**SHADOWFOX INTERSHIP**

**TASK-3 DOCUMENTATION**

**TABLE OF CONTENT**

|  |
| --- |
| **• Introduction**  **• Research Question**  **• Analysis Methodology**  **• Results and Findings**  **• Code**  **• Conclusions**  **• Future Directions**  **• Reference**  **• Outcomes** |

**Climate Change Analysis**

This documentation provides an overview of the analysis conducted on the global climate dataset spanning from January 2020 to May 2024. The analysis aims to explore various climatic features, identify trends, patterns, and relationships within the data, and derive insights into the impact of climate change.

**1.** **Introduction**

The analysis focuses on understanding the dynamics of climate change using a comprehensive dataset sourced from Kaggle. The dataset includes monthly measurements of various climatic factors such as temperature, precipitation, humidity, CO2 concentration, and others. Exploring this dataset provides valuable insights for climate research, weather forecasting, environmental analysis, impact assessment, and policy-making.

**2. Data Exploration**

The exploratory data analysis (EDA) phase involves:

* + - Loading the dataset and inspecting its structure.
    - Checking for missing values and handling them appropriately.
    - Summary statistics to understand the distribution and variability of the data.
    - Visualizations to explore relationships and trends within the dataset.

**3.** **Research Question**

**Question 1:**

"How do urbanization levels affect particulate matter concentration and air quality over time in different regions?"

**Objective:**

**To investigate the relationship between urbanization and air quality, focusing on particulate matter concentration. This analysis will help understand how increasing urbanization impacts air pollution levels and identify regions most affected by this change.**

**Approach:**

* Segment the data based on Urbanization\_Index.
* Analyze trends in Particulate\_Matter (µg/m³) over time across different levels of urbanization.
* Use visualizations such as line plots and bar plots to compare particulate matter concentration in urban vs. rural areas.
* Perform statistical analysis to determine the significance of observed trends.

**Question 2:**

"What is the impact of proximity to water bodies on temperature and humidity levels?"

**Objective:**

To explore how the distance to the nearest water body influences local climatic conditions, particularly temperature and humidity. This research can provide insights into the moderating effects of water bodies on climate.

**Approach:**

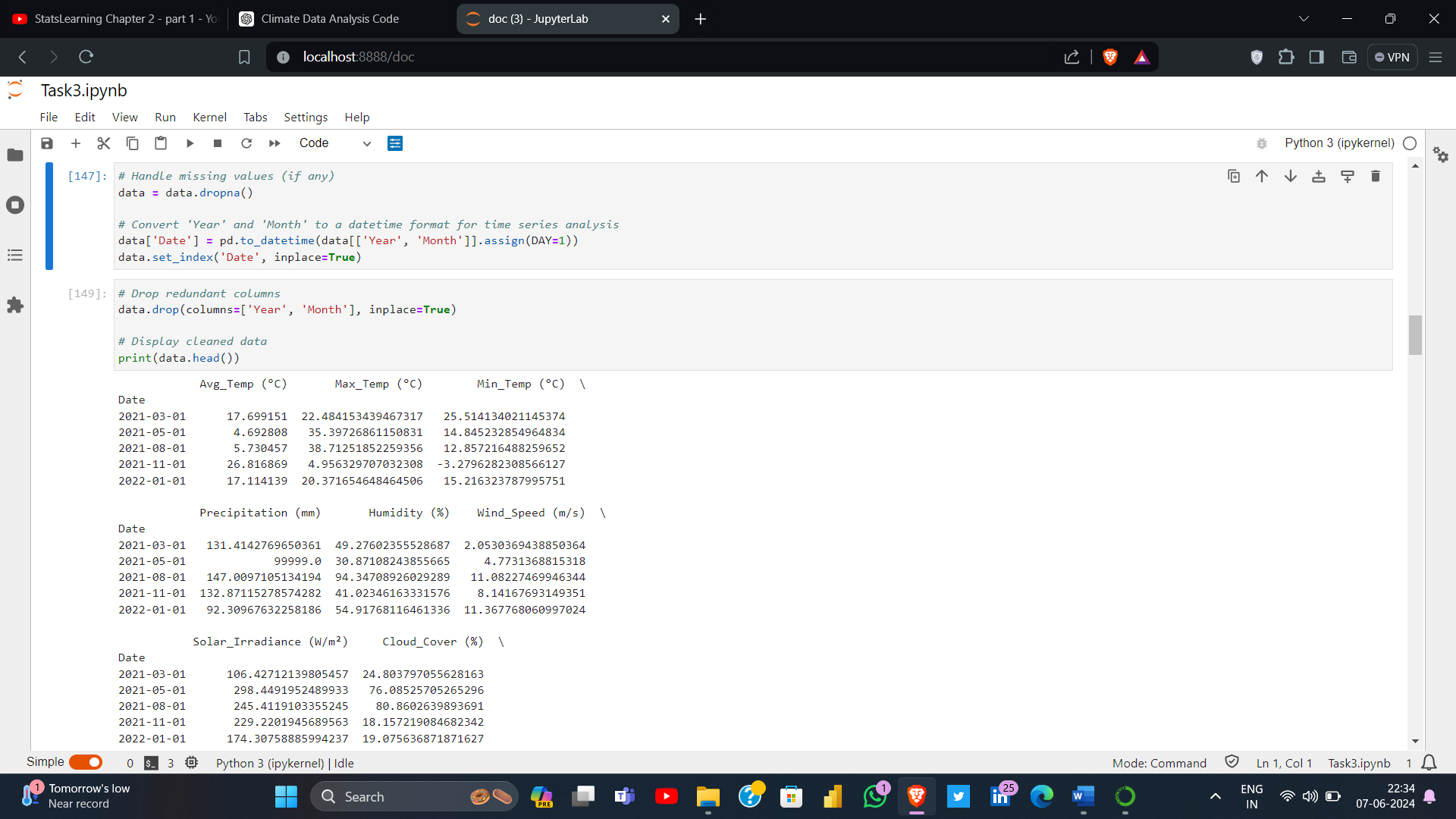
* Segment the data based on Proximity\_to\_Water (km).
* Compare Avg\_Temp (°C), Max\_Temp (°C), Min\_Temp (°C), and Humidity (%) for locations at varying distances from water bodies.
* Use scatter plots, box plots, and heatmaps to visualize the relationship between proximity to water and climatic variables.
* Conduct regression analysis to quantify the impact of proximity to water on temperature and humidity.

