**BÁO CÁO PHÂN TÍCH NHÓM KHÁCH HÀNG TẠI MỘT CỬA HÀNG ĐỒ LƯU NIỆM ONLINE**

**Chủ đề: Customer Relationship Management**

**Người thực hiện : Đặng Trung Kiên**

**Hà Nội – 2023**

**MỤC LỤC**

[**1. Mô tả bài toán** 1](#_Toc135053357)

[**2. Chuẩn bị dữ liệu** 3](#_Toc135053358)

[**2.1 Mô tả bộ dữ liệu** 3](#_Toc135053359)

[**2.2 Xử lý dữ liệu thô** 4](#_Toc135053360)

[**3. Xây dựng kho dữ liệu** 9](#_Toc135053361)

[**3.1 Xác định các đơn vị dữ liệu** 9](#_Toc135053362)

[**3.2 Chọn các chiều dữ liệu (dimension)** 9](#_Toc135053363)

[**3.3 Thiết kế bảng Fact** 12](#_Toc135053364)

[**3.4 Ràng buộc quan hệ giữa các bảng chiều và bảng Fact** 13](#_Toc135053365)

[**3.5 Nhập dữ liệu vào kho dữ liệu** 13](#_Toc135053366)

[**4. Trực quan hóa và khai thác bộ dữ liệu** 14](#_Toc135053367)

[**4.1 Kết nối data warehouse với PowerBI** 14](#_Toc135053368)

[**4.2 Phân tích bộ dữ liệu** 15](#_Toc135053369)

[**4.2.1 Phân tích tổng quan** 15](#_Toc135053370)

[**4.2.2 Phân tích chi tiết tập trung vào khách hàng** 17](#_Toc135053371)

[**5. Kết Luận** 24](#_Toc135053372)

[**Bộ dữ liệu sử dụng trong bài báo cáo** 25](#_Toc135053373)

[**Tài liệu tham khảo** 25](#_Toc135053374)

# **1. Mô tả bài toán**

Customer Relationship Management (CRM) là một chiến lược kinh doanh dựa trên công nghệ, tập trung vào quản lý mối quan hệ khách hàng. Nó giúp các doanh nghiệp tăng cường tương tác với khách hàng, nâng cao chất lượng dịch vụ và tăng doanh số bán hàng.

CRM bao gồm một hệ thống quản lý thông tin khách hàng, quản lý quan hệ khách hàng, và quản lý tương tác khách hàng. Hệ thống quản lý thông tin khách hàng (Customer Information System - CIS) lưu trữ thông tin khách hàng như tên, địa chỉ, số điện thoại, email, lịch sử mua hàng, các hoạt động tương tác, giúp các doanh nghiệp hiểu rõ hơn về khách hàng của mình. Quản lý quan hệ khách hàng (Customer Relationship Management - CRM) bao gồm các chiến lược để giữ chân khách hàng, xây dựng mối quan hệ lâu dài và đáp ứng nhu cầu khách hàng. Quản lý tương tác khách hàng (Customer Interaction Management - CIM) giúp các doanh nghiệp tương tác với khách hàng thông qua nhiều kênh khác nhau như email, điện thoại, chat trực tuyến, mạng xã hội, giúp tăng cường tương tác và nâng cao trải nghiệm khách hàng.

Các ứng dụng của CRM rất đa dạng. Một số ứng dụng nổi trội có thể kể đến như: Quản lý bán hàng, Quản lý dịch vụ khách hàng, Marketing, Quản lý chiến dịch, Quản lý chất lượng.

Nhu cầu của CRM: Các nhu cầu của CRM thường liên quan đến việc tối ưu hóa quá trình quản lý mối quan hệ khách hàng và nâng cao chất lượng dịch vụ, đáp ứng nhu cầu của khách hàng. Cụ thể, một số nhu cầu của CRM bao gồm:

* Tăng cường tương tác khách hàng: Các doanh nghiệp cần nâng cao khả năng tương tác với khách hàng trên nhiều kênh khác nhau để tăng cường sự gắn kết và sự hài lòng của khách hàng.
* Phát triển chiến lược marketing hiệu quả: Các doanh nghiệp cần phát triển chiến lược marketing phù hợp với nhu cầu của khách hàng để tăng cường tương tác và thu hút khách hàng.
* Nâng cao khả năng quản lý thông tin khách hàng: Các doanh nghiệp cần quản lý thông tin khách hàng một cách chuyên nghiệp để tối ưu hóa quá trình quản lý mối quan hệ khách hàng.
* Tối ưu hóa quy trình bán hàng: Các doanh nghiệp cần tối ưu hóa quy trình bán hàng để tăng cường hiệu quả kinh doanh và tăng doanh số bán hàng.
* Quản lý chiến dịch marketing: Các doanh nghiệp cần quản lý các chiến dịch marketing để đạt được hiệu quả kinh doanh cao nhất.
* Tăng cường khả năng giám sát và đánh giá hiệu quả của hoạt động kinh doanh: Các doanh nghiệp cần giám sát và đánh giá hiệu quả của hoạt động kinh doanh để tối ưu hóa các chiến lược kinh doanh và tăng cường hiệu quả kinh doanh.

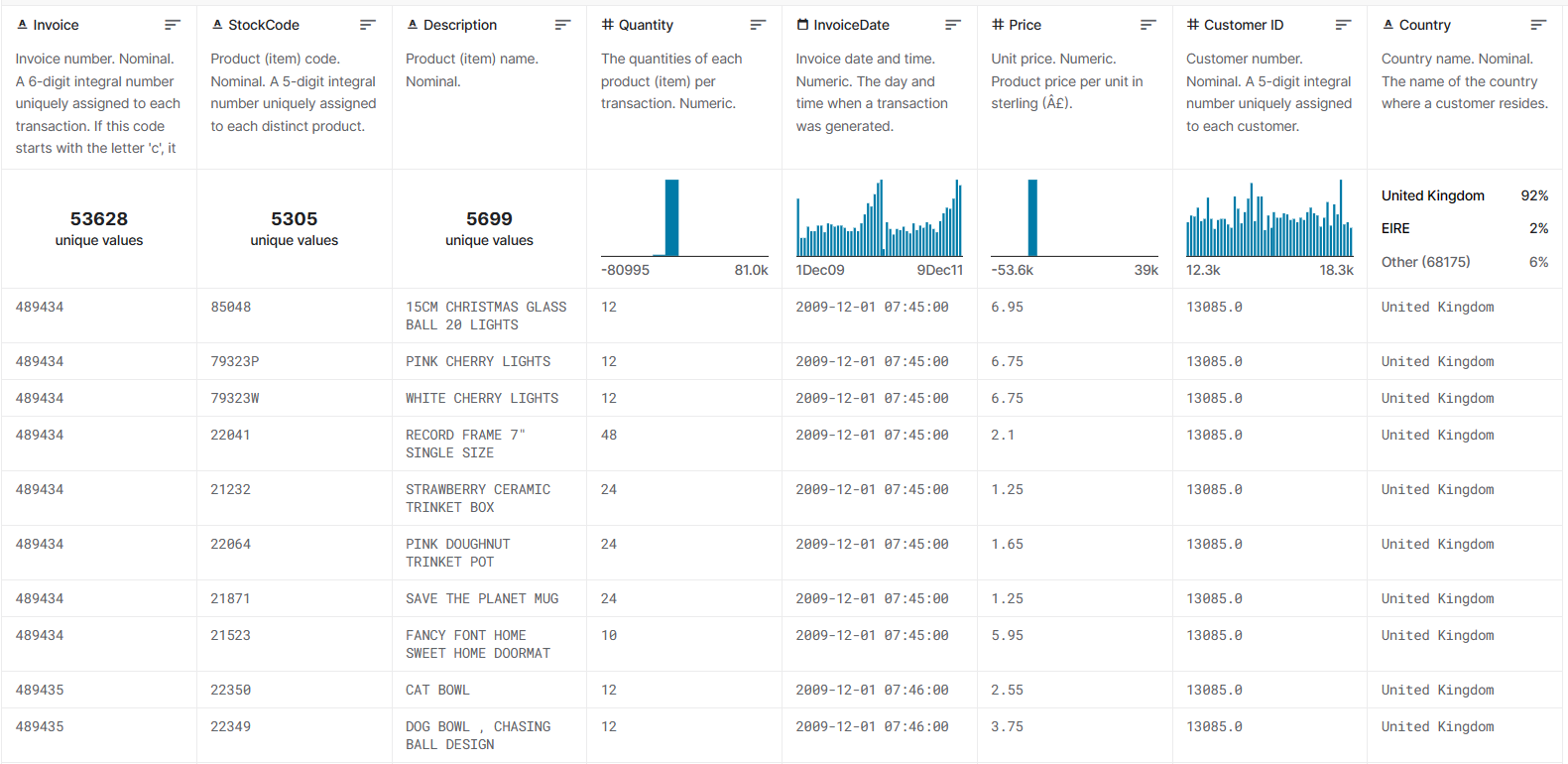
Với việc phân nhóm khách hàng kết hợp cùng thông tin có sẵn trong bộ dữ liệu, các câu hỏi được đặt ra để hiểu hơn về phân khúc khách hàng cũng như thói quen của họ trong bộ dữ liệu như:

* Nhóm khách hàng chính của cửa hàng là những ai? Những khách hàng trung thành, khách mới hay những khách hàng khác?
* Những khách hàng đó đến từ những đâu? Những khách hàng mới, khách hàng trung thành họ thường đến từ những đâu?
* Những khách hàng đó họ tìm kiếm điều gì ở cửa hàng? Những nhóm khách hàng thường mua những gì từ cửa hàng, liệu có gì khác biệt giữa các nhóm không?
* Thói quen mua sắm của họ như thế nào? Họ có xu hướng mua vào những ngày, giờ hay tháng đặc biệt nào không?

# **2. Chuẩn bị dữ liệu**

## **2.1 Mô tả bộ dữ liệu**

Dữ liệu trong dự án được lấy từ bộ dữ liệu Online Retail II UCI [[1]](#footnote-1) được cung cấp trên kaggle bởi tác giả MIYABON. Retail II UCI là bộ dữ liệu chứa thông tin giao dịch của khách hàng tại một cửa hàng bán lẻ online ở Vương Quốc Anh từ 01/12/2009 đến 09/12/2011. Với sản phẩm chủ yếu là quà tặng và khách hàng có thể bao gồm các nhà bán buôn.



***Hình 1****: Tổng quan về bộ dữ liệu*

**Invoice**: Gồm 6 số chỉ định cho duy nhất mỗi giao dịch

**StockCode**: Gồm 5 số biểu thị cho mỗi sản phẩm riêng biệt

**Description**: Mô tả sản phẩm

**InvoiceDate**: Ngày tạo hóa đơn

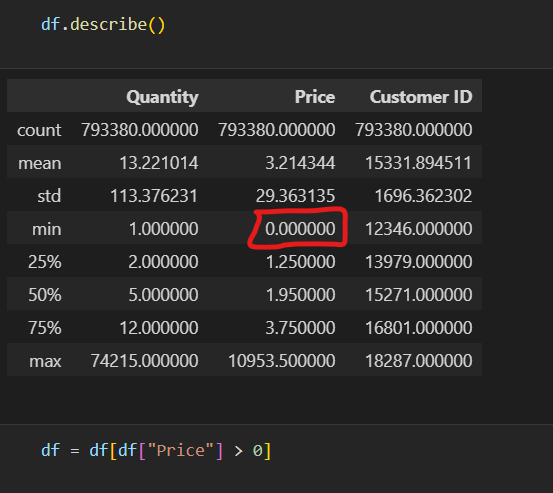
**Price**: giá của một sản phẩm tương ứng

**CustomerID**: mã định danh cho mỗi người dùng

**Country**: Đất nước mà khách hàng cư trú

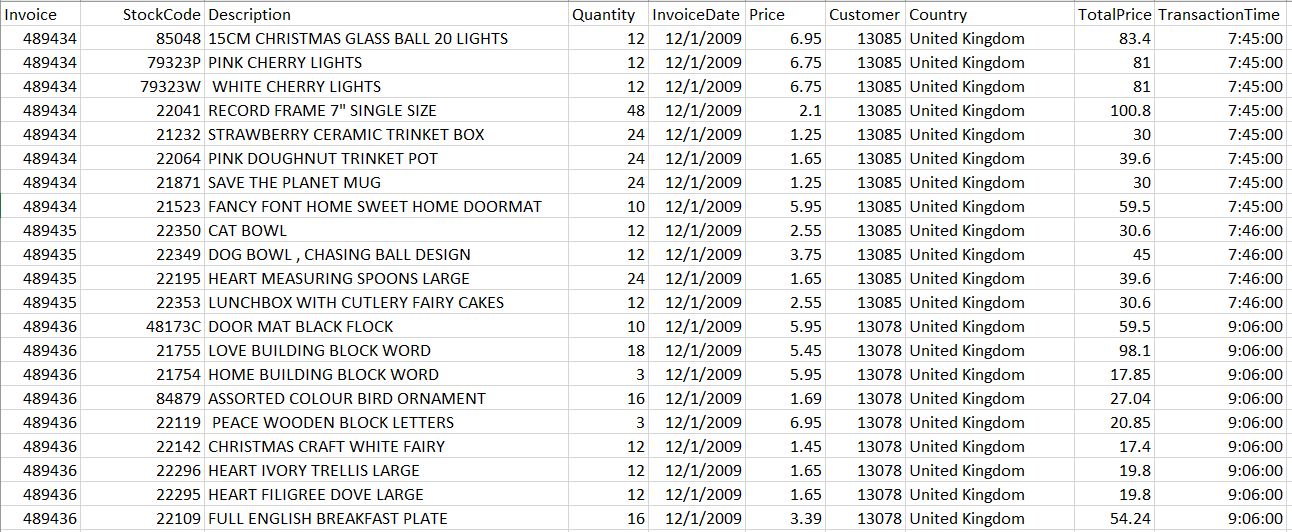
## **2.2 Xử lý dữ liệu thô**

Sử dụng python loại bỏ dữ liệu NaN, dữ liệu nhiễu ra khỏi bộ dữ liệu. Ở hình dưới đây có thể thấy có những sản phẩm có giá bằng 0 điều này có thể ảnh hưởng đến việc tính toán sau này nên các sản phẩm này sẽ bị loại bỏ.



***Hình 2****: Loại bỏ dữ liệu nhiễu*

Sau đó thêm cột ToltalPrice (Quantity \* Price), tách cột InvoiceDate từ định dạng datetime thành định dạng date và time riêng biệt. Sau khi hoàn thành xử lý dữ liệu với Python, xuất dữ liệu ra dạng flat file để tiến hành bước xử lý tiếp theo.



