



ANMART

Dự án phân tích dữ liệu

NỘI DUNG

- 01** Đặt vấn đề
- 02** Phân tích dữ liệu
- 03** Báo cáo
- 04** Kết luận

01

Đặt vấn đề

1.1. Giới thiệu

Anmart là chuỗi cửa hàng gồm nhiều chi nhánh ở Mỹ chuyên cung cấp đa dạng các sản phẩm:

- Văn phòng phẩm
- Nội thất
- Sản phẩm công nghệ

1.2. Thách thức

- Lợi nhuận chưa tương xứng với doanh thu

Tại sao ?



02

Phân tích dữ liệu

2.1. Dataset

SampleSuperstore.csv (1.11 MB)

1	Category	Chủng loại sản phẩm
2	City	Thành phố
3	Country	Quốc gia: United States
4	Customer Name	Tên khách hàng
5	Manufacturer	Nhà sản xuất
6	Order Date	Ngày đặt hàng
7	Order ID	Mã đơn hàng, là giá trị duy nhất
8	Postal Code	Mã bưu chính
9	Product Name	Tên sản phẩm
10	Region	Vùng
11	Segment	Phân khúc khách hàng

12	Ship Date	Ngày giao hàng
13	Ship Mode	Phương thức giao
14	State	Bang
15	Sub-Category	Danh mục con
16	Discount	Chiết khấu
17	Number of Records	Số lượng bản ghi
18	Profit	Lợi nhuận
19	Profit ratio	Tỉ suất lợi nhuận
20	Quantity	Số lượng
21	Sales	Doanh thu

Giai đoạn 2013 –2016

2.2. Công cụ thực hiện



Python

- Làm sạch và chuẩn hóa dữ liệu
- Tạo thêm các cột mới phục vụ phân tích
- Tách dataset gốc thành 3 bảng (sales, customers, products) để trực quan hóa hiệu quả hơn

Power BI

- Kiểm tra, chuyển đổi và tạo mối quan hệ giữa các bảng
- Xây dựng dashboard tương tác, trực quan hóa dữ liệu

Python

```
# Kiểm tra dữ liệu
df.isnull().sum()
```

✓ 0.0s

Category	0
City	0
Country	37
Customer Name	0
Manufacturer	3
Order Date	0
Order ID	0
Postal Code	0
Product Name	5
Region	0
Segment	0
Ship Date	0
Ship Mode	0
State	0
Sub-Category	0
Discount	0
Number of Records	0
Profit	0
Profit Ratio	58
Quantity	0
Sales	0
dtype: int64	

```
# Điền "United States" đối với các dòng thiếu Country
df['Country'] = df['Country'].fillna('United States')
✓ 0.0s
```

```
# Xóa các dòng thiếu Manufacturer hoặc Product Name
df.dropna(inplace=True)
✓ 0.0s
```

→

```
# Điền các dòng thiếu Profit Ratio theo công thức Profit Ratio = Profit/Sales
```

```
# Xóa $, khoảng trắng, chuyển kiểu dữ liệu thành số nguyên để tính toán
df['Profit'] = pd.to_numeric(df['Profit'].astype(str).str.replace('$', '', regex=False).
                             str.replace(',', '', regex=False).str.strip(), errors='coerce')
df['Sales'] = pd.to_numeric(df['Sales'].astype(str).str.replace('$', '', regex=False).
                             str.replace(',', '', regex=False).str.strip(), errors='coerce')
```

```
# Tính lại Profit Ratio cho các giá trị NaN
df.loc[df['Profit Ratio'].isna(), 'Profit Ratio'] = df['Profit'] / df['Sales']
✓ 0.0s
```

```
df.isnull().sum()
```

✓ 0.0s

Category	0
City	0
Country	0
Customer Name	0
Manufacturer	0
Order Date	0
Order ID	0
Postal Code	0
Product Name	0
Region	0
Segment	0
Ship Date	0
Ship Mode	0
State	0
Sub-Category	0
Discount	0
Profit	0
Profit Ratio	0
Quantity	0
Sales	0
dtype: int64	

Xóa dòng trùng lặp, xóa cột không cần thiết, vv...

Python

3.1. Thêm 2 cột Customer ID - Product ID

```
# mã hóa các giá trị duy nhất thành số nguyên
df['Customer ID'] = df['Customer Name'].factorize()[0] + 1
# luôn có dạng: Cxxxx
df['Customer ID'] = df['Customer ID'].apply(lambda x: f"C{x:04}")
```

```
df['Product ID'] = df['Product Name'].factorize()[0] + 1
df['Product ID'] = df['Product ID'].apply(lambda x: f"P{x:04}")
```

3.2. Thêm cột Price

```
# Công thức: Price = Sales / [Quantity * (1 - Discount)]

df['Price'] = round(df['Sales'] / (df['Quantity'] * (1 - df['Discount'])), 2)
```

3.3. Thêm 3 cột: Year - Quarter - Month từ Order Date

```
df['Year'] = df['Order Date'].dt.year
df['Month'] = df['Order Date'].dt.month
df['Quarter'] = df['Order Date'].dt.quarter
```

IV. TÁCH BẢNG

```
sales = df[['Order ID', 'Customer ID', 'Product ID', 'Order Date', 'Ship Date', 'Ship Mode',
            'Price', 'Quantity', 'Discount', 'Sales', 'Profit Ratio', 'Profit', 'Year', 'Quarter', 'Month']]
```

```
customers = df[['Customer ID', 'Customer Name', 'Segment', 'City', 'State', 'Postal Code', 'Region']]
customers.drop_duplicates(inplace=True)
```

```
products = df[['Product ID', 'Product Name', 'Category', 'Sub-Category', 'Manufacturer']]
products.drop_duplicates(inplace=True)
```

V. LƯU BẢNG VỀ MÁY

```
sales.to_csv('sales.csv', index=False)          # Không lưu chỉ mục
customers.to_csv('customers.csv', index=False)
products.to_csv('products.csv', index=False)
```

Kết quả – gồm 3 files csv:

- **sales:** chứa thông tin giao dịch của khách hàng, gồm 15 cột
- **products:** thông tin sản phẩm của Anmart, gồm 5 cột
- **customers:** thông tin khách hàng, gồm 7 cột

Power BI

Import, chuẩn hóa dữ liệu:

The image shows two overlapping 'Query Settings' dialog boxes. The top-left box is for a query named 'customers', showing the 'APPLIED STEPS' list with 'Removed Duplicates' selected. The bottom-right box is for a query named 'sales', showing a longer 'APPLIED STEPS' list including 'Changed Type with Locale1' through 'Changed Type with Locale4', with 'Changed Type1' selected at the bottom.