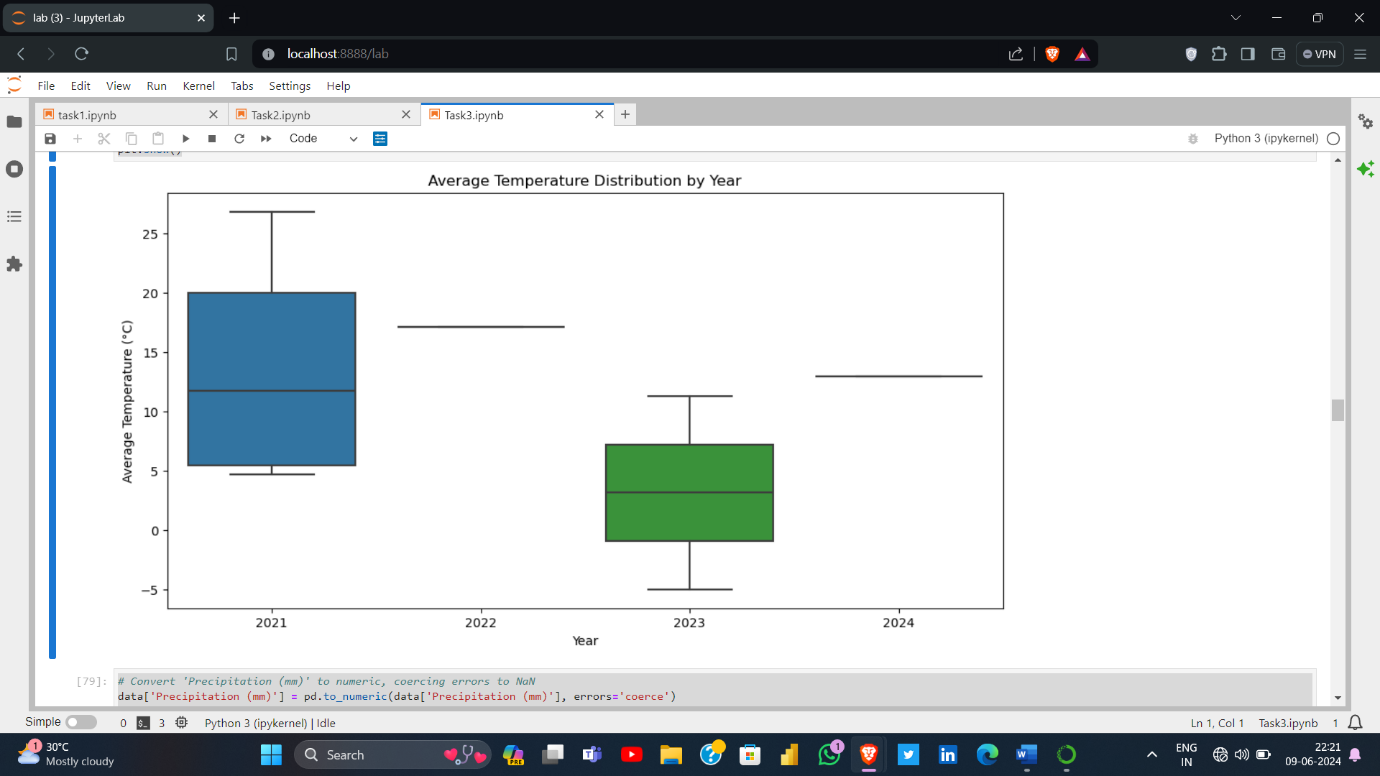
**Question:3**

"What is the relationship between average temperature and CO2 concentration over time, and how does it vary by geographical location?"

**4.** **Analysis Methodology**

The analysis methodology includes:

* Data cleaning: Handling missing values, converting data types, and feature engineering if necessary.
* Visualizations: Using various plotting techniques to analyze the data, including line plots, scatter plots, box plots, bar plots, and heatmaps.

