

Bài 16 - Model RFM phân khúc khách hàng

08 Nov 2019 - phamdinhkhanh

Menu

- I. Mô hình RFM (Recency - Frequency - Monetary model)
 - Tại sao cần chuẩn hóa dữ liệu đầu vào?
 - Tác dụng của chuẩn hóa dữ liệu
 - Phương pháp chuẩn hóa dữ liệu
 - Tính rank cho một khách hàng
 - Thực hiện customer segmentation
 - Đổi dấu để đảm bảo tính đồng biến với rank khách hàng
- II. Thực hành xây dựng mô hình RFM
- 3. Các biến thể khác của RFM
- 4. Tổng kết
- 5. Tài liệu tham khảo

I. Mô hình RFM (Recency - Frequency - Monetary model)

Theo nguyên lý pareto 20% khách hàng sẽ mang lại 80% doanh số. Do đó doanh nghiệp cần xác định được những khách hàng quan trọng nhất để chăm sóc đặc biệt. Những tập khách hàng này được gọi là VIP, Priority hoặc khách hàng cao cấp, tùy từng doanh nghiệp có cách gọi khác nhau. Việc phân chia khách hàng thành các nhóm khác nhau dựa trên nhu cầu mua sắm sẽ giúp doanh nghiệp kinh doanh hiệu quả hơn, marketing target đến đúng tập khách hàng hơn, và khách hàng được phục vụ tốt hơn. Để minh chứng cho nhận định trên, tôi sẽ lấy ra một vài ví dụ đơn giản:

- Giả sử doanh nghiệp của bạn kinh doanh rất nhiều các mặt hàng từ cái bút bi đến xe hơi. Bạn không thể nói rằng doanh nghiệp mình đã chăm sóc tốt khách hàng nếu như doanh nghiệp đó đối xử với những khách hàng mua một chiếc xe hơi ngang bằng với khách hàng mua một chiếc bút bi.
- Bạn cũng không thể suggest một chiếc tivi 50 triệu cho một người lương khoảng 5 triệu trong khi hàng tháng anh ta phải trả tiền nhà trọ 2 triệu và tiền ăn 2 triệu. Trong trường hợp lý tưởng (không ốm đau đột xuất, không hiếu hỉ, du lịch cơ quan, bạn bè), mỗi tháng anh ta tiết kiệm được 1 triệu và cần tới 4 năm tiết kiệm để mua được chiếc tivi đó. Bạn chỉ nên suggest chiếc nào rẻ rẻ thôi, 5 triệu chẳng hạn.
- Và trái lại với một khách hàng thu nhập vài trăm triệu/tháng và cũng muốn mua tivi. Bạn cũng không nên suggest 1 chiếc tivi 5 triệu nếu không muốn khách hàng nổi giận vì tội sỉ nhục khách hàng. Thà ông nói cái tivi 5 triệu là 20 triệu tôi còn vui, ông nghĩ gì mà lại đi bán một chiếc tivi 5 triệu cho tôi? đại khái thế.
- Bạn cũng không nên suggest một người mua iphone một chiếc tai nghe samsung hay một đĩa trẻ rươi bia và thuốc lá.

Qua các ví dụ cụ thể trên chúng ta có thể thấy:

- Chính sách chăm sóc khách hàng cần phải thay đổi để phù hợp với phân khúc.
- Chăm sóc khách hàng cần phải phù hợp với thu nhập. Top
- Chăm sóc khách hàng đồng thời cũng phải phù hợp với nhu cầu, thị hiếu của khách hàng.

Như vậy việc phân khúc khách hàng mang lại rất nhiều các lợi ích cho doanh nghiệp.

Chính vì thế các tập công ty, tập đoàn lớn trong lĩnh vực tài chính-ngân hàng, công nghệ thông tin, du lịch, dịch vụ, vận tải,... đều có những phương án phân chia tập khách hàng của riêng họ và đồng thời phát triển mô hình kinh doanh, bộ máy vận hành của mình định hướng theo các tập khách hàng đó.

Chẳng hạn như tại ngân hàng có thể tìm cách chia khách hàng thành Mass Affluent (khách hàng đại chúng), Affluent (khách hàng khả giả), Priority (khách hàng cao cấp). Một số phòng ban chiến lược được thành lập chỉ để chăm sóc cho một nhóm khách hàng cao cấp hơn như bộ phận chăm sóc khách hàng priority. Tại những bộ phận này, nhân viên tư vấn đòi hỏi phải có kỹ năng cao hơn, ngoại hình ưa nhìn hơn, bàn ghế phục vụ tất nhiên cũng sang trọng hơn. Hoặc phân khúc khách hàng theo thành loại hình bundle (bán thêm sản phẩm), not-bundle (khách hàng mua sản phẩm lần đầu). Trong chứng khoán có thể chia tập khách hàng theo khách hàng thứ cấp, khách hàng thông thường và khách hàng VIP dựa trên giá trị NAV (net asset value) của khách hàng.

Và còn rất nhiều các phương pháp phân chia phân khúc khách hàng khác nữa dựa trên giá trị sử dụng các sản phẩm, dịch vụ, dựa trên loại hình sản phẩm hoặc dựa trên đối tượng khách hàng là cá nhân hay doanh nghiệp.

