



Data Glacier

Your Deep Learning Partner

Forecast Project – EDA Results

Virtual Internship

11-8-2023

Agenda

Executive Summary

Problem Statement

Approach

EDA

EDA Summary

Recommendations

Model proposals

Executive Summary

This project involves forecasting sales through a systematic approach encompassing tasks such as data cleaning, variable relevance sorting, exploratory data analysis (EDA) with a time series focus, feature engineering, and model selection, training, error estimation, and tuning. The main goal is to predict sales accurately for informed business decisions. Data cleaning ensures data quality, followed by sorting variables by their impact on sales. EDA emphasizes time series patterns, aiding feature engineering. Model selection involves choosing the best-fit model, which is trained, evaluated for accuracy, and fine-tuned. The project aims to construct a reliable sales forecasting model for effective resource allocation and decision-making based on y target 'Sales'.

Problem Statement

The Client:

The large company who is into beverages business in Australia. They sell their products through various super-markets and also engage into 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.

Objective:

1. Build at least 4-5 multivariable forecasting model which included ML or Deep Learning based Model in PySpark leveraging parallel computing techniques.
2. Demonstrate best in class forecast accuracy (Forecast Accuracy = $1 - \text{Wt. MAPE}$ where $\text{Wt. MAPE} = \frac{\text{sum(Error)}}{\text{sum(Actual)}}$).
