



## **BÀI TIỂU LUẬN** **HỌC PHẦN PHÂN TÍCH DỮ LIỆU LỚN**

**Đề tài:**

**Tìm hiểu phân tích dữ liệu lớn trong lĩnh vực kinh doanh và áp dụng phân tích tập dữ liệu Superstore trên Power BI**

**Tên nhóm: NHÓM 03**

**Nhóm sinh viên thực hiện:**

- 1. Khuất Đình Lễ**
- 2. Trần Trung Kiên**
- 3. Bùi Quang Huy**

**Giáo viên hướng dẫn: Ngô Hữu Huy**

*Thái Nguyên, tháng 05 năm 2022*

## MỤC LỤC

<b>MỤC LỤC.....</b>	<b>2</b>
<b>PHÂN CÔNG NHIỆM VỤ.....</b>	<b>3</b>
<b>MỞ ĐẦU.....</b>	<b>4</b>
<b>CHƯƠNG 1. CƠ SỞ LÝ THUYẾT.....</b>	<b>4</b>
1.1. Nội dung.....	5
<b>CHƯƠNG 2. CÔNG CỤ POWE BI.....</b>	<b>12</b>
2.1. Nội dung.....	12
<b>CHƯƠNG 3. BÀI TOÁN .....</b>	<b>14</b>
3.1. Mô tả bài toán.....	23
<b>KẾT LUẬN.....</b>	<b>33</b>
<b>TÀI LIỆU THAM KHẢO.....</b>	<b>39</b>

## PHÂN CÔNG NHIỆM VỤ

### Theo tiến độ thực hiện

Stt	Tên nhiệm vụ	Người thực hiện
1	Tìm hiểu về Big Data	Khuất Đình Lễ Bùi Quang Huy
2	Ứng dụng của Big Data trong kinh doanh	Khuất Đình Lễ Bùi Quang Huy Trần Trung Kiên
3	Tìm hiểu về Power BI	Bùi Quang Huy Trần Trung Kiên
4	Thu thập và tổng hợp dữ liệu	Khuất Đình Lễ Bùi Quang Huy Trần Trung Kiên
5	Làm sạch dữ liệu và trực quan hóa dữ liệu	Khuất Đình Lễ Bùi Quang Huy Trần Trung Kiên
6	Viết báo cáo	Bùi Quang Huy Trần Trung Kiên
7	Thuyết trình	Khuất Đình Lễ

## Theo thành viên

Stt	Tên thành viên	Tên nhiệm vụ	Chữ ký
1	Khuất Đình Lễ	Phân tích bài toán. Tìm hiểu công cụ PowerBI	
2	Trần Trung Kiên	Phân tích bài toán. Tìm hiểu công cụ PowerBI	
3	Bùi Quang Huy	Tìm hiểu về cơ sở lý thuyết. Tìm hiểu công cụ PowerBI	

## MỞ ĐẦU

Ngày nay, công nghệ thông tin phát triển đồng nghĩa với việc phát triển các phần mềm ứng dụng và nhu cầu về phân tích dữ liệu đồng thời cũng có thể thể hiện cho những người dùng khác về bài phân tích. Về phần thể hiện dữ liệu đã có phần mềm Microsoft Excel, về phần trình bày cho người xem chúng ta có Microsoft PowerPoint là những phần mềm phổ biến và nổi bật nhưng khi thể hiện dữ liệu và phân tích dữ liệu chúng ta cần một phần mềm khác chuyên sâu hơn và thể hiện dữ liệu rõ hơn. Một trong số những ứng dụng nổi bật về điều này đó là PowerBI.

Một trong những ứng dụng của PowerBI đó là phân tích dữ liệu của sự phát triển kinh tế và đối với những người mới bắt đầu với phân tích dữ liệu lớn thì phổ biến nhất đó là phân tích dữ liệu buôn bán của một cửa hàng. Chính vì

thế trong bài viết này, chúng ta hãy cũng tìm hiểu và phân tích dữ liệu lớn trong lĩnh vực kinh doanh và áp dụng phân tích tập dữ liệu Superstore.

## **CHƯƠNG 1. CƠ SỞ LÝ THUYẾT**

### **1.1. Nội dung**

Khái niệm, nguồn hình thành và đặc trưng của Big Data

#### **1.1.1 Khái niệm về dữ liệu lớn**

- Theo Wikipedia: Dữ liệu lớn (Big data) là một thuật ngữ chỉ bộ dữ liệu lớn hoặc phức tạp mà các phương pháp truyền thống không đủ các ứng dụng để xử lý dữ liệu này.

- Theo Gartner: Dữ liệu lớn là những nguồn thông tin có đặc điểm chung khối lượng lớn, tốc độ nhanh và dữ liệu định dạng dưới nhiều hình thức khác nhau, do đó muốn khai thác được đòi hỏi phải có hình thức xử lý mới để đưa ra quyết định, khám phá và tối ưu hóa quy trình.

## 1.2. Nguồn hình thành dữ liệu và phương pháp khai thác và quản lý dữ liệu lớn.



Qua thống kê và tổng hợp, nguồn dữ liệu lớn được hình thành chủ yếu từ 6 nguồn:

(1) Dữ liệu hành chính (phát sinh từ chương trình của một tổ chức, có thể là chính phủ hay phi chính phủ). Ví dụ, hồ sơ y tế điện tử ở bệnh viện, hồ sơ bảo hiểm, hồ sơ ngân hàng..;

(2) Dữ liệu từ hoạt động thương mại (phát sinh từ các giao dịch giữa hai thực thể). Ví dụ, các giao dịch thẻ tín dụng, giao dịch trên mạng, bao gồm cả từ các thiết bị di động;

(3) Dữ liệu từ các thiết bị cảm biến như thiết bị chụp hình ảnh vệ tinh, cảm biến đường, cảm biến khí hậu;

(4) Dữ liệu từ các thiết bị theo dõi, ví dụ theo dõi dữ liệu từ điện thoại di động, GPS;

(5) Dữ liệu từ các hành vi, ví dụ như tìm kiếm trực tuyến về (một sản phẩm, một dịch vụ hay thông tin khác), đọc các trang mạng trực tuyến..;

(6) Dữ liệu từ các thông tin về ý kiến, quan điểm của các cá nhân, tổ chức, trên các phương tiện thông tin xã hội.

Phương pháp khai thác và quản lý dữ liệu lớn hiện nay được thiết kế phù hợp dựa theo các nguồn hình thành dữ liệu lớn. Mỗi nguồn dữ liệu lớn khác nhau sẽ có phương pháp khai thác và quản lý dữ liệu lớn khác nhau. Tuy

hiện, hiện nay phần lớn các tổ chức trên thế giới đều dùng Hadoop ecosystem là giải pháp tối ưu để khai thác và quản lý dữ liệu lớn.

### 1.3. Đặc trưng 5V của dữ liệu lớn

Dữ liệu lớn có 5 đặc trưng cơ bản như sau (mô hình 5V):



- Khối lượng dữ liệu (Volume) :

Đây là đặc điểm tiêu biểu nhất của dữ liệu lớn, khối lượng dữ liệu rất lớn. Kích cỡ của Big Data đang từng ngày tăng lên, và tính đến năm 2012 thì nó có thể nằm trong khoảng vài chục terabyte cho đến nhiều petabyte (1 petabyte = 1024 terabyte) chỉ cho một tập hợp dữ liệu. Dữ liệu truyền thống có thể lưu trữ trên các thiết bị đĩa mềm, đĩa cứng. Nhưng với dữ liệu lớn chúng ta sẽ sử dụng công nghệ “đám mây” mới đáp ứng khả năng lưu trữ được dữ liệu lớn.

- Tốc độ (Velocity) :

Tốc độ có thể hiểu theo 2 khía cạnh :

- + Khối lượng dữ liệu gia tăng rất nhanh (mỗi giây có tới 72.9 triệu các yêu cầu truy cập tìm kiếm trên web bán hàng của Amazon);
- + Xử lý dữ liệu nhanh ở mức thời gian thực (real-time), có nghĩa dữ liệu được xử lý ngay tức thời ngay sau khi chúng phát sinh (tính đến bằng mili

giây). Các ứng dụng phổ biến trên lĩnh vực Internet, Tài chính, Ngân hàng, Hàng không, Quân sự, Y tế - Sức khỏe như hiện nay phần lớn dữ liệu lớn được xử lý real-time. Công nghệ xử lý dữ liệu lớn ngày nay đã cho phép chúng ta xử lý tức thì trước khi chúng được lưu trữ vào cơ sở dữ liệu.

- Đa dạng (Variety) :

Đối với dữ liệu truyền thống chúng ta hay nói đến dữ liệu có cấu trúc, thì ngày nay hơn 80% dữ liệu được sinh ra là phi cấu trúc (tài liệu, blog, hình ảnh, vi deo, bài hát, dữ liệu từ thiết bị cảm biến vật lý, thiết bị chăm sóc sức khỏe...). Big Data cho phép liên kết và phân tích nhiều dạng dữ liệu khác nhau.

Ví dụ: với các bình luận của một nhóm người dùng nào đó trên Facebook với thông tin video được chia sẻ từ Youtube và Twitter.

- Độ tin cậy/chính xác (Veracity) :

Một trong những tính chất phức tạp nhất của Dữ liệu lớn là độ tin cậy/chính xác của dữ liệu. Với xu hướng phương tiện truyền thông xã hội (Social Media) và mạng xã hội (Social Network) ngày nay và sự gia tăng mạnh mẽ tính tương tác và chia sẻ của người dùng Mobile làm cho bức tranh xác định về độ tin cậy & chính xác của dữ liệu ngày một khó khăn hơn. Bài toán phân tích và loại bỏ dữ liệu thiếu chính xác và nhiễu đang là tính chất quan trọng của BigData.

- Giá trị (Value) :

Giá trị là đặc điểm quan trọng nhất của dữ liệu lớn, vì khi bắt đầu triển khai xây dựng dữ liệu lớn thì việc đầu tiên chúng ta cần phải làm đó là xác định được giá trị của thông tin mang lại như thế nào, khi đó chúng ta mới có quyết định có nên triển khai dữ liệu lớn hay không. Nếu chúng ta có dữ liệu lớn mà chỉ nhận được 1% lợi ích từ nó, thì không nên đầu tư phát triển dữ liệu lớn. Kết quả dự báo chính xác thể hiện rõ nét nhất về giá trị của dữ liệu lớn mang lại.



Ví dụ: từ khối dữ liệu phát sinh trong quá trình khám, chữa bệnh sẽ giúp dự báo về sức khỏe được chính xác hơn, sẽ giảm được chi phí điều trị và các chi phí liên quan đến y tế.

#### **1.4. Sự khác biệt giữa dữ liệu lớn với dữ liệu truyền thống**

Dữ liệu lớn khác với dữ liệu truyền thống (ví dụ, kho dữ liệu - Data Warehouse) ở 4 điểm cơ bản: Dữ liệu đa dạng hơn; lưu trữ dữ liệu lớn hơn; truy vấn nhanh hơn; độ chính xác cao hơn.

