CARNEGIE MELLON UNIVERSITY

DATA ANALYTICS (COURSE 18-899)

ASSIGNMENT 2

Name: Manzi Patrick

Preferred name: Professor Manzi

Andrew_ID: pmanzi

Honor statement

On my honor, as a Carnegie-Mellon Africa student, I have neither given nor received unauthorized assistance on this work.

Language: python

Libraries used:

1. seaborn

3.matplotlib

4. sm

5.pandas

6.numpy

7.statistics

8.mean_absolute_error

9.mean_absolute_percentage_error

10. MinMaxScaler

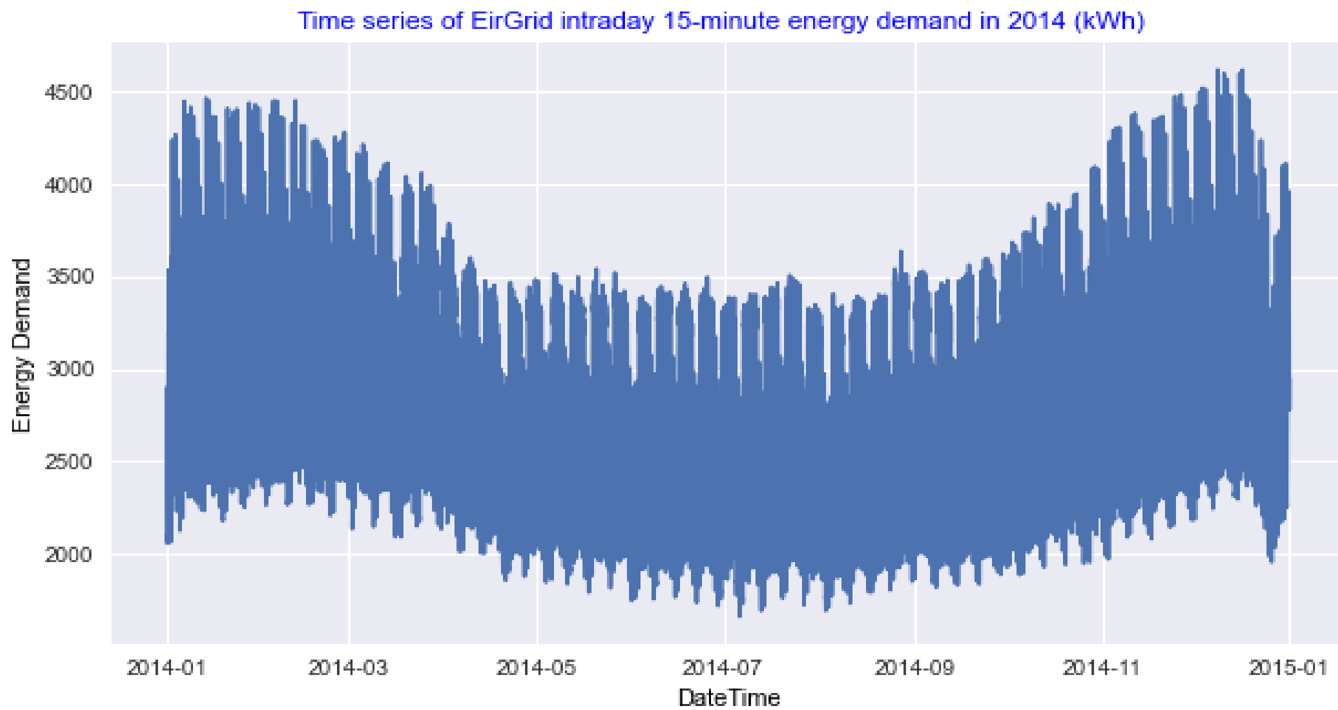
11. datetime

12.ACF

13.t-test

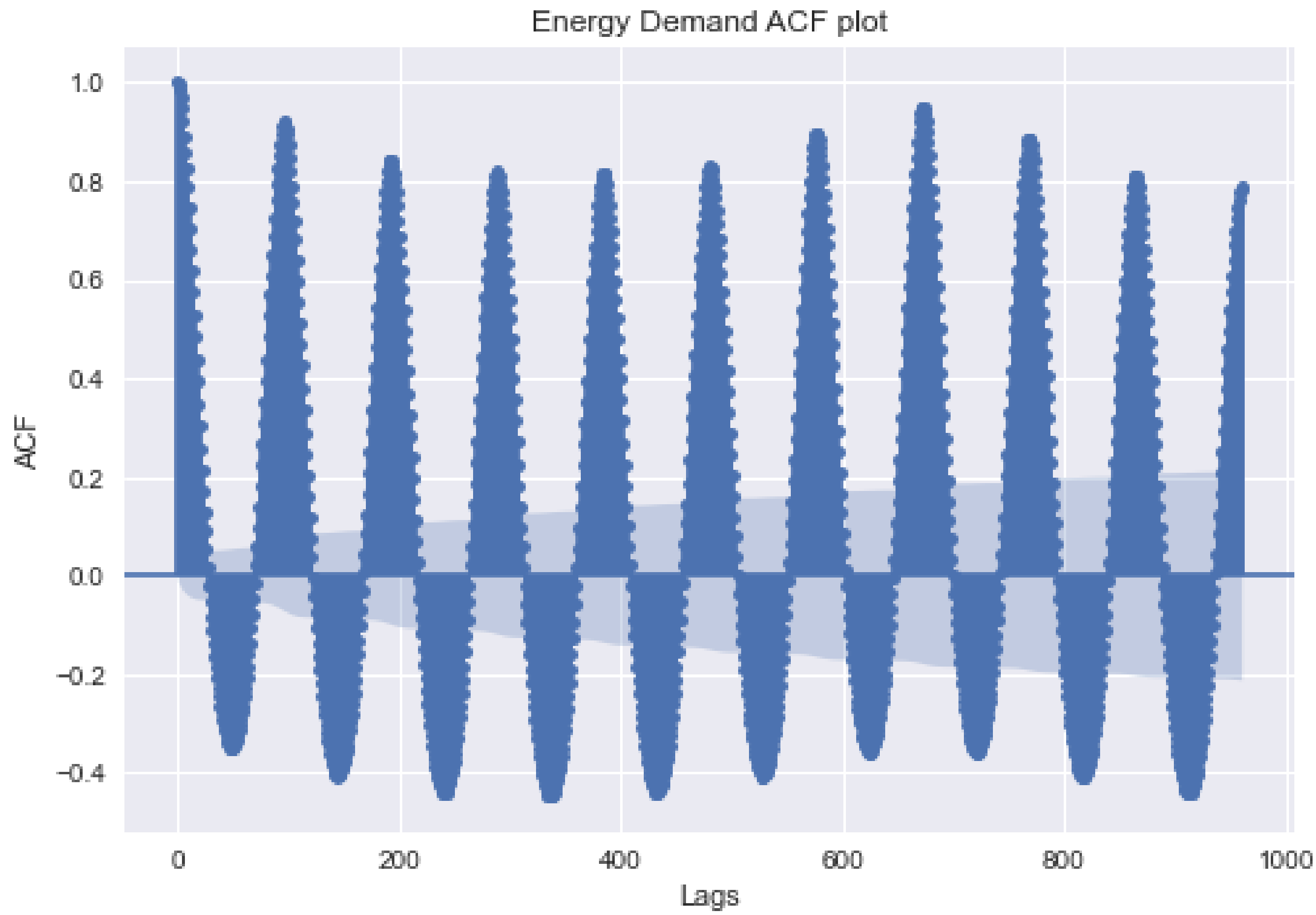
Energy Demand Forecasting

1.Load the EirGrid system demand data into your computer. Fix day-light saving issues and missing values using linear interpolation. Plot and carefully label the time series of energy demand during 2014.



