Project title: Weather condition in Tehran 2023-2024

CMSC6950 Fall 2025   **Mona Hejazi**   **201996350**

Introduction

Project Aim

This project investigates a one-year weather dataset for Tehran, with a primary focus on identifying extreme values in temperature and precipitation. Through statistical analysis, the study aims to uncover insights into Tehran's climatic trends and extreme weather events over the observed period.

Background

Weather data serves as a crucial resource for understanding climatic patterns, forecasting, and decision-making in various sectors. By analyzing meteorological parameters such as temperature, precipitation, and wind patterns, we can identify anomalies, explore correlations, and assess their potential impacts. This project explores the statistical distribution of key weather variables for Tehran, delving into both summary statistics and outlier detection.

Dataset Overview:

The dataset contains daily weather records for Tehran from **October 1, 2023**, to **September 30, 2024**. It comprises 366 entries, representing each day of the year, and includes 11 meteorological variables:
 - `date`: Date of the weather record
 - `tavg`: Average temperature (°C)
 - `tmin`: Minimum temperature (°C)
 - `tmax`: Maximum temperature (°C)
 - `prcp`: Precipitation (mm)
 - `snow`: Snowfall (cm) — with limited data
 - `wdir`: Wind direction (degrees)
 - `wspd`: Wind speed (m/s)
 - `wpgt`: Wind gust (m/s) — with missing data
 - `pres`: Atmospheric pressure (hPa)
 - `tsun`: Sunshine duration (hours) — missing data

Objectives

The project's primary objectives include:
1) Analyzing daily averages, minimum and maximum temperatures, and precipitation trends across the year.
2) Identifying and visualizing extreme values in temperature and precipitation using statistical methods.
3) Exploring correlations between key variables to understand weather dynamics in Tehran.

Hypothesis

 1) The average temperature, wind speed, and precipitation do not follow a normal distribution.
 2) There is no linear correlation between average temperature and precipitation.
 3) A linear correlation exists between wind speed and precipitation.
 4) Removing outliers enhances the reliability of statistical analysis.

Source : https://meteostat.net/en/place/ir/tehran?s=40754&t=2023-10-01/2024-09-30

Methodology

Data Processing and Handling Missing Values

The project began with a detailed inspection of the dataset to evaluate its structure and identify issues such as missing values. Several columns, including **snow**, **wpgt**, and **tsun**, contained substantial gaps in data. To maintain the reliability of the analysis, these variables were excluded from further investigation.

Summary Statistics

Key weather variables—**tavg** (average temperature), **tmin** (minimum temperature), **tmax** (maximum temperature), **wspd** (wind speed), **pres** (atmospheric pressure), and **prcp** (precipitation)—were analyzed to calculate summary statistics. Metrics such as mean, standard deviation, minimum, maximum, and interquartile range (25th, 50th, and 75th percentiles) were computed. These measures provided a foundational understanding of the dataset, highlighting central tendencies, variability, and potential anomalies.

Extreme Temperature Analysis
Workflow Overview

A Python-based analytical pipeline was developed to examine Tehran's annual weather trends. The workflow included importing and preprocessing the dataset from an Excel file. The **date** column was converted to a datetime format and set as the index to enable time-series operations. This preprocessing step facilitated the structured analysis of weather data across time.

Identifying and Highlighting Extreme Values

To detect significant deviations in Tehran's weather patterns:

1. A custom function identified the highest and lowest temperatures along with their corresponding dates.
2. These values were critical for understanding extreme weather conditions in the dataset.

Visualization of Temperature Trends

Visual representations were created to communicate findings effectively:

- **Daily Trends**: A line plot illustrated daily minimum, maximum, and average temperatures, with shaded areas denoting the range between the minimum and maximum.
- **Extreme Highlights**: Extreme temperature values were marked and annotated to emphasize their significance.
- **Annual Mean Temperature**: A horizontal reference line displayed the yearly average temperature for context.

