**Weather Data Analysis Project Documentation**

**Introduction**

This project aims to analyze weather data using Python and various libraries such as pandas, seaborn, and matplotlib. The project involves loading weather data from a CSV file, handling missing values, exploring the data, performing correlation analysis, and visualizing the data using different plots and charts.

**Code Overview**

The project code is divided into several sections, each covering a specific aspect of the analysis. Here's an overview of the code sections:

1. Importing Libraries and Loading Data: Importing the required libraries (pandas, seaborn, matplotlib) and loading the weather data from the 'Weather\_data.csv' file into a pandas DataFrame.
2. Handling Missing Data: Checking for missing values in the DataFrame, removing columns with missing data, and filling missing values using appropriate strategies.
3. Data Exploration: Exploring the dimensions of the DataFrame, viewing the first few rows, checking column names, data types, and generating summary statistics.

1. Data Transformation: Converting data types, renaming column names to improve clarity.
2. Correlation Analysis: Calculating correlation coefficients between variables, creating scatter plots, scatter matrix, and correlation matrix to examine relationships.
3. Data Visualization: Creating various types of plots and charts such as scatter plots, line plots, bar plots, box plots, histograms, scatter matrix, violin plots, line plots with confidence intervals, pair plots, boxen plots, scatter plots with color mapping, stacked area plots, etc.
4. Time-Based Analysis: Extracting time components from datetime data and performing time-based analysis such as box plots for temperature variation by time of the day.

**Key Takeaways from plots visualization**

* **Scatter plot of 'temperature' and 'humidity'**

If the points are tightly clustered around a line, it suggests a strong relationship between temperature and humidity.

On the other hand, if the points are scattered randomly, it indicates a weak or no correlation.

**Possible scenarios:**

Positive correlation: If the data points on the scatter plot exhibit a general upward trend from left to right, it indicates a positive correlation between temperature and humidity.

Negative correlation: If the data points show a general downward trend from left to right, it indicates a negative correlation between temperature and humidity.( In this case, as temperature increases, humidity decreases.) plotted trend in chart.

No correlation: If the data points appear scattered and there is no clear trend or pattern, it suggests that temperature and humidity are not strongly related.

• **Correlation matrix of heatmap**

Correlation matrix, which is a square matrix where each cell represents the correlation between two variables.

values range from -1 to 1.

-1 indicates a strong negative correlation

0 indicates no correlation

1 indicates a strong positive correlation.

Intensity of the colour indicates the strength of the correlation, with darker shades representing stronger correlations. Warmer colours (e.g., red) represent positive correlations, while cooler colours (e.g., blue) represent negative correlations.

* **Scatter plot of 'temp' and 'feelslike'(temp)**

'feelslike'(temp): Perceived temperature experienced by humans, which takes into account factors such as humidity, wind speed, and other environmental conditions.

Positive Correlation: Data points on the scatter plot show a clear, upward-sloping trend, it indicates a strong positive correlation between temperature and "feels like" temperature.

This suggests that as the actual temperature increases, the perceived or "feels like" temperature also tends to increase.

* **Line plot of 'datetime' and 'temp'**

Line plot shows the temperature trend over time. Plot indicates fluctuations in temperature values over the specified time period.

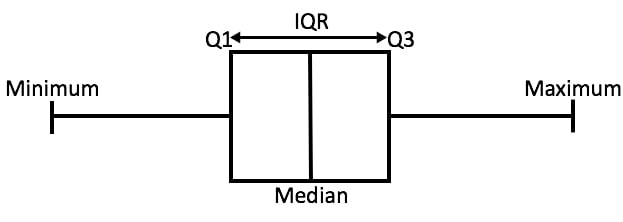
* ***Bar plot of 'conditions' and 'precip'***

