the weekend demand is typically lower than weekdays due to the decreased energy usage in commercial and industrial sectors. Usually, people use less electricity on weekends because they don't need as much energy for work. Also, they may spend more time outside, so they don't need to use as much electricity to light and heat their homes.

Task 7

The goal was to make a graph that shows the demand for each day of the week. We'll find the average demand during a certain hour for each day. Then, we'll plot each day's demand as a separate curve on the graph with respect to the following steps:

- I aggregated the given data set and given it an hour interval for each date of the year, each data was assigned with its corresponding day of the week
- I plotted the demand for every day of the week against the 24 hours of the day using seaborn python library

Results: According to Figure 7, the findings demonstrate that energy demands are elevated on weekdays such as Tuesdays, Wednesdays, Thursdays, Monday and Fridays throughout the day, but on Saturdays and Sundays, they remain lower throughout the day.

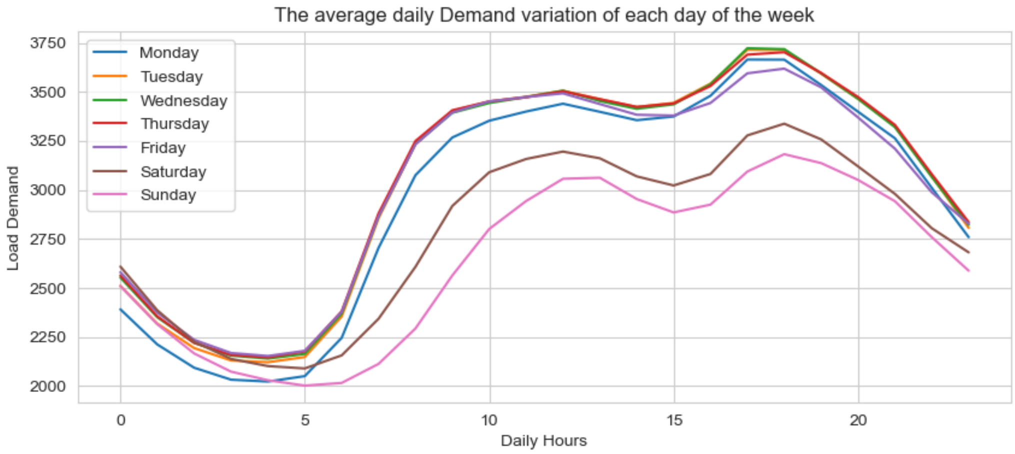

Figure 7: The daily variation of each day of the week load demand

Insight: The observed patterns in energy demand across weekdays and weekends can be attributed to a variety of factors such as work and school schedules, commercial activity, and human behaviours. All of these factors can work together to influence the overall increase in energy demand during weekdays compared to weekends.

Task 8:

This task required us to use a statistical hypothesis test, such as a t-test, to analyse the data and determine if there was a significant difference in the demand for a product or service between the weekend and the working days of the week. To reach a conclusion, we followed these steps.

- I assigned each row that consisted of a date and the demand of that day their corresponding day of the week
- I combined all the weekdays into one data frame and separated them from the weekend days.
- I performed a statistical hypothesis test of the two data frame using a python function" ttest_ind"

Results: the statistical hypothesis test was conducted in two ways, firstly I computed it on the data set had then I also computed by considering the daily average load demand. Both results showed that p-value is below the significant level, do the null hypothesis was rejected. By considering the whole data set the p-value obtained was 0, while considering the daily average load demand the obtained p-value was 0.00109.

Results by considering the given data set

```
T-statistic value:  46.54684714308396
P-Value:  0.0
```

By considering the daily average demand

```
T-statistic value:  6.732306617845197
P-Value:  0.0010963146723274575
```

Insight: Since the p-value is below the significant level $\alpha = 0.05$, we can reject the null hypothesis and conclude that there is a significant difference between the working day of the week and the weekends.

