France Intraday Energy Demand Analysis and Forecasting Report

Rahman Mabano, MSECE
rmabano

Imported Libraries

numpy
pandas
matplotlib.pyplot
statsmodels.graphics.tsaplots [plot_acf]
pingouin
sklearn.metric [mean_absolute_error, mean_absolute_percentage_error]

PART 1

A dataset in form of a CSV file, called consists of 15-minute energy demand data readings for 2014, was imported into the programming environment.

Missing values were immediately handled by interpolating, and the date format was arranged into a timestamp column that will be used to plot a timeseries graph (month-day-year format).

The graph below shows a timeseries plot, showing the variation of electrical energy demand across all the months of the year 2014.

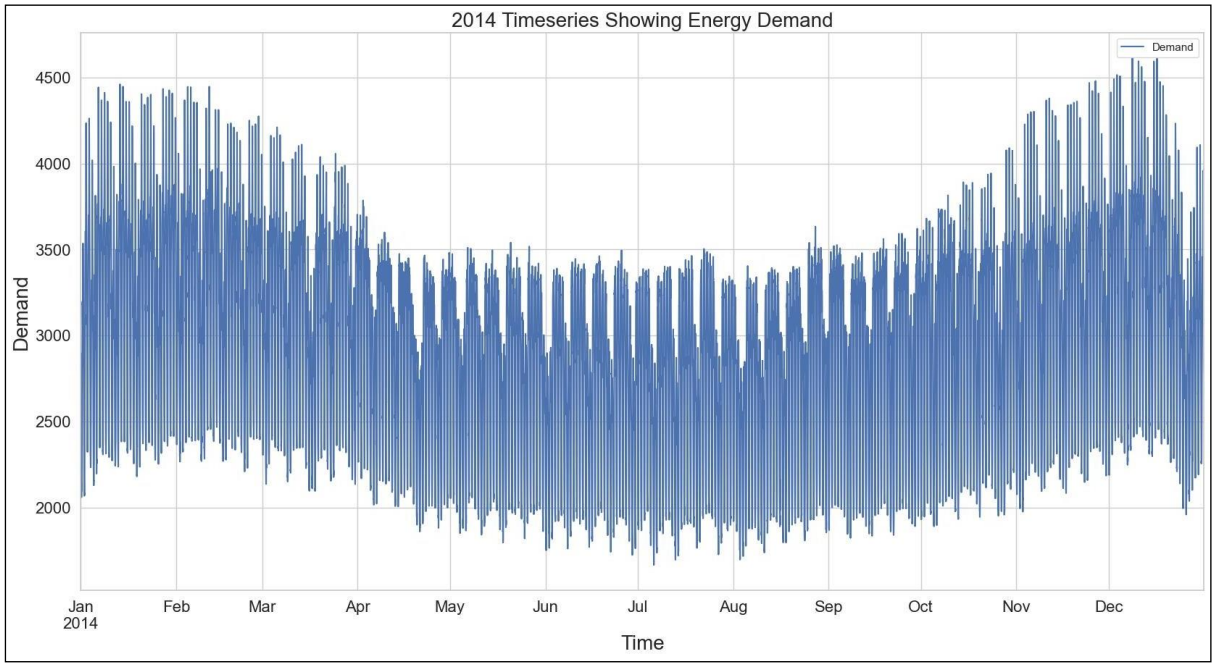

Fig.1: timeseries plot showing intra-15minute energy demand for the year, 2014.

Observations:

The maximum energy demand throughout the whole year was 4614 kWh, and the minimum demand occurs around July, with a demand value of 1665 kWh, as supported the timeseries plot in Fig.1.

The timeseries plot depicts an overall decline in energy demand from January to May, with a steady maintenance of the low energy requirement from May to September, where we start seeing a steady increase in energy demand at the end of October. Energy demand begins to rise back up again and reaches the maximum demand at the end of the year [1].

This obvious display of seasonality suggests that the increase and decrease in energy consumption may have been due to changes in weather conditions, as well as increase in human activity around the festive season at the end of the year. The Summer,

Fall, Winter and Spring seasons in Europe and the Western hemisphere align closely to the patterns shown in Fig.1, where higher temperatures require less utilisation of home heating systems, and encourage more outdoor activities. On the other hand, winter, and fall, catalyse more indoor activities and more utilization of heating systems leading to an increase in energy demands.