This systematic approach provided insights into both extreme events and broader temperature trends in Tehran.
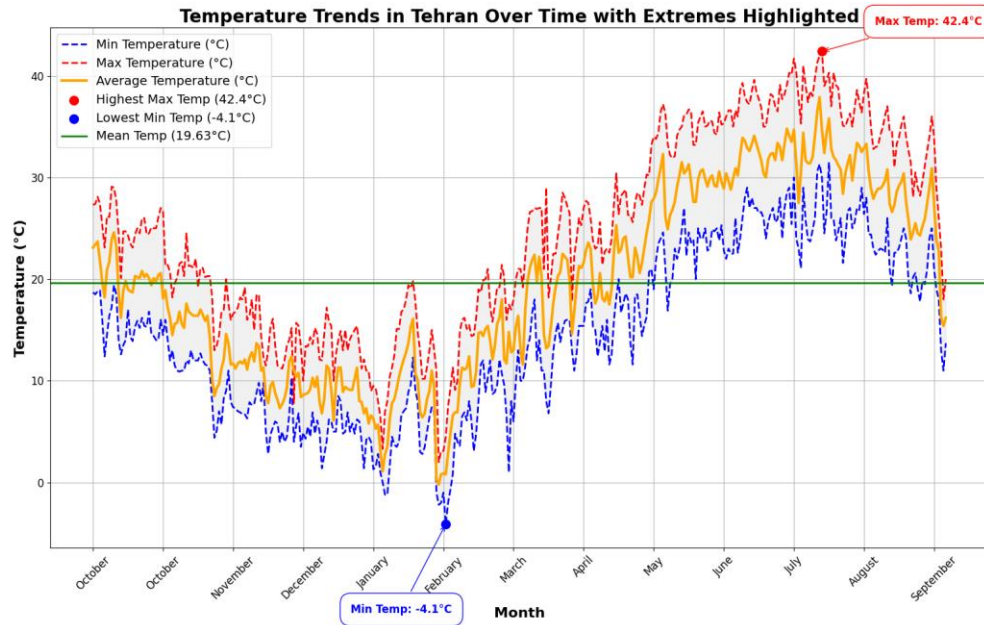
Figure 1 : This plot illustrates the trends in minimum, maximum, and average temperatures in Tehran over time. Highlighted are the extreme temperatures: the highest maximum temperature and the lowest minimum temperature, along with their corresponding dates. Shaded areas between the minimum and maximum temperatures represent the range of temperature variation on each day. The green horizontal line represents the mean temperature.

Analysis of Deviations from Historical Mean

Objective and Methodology

To investigate temperature trends deviating from the historical mean, a threshold of ±1.5 standard deviations (SD) was employed. Daily temperatures were classified into three categories:

- **Above**: Significantly warmer days (+1.5 SD).
- **Below**: Significantly cooler days (-1.5 SD).
- **Normal**: Temperatures within ±1.5 SD of the mean.

Categorization and Monthly Trends

Each day was categorized based on its temperature value:

- Days exceeding the upper threshold were labeled as "Above."
- Days below the lower threshold were labeled as "Below."
- Remaining days were categorized as "Normal."

Monthly counts of these categories were tallied to determine the predominant trend for each month.

Visualization and Key Findings

- **Scatter Plot**: Color-coded markers identified the three temperature categories.
- **Summary Table**: Monthly counts and trends were summarized in a table.

- **Findings**:
  - **Warmest Month**: July had the highest number of "Above" days, marking it as the warmest month.
  - **Coldest Month**: February had the most "Below" days, making it the coldest.

