**CP610 Project Deliverable #1 (Worth 6/30)**

**Due 7 Oct, 2025 at 11:59pm**

**What to Do:**

1. Download Deliverable1Dataset.csv
2. For each **categorical attribute**, do the following:
   1. Identify the type of Missing Data (MCAR/MAR/MNAR). Explain your rationale.
   2. Handle missing data. Justify your choice of the method you chose.
   3. Encode it using one of the techniques discussed in class. Justify your choice.
   4. For the attribute Customer ID, use the *Target Encoding* method where the target is the Total Spent attribute.
   5. You don’t need to encode the attribute Transaction ID as it is not going to be part of the data analysis going forward.
   6. After encoding the categorical features, perform the tasks shown in 3. below.
3. For each **numerical attribute**, do the following:
   1. Identify the type of Missing Data (MCAR/MAR/MNAR). Explain your rationale.
   2. Handle missing data. Justify your choice of the method you chose.
   3. Rescale using normalization, standardization, and robust scaling. Which method you think is most suitable for the attribute at hand and why?
4. You are required to perform the tasks above programmatically using Python (utilizing pandas and numpy).
   1. Make sure to comment your code. I should be able to understand what your code is doing by reading the comments only.
   2. Use descriptive names for functions, variables, classes …etc.
   3. The code implementing each task should be placed in its own file. The file name should clearly indicate which task it implements (e.g., rescaling-3-d.py).
   4. Create a data file that reflects how the data looks After each scaling task. The file name should reflect the task the file is the result of (e.g., data-rescaling-Norm-3-d.csv).

**What to deliver:**

Place the following in a zip file. The file name should follow the format deliverable#1-your-group-name. Submit the zip file at the Deliverable #1 Dropbox on MyLS.

1. All the python code.
2. All the dataset versions.
3. A report that explains what you did and the rationale behind the decisions you made. List any conclusions you made. Provide supporting charts when appropriate.

**Note:**

No email submissions will be accepted. Start early and submit your deliverable on time!