

NHẬN XÉT CỦA GIÁO VIÊN HƯỚNG DẪN

Trà Vinh, ngày tháng năm
Giáo viên hướng dẫn
(Ký tên và ghi rõ họ tên)

NHẬN XÉT CỦA THÀNH VIÊN HỘI ĐỒNG

Trà Vinh, ngày tháng năm

Thành viên hội đồng
(Ký tên và ghi rõ họ tên)

LỜI CẢM ƠN

Em xin cảm ơn cô Hà Thị Thúy Vi đã nhiệt tình hướng dẫn em trong quá trình thực hiện đồ án cơ sở ngành “Tìm hiểu và ứng dụng công nghệ Tableau trong phân tích và trực quan hóa dữ liệu”. Em rất trân quý những lời chỉ bảo mà cô đã đưa ra để giúp đỡ em hoàn thành đồ án cơ sở ngành. Trong quá trình thực hiện em vẫn còn nhiều thiếu sót mong cô thông cảm và bỏ qua cho em. Em rất muốn nghe những ý kiến đóng góp của cô.

Chân thành cảm ơn cô!

Trà Vinh, ngày 8 tháng 1 năm 2024

Sinh viên thực hiện

MỤC LỤC

CHƯƠNG 1: TỔNG QUAN.....	9
CHƯƠNG 2: NGHIÊN CỨU LÝ THUYẾT	9
2.1 Giới thiệu phần mềm Tableau.....	10
2.1.1 Những đặc điểm chính của Tableau.....	10
2.1.2 Các khái niệm cơ bản trong Tableau.....	14
2.2 Các hàm và công cụ trong Tableau	18
2.2.1 Các hàm cơ bản trong Tableau.	18
2.2.2 Filters.....	24
2.2.3 Sort	25
2.2.4 Group.	27
2.2.5 LOD (Level of Detail Expressions).....	29
CHƯƠNG 3: HIỆN THỰC HÓA NGHIÊN CỨU	31
3.1 Cài đặt phần mềm.....	31
3.2 Thu thập dữ liệu.....	32
3.3 Kết nối dữ liệu.....	34
3.4 Cài đặt bài toán.	35
3.4.1 Mô tả tập dữ liệu “sales_06_FY2020-21.csv”	35
3.4.2 .Trả lời câu hỏi”Cho biết doanh thu từng tiểu bang”	38
3.4.3 Tạo worksheet tính doanh thu theo tháng trong năm.	39
3.4.4 Tính doanh thu theo tuổi.	40
3.4.5 Tương quan giữa số lượng và tỷ lệ chiết khấu.....	42
3.4.6 Tính phần trăm doanh thu mỗi khu vực.	43
3.4.7 Doanh thu mỗi danh mục sản phẩm theo giới tính.	44
3.4.8 Tổng doanh thu trung bình mỗi tiểu bang.	44
3.4.9 Tần suất đặt hàng.....	45
3.4.10 Xây dựng bảng Dashboard (Bảng điều khiển)	46
CHƯƠNG 4: KẾT QUẢ NGHIÊN CỨU	49
4.1 Kết quả đạt được sau khi tìm hiểu về phần mềm Tableau.	49
4.2 Tạo bộ lọc Slicing Filters.	52
CHƯƠNG 5: KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN	53
DANH MỤC TÀI LIỆU THAM KHẢO	54
PHỤ LỤC	Error! Bookmark not defined.

DANH MỤC HÌNH ẢNH – BẢNG BIỂU

Hình 2.1.1. Dashboard	10
Hình 2.1.2. Biểu đồ đường.....	12
Hình 2.1.3. Bản đồ.....	13
Hình 2.1.4. Tạo Worksheet	14
Hình 2.1.5. Tạo Dashboard	15
Hình 2.1.6. Trường Measure Values.....	16
Hình 2.1.7. Trường Measure Names.....	16
Hình 2.1.8.Chiều dữ liệu liên tục và chiều dữ liệu rời rạc.....	17
Hình 2.2.1. Chiều dữ liệu liên tục và chiều dữ liệu rời rạc.....	18
Hình 2.2.2. Mở hộp thoại Add Parameter.....	19
Hình 2.2.3. Chọn Reference Line để tạo đường tham chiếu.....	20
Hình 2.2.4. Kéo chọn các ô trong Worksheet	20
Hình 2.2.5. Kéo chọn các ô trong Worksheet	21
Hình 2.2.6. Dùng hàm IF để tính tổng doanh thu của từng mặt hàng.....	22
Hình 2.2.7. Tổng doanh thu theo giới tính.....	23
Hình 2.2.8. Lọc dữ liệu theo phân loại.	24
Hình 2.2.9. Lọc dữ liệu theo định lượng.....	25
Hình 2.2.10. Sắp xếp theo một trục.	25
Hình 2.2.11. Sắp xếp theo tiêu đề.	26
Hình 2.2.12. Sắp xếp theo nhãn.....	26
Hình 2.2.13. Tạo một nhóm dữ liệu.....	27
Hình 2.2.14. Chọn dữ liệu muôn nhóm.	28
Hình 2.2.15. Chọn dữ liệu muôn nhóm.	28
Hình 2.2.16. Sử dụng hàm FIXED để tính tổng doanh thu từng vùng.	29
Hình 2.2.17. Sử dụng hàm FIXED để tính tổng doanh thu từng vùng.	30
Hình 2.2.18. Sử dụng hàm EXCLUDE để tính giá cả của từng khu vực theo tháng.....	30
Hình 3.1.1. Trang chủ Tableau.	31
Hình 3.1.2. Điện thông tin cá nhân.	31
Hình 3.1.3. Tải phần mềm Tableau.	32
Hình 3.2.1. Trang chủ Kaggle.....	32
Hình 3.2.2. Tập dữ liệu “sales_06_FY2020-21.csv”	33
Hình 3.3.1. Join dữ liệu.....	34
Hình 3.3.2. Tạo trang Worksheet mới.	34
Hình 3.3.3. Giao diện màn hình làm việc của Tableau.....	35
Hình 3.4.1. Số liệu doanh thu từng tiểu bang.	38
Hình 3.4.2. Bản đồ thể hiện doanh thu từng tiểu bang.	38
Hình 3.4.3. Chuyển kiểu dữ liệu trường Month.....	39
Hình 3.4.4. Doanh thu theo tháng hàng năm.	39
Hình 3.4.5. Doanh thu theo tháng hàng năm.	40
Hình 3.4.6. Mở hộp thoại Edit Aliases.	40
Hình 3.4.7. Thay đổi giá trị độ tuổi.	41
Hình 3.4.8. Biểu đồ thể hiện doanh thu theo tuổi.	41
Hình 3.4.9. Chuyển đổi kiểu dữ liệu.....	42
Hình 3.4.10. Biểu đồ bong bóng thể hiện tương quan giữa số lượng và tỉ lệ chiết khấu.	42
Hình 3.4.11. Chuyển biểu đồ cột thành biểu đồ tròn.	43
Hình 3.4.12. Biểu đồ tròn thể hiện phân trăm doanh thu từng khu vực.	43

Hình 3.4.13. Tổng doanh thu theo giới tính.....	44
Hình 3.4.14. Doanh thu trung bình của từng vùng.	44
Hình 3.4.15. Chuyển công thức từ Measure về dạng Dimensions.	45
Hình 3.4.16. Chuyển công thức Measure về dạng COUNTD.	45
Hình 3.4.17. Biểu đồ thể hiện tần suất đặt hàng.	46
Hình 3.4.18. Xây dựng Dashboard.	46
Hình 3.4.19. Cài đặt Slicing Filter.	47
Hình 3.4.20. Chọn kiểu dữ liệu muốn hiển thị.	47
Hình 3.4.21. Tạo bảng thả xuống.....	48
Hình 3.4.22. Bảng Dashboard.....	48
Hình 4.1.1. Kết nối dữ liệu bán hàng.....	49
Hình 4.1.2. Kết nối dữ liệu bán hàng.....	49
Hình 4.1.3. Dashboard (Bảng điều khiển) dữ liệu bán hàng (1)	50
Hình 4.1.4. Dashboard (Bảng điều khiển) dữ liệu bán hàng (2)	50
Hình 4.2.1. Bộ lọc Slicing Filters	52

TÓM TẮT NIÊN LUẬN/ĐỒ ÁN CƠ SỞ NGÀNH/ĐỒ ÁN CHUYÊN NGÀNH

Vấn đề nghiên cứu:

- Phân tích và trực quan hóa dữ liệu .
- Tìm hiểu các hàm phân tích dữ liệu.
- Tạo các biểu đồ và trực quan hóa dữ liệu.
- Tạo Dashboard.

Hướng tiếp cận:

- Tìm hiểu các phần mềm trực quan hóa dữ liệu.
- Tìm hiểu qua các tài liệu liên quan đến trực quan hóa dữ liệu.
- Tham khảo tài liệu điện tử.

Cách giải quyết vấn đề:

- Sử dụng phần mềm Tableau để xây dựng các hàm phân tích và trực quan hóa dữ liệu.
- Kết nối tập dữ liệu.
- Tạo biểu đồ so sánh tập dữ liệu.
- Xây dựng Dashboard.

Các kết quả đạt :

- Tạo được biểu đồ so sánh dữ liệu.
- Xây dựng hàm phân tích và trực quan hóa dữ liệu.
- Xây dựng Dashboard điều khiển.

MỞ ĐẦU

1. Lý do chọn đề tài:

Dữ liệu ngày nay đóng vai trò quan trọng và ngày càng trở nên không thể thiếu trong nhiều lĩnh vực của cuộc sống và kinh doanh. Trong các công ty, các tổ chức các nhà lãnh đạo muốn xem xét và có cái nhìn tổng thể về tình hình kinh doanh của doanh nghiệp để có những giải pháp phát triển kịp thời.

Phân tích dữ liệu trực quan có thể được coi là việc "kể câu chuyện về dữ liệu." Thay vì chỉ sử dụng số liệu và thống kê khô khan, trực quan hóa dữ liệu giúp tạo ra một câu chuyện số hóa, biểu diễn thông tin một cách sinh động và dễ hiểu. Có rất nhiều công cụ giúp người dùng phân tích và trực quan dữ liệu hiện nay như Excel, PowerBI, ... Tableau là công cụ phân tích và trực quan hóa dữ liệu với giao diện dễ sử dụng thực hiện các công việc liên quan đến phân tích dữ liệu một cách dễ dàng. Do đó, em chọn đề tài “Tìm hiểu và ứng dụng công nghệ Tableau trong phân tích và trực quan hóa dữ liệu”.

2. Mục đích:

Tìm hiểu và ứng dụng công cụ Tableau để phân tích và trực quan hóa dữ liệu dữ liệu bán hàng.

3. Đối tượng nghiên cứu:

a. **Nghiên cứu lý thuyết:** Tìm hiểu về công cụ Tableau.

b. **Nghiên cứu thực nghiệm:** Phân tích và trực quan hóa tập dữ liệu.

4. Phạm vi nghiên cứu:

Tạo biểu đồ trực quan hóa dữ liệu, tạo Dashboard.

CHƯƠNG 1: TỔNG QUAN

Tableau là một **phần mềm BI** (Business Intelligence) và thông kê đồ họa tổng quan, cho phép người dùng tạo, chia sẻ và xem những bảng điều khiển, biểu đồ và dashboard trực quan từ dữ liệu. Tableau cung cấp các công cụ đồ họa động, giúp người dùng nhanh chóng trực quan hóa dữ liệu và trả lời những câu hỏi kinh doanh quan trọng.

✓ **Biểu đồ và Dashboard:**

Tableau cho phép người dùng tạo ra các biểu đồ, đồ thị, và Dashboard tương tác thông qua dữ liệu đã chọn.

Hỗ trợ nhiều loại biểu đồ như cột, dòng, bong bóng, đường, đồ thị, heatmap, và nhiều loại biểu đồ khác.

✓ **Kết nối dữ liệu đa dạng:**

Tableau có khả năng kết nối đến nhiều nguồn dữ liệu khác nhau như cơ sở dữ liệu SQL, Excel, các dịch vụ đám mây như Azure, Amazon Service,... và nhiều nguồn dữ liệu khác.

✓ **Kết hợp dữ liệu:**

Cung cấp khả năng kết hợp và lọc dữ liệu từ nhiều nguồn khác nhau để tạo ra bảng dữ liệu có thể sử dụng trong các biểu đồ và Dashboard.

✓ **Tính năng tương tác:**

Tableau hỗ trợ tương tác trực quan giữa các yếu tố khác nhau trong biểu đồ. Người dùng có thể tương tác với dữ liệu bằng cách di chuyển chuột, lọc dữ liệu, và thậm chí là tạo các truy vấn động.

✓ **Theo dõi dữ liệu :**

Cung cấp khả năng tạo các Dashboard và biểu đồ để theo dõi sự thay đổi của dữ liệu trong tương lai.

Tableau có thể kết hợp với các công cụ phân tích thống kê và ngôn ngữ lập trình như Python để mở rộng khả năng phân tích.

CHƯƠNG 2: NGHIÊN CỨU LÝ THUYẾT

2.1 Giới thiệu phần mềm Tableau

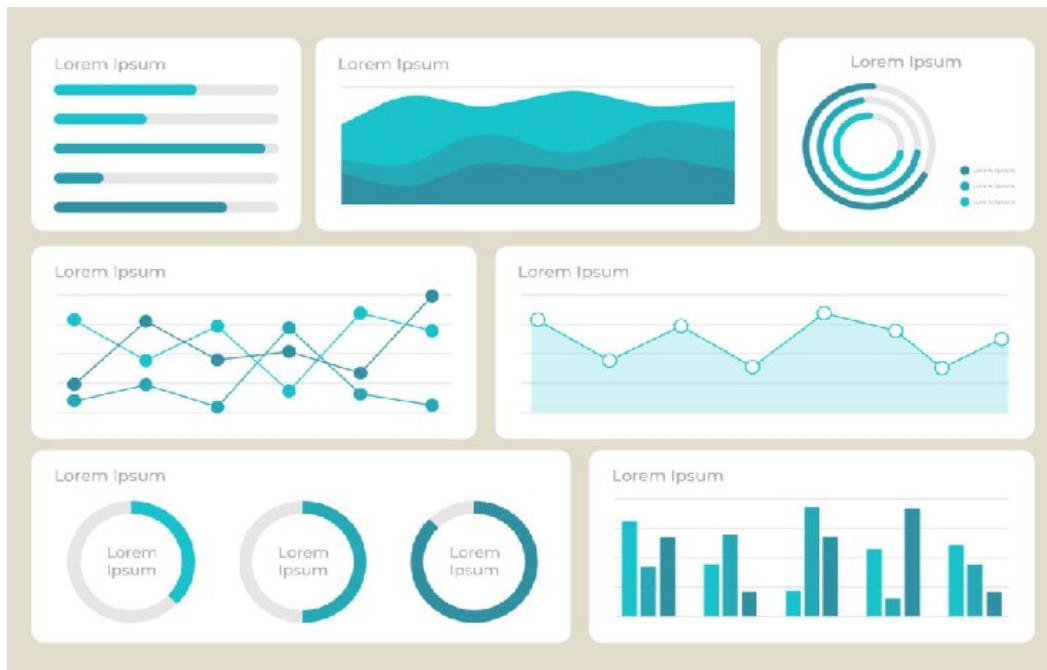
Tableau là phần mềm hỗ trợ phân tích (Data Analyst) và trực quan hóa dữ liệu (Data Visualization), công cụ được dùng nhiều trong ngành BI (Business Intelligence).

Tableau giúp tổng hợp các dữ liệu này từ các dãy số thành những hình ảnh, biểu đồ trực quan, xây dựng các dashboard và các phân tích (Self-services).

2.1.1 Những đặc điểm chính của Tableau.

❖ Tạo Dashboard:

Dashboard cung cấp một cái nhìn toàn diện về dữ liệu doanh nghiệp bằng cách công cụ trực quan hóa, đối tượng trực quan, văn bản, v.v... Dashboard có thể cung cấp nhiều thông tin hữu dụng nhờ tính năng biểu diễn dữ liệu theo định dạng trình tự thời gian, cho phép bổ sung nhiều chế độ xem và đối tượng, cung cấp nhiều loại bộ lọc và định dạng, cho phép doanh nghiệp triển khai các bộ lọc phù hợp. Doanh nghiệp thậm chí còn có thể sao chép một dashboard hay thành phần cụ thể nào đó từ cửa sổ làm việc này sang cửa sổ khác dễ dàng.



Hình 2.1.1. Dashboard.

❖ **Cộng tác và chia sẻ:**

Tableau cung cấp những tiện ích cho phép người dùng cộng tác và chia sẻ dữ liệu tức thì theo định dạng trực quan hóa, bảng tính, dashboard,v..v... tại thời gian thực. Công cụ này cho phép doanh nghiệp chia sẻ dữ liệu an toàn từ nhiều nguồn dữ liệu như on-premise, cloud hoặc hybrid,v..v... Việc nhanh chóng và dễ dàng cộng tác cũng như chia sẻ dữ liệu giúp nhân viên nhận được đánh giá và phản hồi về dữ liệu nhanh hơn, đồng thời cải thiện chất lượng hoạt động phân tích tổng thể.

❖ **Dữ liệu trực tuyến và dữ liệu lưu trữ:**

Tableau đảm bảo kết nối doanh nghiệp đến cả những nguồn dữ liệu trực tuyến hoặc trích xuất dữ liệu từ các nguồn bên ngoài như một dạng dữ liệu lưu trữ (in-memory data). Điều này mang đến cho doanh nghiệp tính linh hoạt khi sử dụng dữ liệu từ nhiều hơn một nguồn và không gặp phải bất kỳ rào cản nào.

Doanh nghiệp có thể sử dụng dữ liệu trực tiếp từ các nguồn bằng cách thiết lập kết nối đến dữ liệu trực tuyến hoặc lưu trữ dữ liệu trong bộ nhớ bằng cách trích xuất dữ liệu từ một nguồn nào đó theo yêu cầu. Tableau cung cấp các tính năng hỗ trợ để hỗ trợ các kết nối dữ liệu như làm mới trích xuất tự động, thông báo đến người dùng nếu kết nối trực tuyến không thành công,v..v...

❖ **Tùy chọn sử dụng nguồn dữ liệu**

Tableau cung cấp vô số tùy chọn nguồn dữ liệu mà doanh nghiệp có thể kết nối và trích xuất dữ liệu. Các nguồn dữ liệu trải dài từ các file dữ liệu cố định, bảng tính, các cơ sở dữ liệu, big data cho đến các dữ liệu lưu trữ trên không gian đám mây đều sẵn có ở Tableau. Doanh nghiệp có thể dễ dàng thiết lập các liên kết an toàn đến bất kỳ nguồn dữ liệu nào từ Tableau và sử dụng song song với dữ liệu từ nhiều nguồn khác để tạo nên một cái nhìn tổng hợp về dữ liệu một cách trực quan hóa. Tableau cũng hỗ trợ nhiều công cụ kết nối dữ liệu khác như Salesforce, Google Analytics, Google Sheets,...

