Phân tích RFM cho data game “Biệt đội Đùn Đùn”

# Goal

Phân tích data game, phân khúc khách hàng, ảnh hưởng của các yếu tố đến từng phân khúc, từ đó có các chiến dịch phù hợp cho từng phân khúc.

# Resources

Link data: <https://drive.google.com/drive/folders/1_QKmXaTUj2PLR_2EaXGzRtWIOjuBTdnf?usp=drive_link>

# Method

1. Phân tích tổng quan về data
2. Bảng websale : (Gồm thông tin về ID tài khoản người chơi, ID giao dịch, Số tiền mà người chơi giao dịch tronng game, …). Dựa vào đó có thể tính:

* Doanh thu từng ngày
* Sự tăng trưởng doanh thu
* Số lượng khách hàng
* Số lượng giao dịch
* Số lượng account cũ/mới
* Tỷ lệ tăng trưởng của account mới
* Tỷ lệ giữ chân account cũ
* Số lượng mỗi loại package

1. Bảng af\_registration: (Gồm thông tin về quốc gia, châu lục, thành phố, platfom (IOS, android), các chiến dịch truyền thông,…)
   * Tài khoản người chơi ở trong/ ngoài nước
   * Tài khoản ở các tỉnh có số lượng như thế nào?
   * Platfom
2. Bảng user\_level (Level của người chơi)
   * Số lượng người chơi ở từng level
3. RFM analysis
4. Phân tích data để tìm ra những feature sử dụng để tính toán R, F, M

* Recency (Tần suất gần nhất).
* Frequency (Tần suất mua hàng).
* Monetary (Giá trị tiêu dùng).

1. Phân khúc

* Phân khúc account vào từng nhóm.
* Đánh giá đặc điểm của từng nhóm.
* Nhận xét và đưa ra cách chăm sóc với từng phân khúc.

# Analysis

### Doanh thu hàng ngày (tổng)

* Công thức tính:

1. FUN= 500 VND.

Revenue = TopupAmount\*500

* Công việc cần làm:
* Tạo một key chung đại diện cho biến ngày trong năm (DayKey)
* Tính doanh thu
* Group doanh thu theo ngày
* In thu theo ngày, vẽ chart.

A graph with blue lines

Description automatically generated

### Tăng trường doanh thu

* Công thức

Doanh thu ngày hiện tại/ Doanh thu ngày trước đó -1

A graph with blue lines

Description automatically generated

### Số lượng Account hàng ngày

* Công thức: Đếm số lượng Account ( không trùng), nhóm theo ngày.

A graph with blue dots

Description automatically generated

### Số lượng giao dịch hàng ngày:

* Công thức: Nhóm theo ngày groupby(['DayKey']) và đếm số giao dịch (['TransID'].count())

A graph with blue lines

Description automatically generated

### Doanh thu trung bình hàng ngày của giao dịch

A graph with blue lines

Description automatically generated

*Nhận xét*:

* Số lượng giao dịch tăng vào ngày 9 tháng 8 thế nhưng doanh thu trung bình giao dịch của ngày hôm đó lại giảm => người chơi giao dịch các gói có giá trị thấp hơn vào ngày hôm đó.
* Ngày hôm đó số lượng khách hàng cũng tăng (Có thể có sự kiện gì mới đã thu hút người chơi thực hiện giao dịch).

### Số lượng người chơi giao dịch lần đầu và người chơi có sự giao dịch lại

* Tìm ngày giao dịch đầu tiên của mỗi tài khoản (AccountID) bằng cách nhóm dữ liệu theo "AccountID" và lấy .min() ngày giao dịch của họ. Sau đó tạo 1 DataFrame mới có tên "retail\_new\_df" chứa thông tin giao dịch lần đầu của mỗi tài khoản.
* Tạo một cột mới 'UserType' trong "retail\_new\_df" và gán giá trị 'New' cho tất cả các hàng.
* Cập nhật giá trị cho 'UserType' với điều kiện: Nếu 'DayKey' (ngày giao dịch) lớn hơn 'MinDayKey' (ngày giao dịch đầu tiên), thì cột 'UserType' được cập nhật thành 'Existing', ngược lại, nó sẽ là 'New.

