NGÂN HÀNG NHÀ NƯỚC VIỆT NAM BỘ GIÁO DỤC VÀ ĐÀO TẠO

**TRƯỜNG ĐẠI HỌC NGÂN HÀNG TP. HỒ CHÍ MINH**

**KHOA HỆ THỐNG THÔNG TIN QUẢN LÝ**

-------------------------------------

**A picture containing text, graphics, font, logo

Description automatically generated**

**ĐỒ ÁN MÔN HỌC**

**TRÍ TUỆ KINH DOANH**

**ĐỀ TÀI: DỰ ĐOÁN DOANH SỐ BÁN HÀNG (SALES FORECASTING)**

**Nhóm thực hiện: Nhóm 06**

**Lớp: BIN701\_241\_1\_D01**

**GVHD: ThS. Vương Trọng Nhân**

*Thành phố Hồ Chí Minh, 12 tháng 11 năm 2024*

# BẢNG PHÂN CÔNG CÔNG VIỆC

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **STT** | **Họ và tên** | **MSSV** | **Chức vụ** | **Công việc** | **Tỉ lệ tham gia** |
| 1 | Nguyễn Ngọc Hồng Hòa | 030238220067 | Trưởng nhóm | * Thu thập và làm sạch dữ liệu * Phân tích dữ liệu và mô hình hóa * Đưa ra đề xuất và chiến lược kinh doanh * Viết bài báo cáo | 100% |
| 2 | Phạm Phương Bình | 030238220013 | Thành viên | * Chọn lĩnh vực kinh doanh và xác định vấn đề cần giải quyết * Phân tích dữ liệu và mô hình hóa * Trực quan hóa dữ liệu * Viết bài báo cáo | 100% |
| 3 | Ngô Bảo Ngân | 030238220142 | Thành viên | * Thu thập và làm sạch dữ liệu * Phân tích dữ liệu và mô hình hóa * Trực quan hóa dữ liệu * Viết bài báo cáo | 100% |
| 4 | Huỳnh Thị Anh Nguyên | 030238220153 | Thành viên | * Chọn lĩnh vực kinh doanh và xác định vấn đề cần giải quyết * Phân tích dữ liệu và mô hình hóa * Trực quan hóa dữ liệu * Viết bài báo cáo | 100% |
| 5 | Nguyễn Phan Thị Khánh Vy | 030238220306 | Thành viên | * Thu thập và làm sạch dữ liệu * Phân tích dữ liệu và mô hình hóa * Đưa ra đề xuất và chiến lược kinh doanh * Viết bài báo cáo | 100% |

# NHẬN XÉT CỦA GIẢNG VIÊN

*…..ngày…..tháng…..năm 2024*

Giảng viên

# MỤC LỤC

[BẢNG PHÂN CÔNG CÔNG VIỆC i](#_Toc207367506)

[NHẬN XÉT CỦA GIẢNG VIÊN ii](#_Toc207367507)

[MỤC LỤC iii](#_Toc207367508)

[DANH MỤC HÌNH iv](#_Toc207367509)

[LỜI MỞ ĐẦU v](#_Toc207367510)

[NỘI DUNG 1](#_Toc207367511)

[1. Chọn lĩnh vực kinh doanh và xác định vấn đề cần giải quyết 1](#_Toc207367512)

[1.1. Mô tả về bộ dữ liệu Retail Data Analytics 1](#_Toc207367513)

[1.2. Xác định vấn đề kinh doanh cần giải quyết 1](#_Toc207367514)

[1.3. Phạm vi phân tích 1](#_Toc207367515)

[1.4. Mục tiêu cần đạt 1](#_Toc207367516)

[1.5. Các câu hỏi nghiên cứu mà BI có thể trả lời 2](#_Toc207367517)

[2. Thu thập và làm sạch dữ liệu 2](#_Toc207367518)

[2.1. Thu thập dữ liệu 2](#_Toc207367519)

[2.2. Làm sạch dữ liệu 3](#_Toc207367520)

[2.2.2. Chuẩn hóa dữ liệu 3](#_Toc207367521)

[2.2.3. Hợp nhất dữ liệu 3](#_Toc207367522)

[2.3. Kết luận 4](#_Toc207367523)

[3. Phân tích dữ liệu và mô hình hóa 4](#_Toc207367524)

[3.1. Phân tích dữ liệu 4](#_Toc207367525)

[3.2. Mô hình hóa 7](#_Toc207367526)

[3.2.1. Phân tích chu kỳ và xu hướng doanh số bán hàng 7](#_Toc207367527)

[3.2.2. Phân Tích và Dự Báo Doanh Số Hàng Tháng 7](#_Toc207367528)

[4. Trực quan hóa dữ liệu 9](#_Toc207367529)

[4.1 Về các chỉ số hiệu suất chính (KPIs) 10](#_Toc207367530)

[4.2 Phân tích các biểu đồ 10](#_Toc207367531)

[4.3 Tổng kết 13](#_Toc207367532)

[5. Đưa ra đề xuất và chiến lược kinh doanh 13](#_Toc207367533)

[KẾT LUẬN 15](#_Toc207367534)

[TÀI LIỆU THAM KHẢO 16](#_Toc207367535)

# DANH MỤC HÌNH

[Hình 1: Features data-set 2](#_Toc182691358)

[Hình 2: Sales data-set 2](#_Toc182691359)

[Hình 3: Stores data-set 3](#_Toc182691360)

[Hình 4: Dữ liệu đã xử lý 4](#_Toc182691361)

[Hình 5: Biểu đồ thể hiện doanh số bán hàng hàng tuần 4](#_Toc182691362)

[Hình 6: Biểu đồ thể hiện sự giảm giá trong kỳ nghỉ lễ 5](#_Toc182691363)

[Hình 7: Biểu đồ doanh thu theo từng phòng ban tại mỗi cửa hàng 5](#_Toc182691364)

[Hình 8: Sales Forecasting Dashboard 9](#_Toc182691365)

[Hình 9: Các chỉ số hiệu suất chính 10](#_Toc182691366)

[Hình 10: Biểu Đồ Nhiệt Độ Tương Quan 10](#_Toc182691367)

[Hình 11: Biểu đồ phân tán 11](#_Toc182691368)

[Hình 12: Biểu đồ cột 11](#_Toc182691369)

[Hình 13: Biểu đồ cột chồng 12](#_Toc182691370)

[Hình 14: Biểu đồ tròn 12](#_Toc182691371)

# LỜI MỞ ĐẦU

Trong môi trường kinh doanh ngày nay, việc ra quyết định dựa trên dữ liệu (Data-Driven Decision Making) đang ngày càng trở nên quan trọng hơn bao giờ hết. Các công cụ phân tích dữ liệu không chỉ giúp doanh nghiệp hiểu rõ hơn về thị trường, khách hàng mà còn cung cấp những thông tin quan trọng để tối ưu hóa chiến lược kinh doanh và giảm thiểu rủi ro. Đặc biệt trong ngành bán lẻ, các cửa hàng phải đối mặt với nhiều yếu tố tác động từ cả bên ngoài và bên trong, bao gồm các chiến dịch khuyến mãi, các yếu tố kinh tế vĩ mô như giá nhiên liệu, tỷ lệ thất nghiệp, chỉ số giá tiêu dùng (CPI) và sự thay đổi hành vi của người tiêu dùng trong các kỳ lễ lớn.

