CAPSTONE PROJECT : MARKETING AND RETAIL ANALYTICS

BY – PRABHAT KUMAR SINGH

PROBLEM DEFINITION

- OList is a 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.
- 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.

DATA EXPLORATION AND CLEANING

- I have used Python notebook for data exploration and cleaning treatment.
- In order table we are interested in orders which are successfully delivered only. rest are not useful for analysis. Hence, filtered and created new data frame :

```
# creating new DF where the status of orders are delivered
orders = orders[orders.order_status =='delivered']
```

- Missing values are identified :

```
# checking missing values
orders.isna().sum().sort_values(ascending = False)
```

- order_approved_at and order_delivered_timestamp had missing values
- Missing values were filled with the values of:
- order_approved_at = order_purchase_timestamp
- order_delivered_timestamp = orders_estimated_delivery_dateMissing values are identified in table Products :

```
# checking missing values
products.isna().sum().sort_values(ascending = False)
```

- Below columns had missing values:
  - product_category_name & product_weight_g
  - product_length_cm & product_height_cm & product_width_cm

- Missing values were filled with the values of:
- product_category_name = products.product_category_name.mode()[0]
  - product_weight_g = product_weight_g.median()
  - product_length_cm = product_length_cm. median()
  - product_height_cm = product_height_cm. median()
  - product_width_cm = product_width_cm. median()

- In customers table I have found duplicates values.

- Duplicates values are identified :

```
# looking for duplicates
customers.customer_id.duplicated().sum()
```

- 3345 duplicates values were present which was removed :

```
#dropping duplicates values
customers.drop_duplicates(subset='customer_id', keep='first',inplace=True)
```

- Data set is now cleaned.

EXPORTING DATA SET

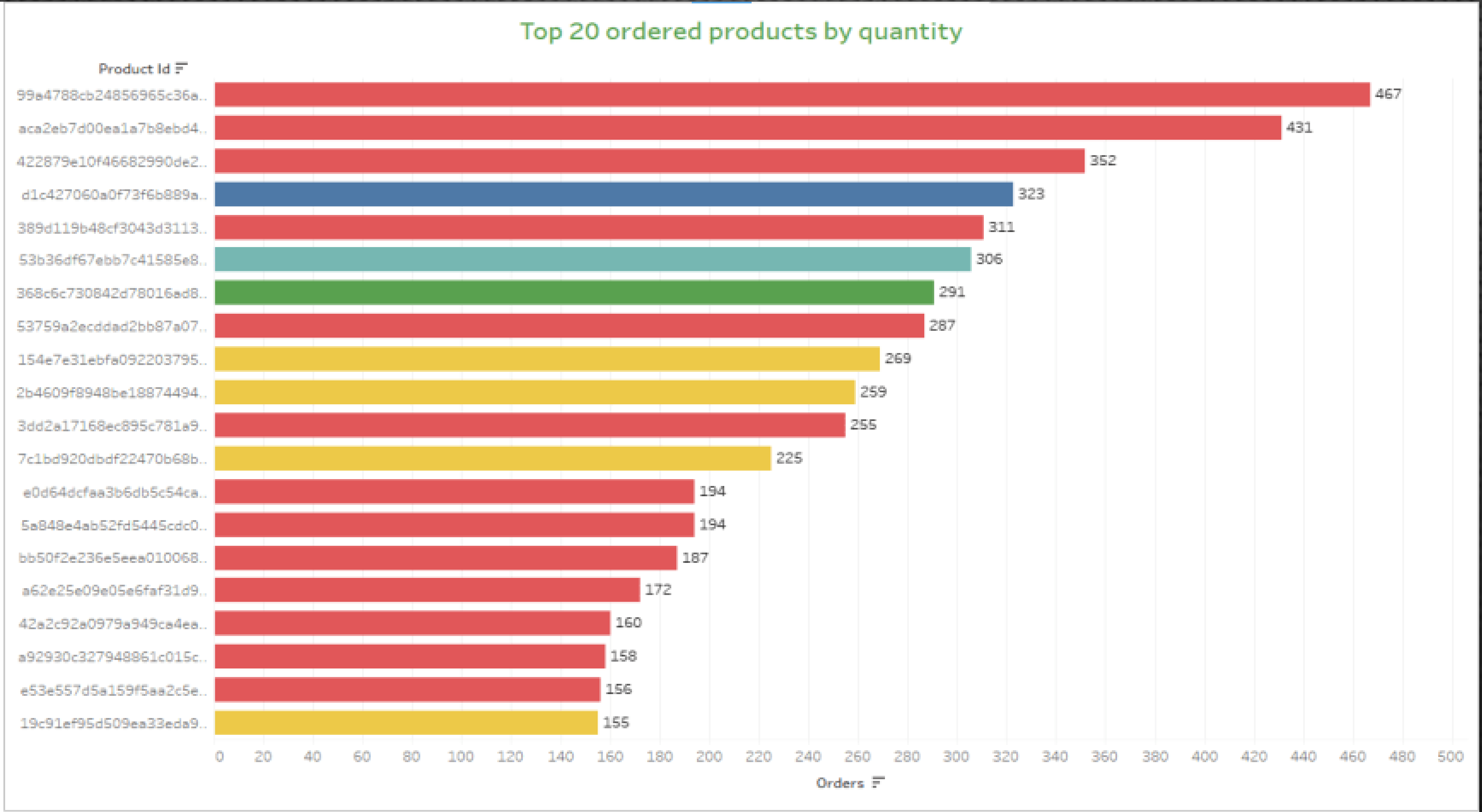
- Exporting to excel file named as "Retail_dataset_cleaned.xls"
- I have used pandas library writer function to export the data set to excel file:

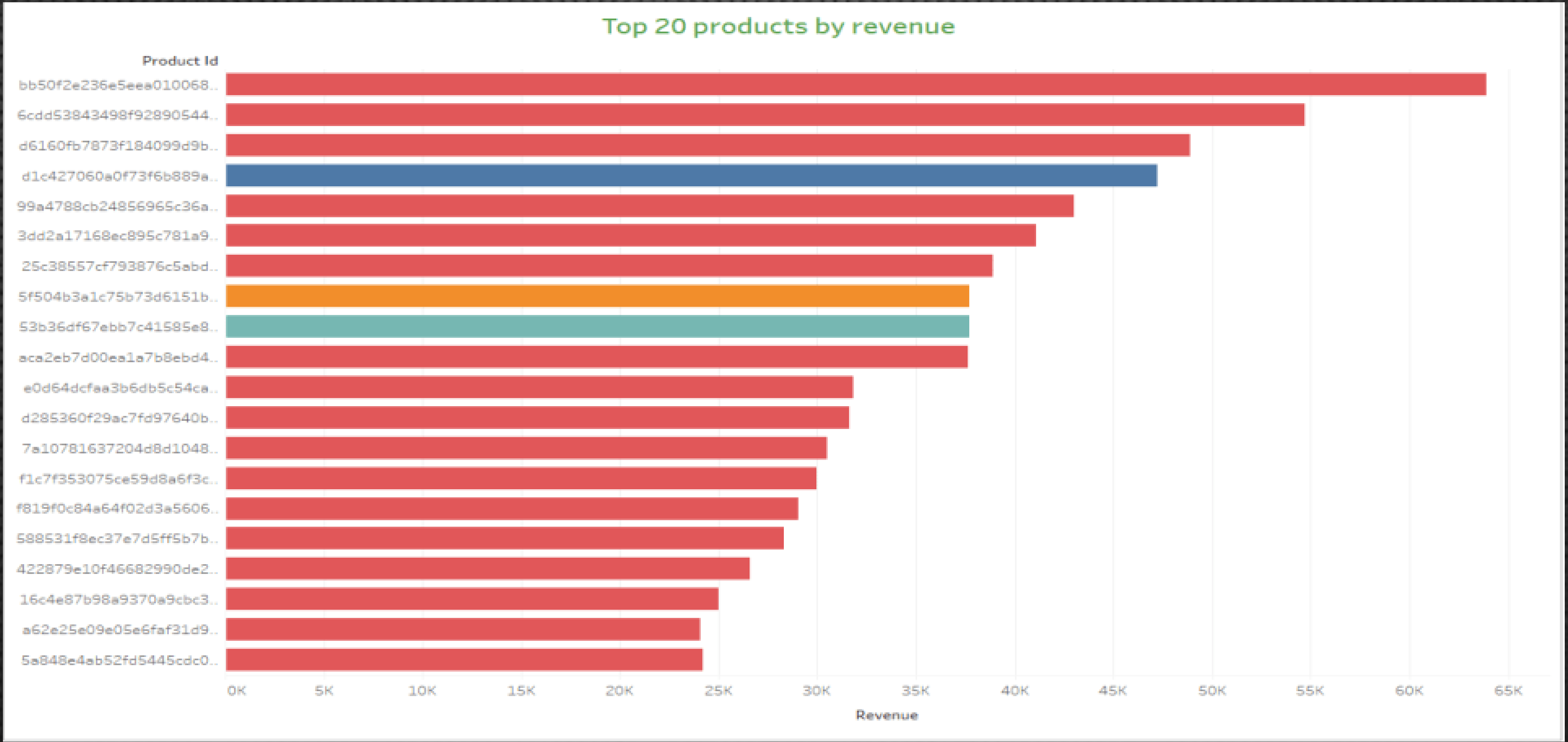
```
writer = pd.ExcelWriter('Retail_dataset_cleaned.xlsx', engine='xlsxwriter')

# Write each data frame to a different worksheet.
orders.to_excel(writer, sheet_name='orders')
order_items.to_excel(writer, sheet_name='order_items')
customers.to_excel(writer, sheet_name='customers')
payments.to_excel(writer, sheet_name='payments')
products.to_excel(writer, sheet_name='products')

# Close the Pandas Excel writer and output the Excel file.
writer.save()
```

