

PHÂN TÍCH KINH DOANH

BI54 - Nhóm 3

Thành viên nhóm

1. Bùi Minh Luân
2. Cao Hồng Nhi
3. Trương Duy Phong
4. Nguyễn Văn Tuấn



ONLINE RETAIL



01
GIỚI THIỆU



02
XỬ LÝ DỮ LIỆU



03
PHÂN TÍCH



04
ĐỀ XUẤT

01 - GIỚI THIỆU

Giới thiệu Dataset - Online Retail

Đây là bộ dữ liệu chứa tất cả các giao dịch xảy ra từ ngày 12/01/2010 đến ngày 12/09/2011 của một cửa hàng bán lẻ trực tuyến và có trụ sở tại Vương quốc Anh. Công ty chủ yếu bán quà tặng độc đáo cho mọi dịp. Nhiều khách hàng của công ty là người bán sỉ.

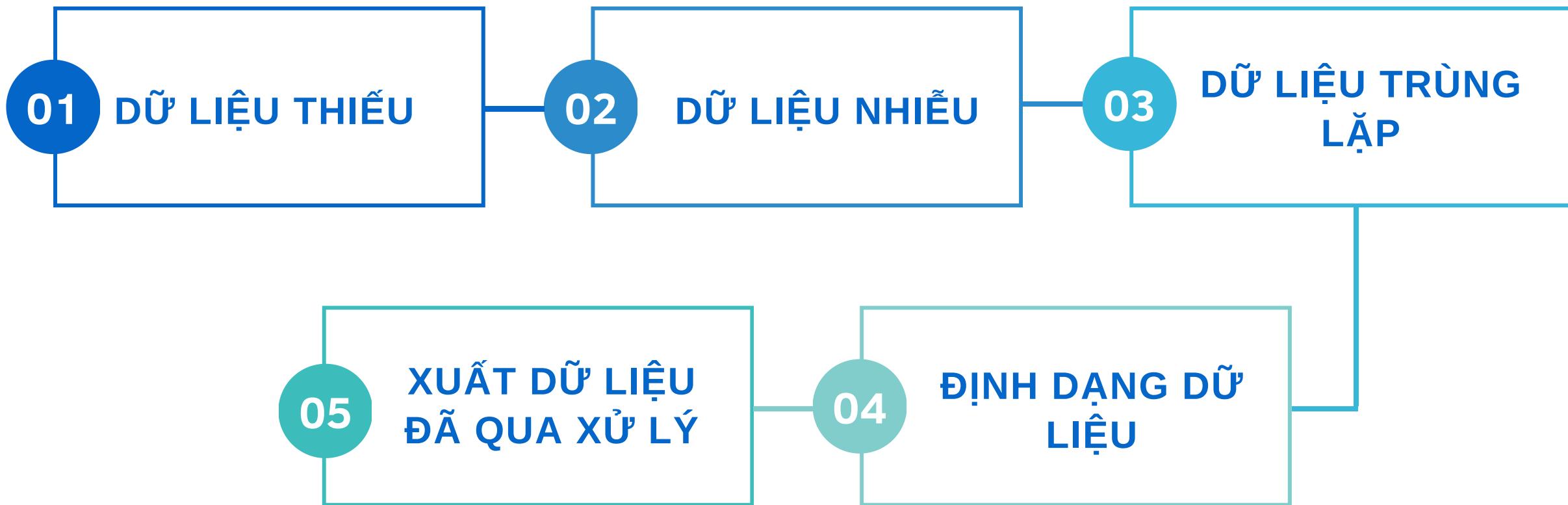
InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
536557	22731	3D CHRISTMAS STAMPS STICKERS	1	12/1/2010 2:41:00 PM	1.25	17841	United Kingdom
536557	22953	BIRTHDAY PARTY CORDON BARRIER TAPE	1	12/1/2010 2:41:00 PM	1.25	17841	United Kingdom
536557	21672	WHITE SPOT RED CERAMIC DRAWER KNOB	1	12/1/2010 2:41:00 PM	1.25	17841	United Kingdom
536557	22686	FRENCH BLUE METAL DOOR SIGN No	1	12/1/2010 2:41:00 PM	1.25	17841	United Kingdom
536557	20963	APPLE BATH SPONGE	1	12/1/2010 2:41:00 PM	1.25	17841	United Kingdom
536984	21446	12 RED ROSE PEG PLACE SETTINGS	1	12/3/2010 2:35:00 PM	1.25	17841	United Kingdom
536984	20963	APPLE BATH SPONGE	1	12/3/2010 2:35:00 PM	1.25	17841	United Kingdom
536984	20961	STRAWBERRY BATH SPONGE	1	12/3/2010 2:35:00 PM	1.25	17841	United Kingdom
537405	22673	FRENCH GARDEN SIGN BLUE METAL	1	12/6/2010 2:51:00 PM	1.25	17841	United Kingdom
537405	20972	PINK CREAM FELT CRAFT TRINKET BOX	1	12/6/2010 2:51:00 PM	1.25	17841	United Kingdom
538163	22679	FRENCH BLUE METAL DOOR SIGN 4	1	12/9/2010 5:27:00 PM	1.25	17841	United Kingdom
538163	20971	PINK BLUE FELT CRAFT TRINKET BOX	1	12/9/2010 5:27:00 PM	1.25	17841	United Kingdom
538163	21867	PINK UNION JACK LUGGAGE TAG	1	12/9/2010 5:27:00 PM	1.25	17841	United Kingdom

01 - GIỚI THIỆU

Giới thiệu Dataset

Giới thiệu Dataset	
InvoiceNo	Số hóa đơn. Danh nghĩa, một số nguyên gồm 6 chữ số được gán duy nhất cho mỗi giao dịch.
StockCode	Mã sản phẩm (mặt hàng). Danh nghĩa, một số nguyên gồm 5 chữ số được gán duy nhất cho từng sản phẩm riêng biệt.
Description	Tên sản phẩm
Quantity	Số lượng của từng sản phẩm trên mỗi giao dịch.
InvoiceDate	Ngày và giờ lập hóa đơn. Số, ngày và giờ mỗi giao dịch được tạo.
UnitPrice	Đơn giá. Số, Giá sản phẩm trên mỗi đơn vị bằng đồng bảng Anh.
CustomerID	Mã số khách hàng. Danh nghĩa, một số nguyên gồm 5 chữ số được gán duy nhất cho mỗi khách hàng.
Country	Tên quốc gia. Danh nghĩa, tên quốc gia nơi mỗi khách hàng cư trú.

