Exploratory Data Analysis

Retail Forecasting Case

Apr-29-2024

Yan-Ping Yu

Agenda

Executive Summary

Problem Statement

EDA

EDA Summary

Recommendations

Executive Summary

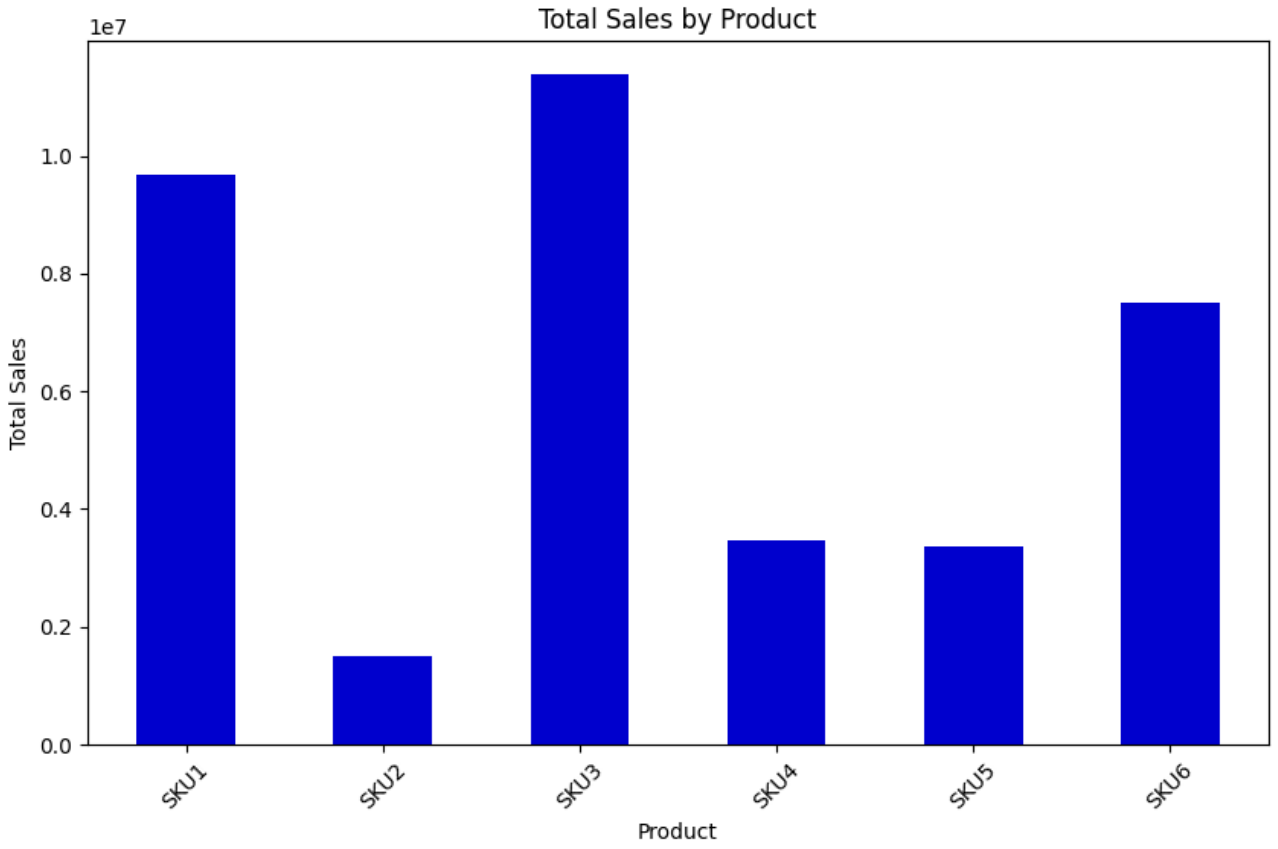
- The large company that is into the beverages business in Australia. They sell their products through various super-markets and also engage in heavy promotions throughout the year. Their demand is also influenced by various factors like holiday, seasonality. They needed forecast of each of products at item level every week in weekly buckets.

- This initiative is driven by the need to analyze historical time series data, incorporating various factors that influence demand, to forecast the quantity of items required by customers each week.

Problem of Statement

- The primary challenge is to develop multivariate forecasting models, utilizing machine learning or deep learning techniques, to accurately predict weekly demand for their products.

Total Sales of Product Analysis



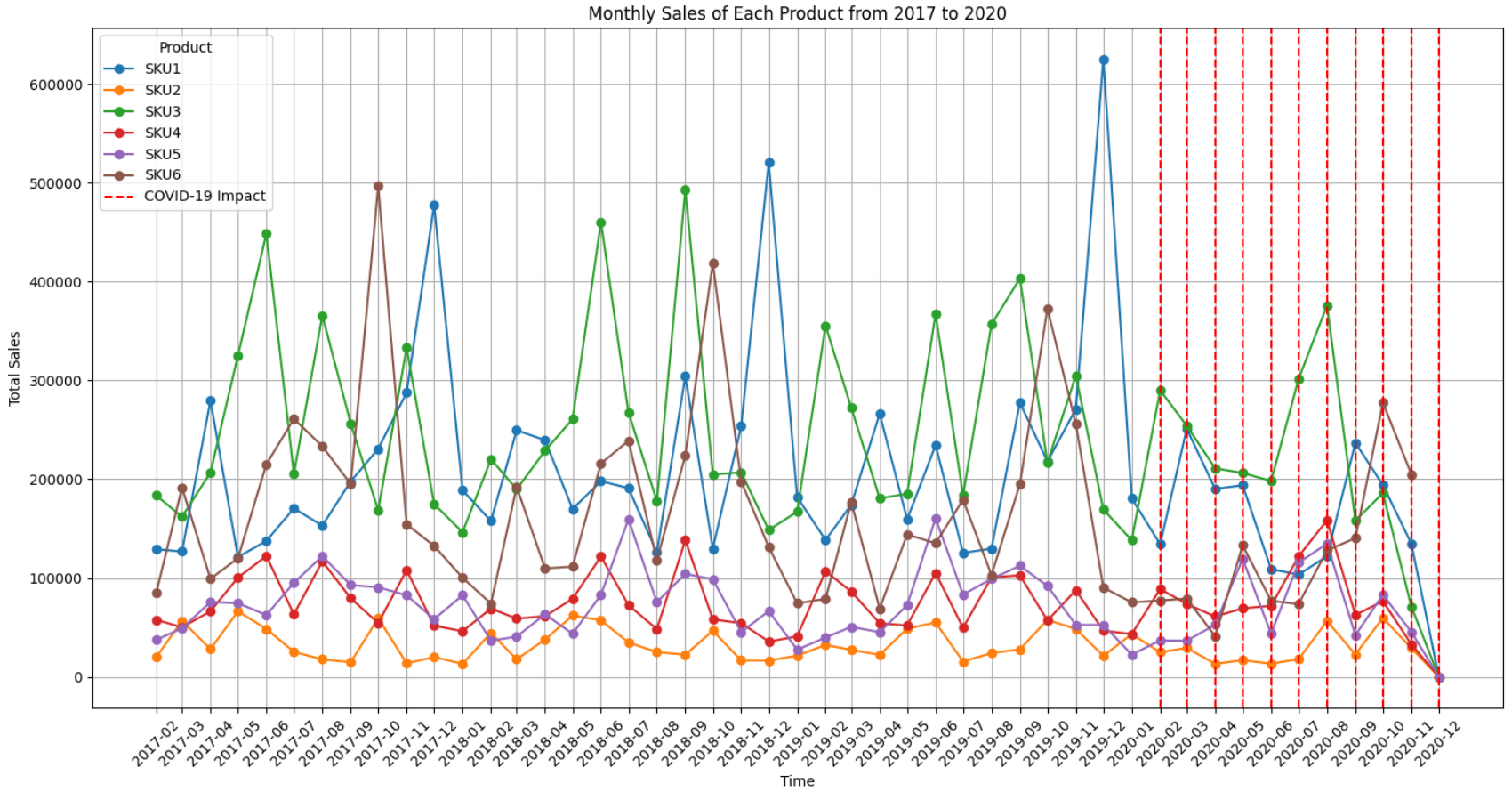
Top-Performing Products:

- SKU3
- SKU1
- SKU6

Customer preference concentrated on specific products

SKU3, SKU1, and SKU6 might be more aligned with market trends

Sales of Products Over Years Analysis

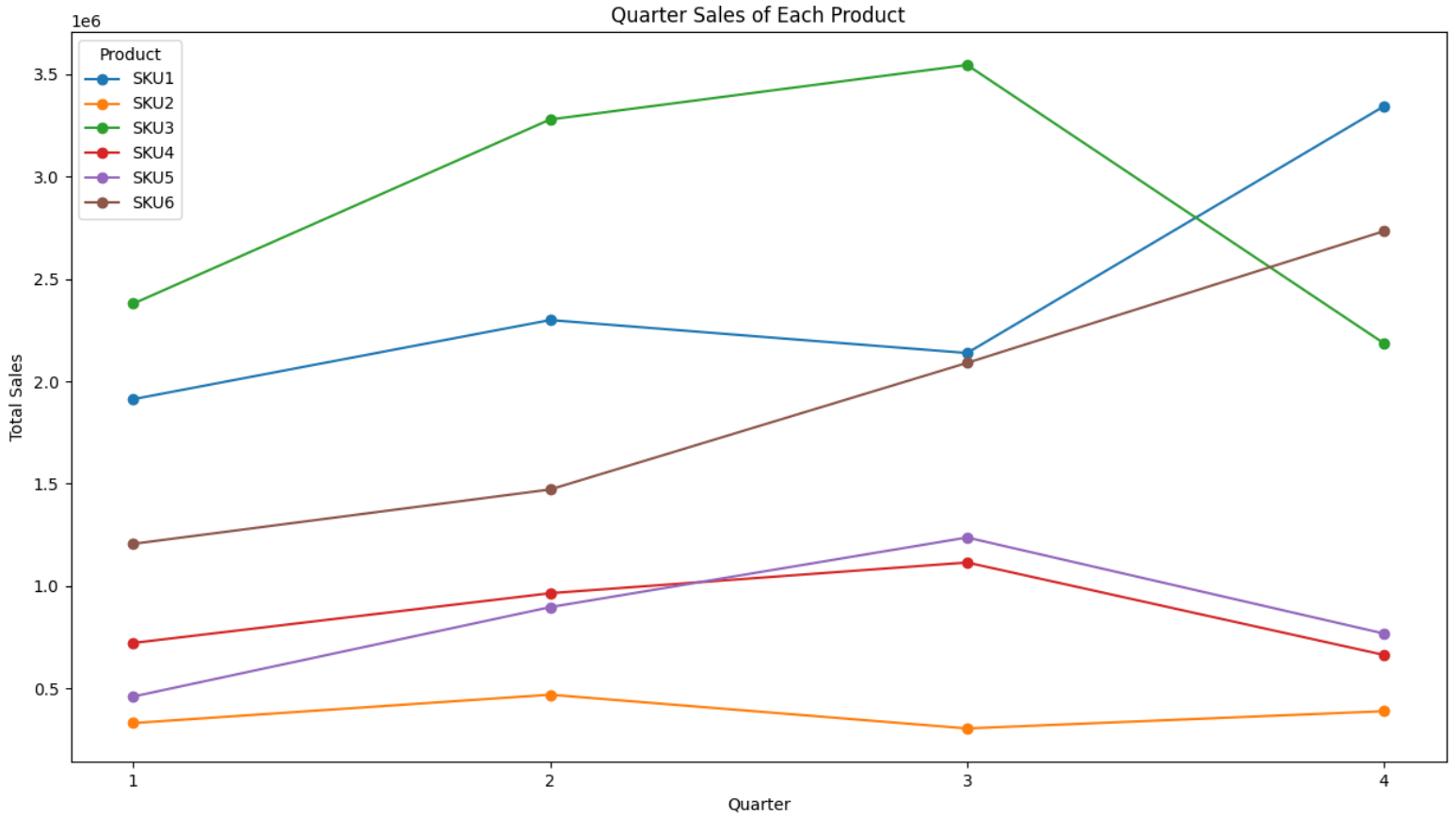


COVID-19 had a significantly negative impact on sales, especially on two highest total sales products, SKU1 and SKU6

