Source Data

Data source:
https://www.kaggle.com/datasets/berkayalan/retail-sales-data

Data format: csv

Number of files: 3

sales.csv - Daily sales data covering 2017-2019. This refers to Stores and Products Data.

product_hierarchy.csv - Data containing the hierarchy and sizes of the products.

store_cities.csv - Data containing the city, type and size information of the stores.

Loading Data

Played with data, reducing data sometimes using where clause, including / excluding fields, joining data various ways, and saving screenshots of charts in the process, when any interesting insights were revealed. The final script is the last script, but not the one used for all charts.

Data Overview

1. Sales data has subset ratio of 92.8% on product_id. All other keys are showing 100% subset ratio in Data Modeler View. This implies that Sales data does not include data for some products. From exploring the individual tables, we already know that 50 product_id's have blank cluster_id and those are the same product_id's for which we have no sales data.

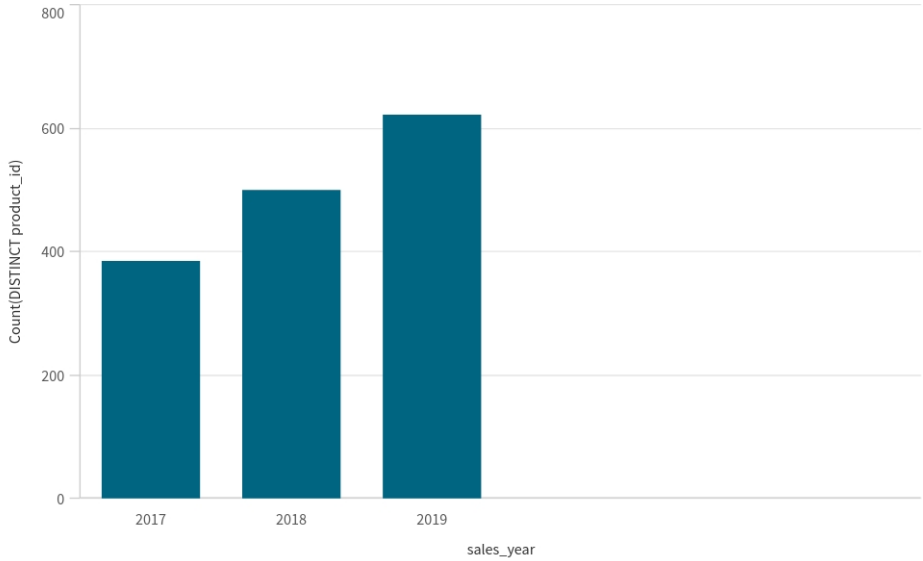
2. There are **15,076,599** rows where revenue is '0' and **15,074,394** rows where sales value is '0'. Expression Sum({<Year={2017}>}Sales) shows that there are **3.57 K sales where revenue generated is 0.**

Data Overview

623 distinct product ids out of 699 have 0 sales and 0 revenue

144 of 144 stores have some products which have both sales and revenue as 0

The number of such products that had 0 sales and 0 revenue increased over the years.



Data Overview

74 product ids have non-zero sales but generated 0 revenue

134 stores have such sales

Below is a chart of Top 10 Products with non-zero sales and zero revenue, over the 3 years.