

ĐẠI HỌC BÁCH KHOA HÀ NỘI



TIỂU LUẬN HỌC PHẦN PHÂN TÍCH DỮ LIỆU VÀ TRI THỨC KINH DOANH

NHÓM 7

Đề bài: Retail Sales

Giáo viên hướng dẫn: ThS. Nguyễn Danh Tú

Sinh viên thực hiện: Nguyễn Hữu Hải

MSSV: 2021344

Lớp: Kế toán 01- K66

Mã HP: MI3600

Mã lớp: 150761

Năm học 2023 - 2024

MỤC LỤC

I. TỔNG QUAN VỀ DATA WAREHOUSE.....	6
1.1 Data Warehoue là gì ?	6
1.2 Vai trò của Data Warehouse	8
1.3 Cấu trúc của Data Warehouse	9
1.3.1 Kiến trúc tổng thể Data Warehouse	11
1.4 Các loại lược đồ dữ liệu	13
1.4.1 Lược đồ hình sao (Star Schema)	13
1.4.2 Lược đồ bông tuyết (Snow Flake Schema)	14
1.4.3 Mô hình thiên hà (Galaxy Schema)	14
1.4.4 Fact Constellation Schema	15
1.5 Thiết kế Data Warehouse	15
II. TỔNG QUAN VỀ BUSINESS INTELLIGENCE (BI)	16
2.1 Khái niệm về BI	16
2.2 Hoạt động chính BI	17
2.3 Các thành phần trong hệ thống BI	18
2.4 Tổng quan về BI	20
2.4.1 Ứng dụng	20
2.4.2 Xu hướng xây dựng BI	22
2.4.3 Lợi ích BI	23
III. ỨNG DỤNG DATAWAREHOUSE VÀ BI VÀO PHÂN TÍCH DỮ LIỆU LĨNH VỰC BÁN LẺ.....	24
3.1 Tổng quan về bài toán	24
3.1.1 Tổng quan về bán lẻ.....	24
3.1.2 Giới thiệu về bài toán	25
3.2 Kiến trúc Data Warehouse	29
3.3 Tiềm xử lý dữ liệu.....	30
3.4 Mô hình dữ liệu OLTP	32

3.4.1	<i>Mô tả cấu trúc trong cơ sở dữ liệu</i>	32
3.5	Mô hình dữ liệu OLAP	34
3.5.1	<i>Mô hình dữ liệu OLAP-ER</i>	34
3.6	Các mẫu DashBoard	35
3.6.1	<i>DashBoard - Doanh thu</i>	35
3.6.2	<i>DashBoard – Lợi nhuận</i>	39
3.6.3	<i>Forecats</i>	44
IV.	KẾT LUẬN	45

DANH MỤC HÌNH ẢNH

Hình 1. 1	Mô hình dữ liệu nhiều chiều.....	9
Hình 1. 2	Đồ thị lược đồ hình sao	13
Hình 1. 3	Lược đồ bông tuyết (Snow Flake Schema)	14
Hình 1. 4	Mô hình thiên hà (Galaxy Schema).....	15
Hình 1. 5	Constellation Schema	15
Hình 3. 1	Retail Sales Model Canvas	25
Hình 3. 2	Data flow	26
Hình 3. 3	Quy trình mua hàng chung	26
Hình 3. 4	Sơ đồ quy trình nghiệp vụ	27
Hình 3. 5	Khám phá dữ liệu	28
Hình 3. 6	Biểu đồ khuynh hướng doanh thu	28
Hình 3. 7	Biểu đồ Scatter Chart giữa doanh thu và lợi nhuận.....	29
Hình 3. 8	Kiến trúc Data Warehouse	29
Hình 3. 9	Định dạng lại ngày.....	31
Hình 3. 10	Sắp xếp lại thứ tự thời gian.....	31
Hình 3. 11	Xóa dữ liệu null	31
Hình 3. 12	Xóa dữ liệu trùng lặp Customer ID	32
Hình 3. 13	Transform dữ liệu	32
Hình 3. 14	Cấu trúc Detail Oder.....	33
Hình 3. 15	Cấu trúc Oder.....	33
Hình 3. 16	Cấu trúc Customer	33
Hình 3. 17	Cấu trúc Product	34

Hình 3. 18 Cấu trúc Category	34
Hình 3. 19 Mô hình dữ liệu OLAP- ER	34
Hình 3. 20 Mô hình dữ liệu OLAP	35
Hình 3. 21 DashBoard Sales 2014.....	37
Hình 3. 22 DashBoard Sales 2015.....	37
Hình 3. 23 DashBoard Sales 2016.....	38
Hình 3. 24 DashBoard Sales 2017.....	39
Hình 3. 25 DashBoard Profit 2014	41
Hình 3. 26 DashBoard Profit 2015	41
Hình 3. 27 DashBoard Profit 2016	42
Hình 3. 28 DashBoard Profit 2017	43
Hình 3. 29 Product Category Detail	43
Hình 3. 30 Detail Profit Sub-Category	44
Hình 3. 31 Forecats.....	45

LỜI MỞ ĐẦU

Trong bối cảnh hội nhập kinh tế hiện nay, sự cạnh tranh khốc liệt đòi hỏi các doanh nghiệp phải đối mặt với nhiều khó khăn để tồn tại và phát triển. Để vượt qua những thách thức này, doanh nghiệp cần không ngừng hoàn thiện và nâng cao hoạt động kinh doanh, đồng thời phải xây dựng các chiến lược phù hợp cho từng giai đoạn phát triển. Do đó, việc phân tích hoạt động kinh doanh thường xuyên sẽ giúp các nhà quản trị đánh giá đầy đủ và chính xác mọi diễn biến và kết quả sản xuất kinh doanh. Qua đó, doanh nghiệp có thể nhận diện điểm mạnh và điểm yếu để củng cố, phát huy hoặc cải tiến quản lý, tận dụng tối đa nguồn lực nhằm đạt hiệu quả cao nhất trong kinh doanh. Kết quả phân tích là cơ sở để đưa ra các quyết định quản trị ngắn hạn và dài hạn, giúp dự báo, đề phòng và hạn chế rủi ro trong kinh doanh.

Hiện nay, với sự phát triển mạnh mẽ của nền kinh tế thế giới và nền kinh tế Việt Nam, các doanh nghiệp cần nỗ lực lớn mới có thể tồn tại và phát triển. Điều này đòi hỏi các nhà quản trị doanh nghiệp phải hiểu rõ thực lực của doanh nghiệp mình để đề ra các phương hướng phát triển phù hợp. Để đạt được điều này, việc phân tích kết quả hoạt động kinh doanh dựa trên dữ liệu doanh nghiệp cần được thực hiện nghiêm túc. Phân tích kết quả hoạt động kinh doanh có vai trò quan trọng đối với mọi doanh nghiệp, quyết định sự sống còn của họ. Để thu hẹp khoảng cách giữa dự tính kế hoạch và thực tế, việc phân tích cần được thực hiện cẩn trọng nhằm đảm bảo đánh giá đúng đắn và chính xác. Từ đó, nhà quản trị có thể định hướng tương lai doanh nghiệp phù hợp với sự phát triển chung của nền kinh tế, tăng khả năng cạnh tranh và giúp doanh nghiệp phát triển bền vững.

Cùng với sự phát triển nhanh chóng của công nghệ, việc phân tích dữ liệu và xây dựng các Dashboard ngày càng trở nên dễ dàng. Vì vậy, nhóm chúng em đã lựa chọn đề tài "Phân tích dữ liệu với bài toán bán lẻ" thông qua hệ cơ sở dữ liệu và quá trình kinh doanh thông minh (BI).

I. TỔNG QUAN VỀ DATA WAREHOUSE

1.1 Data Warehouse là gì ?

Kho dữ liệu (Data Warehouse - DW) là một hệ thống lưu trữ dữ liệu có tính chất chủ đề, tích hợp, theo dõi theo thời gian và có tính bền vững, hỗ trợ quá trình ra quyết định các hoạt động kinh doanh thông minh. Data Warehouse có thể hiểu đơn giản là một tập hợp các dữ liệu và thông tin liên quan đến một chủ đề cụ thể, được tổng hợp từ nhiều nguồn khác nhau và qua nhiều thời điểm khác nhau mà không bị thay đổi. Hệ thống này được sử dụng để hỗ trợ ra quyết định, phân tích dữ liệu và lập báo cáo quản lý.

Kho dữ liệu tập trung và hợp nhất lượng lớn dữ liệu từ nhiều nguồn. Khả năng phân tích của nó cho phép các tổ chức rút ra những hiểu biết kinh doanh có giá trị từ dữ liệu của họ để cải thiện việc ra quyết định. Theo thời gian, nó xây dựng nên một bản ghi lịch sử có thể là vô giá đối với các nhà khoa học dữ liệu và nhà phân tích kinh doanh. Nhờ những khả năng này, kho dữ liệu có thể được coi là “nguồn sự thật duy nhất” của tổ chức.

Một kho dữ liệu điển hình thường bao gồm các yếu tố sau:

- Cơ sở dữ liệu quan hệ để lưu trữ và quản lý dữ liệu
- Giải pháp trích xuất, tải và chuyển đổi (ELT) để chuẩn bị dữ liệu cho phân tích
- Khả năng phân tích thống kê, báo cáo và khai thác dữ liệu
- Công cụ phân tích khách hàng để trực quan hóa và trình bày dữ liệu cho người dùng doanh nghiệp
- Các ứng dụng phân tích khác, phức tạp hơn tạo ra thông tin hữu ích bằng cách áp dụng thuật toán khoa học dữ liệu và trí tuệ nhân tạo (AI) hoặc các tính năng biểu đồ và không gian cho phép nhiều loại phân tích dữ liệu hơn trên quy mô lớn

Bên cạnh đó, Data Warehouse cũng có những đặc tính nhất định của bản thân hệ cơ sở, bao gồm 04 đặc tính: định hướng theo chủ đề, tích hợp, không biến đổi và biến đổi theo thời gian. Bốn đặc điểm độc đáo (được mô tả bởi nhà khoa học máy tính William Inmon, người được coi là cha đẻ của kho dữ liệu) từng đề cập rằng: "Một kho dữ liệu là

một tập hợp dữ liệu hướng chủ đề, tích hợp, biến thể thời gian, ít biến động hỗ trợ cho quá trình đưa ra quyết định của doanh nghiệp"

- **Định hướng theo chủ đề:** Thông tin trong Data Warehouse chủ yếu tập trung vào việc mô hình hóa và phân tích dữ liệu để hỗ trợ ra quyết định, thay vì xử lý các giao dịch hoặc tác nghiệp hàng ngày. Các chủ đề thường liên quan đến khách hàng, sản phẩm và doanh thu. Mục tiêu của kho dữ liệu là phục vụ các yêu cầu phân tích hoặc khai thác thông tin cụ thể, được gọi là chủ đề, bằng cách cung cấp một góc nhìn đơn giản và loại bỏ dữ liệu không cần thiết cho việc ra quyết định.
- **Tích hợp:** Kho dữ liệu tạo ra sự nhất quán giữa các loại dữ liệu khác nhau từ các nguồn khác nhau. Áp dụng các kỹ thuật và quy trình ETL (Extract, Transform, Load) giúp làm sạch và tổng hợp dữ liệu từ nhiều nguồn vào một kho dữ liệu tập trung. Quá trình này cho phép người dùng truy cập và xem xét đồng thời nhiều nhóm chỉ tiêu khác nhau, đảm bảo tính nhất quán và chính xác giữa các nguồn dữ liệu.
- **Không biến đổi:** Khi dữ liệu ở trong kho dữ liệu, nó sẽ ổn định và không thay đổi. Kho dữ liệu luôn được tách biệt về mặt vật lý khỏi dữ liệu được sử dụng trong xử lý giao dịch hàng ngày. Vì sự tách biệt này, kho dữ liệu không cần xử lý các giao dịch, khôi phục hay kiểm soát xử lý đồng thời. Thường thì, kho dữ liệu chỉ yêu cầu thực hiện hai hoạt động chính: tải dữ liệu và làm mới dữ liệu.
- **Dữ liệu theo thời gian:** Data Warehouse lưu trữ dữ liệu lịch sử, cho phép phân tích các xu hướng và thay đổi qua các thời kỳ. Dữ liệu được ghi lại với các dấu thời gian cụ thể, giúp theo dõi sự biến đổi và phân tích theo thời gian, điều này rất quan trọng cho việc ra quyết định dựa trên xu hướng dài hạn và sự thay đổi theo thời gian.

Kho dữ liệu được thiết kế tốt sẽ thực hiện các truy vấn rất nhanh, cung cấp thông lượng dữ liệu cao và cung cấp đủ tính linh hoạt cho người dùng cuối để “cắt và xức xắc” hoặc giảm khối lượng dữ liệu để kiểm tra kỹ hơn nhằm đáp ứng nhiều nhu cầu khác nhau—dù ở mức cao mức độ hoặc ở mức độ rất tốt, chi tiết. Kho dữ liệu đóng vai trò là

nền tảng chức năng cho môi trường BI phần mềm trung gian cung cấp cho người dùng cuối các báo cáo, bảng thông tin và các giao diện khác.

1.2 Vai trò của Data Warehouse

Sự xuất hiện của kho dữ liệu là một phản ứng tự nhiên đối với việc tăng cường nhu cầu xử lý lượng dữ liệu ngày càng lớn. Trong bối cảnh mà nhu cầu lưu trữ dữ liệu liên tục gia tăng, cùng với sự phức tạp ngày càng cao của hệ thống máy tính, kho dữ liệu mang lại nhiều lợi ích quan trọng cho các doanh nghiệp.