A screen shot of a computer

Description automatically generated

### Tổng số tiền mà người chơi new/ existing chi tiêu theo ngày

A graph with blue lines and text

Description automatically generated

*Nhận xét*:

* Dễ thấy, lượng account cũ nhiều hơn lượng account mới. Số lượng account mới còn thấp. Doanh thu mà account cũ mang lại lớn hơn so với account mới.
* Số lượng account quay lại giao dịch khá ổn định.

### Tỷ lệ giữ chân account cũ (người chơi quay lại giao dịch)

**A graph with blue lines

Description automatically generated**

*Nhận xét*:

* Ngày 9 tháng 8 có sự giảm sút rất lớn.
* Cần tìm ra nguyên nhân tại sao ngày 9 tháng 8 lại có giảm sút như vậy.

### Tỷ lệ phân bố server theo vùng

A blue circle with a orange triangle

Description automatically generated

*Nhận xét*:

* Tỷ lệ người chơi trong nước chiếm đa số, lớn hơn rất nhiều so với lượng người chơi ở nước ngoài.

### Tỷ lệ phân bố server theo các quốc gia

* Sử dụng Choropleth map

A map of the world

Description automatically generated

Số lượng ở các quốc gia:

A computer screen shot of a number

Description automatically generated

*Nhận xét:*

* Game thu hút lượng người chơi ở các quốc gia: Việt Nam, United Stated, Japan,…

### Tỷ lệ phân bố server ở các thành phố tại Việt Nam

A graph with blue bars

Description automatically generated*Nhận xét:*

* Đa phần người chơi là ở trong nước, người chơi nước ngoài có số lượng rất thấp.
  + Hồ Chí Minh có lượng người chơi nhiều nhất, sau đó đến Hà Nội. (thành phố lớn)
* Các tỉnh vùng núi như Bắc Kạn, Cao Bằng, Kon Tum… có số lượng người chơi thấp.

# Phân khúc khách hàng

Ở trên, chúng ta đã phân tích dữ liệu của data game và tìm ra những yếu tố ảnh hưởng đến doanh thu ( việc tài khoản nạp tiền vào game). Phần này, chúng ta sẽ tiến hành phân loại khách hàng theo nhóm để phục vụ cho công việc marketing.

### Phương pháp 1: Sử dụng Kmeans để phân khúc khách hàng

A diagram of a company

Description automatically generated with medium confidence

***Tổng quát:***

Data sử dụng trực tiếp để tính RFM là ‘websale’ gồm 145051 dòng, chứa giao dịch của các tài khoản game. Nghiên cứu sẽ tập trung khai thác các thuộc tính gồm:

* Thuộc tính AccountID mỗi một giao dịch (nạp game) chỉ thuộc về 1 tài khoản; tính TopupAmount (số tiền nạp), mỗi một giao dịch sẽ có mã giao dịch (TransID).
* Thuộc tính TrainsID sẽ dùng để tính giá trị Frequency (đếm số TrainsID của mỗi AccountID);
* Thuộc tính TopupAmount với công thức sum(TopupAmout)\*500 có thể xác định được tổng số tiền mà Account nạp vào game( tính giá trị Monetary);
* Thuộc tính TopupDate (ngày nạp tiền) dùng để tính giá trị Recency bằng cách chọn ra TopupDate mới nhất (gần nhất) của từng AccountID.
* Sau quá trình khảo sát và tiền xử lý cũng như loại bỏ các giá trị không cần thiết và giữ lại các giá trị phù hợp, mô hình dữ liệu RFM được thiết lập.

***Chi tiết:***

1. Clean data

* Loại ra các data trước ngày 19/07 do game launch từ ngày 19/07.
* Loại ra data có server là 5001, 1 vì đó là server dùng để test.
* Chỉ giữ lại data có ‘Gamecode’ là ‘bietdoi’.