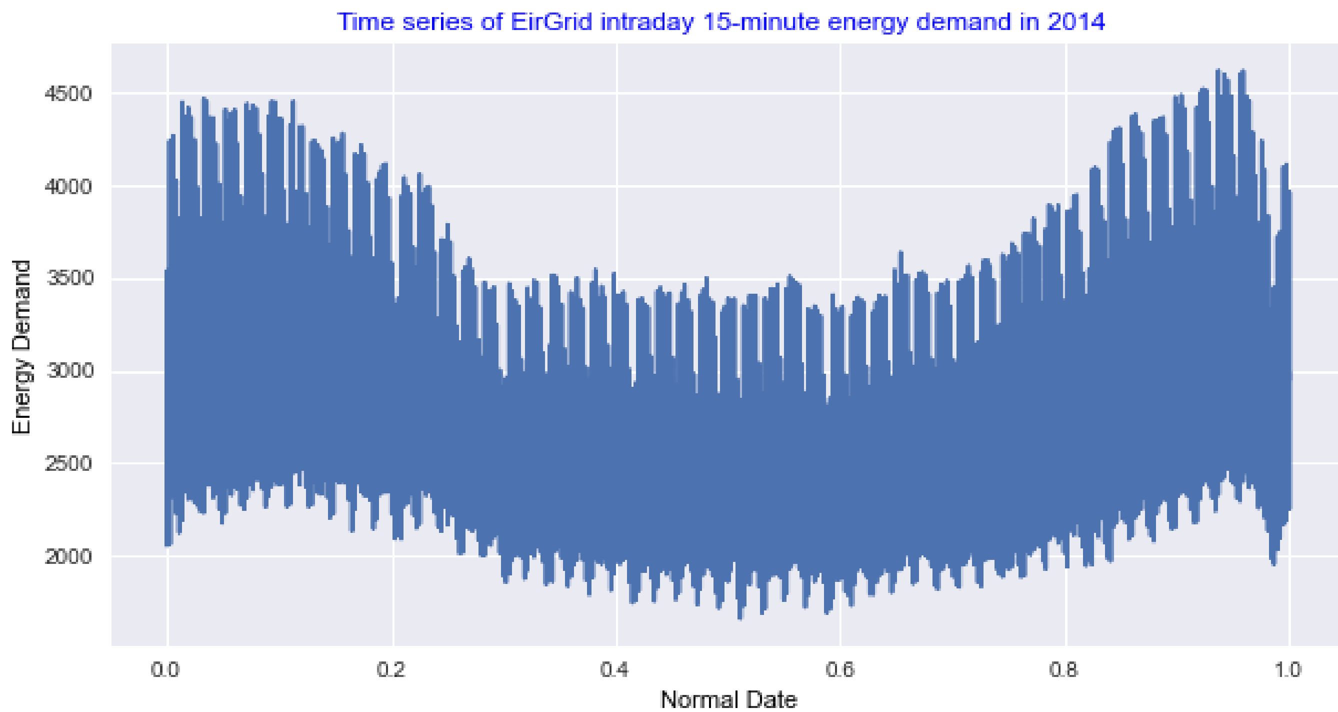
Despite daily fluctuations, the energy demand time series shows that demand tends to be generally high at the beginning of year till the end of march, and then the demand becomes slightly low in the middle of the year starting from April to September and then rise to its pick again. these rise and fall in energy demand might be as a results individual/corporate activities.

2. Estimate autocorrelation coefficients for 10 days and plot the autocorrelation against the lag with axis labelled in days. Comment on the shape of the plot.



