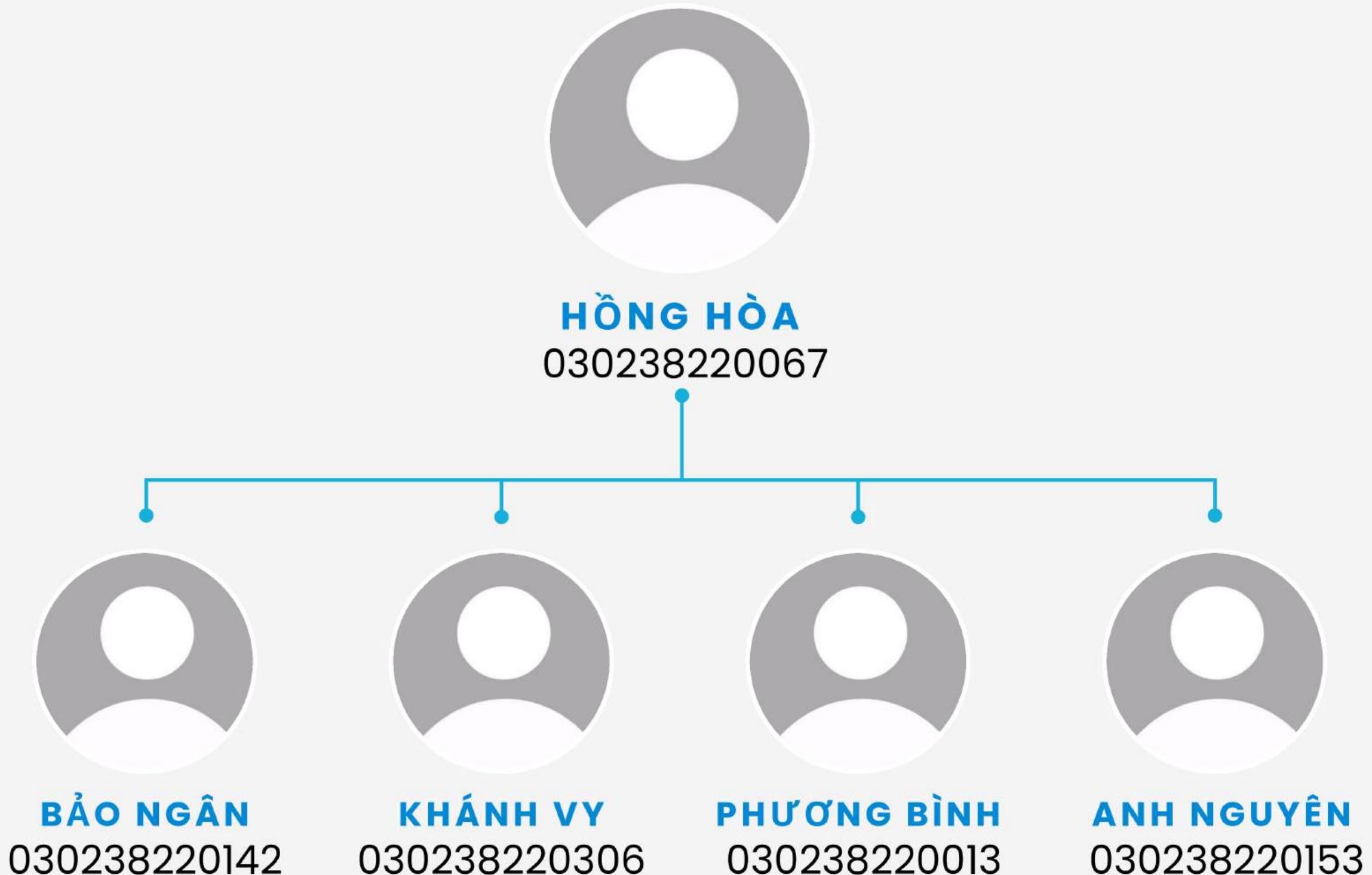


NHÓM 6

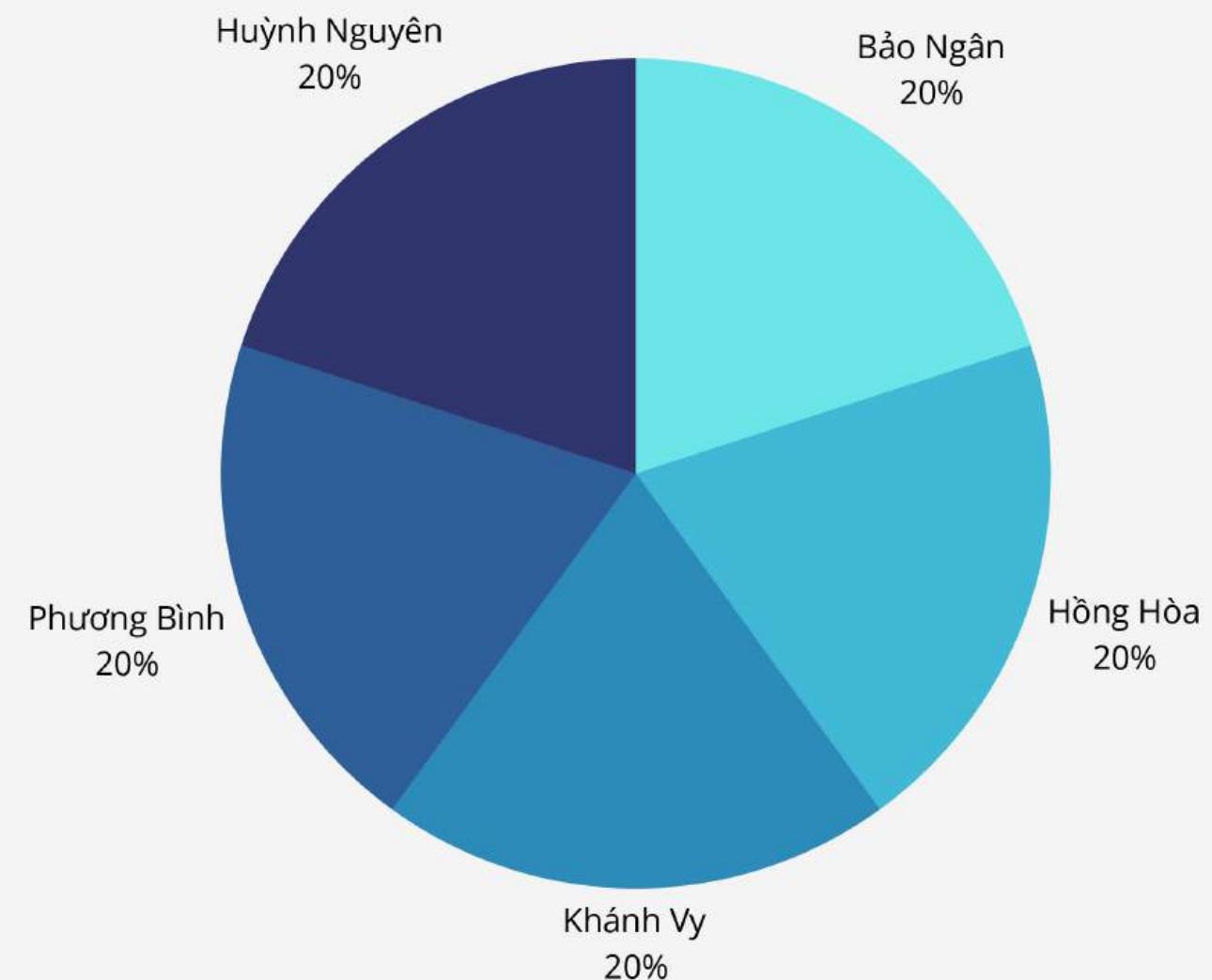
DỰ ĐOÁN DOANH SỐ BÁN HÀNG (SALES FORECASTING)



NHÓM 6



Thành viên - MSSV



Mức độ đóng góp

NỘI DUNG

- 1 CHỌN LĨNH VỰC KINH DOANH VÀ XÁC ĐỊNH VẤN ĐỀ CẦN GIẢI QUYẾT**
- 2 THU THẬP VÀ LÀM SẠCH DỮ LIỆU**
- 3 PHÂN TÍCH DỮ LIỆU VÀ MÔ HÌNH HÓA**
- 4 TRỰC QUAN HÓA DỮ LIỆU**
- 5 ĐỀ XUẤT VÀ CHIẾN LƯỢC KINH DOANH**

1. CHỌN LĨNH VỰC KINH DOANH VÀ XÁC ĐỊNH VẤN ĐỀ CẦN GIẢI QUYẾT

LĨNH VỰC KINH DOANH:

LĨNH VỰC BÁN LẺ

Mục tiêu giải quyết một vấn đề kinh doanh cụ thể thông qua việc phân tích bộ dữ liệu **Retail Data Analytics**

1. CHỌN LĨNH VỰC KINH DOANH VÀ XÁC ĐỊNH VẤN ĐỀ CẦN GIẢI QUYẾT

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

Chương trình giảm giá

Các yếu tố bên ngoài

Tuần lễ đặc biệt

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, giá nhiên liệu, chỉ số giá tiêu dùng, và tỷ lệ thất nghiệp.

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. CHỌN LĨNH VỰC KINH DOANH VÀ XÁC ĐỊNH VĂN ĐỀ CẦN GIẢI QUYẾT

XÁC ĐỊNH VĂN ĐỀ KINH DOANH CẦN GIẢI QUYẾT

- 1 **Dự đoán doanh số bán hàng**
- 2 **Tác động của các chương trình khuyến mãi trong các dịp lễ**
- 3 **Ảnh hưởng của mùa vụ đến doanh thu bán hàng**
- 4 **Ảnh hưởng của các yếu tố kinh tế vĩ mô như giá nhiên liệu , tỷ lệ thất nghiệp , chỉ số giá tiêu dùng ảnh hưởng đến hành vi tiêu dùng và doanh thu.**
- 5 **Phân tích hiệu suất theo khu vực và loại cửa hàng**

1. CHỌN LĨNH VỰC KINH DOANH VÀ XÁC ĐỊNH VĂN ĐỀ CẦN GIẢI QUYẾT

PHẠM VI PHÂN TÍCH

Gồm ba bảng chính: Stores, Features và Sales

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:

- Chương trình giảm giá
- Sự kiện đặc biệt
- Các yếu tố bên ngoài: nhiệt độ, giá nhiên liệu, chỉ số giá tiêu dùng, tỷ lệ thất nghiệp

1. CHỌN LĨNH VỰC KINH DOANH VÀ XÁC ĐỊNH VĂN ĐỀ CẦN GIẢI QUYẾT

MỤC TIÊU CẦN ĐẠT

- Dự đoán doanh thu bán hàng trong các tuần hoặc tháng tiếp theo.
- 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. CHỌN LĨNH VỰC KINH DOANH VÀ XÁC ĐỊNH VẤN ĐỀ CẦN GIẢI QUYẾT

CÁC CÂU HỎI NGHIÊN CỨU MÀ BI CÓ THỂ TRẢ LỜI

- Doanh thu bán hàng trong các tuần hoặc tháng tiếp theo có xu hướng tăng hay giảm?
- Các chương trình giảm giá có ảnh hưởng tích cực hay tiêu cực đến doanh thu bán hàng?
- Liệu có sự tương quan giữa giá xăng, tỷ lệ thất nghiệp và doanh thu bán hàng của các cửa hàng?
- Doanh thu có biến động giữa các tháng cao điểm và thấp điểm hay không?
- 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

THU THẬP DỮ LIỆU

Nguồn dữ liệu: <https://www.kaggle.com/datasets/manjeetsingh/retaildataset>

	Store	Date	Temperature	Fuel_Price	MarkDown1	MarkDown2	MarkDown3	MarkDown4	MarkDown5	CPI	Unemployment	IsHoliday
0	1	5/2/2010	42.31	2.572	NaN	NaN	NaN	NaN	NaN	211.096358	8.106	False
1	1	12/2/2010	38.51	2.548	NaN	NaN	NaN	NaN	NaN	211.242170	8.106	True
2	1	19/02/2010	39.93	2.514	NaN	NaN	NaN	NaN	NaN	211.289143	8.106	False
3	1	26/02/2010	46.63	2.561	NaN	NaN	NaN	NaN	NaN	211.319643	8.106	False
4	1	5/3/2010	46.50	2.625	NaN	NaN	NaN	NaN	NaN	211.350143	8.106	False

Features data set

	Store	Dept	Date	Weekly_Sales	IsHoliday
0	1	1	5/2/2010	24924.50	False
1	1	1	12/2/2010	46039.49	True
2	1	1	19/02/2010	41595.55	False
3	1	1	26/02/2010	19403.54	False
4	1	1	5/3/2010	21827.90	False

Sales data set

	Store	Type	Size
0	1	A	151315
1	2	A	202307
2	3	B	37392
3	4	A	205863
4	5	B	34875

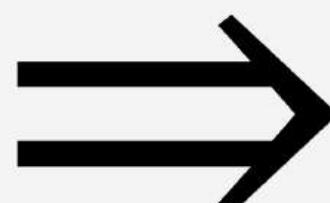
Stores data set

2. THU THẬP VÀ LÀM SẠCH DỮ LIỆU

XỬ LÝ VÀ CHUẨN HÓA DỮ LIỆU

Store	Date	Temperature	Fuel_Price	MarkDown1	MarkDown2	MarkDown3	MarkDown4	MarkDown5	CPI	Unemployment	IsHoliday
0	1	5/2/2010	42.31	2.572	NaN	NaN	NaN	NaN	NaN	211.096358	8.106
1	1	12/2/2010	38.51	2.548	NaN	NaN	NaN	NaN	NaN	211.242170	8.106
2	1	19/02/2010	39.93	2.514	NaN	NaN	NaN	NaN	NaN	211.289143	8.106
3	1	26/02/2010	46.63	2.561	NaN	NaN	NaN	NaN	NaN	211.319643	8.106
4	1	5/3/2010	46.50	2.625	NaN	NaN	NaN	NaN	NaN	211.350143	8.106

Gán giá trị NaN trong các cột Markdown bằng 0



Sử dụng phương pháp nội suy theo thời gian (forward fill) để điền giá trị.

Định dạng ngày tháng: String -> Datetime

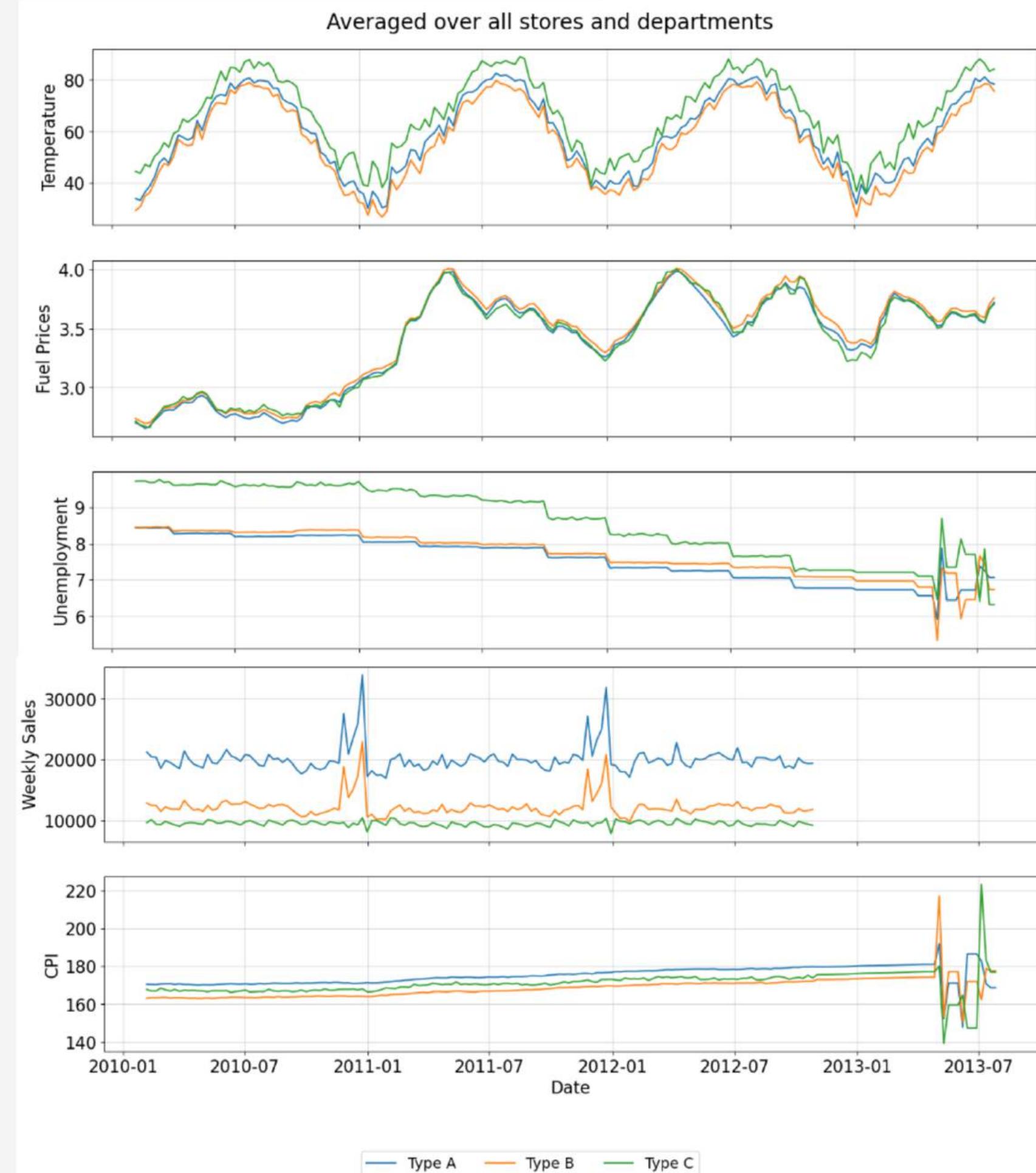
2. THU THẬP VÀ LÀM SẠCH DỮ LIỆU

HỢP NHẤT DỮ LIỆU

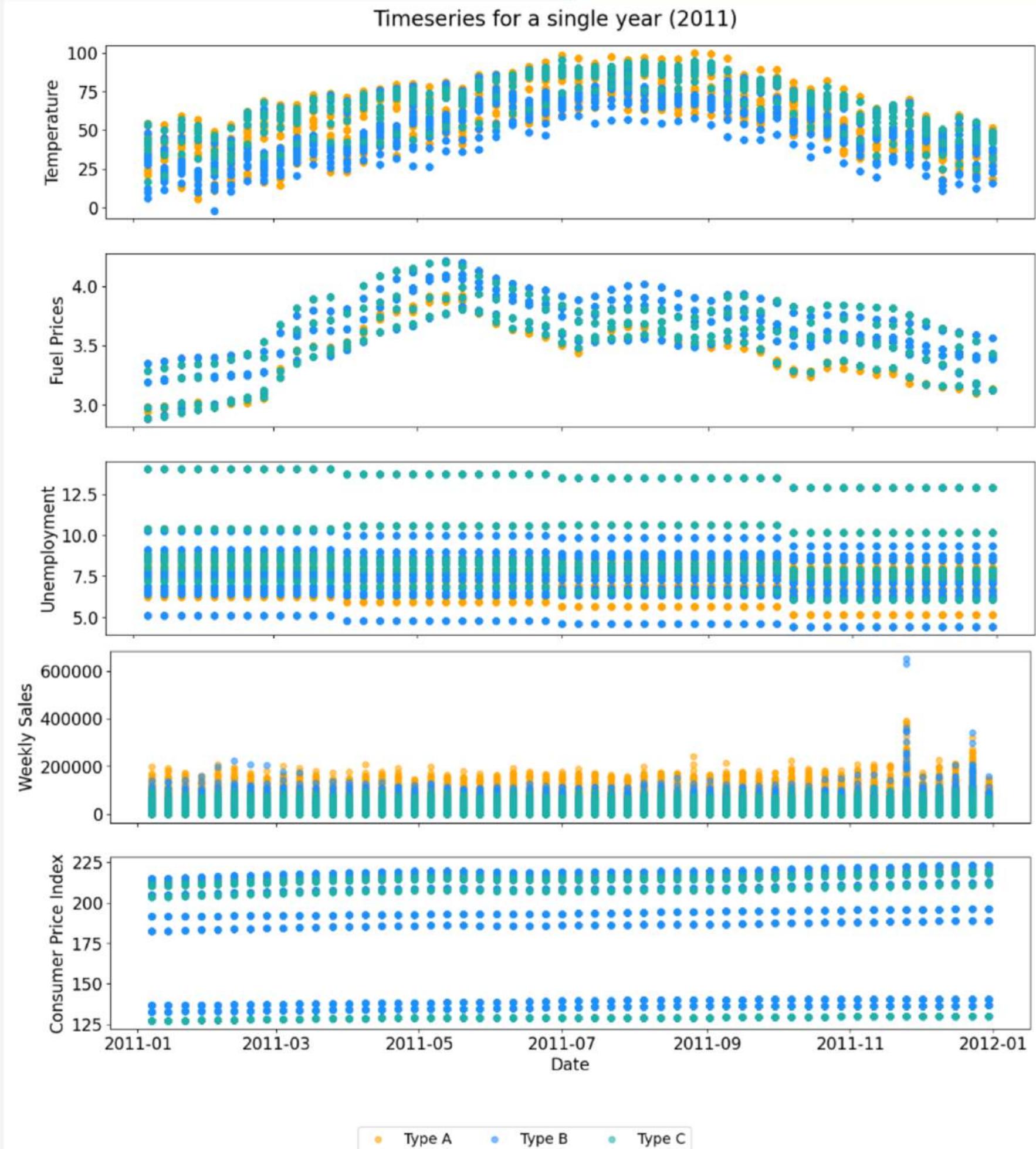
	Store	Date	Temperature	Fuel_Price	MarkDown1	MarkDown2	MarkDown3	MarkDown4	MarkDown5	CPI	Unemployment	IsHoliday	Dept	Weekly_Sales	Type	Size
0	1	2010-02-05	42.31	2.572	0.0	0.0	0.0	0.0	0.0	211.096358	8.106	False	1.0	24924.50	A	151315
1	1	2010-02-05	42.31	2.572	0.0	0.0	0.0	0.0	0.0	211.096358	8.106	False	2.0	50605.27	A	151315
2	1	2010-02-05	42.31	2.572	0.0	0.0	0.0	0.0	0.0	211.096358	8.106	False	3.0	13740.12	A	151315
3	1	2010-02-05	42.31	2.572	0.0	0.0	0.0	0.0	0.0	211.096358	8.106	False	4.0	39954.04	A	151315
4	1	2010-02-05	42.31	2.572	0.0	0.0	0.0	0.0	0.0	211.096358	8.106	False	5.0	32229.38	A	151315

1. Hợp nhất **Features** và **Sales** dựa trên 3 khóa chính
2. Hợp nhất với **Stores**
=> mergedDF là một DataFrame duy nhất

2. THU THẬP VÀ LÀM SẠCH DỮ LIỆU



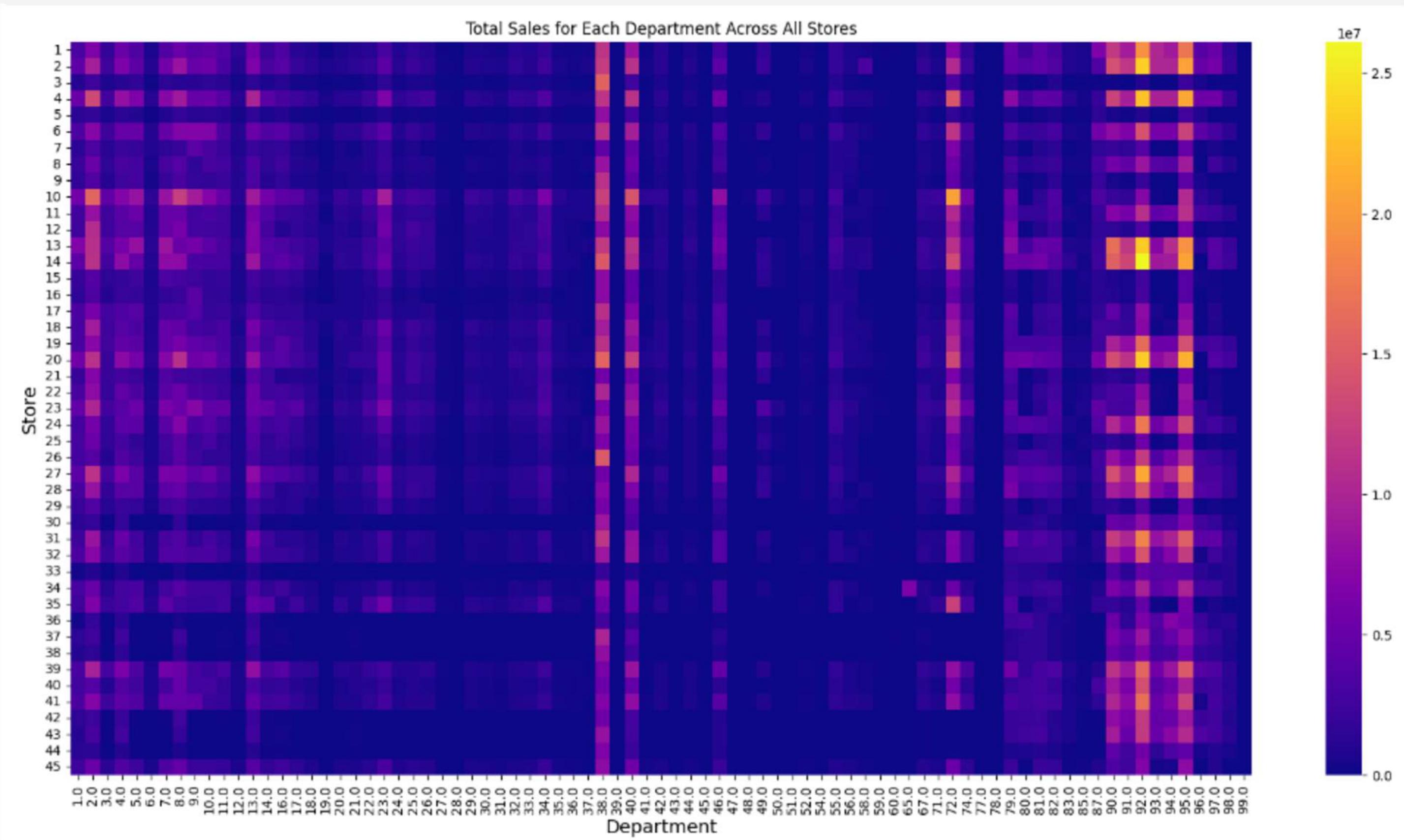
2. THU THẬP VÀ LÀM SẠCH DỮ LIỆU



3. PHÂN TÍCH DỮ LIỆU VÀ MÔ HÌNH HÓA

PHÂN TÍCH DỮ LIỆU

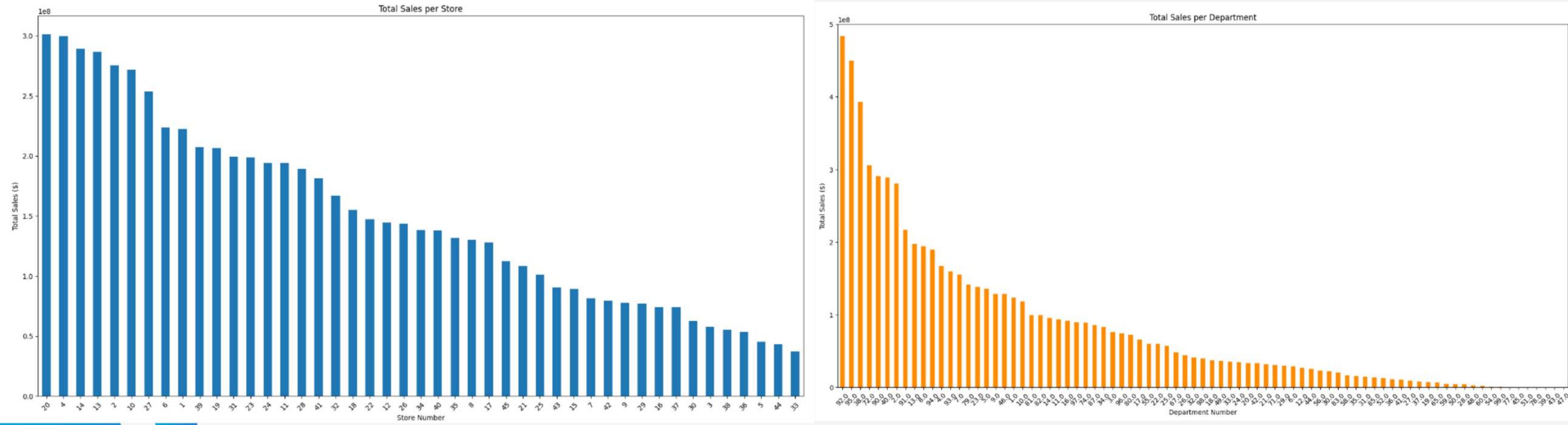
Doanh thu theo từng phòng ban tại mỗi cửa hàng



3. PHÂN TÍCH DỮ LIỆU VÀ MÔ HÌNH HÓA

PHÂN TÍCH DỮ LIỆU

Những cửa hàng nào đang hoạt động tốt nhất tính đến thời điểm hiện tại?