3. Write a code in such a way you run the model in least time.
4. Demonstrate explainability in the form of contribution of each variables
5. Leverage Feature Engineering concepts to derive more variables to gain accuracy improvement

Approach

What's the key feature to boost sales with discounts during drops?

Which top 3 products need investment due to high sales (SKU1, 3, 6)?

When does SKU1 excel with >10% sales during holidays?

How do In-Store promos affect SKU1-5 sales instantly?

Do Catalogue promos increase sales for all products next day?

Does Store-End promo spike SKU4 sales on the next day?

What's the daily sales range (95%) for each SKU (1-6)?

Does Google Mobility impact sales; should sales be closed?

What are optimal discount ranges for each SKU (1-6) to boost sales and profit balance?

Result 1

EDA

Price Discount is the most impactful feature within the dataset to optimize sales for every product, whenever sales are dropping low, it would be the best to offer discounts.



Random Forest:

```
Price Discount (%): 56.14%
year: 13.91%
day: 10.11%
month: 9.55%
Store End Promo: 4.52%
Google_Mobility: 1.75%
In-Store Promo: 1.47%
V_DAY: 1.39%
Catalogue Promo: 0.87%
EASTER: 0.13%
Covid_Flag: 0.09%
CHRISTMAS: 0.07%
```

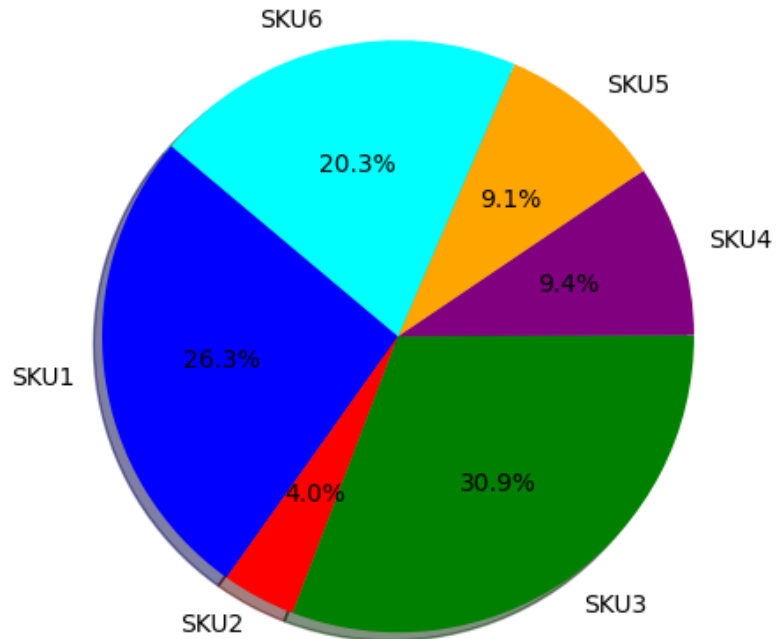
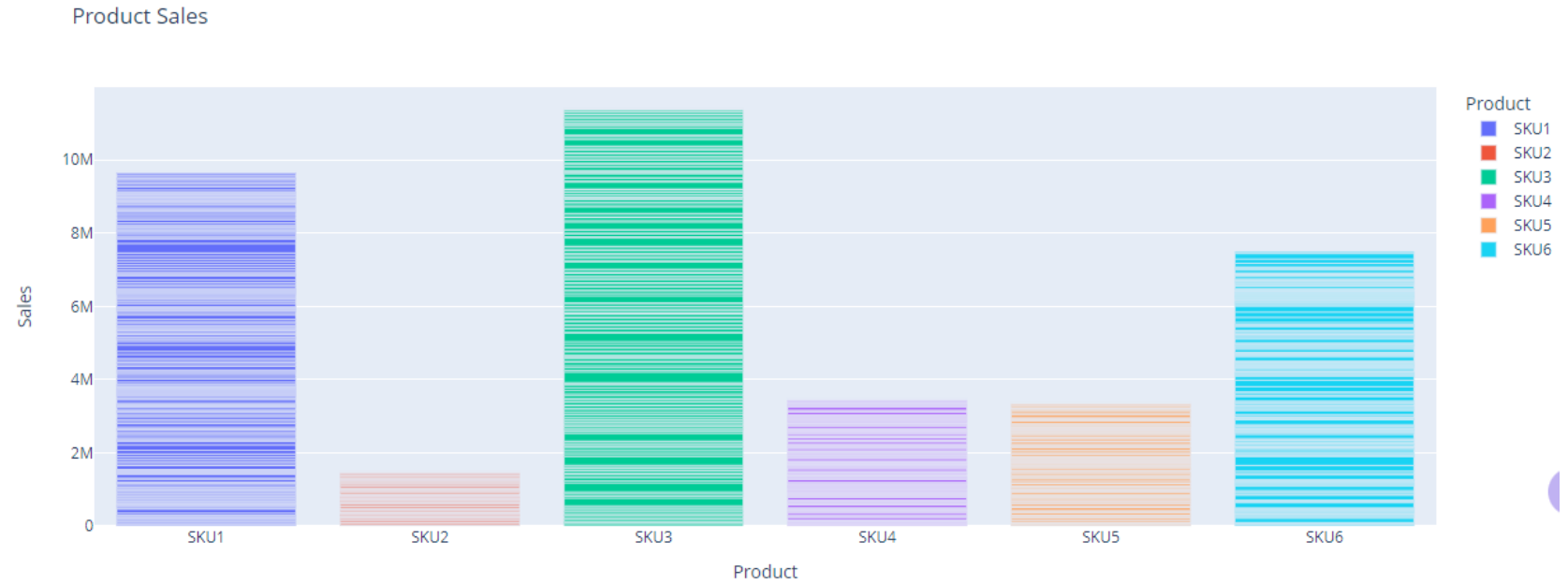
Gradient Boost:

```
Price Discount (%): 42.0%
month: 27.98%
year: 24.25%
day: 4.79%
Google_Mobility: 0.71%
Covid_Flag: 0.12%
Catalogue Promo: 0.08%
Store End Promo: 0.05%
In-Store Promo: 0.02%
EASTER: 0.0%
V_DAY: 0.0%
CHRISTMAS: 0.0%
```

