



TRƯỜNG ĐẠI HỌC
VĂN LANG
Đạo đức - Ý chí - Sáng tạo

MÔN: KHAI THÁC VÀ PHÂN TÍCH DỮ LIỆU

ĐỀ TÀI:

KỸ NĂNG TRỰC QUAN HÓA DỮ LIỆU

Thực hiện: Nguyễn Huỳnh

Giảng viên hướng dẫn: Lương Thái Hà

Contents

1. Giới thiệu về tập dữ liệu.....	3
2. Phân tích dữ liệu	3
2.1) Về tập dữ liệu	3
2.2) Sơ nội dung của bài phân tích:.....	3
2.3) Các cột và biến tham gia trong từng thư mục.....	3
2.4) Biến mới	4
3. Data Visualization.....	4
3.1) Phân tích mô hình RFM.....	4
3.3 a) Thống kê doanh thu và lợi nhuận của các sản phẩm theo từng quốc gia.....	16
3.3 b) Từng kênh bán hàng khác nhau (online và offline).	18
3.4) Phân tích những mặt hàng nào đóng góp nhiều và những mặt hàng nào có đóng góp ít vào phần trăm lợi nhuận của công ty.....	18
3.5) Phân tích kích bản của từng mặt hàng sản phẩm.	19
3.6) Phân tích về ngày giao hàng dựa trên mức độ ưu tiên.	20
3.7)Phân tích các khoản chi phí bất thường.....	23
3.8) Tiên đoán những điều sẽ xảy ra trong tương lai.	25
3.9) Trendline.....	27
4) Kết luận.....	29
4.1) Nhận định chung.....	29
4.2) Suggestions	29

1. Giới thiệu về tập dữ liệu

- Sales_record là tập dữ liệu bán hàng của các siêu thị trên toàn thế giới. Tập dữ liệu bán hàng này được lấy trên một trang web có tên là “data.world”
- Tập dữ liệu này có 14 cột và 5001 dòng, trong 14 cột đó thì có 5 cột dạng chữ, 2 dữ liệu về thời gian (ngày/tháng/năm) và 7 cột dạng số.
- Tập dữ liệu không chứa missing value, nhưng nếu có missing value thì có thể dùng hàm “ISNULL” để loại bỏ.

2. Phân tích dữ liệu

2.1) Về tập dữ liệu

Nhóm em sẽ dùng hầu hết tất cả các cột trong tập dữ liệu để có thể khai thác tối đa và phân tích sâu vào nhiều vấn đề trong tập dữ liệu.

Các cột được đưa vào phân tích: Region, Country, Item type, Sale channel, Order priority, Order date, Order ID, Ship date, Units sold, Unit price, Unit cost, Total Revenue, Total Cost, Total Profit.

2.2) Sơ nội dung của bài phân tích:

- 1) Phân tích mô hình RFM.
- 2) Tổng doanh thu và lợi nhuận của tất cả các sản phẩm qua các năm.
- 3) Thống kê doanh thu và lợi nhuận của các sản phẩm theo a) từng quốc gia, b) từng kênh bán hàng khác nhau (online và offline).
- 4) Phân tích những mặt hàng nào đóng góp nhiều và những mặt hàng nào có đóng góp ít vào phần trăm lợi nhuận của công ty.
- 5) Phân tích kích bản của từng mặt hàng sản phẩm.
- 6) Phân tích về ngày giao hàng dựa trên mức độ ưu tiên.
- 7) Phân tích các khoản chi phí bất thường.
- 8) Tiên đoán những điều sẽ xảy ra trong tương lai.
- 9) Trendline

2.3) Các cột và biến tham gia trong từng thư mục

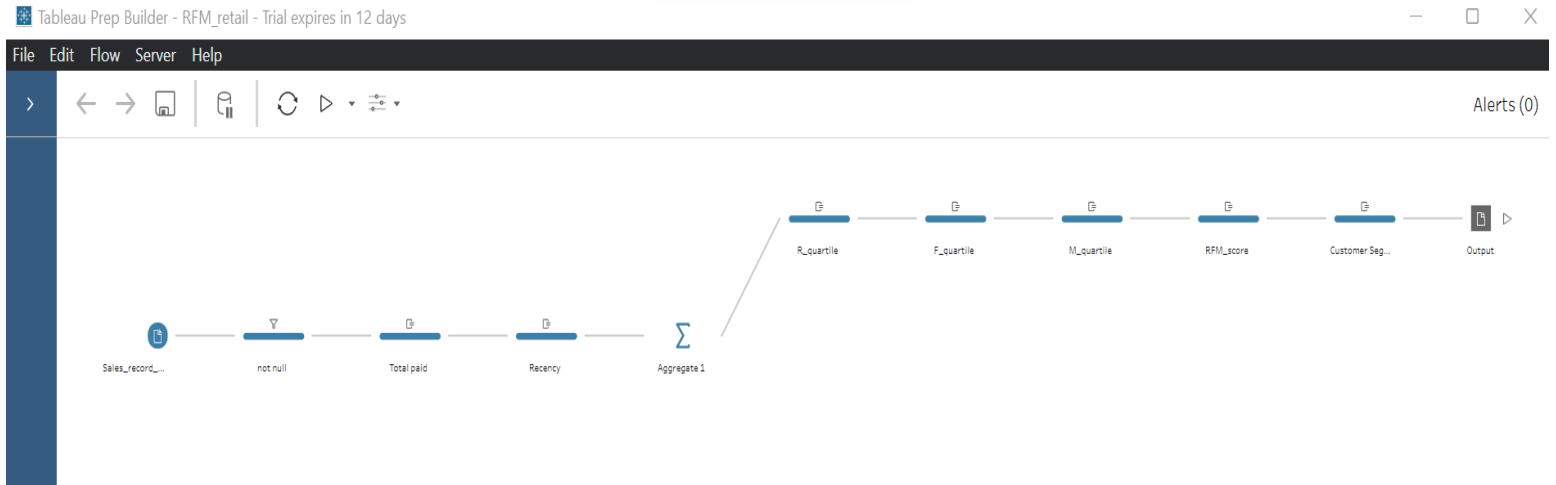
- 1) Cột $F + G + I + J + O$
- 2) Cột $C + F + L + N$
- 3.a) Cột $B + C + L + N$
- 3.b) Cột $C + D + L + N$
- 4) Cột $C + N$
- 5) Cột $C + N$
- 6) Cột $C + E + F + G + H$
- 7) Cột $A + B + M$
- 8) Cột $F + N$
- 9) Cột $C + F + L$

2.4) Biến mới

Tạo thêm cột Total paid (không có trong tập dữ liệu gốc)
Total paid (Cột 15) = Units sold (Cột 9) * Unit price (Cột 10)

3. Data Visualization

3.1) Phân tích mô hình RFM



Ảnh 1: Quá trình phân tích RFM

Để phân tích mô hình RFM nhóm KTPT sẽ sử dụng phần mềm tiền xử lý của Tableau có tên là “Tableau Prep Builder”. Ta sẽ cùng tìm hiểu và phân tích tất cả các bước trong ảnh 1.

Bước 1: Tải tập dữ liệu lên

Bước 2: Loại bỏ missing values (nếu có) bằng phần mềm lọc

Edit Filter

Calculation

Must return a boolean value

not ISNULL([Order ID])

Reference

All

Search

ABS

ACOS

AND

ASC

ASCII

ASIN

ATAN

ATAN2

AVG

CASE

CEILING

CHAR

CONTAINS

COS

COT

ABS(number)

Returns the absolute value of the given number.

Example: ABS(-7) = 7

Calculation is valid

Apply

Save

Ảnh 2: Lệnh not ISNULL() để bỏ đi những hàng chứa missing value

Bước 3 (Total Paid): Tính tổng giá tiền mà khách hàng đã trả

Field Name: Total Paid

Reference: All

Search:

ABS

ACOS

AND

ASC

ASCII

ASIN

ATAN

ATAN2

AVG

CASE

CEILING

CHAR

CONTAINS

COS

COT

COUNT

COUNTA

ABS(number)

Returns the absolute value of the given number.

Example: ABS(-7) = 7

Calculation is valid ^

Apply Save

Ảnh 3: Lấy đơn giá sản phẩm nhân với số lượng bán

Bước 4 (Recency): Trong bước này ta sẽ tính toán xem lần mua hàng cuối cùng của khách hàng là cách đây bao lâu

Edit Field

×

Field Name

last-to-present

`DATEDIFF('day',[Order Date], #2017-07-29#)`

Reference

All

Search

ABS

ACOS

AND

ASC

ASCII

ASIN

ATAN

ATAN2

AVG

CASE

CEILING

CHAR

CONTAINS

COS

COT

COUNT

COUNTA

ABS(number)

Returns the absolute value of the given number.

Example: ABS(-7) = 7

Calculation is valid ^

Apply

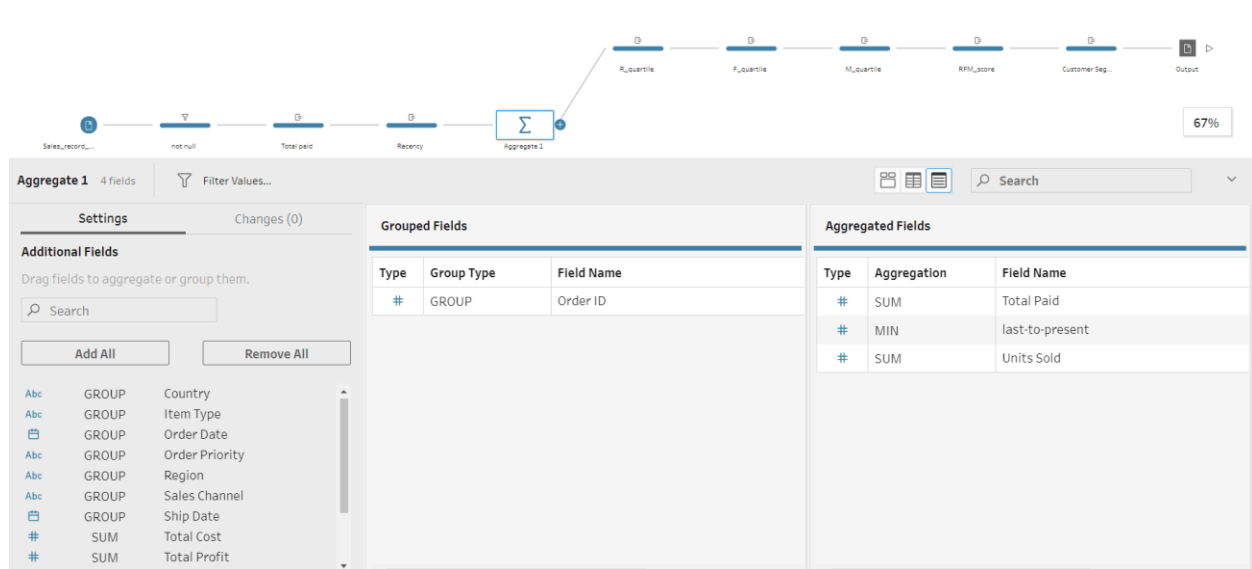
Save

Ảnh 4:

Để có thể tính được ngày mua hàng gần nhất của khách cách đây là bao lâu thì đầu tiên ta sẽ xem ngày cuối cùng trong đơn đặt hàng và sau đó dùng hàm DAYDIFF để tính từ đó đến ngày cần tính là bao nhiêu ngày.

