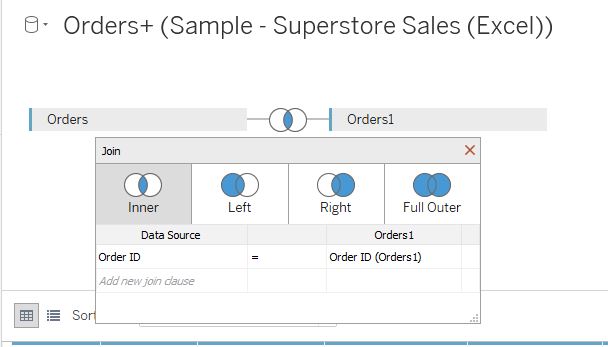
**Project 5 – Market Basket Analysis**

Create a self join, such that both the tables are joined based on Order ID.

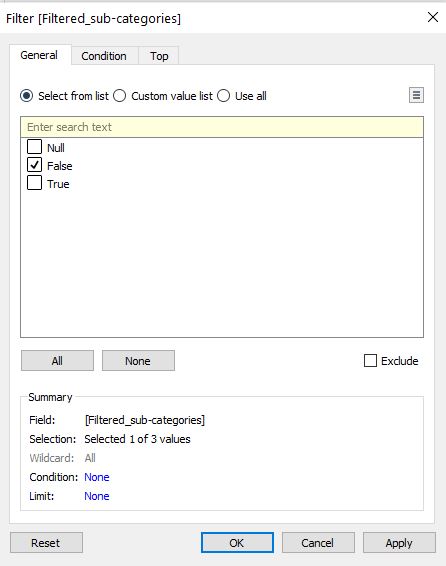


Self-join

Create a calculated field labelled Filtered\_subcategories.



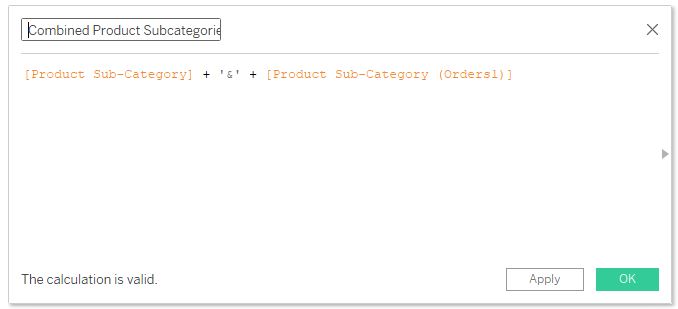
Place this field in the Filter shelf and select False. This will filter out any product subcategories that are the same in both the data sources. It will remove any product subcategory that joins with to a product subcategory with the same name.



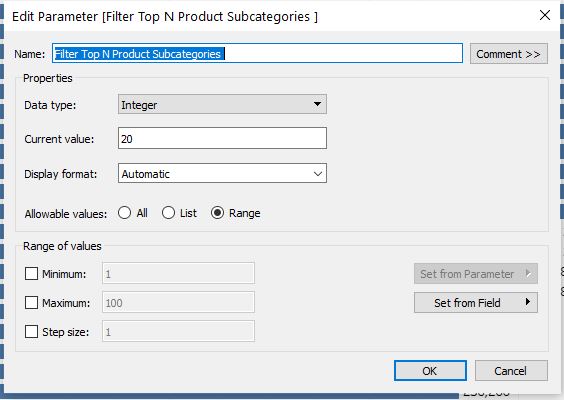
Then we create a total sales calculated field.