1. Tính R, F, M
   1. Tính Recency

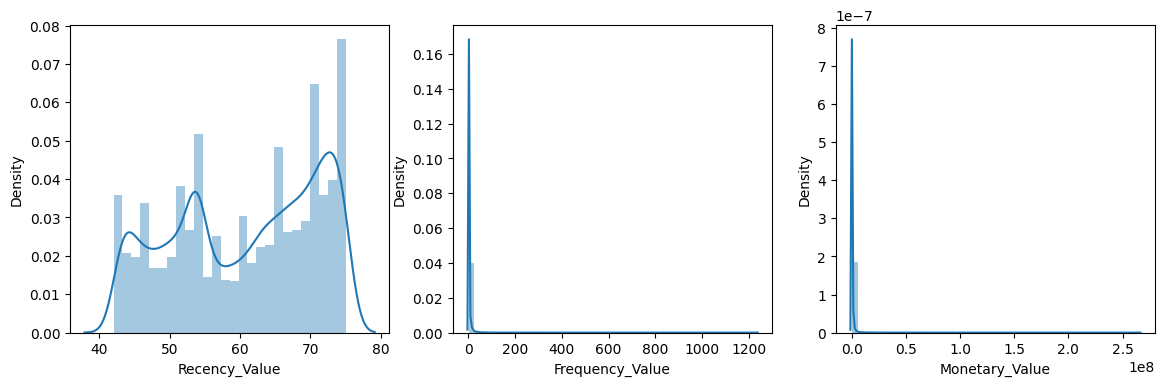
* Gán ngày hiện tại: current\_date = datetime.now()
* Lấy ra ngày giao dịch cuối cùng của mỗi account
* Tính ra khoảng thời gian kể từ lần mua hàng cuối cùng đến thời điểm hiện tại
  1. Tính Frequency
* Tính bằng cách gom nhóm giao dịch (TrainsID) theo AccountID và đếm số lượng.
  1. Tính Monetary
* Tính bằng cách gom nhóm TopupAmout theo AccountID, sum(TopupAmout) \*500.

|  | **AccountID** | **Recency\_Value** | **Frequency\_Value** | **Monetary\_Value** |
| --- | --- | --- | --- | --- |
| **0** | 3650206 | 43 | 88 | 20119000 |
| **1** | 3604503 | 45 | 118 | 29665000 |
| **2** | 3576917 | 43 | 51 | 13360000 |
| **3** | 3732234 | 45 | 13 | 2217500 |
| **4** | 3596920 | 44 | 16 | 2026500 |
| **6** | 3615671 | 46 | 4 | 1002000 |
| **8** | 3727973 | 46 | 1 | 25000 |
| **9** | 3572892 | 42 | 115 | 27190500 |
| **12** | 3727296 | 46 | 4 | 298000 |
| **13** | 3709921 | 45 | 12 | 6380000 |

1. Xử lý bảng RFM

3.1 Transformation

Sau khi kiểm tra Recency\_Value, Frequency\_Value, Monetary\_Value thấy chúng chưa tuân theo phân phối chuẩn => Cần thực hiện Transformation.



Một số phương pháp transformation

* Log transformation
* Square root transfomation
* Box-cox transformation
* Cube root transformation
* Với Frequency\_Value và Recency\_Value:

A group of graphs of different sizes

Description automatically generated A group of graphs showing different values

Description automatically generated

A screenshot of a computer

Description automatically generated A screenshot of numbers

Description automatically generated

A group of graphs with numbers

Description automatically generated*A screenshot of a computer

Description automatically generated*

*Nhận xét* : Dựa vào hình ảnh dưới ta có thể thấy Frequency\_Value, Recency\_Value và Monetary\_Value có 0.65, -0.13 và 0.66 là gần 0 nhất

=> Chọn phương pháp Box-cox transformation

**Kết quả sau khi Transformation:**

A graph of a number of data

Description automatically generated with medium confidence

Recency\_Value, Frequency\_Value, Monetary\_Value đã tuân theo phân phối chuẩn. Thế nhưng, giá trị của chúng lệch nhau rất nhiều. Chính vì sự phân bố giá trị của các yếu tố trong tập dữ liệu và các ảnh hưởng của outlier đến kết quả phân cụm, nên cần quy đổi các giá trị trên về cùng 1 đơn vị.

A screenshot of a graph

Description automatically generated

=> Sử dụng Z-score (Phương pháp phân phối theo hệ số chuẩn standard score)

3.2 Chuẩn hóa Z-score

Với điểm Z-score này sẽ giúp chúng ta hình dung được độ xa của một điểm dữ liệu so với điểm dữ liệu trung bình (điểm chuẩn). Công thức để quy đổi các giá trị theo Z-score như sau:

A mathematical equation with black text

Description automatically generated

Trong đó với x là giá trị của điểm dữ liệu, mean là giá trị trung bình của tập dữ liệu, std (standard deviation) là độ lệch chuẩn của tập dữ liệu.

