**Weather Data Transformations**

**Data Quality Checks**

1. **Data Type Conversion & Standardization**
   * Converted "null" strings to NA.
   * Converted valid column to datetime (ymd\_hm).
   * Extracted date and time from valid.
   * Converted numeric columns (tmpc, dwpc, relh, feel, drct, sped, mslp, p01i, vsby, gust\_mph) to numeric data type.
2. **Duplicate Detection & Removal**
   * Counted duplicate rows.
   * Removed duplicates using distinct().
3. **Missing Value Handling**
   * Checked missing values for each column.
   * Filled tmpc (temperature) missing values with **mean**.
   * Filled gust\_mph missing values with **wind speed (sped)**.
   * Filled wxcodes missing values with "No Weather Event".
4. **Factor Conversion**
   * Converted station and wxcodes into categorical (factor) variables.
5. **Unique Value & Range Checks**
   * Checked **date range** of data.
   * Counted unique stations.
6. **Derived Column Creation**
   * Created new column WEATHER\_ID = date\_station for unique row identification.

**Data Validation Checks**

1. **Structural Validation**
   * Used str() and summary() before and after cleaning to verify column structure, types, and ranges.
2. **Statistical Validation**
   * Computed summary statistics to validate numerical values.
3. **Distribution Checks**
   * Plotted **temperature distribution** (histogram).
   * Plotted **relative humidity distribution** (histogram).
4. **Outlier Detection**
   * Boxplot of wind speed (sped) to identify potential outliers.
5. **Missing Data Validation**
   * Visualized missing values per column using bar chart.
6. **Date Validation**
   * Checked min & max of date to ensure valid range.
   * Converted to proper date format.