Marketing and Retail Analytics

Submitted by : Rithesh Kamath

Agenda

- Objective
- Background
- Data Preparation & Cleaning
- Importing data into Visualization Tool
- Data Analysis
- Recommendation
- Appendix:
  - Data Assumptions
  - Data Methodology

Objective

- Our main objective is to identify list of top products that contribute to the revenue and also use market basket analysis to analyze the purchase behavior of individual customers to estimate with relative certainty, what items are more likely to be purchased individually or in combination with some other products.
- We also need to identify the product categories which we can get rid of without significantly impacting business.

Background

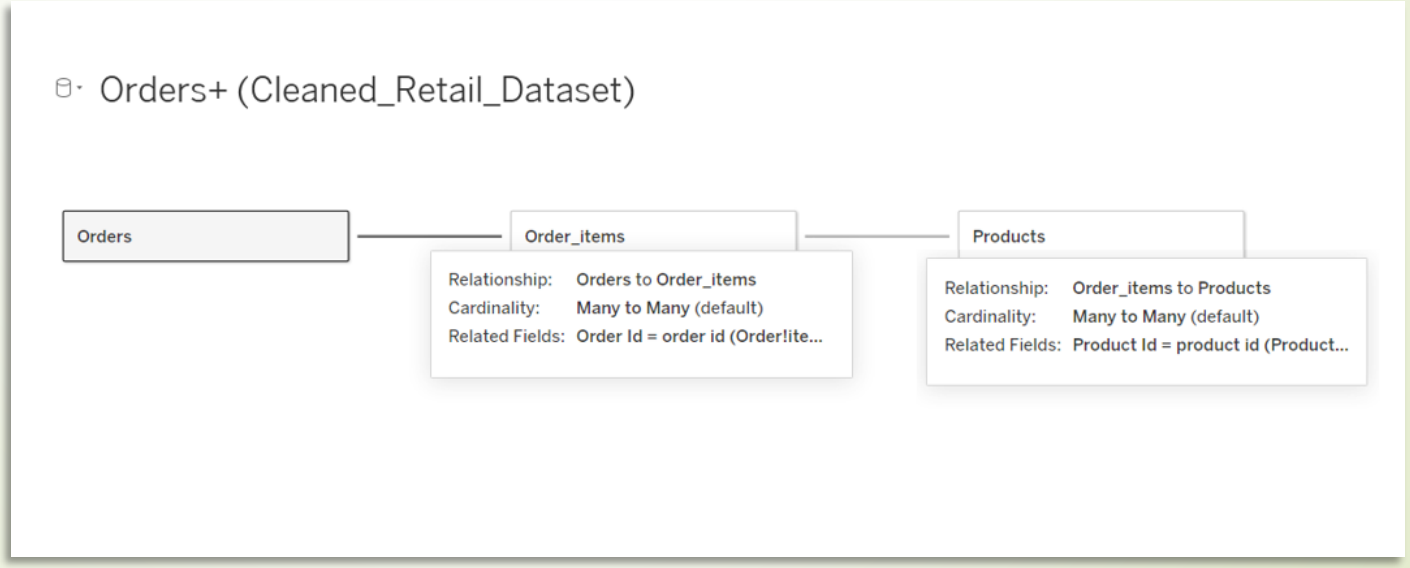
- OList is an e-commerce company that has faced some losses recently and they want to manage their inventory very well so as to reduce any unnecessary costs that they might be bearing.
- Some of products might be fast-moving which sell very quickly while others might be slow-moving. Each of the products being stored in a warehouse incurs cost to the company in terms of space and maintenance. Since storing these products obviously add to the costs that the company incurs, it is absolutely necessary for the organizations to plan their inventory well.

Data Preparation and Cleaning

- Filter & keep records only containing order_status = ‘delivered’ in the ‘orders’ table
- Treat the missing values under the columns ‘order_approved_at’ and ‘order_delivered_timestamp’ of the ‘orders’ table with their respective ‘order_purchase_timestamp’ and ‘order_estimated_delivery_date’ values
- Filter and remove duplicate customer ids from ‘customer_id’ column of the ‘customers’ table
- Treat the missing values under column ‘product_category_name’ of the ‘products’ table by assigning the value ‘toys’ as 75% of the products belong to this category and the mode value for this column also turns out to be ‘toys’
- Treat the missing values under the columns ‘product_weight_g’, ‘product_length_cm’, ‘product_height_cm’ and ‘product_width_cm’ with its respective median values as the data in these columns are skewed
- These cleaned datasets are imported to two excel files - ‘Market_Basket_Dataset.xlsx’ (for market basket analysis) & ‘Cleaned_Retail_Dataset.xlsx’ (for all other visualizations & analysis)