This analysis provided a nuanced understanding of temperature anomalies and seasonal patterns.
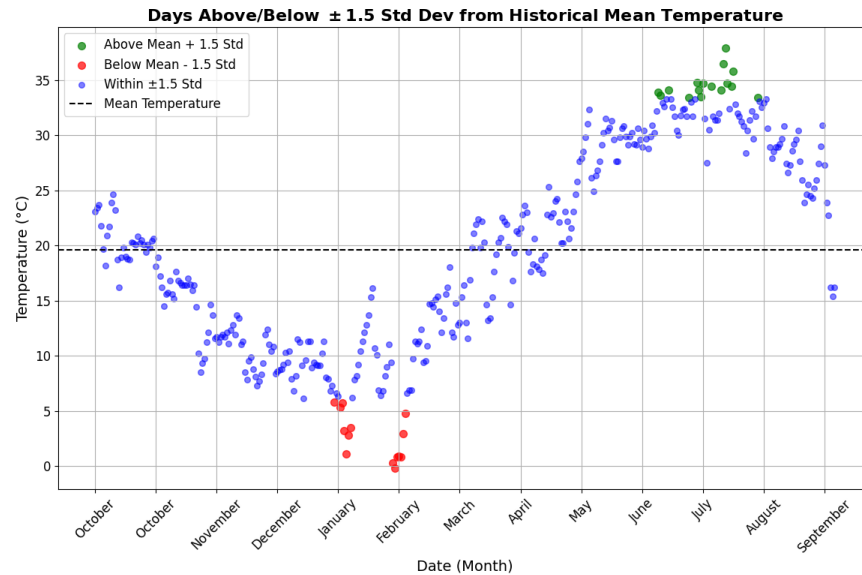

Figure 2 :This plot illustrates the temperature trends over time, showing days where the temperature is above or below the historical mean by ±1.5 standard deviations. The green dots represent days with temperatures above the mean +1.5 standard deviations, the red dots represent days below the mean -1.5 standard deviations, and the blue dots represent days within the ±1.5 standard deviation range.

Table 1: Monthly Trends in Temperature Above/Below ±1.5 Std

month	days_above	days_below	total_days	trend
January	0	3	3	below
February	0	9	9	below
March	0	2	2	below
April	0	0	0	below
May	0	0	0	below
June	0	0	0	below
July	9	0	9	above
August	7	0	7	above
September	0	0	0	below
October	0	0	0	below
November	0	0	0	below
December	0	0	0	below

Seasonal Temperature Analysis

Seasonal Division

The dataset was segmented into four seasons using the following date ranges:

- **Winter**: December 21–March 20
- **Spring**: March 21–June 20
- **Summer**: June 21–September 20
- **Fall**: September 21–December 20

Boxplot Analysis

Two boxplots were generated for each season to examine temperature distributions before and after outlier removal:

- **Outlier Identification**:
    - Z-Score ($\pm$1.5 SD).
    - Interquartile Range (IQR).
- **Visualization Enhancements**:
    - Outliers identified by Z-Score were highlighted in yellow.
    - Outliers identified by IQR were marked in orange.
- **Statistics**: Seasonal mean and standard deviation values were annotated on the plots.

Outlier Removal

Outliers were removed to recalibrate seasonal statistics and assess the impact on data distributions. The percentage of removed data for each season was computed as follows:

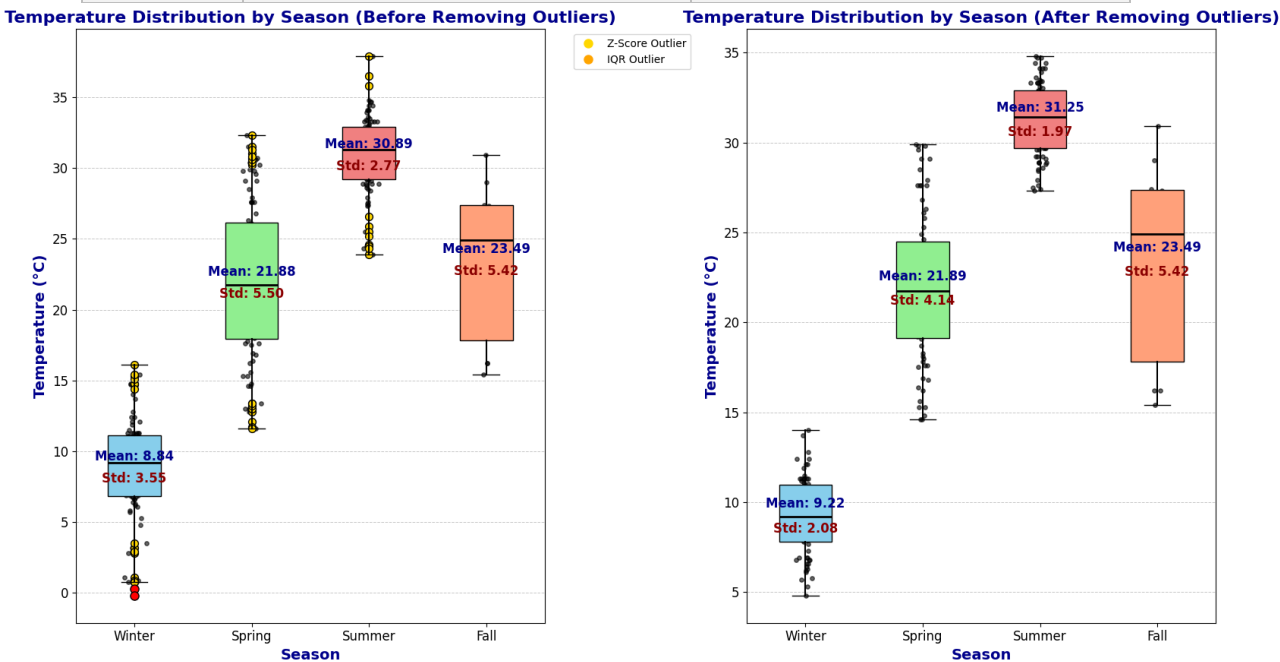| Season	Number of Outliers Removed	Percentage of Data Removed
Winter	17	18.68%
Spring	18	19.57%
Summer	11	11.96%
Fall	0	0.00%
Total	**46**	**16.14%**

Figure 3: Comparison of temperature distributions across seasons.
The left plot shows the original data with Z-Score (gold) and IQR (orange) outliers highlighted.
The right plot shows the cleaned data with outliers removed for better representation of seasonal trends.

Seasonal Heatmap

To visualize temperature variations, daily averages for each season were combined into a matrix. A heatmap, with a color gradient from blue (cold) to red (warm), displayed these seasonal averages.
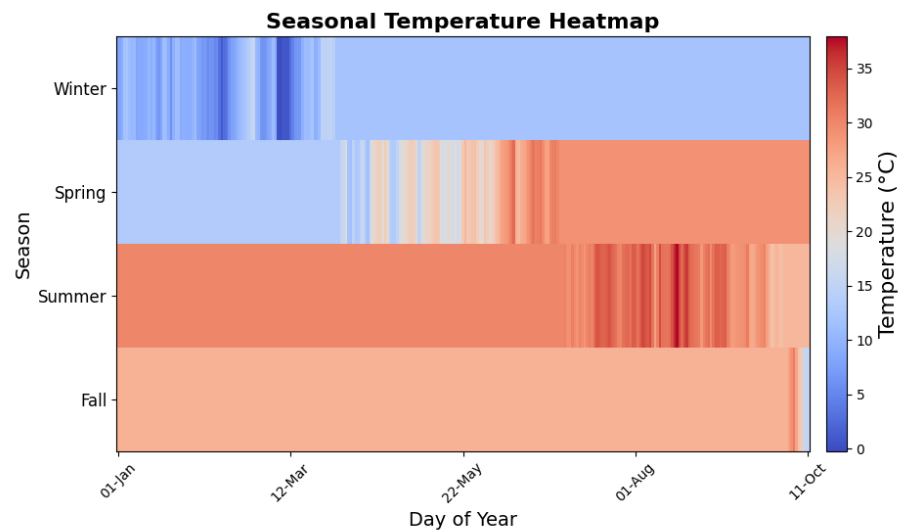

Figure 4 : This heatmap represents the average daily temperatures across seasons in Tehran for the year 2024.