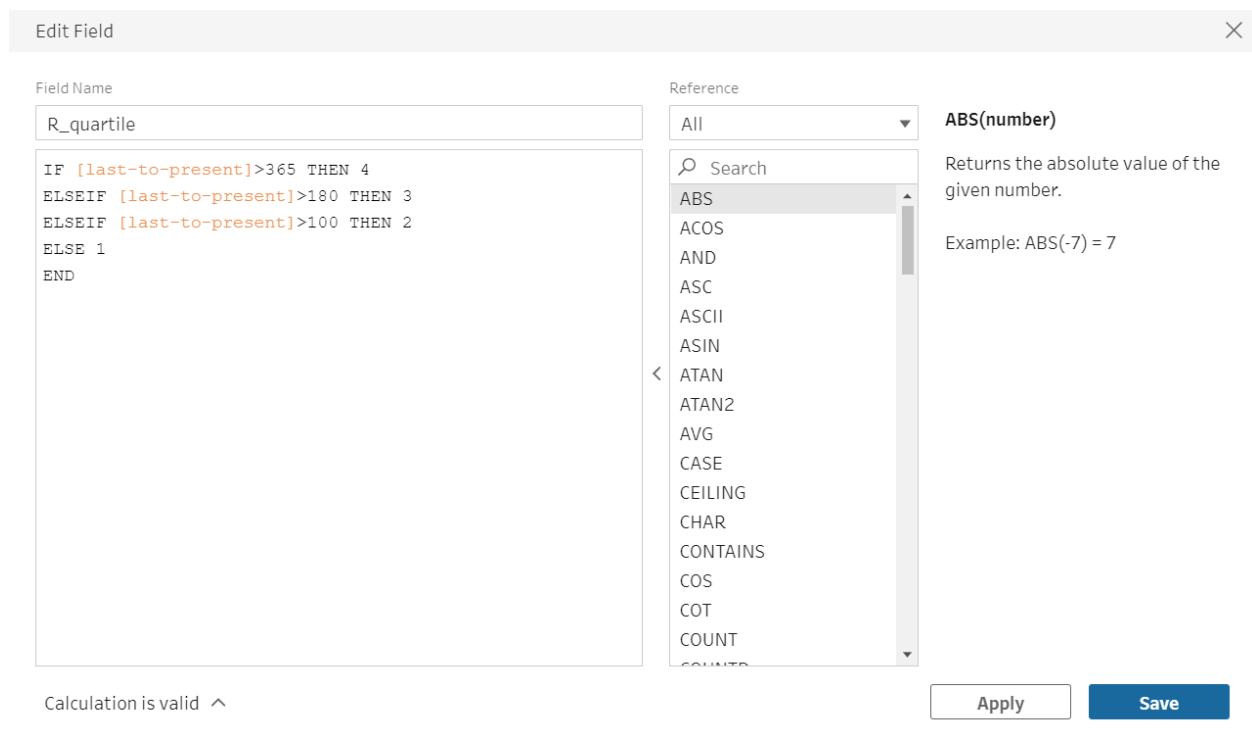
Bước 5: Nhóm các đơn hàng để tính tổng số lượng, tần suất và tổng tiền.

- Đầu tiên sẽ nhóm theo cột OrderID
- Sau đó nhóm Total paid (bước 3) bằng hàm SUM
- Nhóm Recency (bước 4) bằng hàm MIN
- Và nhóm Units sold bằng hàm SUM



Ảnh 5: Sử dụng chức năng Aggregate để nhóm

Bước 6 (R_quartile): Đánh giá khách hàng thông qua lần mua hàng gần nhất qua 4 mức và theo thứ tự mức 1 sẽ là mức cao nhất và 4 sẽ là mức thấp nhất



Ảnh 6

Theo ảnh 6, có thể sắp xếp các loại khách hàng vào 4 mức đó là:

Mức 4 (> 365 ngày): Ở mức này thời gian mua hàng gần nhất của một khách hàng là cách đây 1 năm

Mức 3 (> 180 ngày): khách hàng mua cách đây 6 tháng

Mức 2 (> 100 ngày): khách hàng mua hàng cách đây 3 tháng

Mức 1: khách hàng mua hàng gần đây là dưới 3 tháng

Bước 7 (F_quartile): Ở bước này ta sẽ tính tần suất mua hàng của khách hàng dựa trên các đơn hàng đã bán ở cột Units sold.

Ở bước này ta có thể phân loại khách hàng theo tiêu chí khách hàng trung thành hay khách hàng không trung thành.

Field Name: F_quartile

Reference: All

ABS(number)

Returns the absolute value of the given number.

Example: ABS(-7) = 7

```
IF [Units Sold]>=7576 THEN 1
ELSEIF [Units Sold]>=5123 THEN 2
ELSEIF [Units Sold]>=2453 THEN 3
ELSE 4
END
```

Calculation is valid ^

Apply Save

Ảnh 7: Những số liệu trên được tính bằng hàm QUARTILE dựa theo cột Units sold

Mức 1: Khách hàng mua trên 7576 đơn hàng

Mức 2: khách hàng mua trên 5123 đơn hàng

Mức 3: khách hàng mua trên 2453 đơn hàng

Mức 4: khách hàng mua dưới 2453 đơn hàng

Bước 8 (M_quartile): Đánh giá khách hàng dựa trên số tiền mà khách hàng đã trả

Edit Field

×

Field Name

M_quartile

```

IF [Total Paid]>=1839975 THEN 1
ELSEIF [Total Paid]>=779409 THEN 2
ELSEIF [Total Paid]>=257416 THEN 3
ELSE 4
END

```

Reference

All

Search

ABS

ACOS

AND

ASC

ASCII

ASIN

ATAN

ATAN2

AVG

CASE

CEILING

CHAR

CONTAINS

COS

COT

COUNT

COUNTA

ABS(number)

Returns the absolute value of the given number.

Example: ABS(-7) = 7

Calculation is valid ^

Apply

Save

Ảnh 8: Những số liệu trên được tính bằng hàm QUARTILE dựa theo cột Total paid (2.4)

Mức 1: Khách hàng đã trả trên 1.839.975 USD

Mức 2: Khách hàng đã trả trên 779.409 USD

Mức 3: Khách hàng đã trả trên 257.416 USD

Mức 4: khách hàng mua dưới 257.416 USD

Bước 9 (RFM_score): Gộp 3 bước trên là R_quartile, F_quartile và M_quartile.

Edit Field

Field Name

RFM_score

STR([R_quartile])+STR([F_quartile])+STR([M_quartile])

Reference

All

Search

ABS

ACOS

AND

ASC

ASCII

ASIN

ATAN

ATAN2

AVG

CASE

CEILING

CHAR

CONTAINS

COS

COT

COUNT

ABS(number)

Returns the absolute value of the given number.

Example: ABS(-7) = 7

Calculation is valid ^

Apply

Save

Ảnh 9: Dùng hàm STR để đổi thành dạng string

Sales_record...

not null

Total paid

Recency

Aggregate 1

R_quartile

F_quartile

M_quartile

RFM_score

Customer Seg...

Output

67%

RFM_score

8 fields

5K rows

Filter Values...

Rename Fields...

Create Calculated Field...

Search

Changes (1)

Calculated Field

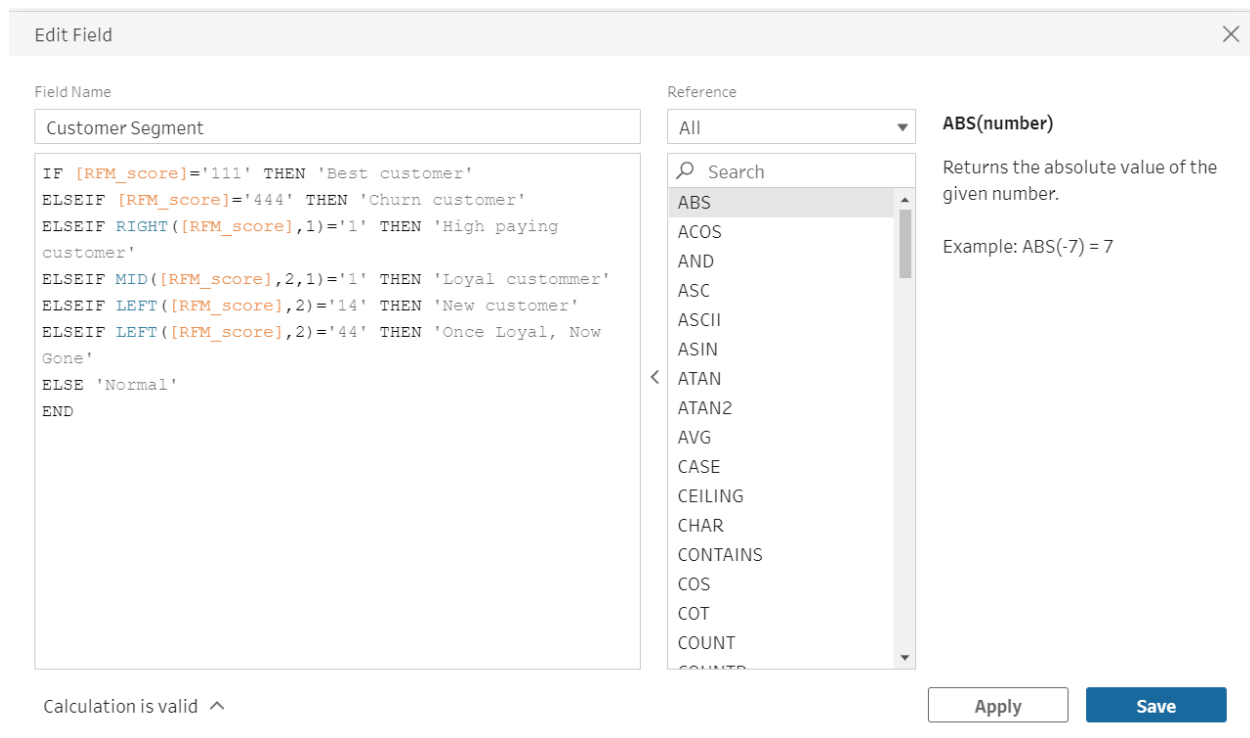
[RFM_score]

STR([R_quartile])+STR([F_quartile])+STR([M_quartile])

Abc	#	#	#	#
RFM_score	F_quartile	R_quartile	M_quartile	Total Paid
414	1	4	4	75,843.57
444	4	4	4	104,517.3
423	2	4	3	281,283.60000000003
422	2	4	2	1,386,936.24
143	4	1	3	260,385.6
411	1	4	1	6,340,180.56
433	3	4	3	459,098.8
431	3	4	1	3,267,120.57
421	2	4	1	3,129,914.8
443	4	4	3	343,305
422	2	4	2	1,670,041.76

Ảnh 10: Sau khi gộp sẽ ra được kết quả như trên

Bước 10 (Customer Segment): Là bước quan trọng nhất vì trong bước này ta sẽ đánh giá tổng quan các khách hàng dựa theo vài tiêu chí như sau (theo lý thuyết thì có đến 64 tiêu chí):



Ảnh 11

Nếu khách hàng có số điểm “111” thì sẽ xếp vào loại “Khách hàng tốt nhất” cần phải chăm sóc kỹ những khách hàng này

Ngược lại nếu khách hàng có số điểm là “444” thì sẽ là “Khách hàng đã ngưng sử dụng dịch vụ của cửa hàng”

Nếu khách hàng có điểm bên phải (M) bằng “1” thì là “Khách hàng chi rất nhiều tiền cho các sản phẩm”

Nếu khách hàng có điểm ở giữa (F) bằng “1” thì là “Khách hàng trung thành” vì có tần suất mua hàng thường xuyên

Nếu khách hàng có điểm bên trái và ở giữa (R & F) bằng “14” thì là “Khách hàng mới”

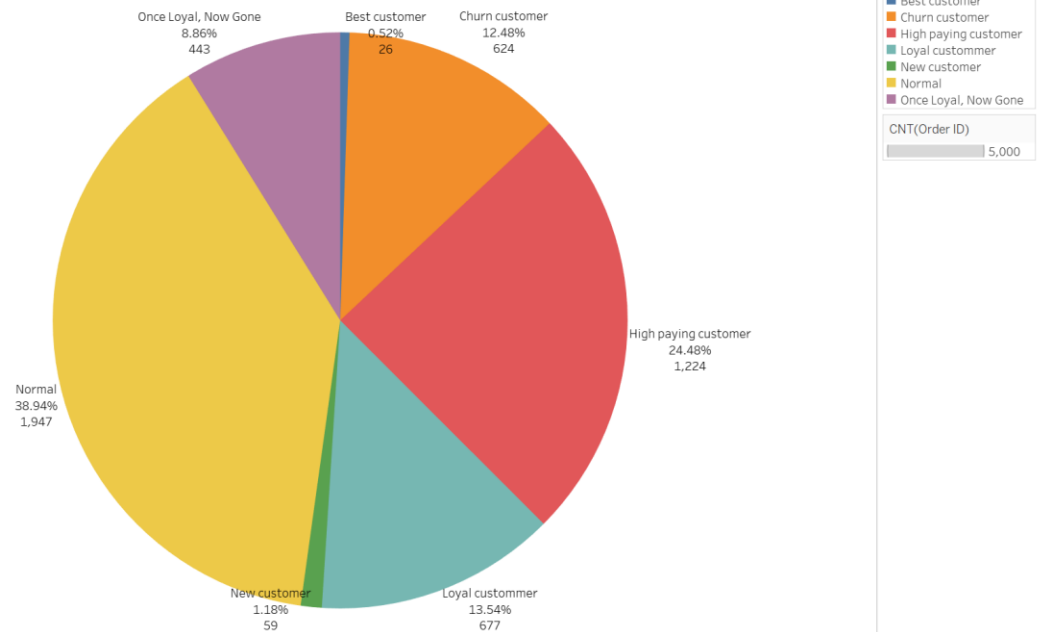
Nếu khách hàng có điểm bên trái và ở giữa (R & F) bằng “44” thì là “Khách hàng đã từng trung thành”

