Data Analyst Associate Practical Exam Submission

Data Validation

The data provided was originally comprised of 880 rows and 9 columns. However, according to the instructions any rows that did not contain the pet types "cat", "dog", "bird" or "fish" were to be excluded, leaving the dataset with 833 rows. I also verified that the rating column was on a ten-point scale and that the rebuy column only contained either a 1 or 0 for its values. The sales column also needed to be converted from currency to whole numbers as requested in the dataset criteria. Lastly, I verified there were no missing values in the entire dataset.

At an overview, the dataset shows us we have:

- Product ID numbers
- 11 different product categories
- Sales number of the product last year and price of the product.
- 4 pet types, excluding hamsters and rabbits.
- 5 pet size categories, ranging from extra-small to extra-large.
- A unique character identifier for each vendor
- A 1-10 rating and rebuy indicator (1 or 0) for each product.

Data Discovery and Visualization

1. **How many products are being purchased more than once?**

Since the marketing team wishes to know whether the list of products for the subscription boxes should only include products purchased more than once, it is a good idea to see how many of their products are bought more than once. This was achieved by taking the count of the re_buy column and splitting it into two categories, "yes" and "no".

As this pie chart shows, the number of products purchased more than once are slightly less than those that are bought once. This could imply that some of these items purchased only once are larger, more expensive items such as a crate for a dog, or a tank for fish. However, the company does mention that they are planning to reduce the cost of customer retention. Since this appears to be an issue for PetMind, I believe the more likely explanation is that the majority of customers simply do not return to PetMind for their pet related supplies.

How Many Products were Re-buys

390

443

■ Yes ■ No
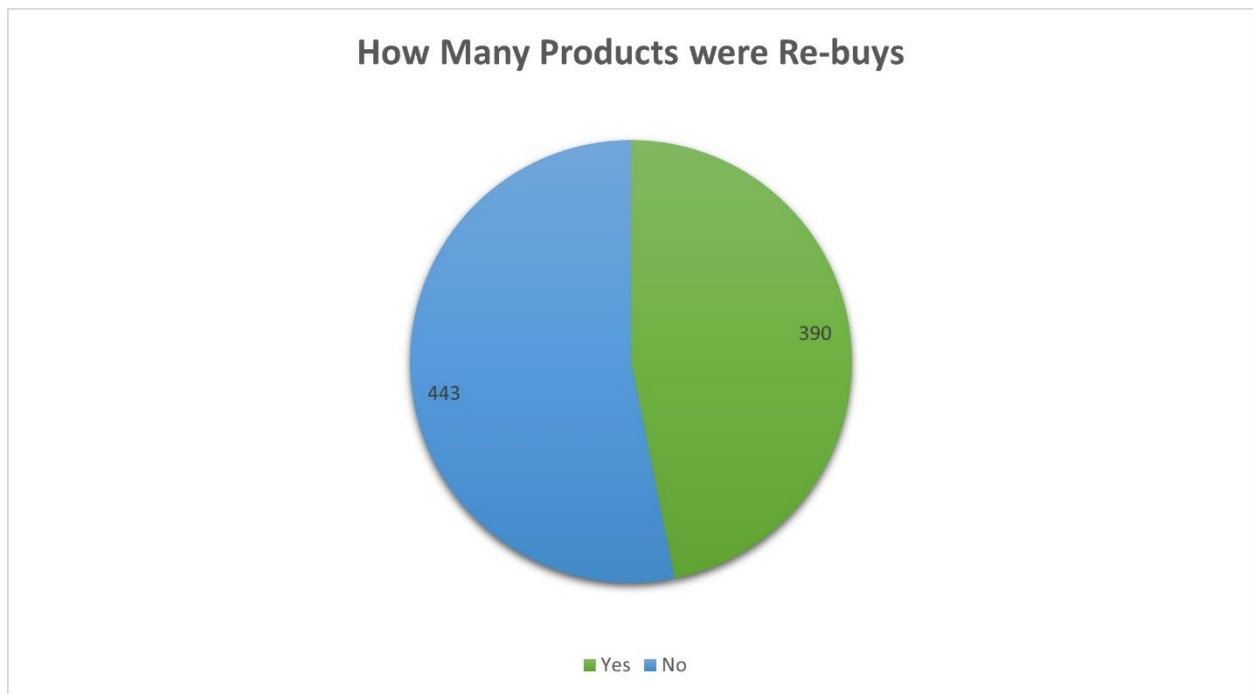
2. Do the products being purchased again have better sales than others?

If we are to gain better insight into which products should be included in the subscriptions, it is important to know what products are bringing in the most sales for PetMind. I calculated the seperate total sales for both repeat and single purchases and created a bar chart to compare the two at a quick glance. The chart provided below shows us that products purchased more than once are bringing in less money than those that are bought once. The difference between the two is not drastic though.

