Case Study Project - Food Claims Process

You can use any tool that you want to do your analysis and create visualizations. You will only need to use DataCamp Workspace to write up your summary in Markdown.

We have included everything you need to complete the project in this document. Read it in detail before you get started.

Company Background

Vivendo is a fast food chain in Brazil with over 200 outlets. As with many fast food establishments, customers make claims against the company. For example, they blame Vivendo for suspected food poisoning.

The legal team, who processes these claims, is currently split across four locations. The new head of the legal department wants to see if there are differences in the time it takes to close claims across the locations.

Customer Question

The legal team has given you a data set where each row is a claim made against the company. They would like you to answer the following questions:

- How does the number of claims differ across locations?
- What is the distribution of time to close claims?
- How does the average time to close claims differ by location?

Dataset

The dataset contains one row for each claim. The dataset can be downloaded from here.

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

Column Name	Criteria
Claim ID	Character, the unique identifier of the claim.
Time to Close	Numeric, number of days it took for the claim to be closed.
Claim Amount	Numeric, initial claim value in the currency of Brazil. For example, "R$50,000.00" should be converted into 50000.
Amount Paid	Numeric, total amount paid after the claim closed in the currency of Brazil.
Location	Character, location of the claim, one of "RECIFE", "SAO LUIS", "FORTALEZA", or "NATAL".

Individuals on Claim	Numeric, number of individuals on this claim.
Linked Cases	Binary, whether this claim is believed to be linked with other cases, either TRUE or FALSE.
Cause	Character, the cause of the food poisoning injuries, one of 'vegetable', 'meat', or 'unknown'. Replace any empty rows with 'unknown'.

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 can use any tools you want to do your analysis and create visualizations.

3. You will need to use DataCamp Workspace to write up your findings and share visualizations. Use the Markdown Guide for your reference to write your report in the DataCamp Workspace.

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

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

Project Task List

Data Validation

1. Check the data matches the criteria in the data dictionary.
2. Describe the validation tasks you performed and what you found. Have you made any changes to the data to enable further analysis?

Data Discovery and Visualization

1. Use exploratory analysis methods to answer the customer questions in the project brief.
2. Create at least two different data visualizations to demonstrate the characteristics of variables.
3. Create at least one data visualization to demonstrate the relationship between two or more variables.
4. Describe what you found in the analysis and how the visualizations answer the business questions in the project brief.

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 to demonstrate the characteristics of data and represent relationships between features.	Has created at least two different types of data visualization that highlight characteristics of individual variables after validation. Has created at least one visualization that shows the relationship between two variables. 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.
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 validation, exploratory analysis).