

OList Marketing and Retail Analytics: Capstone Project

Presented By,

Dr. Snehal Yogesh Hole, Mr. Roy Mathew Benjamin and Mr. Rahul Singh

EPGP - Batch - DSC-49- OCT-2022
upGrad & IIT Bangalore

Agenda

1. Objectives
2. Background
3. Visualizations
4. Insights
5. Recommendations
6. Appendix –
 - Data Sources
 - Data Methodology
 - Data Assumptions

Objective

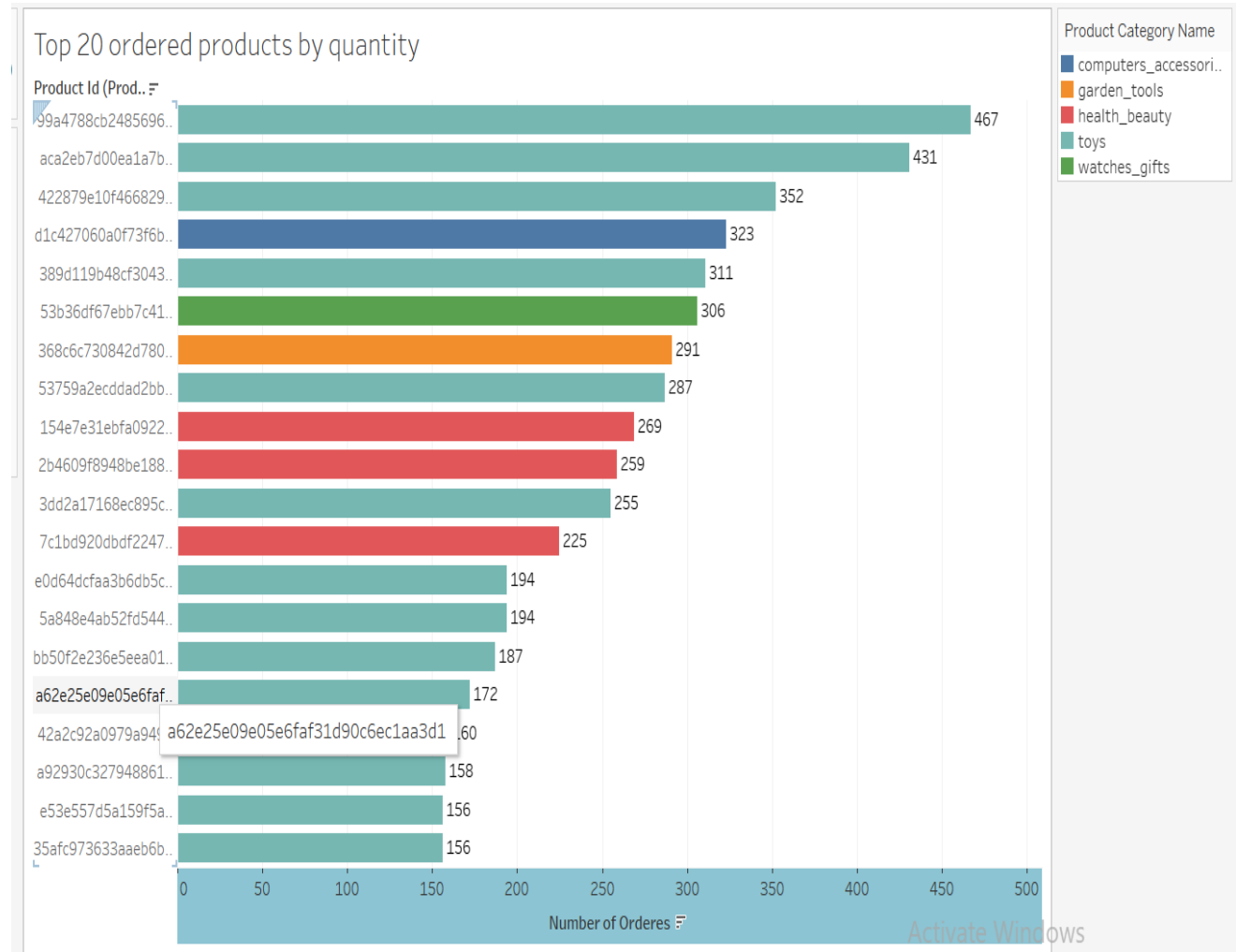
- To identify top products that contribute to the revenue and top product category using Pareto Analysis.
- Improve our understanding with the use of market basket analysis to analyze the purchase behavior of customers.
- Understand what items are most likely to be purchased individually or in combination with some other products.

Background

- OList is an e-commerce company that has faced some losses recently and they want to manage their inventory so as to reduce any unnecessary costs.
- Now to be able to meet the demands of the customers, the company would need to store tons and tons of products in warehouses.
- Since storing these products adds to the costs that the company incurs, it is necessary for the organization to plan their inventory well.

