Practical Exam - Fitness Class

Company Background

GoalZone is a fitness club chain providing five types of fitness classes in Canada. Finally, the fitness classes schedule is back to normal after the COVID-19 restrictions are lifted. However, they have received many complaints from the customers about having a hard time booking a fitness class.

From initial analysis, the program operation team found out that the fitness classes are always fully booked but are having a low attendance rate per class. To improve this situation, they would like to increase the class capacity for customers to sign up if a low attendance rate is predicted.

Customer Questions

The operation manager has asked you to answer the following:

- Can you predict the attendance rate for each fitness class?

Dataset

The dataset contains the attendance information for the class scheduled this year so far. The data is available in a DataCamp Workspace, which you can find from the certification dashboard.

The dataset needs to be validated based on the description below:

Column Name	Criteria
Day of Week	Character, the day of the week the class was scheduled, one of values from “Mon” to “Sun”.
Time	Character, the time of the day the class was scheduled, either AM or PM.
Class Category	Character, the category of the fitness class, one of “Yoga”, “Aqua”, “Strength”, “HIIT”, or “Cycling”.
Days Before	Numeric, number of days the class stayed fully booked, maximum five days.
Class Capacity	Numeric, maximum number of members can sign up for that class, either 15 or 25. The class capacity being 26 needs to be updated to 25.
Attendance	Numeric, number of members actually attended the class.

Average Age	Numeric, average age of the members signing up for that class. Remove rows that average age is smaller than 14 because group fitness class are for members aged 14 and order.
Number of New Students	Numeric, number of new students signing up for this class.
Number of Members Over 6 months	Numeric, number of members signing up for the class have been joining the club more than 6 months.

Submission Requirements

1. You are going to create a written report summarizing your findings. Use the project task list provided below for guidance in the tasks you should complete and information to include in the report.

2. You will need to use DataCamp Workspace to complete your analysis, write up your findings and share visualizations.

3. You must use the data we provide for the analysis.

4. Use the grading rubric provided below to check your work before submitting the report.

Project Task List

Data Validation

1. Check the data against the criteria in the data dictionary.
2. For each column in the data, describe the validation tasks you complete and what you found. Have you made any changes to the data to enable further analysis?

Exploratory Analysis

1. Explore the characteristics of the numerical and categorical variables.
2. Create at least two different data visualizations that include only a single variable.
3. Create at least one data visualization that includes two or more variables.
4. Describe what you found in the exploratory analysis. Have you made any changes to those variables to enable model fitting?

Model Fitting

1. Describe what category of machine learning models are suitable to address the problem (e.g. regression, classification, clustering).

2. Choose and fit a baseline model.
3. Choose and fit a comparison model.
4. Explain the reason for choosing the two models above.

Model Evaluation

1. Evaluate the performance of two models by appropriate metrics.
2. Compare the evaluation results between two models and describe what that means for addressing the business problem.

Grading Rubric

You will be graded against the following criteria. You must pass all criteria to pass this part of the certification.

Domain	***Description***	**Sufficient**	**Insufficient**
Data Validation	Assess data quality and perform validation tasks	Has validated all variables against provided criteria and where necessary has performed cleaning tasks to result in analysis-ready data.	Has not conducted all the required checks and/or has not cleaned the data. May have removed data rather than performed cleaning tasks.
Data Visualization	Create data visualizations in coding language to demonstrate the characteristics of data and represent relationships between features.	Has created at least two different visualizations of single variables (e.g. histogram, bar chart, single boxplot) Has created at least one visualization including two or more variables (e.g. scatterplot, filled barchart, multiple boxplots) Has used visualizations that support the findings being presented	Has used the same visualization throughout. Has not included graphics to represent single variables and relationships. Has not used visualizations that support the findings being presented.
Model Fitting	Implement standard modeling approaches for supervised or unsupervised learning problems	Correctly identified the type of problem (regression, classification or clustering) Has selected and fitted a model for that problem to be used as a baseline.	Has incorrectly identified the type of problem. Has not fitted a baseline model or has used a model for the wrong type of problem. Has not fitted a

		Has selected and fitted a comparison model for the problem that they were provided.	comparison model or has used a model for the wrong type of problem.
Model Evaluation	Use suitable methods to assess the performance of a model	Compared the performance of the two models/approaches using any method appropriate to the type of problem. Has described what the model comparison shows about the selected approaches.	Has selected a method not suitable for the type of problem. Has not described what the results show about the selected approaches.
Communication	Presents data concepts to small, diverse audiences	For each analysis step, has explained their findings and/or the reasoning for selecting approaches.	Has not provided a summary for each step (data quality, exploratory analysis, model fitting and model evaluation).