Data Source

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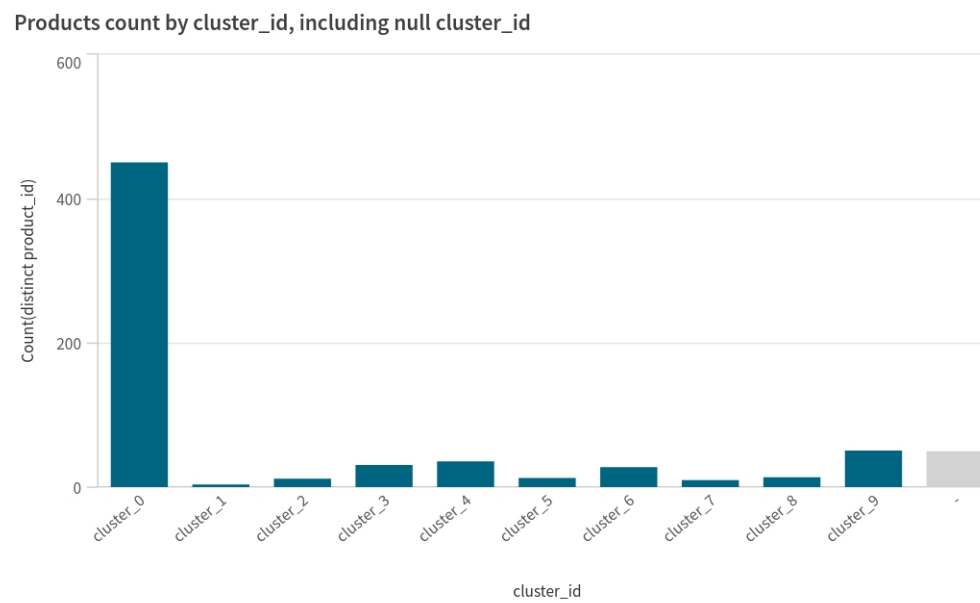
Data format: csv

Number of Files: 1

product_hierarchy.csv - Data containing the hierarchy and sizes of the products. Number of rows is 699
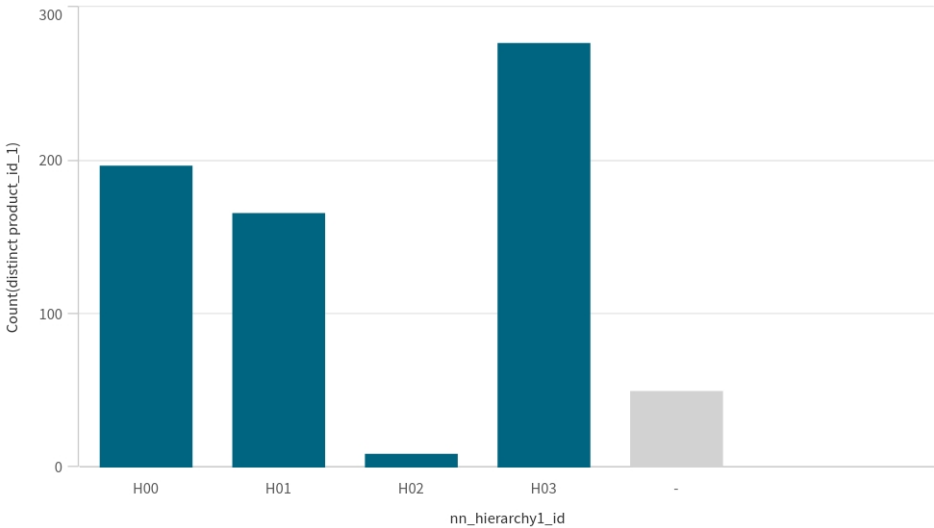
Exploring Products Data

There are 699 distinct product_ids. 50 of them do not have a cluster_id. Replaced blank values of cluster_id with NULL in Load script.
The following chart shows counts of Products by cluster_id, including those for cluster_id that is blank (null).

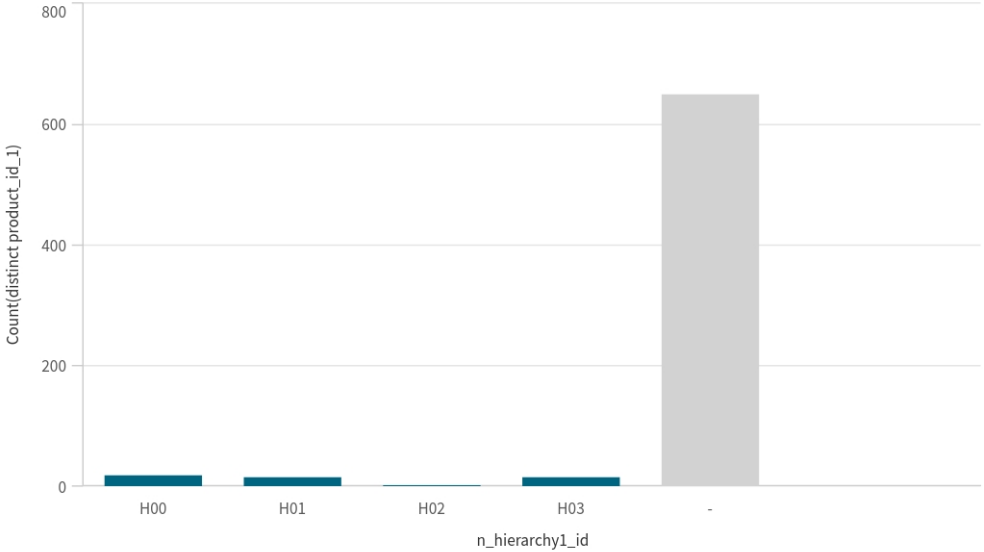**Products count by cluster_id, including null cluster_id**

Exploring Products Data

To experiment with joins and creating tables for the purpose of analysis, in Load script, created two tables, by filtering on cluster_id using where clause - one for products with null cluster_id and another for products with non-null cluater_id.
Created chart showing Product counts over hierarchy, from each of the tables.

Distribution of product ids with cluster_id (not null) across hierarchy

Distribution of product ids with NULL cluster_id across hierarchy

Exploring Products Data

Created a chart to experiment using an expression to filter the values of dimension. Used expression to count productId over hierarchy, only when cluster_id is not null.

Products count by hierarchy for non-null cluster_id using expression