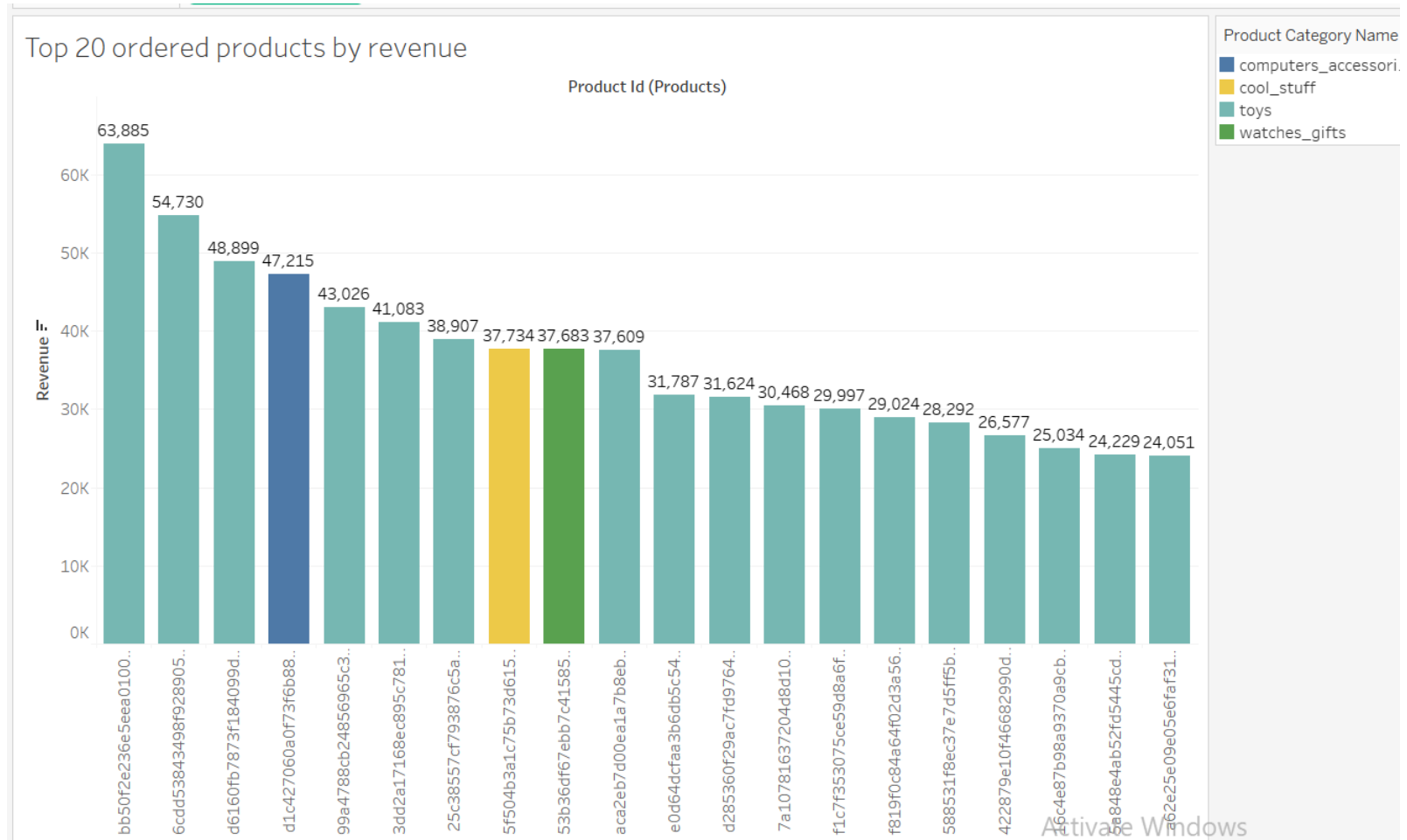
Top 20 Ordered Products by Quantity

- The highest ordered product is from the Toys category and has been ordered 467 times.
- Most of the products in the Top 20 that are frequently ordered belong to the Toys category.



Top 20 Ordered Products by Revenue

- The highest revenue generation is 63,885 which belongs to the Toys Category.
- Most of the products in the Top 20 list generating high revenue belong to the Toys category.



Activate Windows

Percent Running Totals

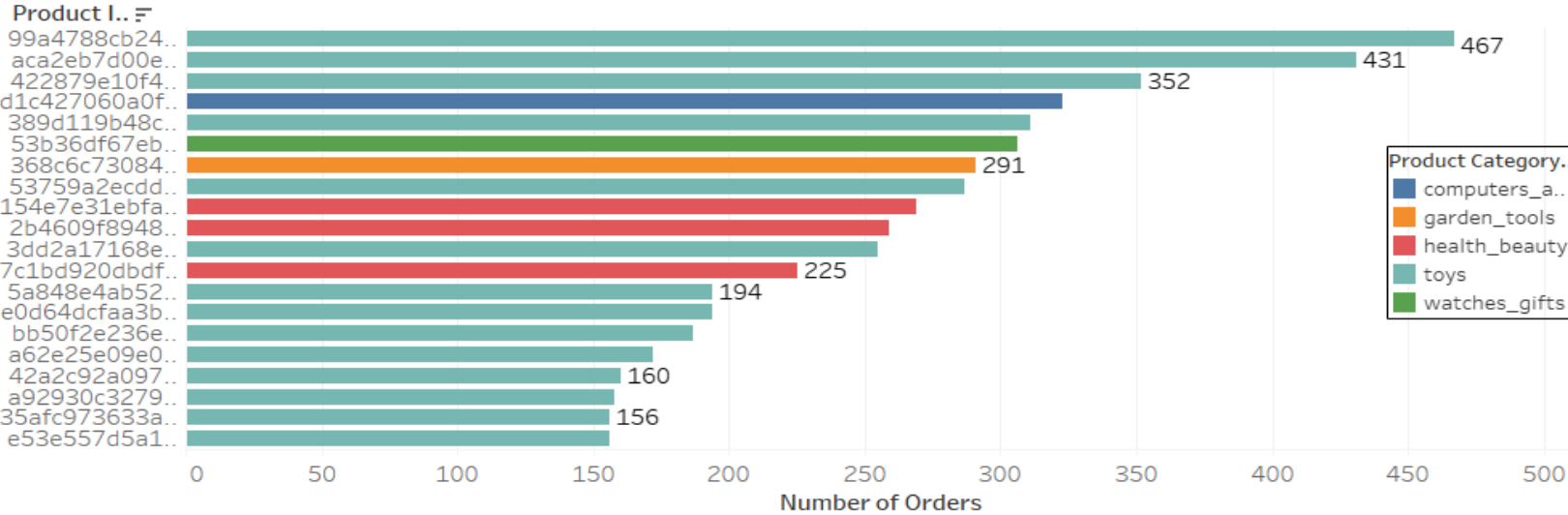
- The Percentage of Total Running Revenue and Quantity Ordered has been broken down by Product Id.
- The contribution of each product towards the total revenue can be identified.

