

ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC BÁCH KHOA
KHOA KHOA HỌC & KỸ THUẬT MÁY TÍNH



DATABASE SYSTEMS (EXTRA) (CO201B)

Đề tài:

Business Intelligence System for sales

GVHD: Phan Trọng Nhân
Nhóm: 1
SV thực hiện: Cù Thanh Bằng – 2012682
Nguyễn Đức An – 2010102
Nguyễn Hoàng Bảo Hùng – 2011312



Bảng phân công nhiệm vụ

STT	Họ và tên	MSSV	Công việc	Tỉ lệ hoàn thành
1	Cù Thanh Bằng	2012682	Tìm hiểu về MySQL, chỉ ra lợi ích trong việc sử dụng BI để đưa ra quyết định	100%
2	Nguyễn Đức An	2010102	Tìm hiểu về Power BI, vẽ các trang cho dashboard, thực hiện chức năng dự đoán	100%
3	Nguyễn Hoàng Bảo Hùng	2011312	Tìm hiểu về hệ thống BI, thực hiện chức năng realtime, viết hướng dẫn sử dụng	100%

Mục lục

1	Giới thiệu đề tài	4
2	Các lược đồ cơ sở dữ liệu liên quan đến đề tài Sales	4
3	Business Intelligence	4
3.1	Business Intelligence là gì ?	4
3.2	Các thành phần chính của hệ thống BI	5
3.3	Lợi ích của hệ thống BI	6
3.4	Đối tượng sử dụng của BI	7
3.5	Các hoạt động chính của BI	7
3.5.1	Hỗ trợ quyết định (Decision support)	7
3.5.2	Truy vấn và báo cáo (Query and reporting)	7
3.5.3	Phân tích thống kê (Statistical analysis)	8
3.5.4	Dự đoán (Forecasting)	8
3.6	Một số tool cho BI	8
3.7	Giới thiệu về Power BI	9
3.7.1	Power BI là gì ?	9
3.7.2	Các loại công cụ của Power BI	9
3.7.3	Một số tính năng của Power BI	9
3.7.4	Ưu điểm và nhược điểm của Power BI	10
4	Design System	11
4.1	Dashboard tổng quan	11
4.1.1	Thống kê tổng doanh thu, tổng chi, tổng số sản phẩm	11
4.1.2	Thống kê số lượng các thông tin liên quan khác	12
4.1.3	Phân bố tổng quan doanh thu chuỗi cửa hàng theo thời gian	12
4.1.4	Phân bố tổng quan doanh thu của từng cửa hàng	13
4.2	Dashboard so sánh doanh thu theo thời gian và cửa hàng	14
4.2.1	Thống kê về tổng doanh thu của các cửa hàng	14
4.2.2	Phân bố doanh thu của các cửa hàng theo thời gian	15
4.2.3	So sánh tổng quan doanh thu giữa các cửa hàng	15
4.2.4	So sánh tổng doanh thu, lợi nhuận giữa các loại cửa hàng	16
5	Develop the proposed system in Power BI	17
5.1	Dashboard sales tổng quan	17
5.2	Dashboard sales trong năm 1997	18
5.3	Dashboard sales trong năm 1998	19
6	Cập nhật tính năng real-time cho hệ thống	19
7	Hướng dẫn sử dụng các tool để sử dụng hệ thống BI	21
7.1	Hướng dẫn sử dụng MySQL	21
7.2	Hướng dẫn sử dụng Power BI	21
7.3	Hướng dẫn hiện thực hệ thống BI	22
8	Một số insights từ dữ liệu nhằm hỗ trợ quá trình ra quyết định	25
8.1	Từ dữ liệu theo thời gian	25
8.2	Từ dữ liệu theo cửa hàng	26
9	Hiện thực tính năng dự đoán (forecast) đối với tập dữ liệu	28



10 Tài liệu tham khảo

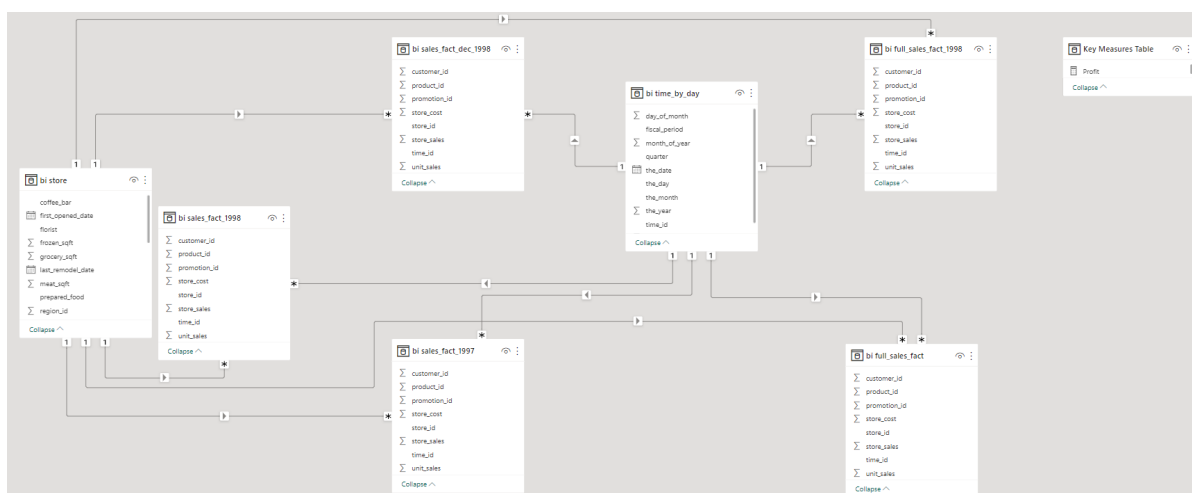
30

1 Giới thiệu đề tài

Trong đề tài này, nhóm sinh viên tập trung phân tích tổng quan các thông tin về dữ liệu sales của chuỗi cửa hàng, cùng với khai thác một số thông tin chuyên sâu về các cửa hàng theo thời gian từ tập dữ liệu mẫu.

2 Các lược đồ cơ sở dữ liệu liên quan đến đề tài Sales

- Dựa trên tập dữ liệu mẫu DB.sql, nhóm chọn 5 bảng chính bao gồm: sales_fact_1997, sales_fact_1998, sales_fact_dec_1998, store, time_by_day.
- Tiến hành thêm các bảng đó vào PowerBI và merge thành bảng full_sales_fact_1998 (là hợp giữa bảng sales_fact_1998 và sales_fact_dec_1998) và combine thành bảng full_sales_fact (là hợp giữa các bảng sales_fact_1997 và full_sales_fact_1998).



Hình 1: Data Model sau khi load dữ liệu vào Power BI

3 Business Intelligence

3.1 Business Intelligence là gì ?

- BI (Business Intelligence) có thể hiểu là hệ thống chuyển đổi các dữ liệu thô thành các dữ liệu có nghĩa và trực quan hóa chúng để phục vụ cho mục đích phân tích trong kinh doanh.
- Mục tiêu cuối cùng của BI là hỗ trợ đưa ra các quyết định tốt hơn nhằm tăng doanh thu, cải thiện năng suất, hiệu quả hoạt động và có được lợi thế cạnh tranh so với các đối thủ. Để đạt được điều đó, BI áp dụng các công cụ phân tích, quản lý và trực quan hóa dữ liệu, các quy trình công nghệ cũng như các phương pháp khác để quản lý và phân tích dữ liệu.
- Khái niệm BI mở rộng ra không chỉ áp dụng cho lĩnh vực kinh doanh mà còn cho nhiều lĩnh vực khác như giáo dục, y tế, ...

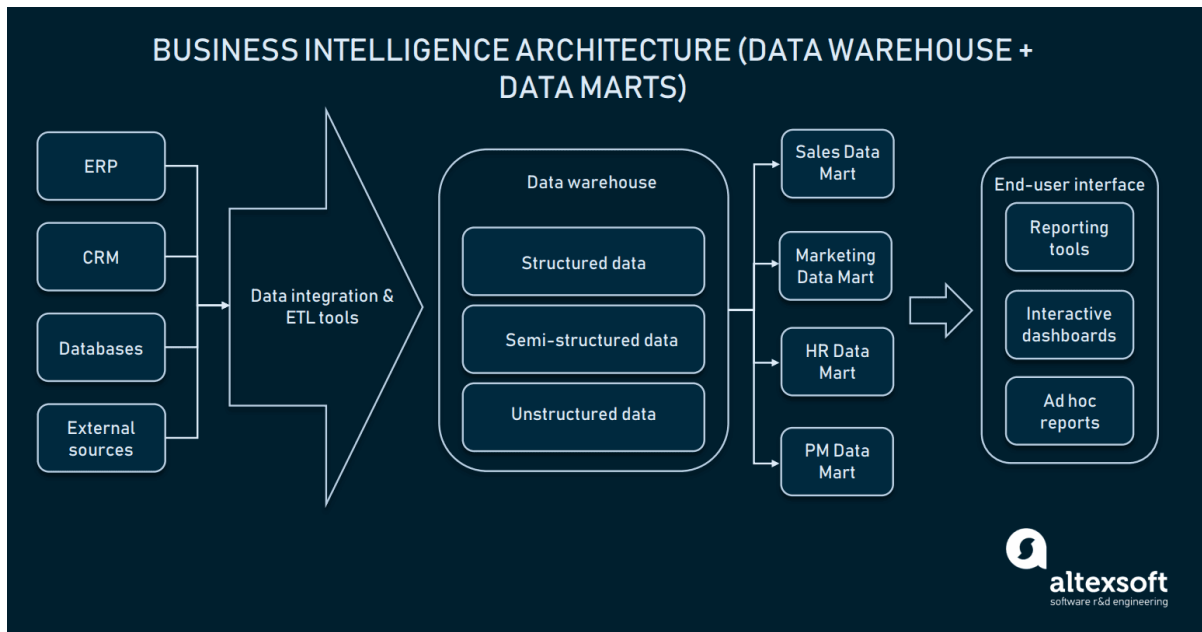


Hình 2: Giới thiệu về hệ thống BI

3.2 Các thành phần chính của hệ thống BI

Hệ thống BI gồm có 7 thành phần chính:

- **Data sources:** Dữ liệu thô được thu thập từ nhiều nguồn.
- **Data warehouse:** Kho dữ liệu để lưu trữ dữ liệu lâu dài.
- **Integrating Server (Máy chủ tích hợp):** Giúp vận hành quy trình ETL (Extract - Transform - Load) để chuyển đổi dữ liệu từ Data sources sang Data warehouse.
- **Analysis Server (Máy chủ phân tích):** Thu nhận dữ liệu và trả về kết quả dựa trên tri thức nghiệp vụ đã định nghĩa sẵn.
- **Reporting Server (Máy chủ báo cáo):** Bộ phận báo cáo các kết quả nhận được từ Analysis Server.
- **Data Mining (Khai thác dữ liệu):** Đây là quá trình dữ liệu đã qua xử lý được đem đi trích xuất, là một trong những khâu khá quan trọng. Bao gồm phân loại (classification), phân cụm (Clustering), hoặc dự đoán (Prediction).
- **Data Presentation (Trình bày dữ liệu):** Cuối cùng, dữ liệu sẽ được tổng hợp thành các biểu đồ từ quá trình data mining được tạo ra từ đây.