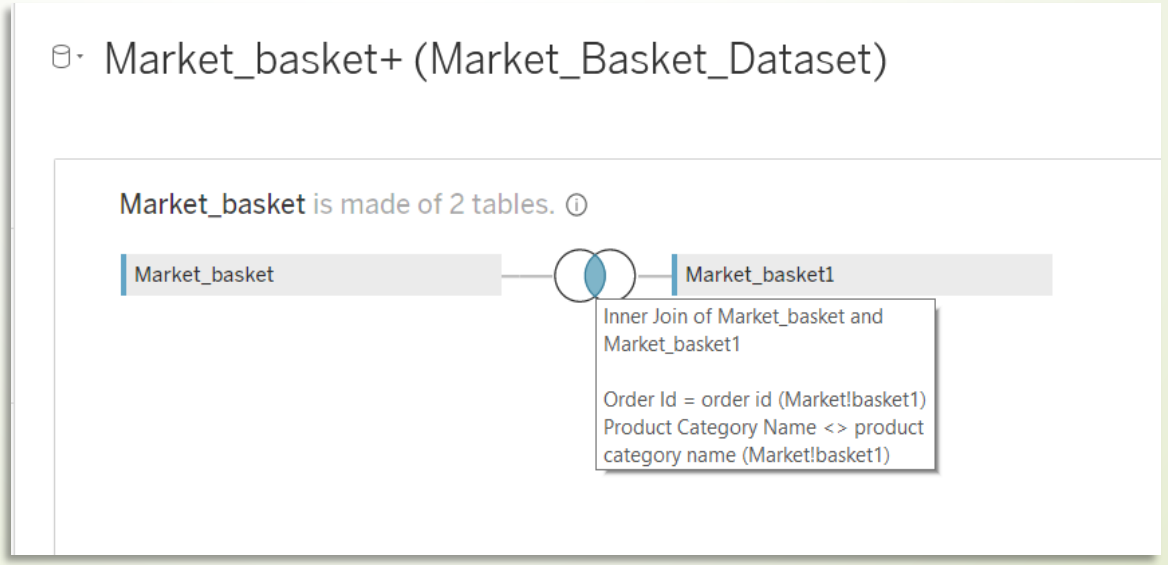
Importing data into Visualization Tool

- For the analysis involving 'Top 20 Ordered Products', 'Top 20 products by Revenue' and 'Product Category Wise orders', we have used the excel file 'Cleaned_Retail_Dataset.xlsx' and the merging of data sets in tableau is done as follows:



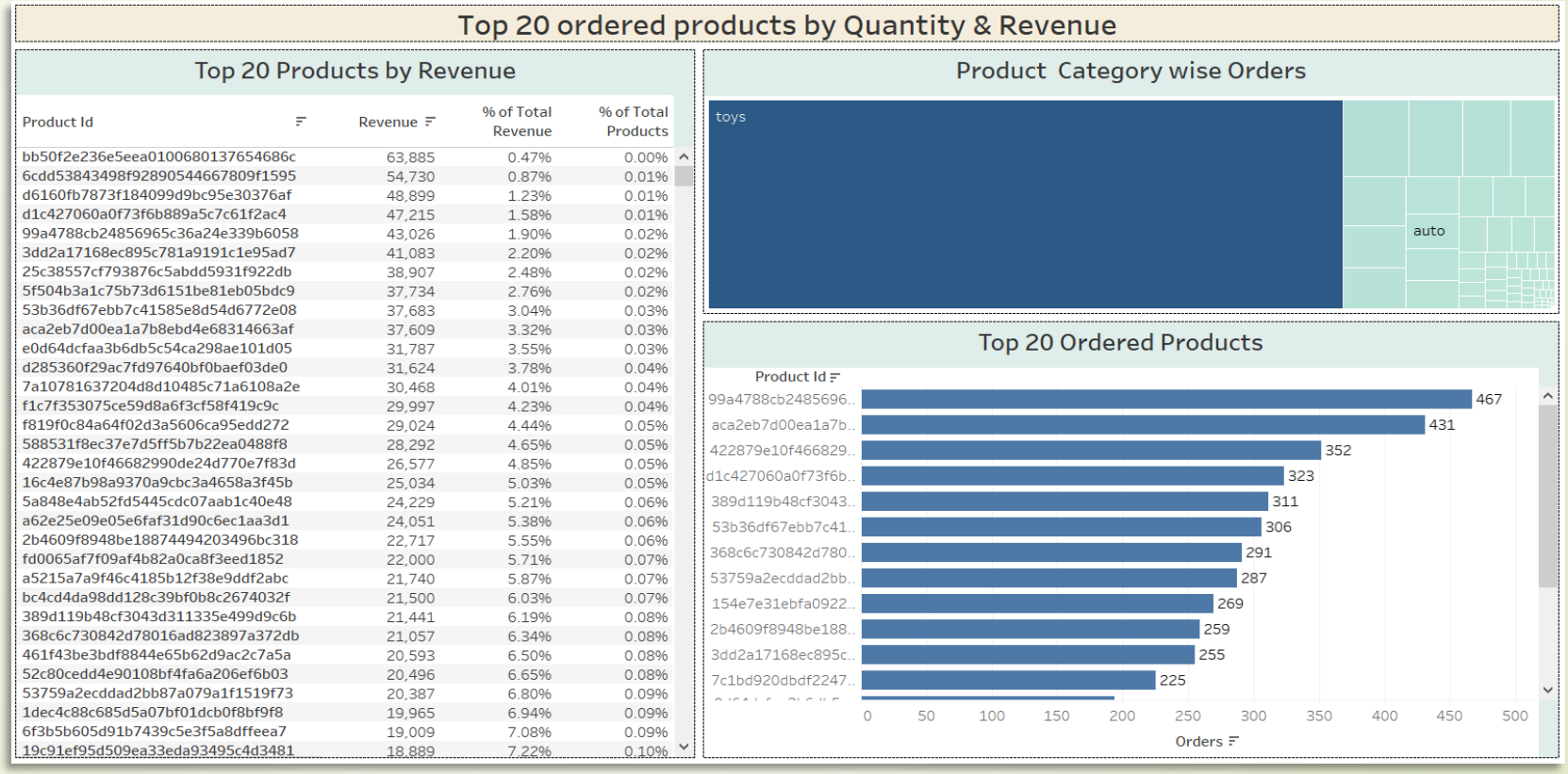
Importing data into Visualization Tool

- For the analysis involving 'Market Basket Summary' and 'Market Basket Category Summary', we have used the excel file 'Market_Basket_Dataset.xlsx' and the self-joining in tableau is done as follows:



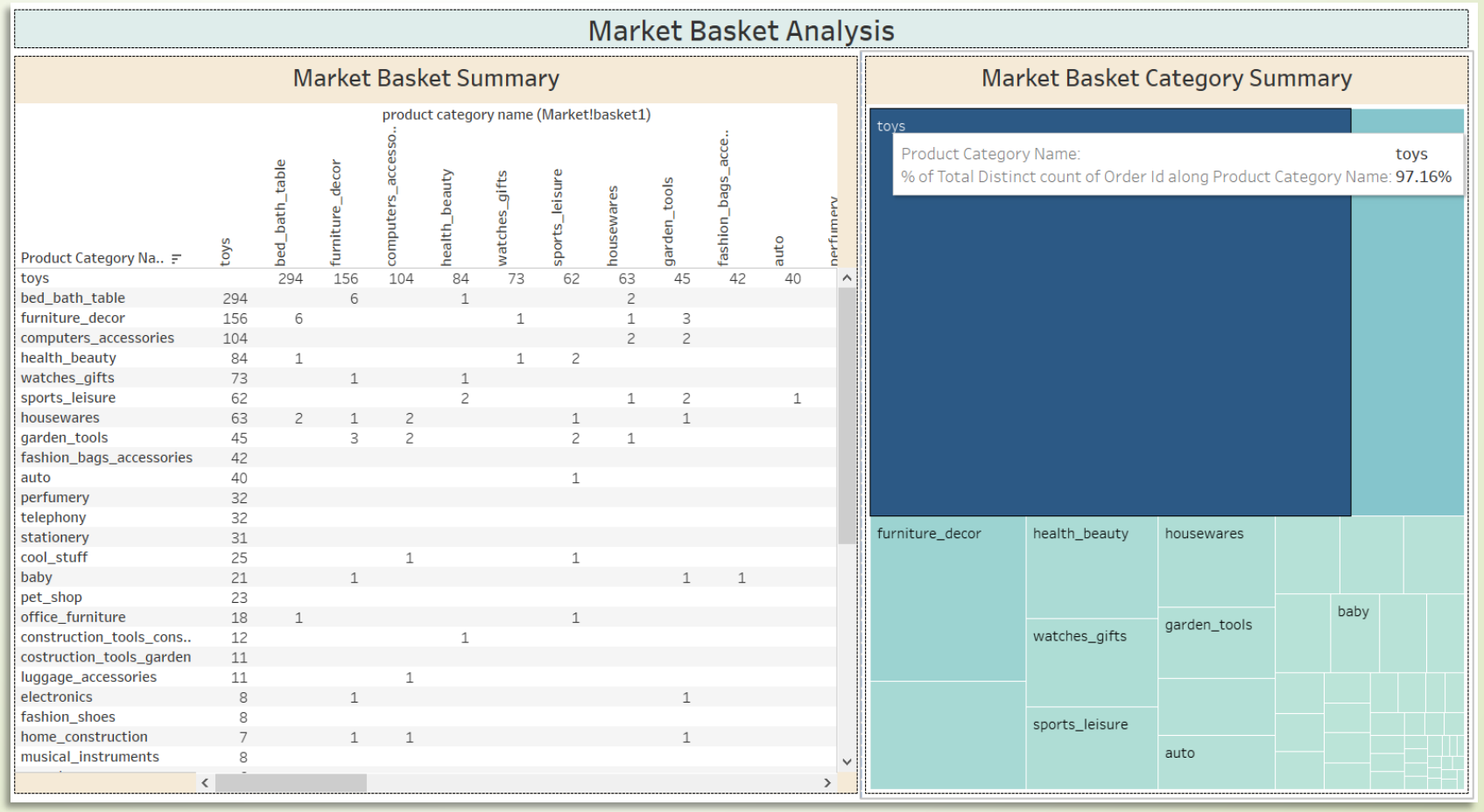
Dashboard 1: Top 20 ordered products by Quantity & Revenue