Percentage Running Totals

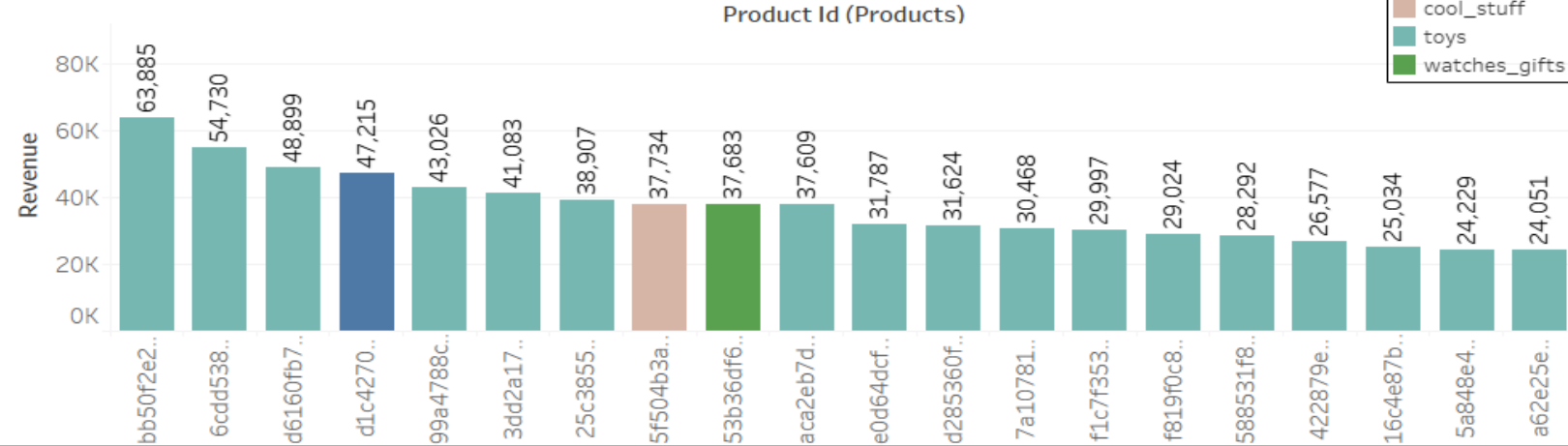
Product Id (Products)	Revenue	% of Tot..	Total Co..	% of Tot..
bb50f2e236e5eea0..	63,885	0.47%	186	0.19%
6cdd53843498f928..	54,730	0.87%	148	0.35%
d6160fb7873f1840..	48,899	1.23%	33	0.38%
d1c427060a0f73f6b..	47,215	1.58%	313	0.70%
99a4788cb2485696..	43,026	1.90%	456	1.18%
3dd2a17168ec895c..	41,083	2.20%	253	1.44%
25c38557cf793876c..	38,907	2.48%	38	1.48%
aca2eb7d00ea1a7b..	37,609	2.76%	425	1.92%
53b36df67ebb7c41..	37,683	3.04%	304	2.23%
5f504b3a1c75b73d..	37,734	3.32%	63	2.30%
e0d64dcfaa3b6db5c..	31,787	3.55%	193	2.50%
d285360f29ac7fd97..	31,624	3.78%	118	2.62%
7a10781637204d8d..	30,468	4.01%	140	2.77%
f1c7f353075ce59d8..	29,997	4.23%	149	2.92%
f819f0c84a64f02d3..	29,024	4.44%	44	2.97%
588531f8ec37e7d5f..	28,292	4.65%	19	2.99%
422879e10f466829..	26,577	4.85%	352	3.35%
16c4e87b98a9370a..	25,034	5.03%	13	3.37%
5a848e4ab52fd544..	24,229	5.21%	187	3.56%
a62e25e09e05e6faf..	24,051	5.38%	170	3.74%
2b4609f8948be188..	22,717	5.55%	254	4.00%
fd0065af7f09af4b8..	22,000	5.71%	10	4.01%
a5215a7a9f46c418..	21,740	5.87%	16	4.03%
389d119b48cf3043..	21,441	6.03%	309	4.35%
bc4cd4da98dd128c..	21,500	6.19%	17	4.36%
368c6c730842d780..	21,057	6.34%	291	4.67%
461f43be3bdf8844e..	20,593	6.50%	142	4.81%
53759a2ecddad2bb..	20,387	6.65%	287	5.11%
52c80cedd4e90108..	20,496	6.80%	111	5.23%
1dec4c88c685d5a0..	19,965	6.94%	35	5.26%
6f3b5b605d91b743..	19,009	7.08%	116	5.38%
19c91ef95d509ea3..	18,889	7.22%	153	5.54%
1a080577618e7fe4..	18,263	7.36%	32	5.57%

