Practical Exam - Recipe Site Traffic

From:	Head of Data Science
Received:	Today
Subject:	New project from the product team

Hey!

I have a new project for you from the product team. Should be an interesting challenge. You can see the background and request in the email below.

I would like you to perform the analysis and write a short report for me. I want to be able to review your code as well as read your thought process for each step. I also want you to prepare and deliver the presentation for the product team - you are ready for the challenge!

They want us to predict which recipes will be popular 80% of the time and minimize the chance of showing unpopular recipes. I don't think that is realistic in the time we have, but do your best and present whatever you find.

You can find more details about what I expect you to do here. And information on the data here.

I will be on vacation for the next couple of weeks, but I know you can do this without my support. If you need to make any decisions, include them in your work and I will review them when I am back.

From:	Product Manager - Recipe Discovery
To:	Head of Data Science
Received:	Yesterday
Subject:	Can you help us predict popular recipes?

Hi,

We haven't met before but I am responsible for choosing which recipes to display on the homepage each day. I have heard about what the data science team is capable of and I was wondering if you could help me choose which recipes we should display on the home page.

At the moment, I choose my favorite recipe from a selection and display that on the home page. We have noticed that traffic to the rest of the website goes up by as much as 40% if I pick a popular recipe. But I don't know how to decide if a recipe will be popular. More traffic means more subscriptions so this is really important to the company.

Can your team:

- Predict which recipes will lead to high traffic?
- Correctly predict high traffic recipes 80% of the time?

We need to make a decision on this soon, so I need you to present your results to me by the end of the month. Whatever your results, what do you recommend we do next?

Look forward to seeing your presentation.We haven't met before but I am responsible for choosing which recipes to display on the homepage each

day. I have heard about what the data science team is capable of and I was wondering if you can help me choose which recipes we should display on the home page?

At the moment, I choose my favorite recipe from a selection and display that on the home page. We have noticed that traffic to the rest of the website goes up by as much as 40% if I pick a popular recipe. But I don't know how to decide if a recipe will be popular. More traffic means more subscriptions so this is really important to the company.

Can your team:

Predict which recipes will lead to high traffic?

Correctly predict high traffic recipes 80% of the time?

We need to make a decision on this soon, so I need you to present your results to me by the end high-traffic. Whatever your results, what do you recommend we do next?

Look forward to seeing your presentation.

About Tasty Bytes

Tasty Bytes was founded in 2020 in the midst of the Covid Pandemic. The world wanted inspiration so we decided to provide it. We started life as a search engine for recipes, helping people to find ways to use up the limited supplies they had at home.

Now, over two years on, we are a fully fledged business. For a monthly subscription we will put together a full meal plan to ensure you and your family are getting a healthy, balanced diet whatever your budget. Subscribe to our premium plan and we will also deliver the ingredients to your door.

Example Recipe

This is an example of how a recipe may appear on the website, we haven't included all of the steps but you should get an idea of what visitors to the site see.

Tomato Soup

Servings: 4
Time to make: 2 hours
Category: Lunch/Snack
Cost per serving: $ (in dollar)

Nutritional Information (per serving)	
Calories	123
Carbohydrate	13g
Sugar	1g
Protein	4g

Ingredients:

- Tomatoes
- Onion
- Carrot
- Vegetable Stock

Method:

1. Cut the tomatoes into quarters…

Data Information

The product manager has tried to make this easier for us and provided data for each recipe, as well as whether there was high traffic when the recipe was featured on the home page.

As you will see, they haven't given us all of the information they have about each recipe.

You can find the data here.

I will let you decide how to process it, just make sure you include all your decisions in your report.

Don't forget to double check the data really does match what they say - it might not

Column Name	Details
recipe	Numeric, unique identifier of recipe
calories	Numeric, number of calories
carbohydrate	Numeric, amount of carbohydrates in grams
sugar	Numeric, amount of sugar in grams
protein	Numeric, amount of protein in grams
category	Character, type of recipe. Recipes are listed in one of ten possible groupings (Lunch/Snacks', 'Beverages', 'Potato',

	'Vegetable', 'Meat', 'Chicken, 'Pork', 'Dessert', 'Breakfast', 'One Dish Meal').
servings	Numeric, number of servings for the recipe
high_traffic	Character, if the traffic to the site was high when this recipe was shown, this is marked with “High”.

Guide to Data Science Projects

1. I would like you to create a written report to summarize the analysis you have performed and your findings. The report will be read by me (Head of Data Science). The list below describes what I expect to see in your written report.

2. You will need to use Google Colab workspace to complete your analysis, write up your findings and share visualizations.

3. You must use the data provided for the analysis.

4. You will also need to prepare and deliver a presentation. You should prepare around 8-10 slides to present to the product manager. The list below describes what they expect to see in your presentation.

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

Written Report

Your written report should include written text summaries and graphics of the following:

- Data validation:
    - Describe validation and cleaning steps for every column in the data
- Exploratory Analysis to answer the customer questions ensuring you include:

    - Two different types of graphic showing single variables only
    - At least one graphic showing two or more variables
    - Description of your findings
- Model Development including:
    - What type of problem this is
    - Fitting a baseline model
    - Fitting a comparison model
- Model evaluation
    - Show how the two models compare
- Definition of a metric for the business to monitor
    - How should the business monitor what they want to achieve?
    - Estimate the initial value(s) for the metric based on the current data?
- Final summary including recommendations that the business should undertake

Presentation

You will give an overview presentation to the product manager who requested the work. 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 the metric to monitor and current estimation
- Your recommendations to the business

Assignment Submission Guidelines

1. Write all your code in a single Jupyter Notebook/Google colab file. You just need to submit the .ipynb file.
2. For Presentation, you need to submit the pptx/pdf file of your presentation and also submit the video recording of your presentation (mp4 file).

Grading

Before submitting your written report or delivering your presentation, remember to check your work against the grading criteria.

You must pass all criteria to pass this part of the certification.

Competency	**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 bar chart, 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	Has incorrectly identified the type of problem Has not fitted a baseline model or has used a model for the wrong type of problem

	problem to be used as a baseline Has selected and fitted a comparison model for the problem that they were provided	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 compared the performance of the two models/approaches using the defined KPI	Has not identified a KPI to compare the model performance to the business problem 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