Bước 11 (Output): Lưu ra file RFM_retail.csv

Bước 12: Tải tệp RFM_retail vào Tableau

Bước 13: Trực quan hóa dữ liệu RFM

RFM

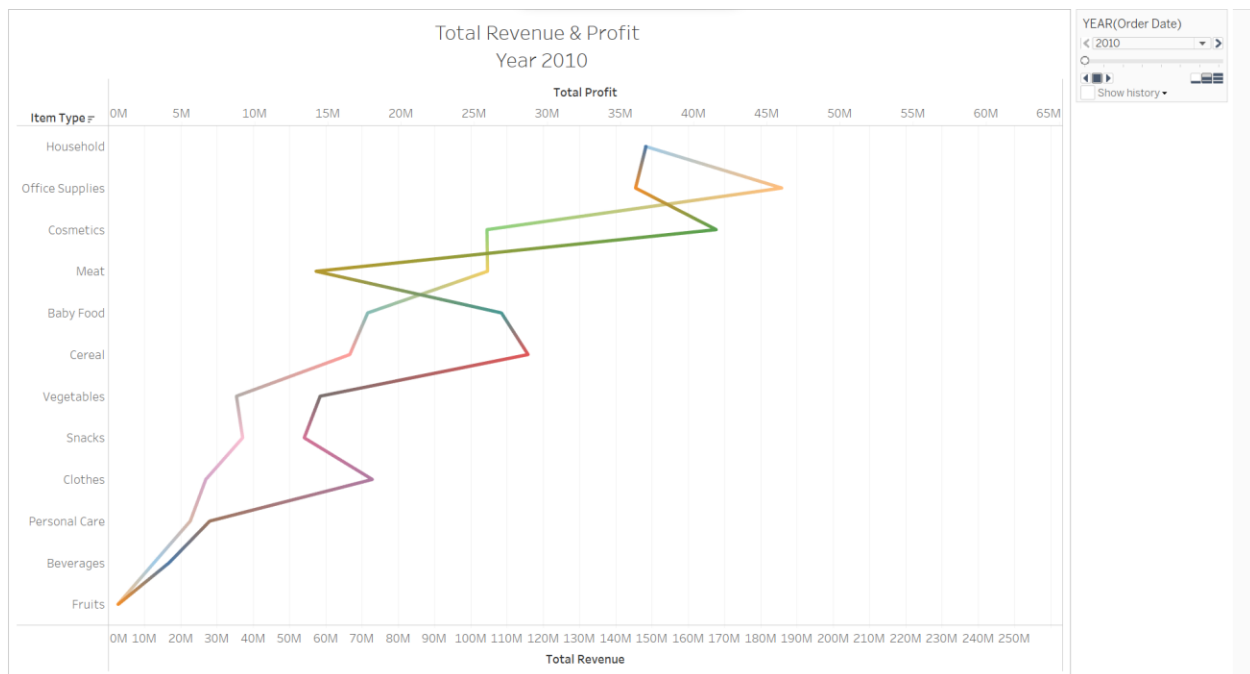


Ảnh 12

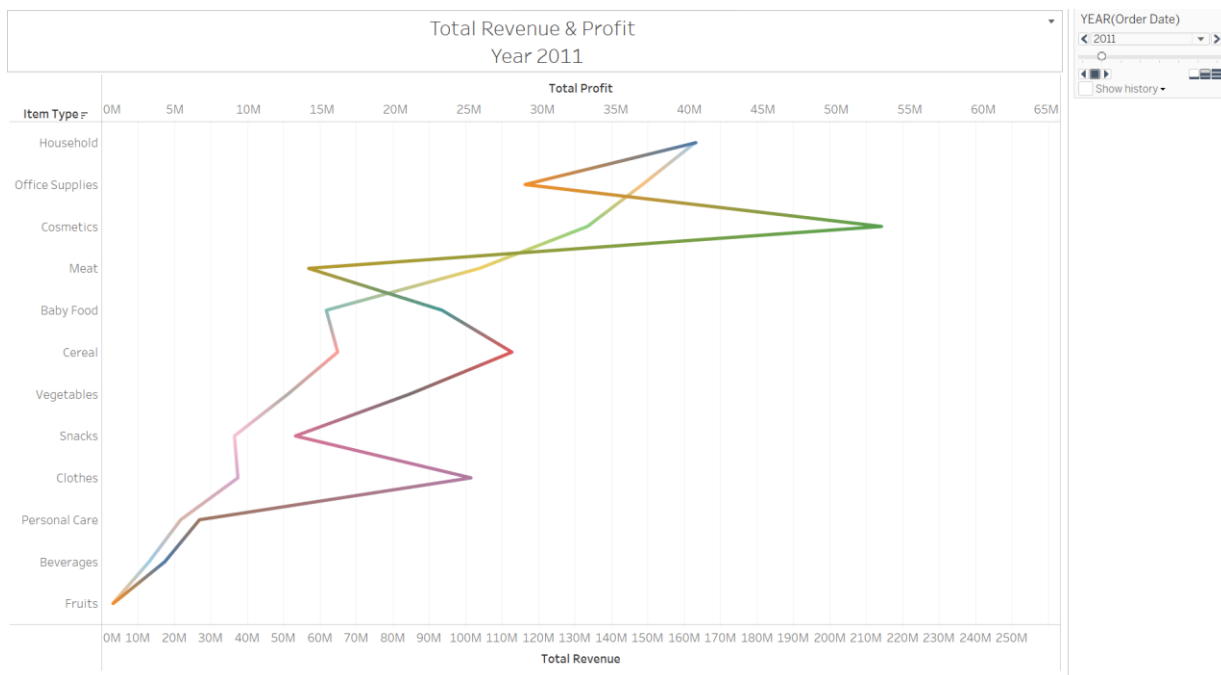
Sau khi vẽ biểu đồ ta dễ dàng thấy được những loại khách hàng và chiếm bao nhiêu phần trăm để ta có thể đưa ra các chiến lược chăm sóc những khách hàng tiềm năng và các chiến lược Marketing để thu hút khách hàng cũ. Ví dụ như phát Voucher, Discount theo từng loại khách hàng.

3.2) Tổng doanh thu và lợi nhuận của tất cả các sản phẩm qua các năm.

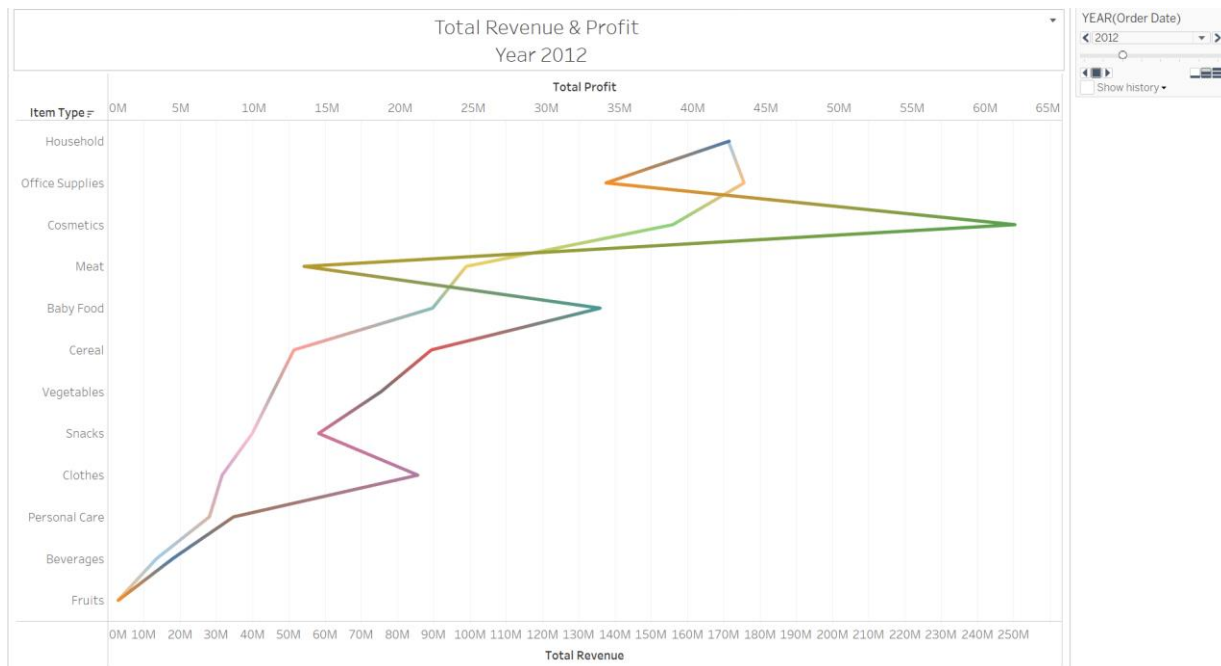
Một chart Animation giúp ta dễ dàng đối chiếu tổng doanh thu và lợi nhuận qua từng năm của tất cả các sản phẩm.