Top 20 Products and Running Totals by Quantity and Revenue

Top 20 Ordered Products by Quantity



Top 20 Ordered Products by Revenue

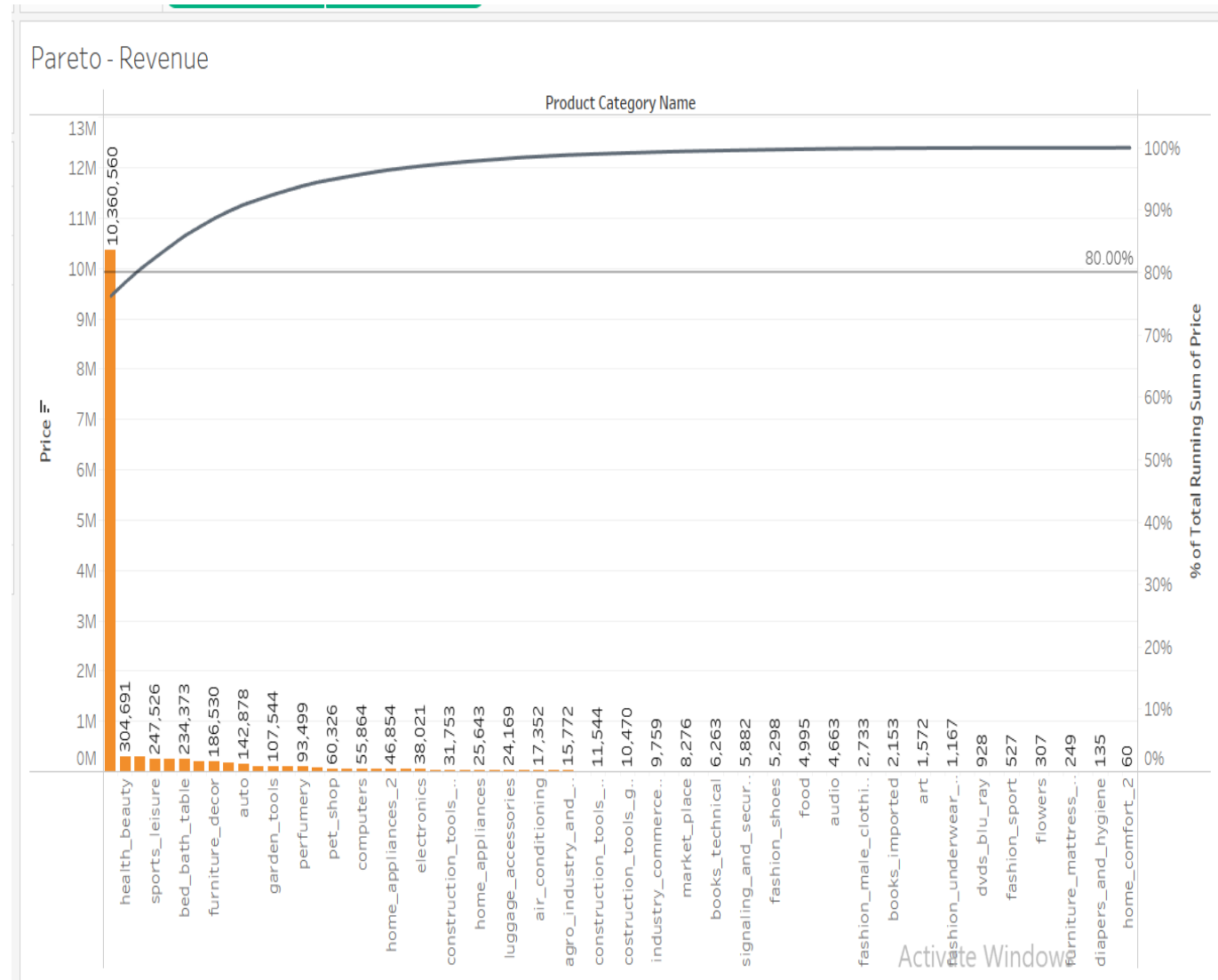


Percentage Running Totals by Revenue and Orders

Product Id	Revenue	% of Total Running ..	Total Quantity ..	% of Total Running ..
bb50f2e236e5..	63,885	0.47%	186	0.19%
6cdd5384349..				
d6160fb7873f..				
d1c427060a0f..				
Product Id: bb50f2e236e5ea010				
Total Quantity Ordered: 186				
99a4788cb248..	43,026	1.90%	456	1.18%
3dd2a17168ec..	41,083	2.20%	253	1.44%
25c38557cf79..	38,907	2.48%	38	1.48%
5f504b3a1c75..	37,734	2.76%	63	1.54%
53b36df67ebb..	37,683	3.04%	304	1.86%
aca2eb7d00ea..	37,609	3.32%	425	2.30%
e0d64dcfaa3b..	31,787	3.55%	193	2.50%
d285360f29ac..	31,624	3.78%	118	2.62%
7a107816372..	30,468	4.01%	140	2.77%
f1c7f353075c..	29,997	4.23%	149	2.92%
f819f0c84a64f..	29,024	4.44%	44	2.97%
588531f8ec37..	28,292	4.65%	19	2.99%
422879e10f46..	26,577	4.85%	352	3.35%
16c4e87b98a9..	25,034	5.03%	13	3.37%
5a848e4ab52f..	24,229	5.21%	187	3.56%
a62e25e09e05..	24,051	5.38%	170	3.74%
2b4609f8948b..	22,717	5.55%	254	4.00%
fd0065af7f09..	22,000	5.71%	10	4.01%
a5215a7a9f46..	21,740	5.87%	16	4.03%
bc4cd4da98dd..	21,500	6.03%	17	4.04%
389d119b48cf..	21,441	6.19%	309	4.36%
368c6c730842..	21,057	6.34%	291	4.67%
461f43be3bdf..	20,593	6.50%	142	4.81%
52c80cedd4e9..	20,496	6.65%	111	4.93%
53759a2ecdda..	20,387	6.80%	287	5.23%
1dec4c88c685..	19,965	6.94%	35	5.26%

