SALES ANALYSIS

LY NGUYEN THIEN BAO
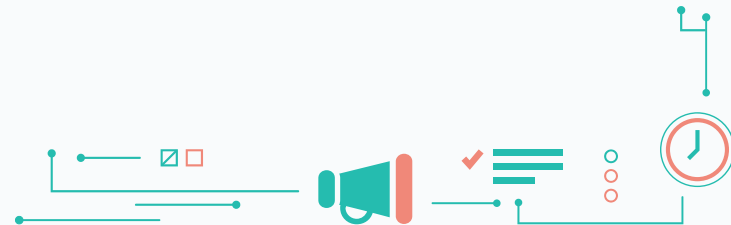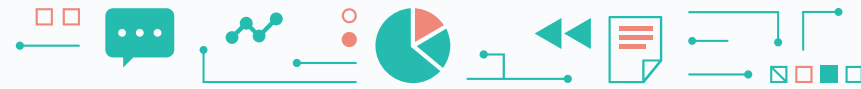
TABLE OF CONTENTS
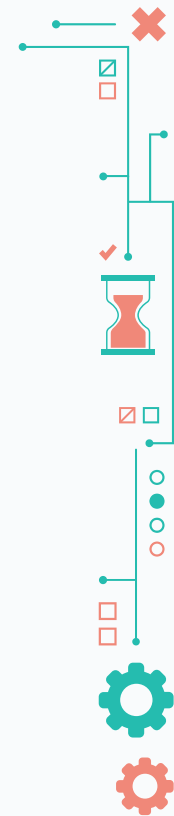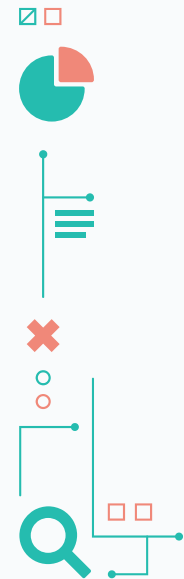
01.

DATSET INFORMATION AND ANALYSIS TOOL

GLOBAL SUPERSTORE DATASET

"Global Superstore is a virtual business operating in the Ecommerce sector, specializing in providing technology products, furniture, and office supplies to individuals and businesses worldwide."

However, in recent years, the company's CEO has assessed that the company's operations have not been efficient and show signs of stagnation. They are puzzled as to why the growth rate is low, profit margins are not high, and the rate of new customers is decreasing. Therefore, they have provided a dataset including the following information:

- *Orders information:* Order ID, order date, delivery date, quantity, profit...
- *Customer information:* Name, address, country, region,...
- *Product information:* Product name, price, product type,...

Dataset: Global Super Store Dataset
Data format: .csv
Source: https://www.kaggle.com/datasets/apoorvaappz/global-super-store-dataset

STORAGE, FORMATTING, CREATING RELEVANT TABLE

```sql
-- create database
CREATE DATABASE Store;

-- select database
use Store;

select * from superstore s ;

-- Date time format
        -- Add column with correct date format
        ALTER TABLE superstore
        ADD COLUMN new_Order_Date DATE,
        ADD COLUMN new_Shipping_Date DATE;

        -- Update data for new column
        UPDATE superstore
        SET new_Order_Date = STR_TO_DATE(`Order Date`, '%m/%d/%Y'),
            new_Shipping_Date = STR_TO_DATE(`Ship Date`, '%m/%d/%Y');

        -- Delete old Order Date and Ship Date
        ALTER TABLE superstore
        DROP COLUMN `Order Date`,
        DROP COLUMN `Ship Date`;

        -- Rename the new column
        ALTER TABLE superstore
        CHANGE COLUMN new_Order_Date `Order Date` DATE,
        CHANGE COLUMN new_Shipping_Date `Ship Date` DATE;
```

```sql
-- create orders_shipping inforamtion table
DROP TABLE IF EXISTS orders_shipping;

create table orders_shipping as
    (
    SELECT
        `Order ID` AS `Order ID`,
        `Customer ID`,
        `Order Date`,
        `Ship Date`,
        `Ship Mode`,
        `Order Priority`,
        CASE
            WHEN Market <> Region THEN CONCAT(Region, ' ', Market)
            ELSE Market
        END AS Market,
        Country,
        State,
        `City`,
        sum(`Shipping Cost`) as Total_Shipping_Cost
    FROM superstore
    GROUP BY `Order ID`
    );

select * from orders_shipping;
```

Storage and formatting

Use MySql to create database, format datetime

Create relevant table

Create table with Aggregate Function, Case When,… functions

DATA MODELLING
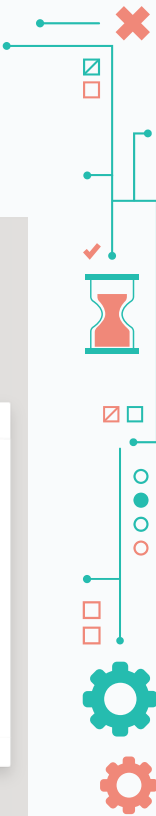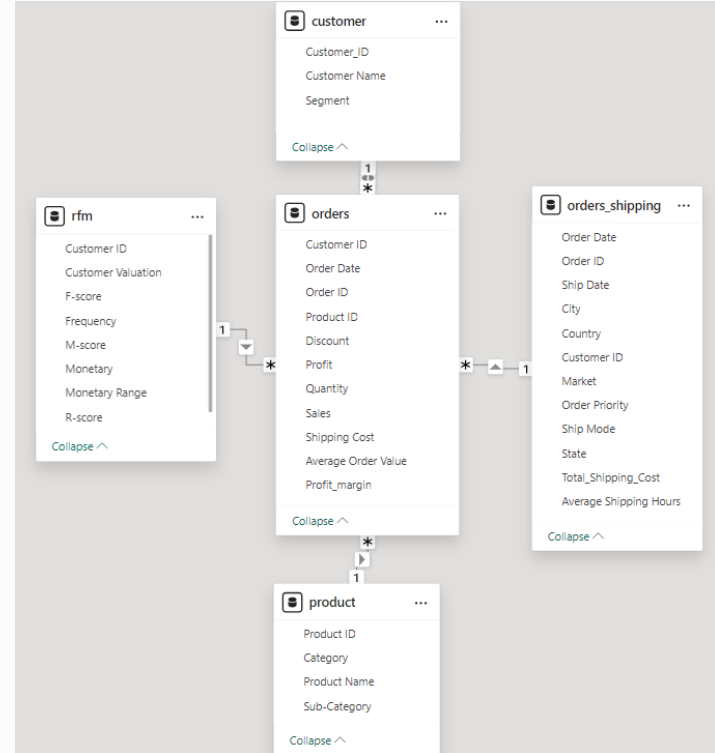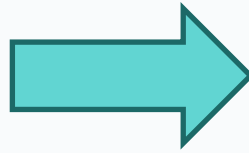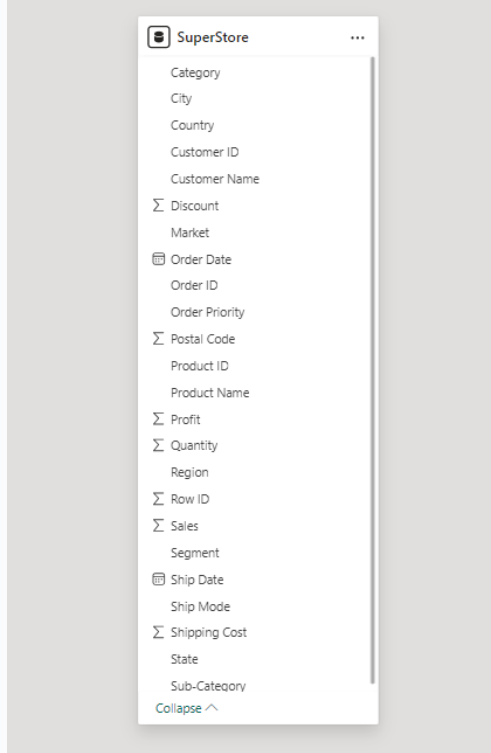
APRIORI DATA MINING

```python
# Applying Apriori
frq_items = apriori(basket, min_support=0.01, use_colnames=True)

# Show results in data frame
rules = association_rules(frq_items, metric="lift", min_threshold=1)
rules
```

|	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric
0	(Appliances)	(Binders)	0.090038	0.262727	0.025953	0.288248	1.097140	0.002298	1.035857	0.097300
1	(Binders)	(Appliances)	0.262727	0.090038	0.025953	0.098784	1.097140	0.002298	1.009705	0.120090
2	(Appliances)	(Furnishings)	0.090038	0.175085	0.015971	0.177384	1.013129	0.000207	1.002794	0.014241
3	(Furnishings)	(Appliances)	0.175085	0.090038	0.015971	0.091220	1.013129	0.000207	1.001301	0.015710
4	(Appliances)	(Paper)	0.090038	0.237772	0.021761	0.241685	1.016458	0.000352	1.005160	0.017793
5	(Paper)	(Appliances)	0.237772	0.090038	0.021761	0.091520	1.016458	0.000352	1.001631	0.021242
6	(Appliances)	(Phones)	0.090038	0.162507	0.015173	0.168514	1.036964	0.000541	1.007224	0.039173
7	(Phones)	(Appliances)	0.162507	0.090038	0.015173	0.093366	1.036964	0.000541	1.003671	0.042563
8	(Art)	(Phones)	0.145937	0.162507	0.024755	0.169631	1.043833	0.001040	1.008578	0.049168
9	(Phones)	(Art)	0.162507	0.145937	0.024755	0.152334	1.043833	0.001040	1.007546	0.050140
10	(Fasteners)	(Binders)	0.042923	0.262727	0.012378	0.288372	1.097611	0.001101	1.036037	0.092919
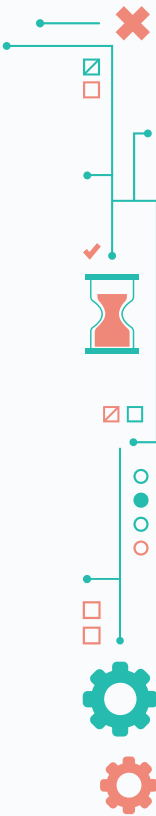
DAX FUNCTION

```
Recency =
VAR
MaxDate = CALCULATE(
        MAX(Global_Superstore_date[Order Date]),
            ALLEXCEPT(Global_Superstore,
            Global_Superstore[Customer ID]
            )
            )
RETURN
DATEDIFF(MaxDate, TODAY(), DAY)
```

```
1 RFM =
2 SUMMARIZE(
3     Global_Superstore,
4     Global_Superstore[Customer ID],
5     "Recency", MAX(Global_Superstore[Recency]),
6     "Frequency", MAX(Global_Superstore[Frequency]),
7     "Monetary", SUM(Global_Superstore[Sales])
8 )
```

```
Customer Valuation =
VAR RFM_Score = 'RFM'[R score] * 100 + 'RFM'[F score] * 10 + 'RFM'[M score]
RETURN
SWITCH(
    TRUE(),
    RFM_Score IN {555, 554, 544, 545, 454, 455, 445}, "Champion",
    RFM_Score IN {543, 444, 435, 355, 354, 345, 344, 335}, "Loyal Customer",
    RFM_Score IN {553, 551, 552, 541, 542, 533, 532, 531, 452, 451, 442, 441,431, 453, 433,
    432,423, 353, 352, 351, 342, 341, 333, 323}, "Potential Loyalist",
    RFM_Score IN {512, 511, 422, 421, 412, 411, 311}, "Recent Customer",
    RFM_Score IN {525, 524, 523, 522, 521, 515, 514, 513, 425, 424, 413, 414, 415, 315, 314,
    313}, "Promising",
    RFM_Score IN {535, 534, 443, 434, 343, 334, 325, 324}, "Need Attention",
    RFM_Score IN {331, 321, 312, 221, 213}, "About to Risk",
    RFM_Score IN {255, 254, 245, 244, 253, 252, 243, 242, 235, 234, 225, 224, 153, 152, 145,
    143, 142, 135, 134, 133, 125, 124}, "At Risk",
    RFM_Score IN {155, 154, 144, 214, 215, 115, 114, 113}, "Cannot Lose Them",
    RFM_Score IN {332, 322, 231, 241, 251, 233, 232, 223, 222, 132, 123, 122, 212, 211},
    "Hibernate",
    "Lost"
)
```

CREATE MEASURE

Calculate "Recency" by finding the last transaction date, using ALLEXCEPT to retain Customer ID, and DATEDIFF for the date calculation

CREATE TABLE RFM

Create RFM table by calculating Recency, Frequency, and the sum of Sales grouped by Customer.

CREATE COLUMN

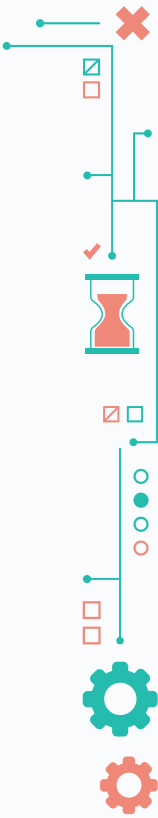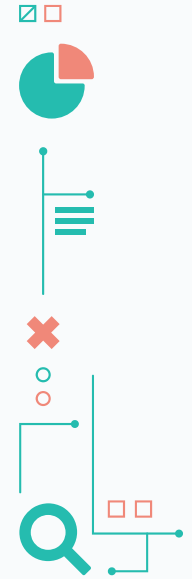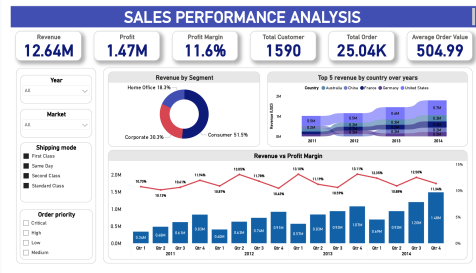Calculate the RFM Score using variables and identify customer segments based on provided conditions using the SWITCH function.

02

DATA VISUALIZATION

DASHBOARD VISUALIZATION

1. OVERVIEW

2. PRODUCT
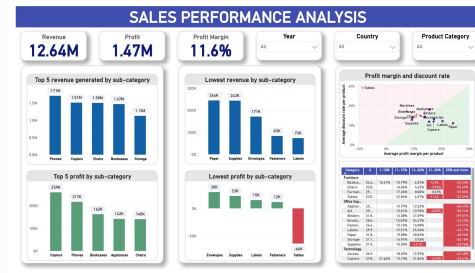
3. CUSTOMERS
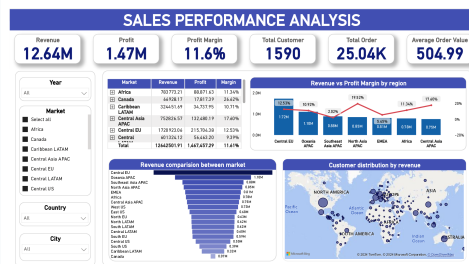
4. LOCATION
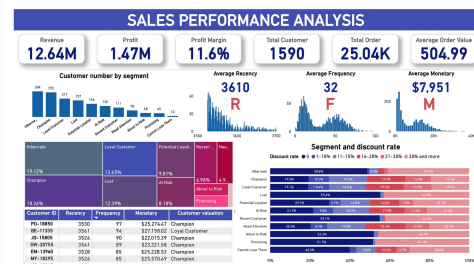
5. SHIPPING

SALES PERFORMANCE ANALYSIS

Revenue	Profit	Profit Margin	Total Customer	Total Order	Average Order Value
12.64M	**1.47M**	**11.6%**	**1590**	**25.04K**	**504.99**

Year
All

Market
All

Shipping mode
- First Class
- Same Day
- Second Class
- Standard Class

Order priority
- [ ] Critical
- [ ] High
- [ ] Low
- [ ] Medium

Revenue by Segment

- Home Office 18.3%
- Corporate 30.3%
- Consumer 51.5%
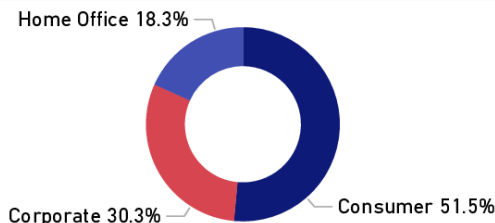
Top 5 revenue by country over years

Country: Australia, China, France, Germany, United States

Revenue (USD)

	2011	2012	2013	2014
	0.5M	0.5M	0.6M	0.7M
	0.2M	0.2M	0.3M	0.3M
	0.2M	0.2M	0.3M	0.3M
	0.2M	0.2M	0.2M	0.2M
		0.2M	0.2M	0.2M

Revenue vs Profit Margin

	Qtr 1	Qtr 2	Qtr 3	Qtr 4	Qtr 1	Qtr 2	Qtr 3	Qtr 4	Qtr 1	Qtr 2	Qtr 3	Qtr 4	Qtr 1	Qtr 2	Qtr 3	Qtr 4
Revenue	0.34M	0.48M	0.61M	0.83M	0.40M	0.63M	0.74M	0.91M	0.57M	0.83M	0.93M	1.07M	0.69M	0.93M	1.20M	1.48M
Profit Margin	10.73%	10.13%	10.61%	11.94%	10.87%	13.05%	11.78%	10.43%	13.10%	11.19%	10.59%	13.11%	12.35%	10.88%	12.50%	11.34%
Year	2011				2012				2013				2014			

SALES PERFORMANCE ANALYSIS

Revenue
12.64M

Profit
1.47M

Profit Margin
11.6%

Year
All

Country
All

Product Category
All

Top 5 revenue generated by sub-category

1.71M	1.51M	1.50M	1.47M	1.13M
Phones	Copiers	Chairs	Bookcases	Storage

Top 5 profit by sub-category

259K	217K	162K	142K	140K
Copiers	Phones	Bookcases	Appliances	Chairs

Lowest revenue by sub-category

244K	243K	171K	83K	73K
Paper	Supplies	Envelopes	Fasteners	Labels

Lowest profit by sub-category

30K	23K	15K	12K	-64K
Envelopes	Supplies	Labels	Fasteners	Tables
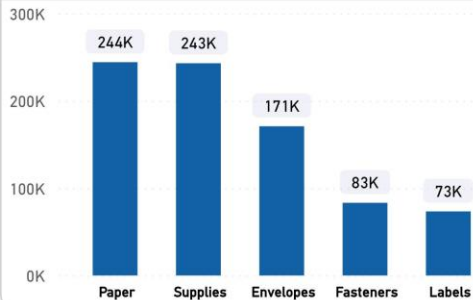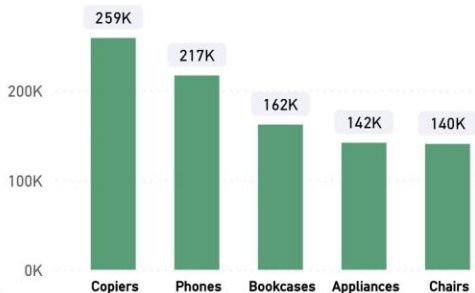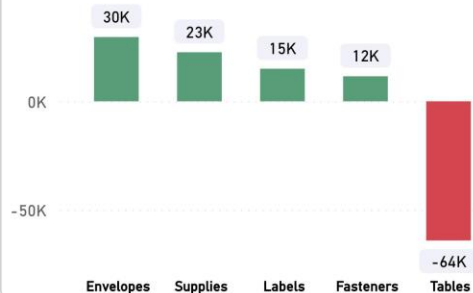
Profit margin and discount rate

Average discount rate per product (y-axis): 0% – 30%
Average profit margin per product (x-axis): -10% – 30%

Tables, Machines, Bookcases, Appliances, Binders, Chairs, Accessories, Storage, Supplies, Art, Labels, Paper, Copiers
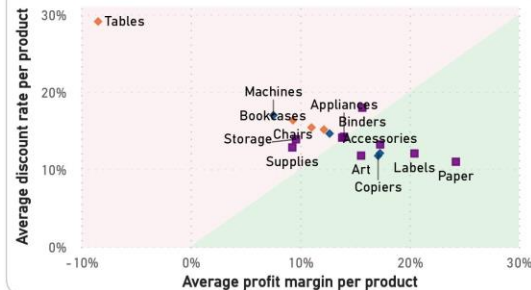
Category	0	1–10%	11–15%	16–20%	21–30%	30% and more
Furniture						
Bookca...	24.4...	16.41%	15.79%	4.53%	-12.98...	-45.20%
Chairs	23.8...		16.06%	4.22%	-3.98%	-55.69%
Furnish...	25...		17.26%	8.00%	0.03%	-55.80%
Tables	23.0...		22.84%	4.57%	-11.42...	-52.59%
Office Sup...						
Applian...	25....		15.77%	12.22%		-60.62%
Art	25....		15.61%	10.70%	-6.83%	-58.74%
Binders	31.8...		16.38%	31.09%		-89.51%
Envelo...	28.6...		15.05%	24.37%		-42.23%
Fasten...	26.4...		15.16%	16.08%		-47.31%
Labels	29.9...		19.51%	25.34%		-42.17%
Paper	31.8...		19.08%	30.83%		-48.96%
Storage	21.1...		14.91%	0.76%		-50.18%
Supplies	21.0...		15.30%	-12.19...		-49.76%
Technology						
Access...	26.9...		18.20%	13.59%		-45.68%
Copiers	27.8...	21.44%	13.10%	31.84%	-3.67%	-37.67%

SALES PERFORMANCE ANALYSIS

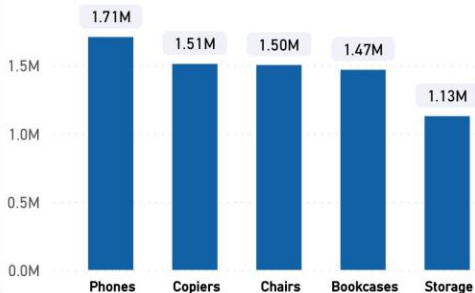| Revenue	Profit	Profit Margin	Total Customer	Total Order	Average Order Value
12.64M	**1.47M**	**11.6%**	**1590**	**25.04K**	**504.99**

Customer number by segment

Hibernate: 304, Champion: 292, Loyal Customer: 217, Lost: 197, Potential Loyalist: 156, At Risk: 130, Recent Customer: 111, Need Attention: 78, About to Risk: 48, Promising: 45, Cannot Lose Them: 12
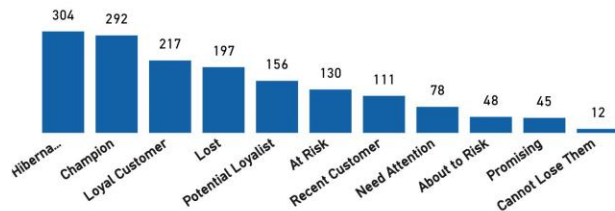
Average Recency
3610
R
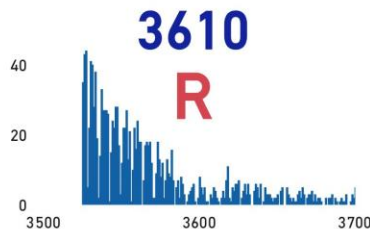
Average Frequency
32
F
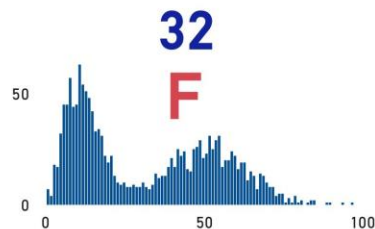
Average Monetary
$7,951
M

Hibernate		Loyal Customer		Potential Loyali...	Recent ...	Nee...
19.12%		13.65%		9.81%	6.98%	4.9...
Champion		Lost		At Risk	About to Risk	
18.36%		12.39%		8.18%	Promising	

Customer ID	Recency	Frequency ▼	Monetary	Customer valuation
PO-18850	3530	97	$25,274.47	Champion
BE-11335	3561	94	$27,158.02	Loyal Customer
JG-15805	3526	90	$22,015.39	Champion
SW-20755	3541	89	$23,321.58	Champion
EM-13960	3528	85	$25,228.53	Champion
MY-18295	3526	85	$25,570.49	Champion

Segment and discount rate

Discount rate: ● 0 ● 1-10% ● 11-15% ● 16-20% ● 21-30% ● 30% and more
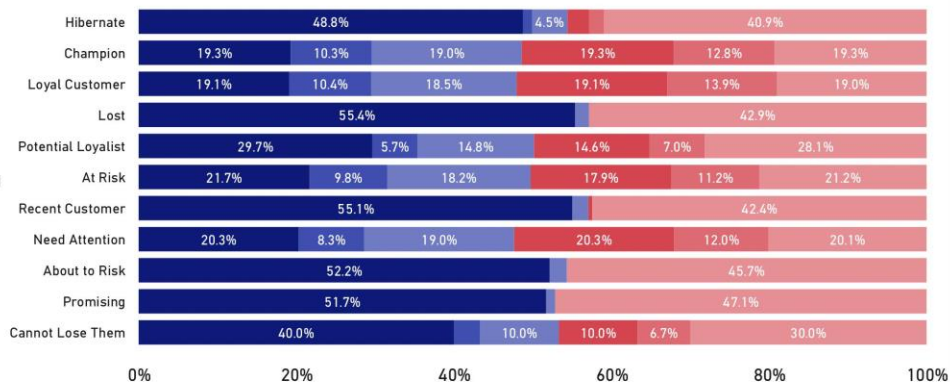
Segment						
Hibernate	48.8%		4.5%		40.9%	
Champion	19.3%	10.3%	19.0%	19.3%	12.8%	19.3%
Loyal Customer	19.1%	10.4%	18.5%	19.1%	13.9%	19.0%
Lost	55.4%			42.9%		
Potential Loyalist	29.7%	5.7%	14.8%	14.6%	7.0%	28.1%
At Risk	21.7%	9.8%	18.2%	17.9%	11.2%	21.2%
Recent Customer	55.1%			42.4%		
Need Attention	20.3%	8.3%	19.0%	20.3%	12.0%	20.1%
About to Risk	52.2%			45.7%		
Promising	51.7%			47.1%		
Cannot Lose Them	40.0%		10.0%	10.0%	6.7%	30.0%

0% 20% 40% 60% 80% 100%

SALES PERFORMANCE ANALYSIS

Revenue	Profit	Profit Margin	Total Customer	Total Order	Average Order Value
12.64M	1.47M	11.6%	1590	25.04K	504.99

Year

All

Market

- Select all
- Africa
- Canada
- Caribbean LATAM
- Central Asia APAC
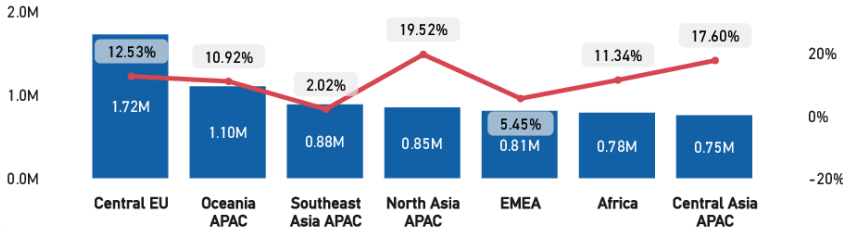- Central EU
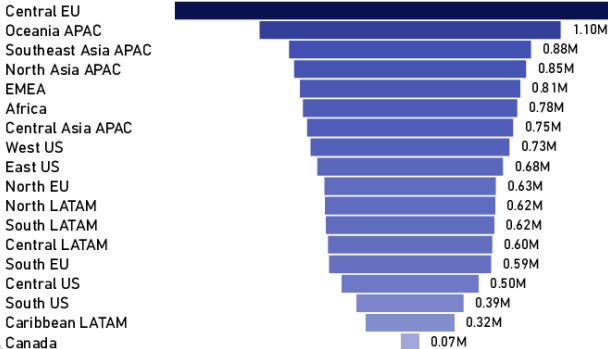- Central LATAM
- Central US

Country

All

City

All

Market	Revenue	Profit	Margin
Africa	783 773.21	88,871.63	11.34%
Canada	66 928.17	17,817.39	26.62%
Caribbean LATAM	324 451.69	34,737.95	10.71%
Central Asia APAC	752 826.57	132,480.19	17.60%
Central EU	1 720 923.06	215,704.38	12.53%
Central LATAM	601 324.12	56,463.20	9.39%
Total	12 642 501.91	1,467,457.29	11.61%

Revenue vs Profit Margin by region

Region	Revenue	Margin
Central EU	1.72M	12.53%
Oceania APAC	1.10M	10.92%
Southeast Asia APAC	0.88M	2.02%
North Asia APAC	0.85M	19.52%
EMEA	0.81M	5.45%
Africa	0.78M	11.34%
Central Asia APAC	0.75M	17.60%

Revenue comparision between market

Market	Revenue
Central EU	
Oceania APAC	1.10M
Southeast Asia APAC	0.88M
North Asia APAC	0.85M
EMEA	0.81M
Africa	0.78M
Central Asia APAC	0.75M
West US	0.73M
East US	0.68M
North EU	0.63M
North LATAM	0.62M
South LATAM	0.62M
Central LATAM	0.60M
South EU	0.59M
Central US	0.50M
South US	0.39M
Caribbean LATAM	0.32M
Canada	0.07M

Customer distribution by revenue
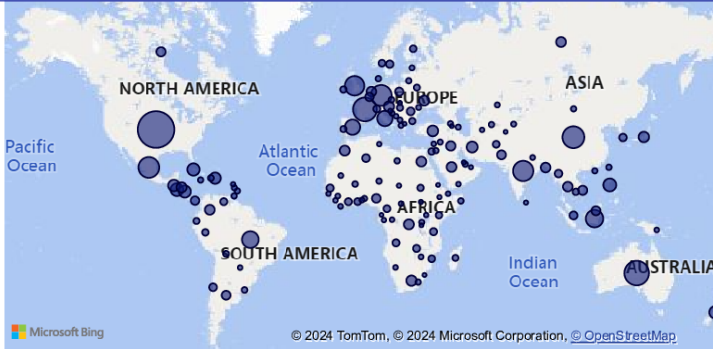
Microsoft Bing | © 2024 TomTom, © 2024 Microsoft Corporation, © OpenStreetMap

SALES PERFORMANCE ANALYSIS

Revenue	Profit	Profit Margin	Total Customer	Total Order	Average Order Value
12.64M	**1.47M**	**11.6%**	**1590**	**25.04K**	**504.99**

Year
All

Market
- ☐ Select all
- ☐ Africa
- ☐ Canada
- ☐ Caribbean LATAM
- ☐ Central Asia APAC
- ☐ Central EU
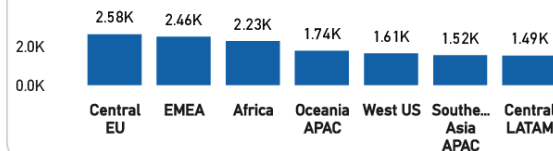- ☐ Central LATAM
- ☐ Central US

Country
All

City
All

1,352,820.69
Total Shipping Cost

95.07
Average Shipping Hours

Ship Mode	Critical	High	Low	Medium	Total
First Class	881	1716		1151	3748
Same Day	349	627		343	1319
Second Class	711	1954		2350	5015
Standard Class		3341	1175	10437	14953
Total	1941	7638	1175	14281	25035

Shipment by region

Central EU	EMEA	Africa	Oceania APAC	West US	Southe... Asia APAC	Central LATAM
2.58K	2.46K	2.23K	1.74K	1.61K	1.52K	1.49K

Average Shipping Hour by Country

Indonesia	United Kingdom	Germany	Turkey	India	United States	Mexico	Brazil	Italy	Australia	China	France
96.6	96.5	96.0	95.3	95.2	95.0	95.0	95.0	94.7	94.5	93.9	93.9

Total Shipment

January	February	March	April	May	June	July	August	September	October	November	December
1440	1090	1471	1478	1767	2534	1647	2319	2891	2293	3012	3093

Delivery by ship mode

- Second Class 20.0%
- First Class 15.0%
- Same Day 5.3%
- Standard Class 59.7%
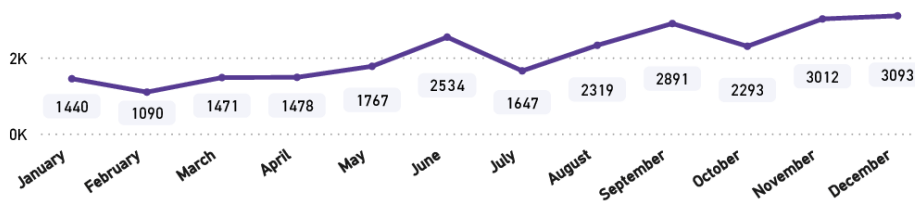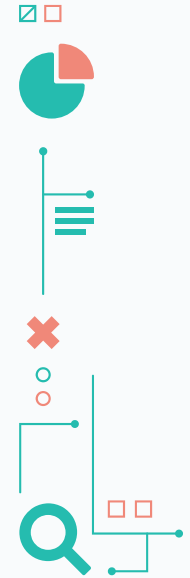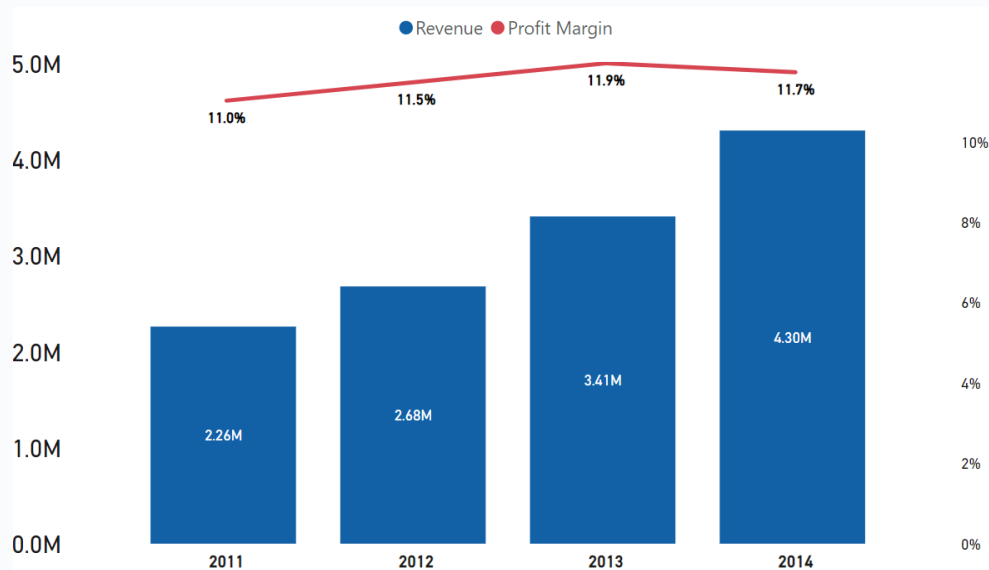
03

PROBLEMS IDENTIFICATION
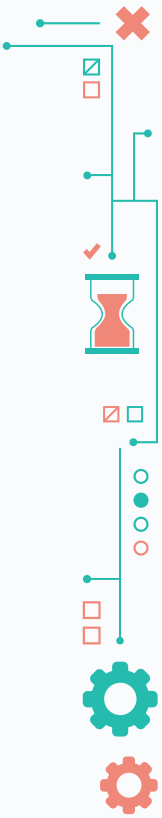
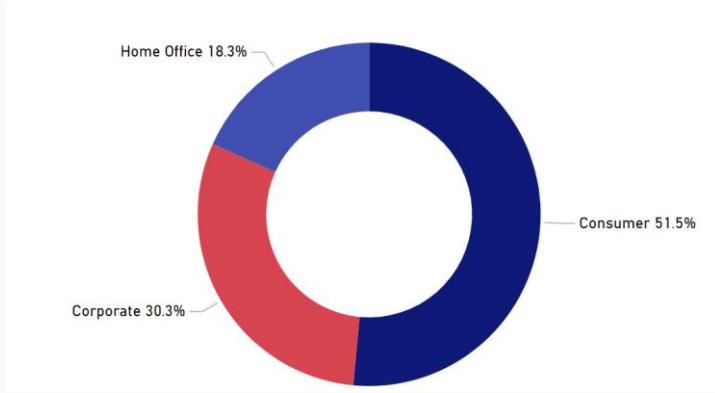# Revenue and Profit Increase Year Over Year

However, the company's
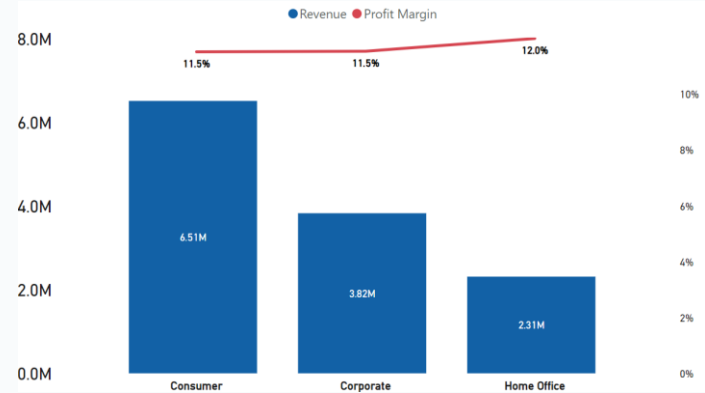**profit margin slightly declined in 2014**

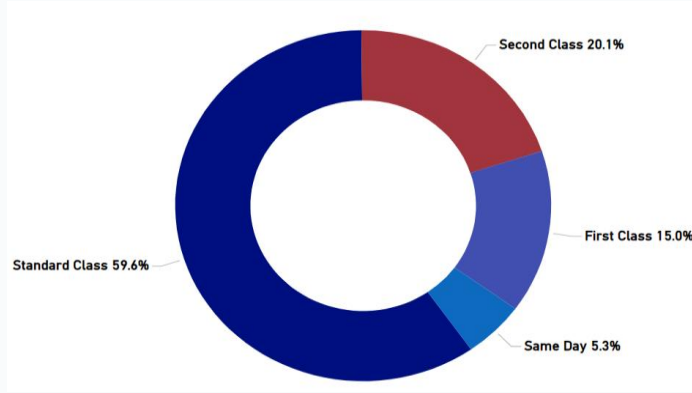So **what** caused the **profit margin to decline**?

CUSTOMER SEGMENT

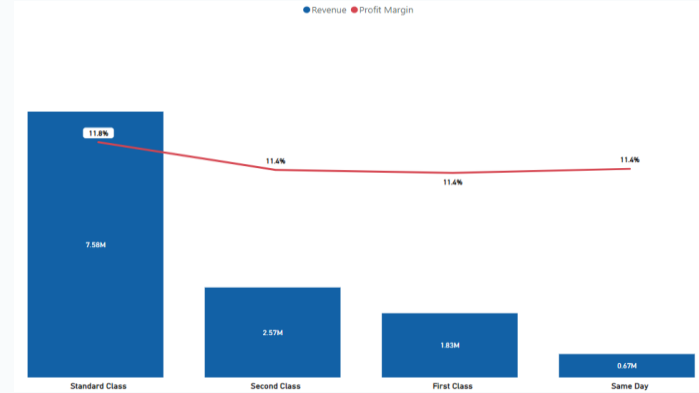More than 50% of the revenue comes from the 'Consumer' segment

Profit margin remains consistent across all segments

CUSTOMER SEGMENT DOESN'T DIRECTLY AFFECT PROFIT

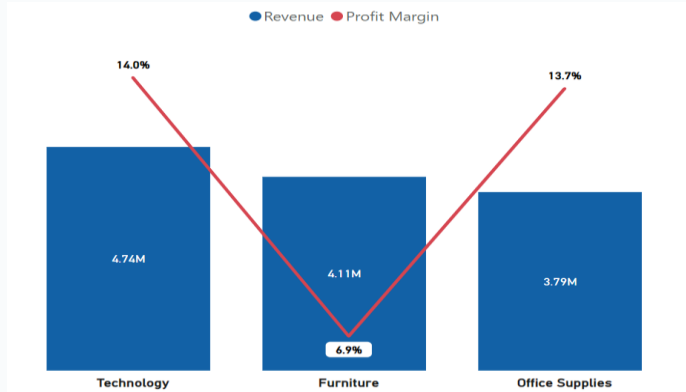# SIMILARLY, SHIP MODE DOES NOT IMPACT PROFIT

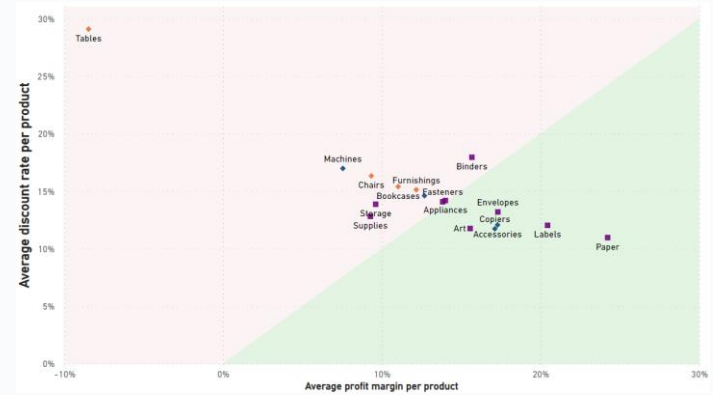More than 50% of the orders used Standard Class

However, since shipping costs are not included in the overall cost, they do not directly impact the profit

SHIP MODE DOES NOT DIRECTLY AFFECT PROFIT

Some product sub-categories have very low or negative profit

Legend: ● Revenue ● Profit Margin

14.0%   13.7%

6.9%

4.74M   4.11M   3.79M

Technology   Furniture   Office Supplies

The profit from the "**furniture**" is significantly lower compared to others

Average discount rate per product (y-axis: 0% to 30%)
Average profit margin per product (x-axis: -10% to 30%)

Tables
Machines
Chairs   Furnishings   Binders
Bookcases   Fasteners
Storage   Appliances   Envelopes
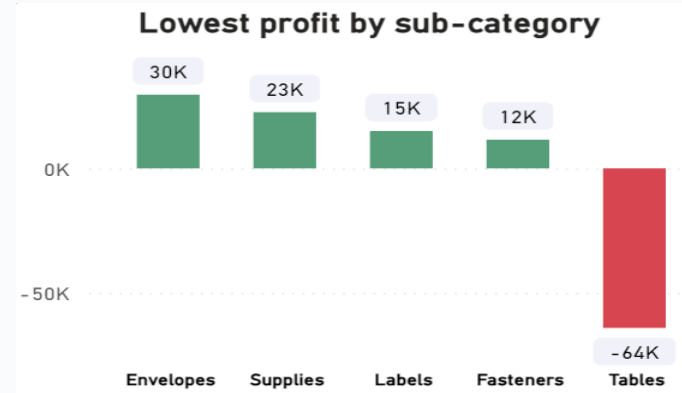Supplies   Copiers
Art   Accessories   Labels   Paper

The discount is very high compared to the profit margins, especially those in the **furniture** category

DISCOUNTS IMPACT THE PROFIT OF MANY SUB-CATEGORIES

Especially, "Discount rate > 20%" generated negative profit margin

Category	0	1-10%	11-15%	16-20%	21-30%	30% and more
Furniture						
Bookca...	24.42%	16.41%	15.79%	4.53%	-12.98%	-45.20%
Chairs	23.88%		16.06%	4.22%	-3.98%	-55.69%
Furnish...	25.39%		17.26%	8.00%	0.03%	-55.80%
Tables	23.09%		22.84%	4.57%	-11.42%	-52.59%
Office Sup...						
Applian...	25.92%		15.77%	12.22%		-60.62%
Art	25.01%		15.61%	10.70%	-6.83%	-58.74%
Binders	31.81%		16.38%	31.09%		-89.51%
Envelo...	28.64%		15.05%	24.37%		-42.23%
Fasten...	26.43%		15.16%	16.08%		-47.31%
Labels	29.94%		19.51%	25.34%		-42.17%
Paper	31.89%		19.08%	30.83%		-48.96%
Storage	21.15%		14.91%	0.76%		-50.18%
Supplies	21.09%		15.30%	-12.19%		-49.76%
Technology						
Access...	26.92%		18.20%	13.59%		-45.68%
Copiers	27.81%	21.44%	13.10%	31.84%	-3.67%	-37.67%
Machin...	25.72%		11.14%	12.90%	-3.51%	-54.34%
Phones	24.01%		16.09%	9.63%	0.48%	-49.01%

The discount rate more than 20% generate negative profit margin

Lowest profit by sub-category

	Envelopes	Supplies	Labels	Fasteners	Tables
	30K	23K	15K	12K	-64K

The 'Tables' sub-category showed the lowest and negative profit

DISCOUNTS IMPACT THE PROFIT OF MANY SUB-CATEGORIES

Moreover, the increase of new customers is not high

Total customer by year

Year	Total
2011	1309
2012	1373
2013	1458
2014	1511
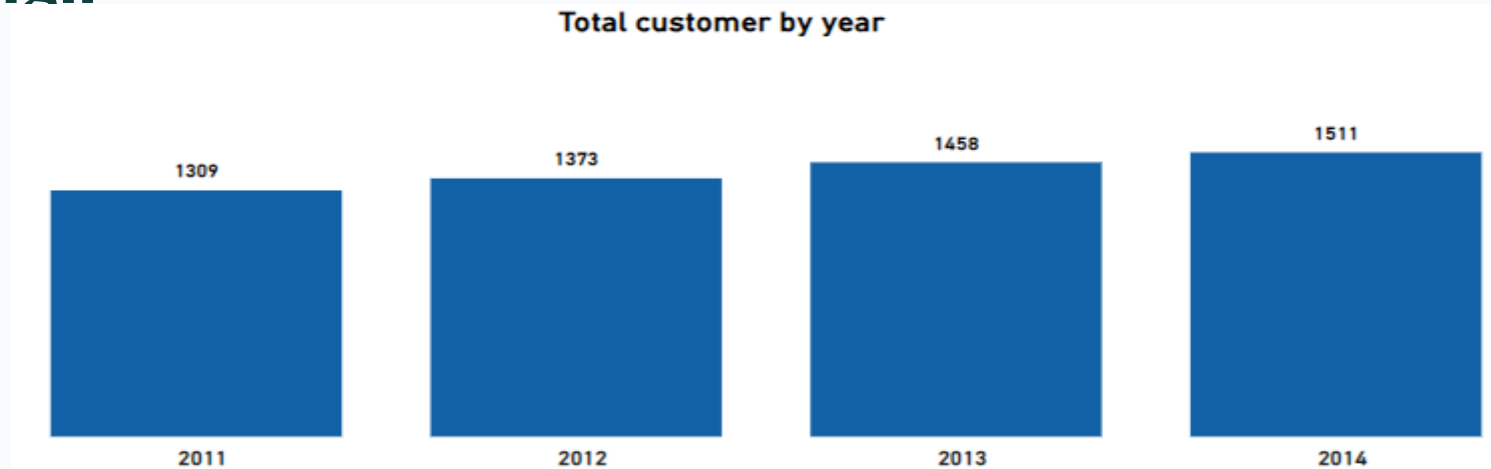
Therefore, the company also needs to **find solutions to increase new customers**

04

RECOMMENDATIONS
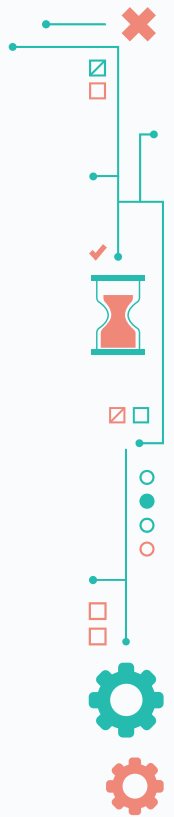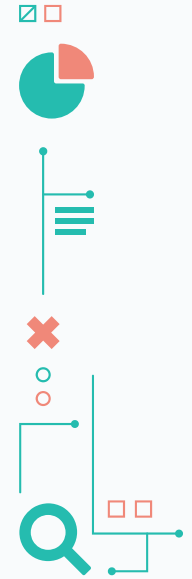
STOP DISCOUNT > 20% for ALL PRODUCTS

Category	0	1-10%	11-15%	16-20%	21-30%	30% and more
Furniture						
Bookca...	24.42%	16.41%	15.79%	4.53%	-12.98%	-45.20%
Chairs	23.88%		16.06%	4.22%	-3.98%	-55.69%
Furnish...	25.39%		17.26%	8.00%	0.03%	-55.80%
Tables	23.09%		22.84%	4.57%	-11.42%	-52.59%
Office Sup...						
Applian...	25.92%		15.77%	12.22%		-60.62%
Art	25.01%		15.61%	10.70%	-6.83%	-58.74%
Binders	31.81%		16.38%	31.09%		-89.51%
Envelo...	28.64%		15.05%	24.37%		-42.23%
Fasten...	26.43%		15.16%	16.08%		-47.31%
Labels	29.94%		19.51%	25.34%		-42.17%
Paper	31.89%		19.08%	30.83%		-48.96%
Storage	21.15%		14.91%	0.76%		-50.18%
Supplies	21.09%		15.30%	-12.19%		-49.76%
Technology						
Access...	26.92%		18.20%	13.59%		-45.68%
Copiers	27.81%	21.44%	13.10%	31.84%	-3.67%	-37.67%
Machin...	25.72%		11.14%	12.90%	-3.51%	-54.34%
Phones	24.01%		16.09%	9.63%	0.48%	-49.01%

Discounting at a rate more than 20% led to low or negative profit margins, in spite of boosting sales.

Therefore, we should stop offering rate of discounts more than 20% on all products.

Discount products by basket analysis

Basket
- (Appliances), (Binders)
- (Appliances), (Furnishings)
- (Appliances), (Paper)
- (Appliances), (Phones)
- (Art), (Phones)
- (Binders, Phones), (Paper)
- (Binders, Storage), (Paper)
- (Fasteners), (Binders)
- (Fasteners), (Paper)
- (Furnishings), (Phones)
- (Furnishings), (Storage)
- (Furnishings, Binders), (Paper)
- (Furnishings, Paper), (Binders)
- (Labels), (Phones)
- (Labels), (Storage)
- (Phones, Paper), (Binders)
- (Storage, Paper), (Binders)
- (Tables), (Phones)
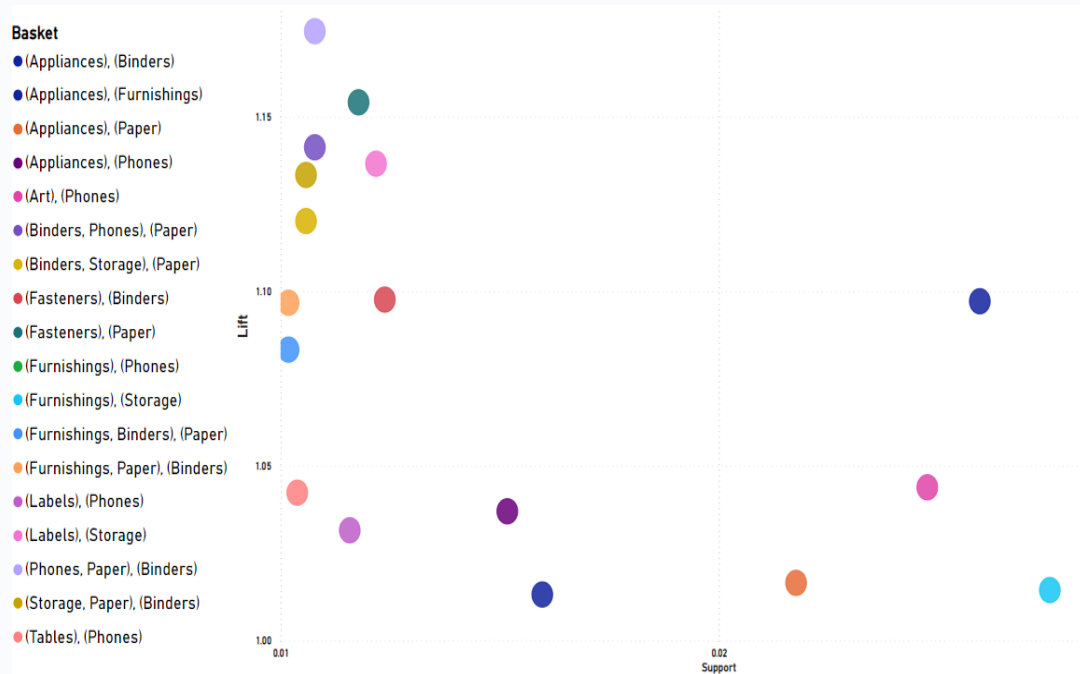
Using Apriori to find product that are frequently purchase toghether. The results showed 18 basket group as illustrated

Discount products by basket analysis

To optimize the additional profit margin of cross-selling, among the 18 baskets listed, we need to implement the following:
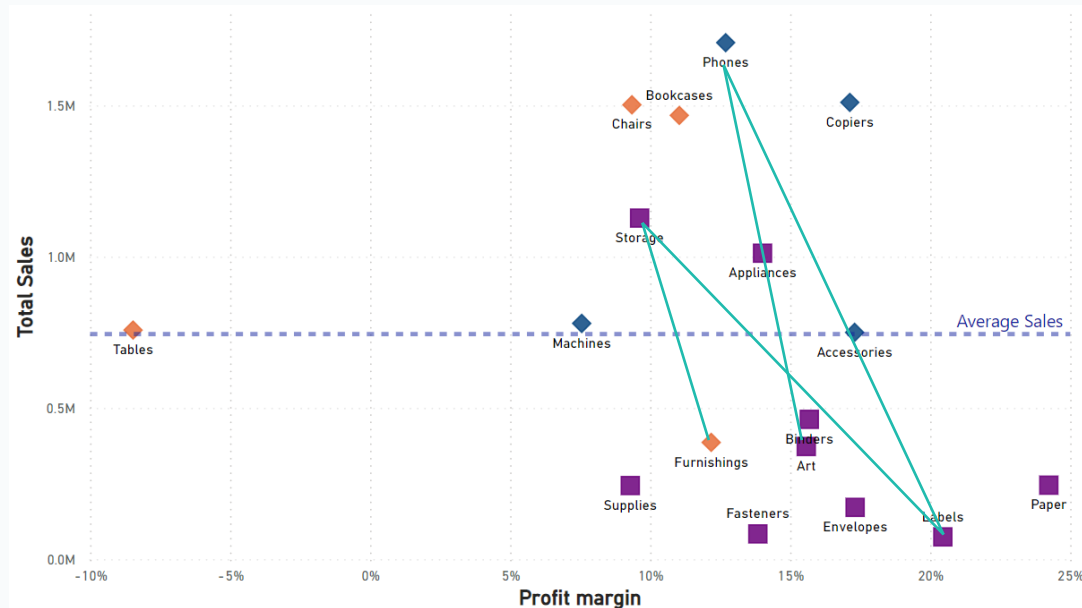
(A) 1 sub-category that sells a lot, but has a low profit margin

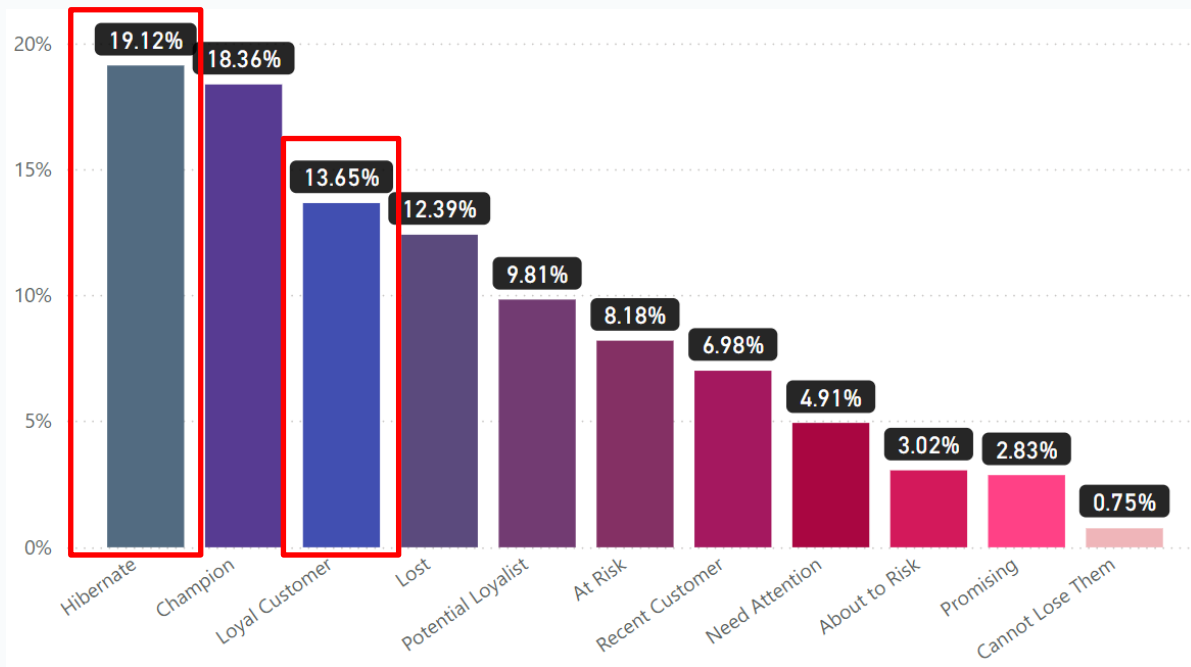**(B)** 1 sub-category that sells less, but has a higher profit margin

→ we can increase the overall profit margin of the order.

