Data Analyst Associate Practical Exam Submission

You can use any tool that you want to do your analysis and create visualizations. Use this template to write up your summary for submission.

You can use any markdown formatting you wish. If you are not familiar with Markdown, read the **Markdown Guide** ↗ before you start.

Task 1

For every column in the data:

- State whether the values match the description given in the table above
- State the number of missing values in the column.
- Describe what you did to make values match the description if they did not match.

Solution

Product_Id: There were 1500 unique values that match the description given. No missing values and no changes made to this column.

Category: There were 6 unique values (Housing, Food, Toys, Equipment, Meidicine, Accessory) that match the description given. 25 missing values filled with "-" were changed to "Unknown".

Animal: There were 4 unique values (Dog, Cat, Fish, Bird) that match the description given. No missing values and no changes made to this column.

Size: There were 3 unique values (Small, Medium, Large) that match the description given. No missing values and no changes made to this column.

Price: There were set of positive numeric values rounded to 2 decimal places that match the description given. 150 missing values filled with "unlisted" were replaced with the median value of the given data rounded to 2 decimal places which was 28.07.

Sales: There were set of positive numeric values rounded to 2 decimal places that match the description given. No missing values and no changes made to this column.

Rating: There were set of positive integers ranging from 1 to 10 that match the description given. 150 missing values filled with "NA" were replaced with 0.

Repeat_purchase: There were values given to be either 0 or 1 matching the description given. No missing values and no changes made to this column.
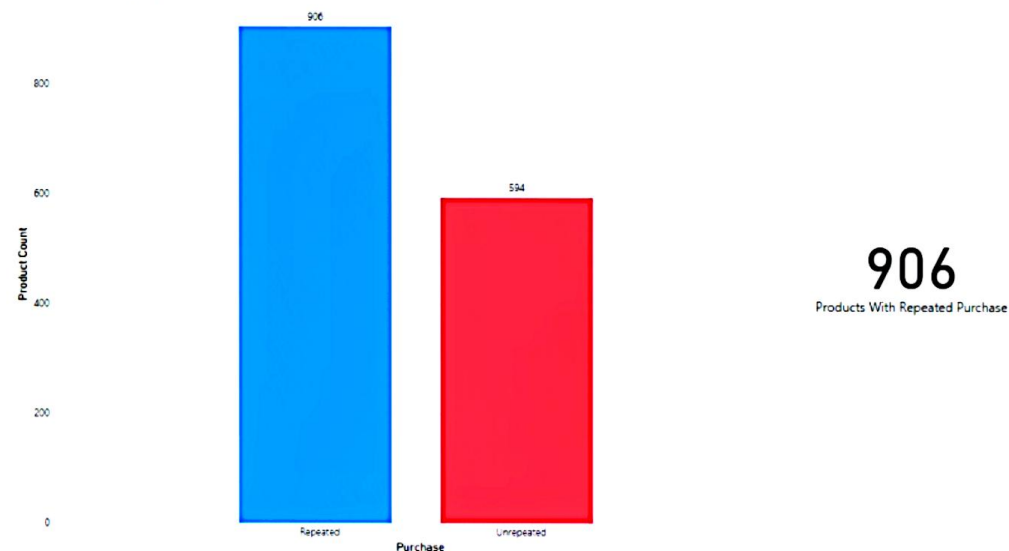
Task 2

Create a visualization that shows how many products are repeat purchases. Use the visualization to:

- State which category of the variable repeat purchases has the most ovservations
- Explain whether the observations are balanced across categories of the variable repeat purchases
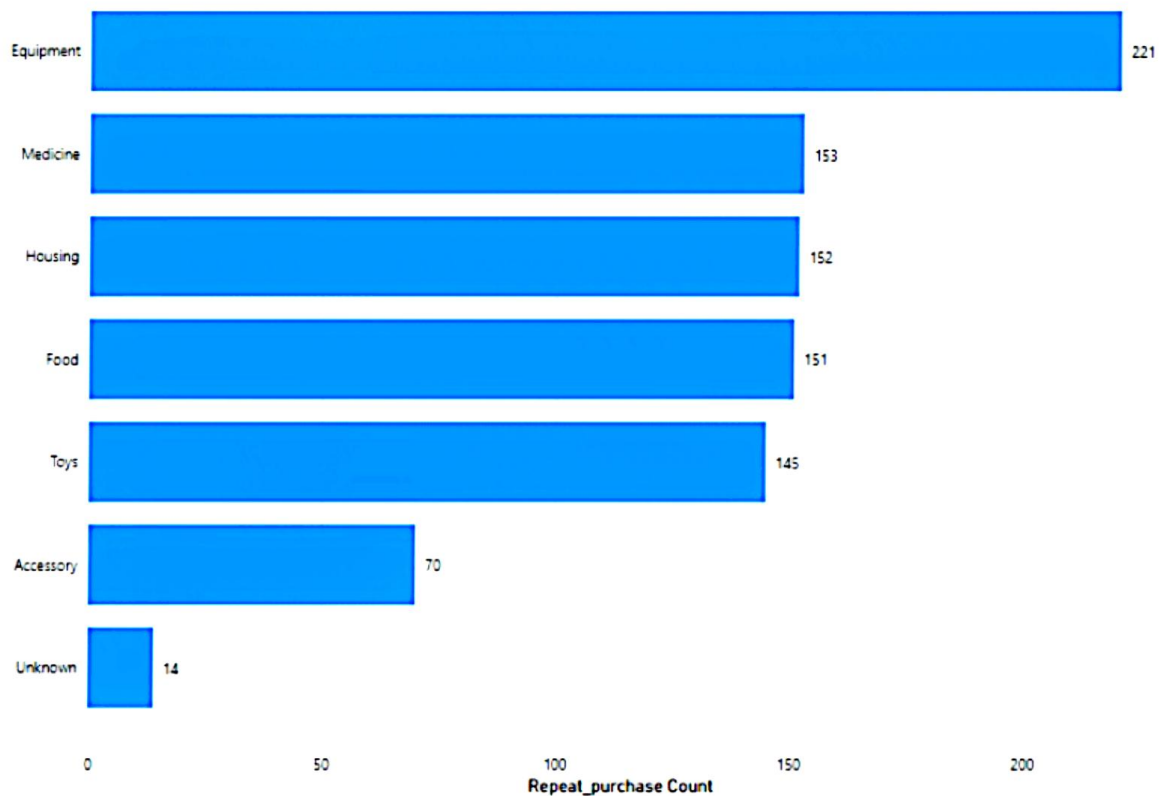
Solution

