Content

Scrub

You are being provided with data about sales at a store. You can access this data by clicking here 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.

- Column F was changed into currency.
- Records missing in row 3, 62, 77, etc.
- Records with similar transaction date and customer id on row 3 and 230; row 8 and 164.

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

- Changed the data type by using the formatting options.
- Removing the rows with missing records using the filter function
- Adding missing data using the duplicates available.

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.

The functions used were
- Average - $26.03
- SUM - $7809.65
- Min - $10.80
- Max - $39.55
- Median - $25.48

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.

=QUERY(B2:J301,"SELECT J,COUNT(J) GROUP BY J")
USED THIS QUERY TO SEE THE NUMBER OF PETS BY TYPE TO ANALYZE SALES REQUIREMENT OF EACH PET.
RESULT:

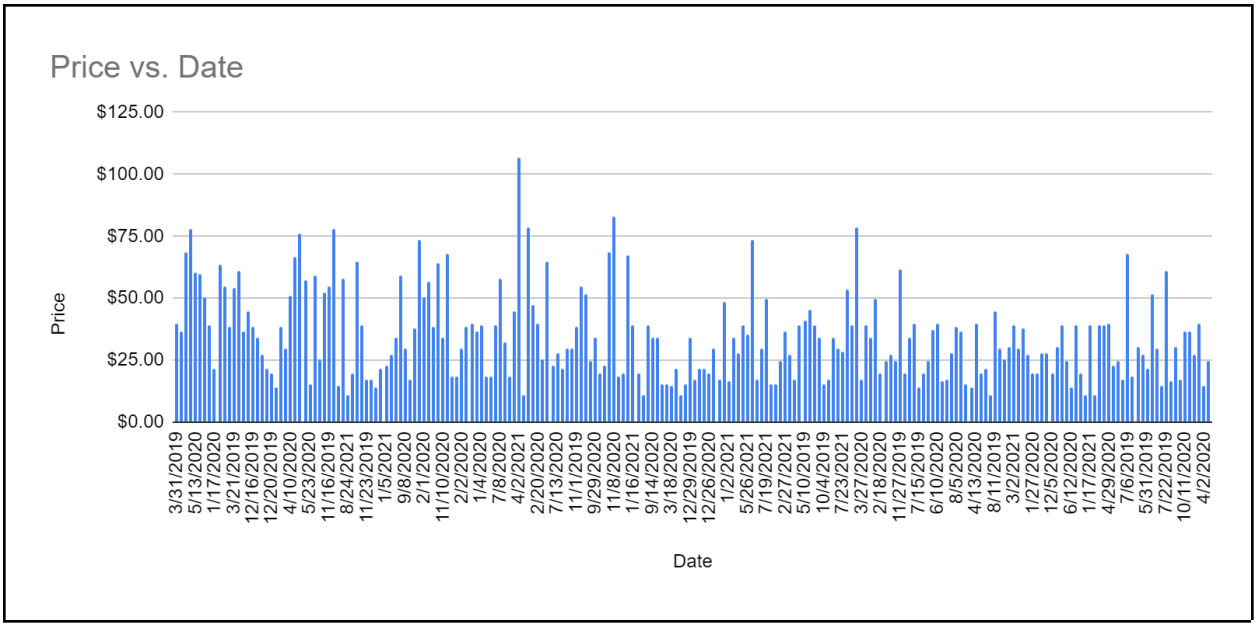
	count
cat	163
dog	135

=QUERY(B2:J301,"SELECT F WHERE F > 26")
TO SEE HOW MANY ORDERS WERE PLACED ABOVE THE AVERAGE PRICE.
RESULT: TOTAL 138 ORDERS ABOVE THE AVERAGE

=QUERY(B2:J301,"SELECT G,COUNT(G) GROUP BY G")
TO SEE WHAT IS THE NUMBER OF ORDERS BY SIZE.
RESULT:

	count
large	22
medium	19
small	30

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.

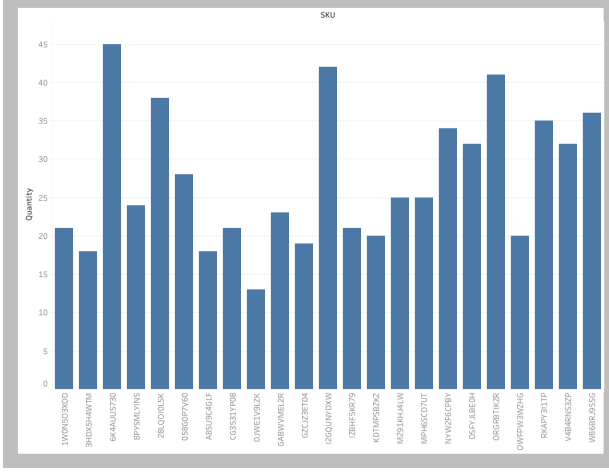
Copy and paste the URL for your published Tableau Public dashboard

https://public.tableau.com/app/profile/parth.punia/viz/Book5_17223485196980/Dashboard1?publish=yes

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

SALES COMPARISON FOR PET STORE

SKU Vs QUANTITY



TIME Vs PRICE