Result 2

EDA

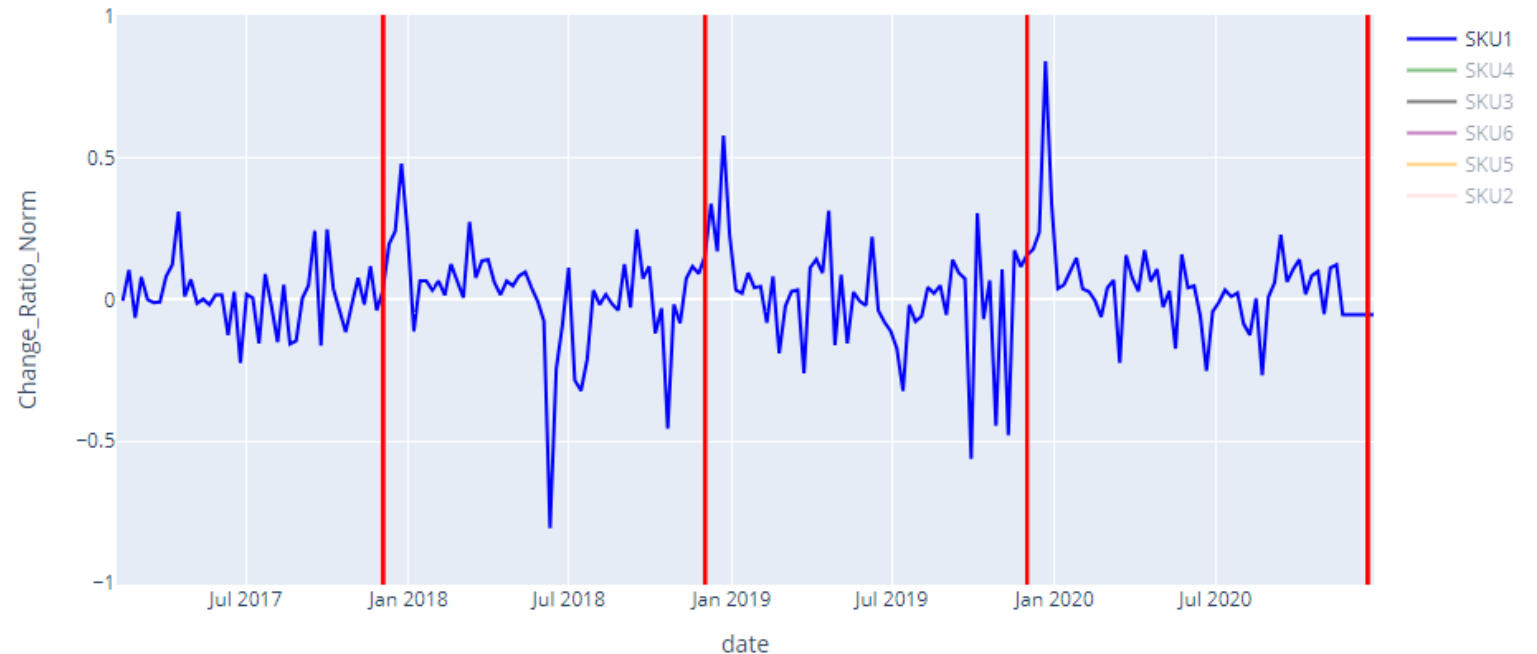
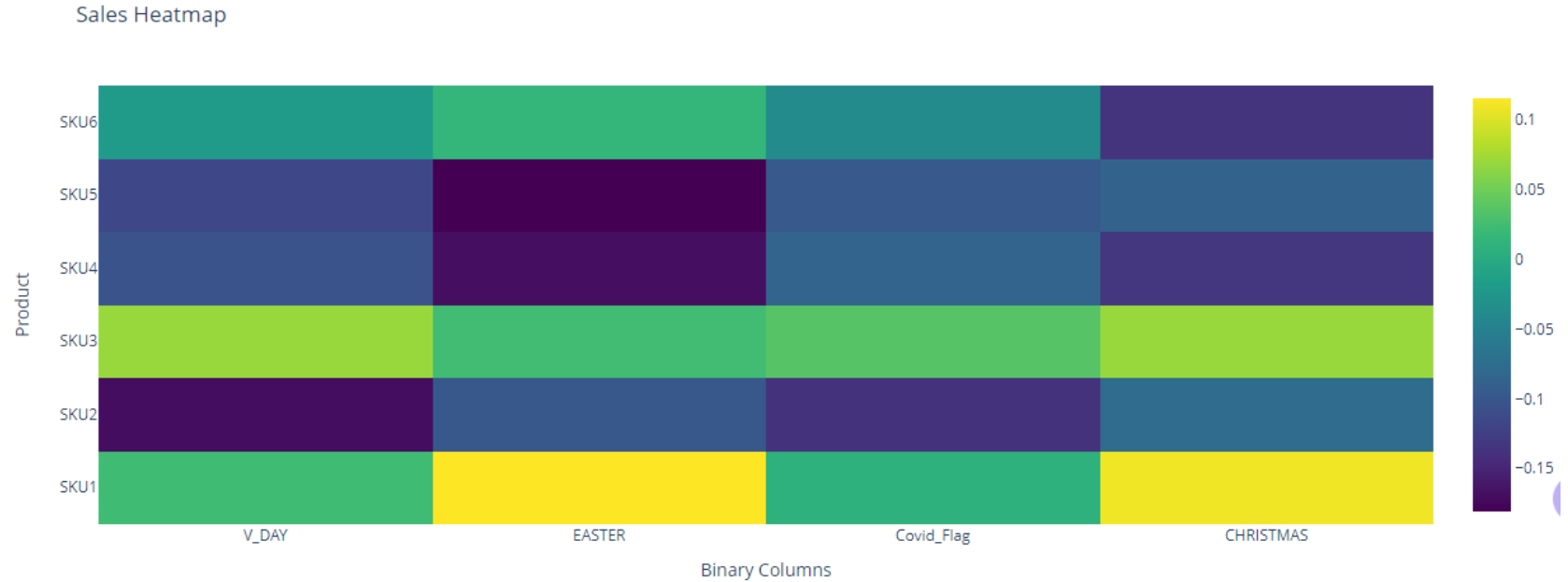
SKU1, 3 and 6 have the majority of the sales, investments should prioritize these 3 products.



Result 3

EDA

During Christmas and Easter, SKU1 is the optimal product for sales, it would have > 10% of it's standard sales rate during that day..