❖ Trực quan hóa nâng cao (Dạng biểu đồ)

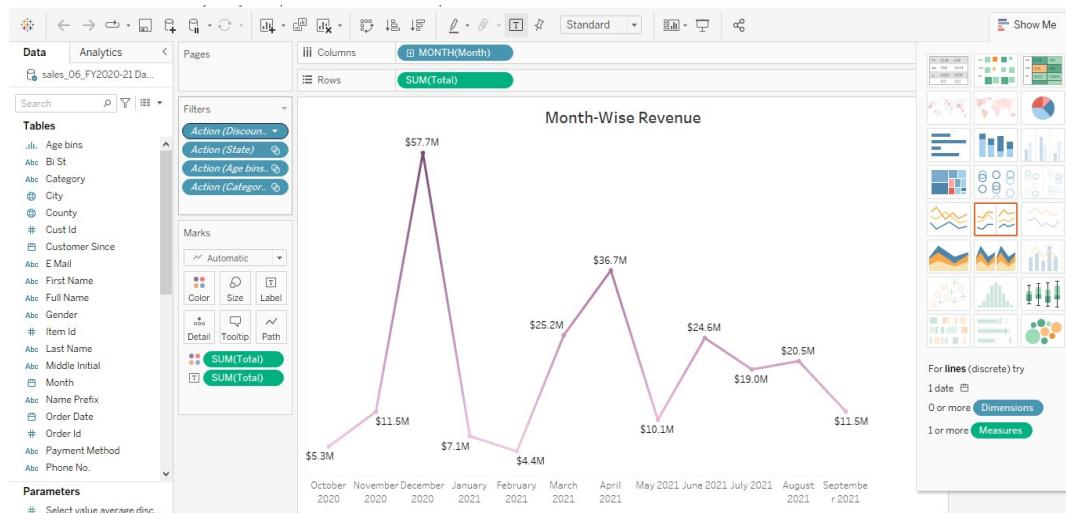
Một trong những tính năng nổi bật của Tableau là khả năng hiển thị dữ liệu đa dạng tới hơn 38000 điểm. Với Tableau, doanh nghiệp có thể biểu diễn dữ liệu từ đơn giản với:

- Biểu đồ cột
- Biểu đồ tròn

Các biểu đồ phức tạp như:

- Biểu đồ tần suất
- Biểu đồ Gantt
- Biểu đồ hình đạn (Bullet chart)
- Biểu đồ chuyển động
- Biểu đồ Treemap
- Biểu đồ hộp
- V..V...

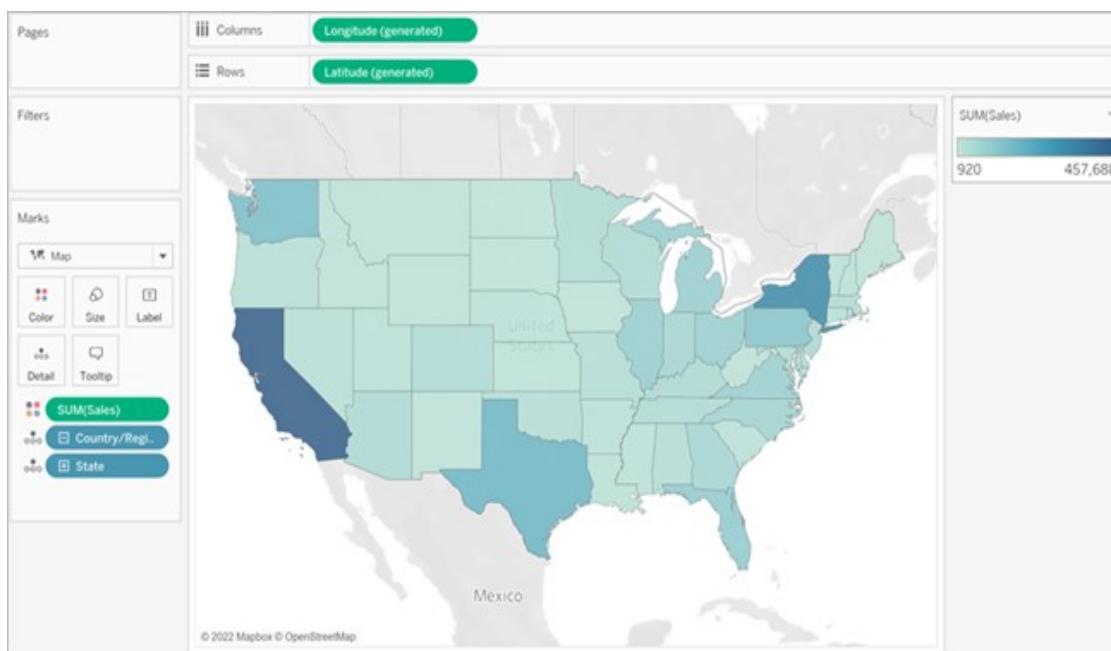
Doanh nghiệp có thể dễ dàng lựa chọn và tạo lập bất kỳ phương thức biểu diễn nào bằng cách nhấn chọn định dạng biểu diễn dữ liệu trên thanh công cụ Show Me.



Hình 2.1.2. Biểu đồ đường.

❖ Bản đồ

Một tính năng nổi bật khác của Tableau chính là bản đồ. Tableau tích hợp sẵn nhiều thông tin trên bản đồ như thành phố, mã bưu chính, ranh giới hành chính,... Điều này khiến cho các bản đồ được tạo lập trên Tableau thường rất chi tiết và đầy đủ thông tin. Doanh nghiệp có thể bổ sung các tầng địa chất khác nhau lên bản đồ theo yêu cầu và tạo lập những bản đồ hữu ích trên Tableau bằng kho dữ liệu. Nhiều loại bản đồ khác nhau sẵn có trên Tableau như Bản đồ nhiệt, Bản đồ phân luồng, Bản đồ chuyên biệt, Bản đồ phân bố điểm, v..v..



Hình 2.1.3. Bản đồ.

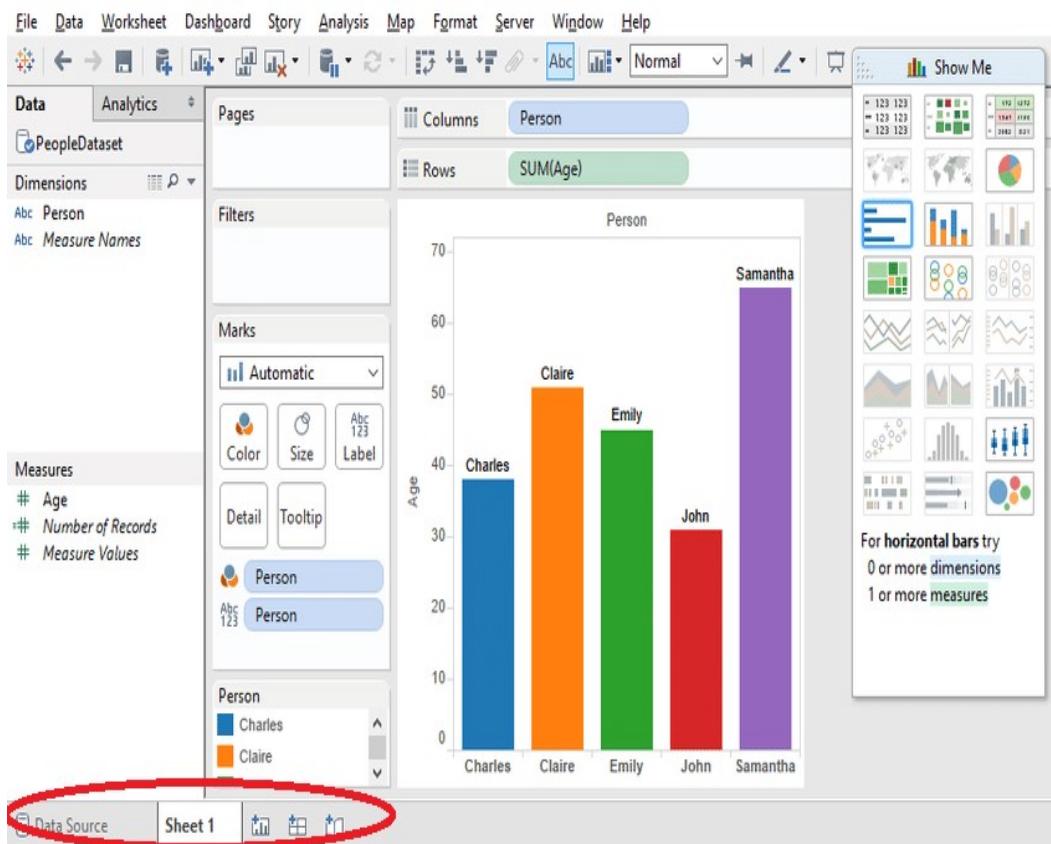
❖ Bảo mật mạnh mẽ

Tableau rất chú trọng đến bảo mật dữ liệu và người dùng. Tableau sở hữu một hệ thống bảo mật chống lừa đảo dựa trên một hệ thống được xác thực và cấp phép. Tableau còn cho phép doanh nghiệp tự do tích hợp các giao thức bảo mật khác như Active Directory, Kerberos, v..v... Một điểm quan trọng cần lưu ý đó là Tableau sẽ tiến hành sàng lọc dữ liệu theo hàng (row-level filtering) nhằm phục vụ cho bảo mật dữ liệu.

2.1.2 Các khái niệm cơ bản trong Tableau.

❖ Worksheet:

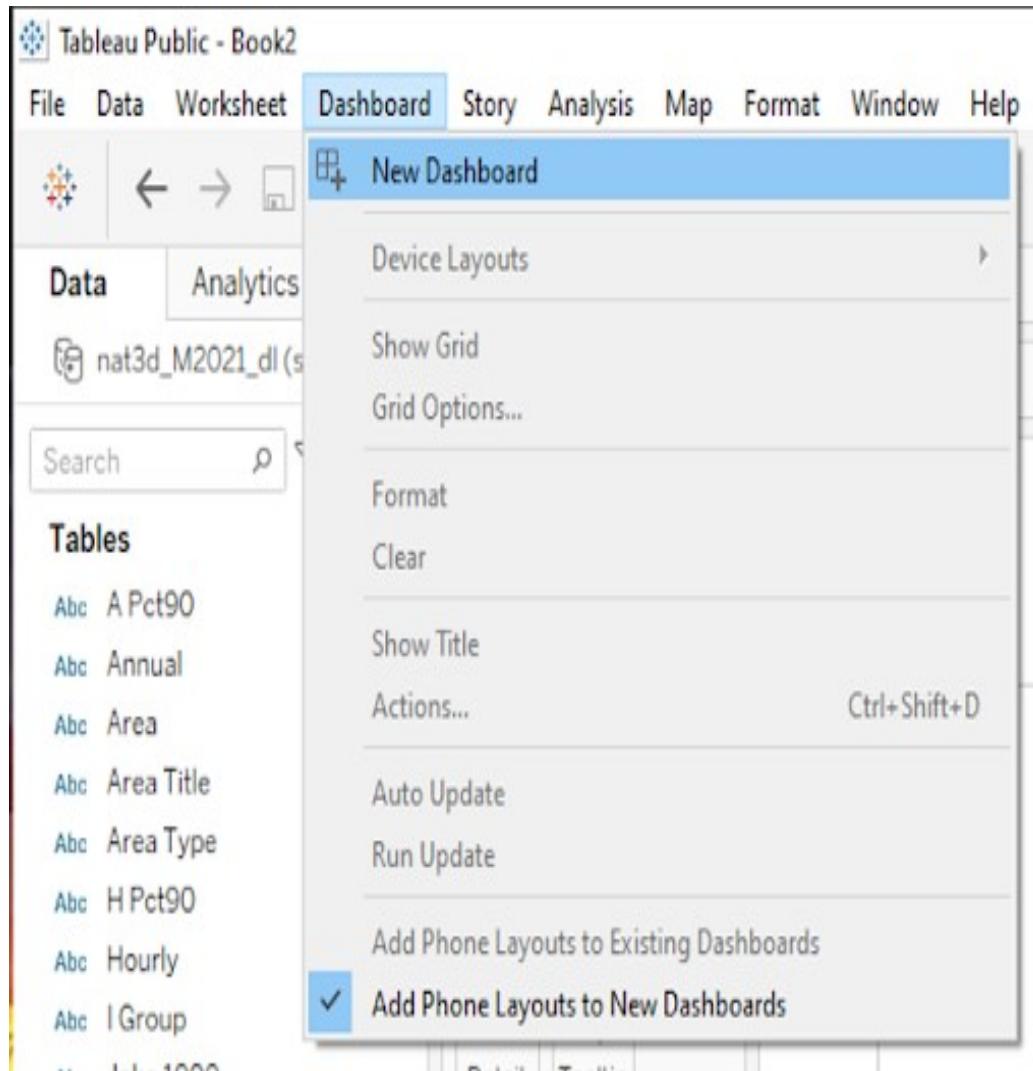
Worksheet là không gian làm việc chính trong Tableau, nơi bạn thiết kế và xây dựng biểu đồ. Worksheet chứa một view đơn với các shelves, cards, legends, và Data, Analytics Pane ở thanh side bar (bên trái).



Hình 2.1.4. Tạo Worksheet.

❖ Dashboard:

Dashboard là tập hợp của các views từ nhiều worksheets khác nhau, cho phép so sánh nhiều views cùng một lúc. Để tạo một Dashboard mới, hãy nhấp vào nút **Dashboard** trên thanh Menu ở đầu màn hình. Thao tác này sẽ hiển thị một cột thả xuống. Chọn **New Dashboard**. Bạn có thể đổi tên theo mong muốn.



Hình 2.1.5. Tạo Dashboard.

❖ **Measures:**

Measures chứa các con số, giá trị định tính mà bạn có thể đo. Measures có thể được tổng hợp, khi bạn kéo một measure vào view, Tableau sẽ áp dụng một tổng hợp cho nó theo mặc định.

Measure Values sẽ chứa tất cả các Measure trong dữ liệu của bạn, thu thập vào một trường duy nhất với các giá trị liên tục. Có thể kéo các trường measure ra khỏi thẻ Measure Values để xóa chúng khỏi view.

The screenshot shows the Tableau interface with the following components:

- Top Bar:** Shows 'Pages', 'Columns', and 'Rows'.
- Filters:** Shows 'Measure Names'.
- Marks:** Shows 'Automatic' and 'Color'.
- Sheet 1:** Title 'Sheet 1' with 'Abc' placeholder.
- Left Shelf:** Labeled 'Khung Measure Values' (Measure Values Frame). It contains a list of measures:
 - SUM(Age)
 - CNT(sales_06_FY20..)
 - SUM(Discount Amo..)
 - SUM(Discount Perc..)
 - SUM(Price)
 - SUM(Qty Ordered)
 - SUM(Total)
 - SUM(Value)
- Right Shelf:** A grid of visualization icons. One icon, a horizontal bar chart, is highlighted with a red box. Below it, text says 'For horizontal bars try 0 or more Dimensions 1 or more Measures'.
- Bottom Right:** 'Activate Windows' button.

Hình 2.1.6. Trường Measure Values.

- **Measure Names** là trường chứa tên của tất cả measure trong dữ liệu của bạn, thu thập vào một trường duy nhất với các giá trị rác.

The screenshot shows the Tableau interface with the following components:

- Top Bar:** Shows 'Pages', 'Columns', and 'Rows'.
- Filters:**
- Marks:** Shows 'Automatic' and 'Color'.
- Sheet 1:** Title 'Sheet 1' with 'Region' and 'Category' dimensions.
- Left Shelf:** Shows 'Category' and 'Region' dimensions, and 'Measure Names'.
- Data View:** A table showing data for three categories (Appliances, Beauty & Grooming, Books) across four regions (Midwest, Northe.., South, West). The data is mostly 'Abc' (rubbish value).
- Right Shelf:** A grid of visualization icons. One icon, a horizontal bar chart, is highlighted with a red box. Below it, text says 'For horizontal bars try 0 or more Dimensions 1 or more Measures'.
- Bottom Right:** 'Activate Windows' button.

Hình 2.1.7. Trường Measure Names.

❖ **Dimensions:**

Dimensions chứa giá trị định tính như tên, ngày tháng, dữ liệu địa lý. Bạn có thể sử dụng dimensions để phân loại, phân chia và tiết lộ chi tiết về dữ liệu của bạn. Dimensions sẽ ảnh hưởng đến mức độ chi tiết bên trong view.

Product Name

Dimension rời rạc.

YEAR(Order Date)

Dimensions liên tục.



Hình 2.1.8. Chiều dữ liệu liên tục và chiều dữ liệu rời rạc.

2.2 Các hàm và công cụ trong Tableau

2.2.1 Các hàm cơ bản trong Tableau.

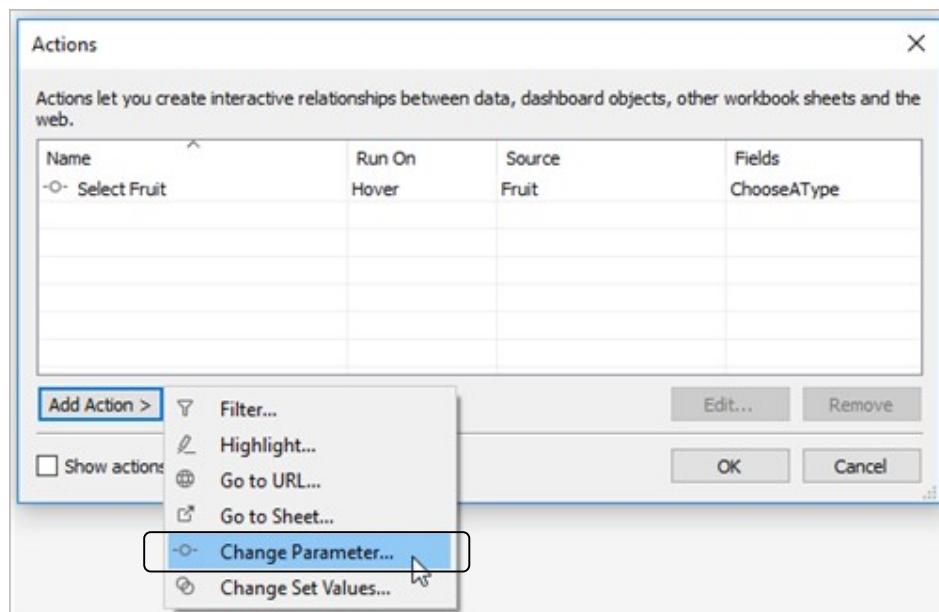
❖ Parameter:

Parameter trong Tableau là các biến số làm việc như số, ngày tháng hoặc chuỗi có thể thay thế giá trị không đổi trong phép tính, bộ lọc hoặc dòng tham chiếu.

