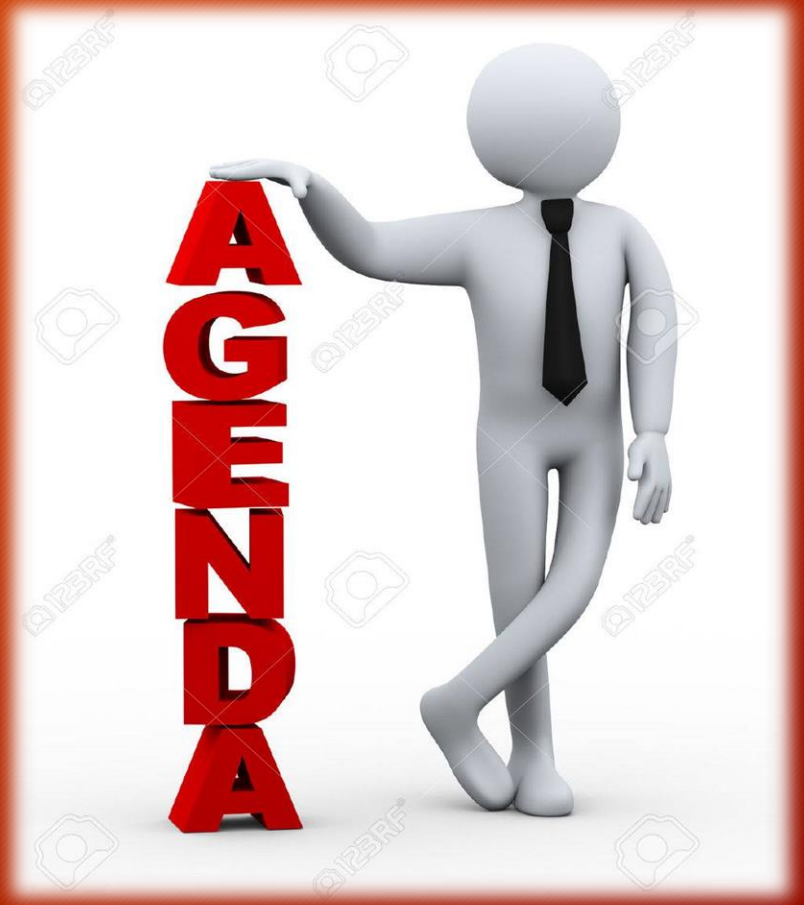
Marketing & Retail
Capstone project

Submitted By : Milan Mithal

Data Science Batch: C22

Marketing & Retail
Agenda

- Background
- Objectives
- Data visualisation
- Market Basket Analysis
- Recommendation
- Appendix
  - Data Source
  - Data Model Assumption
  - Data exploration and cleaning



Marketing & Retail
Background

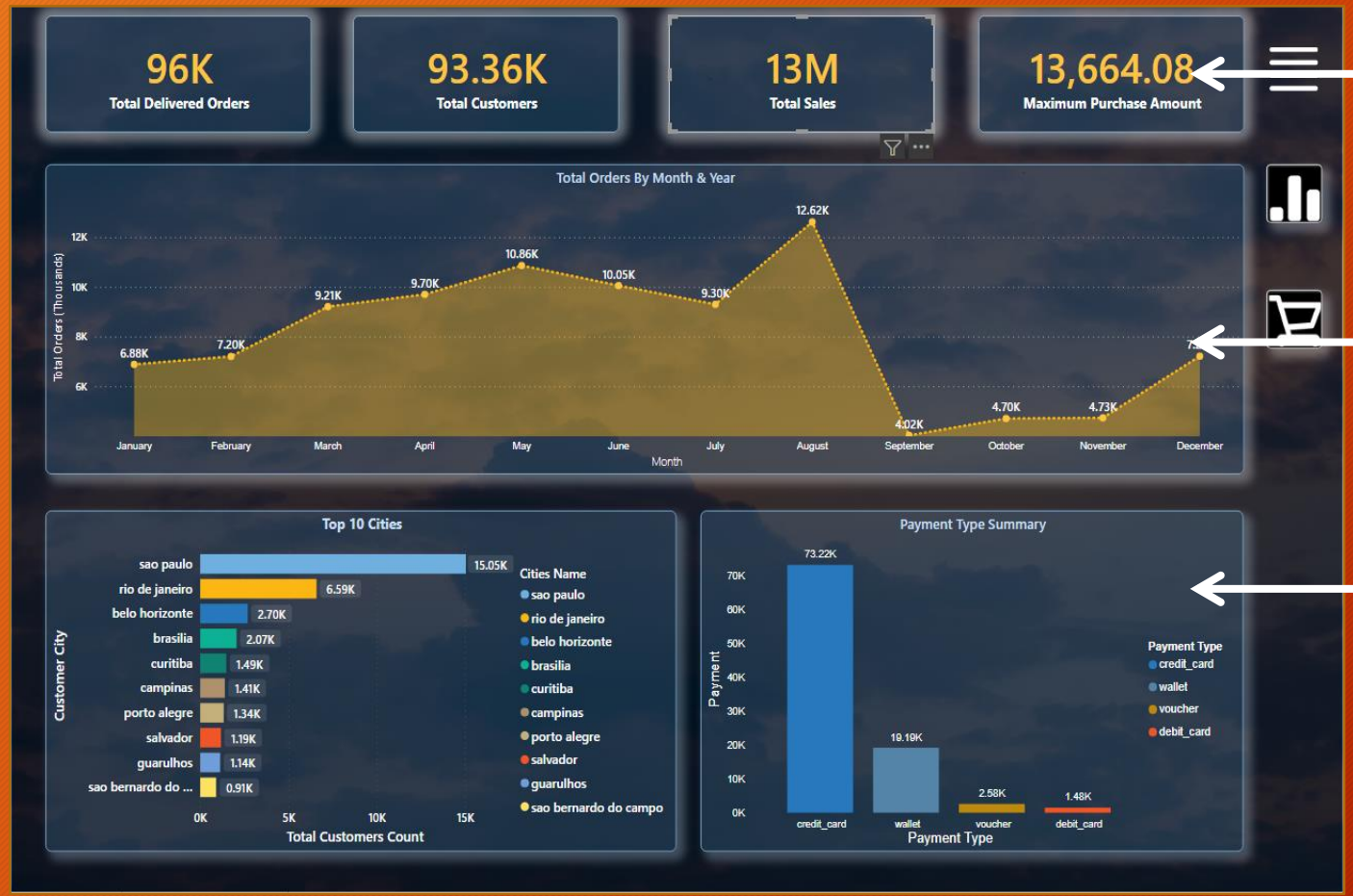
OList is an e-commerce company that sale multiple products like toys, auto accessories, furniture etc. Recently company has faced some losses and they want to manage their inventory, so as to reduce any unnecessary costs that they might be bearing.

Marketing & Retail
Objective

- We have to manage the inventory cost of this e-commerce company OList.
- We need to identify top products that contribute to the revenue and also use market basket analysis to analyse 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.

Marketing & Retail
Data Visualisation

Dashboard : Here we have represented summary of data.



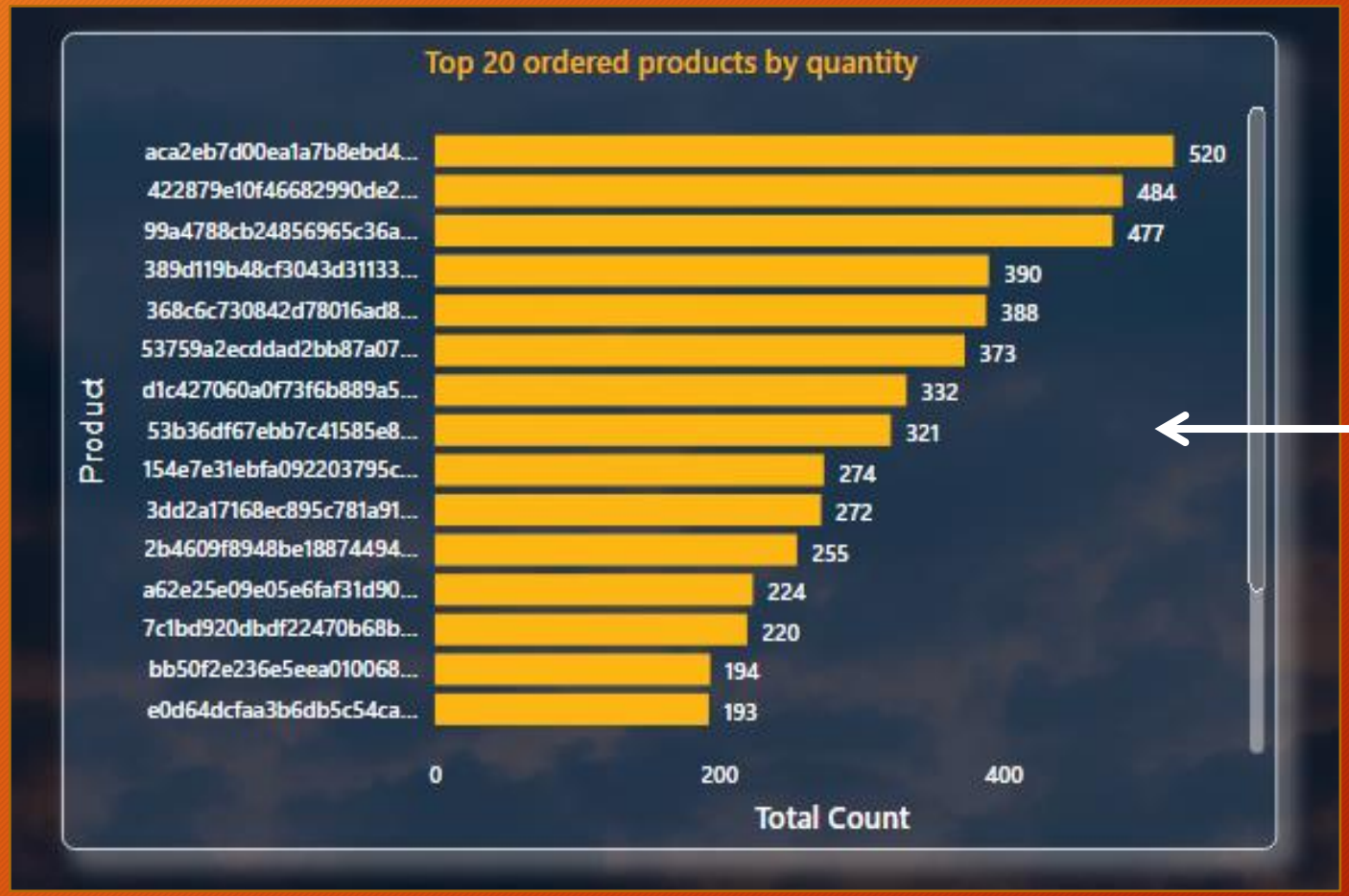
Summary of total orders, total customers, total sales & maximum Purchase amount.

Month & Year wise total orders

Top 10 most order Cities & Preferable payment mode by customer

Marketing & Retail
Data Visualisation

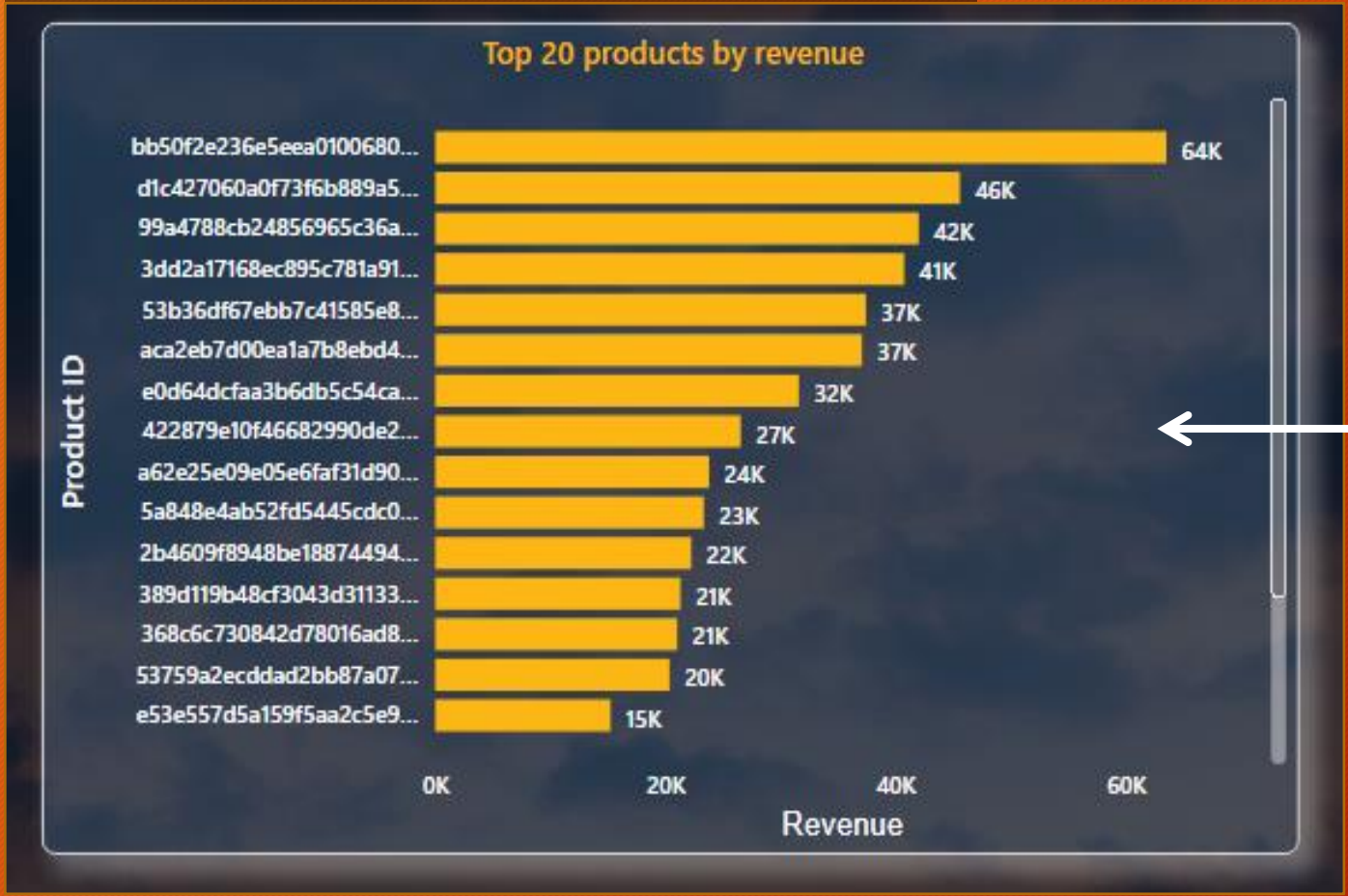
The top 20 ordered products by quantity are identified and visualized



Product id aca2eb7d00ea1a7b8ebd4e68314663af had the highest count of 520

Marketing & Retail
Data Visualisation

The top 20 products by revenue are identified and visualised.