Phương pháp này giúp chúng ta mô tả lại được chính xác và gần hơn giá trị thực ban đầu của dữ liệu, làm giảm đi khoảng cách chênh lệch lớn giữa các yếu tố trong phương pháp RFM mà không làm thay đổi ý nghĩa ban đầu của dữ liệu.

**Kết quả sau khi chuẩn hóa:**

A screenshot of a graph

Description automatically generated

1. Tiến hành phân khúc khách hàng bằng Kmeans
   1. Lựa chọn số cụm tối ưu.
      * + Sử dụng phương pháp Elbow : minh họa dưới dạng đồ thị đường cong với trục hoành là số K cụm (nghĩa là số phân khúc khách hàng dựa trên giá trị từ mô hình dữ liệu RFM), trục tung là chỉ số SSE (Sum of Errors) – tức chỉ số đo lường sự khác biệt giữa các điểm trong cụm.
        + Tiến hành thực hiện phương pháp Elbow có số cụm từ 1 đến 20 trên mô hình RFM thu kết quả như sau: A graph with a line

          Description automatically generated
        + Kiểm tra lại bằng biểu đồ Silhouse Score:

A graph with blue lines

Description automatically generated

*Nhận xét:*

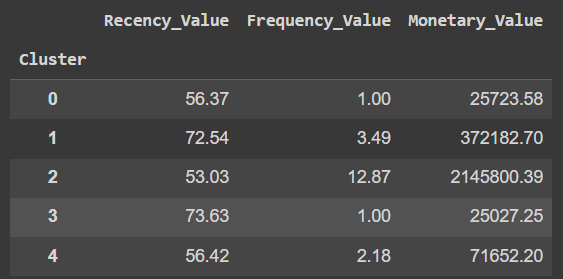
Tại k=5 cho độ chính xác cao nhất => Chọn K =5 là điểm tối ưu cho tham số đầu vào trong phương pháp phân cụm K-means.

* Phân cụm phân khúc khách hàng và trực quan hóa kết quả phân tích
  + - * Số lượng phần tử (khách hàng) trong từng cụm (cluster)

A black background with white numbers

Description automatically generated

* + - * Mean của từng cụm :



Dựa vào bảng mean của các cụm trên có thể thấy:

* + - * **Cluster 0:** Low-value customer: Những khách hàng đóng góp thấp cho hoạt động kinh doanh.
    - Recency\_Value: Trung bình khoảng cách thời gian từ lần mua hàng gần đây là 56.37.
    - Frequency\_Value: Trung bình mua hàng một lần.
    - Monetary\_Value: Trung bình là 25723.58.

Đặc Điểm: Các khách hàng trong nhóm này có thể là những tài khoản giao dịch lần đầu và chưa có nhiều giao dịch. Tổng số tiền mà họ giao dịch (nạp game) không cao.

* + - * **Cluster 1:** High Value Customer: Không thuộc nhóm hàng đầu nhưng vẫn có giá trị, đóng góp lớn đáng kể.
    - Recency\_Value: Trung bình khoảng cách thời gian từ lần giao dịch gần đây là 72.54. (Họ giao dịch cách đây đã lâu)
    - Frequency\_Value: Có giao dịch lại.
    - Monetary\_Value: Trung bình là 372182.70.

Đặc Điểm: Nhóm này có khách hàng có khoảng cách thời gian giao dịch đến hiện tại khá lớn nhưng có sự giao lịch lại và số tiền giao dịch tổng khá cao => Nhóm khách hàng tiềm năng.

* + - * **Cluster 2:** Top Customer: Khách hàng có mức độ gần đây, tần suất và giá trị tiền tệ cao.
    - Recency\_Value: Trung bình khoảng cách thời gian từ lần mua hàng gần đây là 53.03.
    - Frequency\_Value: Trung bình mua hàng 12.87 lần.
    - Monetary\_Value: Trung bình giá trị đơn hàng là 2145800.39.

Đặc Điểm: Nhóm này bao gồm các tài khoản rất thường xuyên giao dịch và có giá trị rất lớn. Đây là nhóm tài khoản trung thành.