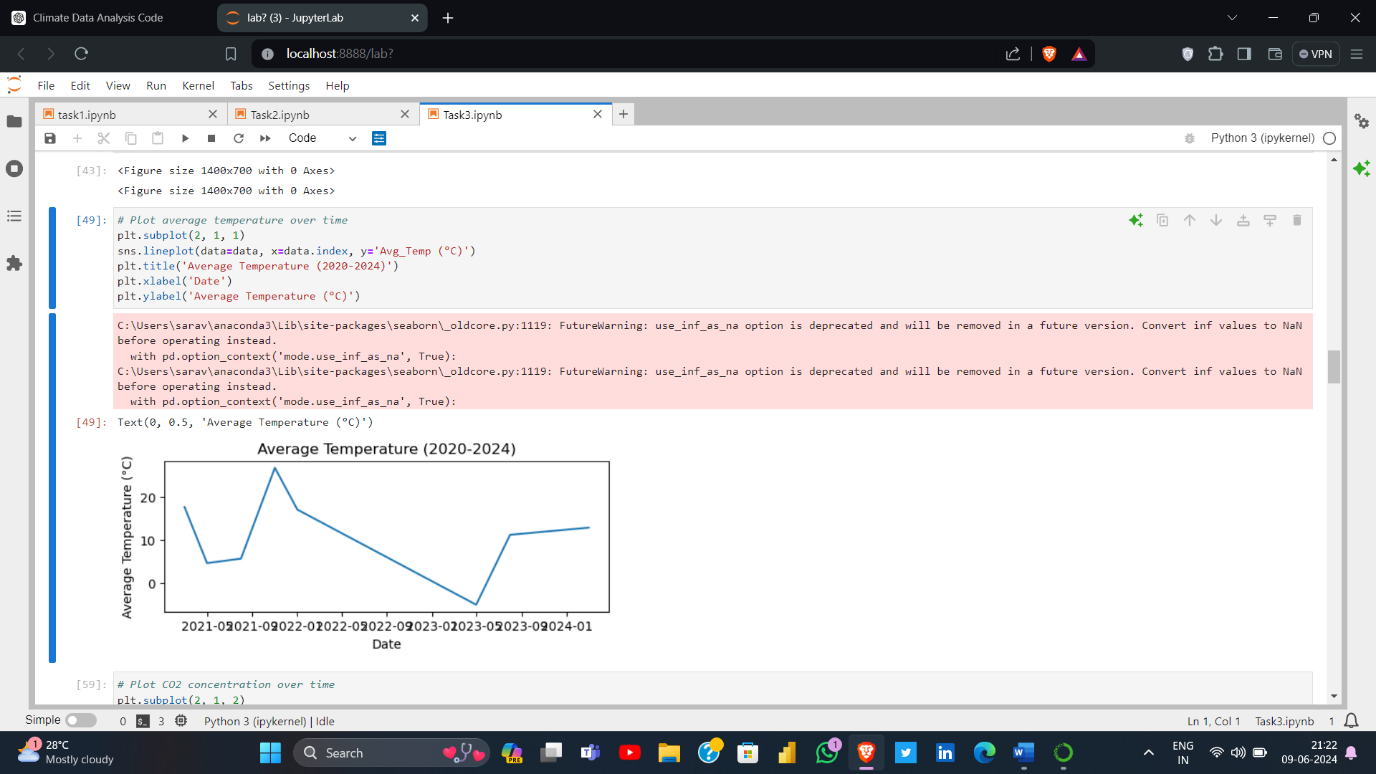


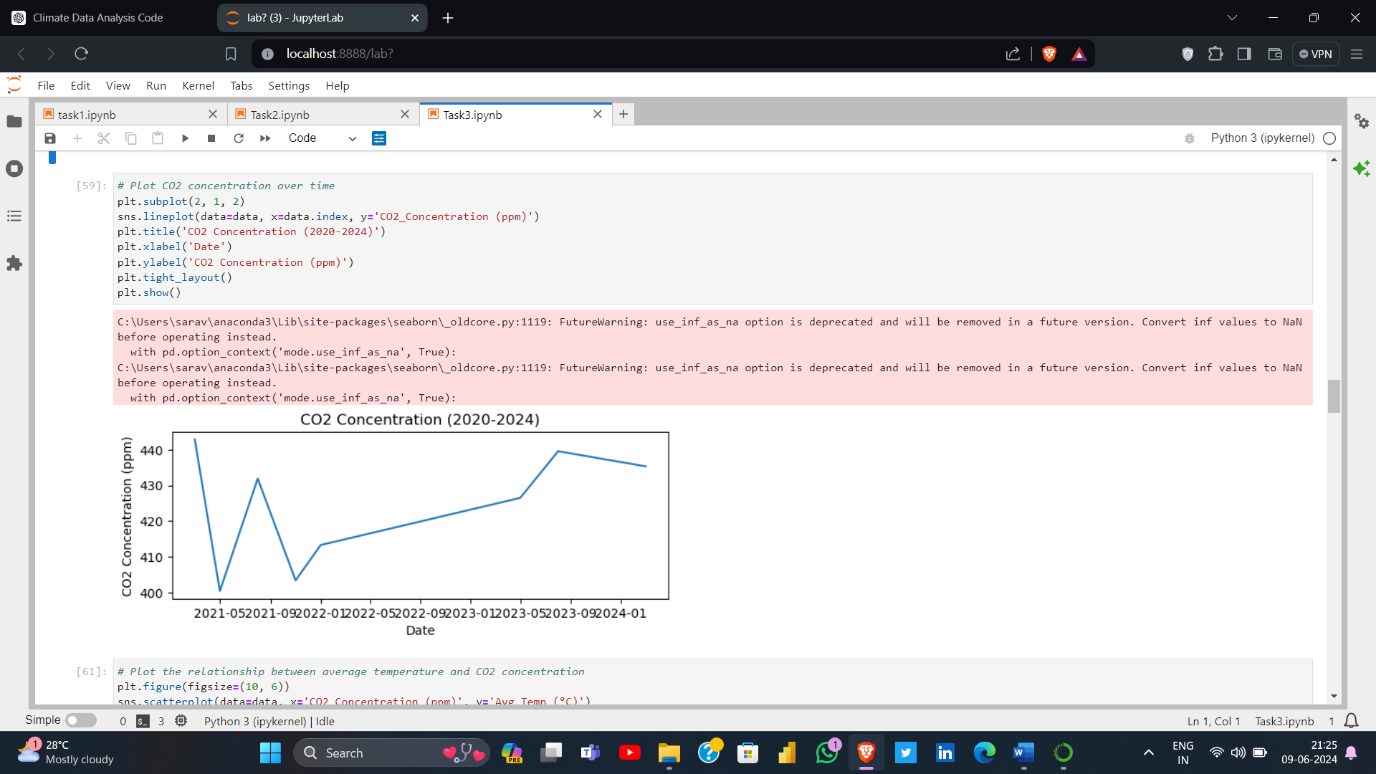
* Statistical analysis: Calculating correlations, trends, and descriptive statistics to derive insights.

