

The background of the entire slide is a dense field of 3D-rendered numbers in various sizes and orientations, creating a sense of depth and data. The numbers are light blue/white against a darker blue background.

**CAPSTONE
PROJECT**

**MARKETING
AND RETAIL
ANALYTICS**

Prepared by -

**Pooja Jantwal ,
Ritik Bhandari**

&

Iranna Chatti

Problem Statement

OList an e-commerce companies have emerged and flourished in the industry. They offer the convenience to order from a wide variety of options from the comfort of one's home. But how do they offer these "wide variety of options or products"? To be able to meet the demands of the customers, any e-commerce company would obviously need to store tons and tons of products in warehouses. Now, some of these warehoused products might be fast-moving products which sell very quickly and some others might be slow-moving. Each of the products being stored incurs a 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 organisations to plan their inventory well.

Now, OList is one such 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. In this assignment, you have to manage the inventory cost of this e-commerce company OList. You need to identify top products that contribute to the revenue and also use market basket analysis to analyse the purchase behaviour 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 should understand the below points for our Analysis:

- *What should be the ideal category depth?*
- *Some product combinations are sold more than others.*
- *This needs to be understood using market basket analysis.*
- *After having specific categories in hand that will be beneficial to the company, how can you reduce the number of items in a particular category?*

Olist Dataset as Retail Dataset

The dataset available is called “Retail_dataset”. It is a xlsx file or simply a spreadsheet.

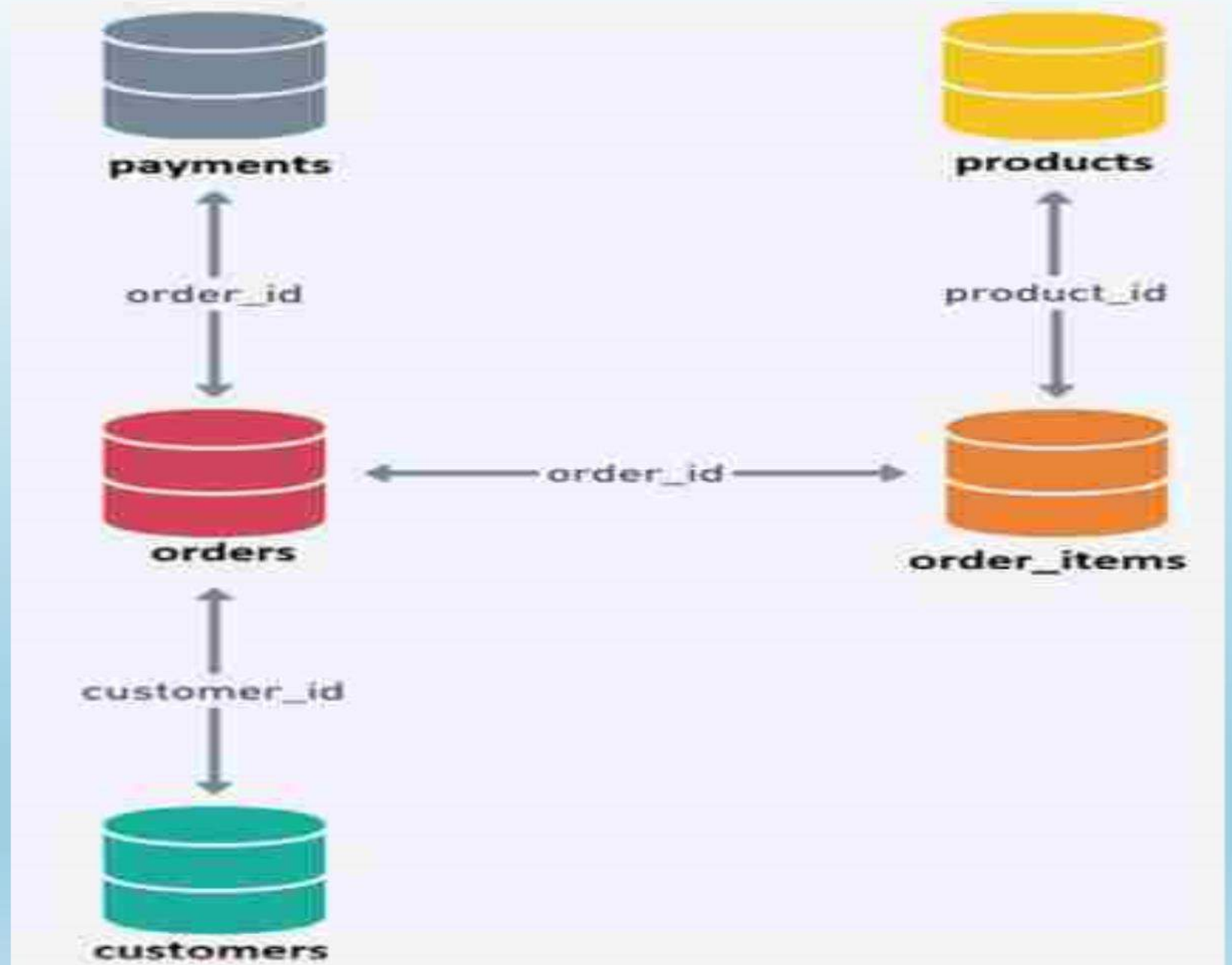
There are five tabs in the excel file.

1. Orders
2. Order_Items
3. Products
4. Payments
5. Customers

Overview of Retails Dataset:

- The total number of Products available at Olist Warehouse is 32950.
- There are total 99440 Orders ordered.
- Unique customers total at 96096.
- There were 3345 repeat customers.
- The dataset contains transactions from September 2016 to September 2018.