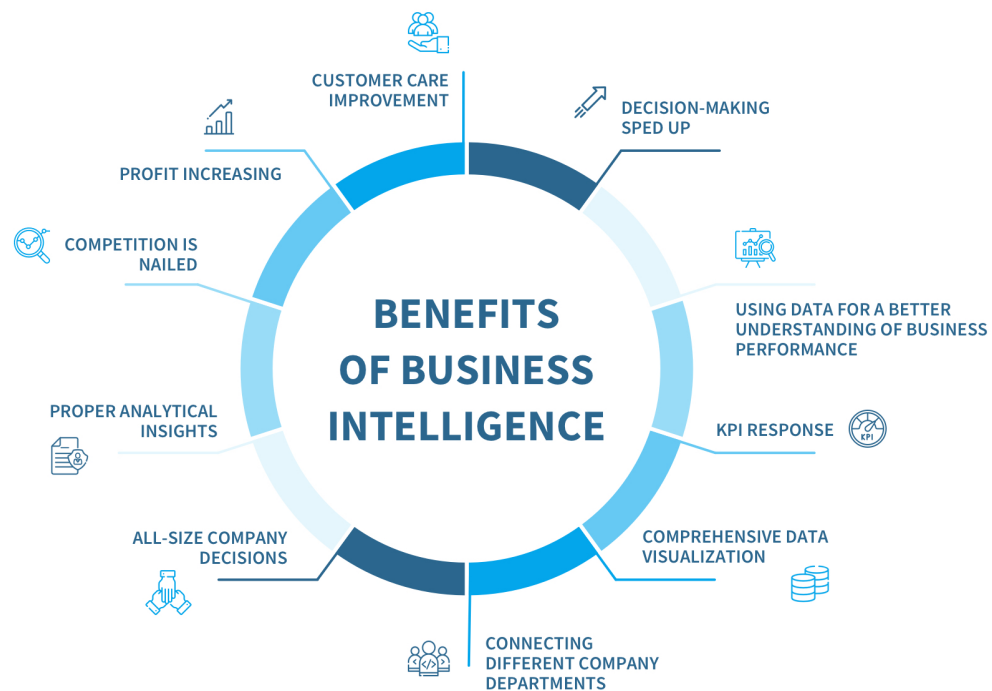
Hình 3: Quy trình làm việc với hệ thống BI

3.3 Lợi ích của hệ thống BI

Hệ thống BI giúp các doanh nghiệp đưa ra các quyết định tốt hơn khi tổng quan hóa và phân tích dữ liệu một cách trực quan và chính xác. Từ đó, các nhà phân tích có thể tận dụng BI để dự đoán về xu hướng của giá cả dịch vụ, hành vi khách hàng, phát hiện khách hàng tiềm năng để đề ra các chiến lược kinh doanh phù hợp nhằm gia tăng khả năng cạnh tranh của doanh nghiệp.

Một số cách mà BI có thể giúp doanh nghiệp đưa ra quyết định tốt hơn như:

- Giúp doanh nghiệp có cái nhìn tổng thể và chính xác về kết quả cũng như những vấn đề mà đang và có thể gặp phải.
- Dự đoán được tình hình tương lai của doanh nghiệp.
- Xác định được vị trí và khả năng của doanh nghiệp trong bức tranh tổng quan của thị trường.
- Giúp doanh nghiệp đưa ra các quyết định và xây dựng các chiến lược kinh doanh phù hợp.
- Tiết kiệm thời gian và chi phí cho công tác quản lý doanh nghiệp.



Hình 4: Một số lợi ích của hệ thống BI

3.4 Đối tượng sử dụng của BI

- Nhà quản trị
- Nhà phân tích
- Người đưa ra các quyết định
- Khách hàng

3.5 Các hoạt động chính của BI

3.5.1 Hỗ trợ quyết định (Decision support)

Các dữ liệu phân tích trong BI sẽ được sử dụng để định hướng và giúp các doanh nghiệp ra những quyết định trong kinh doanh.

3.5.2 Truy vấn và báo cáo (Query and reporting)

Truy vấn được sử dụng trong phân tích dữ liệu, giúp khám phá thêm các thông tin quan trọng trong dữ liệu đó. Báo cáo là hành động trực quan hóa dữ liệu dưới dạng các mô hình, biểu đồ, đường biểu diễn trên một trang tổng quan để tổng kết lại các thông tin thu được.

3.5.3 Phân tích thống kê (Statistical analysis)

Phân tích thống kê là hoạt động giải thích dữ liệu nhằm phát hiện ra các mẫu và các xu hướng. Nó là một khâu quan trọng trong phân tích dữ liệu.

3.5.4 Dự đoán (Forecasting)

Dự đoán tính hình, những vấn đề có thể xảy ra trong tương lai cũng là một hoạt động của BI. Nó giúp doanh nghiệp có thể chuẩn bị tốt các phương án đối phó cũng như định hình kế hoạch dài hạn trong tương lai.

3.6 Một số tool cho BI

- Looker Studio



Hình 5: Looker Studio

Looker Studio, trước đây là Google Data Studio, là một công cụ trực tuyến để chuyển đổi dữ liệu thành các trang tổng quan và báo cáo thông tin có thể tùy chỉnh được Google giới thiệu vào ngày 15 tháng 3 năm 2016 như một phần của bộ Google Analytics 360 dành cho doanh nghiệp

- Metabase



Hình 6: Metabase

Metabase là một nền tảng BI mã nguồn mở, có thể sử dụng để đặt câu hỏi về dữ liệu của mình hoặc nhúng Metabase vào ứng dụng để cho phép khách hàng tự khám phá dữ liệu của họ.

- PowerBI



Hình 7: PowerBI

Microsoft Power BI là một sản phẩm phần mềm trực quan hóa dữ liệu tương tác được phát triển bởi Microsoft với trọng tâm chính là kinh doanh thông minh. Power BI là một phần của Microsoft Power Platform.

- Tableau (Community edition)



Hình 8: Tableau

Tableau là một nền tảng phân tích trực quan sẽ biến đổi cách thức chúng ta sử dụng dữ liệu để giải quyết vấn đề, từ đó giúp mọi người và các tổ chức khai thác tối đa dữ liệu của họ.

3.7 Giới thiệu về Power BI

3.7.1 Power BI là gì ?

- Microsoft Power BI là một loại công cụ, phần mềm, ứng dụng hỗ trợ việc kết nối và trực quan hóa dữ liệu (visualize data) và khám phá dữ liệu (explore insights).
- Power BI có thể làm việc với các nguồn dữ liệu đơn giản như Excel, JSON, SQL,...đến các dữ liệu phức tạp trên nền tảng cloud hoặc Data Warehouse.
- Power BI dễ dàng kết nối đến các nguồn dữ liệu, trực quan hóa, khám phá dữ liệu và chia sẻ và xuất bản thông tin cho mọi người thông qua báo cáo hoặc dashboard trên Power BI Service.

3.7.2 Các loại công cụ của Power BI

- **Power BI Desktop:** đây là phiên bản Power BI có thể cài đặt về máy ở dạng Desktop App, có vai trò xử lý, tập hợp và xây dựng mô hình dữ liệu dùng để trực quan hóa dữ liệu cho các báo cáo. Phần mềm có thể cài đặt trên hệ điều hành Windows.
- **Power BI Mobile:** đây là phiên bản Power BI ứng dụng có thể sử dụng trên các tảng như Android hay iOS.
- **Power BI Service (Power BI Online):** đây là một phiên bản Power BI cho phép người dùng lưu trữ báo cáo, dashboard trực quan trực tiếp trên nền tảng web.

3.7.3 Một số tính năng của Power BI

- Power BI có chức năng vượt trội trong việc tạo báo cáo hoặc các dashboard doanh nghiệp. Đây sẽ là một công cụ hữu hiệu cho người dùng khi có nhu cầu xử lý nguồn dữ liệu lớn và đòi hỏi những thông tin chuyên môn chuyên sâu.



Hình 9: Giới thiệu về Power BI

– Power BI có một số tính năng nổi bật như:

- Cho phép người dùng được truy cập dữ liệu từ nhiều nguồn và tự động hóa quy trình xử lý dữ liệu.
- Hỗ trợ kết nối, chuyển đổi phân tích dữ liệu kích thước lớn: Khả năng xử lý dữ liệu của Power BI vào hàng cực khủng khi có thể làm việc với 8-10 triệu dòng dữ liệu một lần. ‘
- Nâng cao tính năng trực quan hóa dữ liệu
- Dùng Biểu thức phân tích (DAX) để phân tích dữ liệu: DAX là biểu thức phân tích vô cùng mạnh mẽ với tốc độ xử lý nhanh, hiệu quả.
- Xây dựng các mô hình dữ liệu để kết hợp dữ liệu từ nhiều nguồn.

3.7.4 Ưu điểm và nhược điểm của Power BI

• Ưu điểm

- Cung cấp các mẫu báo cáo và dashboard có sẵn.
- Các bảng dashboard luôn được cập nhật real-time, người dùng có thể thao tác trên nhiều thiết bị khác nhau.
- Power Query giúp người dùng dễ dàng hơn trong việc sử dụng và làm gọn dữ liệu.
- Hỗ trợ truy vấn ngôn ngữ tự nhiên khi khai thác dữ liệu.
- Tích hợp code Python và R để hỗ trợ tốt tính năng trực quan hóa, phân tích dữ liệu.
- Được xây dựng từ nền tảng AI (trí tuệ nhân tạo) và Machine Learning (máy học) nên nhạy bén trong việc xử lý dữ liệu.

- Bảo mật cao trong việc kết nối các nguồn dữ liệu (qua đám mây hay hệ thống doanh nghiệp).