Có nhiều cách phân chia phân khúc khách hàng khác nhau mà doanh nghiệp có thể áp dụng. Việc phát triển kinh doanh theo phân khúc khách hàng luôn mang lại lợi ích cho doanh nghiệp vì các khách hàng có hành vi tương đồng được gom lại để phát triển các sản phẩm, dịch vụ đi kèm phù hợp với nhu cầu của họ.

Okie, đến đây bạn đã hiểu được lợi ích của phân chia phân khúc khách hàng. Tiếp theo là sử dụng phương pháp nào để phân chia phân khúc? Đó là lý do tôi viết bài hướng dẫn này.

Phương pháp mà tôi giới thiệu trong bài viết này có tên là RFM, một trong những phương pháp điển hình được sử dụng trong phân khúc khách hàng. Định nghĩa của phương pháp có thể tìm tại wikipedia như sau:

"RFM is a method used for analyzing customer value. It is commonly used in database marketing and direct marketing and has received particular attention in retail and professional services industries. RFM stands for the three dimensions: Recency – How recently did the customer purchase?"

Có nghĩa là:

"RFM là một phương pháp được sử dụng để phân tích giá trị khách hàng. Nó thường được sử dụng trong marketing cơ sở dữ liệu (kiểu như dựa vào dữ liệu về khách hàng để tiếp thị sản phẩm) và marketing trực tiếp và đã nhận được sự chú ý đặc biệt trong ngành bán lẻ và dịch vụ."

RFM định lượng giá trị của một khách hàng dựa trên 3 thông tin chính:

- Recency: Khoảng thời gian mua hàng gần đây nhất là bao lâu. Cho biết khách hàng có đang thực sự hoạt động gần thời điểm đánh giá. Chỉ số này càng lớn càng cho thấy xu hướng rời bỏ của khách hàng càng cao. Đó là một cảnh báo cho doanh nghiệp nên thay đổi sản phẩm để đáp ứng thị hiếu khách hàng hoặc thay đổi chính sách để nâng cao chất lượng phục vụ.
- Frequency: Tần suất mua hàng của khách hàng. Nếu khách hàng mua càng nhiều đơn thì giá trị về doanh số mang lại cho công ty càng cao và tất nhiên giá trị của họ càng lớn. Tuy nhiên nếu chỉ xét dựa trên tần suất mua hàng thì cũng chưa đánh giá được đầy đủ mức độ tác động lên doanh thu bởi bên cạnh đó, giá trị đơn hàng cũng là yếu tố trực tiếp cho thấy khách hàng tiềm năng như thế nào.

Top

- **Monetary:** Là số tiền chi tiêu của khách hàng. Đây là yếu tố trực quan nhất ảnh hưởng tới doanh số. Hay nói cách khác, doanh nghiệp quan tâm nhất là khách hàng đã dành bao nhiêu tiền để mua sản phẩm của công ty? Monetary sẽ tác động trực tiếp tới doanh thu và bị tác động gián tiếp thông qua 2 yếu tố còn lại là Recency và Frequency.

Sau khi đã có đầu vào là 3 nhân tố trên. Chúng ta sẽ sử dụng thuật toán K-Mean clustering, một thuật toán học không giám sát để nhóm các khách hàng có cùng mức độ VIP thành một nhóm. Nhưng trước đó chúng ta cần phải chuẩn hóa dữ liệu đầu vào.

Tại sao cần chuẩn hóa dữ liệu đầu vào?

Hãy tưởng tượng đơn vị của Recency là số ngày, của Frequency là số lần và của Monetary là số tiền. Khó khăn của mô hình học không giám sát là làm thế nào phân cụm chính xác những điểm dữ liệu được tạo thành từ những biến khác nhau về độ đo. Thông thường Monetary sẽ rất lớn so với Recency và Frequency nên trong không gian euclidean, yếu tố khoảng cách giữa các điểm đại diện cho 1 khách hàng sẽ ít bị tác động bởi Recency và Frequency hơn so với Monetary. Nếu không chuẩn hóa dữ liệu sẽ dẫn tới khoảng cách phần lớn bị ảnh hưởng bởi Monetary và ít bị ảnh hưởng bởi 2 biến còn lại.

Tác dụng của chuẩn hóa dữ liệu

- **Chuẩn hóa dữ liệu giúp đồng nhất đơn vị:** Các nhân tố Recency, Frequency, Monetary sẽ không còn khác biệt về đơn vị. Giá trị của nhân tố này sẽ được đưa về cùng một miền biến thiên, phương sai và trung bình. Do đó vai trò của các biến trong việc xác định cụm là bình đẳng.
- **Các cụm sẽ tách biệt nhau hơn:** Trước khi chuẩn hóa dữ liệu, nếu xét theo các trục Recency, Frequency thì các cụm sẽ rất gần nhau do giá trị Recency, Frequency là rất nhỏ. Sau khi chuẩn hóa dữ liệu, đơn vị là đồng nhất nên ranh giới giữa các cụm sẽ trở nên tách biệt.
- **Hạn chế các điểm dữ liệu outliers:** **Outliers** hay còn gọi là điểm dị biệt, là một trong những nguyên nhân khiến thuật toán phân cụm bị **bias** (chệch). Dữ liệu outliers sẽ thường quá lớn hoặc quá nhỏ, điều này khiến khoảng cách giữa các điểm trở nên phân tán. Trong quá trình huấn luyện để thuật toán hội tụ về **centroids** (lõi) của từng **clusters** (cụm), do tác động của các outliers nên các centroids sẽ thường bị lệch cách xa các vị trí hội tụ lõi. Thuật toán cũng cần nhiều thời gian để hội tụ hơn. Việc này gây lãng phí về chi phí tính toán và chi phí thời gian. Do đó chúng ta cần chuẩn hóa dữ liệu để loại bỏ các outliers.

