DAV REPORT
(22ADE01)

On
Weather data analysis

Submitted in partial fulfilment for the completion of

BE-IV Semester

in

INFORMATION TECHNOLOGY

By

160122737150- Sudha Jennifer

160122737151 – Sowmya Sri

160122737164 – Jyothika

Under the guidance of

Dr. N Sudhakar Yadav Sir
Assistant Professor,
Dept. of IT

DEPARTMENT OF INFORMATION TECHNOLOGY

CHAITANYA BHARATHI INSTITUTE OF TECHNOLOGY
(Autonomous)
(Affiliated to Osmania University; Accredited by NBA(AICTE) and NAAC(UGC), ISO Certified 9001:2015)
GANDIPET, HYDERABAD – 500075
Website: www.cbit.ac.in

2023-2024

CERTIFICATE

This is to certify that the project work entitled "**WEATHER DATA ANALYSIS**" submitted to **CHAITANYA BHARATHI INSTITUTE OF TECHNOLOGY,** in partial fulfilment of the requirements for the award of the completion of DAV project (22ADE01) IV semester of B. E in Information Technology, during the academic year 2023-2024, is a record of original work done by **Sudha Jennifer(160122737150), Sowmya Sri(160122737151) and Jyothika(160122737164)** during the period of study in Department of IT, CBIT, HYDERABAD, under our supervision and guidance.

Project Guide
Dr. N. Sudhakar Yadav
Asst. Professor, Dept. of IT,
CBIT, Hyderabad.

ABSTRACT

Weather data analysis plays a crucial role in understanding climate patterns, forecasting, and making informed decisions in various sectors such as agriculture, transportation, and disaster management. This abstract presents a comprehensive overview of methodologies and insights gained from analyzing weather data.

The analysis begins with data collection from diverse sources, including weather stations, satellites, and remote sensors, encompassing variables like temperature, precipitation, humidity, wind speed, and atmospheric pressure. Utilizing statistical techniques, machine learning algorithms, and data visualization tools, we explore temporal and spatial trends, seasonal variations, and extreme weather events.

Key findings reveal long-term climate trends, such as global warming and shifting precipitation patterns, along with regional disparities and anomalies. Seasonal analysis uncovers cyclical patterns and their impact on ecosystems, agriculture, and water resources management. Furthermore, detecting extreme weather events, such as hurricanes, heatwaves, and droughts, aids in assessing their frequency, intensity, and societal implications.

Moreover, correlation analysis elucidates relationships between different weather variables and their influence on phenomena like El Niño Southern Oscillation (ENSO) and Arctic Oscillation. These insights contribute to improved weather forecasting models, early warning systems, and climate adaptation strategies.

OBJECTIVES AND OUTCOMES

1. **Data Exploration and Information Retrieval:** The code aims to load weather data from a CSV file and perform basic exploratory data analysis (EDA) tasks to gain insights into various weather parameters such as temperature, humidity, wind speed, and atmospheric pressure.
2. **Data Filtering and Subsetting:** It involves filtering and subsetting the data based on specific conditions, such as retrieving records for a particular weather condition ('Clear'), specific wind speed, or date/time range (e.g., January 6th).
3. **Categorization of Climate Factors:** The code categorizes climate factors based on temperature, wind speed, and atmospheric pressure, aiming to understand the primary factors contributing to climate variability and change.
4. **Visualization of Climate Trends:** It visualizes climate trends over time, such as the variation in temperature, humidity, and other climate factors, using line plots, scatter plots, and histograms to provide a clear understanding of the data distribution and trends.