- **Nhược điểm**

- Chỉ chấp nhận dữ liệu file nhỏ hơn 250MB hoặc file không nén.
- Các tính năng còn khá hạn chế khi sử dụng trên điện thoại di động.
- Chỉ có thể an toàn bảo mật nếu người dùng chia sẻ dashboard và báo cáo trên cùng một email tên miền.
- Nguồn dữ liệu của dashboard còn khá hạn chế.

4 Design System

4.1 Dashboard tổng quan

Ở trang dashboard tổng quan về sales, tụi em sẽ phân tích một số thông tin sau:

STT	Tiêu chuẩn phân tích	Mô tả
1	Thống kê về tổng doanh thu, tổng chi, tổng số sản phẩm trong cả hai năm 1997 và 1998.	Thống kê tổng doanh thu, tổng chi, tổng số sản phẩm trong cả hai năm 1997 và 1998.
2	Thống kê số lượng các thông tin liên quan khác trong cả hai năm 1997 và 1998.	Thống kê tổng số cửa hàng, tổng số sản phẩm, tổng số chiến dịch quảng cáo, tổng số khách hàng trong cả hai năm 1997 và 1998.
3	Phân bố tổng quan doanh thu chuỗi cửa hàng theo thời gian từ năm 1997 đến năm 1998.	Phân bố doanh thu của chuỗi cửa hàng qua từng ngày, tháng, quý hoặc năm.
4	Phân bố tổng quan doanh thu của từng cửa hàng trong cả hai năm 1997 và 1998.	So sánh trực quan doanh thu của các cửa hàng theo từng ngày, tháng, quý hoặc năm.

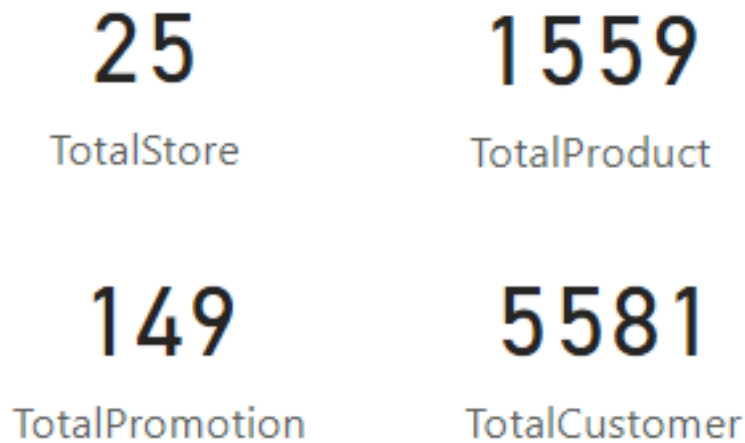
4.1.1 Thống kê tổng doanh thu, tổng chi, tổng số sản phẩm

1.76M **706.295K** **833.489K**
TotalSales TotalCosts TotalUnit

• **Nhận xét:**

- Trong cả hai năm 1997 và 1998, chuỗi cửa hàng đạt được tổng doanh thu là khoảng 1.76 triệu USD.
- Trong khi đó, tổng chi của chuỗi cửa hàng chỉ có 706.295K, ở đây có nghĩa là doanh thu cửa hàng gấp khoảng 2.3 lần so với tổng chi. Điều này có thể cho thấy chuỗi cửa hàng kinh doanh rất có hiệu quả.
- Ngoài ra, tổng số sản phẩm bán ra là 833.489K sản phẩm.

4.1.2 Thống kê số lượng các thông tin liên quan khác

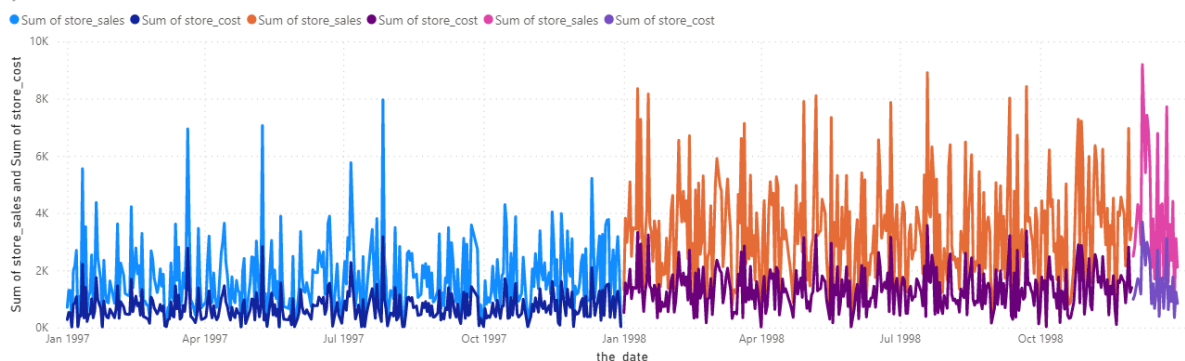


• **Nhận xét:**

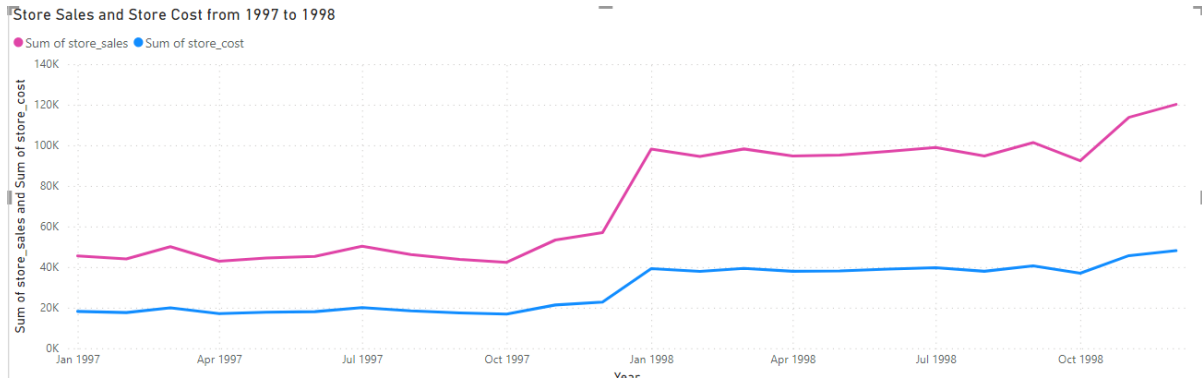
- Trong cả hai năm 1997 và 1998, chuỗi cửa hàng có 25 cửa hàng, 1559 sản phẩm, 149 chiến dịch quảng cáo và 5581 khách hàng.

4.1.3 Phân bố tổng quan doanh thu chuỗi cửa hàng theo thời gian

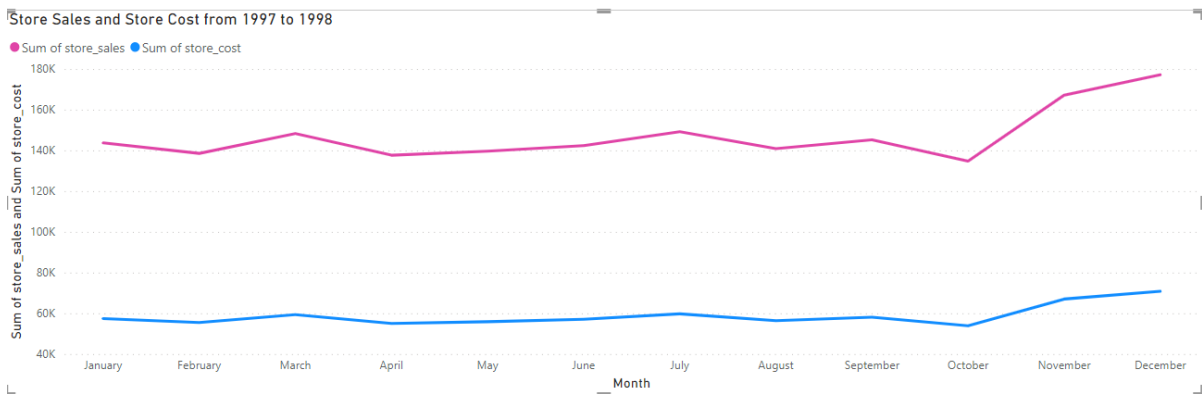
Store Sales and Store Cost from 1997 to 1998



Ngoài ra người dùng có thể tiến hành xem xu hướng theo quý hoặc tháng như sau:



Hình 10: Xu hướng doanh thu theo quý



Hình 11: Xu hướng doanh thu theo tháng

• Nhận xét:

- Nhóm sử dụng biểu đồ đường để mô tả xu hướng của sales.
- Nhìn chung store_sales lớn hơn store_cost nhiều trong hầu hết các giai đoạn.
- Có thể thấy store_sales tỉ lệ thuận với store_cost, nghĩa là cửa hàng đầu tư càng nhiều thì sẽ thu lợi nhuận càng cao, doanh thu dao động theo thời gian, doanh thu cao nhất là hơn 8K USD/ngày.

4.1.4 Phân bố tổng quan doanh thu của từng cửa hàng

• Nhận xét:

- Đây là bảng thống kê tổng chi, tổng doanh thu, lợi nhuận của cửa hàng trong suốt giai đoạn từ năm 1997-1998.

Store

store_id	TotalCostsByStore	TotalSalesByStore	ProfitByStore
2	1,896.62	4,739.23	2,842.61
3	21,121.96	52,896.30	31,774.34
6	18,266.44	45,750.24	27,483.80
7	21,771.54	54,545.28	32,773.74
11	46,073.62	115,334.62	69,261.00
Total	225,627.23	565,238.13	339,610.90

4.2 Dashboard so sánh doanh thu theo thời gian và cửa hàng

Ở trang dashboard tổng quan về sales, tụi em sẽ phân tích một số thông tin sau:

STT	Tiêu chuẩn phân tích	Mô tả
1	Thống kê về tổng doanh thu của các cửa hàng	Thống kê các số liệu về tổng chi, tổng doanh thu, lợi nhuận của các cửa hàng.
2	Phân bố doanh thu của các cửa hàng theo thời gian	Phân bố doanh thu của cửa hàng qua từng ngày, tháng, quý hoặc năm.
3	So sánh tổng quan doanh thu giữa các cửa hàng	So sánh tổng quan doanh thu giữa các cửa hàng với nhau.
4	So sánh tổng doanh thu, lợi nhuận giữa các loại cửa hàng	So sánh xem loại cửa hàng nào đem lại doanh thu nhiều nhất.

