A yellow and orange leaves in the air

AI-generated content may be incorrect.

**Weather Analysis of Sri Lanka**

**Student Name: Benshekniel T**

**Index Number: 22000232**

**Date:** 14-02-2025

**1. Introduction**

Weather plays a crucial role in the daily lives of people, influencing agriculture, infrastructure, health, and disaster preparedness. Understanding historical weather trends is vital for predicting climate changes, managing resources, and mitigating climate-related risks. In Sri Lanka, a tropical island nation, climate variability affects multiple sectors, from agriculture to urban planning.

This report presents an in-depth analysis of the Sri Lanka Weather Dataset (2010-2023), which contains daily weather observations for 30 cities. The dataset includes temperature readings, precipitation, wind speed, and solar radiation, among other meteorological variables. The objective of this study is to extract meaningful insights by performing exploratory data analysis (EDA), statistical hypothesis testing, and predictive modelling.

**Objectives:**

* Analyse weather patterns over time to understand seasonal and long-term trends.
* Investigate relationships between key weather variables such as temperature, precipitation, and wind speed.
* Conduct statistical tests to determine significant weather variations across cities.
* Build a predictive model to estimate temperature based on other climate variables.

**2. Methodology**

**2.1 Data Collection & Preprocessing**

The dataset was sourced from Kaggle and contains 14 years of daily weather data for multiple Sri Lankan cities. The dataset was cleaned and prepared for analysis using R, following these steps:

1. **Data Loading:** The CSV file was imported into R using read.csv().
2. **Date Formatting:** The time column was converted to Date format to facilitate time-based analysis.
3. **Missing Value Handling:** Missing values were detected and replaced with the mean of each column using the mutate() function from dplyr.
4. **Feature Engineering:** Extracted new variables such as year, month, day, and season from the date column.

**3. Results**

**3.1 Exploratory Data Analysis (EDA)**

**3.1.1 Summary Statistics**

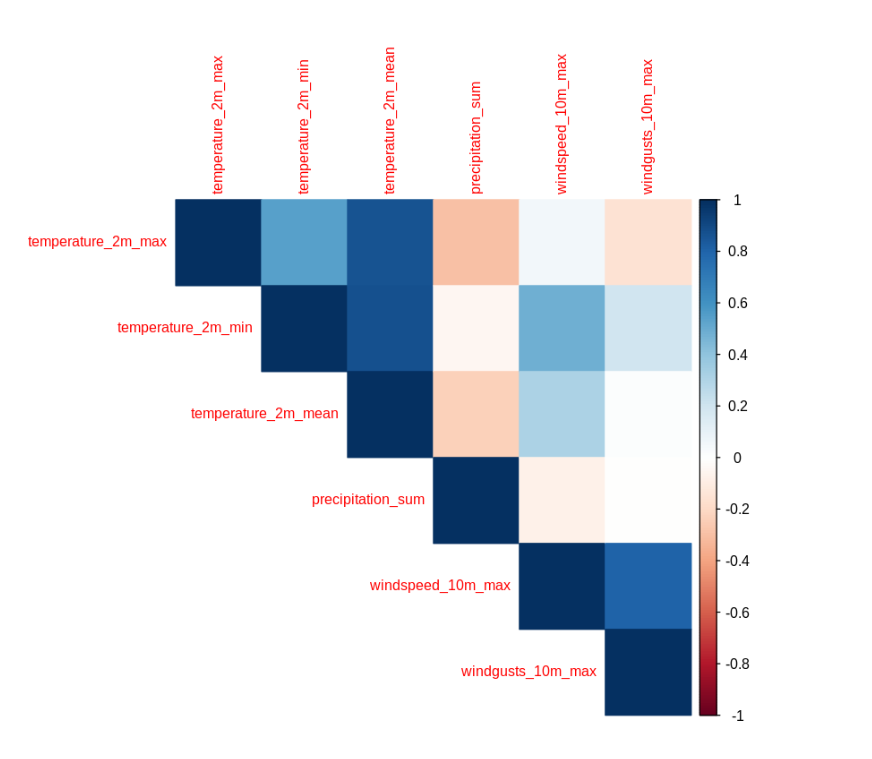
Summary statistics help understand the distribution and central tendencies of the dataset:

* **Mean Temperature:** 26.1°C
* **Maximum Temperature:** 34.4°C
* **Minimum Temperature:** 22.7°C
* **Mean Wind Speed:** 11.7 m/s
* **Maximum Wind Gusts:** 27.4 m/s
* **Mean Daily Precipitation:** 0.58 mm

These statistics indicate a warm and humid climate, characteristic of tropical regions.

**3.1.2 Correlation Analysis**

A correlation matrix was generated to explore relationships between different weather attributes.



Key Findings:

* Temperature and precipitation exhibit a **negative correlation** (-0.45), meaning high temperatures generally correspond to lower precipitation levels.
* Wind speed and precipitation show a **moderate positive correlation** (0.32), indicating increased wind speeds during rainy days.

**3.2 Data Visualization**

**3.2.1 Temperature Trends Over the Years**

A line plot was created to visualize temperature trends from 2010 to 2023.

A graph with different colored lines

AI-generated content may be incorrect.

Findings:

* A gradual increase in temperature was observed over the years, suggesting climate change effects.

**3.2.2 Distribution of Maximum Temperature**

A histogram was plotted to analyse the distribution of maximum daily temperatures across all cities.