Vì vậy chúng em chọn lĩnh vực bán lẻ, với mục tiêu giải quyết một vấn đề kinh doanh cụ thể thông qua việc phân tích bộ dữ liệu Retail Data Analytics. Việc áp dụng các kỹ thuật phân tích dữ liệu sẽ giúp các cửa hàng bán lẻ dự báo chính xác về doanh số trong tương lai, giúp tối ưu hóa quản lý hàng tồn kho và chiến lược kinh doanh.

# NỘI DUNG

## 1. Chọn lĩnh vực kinh doanh và xác định vấn đề cần giải quyết

### 1.1. Mô tả về bộ dữ liệu Retail Data Analytics

Bộ dữ liệu cung cấp thông tin chi tiết về 45 cửa hàng bán lẻ, mỗi cửa hàng có nhiều phòng ban và hoạt động tại các khu vực khác nhau. Dữ liệu bao gồm:

* Doanh thu hàng tuần của các cửa hàng và các phòng ban từ tháng 2/2010 đến tháng 12/2012.
* Thông tin về các chương trình giảm giá (Markdown 1-5) diễn ra trước các ngày lễ lớn như Super Bowl, Ngày Lao Động, Lễ Tạ ơn và Giáng Sinh.
* Các yếu tố bên ngoài ảnh hưởng đến doanh thu, bao gồm: nhiệt độ trung bình khu vực (Temperature), giá nhiên liệu (Fuel\_Price), chỉ số giá tiêu dùng (CPI), và tỷ lệ thất nghiệp (Unemployment).
* Thông tin về các tuần lễ đặc biệt (IsHoliday), trong đó các tuần lễ đặc biệt được đánh giá trọng số cao hơn trong việc xác định doanh thu bán hàng.

### 1.2. Xác định vấn đề kinh doanh cần giải quyết

* Dự đoán doanh số bán hàng (Sales Forecasting): Dự đoán doanh thu của từng cửa hàng và phòng ban trong tương lai dựa trên dữ liệu lịch sử.
* Tác động của các chương trình khuyến mãi (Markdowns) trong các dịp lễ (ví dụ như Giáng Sinh, Lễ Lao động, và Super Bowl) đến doanh thu của cửa hàng
* Ảnh hưởng của mùa vụ (tháng cao điểm và thấp điểm) đến doanh thu bán hàng
* Ảnh hưởng của các yếu tố kinh tế vĩ mô (như giá nhiên liệu (Fuel Price), tỷ lệ thất nghiệp (Unemployment), chỉ số giá tiêu dùng (CPI)) ảnh hưởng đến hành vi tiêu dùng và doanh thu.
* Phân tích hiệu suất theo khu vực và loại cửa hàng: Các cửa hàng nằm ở các khu vực khác nhau có thể có hiệu suất khác nhau. Tìm hiểu sự khác biệt về doanh thu giữa các cửa hàng để đưa ra chiến lược phù hợp cho từng khu vực hoặc loại cửa hàng.

### 1.3. Phạm vi phân tích

Phạm vi phân tích dữ liệu từ ba nguồn chính: Stores, Features và Sales, bao gồm thông tin từ tháng 2/2010 đến tháng 11/2012. Phân tích sẽ tập trung vào việc xác định và đánh giá các yếu tố ảnh hưởng đến doanh số bán hàng hàng tuần của 45 cửa hàng, bao gồm các yếu tố nội bộ như các chương trình giảm giá (MarkDown1-5), và sự kiện đặc biệt (IsHoliday), cùng các yếu tố bên ngoài như nhiệt độ (Temperature), giá nhiên liệu (Fuel\_Price), chỉ số giá tiêu dùng (CPI) và tỷ lệ thất nghiệp (Unemployment). Mục tiêu là đưa ra các dự báo doanh số cho các tuần hoặc tháng tiếp theo, áp dụng cho các cửa hàng ở các khu vực khác nhau.

### 1.4. Mục tiêu cần đạt

* Dự đoán doanh thu bán hàng trong các tuần hoặc tháng tiếp theo, đặc biệt là trong các kỳ lễ lớn và các sự kiện quan trọng, để giúp các cửa hàng lên kế hoạch tồn kho và chiến lược marketing hiệu quả.
* Đánh giá được tác động của mùa vụ đến doanh thu từ đó tối ưu hóa chiến dịch marketing, cải thiện chiến lược quảng cáo.
* Xác định ảnh hưởng của các yếu tố kinh tế vĩ mô (như CPI, tỷ lệ thất nghiệp, giá nhiên liệu) đến hành vi tiêu dùng và doanh thu bán hàng.
* Phân tích hiệu suất bán hàng theo khu vực và theo cửa hàng, nhằm đưa ra các chiến lược tối ưu cho từng khu vực và từng loại cửa hàng.

### 1.5. Các câu hỏi nghiên cứu mà BI có thể trả lời

* Dự đoán doanh thu bán hàng trong các tuần hoặc tháng tiếp theo:

+ Doanh thu bán hàng trong các tuần hoặc tháng tiếp theo có xu hướng tăng hay giảm?

* Tác động của các chương trình khuyến mãi (Markdowns) vào các ngày lễ:

+ Các chương trình giảm giá có ảnh hưởng tích cực hay tiêu cực đến doanh thu bán hàng?

* Ảnh hưởng của yếu tố kinh tế vĩ mô:

+ Liệu có sự tương quan giữa giá xăng, tỷ lệ thất nghiệp và doanh thu bán hàng của các cửa hàng?

* Ảnh hưởng của mùa vụ đến doanh thu bán hàng:

+ Doanh thu có biến động giữa các tháng cao điểm và thấp điểm hay không?

* Phân tích hiệu suất theo khu vực:

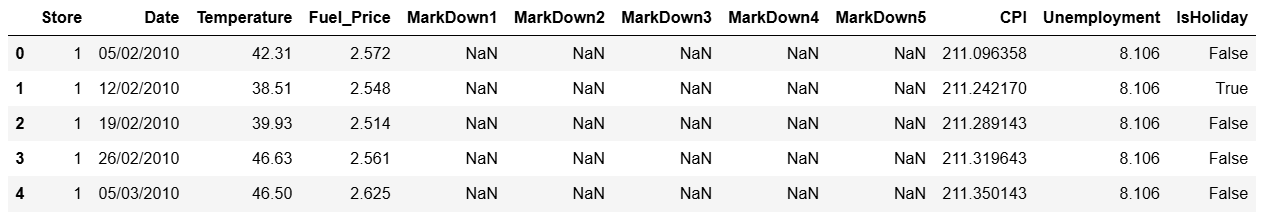
+ Các cửa hàng ở các khu vực khác nhau có hiệu suất doanh thu khác nhau như thế nào?

## 2. Thu thập và làm sạch dữ liệu

### 2.1. Thu thập dữ liệu

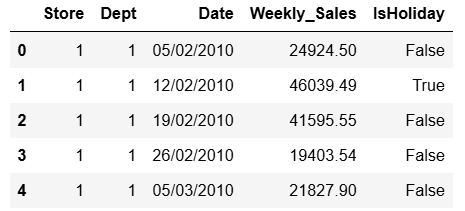
Dữ liệu được thu thập từ ba tệp:

* **Features data-set.csv**: Chứa thông tin về giá nhiên liệu, nhiệt độ, chỉ số CPI, và tỷ lệ thất nghiệp.



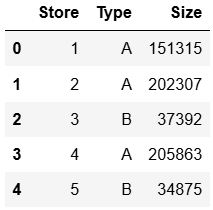
Hình 1: Features data-set

* **Sales data-set.csv**: Chứa thông tin doanh số bán hàng hàng tuần.



