Dr. M.G.R.
EDUCATIONAL AND RESEARCH INSTITUTE
DEEMED TO BE UNIVERSITY
University with Graded Autonomy Status
(An ISO 21001 : 2018 Certified Institution)
Periyar E.V.R. High Road, Maduravoyal, Chennai-95. Tamilnadu, India.

STATISTICAL ANALYSIS OF CLIMATE VARIABILITY ACROSS INDIA'S DIVERSE REGIONS

: A MULTI-STATION STUDY (1980-2023)

Submitted In Partial Fulfillment of The Requirements for The
Award Of the Degree
In
Statistics

By

Mr. Ruturaj Tanaji Saravane
(233522311033)

Department Of Statistics
January 2023

DECLARATION

I, **Ruturaj T. Saravane** (**233522311033**) hereby declare that the Project Report entitles "**Statistical Analysis of Climate Variability Across India's Diverse Regions**" is done by me under the guidance of **Dr. R Nandhinidevi**, Asst. Professor, is submitted in partial fulfillment of the requirements for the award of the degree in MASTER OF SCIENCE in Statistics.

Date:

Place: Chennai **Signature of The Candidate**

Dr. M.G.R.
EDUCATIONAL AND RESEARCH INSTITUTE
DEEMED TO BE UNIVERSITY
University with Graded Autonomy Status
(An ISO 21001 : 2018 Certified Institution)
Periyar E.V.R. High Road, Maduravoyal, Chennai-95. Tamilnadu, India.

DEPARTMENT OF STATISTICS

BONAFIDE CERTIFICATE

This is to certify that the Project Report is the Bonafide work of **Ruturaj T. Saravane (233522311033),** who carried out the project entitled **"STATISTICAL ANALYSIS OF CLIMATE VARIABILITY ACROSS INDIA'S DIVERSE REGIONS"** under my Supervision from October 2024 to December 2024.

Dr. R. Nandhinidevi

Department of Statistics

Dr. M.G.R. Educational and Research Institute

ACKNOWLEDGEMENT

I would first like to express my deepest gratitude to my college's esteemed administration for their unwavering support throughout my academic journey and the successful completion of this project.

I extend my heartfelt thanks to the Chancellor **Thiru A. C. Shanmugam, B.A, B.L** and President **Er. A. C. S. Arun Kumar, B.Tech., M.B.A,** also our Secretary **Thiru A. Ravikumar**, for fostering an encouraging academic environment and providing the necessary resources for my research.

I am profoundly grateful to my Head of Department, **Prof. Dr. T. Johnson**, for their invaluable guidance and insights, which were instrumental in shaping my project on climate data analysis.

A special thanks to my Project Guide, **Dr. R Nandhinidevi**, for their constant support, patience, and expert advice throughout this endeavour. Their mentorship and encouragement have been key to overcoming challenges and achieving success.

I also wish to acknowledge the efforts of the teaching and non-teaching staff of the Department of Statistics, whose assistance and encouragement have motivated me to excel in my work.

Finally, I would like to thank the Indian Meteorological Department (IMD) for providing the necessary datasets that served as the foundation for this research. Their contributions were indispensable for conducting this study and drawing meaningful conclusions.

To my family and friends, who have been my pillars of support during this journey, I owe my sincere appreciation for their unwavering belief in me.

INDEX

Sr. No.	Contents	Page No.
1.	Abstract	
2.	Chapter 1: Introduction	01
3.	Chapter 2: Literature Survey	02-03
4.	Chapter 3: Aim And Objectives of The Project Study	04
	3.1. Specific Objectives 3.2. Scope of the Present Investigation	
5.	Chapter 4: Context to the Selected Regions	05
6.	Chapter 5: Research Methodology	07-15
	5.1 Data Collection 5.2 Data Preparation and Cleaning 5.3 Data Analysis 5.4 Anova With Post Hoc Test 5.5 Tools and Software 5.6 Limitations 5.7 Conclusion of Methodology	
7.	Chapter 6: Analysis, Results, and Interpretation	20-71
	6.1 Descriptive Statistics 6.2 Exploratory Analysis 6.3 Time Series Analysis 6.4 Extreme Value Analysis 6.5 Multivariate Comparison	
8.	Chapter 7: Major Findings and Conclusion	72-74
9.	Chapter 8: References	75

ABSTRACT

This project investigates climate variability across diverse regions of India using daily data from 7 meteorological stations, spanning from 1980 to 2023. Focusing on key parameters such as temperature, rainfall, wind speed, evaporation, this study aims to understand how regional climate and surface conditions interact and affect overall climate across India’s varied climatic zones—ranging from tropical and coastal to mountainous and arid regions.

The significance of this study lies in addressing how seasonal and long-term climate patterns vary across regions and how these patterns influence the weather. This analysis provides valuable insights into critical questions: How do seasonal changes in temperature and rainfall differ across regions? What are the conditions under which marine climate affects the coastal regions?

To answer these questions, the project uses statistical tools such as time series analysis for trend detection in climate data, as well as correlation analysis to study the relationships between weather parameters. Parametric tests are used to define the significant relation between parameters. Also, the study of extreme values conducted through GEV distribution. Data visualization, including histograms, boxplots, trend graphs and heatmaps, presents these findings in an accessible and interpretable format.

The outcomes of this analysis contribute to a better understanding of how climate variability and surface variables influence weather across different regions of India. These insights have practical applications for temperature and rainfall management, urban planning, and public health, especially in areas prone to high temperature or rainfall. By identifying regional patterns in climate variability, this study offers a valuable resource for policymakers, environmental agencies, and researchers focused on promoting sustainable, climate-resilient communities across India.

Chapter 1: INTRODUCTION

India, with its diverse topography and climatic zones, experiences a wide range of weather patterns that significantly influence environmental conditions. Understanding the interplay between climate variability, marine and upper air variables is crucial for effective environmental management and public health protection. This project delves into a comprehensive statistical analysis of these factors across India's diverse regions, utilizing meteorological data from 7 cities spanning the period 1980-2023.

The motivation behind this study stems from the growing concern about climate pollution and its impact on human health and the environment, particularly in rapidly developing nations like India. By examining long-term trends and regional variations in key meteorological parameters such as temperature, rainfall, wind speed, and evaporation, I aim to uncover how these factors contribute to climate stability.

This research employs a robust statistical framework, including time series analysis to detect trends in climate data, correlation analysis to investigate relationships between weather parameters and parametric tests to identify periods of stable atmospheric conditions conducive to extreme event accumulation.

This analysis has significant implications for climate variables management, urban planning, and public health policies in India. By identifying regional vulnerabilities and the influence of climate variability on earth’s surface, this research can contribute to the development of effective strategies for mitigating inconsistencies in climate and promoting sustainable environmental practices.

Chapter 2: LITERATURE SURVEY

This literature survey examines existing research on climate variability and atmospheric stability in India, with a focus on the selected regions of New Delhi, Jodhpur, Guwahati, Bhopal, Mumbai, Chennai, and Thiruvananthapuram. This provides a foundation for understanding the long-term changes and regional nuances in meteorological conditions that influence atmospheric stability and, consequently, air quality.

2.1 CLIMATE VARIABILITY IN INDIA –

- **Regional Variations:**
  Studies have highlighted significant regional variations in climate trends across India. For example, the Himalayan region, which influences Guwahati's climate, exhibits a higher warming trend compared to peninsular India (Krishnan et al., 2020). The arid region of Jodhpur experiences distinct temperature and rainfall patterns compared to the coastal cities of Mumbai and Chennai.

- **Monsoon Dynamics:**
  The Indian monsoon plays a crucial role in climate variability, particularly in Mumbai, Chennai, and Thiruvananthapuram. Research has focused on understanding changes in monsoon patterns, including rainfall intensity, frequency, and onset (Goswami et al., 2006).

- **Extreme Events:**
  Several studies have investigated the increasing frequency and intensity of extreme weather events such as droughts, floods, and heat waves in India, linking them to climate change (Guhathakurta & Rajeevan, 2008). These events can have significant impacts on atmospheric stability and air quality in cities like New Delhi and Bhopal.

2.2 UPPER AIR DATA:

- **Atmospheric Variables**:

  Studies on upper air data, specifically focusing on parameters like dry bulb temperature, wet bulb temperature, cloud height, and cloud amount, have provided valuable insights into atmospheric processes and climate variability.

- Research has shown a strong correlation between rising dry bulb and wet bulb temperatures in the upper atmosphere and increasing global temperatures (Mokhov & Semenov, 2000). This trend has significant implications for understanding the greenhouse effect and predicting future climate scenarios.

- **Cloud Height and Amount**:

  Analysis of cloud height and cloud amount data has revealed complex interactions between cloud formation, atmospheric circulation patterns, and regional climate variability (Norris & Wild, 2007). Changes in cloud cover and height can influence radiative forcing, precipitation patterns, and overall energy balance in the atmosphere.

2.3 FOCUS ON SURFACE DATA:

- **Temperature and Wind Profiles:** Several studies have analyzed surface temperature and wind data to understand atmospheric stability and boundary layer characteristics in different regions of India (Sikka & Gadgil, 1980).
- **Stability Indices:** Research has utilized surface meteorological data to calculate atmospheric stability indices and assess pollution potential in various Indian cities (Arya, 1999).

This project will contribute to the existing literature by conducting a detailed statistical analysis of climate variability and atmospheric stability across the selected regions using a long-term dataset (1980-2023) of surface meteorological data. The findings will enhance our understanding of the regional nuances in weather dynamics and provide valuable insights for urban planning and environmental management in India.

Chapter 3: AIM AND OBJECTIVES OF THE PROJECT STUDY

The primary aim of this investigation is to conduct a comprehensive statistical analysis of climate variability across diverse regions of India, utilizing a multi-station approach and a long-term dataset (1980-2023). This analysis will focus on understanding how meteorological factors contribute to surface conditions.

3.1 Specific Objectives:

1. To Explore the Climate Variability Using Descriptive Statistics.
2. Investigate The Relationship Between Temperature, Rainfall, Wind Speed and Evaporation Patterns.
3. Quantify Regional Climate Trends in Temperature, Rainfall, And Other Parameters.
4. Examine Inter-Annual Variability and Climate Seasonality of Temperature and Rainfall Over Selected Cities.
5. To Conduct Extreme Event Analysis in Case of High Temperature and Heavy Rainfall and To Predict Possible Future Extreme Magnitudes.
6. To Test the Relationship and Significance of Dry Bulb Temperature and Wet Bulb Temperature Parameters with Sea Surface Temperature Using Marine Data.

3.2 Scope of the Present Investigation:

This investigation will utilize daily meteorological data from seven stations representing diverse climatic zones across India, covering a period of 43 years (1980-2023). The scope encompasses,

- **Spatial Coverage:** Analysis of seven geographically diverse regions,
- **Temporal Coverage:** Analysis of daily, monthly, and annual data over a 43-year period (1980-2023), allowing for the detection of long-term trends and cyclical patterns.

By combining a multi-station approach with a long-term dataset and robust statistical analysis, this investigation aims to provide valuable insights into the complex relationship between climate variability and atmospheric stability across India. The findings will contribute to a better understanding of regional differences in atmospheric dynamics and have implications for urban planning, and environmental policy in India.

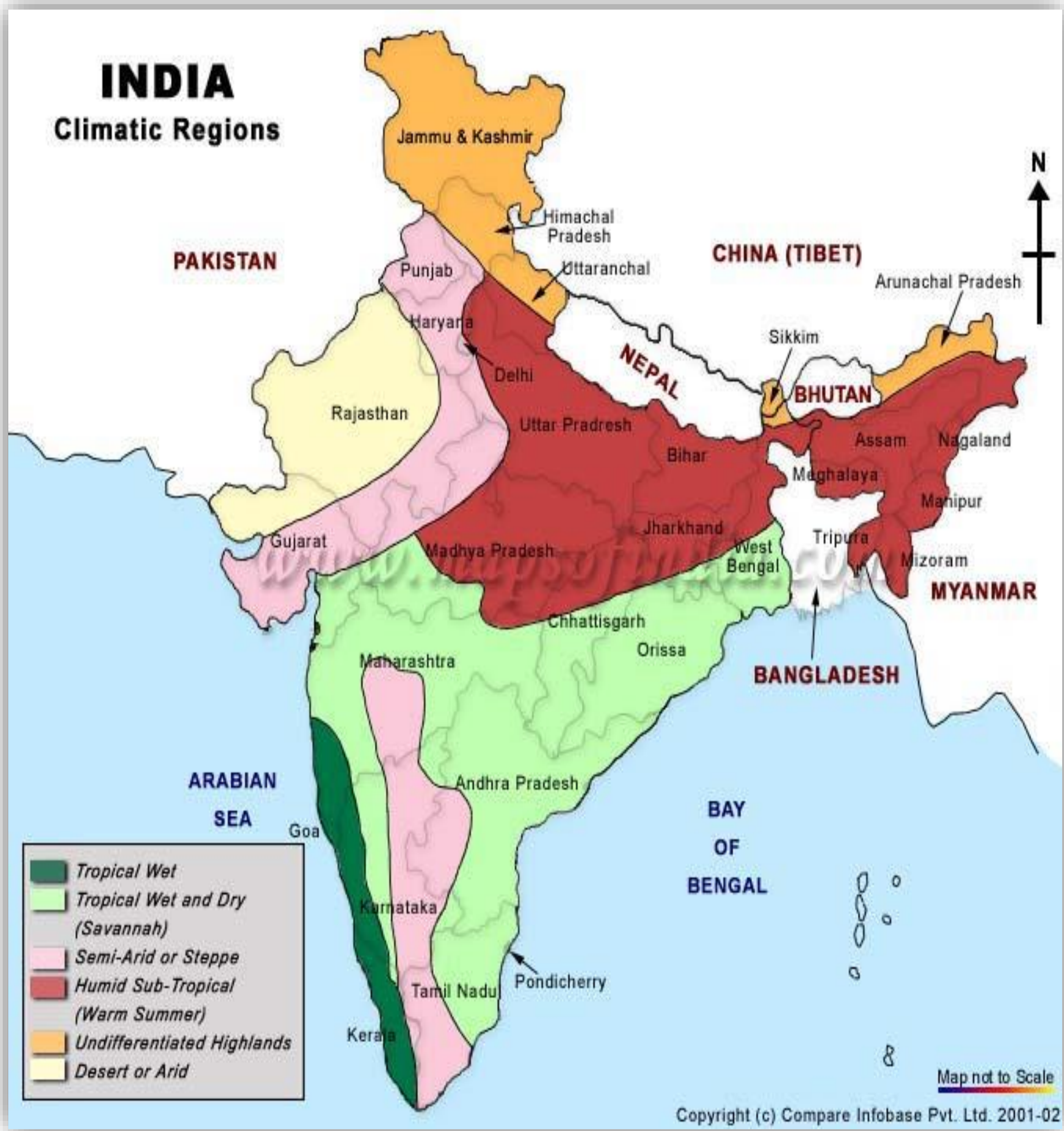
Chapter 4: CONTEXT TO THE SELECTED REGIONS

This study focuses on analysing climate variability across seven geographically diverse cities in India. These cities were strategically selected to represent a range geographical features, ensuring a comprehensive assessment of climate variability across the country.

- **New Delhi**: The capital city, located in northern India, experiences a humid subtropical climate with hot, dry summers and cool, relatively dry winters. The city lies within the Indo-Gangetic Plain and is influenced by the monsoon season, which brings significant rainfall from June to September.
- **Jodhpur**: Situated in the western state of Rajasthan, Jodhpur is characterized by a hot desert climate with extremely high temperatures and low precipitation. The city lies on the edge of the Thar Desert, a vast, arid region with sandy terrain and sparse vegetation.
- **Guwahati**: Located in the northeastern state of Assam, Guwahati experiences a humid subtropical climate with high humidity and heavy rainfall throughout the year, particularly during the monsoon season.
- **Bhopal**: Situated in central India, Bhopal has a tropical savanna climate (Aw) with distinct wet and dry seasons. The city is known as the "City of Lakes" due to its numerous artificial and natural lakes, which contribute to its moderate climate.
- **Mumbai**: Located on the western coast of India, Mumbai experiences a tropical wet and dry climate with warm temperatures and high humidity. The city's coastal location moderates temperatures and brings significant rainfall during the monsoon season.
- **Chennai**: Situated on the southeastern coast, Chennai has a hot semi-arid climate with hot, humid weather for most of the year. The city's coastal location influences its climate, with sea breezes providing some relief from the heat.
- **Thiruvananthapuram**: Located in the southernmost state of Kerala, Thiruvananthapuram experiences a tropical monsoon climate with high humidity and heavy rainfall, especially during the monsoon season. The city is situated on the Malabar Coast, surrounded by hills and lush greenery.

A map illustrating the geographical locations of these cities is presented in Figure below. The diverse climatic conditions and geographical settings of these cities provide a representative sample for investigating the complexities of climate variability across India.

Map of Climatic Zones of India



Chapter 5: RESEARCH METHODOLOGY

5.1 DATA COLLECTION –

This study on climate variability utilizes datasets obtained from the Indian Meteorological Department (IMD) through its data request portal. The datasets include surface data, marine data, and upper air data. Due to significant gaps in certain datasets, only the following key parameters were considered for analysis:

- Maximum and Minimum Temperature
- Rainfall
- Wind Speed
- Evaporation

Marine Data by Longitude and Latitude –

- Dry Bulb Temperature
- Wet Bulb Temperature

5.2 DATA PREPARATION AND CLEANING –

Before conducting any statistical analysis, the raw meteorological data obtained from the Indian Meteorological Department (IMD) underwent a series of preprocessing steps to ensure data quality and consistency.

The acquired data was provided in .txt format, making direct analysis challenging. To facilitate processing, the datasets were converted into Excel files. Handling of missing and irrelevant values was carried out using Python programming, specifically employing a combination of **interpolation** and **weighted moving average (WMA)** techniques. This approach ensured the preservation of data integrity and continuity.

Subsequently, the datasets were formatted into a usable structure by:

1. Extracting monthly and annual averages from the original robust datasets.
2. Identifying and eliminating irrelevant parameters.
3. Removing inconsistent columns that could compromise analysis accuracy.

5.2.1 Use of Interpolation with WMA –

In this combined method, all the missing values indexed in a single array. Using Linear Interpolation, all the missing values were calculated and then applied WMA method on computed values and then replaced those new values with respective indexed missing places.

```python
import numpy as np
import pandas as pd

df = pd.read_excel("C:/Users/rutur/Desktop/M.Sc Project/for code on big data/For Code BigData.xlsx")
df2 = df[0:]
var = df2.columns.to_list()

for v in var:
    #print(f'\n\nFor {v}')
    mainsr = df2[v]
    nindex = mainsr[mainsr.isna()].index.tolist()
    #print(nindex)
    intpsr = mainsr.interpolate(method = 'linear')
    if np.isnan(intpsr[0]) == True:
        intpsr[0] = intpsr[1:4].mean()
    #print(intpsr)

    wts = [1/4,1/3,1/2,1,1/2,1/3,1/4]
    finvs = []
    for n in nindex:
        #print(f'\nfor {n}')
        if n == 0:
            finv = intpsr[0]

        elif 0 < n < 3:
            surv = intpsr[:n+4].to_numpy()
            #print(f'surrounding_values for {n}: \n' , surv)
            wsum = sum(wts[n-1:])
            wdsum = sum(w*v for w, v in zip(wts[n-1:], surv))
            #print('\nweight_sum: ',wsum, '\nweighted_sum: ', wdsum)
            finv = wdsum/wsum
        else:
            surv = intpsr[n-3:n+4].to_numpy()
            l = len(surv)
            #print(f'surrounding_values for {n}: \n' , surv)
            wdsum = sum(w*v for w, v in zip(wts, surv))
            if l < 7:
                    wsum = sum(wts[:l])
            else:
                wsum = sum(wts)
            #print('\nweight_sum: ',wsum, '\nweighted_sum: ', wdsum)
            finv = wdsum/wsum
        #print('\nfinal value for {n}:', finv)
        finvs.append(finv)

        #print('final values for {v} : ', finvs)
        for i in range(len(finvs)):
                a = nindex[i]
                df[v][a] = finvs[i]

print(df)
path = r'C:/Users/rutur/Desktop/M.Sc Project/for code on big data/1-Exported BigData.xlsx'
df.to_excel(path)
```

5.2.2 Ideology behind this method –

Let us suppose y is the missing value and x is the index of y, and the missing value y is calculated by the formula,

$$y = y_1 + \frac{(x - x_1)(y_2 - y_1)}{x_2 - x_1}$$

Index	Sample with Missing Value
x0	y0
x1	y1
x	-
x2	y2
x3	y3

After applying interpolation, we consider the sample with imputed value y and assign weights for applying WMA method as follow,

Index	Sample with Imputed Values	Weights
x_0	y_0	1/3
x_1	y_1	1/2
x	y	1
x_2	y_2	1/2
x_3	y_3	1/3

Applying WMA on imputed value y with index x and corresponding weights,

$$\text{Final value,}\quad \mathrm{y} = \frac{(y_0 * 0.33) + (y_1 * 0.5) + (y * 1) + (y_2 * 0.5) + (y_3 * 0.33)}{(1/3 + 1/2 + 1 + 1/2 + 1/3)}$$

By using the combination of both methods, it is observed that we get closer to the most likely calculated value.

Sample	With Missing	After Linear Interpolation and WMA
25.37	25.37	25.37
28.16	28.16	28.16
37.73	-->	34.70
41.23	41.23	41.23
38.52	38.52	38.52
33.51	33.51	33.51
34.30	34.30	34.30
34.66	34.66	34.66
33.69	33.69	33.69
28.00	-->	28.12
22.55	22.55	22.55

Furthermore, monthly and annual averages were computed from the cleaned data to facilitate the investigation of seasonal and annual climate patterns. Finally, irrelevant and inconsistent columns were removed from the dataset to maintain focus on the key climate parameters relevant to this study. This preprocess ensured a robust and reliable dataset suitable for subsequent statistical analysis and the extraction of meaningful insights into climate variability in India.

5.3 DATA ANALYSIS –

The processed data was analysed to investigate seasonal and annual climate patterns using a series of statistical techniques and tools. These methods include:

5.3.1 Descriptive Statistics –

To gain an initial understanding of the climate data, descriptive statistics were calculated. This involved computing summary statistics such as the mean, standard deviation, and quartiles for each climate variable. Histograms and boxplots were then generated to visualize the distribution of the data and identify any potential outliers.

Summary Statistics:

Understanding the distribution of key climate parameters is crucial for assessing climate variability in each city, summary statistics were calculated for five important and consistent parameters. These parameters include Max Temperature, Min Temperature, Wind Speed, Rainfall, Evaporation. The following tables presents the summary statistics for each city, offering insights into the central tendency, variability, and range of these climate variables.

Histograms and Boxplots:

Histograms and boxplots are valuable tools for visualizing the distribution of data and identifying key characteristics such as central tendency, spread, and the presence of outliers. This chapter presents histograms and boxplots for the key climate parameters in each of the seven selected cities, providing a visual representation of the patterns observed in the descriptive statistics.

5.3.2 Exploratory Analysis –

To explore the data better Correlation heatmaps are used. It is a powerful tool for visualizing the relationships between multiple variables. This chapter presents correlation heatmaps for each city, depicting the strength and direction of the correlations between pairs of climate parameters. By examining these heatmaps, we can gain insights into how different climate variables interact and influence each other, contributing to a deeper understanding of climate variability patterns.

5.3.3 Time Series Analysis –

Time series decomposition is a valuable technique for separating a time series into its constituent components: trend, seasonality, and residuals. This chapter applies decomposition methods to the climate data for each city, allowing for a detailed examination of long-term trends, seasonal patterns, and any irregular fluctuations that may be present. This analysis provides a comprehensive understanding of the temporal dynamics of the climate variables.