02 - XỬ LÝ DỮ LIỆU



02 - XỬ LÝ DỮ LIỆU

Kiểm tra và xử lý dữ liệu thiếu

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 541909 entries, 0 to 541908
Data columns (total 8 columns):
 #   Column      Non-Null Count  Dtype  
--- 
 0   InvoiceNo    541909 non-null   object  
 1   StockCode     541909 non-null   object  
 2   Description   540455 non-null   object  
 3   Quantity      541909 non-null   int64   
 4   InvoiceDate   541909 non-null   datetime64[ns]
 5   UnitPrice     541909 non-null   float64 
 6   CustomerID    406829 non-null   float64 
 7   Country       541909 non-null   object  
dtypes: datetime64[ns](1), float64(2), int64(1), object(4)
memory usage: 33.1+ MB
```



```
df.isnull().sum()
```

```
InvoiceNo          0
StockCode          0
Description        1454
Quantity           0
InvoiceDate        0
UnitPrice          0
CustomerID        135080
Country            0
dtype: int64
```



```
df = df.dropna(subset= ['CustomerID'])
df
```

- Dữ liệu bao gồm 541,909 dòng
- Cột Description thiếu 1454 dòng
- Cột CustomerID thiếu 135,080 dòng



- Xóa các dòng CustomerID chứa giá trị null

02 - XỬ LÝ DỮ LIỆU

Xử lý dữ liệu nhiễu

```
[ ] result = df[ (df['Quantity'] >= 0) & (df['UnitPrice'] >= 0) ]  
result
```

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	2010-12-01 08:26:00	2.55	17850.0	United Kingdom
1	536365	71053	WHITE METAL LANTERN	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	2010-12-01 08:26:00	2.75	17850.0	United Kingdom
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom
...
541904	581587	22613	PACK OF 20 SPACEBOY NAPKINS	12	2011-12-09 12:50:00	0.85	12680.0	France
541905	581587	22899	CHILDREN'S APRON DOLLY GIRL	6	2011-12-09 12:50:00	2.10	12680.0	France
541906	581587	23254	CHILDRENS CUTLERY DOLLY GIRL	4	2011-12-09 12:50:00	4.15	12680.0	France
541907	581587	23255	CHILDRENS CUTLERY CIRCUS PARADE	4	2011-12-09 12:50:00	4.15	12680.0	France
541908	581587	22138	BAKING SET 9 PIECE RETROSPOT	3	2011-12-09 12:50:00	4.95	12680.0	France

397924 rows × 8 columns

- Loại bỏ các dòng dữ liệu có Quantity(số lượng sản phẩm) âm và UnitPrice (Giá) âm

02 - XỬ LÝ DỮ LIỆU

Xử lý dữ liệu dữ liệu trùng lặp

```
[ ] df = df.drop_duplicates()  
df
```

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	2010-12-01 08:26:00	2.55	17850.0	United Kingdom
1	536365	71053	WHITE METAL LANTERN	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	2010-12-01 08:26:00	2.75	17850.0	United Kingdom
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom
...

- Loại bỏ các dữ liệu trùng lặp

02 - XỬ LÝ DỮ LIỆU

Định dạng dữ liệu

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 541909 entries, 0 to 541908
Data columns (total 8 columns):
 #   Column      Non-Null Count   Dtype  
--- 
 0   InvoiceNo   541909 non-null    object  
 1   StockCode    541909 non-null    object  
 2   Description  540455 non-null    object  
 3   Quantity     541909 non-null    int64  
 4   InvoiceDate  541909 non-null    datetime64[ns]
 5   UnitPrice    541909 non-null    float64 
 6   CustomerID   406829 non-null    float64 
 7   Country      541909 non-null    object  
dtypes: datetime64[ns](1), float64(2), int64(1), object(4)
memory usage: 33.1+ MB
```

