

TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
Khoa Công nghệ thông tin

o0o



BÁO CÁO
ĐỒ ÁN CUỐI KỲ

Môn học	:	Trực quan hóa dữ liệu
Giảng viên phụ trách	:	TS. Bùi Tiến Lên
Nhóm sinh viên thực hiện	:	Nhóm 23
Các thành viên	:	Nguyễn Thiên An – 20120030 Nguyễn Thị Châu Ngọc – 20120146 Hồng Nhất Phương – 20120165 Nguyễn Nhật Tiến – 20120209 Trần Khánh Hoàng – 20120290

TP.Hồ Chí Minh - 06/2023

Mục lục

1	Tổng quan về bài làm.	2
2	Tổng quan về tập dữ liệu.	2
2.1	Thu thập dữ liệu	2
2.2	Khám phá và tiền xử lý dữ liệu	2
3	Trực quan hóa dữ liệu.	3
3.1	Phòng Kinh tế thị trường	3
3.2	Phòng Tài chính đầu tư	14
3.3	Phòng Marketing	30
3.4	Phòng Chăm sóc khách hàng	39
3.5	Việt Nam	45
4	Dự đoán doanh thu và lợi nhuận Superstore.	51
4.1	Lợi ích khi trả lời câu hỏi	51
4.2	Các bước thực hiện	51
4.3	Kết quả mô hình	51
4.4	Nhận xét	52
5	Kết luận.	53
6	Thông tin nhóm và phân công nhiệm vụ.	54
7	Dánh giá mức độ hoàn thành.	55
8	Tài liệu tham khảo.	56

1 Tổng quan về bài làm.

Công cụ trực quan hóa dữ liệu được sử dụng trong Đồ án là **Tableau**, cụ thể hơn là Tableau Desktop, phiên bản **2023.1**. Riêng với phần **4. Dự đoán doanh thu và lợi nhuận Superstore**, nhóm sử dụng **Python** và các thư viện **Python** để xây dựng mô hình dự đoán xu hướng. Bài làm của nhóm được chia thành các thư mục sau:

- Thư mục **products**: gồm *superstore.twbx* là file bài làm Tableau và *link_dashboard.txt* chứa đường liên kết đến dashboard chia sẻ qua Tableau Public.
- Thư mục **datasets**: bao gồm tập dữ liệu gốc *global_superstore_2016.xlsx* và tập dữ liệu sau khi tiền xử lý *global_superstore_clean.xlsx*.
- Thư mục **docs**: chứa file báo cáo *Report.pdf* và slide thuyết trình *Slide.pptx*.
- Thư mục **source_codes**: chứa các file notebook sau:
 - File *data_collection.ipynb*: nêu các thông tin cơ bản về tập dữ liệu superstore, kèm theo ảnh minh họa *superstore.png*.
 - File *data_preprocessing.ipynb*: tiền xử lý dữ liệu trước khi đưa vào Tableau.
 - File *forecasting.ipynb*: dự đoán doanh thu, lợi nhuận superstore.
 - File *linear_regression.ipynb*: tìm ra công thức thể hiện mối quan hệ giữa doanh thu và lợi nhuận của Superstore bằng hồi quy tuyến tính, phục vụ cho vẽ biểu đồ "What if" (phân tích kịch bản) của phòng Tài chính đầu tư.
 - File *association_rule.ipynb*: khai thác luật kết hợp phục vụ cho chiến lược quảng cáo, tiếp thị sản phẩm của phòng Marketing.

2 Tổng quan về tập dữ liệu.

2.1 Thu thập dữ liệu

Các thông tin về tập dữ liệu được trình bày trong file *data_collection.ipynb*. Tập dữ liệu *global_superstore_2016.xlsx* gồm 3 sheet là Orders, Returns và People. Nhận thấy sheet People không có nhiều ý nghĩa trong phân tích và trực quan hóa dữ liệu, nhóm quyết định chỉ sử dụng 2 sheet đầu tiên là Orders và Returns trong đồ án này.

2.2 Khám phá và tiền xử lý dữ liệu

Các thao tác tiền xử lý dữ liệu được trình bày trong file *data_preprocessing.ipynb*. Nhìn chung, tập dữ liệu khá sạch và nhóm chỉ tiến hành bỏ đi cột **Postal Code** vì thiếu dữ liệu quá nhiều; đồng thời chuyển các cột có kiểu **Object** về **string** để tiếp tục khám phá thêm các cột này.

3 Trực quan hóa dữ liệu.

Trong phần này, nhóm chúng em sẽ làm công việc của một DA (Data Analysis), với nhiệm vụ được giao là sử dụng tập dữ liệu Superstore để trực quan hóa theo nhu cầu của từng phòng ban trong công ty, từ đó đưa ra các đề xuất, chiến lược hành động cho mỗi phòng ban cụ thể, giúp công ty Superstore thu được nhiều lợi nhuận nhất. Có 4 phòng ban là: phòng Kinh tế - thị trường, phòng Tài chính - đầu tư, phòng Marketing và phòng Chăm sóc khách hàng. Chúng em nhận thấy rằng mỗi phòng ban có nhu cầu sử dụng thông tin khác nhau từ dữ liệu, và việc trực quan hóa sẽ giúp người xem hiểu rõ hơn về các mô hình, xu hướng và thị trường mà từng phòng ban đang tập trung vào. Người xem ở đây có thể là nhân viên, trưởng phòng hoặc giám đốc phòng ban đó. Bằng cách áp dụng các kỹ thuật trong Tableau, chúng em hy vọng sẽ mang lại cái nhìn toàn diện và dễ hiểu về hoạt động và hiệu suất của mỗi phòng ban trong công ty.Thêm vào đó, với vai trò là nhân viên, chúng em cũng nghiên cứu thêm về tầm ảnh hưởng của "công ty Superstore" tại thị trường Việt Nam, từ đó quyết định chiến lược và phát triển dài hạn của công ty trong thị trường Việt Nam.

Dashboard gồm 5 tabs tương ứng với 4 phòng ban và Việt Nam, có background thay đổi theo thời gian trong ngày (4 buổi sáng - trưa - chiều - tối). Đây là một điểm nhấn của dashboard nhóm muốn thực hiện để tạo hiệu ứng thị giác hấp dẫn và thu hút sự chú ý của người xem, đồng thời tăng tính tương tác và trải nghiệm của người dùng với dashboard.

3.1 Phòng Kinh tế thị trường

Nhu cầu: Phòng ban Kinh tế thị trường cần thông tin về cấu trúc và phân bố của thị trường, tình hình tiêu thụ, lợi nhuận, xu hướng.

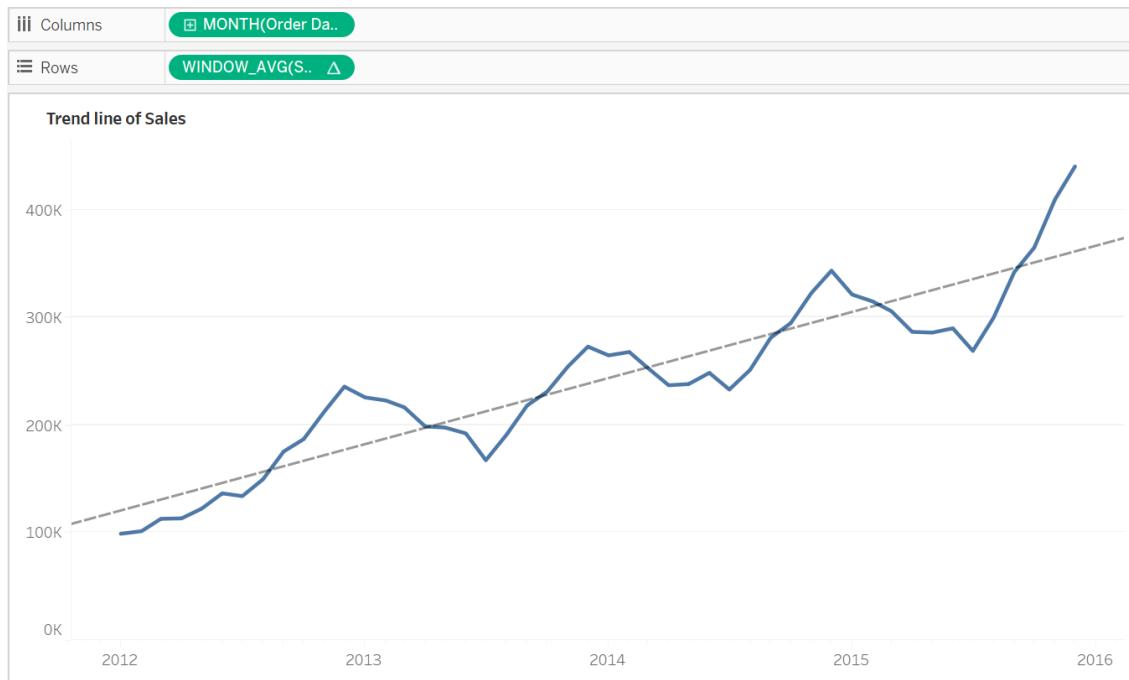
Tình huống: Công ty nhận được một khoản đầu tư, yêu cầu phòng Kinh tế - Thị trường xem xét và đưa ra ý kiến có nên đầu tư thêm hay không và nên đầu tư vào thị trường nào?

1. Phân tích xu thế của doanh số bán hàng: Chọn biểu đồ đường kết hợp đường xu hướng (trend line) cùng với phương pháp trung bình trượt

Biểu đồ line thường được sử dụng bởi vì nó liên kết các điểm theo thứ tự thời gian, dễ dàng thể hiện lợi nhuận qua các khoảng thời gian.

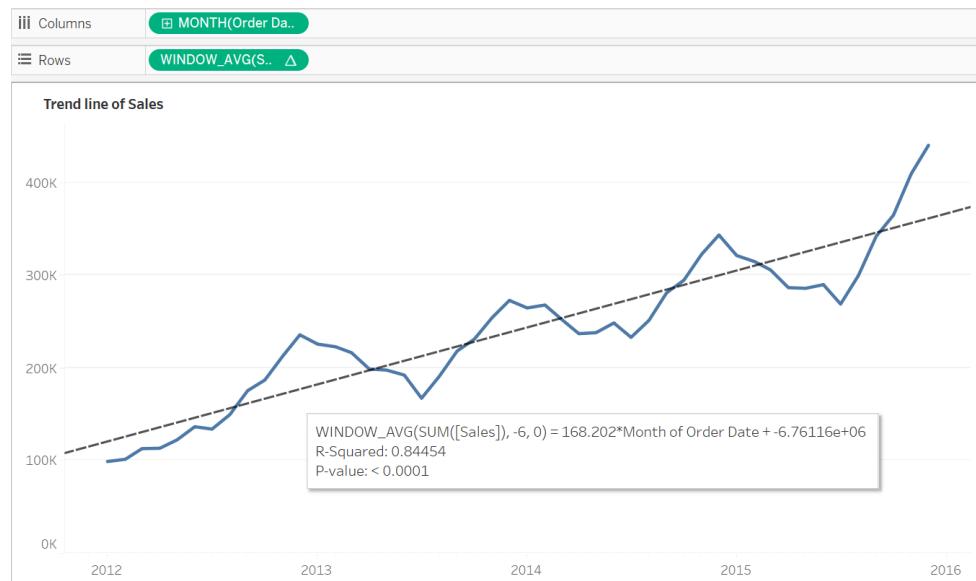
Trong Tableau có thể hiển thị các đường xu hướng (trend line), các đường này sẽ hiển thị hướng mà các giá trị dữ liệu di chuyển theo thời gian. Đường xu hướng cùng với các tham số thống kê của nó cho biết xu hướng quan sát được có ý nghĩa thống kê hay không.

Thay vì sử dụng doanh số bán hàng, chúng ta có thể sử dụng trung bình trượt (moving average) của doanh số bán hàng, điều này sẽ làm trơn chuỗi thời gian, dễ dàng nhìn thấy xu hướng.



Hình KTTT-1. Tạo biểu đồ xu hướng doanh thu

Nhận xét: đường xu hướng cho thấy tổng doanh số tăng lên theo thời gian. Trend line có R-square bằng 0.84 cho thấy có mối quan hệ chặt chẽ giữa đường trung bình trượt của doanh số và thời gian. Giá trị p-value rất thấp nên có thể tin tưởng vào mô hình.

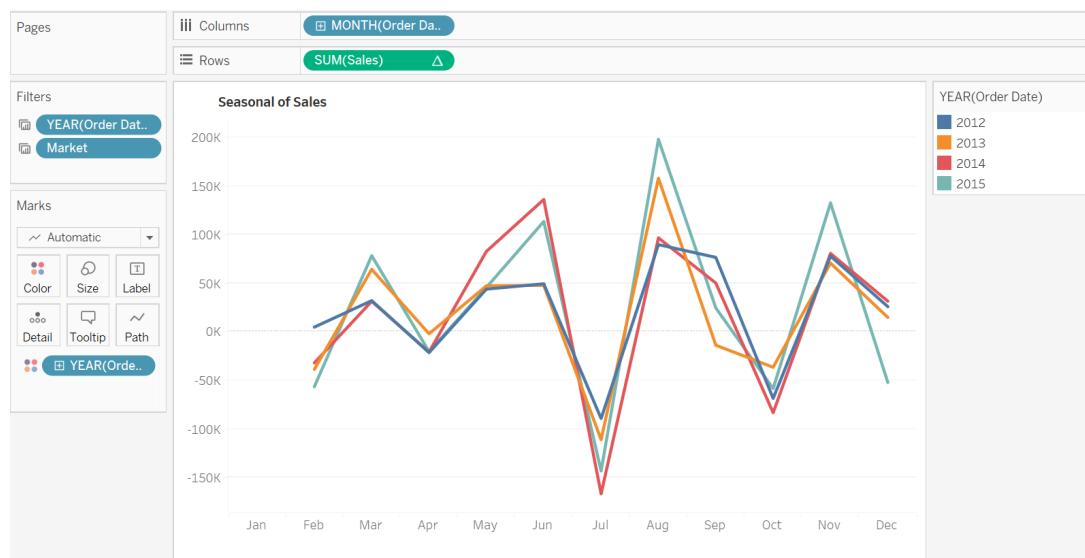


Hình KTTT-2. Đánh giá trend line

Từ các phân tích trên, có thể kết luận rằng có một xu hướng tích cực trong dữ liệu: Doanh số bán hàng tăng dần theo thời gian, thể hiện rằng Superstore trong những năm qua phát triển tốt, thái độ phục vụ ân cần, mặt hàng đa dạng nên nhận được sự chú ý và tin dùng của khách hàng.

2. Phân tích tính mùa vụ: sử dụng biểu đồ đường theo từng năm

Từ dữ liệu, ta nhận thấy tính mùa vụ trong việc bán hàng, có nghĩa là doanh số bán hàng đã được lặp lại qua các năm. Ta có thể quan sát tính mùa vụ một cách tốt hơn trong biểu đồ đường.



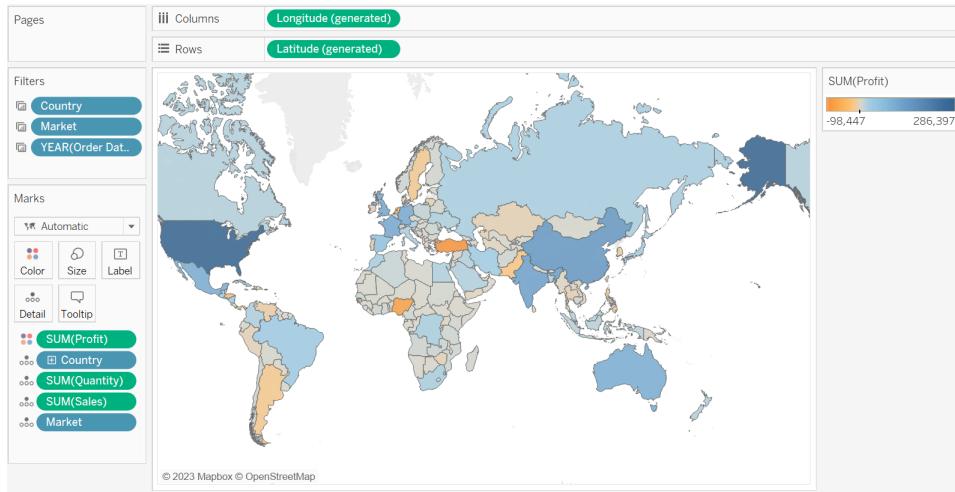
Hình KTTT-3. Biểu đồ mùa vụ

Nhận xét: Từ biểu đồ trên, có thể thấy rằng doanh số bán hàng đã tăng vọt trong tháng 3 so với mức thấp ở tháng 2. Doanh số bán hàng giảm đột biến vào tháng 4 và tháng 7, sau đó tăng mạnh vào tháng 8 và giảm trở lại vào tháng 10, rồi lại tăng vọt vào tháng 11. Đặc điểm này lặp lại mỗi năm, từ đó ta có thể kết luận có tính mùa vụ trong bán hàng ở Superstore.

Dựa vào phân tích trên, các nhà quản lý Superstore có thể điều chỉnh chiến lược kinh doanh, đẩy mạnh quảng cáo, khuyến mãi, nhanh chóng tìm nguồn hàng chất lượng và kịp thời.

3. Phân tích lợi nhuận: sử dụng biểu đồ map

Biểu đồ map thường được sử dụng để thể hiện dữ liệu trên bản đồ, cho phép chúng ta có cái nhìn tổng quan về sự phân bố địa lý của dữ liệu. Trong trường hợp này, chúng ta sử dụng biểu đồ map để thể hiện sự phân bố địa lý của lợi nhuận trong Superstore theo từng nước trên thế giới.