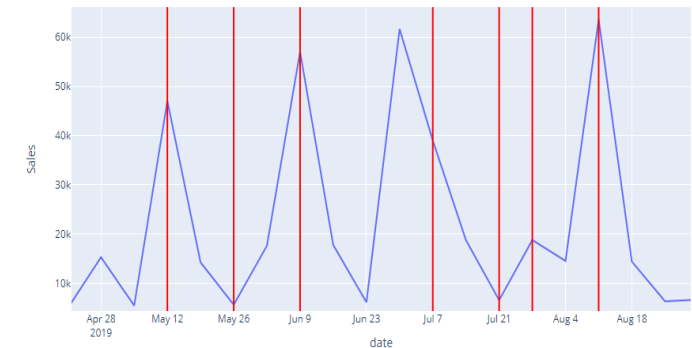
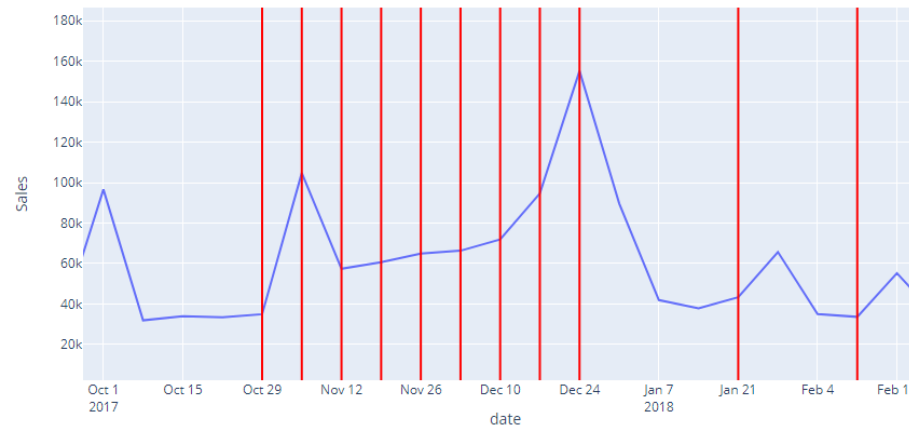
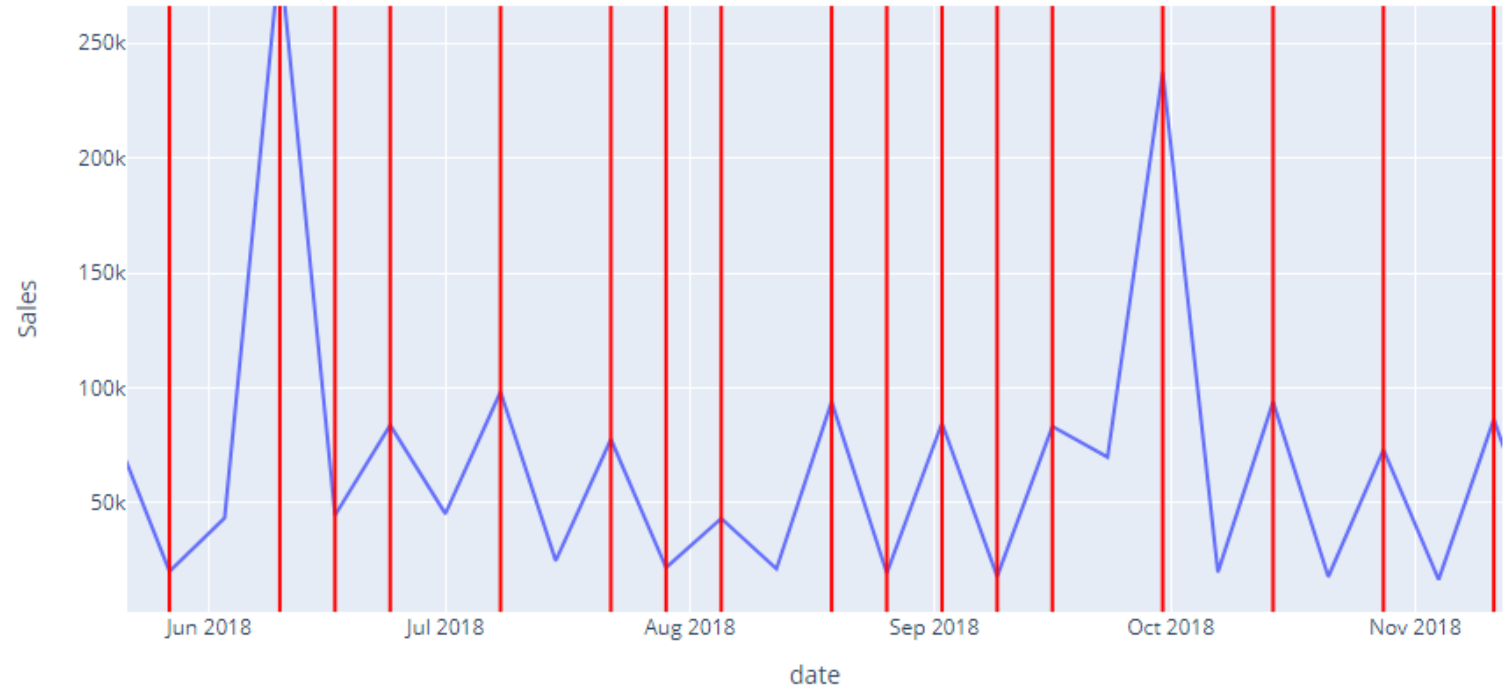


Result 4

EDA

For products from SKU1 to SKU5, offering In-Store promotions would increase sales the same day or the day after the promotion.