- Dữ liệu đa dạng hơn:

Khi khai thác dữ liệu truyền thống (Dữ liệu có cấu trúc), chúng ta thường phải trả lời các câu hỏi: Dữ liệu lấy ra kiểu gì? định dạng dữ liệu như thế nào? Đối với dữ liệu lớn, không phải trả lời các câu hỏi trên. Hay nói khác, khi khai thác, phân tích dữ liệu lớn chúng ta không cần quan tâm đến kiểu dữ liệu và định dạng của chúng; điều quan tâm là giá trị mà dữ liệu mang lại có đáp ứng được cho công việc hiện tại và tương lai hay không.

- Lưu trữ dữ liệu lớn hơn:

Lưu trữ dữ liệu truyền thống vô cùng phức tạp và luôn đặt ra câu hỏi lưu như thế nào? Dung lượng kho lưu trữ bao nhiêu là đủ? Gắn kèm với câu hỏi đó là chi phí đầu tư tương ứng. Công nghệ lưu trữ dữ liệu lớn hiện nay đã phần nào có thể giải quyết được vấn đề trên nhờ những công nghệ lưu trữ đám mây, phân phối lưu trữ dữ liệu phân tán và có thể kết hợp các dữ liệu phân tán lại với nhau một cách chính xác và xử lý nhanh trong thời gian thực.

- Truy vấn dữ liệu nhanh hơn:

Dữ liệu lớn được cập nhật liên tục, trong khi đó kho dữ liệu truyền thống thì lâu lâu mới được cập nhật và trong tình trạng không theo dõi thường xuyên gây ra tình trạng lỗi cấu trúc truy vấn dẫn đến không tìm kiếm được thông tin đáp ứng theo yêu cầu.

- Độ chính xác cao hơn:

Dữ liệu lớn khi đưa vào sử dụng thường được kiểm định lại dữ liệu với những điều kiện chặt chẽ, số lượng thông tin được kiểm tra thông thường rất lớn, và đảm bảo về nguồn lấy dữ liệu không có sự tác động của con người vào thay đổi số liệu thu thập.

### **1.5 Ứng dụng của big data trong lĩnh vực kinh doanh**

Big Data (Dữ Liệu Lớn) có ứng dụng rất lớn trong thế giới kinh doanh. Chính vì thế những người làm kinh doanh cần tìm hiểu về khoa học dữ liệu để hiểu đầy đủ về tác động của Big Data đến môi trường kinh doanh và cách tạo ra giá trị thực thông qua phân tích.

#### **Quản lý rủi ro**

Mỗi doanh nghiệp cần có một cách tiếp cận chiến lược để quản lý rủi ro. Có rất nhiều cách tiếp cận, nhưng dữ liệu lớn là một trong những cách quan trọng nhất. Nó cho phép các công ty định lượng và mô hình hóa các rủi ro có thể gặp trong hoạt động của công ty. Với việc sử dụng Big Data, các công ty có thể dựa vào phân tích dự đoán để có tầm nhìn xa về rủi ro. Nhờ đó, doanh nghiệp sẽ dễ dàng thực hiện các chiến lược giảm thiểu rủi ro hơn.

#### **Hiểu khách hàng tốt hơn**

Các doanh nghiệp phải hiểu khách hàng của mình và Big Data sẽ giúp điều này trở nên khả thi. Từ lượt truy cập trang web đến các tương tác trên mạng xã hội, Big Data sẽ tiết lộ nhiều điều về khách hàng. Nó cũng sẽ cho phép doanh nghiệp tạo hồ sơ khách hàng và tính cách người mua. Khi doanh nghiệp hiểu họ hơn, doanh nghiệp có thể cải thiện sản phẩm và dịch vụ của mình. Nó cho phép doanh nghiệp cung cấp mức độ hài lòng cao nhất, giúp xây dựng lòng trung thành của khách hàng.

### **Hiểu đối thủ**

Dữ liệu lớn cũng mang lại cơ hội tìm hiểu đối thủ cạnh tranh của doanh nghiệp. Ví dụ: nó có thể cung cấp thông tin về các mô hình giá cả của đối thủ và cảm nhận của khách hàng. Doanh nghiệp cũng sẽ hiểu được nhận thức của khách hàng về đối thủ cạnh tranh. Thêm vào đó, nó sẽ giúp doanh nghiệp xác định cách hoạt động trực tuyến, chẳng hạn như mức độ tương tác trên mạng xã hội.

### **Cá nhân hóa hoạt động tiếp thị**

Tiếp thị có thể tạo nên hoặc phá vỡ doanh nghiệp. Để thành công, một trong những điều quan trọng nhất là cá nhân hóa. Đây là một lĩnh vực khác mà doanh nghiệp có thể sử dụng Big Data. Bởi vì nó cho phép doanh nghiệp hiểu khách hàng của mình, từ đó có thể tạo các chiến dịch tiếp thị nhắm mục tiêu đến một thị trường ngách hoặc phân khúc cụ thể. Nó có thể cung cấp thông tin chi tiết cho phép các nhà tiếp thị tạo ra các chiến dịch có khả năng chuyển đổi cao.

### **Xác định xu hướng**

Thông qua dữ liệu lớn, doanh nghiệp cũng có thể xác định được xu hướng, có thể hữu ích trong việc nghiên cứu và phát triển sản phẩm. Từ hành vi của khách hàng đến mô hình mua hàng, Big Data sẽ cung cấp cho ban quản lý thông tin cần thiết để phân tích các xu hướng sẽ thay đổi như thế nào theo thời gian và do đó sẽ cho doanh nghiệp thời gian để chuẩn bị cho những thay đổi này.

### **Quản lý nguồn nhân lực**

Việc sử dụng dữ liệu lớn sẽ giúp doanh nghiệp phát triển một cách tiếp cận mạnh mẽ để quản lý nguồn nhân lực. Từ quyết định tuyển dụng đến đào tạo và phát triển, Big Data sẽ giúp doanh nghiệp có những nhân viên giỏi nhất. Từ sự

hài lòng của nhân viên đến dữ liệu năng suất, doanh nghiệp có quyền truy cập vào thông tin mà doanh nghiệp cần để hiểu và cải thiện lực lượng lao động.

Big Data có vô số ứng dụng trong các doanh nghiệp. Các doanh nghiệp cần khám phá các mô hình, mối tương quan ẩn và có được thông tin chi tiết từ dữ liệu lớn không có cấu trúc để đưa ra các quyết định kinh doanh đúng đắn. Từ quản lý rủi ro đến quản lý nguồn nhân lực, doanh nghiệp có thể sử dụng dữ liệu lớn theo nhiều cách để cải thiện hoạt động kinh doanh. Vì vậy, hãy đảm bảo có cơ sở hạ tầng phù hợp để quản lý dữ liệu lớn.

## **CHƯƠNG 2. CÔNG CỤ POWER BI**

### **2.1. Nội dung**

#### **2.1.1 Power BI là gì?**

Power BI là tên gọi chung của một bộ các ứng dụng và dịch vụ trên nền tảng đám mây, có giao diện vô cùng thân thiện với người dùng. Nó là một công cụ BI và trực quan hóa dữ liệu từ các nguồn khác nhau để trở thành một bảng dashboard có thể tương tác trực tiếp và các báo cáo phân tích.

Các công cụ BI như Power BI có thể được dùng với nhiều mục đích khác nhau:

- + Power BI có thể kéo và tổng hợp các dữ liệu về lại 1 nơi và xử lý các dữ liệu đó trở thành các thông tin dễ hiểu hơn (thông thường là các hình ảnh bắt mắt,

đồ thị, biểu đồ). Điều này giúp cho người dùng có thể xây dựng và chia sẻ những gì đang diễn ra trong doanh nghiệp của họ.

- + Power BI có thể kết nối với nhiều nguồn dữ liệu khác nhau, từ Excel cho đến các cơ sở dữ liệu trên đám mây, cũng như trên các ứng dụng.

- + Bạn có thể sử dụng Power BI trên màn hình desktop với (Power BI desktop), dịch vụ SaaS online tên là Power BI service và các ứng dụng Power BI trên hệ điều hành Windows, iOS và Android.

- + Power BI được xây dựng trên nền tảng của Excel, nên nếu bạn có thể sử dụng Excel thì việc sử dụng Power BI cũng không quá khó khăn.

Những ai cần sử dụng Power BI?

Power BI là công cụ dành cho những chuyên gia như:

- + PMO – Project and Portfolio Manager (Quản lý dự án và danh mục đầu tư).

- + Business & Data Analyst (Chuyên viên phân tích dữ liệu và doanh nghiệp)

- + Developer & Database Administrator (Chuyên viên cơ sở dữ liệu và phát triển).

- + IT Team, IT Professional .

- + Consumer for End User Report (Tạo báo cáo cho Người dùng cuối).

- + Data Scientist

Ưu điểm của Power BI :

- + Cung cấp bảng dashboard và báo cáo được xây dựng sẵn, là giải pháp dành cho SaaS.

- + Đưa ra các bảng dashboard được cập nhật real-time.

- + Việc kết nối đến nguồn dữ liệu qua đám mây hoặc ngay trên hệ thống của doanh nghiệp, được bảo mật và đáng tin cậy.

- + Triển khai nhanh, đa cấu hình, và có môi trường bảo mật.

- + Khai thác dữ liệu thông qua các ngôn ngữ truy vấn tự nhiên.

- + Có khả năng xây dựng dashboard được trực quan hóa.
- + Có nhiều tính năng phù hợp với người dùng Excel.
- + Các câu hỏi về khả năng kết nối với cơ sở dữ liệu mở rộng được công bố trên website.
- + Tích hợp cả các đoạn code Python và R dùng để trực quan hóa.
- + Power Query cung cấp nhiều lựa chọn liên quan đến việc sử dụng và làm sạch dữ liệu.
- + Khi bạn tải các dữ liệu vào dịch vụ web của Power BI, bạn có thể thiết lập lịch trình để cập nhật dữ liệu, mà không cần phải can thiệp thủ công.
- + Power BI được tạo nên từ nền tảng của AI (trí tuệ thông minh) và ML (máy học).

Nhược điểm của Power BI :

- + Dashboard và báo cáo chỉ có thể được chia sẻ với những người dùng có cùng tên miền email.
- + Power BI sẽ không kết hợp các dữ liệu được nhập vào từ những kết nối theo thời gian thực.
- + Power BI không chấp nhận những file lớn hơn 250MB và các file được nén bởi các dữ liệu X trong các cơ sở dữ liệu trong bộ nhớ.
- + Power BI không chấp nhận những file lớn hơn 1GB.
- + Các dashboard này sẽ không thể chia sẻ cho người dùng, tài khoản khác hoặc các thông số thực tế khác.
- + Có rất ít nguồn dữ liệu cho phép kết nối real-time với các báo cáo và dashboard trên Power BI.

### **2.1.2 DAX là gì ?**

DAX là viết tắt của cụm từ Data Analysis Expressions là tập hợp các hàm, toán tử và hằng số, được sử dụng để giải quyết những bài toán từ cơ bản đến phức tạp. Từ đó, trả về một hoặc nhiều giá trị từ dữ liệu có sẵn. Một cách dễ hiểu, DAX được coi là hàm trong Power BI. Tuy nhiên, vượt xa khả năng của

các hàm thuần túy, với các chức năng được xây dựng chuyên biệt để làm việc với các data model, DAX rất mạnh mẽ và linh hoạt



### 03 yếu tố của DAX :

Bạn cần biết rằng có rất nhiều yếu tố quan trọng khác nhưng đây là 3 vấn đề quan trọng mà nhất định phải nắm. Đó là **Syntax**(cú pháp), **Function**(chức năng) và **Context**(ngữ cảnh).