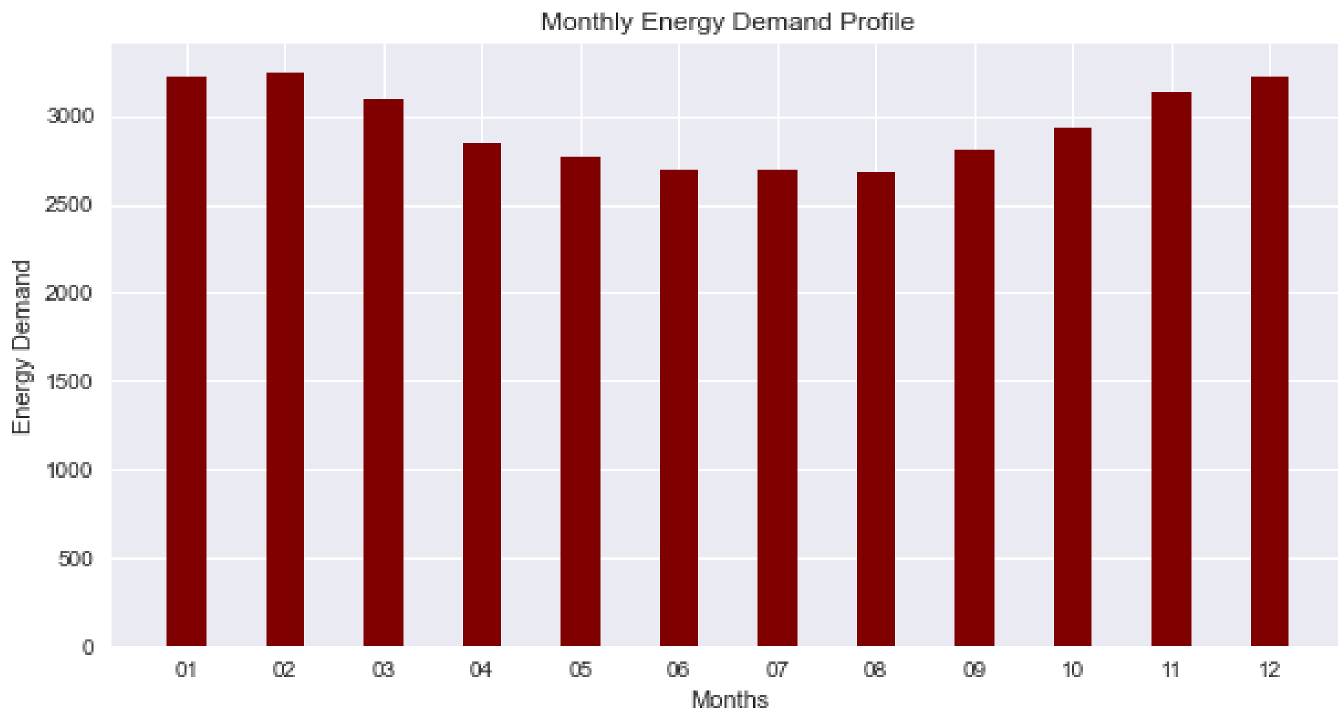
The autocorrelation function plot shows that there is significantly high interdependence between values of energy demand at time (t) and their lags at (t-1). Generally, those autocorrelations coefficients that are very high for example from 0.8 confirms that the correlation between today's demand values and prior values is higher. There is also a repeating pattern of seasonality that is depicted by these positive and negative fluctuations.

3. Create a time of year variable that ranges between 0 and 1 and show how the demand varies over the course of the year using a graphic.



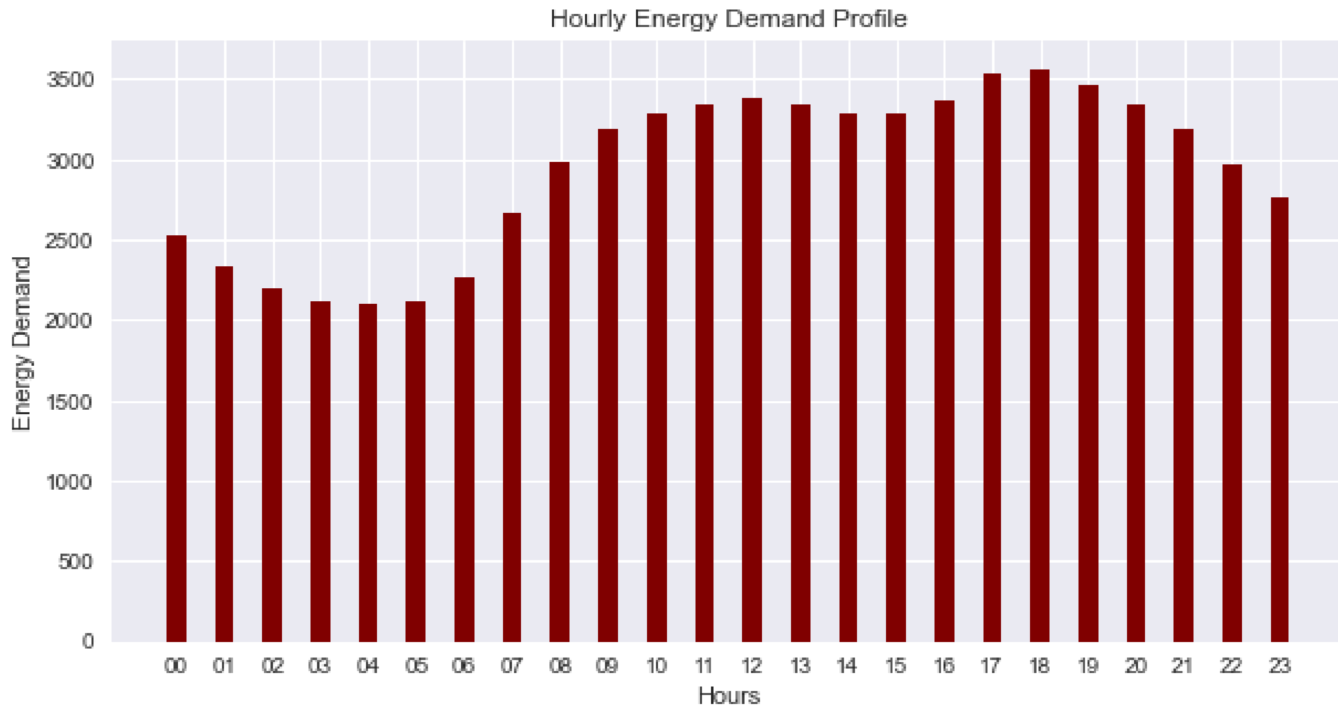
This graph shows the variation of energy demand from January (0) to December (1) in 2014. It highlights significantly high energy demand around January till the beginning of April, followed by a constant low demand that starts around April till September and then a high demand throughout the rest of the year. This variation might be caused by several factors such as weather conditions, vacations of employees in factories and many more.