OUTCOMES

1. **Insights into Weather Patterns:** By analyzing the data, the code provides insights into various weather patterns, including temperature variations, humidity levels, wind speeds, and atmospheric pressure changes over time.
2. **Identification of Extreme Weather Events:** The code enables the identification and analysis of extreme weather events, such as hurricanes, heatwaves, and storms, through data filtering and visualization techniques.
3. **Understanding Climate Change Impacts:** By categorizing climate factors and visualizing their trends over time, the code helps in understanding the impacts of climate change, such as increasing temperatures, changing precipitation patterns, and variations in atmospheric pressure.
4. **Enhanced Decision Making:** The insights gained from the data analysis support informed decision-making processes in sectors such as agriculture, transportation, disaster management, and urban planning by providing valuable information

DATASET COLLECTION AND PREPARATION

Data Collection: The first step in weather data analysis is collecting relevant datasets from reliable sources. These sources may include government agencies, meteorological organizations, research institutions, or commercial weather data providers. The dataset should encompass a wide range of weather parameters such as temperature, humidity, wind speed, precipitation, atmospheric pressure, and visibility, recorded at regular intervals (e.g., hourly, daily) over a significant time period.

Data Quality Check: Once the data is collected, it's essential to perform a thorough quality check to ensure data integrity and accuracy. This involves checking for missing values, inconsistencies, outliers, and errors in the dataset. Missing values may be imputed or interpolated based on neighbouring values or using statistical techniques.

Data Cleaning and Preprocessing: Data cleaning involves removing duplicates, outliers, and irrelevant columns from the dataset. Additionally, data may need to be transformed or standardized for consistency and compatibility with analysis tools and algorithms. For example, categorical variables may be encoded into numerical format, and timestamps may be converted into datetime objects.

Exploratory Data Analysis (EDA): EDA is performed to gain insights into the dataset's characteristics, distributions, correlations, and trends. This involves generating summary statistics, visualizations (e.g., histograms, scatter plots, time series plots), and exploring relationships between different weather parameters.

Data Partitioning: Depending on the analysis goals, the dataset may be partitioned into training, validation, and test sets. The training set is used to train the model, the validation set is used to tune hyperparameters and evaluate model performance during training, and the test set is used to evaluate the final model's performance.

Data Integration : In some cases, additional datasets such as geographical data, land use data, or socioeconomic data may be integrated with weather data to enrich the analysis and provide more context for understanding the impacts of weather on various phenomena.