Entity Relationship Diagram



Step 1: Importing Libraries and Reading Data

Used many python libraries throughout the project like- pandas, numpy, seaborn, Sklearn, matplotlib.pyplot etc. I formed a dataframe for each sheet in the excel file using the pandas 'read_excel' function to read all the different sheets in the "Retail_dataset".

Step 2: Data Cleaning and EDA

The data available to us contains many duplicates and Null values. We will first make the data suitable for evaluation by performing Exploratory Data Analysis. As per requirement, we only consider orders having statuses as 'delivered'. We will filter down the values in the jupyter notebook and create a separate cleaned file which we can use to create visualizations in Tableau.

Step 3: Creating a Cleaned Excel File

After having gone through the whole dataset and performing necessary cleanup and EDA. I converted the dataframes into a new excel file using the 'ExcelWriter' function of pandas. I combined all the dataframe into a single sheet to be used further in the Tableau environment.

Step 4: Merging Dataframes

Using pandas merge function to join all the sheets using the common fields as depicted in the Entity Relationship Diagram.

Step 5: Data Visualisation in Python

Using bar plot and pandas value_counts functions, I come to know that:

- Top 20 most ordered product categories
- Top 20 most revenue generating product categories

The most frequent purchases from a product category are in the following order:

- 1.Toys
2. Health and Beauty
3. Bed Bath Table
4. Sports and Leisure and
- 5.Computer Accessories

I observed that Toys health and beauty and watches and gifts make up more than 80% of all revenue.

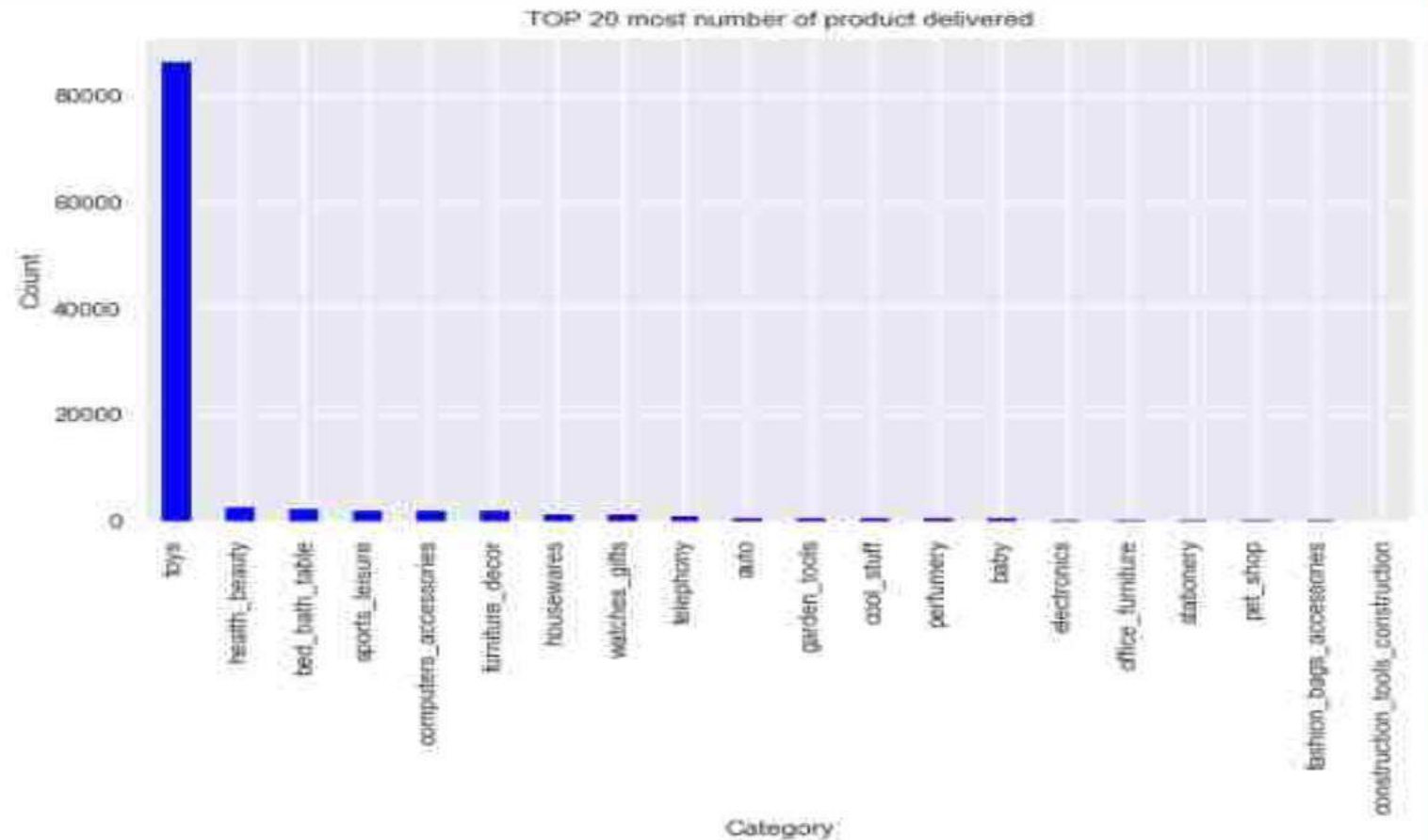
EDA Analysis Observations

- We find that 97 % of orders were of 'delivered' status.
- Dropped all the other orders for the purpose of this project.
- There were 160 'order_approved_at' null values and 2965 'order_delivered_timestamp' null values in the order's sheet.
- Filled the Nan values of 'order_approved_at' and 'order_delivered_timestamp' with the appropriate values derived from the same dataframe.