Phương pháp chuẩn hóa dữ liệu

Có rất nhiều phương pháp khác nhau để chuẩn hóa dữ liệu cho các biến Recency, Frequency, Monetary. Một trong những phương pháp được sử dụng phổ biến nhất là chia độ lớn của các biến theo rank. Chẳng hạn Recency, Frequency, Monetary sẽ được chia thành 5 ranks hoặc 10 ranks tùy theo mục tiêu và định nghĩa từ trước về thang đo. Trong trường hợp bạn muốn chấm điểm khách hàng của mình chi tiết hơn, lựa chọn 10 ranks sẽ phù hợp hơn. Trái lại, bạn muốn các nhóm khái quát hơn thì lựa chọn rank 5 sẽ hợp lý hơn. Ngoài ra, các biến cũng nên cùng khoảng rank (cùng là 5 hoặc 10) để đồng nhất mức độ tác động lên rank tổng hợp của khách hàng. Không nên chọn Recency, Frequency, Monetary mỗi biến một khoảng rank.

Tính rank cho một khách hàng

Top

Rank của một khách hàng sẽ được tính bằng trung bình cộng của 3 giá trị rank của Recency, Frequency và Monetary. Giá trị này càng cao thì khách hàng của chúng ta càng có giá trị đối với công ty.

Để mapping được giá trị của một biến cụ thể sang rank, rất đơn giản chúng ta sẽ sử dụng hàm `pandas.qcut()` với giá trị khai báo là số lượng ranks cần chia. Chẳng hạn cần tạo ranks 10 thì truyền vào `pandas.qcut(10)`. Hàm `qcut()` sẽ tự động tìm ra các khoảng giá trị sao cho số lượng các quan sát được chia đều vào mỗi khoảng. Bạn có thể lựa chọn giá trị mapping là một số nguyên bất kỳ ứng với mỗi rank thông qua thiết lập tham số `labels = False` hoặc giữ nguyên `labels` là các khoảng khi `labels = True`. Cụ thể xem hướng dẫn về hàm `pandas.qcut()` (<https://pandas.pydata.org/pandas-docs/version/0.23.4/generated/pandas.qcut.html>).

Lưu ý một số trường hợp bạn sử dụng hàm `pandas.qcut()` và gặp lỗi `bins edge duplicate`. Tức điểm đầu và điểm cuối của các khoảng bằng nhau. Có thể solve vấn đề như hướng dẫn [stackoverflow - qcut non unique bin edges](https://stackoverflow.com/questions/20158597/how-to-qcut-with-non-unique-bin-edges) (<https://stackoverflow.com/questions/20158597/how-to-qcut-with-non-unique-bin-edges>).

Thực hiện customer segmentation

Phân khúc khách hàng có thể được thực hiện thông qua nhiều phương pháp khác nhau.

1. Tính trung bình rank các biến: Một cách khá đơn giản là tính trung bình ranks của các biến ta thu được giá trị rank tổng hợp ứng với mỗi khách hàng. Ta có thể giữ nguyên mỗi rank tổng hợp là một nhóm hoặc gộp nhiều rank thành một nhóm theo khoảng giá trị như bên dưới:
 - Khách hàng VIP: rank từ 8-10.
 - Khách hàng đại chúng: rank từ 5-7.
 - Khách hàng thứ cấp: rank < 5.
1. Sử dụng thuật toán K-mean clustering để phân cụm các khách hàng dựa vào input là 3 biến giá trị ranks của Recency, Frequency, Monetary. Theo cách này ta sẽ tự động phân cụm toàn bộ khách hàng về 3 nhóm mà không cần phải tạo khoảng rank. Tuy nhiên hạn chế là các clustering không có một định nghĩa rõ ràng (VIP, đại chúng, thứ cấp). Chúng ta có thể cần thêm một bước đặt tên cho các nhóm này dựa trên phân tích centroid để tìm ra đặc tính của từng nhóm. Ngoài ra việc xác định số lượng nhóm để phân cụm cũng là một vấn đề của thuật toán. Khi đó ta sẽ cần quan sát biểu đồ sai số MRSE để tìm ra điểm elbow. Kỹ thuật này tôi sẽ không trình bày quá chi tiết trong bài viết, bạn đọc có thể search rất nhiều tài liệu trên mạng để tìm hiểu thêm.

Đổi dấu để đảm bảo tính đồng biến với rank khách hàng

Ta biết rằng rank của khách hàng càng cao thì khách hàng đó càng VIP. Một số trường hợp biến đầu vào có thể ngược chiều với rank, chính vì vậy ta sẽ phải điều chỉnh lại dấu các biến sao cho chúng đồng biến với rank để phản ánh đúng rank càng cao thì khách hàng càng VIP. Nhận thấy rằng biến Recency có quan hệ ngược chiều vì khách hàng càng lâu mua hàng (tương ứng với Recency lớn) thì xếp hạng khách hàng càng thấp. Do đó ta sẽ cần phải đổi dấu sang âm để rank của biến tỷ lệ thuận với rank của khách hàng.