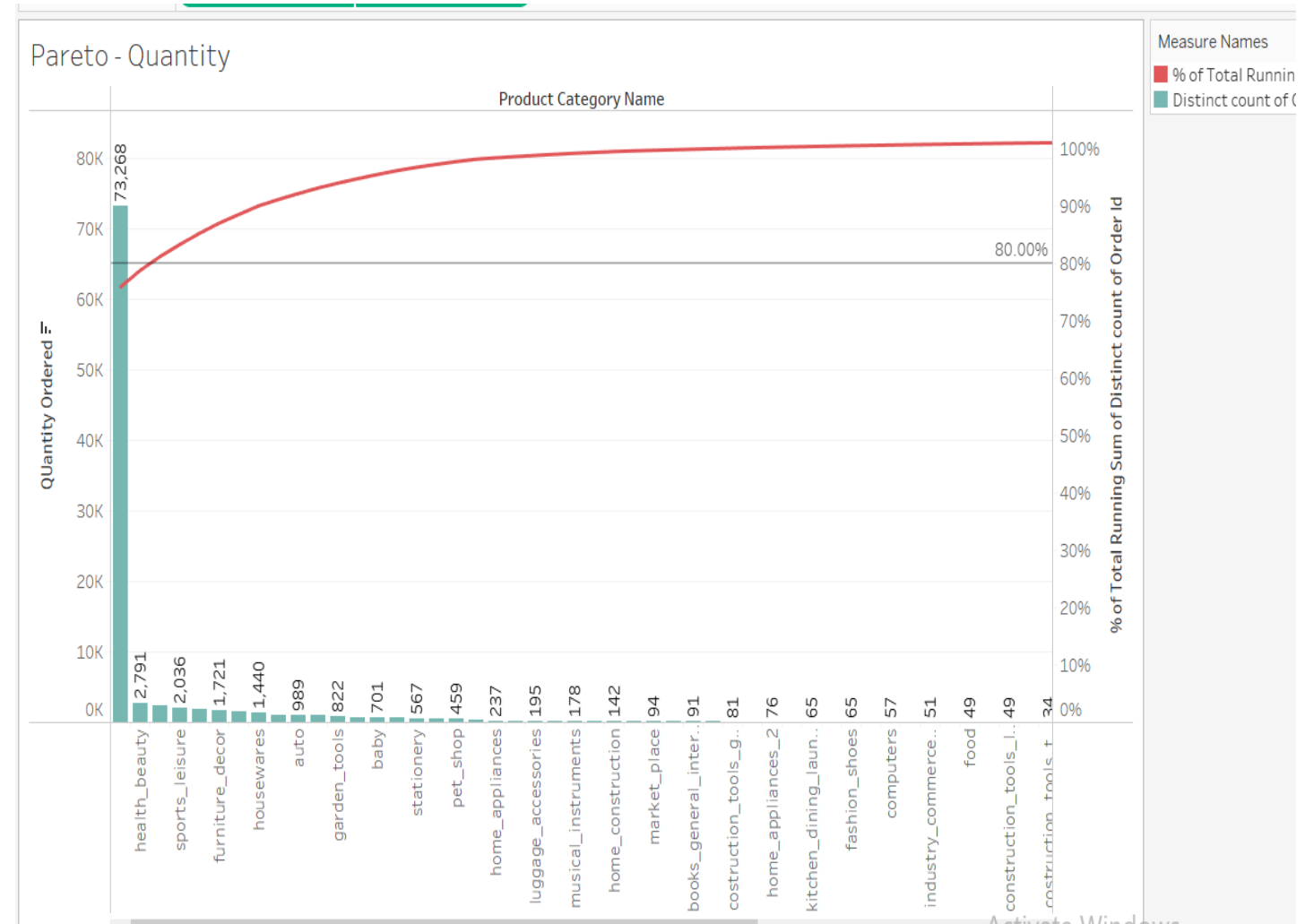
Revenue Pareto Analysis

- Toys, health_beauty and watches_gift combine generate 80.56% of the revenue.
- Toys alone generates 76.23% of the revenue.
- The rest of the 70+ product categories generates 19.44% of the revenue.

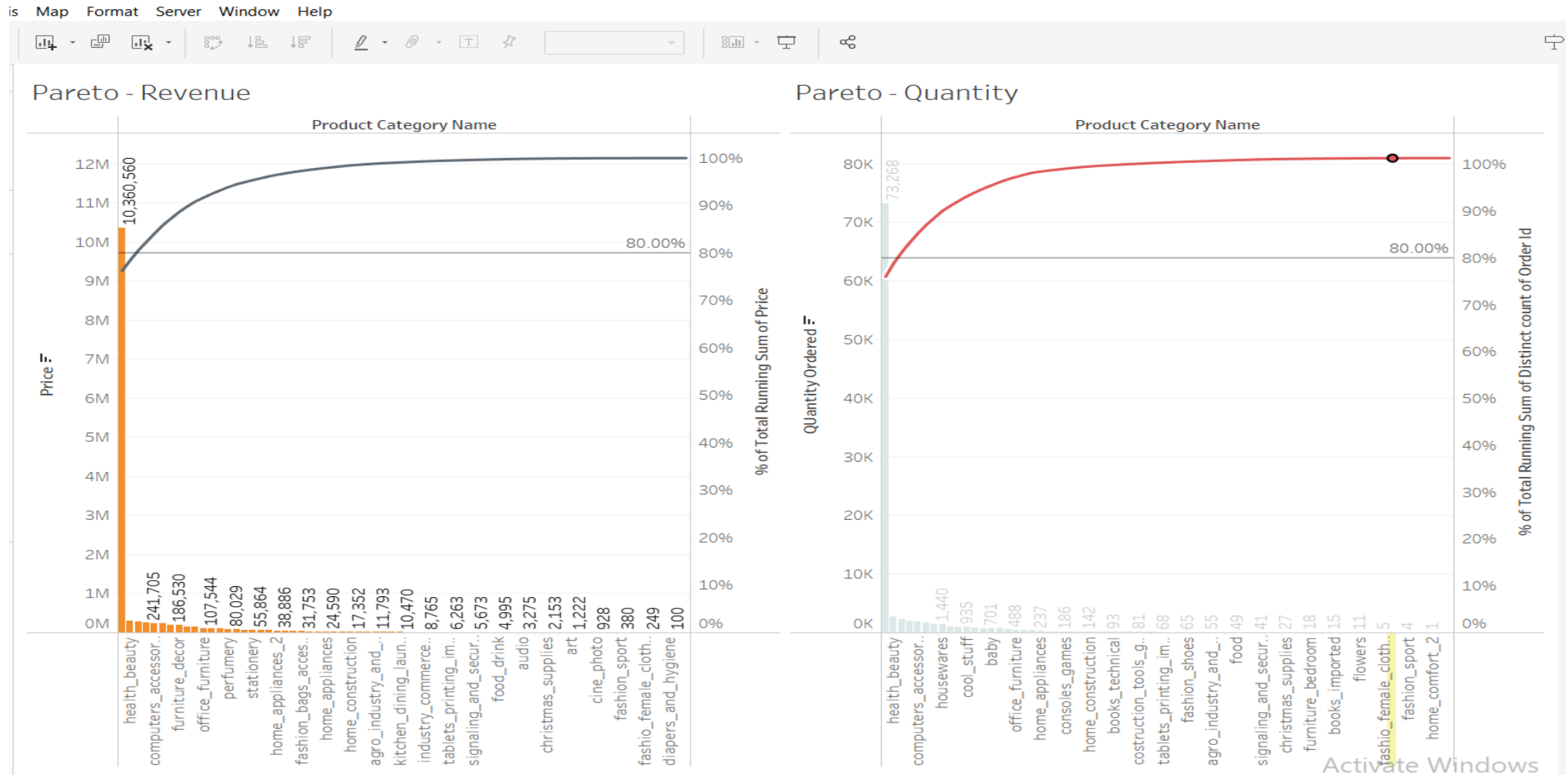


Quantity Pareto Analysis

- Toys, health_beauty and bed_bath_table make up 80.38% of the total orders.
- Toys alone has 75.94% of the total orders.
- The rest of the 70+ product categories generate 19.62% of the total orders.



