**🧮 Detailed Statistical Operations Task List (With Instructions)**

**🔹 1. Central Tendency Measures**

**Objective**: Understand where the data tends to cluster.

* **Task 1.1**: Calculate the **mean**, **median**, and **mode** for each of the following columns:
  + Temperature
  + Fuel\_Price
  + CPI
  + Unemployment
* **Instruction**: Use df['column'].mean(), median(), and mode() for each column. Interpret what the results say about the dataset.

**🔹 2. Dispersion Measures**

**Objective**: Measure how spread out the values are.

* **Task 2.1**: Calculate the **range** (max - min) of each numerical column.
* **Task 2.2**: Compute the **standard deviation** and **variance** for:
  + Temperature, Fuel\_Price, CPI, Unemployment, MarkDown1–5
* **Instruction**: Use .std() and .var() methods. Comment on which variable has the most/least variability.

**🔹 3. Percentiles & IQR**

**Objective**: Identify data concentration and outliers.

* **Task 3.1**: Calculate the **25th, 50th, and 75th percentiles** for Fuel\_Price, CPI, and all MarkDown columns.
* **Task 3.2**: Compute the **Interquartile Range (IQR)** for the above.
* **Instruction**: Use np.percentile() or df.quantile([0.25, 0.5, 0.75]). IQR = Q3 - Q1.

**🔹 4. Z-score & Outlier Detection**

**Objective**: Standardize values and detect anomalies.

* **Task 4.1**: Compute the **z-score** for Fuel\_Price and Temperature columns.
* **Task 4.2**: Identify and count **outliers** using z-score > 3 or < -3.
* **Instruction**: Use scipy.stats.zscore() or manual formula:

z=(x−mean)/std

**🔹 5. Coefficient of Variation (CV)**

**Objective**: Assess relative variability.

* **Task 5.1**: Calculate CV for Temperature, Fuel\_Price, CPI, Unemployment:

CV= (Standard Deviation/Mean)\*100%

* **Instruction**: Compare which feature is most volatile (high CV).

**🔹 6. Skewness and Kurtosis**

**Objective**: Measure asymmetry and peakness of data.

* **Task 6.1**: Calculate **skewness** for numerical columns.
* **Task 6.2**: Calculate **kurtosis** to understand tail distribution.
* **Instruction**: Use df.skew() and df.kurtosis(). Interpret:
  + Skewness > 0: Right-skewed
  + Skewness < 0: Left-skewed
  + Kurtosis > 3: Heavy tails

**🔹 7. Statistical Summary by Groups**

**Objective**: Compare distributions across subgroups.

* **Task 7.1**: Group by IsHoliday and calculate:
  + Mean and std of all markdown columns.
  + Mean Fuel\_Price and Temperature.
* **Task 7.2**: Group by Store and calculate average CPI, Fuel\_Price, and Unemployment.
* **Instruction**: Use groupby() followed by .agg(['mean', 'std']).

**🔹 8. Handling Missing Data – Statistical Approach**

**Objective**: Explore missingness and treat it statistically.

* **Task 9.1**: Count missing values for MarkDown1–5, CPI, Unemployment.
* **Task 9.2**: Fill missing values using:
  + Mean
  + Median
  + Forward fill
* **Instruction**: Compare before vs after fill using .describe() to see changes in stats.

**🔹 9. Visual Statistical Insights**

**Objective**: Combine statistics with visualization.

* **Task 10.1**: Plot **box plots** for each numerical column (to visualize spread and outliers).
* **Instruction**: Use seaborn.boxplot().