Most Frequent Purchases from Product Category :

| | |
|---------------------------------|-------|
| toys | 86777 |
| health_beauty | 3047 |
| bed_bath_table | 2690 |
| sports_leisure | 2332 |
| computers_accessories | 2209 |
| furniture_decor | 2193 |
| housewares | 1718 |
| watches_gifts | 1555 |
| telephony | 1229 |
| auto | 1076 |
| garden_tools | 1070 |
| cool_stuff | 1020 |
| perfumery | 856 |
| baby | 788 |
| electronics | 722 |
| office_furniture | 702 |
| stationery | 638 |
| pet_shop | 513 |
| fashion_bags_accessories | 445 |
| construction_tools_construction | 254 |

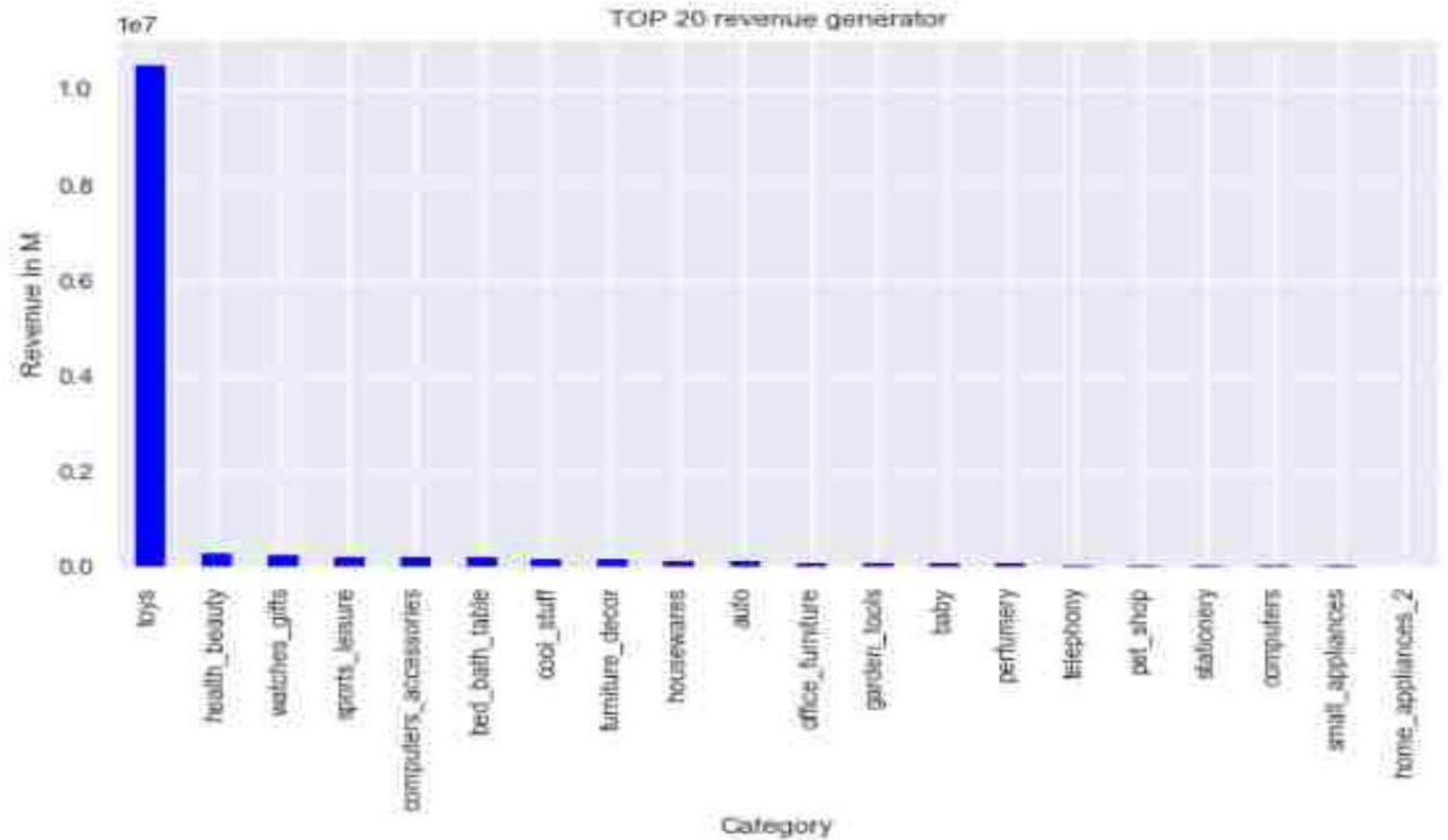
Name: product_category_name, dtype: int64



The most frequent purchases from a product category are in the following order:

Most Revenue Generator Products :