Hình 2: Sales data-set

* **Stores data-set.csv**: Cung cấp thông tin chi tiết về từng cửa hàng.



Hình 3: Stores data-set

Cách thực hiện:

* Tải dữ liệu vào các DataFrame.
* Xem xét dữ liệu ban đầu để kiểm tra cấu trúc và nội dung đang có vấn đề gì.

### 2.2. Làm sạch dữ liệu

**2.2.1 Xử lý giá trị thiếu**

* Khi xem xét dữ liệu từ **Features**, nhận thấy một số cột như **MarkDown1, MarkDown2**,... chứa các giá trị NaN (không có dữ liệu). Đây là các cột liên quan đến các chiến dịch khuyến mãi và không phải tuần nào cũng có khuyến mãi. Cách xử lý là gán giá trị NaN trong các cột MarkDown bằng **0**, vì khi không có khuyến mãi, giá trị tương ứng có thể xem như không tồn tại.
* Các cột **CPI** (chỉ số giá tiêu dùng) và **Unemployment** (tỷ lệ thất nghiệp): Đây là các dữ liệu kinh tế vĩ mô có thể bị thiếu ở một số thời điểm. Ta sử dụng phương pháp nội suy theo thời gian (**forward fill**) để điền giá trị.

#### 2.2.2. Chuẩn hóa dữ liệu

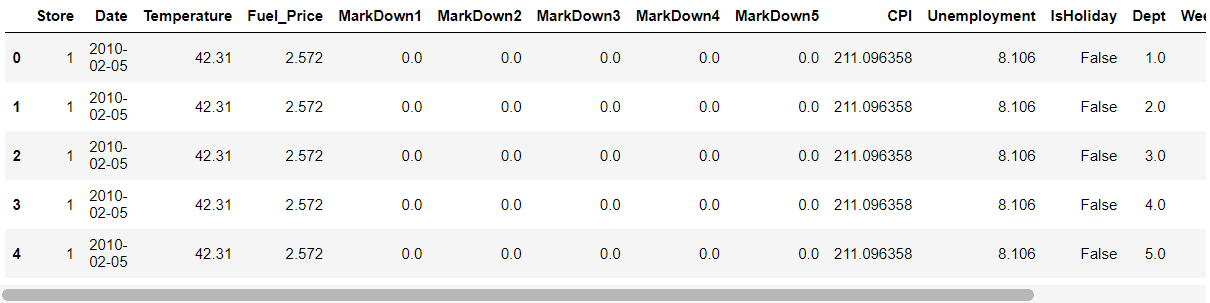
* Định dạng cột ngày tháng: Trong cả ba tệp dữ liệu, cột ngày tháng đang ở dạng chuỗi (string). Để phân tích dữ liệu theo thời gian, cần chuẩn hóa về định dạng **datetime**.

#### 2.2.3. Hợp nhất dữ liệu

* Sau khi làm sạch và chuẩn hóa từng DataFrame, cần hợp nhất chúng để tạo thành một bộ dữ liệu hoàn chỉnh.

**+ Bước 1:** Hợp nhất **Features** và **Sales** dựa trên ba cột chung: **Store**, **Date**, và **IsHoliday**. Vì các cột này là chìa khóa chính để liên kết thông tin về đặc điểm và doanh số bán hàng.

**+ Bước 2:** Hợp nhất kết quả trên với **Stores** để thêm thông tin kích thước và loại cửa hàng. Khi đó **mergedDF** là một DataFrame duy nhất, chứa tất cả thông tin từ ba nguồn.



Hình 4: Dữ liệu đã xử lý

### 2.3. Kết luận

* Chất lượng dữ liệu tốt hơn: Giá trị thiếu được xử lý, dữ liệu được chuẩn hóa và loại bỏ ngoại lệ.
* Tính đồng nhất: Các bảng dữ liệu từ nhiều nguồn đã được hợp nhất thành một bộ duy nhất.
* Phân tích hiệu quả hơn: Các bước trực quan hóa giúp phát hiện xu hướng và chuẩn bị cho các bước phân tích tiếp theo.

## 3. Phân tích dữ liệu và mô hình hóa

### 3.1. Phân tích dữ liệu

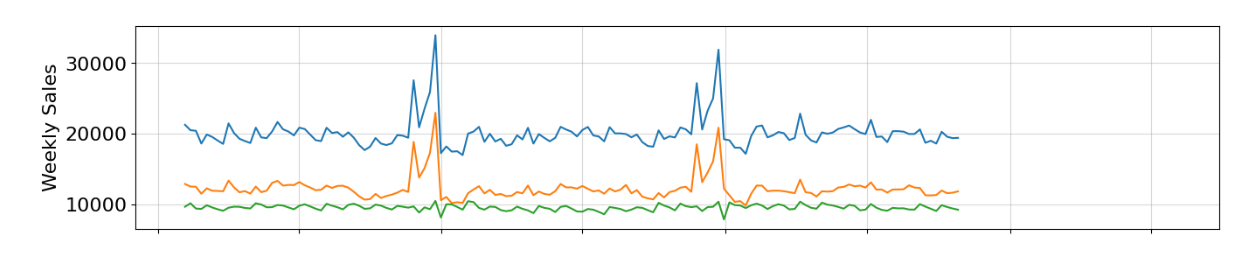
**- Phân tích mô tả dữ liệu**

+ Dữ liệu chứa các yếu tố như mã của hàng, doanh số hàng tuần, nhiệt độ, giá nhiên liệu, CPI, tỷ lệ thất nghiệp,….

+ Dựa vào các loại dữ liệu đã xử lý ở trên ta có thể phân tích mô tả sau:

**- Doanh số theo loại cửa hàng**:

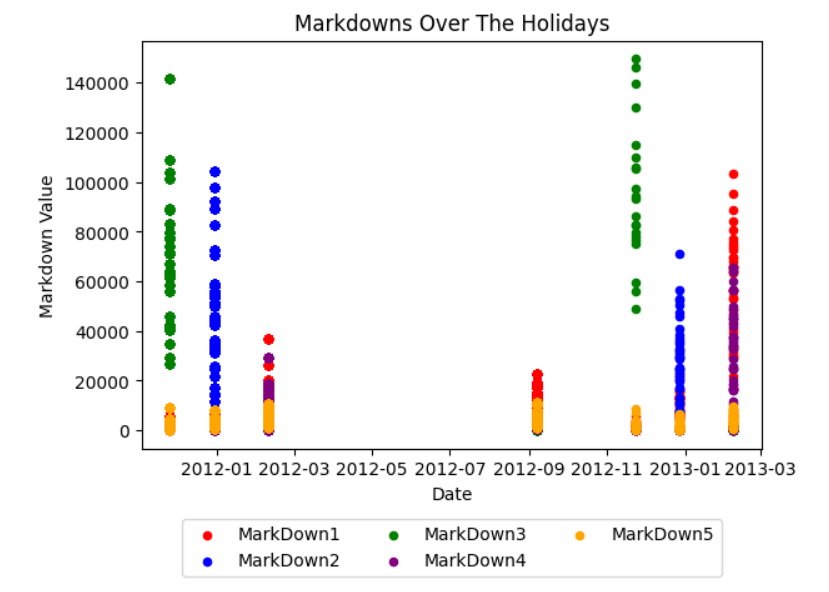
* Cửa hàng loại **A** đạt doanh số cao nhất, phản ánh hiệu suất vượt trội so với loại **B** và **C**.
* Cửa hàng loại **B** và **C** có mức doanh thu thấp hơn nhưng đóng góp ổn định.