3. PHÂN TÍCH DỮ LIỆU VÀ MÔ HÌNH HÓA

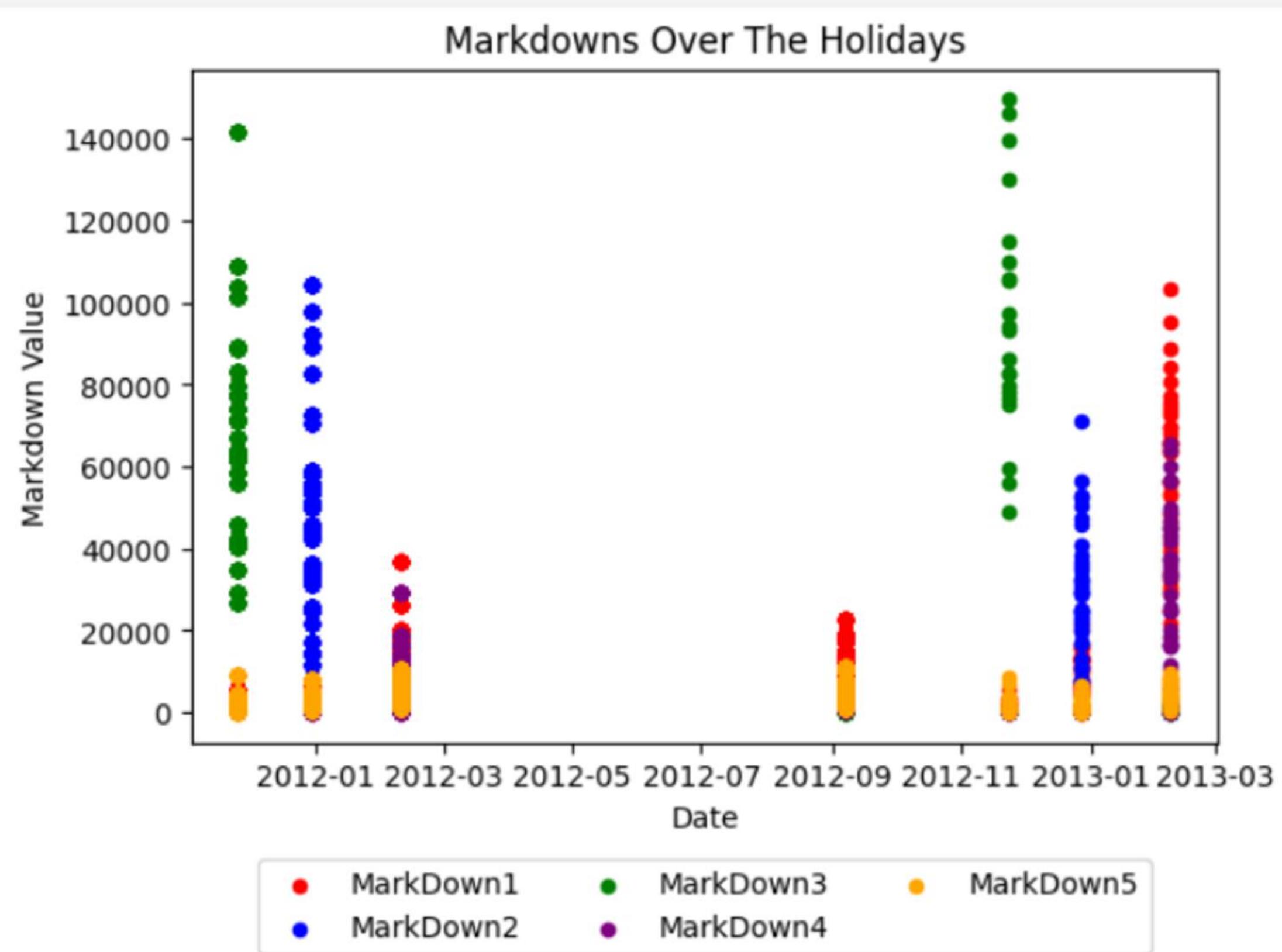
PHÂN TÍCH DỮ LIỆU

Những tác động của việc giảm giá (Markdowns) trong các kỳ nghỉ lễ

	Store	Date	Temperature	Fuel_Price	MarkDown1	MarkDown2	MarkDown3	MarkDown4	MarkDown5	CPI	Unemployment	IsHoliday	Dept	We
73	1	2010-02-12	38.51	2.548	0.00	0.00	0.00	0.00	0.00	211.242170	8.106	True	1.0	
74	1	2010-02-12	38.51	2.548	0.00	0.00	0.00	0.00	0.00	211.242170	8.106	True	2.0	
75	1	2010-02-12	38.51	2.548	0.00	0.00	0.00	0.00	0.00	211.242170	8.106	True	3.0	
76	1	2010-02-12	38.51	2.548	0.00	0.00	0.00	0.00	0.00	211.242170	8.106	True	4.0	
77	1	2010-02-12	38.51	2.548	0.00	0.00	0.00	0.00	0.00	211.242170	8.106	True	5.0	
...
422815	45	2012-09-07	75.70	3.911	11024.45	12.80	52.63	1854.77	2055.70	191.577676	8.684	True	97.0	
422816	45	2012-09-07	75.70	3.911	11024.45	12.80	52.63	1854.77	2055.70	191.577676	8.684	True	98.0	
423289	45	2012-11-23	43.08	3.748	1005.79	0.00	72542.01	484.70	620.12	192.283032	8.667	True	NaN	
423294	45	2012-12-28	35.96	3.563	12751.77	24656.74	230.07	31.46	1654.14	192.559264	8.667	True	NaN	
423300	45	2013-02-08	28.99	3.753	53311.88	531.33	78.26	24823.94	3233.44	192.897089	8.625	True	NaN	

29796 rows × 16 columns

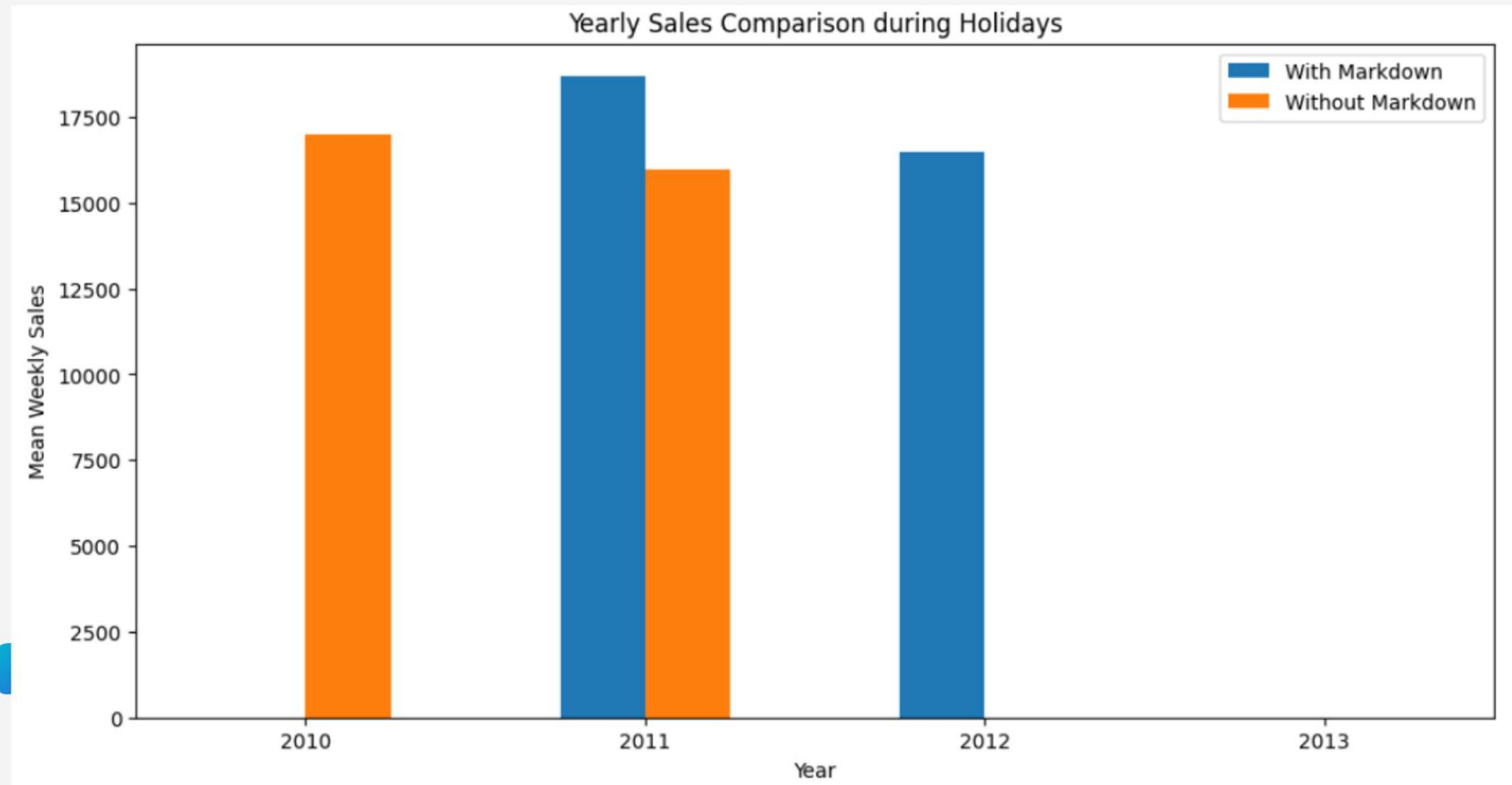
Những tác động của việc giảm giá (Markdowns) trong các kỳ nghỉ lễ



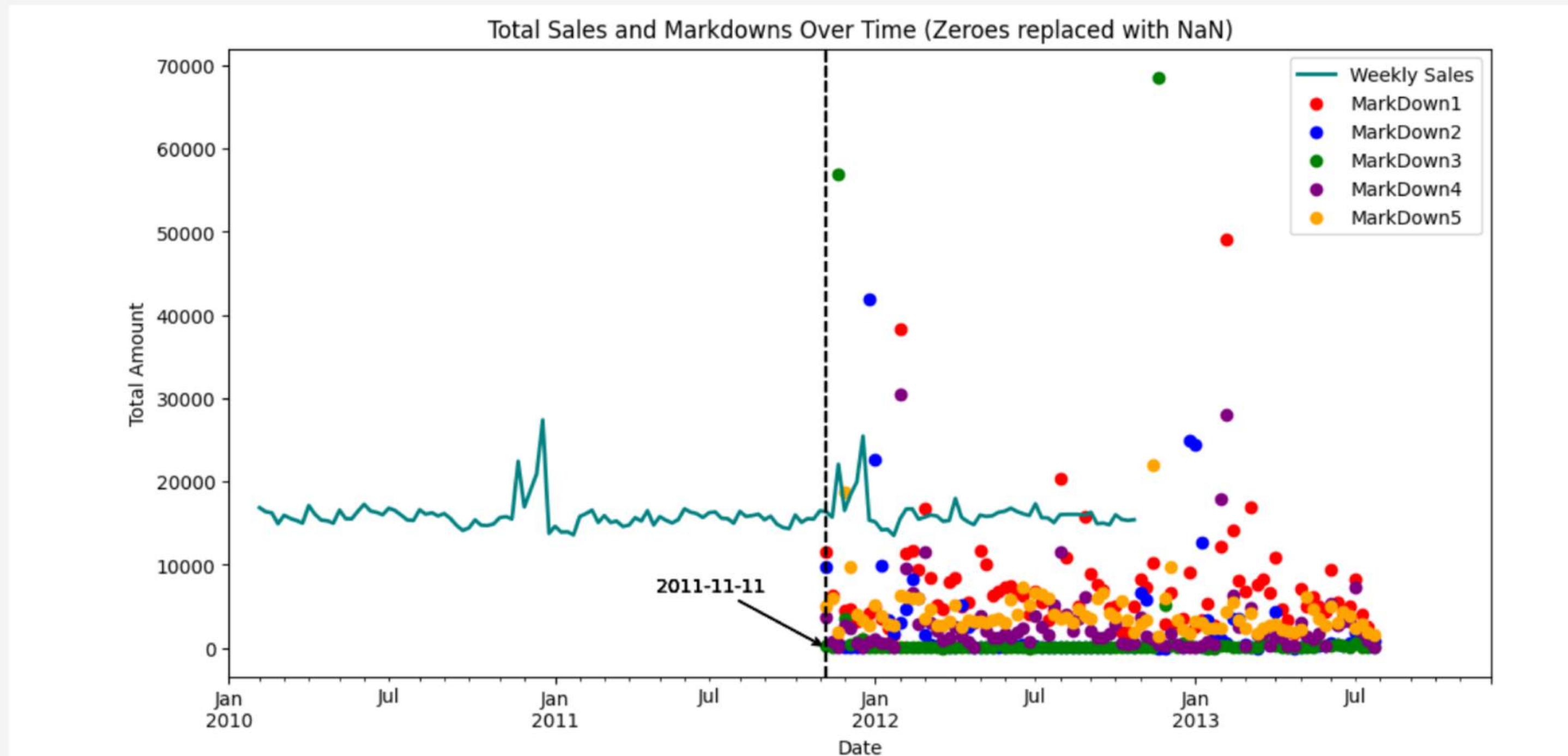
3. PHÂN TÍCH DỮ LIỆU VÀ MÔ HÌNH HÓA

PHÂN TÍCH DỮ LIỆU

Những tác động của việc giảm giá (Markdowns) trong các kỳ nghỉ lễ

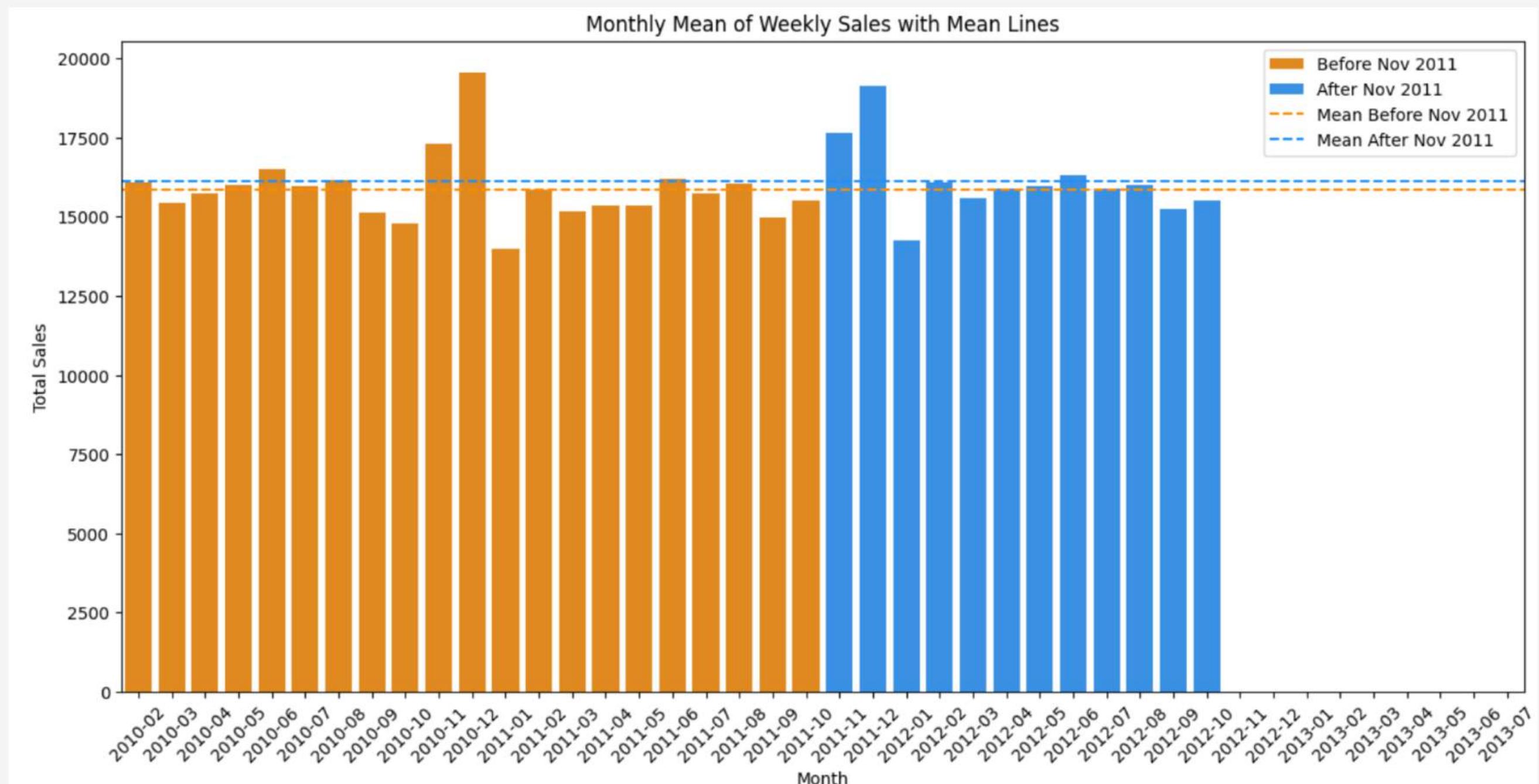


Những tác động của việc giảm giá (Markdowns) trong các kỳ nghỉ lễ



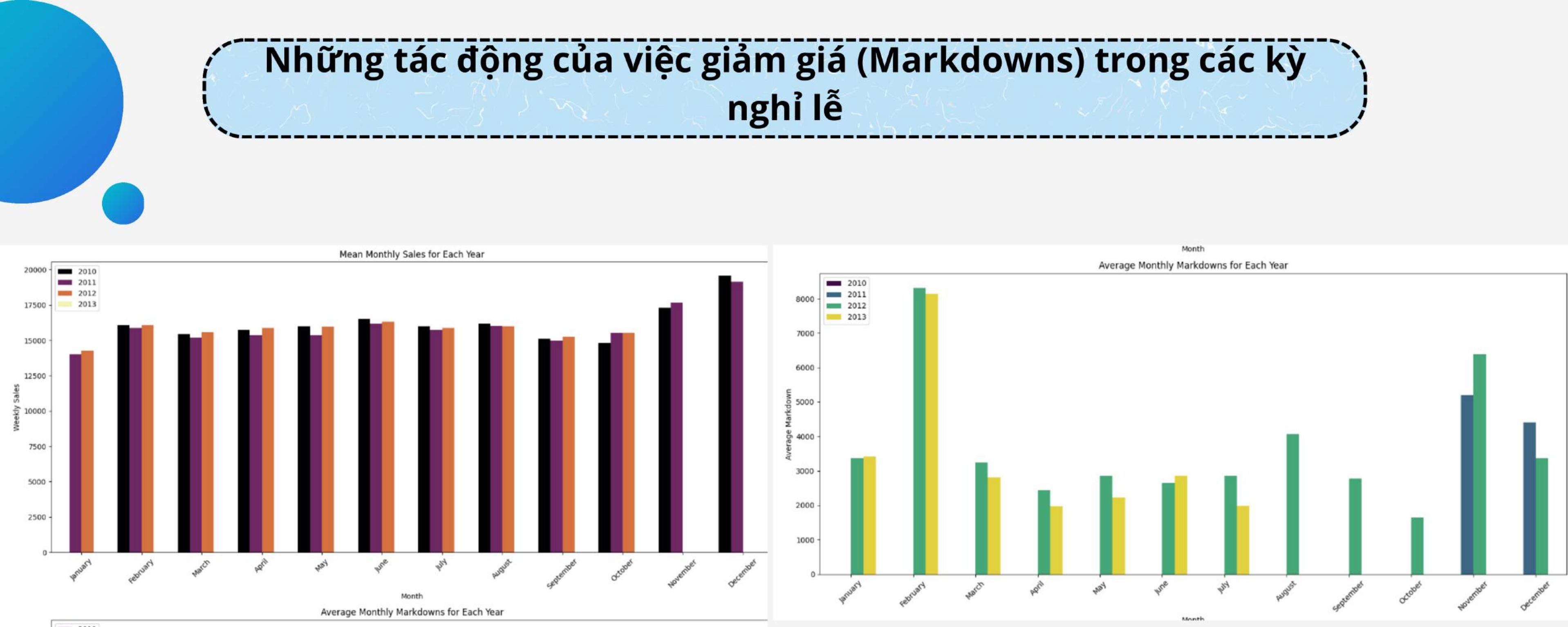
Trước và sau ngày 11-11-2011, doanh số hàng tuần (Weekly Sales) duy trì ổn định, với một vài đỉnh tăng cao vào các dịp lễ hoặc sự kiện đặc biệt. Sau ngày này, các markdown (giảm giá) bắt đầu xuất hiện, nhưng không tạo ra sự tăng trưởng đột biến trong doanh số, mà vẫn duy trì ở mức ổn định hoặc giảm nhẹ ở một số giai đoạn

Những tác động của việc giảm giá (Markdowns) trong các kỳ nghỉ lễ



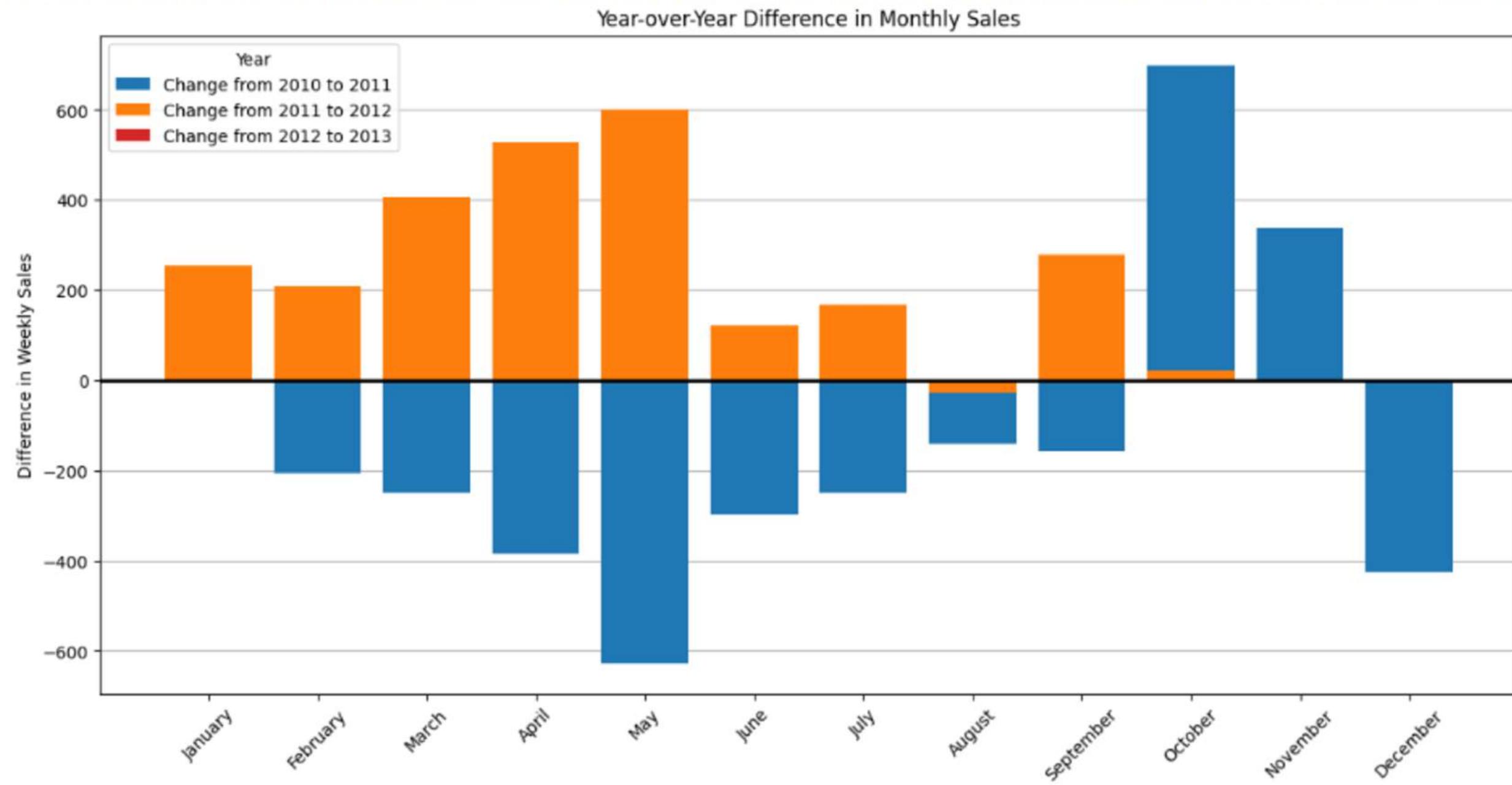
Biểu đồ Monthly Mean of Weekly Sales with Mean Lines cho thấy doanh số trung bình sau tháng 11/2011 (màu xanh) cao hơn so với trước đó (màu cam). Doanh số có sự dao động rõ rệt theo tháng, với các đỉnh cao vào các tháng lễ lớn và giảm vào mùa hè. Sau khi triển khai chiến dịch Markdown, doanh số trung bình hàng tháng đã cải thiện, chứng tỏ hiệu quả tích cực từ các chương trình giảm giá.

