PROJECT EXERCISE #1

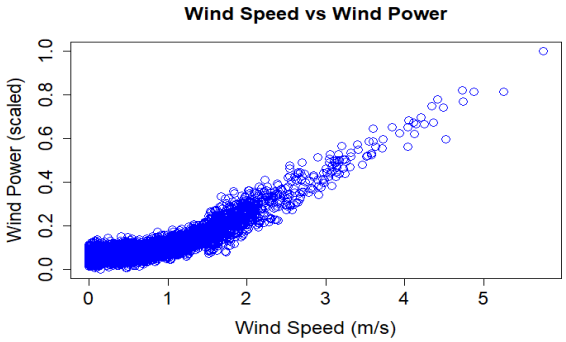
Data Bosses
Anne Christine Domercant
Tina Truc Hoang
Tasnim Quayum

Contribution:

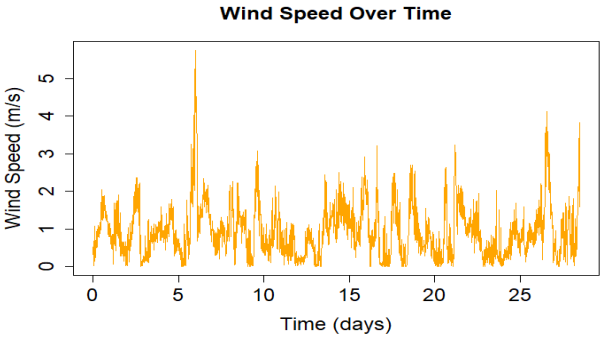
Anne Christine	Task 3, report
Tina Truc	Task 2, report
Tasnim	Task 2, report

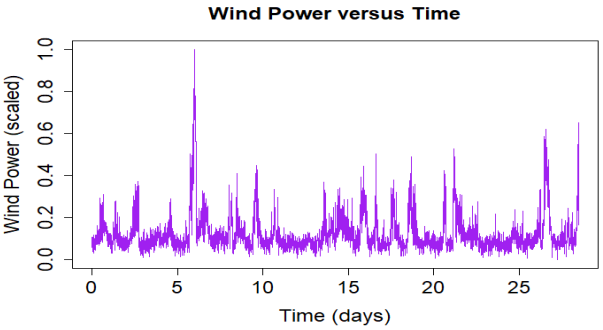
Course: Industrial Informatics (14:540:485:01)
Instructor: Dr. Ahmed Aziz Ezzat
Rutgers, The State University of New Jersey

Task 2: Data Visualization and Preliminary Analysis

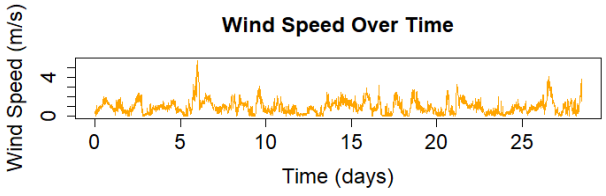


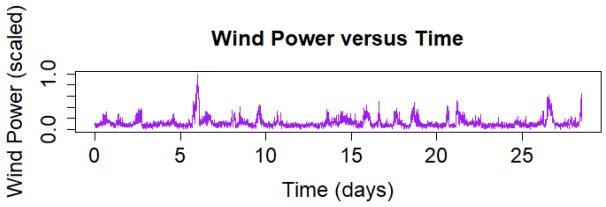
(i) The scatter plot shows the positive correlation between wind speed and wind power. There is also a higher concentration of data points at lower wind speeds and wind speed. The cut-off wind speed is the point where the wind power starts to increase rapidly with increasing wind speed. So, an eyeball estimate of the cut-off wind speed could be from 0 to 0.5 m/s.





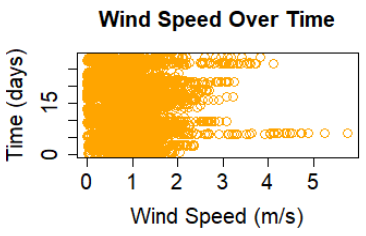
(ii) The wind speed fluctuates over time, with some periods of high wind speed. A majority of the wind speeds over time are less than 2.5 m/s. Around day 5, we see the maximum wind speed recorded at 5.757 m/s. Some seasonal patterns and trends from historical data that can be useful for forecasting future wind speed values.
(iii) A majority of wind power values are from 0.0 to 0.2. The power does not fluctuate as much as speed over the period of time. There are periods when the power output is high and other periods when it is low.