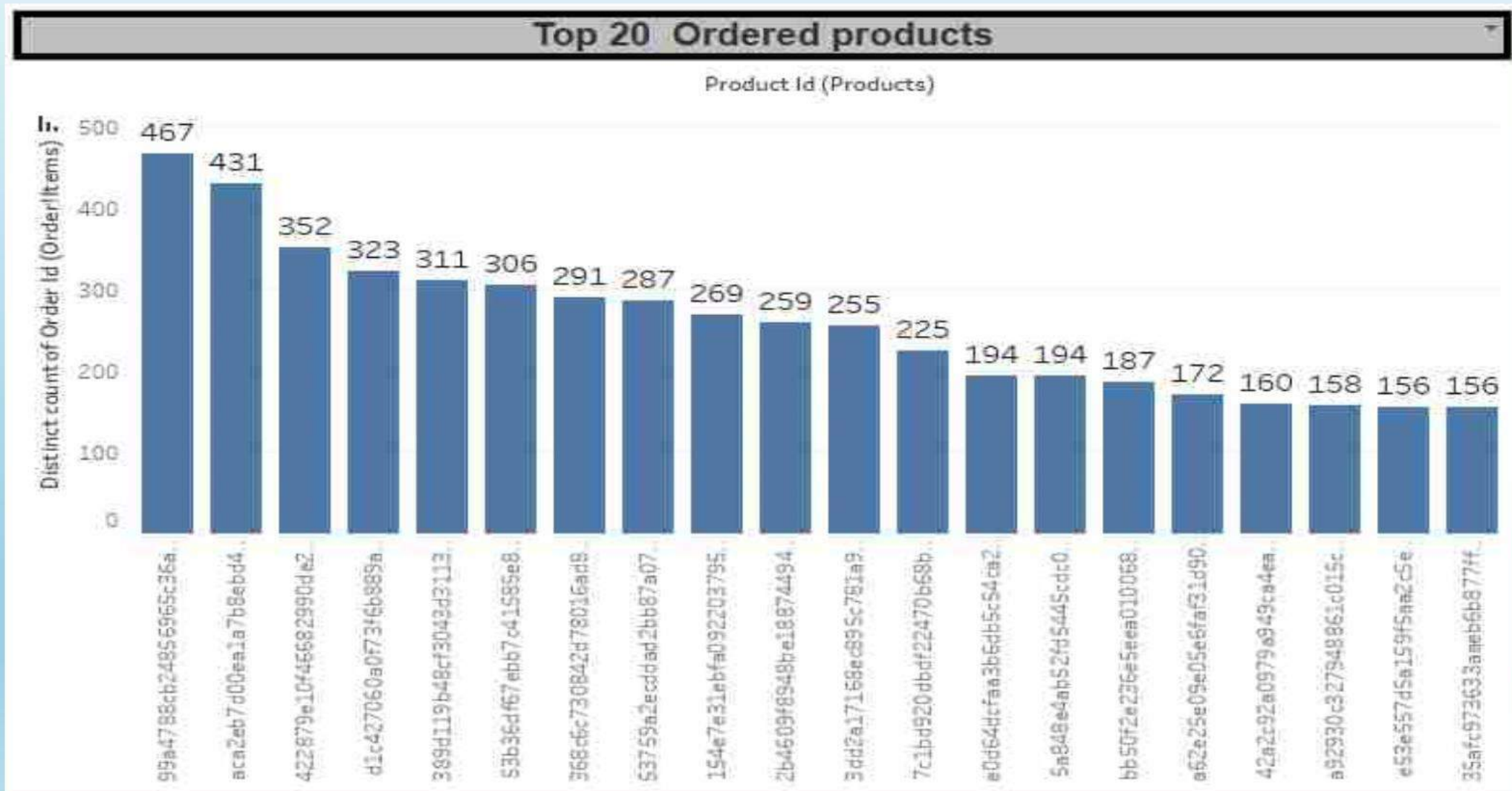
```
0      toys
1  health_beauty
2  watches_gifts
3  sports_leisure
4  computers_accessories
5  bed_bath_table
6  cool_stuff
7  furniture_decor
8  housewares
9  auto
10 office_furniture
11 garden_tools
12 baby
13 perfumery
14 telephony
15 pet_shop
16 stationery
17 computers
18 small_appliances
19 home_appliances_2
Name: product_category_name, dtype: object
```