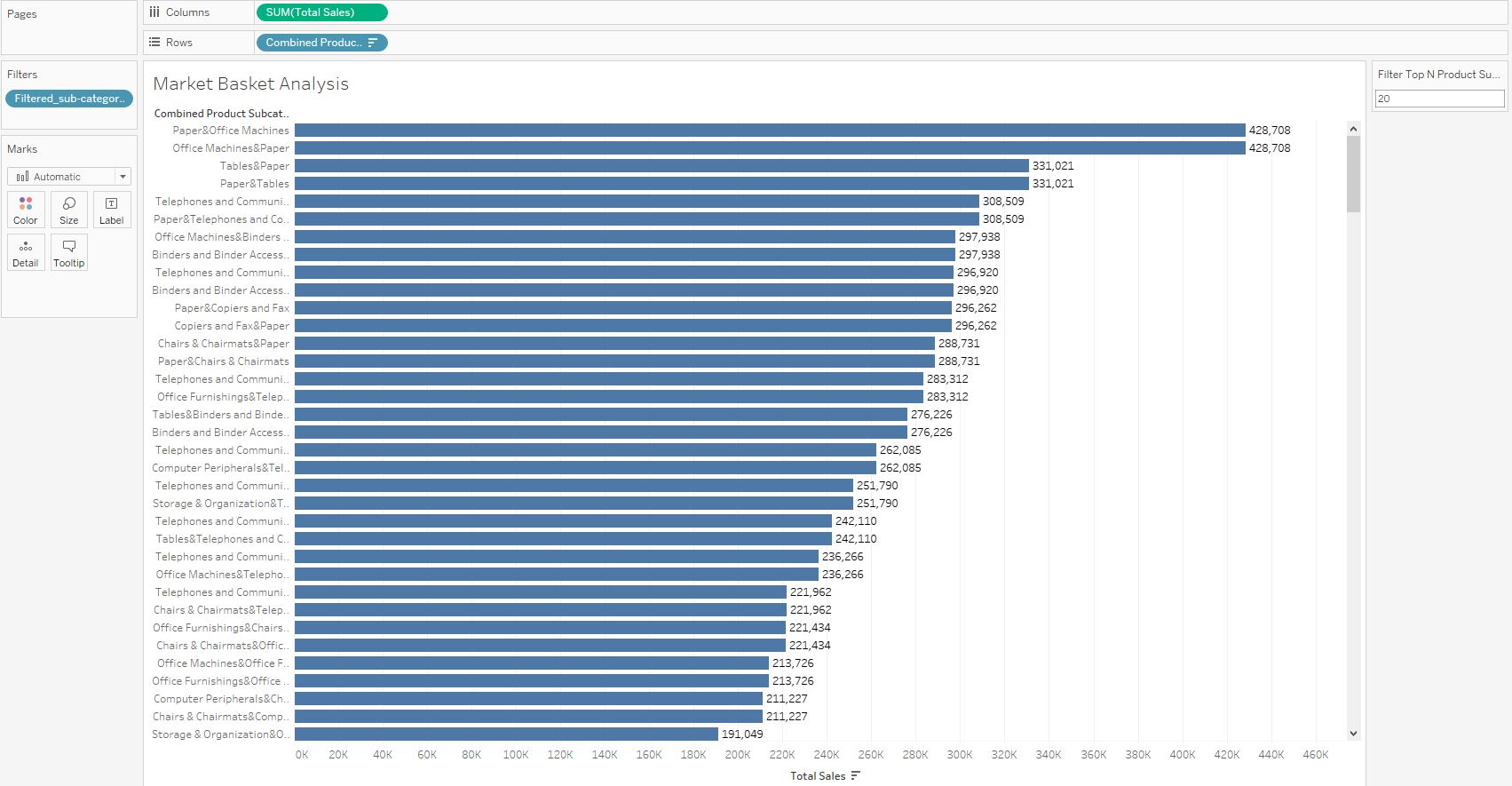
We will now create a calculated field to combine the product subcategories to get a list of the products that are sold together.



Then we create a top n filter to show the top n product sub categories based on total sales.

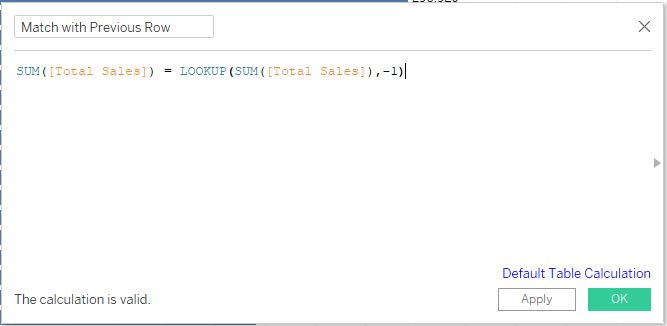


Finally we create a bar chart with the combined product subcategories based on Total Sales.