**5.** **Results and Findings**

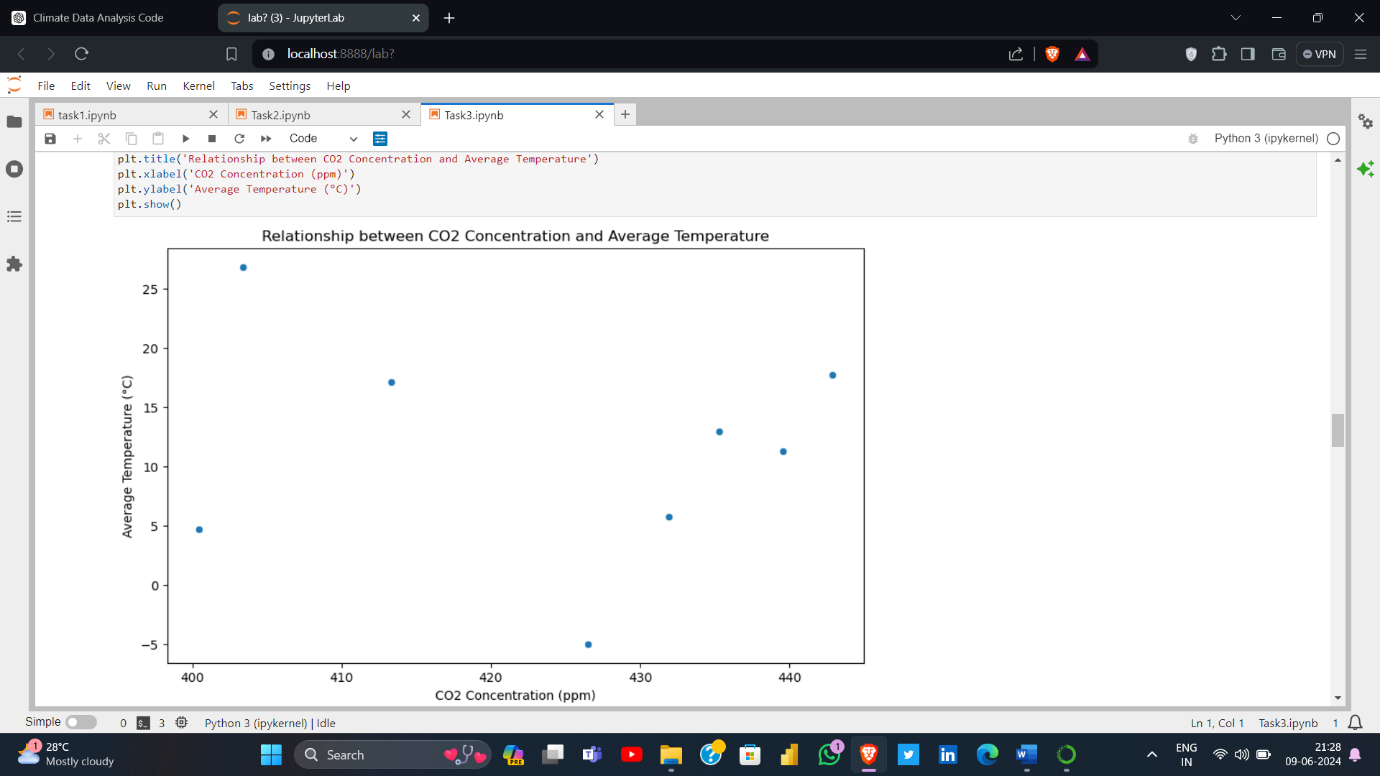
Key findings from the analysis include:

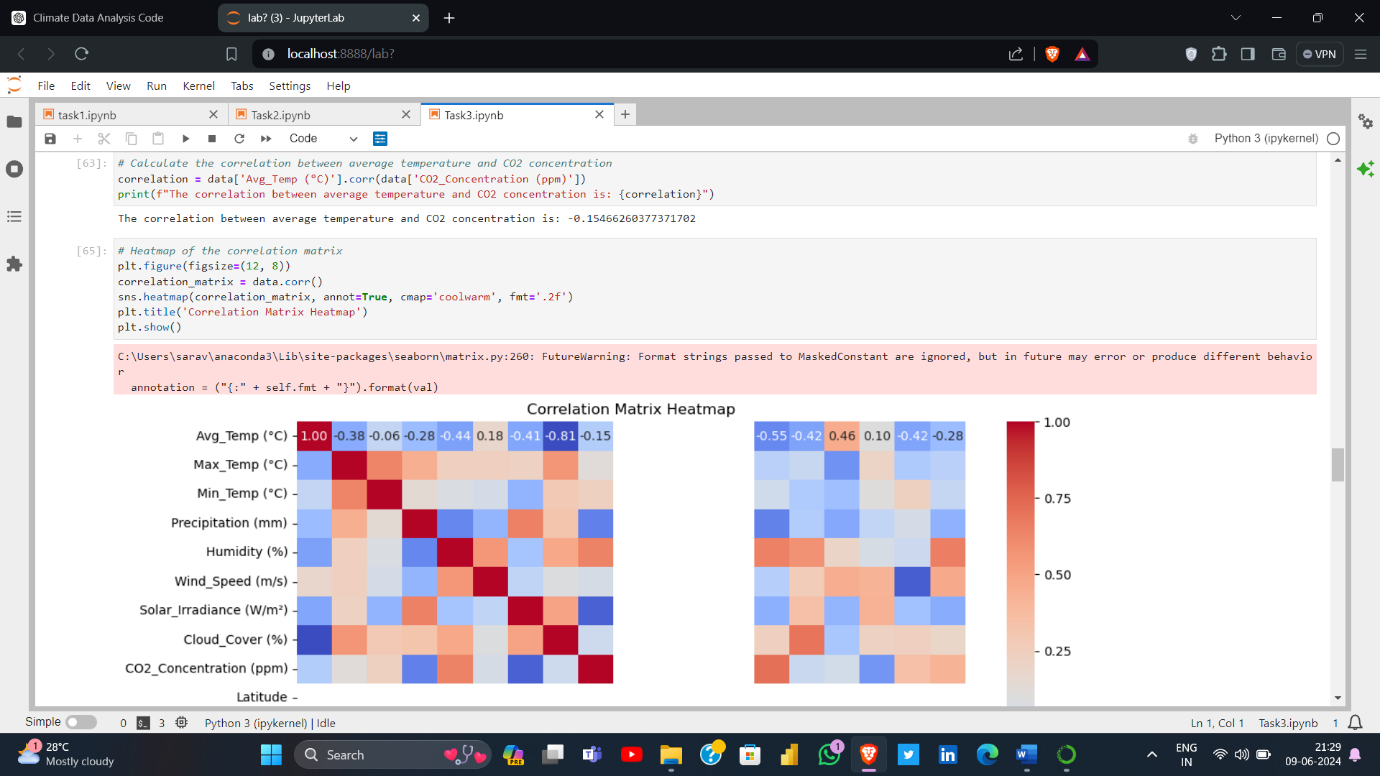
* Trends in average temperature and CO2 concentration over time.