Statistical Analysis

Distributions of Weather Variables
- Variables analyzed: **tavg**, **wspd**, **prcp**.
- Methods:
  - Histograms and kernel density estimates (KDE).
  - Skewness, kurtosis, and Shapiro-Wilk normality tests.
- **Results**: None of the analyzed variables followed a normal distribution.

```
Statistical Analysis Table:
      Skewness  Kurtosis Shapiro-Wilk Test Stat Shapiro-Wilk P-value
tavg   0.00983  -1.15549               0.956811                  0.0
wspd   1.18579  2.034058               0.927751                  0.0
prcp  5.042631   28.8941                0.29523                  0.0

      Normal Distribution
tavg                   No
wspd                   No
prcp                   No
```
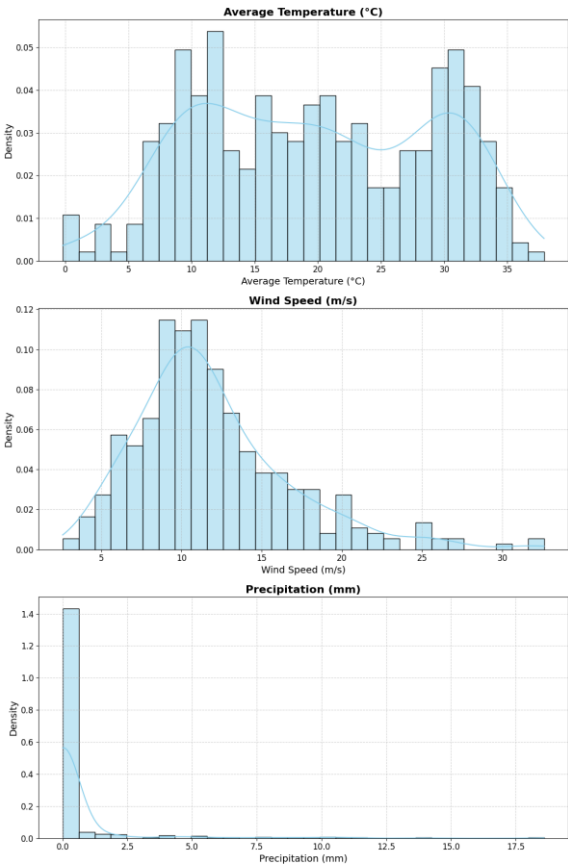

Figure 6: Distributions of temperature, wind speed, and precipitation with KDE.

Temperature and Precipitation Trends

Temperature and precipitation trends were jointly analyzed using a time-series plot:

- **Temperature**: Daily minimum, maximum, and average values.
- **Precipitation**: Represented as dark purple bars.
- **Findings**:
    - Maximum precipitation (18.6 mm) occurred in September.
    - Driest months: October, December, June, July, and August.