- This dashboard gives the top ordered products by quantity as well as by revenue.



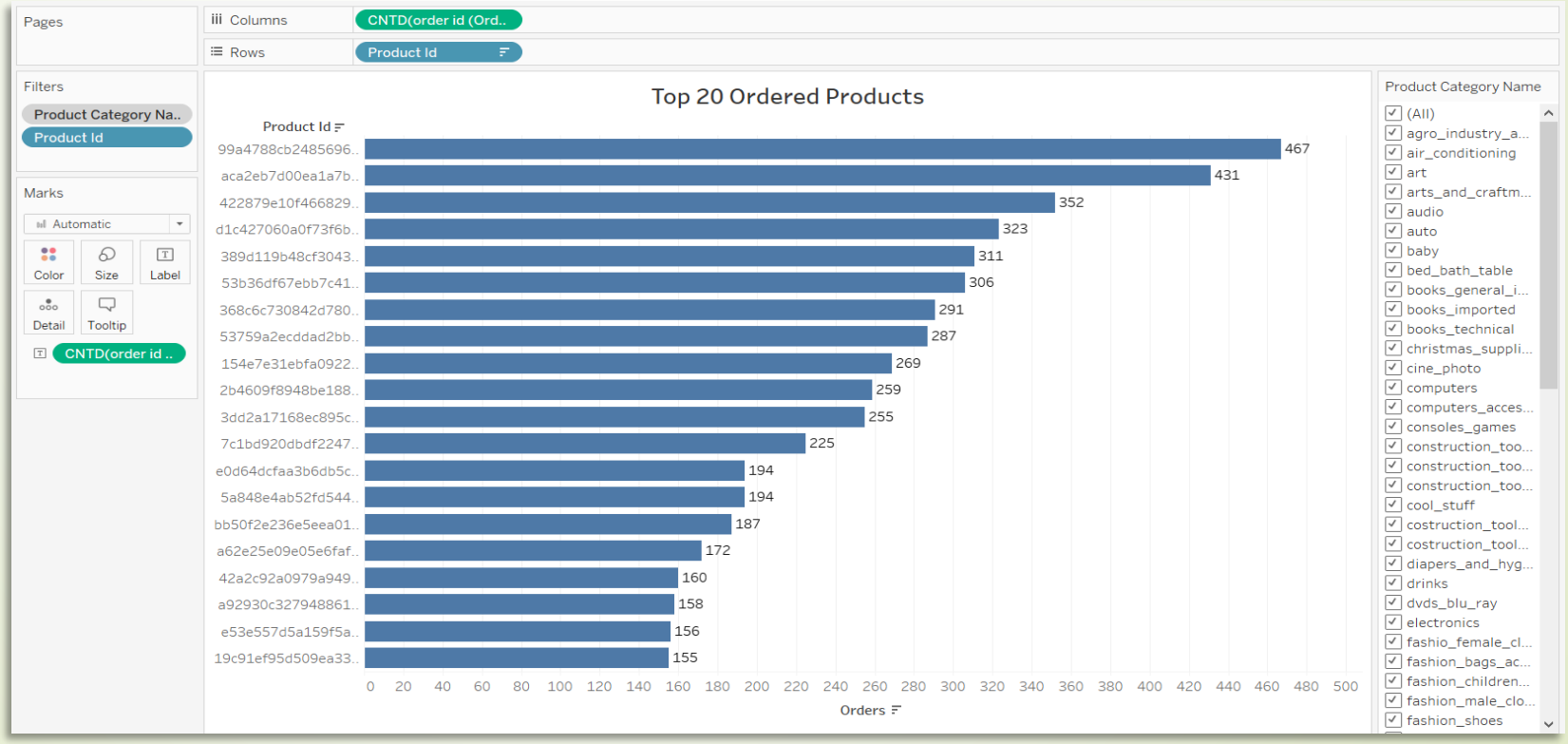
Dashboard 2: Market Basket Analysis

- This dashboard gives the best combinations of product categories which are frequently ordered together



