



Contents

[**1-**](#_heading=h.gjdgxs) **Introduction** 2

[**2-**](#_heading=h.30j0zll) **Collect historical data** 3

[**3-**](#_heading=h.1fob9te) **Analyze the data** 3

[**4-**](#_heading=h.tyjcwt) **Choose a forecasting model:** 5

[1. Linear Regression 5](#_heading=h.3dy6vkm)

[2. SARIMA 5](#_heading=h.1t3h5sf)

[3. Simple Naive 5](#_heading=h.4d34og8)

[4. RNN 5](#_heading=h.2s8eyo1)

[5. Holt Winter's Seasonal 5](#_heading=h.17dp8vu)

[**5-**](#_heading=h.3rdcrjn) **Make the forecast** 5

[**6-**](#_heading=h.lnxbz9) **Monitor and adjust the forecast** 6

[**Conclusion:** 6](#_heading=h.35nkun2)

[**Figure 1 Sales distribution 4**](#_heading=h.3znysh7)

[**Figure 2 Sales and Product relationship 4**](#_heading=h.2et92p0)

[**Figure 3 SARIMA's forecasting result for one of the products 6**](#_heading=h.26in1rg)

## **Introduction**

A large beverage company in Australia sells its products for a whole year, and the selling of those products is influenced by some factors like seasons or holidays. They need a weekly forecast for each of their products, which can predict how much of a product will sell in the current week.

Retail forecasting is the process of predicting future sales and consumer behavior in the retail industry. Accurate forecasting is crucial for retail businesses to make informed decisions regarding inventory management, pricing strategies, and overall business strategy.

One popular method of retail forecasting is time series analysis, which uses historical sales data to predict future sales. This method involves analyzing trends, seasonality, and other patterns in the data to make predictions. Other forecasting methods include machine learning algorithms, which can incorporate additional data sources such as weather, economic indicators, and social media trends to make more accurate predictions.

One of the main challenges of retail forecasting is the unpredictability of consumer behavior. External factors such as changes in the economy, weather, and social trends can greatly impact consumer behavior and make it difficult to accurately predict sales. Additionally, the rise of e-commerce and online retail has created new challenges for forecasting, as traditional methods may not be as effective in predicting online sales.

Despite these challenges, retail forecasting is a crucial aspect of retail business strategy. Accurate forecasting can help businesses optimize inventory, adjust pricing strategies, and make informed decisions regarding expansion and other business decisions. As technology continues to evolve, the methods and tools used for retail forecasting are likely to become even more sophisticated and effective.

The Australian beverages company has 6 products SKU1, SKU2, SKU3, SKU4, SKU5, SKU6. To forecast demand for each product at an item level every week, these are the steps that can be taken.

1. Collect historical data
2. Analyze the data
3. Choose a forecasting model
4. Make the forecast
5. Monitor and adjust the forecast

## **Collect historical data**

The first step is to gather data on sales for each product at the item level for at least the past few years. This data should include information on the time period (week), the number of units sold, and any other relevant factors that may have influenced demand (such as holidays, seasonality, promotions, etc.).