This comprehensive methodology ensured robust insights into Tehran's weather dynamics over the observed period.
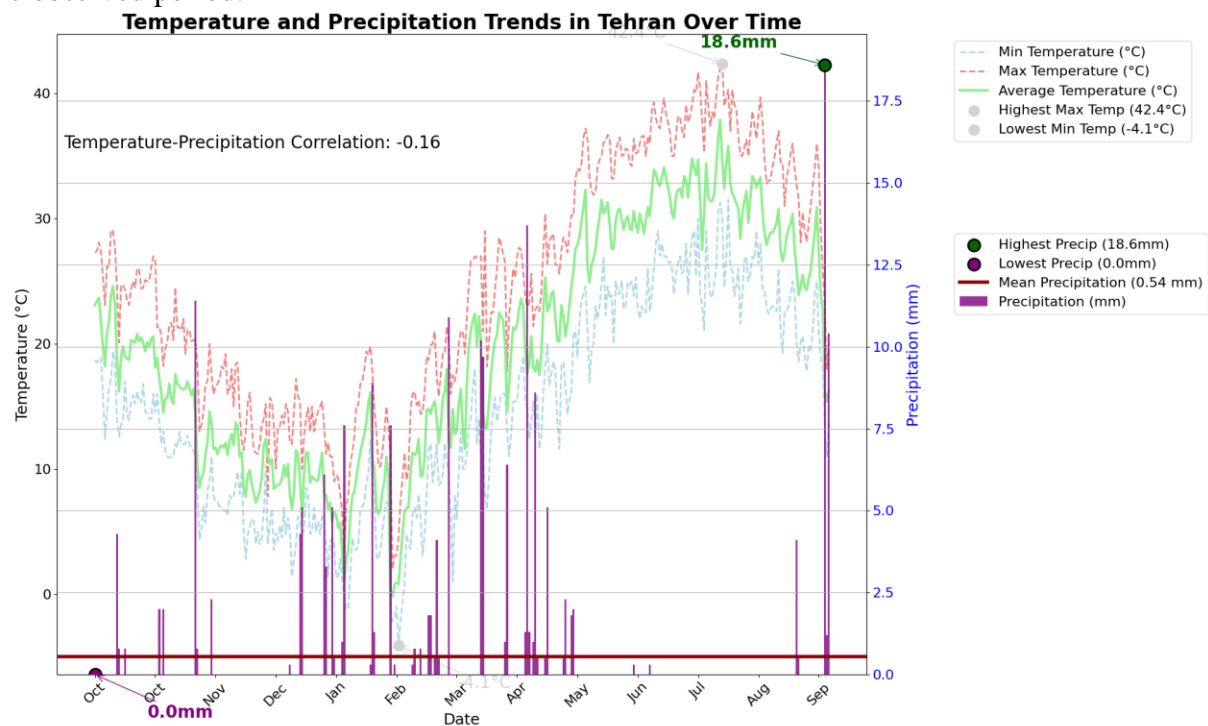

Figure 7 :This plot illustrates the trends in minimum, maximum, and average temperatures in Tehran over time. Temperature trends are shown in light colors, while precipitation is represented with darker purple bars. Extreme values for both temperature and precipitation are highlighted, and the correlation between temperature and precipitation is displayed in the top left corner.

Visualizing Temperature and Precipitation Trends

To analyze seasonal trends, weather data was processed to compute key statistics—minimum, maximum, mean, median, and standard deviation—for temperature, precipitation, and wind speed. These metrics were organized into a comprehensive table for better interpretation.

Visualization Technique (Grouped Bar Chart): A bar chart compared the average values of these variables across the four seasons. Precise annotations were included to enhance readability.

Key Findings

- **Wind Speed**: Remained consistent across seasons, ranging between 11 and 14 m/s.
- **Precipitation**: Minimal during summer, averaging only 0.1 mm.
- **Temperature**: Summer recorded the highest average temperature of 30.9°C.
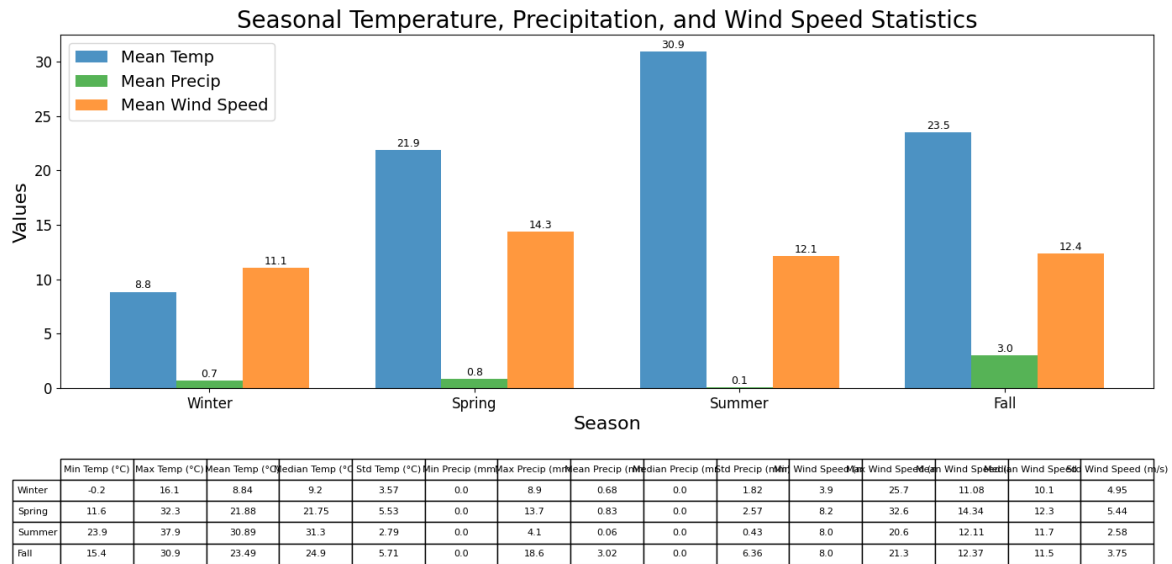
Figure 8 and Table 2: Seasonal Temperature, Precipitation, and Wind Speed Statistics