+ Syntax có thể hiểu là cú pháp, trước khi tạo ra các công thức tính toán, bạn cần biết cú pháp của Power BI. Dưới đây là một ví dụ công thức DAX đơn giản cho một measure.

+ Function được định nghĩa là các công thức được xác định trước, thực hiện những phép tính bằng các giá trị cụ thể được gọi là đối số, theo một thứ tự hoặc cấu trúc cụ thể.

+ Context là ngữ cảnh sử dụng DAX. Có 2 loại Context trong DAX đó là Row Context và Filter Context.

- Row Context áp dụng bất cứ khi nào một công thức có một hàm áp dụng bộ lọc để xác định một hàng trong bảng

- Filter Context giống như là một hoặc nhiều bộ lọc được áp dụng trong một phép tính mà được xác định kết quả hoặc giá trị.

\* So sánh : DAX (Power BI) và Hàm (Excel)

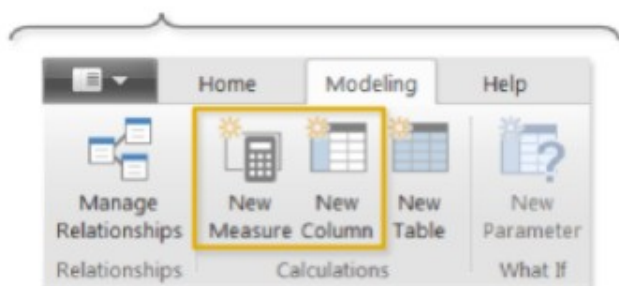
<b>DAX</b>	<b>Hàm</b>
Thường quy định kiểu dữ liệu đầu vào	Thường không quy định, nếu sai thì báo lỗi
Chỉ tham chiếu theo Cột hoặc Bảng	Tham chiếu theo Ô-vùng và cả Cột-Bảng
Kết quả trả về có thể là một Giá trị, một Bảng	Kết quả trả về là một Giá trị



DAX yêu cầu tạo ra các mối quan hệ giữa các bảng để sử dụng	Có thể làm việc với các bảng độc lập
---	--------------------------------------

## THÊM COLUMNS & MEASURES

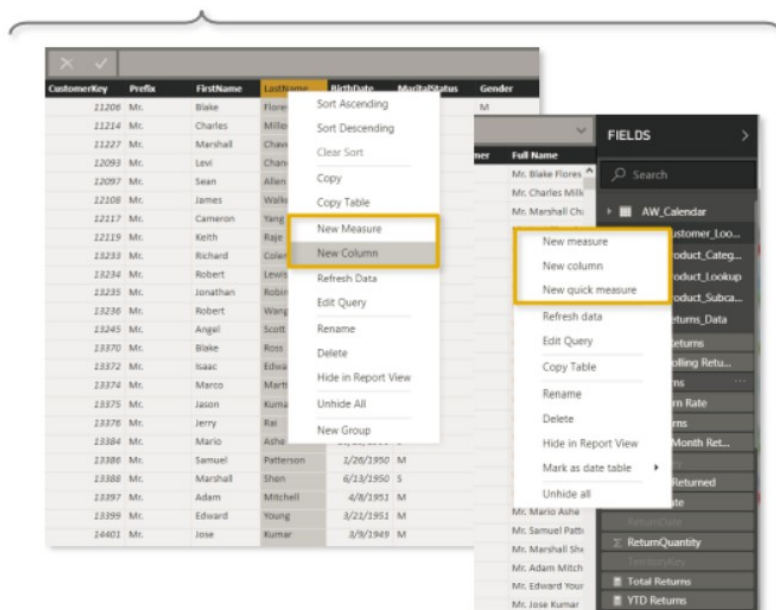
Option 1: Chọn “New Measure” hoặc "New Column" trong tab Modeling



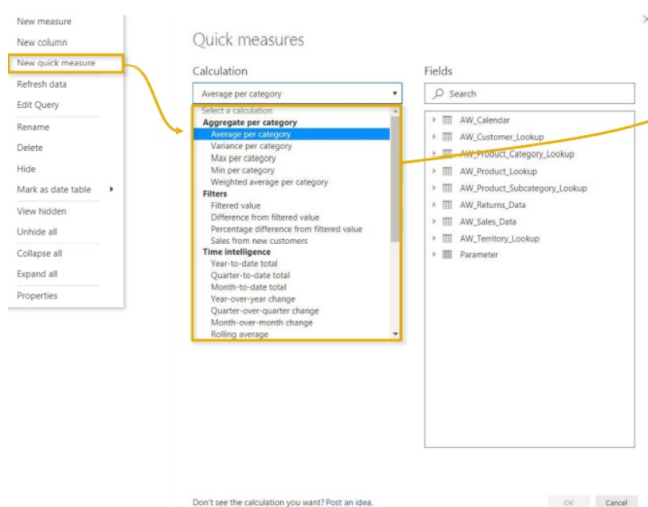
Khi bạn chèn một Columns hoặc Measures sử dụng tab **Modeling**, PowerBI sẽ mặc định bảng đang chọn làm dữ liệu tham chiếu, hoặc bảng đầu tiêu trong list thông thường

- Measures có thể chọn lại tables (trong phần “Properties” options trong tab **Modeling**), nhưng Option 2 sẽ cho phép bạn có sự chủ động hơn về vị trí và về bảng

Option 2: Chuột phải trong table (trong Data view) hoặc trong Field List (cả Data và Report view)



## QUICK MEASURES



**Quick Measures** là các công thức dựng trước dạng templates, cho phép bạn kéo và thả vào các trường, hơn là phải viết DAX từ đầu.

Mặc dù công cụ này khá hữu ích khi sử dụng, đặc biệt là với các hàm có tính chất phức tạp (như weighted averages hay time intelligence...), nhưng nó sẽ khiến bạn lười hơn và đôi khi không hiểu rõ về DAX.

## CALCULATED COLUMNS

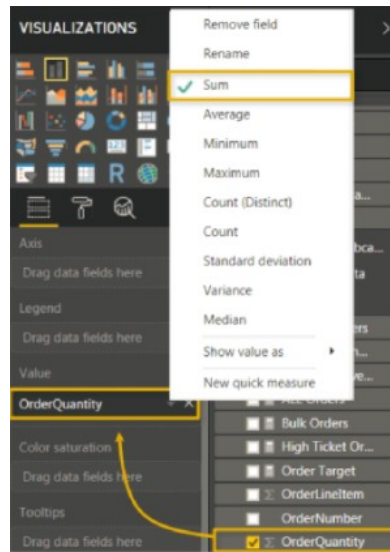
**Calculated columns** cho phép bạn thêm mới cột vào data, dựa trên các biểu thức tính toán từ các cột đã có sẵn.

- Giống như calculated columns, measures cũng tham chiếu tới **tables** và **columns**.
- Tuy nhiên, **measure** không hiện hữu giá trị trong bảng; nó chỉ có thể “nhìn thấy” khi visualize như khi sử dụng chart hay matrix(giống như một trường tính toán trong Excel pivot).
- Measures tính toán dựa trên các dữ liệu lọc từ bảng, nghĩa là nó sẽ cập nhật lại khi các trường hoặc dữ liệu lọc từ bảng thay đổi.

## IMPLICIT vs EXPLICIT MEASURES

**Measures ẩn** là các measures được tạo ra khi bạn kéo các trường dữ liệu số (như “OrderQuantity”) trong trình tạo visual và chọn một dạng tính toán nào đó (Sum, Average, Min/Max, ...)

**Measures hiện** là các measures được tạo ra bằng việc nhập các hàm DAX (hoặc thêm một “quick measures”) để xác định giá trị.



Ví dụ một *implicit measure*

## RECAP: CALCULATED COLUMNS vs MEASURES

CALCULATED COLUMNS	MEASURES
Giá trị tính toán dựa trên tham chiếu tới vị trí theo hàng trong cột	Giá trị được tính toán dựa vào kết quả lọc từ các dữ liệu trong
Nối các giá trị tính trong bảng và lưu trữ trong model( làm tăng kích thước file)	Không tạo ra các data mới trong model( không làm tăng kích thước file)
Cập nhật khi data source được refresh	Cập nhật khi các kết quả lọc từ các

hoặc thay đổi được tạo ra trong bảng.	dữ liệu thay đổi trong báo cáo
Thường sử dụng để tạo thêm các columns, slicers or filters	hường được sử dụng trong các values lên báo cáo hoặc biểu đồ

## HIỂU VỀ FILER CONTEXT

Chúng ta luôn nhớ rằng, measures được tính toán dựa trên filter context, có nghĩa là nó sẽ tính toán lại khi các trường hoặc filter thay đổi.

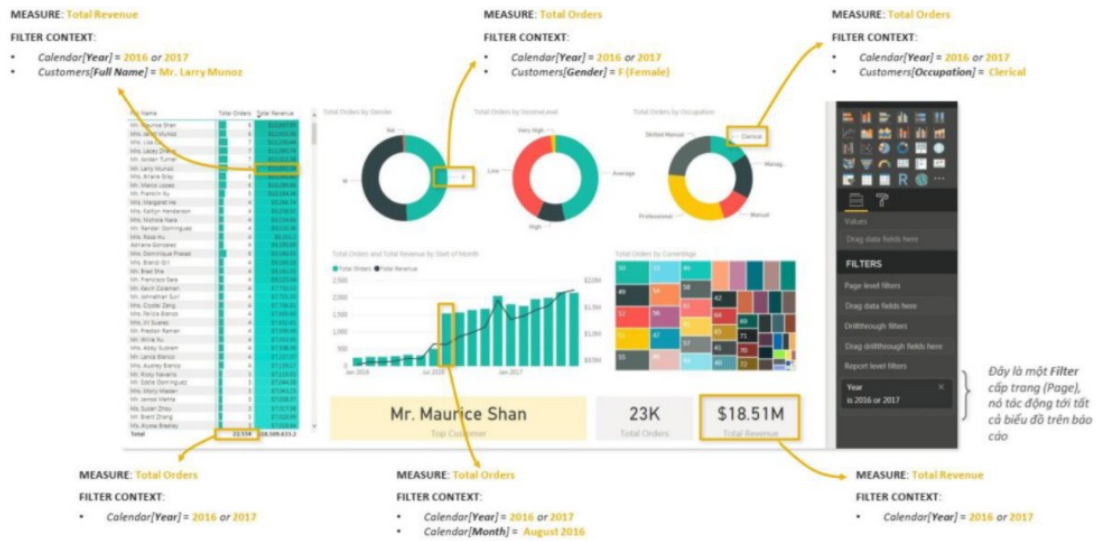
ProductName	Total Orders	Return Rate
Water Bottle - 30 oz.	1,344	1.98 %
Road Tire Tube	829	1.63 %
AWC Logo Cap	603	1.93 %
Patch Kit/8 Patches	789	1.97 %
Sport-100 Helmet, Red	729	2.79 %
Touring Tire Tube	722	1.35 %
Sport-100 Helmet, Blue	668	3.15 %
Sport-100 Helmet, Black	626	3.67 %
Road Bottle Cage	560	1.58 %
Mountain Tire Tube	554	1.95 %
Mountain Bottle Cage	539	1.38 %
Touring Tire	427	1.16 %
LL Road Tire	421	2.02 %
Fender Set - Mountain	378	1.82 %
ML Road Tire	297	1.72 %
ML Mountain Tire	286	2.94 %
HL Mountain Tire	208	2.40 %
Mountain-200 Silver, 46	199	1.51 %
Mountain-200 Black, 46	196	3.06 %
LL Mountain Tire	195	2.09 %
Mountain-200 Silver, 38	189	2.65 %
Bike Wash - Dissolver	187	2.38 %
Mountain-200 Black, 42	182	3.85 %
Mountain-200 Black, 38	180	3.33 %
Long-Sleeve Logo Jersey, M	161	4.95 %
HL Road Tire	151	5.06 %
Mountain-200 Silver, 42	151	1.28 %
Hydration Pack - 70 oz.	147	4.08 %
Long-Sleeve Logo Jersey, L	147	2.72 %
Long-Sleeve Logo Jersey, S	146	2.31 %
Total	7,389	2.17 %