Hình KTTT-4. Biểu đồ map

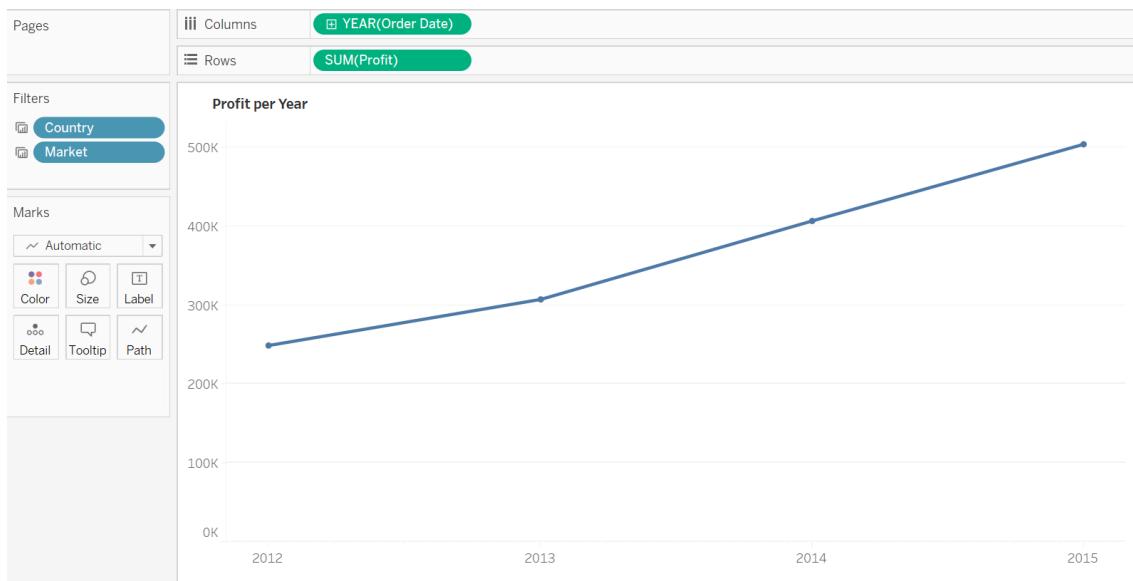
Nhận xét: Dựa vào biểu đồ và sử dụng các tương tác với map, ta thấy tổng lợi nhuận giai đoạn 2014-2017 ở United State thu được là lớn nhất vì United State cũng là nước có tổng số đơn hàng và tổng doanh thu cao nhất trong giai đoạn này. Tiếp đó là China có tổng lợi nhuận nhiều thứ 2. Những vùng có màu cam như Turkey, Nigeria là những khu vực tạo ra lợi nhuận âm.

Từ việc phân tích trên, các nhà quản lý có thể kịp thời xem xét hiệu quả kinh doanh tại những bang có lợi nhuận thấp như thái độ phục vụ, chăm sóc khách hàng, chất lượng hàng hóa, cách thức bảo quản hàng hóa, cách trưng bày hàng lên kệ,... đồng thời đẩy mạnh, phát triển kinh doanh cho các bang lợi nhuận cao.

4. Phân tích lợi nhuận theo năm: Sử dụng biểu đồ đường

Biểu đồ lợi nhuận theo từng năm chỉ ra sự biến động của lợi nhuận qua từng năm, từ đó giúp các nhà đầu tư có cái nhìn tổng quát khi đưa ra quyết định đầu tư

Sử dụng biểu đồ đường với thuộc tính **Order date** theo Year và **Profit** để có được biểu đồ thể hiện lợi nhuận qua các năm

**Hình KTTT-5. Tạo biểu đồ lợi nhuận**

Nhận xét: Biểu đồ lợi nhuận hàng năm cho thấy tổng lợi nhuận đã tăng dần theo các năm, khẳng định sự hiện diện của xu hướng dài hạn. Dựa vào biểu đồ trên, ta thấy SuperStore đã kinh doanh tốt và có hiệu quả

5. Giải quyết tình huống

Để giải quyết tình huống, cần xem xét lợi nhuận của từng thị trường trong những năm vừa qua cũng như tìm hiểu ưu điểm và nhược điểm ảnh hưởng đến lợi nhuận của từng thị trường.

(a) Khu vực USCA

USCA gồm 2 nước United State và Canada. Đây là thị trường có lợi nhuận tăng qua từng năm và các nước đều có lợi nhuận dương, United State là nước tạo ra lợi nhuận cao nhất trên thế giới

Thuận lợi: Tiếp cận nguồn khách hàng lớn và đa dạng. Cơ sở hạ tầng tài chính phát triển, tạo thuận lợi cho các giao dịch kinh doanh. Thị trường cạnh tranh cao, có thể thúc đẩy đổi mới và tạo cơ hội phát triển.

Khó khăn: Mức độ cạnh tranh cao, có thể gây khó khăn cho các doanh nghiệp mới tham gia thị trường. Thuế và chi phí lao động cao, có thể ảnh hưởng đến lợi nhuận



Hình KTTT-6a và KTTT-6b. Khu vực USCA

Kết luận: Dựa vào lợi nhuận của USCA tăng qua từng năm, tổng lợi nhuận các năm dương cùng với những ưu thế của khu vực, công ty có thể đầu tư vào USCA

(b) Khu vực Europe

Khu vực Europe có nước Anh mang lại lợi nhuận cao nhất trong khu vực, đa số các nước đều có lợi nhuận dương. Một số ít nước như Hà Lan, Thụy Điển mang lại lợi nhuận âm

Thuận lợi: Sức mua mạnh và thu nhập khả dụng. Tiếp cận với lao động có trình độ và tay nghề cao. Tiếp cận các nhà đầu tư và thị trường toàn cầu.

Khó khăn: Mức độ cạnh tranh cao, có thể gây khó khăn cho các doanh nghiệp mới tham gia thị trường. Thuế và chi phí lao động cao, có thể ảnh hưởng đến lợi nhuận



Hình KTTT-7a và KTTT-7b. Khu vực Europe

Kết luận: Có thể cân nhắc đầu tư, nhưng chỉ đầu tư vào những nước có lợi nhuận tăng trưởng hằng năm

(c) Khu vực Asia Pacific

Khu vực Asia Pacific có Turkey là nước có lợi nhuận thấp nhất thế giới nhưng bên cạnh đó cũng có China là nước có lợi nhuận cao thứ 2 thế giới. Số lượng các nước có lợi nhuận dương (tông màu xanh) và lợi nhuận âm (tông màu cam) không chênh lệch nhiều.

Thuận lợi: Thị trường rộng lớn, đa dạng với nhu cầu hàng hóa và dịch vụ ngày càng cao. Các chính sách ưu đãi của chính phủ đối với các nhà đầu tư nước ngoài. Có sẵn chi phí lao động và chi phí sản xuất thấp.

Khó khăn: Thị trường phức tạp và đa dạng đòi hỏi phải nghiên cứu và thích ứng thị trường sâu rộng. Hiểu biết hạn chế về văn hóa và ngôn ngữ địa phương. Cơ sở hạ tầng kém phát triển ở một số khu vực. Cạnh tranh từ các doanh nghiệp trong nước và các tập đoàn đa quốc gia lâu đời.



Hình KTTT-8a và KTTT-8b. Khu vực Asia Pacific

Kết luận: Có thể cân nhắc đầu tư, nhưng chỉ đầu tư vào những nước có lợi nhuận tăng trưởng hằng năm.

(d) Khu vực Africa

Khu vực Africa có Nigeria có lợi nhuận thấp thứ hai thế giới, các nước khác đều có lợi nhuận khá thấp so với các khu vực khác.

Thuận lợi: Đây là thị trường chưa được khai thác với dân số ngày càng tăng, mang đến cơ hội cho các doanh nghiệp muốn mở rộng. Nhiều quốc gia có mức độ cạnh tranh thấp trong các ngành và lĩnh vực khác nhau, tạo cơ hội cho các doanh nhân thành lập doanh nghiệp mà không phải đối mặt với sự cạnh tranh cao. Nhiều quốc gia có nguồn tài nguyên thiên nhiên phong phú như khoáng sản, dầu mỏ và khí đốt giúp tiếp cận nguồn nguyên liệu thô giá cả phải chăng. **Khó khăn:** Cơ sở hạ tầng yếu kém: Nhiều quốc gia châu Phi có cơ sở hạ tầng yếu kém có thể hạn chế sự di chuyển của hàng hóa và con người, cản trở tăng trưởng kinh doanh. Khí hậu khắc nghiệt ảnh hưởng đến việc bảo quản và lưu thông hàng hóa.



Hình KTTT-9a và KTTT-9b. Khu vực Africa

Kết luận: Có thể cân nhắc đầu tư vì dựa vào biểu đồ ta thấy lợi nhuận những năm sau tăng vọt nhưng vẫn chỉ đầu tư vào những nước có lợi nhuận tăng trưởng hằng năm.

(e) Khu vực LATAM

Khu vực LATAM có nước Mexico mang lại lợi nhuận cao nhất trong khu vực. Số lượng các nước có lợi nhuận dương (tông màu xanh) và lợi nhuận âm (tông màu cam) không chênh lệch nhiều.

Thuận lợi: Đây là một thị trường mới nổi với tiềm năng tăng trưởng đáng kể. Có nguồn tài nguyên thiên nhiên dồi dào có thể khai thác để phát triển kinh tế. Sự đa dạng văn hóa của khu vực tạo cơ hội cho các doanh nghiệp tạo ra các sản phẩm và dịch vụ sáng tạo. Vị trí chiến lược của thị trường LATAM khiến nó trở thành điểm đến hấp dẫn cho các doanh nghiệp muốn mở rộng sang các thị trường mới. Chi phí lao động tại thị trường LATAM thấp hơn so với các thị trường phát triển khác như Châu Âu và Bắc Mỹ.

Khó khăn: Thị trường LATAM bao gồm các quốc gia có nhiều ngôn ngữ khác nhau, khiến các doanh nghiệp gặp khó khăn trong việc giao tiếp hiệu quả. Nhiều quốc gia trong thị trường LATAM thiếu cơ sở hạ tầng đáng tin cậy và hiệu quả, điều này có thể làm tăng chi phí kinh doanh.



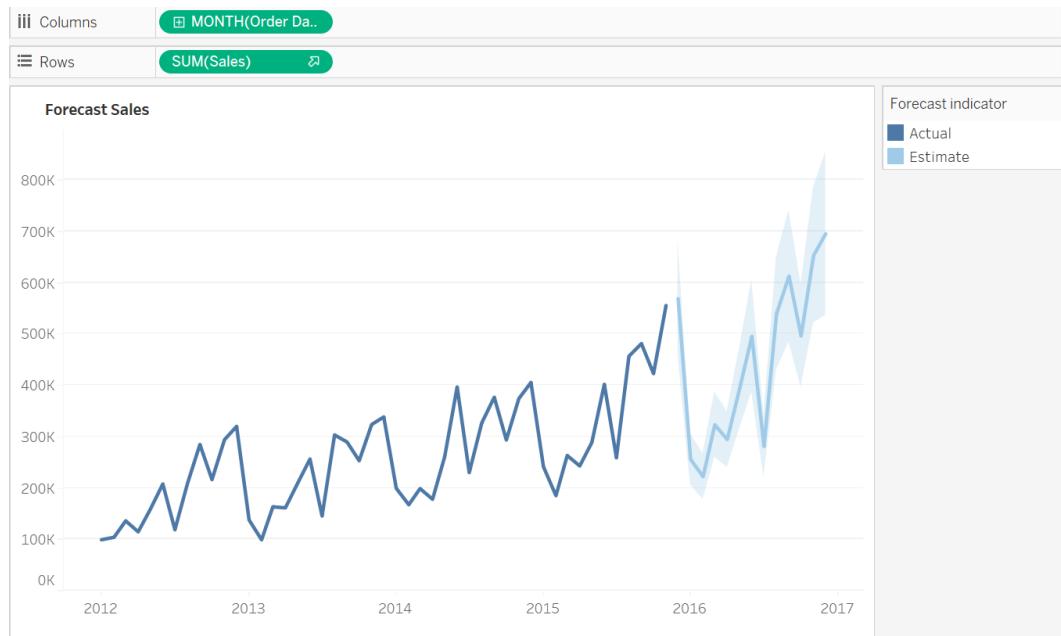
Hình KTTT-10a và KTTT-10b. Khu vực LATAM

Kết luận: Có thể cân nhắc đầu tư vào những nơi có lợi nhuận tăng hằng năm và có ưu thế về vị trí chiến lược

6. Dự báo doanh thu và lợi nhuận: sử dụng biểu đồ đường kết hợp với forecasting

Phòng Kinh tế - Thị trường đưa ra biểu đồ dự báo doanh thu và lợi nhuận.

Việc dự báo trước doanh thu có rất nhiều lợi ích cho các đối tượng khác nhau: doanh nghiệp có thể điều chỉnh các chiến lược bán hàng để tối đa hóa doanh thu, phát hiện ra các khuynh hướng tiêu dùng và hành vi mua hàng của khách hàng, từ đó đưa ra các quyết định kinh doanh phù hợp để thu hút và giữ chân khách hàng; các nhà quản lý có thể hiểu rõ hơn về sản phẩm và dịch vụ được ưa chuộng nhất để tối ưu hóa sản xuất và phân phối, cải thiện hiệu quả quản lý đơn hàng và quản lý kho.



Hình KTTT-11a. Biểu đồ dự báo doanh thu

Nhận xét: Từ hình trên có thể thấy rằng dự báo cũng có tính mùa vụ và nó tuân theo quy luật bán hàng chung trong những năm qua. Xem xét hiệu suất mô hình của mô hình dự báo chuỗi thời gian ta thấy: mô hình dự báo được xây dựng dựa trên doanh số hàng tháng trong giai đoạn từ tháng 1 năm 2012 đến tháng 11 năm 2015, mô hình dự báo đã được tính toán bằng cách sử dụng làm trơn hàm mũ. Mô hình dựa trên Level, Season và xu hướng. Trong mô hình này có MAPE là khoảng 7.3%, điều này có nghĩa là giá trị dự báo có thể chênh lệch trung bình 7.3% so với giá trị thực tế. Nhờ vào biểu đồ dự báo này, các nhà quản lý Superstore đề xuất các chiến lược kinh doanh, nắm bắt thị hiếu của khách hàng.



Hình KTTT-12a. Thông tin dự báo doanh thu

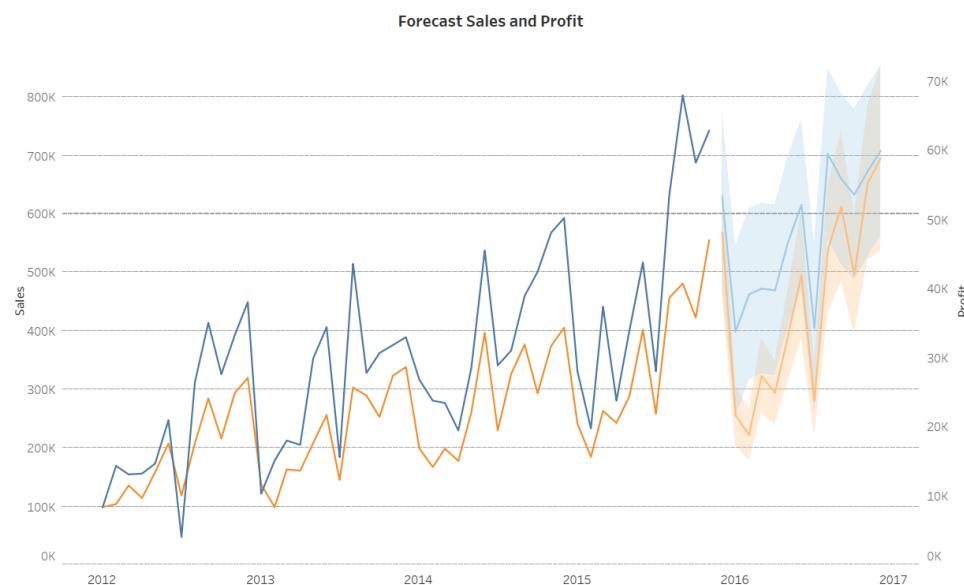


All forecasts were computed using exponential smoothing.

Sum of Sales

Level	Model			Quality Metrics					Smoothing Coefficients		
	Trend	Season		RMSE	MAE	MASE	MAPE	AIC	Alpha	Beta	Gamma
Multiplicative	Additive	Multiplicative		24,987	18,724	0.26	7.3%	986	0.041	0.364	0.000

Hình KTTT-13a. Thông tin dự báo doanh thu



Hình KTTT-11b. Thêm vào đường dự báo lợi nhuận



Hình KTTT-12b. Thông tin dự báo lợi nhuận



Hình KTTT-13b. Thông tin dự báo lợi nhuận

Nhận xét:

- Sự biến động của đường lợi nhuận có tương quan mật thiết với đường doanh thu.
- Khi sử dụng forecast để dự đoán lợi nhuận ta vẫn thấy sự tương quan trong biến đổi của lợi nhuận với doanh thu vẫn được thể hiện rõ.
- Theo dự đoán khi sử dụng forecast, có sự giảm mạnh trong lợi nhuận từ tháng December 2015 đến February 2016.
- Từ February 16 đến December 2016 lợi nhuận tuy có sự nhiều sự biến động nhưng nhìn chung vẫn có xu hướng tăng.

7. Doanh thu và lợi nhuận qua các khu vực cụ thể: sử dụng bubble chart

Hình KTTT-14. Biểu đồ Doanh thu và lợi nhuận qua các khu vực

Nhận xét:

- Các thị trường nổi bật mang lại nhiều doanh thu và lợi nhuận bao gồm: Western Europe, Oceania, Central America, Eastern Asia, Southern Asia.

- Western Europe là khu vực mang lại nhiều doanh thu (1.731.930\$) và lợi nhuận nhất (218,431\$).
- Western Asia, Western Africa, Central Asia là 3 khu vực có lợi nhuận âm, trong đó Western Asia thua lỗ nhiều nhất (-53,922\$).
- Riêng Southeastern Asia đây là khu mang lại doanh thu đáng kể, đứng thứ tư về doanh thu giữa các khu vực nhưng lợi nhuận mang lại rất thấp (17,852\$).

3.2 Phòng Tài chính đầu tư

Nhu cầu: Phòng ban Tài chính - Đầu tư cần thông tin về dự báo doanh thu, lợi nhuận của Superstore, cần đánh giá hiệu quả đầu tư, nên tập trung đầu tư vào sản phẩm nào, xác định chiến lược đầu tư và kế hoạch tài chính phù hợp.

1. Phân tích doanh thu - lợi nhuận dựa trên chỉ số YOY: Chọn biểu đồ đường kết hợp sử dụng chỉ số YOY (Year Over Year)

Lý do lựa chọn biểu đồ:

- Biểu đồ đường thể hiện tốt xu hướng và biến động của tốc độ tăng trưởng các quý so với cùng kỳ năm trước. Nó giúp ta nhìn thấy sự thay đổi liên tục và phân tích các xu hướng dễ dàng hơn so với các loại biểu đồ khác.

- Biểu đồ đường cũng có thể được sử dụng để so sánh và phân tích mối quan hệ tốc độ tăng trưởng của 3 Category trong cùng một khoảng thời gian. Bên cạnh đó, sử dụng màu sắc hoặc các đường kẻ khác nhau để phân biệt 3 Category.