Hình 5: Biểu đồ thể hiện doanh số bán hàng hàng tuần

Weekly Sales (Doanh số bán hàng hàng tuần): Có những đỉnh cao rõ ràng vào cuối năm 2010, 2011, trùng với các mùa lễ hội Type A luôn có doanh số cao hơn hẳn so với Type B và C. Type B và C có doanh số thấp hơn đáng kể và khá ổn định.

**- Giảm giá trong các kỳ nghỉ lễ:**

* Biểu đồ này giúp quan sát sự thay đổi mức giảm giá trong các kỳ nghỉ lễ, qua đó xác định loại giảm giá nào có giá trị lớn nhất hoặc biến động đáng kể theo thời gian.

Hình 6: Biểu đồ thể hiện sự giảm giá trong kỳ nghỉ lễ

Điều này giúp nhận diện các thời điểm giảm giá trong kỳ nghỉ lễ, tạo cơ sở để phân tích tác động của markdown đối với doanh số bán hàng.

**- Doanh số trong các ngày lễ khi áp dụng việc giảm giá**

+ Tổng doanh thu từ các ngày lễ có giảm giá: Phản ánh đóng góp của các chương trình giảm giá khi doanh nghiệp triển khai trong các ngày lễ đặc biệt.

+ Doanh thu trung bình hàng tuần trong các ngày lễ có giảm giá: Đo lường mức độ hiệu quả của các ngày lễ trong việc thúc đẩy doanh số.

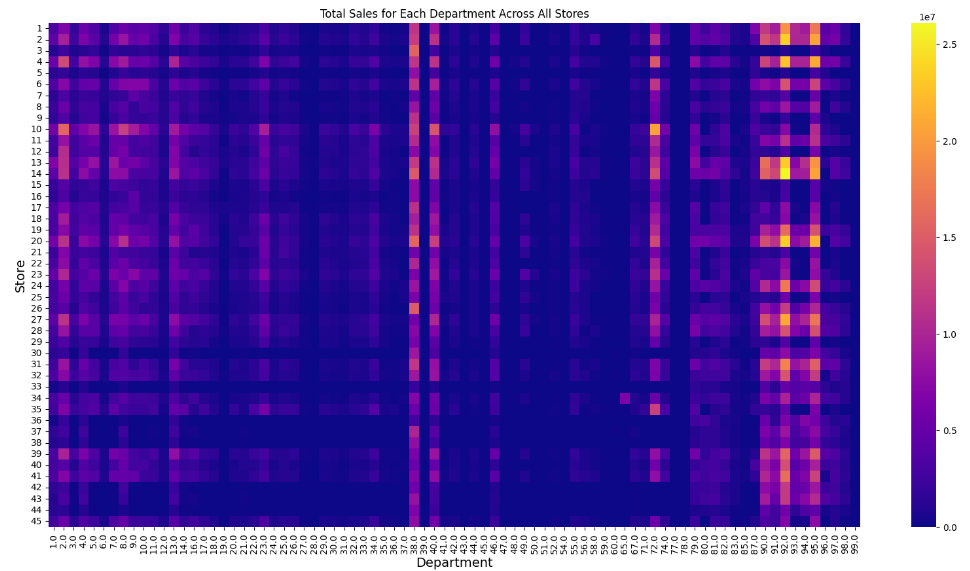
* Việc giảm giá góp phần làm tăng doanh số, đặc biệt là trong các kỳ nghỉ lễ. Làm nổi bật cách doanh số hàng tuần thay đổi theo thời gian và nhấn mạnh tác động tích cực của việc giảm giá đối với doanh số trong các mùa cao điểm của kỳ nghỉ lễ.

**- Doanh số bán hàng khi áp dụng chương trình khuyến mãi**

Doanh số đã cho thấy sự gia tăng nhẹ sau khi các chương trình giảm giá được triển khai.

**- Doanh thu theo từng phòng ban tại mỗi cửa hàng**

* Có 45 của hàng riêng biệt trong hệ thống.
* Có 81 phòng ban đang hoạt động.



Hình 7: Biểu đồ doanh thu theo từng phòng ban tại mỗi cửa hàng

Biểu đồ heatmap thể hiện tổng doanh thu của mỗi phòng ban tại các cửa hàng. Màu sắc sáng hơn biểu thị doanh thu cao, trong khi màu tối hơn cho thấy doanh thu thấp hoặc không có doanh thu (được điền bằng 0). Biểu đồ này giúp dễ dàng nhận diện các cửa hàng và phòng ban có doanh thu cao hoặc cần cải thiện.

**- Doanh số và giảm giá theo thời gian**

+ Khi áp dụng giảm giá, doanh số tăng cho thấy xu hướng bán hàng tích cực.

+ Mức giảm giá cao hơn tương quan với mức tăng đột biến của doanh số, cho thấy mức giảm giá góp phần vào tăng trưởng doanh số.

**- Hiệu suất của các cửa hàng và phòng ban hàng đầu**

+ Các cửa hàng hàng đầu có doanh số cao và ổn định hơn. Có những đỉnh cao rõ rệt vào khoảng thời gian cuối năm (dịp lễ hội): cuối năm 2010, cuối năm 2011. Xu hướng doanh số vẫn ổn định với mức dao động không quá lớn giữa các thời kỳ ngoài mùa lễ hội.

+ Ngược lại các cửa hàng cuối bảng có doanh số thấp, chỉ đạt khoảng 1/3 so với nhóm Top 10 Store. Biên độ dao động thấp hơn, cho thấy doanh số khá ổn định ở mức thấp.

* Tổng thể, các cửa hàng hàng đầu vượt trội hơn về doanh số và sự ổn định.

**- 10 cửa hàng hàng đầu và những yếu tố ảnh hưởng đến hiệu suất của chúng**

* Doanh số không có mối tương quan mạnh với các yếu tố như nhiệt độ, giá nhiên liệu, CPI, hoặc tỷ lệ thất nghiệp.
* Ngoài ra, loại hình cửa hàng (Type) có tác động đáng kể đến hiệu suất doanh thu. Phân tích cho thấy rằng loại hình cửa hàng (Type A, B, C) đóng vai trò quan trọng trong việc ảnh hưởng đến doanh thu và hiệu suất hoạt động của từng cửa hàng.

Doanh số không bị ảnh hưởng trực tiếp bởi các yếu tố như nhiệt độ, giá nhiên liệu, CPI, hay tỷ lệ thất nghiệp. Các biến động lớn chủ yếu do mùa lễ hội. Giá nhiên liệu, CPI, và tỷ lệ thất nghiệp có mối tương quan chặt chẽ với nhau, phản ánh tình hình kinh tế vĩ mô.

**- Nguyên nhân gì đã gây ra sự sụt giảm doanh số từ năm 2011 đến năm 2012?**

Mô tả sự thay đổi phần trăm theo từng năm của các chỉ số để xác định xu hướng tăng hoặc giảm.

* Doanh thu hàng tuần (Weekly\_Sales).
* Giá nhiên liệu (Fuel\_Price).
* Chỉ số giá tiêu dùng (CPI).
* Tỷ lệ thất nghiệp (Unemployment).

Chúng ta nhận thấy rằng giữa năm 2011 và 2012, việc giảm tỷ lệ thất nghiệp, tăng chỉ số giá tiêu dùng (CPI) và các chương trình giảm giá đã giúp giảm tốc độ doanh thu hàng tuần chậm lại, vì vậy yếu tố này có thể góp phần làm giảm sự biến động và giúp doanh nghiệp ổn định hơn.

