

# notebook

September 11, 2022



## 1 Introduction:

**PetMind** is a pet goods shop with nationwide presence in the United States. The company wants to cut its customer retention costs by strengthening brand loyalty. Therefore, in the next three months, the strategy is establish a monthly pet box subscription. PetMind marketing team prepared a list of popular products for the pet box subscription. # Purpose: The chief marketing officer of PetMind wants to know if the list of popular products for the pet box subscription should only contain products that have been purchased several time. # Questions: - 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?

## 2 Analysis

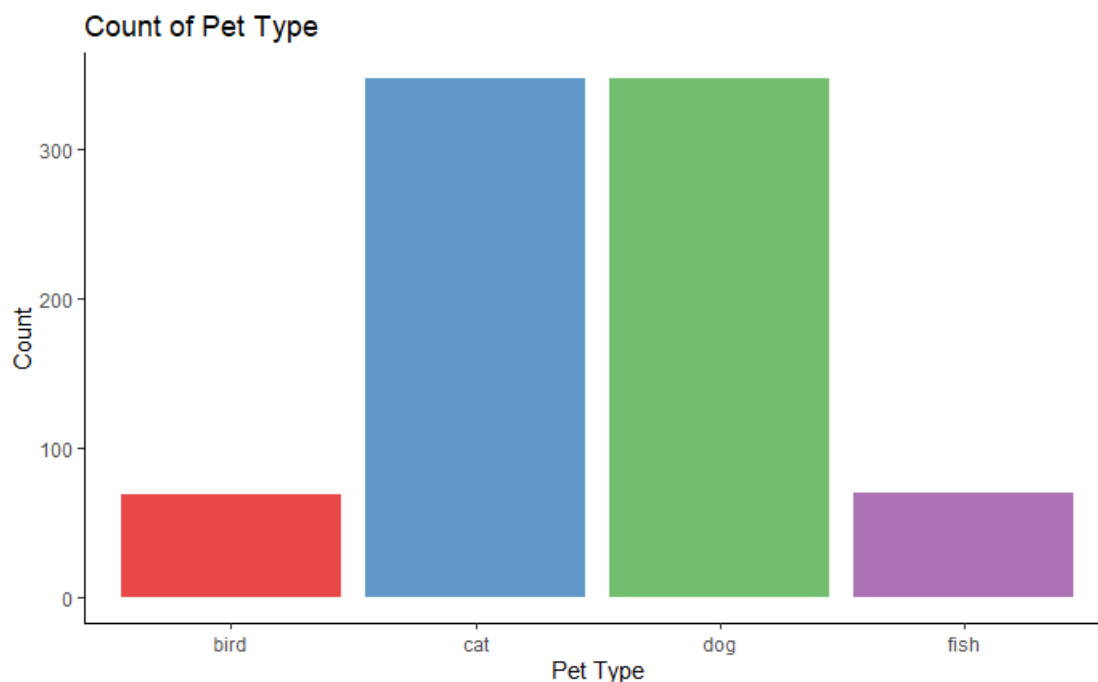
### 2.1 Data Validation

After uploading the data and the necessary packages, the `str()` function is used to have glimpse at the data. The data contains 879 observation and 9 variables. - **product\_id** and **vendor\_id**: Since the data has 879 rows, counting the number of unique values **product\_id** and **vendor\_id** helps determine if product id and vendor id are repeated or not. Indeed each has their unique identifier. Also, to meet the characteristics of **product\_id**, it was converted to character using `as.character()`. - **Sales** has dollar sign. Thus, for easier and accurate analysis it was removed. Then, sales converted to numeric. - **price** is also numeric so no further manipulation was needed. We investigated if the data contains 11 product categories, 5 pet sizes, customer rating of the product on 10 point scale, determined number of pet types, `re_buy` either 1 or 0. - **11 product category**: is character with these categories: “Equipment”, “Toys”, “Snack”, “Supplements”, “Bedding”, “Medicine”, “Housing”, “Food”, “Clothes”, “Accessory”, “Grooming”; - **5 pet sizes**: is character with these categories: “small”, “large”, “extra\_small”, “medium”, “extra\_large”; - **6 pet types**: is character with these categories: “fish”, “cat”, “hamster”, “dog”, “bird”, “rabbit”; - **ratings**: from 1 to 10, converted to numeric; - **re\_buy**: is binary, converted to factor.