4. For each of the 12 months of the year, calculate the average demand and display them as a bar chart, and label them appropriately.



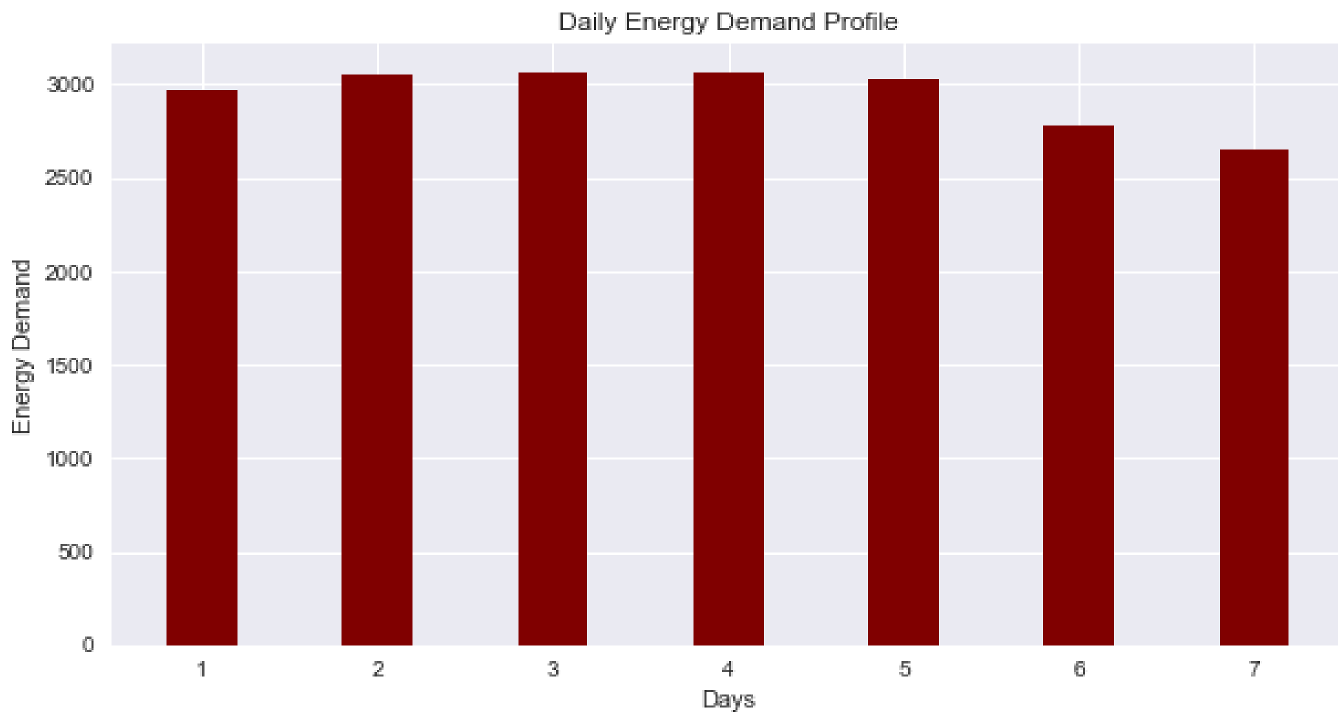
The variance in energy demand from January (0) to December (1) in 2014 is depicted in this graph. It shows a period of significantly high energy demand from January to the beginning of April, followed by a period of consistently low demand from April to September, and then a period of consistently high demand for the rest of the year. This fluctuation could be caused by a variety of variables, including weather conditions, factory employee vacations, and so on.