5.3.4 Extreme Value Analysis –

This chapter applies the Generalized Extreme Value distribution to model the extremes of temperature, rainfall, and wind speed in the selected cities. For this chapter **daily records of parameter** are considered to estimate extreme values for specific periods using GEV model as discussed below.

Generalised Extreme Value Distribution –

The Generalized Extreme Value distribution is a widely used statistical model for analyzing extreme values in a dataset. By the extreme value theorem, the GEV distribution is the only possible limit distribution of properly normalized maxima of a sequence of independent and identically distributed random variables.

The theoretical Cumulative Distribution Function, TCDF value for the Generalized Extreme Value distribution for plotting probability distribution of above values is computed using the formula:

$$F_T(x) = \begin{cases} \exp\left(-\left[1 + \varepsilon\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\varepsilon}}\right), & if\ \varepsilon \neq 0, 1 + \varepsilon\left(\frac{x-\mu}{\sigma}\right) > 0 \\ \exp\left(-exp\left(-\frac{x-\mu}{\sigma}\right)\right), & if\ \varepsilon = 0 \end{cases}$$

Where,

- x: The value for which we calculate the theoretical CDF.

- μ: The location parameter.
- σ: The scale parameter (σ > 0).
- ε: The shape parameter.

K-S Test for Goodness of Fit:

The Kolmogorov-Smirnov (K-S) test is used to evaluate the goodness-of-fit of the GEV model, ensuring that it accurately represents the observed extreme values. The likelihood function for a GEV distribution is the product of the probability density functions (PDFs) of the observed data points which will help us to choose optimal values of GEV parameters to compare with actual data and it is calculated by,

$$L(\xi, \mu, \sigma) = \prod_{i=1}^{n} f(x_i; \xi, \mu, \sigma)$$

Hypothesis Testing:

Our claim for this test will to check the goodness of fit of the GEV distribution so we can take claim as the null hypothesis and complement as the alternative hypothesis at 5% level of significance.

Thus, we will test following assumptions,

H_0 : GEV distribution may provide a good fit for the data.

I.e. $F(x) = F_T(x)$ for all values of x

H_1 : GEV distribution may not provide a good fit for the data.

I.e. $F(x) \neq F_T(x)$ for at least one value of x

For testing the null hypothesis, the test statistic is given by,

$$D = \max | F_n(x) - F_T(x) |$$

Where, $F_n(x)$ is the empirical cumulative distribution function (ECDF) bases on observed sample and it is given by,

$$F_n(x) = \frac{Number\ of\ data\ points\ \leq x}{Total\ data\ points}$$

For obtaining the empirical cumulative distribution function $F_n(x)$, first we must obtain cumulative frequencies and then by the definition of theoretical cumulative distribution function we obtain $F_T(x)$.

And to find the critical value for our sample at a given significance level (α),

$$D_\alpha = \frac{c(\alpha)}{\sqrt{n}}$$

Where, c(α) is a constant depending on α and its value when α=0.05 is 1.36.

GEV Model Prediction on Return Levels:

Return levels refer to the estimated values of a variable like temperature, rainfall, or wind speed that are expected to be equaled or exceeded on average once every certain number of years. This "certain number of years" is the **return period**. Analyzing trends in return levels over time can provide insights into how the frequency and magnitude of extreme events are changing due to climate change.

A table with magnitude of extreme event for the period of 10 years, 50 years and 100 years created for every city in the study using genextreme-Scipy module considering the parameters temperature, rainfall and wind speed.

City	Temperature (^oC)	Rainfall (mm)	Wind Speed (Km/H)
ABC	(X, Y, Z)	(X, Y, Z)	(X, Y, Z)

Where,

ABC: The name of City

X: The magnitude of the event associated with a return period of 10 years that is expected to be equaled or exceeded on average once every 10 years.

Y: The magnitude of the event associated with a return period of 50 years that is expected to be equaled or exceeded on average once every 50 years.

Z: The magnitude of the event associated with a return period of 100 years that is expected to be equaled or exceeded on average once every 100 years.

5.4 MULTIVARIATE COMPARISON TEST –

In this chapter, parameters from selected sample data are considered for the multivariate comparison test which is conducted using ANOVA, analysis of variance followed by Post hoc test.

5.4.1 Anova Method

Analysis of variance (ANOVA) is a collection of statistical models and their associated estimation procedures (such as the "variation" among and between groups) used to analyze the differences between groups.

Source of variation	Sum of Squares	*df*	*MS*	*f*	P-value	*f.cri.*
among or between "groups"	SSG	k-1	SSG/DFG	MSG/MSE	Tail area above *f*	Value of *f* for α
within groups or "error"	SSE	N-k	SSE/DFE	-	-	-
total	SST=SSG+SSE	N-1	-	-	-	-

It is assumed that mean of all the groups are equal hence there arise two hypotheses as follows,

Null hypothesis, (H_0): Mean of all the groups is equal,

$$\text{i.e. } \mu_1 = \mu_2 = \mu_3$$

Alternative hypothesis, (H_1): At least one of the groups has different mean

$$\text{i.e. } \mu_i \neq \mu_j \text{ for at least one i, j}$$

The level of significance (α) for the test is 5% i.e. α = 0.05.

5.4.2 Post Hoc Test

A post-hoc analysis refers to a statistical analysis specified after a study has been concluded and the data collected. When null hypothesis is rejected that is at least one of the groups has different mean then a post-hoc test is done to identify exactly which groups differ from each other.

Here, Tukey's Honest Significant Difference (HSD) test is used as a post-hoc test that compares the means of all treatments to the mean of every other treatment. The test statistic is given by,

$$q = \frac{\bar{x}_{max} - \bar{x}_{min}}{s.e.} \text{ where } s.e. = \sqrt{MS_W/n}$$

where n = the size of each of the group samples. The statistic q has a distribution called the studentized range q (see Studentized Range Distribution). The critical values for this distribution are presented in the Studentized Range q Table based on the values of α, k (the number of groups), and df_W. If $q > q_{crit}$ then the two means are significantly different.

Further the HSD is calculated by the formula,

$$HSD = q\sqrt{\frac{MSw}{n}}$$

5.5 TOOLS AND SOFTWARE –

- **Data Processing:** Advanced Excel and Python programming was extensively used for data cleaning, transformation, and analysis.
- **Statistical Analysis:** Excel and Python with statistical packages
- **Data Visualization:** Libraries like Matplotlib, Seaborn (Python).

5.6 LIMITATIONS –

- The study is limited by the extent of missing data in the original IMD datasets.
- The analysis primarily focuses on temporal variability and may not fully capture spatial variations in climate across India.

5.7 CONCLUSION OF METHODOLOGY

This methodology enabled a structured approach to data handling and analysis, ensuring the reliability of results and facilitating a comprehensive understanding of climate variability patterns across temporal scales.

Certificate of Undertaking Requested by IMD

Enclosure-IV (A)

- *To be made and duly signed on the letter head of the Institute / Company / Party receiving IMD data.*
- *The certificate is valid for all data procurements made during next three years from the date of enrolment or till the end of the validity of the enrolment, whichever is earlier.*

CERTIFICATE OF UNDERTAKING

(To be signed by the party receiving meteorological data through Data Supply Portal)

I / We certify that :

1. The data supplied by India Meteorological Department will be used only for the purpose for which it has been asked for.
2. The data are meant exclusively for our own use and shall not be passed or transmitted on to any other party or agency (Indian or foreign) / media (Internet / electronic), either in part or in full. If so needed, prior approval in writing will be taken from India Meteorological Department for the same.
3. The data shall not be used for commercial purpose or to earn consultancy fees, honoraria etc.(applicable for parties other than commercial parties only).
4. Due acknowledgement shall be given to India Meteorological Department for the source of data in all reports / publications etc. brought out by us.

I / We agree to abide by all the terms and conditions as laid down by IMD with regard to procurement of Meteorological Data.

Signature :

Place : Maduravoyal

Date : 07/11/2024

Name : Mr. Ruturaj Tanaji Saravane

Organisation : Dr. M.G.R. Educational and Research Institute

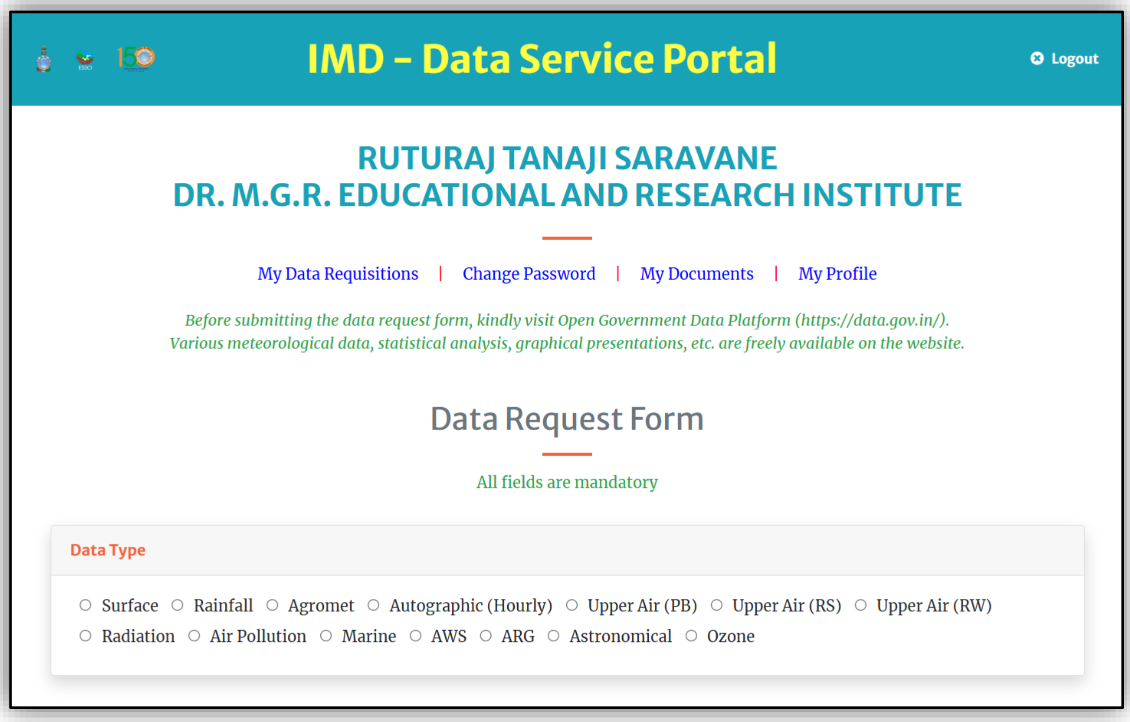
8.11.24

Head of the Office/Department/Competent Authority

(Stamp/Seal of Requesting Party)

ADDL. REGISTRAR
Dr. M.G.R.
EDUCATIONAL AND RESEARCH INSTITUTE
Deemed to Be University
Periyar E.V.R. Road,
Maduravoyal, Chennai-600 095.

Registration on Data Supply Portal



IMD – Data Service Portal

Logout

RUTURAJ TANAJI SARAVANE
DR. M.G.R. EDUCATIONAL AND RESEARCH INSTITUTE

My Data Requisitions | Change Password | My Documents | My Profile

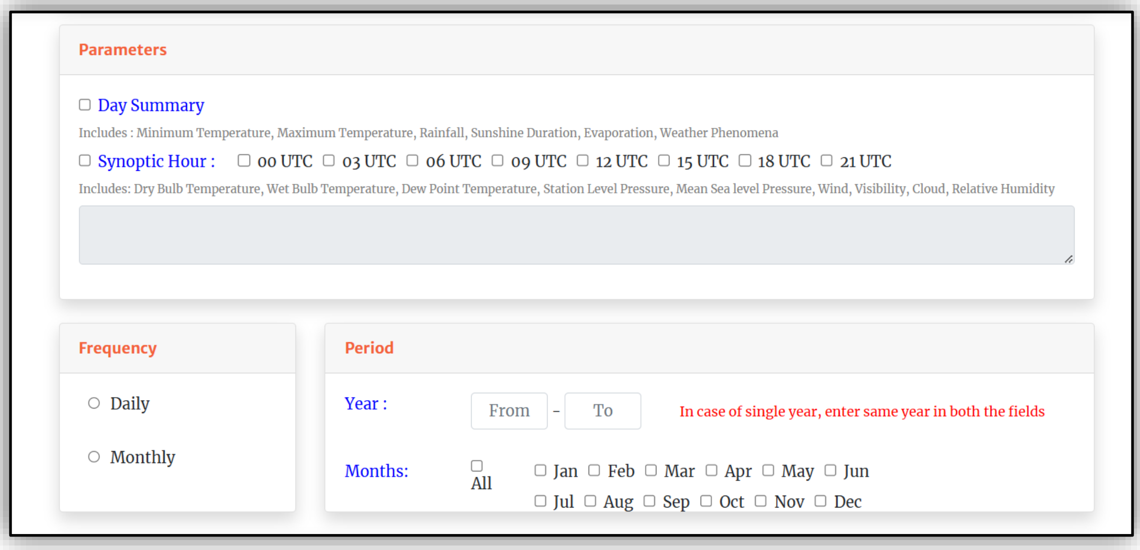
Before submitting the data request form, kindly visit Open Government Data Platform (https://data.gov.in/).
Various meteorological data, statistical analysis, graphical presentations, etc. are freely available on the website.

Data Request Form

All fields are mandatory

Data Type

○ Surface ○ Rainfall ○ Agromet ○ Autographic (Hourly) ○ Upper Air (PB) ○ Upper Air (RS) ○ Upper Air (RW)
○ Radiation ○ Air Pollution ○ Marine ○ AWS ○ ARG ○ Astronomical ○ Ozone



Parameters

☐ Day Summary

Includes : Minimum Temperature, Maximum Temperature, Rainfall, Sunshine Duration, Evaporation, Weather Phenomena

☐ Synoptic Hour : ☐ 00 UTC ☐ 03 UTC ☐ 06 UTC ☐ 09 UTC ☐ 12 UTC ☐ 15 UTC ☐ 18 UTC ☐ 21 UTC

Includes: Dry Bulb Temperature, Wet Bulb Temperature, Dew Point Temperature, Station Level Pressure, Mean Sea level Pressure, Wind, Visibility, Cloud, Relative Humidity

Frequency

○ Daily

○ Monthly

Period

Year : From - To In case of single year, enter same year in both the fields

Months: ☐ All ☐ Jan ☐ Feb ☐ Mar ☐ Apr ☐ May ☐ Jun ☐ Jul ☐ Aug ☐ Sep ☐ Oct ☐ Nov ☐ Dec

Original Dataset Formats

```
                    DAILY SURFACE MARINE DATA (IN IMMT1 FORMAT)
---------------------------------------------------------------------------------------------------------------------------
....Date..... ..Position.. I       .Wind. ..Air Temp. Press.   WW NCCC      II PP ..HH ...Swell.. IEER            I     T      P     D V ...Swell.. CSBDZ
YEAR MN DT HR Q .LAT .LONG 1 h VV N dd fff ..DBT ..DPT PPPPPP ww12 hLMH ..SST 23 ww ..ww d1 p1 ..h1 4nnn S O CALLSGN CO 5 .RAIN r ..WBT t pppp s s d2 p2 ..h2 iiiii
---------------------------------------------------------------------------------------------------------------------------
1982 02 16 18 1 22.6  60.0 0 9 98 0 05  06  21.8  19.8 1015.0 0100 0000  24.0 10       00 00  .0     1 1 YTYP   HR 1          7  .5 1 3
1982 02 25 00 1  7.4  75.4 0 9 98 0 00  00  24.6  19.8 1009.5 0200 0000  29.5 10 00   .0 00 00  .0     1 1 YTGH   HR 1          4  .0 2 4
1982 02 25 12 1  6.2  78.9 0 4 99 3 35  08  26.4  22.6 1008.4 0210 3200  30.1 10 03  1.0 01 07  1.0    1 1 YTGH   HR 1          7  1.0 2 4
1982 03 01 00 1  7.0  76.7 0 4 98 4 00  00  29.2  25.8 1010.0 0211 3280  31.0 00 00   .0 00 00  .0     1 1   YTQV HR 2          7  1.0 6 4
1982 03 02 06 1  9.3  66.4 0 9 99 0 00  00  29.2  22.0 1007.0 0200      29.5 10 00  .0 00 00  .0    1 1 YTQV   51 1         7  3.0 6 4
1982 03 12 18 1  5.9  60.2 0 5 98 3 09  10  28.0  23.0 1019.0 0200 3200  29.0 10 05  1.0 07 06  1.0    1 1 YTUW   HR 1          2  1.0 3 3
1982 03 13 00 1  5.1  61.4 0 5 98 1 09  12  27.4  22.0 1017.0 0200 1200  29.0 10 05  1.0 07      .5   1 1 YTUW   HR 1          7  2.0 3 3
1982 03 14 00 1 20.5  60.8 0 7 98 2 11  11  24.6  17.0 1011.9 0200 2010  27.0 10 03  1.0 00 00  .0     1 1 YTEE   51 1          7  3.4 1 3
1982 03 14 06 1 21.2  61.8 0 9 98 0 05  11  24.2  19.0 1014.0 0200 0000  27.0 10 03  1.0 00 00  .0     1 1 YTEE   51 1          2  1.0 1 3
1982 03 14 12 1 21.8  62.7 0 9 99 0 05  11  25.0  17.0 1011.7 0200 0000  27.0 10 03  1.0 00 00  .0     1 1 YTEE   51 1          7  2.3 1 3
1982 03 14 18 1 22.7  63.8 0 9 98 0 05  11  24.6  17.0 1013.7 0200 0000  27.0 10 03  1.0 00 00  .0     1 1 YTEE   51 1          2  .7 1 3
1982 03 15 00 1 23.4  64.8 0 9 98 0 00  00  24.2  18.0 1011.3 0200 0000  27.0 10 00   .0 00 00  .0     1 1 YTEE   51 1          7  .3 1 3
1982 03 15 06 1 24.2  65.9 0 9 98 0 27  08  23.8  20.0 1012.7 0200 0000  24.0 10 03  1.0 00 00  .0     1 1 YTEE   51 1          2  .7 1 3
1982 03 15 18 1 24.7  64.9 0 9 97 0 25  09  23.0  20.0 1007.7 0200 0000  24.0 10 04  1.5 00 00  .0     1 1 YTEE   51 1          4  .0 1 3
1982 03 19 00 1 24.7  63.5 0 9 96 0 00  00  23.0  19.0 1007.9 0200 0000  24.0 10 00   .0 00 00  .0     1 1 YTEE   51 1          2  .2 6 3
1982 03 19 06 1 24.6  62.3 0 9 98 0 18  05  24.8  21.0 1010.3 0200 0000  24.0 10 02   .5 00 00  .0     1 1 YTEE   51 1          2  1.3 6 3
1982 03 19 12 1 10.1  61.2 0 5 99 2 07  08  28.0  18.6 1007.8 0200 2100  29.0 10 03   .5 09 05  1.5    1 1 YTUG   HR 1          7  3.0 2 4
1982 03 19 18 1  9.5  62.8 0 4 97 2 14  06  26.0  20.0 1013.7 0200 2140  28.0 10 02   .5 00 00  .0     1 1 YTJJ   51 1          7  2.0 2 3
1982 03 19 18 1  9.8  62.8 0 5 99 4 07  08  27.0  18.6 1010.0 0311 4400  30.0 10 03   .5 09 05  1.5    1 1 YTUG   HR 1          2  2.2 2 4
1982 03 20 00 1  9.6  63.7 0 4 98 3 00  00  26.0  19.0 1011.0 0200 2240  29.0 10 00   .0 00 00  .0     1 1 YTJJ   51 1          7  2.7 2 3
1982 03 20 00 1  9.6  64.6 0 5 99 2 07  07  26.8  20.4 1008.6 0111 2200  30.0 10 03   .5 09 05  1.5    1 1 YTUG   HR 1          7  .8 2 4
1982 03 20 06 1  7.2  78.7 0 7 98 3 04  06  28.0  23.0 1010.4 0300 2019  30.0 10 02   .5 00 00  .0     1 1 YTQV   51 1          4  .0 6 4
1982 03 20 06 1  9.1  66.2 0 5 99 1 07  09  30.0  21.0 1010.6 0200 1000  30.0 10 03   .5 09 06  2.0    1 1 YTUG   HR 1          2  2.0 2 4
1982 03 20 06 1  9.3  63.4 0 9 99 0 09  06  29.0  22.0 1014.7 0100 0000  29.0 10 01   .5 00 00  .0     1 1 YTJJ   51 1          2  2.0 2 3
1982 03 20 12 1  7.3  77.5 0 9 99 0 00  00  28.0  24.0 1006.0 0200      30.0 10 00  .0 00 00  .0    1 1 YTQV   51 1         7  4.4 6 4
1982 03 20 12 1  8.9  67.2 0 4 99 3 36  05  30.0  20.0 1009.0 0100 3200  29.0 10 02   .5 00 00  .0     1 1 YTJJ   51 1          7  5.7 2 3
1982 03 20 12 1  8.9  67.9 0 5 98 3 01  09  28.6  22.0 1007.4 0200 3100  31.0 10 04  1.0 05 06  2.0    1 1 YTUG   HR 1          7  1.5 2 4
1982 03 20 18 1  7.7  74.7 0 9 99 0 00  00  28.0  23.0 1010.4 0200      30.0 10 00  .0 00 00  .0    1 1 YTQV   51 1         2  4.4 6 4
1982 03 20 18 1  8.6  68.7 0 5 98 1 14  10  28.0  22.0 1013.7 0100 1040  29.0 10 02   .5 00 00  .0     1 1 YTJJ   51 1          2  3.0 2 3
```