* Cách tính chỉ số YOY đối với doanh thu: $(\text{Doanh thu năm hiện tại} - \text{Doanh thu năm trước}) / \text{Doanh thu năm trước} * 100$

* Cách tính chỉ số YOY đối với lợi nhuận: $(\text{Lợi nhuận năm hiện tại} - \text{Lợi nhuận năm trước}) / \text{Lợi nhuận năm trước} * 100$

Cách thực hiện đối với doanh thu:

- Kéo thả các thuộc tính Order date, Category, Sales (Order date được kéo thả 2 lần: lần 1 - lấy Year; lần 2 - lấy Quarter để phân Quý trong năm).

- Nhấp chuột vào Sum(Sales), di chuyển xuống Quick Table Calculation rồi nhấn chọn Year Over Year Growth.

- Chọn biểu đồ đường để trực quan

Cách thực hiện đối với lợi nhuận: quy trình thực hiện như phần doanh thu

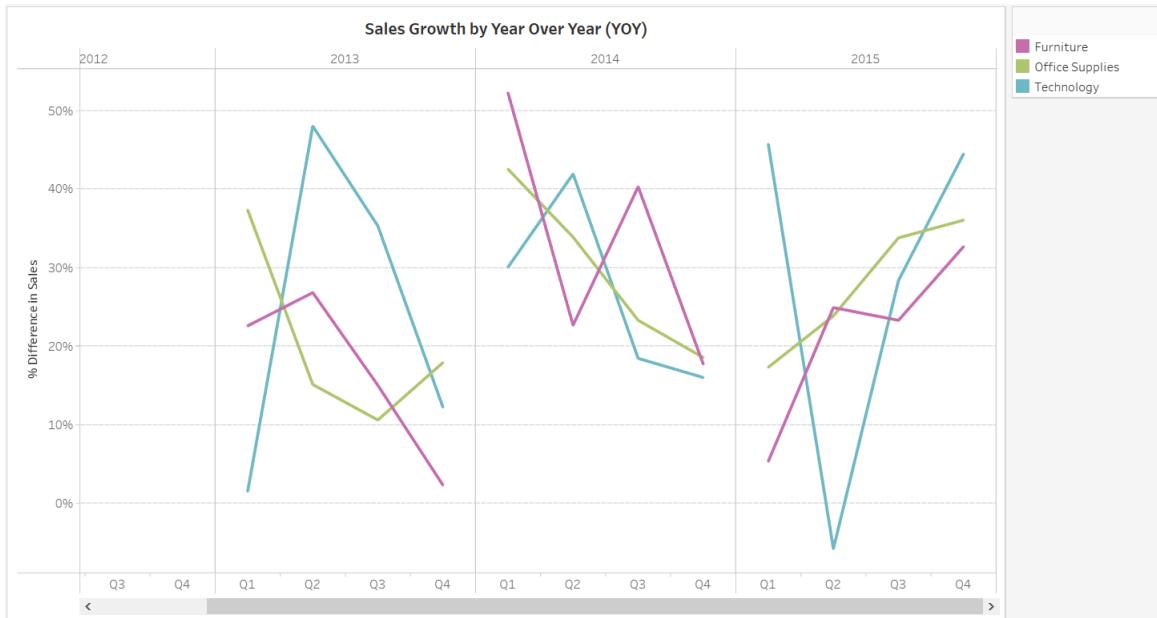
Trực quan:



Hình TCDT-1. Biểu đồ trực quan tốc độ tăng trưởng doanh thu – lợi nhuận bằng chỉ số YOY

Nhận xét:

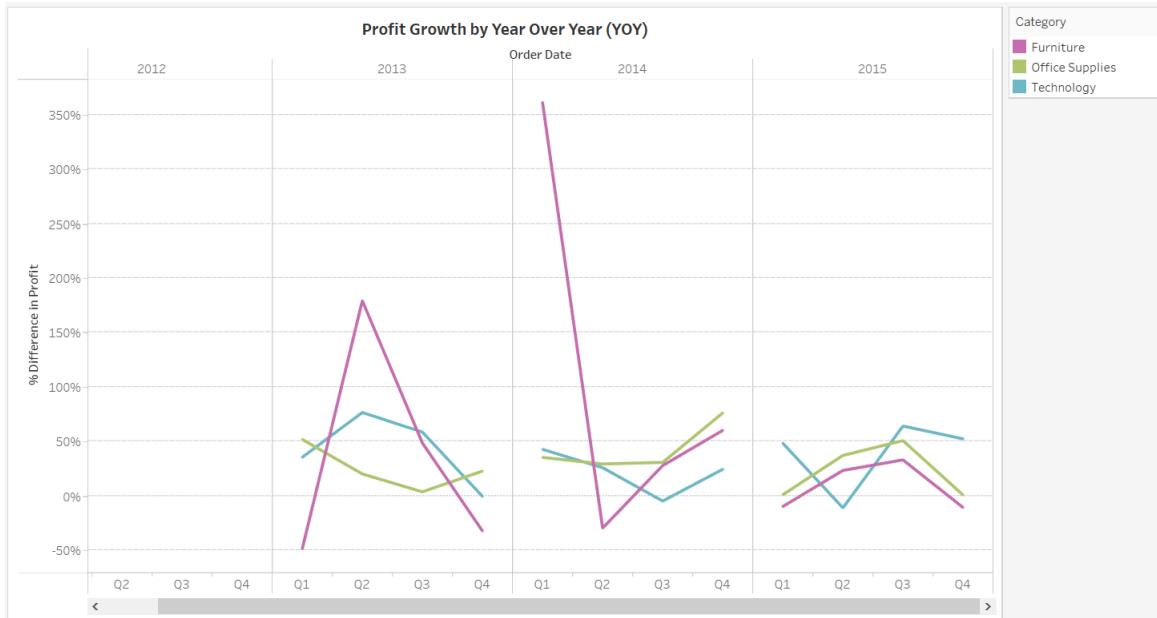
- Dựa vào biểu đồ về tăng trưởng doanh thu - lợi nhuận với chỉ số YOY Growth cần xem xét các câu hỏi sau để vận dụng tốt cho từng biểu đồ:
 - + Qua biểu đồ thể hiện chỉ số YOY, theo cùng kỳ 1 Quý năm trước so với năm hiện tại thì tốc độ tăng trưởng doanh thu theo từng Category tốt như thế nào? Tăng trưởng âm hay dương?
 - + Đánh giá được quá trình phát triển từng Category của SuperStore trong từng giai đoạn như thế nào? Tiềm năng và định hướng phát triển của Category đó trong tương lai?
 - + Dựa vào chỉ số YOY có thể đánh giá được hiệu quả kinh doanh của SuperStore trong quá trình kinh doanh theo quý, năm và đưa ra những quyết định phù hợp.
- Phân tích cụ thể trong giai đoạn 2012 - 2015:



Hình TCDT-2. Biểu đồ trực quan tốc độ tăng trưởng doanh thu bằng chỉ số YOY

+ Về doanh thu: dấu hiệu đáng mừng là doanh thu trong giai đoạn này ở tất cả các Category đều tăng trưởng dương, duy chỉ có category Technology vào Q2/2015 giảm 5.82% so với cùng kỳ năm trước

- Xét Category Technology ở Quý 2/2013 có doanh thu tăng trưởng cao nhất 47.99% so với cùng kỳ năm 2012.
- Xét Category Furniture Quý 1/2014 có doanh thu tăng vượt trội nhất 52.23% so với cùng kỳ năm 2013.
- Xét Category Office Supplies Quý 3/2014 có doanh thu tăng vượt trội nhất 42.51% so với cùng kỳ năm 2013.



Hình TCDT-3. Biểu đồ trực quan tốc độ tăng trưởng lợi nhuận bằng chỉ số YOY

+ Về lợi nhuận:

- Xét Category Technology ở Quý 2/2013 có lợi nhuận tăng nhanh 76.2% so với cùng kỳ năm 2012; còn ở Quý 2/2015 có lợi nhuận tăng trưởng âm -11.3% so với cùng kỳ năm 2014.
 - Xét Category Furniture ở Quý 1/2014 có lợi nhuận tăng mạnh nhất 360.8% so với cùng kỳ năm 2013; còn ở Quý 1/2013 có lợi nhuận tăng trưởng âm -48.9% so với cùng kỳ năm 2012.
 - Xét Category Office Supplies có tốc độ tăng trưởng lợi nhuận khá bền vững không có biến động nhiều và không có quý nào tăng trưởng âm; có Quý 4/2014 có lợi nhuận tăng vượt trội nhất 75.8% so với cùng kỳ năm 2013.
- + Dựa vào chỉ số tăng trưởng về doanh thu và lợi nhuận trong từng giai đoạn Quý qua các năm SuperStore cần đánh giá được tính hình kinh doanh của từng nhóm Category để có chính sách đầu tư kinh doanh phù hợp cho từng Category ở mỗi Quý.

2. Phân tích tỷ suất lợi nhuận theo từng SubCategory: biểu đồ cột ngang kết hợp phân nhóm lợi nhuận bằng việc sử dụng chỉ số tỷ suất lợi nhuận

Lý do lựa chọn biểu đồ:

- Biểu đồ cột ngang được sử dụng để trực quan hóa dữ liệu dạng số, tỷ lệ, cho phép so sánh mức độ chênh lệch tỷ suất lợi nhuận giữa các SubCategory bằng cách sử dụng chiều ngang của cột để đại diện cho giá trị.

- Hơn nữa tỷ suất lợi nhuận có giá trị âm vì vậy thể hiện trên trục ngang Ox giúp người đọc dễ dàng nhận biết giá trị âm dương.
- Giúp tiết kiệm không gian, diện tích, dữ liệu được trực quan rõ ràng, không chồng đe lên nhau.

Cách thực hiện:

- Tạo Calculated Field là Profit Ratio (tỷ suất lợi nhuận) để tính toán cho từng SubCategory.
- Profit Ratio cho mỗi SubCategory được tính bằng cách: Tổng lợi nhuận/Tổng doanh thu * 100 (%).



Hình TCĐT-4. Công thức tính Calculated Field “Profit Ratio”

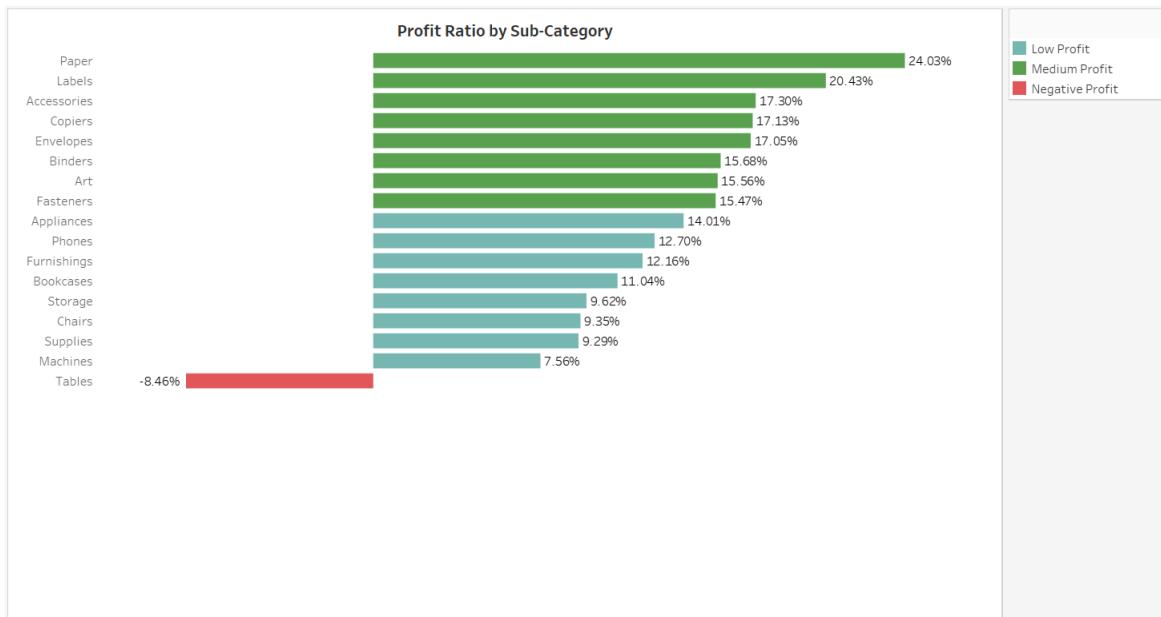
- Tạo Calculated Field là Profit Level (phân nhóm lợi nhuận) cho từng SubCategory. Phân loại thành 4 nhóm tỷ suất lợi nhuận (đơn vị: %):
 - + High Profit: có tỷ suất lợi nhuận ($\geq 30\%$)
 - + Medium Profit: có tỷ suất lợi nhuận ($15\% \leq x < 30\%$)
 - + Low Profit: có tỷ suất lợi nhuận ($0\% \leq x < 15\%$)
 - + Negative Profit: có tỷ suất lợi nhuận ($< 0\%$)



Hình TCDT-5. Công thức phân nhóm lợi nhuận cho Calculated Field “Profit Level”

- Tiếp theo kéo thả các thuộc tính Sub category, Profit Ratio, Profit Level và chọn biểu đồ cột ngang để trực quan hóa tỷ suất lợi nhuận.

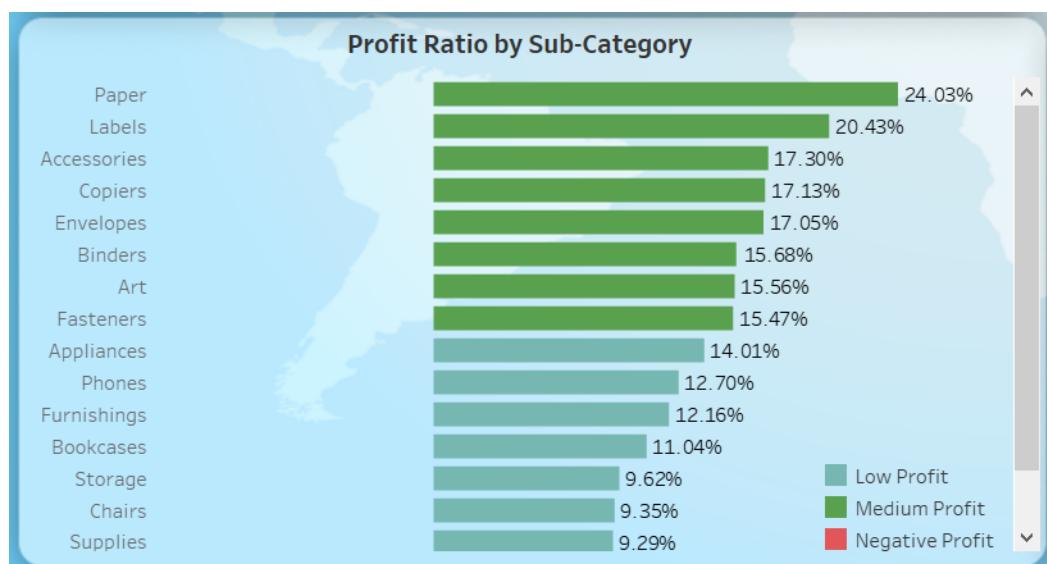
Trực quan:



Hình TCDT-6. Biểu đồ trực quan tỷ suất lợi nhuận của các SubCategory

Nhận xét:

- Không có SubCategory nằm trong nhóm “High Profit” mang lại tỷ suất lợi nhuận cao cho SuperStore cần tập trung đầu tư phát triển SubCategory đó mạnh hơn tuy nhiên cần chú trọng đến phát triển bền vững hơn là phát triển nhanh một cách ồ ạt.
- SubCategory (Paper, Labels, Accessories, Copiers, Envelopes, Binders, Art, Fasteners) nằm trong nhóm “Median Profit” mang lại tỷ suất lợi nhuận trung bình cho SuperStore cũng cần tập trung đầu tư phát triển.
- SubCategory (Appliances Furnishings, Phones, Storage, Chairs, Machines, Supplies, Bookcases) nằm trong nhóm “Low Profit” mang lại tỷ suất lợi nhuận thấp cho SuperStore nếu muốn phát triển sản phẩm này thì cần xem xét, nghiên cứu thị trường một cách kĩ lưỡng hơn, cần đưa ra thêm những chiến lược sale, marketing phù hợp để thúc đẩy doanh thu lợi nhuận của nhóm subcategory này tốt hơn.
- SubCategory (Tables) nằm trong nhóm “Negative Profit” mang lại lợi nhuận âm cho SuperStore. Cần nghiên cứu nguyên nhân vì sao những SubCategory này lại mang về lợi nhuận âm qua đó có sự điều chỉnh cho phù hợp với thực tế (giảm kinh doanh những subcategory này).



Hình TCDT-7. Biểu đồ trực quan tỷ suất lợi nhuận của các SubCategory

3. Phân tích kịch bản doanh thu - lợi nhuận theo từng Quý: Chọn biểu đồ đường kết hợp việc sử dụng chỉ số tăng trưởng doanh thu

Lý do lựa chọn biểu đồ:

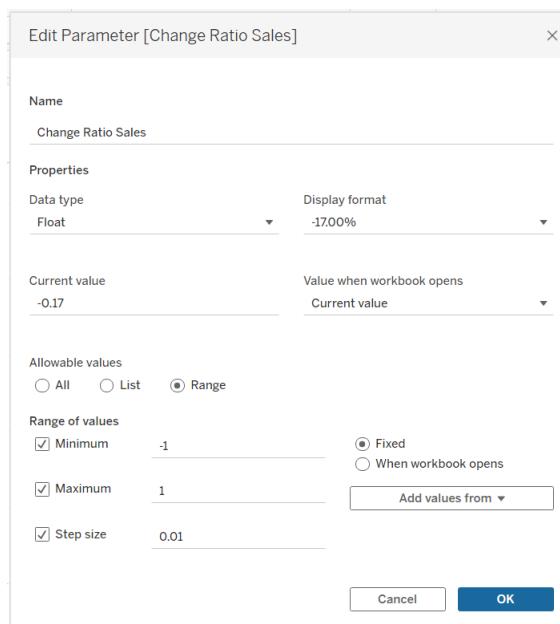
- Biểu đồ đường thể hiện tốt xu hướng và biến động của doanh theo thời gian. Nó giúp ta nhìn thấy sự thay đổi liên tục và phân tích các xu hướng dễ dàng hơn so với các loại biểu đồ khác.