Toys, Health and beauty, Watches and gift make up more than 80% of all revenue

Tableau Visualizations

Top 20 Most Ordered Products by Product Id



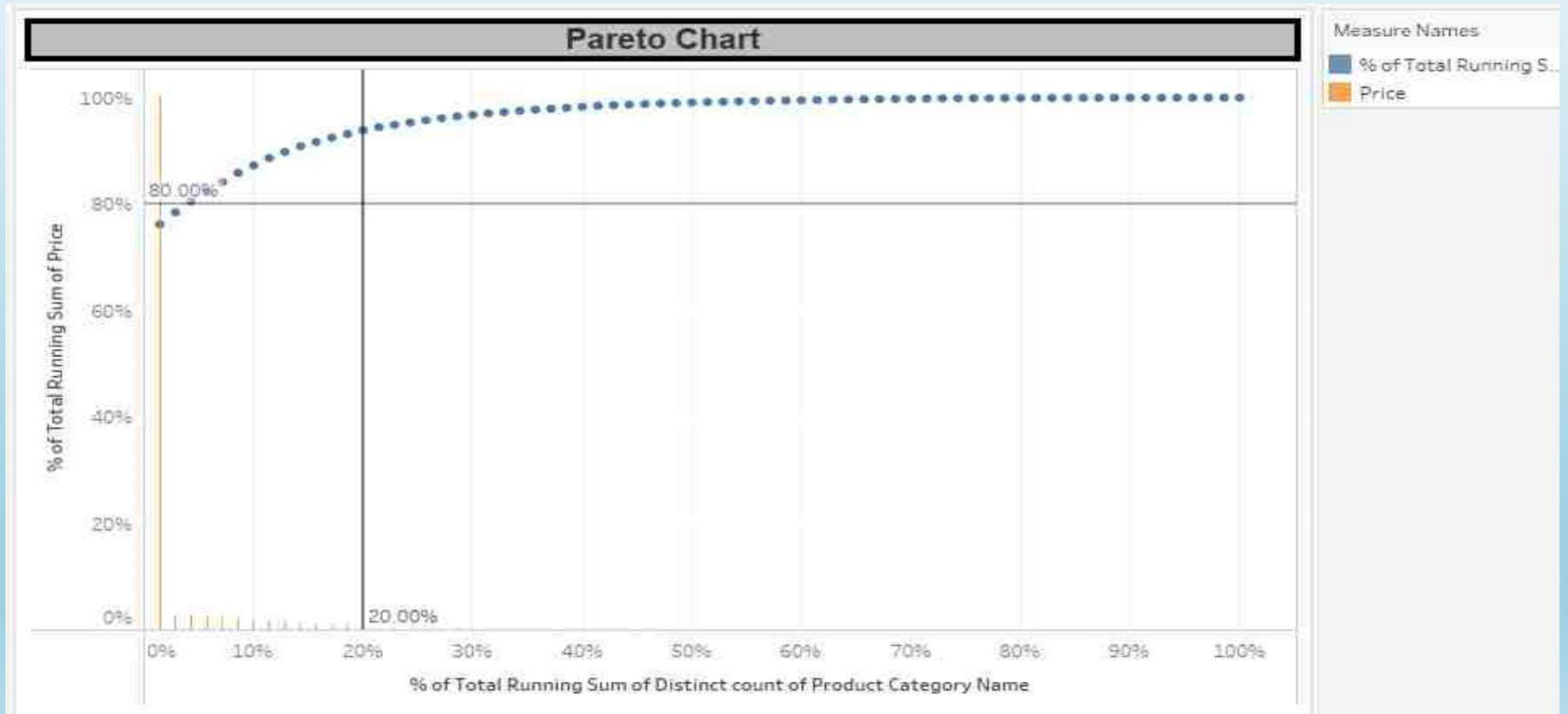
Top 20 Most Revenue Generating Product Categories