- **Ví dụ: Tạo một Parameter Action.**

Bước 1: Trong một worksheet, chọn **Worksheet>Action**.

Bước 2: Trong hộp thoại Actions, nhấn **Add Action** và sau đó chọn **Change Parameter**.



Hình 2.2.1. Chiều dữ liệu liên tục và chiều dữ liệu rời rạc.

Bước 3: Trong hộp thoại Actions, đặt một cái tên để bạn có thể nhận biết.

Bước 4: Chọn một source sheet hoặc data source. Sheet hiện tại được chọn theo mặc định. Nếu bạn chọn một data source hoặc dashboard bạn có thể chọn từng sheet bên trong.

Bước 5: Chọn cách mà người dùng sẽ chạy action.

- **Hover** – Chạy khi một người dùng rê chuột qua một mark trong view. Cũng hiển thị một tùy chọn cho action trong tooltip của mark.

- **Select** – Chạy khi một người dùng nhấp vào một mark hoặc chọn nhiều marks trong view.
- **Menu** – Chạy khi người dùng nhấp chuột phải (control click trên Mac) vào một mark trong view, sau đó chọn một tùy chọn trên menu ngữ cảnh. Cũng hiển thị một tùy chọn cho action trong tooltip của mark.

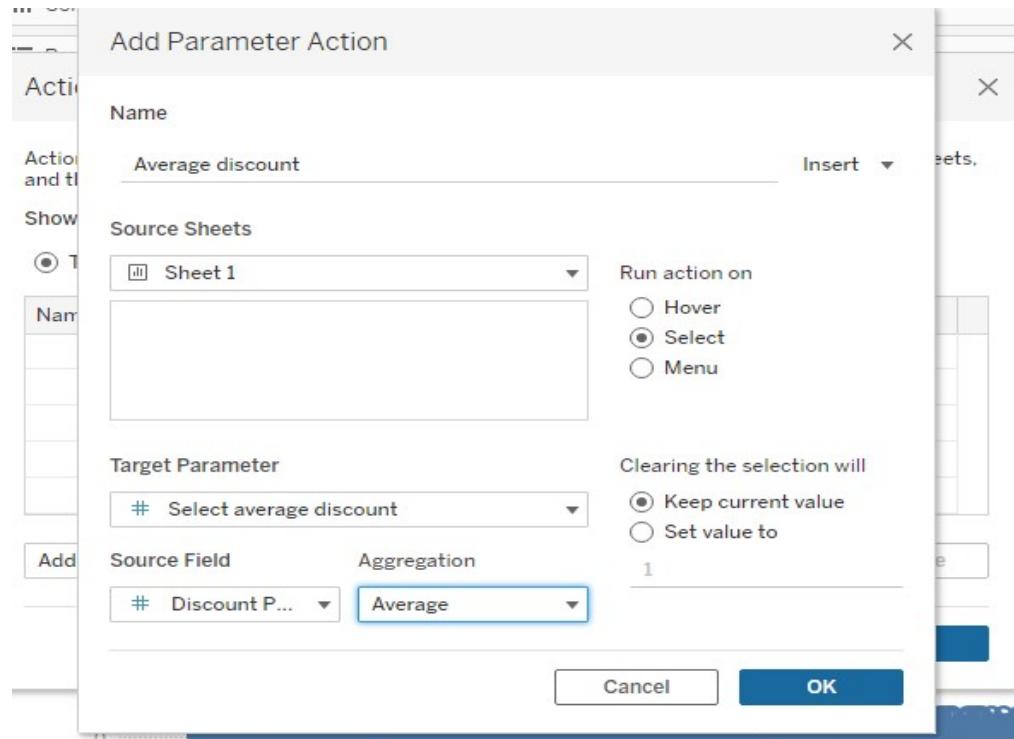
Bước 6: Bên dưới **Target**, chỉ định target parameter và một source field. Tất cả parameter trong workbook có sẵn trong danh sách **Parameter**.

Bước 7: Chỉ định điều gì sẽ xảy ra khi lựa chọn bị xóa

- **Keep current value** – Giá trị parameter hiện tại sẽ vẫn còn trong view.
- **Set value to** – Parameter value sẽ hoàn nguyên về giá trị được chỉ định.

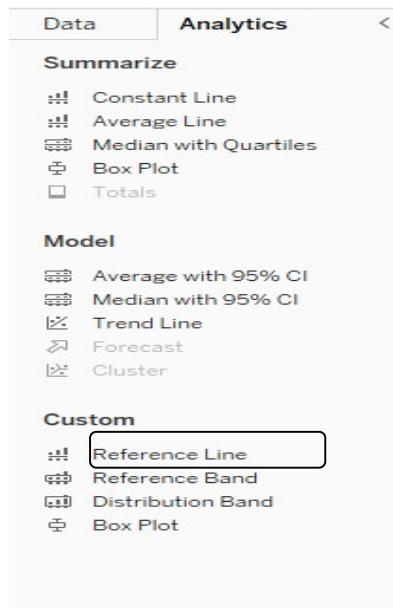
Bước 8: Nhấn **OK** để lưu các thay đổi của bạn và trở về view.

Bước 9: Kiểm tra parameter action bằng cách tương tác với trực quan. Chính sửa một số cài đặt cho hành động để điều chỉnh hành vi khi cần.



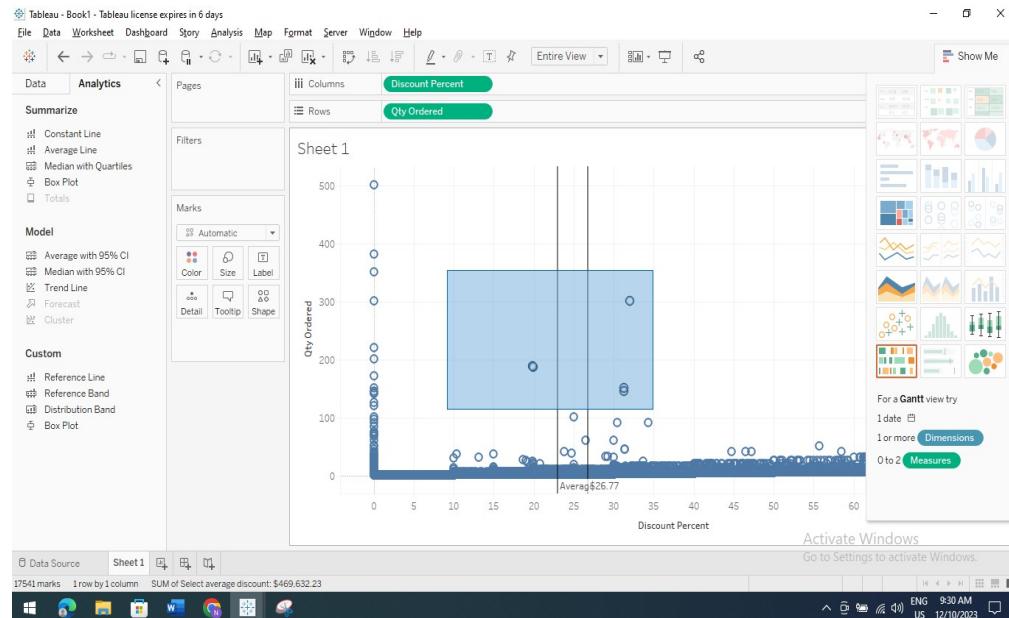
Hình 2.2.2. Mở hộp thoại Add Parameter.

Chọn mục Analytics, kéo thả mục Reference Line vào Worksheet. Thả mục Reference Line vào Discount Percent của Table.



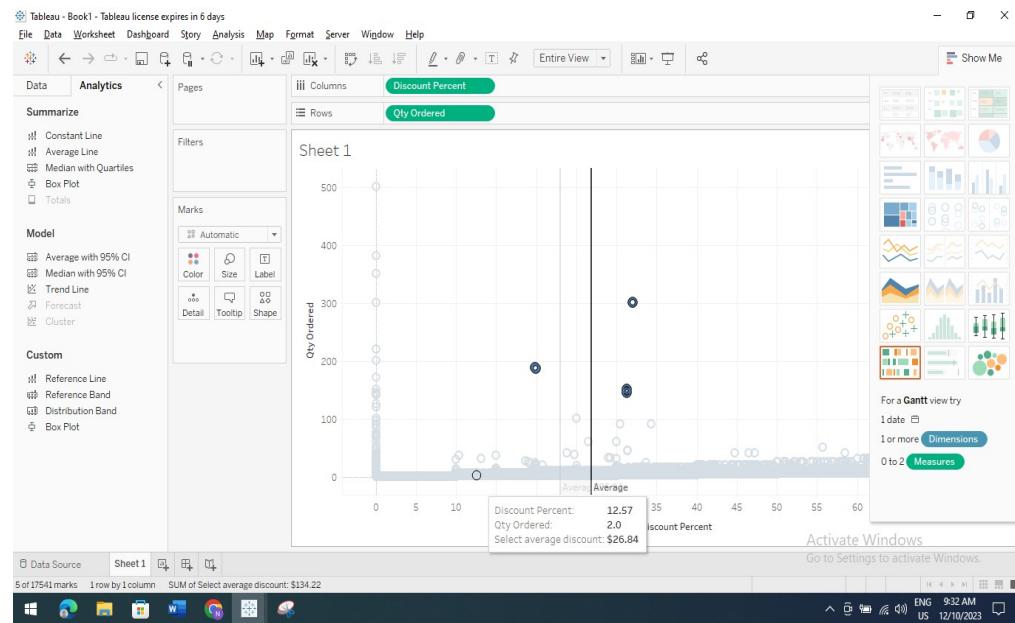
Hình 2.2.3. Chọn Reference Line để tạo đường tham chiếu.

Hộp thoại Edit Reference xuất hiện, sau khi chọn các mục trong hộp thoại nhấn nút OK.



Hình 2.2.4. Kéo chọn các ô trong Worksheet.

Khi kéo chọn các ô trong worksheet nó sẽ giữ nguyên phần đã chọn và tính ra giá trị trung bình của ô đó.



Hình 2.2.5. Kéo chọn các ô trong Worksheet.

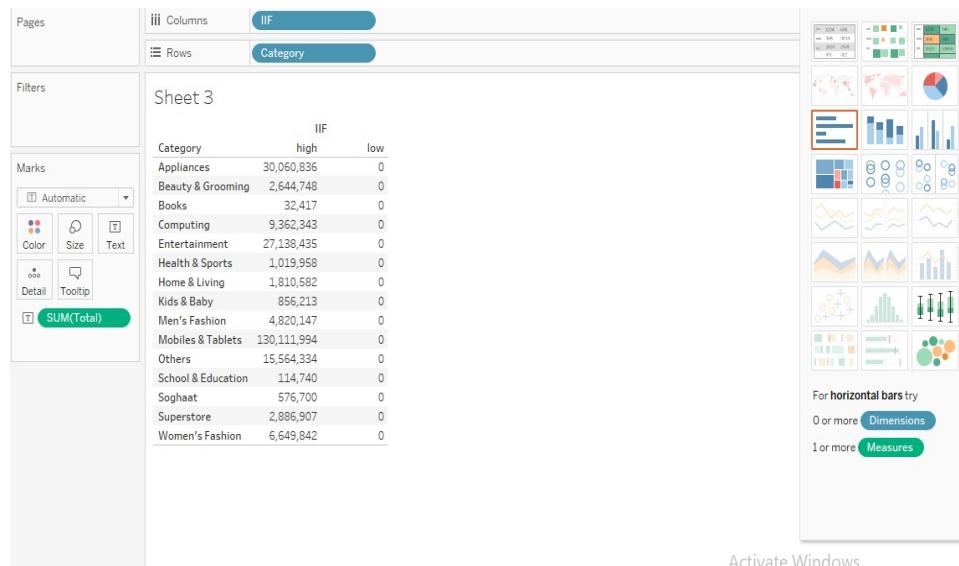
❖ Hàm IIF

Hàm IIF là phiên bản đơn giản của hàm If Else. Nếu cả hai điều kiện là đúng, thì nó sẽ trả về Câu lệnh đầu tiên, ngược lại, nó sẽ trả về câu lệnh thứ hai.

Cú pháp:

IIF(<test>,<then>,<else>,[<unknown>])

Kiểm tra xem một điều kiện có được đáp ứng hay không (<test>) và trả về <then>nếu kiểm tra là đúng, <else>nếu kiểm tra là sai và một giá trị tùy chọn <unknown>nếu kiểm tra là null. Nếu ẩn số tùy chọn không được chỉ định, IIF trả về null.



Hình 2.2.6. Dùng hàm IF để tính tổng doanh thu của từng mặt hàng.

Kiểm tra một loạt biểu thức và trả về <then>giá trị cho giá trị true đầu tiên <test>.

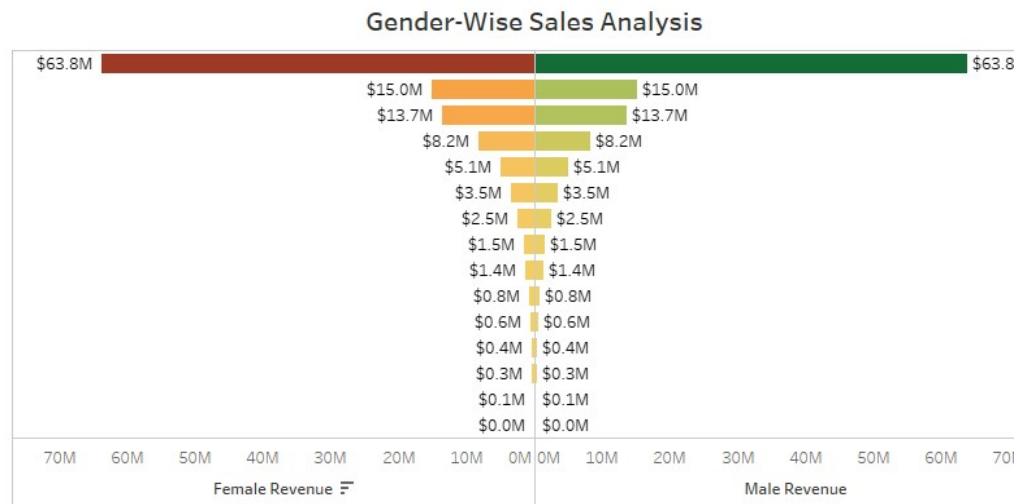
Cú pháp:

IF <test1> THEN <then1>

[ELSEIF <test2> THEN <then2>...]

[ELSE <default>]

END



Hình 2.2.7. Tổng doanh thu theo giới tính.

2.2.2 Filters.

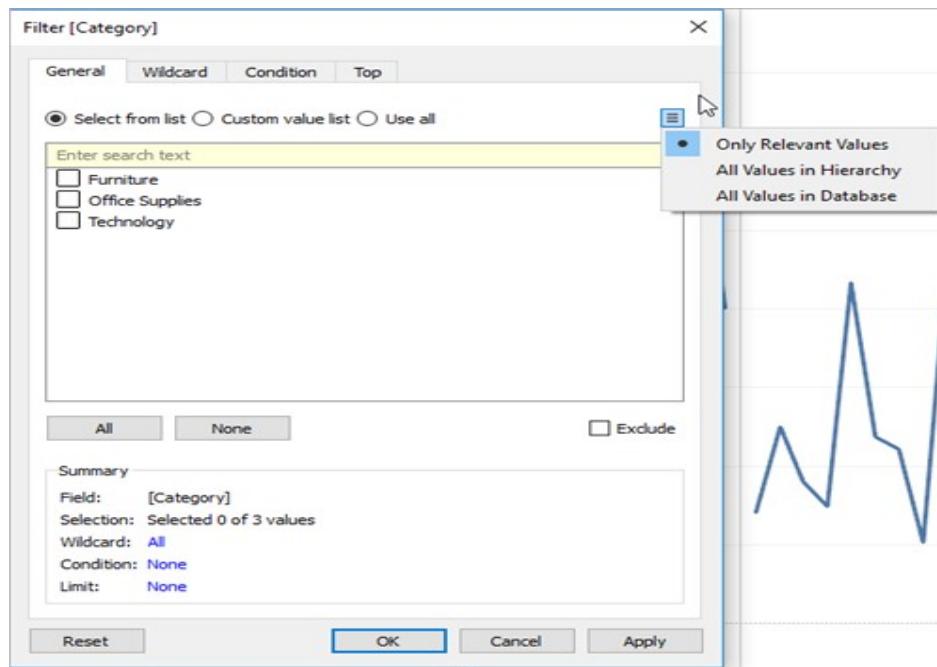
Tableau Filters được dùng để thực hiện điều đó, chúng đơn giản và lọc dữ liệu không liên quan, nhằm phân tích dữ liệu một cách tối ưu.

Tableau thực hiện các hành động trên view trong một trật tự cụ thể, được gọi là Order of Operations (trật tự hoạt động). Bộ lọc được thực thi theo thứ tự sau:

- Extract filters
- Data source filters
- Context filters
- Filters trên dimensions (bất kỳ đâu trên ngăn Filters hoặc các thẻ filter trong view)
 - Filters trên measures (bất kỳ đâu trên ngăn Filters hoặc thẻ các filter trong view)

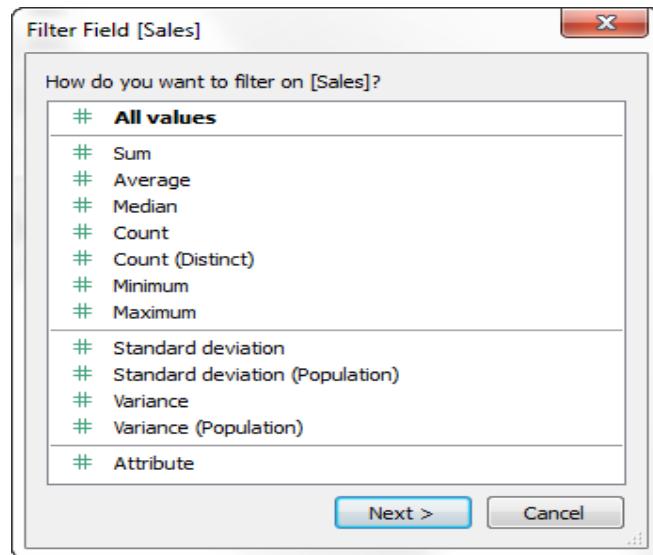
Ví dụ:

- Lọc dữ liệu phân loại.