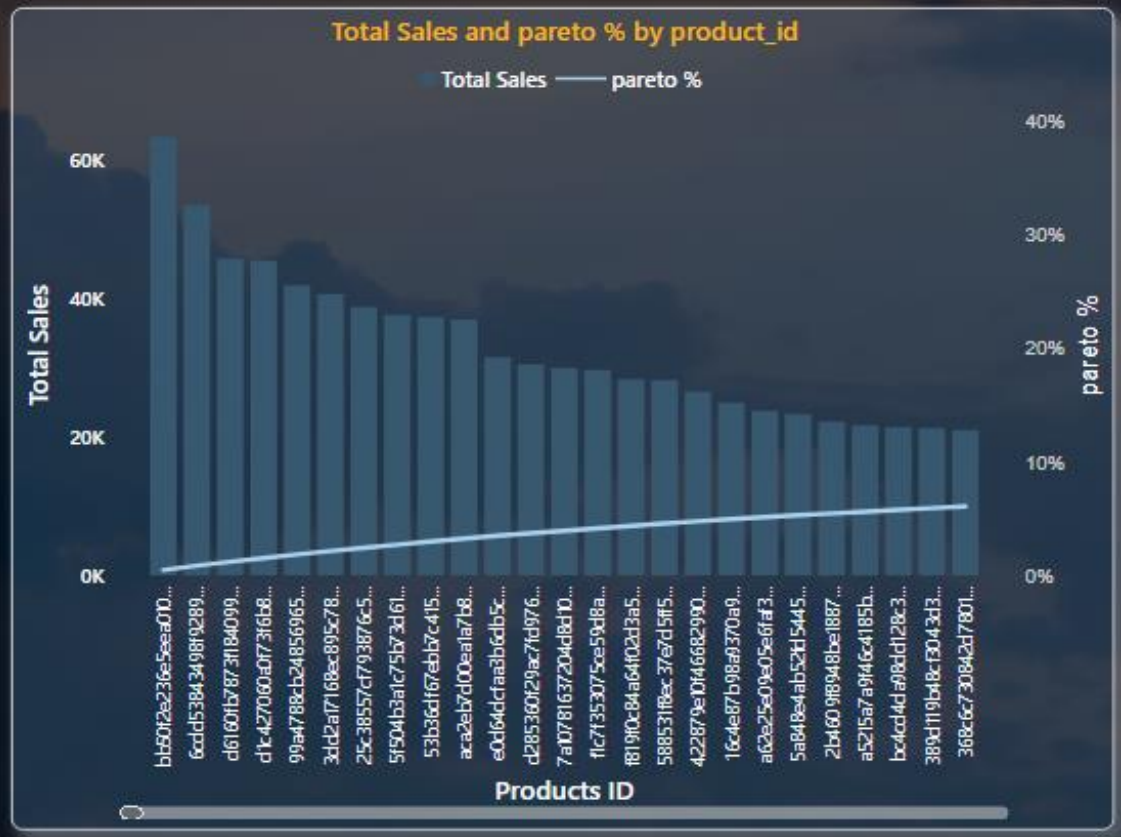
Product id bb50f2e236e5eea0100680137654686c had the highest revenue of approx. 64K

Marketing & Retail
Data Visualisation

The percentage running totals by revenue and number of orders.

