Case Study Project - Pet Box Subscription

Company Background

PetMind is a nationwide pet product retailer in the United States. With inflation hitting 41-year highs, the company is planning to reduce the cost of customer retention by improving brand loyalty. The first strategy is to launch a monthly pet box subscription in three months.

The marketing team is preparing a list of popular products for the pet box subscription. The chief marketing officer wants to know whether the list should only include the products being purchased more than once.

Customer Questions

The marketing team would like to answer the following questions to help with the decision:

- How many products are being purchased more than once?
- Do the products being purchased again have better sales than others?
- What products are more likely to be purchased again for different types of pets?

Dataset

The dataset contains the sales records in the stores last year. The dataset can be downloaded from here.

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

Column Name	Criteria
Product ID	Character, the unique identifier of the product.
Product Category	Character, the category of the product, one of 11 categories.
Sales	Numeric, the sales of the product in dollars last year. For example, "$120,000" should be converted into 120000.
Price	Numeric, the price of the product in dollars.
Vendor ID	Character, the unique identifier of the vendor for the product.
Pet Size	Character, the pet size category the product is designed for, one of 5 size categories.

Pet Type	Character, the pet type the product is designed for, one of “cat” ,“dog”, “fish”, “bird”. Rows that indicate other pet types should not be included in the analysis.
Rating	Numeric, customer’s rating of the product (should be on a 10 point scale)
Rebuy	Binary, whether the product is purchased more than once or not (should be 1 or 0)

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 complete 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 business 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 visualizaitons 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).