Những tác động của việc giảm giá (Markdowns) trong các kỳ nghỉ lễ



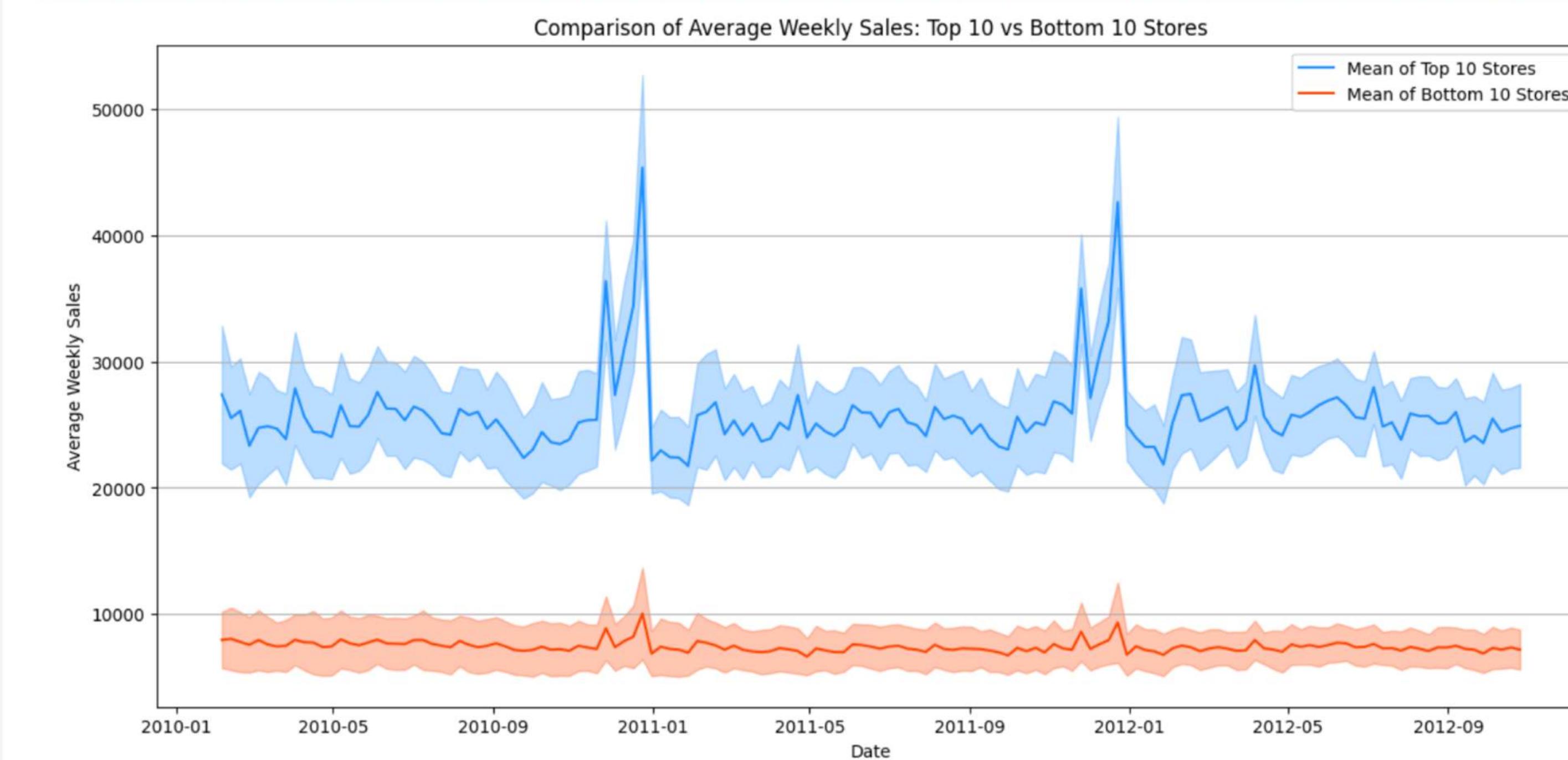
Biểu đồ thanh cho thấy doanh số tăng sau khi áp dụng giảm giá. Biểu đồ đầu tiên nêu bật xu hướng bán hàng tích cực. Biểu đồ thứ hai cho thấy mức giảm giá cao hơn tương quan với mức tăng đột biến của doanh số, cho thấy mức giảm giá góp phần vào sự tăng trưởng doanh số.

Những tác động của việc giảm giá (Markdowns) trong các kỳ nghỉ lễ



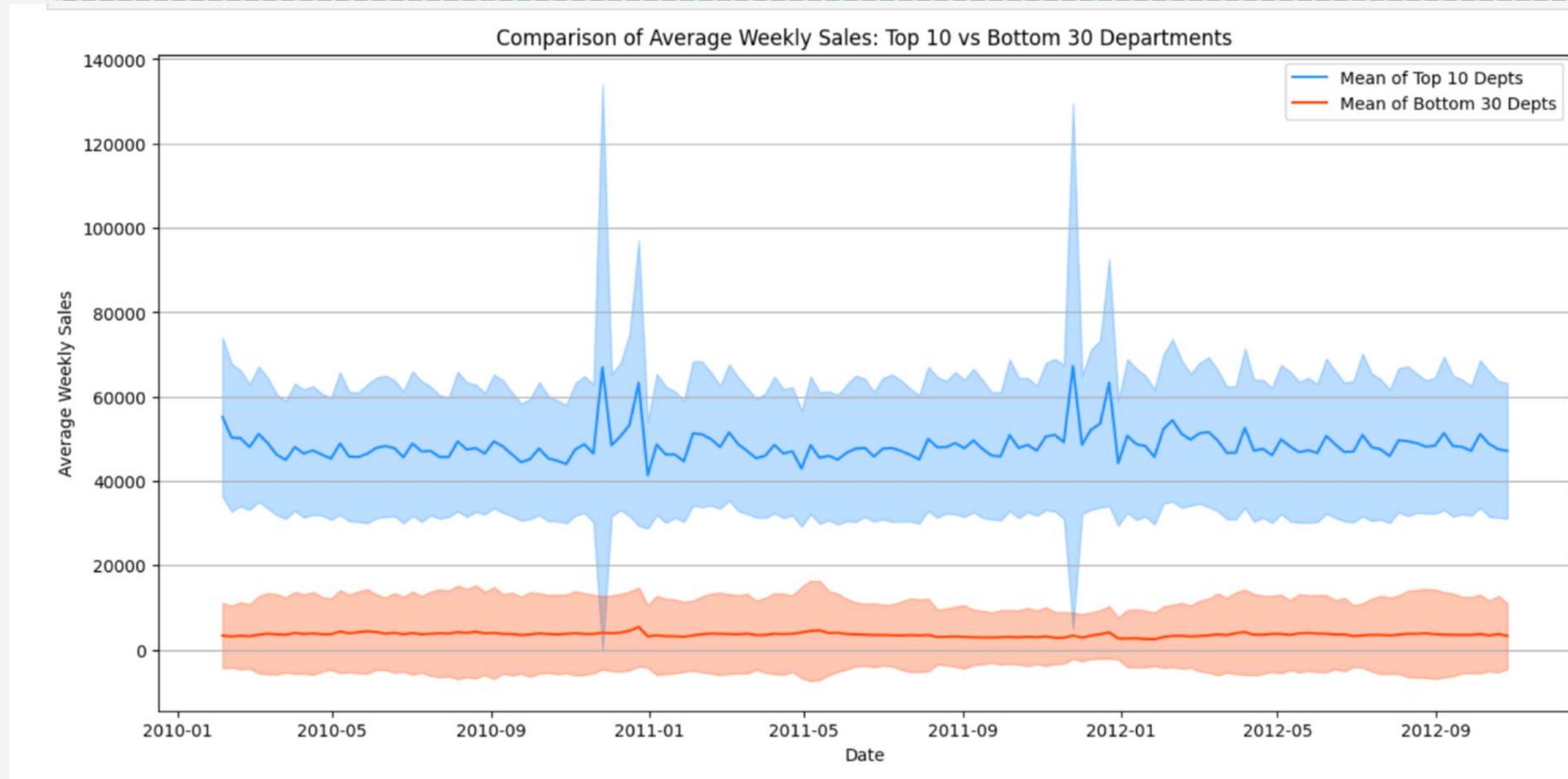
Biểu đồ trên cho thấy từ khi các chương trình giảm giá được giới thiệu vào tháng 11 năm 2011, doanh số đã tăng trưởng ổn định, đặc biệt là từ tháng 1 đến tháng 5 năm 2012. Tuy nhiên, doanh số tháng 12 năm 2011 lại thấp hơn so với tháng 12 năm 2010. Tác động của giảm giá có vẻ rõ rệt hơn trong các tháng không phải kỳ nghỉ, với ảnh hưởng lớn nhất được quan sát trong giai đoạn từ tháng 1 đến tháng 5 năm 2012.

Hiệu suất của các cửa hàng và phòng ban hàng đầu



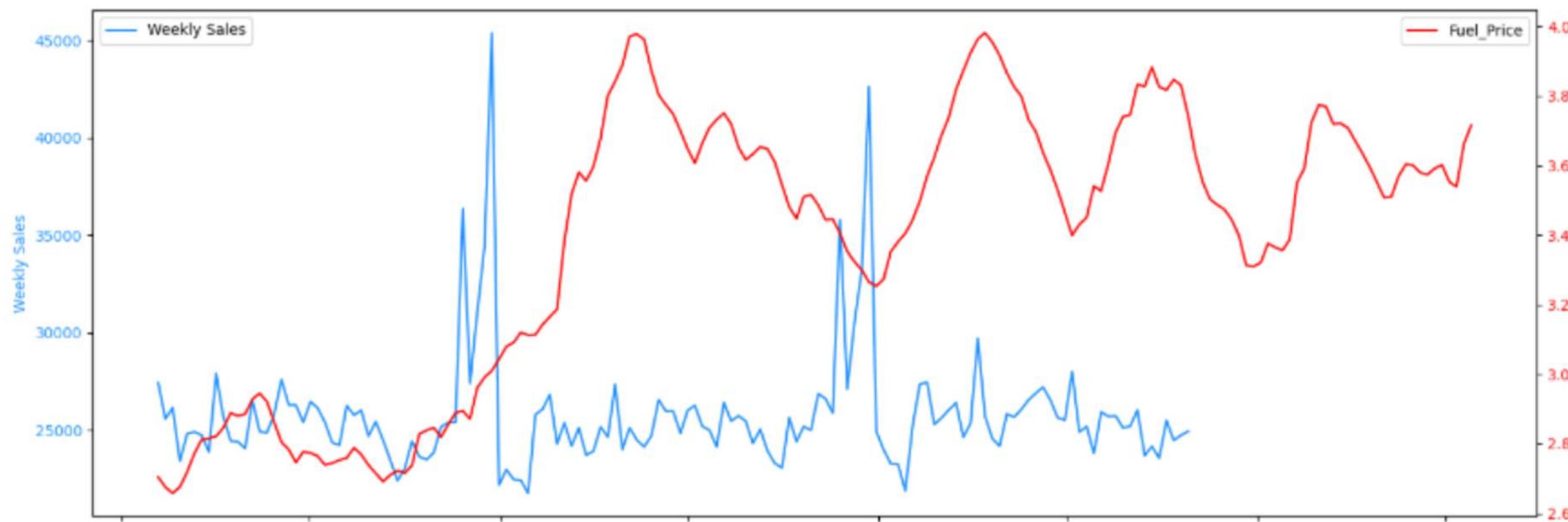
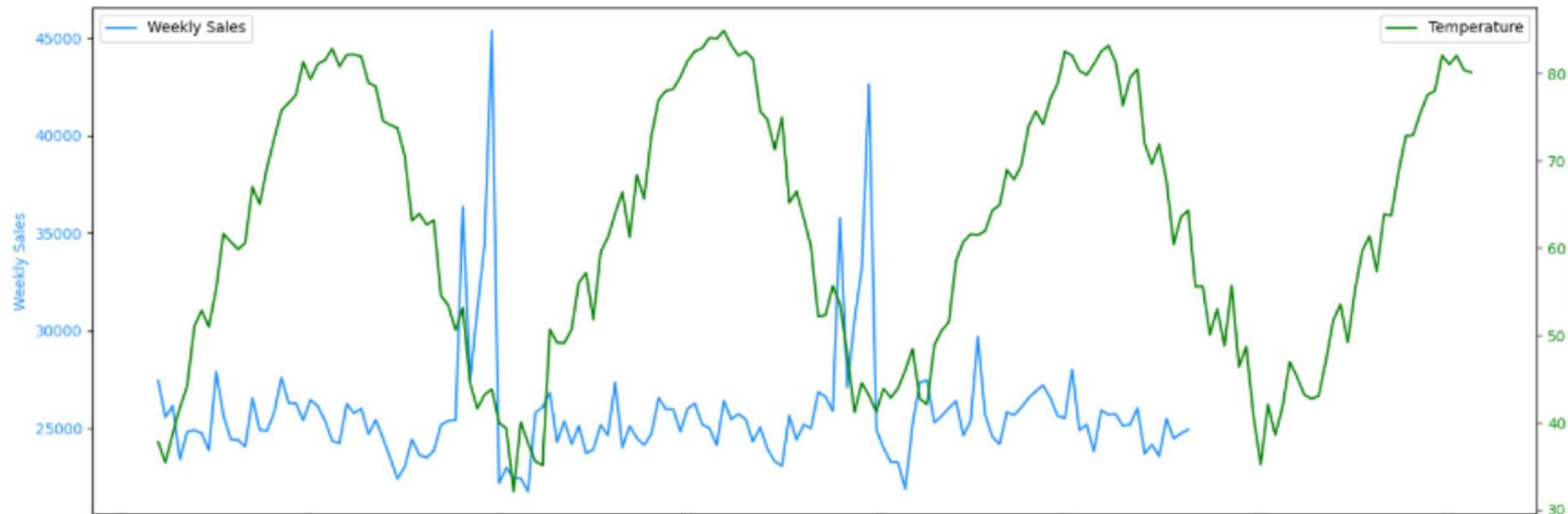
Biểu đồ cho thấy các cửa hàng hàng đầu có doanh số cao và ổn định hơn. Có những đỉnh cao rõ rệt vào khoảng thời gian cuối năm (dịp lễ hội): cuối năm 2010, cuối năm 2011. Xu hướng doanh số vẫn ổn định với mức dao động không quá lớn giữa các thời kỳ ngoài mùa lễ hội. Ngược lại các cửa hàng cuối bảng có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

Hiệu suất của các cửa hàng và phòng ban hàng đầu

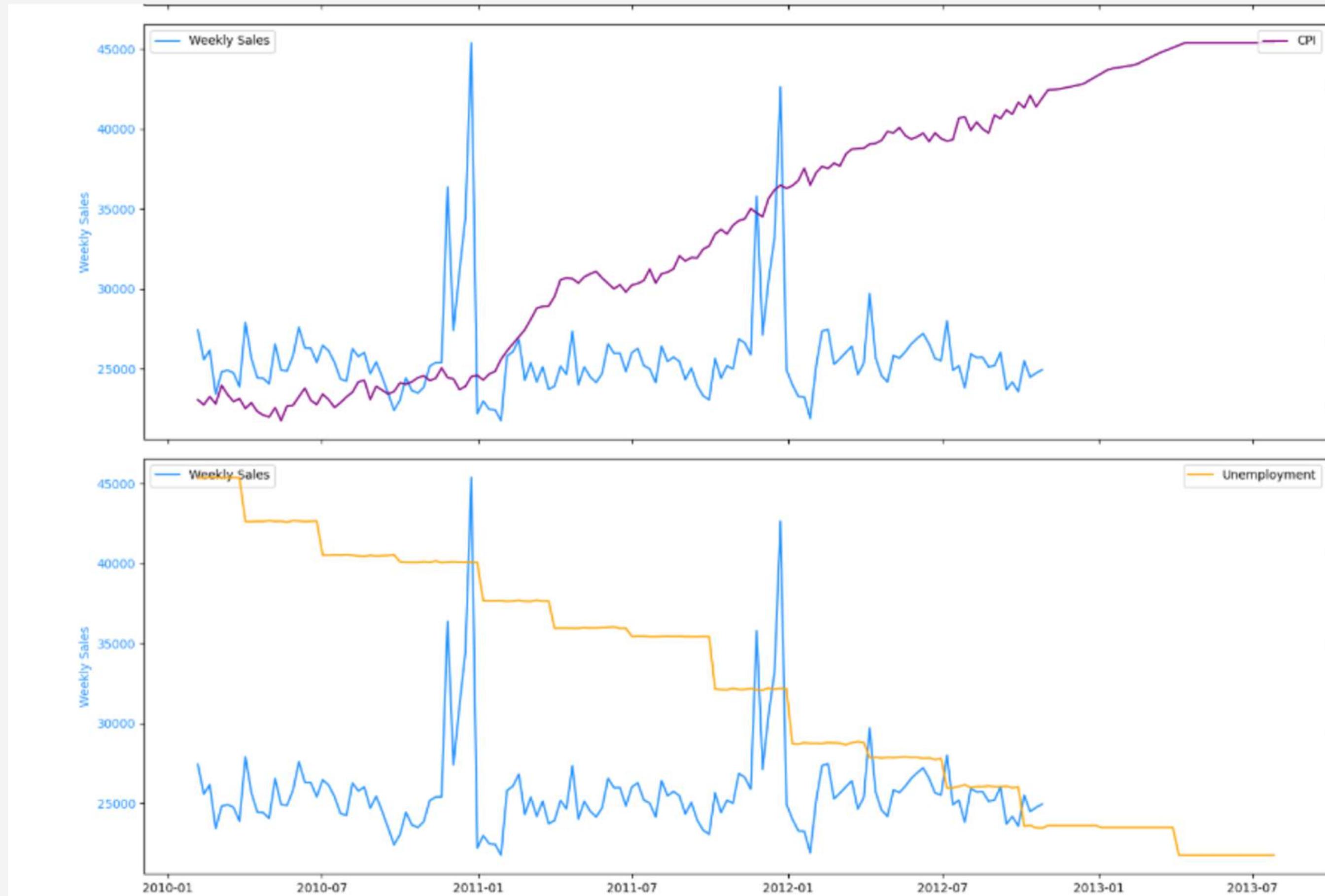


Biểu đồ cho thấy sự khác biệt giữa 10 phòng ban đầu và 30 phòng ban cuối. Các phòng ban đầu có doanh thu cao hơn và biến động lớn, trong khi các phòng ban cuối có doanh thu thấp và ổn định. Điều này phản ánh hiệu suất kinh doanh không giống nhau giữa các phòng ban trên và các phòng ban dưới.

Nhận xét 10 cửa hàng hàng đầu và những yếu tố ảnh hưởng đến hiệu suất của chúng

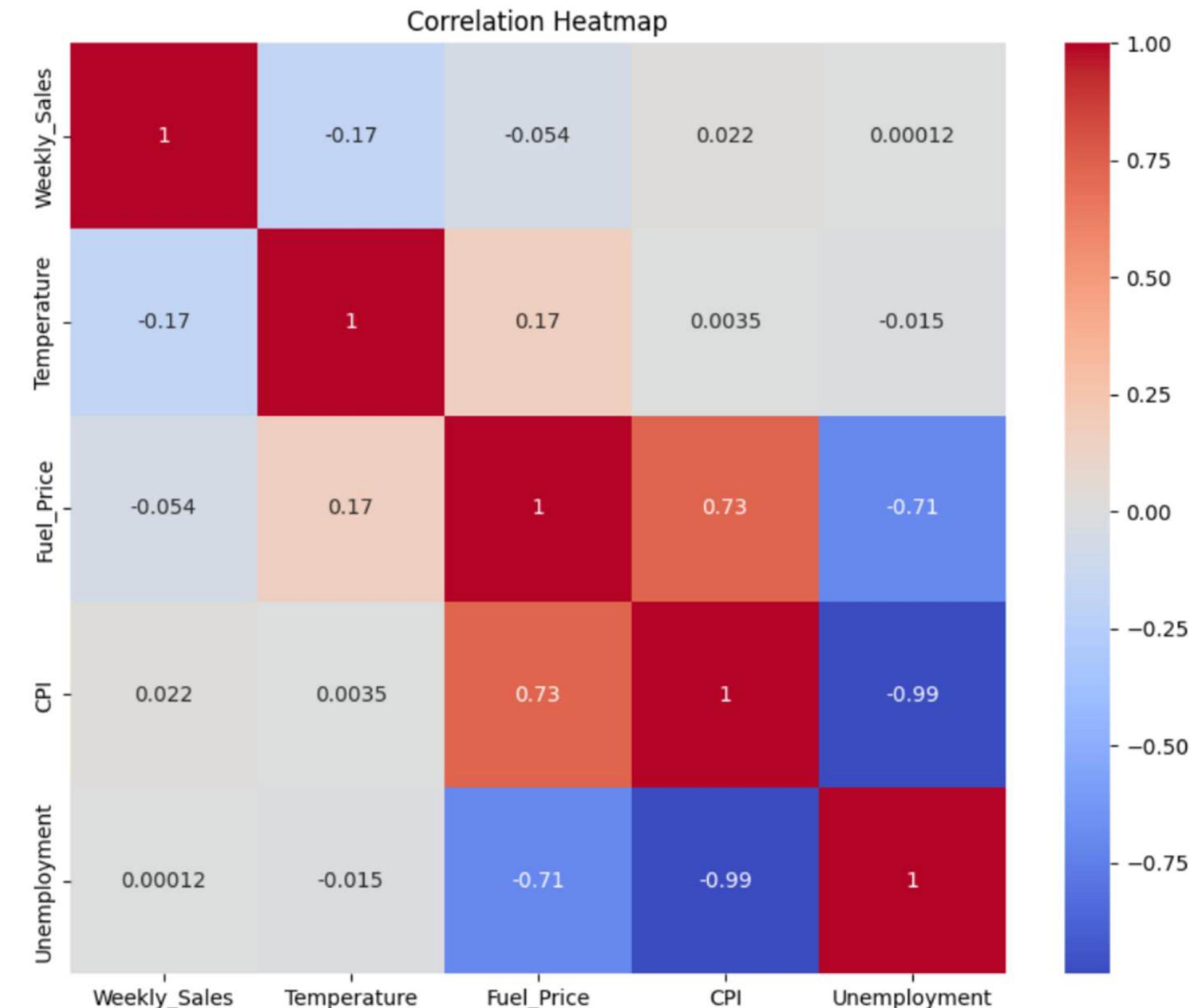


Nhận xét 10 cửa hàng hàng đầu và những yếu tố ảnh hưởng đến hiệu suất của chúng

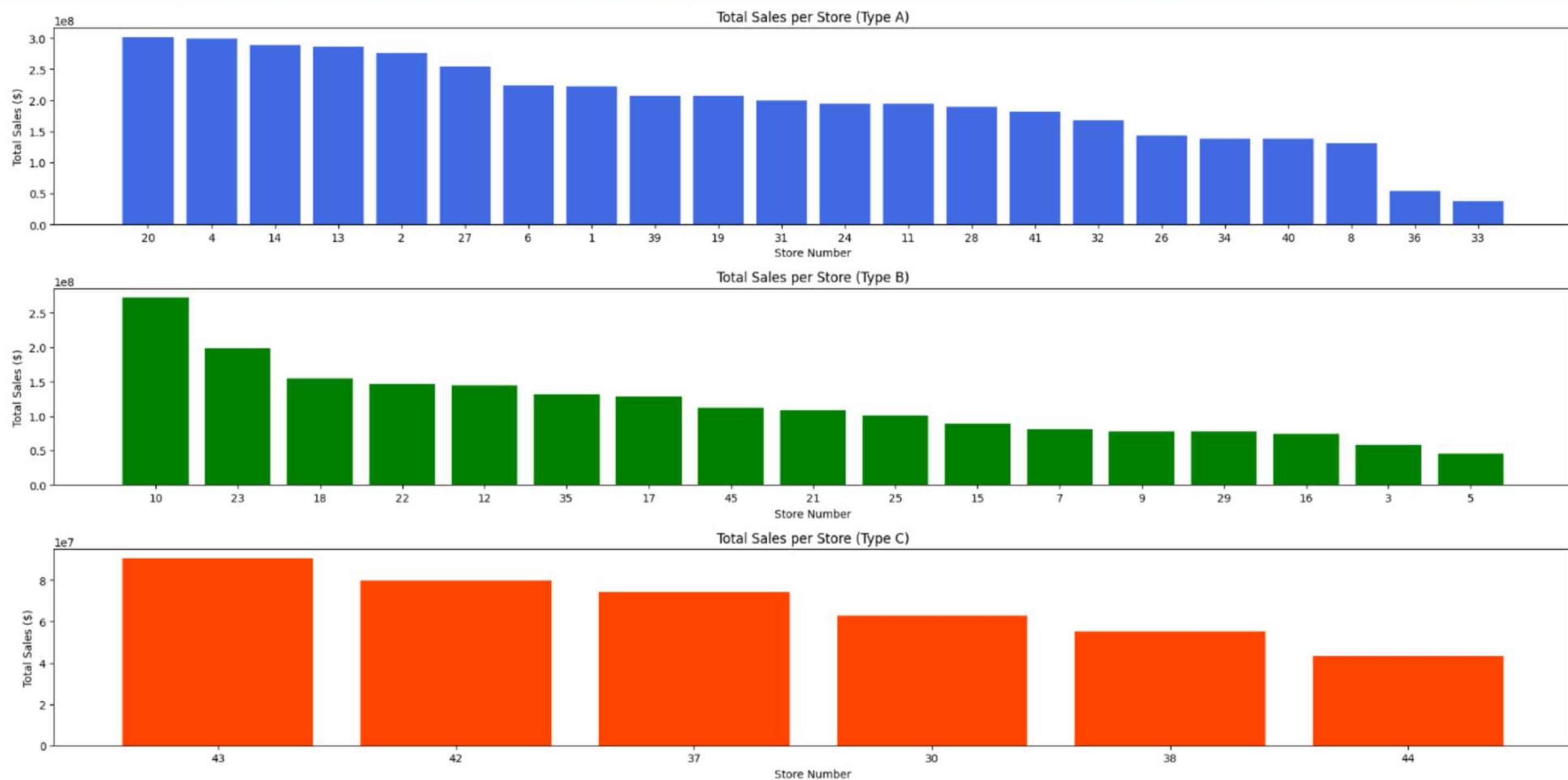


Nhận xét 10 cửa hàng hàng đầu và những yếu tố ảnh hưởng đến hiệu suất của chúng

- Doanh số không bị ảnh hưởng trực tiếp bởi các yếu tố như nhiệt độ, giá nhiên liệu, CPI.
- 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.