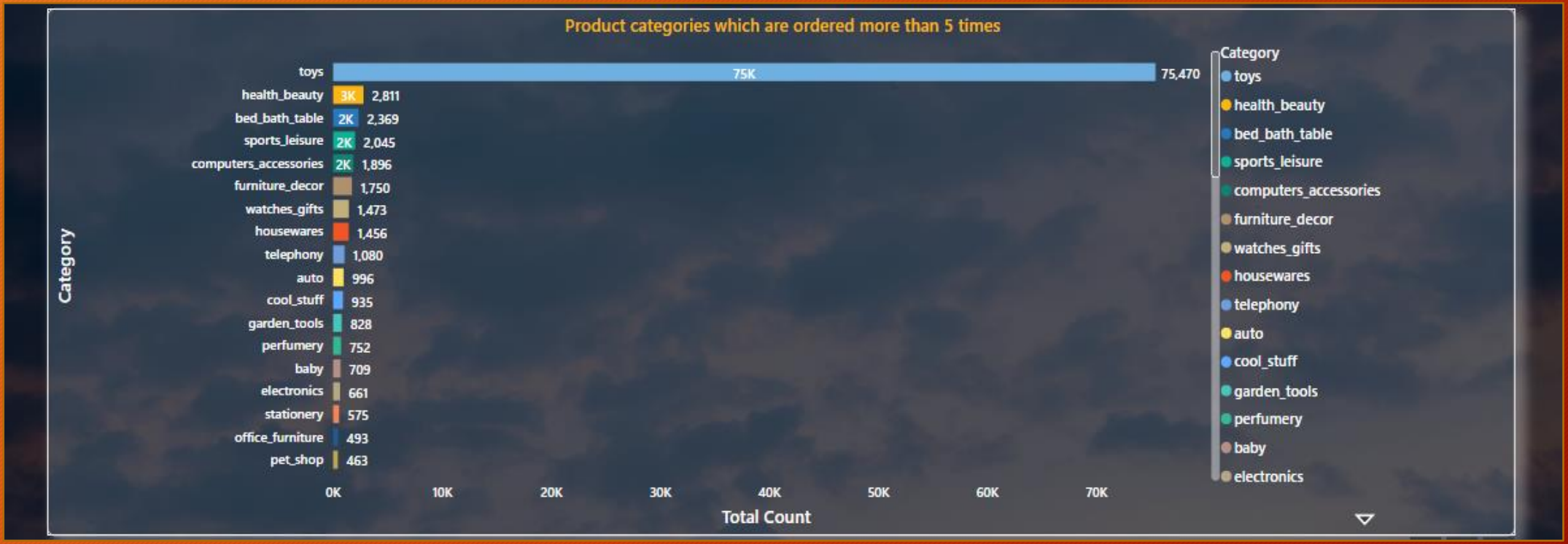
Percentage running totals by revenue and number of orders

product_id	Total Sales	Cummulative	Sales Rank	order_count	pareto %
bb50f2e236e5eea0100680137654686c	63560	63560.000	1	194	0.47%
6cdd53843498f92890544667809f1595	53652	117212.300	1	153	0.86%
d6160fb7873f184099d9bc95e30376af	45949	163161.650	1	33	1.20%
d1c427060a0f73f6b889a5c7c61f2ac4	45621	208782.210	1	332	1.54%
99a4788cb24856965c36a24e339b6058	42050	250831.870	1	477	1.85%
3dd2a17168ec895c781a9191c1e95ad7	40783	291614.670	1	272	2.15%
25c38557cf793876c5abdd5931f922db	38907	330521.990	1	38	2.43%
5f504b3a1c75b73d6151be81eb05bdc9	37734	368255.890	1	63	2.71%
53b36df67ebb7c41585e8d54d6772e08	37455	405710.520	1	321	2.98%
aca2eb7d00ea1a7b8ebd4e68314663af	37104	442814.820	1	520	3.26%
e0d64dcfaa3b6db5c54ca298ae101d05	31622	474436.640	1	193	3.49%
d285360f29ac7fd97640bf0baef03de0	30574	505010.550	1	119	3.72%
7a10781637204d8d10485c71a6108a2e	30045	535055.050	1	141	3.94%
f1c7f353075ce59d8a6f3cf58f419c9c	29792	564847.410	1	153	4.16%
f819f0c84a64f02d3a5606ca95edd272	28452	593298.990	1	44	4.37%
588531f8ec37e7d5ff5b7b22ea0488f8	28292	621590.980	1	20	4.57%
422879e10f46682990de24d770e7f83d	26577	648168.200	1	484	4.77%
16c4e87b98a9370a9cbc3a4658a3f45b	25034	673202.200	1	13	4.95%
a62e25e09e05e6faf31d90c6ec1aa3d1	23835	697037.200	1	224	5.13%
5a848e4ab52fd5445cdc07aab1c40e48	23368	720405.300	1	190	5.30%
2b4609f8948be188744942034964bc318	22277	742682.570	1	255	5.46%
a5215a7a9f46c4185b12f38e9ddf2abc	21740	764422.470	1	17	5.62%
bc4cd4da98dd128c39bf0b8c2674032f	21500	785922.060	1	17	5.78%
389d119b48cf3043d311335e499d9c6b	21337	807258.850	1	390	5.94%
Total	13221498	363996539941.442	1198935919	110197	2678090.66%



Marketing & Retail
Market Basket Analysis

The product categories which are ordered more than 5 times.

