CARNEGIE MELLON UNIVERSITY
DATA ANALYTICS (COURSE 18-787)
ASSIGNMENT 3

SUBMISSION INSTRUCTIONS:

- Submission process:
  - Put **code files and data files** in a single folder
  - The name of the folder should be the same as your Andrew ID
  - The name of the code and report files should follow the format: **Student_ID-Name-DAassignment3**
  - **Zip this folder and attach the zipped file on the assignment submission page (CANVAS)**
  - After attaching the zipped file, click on "Add Another File" from the assignment submission page
  - **Attach your report**
- All graphs should be followed by observational comments
- If you prefer to use R and Python for this assignment, the report should provide a list of the non-built-in libraries you used in your code.
- The submission deadline is **23:59 Rwandan Time (CAT)** on **Monday 28th February 2021.**

1. Historical intraday demand data can be used to study human behavior and seasonality in activity. Download the EirGrid intraday 15-minute energy demand data for the year 2014 from Canvas and save the data as a csv file with three columns for Date, Time, and Demand. Note the null values in the spreadsheet. 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.

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.

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.

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.

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.

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?

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.

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.

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.

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.