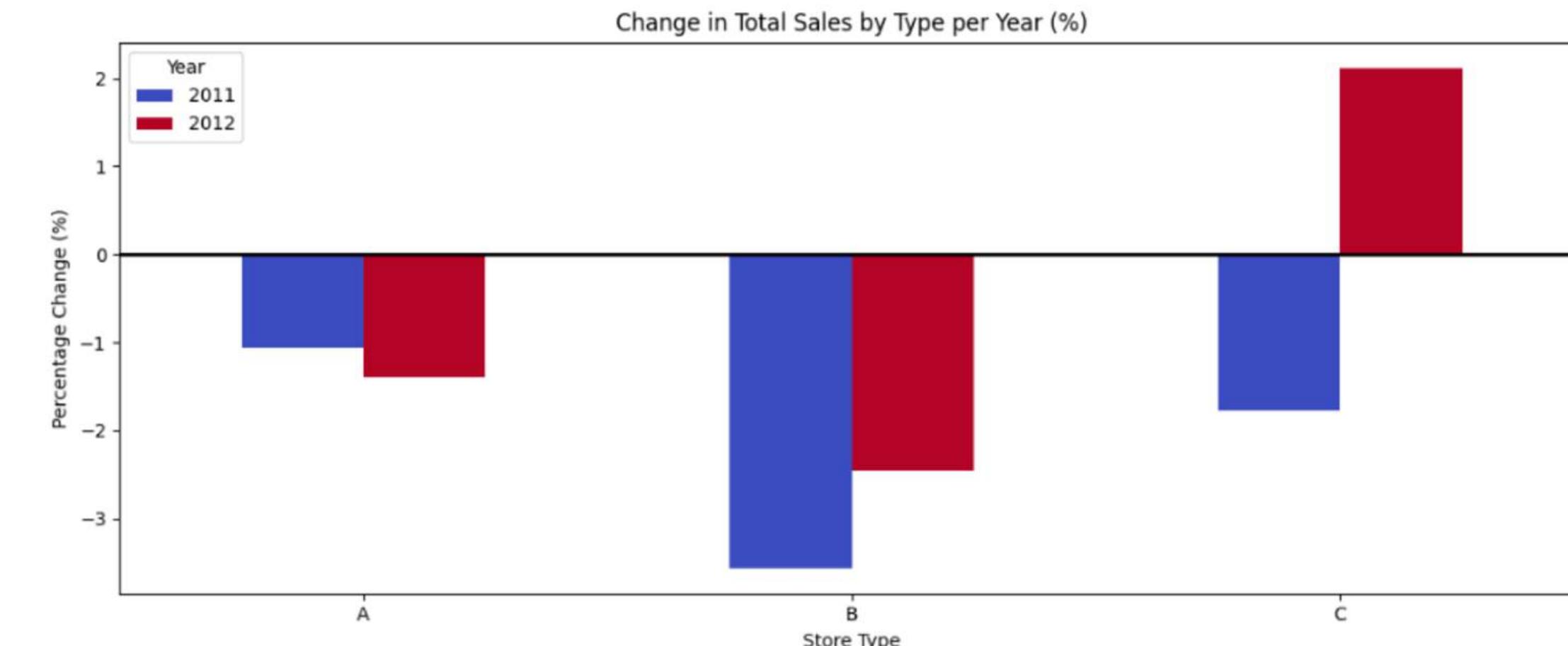
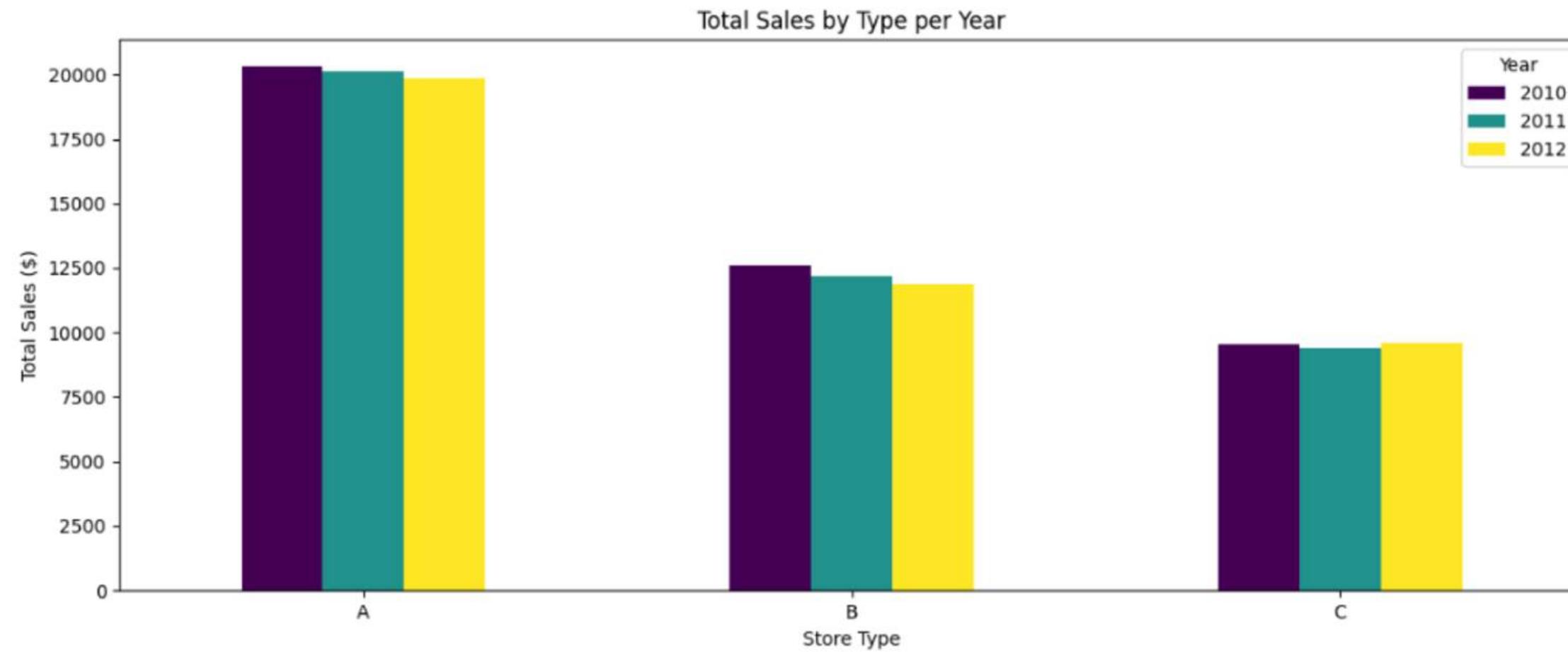
Nhận xét 10 cửa hàng hàng đầu và những yếu tố ảnh hưởng đến hiệu suất của chúng



Biểu đồ thể hiện doanh thu theo từng cửa hàng cho các loại A, B và C, sắp xếp theo doanh thu giảm dần. Cửa hàng loại A có doanh thu cao nhất, với các cửa hàng như số 20, 4, 14 và 13 vượt mốc 300 triệu USD, trong khi cửa hàng số 36 và 33 có doanh thu dưới 100 triệu USD. Cửa hàng loại B có sự phân hóa rõ rệt, với cửa hàng số 10 dẫn đầu trên 250 triệu USD và cửa hàng số 5 dưới 50 triệu USD. Nhóm Type C có doanh thu thấp, đồng đều hơn, với cửa hàng số 43 đạt 80 triệu USD và cửa hàng số 44 đạt 60 triệu USD.

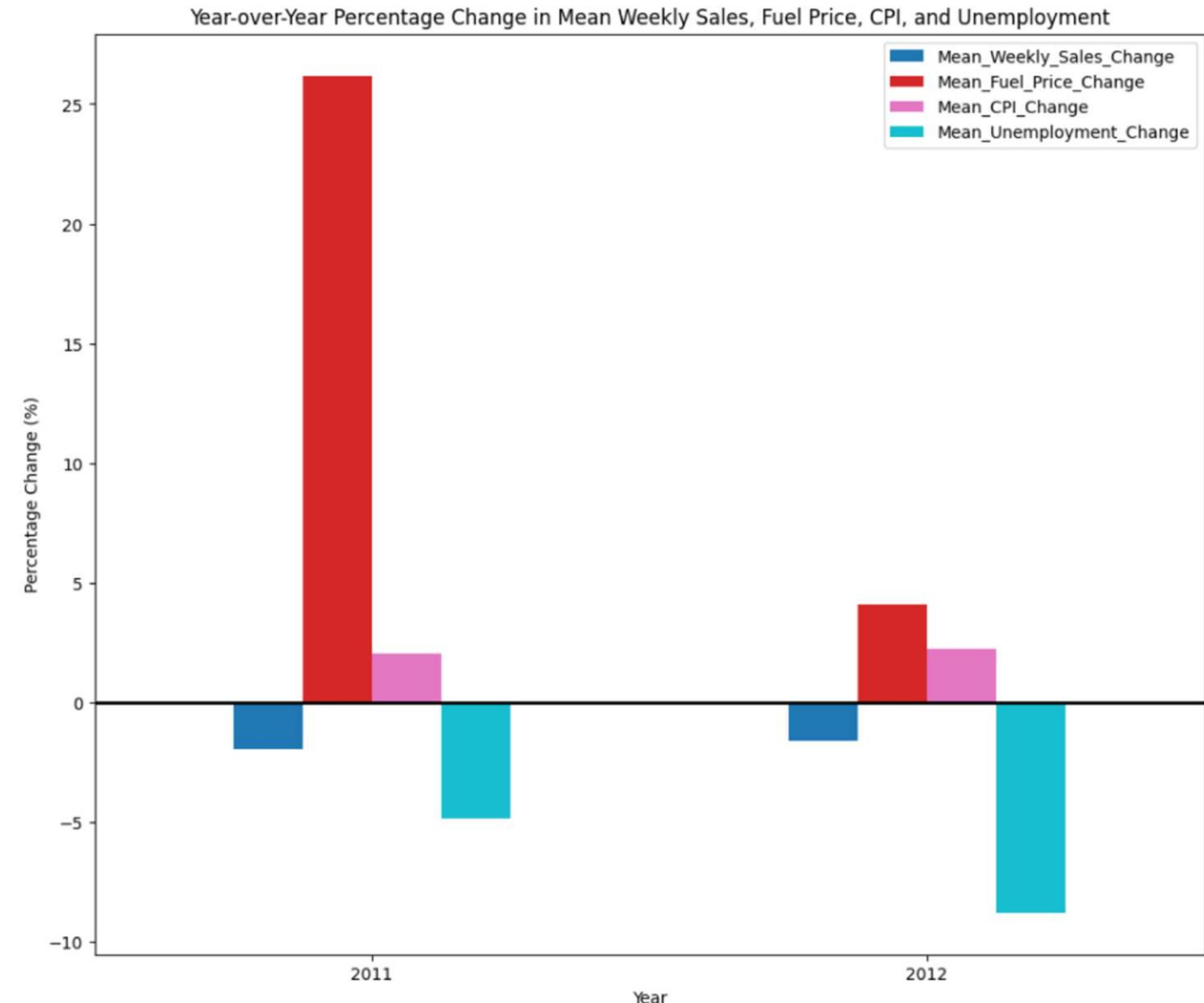
Nhận xét 10 cửa hàng hàng đầu và những yếu tố ảnh hưởng đến hiệu suất của chúng

- Qua biểu đồ trên có thể thấy được rằng doanh thu có chiều hướng giảm qua các năm.
- Các cửa hàng loại A có doanh thu cao nhất mỗi năm, tiếp theo là các cửa hàng loại B và C.
- Ngoài ra, các cửa hàng loại A và B đã trải qua sự sụt giảm doanh thu, trong khi cửa hàng loại C là nhóm duy nhất có có sự gia tăng doanh thu 2% vào năm 2012, cho thấy tiềm năng phát triển nếu được đầu tư đúng hướng.



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

- Các yếu tố như **sự tăng mạnh giá nhiên liệu** năm 2011 có thể làm **thay đổi hành vi mua sắm**, trong khi **sự giảm mạnh của tỷ lệ thất nghiệp** hoặc **CPI** có thể thúc đẩy tiêu dùng nhờ khả năng chi tiêu cao hơn.



3. PHÂN TÍCH DỮ LIỆU VÀ MÔ HÌNH HÓA

3.2. MÔ HÌNH HÓA

Phân Tích Chu Kỳ và Xu Hướng Doanh Số Bán Hàng

Phân Tích và Dự Báo Doanh Số Hàng Tháng

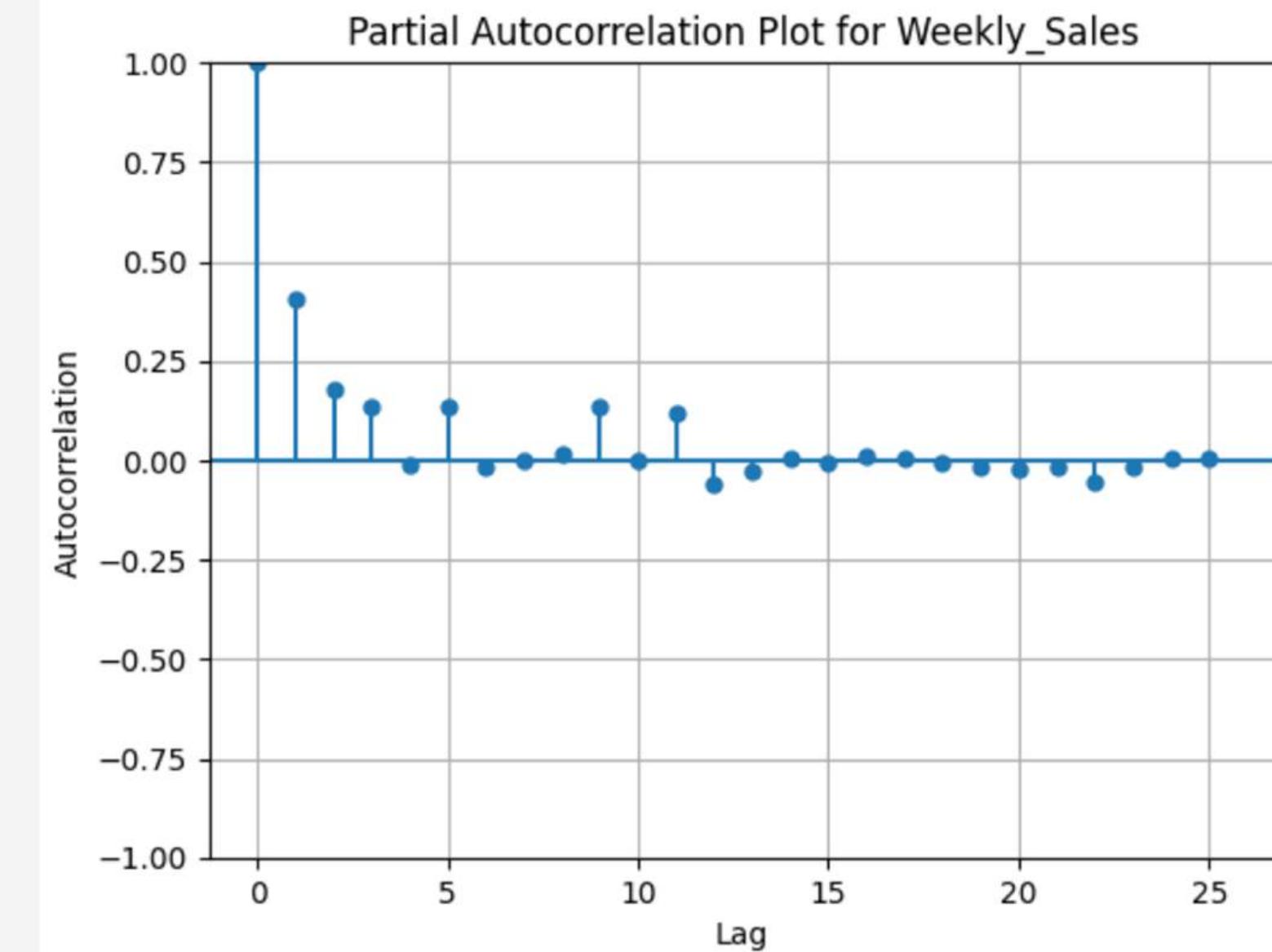
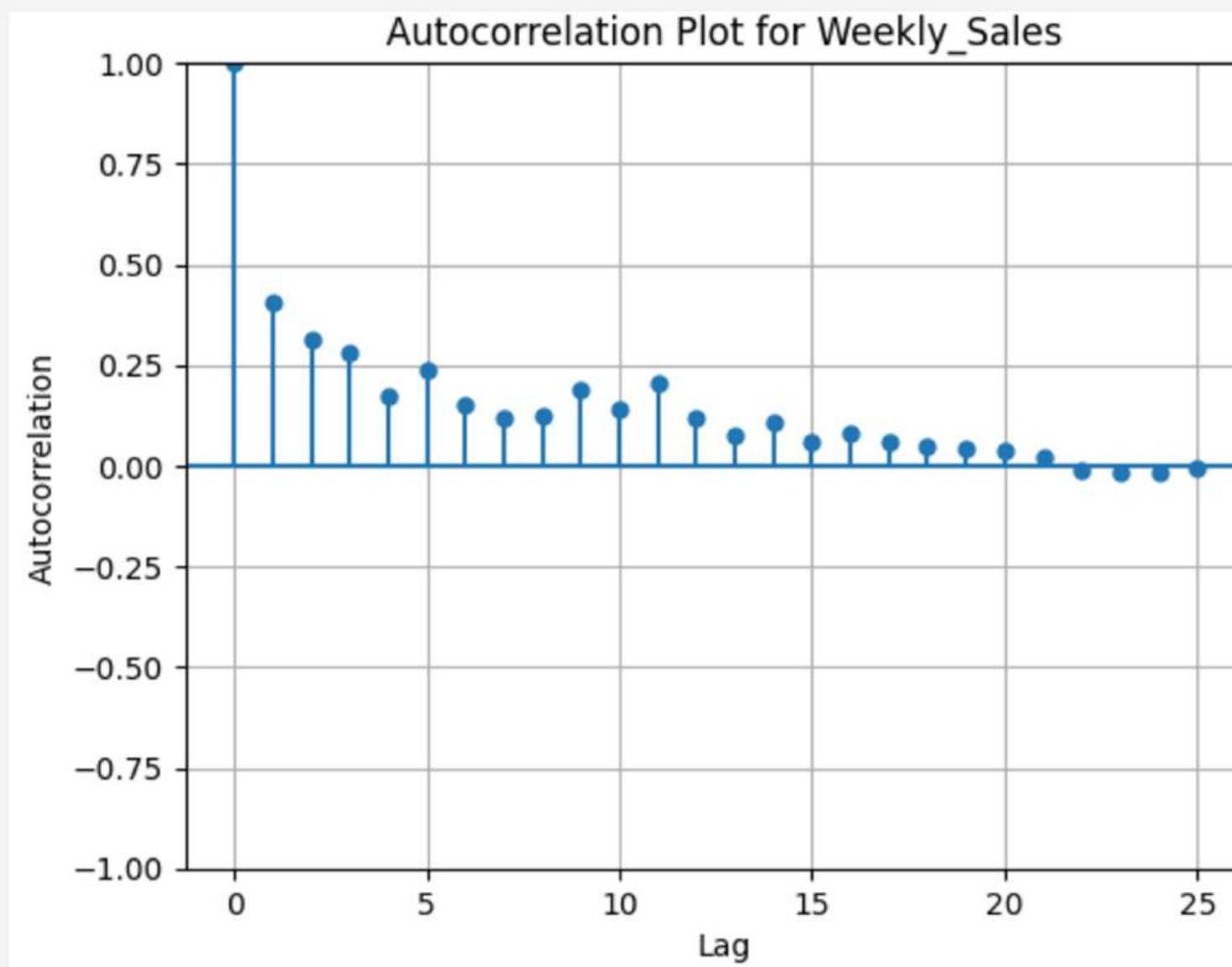
Áp Dụng ARIMA cho Doanh Số Bán Hàng

Áp Dụng ARIMA và SARIMA cho Dự Báo Doanh Số

3. PHÂN TÍCH DỮ LIỆU VÀ MÔ HÌNH HÓA

Phân Tích Chu Kỳ và Xu Hướng Doanh Số Bán Hàng

3.2. MÔ HÌNH HÓA



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.

3. PHÂN TÍCH DỮ LIỆU VÀ MÔ HÌNH HÓA

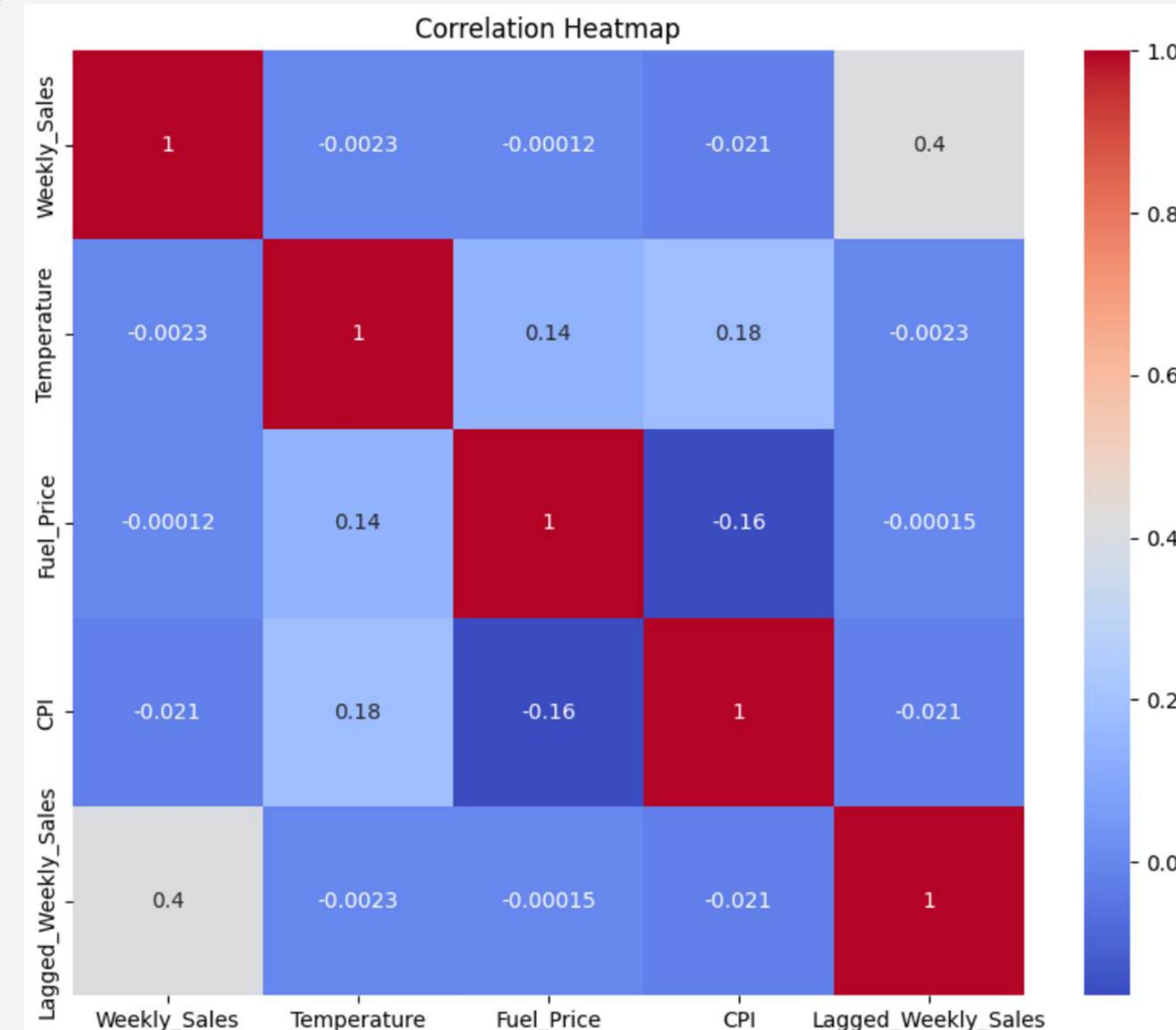
Phân Tích Chu Kỳ và Xu Hướng Doanh Số Bán Hàng

Lagged_Weekly_Sales, chứa giá trị doanh số của ngày trước đó (lag = 1), xác định mức độ phụ thuộc ngắn hạn của doanh số

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.

3.2. MÔ HÌNH HÓA



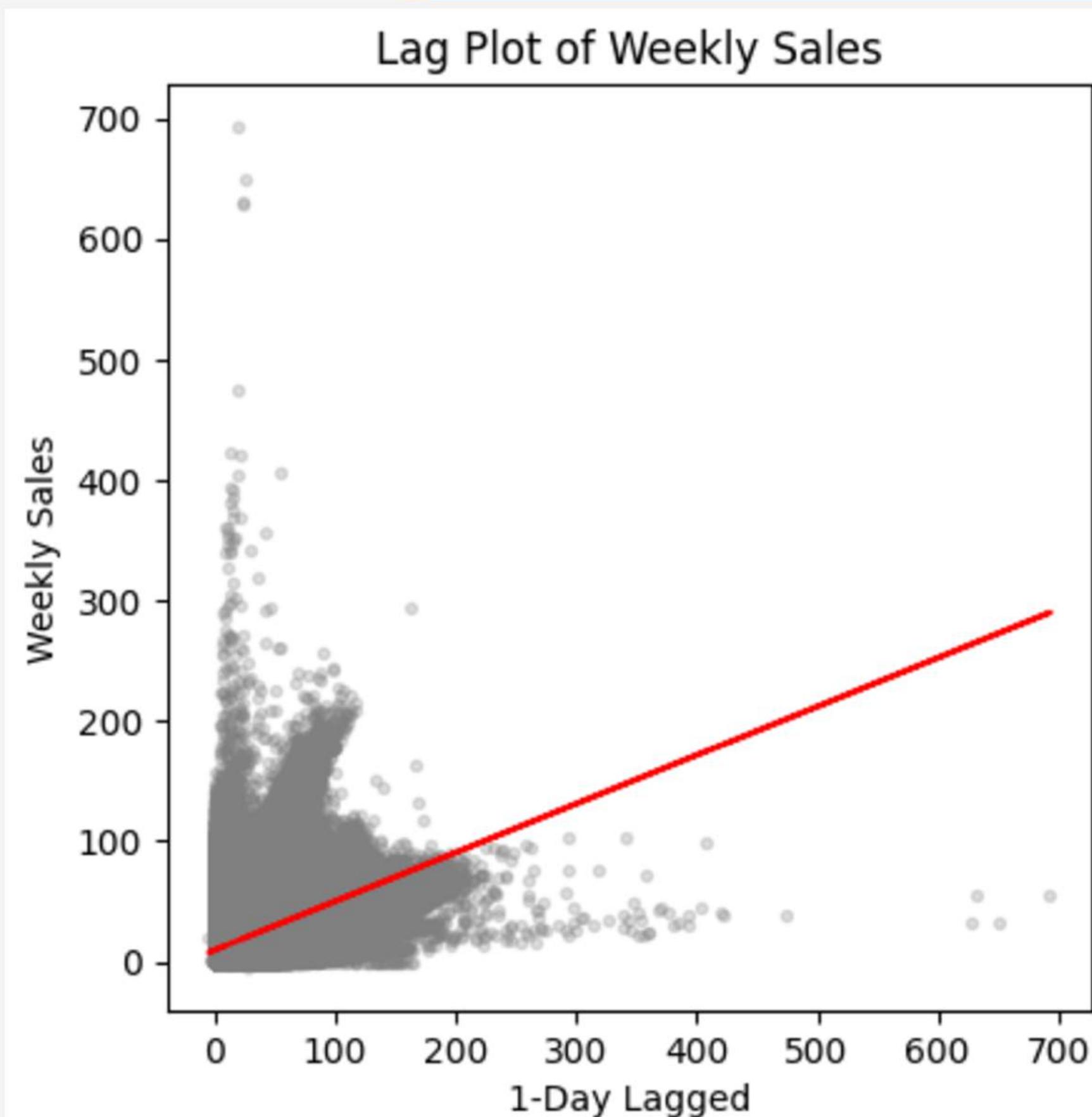
Phân Tích Chu Kỳ và Xu Hướng Doanh Số Bán Hàng

Xây dựng một mô hình hồi quy tuyến tính (Linear Regression) để dự đoán doanh số bán hàng dựa trên đặc trưng Lagged_Weekly_Sales

- Phân lõn các điểm dữ liệu tập trung ở khu vực giá trị thấp
- Phân tán rộng và không tập trung thành một đường thẳng, đặc biệt khi giá trị trễ tăng cao

=> Mô hình hồi quy tuyến tính có thể không phải là lựa chọn tối ưu.