CODING

```python
import numpy as np
import pandas as pd
data = pd.read_csv(r"Weather Data.csv")
data
data.info()
data.head()
data['Weather'].unique()
data['Weather'].value_counts()
data[(data.Weather=='Clear')].count()[1]
data[data["Wind Speed_km/h"]==4]
data['Wind Speed_km/h'].unique()
data[['Weather','Temp_C']].head()
data['Date/Time'] = pd.to_datetime(data['Date/Time'])
jan_6_data = data[data['Date/Time'].dt.month == 1]
jan_6_data = jan_6_data[jan_6_data['Date/Time'].dt.day == 6]
first_5_pressure_values = jan_6_data['Press_kPa'].head(5)
print(first_5_pressure_values)
data[(data['Wind Speed_km/h']>30) & (data['Visibility_km']==25)]
import pandas as pd
data = pd.read_csv('Weather Data.csv')
def categorize_climate_factors(row):
factors = []
# Check temperature
if row['Temp_C'] > 25:
factors.append('High temperature')
```

```python
    elif row['Temp_C'] < 5:
        factors.append('Low temperature')
    # Check wind speed
    if row['Wind Speed_km/h'] > 20:
        factors.append('High wind speed')
    # Check pressure
    if row['Press_kPa'] < 95:
        factors.append('Low pressure')
    return factors
# Apply the function to each row of the DataFrame to categorize climate factors
data['Climate_Factors'] = data.apply(categorize_climate_factors, axis=1)
# Display the DataFrame with categorized climate factors
print(data[['Date/Time', 'Climate_Factors']])
import pandas as pd
import matplotlib.pyplot as plt# Load the dataset into a DataFrameimport pandas as pd
import matplotlib.pyplot as plt
data = pd.read_csv('Weather Data.csv')
# Define a function to categorize the primary factors contributing to climate change
def categorize_climate_factors(row):
    factors = []
    # Check temperature
    if row['Temp_C'] > 25:
        factors.append('High temperature')