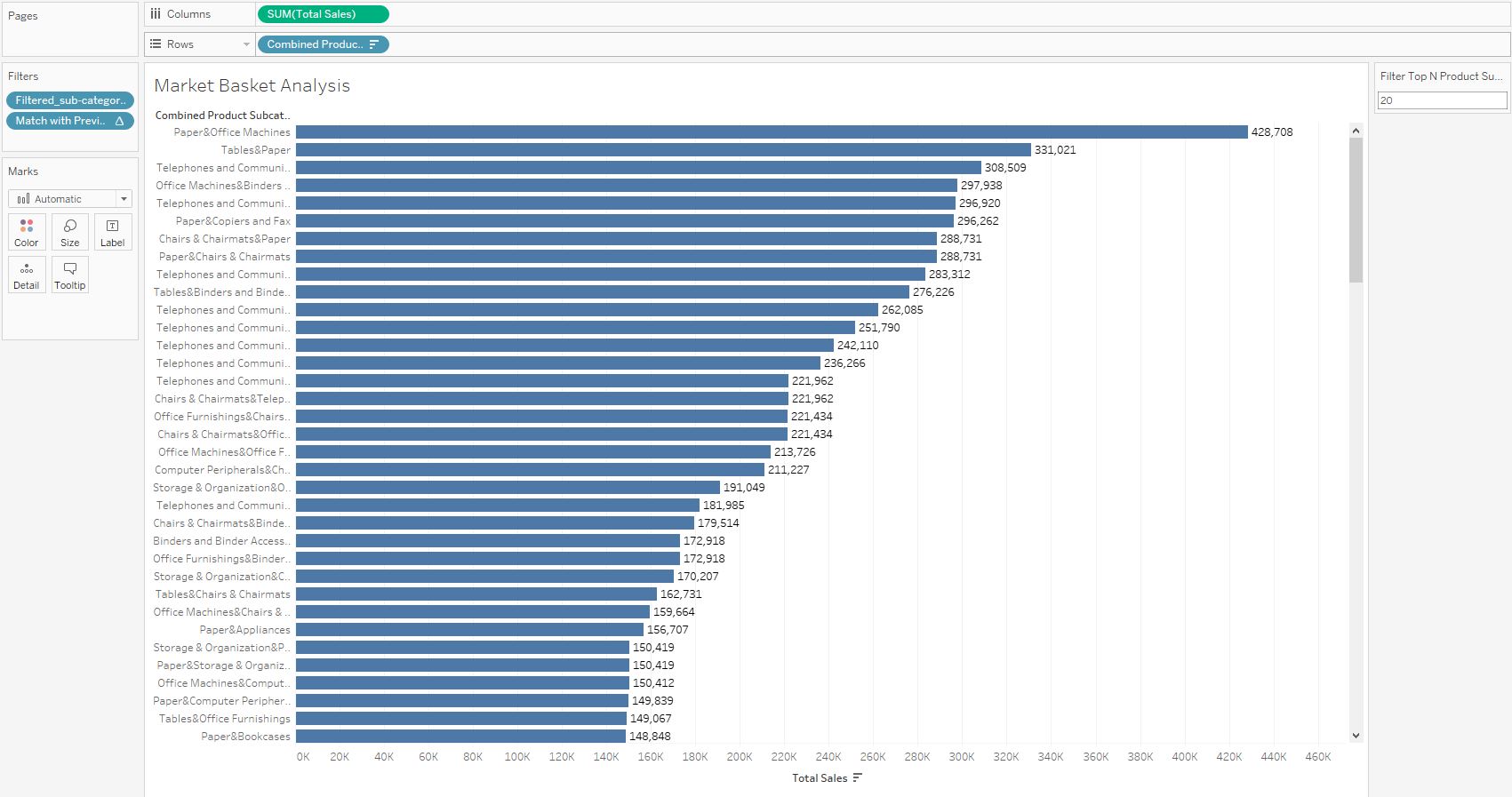


However, the bar graph has duplicated values. This can be fixed with some calculated field calculations.