Hình 2.2.8. Lọc dữ liệu theo phân loại.

- **Lọc dữ liệu định lượng.**

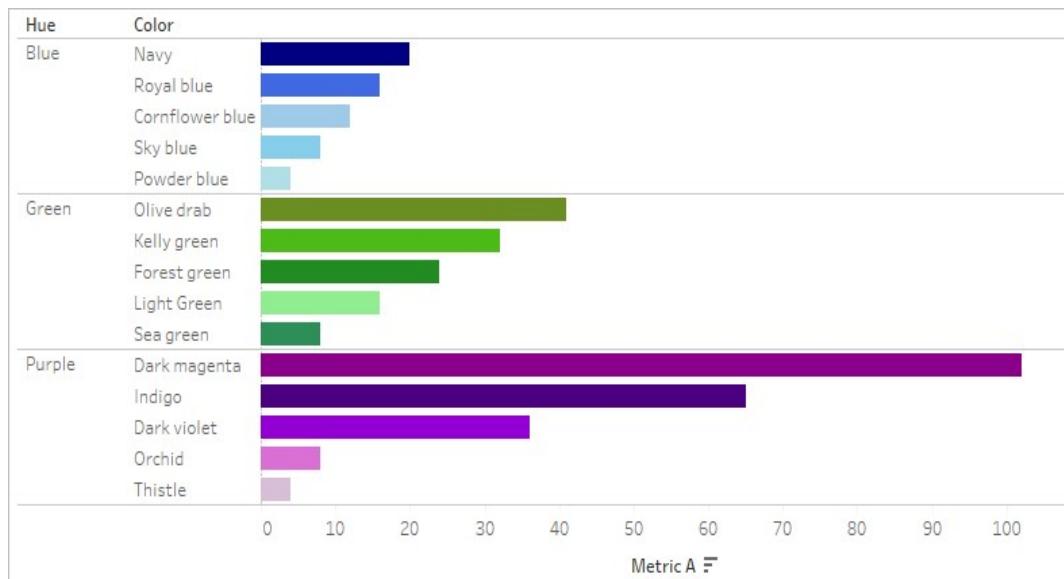


Hình 2.2.9. Lọc dữ liệu theo định lượng.

2.2.3 Sort

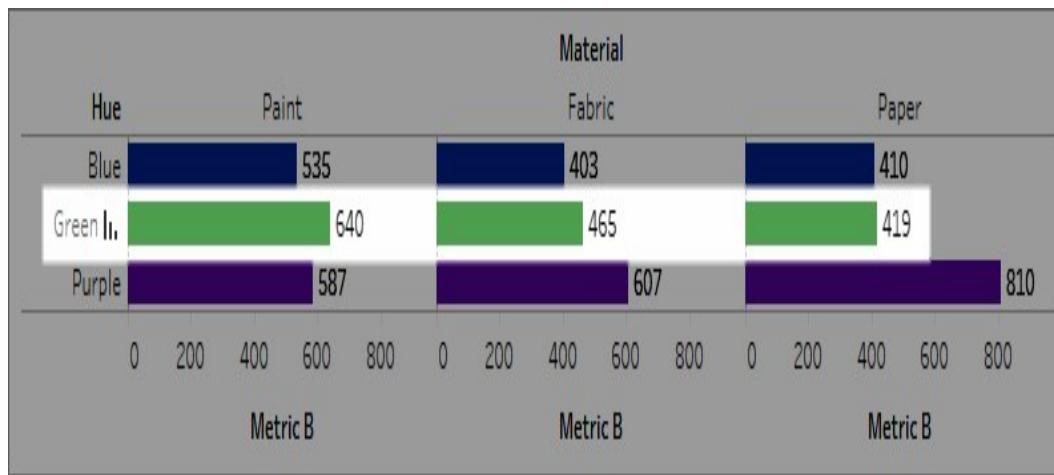
Sort được sử dụng để sắp xếp dữ liệu trong các trường dữ liệu. Có thể sắp xếp dữ liệu theo thứ tự tăng dần hoặc giảm dần dựa trên giá trị của trường cù the. Sắp xếp là một phần quan trọng của việc hiển thị và phân tích dữ liệu trong Tableau.

- **Sắp xếp từ một trục.**



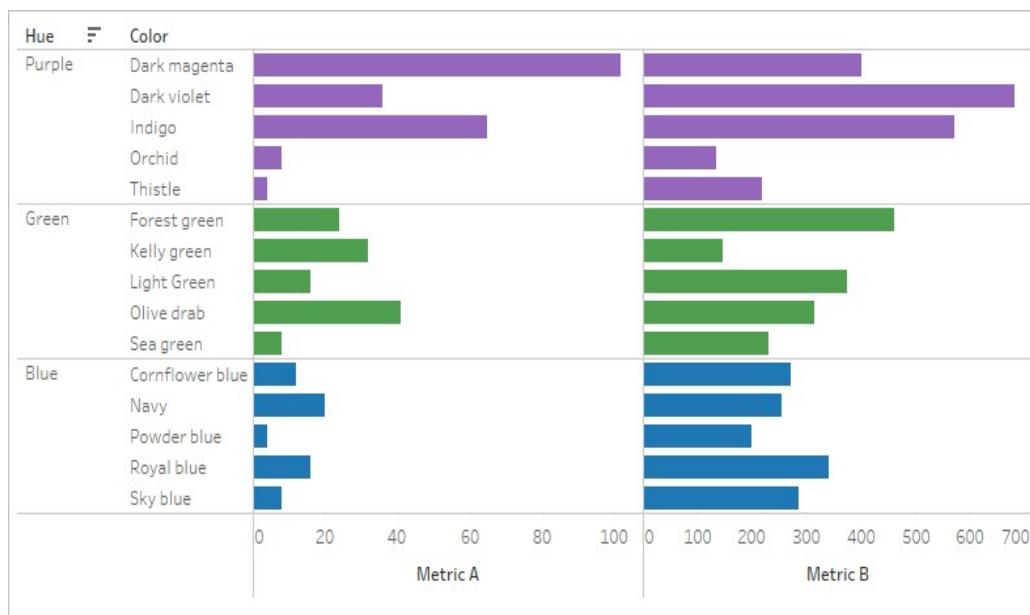
Hình 2.2.10. Sắp xếp theo một trục.

- **Sắp xếp từ một tiêu đề**



Hình 2.2.11. Sắp xếp theo tiêu đề.

- **Sắp xếp từ một nhãn trường**

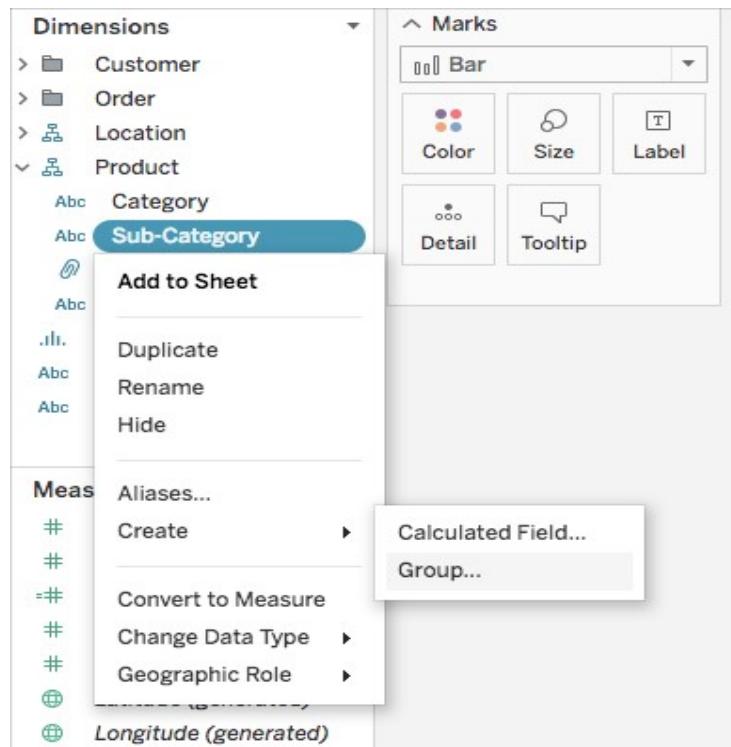


Hình 2.2.12. Sắp xếp theo nhãn.

2.2.4 Group.

- **Tạo một nhóm bằng cách chọn dữ liệu trong view**

Trong view đang làm việc, chọn một hoặc nhiều điểm dữ liệu, trên thanh công cụ xuất hiện nhập vào biểu tượng gộp nhóm.

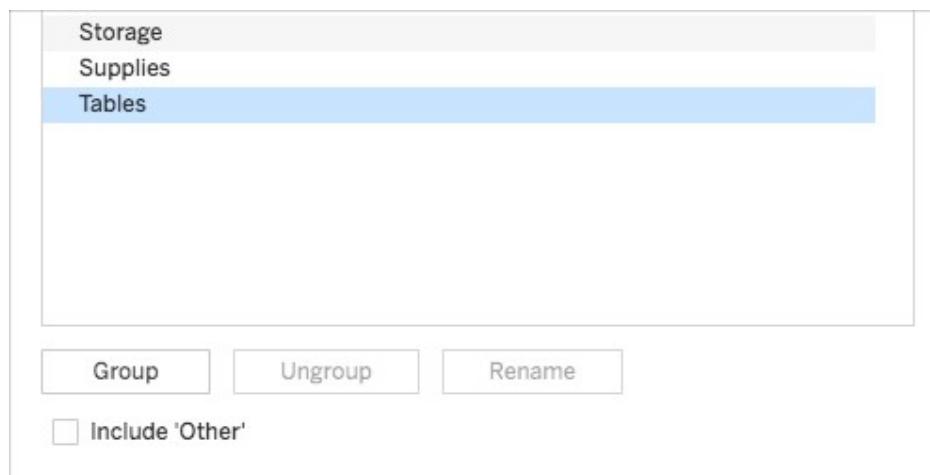


Hình 2.2.13. Tạo một nhóm dữ liệu.

- **Tạo một nhóm từ một Field trong Data pane:**

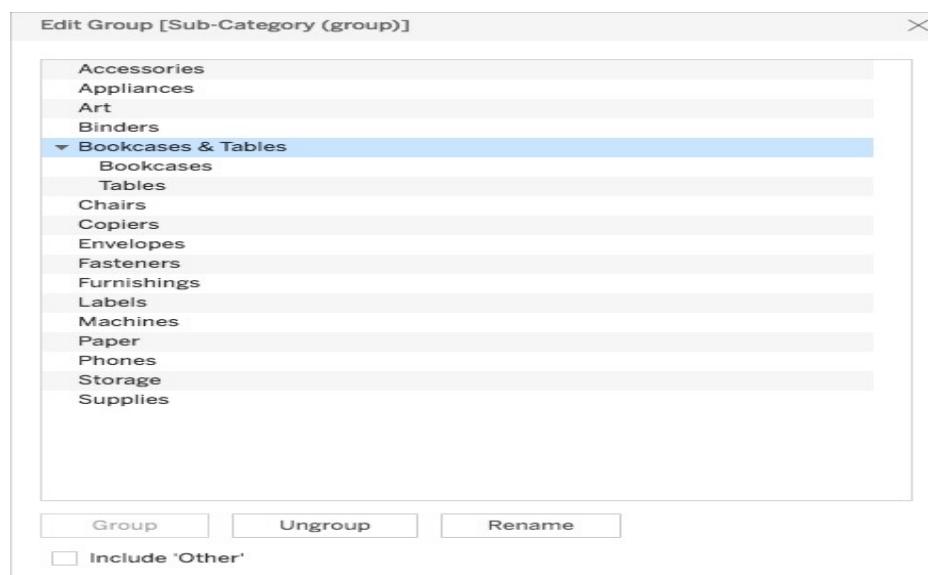
Trong **Data pane**, nhấp chuột phải vào một trường và chọn **Create => Group**.

Tại hộp thoại tạo nhóm, chọn các thành viên mà bạn muốn nhóm và nhấn vào nút **Group**.



Hình 2.2.14. Chọn dữ liệu muốn nhóm.

Các dữ liệu thành viên được chọn sẽ được gộp vào một nhóm với một cái tên mặc định được tạo tự động. Để đổi tên nhóm, chọn nó trong danh sách và nhấn **Rename**.



Hình 2.2.15. Chọn dữ liệu muốn nhóm.

2.2.5 LOD (Level of Detail Expressions)

Biểu thức LOD là một cách để tính toán hiệu quả các tập hợp không ở mức độ chi tiết của trực quan hóa. Một ví dụ đơn giản là thêm thứ nguyên vào giá trị tổng hợp đã được tính toán. Chúng giúp bạn nhanh chóng nhận được câu trả lời liên quan đến nhiều cấp độ chi tiết trong một lần trực quan hóa. Biểu thức LOD bao gồm các hàm:

- **Fixed:** Fixed tính toán một giá trị bằng cách sử dụng các chiều dữ liệu đã chỉ định mà không tham chiếu đến các dimensions trong view.

Sheet 4

Region	State	
Midwest	WI	4,439,681
Northeast	CT	2,211,472
	MA	4,456,565
	ME	3,039,019
	NH	1,467,672
	NJ	5,148,934
	NY	11,747,764
	PA	10,789,702
	RI	323,768
	VT	2,924,481
South	AL	4,187,979
	AR	4,093,113
	DC	1,478,855
	DE	551,209
	FL	9,157,738
	GA	5,269,134
	KY	7,357,426
	LA	4,118,067
	MD	3,322,195
	MS	5,383,165
	NC	6,829,618
	OK	4,541,635

Hình 2.2.16. Sử dụng hàm FIXED để tính tổng doanh thu từng vùng.

- **INCLUDE:** INCLUDE tính toán các giá trị bằng cách sử dụng các chiều dữ liệu được chỉ định bên cạnh bất kỳ dimensions nào có trong view.



Hình 2.2.17. Sử dụng hàm FIXED để tính tổng doanh thu từng vùng.

- **EXCLUDE:** EXCLUDE khai báo các các chiều dữ liệu để bỏ qua mức độ chi tiết của view.

The table displays monthly sales figures for four regions: Midwest, Northeast, South, and West. The rows represent months from January to December. The columns are labeled "Month of Or.", "Midwest", "Northeast", "South", and "West". The values are formatted with commas and include a trailing 'M' for millions. The table uses the FIXED function to calculate the total sales for each region.

Month of Or..	Region			
	Midwest	Northeast	South	West
January	1,830,277	1,180,362	2,734,422	1,148,106
February	1,414,548	879,289	2,104,816	1,007,454
March	6,953,495	3,773,575	8,869,475	4,143,258
April	8,996,808	5,512,622	11,711,432	5,549,613
May	2,300,999	1,442,123	3,371,850	1,459,162
June	8,731,384	5,944,165	12,210,477	6,420,757
July	4,314,293	2,975,564	6,833,711	2,999,782
August	3,788,039	2,886,740	6,974,350	3,033,254
September	4,056,003	2,213,244	4,101,094	1,668,561
October	1,268,984	816,990	2,151,981	802,228
November	2,647,670	1,586,121	4,118,644	2,037,468
December	19,282,605	12,898,584	27,291,571	13,391,978
	0M	20M	40M	0M
	Price	Price	Price	Price

Hình 2.2.18. Sử dụng hàm EXCLUDE để tính giá cả của từng khu vực theo tháng.

CHƯƠNG 3: HIỆN THỰC HÓA NGHIÊN CỨU

3.1 Cài đặt phần mềm.

- **Bước 1:** Truy cập vào trang web <http://www.tableau.com/products/desktop>.



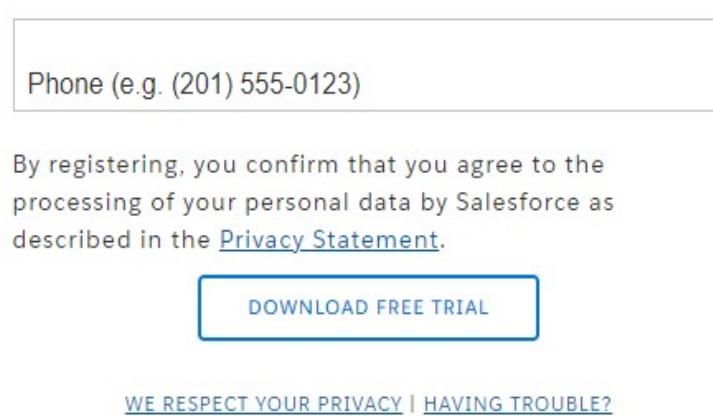
Hình 3.1.1. Trang chủ Tableau.

- **Bước 2:** Điền thông tin đăng nhập.

A screenshot of a sign-up form for Tableau Desktop. The title 'Tableau Desktop: Start your free 14-day trial' is at the top. Below it, a message says 'Almost there!'. It informs the user that it only takes 15 seconds to fill out and provides a link to 'sign in' if they are already registered. There are two input fields: 'First Name' and 'Last Name', both currently empty.

Hình 3.1.2. Điền thông tin cá nhân.

Sau đó, chọn “Download free trial”.

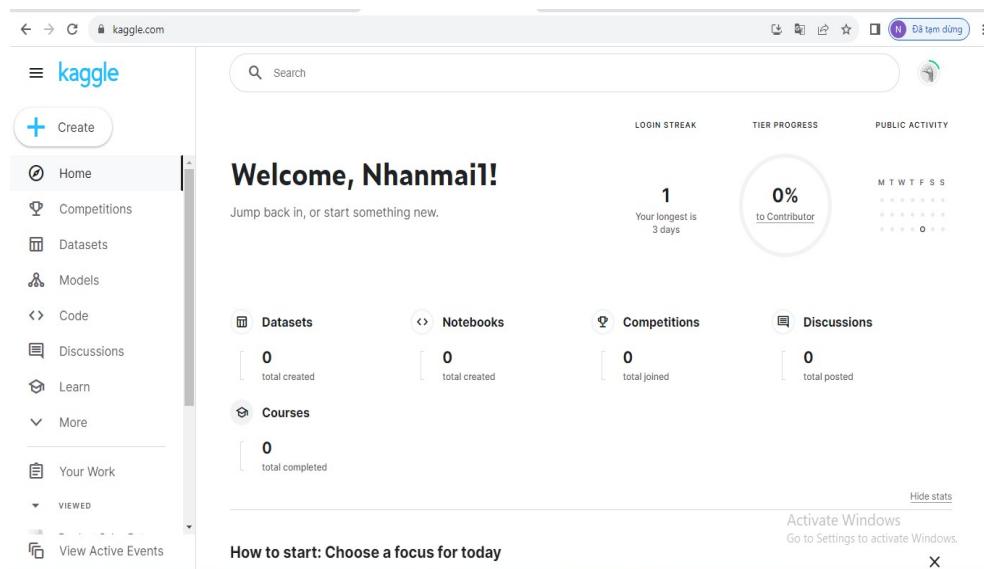


Hình 3.1.3. Tải phần mềm Tableau.

