Australian beverage retail forecasting

Business Understanding

Overview

Company XYZ is an Australian based beverage company.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.

At the time, they were using their own software, written in-house, but it often produced forecasts that did not seem sensible. Company wanted to explore power of AI/ML based forecasting to replace their in house local solution.The project will be based on using time series forecasting using Machine Learning algorithms.We will forecast the retail for future periods.

Retail forecasting helps in maintaining a good stock inventory management .Where by we are able to know the demand by forecasting and therefore we can maintain a optimum inventory level for avoid stockouts therefore the customers loyalty is maintain by finding their required products always ready.Retail forecasting also aids in avoid wastage by having excess amounts of stock the resources could used in another productive way to increase revenue in an efficient way.Reducing stockouts and reducing wastage increases the amount of profit for the business.

The Project success criteria will be a forecast accuracy of 90% meaning that the forecast error is than 10% of the actual sales.

Problem Statement

 Inaccurate forecasting can lead to overstocking or understocking of products, which can result in lost sales, increased costs, and reduced customer satisfaction. The company needs an accurate and reliable forecasting model that takes into account the various factors that influence demand and can provide timely and actionable insights to support decision-making.

Objectives

1. The main objective is to forecast accurate sales

2. To determine the product with the most sales

3. To determine which holiday has the most sales

4. To determine which month has the most sales

Data Understanding

The data has 12 columns and 1218 rows

The data was collected from 2017 to 2020 about the sales for the Australian Company

The description for the data is found below:

 \* Product: This column  refers to the unique identifier or code for a particular product.

\* Date: This column indicates the date of sale for the product.

\* Sales: This column shows the total sales revenue generated by the product on the given date.

\* Price Discount (%): This column indicates the percentage discount that was applied to the product's price at the time of sale.

\* In-Store Promo: This column indicates whether there was a promotion or discount offered specifically for in-store purchases.

\* Catalogue Promo: This column indicates whether there was a promotion or discount offered through a catalogue or other print or digital media.

\* Store End Promo: This column indicates whether there was a promotion or discount offered specifically at the end of a store's promotional period.

\* Google\_Mobility: This column is a measure of the relative mobility or foot traffic in the area surrounding the store where the product was sold, based on Google's location data.

 \* Covid\_Flag: This column  indicates whether there were any COVID-related restrictions or concerns in effect on the date of sale.

\* V\_DAY: This column  indicates whether the date of sale was close to or on Valentine's Day.

\* EASTER: This column  indicates whether the date of sale was close to or on Easter.

\* CHRISTMAS: This column  shows the date of sale was close to or on Christmas.

Data Preparation

This will involve checking for :

\* Completeness - Checking for missing values in the data

\* Uniformity - Here we check for the uniformity of the data and this includes looking at the data types of the different columns and   ensuring they are correct.we check to ensure the labels of the columns are uniform and explicit

\* Consistency - Involves checking for duplicated entries in the data

\* Validity - Involves checking for outliers in the data

Incase any of these are not met we will do the necessary process to make the data complete,valid,uniform and consistent

The data has no missing values

The data has no duplicates

Feature engineering

We will create new columns to help build the accuracy for the model and for analysis.

The new columns will include:

Month

Year

Day of the week

Australian Day

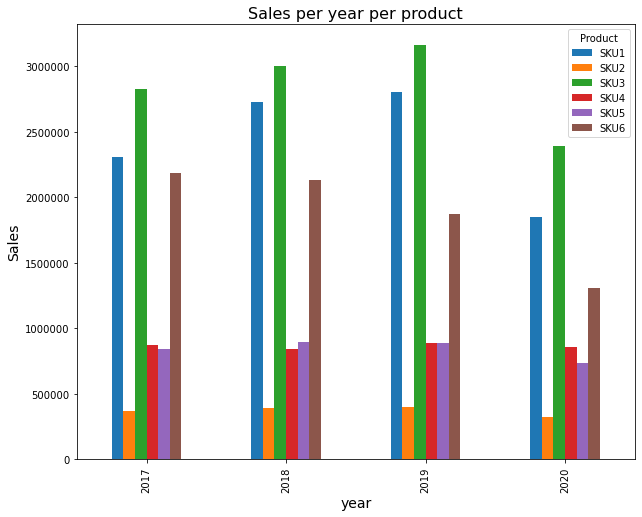
Anzac Day

Seasons

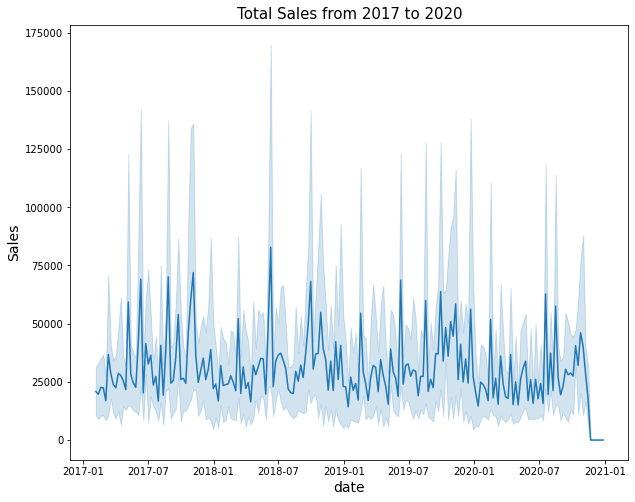
Data Analysis

The analysis bit involves:

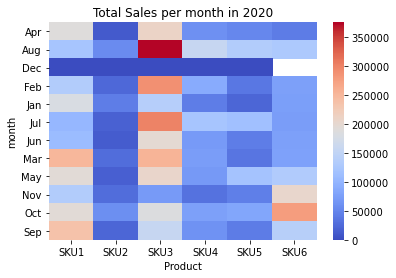
* Univariate analysis
* Bivariate analysis
* Multivariate analysis

[](https://docs.google.com/spreadsheets/d/10QS3G4BrATPi_OEm6nh8N9F_vGRrCuOLf4nfRgLgYqg/copy)

* SKU1 and SKU3 had the highest sales in all the years
* SKU2 has the lowest sales in all the years
* SKU6 had the third highest Sales in all the years

[](https://docs.google.com/spreadsheets/d/10QS3G4BrATPi_OEm6nh8N9F_vGRrCuOLf4nfRgLgYqg/copy)

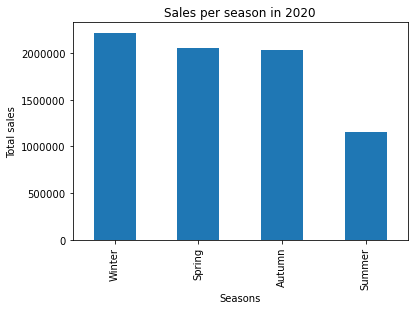
* There is a hike in sales around July every year
* There is downward trend in sales around September in 2020
* The sales have a constant trend

[](https://docs.google.com/spreadsheets/d/10QS3G4BrATPi_OEm6nh8N9F_vGRrCuOLf4nfRgLgYqg/copy)

SKU3 has the highest sales in August and February

SKU6 has the highest sales in October 2020

December had the least sales in 2020



Winter has the highest sales in terms of season

Summer has the least sales in 2020

Modelling

The following models were used in the project:

1.ARIMA (AutoRegressive Integrated Moving Average)

2.ARIMAX (AutoRegressive Integrated Moving Average with exogoneous variables)

3.SARIMA (Seasonal AutoRegressive Integrated Moving Average)

4.SARIMAX(Seasonal AutoRegressive Integrated Moving Average with exogoneous variables )

5.Facebook prophet

The models are trained and evaluated using mean absolute error and class forecast accuracy.

Model evaluation

|  |  |  |
| --- | --- | --- |
|  | Forecast  accuracy | Mean absolute  error |
| ARIMA | 0.92 | 25407 |
| ARIMAX | 0.96 | 24941 |
| SARIMA | 0.94 | 31642 |
| SARIMAX | 1 | 20662 |
| Facebook  prophet | 1 | 17769 |

The success criteria was a forecast accuracy greater than 0.9 and with the least mean absolute error

The model that performed best is the facebook prophet model

Conclusion

Based on the analysis we see that:

- SKU1, SKU3 has the highest sales every year

- SKU2 had lowest sales every year

- July has the most Sales every year

We picked the facebook prophet model with exogenous variables i.e model\_tune

The model demostrated a good forecast accuracy of 1 which means that it is almost close to perfect with the least mean absolute error

The major limitation was working with limited data for different products each being modelled separately.The other limitation is just trying out more models

Based on our findings, we recommend that the company use the facebook prophet model(model\_tune) to inform their decision-making around inventory management and production planning, and to continue to refine and improve the model over time

Next Steps

To use deep learning models that is RNNs and LSTMs to train the data for better accuracy

To scale the facebook prophet model to be able to handle big data

To do more analysis on the data