3.2. MÔ HÌNH HÓA



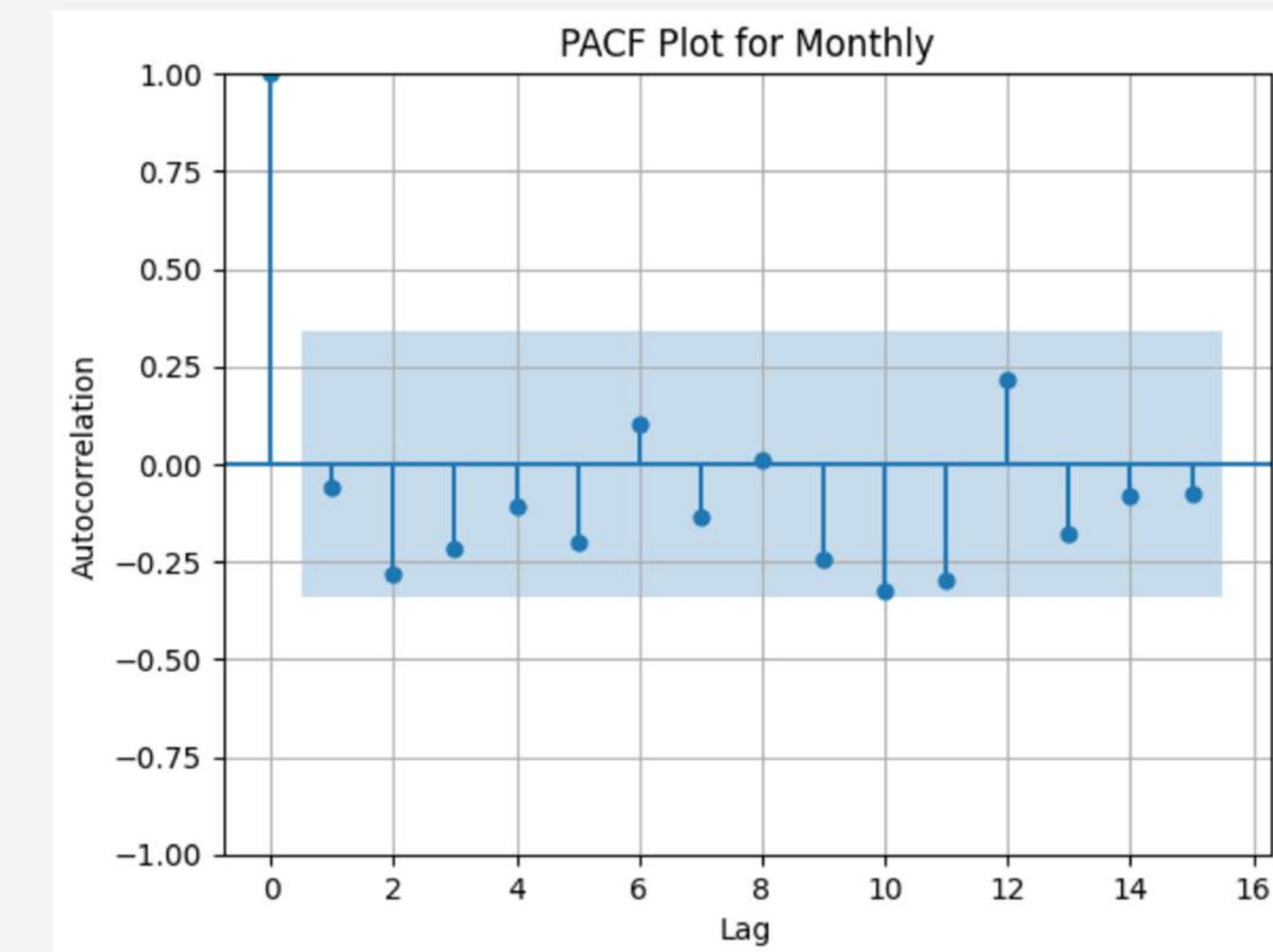
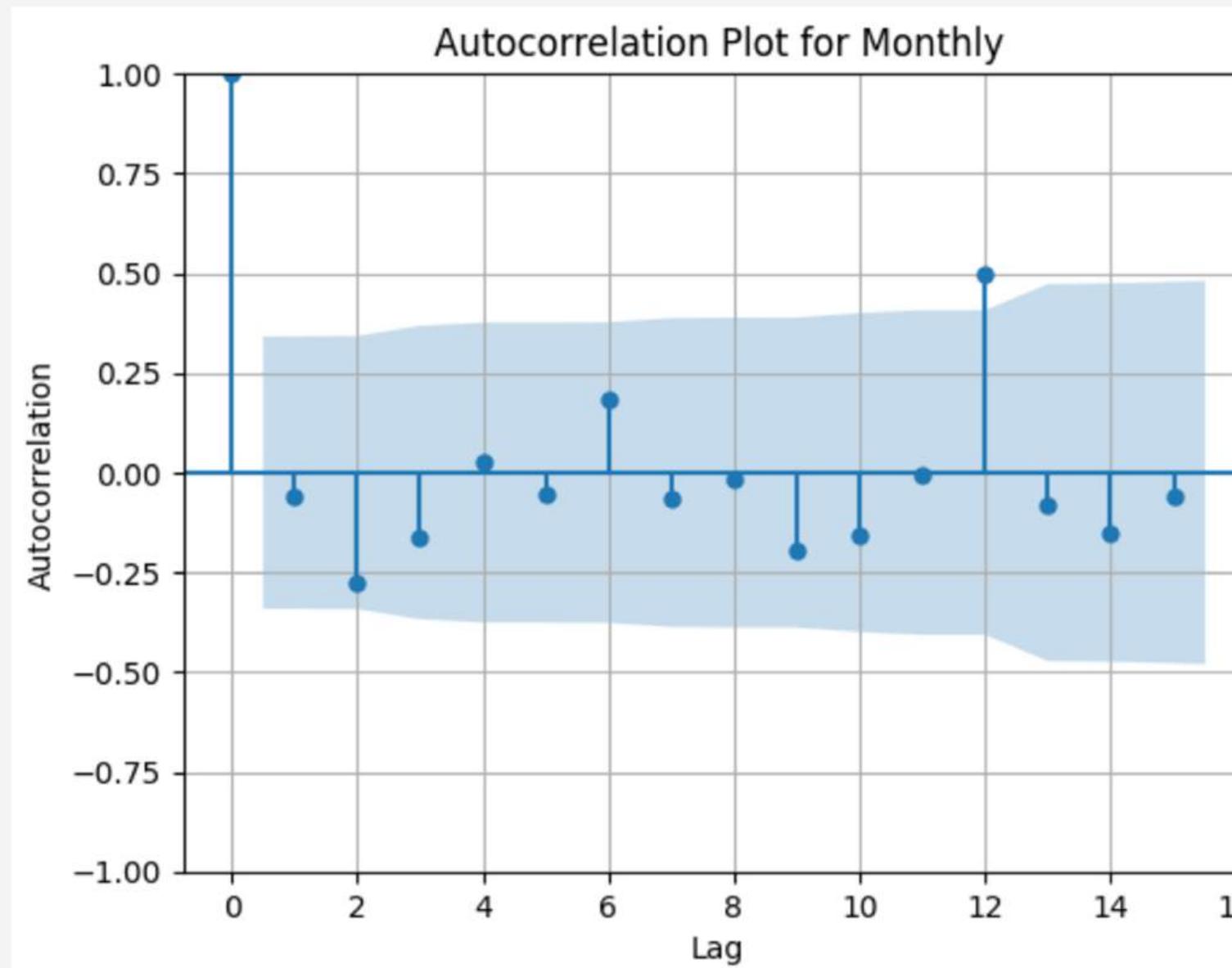
3. PHÂN TÍCH DỮ LIỆU VÀ MÔ HÌNH HÓA

Phân Tích và Dự Báo Doanh Số Hàng Tháng

3.2. MÔ HÌNH HÓA

Chúng ta có thể làm tốt hơn với dữ liệu trung bình hàng tháng không?

Sử dụng DataFrame monthlySales ở phần 3.1



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.

3. PHÂN TÍCH DỮ LIỆU VÀ MÔ HÌNH HÓA

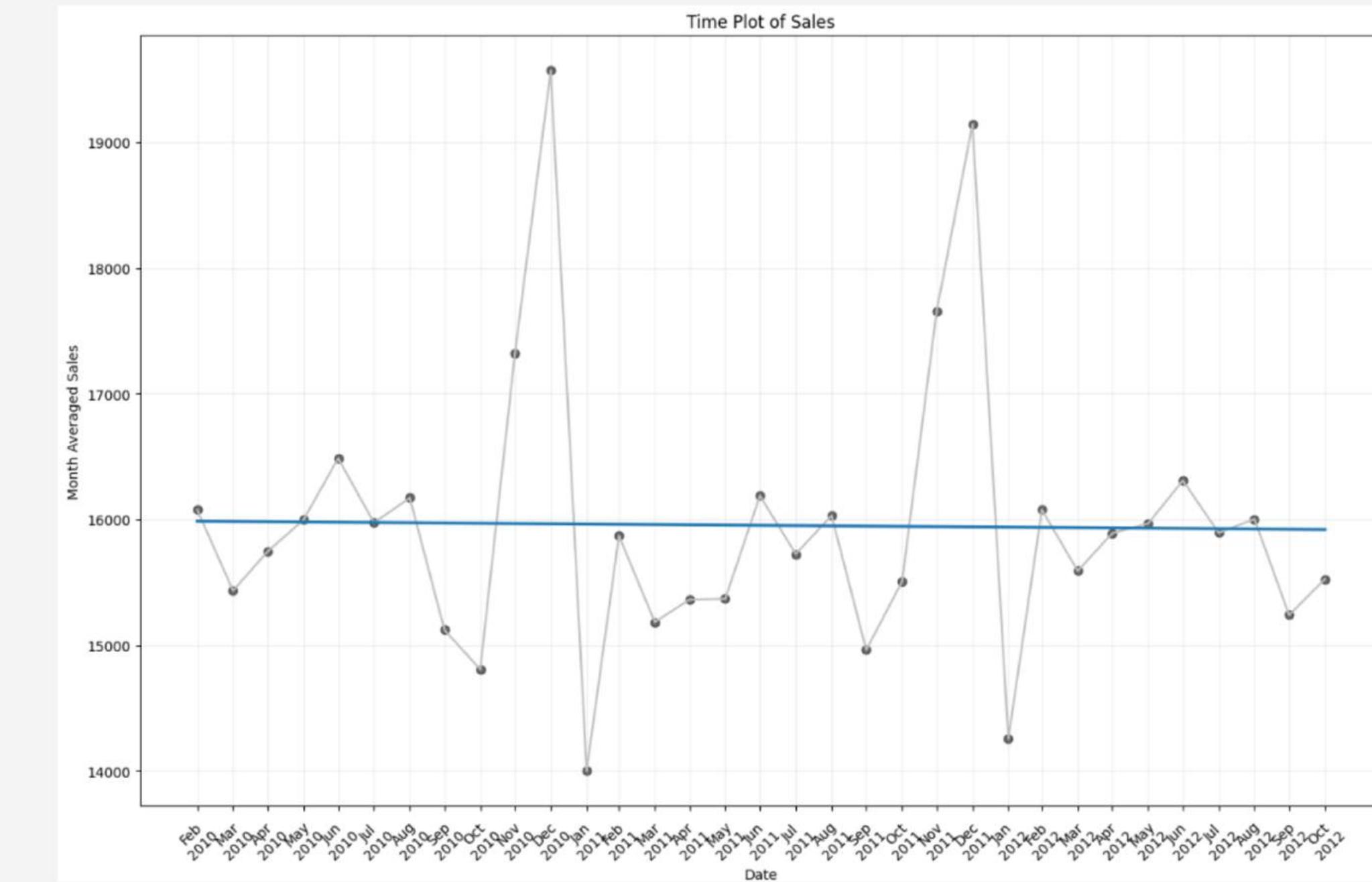
Phân Tích và Dự Báo Doanh Số Hàng Tháng

3.2. MÔ HÌNH HÓA

Chúng ta có thể làm tốt hơn với dữ liệu trung bình hàng tháng không?

- Tổng hợp dữ liệu hàng tuần thành hàng tháng để phân tích xu hướng.
- Sử dụng phương pháp resample('M') và vẽ biểu đồ đường với đường hồi quy.

Doanh số biến động mạnh có biến động mạnh vào các cuối tháng 12 đầu tháng 1, nhưng xu hướng tổng thể ổn định.



3. PHÂN TÍCH DỮ LIỆU VÀ MÔ HÌNH HÓA

Phân Tích và Dự Báo Doanh Số Hàng Tháng

3.2. MÔ HÌNH HÓA

Chúng ta có thể làm tốt hơn với dữ liệu trung bình hàng tháng không?

- Dự báo doanh số tháng tiếp theo dựa trên trung bình động.
- Sử dụng SMA 3 kỳ để dự báo và vẽ biểu đồ so sánh.

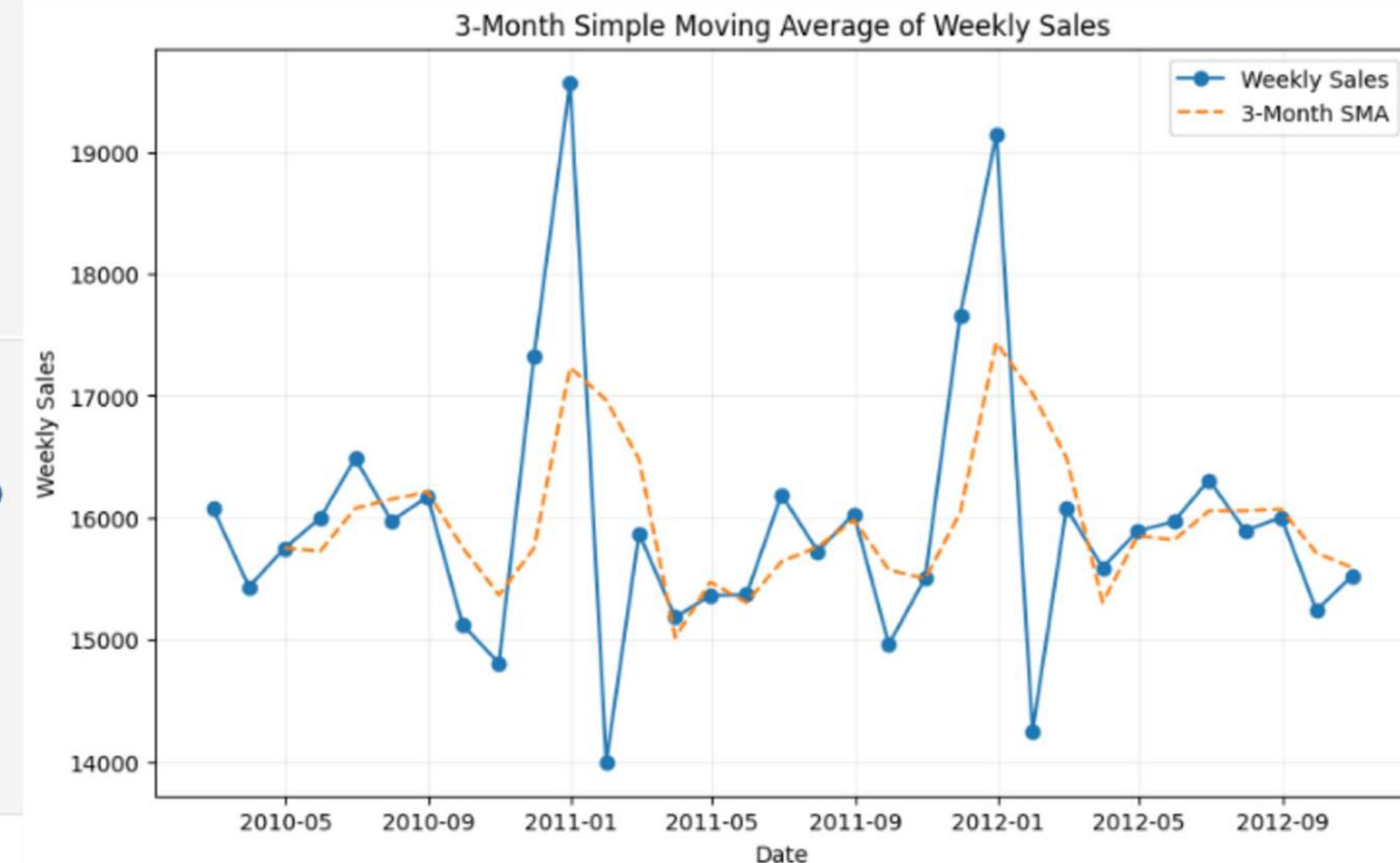
```
monthlySales.set_index('Date', inplace=True)

# Tính toán trung bình động 3 kỳ (SMA 3)
monthlySales['SMA_3'] = monthlySales['Weekly_Sales'].rolling(window=3).mean()

# Dự báo doanh số của kỳ tiếp theo
next_period_forecast = monthlySales['SMA_3'].iloc[-1]

print(f"The forecast for Nov 2012 sales is {next_period_forecast}")

The forecast for Nov 2012 sales is 15588.91624099263
```



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

3. PHÂN TÍCH DỮ LIỆU VÀ MÔ HÌNH HÓA

- Kiểm tra tính dừng của chuỗi thời gian và loại bỏ xu hướng nếu cần
- Sử dụng kiểm định ADF và phép lấy sai phân

3.2. MÔ HÌNH HÓA

ADF Statistic: -1.524424

p-value: 0.521397

Critical Values:

1%: -3.479

5%: -2.883

10%: -2.578

BEFORE

ADF Statistic: -13.616157

p-value: 0.000000

Critical Values:

1%: -3.479

5%: -2.883

10%: -2.578

AFTER

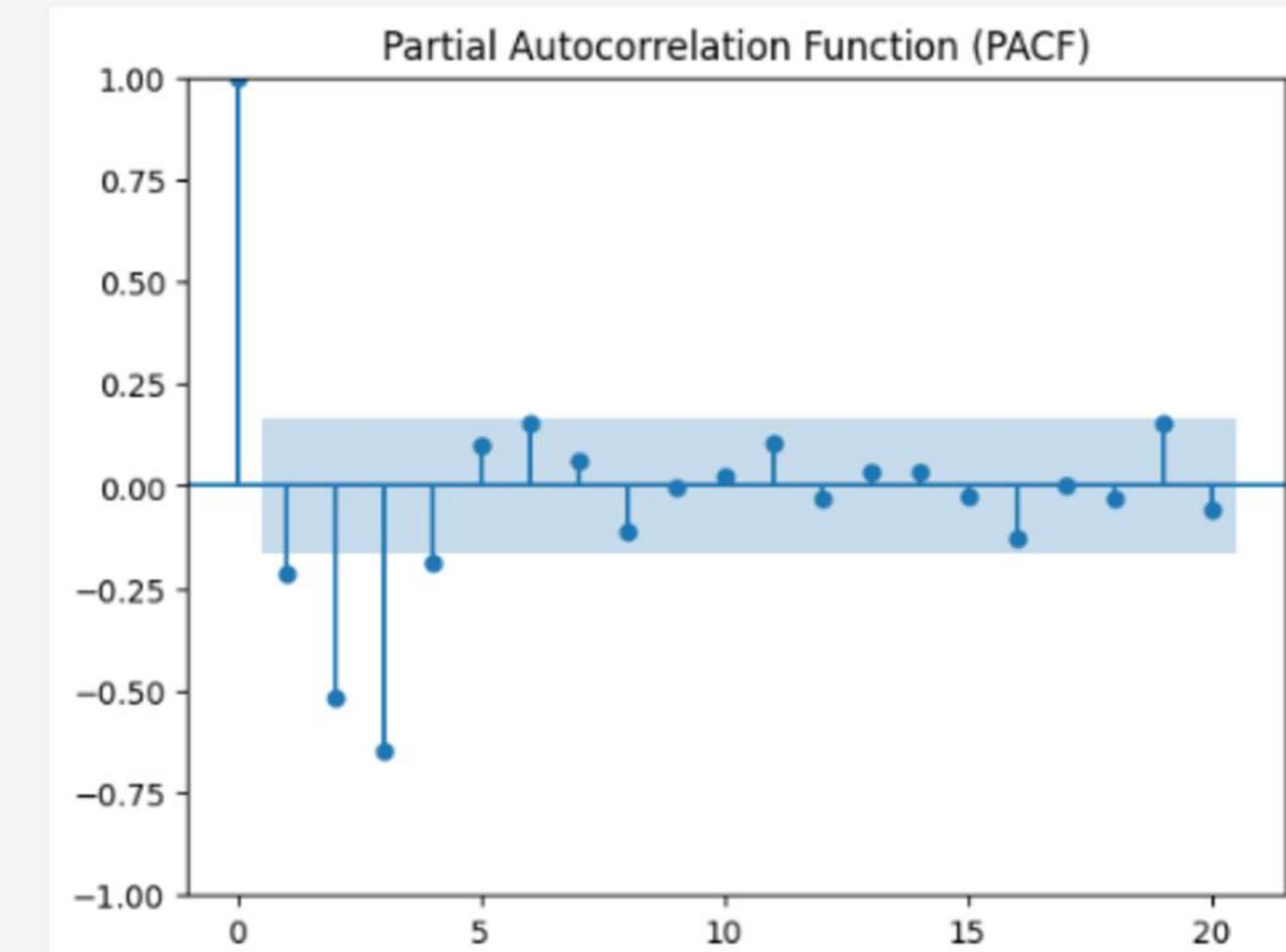
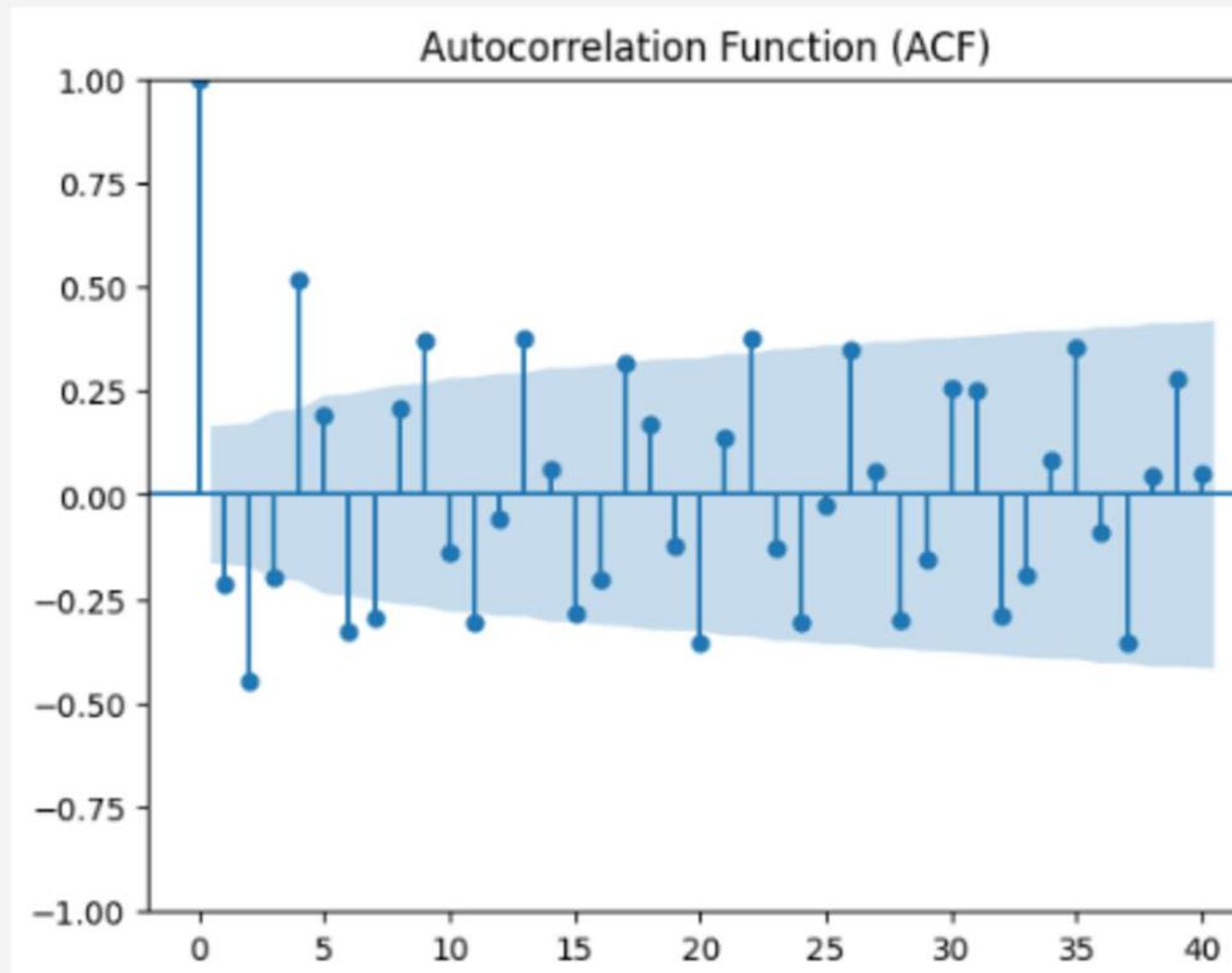
Chuỗi ban đầu không dừng, sau khi lấy sai phân thì chuỗi trở nên dừng

3. PHÂN TÍCH DỮ LIỆU VÀ MÔ HÌNH HÓA

Từ các phần trước, chúng ta biết rằng

Cửa hàng số 20 và Phòng ban số 92 có doanh số bán hàng cao nhất

3.2. MÔ HÌNH HÓA



Kết quả: Chọn $p = 3$ và $q = 4$

Áp Dụng ARIMA cho Doanh Số Bán Hàng

3. PHÂN TÍCH DỮ LIỆU VÀ MÔ HÌNH HÓA

p = 3 từ biểu đồ PACF

d = 1 vì chúng ta đã thực hiện phép lấy sai phân một lần

q = 4 dựa trên biểu đồ ACF

Chia dữ liệu và huấn luyện mô hình ARIMA(3, 1, 4) bằng statsmodels

SARIMAX Results						
Dep. Variable:	Weekly_Sales	No. Observations:	107			
Model:	ARIMA(3, 1, 4)	Log Likelihood	-1176.362			
Date:	Sat, 16 Nov 2024	AIC	2368.723			
Time:	14:53:29	BIC	2390.031			
Sample:	02-05-2010 - 02-17-2012	HQIC	2377.359			
Covariance Type:	opg					
	coef	std err	z	P> z	[0.025	0.975]
ar.L1	-0.7192	0.447	-1.610	0.107	-1.594	0.156
ar.L2	-0.7465	0.112	-6.648	0.000	-0.967	-0.526
ar.L3	-0.9764	0.444	-2.199	0.028	-1.847	-0.106
ma.L1	-0.0061	0.479	-0.013	0.990	-0.944	0.932
ma.L2	0.2332	0.494	0.473	0.637	-0.734	1.201
ma.L3	0.4567	0.526	0.869	0.385	-0.574	1.487
ma.L4	-0.6892	0.306	-2.249	0.024	-1.290	-0.089
sigma2	2.182e+08	8.53e-10	2.56e+17	0.000	2.18e+08	2.18e+08
Ljung-Box (L1) (Q):	0.13	Jarque-Bera (JB):	79.61			
Prob(Q):	0.72	Prob(JB):	0.00			
Heteroskedasticity (H):	1.14	Skew:	-0.63			
Prob(H) (two-sided):	0.71	Kurtosis:	7.06			
...						

