

**BỘ GIÁO DỤC VÀ ĐÀO TẠO  
TRƯỜNG ĐẠI HỌC MỎ - ĐỊA CHẤT**

# **BÁO CÁO SEMINAR**

**TÊN ĐỀ TÀI  
PHÂN CHIA PHÂN KHÚC KHÁCH HÀNG DỰA TRÊN  
MÔ HÌNH RFM**

Người thực hiện: PGS. TS. Lê Văn Hưng (Mã cán bộ: 0801-01)

Đơn vị: Bộ môn Công nghệ phần mềm  
Khoa Công nghệ Thông tin

**Hà Nội - 2022**

## MỤC LỤC

DANH MỤC CÁC HÌNH VẼ .....	2
1. Đặt vấn đề .....	3
2. Mô hình RFM.....	4
2.1 Chuẩn hóa dữ liệu đầu vào.....	5
2.2 Phương pháp điểm trung bình.....	6
2.3 Sử dụng thuật toán K-mean clustering.....	6
3. Ví dụ về xây dựng mô hình RFM .....	6
3.1 Thu thập và xử lý dữ liệu .....	6
3.2 Phương pháp điểm trung bình.....	7
3.3 Mã nguồn ví dụ .....	12
4. Kết luận .....	14
5. TÀI LIỆU THAM KHẢO.....	14

## **DANH MỤC CÁC HÌNH VẼ**

Hình 1. Biểu đồ phân phối điểm của các khách hàng.....	7
Hình 2. Thống kê phân trăm khách hàng theo mỗi phân khúc .....	8
Hình 3. Phân trăm tổng doanh số theo nhóm khách hàng .....	9
Hình 4. Doanh số trung bình theo phân khúc .....	10
Hình 5. Tần suất mua hàng theo các phân khúc .....	11
Hình 6. Trung bình số ngày mua hàng gần nhất theo phân khúc .....	11

## 1. Đặt vấn đề

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, đặt mục tiêu marketing đến đúng tập khách hàng hơn và khách hàng được phục vụ tốt hơn. Để minh chứng cho nhận định trên, ta có một vài ví dụ đơn giản sau:

- Giả sử doanh nghiệp của ta kinh doanh rất nhiều các mặt hàng từ cái bút bi đến xe hơi. Ta 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.

- Ta không thể gợi ý 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, ta bè), mỗi tháng anh ta tiết kiệm được 1 triệu và cần tới 4 năm tiết kiệm để mua được chiếc tivi đó. Ta chỉ nên gợi ý chiếc nào rẻ rẻ thôi, 5 triệu chẳng hạn. Trái lại, với một khách hàng thu nhập vài trăm triệu/tháng và cũng muốn mua tivi. Ta cũng không nên gợi ý 1 chiếc tivi 5 triệu nếu không muốn khách hàng nổi giận vì tội sỉ nhục khách hàng.

- Ta cũng không nên gợi ý một người mua iPhone với một chiếc tai nghe Samsung hay một đứa trẻ với rượu 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.
- Chăm sóc khách hàng đồng thời cũng phải phù hợp với nhu cầu, thị hiếu của khách hàng.

- Như vậy việc phân chia 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 khách hàng đại chúng (Mass Affluent), khách hàng khả giả (Affluent), khách hàng cao cấp (Priority). 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. Trong chúng khoán có thể chia tập khách hàng theo khách hàng thứ cấp, khách hàng thông thường và khách hàng VIP dựa trên giá trị giá trị tài sản ròng (NAV) 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.

## **2. Mô hình RFM**

Phương pháp phân chia phân khúc khách hàng được giới thiệu ở đây có tên là RFM (Recency- Frequency- Monetary), một trong những phương pháp điển hình được sử dụng trong phân khúc khách hàng. RFM định lượng giá trị của một khách hàng dựa trên 3 thông tin chính:

Recency: Khoảng thời gian mua hàng gần đây nhất đến nay. 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.

