ENSF 692 Spring 2024 – Project Grading

Project Rubric (100 marks, 30% of overall grade)

Your code must successfully compile to be graded. Code that does not compile will be given a grade of zero. For compiled code, partial marks may be given for each criterion listed below. You must pass the project to pass the course.

Data Handling (20%):

At least three separate datasets are merged into a large set with a minimum of ten columns and 200 rows. Program does not modify the Excel files directly. No information is hard-coded/copy-pasted except for the Excel column names. Data is stored as a multi-indexed DataFrame. Two merge/join operations are used, and all duplicated columns/rows are deleted. Data is sorted according to the indices.

Code Implementation (25%):

Program solution uses the describe method to print aggregate stats for the entire dataset. At least two columns are added to the combined dataset. An aggregation computation is used for a subset of the data. A masking operation, groupby operation, and pivot table are all used correctly. The code structure includes at least two user-defined functions or a class that contains two methods. No global variables are used.

User Interface and Execution (25%):

User is given clear guidance on how to enter the two required input values. If invalid input is provided, an exception is used to prompt for re-entry without terminating the program. Clear headers are used to separate all output and data is presented in the correctly sorted order. Screenshots show the expected execution, including handling of incorrect input. An exported Excel sheet shows the entire indexed dataset, and a plot is shown that correctly depicts an aspect of the data.

Commenting and Syntax (10%):

All team members' names and group number are included on all submitted files. Comments are included throughout the code to explain the functionality. All classes, methods, and functions are fully documented using docstrings (including summary, parameters, and return values). All variables and functions have clear and useful names that use lowercase words separated by an underscore, all classes have CamelCase names. Code is clearly indented, and spaces are included between variables and operators.

Report (10%):

Report briefly describes the chosen dataset, summarizes the user interface input/output, and explains how the solution meets the given specifications. Minimal spelling or grammatical errors. An IEEE-style citation is provided for the chosen dataset.

Presentation (10%): (to be given June 14th during the class)

Demonstration clearly explains how the solution meets the requirements including the user input/output. The dataset analysis is explained, and the plot result is shown. Speakers are clear and audible. All team members participate in the demonstration and duration is less than 5 minutes. All members are able to answer questions about the program.