- **Bước 3:** Mở File vừa được Download lên. Click vào “I have read and accept the terms of the license agreement” > Click “Install”.

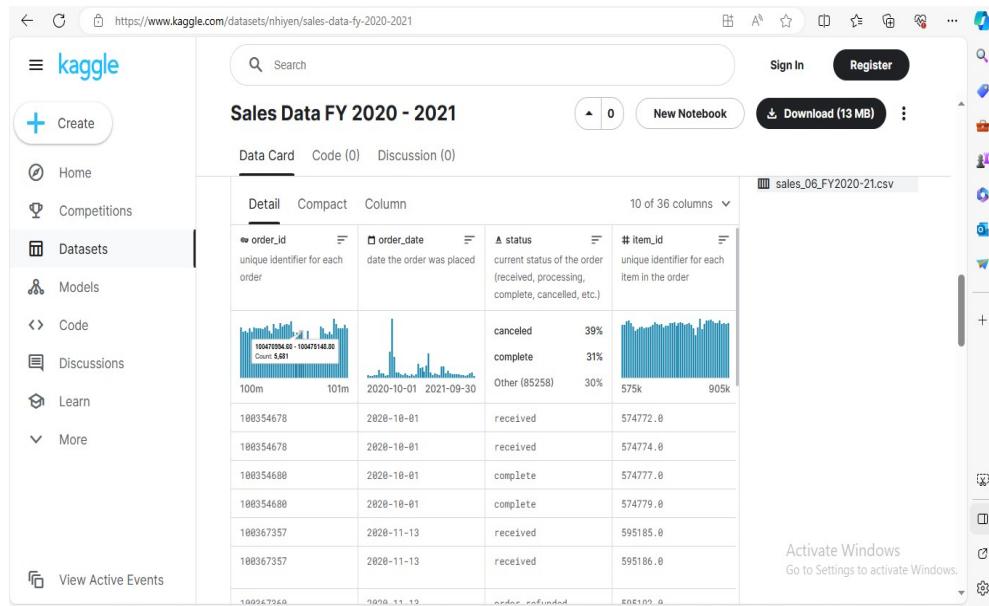
3.2 Thu thập dữ liệu.

Sử dụng dữ liệu trên trang <http://www.Kaggle.com>.



Hình 3.2.1. Trang chủ Kaggle.

Để chọn dữ liệu phù hợp vào Dataset và nhập tên dữ liệu muốn sử dụng.



Hình 3.2.2. Tập dữ liệu “sales_06_FY2020-21.csv”

3.3 Kết nối dữ liệu.

Làm thế nào để kết nối để kết nối dữ liệu?

Tableau cung cấp hầu hết các kiểu dữ liệu từ Excel, Microsoft Access, Microsoft SQL Server,... Để kết nối dữ liệu cần kết nối với tập dữ liệu. Sau đó, chọn kiểu join dữ liệu phù hợp trong cửa sổ làm việc của Tableau.

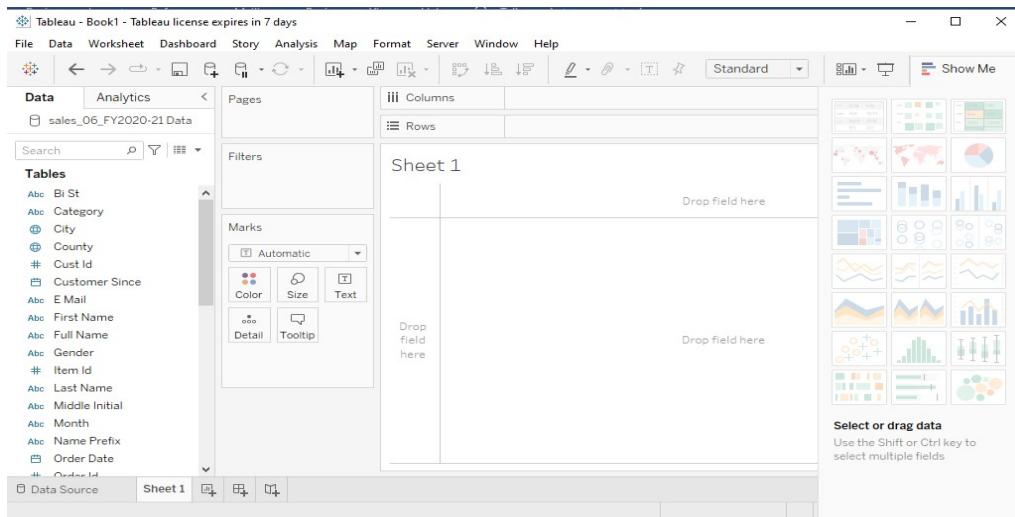
#	sales_06_FY2020-21 Data.csv	sales_06_FY2020-21 Data.csv	Akt	sales_06_FY2020-21 Data.csv
	Order Id	Order Date	Status	Item Id
100412903	12/27/2020	cancelled	67402700	
100416476	12/27/2020	cancelled	679133.00	
100416485	12/27/2020	cancelled	679142.00	
100416493	12/27/2020	cancelled	679151.00	

Hình 3.3.1. Join dữ liệu.

Chọn worksheet để mở ra trang làm việc trong Tableau.

Hình 3.3.2. Tạo trang Worksheet mới.

Trang Worksheet có giao diện đầy đủ như sau.



Hình 3.3.3. Giao diện màn hình làm việc của Tableau.

3.4 Cài đặt bài toán.

Bài toán mô tả dữ liệu bán hàng qua từng năm của công ty. Cần xây dựng bài dựng bài toán để tính doanh thu của công ty qua từng năm, với số liệu từ tập dữ liệu “sales_06_FY2020-21.csv”.

3.4.1 Mô tả tập dữ liệu “sales_06_FY2020-21.csv” .

Tập dữ liệu này chứa thông tin về các đơn đặt hàng của khách hàng trên một trang web thương mại điện tử. Nó bao gồm 32 thuộc tính và 286.393 bộ dữ liệu.

STT	Thuộc tính	Diễn giải
1	Order_id	Mã đơn hàng
2	Order_date	Ngày đặt hàng
3	Status	Trạng thái
4	Item_id	Mã mặt hàng
5	Qty_ordered	Số lượng đặt hàng
6	Price	Giá
7	Value	Giá trị
8	Discount_amount	Số tiền chiết khấu
9	Total	Tổng số
10	Category	Danh mục hàng hóa

11	Payment_method	Phương thức thanh toán
12	Customer_id	Mã khách hàng
13	Year	Năm
14	Month	Ngày tháng
15	Ref_num	Số tham chiếu
16	Name Prefix	Danh xưng
17	First Name	Tên
18	Middle Initial	Tên đệm
19	Last Name	Họ
20	Gender	Giới tính
21	Age	Tuổi
22	E Mail	Địa chỉ thư điện tử
23	Customer Since	Ngày đăng ký khách hàng
24	Phone No.	Số điện thoại
25	Place Name	Tên địa điểm
26	County	Quận
27	City	Thành phố
28	State	Tiểu bang
29	zip	Mã Zip
30	Region	Khu vực
31	User Name	Tên người dùng
32	Discount_Percent	Phần trăm chiết khấu

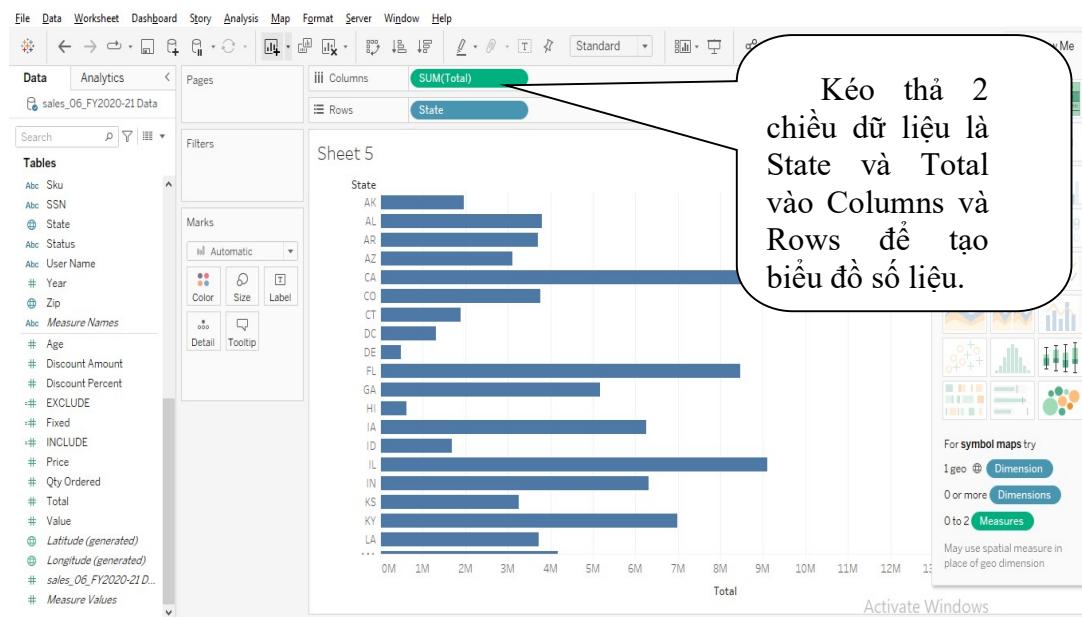
Bộ dữ liệu này rất hữu ích để phân tích mô hình mua hàng của khách hàng trên trang web thương mại điện tử. Nó có thể giúp xác định các sản phẩm phổ biến, hành vi của khách hàng và xu hướng mua hàng. Ngoài ra, tập dữ liệu này có thể được sử dụng để phát triển và đào tạo các thuật toán học máy nhằm dự đoán hành vi của khách hàng và xu hướng bán hàng.

- Nhìn chung, tập dữ liệu này cung cấp những hiểu biết có giá trị về mô hình mua hàng của khách hàng và có thể được các doanh nghiệp sử dụng để đưa ra quyết định sáng suốt.

Một số câu hỏi đặt ra như sau:

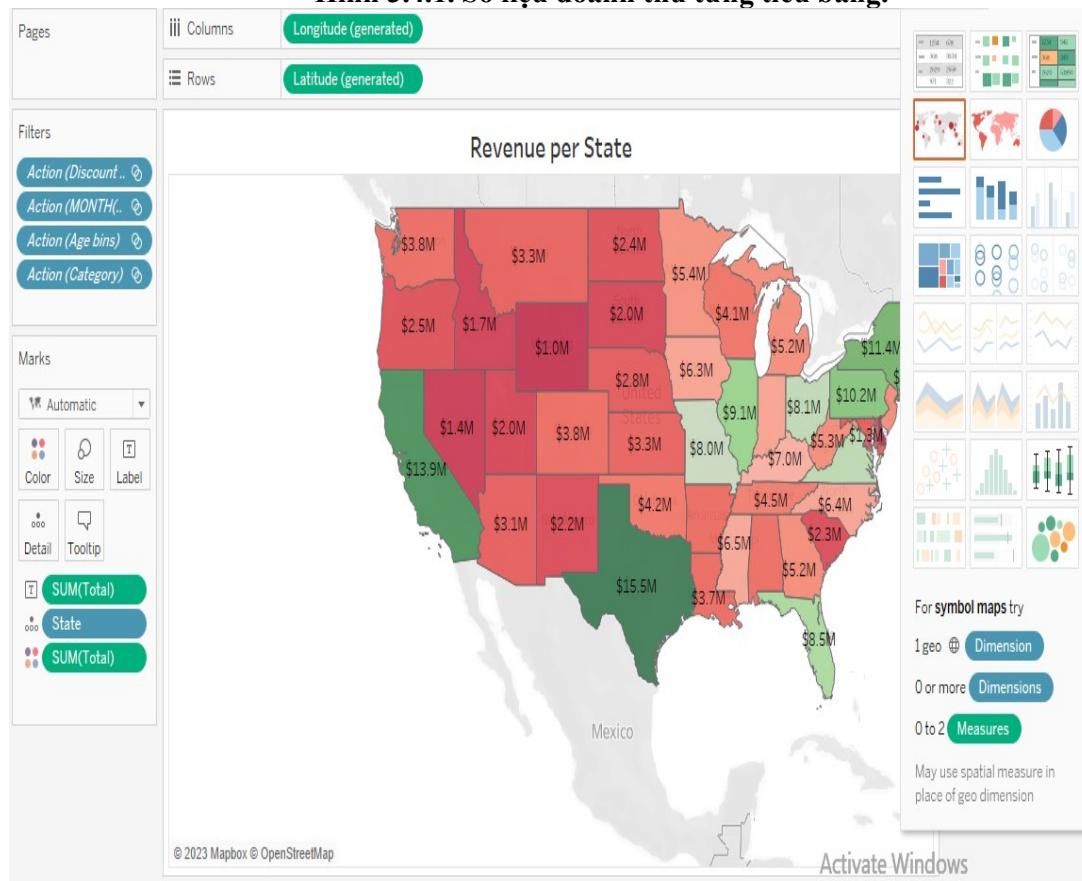
- Cho biết doanh thu bán hàng ở từng tiểu bang ?
- Doanh thu của công ty hàng tháng là bao nhiêu ?
- Cho biết doanh thu theo tuổi ?
- Cho biết tương quan giữa số lượng và tỷ lệ chiết khấu..?
- Phần trăm mỗi khu vực là bao nhiêu?
- Doanh thu mỗi doanh mục sản phẩm là bao nhiêu?
- Cho biết tổng doanh thu trung bình của mỗi tiểu bang?
- Cho biết tần suất đặt hàng?
- Xây dựng Dashboard như thế nào?

3.4.2 Trả lời câu hỏi "Cho biết doanh thu từng tiểu bang".



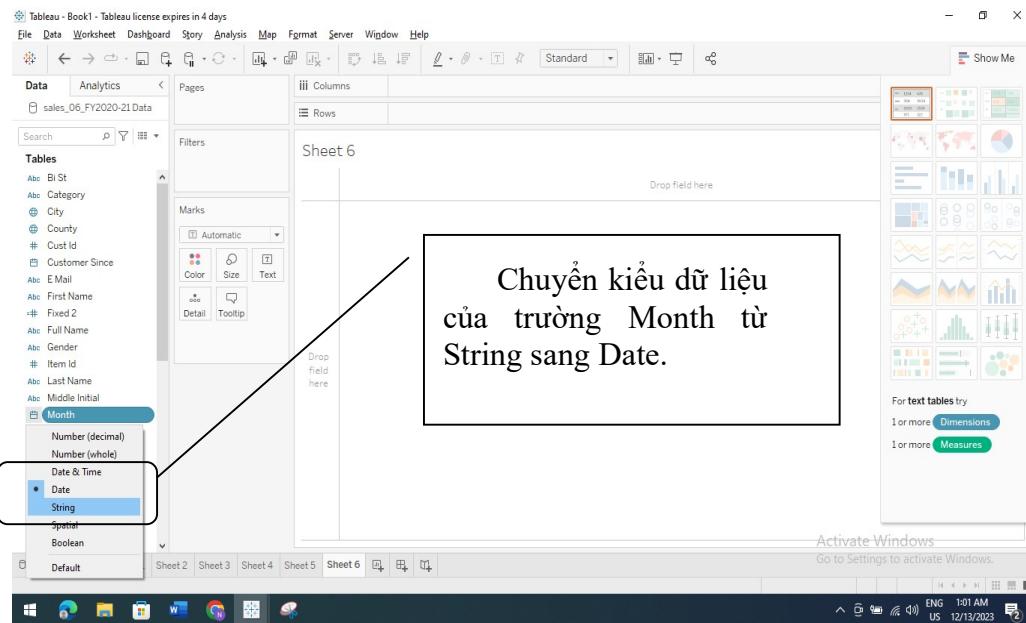
Kéo thả 2 chiều dữ liệu là State và Total vào Columns và Rows để tạo biểu đồ số liệu.

Hình 3.4.1. Số liệu doanh thu từng tiểu bang.



Hình 3.4.2. Bản đồ thể hiện doanh thu từng tiểu bang.

3.4.3 Tạo worksheet tính doanh thu theo tháng trong năm.



Hình 3.4.3. Chuyển kiểu dữ liệu trường Month.

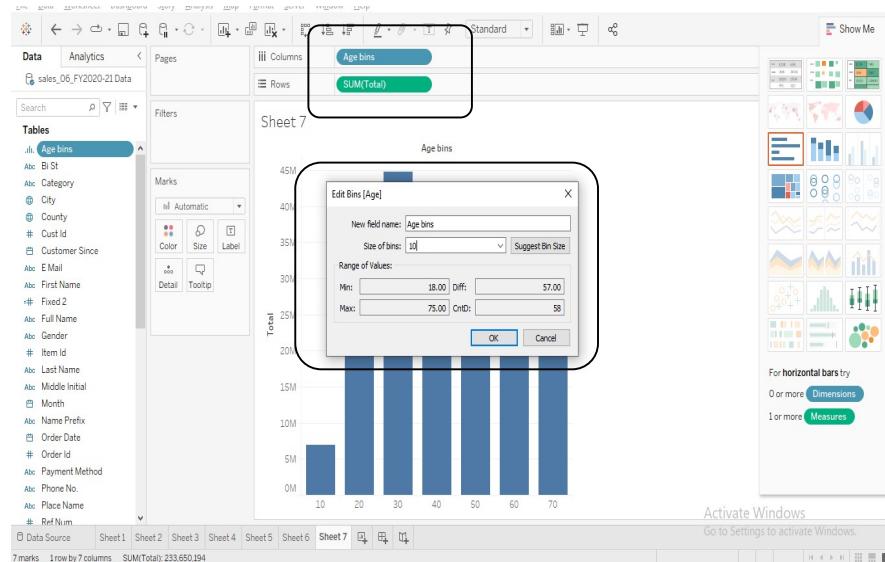
Sau đó, kéo thả trường Month và Total lần lượt vào Rows và Columns trong vùng làm việc của Tableau.



Hình 3.4.4. Doanh thu theo tháng hàng năm.