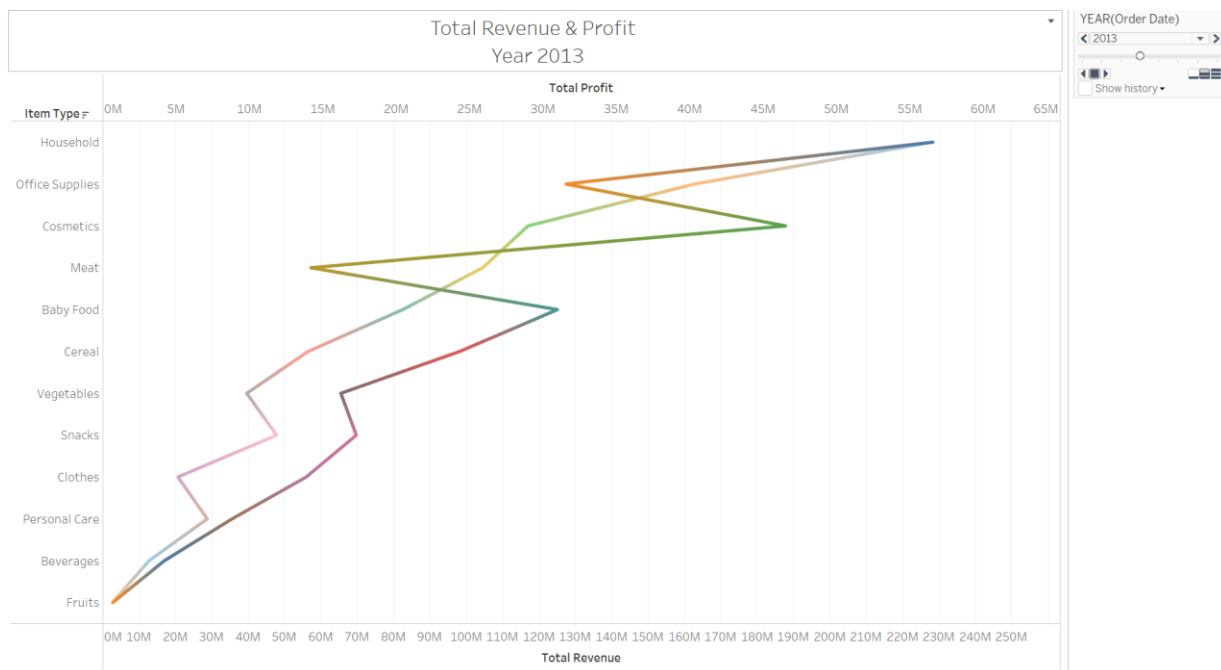
Ảnh 13



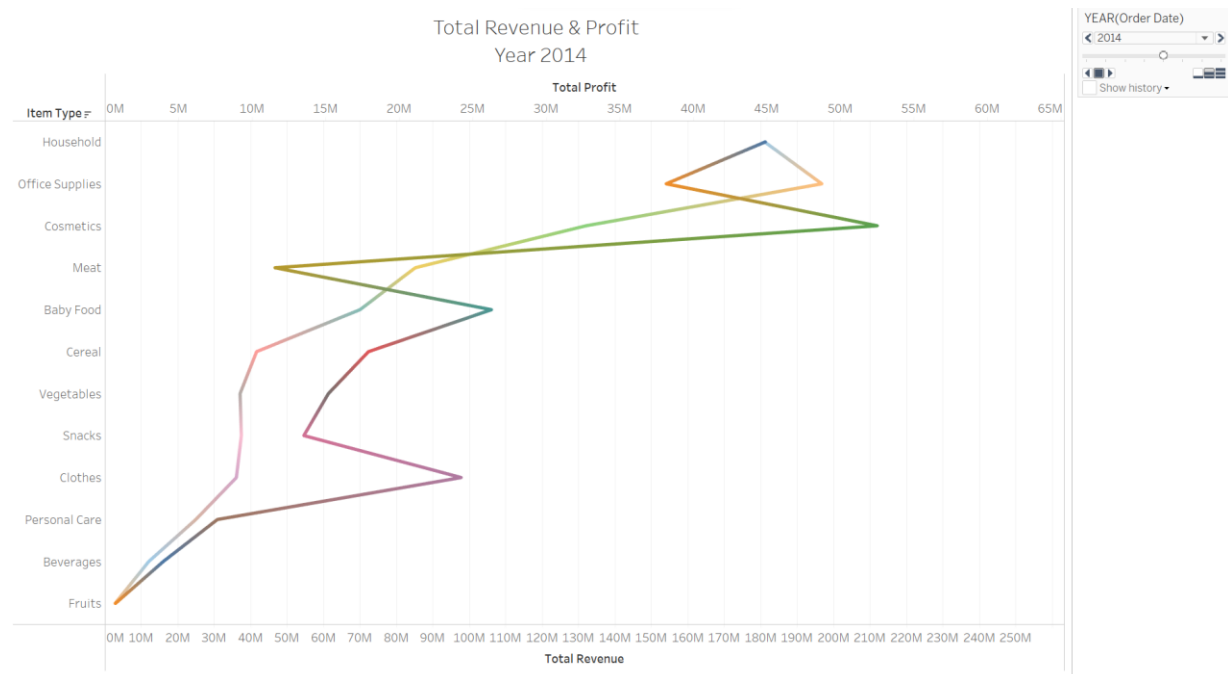
Ảnh 14



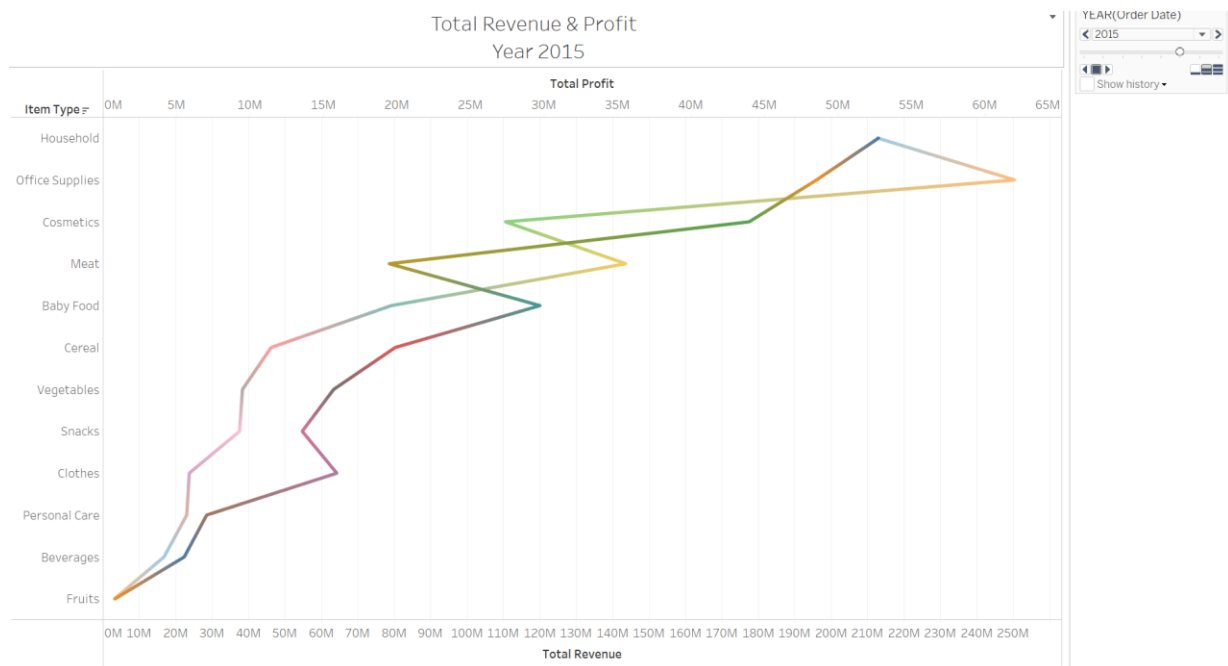
Ảnh 15



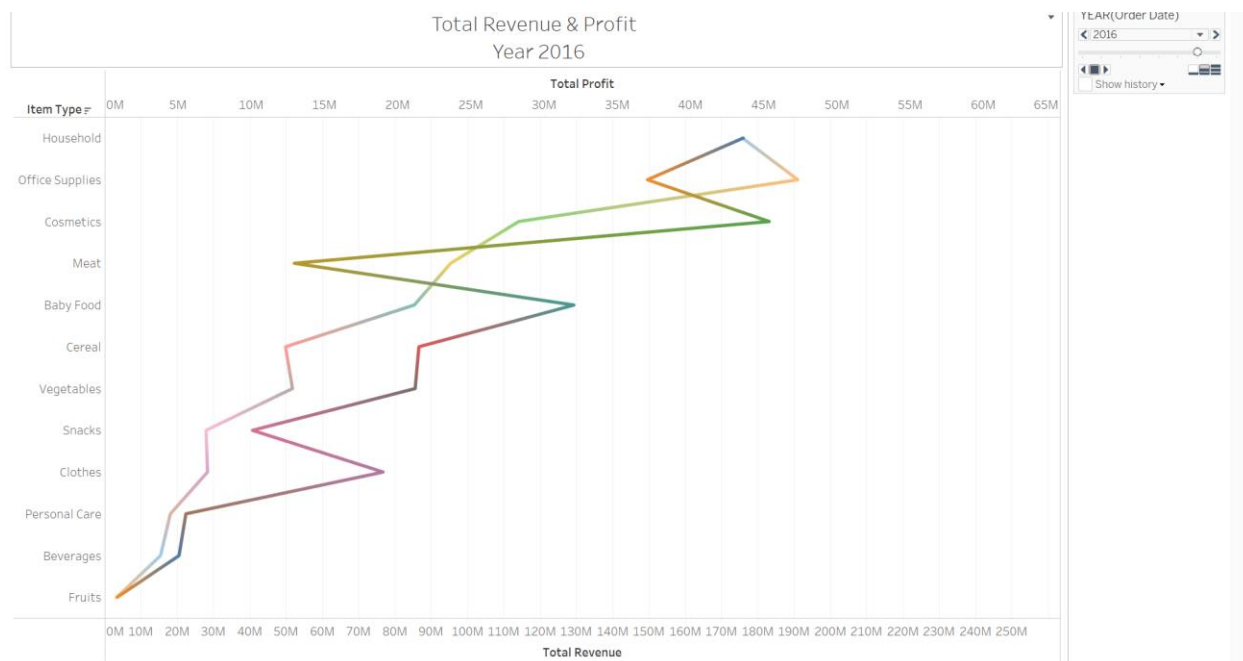
Ảnh 16



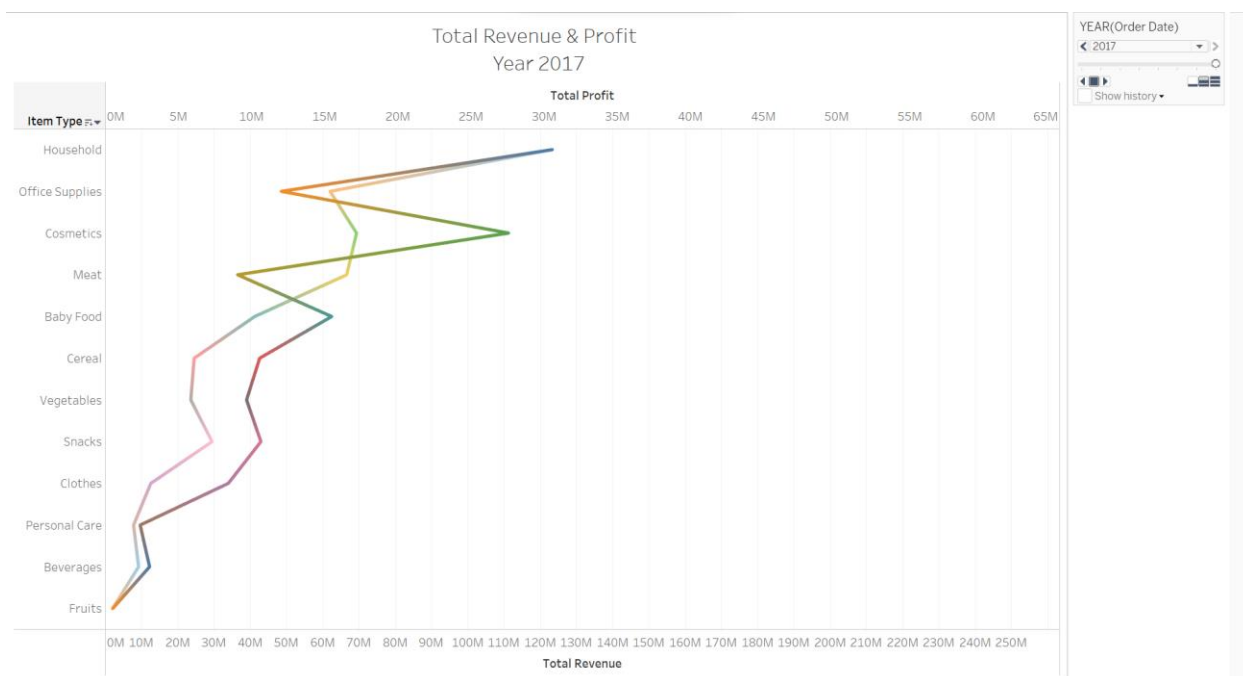
Ảnh 17



Ảnh 18



Ảnh 19



Ảnh 20

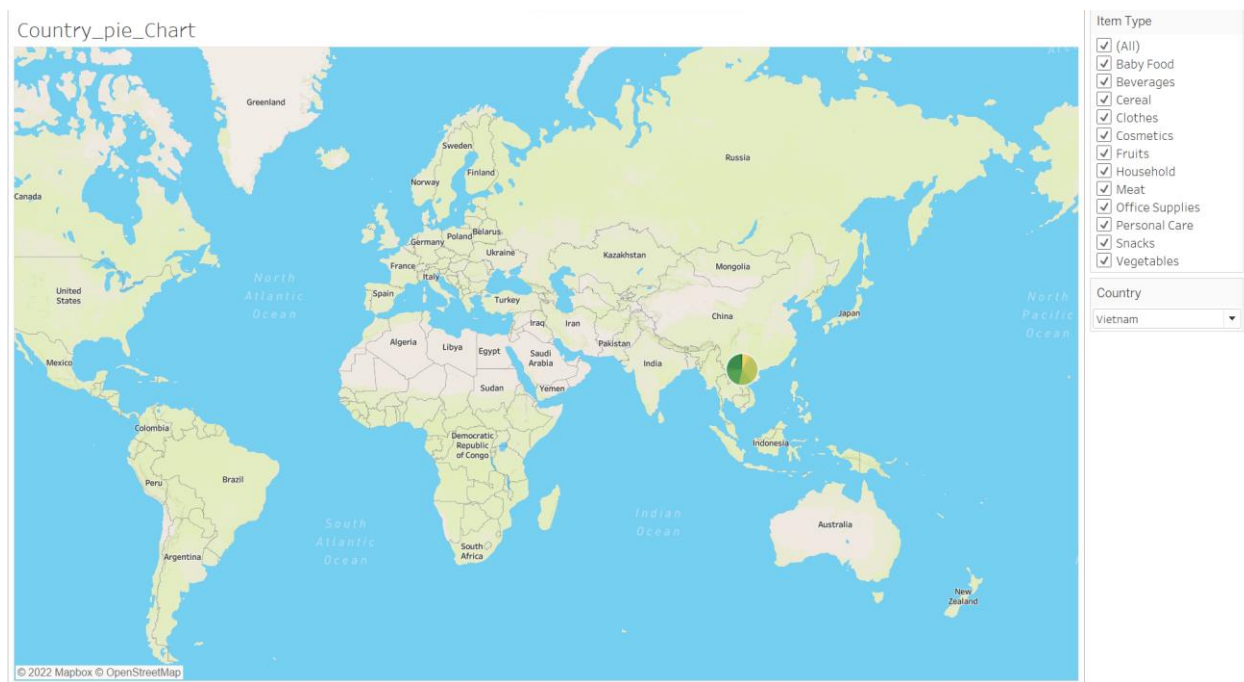
Tổng doanh thu và lợi nhuận giảm đáng kể so với những năm đầu

3.3 a) Thống kê doanh thu và lợi nhuận của các sản phẩm theo từng quốc gia

Country_pie_Chart thể hiện tổng doanh thu và lợi nhuận của tất cả các sản phẩm trên từng quốc gia và độ lớn của những hình tròn phụ thuộc vào tổng doanh thu và lợi nhuận của nước đó.