- Kiểu dữ liệu đã hợp lý

Xuất dữ liệu

```
[ ] result.to_csv('Online_retail_final.csv', index= False)
```

- Xuất dữ bộ dữ liệu mới bao gồm 397,924 dòng và 8 cột

03 - PHÂN TÍCH

OVERVIEW

DATETIME

All

TOTAL REVENUE

8.91M

AOV

480.76

TOTAL ORDERS

18.536K

TOTAL CUSTOMERS

4.339K

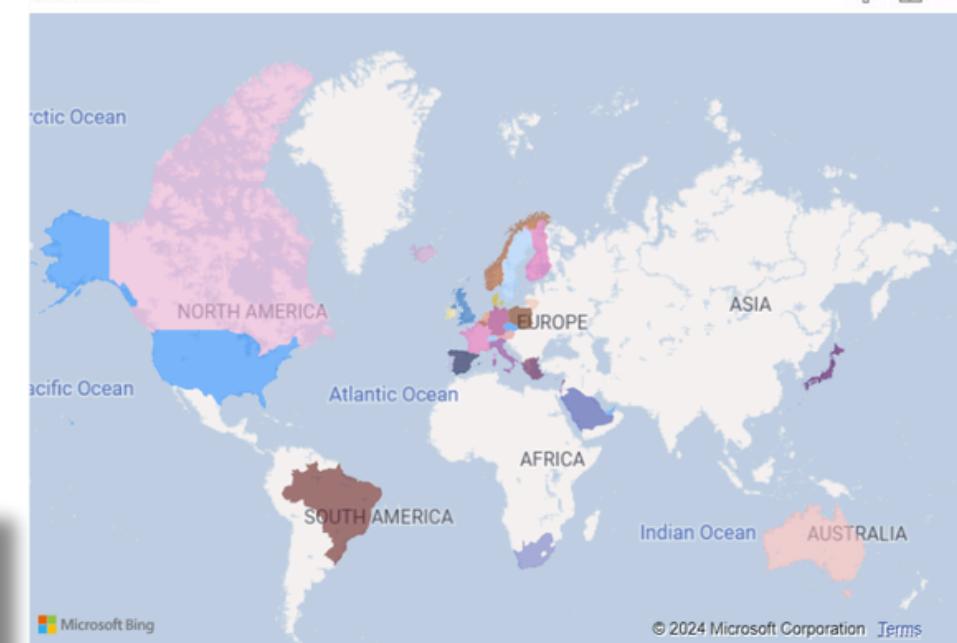
TOTAL PRODUCTS

3.868K

TOTAL REVENUE OVER TIME



COUNTRY

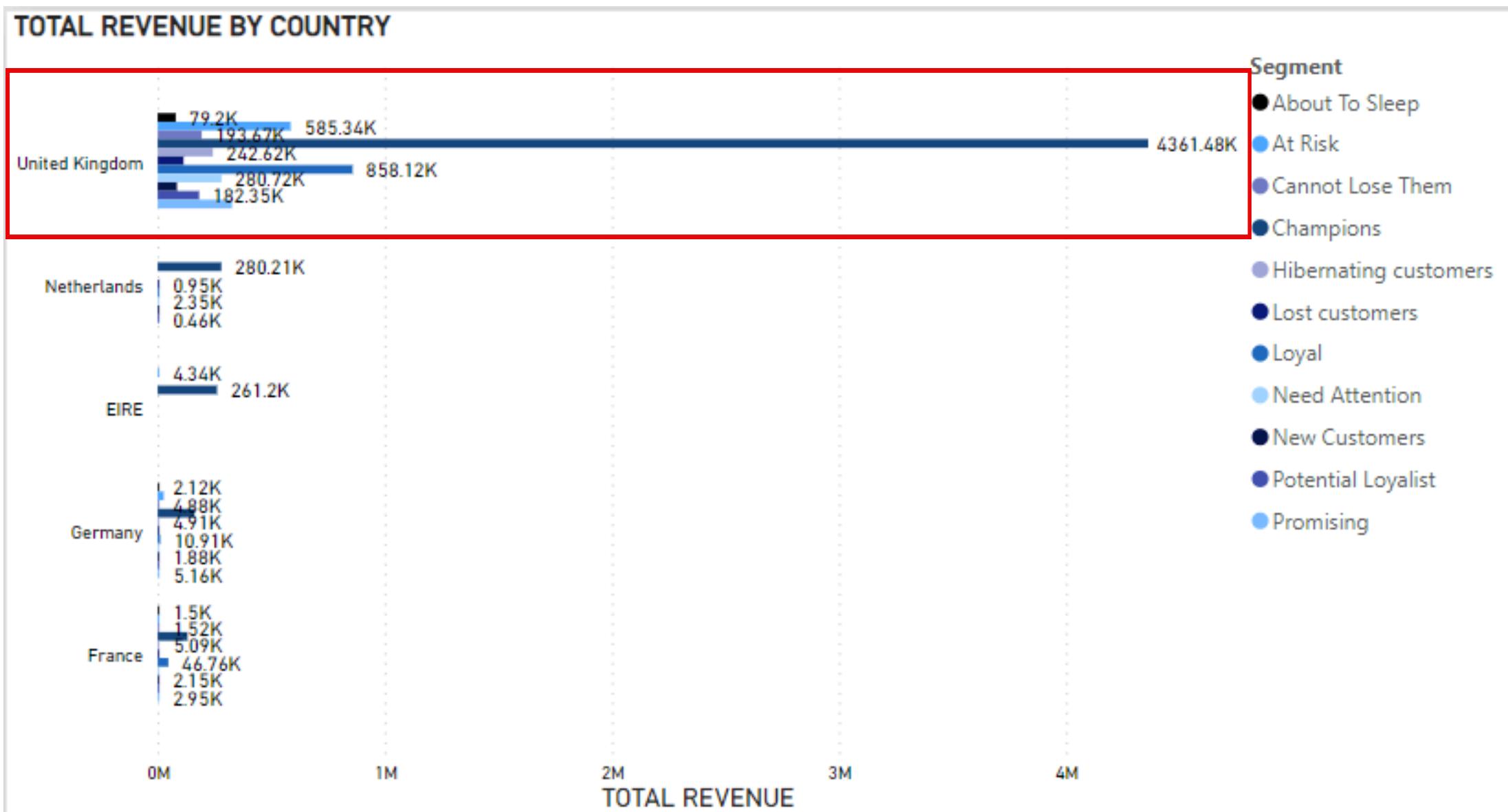


Product

Product	Total Quantity	Total Revenue
PAPER CRAFT , LITTLE BIRDIE	80995	168,469.60
REGENCY CAKESTAND 3 TIER	12412	142,592.95
WHITE HANGING HEART T-LIGHT HOLDER	36725	100,448.15
JUMBO BAG RED RETROSPOT	46181	85,220.78
MEDIUM CERAMIC TOP STORAGE JAR	77916	81,416.73
POSTAGE	3120	77,803.96
PARTY BUNTING	15295	68,844.33
ASSORTED COLOUR BIRD ORNAMENT	35362	56,580.34
Manual	7179	53,779.93
RABBIT NIGHT LIGHT	27202	51,346.20
CHILLI LIGHTS	9650	46,286.51
PAPER CHAIN KIT FOR CHRISTMAS	15117	42,660.00
Total	5181696	8,911,407.90