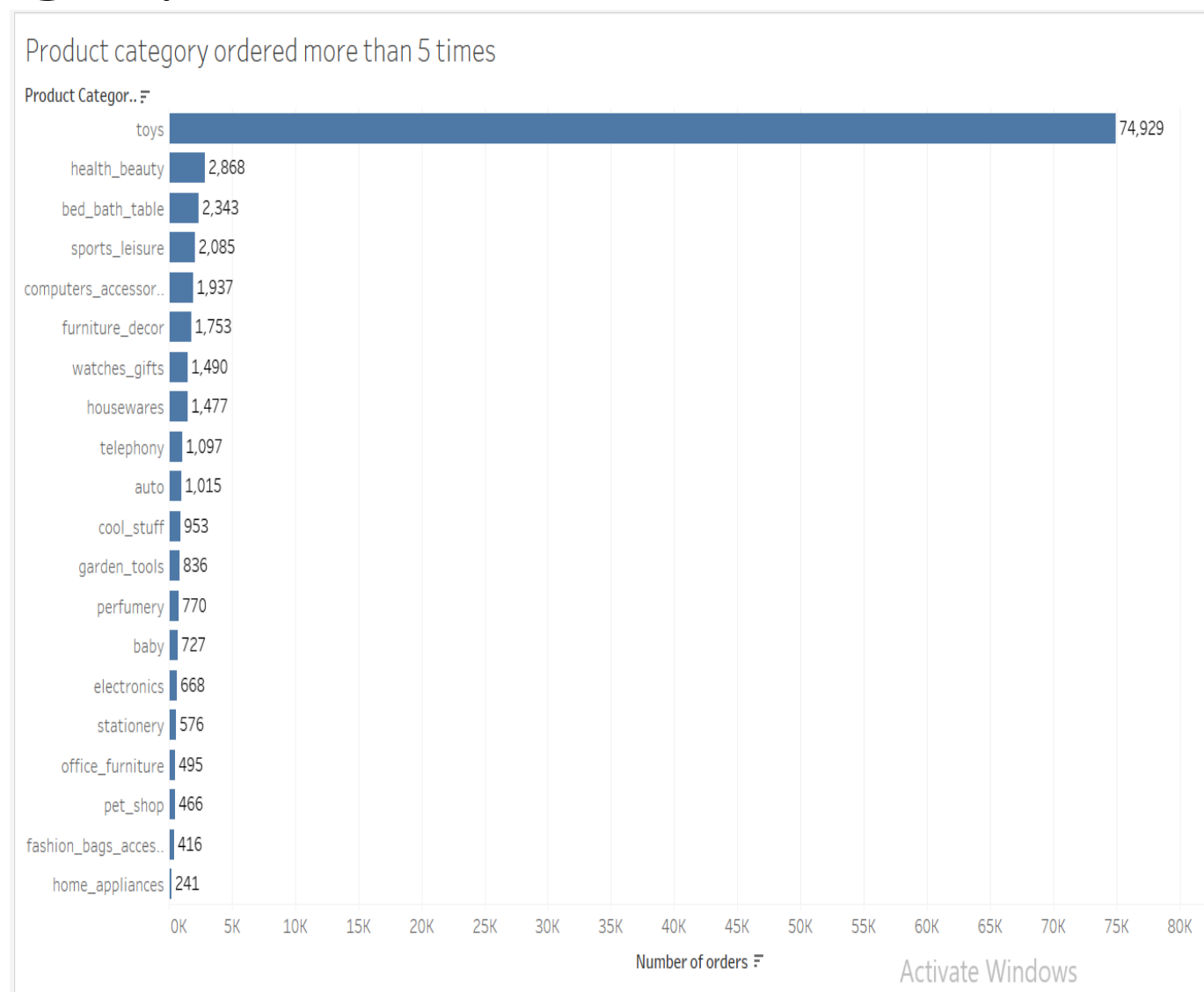
Pareto Analysis Dashboard



Activate Windows

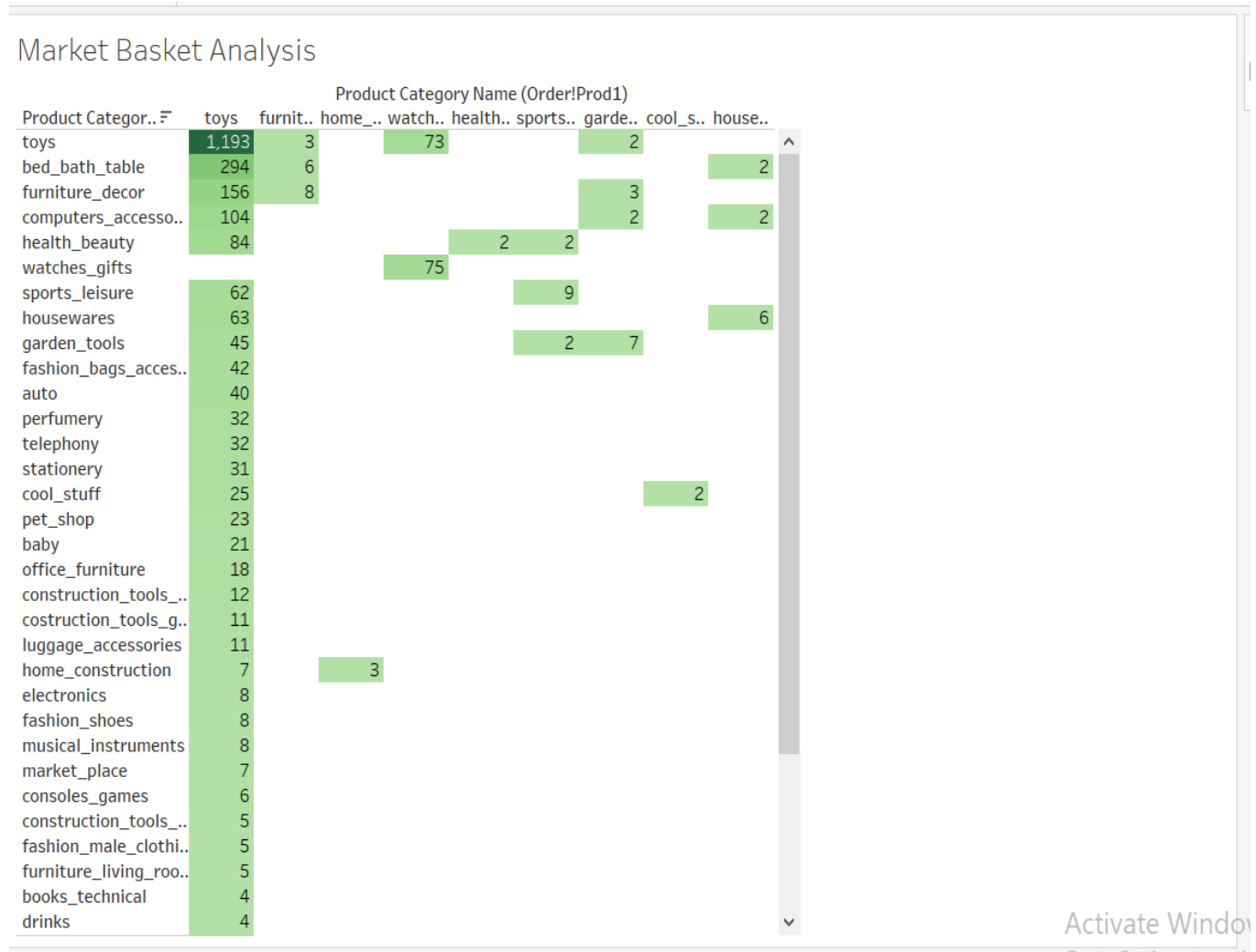
Product Category Ordered >5 Times

- Toys category is the most ordered category with a total of 74,929 orders.
- Health_beauty, bed_bath_table and sports_leisure are the next most ordered category.



