Tasks :-
Submit your answers directly in the workspace provided.

Data validation

- The original dataset: pet_supplies_2212.csv and stored it as cleanedpetsupplies.csv.
- **Worth noticing :-**
  - This dataset has 8 columns and 1500 unique rows.
  - There are no duplicates.
  - 150 rows or 10% of dataset rows contain BOTH **"unlisted"** in price column and **NA** in rating column, they will be cleaned per the data description, none will be removed.
  - The visualizations were designed with consideration for individuals with color blindness.

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.

1. Category column: converted 25 missing “category” data to “Unknown” per the data description.
2. Animal column : We do not have anything other than this type so we are good to go.
3. Size column : had half of the data all caps and half lowercase which was returning 3 distinct inputs; converted all to proper case, leaving three distinct sizes as per the data description : Small, Medium, Large
4. Price column : converted to float data type rounded to 2 digits to perform aggregations, replaced 150 "unlisted" values to 'overall median price' as per the data description
5. Sales column : converted to float data type rounded to 2 digits.
6. Rating column : replaced 150 missing values as ”NA” with 0 per the data description
7. Repeat_purchase column : do not have any missing value so kept as is.

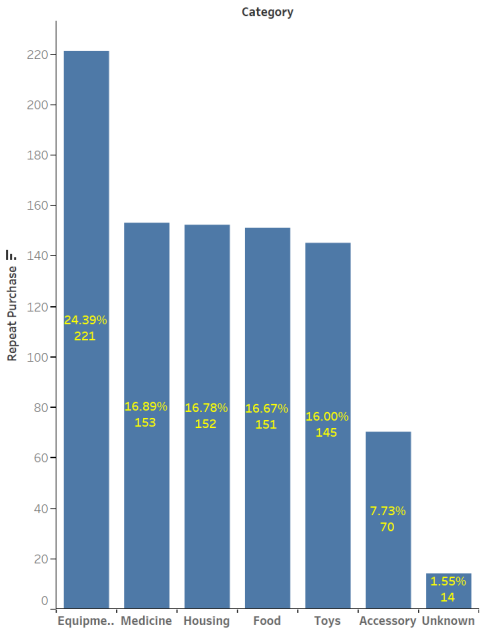
2. Create a visualization that shows how many products are repeated purchases. Use the visualization to:
a. State which category of the variable repeat purchases has the most observations
b. Explain whether the observations are balanced across categories of the variable repeat purchases

PetMind has been trying to increase their repeat business. This approach seems like an intelligent business move. I have explored the data and will highlight how the approach is working thus far.

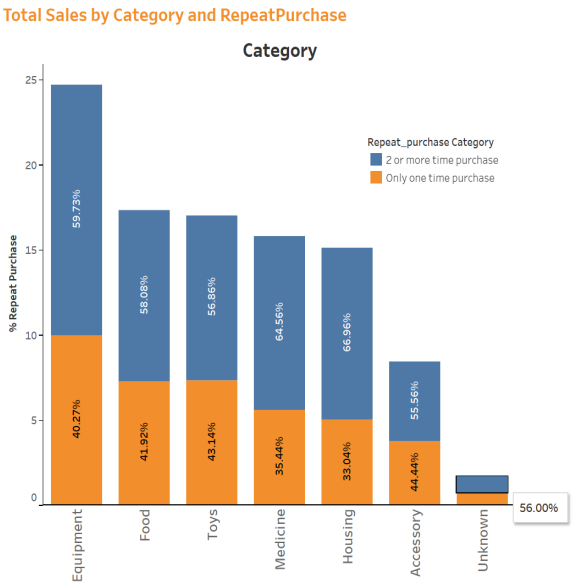
Repeat_purchase Category	% cal along Table (Down)	Count of Repeat Purchase
2 or more time purchase	60.40%	906
Only one time purchase	39.60%	594