- Biểu đồ đường cũng có thể được sử dụng để so sánh và phân tích mối quan hệ giữa Sales – Change Sales và Profit – Change Profit theo thời gian để so sánh sự thay đổi giữa Sales – Change Sales và Profit – Change Profit trong cùng một khoảng thời gian. Bên cạnh đó, sử dụng màu sắc hoặc các đường kẻ khác nhau để phân biệt giữa Sales – Change Sales và Profit – Change Profit.

- Dễ dàng đọc giá trị của biến tại bất kỳ thời điểm nào.

Cách thực hiện:

- Tạo Parameter là Change Ratio Sales (tỷ lệ tăng/giảm doanh thu a %).



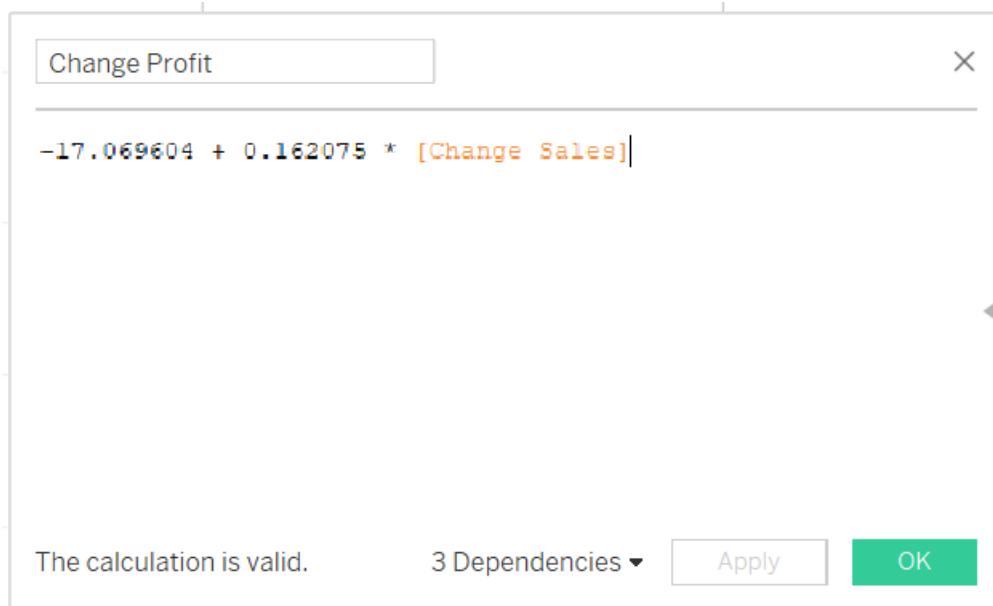
Hình TCDT-8. Cách tạo Parameter “Change Ratio Sales”

- Tạo Calculated Field là Change Sales (doanh thu sau khi tăng/giảm a %) theo từng Quý.
- Change Sales cho từng Quý được tính bằng cách: Sales * (1 + Change Ratio Sales)



Hình TCDT-9. Công thức tạo Calculated Field “Change Sales”

- Change Profit cho từng Quý được tính bằng cách: $-17.069604 + 0.162075 * \text{Change Sales}$

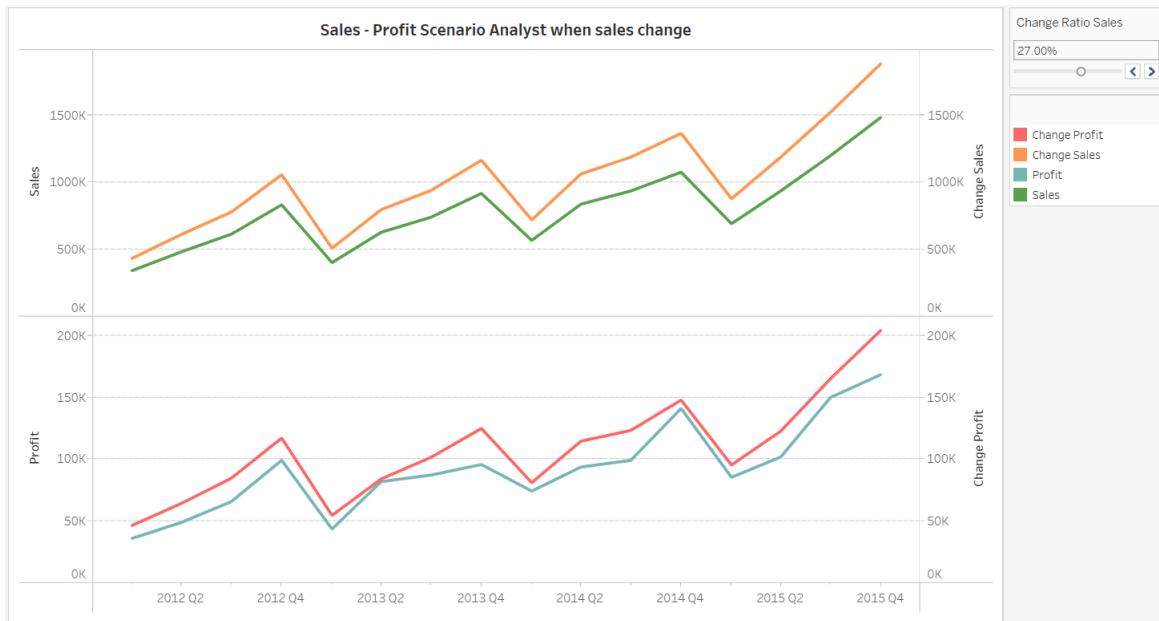


Hình TCDT-10. Công thức tạo Calculated Field “Change Profit”

- Tiếp theo kéo thả các thuộc tính Order Date, Sales, Change Sales, Profit, Change Profit và chọn biểu đồ đường để trực quan hóa doanh thu thực tế và lợi nhuận kì vọng.

- Đồng bộ hóa 2 biểu đồ Sales – Change Sales và Profit – Change Profit

Trực quan:

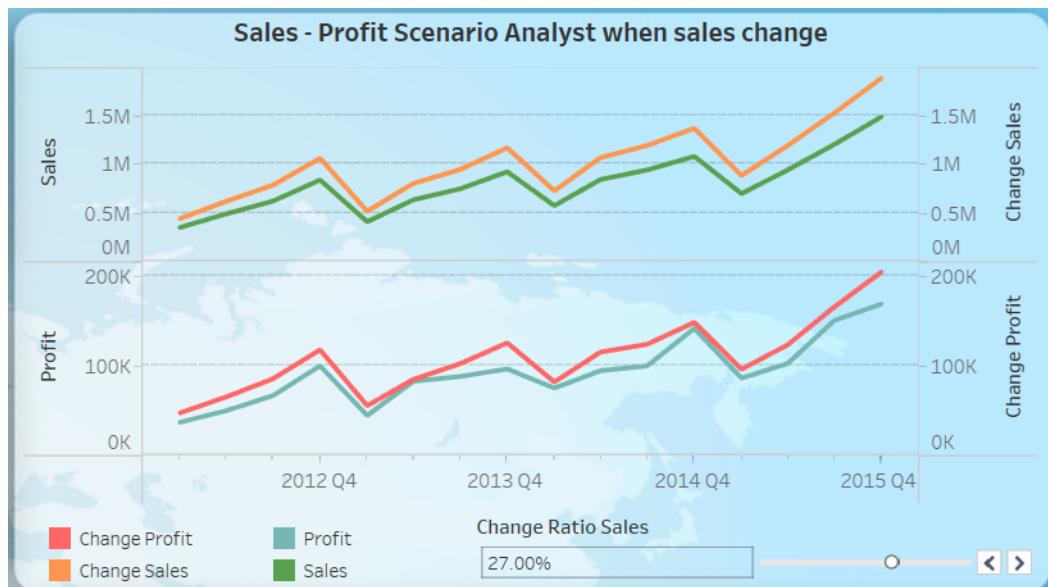


Hình TCDT-11. Biểu đồ trực quan Phân tích kịch bản doanh thu lợi nhuận khi tăng doanh thu

Nhận xét:

- Nếu đường Change Sales nằm trên đường Sales thì tổng doanh thu kỳ vọng đang tăng. Nếu đường Change Sales và đường Sales nằm càng xa nhau thì doanh thu thực tế và doanh thu kỳ vọng biến động khá mạnh và tăng nhanh. Cần chuẩn bị kho bãi, lượng hàng nhiều để kịp cung ứng cho khách hàng. Cần xem xét đây có phải là sự tăng trưởng bền vững không?
- Nếu đường Change Profit nằm trên đường Profit thì tổng lợi nhuận kỳ vọng đang tăng. Nếu đường Change Profit và đường Profit nằm càng xa nhau thì lợi nhuận thực tế và lợi nhuận kỳ vọng biến động khá mạnh và đang phát triển tốt. Điều này cho thấy việc tăng doanh thu như kịch bản này đang có lợi về cả doanh thu và lợi nhuận và nên phát triển theo hướng đi này.
- Nếu đường Change Sales nằm dưới đường Sales thì tổng doanh thu kỳ vọng đang giảm. Nếu đường Change Sales và đường Sales nằm càng xa nhau thì doanh thu thực tế và doanh thu kỳ vọng khác biệt và đang giảm mạnh ở mức nghiêm trọng. Cần phải chuẩn bị những kịch bản để ứng phó một cách tốt nhất với những vấn đề sụt giảm doanh thu. Cần có cái nhìn chính xác, những đánh giá chuyên sâu để có kế hoạch chuẩn bị tốt cho sự đối phó của SuperStore nếu gặp tình trạng thực tế này.

- Nếu đường Change Profit nằm dưới đường Profit thì tổng lợi nhuận kỳ vọng đang giảm. Nếu đường Change Profit và đường Profit nằm càng xa nhau thì lợi nhuận thực tế và lợi nhuận kỳ vọng biến động khá mạnh và lợi nhuận đang sụt giảm nghiêm trọng. Điều này cho thấy việc thay đổi doanh thu như kịch bản này đang có vấn đề về lợi nhuận cần xem xét kĩ lưỡng.



Hình TCDT-12. Biểu đồ trực quan Phân tích kịch bản doanh thu lợi nhuận khi tăng doanh thu trong dashboard

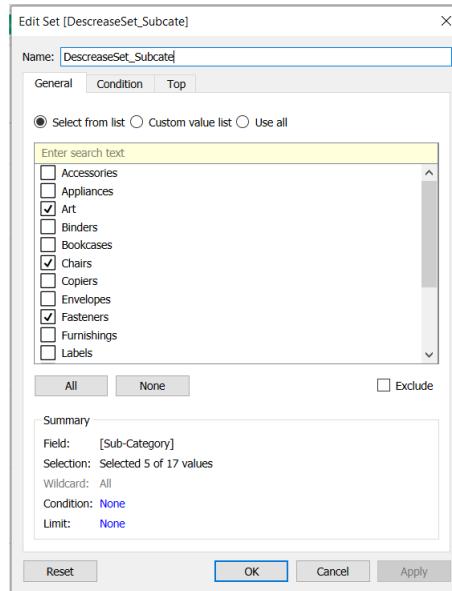
4. Phân tích kịch bản lợi nhuận theo từng Quý phụ thuộc vào sự tăng/giảm lợi nhuận của một số Subcategory: Chọn biểu đồ đường kết hợp việc sử dụng chỉ số tăng/giảm lợi nhuận một số SubCategory

Lý do lựa chọn biểu đồ:

- Biểu đồ đường thể hiện tốt xu hướng và biến động của doanh theo thời gian. Nó giúp ta nhìn thấy sự thay đổi liên tục và phân tích các xu hướng dễ dàng hơn so với các loại biểu đồ khác.
- Biểu đồ đường cũng có thể được sử dụng để so sánh và phân tích mối quan hệ giữa Profit và Scenario Profit theo thời gian để so sánh sự thay đổi giữa Profit và Scenario Profit trong cùng một khoảng thời gian. Bên cạnh đó, sử dụng màu sắc hoặc các đường kẻ khác nhau để phân biệt giữa Profit và Scenario Profit.
- Dễ dàng đọc giá trị của biến tại bất kỳ thời điểm nào.

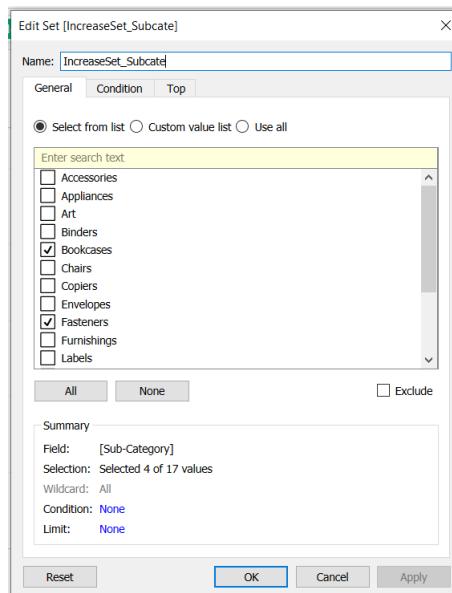
Cách thực hiện:

- Tạo set DescreaseSet_Subcate là danh sách những SubCategory giảm lợi nhuận



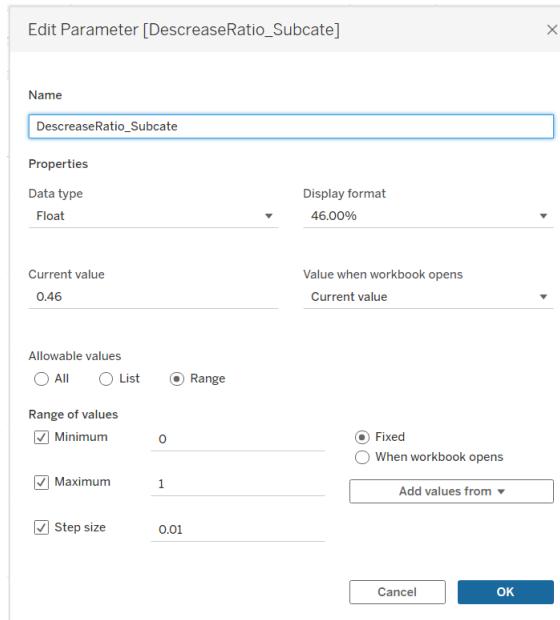
Hình TCDT-13. Cách tạo Set “DescreaseSet _ Subcate”

- Tạo set IncreaseSet _ Subcate là danh sách những SubCategory tăng lợi nhuận.



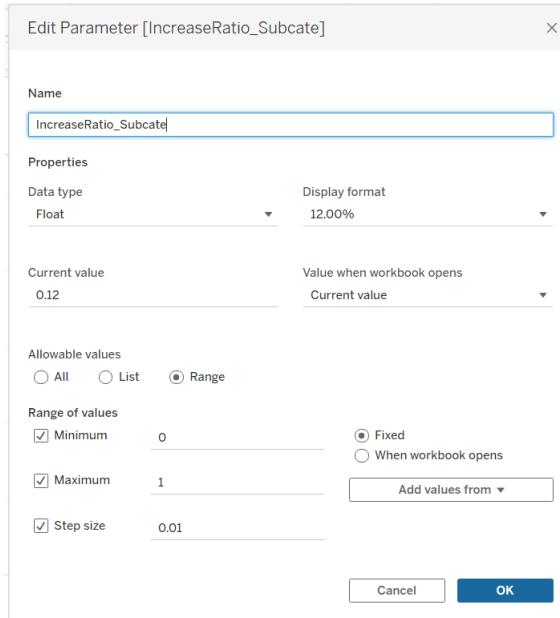
Hình TCDT-14. Cách tạo Set “IncreaseSet _ Subcate”

- Tạo Parameter là DescreaseRatio _ Subcate (Những subCategory giảm lợi nhuận với tỷ lệ a%); parameter này có thể cho người dùng chọn những subCategory sẽ giảm và thay đổi tỉ lệ giảm.



Hình TCDT-15. Cách tạo Paramater “DescreaseRatio _ Subcate”

- Tạo Parameter là IncreaseRatio_Subcate (Những subCategory tăng lợi nhuận với tỷ lệ b%); parameter này có thể cho người dùng chọn những subCategory sẽ tăng và thay đổi tỉ lệ tăng.



Hình TCDT-16. Cách tạo Paramater “IncreaseRatio _ Subcate”

- Tạo Calculated Field là Scenario Profit (lợi nhuận sau khi điều chỉnh tăng giảm một số Subcategory một tỷ lệ nhất định).

- Đoạn lệnh tính toán Scenario Profit:

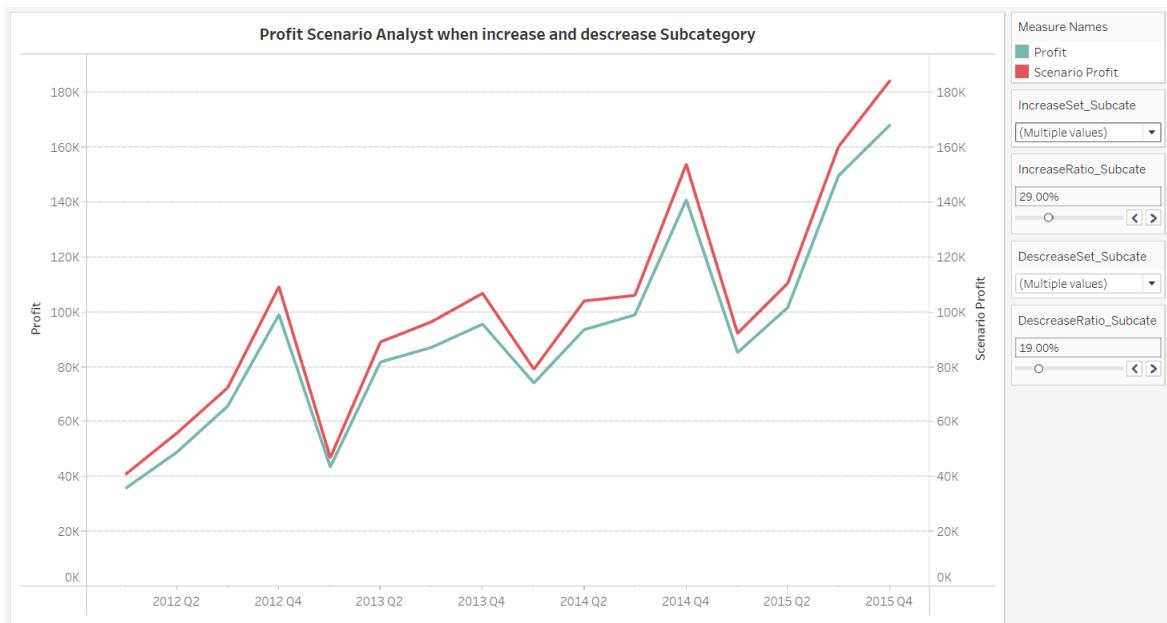
```
IF [IncreaseSet_Subcate] THEN [Profit]*(1+[IncreaseRatio_Subcate])
ELSEIF [DescreaseSet_Subcate] THEN [Profit]*(1-[DescreaseRatio_Subcate])
ELSE [Profit]
END
```



Hình TCDT-17. Công thức tạo Calculated Field “Scenario Profit”

- Tiếp theo kéo thả các thuộc tính Order date, Profit, Scenario Profit và hiển thị các parameter, các set được tạo và chọn biểu đồ đường để trực quan hóa lợi nhuận thực tế và lợi nhuận sau khi điều chỉnh tăng giảm một số Subcategory, tính tổng Profit và Scenario Profit theo từng Quý (Lưu ý dùng "dual axis" và đồng bộ 2 trục).