Trong số các lợi ích đó, có thể kể đến:

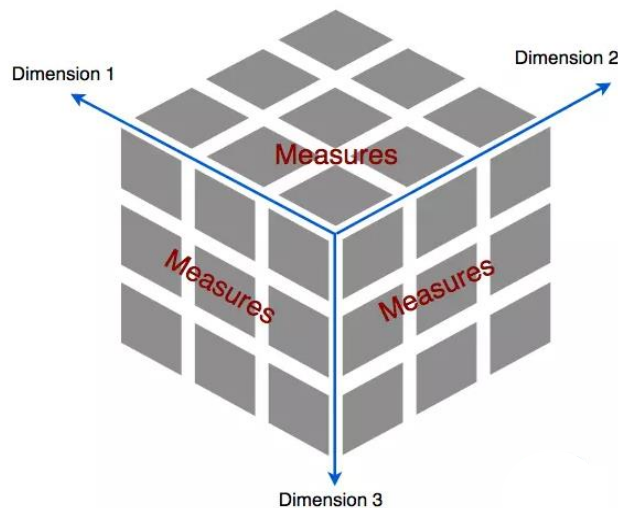
- **Tích hợp dữ liệu:** Kho dữ liệu giúp tổ chức và tích hợp dữ liệu từ nhiều nguồn khác nhau vào một nguồn duy nhất, và thường được chuẩn hóa về cùng một định dạng. Điều này giải quyết vấn đề phân mảnh và mất cân bằng dữ liệu, đồng thời giúp đáp ứng nhu cầu thông tin của tất cả các người dùng trong tổ chức.
- **Tiết kiệm thời gian và tăng hiệu suất:** Kho dữ liệu giúp tiết kiệm thời gian trong việc tìm kiếm dữ liệu cần thiết bằng cách tổ chức và lưu trữ dữ liệu một cách có tổ chức giúp người dùng nhanh chóng truy xuất thông tin mà họ cần mà không cần phải tìm kiếm qua nhiều nguồn dữ liệu.
- **Phân tích dữ liệu hiệu quả hơn:** Data Warehouse không chỉ là nơi lưu trữ dữ liệu mà còn là công cụ quan trọng cho việc phân tích dữ liệu. Thông qua việc xử lý và phân tích dữ liệu, Data Warehouse giúp tổ chức và lọc ra những thông tin quan trọng và hữu ích từ dữ liệu nguyên thô, từ đó hỗ trợ quyết định kinh doanh và chiến lược phát triển.
- **Hỗ trợ ra quyết định:** Kho dữ liệu cung cấp các báo cáo và dữ liệu phân tích phong phú, giúp người dùng đưa ra các quyết định kinh doanh hợp lý, nhanh chóng và hiệu quả. Việc dựa vào thông tin chính xác và chi tiết từ kho dữ liệu giúp tối ưu hóa quy trình ra quyết định và đưa ra các quyết định có tính chiến lược.
- **Quản lý dự án và hoạt động:** Kho dữ liệu cung cấp một cơ sở dữ liệu đồng nhất và tin cậy cho các dự án và hoạt động của doanh nghiệp. Việc tổ chức, xác định và quản

lý dữ liệu một cách hiệu quả giúp tăng cường khả năng quản lý và thực hiện các dự án, hoạt động một cách chính xác và hiệu quả.

- **Xử lý lượng dữ liệu lớn:** Với sự tăng trưởng không ngừng của dữ liệu, kho dữ liệu giúp doanh nghiệp quản lý và xử lý lượng dữ liệu lớn một cách hiệu quả hơn, đáp ứng nhu cầu ngày càng tăng về lưu trữ và xử lý dữ liệu. Điều này là quan trọng đặc biệt trong bối cảnh thị trường ngày càng cạnh tranh và yêu cầu về thông tin kinh doanh ngày càng cao.

1.3 Cấu trúc của Data Warehouse

Trong kho dữ liệu, lượng dữ liệu thường rất lớn và không có các hoạt động sửa đổi hoặc tạo mới, mà thường được tối ưu cho việc phân tích và báo cáo. Các thao tác với dữ liệu trong kho dữ liệu thường dựa trên mô hình dữ liệu đa chiều (multidimensional data model), trong đó dữ liệu được tổ chức thành các đối tượng được gọi là data cube (khối dữ liệu). Data cube là trung tâm của việc phân tích, bao gồm các dữ liệu thực tế (facts) và các thuộc tính tạo thành các chiều dữ liệu khác nhau (dimensions).



Hình 1. 1 Mô hình dữ liệu nhiều chiều

Trong kho dữ liệu Data Warehouse, chúng ta thường gặp ba loại dữ liệu: Additive (Cộng hưởng): Đây là loại dữ liệu có thể được tổng hợp thông qua tất cả các chiều dữ liệu (Dimension) trong bảng Fact. Ví dụ, doanh số bán hàng có thể được tổng hợp theo khách hàng, sản phẩm và thời gian. Semi-Additive (Bán cộng hưởng): Đây là loại dữ liệu có thể được tổng hợp cho một số chiều dữ liệu trong bảng Fact, nhưng không phải cho tất cả các chiều. Ví dụ, số lượng hàng tồn kho có thể được tổng hợp theo sản phẩm và thời gian, nhưng không thể tổng hợp theo vị trí lưu trữ. Non-Additive (Không cộng hưởng): Đây là loại dữ liệu không thể tổng hợp cho bất kỳ chiều dữ liệu nào trong bảng Fact. Ví dụ, giá trị trung bình của một sản phẩm không thể được tổng hợp theo thời gian.

Hệ thống kho dữ liệu thường được tổ chức thành ba tầng:

Tầng đáy: Tầng này cung cấp các dịch vụ để lấy dữ liệu từ nhiều nguồn khác nhau, sau đó chuẩn hoá, làm sạch, tối ưu hóa và lưu trữ dữ liệu đã được tập trung. Tầng giữa: Tầng này cung cấp các dịch vụ để thực hiện các thao tác với dữ liệu, bao gồm dịch vụ OLAP (Online Analytical Processing), để phục vụ cho việc phân tích dữ liệu. Tầng trên: Tầng này là nơi chứa các câu truy vấn, báo cáo và công cụ phân tích dữ liệu, giúp người dùng trực tiếp tương tác với dữ liệu và lấy thông tin cần thiết.

- Trong quá trình phân tích dữ liệu, chúng ta thường gặp các hoạt động OLAP chính:
- Thu nhỏ (Roll-up): Tổng hợp dữ liệu thành phạm vi lớn hơn. Ví dụ: tổng hợp theo năm thay vì theo quý.
- Mở rộng (Drill-down): Phân chia dữ liệu thành các phạm vi nhỏ hơn. Ví dụ: phân chia theo tháng thay vì theo quý.
- Cắt lát (Slice): Xem dữ liệu theo từng lớp một. Ví dụ: chỉ xem doanh số bán hàng của một quý cụ thể.
- Thu nhỏ (Dice): Giảm bớt một phần của dữ liệu.
- Pivot (Rotate): Trực quan hóa dữ liệu, chẳng hạn chuyển từ một biểu đồ 3D sang một biểu đồ mặt phẳng 2D.

1.3.1 Kiến trúc tổng thể Data Warehouse

Data Source Layer (Lớp dữ liệu nguồn)

Mô tả: Đây là lớp đầu vào của hệ thống DWH (Data Warehouse), nơi thu thập dữ liệu từ nhiều nguồn khác nhau. Các nguồn dữ liệu có thể bao gồm:

- Cơ sở dữ liệu quan hệ (relational databases) như MySQL, PostgreSQL, Oracle, SQL Server.
- File text (như CSV, TSV).
- File XML (Extensible Markup Language).
- Dữ liệu từ các dịch vụ web (web services) như API REST, SOAP.
- Lịch sử truy cập và hoạt động của người dùng (user activity logs).

Vai trò: Cung cấp dữ liệu thô và đa dạng cho hệ thống DWH, đảm bảo rằng tất cả các nguồn thông tin cần thiết đều được đưa vào quy trình xử lý và phân tích.

ETL Layer

Mô tả: ETL (Extract, Transform, Load) là quy trình quan trọng trong việc xây dựng và duy trì DWH.

- Extract (Trích xuất): Lấy dữ liệu từ các nguồn dữ liệu khác nhau.
- Transform (Biến đổi): Chuyển đổi dữ liệu từ định dạng nguồn sang định dạng đích, thực hiện các thao tác làm sạch, chuẩn hóa, và tích hợp dữ liệu.
- Load (Tải): Đưa dữ liệu đã biến đổi vào DWH.

Vai trò: Đảm bảo dữ liệu từ các nguồn khác nhau được xử lý một cách nhất quán và hợp lệ trước khi lưu trữ vào DWH.

Staging Area

Mô tả: Là nơi tạm thời lưu trữ dữ liệu trong quá trình chuyển từ nguồn vào DWH. Tại đây, các trường dữ liệu thừa hoặc không cần thiết được loại bỏ và dữ liệu được chuẩn hóa theo chuẩn của DWH.

Vai trò: Tạo một không gian trung gian giúp xử lý và làm sạch dữ liệu trước khi lưu trữ lâu dài trong DWH, đảm bảo chất lượng dữ liệu.

Data Storage Layer

Mô tả: Lớp lưu trữ dữ liệu cuối cùng sau khi đã được làm sạch và chuẩn hóa. Dữ liệu ở đây đã sẵn sàng cho các mục đích truy vấn và phân tích.

Vai trò: Lưu trữ dữ liệu đã được chuẩn hóa, đảm bảo rằng dữ liệu có chất lượng cao và có thể dễ dàng truy xuất cho các ứng dụng và người dùng.

Data Logic Layer

Mô tả: Lớp lưu trữ các quy luật và logic của doanh nghiệp. Các quy luật này không thay đổi dữ liệu trong DWH nhưng ảnh hưởng đến cách dữ liệu được sử dụng và hiển thị trong các báo cáo và phân tích.

Vai trò: Đảm bảo rằng các quy luật nghiệp vụ được áp dụng một cách nhất quán và đúng đắn trong các báo cáo và phân tích dữ liệu.

Data Presentation Layer

Mô tả: Lớp này chịu trách nhiệm hiển thị thông tin và dữ liệu cho người dùng cuối. Các thông tin này có thể được trình bày dưới dạng bảng, báo cáo đồ họa, biểu đồ, hoặc email báo cáo tự động.

Vai trò: Cung cấp thông tin hữu ích cho người dùng cuối theo yêu cầu của doanh nghiệp, đảm bảo dữ liệu được trình bày một cách trực quan và dễ hiểu.

Metadata Layer

Mô tả: Lưu trữ thông tin về dữ liệu trong DWH, bao gồm thông tin về nguồn gốc dữ liệu, các quy tắc biến đổi, cấu trúc dữ liệu, và các chỉ số hiệu suất.

Vai trò: Cung cấp thông tin chi tiết về dữ liệu trong DWH, hỗ trợ quá trình quản lý, truy xuất và phân tích dữ liệu.

System Operations Layer

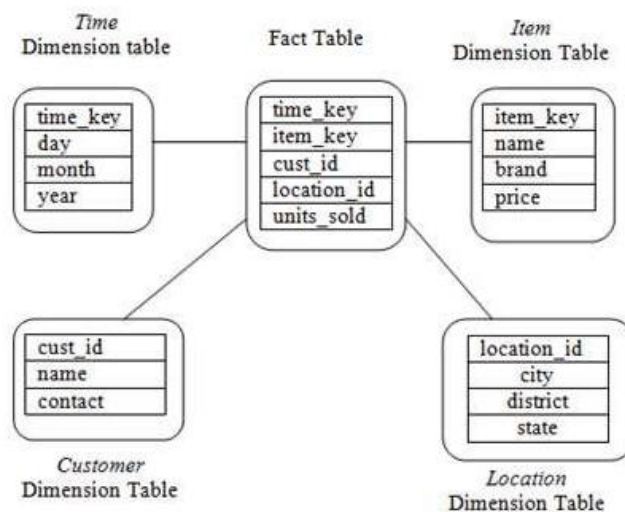
Mô tả: Lưu trữ thông tin về hoạt động của hệ thống DWH như tình trạng của tiến trình ETL, hiệu suất hệ thống, và lịch sử truy cập của người dùng.

Vai trò: Giám sát và quản lý hoạt động của hệ thống DWH, đảm bảo hệ thống hoạt động hiệu quả và ghi lại các sự kiện quan trọng để phục vụ cho việc quản lý và bảo trì.

1.4 Các loại lược đồ dữ liệu

1.4.1 Lược đồ hình sao (Star Schema)

Gồm 1 bảng Fact (bảng sự kiện) nằm ở trung tâm và được bao quanh bởi những bảng Dimension (bảng chiều). Dữ liệu của lược đồ hình sao không được chuẩn hoá. Các câu hỏi nhằm vào bảng Fact và được cấu trúc bởi các bảng Dimension.



Hình 1. 2 Đồ thị lược đồ hình sao

Đồ thị lược đồ hình sao

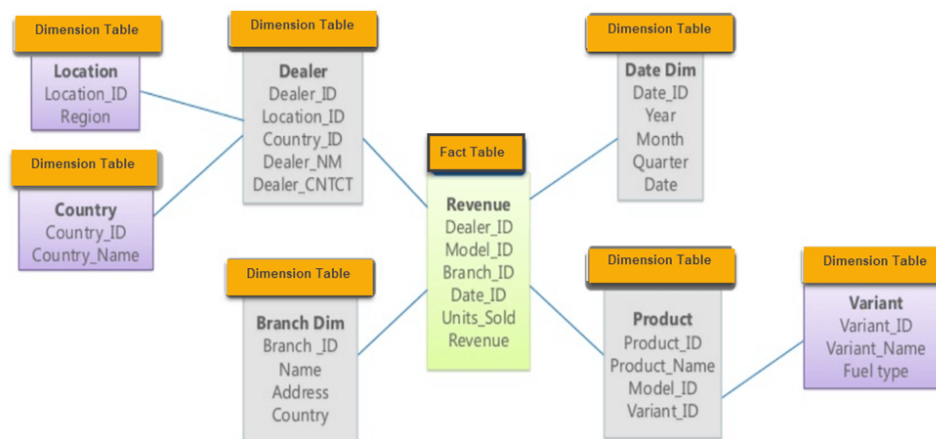
Ưu điểm: Bảng Fact, Dimension được mô tả rõ ràng, dễ hiểu. Bảng Dim là bảng dữ liệu tĩnh, và bảng Fact là dữ liệu động được nạp bằng các thao tác. Khoá của Fact được tạo

bởi khoá của các bảng Dim. Nghĩa là khoá chính của các bảng Dim chính là khoá của bảng Fact.

Nhược: Dữ liệu không được chuẩn hoá.