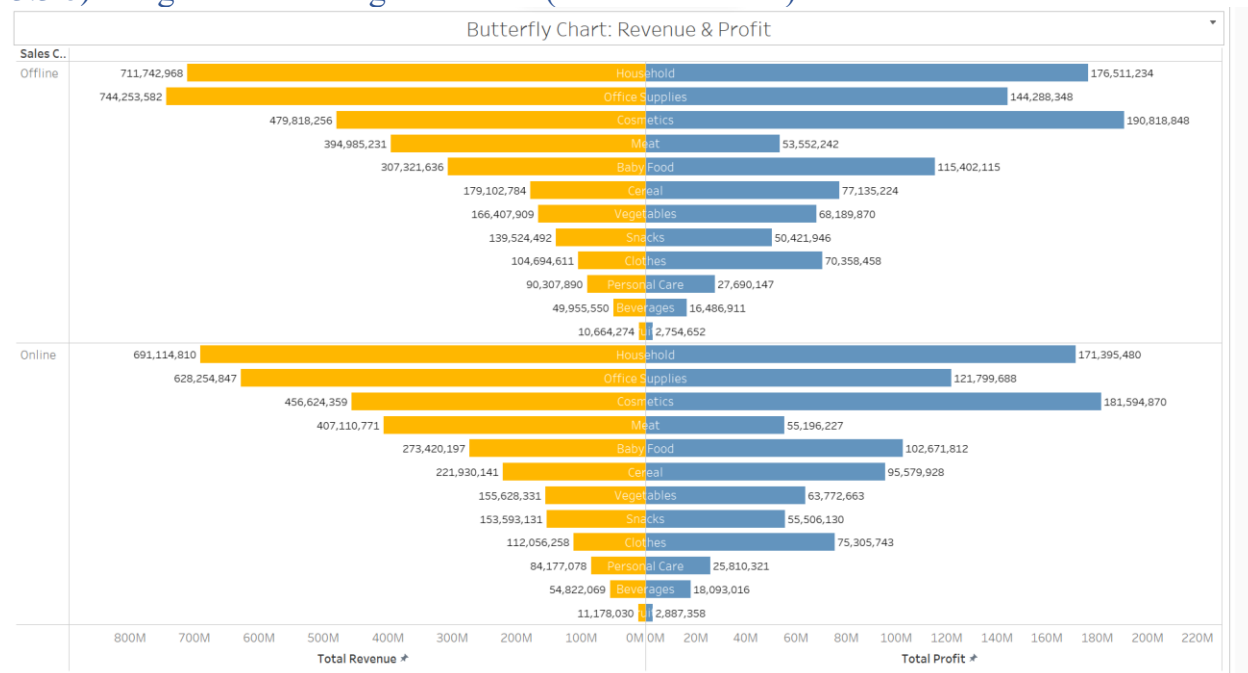
Ảnh 21: Country_pie_Chart



Ảnh 22

Theo ảnh 22 ta có thể sử dụng bộ lọc để xem các quốc gia nào là quốc gia bán hàng tiềm năng hoặc cũng có thể xem doanh thu và lợi nhuận của từng sản phẩm được bán trên toàn quốc gia

3.3 b) Từng kênh bán hàng khác nhau (online và offline).

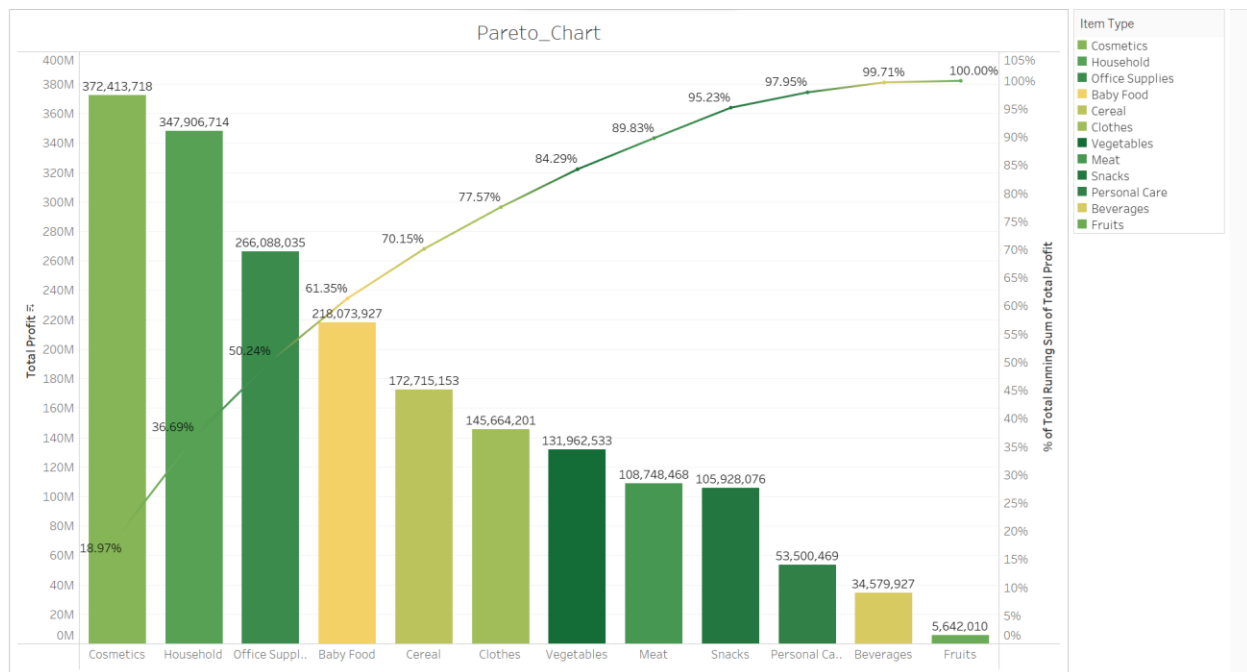


Ảnh 23

Sử dụng Butterfly Chart cho ta nhìn thấy được sự so sánh của tổng doanh thu và lợi nhuận được thể hiện qua 2 kênh bán hàng online và offline.

Và cũng cho ta thấy được có những mặt hàng có doanh thu cao nhưng lợi nhuận lại thấp hơn những mặt hàng có doanh thu thấp. Việc này có thể hiểu được là do các sản phẩm như Mỹ phẩm (Cosmetics) hay thời trang (clothes) là những sản phẩm đắt tiền và bán có lời nhiều hơn các sản phẩm khác.

3.4) Phân tích những mặt hàng nào đóng góp nhiều và những mặt hàng nào có đóng góp ít vào phần trăm lợi nhuận của công ty



Ảnh 24: Pareto_Chart

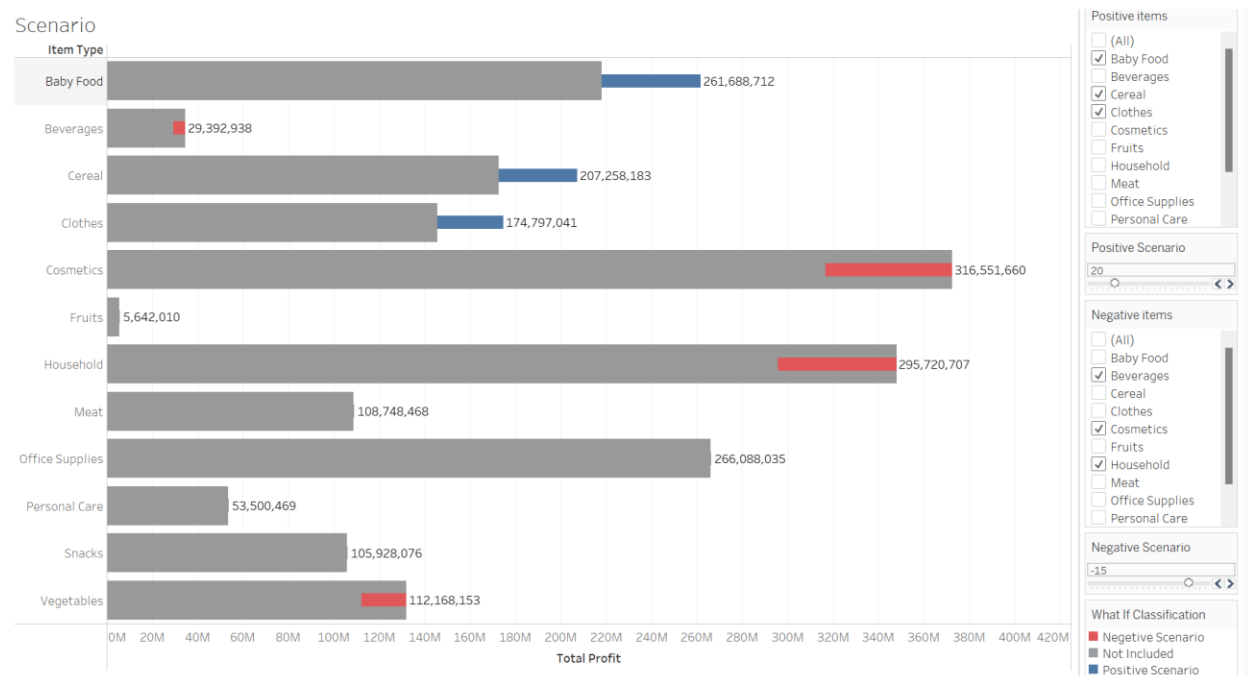
Đồ thị Pareto thường để dùng trong nghiên cứu kinh doanh và để tìm ra các nguyên nhân hay yếu tố tác động lên mục tiêu nghiên cứu, xem các yếu tố đóng góp bao nhiêu phần trăm vào tổng thể.

Trong ảnh 24 thì ta thấy Cosmetics đóng góp nhiều nhất về phần trăm lợi nhuận và fruits đóng góp ít nhất.

Ta thấy được 5 sản phẩm đầu tiên chiếm 70% lợi nhuận của công ty đó là Cosmetics, Household, Officesupplies, Babyfood, Cereal còn 7 sản phẩm còn lại chỉ chiếm 30% trong tổng thể.

3.5) Phân tích kịch bản của từng mặt hàng sản phẩm.

Phân tích kịch bản hay được sử dụng trong phân tích kinh doanh, thị trường, đầu tư, khách hàng,...



Ảnh 25: Scenario

Ta sẽ dùng Scenario để phân tích lợi nhuận của công ty bán hàng xem những mặt hàng nào tăng hoặc giảm và nó ảnh hưởng như thế nào đến sự tăng giảm của tổng lợi nhuận công ty.

Trường hợp trong ảnh 25 dự đoán những sản phẩm như Babyfood, Cereal và Clothes là những sản phẩm tăng lên 20% lợi nhuận

Còn những sản phẩm như Beverage, Cosmetic, Household và Vegetable giảm 15% lợi nhuận.

Và như vậy sẽ tổng lợi nhuận của siêu thị sẽ thay đổi theo tổng phần trăm đóng góp của những sản phẩm (phần 4)

3.6) Phân tích về ngày giao hàng dựa trên mức độ ưu tiên.

Ta sẽ phân tích tình hình giao hàng của siêu thị là giao hàng đúng giờ hoặc giao hàng sớm hay trễ dựa trên cột “Order priority”

Ta có 4 mức độ ưu tiên trong tập dữ liệu là:

Day to ship priority

Sales_record_data

×

CASE [Order Priority]
WHEN "I" THEN 0
WHEN "H" THEN 1
WHEN "M" THEN 3
WHEN "L" THEN 6
END

The calculation is valid.