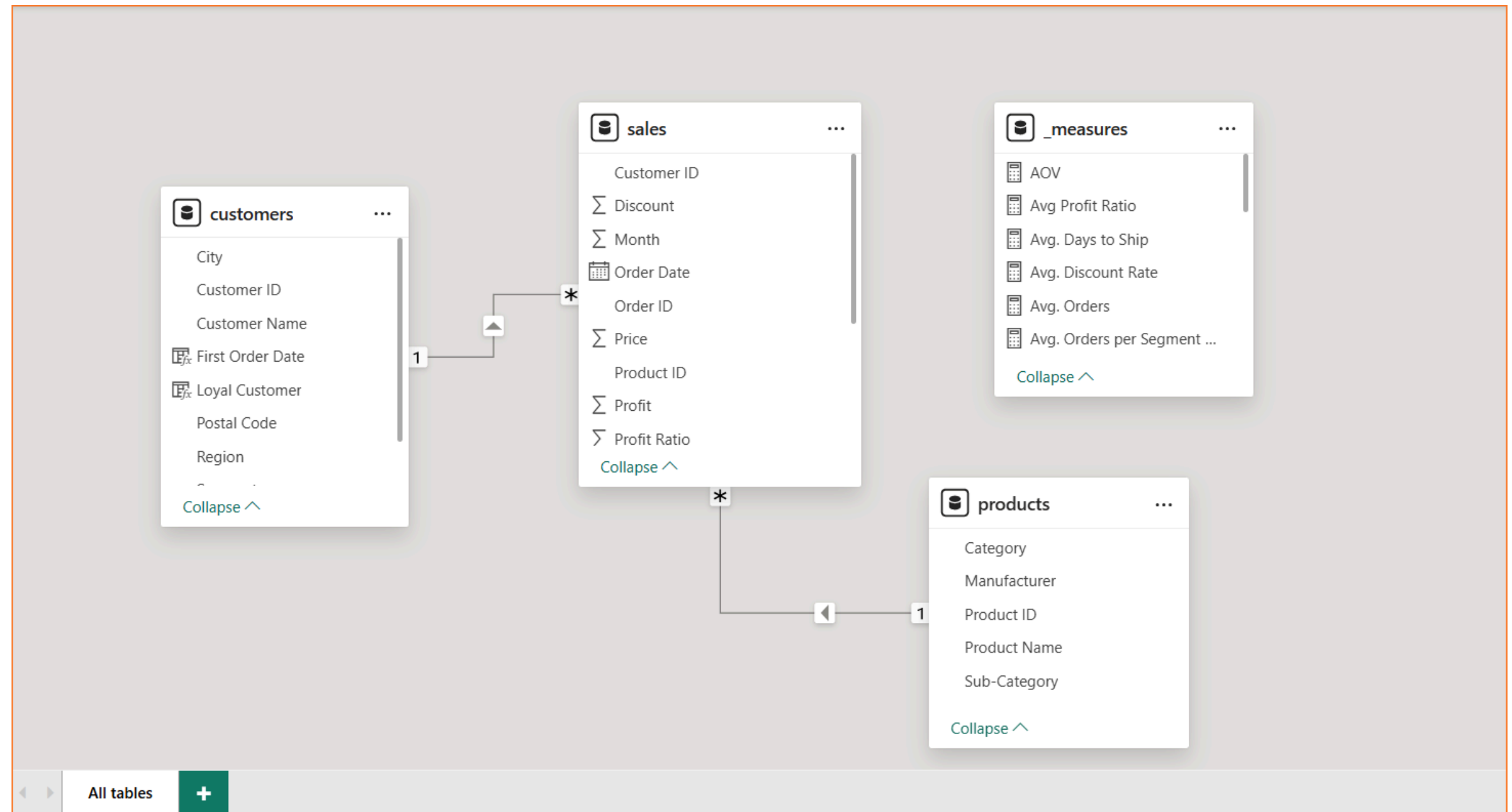
Query Settings (customers)

- PROPERTIES**
 - Name: customers
- APPLIED STEPS**
 - Source
 - Promoted Headers
 - Changed Type
 - Removed Duplicates

Query Settings (sales)

- PROPERTIES**
 - Name: sales
- APPLIED STEPS**
 - Source
 - Promoted Headers
 - Changed Type
 - Changed Type with Locale
 - Changed Type with Locale1
 - Changed Type with Locale2
 - Changed Type with Locale3
 - Changed Type with Locale4
 - Changed Type1

Kiểm tra data model:



Power BI

Tạo các measures

- ✓ measures
 - ☐ AOV
 - ☐ Avg Profit Ratio
 - ☐ Avg. Days to Ship
 - ☐ Avg. Discount Rate
 - ☐ Avg. Orders
 - ☐ Avg. Orders per Segment Customer
 - ☐ Avg. Price (Discounted)
 - ☐ Customers
 - ☐ Discount Amount
 - ☐ Discounted Orders
 - ☐ New Customers
 - ☐ Orders
 - ☐ Profit
 - ☐ Profit from Discounted
 - ☐ Quantity Sold
 - ☐ Repeat Customer
 - ☐ Sales
 - ☒ Sales per Shipped Order
 - ☐ Sub-Category
 - ☐ Top Sub-Category
 - ☐ Top-Profit Product
 - ☐ Top-Sale Product
 - ☐ Units Shipped

```
Sales = SUM(sales[Sales])
```

```
Profit = SUM(sales[Profit])
```

```
Discount Amount =  
ROUND(SUMX(  
    sales,  
    sales[Sales] * sales[Discount]), 2)
```

```
Avg. Days to Ship =  
AVERAGEX(  
    FILTER(sales, NOT ISBLANK(sales[Ship Date])),  
    DATEDIFF(sales[Order Date], sales[Ship Date], DAY)  
)
```

```
Repeat Customer =  
CALCULATE(  
    DISTINCTCOUNT(customers[Customer ID]),  
    FILTER(  
        SUMMARIZE(sales, customers[Customer ID], "OrderCount",  
            DISTINCTCOUNT(sales[Order ID])),  
        [OrderCount] > 1  
    )
```

Power BI

Một số công thức được sử dụng

- $AOV \text{ (Average Order Value)} = \text{Sales} / \text{Orders}$
- $\text{Avg. Days to ship} = \text{Day}[\text{Ship Date}] - \text{Day}[\text{Order Date}]$
- $\text{Avg. Orders} = \text{Orders} / \text{Customers}$
- $\text{Discount Amount} = \text{Sales} * \text{Discount}$
- Repeat Customer: $\text{OrderCount} > 1$