I believe it is safe to assume that pet owners, over time, spend more on items like food, toys, medicine, etc. than on one-time purchases such as a crate or fish tank. With this assumption in mind, one would expect PetMind's sales to reflect that products that are repeat purchases would outweigh the sales of one-time items. Yet, this is not the case with PetMind, which could imply they have issues retaining customers.
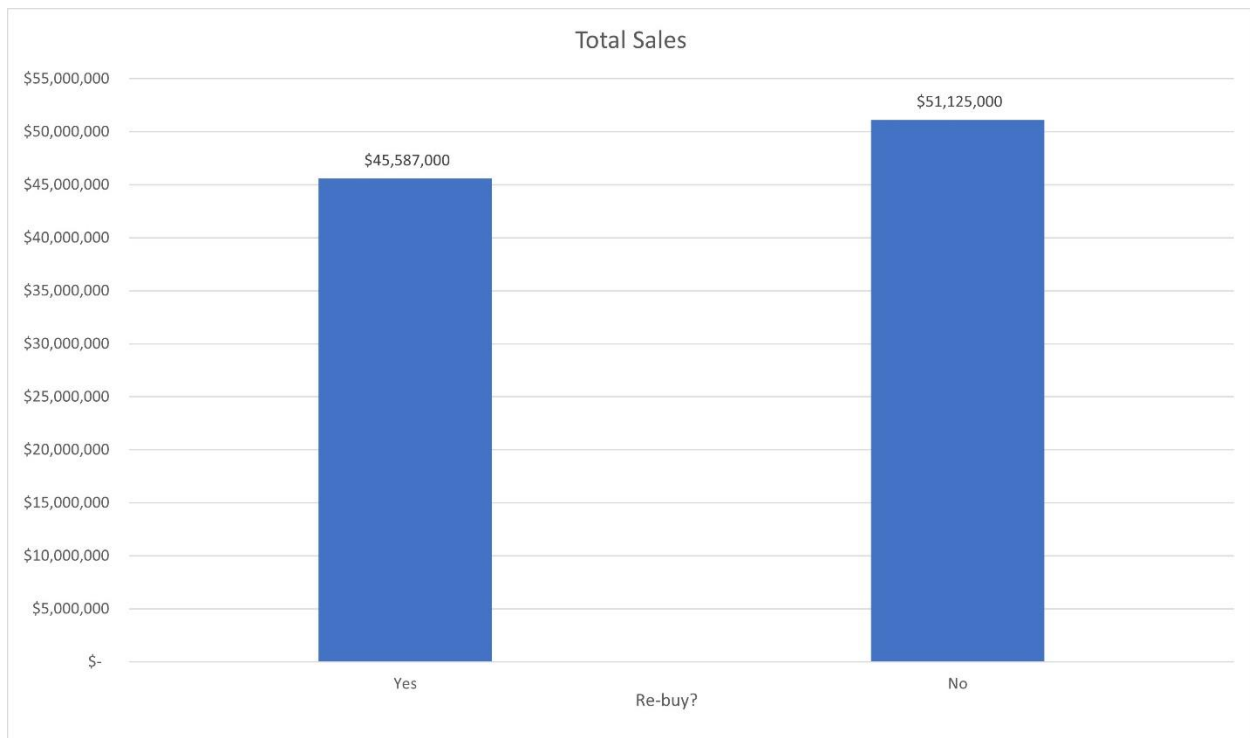
3. What products are more likely to be purchased again for different types of pets?

For the final question I started by creating a pivot table. The pet types were used as the columns, product categories for the rows and the table values were the count of re-buy products for each pet type, broken into the product categories. This pivot table was then visualized with a cluster column chart so we can easily determine which product categories have higher repeat purchases among the four pet types.

product_category	fish	dog	cat	bird
Accessory	0	10	9	1
Bedding	0	14	9	1
Clothes	0	8	10	1
Equipment	7	28	27	7
Food	1	17	18	1
Grooming	0	10	7	0
Housing	2	8	9	1
Medicine	1	20	15	1
Snack	8	25	19	7
Supplements	1	11	17	1
Toys	7	19	25	7

By far the most likely products to be re-buys are equipment, snacks and toys. This is reflected on the clustered column chart for all 4 pet types shown below. Food is also quite popular for dogs and cats, but not fish and birds. Based on this chart, dogs and cat make up the majority of sales across all categories.

After considering the data and the information gained from these insights, PetMind may wish to begin their subscription box service by providing the most popular items for dogs and cats. The owners of these pet types make up far more of their repeat customer base compared to other pet types and would be more likely to subscribe than fish or bird owners. This could allow PetMind to collect more data and initial feedback from customers on the service, giving insight before deciding to expand the service to other pet types and/or products.

Lastly, the idea of a subscription-based service may appeal to many customers who might not have considered coming back to PetMind in the past, an issue PetMind seems to be struggling with. Plenty of pet owners today would be willing to pay for the convenience of having their pet supplies delivered to their doorstep than head to the store.

Count of product category re-buys per pet type