-

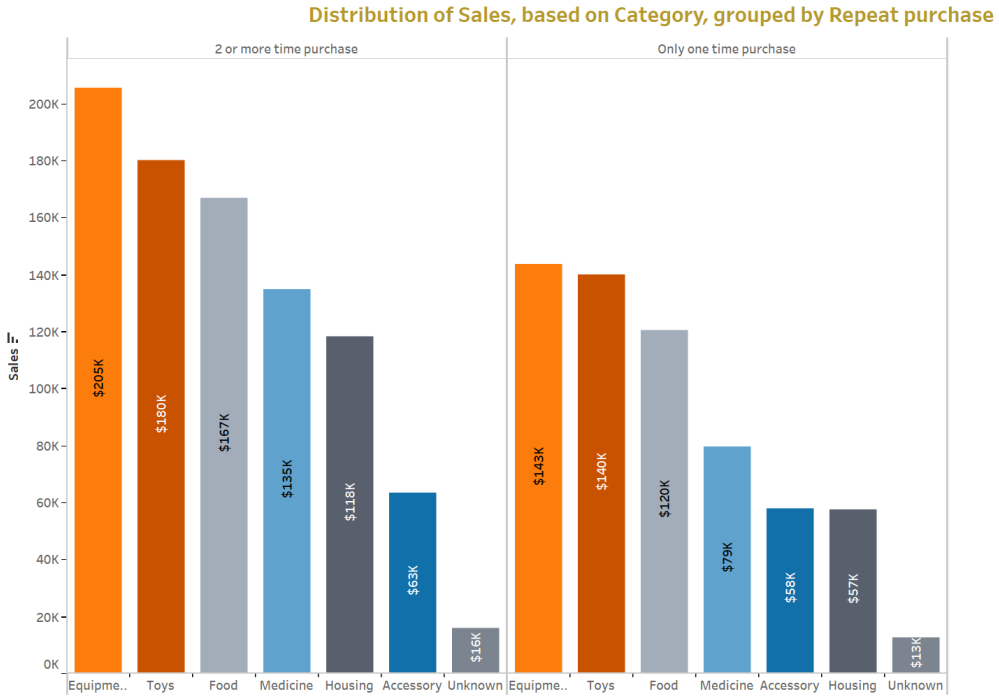
- Equipment ranges a little higher with the 221 observation count which is 24.39% of the total repeat purchase and represents 59.73% of repeat purchase within its category.



- However, items in the Medicine, Housing, Food and Toys categories are all maintaining a consistent repeat count of sales 153, 152, 151 and 145 respectively of the total repeat purchase and represents 64.56%, 66.96%, 58.08% and 56.86% respectively of repeat purchase within its category.
- Accessory is the category with only 70 observations which makes 55.56% of repeat purchase within its own category.
- .Unknown category has the lowest observations (14) which is 56% of its own category.



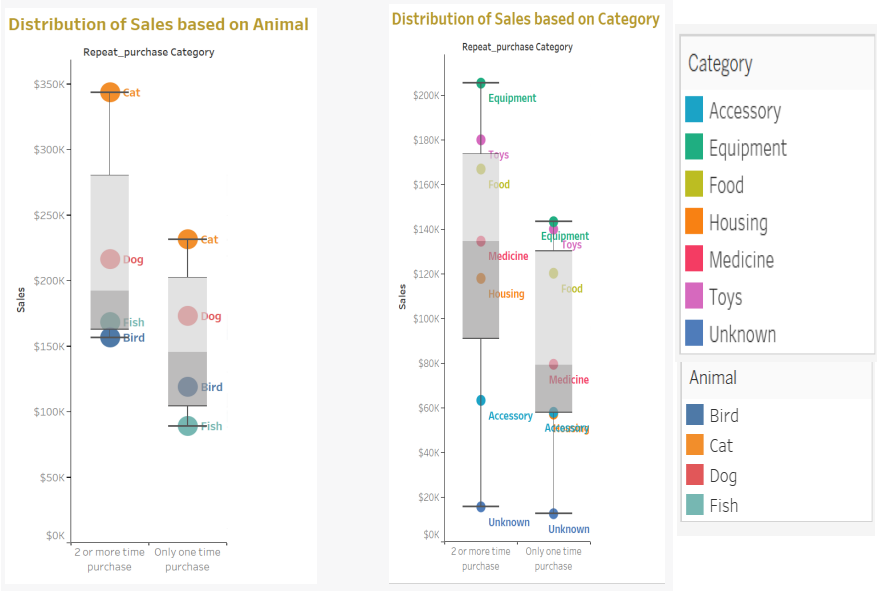
3. Describe the distribution of all of the sales. Your answer must include a visualization that shows the distribution.