1.4.2 Lược đồ bông tuyết (Snow Flake Schema)

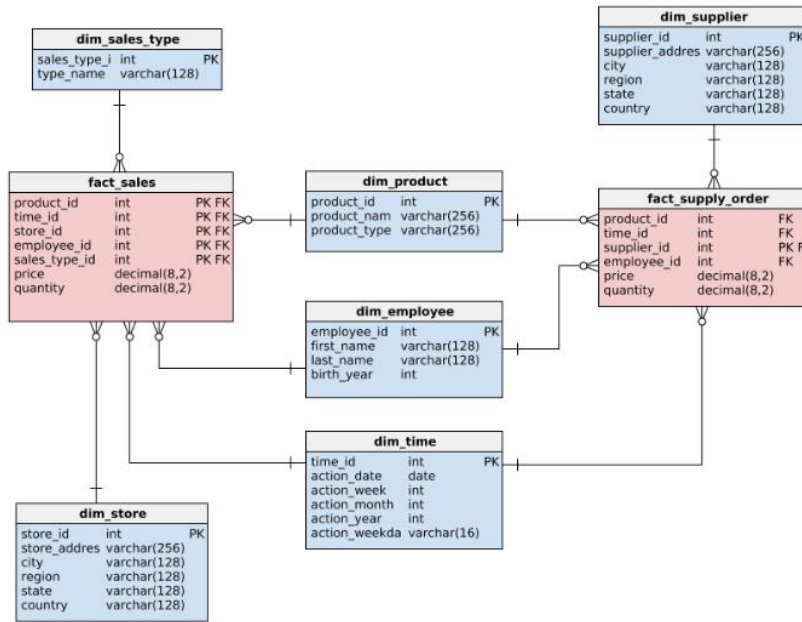
Lược đồ bông tuyết là dạng mở rộng của lược đồ hình sao bằng các bổ sung các Dim. Bảng Fact như lược đồ hình sao, bảng Dim được chuẩn hoá. Các chiều được cấu trúc rõ ràng. Bảng Dim được chia thành chiều chính hay chiều phụ.



Hình 1. 3 Lược đồ bông tuyết (Snow Flake Schema)

1.4.3 Mô hình thiên hà (Galaxy Schema)

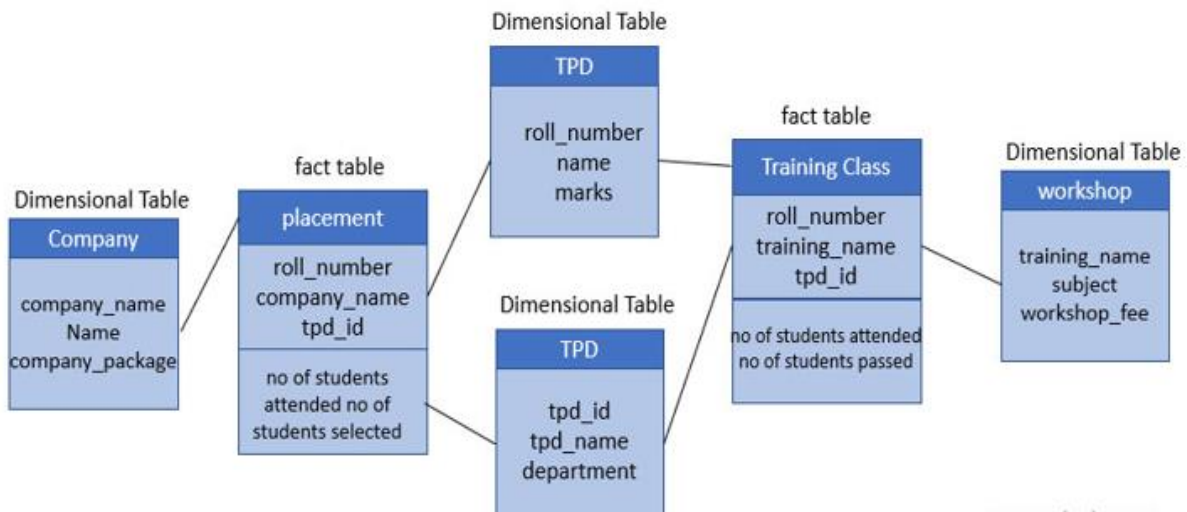
Mỗi cái đó bao gồm chỉ một bảng fact và một vài bảng dimension. Nếu chúng ta muốn kết hợp 2 data mart vào làm một. Đây là kiểu mô hình chứa vài bảng fact và chia sẻ các bảng dimension nó được gọi là mô hình galaxy (galaxy schema). Chia sẻ các bảng dimension có thể giảm kích thước của cơ sở dữ liệu đặc biệt là khi chia sẻ các bảng dimension có nhiều giá trị.



Hình 1. 4 Mô hình thiên hà (Galaxy Schema)

1.4.4 Fact Constellation Schema

Dimension trong lược đồ được tách thành các Dimension độc lập dựa trên các cấp độ của hệ thống phân cấp.



Hình 1. 5 Constellation Schema

1.5 Thiết kế Data Warehouse

Trong quá trình thiết kế Data Warehouse, ta cần để ý tới 03 vấn đề chính bao gồm: Data, Structure và Process; cụ thể, ta cần trả lời một số câu hỏi như sau:

Data: Cần thông tin gì để hỗ trợ ra quyết định? Ở cấp độ nào? Dữ liệu được lấy từ đâu? định dạng như thế nào? Độ lớn dữ liệu? Mức độ tăng trưởng dữ liệu như thế nào? cần bao nhiêu không gian để chứa?

Structure: Dữ liệu cần được xây dựng theo những chiều nào để phục vụ phân tích? Cấu trúc dữ liệu nào là phù hợp với yêu cầu? (Relational OLAP, Multi-Dimensional OLAP, Hybrid OLAP).

Process: Tiến trình Extract (truy cập hệ thống nguồn để trích xuất dữ liệu) -Transform (Kiểm tra, làm sạch, điều chỉnh phù hợp các yêu cầu của DW) -Load (cập nhật DW với dữ liệu được cung cấp từ load Data) được thực hiện như thế nào? Đặt ở đâu? Lập lịch như nào? Có những ngoại lệ nào cần xử lý? ở cấp độ nào?

II. TỔNG QUAN VỀ BUSINESS INTELLIGENCE (BI)

2.1 Khái niệm về BI

Business Intelligence – BI (hay còn được gọi là Kinh doanh thông minh) là một dạng công nghệ giúp doanh nghiệp hiểu biết về quá khứ, qua đó đưa ra quyết định, hành động và dự đoán tương lai.

Trong đó BI bao gồm các kỹ năng, quy trình, công nghệ hay ứng dụng để hỗ trợ ra quyết định. BI là các hoạt động kết hợp giữa phân tích kinh doanh, khai thác dữ liệu, trực quan hóa dữ liệu, công cụ dữ liệu và cơ sở hạ tầng và thực tiễn để giúp tổ chức ra các quyết định-dựa trên data (data-driven decision). BI là công cụ để chuyển đổi những dữ liệu thô thành những thông tin có nghĩa, phục vụ cho mục tiêu phân tích kinh doanh. BI là công nghệ giúp doanh nghiệp hiểu biết về quá khứ và dự đoán tương lai

BI (Business Intelligence) là quy trình và tập hợp các công nghệ mà các doanh nghiệp sử dụng để quản lý khối lượng dữ liệu lớn, khai thác thông tin và tri thức, từ đó giúp doanh nghiệp đưa ra các quyết định kinh doanh hiệu quả hơn. Công nghệ BI mang đến một cái nhìn toàn diện về hoạt động của doanh nghiệp từ quá khứ, hiện tại đến dự đoán tương lai. Mục tiêu của BI là hỗ trợ doanh nghiệp trong việc ra quyết định chính xác và kịp thời. Do đó, hệ thống BI thường được xem như là một hệ thống hỗ trợ quyết định (Decision Support System - DSS).

2.2 Hoạt động chính BI

BI (Business Intelligence) được áp dụng cho mọi loại hình tổ chức và doanh nghiệp nhằm đáp ứng nhu cầu tích hợp dữ liệu và phân tích thông tin. Thông thường, khi ban lãnh đạo doanh nghiệp cần thông tin để thống kê, phân tích và đưa ra quyết định chiến lược, họ sẽ yêu cầu các bộ phận hỗ trợ (thường là phòng tài chính) chuẩn bị báo cáo. Quá trình này thường mất nhiều thời gian để liên hệ và tạo lập báo cáo, và kết quả có thể không hoàn toàn chính xác do công việc chủ yếu được thực hiện thủ công với sự hỗ trợ của các công cụ văn phòng như MS Excel.

BI xuất hiện để giải quyết các vấn đề này, cho phép doanh nghiệp trả lời các câu hỏi quan trọng như:

- Top 5 khách hàng quan trọng nhất của doanh nghiệp hiện nay?
- Những thị trường nào tạo ra lợi nhuận lớn?
- Những mặt hàng nào có tiềm năng?
- Làm thế nào để tiết kiệm thời gian và tối ưu chi phí quản trị doanh nghiệp ?

Với những doanh nghiệp đã triển khai hệ thống quản trị nguồn lực doanh nghiệp (ERP) hoặc các hệ thống thông tin chuyên biệt khác, BI có thể tự động kết nối với các nền tảng này để cung cấp báo cáo phân tích và thống kê tức thời, hỗ trợ cho việc quản lý và ra quyết định nhanh chóng. BI giúp tạo ra các báo cáo động theo nhiều chiều thông tin khác nhau, bao gồm:

- Hệ thống các biểu đồ động (graphical chart);
- Các chỉ số quản lý hiệu suất (KPIs);
- Hệ thống phân tích và dự báo báo cáo (Simulation and Forecasting reports);
- Phân tích xử lý trực tuyến (Online Analytical Processing - OLAP);
- Khai thác dữ liệu (Data mining);
- Phân tích thống kê (Statistical analysis);
- Dự đoán (Forecasting);
- Xác định xu hướng thị trường (Market trend analysis);
- Đánh giá hành vi khách hàng (Customer behavior analysis);
- Tối ưu hóa chuỗi cung ứng (Supply chain optimization).

Sự ra đời của BI giúp doanh nghiệp tối ưu hóa quá trình ra quyết định, tiết kiệm thời gian và nâng cao độ chính xác của các báo cáo và phân tích, từ đó nâng cao hiệu quả hoạt động kinh doanh.

2.3 Các thành phần trong hệ thống BI

Data Sources (Nguồn dữ liệu)

Bước đầu tiên trong quy trình BI là thu thập dữ liệu từ các nguồn khác nhau. Dữ liệu có thể đến từ nhiều hệ thống khác nhau như hệ thống quản lý nhân sự (HRM), quản lý quan hệ khách hàng (CRM), các trang thương mại điện tử, hệ thống quản lý tài chính, các phần mềm kế toán, và nhiều nguồn khác. Việc này đảm bảo rằng tất cả các thông tin cần thiết được thu thập đầy đủ để phục vụ cho các bước tiếp theo.

Data Warehousing (Kho dữ liệu)

Dữ liệu sau khi thu thập sẽ được chuyển đến kho dữ liệu (Data Warehouse). Đây là nơi lưu trữ lâu dài dữ liệu bằng các hệ thống thiết bị điện tử của doanh nghiệp. Kho dữ liệu được thiết kế để lưu trữ khối lượng dữ liệu lớn một cách có hệ thống và dễ dàng truy xuất, đảm bảo tính nhất quán và toàn vẹn của dữ liệu.

Integrating Server (Tích hợp máy chủ)

Trong giai đoạn này, công cụ ETL (Extract, Transform, Load) được sử dụng để chuyển đổi dữ liệu từ các nguồn dữ liệu ban đầu (Data Sources) sang kho dữ liệu (Data Warehouse). Quá trình ETL bao gồm:

- Extract (Trích xuất): Lấy dữ liệu từ các nguồn khác nhau.
- Transform (Chuyển đổi): Chuyển đổi dữ liệu sang định dạng chuẩn hóa và phù hợp cho việc lưu trữ trong kho dữ liệu.
- Load (Nạp): Nạp dữ liệu đã chuyển đổi vào kho dữ liệu.

Analysis Server (Máy chủ phân tích)

Dữ liệu trong kho dữ liệu sau đó sẽ được đưa vào máy chủ phân tích (Analysis Server). Tại đây, dữ liệu sẽ được xử lý và phân tích dựa trên các nghiệp vụ đã được định nghĩa sẵn. Máy chủ phân tích sẽ tiến hành các thao tác như tổng hợp, so sánh, và phân tích dữ liệu để đưa ra các kết quả có ý nghĩa, giúp hỗ trợ cho việc ra quyết định.

Reporting Server (Máy chủ báo cáo)

Máy chủ báo cáo nhận các output từ máy chủ phân tích và tạo ra các báo cáo chi tiết. Các báo cáo này có thể dưới dạng bảng biểu, biểu đồ, hoặc các dạng trực quan khác. Máy chủ báo cáo giúp cung cấp các thông tin phân tích chi tiết và dễ hiểu cho người dùng cuối, hỗ trợ cho việc ra quyết định.

Data Mining (Khai thác dữ liệu)

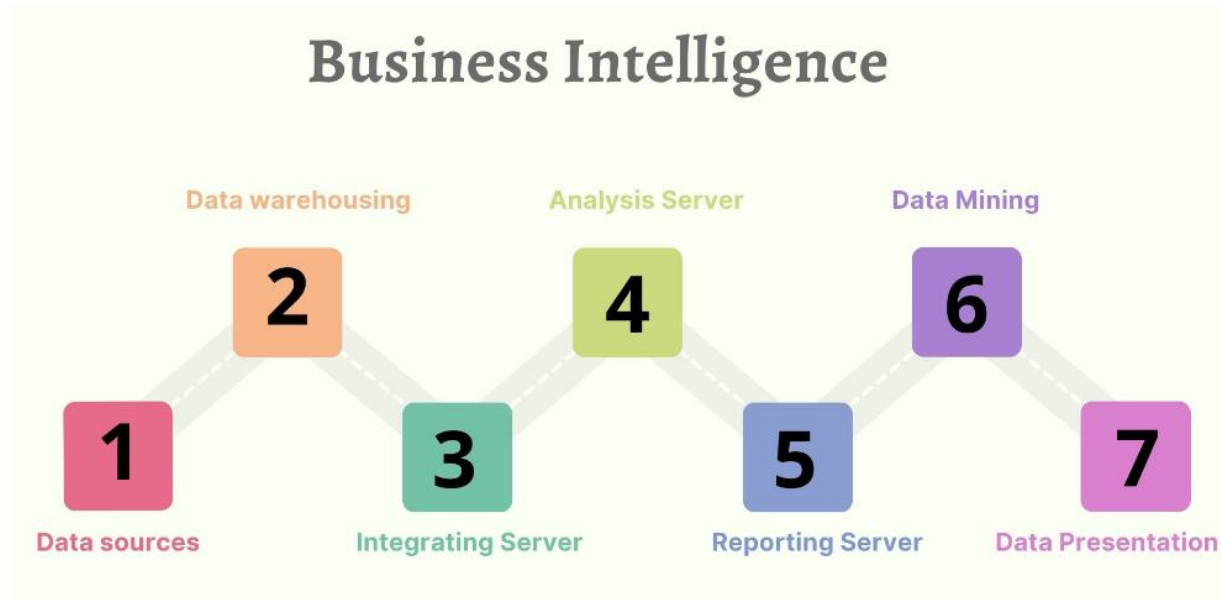
Khai thác dữ liệu là một bước quan trọng trong hệ thống BI. Dữ liệu đã được xử lý và phân tích sẽ được khai thác để tìm ra các mẫu và xu hướng tiềm ẩn. Quá trình này bao gồm:

- Classification (Phân loại): Phân loại dữ liệu vào các nhóm khác nhau.
- Clustering (Phân cụm): Nhóm các đối tượng dữ liệu có đặc điểm tương tự nhau vào cùng một cụm.