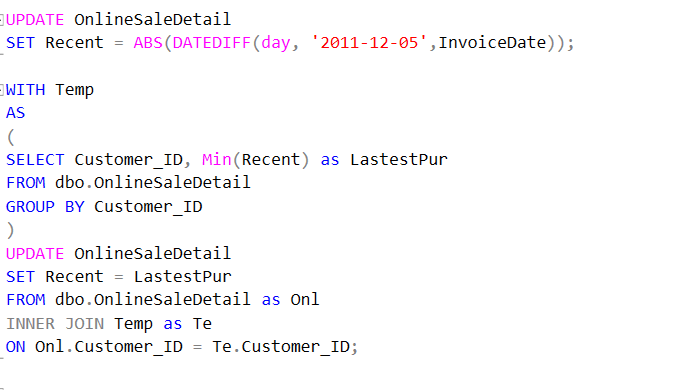
***Hình 3****: Tổng quan về dữ liệu sau khi loại bỏ dữ liệu nhiễu với Python*

Tạo một database để lưu dữ liệu chưa xử lý, tiến hành thêm dữ liệu vào như một bảng trong database. Thêm các cột Frequency, Recent, FrequencyRating, RecentRating, MonetaryRating, RFM\_Score phục vụ cho phân khúc khách hàng và phân tích RFM. Trong đó:

Frequency: Số lần mỗi người dùng mua hàng tại cửa hàng

Recent: Số ngày kể từ ngày gần nhất người dùng mua hàng tại cửa hàng

MonetaryRating: Xếp hạng dựa trên tổng số tiền mà một khách hàng đã chi tiêu



***Hình 4****: Cập nhật dữ liệu cho cột Recent*

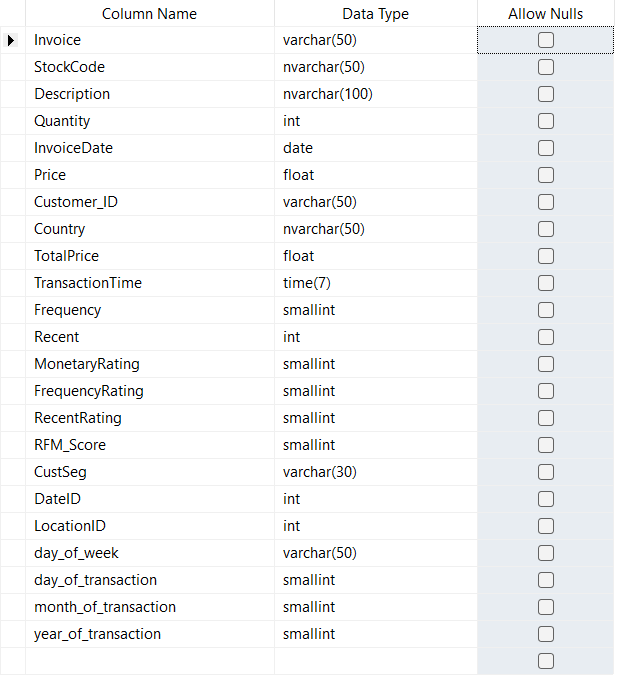
Sau khi có được các cột trên tiến hành tính điểm RFM và phân loại khách hàng thành các nhóm khách hàng khác nhau

|  |  |
| --- | --- |
| VIP | Nhóm khách hàng mua gần đây, thường xuyên và chi nhiều nhất |
| Loyal | Nhóm khách hàng chi số tiền khá lớn và mua thường xuyên |
| Potential | Mới mua sắm gần nhưng chi số tiền khá lớn và mua nhiều hơn 1 lần |
| Recent | Mới mua gần đây nhưng không thường xuyên |
| Promising | Mới mua gần đây nhưng không chi tiêu quá nhiều |
| Need Attention | Nhóm khách hàng có điểm RFM chung trên trung bình |
| Sleep | Nhóm khách hàng có điểm RFM chung trên dưới bình |
| At Risk | Nhóm đã từng mua nhiều lần với số tiền lớn |
| Can’t lose | Nhóm khách hàng tiêu số tiền lớn và thường xuyên nhưng đã lâu không mua hàng |
| Hibernating | Nhóm khách hàng đã lâu không mua hàng, chi tiêu thấp và không mua hàng nhiều |
| Lost | Nhóm khách hàng có điểm RFM thấp nhất |

***Bảng 1****: Phân nhóm khách hàng*

Sau khi phân nhóm khách hàng, tiếp tục tách nhỏ ngày tạo đơn thành ngày, tháng, năm và ngày trong tuần, thêm ID cho ngày và địa điểm để phục vụ việc thiết kế data warehouse. Tuy nhiên, với cùng một id sản phẩm tồn tại nhiều description khách nhau, nên sau đó tiến hành đánh lại id sản phẩm theo từng description riêng biệt.

Sau cùng ta sẽ thu bộ dữ liệu đã được làm sạch và sẵn sàng cho việc thiết kế data warehouse với cấu trúc như hình 5.



***Hình 5****: Cấu trúc bộ dữ liệu sau khi đã được điều chỉnh*

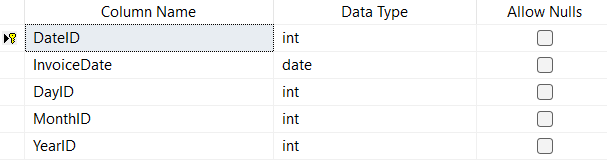
# **3. Xây dựng kho dữ liệu**

## **3.1 Xác định các đơn vị dữ liệu**

Bảng Fact lưu trữ thông tin về giao dịch của khách hàng với sản phẩm cụ thể. Với mỗi dòng dữ liệu trong bảng Fact tương ứng với một lần khách hàng tiến hành giao dịch một sản phẩm cụ thể tại cửa hàng.

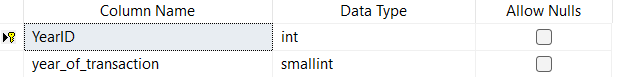
## **3.2 Chọn các chiều dữ liệu (dimension)**

**Chiều thời gian(dim\_date)**: cung cấp thông tin chi tiết về thời gian mà giao dịch được thực hiện như: ngày trong tuần, ngày, tháng, năm.



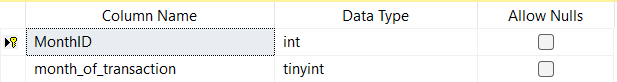
***Hình 6****: Cấu trúc chiều thời gian*

**Chiều năm(dim\_year)**: Chứa thông tin về năm thực hiện giao dịch



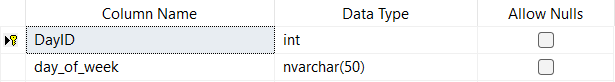
***Hình 7****: Cấu trúc chiều năm giao dịch*

**Chiều tháng(dim\_month)**: chứa thông tin về tháng thực hiện giao dịch



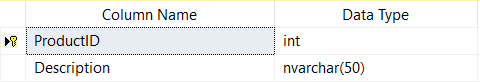
***Hình 8****: Cấu trúc chiều tháng giao dịch*

**Chiều ngày(dim\_day)**: chứa thông tin về ngày thực hiện giao dịch trong tuần

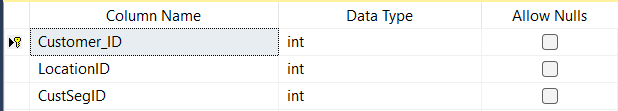


***Hình 9****: Cấu trúc chiều tháng giao dịch*

**Chiều sản phẩm(dim\_product)**: Chứa thông tin liên quan đến một sản phẩm nhất định như productID, description.

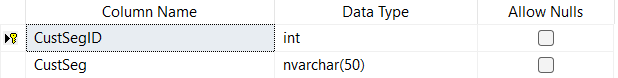
***Hình 10****: Cấu trúc chiều sản phẩm*

**Chiều khách hàng(dim\_customer)**: Chứa thông tin liên quan đến một khách hàng.

******

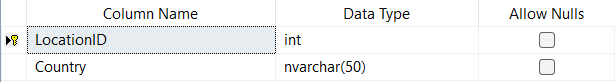
***Hình 11****: Cấu trúc chiều khách hàng*

**Chiều phân loại khách hàng(dim\_customer\_segmentation)**: Chứa thông tin liên quan đến phân loại của một khách hàng được tính bằng chỉ số RFM.

******

***Hình 12****: Cấu trúc chiều phân loại khách hàng*

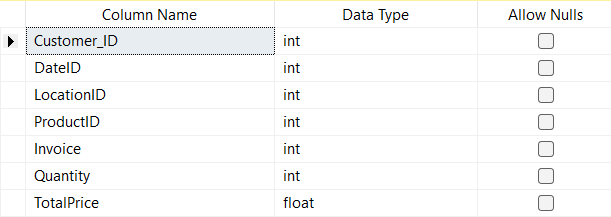
**Chiều địa điểm của khách hàng(dim\_location)**: Chứa thông tin liên quan đến phân loại của một khách hàng được tính bằng chỉ số RFM.