INDEX	YEAR	MN	DT	MAX	MIN	AW	RF	EVP	DRNRF(hrs	DRNRF(mr	SSH	TS	N	FFF
42027	1980	1	1	2.7	0.1	0	11.6	0.3						
42027	1980	1	2	2.2	0	0	7.8	0.3						
42027	1980	1	3	2.4	-1.2	0	8.8							
42027	1980	1	4	5.4	-1.4	0	0	0.4						
42027	1980	1	5	7.8	-2.9	0	0	0.4						
42027	1980	1	6	6.8	-2.7	0	0							
42027	1980	1	7	5.4	-4.1	0	0							
42027	1980	1	8	6	-1.9	0	0							
42027	1980	1	9	5.7	-3.5		0							
42027	1980	1	10	9.5	-4.5	0	0	0.3						
42027	1980	1	11			0	0							
42027	1980	1	12	9.5	-2.8	0	0							
42027	1980	1	14											
42027	1980	1	15	8.3	-2.6	0	0							
42027	1980	1	16	2.5	0.2	0	1.1							
42027	1980	1	17	5.3	1.1	0	0	0.3						
42027	1980	1	18	9.5	2.4		0	0.4						
42027	1980	1	19	7.6	2	0	0	0.1						
42027	1980	1	20	10.4	-0.5	0	0							
42027	1980	1	21	6.3	1.7	0	0	0.5						
42027	1980	1	22	9.7	0.9	0	0	0.5						
42027	1980	1	23	10.4	0.8	0	0	0.5						
42027	1980	1	24	11.5	-0.2	0	0	0.5						
42027	1980	1	25	7.4	1.2	0	4							
42027	1980	1	26	4	1	0	1.3							
42027	1980	1	27	6.6	1.5	0	6.2							

```
---------------------------------------------------------------------------------------------------------------------------
             .AGL. ..................................DATA  FOR  LEVELS  IN  KM  ABOVE  MEAN  SEA  LEVEL........................................ ...ADDITIONAL LEVELS...
           ..GL.  0.15  0.3  0.6  0.9  1.5  2.1  3.1  3.6  4.5  5.8  6.0  7.6  9.5  10.6  12.3  14.1  16.6  18.5  20.5  24.0  26.5  31.0  33.0  37.0  11.2  13.0  15.0  17.0
INDEX YEAR MN DT HR DDDFF DDDFF DDDFF DDDFF DDDFF DDDFF DDDFF DDDFF DDDFF DDDFF DDDFF DDDFF DDDFF DDDFF DDDFF DDDFF DDDFF DDDFF DDDF
DDDFF DDDFF DDDFF DDDFF DDDFF DDDFF DDDFF DDDFF DDDFF DDDFF
---------------------------------------------------------------------------------------------------------------------------
42591 1980 01 01 00 00000 11206 11205 15603 24606 27507
42591 1980 01 02 00 09003 09006 09806 11306 20303 24006
42591 1980 01 03 00 00000
42591 1980 01 04 00 00000
42591 1980 01 05 00 00000 08102 07802 04503 01503 26510 23507
42591 1980 01 06 00 00000 26502 34202 31802 30802 31703 32505 32506
42591 1980 01 07 00 00000 28005 29105 29205 29306 30207 30008
42591 1980 01 08 00 00000 28803 32702 31002 30005 31808 30208 30511
42591 1980 01 09 00 00000 26605 26605 28604 28608 30508 30109
42591 1980 01 10 00 00000 29804 29804 30804 31207 31608 28810
42591 1980 01 11 00 00000 26605 28505 29507 30308 30613 29813
42591 1980 01 12 00 00000 12004 09803 04802 24001 27508 29512 27510
42591 1980 01 13 00 00000 35502 35501 33803 28505 30807
42591 1980 01 14 00 00000 29203 29203 31304 31403 30503
42591 1980 01 15 00 00000 34201 34201 30002 26203 27209 28011
42591 1980 01 16 00 00000 11804 13004 07001 28002 15502 15003 15305
42591 1980 01 17 00 00000 07603 09403 13801 26302 29209 30409
42591 1980 01 18 00 00000 10704 10504 04401 29301 28008 29609
42591 1980 01 19 00 00000 12103 11303 02002 30203 30409 30811
42591 1980 01 20 00 18001 24404 26804 30007 31007 31210 28811 29011
42591 1980 01 21 00 00000 27305 27305 32006 29606 29307 28506
42591 1980 01 22 00 00000 30202 31503 32504 29005 28508 29213 30011
42591 1980 01 23 00 27003 30909 30909 31708 29806 29005 25610
42591 1980 01 24 00 22502 26007 28007 30508 30807 30309 30016
42591 1980 01 25 00 00000 18801 18802 32701 28403 30807 30509
42591 1980 01 26 00 00000 21502 21502 33501 28503 27807 29008 29813
42591 1980 01 27 00 13501 15304 14803 17801 31404 28308
```

Datasets after Pre-Processing, Cleaning and Transformation –

Months	Average of MAX	Average of MIN	Average of AW	Average of RF	Average of EVP	Average of DRNRF(hrs)	Average of DRNRF(mnts)	Average of SSH
Jan-80	21.10	7.28	7.94	0.20	1.73	0.21	4.26	7.56
Feb-80	25.37	11.03	10.28	0.57	3.21	0.46	5.19	8.55
Mar-80	28.16	14.99	9.19	1.51	4.05	1.08	5.71	6.95
Apr-80	37.73	23.05	10.50	0.06	8.01	0.08	5.61	9.35
May-80	41.23	27.76	10.48	0.65	9.23	0.23	5.46	9.28
Jun-80	38.52	28.19	8.07	5.33	7.72	0.81	11.45	8.46
Jul-80	33.51	26.03	6.10	15.72	3.90	3.40	26.32	4.95
Aug-80	34.30	26.47	5.94	4.99	5.46	1.18	16.10	7.71
Sep-80	34.66	24.71	7.03	4.26	4.97	0.90	7.46	7.94
Oct-80	33.69	20.33	5.19	0.54	4.39	0.23	6.45	8.53
Nov-80	28.00	13.05	4.30	0.09	2.46	0.12	3.49	8.28
Dec-80	22.55	9.08	6.69	0.39	1.83	0.77	4.55	7.00
Jan-81	20.51	8.53	6.70	1.73	1.53	0.84	11.19	6.46
Feb-81	24.35	11.28	9.54	0.94	3.08	1.19	5.62	8.00
Mar-81	27.68	15.26	8.55	0.76	3.59	1.30	15.15	7.44
Apr-81	36.43	22.07	10.80	0.12	7.95	0.05	1.17	9.26
May-81	39.19	26.34	8.90	0.29	8.72	0.19	13.32	9.34
Jun-81	40.90	28.24	7.60	10.36	8.67	1.17	10.36	7.64
Jul-81	33.23	26.54	4.26	8.40	4.23	2.68	28.24	5.46
Aug-81	34.18	26.70	10.45	2.86	5.00	0.79	15.37	7.10
Sep-81	35.83	25.57	8.63	4.59	5.56	0.79	15.63	7.69
Oct-81	33.52	18.91	6.81	0.63	4.39	0.09	0.74	9.16
Nov-81	26.25	13.12	7.03	1.63	2.04	1.24	9.24	7.72
Dec-81	22.78	7.83	6.39	0.14	1.84	0.14	2.71	7.65
Jan-82	20.65	8.38	7.94	1.24	1.52	1.34	9.39	6.38
Feb-82	21.38	10.08	10.39	0.50	2.10	1.27	21.91	5.28

DATE	LAT.	LONG.	CL. HEIGHT	CL. AMT.	WIND DIR.	WIND SPD	DBT	WBT	AIR PRESS.	LOW. CL. AMT.	SST
16-02-1982	22.6	60	9	0	5	6	21.8	19.8	1015	0	24
25-02-1982	7.4	75.4	9	0	0	0	24.6	19.8	1009.5	0	29.5
25-02-1982	6.2	78.9	4	3	35	8	26.4	22.6	1008.4	3	30.1
01-03-1982	7	76.7	4	4	0	0	29.2	25.8	1010	3	31
02-03-1982	9.3	66.4	9	0	0	0	29.2	22	1007		29.5
12-03-1982	5.9	60.2	5	3	9	10	28	23	1019	3	29
13-03-1982	5.1	61.4	5	1	9	12	27.4	22	1017	1	29
14-03-1982	22.7	63.8	9	0	5	11	24.6	17	1013.7	0	27
14-03-1982	21.8	62.7	9	0	5	11	25	17	1011.7	0	27
14-03-1982	21.2	61.8	9	0	5	11	24.2	19	1014	0	27
14-03-1982	20.5	60.8	7	2	11	11	24.6	17	1011.9	2	27
15-03-1982	24.7	64.9	9	0	25	9	23	20	1007.7	0	24
15-03-1982	24.2	65.9	9	0	27	8	23.8	20	1012.7	0	24
15-03-1982	23.4	64.8	9	0	0	0	24.2	18	1011.3	0	27
19-03-1982	24.7	63.5	9	0	0	0	23	19	1007.9	0	24
19-03-1982	24.6	62.3	9	0	18	5	24.8	21	1010.3	0	24
19-03-1982	10.1	61.2	5	2	7	8	28	18.6	1007.8	2	29
19-03-1982	9.8	62.8	5	4	7	8	27	18.6	1010	4	30
19-03-1982	9.5	62.8	4	2	14	6	26	20	1013.7	2	28
20-03-1982	9.6	63.7	4	3	0	0	26	19	1011	2	29
20-03-1982	9.6	64.6	5	2	7	7	26.8	20.4	1008.6	2	30
20-03-1982	9.3	63.4	9	0	9	6	29	22	1014.7	0	29
20-03-1982	9.1	66.2	5	1	7	9	30	21	1010.6	1	30
20-03-1982	8.9	67.2	4	3	36	5	30	20	1009	3	29
20-03-1982	8.9	67.9	5	3	1	9	28.6	22	1007.4	3	31

Chapter 6: ANALYSIS, RESULTS AND INTERPRETATION

6.1 DESCRIPTIVE STATISTICS –

Table: Summary Stats of New Delhi

Region	Parameter	Mean	std.	Min	Max	25%	50%	75%
New Delhi	Max Temp.	31.45	6.27	17.35	42.16	26.19	33.37	36
	Min Temp.	19.01	7.46	5.34	30.63	11.99	20.29	26.25
	Wind Speed	6.95	2.24	1.31	15.83	5.26	6.95	8.46
	Rainfall	3.04	4.11	0	26.95	0.21	1.27	4.28
	Evaporation	4.18	2	0	12.61	2.79	3.88	5.03

Key Observations –

- New Delhi experiences a wide range of temperatures, with a mean maximum temperature of 31.45°C and a mean minimum temperature of 19.01°C. The standard deviation for both maximum and minimum temperatures is relatively high, indicating significant temperature variability throughout the year.
- The average wind speed in New Delhi is 6.95 with moderate variability.
- Rainfall in New Delhi is highly variable, with a mean of 3.04 and a large standard deviation (4.11). This suggests that rainfall can range significantly from year to year, with some years experiencing very little rainfall and others experiencing heavy rainfall events.
- Evaporation also shows some variability, with a mean of 4.18.

Table: Summary Stats of Jodhpur

Region	Parameter	Mean	std.	Min	Max	25%	50%	75%
Jodhpur	Max Temp.	33.71	5.79	0	43.09	29.97	34.5	37.6
	Min Temp.	19.99	7.02	0	30.27	13.28	21.69	26.3
	Wind Speed	2.92	2.76	0	12.8	0	2.78	4.71
	Rainfall	1.64	2.84	0	19.11	0	0.25	1.93
	Evaporation	0	0.04	0	0.73	0	0	0

Key Observations –

- Jodhpur experiences very high temperatures, with a mean maximum temperature of 33.71°C. The maximum recorded temperature is very high at 43.09°C. This is consistent with its hot desert climate.

- While hot, Jodhpur also experiences a wide range of temperatures, as indicated by the relatively high standard deviation for both maximum and minimum temperatures (5.79 and 7.02 respectively). This suggests significant daily and seasonal temperature fluctuations.
- Jodhpur generally has low wind speeds, with a mean of 2.92.
- Rainfall in Jodhpur is very low, with a mean of 1.64 and a maximum recorded rainfall of 19.11. The low rainfall and high temperatures are characteristic of its arid desert environment.
- Evaporation is also very low, likely due to the limited availability of water.

Table: Summary Stats of Guwahati

Region	Parameter	Mean	std.	Min	Max	25%	50%	75%
Guwahati	Max Temp.	29.91	3.21	21.61	35.93	27.54	30.87	32.3
	Min Temp.	19.71	5.4	8.74	26.81	14.88	20.89	24.97
	Wind Speed	2.26	1.01	0.74	5.93	1.5	2.02	2.84
	Rainfall	5.94	5.46	0	27.59	0.9	5.03	9.67
	Evaporation	3.55	1.26	0	9.73	2.73	3.46	4.32

Key Observations –

- Guwahati experiences a moderate range of temperatures, with a mean maximum temperature of 29.91°C and a mean minimum temperature of 19.71°C. This is consistent with its humid subtropical climate.
- The standard deviation for both maximum and minimum temperatures is relatively moderate, suggesting a less extreme range of temperatures compared to cities like New Delhi or Jodhpur.
- Wind speeds in Guwahati are generally low, with a mean of 2.26.
- Guwahati receives high rainfall, with a mean of 5.94 and a maximum recorded rainfall of 27.59. This is expected given its location in a region influenced by the monsoon.
- Evaporation levels are moderate, likely influenced by the high humidity and rainfall.

Table: Summary Stats of Bhopal

Region	Parameter	Mean	std.	Min	Max	25%	50%	75%
Bhopal	Max Temp.	31.85	4.86	21.53	42.72	28.47	30.98	34.66
	Min Temp.	19.08	5.4	8.18	28.15	13.81	20.38	23.4
	Wind Speed	7.71	4.6	0	24.17	4.36	6.85	9.81
	Rainfall	3.96	5.99	0	33.95	0.04	0.81	6.16
	Evaporation	0.01	0.07	0	1.33	0	0	0

Key Observations –

- Bhopal experiences high temperatures overall, with a mean maximum temperature of 31.85°C and a high maximum recorded temperature of 42.72°C.
- The standard deviation for maximum temperature is moderate (4.86), suggesting a decent amount of fluctuation but not as extreme as some other cities in study.
- Bhopal has moderate wind speeds on average (mean of 7.71).
- Rainfall in Bhopal shows considerable variability (standard deviation of 5.99), with a mean of 3.96. This indicates that rainfall can fluctuate significantly from year to year.
- Evaporation appears to be very low in Bhopal, likely due to factors like relatively lower humidity compared to coastal cities.

Table: Summary Stats of Mumbai

Region	Parameter	Mean	std.	Min	Max	25%	50%	75%
Mumbai	Max Temp.	32.2	1.57	28.4	35.5	31	32.3	33.5
	Min Temp.	22.82	3.35	14.2	28.9	19.98	24.1	25.3
	Wind Speed	6.75	3.15	0	17.3	4.7	6.3	8.6
	Rainfall	7.33	11.71	0	57.7	0	0.2	11.43
	Evaporation	0	0.03	0	0.7	0	0	0

Key Observations –

- Mumbai experiences moderate temperatures with a mean maximum temperature of 32.2°C. The standard deviation for both maximum and minimum temperatures is relatively low, suggesting less temperature variability compared to inland cities. This is likely due to the moderating influence of the coastal location.
- Wind speeds in Mumbai are moderate, with a mean of 6.75.

- Mumbai experiences high rainfall, with a mean of 7.33. However, the high standard deviation (11.71) indicates significant variability in rainfall amounts. This is expected given the influence of the monsoon season, which can bring heavy rainfall but also periods of dry weather.
- Evaporation appears to be very low in Mumbai. This might seem counterintuitive given the coastal location, but it could be due to factors like high humidity, cloud cover, or specific measurement methods used in the dataset.

Table: Summary Stats of Chennai

Region	Parameter	Mean	std.	Min	Max	25%	50%	75%
Chennai	Max Temp.	33.21	2.82	27.43	39.79	30.69	33.62	35.16
	Min Temp.	25.01	2.31	19.77	29.75	23.07	25.38	26.79
	Wind Speed	6.27	1.32	2.81	11.03	5.42	6.22	7.2
	Rainfall	5.28	6.31	0	41.33	0.3	3.27	8.08
	Evaporation	0	0	0	0	0	0	0

Key Observations –

- Chennai experiences high temperatures with a mean maximum temperature of 33.21°C. The standard deviation for maximum temperature is moderate (2.82), suggesting some variability but not as extreme as in desert climates.
- Minimum temperatures are also quite high, with a mean of 25.01°C and relatively low variability. This indicates consistently warm nights throughout the year.
- Wind speeds in Chennai are moderate, with a mean of 6.27.
- Chennai receives a high amount of rainfall on average (mean of 5.28). However, the high standard deviation (6.31) indicates significant variability in rainfall amounts from year to year, likely influenced by the monsoon season.
- Evaporation appears to be very low in Chennai, similar to what was observed in Mumbai. This could be due to factors like high humidity, cloud cover, or specific measurement methods used in the dataset.
-

Table: Summary Stats of Thiruvananthapuram

Region	Parameter	Mean	std.	Min	Max	25%	50%	75%
Thiruvananthapuram	Max Temp.	31.91	1.37	28.8	35.8	30.9	31.8	33
	Min Temp.	23.78	0.99	21.4	27.3	23.2	23.65	24.4
	Wind Speed	6.3	2.67	0	14.7	4.6	5.9	8.1
	Rainfall	6.3	5.21	0	30.9	1.8	5.5	9.3
	Evaporation	3.77	1.59	1.6	19.4	3	3.5	4.1

Key Observations –

- Thiruvananthapuram experiences consistently high temperatures, with a mean maximum temperature of 31.91°C and a relatively low standard deviation (1.37). This suggests less variability in maximum temperatures compared to other cities.
- Minimum temperatures are also warm, with a mean of 23.78°C and low variability. This indicates consistently warm nights throughout the year.
- Wind speeds in Thiruvananthapuram are moderate, with a mean of 6.3.
- Thiruvananthapuram receives high rainfall, with a mean of 6.3. However, the standard deviation (5.21) indicates variability in rainfall amounts, likely influenced by the monsoon season.
- Evaporation levels are moderate (mean of 3.77), which is higher than observed in the other coastal cities (Mumbai and Chennai). This could be due to factors like higher temperatures or differences in local humidity and wind patterns.

Histograms and Boxplots: Visualizing Data Distribution –

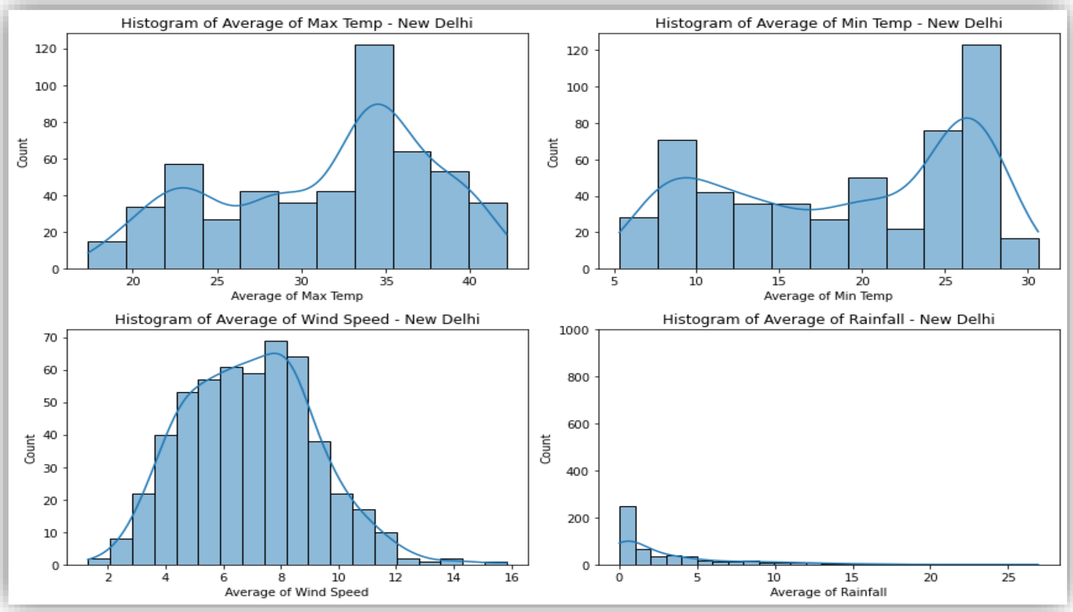


Fig. Histogram of Parameters – New Delhi

Histogram of Average of Max Temp -

- **Shape:** The distribution appears roughly normal, with a slight skew to the right (meaning there are slightly higher average maximum temperatures).

- **Central Tendency and Spread:** The data is concentrated around 30-35°C indicating variability in average maximum temperatures.
- **Outliers:** There might be a few outliers on the higher end (above 40°C), which could represent particularly hot years.

Histogram of Average of Min Temp -

- **Shape:** The distribution also seems roughly normal, with slight skew to the right.
- **Central Tendency and Spread:** Most of the data falls between 15-25°C, indicating the typical range of average minimum temperatures.
- **Outliers:** There might be a few outliers on the lower end (below 10°C), representing unusually cool years.

Histogram of Average of Wind Speed -

- **Shape:** This distribution is skewed to the right, with most of the data clustered around lower wind speeds (2-6).
- **Central Tendency and Spread:** The data is concentrated around lower values, indicating that lower wind speeds are more common in New Delhi.
- **Outliers:** There are some higher wind speed values (above 10), suggesting occasional strong wind events.

Histogram of Average of Rainfall -

- **Shape:** This distribution is heavily skewed to the right, with most of the data clustered around low rainfall values.
- **Central Tendency and Spread:** Most of the data points represent lower rainfall amounts, indicating that New Delhi generally receives less rainfall.
- **Outliers:** There are values with very high rainfall (above 20), suggesting occasional heavy rainfall events or years with unusually high precipitation.

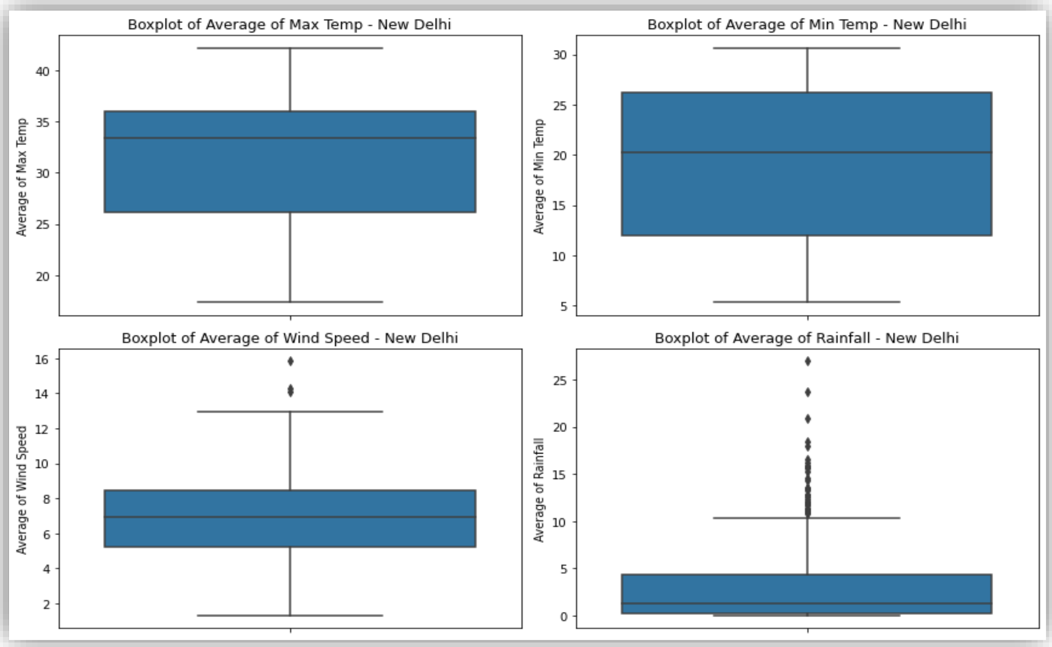


Fig. Boxplots of Parameters – New Delhi

Boxplot of Average of Max Temp -

- This boxplot summarizes New Delhi's average maximum temperatures. The median is ~33°C, with most years in the IQR of 30°C–36°C. Whiskers span 26°C–40°C, capturing typical values, while outliers above 40°C mark exceptionally hot years.