### 3.2. Mô hình hóa

#### 3.2.1. Phân tích chu kỳ và xu hướng doanh số bán hàng

* **Phân Tích Tự Tương Quan (ACF) và Tự Tương Quan Một Phần (PACF):**

+ Mục tiêu: Xác định mối quan hệ giữa doanh số hiện tại và doanh số trong quá khứ.

+ Quy trình: Sử dụng biểu đồ ACF và PACF để nhận biết tính chu kỳ và các giá trị lag quan trọng.

+ Kết quả: Doanh số hiện tại có mối quan hệ mạnh với doanh số trong 1-2 tuần trước, đặc biệt là các giá trị lag 1 và lag 2.

* **Tính Toán Ma Trận Tương Quan:**

+ Mục tiêu: Phân tích mức độ phụ thuộc của doanh số hiện tại vào các yếu tố quá khứ như Lagged\_Weekly\_Sales, CPI, Temperature và Fuel\_Price.

+ Quy trình: Sử dụng .corr() và heatmap để tính và trực quan hóa ma trận tương quan.

+ Kết quả: Lagged\_Weekly\_Sales có tương quan cao nhất với Weekly\_Sales, trong khi Temperature, Fuel\_Price và CPI có tương quan yếu hoặc không đáng kể.

* **Mô Hình Hồi Quy Tuyến Tính (Linear Regression):**

+ Mục tiêu: Sử dụng Lagged\_Weekly\_Sales để dự đoán Weekly\_Sales.

+ Huấn luyện mô hình: Huấn luyện mô hình hồi quy tuyến tính và vẽ biểu đồ Lag Plot.

+ Kết quả: Dữ liệu phân tán rộng và không tập trung thành một đường thẳng, đặc biệt khi giá trị trễ tăng cao, cho thấy hồi quy tuyến tính có thể không phải là lựa chọn tối ưu.

#### 3.2.2. Phân Tích và Dự Báo Doanh Số Hàng Tháng

* **Phân Tích Tự Tương Quan (ACF) và Tự Tương Quan Một Phần (PACF):**

+ Mục tiêu: Xác định mối quan hệ giữa doanh số hiện tại và doanh số các kỳ trước.

+ Thực hiện: Sử dụng biểu đồ ACF và PACF để phân tích.

+ Kết quả: ACF cho thấy tự tương quan cao tại lag 12 (chu kỳ hàng năm). PACF cho thấy tương quan yếu nhưng có ý nghĩa tại lag 10.

* **Phân Tích Xu Hướng Doanh Số:**

+ Mục tiêu: Tổng hợp dữ liệu hàng tuần thành hàng tháng để phân tích xu hướng.

+ Thực hiện: Sử dụng phương pháp resample('M') và vẽ biểu đồ đường với đường hồi quy.

+ Kết quả: Doanh số biến động mạnh do yếu tố mùa vụ hoặc sự kiện đặc biệt, nhưng xu hướng tổng thể ổn định.

* **Áp Dụng Trung Bình Động Đơn Giản (SMA):**

+ Mục tiêu: Dự báo doanh số tháng tiếp theo dựa trên trung bình động.

+ Thực hiện: Sử dụng SMA 3 kỳ để dự báo và vẽ biểu đồ so sánh.

+ Kết quả: Dự báo doanh số tháng 11 năm 2012 là 15588.92. SMA giúp nhận diện xu hướng rõ hơn.

**- Áp Dụng ARIMA và Seasonal ARIMA cho Doanh Số Bán Hàng**

* **Kiểm Tra Tính Dừng và Lấy Sai Phân:**

+ Mục tiêu: Kiểm tra tính dừng của chuỗi thời gian và loại bỏ xu hướng nếu cần.

+ Thực hiện: Sử dụng kiểm định ADF và phép lấy sai phân.

+ Kết quả: Chuỗi ban đầu không dừng, sau khi lấy sai phân thì chuỗi trở nên dừng.

* **Phân Tích ACF và PACF:**

+ Mục tiêu: Xác định các giá trị p (AR) và q (MA).

+ Thực hiện: Sử dụng biểu đồ ACF và PACF.

+ Kết quả: Chọn p = 3 và q = 4.

* **Huấn Luyện Mô Hình ARIMA Thủ Công:**

+ Mục tiêu: Huấn luyện mô hình ARIMA(3, 1, 4) trên tập huấn luyện.

+ Thực hiện: Chia dữ liệu và huấn luyện mô hình bằng statsmodels.

+ Kết quả: Mô hình ARIMA(3, 1, 4) được huấn luyện thành công với các hệ số có ý nghĩa thống kê.

* **Dự Đoán và Đánh Giá Mô Hình ARIMA Thủ Công:**

+ Mục tiêu: Dự đoán doanh số trên tập kiểm tra và so sánh với giá trị thực tế.

+ Thực hiện: Dự đoán và tính toán các chỉ số MSE, RMSE.

+ Kết quả: Mô hình dự đoán tốt xu hướng nhưng gặp khó khăn với các đột biến mạnh.

* **Tìm Mô Hình Auto ARIMA và So Sánh:**

+ Mục tiêu: Sử dụng Auto ARIMA để tự động chọn tham số và so sánh với mô hình thủ công.

+ Thực hiện: Huấn luyện Auto ARIMA và so sánh kết quả dự đoán.

+ Kết quả: Auto ARIMA kém hơn so với mô hình thủ công trong việc dự đoán dài hạn và các dao động lớn.

**- Áp Dụng ARIMA và SARIMA cho Dự Báo Doanh Số**

* **Tách Chuỗi Thời Gian Thành Các Thành Phần:**

+ Mục tiêu: Tách Weekly\_Sales thành trend, seasonal, và residual.

+ Thực hiện: Sử dụng phương pháp STL và vẽ biểu đồ.

+ Ý nghĩa: Hiểu rõ cấu trúc chuỗi thời gian, bao gồm xu hướng dài hạn và tính chu kỳ.

* **Phân Tích ACF cho Seasonal và Trend:**

+ Mục tiêu: Đánh giá mối liên quan giữa các giá trị chuỗi thời gian.

+ Thực hiện: Vẽ biểu đồ ACF cho seasonal và trend.

+ Ý nghĩa: Seasonal ACF cho thấy tính mùa vụ rõ rệt, Trend ACF phản ánh xu hướng tăng.

* **Grid Search để Chọn Tham Số SARIMA Tối Ưu:**

+ Mục tiêu: Tìm tham số tối ưu cho mô hình SARIMA.

+ Thực hiện: Thử các tổ hợp tham số và chọn dựa trên chỉ số BIC.

+ Kết quả: Tham số tối ưu là (1, 0, 2) và (0, 1, 0, 4) với BIC = 3089.516934.

* **Huấn Luyện Mô Hình SARIMA:**

+ Mục tiêu: Huấn luyện mô hình SARIMA với tham số tối ưu.

+ Thực hiện: Xây dựng mô hình và kiểm tra phần dư.

+ Kết quả: Mô hình hợp lý nhưng cần cải thiện phần dư không tuân theo phân phối chuẩn.

* **Dự Đoán và So Sánh Các Mô Hình:**

+ Mục tiêu: So sánh dự báo của ARIMA thủ công, Auto ARIMA, và SARIMA.

+ Thực hiện: Dự đoán và vẽ biểu đồ so sánh với dữ liệu thực tế.

+ Kết quả: Cả ba mô hình đều có sai số, nhưng ARIMA thủ công cho kết quả tốt nhất.

