Report

Retail Forecasting

September 1, 2023

Team Members

Kierra Dangerfield

kierradachelle@yahoo.com

United States of America

Freelance

Specialization: Data Science

Agenda

Background

EDA

Model Selection

Recommendations

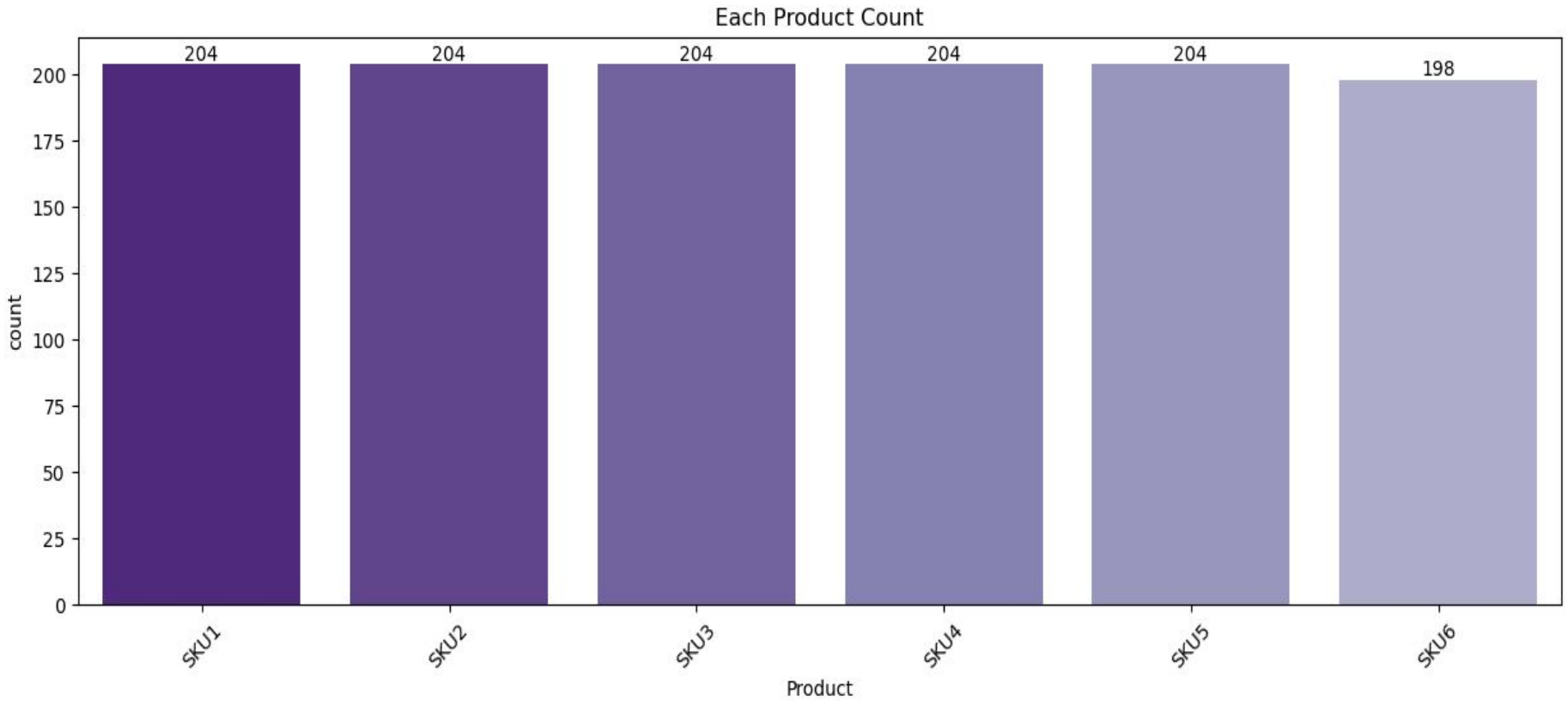
Background

Background

- The large company which 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 a forecast of each of the products at item level every week in weekly buckets.