Market Basket Analysis

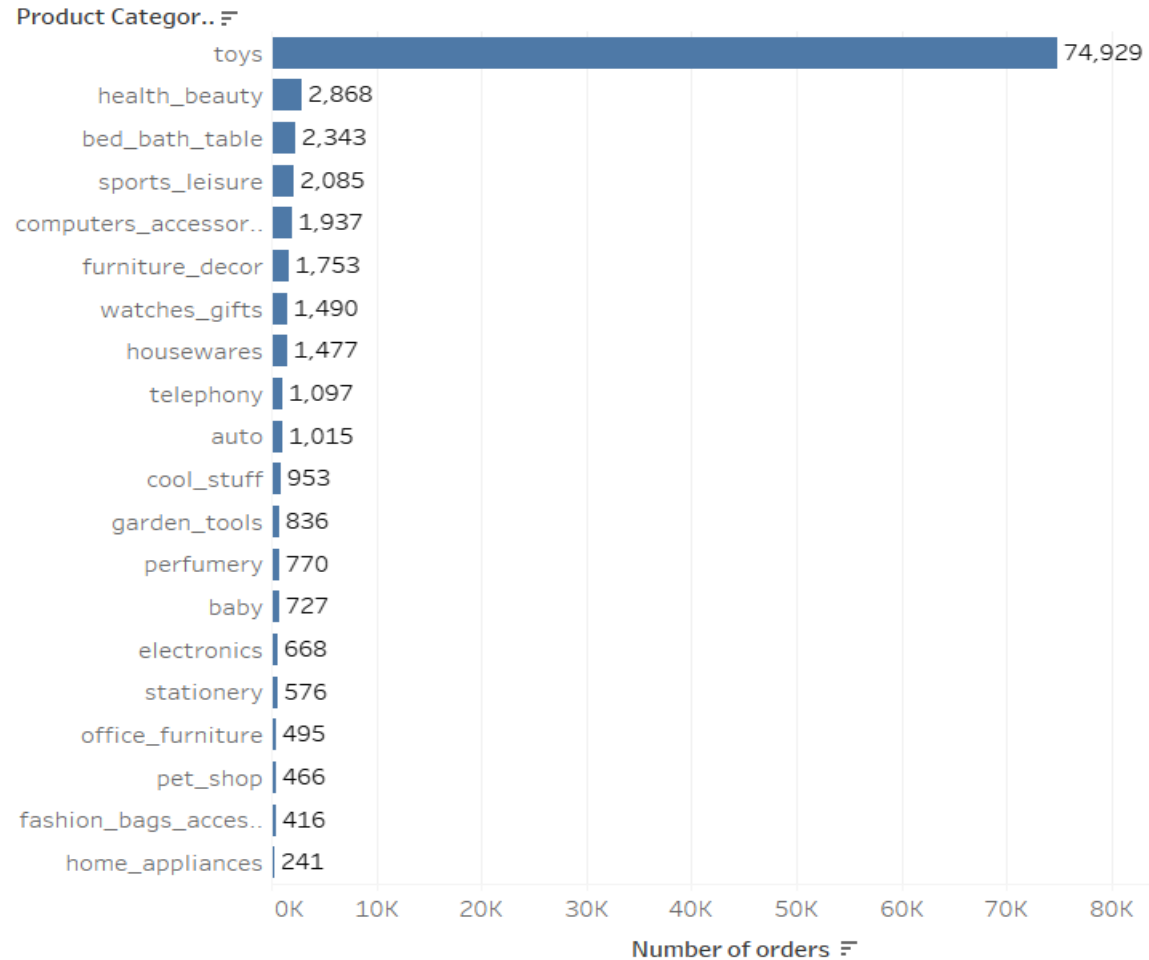
- Market Basket Analysis is performed to identify the frequently ordered category association.
- Toys are the most ordered category along with the categories of bed_bath_table, furniture_decor, computers_accessories and health_beauty.