* **Đánh Giá và So Sánh Hiệu Suất Mô Hình:**

+ Mục tiêu: So sánh độ chính xác của các mô hình dựa trên RMSE.

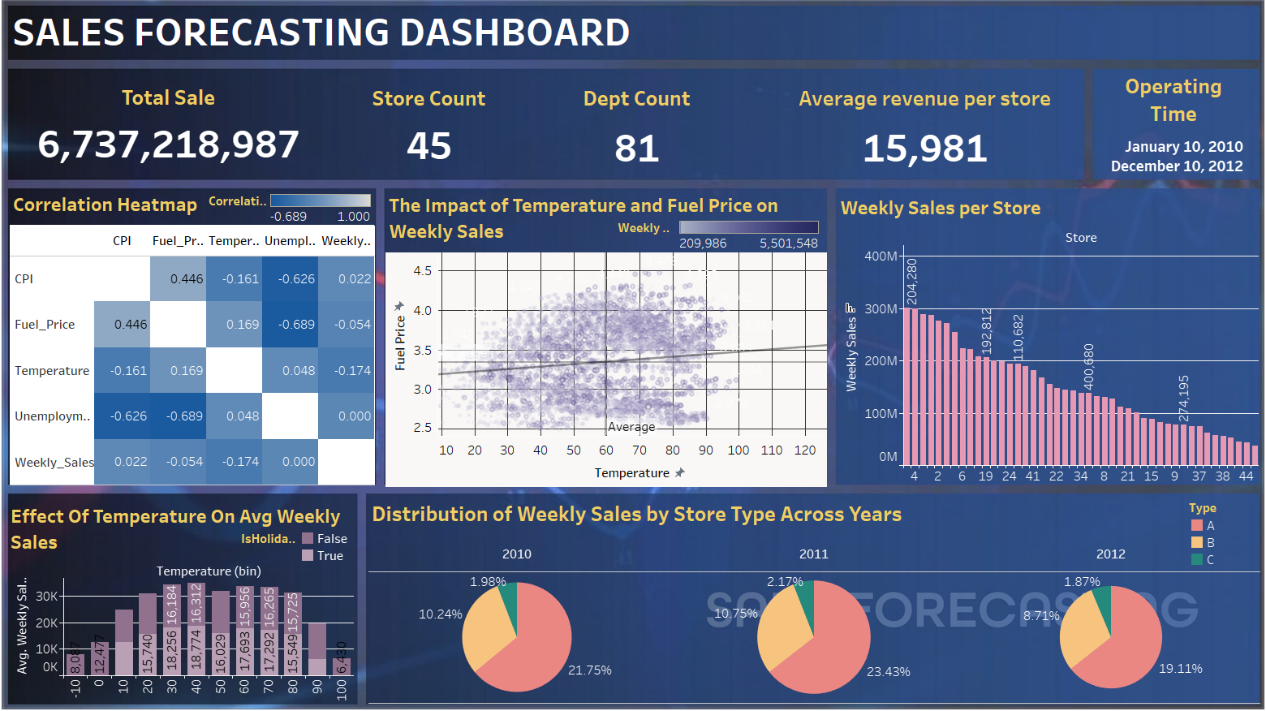
+ Thực hiện: Tính RMSE và vẽ biểu đồ cột để so sánh.

+ Kết quả: ARIMA thủ công có RMSE nhỏ nhất so với hai mô hình còn lại, chứng tỏ là mô hình chính xác nhất.

* Kết luận lại, ARIMA thủ công là mô hình được khuyến nghị sử dụng vì có hiệu suất dự báo tốt nhất (RMSE nhỏ nhất) để dự báo doanh số bán hàng trong tương lai.

## 4. Trực quan hóa dữ liệu

Với đề tài **Sales Forecasting**, nhóm tụi em tiến hành dùng Tableau để trực quan hóa phân tích bảng điều khiển dự báo doanh số được cung cấp, làm nổi bật các chỉ số chính, xu hướng và thông tin chi tiết từ dữ liệu. Mục tiêu là cung cấp một cái nhìn tổng quan toàn diện về bảng điều khiển và đề xuất các lĩnh vực cần cải thiện.



Hình 8: Sales Forecasting Dashboard

### 4.1. Về các chỉ số hiệu suất chính (KPIs)

Bảng điều khiển cung cấp các chỉ số chính sau:

* Tổng Doanh Số: 6,737,218,987
* Số Lượng Cửa hàng: 45
* Số Lượng Phòng ban: 81
* Doanh Thu Trung Bình Mỗi Cửa Hàng: 15,981
* Thời Gian Hoạt Động: 10/01/2010 - 10/12/2012



Hình 9: Các chỉ số hiệu suất chính

### 4.2. Phân tích các biểu đồ

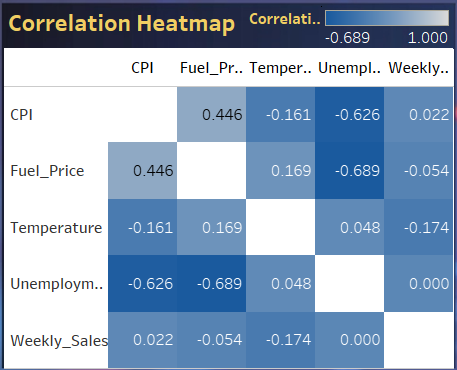
* **Biểu đồ nhiệt độ tương quan (Correlation Heatmap)**

+ Giá nhiên liệu và tỷ lệ thất nghiệp (-0.689): Mối quan hệ tương quan âm mạnh cho thấy khi giá nhiên liệu tăng, tỷ lệ thất nghiệp có xu hướng giảm.

+ Nhiệt độ và doanh số hàng tuần (0.169): Mối tương quan dương yếu thể hiện rằng nhiệt độ cao hơn có thể dẫn đến sự gia tăng nhẹ trong doanh số hàng tuần.

+ CPI và doanh số hàng tuần (0.022): Chỉ số giá tiêu dùng (CPI) có tác động rất nhỏ đến doanh số hàng tuần, gần như không đáng kể.

+ Thông qua Biểu Đồ Nhiệt Độ Tương Quan có thể thấy, giá nhiên liệu và nhiệt độ là hai yếu tố có mức ảnh hưởng đáng kể hơn đến doanh số hàng tuần, trong khi CPI chỉ có ảnh hưởng nhẹ và không rõ ràng.



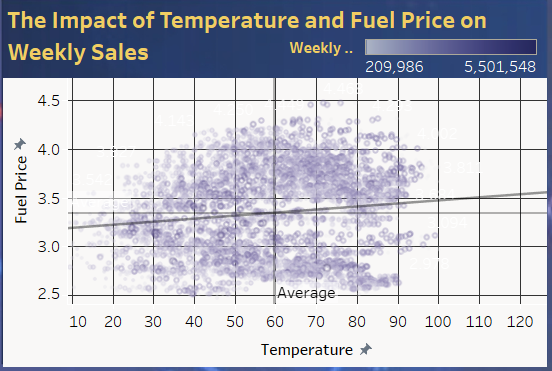
Hình 10: Biểu đồ nhiệt độ tương quan

* **Biểu đồ phân tán (Impact of Temperature and Fuel Price on Weekly Sales)**

+ Biểu đồ phân tán cung cấp thông tin chi tiết về mối quan hệ giữa nhiệt độ, giá nhiên liệu và doanh số hàng tuần. Doanh số hàng tuần có xu hướng giảm khi giá nhiên liệu tăng.