EDA

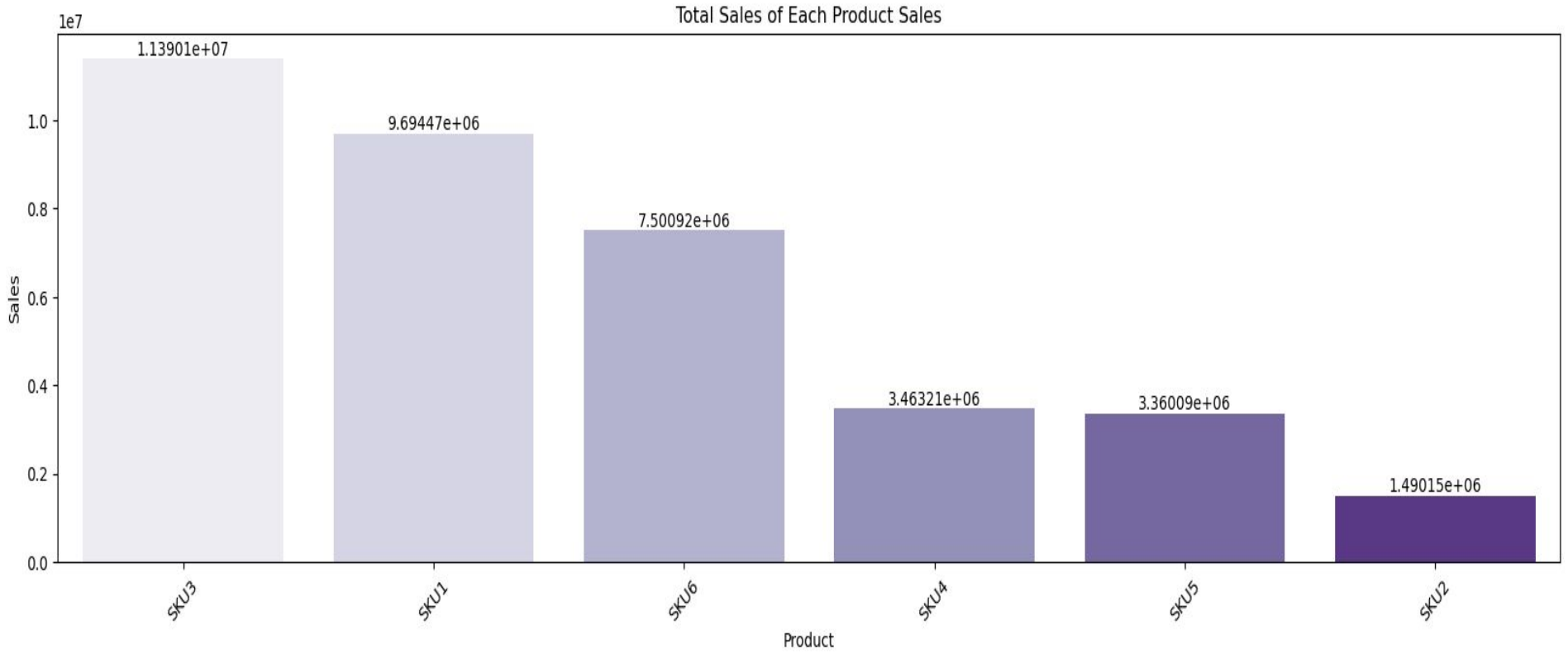
Products Sold



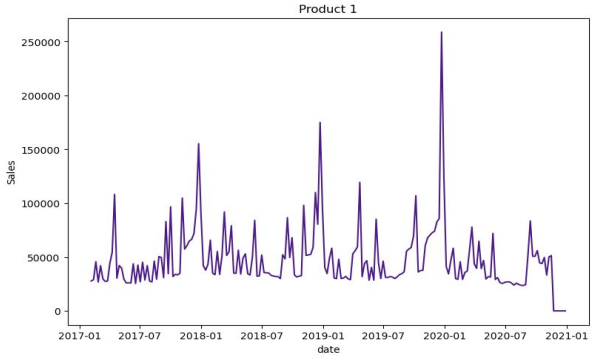
- These are the products sold count.
- Product 6 has slightly less amount of products sold, but it is still around the same amount of beverages sold.

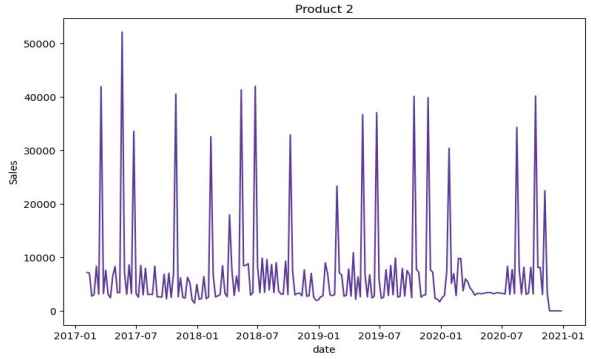
Total Sales of Each Product

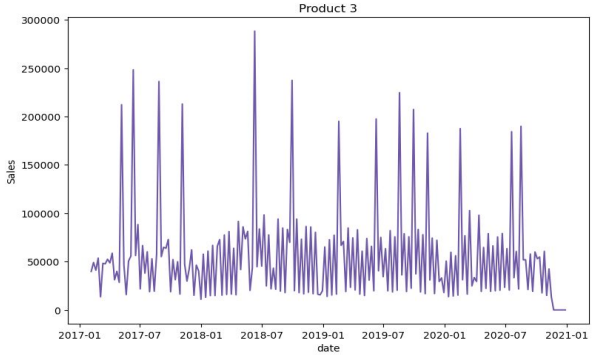
- Product 3 has the most amount in total sales, and product 2 has the least amount in total sales.
- Product 3 could have a higher price compared to the others.