03 - PHÂN TÍCH

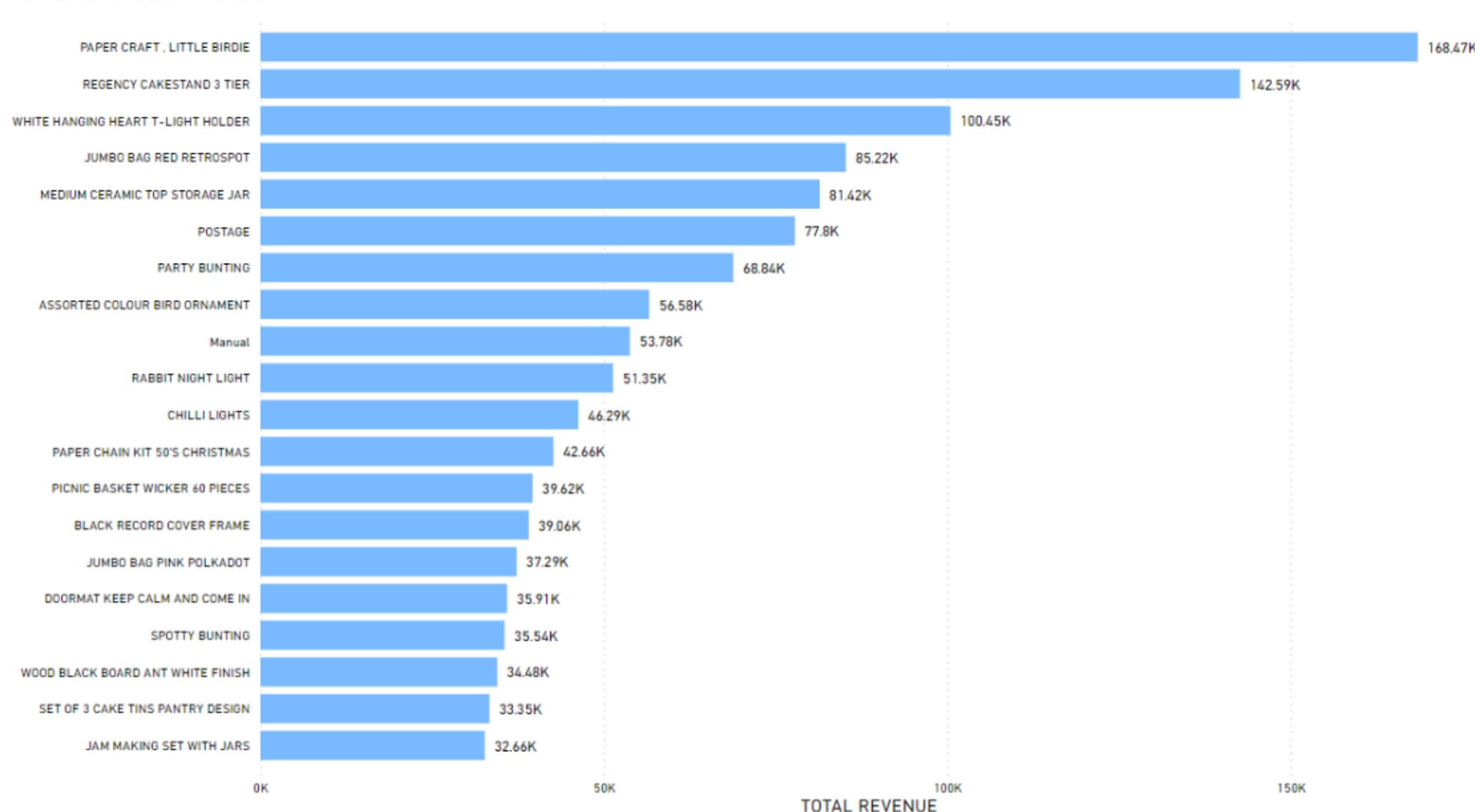
COUNTRY



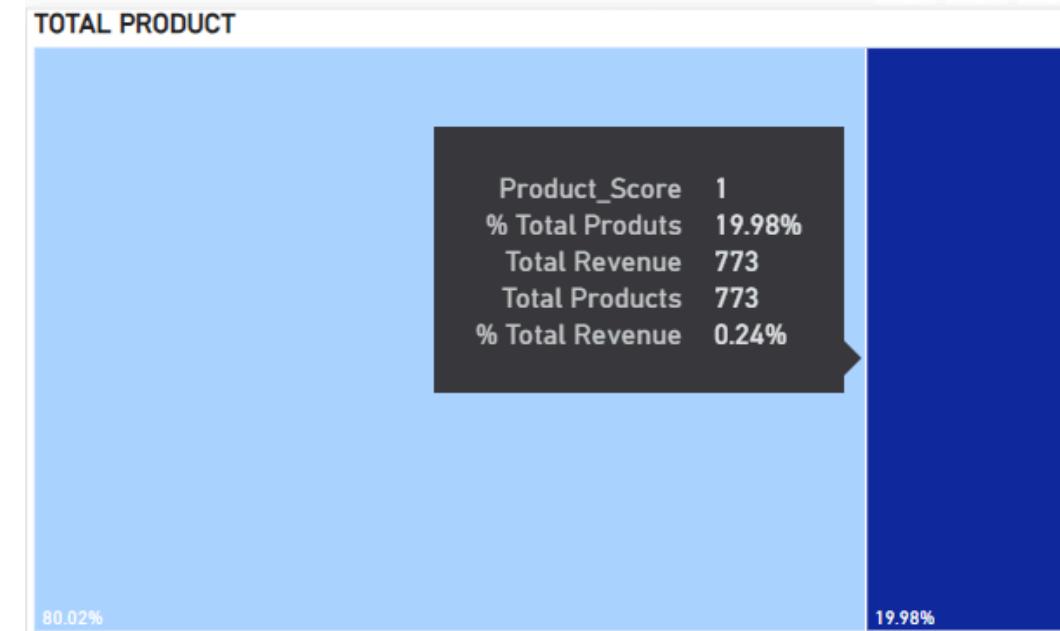
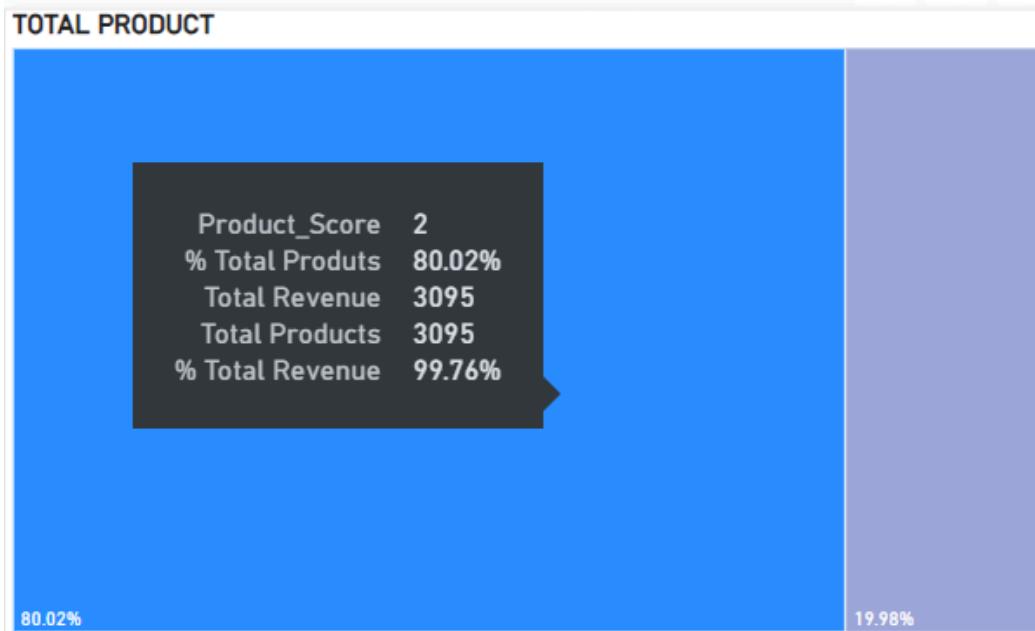
03 - PHÂN TÍCH