******

***Hình 13****: Cấu trúc chiều địa điểm khách hàng*

## **3.3 Thiết kế bảng Fact**

Bảng Fact sẽ lưu trữ thông tin về một giao dịch của từng khách hàng với mỗi sản phẩm cụ thể với đầy đủ thông tin về thời gian, số lượng, tổng giá trị của đơn.



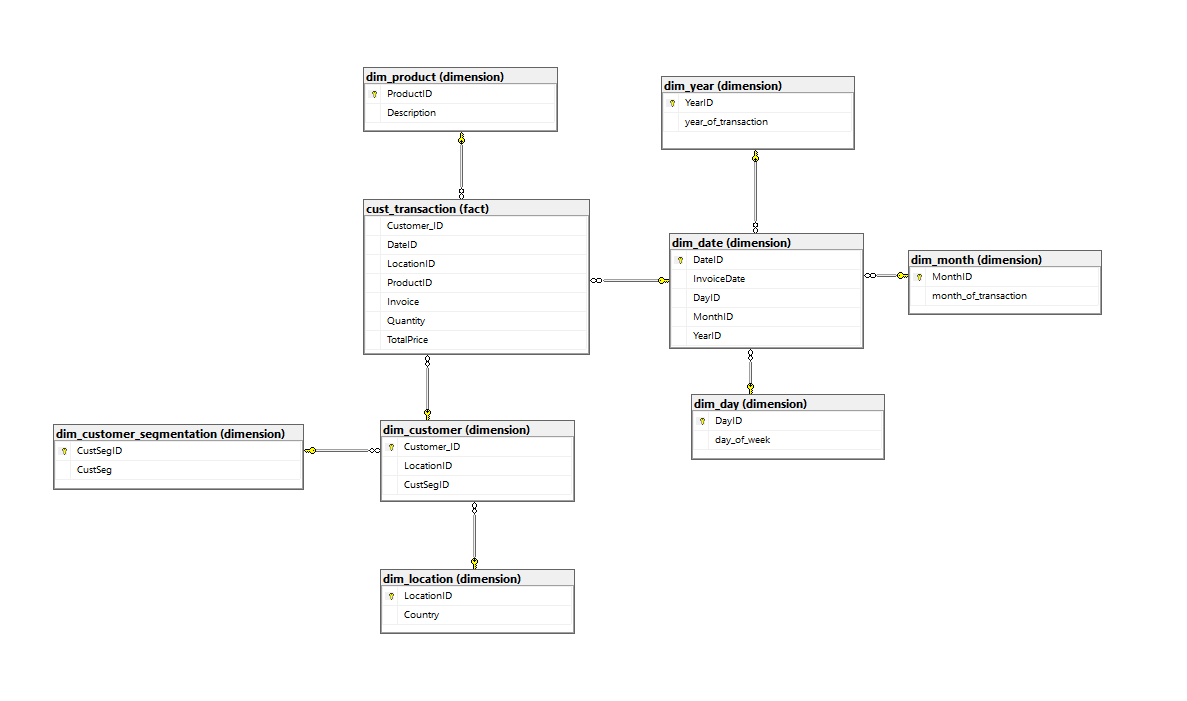
***Hình 14****: Cấu trúc bảng Fact*

**Invoice**: Số hóa đơn của giao dịch

**Quantity**: Số lượng sản phẩm được mua

**TotalPrice**: Giá của số lượng sản phẩm tương ứng

## **3.4 Ràng buộc quan hệ giữa các bảng chiều và bảng Fact**



***Hình 15****: Ràng buộc quan hệ giữa bảng Fact và các bảng Dimension*

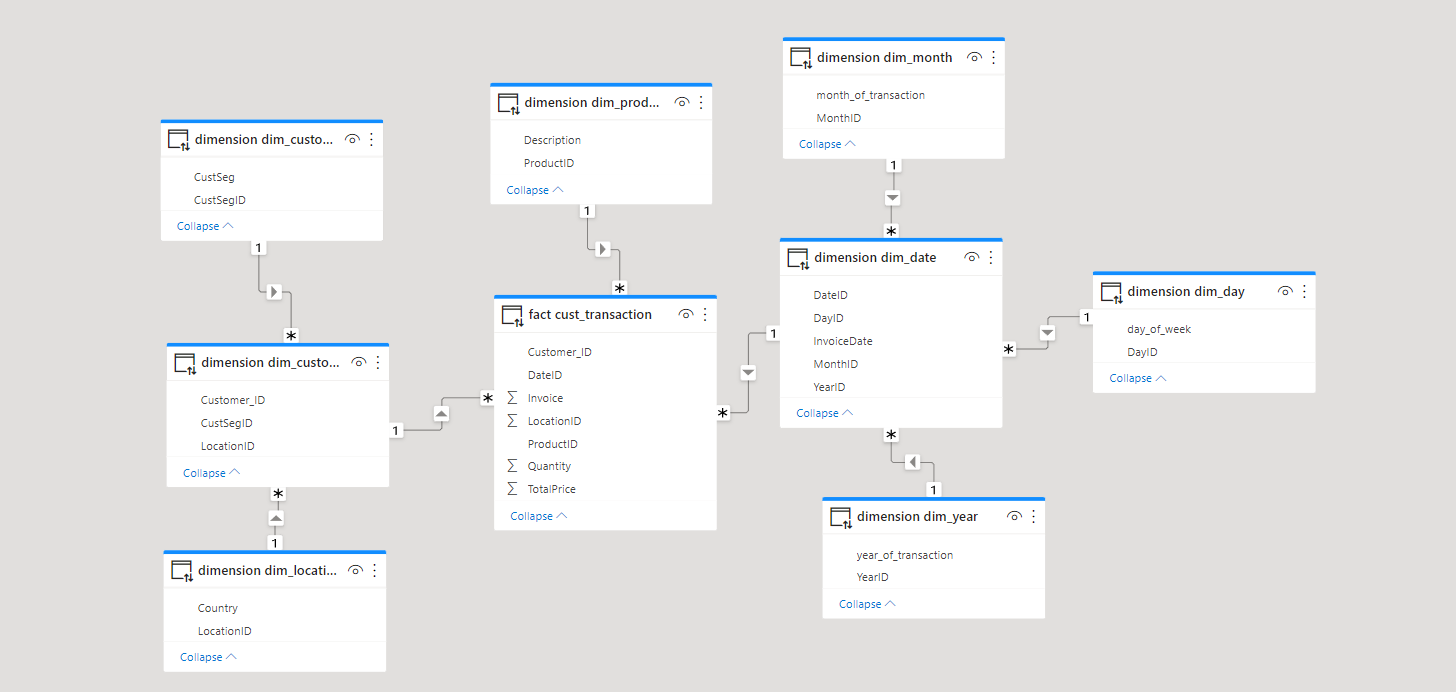
## **3.5 Nhập dữ liệu vào kho dữ liệu**

Sau khi xây dựng kho dữ liệu, tiến hành truy vấn thông tin cần thiết trong database chứa data đã được xử lý trước đó, xuất ra dạng flat file và load vào các bảng, các trường tương ứng trong kho dữ liệu. Để thuận tiện cho việc quản lý, việc định dạng lại một số kiểu dữ liệu tại một số cột là rất cần thiết, ví dụ như chuyển customer\_ID từ nvarchar sang int.