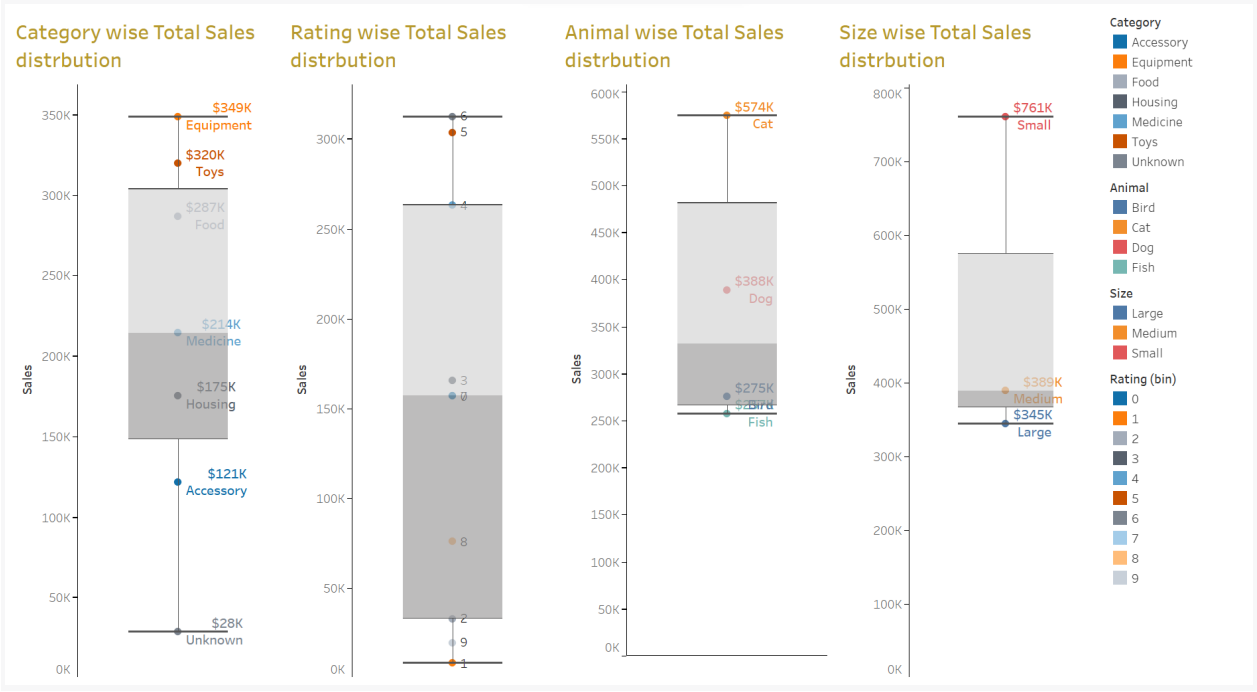


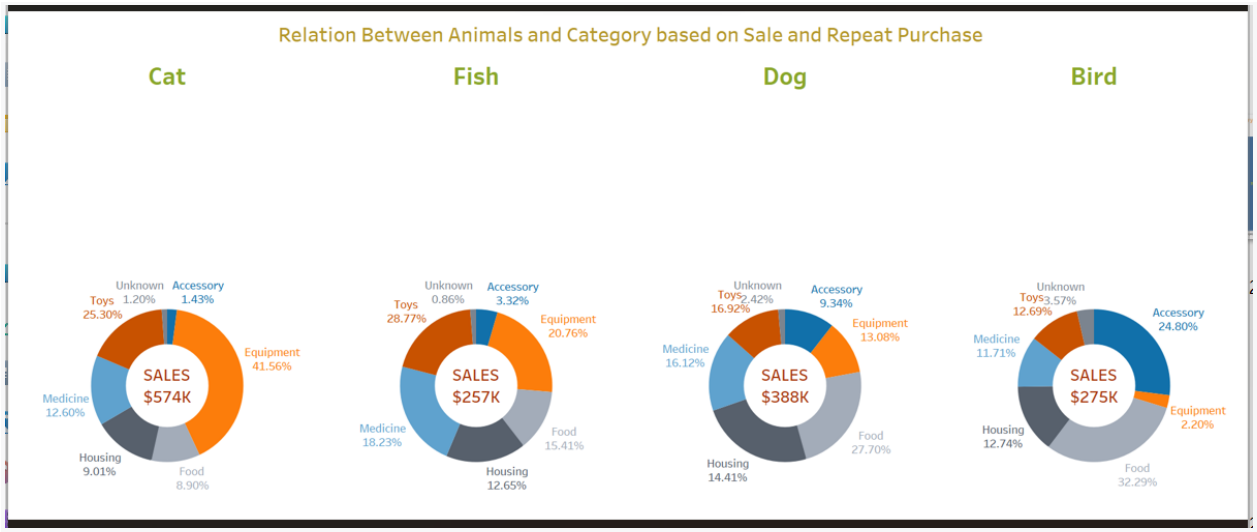
According to the information provided, the luxury items category includes Toys while Food is classified as an everyday item. Looking at the graph above, it is apparent that Equipment is the top-selling item in both bins. However, it's worth noting that Toys and Food are closely ranked, and are in competition with each other. Despite the company's aim to increase sales of everyday products, it should not be disheartened by the fact that luxury items such as Toys are leading in sales. Instead, the company can view the Toys category as a wildcard option, in case of future needs.