The marketing team requested to select only **cat**, **dog**, **bird** and **fish** products. Therefore, **hamster** and **rabbit** were removed from the analysis. We investigated on why the marketing team wanted only four pets for the subscription box, and we found out that hamster and rabbit products were not frequently bought by customers. After cleaning the data, the new data contains **833 observations**. We checked for missing values and there was none.

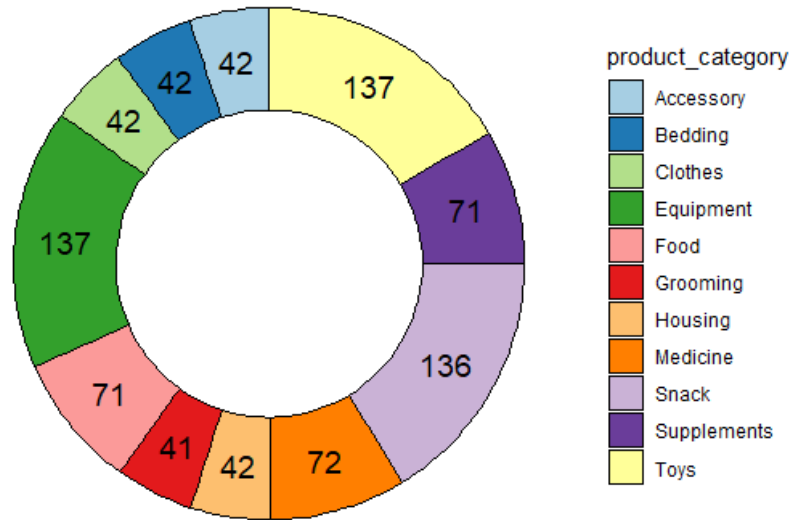
### 2.2 Data Discovery and Visualization

Before answering the questions of the marketing team, let's get to know the data better:



Based on the bar chart which gives the count of each pet category, most customer bought products for dogs and cats equally.

Product Category



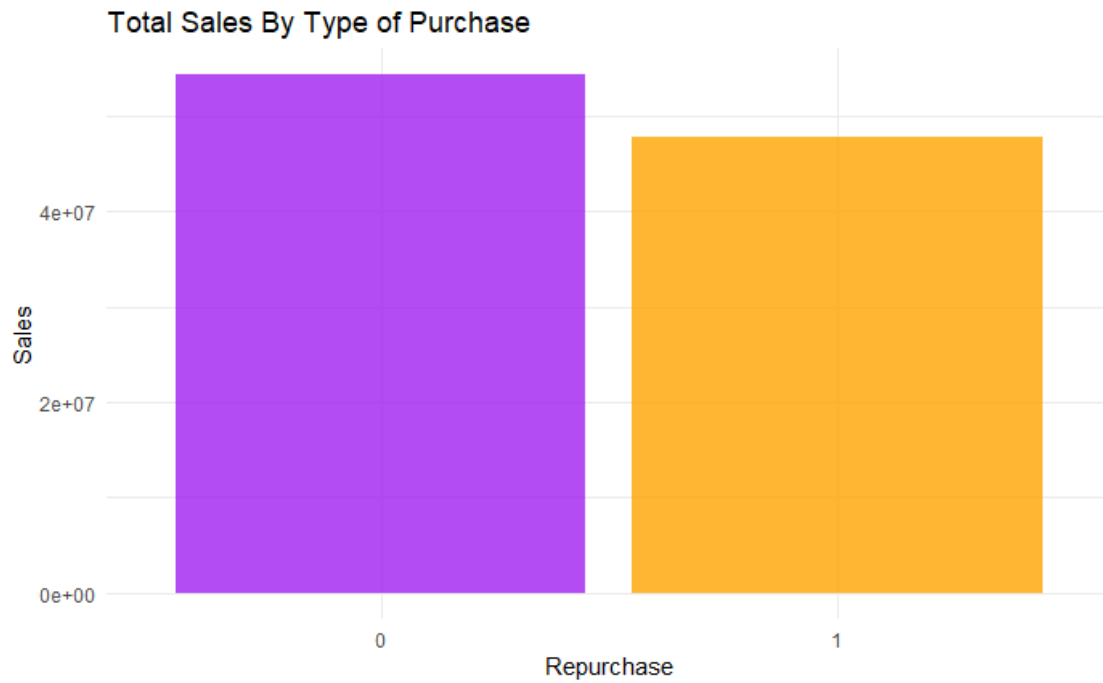
The donut chart reveals that the most common product categories are **equipment, snacks, and toys**. Whereas, the least common are grooming, housing, clothes, bedding and accessory.