3.2. MÔ HÌNH HÓA

Các hệ số AR(2), AR(3), MA(4) có p-value < 0.05, cho thấy chúng đóng vai trò quan trọng trong việc giải thích dữ liệu.

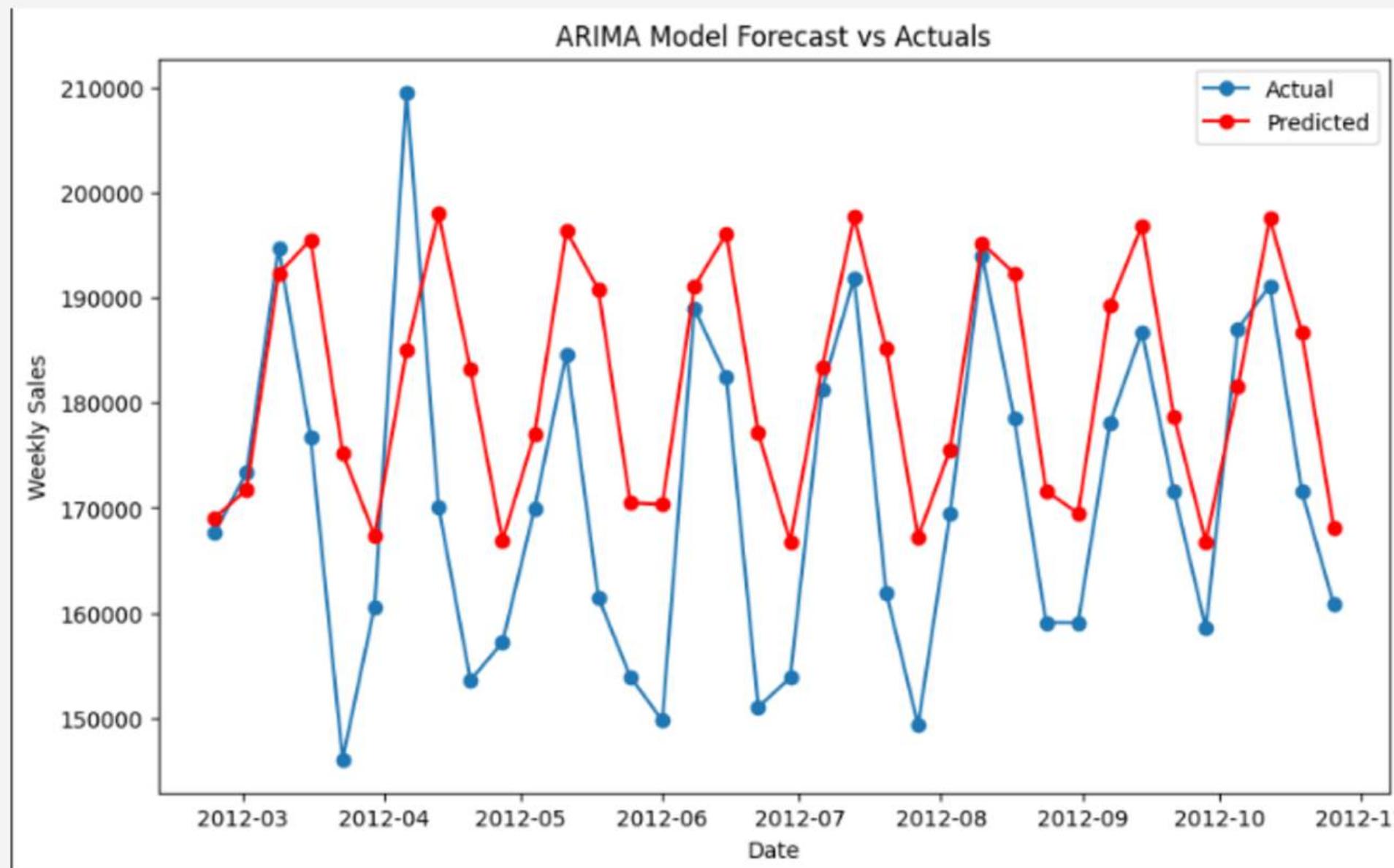
Kiểm định Ljung-Box cho Prob(Q) = 0.72 > 0.05, xác nhận phần dư đã ngẫu nhiên.

Kiểm định Heteroskedasticity cho Prob(H) = 0.71 > 0.05, chứng tỏ mô hình ổn định.

3. PHÂN TÍCH DỮ LIỆU VÀ MÔ HÌNH HÓA

3.2. MÔ HÌNH HÓA

Đánh Giá Mô Hình ARIMA Thủ Công



RMSE: 15357.575

- Mô hình dự đoán tốt xu hướng nhưng gặp khó khăn với các đột biến mạnh.

3. PHÂN TÍCH DỮ LIỆU VÀ MÔ HÌNH HÓA

3.2. MÔ HÌNH HÓA

SARIMAX Results						
Dep. Variable:	y	No. Observations:	107 <th data-cs="3" data-kind="parent"></th> <th data-kind="ghost"></th> <th data-kind="ghost"></th>			
Model:	SARIMAX(3, 1, 1)	Log Likelihood	-1186.954			
Date:	Sat, 16 Nov 2024	AIC	2385.908			
Time:	21:00:33	BIC	2401.888			
Sample:	02-05-2010 - 02-17-2012	HQIC	2392.385			
Covariance Type:	opg					
	coef	std err	z	P> z	[0.025	0.975]
intercept	494.3981	1202.992	0.411	0.681	-1863.422	2852.219
ar.L1	-0.4682	0.098	-4.793	0.000	-0.660	-0.277
ar.L2	-0.5359	0.053	-10.051	0.000	-0.640	-0.431
ar.L3	-0.4255	0.040	-10.699	0.000	-0.503	-0.348
ma.L1	-0.3078	0.123	-2.509	0.012	-0.548	-0.067
sigma2	2.344e+08	0.007	3.25e+10	0.000	2.34e+08	2.34e+08
Ljung-Box (L1) (Q):	1.05	Jarque-Bera (JB):	30.80			
Prob(Q):	0.31	Prob(JB):	0.00			
Heteroskedasticity (H):	0.73	Skew:	-0.21			
Prob(H) (two-sided):	0.35	Kurtosis:	5.61			

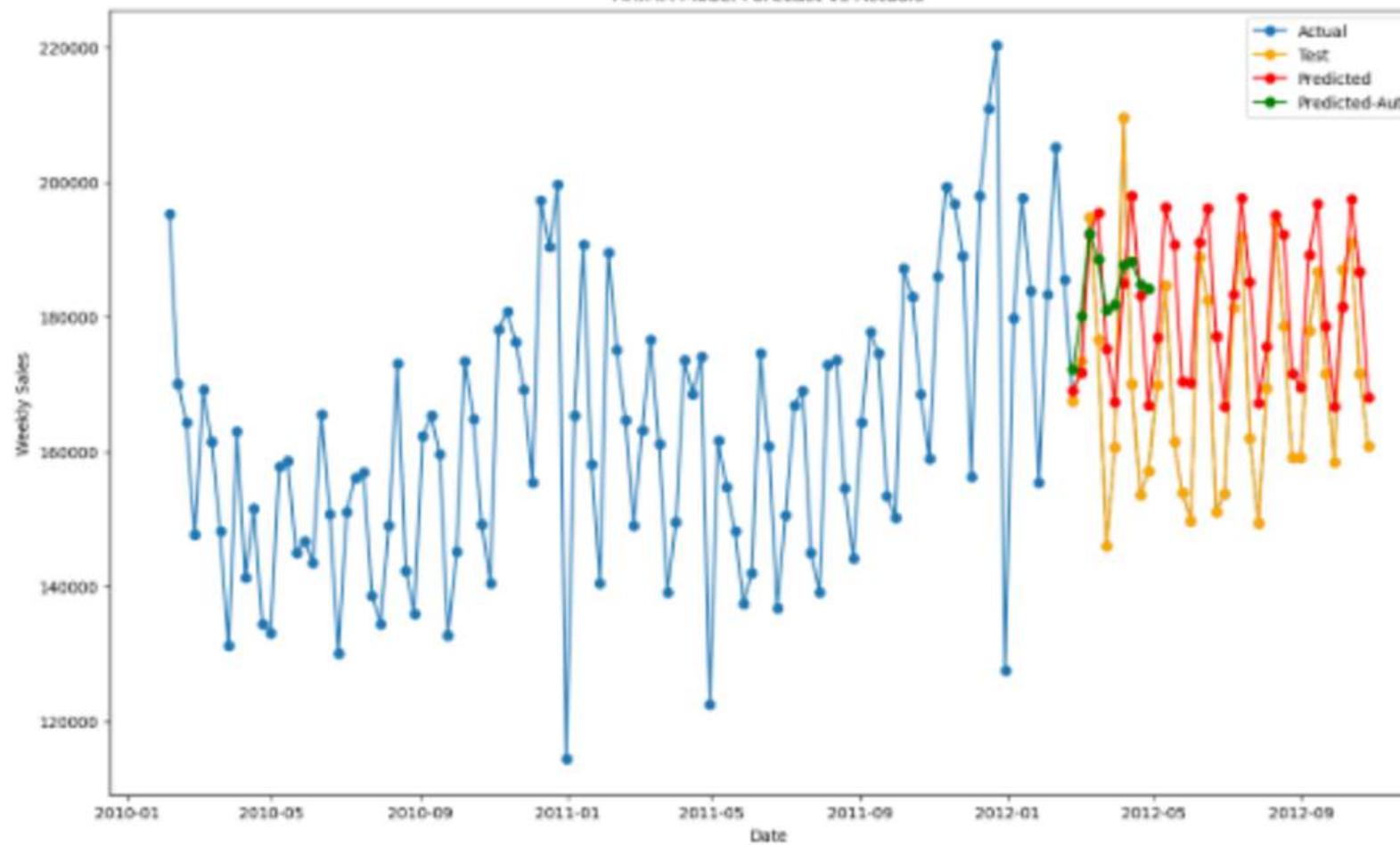
Mô hình SARIMAX(3, 1, 1) thể hiện khả năng giải thích biến động dữ liệu khá tốt.

Các hệ số AR và MA có ý nghĩa thống kê cao.
Phần dư không có tự tương quan đáng kể.

Hạn chế:

Giá trị AIC và BIC cao

Phần dư không tuân theo phân phối chuẩn



3. PHÂN TÍCH DỮ LIỆU VÀ MÔ HÌNH HÓA

3.2. MÔ HÌNH HÓA

Khi so sánh với tập kiểm tra, kết quả của Auto-ARIMA tệ hơn so với phiên bản được điều chỉnh thủ công.

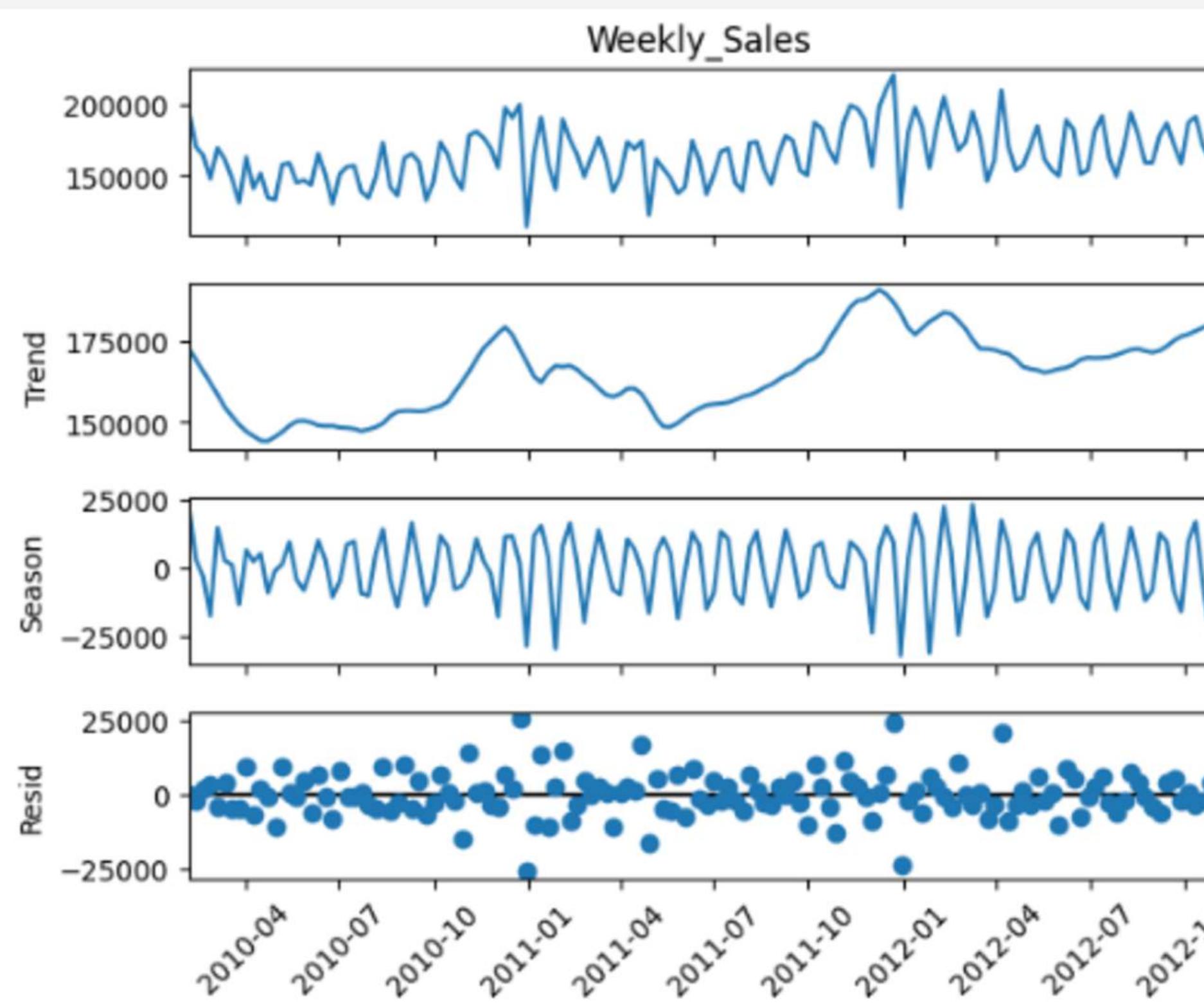
Mô hình ARIMA điều chỉnh thủ công thể hiện tính linh hoạt cao hơn trong việc thích nghi với toàn bộ tập dữ liệu kiểm tra, khiến nó phù hợp hơn cho các dự đoán dài hạn.

Mặc dù cả hai mô hình đều nắm bắt được xu hướng tổng thể, phiên bản điều chỉnh thủ công cho thấy sự phù hợp tốt hơn với các dao động lớn, mặc dù vẫn có một số sai số.

Tách Chuỗi Thời Gian Thành Các Thành Phần

3. PHÂN TÍCH DỮ LIỆU VÀ MÔ HÌNH HÓA

3.2. MÔ HÌNH HÓA



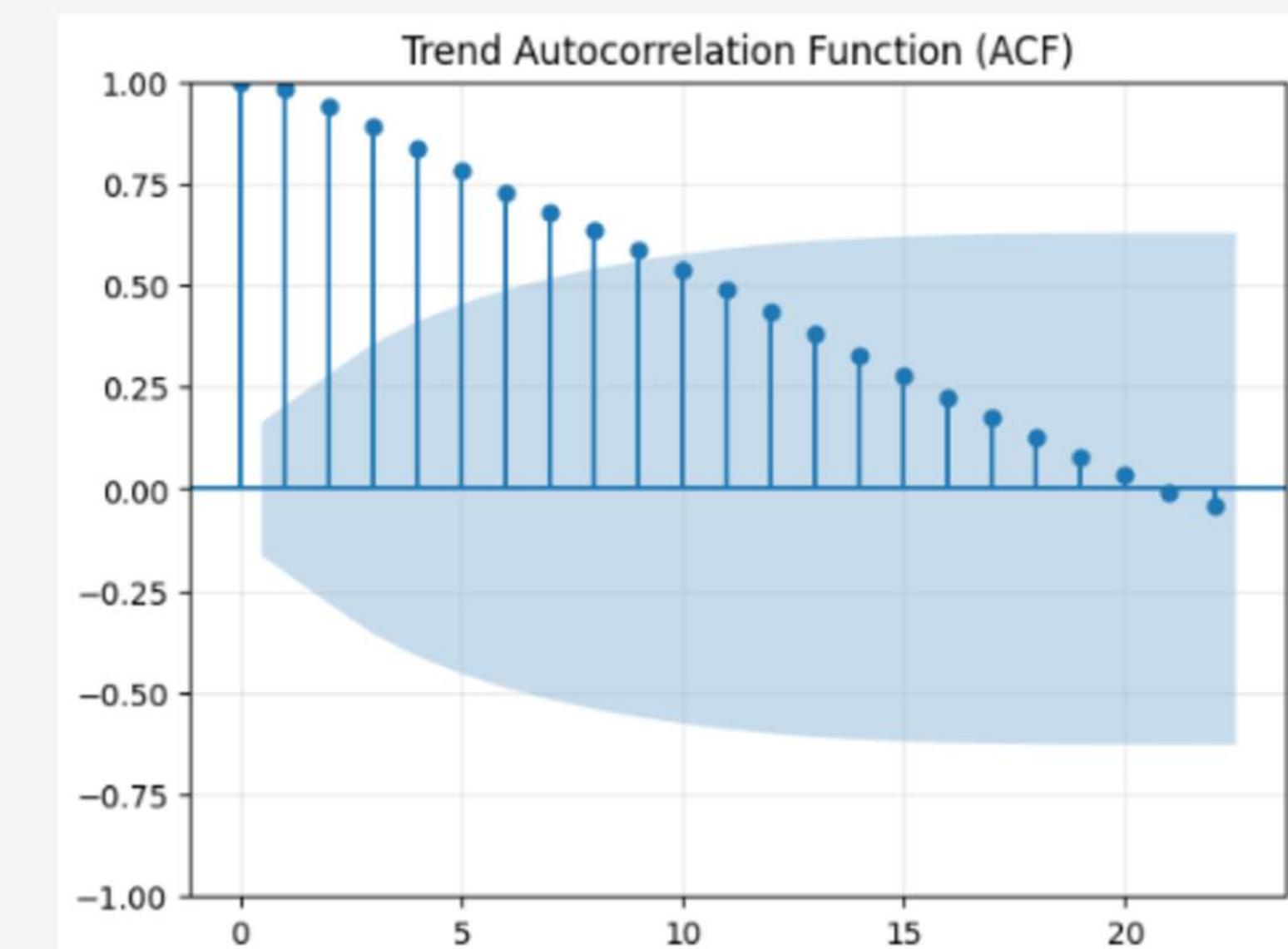
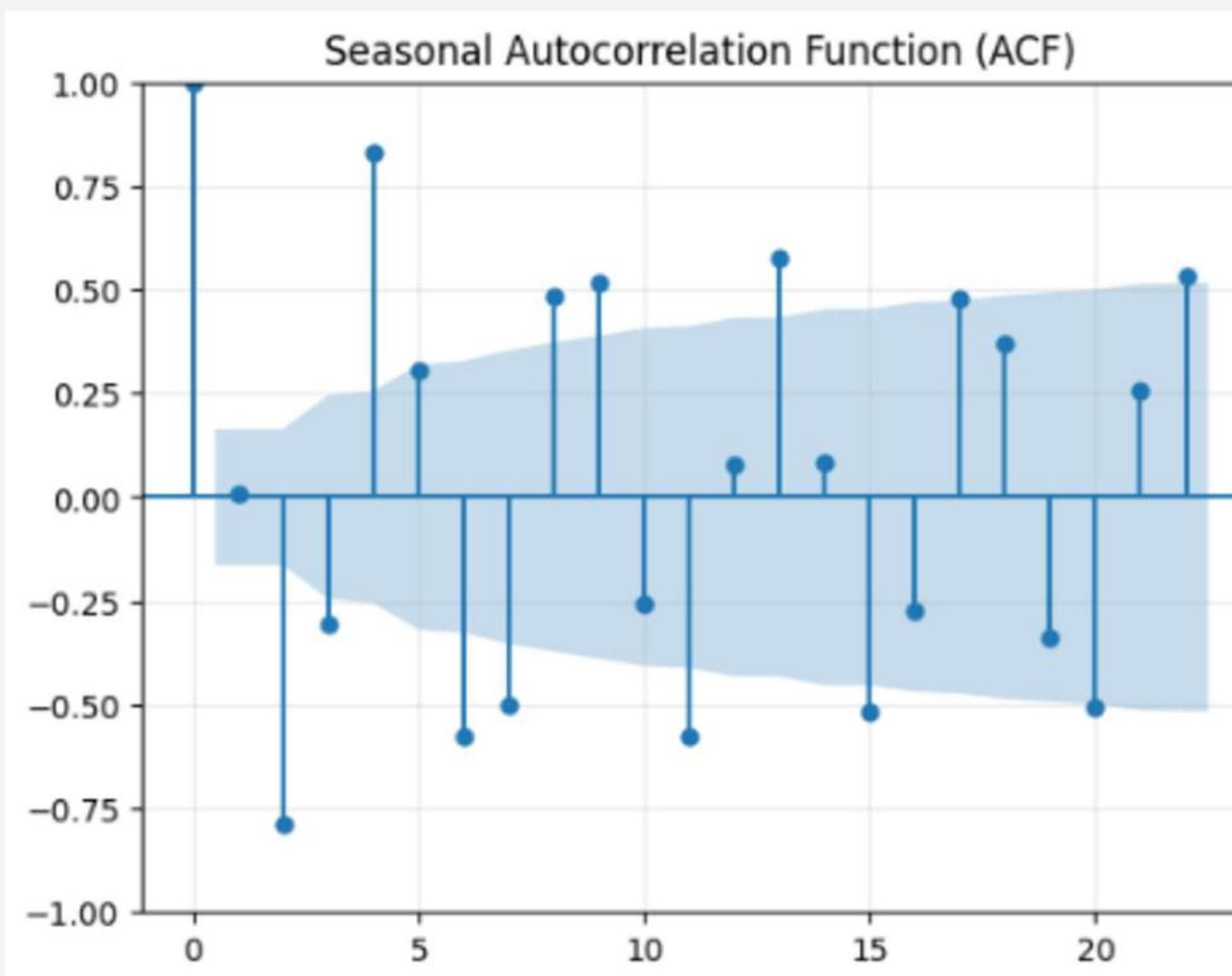
- Doanh số bán hàng cho thấy xu hướng tăng trưởng theo thời gian.
- Mùa vụ có dao động chu kỳ, có thể do yếu tố mùa hoặc khuyến mãi.
- Phần dư không có mô hình rõ rệt, chứng tỏ mô hình đã giải thích tốt các yếu tố xu hướng và mùa vụ.

Áp Dụng ARIMA và SARIMA cho Dự Báo Doanh Số

3. PHÂN TÍCH DỮ LIỆU VÀ MÔ HÌNH HÓA

Phân Tích ACF cho Seasonal và Trend

3.2. MÔ HÌNH HÓA



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

Áp Dụng ARIMA và SARIMA cho Dự Báo Doanh Số

3. PHÂN TÍCH DỮ LIỆU VÀ MÔ HÌNH HÓA

Grid Search để Chọn Tham Số SARIMA Tối Ưu

3.2. MÔ HÌNH HÓA

	pdq	pdqs	bic
147	(1, 0, 2)	(0, 1, 0, 4)	3089.516934
237	(2, 0, 1)	(0, 1, 0, 4)	3090.748048
4	(0, 0, 0)	(0, 1, 1, 4)	3091.313420
9	(0, 0, 0)	(1, 1, 0, 4)	3091.670812
3	(0, 0, 0)	(0, 1, 0, 4)	3091.986025

