

MARKET & RETAIL ANALYSIS

BY: BHAVENDRA SAHU



AGENDA

Objectives

Background

Key findings

Recommendations

Appendix

- Data sources
- Data methodology
- Data Assumptions & Executive Summary

OBJECTIVE



To manage the inventory cost of E-commerce company Olist.



To help OList to identify the product categories which they can get rid of without significantly impacting business.



To identify top products that contribute to the revenue.



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.

BACKGROUND



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.



They need to identify top products that contribute to the revenue .

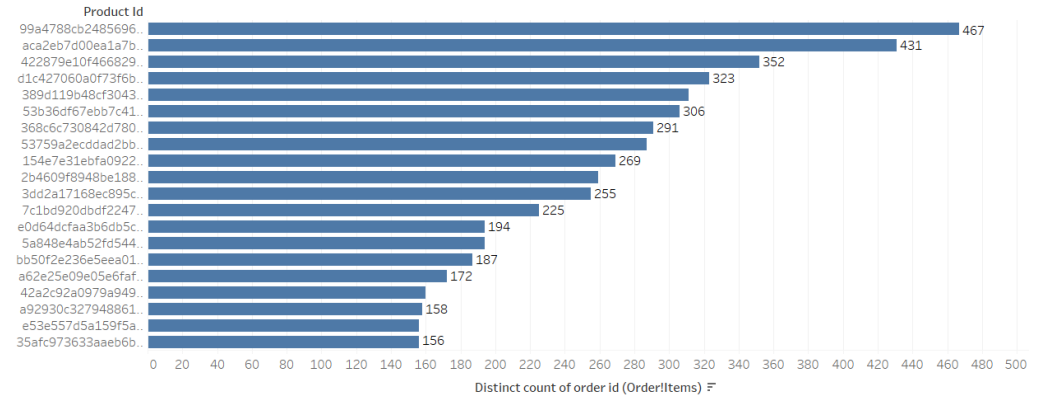


They want to analyze the purchase behavior of individual customers to estimate with relative certainty, what items are more likely to be purchased individually or in combination with some other products.

KEY FINDINGS

- The graph shows the top 20 products ordered by quantity.
- The most ordered product has been purchased **467** times.
- Most of the top products ordered are from 'toys' category.

Top 20 Products By Order Quantity

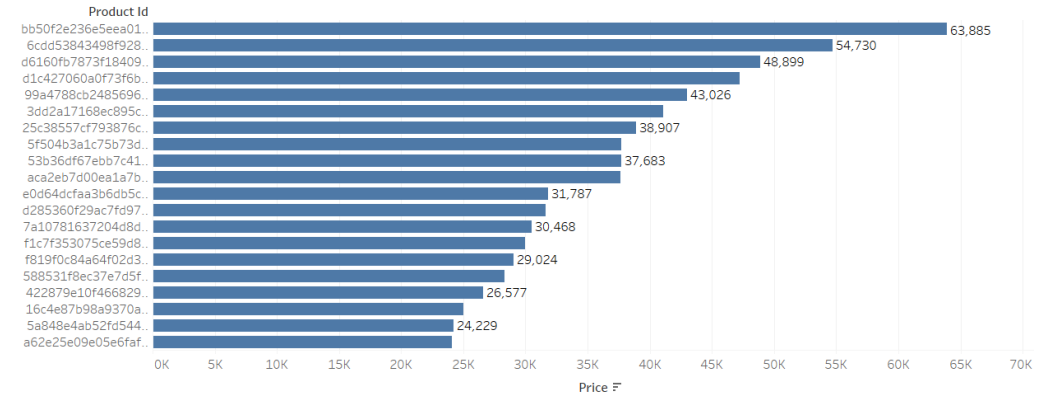


Distinct count of order id (OrderItems) for each Product Id. The marks are labeled by distinct count of order id (OrderItems). The data is filtered on Product Category Name, which keeps 20 of 70 members. The view is filtered on Product Id, which keeps 20 of 32,951 members.

TOP 20 PRODUCTS BY REVENUE

- The highest revenue generated by the top product is **63,885**.
- The product that generated highest revenue belongs to the **‘Toys’** category.
- Most of the products belongs to the **‘toys’** category contributing to higher revenue.

Top 20 Products by revenue



Sum of Price for each Product Id. The marks are labeled by sum of Price. The data is filtered on Product Category Name, which keeps 20 of 70 members. The view is filtered on Product Id, which has multiple members selected.

REVENUE PARETO

- This Revenue Pareto shows the number of orders , percentage of total running revenue and revenue generated for each product id.
- About 25% of the top products having higher revenues contributing to around 80% of total revenue.
- It helps to identify the contribution of the products towards total revenue.