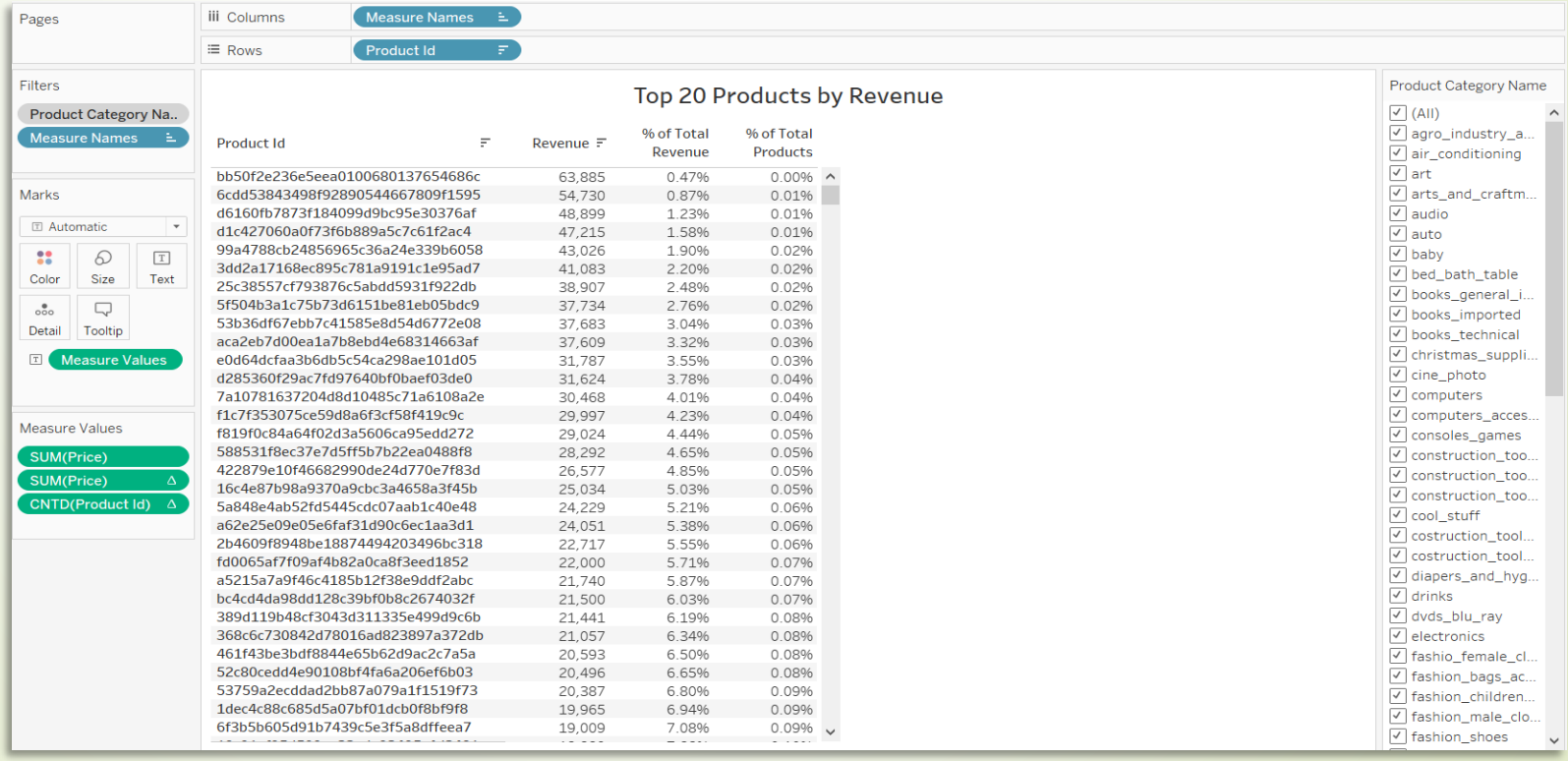
Top 20 Ordered Products

- Visualizes top 20 products ordered by quantity
- Using the filter on the right side, either select a specific or a group of product categories to show their top 20 ordered products by quantity