Một ví dụ cụ thể trong matrix, **Total Orders** measure được tính dựa trên **filter context** sau: **Products[ProductName] = "Touring Tire Tube"**

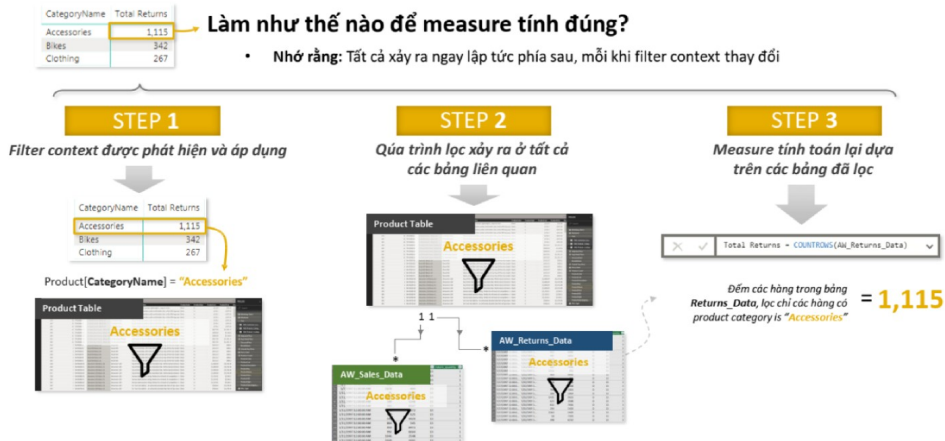
- Nó cho phép measure trả về kết quả **Total order** cho mỗi sản phẩm cụ thể theo tên sản phẩm (hoặc bất kỳ hàng hoặc cột có nhãn riêng biệt – *product categories, customer names, ...*)

Giá trị này **không** tính toán bởi nguyên tắc như ở trên; nó vẫn là một measure, không dùng **filter context** (nghĩa là chúng ta không thể tính toán cho một sản phẩm cụ thể)

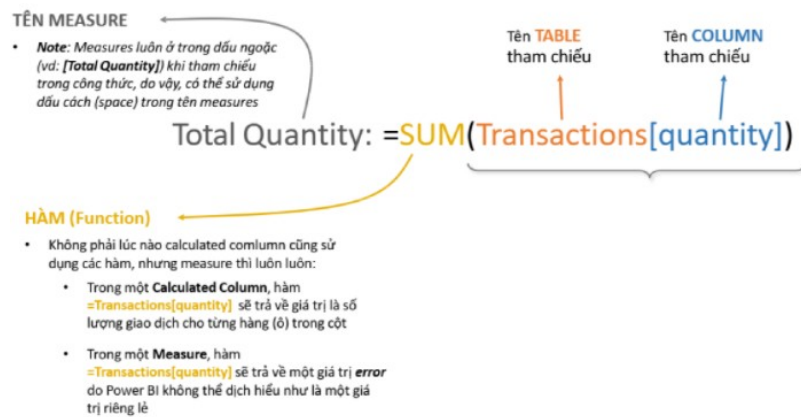
## FILTER CONTEXT (ví dụ)



## HIỂU CÁC BƯỚC KHI CÓ MEASURE CALCULATION



## Cấu trúc của Dax



## Dấu câu trong Dax

Loại số học	Ý nghĩa	Ví dụ
+	Cộng	2 + 7
-	Trừ	5 - 3
*	Nhân	2 * 6
/	Chia	4 / 2
^	Mũ	2 ^ 5

Loại so sánh	Ý nghĩa	Ví dụ
=	Bằng	[City]="Boston"
>	Lớn hơn	[Quantity]>10
<	Nhỏ hơn	[Quantity]<10
>=	Lớn hơn hoặc bằng	[Unit_Price]>=2.5
<=	Nhỏ hơn hoặc bằng	[Unit_Price]<=2.5
<>	Khác	[Country]<>"Mexico"

*Chú ý tới 02 biểu thức này*

Loại Text/Logic	Ý nghĩa	Ví dụ
&	Nối 02 giá trị dạng text thành 01 giá trị	[City] & " " & [State]
&&	Tạo một điều kiện VÀ giữa 02 biểu thức logic	(([State]="MA") && ([Quantity]>10))
(double pipe)	Tạo một điều kiện HOẶC giữa 02 biểu thức logic	(([State]="MA")    ([State]="CT"))
IN	Tạo một điều kiện HOẶC logic dựa trên danh sách đã cho (sử dụng dấu ngoặc)	'Store Lookup'[State] IN { "MA", "CT", "NY" }

## Các nhóm hàm Dax phổ biến

MATH & STATS functions : Hàm tính toán cơ bản cũng như tính toán trên vòng lặp sử dụng ngữ cảnh hàng.

Ví dụ : SUM, AVERAGE, MAX/MIN, DIVIDE,....

LOGICAL functions : Hàm trả về thông tin về các giá trị trong biểu thức dựa trên điều kiện đã cho

Ví dụ : IF, IFERROR, AND, OR, NOT, SWITCH, TRUE, FALSE

TEXT functions : Các hàm để thao tác các chuỗi văn bản hoặc các định dạng điều kiện cho ngày, giờ hoặc số.

Ví dụ : LEN, TRIM, REPT, FORMAT, CONCATENATE,...

FILTER functions : Các hàm tra cứu dựa trên các bảng có liên quan và các hàm lọc để tính toán

Ví dụ : CALCULATE, FILTER, ALL, VALUES, DISTINCT,...

DATE & TIME functions : Các chức năng này và giờ cơ bản cũng như các hàm thời gian nâng cao

Ví dụ : DATEDIFF, YEARRAC, YEAR/MONTH/DAY, TODAY/NOW



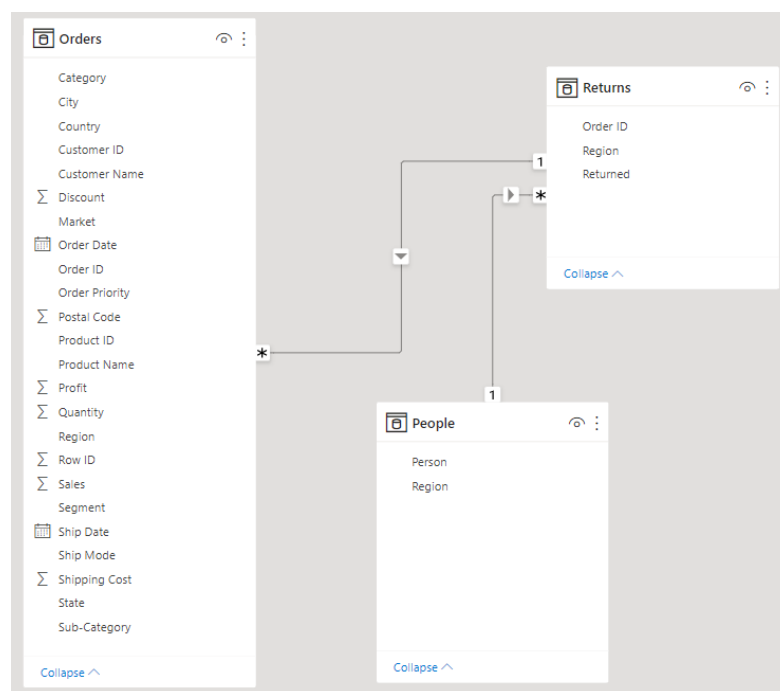
# CHƯƠNG 3: Bài toán phân tích tập dữ liệu Superstore trên Power BI

## 3.1 Mô tả bài toán

Đây là tập dữ liệu bán hàng của một cửa hàng từ năm 2012 đến năm 2015. Nó cho thấy sự chuyển dịch cơ cấu của hàng xuất ra hàng năm và sự phát triển của cửa hàng theo từng mức thời gian. Và cho thấy sự thay đổi nhu cầu của người tiêu dùng theo thời gian dựa vào đó ta dự đoán được hướng phát triển của cửa hàng trong tương lai

## 3.2 Mô tả và phân tích tập dữ liệu

Tập dữ liệu gồm 3 bảng : Orders, Returns, People.



Hình Relationship giữa các bảng

Mối quan hệ giữa các bảng :

- Return với Orders : 1 – any ( liên kết qua trường Order ID )
- People với Return : 1 – any ( liên kết qua trường Region )

Thông tin chi tiết của các bảng :

Bảng Order : Nêu các thông tin như order id, row id, customer id, product id,

Các thông tin ngày tháng của đơn hàng (ngày, tháng, năm) nhận và giao hàng (Order date, Ship date), Mức độ yêu tiên giao hàng (Ship mode), Các thông tin của khách hàng (Customer Name, Segment, City, State, Country, Region) , Thị trường (Market ) ,Thông tin mặt hàng (Product name, Sales, Quantity, Discount, Profit, Shipping Cost, Order Priority, Sub-category)

Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Postal Code	City	State	Country	Region	Market	Pro
12355	ES-2013-MC1813045-41597	19 November 2013	24 November 2013	Standard Class	MC-1813045	Mike Caulfe	Corporate	0	Argenteuil	Ile-de-France	France	Western Europe	Europe	
11090	ES-2015-SG2060545-42199	14 July 2015	20 July 2015	Standard Class	SG-2060545	Speros Goranitis	Consumer	0	Vitry-sur-Seine	Ile-de-France	France	Western Europe	Europe	
17322	ES-2014-AS1009045-41905	23 September 2014	27 September 2014	Standard Class	AS-1009045	Adam Shillingsburg	Consumer	0	Paris	Ile-de-France	France	Western Europe	Europe	
17388	ES-2013-JK1537045-41496	10 August 2013	17 August 2013	Standard Class	JK-1537045	Jay Kimmel	Consumer	0	Domont	Ile-de-France	France	Western Europe	Europe	
19888	ES-2012-AB1015045-40933	25 January 2012	30 January 2012	Standard Class	AB-1015045	Aimee Bixby	Consumer	0	Maisons-Alfort	Ile-de-France	France	Western Europe	Europe	
14449	ES-2014-ED1388545-41907	25 September 2014	29 September 2014	Standard Class	ED-1388545	Emily Ducich	Home Office	0	Neuilly-sur-Marne	Ile-de-France	France	Western Europe	Europe	
17939	ES-2014-ND1837045-41977	04 December 2014	08 December 2014	Standard Class	ND-1837045	Natalie DeCherney	Consumer	0	Eragny	Ile-de-France	France	Western Europe	Europe	
19668	ES-2013-SN2071045-41544	27 September 2013	03 October 2013	Standard Class	SN-2071045	Steve Nguyen	Home Office	0	Pontault-Combault	Ile-de-France	France	Western Europe	Europe	
13813	ES-2014-PM1913545-41909	27 September 2014	03 October 2014	Standard Class	PM-1913545	Peter McVee	Home Office	0	Franconville	Ile-de-France	France	Western Europe	Europe	
12790	ES-2015-JC1534045-42335	27 November 2015	04 December 2015	Standard Class	JC-1534045	Jasper Cacioppo	Consumer	0	Paris	Ile-de-France	France	Western Europe	Europe	
11661	IT-2015-TB2105545-42334	26 November 2015	03 December 2015	Standard Class	TB-2105545	Ted Butterfield	Consumer	0	Vincennes	Ile-de-France	France	Western Europe	Europe	
13862	ES-2014-MW1823545-41965	22 November 2014	26 November 2014	Standard Class	MW-1823545	Mitch Willingham	Corporate	0	Le Blanc-Mesnil	Ile-de-France	France	Western Europe	Europe	
11306	ES-2015-YC2189545-42326	18 November 2015	22 November 2015	Standard Class	YC-2189545	Yoseph Carroll	Corporate	0	Le Plessis-Robinson	Ile-de-France	France	Western Europe	Europe	
16137	ES-2014-CM1244545-41973	30 November 2014	07 December 2014	Standard Class	CM-1244545	Chuck Magee	Consumer	0	Paris	Ile-de-France	France	Western Europe	Europe	
11936	ES-2014-FW1439545-41957	14 November 2014	18 November 2014	Standard Class	FW-1439545	Fred Wasserman	Corporate	0	Bobigny	Ile-de-France	France	Western Europe	Europe	
16528	ES-2013-SC2009545-41391	27 April 2013	01 May 2013	Standard Class	SC-2009545	Sanjit Chand	Consumer	0	Chelles	Ile-de-France	France	Western Europe	Europe	
10330	ES-2014-SC2002045-41790	31 May 2014	05 June 2014	Standard Class	SC-2002045	Sam Craven	Consumer	0	Chaville	Ile-de-France	France	Western Europe	Europe	
13826	ES-2013-AH1021045-41312	07 February 2013	12 February 2013	Standard Class	AH-1021045	Alan Hwang	Consumer	0	Argenteuil	Ile-de-France	France	Western Europe	Europe	
13968	ES-2013-AD1018045-41614	06 December 2013	10 December 2013	Standard Class	AD-1018045	Alan Dominguez	Home Office	0	Antony	Ile-de-France	France	Western Europe	Europe	
13866	ES-2014-MW1823545-41965	22 November 2014	26 November 2014	Standard Class	MW-1823545	Mitch Willingham	Corporate	0	Le Blanc-Mesnil	Ile-de-France	France	Western Europe	Europe	
15163	ES-2015-GA1451545-42290	13 October 2015	17 October 2015	Standard Class	GA-1451545	George Ashbrook	Consumer	0	Paris	Ile-de-France	France	Western Europe	Europe	
15186	IT-2012-CM1219045-41269	26 December 2012	31 December 2012	Standard Class	CM-1219045	Charlotte Melton	Consumer	0	Villeparisis	Ile-de-France	France	Western Europe	Europe	
14095	ES-2015-GZ1447045-42107	13 January 2015	17 January 2015	Standard Class	GZ-1447045	Gary Zandusky	Consumer	0	Paris	Ile-de-France	France	Western Europe	Europe	
18000	ES-2015-KA1652545-42153	29 May 2015	02 June 2015	Standard Class	KA-1652545	Kelly Andreada	Consumer	0	Clichy	Ile-de-France	France	Western Europe	Europe	
11819	ES-2014-PM1894045-41723	25 March 2014	01 April 2014	Standard Class	PM-1894045	Paul Macintyre	Consumer	0	Versailles	Ile-de-France	France	Western Europe	Europe	
11660	IT-2015-TB2105545-42334	26 November 2015	03 December 2015	Standard Class	TB-2105545	Ted Butterfield	Consumer	0	Vincennes	Ile-de-France	France	Western Europe	Europe	
19515	ES-2012-CC1255045-41038	09 May 2012	14 May 2012	Standard Class	CC-1255045	Clay Cheatham	Consumer	0	Champigny-sur-Marne	Ile-de-France	France	Western Europe	Europe	

Bảng Return : Nêu các thông tin về đơn hàng được trả lại (Order ID, Region),  
 Tình trạng đơn hàng ( Returned )

Returned ▾	Order ID ▾	Region ▾
Yes	ES-2015-RA1994545-42218	Western Europe
Yes	ES-2012-SC208458-41070	Western Europe
Yes	ES-2015-CC1210045-42182	Western Europe
Yes	ES-2015-MM1792045-42199	Western Europe
Yes	ES-2015-BB1154548-42336	Western Europe
Yes	ES-2014-EG1390045-41644	Western Europe
Yes	ES-2013-DB1297045-41458	Western Europe
Yes	IT-2012-MP1817545-41040	Western Europe
Yes	IT-2015-AR1082545-42290	Western Europe
Yes	ES-2015-BN1151545-42304	Western Europe
Yes	ES-2012-AG1049545-41002	Western Europe
Yes	ES-2015-LR1691545-42237	Western Europe
Yes	ES-2012-TT2146045-41014	Western Europe
Yes	ES-2015-RA1994548-42249	Western Europe
Yes	IT-2015-KD1634545-42238	Western Europe

Bảng People : Nêu họ và tên chủ chi nhánh và khu vực quản lý  
 ( Person , Region)

Person ▾	Region ▾
Marilène Rousseau	Caribbean
Andile Ihejirika	Central Africa
Nicodemo Bautista	Central America
Cansu Peynirci	Central Asia
Lon Bonher	Central US
Wasswa Ahmed	Eastern Africa
Hadia Bousaid	Eastern Asia
Lynne Marchand	Eastern Canada
Oxana Lagunov	Eastern Europe
Dolores Davis	Eastern US
Lindiwe Afolayan	North Africa
Miina Nylund	Northern Europe
Kauri Anaru	Oceania
Vasco Magalhães	South America
Preecha Metharom	Southeastern Asia
Nora Cuijper	Southern Africa

### 3.2 Phân tích bài toán.

- **Tổng số khách hàng:**



-Sử dụng hàm Count để đếm tên những khách hàng có thông tin cá nhân khác nhau (nếu trùng tên thì so sánh các thông tin khác nếu giống nhau hoàn toàn thì là cùng một khách hàng).

-Cho thấy số lượng khách hàng đã đặt hàng.

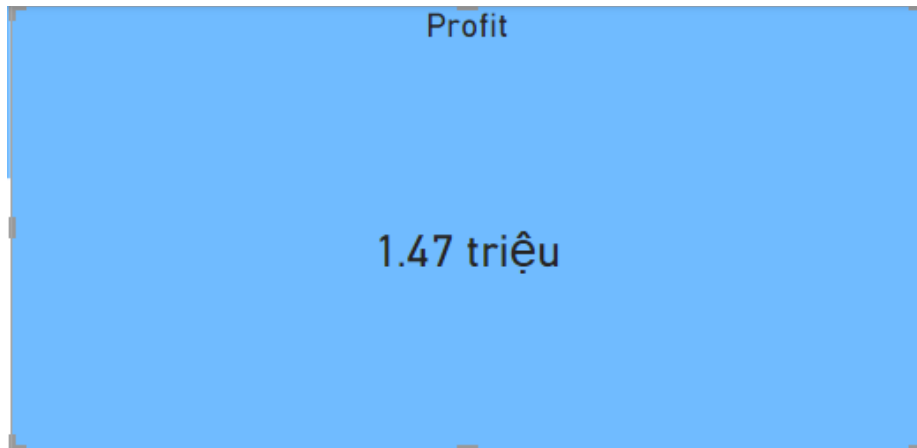
- **Số nơi đã đặt hàng:**



-Sử dụng hàm Count để đếm những nơi đặt hàng trên thế giới.

-Cho thấy số lượng những nơi đã và đang đặt hàng từ cửa hàng trên thế giới.

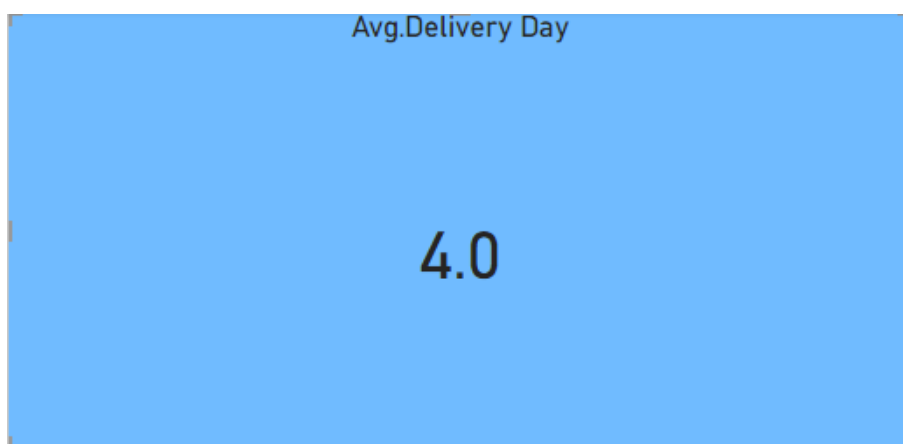
- **Lợi nhuận của cửa hàng**



-Sử dụng hàm Sum để cộng lại tổng giá trị hàng bán được và trừ đi chi phí chế tạo và chi phí vận chuyển.

-Cho thấy lời nhuận của cửa hàng.

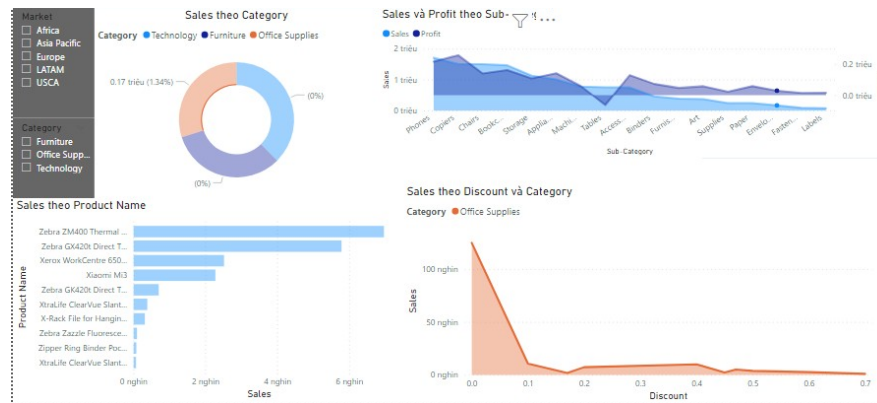
- **Số ngày giao hàng trung bình**



-Sử dụng hàm Averagex để tính trung bình cộng số ngày để sản phẩm tới tay khách hàng.

