## Declaration on Plagiarism

*This form must be filled in and completed by the student(s) submitting an assignment*

|  |  |
| --- | --- |
| **Name:** | Anubhav Gupta |
| **Student Number:** | 19211171 |
| **Programme:** | MSc. In Computing |
| **Module Code:** | CA682 |
| **Assignment Title:** | Data Visualisation |
| **Submission Date:** | 13 Dec 2019 |
| **Module Coordinator:** | Dr Suzanne Little |

I declare that this material, which I now submit for assessment, is entirely my own work and has not been taken from the work of others, save and to the extent that such work has been cited and acknowledged within the text of my work. I understand that plagiarism, collusion, and copying are grave and serious offences in the university and accept the penalties that would be imposed should I engage in plagiarism, collusion or copying. I have read and understood the Assignment Regulations. I have identified and included the source of all facts, ideas, opinions, and viewpoints of others in the assignment references. Direct quotations from books, journal articles, internet sources, module text, or any other source whatsoever are acknowledged and the source cited are identified in the assignment references. This assignment, or any part of it, has not been previously submitted by me or any other person for assessment on this or any other course of study.

I have read and understood the referencing guidelines found at <http://www.dcu.ie/info/regulations/plagiarism.shtml>, <https://www4.dcu.ie/students/az/plagiarism> and/or recommended in the assignment guidelines

Name: Anubhav Gupta Date: 13/12/2019

**Australia Weather Analysis**

Abstract (max 200 words)

*What is the question or story you are trying to tell?*

*The question that I was aiming to, was the weather condition in Australia and how the weather changed in different cities at different years. I tried to visualize the weather conditions like windspeed, temperature and rainfall in a period from 2008 to 2017.*

*What is the conclusion that you reached?*

Looking at different attributes I found rainfall, windspeed and temperature to be the most important aspects so keeping in mind those three aspects I used them to build my graph also I want to show rainfall temperature and windspeed of each city individually at every point of time according to the dataset.

1. Dataset [½ page]

*Where/how did you retrieve it or them*

*The dataset was taken from Kaggle.*

<https://www.kaggle.com/jsphyg/weather-dataset-rattle-package>

*Describe the data - size (GB or attributes), number of rows, attributes, data types present*

*The data contains weather observations from numerous weather stations in Australia. The data was of about 15 Mb and it was having around 142000 rows and the dataset comprised of different attributes on which the visualisation was made like max temperature, minimum temperature, location, rainfall, date, etc.*

*What aspects (if any) of big data (volume, variety, velocity) are present in your data*

The data is comprised of the volume aspect of the big data as the data set is of 140,000 rows and is a large dataset also I was looking to integrate some other datasets to increase the variety but couldn’t able to do so.

