**BÀI TEST VỊ TRÍ TRỢ LÝ PHÂN TÍCH DỮ LIỆU**

[**Link file bài làm**](https://drive.google.com/drive/folders/1wcdbKKaM8y4hbK3bN_94--vD2RLkIM2J?usp=sharing)*(Link này em có để file dữ liệu đã xử lý và file PowerBI bài làm của mình)*

**Phần I: Xử lý dữ liệu**

Dữ liệu mẫu có các ô không chứa thông tin, các cột payment\_status, total\_fee, actually\_received có thể bị sai thông tin

**\*Xử lý cột cus\_mode\_id**

-Ở bảng mode chỉ có 4 phương thức quảng cáo với id lần lượt là 1,2,3,4 nhưng trong bảng dữ liệu order của khách hàng lại chứa 1 giá trị 0 và nhiều giá trị trống.

+Cách xử lý:Tạo thêm 1 dòng với id = 0, mode = “Unknown” trong bảng mode

+Tìm kiếm các khách hàng mua hàng nhiều lần, lấy thông tin ở lần mua hàng không có giá trị rỗng để điền vào ô trống trong các lần mua hàng khác.

+Đổi cách giá trị trống còn lại trong bảng order thành 0

**payment\_amount**: Số tiền khách đã thanh toán,

**total\_fee**:tổng số phí mà khách hàng phải trả

**actually\_received**: số tiền thực tế công ty nhận được

**\*Xử lý cột payment\_status:**

- **Các khách hàng đã đặt đơn nhưng không thanh toán** (payment\_amount = 0)

+sẽ thay payment\_status = past due

+cột acctually\_received =0, total\_fee = 0

+có thể cắt toàn bộ số khách hàng này ra khỏi order, lưu vào một data khác để khai thác tiếp hoạt động kinh doanh

- **Các khách hàng đã thanh toán nhưng cột payment\_status còn trống:**

+Nếu payment\_amount=total\_fee thì payment\_status=Fully paid

+Nếu payment\_amount<total\_fee thì payment\_status=deposit

**-Các khách hàng đã thanh toán, cột payment\_status=past due:**

+Kiểm tra lại bằng cách nếu payment\_amount=total\_fee thì đổi thành fully-paid

+Tạo thêm 1 cột tính remain\_payment=total\_fee-payment\_amount

+Nếu =0 thì nhận fully-paid

+Nếu >0 thì past due

+Nếu <0 nghĩa là khách hàng đang thanh toán thừa, nhận fully-paid

**-Các khách hàng đã thanh toán, cột payment\_status=deposit**

+Kiểm tra lại nếu payment\_amount=total\_fee thì đổi thành fully-paid

**-Các khách hàng đã thanh toán, cột payment\_status = fully-paid**

+Tính remain\_payment

+Nếu>0 đổi payment\_status =deposit

+Riêng có khách hàng thanh toán thiếu 1đ=>có thể bỏ qua, vẫn để là fully-paid

**\*Xử lý cột actually\_received:**

-Tính chênh lệch giữa (total\_fee - actually\_received):

+Nếu <0 thì thay thế actually\_received=total\_fee. Vì coi như khoản khách hàng trả thừa sẽ kh tính vào doanh thu

+Với các giá trị = 0, đối chiếu sang cột payment\_status. Nếu payment\_status = deposit thì đổi cột actually\_received=payment\_amount

+Với các giá trị >0, thì actually\_received=payment\_account

- Các ô trống của cột actually\_received và total\_fee thay bằng payment\_amount

**\*Xử lý cột cus\_location\_id:**

-Thay thế các ô trống bằng giá trị ‘5’ (Toàn quốc)