5. For each of the 24 hours of the day, calculate the average demand and display them as a bar chart, indicating the different hours of the day. This graphic is often referred to as the daily demand profile.



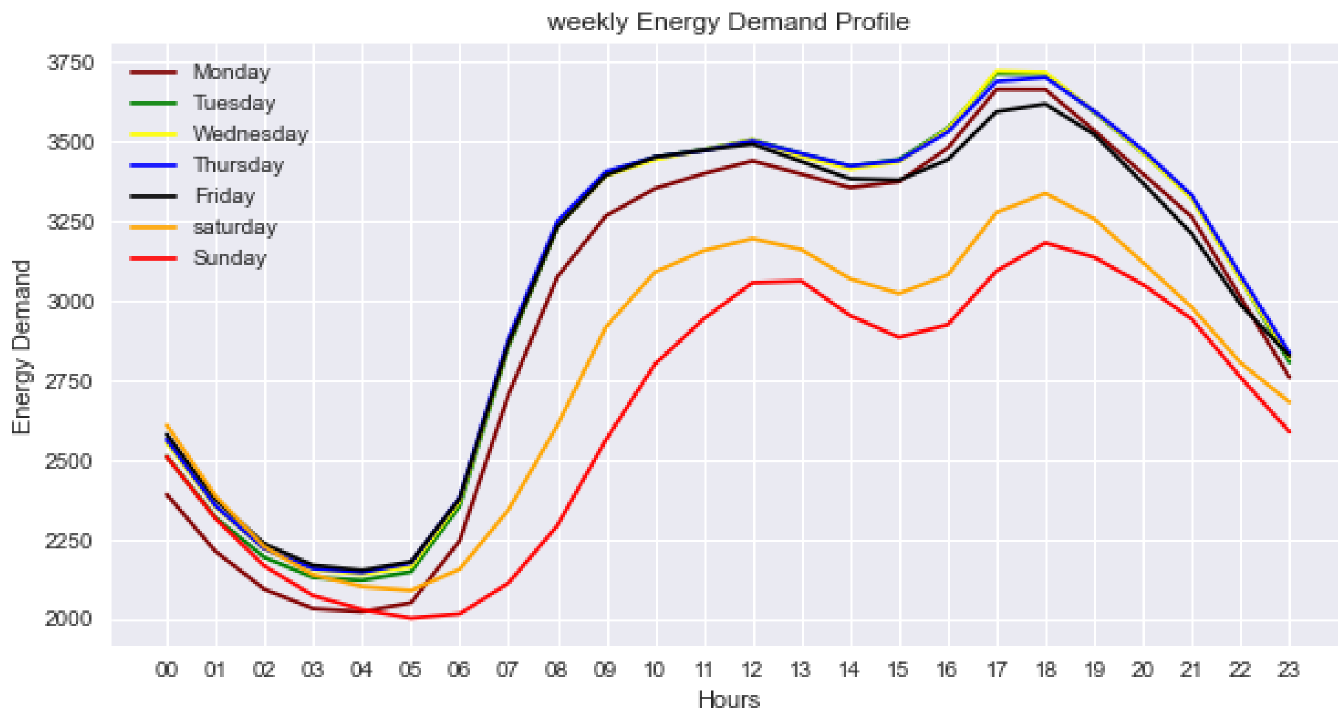
The graph above depicts hourly energy demand profile in 2014. it shows that the energy requirement is low at night(22h-7h) and then high throughout a day. this is basically due to fact that many activities stop at night.

6. For each of the seven days of the week, calculate the average demand and display them as a bar chart. Does the result make sense based on intuition about electricity consumption?



This graph illustrates daily energy demand profile in 2014. It shows that their significantly higher demand over the working days which compared to weekends. This reduction in weekend is due to the fact that many employees don't go to work and of course many activities do not open.

7. Calculate a daily demand profile for each day of the week. This can be achieved by selecting a specific hour for each day and computing the average. Show the results on a graphic with a separate profile curve for each day.