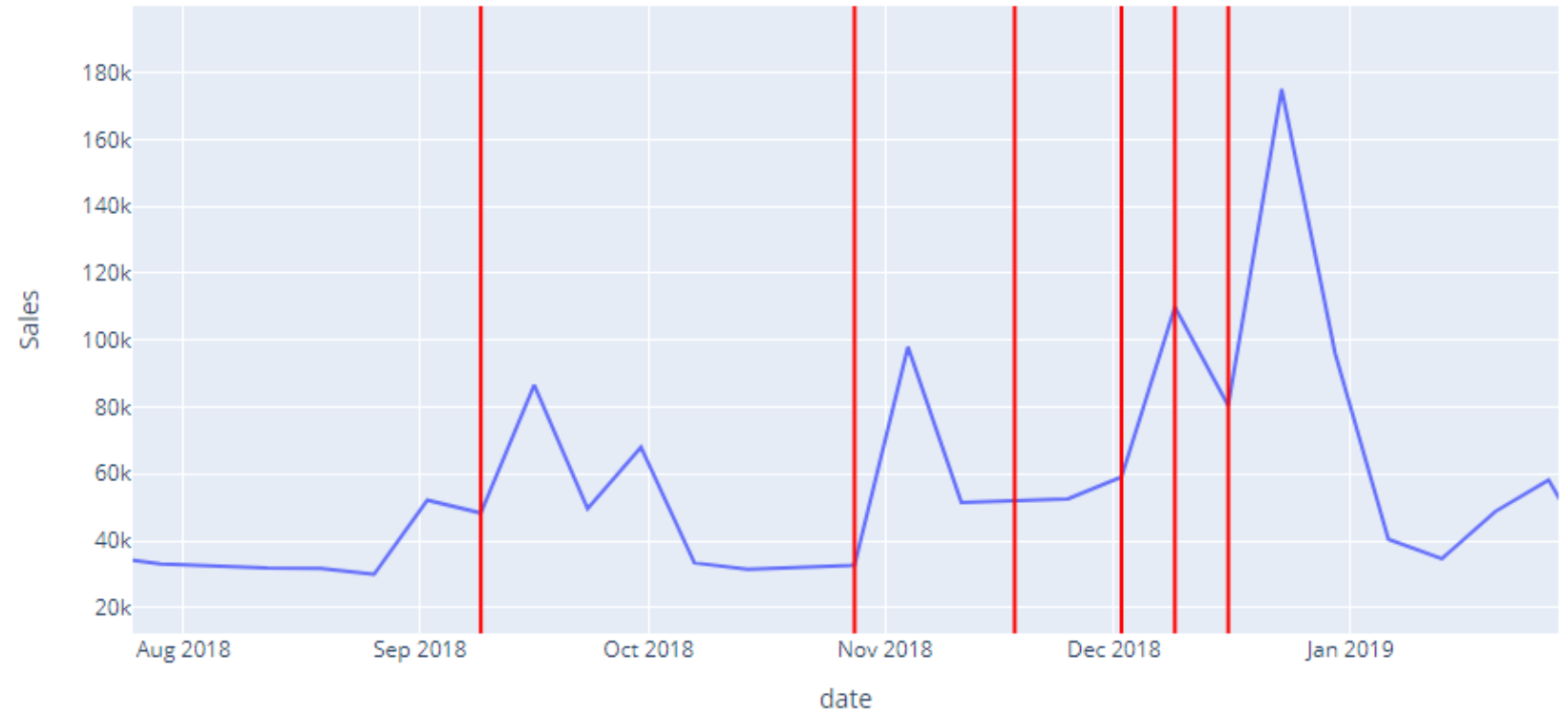
Sales Trend with Promotion Periods



Result 5

EDA

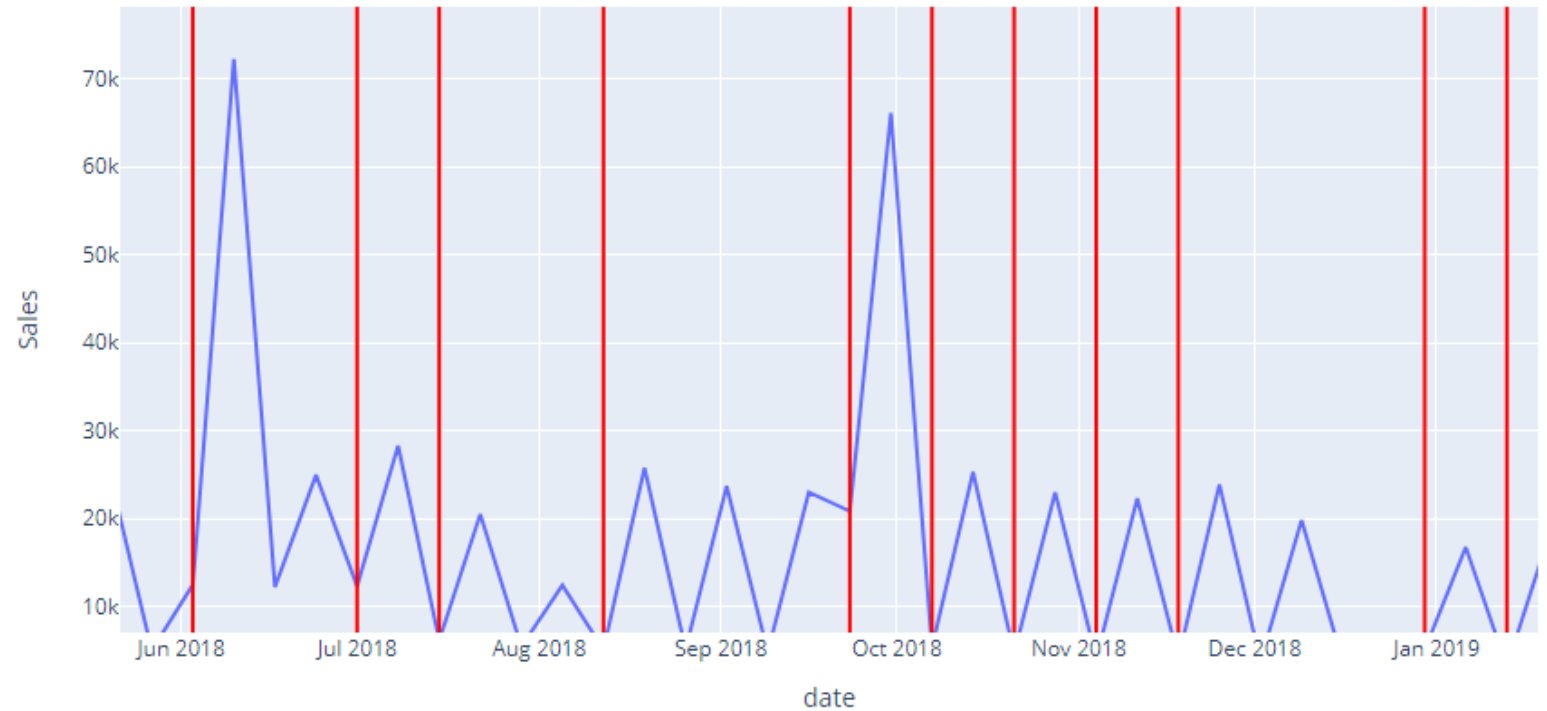
For all products, offering Catalogue promotions would increase sales the day after the promotion.