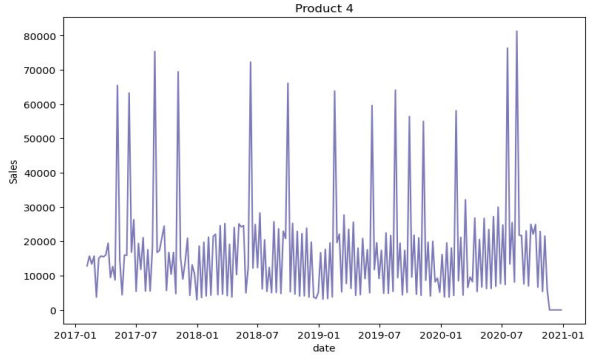


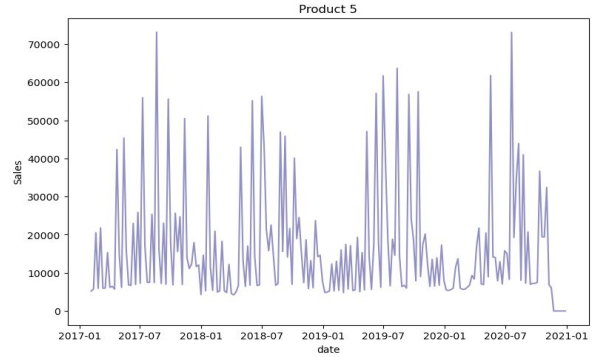
Product Sales Over Time

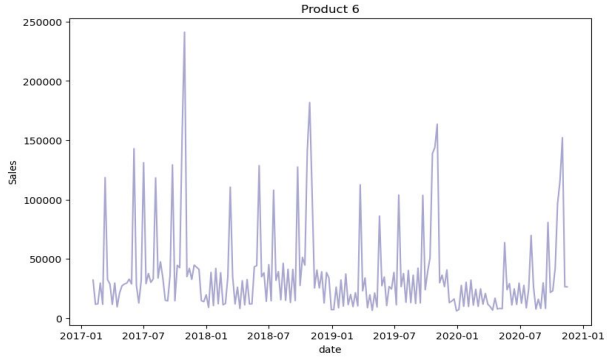










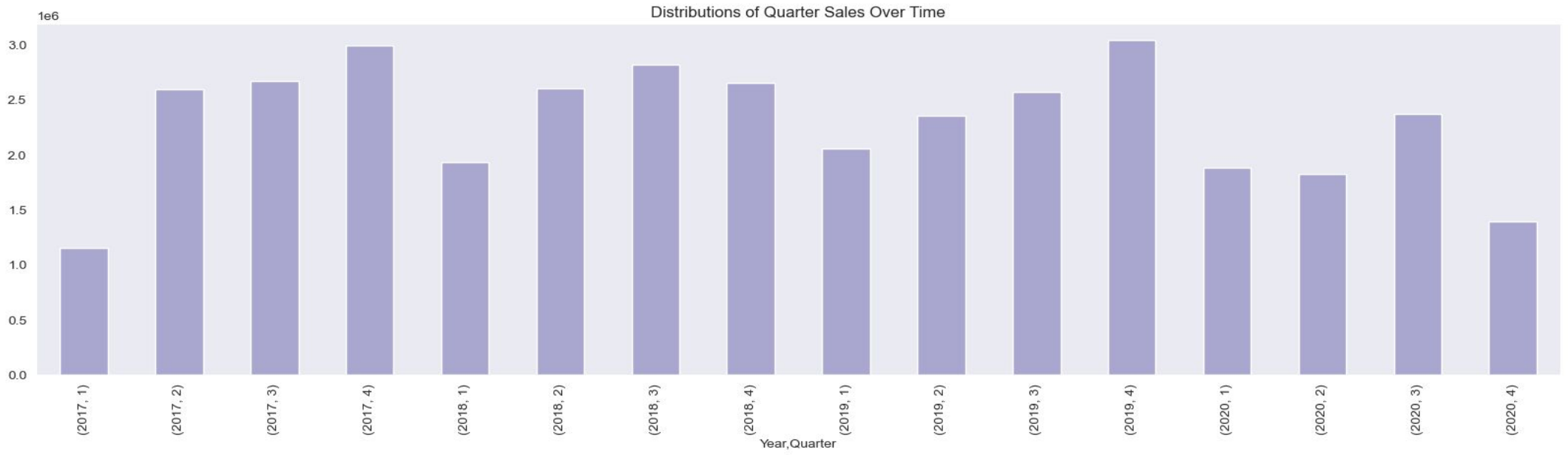


- The line plots show the sales over time for each product.
- For each product we can see when the product has high points and low points of sales.

Quarter Sales Over Time

The first quarter had the lowest sales of the year except for quarter 1 in 2020.

In 2019 there was a steady, linear increase.