PRODUCT - TOP 20

TOTAL REVENUE BY PRODUCT



03 - PHÂN TÍCH PRODUCT



- 80% sản phẩm, cụ thể là 3095 sản phẩm của cửa hàng chiếm 99,76% doanh thu
- 20% sản phẩm, cụ thể là 773 sản phẩm của cửa hàng chiếm 0.24% doanh thu
- Ngược lại, 10% sản phẩm có doanh thu cao nhất chiếm hơn 93% tổng doanh thu.
- Loại bỏ 20% sản phẩm có doanh thu thấp, để tập trung phát triển các sản phẩm còn lại. Ngược lại, đẩy mạnh chương trình khuyến mãi phát triển cho 10% sản phẩm đang có doanh thu cao nhất.

03 - PHÂN TÍCH

RFM

TOTAL REVENUE

8.91M

TOTAL ORDERS

18.536K

TOTAL CUSTOMERS

4.339K

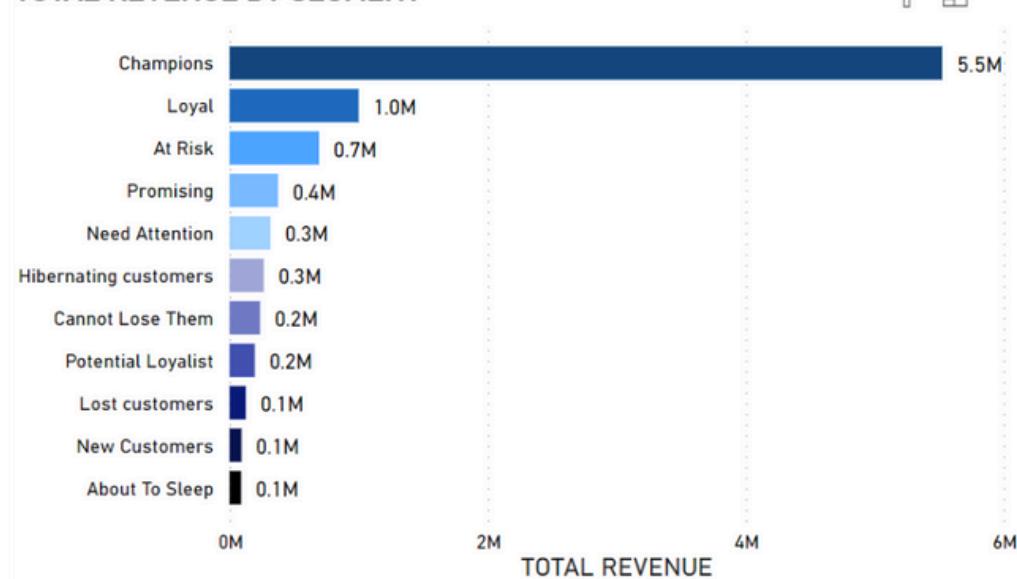
TOTAL COUNTRIES

37

TOTAL REVENUE BY SEGMENT



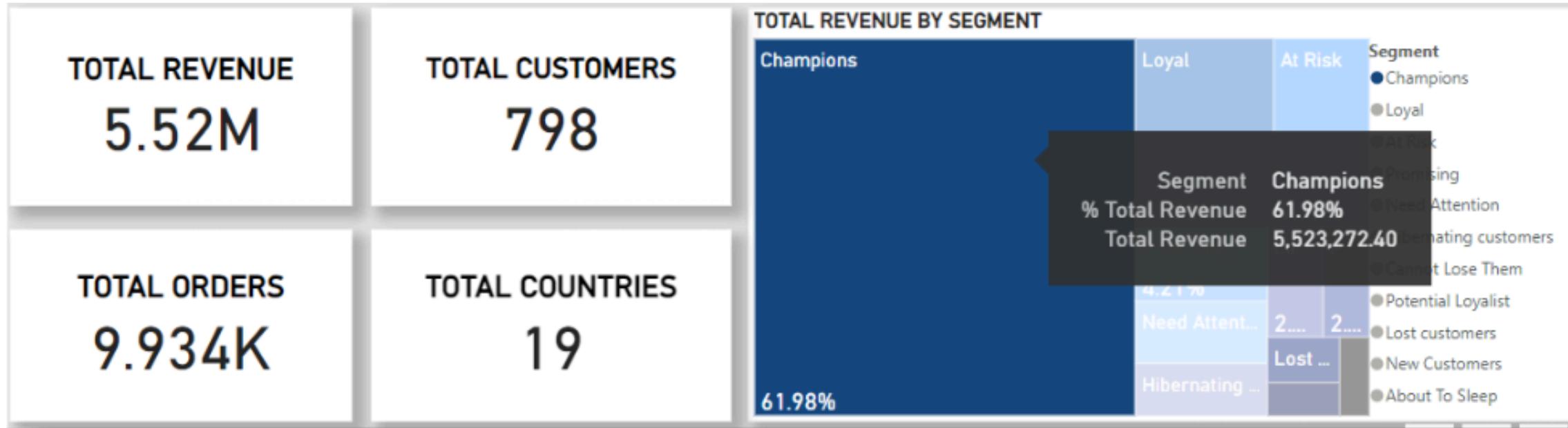
TOTAL REVENUE BY SEGMENT



Customer ID	RFM Score	Last day order	Segment	Quantity	TotalSales
14646	555	December 8, 2011	Champions	74	280,206.02
18102	555	December 9, 2011	Champions	60	259,657.30
17450	555	December 1, 2011	Champions	46	194,550.79
