Programming for Data Analytic

1.1 Task Specification

This project contains five individual tasks on a given dataset called *weatherAUS.csv*. Please note that each task is required to be implemented as one function. A template file (*template.py*) is provided where you need to implement all the functions in that file.

1.1.1 Task 1

Read the dataset into an appropriate format and calculate the average rainfall for each individual location in the country. Sort the locations according to their average rain fall in descending order and print them all where each line contains the name of the location and the average of rainfall.

1.1.2 Task 2

Read the dataset into an appropriate format and identify the maximum temperature for each individual location in the country. Sort the locations according to their maximum temperature in descending order Use an appropriate visualization and only visualize the locations that their maximum temperature is greater than 46.6; see Figure 1.

1.1.3 Task 3

Read the dataset and create two sub-datasets as follows:

1. **DataSet1:** *WindSpeed9am*, *Humidity9am*, *Pressure9am*, , *RainTomorrow*.
2. **DataSet2:** *WindSpeed3pm*, *Humidity3pm*, *Pressure3pm*, *RainTomorrow*

Run a supervised learning algorithm (Decision Tree Classifier) and predict *RainTomorrow* (target attribute). Use 0.33 of the data as Test set and the rest as Training set.

Report the accuracy for both datasets and discuss which of them is better and why? Use comment for discussion.

1.1.4 Task 4

Apply K-Neighbors-Classifier and decide which of the following can predict *RainTomorrow* better?:

- **MaxTemp**
- **MinTemp**
- **WindGustSpeed**
- **Rainfall**

Explain the process as the comment below the function.

1.1.5 Task 5

Apply an unsupervised clustering algorithm on a dataset with the following attributes:

- **WindGustDir**,
- **WindGustSpeed**
- **WindSpeed9am**
- **WindSpeed3pm**

and decide what would be the best number of clusters using elbow method. Visualize the elbow method and explain your finding as a comment below the function.

Note that visualization plots need to have proper labels and annotations.

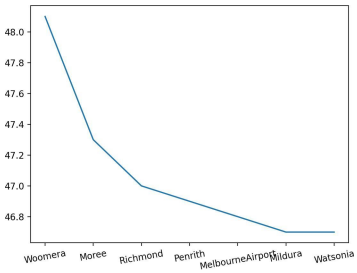

Figure 1: Maximum Temperature.