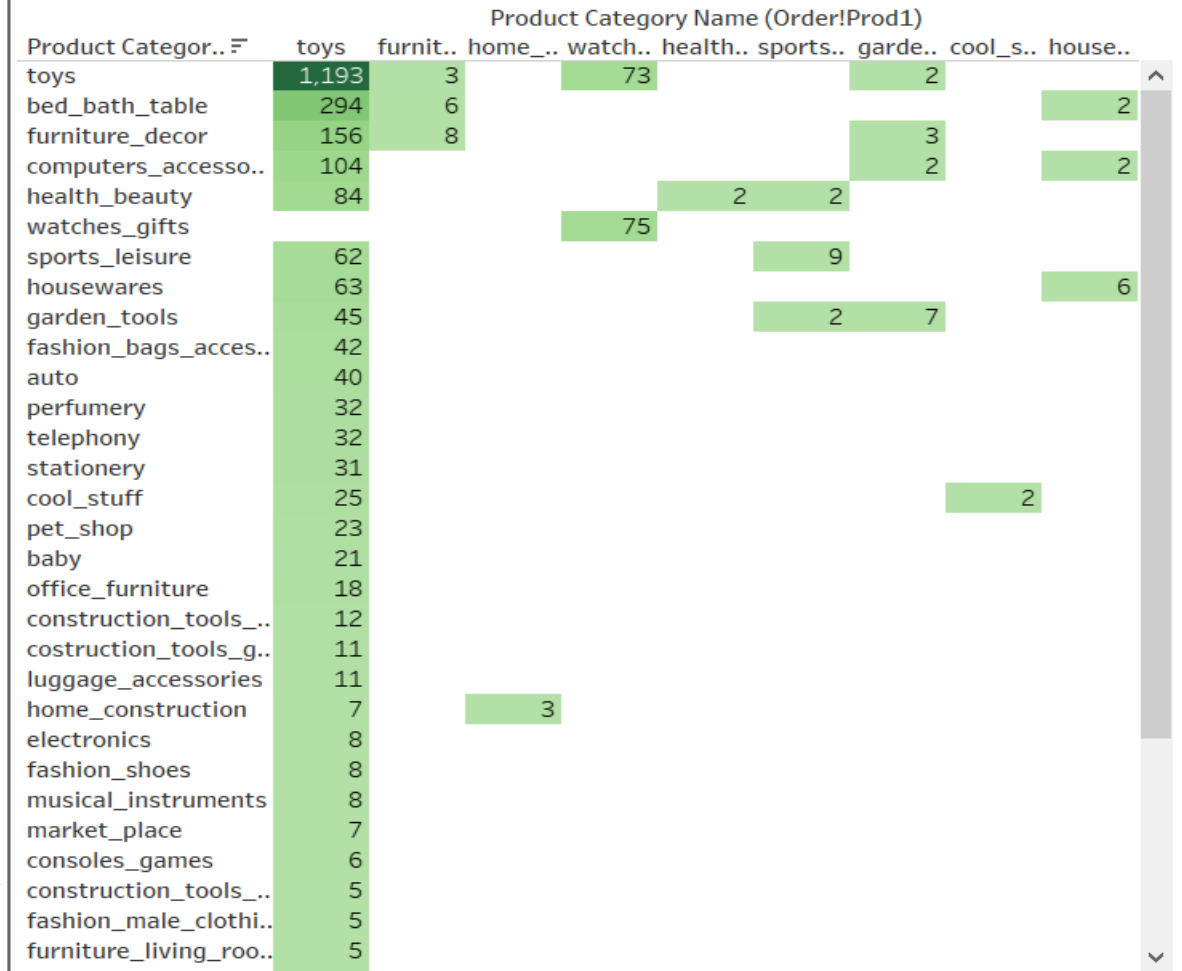
Market Basket Analysis Dashboard

Market Basket Analysis _Dashboard

Product category ordered more than 5 times



Market Basket Analysis



Insights

- The category Toys constitute 20% of the products which generates 80% of the revenue.
- It can be seen that even if the price of the certain products is high, it is still bought by the customer more often.
- Apart from Toys, the products from the categories of bed_bath_table, furniture_decor, computers_accessories and health_beauty are the most frequently ordered. The above categories with Toys or/and with each other are most frequent in customers' basket.
- It is observed that despite of the high price, some products are frequently purchased by the customers.

Recommendations

- The company should focus on the categories which generate more than 80% of the revenue by always keeping them in stock.
- The company should target customers who are more likely to buy toys to boost sales as the category toys is the most ordered category.
- Offer promo-codes or discounts on the frequently ordered category associations to encourage cross selling among the products.
- The company can reduce some of the sub categories which have very low sales.

Appendix - Data Sources

- Here is a snapshot of our data dictionary:
 - Order details such as order id, order status, order purchased timestamp, etc.
 - Order Items detail such as order item id, seller id, price, shipping charges, etc.
 - Customer details such as customer id, customer city, customer state, etc.
 - Payment details such as payment type, payment value, etc.
 - Product details such as product id, product category name, product dimensions, etc.
- The following data sources were used:
 - OList retail dataset containing order-related information.
 - The data consisted for the year 2016 to 2018.

Appendix - Data Methodology

A thorough analysis of the OList Retail Dataset was conducted. The process included:

- The dataset was cleaned and transformed using the python libraries of Pandas and Numpy in the Jupyter Notebook.
- The missing values for the various columns were replaced with the best values.
- The redundant and duplicate records were discarded and only first occurrence is kept.
- Exploratory data analysis was done using the python libraries of Matplotlib and Seaborn in the Jupyter Notebook.
- A new dataset consisting of order id and product category name was created for Market Basket Analysis.
- Various visualizations and Market Basket Analysis was conducted in Tableau.

Step 1- EDA – Data analysis and cleaning

As we have provided the data in different excel sheets for the orders, order_items, customers, payments and products.. Need to check the null values, duplicates or any outliers in python

Step-2 – after cleaning the data need to download the new excel sheet for the further analysis and visualization in tableau

need to identify top products that contribute to the revenue and also we have use market basket analysis to analyze the purchase behaviour of individual customers to estimate with relative certainty.

We have to use the 80-20 rule of the Pareto analysis to calculate the ideal category

Identify the top 20 ordered products by quantity

the top 20 ordered products by revenue

the percentage running totals by revenue and number of orders

Market Basket Analysis

Identify the product categories which are ordered more than 5 times

Combinations of product categories which are frequently ordered together

Appendix - Data Assumptions

- Only the cases having order status as 'delivered' are considered.
- We assumed that the data provided was achieving the desired revenue.
- We assumed that the company does not want to expand to new warehouses.
- The company's strategies are decided considering there is constant growth in sales.

Conclusion

The OList marketing and retail data analysis provides the valuable insights of product performance, customer behavior and revenue. By implementing the given recommendations OList can optimize their inventory, boost sales and improve the profitability to ensure the more sustainable future of the company

Thank You