    elif row['Temp_C'] < 5:
        factors.append('Low temperature')
    # Check wind speed
    if row['Wind Speed_km/h'] > 20:
        factors.append('High wind speed')
    # Check pressure
    if row['Press_kPa'] < 95:
```

```python
        factors.append('Low pressure')
    return factors
# Apply the function to each row of the DataFrame to categorize climate factors
data['Climate_Factors'] = data.apply(categorize_climate_factors, axis=1)
# Convert Date/Time column to datetime type
data['Date/Time'] = pd.to_datetime(data['Date/Time'])
# Create a figure and axis
fig, ax = plt.subplots(figsize=(10, 6))
# Plot each category of climate factor over time
for factor in data['Climate_Factors'].explode().unique():
    factor_data = data[data['Climate_Factors'].apply(lambda x: factor in x)]
    ax.plot(factor_data['Date/Time'], [factor]*len(factor_data), linestyle='', marker='o',
label=factor)
# Set labels and title
ax.set_xlabel('Date/Time')
ax.set_ylabel('Climate Factors')
ax.set_title('Categorized Climate Factors Over Time')
ax.legend()
# Rotate x-axis labels for better readability
plt.xticks(rotation=45)
# Show plot
plt.tight_layout()#gives padding between plot elements
plt.show()
#timeseries plot
import pandas as pd
import matplotlib.pyplot as plt


# Load the weather dataset into a DataFrame
data = pd.read_csv('Weather Data.csv')


# Convert Date/Time column to datetime type
```

```python
data['Date/Time'] = pd.to_datetime(data['Date/Time'])

# Extract year and month from Date/Time
data['Year'] = data['Date/Time'].dt.year
data['Month'] = data['Date/Time'].dt.month

# Group by year and month, calculate average temperature
monthly_temp_avg = data.groupby(['Year', 'Month'])['Temp_C'].mean()

# Reset index to access 'Year' and 'Month' as columns
monthly_temp_avg = monthly_temp_avg.reset_index()