- Prediction (Dự đoán): Dự đoán các xu hướng hoặc kết quả trong tương lai dựa trên dữ liệu hiện tại và quá khứ.

Data Presentation (Trình bày dữ liệu)

Cuối cùng, dữ liệu đã được khai thác và phân tích sẽ được trình bày dưới dạng các biểu đồ và báo cáo trực quan. Giai đoạn này giúp chuyển đổi các thông tin phức tạp thành các dạng dễ hiểu và có thể hành động được, chẳng hạn như các biểu đồ động (graphical chart), các chỉ số KPI, và các báo cáo phân tích. Việc trình bày dữ liệu một cách trực quan giúp người dùng cuối dễ dàng nắm bắt thông tin và ra quyết định một cách hiệu quả hơn.



2.4 Tổng quan về BI

2.4.1 Ứng dụng

2.4.1.1 Trực quan hóa dữ liệu

Phần mềm BI sử dụng một loạt các công cụ phân tích dữ liệu được thiết kế để phân tích và quản lý thông tin liên quan đến hoạt động kinh doanh của bạn. Các dữ liệu này, khi được trình bày dưới dạng trực quan, giúp tổ chức theo dõi hiệu suất hoạt động hậu cần, bán hàng, năng suất và nhiều khía cạnh khác.

Một số nền tảng BI cung cấp khả năng báo cáo tùy chỉnh, cho phép người dùng chỉ định các thông số riêng của họ. Trong khi đó, các nền tảng khác cung cấp các mẫu báo cáo sẵn có với các chỉ số tiêu chuẩn ngành.

Bằng cách trình bày dữ liệu dưới dạng hình ảnh trực quan và định dạng dễ hiểu, hệ thống BI cho phép ngay cả những nhân viên ít kinh nghiệm nhất cũng có thể hiểu và rút ra thông tin chi tiết từ dữ liệu. Thay vì phụ thuộc vào các nhà khoa học dữ liệu được đào tạo chuyên sâu, bạn có thể tự phân tích và trình bày dữ liệu của mình cho cổ đông, các bộ phận khác hoặc nhóm của bạn.

2.4.1.2 Quản lý hiệu suất

Với các ứng dụng BI, các tổ chức có thể theo dõi tiến độ mục tiêu dựa trên các khung thời gian đã được xác định trước hoặc có thể tùy chỉnh. Các mục tiêu hướng tới dữ liệu bao gồm thời hạn hoàn thành dự án, mục tiêu thời gian giao hàng, hoặc mục tiêu bán hàng. Chẳng hạn, nếu bạn muốn đạt một mục tiêu bán hàng nhất định, hệ thống BI của bạn có thể phân tích dữ liệu từ các tháng trước đó và đề xuất một mục tiêu hợp lý để hướng tới dựa trên hiệu suất trong quá khứ.

Bảng điều khiển Power BI cung cấp thông tin chi tiết về các chỉ số quản lý hiệu suất. Các mục tiêu này có thể được theo dõi chặt chẽ để cung cấp thông tin cập nhật thường xuyên về tiến độ, giúp bạn nhận ra những khoảng trống còn lại. Người dùng có thể thiết lập hệ thống cảnh báo khi họ sắp đạt được mục tiêu hoặc nếu thời hạn kết thúc mà vẫn chưa đạt mục tiêu. Điều này giúp các nhà quản lý và nhân viên luôn cập nhật tiến độ của mình và giúp các nhóm tập trung vào mục tiêu.

Người dùng cũng có thể theo dõi việc hoàn thành mục tiêu và sử dụng dữ liệu tiến độ để đánh giá năng suất tổng thể của tổ chức. Thay vì tốn nhiều thời gian để theo dõi hoặc sắp xếp dữ liệu khẩn cấp, thông tin luôn dễ dàng truy cập. Điều này giúp doanh nghiệp tiết kiệm thời gian và chi phí.

2.4.1.3 Báo cáo

Một ứng dụng quan trọng của BI trong kinh doanh là báo cáo. Các công cụ BI thu thập và phân tích các tập dữ liệu phi cấu trúc, sau đó tổ chức và sử dụng chúng để tạo ra nhiều loại báo cáo khác nhau, bao gồm các báo cáo về nhân sự, chi phí, bán hàng, dịch vụ khách hàng và các quy trình khác.

Mặc dù báo cáo và phân tích dữ liệu có nhiều điểm tương đồng, chúng khác nhau đáng kể về mục đích, phân phối, nhiệm vụ và giá trị. Báo cáo là quá trình sắp xếp dữ liệu dưới dạng tóm tắt nhằm theo dõi hoạt động kinh doanh. Ngược lại, phân tích là quá trình khám phá dữ liệu để rút ra những hiểu biết sâu sắc có thể được áp dụng để cải thiện thực tiễn kinh doanh.

Về cơ bản, báo cáo biến dữ liệu thành thông tin đơn giản, trong khi phân tích biến dữ liệu thành thông tin chi tiết hữu ích. Cả hai phương pháp đều giúp doanh nghiệp cải thiện hiệu suất và giám sát hoạt động, nhưng chúng sử dụng các phương pháp khác nhau để đạt được mục tiêu này. Báo cáo cung cấp cho người dùng thông tin về những gì đang xảy ra, trong khi phân tích giải thích lý do tại sao điều đó xảy ra. Cả hai quy trình đều có thể sử dụng hình ảnh hóa dữ liệu, nhưng không nhất thiết phải luôn như vậy.

Các công cụ BI lý tưởng để xử lý dữ liệu động. Trước đây, việc trực quan hóa dữ liệu thường tĩnh và phải tạo mới mỗi khi có thay đổi biến. Phần mềm BI hiện đại cung cấp các bảng điều khiển tương tác có thể cập nhật theo thời gian thực, mang lại một cấp độ mới về khả năng sử dụng và sự linh hoạt trong phân tích dữ liệu.

2.4.2 Xu hướng xây dựng BI

Bên cạnh các nhà quản lý BI, nhóm ứng dụng BI thường bao gồm các kiến trúc sư BI, nhà phát triển, nhà phân tích nghiệp vụ và các chuyên gia quản lý dữ liệu BI. Người sử dụng nghiệp vụ cũng tham gia vào nhóm dự án, đại diện cho phía nghiệp vụ và đảm bảo các yêu cầu nghiệp vụ cần thiết được đáp ứng trong quá trình phát triển BI.

Để hỗ trợ quy trình này, ngày càng nhiều tổ chức chuyển từ mô hình phát triển kiểu thác nước sang Agile BI. Các phương pháp tiếp cận Agile BI và kho dữ liệu sử dụng kỹ thuật phát triển phần mềm Agile để chia nhỏ dự án BI thành các phần nhỏ và phát hành

các chức năng cho phân tích nghiệp vụ trên cơ sở lặp và nâng cấp dần. Cách tiếp cận này cho phép các doanh nghiệp đưa các tính năng BI vào thực tiễn nhanh hơn, cũng như làm mịn hoặc điều chỉnh các kế hoạch phát triển khi có thay đổi nghiệp vụ cần thiết hoặc khi xuất hiện các yêu cầu mới, ưu tiên cao hơn các vấn đề cũ.

Việc chuyển sang Agile BI giúp các tổ chức linh hoạt hơn, có khả năng phản ứng nhanh hơn với những thay đổi và nhu cầu mới trong môi trường kinh doanh, từ đó tăng hiệu quả và khả năng cạnh tranh.

2.4.3 Lợi ích BI

Khả năng hiển thị dữ liệu

Một nền tảng BI tốt cung cấp khả năng hiển thị đầy đủ tất cả dữ liệu quan trọng mà doanh nghiệp có. Không chỉ dừng lại ở việc xem dữ liệu, mà còn cung cấp cái nhìn chi tiết hơn về dữ liệu và hiển thị theo thời gian thực về mọi khía cạnh của doanh nghiệp. Với khả năng hiển thị chi tiết này, các nhà lãnh đạo có thể thích ứng nhanh chóng với những thay đổi trong doanh nghiệp.

Báo cáo chính xác

BI cho phép theo dõi chính xác tất cả các chỉ số KPI và tạo ra các báo cáo chính xác ở mọi cấp độ. Nhờ vào dữ liệu được trực quan hóa, các nhà lãnh đạo có thể phân tích các xu hướng, bao gồm dữ liệu quá khứ và thực hiện một số phân tích dự đoán cho tương lai của doanh nghiệp. BI cũng cho phép xây dựng các chế độ xem cụ thể cho các bên liên quan chính, giúp họ có thể truy cập và xem những dữ liệu quan trọng một cách dễ dàng.

Hợp lý hóa các quy trình

BI giúp loại bỏ sự phức tạp liên quan đến các quy trình và tự động hóa việc phân tích. Nó cung cấp các phân tích dự đoán, lập mô hình máy tính, đo điểm chuẩn và các phương pháp luận khác, giúp tối ưu hóa và đơn giản hóa quy trình làm việc.

Ra quyết định rõ ràng

BI nâng cao tính minh bạch của dữ liệu, từ đó cải thiện chất lượng của việc ra quyết định. Ngay cả những người không chuyên về kỹ thuật hoặc không phải là nhà phân tích cũng có thể nhanh chóng thu thập và xử lý dữ liệu. Điều này giúp mở rộng sức mạnh của phân tích đến nhiều người dùng hơn, tạo điều kiện cho việc ra quyết định dựa trên dữ liệu trở nên dễ dàng và hiệu quả hơn.

III. ỨNG DỤNG DATAWAREHOUSE VÀ BI VÀO PHÂN TÍCH DỮ LIỆU LĨNH VỰC BÁN LẺ

3.1 Tổng quan về bài toán

3.1.1 Tổng quan về bán lẻ

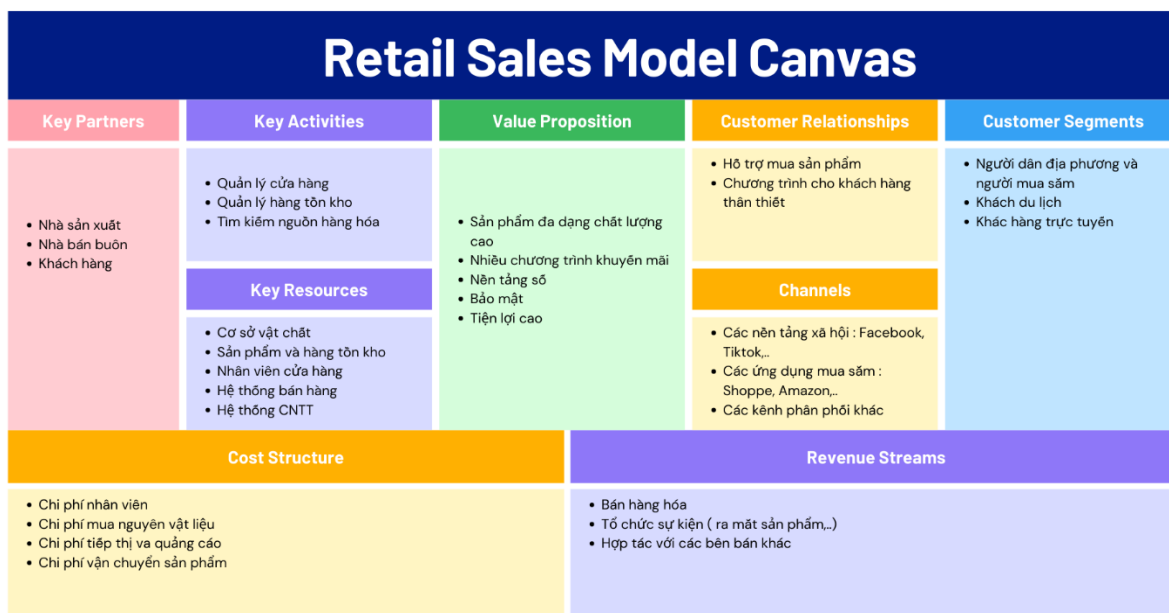
Retail sales là hoạt động bán hàng trực tiếp cho người tiêu dùng cuối cùng thông qua các cửa hàng, siêu thị, trang web bán lẻ, và các kênh phân phối khác. Nó bao gồm việc bán các sản phẩm và dịch vụ từ nhà sản xuất hoặc nhà cung cấp đến người mua hàng.

Vai trò:

- **Kinh tế:** Retail sales đóng vai trò quan trọng trong kinh tế bằng cách tạo ra doanh số bán hàng, tăng trưởng kinh tế và cung cấp việc làm cho một số lượng lớn người lao động.
- **Tiếp thị và Bán hàng:** Retail sales là cơ hội để tiếp cận và tương tác trực tiếp với khách hàng. Các doanh nghiệp sử dụng các chiến lược tiếp thị và bán hàng để thu hút và duy trì khách hàng.
- **Phân phối:** Retail sales cung cấp một kênh phân phối quan trọng giữa nhà sản xuất và người tiêu dùng. Nó giúp đưa sản phẩm từ nguồn cung đến người tiêu dùng ở nhiều địa điểm khác nhau.
- **Dịch vụ khách hàng:** Retail sales cung cấp một nền tảng để cung cấp dịch vụ khách hàng, bao gồm hỗ trợ sau bán hàng, đổi trả và bảo hành.

