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

Final report title: Analyzing Weather Data in R

Group ID: A82

Dataset number: DS031

Prepared by: Sohail Ahmad

University of Hertfordshire
Hatfield, 2024

Table of Contents

1. Introduction

1.1. Problem statement and research motivation

This report aims to examine the findings relevant to a particular research question using quantitative analysis and visualization of data employing statistical tools to verify the existing hypotheses and analyze the characteristics of the dataset. This research seeks to fill the above gaps by conducting a strong analysis of a well-defined problem in the field. As for the structure of this study, it has been designed in a systematic manner which makes it possible to assess the applicability of research objectives satisfactorily, and review all hypotheses, to derive the results that could be genuinely valuable to the existing body of knowledge.

1.2. The data set

The data set used in this study comprises global climate data that include climatic parameters including temperature, rainfall, humidity, wind speed, and pressure. The data include a trend for different geographic locations and periods, which gives an idea of the climate changes. Such a structure allows working with data statistically and visually to reveal tendencies and outliers. Data pre-processing methods are useful in maintaining high data quality that is suitable for hypothesis testing.

1.3. Research question

Is there a correlation between temperature and the air quality index (US EPA) in different global locations?

1.4. Null hypothesis and alternative hypothesis (H0/H1)

Null Hypothesis (H0): As there is no statistical significance between temperature and air_quality_us-epa-index these variables are not correlated.

Alternative Hypothesis (H1): As there exists a statistical significance between temperature and air_quality_us-epa-index these variables are correlated.