-Cho thấy số ngày trung bình để sản phẩm đến tay khách hàng.

- **Bảng so sánh giữa các khu vực.**



-Phần market để chọn thị trường, phần Category để chọn loại hàng hóa muốn xem:

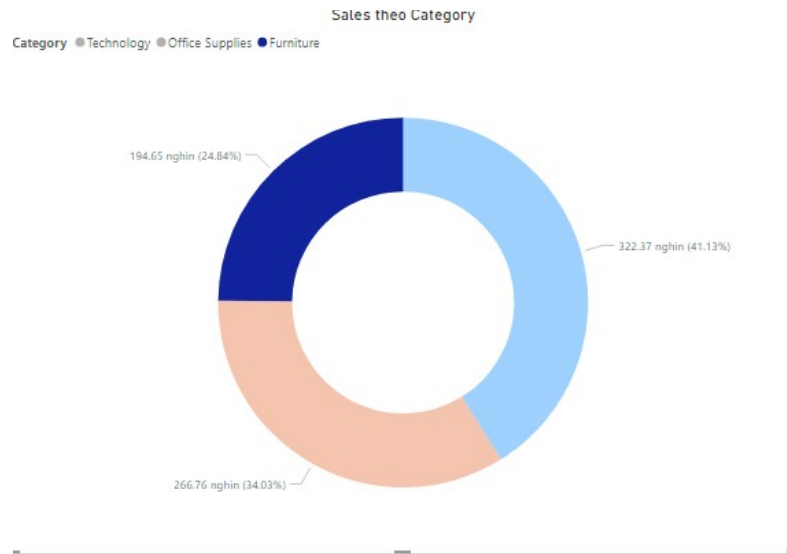
Market

☐ Africa
 ☐ Asia Pacific
 ☐ Europe
 ☐ LATAM
 ☐ USCA

Category

☐ Furniture
 ☐ Office Supp...
 ☐ Technology

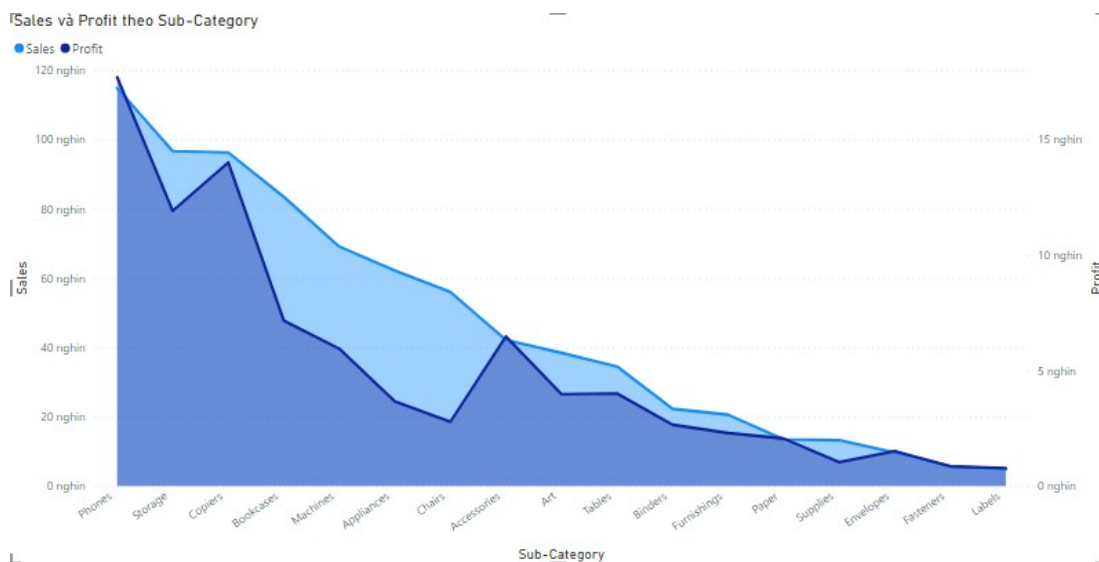
-Biểu đồ sales theo category cho thấy tỉ lệ người dùng mua hàng theo loại hàng của khu vực đã chọn.



+Biểu đồ người dùng mua hàng theo loại hàng tại khu vực (trên là ví dụ tại khu vực africa).

+Biểu đồ này dùng hàm Sum để tính tổng số hàng trong một loại khách hàng đã mua và so sánh trong tổng số hàng bán được trong biểu đồ vòng.

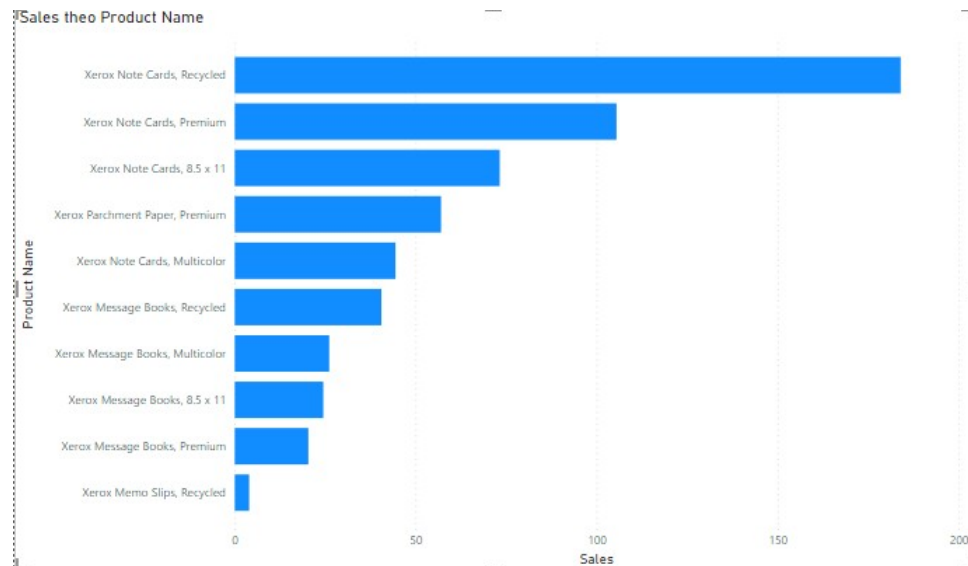
-Biểu đồ so sales và frofit theo sub-category:



+Cho thấy sự chênh lệch giữa giá trị hàng hóa và lời nhuận của hàng hóa đối với mỗi loại hàng cụ thể trong một khu vực (trên hình là ví dụ tại khu vực Africa).

+Sử dụng hàm sum để tính profit và sales sau đó sử dụng biểu đồ vùng để hiển thị.

-Biểu đồ sales theo product name:

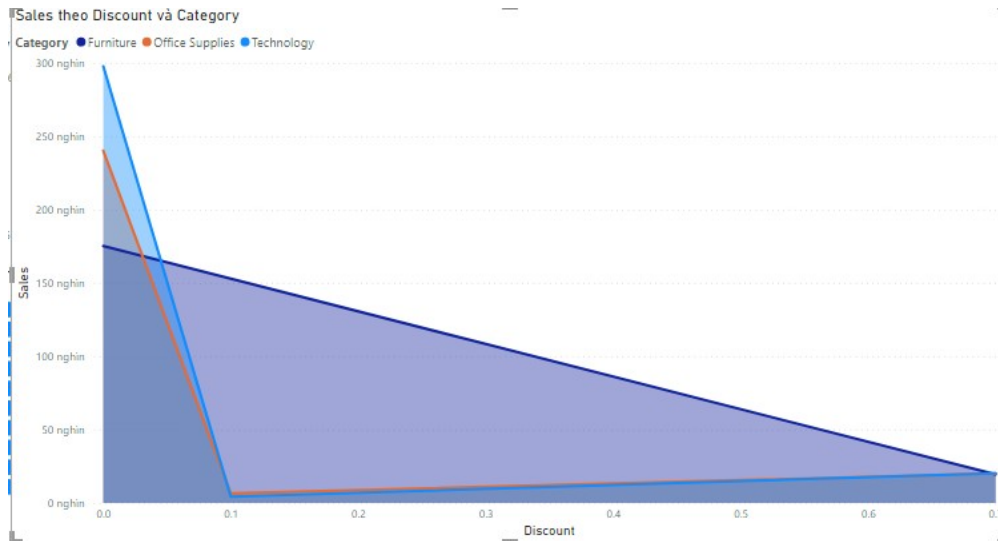


+Cho thấy số lượng hàng đã bán ra của một mặt hàng và so sánh với các mặt hàng khác trong một khu vực (trên hình là khu vực Africa).

+Sử dụng hàm sum để tính tổng số lượng hàng bán ra của từng sản phẩm và sử dụng biểu đồ thanh bó cụm để thống kê.

-Biểu đồ sales theo discount và category:

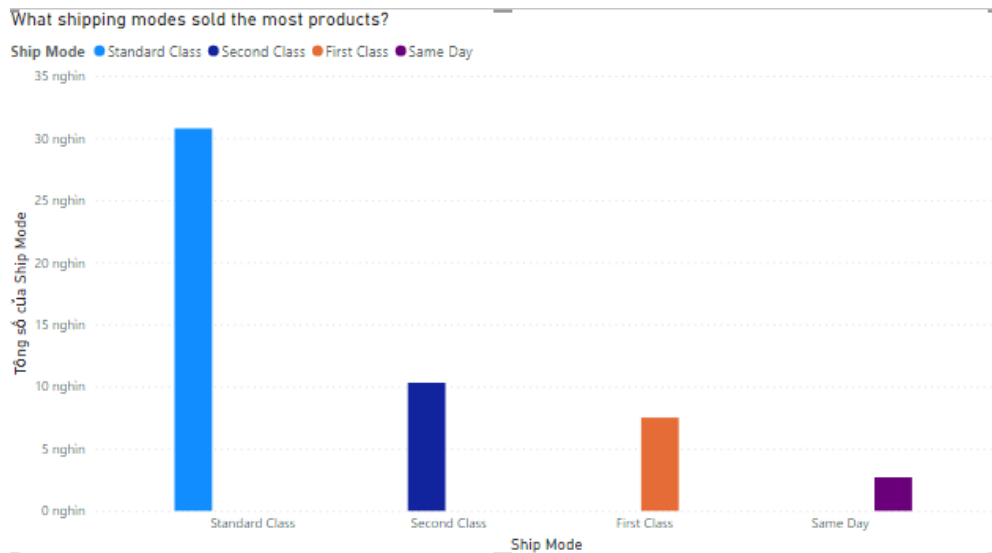




+Cho thấy tỉ lệ mua hàng khi được giảm giá và mức giảm giá của từng loại sản phẩm tại khu vực (trên hình là của khu vực Africa).

+Sử dụng hàm sum để tính giá trị hàng và số mặt hàng được giảm giá theo từng loại sản phẩm sau đó sử dụng biểu đồ vùng để thống kê.

