Preppin Data 2022-WK31: Preppin' Parameters

> (*) No parameters but using the notebook filter instead (check the final output section)

📓 Table of Content

🔗 Link to challenge

https://preppindata.blogspot.com/2022/08/2022-week-31-c-preppin-parameters.html

🎯 Requirements

- Input the data
- Split the Product Name field into Product Type and Size
- Only keep the Liquid products
- Total up the sales for each Product Size and Scent for each Store
- Rank each of the Product Size and Scent combinations for each Store
- Only leave the top 10 based on total sales value calculated above
- Round the Sales Values to the nearest 10 value (ie 1913 becomes 1910)
- Create a parameter to select the store (*)
- Ensure the output only contains the chosen store (*)
- Output the data and include the Store Name in the file name (*)

📸 Screenshot solution

The below is a screenshot of what the solution looks like in the notebook. You can view and download the final output here. Check my Github repository for the compiled SQL and a copy of the notebook markdown.

Notebook cells

all_data

```
-- same data as for week 27

SELECT * FROM preppindata.`2022w27_cbsco`;
```

Untitled

Sale_Date	Order_ID	Sale_Value	Product_Name	Store_Name	Region
2022-12-15T00:00:00.000Z	770	143.84	Liquid - 750ml	Chelsea	West
2022-12-15T00:00:00.000Z	770	143.84	Liquid - 1L	Chelsea	West
2022-12-15T00:00:00.000Z	770	143.84	Bar - 2x	Chelsea	West
2022-08-04T00:00:00.000Z	515	90.17	Liquid - 250ml	Chelsea	West
2022-08-04T00:00:00.000Z	515	90.17	Bar - 1x	Chelsea	West

7 columns · 4,063 rows 1.5 seconds (23 minutes ago)

step_1

```
-- split productname into type and size
-- only keep rows for liquid
-- total up sales by size  & scent combo for each store

SELECT
Store_Name,
CONCAT(TRIM(SPLIT(Product_Name,"-") [ORDINAL(2)]),"-",Scent_Name) AS Size_Scent,
SUM(Sale_Value) AS Total_Sales
FROM preppindata.`2022w27_cbsco`
WHERE TRIM(SPLIT(Product_Name,"-") [ORDINAL(1)]) = "Liquid"
GROUP BY Store_Name, Size_Scent;
```

Untitled

Store_Name	Size_Scent	Total_Sales
Chelsea	750ml-Rosemary	1,503.47
Chelsea	1L-Apricot	782.36
Chelsea	250ml-Banana	947.85
Chelsea	50ml-Lavender	401.52
Chelsea	1L-Watermelon	406.31

3 columns · 420 rows 0.8 seconds (23 minutes ago)

step_2

```
-- rank size & scent combo for each store

SELECT
*,
RANK() OVER(PARTITION BY Store_Name ORDER BY Total_Sales DESC) AS Sales_Ranking
FROM step_1;
```

Untitled

Store_Name	Size_Scent	Total_Sales	Sales_Ranking
Shoreditch	250ml-Mint	1,941.99	1
Shoreditch	25ml-Rose	1,847.59	2
Shoreditch	1L-Rose	1,830.12	3
Shoreditch	750ml-Lemon	1,672.88	4
Shoreditch	1L-Lemongrass	1,661.18	5

4 columns · 420 rows 1.3 seconds (20 minutes ago)

checking_distinct_size_scent_combos

step_3

```
-- only keep the top 10
-- round sales value to the nearest '10' value (1913 to 1910)
-- no parameter but table can be filtered on store_name via the notebook
-- https://count.co/notebook/ZVkjh4yRKot?block=DywlXTjSigg

SELECT
Store_Name,
Sales_Ranking,
SPLIT(Size_Scent,"-") [ORDINAL(2)] AS Scent_Name,
SPLIT(Size_Scent,"-") [ORDINAL(1)] AS Size,
ROUND(Total_Sales, -1) AS Sales_Value
FROM step_2
WHERE Sales_Ranking <= 10;
```

Untitled

Store_Name	Sales_Ranking	Scent_Name	Size	Sales_Value
Wimbledon	1	Rosemary	100ml	2,270
Wimbledon	2	Lemon	500ml	1,980
Wimbledon	3	Rose	500ml	1,830
Wimbledon	4	Banana	250ml	1,810
Wimbledon	5	Lemongrass	250ml	1,800

5 columns · 60 rows 1.2 seconds (20 minutes ago)

🏇 Final output

output_filter

Cell/table step_3

Filter records by Store Name

Store_Name	Sales_Ranking	Scent_Name	Size	Sales_Value
Chelsea	1	Watermelon	750ml	2,200
Chelsea	2	Watermelon	500ml	1,850
Chelsea	3	Lemon	25ml	1,790
Chelsea	4	Strawberry	50ml	1,730
Chelsea	5	Mint	50ml	1,520

5 columns · 60 rows 1.2 seconds (20 minutes ago)