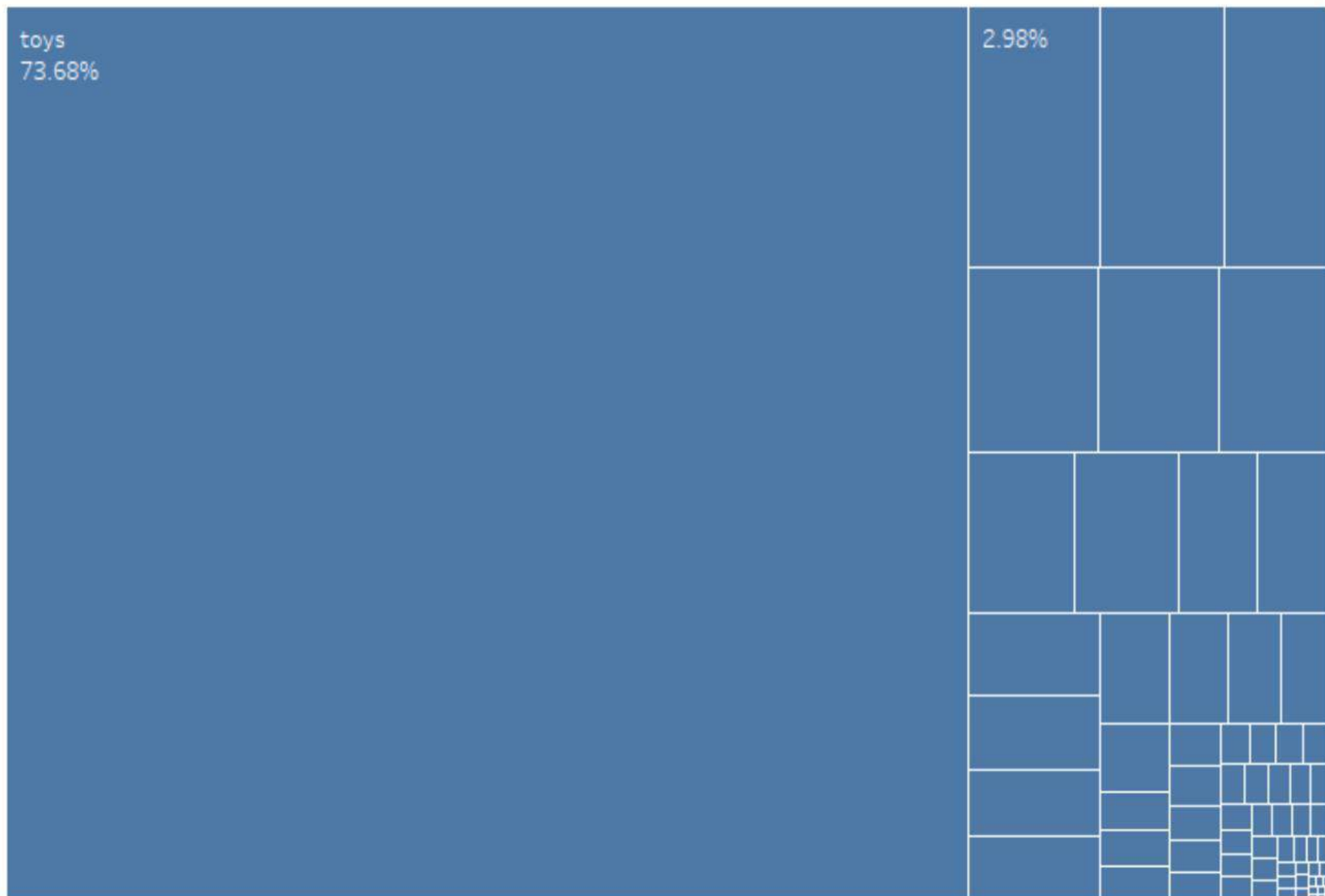
Top 10 Most Ordered Product Category



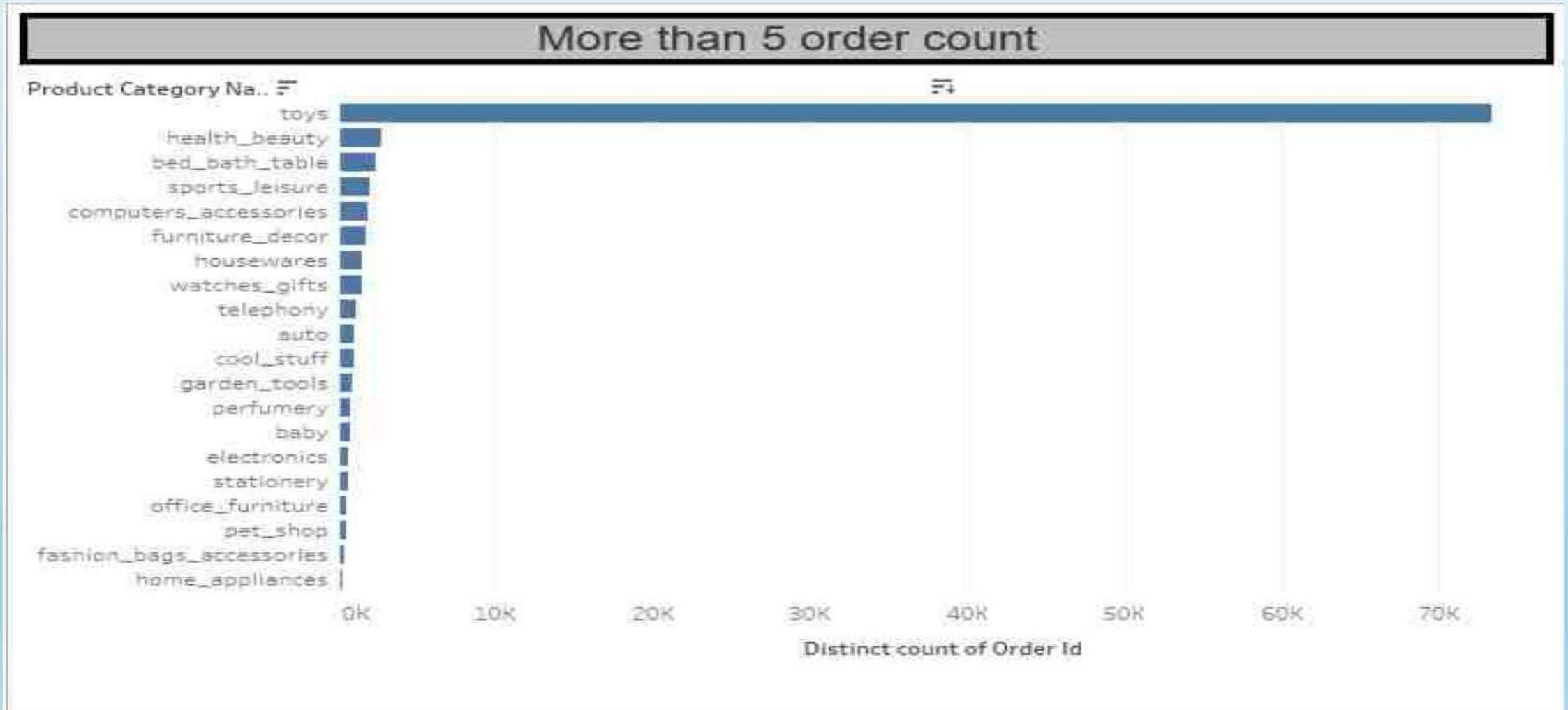
Pareto Chart – Between total running Sum Price & Distinct count of Product



Share of each product category by Number of Orders



Considering more than 5 order count by plotting bar chart between product category & Order ID