Nếu không để ý tới yếu tố đồng biến thì có thể bạn sẽ tạo ra những mô hình kém chất lượng. Khi đó hậu quả cũng siêu to khổng lồ lắm. Sẽ có vài câu chuyện cười như sau: Tập khách hàng VIP là những khách thu nhập 5-6 triệu/tháng và được suggest tivi 50 triệu, trong khi khách hàng thu nhập vài trăm triệu/tháng rơi vào tập thứ cấp nên được suggest tivi 5 triệu.

Top

II. Thực hành xây dựng mô hình RFM

Trước tiên để xây dựng mô hình RFM chúng ta cần thu thập bảng dữ liệu các đơn hàng với những trường dữ liệu như sau:

- CustomerID: Giống như CMND, CustomerID giúp định danh khách hàng đó là ai?
- OrderDate: Ngày đặt hàng. Từ ngày đặt hàng ta sẽ biết được ngày đặt hàng gần nhất của mỗi khách hàng để từ đó suy ra Recency.
- OrderID: Là trường key của bảng giúp xác định một đơn hàng. Đếm số lượng OrderID ta sẽ suy ra số Frequency.
- Amount: Giá trị đơn hàng. Tổng giá trị đơn hàng theo CustomerID sẽ chính bằng Monetary.

Chúng ta sẽ lần lượt đi qua các bước:

Bước 1: Thu thập dữ liệu đơn hàng. Dữ liệu được tôi sử dụng cho bài viết này được lấy từ một cuộc thi trên kaggle. Bạn đọc có thể download tại dataCustomerRFM

(<https://drive.google.com/file/d/1fAF2ZJC9F7N5qjacQ0tAn7YS8hQU4wFU/view?usp=sharing>)

```
1 import pandas as pd
2 import pyodbc
3
4 df = pd.read_csv('dataCustomerRFM.csv')
5 df.head()
```

	CustomerID	OrderDate	OrderID	Amount
0	b5731a6437	2019-07-09 23:56:44	f996cde8437b432580ae2d9c61b1470c	6249000.0
1	ba9a96efd7	2019-07-27 22:25:52	f9e7114611c34377b456774a05310bb5	816200.0
2	3bb7d08ac6	2019-09-30 06:29:39	f9f793f21dc24607a7ca4dae6c3abed3	77000.0
3	9cf0875c93	2019-07-31 16:12:51	83ccdf7eb8944a23b0ef706551c1a818	997000.0
4	b50d95bb37	2019-08-29 06:11:15	7fc8a7e9f6da4d76b9e562ce2ca58e97	380700.0

Bước 2 : Tính các giá trị Recency, Frequency và Monetary

Top

```

1  from datetime import datetime
2  # 1. Calculate Recency
3  # 1.1. Find the most recent orderDate.
4
5  dfRecentOrder = pd.pivot_table(data = df,
6                                  index = ['CustomerID'],
7                                  values = ['OrderDate'],
8                                  aggfunc = {'OrderDate':max}
9                                  )
10
11  dfRecentOrder.columns = ['RecentOrderDate']
12  df = pd.merge(df, dfRecentOrder.reset_index(), on = ['CustomerID'])
13  df['RecentOrderDate'] = df['RecentOrderDate'].apply(lambda x: datetime
14  df['Recency'] = df['RecentOrderDate'].apply(lambda x: (datetime.now()

```

Do số ngày gần nhất kể từ lần mua hàng cuối cùng càng cao thì mức độ active của khách hàng càng thấp. Do đó giá trị này càng nhỏ thì khách hàng đó càng có thứ hạng cao. Chính vì vậy ta cần đổi dấu của Recency để giá trị của biến đồng biến với rank của khách hàng.

```

1  # Đổi dấu recency
2  df['Recency'] = - df['Recency']

```

Tiếp theo ta sẽ tính tần suất của khách hàng mua hàng trong toàn bộ thời gian nghiên cứu.

```

1  # 2. Calculate Frequency
2  dfFrequency = df.groupby('CustomerID').OrderID.nunique().to_frame()
3  dfFrequency.columns = ['Frequency']
4  df = pd.merge(df, dfFrequency.reset_index(), on = 'CustomerID')

```

Cuối cùng là tính tổng số tiền mà khách hàng đã chi tiêu.

```

1  # 3. Calculate Monetary
2  dfMonetary = df.groupby('CustomerID').Amount.sum().to_frame()
3  dfMonetary.columns = ['Monetary']
4  df = pd.merge(df, dfMonetary.reset_index(), on = 'CustomerID')

```

Bước 3: Mapping giá trị các trường Recency, Frequency, Monetary với rank tương ứng trong ngưỡng ranks là 10.

```

1  orderFrequencies = df['Frequency'].rank(method='first')
2  df['rFrequency'] = pd.qcut(orderFrequencies, 10, labels = False)
3  df[['rRecency', 'rMonetary']] = df[['Recency', 'Monetary']].apply(lambda
4  df['rank'] = (df['rFrequency'] + df['rRecency'] + df['rMonetary'])/3
5  df['FinalRank'] = df['rank'].apply(int)