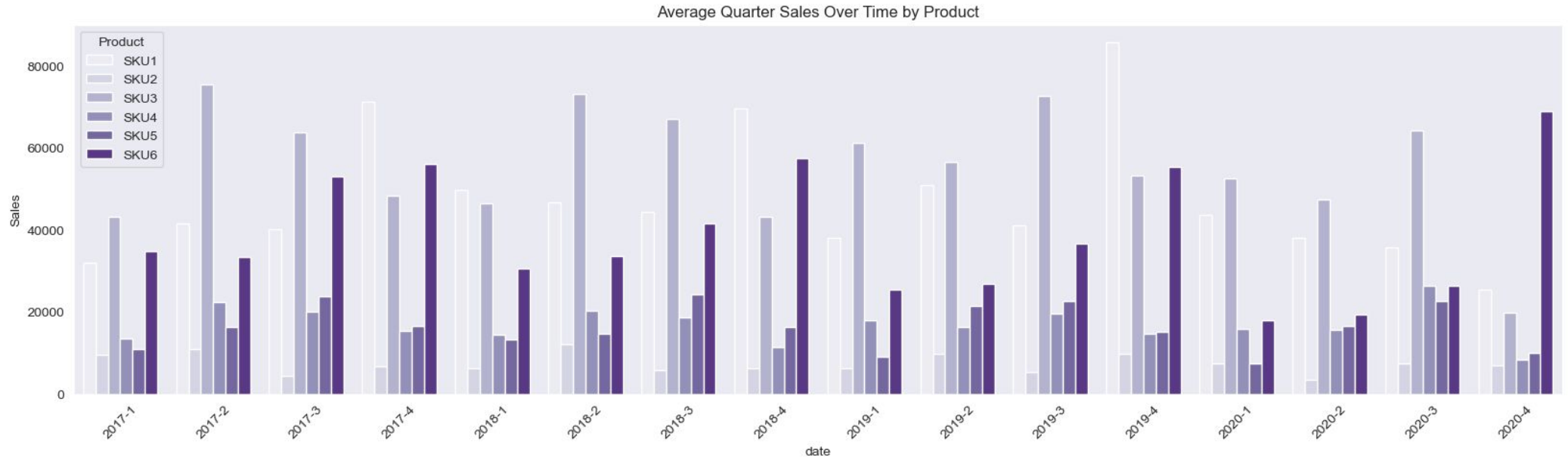


Average Quarter Sales

The chart shows the average quarter sales over time by product.

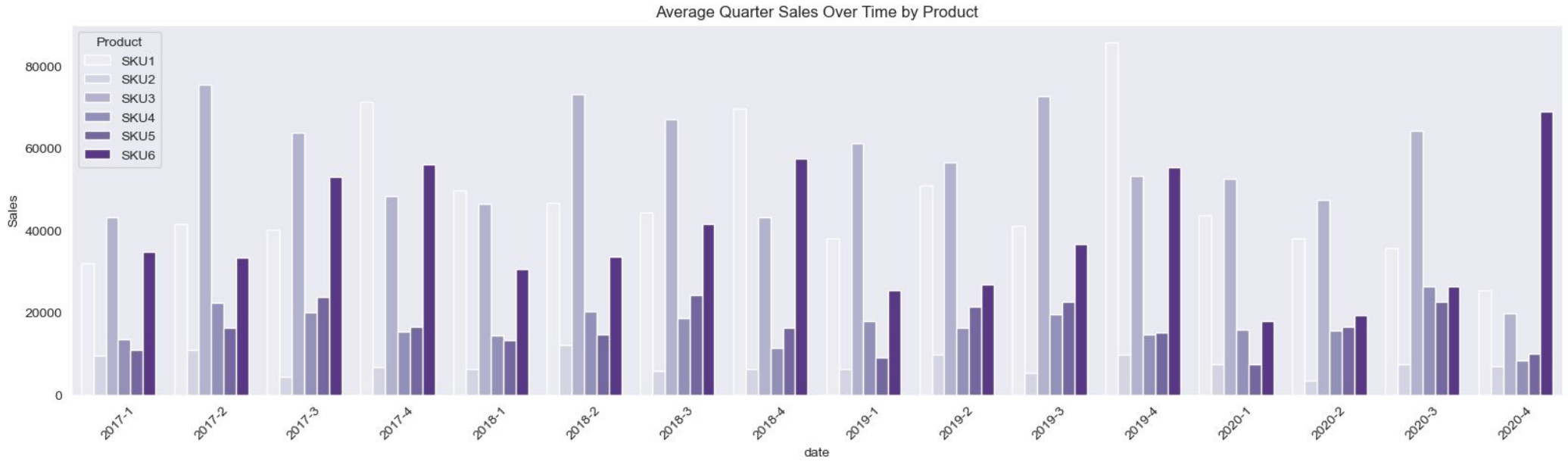
There is a significant difference between Product 1 in 2019-4 compared to 2020-4.

The average quarter sales for product 1 has consistently gone down since 2019-4.



Average Quarter Sales Cont.

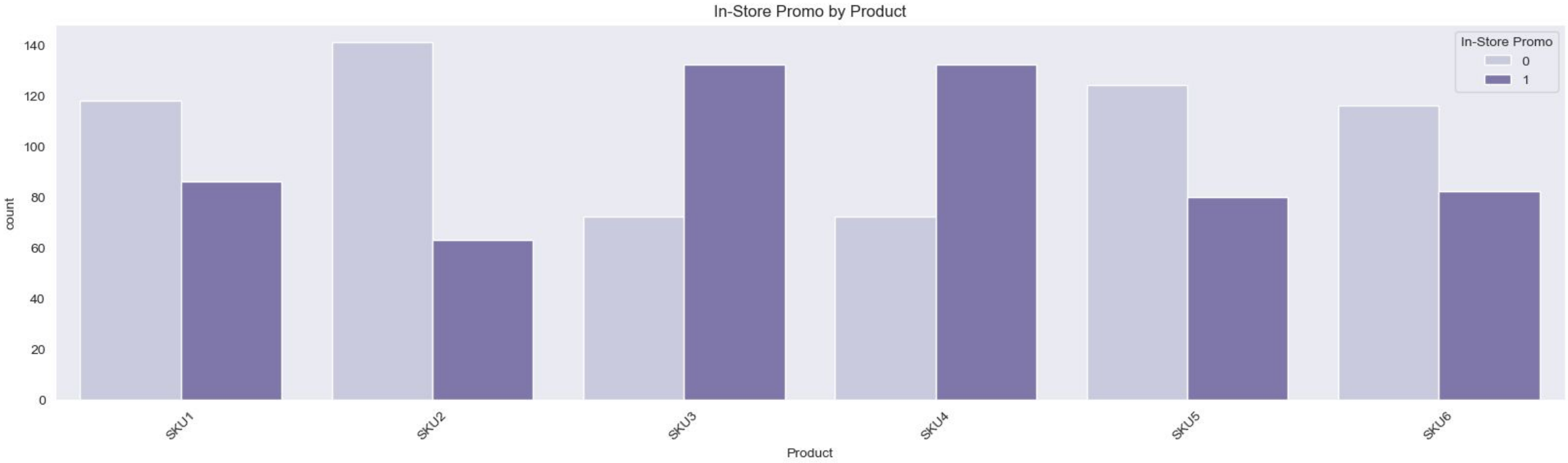
- Product 3 is one of the best performing products.
- Product 2 has about the same average except for their low point in 2020, quarter 2.
- Product 4 and 5 were about the same except for quarter 1 in 2020 for Product 5.
- Product 6 had the best quarter in 2020, quarter 4.



