Content

Scrub

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

- Duplicate records
- Missing values
- Obviously wrong values

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.

Missing product names from row 2 to 30. Missing product lines in row 2 and 3. Multiple rows are missing product_categories. There are no duplicate records in the data. There are no obviously wrong values. Multiple rows are present with missing customer_ids. Missing size records.

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

Use filter and group by action to add appropriate values to missing cells. Deleted rows with missing customer_ids. Added N/A at missing size records as the size does not affect the methods of further data analysis.

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.

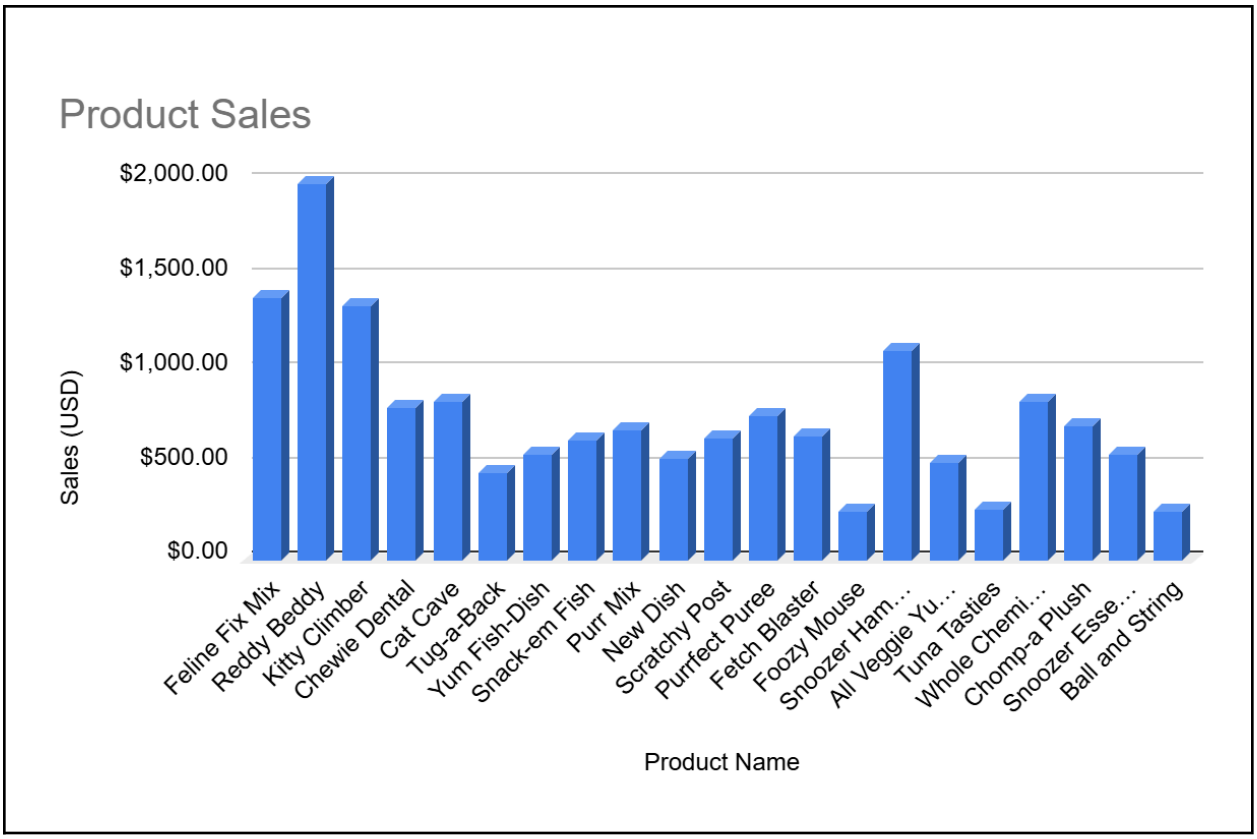
What spreadsheet functions did you use and what results did you get? For example: “The AVERAGE of the “sales” column was $35.55.” You do not need to include every function you used, but include at least three.

The Average and maximum item price was $26.02 and $39.55 respectively. The Average order quantity is 2.1 and the average order value is $54.65. Added a Sales column using Product function and multiplying Item price and quantity sold. Added a new day column to establish the day of the week at which the sale was made, using format and =WEEKDAY functions. Used the CORREL Function to find out correlation constant between sales and week of the day, which turned out to be -0.078, meaning there is little to no correlation between them.

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.

Used “=QUERY('transactions-pet_store-small'!B2:L289,"SELECT D,F,K ORDER BY K DESC LIMIT 10")” query command to figure out the top 10 order values. Used “=QUERY('transactions-pet_store-small'!A2:L289,"SELECT D,COUNT(D) GROUP BY D ORDER BY COUNT(D) DESC")” query command to find out the respective numbers of items sold and arrange them in descending order. Used “=QUERY('transactions-pet_store-small'!A2:L289,"SELECT D, SUM(K) GROUP BY D ORDER BY SUM(K) DESC LIMIT 10")” query command to find out the top 10 selling items.

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


Product Sales
$2,000.00
$1,500.00
$1,000.00
$500.00
$0.00
Sales (USD)
Feline Fix Mix
Reddy Beddy
Kitty Climber
Chewie Dental
Cat Cave
Tug-a-Back
Yum Fish-Dish
Snack-em Fish
Purr Mix
New Dish
Scratchy Post
Purrfect Puree
Fetch Blaster
Foozy Mouse
Snoozer Ham…
All Veggie Yu…
Tuna Tasties
Whole Chemi…
Chomp-a Plush
Snoozer Esse…
Ball and String
Product Name

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/palash.gandhi/viz/SalesDashboard_17618350797700/SalesDashboard?publish=yes

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