PART 2

Analysing the energy demand data over a 10-day period revealed a significant influence of past and future consumption on current usage. This connection is indicated by the high average autocorrelation coefficient of 0.813, as shown in the figure below.

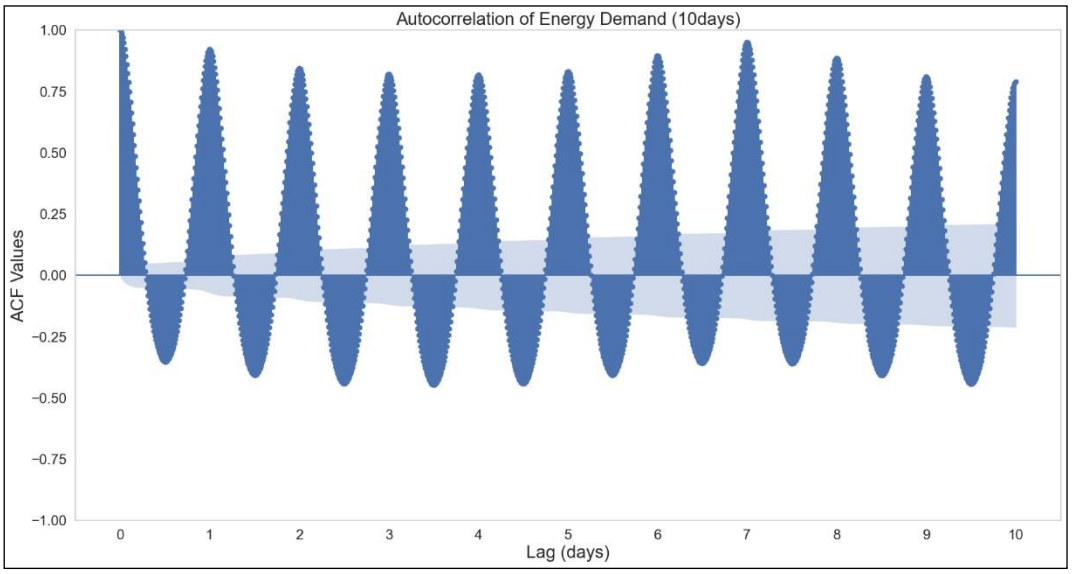

Fig.2: 10 Day energy demand autocorrelation coefficients.

Further illustrated in Fig.2, the plotted autocorrelation against lag (in days) depicts a daily cyclical pattern. The correlation between data points weakens as the time difference increases within a day but strengthens again as the day nears its end. This trend reinforces the presence of a distinct daily seasonality in energy demand [2].

PART 3

A new column called Time of Year, was created after normalizing and scaling the data frame values based on the shape of the data frame. The resulting timeseries shows how the demand varies over the course of the year, with respect to the time of the year, as shown in the diagram below:

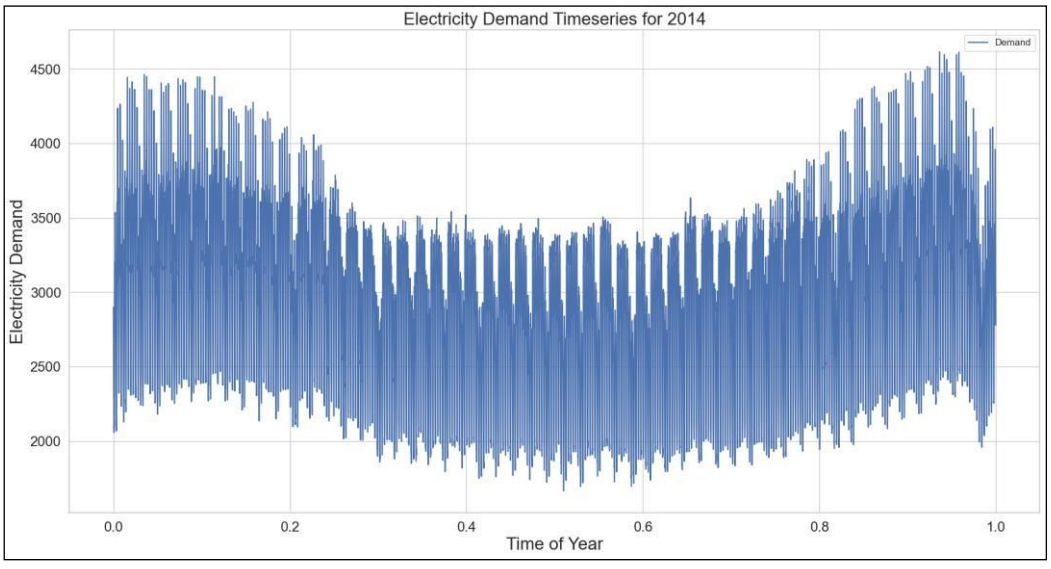

Fig.3: varying energy demand with time of the year.