In-Store Promo

Majority of product 3 and 4 sales had in-store promo.

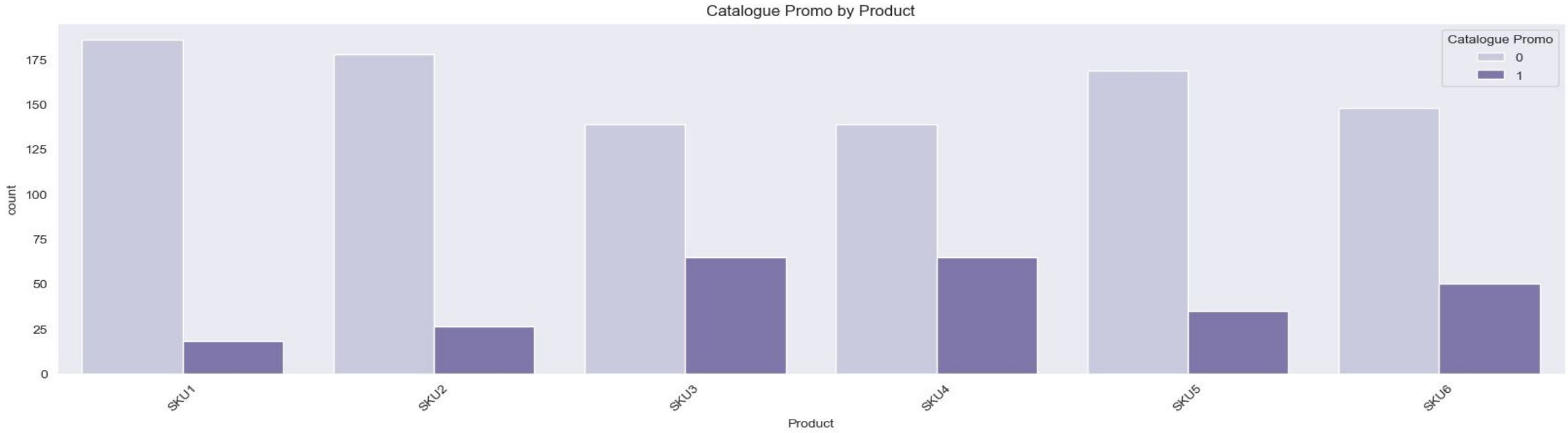
It seems customers respond well to in-store promo.



Catalog Promo

Majority of the sales did not have catalog promo. However, of the sales that had catalog promo, product 3 and 4 sales were the highest amount.

It seems customers respond pretty well to catalog promo.

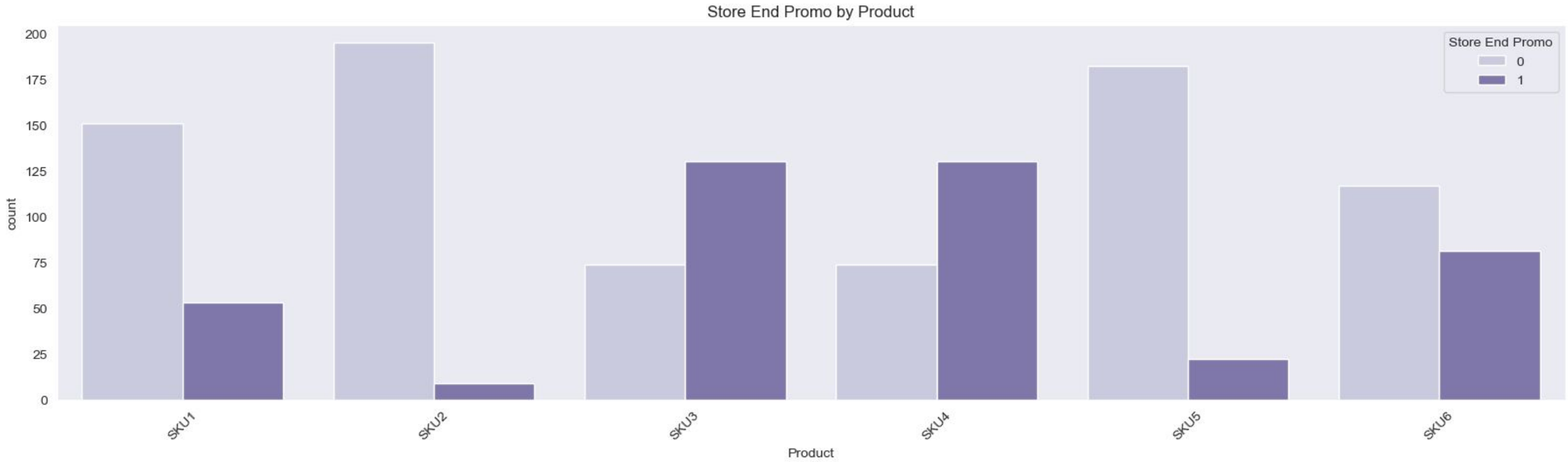


Store End Promo

Majority of the sales did not have store end promo except for product 3 and 4.