Task 9:

In this section we aimed to determine the average absolute deviation (MAE) in our forecasts and present a visual representation of the fluctuations in the MAE for various prediction periods, ranging up to 24 hours in advance. Our primary objective is to compute and analyse the error rates associated with our predictions and generate a graphical depiction of the variations in these rates over different time horizons.

- I divided the data set into two equal halves based on the index but before I calculated the persistence to forecast using the benchmark forecasting approach using a formula $y\_hat(\mathrm{t}+\mathrm{k})=\mathrm{y}(\mathrm{t})$ for 1 day horizon
- Then I divided the prediction into two halves and used the second half of the data as actual value and the second half of the predicted data as predicted value to calculate the mean absolute error
- Then the calculated the mean absolute error was plotted with respect to time

Results: The prediction of one day horizon was a success as shown in the figure below. Among the different time intervals, the prediction for one hour from now had the least amount of error, whereas the forecast for the mid-day hour had a higher degree of error.

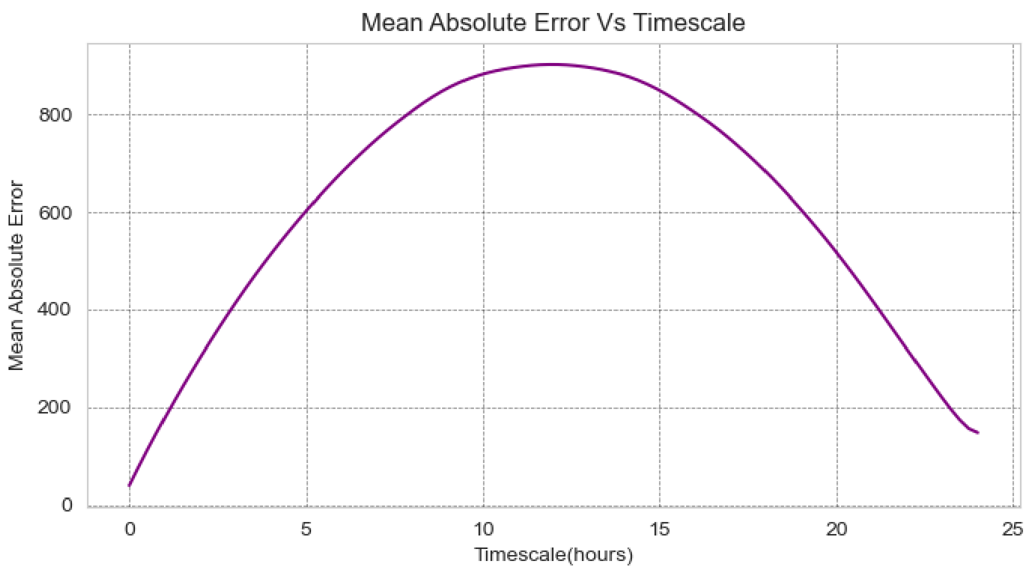

Figure 8: This figure shows the mean absolute error resulted in a simple benchmark forecast of the EIR 2014 load demand

Insight: The accuracy of predictions can be affected by weather changes, data availability, and the complexity of the forecasting model [4]. Simple models used in benchmark forecasting may not always provide accurate predictions. To improve accuracy, it is advisable to combine benchmark forecasting with other methods such as machine learning or time-series analysis. This can lead to better predictions and a more comprehensive understanding of the situation.

Task 10:

Our objective is to determine the mean absolute percentage error for the persistence method and create a graph that shows its performance for forecast horizons up to one day in advance. Then, we need to analyse and clarify the shapes of the resulting curves, which indicate how well the method performs for different forecast horizons.

- To calculate this, I firstly divided the data set into two halves as I did in the previous task
- That whole data set was used to persistence to forecast using the benchmark forecasting approach as I did also in the previous task and divided the result prediction into two halves
- I used the both second halves one as actual value and the other as predicted value to calculate the mean absolute percentage error and then plotted the result

Results: The graph displayed in Figure 9 exhibits a similar shape to the one produced in the previous task. However, in this case, the y-axis represents the percentage of errors generated by the forecast. The depicted graph indicates that the percentage of error falls between the range of 1.4% to 33.3%. From midnight, the errors went up from 1.4% to 33.3%, but started to decrease at 11:30 until they reached 5% as show in figure 9 below.