Pandemic might have led to changes in consumer spending

SKU1 and SKU6 may be more sensitive to market changes and external disruptions

Quarter Sales Analysis

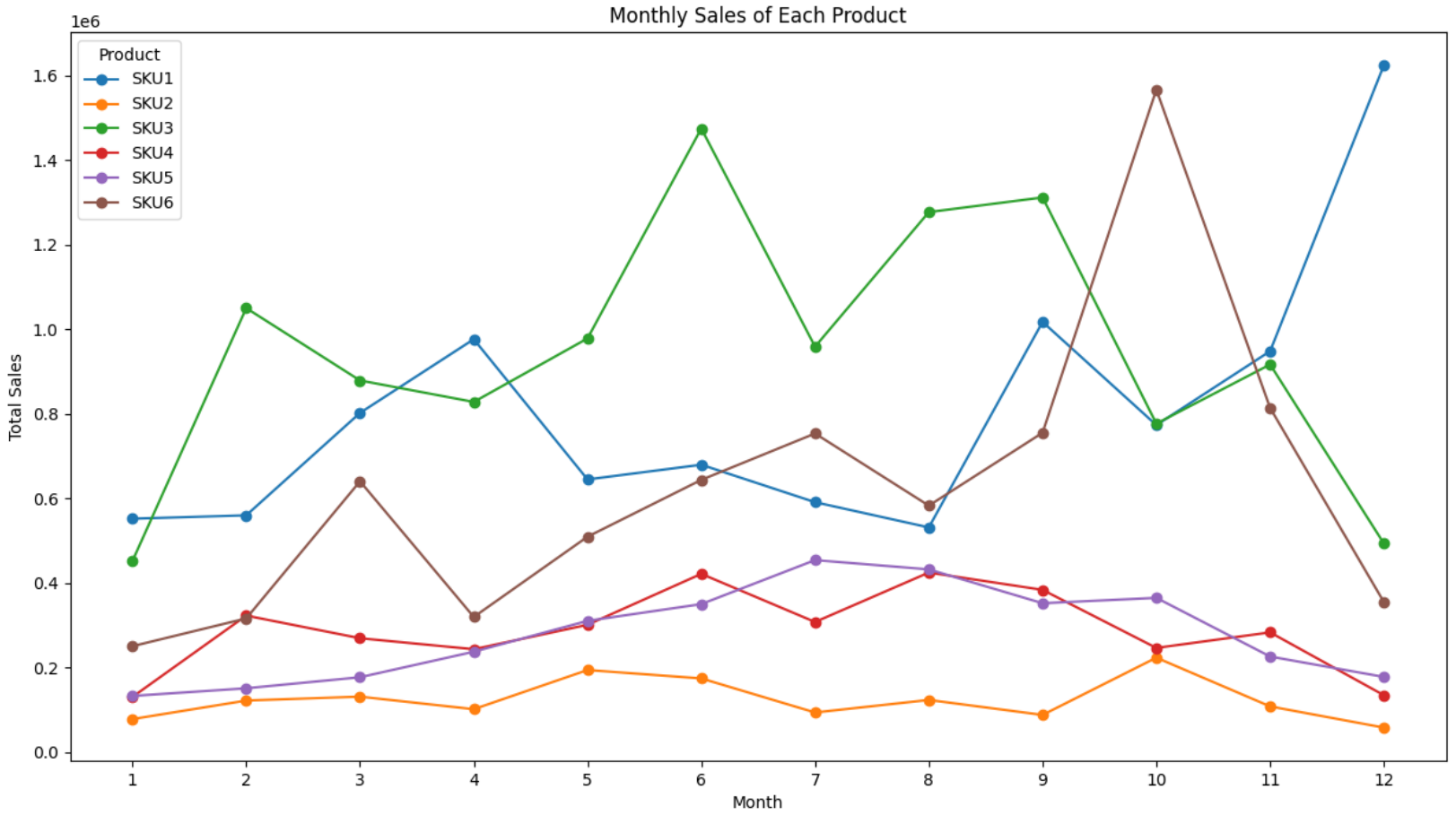


Each Product has its own Seasonality Patterns in Sales:

Peak Performance:

- SKU1: Q4
- SKU2: Q2
- SKU3: Q3
- SKU4: Q3
- SKU5: Q3
- SKU6: Q4

Monthly Sales Analysis

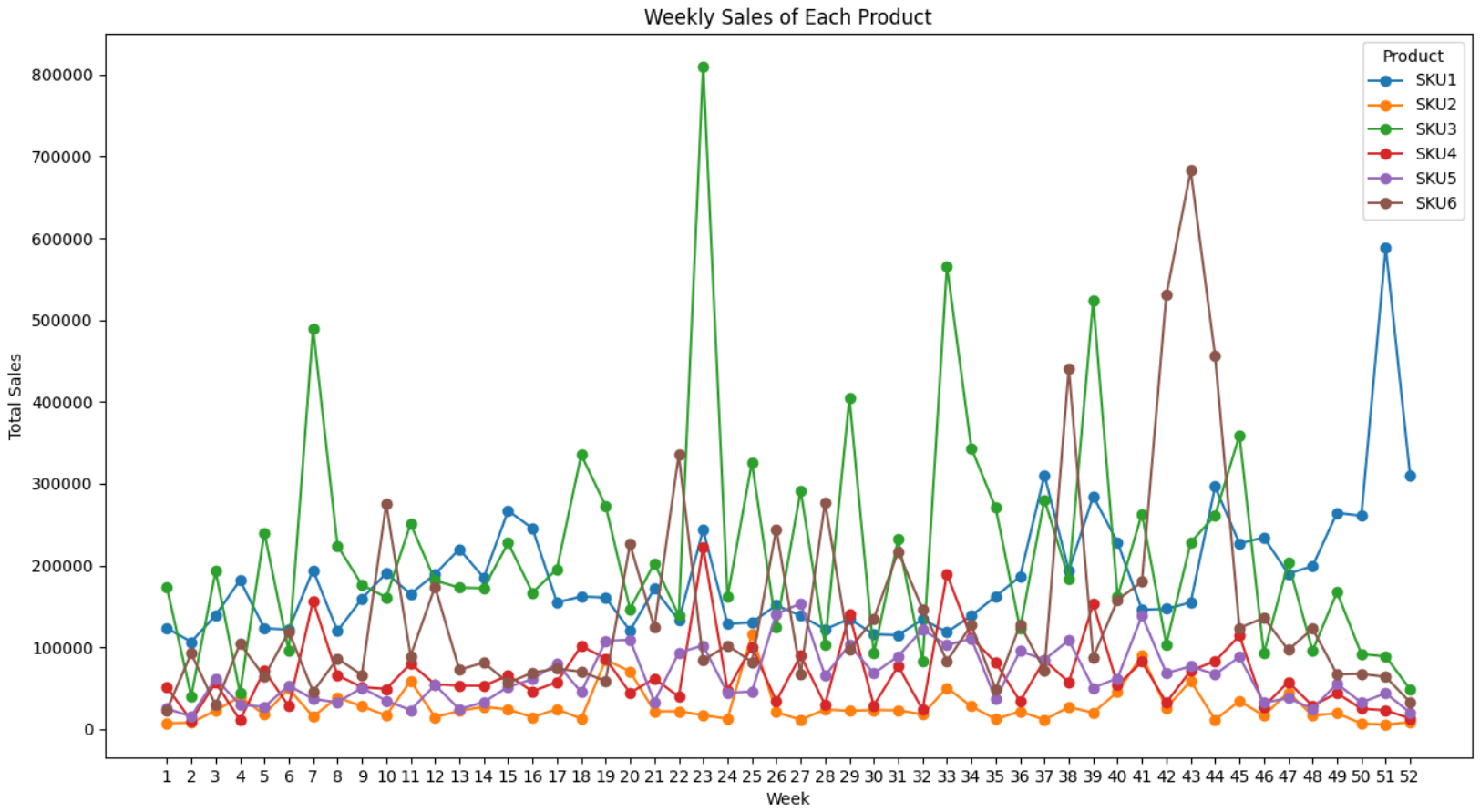


Each Product has its own Seasonality Patterns in Sales:

Peak Performance:

- **SKU1: Dec**
- SKU2: Oct
- SKU3: Jun
- SKU4: Jun
- SKU5: Jul
- **SKU6: Oct**

Weekly Sales Analysis



Each Product has its own Seasonality Patterns in Sales:

Peak Performance for Week:

- SKU1: 51st
- SKU2: 25th
- SKU3: 23rd
- SKU4: 23rd
- SKU5: 27th
- SKU6: 43rd

If sales indicate seasonality, we can…

1. Predict Sales Pattern

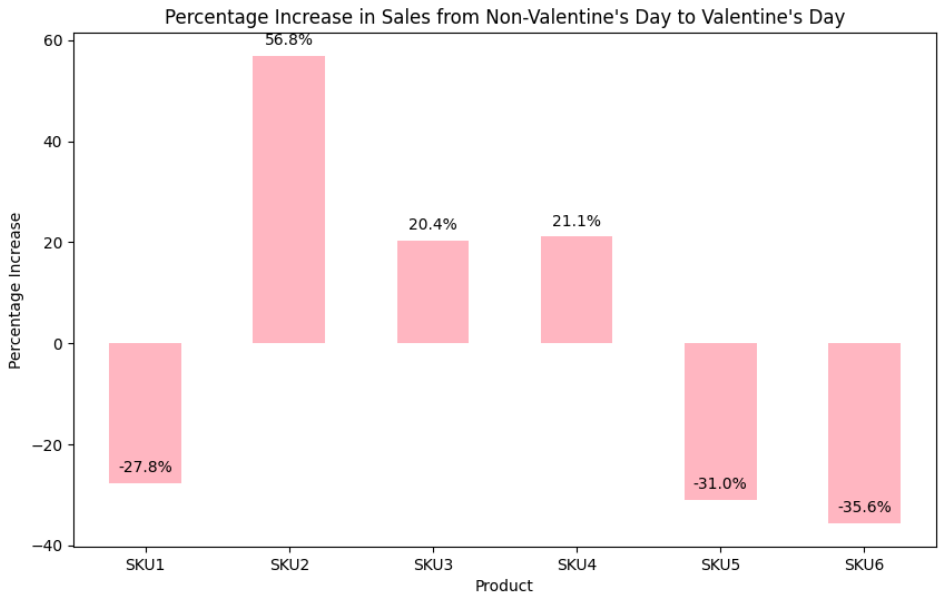
2. Improve Inventory Management

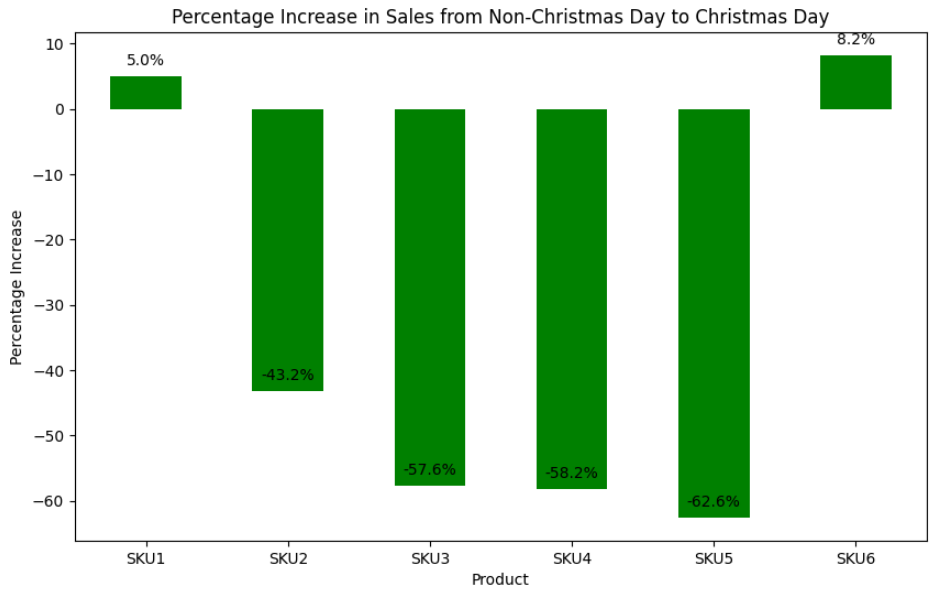
E.g. prompting the company to stock up ahead of anticipated sales increases and manage stock more conservatively during slower periods

3. Design Marketing & Promotions Plan

E.g. Marketing efforts can be timed to coincide with seasonal peaks to maximize their impact. Promotional activities could be planned to boost sales during typically slow periods.

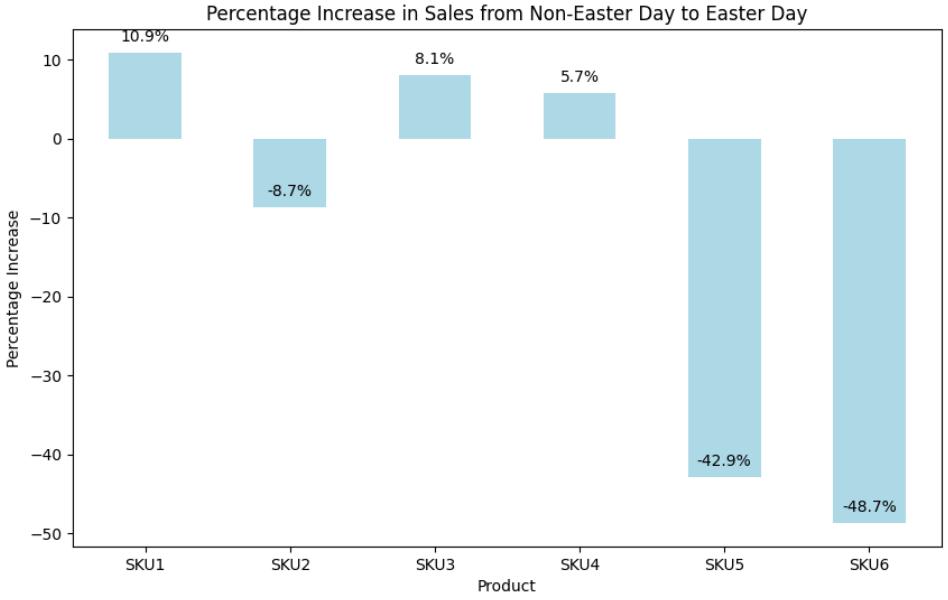
Sales in Non-holiday v.s. Holiday





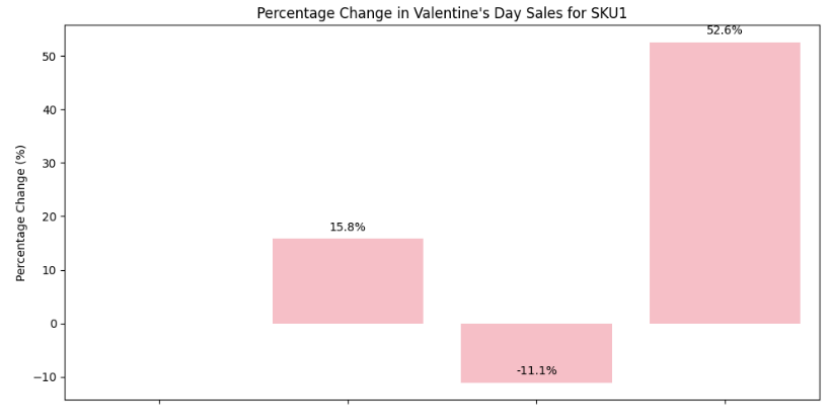
1. SKU2 sales increase the most during Valentine's Day.
2. SKU6 sales decrease the most during Easter Day.
3. Most product sales decrease significantly during Christmas.

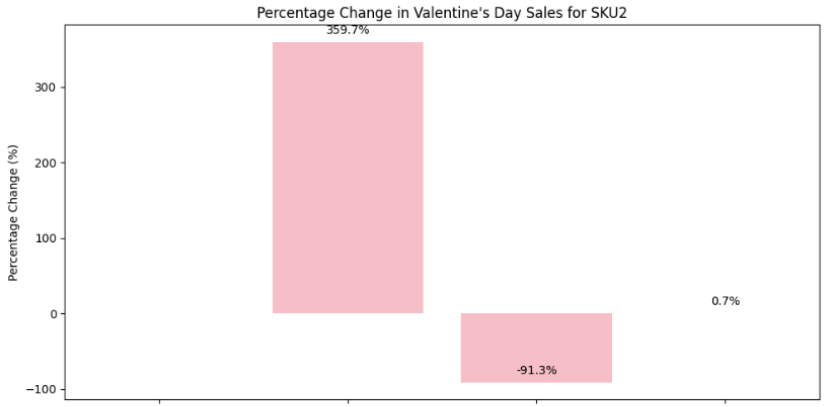
A sales increase/decrease during holidays can inform inventory & marketing decisions