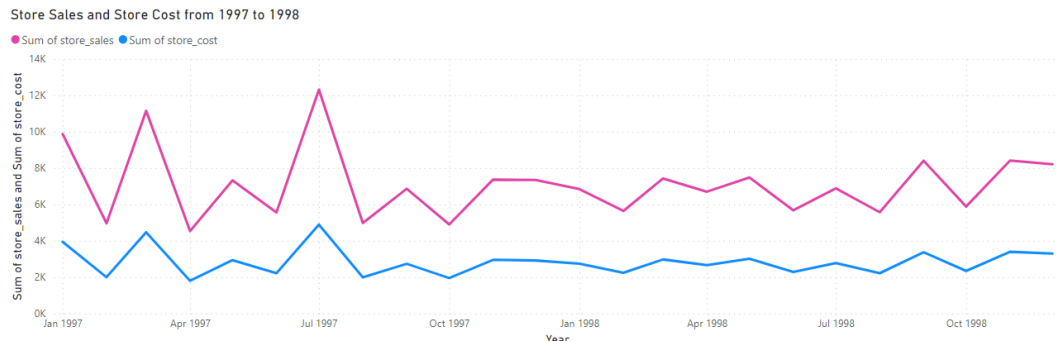
4.2.1 Thống kê về tổng doanh thu của các cửa hàng

store_id	TotalCostsByStore	TotalSalesByStore	ProfitByStore
13	68,160.59	170,398.94	102,238.35
17	63,127.09	157,695.90	94,568.81
15	46,327.38	115,981.46	69,654.08
11	46,073.62	115,334.62	69,261.00
Total	706,294.73	1,764,546.44	1,058,251.71

- **Nhận xét:**

- Đây là bảng sắp thứ tự giảm dần về lợi nhuận thu được của các cửa hàng theo thời gian.
- Có thể thấy store 13 đạt được lợi nhuận cao nhất.

4.2.2 Phân bố doanh thu của các cửa hàng theo thời gian



- **Nhận xét:**

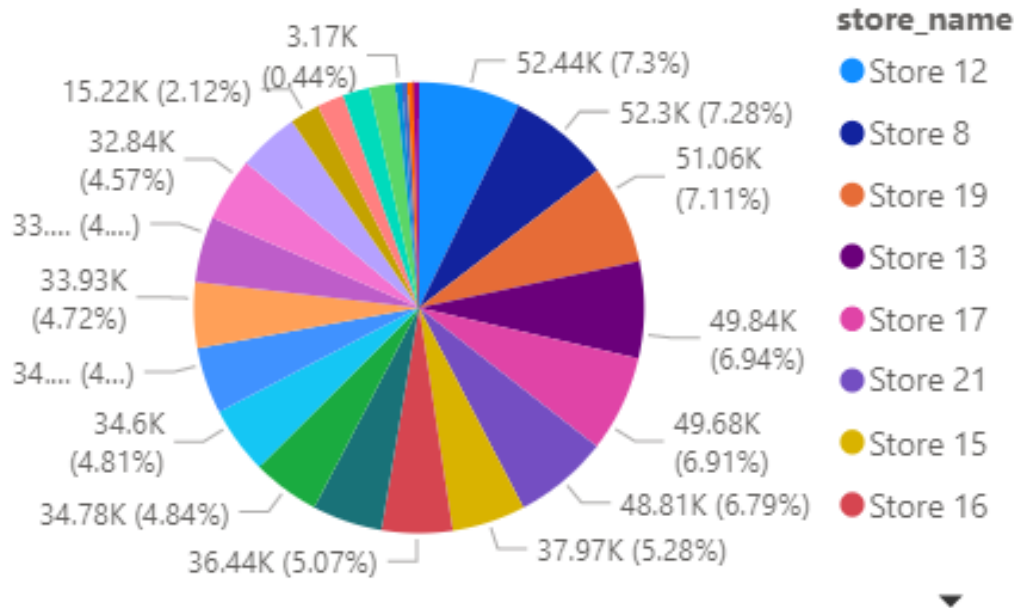
- Đây là biểu đồ đường mô tả doanh thu của cửa hàng thứ 13 theo thời gian.
- Như trên hình chúng ta có thể thấy doanh thu của cửa hàng 13 đạt đỉnh vào tháng 7 năm 1997.

4.2.3 So sánh tổng quan doanh thu giữa các cửa hàng

Sum of store_sales and Sum of store_cost by store_name



Profit1998 by store_name

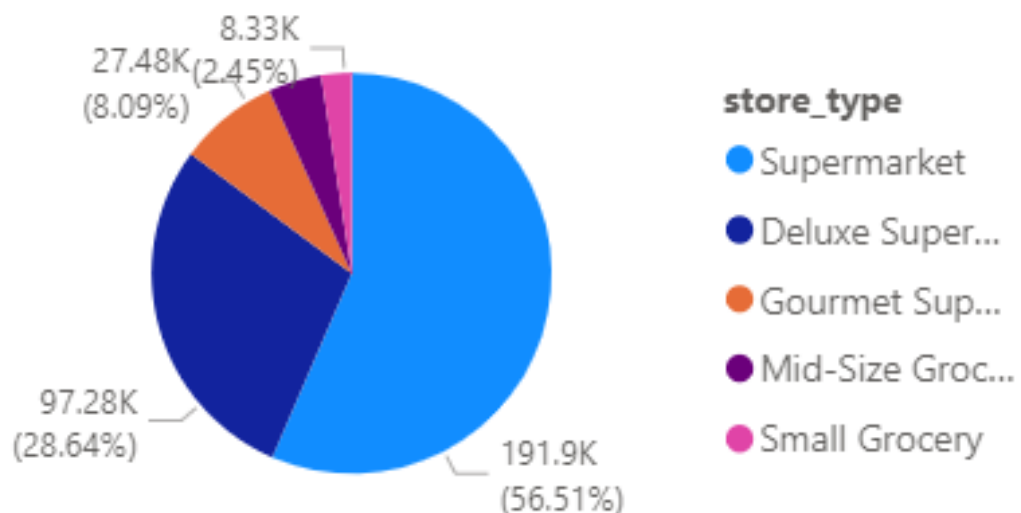


• Nhận xét:

- Nhóm sử dụng biểu đồ cột và biểu đồ tròn để so sánh doanh thu và lợi nhuận giữa các cửa hàng với nhau
- Như trên hình chúng ta có thể thấy overall thì cửa hàng 13 có doanh thu lớn nhất trong các trường hợp. Tuy nhiên, vào năm 1998, store 12 có lợi nhuận đạt được lớn nhất trong các store.

4.2.4 So sánh tổng doanh thu, lợi nhuận giữa các loại cửa hàng

Profit1997 by store_type

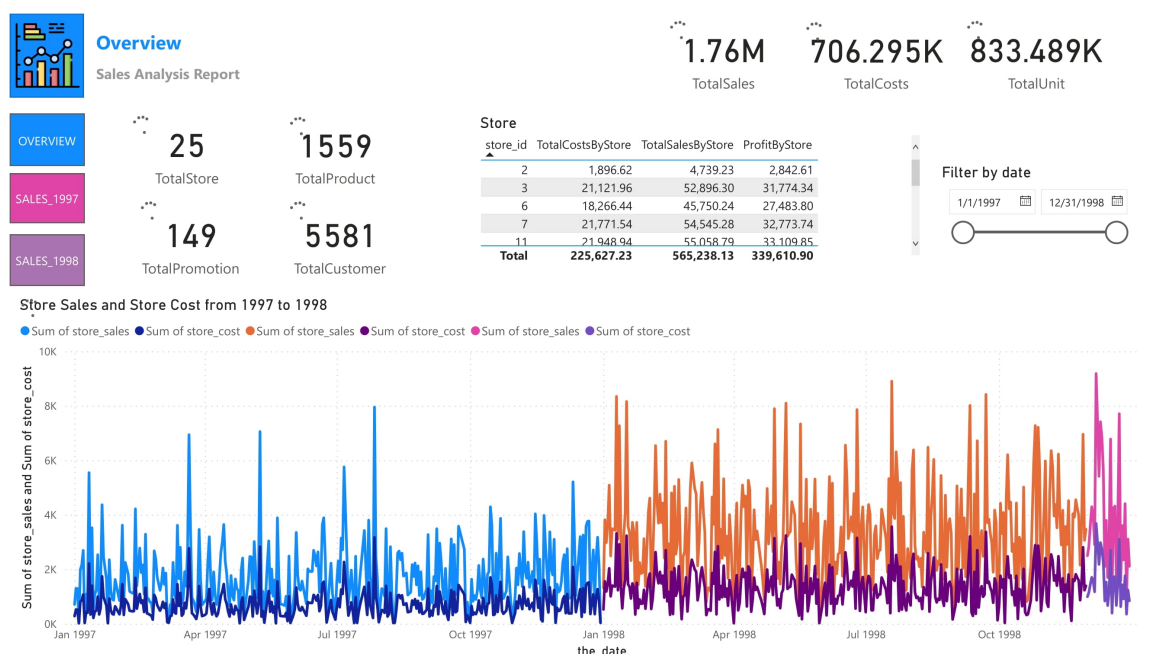


• **Nhận xét:**

- Đây là bảng so sánh lợi nhuận giữa các loại cửa hàng trong năm 1997.
- Như trên hình chúng ta có thể thấy lợi nhuận từ supermarket là lớn nhất trong các loại cửa hàng.