As can be seen in the graph above, there seems to be a relatively low demand for energy between 0.3 to 0.7 for energy as compared to other times of the year, relative to the time of the year. We can also see the nonlinear relationship between time of the year and the energy demand.

PART 4

The monthly average energy demand is shown in the bar plot below. The numbers on the x-axis represent the chronological months of the year, January being the 1st and December being the 12th month of the year.

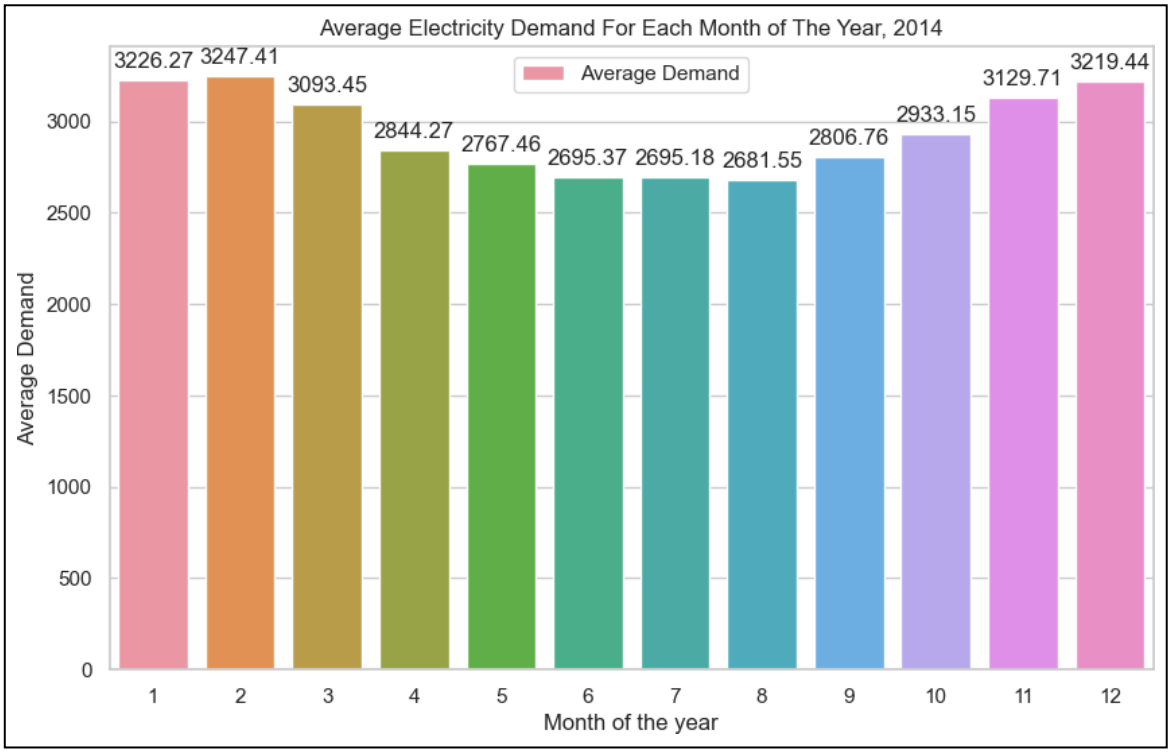

Fig.4: average monthly energy demand for each month of the year.

From the Fig.4, it can be seen that our previous deductions about seasonal energy demands are accurate. February has the highest average energy demand than all months in the year, along with January, December, then November. June, July, and August are the months with the lowest demand in energy respectively. These months correspond to the summer and winter months respectively.

PART 5

We constructed the daily demand profile by analysing energy consumption trends throughout the year. We achieved this by grouping hourly data points and calculating the average energy demand for each hour. This process yielded the hourly breakdown of typical energy consumption, visualized in the attached graph below.