Kết luận:

Retail sales không chỉ là quá trình bán hàng, mà còn là một phần quan trọng của nền kinh tế và xã hội. Vai trò của nó không chỉ là tạo ra doanh số và lợi nhuận, mà còn là cơ hội để tạo ra mối quan hệ với khách hàng và cung cấp giá trị cho cả người tiêu dùng và doanh nghiệp.



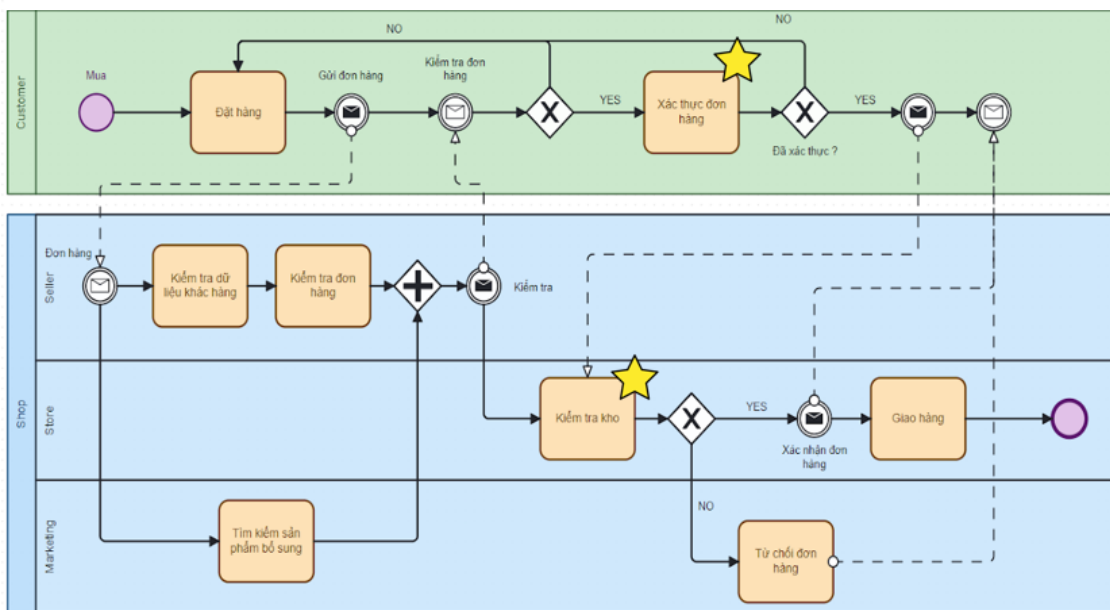
Hình 3. 1 Retail Sales Model Canvas

3.1.2 Giới thiệu về bài toán

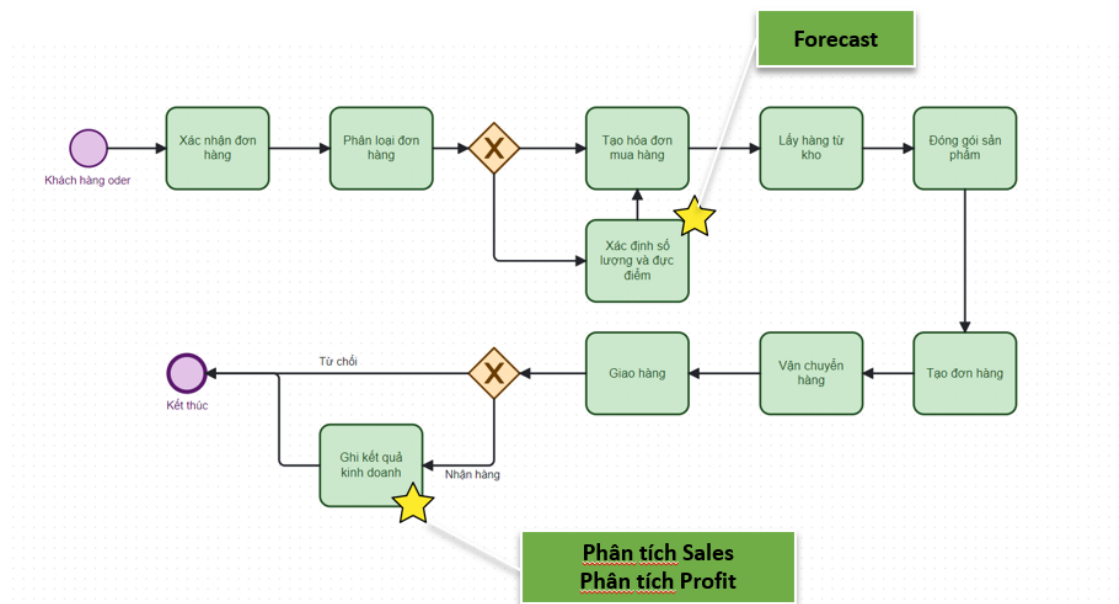
Tập dữ liệu được sử dụng trong bài phân tích thuộc lĩnh vực bán hàng tại thị trường trong giai đoạn 2014 – 2017. Dữ liệu gồm 3 sheet, Order gồm 22 thuộc tính và 9994 dữ liệu, Returned có 2 thuộc tính và 296 dữ liệu, People gồm 2 thuộc tính và 4 dữ liệu. Đơn hàng được thực hiện trong suốt 04 năm bao gồm 793 khách hàng khác nhau được chia ra làm 03 bộ phận khách hàng chính: Người tiêu dùng (Consumer), tổ chức (Corporate), văn phòng (Home Office).

Product				Data			District			People
Product Name			Oder ID	Moth	Quarter	Year	City	State	Regison	Name
Furniture	Office Supplie	Technology	5009	12	4	4	532	49	4	4
567	432	851	CA-2016-152156	1	1	2014	Henderson	Kentucky	West	Anna Andreadi
Bush Somerse	Self-Adhesive	Mitel 5320 II	CA-2016-138688	2	2	2015	Los Angeles	California	East	Cassandra Brandow
Hon Deluxe F	Eldon Fold 'N	Konftel 250	US-2015-108966	3	3	2016	Fort Lauderc	Florida	Central	Kelly Williams
Bretford CR4	Newell 322	Cisco SPA 5	CA-2014-115812	4	4	2017	Concord	Texas	South	Chuck Magee
Eldon Express	DXL Angle-Vi	Imation 8GB	CA-2017-114412	5			Seattle	Wisconsin		
Chromcraft R	Belkin F5C206	GE 30524EE	CA-2016-161389	6			Fort Worth	Utah		
Global Deluxe	Xerox 1967	Plantronics F	US-2015-118983	7			Madison	Pennsylvania		
Bretford CR4	Fellowes PB2	Panasonic K	CA-2014-105893	8			West Jordan	Illinois		

Hình 3. 2 Data flow



Hình 3. 3 Quy trình mua hàng chung



Hình 3. 4 Sơ đồ quy trình nghiệp vụ

Sau khi xác định được dữ liệu, nhóm bắt đầu xác định các chủ điểm phân tích (Requirement), bao gồm: Sale, Profit, Forecast : các FACT được phân tích dựa trên các DIM như: Order Date (Month, Quarter, Year), Ship Mode, Segment, Region, State, City, Category, Sub - Category, Product.

Việc phân tích các chủ điểm dựa vào các DIM trên sẽ giúp người sử dụng có cái nhìn tổng quát nhất về kết quả kinh doanh theo nhiều hướng khác nhau , từ đó cung cấp những thông tin cần thiết cho việc ra quyết định (chân dung khách hàng tiềm năng, dòng sản phẩm được sử dụng nhiều, khu vực bán chạy nhất, dự báo xu hướng sử dụng sản phẩm, doanh thu của sản phẩm...) từ đó cải thiện hiệu suất kinh doanh.

3.1.3 Khám phá dữ liệu

Sales
\$ 2.297.200,86

Profit
\$ 286.397,02

Returnd
296

Oder ID
5009

Customer ID
793

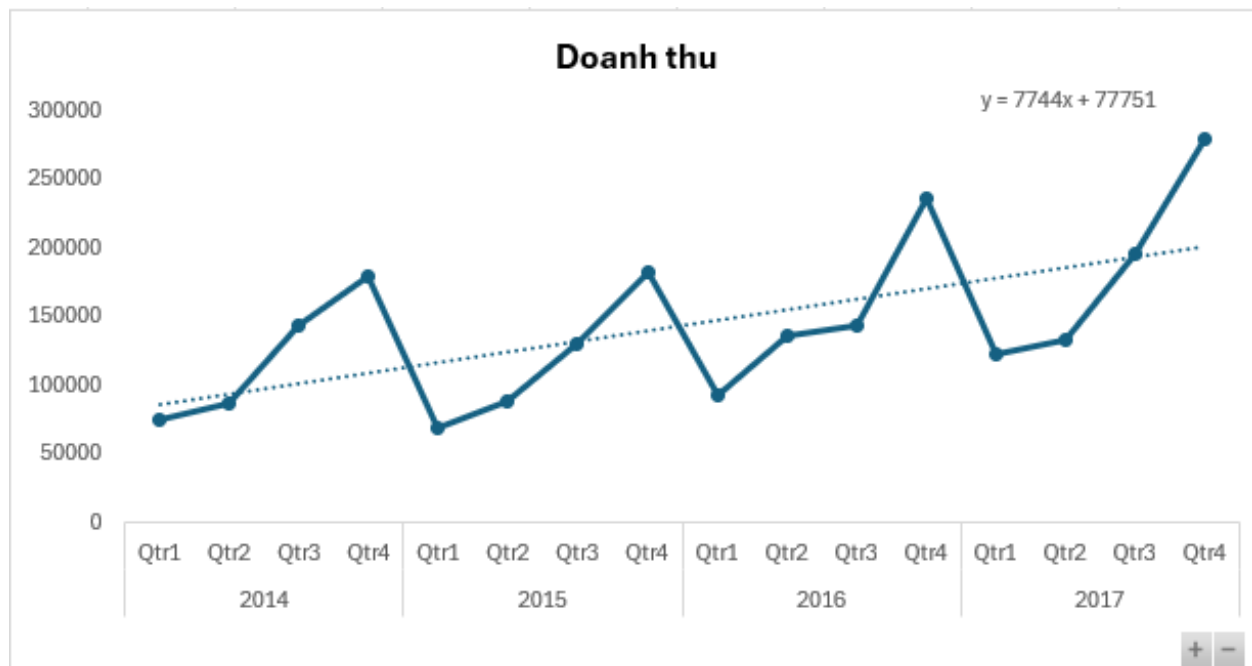
Product Name
1850

Region
4

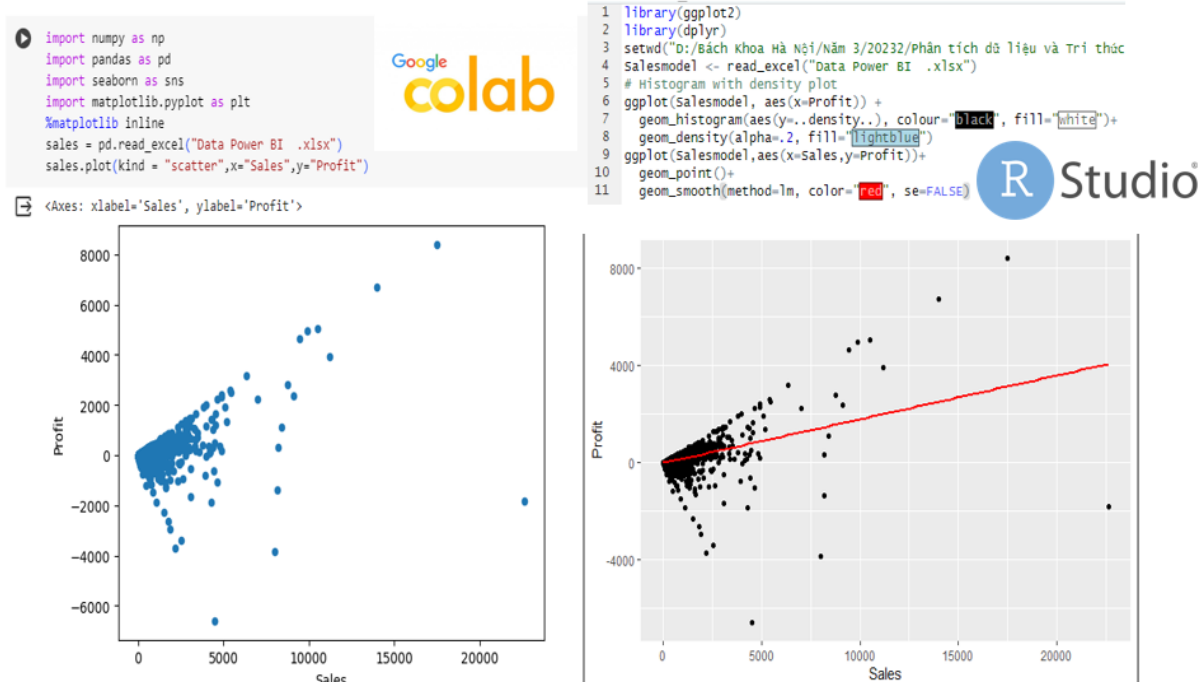
State
49

City
532

Hình 3. 5 Khám phá dữ liệu

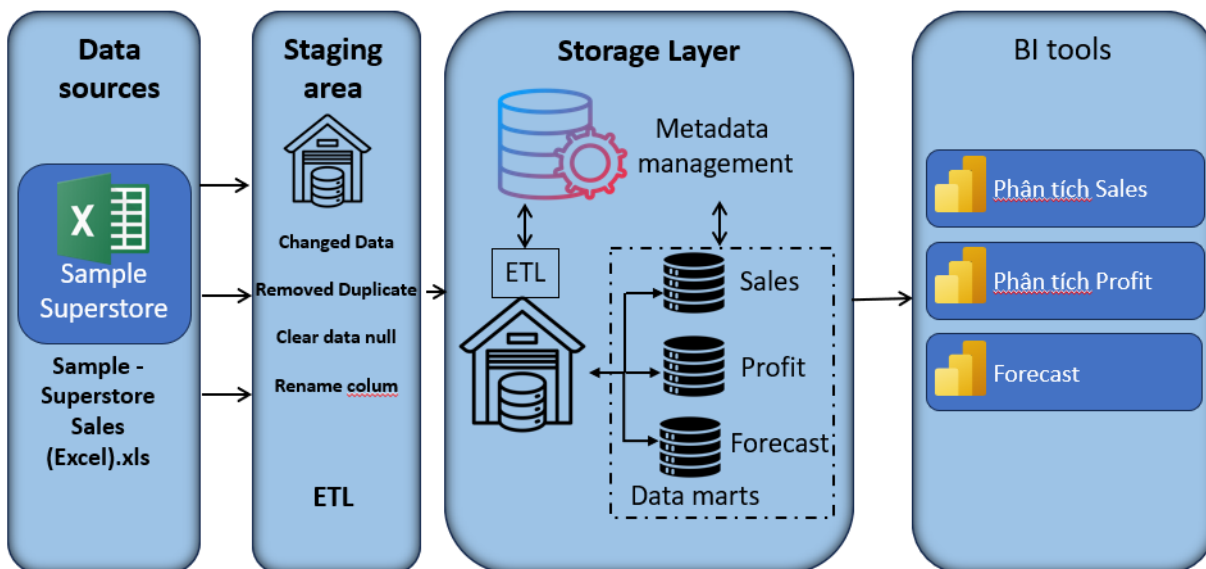


Hình 3. 6 Biểu đồ khuynh hướng doanh thu



Hình 3. 7 Biểu đồ Scatter Chart giữa doanh thu và lợi nhuận

1.1 Kiến trúc Data Warehouse



Hình 3. 8 Kiến trúc Data Warehouse

Tầng đầu tiên của kiến trúc là Data Source - “Sample Superstore Sales” được lấy từ hệ cơ sở dữ liệu hoạt động từ các nguồn bên ngoài (file csv, excel...). Những công cụ và tiện ích này thực hiện việc loại bỏ dữ liệu thừa, làm sạch dữ liệu, chuyển đổi dữ liệu, cập nhật dữ liệu. Dữ liệu được lưu trữ vào 3 bảng:

(1) Bảng Detail Order

(2) Bảng Customer

(3) Bảng Order

(4) Bảng Product

(5) Bảng Category

Tầng tiếp theo là việc sử dụng các công cụ tích hợp trên Excel hoặc Power Query nhằm ETL - tiền xử lý dữ liệu trước khi đưa vào Data Warehouse như: Changed Data, Removed Duplicate, Clear data null, Rename column, ...