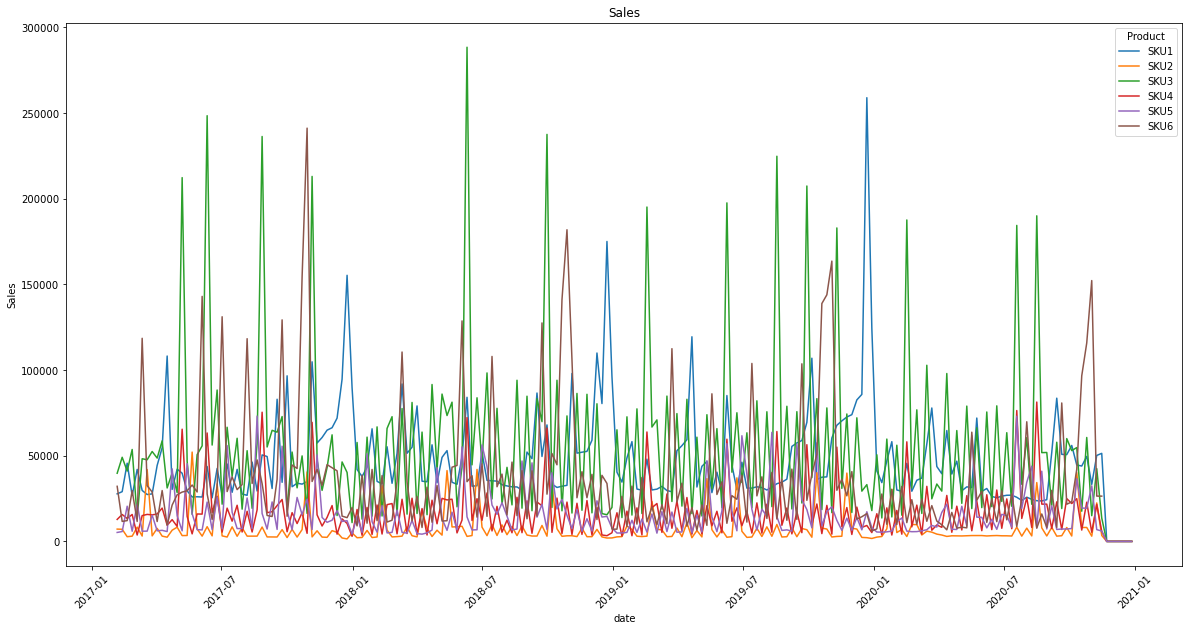
## **Analyze the data**

## 

The given data was explored, manipulated, analyzed and cleansed using multiple EDA techniques, and the results depicts that the given dataset is a time series data from February 2017 till November 2020. Having information about : Product, date, Sales, Price Discount (%), In-Store Promo, Catalogue Promo, Store End Promo, Google\_Mobility, Covid\_Flag, V\_DAY, EASTER, CHRISTMAS. The dataset has 1219 rows and 12 columns. We have three numerical attributes: sales, price discount, and Google mobility, among which sales is the outcome we want to predict. The rest nine attributes are categorical. There is not much correlation between features, only 'Covid\_Flag' & 'Google\_Mobility' have relatively higher correlation. There are no NA values in our dataset.

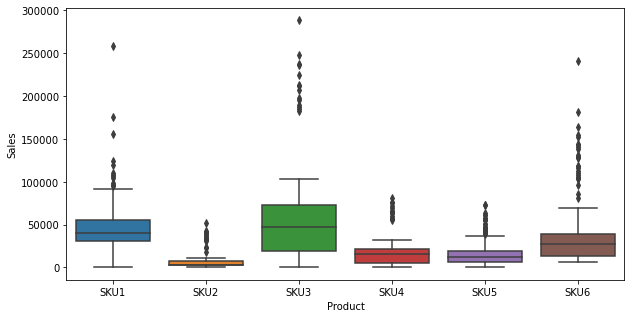
For all numerical fields, there is a skewness problem. Besides the Price Discount field, Sales and Google\_Mobility fields have the problem of outliers. Also, there is a percentage of zeros in the Sales field, which may need further manipulation. For the other two fields, a relatively large proportion of zeros will not affect further works.

Here is the overall distribution of the sales of each product:



*Figure 1 Slaes distribution*

The following box plot represent the relationship between sales and products, we can also observe the outliers for each product:



*Figure 2 Sales and Product relationship*

At product level, it is evident that SKU1, SKU3, and SKU6 are the top three products that generate major sales for the company. So we will recommend the company should put their main sources to produce those three products.

## **Choose a forecasting model**

Based on the exploratory data analysis of the historical data, the company can choose a forecasting model that best fits their needs. For that reason, following models were built:

1. Linear Regression: Linear regression is a statistical method for modeling the relationship between a dependent variable and one or more independent variables. In retail forecasting, linear regression can be used to identify the key drivers of sales, such as price, promotions, seasonality, and marketing campaigns. By analyzing historical data on these factors, a linear regression model can estimate how changes in these variables will affect future sales of the products.
2. SARIMA: Seasonal Autoregressive Integrated Moving Average (SARIMA) is a time series forecasting method that takes into account both seasonality and trends in the data. SARIMA models can be used to identify patterns and trends in historical sales data for the products, and make predictions about future sales based on these patterns. SARIMA models are often used in retail forecasting to account for the seasonal variations in demand for products.
3. Simple Naive: Simple naive forecasting is a basic method that involves assuming that future sales will be the same as past sales. This method can be useful for forecasting short-term sales trends, but may not be accurate for longer-term forecasts or when there are significant changes in demand for the products.
4. RNN: Recurrent Neural Networks (RNNs) are a type of machine learning model that can be used for time series forecasting. RNNs are particularly useful for forecasting complex patterns and trends in sales data for the products. RNNs can be trained on historical data to learn the relationships between different variables and make predictions about future sales based on this learning.
5. LSTM: Long Short-Term Memory networks are used in Deep Learning. They are able to learn long term dependencies particularly in sequence prediction problems.
6. GRU: Gated Recurrent Units (GRU) is a gating mechanism in RNNs.
7. Holt-Winter's Seasonal: Holt-Winter's Seasonal method is another time series forecasting method that can be used to account for seasonal variations in sales data. This method involves estimating three parameters: the level of sales, the trend in sales, and the seasonal component of sales. These parameters can then be used to make predictions about future sales of the products.