03

Báo cáo

Power BI

01



SALES DASHBOARD

Product
AllSegment
AllYear
AllQuarter
AllRegion
All

\$2,3M

Sales



\$286,77K

Profit



5008

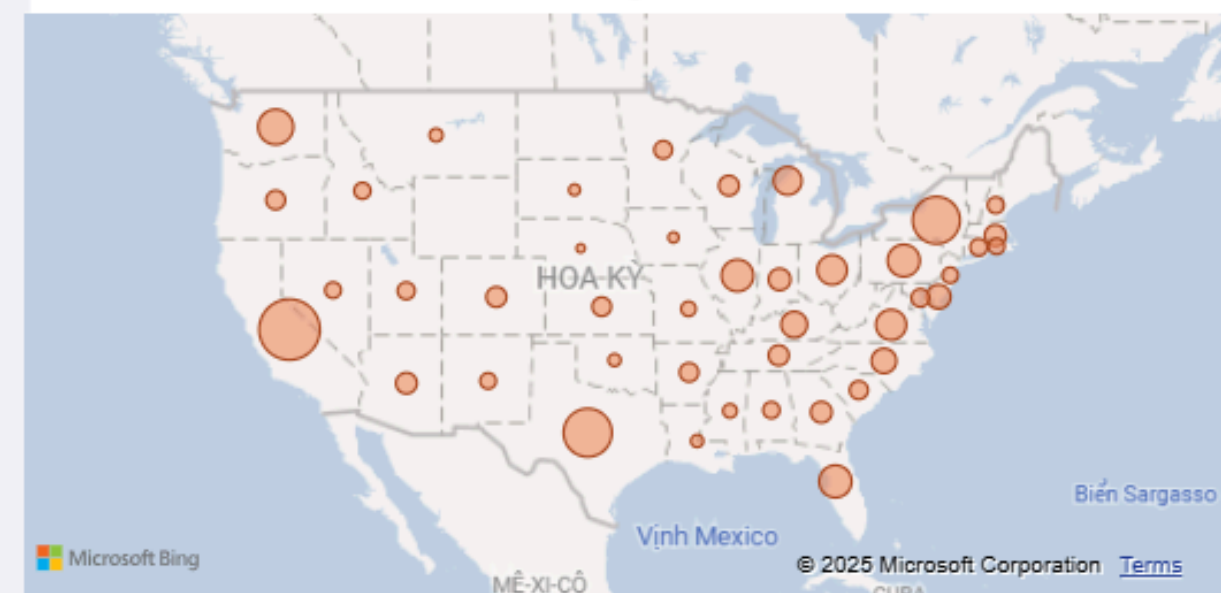
Orders



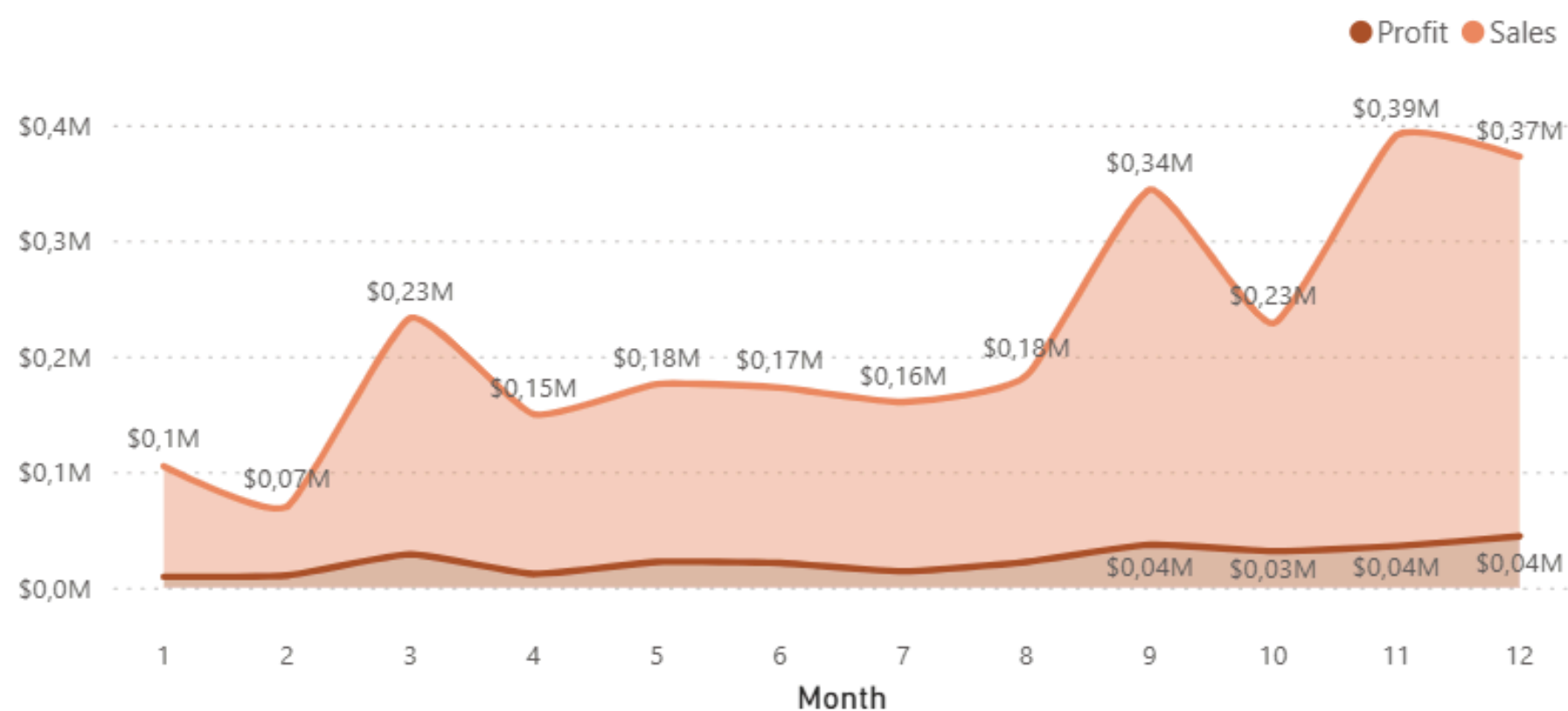
12,09%

Avg Profit Ratio

Sales by State

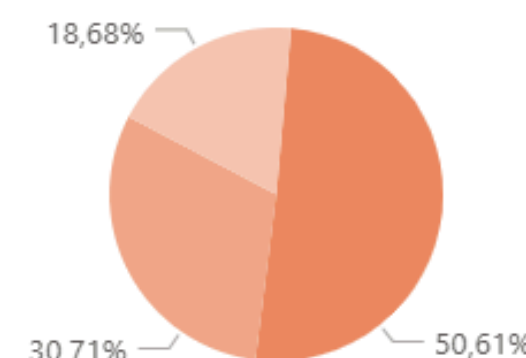


Profit Ratio

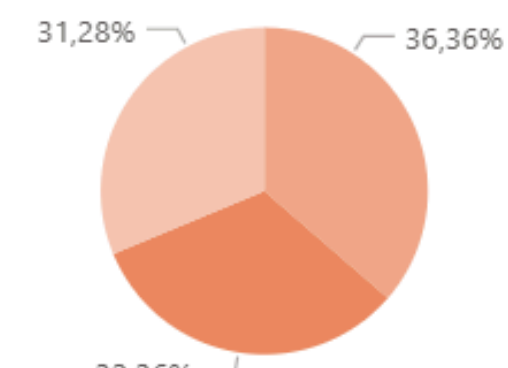


Sale Ratio

by Segment



by Category



INTRO

SALES^x

PRODUCT

CUSTOMER

SHIPMENT

PROMOTION



02



PRODUCT DASHBOARD

Product
AllSegment