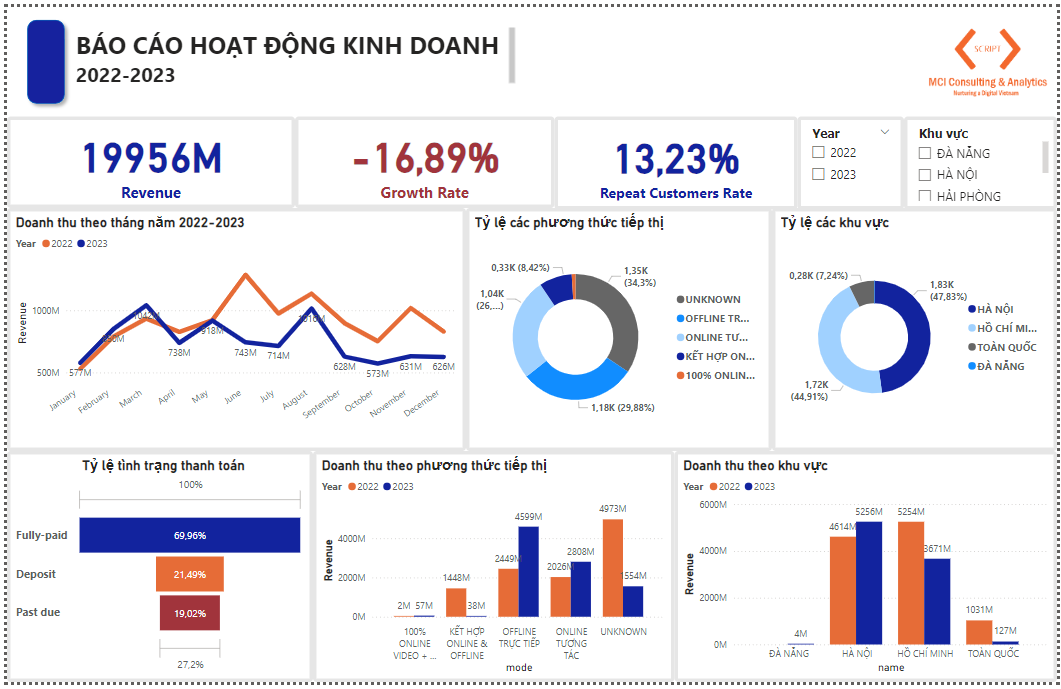
**\*Xử lý cột lead\_source, payment\_method:**

-Thay thế các giá trị trống bằng đối tượng xuất hiện nhiều nhất là “FB”

**\*Các cột còn lại cần có thêm thông tin mô tả về dữ liệu**

**Phần II.Phân tích dữ liệu**

**1.Lập dashboard**

****

**\*Tổng quan:**

- Tổng doanh thu từ đầu năm 2022 đến cuối năm 2023 là hơn 19 tỷ 956 triệu

- Tuy nhiên tỷ lệ tăng trưởng năm 2023 so với năm 2022 lại giảm đi 16.89%

- Trong tổng số khách hàng mà công ty sở hữu, tỷ lệ khách hàng quay lại là 13.23%

**\*Chi tiết:**

*- Biểu đồ thứ nhất: Doanh thu theo tháng năm 2022-2023*

+ Tuy doanh thu từ tháng 1 đến giữa tháng 3 của năm 2023 cao hơn năm 2022 nhưng từ cuối tháng 3 đến hết năm thì doanh thu năm 2023 luôn nhỏ hơn năm 2022

+ Vì thế dẫn tới tỷ lệ tăng trưởng lại giảm như trên

* *Biểu đồ thứ 2:Tỷ lệ tình trạng thanh toán*

+ Có tới 69.96% khách hàng đã thanh toán hết đơn hàng họ mua, 21.49% khách hàng đã đặt cọc

+ Các khách hàng còn lại đã thanh toán một phần nhưng khoản còn lại đã hết hạn thanh toán. Những người này có thể trong lúc đợi hay dùng thử sản phẩm không ưng ý nên đã bỏ ngang việc mua hàng.

* *Biểu đồ thứ 3,4: Liên quan đến các hình thức tiếp thị*

+Có 2 phương thức tiếp thị hiệu quả nhất đó là Online tương tác, Offline trực tiếp. Doanh thu năm 2023 của 2 phương thức này đều tăng so với năm 2022

+Tuy nhiên có tới hơn 1300 khách hàng không có thông tin về cách tiếp thị. Doanh thu của nhóm khách hàng này cao nhất vào năm 2022 là 4.79 tỷ, nhưng đến năm 2023 lại giảm còn 554 triệu.

*-Biểu đồ thứ 5,6: Liên quan đến khu vực sinh sống của khách hàng*

+Không có khách hàng nào đến từ Hải Phòng

+2 thành phố có lượng khách hàng nhiều nhất là Hà Nội, HCM chiếm hơn 90% tổng số khách hàng.

+Vì thế doanh thu mà 2 thành phố này đem lại cũng nhiều nhất. Ở Hà Nội doanh thu có tăng lên từ năm 2022-2023, trong khi Hồ Chí Minh lại giảm đi đáng kể.