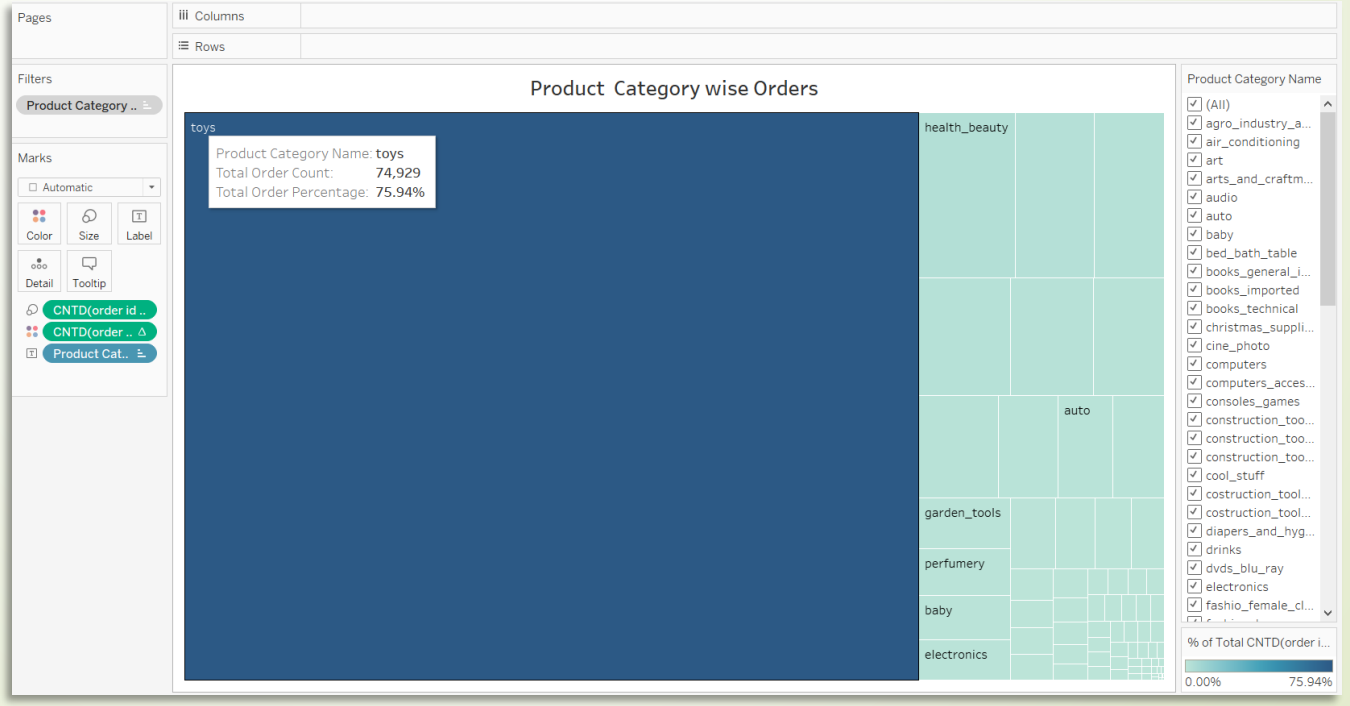
Top 20 Products by Revenue

- Visualizes top products based on revenue
- Using the filter on the right side, either select a specific or a group of product categories to show their top products based on revenue
- The percentage running totals by revenue and by total number of orders are also depicted in this visualization