# Plotting
plt.figure(figsize=(10, 6))
for year in monthly_temp_avg['Year'].unique():
    year_data = monthly_temp_avg[monthly_temp_avg['Year'] == year]
    plt.plot(year_data['Month'], year_data['Temp_C'], label=year, marker='o')

plt.title('Monthly Average Temperature Variation')
plt.xlabel('Month')
plt.ylabel('Average Temperature (°C)')
plt.xticks(range(1, 13), ['Jan', 'Feb', 'Mar', 'Apr', 'May', 'Jun', 'Jul', 'Aug', 'Sep', 'Oct', 'Nov', 'Dec'])
plt.legend(title='Year')
plt.grid(True)
plt.show()
import pandas as pd
import matplotlib.pyplot as plt
# Load the weather dataset into a DataFrame
data = pd.read_csv('Weather Data.csv')
# Convert Date/Time column to datetime type
```

```
data['Date/Time'] = pd.to_datetime(data['Date/Time'])
# Plotting temperature and humidity over time
plt.figure(figsize=(10, 6))
plt.plot(data['Date/Time'], data['Temp_C'], label='Temperature (°C)', color='red')
plt.plot(data['Date/Time'], data['Rel Hum_%'], label='Relative Humidity (%)', color='blue')
plt.xlabel('Date/Time')
plt.ylabel('Value')
plt.title('Temperature and Humidity Variation Over Time')
plt.legend()
plt.grid(True)
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```

RESULT VISUALIZATION, ANALYSIS AND CONCLUSION

Weather forecasting is defined as prediction of atmospheric conditions like air

temperature, humidity, sky conditions, air pressure and general circulation of the atmosphere of a particular place or a region using scientific tools and technological knowledge. During ancient times, much of the forecasting was done in a crude manner like sky colour, wind direction, cloud colour and its cover, lightening, thunder, behaviour of some animals and birds, certain folklores etc. Systematic weather data recording started with the advent of instruments like thermometer, barometer, telegraph, radiotelegraphy, radar technology and finally satellites namely sun synchronous and geosynchronous satellites.