The graph above shows a snapshot of weekly energy demand profile across 24 hours of a day in 2014. it clearly shows low energy demand overnight hours and a significant rise of demand in a daytime, also it clearly shows that the energy demand is significantly row on Saturday and Sunday compared to other working days.

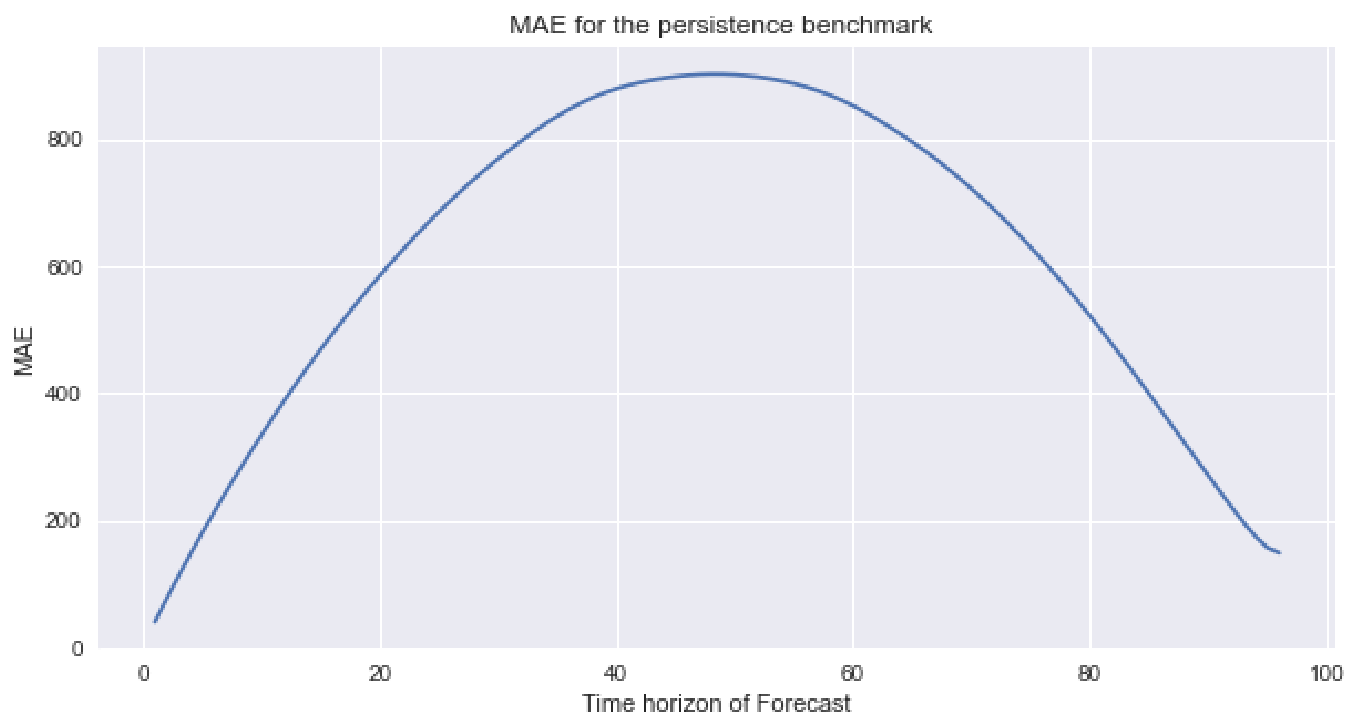
8. Is there a statistically significant difference between demand during the weekend (Saturday and Sunday) and during the working week (Monday through Friday)? Perform a statistical hypothesis test, such as a t-test, in order to reach a conclusion.

```
T-statistic=46.54684714308395,
pvalue=0.0
```

Null hypothesis: Energy demand distribution in working days is equal to the energy demand over the wee kend.
Here **we reject null hypothesis since p-value of (0.0) is less that alpha (0.05).**

