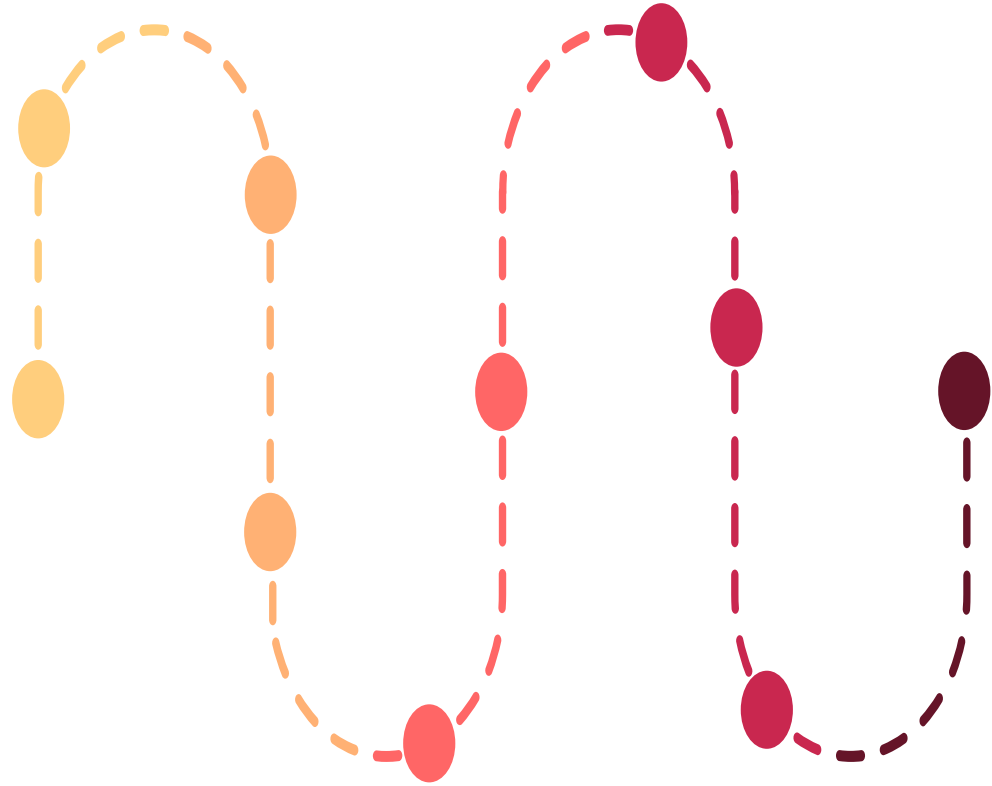


Customer Insights for E-Commerce Company

Presenter: Hoang Thi Ngoc Han



NỘI DUNG

- 01 **Giới thiệu**
- 02 **Phân tích yêu cầu**
- 03 **Làm sạch dữ liệu**
- 04 **Phân tích**
- 05 **Kết luận/Insights**

1. Giới thiệu

Lĩnh vực: E-commerce

Dataset: kết quả hoạt động bán hàng trên sàn thương mại điện tử của một doanh nghiệp

[Kaggle](#)

Dữ liệu giao dịch đã được cung cấp trong khoảng thời gian từ ngày 1 tháng 1 năm 2019 đến ngày 31 tháng 12 năm 2019.

Bán 20 danh mục sản phẩm khác nhau liên quan đến các vật dụng cần thiết trong hoạt động sống, mỗi danh mục sản phẩm thì có những mặt hàng liên quan.



Dataset

Customers

```
df_customer.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1468 entries, 0 to 1467
Data columns (total 4 columns):
#   Column          Non-Null Count  Dtype
---  ---
0   CustomerID      1468 non-null  int64
1   Gender          1468 non-null  object
2   Location        1468 non-null  object
3   Tenure_Months   1468 non-null  int64
dtypes: int64(2), object(2)
```

Marketing_Spend

```
df_marketing.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 365 entries, 0 to 364
Data columns (total 3 columns):
#   Column          Non-Null Count  Dtype
---  ---
0   Date            365 non-null   object
1   Offline_Spend   365 non-null   int64
2   Online_Spend    365 non-null   float64
```

Discount_Coupon

```
df_discount.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 204 entries, 0 to 203
Data columns (total 4 columns):
#   Column          Non-Null Count  Dtype
---  ---
0   Month           204 non-null   object
1   Product_Category 204 non-null   object
2   Coupon_Code     204 non-null   object
3   Discount_pct    204 non-null   int64
```

Tax

```
df_tax.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 20 entries, 0 to 19
Data columns (total 2 columns):
#   Column          Non-Null Count  Dtype
---  ---
0   Product_Category 20 non-null     object
1   GST             20 non-null     object
..  ..             ..             ..
```

Online_Sales

```
df_sale.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 52924 entries, 0 to 52923
Data columns (total 10 columns):
#   Column          Non-Null Count  Dtype
---  ---
0   CustomerID      52924 non-null  int64
1   Transaction_ID   52924 non-null  int64
2   Transaction_Date  52924 non-null  object
3   Product_SKU     52924 non-null  object
4   Product_Description 52924 non-null  object
5   Product_Category 52924 non-null  object
6   Quantity        52924 non-null  int64
7   Avg_Price       52924 non-null  float64
8   Delivery_Charges 52924 non-null  float64
9   Coupon_Status   52924 non-null  object
dtypes: float64(2), int64(3), object(5)
memory usage: 4.0+ MB
```

2. Phân tích yêu cầu



3. Làm sạch dữ liệu: Python

Kiểm tra thông tin

Định dạng kiểu dữ liệu

```
df_discount["Discount_pct"] = df_discount["Discount_pct"].astype(float) / 100
df_discount["Discount_pct"]
```

```
# Transaction_Date chưa ở định dạng ngày tháng
df_sale["Transaction_Date"] = pd.to_datetime(df_sale["Transaction_Date"])
df_sale.info()
```

```
# Thêm cột Month và chỉnh định dạng Month về giống Month của Discount
df_sale["Month"] = df_sale["Transaction_Date"].dt.month_name().str.upper().str[:3]
df_sale
```

```
df_tax["GST"] = df_tax["GST"].str.replace("%", "", regex=True)
df_tax
```

```
# Date chưa ở định dạng ngày tháng
df_marketing["Date"] = pd.to_datetime(df_marketing["Date"])
df_marketing
```

Đưa Discount vào cho từng danh mục sản phẩm

```
merged = pd.merge(df_sale, df_discount, on=["Month", "Product_Category"], how="left")
merged
```

Đưa mức Tax vào cho từng danh mục sản phẩm

```
fact = pd.merge(merged, df_tax, on=["Product_Category"], how="left")
fact
```

Sửa null

```
fact["Discount_pct"].fillna(0, inplace=True)
fact["Discount_pct"]
```

3. Làm sạch dữ liệu

- ❖ Các sản phẩm không được Discount lần nào trong năm: **Backpacks, Fun, Google, More Bags**

```
df_discount['Product_Category'].unique()
```

```
array(['Apparel', 'Nest-USA', 'Office', 'Drinkware', 'Lifestyle', 'Bags',  
      'Notebooks', 'Headgear', 'Nest', 'Waze', 'Bottles', 'Nest-Canada',  
      'Housewares', 'Accessories', 'Gift Cards', 'Notebooks & Journals',  
      'Android'], dtype=object)
```

```
df_sale['Product_Category'].unique()
```

```
array(['Nest-USA', 'Office', 'Apparel', 'Bags', 'Drinkware', 'Lifestyle',  
      'Notebooks & Journals', 'Headgear', 'Waze', 'Fun', 'Nest-Canada',  
      'Backpacks', 'Google', 'Bottles', 'Gift Cards', 'More Bags',  
      'Housewares', 'Android', 'Accessories', 'Nest'], dtype=object)
```

Tính doanh thu

```
#Thêm cột revenue: phụ thuộc vào có Discount hay không
def calculate_revenue(i):
    if i["Coupon_Status"] == "Used":
        return (i["Avg_Price"] * i["Quantity"]) * (1 - i["Discount_pct"]) * (1 + i["GST"]) + i["Delivery_Charges"]
    else:
        return (i["Avg_Price"] * i["Quantity"]) * (1 + i["GST"]) + i["Delivery_Charges"]
fact["Revenue"] = fact.apply(calculate_revenue, axis=1)
fact
```

4. Phân tích



Kết quả phân tích

4. Phân tích

SQL: Cohort Analysis

```
--BUOC 3:tao bang cohort so retained customer theo month
retained_table as(
select first_month, count(distinct CustomerID) as number_of_user,
COUNT(CASE WHEN month_diff = 0 then CustomerID END) month_0,
COUNT(CASE WHEN month_diff = 1 then CustomerID END) month_1,
COUNT(CASE WHEN month_diff = 2 then CustomerID END) month_2,
COUNT(CASE WHEN month_diff = 3 then CustomerID END) month_3,
COUNT(CASE WHEN month_diff = 4 then CustomerID END) month_4,
COUNT(CASE WHEN month_diff = 5 then CustomerID END) month_5,
COUNT(CASE WHEN month_diff = 6 then CustomerID END) month_6,
COUNT(CASE WHEN month_diff = 7 then CustomerID END) month_7,
COUNT(CASE WHEN month_diff = 8 then CustomerID END) month_8,
COUNT(CASE WHEN month_diff = 9 then CustomerID END) month_9,
COUNT(CASE WHEN month_diff = 10 then CustomerID END) month_10,
COUNT(CASE WHEN month_diff = 11 then CustomerID END) month_11
from final_table
group by first_month)
--order by first_month
```

```
select * from fact
SELECT COUNT(DISTINCT(month(Transaction_Date))) AS distinct_transaction_dates
FROM fact;
--BUOC 1: tinh thang mua hang dau tien cua CustomerID
with raw1 as(
select CustomerID, min(month(Transaction_Date)) as fist_month from fact
where year(Transaction_Date) = 2019
group by CustomerID),
--BUOC 2: join lay thong tin month(Transaction_Date) va fist_month theo customerID
raw2 as(
select CustomerID, month(Transaction_Date) as order_month from fact
where year(Transaction_Date) = 2019),
final_table as(
select distinct a.CustomerID, fist_month, order_month, order_month - fist_month as month_diff from raw1 a
left join raw2 b
on a.CustomerID = b.CustomerID),
--select * from final_table
--order by CustomerID
```

```
--BUOC 4: tinh bang % retained customer theo month previous_month
select fist_month,
cast(month_0 as float)/month_0 as month_0,
round(CASE WHEN month_0 <> 0 then cast(month_1 as float)/month_0 END,2) as month_1,
round(CASE WHEN month_1 <> 0 then cast(month_2 as float)/month_1 END,2) as month_2,
round(CASE WHEN month_2 <> 0 then cast(month_3 as float)/month_2 END,2) as month_3,
round(CASE WHEN month_3 <> 0 then cast(month_4 as float)/month_3 END,2) as month_4,
round(CASE WHEN month_4 <> 0 then cast(month_5 as float)/month_4 END,2) as month_5,
round(CASE WHEN month_5 <> 0 then cast(month_6 as float)/month_5 END,2) as month_6,
round(CASE WHEN month_6 <> 0 then cast(month_7 as float)/month_6 END,2) as month_7,
round(CASE WHEN month_7 <> 0 then cast(month_8 as float)/month_7 END,2) as month_8,
round(CASE WHEN month_8 <> 0 then cast(month_9 as float)/month_8 END,2) as month_9,
round(CASE WHEN month_9 <> 0 then cast(month_10 as float)/month_9 END,2) as month_10,
round(CASE WHEN month_10 <> 0 then cast(month_11 as float)/month_10 END,2) as month_11
from retained_table
order by fist_month
```

4. Phân tích

DAX

- Chỉ số AOV: mức chi tiêu trung bình/đơn hàng

```
AOV = DIVIDE(sum('fact'[Revenue]), sum('fact'[Quantity]))
```

- Phí vận chuyển trung bình/đơn hàng

```
Delivery_Charges per Quantity = DIVIDE('fact'[Delivery_Charges], 'fact'[Quantity])
```

- Phân nhóm thời gian gắn bó trung bình

```
Tenure_Months_Group =  
SWITCH(TRUE(),  
customer[Tenure_Months] < 12, "Dưới 1 năm",  
customer[Tenure_Months] < 24, "1-2 năm",  
customer[Tenure_Months] < 36, "2-3 năm",  
"Trên 3 năm")
```

- Phân nhóm phí vận chuyển trung bình

```
Delivery_Group =  
SWITCH(TRUE(),  
'fact'[Delivery_Charges] = 0, "NOT",  
'fact'[Delivery_Charges] <= 50, "1-50",  
'fact'[Delivery_Charges] <= 100, "51-100",  
>100")
```

4. Phân tích

DAX: RFM Analysis

- Tạo bảng RFM

```
RFM =  
SUMMARIZE(fact, fact[CustomerID],  
          "R", DATEDIFF(MAX('fact'[Transaction_Date]), TODAY(), DAY),  
          "F", DISTINCTCOUNT('fact'[Transaction_ID]),  
          "M", SUM('fact'[Revenue]))
```

- Tính các điểm R, F, M

```
F_Score =  
var percent_80 = PERCENTILE.EXC(RFM[F], 0.8)  
var percent_60 = PERCENTILE.EXC(RFM[F], 0.6)  
var percent_40 = PERCENTILE.EXC(RFM[F], 0.4)  
var percent_20 = PERCENTILE.EXC(RFM[F], 0.2)
```

```
return  
SWITCH(TRUE(),  
        RFM[F] >= percent_80 , 5,  
        RFM[F] >= percent_60 , 4,  
        RFM[F] >= percent_40 , 3,  
        RFM[F] >= percent_20 , 2,  
        1)
```

```
RFM_Score = RFM[R_Score] & RFM[F_Score] & RFM[M_Score]
```



Phân loại theo Segment: `Segment = RELATED(RankRFM[Segment])`

Overview



Tổng doanh thu năm 2019: 38,41 tỷ

Số lượng khách hàng: 1468

Khách hàng nữ có doanh thu nhiều hơn nam

Khách hàng đến từ Chicago đem lại doanh thu lớn nhất (35,06%), tiếp theo là California: 31% và thấp nhất là Washington DC (4,78%)

OVERVIEW

Revenue

Customer

Quantity

Discount

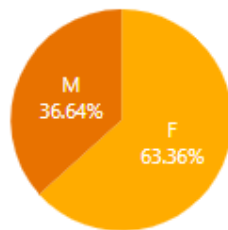
38.41bn

1468

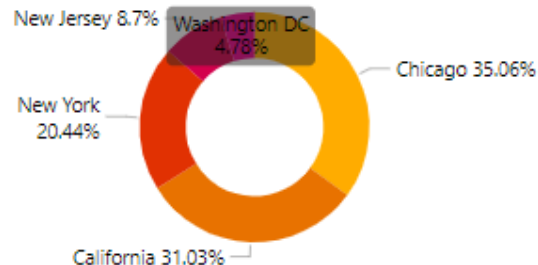
238K

52.52K

Revenue by Gender



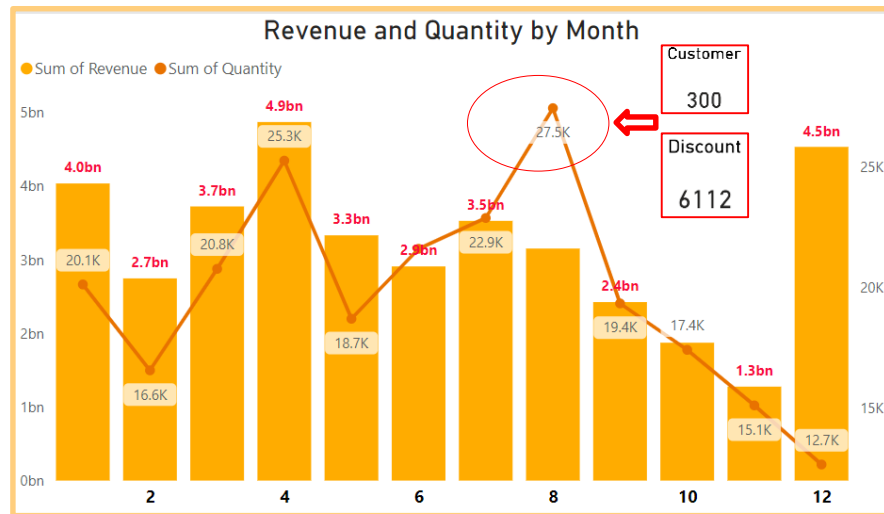
Revenue by Location



Overview

Doanh thu và số lượng sản phẩm bán ra theo tháng:

- Doanh thu cao nhất vào các tháng: 1, 4, 12 => tập trung chủ yếu vào các nhóm sp: **Apparel**, **Nest-Usa**, **Nest**, **Drinkware**, **Bags**. Đây cũng chính là 5 nhóm sản phẩm đem lại doanh thu cao nhất cho sàn.
- Tháng 8 thì có số lượng sản phẩm bán là lớn nhất: 27,5K (số lượng khách hàng, số sản phẩm được discount vào tháng này cũng là cao nhất)



Top Product Category

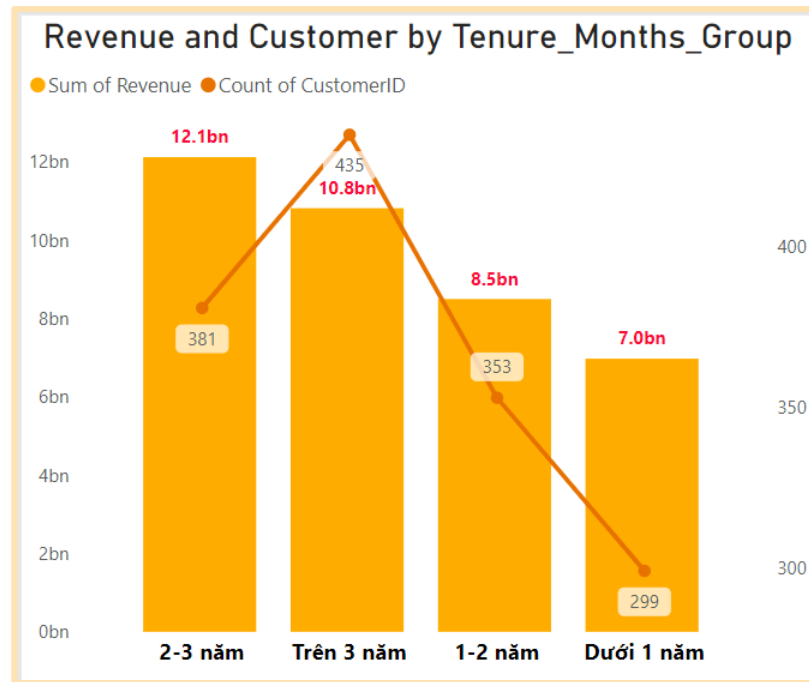
Apparel	
22,280,629,845.00	32438
Sum of Revenue	Sum of Quantity
Drinkware	
4,975,850,118.00	30501
Sum of Revenue	Sum of Quantity
Bags	
3,893,675,634.00	15273
Sum of Revenue	Sum of Quantity
Nest-USA	
1,681,784,181.00	21430
Sum of Revenue	Sum of Quantity
Nest	
1,502,851,450.00	2837
Sum of Revenue	Sum of Quantity

Overview

Và với tổng doanh thu như trên thì khách hàng lâu lắm hay khách hàng dưới 1 năm sẽ có xu hướng mua nhiều hơn?

⇒ Các khách hàng mà có thời gian gắn bó:

- 2-3 năm: đem lại doanh thu cao nhất
- Trên 3 năm: số lượng khách hàng đông nhất, doanh thu thấp hơn



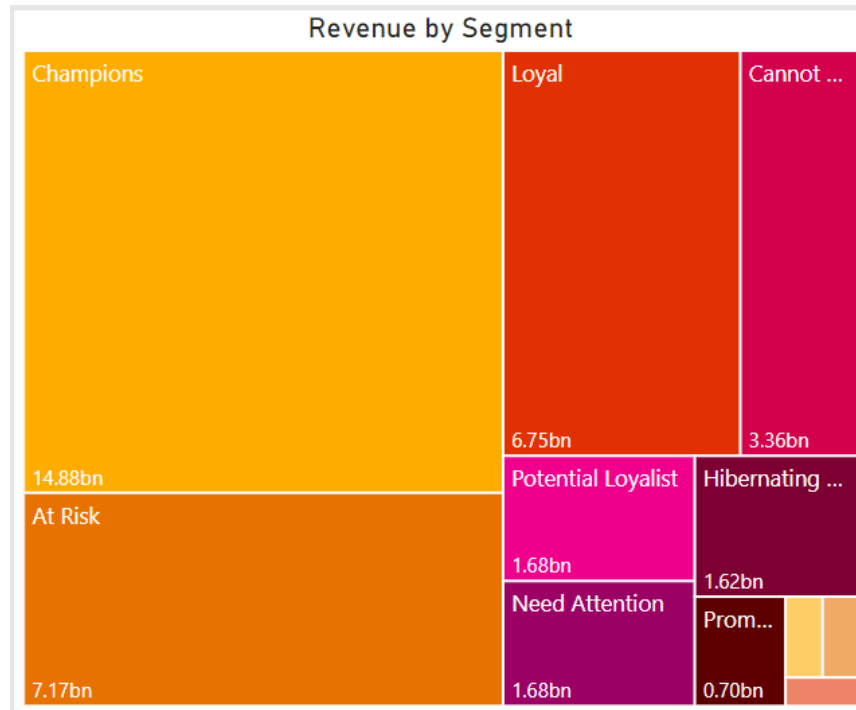
Customer Segment

3 nhóm Khách hàng tiềm năng của sản:

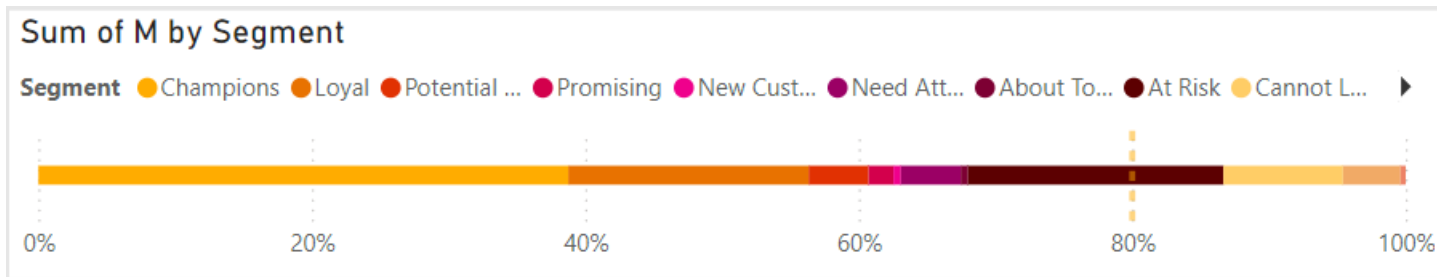
- Champions: Đã mua gần đây, đặt hàng thường xuyên và chi tiêu nhiều nhất.
- Loyal : Đặt hàng thường xuyên. Có tương tác các chương trình khuyến mãi.
- Potential Loyalist

Ngoài ra:

- Can't lose them (Thực hiện các đơn đặt hàng lớn nhất và thường xuyên. Nhưng đã lâu không trở lại)
- At Risk (khách hàng có nguy cơ rời bỏ)
=> chiếm doanh thu cao chỉ sau 2 nhóm trên

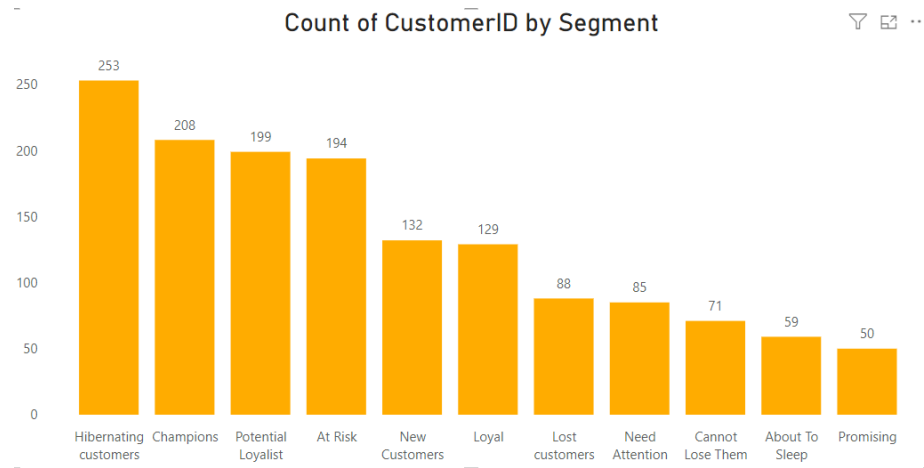


Customer Segment



Nhóm best customers: Champion, Loyal, Potential Loyalist chưa thể đem lại doanh thu kỳ vọng trên tổng số doanh thu, để mà đạt được mức 80% doanh thu thì cần thêm các nhóm khách hàng: At risk, promising => nhóm khách hàng rời bỏ tăng

Hibernating customers: 253 khách hàng
=> Tiềm năng quay trở lại



Retention

% of retained customers in lifetime month												
first_month	0	1	2	3	4	5	6	7	8	9	10	11
1	100%	6%	185%	142%	68%	191%	80%	134%	49%	122%	71%	170%
2	100%	7%	129%	178%	106%	129%	86%	79%	80%	92%	145%	
3	100%	10%	194%	71%	128%	103%	67%	100%	68%	127%		
4	100%	9%	171%	100%	75%	83%	67%	160%	75%			
5	100%	11%	75%	144%	77%	130%	108%	57%				
6	100%	15%	110%	55%	92%	127%	79%					
7	100%	14%	31%	150%	183%	82%						
8	100%	10%	107%	67%	80%							
9	100%	8%	50%	67%								
10	100%	7%	67%									
11	100%	10%										
12	100%											
		10%	12%	12%	12%	15%	12%	14%	10%	12%	12%	16%

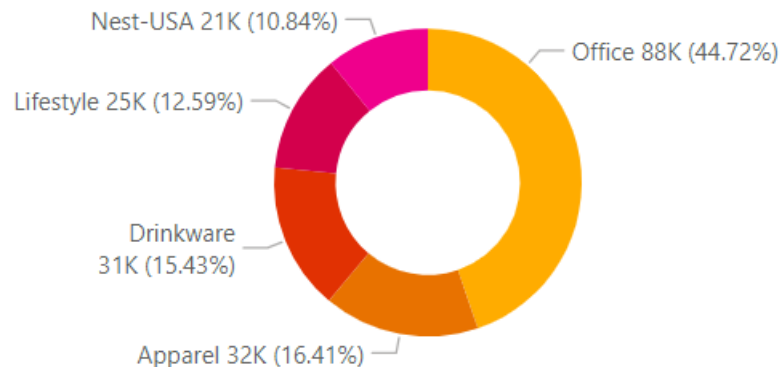
Tỷ lệ khách hàng quay trở lại theo các tháng (tính theo tháng trước) là cao, có những tháng tỷ lệ trên 100%: tháng 2,3,5

Tỷ lệ khách hàng mua hàng trên tổng số khách hàng của các tháng đang rất thấp: dưới 20%

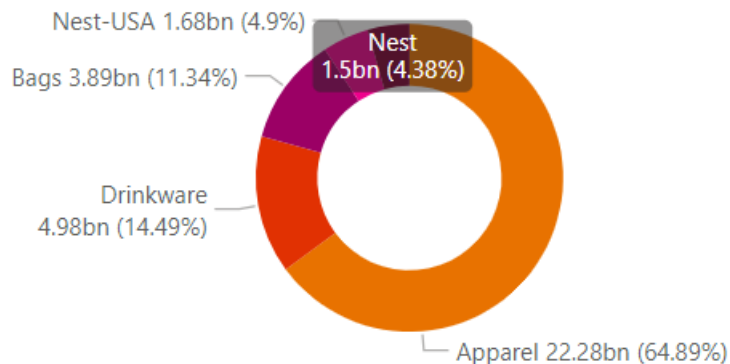
Product

- Nhóm Sản phẩm bán chạy nhất là Office, thứ 2 là Apparel, thấp nhất Nest-USA
 - Doanh thu cao nhất thuộc về nhóm sản phẩm Apparel, thấp nhất là Nest
- => Doanh thu và số lượng sản phẩm bán ra không tỷ lệ thuận với nhau

Quantity by Product_Category



Revenue by Product_Category



Product

Giá trị trung bình trên một đầu đơn: 161,36K/đơn hàng

Các nhóm sản phẩm mà có AOV cao hơn mức TB của sàn:

Apparel, Nest, Waze, Bags...

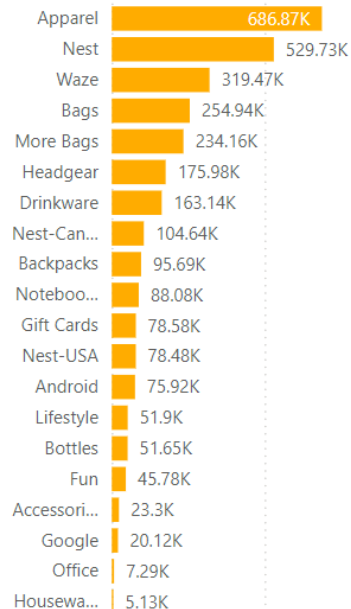
=> sản phẩm có hiệu quả bán cao

Các nhóm sản phẩm mà có AOV thấp: Housewares, Office, Google Accessories

=> hiệu quả bán thấp

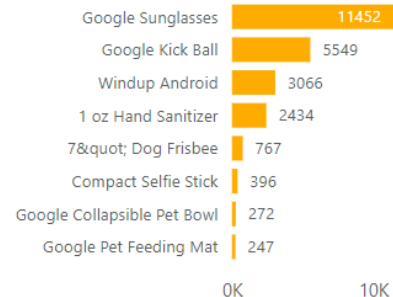
Office, Lifestyle bán chạy nhất nhưng hiệu quả bán thấp

AOV by Product_Category



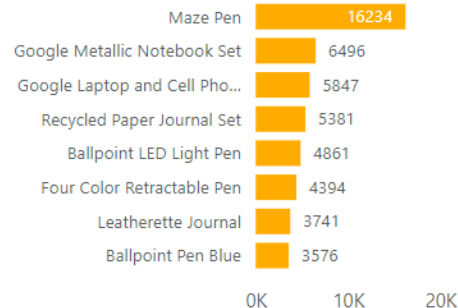
Lifestyle

Sum of Quantity by Product_Description



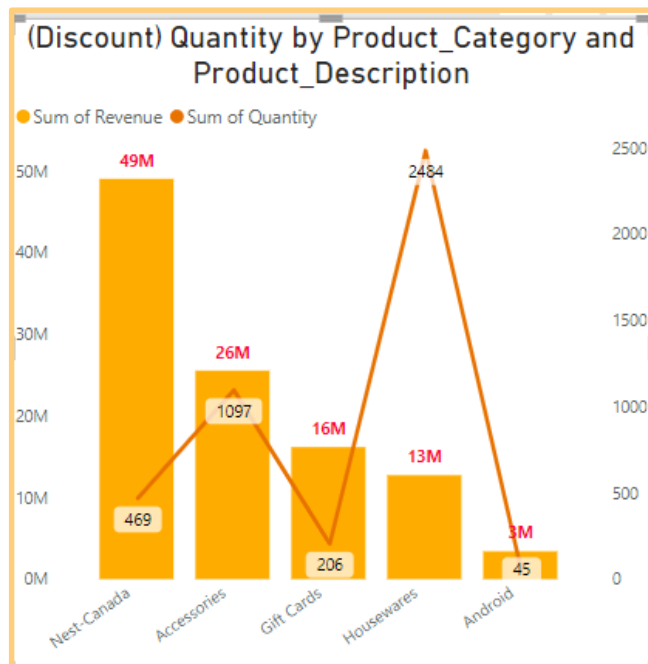
Office

Sum of Quantity by Product_Description

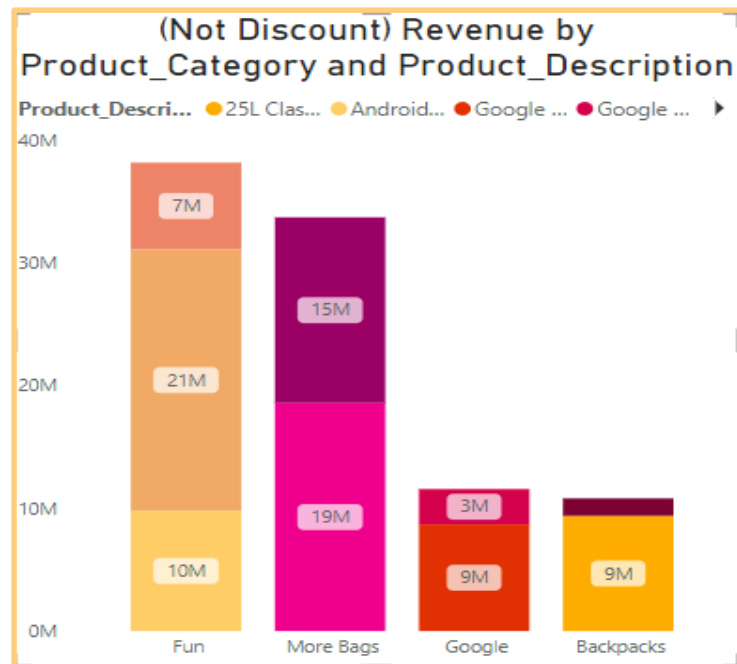


Product

Các nhóm sản phẩm có Discount nhưng mang lại doanh thu thấp: Housewares, Android, Gift Cards



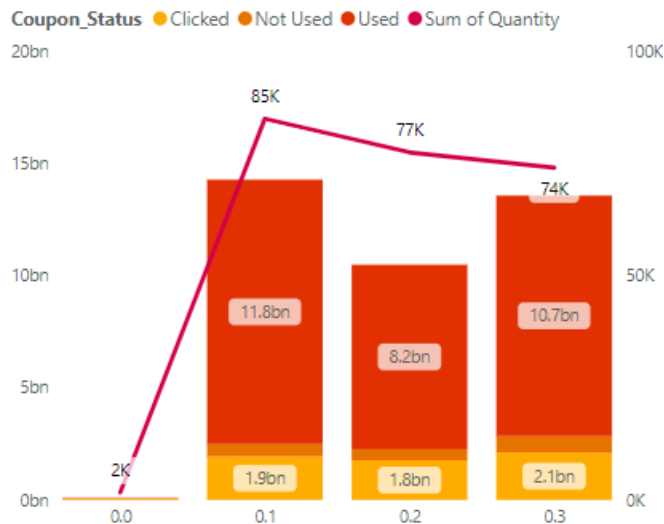
Nhóm sản phẩm không có Discount và doanh thu thấp: Google, Backpacks



Discount/Delivery

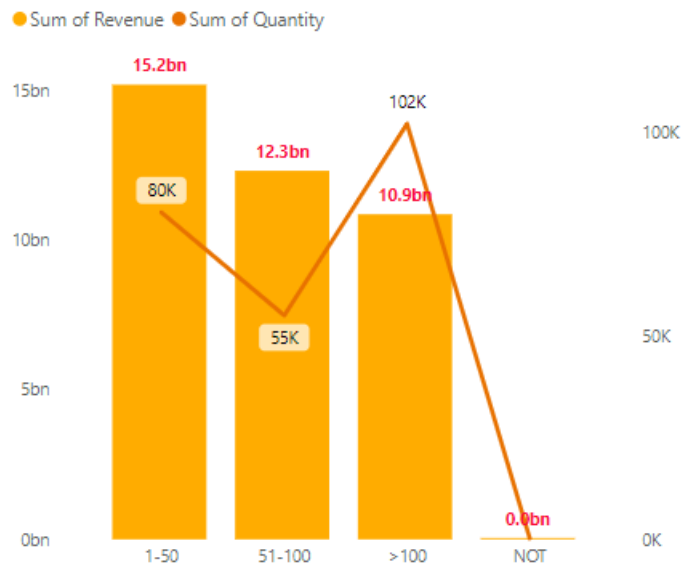
Các sản phẩm được Discount 10% mang lại doanh thu cao nhất, trạng thái có sử dụng đem lại doanh thu cao cho mỗi giá trị Discount

Revenue and Quantity by Discount_pct and Coupon_Status



Phí vận chuyển 1-50 đem lại doanh thu cao nhất
Mức free phí vận chuyển thấp

Revenue and Quantity by Delivery_Group



■ 5. Kết luận/Insights

Kết luận



Doanh thu mà Khách hàng đem lại cao nhất vào các tháng đầu năm và tháng 12, các tháng giữa năm và gần cuối có doanh thu thấp, tháng 8 có số lượng sản phẩm bán ra là cao nhất



Nhóm khách hàng trung thành: Potential Loyalist chiếm doanh thu chưa cao, các nhóm Can't lose them, At Risk chiếm doanh thu cao hơn => số khách hàng có nguy cơ rời bỏ tăng cao



**Nhóm Best Customers chưa đem lại doanh thu đáng kể cho sàn, phụ thuộc vào các nhóm khách hàng có nguy cơ rời bỏ
Tuy nhiên nhóm khách hàng tiềm năng chiếm đa số**

5. Kết luận/Insights

Kết luận

- Tỷ lệ khách hàng quay trở lại theo các tháng trước cao, hầu hết trên 60%. Nhưng tỷ lệ khách hàng quay trở lại trong các tháng trên tổng số khách hàng thấp
- Office, Lifestyle bán chạy nhất nhưng hiệu quả bán thấp (giá trị AOV thấp), các sản phẩm được khách hàng ưa chuộng trong 2 mặt hàng này: Google Sunglasses, Maze Pen
- Có những sản phẩm có Discount nhưng mang lại doanh thu thấp như: Housewares, Android, Gift Cards
Có những sản phẩm không có Discount nhưng mang lại doanh thu cao như: Fun, More Bags
- Các sản phẩm được Discount 10% mang lại doanh thu cao nhất, trạng thái có sử dụng đem lại doanh thu cao cho mỗi giá trị Discount => Hoạt động marketing tốt
Phí vận chuyển thấp nhất lại đem lại doanh thu cao nhất

5. Kết luận/Insights



Gia tăng nguồn cung, quản lý tồn kho hiệu quả để đáp ứng nhu cầu tăng cao vào các tháng đầu năm và tháng 12.



Thực hiện các chiến dịch marketing và khuyến mãi hấp dẫn để thu hút khách hàng và kích thích tiêu dùng vào các tháng mùa thấp, và thu hút khách hàng tiềm năng tiếp tục mua hàng



Cân nhắc mở rộng sang các sản phẩm mới phù hợp với nhu cầu của khách hàng trong các tháng mùa thấp.



Thiết kế các chương trình khách hàng thân thiết, ưu đãi dành riêng cho các khách hàng lâu năm.



Upsell các sản phẩm giá trị cao => Thu hút **Potential Loyalist**
Xin ý kiến đánh giá từ họ và thực hiện các chiến dịch gia tăng gắn kết cung cấp chương trình thành viên, chương trình khách hàng thân thiết hoặc đề xuất các sản phẩm liên quan để nâng cấp họ thành khách hàng Champions.

■ 5. Kết luận/Insights



Kết hợp bán theo set các sản phẩm không bán chạy và bán chạy (có Discount)

=> Thúc đẩy doanh thu của nhóm sản phẩm có Discount, tránh trường hợp Hàng tồn kho ứ đọng



Nếu các sản phẩm được Discount 10% mang lại doanh thu cao nhất, hãy xem xét mở rộng danh mục các sản phẩm có mức giảm giá này => Xem xét áp dụng Discount cho các nhóm sản phẩm không bán chạy



Phí vận chuyển thấp nhất lại mang lại doanh thu cao nhất, hãy xem xét duy trì hoặc thậm chí giảm thêm phí vận chuyển.

THANK YOU FOR LISTENING

Thông tin liên hệ

SDT: 0335496272

Email: ngochan3009.work@gmail.com