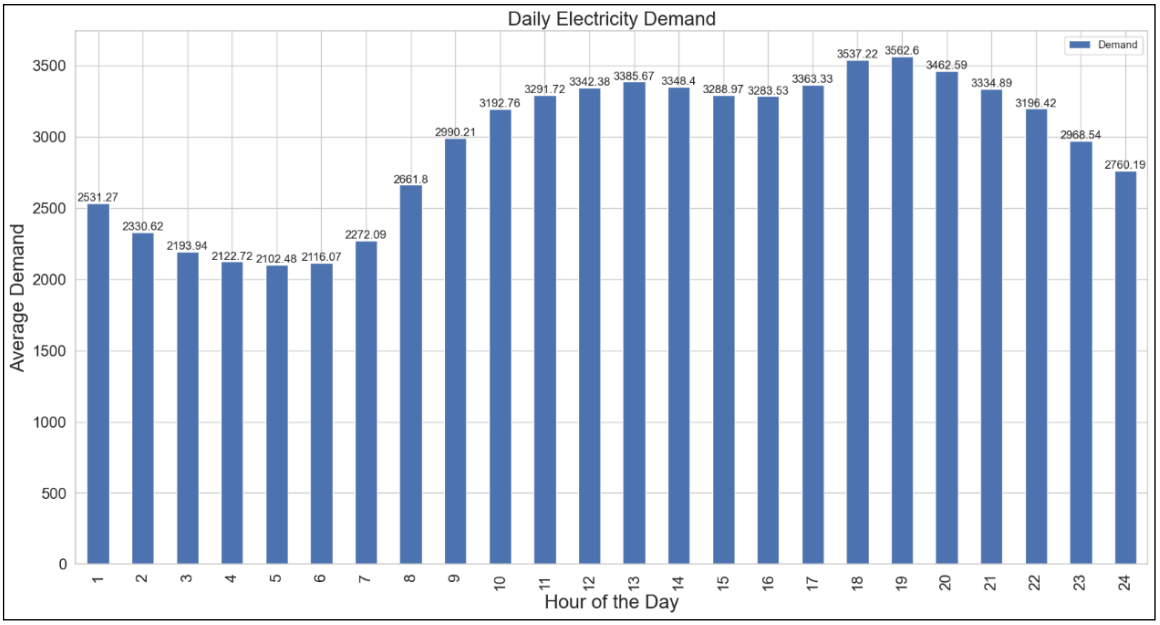

Fig.5: daily energy demand profile, showing average hourly demand.

The bar graph above displays the varying energy demand profile for each hour of the year.

Fig.5 illustrates a clear daily cycle in energy consumption, peaking at 7 PM and reaching its lowest point around 5 AM. From late evening onwards, demand steadily declines as businesses wind down and households settle in. This continues until the early morning, when activity is minimal and lighting needs are primarily limited to streetlights, and essential services that require energy. As morning progresses, demand gradually rises again, culminating in the evening peak coinciding with increased activity across homes and businesses. This pattern aligns closely with human behaviour, reflecting increased lighting, cooking, and general activity during evening hours, followed by a decline as households sleep and businesses close. The lowest point at 5 AM reflects the minimal activity during the night. This relationship between human behaviour and energy demand suggests that further analysis of hourly activity patterns could reveal other interesting trends.

PART 6

On the x-axis the days of the week, numbered from 1 to 7, Monday being the 1st day of the week, all the way to Sunday the 7th day of the week. The arrangement is follows:

Day of the week	Number on x-axis
Monday	1
Tuesday	2
Wednesday	3
Thursday	4
Friday	5
Saturday	6
Sunday	7

This was done by creating a column called 'day_of_the_week' and filling it up days corresponding to the dates from the timestamp column.