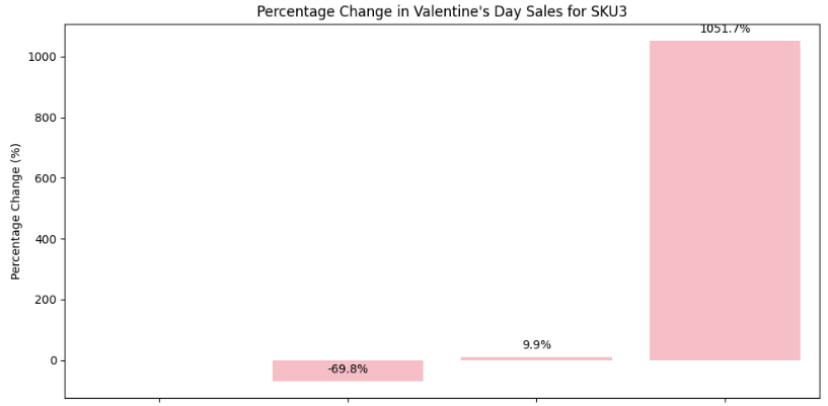


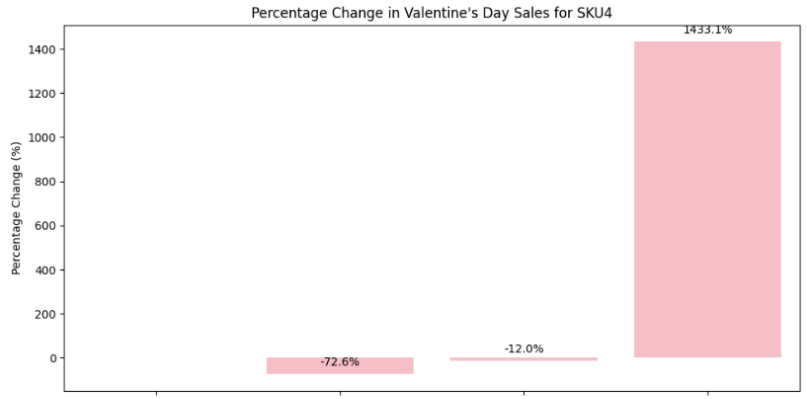
	Sales in Valentine's	Sales in Easter	Sales in Christmas
SKU1	Decrease	**Increase**	**Increase**
SKU2	**Increase**	Decrease	Decrease
SKU3	**Increase**	**Increase**	Decrease
SKU4	**Increase**	**Increase**	Decrease
SKU5	Decrease	Decrease	Decrease
SKU6	Decrease	Decrease	**Increase**

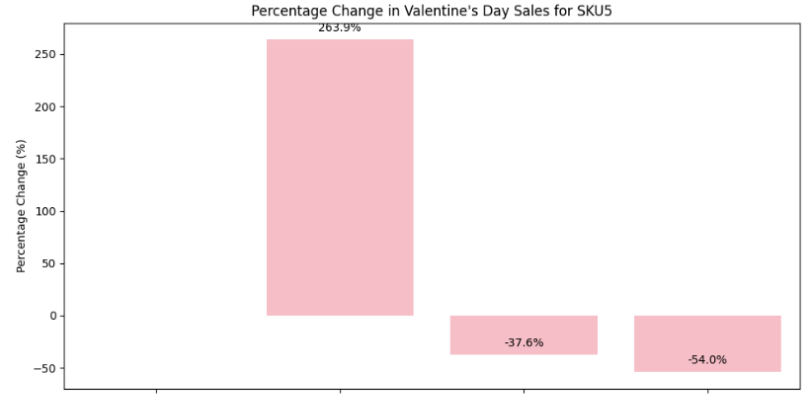
Yearly Sales Change in Valentine's Day

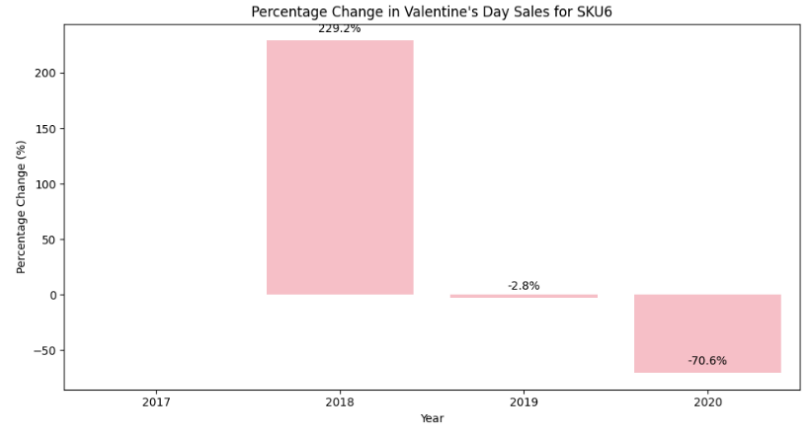








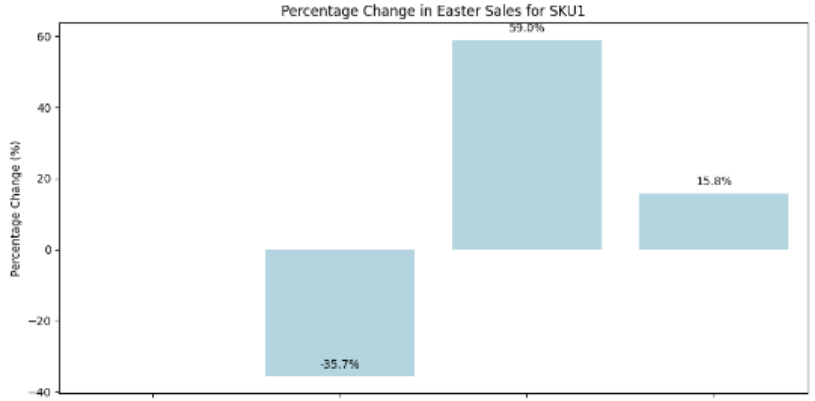


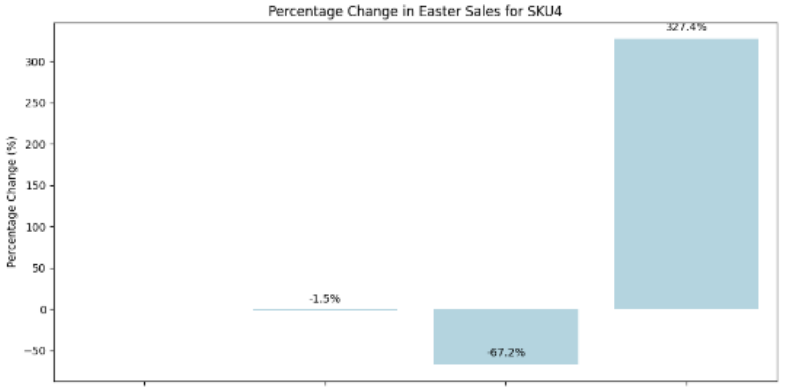


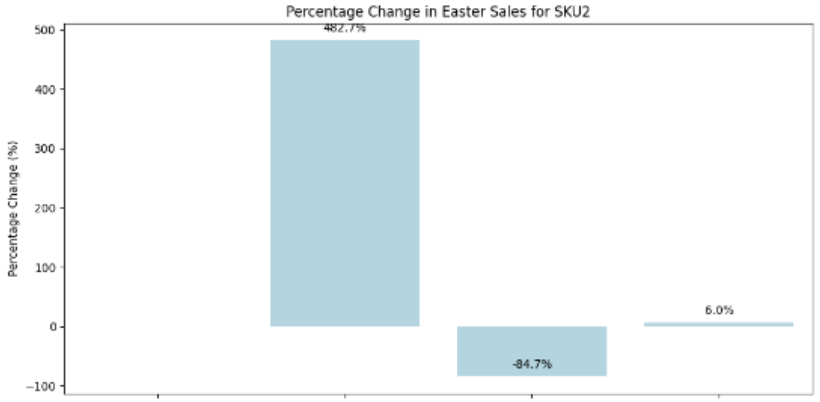
	Yearly Sales Trend
SKU1	**Increase**
SKU2	Decrease
SKU3	**Increase**
SKU4	**Increase**
SKU5	Decrease
SKU6	Decrease