Tham số tối ưu là (1, 0, 2) và (0, 1, 0, 4) với BIC = 3089.516934

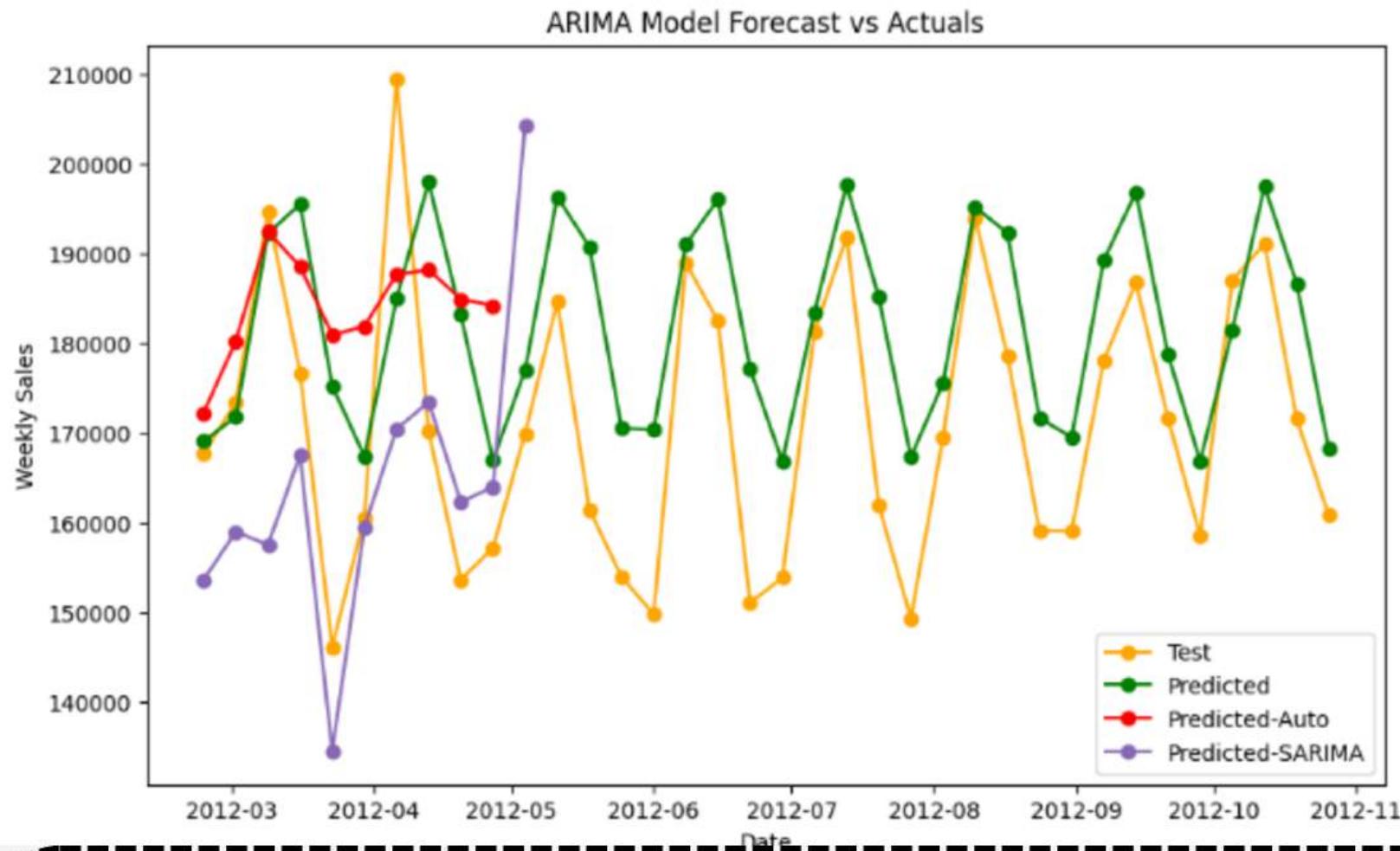
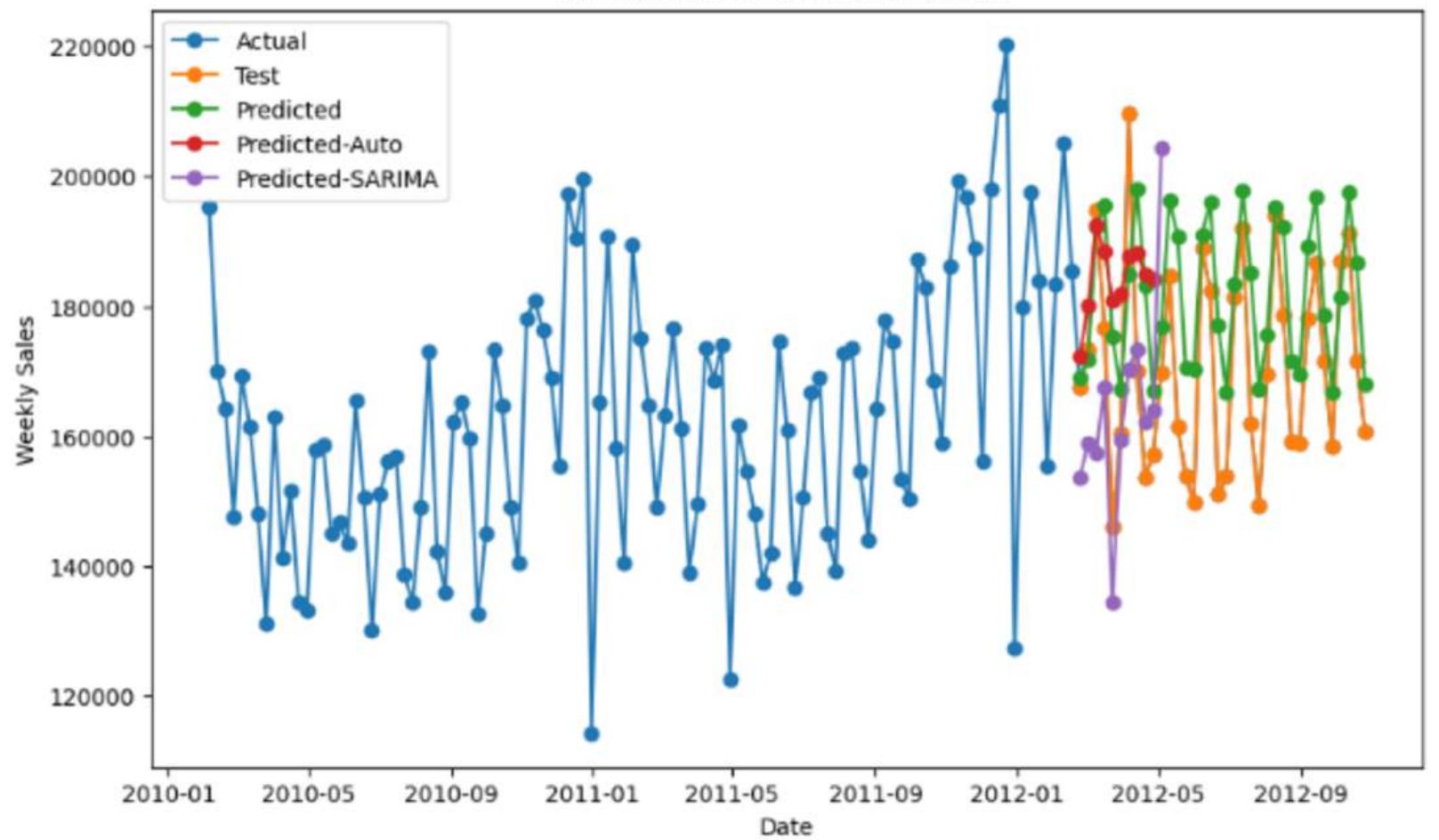
3. PHÂN TÍCH DỮ LIỆU VÀ MÔ HÌNH HÓA

Xây dựng mô hình SARIMAX với các tham số tối ưu

3.2. MÔ HÌNH HÓA

SARIMAX Results						
<hr/>						
Dep. Variable:	Weekly_Sales	No. Observations:	107			
Model:	SARIMAX(1, 0, 2)x(0, 1, [], 4)	Log Likelihood	-1174.441			
Date:	Sat, 16 Nov 2024	AIC	2356.882			
Time:	21:01:56	BIC	2367.302			
Sample:	02-05-2010 - 02-17-2012	HQIC	2361.099			
Covariance Type:	opg					
<hr/>						
	coef	std err	z	P> z	[0.025	0.975]
<hr/>						
ar.L1	1.1400	0.138	8.243	0.000	0.869	1.411
ma.L1	1.6589	0.088	18.882	0.000	1.487	1.831
ma.L2	-1.8894	0.194	-9.738	0.000	-2.270	-1.509
sigma2	2.809e+08	3.49e-10	8.05e+17	0.000	2.81e+08	2.81e+08
<hr/>						
Ljung-Box (L1) (Q):	18.55	Jarque-Bera (JB):	52.77			
Prob(Q):	0.00	Prob(JB):	0.00			
Heteroskedasticity (H):	0.49	Skew:	-0.15			
Prob(H) (two-sided):	0.04	Kurtosis:	6.55			
<hr/>						

Mô hình đã nắm bắt được các mẫu quan trọng trong dữ liệu, nhưng có thể có một số vấn đề với tự tương quan phần dư và không tuân theo phân phối chuẩn.



3. PHÂN TÍCH DỮ LIỆU VÀ MÔ HÌNH HÓA

3.2. MÔ HÌNH HÓA

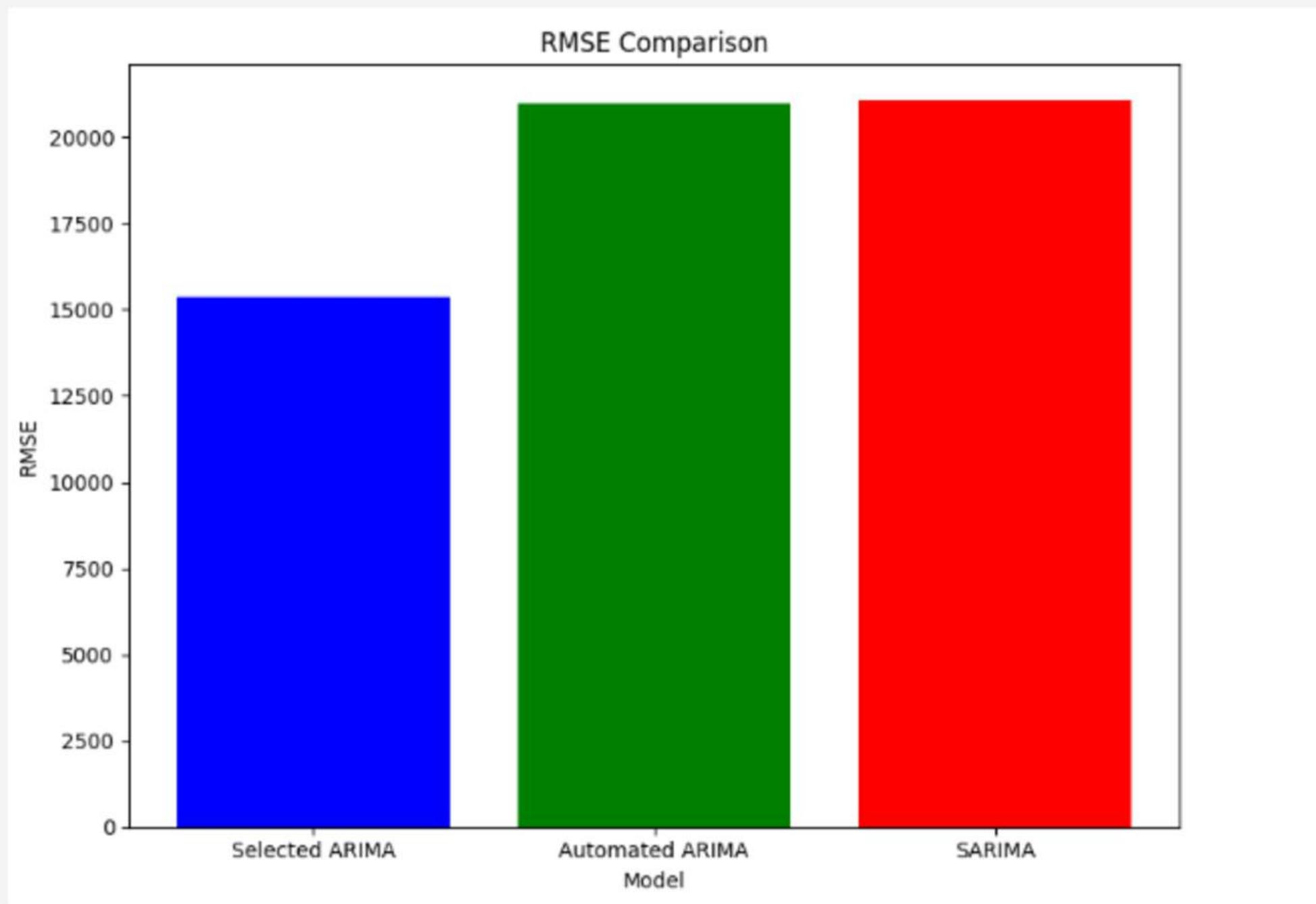
Dự Đoán và So Sánh Các Mô Hình

Cả ba mô hình ARIMA (Predicted, Predicted-Auto, Predicted-SARIMA) đều không dự báo chính xác tuyệt đối so với dữ liệu thực tế. Điều này cho thấy dự báo doanh số bán hàng là một bài toán phức tạp và luôn tồn tại sai số.

3. PHÂN TÍCH DỮ LIỆU VÀ MÔ HÌNH HÓA

Đánh giá mô hình dựa trên RMSE

3.2. MÔ HÌNH HÓA



Selected ARIMA (Thủ công) là mô hình được khuyến nghị sử dụng vì có hiệu suất dự báo tốt nhất (RMSE nhỏ nhất)

Dự đoán doanh số bán hàng (Sales Forecasting)



 + a b l e a u®
S O F T W A R E

Chỉ số chính



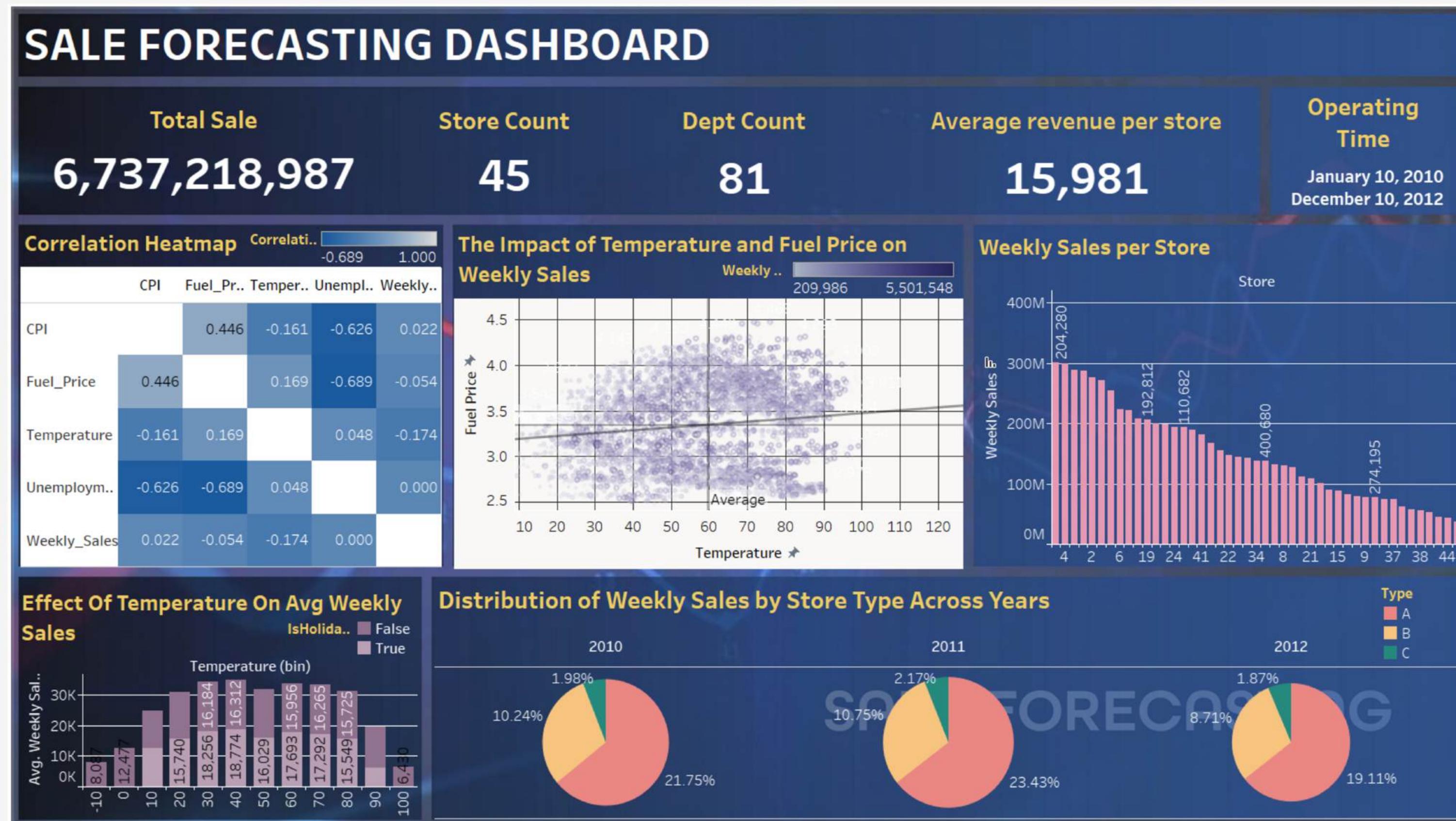
Xu hướng

Thông tin chi tiết

Cung cấp một cái nhìn tổng quan toàn diện và đề xuất các lĩnh vực cần cải thiện.

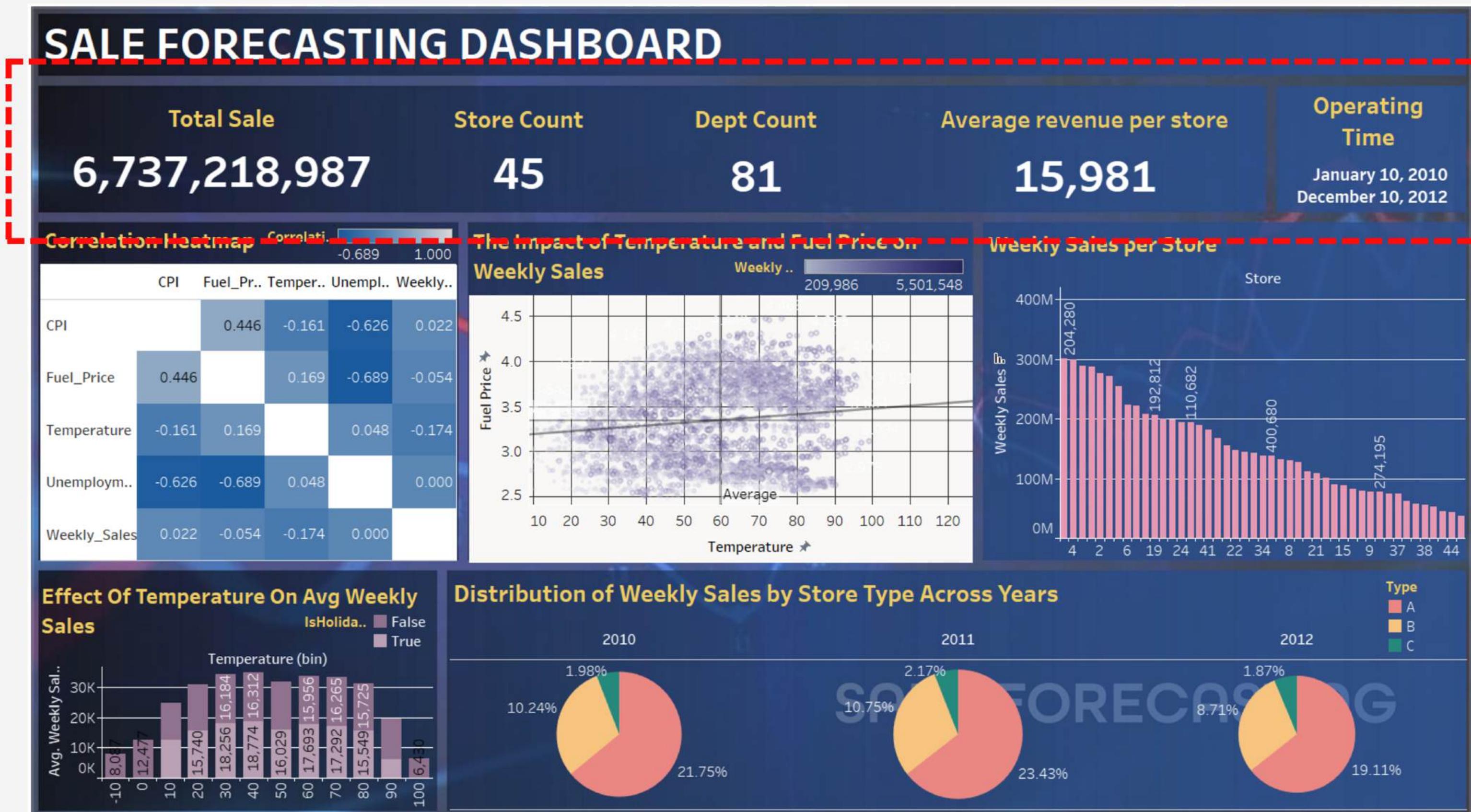


4. TRỰC QUAN HÓA DỮ LIỆU



Hình: Sales Forecasting Dashboard

4. TRỰC QUAN HÓA DỮ LIỆU



Hình: Sales Forecasting Dashboard

4. TRỰC QUAN HÓA DỮ LIỆU

TỔNG DOANH SỐ: 6,737,218,987

SỐ LƯỢNG CỬA HÀNG: 45

SỐ LƯỢNG PHÒNG BAN: 81

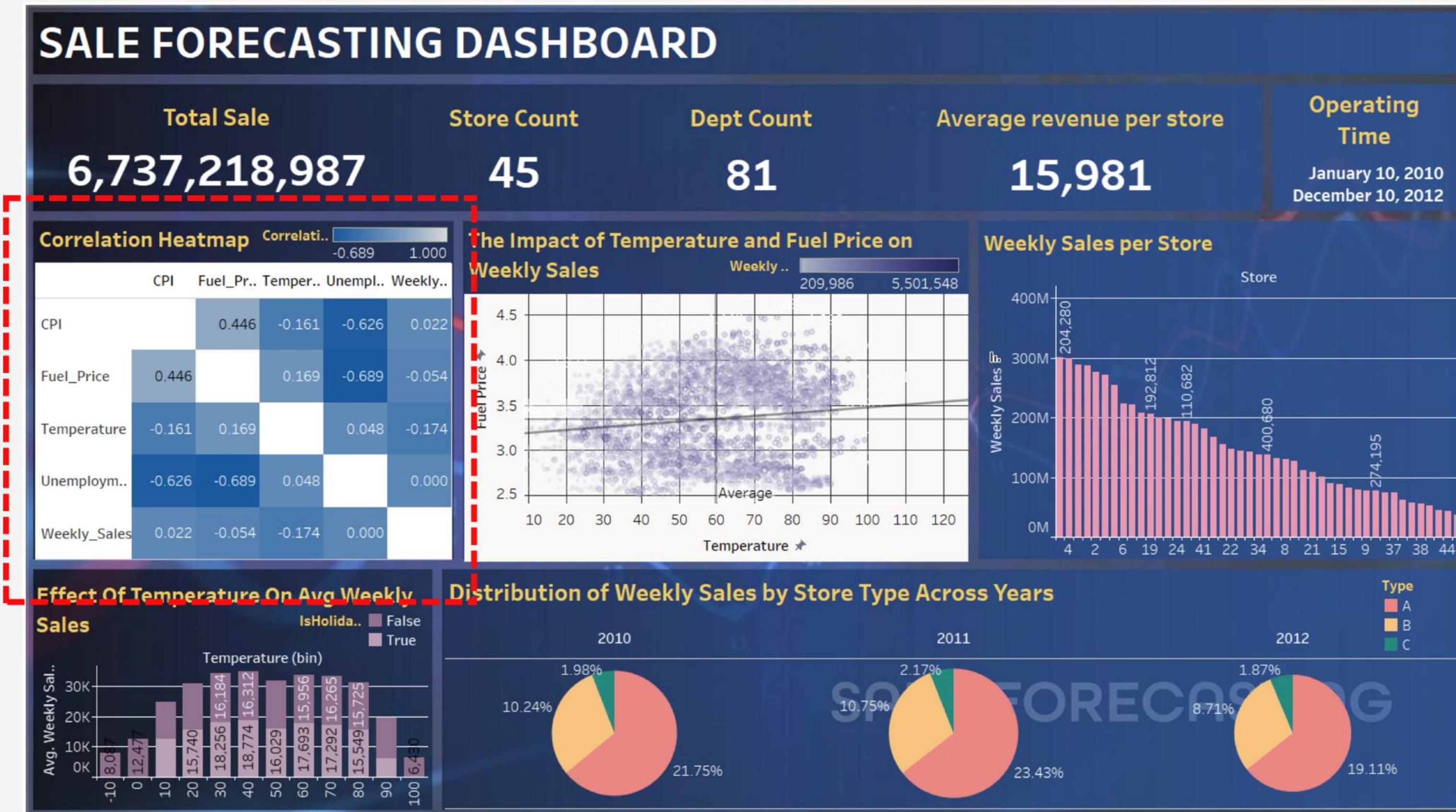
DOANH THU TRUNG BÌNH MỖI CỬA
HÀNG: 15,981

THỜI GIAN HOẠT ĐỘNG: 10/01/2010 - 10/12/2012

Total Sale	Store Count	Dept Count	Average revenue per store	Operating Time
6,737,218,987	45	81	15,981	January 10, 2010 December 10, 2012

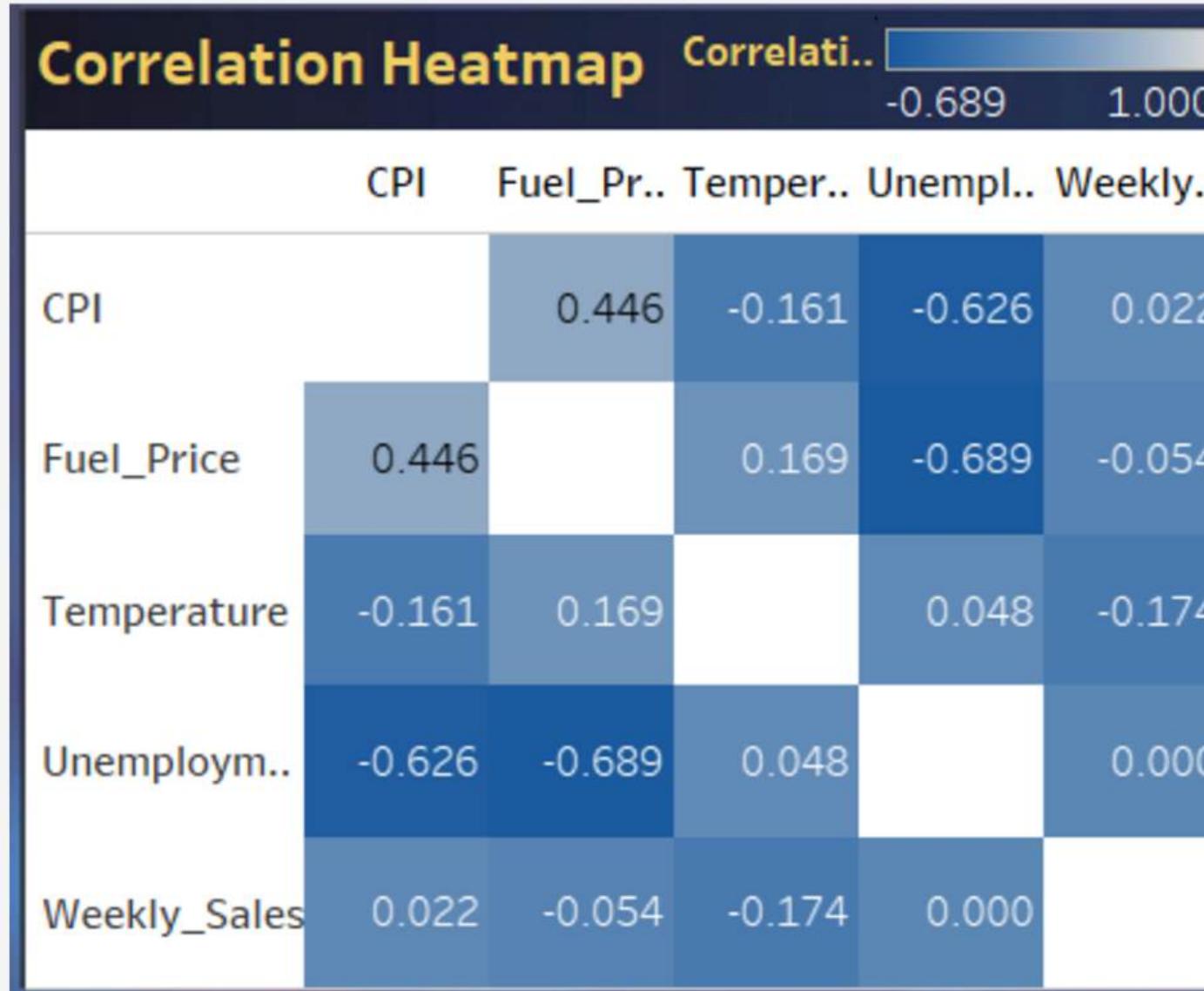
Hình: Các chỉ số hiệu suất chính

4. TRỰC QUAN HÓA DỮ LIỆU



Hình: Sales Forecasting Dashboard

4. TRỰC QUAN HÓA DỮ LIỆU



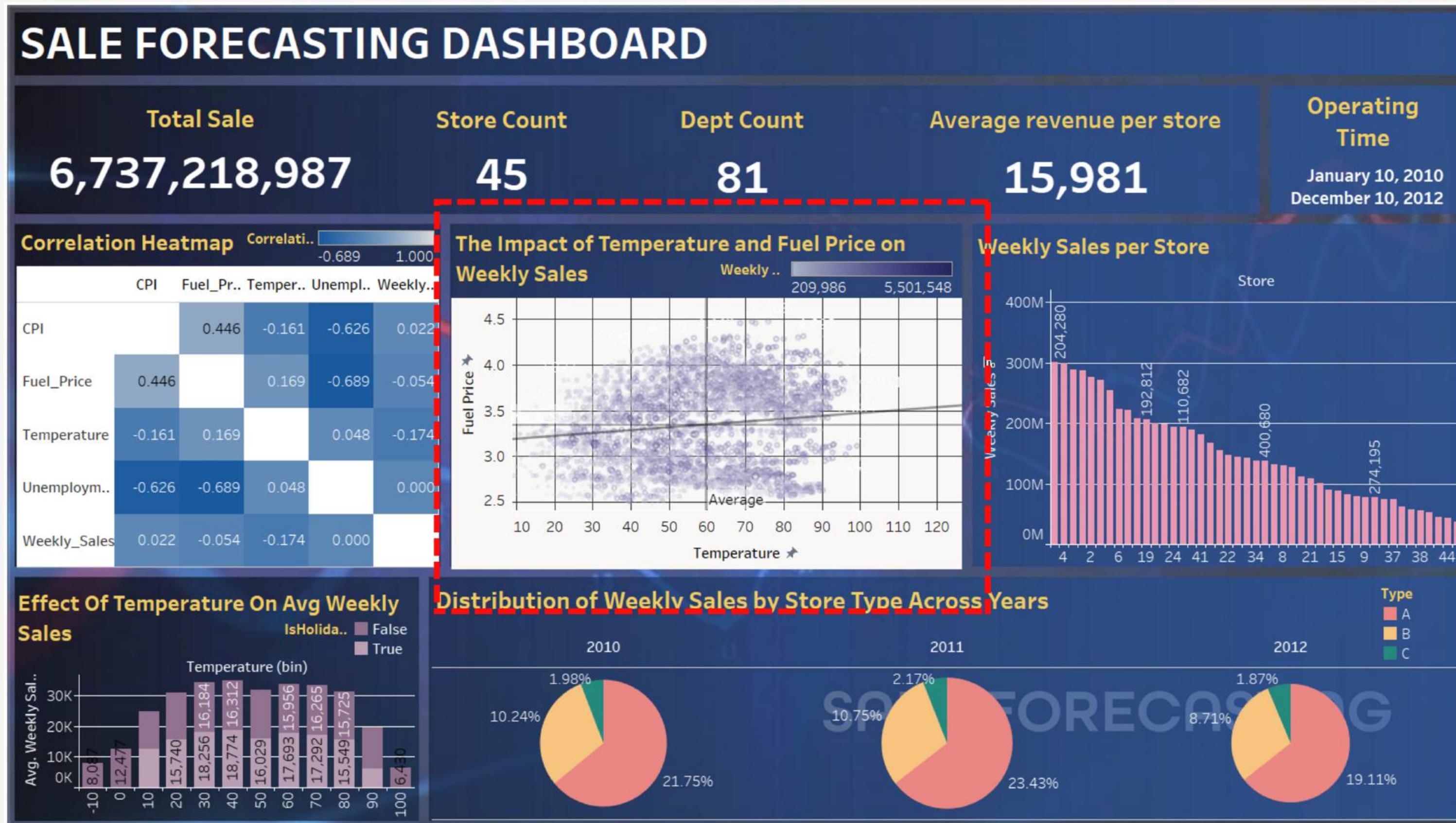
+ Giá nhiên liệu và tỷ lệ thất nghiệp (-0.689): Mỗi quan hệ tương quan âm mạnh cho thấy khi giá nhiên liệu tăng, tỷ lệ thất nghiệp có xu hướng giảm.