```

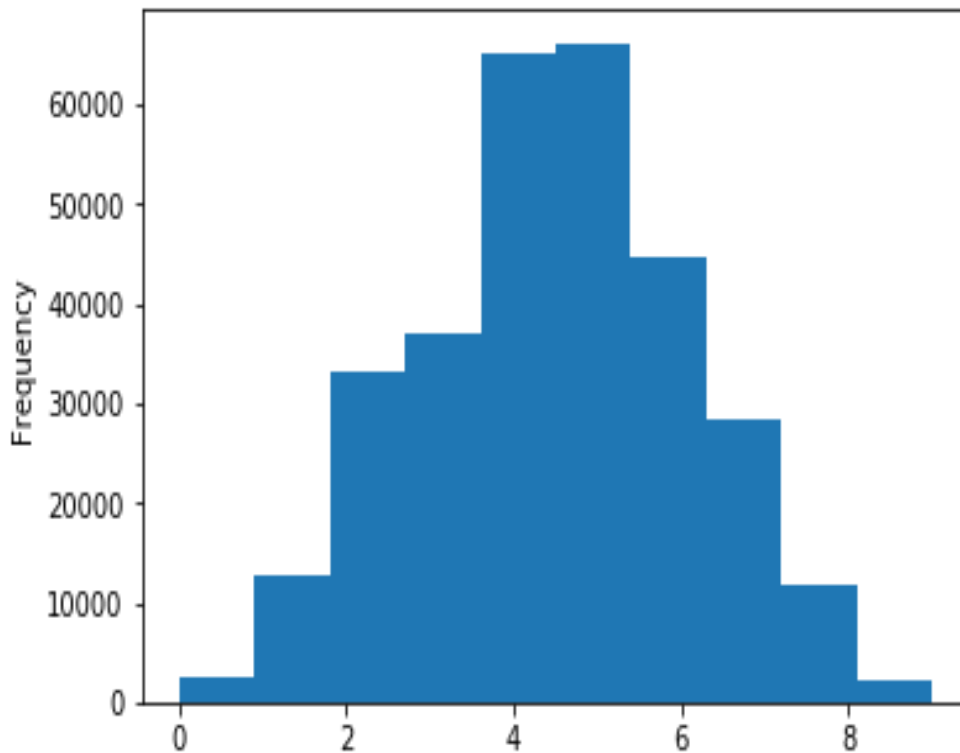
Như vậy chúng ta đã tính xong rank cho từng khách hàng. Hãy cùng xem phân phối điểm rank của khách hàng như thế nào thông qua biểu đồ histogram với số lượng bins = 10.

Top

```

1 import matplotlib.pyplot as plt
2
3 df['rank'].plot.hist(bins = 10)
4 plt.show()

```



Ta nhận thấy biểu đồ có hình dạng phân phối chuẩn. Điều này cho thấy tập khách hàng của công ty đa phần sẽ nằm ở những điểm rank nằm trong khoảng trung bình, chẳng hạn từ 4-6. Với các điểm rank quá cao hoặc quá thấp thì số lượng khách hàng tập trung càng thấp.

Dựa vào biểu đồ, ta cũng có thể phân tập khách hàng thành 3 nhóm:

- Khách hàng ít tiềm năng - nhãn **Low**: Rank từ [0, 4)
- Khách hàng thông thường - **Normal**: Rank từ [4, 7)
- Khách hàng VIP - nhãn **VIP**: Rank [7, 9]

```

1 # help(pd.qcut)
2 # df['rank'].min()
3
4 df['Segment'] = 'Low'
5 df.loc[(df['rank'] < 7) & (df['rank'] >= 4), 'Segment'] = 'Normal'
6 df.loc[df['rank'] >= 7, 'Segment'] = 'VIP'

```

Thống kê số lượng khách hàng theo mỗi Segment.

```

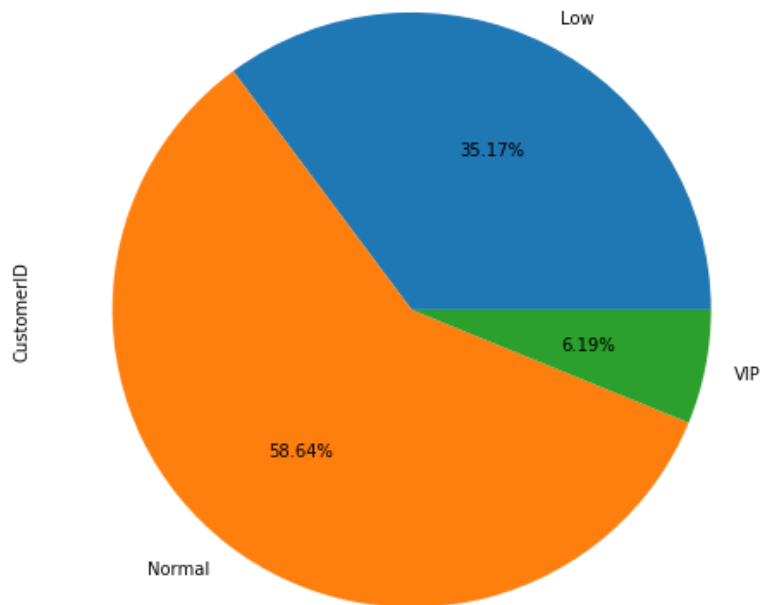
1 df.groupby('Segment').CustomerID.count().plot.pie(autopct = '%.2f%%',
2 plt.title('Tỷ lệ số lượng khách hàng theo segment KH')

