### **Dataset Information**

* Dataset File: Weather Records.xlsx
* Prediction Column: Precipitation(in)
* Dataset Columns: 13 columns, including the prediction column.

### **Tasks**

#### **Task A: Data Cleaning and Analysis (50 Marks)**

1. Handling Missing Values:
   * Fill NaN values using information from neighboring rows or columns. Only drop NaN values if they cannot be filled.
   * List the columns for which you dropped NaN values and provide a one-line reason for each column.
2. Feature Engineering for Date Column:
   * Create new features from the StartTime(UTC) column, such as:
     + Hour of the day
     + Day of the week
     + Month
   * Explain why these features might be useful for analysis.

#### **Task B: Visualization (10 Marks)**

* Perform one visualization for each column in the dataset.
* Use appropriate visualization types (e.g., bar chart, histogram, scatter plot, etc.).
* Write a short explanation for each visualization, highlighting key insights.

#### **Task C: Advanced Feature Engineering (20 Marks)**

1. Create New Features:
   * Add a Duration feature that calculates the time difference (in hours) between StartTime(UTC) and EndTime(UTC).
   * Add a Season feature based on the month extracted from StartTime(UTC).
2. Encoding Categorical Variables:
   * Convert the Severity column into numerical values using label encoding or one-hot encoding. Explain your choice of method.

#### **Task D: Advanced Analysis (20 Marks)**

1. Correlation Analysis:
   * Calculate the correlation between Precipitation(in) and the Duration feature.
   * Visualize the correlation using a heatmap.
   * Write a short explanation of the insights derived from the correlation analysis.
2. Outlier Detection:
   * Identify outliers in the Precipitation(in) column using the IQR (Interquartile Range) method.
   * Visualize the outliers using a boxplot.
   * Explain how you would handle these outliers.
3. Geospatial Analysis:
   * Plot the locations of all weather events on a map using the LocationLat and LocationLng columns.
   * Use a scatter plot for this task.
   * Write a short analysis of the geographical distribution of weather events.

### **Additional Notes**

* Ensure your code is well-commented and easy to follow.
* Submit your Jupyter Notebook along with the generated visualizations.

### **Good Luck!**