**Monetary:** Là số tiền mua hàng 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 sắm 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. Ta có thể một số phương pháp, chẳng hạn thuật toán phân cụm K-Mean clustering, để nhóm các khách hàng có cùng mức độ VIP thành một nhóm.

## **2.1 Chuẩn hóa dữ liệu đầu vào**

Ta thấy rằng biến Recency có quan hệ ngược chiều với mức độ quan trọng của khách hàng vì khách hàng càng lâu mua hàng (trùng ứng với Recency lớn) thì xếp hạng khách hàng càng thấp. Vì vậy, ta cần đổi dấu của biến Recency sang âm để đảm bảo giá trị của nó càng cao thì khách hàng đó càng VIP.

Ta cần phải chuẩn hóa dữ liệu đầu vào vì một đầu vào có giá trị quá lớn so với những đầu vào còn lại sẽ ảnh hưởng rất nhiều đến kết quả. Trong bài toán này, thông thường Monetary sẽ có giá trị lớn hơn rất nhiều so với Recency và Frequency. Việc chuẩn hóa dữ liệu sẽ giúp vai trò của các biến trong xác định cụm là bình đẳng.

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 đưa giá trị của các biến về thang điểm 10. Sau đó, ta

sẽ tính điểm của từng khách hàng bằng trung bình cộng của các giá trị của ba biến Recency, Frequency và Monetary. Giá trị này càng cao thì khách hàng càng có giá trị đối với công ty.

Để chuyển giá trị của một biến cụ thể sang thang điểm 10, ta sẽ sử dụng hàm `pandas.qcut(10)`. Hàm `qcut()` sẽ tự động tìm ra các khoảng giá trị sao cho số lượng các giá trị được chia đều vào mỗi khoảng.

## 2.2 Phương pháp điểm trung bình

Một cách khá đơn giản là ta tính điểm trung bình của từng khách hàng theo điểm của các biến đầu vào và xếp nhóm khách hàng dựa trên giá trị này. Ví dụ, ta có thể xếp nhóm khách hàng như sau:

Khách hàng VIP: điểm từ 8-10.

Khách hàng đại chúng: điểm từ 5-7.

Khách hàng thứ cấp: điểm < 5.

## 2.3 Sử dụng thuật toán K-mean clustering

Sử dụng thuật toán K-mean clustering để phân cụm các khách hàng với đầu vào là giá trị của ba biến Recency, Frequency, Monetary đã chuẩn hóa. Theo cách này ta sẽ tự động phân cụm toàn bộ khách hàng thành 3 nhóm. Tuy nhiên hạn chế là các thuật toán phân cụm là sau khi có các cụm, ta không biết được cụm nào là cụm khách hàng VIP, đại chúng hay thứ cấp. Ta có thể cần thêm một bước nữa là đặt tên cho các cụm này dựa trên phân tích tâm cụm để tìm ra đặc tính. Ngoài ra, việc xác định trước số lượng cụm để phân cũng là một vấn đề của thuật toán K-mean clustering.

## 3. Ví dụ về xây dựng mô hình RFM

### 3.1 Thu thập và xử lý dữ liệu

**Bước 1:** Để xây dựng mô hình RFM, ta cần thu thập dữ liệu mua hàng của khách hàng với những trường dữ liệu sau:

CustomerID: Mã khách hàng

OrderDate: Ngày đặt hàng. Từ ngày đặt hàng ta sẽ biết được ngày đặt hàng gần nhất của mỗi khách hàng để từ đó tính ra giá trị Recency.

OrderID: Mã đơn hàng. Đếm số lượng OrderID ta sẽ tính ra giá trị Frequency.

Amount: Giá trị đơn hàng. Tổng giá trị đơn hàng theo CustomerID sẽ chính là giá trị Monetary.