Weather forecasting involves some well-defined steps like recording, collecting, transmitting, compilation, plotting, analyzing and then the final forecasting of the weather related information. The tools used in weather forecasting are thermometers, barometers, hygrometers, rain gauge, radiosondes, rawinsondes, aeroplanes, satellites and radars etc.

Weather forecasting is done for different temporal scales which has different uses and have different parameters for accuracy like long range weather forecast, medium range weather forecast, short range weather forecast and nowcast. There are several different methods used for weather forecasting. Some of the well-defined forecasting methods used by climatologists are persistence method, trends method, climatology method, analog method, numerical weather prediction method.
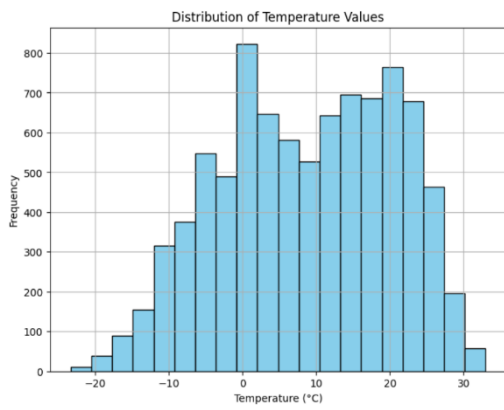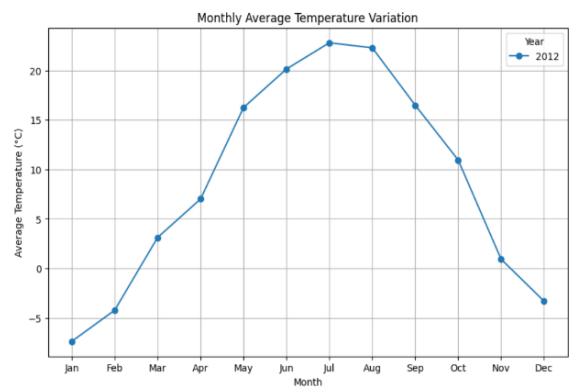
Fig 1. Distribution of Temperature Values

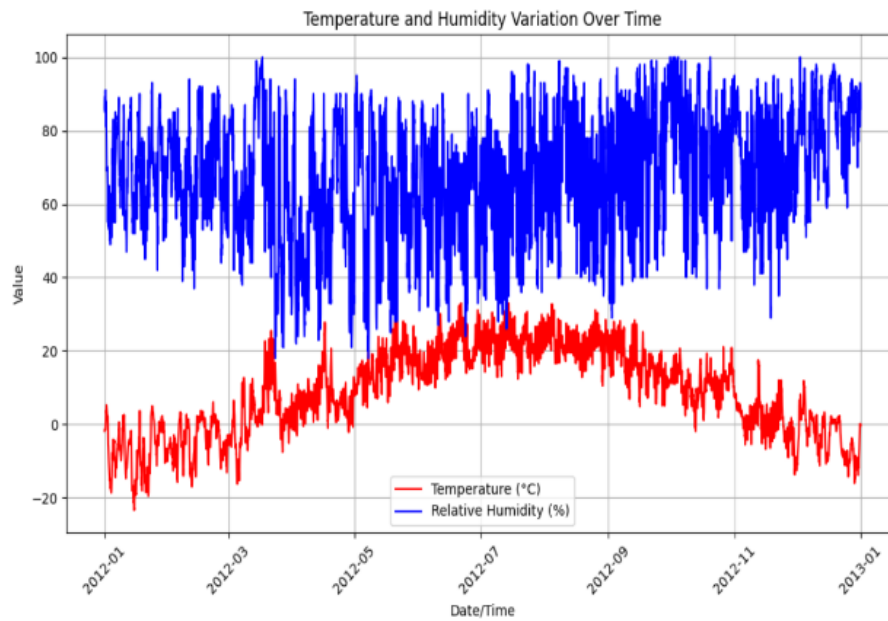**Fig 2.** Monthly Average Temperature Variation

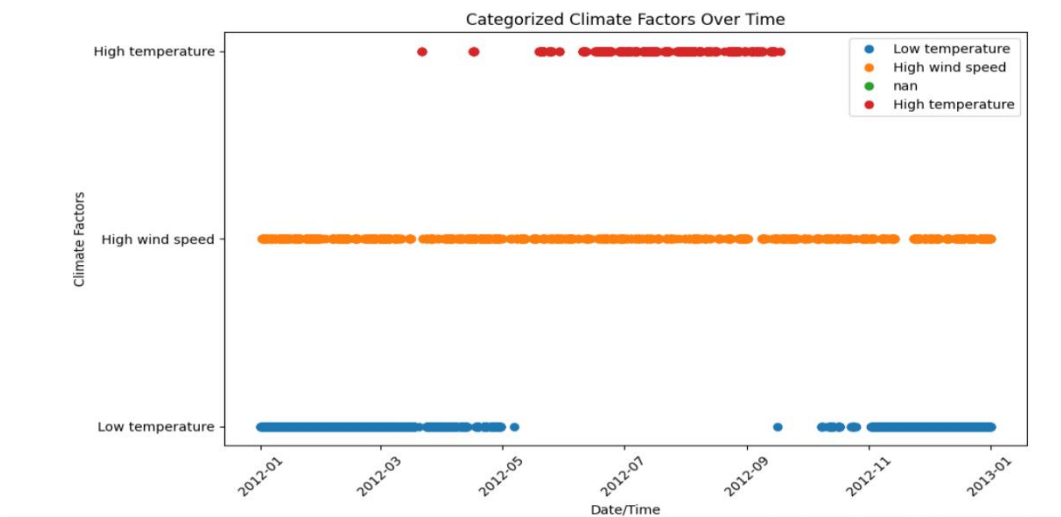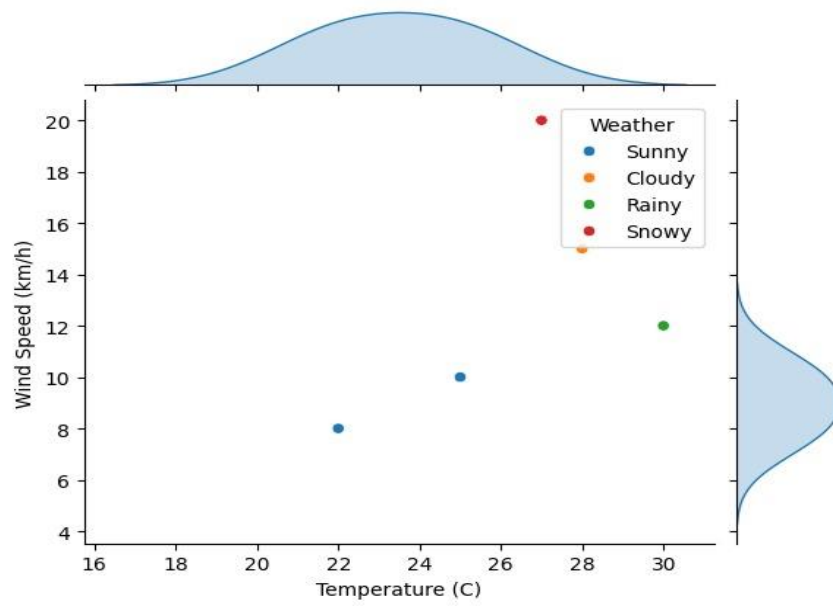**Fig 3**. Temperature and Humidity Variation Over Time

Fig 4. Categorized climate over Time

Joint Plot of Temperature vs Wind Speed by Weather

Fig 5. Temperature over Wind Speed

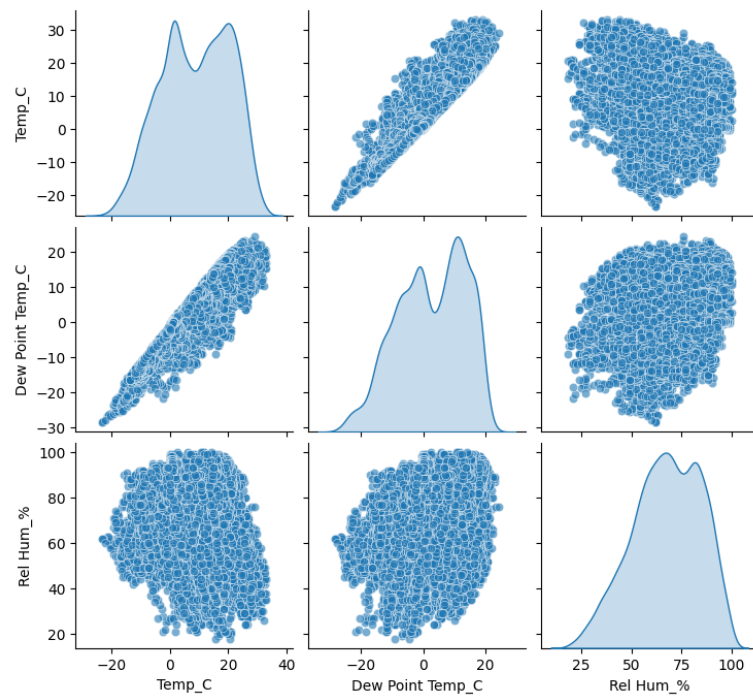Pair Plot of Temperature, Dew Point Temperature, and Relative Humidity

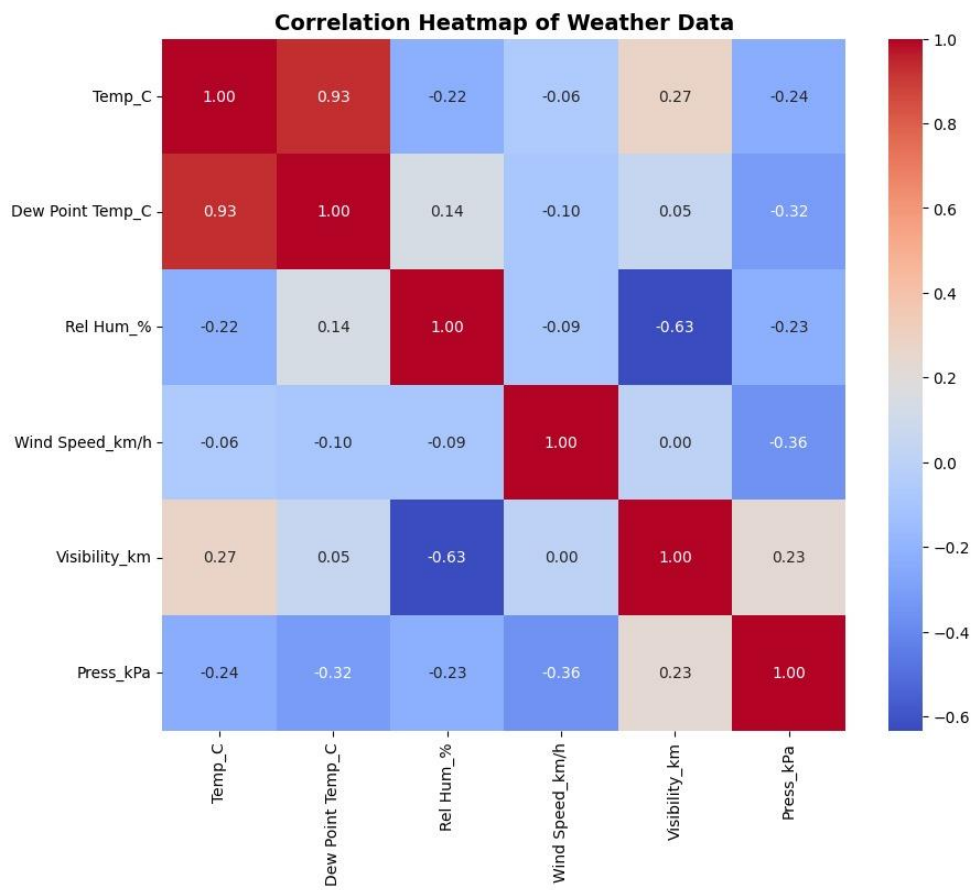**Fig 6**. Pair plot graph to visualize Temperature, Dew Point Temperature and Relative Humidity

Fig 7. Correlation heatmap of Weather Data using Seaborn