Relation showing between the two products

Product Category Association

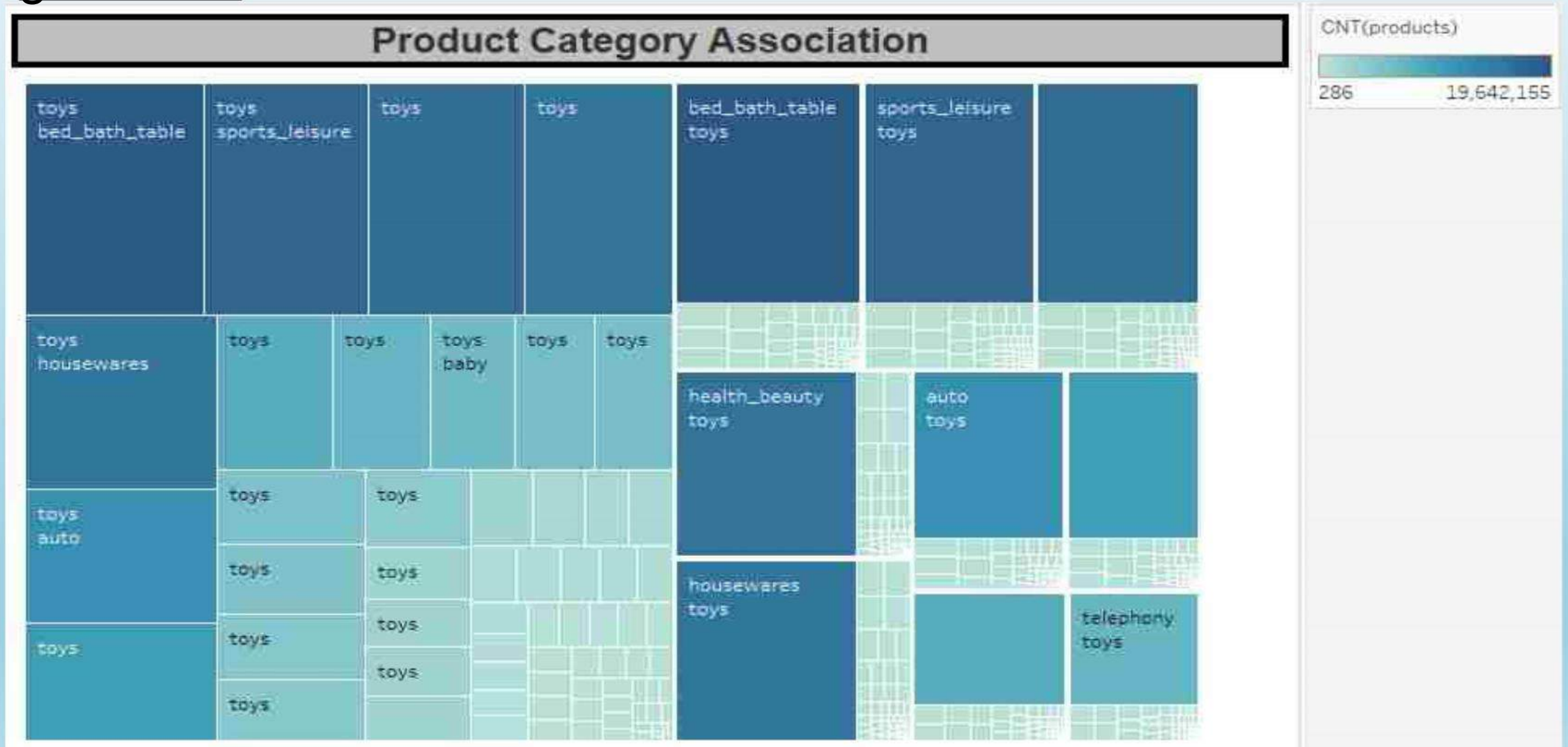
| Product Cat. | ath_t. | computers_... | furniture_d... | health_beau.. | housewares | sports_leisu.. | telephony | toys | watches_gif.. |
|-----------------|--------|---------------|----------------|---------------|------------|----------------|-----------|------------|---------------|
| kitchen_din... | 19,725 | 10,625 | 17,175 | 16,200 | 15,775 | 18,025 | 7,150 | 622,375 | 8,400 |
| la_cuisine | 2,367 | 1,275 | 2,061 | 1,944 | 1,893 | 2,163 | 858 | 74,685 | 1,008 |
| luggage_ac... | 31,267 | 43,775 | 70,761 | 66,744 | 64,993 | 74,263 | 29,458 | 2,564,185 | 34,608 |
| market_pla... | 24,459 | 13,175 | 21,297 | 20,088 | 19,561 | 22,351 | 8,866 | 771,745 | 10,416 |
| music | 4,734 | 2,550 | 4,122 | 3,888 | 3,786 | 4,326 | 1,716 | 149,370 | 2,016 |
| musical_ins... | 59,964 | 32,300 | 52,212 | 49,248 | 47,956 | 54,796 | 21,736 | 1,892,020 | 25,536 |
| office_furni... | 76,533 | 41,225 | 66,639 | 62,856 | 61,207 | 69,937 | 27,742 | 2,414,815 | 32,592 |
| party_supp... | 7,101 | 3,825 | 6,183 | 5,832 | 5,679 | 6,489 | 2,574 | 224,055 | 3,024 |
| perfumery | 77,525 | 95,625 | 154,575 | 145,800 | 141,975 | 162,225 | 64,350 | 5,601,375 | 75,600 |
| pet_shop | 35,708 | 73,100 | 118,164 | 111,456 | 108,532 | 124,012 | 49,192 | 4,281,940 | 57,792 |
| security_an... | 789 | 425 | 687 | 648 | 631 | 721 | 286 | 24,895 | 336 |
| signaling_a... | 24,459 | 13,175 | 21,297 | 20,088 | 19,561 | 22,351 | 8,866 | 771,745 | 10,416 |
| small_appli... | 48,918 | 26,350 | 42,594 | 40,176 | 39,122 | 44,702 | 17,732 | 1,543,490 | 20,832 |
| small_appli... | 5,523 | 2,975 | 4,809 | 4,536 | 4,417 | 5,047 | 2,002 | 174,265 | 2,352 |
| sports_leis... | 58,869 | 306,425 | 495,327 | 467,208 | 454,951 | | 206,206 | 17,949,295 | 242,256 |
| stationery | 36,204 | 100,300 | 162,132 | 152,928 | 148,916 | 170,156 | 67,496 | 5,875,220 | 79,296 |
| tablets_pri... | 3,156 | 1,700 | 2,748 | 2,592 | 2,524 | 2,884 | 1,144 | 99,580 | 1,344 |
| telephony | 25,654 | 121,550 | 196,482 | 185,328 | 180,466 | 206,206 | | 7,119,970 | 96,096 |
| toys | 42,155 | 10,530,375 | 17,102,865 | 16,131,960 | 15,708,745 | 17,949,295 | 7,119,970 | | 8,364,720 |
| watches_gi... | 55,104 | 142,800 | 230,832 | 217,728 | 212,016 | 242,256 | 96,096 | 8,364,720 | |