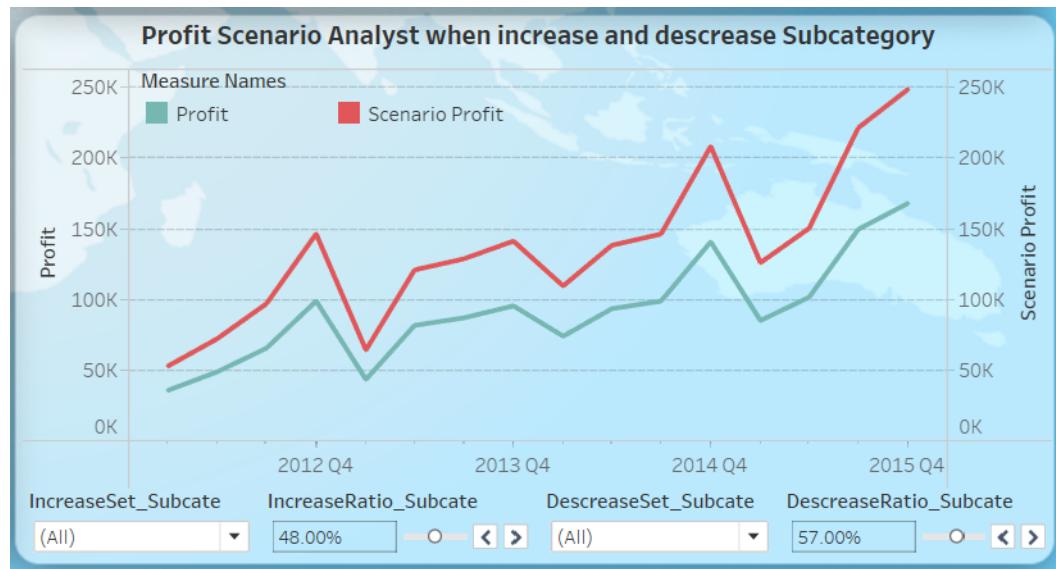
Trực quan:



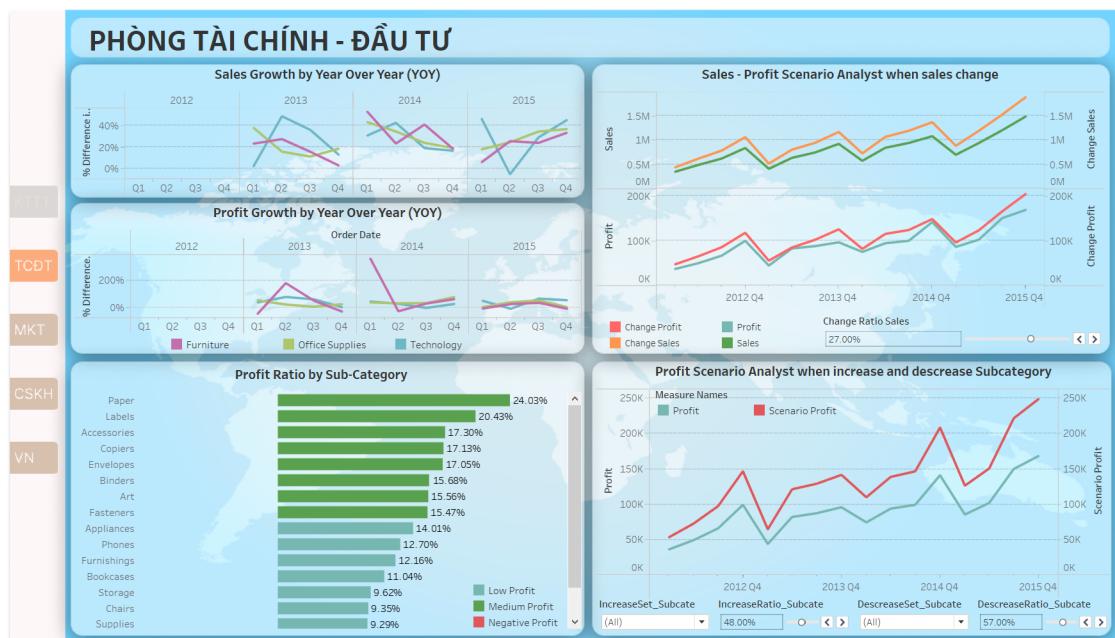
Hình TCDT-18. Biểu đồ trực quan Phân tích kịch bản lợi nhuận một số Subcate sẽ tăng/giảm

Nhận xét:

- Nếu đường Scenario Profit nằm trên đường Profit thì ta việc tăng/giảm lợi nhuận của một số subCategory theo định hướng này đang có lợi nhất cho SuperStore. Cần điều chỉnh hướng kinh doanh theo kịch bản này, tập trung đầu tư vào những subCategory làm tăng lợi nhuận, cần giảm những subCategory nào để tránh tạo lợi nhuận âm cho SuperStore.
- Nếu đường Scenario Profit nằm dưới đường Profit thì ta việc tăng/giảm lợi nhuận của một số subCategory theo định hướng này không tốt cho lợi nhuận của SuperStore. Cần tránh kinh doanh – đầu tư theo kịch bản này, cần tránh những subCategory ảnh hưởng xấu đến lợi nhuận, cần chuẩn bị sẵn những phương án để đối phó nếu gặp kịch bản này trong thực tế.



Hình TCDT-19. Biểu đồ trực quan Phân tích kịch bản lợi nhuận một số Subcate sẽ tăng/giảm trong dashboard



Hình TCDT-20. Dashboard Phòng tài chính đầu tư

Đề xuất chiến lược để cải thiện kinh doanh:

- Sát sao trong việc đánh giá tình hình kinh doanh của từng nhóm danh mục để có chính sách đầu tư kinh doanh phù hợp với từng danh mục ở mỗi quý, mỗi năm.
- Tập trung đầu tư, phát triển bền vững và tăng cường quảng bá cho các Subcategory thuộc nhóm “High Profit” và “Median Profit”.

- Kết hợp với các đội nhóm khác như bộ phận marketing để tạo ra các chiến lược quảng cáo mang tính sáng tạo giúp tăng doanh số và lợi nhuận trong các mặt hàng mang lợi nhuận cao và trung bình.
- Đối với các Subcategory nằm trong nhóm “Low Profit” thì cần xem xét, nghiên cứu thị trường một cách kĩ lưỡng hơn, đưa ra thêm các chiến lược quảng bá và tiếp thị phù hợp để giới thiệu và tăng khả năng tiếp cận của khách hàng đến các sản phẩm này.
- Với nhóm “Negative Profit” cần nghiên cứu nguyên nhân vì sao những Subcategory lại mang lại lợi nhuận âm và điều chỉnh phù hợp với thực tế như giảm kinh doanh những Subcategory này.
- Thông qua việc phân tích kịch bản doanh thu theo từng Quý: Tùy theo từng biểu hiện của đường SaleChange và đường Sales trên biểu đồ mà có hai cách ứng phó khác nhau như:
 - Cần chuẩn bị kho bãi, và tăng cường lượng hàng để đáp ứng nhu cầu cung ứng cho khách hàng, tuy nhiên bên cạnh đó cần xem xét kĩ đây có phải là sự phát triển có bền vững không.
 - Cần chuẩn bị kịch bản để ứng phó một cách tốt nhất với những vấn đề trong việc sụt giảm doanh thu. Không những thế cần có những đánh giá chuyên sâu để lên kế hoạch chuẩn bị tốt để đối phó khi gặp vấn đề trong thực tế.
- Thông qua việc phân tích kịch bản lợi nhuận theo từng Quý: Tùy theo từng biểu hiện của đường ScenarioProfit và đường Profit thì có hai cách ứng phó như sau:
 - Tập trung điều chỉnh hướng kinh doanh theo kịch bản hiện tại, đầu tư vào các Subcategory làm tăng lợi nhuận và giảm những Subcategory mang lại lợi nhuận âm cho các Superstore nếu việc tăng/ giảm lợi nhuận của một số Subcategory theo định hướng hiện tại đang có lợi nhất cho Superstore.
 - Nếu việc tăng/giảm lợi nhuận của một số Subcategory theo định hướng hiện tại không tốt cho lợi nhuận của các Superstore thì cần tránh kinh doanh, đầu tư theo kịch bản này và cũng tránh những Subcategory đang ảnh hưởng xấu đến lợi nhuận. Bên cạnh đó, cần chuẩn bị sẵn những phương án để đối phó nếu gặp kịch bản này trong thực tế.

3.3 Phòng Marketing

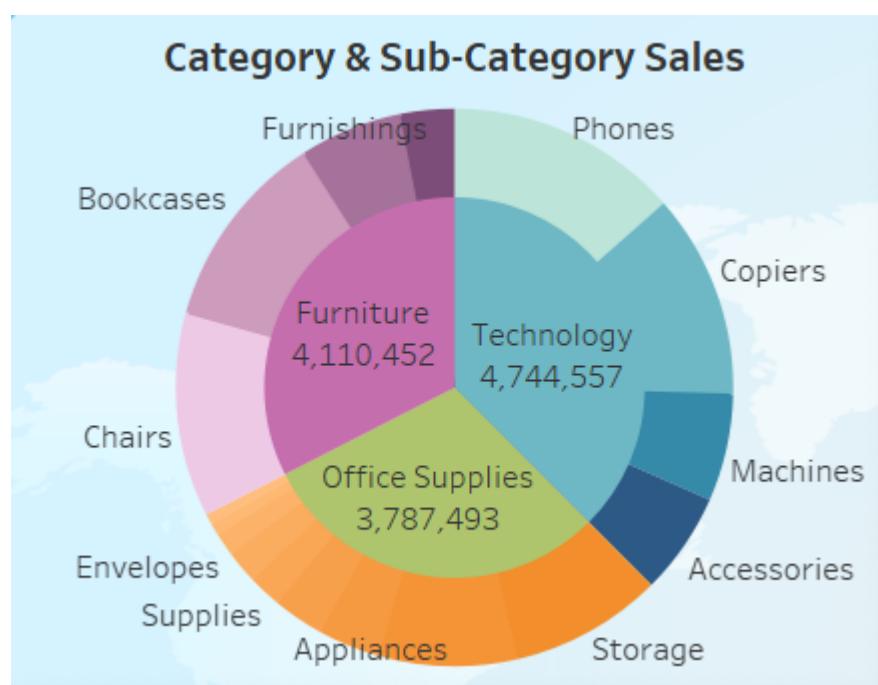
Nhu cầu: Phòng ban Marketing cần thông tin về biểu đồ thể hiện hiệu quả của chiến dịch quảng cáo, chương trình khuyến mãi, mặt hàng nào đem lại lợi nhuận cao nhất để tập trung marketing.

1. **Tìm hiểu doanh số các Category và Sub Category của các superstore. (Sunburst Chart)**

Lý do lựa chọn biểu đồ:

- Sunburst Diagram được thiết kế để trực quan hóa các dữ liệu có tính chất phân cấp, nó cho phép ta hiển thị dữ liệu với mức độ chi tiết khác nhau từ toàn bộ tới mức con nhỏ nhất. Và khi sử dụng hình tròn để biểu diễn các category và sub-category, ta có thể dễ dàng theo dõi tỷ lệ phần trăm và mối quan hệ giữa các category và sub-category một cách trực quan và dễ hiểu
- Trong biểu đồ dưới đây, khi dùng Sunburst Diagram để biểu diễn dữ liệu phân cấp, ta thường vẽ các category là các lát cắt hình tròn nằm ở trong và các sub-category của chúng được đặt nằm bên ngoài hình tròn này. Điều này rất hữu ích trong việc biểu diễn dữ liệu phân cấp, vì nó cho phép ta mã hóa các category và các sub-category của nó theo tỷ lệ phần trăm tương ứng với kích thước của lát cắt hình tròn. Và khi biểu diễn các sub-category bên ngoài hình tròn của category cha, ta có thể thấy rõ hơn tỷ lệ phần trăm của chúng so với category cha và giữ cho cấu trúc dữ liệu trở nên rõ ràng hơn.

Trực quan:



Hình MKT-1. Sunburst Chart

Nhận xét:

- Doanh thu của các sản phẩm đến từ danh mục Technology chiếm phần lớn nhất trong ba danh mục, tiếp đến là Furniture và Office Supplies.
- Trong Technology, Phones là danh mục con mang lại phần lớn doanh thu(1.706.824\$).

- Đối với Furniture là Chairs(1.501.682\$) và Office Supplies là Storage(1.126.813\$).

Đề xuất chiến lược để cải thiện kinh doanh:

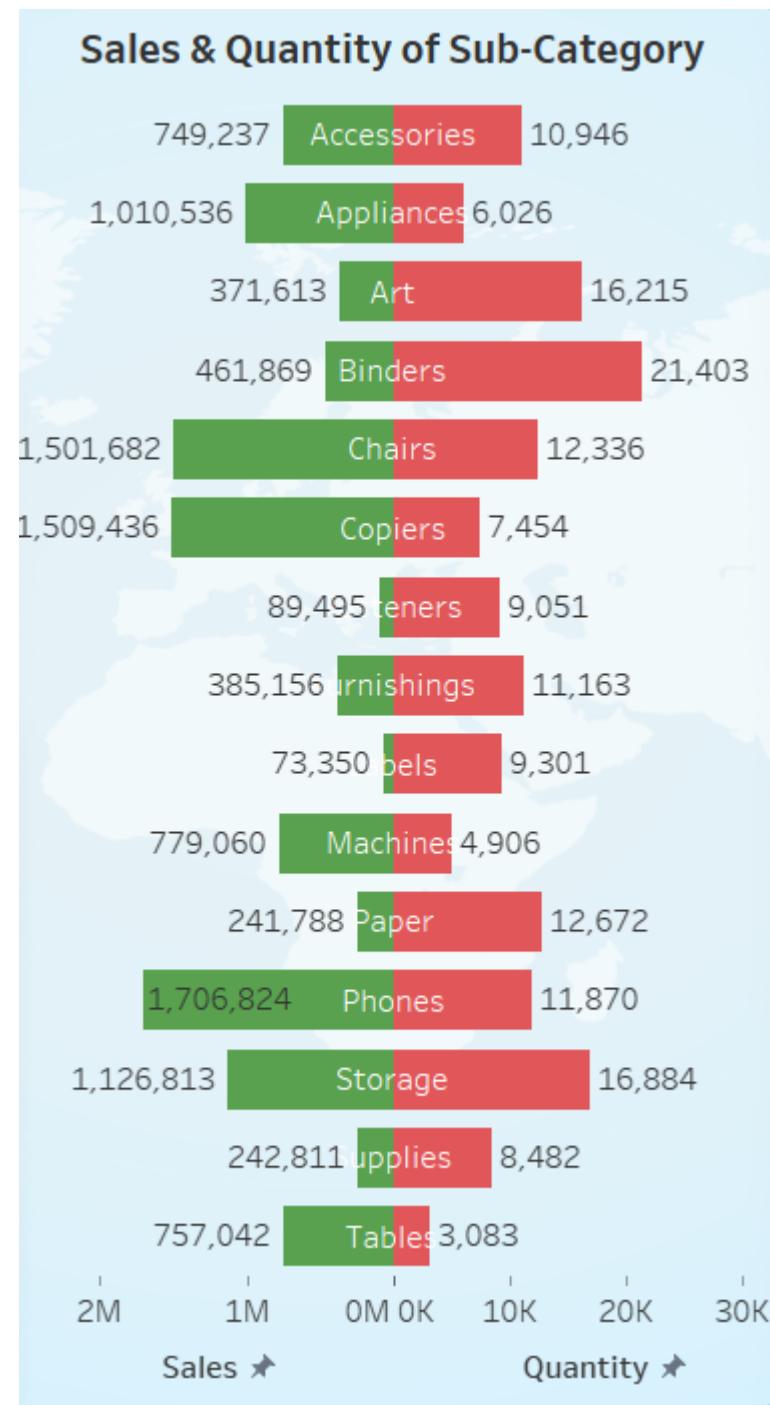
- Tập trung nâng cao doanh số bán hàng với các chiến dịch quảng cáo, khuyến mãi cho các sản phẩm đem lại doanh số cao.
- Sử dụng các kênh quảng cáo hiệu quả như : facebook, tiktok, website... để giới thiệu và quảng bá các sản phẩm đến với người tiêu dùng.
- Theo dõi các xu hướng và cập nhật sản phẩm mới trong các danh mục sản phẩm để đáp ứng nhu cầu của khách hàng.
- Tăng cường tính tương tác với khách hàng, nhận phản hồi và sử dụng các thông tin này để cải thiện chất lượng sản phẩm và dịch vụ.

2. Doanh số và số lượng mặt hàng đã bán của từng Sub Category.(Butterfly Chart)

Lý do lựa chọn biểu đồ:

- Butterfly chart là một biểu đồ thống kê được sử dụng để so sánh và hiển thị sự khác biệt giữa hai nhóm các dữ liệu số. Butterfly chart thường được sử dụng để so sánh đại lượng hoặc tỷ lệ của các biến giữa hai nhóm, và giúp tăng cường khả năng so sánh giữa các nhóm. Butterfly chart thường được biểu diễn dưới dạng hai cột đối xứng nhau, mỗi cột biểu thị cho một nhóm dữ liệu. Dữ liệu của mỗi cột được chia thành các khoảng giá trị đều nhau, và các biểu đồ cột của mỗi nhóm được đặt lên đối diện nhau.
- Butterfly chart là một công cụ tuyệt vời để so sánh sự khác biệt giữa các subcategory và đánh giá hiệu quả kinh doanh của chúng. Với số liệu doanh thu và số lượng bán ra của các subcategory trong superstore, butterfly chart có thể giúp bạn hiểu rõ hơn về sự khác biệt giữa về doanh số và số lượng sản phẩm bán ra của các subcategory từ đó tìm ra cách để tăng cường hiệu quả kinh doanh của các subcategory đó.

