Data Report: Analysis of Weather Trends and Patterns for Prediction

By: Almas Ali Pinto (23123639)

Research Question

How do weather patterns differ across regions, and what insights can improve prediction accuracy?

Objectives

1. Analyze trends in weather data (temperature, precipitation, etc.).
2. Compare weather patterns across geographical locations.
3. Provide actionable insights to improve weather forecasting accuracy.

Datasets

1. Seattle Weather Prediction Dataset

- **Why Chosen:**
  Contains localized weather data for Seattle, ideal for understanding trends and patterns in a specific region.
- **Source:** Kaggle
- **URL:**https://www.kaggle.com/api/v1/datasets/download/petalme/seattle-weather-prediction-dataset
- **Data Content:**
  Includes columns such as "Date," "Temperature (High/Low)," "Precipitation," and "Weather Condition."
- **Data Quality:**
  - Missing values are handled using mean or median imputation.
  - Outliers and anomalies (e.g., extreme weather) were addressed using statistical thresholds.
- **License:** Publicly accessible via Kaggle's open license policy.

2. 380,000 Weather Data Records

- **Why Chosen:**
  Provides a comprehensive dataset across various locations, enabling a broader comparative analysis.
- **Source:** Kaggle

- **URL:**https://www.kaggle.com/api/v1/datasets/download/pinto391/380000-weather-data
- **Data Content:**
  Contains fields like "Location," "Temperature," "Humidity," "Wind Speed," and "Date."
- **Data Quality:**
  - Addressed missing data using interpolation and median replacement.
  - Ensured consistency in date and location formats.
- **License:** Publicly available under Kaggle's terms.

Methodology

1. Data Gathering:

- Downloaded the datasets and documented their structure, coverage, and quality.

2. Data Preparation:

- Cleaned missing data by filling numeric fields with medians and removing redundant entries.
- Standardized temperature units (Celsius/Fahrenheit).
- Unified date formats using `pandas.to_datetime`.
- Normalized categorical labels for "Weather Condition" fields across datasets.

3. Analysis:

- Examined temperature, precipitation, and seasonal patterns.
- Calculated critical metrics:
  - **Average Precipitation:** Mean precipitation per month/year.
  - **Temperature Variability:** Range and standard deviation in daily highs/lows.
  - **Extreme Events:** Frequency of days with high wind speeds or heavy rainfall.

4. ETL Pipeline:

- Built a Python-based automated pipeline using Pandas and SQLite.
- Filtered essential columns: "Date," "Location," "Temperature (High/Low)," "Precipitation, "Wind Speed."
- Stored cleaned and unified data in an SQLite database for efficient querying and analysis.

5. Documentation:

- Created visual charts for trends in temperature, precipitation, and wind speed.
- Summarized results for actionable weather prediction insights.

Results:

Temperature Trends:

- **Seattle:**
  - Moderate annual temperature variability, with the highest average highs in July and lowest in January.
  - Temperature extremes (e.g., heatwaves) observed in recent years.
- **General Patterns (from 380,00 records):**
  - Diverse variability in temperature across locations, with coastal regions experiencing milder fluctuations compared to inland areas.

Precipitation and Weather Events:

- **Seattle:**
  - Frequent rainfall, especially during winter months (November to February).
  - Low occurrence of extreme weather events like snowstorms.
- **National Trends:**
  - Regions in the dataset showed higher precipitation in tropical climates, with occasional extreme storms.

Wind and Other Insights:

- Wind speed generally peaked during colder months, indicating higher variability due to seasonal storms.
- The 380,00-record dataset revealed notable differences in weather patterns based on elevation and proximity to water bodies.

Limitations

1. **Geographic Scope:**
   - Seattle dataset is region-specific, while the second dataset spans multiple regions, introducing potential inconsistencies in granularity.
2. **Data Completeness:**
   - Missing or incomplete entries for certain dates or locations could slightly bias the results.
3. **Generalization:**
   - Insights derived may not apply universally due to differences in topography, climate, and data recording standards.

Conclusion

The analysis revealed that weather patterns exhibit significant regional variability, with Seattle characterized by consistent rainfall and mild temperature changes. The comprehensive dataset allowed comparisons, highlighting factors like proximity to coasts and elevation as key determinants of variability. These insights can guide more accurate weather prediction models, particularly in regions prone to extreme weather conditions.