Result 6

EDA

For SKU4, there is a huge increase on sales the day after a Store-End promotion was chosen.



Result 7

EDA

Each product has a 95% probability of generating these amount of sales daily (to prepare budget before hand):

SKU1: [43574.16, 52112.24]

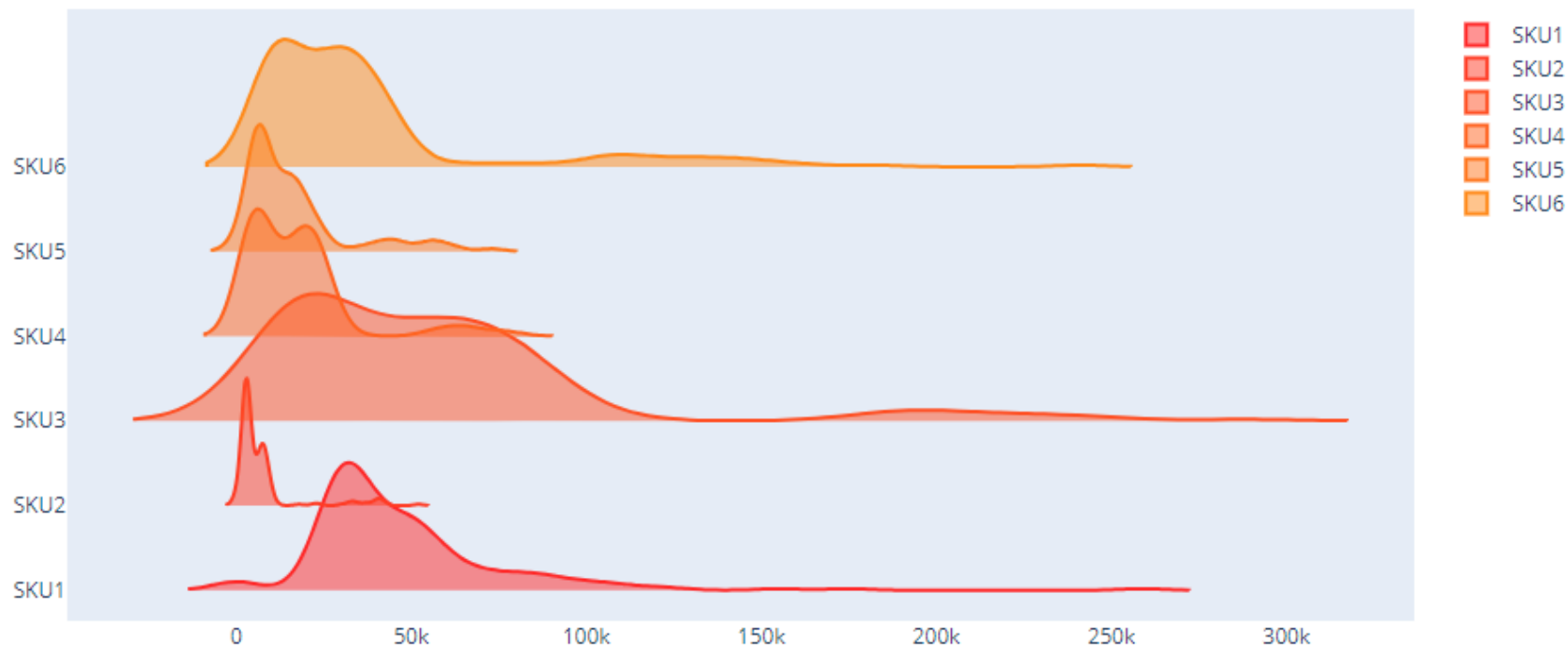
SKU2: [6107.43, 8654.14]

SKU3: [49403.63, 63337.02]

SKU4: [14931.67, 19247.07]

SKU5: [14460.09, 18752.23]