2 Dependencies ▾

Apply

OK

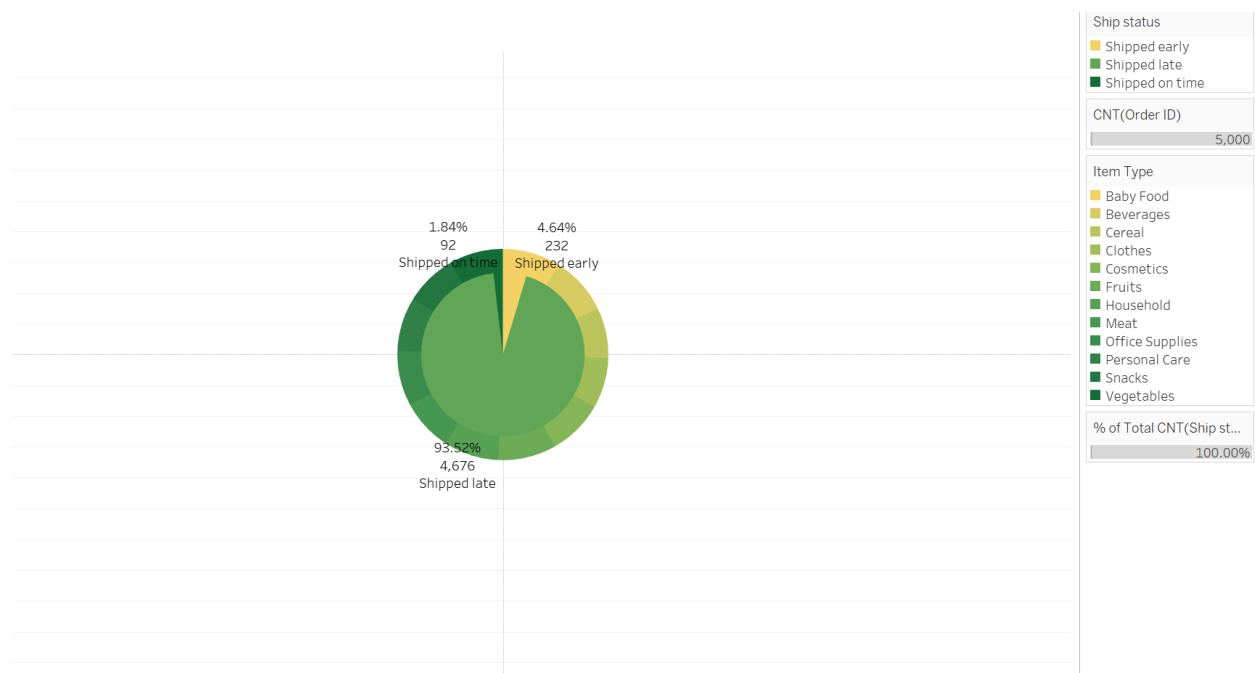
Ảnh 25: Order priority

“I”: Immediately: tức là giao ngay

“H”: High: giao 1 ngày sau đó

“M”: Medium: giao sau 3 ngày

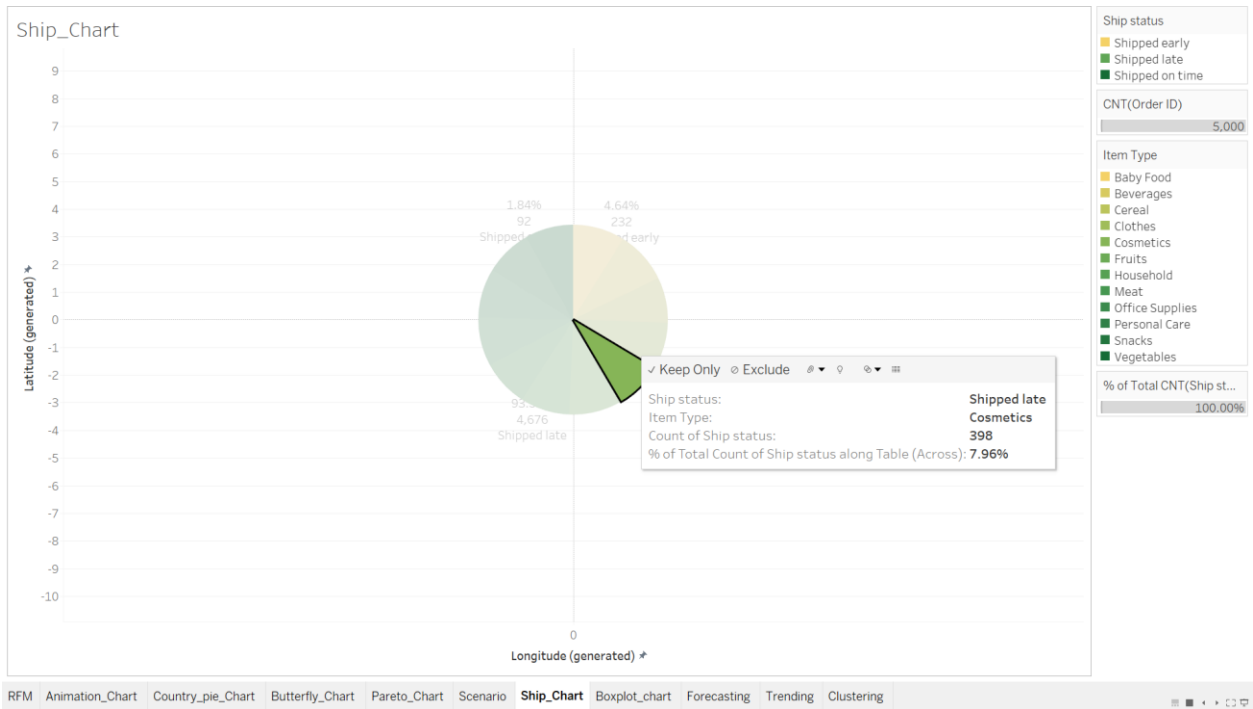
“L”: Low: giao sau 6 ngày



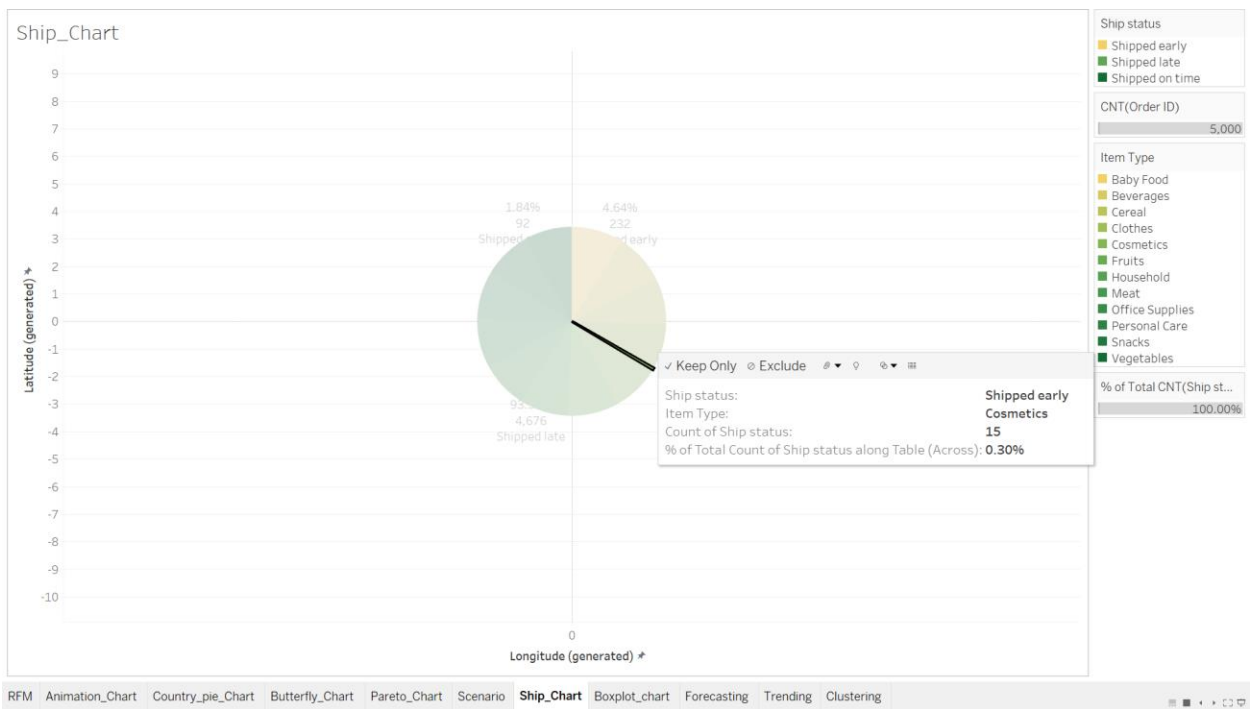
Ảnh 26: Ship_Chart

Theo dữ liệu ảnh 26 cho thấy có 92 đơn hàng, chiếm 1.84% của tổng thể là giao đúng giờ. 232 đơn hàng, chiếm 4.64% là giao sớm. 4676 đơn hàng, chiếm 93.52% là giao hàng trễ.

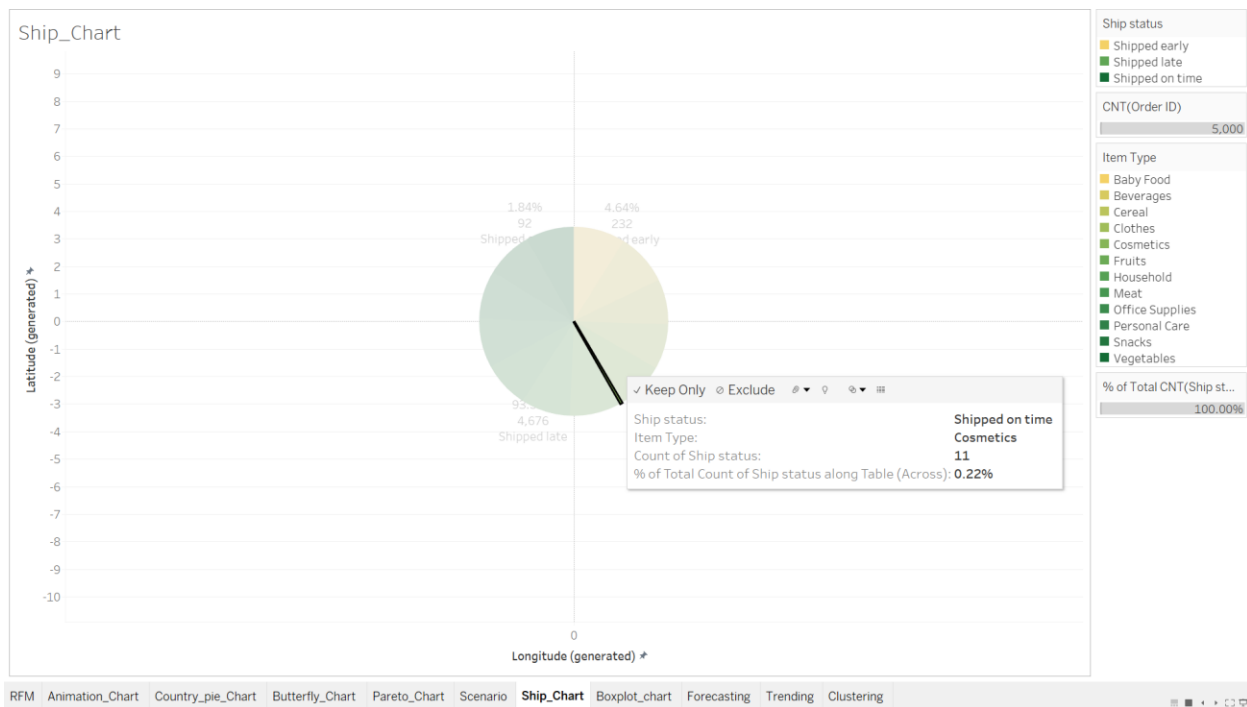
Ta cũng có thể xem được cụ thể những đơn hàng và xem trong đó một đơn hàng chiếm bao nhiêu phần trăm giao sớm, trễ hay đúng giờ



Anh 27



Anh 28



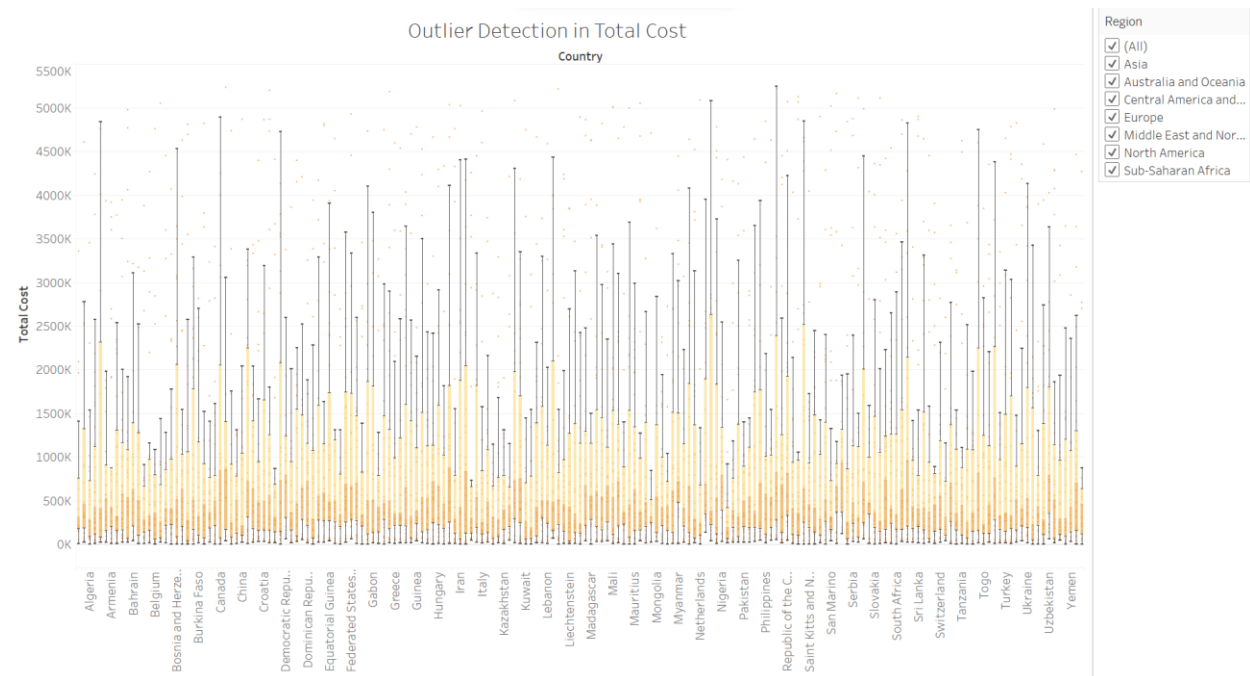
Ảnh 29

Ví dụ như theo hình 27, 28 và 29 ta thấy được loại hàng mỹ phẩm có 7.96% trên 93.52% là giao hàng trễ còn 0.3% trên 4.64% là giao hàng sớm. Và 0.22% trên 1.84% là giao hàng đúng giờ.