16446	525	December 9, 2011	Promising	2	168,472.50
14911	555	December 8, 2011	Champions	201	143,825.06
12415	455	November 15, 2011	Champions	21	124,914.53
14156	555	November 30, 2011	Champions	55	117,379.63
17511	555	December 7, 2011	Champions	31	91,062.38
16029	355	November 1, 2011	Loyal	63	81,024.84
12346	115	January 18, 2011	Cannot Lose Them	1	77,183.60
16684	555	December 5, 2011	Champions	28	66,653.56
14096	555	December 5, 2011	Champions	17	65,164.79
Total				18536	8,911,407.90

03 - PHÂN TÍCH

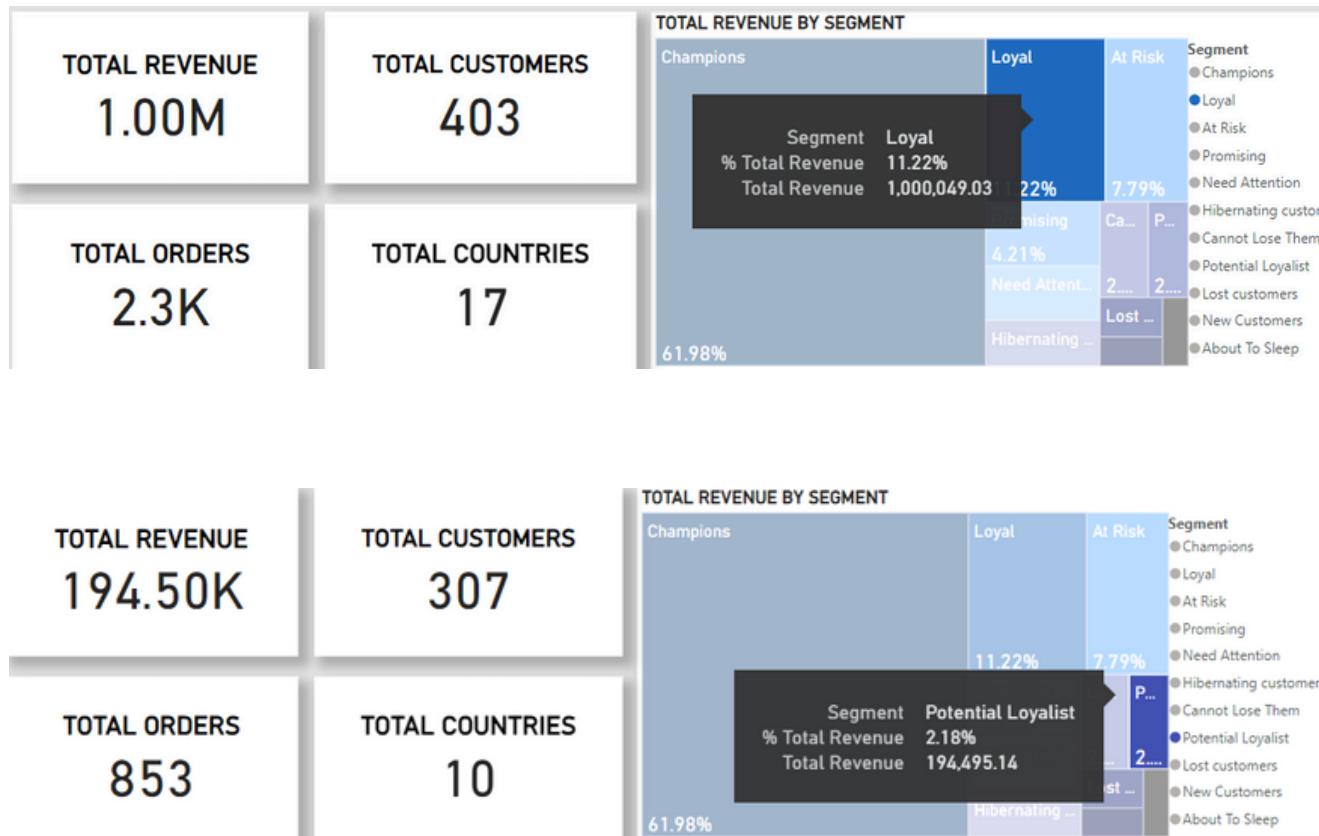
Phân khúc khách hàng Champions



- Nhóm khách hàng thường xuyên mua hàng, với giá trị giỏ hàng lớn và vẫn có giao dịch gần đây chiếm gần 62% doanh thu, với hơn 5,5 triệu bảng Anh.
- Nên tạo chương trình khuyến mãi và ưu đãi đặc biệt cho nhóm Champions để tăng trung thành và doanh thu. Tối ưu hóa trải nghiệm mua hàng và tiếp thị để tăng tương tác và hiệu quả với nhóm Champions.

