

Practical Exam: Pet Box Subscription

Instructions

- Use any tools that you are comfortable with to perform the tasks required (for example Tableau, Power BI, MS Excel, Python, R).
- Write your solutions in the workspace provided from your certification page.
- Include all of the visualizations you create to complete the tasks.
- Visualizations must be visible in the published version of the workspace. Links to external visualizations will not be accepted.
- You do not need to include any code.
- You must pass all criteria to pass this exam. The full criteria can be found [here](#).

Background

PetMind is a retailer of products for pets. They are based in the United States.

PetMind sells products that are a mix of luxury items and everyday items. Luxury items include toys. Everyday items include food.

The company wants to increase sales by selling more everyday products repeatedly.

They have been testing this approach for the last year.

They now want a report on how repeat purchases impact sales.

Data

The dataset contains the sales records in the stores last year.

The dataset can be downloaded from [here](#).

Column Name	Criteria
product_id	Nominal. The unique identifier of the product. Missing values are not possible due to the database structure.
category	Nominal. The category of the product, one of 6 values (Housing, Food, Toys, Equipment, Medicine, Accessory). Missing values should be replaced with "Unknown".
animal	Nominal. The type of animal the product is for. One of Dog, Cat, Fish, Bird. Missing values should be replaced with "Unknown".
size	Ordinal. The size of animal the product is for. Small, Medium, Large. Missing values should be replaced with "Unknown".
price	Continuous. The price the product is sold at. Can be any positive value, round to 2 decimal places. Missing values should be replaced with the overall median price.
sales	Continuous. The value of all sales of the product in the last year. This can be any positive value, rounded to 2 decimal places Missing values should be replaced with the overall median sales.
rating	Discrete. Customer rating of the product from 1 to 10. Missing values should be replaced with 0.
repeat_purchase	Nominal. Whether customers repeatedly buy the product (1) or not (0) Missing values should be removed.

Tasks

Submit your answers directly in the workspace provided.

1. For every column in the data:
 - a. State whether the values match the description given in the table above.
 - b. State the number of missing values in the column.
 - c. Describe what you did to make values match the description if they did not match.
2. Create a visualization that shows how many products are repeat purchases. Use the visualization to:
 - a. State which category of the variable repeat purchases has the most observations
 - b. Explain whether the observations are balanced across categories of the variable repeat purchases
3. Describe the distribution of all of the sales. Your answer must include a visualization that shows the distribution.
4. Describe the relationship between repeat purchases and sales. Your answer must include a visualization to demonstrate the relationship.