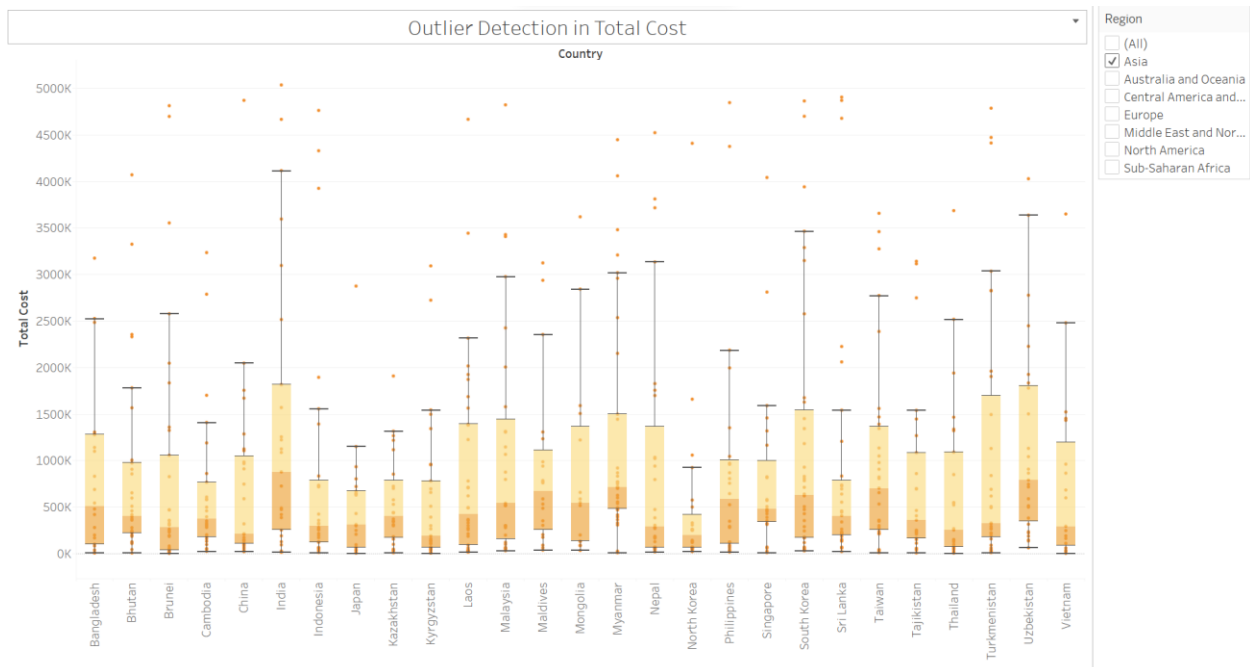
3.7) Phân tích các khoản chi phí bất thường.

Sử dụng đồ thị Boxplot để phân tích các khoản chi phí bất thường của các cửa hàng trên tất cả các nước và có thể lọc theo từng vùng. Các khoản chi phí bất thường thường là các khoản chi phí quá cao hoặc quá thấp so với khoản chi phí trung bình.

Đây là giai đoạn tiền xử lý trước khi phân tích dữ liệu nhằm loại bỏ các khoản bất thường.



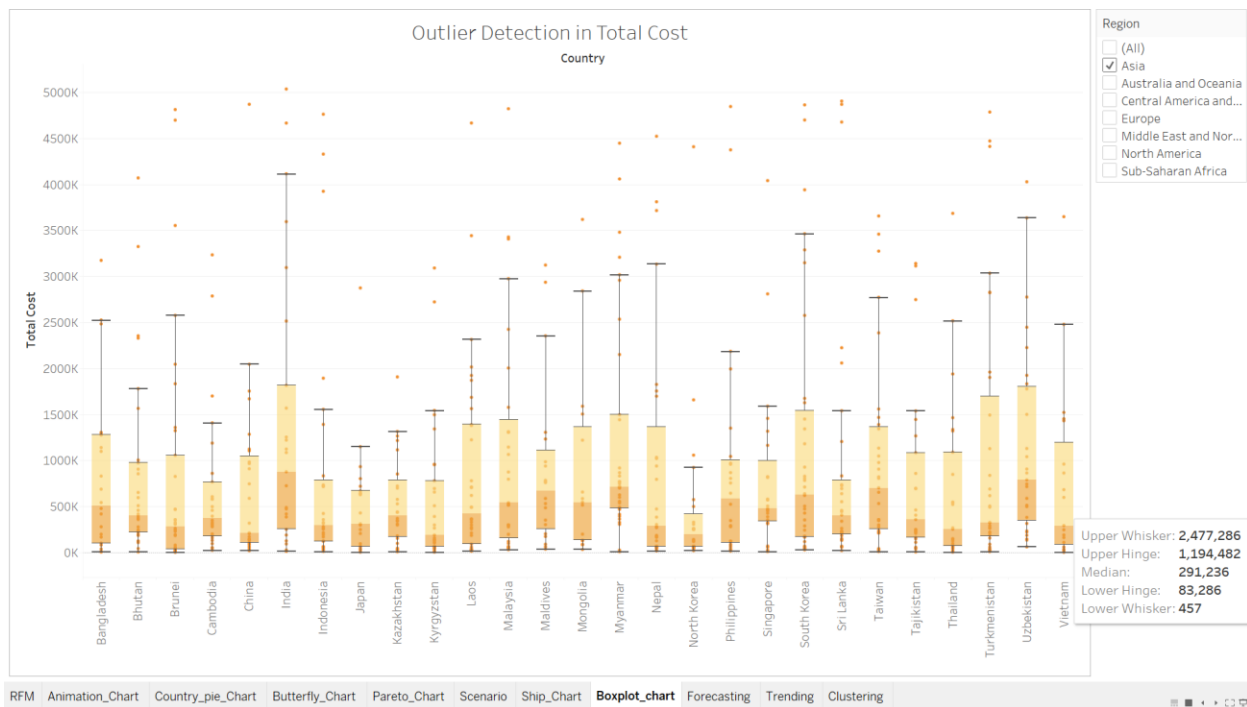
Ảnh 30



Ảnh 31

Ví dụ như ở ảnh 31 là lọc xem các khoản chi phí bất thường của các cửa hàng ở khu vực châu á.

Ta cũng có thể xem được điểm trung vị của nước nào là cao nhất trong khu vực châu á. Ví dụ theo ảnh 31 ta thấy được điểm trung vị của Ấn độ là cao nhất nghĩa là Ấn độ có chi phí trung bình cao nhất.

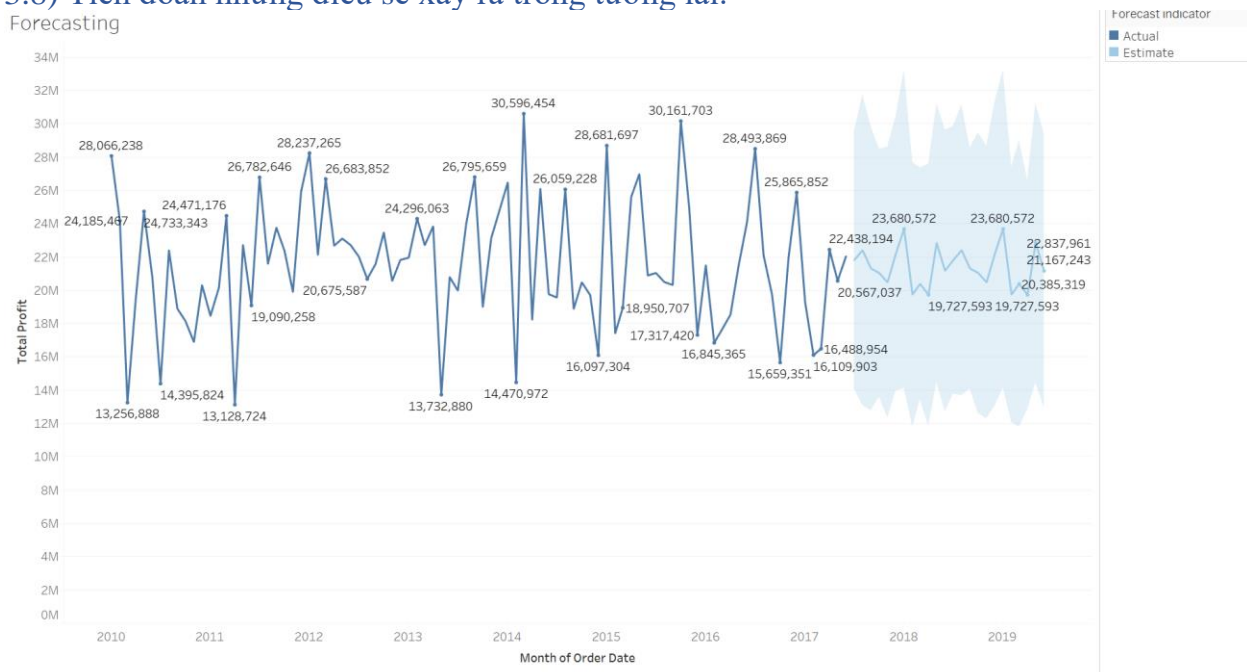


Ảnh 32: Total cost in VietNam

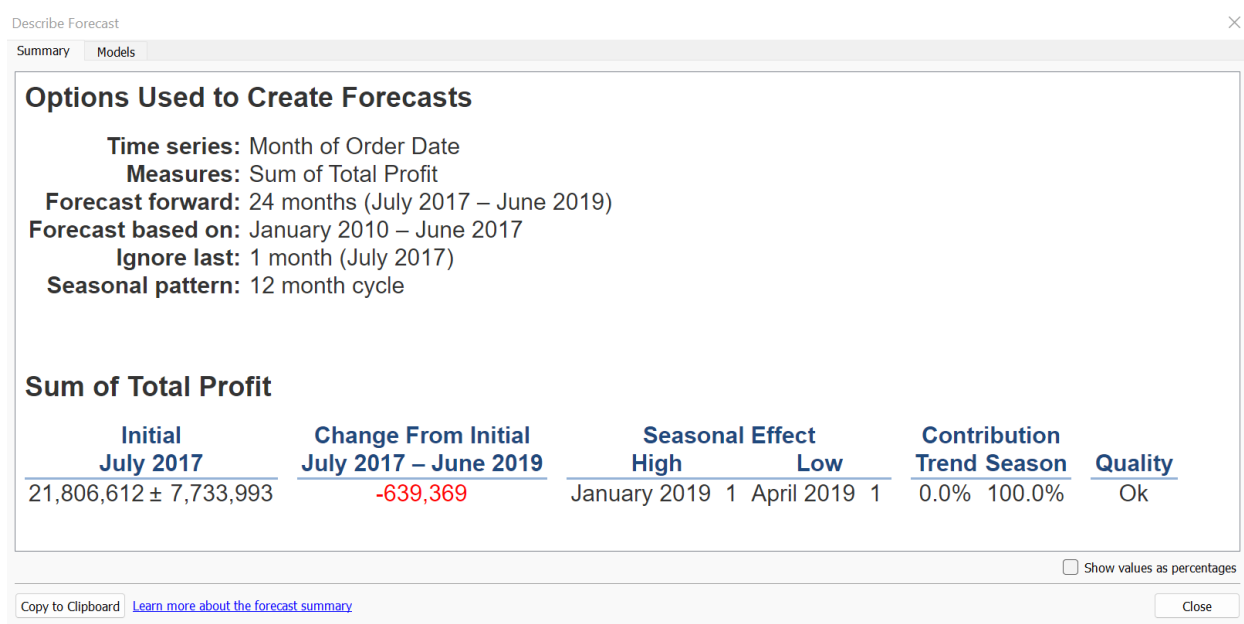
Ví dụ như đối với Việt Nam thì các khoản chi phí trên 2,477,286 hay dưới 457 là những khoản chi phí bất thường

3.8) Tiên đoán những điều sẽ xảy ra trong tương lai.

Forecasting



Ảnh 33: Forecasting



Ảnh 34: Summary describe Forecast

Dựa vào đường dự báo lợi nhuận ở hình 33 ta thấy được lợi nhuận trong 24 tháng tiếp theo là tăng hay giảm và những dự báo đó được tính theo mô hình như hình 34

Dự báo được dựa trên thời gian từ January 2010 – June 2017 và bỏ qua 1 tháng là July 2017

Ta có khoảng thời gian dự báo đầu tiên là July 2017 sau đó từ July 2017 đến June 2019 thì giảm 639,369 tương ứng giảm 2,9% lợi nhuận

Seasonal Effect hiển thị giá trị cao và thấp của thành phần theo mùa của chu kỳ mùa đầy đủ cuối cùng trong chuỗi thời gian kết hợp của các giá trị thực tế và giá trị dự báo. Thành phần theo mùa thể hiện độ lệch so với xu hướng và do đó thay đổi xung quanh 0 và tổng bằng 0 trong suốt một mùa.