DATA VISUALISATION





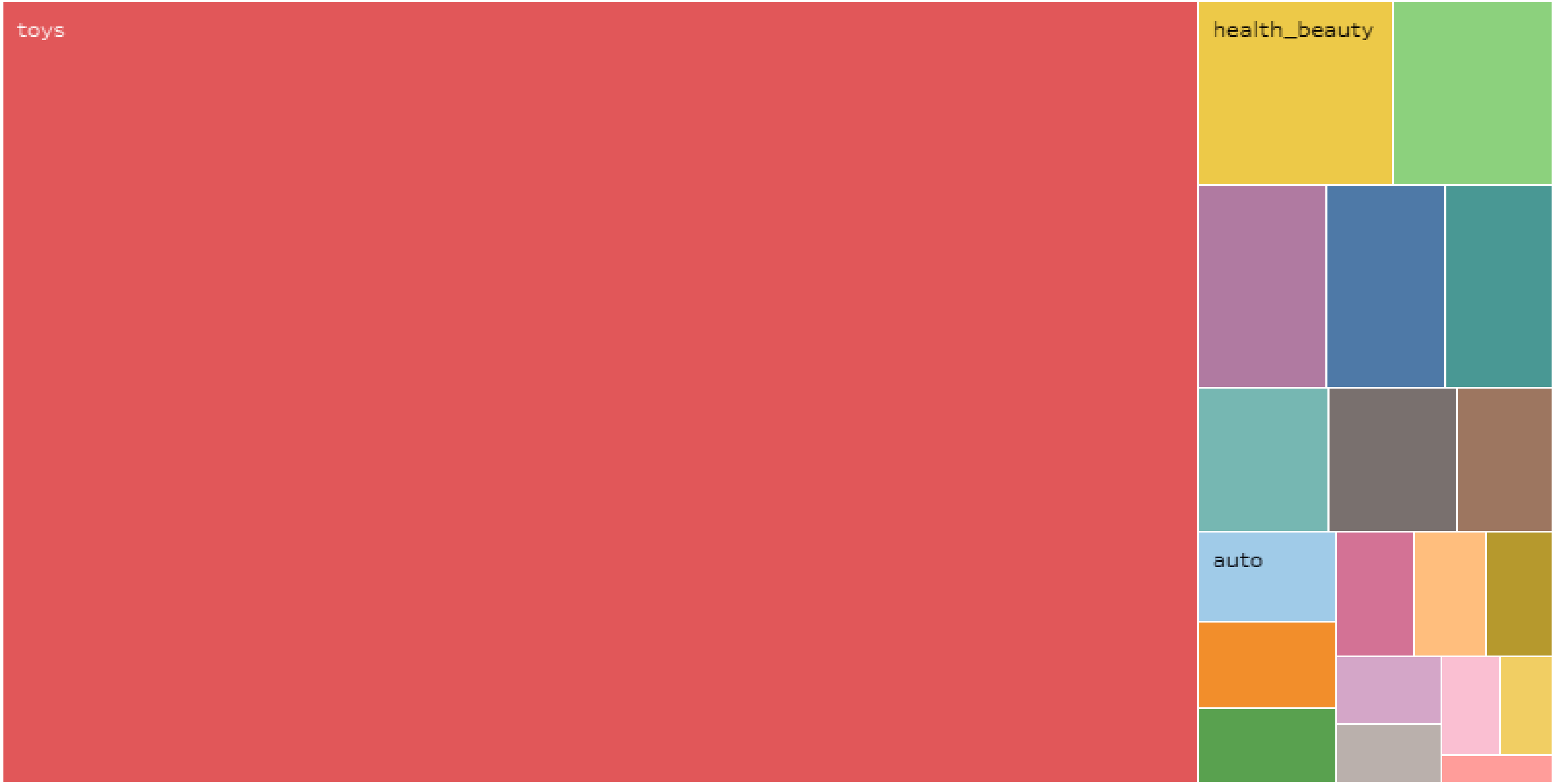
Toys in the product categories contributes maximum towards revenue generation.