# **4. Trực quan hóa và khai thác bộ dữ liệu**

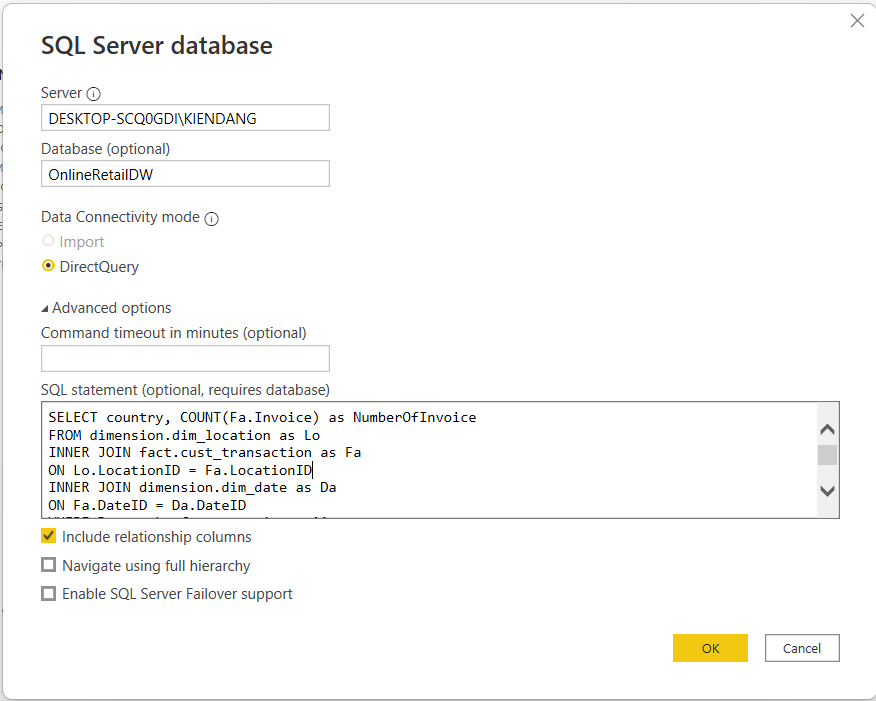
## **4.1 Kết nối data warehouse với PowerBI**

Chọn SQL Server trên thanh công cụ của PowerBI. Tiếp tục điền tên server và tên data warehouse sau đó chọn DirectQuery, việc chọn DirectQuery sẽ thực hiện truy vấn lấy dữ liệu từ data warehouse thay vì lưu vào bộ nhớ trên PowerBI như việc sử dụng import. Khi kết nối thành công data source và PowerBI phần model sẽ hiển các bảng được chọn và mối quan hệ giữa các bảng đó.



***Hình 16****: Mối quan hệ giữa các bảng sau khi kết nối với data source*

Ngoài việc truy vấn mặc định PowerBI cũng cung cấp khả năng truy theo yêu cầu của người dùng để tương tác với data warehouse

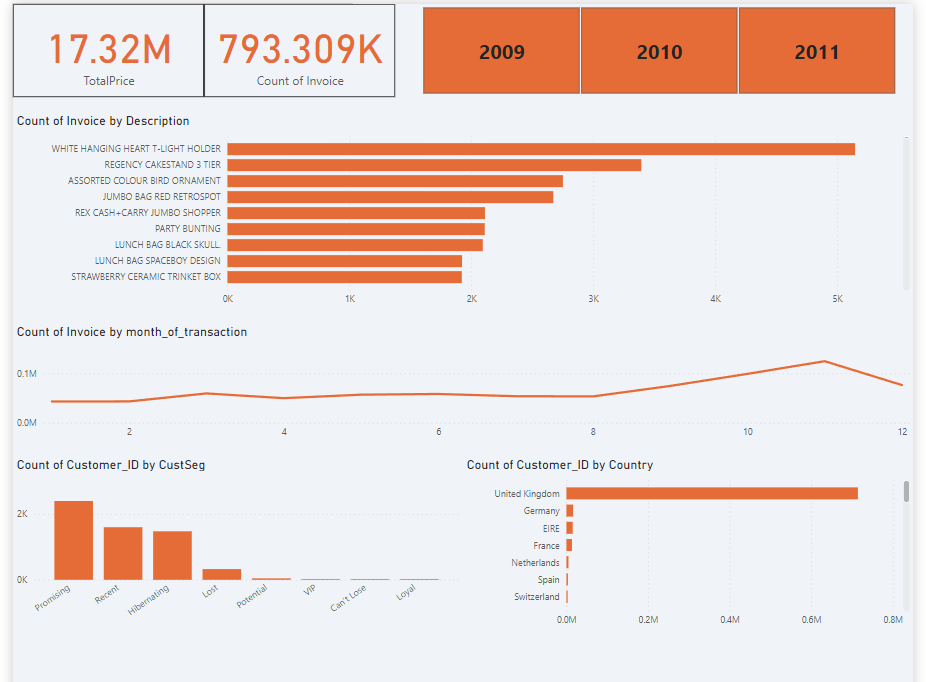


***Hình 17****: Thực hiện truy vấn trên PowerBI*

## **4.2 Phân tích bộ dữ liệu**

### **4.2.1 Phân tích tổng quan**

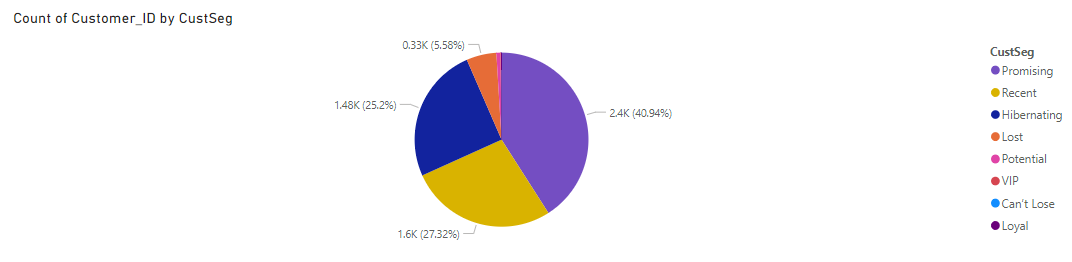
Trước khi phân tích chi tiết, ta sẽ nhìn tổng quan về bộ dữ liệu từ những thông tin bộ dữ liệu cung cấp kết hợp với những thông được nhóm thêm mới như phân khúc khách hàng

*** Hình 18****: Phân tích tổng quan về bộ dữ liệu*

Từ báo cáo trên có thể thấy trong 2 năm từ 12/2009 đến 12/2011 cửa hàng bán được tổng cộng 793,309 đơn hàng với tổng doanh thu là 17,32 triệu bảng anh với sản phẩm bán chạy nhất là “White Hanging heart t-light holder” với 5147 đơn đặt hàng. Nhóm khách mới chiếm tỉ trọng cao và chủ yếu là khách hàng đến từ Vương Quốc Anh. Dựa vào biểu đồ đơn hàng theo tháng nhóm nhận thấy càng về cuối năm xu hướng mua hàng của người dùng càng tăng cao và có 2 đợt xu hướng này tăng lên đột ngột là tháng 3 và bắt đầu tăng mạnh từ tháng 9 đến tháng 11 với tháng 11 là đỉnh điểm mua sắm.