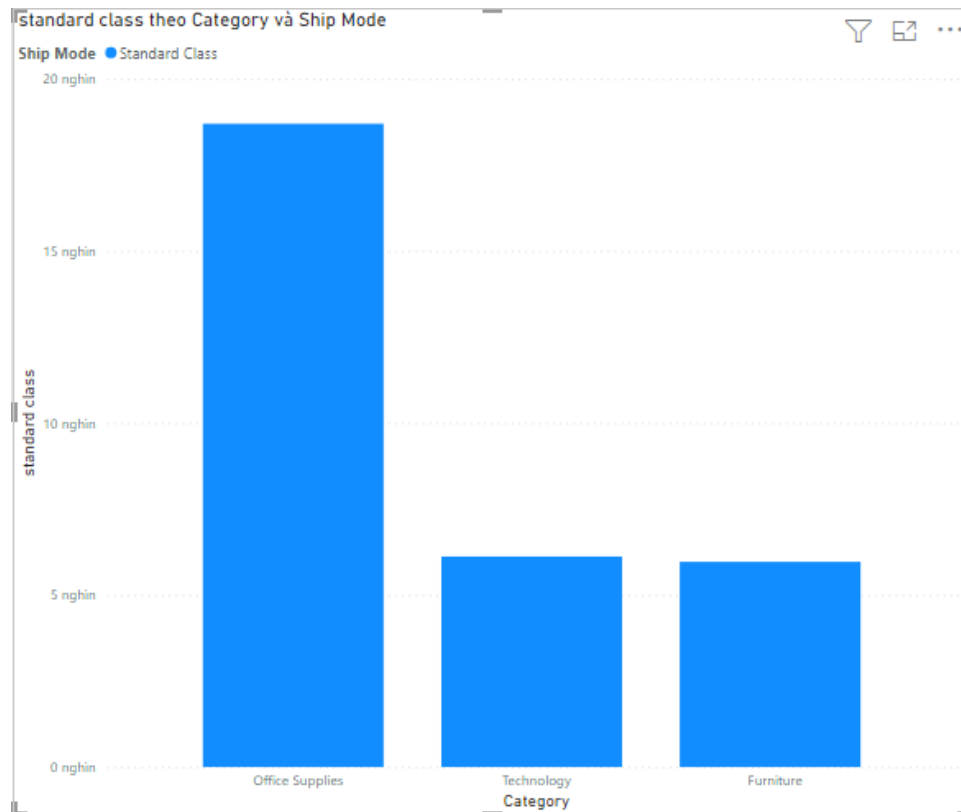
- **Biểu đồ what shipping modes sold the most product**



-Chỉ ra loại giao hàng và khách hàng hay sử dụng nhất.

-Sử dụng hàm sum để tính tổng đơn hàng theo cách ship hàng sau đó sử dụng biểu đồ cột bó cụm để biểu diễn.

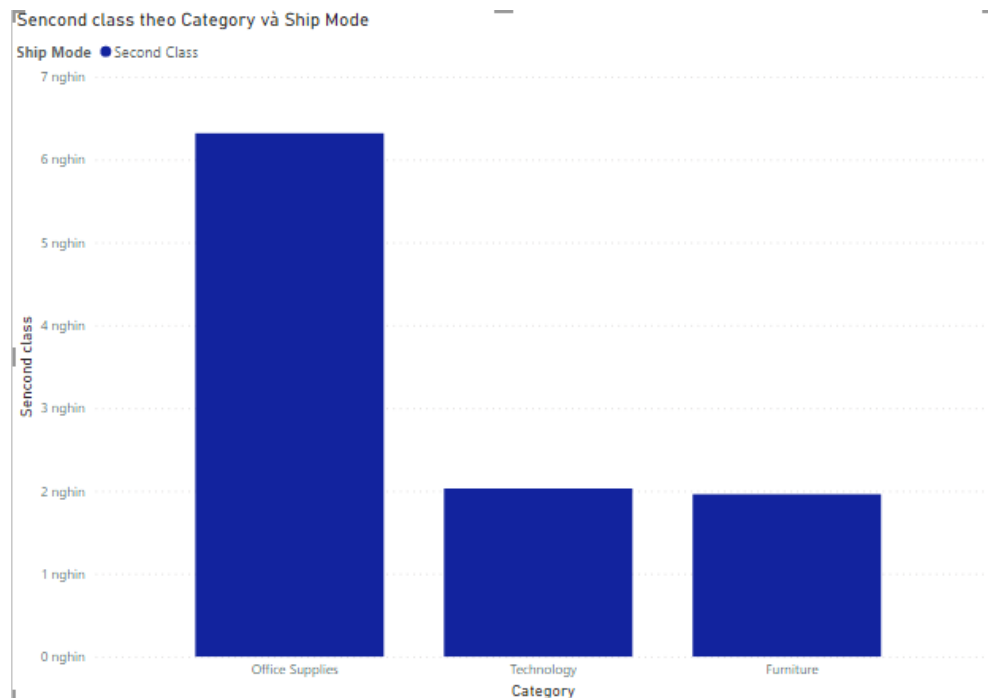
- **Biểu đồ standard class theo category và ship mode.**



-Cho thấy số lượng mặt hàng ship mode thuộc standard class theo từng loại hàng.

-Sử dụng hàm sum tính tổng số standard class theo từng loại hàng và dùng biểu đồ cột bó cụm để biểu diễn

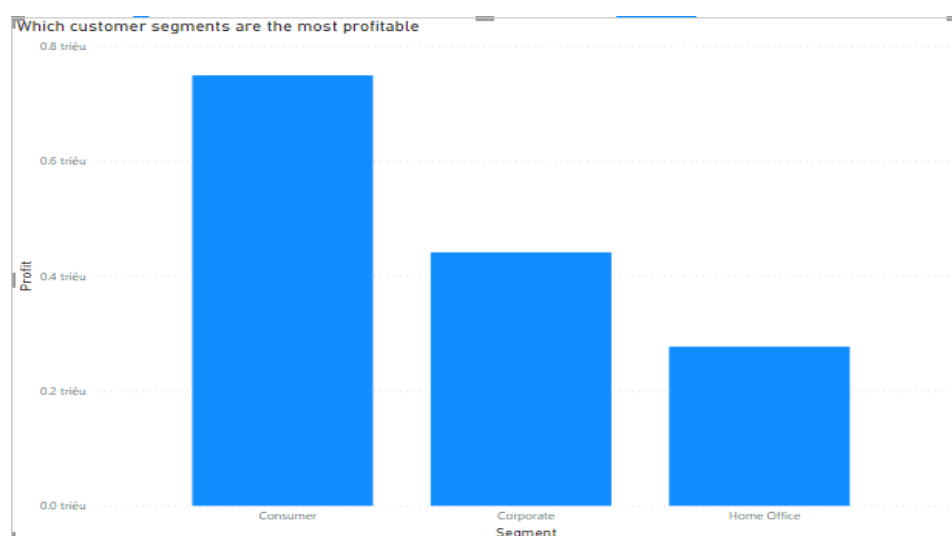
- **Biểu đồ second class theo category và ship mode.**



-Cho thấy số lượng mặt hàng ship mode thuộc second class theo từng loại hàng.

-Sử dụng hàm sum tính tổng số second class theo từng loại hàng và dùng biểu đồ cột bó cụm để biểu diễn

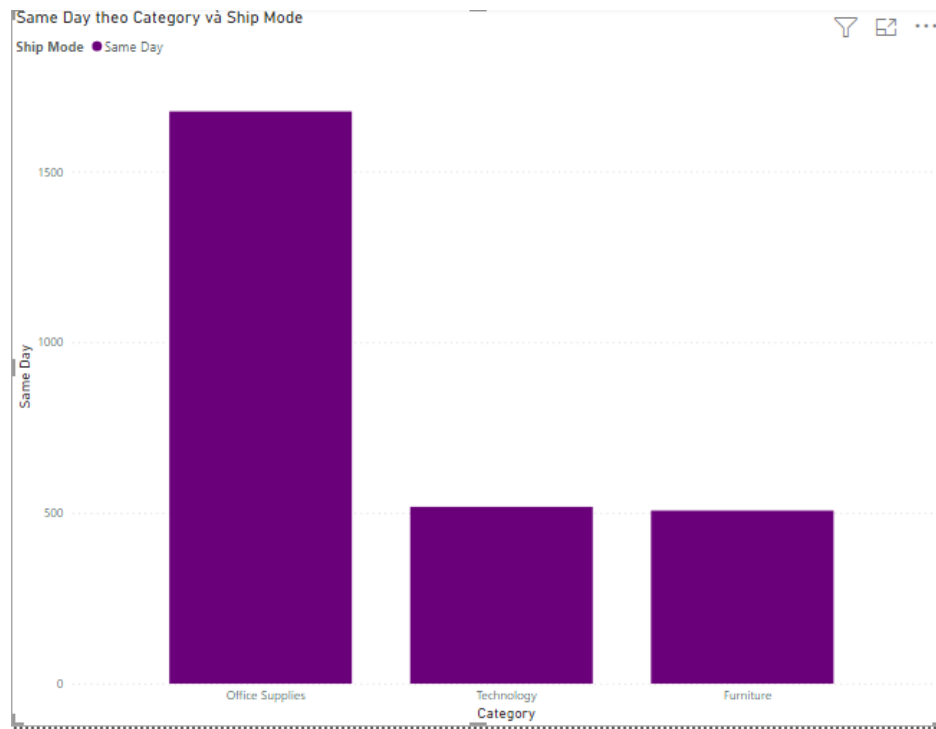
- **Biểu đồ Which customer segments are the most profitable**



-Cho thấy lớp khách hàng mang lại lời nhuận lớn nhất cho cửa hàng.

-Sử dụng hàm sum tính tổng số hác hàng thuộc từng lớp và sử dụng biểu đồ cột bó cụm để biểu diễn.

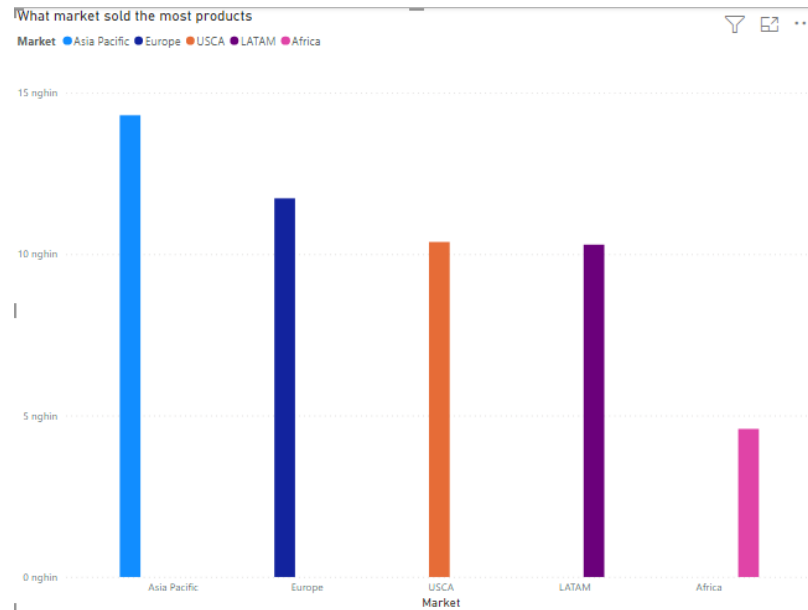
- **Biểu đồ same day thoe category và shipmode**



-Cho thấy số lượng đơn hàng giao trong cùng ngày của từng loại ship.