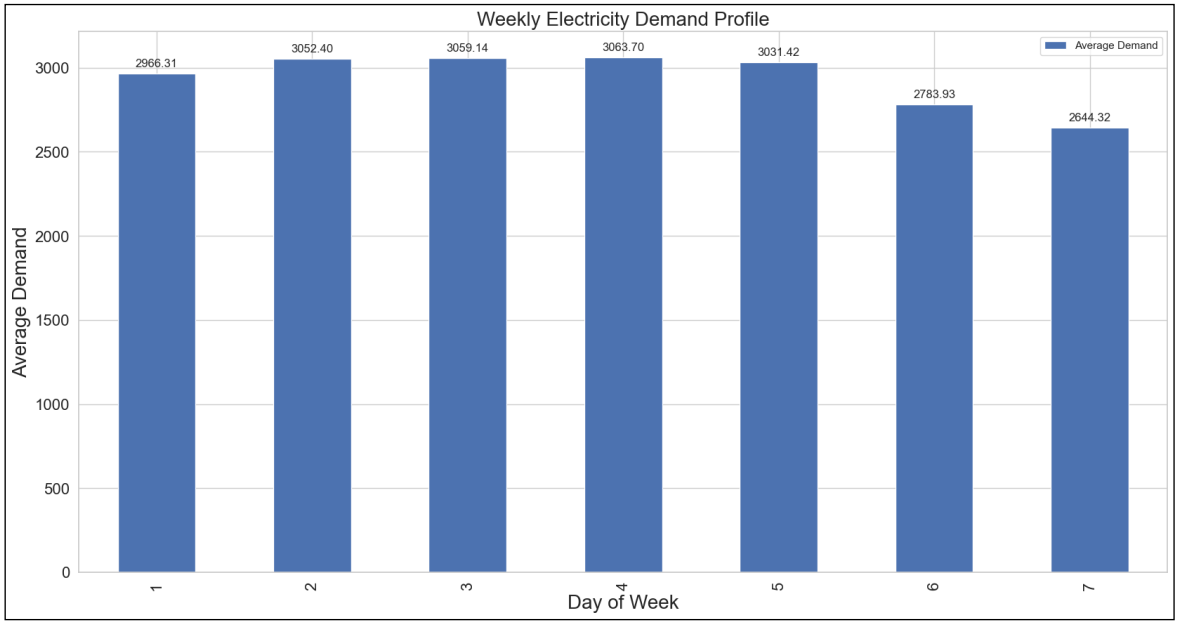

Fig.6: average energy demand for each of the seven days of the week.

The graph unveils a distinct difference in energy demand between weekdays and weekends. Workdays show consistently higher demand compared to both afternoons and Saturdays, likely due to reduced business activity during those periods. By midweek (Tuesday-Thursday), demand reaches its peak, reflecting the heightened bustle of daily routines. Notably, a significant drop occurs from Friday to Sunday, mirroring the widespread shutdown of businesses and reduced activity on Sundays. This observed pattern aligns perfectly with intuitive expectations regarding energy consumption, validating these findings.

PART 7

The datetime library can be effectively used to extract accurate day names from provided dates, using the day_name() function. The output is then stored in a column called day_of_week, used to plot the graph as shown below:

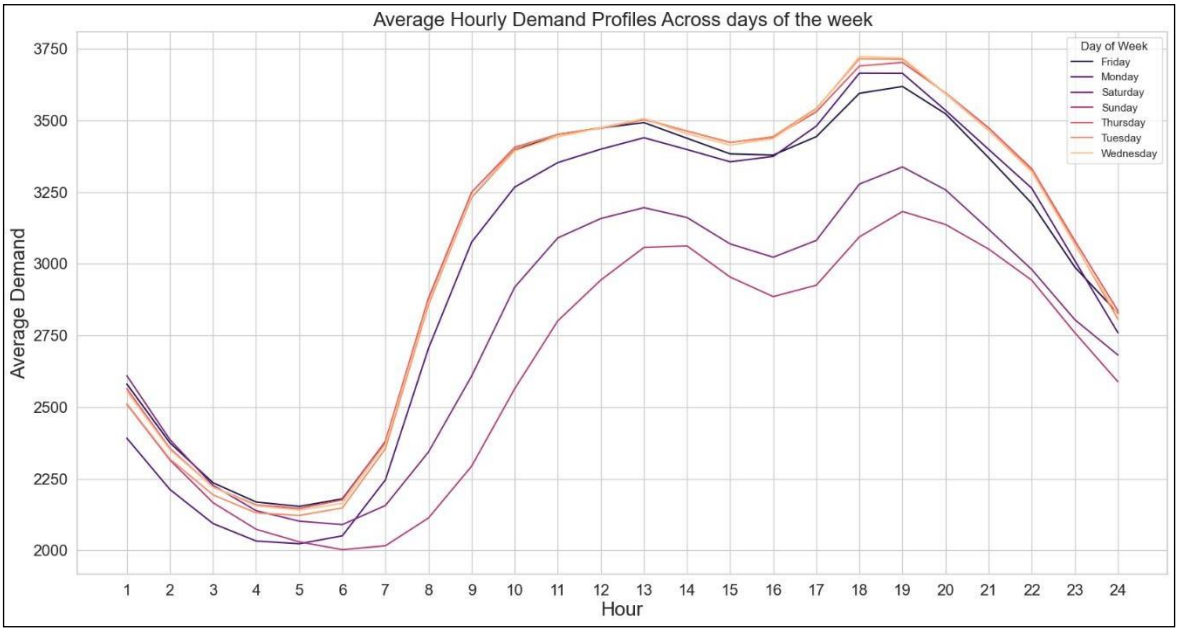

Fig.7: daily demand profile for each day of the week

Fig.7 depicts the average hourly energy consumption for each day of the week in 2014. It reveals a consistent daily pattern across the year. Energy demand starts low in the early morning hours (4-6 am) and steadily climbs until an afternoon peak around 1 pm. Following a slight dip around 2:30 pm, it surges again to reach its highest point between 7 and 8 pm. Finally, it gradually declines throughout the evening, remaining low overnight. This cyclical pattern highlights a relatively quiet early morning, escalating activity throughout the day, reaching its peak around sunset, and finally settling down overnight.

PART 8

In this case the 2 sample T-test analysis was conducted as follows:

- Null Hypothesis: average demand over the weekend == average demand during weekdays
- Alternative Hypothesis: average demand over the weekend != average demand during weekdays

Again, the effective use of the datetime library is crucial in helping to segregate weekends from weekdays and storing the values in appropriate columns.

In the end, if the p-value is less than alpha (level of significance), which is 0.005, we reject the Null hypothesis case.

To carry this analysis out, the Scipy.stats library comes equipped with a two-sample t-test function, which was used to compare the mean demand for the same minute and hour on weekends versus on weekdays. The results of the two-sample t-test function are as follows:

```
The t-test statistic is 46.54684714308395 with a p-value 0.0
```

From the T Test results, the p-value alone is sufficient with facilitating us to reject or approve the null hypothesis, as previously stated. Given a p-value of 0.0, which is significantly lower than the chosen significance level (alpha) of 0.05, we can reject the null hypothesis. This means there is strong evidence to conclude that the average demand on weekends differs from the average demand on weekdays. The obtained t-test statistic ais approximately 46.55.

PART 9

To assess the persistence benchmark, the intraday energy demand data was split in half. This benchmark uses the present value to predict future values at specific time intervals (96 forecasts at 15-minute intervals), using the formula shown below:

$$y\_hat(t+k) = y(t)$$

where:

- y_hat(t+k) is the forecast at time t horizon in the future.
- y(t) is the current observation.

In simpler terms, we took current energy demand measurements and used them to predict demand 15 minutes, 30 minutes, and so on up to a full day ahead. The graph below shows evaluation of forecasted horizons, by plotting the Mean Absolute Error (MAE) against the Forecast Horizons (hourly).