### 2.2.1 How many products are being purchased more than once?



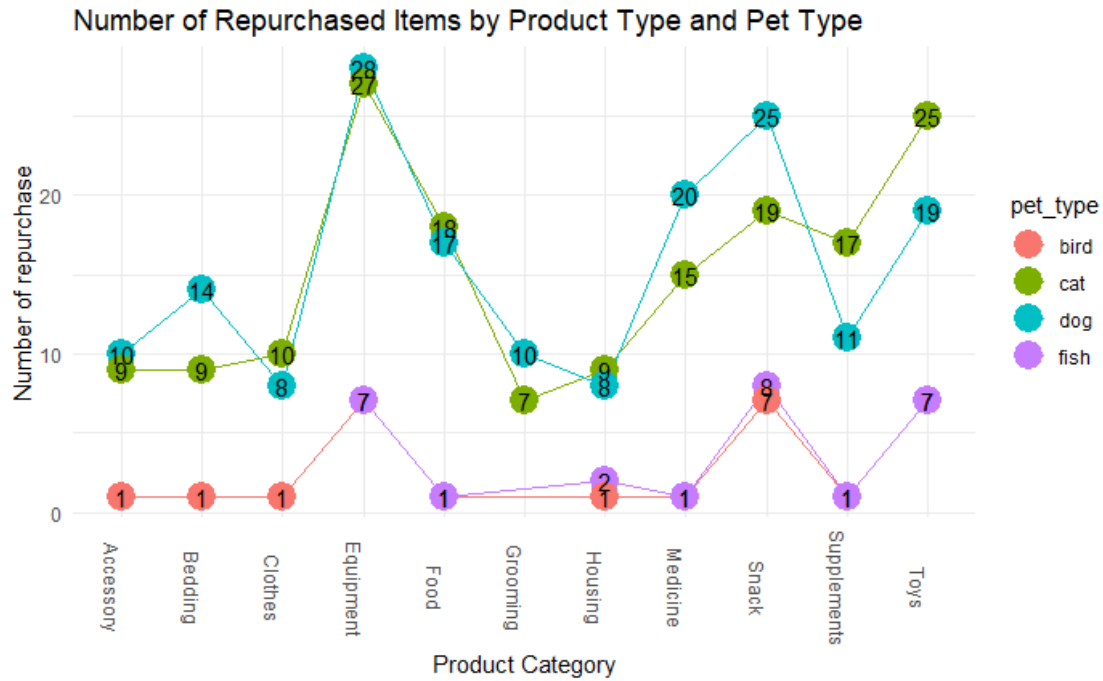
*By creating a **bar plot** for the **re\_buy** features, we discovered that less than half of the products are purchased again. Exactly, 390 products are repurchased, compared to 443 that were not repurchased.*

### 2.2.2 Do the products being purchased again have better sales than others?



Using a **column chart** we determined that products that were purchased only once generated 51,125,000 USD in total sales, while product that were purchased more than once generated cumulative sales of 45,587,000 USD. The products that were not repurchased produced more sales.

