7COM1079-0901-2024 - Team Research and Development Project

Final report title: Analyzing Weather Data in R

Group ID: A82

Dataset number: DS031

Prepared by: Mohsin Raza Naqvi 23031814

University of Hertfordshire
Hatfield, 2024

Table of Contents

1. Conclusions

1.1. Results explained

The study showed some interesting features of the dataset. There was a fairly good agreement observed between temperature and humidity data by plotting them in the form of scatter plots. Retrospective analysis of the temperature using the boxplots has depicted the fluctuations between countries and the heatmap has shown a strong correlation between the air quality parameters. On the other hand, the hypothesis tests, Pearson correlation and Chi-square test, conducted to answer the research question reveals that the variables temperature and air quality index (US EPA) are correlated. This implies that any changes in the either one of these two variables reflect a change in the other one. As can be deduced from the graphs above the temperature process is a stationary process and thus the ARIMA model was able to forecast the trends.

1.2. Interpretation of the results

This study shows that temperature trends in regions and humidity all have an important bearing on the weather. The analysis utilizing the ARIMA model showed the possibility of precise temperature predictions will go a long way in climate change variability research. The relationships between the variables located on the heatmap could be indicative of linkages between the factors in the environment and may well be signals of large-scale interactions in ecosystem management.

1.3. Reasons and/or implications for future work, limitations of your study

In future work it may be useful to investigate higher-level time series models such as seasonal ARIMA to further refine forecast performance. Moreover, some of the measures that have to be done involve determining if there is a missing value of the data set and if there are outliers to be dealt with. Failures of the study include; the absence of finer weather data resolution and possible real inaccuracies in the temperature measurement.