	Min Temp (°C)	Max Temp (°C)	Mean Temp (°C)	Median Temp (°C)	Std Temp (°C)	Min Precip (mm)	Max Precip (mm)	Mean Precip (mm)	Median Precip (mm)	Std Precip (mm)	Min Wind Speed	Max Wind Speed	Mean Wind Speed	Median Wind Speed	Std Wind Speed (m/s)
Winter	-0.2	16.1	8.84	9.2	3.57	0.0	8.9	0.68	0.0	1.82	3.9	25.7	11.08	10.1	4.95
Spring	11.6	32.3	21.88	21.75	5.53	0.0	13.7	0.83	0.0	2.57	8.2	32.6	14.34	12.3	5.44
Summer	23.9	37.9	30.89	31.3	2.79	0.0	4.1	0.06	0.0	0.43	8.0	20.6	12.11	11.7	2.58
Fall	15.4	30.9	23.49	24.9	5.71	0.0	18.6	3.02	0.0	6.36	8.0	21.3	12.37	11.5	3.75

The figure shows the average temperature (°C), precipitation (mm), and wind speed (m/s) across seasons (Winter, Spring, Summer, and Fall) as grouped bar charts. The table below summarizes detailed seasonal statistics, including minimum, maximum, mean, median, and standard deviation for each variable. The data is derived from Tehran's weather dataset, highlighting seasonal variations in climate parameters.

Correlation Analysis

Correlation Heatmap

A correlation heatmap was developed to examine relationships between key numerical variables, including: **tavg** (average temperature), **prcp** (precipitation), **wspd** (wind speed), and **pres** (atmospheric pressure).

Variables with significant missing data, such as snow, wind gust, and solar radiation, were excluded. The correlation matrix was computed, and a heatmap with annotated coefficients and a color scale was generated to visually represent interdependencies.

Findings

- **Temperature and Pressure**: Moderate negative correlation (-0.76).
- **Temperature and Precipitation**: Very weak negative correlation (-0.16).
- **Wind Speed and Pressure**: Weak negative correlation (-0.31).
- **Wind Speed and Temperature**: Weak positive correlation (0.12).

These results highlight a strong inverse relationship between temperature and pressure, while other pairs exhibited minimal correlation.

Pairplot Analysis

A pairplot was created to further explore pairwise relationships among the variables. This visualization included:

- **Kernel Density Estimations (KDE)**: Displayed on the diagonal to represent individual variable distributions.
- **Scatterplots**: Showcased pairwise relationships for each combination of variables.

Findings

- **Distributions**: Temperature and pressure showed relatively even distributions, while precipitation and wind speed were highly skewed.
- **Scatterplots**: Revealed weak or no significant trends, except for a slight negative correlation between temperature and wind speed/pressure.