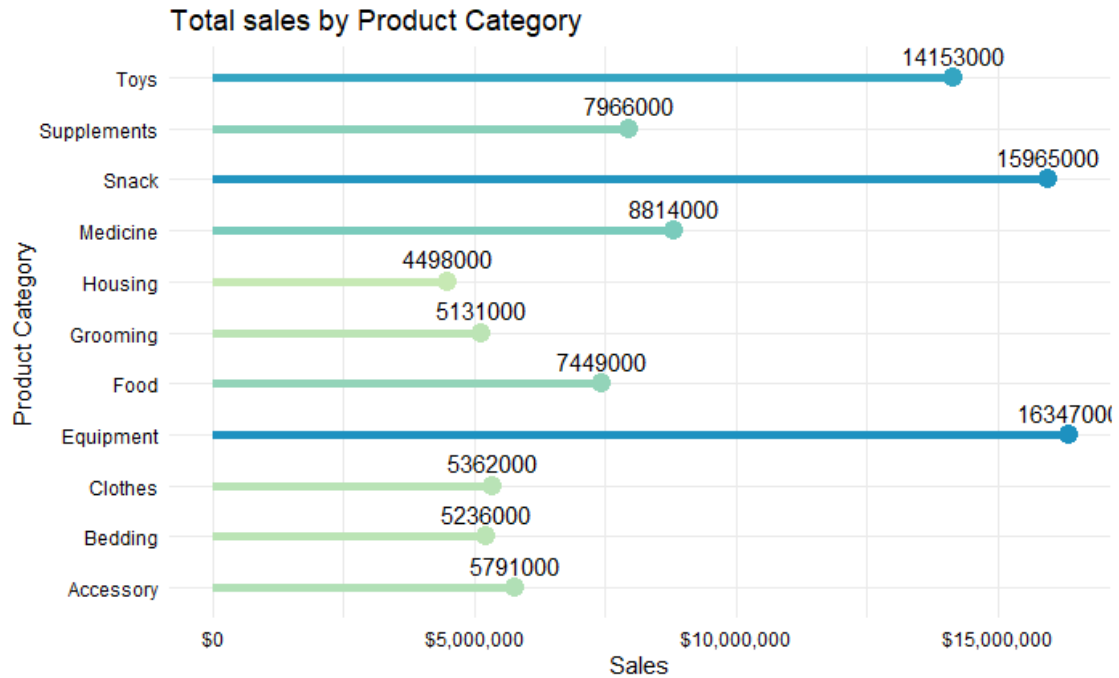
### 2.2.3 What products are more likely to be purchased again for different types of pets?



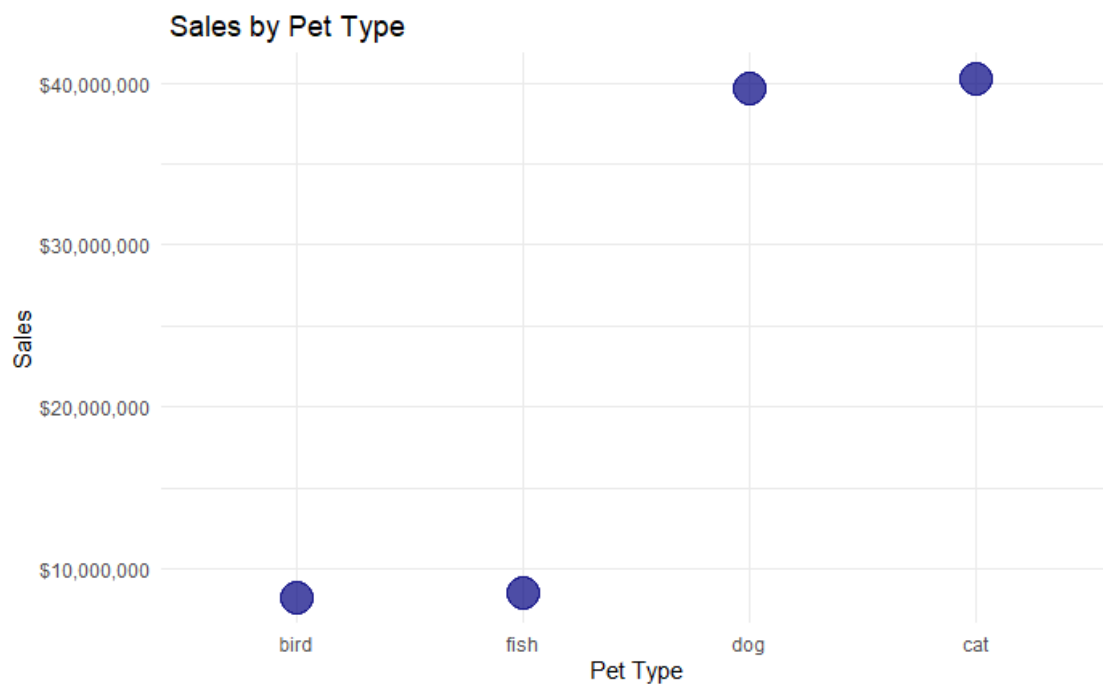
After counting the number of repurchased products for each pet type, a line graph is created. Each line represent a pet type. On a first look on the **line graph** and as expected **dog** and **cat** are dominating **bird** and **fish**.