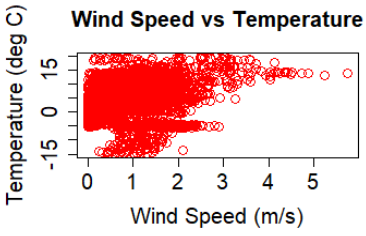


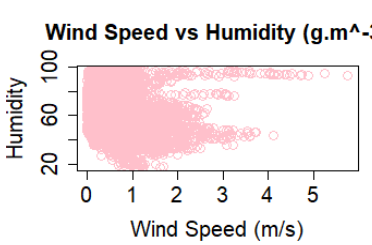


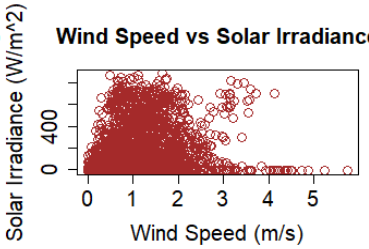
(iv) From the comparison of the wind speed and wind power time series plots, we can observe the strong correlation between wind speed and wind power and their similar pattern. However, it is easier to see changes in wind speed than wind power.

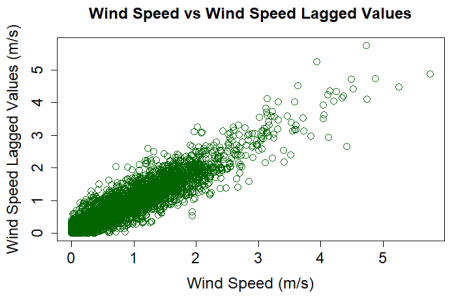
(v) There is a relationship that can be seen between wind speed and time, which increases and decreases over time. There does not seem to be a clear relationship between wind speed and humidity, and wind speed and solar irradiance. There appears to be a slight relationship between wind speed and temperature, because we observe that higher wind speeds tend to occur at higher temperatures.

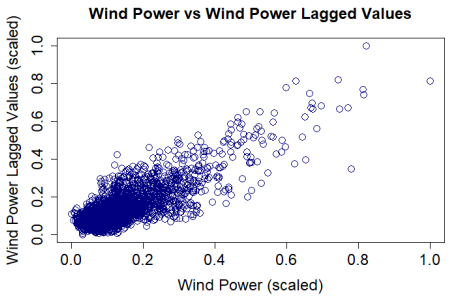










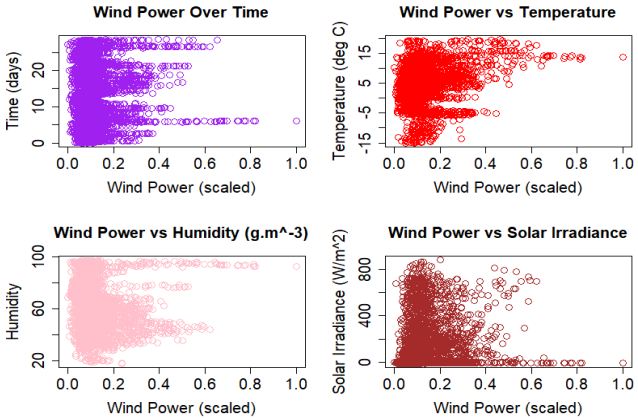


(vi & vii) At a glance, the wind speed and wind power plotted with their respective lagged values appear to be positively correlated with a linear relationship. Most of data is concentrated at lower values of wind speed and power.

```
> cor(data_new)
                    Speed_ms Temperature_C     Humidity Solar_Irradiance_Wm2 Power_scaled
Speed_ms           1.0000000   0.274028779 -0.169861218            0.2765546    0.8885186
Temperature_C      0.2740288   1.000000000 -0.004808583            0.2133071    0.2570446
Humidity          -0.1698612  -0.004808583  1.000000000           -0.4658484   -0.0675234
Solar_Irradiance_Wm2 0.2765546   0.213307149 -0.465848366            1.0000000    0.1958399
Power_scaled       0.8885186   0.257044591 -0.067523396            0.1958399    1.0000000
```

(viii) There is a strong positive correlation between wind speed and wind power at a value of 0.89. There is also a moderate negative correlation between humidity and solar irradiance. There is a weak correlation between all other variables.

(ix) One interesting trend and finding is that the correlation of wind speed and wind power with their respective lagged values is strong, at values of 0.9155812 and 0.8199444. This shows that the lagged values may be useful to accurately forecast. Wind speed also has a strong correlation with wind power at a value of 0.8885186, so these may also be useful to forecast speed and power.



(x) The plots seen previously show a relationship between wind power and time. Similar to the plots of wind speed versus all other variables, there does not seem to be a clear relationship between wind power and humidity and wind power and solar irradiance. There appears to be a slight relationship between wind power and temperature, because we observe that higher wind speeds tend to occur at higher temperatures.

Task 3: Pseudocode: Predicting wind power given wind speed

The pseudocode allowed the team to define the blueprint and methodical approach to predicting wind power. Visualizing the data showed which inputs and correlations could be used within the models chosen to train the data. Moving forward, most efforts will be dedicated to finding the model that encompasses the insights from task 2.