AllYear
AllQuarter
AllRegion
All

38K

Quantity Sold

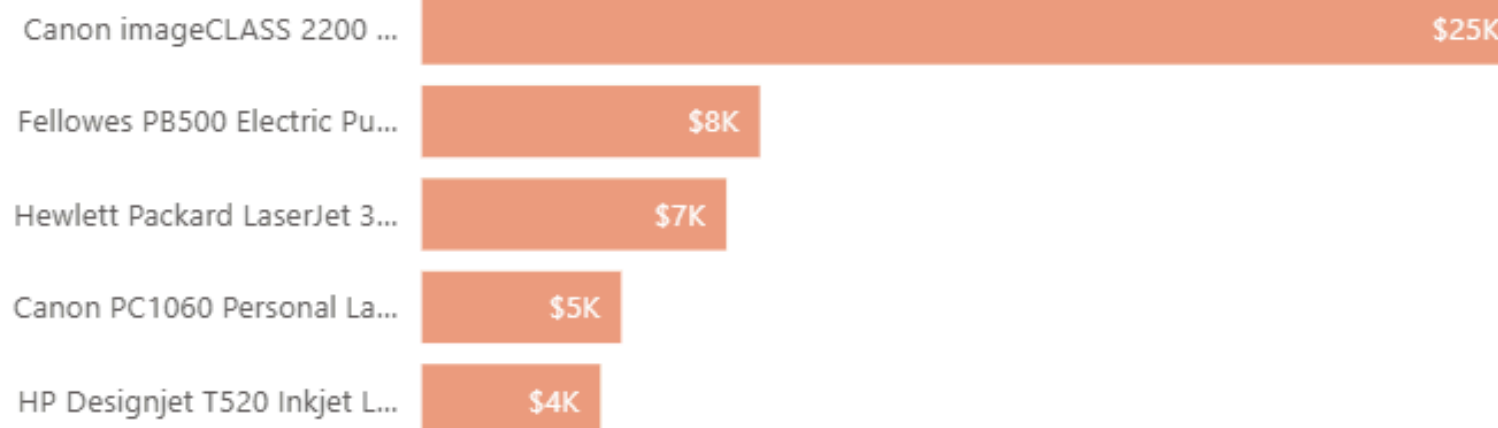
Canon
imageCLASS
2200

Top-Sale Product

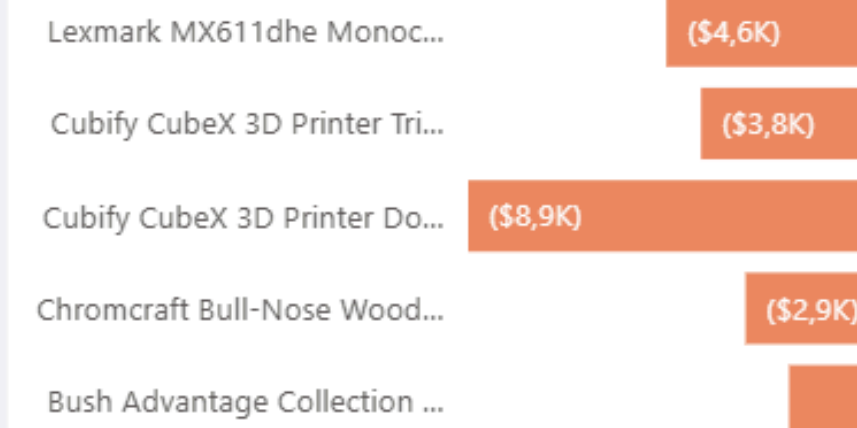
Chairs

Top Sub-Category

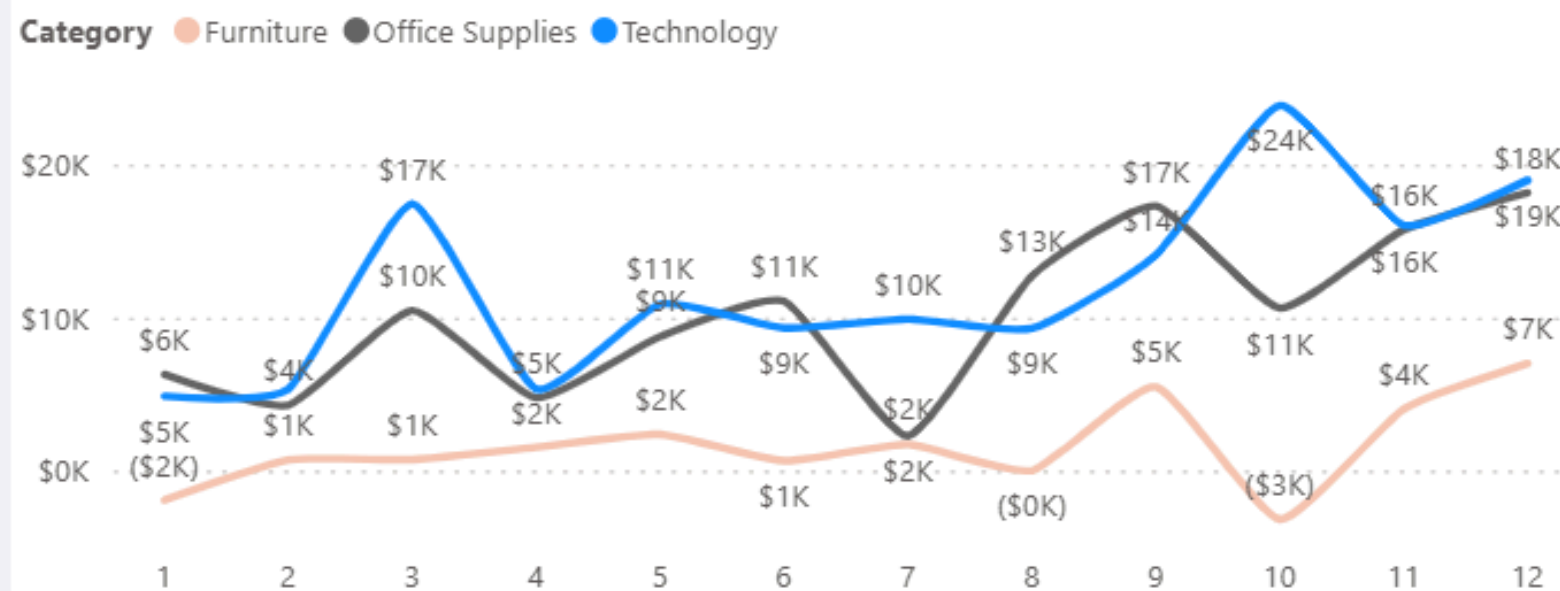
Top 5 most profitable products



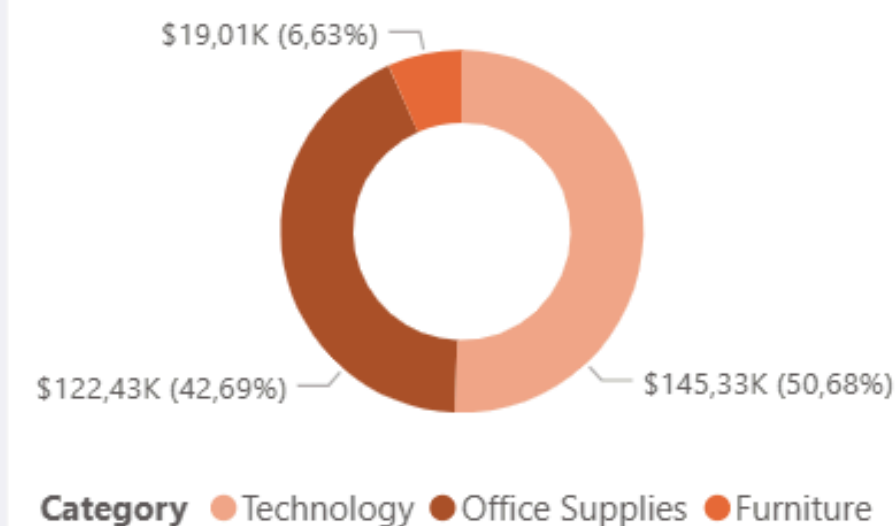
Top 5 most loss making products



Profit Trend by Category



Profit Distribution by Category



INTRO

SALES

PRODUCT ^x

CUSTOMER

SHIPMENT

PROMOTION

+

03



CUSTOMER DASHBOARD

Loyal Customer