```



Top

Tỷ lệ số lượng khách hàng theo segment KH

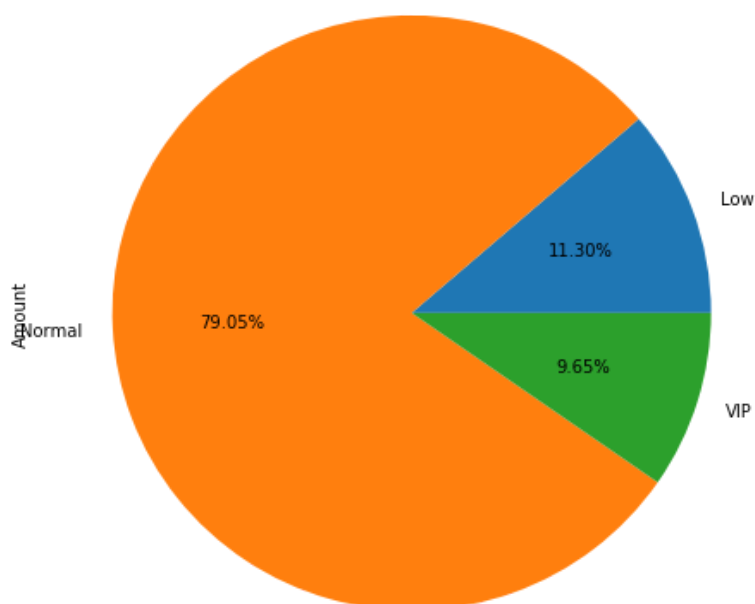


Kết quả cho thấy số lượng khách hàng VIP của công ty rất ít. Công ty cần có chính sách thay đổi sản phẩm để bắt kịp thị hiếu thị trường hoặc tăng cường marketing để thu hút khách hàng tiêu dùng nhiều hơn.

Tiếp theo chúng ta sẽ visualize biểu đồ theo doanh số, tần suất mua hàng, thời gian quay trở lại của các nhóm khách hàng sau khi đã segment.

```
1 df.groupby('Segment').Amount.sum().plot.pie(autopct = '%.2f%%', figsize=(10, 10))
2 plt.title('Tỷ lệ doanh số theo segment KH')
```

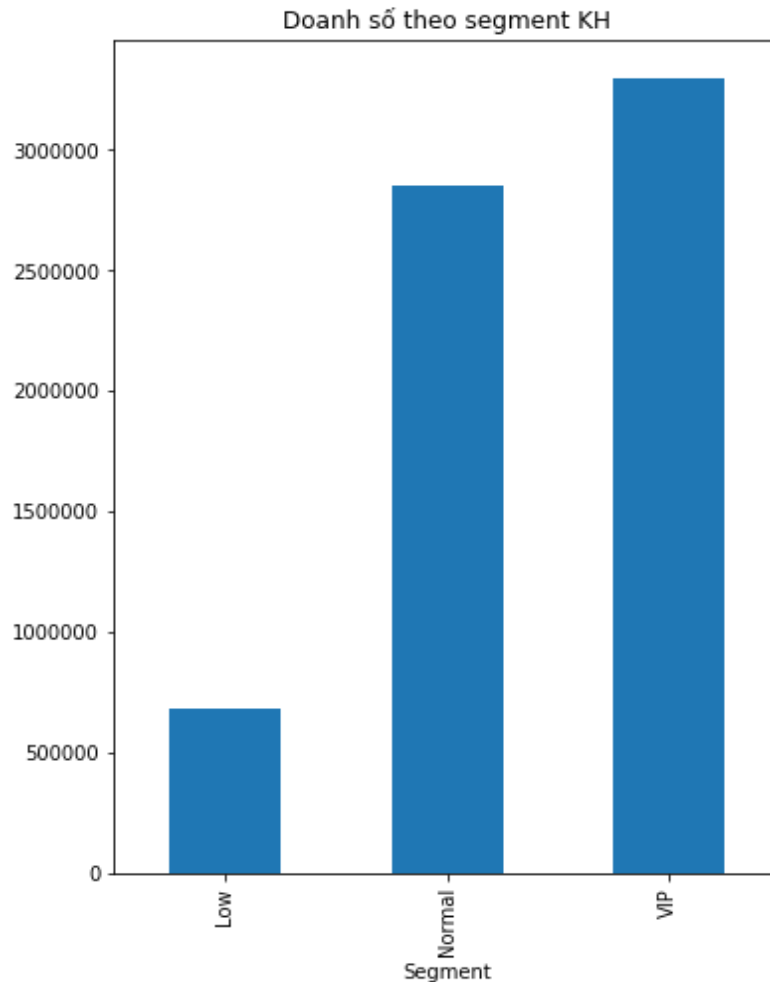
Doanh số theo segment KH



Top

Khách hàng normal lại là khách hàng mang lại doanh số lớn nhất cho công ty, chiếm tới gần 80%. Công ty vẫn chưa thể đạt được mục tiêu 20% khách hàng VIP mang lại 80% lợi nhuận.