- Percentage Running total with Product ID, count of Order Items and Revenue has been broken down by Product ID.
- I have applied filter to see the TOP 20 products.
- Toys in the product categories contributes more than 80% in revenue generation.

Revenue Pareto

Product Id	Revenue_	% of Total Revenue	Count of Products
bb50f2e236e5eea010068..	63,885	8.73%	195
6cdd53843498f92890544..	54,730	7.48%	156
d6160fb7873f184099d9b..	48,899	6.68%	35
d1c427060a0f73f6b889a..	47,215	6.45%	343
99a4788cb24856965c36a..	43,026	5.88%	488
3dd2a17168ec895c781a9..	41,083	5.61%	274
25c38557cf793876c5abd..	38,907	5.32%	38
5f504b3a1c75b73d6151b..	37,734	5.16%	63
53b36df67ebb7c41585e8..	37,683	5.15%	323
aca2eb7d00ea1a7b8ebd4..	37,609	5.14%	527
e0d64dcfaa3b6db5c54ca..	31,787	4.34%	194
d285360f29ac7fd97640b..	31,624	4.32%	123
7a10781637204d8d1048..	30,468	4.16%	143
f1c7f353075ce59d8a6f3c..	29,997	4.10%	154
f819f0c84a64f02d3a560..	29,024	3.97%	45
588531f8ec37e7d5ff5b7..	28,292	3.87%	20
422879e10f46682990de2..	26,577	3.63%	484
16c4e87b98a9370a9cbc3..	25,034	3.42%	13
5a848e4ab52fd5445cdc0..	24,229	3.31%	197
a62e25e09e05e6faf31d9..	24,051	3.29%	226