03 - PHÂN TÍCH

Phân khúc khách hàng Loyal và Potential Loyalist



- Nhóm khách hàng thường xuyên mua hàng, với giá trị giỏ hàng nhỏ chiếm khoảng 14%, với khoảng 1,2 triệu bảng Anh.
- Nên tạo chương trình ưu đãi và tiếp thị tập trung để kích thích trung thành và mua sắm thường xuyên của khách hàng Loyal và Potential Loyalist.

03 - PHÂN TÍCH

Phân khúc khách hàng New và Promising

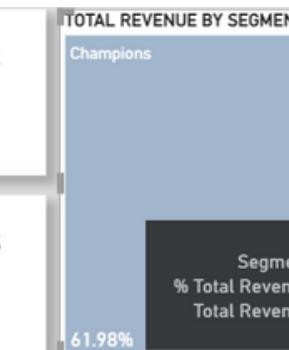
- Nhóm khách hàng chưa mua hàng thường xuyên và vẫn đang trong giai đoạn đánh giá, trải nghiệm với giá trị giỏ hàng nhỏ chiếm khoảng 5%, với khoảng gần 500 ngàn bảng Anh.
- Nên tập trung vào việc tạo trải nghiệm mua hàng, các chương trình quảng cáo, truyền thông, khuyến mãi đặc biệt cho khách hàng mới nhằm thu hút để kích thích mua sắm và tăng cường tích cực của khách hàng New và Promising

TOTAL REVENUE
91.93K

TOTAL ORDERS
484

TOTAL CUSTOMERS
408

TOTAL COUNTRIES
13

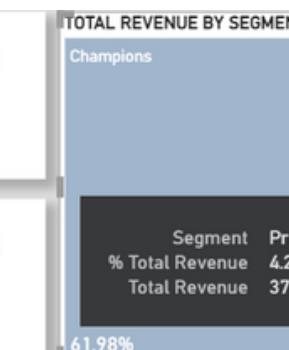


TOTAL REVENUE
375.32K

TOTAL ORDERS
389

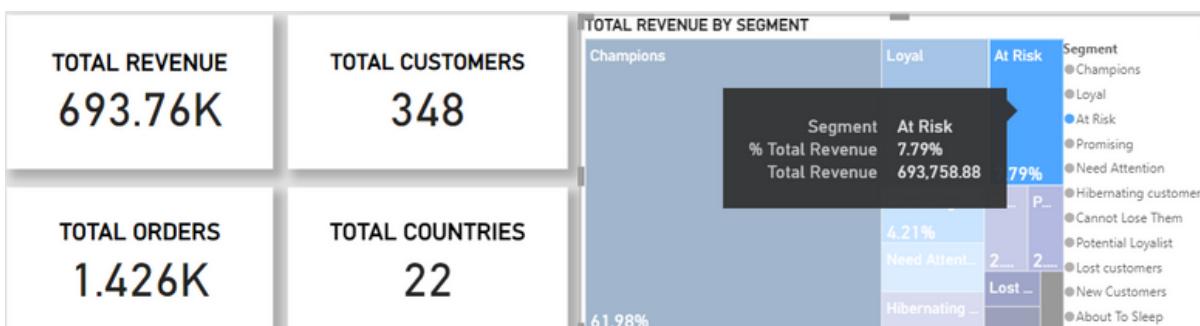
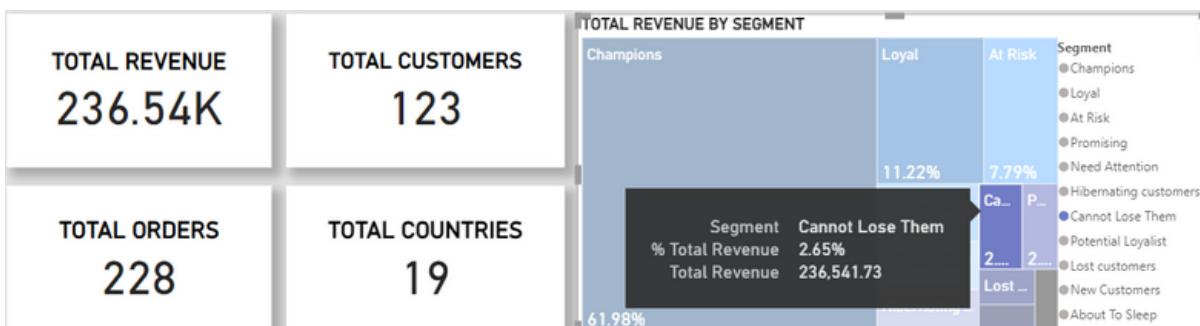
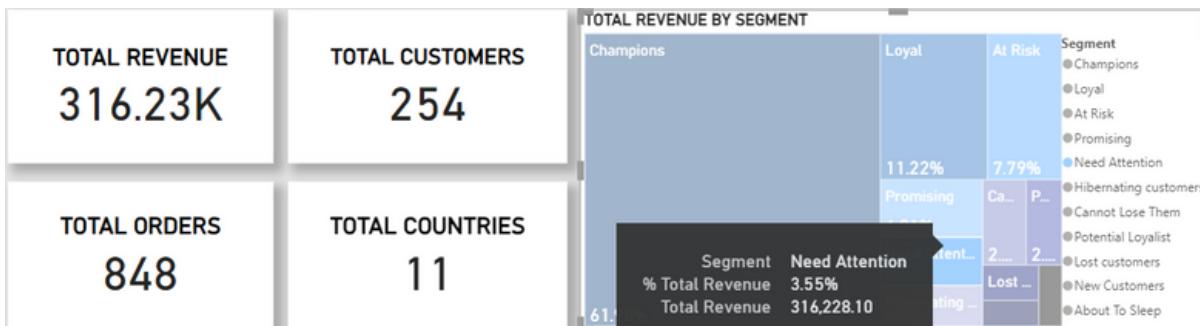
TOTAL CUSTOMERS
243