Trực quan:



Hình MKT-2. Butterfly Chart

Nhận xét:

- Chairs, Phones, Copiers là những subcategory mang lại doanh thu lớn cho các superstore.

- Label, Fasteners là những subcategory ít mang lại doanh thu cho các superstore.
- Paper và Binders là hai subcategory được bán với số lượng lớn nhất.
- Binders và Storage là hai subcategory được bán với số lượng lớn nhất. Trong khi đó, Table, Machine và Appliances có số lượng bán ra thị trường khá ít.
- Nhiều sản phẩm có lượng bán ra ở tầm trung nhưng lại mang lại nhiều doanh thu như Chairs và Phones. Tuy nhiên bên cạnh đó, paper tuy có doanh thu ít nhưng số lượng bán ra rất lớn. Điều này có thể dễ dàng được giải thích bởi sự chênh lệch trong giá bán của các sản phẩm.

Đề xuất chiến lược để cải thiện kinh doanh:

- Tập trung đầu tư và phát triển các subcategory mang lại doanh thu lớn và có số lượng bán ra lớn.
- Xem xét chiến lược như thay đổi phương thức bán hàng, tăng cường quảng cáo... để tăng doanh thu của các subcategory chưa mang lại doanh số cao.
- Nghiên cứu thị trường và khảo sát khách hàng để tìm hiểu lí do vì sao các subcategory như Tables, Machine và Appliances có số lượng bán ra thấp từ đó có những chiến lược tiếp thị phù hợp.

3. Lợi nhuận của các Sub Category và áp dụng nguyên lý Pareto cho các danh mục mặt hàng trên.(Pareto Chart)

Lý do lựa chọn biểu đồ:

- Pareto Chart là một công cụ thống kê được sử dụng để phân tích và biểu đồ hóa dữ liệu theo tỷ lệ phần trăm hoặc tần suất của từng nhóm trong một tập dữ liệu. Đây cũng là một công cụ mạnh mẽ trong việc phân tích và quản lý dữ liệu, giúp các tổ chức tăng cường hiệu quả và nâng cao chất lượng sản phẩm hoặc dịch vụ của mình.
- Với dữ liệu Superstore, chúng ta sử dụng Pareto Chart để phân tích và biểu diễn lợi nhuận của từng danh mục sản phẩm con. Bằng cách này, chúng ta có thể tìm ra những danh mục con quan trọng nhất và tập trung vào việc cải thiện chúng, nâng cao lợi nhuận, đưa ra các giải pháp phù hợp để tối ưu hóa dịch vụ kinh doanh.

Trực quan:



Hình MKT-3. Pareto Chart

Nhận xét:

- Copiers là danh mục con mang lại nhiều lợi nhuận nhất với 258.568\$ chiếm 17.62% trong tổng số các danh mục con.
- Trong khi đó, Tables là danh mục con duy nhất mang lợi nhuận âm (- 64.083\$)
- Khoảng 80% lợi nhuận đến từ các danh mục con là: Copier, Phones, Bookcases, Appliances, Chair, Accessories và Storage.

Đề xuất chiến lược để cải thiện kinh doanh:

- Tập trung vào các danh mục sản phẩm con được liệt kê ở trên để mang lại nhiều lợi nhuận, trong đó đặc biệt cần chú trọng với danh mục Copiers.
- Xây dựng các chiến lược tiếp thị, quảng cáo, và khuyến mãi tập trung nhiều vào các danh mục con mang lại 80% lợi nhuận.
- Nghiên cứu thị trường để hiểu rõ hơn về nhu cầu của khách hàng và điều chỉnh chiến lược tiếp thị sao cho phù hợp hơn với thị trường.

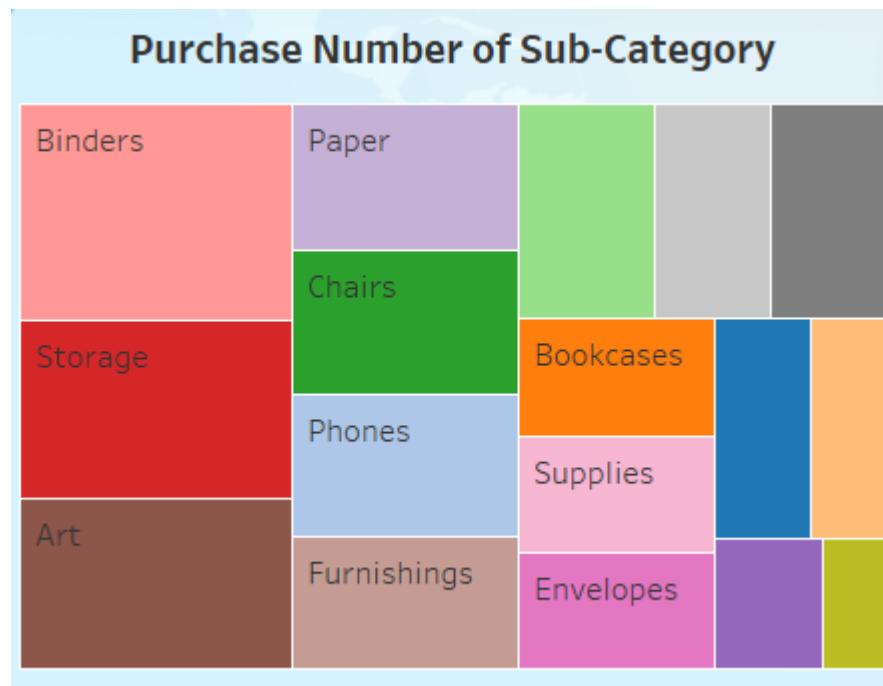
4. Tìm hiểu về số lượt mua hàng từ 2014 - 2018 của các Sub Category. (Tree Map)

Lý do lựa chọn biểu đồ:

- Treemap cho phép hiển thị tỷ lệ giữa các thành phần một cách rõ ràng, làm cho việc so sánh và phân tích dữ liệu dễ dàng hơn. Cụ thể, ta dễ dàng so sánh sản phẩm nào được mua nhiều nhất, ít nhất thông qua kích thước của chúng.

- Treemap cho phép nhìn thấy cấu trúc của dữ liệu phân cấp một cách dễ dàng, nhìn vào người dùng sẽ thấy ngay ta đang so sánh số lượng sản phẩm được mua tại Superstore. Nói cách khác, treemap cung cấp cái nhìn tổng quan về dữ liệu một cách nhanh chóng.
- Treemap có khả năng hiển thị màu sắc để tô đậm giá trị cao hơn (ứng với các sản phẩm được mua nhiều sẽ được vẽ với kích thước lớn và màu sắc đậm), giúp tăng tính trực quan và thu hút sự chú ý của người xem.
- Treemap hiển thị các thông tin chi tiết khác về mỗi đối tượng thông qua labels. Ở đây khi người dùng quan tâm đến một sản phẩm nào đó, họ có thể đọc các thông tin chi tiết khác ở labels như số lần mua, mức giảm giá cao nhất, doanh thu và lợi nhuận trung bình.

Trực quan:



Hình MKT-4.Tree Map

Nhận xét:

- Binders là sản phẩm có số lượt mua nhiều nhất (6146 lượt mua). Theo sau đó là Storage với số lượt mua là 5049.
- Tables có số lượt mua ít nhất (chỉ với 861 lượt) có doanh thu khá lớn tuy nhiên lợi nhuận lại là âm.
- Binders doanh thu nhiều nhưng lợi nhuận ít, có thể do mức giảm giá quá cao.

- Ngược lại, copiers (máy photo) tuy ít được mua (với 2223 lượt) trong vòng 4 năm nhưng đem lại lợi nhuận trung bình cao nhất: 116.31\$.

Đề xuất chiến lược để cải thiện kinh doanh:

- Với số lượt mua nhiều và mức giảm giá hấp dẫn, các sản phẩm Binders và Storage có tiềm năng thu hút nhiều khách hàng vì vậy nên tăng cường tiếp thị và quảng cáo đặc biệt cho hai sản phẩm này có thể giúp tăng độ nhận thức và tăng cường doanh số bán hàng.
- Xem xét điều chỉnh mức giảm giá để tối ưu hóa lợi nhuận mà vẫn giữ được sự hấp dẫn đối với khách hàng tránh trường hợp sản phẩm có doanh thu nhiều nhưng lợi nhuận quá ít.
- Tăng cường quảng cáo và áp dụng các chiến dịch thúc đẩy mua sắm nhằm tăng sự nhận thức và tăng cường doanh số bán hàng cho sản phẩm có doanh thu lớn như Tables và sản phẩm mang lại lợi nhuận cao như Copies.

5. Tìm hiểu các mặt hàng sản phẩm nào thường được khách hàng mua cùng nhau (Association Rule)

Lợi ích khi trả lời câu hỏi:

- Khi các doanh nghiệp, quản lý cửa hàng và nhà đầu tư của Superstore biết được các sản phẩm thường được khách hàng mua cùng nhau, từ đó áp dụng các chiến lược để tăng cường trải nghiệm mua hàng của khách hàng và tăng doanh số bán hàng cho doanh nghiệp.
- Ngoài ra, việc biết được các sản phẩm thường được mua cùng nhau cũng giúp doanh nghiệp lên kế hoạch sản xuất và kinh doanh hiệu quả hơn, đồng thời tối ưu hóa quy trình đặt hàng, lưu kho và vận chuyển hàng hóa.

Các bước thực hiện:

Các bước thực hiện được trình bày trong file *association_rule.ipynb* nằm trong thư mục **source_codes**.

Trực quan:

antecedents	consequents	support	confidence
(Paper)	(Binders)	0.025653	0.206314
(Storage)	(Binders)	0.034787	0.196057
(Accessories)	(Binders)	0.021883	0.194608
(Phones)	(Binders)	0.023710	0.194206
(Furnishings)	(Binders)	0.022155	0.192828
(Art)	(Binders)	0.032299	0.189122
(Chairs)	(Binders)	0.022427	0.180538
(Art)	(Storage)	0.030356	0.177742
(Storage)	(Art)	0.030356	0.171084
(Binders)	(Storage)	0.034787	0.164764

Hình MKT-5. Association Rules

Nhận xét:

- Ta rút trích được 16 luật kết hợp, nhưng khi thống kê trên Phòng Marketing ta chỉ lấy top 10 luật kết hợp có chỉ số confidence cao trong đó có thể kể đến là:

- Paper → Binders: Nếu mua giấy thì có thể sẽ mua tập bìa còng chứa tài liệu.
- Storage → Binders: Nếu mua các vật dụng chứa đồ(tủ, giá sách, ngăn kéo...) thì có thể sẽ mua tập bìa còng chứa tài liệu.
- Accessories → Binders: Nếu mua phụ kiện công nghệ thì có thể sẽ mua bìa còng chứa tài liệu.
- Phones → Binders: Nếu mua điện thoại thì có thể sẽ mua tập bìa còng chứa tài liệu.
- Furnishings → Binders: Nếu mua các thiết bị nội thất thì có thể sẽ mua tập bìa còng chứa tài liệu.
- Art → Binders: Nếu mua các dụng cụ vẽ thì có thể sẽ mua bìa còng chứa tài liệu.
- Chairs → Binders: Nếu mua ghế thì có thể sẽ mua tập bìa còng chứa tài liệu.
- Art → Storage: Nếu mua các dụng cụ vẽ thì sẽ có thể mua các vật dụng chứa đồ và ngược lại.

- Binders → Storage: Nếu mua tập bìa còng chứa tài liệu thì có thể sẽ mua các vật dụng chứa đồ.
- Một số luật được kể đến trên khá hợp lý theo logic thông thường, tuy nhiên một vài luật cần có sự kiểm chứng thêm trước khi áp dụng chúng cho các chiến lược kinh doanh của các superstore.

Đề xuất chiến lược để cải thiện kinh doanh:

- **Product Bundling:** Nếu một số sản phẩm thường được mua cùng nhau, phòng marketing có thể đề xuất tạo ra các product bundles hoặc gói combo để khuyến khích khách hàng mua nhiều sản phẩm cùng một lúc. Ví dụ: ta có thể thực hiện đóng gói giấy, và tập bìa còng chứa tài liệu vào một bundle có giá ưu đãi, có thể tăng doanh số bán hàng và khách hàng sẽ có lợi.
- **Cross-Selling:** là một kĩ thuật bán hàng được sử dụng để khiến khách hàng chi tiêu nhiều hơn bằng cách mua một sản phẩm có liên quan đến những gì đã được mua. Phòng marketing có thể sử dụng chiến lược cross-selling để khuyến khích khách hàng mua sản phẩm đi kèm khi biết chúng có mối liên hệ mạnh với nhau. Ví dụ: nếu khách hàng thường mua dụng cụ vẽ, có thể đề xuất đồng thời bán thêm giấy để tăng giá trị cho đơn hàng.
- **Personalization:** Phân tích association rule có thể giúp phòng marketing hiểu được nhóm khách hàng và sở thích của họ. Dựa trên thông tin này, superstore có thể tạo ra các chiến dịch tiếp thị cá nhân hóa dành riêng cho từng nhóm khách hàng. Ví dụ: nếu có một nhóm khách hàng thường mua các dụng cụ vẽ cùng với sản phẩm tủ hay giá sách, có thể gửi các ưu đãi đặc biệt hoặc tin tức liên quan đến cả hai loại sản phẩm này.
- **Recommendation system:** dựa trên những association rule, khi khách hàng chọn mua một sản phẩm, hệ thống sẽ gợi ý cho khách hàng các sản phẩm khác thường được mua cùng sản phẩm đó. Điều này giúp khách hàng tìm được các sản phẩm liên quan và tương tự nhau, cải thiện trải nghiệm mua sắm của khách hàng, đồng thời cũng giúp doanh nghiệp tăng doanh số và tính hiệu quả kinh doanh.

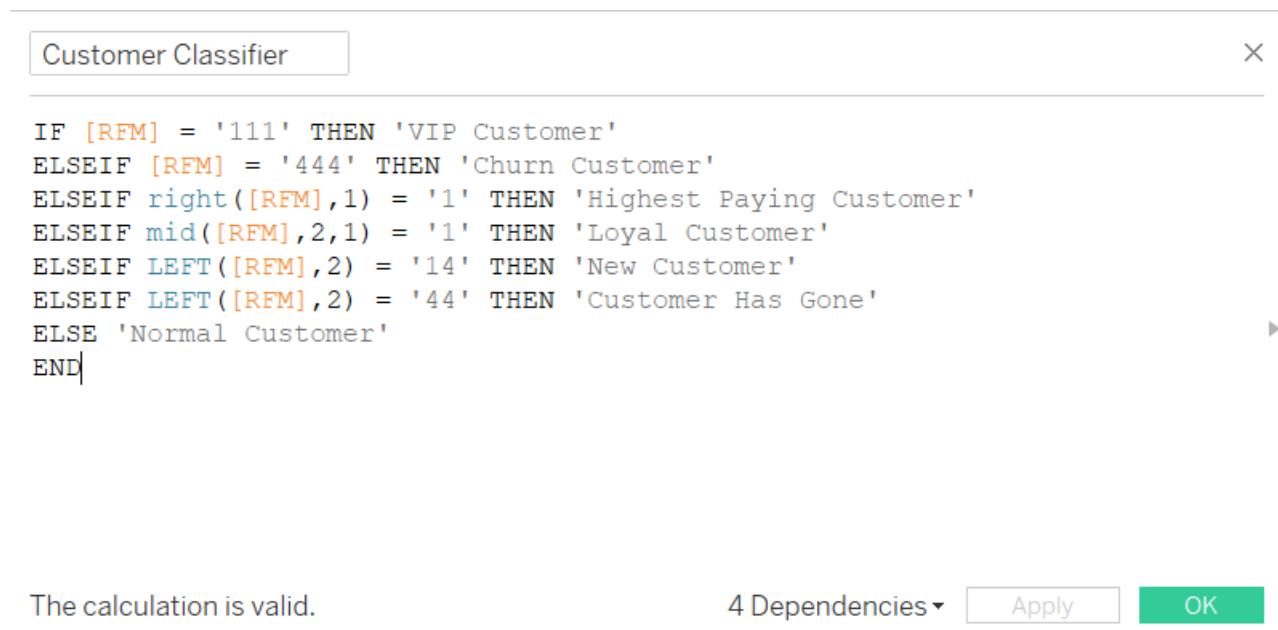
3.4 Phòng Chăm sóc khách hàng

Nhu cầu : Phòng ban CSKH cần thông tin về phản hồi về chất lượng dịch vụ, phân tích tương tác KH, thời gian giải quyết khiếu nại, hồ sơ KH.

1. Phân loại khách hàng tìm năng bằng RFM

- Giải thích về RFM:
 - + Phân tích RFM được dựa vào 3 chỉ số là R (Recency), F(Frequency), M (Mone-tary) hay Thời gian mua gần đây, tần suất mua, và số tiền mua hàng.
 - + Từng khách hàng sẽ được tính các chỉ số này, được chia làm 4 bậc, bậc 1 là cao nhất và 4 là thấp nhất.

