Model Tester Programming Exercise

Before beginning, be sure to create a clean new directory that will hold all the output you produce as part of this exercise (CSV files, images, etc.).

- Task 1: Import data.
  - There are two CSV files in the data folder, PriceData_A.csv and PriceData_B.csv. Each of these files contains data that is representative of simulated power prices generated by the cQuant platform. Import both files as separate data frames.
- Task 2: Pivot the data.
  - Transform both data sets into long format. The resulting data frames should have 4 columns: Name, Date, Run (containing the original column names Run_1, Run_2, … Run_5), and Value (containing the numeric values for the corresponding runs), and should have 5 times as many rows as the original data sets imported in Task 1.
- Task 3: Join the data.
  - Combine the two data sets into a single data frame by joining on the 'Name', 'Date' and 'Run' columns. The result should be a single data frame with columns for Name, Date, Run, and two separate columns for the values from each table.
- Task 4: Calculate differences between the sets of results
  - Calculate 4 metrics describing the differences between the two sets of results:
    1. Difference (A – B)
    2. Percent difference (A – B) / A
    3. Absolute difference (absolute value of the result of [1])
    4. Absolute percent difference (absolute value of the result of [2])
- Task 5: Calculate summary statistics
  - Power prices vary in important ways by year, month, day of week, and hour of day. Calculate the minimum, maximum, and average value of the absolute difference metric (calculated in step 3 of Task 4), grouped by the relevant date and time dimensions. The result should be 4 separate summary tables:
    1. Min / max / mean by year
    2. Min / max / mean by month
    3. Min / max / mean by day of week
    4. Min / max / mean by hour of day
  - If you have time, you may also wish to generate additional tables that group the results by multiple dimensions, such as by month and hour of day, to support Task 7.
- Task 6: Write the summary tables to file
  - For each of the 4 summary tables generated in Task 5 (and any others you decide to create), write the results to a csv file in your output folder that is intuitively named.
- Task 7: Generate visualizations describing the differences
  - This one is open-ended. Using the results of Tasks 4 and 5, generate figures that you feel help explain the overall magnitude of the differences between results A and B, as well as how those differences vary over the time dimensions from Task 5. Save these figures in your output folder.