TOTAL COUNTRIES
16



03 - PHÂN TÍCH

Phân khúc khách hàng Need Attention, Can't Lose Them và At Risk

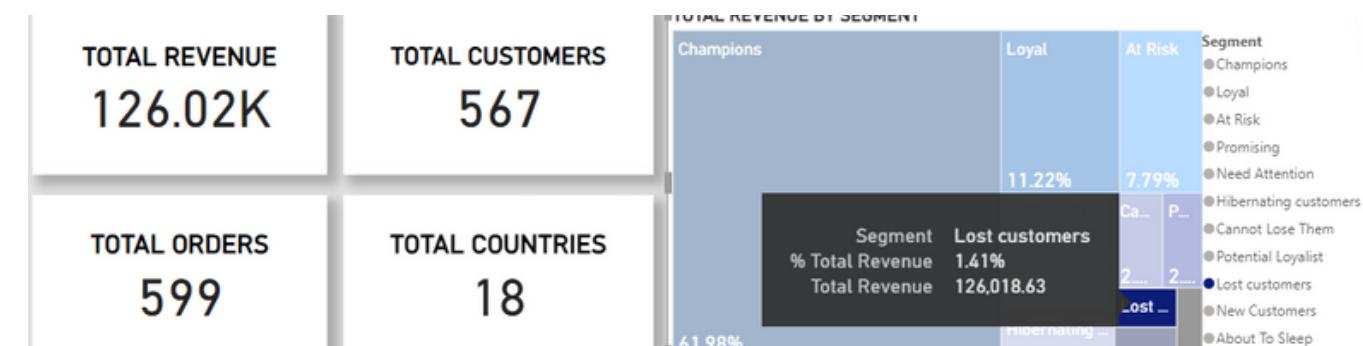
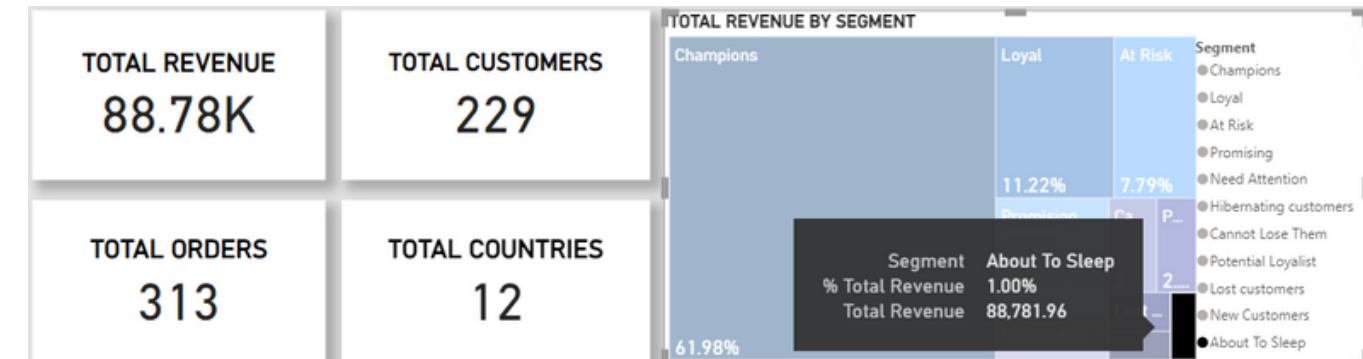


- Nhóm khách hàng không mua lại gần đây và có sức mua trung bình - khá chiếm khoảng gần 14%, với khoảng gần 1,3 triệu bảng Anh
- Nên cải thiện dịch vụ hậu mãi và tạo trải nghiệm mua hàng đặc biệt để khuyến khích khách hàng không mua lại gần đây quay trở lại và tăng doanh số bán hàng.

03 - PHÂN TÍCH

Phân khúc khách hàng About To Sleep, Hibernating và Lost

- Nhóm khách hàng đã lâu không quay lại và sức mua yếu chiếm khoảng gần 6%, với khoảng gần 500 ngàn bảng Anh.
- Đồng thời, sử dụng các chiến lược tiếp thị nhắm đến họ thông qua email marketing, quảng cáo trực tuyến, hoặc thông báo trên mạng xã hội để thu hút họ quay lại mua sắm. Đảm bảo rằng trải nghiệm mua hàng của họ là thuận tiện và hấp dẫn để tăng khả năng quay lại trong tương lai.



04 - ĐỀ XUẤT

TẬP TRUNG CHƯƠNG TRÌNH THEO MÙA VỤ

Tập trung các chương trình thu hút khách hàng vào các tháng 9-10-11

CHƯƠNG TRÌNH KHUYẾN MÃI ĐẶC BIỆT

Tạo ưu đãi đặc biệt và giảm giá để thu hút khách hàng mới và khách mua trở lại.

TRẢI NGHIỆM MUA HÀNG TỐT HƠN

Cải thiện giao diện và hỗ trợ khách hàng

TẬP TRUNG CHƯƠNG TRÌNH THEO QUỐC GIA

Tập trung các chương trình và các sản phẩm cho thị trường châu âu, xử lạnh

TIẾP THỊ TÁI KÍCH HOẠT

Sử dụng email và quảng cáo để thu hút quay lại

THU THẬP PHẢN HỒI

Tìm hiểu nguyên nhân và cải thiện sản phẩm dịch vụ, tập trung vào các sản phẩm có phản hồi và doanh thu tốt (10% tổng lượng sản phẩm).



**THANKS FOR
WATCHING**