Reminders…

- Requirements:
  - Answer all the questions in this document
  - When complete, download a PDF copy of this document for future reference. You can also save it in your Google Drive.
  - Don't know how to download it as a PDF? You can find more information about downloading this by clicking here.

Content

Scrub

You are being provided with data about sales at a store. You can access this data by clicking here [ADD LINK] and clicking on Use Template in the upper right corner.

This data has issues that need to be fixed before it can be used for your data analysis project. some issues may include:

- Duplicate records
- Missing values
- Obviously wrong values

Using Google Sheets, review the data, find the issues, and clean the dataset for your use in this project.

What are some errors you identified in this dataset? Identify the column or row where possible. For example: “Duplicate transaction data on rows 45 and 46.” You do not need to include every error, but include at least three.

1. Missing ‘$’ sign in Price
2. Missing Product_Name data from row row 2 to 30
3. Missing Product_Category data from row 2 to 146

How did you fix the errors that you identified in the previous question? For example: “Deleted the duplicate transaction data on row 46.”

1. Format Price as a currency data type
2. Create a helper column in column K called Filled_Product_Name and use VLOOKUP to find the Product_Name with the same SKU.
3. Create a helper column in column L called Filled_Product_Category and use VLOOKUP to find the Product_Category with the same SKU.

Explore

Using the spreadsheet tools in Google Sheets, explore the data. You are encouraged to use spreadsheet functions like AVERAGE and CORREL as well as SQL queries like ORDER BY and LIMIT.

When you have used these tools, create a chart that highlights a relationship you discovered in the data. For instance, you might create a bar chart that shows sales of a particular item in different months of the year to showcase how well it sells in warmer months.

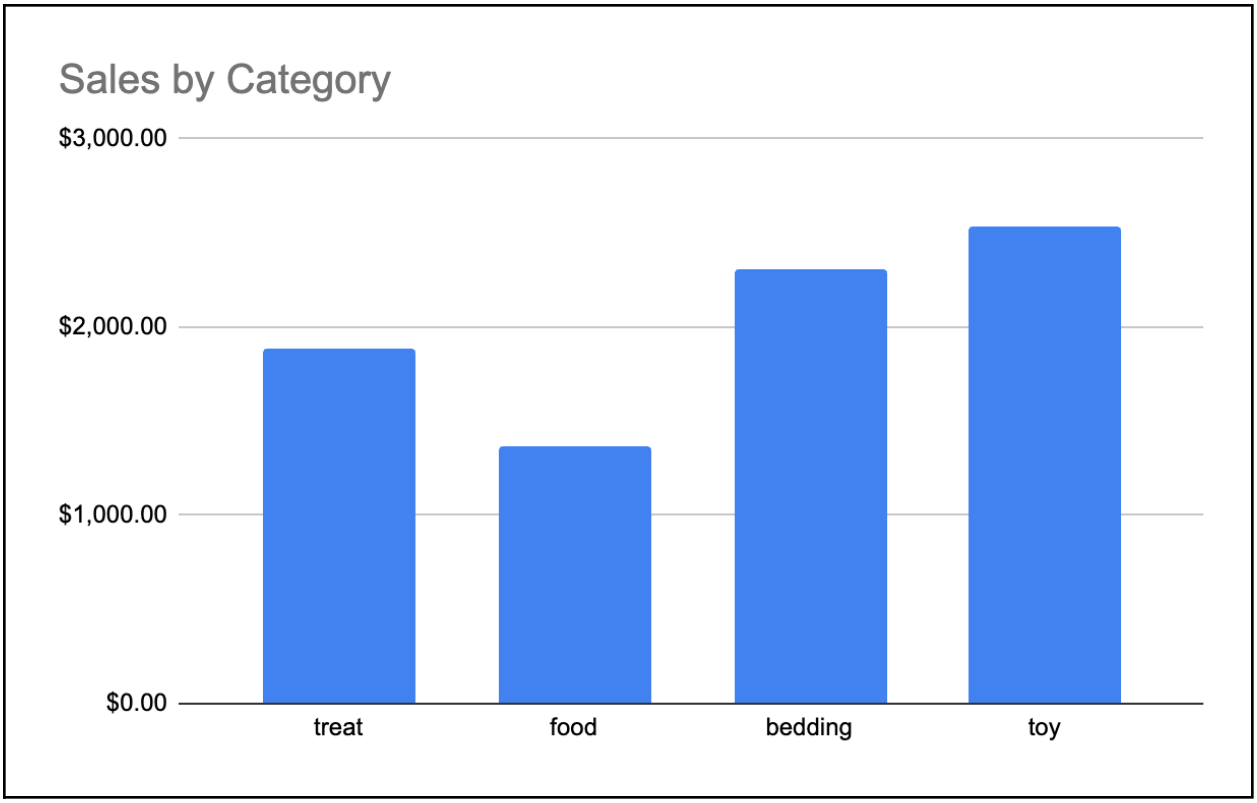
What spreadsheet functions did you use and what results did you get? For example: “The AVERAGE of the “sales” column was $35.55.” You do not need to include every function you used, but include at least three.

1. Using SUMIF, the total sales for treats was $2095.79.
2. The AVERAGE of the “Price” column for treats was $25.25.
3. Using QUERY, the best-selling product was Reddy Beddy, with 75 units sold.

What SQL queries did you use and why? For example: “Used the AND clause to see sales that were for a certain item and above a certain price” You do not need to include every query you used, but include at least three.

1. SELECT D, SUM(H) displays each product and how many units it sold in total.
2. WHERE D IS NOT NULL filters out any rows where the Product_Name is blank.
3. GROUP BY D combines all rows of the same product so we can total their quantities.
4. ORDER BY SUM(H) DESC puts the product with the most sales at the top.
5. LIMIT 1 shows only the first row after sorting, in this case, the top-selling product.

Copy and paste at least one chart into this document that was created from the dataset.



Visualize

Using the dataset you have scrubbed and explored, create a dashboard with at least two charts and at least one interaction.

Copy and paste the URL for your published Tableau Public dashboard

https://public.tableau.com/views/PetStore_17604161371740/PetStoreDashboard?:language=en-GB&publish=yes&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link

Copy and paste an image of the dashboard downloaded from Tableau Public