* + - * **Cluster 3:** Lost Customer: Khách hàng rất ít giao dịch, giao dịch thấp và rất lâu chưa giao dịch lại, cần nỗ lực tương tác lại.
    - Recency\_Value: Trung bình khoảng cách thời gian từ lần mua hàng gần đây là 71.63.
    - Frequency\_Value: Trung bình mua hàng một lần.
    - Monetary\_Value: Trung bình là 25027.25.

Đặc Điểm: Các khách hàng trong nhóm này có thể là những tài khoản giao dịch lần đầu và chưa có nhiều giao dịch. Tổng số tiền mà họ giao dịch (nạp game) không cao. Bên cạnh đó, rất lâu rồi họ chưa giao dịch lại.

* + - * **Cluster 4:** Medium value customer: Những người đóng góp vững chắc cho hoạt động kinh doanh của doanh nghiệp, chiếm vị trí trung bình.
    - Recency\_Value: Trung bình khoảng cách thời gian từ lần giao dịch gần đây là 56.42. (Họ giao dịch cách đây đã lâu)
    - Frequency\_Value: Có giao dịch lại, trung bình giao dịch hơn 2 lần.
    - Monetary\_Value: Trung bình là 372182.70.

Đặc Điểm: Nhóm này có khách hàng có khoảng cách thời gian giao dịch đến hiện tại khá gần so với các nhóm còn lại, nhưng sự giao lịch lại và số tiền giao dịch tổng chưa cao cao => Nhóm khách hàng bình thường.

* Chiến dịch tiếp thị đưa ra có thể được tập trung vào từng nhóm tài khoản dựa trên đặc điểm và giá trị của họ. Ví dụ, với Cluster 2, có thể chạy các chiến dịch tiếp thị dựa trên giá trị để tạo thêm doanh số. Với Cluster 2, có thể tập trung vào việc duy trì và tạo thêm sự trung thành của họ qua các chương trình thưởng đối với nhóm tài khoản trung thành. ( Những người chơi này có thể được nhận thưởng vào các ngày đặc biệt…)

1. Một số biểu đồ trực quan sau khi đã phân nhóm khách hàng
   1. Tỷ lệ phân bổ của các nhóm

A colorful pie chart with numbers

Description automatically generated

* Phân bổ của Cluster 2 thấp, 15.98%.
* Phân bố của Cluster 1 cũng thấp.
* Cluster 3 và 0 chiếm % rất lớn nhưng giá trị mang lại lại thấp. => Cần đưa ra giải pháp để tăng số lượng của Cluster 5 và 2 lên.
* Phân bổ của Cluster 0+1 chiếm số lượng lớn, nhưng nhóm khách hàng này lại mang lại ít giá trị.
  1. Tỷ lệ account và doanh thu tương ứng theo từng phân khúc

A graph with blue and purple bars

Description automatically generated

A graph with numbers and lines

Description automatically generated

* 1. Phân tích tỷ lệ duy trì khách hàng (Customer Retention)
     + - Sử dụng phương pháp phân tích Cohort (hay còn được hiểu là phân tích theo nhóm một cách tuần tự theo khoảng thời gian). Phương pháp này thường dùng để đo mức độ tương tác của người dùng theo thời gian.

r = số account duy nhất tại thời điểm/ tổng account duy nhất trong Cohort

* + - * Sau khi phân tích, thu được kết quả dưới dạng ma trận và biểu đồ nhiệt

A graph with numbers and a number of percentages

Description automatically generated with medium confidence

* 1. Phân tích package theo từng phân khúc.
     + - Dựa vào link Package theo PackageID:

<https://docs.google.com/spreadsheets/d/1rQqjrrspyUbC0ed9A452OkEloalygmB7MCtjZwucY9I/edit?usp=sharing>

+ Tạo df tính số lượng của từng package:  
<https://drive.google.com/file/d/1-3czw5sNpdACOUgM-C7QCd7vqstfbA-W/view?usp=sharing>

+ Tạo df tính số lượng package theo từng phân khúc:

Từ bảng websale đã được lọc bỏ điều kiện về thời gian và server, chúng ta tạo data\_revenue\_df có dạng:

A screenshot of a computer

Description automatically generated

Link: <https://drive.google.com/file/d/1-211LYaVbbARjGBkxYF5yLVlAI9t1w2J/view?usp=sharing>

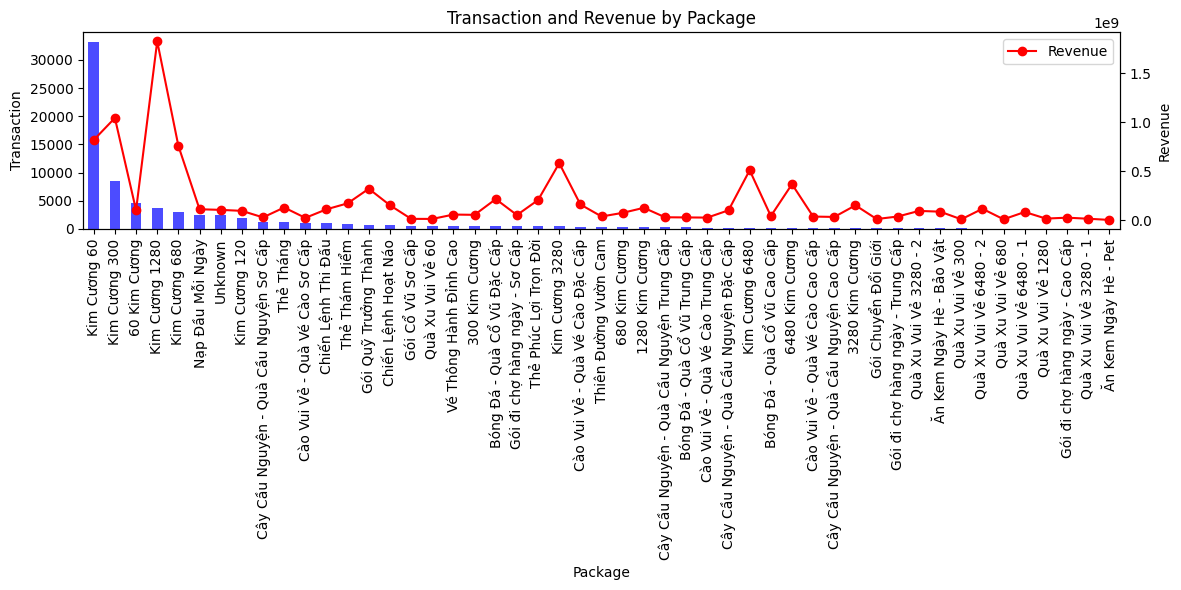
Sau đó, tạo combined\_df bằng cách kết hợp rfm\_df ( bảng tính Recency\_Value, Frequency\_Value và Monetay\_Value của Account và đã nhóm account vào từng phân khúc) với package từ data\_revenue tạo bên trên.

* Biểu đồ thể hiện số lượng của mỗi package theo phân khúc:

A screenshot of a data

Description automatically generated

* Tổng giao dịch và doanh thu theo package



*Nhận xét:*

* + - * Package được trao đổi nhiều nhất là kim cương 60. => Đây là package dễ tiếp cận với các phân khúc. Điều này là dễ hiểu vì đây là gói packgae có giá trị tương đối thấp, còn những package có giá trị cao chỉ tiếp cận được với những phân khúc người chơi cao hơn.
      * Ở phân khúc vip, package được mua nhiều cũng là kim cương 60, ngoài ra còn có các package như: kim cương 7196, … => Những phân khúc cao có sự giao dịch các gói package đa dạng hơn, số lượng mỗi package giao dịch cũng nhiều hơn.

**Formulation of hypotheses**

* Level của người chơi liệu có ảnh hưởng gì ( Sử dụng bảng user\_level.csv)

**A graph of blue squares

Description automatically generated with medium confidence**

Level 1 có lượng người chơi nhiều nhất, sau đó từ level 2 lượng người chơi giảm rất lớn so với level 1. Nếu lượng user chỉ dừng ở level 1 thì họ sẽ không nạp tiền vào game nhiều, ảnh hưởng đến doanh thu. => Cần tìm ra lý do, đặt các giả thuyết như:

* Khó thắng => có thể khiến người chơi nhanh chóng mất hứng thú và rời khỏi game.
* Game chưa đủ thu hút nên người chơi chỉ vào chơi thử rồi out game.