Tầng giữa là “Storage Layer” bao gồm các Data Warehouse và Data marts được cài đặt dùng mô hình quan hệ OLAP.

Tầng trên cùng là tầng người dùng cuối, gồm các câu truy vấn và các công cụ làm báo cáo, phân tích, công cụ khai thác dữ liệu, thông qua công cụ Power BI về:

(1) Forecast: Dự báo số lượng

(2) Sale: Doanh thu

(3) Profit: Lợi nhuận thu được

3.3 Tiềm xử lý dữ liệu

Đối với dữ liệu được sử dụng trong bài toán, nhóm tiến hành các thao tác như: Remove Column, Data type, xóa dữ liệu null, xóa dữ liệu trùng lặp, xử dụng các hàm trong excel, transform dữ liệu sang Power BI.

Data Type

A _C Order ID	Order Date	Ship Date	A _C Ship Mode
CA-2016-152156	11/8/2016 12:00:00 AM	11/11/2016 12:00:00 AM	Second Class
CA-2016-152156	11/8/2016 12:00:00 AM	11/11/2016 12:00:00 AM	Second Class
CA-2016-138688	6/12/2016 12:00:00 AM	6/16/2016 12:00:00 AM	Second Class
US-2015-108966	10/11/2015 12:00:00 AM	10/18/2015 12:00:00 AM	Standard Class
US-2015-108966	10/11/2015 12:00:00 AM	10/18/2015 12:00:00 AM	Standard Class
CA-2014-115812	6/9/2014 12:00:00 AM	6/14/2014 12:00:00 AM	Standard Class
CA-2014-115812	6/9/2014 12:00:00 AM	6/14/2014 12:00:00 AM	Standard Class
CA-2014-115812	6/9/2014 12:00:00 AM	6/14/2014 12:00:00 AM	Standard Class
CA-2014-115812	6/9/2014 12:00:00 AM	6/14/2014 12:00:00 AM	Standard Class
CA-2014-115812	6/9/2014 12:00:00 AM	6/14/2014 12:00:00 AM	Standard Class
CA-2014-115812	6/9/2014 12:00:00 AM	6/14/2014 12:00:00 AM	Standard Class

Order Date	Ship Date
11/8/2016	11/11/2016
11/8/2016	11/11/2016
6/12/2016	6/16/2016
10/11/2015	10/18/2015
10/11/2015	10/18/2015
6/9/2014	6/14/2014
6/9/2014	6/14/2014

Hình 3. 9 Định dạng lại ngày

Remove Colum

A _C Merged	A _C Merged.1	A _C Ship Mode
8/11/2016	11/11/2016	Second Class
8/11/2016	11/11/2016	Second Class
12/6/2016	16/6/2016	Second Class
11/10/2015	18/10/2015	Standard Class
11/10/2015	18/10/2015	Standard Class
9/6/2014	14/6/2014	Standard Class
9/6/2014	14/6/2014	Standard Class
9/6/2014	14/6/2014	Standard Class

i ₂₃ Order Date.1	i ₂₃ Order Date.2	i ₂₃ Order Date.3
11	8	2016
11	8	2016
6	12	2016
10	11	2015
10	11	2015
6	9	2014
6	9	2014

Hình 3. 10 Sắp xếp lại thứ tự thời gian

Xóa dữ liệu null

A _C Customer ID	A _C Customer Name
null	null
null	null
null	null
null	null
null	null
null	null

A _C Customer ID	A _C Customer Name
TS-21655	Trudy Schmidt
VD-21670	Valerie Dominguez
VM-21685	Valerie Mitchum
VT-21700	Valerie Takahito
VF-21715	Vicky Freymann
VP-21730	Victor Preis

Hình 3. 11 Xóa dữ liệu null

Xóa dữ liệu trùng lặp

Customer ID	Customer Name
AB-10015	Aaron Bergman
AB-10015	Aaron Bergman
AB-10015	Aaron Bergman
AB-10015	Aaron Bergman
AB-10015	Aaron Bergman

Customer ID	Customer Name
AB-10015	Aaron Bergman
AH-10030	Aaron Hawkins
AS-10045	Aaron Smayling
AB-10060	Adam Bellavance
AH-10075	Adam Hart
AS-10090	Adam Shillingsburg

Hình 3. 12 Xóa dữ liệu trùng lặp Customer ID

Transform dữ liệu sang Power BI

Queries

fx = Table.RemoveColumns(#"Changed Type",{ "Returned_1", "Column24", "Column23" })

	Order ID	Returned	Order Date	Ship Date
1	CA-2016-152156	0	11/8/2016	11/11/2016
2	CA-2016-152156	0	11/8/2016	11/11/2016
3	CA-2016-138688	0	6/12/2016	6/16/2016
4	US-2015-108966	0	10/11/2015	10/18/2015
5	US-2015-108966	0	10/11/2015	10/18/2015
6	CA-2014-115812	0	6/9/2014	6/14/2014
7	CA-2014-115812	0	6/9/2014	6/14/2014
8	CA-2014-115812	0	6/9/2014	6/14/2014
9	CA-2014-115812	0	6/9/2014	6/14/2014
10	CA-2014-115812	0	6/9/2014	6/14/2014
11	CA-2014-115812	0	6/9/2014	6/14/2014

Hình 3. 13 Transform dữ liệu

3.4 Mô hình dữ liệu OLTP

3.4.1 Mô tả cấu trúc trong cơ sở dữ liệu

Bảng FACT_DetailOder

STT	Tên thuộc tính	Kiểu
1	Oder ID	Nvarchar(255)
2	Product ID	Nvarchar(255)
3	Customer ID	Nvarchar(255)
4	Discount	Float
5	PriceBuy	Money

6	Profit	Money
7	Quantity	Int
8	Sale	Money

Hình 3. 14 Cấu trúc Detail Oder

Bảng DIM_Oder

STT	Tên thuộc tính	Kiểu
1	Customer ID	Nvarchar(255)
2	Oder ID	Nvarchar(255)
3	Oder Date	Nvarchar(255)
4	Ship Date	Nvarchar(255)
5	Ship Mode	Nvarchar(255)

Hình 3. 15 Cấu trúc Oder

Bảng DIM_Customer

STT	Tên thuộc tính	Kiểu
1	Customer ID	Nvarchar(255)
2	Customer Name	Nvarchar(255)
3	City	Nvarchar(255)
4	Country	Nvarchar(255)
5	Region	Nvarchar(255)
6	State	Nvarchar(255)
7	Segment	Nvarchar(255)
8	Postal Code	Int

Hình 3. 16 Cấu trúc Customer

Bảng DIM_Product

STT	Tên thuộc tính	Kiểu
1	Category ID	Nvarchar(255)

2	Product ID	Nvarchar(255)
3	Category	Nvarchar(255)
4	Product Name	Nvarchar(255)
5	PriceBuy	Money

Hình 3. 17 Cấu trúc Product

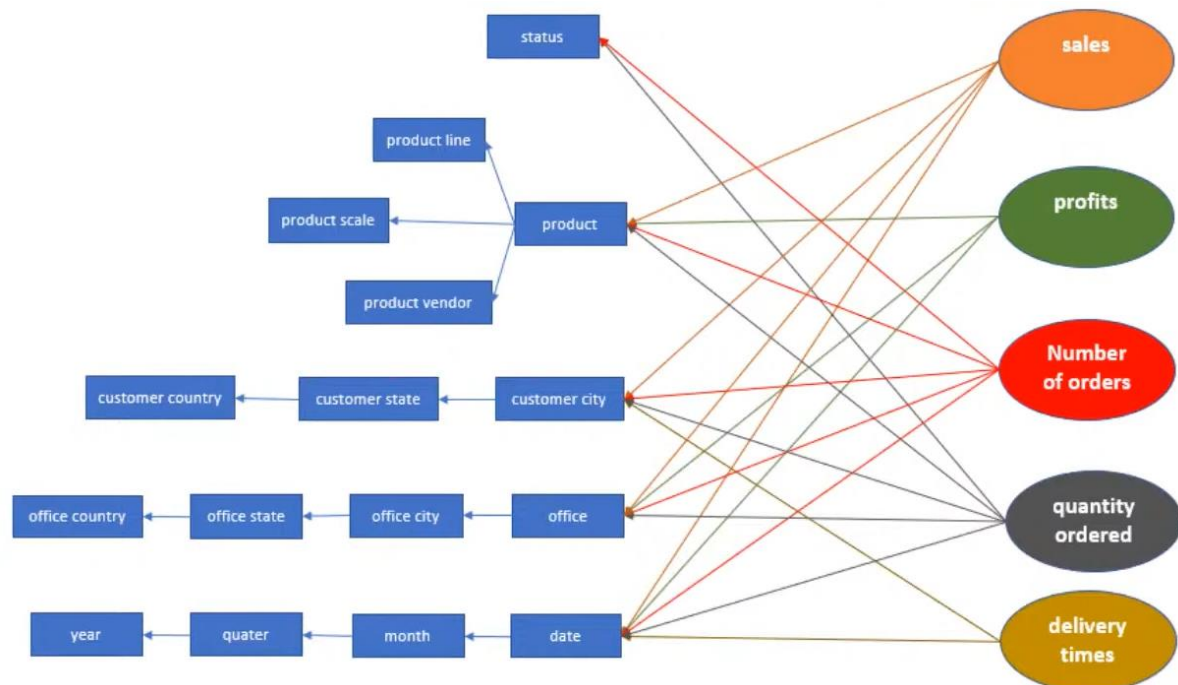
Bảng DIM_Category

STT	Tên thuộc tính	Kiểu
1	Category ID	Nvarchar(255)
2	Category	Nvarchar(255)
3	Sub - Category	Nvarchar(255)

Hình 3. 18 Cấu trúc Category

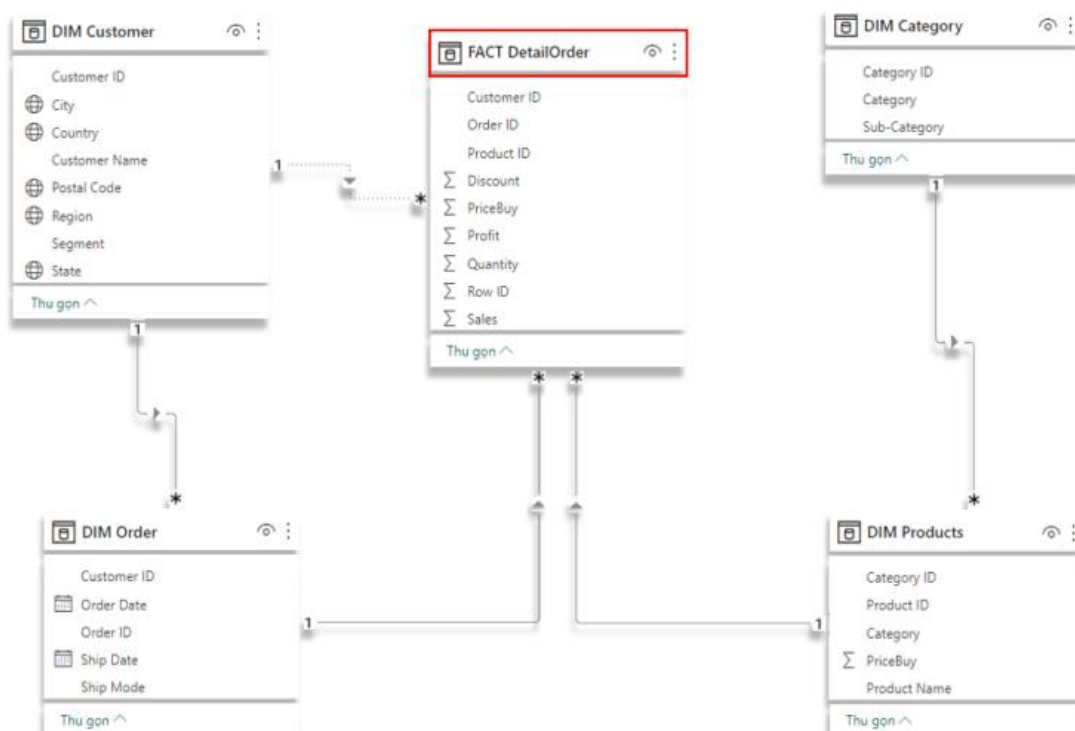
3.5 Mô hình dữ liệu OLAP

3.5.1 Mô hình dữ liệu OLAP-ER



Hình 3. 19 Mô hình dữ liệu OLAP- ER

3.5.2 Mô hình dữ liệu OLAP



Hình 3. 20 Mô hình dữ liệu OLAP

3.6 Các mẫu DashBoard

3.6.1 DashBoard - Doanh thu

Nhìn chung, doanh thu có xu hướng tăng qua các năm, đáng chú ý là doanh thu tập trung chủ yếu vào vào 04 tháng cuối năm do thời điểm này bắt đầu có nhiều chương trình giảm giá các dịp lễ cũng như mọi người đi mua sắm cho những dịp nghỉ cuối năm. Tổng doanh thu trong cả 4 năm là 2.297.201 \$ và đã bán được 9.994 nghìn sản phẩm. Mức ROA 0.12 chỉ số này cho biết 100 đồng vốn đầu tư vào kinh doanh tạo ra 0.12 đồng vốn chủ sở hữu. West chiếm tỷ trọng doanh thu cao nhất với mức 31,58% sát nút sau đó là East với 29,55% kế tiếp là 21,82% của Central và cuối cùng là 17,05% South, điều đó cho thấy doanh thu phân bố khá đồng đều các khu vực. Doanh thu chủ yếu danh mục là Phones và Chairs với doanh số gần tương đương nhau.