Boxplot of Average of Min Temp -

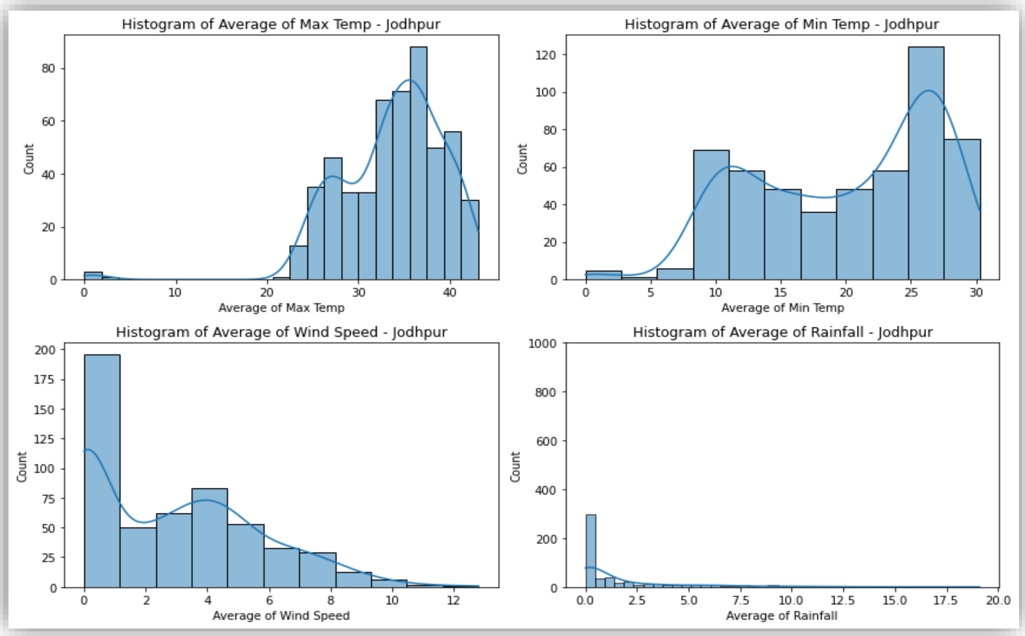
- The median minimum temperature is ~20°C, with most years in the IQR of 15°C–25°C. Whiskers reach ~10°C, indicating cooler years, while outliers below 10°C reflect unusually cold years

Boxplot of Average of Wind Speed -

- The boxplot shows the distribution of average wind speeds. The median is ~7, with most values in the IQR of 5–9. Whiskers span 2–14, and outliers above 14 suggest occasional high wind speeds, likely from storms or specific weather patterns.

Boxplot of Average of Rainfall -

- The median rainfall is ~1, with most values in the IQR of 0–5. The upper whisker reaches ~10, and outliers above this indicate significant variability, likely influenced by the monsoon and extreme rainfall events.



Histograms of Parameters - Jodhpur

Histogram of Average of Max Temp -

- **Shape:** The distribution of average maximum temperatures appears roughly normal and is centered around 34-35°C.
- **Central Tendency and Spread:** The data is concentrated around higher values, which indicates consistently high temperatures with some variability.
- **Outliers:** There are no significant outliers in the data.
- The observation aligns with the high mean max. temperature observed in the summary stats.

Histogram of Average of Min Temp -

- **Shape:** The distribution of average minimum temperatures appears roughly normal and is centered around 21-22°C.
- **Central Tendency and Spread:** The data is concentrated around 21-22°C, which indicates consistently warm nighttime temperatures.
- **Outliers:** There are no significant outliers in the data.

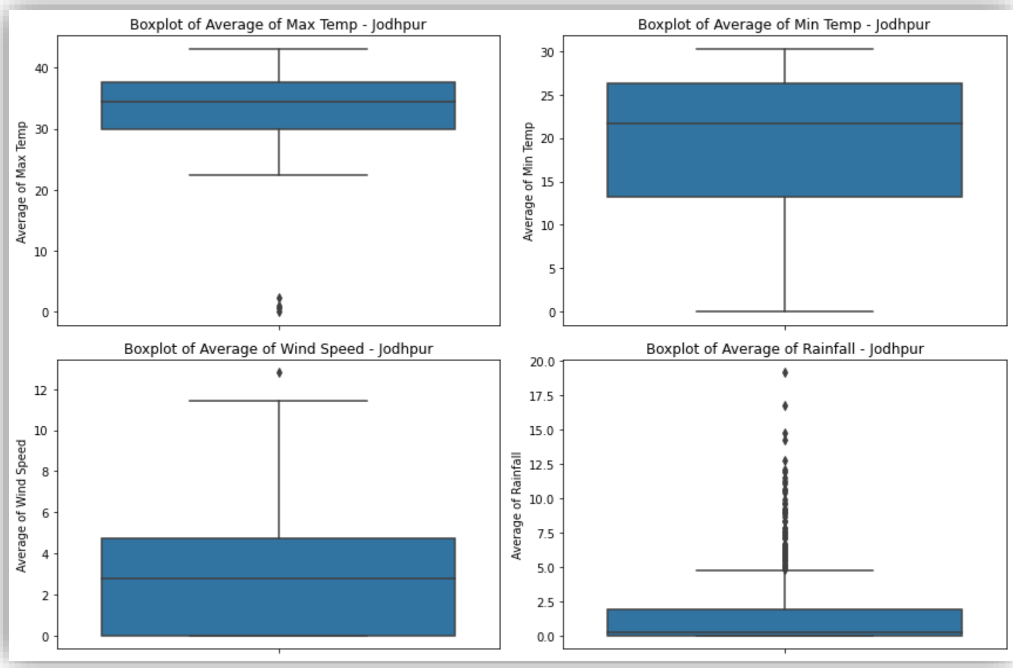
Histogram of Average of Wind Speed -

- **Shape:** The distribution of average wind speeds is right-skewed.
- **Central Tendency and Spread:** Most data points are clustered around lower wind speeds (2-4).

- **Outliers:** There are a few higher values, which suggest occasional stronger winds.
- This observation aligns with the low mean wind speed in the summary statistics.

Histogram of Average of Rainfall -

- **Shape:** The distribution of average rainfall is heavily right-skewed.
- **Central Tendency and Spread:** Most data points represent very low rainfall amounts.
- **Outliers:** There are a few instances of higher rainfall.
- This observation is consistent with the very low mean rainfall and the arid climate of Jodhpur.



Boxplot of Parameters - Jodhpur

Boxplot of Average of Max Temp -

- Median average maximum temperature is around 34°C, with most years falling between 30°C and 38°C. A few exceptionally hot years are observed with average maximum temperatures above 38°C.

Boxplot of Average of Min Temp -

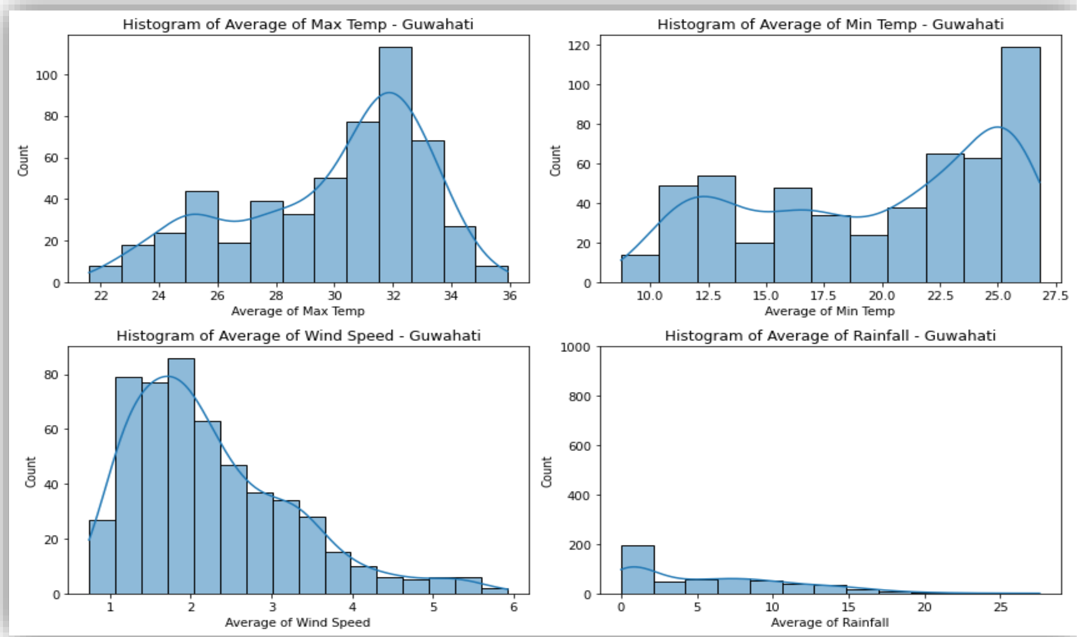
- Median average minimum temperature is around 20°C, with most years between 15°C and 25°C. A few unusually cool years have average minimum temperatures below 15°C.

Boxplot of Average of Wind Speed -

- Median average wind speed is around 3, with most values falling between 1 and 5. Some years experience higher average wind speeds, exceeding 10.

Boxplot of Average of Rainfall -

- Median average rainfall is very low, close to 0. Most years have average rainfall between 0 and 2. A few years show higher rainfall amounts, exceeding 15.



Histogram of Parameters - Guwahati

Histogram of Average of Max Temp -

- **Shape:** The distribution appears roughly normal, with a slight right skew.
- **Central Tendency and Spread:** The data is concentrated around 29-30°C, with a moderate spread indicating some variability in average maximum temperatures.
- **Outliers:** There may be a few outliers on the higher end (above 34°C), representing warmer years.

Histogram of Average of Min Temp -

- **Shape:** The distribution appears roughly normal, with a slight left skew.
- **Central Tendency and Spread:** Most of the data falls between 20-25°C, indicating the typical range of average minimum temperatures.
- **Outliers:** There might be a few outliers on the lower end (below 17.5°C),

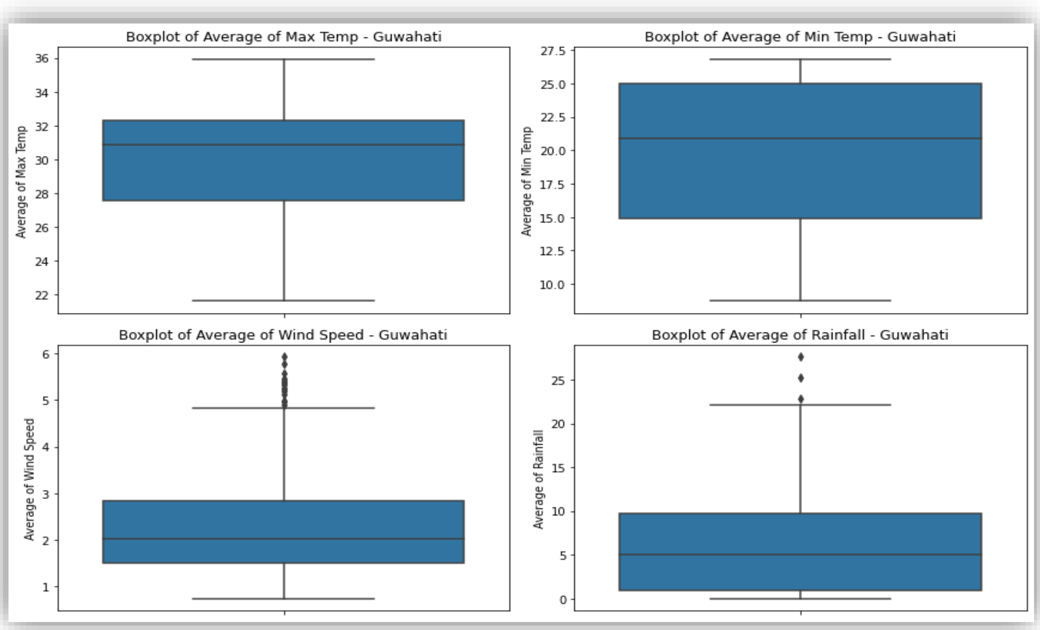
representing unusually cool years.

Histogram of Average of Wind Speed -

- **Shape:** This distribution is skewed to the right, with most of the data clustered around lower wind speeds (1-3).
- **Central Tendency and Spread:** The data is concentrated around lower values, indicating that lower wind speeds are more common in Guwahati.
- **Outliers:** There are some higher wind speed values (above 4), suggesting occasional strong wind events.

Histogram of Average of Rainfall -

- **Shape:** This distribution is heavily skewed to the right, with most of the data clustered around lower rainfall values.
- **Central Tendency and Spread:** The majority of data points represent lower rainfall amounts, but there is a wide range, indicating variability in rainfall patterns.
- **Outliers:** There are a few instances of very high rainfall (above 20), suggesting occasional heavy rainfall events or years with unusually high precipitation.



Boxplot of Parameters - Guwahati

Boxplot of Average of Max Temp -

- The median average maximum temperature in Guwahati is around 30°C, with most years falling within the interquartile range (IQR) between approximately 28°C and

32°C. The whiskers extend to about 26°C and 34°C, capturing the typical range of average maximum temperatures.

Boxplot of Average of Min Temp -

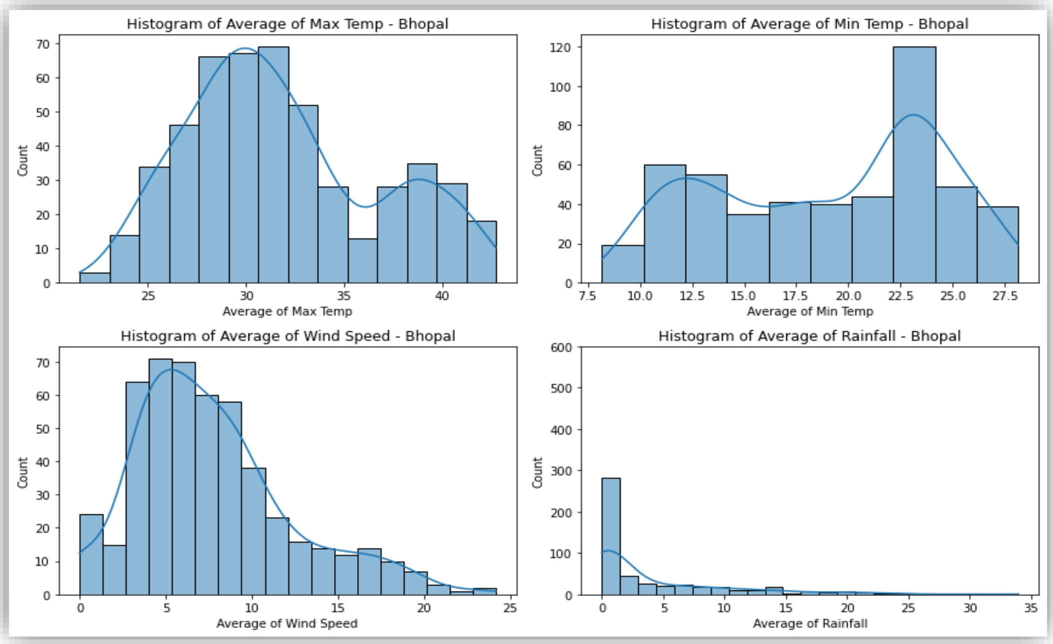
- It shows that most years falling within the IQR of roughly 18°C to 22°C. The whiskers extend down to about 16°C and up to 24°C, showing the typical range of average minimum temperatures.

Boxplot of Average of Wind Speed -

- The most values falling within the IQR between approximately 1 and 3. The upper whisker extends to about 5, while the lower whisker is close to 0. The presence of several outliers above 5 indicates occasional occurrences of high wind speeds, possibly due to storms or specific weather patterns.

Boxplot of Average of Rainfall -

- The median average rainfall is around 10, with most values falling within the IQR between approximately 5 and 15. The upper whisker extends to about 25, and there are a few outliers above this, indicating a moderate degree of variability in rainfall.



Histogram of Parameters – Bhopal

Histogram of Average of Max Temp -

- **Shape:** The distribution of average maximum temperatures in Bhopal appears to follow a roughly normal distribution, centered around 32-33°C.
- **Central Tendency and Spread:** The data is concentrated around the mean, with a moderate spread indicating some variability in average maximum temperatures.
- **Outliers:** There may be a few outliers on the higher end of the distribution (above 40°C), which could represent particularly hot years.
- This observation aligns with the summary statistics, which showed a mean maximum temperature of 31.85°C.

Histogram of Average of Min Temp -

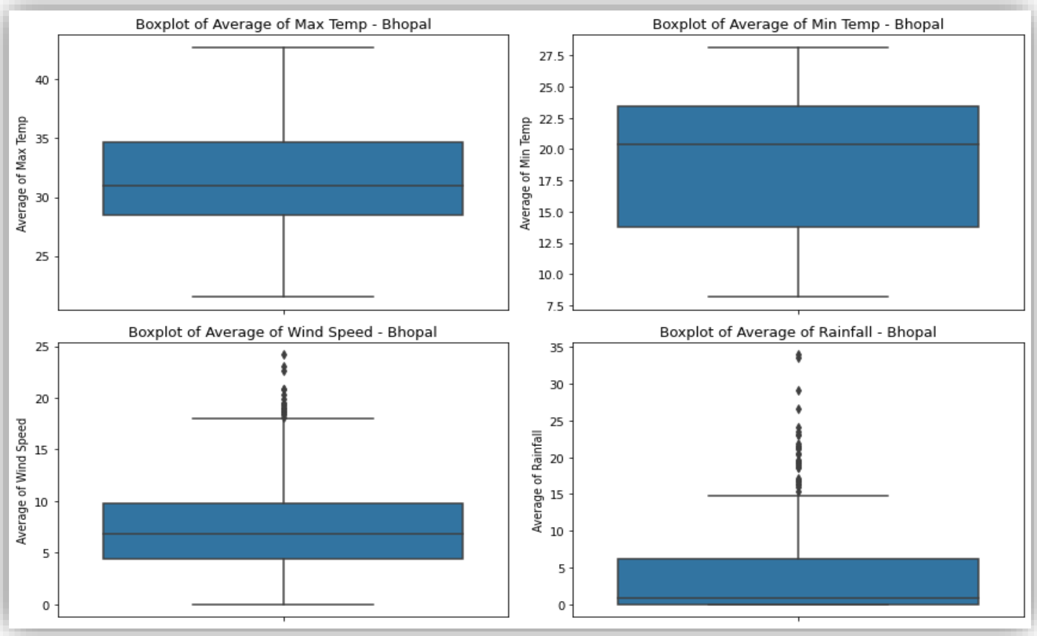
- **Shape:** The distribution of average minimum temperatures in Bhopal appears to follow a roughly normal distribution, centered around 20-21°C.
- **Central Tendency and Spread:** The data is concentrated around the mean, indicating consistency in average minimum temperatures.
- **Outliers:** There may be a few outliers on the lower end of the distribution (below 10°C), which could represent unusually cool years.

Histogram of Average of Wind Speed -

- **Shape:** The distribution of average wind speeds in Bhopal is right-skewed.
- **Central Tendency and Spread:** The data is concentrated around lower wind speeds (4-8), with some variability.
- **Outliers:** Some higher wind speed values (above 15) suggest the occurrence of occasional strong wind events.

Histogram of Average of Rainfall -

- **Shape:** The distribution of average rainfall in Bhopal is heavily right-skewed.
- **Central Tendency and Spread:** Most data points are clustered around lower rainfall values, indicating that Bhopal generally receives lower rainfall.
- **Outliers:** A few instances of very high rainfall (above 25) suggest the occurrence of occasional heavy rainfall events.



Boxplot of Parameters – Bhopal

Boxplot of Average of Max Temp -

- This boxplot shows that the median average maximum temperature in Bhopal is around 31°C, with most years falling within the interquartile range (IQR) between approximately 29°C and 34°C. The whiskers extend to about 27°C and 37°C, capturing the typical range of average maximum temperatures. There are a few outliers above 37°C, indicating some exceptionally warm years.

Boxplot of Average of Min Temp -

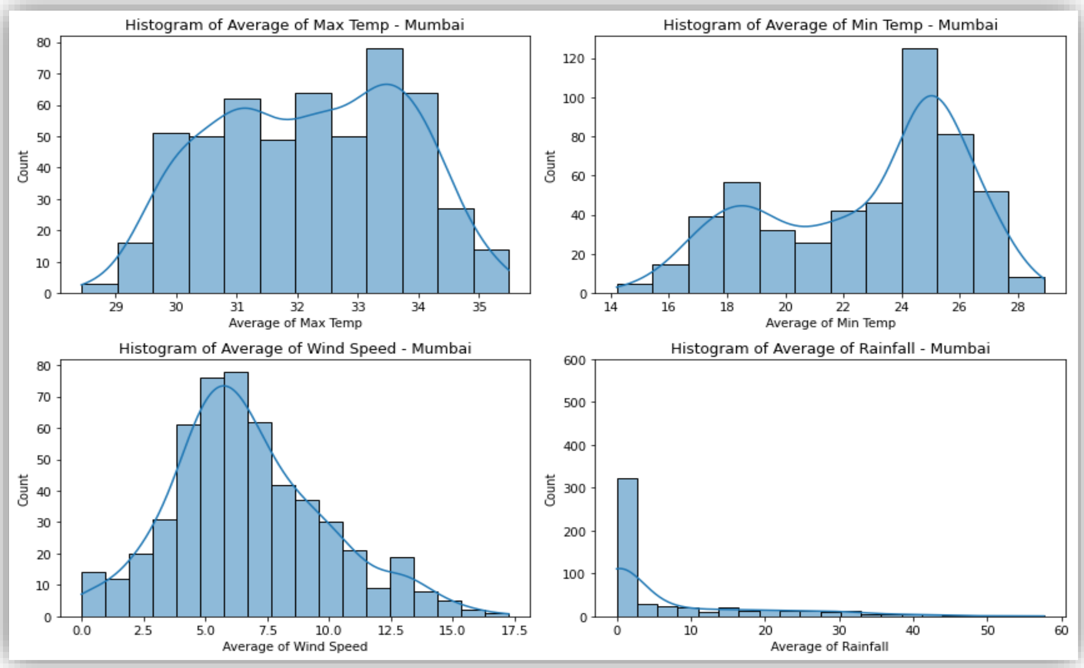
- The most years falling within the IQR of roughly 17°C to 23°C. The whiskers extend down to about 13°C and up to 26°C, showing the typical range of average minimum temperatures. There are a few outliers below 13°C, suggesting some unusually cool years.

Boxplot of Average of Wind Speed -

- The median average wind speed is around 7, with most values falling within the IQR between approximately 4-10. The upper whisker extends to about 20, while the lower whisker is close to 0. The presence of several outliers above 20 indicates occasional occurrences of high wind speeds, possibly due to storms or specific weather patterns.

Boxplot of Average of Rainfall -

- The median average rainfall is around 5, with most values falling within the IQR between approximately 0 and 10. The upper whisker extends to about 20, and there are a few outliers above this, indicating a moderate degree of variability in rainfall and the occurrence of years with significantly higher precipitation. This variability is likely due to the influence of the monsoon season and the potential for extreme rainfall events.



Histogram of Parameters - Mumbai

Histogram of Average of Max Temp -

- **Shape:** The distribution of average maximum temperatures appears roughly normal, with a slight right skew.
- **Central Tendency and Spread:** The data is concentrated around 31-33°C, with a moderate spread indicating some variability in average maximum temperatures.
- **Outliers:** There may be a few outliers on the higher end (above 35°C), representing warmer years.