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

PetMind's strategy of promoting repeat purchases, particularly for everyday items such as food, has proven successful, with such sales accounting for a majority of total sales. The gap between repeat and non-repeat purchases is significant, with a difference of approximately 20% in total sales. Repeat food sales have been particularly positive for PetMind, as have toy sales, which have shown a considerable potential for growth, especially in the cat category. These results demonstrate that the company's focus on promoting repeat sales of everyday items has been effective in driving revenue. Going forward, PetMind should continue to pursue this objective, and could consider expanding their focus to include repeat purchases of select luxury items. Overall, this data shows that products purchased repeatedly generate better sales than others.


Distribution of Sales based on Animal
Repeat_purchase Category
Sales
$350K
$300K
$250K
$200K
$150K
$100K
$50K
$0K
Cat
Dog
Fish
Bird
2 or more time purchase
Only one time purchase
Distribution of Sales based on Category
Repeat_purchase Category
Sales
$200K
$180K
$160K
$140K
$120K
$100K
$80K
$60K
$40K
$20K
$0K
Equipment
Toys
Food
Medicine
Housing
Accessory
Unknown
2 or more time purchase
Only one time purchase
Category
Accessory
Equipment
Food
Housing
Medicine
Toys
Unknown
Animal
Bird
Cat
Dog
Fish


Category wise Total Sales distrbution
Sales
350K
300K
250K
200K
150K
100K
50K
0K
$349K
Equipment
$320K
Toys
$287K
Food
$214K
Medicine
$175K
Housing
$121K
Accessory
$28K
Unknown
Rating wise Total Sales distrbution
Sales
300K
250K
200K
150K
100K
50K
0K
6
5
4
3
8
2
9
1
Animal wise Total Sales distrbution
Sales
600K
550K
500K
450K
400K
350K
300K
250K
200K
150K
100K
50K
0K
$574K
Cat
$388K
Dog
$275K
Fish
Size wise Total Sales distrbution
Sales
800K
700K
600K
500K
400K
300K
200K
100K
0K
$761K
Small
$389K
Medium
$345K
Large
Category
Accessory
Equipment
Food
Housing
Medicine
Toys
Unknown
Animal
Bird
Cat
Dog
Fish
Size
Large
Medium
Small
Rating (bin)
0
1
2
3
4
5
6
7
8
9


Relation Between Animals and Category based on Sale and Repeat Purchase
Cat
Unknown 1.20%
Accessory 1.43%
Toys 25.30%
Equipment 41.56%
SALES $574K
Medicine 12.60%
Housing 9.01%
Food 8.90%
Fish
Unknown 0.86%
Accessory 3.32%
Toys 28.77%
Equipment 20.76%
SALES $257K
Medicine 18.23%
Food 15.41%
Housing 12.65%
Dog
Unknown 2.42%
Toys 16.92%
Accessory 9.34%
Equipment 13.08%
Medicine 16.12%
SALES $388K
Food 27.70%
Housing 14.41%
Bird
Unknown 3.57%
Toys 12.69%
Accessory 24.80%
Medicine 11.71%
SALES $275K
Equipment 2.20%
Housing 12.74%
Food 32.29%