All



Segment



All



Year



All



Quarter



All



Region



All



800

Customers

781

Repeat Customer

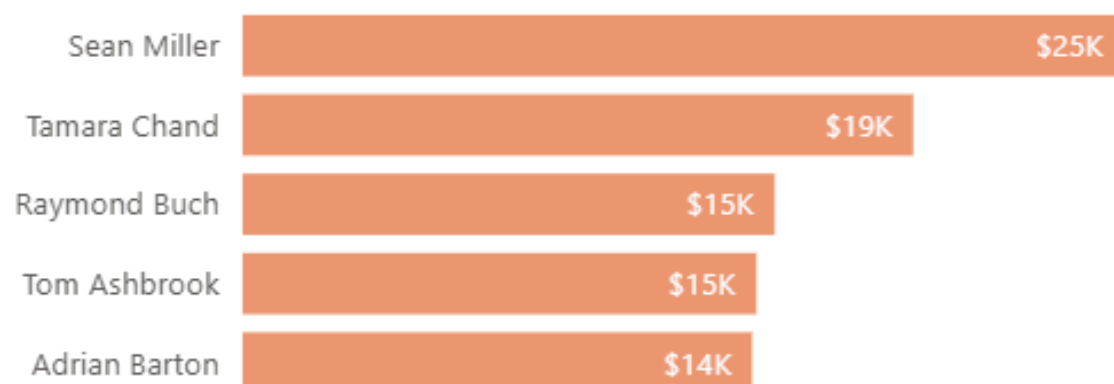
\$459,07

AOV

6,26

Avg. Orders

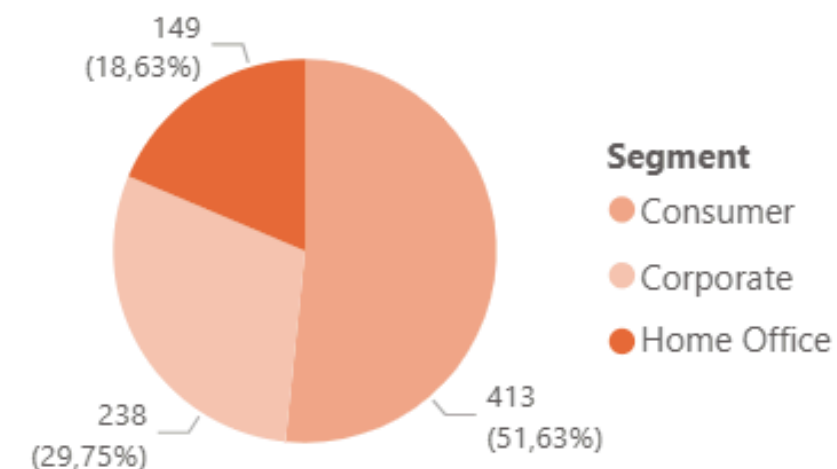
Top 5 Customers by Sales



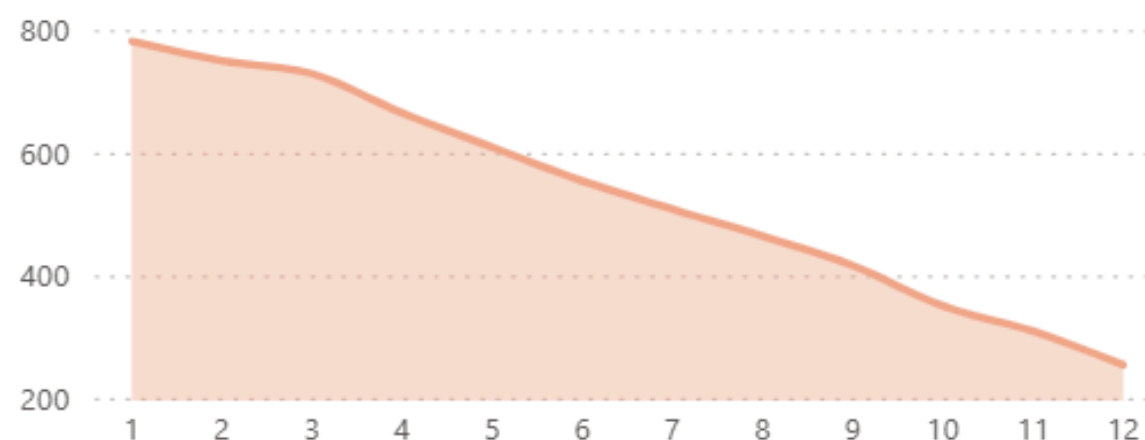
Avg. Orders per Segment



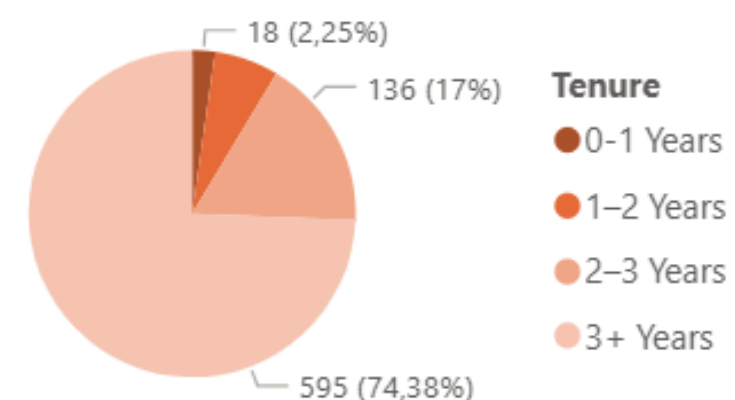
Customers by Segment



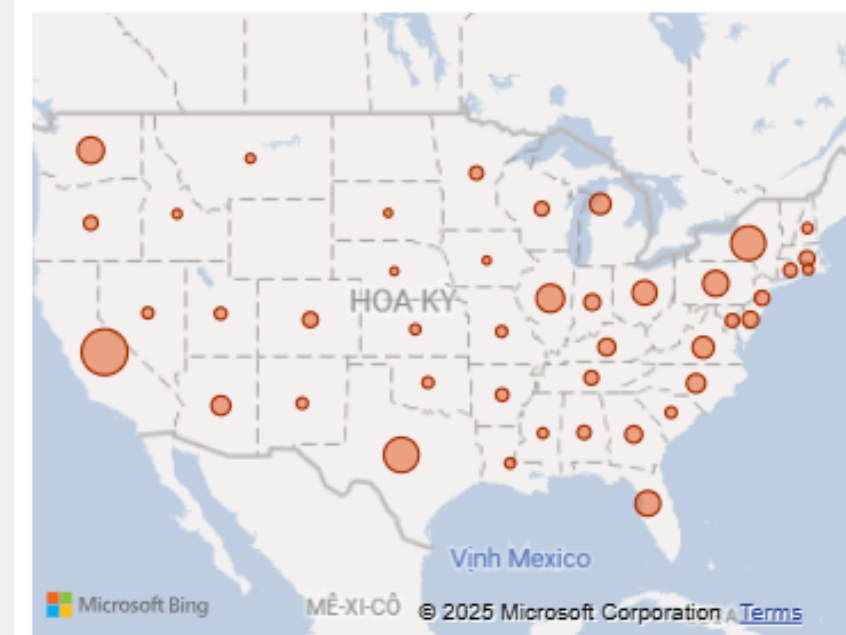
New Customers by Month



Customer Tenure Groups



Customers by State



INTRO

