




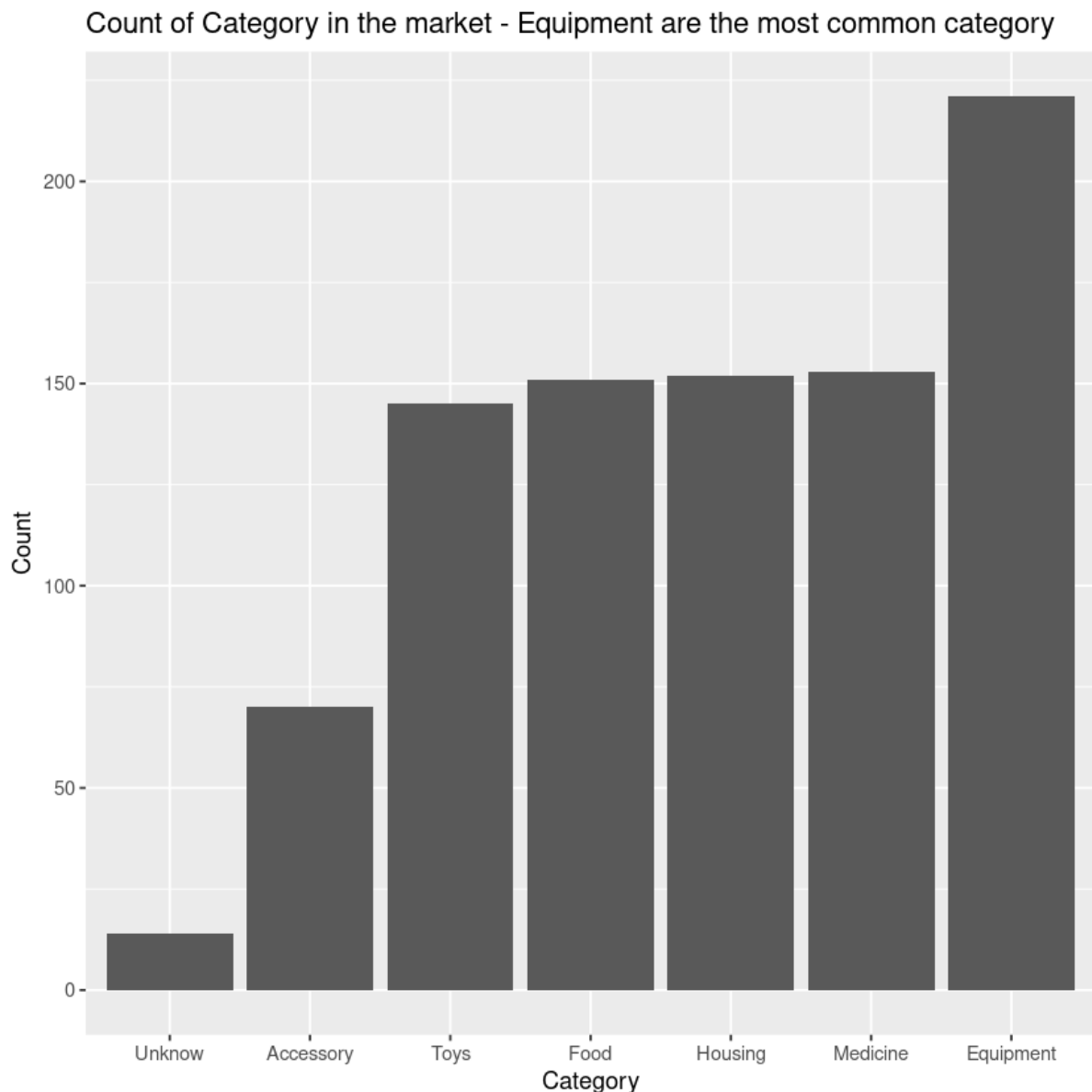
# DA101 - PET\_SUPPLIES\_2212

 Owner	 Khanh Huyen
 Tags	

## TASK 1

1. Product\_id: Missing values are not possible due to the database structure.
2. Category: There were 6 unique values. 25 values (-) were missing. The missing values were replaced with "Unknow" as per the data description.
3. Animal: There were 4 unique values that matches the four given in the data dictionary. There were no missing values so no changes were made to this column.
4. Size: There were 3 unique values, no missing value. However, "Small" including "small" and "SMALL", "Medium" including "medium" and "MEDIUM", "Large" including "large" and "LARGE"; values that do not match the description have been converted is "Small", "Medium", "Large".
5. Price: There were 150 missing values("unlisted"). The missing values were replace with the overall median price and round to 2 decimal places.
6. Sale: There were no missing values, no changes were made to this column.
7. Rating: There were 150 missing values ('NA'), and the missing values were replace with 0 as per the data description.
8. Repeat\_purchase: There were no missing values, no changes were made to this *column*.