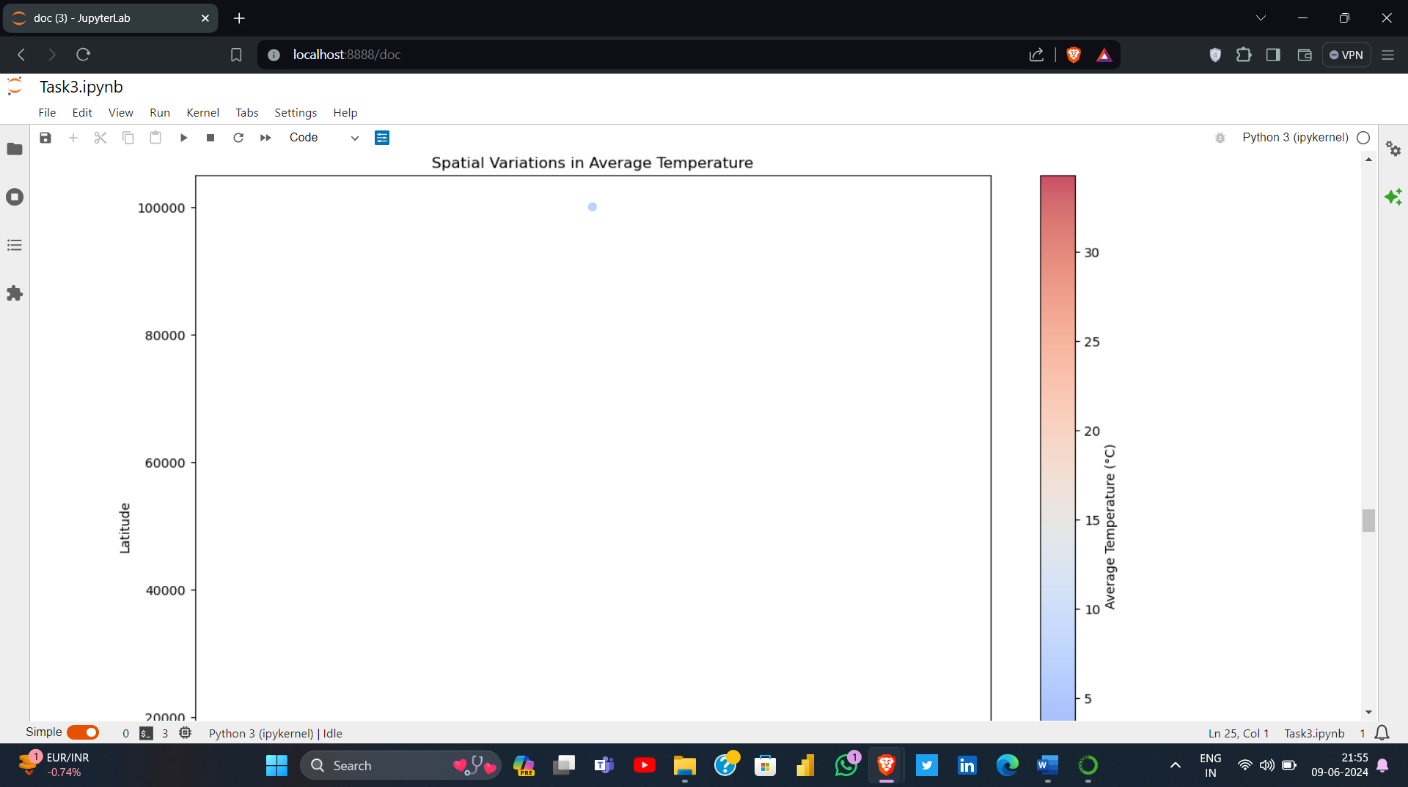
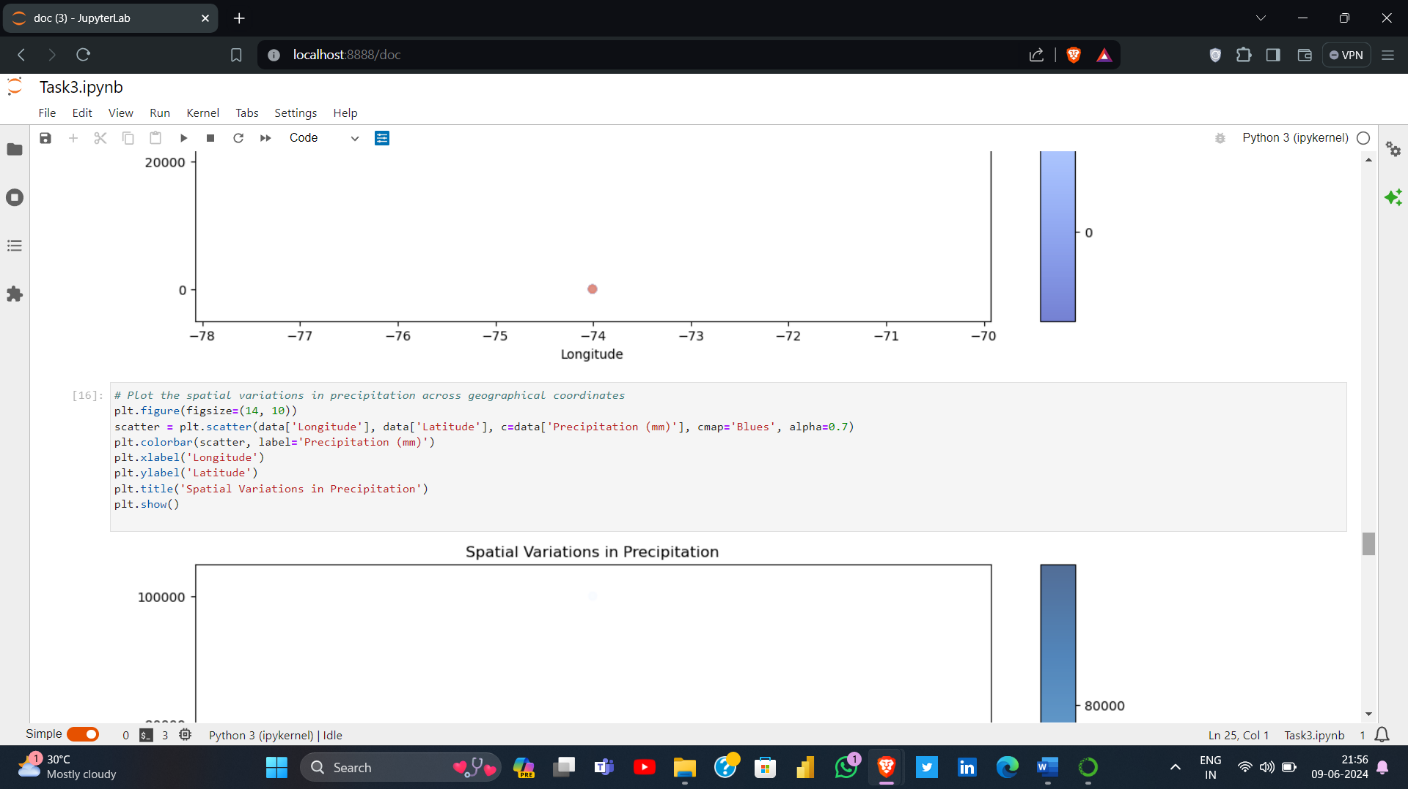