Overall, the pairplot emphasized the lack of strong inter-variable correlations in the dataset.
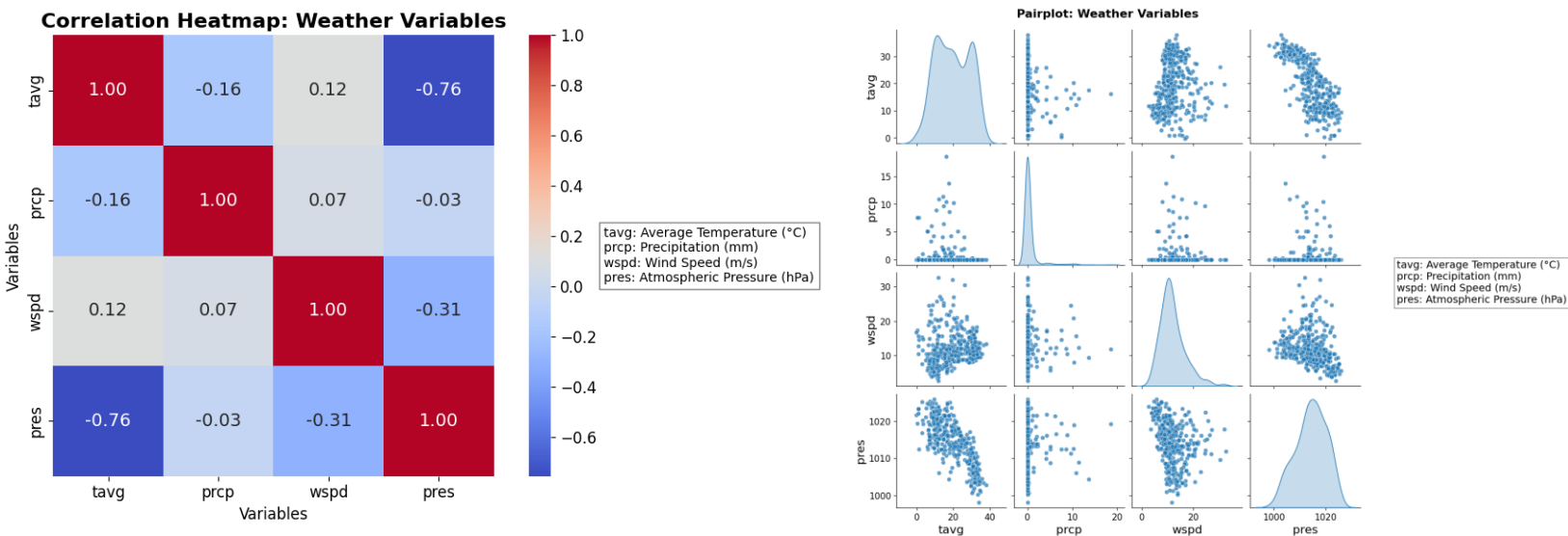

Figure 8: The correlation heatmap and pairplot illustrate relationships among selected weather variables in Tehran.

Discussion

The analysis of Tehran's weather data for one year has provided meaningful insights into the city's climatic patterns and extreme weather events. By employing statistical techniques—such as summary statistics, correlation analysis, and outlier detection—combined with advanced visualization methods, the study successfully identified key trends and anomalies in temperature and precipitation.

The findings revealed distinct seasonal characteristics of Tehran's climate. For instance, July was identified as the warmest month, while February was the coldest. These seasonal variations were further validated through the analysis of temperature trends, which highlighted pronounced deviations during these extreme months. Correlation analysis uncovered minimal linear relationships between average temperature and precipitation, supporting the hypothesis that these variables are weakly correlated. Furthermore, the removal of outliers improved the dataset's reliability, resulting in cleaner and more accurate distributions for analysis.

However, the study faced limitations. Significant gaps in data, particularly for variables like wind gust and sunshine duration, restricted the scope of the investigation. These limitations highlight the need for future studies to integrate additional data sources to address missing information and enable a more comprehensive analysis of Tehran's climatic dynamics.

Conclusion

This project demonstrates the value of statistical and visualization techniques in climate research. Through the analysis of Tehran's weather patterns, it has provided critical insights into seasonal trends and extreme events. The identification of outliers and anomalies contributes to a better understanding of climatic variability, offering valuable information for decision-makers to prepare for and mitigate the impacts of extreme weather events.

Overall, the findings not only enhance the understanding of Tehran's weather but also establish a strong foundation for future research aimed at exploring its climate more comprehensively.