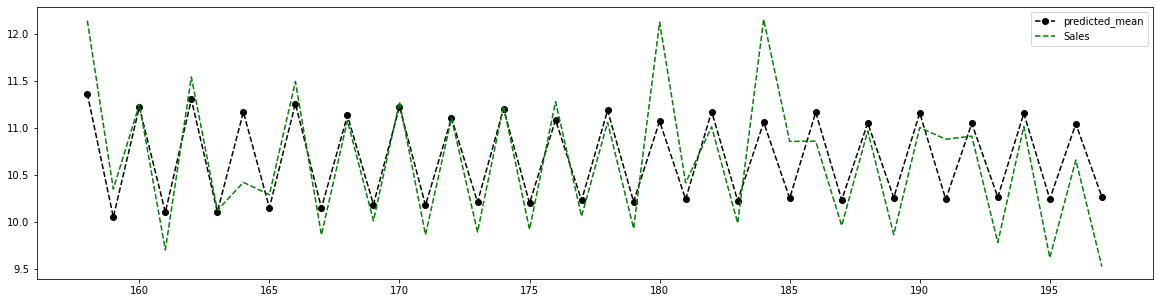
## **Make the forecast**

Once the forecasting models are chosen, the company can use it to generate forecasts for each product at the item level for the upcoming weeks. The performance of these forecasting models were scrutinized by MAPE (Mean Absolute Percentage Error). MAPE is a percentage error metric where the value corresponds to the average amount of error that predictions have. Therefore, a lower MAPE is better, where **the lower the value the more accurate the model is**. The forecasting results (1-MAPE) are presented in the form of a table below:

| **Products** | **Linear Regression** | **Simple Naive** | **SARIMA** | **RNN** | **LSTM** | **GRU** | **Holt-Winter’s Seasonal** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **SKU1** | 0.79 | 0.97 | 0.97 | 0.73 | 0.87 | 0.85 | 0.97 |
| **SKU2** | 0.21 | 0.92 | 0.94 | 0.97 |
| **SKU3** | 0.35 | 0.91 | 0.97 | 0.95 |
| **SKU4** | 0.34 | 0.88 | 0.97 | 0.94 |
| **SKU5** | 0.65 | 0.94 | 0.94 | 0.93 |
| **SKU6** | 0.17 | 0.92 | 0.93 | 0.96 |

*Table 1: Forecasting Model's Performance*

Linear regression, Simple Naïve, SARIMA and Holt-Winter’s Seasonal have been modeled individually for each product, whereas RNN, LSTM and GRU made the forecasting collectively. Below given graph is forecasting result from SARIMA, for one of the products.



*Figure 3 SARIMA's forecasting result for one of the products*

## **Monitor and adjust the forecast**

It is important to monitor the accuracy of the forecasts and adjust them as necessary. This can be done by comparing the actual sales data to the forecasted values and making adjustments to the forecasting model as needed.

## **Conclusion**

In conclusion, each of these forecasting methods (Linear Regression, SARIMA, Simple Naive, RNN, and Holt Winter's Seasonal) can be useful in different ways for forecasting retail sales of products at product level. The choice of method depends on the specific characteristics of the data, the length of the forecasting horizon, and the goals of the forecast. For our case and according to the MAPE accuracy results, SARIMA and Holt-Winter’s Seasonal appear to be the best fit on the test data for log transformed sales data. SARIMA has performed best among all and LSTM is the best performing model among deep learning models. Surely, we should further investigate the model’s usability and overfitting problem in the future with updated information.