- **Dog:** most repurchased items are **equipment** with (28), and **snack** (25);
- **Cat:** most repurchased items are **equipment** (27), and **toys** (25);
- **fish:** most repurchased items are **snack** (8), **equipment**, and **toys** (7);
- **bird:** most repurchased items are **equipment**, **snacks**, **toys** (all 7);

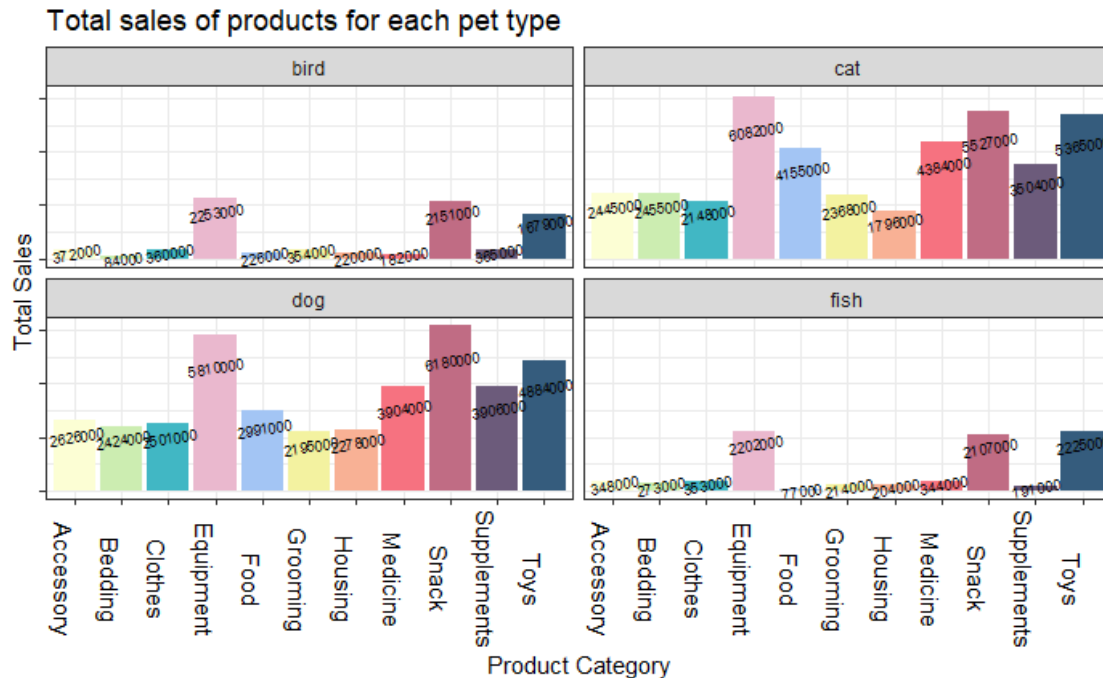
*Snacks, toys and equipment are also dominating other product categories just like product category graph showed .*



Overall **equipment, snacks, and toys** brought in the best sales.