5 Develop the proposed system in Power BI

5.1 Dashboard sales tổng quan

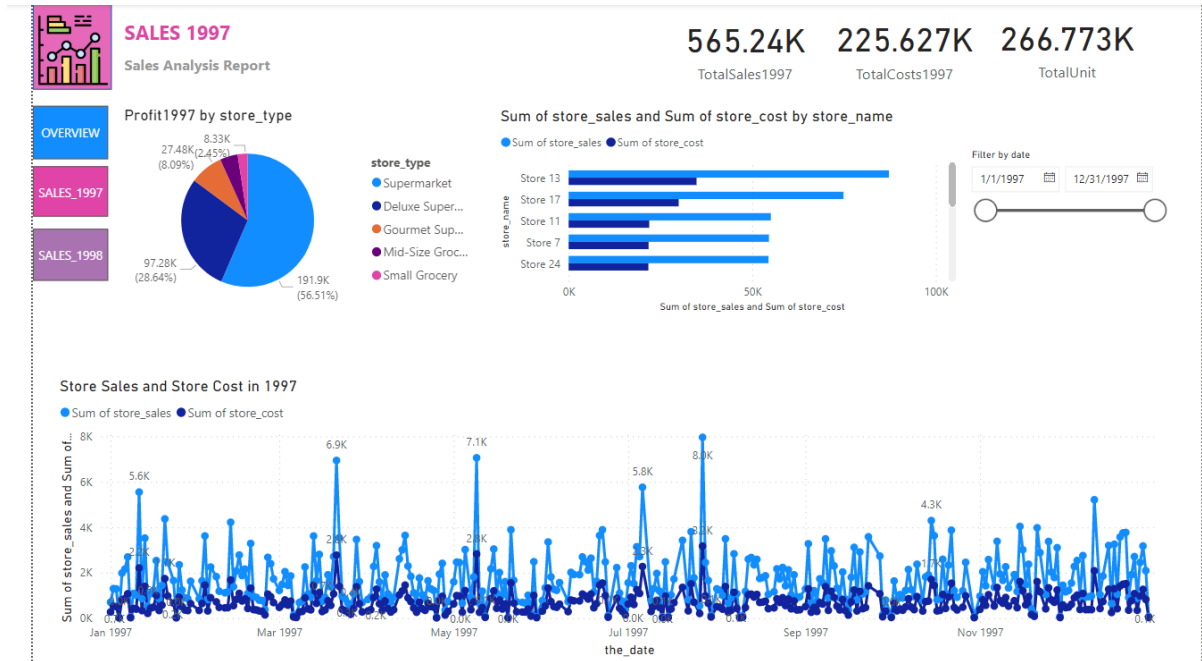


Hình 12: Dashboard Overview

Trong Dashboard thống kê chung, người dùng có thể thực hiện được một số chức năng sau:

- Xem được tổng quan các số liệu về tổng thu, tổng chi, tổng số sản phẩm bán ra của chuỗi cửa hàng.
- Người dùng có thể xem được sự biến đổi trong xu hướng của sales dựa vào ngày, tháng, quý hoặc năm,...
- Người dùng có thể so sánh doanh thu, tổng chi, lợi nhuận của các cửa hàng với nhau.
- Người dùng có thể filter theo time để xem các thông tin trong một khoảng thời gian nhất định.
- Người dùng có thể filter theo store để xem các thông tin về doanh thu của cửa hàng đó theo thời gian.

5.2 Dashboard sales trong năm 1997

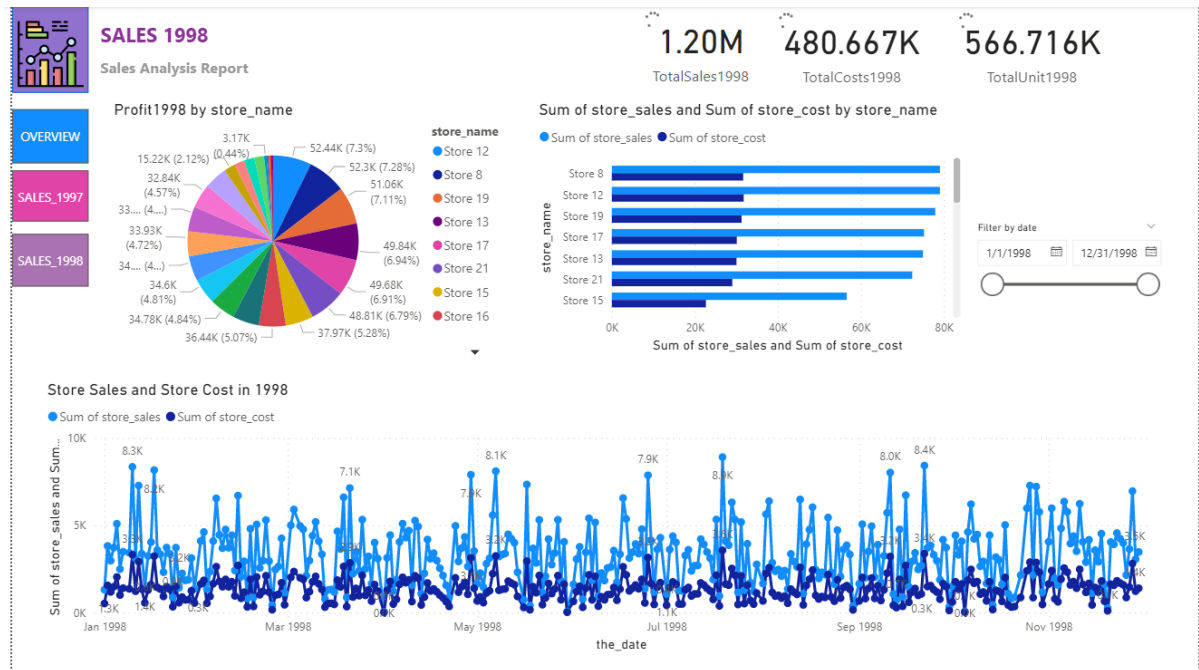


Hình 13: Dashboard Sales In 1997

Trong Dashboard năm 1997, người dùng có thể thực hiện được một số chức năng sau:

- Xem được tổng quan các số liệu về tổng thu, tổng chi, tổng số sản phẩm bán ra của chuỗi cửa hàng trong năm 1997.
- Người dùng có thể xem được sự biến đổi trong xu hướng của sales dựa vào ngày, tháng, quý của năm 1997
- Người dùng có thể so sánh lợi nhuận của các loại cửa hàng với nhau.
- Người dùng có thể filter theo time để xem các thông tin trong một khoảng thời gian nhất định trong năm 1997.
- Người dùng có thể so sánh và filter theo store để xem doanh thu và lợi nhuận của một số cửa hàng mong muốn.

5.3 Dashboard sales trong năm 1998



Hình 14: Dashboard Sales In 1998

Trong Dashboard năm 1998, người dùng có thể thực hiện được một số chức năng sau:

- Xem được tổng quan các số liệu về tổng thu, tổng chi, tổng số sản phẩm bán ra của chuỗi cửa hàng trong năm 1998.
- Người dùng có thể xem được sự biến đổi trong xu hướng của sales dựa vào ngày, tháng, quý của năm 1998
- Người dùng có thể so sánh lợi nhuận của các loại cửa hàng với nhau.
- Người dùng có thể filter theo time để xem các thông tin trong một khoảng thời gian nhất định trong năm 1998.
- Người dùng có thể so sánh và filter theo store để xem doanh thu và lợi nhuận của một số cửa hàng mong muốn.

6 Cập nhật tính năng real-time cho hệ thống

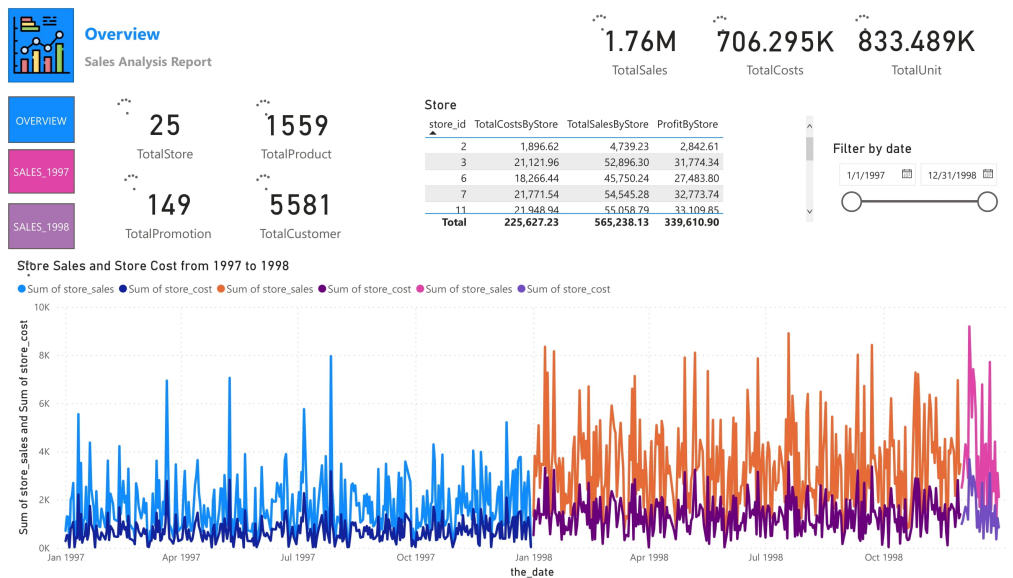
Qua tìm hiểu, nhóm quyết định sử dụng chức năng tự động làm mới trang trong PowerBI để hiển thị kết quả gần theo thời gian thực. Tuy nhiên chức năng này cần database được kết nối với PowerBI bằng chế độ DirectQuery và MySQL lại không hỗ trợ chế độ kết nối này. Do đó, nhóm sử dụng Connector MaridaDB để làm trung gian kết nối MySQL với PowerBI.

Chế độ DirectQuery cho phép mỗi lần làm mới trang thì PowerBI sẽ thực hiện truy xuất lại trực tiếp vào database để lấy dữ liệu mới nhất. Nhóm cài đặt mỗi 15 giây trang sẽ tự động được

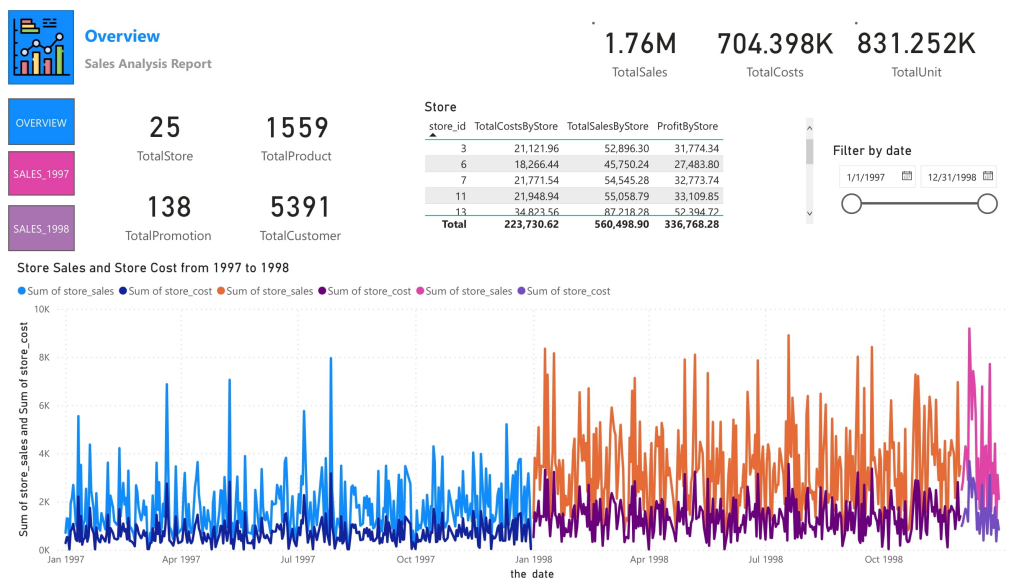
làm mới và cập nhật dữ liệu mới nhất từ database.