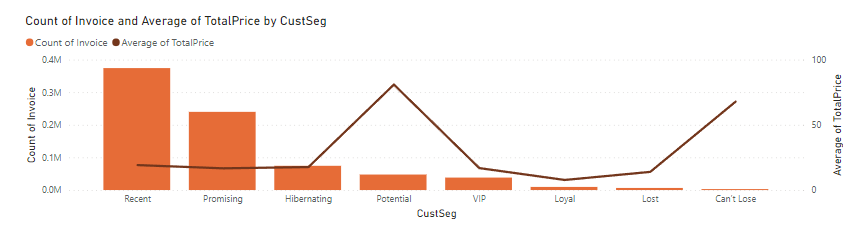
### **4.2.2 Phân tích chi tiết tập trung vào khách hàng**

*Nhóm khách hàng của cửa hàng là những ai? Những khách hàng trung thành, khách mới hay những khách hàng khác?*

******

***Biểu đồ 1****: Phân bố nhóm khách hàng*

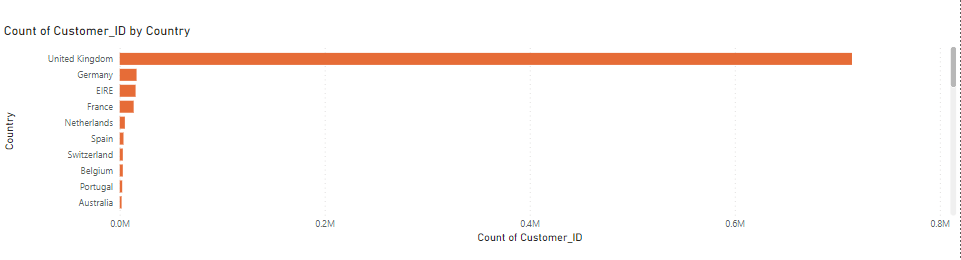
Với biểu đồ trên, 5 nhóm khách hàng chiếm số lượng lớn là Promising, Recent, Hibernating, lost và Potential thì hầu hết là những nhóm khách hàng đã rời bỏ hoặc đã lâu không mua hàng chỉ có Promising, Potential và Recent là những nhóm khách hàng mới tương tác với cửa hàng thời gian gần đây vậy thì có thể kết luận rằng nhóm khách hàng chủ yếu tại cửa hàng là những khách hàng mới.



***Biểu đồ 2****: Số lượt mua và giá trị trung bình đơn hàng của mỗi nhóm khách hàng*

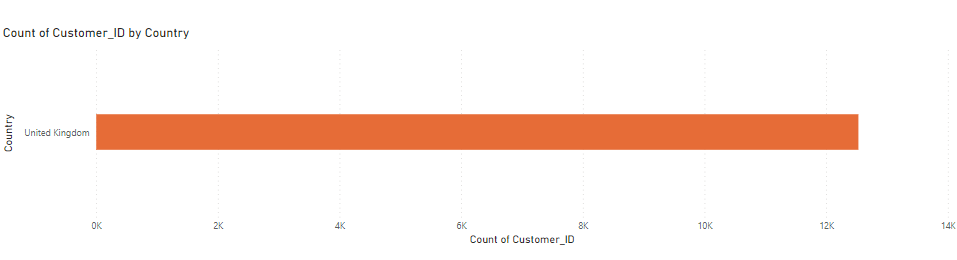
Với biểu đồ 2, những nhóm khách hàng mới như Recent, Promising họ mua rất nhiều tuy nhiên giá trị đơn hàng trung bình của họ tương đối thấp. Trong khi đó, Với các nhóm như Potential và Can’t lose họ mua không quá nhiều đặc biệt là nhóm Can’t lose họ mua ít nhất trong tất cả các nhóm nhưng giá trị đơn hàng trung bình của 2 nhóm này lại là cao hơn rất nhiều so với các nhóm khác.

*Những khách hàng của cửa hàng đến từ những đâu? Những khách hàng mới, khách hàng trung thành họ thường đến từ đâu?*

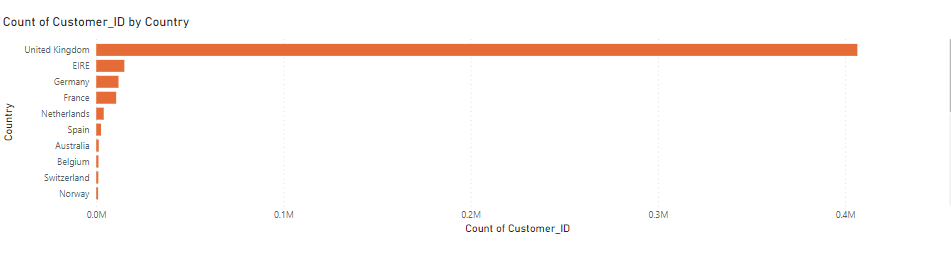
**

***Biểu đồ 3****: Phân bố khách hàng tại các quốc gia*

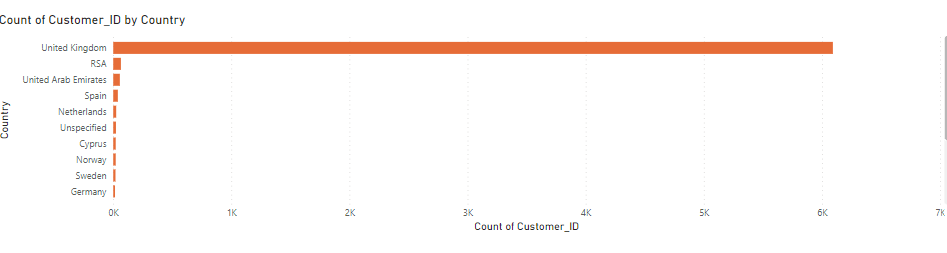
Với biểu đồ 3, nhận thấy khách hàng hầu hết là đến từ Vương Quốc Anh, có khách hàng đến từ nhiều quốc gia khác nhau nhưng số lượng không nhiều. Tuy nhiên với từng nhóm khách hàng cụ thể phân bố này sẽ có sự khác biệt nhất định.



***Biểu đồ 4****: phân bố đất nước của nhóm khách hàng Can’t lose và Loyal*

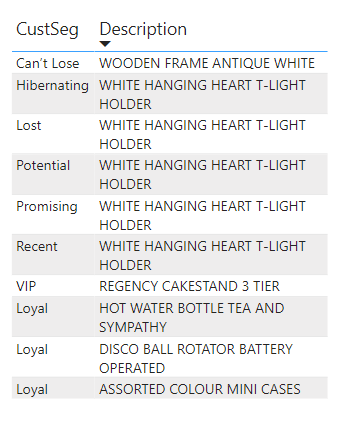
**

***Biểu đồ 5****: Phân bố đất nước của nhóm khách hàng VIP, Potential và Recent*

**

***Biểu đồ 6****: Phân bố đất nước của nhóm khách hàng Lost*

Có thể nhận thấy nhóm khách hàng Loyal và Can’t lose chỉ đến từ nước Anh điều này có thể lý giải dễ hiểu và có căn cứ rằng cửa hàng bán hàng Online cũng ở Vương Quốc Anh nên lượng khách hàng trong nước đóng vai trò như những khách hàng trung thành và thường xuyên mua sản phẩm. Trong đó với khách VIP, Potential và Recent thì ngoài khu vực Vương Quốc Anh thì hầu hết đều đến từ các nước châu âu như EIRE(IreLand), Đức, Pháp, Hà lan, … Có thể thấy gần cửa hàng đã thu hút được không chỉ khách hàng nội địa mà cả các quốc gia khác bên ngoài lãnh thổ Vương Quốc Anh. Qua đến biểu đồ phân bố của nhóm khách hàng Lost thì có thể thấy chính sách của cửa hàng hoặc sản phẩm không phù hợp với các quốc gia ở châu Phi hoặc Ả rập điều này được minh chứng rõ rằng tỉ lệ khách hàng rời bỏ của 2 khu vực này chỉ đứng sau mỗi Vương Quốc Anh.