Dữ liệu xây dựng mô hình ở đây được chúng tôi thu thập trong file dataCustomerRFM.

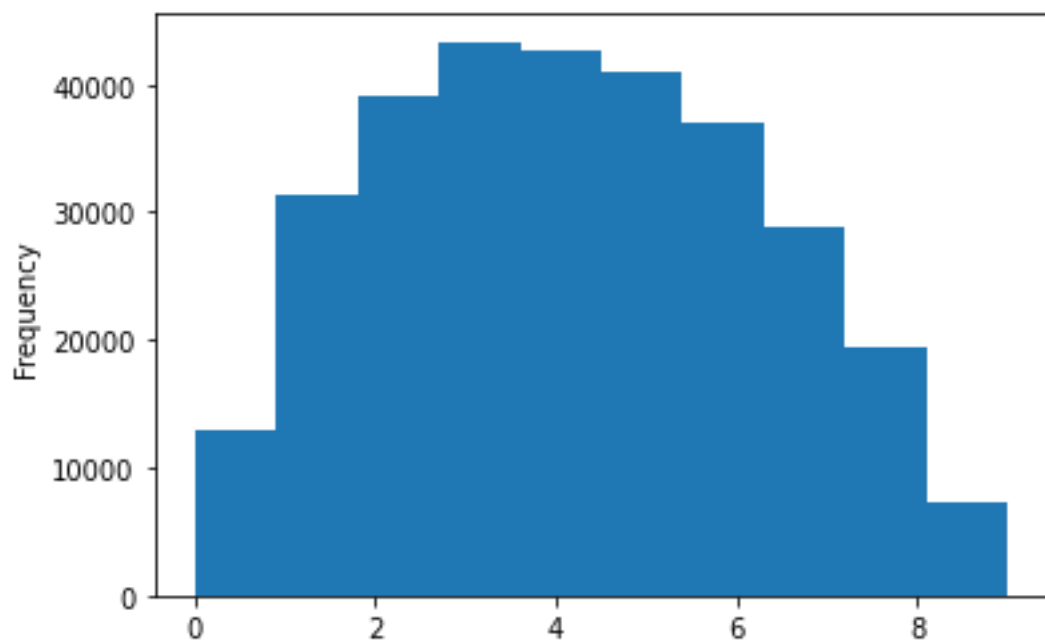
**Bước 2:** Tính các giá trị Recency, Frequency và Monetary và đổi dấu giá trị của biến Recency.

**Bước 3:** Chuyển đổi giá trị các trường Recency, Frequency, Monetary thành giá trị tương ứng trong thang điểm 10.

### 3.2 Phương pháp điểm trung bình

Tính điểm xếp hạng của mỗi khách hàng bằng trung bình cộng của các trường Recency, Frequency và Monetary.

Biểu đồ phân phối điểm xếp hạng của các khách hàng như Hình 1.



Hình 1. Biểu đồ phân phối điểm của các khách hàng

Ta nhận thấy biểu đồ có dạng phân phối chuẩn. Điều này cho thấy tập khách hàng của công ty đa phần nằm ở những điểm nằm trong khoảng trung

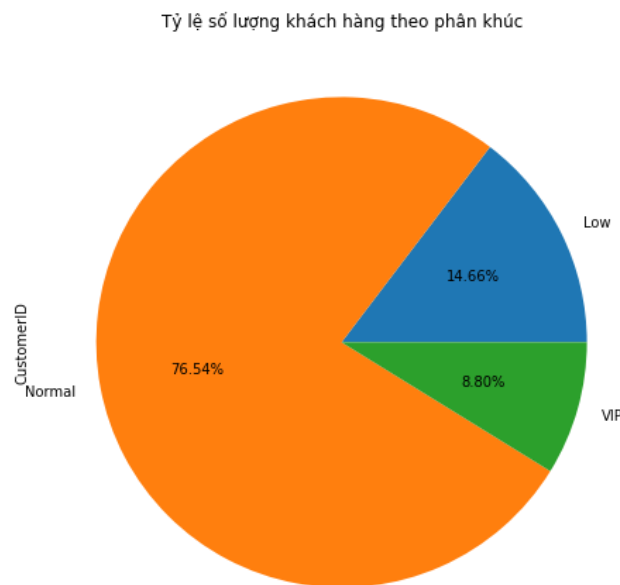


bình, chẳng hạn từ 3-7. Với các điểm quá cao hoặc quá thấp thì số lượng khách hàng càng thấp.