It seems customers respond pretty well to store end promo.

Even with store end promo customers still did not respond to buy product 2.



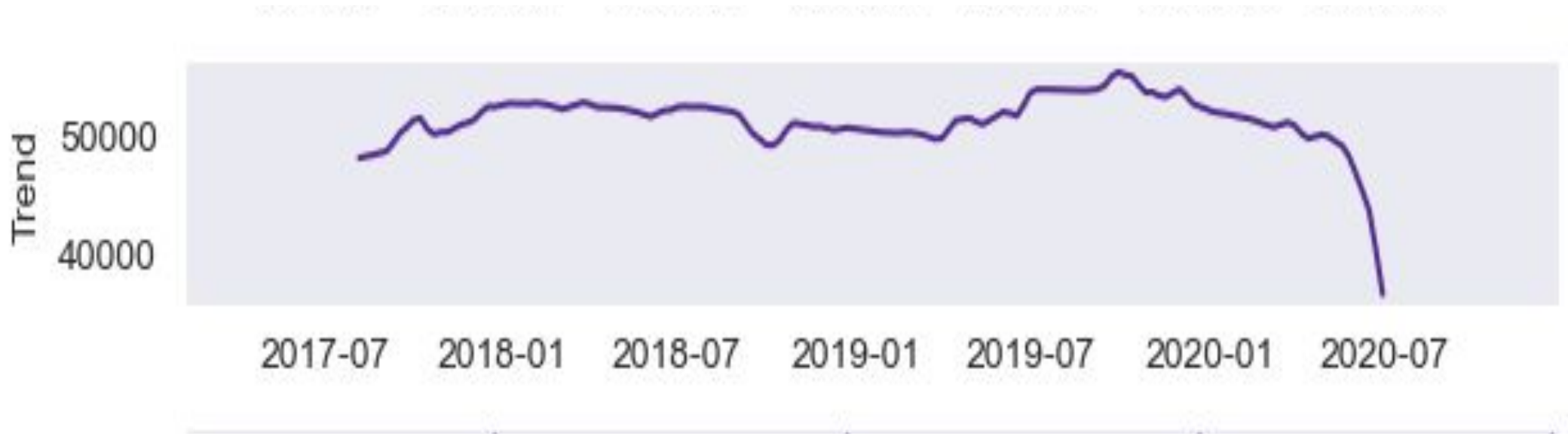
Time Series Decompose

- Time series decomposition is a technique that splits a time series into several components, each representing an underlying pattern category, trend, seasonality, and noise.
- I focused on the trend for each product. I will show the highlights
- Trend: describes the increasing or decreasing trend in data.

Product 1

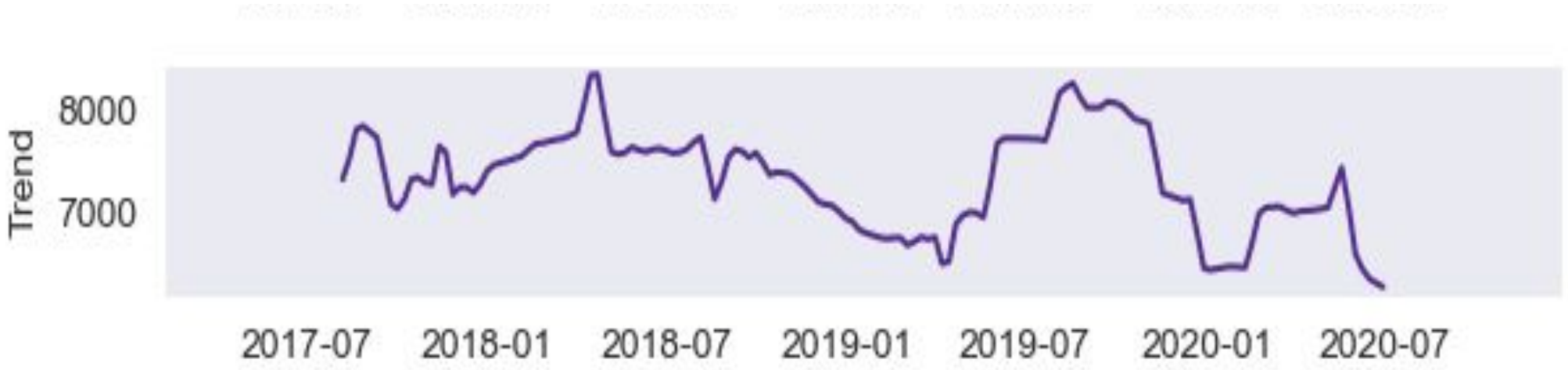
Around the end of 2019 and January 2020 was the highest amount of average sales.

After 2020, it went down.



