Practical Exam - Recipe Website

Company Background

Tasty Bytes , an online recipe startup, has hired you as a data scientist. The website features new recipes on the homepage every day. The owner has told you that on days that they feature a popular recipe, traffic increases by as much as 40%. However, it is difficult to predict in advance which recipes will be popular.

Recipes are considered to be popular if they receive a high score. The data team has collected data from previously published recipes.

Customer Question

The owner wants to know:

- Can you use information on previously published recipes to predict whether a recipe will receive a high score?

Success Criteria

The owner estimates that of all low scoring recipes, they currently correctly categorize 75% of them. They want to know how your approach compares to this.

Data Dictionary

The data is available in a DataCamp Workspace, which you can find from the certification dashboard. The data set has the following columns:

Column Name	Details
RecipeId	Numeric, unique identifier of recipe
Name	Character, name of recipe
RecipeCategory	Character, type of recipe (e.g., Dessert, Breakfast, etc)
Calories	Numeric, number of calories
CholesterolContent	Numeric, amount of cholesterol in milligrams
CarbohydrateContent	Numeric, amount of carbohydrates in grams
SugarContent	Numeric, amount of sugar in grams
ProteinContent	Numeric, amount of protein in grams
RecipeServings	Numeric, number of servings
HighScore	Numeric, where 1 is a high score (i.e., popular), and 0 is a low score (i.e., unpopular)

Submission Requirements

1. You are going to create a written report to summarize the analysis you have performed and your findings. This report should be for the data science manager. The task list below describes what they expect to see in your 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. You will also need to prepare and deliver an oral presentation. You should prepare around 8-10 slides to present to the non-technical customer. The task list below describes what they expect to see in the presentation.

5. Your presentation must be no longer than 10 minutes.

Task List- Written Report

Your written report should include both code, output and written text summaries of the following:

- Data validation, including a summary of any changes you make to the data
- Exploratory Analysis, including graphics to support your findings
- Model Development, including justification for your choice of models
- Model Evaluation, including explanation of what this means about your models
- Comparision to the business success criteria
- Final summary including recommendations for future work that the business should undertake

Task List - Oral Presentation

Your presentation should be targetted at the non-technical customer who requested the work you have completed. The presentation should include:

- An overview of the project and business goals
- A summary of the work you undertook and how this addresses the problem
- Your key findings including how your work compared to the business success criteria
- Your recommendations to the business for future work

Grading

Before submitting your written report or delivering your oral presentation, remember to check your work against the following grading 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 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	Has used the same visualization throughout. Has not included graphics to represent single variables and relationships. Has not used visualizations that

		barchart, multiple boxplots) Has used visualizations that support the findings being presented	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 selected and fitted a comparison model for the problem that they were provided.	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 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.
Business Focus	Make recommendations for analytic approaches based on business goals	Has described at least one of the business goals of the project Has explained how their work has addressed the business problem Has provided at least one recommendation for future action to be taken based on the outcome of the work done	Has not identified any business goals Has not explained how their work has addressed the business problem Has not provided any recommendations for future actions
Business Metrics	Judge performance of analytic results against relevant business criteria	Has defined a KPI to compare model performance to business criteria in the problem	Has not identified a KPI to compare the model performance to the business problem

		Has compared the performance of the two models/approaches using the defined KPI	Has not compared the performance of the two approaches using the defined KPI
Communication	Employs multiple tactics (written and verbal) to communicate to business leaders	For each analysis step, has provided a written explanation of their findings and/or reasoning for selecting approaches Has delivered a verbal presentation addressing the business goals, outcomes and recommendations	Has not provided a written summary for each step Has not delivered a verbal presentation