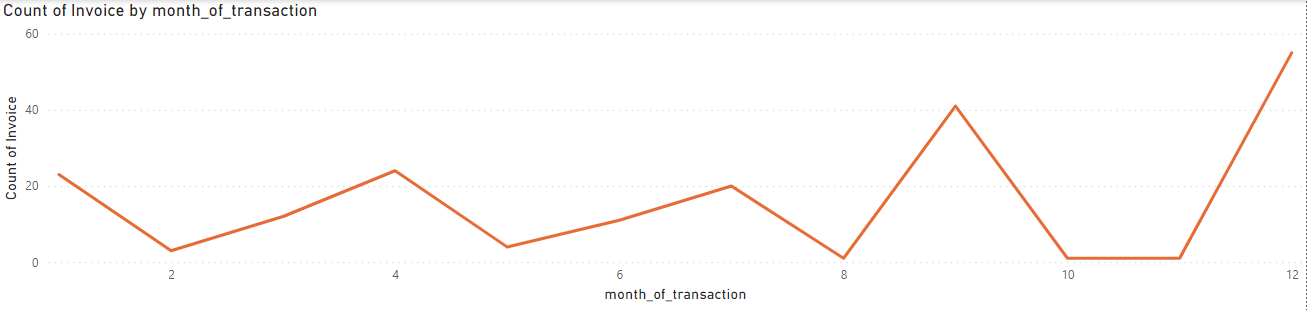
*Những khách hàng đó họ tìm kiếm điều gì ở cửa hàng? Những nhóm khách hàng thường mua những gì từ cửa hàng, liệu có gì khác biệt giữa các nhóm không?*

***Bảng 2****: Sản phẩm được mua nhiều nhất theo từng nhóm khách hàng*

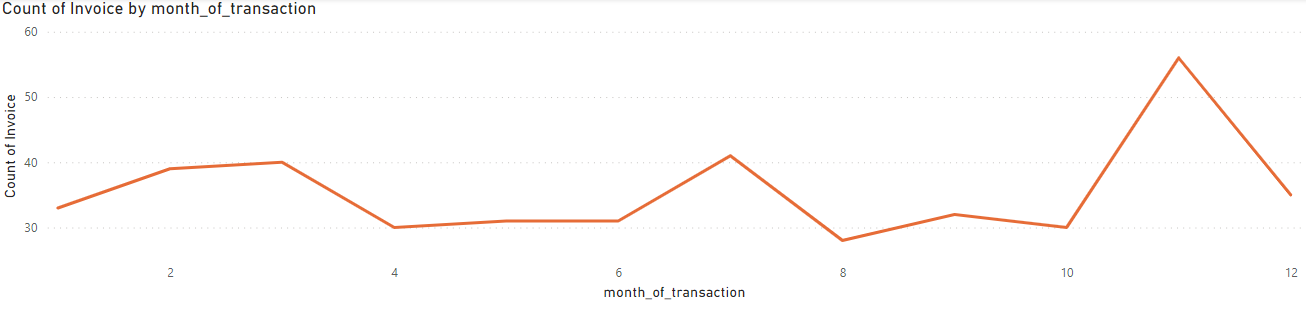
Với dữ liệu truy vấn được ở bảng 2 có thể thấy lý khách hàng thuộc nhóm Can’t lose không mua nhiều nhưng giá trị đơn hàng của họ lại lớn chính là bởi họ mua những sản phẩm có giá trị lớn như các sản phẩm đồ gỗ. Những nhóm khách hàng mới và nhóm khách hàng lâu không mua hàng có xu hướng mua giống nhau là những sản phẩm tiêu dùng giá rẻ, có thể đây là dòng sản phẩm thu hút được khách hàng của cửa hàng nhưng vì một lý do nào đó mà những sản phẩm này lại không thể giữ chân khách hàng và mang khách hàng trở lại cửa hàng. Trong khi đó nhóm khách hàng VIP, Loyal lại ưa chuộng những dụng cụ tiện ích (hộp đựng đồ, đế làm bánh) hay đồ điện tử. Đặc biệt nhóm khách hàng loyal có sở thích khá đồng đều khi họ mua 3 loại sản phẩm khác nhau với số lượng bằng nhau. Điều này có thể thấy rằng những mặt hàng như dụng cụ tiện ích hay đồ điện tử không có khả năng thu hút khách hàng lớn như những mặt hàng tiêu dùng tuy nhiên chúng lại là những mặt hàng có thể giữ chân khách hàng và đem khách hàng trở lại với cửa hàng.

*Thói quen mua sắm của họ như thế nào? Họ có xu hướng mua vào những ngày, giờ hay tháng đặc biệt nào không?*

Nhìn chung thói quen mua sắm của khách hàng của khách hàng thường có xu hướng tăng lên vào tháng 3 và tháng 11 tuy nhiên với mỗi nhóm khách hàng xu hướng này sẽ có chút khách biệt.



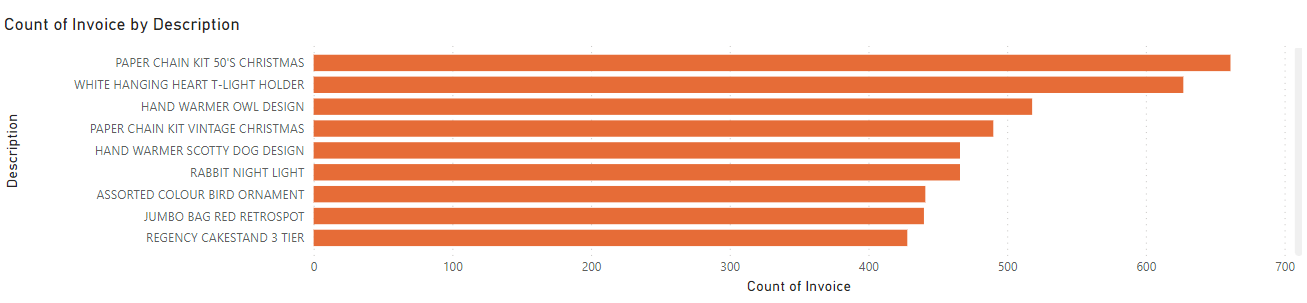
***Biểu đồ 7****: Thói quen mua sắm theo tháng của nhóm khách hàng Can’t lose*



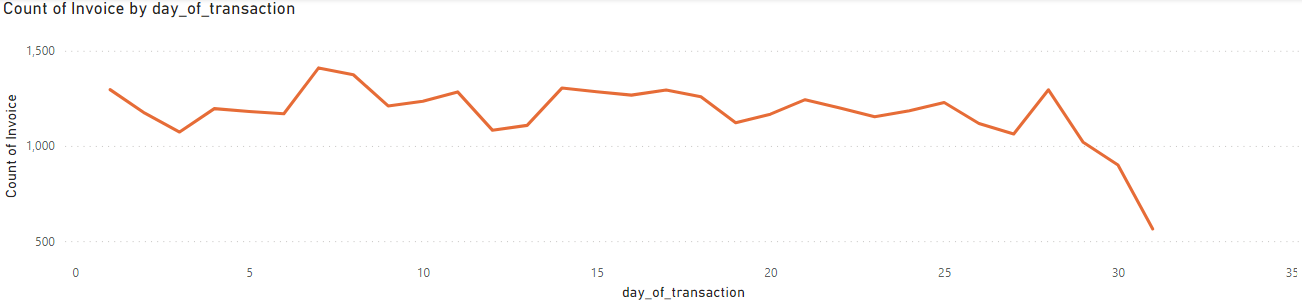
***Biểu đồ 8****: Thói quen mua sắm theo tháng của nhóm khách hàng Loyal*

Với hai nhóm khách hàng Can’t lose và Loyal tuy đều ở cùng trong một khu vực địa lý cụ thể là tất cả đều đến từ Vương Quốc Anh nhưng thói quen mua sắm của hai nhóm này có sự khách biệt rõ rệt. Nhóm khách hàng Can’t lose họ không mua sắm nhiều và thường xuyên dẫn đến biểu đồ thói quen mua sắm của họ tạo thành nhiều đỉnh và vùng trũng, nhóm khách hàng này thường có thói quen mua sắm vào đầu và cuối mỗi quý trừ quý 4 và giảm lượng chi tiêu và giữa các quý. Đặc biệt là dịp cuối năm so với những nhóm khác thường có xu hướng mua nhiều nhất và tháng 11 thì nhóm Can’t lose lại tập trung nhiều vào tháng 12 và dường như rất ít chi tiền trong tháng 10 và 11. Còn đối với nhóm khách hàng Loyal, họ mua hàng thường xuyên hơn nên biểu đồ của họ sẽ không xuất hiện nhiều đỉnh và vùng trũng. Nhóm này thường có xu hướng mua nhiều hơn vào đầu và cuối năm, đặc biệt là tháng 11.