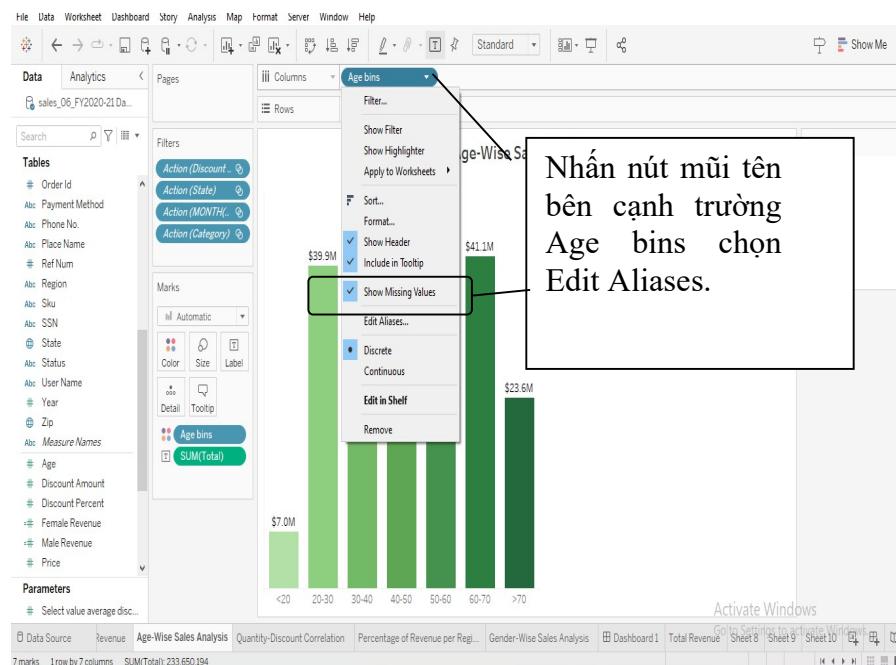
3.4.4 Tính doanh thu theo tuổi.

Tạo một field name (Age bins) trong Age.

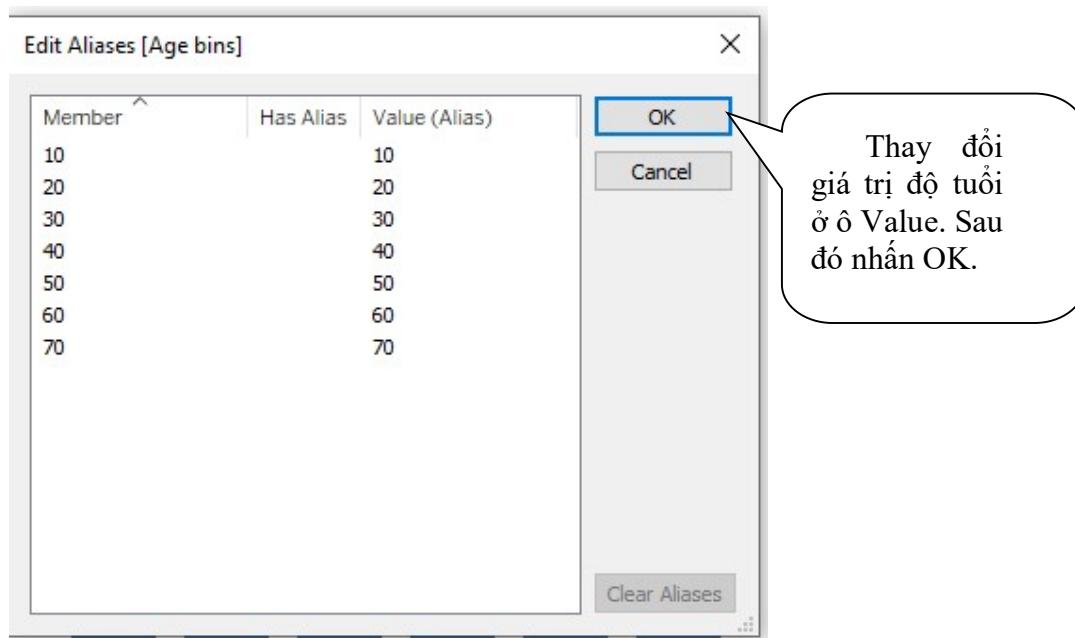


Hình 3.4.5. Doanh thu theo tháng hàng năm.

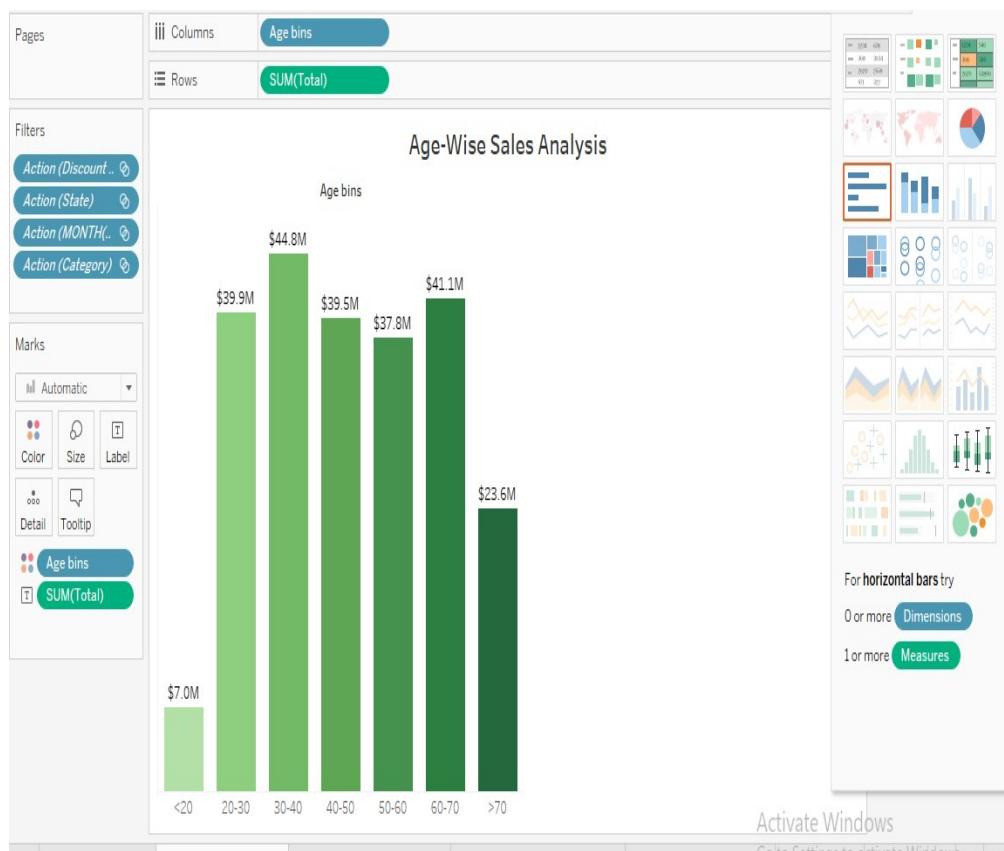
Kéo thả trường Age bins và Total vào vùng làm việc của Tableau.



Hình 3.4.6. Mở hộp thoại Edit Aliases.



Hình 3.4.7. Thay đổi giá trị độ tuổi.



Hình 3.4.8. Biểu đồ thể hiện doanh thu theo tuổi.

3.4.5 Tương quan giữa số lượng và tỷ lệ chiết khấu

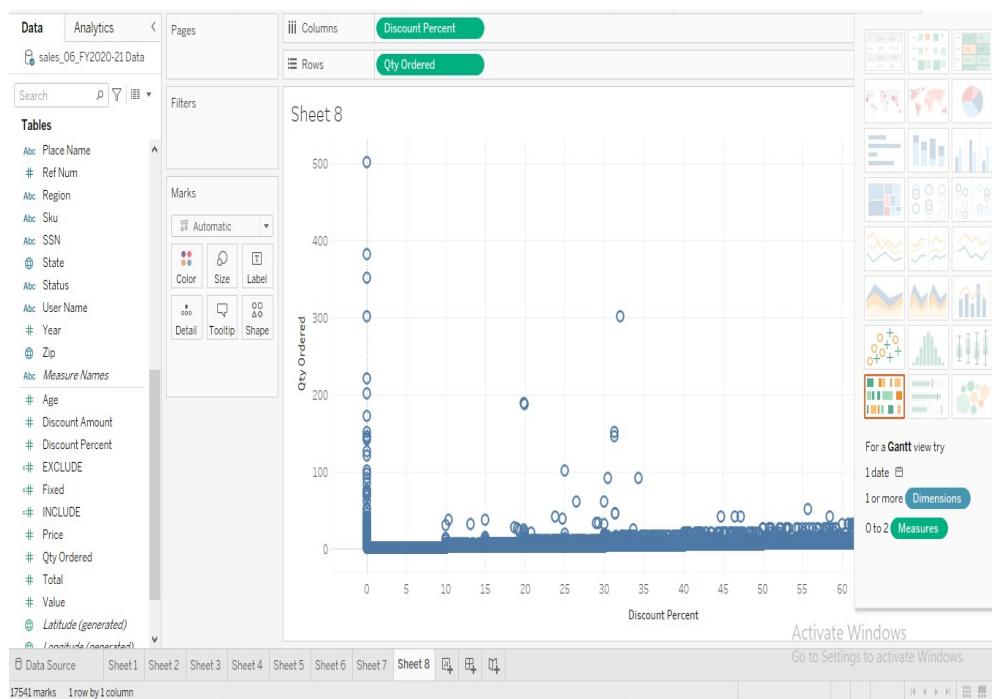
Tạo biểu đồ bubble để thể hiện tương quan giữa 2 measure Qty Ordered và Discount Percent.

Chuyển đổi kiểu dữ liệu Qty Ordered và Discount Percent về dạng Dimensions.

Context menu (right-clicked on measures):

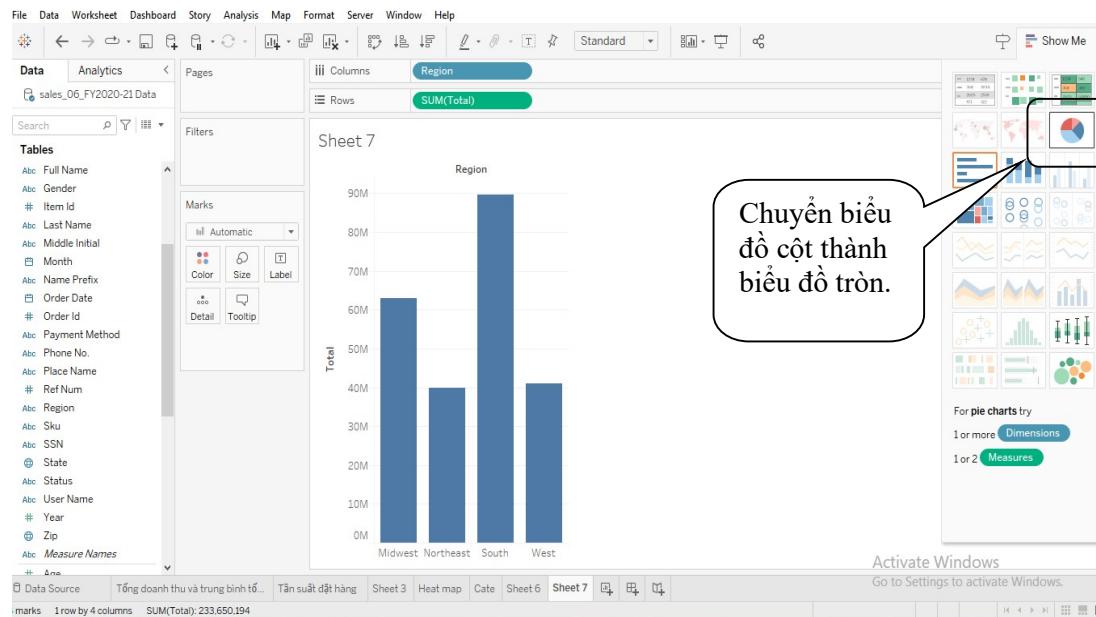
- Dimension (selected)
- Attribute
- Measure (Sum)
- Discrete
- Continuous
- Quick Table Calculation
- Swap
- Remove

Hình 3.4.9. Chuyển đổi kiểu dữ liệu.



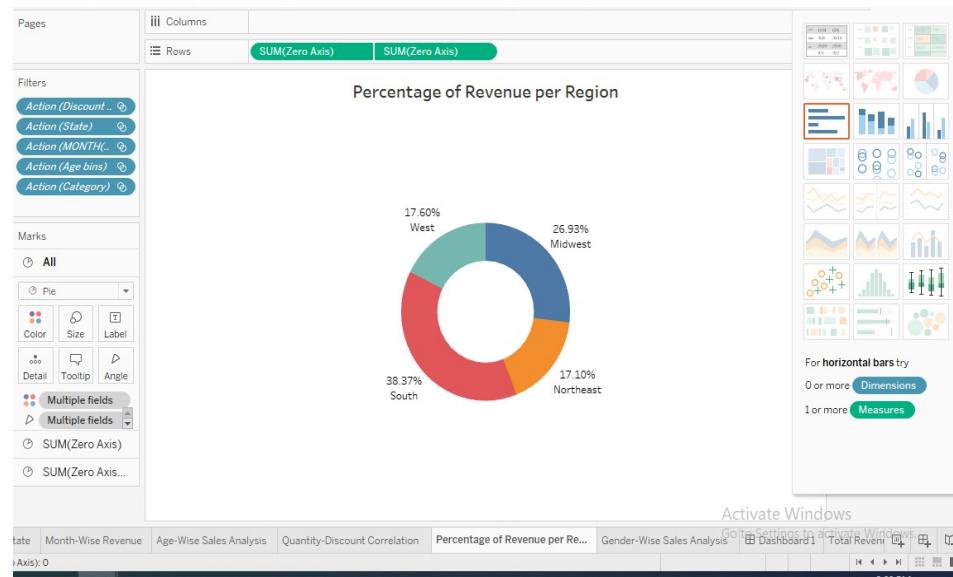
Hình 3.4.10. Biểu đồ bong bóng thể hiện tương quan giữa số lượng và tỷ lệ chiết khấu.

3.4.6 Tính phần trăm doanh thu mỗi khu vực.



Hình 3.4.11. Chuyển biểu đồ cột thành biểu đồ tròn.

Tạo biểu đồ tròn giữa Region và Total để tính phần trăm doanh thu của mỗi khu vực.



Hình 3.4.12. Biểu đồ tròn thể hiện phần trăm doanh thu từng khu vực.

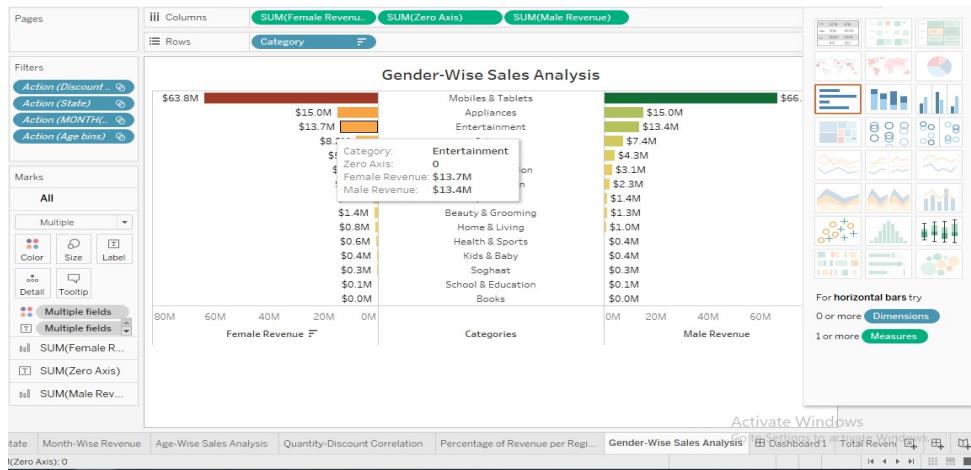
3.4.7 Doanh thu mỗi danh mục sản phẩm theo giới tính.

Để tính doanh thu theo giới tính nam và nữ cần phải làm gì?

Dùng hàm IF để tính doanh thu theo giới tính. Để tính doanh thu khách hàng cần cài đặt hàm như sau:

IF [Gender] = 'M' THEN [Total] END (để tính doanh thu số khách hàng nam).

IF [Gender] = 'F' THEN [Total] END (để tính doanh thu số khách hàng nữ).



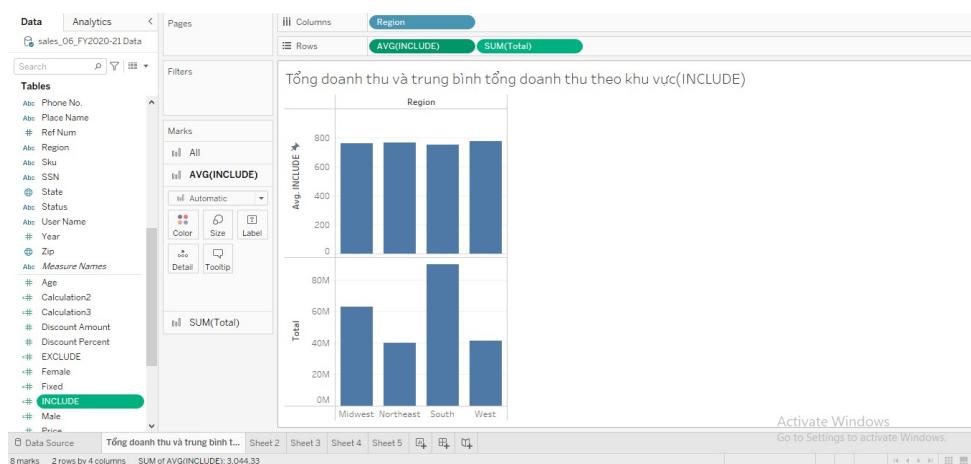
Hình 3.4.13. Tổng doanh thu theo giới tính.

3.4.8 Tổng doanh thu trung bình mỗi tiểu bang.

Tính doanh thu trung bình với mỗi khách hàng.

{ INCLUDE [User Name] : AVG([Total]) }

Kéo Region vào vùng làm việc của Tableau.