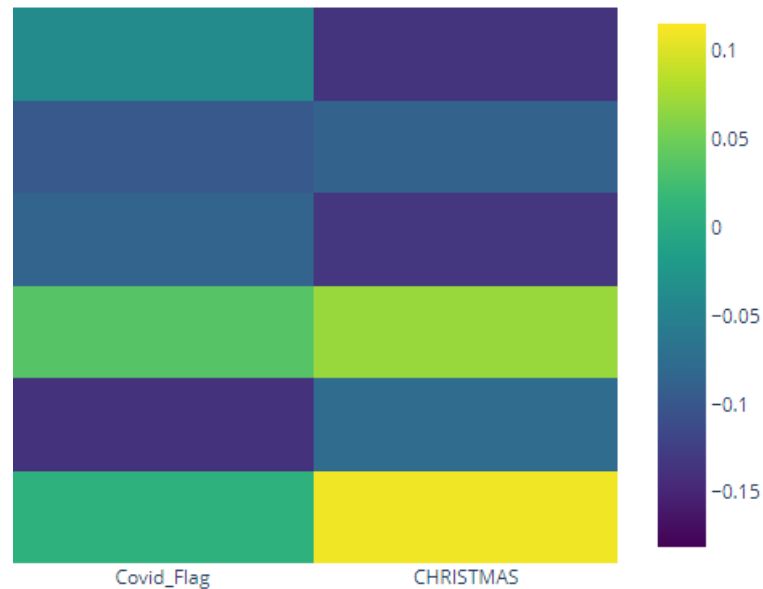
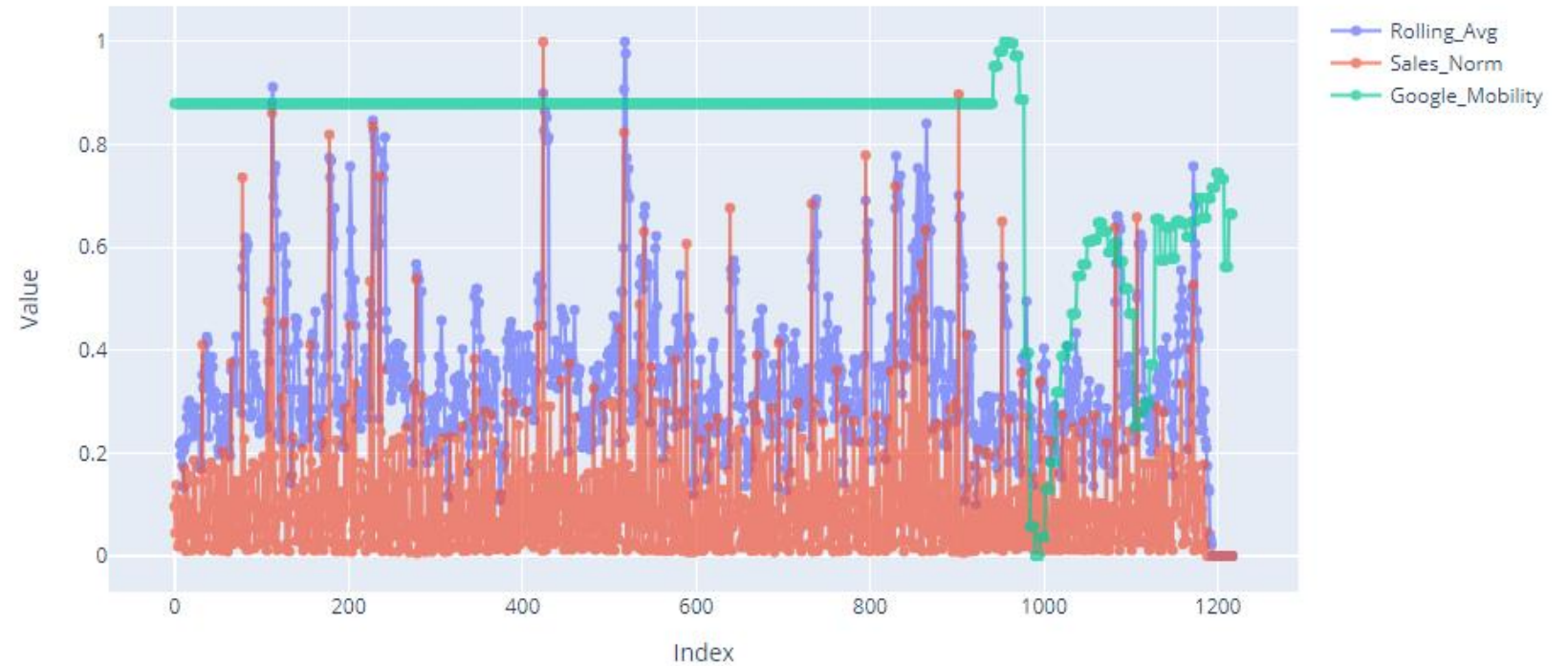
SKU6: [32916.42, 43159.34]



Result 8

EDA

Google Mobility has no effect over the sales of any product, so there should be no need to close off sales given this situation.



STORE END PROMO: 4.52%
Google_Mobility: 1.75%

Google_Mobility: 0.71%
Covid_Flag: 0.12%

Result 9

EDA

The following discounts ($[0,1] \Leftrightarrow [0\%, 100\%]$) are optimal for product sales increase and company's profit balance:

SKU1: [0.11, 0.15]