-Sử dụng hàm sum tính tổng số đơn hàng giao trong cùng ngày thuộc từng loại shipmode và sử dụng biểu đồ cột bó cụm để biểu diễn.

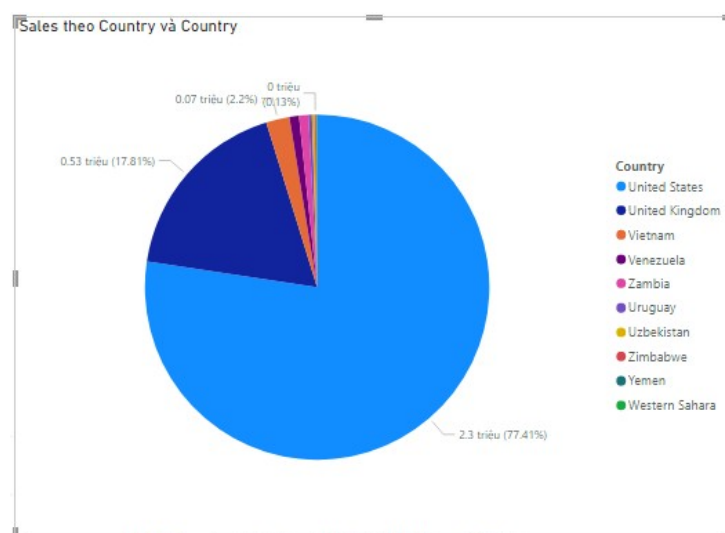
- **Biểu đồ what market sales the most products**



-Cho thấy thị trường tại khu vực nào bán được nhiều sản phẩm nhất.

-Sử dụng hàm sum để tính tổng số lượng sản phẩm đã bán tại mỗi khu vực và sử dụng biểu đồ cột để biểu diễn.

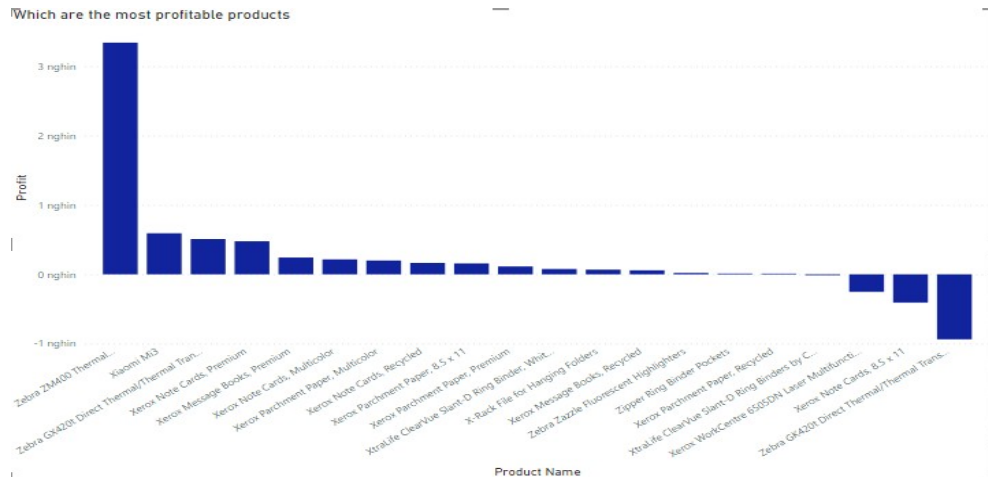
- **Biểu đồ sales theo country**



-Cho thấy số sản phẩm và tỉ lệ sản phẩm tiêu thụ tại từng quốc gia và so sánh với thế giới.

-Sử dụng hàm sum để tính tổng số sản phẩm tiêu thụ tại mỗi quốc gia và sử dụng biểu đồ tròn để biểu diễn.

- **Biểu đồ which are the most profitable products**



-Cho thấy đâu là mặt hàng mang lại lời nhuận nhiều nhất cho của hàng.

-Sử dụng hàm sum tính tổng lợi nhuận của từng mặt hàng sau đó sử dụng biểu đồ cột bó cụm để biểu diễn.

- **Biểu đồ sales và profit theo category tại từng thị trường**

Region	Canada			Caribbean			Central Africa			Central America			Western Europe			Western US			T'ng Sales	Profit
	Sales	Profit		Sales	Profit		Sales	Profit		Sales	Profit		Sales	Profit		Sales	Profit			
Accessories	4,473.21	1,293.31		20,191.52	5,345.94		7,666.60	1,649.25		84,218.66	17,546.20		94,855.76	19,085.57		61,114.12	16,484.60		749,237.02	129,626.31
Appliances	7,949.43	2,233.89		30,064.56	8,078.16		9,878.16	1,678.74		100,323.02	16,325.02		141,358.99	28,937.32		81,551.12	21,917.12		1,010,535.53	141,562.12
Art	4,119.54	913.35		6,083.81	963.53		7,136.49	1,895.04		23,731.29	5,397.55		95,564.14	19,705.25		9,089.91	2,338.17		371,613.15	57,829.86
Binders	3,089.34	786.18		8,465.38	1,145.20		3,384.51	949.49		24,656.92	5,342.52		62,446.92	12,909.09		55,454.57	16,091.01		461,869.39	72,433.15
Bookcases	5,734.02	1,342.53		39,358.36	1,949.18		15,561.42	3,565.26		172,228.15	16,144.17		196,519.30	20,336.14		36,004.12	11,645.51		1,466,572.24	161,924.42
Chairs	3,203.37	856.74		4,571.12	1,816.34		8,919.25	2,696.85		161,573.83	13,492.65		167,387.34	22,396.83		101,781.33	4,027.58		1,501,681.76	140,396.27
Copiers	7,465.53	2,663.64		50,040.33	7,178.99		13,376.43	3,051.45		27,883.26	20,583.66		22,673.36	9,749.29		19,327.24			1,509,436.27	258,567.55
Envelopes	567.18	171.18		6,331.28	993.88		1,346.41	425.70		22,969.94	4,932.72		21,581.36	4,796.91		3,814,008	1,768.03		169,217.49	28,849.49
Filing	40.13	14.13		1,304.33	1,004.33		1,004.33	1,004.33		12,457.33	1,004.33		12,457.33	1,004.33		12,457.33	1,004.33		39,449.43	28,849.43
Furnishings	808.53	113.79		10,165.61	1,205.49		4,731.21	1,063.49		32,801.70	1,087.06		49,635.70	11,320.05		29,844.55	7,566.10		385,195.97	46,845.43
Labels	421.74	129.33		2,197.38	3,060.00		1,060.29	268.35		7,919.30	1,604.50		1,111.20	2,013.67		5,061.39	2,294.97		73,350.28	14,988.92
Machines	4,274.46	607.80		51,640.56	2,603.66		15,184.83	2,920.56		22,638.91	4,120.55		136,996.16	12,406.78		42,444.12	11,619.93		779,060.07	88,867.87
Paper	1,519.53	373.86		7,185.60	1,292.10		2,523.96	645.69		20,431.14	3,873.14		24,104.33	4,898.72		26,121.22	11,883.88		247,783.58	58,111.65
Phones	10,085.61	2,679.87		40,876.70	5,608.32		2,516.13	7,985.04		17,867.52	25,129.24		214,201.57	27,925.57		98,684.35	9,110.74		1,706,824.14	216,717.01
Storage	10,587.30	2,912.43		18,670.07	2,616.15		17,962.86	4,029.24		86,216.23	12,336.35		203,078.27	25,046.23		70,500.81	8,636.67		1,126,812.97	108,416.68
Supplies	1,309.59	277.87		22,280.98	1,378.12		1,958.40	425.13		22,599.35	4,697.23		30,021.66	6,491.51		10,504.72	615.98		22,811.13	2,250.13
Tapes	30.87	30.87		1,042.96	1,042.96		1,042.96	1,042.96		1,042.96	1,042.96		1,042.96	1,042.96					22,811.13	2,250.13
T'ng	66,938.17	17,817.09		324,888.94	34,571.37		143,630.19	35,383.71		1,223,100.36	158,881.93		1,073,627.67	216,433.31		725,457.92	108,418.45		12,642,501.91	1,467,429.97

-Cho thấy tổng tiền bán ra và lợi nhuận của từng mặt hàng tại từng khu vực thì trường.

-Sử dụng hàm sum để tính sales và profit sau đó sử dụng biểu đồ ma trận để biểu diễn dữ liệu.

## KẾT LUẬN

Sau thời gian nghiên cứu đề tài, nhóm đã nhận được sự giúp đỡ nhiệt tình của giáo viên hướng dẫn, giáo viên bộ môn và các bạn trong lớp. Nhờ thế nhóm đã có được một số thành quả nhất định đối với phân tích dữ liệu lớn và phần mềm PowerBI :

- Đã bước đầu nắm được các phân tích dữ liệu.
- Đã nắm giữ những kiến thức cơ bản về PowerBi cũng như các sử dụng.
- Phân tích và chỉ ra những bước phân tích và kết quả phân tích cơ bản.
- Bên cạnh đó cũng có những nhược điểm và thiếu sót trong quá trình thực hiện bài phân tích:
- Chưa tận dụng được hết những dữ liệu đã có.
- Để sót một vài thông tin quan trọng trong quá trình phân tích.
- Do thời gian tiếp xúc không quá dài nên còn thiếu sót trong những kiến thức cơ bản cần phải bổ sung.

Trong thời gian tới em sẽ tìm hiểu, bổ sung kỹ năng chuyên môn để sửa lỗi mà nhóm còn mắc phải. Do thời gian học tập, kỹ năng chuyên môn của nhóm em còn hạn chế nên gặp lỗi là điều không thể tránh khỏi, còn nhiều vấn đề chưa được giải quyết hoàn chỉnh. Vì vậy nhóm em rất mong nhận được những ý kiến đóng góp của quý thầy cô và các bạn để có thể hoàn thiện và phát triển đề tài hơn.

## TÀI LIỆU THAM KHẢO

- [1] Big Data - Tìm hiểu về dữ liệu lớn trong thời kỳ 4.0  
<https://bizflycloud.vn/tin-tuc/big-data-la-gi-20210709142033297.htm>
- [2] Tìm hiểu về văn hóa dữ liệu trong power BI  
<https://powerbi.soft365.vn/tim-hieu-ve-van-hoa-du-lieu-trong-power-bi/>
- [3] Các hàm dax trong power BI :  
<https://elearning.drb.com.vn/product/cac-ham-dax-trong-power-bi/>
- [4] Website Gitiho: <https://gitiho.com/blog/cung-tim-hieu-ve-giao-dien-va-quy-trinh-lam-viec-cua-powerbi.html>
- [5] Website KPIM: <https://kpim.vn/dax/>
- [6] Website slide share: <https://www.slideshare.net/tantrieuf31/tng-quan-v-d-liu-ln-bigdata>
- [7] Kênh youtube Easy : <https://www.youtube.com/watch?v=rPRtUBC-ciQ>
- [8] Kênh youtube Gà Exel: <https://www.youtube.com/watch?v=F7JRKUlim-0>



## NHẬN XÉT CỦA GIÁO VIÊN HƯỚNG DẪN

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

*Thái Nguyên, ngày .... tháng .... năm .....*

**GIÁO VIÊN HƯỚNG DẪN**