* Correlation between climatic features such as temperature, precipitation, and CO2 concentration.

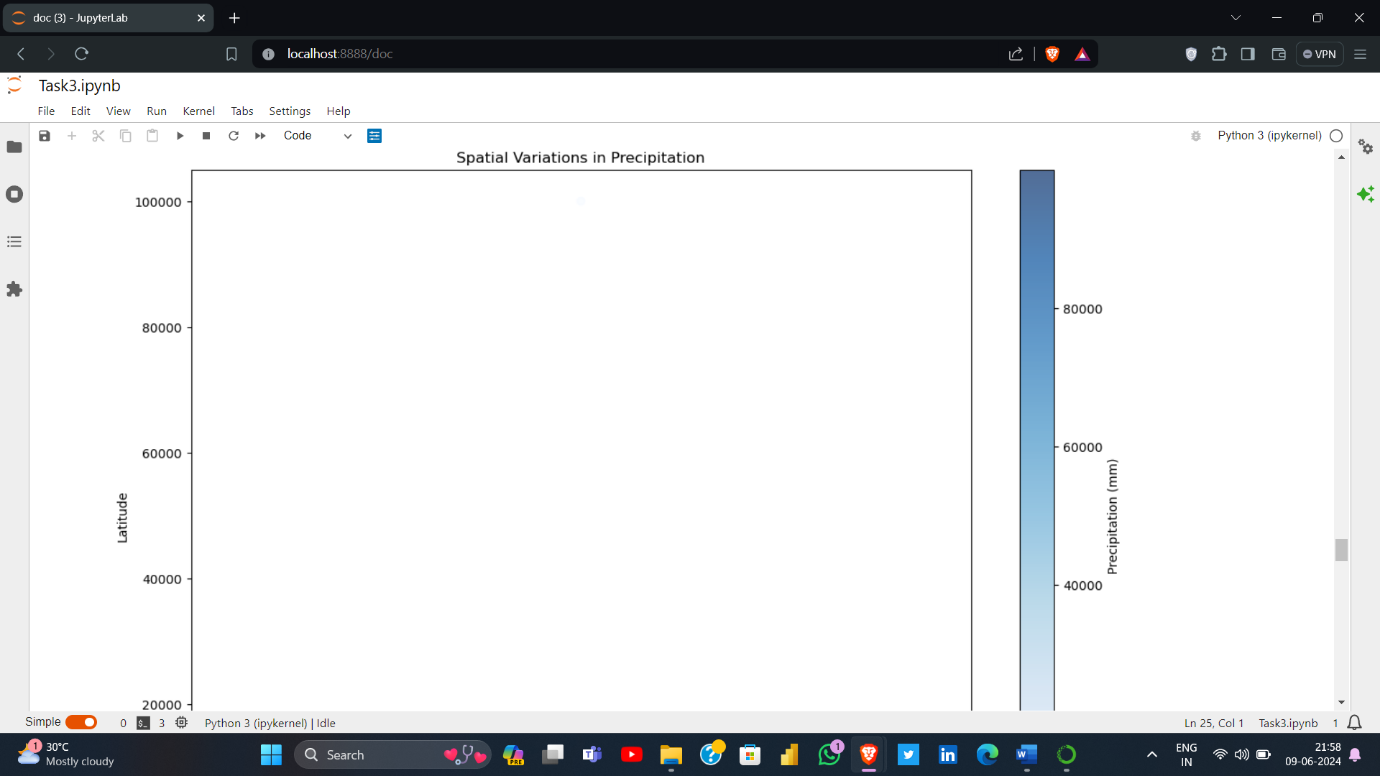
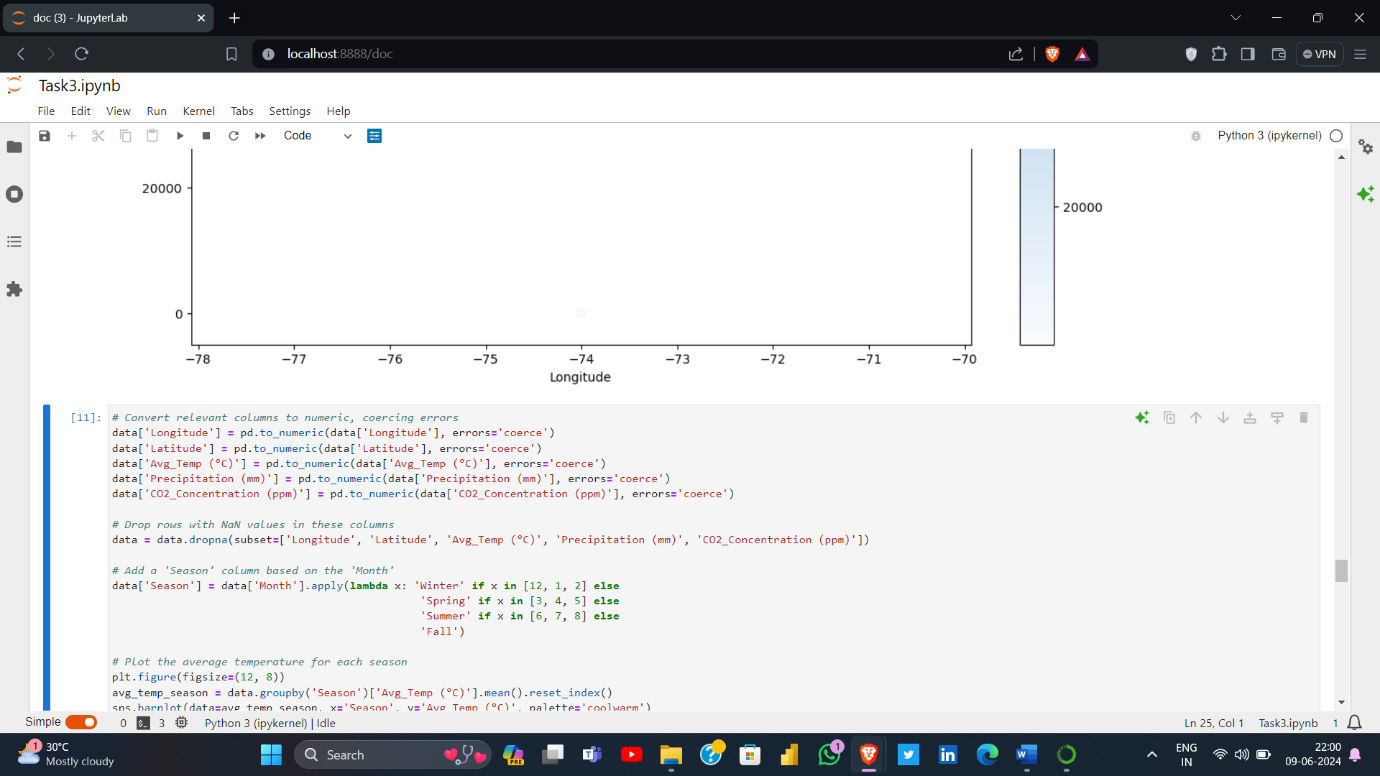




