

Content

[Scrub](#)

[Explore](#)

[Visualize](#)

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 Data in Customer_ID, Product_Name, Size, Product_Category and Product_Line.
2. Some Product Name requires changes like New Dish
3. Look into Product Category and Product Line for refinement basis

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

1. Either remove it or make an educated guess based on existing data
2. Replace New Dish with a proper name
3. Link product category and product line to respective Product Names

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.

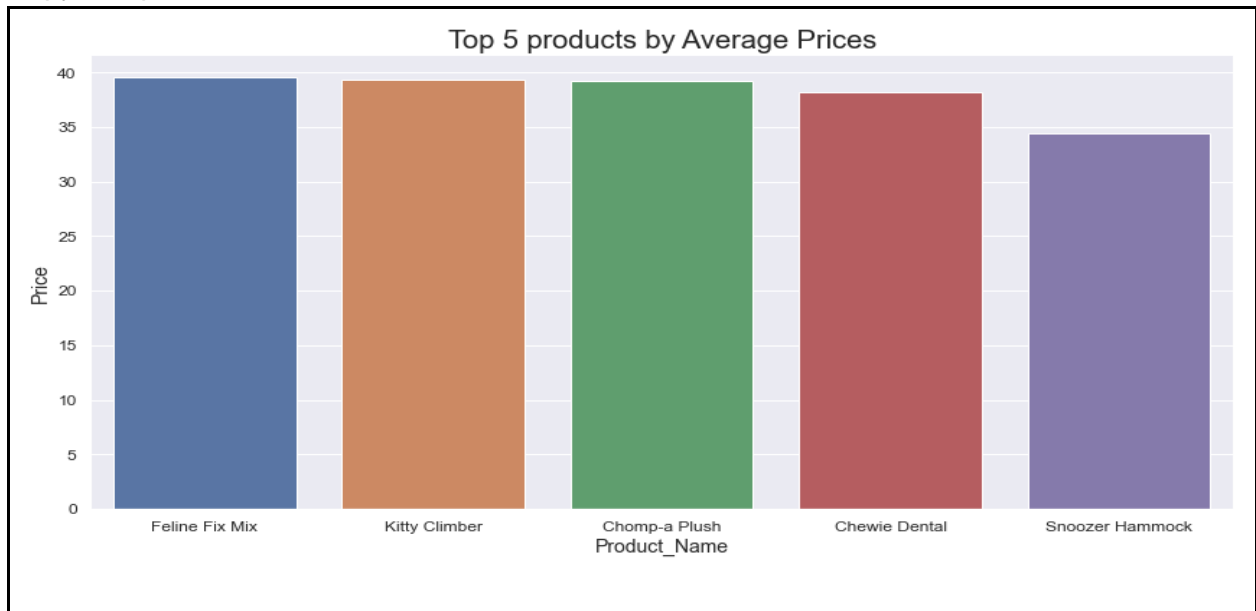
What spreadsheet functions did you use and what results did you get? For example: "The AVERAGE of the "sales" column was \$35.55." You do not need to include every function you used, but include at least three.

```
AVERAGE("Price", "Quantity")
CORREL(("Price", "Quantity"))
MEDIAN(("Price", "Quantity"))
```

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.

```
# Check which product line prices more than $30:
SELECT Product_Line FROM petstore WHERE Price > 30 LIMIT 10;
# Check which sizes where price is less than $15:
SELECT Size FROM petstore WHERE Price < 15 LIMIT 10;
# Select the product name and order the results by the average price of the products:
SELECT Product_Name FROM petstore ORDER BY AVG(Price);
```

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

I am using PowerBI which is equivalent to Tableau

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