Contribute season có xu hướng là không được xem xét và mùa vụ là 100%

Quality: OK có nghĩa là dự báo có khả năng có ít lỗi hơn

Trong nhiều trường hợp khác thì có GOOD có nghĩa là dự báo có ít sai sót hơn một nửa và POOR có nghĩa là dự báo có nhiều lỗi hơn.

Describe Forecast

SummaryModels

All forecasts were computed using exponential smoothing.

Sum of Total Profit

Level	Model	Season	RMSE	Quality Metrics				Smoothing Coefficients		
	Trend			MAE	MASE	MAPE	AIC	Alpha	Beta	Gamma
Multiplicative	None	Multiplicative	4,094,713	3,272,338	0.75	15.8%	2,771	0.020	0.000	0.143

Copy to Clipboard

[Learn more about the forecast models](#)

Close

Ảnh 35: Model describe Forecast

Theo ảnh 35 thì đường dự đoán này sử dụng mô hình nhân lên nhiều lần với 2 thành phần là Level và Season và không có xu hướng

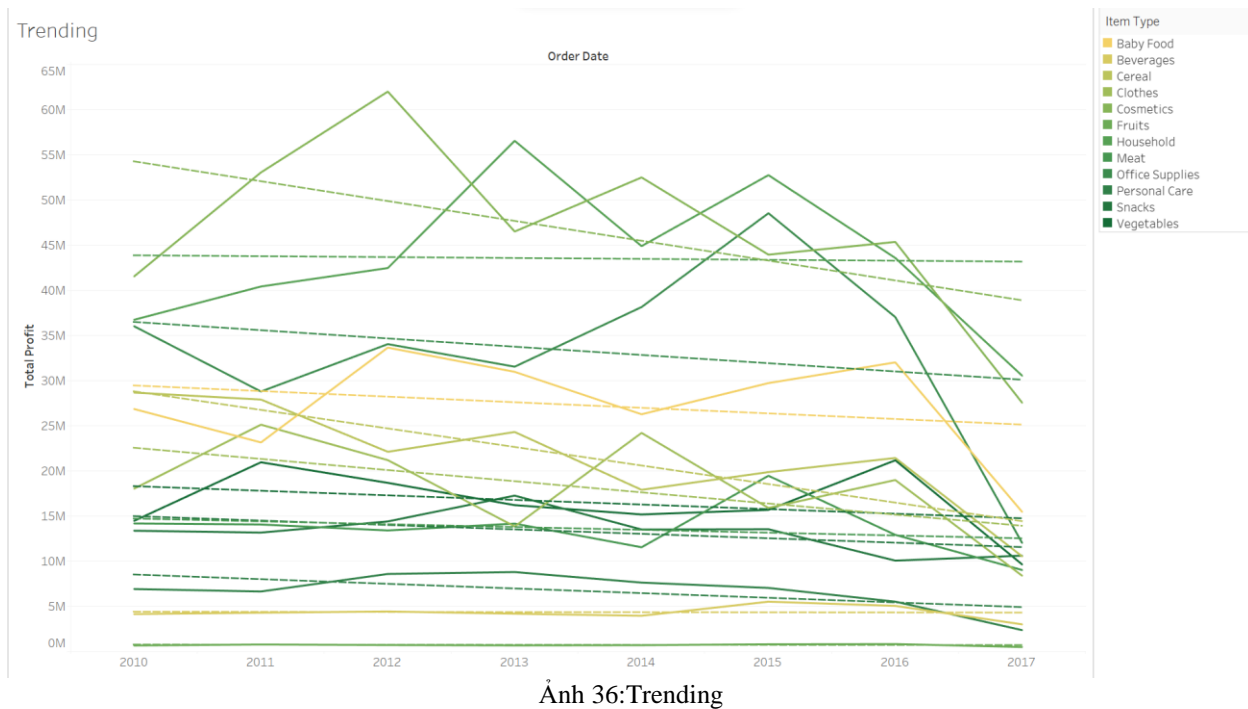
Ta thấy RMSE (Sai số trung bình bình phương gốc) và MAE (Sai số trung bình tuyệt đối) tương đối lớn tức là sai số càng lớn thì mức độ ước lượng cho thấy độ tin cậy của mô hình là thấp.

MASE (Sai số tỷ lệ tuyệt đối) là 0,75 có nghĩa là dự báo có khả năng có sai số hơn một nửa

MAPE (Sai số phần trăm tuyệt đối trung bình) là 15,8% tức là kết quả dự báo có thể sai số trung bình 15,8%

Ta thấy Alpha là hệ số làm trơn hàm mũ của biến level, Beta là hệ số làm trơn của Trend và Gamma là hệ số làm trơn của Season và hệ số hàm mũ càng ít gần 1 thì việc làm trơn càng được thực hiện nhiều

3.9) Trendline



Đây là một kỹ thuật phân tích phổ biến để ước tính xu hướng hoặc mối quan hệ giữa các biến liên tục.

Theo ảnh 36 đường xu hướng ở đây là đường Linear để xác định sự hiện diện hay thiếu mối quan hệ tuyến tính giữa các biến liên tục. Thể hiện mối tương quan giữa tổng lợi của các sản phẩm qua các năm.

Nhìn qua ta thấy đường xu hướng trong ảnh 36 là đường xu hướng tiêu cực (giảm) vì đường dốc của đường xu hướng là đường đi xuống

Describe Trend Model

Trend Lines Model

A linear trend model is computed for sum of Total Profit given Order Date Year. The model may be significant at $p \leq 0.05$. The factor Item Type may be significant at $p \leq 0.05$.

Model formula: Item Type*(Year of Order Date + Intercept)

Number of modeled observations: 96

Number of filtered observations: 0

Model degrees of freedom: 24

Residual degrees of freedom (DF): 72

SSE (sum squared error): 2.33582e+15

MSE (mean squared error): 3.2442e+13

R-Squared: 0.893488

Standard error: 5.69579e+06

p-value (significance): < 0.0001

Analysis of Variance:

Field	DF	SSE	MSE	F	p-value
Item Type	22	1.9313163e+16	8.77871e+14	27.0597	< 0.0001

Individual trend lines:

Panes	Row	Column	Color	Line	Coefficients	Value	StdErr	t-value	p-value	
			Item Type	p-value	DF	Term				
Total Profit	Year of Order Date	Vegetables		0.426013	6	Year of Order Date	-508309	595389	-0.853742	0.426013
						Intercept	1.03998e+09	1.19882e+09	0.867502	0.419005
Total Profit	Year of Order Date	Snacks		0.167672	6	Year of Order Date	-490126	312358	-1.56911	0.167672
						Intercept	1.00011e+09	6.28934e+08	1.59017	0.1629
Total Profit	Year of Order Date	Personal Care		0.102194	6	Year of Order Date	-515342	267360	-1.92753	0.102194
						Intercept	1.04433e+09	5.38329e+08	1.93995	0.100449
Total Profit	Year of Order Date	Office Supplies		0.607989	6	Year of Order Date	-913822	1.68912e+06	-0.541005	0.607989
						Intercept	1.87324e+09	3.40104e+09	0.550784	0.601671
Total Profit	Year of Order Date	Meat		0.53447	6	Year of Order Date	-312778	474765	-0.658807	0.53447
						Intercept	6.43373e+08	9.5594e+08	0.673026	0.525989
Total Profit	Year of Order Date	Household		0.944932	6	Year of Order Date	-99580.1	1.38281e+06	-0.0720127	0.944932
						Intercept	2.43993e+08	2.78429e+09	0.0876318	0.933021
Total Profit	Year of Order Date	Fruits		0.669644	6	Year of Order Date	-7739.57	17262.8	-0.448339	0.669644
						Intercept	1.62889e+07	3.47586e+07	0.468629	0.655877
Total Profit	Year of Order Date	Cosmetics		0.172744	6	Year of Order Date	-2.19735e+06	1.42007e+06	-1.54735	0.172744
						Intercept	4.47091e+09	2.85932e+09	1.56363	0.168937
Total Profit	Year of Order Date	Clothes		0.160947	6	Year of Order Date	-1.23389e+06	771685	-1.59895	0.160947
						Intercept	2.50264e+09	1.55379e+09	1.61067	0.158378
Total Profit	Year of Order Date	Cereal		0.0055116	6	Year of Order Date	-2.05342e+06	485659	-4.22812	0.0055116
						Intercept	4.15616e+09	9.77875e+08	4.2502	0.005379

Copy

Close

Ảnh 37: Describe trendlines model

Đầu ra của mô hình đường xu hướng tuyến tính cho thấy mối quan hệ rất mạnh. Chỉ số R-Squared bằng 89% có nghĩa là tác động của thời gian đến tổng lợi nhuận là rất lớn. Chỉ số p-value nhỏ hơn 0,0001 cho thấy đường xu hướng này có ý nghĩa thống kê

4) Kết luận

4.1) Nhận định chung

Qua quá trình phân tích RFM ta thấy chỉ có 26 khách hàng chiếm 0,52% trên tổng số 5000 khách hàng là “Best customer” và ngược lại thì “Churn Customer” chiếm đến 12,48%. Tổng doanh thu và lợi nhuận giảm đáng kể so với những năm đầu và không có sự khác biệt lớn giữa 2 kênh bán hàng Online và Offline. Tổng lợi nhuận của công ty phụ thuộc phần lớn vào 5 sản phẩm là: Cosmetics, Household, Officesupplies, Babyfood, Cereal. Tỷ lệ giao hàng trễ chiếm hơn 93% và có nảy sinh nhiều khoản chi phí bất thường ở tất cả các khu vực. Theo như 2 đường dự báo và xu hướng thì doanh thu và lợi nhuận sẽ giảm theo thời gian, điều này rất đáng lo ngại cho việc kinh doanh.

4.2) Suggestions

Một số cách giảm thiểu “Churn customer”:

- Tập trung vào khách hàng đang hài lòng về sản phẩm

Thay vì cố gắng khắc phục và níu giữ những khách hàng muốn rời đi, ta nên tập trung vào những khách hàng đang hài lòng về sản phẩm dịch vụ của mình. Họ sẽ là những người tạo ra khối lợi nhuận cao cũng có thể mang đến lượng khách hàng mới cho doanh nghiệp.

- Phân tích tỷ lệ Customer Churn ngay khi nó xảy ra

Phân tích cách thức và thời điểm khách hàng rời đi trong thời gian tồn tại của khách hàng với công ty và sử dụng dữ liệu đó để đưa ra các biện pháp ưu tiên.

- Cho các khách hàng thấy họ được doanh nghiệp quan tâm rất nhiều

Trao đổi với họ tất cả các đặc quyền công ty cung cấp và cho họ thấy được sự quan tâm từ công ty đến trải nghiệm của họ và họ chắc chắn sẽ tiếp tục.

➔ Một trải nghiệm khó chịu cho khách hàng có thể khiến họ một đi không trở lại, vì vậy luôn cung cấp và hỗ trợ khách hàng 24/7 để đem tới những kết quả khả quan nhất cho doanh nghiệp.

Nên đầu tư vào kênh bán hàng online nhiều hơn vì nó sẽ là nguồn thu nhập lớn trong tương lai. Nên áp dụng bán hàng trên các sàn thương mại điện tử, quảng bá sản phẩm bằng các trang mạng xã hội như: Instagram, Tiktok,... Đó là những trang mạng xã hội rất có tầm ảnh hưởng trên toàn thế giới. Thuê đại diện là các ca sĩ, KOL để tăng sự thu hút và độ tin cậy của khách hàng.

Hạn chế khả năng chậm trễ trong việc giao hàng, một số giải pháp giúp khắc phục tình trạng giao hàng trễ cho khách hàng:

- Sử dụng dịch vụ giao hàng của các công ty vận chuyển

Sử dụng dịch vụ của các công ty vận chuyển tốt nhất trên thế giới hiện nay như: DTDC (Ấn Độ), YRC (Mỹ), TNT (Châu Âu),...

- Đặt khoảng thời gian giao hàng

Đặt khoảng thời gian giao hàng tối thiểu để khách hàng biết được thời gian nhận hàng một cách rõ ràng. Khoảng thời gian đó phải thực tế theo nghĩa là bạn sẽ không đặt mình dưới áp lực phải giao hàng và không quá lâu làm nản lòng khách hàng.

- Lên kế hoạch trước

Nên lên kế hoạch trước về việc đảm bảo doanh nghiệp chuẩn bị sẵn sàng cho số lượng đơn hàng lớn hơn.

Các khoảng chi phí bất thường thường chủ yếu do vận chuyển có sự cố, vấn đề bảo hiểm, thuế hoặc cũng có thể do các nhân viên không đáng tin cậy.

Nên củng cố vấn đề vận chuyển hơn (có nhắc đến phía trên) và chỉ tuyển dụng các nhân viên đáng tin cậy.

-- HẾT --