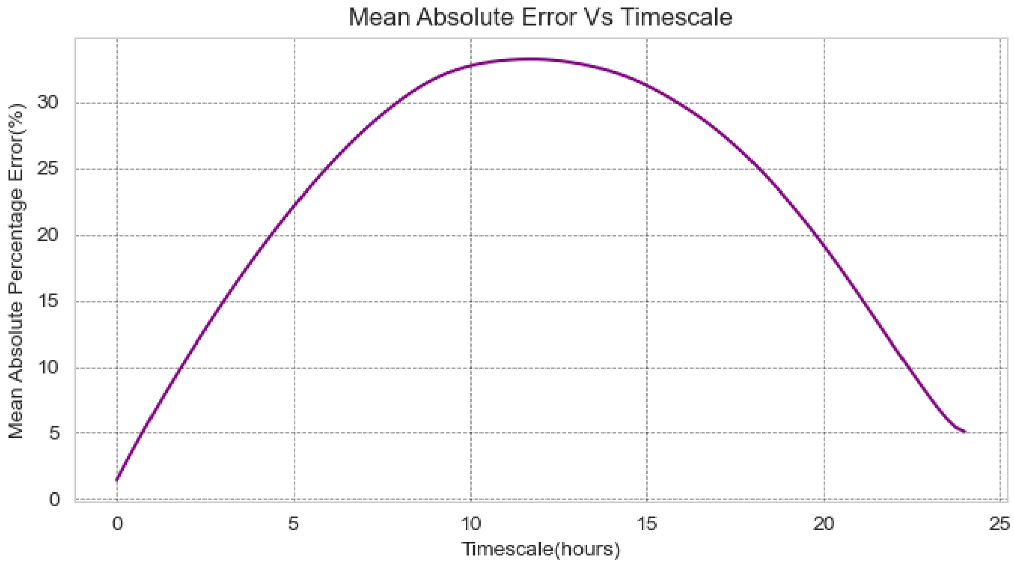

Figure 9: The mean absolute percentage error resulted from simple benchmark forecast

Insight: As shown in the above graph above, in my opinion the prediction on the next hour is the one which seem to be accurate using this approach. Since the errors varies within the 1.4-33.3% range, the acceptability of a forecasting error will depend on the specific context and the consequences of the error. Here MAPE displays an upward trend followed by a downward trend when plotted, it indicates that the forecasting model's accuracy fluctuates over time and investigating the reasons behind these changes is crucial for ensuring dependable and precise forecasting. Exploring other forecasting models should also be considered to determine if they perform better than the current model.

References

[1] "EirGrid," Wkipedia A, [Online]. Available: https://en.wikipedia.org/wiki/EirGrid. [Accessed 25 February 2023].

[2] "WHY IS MY ELECTRICITY BILL HIGHER IN THE WINTER?," Mister Spaky, 14 December 2022. [Online]. Available: https://www.mistersparkytulsa.com/higher-electricity-bill/#:~:text=Trends%20show%20that%20consumers%20use,higher%20demand%20for%20hot%20water.. [Accessed 20 February 2023].

[3] IEA, "Power Systems in Transition," IEA, May 2020. [Online]. Available: https://www.iea.org/reports/power-systems-in-transition. [Accessed February 2023].

[4] L. Nguyen, "Measuring Forecasting Accuracy: Formula, Best Practices And Tips," REVENUEGRID, 20 June 2022. [Online]. Available: https://revenuegrid.com/blog/forecasting-accuracy/#:~:text=The%20Mean%20Absolute%20Error%20(MAE,root%20of%20your%20sample%20size.. [Accessed 23 February 2023].