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

- Khách hàng ít tiềm năng - nhãn Low: Điểm từ [0, 3)
- Khách hàng thông thường – nhãn Normal: Điểm từ [3, 7)
- Khách hàng VIP - nhãn VIP: Điểm [7, 9]

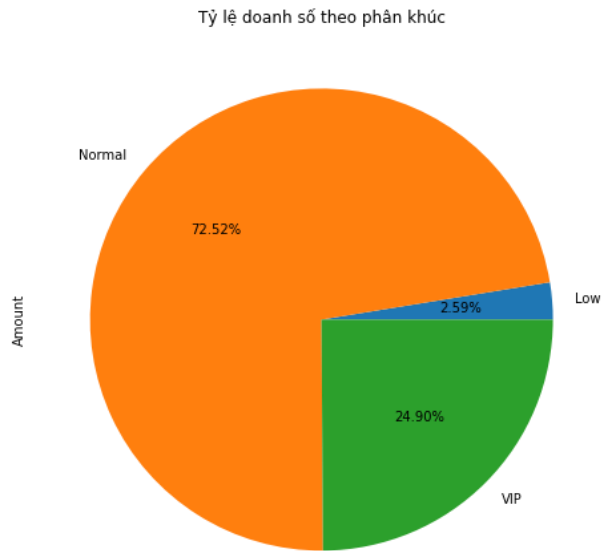
Hình 2 sau đây thống kê số lượng khách hàng theo mỗi phân khúc.



Hình 2. Thống kê phân trăm khách hàng theo mỗi phân khúc

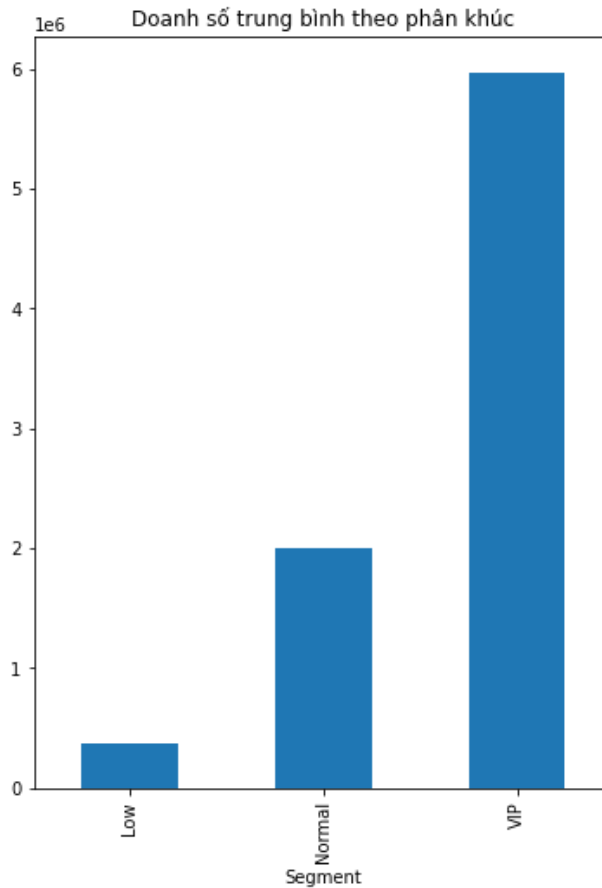
Kết quả cho thấy số lượng khách hàng VIP của công ty rất ít (8,8%). Công ty cần có chính sách thay đổi sản phẩm để bắt kịp thị hiếu thị trường hoặc tăng cường marketing để thu hút khách hàng tiêu dùng nhiều hơn.