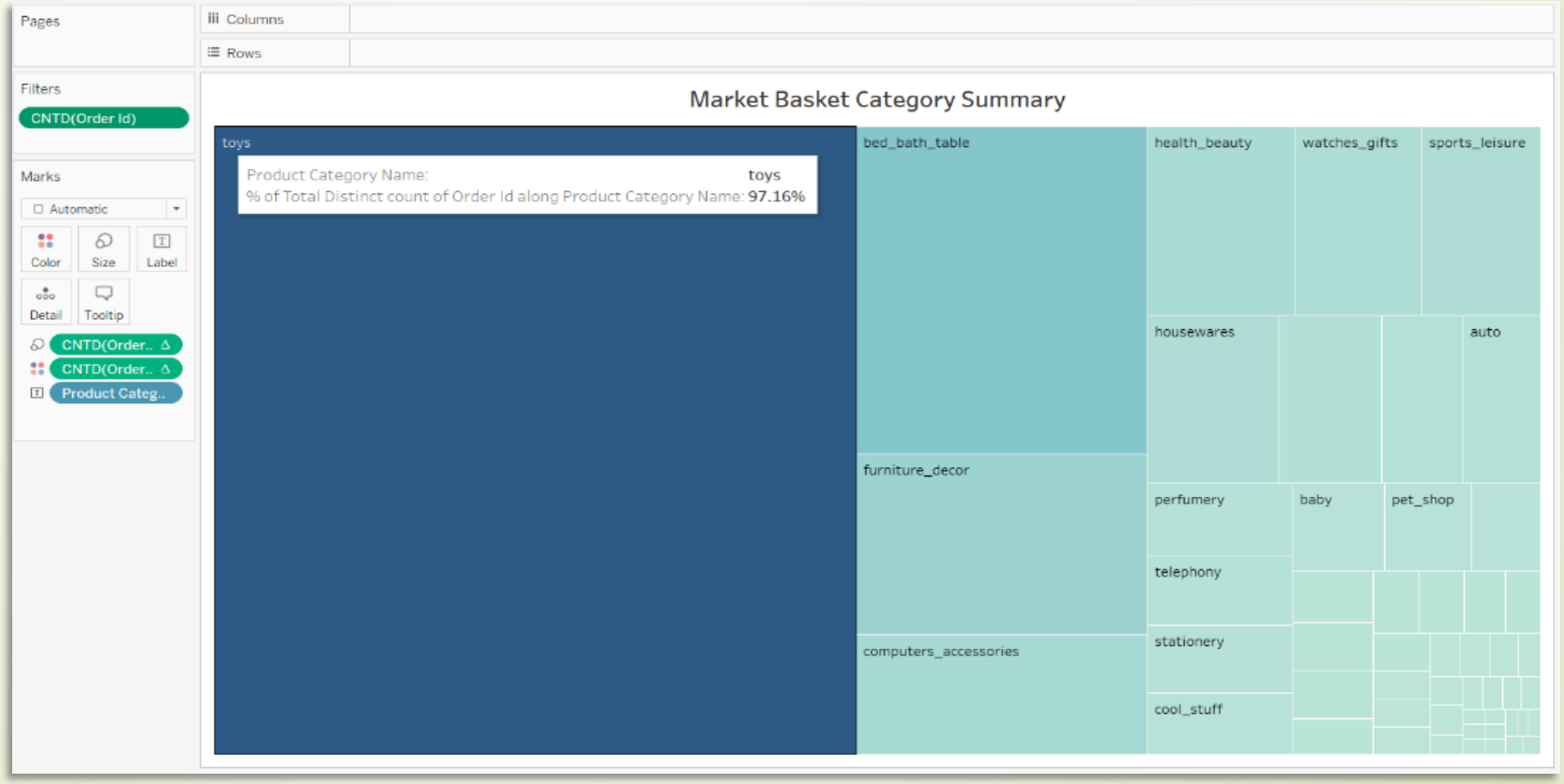
Product Category wise Orders

- Visualizes product category wise total order count and percentage when the mouse is hovered over a particular product category
- Using the filter on the right side, select a group of product categories to show their total order count & percentage of total among these categories
- Toys have the highest order count & percentage