Video chạy thử được đính kèm hoặc xem ở link sau: https://drive.google.com/file/d/118JeIbfy0a7LiP58sinhb00Iovto7igG/view?usp=share_link

Ví dụ minh họa:



Hình 15: Overview ban đầu

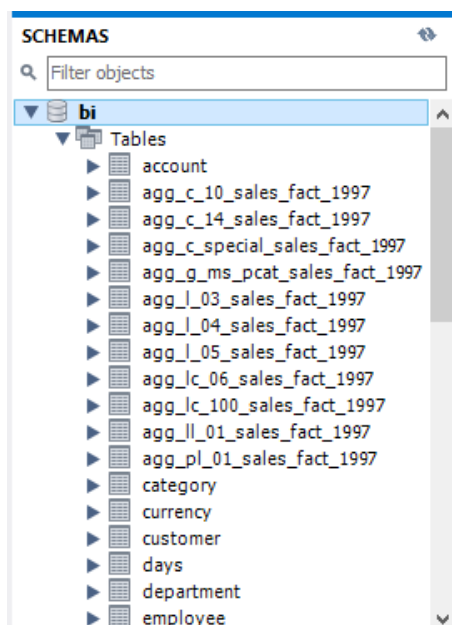


Hình 16: Sau khi xóa toàn bộ thông tin sale của cửa hàng có store_id là 2 trong database

7 Hướng dẫn sử dụng các tool để sử dụng hệ thống BI

7.1 Hướng dẫn sử dụng MySQL

Chúng ta sẽ mở file DB.sql và chạy trong MySQL WorkBench để sinh ra cơ sở dữ liệu phục vụ cho BI.

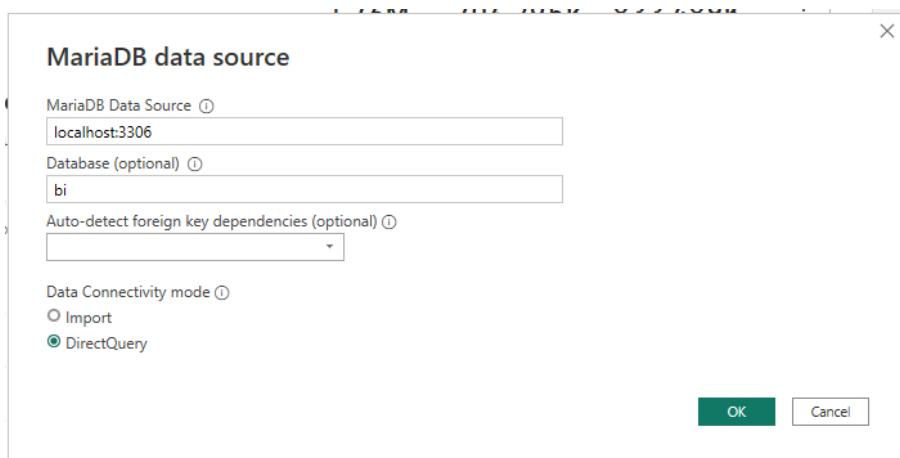


7.2 Hướng dẫn sử dụng Power BI

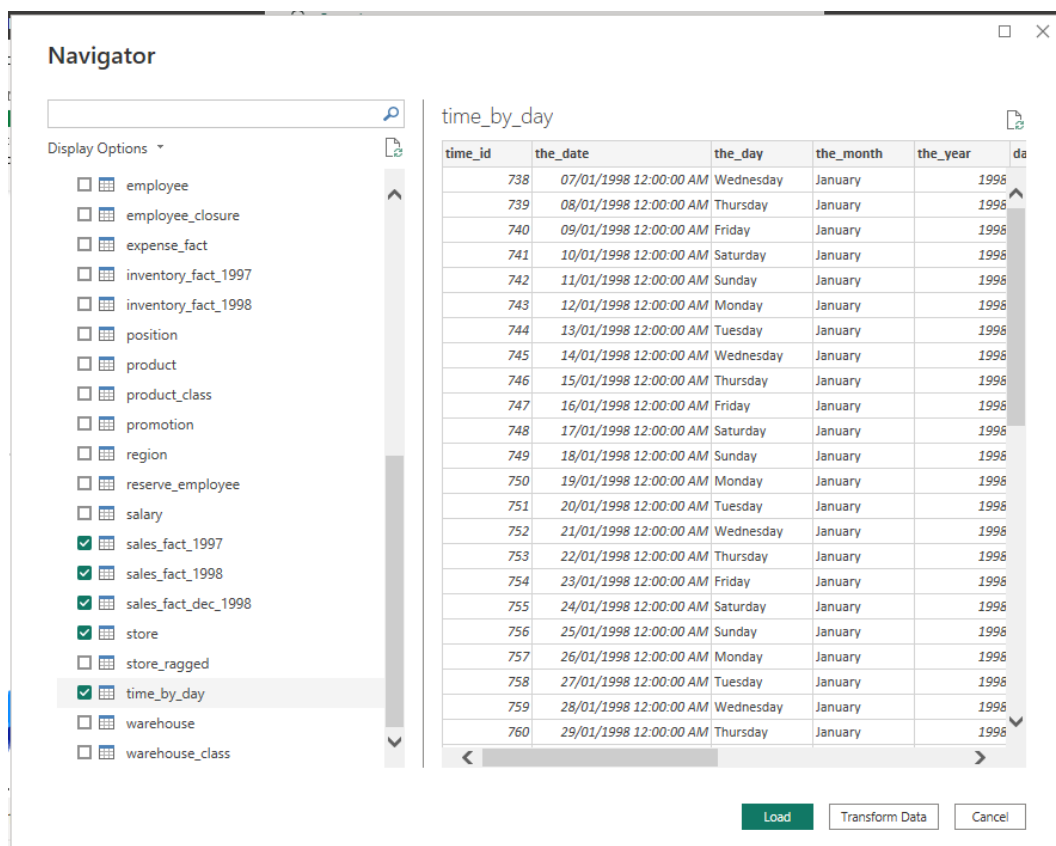
Trước hết, chúng ta cần cài đặt MariaDB theo hướng dẫn ở trang <https://mariadb.com/docs/server/connect/power-bi>

Sau đó sẽ kết nối cơ sở dữ liệu **bi** trong MySQL với PowerBI thông qua MariaDB.

Nhập Data Source và Database, chọn Data Connectivity Mode là DirectQuery và nhập Username, Password để kết nối.

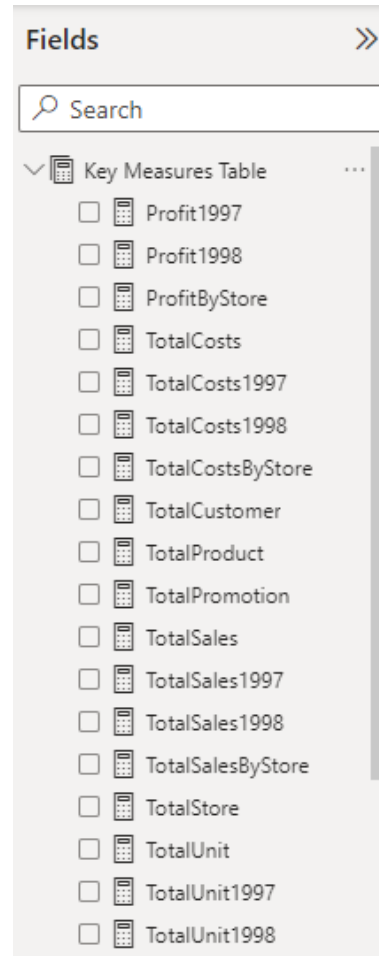


Sau khi kết nối thành công, chúng ta sẽ chọn các bảng cần thiết để tải vào PowerBI.



7.3 Hướng dẫn hiện thực hệ thống BI

Chúng ta tạo các Measure biến những số liệu ở các bảng vừa tải thành các thông tin có ích để hiện thị lên dashboard của PowerBI.



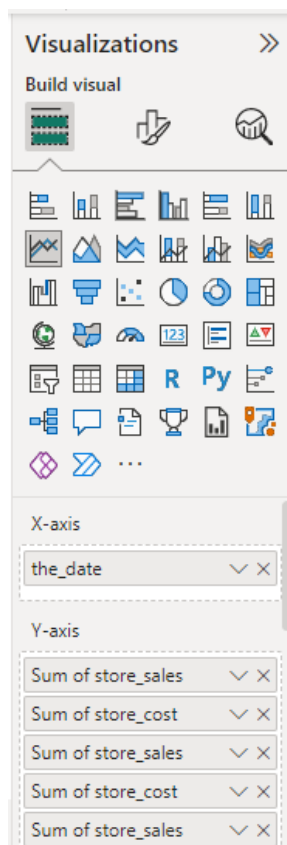
```
1 TotalSales1997 = SUMX('sales_fact_1997', 'sales_fact_1997'[store_sales])
```

Hình 17: Nội dung Measure TotalSale1997

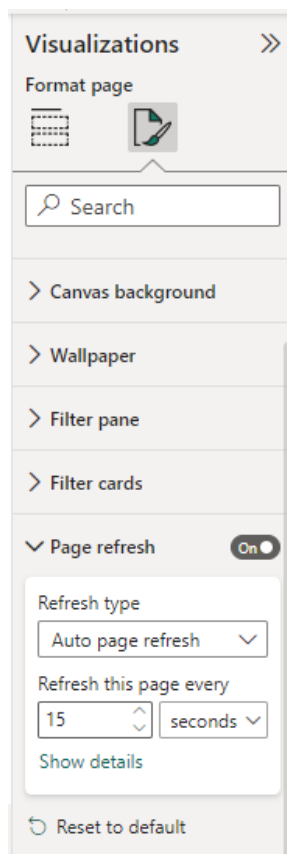
```
1 TotalSales1998 = SUMX('sales_fact_1998', 'sales_fact_1998'[store_sales]) + SUMX('sales_fact_dec_1998', 'sales_fact_dec_1998'[store_sales])
```

Hình 18: Nội dung Measure TotalSale1998

Chúng ta tiếp tục tạo ra các đối tượng hiện thị trên Dashboard như biểu đồ cột, đường biểu diễn,... và truyền các trường giá trị hay measure vào cho nó.