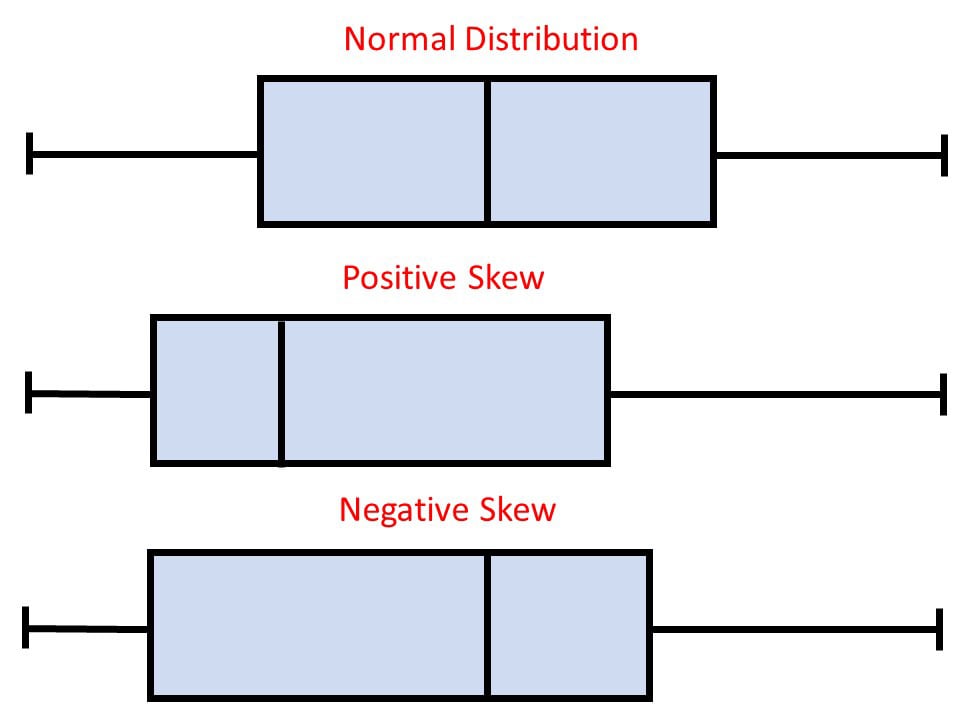
The bar plot shows distribution of precipitation levels for each condition and allows for easy comparison between conditions.

* ***Box plot of 'conditions' and 'temp'(*** useful when dealing with multiple categories or groups)

Box:

Medians: It indicates central tendency of the data





Skewness: The position of the median within the box can indicate the skewness of the data

Symmetric: Whiskers are the same on both sides of the box and median is in the middle of the box.

Positively skewed: Whisker is shorter on the lower end of the box and median is closer to the bottom of the box.

Negatively skewed: Whisker is shorter on the upper end of the box and median is closer to the top of the box.

Quartiles (Q1- represents 25th percentile, Median(Q2), Q3- represents 75th percentile.)

Whisker ranges: minimum and maximum values of the dataset are represented by lines

Outliers: datapoints beyond whiskers

* ***Histogram of 'temp'***

The shape of the histogram indicates the pattern of the temperature distribution. A higher frequency in a particular bin suggests that more data points fall within that temperature range.

* **Scatter Matrix: Visualizing pairwise relationships between multiple variables.**

Scatter matrix or pairwise plot, is a grid of scatter plots that visualizes the relationships between pairs of selected variables in a dataset.

Pairwise scatter plots are valuable in understanding the relationships and dependencies between variables.

Observed patterns: upward-sloping pattern

downward-sloping pattern

No Relationship(datapoints scattered randomly)

Strength of Relationship: A tight cluster of points suggests a strong relationship, while a more scattered distribution implies a weak or no relationship.

* **Violin plot of 'conditions' and 'temp'**

Violin plot provides insights into the distribution of the 'temp' variable across different categories of the 'conditions' variable.

The width of each violin represents the density or frequency of temperature values, with wider parts indicating higher density (at particular temp).

* **Boxen plot of 'conditions' and 'temp'**

Similar to box plots with more detailed information about the data distribution using Tiers.

Whiskers: The length of the whiskers shows the range of the data. Width of the whiskers in boxen plots varies based on the density of the data points.

**Conclusion**

The Weather Data Analysis project demonstrates the application of Python and data analysis libraries for exploring and visualizing weather data. By following the provided code and documentation, users can gain insights into the weather patterns and trends. The project can be further extended by incorporating additional analysis techniques, exploring different visualizations, or applying machine learning algorithms for weather forecasting.