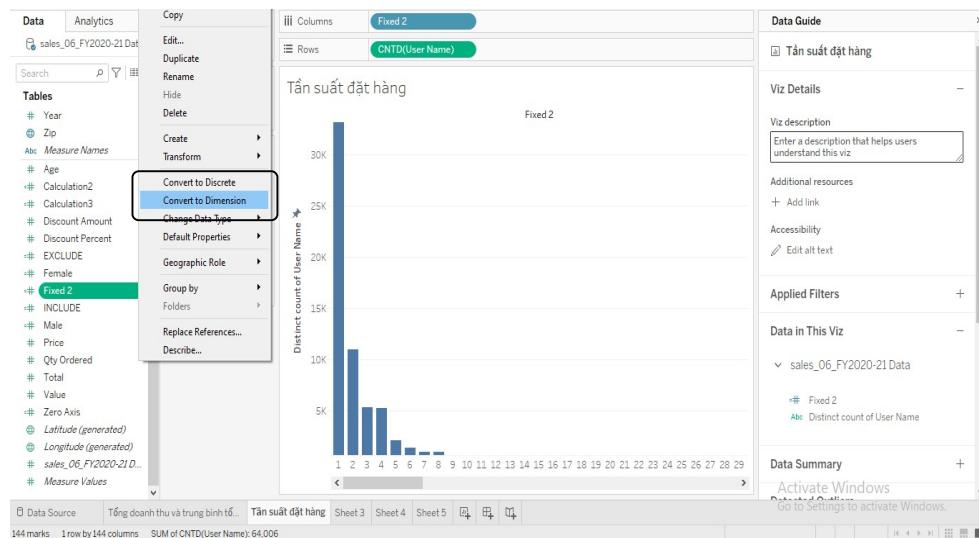
Hình 3.4.14. Doanh thu trung bình của từng vùng.

3.4.9 Tần suất đặt hàng.

Tính tần suất đặt hàng.

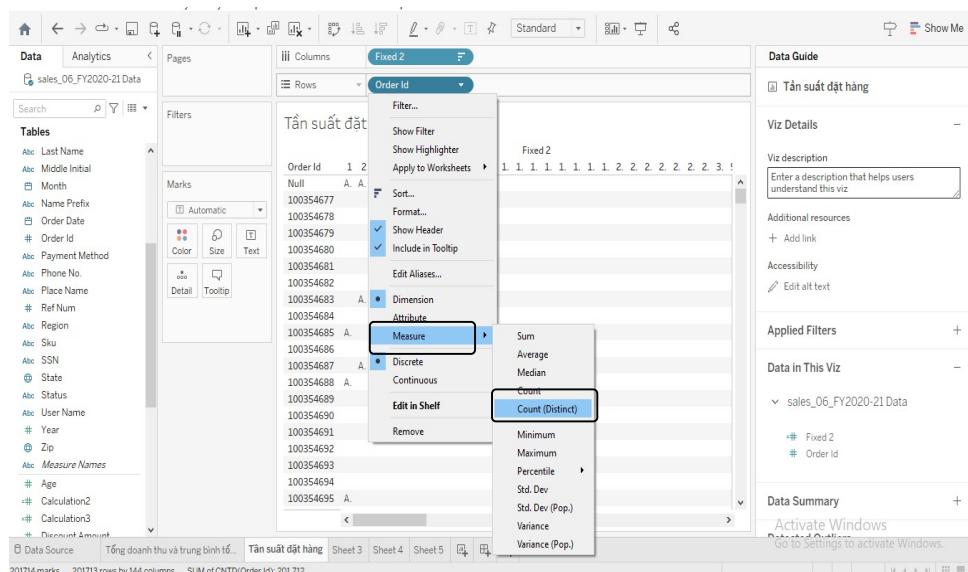
{ FIXED [User Name]: COUNTD([Order Id]) }

Chuyển công thức vừa tạo từ Measure thành Dimension. Nhấp chuột vào measure và chọn Convert to Dimension.

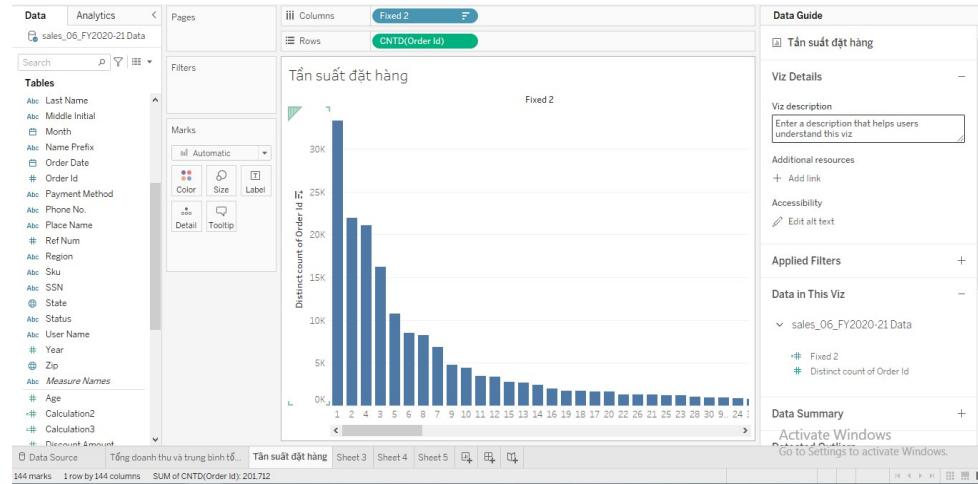


Hình 3.4.15. Chuyển công thức từ Measure về dạng Dimensions.

Kéo Order ID vào Rows và chuyển Measure Values thành COUNTD.



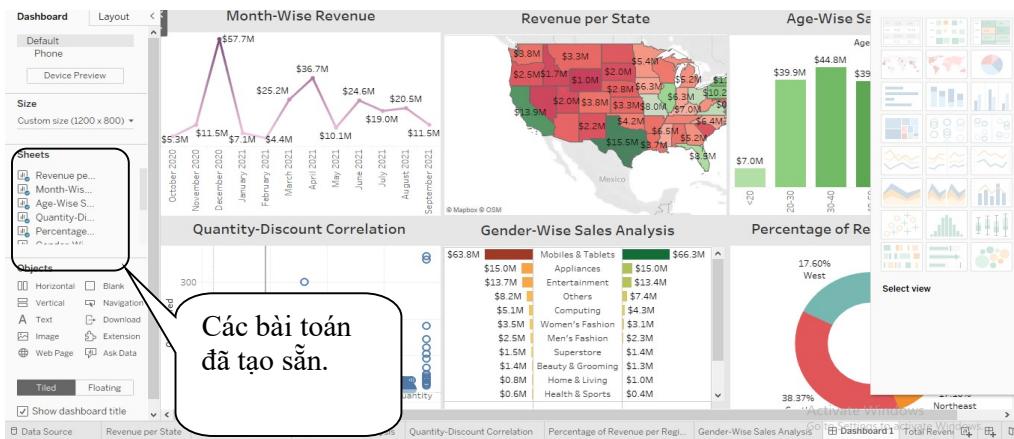
Hình 3.4.16. Chuyển công thức Measure về dạng COUNTD.



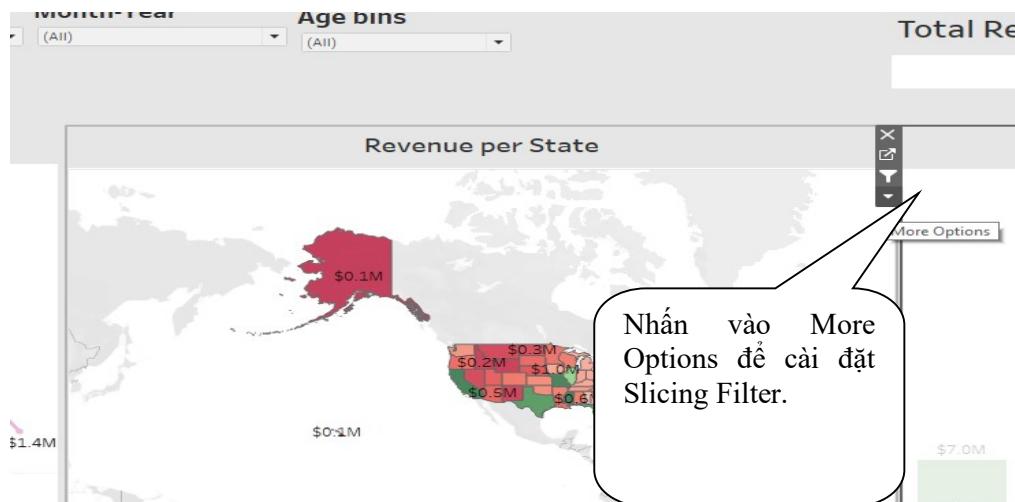
Hình 3.4.17. Biểu đồ thể hiện tần suất đặt hàng.

3.4.10 Xây dựng bảng Dashboard (Bảng điều khiển)

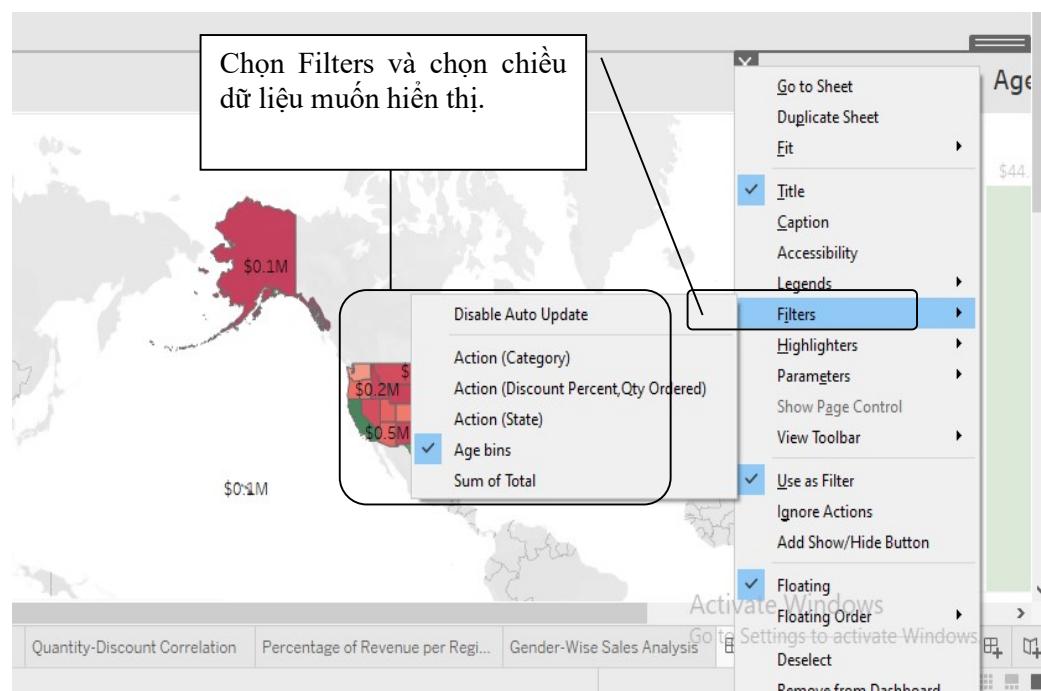
Tạo một Dashboard thả vào các biểu đồ đã thực hiện từ Sheets vào vùng làm việc của Dashboard.



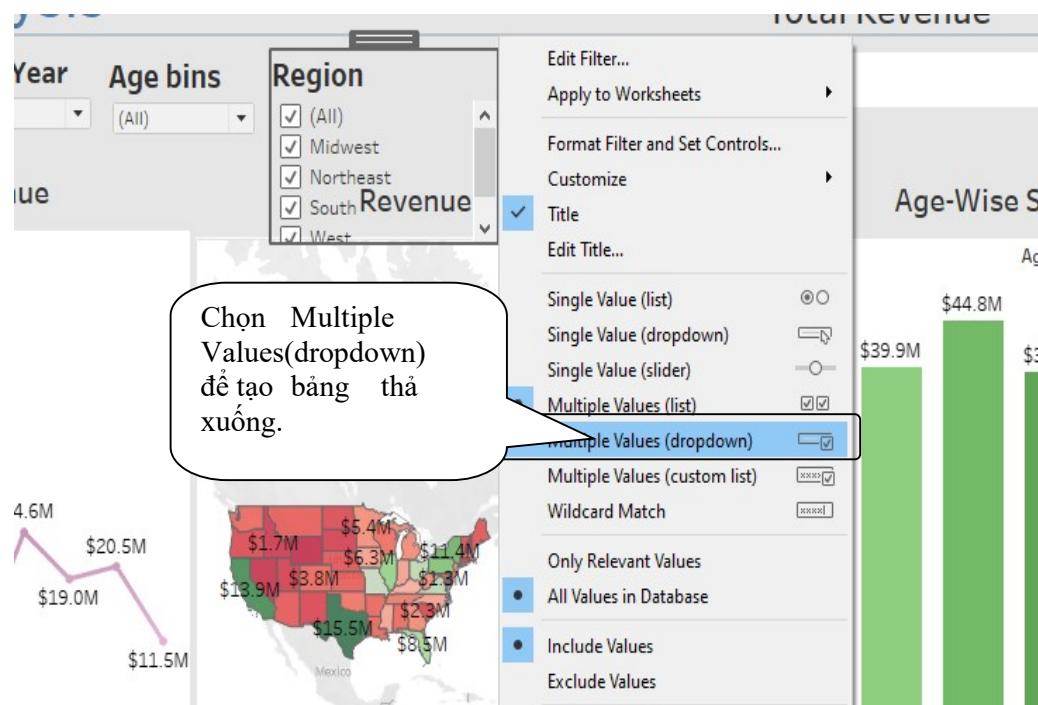
Hình 3.4.18. Xây dựng Dashboard.



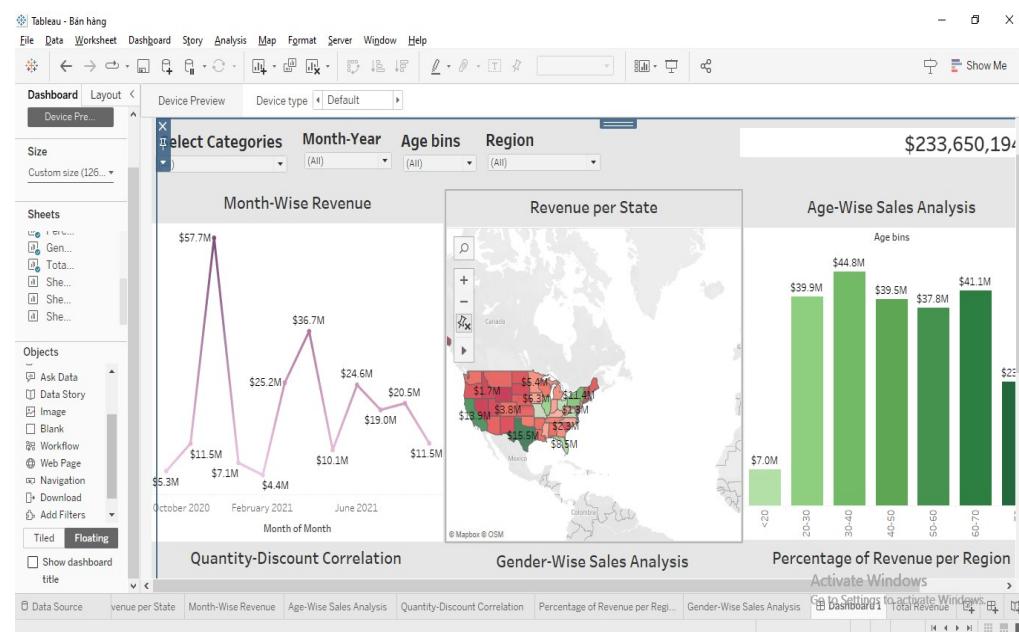
Hình 3.4.19. Cài đặt Slicing Filter.



Hình 3.4.20. Chọn kiểu dữ liệu muốn hiển thị.



Hình 3.4.21. Tạo bảng thả xuống.



Hình 3.4.22. Bảng Dashboard.

CHƯƠNG 4: KẾT QUẢ NGHIÊN CỨU

4.1 Kết quả đạt được sau khi tìm hiểu về phần mềm Tableau.

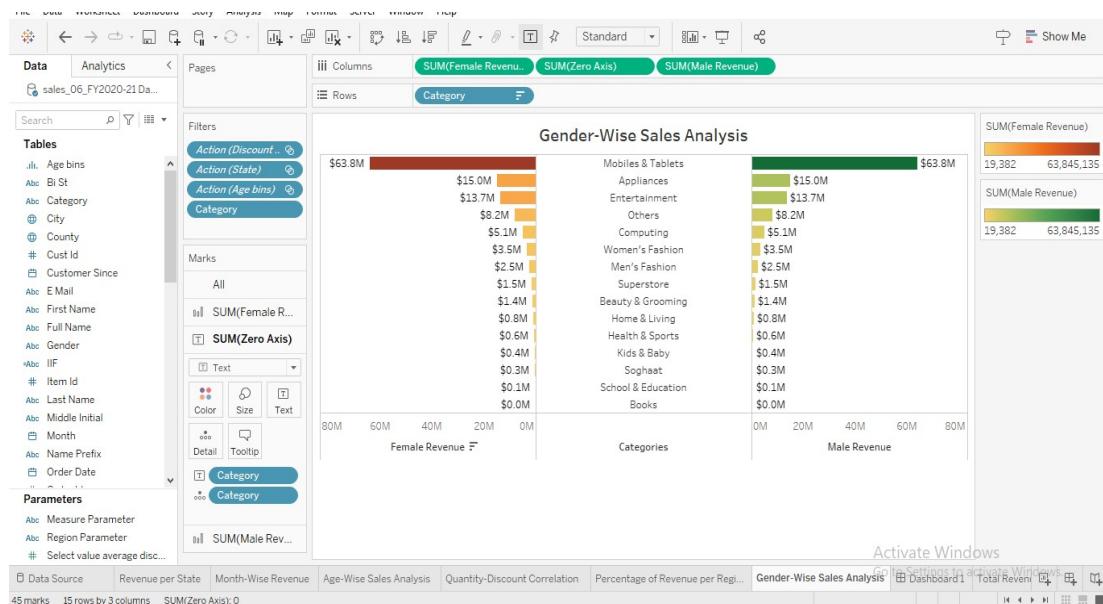
Sau khi thực hiện bài toán, kết quả thực hiện nghiên cứu ứng dụng Tableau thu được kết quả như sau:

- Kết nối dữ liệu bán hàng vào phần mềm Tableau:

The screenshot shows the Tableau Data Source interface. On the left, under 'Connections', there is one entry: 'sales_06_FY2020-21 Data (1)'. Below it, under 'Files', there is a single file: 'sales_06_FY2020-21 Data.csv'. A callout bubble points to this file with the text: 'Tập dữ liệu “sales_06_FY2020-21.csv” được kết nối.' (The dataset "sales_06_FY2020-21.csv" is connected). At the bottom of the interface, there are several tabs: 'Data Source', 'Revenue per State', 'Month-Wise Revenue', 'Age-Wise Sales Analysis', 'Quantity-Discount Correlation', 'Percentage of Revenue per Region', 'Gender-Wise Sales Analysis', 'Dashboard 1', 'Total Revenue', and 'Activate Windows'.