Histogram of Average of Min Temp -

- **Shape:** The distribution of average minimum temperatures appears roughly normal, with a peak around 24-25°C.
- **Central Tendency and Spread:** Most of the data falls between 20-27°C, indicating the typical range of average minimum temperatures.
- **Outliers:** There might be a few outliers on the lower end (below 18°C), representing

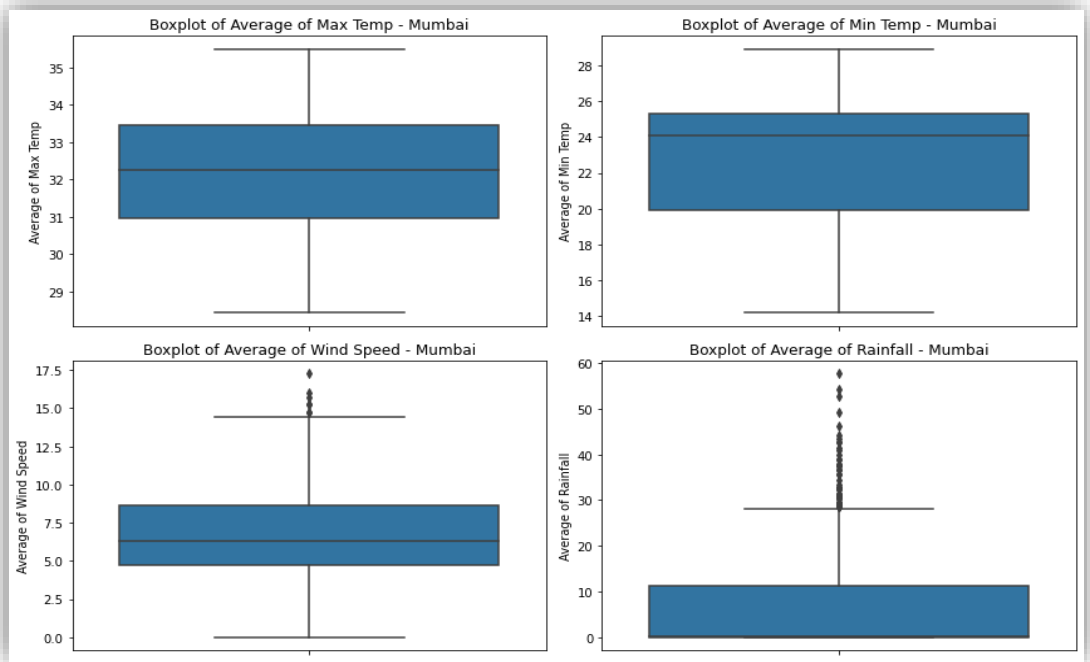
unusually cool years.

Histogram of Average of Wind Speed -

- **Shape:** This distribution is skewed to the right, with most of the data clustered around lower wind speeds (4-8).
- **Central Tendency and Spread:** The data is concentrated around lower values, indicating that lower wind speeds are more common in Mumbai.
- **Outliers:** There are some higher wind speed values (above 12), suggesting occasional strong wind events.

Histogram of Average of Rainfall -

- **Shape:** This distribution is heavily skewed to the right, with most of the data clustered around lower rainfall values.
- **Central Tendency and Spread:** The majority of data points represent lower rainfall amounts, but there is a wide range, indicating variability in rainfall patterns.
- **Outliers:** There are a few instances of very high rainfall (above 40), suggesting occasional heavy rainfall events or years with unusually high precipitation.
- This aligns with the summary statistics, which showed a high mean rainfall (7.33) but also a high standard deviation (11.71).



Boxplot of Parameters - Mumbai

Boxplot of Average of Max Temp -

- This boxplot shows that the median average maximum temperature in Mumbai is around 33°C, with most years falling within the interquartile range (IQR) between approximately 31°C and 34°C. The whiskers extend to about 29°C and 35°C, capturing the typical range of average maximum temperatures.

Boxplot of Average of Min Temp -

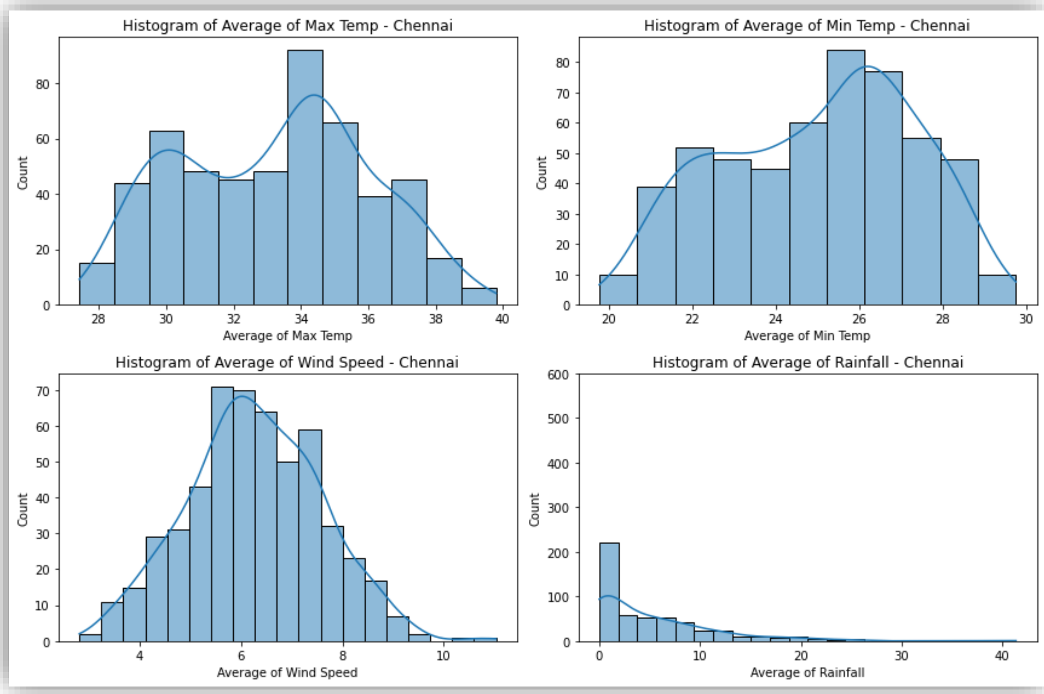
- The median average minimum temperature is around 24°C, with most years falling within the IQR of roughly 22°C to 26°C. The whiskers extend down to about 20°C and up to 28°C, showing the typical range of average minimum temperatures. There are no significant outliers, suggesting a consistent range of minimum temperatures.

Boxplot of Average of Wind Speed -

- The median average wind speed is around 7, with most values falling within the IQR between approximately 5-10. The upper whisker extends to about 15, while the lower whisker is close to 0. The presence of several outliers above 15 indicates occasional occurrences of high wind speeds, possibly due to storms or specific weather patterns.

Boxplot of Average of Rainfall -

- The median average rainfall is around 10, with most values falling within the IQR between approximately 0 and 20. The upper whisker extends to about 40, and there are a few outliers above this, indicating a moderate degree of variability in rainfall and the occurrence of years with significantly higher precipitation. This variability is likely due to the influence of the monsoon season and the potential for extreme rainfall events.



<u>Histogram of Parameters - Chennai</u>

Histogram of Average of Max Temp -

- **Shape:** The distribution of average maximum temperatures in Chennai appears to follow a roughly normal distribution, centered around 33-34°C.
- **Central Tendency and Spread:** The data is concentrated around the mean, with a moderate spread indicating some variability in average maximum temperatures.
- **Outliers:** There may be a few outliers on the higher end of the distribution (above 38°C), which could represent particularly warm years.

Histogram of Average of Min Temp -

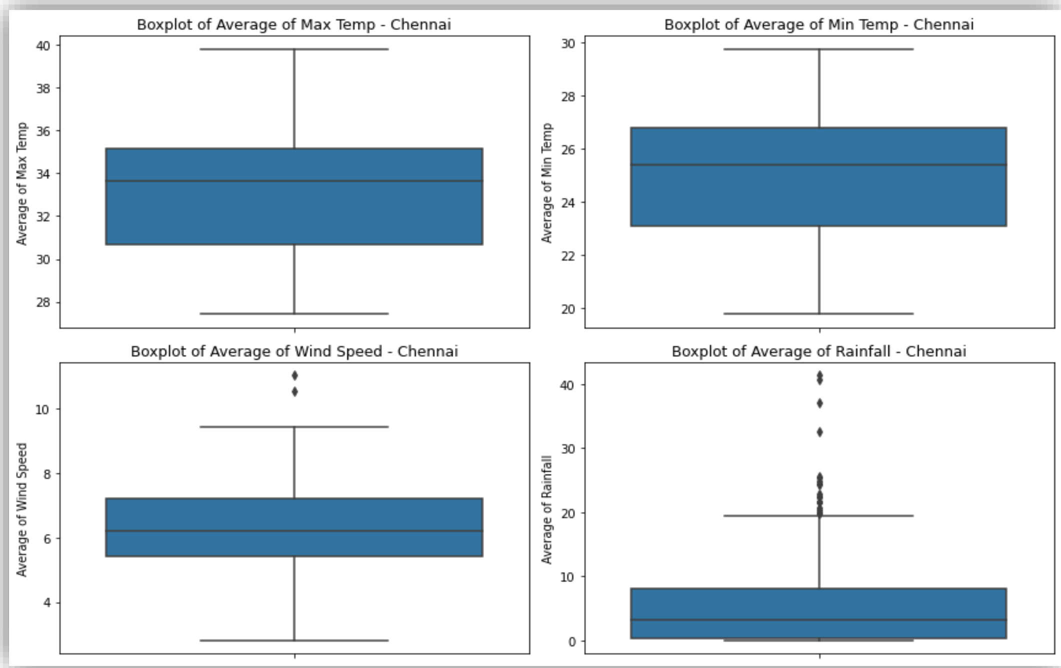
- **Shape:** The distribution of average minimum temperatures in Chennai appears to follow a roughly normal distribution, with a peak around 25-26°C.
- **Central Tendency and Spread:** The data is concentrated around the mean, indicating consistency in average minimum temperatures with less variability compared to maximum temperatures.
- **Outliers:** There may be a few outliers on the lower end of the distribution (below 22°C), which could represent unusually cool years.
- This observation matches the summary statistics, which showed a mean minimum temperature of 25.01°C.

Histogram of Average of Wind Speed -

- **Shape:** The distribution of average wind speeds in Chennai is right-skewed.
- **Central Tendency and Spread:** The data is concentrated around lower wind speeds (5-8), with some variability.
- **Outliers:** Some higher wind speed values (above 10) suggest the occurrence of occasional strong wind events.
- This observation is consistent with the summary statistics, which showed a mean wind speed of 6.27.

Histogram of Average of Rainfall -

- **Shape:** The distribution of average rainfall in Chennai is heavily right-skewed.
- **Central Tendency and Spread:** Most data points are clustered around lower rainfall values, indicating generally lower rainfall with high variability.
- **Outliers:** A few instances of very high rainfall (above 30) suggest the occurrence of occasional heavy rainfall events.
- This observation aligns with the summary statistics, which showed a mean rainfall of 5.28 and a high standard deviation (6.31).



Boxplot of Parameters - Chennai

Boxplot of Average of Max Temp -

- The median average maximum temperature is about 32°C, with most values falling between 31°C and 33°C. The typical range extends from 30°C to 35°C, with no significant outliers, indicating consistent temperatures.

Boxplot of Average of Min Temp -

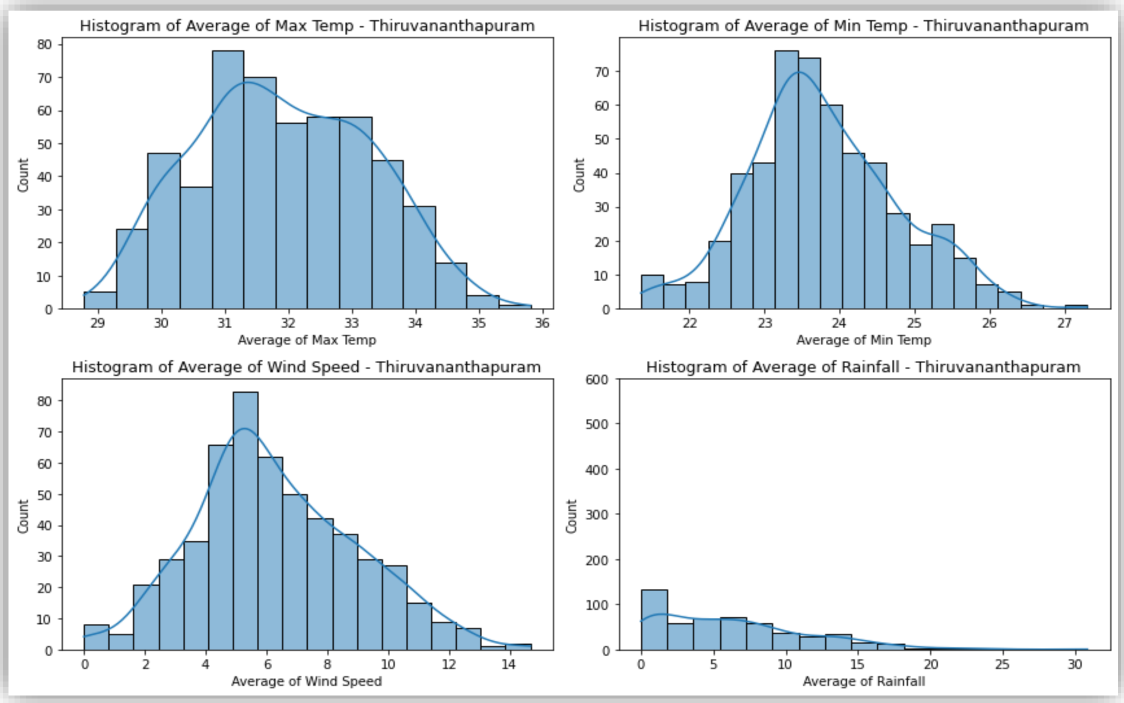
- The most years are falling within the IQR of roughly 23°C to 27°C. The whiskers extend down to about 21°C and up to 29°C, showing the typical range of average minimum temperatures. There are no significant outliers, suggesting a consistent range of minimum temperatures.

Boxplot of Average of Wind Speed -

- The boxplot shows a median average wind speed of roughly 6, with most values between 4 and 8. The typical range extends from about 2 to 12, with a few instances of higher wind speeds likely due to storms or weather patterns.

Boxplot of Average of Rainfall -

- The boxplot shows a median average rainfall of about 10, with most values between 2 and 18. The typical range extends to 35, with a few years experiencing significantly higher rainfall, likely due to monsoons and extreme events.



Histogram of Parameters - Thiruvananthapuram

Histogram of Average of Max Temp -

- **Shape:** The distribution of average maximum temperatures appears roughly normal, centered around 31-32°C.
- **Central Tendency and Spread:** The data is concentrated around the mean, with a relatively small spread, indicating less variability in average maximum temperatures compared to some other cities.
- **Outliers:** There are no significant outliers.

Histogram of Average of Min Temp -

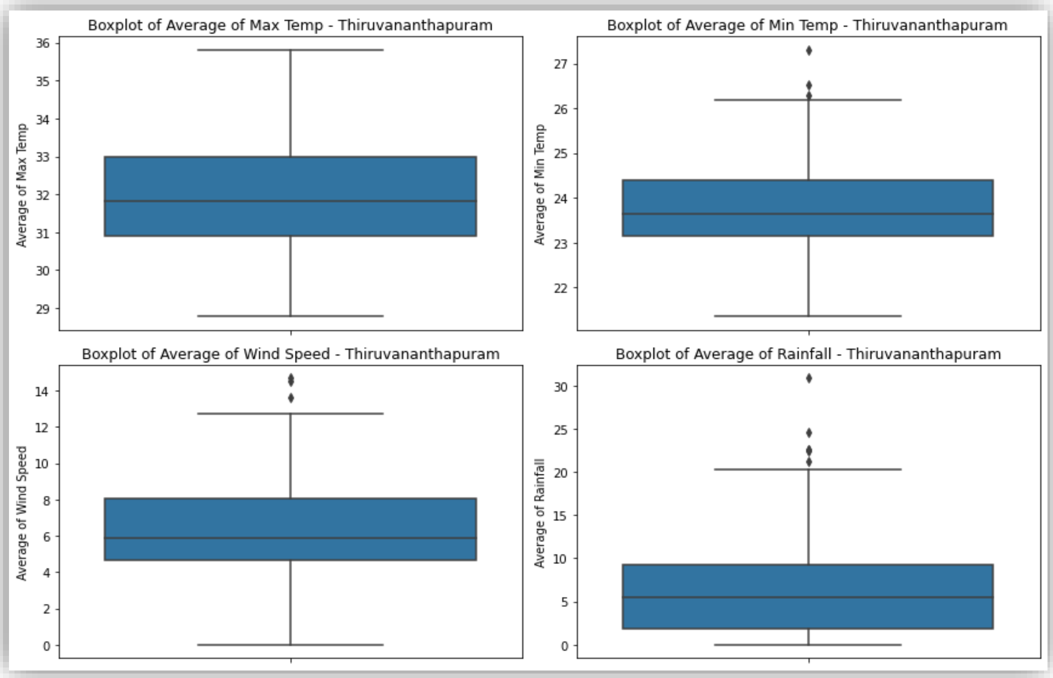
- **Shape:** The distribution of average minimum temperatures appears roughly normal, with a peak around 23-24°C.
- **Central Tendency and Spread:** Most of the data falls between 22-26°C, indicating the typical range of average minimum temperatures.
- **Outliers:** There are no significant outliers.

Histogram of Average of Wind Speed -

- **Shape:** This distribution is skewed to the right, with most of the data clustered around lower wind speeds (4-7).
- **Central Tendency and Spread:** The data is concentrated around lower values, indicating that lower wind speeds are more common in Thiruvananthapuram.
- **Outliers:** There are some higher wind speed values (above 10), suggesting occasional strong wind events.

Histogram of Average of Rainfall -

- **Shape:** This distribution is right-skewed, with most of the data clustered around lower rainfall values.
- **Central Tendency and Spread:** The majority of data points represent lower rainfall amounts, but there is a wide range, indicating variability in rainfall patterns.
- **Outliers:** There are a few instances of higher rainfall (above 20), suggesting occasional heavy rainfall events or years with unusually high precipitation.



<u>Boxplot of Parameters - Thiruvananthapuram</u>

Boxplot of Average of Max Temp -

- This boxplot shows that the median average maximum temperature in Thiruvananthapuram is around 32°C, with most years falling within the interquartile range (IQR) between approximately 31°C and 33°C. The whiskers extend to about 30°C and 35°C, capturing the typical range of average maximum temperatures. There are no significant outliers, indicating consistent maximum temperatures.

Boxplot of Average of Min Temp -

- The median average minimum temperature is around 24°C, with most years falling within the IQR of roughly 23°C to 25°C. The whiskers extend down to about 22°C and up to 27°C, showing the typical range of average minimum temperatures. There are a few outliers above 27°C, suggesting some unusually warm nights.

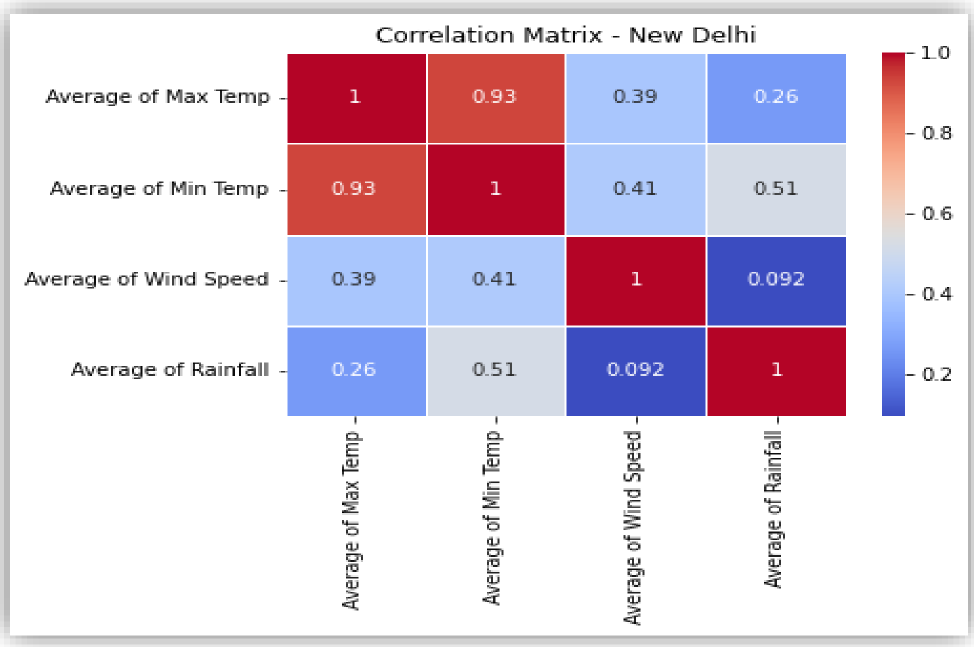
Boxplot of Average of Wind Speed -

- The median average wind speed is around 6, with most values falling within the IQR between approximately 4 and 8. The upper whisker extends to about 14, while the lower whisker is close to 0. The presence of several outliers above 14 indicates occasional occurrences of high wind speeds, possibly due to storms or specific weather patterns.

Boxplot of Average of Rainfall -

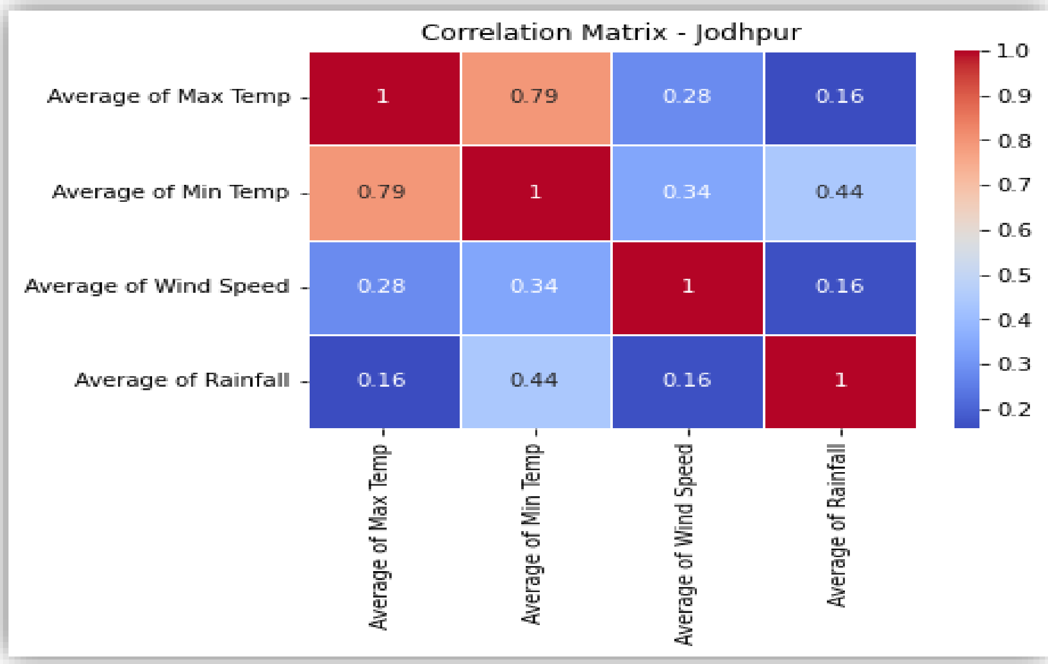
- The median average rainfall is around 10, with most values falling within the IQR between approximately 5 and 15. The upper whisker extends to about 25, and there are a few outliers above this, indicating a moderate degree of variability in rainfall and the occurrence of years with significantly higher precipitation. This variability is likely due to the influence of the monsoon season and the potential for extreme rainfall events.