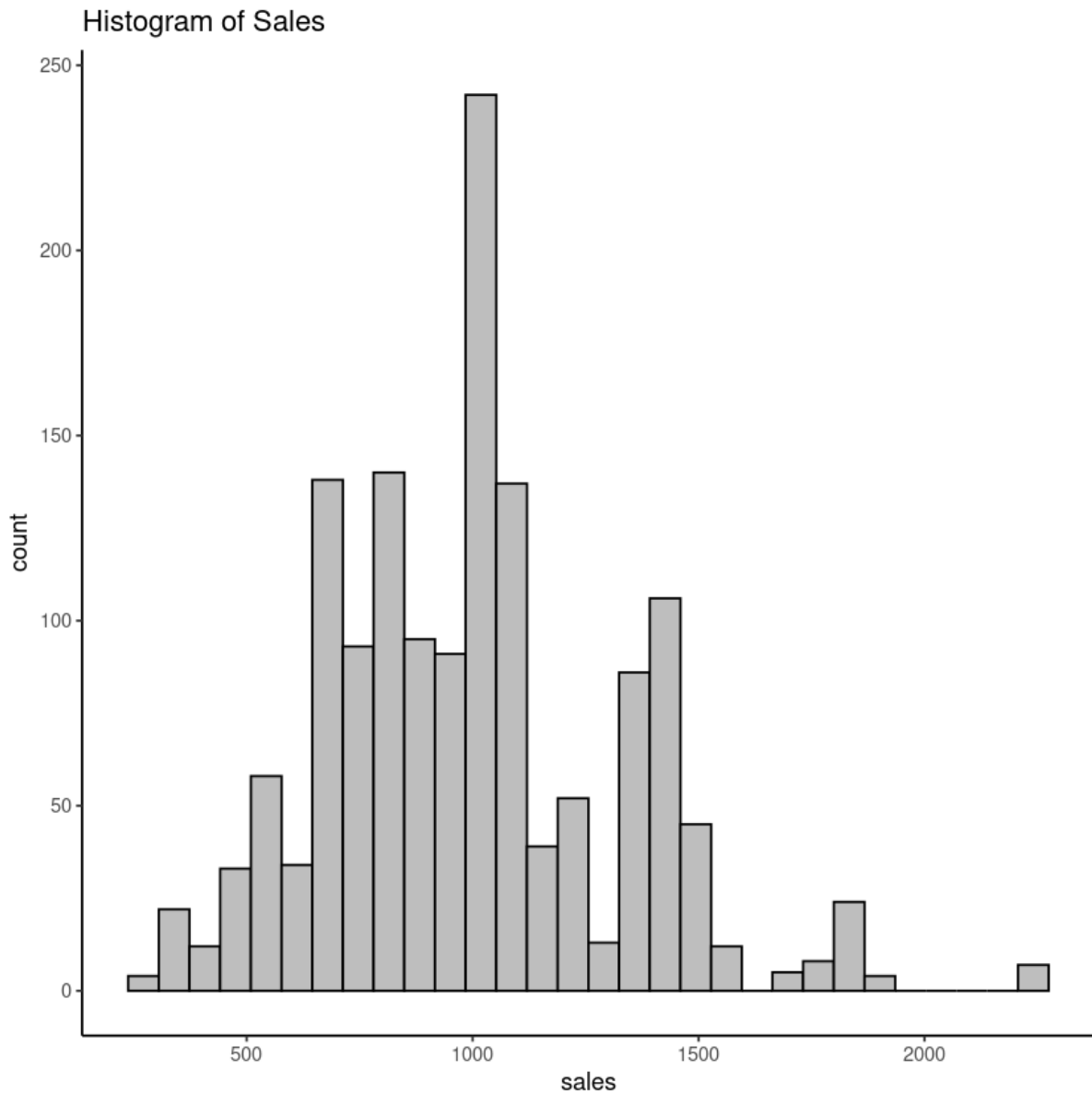
## TASK 2



There are six possible categories in this data, the most popular being Equipment with 221 acquisitions. Next are Medicine (153), Housing (152), Food (151) and Toys (145) repurchases. These categories do not have too much difference in the number of repurchases, the lowest is Accessory (70). acquisition times. PetMind has 2 options: (1) improve the Accessory category (2) Continue to promote the Equipment category.

However, it is found that the Accessory category has less perishable and durable characteristics, thus reducing the frequency of repurchasing the product. So PetMind should promote the Equipment category with promotional campaigns, sell more products, and boost revenue.