Tiếp theo ta sẽ vẽ các biểu đồ theo doanh số, tần suất mua hàng, thời gian quay trở lại của các nhóm khách hàng để phân tích như sau.



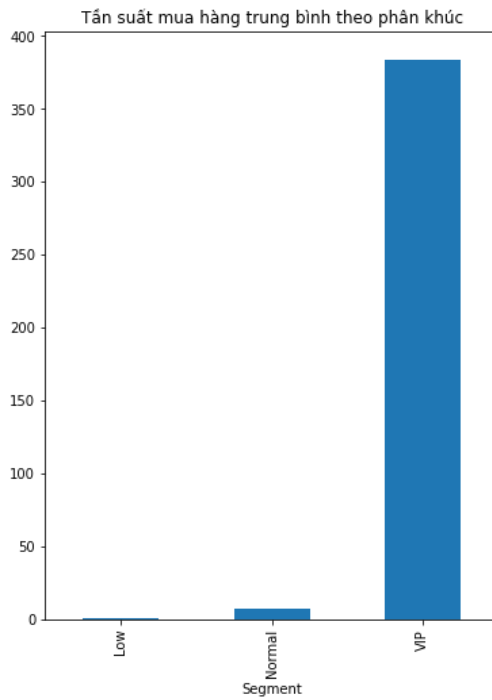
Hình 3. Phần trăm tổng doanh số theo nhóm khách hàng

Hình 3 cho thấy khách hàng Normal lại là khách hàng mang lại doanh số lớn nhất cho công ty, chiếm tới gần 73%. Công ty vẫn chưa thể đạt được mục tiêu 20% khách hàng VIP mang lại 80% lợi nhuận.



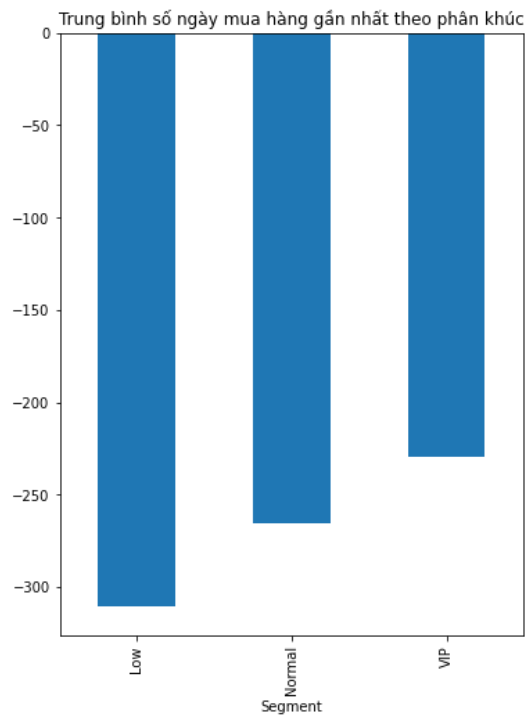
Hình 4. Doanh số trung bình theo phân khúc

Hình 4 cho thấy trung bình một khách hàng VIP tiêu khoảng 6 triệu, khách hàng thông thường tiêu khoảng 2 triệu và khách hàng ít tiềm năng tiêu dưới 0,3 triệu. Như vậy, chênh lệch mức tiêu tiền giữa các phân khúc khách hàng là rất lớn. Công ty có cần có các chính sách sản phẩm và chăm sóc khách hàng phù hợp để tăng tập khách hàng VIP.



Hình 5. Tần suất mua hàng theo các phân khúc

Hình 5 cho thấy khách hàng VIP có tần suất mua hàng rất cao so với các phân khúc khác. Khách hàng ít tiềm năng mua quá ít đơn hàng.