* Spatial variations in climatic patterns based on geographical coordinates.



* Seasonal fluctuations in weather conditions and their impact on environmental factors, Plots the spatial variations in temperature and precipitation



6.Code:

# Import necessary libraries

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

# Load the dataset from local storage

data = pd.read\_csv(r"C:\Users\sarav\OneDrive\Desktop\intern\climate\_change\_dataset.csv")

print(data.head())

print(data.info())

print(data.isnull().sum())

# Summary statistics of the dataset

print(data.describe())

# Print column names to verify

print(data.columns)

# Handle missing values (if any)

data = data.dropna()

# Convert 'Year' and 'Month' to a datetime format for time series analysis

data['Date'] = pd.to\_datetime(data[['Year', 'Month']].assign(DAY=1))

data.set\_index('Date', inplace=True)

# Drop redundant columns

data.drop(columns=['Year', 'Month'], inplace=True)

# Display cleaned data

print(data.head())

# Plot average temperature and CO2 concentration over time

plt.figure(figsize=(14, 7))

# Plot average temperature over time

plt.subplot(2, 1, 1)

sns.lineplot(data=data, x=data.index, y='Avg\_Temp (°C)')

plt.title('Average Temperature (2020-2024)')

plt.xlabel('Date')

plt.ylabel('Average Temperature (°C)')

# Plot CO2 concentration over time

plt.subplot(2, 1, 2)

sns.lineplot(data=data, x=data.index, y='CO2\_Concentration (ppm)')

plt.title('CO2 Concentration (2020-2024)')

plt.xlabel('Date')

plt.ylabel('CO2 Concentration (ppm)')

plt.tight\_layout()

plt.show()

# Plot the relationship between average temperature and CO2 concentration

plt.figure(figsize=(10, 6))

sns.scatterplot(data=data, x='CO2\_Concentration (ppm)', y='Avg\_Temp (°C)')

plt.title('Relationship between CO2 Concentration and Average Temperature')

plt.xlabel('CO2 Concentration (ppm)')

plt.ylabel('Average Temperature (°C)')

plt.show()

# Calculate the correlation between average temperature and CO2 concentration

correlation = data['Avg\_Temp (°C)'].corr(data['CO2\_Concentration (ppm)'])

print(f"The correlation between average temperature and CO2 concentration is: {correlation}")

# Heatmap of the correlation matrix

plt.figure(figsize=(12, 8))

correlation\_matrix = data.corr()

sns.heatmap(correlation\_matrix, annot=True, cmap='coolwarm', fmt='.2f')

plt.title('Correlation Matrix Heatmap')

plt.show()

# Extract year from Date for boxplot

data['Year'] = data.index.year

# Boxplot of average temperature across different years

plt.figure(figsize=(12, 6))

sns.boxplot(x='Year', y='Avg\_Temp (°C)', data=data)

plt.title('Average Temperature Distribution by Year')

plt.xlabel('Year')

plt.ylabel('Average Temperature (°C)')

plt.show()

# Convert 'Precipitation (mm)' to numeric, coercing errors to NaN

data['Precipitation (mm)'] = pd.to\_numeric(data['Precipitation (mm)'], errors='coerce')

# Drop rows with NaN values in 'Precipitation (mm)'

data = data.dropna(subset=['Precipitation (mm)'])

# Bar plot for average Precipitation by Year

plt.figure(figsize=(12, 8))

avg\_precipitation = data.groupby('Year')['Precipitation (mm)'].mean().reset\_index()

sns.barplot(data=avg\_precipitation, x='Year', y='Precipitation (mm)', palette='viridis')

plt.title('Average Precipitation by Year')

plt.xlabel('Year')

plt.ylabel('Average Precipitation (mm)')

plt.show()