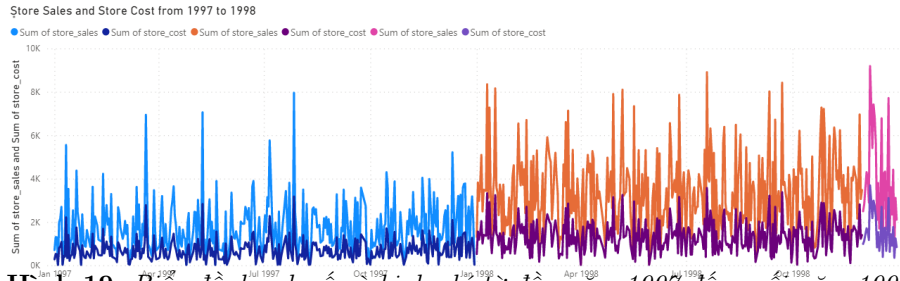


Cuối cùng, chúng ta sẽ cài đặt cho mỗi trang sẽ tự động được làm mới mỗi 15 giây.



8 Một số insights từ dữ liệu nhằm hỗ trợ quá trình ra quyết định

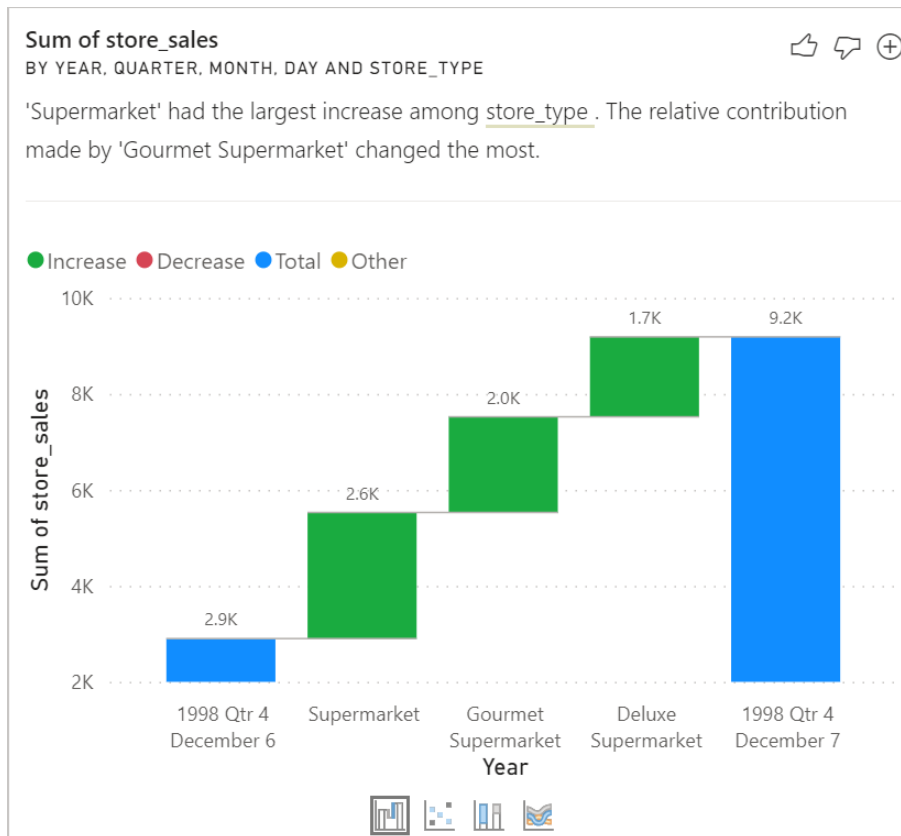
8.1 Từ dữ liệu theo thời gian



Hình 19: Biểu đồ doanh số và kinh phí từ đầu năm 1997 đến cuối năm 1998

Từ thông tin ở biểu đồ tổng quát theo thời gian, ta có thể đưa ra kết luận về xu hướng tăng của cả doanh số và kinh phí của các cửa hàng trên cả giai đoạn. Tuy nhiên, việc tăng doanh số đó không ổn định mà mang tính biến động theo từng giai đoạn thời gian (có dấu hiệu yếu đi rõ rệt ở giai đoạn khoảng từ tháng 8 năm 1997 đến tháng 12 năm 1997).

Hệ thống cung cấp cho người xem những cái nhìn chi tiết hơn ở những mốc thời gian cụ thể. Ví dụ như dưới đây là thống kê cụ thể vào ngày 7 tháng 12 (ngày có số liệu doanh số và kinh phí tăng trưởng lớn bậc nhất trong cả giai đoạn).



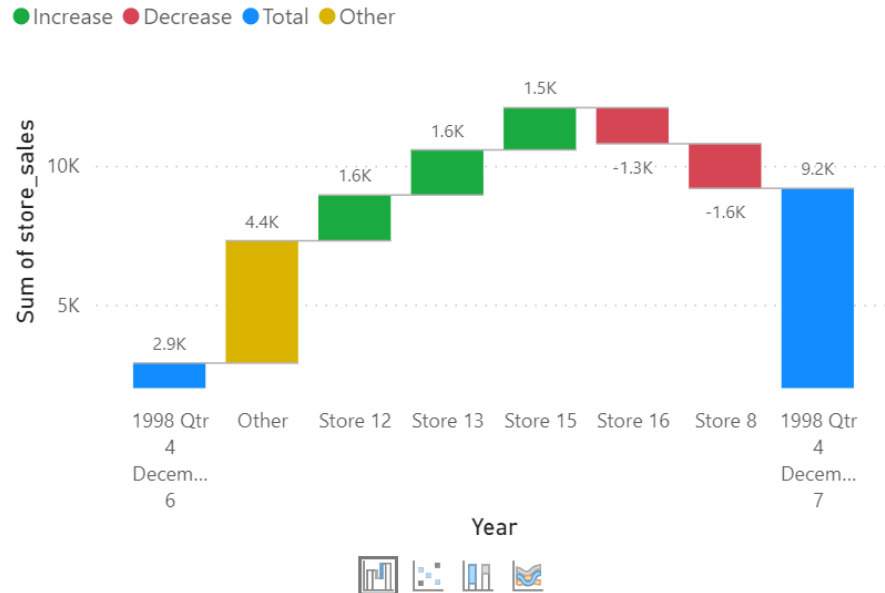
Hình 20: Phân tích số liệu kinh doanh ngày 7/12/1998 theo loại cửa hàng

Sum of store_sales

BY YEAR, QUARTER, MONTH, DAY AND STORE_NAME



'Store 12', 'Store 13', and 'Store 15' accounted for the majority of the increase among store_name, offsetting the decrease of 'Store 8'. The relative contributions made by 'Store 8', 'Store 16', and 'Store 12' changed the most.



Hình 21: Phân tích số liệu kinh doanh ngày 7/12/1998 theo tên cửa hàng

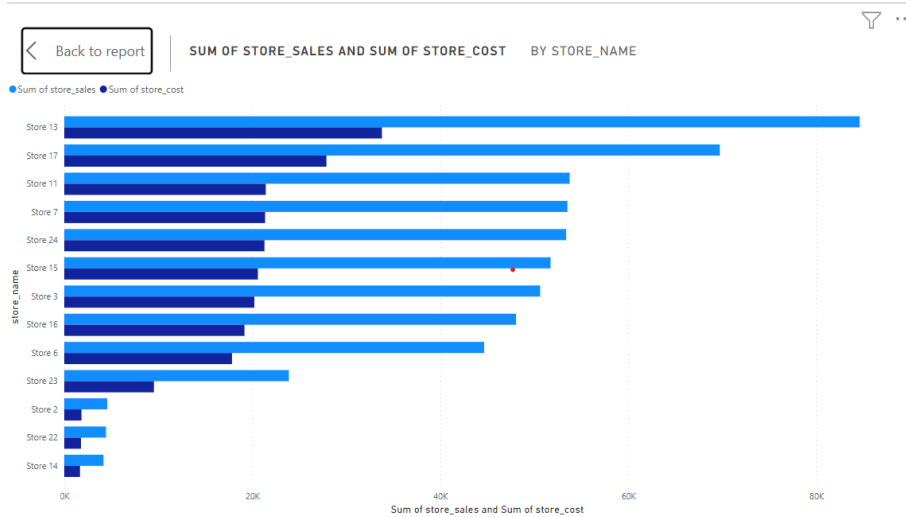
Qua những dữ liệu thống kê trên, ta có thể thấy:

- Thời điểm kinh phí và doanh số có sự tăng trưởng lớn nhất là khoảng từ tháng 10 năm 1998 đến tháng 12 năm 1998.
- Ở cả năm 1997 và 1998 đều thấy có sự giảm doanh số trong khoảng thời gian tháng 3 đến tháng 4 và tháng 7 đến tháng 8.
- Doanh số vào những ngày đạt đỉnh tập trung nhiều ở cửa hàng thuộc loại siêu thị (supermarket). Chứng tỏ việc kinh doanh ở siêu thị đạt hiệu quả cao và mang lại doanh thu lớn
- Doanh số ở cửa hàng số 12, 13 và 15 có mức tăng trưởng lớn, bên cạnh đó thì doanh số cửa hàng 16 và 8 lại có xu hướng giảm mạnh, ta có thể kết luận ở những ngày đạt đỉnh thì người dùng không có xu hướng đến mua sắm ở những cửa hàng có cùng loại/đặc điểm giống các cửa hàng 16 và 8.