+ Doanh số bán hàng có xu hướng cao nhất ở mức nhiệt độ trung bình, đặc biệt là trong khoảng từ 50 đến 70°F. Nhiệt độ quá cao hoặc quá thấp có thể làm giảm doanh số, như thấy ở hai đầu của biểu đồ (nhiệt độ dưới 20°F hoặc trên 80°F).

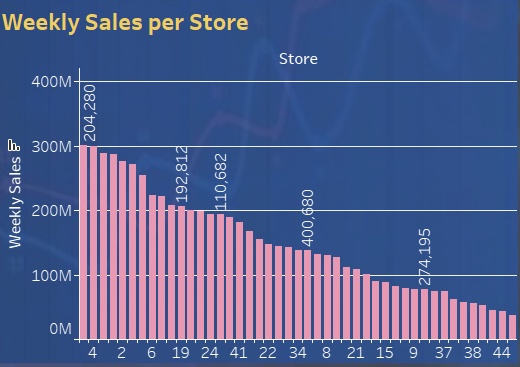
+ Điều này cho thấy rằng thời tiết ôn hòa có thể là điều kiện lý tưởng để thúc đẩy doanh số bán hàng. Biểu đồ này thể hiện mối quan hệ phức tạp nhưng không quá mạnh mẽ giữa nhiệt độ, giá nhiên liệu và doanh số. Sử dụng các yếu tố nhiệt độ và giá nhiên liệu để tối ưu hóa hoạt động kinh doanh là rất quan trọng.



Hình 11: Biểu đồ phân tán

* **Biểu đồ cột (Weekly Sales per Store)**

Biểu đồ hiển thị tổng doanh số hàng tuần của từng cửa hàng, với các cửa hàng được sắp xếp theo thứ tự giảm dần về doanh số. Trong khi một số cửa hàng đạt hiệu suất vượt trội, nhiều cửa hàng khác vẫn chưa phát huy tiềm năng. Việc phân tích sâu hiệu suất của từng cửa hàng và triển khai các chiến lược cải thiện phù hợp sẽ giúp tối ưu hóa tổng doanh thu.

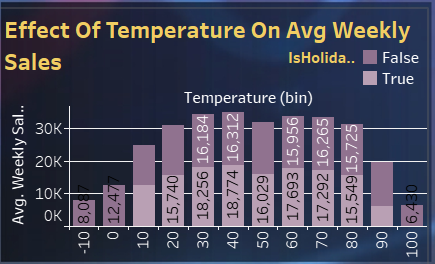


Hình 12: Biểu đồ cột

* **Biểu đồ cột chồng (Effect of Temperature on Average Weekly Sales)**

+ Biểu đồ cung cấp thông tin chi tiết về tác động của nhiệt độ đến doanh số hàng tuần trung bình và so sánh giữa ngày thường và kỳ nghỉ lễ.

+ Trong tất cả các khoảng nhiệt độ, doanh số trong kỳ nghỉ lễ luôn cao hơn đáng kể so với ngày thường. Biểu đồ cho thấy nhiệt độ và kỳ nghỉ lễ có tác động đáng kể đến doanh số hàng tuần trung bình. Bên cạnh đó, cần chuẩn bị chiến lược phù hợp cho các giai đoạn thời tiết khắc nghiệt để duy trì hiệu quả kinh doanh.

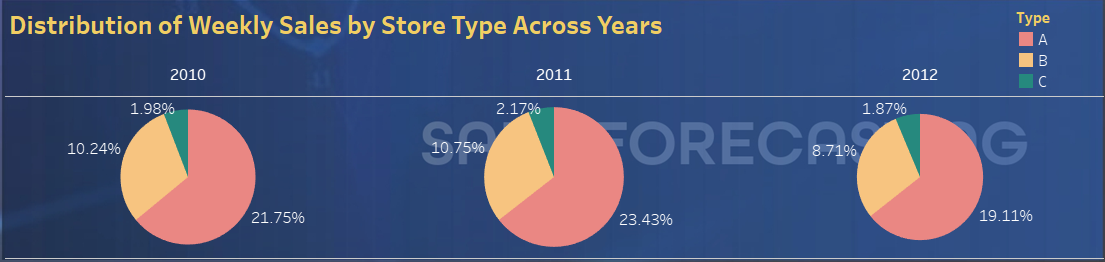


Hình 13: Biểu đồ cột chồng

* **Biểu đồ tròn (Distribution of Weekly Sales by Store Type Across Years)**

+ Biểu đồ này minh họa sự phân bố doanh số hàng tuần theo loại cửa hàng (A, B, C) qua các năm 2010, 2011 và 2012.

+ Biểu đồ cho thấy loại cửa hàng A là nguồn đóng góp doanh thu chính, trong khi loại B đóng vai trò hỗ trợ và loại C có hiệu suất rất thấp. Doanh nghiệp cần tập trung cải thiện loại A và B, đồng thời đưa ra chiến lược phù hợp để tận dụng tiềm năng hoặc tái cấu trúc loại C.



Hình 14: Biểu đồ tròn

### 4.3. Tổng kết

Thông qua **Sales Forecasting Dashboard** giúp cung cấp cái nhìn tổng quan về doanh số, các yếu tố ảnh hưởng và hiệu suất theo từng cửa hàng. Nó cho thấy sự chênh lệch rõ rệt giữa các cửa hàng, tầm quan trọng của kỳ nghỉ lễ và nhiệt độ lý tưởng trong việc thúc đẩy doanh số. Doanh nghiệp có thể sử dụng thông tin này để tối ưu hóa chiến lược kinh doanh, cải thiện hiệu suất và nâng cao hiệu quả hoạt động.

## 5. Đưa ra đề xuất và chiến lược kinh doanh

Trong bối cảnh kinh doanh nhiều biến động, các chiến lược được xây dựng từ phân tích dữ liệu, bao gồm tối ưu hóa giảm giá, khai thác tiềm năng theo mùa, nâng cao hiệu suất đơn vị kinh doanh và phát triển marketing nhắm mục tiêu, đóng vai trò then chốt trong việc gia tăng doanh thu và kiểm soát chi phí.

**Tối ưu hóa chiến lược giảm giá (Markdowns)**

* **Cơ sở đề xuất chiến lược**: Markdowns có tác động đáng kể đến doanh thu trong các dịp lễ và các giai đoạn mua sắm cao điểm.
* **Cách thực hiện**:

+ Tăng cường markdowns vào các dịp lễ lớn và quảng bá mạnh mẽ qua các kênh truyền thông.

+ Lập kế hoạch giảm giá chính xác và điều chỉnh mức giảm phù hợp với từng nhóm sản phẩm và khách hàng.

+ Kết hợp markdowns với các chương trình khuyến mãi khác như tặng quà hoặc miễn phí vận chuyển để nâng cao giá trị đơn hàng.

* **Kết quả kỳ vọng**: Doanh thu tăng mạnh trong các dịp lễ, chi phí khuyến mãi được tối ưu hóa, và tồn kho không tiêu thụ được giảm.

**Chiến lược tối ưu hóa hiệu suất theo mùa**

* **Cơ sở đề xuất chiến lược**: Doanh thu thay đổi theo mùa, đặc biệt trong các tháng cao điểm và thấp điểm.
* **Cách thực hiện**:

+ Tổ chức các chiến dịch khuyến mãi mạnh vào các tháng thấp điểm (như tháng 6, 8, 12).

+ Đảm bảo cung cấp đủ nguồn lực và hàng hóa cho các tháng cao điểm (tháng 4, 5, 10).

+ Tối ưu hóa chiến dịch marketing, sử dụng công cụ phân tích hiệu quả để cải thiện chiến lược quảng cáo.