Product Count by Repeat Purchase

906
Products With Repeated Purchase

906 products are repeat purchase.

Repeat Purchase Count by Category

| Category	Repeat_purchase Count
Equipment	221
Medicine	153
Housing	152
Food	151
Toys	145
Accessory	70
Unknown	14

The category "Equipment" of the variable repeat purchases has the most observation. There are seven categories: Equipment, Medicine, Housing, Food, Toys, Accessory and Unknown. Since Equipment appears to have the highest observations, followed by Medicine, Housing and Food differing at close range and Unknown category having the lowest observation, the observations are hence not balanced across all cateogories of variable repeat purchases.

Task 3

Describe the distribution of all of the sales. Yours answer must include a visualization that shows the distibution.

Solution

Sales Distribution
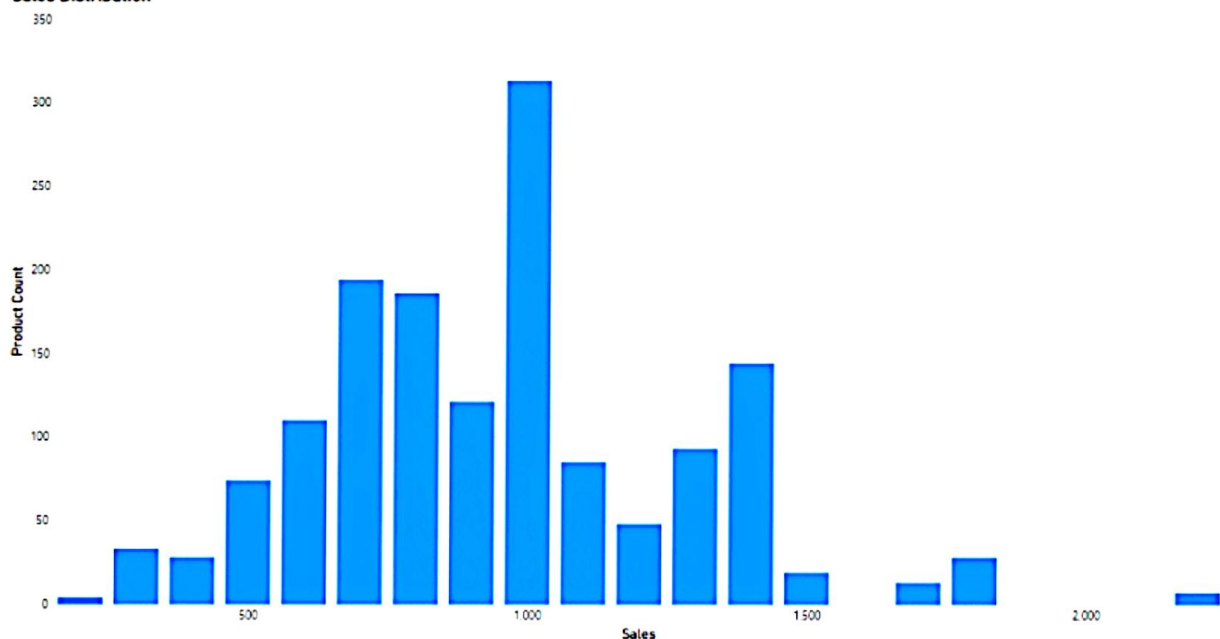
The sales distribution appears normal with most of the sales falling below 1,500 and fewer extreme values above 2,000. This shows that sales with lesser amount made more counts than those with higher amount. For the purpose of precision in the course of futher decision-making, more attentions has to be paid to the values below 1,500 for insights on what to intesify on and more attentions on values above it for insights on what to correct.

Task 4

Describe the relationship between repeat purchases and sales. Your answer must include a visualization to demonstrate the relationship.

Solution

Total Sales for Repeat Purchases

884.05K

Total Sales for Non Repeat Purchases

610.85K

Sales by Repeat Purchase

Looking at the box plots, even though "Not Repeated" Purchases have bulk of their sales to be totally above those of "Repeated" Purchases on the scale, "Repeated" Purchases have managed to record more sales. This draws insights into the impact of price as it appear that lower prices made more repeated purchases than higher price.

Conclusion: Repeat Purchase recorded more sales and Cheapness facilitated Repeat Purchases.

Recommendation: Since Petmind aims at increasing their sales by selling more everyday products and last year sales have shown that cheapness is a major facilitator for repeat purchase, Petmind might want to review their pricing across everyday products involving categories like Food and check the feasibility tilting them towards more affordability.