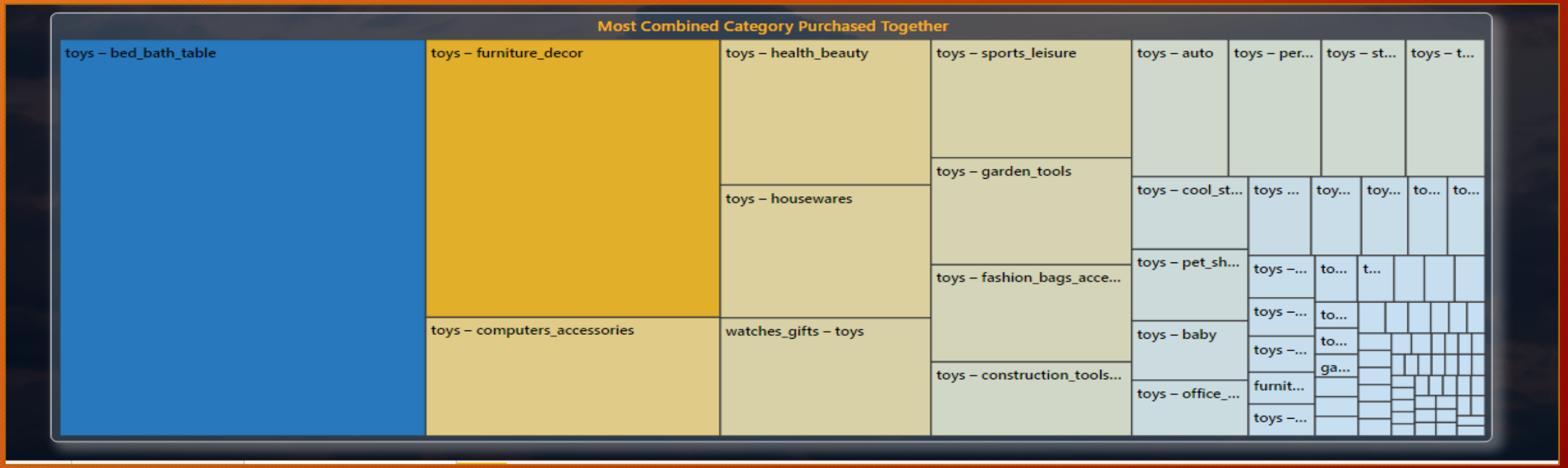


Toys, health_beauty, bed_bath_table,sport_leisure etc are among top categories which are ordered more than 5 times.

Marketing & Retail

Market Basket Analysis

Combinations of product categories which are frequently ordered together.



Toys-bed_bath_table, toys-furniture_decore, toys-computers_accessories etc are product categories which are frequently ordered together.

Marketing & Retail
Recommendation

- Product id **aca2eb7d00ea1a7b8ebd4e68314663af** is most favourite product among buyers. Company should promote this product.
- Product id **bb50f2e236e5eea01006801376546 86c** generate most revenue. Company should focus & promote this product.
- Customers mostly purchased products under **Toys** category. Company should promote this category more to generate revenue.
- Customers like to purchase **Toys-bed_bath_table, toys-furniture_decore** categories products together. Company should promote products under these categories.

Marketing & Retail
Appendix

- We have used Retail Dataset provided by OList.
- The important fields in the data **dictonary are order_id, product_id, price, product_category_name** & **order_status**.
- We have considered only delivered order for our data analysis.
- We have not considered shipping charges in revenue calculation since it is no profit no loss.