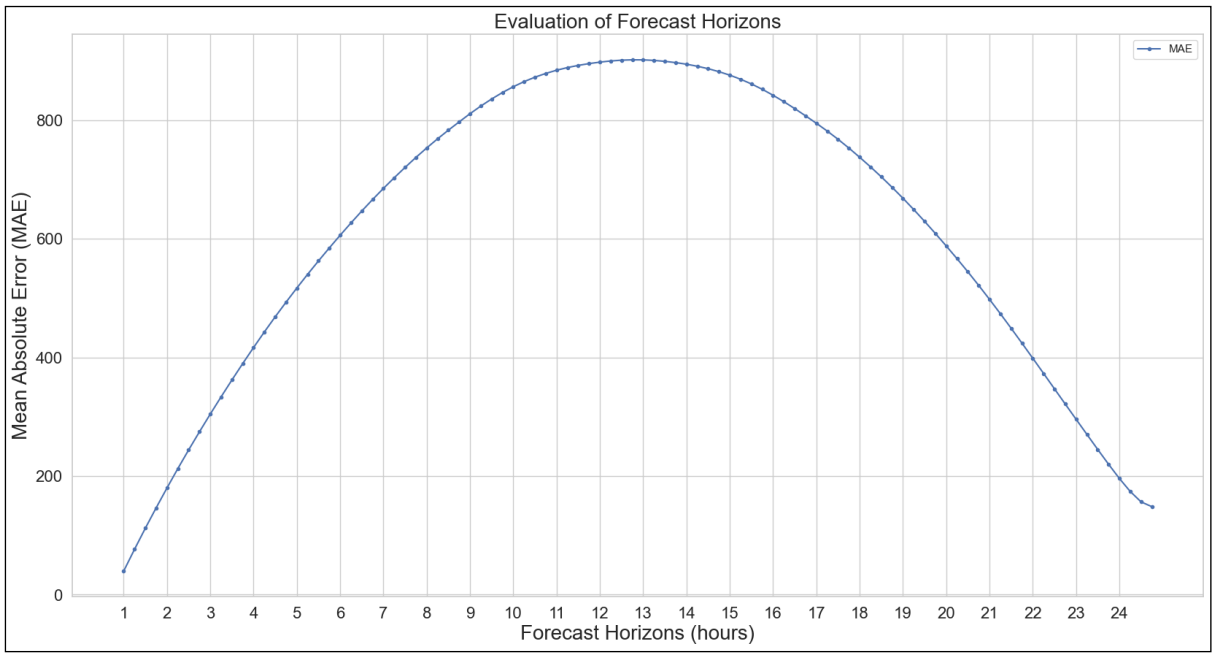

Fig.8: MAE against Forecast Horizons (Persistence Benchmark analysis).

Evaluating the Persistence Benchmark:

To measure how well the persistence benchmark (predicting the future based on the present) performs, we used a metric called **mean absolute error (MAE)**. Fig.8 shows this error plotted against different prediction timeframes (forecast horizons).

Key Findings:

- **Shorter forecasts are more accurate:** The 15-minute forecast had the lowest error (39.706), followed by 30 minutes (76.39). This makes sense, as predicting just a few minutes ahead is easier than predicting hours or days ahead.
- **Exponential error increase:** As we try to predict further into the future (longer forecast horizons), the error increases very quickly, reaching a peak of 901.886 between and 1pm.
- **Unexpected error decrease:** Interestingly, the error then goes significantly down (148.958) when trying to predict a full day ahead.

This error turnaround is explained by another concept called **autocorrelation**. This basically measures how similar the energy demand is at different points in time. Imagine comparing energy use at 10 AM today to energy use at 10 AM last Monday. The closer these time points are, the more similar the demand tends to be (high autocorrelation). However, as we compare times further apart, the similarity usually decreases (low autocorrelation).

The graph mentioned in question 2 (autocorrelation graph) shows that this similarity starts low in the morning, decreases until midday, and then starts increasing again towards the end of the day. This "turnaround" in autocorrelation also explains the error turnaround:

- Early forecasts (minutes/hours) benefit from high autocorrelation (similar time points).
- Longer forecasts (up to 12 hours) suffer from low autocorrelation (less similar time points).
- Full-day forecasts benefit from the rising autocorrelation again towards the end of the day.

In essence, the error pattern reflects the changing similarities between energy demand at different points in time throughout the day. Choosing a lag value close to 0 or close to 24h may provide highly accurate results.

PART 10

A data frame was created to store the Mean Absolute Error (MAE) values as well as the Mean Absolute Percentage Error (MAPE) values for the same number of forecasted horizons.

The key distinction lies in how they handle the magnitude of errors. MAE, focusing on the absolute difference between predicted and actual values, remains unaffected by the size of the actual value itself. This makes it easy to interpret in terms of real-world units, like dollars or watts. MAPE, on the other hand, expresses the error as a percentage of the actual value, offering a scale-independent view. This makes it ideal for comparing errors across different scales, where absolute differences might be misleading. However, this sensitivity to smaller values can make MAPE less intuitive when interpreting specific errors [3].

The graph showing the variation of the MAPE against the forecasted time horizons.