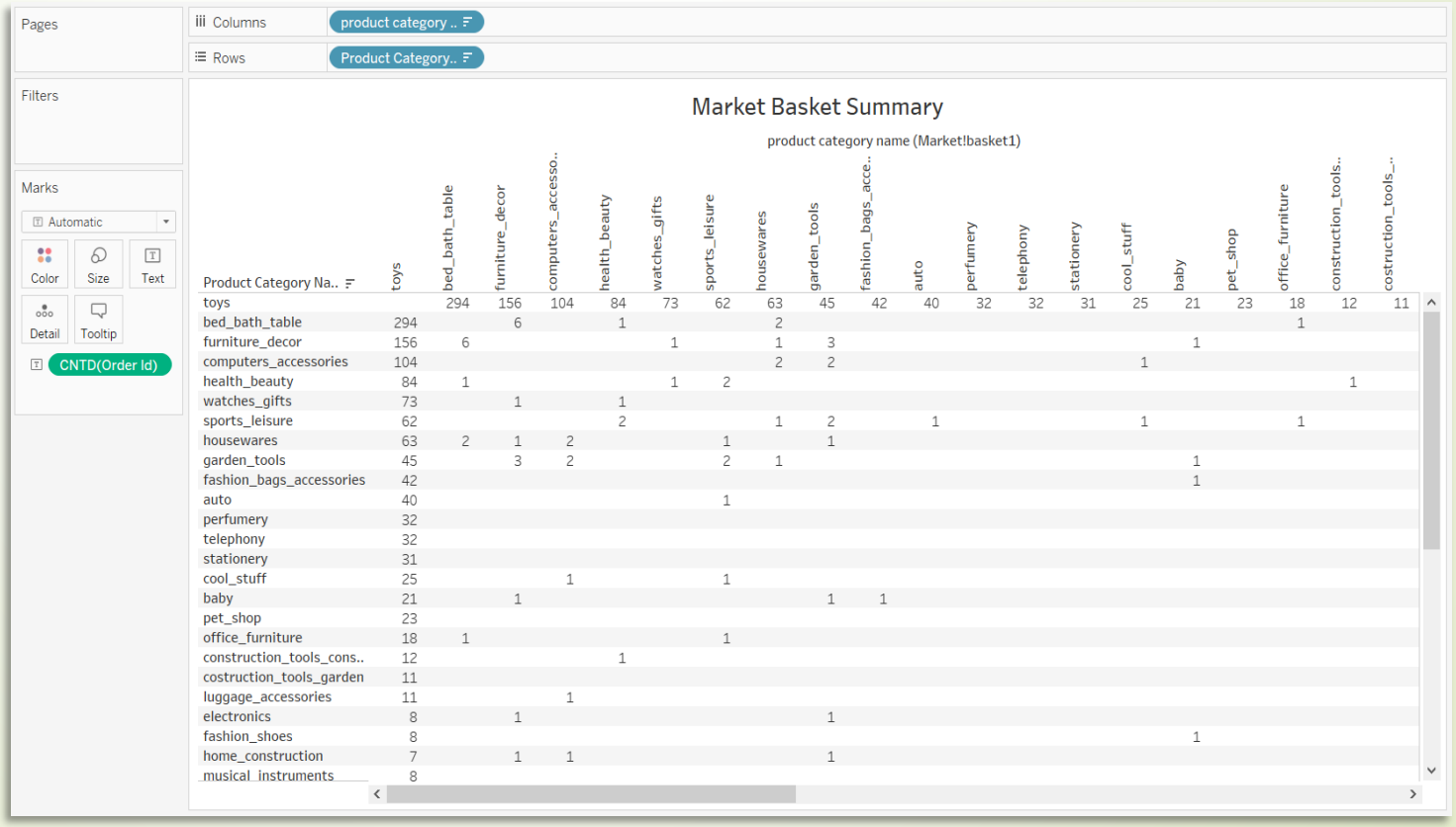
Market Basket Category Summary

- Visualizes the total distinct count of orders along product categories when the mouse is hovered over a particular product category
- Toys lead with the highest total distinct count of orders



Market Basket Summary

- Combinations of product categories which are frequently ordered together are visualized
- Combination of products under ‘toys’ & ‘bed_bath_table’ categories are ordered highest number of times with a count of 294



Product Category Na..	toys	bed_bath_table	furniture_decor	computers_accesso..	health_beauty	watches_gifts	sports_leisure	housewares	garden_tools	fashion_bags_acce..	auto	perfumery	telephony	stationery	cool_stuff	baby	pet_shop	office_furniture	construction_tools..	costruction_tools_..
toys		294	156	104	84	73	62	63	45	42	40	32	32	31	25	21	23	18	12	11
bed_bath_table	294		6		1			2										1		
furniture_decor	156	6				1		1	3							1				
computers_accessories	104							2	2						1					
health_beauty	84	1				1	2												1	
watches_gifts	73		1		1															
sports_leisure	62				2			1	2		1				1			1		
housewares	63	2	1	2			1		1											
garden_tools	45		3	2			2	1								1				
fashion_bags_accessories	42															1				
auto	40						1													
perfumery	32																			
telephony	32																			
stationery	31																			
cool_stuff	25			1			1													
baby	21		1						1	1										
pet_shop	23																			
office_furniture	18	1					1													
construction_tools_cons..	12				1															
costruction_tools_garden	11																			
luggage_accessories	11			1																
electronics	8		1						1											
fashion_shoes	8															1				
home_construction	7		1	1					1											
musical_instruments	8																			

Recommendations

- Toys have the highest order count & percentage (~76%), so it is evident that this is a fast moving product category. Hence, the warehouse with toys will not incur much cost in terms of space & maintenance
- Provide discounts & offers on purchase of Toys
- In comparison to products with shelf life, toys are expected to last longer
- Show recommendation of a toy purchase when baby products are added to cart. Even though this is not a insight from the provided dataset, customers who are intending to buy baby products might be interested in purchasing toys for their baby
- High revenue & order based product categories to be focused more
- Reduce the inventory of low demand products & remove such products from offering which have very low orders
- Toys with bed_bath_table, furniture_decor & computer_accessories are frequently ordered together. So we can recommend shoppers to buy products under these categories who are planning to order Toys

Appendix – Data Assumptions

- First assumption was that when an order is placed online there is very minimal time for it to get approved. So we have filled the missing values present under 'order_approved_at' column of the 'orders' table with its equivalent 'order_purchase_timestamp' values
- Second assumption was that the order usually gets delivered very close to or on the estimated delivery date. So we have filled the missing values under 'order_delivered_timestamp' column of the 'orders' table with its equivalent 'order_estimated_delivery_date' values

Appendix – Data Methodology

- We have made use of Tableau to get insights for our data analysis & visualization
- Please find the detailed Methodology Documents by clicking here

Thank You!