Quality and Supply Chain Management GRA Assignment

Organizing assignment files

1. Please create the git repository in GitHub
2. Clone the repository to your local computer
3. Add your assignment files to this repository
4. For each assignment question,
    a. Create a new branch from master/main branch
    b. Add necessary code for answering the questions
    c. Commit your changes to the new branch
    d. Push your commits to remote
    e. Create a pull request
    f. Merge it with your master/main branch (Usually, pull requests are reviewed and approved by a peer but in this case, you can just approve and merge by yourself)

Python programming instructions

1. All the questions' answering steps must be demonstrated in the Python code
2. Please use Python 3.8 or above
3. You are welcome to use open-sourced Python packages
4. Please use easily interpretable variable names
5. Organize your code neatly with comments and/or markdown text wherever necessary
6. Create one Jupyter Notebook (.ipynb file) per assignment question
7. Include all the input and output data files in the repository

SQL programming instructions

1. Question – 2 requires writing SQL queries for answering the sub-questions
2. Create a sqlite database and load sheets from question_1_2.xlsx in different tables (One table per sheet) using https://sqliteonline.com/
3. Use the online editor to test the queries while answering the questions.
4. Create a .sql file in the same GitHub repository and add your final queries to the file
5. Organize your queries neatly with comments wherever necessary
6. Commit all the files in the repository

Question – 1 (Using Python)

Data file – question_1_2.xlsx

1. Can you find out the mean and standard deviation of product height and weight based on each product code?
2. Which product family produces higher number of defective parts? (Defect - Any unit with measurement greater than upper limit) – Please demonstrate using a pareto chart
3. For the variant Y, which of the two product families show better process control? How will you visualize this?
4. Visualize how height and weight measurements are varying between lots. Also, how can we see the defect count by lot and product family?
5. Is it possible to project the test variation for future orders 7 and 8 based on the given data? Why or why not? Is there some additional information that will be needed for this purpose?

If there were missing data, please demonstrate how you handled them.

Question – 2 (Q1 and Q2 - Using SQL)

Data file – question_1_2.xlsx

1. Can you find out the mean and standard deviation of product height and weight based on each product code?
2. Filter the list of items from the measurements table with weight less than the lowest LSL value in the specifications table.
3. Write the SQL query to order the products based on their decreasing mean height and extract the top 2 products.

If there were missing data, please demonstrate how you handled them.

Question – 3 (Using Python)

Data file – question_3.csv

1. Generate a matrix of charts with 3 rows and 4 columns containing the histograms of all the data columns (except ID and quality)
2. By looking at the histograms generated in (1), what comments can you make about the skewness of distribution?
3. Calculate the Pearson's correlation coefficient for all the combination of variables in the dataset (except ID) and visualize it as cross matrix
4. Extract and list top 5 variable pairs that show positive correlation (if any) and top 5 variable pairs that show negative correlation (if any)
5. Generate a matrix of charts with 2 rows and 5 columns (max.) containing the scatter plots of top 5 variable pairs that show positive correlation (1st row) and top 5 variable pairs that show negative correlation (2nd row)

If there were missing data, please demonstrate how you handled them.