CNT(products)

286

19,642,155

Products which are more compatible in terms of revenue generator



Olist Retail Analytics Dashboard

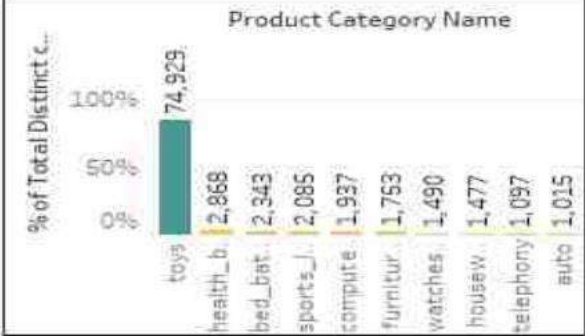
Top 20 Ordered products



TOP 20 Most Revenue Generating Product Category

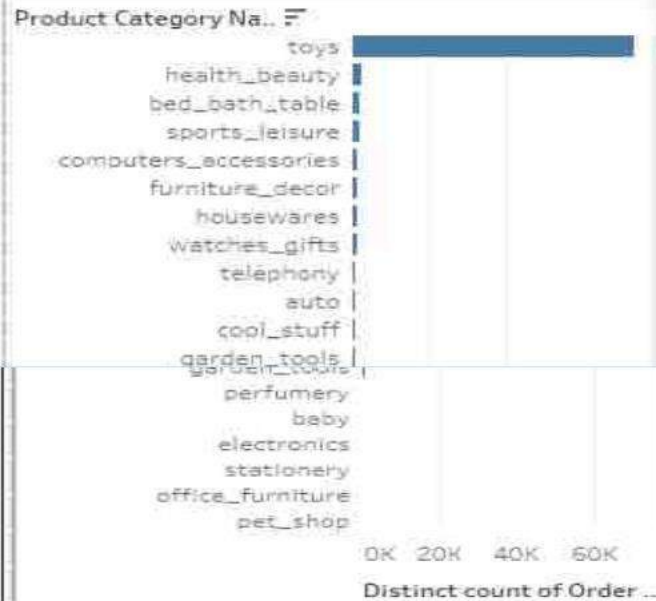


Category wise Orders

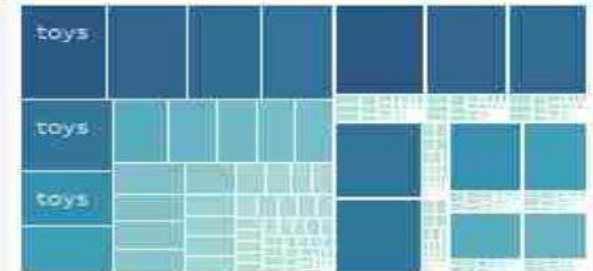


| Product I.. | Distinct .. | % of Tot.. | Price |
|-----------------|-------------|------------|--------|
| bb50f2e236e5.. | 187 | 0.47% | 63,895 |
| 6cdd53843498.. | 151 | 0.87% | 54,730 |
| d6160fb7873f.. | 35 | 1.23% | 48,899 |
| d1c427060a0f.. | 323 | 1.58% | 47,219 |
| 99a4788cb248.. | 467 | 1.90% | 43,026 |
| 3dd2a17168ec.. | 255 | 2.20% | 41,083 |
| 25c38557cf79.. | 38 | 2.48% | 38,907 |
| 5f504b3a1c75.. | 63 | 2.76% | 37,734 |
| 53b36df67ebb.. | 306 | 3.04% | 37,683 |
| aca2eb7d00ea.. | 431 | 3.32% | 37,609 |
| e0d64dcfaa3b.. | 194 | 3.55% | 31,787 |
| d285360f29ac.. | 122 | 3.78% | 31,624 |
| 7a1078163720.. | 142 | 4.01% | 30,468 |
| f1c7f353075ce.. | 150 | 4.23% | 29,997 |
| f819f0c84a64f.. | 45 | 4.44% | 29,024 |
| 588531f8ec37.. | 19 | 4.65% | 28,292 |
| 422879e10f46.. | 352 | 4.85% | 26,577 |
| 16c4e87b98a9.. | 13 | 5.03% | 25,034 |
| 5a848e4ab52f.. | 194 | 5.21% | 24,229 |
| a62e25e09e05.. | 172 | 5.38% | 24,051 |

More than 5 order count



Product Category Association



Pareto Chart



Observation

- It is observed that around 20 product categories account for 80% of Overall Sale count.
- Close to 40-50 % of the sales of furniture decor, bed bath table, sports_leisure, fashion bags and accessories and auto have low frequency or hardly any sale products
- Flowers, home comfort, fashion children's clothing , furniture mattress and upholstery , security and services and many more have hardly any sale

Observations In Market Basket Analysis

Top five products categories in groups of two's are:

1. Toys and Bed Bath Table
2. Toys and Fashion Bags Accessories
3. Toys and Auto
4. Toys and Watches Gift
5. Toys and Health & Beauty

Top five products categories in groups of three's are:

1. Toys, Cine photos and Telephony
2. Toys, Home Construction and Computer Accessories
3. Toys, Garden Tools and Computer Accessories
4. Toys Furniture Decor and Electronics
5. Toys, Furniture Decor and Health and Beauty

Thank You