- + Để đánh giá là bậc nào ta cần so sánh từng chỉ số của khách hàng đó với các tứ phân vị (Quantile). Sau đó tiến hành tính các chỉ số R,F,M riêng lẻ.
- + Chỉ số RFM cuối cùng là trường được tính bằng cách cộng chuỗi R,F,M. Sau đó dựa vào đó để phân loại khách hàng.
- + Cách phân loại bằng chuỗi này:



```

Customer Classifier
X

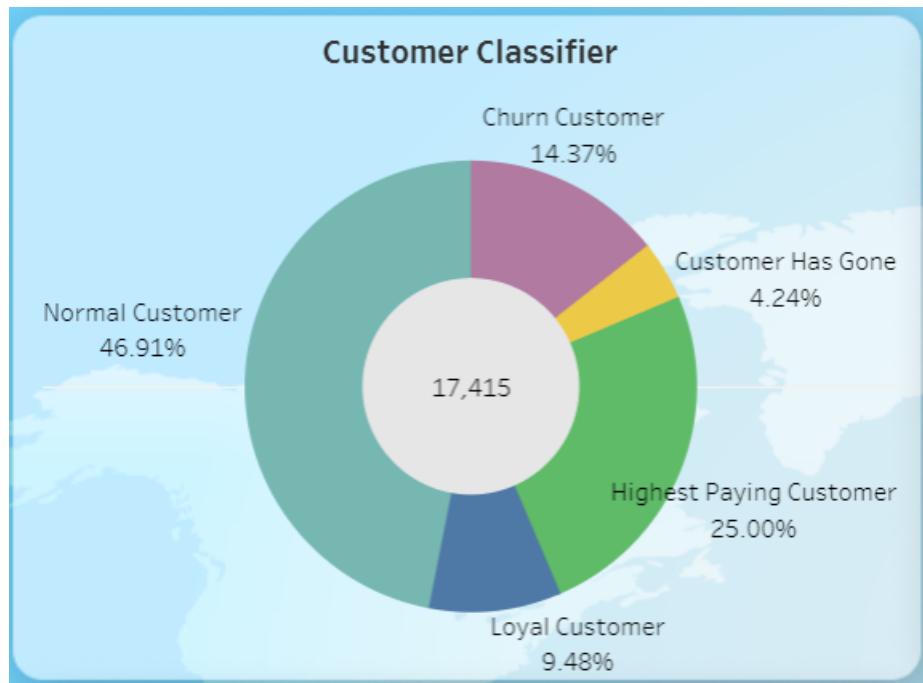
IF [RFM] = '111' THEN 'VIP Customer'
ELSEIF [RFM] = '444' THEN 'Churn Customer'
ELSEIF right([RFM], 1) = '1' THEN 'Highest Paying Customer'
ELSEIF mid([RFM], 2, 1) = '1' THEN 'Loyal Customer'
ELSEIF LEFT([RFM], 2) = '14' THEN 'New Customer'
ELSEIF LEFT([RFM], 2) = '44' THEN 'Customer Has Gone'
ELSE 'Normal Customer'
END
  
```

The calculation is valid. 4 Dependencies ▾ Apply OK

Hình CSKH-1. RFM**Lý do lựa chọn biểu đồ:**

- Pie chart là một loại biểu đồ hai chiều hiển thị dữ liệu dưới dạng các phần tương ứng với phần trăm hoặc tỷ lệ của mỗi phần so với toàn bộ. Nó thường được sử dụng để biểu diễn phân phối dữ liệu theo các nhóm hoặc các thành phần khác nhau.
- Biểu đồ pie chart có hình dạng giống như một cái bánh, trong đó tổng góc được chia thành các phần tương ứng với quy mô tương ứng của từng nhóm. Kích thước của các phần sẽ phản ánh tỷ lệ phần trăm hoặc tỷ lệ số lượng mỗi nhóm so với tổng số.
- Pie chart giúp cho việc so sánh và phân tích các phần trong một tập hợp dữ liệu dễ dàng hơn, và nhanh chóng truyền tải thông tin một cách trực quan.

Trực quan:

**Hình CSKH-2. Customer Classifier****Nhận xét:**

- Có tổng cộng 17415 khách hàng được phân thành 5 loại, trong đó Normal Customer chiếm tỉ lệ nhiều nhất với 46.91% và Customer Has Gone chiếm tỉ lệ ít nhất 4.24%.
- Đa số là những khách hàng bình thường và Highest Paying Customer (những khách hàng trả rất nhiều để mua hàng). Tỉ lệ khách hàng trung thành chỉ chiếm 1 phần nhỏ, còn tỉ lệ khách hàng rời đi và đã dừng mua cũng chiếm 1 phần tương đối. Điều này cho thấy chất lượng dịch vụ chưa thật sự nổi bật, chỉ ở mức ổn.

Đề xuất chiến lược để cải thiện kinh doanh:

- Nên có thêm các chiến dịch phục vụ khách hàng như tích điểm giảm giá, đổi thưởng... cho các khách hàng trung thành trở lên.
- Với các khách hàng trung thành và trả nhiều nên có các ưu đãi tốt để giữ được các khách hàng tiềm năng nhằm gia tăng lợi nhuận.

2. Tìm hiểu về các đơn hàng và tình trạng ship.**Lý do lựa chọn biểu đồ:**

- Stacked bar chart (biểu đồ cột chồng) là một loại biểu đồ hai chiều được sử dụng để hiển thị sự phân bổ và so sánh các giá trị của các thành phần khác nhau trong một tập dữ liệu. Nó cho phép bạn xem tổng giá trị của mỗi thành phần cũng như phần trăm hoặc tỷ lệ mà từng thành phần chiếm trong tổng.

- Stacked bar chart phổ biến trong việc biểu diễn các dữ liệu có sự phụ thuộc.
- Thông qua việc chồng các cột lên nhau, stacked bar chart cho phép nhìn thấy sự tích luỹ của các giá trị và cũng giúp phân loại dễ dàng hơn. Từ đó có thể dễ dàng so sánh tổng giá trị và cũng như mức đóng góp của từng thành phần trong mỗi nhóm.

Trực quan:



Hình CSKH-3. Shipping status by months

Nhận xét:

- Số lượng đơn hàng có xu hướng tăng cao vào cuối năm, nhiều nhất vào tháng 9,11,12. Ít nhất vào khoảng thời gian đầu năm.
- Ở mỗi tháng, tỉ lệ ship sớm vẫn chiếm phần lớn, tuy nhiên vẫn còn tồn tại nhiều đơn hàng với tình trạng ship trễ với số lượng đáng kể. (khoảng 25%)

Đề xuất chiến lược để cải thiện kinh doanh:

- Cần xem xét các trường hợp ship trễ 1 cách kĩ càng, như xem về mặt hàng, khu vực, số lượng,... để có thể tìm hiểu nguyên nhân làm ship trễ, lắng nghe phản hồi từ khách hàng, từ đó giảm bớt tình trạng ship hàng trễ.
- Ngoài ra còn có thể hợp tác thêm với các dịch vụ giao hàng khác để có thể cải thiện tình trạng giao hàng trễ.

3. Tìm hiểu về số lượng hàng trả về từ các khu vực.

Lý do lựa chọn biểu đồ:

- Biểu đồ cột ngang (horizontal bar chart) là một loại biểu đồ hai chiều được sử dụng để biểu diễn dữ liệu theo các nhóm hoặc các thành phần khác nhau. Trong biểu đồ

này, các nhóm được hiển thị trên trục y (trục dọc) và độ dài của mỗi cột biểu thị giá trị hoặc số lượng tương ứng với từng nhóm. Biểu đồ cột ngang thường được sử dụng để so sánh giá trị giữa các nhóm và hiển thị dữ liệu theo thứ tự từ trái sang phải.

- Biểu đồ cột ngang rất linh hoạt và có thể áp dụng cho nhiều ngành nghề và tình huống khác nhau. Nó giúp trực quan hóa dữ liệu, làm nổi bật các thông tin quan trọng, và mang lại cái nhìn tổng quan về dữ liệu một cách dễ dàng.

Trực quan:



Hình CSKH-4. Top 10 most returned region

Nhận xét:

- Top 10 khu vực trả về đều thuộc châu Á, Âu và Mỹ. Trong đó Tây Âu và Trung Mỹ là 2 khu vực với số lượng trả về nhiều nhất và là 2 khu vực duy nhất với hơn 100 đơn hàng trả về, cách xa các khu vực còn lại.
- Các khu vực còn lại có lượng trả về không chênh lệch nhiều đối với 2 khu vực top liên tiếp nhau.

Đề xuất chiến lược để cải thiện kinh doanh:

- Cần khảo sát về lí do trả về, đặc biệt với những khu vực có lượng trả về nhiều (như top 10) hoặc những khu vực đặc biệt cao so với phần còn lại.

- Tiếp nhận các feedbacks nhằm đánh giá lí do cụ thể là gì, từ đó nâng cao dịch vụ nhằm làm hài lòng khách hàng, từ đó giảm số lượng return cũng như tăng doanh thu cho cửa hàng.

4. Tìm hiểu khả năng khách hàng quay lại(Cohort analysis)

- Giải thích về Cohort analysis:

+ Cohort analysis xem xét việc 1 khách hàng quay lại mua hàng sau lần mua hàng đầu tiên.

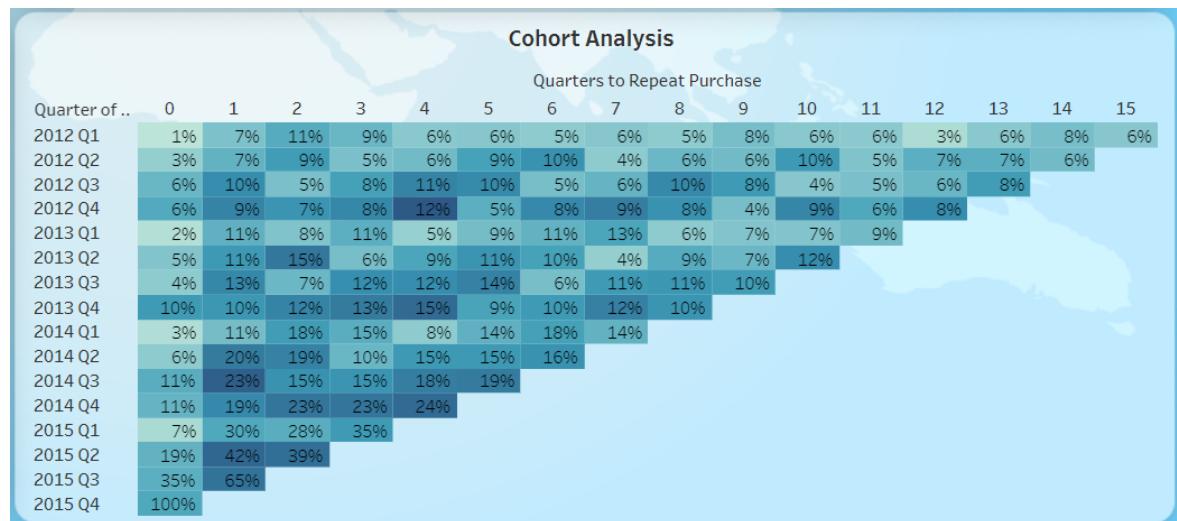
+ Để làm việc này ta sẽ tìm lần mua hàng đầu tiên với mỗi khách hàng, và những lần sau đó đều tính là quay lại. (tính theo quý)

Lý do lựa chọn biểu đồ: - Sparse heatmap là một biểu đồ heatmap được sử dụng để biểu diễn dữ liệu thừa, trong đó hầu hết các ô có giá trị là zero (0) hoặc thiếu thông tin.

- Sparse heatmap nhấn mạnh vào các giá trị không phải zero hoặc không bị thiếu để tạo ra một biểu đồ tập trung vào những điểm dữ liệu quan trọng. Các ô có giá trị không phải zero hoặc không bị thiếu thường được biểu thị bằng các màu sắc khác nhau hoặc các ký hiệu đặc biệt, trong khi các ô có giá trị zero hoặc thiếu thông tin thường được để trống hoặc được hiển thị bằng một màu trung tính.

- Kỹ thuật này giúp tạo ra một biểu đồ tập trung vào những điểm dữ liệu quan trọng và giảm thiểu việc hiển thị các ô trống hoặc có giá trị zero.

Trực quan:



Hình CSKH-5. Cohort analysis

Nhận xét:

- Có thể thấy, lượng khách quay lại trong năm nhiều tập trung vào các quý đầu hơn là các quý sau.
- Lượng khách quay lại nhiều nhất vào quý 4/2012 . So với các năm khác thì có vẻ 2014 là năm có lượng khách quay lại cao nhất.

Đề xuất chiến lược để cải thiện kinh doanh:

- Đề xuất: Tìm hiểu kỹ các loại mặt hàng mà khách quay lại mua nhiều nhằm gia tăng lượng giữ chân khách hàng.
- Xem lại các quý có lượng khách giữ chân cao nhằm so sánh, đổi chiếu nhằm tìm ra giải pháp giữ chân hiệu quả.
- Ngoài ra có thể tìm hiểu thêm một số nguyên nhân ngoài sản phẩm (thời gian ship, tính toàn vẹn sản phẩm,...) nhằm cải thiện để giữ chân nhiều khách hàng hơn.

3.5 Việt Nam

Nhu cầu: nhóm mong muốn hiểu rõ hơn về tình hình kinh doanh của Global Superstore tại thị trường Việt Nam, so sánh hiệu suất kinh doanh của Global Superstore tại thị trường Việt Nam với các thị trường khác, từ đó đánh giá hiệu quả của chiến lược kinh doanh hiện tại và đặt ra các mục tiêu kinh doanh mới cho thị trường này. Chúng em hi vọng Global Superstore sẽ tối ưu hóa hoạt động kinh doanh và đạt được thành công tại Việt Nam.

3.5.1 Tổng quan về doanh thu, lợi nhuận và số đơn hàng tại Việt Nam ở mỗi năm? Tăng hay giảm so với năm trước đó?

1. Lý do lựa chọn biểu đồ

- **Số liệu to, rõ ràng:** hiển thị các thông tin chính dưới dạng số liệu về doanh thu, lợi nhuận và số đơn hàng của các năm. Chỉ số YoY được sử dụng để so sánh mức độ tăng trưởng so với năm trước đó. Điều này góp phần tạo sự chú ý, giúp người xem nhanh chóng có cái nhìn tổng quan về sự tăng trưởng và phát triển qua các năm.
- **Line chart cho năm hiện tại:** theo dõi sự thay đổi của doanh thu, lợi nhuận và số đơn hàng trong năm hiện tại qua các tháng trong năm. Line chart thường được sử dụng để biểu diễn xu hướng và biến động theo thời gian, giúp người xem nhận ra sự tăng trưởng hoặc giảm giá trị trong quá trình thời gian.
- **Area chart cho năm trước:** theo dõi sự thay đổi của doanh thu, lợi nhuận và số đơn hàng trong năm trước đó qua các tháng và dễ dàng nhìn thấy sự chênh lệch so với năm hiện tại.

2. Trực quan



Hình VN-1. Doanh thu, lợi nhuận và số đơn hàng tại Việt Nam năm 2015 so với 2014

3. Nhận xét

- Nhìn chung trong năm 2015 ở Việt Nam, cả doanh thu và lợi nhuận đều giảm rõ rệt so với 2014, tuy nhiên tổng số đơn hàng không đổi (33 đơn hàng).
- Doanh thu cao nhất vào năm 2014 với \$19.111, thấp nhất năm 2012 với \$12.857. Có thể thấy doanh thu giữ mức tăng trưởng đều hằng năm tuy nhiên đến 2015 lại có dấu hiệu suy giảm dù không nhiều (\$18.657).
- Tuy nhiên, lợi nhuận có xu hướng gần như trái ngược với doanh thu: cao nhất năm 2012 với \$976.7, cũng là năm có lợi nhuận dương duy nhất. Các năm còn lại đều thua lỗ, nhất là năm 2013 với \$-1332, sau đó con số này đã tăng lên nhưng vẫn còn thua lỗ.
- Có vẻ như Superstore vẫn còn chưa phổ biến ở thị trường Việt Nam khi trong vòng một năm mà có chưa đến 40 đơn đặt hàng. Số lượng đơn hàng tăng qua các năm nhưng mức tăng rất khiêm tốn, trong năm 2015 chỉ đạt 33 đơn hàng.

3.5.2 Doanh thu và lợi nhuận khác nhau như thế nào giữa các danh mục con sản phẩm?

1. Lý do lựa chọn biểu đồ

- **Bubble chart:** thích hợp để so sánh và tương quan giữa 2 biến số doanh thu và lợi nhuận, làm rõ mức độ và xu hướng ở mỗi danh mục con.

- **Kích thước** bubble thể hiện lợi nhuận và **màu sắc** đậm nhạt thể hiện doanh thu, tạo nên sự đa dạng và hấp dẫn trong việc trình bày dữ liệu.
- **Dễ hiểu** và thu hút sự chú ý của người xem. Có tính năng **tương tác**, cho phép người dùng di chuột qua các bubble để xem thông tin chi tiết về giá trị của doanh thu và lợi nhuận.

2. Trực quan



Hình VN-2. Doanh thu và lợi nhuận giữa các danh mục sản phẩm

3. Nhận xét

- Appliances là sub-category có lợi nhuận (\$1327) lớn nhất ở thị trường Việt Nam. Điều này dễ hiểu bởi sự phát triển kinh tế, hàng loạt công ty được xây dựng, nhu cầu tiêu dùng cho các sản phẩm này đã tăng cao, không những thế các sản phẩm trong Appliances, như máy tính, máy in, máy chiếu, là công cụ cần thiết cho hoạt động văn phòng và kinh doanh. Vì không thể thay thế bằng các sản phẩm khác, việc mua sắm và sử dụng các sản phẩm này trở thành một yếu tố quan trọng và thiết yếu. Ngoài ra, Sản phẩm trong Appliances thường có giá trị tương đối cao. Vì vậy, mỗi giao dịch sẽ mang lại doanh thu lớn hơn so với các sub-category khác, đồng thời tạo ra lợi nhuận cao.

- Doanh thu của Tables ở thị trường việt nam là lớn nhất (\$11282) điều này khá dễ hiểu vì bàn là mặt hàng thiết yếu trong cuộc sống nên nhu cầu tiêu dùng mặt hàng này khá lớn dẫn đến doanh thu của nó cao, tuy nhiên lợi nhuận đến từ sản phẩm này khá thấp so với những mặt hàng khác (\$143) điều này có thể đến từ sự cạnh tranh của thị trường, và để sản xuất vận chuyển sản phẩm cần chi phí lớn đã ảnh hưởng nhiều đến lợi nhuận cuối cùng của sản phẩm.

- Ngược lại, Fasteners mang lại doanh thu thấp nhất (\$549) còn sản phẩm có lợi nhuận thấp nhất là Label (\$76) bởi lẽ các sản phẩm này thường có giá trị thấp và không tạo qua nhiều lợi nhuận, sản phẩm Fasteners và Label không phải là các mặt hàng thiết yếu, và nhu cầu cho chúng có thể thấp hơn so với các loại sản phẩm khác.

3.5.3 Doanh thu thay đổi như thế nào qua các thành phố? Thành phố nào có doanh thu cao nhất, thấp nhất?

1. Lý do lựa chọn biểu đồ

- **Biểu đồ cột ngang:** trực quan dữ liệu dạng số, cho phép so sánh mức độ chênh lệch doanh thu giữa các thành phố bằng cách sử dụng chiều ngang của cột để đại diện cho giá trị
- Tiết kiệm không gian, diện tích, dữ liệu được trực quan rõ ràng, không chồng đè lên nhau
- Giá trị trên trục x được log lại để đỡ chiếm diện tích vì chênh lệch doanh thu giữa TP.HCM, Hà Nội và các thành phố khác khá lớn.

