MMA 860 Assignment 1

**General Instructions**

This assignment consists of the following four questions and the data found in the workbook “MMA860\_Assignment1\_Data.xlsx”.

Please upload a single Word or Jupyter Notebook (as PDF) document that includes **your code, answers, and visualizations as required. Only your Word document (and PowerPoint for question 2) will be graded.**

When asked to explain something, assume that person has a non-technical background and requires a brief but complete explanation of the issue. Use this as an opportunity to demonstrate your thorough understanding of the material and your ability to communicate it effectively.

1. The local electricity utility company is thinking about installing 49 wind turbines in Lake Ontario, just south of the Toronto Islands. Despite how you feel about wind turbines, you have been asked to calculate the amount of power that the station could potentially generate using the latest weather data from January. You are provided with three data sets: “Windfarm Turbine Locations”, “Turbine Specifications” and “Island Airport Weather”. Answer the following questions.

**Show all your code and a snippet of the answer for each part of the question.**

* 1. Since our weather data doesn’t include “air density”, you will need to calculate it for each hour of data using the following formula and save it to a new variable and print out the first 5 records. **2 marks**
  2. Windspeed [Wind Spd (km/h)] will need to be converted from km/h to meters per second for each hour of data (1 km/h ≈ 0.277778 m/s). Save this to a new variable and print out the first 5 results. Include entries even if no power was generated. **2 marks**
  3. To calculate the power production at each turbine, use the following formula, keeping the following points in mind, save it to a new variable and print out the first 5 records. **3 marks**

Points:

1. If the windspeed is above or below the minimum and maximum cutoff speeds, the wind turbines will not operate
2. Each turbine cannot produce more than [Turbine nominal power] at any given point in time. Note that [Turbine nominal power] is given in Megawatts, where your power calculation is in watts (1 MW = 1,000,000 watts)
   1. What is the total amount of electricity produced for the entire windfarm in January in Megawatts? *Hint: it’s the sum of all power produced part c, above, converted to Megawatts.* ***1 mark***
   2. Create a visualization using the best practices we discussed showing power produced per day at the windfarm in January. **2 marks**
3. Present the data and a story to the executive team at Ontario Power Generation.
   1. Create either: **3 marks**
      1. A consulting-style PowerPoint presentation (maximum 3 slides) **OR**
      2. A business memo (maximum 1 page, not including visuals).
   2. Produce at least 2 visualizations and include them in your presentation or memo. Do not re-use the visualization from question 1. **4 marks**
   3. Make sure you include recommendations as part of the story. **3 marks**

The output should not contain *any* overtlytechnical material. Given you do not have any organizational or strategy information, feel free to make any assumptions you need and state them. If you are using PowerPoint submit it as a separate file. Any Tableau visualizations should be pasted into PowerPoint or Word and submitted that way.

Hint: we learned two storytelling narrative techniques – we’ll be looking for evidence of those!

Extra hint: any reasonable assumptions are fair. Perhaps you have been contracted to determine if the turbines are worth installing, or that the turbines have been purchased but now need a location. I am flexible here.

1. Your manager has provided a dataset on product sales, but says she knows little about it. She has told you that it is a bit messy and too difficult for her to work on. Now that you are an expert in Python, she would like you to do the following. Show all your code and a **small** snippet of relevant output.
   1. Make sure all variable formats are correct, and that Product\_IDs are all of the same length, and a string. **2 marks**
   2. Tidy the dataset (i.e., make sure all columns are unique variables, all rows are unique observations, and there is a single data point in each cell). **2 marks**
   3. Create a visualization of price with appropriate labels. **2 marks**
   4. Calculate the correlation between price and number of retailers (In python: np.corrcoef()). Produce a scatterplot to visualize it, and explain what the correlation means in practical terms (i.e., to a business owner) **3 marks**
   5. Tell a story with the data - produce an easy to understand visualization and describe the insights you’ve identified to your manager. Make sure the visualization can stand on its own without explanation, and that your story has a **specific business purpose or outcome.** **3 marks**
2. Collinearity: collections of variables that tend to move together, such as height and weight, are called collinear. This creates some challenges for analysis in that individual t-statistics tend to be less informative. Using the data found on the tab ‘Collinear’:
   1. Filter the data to consider only the first 25 observations then run the following models; **repeat the analysis with all 100 observations**. You do not need to worry about standard data problems such as heteroscedasticity, etc. You must run 6 regressions in total – i, ii, and iii with 25 observations + i, ii, iii with 100 observations. Include a summary of your models and how you reached your conclusion.
      1. Run a linear regression to explain y in terms of experience and height. Does height appear to explain y? **2 marks x2**
      2. Run a linear regression to explain y in terms of experience and weight. Does weight appear to explain y? **2 marks x2**
      3. Run a linear regression to explain y in terms of experience and height and weight. Do height and weight appear to explain y? **2 marks x2**
   2. Consider the results you have found from the work in a. Write a paragraph or two to explain to your manager the patterns you observed with respect to the significance of the t-statistics, why these results occurred, and the strategies for using explanatory variables that exhibit collinearity. **3 marks**