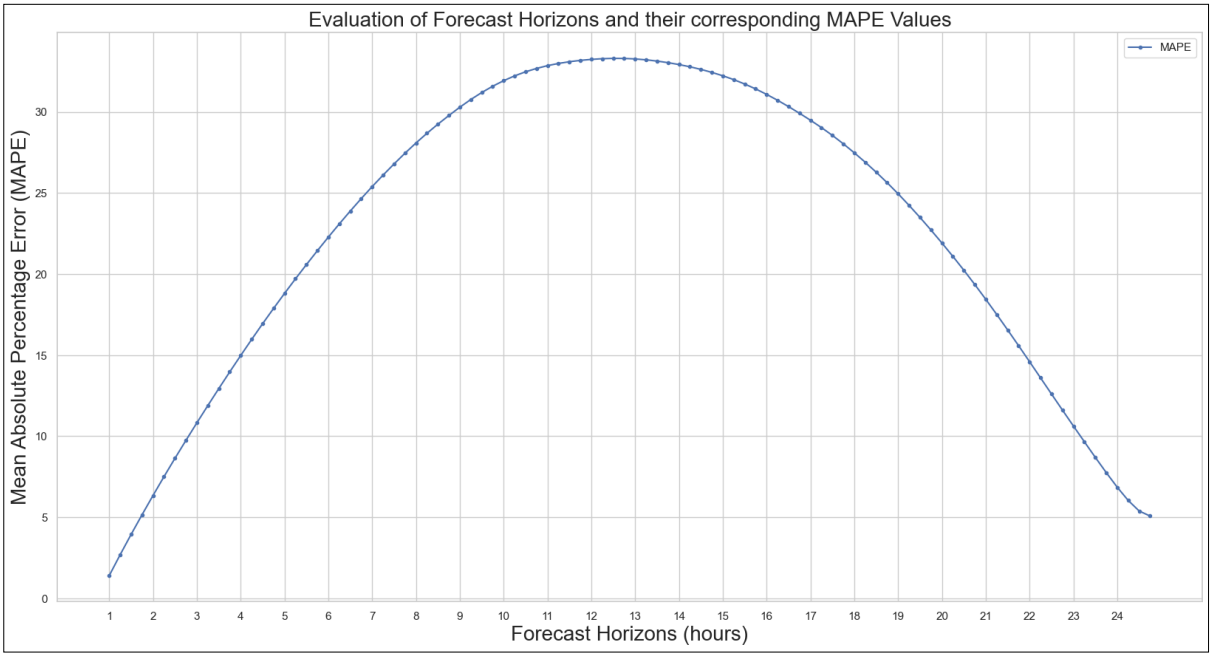

Fig.10: MAPE against Forecast Horizons (Persistence Benchmark analysis).

Persistence Benchmark MAPE Analysis:

The Mean Absolute Percentage Error (MAPE) of the persistence benchmark was evaluated for various forecast horizons up to one day, as shown in Figure 10. Notably, the most accurate forecast was achieved with a 15-minute horizon, exhibiting a MAPE of only 1.4%. Extending the horizon to 30 minutes slightly increases the error to 2.69%.

Figure 10 reveals a key trend: MAPE increases steadily with longer forecast horizons, reaching a maximum of 33.31% around midday. Interestingly, the error then decreases significantly, reaching 5.1% by the end of the day. This distinct pattern forms a concave curve in the MAPE VS Forecast Horizon graph.

This intriguing turnaround in error can be attributed to the behaviour of autocorrelation coefficients. These coefficients, highest at shorter lags, gradually decline until midday. However, they then reverse course, increasing towards the end of the day (as shown in question 2's autocorrelation graph). Crucially, this reversal in autocorrelation mirrors the turnaround in MAPE, suggesting it as the underlying cause.

Key Finding

The persistence benchmark exhibits a non-linear relationship between forecast horizon and error, with the highest error occurring around midday. This behaviour may be directly linked to the changing patterns of autocorrelation throughout the day. Choosing a lag value close to 0 or close to 24h may provide highly accurate results.

REFERENCES

[1] Bett, P. E., Thornton, H. E., Troccoli, A., De Felice, M., Suckling, E., Dubus, L., ... & Brayshaw, D. J. (2022). A simplified seasonal forecasting strategy, applied to wind and solar power in Europe. *Climate services*, *27*, 100318.

[2] Knight, K. M., Klein, S. A., & Duffie, J. A. (1991). A methodology for the synthesis of hourly weather data. *Solar Energy*, *46*(2), 109-120.

[3] Ostertagova, E., & Ostertag, O. (2012). Forecasting using simple exponential smoothing method. *Acta Electrotechnica et Informatica*, *12*(3), 62.