Part I: Research Question

A. Describe the purpose of this data mining report by doing the following:
   1. Propose **one** question relevant to a real-world organizational situation that you will answer using **one** of the following classification methods:
      - *k*-nearest neighbor (KNN)
      - Naive Bayes
   2. Define **one** goal of the data analysis. Ensure that your goal is reasonable within the scope of the scenario and is represented in the available data.

Part II: Method Justification

B. Explain the reasons for your chosen classification method from part A1 by doing the following:
   1. Explain how the classification method you chose analyzes the selected data set. Include expected outcomes.
   2. Summarize **one** assumption of the chosen classification method.
   3. List the packages or libraries you have chosen for Python or R, and justify how *each* item on the list supports the analysis.

Part III: Data Preparation

C. Perform data preparation for the chosen data set by doing the following:
   1. Describe **one** data preprocessing goal relevant to the classification method from part A1.
   2. Identify the initial data set variables that you will use to perform the analysis for the classification question from part A1, and classify *each* variable as continuous or categorical.
   3. Explain *each* of the steps used to prepare the data for the analysis. Identify the code segment for *each* step.
   4. Provide a copy of the cleaned data set.

Part IV: Analysis

D. Perform the data analysis and report on the results by doing the following:
   1. Split the data into training and test data sets and provide the file(s).
   2. Describe the analysis technique you used to appropriately analyze the data. Include screenshots of the intermediate calculations you performed.
   3. Provide the code used to perform the classification analysis from part D2.

Part V: Data Summary and Implications

E. Summarize your data analysis by doing the following:
   1. Explain the accuracy and the area under the curve (AUC) of your classification model.
   2. Discuss the results and implications of your classification analysis.
   3. Discuss **one** limitation of your data analysis.
   4. Recommend a course of action for the real-world organizational situation from part A1 based on your results and implications discussed in part E2.