**2. Mô hình RFM phân khúc khách hàng**

Phân tích RFM là một mô hình phân khúc khách hàng theo các đặc điểm hành tiêu dùng dựa trên các dữ liệu giao dịch trong lịch sử

-Recency: Khoảng thời gian kể từ lần mua hàng gần đây nhất của khách hàng

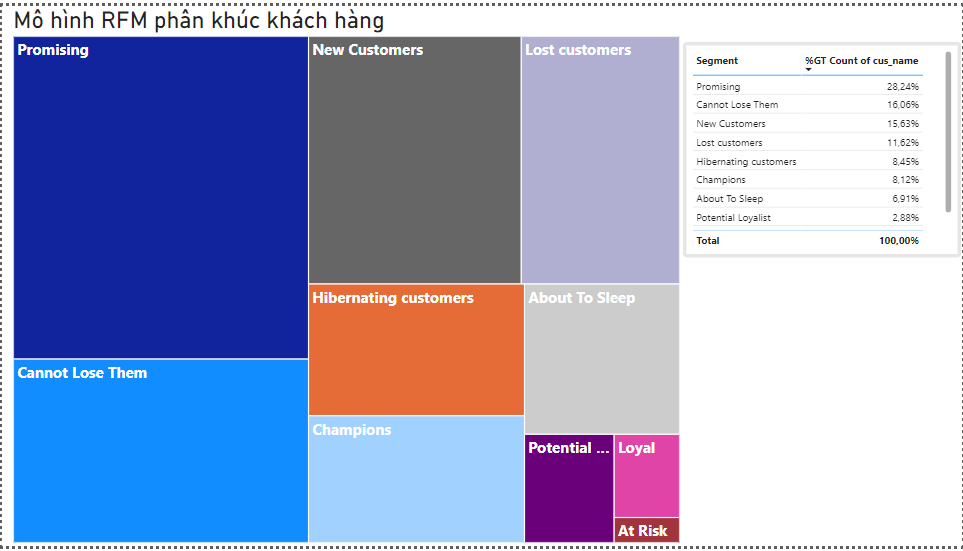
-Frequency:Tần suất giao dịch của khách hàng (trong khoảng thời gian nghiên cứu)

-Montetary:Tổng số tiền mà khách hàng đã chi tiêu qua tất cả giao dịch hay là tổng doanh thu mà công ty nhận được từ mỗi khách hàng

Sau khi tính 3 chỉ số trên, sắp xếp theo thứ tự và chia dữ liệu thành 5 nhóm bằng nhau. Xếp hạng những khách hàng theo giá trị Recency với thang đo 5-4-3-2-1 và giá trị Frenquency, Montetary với thang đo 1-2-3-4-5. Cuối cùng tính RFM-Score và phân loại khách hàng vào các phân khúc như bảng dưới đây



Visual phân khúc khách hàng trên Power BI



-Trong bộ dữ liệu này, phân khúc Promising chiếm tới 28.24% và phân khúc New Customers chiếm 15.63%, đây là các khách hàng mới mua hàng gần đây và tiềm năng sức mua có thể lớn nhưng chưa thường xuyên.

Với tệp khách hàng này, doanh nghiệp phải làm thế nào để giúp họ hài lòng từ giao dịch đầu tiên, và quay lại mua hàng. Doanh nghiệp sẽ cần quan tâm đến dịch vụ chăm sóc khách hàng (Lời cảm ơn, lấy ý kiến về trải nghiệm mua hàng,..) kèm theo những ưu đãi cho những lần mua hàng tiếp theo. Sau đó thường xuyên đề xuất các sản phẩm tương tự với lần mua hàng trước đó của khách hàng, cross-sell/upsell với những gói hàng hấp dẫn.

-Đối với tệp khách Cannot Lose Them là những khách hàng từng mua hàng thường xuyên và chi tiêu nhiều. Doanh nghiệp cần tái kích hoạt để giữ chân họ với các chương trình cho khách hàng thân thiết cũng như đi kèm với việc khảo sát ý kiến.

-3 Phân khúc Lost Customers, Hibernating customers và About to sleep chiếm 26.98% tổng số khách hàng, đây là những khách hàng đã lâu không quay lại, tần suất không thường xuyên, sức mua yếu. Lí do có thể trải nghiệm mua hàng của họ không tốt hoặc đã tìm được sản phẩm thay thế khác. Vì thế doanh nghiệp có thể kết nối lại với những khách hàng này, khảo sát trải nghiệm của họ, có những chương trình trải nghiệm thử sản phẩm và ưu đãi.

-Tệp khách hàng Champions 8.12% là những khách hàng mới giao dịch, mua hàng thường xuyên và chi tiêu nhiều. Doanh nghiệp có thể không cần chương trình giảm giá với tệp khách hàng này mà thay vào đó đề xuất những mặt hàng giá trị hơn hay những combo sản phẩm dựa trên lịch sử mua hàng của họ.