Marketing & Retail
Appendix

- We have used Python Language for exploration & cleaning of data. After cleaning data, we have exported all dataset into an excel workbook.
- For visualisation & Market basket analysis PowerBI tool has been used.

Orders Sheet

In [2]:
```
# Reading Orders Sheet

orders_data = pd.read_excel("Retail_dataset.xlsx", sheet_name="orders")
orders_data.head()
```

In [8]:
```
# As we have to considered only Delivered data, so we save Delivered data only.

orders_data = orders_data[orders_data.order_status == 'delivered']
orders_data.head()
```

Out[8]:

	order_id	customer_id	order_status	order_purchase_timestamp	order_approved_at	order_delivered_timestam
0	e481f51cbdc54678b7cc49136f2d6af7	7c396fd4830fd04220f754e42b4e5bff	delivered	2017-10-02 10:56:33	2017-10-02 11:07:15	2017-10-10 21:25:1
1	53cdb2fc8bc7dce0b6741e2150273451	af07308b275d755c9edb36a90c618231	delivered	2018-07-24 20:41:37	2018-07-26 03:24:27	2018-08-07 15:27:4
2	47770eb9100c2d0c44946d9cf07ec65d	3a653a41f6f9fc3d2a113cf8398680e8	delivered	2018-08-08 08:38:49	2018-08-08 08:55:23	2018-08-17 18:06:2
3	949d5b44dbf5de918fe9c16f97b45f8a	7c142cf63193a1473d2e66489a9ae977	delivered	2017-11-18 19:28:06	2017-11-18 19:45:59	2017-12-02 00:28:4
4	ad21c59c0840e6cb83a9ceb5573f8159	72632f0f9dd73dfee390c9b22eb56dd6	delivered	2018-02-13 21:18:39	2018-02-13 22:20:29	2018-02-16 18:17:0

In [9]:
```
# Checking total entries in new delivery dataset

orders_data.order_status.value_counts()
```

Out[9]:
```
delivered    96478
Name: order_status, dtype: int64
```

In [10]:
```
# Checking Null values in Delivery dataset

orders_data.isna().sum().sort_values(ascending=False)
```

Out[10]:
```
order_approved_at               14
order_delivered_timestamp        8
order_estimated_delivery_date    0
order_purchase_timestamp         0
order_status                     0
customer_id                      0
order_id                         0
dtype: int64
```

In [11]:
```
# Handling Null values
## Handing Approved_at Null value by Purchase timestamp
orders_data.order_approved_at.fillna(orders_data.order_purchase_timestamp, inplace=True)
## Handing Delivery_timestamp Null values by estimated_delivery_date
orders_data.order_delivered_timestamp.fillna(orders_data.order_estimated_delivery_date, inplace=True)
```

In [12]:
```
#Checking again for Null Value after Handling Null Values

orders_data.isna().sum().sort_values(ascending=False)
```

Out[12]:
```
order_estimated_delivery_date    0
order_delivered_timestamp        0
order_approved_at                0
order_purchase_timestamp         0
order_status                     0
customer_id                      0
order_id                         0
dtype: int64
```

All sheets are cleaned. Saving cleaned sheets in Retail_Dataset_Cleaned.xlsx

In [49]:
```
# Using Excel writer using XlsxWriter as the engine.
final_sheet = pd.ExcelWriter('Retail_Dataset_Cleaned.xlsx', engine='xlsxwriter')

# Write each cleaned dataframe to a different worksheet.
orders_data.to_excel(final_sheet, sheet_name='orders',index=False)
order_items_data.to_excel(final_sheet, sheet_name='order_items',index=False)
customers_data.to_excel(final_sheet, sheet_name='customers',index=False)
payments_data.to_excel(final_sheet, sheet_name='payments',index=False)
products_data.to_excel(final_sheet, sheet_name='products',index=False)

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

Marketing & Retail
Capstone project

Thank You