Evaluation Criteria

Q1.1: Perform the Exploratory Data Analysis of the Data (20 points)

Completeness of Analysis (10 points): The student should cover a wide range of EDA techniques, including univariate and bivariate analyses, to uncover patterns, trends, outliers, and anomalies in the data. This includes using summary statistics, distributions, remove NaN, outliers, removal, and various types of plots (histograms, scatter plots, box plots, etc.).

Quality of Visualization (5 points): Visualizations should be clear, appropriately labeled, and effectively convey the intended information. The use of different types of charts to highlight specific data characteristics is essential.

Insightfulness of Observations (5 points): Beyond basic observations, students should demonstrate the ability to derive meaningful insights from their exploratory analysis. This involves interpreting the data in the context of the business and suggesting potential implications of the observed patterns.

Q1.2: Describe Main Data Insights and Connect with Business (10 points)

Clarity and Conciseness of Insights (4 points): The description of the main insights should be clear and direct, avoiding unnecessary complexity. Students should succinctly summarize the key findings from their EDA in a way that is accessible to both technical and non-technical audiences.

Relevance to Business (3 points): Insights should be directly connected to the business context of ABC Insurance. This includes how the findings could impact customer segmentation, policy offerings, marketing strategies, or other business decisions.

Actionability of Insights (3 points): The insights described should not only be interesting but also actionable. This means students should articulate how ABC Insurance could use these insights to inform decision-making, improve customer engagement, tailor products, or any other strategic actions that could benefit the company.

Q2.1 Build a Baseline Model to Estimate the Customer Value Based on Their Characteristics (15 points)

Model Selection and Justification (3 points): The choice of the baseline model should be justified based on the data characteristics and the predictive task. The student should explain why the selected model is appropriate as a starting point for estimating customer value.

Data Preprocessing and Feature Engineering (4 points): Evaluation will consider how the student prepares the data for modeling, including handling missing values, encoding categorical variables, and selecting or creating relevant features that could influence the prediction of customer value. Also the need to create the target variable as the sum of the values (just sum? There are negatives values).

Model Training and Validation (4 points): The process of training the model and the approach used for validation (e.g., cross-validation, train-test split) should be clearly described. The student should demonstrate an understanding of avoiding overfitting and the importance of model generalization.

Performance Evaluation (4 points): The student must present the performance of the baseline model using appropriate metrics (e.g., RMSE, MAE for regression tasks). A discussion on how these metrics reflect the model's ability to predict customer value accurately is essential.

Q2.2 Build a Second Model to Estimate the Customer Value Based on Their Characteristics and Compare Model Performance (10 points)

Advanced Model Selection and Justification (3 points): The rationale for choosing a second, potentially more complex model should be based on the baseline model's limitations. The student should explain how the new model could address these limitations.

Comparison of Model Performances (7 points): The student must compare the performance of the baseline and the second model, using the same evaluation metrics. The comparison should highlight improvements in prediction accuracy and discuss any trade-offs (e.g., complexity, computational cost and explainability).

Q2.3 Based on the Analysis, Suggest a Model to the Head of Marketing and an Application Scenario. Please Justify Your Decisions (10 points)

Recommendation of Model (4 points): The student should convincingly argue why their chosen model (between the baseline and the second model) is best suited for estimating customer value at ABC Insurance. This involves summarizing the strengths of the model in the context of the business's needs.

Application Scenario (3 points): A clear, compelling application scenario for the recommended model should be described. This scenario should illustrate how the model can be used by the marketing team to enhance customer value through targeted marketing, personalized product offerings, or other strategic initiatives.

Justification and Strategic Alignment (3 points): The justification should convincingly align the model's capabilities with ABC Insurance's strategic goals. This includes how the model's

insights can lead to actionable strategies for optimizing marketing efforts, improving product offerings, and ultimately increasing customer value.

Q3.1 Build a Baseline Model to Predict Customers That Will Churn (15 points)

Model Selection and Justification (3 points): The student must explain why the chosen baseline model is appropriate for predicting customer churn, considering the nature of the dataset and the prediction task. The rationale should include any assumptions made by the model that align with the data characteristics.

Data Preprocessing and Feature Selection (4 points): Evaluation will focus on the steps taken to prepare the dataset for modeling, including handling missing values, feature engineering, and the selection process for variables included in the model. The relevance of chosen features to the churn prediction task should be clearly articulated.

Model Training and Validation (4 points): The methodology for training the baseline model and the validation technique(s) used should be well-documented. This includes detailing the approach to mitigate overfitting and ensuring the model's generalizability to unseen data.

Performance Evaluation (4 points): The student should present and interpret the model's performance using appropriate metrics (e.g., accuracy, precision, recall, F1-score, AUC-ROC for classification tasks). The discussion should also cover the model's ability to identify churn accurately within the context of ABC Insurance's operational goals.

Q3.2 Build a Second Model and Compare the Model Performance (15 points)

Advanced Model Selection and Justification (3 points): Justification for selecting a more advanced or different model as a comparison to the baseline should be based on potential areas for improvement identified in Q3.1. The student must explain how the second model could better address the churn prediction challenge.

Comparison of Model Performances (7 points): A critical comparison between the baseline and the second model's performance metrics should be provided. The student must discuss any improvements in churn prediction accuracy, as well as any trade-offs involved (e.g., model complexity, interpretability, computational requirements).

Q3.3 Based on the Analysis, Suggest a Model to the Head of Marketing and an Application Scenario. Please Justify Your Decisions (10 points)

Recommendation of Model (4 points): The student should provide a clear recommendation for one of the models, supported by a concise summary of why this model is preferred for

ABC Insurance's churn prediction needs. This includes highlighting the model's strengths and how it aligns with the objective of improving customer retention.

Application Scenario (3 points): A detailed scenario should be described where the recommended model can be applied by the marketing team to identify and address potential churn risks. This scenario should demonstrate an understanding of how predictive insights can inform targeted retention strategies.

Justification and Strategic Impact (3 points): The recommendation must be justified in terms of strategic value to ABC Insurance, including how the model and its application can lead to effective churn prevention measures. The student should articulate how the model's use will align with broader business objectives, such as improving customer satisfaction, increasing retention rates, and ultimately contributing to the company's bottom line.