Data Science Project Report (Week 10)

Number of team members: 1
Member Name: Yan-Ping Yu
Email: yy3407@columbia.edu
Nationality: Taiwan
College: Columbia University
Specialization: Data Science

1. Problem description:

Our client operates within the beverage industry in Australia. The primary challenge is to develop multivariate forecasting models, utilizing machine learning or deep learning techniques, to accurately predict weekly demand for their products. 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.

2. Github Repo link

https://github.com/YanPing0227/yanping-my-repo

3. EDA performed on the data

1. Correlation Analysis:

Created a correlation matrix to understand the relationships between various numeric variables like sales, price discounts, Google mobility, and different types of promotions.

2. Sales Analysis:

Calculated total sales by product using a bar chart.
Grouped monthly sales by year and product, summing up to observe patterns and marked COVID-19 impact periods.
Conducted a quarterly and monthly sales analysis, grouping the data accordingly and visualizing it with line charts to identify trends and seasonal patterns.
Performed a weekly sales analysis to observe finer time-based trends.

3. Holiday Sales Impact:

Analyzed the percentage increase in sales from non-holiday to holiday periods for Valentine's Day, Easter, and Christmas and represented this using bar charts.
Utilized time adjustment based on the occurrence of these holidays to make year-over-year comparisons accurate.

4. Promotion Analysis:

Investigated the impact of different types of promotions (In-Store, Catalogue, and Store End) on mean sales.

Generated pie charts to represent the proportion of sales with and without promotions.

Summarized sales data by promotion status over time to analyze the effect of promotions on sales.

5. Time Series Adjustments:

Adjusted the year based on the occurrence of specific holidays like Valentine's Day, Easter, and Christmas to better align the sales data with the actual periods they affected.

6. Variable Visualization and Summarization:

Melted the dataset to long format to visualize the effect of promotions on sales with bar charts.

Generated pie charts to compare the proportion of sales affected by different promotion types.

Plotted line charts to show sales over time in relation to active promotions.

7. Sales Percentage Change Analysis:

Computed year-over-year percentage change in sales for significant holidays to assess the impact on each product.

Visualized these changes with bar charts to highlight which products were most and least affected during these periods.

Recommendations:

1. 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.

2. 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.

3. 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.

4. 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.

5. Leverage Data for Forecasting: Use historical sales data to predict future trends and prepare for upcoming seasonal peaks and lows in demand.