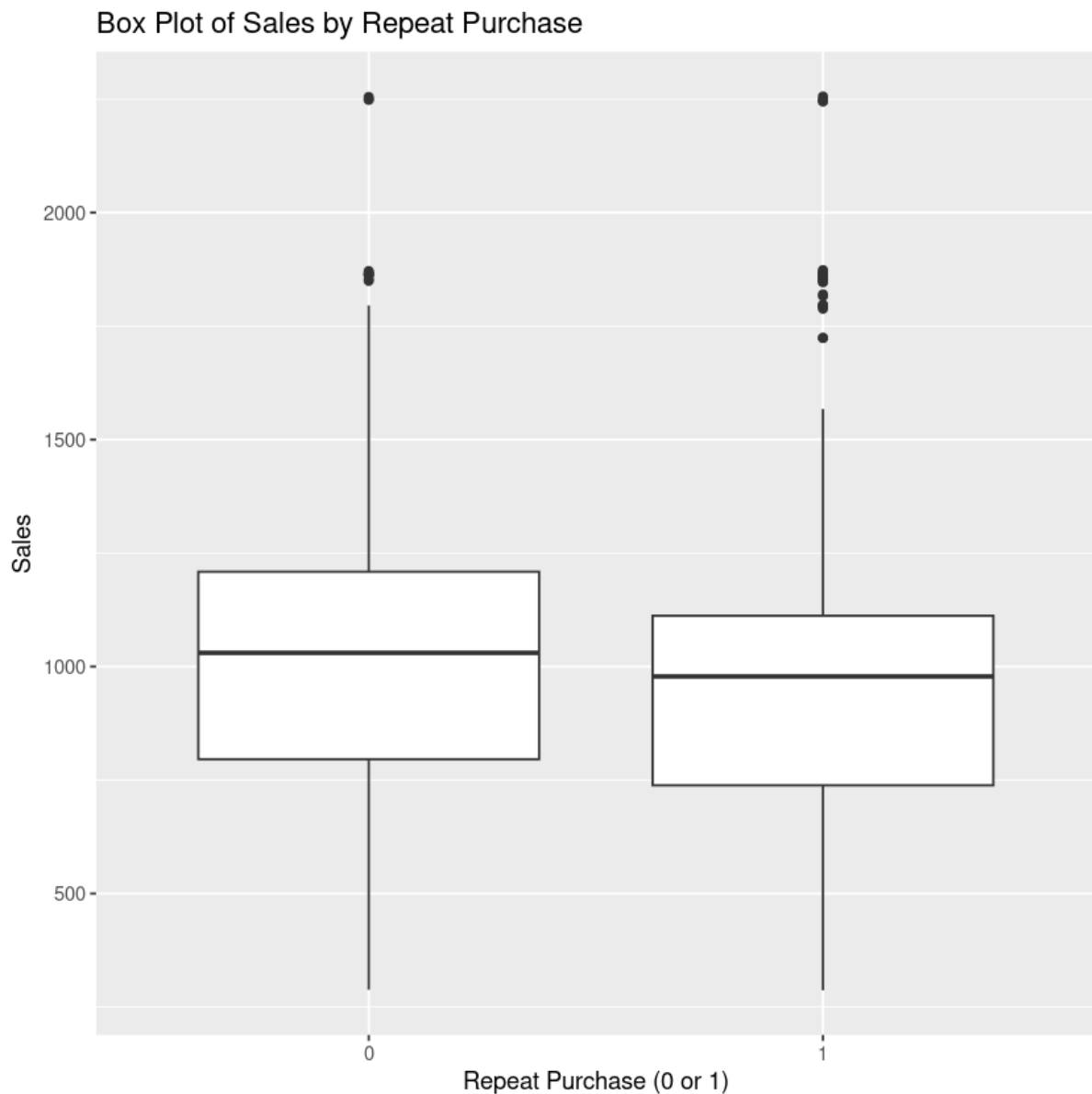
## TASK 3



The graph is skewed to the right, the sales of products with sales around 1000 are the highest, the two adjacent ranges are even, however products with sales around 1400 are unexpectedly low. There are some exceptions that receive more than 2000 sales but this is very rare.

To boost sales in the last months of the year, the team should target products with sales around 1000 but note that it is necessary to pay attention to products with sales around 1400 low.

## TASK 4



From the chart we can see the relationship between sales and repeat\_purchase. Products with sales of 1000 are repurchased less than products with sales of 1000 that are not repurchased.

Besides, the outlet value of repurchased products is higher than that of non-repurchased products,

Suggestion: focus on 1000 sales products, especially actively communicate with 1000 sales products and are repurchased regularly; Besides, find out which products with sales of 1000 cannot be bought again and why.