Hình 6. Trung bình số ngày mua hàng gần nhất theo phân khúc

Hình 6 cho thấy khách hàng VIP có tần suất mua hàng lớn nên thời gian mua cũng gần đây hơn.

Kết luận:

- Công ty 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.

### 3.3 Mã nguồn ví dụ

```
import pandas as pd
df = pd.read_csv('dataCustomerRFM.csv')
from datetime import datetime
# 1. Calculate Recency
# 1.1. Find the most recent orderDate.
dfRecentOrder = pd.pivot_table(data = df,
                                index = ['CustomerID'],
                                values = ['OrderDate'],
                                aggfunc = {'OrderDate':max}
                                )
dfRecentOrder.columns = ['RecentOrderDate']
df = pd.merge(df, dfRecentOrder.reset_index(), on = ['CustomerID'])
df['RecentOrderDate'] = df['RecentOrderDate'].apply(lambda x: datetime.strptime(x, '%Y-%m-%d %H:%M:%S'))
df['Recency'] = df['RecentOrderDate'].apply(lambda x: (datetime.now() - x).days)
# Đổi dấu recency
df['Recency'] = - df['Recency']
# 2. Calculate Frequency
dfFrequency = df.groupby('CustomerID').OrderID.nunique().to_frame()
dfFrequency.columns = ['Frequency']
df = pd.merge(df, dfFrequency.reset_index(), on = 'CustomerID')
```

```

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

# 4. Quy đổi và tính điểm trung bình
orderFrequencies = df['Frequency'].rank(method='first')
df['rFrequency'] = pd.qcut(orderFrequencies, 10, labels = False)
df[['rRecency', 'rMonetary']] = df[['Recency', 'Monetary']].apply(lambda x: pd.qcut(x, 10, labels = False))

df['rank'] = (df['rFrequency'] + df['rRecency'] + df['rMonetary'])/
3

df['FinalRank'] = df['rank'].apply(int)

# 5. Vẽ phân phối theo điểm
import matplotlib.pyplot as plt
df['FinalRank'].plot.hist(bins = 10)
plt.show()

# 6. Xác định phân khúc khách hàng
df['Segment'] = 'Low'
df.loc[(df['rank'] < 8) & (df['rank'] >= 2), 'Segment'] = 'Normal'
df.loc[df['rank'] >= 8, 'Segment'] = 'VIP'

# 7. Vẽ các biểu đồ
df.groupby('Segment').CustomerID.count().plot.pie(autopct = '%.2f%%',
', figsize = (8, 8))

plt.title('Tỷ lệ số lượng khách hàng theo phân khúc')

df.groupby('Segment').Amount.sum().plot.pie(autopct = '%.2f%%', fig
size = (8, 8))

plt.title('Tỷ lệ doanh số theo phân khúc')

df.groupby('Segment').Amount.mean().plot.bar(figsize = (6, 8))
plt.title('Doanh số trung bình theo phân khúc')

df.groupby('Segment').Frequency.mean().plot.bar(figsize = (6, 8))
plt.title('Tần suất mua hàng trung bình theo phân khúc')

df.groupby('Segment').Recency.mean().plot.bar(figsize = (6, 8))
plt.title('Trung bình số ngày mua hàng gần nhất theo phân khúc')

```

## 4. Kết luận

Phương pháp RFM là một phương pháp khá quan trọng và phổ biến trong phân chia 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.

## 5. TÀI LIỆU THAM KHẢO

- [1] RMF model with  $\sigma$ -scaled hadron masses and couplings for description of heavy-ion collisions below 2A GeV
- [2] Using RFM to Identify Your Best Customers
- [3] Making Your Database Pay Off Using Recency Frequency and Monetary Analysis
- [4] Series marketing với python - Khôi Nguyễn