We need to check if the current row matches with the previous row.

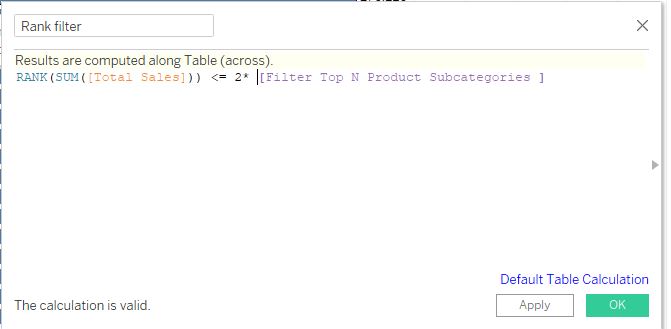


If I get True back as result, which means it’s matching with the previous row and hence it’s a duplicate. Place this in the Filter shelf and exclude the true values to remove the duplicates.

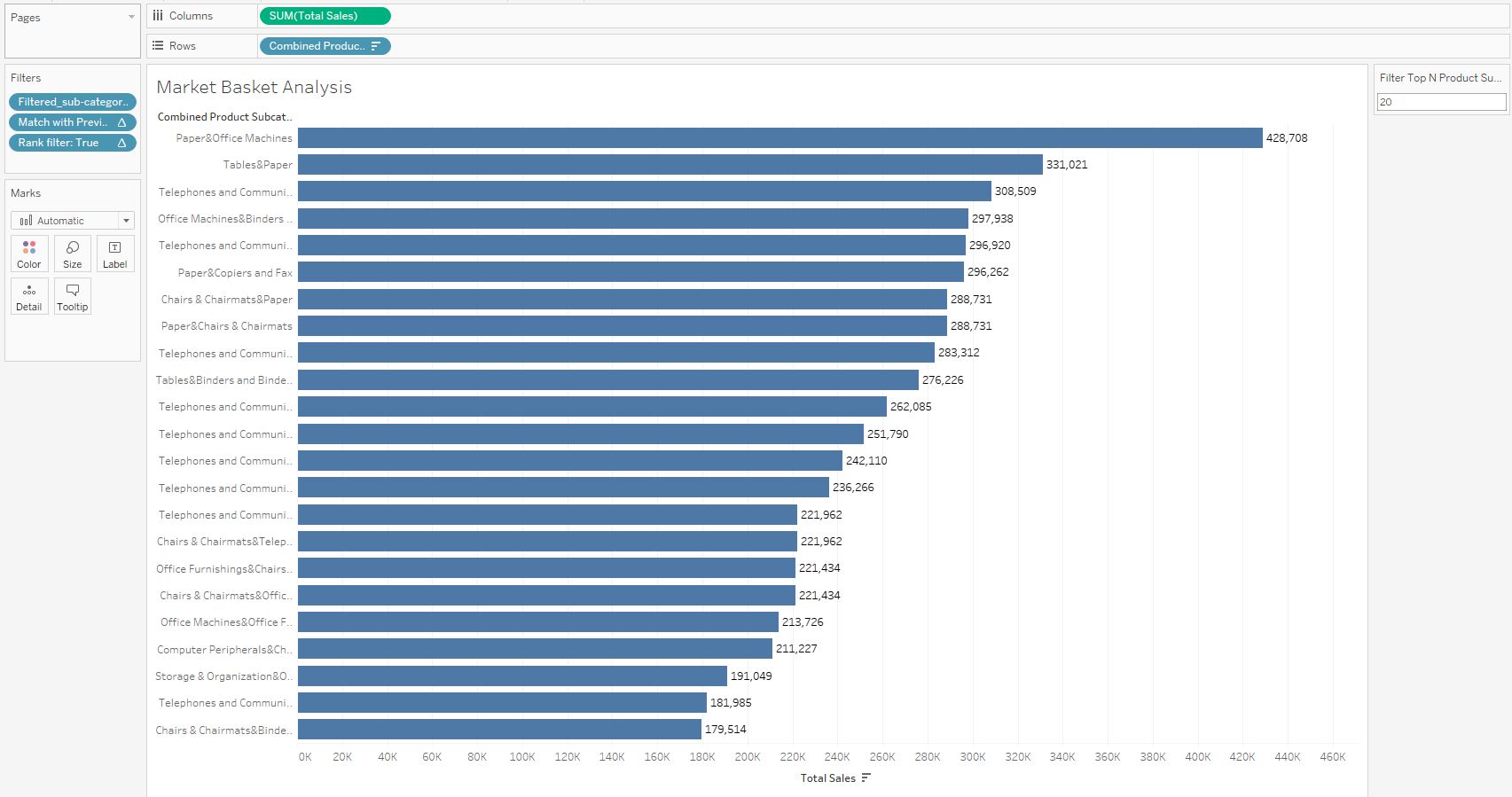


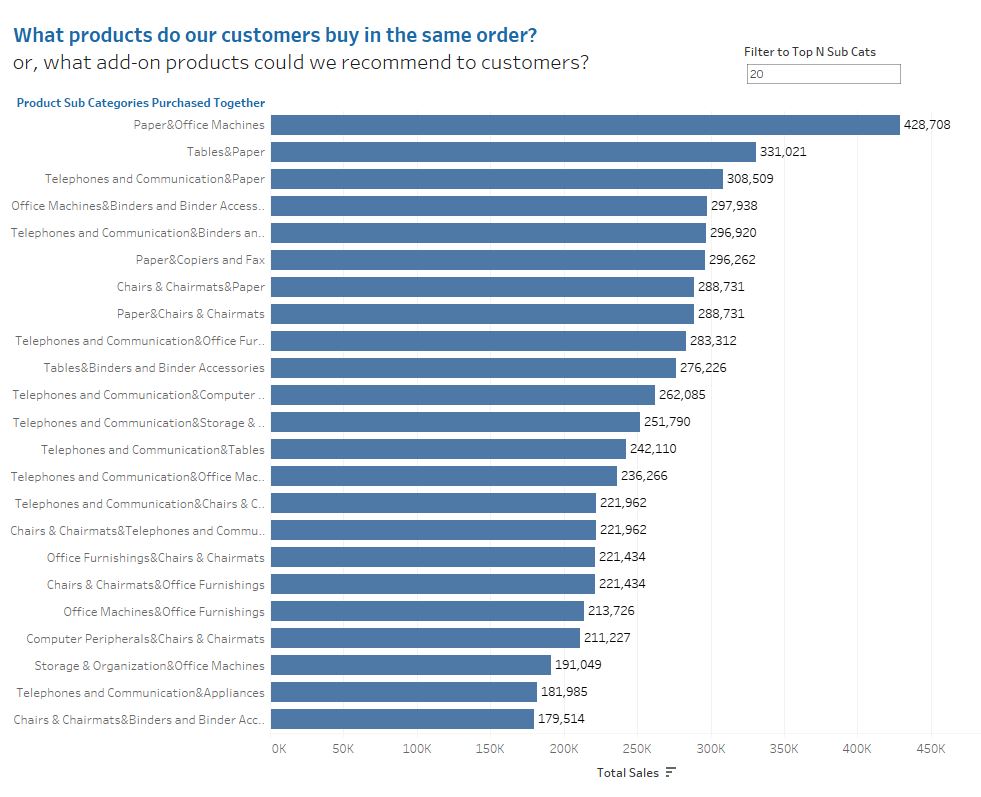
Duplicates are removed

Finally to filter to the top N subcategories, another calculated field is created.



We will rank based on Total Sales. Since we have twice as many records (the previous calculation just hides them the other rows), we take twice as many rows as indicated in the parameter. Place the Rank filter in the filter shelf and select True.





Final dashboard based on the requirement.