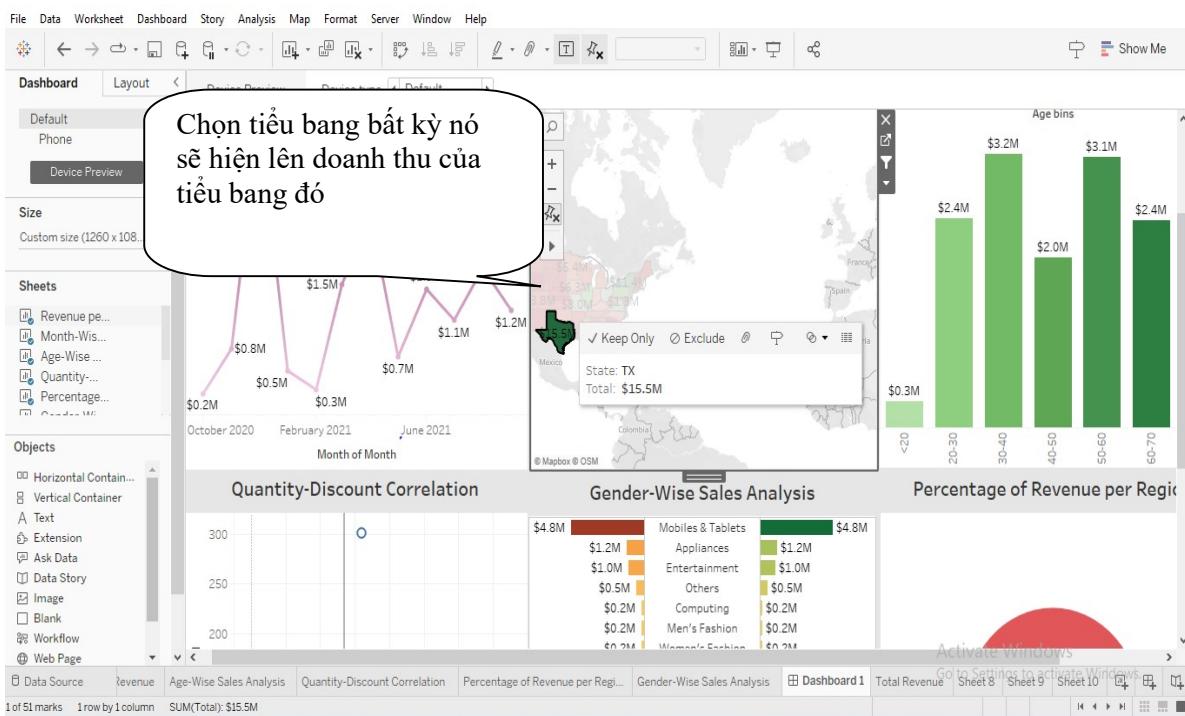
Hình 4.1.1. Kết nối dữ liệu bán hàng.

- Áp dụng các hàm phân tích và trực quan hóa dữ liệu:

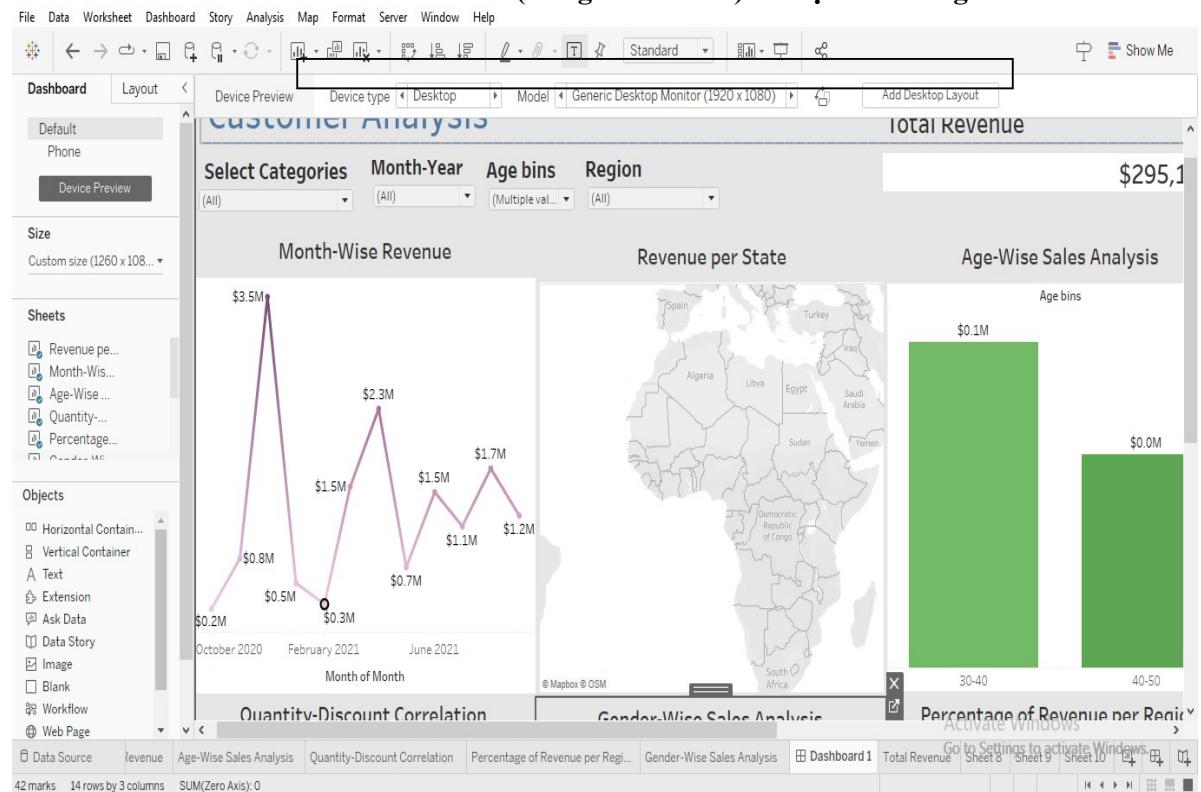


Hình 4.1.2. Kết nối dữ liệu bán hàng.

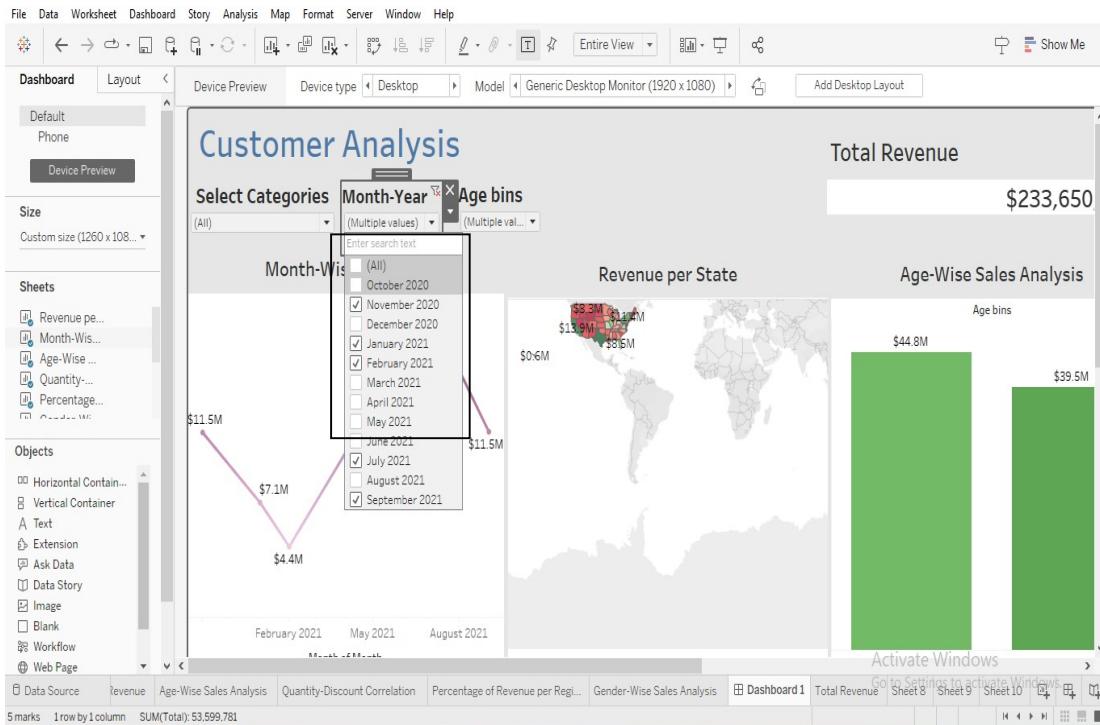
- Xây dựng Dashboard dữ liệu bán hàng:



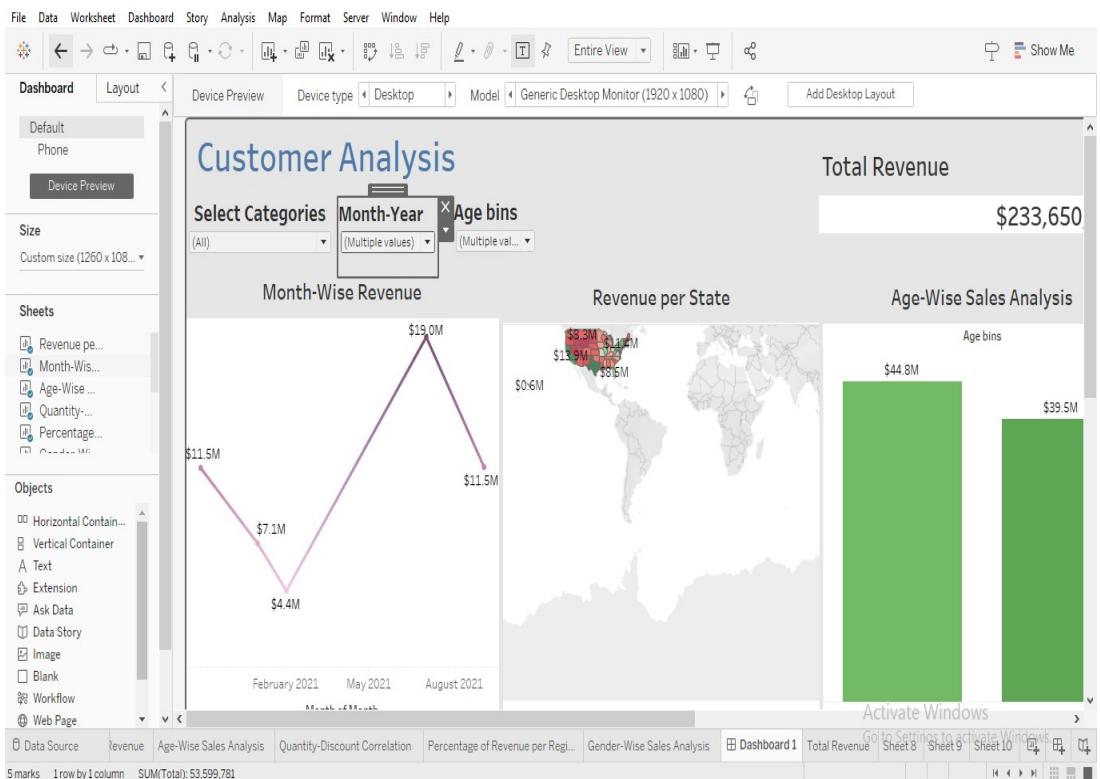
Hình 4.1.3. Dashboard (Bảng điều khiển) dữ liệu bán hàng .



Hình 4.1.4. Dashboard thiết lập trên màn hình máy tính .

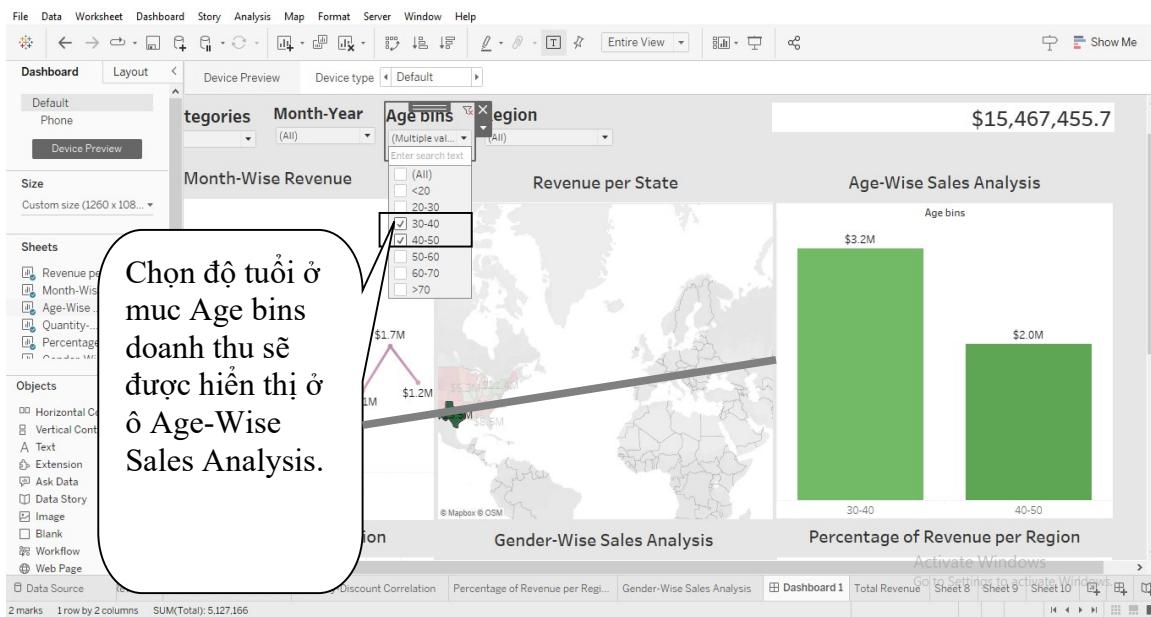


Hình 4.1.5. Chọn ngày tháng bất kỳ trong mục Month-Year.

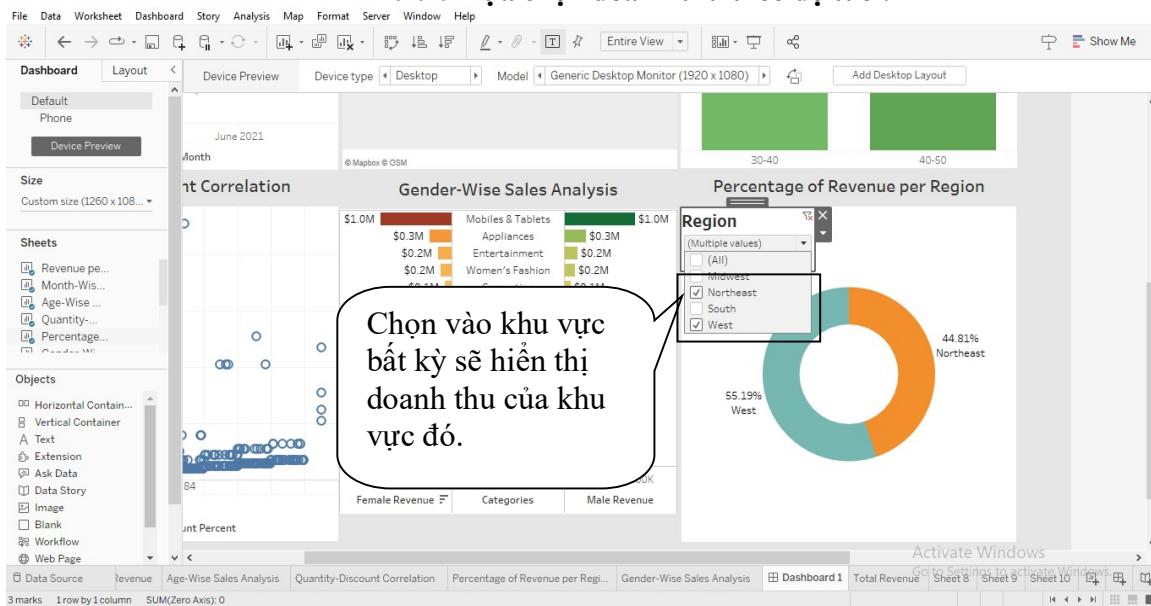


Hình 4.1.6. Hiển thị doanh thu sau khi chọn ngày tháng.

4.2 Bộ lọc Slicing Filters.



Hình 4.2.1. Lựa chọn doanh thu theo độ tuổi.



Hình 4.2.2. Lựa chọn doanh thu theo khu vực.

CHƯƠNG 5: KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

❖ Kết quả đạt được sau khi thực hiện đồ án nghiên cứu:

➤ Về kiến thức:

Làm quen với công cụ hỗ trợ phân tích và trực quan hóa dữ liệu Tableau nhằm tạo bước đà cho các môn học liên quan đến phân tích dữ liệu. Hiểu rõ hơn các công việc cần làm khi làm việc với các tập dữ liệu bán hàng.

➤ Về kỹ năng :

- Xây dựng được các hàm phân tích dữ liệu.
- Tạo biểu đồ so sánh dữ liệu.
- Xây dựng được bảng điều khiển Dashboard.

❖ Hạn chế:

- Chưa nắm rõ một số hàm quan trọng, việc nghiên cứu bài toán chưa được chỉnh chu.
- Xây dựng các hàm xử lý dữ liệu còn đơn giản chưa thể hiện được hết vấn đề quan trọng của bài toán.
- Dashboard hiển thị trên thiết bị điện thoại, máy tính chưa phù hợp với Tableau.

❖ Hướng phát triển:

- Xây dựng bài toán với tập dữ liệu nhiều thành phần hơn .
- Tương thích phần mềm Tableau trên thiết bị như: điện thoại, máy tính, tablet,... với giao diện dễ nhìn.
- Xây dựng đa dạng hóa các loại biểu đồ để thể hiện rõ các công việc muốn làm rõ với dữ liệu bán hàng.
- Sử dụng các hàm nâng cao hơn để hoàn chỉnh bài toán bán hàng.
- Xây dựng Dashboard với giao diện bắt mắt hơn.

DANH MỤC TÀI LIỆU THAM KHẢO

1. <https://www.tableau.com/>
2. <https://tableau.edu.vn/>
3. <https://youtu.be/-Aj8IIC0IEA>