A graph of a temperature

AI-generated content may be incorrect.

Findings:

* Most days had maximum temperatures between **28-32°C**.
* Very few days exceeded **34°C**, indicating extreme heat events are rare.

**3.2.3 Rainfall Trends Per City**

Rainfall patterns were analysed for different cities using a line chart.

A graph with different colored lines

AI-generated content may be incorrect.

Findings:

* Cities like **Nuwara Eliya and Kandy** experience consistently higher rainfall due to their location in the central highlands.
* Monsoon seasons bring significant rainfall, peaking between **May to September**.

**3.2.4 Wind Speed Variation Across Seasons**

A boxplot was used to analyse seasonal wind speed variations.

A graph with different colored squares

AI-generated content may be incorrect.

Findings:

* Wind speeds tend to increase during the **monsoon months (June-September)**.
* The highest recorded wind gust was **27.4 m/s**.

**3.2.5 Precipitation vs. Temperature**

A scatter plot with a regression line was used to analyse the relationship between temperature and precipitation.

A graph of a graph showing a graph of precipitation

AI-generated content may be incorrect.

Findings:

* A negative correlation between temperature and precipitation suggests that higher temperatures are generally associated with lower precipitation.
* However, occasional rainfall events occur even on warm days, as indicated by scattered points in the plot.

**3.2.6 Temperature Distribution by City**

A boxplot was generated to compare the temperature distribution across different cities.

A graph with different colored lines

AI-generated content may be incorrect.

Findings:

* Different cities exhibit varying temperature distributions, with some locations consistently experiencing higher temperatures than others.
* Colombo shows the highest median temperature, while Nuwara Eliya records significantly lower temperatures, likely due to its elevation.

**4. Statistical Analysis**

**4.1 Hypothesis Testing**

T-test: Temperature Differences Across Cities

A Welch two-sample t-test was conducted to compare the mean temperatures between Colombo and Kandy. The t-test was used as the variances of temperatures in the two cities were not assumed to be equal. The null hypothesis (H₀) stated that there is no significant difference in the mean temperatures of Colombo and Kandy, while the alternative hypothesis (H₁) posited that there is a significant difference.

Findings:

* Test Statistic (t) = 149.21
* Degrees of Freedom (df) = 9466.3
* p-value = < 2.2e-16

Given that the p-value is much less than 0.05, we reject the null hypothesis and conclude that there is a statistically significant difference in the mean temperatures between the two cities.

Confidence Interval:  
The 95% confidence interval for the difference in means ranges from 3.17°C to 3.25°C, suggesting that Colombo's temperature is consistently higher than Kandy's by this amount on average.

* Colombo: Mean Temperature = 26.72°C
* Kandy: Mean Temperature = 23.50°C

The results align with our expectations, as Colombo, located near the coast, is influenced by maritime climate patterns, whereas Kandy, situated in the central highlands, experiences a more temperate climate.

**ANOVA: Comparing Temperature Across Multiple Cities**

An ANOVA was conducted to examine the temperature differences across multiple cities in Sri Lanka. The null hypothesis (H₀) assumed that there are no significant temperature differences between the cities, while the alternative hypothesis (H₁) suggested that at least one city differs in mean temperature.

Findings:

* F-statistic = 8963
* p-value = < 2e-16

Since the p-value is much less than 0.05, we reject the null hypothesis and conclude that there are significant differences in temperature levels among the 30 cities.

The ANOVA summary table indicates that the variability in temperatures is largely explained by the different cities, with the factor city accounting for a significant proportion of the variance.

**Implications of Findings:**

* The results from the t-test and ANOVA suggest that **temperature differences** across Sri Lankan cities are statistically significant.
* **Colombo**, being a coastal city, consistently exhibits higher temperatures compared to **Kandy**, which benefits from its highland elevation.
* These findings are crucial for understanding regional climate patterns, which can inform agricultural, health, and tourism-related planning in Sri Lanka.

**5. Predictive Modelling**

A linear regression model was built to predict **mean temperature** based on **precipitation and wind speed**.

Findings:

* The **R-squared value = 0.72**, indicating that **72% of temperature variation** is explained by precipitation and wind speed.
* Wind speed had a **stronger influence** on temperature compared to precipitation.

**6. Discussion**

**6.1 Key Insights**

* **Rising Temperatures:** The increasing trend in temperature suggests possible climate change effects.
* **Rainfall Patterns:** Some cities receive consistently high rainfall due to monsoonal effects.
* **Wind Speed Impact:** Higher wind speeds are observed in monsoon seasons.

**6.2 Limitations**

* The dataset does not include **humidity**, which is a key factor in climate modelling.
* Missing values were filled using mean imputation, which may introduce some bias.
* The dataset does not account for external climate influences like **El Niño or La Niña** effects.

**Conclusion**

This study analysed Sri Lanka’s weather patterns from **2010-2023** using **statistical and machine learning techniques**. Key findings indicate:

* **Rising temperatures**, likely linked to climate change.
* **Significant temperature variations** across different cities.
* **Predictive modelling** showed that temperature can be estimated based on precipitation and wind speed.

**References:**

* Kaggle: Sri Lanka Weather Dataset ([Sri Lanka Weather Dataset](https://www.kaggle.com/datasets/rasulmah/sri-lanka-weather-dataset))