6.2 EXPLORATORY ANALYSIS: CORRELATION HEATMAPS



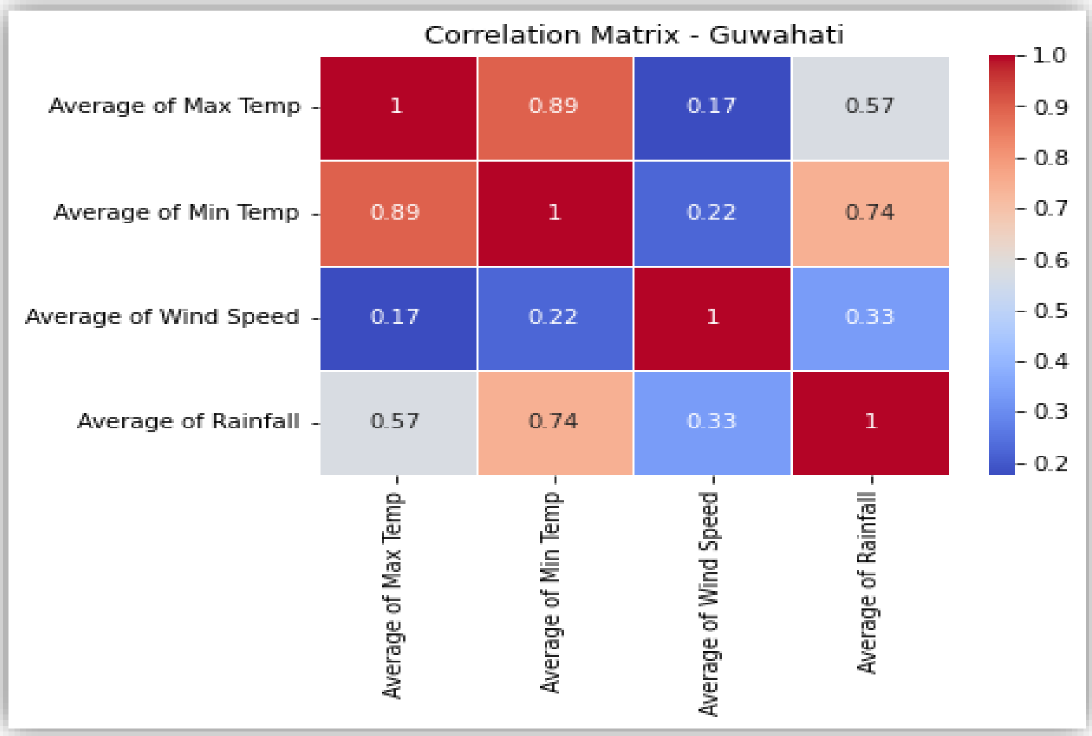
Correlation Matrix – New Delhi

- The average maximum and minimum temperatures in New Delhi are closely related, with a high correlation of 0.93.
- There is a moderate positive relationship between the minimum temperature and rainfall (0.51), suggesting that higher min temperatures are slightly associated with increased rainfall.
- The average max temperature has a weak positive correlation with rainfall (0.26), implying that changes in max temperature have minimal influence on rainfall.
- Wind speed shows weak positive correlations with max temperature (0.39) and min temperature (0.41), hinting at some relationship, possibly due to seasonal or climatic patterns, though not very strong.



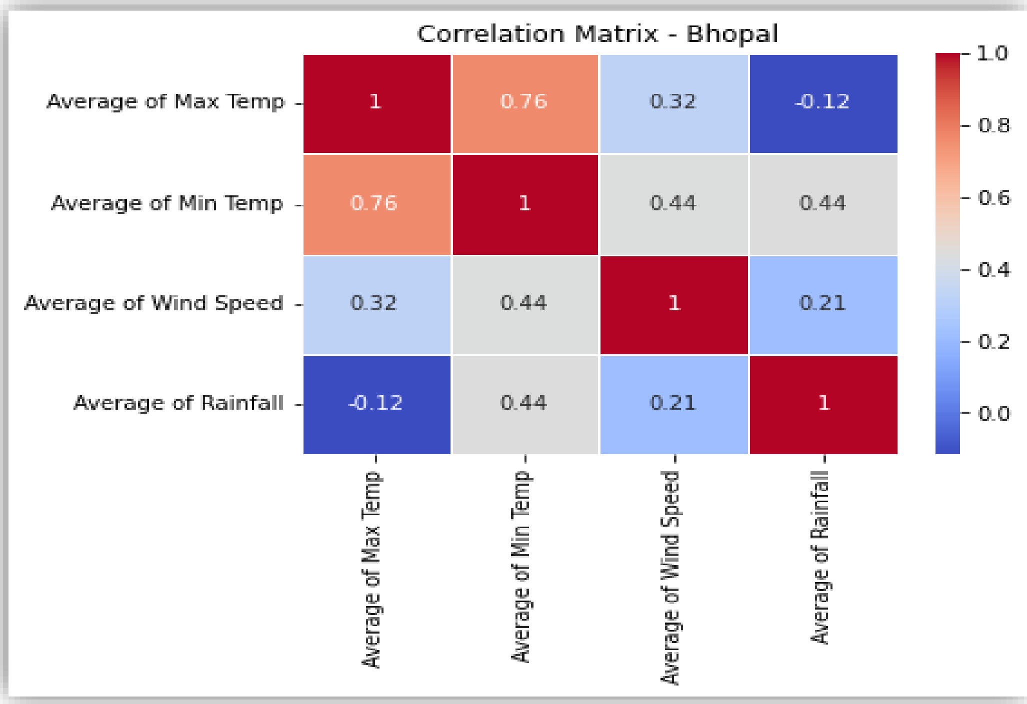
Correlation Matrix - Jodhpur

- The average maximum and minimum temperatures in Jodhpur have a strong positive correlation of 0.79.
- The correlation between maximum temperature and rainfall is weak (0.16), showing minimal influence of maximum temperature on rainfall.
- Wind speed has a weak positive correlation with both minimum temperature (0.34) and maximum temperature (0.28), indicating a slight relationship between these variables.
- The negligible correlation between wind speed and rainfall (0.16) suggests that these variables are mostly independent of each other.



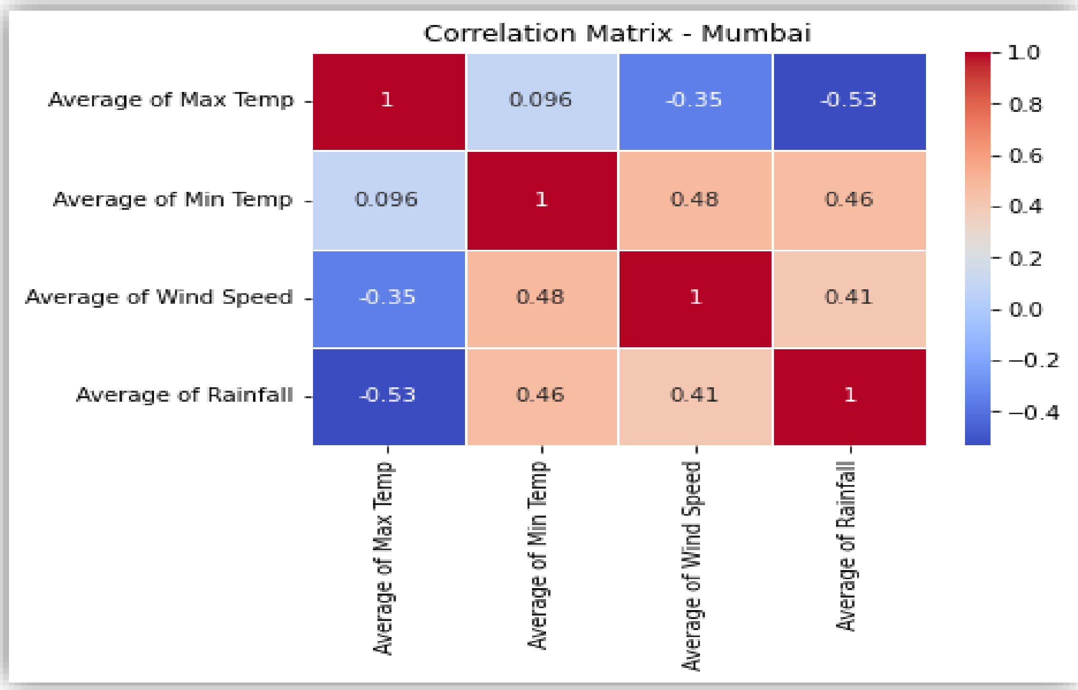
Correlation Matrix - Guwahati

- The average maximum and minimum temperatures in Guwahati show a strong positive correlation of 0.89, indicating that they tend to rise and fall together.
- There is a strong positive correlation (0.74) between minimum temperature and rainfall, suggesting that higher minimum temperatures are associated with increased rainfall.
- The average maximum temperature and rainfall have a moderate positive correlation (0.57), showing some influence of maximum temperature on rainfall.
- Wind speed has weak positive correlations with maximum temperature (0.17), minimum temperature (0.22), and rainfall (0.33), indicating only slight relationships with these variables.
- The overall relationships suggest that temperature variables are more closely tied to rainfall, while wind speed shows relatively lower correlations with other factors.



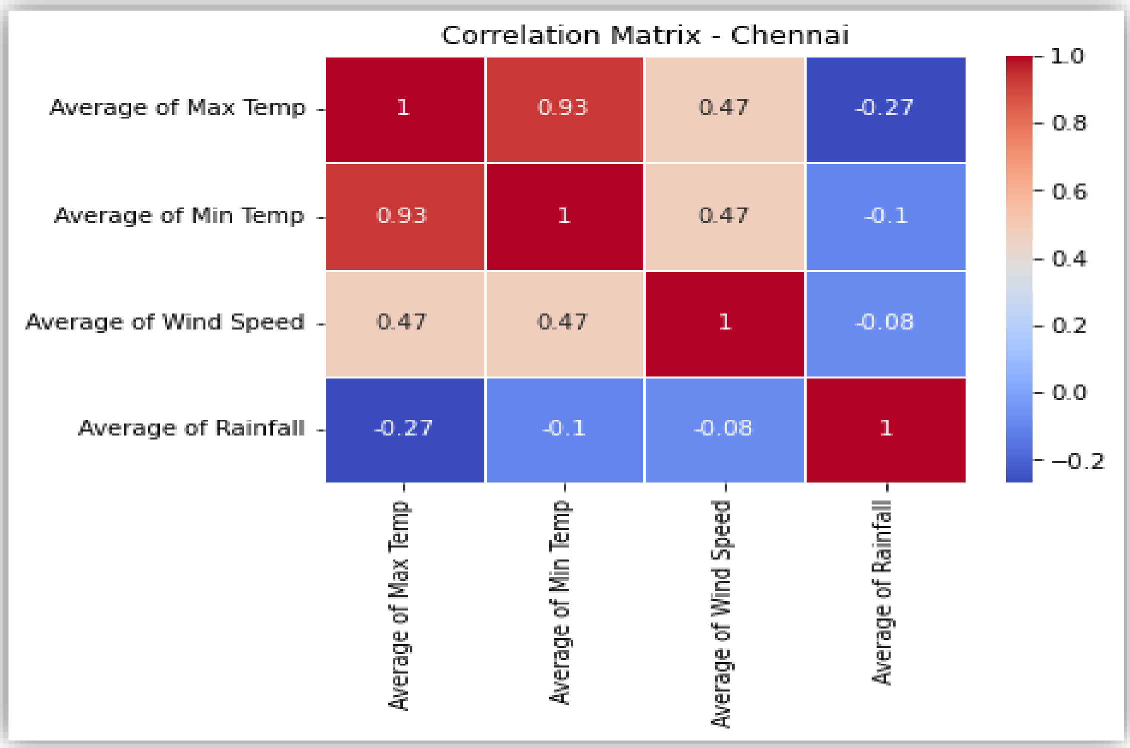
Correlation Matrix - Bhopal

- The average maximum and minimum temperatures in Bhopal have a strong positive correlation of 0.76, indicating that they generally rise or fall together.
- There is a moderate positive correlation (0.44) between minimum temperature and rainfall, suggesting that higher minimum temperatures are somewhat associated with increased rainfall.
- The correlation between maximum temperature and rainfall is weak (-0.12), showing that maximum temperature has minimal and slightly inverse influence on rainfall.
- Wind speed has a moderate positive correlation with minimum temperature (0.44) and a weak positive correlation with maximum temperature (0.32), indicating some relationship between wind speed and temperature.
- The correlation between wind speed and rainfall is weak (0.21), suggesting a slight relationship between these variables.



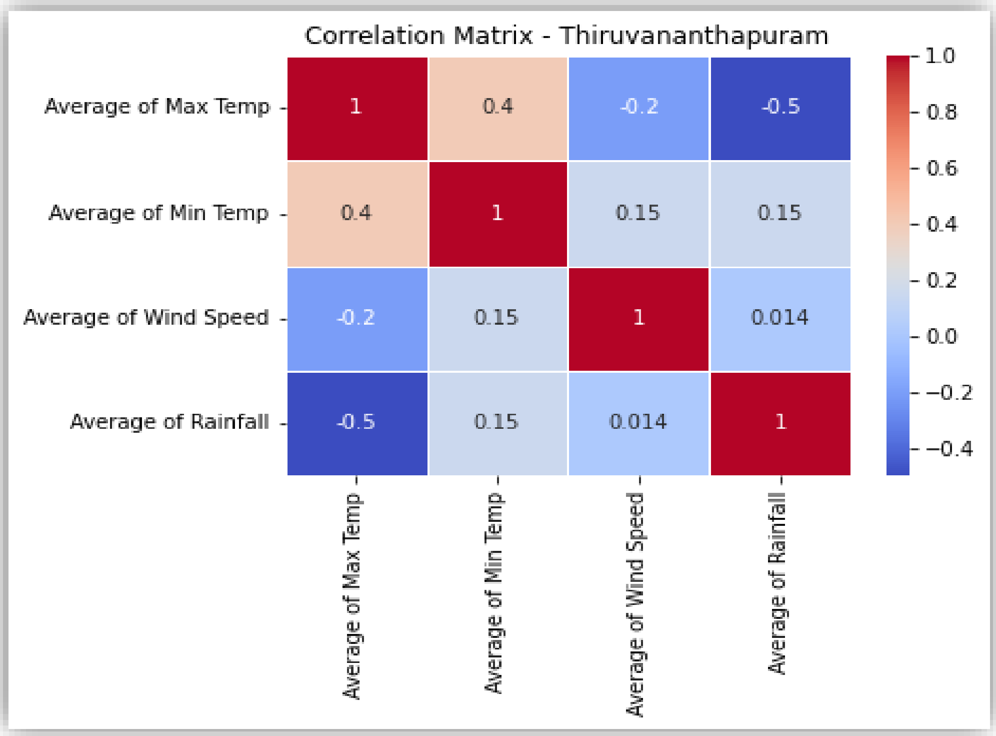
Correlation Matrix -Mumbai

- The average maximum and minimum temperatures in Mumbai have a weak positive correlation of 0.096, indicating minimal interdependence between these variables.
- There is a moderate positive correlation (0.46) between minimum temperature and rainfall, suggesting that higher minimum temperatures are somewhat associated with increased rainfall.
- The correlation between maximum temperature and rainfall is moderate and negative (-0.53), indicating that higher maximum temperatures are associated with lower rainfall.
- Wind speed has a moderate positive correlation with minimum temperature (0.48) and a weak negative correlation with maximum temperature (-0.35), showing some relationship with minimum temperatures but an inverse one with maximum temperatures.
- Wind speed and rainfall have a moderate positive correlation (0.41), suggesting that higher wind speeds are somewhat linked with increased rainfall.



Correlation Matrix - Chennai

- The average maximum and minimum temperatures in Chennai show a strong positive correlation of **0.93**, indicating that they tend to rise and fall together.
- The average maximum temperature and rainfall have a weak negative correlation (**-0.27**), indicating that higher temperatures are slightly associated with reduced rainfall.
- The average minimum temperature and wind speed also exhibit a moderate positive correlation (**0.47**), like the maximum temperature's relationship with wind speed.
- Wind speed and rainfall have a negligible negative correlation (**-0.08**), suggesting little to no connection between these two factors.
- Temperature variables (both maximum and minimum) show stronger relationships with each other and with wind speed, while rainfall shows weaker or negative correlations with the other factors.

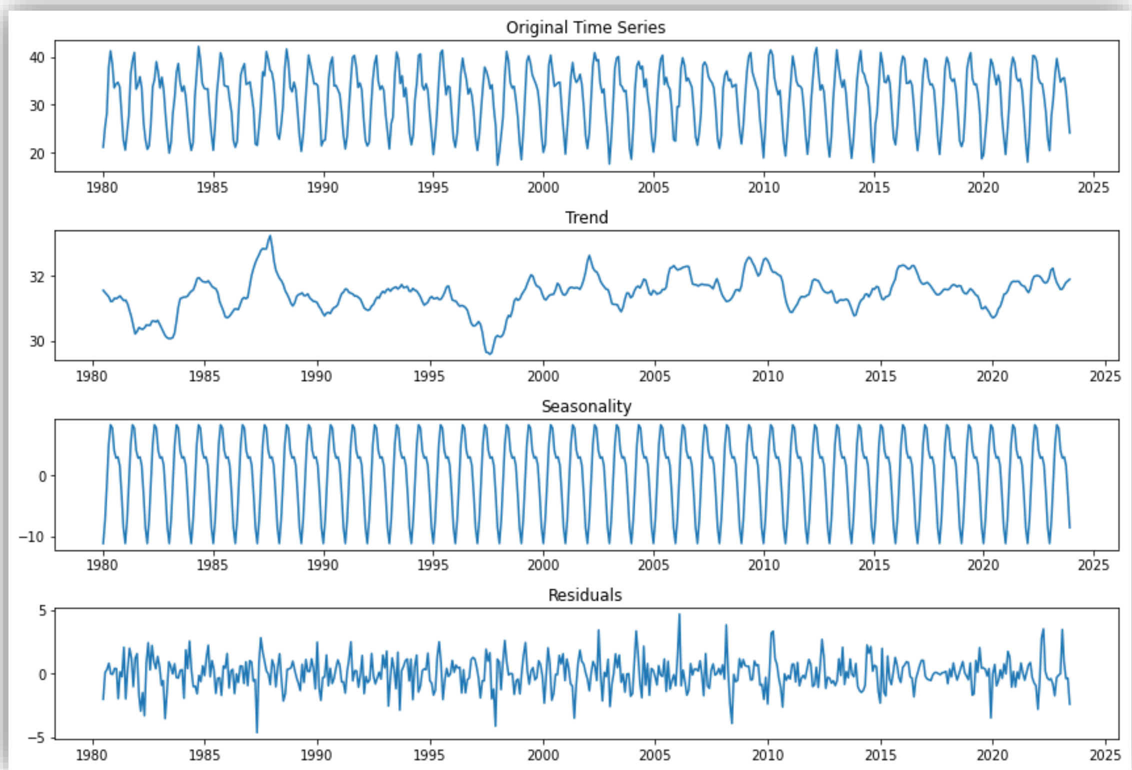


Correlation Matrix - Thiruvananthapuram

- The average maximum and minimum temperatures in Thiruvananthapuram show a moderate positive correlation of **0.4**, indicating that they tend to increase together but with a weaker relationship compared to Chennai.
- The average maximum temperature and rainfall exhibit a moderate negative correlation (**-0.5**), suggesting that higher maximum temperatures are moderately associated with lower rainfall.
- Wind speed has a very weak negative correlation (**-0.2**) with maximum temperature, indicating a negligible inverse relationship.
- Rainfall shows a weak positive correlation (**0.15**) with minimum temperature, indicating a slight tendency for higher rainfall to coincide with higher minimum temperatures.
- Maximum temperature shows the strongest negative relationship with rainfall, while other variables display weaker or negligible correlations with one another.

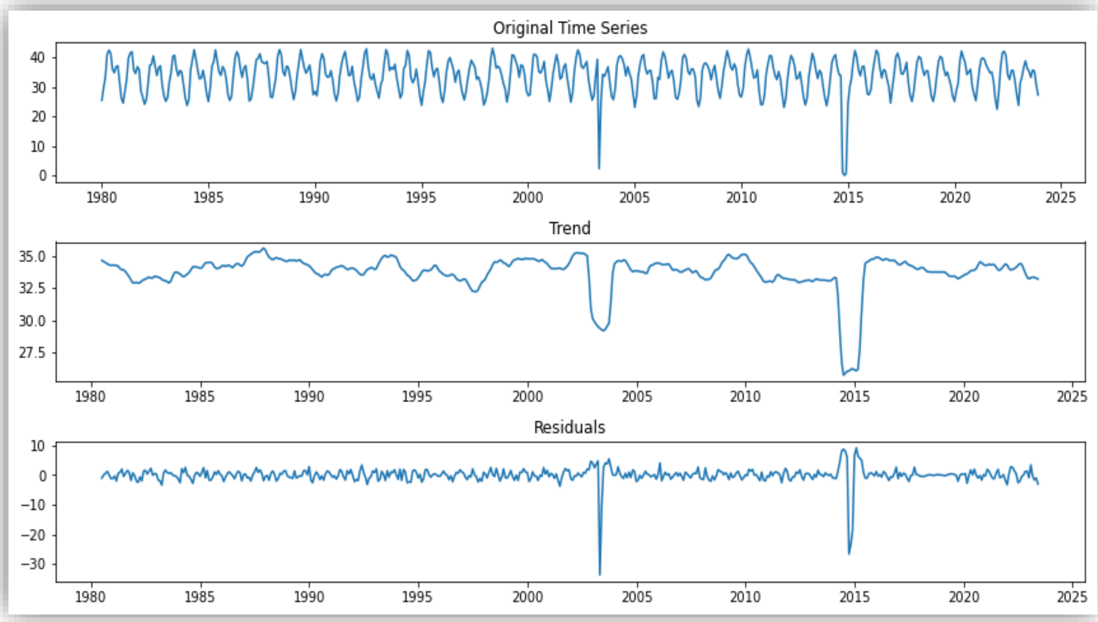
6.3 TIME SERIES ANALYSIS –

6.3.1 Time Series Analysis of Average Daily Temperature –



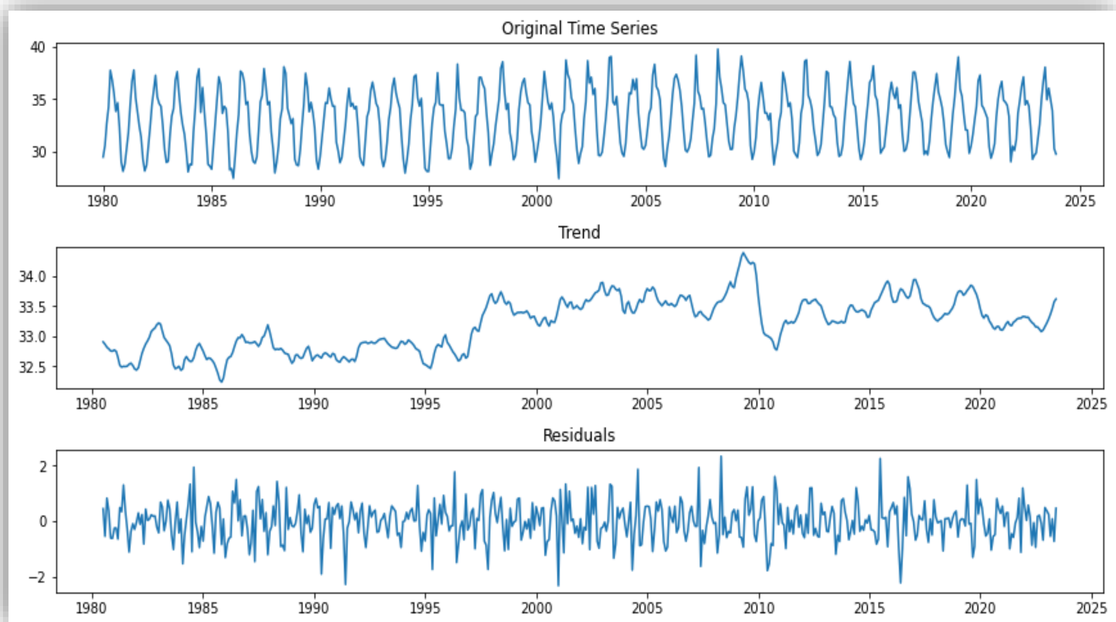
Time Series Decomposition of Monthly Temperature – New Delhi

- **Original Time Series:** The top plot displays the raw data over time, showing clear seasonality of monthly average temperature (regular up-and-down patterns).
- **Trend:** The second plot isolates the overall trend, indicating a slight increase in temperature over time with some fluctuations.
- **Seasonality:** The third plot highlights the repeating seasonal pattern of recorded temperature, likely annual, with regular peaks and troughs.
- **Residuals:** The bottom plot shows the remaining "noise" or irregular fluctuations after the trend and seasonality are removed. There are no significantly large deviations from the expected pattern, indicating relatively stable temperatures.