SALES

PRODUCT

CUSTOMER^x

SHIPMENT

PROMOTION



04



SHIPMENT DASHBOARD

Segment ▾

All ▾

Year ▾

All ▾

Quarter ▾

All ▾

Region ▾

All ▾

5008

Orders

38K

Units Shipped

3,96

Avg. Days to Ship

\$459,07

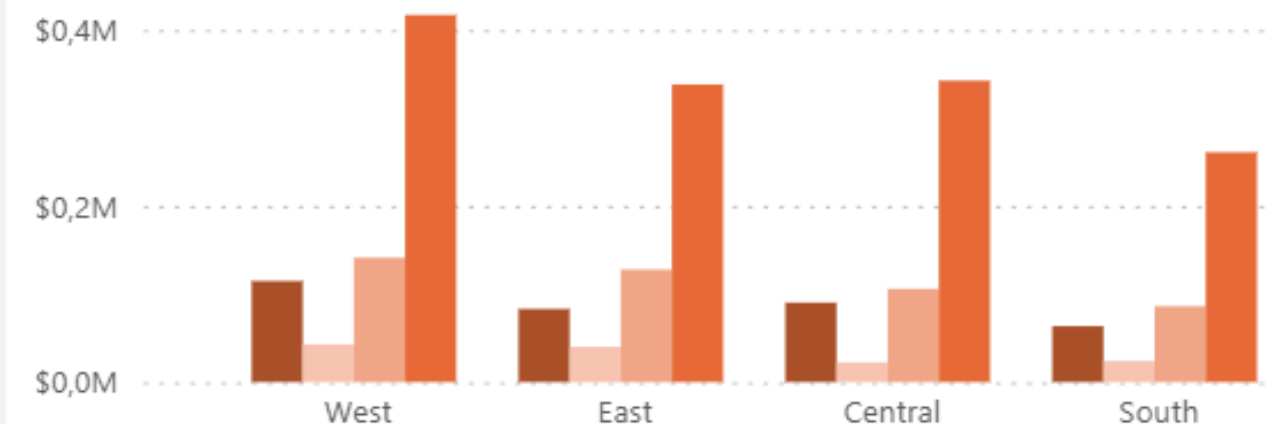
Sales per Shipped Order

Monthly Order Volume and Avg. Days to Ship Trend

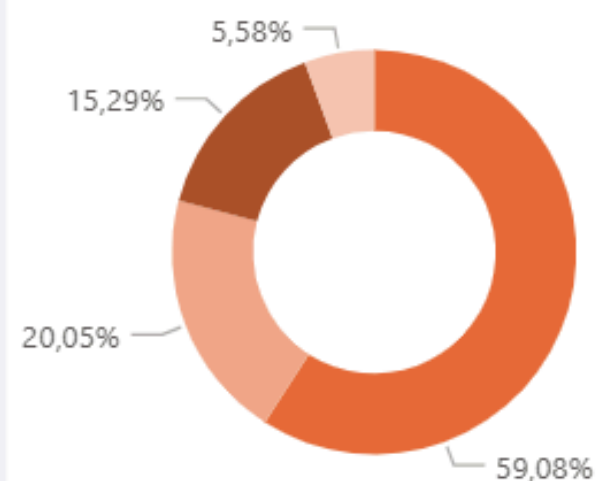


Total Sales by Region and Shipment

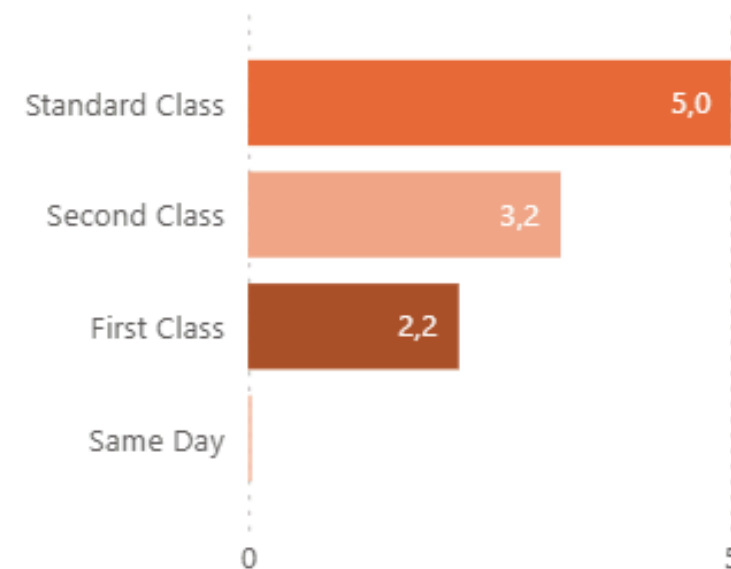
Ship Mode ● First Class ● Same Day ● Second Class ● Standard Class



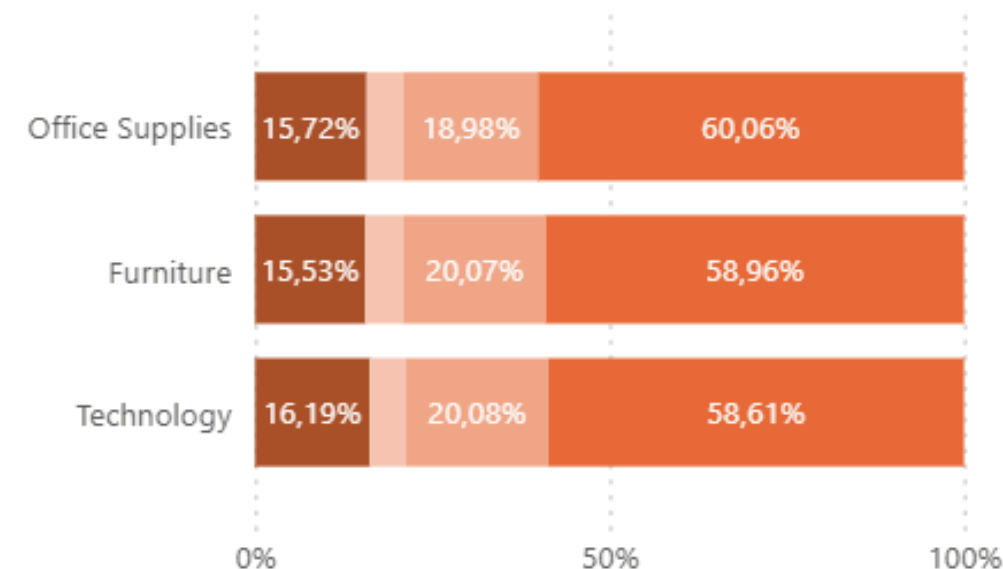
Total Sales by Shipment



Avg. Days to Ship by Shipment



Order Mix by Category and Shipment



INTRO

SALES

PRODUCT

CUSTOMER

SHIPMENT ^x

PROMOTION



05



PROMOTION DASHBOARD

Product ▼
All ▼

Segment ▼
All ▼

Year ▼
All ▼

Quarter ▼
All ▼

Region ▼
All ▼

\$322,57K
Discount Amount

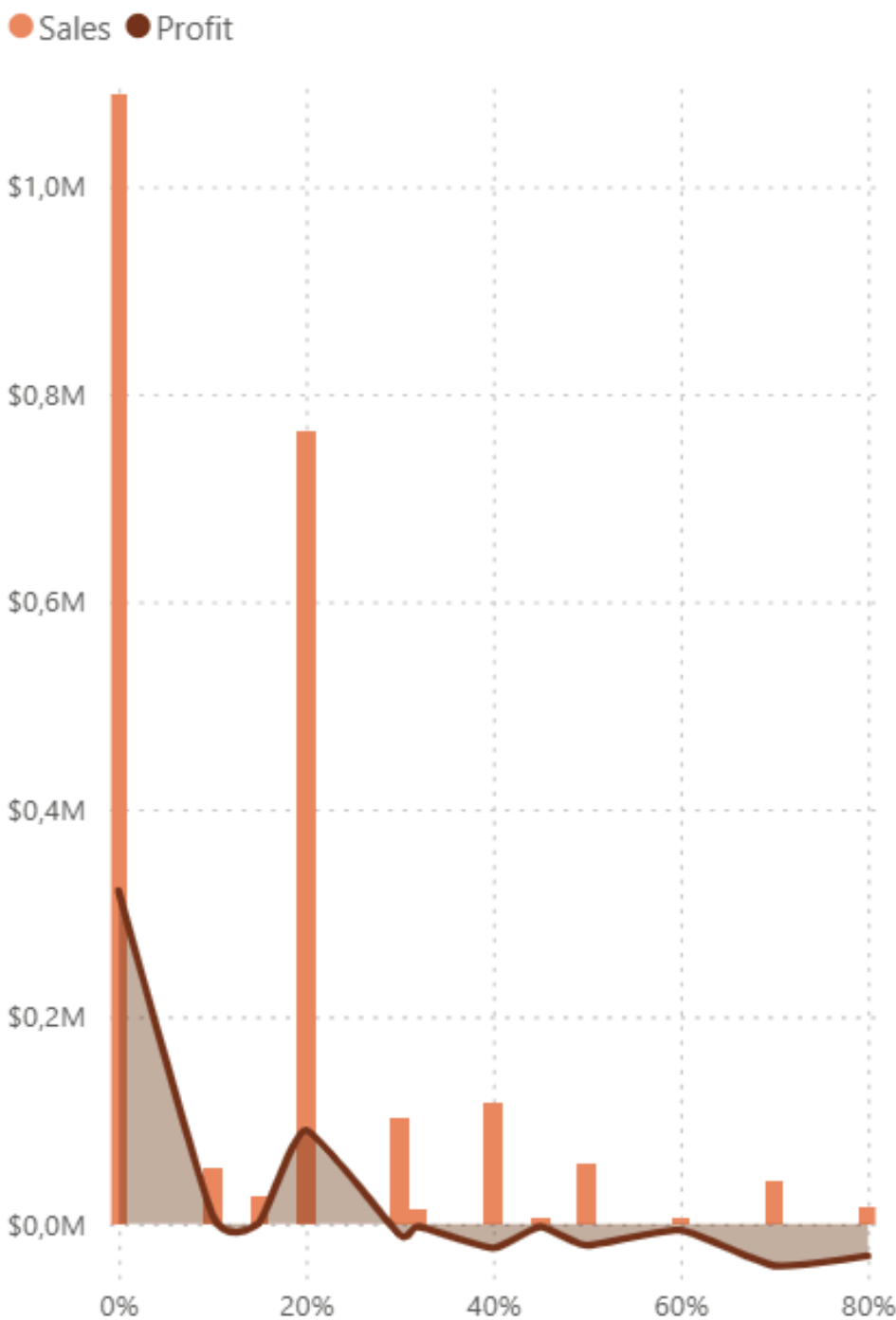
15,60%
Avg. Discount Rate

\$91,02
Avg. Price (Discounted)

2953
Discounted Orders

(\$34,56K)
Profit from Discounted

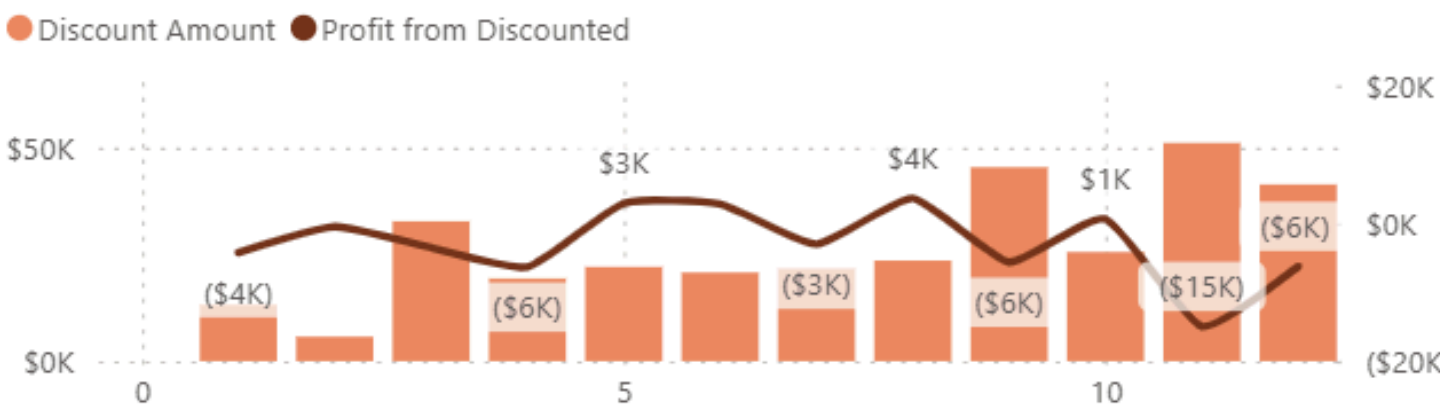
Sales & Profit vs. Discount Rate



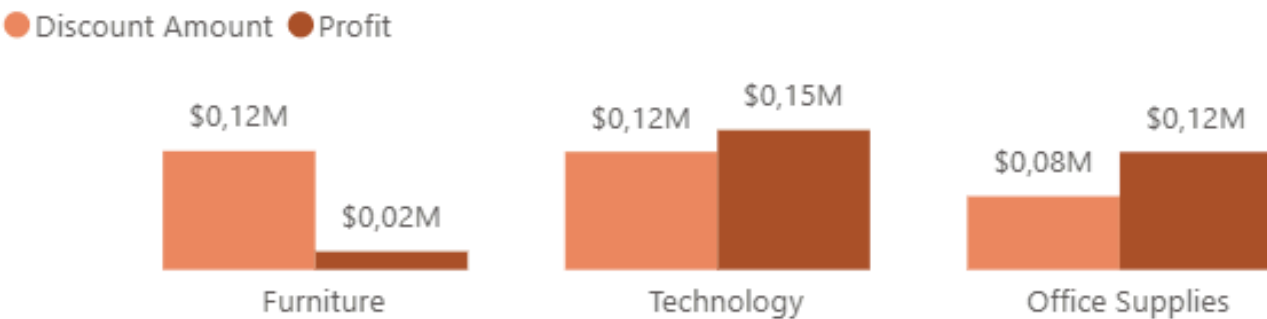
Top 3 Products by Avg. Discount Rate

Product ID	Avg. Discount Rate	Avg. Price (Discounted)	Quantity Sold	Profit
P1468	72,86%	\$11,22	23	(\$70)
P1499	72,50%	\$121,39	15	(\$410)
P1818	80,00%	\$5	2	(\$4)

Monthly Discount Trend & Profit Impact



Discount Impact by Category



INTRO

SALES

PRODUCT

CUSTOMER

SHIPMENT

PROMOTION x



04

Kết luận

**Lợi nhuận
chưa
tương
xứng với
doanh thu
?**

Lí do

- 01** Tỷ suất lợi nhuận gộp thấp: chỉ đạt 12,09%
- 02** Danh mục Nội thất là nguyên nhân chính gây lỗ
- 03** Thời gian giao hàng kéo dài trong mùa cao điểm
- 04** Chương trình chiết khấu làm giảm lợi nhuận

Đề xuất



01 – Chiến lược chiết khấu

Cân nhắc điều chỉnh mức độ và cách thức áp dụng để đảm bảo chúng thực sự mang lại lợi nhuận, thay vì chỉ tập trung đẩy mạnh doanh số

02 – Danh mục và hiệu suất sản phẩm

Nghiên cứu danh mục Nội thất (chi phí, giá bán, marketing, ..) và điều chỉnh cần thiết
Xem xét lại những mặt hàng không hiệu quả để tối ưu hóa danh mục

03 – Vận chuyển

Rà soát lại các quy trình vận chuyển và giao hàng, đặc biệt là trong các giai đoạn cao điểm

04 – Tăng cường văn hóa ra quyết định dựa trên dữ liệu

Theo dõi chặt chẽ các chỉ số quan trọng → Kịp thời nhận diện vấn đề và đưa ra giải pháp

Thanks for
your listening

The logo consists of a horizontal rectangle with a black border. The interior of the rectangle has a horizontal color gradient, transitioning from a bright yellow on the left to a light orange on the right. The word "ANMART" is centered within this rectangle in a dark orange, sans-serif, uppercase font.

ANMART