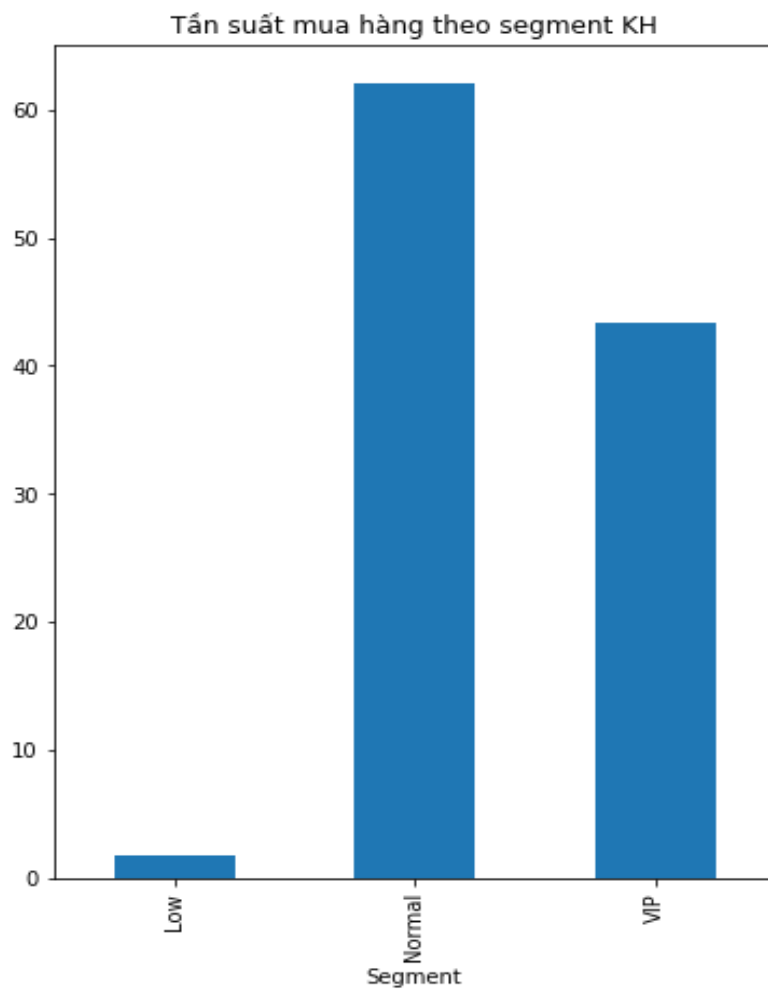
```
1 df.groupby('Segment').Amount.mean().plot.bar(figsize = (6, 8))
2 plt.title('Doanh số theo segment KH')
```



Trung bình một khách hàng VIP tiêu khoảng 3.5 triệu. Khách hàng thông thường tiêu khoảng 2.9 triệu và khách hàng Thấp tiêu dưới 1 triệu. Chênh lệch tiêu thụ giữa khách hàng thông thường và khách hàng VIP không quá lớn. Trong khi chênh lệch này so với khách hàng ít tiềm năng là rất lớn. Công ty đang chưa có một tập khách hàng thực sự gọi là VIP hẳn. Đừng buồn vì các điều chỉnh về chính sách sản phẩm và chăm sóc khách hàng phù hợp sẽ cải thiện dần các chỉ số này.

```
1 df.groupby('Segment').Frequency.mean().plot.bar(figsize = (6, 8))
2 plt.title('Tần suất mua hàng theo segment KH')
```

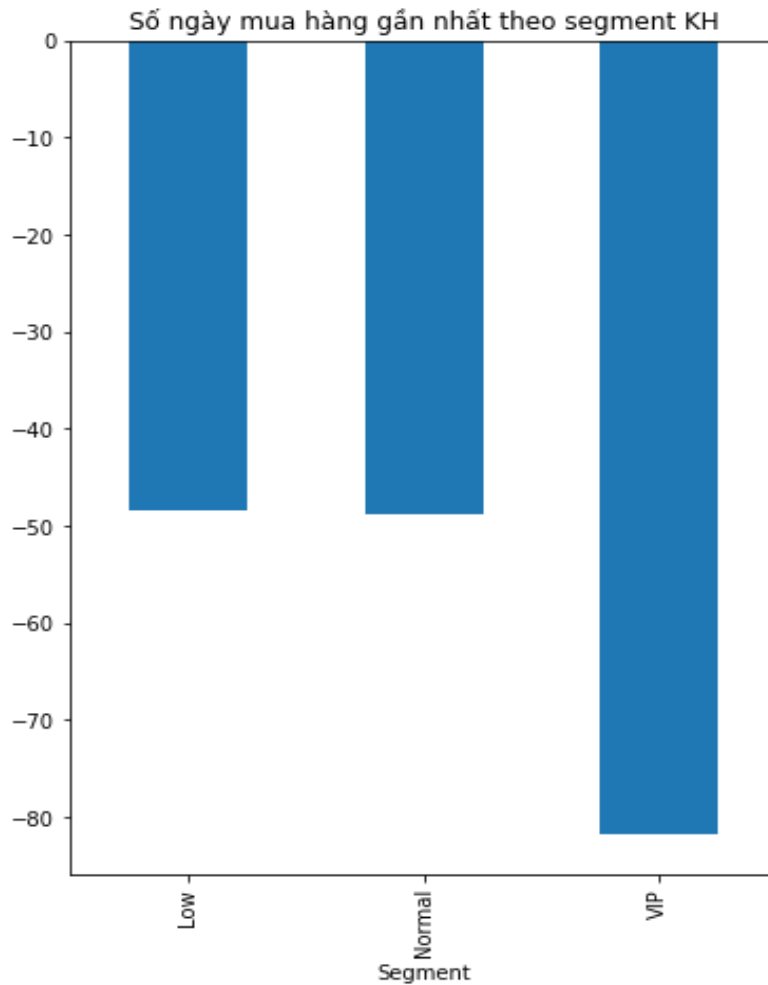
Top



Khách hàng thông thường mua nhiều đơn hàng hơn so với các khách hàng VIP. Khách hàng ít tiềm năng mua quá ít đơn hàng.

```
1 df.groupby('Segment').Recency.mean().plot.bar(figsize = (6, 8))  
2 plt.title('Số ngày mua hàng gần nhất theo segment KH')
```