### Phương pháp 2: Tính theo ranking

* Chọn current\_data là sau ngày đặt hàng cuối 1 ngày

1. Tính Recency, Frequency, Monetary như công thức của c1.
2. Chia khoảng cho Recency, Frequency, Monetary.

* Với Recency:

Chia thành các nhóm theo từng tuần; Recency thuộc từ ngày 1->8, 8-> 15, 15 ->22, 22-> 29, 29 -> 34. Biểu đồ thể hiện số lượng Recency mỗi tuần.

A graph of blue bars

Description automatically generated

*Nhận xét*: Số lượng lần giao dịch cuối của người chơi thường cách đây hơn 3 tuần.

* Với frequency:

A black and white image of numbers

Description automatically generated

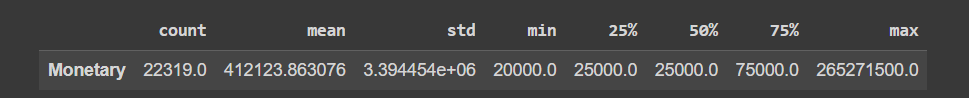
*Nhận xét:* Các giá trị của freequency 50% là 1 => Chỉ giao dịch 1 lần chiếm số lượng rất lớn, ở mức 75% vẫn là những người chơi giao dịch 1-2 lần, và trung bình số giao dịch chỉ hơn 3. Bên cạnh đó max lại lên tới 1232 lần và chỉ có 1 người giao dịch nhiều như vậy.

Dựa vào đó, chúng ta có thể chia Frequency thành các khoảng như biểu đồ bên dưới.

A graph with blue squares

Description automatically generated

* Monetaty:



*Nhận xét:* Tương tự như frequency, số lượng giao dịch của 1 tài khoản chỉ ở mức 20k – 75k. Mức giao dịch trung bình của 1 tài khoản là 412k, trong khi đó max lên tới 265271500 (hơn 265 triệu) , lớn gấp nhiều lần so với mức trung bình. Vì thế, tương tự frequency, với monetary, chúng ta không lên chia khoảng giá trị thành các phần bằng nhau.

A graph with blue squares

Description automatically generated

**Chú ý:** Cách chia này chỉ là ước lượng, khi bên marketing có định nghĩa với các nhóm khách hàng cụ thể ở khoảng nào, chúng ta có thể thay đổi khoảng chia lại sao cho phù hợp với yêu cầu của họ.

1. Tính toán và xếp hạng khách hàng dựa trên các chỉ số Recency, Frequency và Monetary (RFM)

* Xếp hạng
  + - * R\_rank: Tính dựa trên cột Recency: Xếp hạng lớn nhất sẽ có giá trị thấp nhất.
      * F\_rank, M\_rank tính dựa trên cột Frequency, Monetary: Xếp hạng nhỏ nhất sẽ có giá trị thấp nhất.
* Chuẩn hóa xếp hạng:
  + - * Bằng cách chia cho rank lớn nhất và nhân với 100 để giá trị R\_rank, F\_rank, M\_rank thuộc trong khoảng từ 0 đến 100.
* Tính RFM\_score :
  + - * Từ R\_rank\_norm, F\_rank\_norm, M\_rank\_norm ( trị R\_rank, F\_rank, M\_rank được chuẩn hóa ở bên trên). Ở đây, em lấy tỷ lệ của R, F, M là 1/3 ( ảnh hưởng của R, F, M là như nhau đối với việc phân khúc khách hàng).
      * Sau đó đưa RFM\_score về trong khoảng [0,5] bằng cách nhân RFM\_score với 0.05 để tiện cho việc phân nhóm khách hàng.

**Chú ý:** Tỷ lệ này có thể thay đổi, do có thể mỗi công ty có sự quan tâm khác nhau ở các giá trị R, F, M. Cũng có thể chỉ sử dụng R và F, F và M,… chứ không nhất thiết phải sử dụng cả 3 yếu tố R, F, M.

1. Phân khúc khách hàng:

Dựa trên các mức của RFM\_score chia khách hàng thành các phân khúc:

* rfm score > 4.5 **: Top Customer**: Khách hàng có mức độ gần đây, tần suất và giá trị tiền tệ cao.
* 4.5 > rfm score > 4 **: High Value Customer**: Không thuộc nhóm hàng đầu nhưng vẫn có giá trị, đóng góp lớn đáng kể.
* 4 > rfm score > 3 : **Medium value customer**: Những người đóng góp vững chắc cho hoạt động kinh doanh của doanh nghiệp, chiếm vị trí trung bình.
* 3 > rfm score > 1.6 **: Low-value customer**: Những khách hàng đóng góp hơi thấp cho hoạt động kinh doanh.
* rfm score < 1.6 **: Lost Customer**: Khách hàng rất ít giao dịch, rất lâu chưa giao dịch lại, cần nỗ lực tương tác lại.
* Với mỗi nhóm khách hàng, cần đưa ra những chiến dịch phù hợp.

1. Biểu đồ thể hiện 1 số kết quả sau khi phân khúc

* Phân bổ phần trăm của các phân khúc

A pie chart with numbers and text

Description automatically generated

*Nhận xét*:

* Phân khúc Top Customers và High Value còn thấp.
* Phân khúc Lost Customer và Low Value Customers lại chiếm tỷ lệ rất lớn.
* Tỷ lệ account và doanh thu theo phân khúc

*Nhận xét:*

* Phân khúc Top Customers và High Value chiếm số lượng ít nhưng lại mang lại doanh thu lớn.
* Phân khúc Lost Customer và Low Value Customers lại chiếm tỷ lệ số lượng rất lớn nhưng doanh thu từ các phân khúc này mang lại không cao.
* Cần có biện pháp để tăng số lượng ở các phân khúc mang lại giá trị lớn.

A graph with blue lines and numbers

Description automatically generatedA graph with blue rectangles

Description automatically generated

**Sử dụng cách khác để tính Recency\_Value**

Ngoài cách tính Requency bằng Topupup, chúng ta có thể sử dụng các yếu tố khác để tính. Ví dụ như: LoginDate – VwAccountLogin (Thời gian đăng nhập của tài khoản).

* + - * Sắp xếp lại VwAccountLogin với điều kiện: Cùng 1 AccountID sẽ có LoginDate được sắp xếp tăng dần từ nhỏ đến lớn.
      * Tính recency bằng cách:
        + Chọn current\_data là sau ngày đặt hàng cuối 1 ngày.
        + Tính theo công thức như ở cách 1.
* df\_frequency và df\_monetary lấy từ cách 2.
* Sau đó, kết hợp Recency, Frequency và Monetary dựa trên 'AccountID' để tạo thành RFM\_score . Tiến hành phân cụm khách hàng theo 1 trong 2 cách trên.

## Các biến thể của RFM

**RFD (Recency, Frequency, Duration):** Thay vì phân tích giá trị khách hàng( ở đây là người chơi), chúng ta phân tích hành vi của họ.Có thể tính dựa trên(lần truy cập gần đây, tần suất, thời lượng). RFD khác RFM: tổng thời lượng người dùng chơi game thay vì tổng số tiền người dùng nạp vào game.

**RFE (Recency, Frequency, Engagement):** Là mở rộng của RFD, Engagement (E): Đo mức độ tham gia hoặc tương tác của người chơi với game, bao gồm việc tham gia vào cộng đồng, tạo nội dung, hoặc tham gia sự kiện.

**RFM-I (Recency, Frequency, Monetary Value- Interactions):** Tương tự RFM nhưng có thêm yếu tố Interactions. Interaction (I): Thể hiện mức độ tương tác của người chơi với trò chơi, bao gồm việc tham gia vào các sự kiện trong trò chơi, gửi phản hồi, tương tác với cộng đồng game,….

**RFMTC (Recency, Frequency, Monetary Value, Time, Churn rate):** Yếu tố "Time có thể liên quan đến thời gian thực hiện giao dịch, thời gian chơi, hoặc thời gian tích luỹ thành tích trong trò chơi. Nó có thể cung cấp thông tin về sự cam kết và đóng góp của người chơi. Churn rate (Churn - C): Có thể đo tỷ lệ người chơi bỏ cuộc hoặc dừng tham gia trò chơi. Nó biểu thị khả năng của người chơi tiếp tục tương tác với trò chơi trong tương lai.