* **Kết quả kỳ vọng**: Doanh thu ổn định trong các tháng thấp điểm và cao điểm, gia tăng sự trung thành của khách hàng.

**Chiến lược tối ưu hóa các đơn vị kinh doanh**

* **Cơ sở đề xuất chiến lược**: Dựa trên hiệu suất hoạt động của các đơn vị kinh doanh (cửa hàng, phòng ban), nhu cầu thị trường và khả năng khai thác các sản phẩm chủ lực.
* **Cách thực hiện**:

+ Đầu tư vào các cửa hàng có hiệu suất cao, đồng thời hỗ trợ các cửa hàng yếu kém bằng các chiến dịch marketing và khuyến mãi.

+ Điều chỉnh tồn kho phù hợp với nhu cầu của từng khu vực.

+ Quảng bá mạnh mẽ các sản phẩm chủ lực để tối đa hóa lợi nhuận.

* **Kết quả kỳ vọng**: Doanh thu tăng trưởng, chi phí tồn kho giảm, khách hàng hài lòng và tỷ lệ quay lại cao.

**Chiến lược phát triển marketing nhắm mục tiêu**

* **Cơ sở đề xuất chiến lược**: Các cửa hàng có doanh thu thấp ổn định, trong khi các cửa hàng hàng đầu có sự biến động mạnh, cho thấy cần triển khai chiến dịch marketing riêng biệt cho từng khu vực.
* **Cách thực hiện**:

+ Tập trung vào các cửa hàng yếu kém, thực hiện marketing địa phương và phân tích nhu cầu khách hàng để cung cấp sản phẩm phù hợp.

+ Sử dụng các kênh marketing truyền thống và online, nhắm mục tiêu đúng đối tượng.

+ Cung cấp ưu đãi đặc biệt cho từng cửa hàng để tăng cường doanh thu.

* **Kết quả kỳ vọng**: Tăng trưởng doanh thu tại các cửa hàng yếu, nâng cao nhận diện thương hiệu và thu hút khách hàng mới.

**Chiến lược tối ưu hóa doanh thu qua quản lý mùa vụ, tác động kinh tế và các giai đoạn cao điểm**

* **Cơ sở đề xuất chiến lược**: Chiến lược này dựa trên các yếu tố như biến động nhu cầu theo mùa, tác động của các yếu tố kinh tế như giá nhiên liệu, CPI, và tỷ lệ thất nghiệp. Mặc dù biểu đồ và ma trận tương quan cho thấy một mối tương quan thấp giữa giá nhiên liệu và doanh số, nhưng có sự tương quan chặt chẽ giữa giá nhiên liệu với CPI và tỷ lệ thất nghiệp. Và trong thực tế khi kết hợp các yếu tố này lại thì vẫn có khả năng tăng doanh thu.
* **Cách thực hiện**:

**+ Theo mùa**: Tăng cường quảng bá các sản phẩm đặc trưng cho từng mùa, điều chỉnh mức tồn kho để tránh tình trạng thiếu hoặc thừa hàng.

**+ Ứng phó với biến động kinh tế**: Khi giá nhiên liệu giảm hoặc CPI thấp, tăng cường nhập kho và điều chỉnh giá sản phẩm để kích cầu. Ngược lại, khi giá nhiên liệu cao hoặc CPI tăng, cần tối ưu hóa quy trình vận chuyển và sản xuất để giảm chi phí.

**+ Giai đoạn cao điểm**: Chuẩn bị cho các mùa mua sắm quan trọng (Giáng Sinh, Super Bowl, Lễ Tạ ơn, v.v.), kết hợp các chương trình giảm giá để thu hút khách hàng.

* **Kết quả kỳ vọng**: Doanh thu tăng trưởng, chi phí được tối ưu hóa, lợi nhuận duy trì ổn định qua các giai đoạn biến động.

**Tổng kết về đề xuất và chiến lược kinh doanh**

Các chiến lược trên tập trung vào việc tối ưu hóa doanh thu và chi phí, tối ưu hóa marketing, sản phẩm và dịch vụ khách hàng để duy trì sự tăng trưởng bền vững. Kết quả kỳ vọng bao gồm tăng trưởng doanh thu ổn định, giảm chi phí, cải thiện trải nghiệm khách hàng và tạo ra lợi thế cạnh tranh lâu dài.

# KẾT LUẬN

Bài phân tích tập trung vào dự báo doanh số bán lẻ và xem xét các yếu tố ảnh hưởng đến doanh thu từ các góc độ kinh doanh, kinh tế vĩ mô và xu hướng mùa vụ. Dữ liệu từ 45 cửa hàng được làm sạch và xử lý để đảm bảo chất lượng, sau đó được phân tích bằng các mô hình mô tả, thống kê và dự báo như ARIMA, SARIMA, và hồi quy tuyến tính. Kết quả phân tích làm sáng tỏ mối liên hệ giữa doanh thu và các yếu tố nội tại (giảm giá, loại cửa hàng) cũng như ngoại vi (thời tiết, kinh tế).

Những phát hiện này tạo nền tảng cho các chiến lược tối ưu hóa, bao gồm điều chỉnh chương trình giảm giá theo mùa, tận dụng tiềm năng doanh thu trong các kỳ lễ lớn, cải thiện hiệu suất của từng cửa hàng và đẩy mạnh các chiến dịch marketing có định hướng. Bài phân tích không chỉ mang lại cái nhìn sâu sắc về xu hướng và động lực kinh doanh mà còn cung cấp các giải pháp thực tiễn giúp doanh nghiệp tối ưu hóa chi phí, nâng cao trải nghiệm khách hàng và xây dựng lợi thế cạnh tranh bền vững. Đây là một lộ trình rõ ràng để thúc đẩy tăng trưởng doanh thu và hiệu quả hoạt động trong dài hạn.

# TÀI LIỆU THAM KHẢO

Gupta, S. K. (2023, September 29). *Retail\_Analytics*. Retrieved from Kaggle: https://www.kaggle.com/code/soorajgupta7/retail-analytics

Ouf, M. T. (2023, June 27). *Sales Store Dataset*. Retrieved from Kaggle: https://www.kaggle.com/code/mohamedtahaouf/sales-store-dataset

Rahman, A. (2024, July 28). *Ritel Sales Report*. Retrieved from Kaggle: https://www.kaggle.com/code/aoeraman/ritel-sales-report

Samaddar, A. (2024, January 23). *Retail Data Analytics*. Retrieved from Kaggle: https://www.kaggle.com/code/arkayansamaddar/retail-data-analytics#Predicting-future-sales

Singh, M. (2017, September 01). *Retail Data Analytics*. Retrieved from Kaggle: https://www.kaggle.com/datasets/manjeetsingh/retaildataset

*Tối ưu giảm giá sản phẩm markdown giúp tăng 10% lợi nhuận*. (2022, March 18). Retrieved from FPT Digital: https://digital.fpt.com/linh-vuc/giam-gia-san-pham-tang-loi-nhuan.html

Tutorials, D. (2022, December 08). *Tableau Dashboard from start to end (Part 1) | HR dashboard | Beginner to pro | Tableau project*. Retrieved from YouTube: https://www.youtube.com/watch?v=oAIubTqg-Kw

Tutorials, D. (2022, December 09). *Tableau Dashboard from Start to End (Part 2)| HR Dashboard | Beginner to Pro | Tableau Project*. Retrieved from Youtube: https://www.youtube.com/watch?v=oTyCZVnNVZA&t=1731s