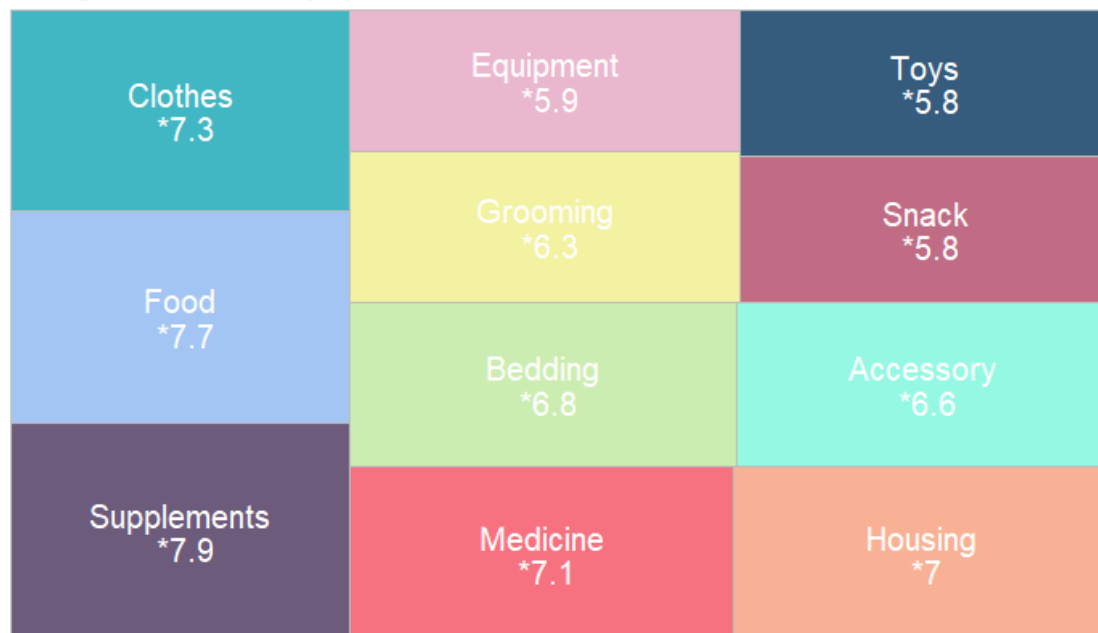


As expected, dog and cat generated most sales with 40,229,000 USD for cats products, slightly less so 39,699,000 USD for dogs, 8,538,000 USD for fish and 8,246,000 USD for birds.



- The sales for **cat** and **dog** is more than other pets, which is expected since most customer of PetMind are owners of cats and dogs.
- **All pet products** resulted great sales with **equipment, snacks and toys**.
- Supplements, food and medicine resulted good sales for **dogs and cats** products.

#### Rating of Product Category



Average rating of all products is 6.7 which means overall the customers are satisfied with PetMind products. The highest rated product categories are the once that were not frequently bought by cus-



tomers. Average rating for *Supplements 7.9, food 7.7, clothes 7.3, medicine 7.1, housing 7*. While, the most bought categories *equipment, snacks and toys on average are rated 5.9, 5.8 and 5.8 respectively*.

## 2.3 Results and Key findings

- More than half of the products are not repurchased, and generated more sales.
- The majority of customers of PetMind are pet and dog owners.
- Most common product categories are equipment, toys and snacks also with the most sales for all types of pets.

## 3 Conclusion

*In this report we answered the following three main questions* - How many products are being purchased more than once? => 390 product - Do the products being purchased again have better sales than others? => No they do not - What products are more likely to be purchased again for different types of pets? => equipment, toys, and snacks for all types of pets *To summarize I would recommend PetMind to:* - Include products that were repurchased and not repurchased, because the products that were purchased only once are more frequent and generated more sales. - Create subscription box for all pets and should include equipment, toys, and snacks. If the team wanted to include more categories to the box subscription I would advise them add medicine, supplements, food for cats and dogs not only did they generated good sales they were also rated on average greater than 7.0. Also, they should stay clear of accessory, bedding, clothes and housing.

**Markdown file** R code: <https://drive.google.com/file/d/1QJQpwEcCrA5an1VWhR5LAu1O-2EW8OgJ/view?usp=sharing>