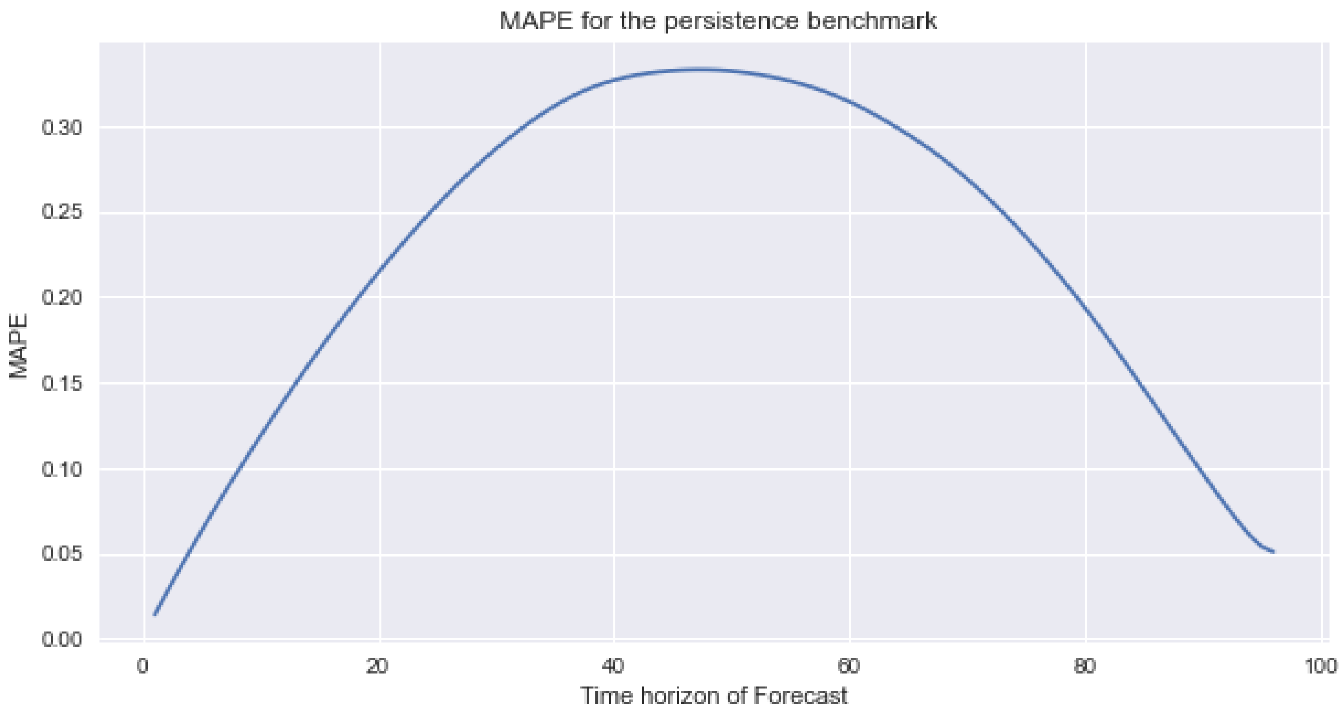
That means there is a significant difference between energy consumption during working days and weeke nds.

9. Divide the data into two halves and use the second half for evaluation purposes. Study the simple benchmark forecasting approach known as persistence. For data that does not change much from one time step to the next, we can assume that the most recent observation is a good forecast of the future. The forecast issued at time t for k periods ahead is simply given by y_hat(t+k) = y(t). Calculate the mean absolute error (MAE) and plot it against forecast horizons for lead times up to one day ahead.



This graph above depicts the forecast of persistence benchmark mean absolute error across 24 hours timeframe. it reveals that the more we predict to the close time horizons we are likely to get accurate forecast than when we predict far time horizons. Again, there is a point that we react (in the middle of a graph) and start repeating a similar pattern, here at that point, the error again decreases even if we are forecasting data points that are far from the basis. finally, the graph shows a repetitive pattern at 96th time horizon that means for example if our basis was 12am and at the end of graph it is going to be 12am which entails that the model is going to be accurate because of repeatability of magnitude of demand over a day.

10. Calculate the mean absolute percentage error for the persistence and plot this against the forecast horizon up to one day ahead. Discuss and explain the shapes of the curves showing performance against forecast horizons.



This graph above depicts the forecast of persistence benchmark *mean absolute percentage error* across 24 hours timeframe. it reveals that the more we predict to the close time horizons we are likely to get accurate forecast than when we predict far time horizons. Again, there is a point that we react (in the middle of a graph) and start repeating a similar pattern, here at that point, the error again decreases even if we are forecasting data points that are far from the basis. finally, the graph shows a repetitive pattern at 96th time horizon that means for example if our basis was 12am and at the end of graph it is going to be 12am which entails that the model is going to be accurate because of repeatability of magnitude of demand over a day.