Đối với các nhóm khác như Recent, VIP hay potential, xu hướng mua của họ không có gì khác biệt so với xu hướng chung đó là mua ít vào đầu năm và càng về cuối năm xu hướng mua sắm càng tăng, đạt đỉnh điểm mua sắp vào tháng 11. Việc các nhóm khách hàng có xu hướng mua nhiều hơn vào tháng 11 có thể bởi vì đây là lúc chuẩn bị bước vào các dịp lễ lớn bên các nước phương tây như giáng sinh hay năm mới, điều này được thể hiện rõ ở dữ liệu sản phẩm mà khách hàng thường mua trong tháng này.

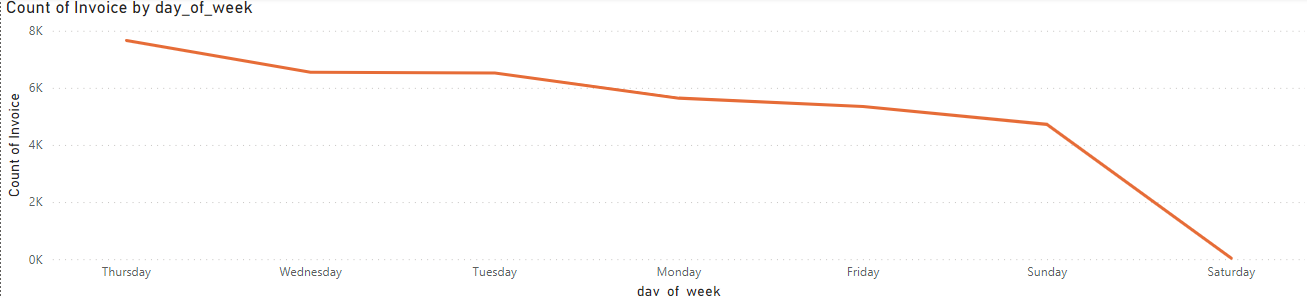


***Biểu đồ 9****: Sản phẩm được khách hàng mua nhiều vào tháng 11*



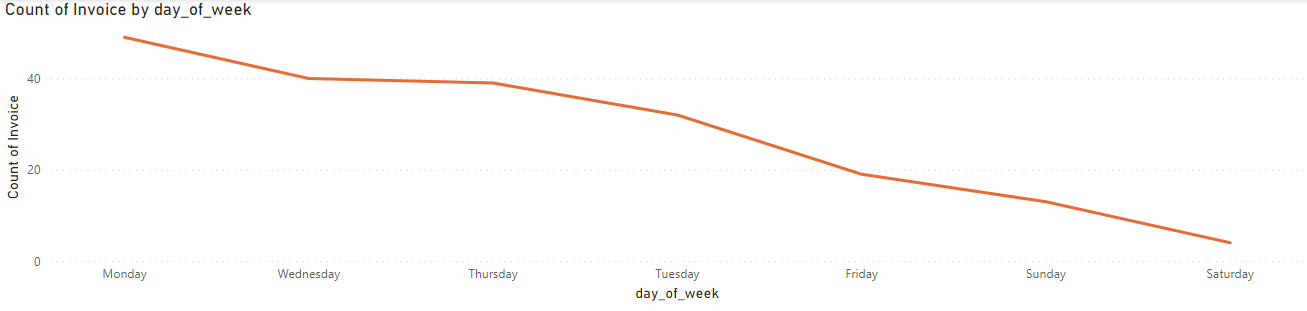
***Biểu đồ 10****: Xu hướng mua theo ngày của các nhóm khách hàng*

Đối với xu hướng mua theo ngày, các nhóm không có nhiều khác biệt, hầu hết đều tuân theo xu hướng chung là sẽ mua nhiều hơn vào những ngày đầu tháng và giảm mạnh vào những ngày cuối tháng.



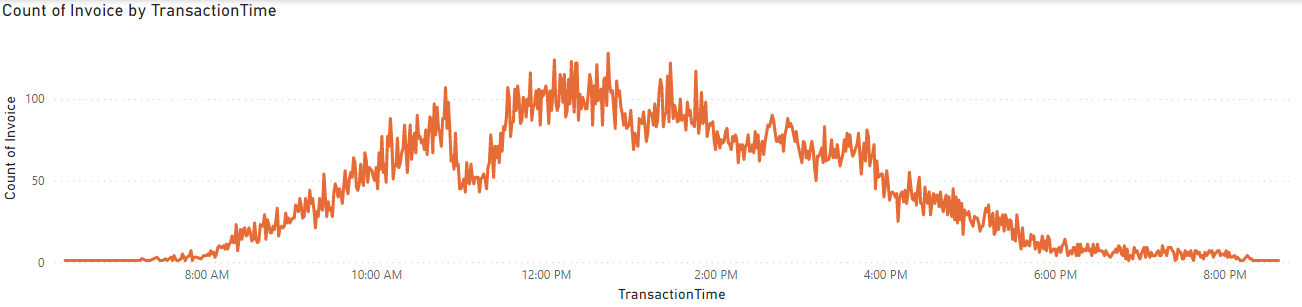
***Biểu đồ 11****: xu hướng mua hàng theo ngày trong tuần của các nhóm khách hàng*

Về xu hướng mua theo ngày trong tuần, các nhóm đều có xu hướng chi tiêu nhiều vào các ngày trong tuần và giảm vào các ngày đầu, cuối tuần. Giữa các nhóm có một điểm chung đó là đặc biệt không chi tiêu vào thứ bảy. Tuy nhiên quy luật này đúng với hầu hết các nhóm, ngoại trừ nhóm Can’t lose.



***Biểu đồ 12****: Xu hướng mua hàng theo ngày của nhóm Can’t lose*

Nhóm khách hàng Can’t lose cũng không thường xuyên mua hàng vào cuối tuần đặc biệt là thứ bảy, tuy nhiên khác với những nhóm khác, nhóm Can’t lose có xu hướng mua sắm nhiều nhất vào ngày đầu tuần.



***Biểu đồ 13****: Xu hướng mua hàng theo giờ trong ngày*

Nhìn vào biểu đồ 13 có thể thấy các nhóm khách hàng có xu hướng mua sắm nhiều trong giờ hành chính, đặc biệt là khoảng từ giữa trưa đến giữa buổi chiều xu hướng mua sắm của khách hàng tăng cao hơn so với các giờ khác. Các khách hàng cũng không thường xuyên mua sắm vào buổi tối bắt đầu từ sau khi kết thúc giờ hành chính. Quy luật này không có sự khác biệt giữa các nhóm.

# **5. Kết Luận**

Từ bộ dữ liệu Online Retail II UCI ta đã tiến hành xây dựng kho dữ liệu dựa trên giao dịch của khách hàng, sau đó tiến hành phân tích bộ dữ liệu để tìm ra quy luật, điểm bất thường để nhằm mục đích cải thiện quan hệ khách hàng và tăng doanh thu cho cửa hàng. Sau khi phân tích, có thể đưa đến kết luận rằng có sự khác biệt về địa lý, thói quen mua sắm giữa các nhóm khách hàng khác nhau. Tuy nhiên, để thu được nhiều kết luận và giúp phân tích chính xác hơn thì bộ dữ liệu sẽ cần được mở rộng và cung cấp nhiều thông tin hơn về khách hàng cũng như cửa hàng.

# **Bộ dữ liệu sử dụng trong bài báo cáo**

<[Online Retail II UCI | Kaggle](https://www.kaggle.com/datasets/mashlyn/online-retail-ii-uci)>

# **Tài liệu tham khảo**

Ralph Kimball, Margy Ross. The Data Warehouse Toolkit.

<[RFM Analysis For Successful Customer Segmentation - Putler](https://www.putler.com/rfm-analysis/#Visualizing_RFM_data)>.

1. [Online Retail II UCI | Kaggle](https://www.kaggle.com/datasets/mashlyn/online-retail-ii-uci) [↑](#footnote-ref-1)