midterm

August 19, 2024

1 STAT40800 Midterm Assignment

1.1 Exploratory data analysis of the Irish weather

For this assignment you will perform an exploratory data analysis (EDA) of historic weather data from Met Eireann, Ireland's main meteorological service.

1.1.1 Instructions

- Solutions must be submitted on Brightspace under *Assessments* → *Midterm Assignment*.
- Your submission must include your completed Jupyter notebook in .ipynb **and** PDF format. To create a PDF of your notebook select *File* → *Download as* → *PDF via LaTeX (.pdf)*. [Alternatively *Save and export Notebook as* → *PDF*]
- All of the results that you wish to include should be viewable without running the Python code. Note that the code may still be run by the grader to check that it functions properly and as intended.
- Marks will be awarded for complete and correct answers to all 9 questions. An additional 10 marks will be reserved for organisation, presentation and conciseness.
- For full marks, you must justify your answers, clearly explain all steps and computations, label your figures, and write concise code.
- The assignment must be completed individually.

This assignment is **due at 11:59pm on November 3rd, 2024**.

[ ]:
```
# Load in necessary packages
```

1. Load in the weather_1819.csv dataset into Python as a pandas DataFrame. Describe the data. How many years of recordings are included? What is the temporal resolution of the data? Which weather measurements are reported? (8 marks)

[ ]:
```

```

2. Determine how many missing values there are in each column of the dataset. Can you think of a reason why these values are missing? Discuss different strategies for filling the missing values, highlighting the advantages and disadvantages of each strategy, in the context of this dataset. (8 marks)

   Note: You do not need to implement any of your suggested strategies.

[ ]:
```

```

3. Write code to answer the following questions: (15 marks)
    1. At what station and on what date was the highest wind speed recorded?
    2. At what station and on what date was the highest maximum air temperature recorded?
    3. At what station and on what date was the largest amount of rain recorded?

[ ]:

4. Create a numerical summary (mean, standard deviation, minimum, maximum, etc.) for each of the weather measurements. Discuss and interpret your results. (8 marks)

[ ]:

5. Create a graphical summary for each of the weather measurements. Discuss your plots in relation to the summary statistics found in question 4. (10 marks)

[ ]:

6. Produce a scatter plot of the mean wind speed versus the highest gust and colour your points based on month. Interpret your plot. (8 marks)

[ ]:

7. Compute the daily temperature range, and add this as an additional variable to your DataFrame. Print out the last 10 rows of your DataFrame to show that the column has been added correctly. (5 marks)

[ ]:

8. Plot the daily temperature range versus the hours of sunlight per day, colouring the points based on month. Interpret your plot. (8 marks)

[ ]:

9. Perform a comparative analysis of the weather at Dublin Airport, Shannon Airport and Cork Airport. (20 marks)

    For full marks on this question you should create numerical and graphical summaries of the weather measurements at each weather station and discuss how the weather differs (or is similar) across these locations.

[ ]: