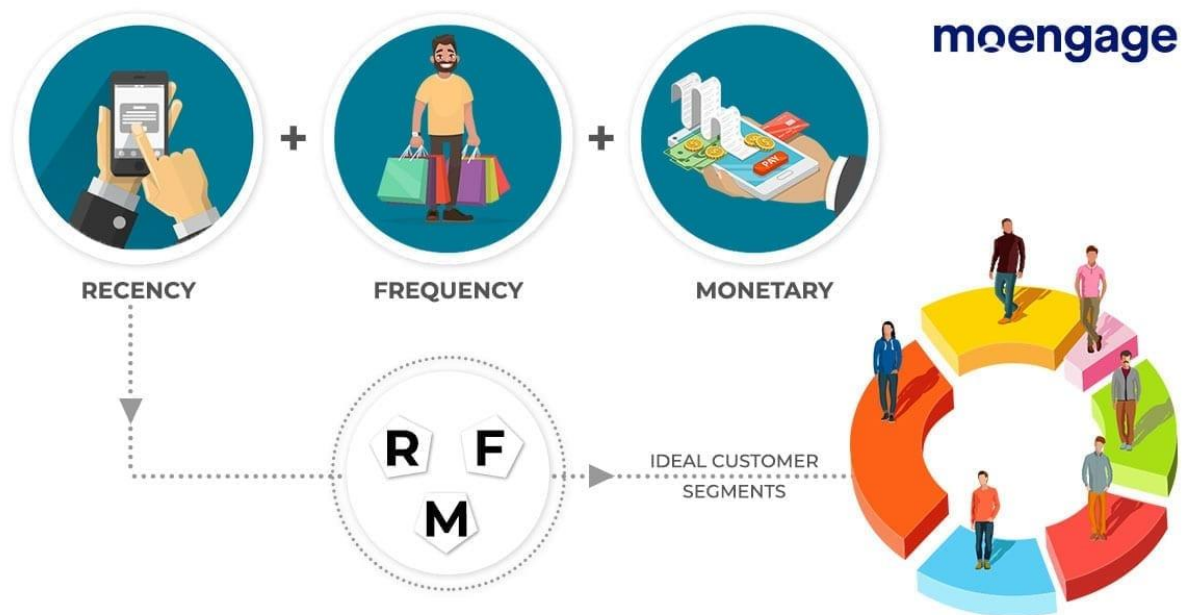


CUSTOMER SEGMENTATION ANALYSIS

USING RFM MODEL

Tác giả: Hỏa Ngọc Tú.

****Lưu ý:** Xin lưu ý rằng tập dữ liệu thu được chỉ có thể được sử dụng cho mục đích nghiên cứu và mô phỏng, và không mang giá trị thương mại. Có thể xuất hiện những sai sót; xin vui lòng thông cảm và bỏ qua những thiếu sót đó.*



MỤC LỤC

Chương 1: Giới Thiệu.....	3
1.1. Định nghĩa.....	3
1.1.1. Phân khúc khách hàng.....	3
1.1.2. Phân đoạn khách hàng bằng phương pháp phân tích RFM.....	3
1.2. Tập dữ liệu	4
Chương 2: Xây dựng mô hình RFM	4
2.1. Tiền xử lý dữ liệu.....	4
2.2 Trích xuất tính năng.	5
2.3 Mô hình hóa với Thuật toán KMeans	6
Chương 3: Phân tích những hiểu biết từ RFM	7
3.1 phân phối.....	7
3.2 Hiệu quả kinh doanh	7
Chương 4: Đề xuất giải pháp.	9
4.1 Giải pháp cho mỗi nhóm khách hàng.....	9
4.2 Đối với hoạt động kinh doanh.....	9

Chương 1: Giới Thiệu

1.1. Định nghĩa

1.1.1. Phân khúc khách hàng

Phân đoạn khách hàng là một phương pháp chia nhóm hoặc phân cụm khách hàng dựa trên các đặc điểm chung. Nhà nghiên cứu thị trường có thể phân đoạn khách hàng trong mô hình B2C bằng cách sử dụng các đặc điểm dân số học của khách hàng như nghề nghiệp, giới tính, tuổi tác, địa điểm và tình trạng hôn nhân. Các đặc điểm tâm lý học như giai cấp xã hội, lối sống và tính cách cũng như các đặc điểm hành vi như chi tiêu, thói quen tiêu dùng, sử dụng sản phẩm/dịch vụ và các sản phẩm đã mua trước đó. Trong mô hình B2B, sử dụng các đặc điểm của các công ty như kích thước của công ty, loại ngành nghề và địa điểm.

1.1.2. Phân đoạn khách hàng bằng phương pháp phân tích RFM.

Phân tích RFM (Recency, Frequency, Monetary) là một phương pháp dựa trên hành vi để phân nhóm khách hàng thành các đoạn khác nhau. Nó phân nhóm khách hàng dựa trên các giao dịch mua hàng trước đó của họ. Đặc điểm này xác định xem khách hàng mua hàng gần đây, thường xuyên như thế nào và chi tiêu bao nhiêu.

- Recency (R): Ai đã mua hàng gần đây? Số ngày kể từ lần mua hàng cuối cùng (ít recency nhất).
- Frequency (F): Ai đã mua hàng thường xuyên? Điều này có nghĩa là tổng số lần mua hàng (tần suất cao).
- Giá trị tiền (M): Ai có số tiền mua hàng cao? Điều này có nghĩa là tổng số tiền khách hàng đã chi tiêu (giá trị tiền cao).

Trong báo cáo này, mỗi trong ba biến số (Recency, Frequency và Monetary) được chia thành năm nhóm bằng nhau, tạo ra 125 (5x5x5) đoạn khách hàng khác nhau và trong trường hợp này, tôi sẽ phân chia thành 11 nhóm lớn của phân đoạn khách hàng (chi tiết trong Chương 2). Mô hình RFM gán một điểm từ 1 đến 5 cho mỗi tiêu chí. Recency gán 1 điểm cho khoảng thời gian dài nhất kể từ lần cuối cùng một người tiêu dùng sử dụng dịch vụ, và 5 điểm cho khoảng thời gian ngắn nhất. Càng cao điểm số cho Frequency và Monetary, càng tốt cho mật độ sử dụng dịch vụ và tổng thu nhập.

1.2. Tập dữ liệu

Tập dữ liệu gồm 541,909 bản ghi được tạo bởi khách hàng từ 38 quốc gia. Thu thập từ tháng 12 năm 2010 đến năm 2011 với tổng doanh số bán hàng trên 9.7 triệu đô và khoảng 4070 sản phẩm.

Chương 2: Xây dựng mô hình RFM

2.1. Tiền xử lý dữ liệu

Bảng dữ liệu mà tôi sử dụng có 8 đặc điểm về thông tin giao dịch của khách hàng và quốc gia của họ.

```
-----Data info-----
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 541909 entries, 0 to 541908
Data columns (total 8 columns):
#   Column          Non-Null Count  Dtype
---  -
0   InvoiceNo        541909 non-null object
1   StockCode       541909 non-null object
2   Description     540455 non-null object
3   Quantity       541909 non-null int64
4   InvoiceDate     541909 non-null object
5   UnitPrice      541909 non-null float64
6   CustomerID     406829 non-null float64
7   Country        541909 non-null object
dtypes: float64(2), int64(1), object(5)
memory usage: 33.1+ MB
None
```

Nhìn chung, tất cả các tính năng đều phù hợp với kiểu dữ liệu phù hợp và có một số bản ghi trùng lặp. Trên hết, chênh lệch Mô tả và ID khách hàng có số lượng ít hơn do có giá trị NULL.

```
-----Data types-----
object      5
float64     2
int64       1
dtype: int64

-----Missing values-----
CustomerID  135080
Description  1454
dtype: int64

-----Duplicate values-----
5268
```

Để có thông tin chi tiết hơn, chúng ta có thể thấy rằng có các giá trị âm trong ngữ cảnh bán hàng. Sau khi kiểm tra khả năng trả lại, hoàn tiền và chiết khấu, không có dấu vết cho những trường hợp đó và kết quả là có dữ liệu sai lệch cần phải loại bỏ

khỏi tập dữ liệu. Sau khi tổng hợp các lỗi, việc tiền xử lý dữ liệu được thực hiện như sau:

- Loại bỏ các giá trị trùng lặp và âm.
- Loại bỏ các giá trị NULL trong các biến Mô tả và Mã khách hàng.
- Làm sạch ngoại lai.

2.2 Trích xuất tính năng.

Ở phần này tôi tạo ra 2 tính năng mới là Revenue. Revenue tính toán dựa vào 2 cột UnitPrice và Quantity.

```
#Tạo tính năng mới Doanh thu
data["Revenue"] = data["Quantity"] * data["UnitPrice"]
```

Như định nghĩa đã cho ở trên, tôi tính toán các phần tử cho mô hình RFM như sau:

- Recency : được tính bằng số ngày tối đa trong tập dữ liệu cộng với 1 ngày, do đó thời gian ngắn nhất thời gian mua hàng cho đến ngày hiện tại sẽ là 1 đơn vị.
- Frequency : được tính bằng cách đếm số lượng hoá đơn của từng khách hàng.
- Monetary values: được tính bằng tổng số tiền thanh toán của từng khách hàng.

Báo cáo này giới thiệu một thuộc tính mới gọi là thời gian giữa các lần mua (Inter-Purchase Time) vào mô hình RFM (Recency, Frequency và Monetary) truyền thống. Thuộc tính mới sẽ bao gồm hành vi mua hàng động của khách hàng trong chu kỳ mua sắm của họ.

Biến thứ tư của RFM, thời gian giữa các lần mua (Interpurchase Time), là một chỉ số đo khoảng thời gian trung bình giữa các lần mua hàng của một khách hàng. Thời gian giữa các lần mua được tính như sau:

$$T = L/(F-1) = (T_n - T_1)/(F-1)$$

T --> Thời Gian Giữa Các Lần Mua

L --> Chu kỳ mua sắm

F --> Tần suất

T1 --> Lần mua đầu tiên

Tn --> Lần mua cuối cùng

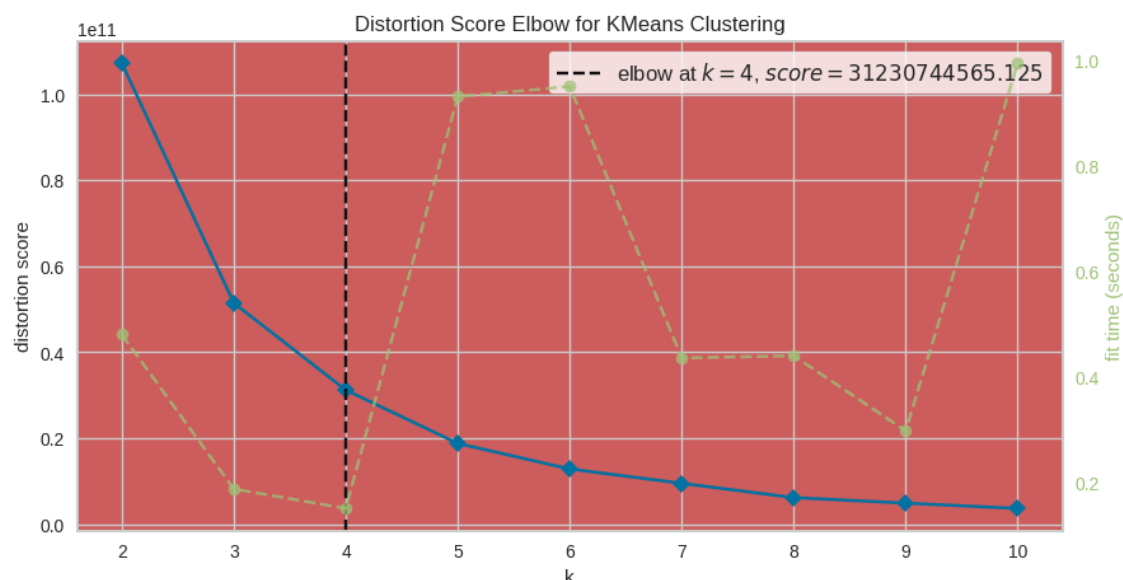
Sau cùng sẽ có RFMT như sau:

```
RFM["Interpurchase_Time"] = RFM["Shopping_Cycle"] // RFM["Frequency"]
RFMT = RFM[["Recency", "Frequency", "Monetary", "Interpurchase_Time"]]
RFMT.head().style.set_properties(**{"background-color": "#cd5c5c", "color": "black", "border-color": "black"})
```

	Recency	Frequency	Monetary	Interpurchase_Time
CustomerID				
12347.000000	2	7	4310.000000	52
12348.000000	75	4	1797.240000	70
12352.000000	36	8	2506.040000	32
12356.000000	22	3	2811.430000	100
12358.000000	1	2	1168.060000	74

2.3 Mô hình hóa với Thuật toán KMeans

Sau khi tính điểm RFM, tôi sử dụng phân cụm k-means để nhóm khách hàng dựa trên các số liệu này. Phân cụm k-means là một thuật toán học máy không giám sát, nhóm các điểm dữ liệu thành k cụm dựa trên độ giống nhau của chúng. Trong trường hợp này, k đại diện cho số lượng phân khúc khách hàng mà chúng tôi muốn tạo. Tôi sử dụng phương pháp khuỷu tay để xác định số cụm tối ưu cho dữ liệu của mình. Dưới đây là tập lệnh tôi sử dụng để thực hiện phân cụm k-means trên dữ liệu của mình. Trong trường hợp của tôi, tôi chọn bốn cụm làm số cụm tối ưu. Tôi điều chỉnh thuật toán phân cụm k-mean cho dữ liệu của mình và chỉ định từng khách hàng vào một cụm.

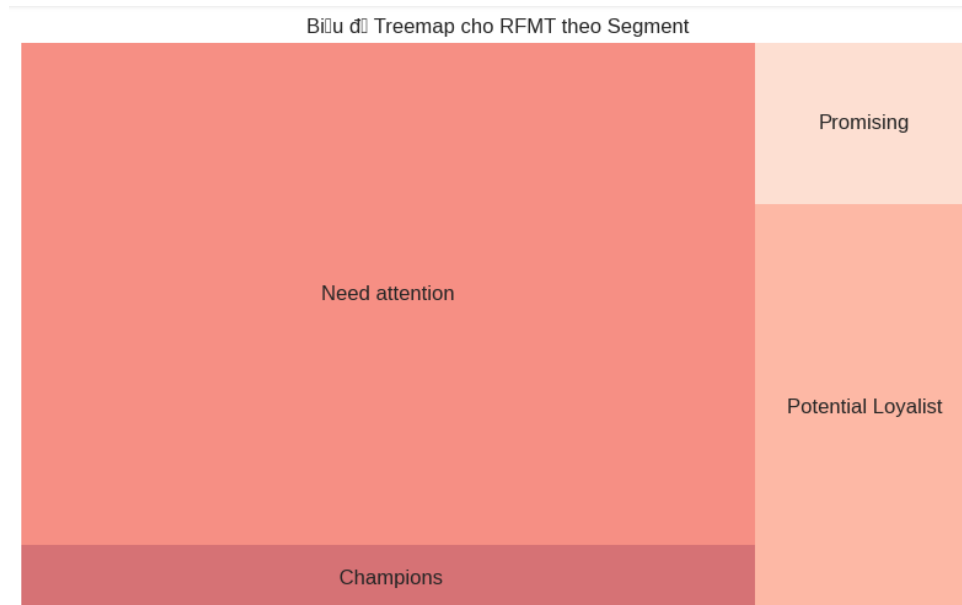


```
[ ] RFMT["Clusters"]=kmeans.labels_
RFMT.head().style.set_properties(**{"background-color": "#cd5c5c", "color": "black", "border-color": "black"})
```

	Recency	Frequency	Monetary	Interpurchase_Time	Clusters
CustomerID					
12347.000000	2	7	4310.000000	52	0
12348.000000	75	4	1797.240000	70	0
12352.000000	36	8	2506.040000	32	0
12356.000000	22	3	2811.430000	100	0
12358.000000	1	2	1168.060000	74	0

Chương 3: Phân tích những hiểu biết từ RFM

3.1 phân phối

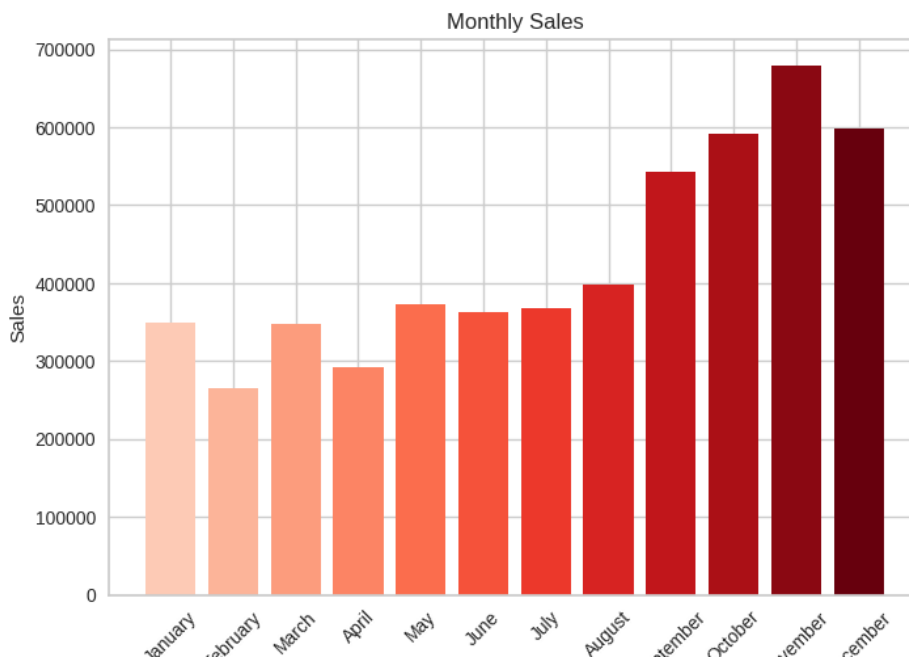
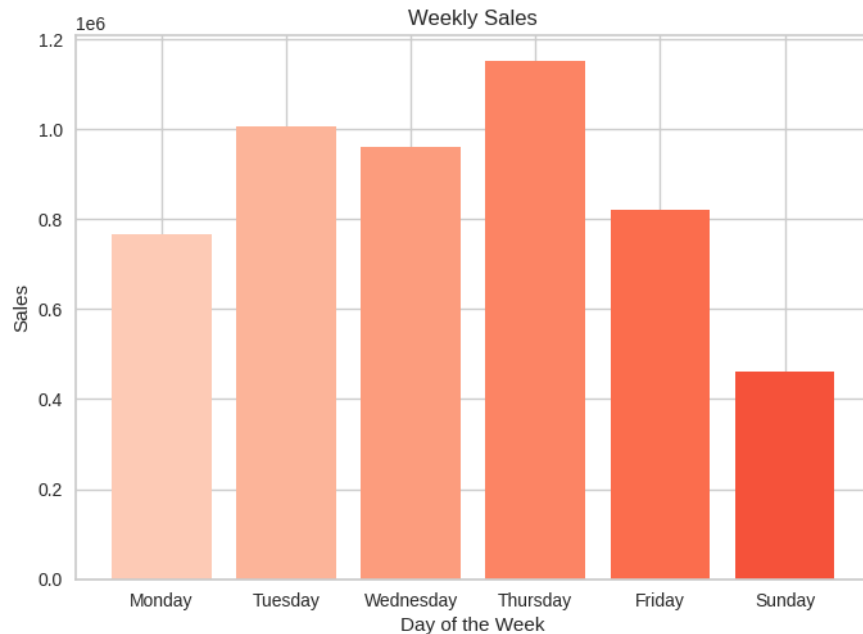


Duy trì nhóm champions và chăm sóc nhiều hơn các nhóm đang trong quá trình chuyển đổi như Need attention, promising.

Nhóm champions là rất quan trọng đối với công ty, chiếm tỷ lệ 18,53% và đóng góp doanh thu cao nhất cho doanh nghiệp. Ngoài ra, các nhóm khác đang trong quá trình chuyển đổi như Need attention, chúng ta cần chú ý hơn đến họ để chuyển họ sang nhóm Potential Loyalist và champions.

3.2 Hiệu quả kinh doanh

Phân tích chuỗi thời gian là phương pháp phân tích tập hợp các điểm dữ liệu trong một khoảng thời gian của thời gian. Thay vì ghi lại các điểm dữ liệu không liên tục hoặc ngẫu nhiên, chuỗi thời gian các nhà phân tích ghi lại các điểm dữ liệu theo những khoảng thời gian nhất quán trong một khoảng thời gian nhất định. Dưới đây là biểu đồ hiển thị hiệu suất bán hàng thông qua các danh mục khác nhau bao gồm hàng tuần, hàng tháng và hàng quý.



Nhìn chung, chúng ta có thể thấy rằng công ty đạt mức cao nhất vào các ngày trong tuần (Thứ Hai đến thứ Sáu). Ngoài ra, những tháng bận rộn trong năm bắt đầu từ tháng 9 đến tháng 12 là quý 4 với doanh thu chiếm gần như 37% trong tổng tỷ lệ phần trăm trong năm.

Mặt khác, số lượng bán ra thấp thường xuyên giảm gần gấp đôi vào thời điểm cuối tuần là 700.000 đô la và tăng dần vào những ngày tiếp theo. Trên hết, kết quả hoạt động cũng không nổi bật trong 3 quý đầu năm, tăng giảm nhiều lần trong khoảng từ 1,3 đến 2 triệu đô la và rõ ràng là dưới mức doanh thu trung bình 2,23 triệu đô la.

Chương 4: Đề xuất giải pháp.

4.1 Giải pháp cho mỗi nhóm khách hàng

Phân khúc	Giải pháp
Need Attention	<ul style="list-style-type: none"> Thực hiện các ưu đãi có giới hạn thời gian. Đưa ra các khuyến nghị được cá nhân hóa.
Champions	<ul style="list-style-type: none"> Tặng quà cho họ. Có thể là người trải nghiệm sớm các sản phẩm mới. Những khách hàng này đã được chứng minh là có mức sẵn sàng chi trả cao hơn, vì vậy đừng sử dụng giá chiết khấu để tăng doanh số bán hàng.
Potential Loyalists	<ul style="list-style-type: none"> Cung cấp chương trình thành viên/khách hàng thân thiết. Đưa ra các khuyến nghị được cá nhân hóa.
Promising	<ul style="list-style-type: none"> Cung cấp phiếu giảm giá. Đưa họ trở lại diễn đàn và thu hút họ tham gia. Đưa ra các khuyến nghị được cá nhân hóa.

4.2 Đối với hoạt động kinh doanh.

❖ Hiệu suất thấp vào ngày thường

Tận dụng chiết khấu cho hàng tồn kho cũ vào cuối tuần:

Tận dụng triệt để tư duy giảm giá và bán bớt hàng tồn kho để lưu kho lâu dài hoặc nhu cầu cực kỳ thấp để thu hút khách hàng và đạt được nhiều doanh thu và sự công nhận hơn vào cuối tuần.

Chiến dịch tiếp thị mục tiêu:

Tạo các chiến dịch tiếp thị có mục tiêu để giải quyết cụ thể các ngày trong tuần. Lời đề nghị chương trình khuyến mãi, giảm giá hoặc khuyến khích cụ thể vào các ngày trong tuần để thu hút khách hàng trong thời gian những khoảng thời gian đó. Sử dụng tiếp thị qua email, phương tiện truyền thông xã hội và các kênh quảng cáo khác để tiếp cận đối tượng mục tiêu một cách hiệu quả.

Mở rộng cung cấp sản phẩm:

Xem xét việc giới thiệu các sản phẩm hoặc dịch vụ mới hấp dẫn cụ thể trong thời gian các ngày trong tuần. Ví dụ: chúng tôi có thể cung cấp bữa trưa đặc biệt hoặc giới thiệu thời gian giới hạn các mặt hàng chỉ có trong tuần. Đa dạng hóa dịch vụ của chúng tôi có thể thu hút khách hàng mới và khuyến khích những người hiện có đến thăm vào các ngày trong tuần.

Chương trình khách hàng thân thiết:

Triển khai chương trình khách hàng thân thiết để thưởng cho những người mua sắm thường xuyên vào các ngày trong tuần. Cung cấp giảm giá độc quyền, quyền truy cập sớm vào các chương trình khuyến mãi hoặc lợi ích bổ sung cho khách hàng trung thành khách hàng. Điều này có thể khuyến khích khách hàng lựa chọn doanh nghiệp của chúng tôi vào các ngày trong tuần.

❖ Tận dụng khối lượng tiềm năng của ngày Thứ Năm và các hoạt động trong thời gian vàng giờ.

Ưu đãi đặc biệt cho các sản phẩm trong khung thời gian:

Tùy thuộc vào số lượng giờ cụ thể lớn trong ngày, chúng tôi có thể thực hiện giảm giá và ưu đãi giới hạn trong khoảng thời gian 2 hoặc 3 giờ đối với một số sản phẩm cần bán thêm hoặc được khách hàng quan tâm nhất. Những điều đó sẽ giúp giải phóng hàng tồn kho của chúng tôi và tiếp cận nhiều khách hàng hơn do thói quen của họ.

Hoạt động phát trực tiếp vào thứ năm:

Không còn nghi ngờ gì nữa, khách hàng của chúng ta ngày nay thích xem mua sắm trực tuyến theo thứ tự để chứng kiến sản phẩm thực tế và nhận những ưu đãi đặc biệt của bất kỳ sản phẩm nào. Tận dụng điều đó, chúng tôi nên nỗ lực phát trực tiếp ở dạng phẳng có thể tiếp cận không chỉ cũ mà còn khách hàng mới vào ngày có doanh số cao này.

❖ Tăng cường khai thác thị trường Anh và Đức.

Dịch vụ riêng biệt và các cơ quan tại chỗ cho các quốc gia đó:

Vì khách hàng ở những quốc gia đó đóng góp lượng doanh thu và tác động rất lớn. Cần đảm bảo các dịch vụ khách hàng dành cho họ là tốt nhất và mở rộng phạm vi sự hiện diện của công ty trong khu vực của họ sẽ là một cơ hội tuyệt vời để giao tiếp và nghiên cứu sâu hơn cho cộng đồng của họ.

Phát triển thêm nhiều cơ sở sản xuất thực tế ở những quốc gia đó:

Điều này sẽ cải thiện sự nhận biết của thương hiệu và tạo ra sự thoải mái hơn cho người già và những khách hàng mới. Trên hết, chúng tôi có thể nâng cao khả năng cạnh tranh của mình tại các thị trường đó và tạo dựng niềm tin cho khách hàng tiềm năng trong tương lai.

HẾT.