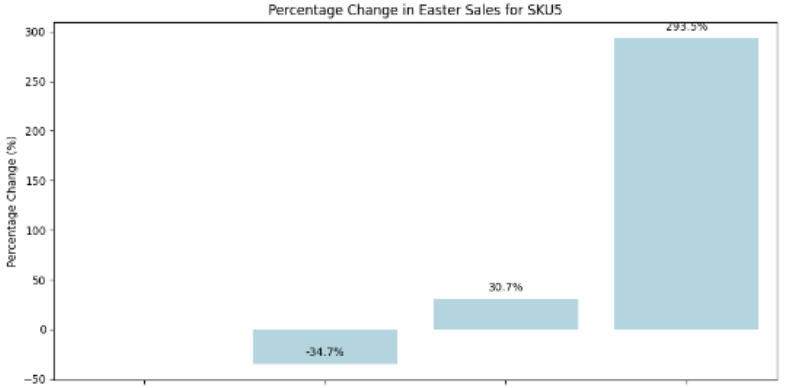
SKU1, SKU3, and SKU4 seem to align more closely with customer preferences these days and also demonstrate resilience during the COVID-19 period

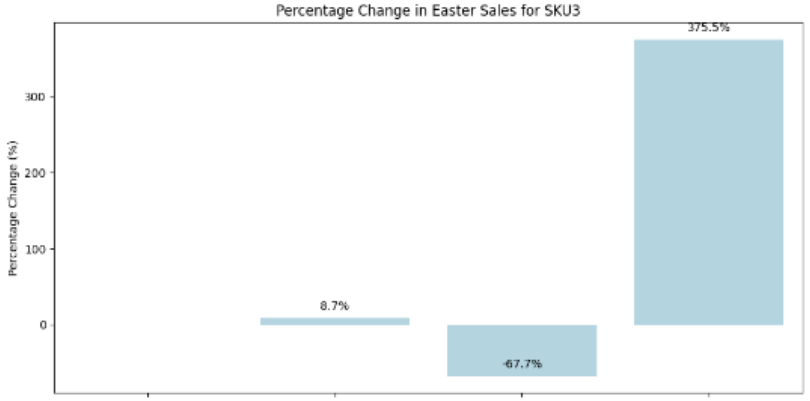
Yearly Sales Change in Easter Day

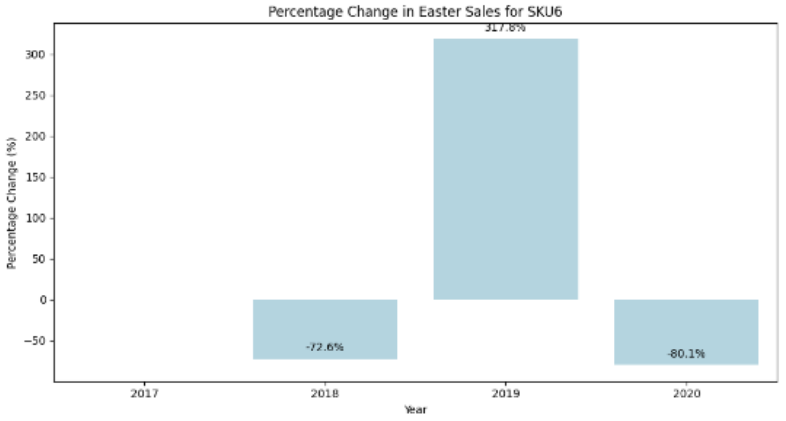








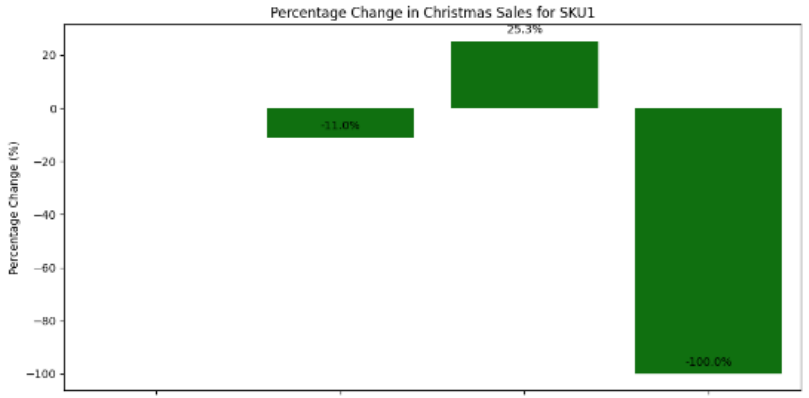


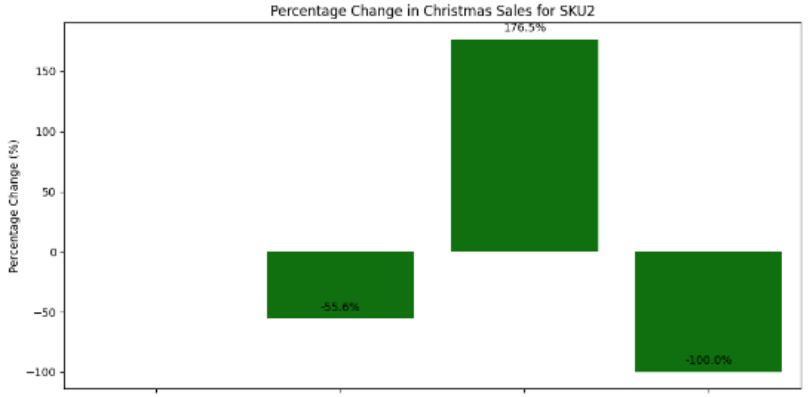


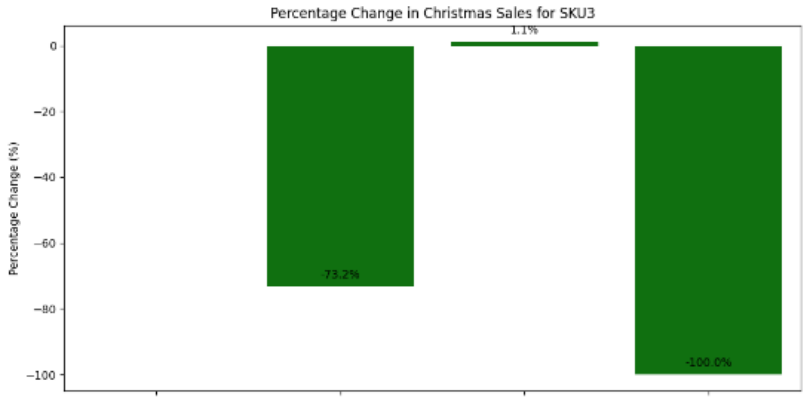
	Yearly Sales Trend
SKU1	**Increase**
SKU2	Decrease
SKU3	**Increase**
SKU4	**Increase**
SKU5	**Increase**
SKU6	Fluctuate

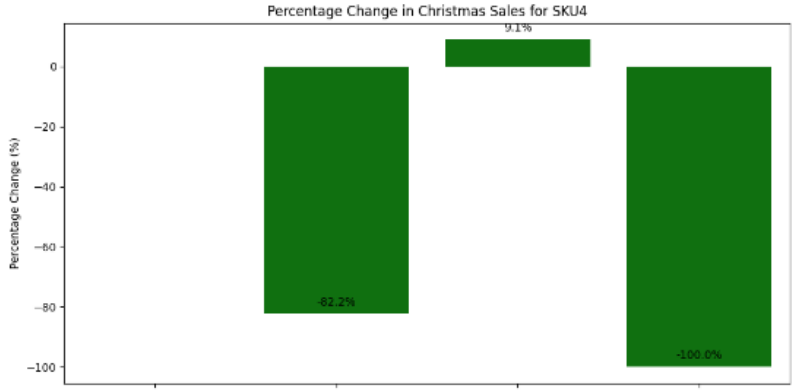
SKU3, SKU4, and SKU5 show high sales during COVID-19 period, indicating these products might be considered essential or high-need items during the pandemic