Tỷ lệ hoàn trả đơn hàng bằng 8% đó là 1 con số khá cao cho thấy nhà quản trị cần xem xét về vấn đề hoàn trả đơn hàng. Còn về tổng số đơn bị trả theo từng phương thức vận chuyển thì Standard Class chiếm nhiều nhất với 59,72% tổng số đơn bị trả, kế tiếp là 19,46% của phương thức Second Class, còn lại 2 phương thức chiếm 20,82%

Năm 2014 đạt 484.247 \$. Tại giai đoạn này thì danh mục Phones ghi nhận doanh số cao nhất với 77.390 \$. Tỷ suất sinh lời ROA với mức là 0.10 cũng khá thấp. Tỷ lệ hoàn

trả đơn hàng lên đến 40% khá cao cho thấy gặp vấn đề trong năm này. Doanh thu theo từng tháng có xu hướng tăng nhưng khá biến động, doanh thu đạt cao nhất tại tháng 9

Năm 2015 doanh thu đạt 470.533 \$ có giảm một chút so với năm ngoái với tổng số lượng sản phẩm bán ra là 2.102 nghìn sản phẩm tăng gần gấp đôi so với năm 2014, tuy nhiên doanh thu lại giảm cho thấy gặp vấn đề trong việc bán sản phẩm. Về tỷ lệ hoàn trả đơn hàng ở năm 2015 cũng khá cao. Mức sinh lời ROA năm 2015 là 0.13 cao hơn năm ngoái. Về doanh thu theo doanh mục sản phẩm đứng đầu là Chair với 71.734 \$ và doanh thu thấp nhất thuộc về Fasteners với 545 \$, Từ đó thì cần phải xác định nguyên nhân tại sao doanh thu sản phẩm Fasteners lại có doanh thu thấp để đưa ra quyết định có nên dừng sản xuất sản phẩm đó nữa không

Tiếp theo là doanh thu năm 2016 với mức 609.206 \$ tăng so với 2 năm trước với tổng số lượng sản phẩm bán ra là 2.57 sản phẩm. Trong đó doanh mục Chair vẫn chiếm doanh thu cao nhất cùng với Phones tiếp sau đó lần lượt là 83.918 \$ và 78.962 \$. Với mức tỷ lệ hoàn trả đơn hàng thấp hơn so với 2 năm trước cho thấy đã cải thiện về mặt hoàn trả đơn hàng nhưng không có ý nghĩa là không quan tâm đến. Tỷ lệ sinh lời ROA vẫn giữ nguyên mức 0.13 so với năm ngoái

Năm 2017 là năm có doanh thu cao nhất đạt 733.215 \$ với lượng tiêu thụ cũng khá là cao so với các năm còn lại là 3.312 sản phẩm. Cùng với doanh thu 2 danh mục chiếm cao nhất vẫn là Phones và Chair cho thấy 2 sản phẩm này đang rất được ưa chuộng từ đó đưa ra các chiến lược cải thiện hơn nữa tuy nhiên cũng cần phải xem xét đến danh mục Fasteners tại sao doanh thu khá thấp và thấp trong 4 năm liên tiếp. Về tỷ lệ hoàn trả đơn hàng đã cải thiện đáng kể với 24% những con số này vẫn còn cao. Doanh thu sẽ tăng dần vào 4 tháng cuối năm cho thấy doanh nghiệp cần đưa ra những chương trình khuyến mại để thu hút hơn nữa khách hàng mua sản phẩm



Hình 3. 21 DashBoard Sales 2014



Hình 3. 22 DashBoard Sales 2015



Hình 3. 23 DashBoard Sales 2016



Hình 3. 24 DashBoard Sales 2017

3.6.2 DashBoard – Lợi nhuận

Nhìn chung thì lợi nhuận qua các năm không thay đổi nhiều. Đáng chú ý là chủ yếu gần cuối năm lợi nhuận tăng nhẹ do thời điểm này bắt đầu có nhiều chương trình giảm giá các dịp lễ cũng như mọi người đi mua sắm cho những dịp nghỉ cuối năm. Tổng lợi nhuận từ năm 2014 – 2017 là 286.397 \$ với mức sản phẩm bán ra là 9.994 sản phẩm . Mặt hàng Technology chiếm lợi nhuận cao nhất là 145.454 \$ tiếp sau đó là Office Supplies với lợi nhuận là 122.491 \$ cuối cùng 18.451\$ của Furniture. Mức ROA của 4 năm 0.12 chỉ số này cho biết 100 đồng vốn đầu tư vào kinh doanh tạo ra 0.12 đồng vốn chủ sở hữu. Lợi nhuận phân bổ chủ yếu ở West chiếm 37,86% và boj phận chiếm lợi nhuận cao nhất Consumer với 46,83%.

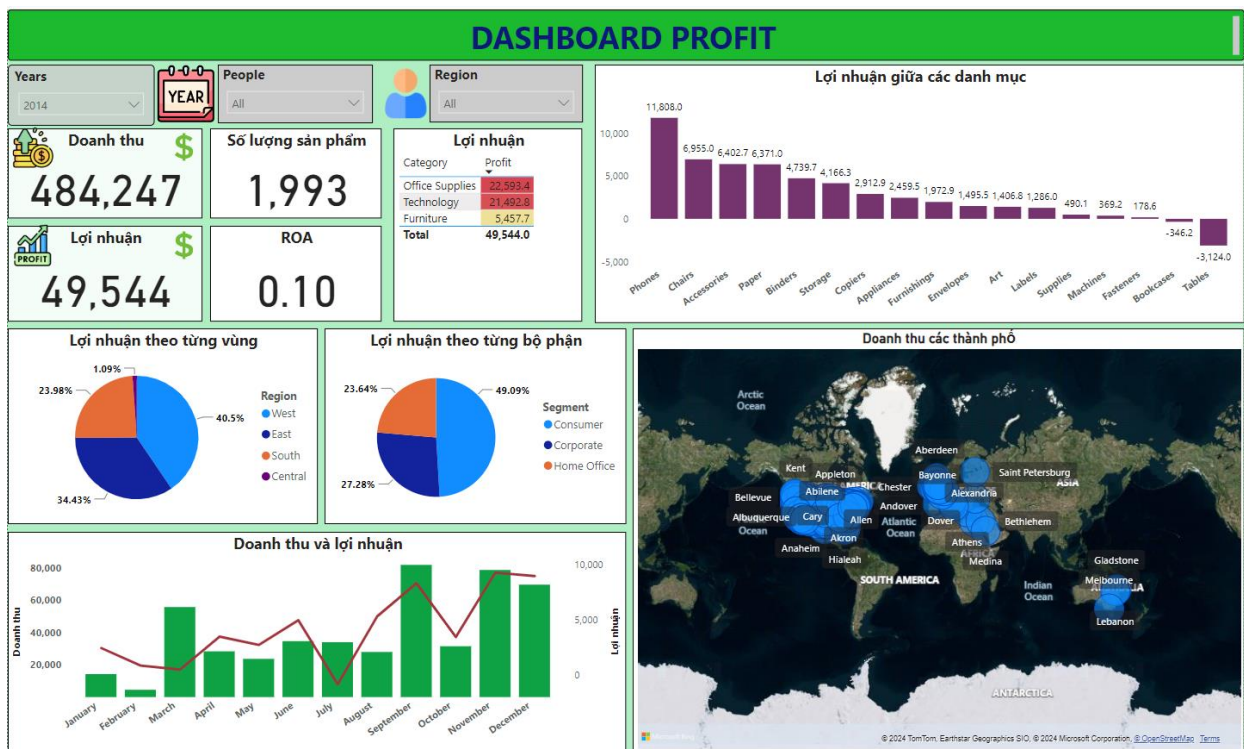
Năm 2014 ghi nhận được mức lợi nhuận 49.544 \$ với mức ROA 0.10 cho thấy lợi nhuận cũng khá thấp. Tổng số lượng sản phẩm bán ra trong năm này là 1.993 sản phẩm theo cơ cấu lợi nhuận thì năm này lợi nhuận đạt cao nhất tại Office Supplies tiếp sau đó là

Technology và cuối cùng là Furniture. Lợi nhuận theo từng vùng phân bố tại West chiếm 40,5% tổng số nhưng khu vực Central chỉ chiếm 1,09% tổng số đây là con số khá thấp. Lợi nhuận đầu năm 2014 sẽ thấp nhất và những tháng cuối năm sẽ tăng

Năm 2015 lợi nhuận tăng 1 chút so với năm 2014 với mức lợi nhuận là 61.619\$ tỷ lệ ROA tăng 1 chút là 0,13. Lợi nhuận giữa các doanh mục có sự phân bố không đều, cụ thể danh mục Phones chiếm doanh thu cao nhất với 10.399\$ sau đó là Accessories nhưng có 2 mặt hàng có lợi nhuận bị âm là Bookcases và tables

Lợi nhuận năm 2016 cũng tăng một chút so với năm 2015 là 81.795\$ cùng với mức ROA tương đương năm 2015 là 0,13. Số lượng sản phẩm bán ra 2.587 sản phẩm lợi nhuận theo Technology chiếm nhiều nhất với 39.774\$. Mặt hàng bán chạy nhất trong năm này là Copiers với mức lợi nhuận 17.742\$ thấp nhất đó là Furniture với 6.960\$. Lợi nhuận theo từng vùng phân bố khá đều trong năm này còn theo từng bộ phận thì Consumer chiếm 43,7% tiếp theo là Corporate với 37,89% cuối cùng 18,37% là Home Office

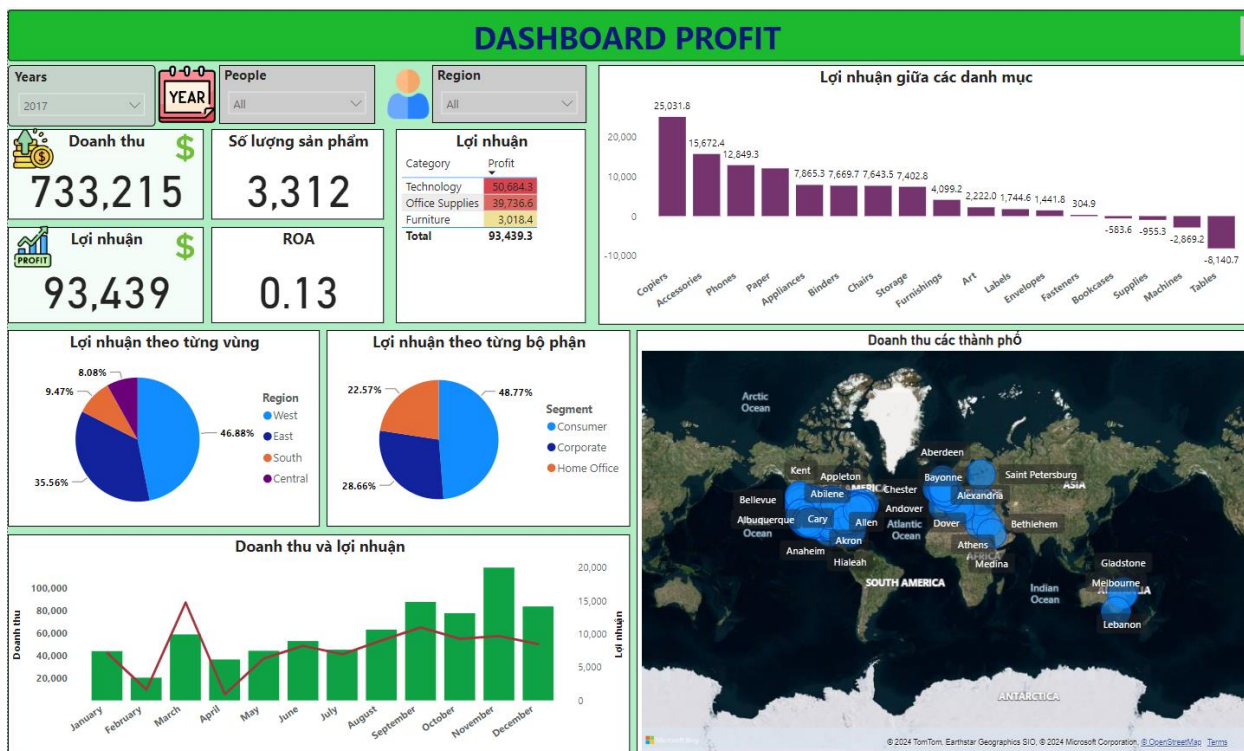
Năm 2017 với mức lợi nhuận là 93.215\$ cùng với 3.312 sản phẩm bán ra cho thấy năm này có lợi nhuận cao nhất trong 4 năm. Danh mục Copiers vẫn giữ lợi nhuận cao nhất 25.031\$ còn lại 2 danh mục làm cho lợi nhuận âm là Machines và Tables lần lượt là -2.869\$ và -8.140\$. Nhìn chung thì lợi nhuận năm này tăng chủ yếu vào cuối năm.