Product 2

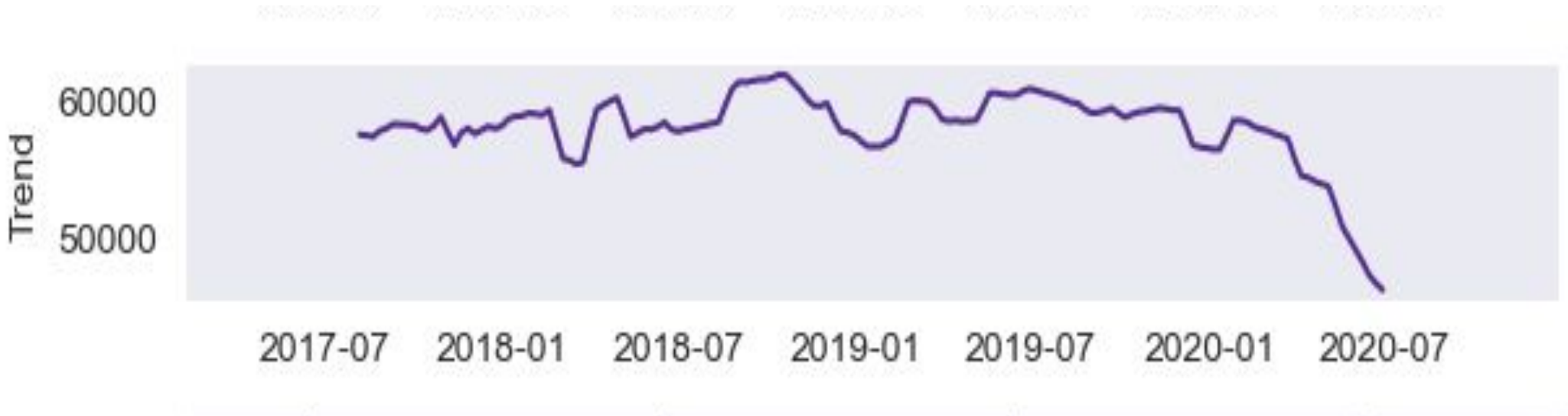
There is a consistent dip between 2018-07 and January 2019, rose again and then fell again around January 2020.



Product 3

Product 3's trend has started to go downward around January 2020.

Before it had minimal dips until around 2020.



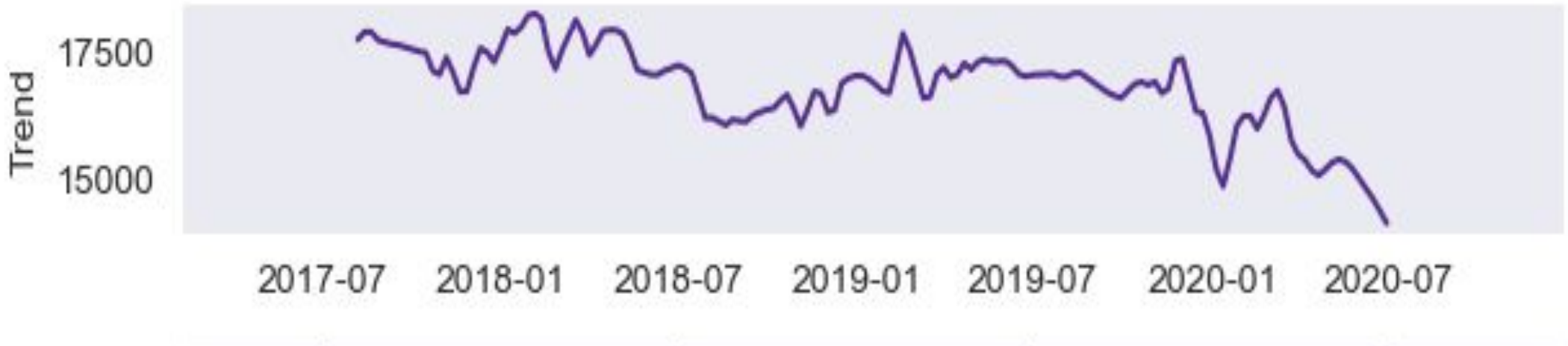
Product 4

Around January 2018 it started to turn downward and spiked in 2020 and started to take a sharp decrease.

Product 5

Based on its trend, it has gone down.

It might need to be discontinued in the future

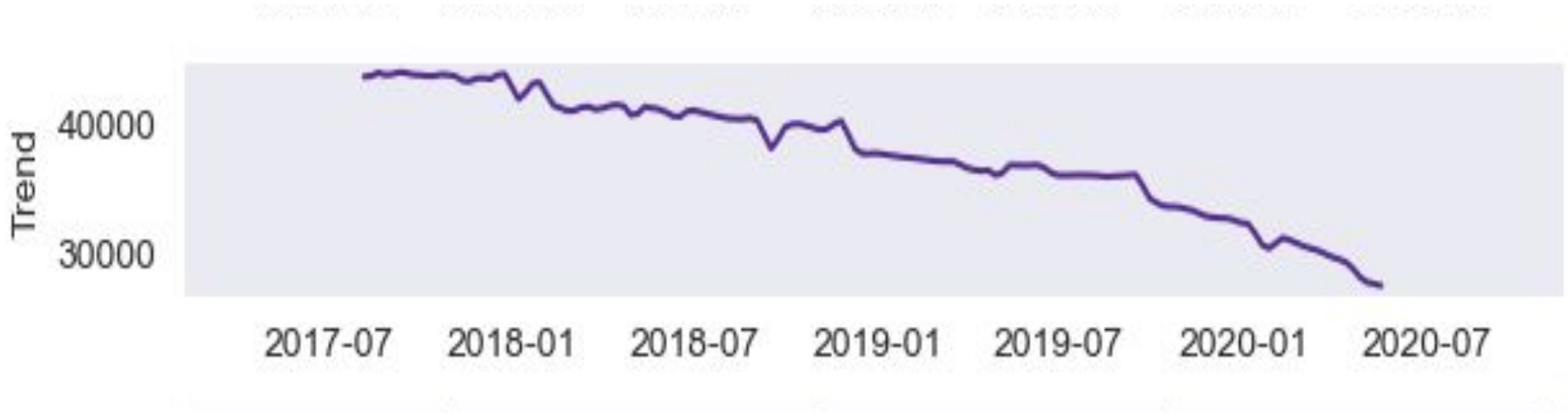


Product 6

This is product 6 decompose

Based on its trend, it has consistently gone down.