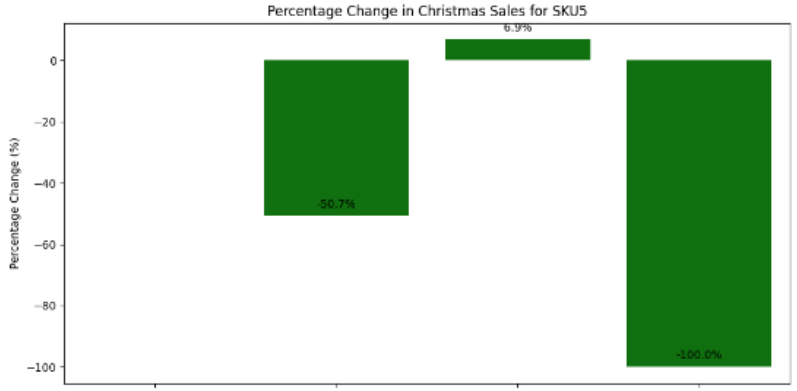
Yearly Sales Change in Christmas Day

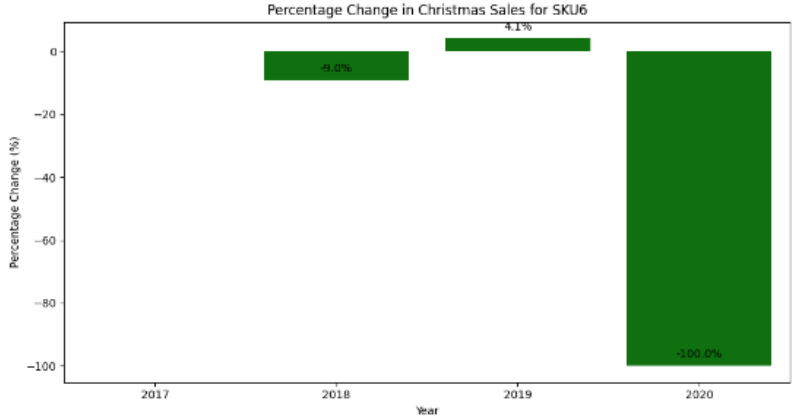










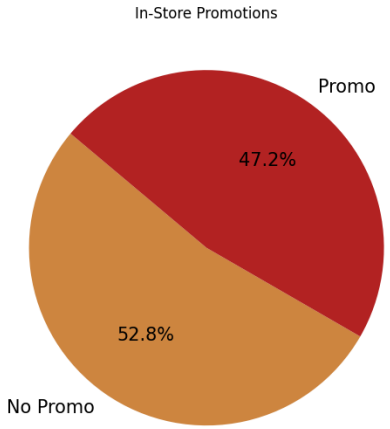


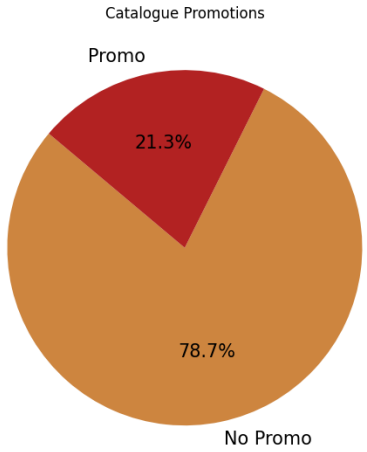
	Yearly Sales Trend from 2018 to 2019	Yearly Sales Trend considered 2020 COVID-19 period
SKU1	Increase	Decrease
SKU2	Increase	Decrease
SKU3	Increase	Decrease
SKU4	Increase	Decrease
SKU5	Increase	Decrease
SKU6	Increase	Decrease

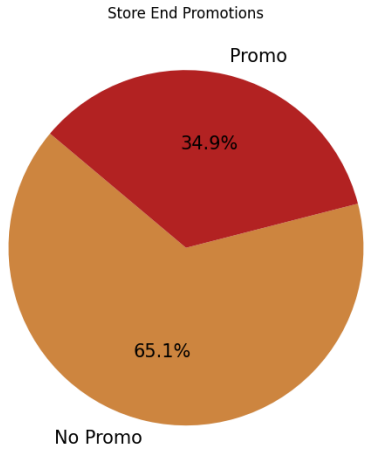
The sales for all products on Christmas day were completely influenced by COVID-19. Our client's physical stores might shut down during this time

Promotion Analysis

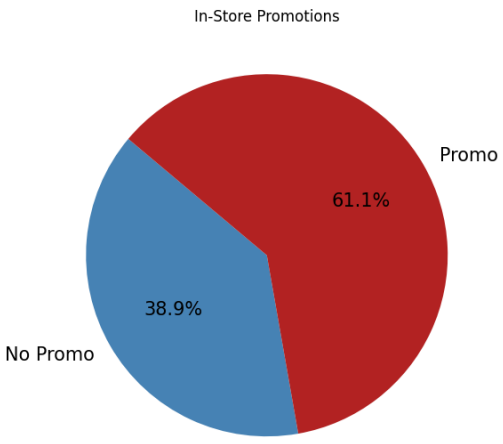
Percentage of Count for Different Types of Promotion