# Pair plot for selected features

selected\_features = ['Avg\_Temp (°C)', 'Max\_Temp (°C)', 'Min\_Temp (°C)', 'CO2\_Concentration (ppm)', 'Precipitation (mm)']

sns.pairplot(data[selected\_features])

plt.suptitle('Pair Plot of Selected Features', y=1.02)

plt.show()

# Plot the spatial variations in temperature across geographical coordinates

plt.figure(figsize=(14, 10))

scatter = plt.scatter(data['Longitude'], data['Latitude'], c=data['Avg\_Temp (°C)'], cmap='coolwarm', alpha=0.7)

plt.colorbar(scatter, label='Average Temperature (°C)')

plt.xlabel('Longitude')

plt.ylabel('Latitude')

plt.title('Spatial Variations in Average Temperature')

plt.show()

# Plot the spatial variations in precipitation across geographical coordinates

plt.figure(figsize=(14, 10))

scatter = plt.scatter(data['Longitude'], data['Latitude'], c=data['Precipitation (mm)'], cmap='Blues', alpha=0.7)

plt.colorbar(scatter, label='Precipitation (mm)')

plt.xlabel('Longitude')

plt.ylabel('Latitude')

plt.title('Spatial Variations in Precipitation')

plt.show()

# Plot the average temperature for each season

plt.figure(figsize=(12, 8))

avg\_temp\_season = data.groupby('Season')['Avg\_Temp (°C)'].mean().reset\_index()

sns.barplot(data=avg\_temp\_season, x='Season', y='Avg\_Temp (°C)', palette='coolwarm')

plt.title('Average Temperature by Season')

plt.xlabel('Season')

plt.ylabel('Average Temperature (°C)')

plt.show()

# Plot the average precipitation for each season

plt.figure(figsize=(12, 8))

avg\_precipitation\_season = data.groupby('Season')['Precipitation (mm)'].mean().reset\_index()

sns.barplot(data=avg\_precipitation\_season, x='Season', y='Precipitation (mm)', palette='Blues')

plt.title('Average Precipitation by Season')

plt.xlabel('Season')

plt.ylabel('Precipitation (mm)')

plt.show()

# Plot the impact of seasons on CO2 concentration

plt.figure(figsize=(12, 8))

avg\_co2\_season = data.groupby('Season')['CO2\_Concentration (ppm)'].mean().reset\_index()

sns.barplot(data=avg\_co2\_season, x='Season', y='CO2\_Concentration (ppm)', palette='Greens')

plt.title('Average CO2 Concentration by Season')

plt.xlabel('Season')

plt.ylabel('CO2 Concentration (ppm)')

plt.show()

7. Conclusion

The analysis provides valuable insights into the dynamics of climate change and its impact on various aspects of the environment. By understanding the patterns and trends in climatic data, stakeholders can make informed decisions regarding climate resilience, disaster management, and sustainable development.

8. Future Directions

Future analysis could explore additional datasets or incorporate advanced modeling techniques to predict future climate trends and assess the effectiveness of mitigation strategies.

9. References

Kaggle Dataset: Climate Change Dataset 2020-2024

Python libraries: pandas, numpy, matplotlib, seaborn

This documentation provides an overview of the climate change analysis task, detailing the methodology, findings, and potential avenues for further research.

GITHUB LINK(TASK):

|  |
| --- |
| 1. <https://github.com/Saravana999/Shadow-Fox> 2. <http://localhost:8888/lab/tree/%20climate_chanage_analysis.ipynb> |

**10. Outcome:**

1. **Data Cleaning and Preprocessing**
   * Data Type Conversion: Ensured that the relevant columns are converted to numeric types to facilitate analysis and visualization.
   * Handling Missing Values: Dropped rows with missing values in critical columns to maintain data integrity for analysis.
2. **Spatial Analysis**
   * **Visualization**: Used scatter plots to visualize spatial variations in average temperature and precipitation across geographical coordinates.
   * **Color Mapping**: Applied color mapping to indicate the intensity of the climatic parameter (e.g., temperature or precipitation) at different geographical locations.
3. **Seasonal Analysis**
   * **Season Classification**: Created a 'Season' column based on the 'Month' column to classify data into Winter, Spring, Summer, and Fall.
   * **Bar Plots**: Used bar plots to compare average values of temperature, precipitation, and CO2 concentration across different seasons.
4. **Library and Tools**
   * **Pandas**: For data manipulation and preprocessing.
   * **Matplotlib**: For creating scatter plots to visualize spatial variations.
   * **Seaborn**: For creating bar plots to visualize seasonal variations.

**Knowledge and Skills Gained**

1. **Data Manipulation with Pandas**
   * Efficiently cleaning and preparing datasets for analysis.
   * Handling missing values and converting data types.
   * Grouping data and performing aggregate calculations
2. **Data Visualization with Matplotlib and Seaborn**
   * Creating scatter plots to visualize spatial data.
   * Using color maps to represent data intensity in scatter plots.
   * Designing bar plots to compare average values across different categories (seasons in this case).
3. **Understanding Climatic Patterns**
   * Analyzing spatial variations in climatic parameters to identify geographical patterns.
   * Studying seasonal variations to understand the impact of seasons on environmental factors.
   * Using geographical coordinates (latitude and longitude) to map climatic data.
4. **Project Management and Execution**
   * Defining clear objectives for climatic data analysis.
   * Planning and executing data cleaning, analysis, and visualization steps systematically.
   * Drawing insights and conclusions from the visualized data.

**Potential Insights**

1. **Geographical Insights**
   * Identify regions with extreme temperatures or high precipitation.
   * Understand the spatial distribution of climatic parameters and their potential impact on local environments.
2. **Seasonal Insights**
   * Determine how different seasons affect average temperature and precipitation levels.
   * Analyze the seasonal impact on CO2 concentration, which can have implications for environmental and climate studies.
3. **Environmental Impact**
   * Use insights to study the impact of climatic patterns on urbanization, vegetation, and other environmental indices.
   * Provide data-driven recommendations for climate resilience and environmental planning.