It might be a product that gets discontinued in the future.



Model Selection

Models

These are the 5 models I decided to test between.

These results are for product 1.

The Random Forest tested the best based on it mean absolute error.

Model	Mean Absolute Error
Linear	$12,602.134
Random Forest	**$7,520.47**
ARIMA	$26,532.32
SARIMA	$26,299.80
SimpleRNN	$13,856.30

Model Results

These are the results using Random Forest Regressor for each product.

Product 2 performed the best.

Product 3 performed the worst.

Product	Mean Absolute Error
SKU1	$7,256.62
SKU2	**$1,067.46**
SKU3	**$13,994.45**
SKU4	$4,536.77
SKU5	$1,588.22
SKU6	$7,980.13

Model Results Cont.

Product 2 had the least amount of total sales which could be a contributing factor.

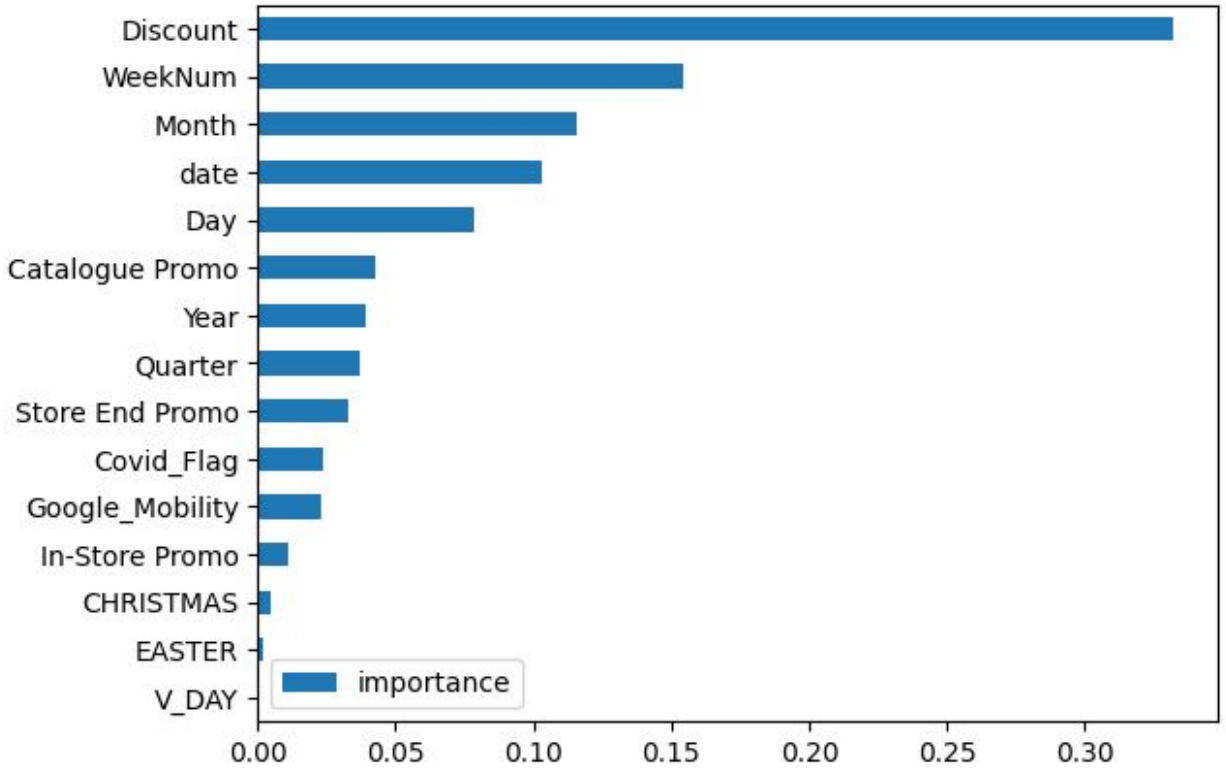
Product 3 also had the highest amount of total sales which could be a contributing factor.

Product	Mean Absolute Error
SKU1	$7,256.62
SKU2	**$1,067.46**
SKU3	**$13,994.45**
SKU4	$4,536.77
SKU5	$1,588.22
SKU6	$7,980.13

Feature Importance

The chart shows the feature importance for Product 1.

Discount is the most important feature.



Recommendations

Final Recommendations

The ARIMA model would be great to use to predict upcoming weeks; however, it is not the best performing model.

I would suggest getting more data for each product to help build a better model. Instead of getting sales weekly, maybe collect daily sales in order to get more data.

If I would discontinue any product it would be product 6 since it has a decreasing trend.

Thank You