Time Series Decomposition of Monthly Temperature – Jodhpur

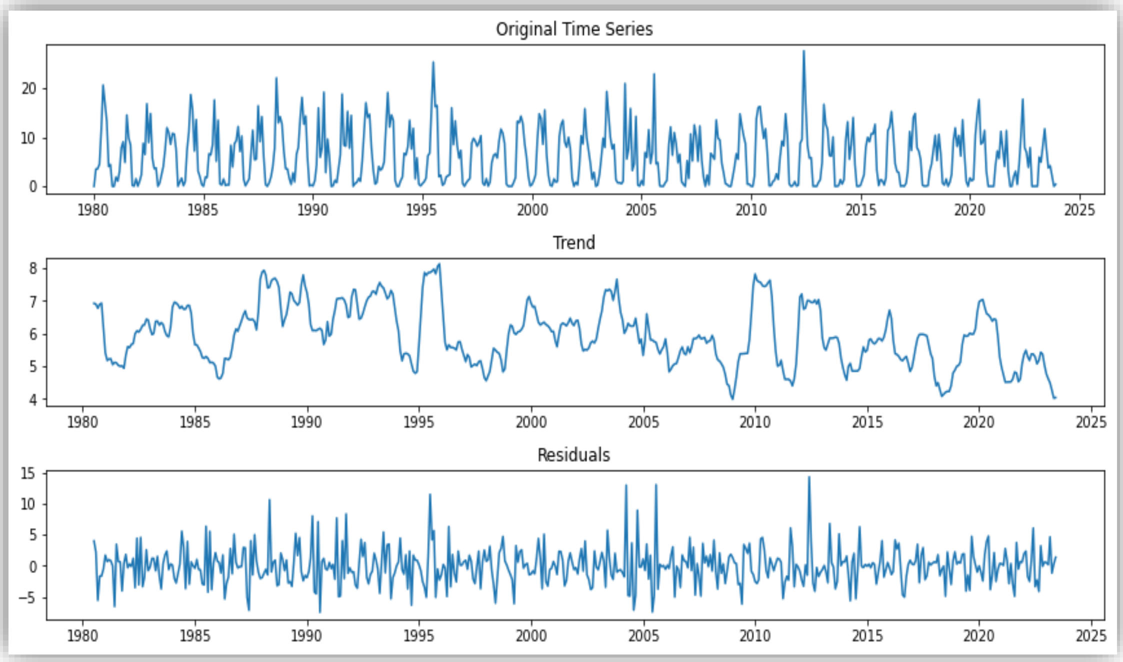
- **Anomalies**: The prominent spikes around 2004-05 and 2015-17 indicate that these are due to missing data for longer period and replaced with null values.
- **Trend:** This plot shows a general upward trend in average temperatures, suggesting a possible warming trend in Jodhpur.
- **Residuals:** This remaining variability shows the significant trend in average temperature but we can notice a significant drop due to anomalies.



Time Series Decomposition of Monthly Temperature – Chennai

- **Original Time Series:** We can see a clear cyclical pattern indicating seasonality, with consistently warm temperatures throughout the year in Chennai city.
- **Trend:** The plot shows a relatively stable trend with a slight upward tilt, suggesting a possible gradual warming in Chennai's average temperatures.
- **Residuals:** The residuals plot represents that there are no significantly large deviations from the expected pattern, indicating relatively stable temperatures in Chennai.

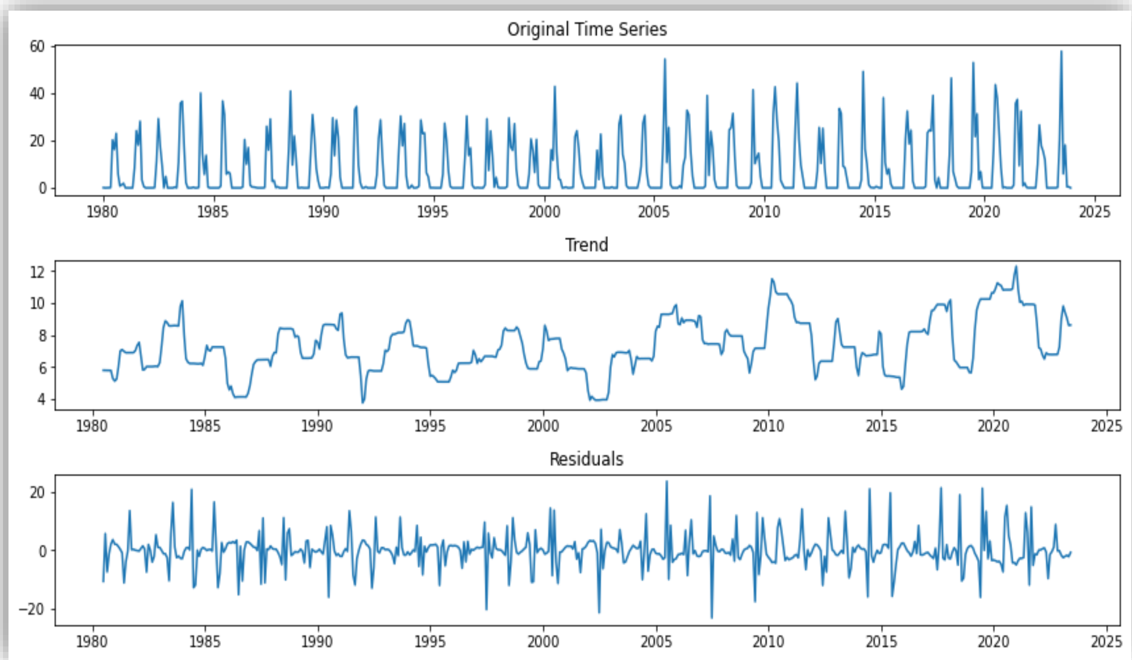
6.3.2 Time Series Analysis of Average Rainfall –



Time Series Analysis of Average Rainfall – Guwahati

- **Original Time Series:** we can see a strong cyclical pattern, which is expected due to the monsoon season that brings heavy rainfall to this region. There are also some peaks and troughs that suggest variability in the amount of rainfall from year to year.
- **Trend:** The overall trend appears to show a relatively stable trend with a slight upward tilt, suggesting a possible gradual increase in average monthly rainfall in Guwahati over time.
- **Residuals:** There are some noticeable spikes, indicating periods where rainfall was significantly higher or lower than expected based on the trend and seasonal pattern. These could represent years with particularly intense monsoons or unusually dry

spells.



Time Series Decomposition of Average Rainfall – Mumbai

- **Original Time Series:** There is a strong cyclical pattern of average rainfall in Mumbai region due to the monsoon season, with peaks typically occurring in June-July. The variability in peak heights suggests fluctuations in monsoon intensity from year to year.
- **Trend:** This shows a relatively stable trend with no strong upward or downward direction, suggesting that Mumbai's average rainfall has remained relatively consistent over the years. But as we join the top peak points then we can notice a slight uptrend showing an increasing average rainfall on this coastal region.
- **Residuals:** There are some noticeable spikes, indicating periods with unusually high or low rainfall, possibly due to particularly strong or weak monsoons or other weather events. Overall residuals are stabilized because of stable trends in rainfall.

6.4 EXTREME VALUE ANALYSIS –

The GEV distribution is fitted for the cities Jodhpur, Mumbai and Bhopal for the extreme event analysis like high temperature, heavy rainfall, and windiest situations respectively. The extreme value probability plot, “Parameter vs Probability Density” is created with GEV distribution function using the daily data records for mentioned cities.

6.4.1 Extreme Value Analysis of Temperature in Jodhpur –

Table: Extreme values of temperature in Jodhpur city to be considered (in ^0C)

41.1	41.1	42.9	43.0
43.0	43.5	44.0	44.0
44.1	44.2	44.4	44.5
44.6	44.7	44.7	44.8
44.8	44.8	44.8	45.0
45.0	45.1	45.2	45.4
45.5	45.6	45.6	45.7
45.7	45.8	45.8	45.9
46.3	46.4	46.4	46.6
46.8	46.8	47.5	47.8
47.8	48.0	48.4	48.8

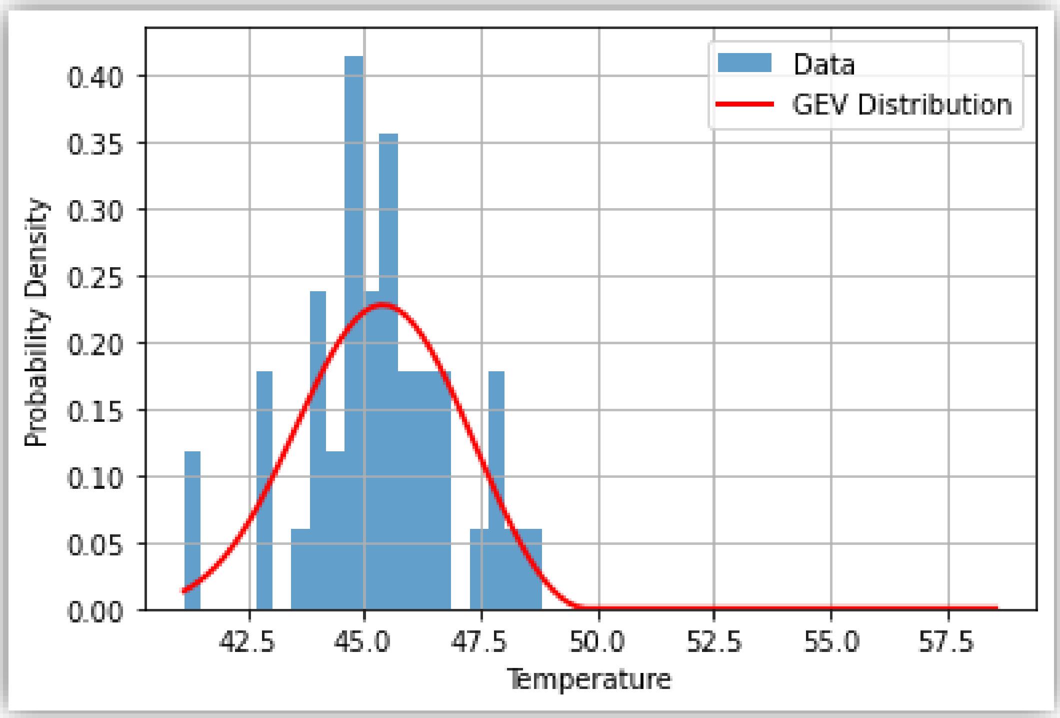
The theoretical cumulative distribution function, TCDF value for the Generalized Extreme Value distribution for plotting probability distribution of above values is computed using the formula:

$$F_T(x) = \begin{cases} \exp\left(-\left[1+\varepsilon\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\varepsilon}}\right), & if\ \varepsilon \neq 0, 1+\ \varepsilon\left(\frac{x-\mu}{\sigma}\right) > 0 \\ \exp\left(-exp\left(-\frac{x-\mu}{\sigma}\right)\right), & if\ \varepsilon = 0 \end{cases}$$

Where,

- x: The value for which we calculate the theoretical CDF.
- μ: The location parameter.
- σ: The scale parameter (σ > 0).
- ε: The shape parameter.

Using the above formula, probability density plot of temperature values is drawn below,



GEV Distribution Fit for Temperature – Jodhpur

GEV Distribution Insights –

- The blue bars in the graph represent the observed frequency distribution of daily maximum temperatures. The red curve shows the fitted GEV probability density function.
- The close alignment of the bars with the curve indicates that the GEV model adequately represents the underlying structure of the data.
- The peak of the red curve i.e. mode aligns well with the most frequently occurring temperature range, suggesting that the location parameter of the GEV model has been estimated correctly.
- The right-hand tail of the red curve captures the rare and extreme temperature events, which is a crucial feature of the GEV distribution for modeling extremes.
- The graph shows a slight right skewness, indicating a tendency for rare extreme temperatures to occur, which the GEV model captures accurately through its shape parameter.

K-S Test for Goodness of Fit:

The likelihood function for a GEV distribution is the product of the probability density functions (PDFs) of the observed data points which will help us to choose optimal values of GEV parameters to compare with actual data and it is calculated by,

$$L(\xi, \mu, \sigma) = \prod_{i=1}^{n} f(x_i; \xi, \mu, \sigma)$$

From the likelihood function for GEV distribution desired parameters can be extracted. The location parameter is 44.73 with scale 1.72 and shape parameter is taken as 0.3414.

Our claim is to check the goodness of fit of the GEV distribution so we can take claim as the null hypothesis and complement as the alternative hypothesis at 5% level of significance.

Thus,

H_0 : GEV distribution may provide a good fit for the data.

I.e. $F(x) = F_T(x)$ for all values of x

H_1 : GEV distribution may not provide a good fit for the data.

I.e. $F(x) \neq F_T(x)$ for at least one value of x

For testing the null hypothesis, the test statistic is given by,

$$D = \max | F_n(x) - F_T(x) |$$

Where, $F_n(x)$ is the empirical cumulative distribution function (ECDF) bases on observed sample and it is given by,

$$F_n(x) = \frac{Number\ of\ data\ points\ \leq x}{Total\ data\ points}$$

For obtaining the empirical cumulative distribution function $F_n(x)$, first we must obtain cumulative frequencies and then by the definition of theoretical cumulative distribution function we obtain $F_T(x)$.

Calculation Table

x	Observed Freq.	Cumulative Frequency	Fn(x)	Ft(x)	D
41.14	1	1	0.022727	0.008024	0.014704
41.14	1	2	0.045455	0.008024	0.037431
42.90	1	3	0.068182	0.083947	0.015765
43.00	1	4	0.090909	0.093141	0.002232
43.00	1	5	0.113636	0.093141	0.020495
43.50	1	6	0.136364	0.150038	0.013674
44.00	1	7	0.159091	0.22584	0.066749
44.00	1	8	0.181818	0.22584	0.044022
44.10	1	9	0.204545	0.243208	0.038663
44.20	1	10	0.227273	0.261266	0.033993
44.40	1	11	0.25	0.299322	0.049322
44.50	1	12	0.272727	0.319247	0.04652
44.60	1	13	0.295455	0.339713	0.044258
44.70	1	14	0.318182	0.360671	0.04249
44.70	1	15	0.340909	0.360671	0.019762
44.80	1	16	0.363636	0.38207	0.018434
44.80	1	17	0.386364	0.38207	0.004293
44.80	1	18	0.409091	0.38207	0.027021
44.80	1	19	0.431818	0.38207	0.049748
45.00	1	20	0.454545	0.425965	0.02858
45.00	1	21	0.477273	0.425965	0.051308
45.10	1	22	0.5	0.448341	0.051659
45.20	1	23	0.522727	0.470918	0.051809
45.40	1	24	0.545455	0.516418	0.029037
45.50	1	25	0.568182	0.539206	0.028976
45.60	1	26	0.590909	0.56193	0.028979
45.60	1	27	0.613636	0.56193	0.051706
45.70	1	28	0.636364	0.584524	0.051839
45.70	1	29	0.659091	0.584524	0.074566
45.80	1	30	0.681818	0.606923	0.074896
45.80	1	31	0.704545	0.606923	0.097623
45.90	1	32	0.727273	0.62906	**0.098213**
46.30	1	33	0.75	0.71379	0.03621
46.40	1	34	0.772727	0.733739	0.038989
46.40	1	35	0.795455	0.733739	0.061716
46.60	1	36	0.818182	0.771812	0.046369
46.80	1	37	0.840909	0.807189	0.03372
46.80	1	38	0.863636	0.807189	0.056447
47.50	1	39	0.886364	0.906734	0.020371
47.80	1	40	0.909091	0.937234	0.028143
47.80	1	41	0.931818	0.937234	0.005416
48.00	1	42	0.954545	0.953621	0.000924
48.40	1	43	0.977273	0.977673	0.0004
48.80	1	44	1	0.991718	0.008282

From the above calculation table we have,

$$D = \max | F_n(x) - F_T(x) | = 0.098213$$

The critical value of test statistic at 5% level of significance is,

$$D\alpha = \frac{1.36}{\sqrt{n}}$$

$$D\alpha = \frac{1.36}{\sqrt{44}}$$

$$D\alpha = 0.2050$$

(We can also obtain this critical value from K-S Statistic table)

Since calculated value of test statistic D (0.0982) is less than critical value (0.2050) so we do not reject the null hypothesis i.e. we accept and support the claim at 5% level of significance.

Thus, we conclude that the GEV distribution provides a good fit to the extreme values from the sample.

GEV Model Prediction Return Levels:

Return Period	Return Level
5-year	46.75801301
10-year	47.44204288
30-year	48.19395049
50-year	48.45159257
100-year	48.73426889

Key Insights –

- The distribution model provides us future return levels of estimated extreme values for the specified periods.
- For a 5-year return period, the expected return level is **46.76**, meaning this value is expected to be equalled or exceeded once every 5 years on average. Similarly, for a

10-year return period, the return level rises to **47.44**, representing an event of greater magnitude that occurs less frequently.

- A more extreme event, with a return level of **48.19**, is expected once every 30 years. For rarer events, the 50-year return level is **48.45**, while the most extreme event in the dataset, with a magnitude of **48.73**, is associated with a 100-year return period, occurring only once in a century on average.
- These increasing return levels align with the expectation that longer return periods are linked to more severe events.

6.4.2 Extreme Value Analysis of Rainfall in Mumbai –

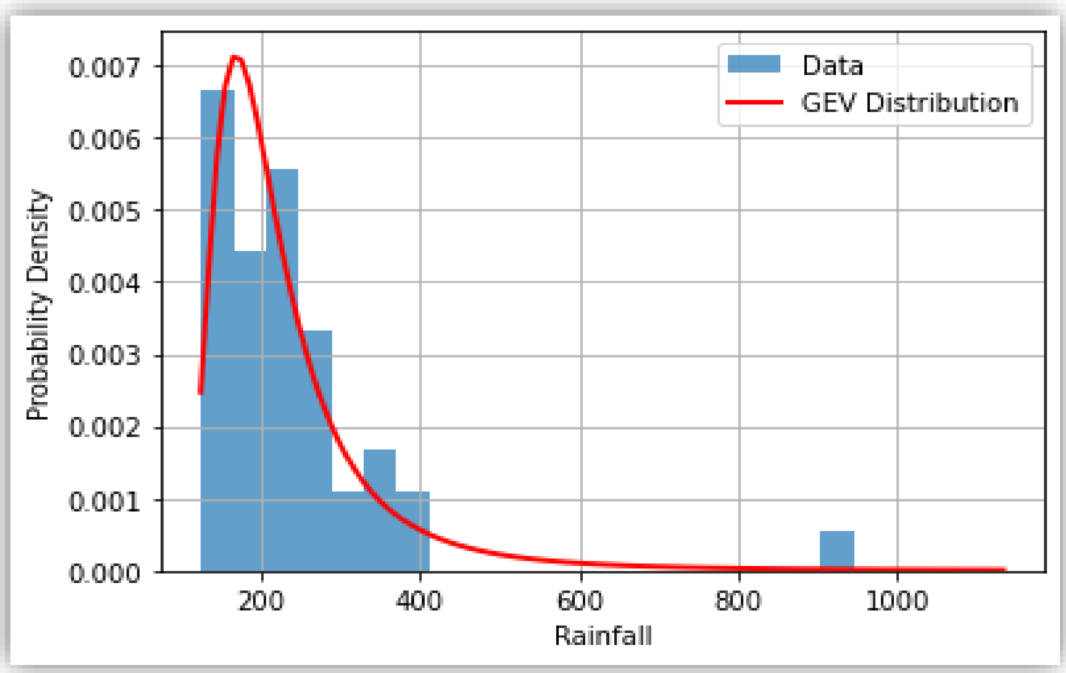
Table. Extreme values of Rainfall in Mumbai city to be considered (in mm)

125.0	**125.7**	**129.7**	**134.4**
142.6	**144.9**	**150.2**	**151.1**
157.2	**158.2**	**161.0**	**161.7**
171.7	**180.0**	**185.3**	**186.0**
192.1	**192.9**	**194.5**	**203.7**
207.2	**208.5**	**211.5**	**215.4**
215.6	**220.4**	**223.6**	**231.0**
231.4	**240.1**	**253.3**	**253.4**
256.6	**274.1**	**283.4**	**286.4**
312.4	**318.2**	**331.4**	**346.2**
351.5	**375.2**	**399.0**	**944.2**

The theoretical cumulative distribution function value for the Generalized Extreme value distribution plotting is computed using the formula,

$$F_T(x) = \begin{cases} \exp\left(-\left[1+\varepsilon\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\varepsilon}}\right), & if\ \varepsilon \neq 0, 1+\varepsilon\left(\frac{x-\mu}{\sigma}\right) > 0 \\ \exp\left(-exp\left(-\frac{x-\mu}{\sigma}\right)\right), & if\ \varepsilon = 0 \end{cases}$$

Hence the probability density plot of GEV distribution of extreme rainfall values,



GEV Distribution Fit for Rainfall – Mumbai

GEV Data Distribution:

- Unlike temperature data, rainfall distributions tend to exhibit a highly skewed structure with a long right tail, capturing the occurrence of rare and extreme rainfall events.
- The peak of the red curve corresponds to the most frequently observed rainfall values, indicating the GEV model effectively captures the central tendency of the data.
- The extended tail of the GEV curve accommodates extreme rainfall events (e.g., floods or unusual heavy downpours), which are critical for hydrological risk assessments.
- The distribution captures a significant range of rainfall data, particularly the low-to-moderate rainfall values, which dominate the dataset.
- The tail behaviour of the GEV curve effectively models rare, high-intensity rainfall events that are key for return-level analysis.

K-S Test for Goodness of Fit:

The likelihood function for a GEV distribution will help us to choose optimal values of GEV parameters to compare with actual data and it is given by,

$$L(\xi, \mu, \sigma) = \prod_{i=1}^{n} f(x_i; \xi, \mu, \sigma)$$

For comparing GEV distribution, most likely parameters are extracted. The location parameter is 183.31 with scale 53.74 and shape parameter is taken as -0.3070.

Our claim is to check the goodness of fit of the GEV distribution so we can take claim as the null hypothesis and complement as the alternative hypothesis at 5% level of significance.

Thus,

H_0 : GEV distribution may provide a good fit for the data.

I.e. $F(x) = F_T(x)$ for all values of x

H_1 : GEV distribution may not provide a good fit for the data.