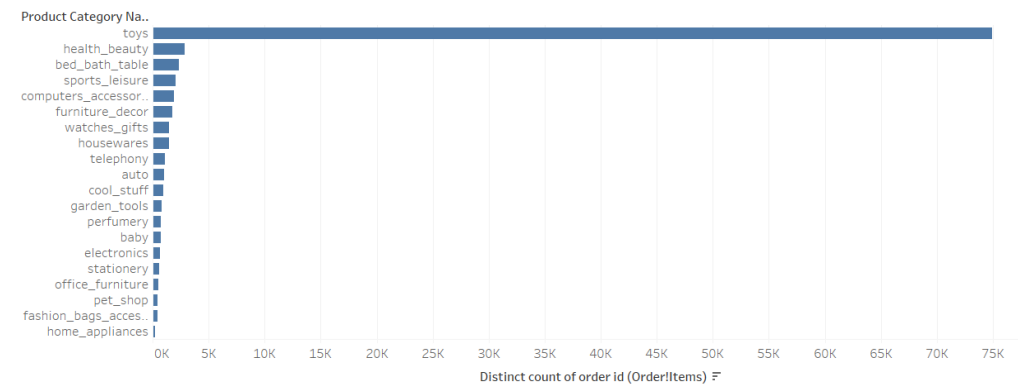
Revenue Analysis

Product Id	Count of ..	% of Tot..	Price
bb50f2e236e5eea0..	186	0.49%	63,885
6cdd53843498f928..	148	0.91%	54,730
d6160fb7873f1840..	33	1.28%	48,899
d1c427060a0f73f6b..	313	1.64%	47,215
99a4788cb2485696..	456	1.97%	43,026
3dd2a17168ec895c..	253	2.29%	41,083
25c38557cf793876c..	38	2.59%	38,907
aca2eb7d00ea1a7b..	425	2.87%	37,609
53b36df67ebb7c41..	304	3.16%	37,683
5f504b3a1c75b73d..	63	3.45%	37,734
e0d64dcfaa3b6db5c..	193	3.70%	31,787
d285360f29ac7fd97..	118	3.94%	31,624
7a10781637204d8d..	140	4.17%	30,468
f1c7f353075ce59d8..	149	4.40%	29,997
f819f0c84a64f02d3..	44	4.62%	29,024
588531f8ec37e7d5f..	19	4.84%	28,292
422879e10f466829..	352	5.04%	26,577
16c4e87b98a9370a..	13	5.24%	25,034
5a848e4ab52fd544..	187	5.42%	24,229
a62e25e09e05e6faf..	170	5.61%	24,051
2b4609f8948be188..	254	5.78%	22,717
fd0065af7f09af4b8..	10	5.95%	22,000
a5215a7a9f46c418..	16	6.11%	21,740
389d119b48cf3043..	309	6.28%	21,441
368c6c730842d780..	291	6.44%	21,057
461f43be3bdf8844e..	142	6.60%	20,593
53759a2ecddad2bb..	287	6.75%	20,387
52c80cedd4e90108..	111	6.91%	20,496
1dec4c88c685d5a0..	35	7.06%	19,965
6f3b5b605d91b743..	116	7.21%	19,009
19c91ef95d509ea3..	153	7.35%	18,889
1a080577618e7fe4..	32	7.49%	18,263
165f86fe8b799a708..	104	7.63%	17,821

PRODUCT CATEGORIES ORDERED MORE THAN 5 TIMES

- The graph shows the top 20 categories ordered more than 5 times.
- Toys, health_beauty , bed_bath_tables, sports_leisure & computer_accessories etc. are the top products ordered more than 5 times.
- The 'Toys' category is the most ordered category with a total of 74,929 orders.

Products Ordered more than 5 times



Distinct count of order id (OrderItems) for each Product Category Name. The view is filtered on Product Category Name, which keeps 20 of 70 members.

MARKET BASKET ANALYSIS

- Market basket analysis is performed to identify the frequently ordered category association.
- 'bed_bath_table' associated very well with 'toys'.
- 'Furniture_decor', 'computers_accessories' go along pretty well with 'toys'.
- 'health_beauty', 'housewares', 'watches_gifts' etc. are also be considered to order with 'toys'.
- Hence the combination of all of these few categories with toys & each other are frequently ordered.

Market Basket Summary

Product Category Name	watch..	furnit..	toys	Distinct count of O..
bed_bath_table	6	294	6	294
furniture_decor		156		
computers_accesso..		104		
health_beauty		84		
housewares		63		
sports_leisure		62		
garden_tools		45		
fashion_bags_acces..		42		
auto		40		
telephony		32		
perfumery		32		
stationery		31		
cool_stuff		25		
pet_shop		23		
baby		21		
office_furniture		18		
construction_tools_..		12		
luggage_accessories		11		
costruction_tools_g..		11		
musical_instruments		8		
fashion_shoes		8		
electronics		8		
market_place		7		
home_construction		7		
consoles_games		6		
toys	73			

Distinct count of Order Id broken down by product category name (Order!ProductCat!2) vs. Product Category Name. Color shows distinct count of Order Id. The marks are labeled by distinct count of Order Id. The view is filtered on distinct count of Order Id, which ranges from 6 to 294.