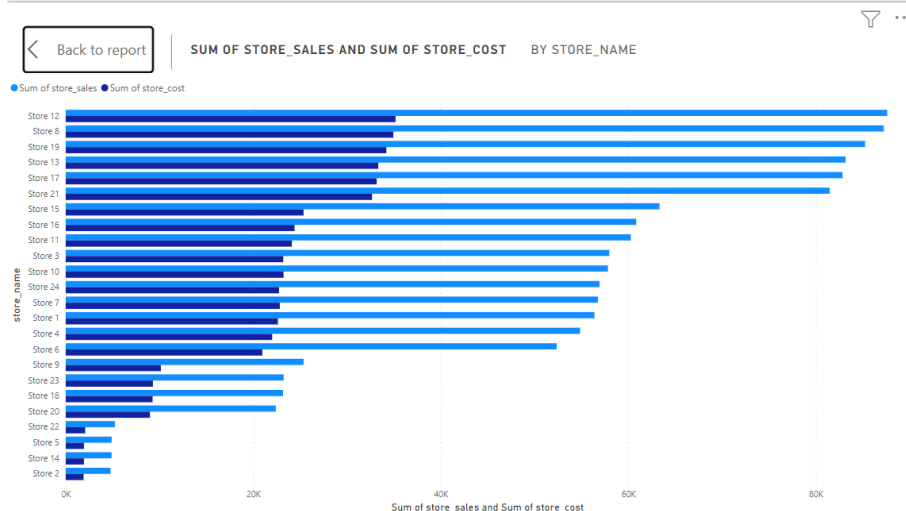
8.2 Từ dữ liệu theo cửa hàng

Hệ thống cho ta thông tin cụ thể về 2 số liệu quan trọng là doanh số và kinh phí, được thống kê theo loại cửa hàng, tên cửa hàng.

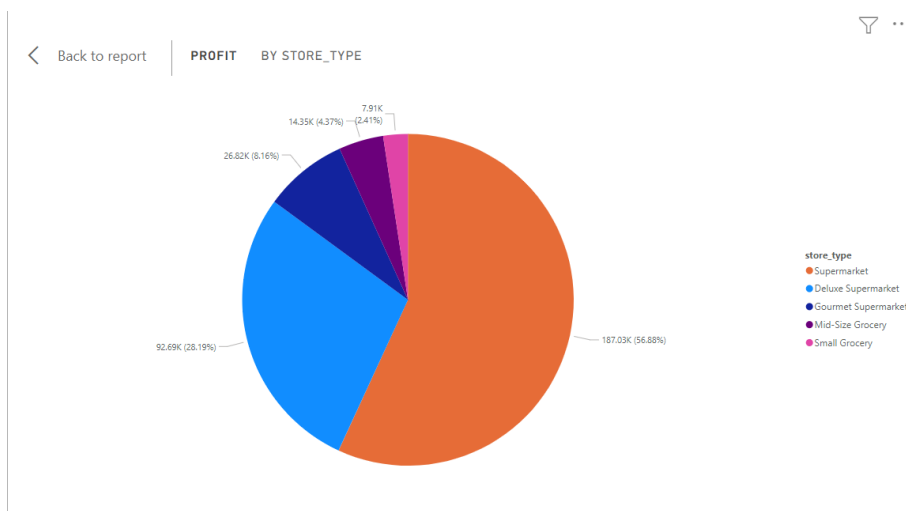
Từ thông tin ở các biểu đồ, ta có thể đưa ra kết luận về sự chênh lệch doanh số theo từng cửa hàng và loại cửa hàng.



Hình 22: Phân tích số liệu theo tên cửa hàng trong năm 1997



Hình 23: Phân tích số liệu theo tên cửa hàng trong năm 1998



Hình 24: Phân tích số liệu lợi nhuận theo loại cửa hàng

Qua những dữ liệu thống kê trên, ta có thể thấy:

- Các cửa hàng thuộc loại siêu thị (supermarket) có lợi nhuận chiếm tỷ lệ rất lớn (hơn 50% tổng số lợi nhuận).

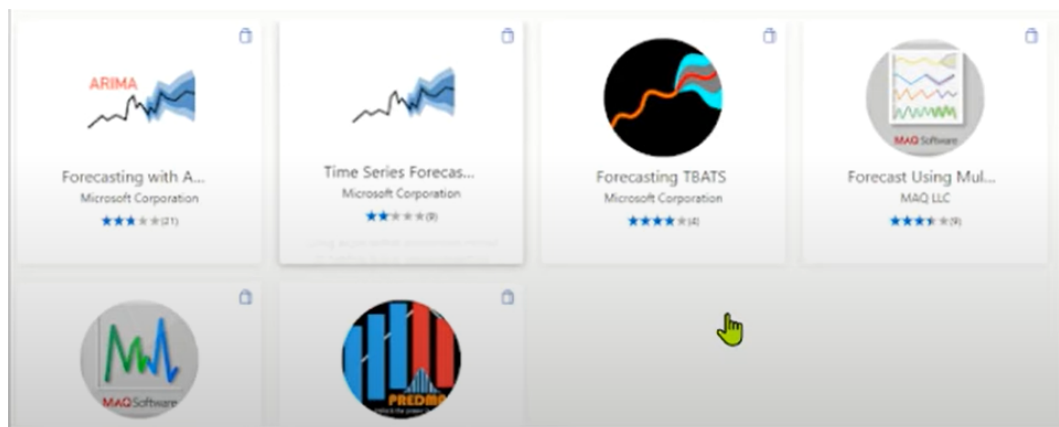
- Vào năm 1997, doanh số ở cửa hàng số 13 đạt cao nhất (84,622.40) và cửa hàng số 14 đạt thấp nhất (4,174.44). Lợi nhuận ở cửa hàng số 13 đạt lớn nhất.

- Vào năm 1998, doanh số ở cửa hàng số 12 đạt cao nhất (87,623.61) và cửa hàng số 2 đạt thấp nhất (4,779.90). Lợi nhuận ở cửa hàng số 12 đạt lớn nhất.

→ Từ những thông tin trên, ta có thể tạm thời quyết định vào những thời điểm cụ thể trong năm, ta nên đầu tư về hàng hóa, quảng cáo, dịch vụ ở những loại cửa hàng và địa chỉ cửa hàng cụ thể để mang lại hiệu quả kinh doanh cao, ngược lại có bước điều chỉnh phù hợp để khắc phục doanh số ở những loại cửa hàng có doanh số thấp.

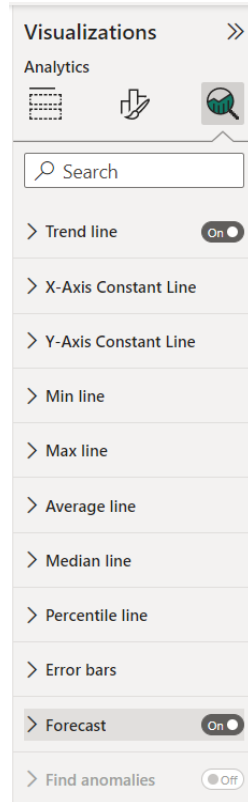
9 Hiện thực tính năng dự đoán (forecast) đối với tập dữ liệu

- Để thực hiện tính năng dự đoán trong Power BI, nhóm sẽ sử dụng hai tính năng là **Trend** và **Forecast** của Power BI.
- Một số thuật toán dùng để dự đoán (forecast) trong Power BI có thể sử dụng như: Arima, Time Series, Regression,...



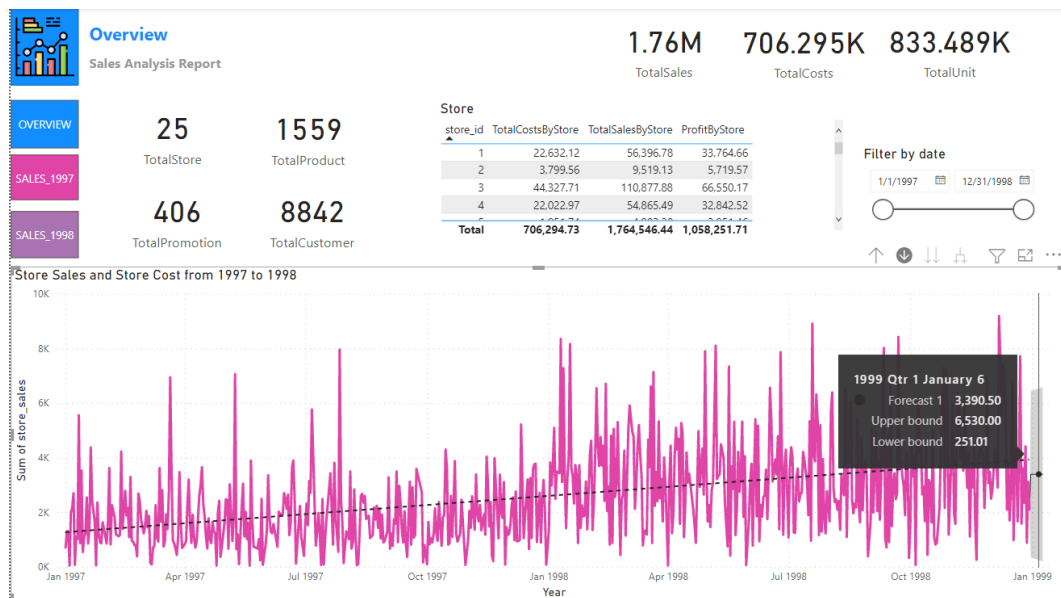
Hình 25: Một số thuật toán forecast trong Power BI

- Để thực hiện tính năng forecast trong Power BI chúng ta sẽ vào phần **Add further analytics to your visual** để enable tính năng forecast và trend.



Hình 26: Bật tính năng forecast trong Power BI

- Power BI sẽ tiến hành sử dụng các thuật toán forecast để train tập dữ liệu đầu vào theo thời gian, ở đây nhóm chọn tập dữ liệu đầu vào là tập hợp của cả hai bảng sales_fact 1997 và sales_fact_1998. Kết quả được Power BI dự đoán như sau:



Hình 27: Kết quả dự đoán doanh thu của chuỗi cửa hàng vào ngày 6/1/1999

10 Tài liệu tham khảo

1. What is business intelligence?, Micosoft, truy cập từ <https://powerbi.microsoft.com/en-us/what-is-business-intelligence/>
2. What is Power BI? Why to use Power BI ?, truy cập từ <https://www.thegioididong.com/game-app/power-bi-la-gi-tai-sao-cac-doanh-nghiep-nen-su-dung-1370098>
3. Power BI Crash Course, Simon Sez IT Channel, Youtube, truy cập từ <https://youtu.be/sIxJCKsa0K0>
4. Pentaho Mondiran Foodmart dataset, truy cập từ <https://mondrian.pentaho.com/documentation/olap.php>
5. How to make real-time data in Power BI using Maria DB ?, truy cập từ <https://community.powerbi.com/t5/Report-Server/Live-connection-with-Maria-DB/m-p/1154910>
6. How to forecast data in Power BI ?, truy cập từ <https://youtu.be/rwnDZxrzPCQ>