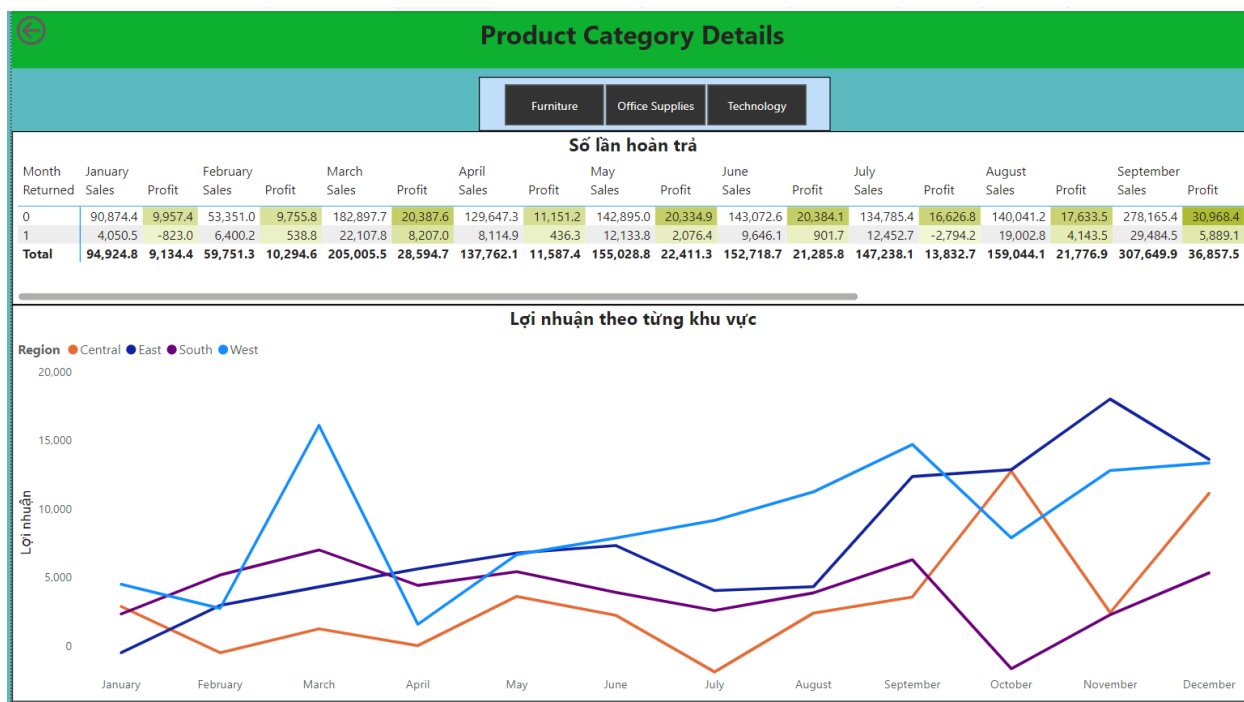
Hình 3. 25 DashBoard Profit 2014



Hình 3. 26 DashBoard Profit 2015



Hình 3. 28 DashBoard Profit 2017



Hình 3. 29 Product Category Detail

←

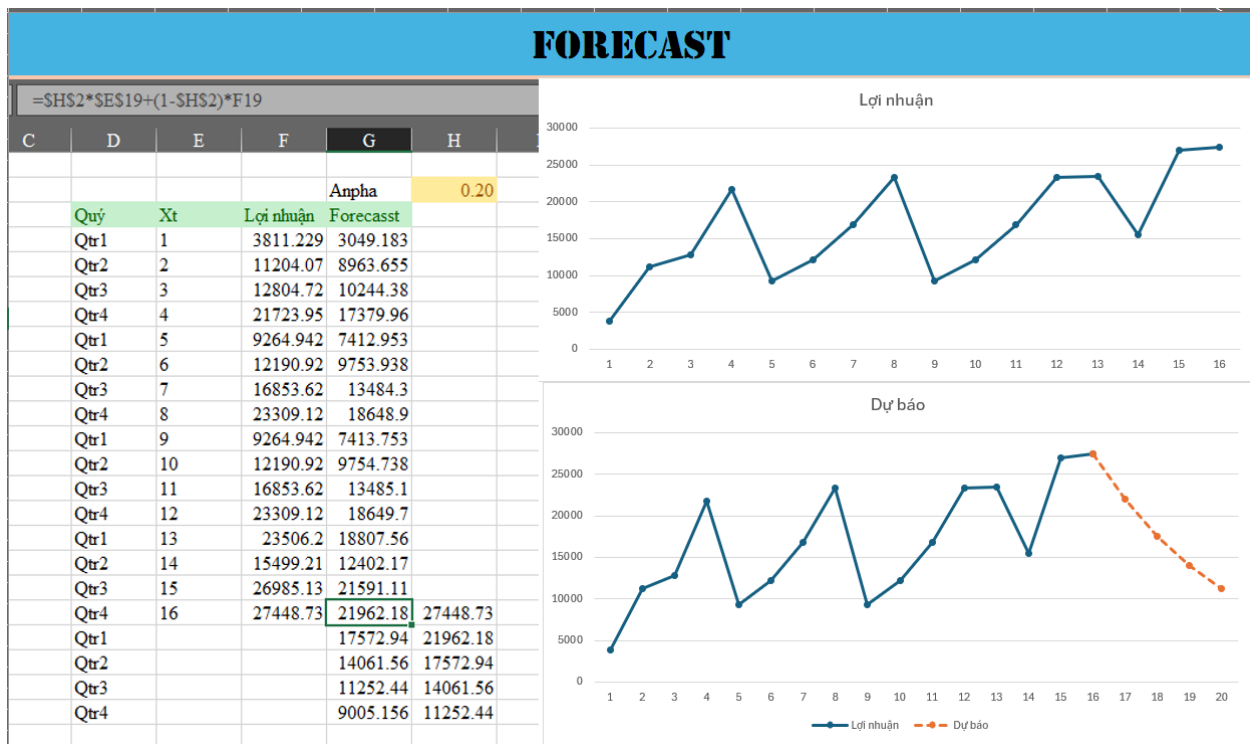
Detail Profit Sub -Category

	Select all	Accessories	Appliances	Art	Binders	Bookcases	Chairs	Copiers	Envelopes	Fasteners	Furnishings	Labels												
Month	January	February	March	April	May	June	July																	
Sub-Category	Sales	Profit	ROA	Sales	Profit	ROA	Sales	Profit	ROA	Sales	Profit	ROA	Sales	Profit										
Accessories	5,477.9	1,337.8	0.24	5,369.3	1,481.8	0.28	8,766.6	2,303.4	0.26	7,951.7	1,832.0	0.23	9,613.1	2,471.7	0.26	8,908.4	2,233.6	0.25	17,126.4	4,784.0	0.28	17,126.4	4,784.0	0.28
Appliances	3,176.2	598.8	0.19	4,933.0	1,111.5	0.23	6,733.6	204.4	0.03	6,041.8	-292.4	-0.05	7,526.0	1,206.5	0.16	7,479.5	1,489.1	0.20	3,823.3	604.3	0.16	3,823.3	604.3	0.16
Art	914.4	232.8	0.25	1,117.5	261.4	0.23	1,302.0	311.0	0.24	2,407.2	686.6	0.25	2,230.8	490.6	0.22	2,217.7	575.8	0.26	2,065.9	497.4	0.24	2,065.9	497.4	0.24
Binders	12,214.2	2,851.7	0.23	4,237.0	690.0	0.16	13,888.7	5,130.1	0.37	13,364.9	786.0	0.06	9,159.4	2,227.6	0.24	13,286.8	3,598.7	0.27	8,557.3	-3,397.1	0.27	8,557.3	-3,397.1	0.27
Bookcases	5,352.4	-551.2	-0.10	1,649.6	250.0	0.15	7,351.9	-681.6	-0.09	4,720.5	38.4	0.01	6,289.6	-238.5	-0.04	9,147.6	231.6	0.03	8,589.0	561.6	0.03	8,589.0	561.6	0.03
Chairs	11,285.2	901.6	0.08	7,582.9	764.0	0.10	21,344.3	1,752.0	0.08	18,526.6	1,659.6	0.09	25,893.8	1,769.2	0.07	21,523.4	1,483.8	0.07	23,016.4	765.5	0.07	23,016.4	765.5	0.07
Copiers	3,959.9	1,500.0	0.38				25,589.7	11,044.4	0.43	3,879.9	1,287.0	0.33	18,399.8	3,719.9	0.20	900.0	315.0	0.35	9,779.8	2,761.9	0.35	9,779.8	2,761.9	0.35
Envelopes	749.9	300.0	0.40	668.6	303.2	0.45	1,656.5	684.9	0.41	945.3	424.1	0.45	1,096.1	467.0	0.43	513.8	219.7	0.43	1,200.1	545.4	0.43	1,200.1	545.4	0.43
Fasteners	87.7	15.0	0.17	159.3	53.2	0.33	149.6	34.2	0.23	258.4	55.4	0.21	109.3	26.8	0.25	115.5	37.2	0.32	181.5	66.8	0.32	181.5	66.8	0.32
Furnishings	3,979.6	404.1	0.10	2,315.7	235.0	0.10	5,159.0	901.2	0.17	7,538.4	1,200.0	0.16	6,893.4	1,147.1	0.17	5,923.5	359.0	0.06	7,609.1	785.1	0.06	7,609.1	785.1	0.06
Labels	206.7	88.9	0.43	299.9	134.8	0.45	940.1	437.7	0.47	430.0	189.3	0.44	863.1	368.3	0.43	1,207.0	547.1	0.45	1,692.0	809.4	0.45	1,692.0	809.4	0.45
Machines	7,214.6	69.9	0.01	8,989.9	2,839.0	0.32	35,051.9	975.4	0.03	18,189.8	96.0	0.01	11,267.7	1,626.1	0.14	12,182.5	2,477.0	0.20	4,064.8	37.5	0.20	4,064.8	37.5	0.20
Paper	2,263.6	997.3	0.44	2,812.7	1,223.7	0.44	6,285.5	2,743.4	0.44	3,964.1	1,678.6	0.42	6,213.2	2,721.3	0.44	6,722.2	2,823.2	0.42	4,180.2	1,775.5	0.42	4,180.2	1,775.5	0.42
Phones	13,469.3	1,927.2	0.14	8,984.2	1,006.2	0.11	28,443.3	3,043.9	0.11	17,608.9	2,164.5	0.12	24,361.5	3,261.2	0.13	26,313.7	4,204.5	0.16	23,883.0	2,517.2	0.16	23,883.0	2,517.2	0.16
Storage	9,217.8	1,029.1	0.11	6,124.7	478.1	0.08	14,792.8	1,753.6	0.12	15,805.7	1,303.0	0.08	14,670.0	1,392.2	0.09	18,605.7	1,709.9	0.09	12,490.8	990.5	0.09	12,490.8	990.5	0.09
Supplies	4,403.4	130.2	0.03	289.4	18.0	0.06	10,636.7	-843.6	-0.08	6,215.9	-4.9	0.00	1,153.6	129.9	0.11	1,266.6	72.7	0.06	8,815.9	426.1	0.06	8,815.9	426.1	0.06
Tables	10,952.1	-2,698.7	-0.25	4,217.7	-555.4	-0.13	16,913.0	-1,199.7	-0.07	9,913.2	-1,437.7	-0.15	9,288.3	-375.5	-0.04	16,405.0	-1,092.1	-0.07	10,162.6	-699.3	-0.07	10,162.6	-699.3	-0.07
Total	94,924.8	9,134.4	0.10	59,751.3	10,294.6	0.17	205,005.5	28,594.7	0.14	137,762.1	11,587.4	0.08	155,028.8	22,411.3	0.14	152,718.7	21,285.8	0.14	147,238.1	13,832.7	0.14	147,238.1	13,832.7	0.14

Hình 3. 30 Detail Profit Sub-Category

3.6.3 Forecats

Biểu đồ dự báo cho thấy doanh thu trong các quý tiếp theo có xu hướng giảm xuống do các yếu tố bên ngoài nào đó. Từ đó cần phải đưa ra những biện pháp cụ thể để làm giảm đáng kể doanh thu. Xây dựng các chiến lược bán hàng với những mặt hàng bán chạy với lợi nhuận cao từ đó cải thiện được doanh thu



Hình 3. 31 Forecats

II. KẾT LUẬN

Kết luận phân tích dữ liệu trong lĩnh vực Bán lẻ bao gồm việc hiểu rõ về khách hàng thông qua phân khúc và dự đoán hành vi mua sắm, từ đó tạo ra các chiến lược tiếp thị và khuyến mãi hướng đến từng phân khúc cụ thể. Việc tối ưu hóa hàng tồn kho và cải thiện chuỗi cung ứng giúp quản lý hàng tồn kho hiệu quả, giảm thiểu tình trạng thiếu hàng hoặc dư thừa. Ngoài ra, dữ liệu giúp cá nhân hóa trải nghiệm khách hàng và cải thiện dịch vụ, đồng thời xác định các sản phẩm bán chạy và đánh giá hiệu quả chương trình khuyến mãi, từ đó điều chỉnh chiến lược kinh doanh. Cuối cùng, ra quyết định dựa trên dữ liệu về chiến lược giá cả và mở rộng thị trường giúp các nhà bán lẻ duy trì và phát triển trong môi trường cạnh tranh.

Thông qua bài báo cáo này, nhóm đã phần nào có thêm cho bản thân những kiến thức quý giá, giúp ích rất nhiều cho quá trình học tập cũng như những định hướng làm việc trong tương lai. Việc phân tích dữ liệu trong lĩnh vực bán lẻ không chỉ cung cấp cái nhìn sâu sắc về hành vi và sở thích của khách hàng mà còn giúp chúng tôi hiểu rõ hơn về cách tối ưu hóa quản lý hàng tồn kho, nâng cao hiệu quả chuỗi cung ứng và cải thiện trải nghiệm khách hàng. Những hiểu biết này không chỉ có giá trị trong học tập mà còn mở ra nhiều cơ hội nghề nghiệp, cho phép chúng tôi áp dụng các kỹ năng phân tích dữ liệu vào thực tế công việc. Hơn nữa, việc nắm vững các chiến lược dựa trên dữ liệu cũng giúp chúng tôi tự tin hơn trong việc ra quyết định kinh doanh, từ việc xác định chiến lược giá cả cạnh tranh đến việc tìm kiếm và mở rộng các cơ hội thị trường mới. Nhìn chung, những kiến thức thu được từ bài báo cáo này sẽ là nền tảng vững chắc giúp chúng tôi phát triển sự nghiệp trong lĩnh vực bán lẻ cũng như các lĩnh vực liên quan khác trong tương lai.