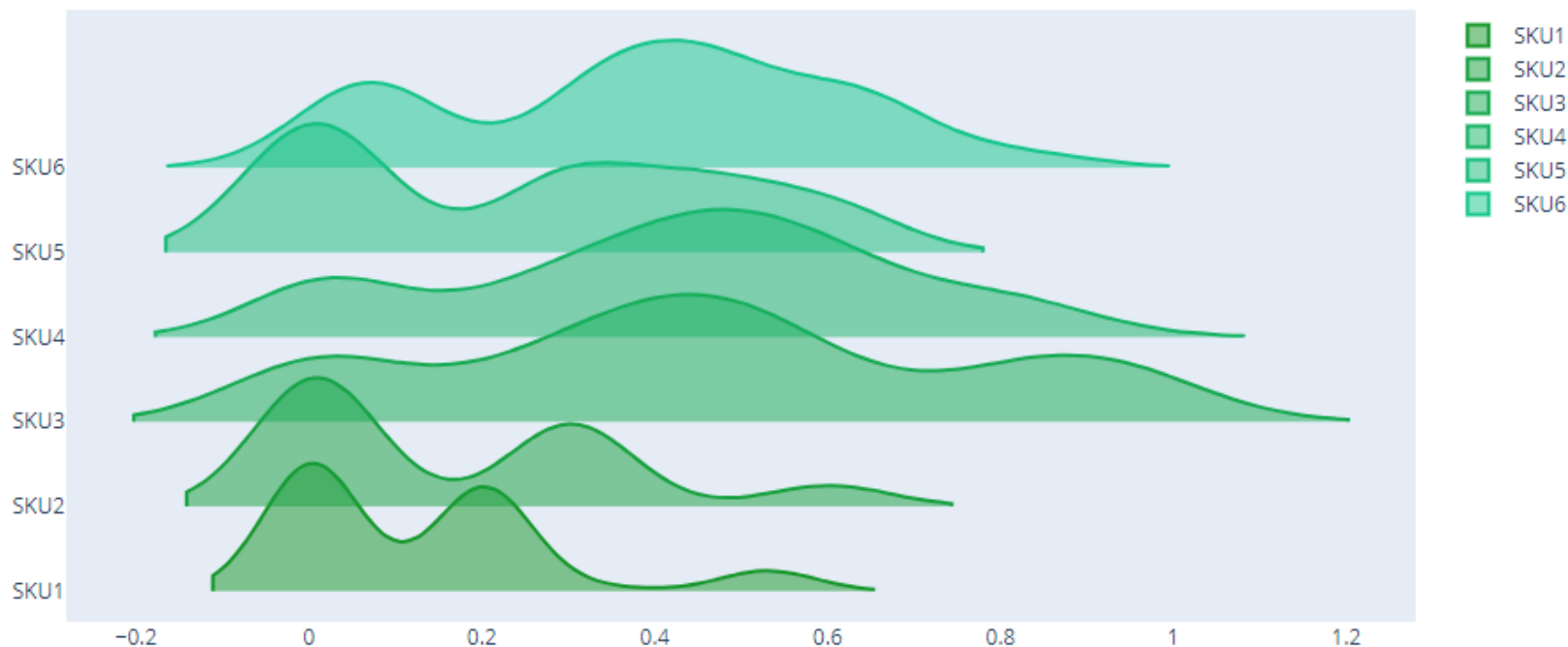
SKU2: [0.14, 0.19]

SKU3: [0.42, 0.50]

SKU4: [0.39, 0.46]

SKU5: [0.22, 0.28]

SKU6: [0.36, 0.42]



EDA Summary

This forecast project has yielded valuable insights and recommendations. The analysis has shown that offering price discounts, especially during sales drops, can significantly impact sales. Prioritizing investments in top-selling products (SKU1, 3, 6) is crucial. SKU1 performs exceptionally well during Christmas and Easter, with over 10% increased sales. In-store promotions have an immediate effect on SKU1-5 sales, while Catalogue promotions consistently increase sales the next day for all products. Store-End promotions notably enhance SKU4 sales the day after implementation. Each SKU's daily sales range (95%) aids budget planning. Google Mobility has no sales impact. Optimal discount ranges to maximize sales and profit balance vary for each SKU (1-6). These findings offer actionable insights for effective sales strategies and resource allocation.

Holiday Recommendations:

Leverage SKU1's exceptional performance during Christmas and Easter by strategically aligning marketing efforts and product availability to capture the increased demand during these holiday periods.

Promotional Recommendations:

Utilize end-store promotions to stimulate immediate sales for SKU4. These promotions have shown to yield instant results and can be employed strategically to enhance sales performance.

Utilize in-store promotions to stimulate immediate sales for SKU1-5. These promotions have shown to yield instant results and can be employed strategically to enhance sales performance.

Utilize catalogue promotions to generate increased sales for all products. These promotions consistently lead to higher sales figures on the day following the promotion.

General Recommendations:

Focus investment efforts on SKU1, 3, and 6 due to their high sales volume. Allocating resources to these products can yield substantial returns and sustain the company's growth.

Utilize the provided daily sales probability ranges (95%) for each SKU (1-6) to facilitate budget planning and resource allocation, ensuring readiness for various sales scenarios.

Given the lack of sales impact from Google Mobility, there is no need to make drastic changes to sales strategies based on this factor. The focus can remain on other proven strategies.

Implement the recommended optimal discount ranges for each SKU (1-6) to strike a balance between boosting sales and maintaining a healthy profit margin.

Model proposals

ARIMA (AutoRegressive Integrated Moving Average):

Can be parallelized in PySpark for improved performance.

Well-suited for univariate or multivariate time series forecasting tasks.

Prophet model:

It can be integrated into PySpark and leverages additive modeling to capture different seasonalities.

LSTM (Long Short-Term Memory) Neural Networks:

PySpark can be used to distribute the training of LSTM models across clusters, enabling efficient processing of large datasets.

Particularly effective for time series with long-range dependencies and intricate relationships.

XGBoost Time Series:

Utilizes gradient boosting techniques for predictive modeling.

PySpark can parallelize XGBoost training across nodes, improving performance for large datasets.

N-BEATS (Optional, haven't tried it yet):

Interpretable and flexible forecasting models.

Forecast Project – EDA Results

Thank You