Category Wise order

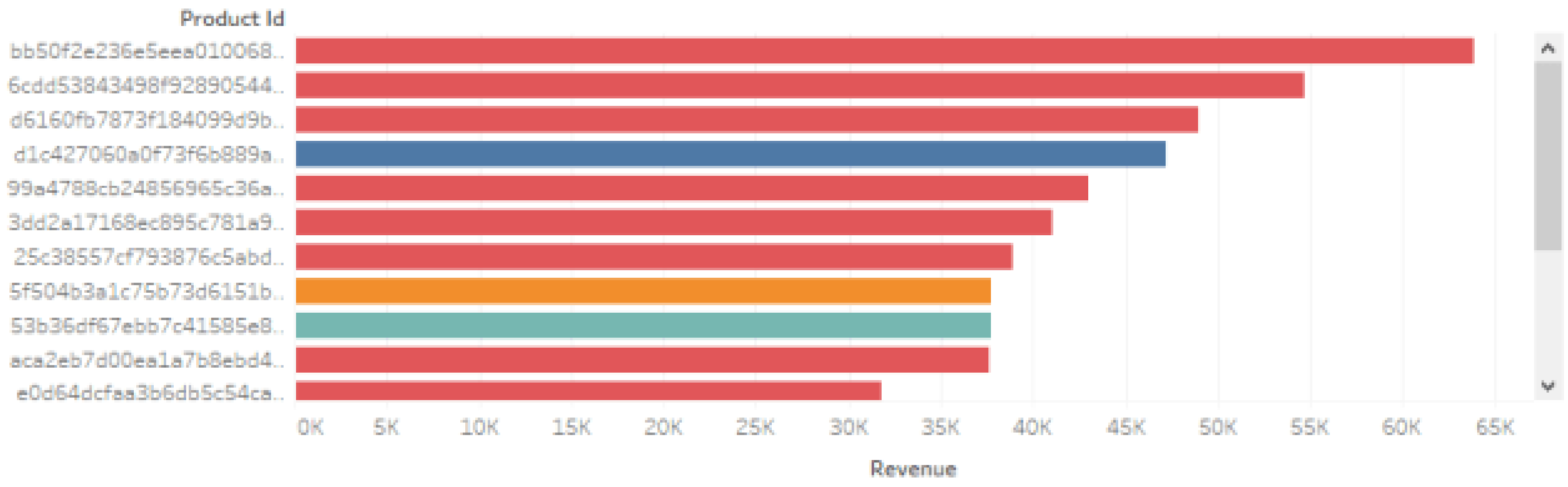


DASHBOARD

Top Products and Revenue Pareto

Top 20 ordered products by quantity

Top 20 products by revenue

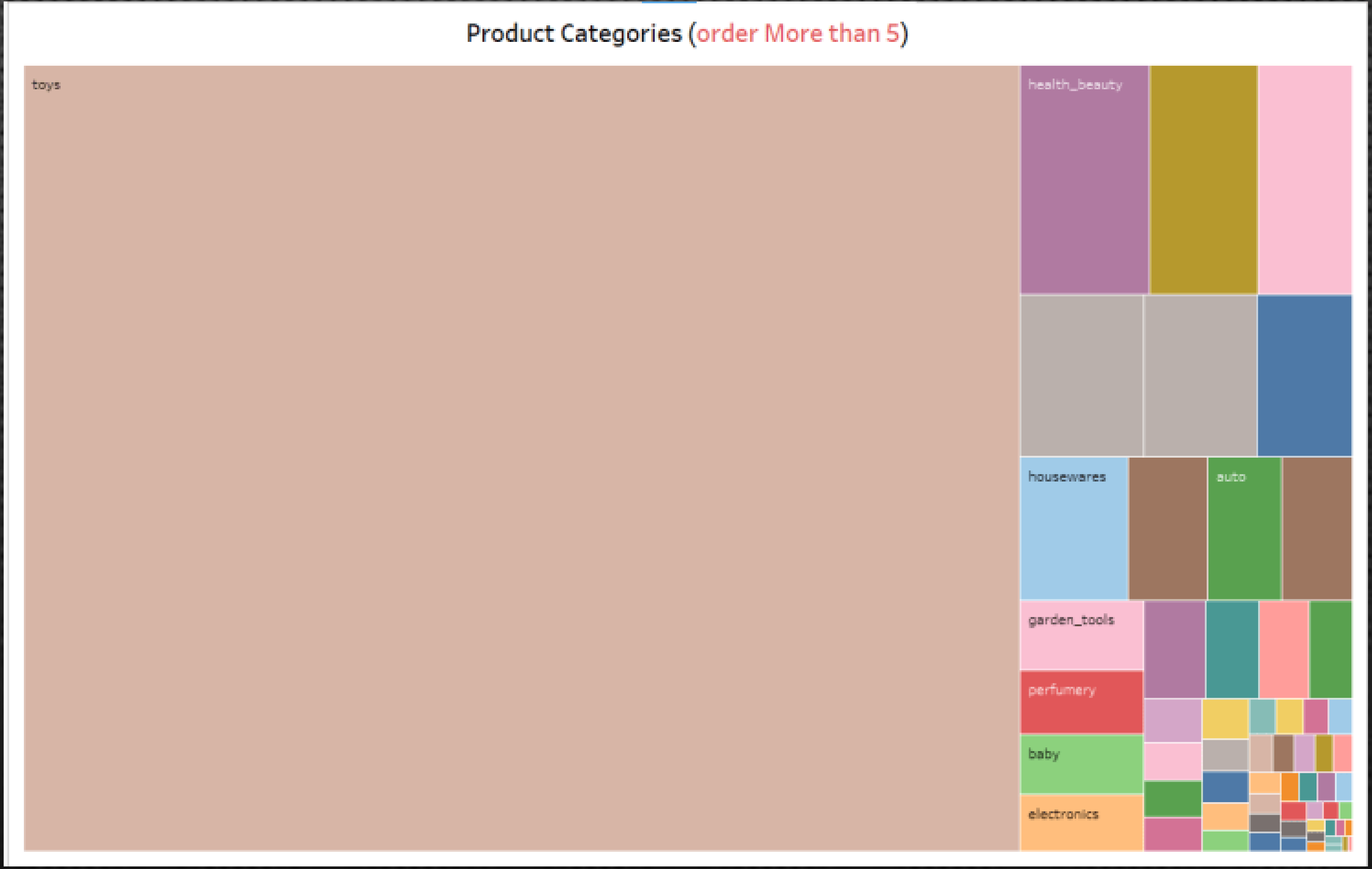


Revenue Pareto

Product Id	Revenue_	% of Total Revenue	Count of Products
bb50f2e236e5..	63,885	8.73%	195
6cdd53843498..	54,730	7.48%	156
d6160fb7873f..	48,899	6.68%	35
d1c427060a0f..	47,215	6.45%	343
99a4788cb248..	43,026	5.88%	488
3dd2a17168ec..	41,083	5.61%	274
25c38557cf79..	38,907	5.32%	38
5f504b3a1c75..	37,734	5.16%	63
53b36df67ebb..	37,683	5.15%	323
aca2eb7d00ea..	37,609	5.14%	527
e0d64dcfaa3b..	31,787	4.34%	194
d285360f29ac..	31,624	4.32%	123
7a1078163720..	30,468	4.16%	143
f1c7f353075ce..	29,997	4.10%	154
f819f0c84a64f..	29,024	3.97%	45
588531f8ec37..	28,292	3.87%	20
422879e10f46..	26,577	3.63%	484
16c4e87b98a9..	25,034	3.42%	13
5a848e4ab52f..	24,229	3.31%	197
a62e25e09e05..	24,051	3.29%	226

MARKET BASKET ANALYSIS

This Treemap shows Product Categories which Are ordered more than 5 Times.



Combinations of product categories (ordered together)