2. Data Exploration, Processing, Cleaning and/or Integration [½ page]

*What did you need to do to prepare the dataset(s) to create your graph/chart?*

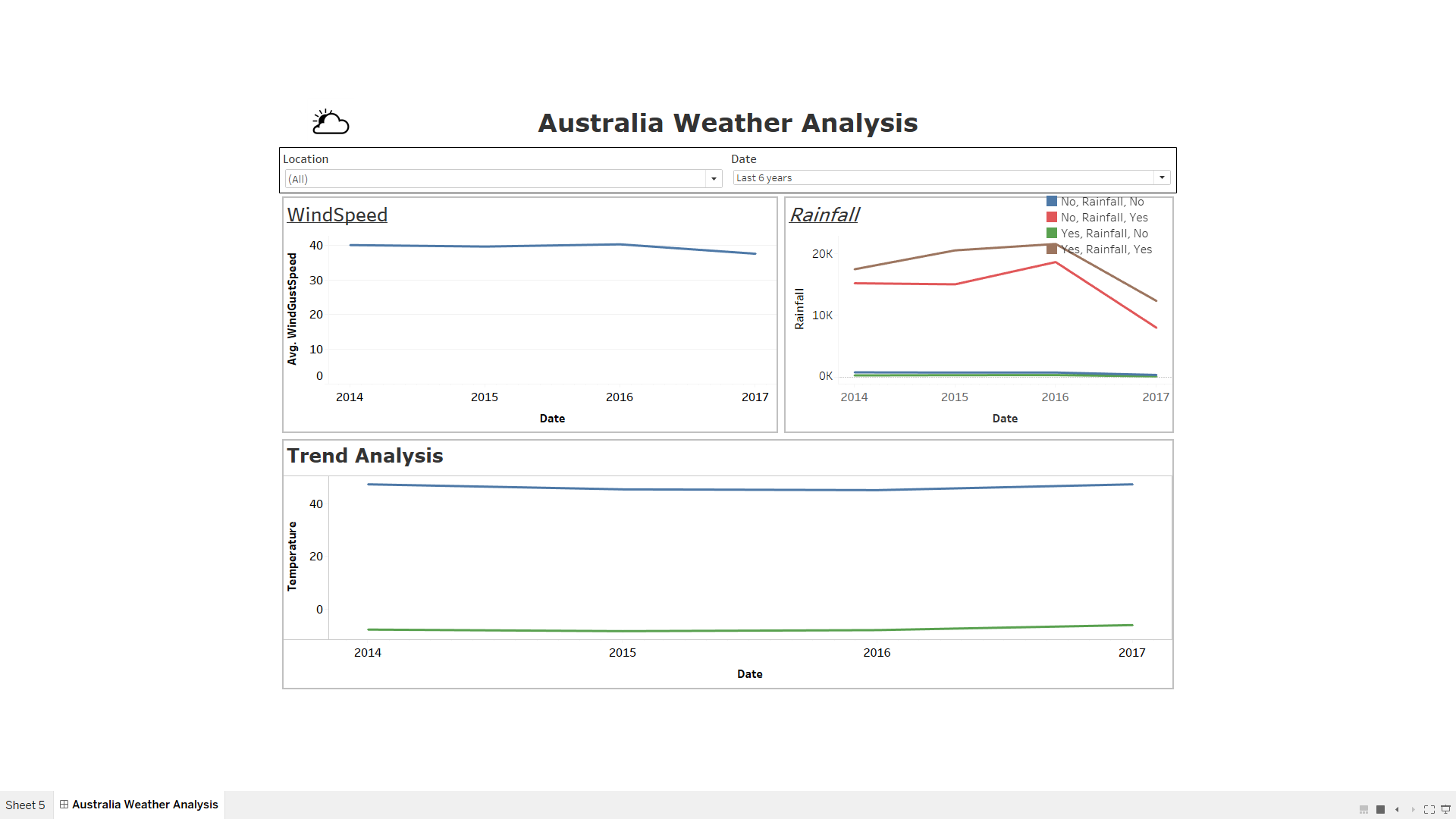
*I processed the data using python and I cleaned the dataset using excel, python and tableau to remove unwanted columns and null values and studied the dataset to make different visualisations and finally using tableau I plot the graphs using different filters.*

*How did you choose the attributes to visualise?*

I choose the attributes based on what I want to visualize, I used different locations as a filter and graphs were plotted to visualize the weather conditions in every city individually or combined.

3. Visualisation [½-1 page]

*Screenshot or image of visualisation*



*Explain your choice of chart or graph type - what relationship or data type are you showing?*

*I choose line charts for my visualisation because they provide a very precise format for anyone to understand the visualisation. Using different colours within line charts makes them more attractive and easier to identify different trends and analyse the visualisation.*

*Design choices - justify your use of colour, shapes, marks, layout, structure, font, labels*

*I used colours in my line chart for differentiating maximum and minimum temperature and differentiating rainfall trends.*

*Any interactivity or animation and how it helps answer your question*

*I used location and date as filters to make it more interactive to visualize the trends. It tells weather conditions of every city individually in a specific interval of time and all cities combined or even 2 or more cities combined.*

*List of tools or libraries used*

*I used Python (NumPy and pandas) libraries to clean the data set and tableau to visualize the data.*

4. Conclusion [½ page]

*Critically analyse the outcome of your visualisation.*

*The data I choose was based on the weather conditions in Australia as recorded by different Australian weather stations over a period of 10 years, the outcome of my visualisation was to identify the trend in weather conditions for example comparing the temperature difference in those 10 years the amount of rainfall received and the windspeed. The visualisation can be used to predict the future weather conditions and for comparison in future*

*Were there aspects that you think could be improved upon?*

*Yes, I could make it more interactive or add a map of Australia that can look more easier to understand.*

*Were there effects or functionality that you were technically unable to achieve?*

*Yes, while I was visualising the dataset, I wanted to make a graph that could predict the future weather conditions based on the data but couldn’t able to produce one.*

References

*Include any citation of the dataset*

<https://www.kaggle.com/jsphyg/weather-dataset-rattle-package>

*Include links to any tutorial or example that contributed significantly to your work*

<https://www.coursera.org/specializations/data-visualization?>

<https://community.tableau.com/community/ideas/content>

*Include any articles or web resources supporting your design choices*