2. Trực quan



Hình VN-3. Doanh thu qua các thành phố

3. Nhận xét

- TP.HCM và Hà Nội có tổng doanh thu dẫn đầu với \$42.670 và \$11.054, chênh lệch khá lớn với doanh thu ở các thành phố khác.
- Sử dụng filter “Ship mode” và “Segment” cho thấy: tổng doanh thu ở TP.HCM vẫn dẫn đầu. Điều đó cho thấy TP.HCM là thị trường sôi động với đa dạng chế độ giao hàng và phân khúc khách hàng.

- Chỉ có 6 thành phố lớn ở Việt Nam tiếp cận được tới các sản phẩm của Superstore, ngoài TP.HCM và Hà Nội ra thì doanh thu ở các thành phố khác không đáng kể.

3.5.4 Tổng số lượng sản phẩm bán được theo danh mục con sản phẩm và phân khúc khách hàng. Các khuyến mãi (KM) đã thực sự hiệu quả chưa?

1. Lý do lựa chọn biểu đồ

- **Table:** là một cách hiển thị dữ liệu có cấu trúc và chi tiết, cho phép người xem xem thông tin chi tiết và số liệu cụ thể về số lượng sản phẩm bán được trong từng danh mục con sản phẩm và phân khúc khách hàng.

- **Heatmap:** thể hiện mức độ và tương quan của dữ liệu thông qua màu sắc. Người xem dễ dàng nhận ra xu hướng KM: các ô màu sáng thể hiện KM trung bình ít, các ô có màu tối hơn thể hiện KM trung bình cao hơn.

- **Kết hợp:** góp phần tạo ra một bức tranh tổng quan và chi tiết về số lượng sản phẩm bán được và hiệu quả KM. Người xem dễ dàng theo dõi và so sánh số liệu, tạo ra một trải nghiệm trực quan và thú vị.

2. Trực quan

Sum of Quantity and Average Discount by Sub-category and Segment					
Category	Sub-Category	Consumer	Corporate	Home Office	Grand Total
Furniture	Bookcases	37	12	6	55
	Chairs	50	8	28	86
	Furnishings	49	11	16	76
	Tables		6	4	10
Office Supplies	Appliances	6	16	6	28
	Art	27	9	21	57
	Binders	35	16	9	60
	Envelopes	34	2	6	42
	Fasteners	16	17	9	42
	Labels	43	19	13	75
	Paper	17	19	15	51
	Storage	46	35	15	96
	Supplies	11	23	38	72
	Accessories	81	6	18	105
Technology	Copiers	17	16	20	53
	Machines	6	6	3	15
	Phones	22	21	25	68
	Grand Total	497	242	252	991

Hình VN-4. Bảng số liệu tổng sản phẩm và trung bình KM theo danh mục con sản phẩm và phân khúc khách hàng

3. Nhận xét

- Accessories là sub-category nằm trong danh mục Technology có tổng số lượng bán lớn nhất trong các sub-category.

- Tables là sub-category nằm trong danh mục Furniture có tổng số lượng bán thấp nhất trong các sub-category.

- Ở phân khúc khách hàng consumer, sản phẩm được mua nhiều nhất là accessories. Ngược lại tables là sản phẩm không được ưa chuộng ở phân khúc này khi không có sản phẩm nào được bán ra.
- Ở phân khúc khách hàng Corporate, sản phẩm được ưa chuộng là Storage còn sản phẩm được mua ít nhất là envelopes.
- Supplies là sub-category được mua nhiều nhất ở phân khúc khách hàng Home Ofice. Còn Machines là sản phẩm ít được ưa chuộng nhất.
- Tuy được bán ra với số lượng khá ít nhưng tables lại là sản phẩm có tỉ lệ giảm giá trung bình cao nhất trong các sản phẩm (57%) theo sau là accessories (47%) và Copier, Bookcases (37%).

3.5.5 Đề xuất chiến lược để cải thiện kinh doanh

- Tăng cường hoạt động marketing và quảng bá ở Việt Nam: Việt Nam thuộc thị trường Châu Á - Thái Bình Dương, đây là thị trường năng động nhưng khá phức tạp, nhu cầu hàng hóa và dịch vụ cao. Tuy nhiên, tình hình kinh doanh của superstore tại Việt Nam không khả quan, khi doanh thu và lợi nhuận đã giảm trong năm 2015. Để đảo ngược tình thế, công ty cần thúc đẩy các hoạt động marketing và quảng bá sản phẩm về Việt Nam để thu hút khách hàng và tăng doanh số bán hàng.

- Tập trung vào các sản phẩm và danh mục có doanh thu và lợi nhuận cao: Các sub-category như Appliances và Tables có doanh thu và lợi nhuận cao tại thị trường Việt Nam. Superstore có thể tập trung phát triển và quảng bá những sản phẩm trong các danh mục này để tăng doanh thu và lợi nhuận.

- Mở rộng thị trường tiềm năng: Superstore cần xem xét mở rộng hoạt động kinh doanh đến các thành phố khác trong Việt Nam, không chỉ tập trung ở TP.HCM và Hà Nội. Điều này giúp tăng khả năng tiếp cận khách hàng mới và mở rộng doanh thu.

- Cải thiện hiệu quả khuyến mãi: Dựa trên thông tin về tỷ lệ giảm giá trung bình của các sản phẩm, Superstore có thể xem xét cải thiện hiệu quả của các chương trình khuyến mãi và giảm giá để thu hút khách hàng và tăng doanh số bán hàng. Tuy nhiên cần đảm bảo khuyến mãi trong một mức độ cho phép để tối ưu hóa lợi nhuận.

- Nâng bắt xu hướng tiêu dùng: tiến hành các biện pháp nghiệp vụ, hiểu rõ tâm lý, sở thích và nhu cầu tiêu dùng của từng đối tượng khách hàng để điều chỉnh chiến lược kinh doanh và cung cấp các sản phẩm phù hợp.

- Tăng cường hoạt động đối tác và nhà cung cấp: Để tối ưu hóa hiệu quả kinh doanh, Superstore cần thiết lập và duy trì mối quan hệ tốt với đối tác và nhà cung cấp, có thể cân nhắc tìm các đối tác và nhà cung cấp tại Việt Nam. Điều này đảm bảo cung cấp hàng hóa chất lượng, linh hoạt trong giá cả và mở rộng hợp tác tốt hơn.

- Theo dõi và phân tích thị trường: Superstore cần thường xuyên theo dõi và phân tích các chỉ số kinh doanh, xu hướng thị trường và sự cạnh tranh để điều chỉnh chiến lược kinh doanh và đáp ứng nhanh chóng với biến động thị trường, đặc biệt là thị trường phức tạp như Châu Á - Thái Bình Dương.

4 Dự đoán doanh thu và lợi nhuận Superstore.

Bài toán được đặt ra là: "Dựa trên các đặc trưng về ngày đặt/giao hàng, chế độ vận chuyển, phân khúc khách hàng, thị trường, sản phẩm, số lượng, mức giảm giá... hãy dự đoán doanh thu và lợi nhuận Superstore thu được ứng với mặt hàng đó".

4.1 Lợi ích khi trả lời câu hỏi

Việc dự đoán trước doanh thu và lợi nhuận có rất nhiều lợi ích cho các đối tượng khác nhau như sau:

- Đối với Doanh nghiệp: điều chỉnh các chiến lược bán hàng để tối đa hóa lợi nhuận; phát hiện ra các khuynh hướng tiêu dùng và hành vi mua hàng của khách hàng, từ đó đưa ra các quyết định kinh doanh phù hợp để thu hút và giữ chân khách hàng.
- Đối với Nhà quản lý: hiểu rõ hơn về sản phẩm và dịch vụ được ưa chuộng nhất để tối ưu hóa sản xuất và phân phối; cải thiện hiệu quả quản lý đơn hàng, quản lý kho.
- Đối với Nhà đầu tư: chuẩn bị ngân sách và dự trù tài nguyên tốt hơn để đáp ứng nhu cầu kinh doanh trong tương lai.

4.2 Các bước thực hiện

Ta sử dụng 3 mô hình học máy là **Decision Tree Regression**, **Random Forest Regression** và **TensorFlow neural network**, nhìn chung có 4 bước như sau:

1. Lựa chọn các đặc trưng để đưa vào mô hình.
2. Tiền xử lý trước khi chạy mô hình.
3. Xây dựng các mô hình.
4. Đánh giá các mô hình.

Các bước trên được trình bày chi tiết trong file *forecasting.ipynb*, thư mục **source_code**.

4.3 Kết quả mô hình

Bảng 1: Sales and Profit of Models

Model	Sales		Profit	
	MSE	R-squared	MSE	R-squared
Dicision Tree Regression	102349.282458	0.620303	24913.643018	0.297160
Random Forest Regression	87557.755217	0.675177	18899.983340	0.466812
TensorFlow Neural Network	273440.450636	-0.014413	24805.592616	0.300208

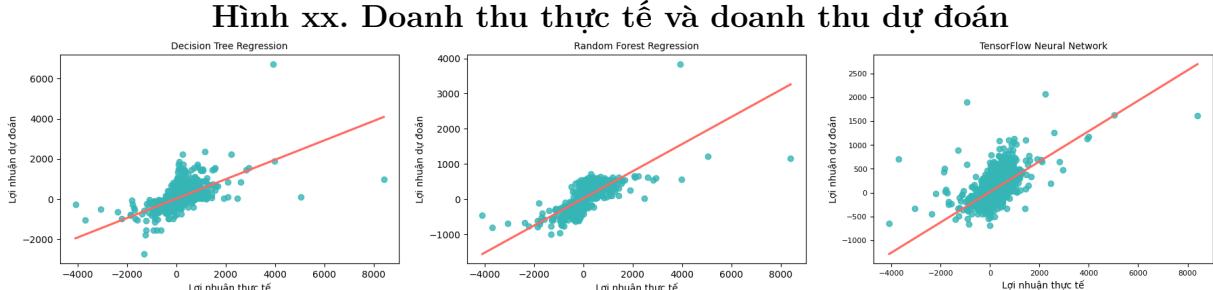
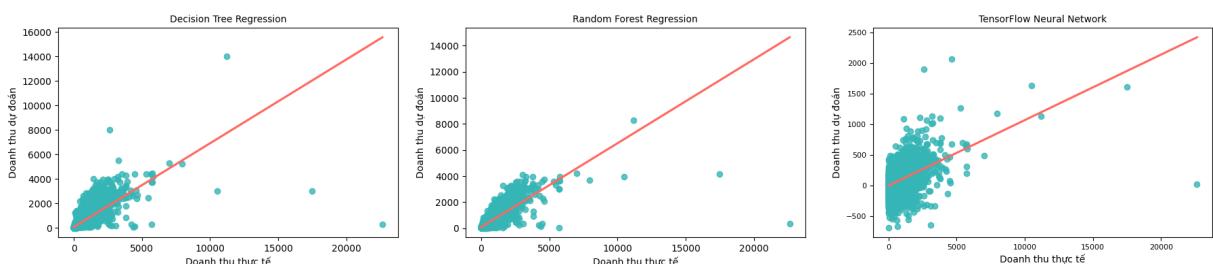
4.4 Nhận xét

- Mean Squared Error (MSE) là một độ đo thường được sử dụng để đánh giá chất lượng của một mô hình dự đoán. Nó tính toán trung bình bình phương của sai số giữa các giá trị dự đoán và các giá trị thực tế. MSE càng nhỏ thì mô hình dự đoán càng tốt. Như vậy mô hình Random Forest dự đoán tốt nhất cả doanh thu và lợi nhuận.

- R-square là một thước đo đánh giá chất lượng của mô hình hồi quy và biểu thị tỉ lệ phần trăm của sự biến động của biến phụ thuộc (y) có thể được giải thích bởi biến độc lập (x) đã chọn cho mô hình. Giá trị R-square nằm trong khoảng từ 0 đến 1, với giá trị càng gần 1 thì mô hình càng tốt. Trong thực tế, mô hình hồi quy với R-square lớn hơn 0,7 được coi là có độ chính xác tốt. Như vậy:

- Đối với Doanh thu: mô hình TensorFlow Neural Network có $R^2 = -0.014 < 0$, cho thấy mô hình không tốt và dự đoán kém hơn việc dự đoán trung bình. Do đó ta không nên sử dụng mô hình này khi dự đoán doanh thu. Trong khi đó, hai mô hình Dicision Tree Regression và Random Forest Regression có R^2 lần lượt là 0.62 và 0.675, chứng tỏ chúng dự đoán doanh thu khá tốt và Random Forest Regression dự đoán doanh thu tốt nhất.
- Đối với Lợi nhuận: Mô hình Random Forest Regression dù khả năng dự đoán lợi nhuận ở mức trung bình ($R^2 = 0.467$) nhưng cho kết quả tốt hơn hai mô hình còn lại. Theo sau đó là TensorFlow Neural Network ($R^2 = 0.3$) và sau cùng là Dicision Tree Regression ($R^2 = 0.297$).

Dưới đây là hình ảnh trực quan cho giá trị mô hình dự đoán và giá trị thực tế ở ba mô hình.



5 Kết luận.

Ý tưởng của đồ án đã được nhóm thực hiện trong Lab 3, và trong đồ án lần này, nhóm chúng em muốn đẩy mạnh ý tưởng đó hơn nữa. Xuất phát từ bài toán thực tế, nhóm đã hoàn thành tương đối tốt vai trò của mình từ vị trí là nhân viên các phòng ban. Nhờ đó, nhóm đã có những phân tích, trực quan rất đặc thù và có tính chuyên môn ở mỗi phòng ban. Điểm nhấn của dashboard là chính là những phân tích, góc nhìn rất thực tế đó. Chúng em cũng đã tham khảo rất nhiều biểu đồ đẹp mắt, sinh động trên Tableau Public, từ đó học hỏi cách phân tích và trực quan dữ liệu. Ý tưởng background thay đổi theo thời gian cũng được tham khảo trên Tableau Public.

Một lần nữa, nhóm chúng em xin gửi lời cảm ơn chân thành nhất đến các thầy cô phụ trách môn **Trực quan hóa dữ liệu**, đặc biệt là thầy Bùi Tiến Lên và thầy Lê Nhựt Nam đã giúp đỡ chúng em rất nhiều trong quá trình thực hiện các lab và đồ án môn học. Các lab và đồ án rất có tính thực tế, giúp chúng em vận dụng tốt những kiến thức được học trên lớp. Trong đó, đồ án cuối kỳ là một dự án thực tế tuyệt vời để chúng em tiếp tục khám phá và phát triển kỹ năng của mình trong lĩnh vực trực quan hóa dữ liệu.

Nhóm 23.

6 Thông tin nhóm và phân công nhiệm vụ.

STT	MSSV	Họ và tên	Email
1	20120030	Nguyễn Thiên An	20120030@student.hcmus.edu.vn
2	20120146	Nguyễn Thị Châu Ngọc	20120146@student.hcmus.edu.vn
3	20120165	Hồng Nhất Phương	20120165@student.hcmus.edu.vn
4	20120209	Nguyễn Nhật Tiến	20120209@student.hcmus.edu.vn
5	20120290	Trần Khánh Hoàng	20120290@student.hcmus.edu.vn

MSSV	Phân công nhiệm vụ
20120030	<ul style="list-style-type: none"> - Tạo dashboard tab Việt Nam. - Thu thập và tiền xử lý dữ liệu. - Dự đoán doanh thu, lợi nhuận superstore. - Viết báo cáo và soạn slide thuyết trình.
20120146	<ul style="list-style-type: none"> - Tạo dashboard tab Marketing. - Khai thác luật kết hợp. - Viết báo cáo và soạn slide thuyết trình.
20120165	<ul style="list-style-type: none"> - Tạo dashboard tab Kinh tế - thị trường. - Viết báo cáo và soạn slide thuyết trình.
20120209	<ul style="list-style-type: none"> - Tạo dashboard tab Chăm sóc khách hàng. - Viết báo cáo và soạn slide thuyết trình.
20120290	<ul style="list-style-type: none"> - Tạo dashboard tab Tài chính - đầu tư. - Viết báo cáo và soạn slide thuyết trình.

7 Đánh giá mức độ hoàn thành.

STT	Tiêu chí	Minh chứng	Tỷ lệ
1	Trực quan dữ liệu	Công việc trực quan hóa dữ liệu được chia thành 5 tab gồm 4 phòng ban và Việt Nam: ở mỗi tab giới thiệu ngắn gọn về nhu cầu (cần thông tin gì ở tập dữ liệu), các câu hỏi đặt ra, biểu đồ trực quan để trả lời câu hỏi đó và các nhận xét dưới mỗi biểu đồ.	60%
2	Thông kê mô tả	Ở mỗi phòng ban đều có các biểu đồ thống kê mô tả đặc trưng của phòng ban đó. Tab Việt Nam có thống kê mô tả nhiều nhất vì cần cung cấp cho người xem cái nhìn tổng quan nhất về tình hình kinh doanh của superstore tại thị trường Việt Nam.	10%
3	Phân tích dữ liệu đơn giản	Ở mỗi phòng ban đều có các biểu đồ phân tích đặc trưng của phòng ban đó (pareto chart ở phòng Marketing, RFM và Cohort analysis ở phòng CSKH...), ngoài ra còn có những phân tích, đề xuất mang tính đặc thù và chuyên môn.	10%
4	Phân tích hồi quy, giải thích và dự đoán	Phân tích hồi quy và dự đoán về doanh thu, lợi nhuận Superstore ở phòng ban Kinh tế - thị trường bằng phân tích tự động của Tableau và ở file <i>forecasting.ipynb</i> bằng các mô hình học máy.	10%
5	Đề xuất giải pháp	Ở mỗi phòng ban đều có nhận xét về các biểu đồ và đưa ra kế hoạch hành động, chiến lược cụ thể cho từng phòng ban.	10%
Tổng			100%

8 Tài liệu tham khảo.

- [1] Official Tableau Desktop document
- [2] *Data Tutorials*, 06/03/2023, Tableau Dashboard from Start to End (Part 2) | Road Accident Dashboard | Beginner to Pro | @tableau
- [3] *James Goodall*, 26/11/2022, Coronavirus (COVID-19) in England
- [4] *Joris van den Berg*, 06/06/2023, #VOTD | SunCycle Sales Dashboard
- [5] *Joris van den Berg*, 16/05/2023, Dashboard that changes through time