Product Category Name

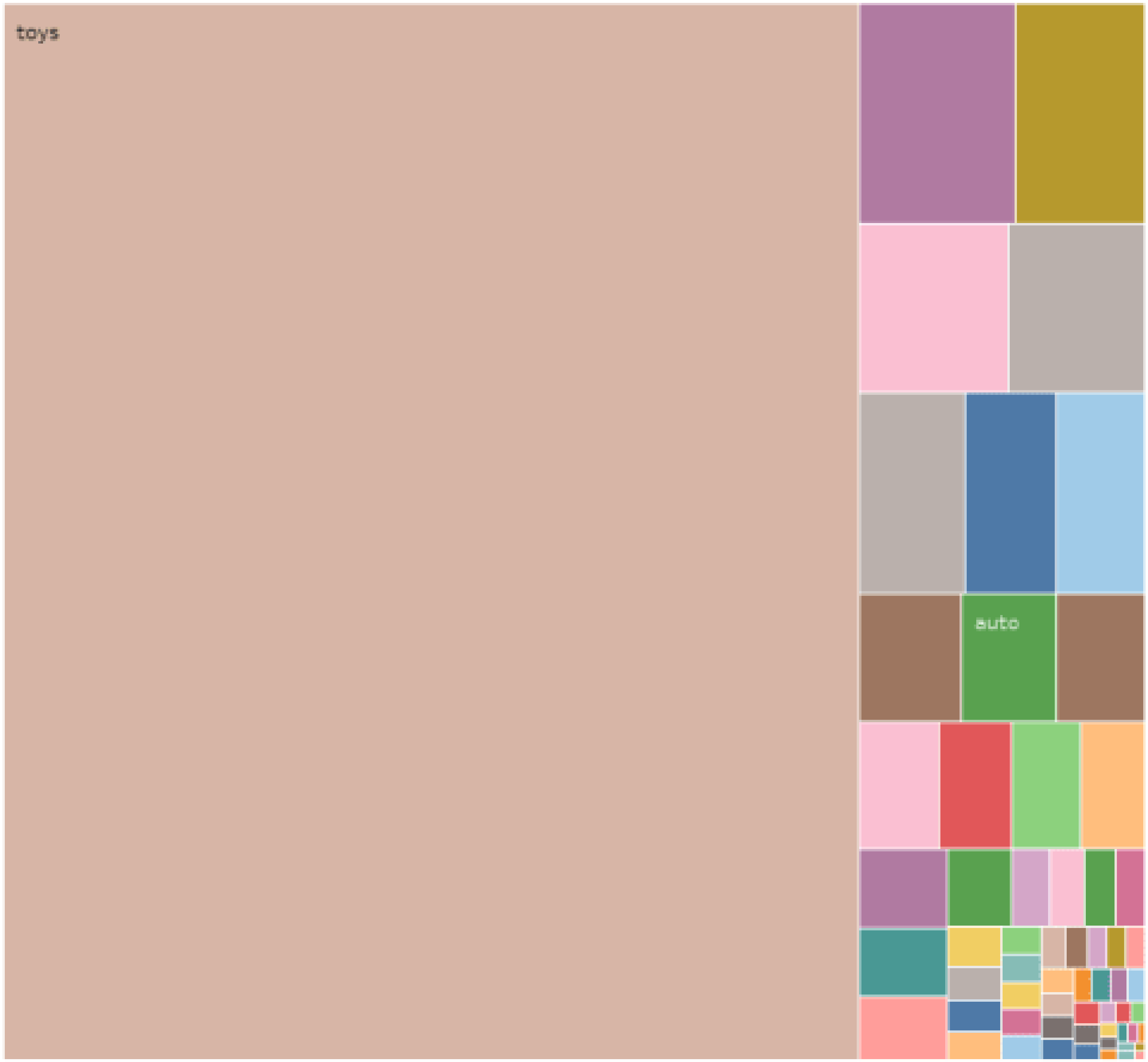
Product Cate..	c..	housew..	industr..	kitchen..	luggage..	market..	music	musical..	office_f..	party_s..	perfum..	pet_sh..	signalin..	small_a..	small_a..	sports_..	station..	tablets..	telepho..	toys	watche..
toys	7	63			11	7		8	18		32	23				62	30		30		73
bed_bath_table																				291	
furniture_decor																				153	
computers_acces..																				103	
health_beauty																				81	
watches_gifts																				73	
sports_leisure																				62	
housewares																				63	
garden_tools																				45	
fashion_bags_acc..																				41	
auto																				39	
perfumery																				32	
telephony																				30	
stationery																				30	
cool_stuff																				25	
pet_shop																				23	
baby																				20	
office_furniture																				18	
construction_tool..																				12	
luggage_accessor..																				11	
costruction_tools..																				11	
home_construction																				7	
fashion_shoes																				8	
electronics																				8	
musical_instrume..																				8	
market_place																				7	
consoles_games																				6	
Null	33	1,371	49	65	183	87	7	170	469	11	714	435	38	141	15	1,965	537	65	1,043	72,020	1,390

Market Basket Analysis Dashboard

Combinations of product categories (ordered together)

	Product Category Name									
Product Cate..	..	signalin..	small_a..	small_a..	sports_..	station..	tablets..	telepho..	toys	watche..
toys	3				62	30		30		73
bed_bath_table									291	
furniture_decor									153	
computers_acces..									103	
health_beauty									81	
watches_gifts									73	
sports_leisure									62	
housewares									63	
garden_tools									45	
fashion_bags_acc..									41	
auto									39	
perfumery									32	
telephony									30	
stationery									30	
cool_stuff									25	
pet_shop									23	
baby									20	
office_furniture									18	
construction_tool..									12	
luggage_accessor..									11	
costruction_tools..									11	
home_construction									7	
fashion_shoes									8	
electronics									8	
musical_instrume..									8	
market_place									7	
consoles_games									6	
Null	5	38	141	15	1,965	537	65	1,043	72,020	1,390

Product Categories (order More than 5)



INSIGHTS

- The category 'Toys' constitutes 20% of the products which generates 80% of revenue.
- It can be seen that even if the price of the certain products are high, it is still bought by the customer.
- The categories apart from Toys which are at least ordered 5 times are health_beauty, bed_bath_table, Sports_leisure, computer_accessories, furniture_decor, watches_gifts and have high order quantity.
- Hence, the combination of all these few categories with Toys and each other are frequently high.

RECOMMENDATIONS & BUSINESS IMPACT

- Company should focus on giving offers for the customers who are frequent buys to retain them.

- Since, there are many categories of products, company can reduce sub categories for ease of access and should cut down few of products which least contributing towards revenue generation.

- As Toys are most selling product company can target customers who are more likely to buy it like – parents of newly born and young children.

- Company can given some good discounts or offers to gain new customers.