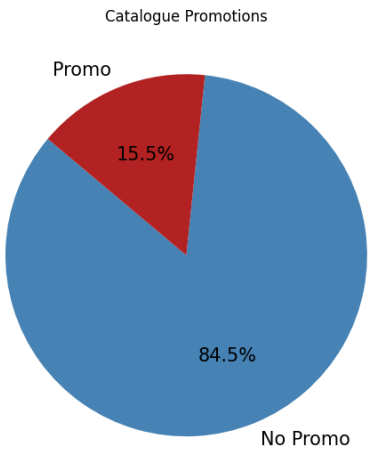


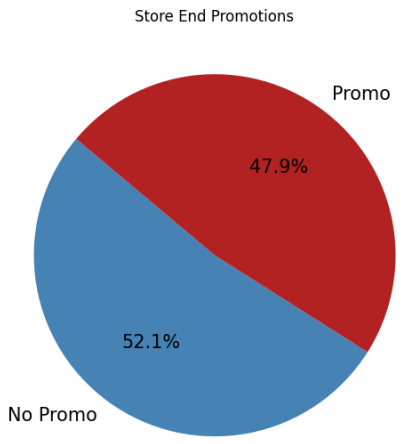




Percentage of Sales for Different Types of Promotion





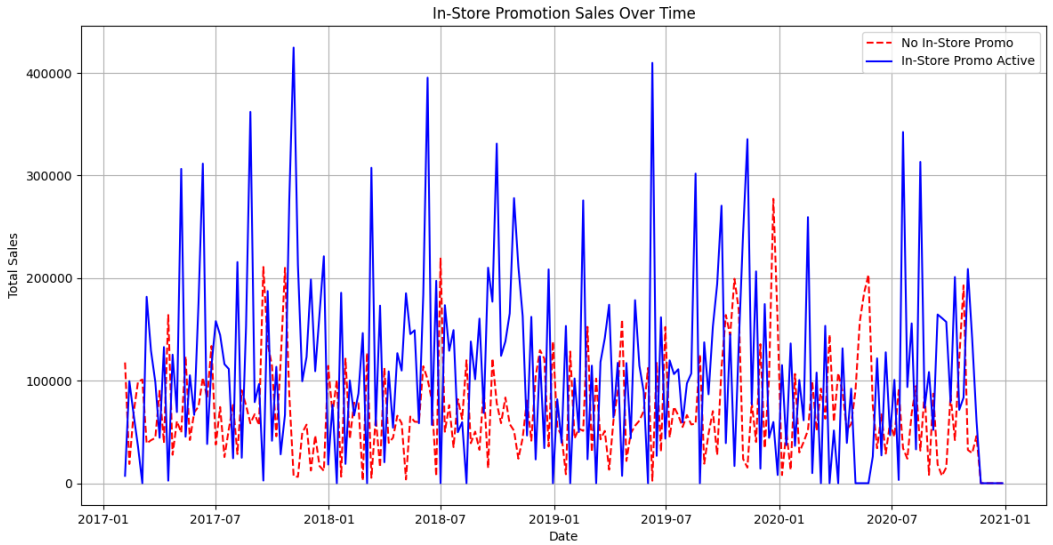


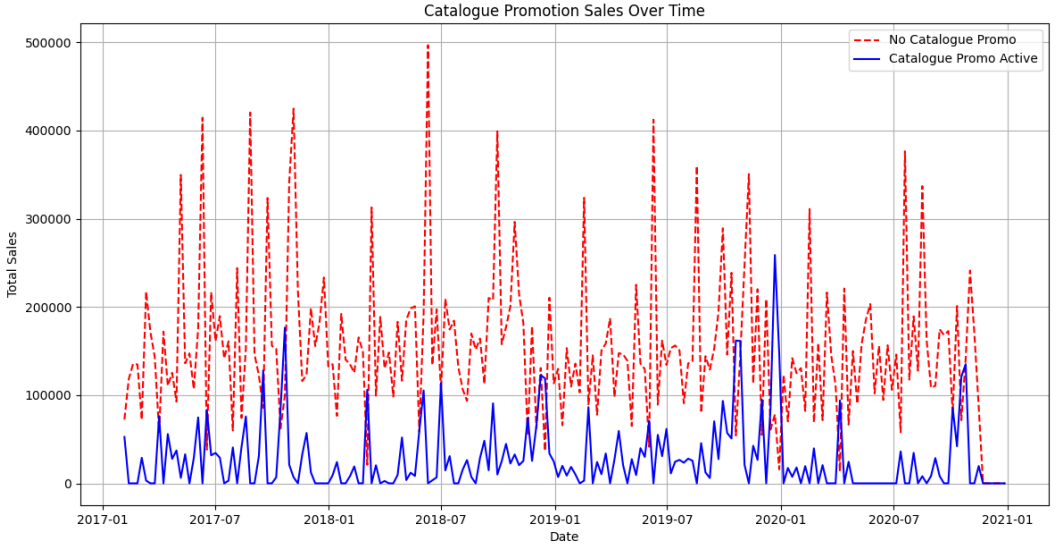
In-Store Promotion:
Sales per promotion transaction are the high

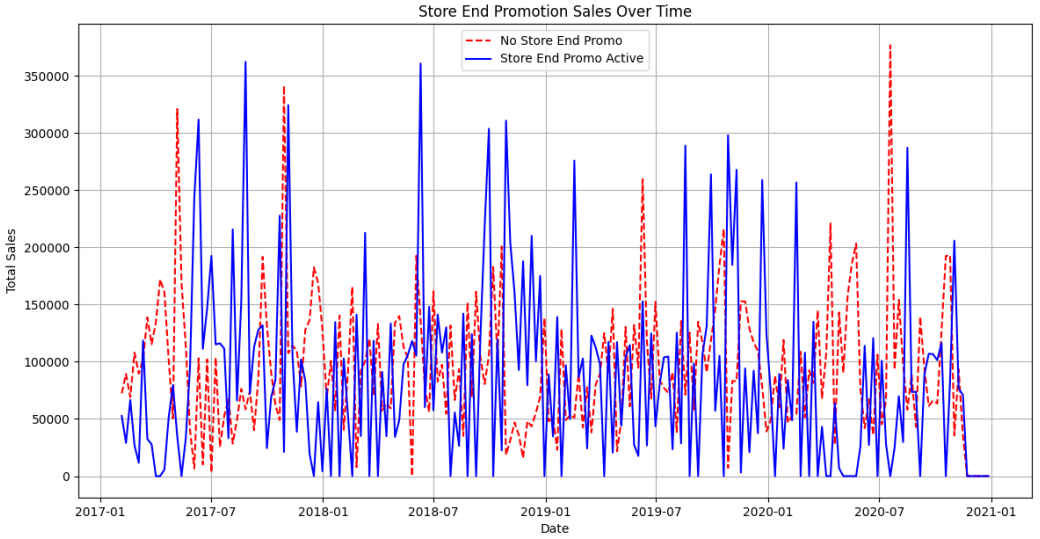
Catalog Promotion:
Sales per promotion transaction are the lowest

Store-End Promotion:
Sales per promotion transaction are high

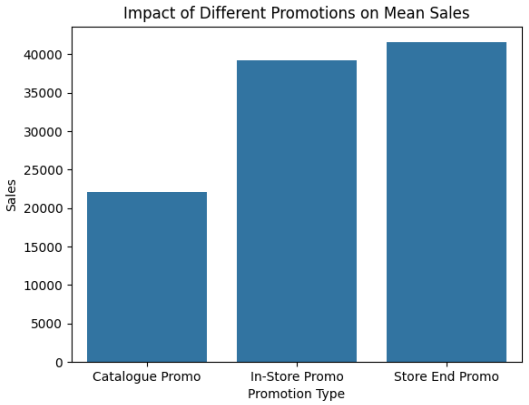
Promotion Analysis Over Time







The sales decreased significantly when catalog promotions were active, suggesting this promotion strategy might be ineffective and may lead to high cost.



Recommendation

- **Targeted Inventory Management**: Align inventory levels with seasonal demand patterns, ensuring higher stock during peak sales periods like Q4 for SKU1 and SKU6, and Q3 for SKU3 and SKU4.
- **Refined Promotional Strategies**: Given the indication that catalog promotions may be ineffective, consider reallocating marketing budgets to more effective promotion channels or redesigning the catalog strategy to better capture customer interest.
- **Diversified Product Strategy**: For products like SKU2, which show decreased sales during certain holidays, develop strategies to diversify their appeal or bundle them with more popular items.
- **Enhanced Digital Presence**: With physical store sales impacted by COVID-19, enhance the online shopping experience and digital presence to capture lost in-store sales and meet changing consumer behaviors.
- **Leverage Data for Forecasting:** Use historical sales data to predict future trends and prepare for upcoming seasonal peaks and lows in demand.

Model Recommendations

1. **ARIMA (AutoRegressive Integrated Moving Average)**

   Model that is well-suited for time series data with trends and seasonality.

2. **Decision Tree:**

   This model can capture complex non-linear relationships.

3. **Gradient Boosting Machines (GBM), e.g., XGBoost, LightGBM:**

   Boosting models that handle varied data types and distributions effectively.

4. **Recurrent Neural Networks (RNNs), particularly LSTM (Long Short-Term Memory) or GRU (Gated Recurrent Units):**

   Deep learning models designed to work with sequential data, capturing time dependencies.

Thank You