I.e. $F(x) \neq F_T(x)$ for at least one value of x

For testing the null hypothesis, the test statistic is given by,

$$D = \max | F_n(x) - F_T(x) |$$

Where, $F_n(x)$ is the empirical cumulative distribution function (ECDF) bases on observed sample and it is given by,

$$F_n(x) = \frac{Number\ of\ data\ points\ \leq x}{Total\ data\ points}$$

By obtaining the empirical cumulative distribution function $F_n(x)$ and by the definition of theoretical cumulative distribution function we calculated the values in following table,

Calculation Table

x	Observed Freq.	Cumulative Frequency	Fn(x)	Ft(x)	D
125	1	1	0.022727273	0.02368751	-0.000960237
125.7	1	2	0.045454545	0.02546088	0.019993665
129.7	1	3	0.068181818	0.03721074	0.030971078
134.4	1	4	0.090909091	0.05457488	0.036334211
142.6	1	5	0.113636364	0.09360264	0.020033724
144.9	1	6	0.136363636	0.10632529	0.030038346
150.2	1	7	0.159090909	0.13807984	0.021011069
151.1	1	8	0.181818182	0.14376575	0.038052432
157.2	1	9	0.204545455	0.18405167	0.020493785
158.2	1	10	0.227272727	0.19089334	0.036379387
161	1	11	0.25	0.21031037	0.03968963
161.7	1	12	0.272727273	0.21521508	**0.057512193**
171.7	1	13	0.295454545	0.28635014	0.009104405
180	1	14	0.318181818	0.34501154	-0.026829722
185.3	1	15	0.340909091	0.38141878	-0.040509689
186	1	16	0.363636364	0.38614354	-0.022507176
192.1	1	17	0.386363636	0.42635197	-0.039988334
192.9	1	18	0.409090909	0.43148772	-0.022396811
194.5	1	19	0.431818182	0.44165763	-0.009839448
203.7	1	20	0.454545455	0.49736564	-0.042820185
207.2	1	21	0.477272727	0.51726987	-0.039997143
208.5	1	22	0.5	0.52447832	-0.02447832
211.5	1	23	0.522727273	0.540731	-0.018003727
215.4	1	24	0.545454545	0.56106491	-0.015610365
215.6	1	25	0.568181818	0.5620836	0.006098218
220.4	1	26	0.590909091	0.5858359	0.005073191
223.6	1	27	0.613636364	0.60093862	0.012697744
231	1	28	0.636363636	0.63369595	0.002667686
231.4	1	29	0.659090909	0.63538287	0.023708039
240.1	1	30	0.681818182	0.67005265	0.011765532
253.3	1	31	0.704545455	0.71582908	-0.011283625
253.4	1	32	0.727272727	0.716147	0.011125727
256.6	1	33	0.75	0.72610657	0.02389343
274.1	1	34	0.772727273	0.77384337	-0.001116097
283.4	1	35	0.795454545	0.79515501	0.000299535
286.4	1	36	0.818181818	0.8015064	0.016675418
312.4	1	37	0.840909091	0.84757039	-0.006661299
318.2	1	38	0.863636364	0.8559772	0.007659164
331.4	1	39	0.886363636	0.87305504	0.013308596
346.2	1	40	0.909090909	0.88929171	0.019799199
351.5	1	41	0.931818182	0.89446387	0.037354312
375.2	1	42	0.954545455	0.91417739	0.040368065
399	1	43	0.977272727	0.92948883	0.047783897
944.2	1	44	1	0.99575912	0.00424088

From the above calculation table we have,

$$D = \max | F_n(x) - F_T(x) | = 0.05751$$

The critical value of test statistic at 5% level of significance is,

$$D\alpha = \frac{1.36}{\sqrt{n}}$$

$$D\alpha = \frac{1.36}{\sqrt{44}}$$

$$D\alpha = 0.2050$$

Since calculated value of test statistic D (0.05751) is less than critical value (0.2050) so we do not reject the null hypothesis i.e. we accept and support the claim at 5% level of significance.

Thus, we conclude that the GEV distribution provides a good fit to the extreme values from the sample.

GEV Model Prediction on Return Levels:

Return Period	Return Level
5-year	285.6780
10-year	357.5498
30-year	502.9940
50-year	588.2096
100-year	726.8565

Key Insights –

- The return period and corresponding return levels from the GEV model highlight the increasing severity of events with decreasing frequency.
- For a **5-year** return period, the return level is **285.68**, indicating that such an extreme value is expected to be equaled or exceeded once every 5 years on average. For a **10-year** return period, the return level rises significantly to **357.55**, reflecting a rarer but more intense event.
- At a **30-year** return period, the return level increases further to **502.99**, representing a more severe event that occurs less frequently. For rarer extremes, the **50-year** return level is **588.21**, and the most extreme events, occurring on average once in a century, have a **100-year** return level of **726.86**.

6.4.3 Extreme Value Analysis of Wind Speed in Bhopal –

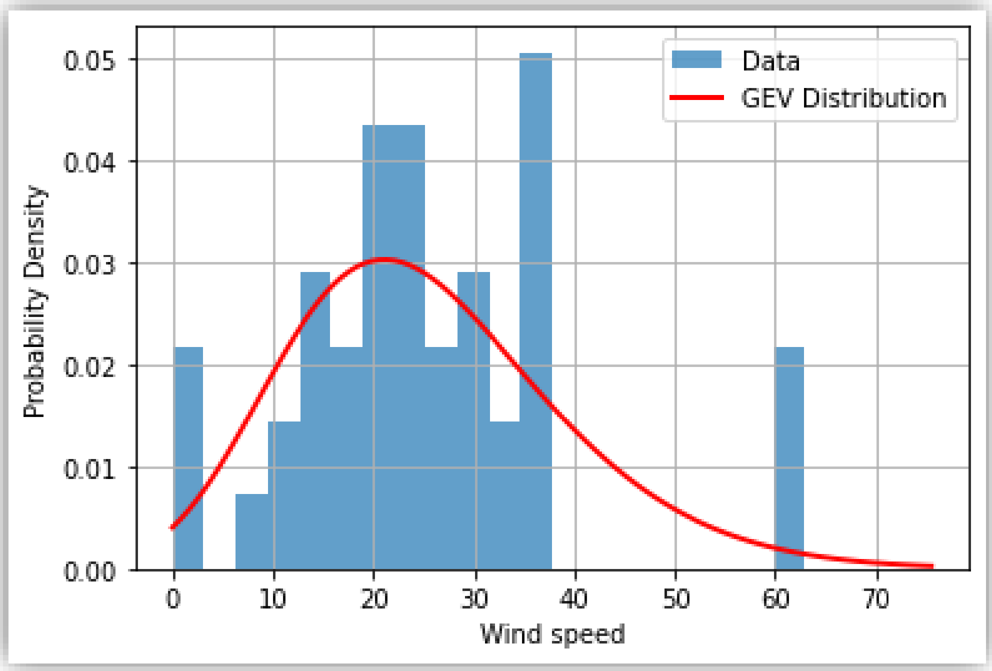
Table. Extreme values of Wind Speed in Bhopal city to be considered (km/h)

0	**0**	**0**	**7**
12	**12**	**13**	**15**
15	**15**	**16**	**17**
18	**19**	**20**	**21**
22	**22**	**22**	**23**
24	**24**	**24**	**24**
25	**26**	**28**	**28**
29	**29**	**30**	**31**
32	**34**	**35**	**35**
35	**36**	**36**	**36**
36	**60**	**61**	**63**

The theoretical cumulative distribution function value for the Generalized Extreme value distribution plotting is computed using the formula,

$$F_T(x) = \begin{cases} \exp\left(-\left[1+\varepsilon\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\varepsilon}}\right), & if\ \varepsilon \neq 0, 1+\ \varepsilon\left(\frac{x-\mu}{\sigma}\right) > 0 \\ \exp\left(-exp\left(-\frac{x-\mu}{\sigma}\right)\right), & if\ \varepsilon = 0 \end{cases}$$

Hence the probability density plot of GEV distribution of extreme rainfall values,



GEV Distribution Fit for Wind Speed – Bhopal

GEV Data Distribution:

- Unlike rainfall and temperature data, wind speed distributions exhibit a more complex pattern, with a potential bimodal structure reflecting different wind regimes.
- The peak of the red curve aligns with the most frequently observed wind speeds (30km/h), demonstrating the GEV model's ability to represent the central tendency of the dataset.
- The extended right tail of the GEV curve captures rare, extreme wind speeds (70km/h), which are crucial for risk assessments in structural engineering and disaster management.
- The GEV distribution effectively represents a broad range of wind speed data, particularly the low-to-moderate speeds, which constitute many of the observations.
- The tail behaviour of the GEV curve provides essential insights for return-level analysis, enabling predictions of extreme wind speeds for long return periods critical for safety and resilience planning.

K-S Test for Goodness of Fit:

The likelihood function for a GEV distribution will help us to choose optimal values of GEV parameters to compare with actual data and it is given by,

$$L(\xi, \mu, \sigma) = \prod_{i=1}^{n} f(x_i; \xi, \mu, \sigma)$$

For comparing GEV distribution, most likely parameters are extracted. The location parameter is 19.5448 with scale 12.2759 and shape parameter is taken as 0.1199.

Our claim is to check the goodness of fit of the GEV distribution so we can take claim as the null hypothesis and complement as the alternative hypothesis at 5% level of significance.

Thus,

H_0 : GEV distribution may provide a good fit for the data.

I.e. $F(x) = F_T(x)$ for all values of x

H_1 : GEV distribution may not provide a good fit for the data.

I.e. $F(x) \neq F_T(x)$ for at least one value of x

For testing the null hypothesis, the test statistic is given by,

$$D = \max | F_n(x) - F_T(x) |$$

Where, $F_n(x)$ is the empirical cumulative distribution function (ECDF) bases on observed sample and it is given by,

$$F_n(x) = \frac{Number\ of\ data\ points\ \leq x}{Total\ data\ points}$$

By obtaining the empirical cumulative distribution function $F_n(x)$ and by the definition of theoretical cumulative distribution function we calculated the values in following table,

Calculation Table

x	Observed Freq.	Cumulative Frequency	Fn(x)	Ft(x)	D
0	1	1	0.022727	0.013657	0.00907
0	1	2	0.045455	0.013657	0.031797
0	1	3	0.068182	0.013657	0.054525
7	1	4	0.090909	0.069782	0.021127
12	1	5	0.113636	0.163746	-0.05011
12	1	6	0.136364	0.163746	-0.02738
13	1	7	0.159091	0.186999	-0.02791
15	1	8	0.181818	0.230392	-0.04857
15	1	9	0.204545	0.230392	-0.02585
15	1	10	0.227273	0.230392	-0.00312
16	1	11	0.25	0.264935	-0.01493
17	1	12	0.272727	0.293099	-0.02037
18	1	13	0.295455	0.322052	-0.0266
19	1	14	0.318182	0.3516	-0.03342
20	1	15	0.340909	0.381546	-0.04064
21	1	16	0.363636	0.411698	-0.04806
22	1	17	0.386364	0.441869	-0.05551
22	1	18	0.409091	0.441869	-0.03278
22	1	19	0.431818	0.441869	-0.01005
23	1	20	0.454545	0.471882	-0.01734
24	1	21	0.477273	0.501571	-0.0243
24	1	22	0.5	0.501571	-0.00157
24	1	23	0.522727	0.501571	0.021156
24	1	24	0.545455	0.501571	0.043883
25	1	25	0.568182	0.530784	0.037398
26	1	26	0.590909	0.559385	0.031524
28	1	27	0.613636	0.61428	-0.00064
28	1	28	0.636364	0.61428	0.022083
29	1	29	0.659091	0.640384	0.018707
29	1	30	0.681818	0.640384	0.041434
30	1	31	0.704545	0.66549	0.039056
31	1	32	0.727273	0.689542	0.037731
32	1	33	0.75	0.712499	0.037501
34	1	34	0.772727	0.755029	0.017698
35	1	35	0.795455	0.774582	0.020873
35	1	36	0.818182	0.774582	0.0436
35	1	37	0.840909	0.774582	0.066327
36	1	38	0.863636	0.792998	0.070639
36	1	39	0.886364	0.792998	0.093366
36	1	40	0.909091	0.792998	0.116093
36	1	41	0.931818	0.792998	**0.13882**
60	1	42	0.954545	0.985013	-0.03047
61	1	43	0.977273	0.986903	-0.00963
63	1	44	1	0.990069	0.009931

From the above calculation table we have,

$$D = \max | F_n(x) - F_T(x) | = 0.13882$$

The critical value of test statistic at 5% level of significance is,

$$D\alpha = \frac{1.36}{\sqrt{n}}$$

$$D\alpha = \frac{1.36}{\sqrt{44}}$$

$$D\alpha = 0.2050$$

Since calculated value of test statistic D (0.13882) is less than critical value (0.2050) so we do not reject the null hypothesis i.e. we accept and support the claim at 5% level of significance.

Thus, we conclude that the GEV distribution provides a good fit to the extreme values from the sample.

GEV Model Prediction on Return Levels:

Return Period	Return Level
5-year	36.3971
10-year	43.7569
30-year	53.6942
50-year	57.8005
100-year	62.9505

Key Insights –

- The return levels represent the estimated wind speeds (in m/s) expected to occur once within specific return periods. For example, a **5-year return level** of **36.40 km/h** suggests that wind speeds of this magnitude are likely to occur at least once every 5 years.
- As the return period increases, the return levels also rise, reflecting the rare but extreme nature of such events. For instance, the **100-year return level** is **62.95 km/h**, indicating the rare occurrence of such high wind speeds.

- The increasing trend of return levels across return periods highlights the ability of the GEV distribution to quantify extreme wind speed events, crucial for long-term planning in infrastructure design and disaster management.

This return table of future extreme magnitude is created using daily records of mentioned parameters.

City	Temperature (^{0}C)	Rainfall (mm)	Wind Speed (Km/H)
New Delhi	(45, 46, 46)	(149, 184, 196)	(32, 40, 43)
Jodhpur	(47, 48, 48)	(132, 204, 238)	(30, 51, 61)
Guwahati	(39, 40, 40)	(141, 174, 187)	(17, 22, 23)
Bhopal	(45, 46, 46)	(184, 247, 271)	(43, 57, 62)
Mumbai	(40, 41, 41)	(357, 588, 726)	(27, 35, 39)
Chennai	(43, 44, 44)	(266, 336, 358)	(23, 36,45)
Thiruvananthapuram	(37, 38, 38)	(171, 217, 234)	(25, 35, 40)

From the table above we conclude that Jodhpur city experiences high temperatures as compared to others and Mumbai is the city with heavy rainfall records. Similarly, we can consider Bhopal for the extreme events of heavy wind speed.

6.5 MULTIVARIATE COMPARISON

In this chapter, three parameters from marine data have been considered which are dry bulb temperature, wet bulb temperature and sea surface temperature. The multivariate comparison test is conducted using ANOVA, analysis of variance followed by Post hoc test.

6.5.1 ANOVA

The following data sample is considered for the comparison test between mentioned parameters.

Table: Marine Data Sample

DATE	DBT	WBT	SST
10-08-1982	33.0	24.0	29.0
16-04-1986	30.0	27.4	4.1
17-04-1986	27.0	25.7	21.3
18-04-1986	26.0	24.7	22.9
20-04-1986	25.0	23.6	24.1
21-04-1986	26.0	24.7	23.9
22-04-1986	26.0	24.7	15.2
22-04-1986	32.0	30.7	14.3
.	.	.	.
.	.	.	.
.	.	.	.
26-06-2017	29.6	25.6	30.2
26-06-2017	29.6	26.2	29.9
24-08-2017	30.4	26.1	28.0
01-09-2017	27.7	24.7	28.8
01-09-2017	27.0	24.5	28.1
02-09-2017	28.2	25.0	28.2
06-10-2017	26.1	24.7	27.8
07-10-2017	29.2	25.5	28.0
07-10-2017	29.6	25.1	28.8
09-11-2017	27.1	20.0	28.9
10-11-2017	32.6	14.2	29.2
30-11-2017	29.1	18.2	26.1

Let us assume,

Null hypothesis (H_0) : Mean of all the groups is equal,

$$\text{i.e. } \mu_1 = \mu_2 = \mu_3$$

Alternative hypothesis (H_1) : At least one of the groups has different mean

$$\text{i.e. } \mu_i \neq \mu_j \text{ for at least one i, j}$$

The level of significance (α) for the test is 5% i.e. α = 0.05.

Summary Table

Groups	Count	Sum	Average	Variance
DBT	172	4732.553704	27.51484711	6.22442112
WBT	172	4092.708333	23.79481589	14.89673649
SST	172	3914.618981	22.75941268	57.81765429

Anova Table

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	2151.424661	2	1075.71233	40.8814994	3.35647E-17	3.013294559
Within Groups	13498.53684	513	26.3129373			
Total	15649.9615	515				

For this test p-value is 3.35647E-17.

➔ p-value < α

Hence, we reject the null hypothesis.

So, there is at least one group which has different mean.

6.5.2 Post Hoc Test (Tukey's HSD Test)

To find which group has different mean we will conduct Post Hoc test on above data sample. Here Tukeys's HSD test is used to comparing the means of all the three groups.

$$q = \frac{\bar{x}_{max} - \bar{x}_{min}}{s.e.} \text{ where } s.e. = \sqrt{MS_W/n}$$

where n = the size of each of the group samples. The statistic q has a distribution called the studentized range q (see Studentized Range Distribution). The critical values for this distribution are presented in the Studentized Range q Table based on the values of α, k (the number of groups), and df_W. If $q > q_{crit}$ then the two means are significantly different.

The value of q for alpha = 0.05 with k and degree of freedom(w) 3 and 513 respectively is 3.32.

Now Tukey's HSD is given by,

$$HSD = q\sqrt{\frac{MSw}{n}}$$

$$HSD = 3.32\sqrt{\frac{26.313}{172}}$$

$$HSD = 1.2986$$

Let,

mean of group DBT, $\mu_1 = 27.5148$,

mean of group WBT, $\mu_2 = 23.7948$,

mean of group SST, $\mu_3 = 22.7594$.

Therefore, the differences are

$|\mu_1 - \mu_2| = 3.72$ > HSD,

$|\mu_2 - \mu_3| = 1.0354$ < HSD,

$|\mu_1 - \mu_3| = 4.7554$ > HSD.

Since both pairs with μ_1 have absolute difference greater than HSD therefore group 1 has different mean than group 2 and 3.

Therefore, average dry bulb temperature is significantly different than average wet bulb temperature and average sea surface temperature.

Chapter 7: MAJOR FINDINGS AND CONCLUSION

7.1 MAJOR FINDINGS

This study provides a comprehensive analysis of climate variability across seven meteorological stations in India, representing diverse climatic zones. The key findings from the research are summarized below:

7.1.1. Descriptive Statistics

- **Temperature Variability:**
  - Jodhpur recorded the highest maximum temperatures (mean of 33.71°C, maximum of 43.09°C), indicative of its desert climate.
  - Guwahati and Thiruvananthapuram exhibited consistent moderate temperatures, reflecting their humid subtropical and monsoon climates, respectively.
  - Minimum temperatures were highest in Chennai (mean of 25.01°C), showing a persistent warm climate.
- **Rainfall Patterns:**
  - Mumbai experienced the highest rainfall variability (mean of 7.33 mm, standard deviation of 11.71), linked to intense monsoon events.
  - Jodhpur had the lowest rainfall (mean of 1.64 mm), consistent with its arid environment.
- **Wind Speeds:**
  - Wind speeds were highest in Bhopal (mean of 7.71 km/h) and lowest in Guwahati (mean of 2.26 km/h), reflecting regional climatic dynamics.
- **Evaporation Rates:**
  - Minimal evaporation was observed in coastal cities like Mumbai and Chennai, likely due to high humidity and cloud cover.

7.1.2. Exploratory Analysis

- Correlation heatmaps revealed:
  - A strong positive correlation between maximum and minimum temperatures across all cities (e.g., 0.93 in Chennai).
  - Moderate to strong correlations between rainfall and minimum temperatures in humid regions, such as Guwahati (0.74).
  - Weak or negligible correlations between wind speeds and rainfall,

underscoring independent dynamics.

7.1.3. Time Series Analysis

- **Temperature Trends:**
    - Long-term warming trends were observed in most cities, with significant seasonal patterns, particularly in New Delhi and Chennai.
    - Anomalies in Jodhpur around 2004-2005 and 2015-2017 suggest data gaps or unusual climatic events.
- **Rainfall Trends:**
    - Guwahati and Mumbai displayed cyclical patterns driven by monsoons, with Mumbai showing a slight increasing trend in average rainfall over the decades.

7.1.4. Extreme Value Analysis

- **Temperature (Jodhpur):**
    - Extreme temperature events are predicted to exceed 48°C once every 50 years, highlighting the increasing frequency of heatwaves.
- **Rainfall (Mumbai):**
    - Extreme rainfall events (>150 mm) are becoming more frequent, with implications for urban flooding.
- **Wind Speed (Bhopal):**
    - Occasional extreme wind speeds (>20 km/h) indicate susceptibility to strong wind events.

7.1.5. Multivariate Analysis

- **ANOVA and Post Hoc Tests:**
    - Significant differences in Dry Bulb Temperature and Wet Bulb Temperature patterns were observed across Sea Surface Levels, confirming distinct climatic behaviours.
    - Tukey's HSD test highlighted that Dry Bulb Temperature significantly differs from Wet Bulb Temperature and Sea Surface Temperature.

7.2. CONCLUSION

The study effectively demonstrates how climatic parameters vary across India's diverse regions and highlights the interplay between temperature, rainfall, wind speed, and evaporation. Key conclusions include:

7.2.1. Regional Climate Vulnerabilities:

- Arid regions like Jodhpur face increasing risks of extreme temperatures, necessitating adaptive strategies for water conservation and heat management.
- Coastal cities such as Mumbai and Chennai, while moderated by their proximity to the sea, are increasingly vulnerable to extreme rainfall events.

7.2.2. Impact of Monsoons:

- The variability in rainfall across regions underscores the critical role of monsoons in shaping India's climate. Coastal and northeastern cities experience significant monsoonal impacts, influencing agriculture, urban planning, and disaster management.

7.2.3. Trends in Climate Extremes:

- The upward trend in extreme temperature and rainfall events indicates ongoing climatic shifts likely driven by global warming.

7.2.4. Policy Implications:

- Findings provide a strong basis for policymakers to prioritize climate-resilient urban planning, particularly in regions prone to extreme weather events.
- Enhanced early warning systems for heatwaves and heavy rainfall are crucial to mitigate their impact on vulnerable populations.

7.2.5. Future Research Directions:

- Expanding the dataset to include more stations and variables (e.g., humidity, cloud cover) would provide deeper insights.
- Integration of machine learning models for predictive analysis can improve the accuracy of extreme event forecasting.

This research serves as a vital resource for understanding regional climatic behaviors in India. By offering detailed statistical insights, it lays the foundation for developing sustainable and adaptive strategies to address the challenges posed by climate variability.

Chapter 8: REFERENCES

8.1 Bibliography

- “Statistical Methods in the Atmospheric Sciences” by Daniel S. Wilks
- Krishnan, R., et al. (2020). "Assessment of Climate Change over the Indian Region." Ministry of Earth Sciences, Government of India.
- Goswami, B. N., et al. (2006). "Increasing Trend of Extreme Rain Events Over India in a Warming Environment." *Science*, 314(5804), 1442-1445.
- Guhathakurta, P., & Rajeevan, M. (2008). "Trends in the Rainfall Pattern Over India." *International Journal of Climatology*, 28(11), 1453-1469.
- Mokhov, I. I., & Semenov, V. A. (2000). "Climate Variability and Changes." *Izvestiya Atmospheric and Oceanic Physics*, 36(1), 1-13.
- Indian Meteorological Department (IMD). "Climatic Data Archive." Retrieved from IMD Website.
- Python Software Foundation. "Python Programming Language." Retrieved from Python.org.
- Seaborn and Matplotlib Documentation. "Data Visualization Tools."

8.2 Websites:

- https://imdpune.gov.in (Indian Meteorological Department, Pune)
- https://dsp.imdpune.gov.in (Data Supply Portal)
- https://cdsp.imdpune.gov.in (Climate Data Service Portal)
- https://www.researchgate.net (Publications)
- Python.org (Visualization Tools)