We have **4 suitable basket** :
- Phones + Art
- Phones + Labels
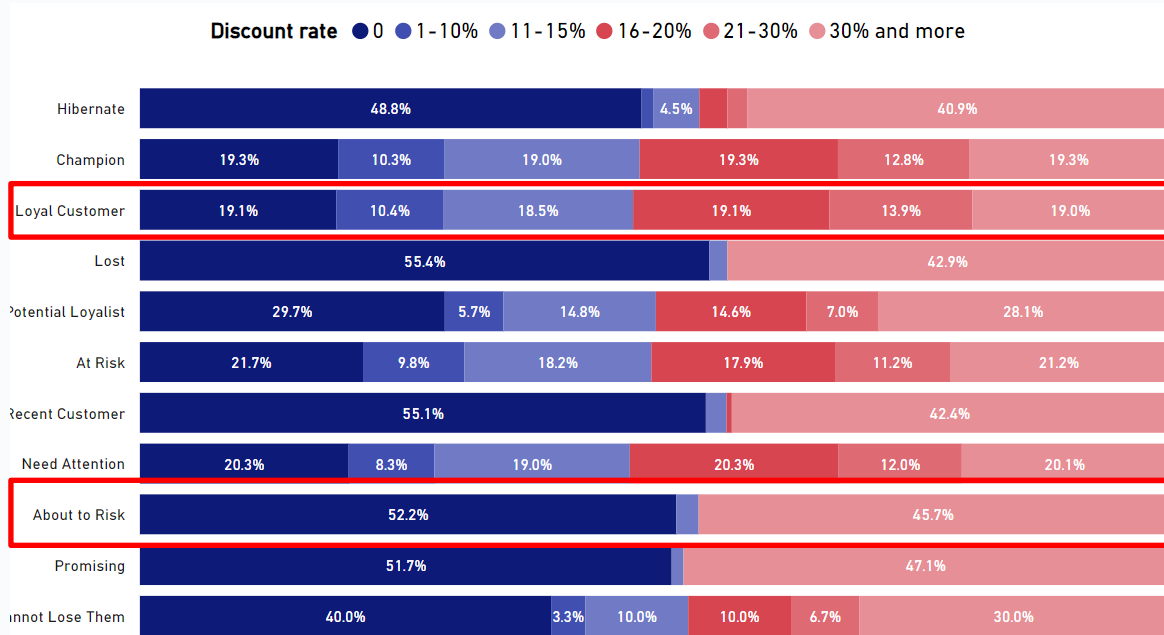- Storage + Label
- Furnishing + Storage

Discount products by Customer RFM

Focusing on **Hibernate** and **Loyal Customer** to boost Sales because they purchase more and easy to attract

Category	Value
Hibernate	19.12%
Champion	18.36%
Loyal Customer	13.65%
Lost	12.39%
Potential Loyalist	9.81%
At Risk	8.18%
Recent Customer	6.98%
Need Attention	4.91%
About to Risk	3.02%
Promising	2.83%
Cannot Lose Them	0.75%

Discount products by Customer RFM

Discount rate ● 0 ● 1-10% ● 11-15% ● 16-20% ● 21-30% ● 30% and more

| Segment						
Hibernate	48.8%	4.5%				40.9%
Champion	19.3%	10.3%	19.0%	19.3%	12.8%	19.3%
Loyal Customer	19.1%	10.4%	18.5%	19.1%	13.9%	19.0%
Lost	55.4%					42.9%
Potential Loyalist	29.7%	5.7%	14.8%	14.6%	7.0%	28.1%
At Risk	21.7%	9.8%	18.2%	17.9%	11.2%	21.2%
Recent Customer	55.1%					42.4%
Need Attention	20.3%	8.3%	19.0%	20.3%	12.0%	20.1%
About to Risk	52.2%					45.7%
Promising	51.7%					47.1%
Cannot Lose Them	40.0%	3.3%	10.0%	10.0%	6.7%	30.0%

Discount more for "About to risk" to attract and retain them

Less discount for loyal customer because they may not need as steep of discounts compared to new or less engaged customers
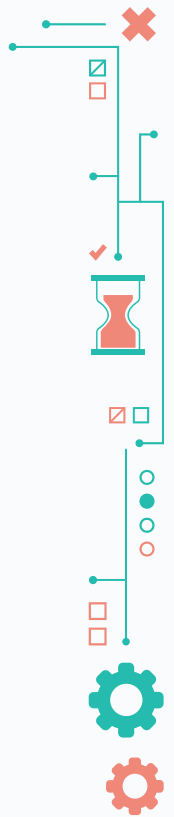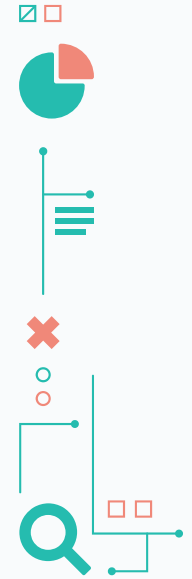
DISCOUNT BY SEASON
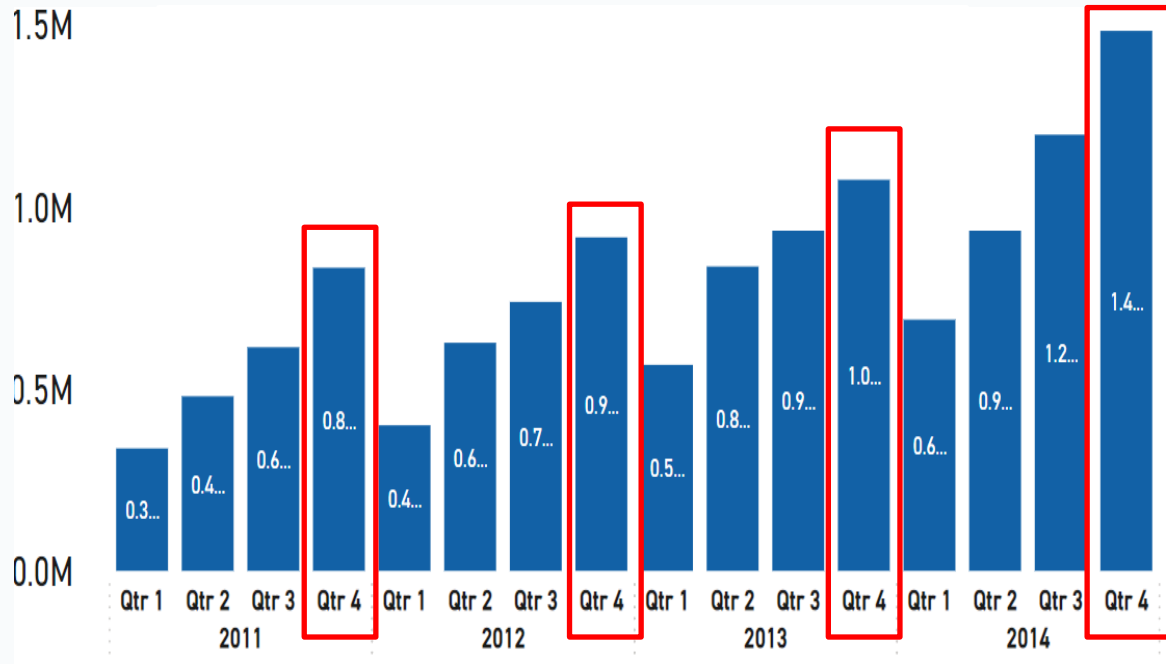
The revenue achieved the highest level at the final quarter of the year because the demand for big holiday such as Christmas, New Year,...

Discount more in the last season of the year

05

EXECUTIVE SUMMARY

EXECUTIVE SUMMARY

PROBLEMS IDENTIFICATION

01 NEGATIVE PROFIT MARGIN

02 LOW RATE OF NEW CUSTOMERS

REASONS

01 DISCOUNT RATE

02 PURCHASE POLICY

RECOMMENDATIONS

01 Stop discount >20%

02 Discount by basket

03 Discount by customer segment