RECOMMENDATIONS

Target customers who are more likely to buy '**toys**' such as parents of newly born and 1-2 years child's.

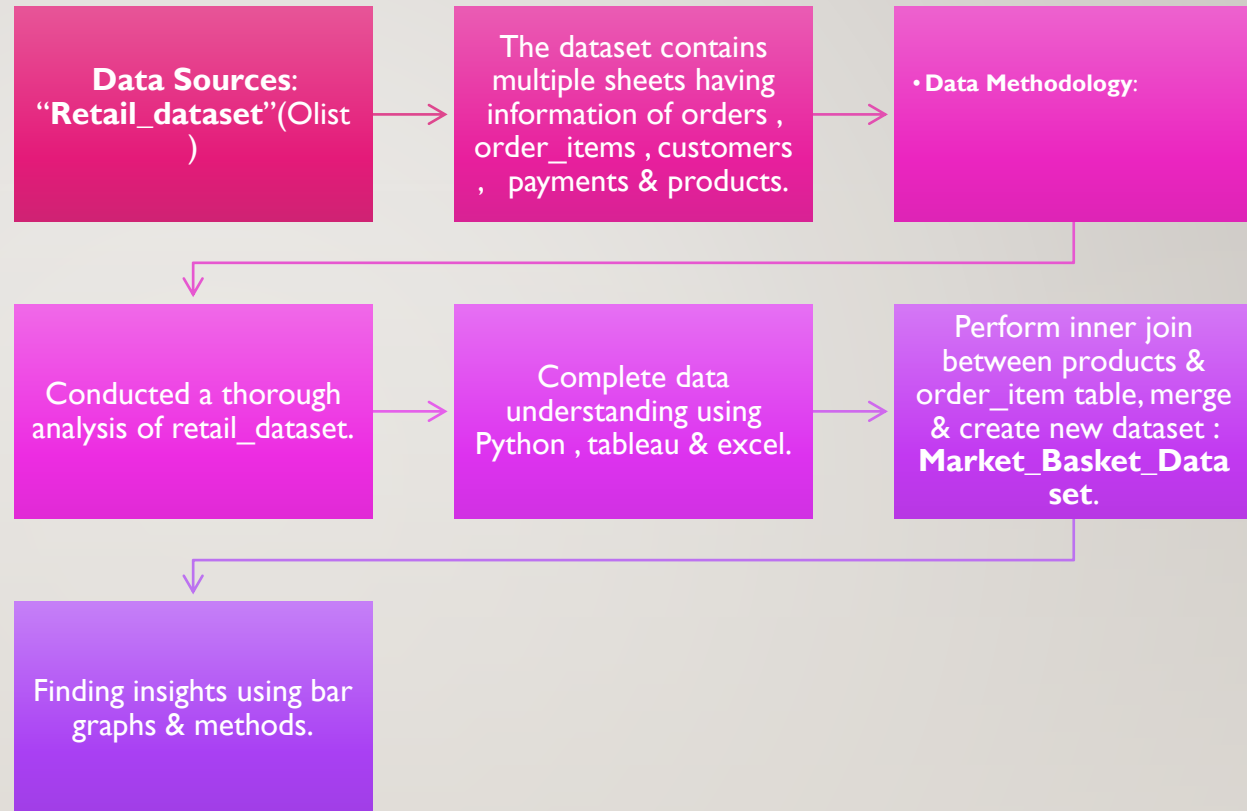
Offer some discounts or coupons on frequently ordered category associations and the most ordered products to attract more customers.

Olist can consider removing unnecessary products & categories which are not much of a major contributors neither in terms of maximum number of orders nor in terms of revenue.

Olist can optimize average price & shipping charges for some of the products which are not much of contributing to revenue.

'**Toys**' should be held in enough stocks as they are contributing to significant amount of revenue.

APPENDIX



DATA ASSUMPTIONS & EXECUTIVE SUMMARY

- Only the cases having order status as 'delivered' are to be considered.
- The columns 'order_approved_at' and 'order_delivered_timestamp' are assumed to be equivalent to/same as the columns 'order_purchase_timestamp' and 'order_estimated_delivery_date' respectively.

Data Cleaning using Python:

- **ORDERS** : Only the orders with order status as 'delivered' are considered for this case study since 97% of the records are successfully delivered. The missing values for 'order_approved_at' are replaced with respective 'order_purchase_timestamp'. Similarly, the missing values for 'order_delivered_timestamp' are replaced with respective 'order_estimated_delivery_date'.
- **Order Items** : No missing values , no duplicate values , No outlier treatment done as it may lead to data loss.
- **Customers** : No missing values , Duplicate customer id's were dropped , No outliers detected.
- **Payments**: No missing values , Duplicate order id's were dropped , No outlier treatment done as it may lead to data loss.
- **Products**: The missing values of product category name has been imputed with the mode , product weight , height , length & width imputed with median , No duplicate values .