Top



Khách hàng VIP khi đã mua hàng thì lâu lâu mới mua một lần nữa. Khách VIP thì cũng khó tính phải không nào? Nhưng cũng chứng tỏ rằng sản phẩm chưa đủ hấp dẫn đối với những nhóm khách hàng này để thuyết phục họ quay lại.

Kết luận:

- Chúng ta cần thay đổi chính sách về sản phẩm để chúng trở nên đa dạng và hấp dẫn khách hàng VIP hơn.
- Gia tăng số lượng khách hàng VIP để đảm bảo mục tiêu 20% khách hàng là VIP mang lại 80% doanh số.
- Đưa ra các chiến lược chăm sóc và giữ chân khách hàng hiệu quả.

3. Các biến thể khác của RFM

Ngoài ra RFM còn có các biến thể khác dựa trên sự thay đổi thêm bớt các biến trong 3 biến Recency, Frequency và Monetary. Bên dưới là một trong những phương pháp đó:

- RFD - Recency, Frequency, Duration (thời gian) là phiên bản đã được modified của phân tích RFM. Nhưng thay vì phân tích giá trị khách hàng, RFD được sử dụng để phân tích hành vi khách hàng theo các nhóm người xem/người đọc/người lướt web.
- RFE - Recency, Frequency, Engagement(mức độ cam kết) là phiên bản mở rộng của phân tích RFD nhằm xác định mức độ gắn bó của khách hàng đối với một nền tảng web, app. Trường Engagement (mức độ cam kết) được xác định thông qua thời lượng truy cập, số trang trên mỗi lượt truy cập và các chiều dữ liệu khác tương tự. Mô hình RFE có thể được sử dụng để phân tích hành vi khách hàng theo các nhóm người xem/người đọc/người lướt web.

Top

- RFM-I - Recency, Frequency, Monetary Value – Interactions (Giá trị tiền tệ - Tương tác) là một phiên bản khác của RFM để đánh giá chi phí tương tác marketing trong tiếp cận khách hàng.
- RFMTC - Recency, Frequency, Monetary Value, Time (Thời gian), Churn rate (Tỷ lệ rời bỏ) là một mô hình RFM mở rộng được đề xuất bởi I-Cheng và cộng sự(2009). Mô hình sử dụng chuỗi Bernoulli trong lý thuyết xác suất nhằm dự báo xác suất mua hàng tại những chiến dịch marketing tiếp theo.

4. Tổng kết

Như vậy thông qua bài viết này chúng ta đã nắm được phương pháp RFM, một phương pháp khá quan trọng và phổ biến trong phân khúc khách hàng dựa trên giá trị mang lại cho công ty. Thông qua phân khúc khách hàng, Chúng ta có thể phát triển các chiến lược kinh doanh và bộ máy tổ chức doanh nghiệp phù hợp với mục tiêu phục vụ trên từng phân khúc. Đồng thời theo dõi chuyển dịch cơ cấu khách hàng trên từng phân khúc qua thời gian cũng giúp đánh giá được mức độ phát triển tập khách hàng của công ty ra sao? Công ty cần tung ra những điều chỉnh, chiến lược như thế nào để phát triển theo hướng gia tăng tỷ trọng khách hàng VIP, giữ chân khách hàng quay lại mua sắm thường xuyên hơn và nâng cao giá trị đơn hàng trên mỗi lượt mua sắm.

Ở đây tôi vẫn chưa sử dụng thuật toán K-mean clustering để phân cụm khách hàng. Sử dụng thuật toán này không quá khó và xin dành cho bạn đọc.

5. Tài liệu tham khảo

Và cuối cùng không thể thiếu sau mỗi bài viết của tôi là tài liệu tham khảo.

1. RMF model wikipedia ([https://en.wikipedia.org/wiki/RFM_\(customer_value\)](https://en.wikipedia.org/wiki/RFM_(customer_value)))
2. RMF model with σ -scaled hadron masses and couplings for description of heavy-ion collisions below 2A GeV (<https://arxiv.org/abs/1902.09016>)
3. Using RFM to Identify Your Best Customers (<https://www.eightleaves.com/2011/01/using-rfm-to-identify-your-best-customers/>)
4. Making Your Database Pay Off Using Recency Frequency and Monetary Analysis (<http://www.dbmarketing.com/2010/03/making-your-database-pay-off-using-recency-frequency-and-monetary-analysis/>)
5. Series marketing với python - Khôi Nguyễn (<https://viblo.asia/p/marketing-voi-python-part-2-phan-khuc-khach-hang-63vKjXXdl2R>)

Top