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

Hình: Biểu đồ nhiệt độ tương quan

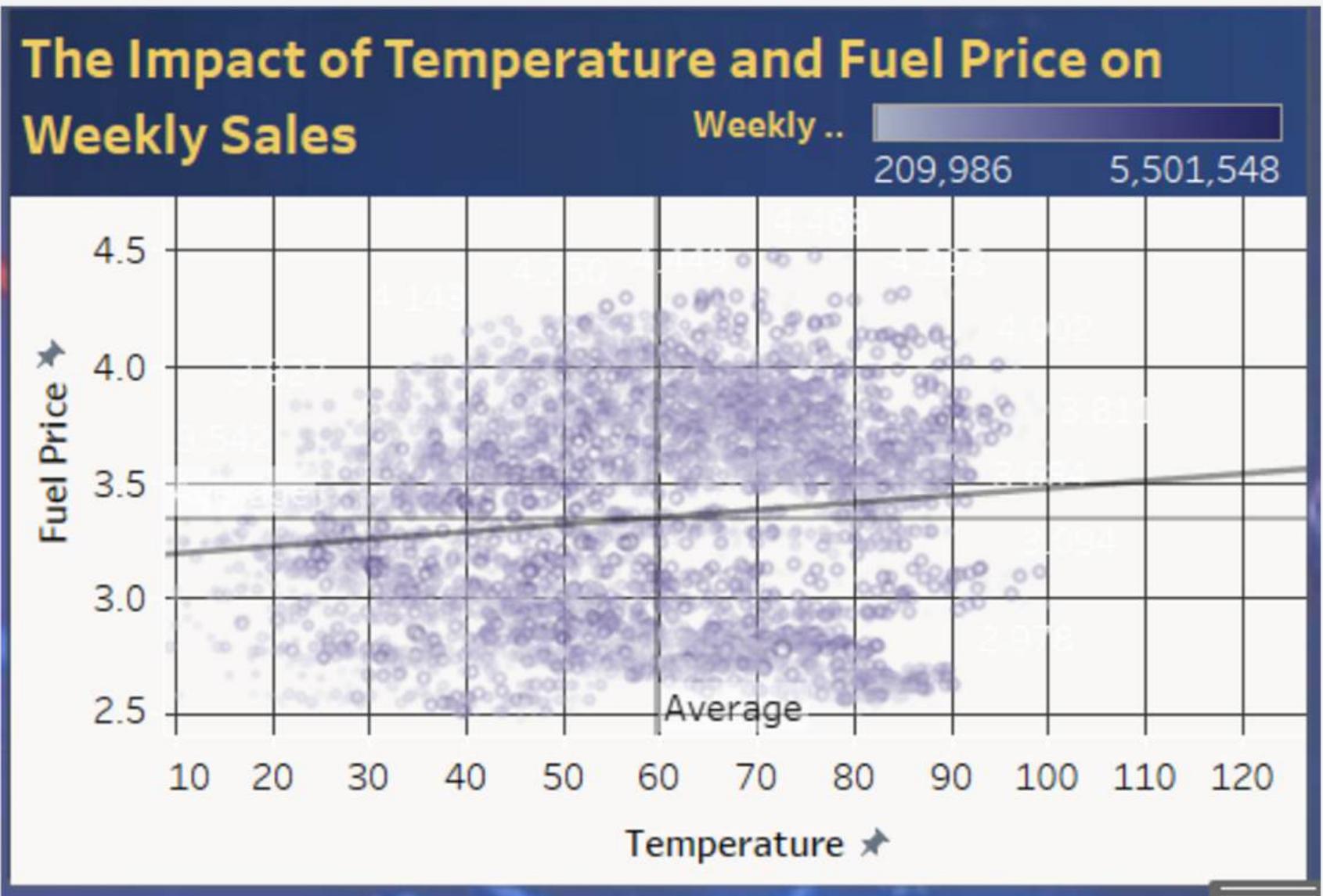
Giá nhiên liệu và nhiệt độ có mức ảnh hưởng đáng kể hơn đến doanh số hàng tuần, CPI chỉ có ảnh hưởng nhẹ và không rõ ràng.

4. TRỰC QUAN HÓA DỮ LIỆU



Hình: Sales Forecasting Dashboard

4. TRỰC QUAN HÓA DỮ LIỆU

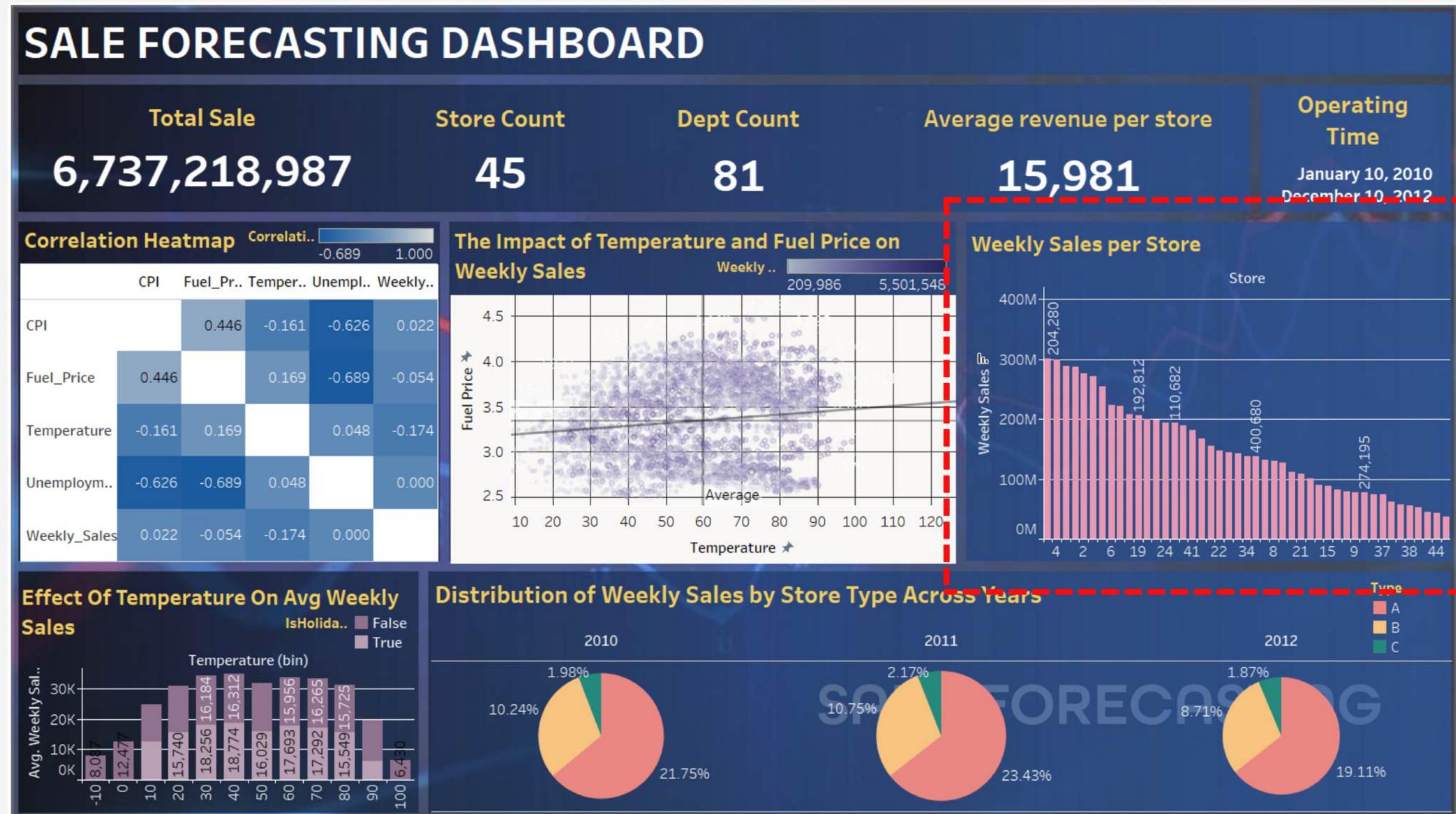


Hình: Biểu đồ phân tán

- + 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, điều này cho thấy rằng thời tiết ôn hòa có thể là điều kiện lý tưởng để thúc đẩy doanh số bán hàng

MQH 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.

4. TRỰC QUAN HÓA DỮ LIỆU



Hình: Sales Forecasting Dashboard

4. TRỰC QUAN HÓA DỮ LIỆU

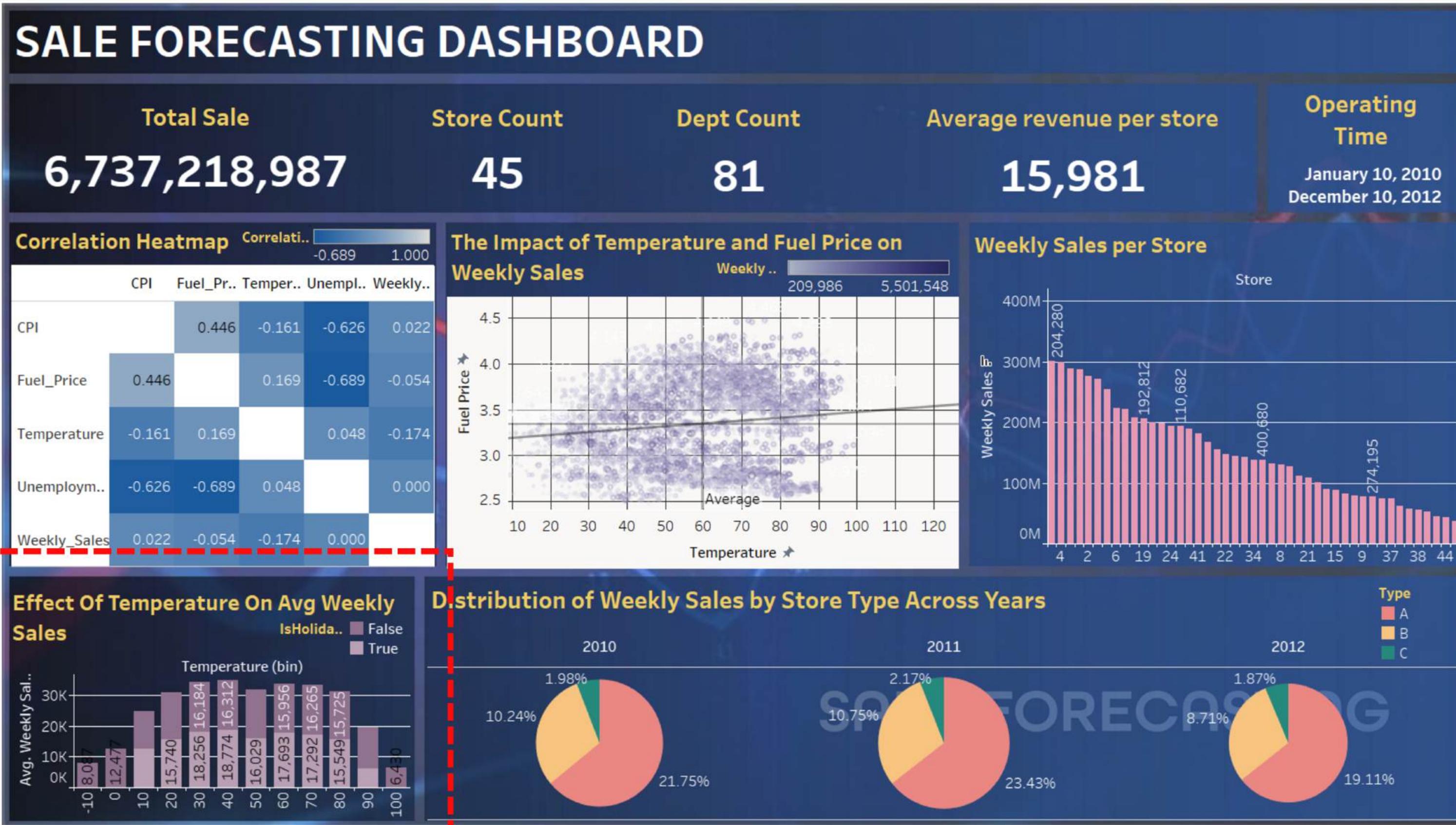
- + Biểu đồ hiển thị tổng doanh số hàng tuần của từng cửa hàng, với các cửa hàng được sắp xếp theo thứ tự giảm dần về doanh số.
- + Một số cửa hàng đạt hiệu suất vượt trội, nhiều cửa hàng khác vẫn chưa phát huy tiềm năng.



Hình: Biểu đồ cột

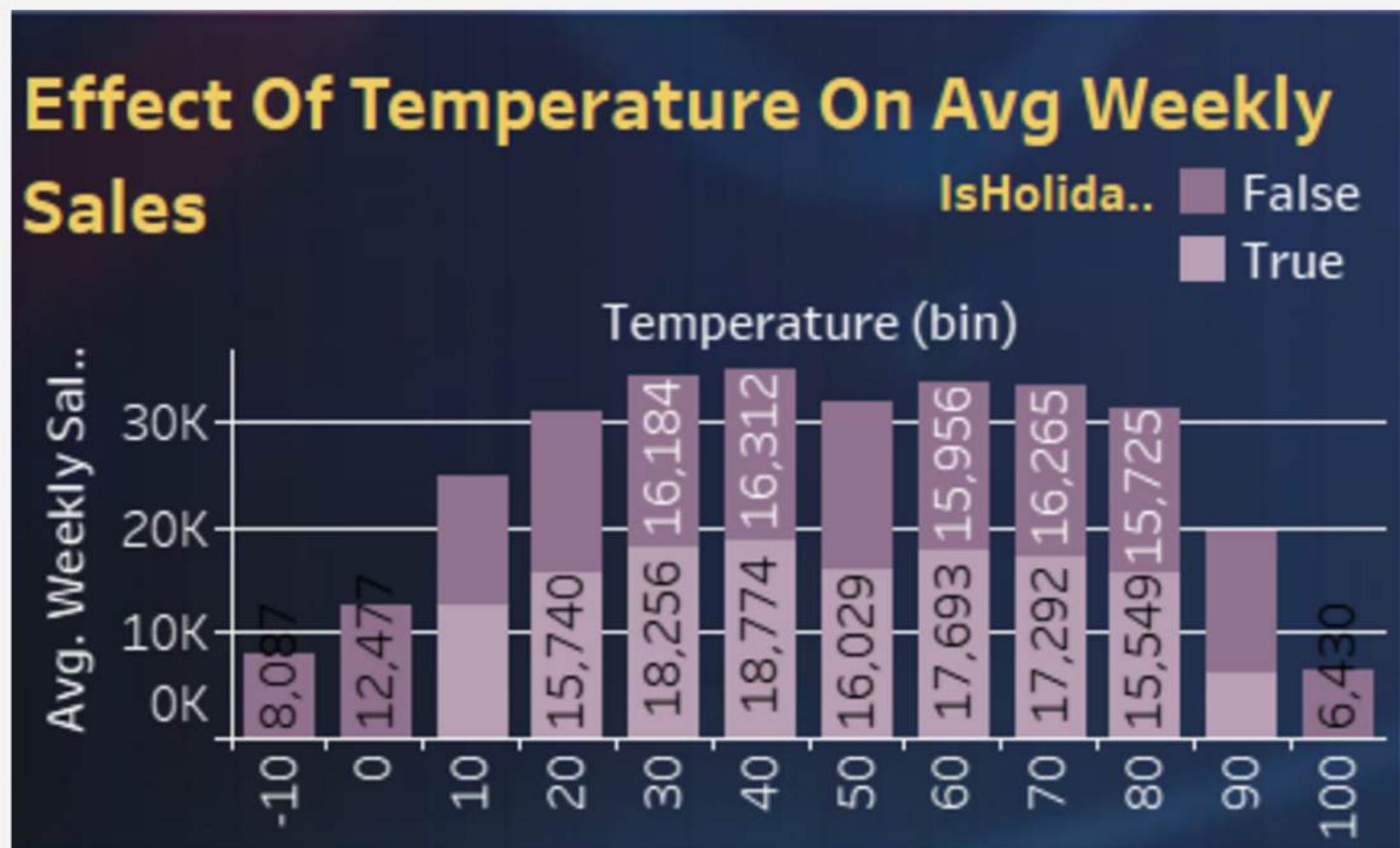
Việc phân tích sâu hiệu suất của từng cửa hàng và triển khai các chiến lược cải thiện phù hợp sẽ giúp tối ưu hóa tổng doanh thu.

4. TRỰC QUAN HÓA DỮ LIỆU



Hình: Sales Forecasting Dashboard

4. TRỰC QUAN HÓA DỮ LIỆU

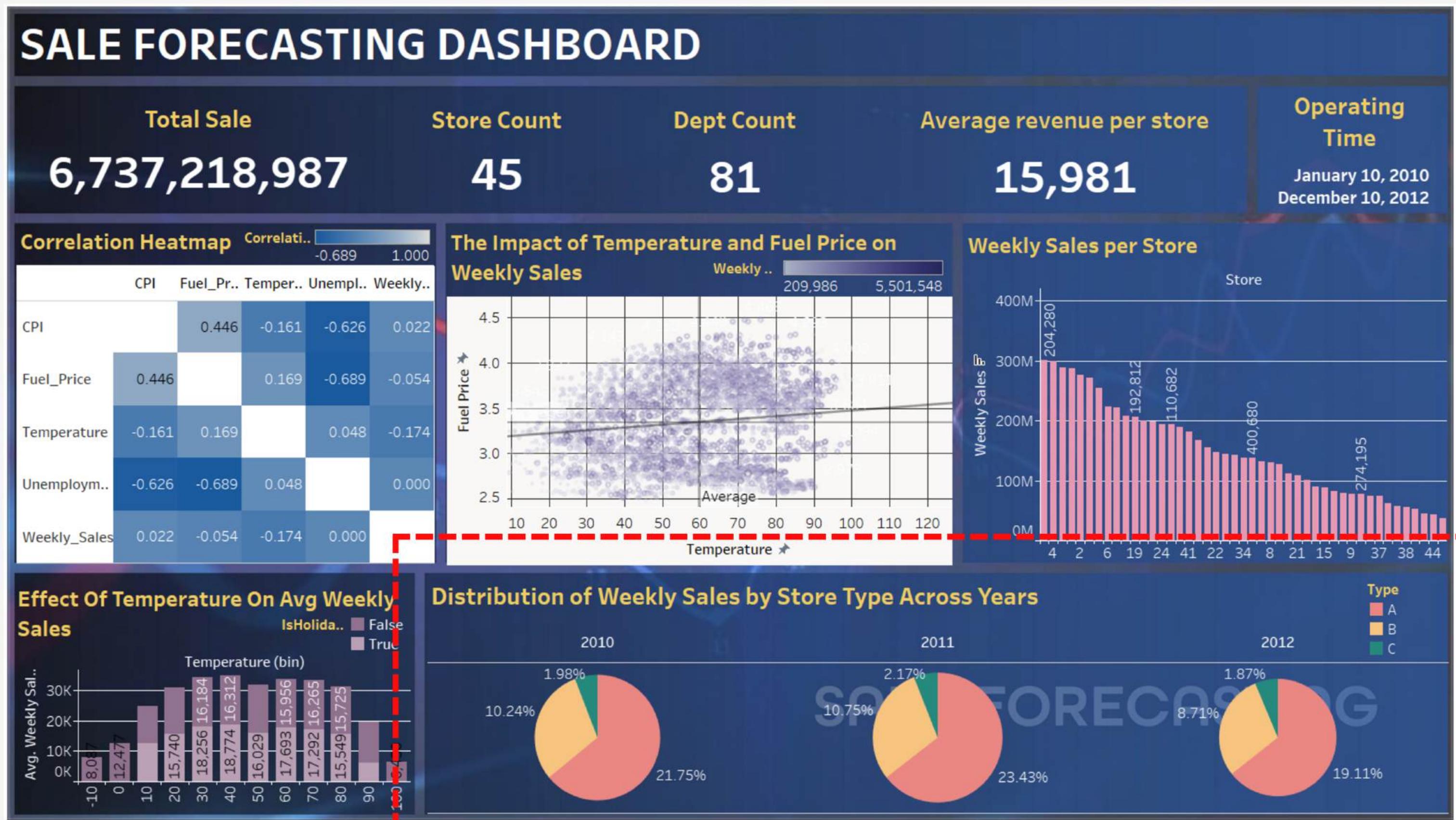


Hình: Biểu đồ cột chồng

- + 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.
- + Doanh số trong kỳ nghỉ lễ luôn cao hơn đáng kể so với ngày thường => nhiệt độ và kỳ nghỉ lễ có tác động đáng kể đến doanh số hàng tuần trung bình.

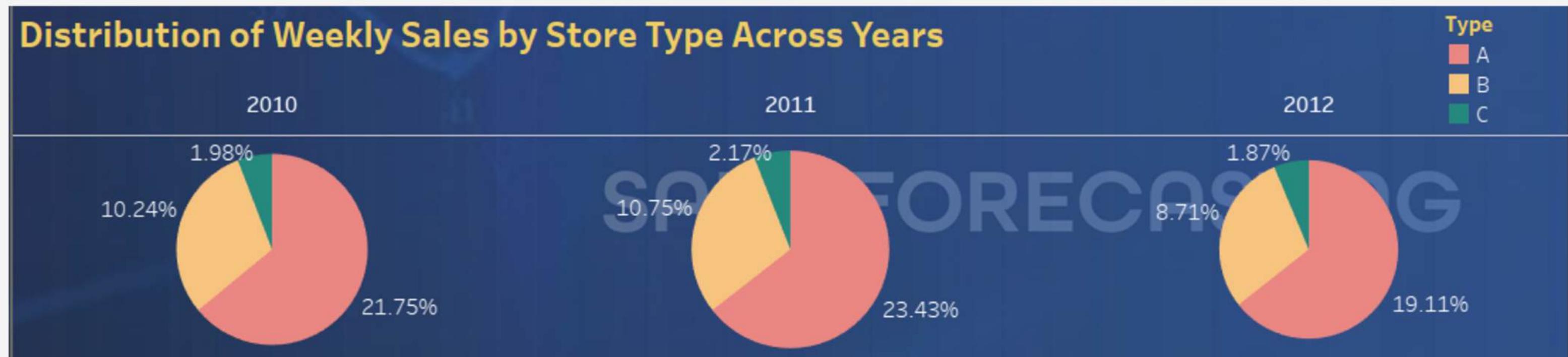
Cần chuẩn bị chiến lược phù hợp cho các giai đoạn thời tiết khắc nghiệt để duy trì hiệu quả kinh doanh.

4. TRỰC QUAN HÓA DỮ LIỆU



Hình: Sales Forecasting Dashboard

4. TRỰC QUAN HÓA DỮ LIỆU



Hình: Biểu đồ tròn

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.

+ 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.

4. TRỰC QUAN HÓA DỮ LIỆU

Tổng quan về doanh số, các yếu tố ảnh hưởng và hiệu suất theo từng cửa hàng.

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.

Tối ưu hóa chiến lược kinh doanh, cải thiện hiệu suất và nâng cao hiệu quả hoạt động.

5. ĐỀ XUẤT VÀ CHIẾN LƯỢC KINH DOANH

Tối ưu hóa chiến lược giảm giá (Markdowns)

- Tăng cường markdowns vào các dịp lễ lớn và quảng bá mạnh mẽ.
- Lập kế hoạch giảm giá chính xác và điều chỉnh mức giảm phù hợp.
- Kết hợp markdowns với các chương trình khuyến mãi khác.

Chiến lược tối ưu hóa hiệu suất theo mùa

- 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 cải thiện chiến lược quảng cáo.

Chiến lược tối ưu hóa các đơn vị kinh doanh

- Đầu tư vào các cửa hàng có hiệu suất cao, đồng thời hỗ trợ các cửa hàng yếu kém.
- 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.

5. ĐỀ XUẤT VÀ CHIẾN LƯỢC KINH DOANH

Chiến lược phát triển marketing nhắm mục tiêu

- Tập trung vào cửa hàng yếu kém, marketing địa phương, phân tích nhu cầu khách hàng.
- 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.

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

- Theo mùa: Tăng cường quảng bá các sản phẩm cho từng mùa, điều chỉnh mức tồn kho.
- Ứng phó với biến động kinh tế: Giá nhiên liệu giảm hoặc CPI thấp, nhập kho và điều chỉnh giá sản phẩm để kích cầu và ngược lại
- 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.

5. 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.

TỔNG KẾT

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.

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.

**CẢM ƠN THẦY VÀ CÁC BẠN ĐÃ
THEO DÕI!**