Objective

The objective of forecasting is to prepare the suppliers and warehouse about the incoming orders and stocks respectively.

Experiments

Approach:

We did brand wise forecasting which will help pepperfry to know which brand is performing well and which supplier to rely upon well in advance.

Data Preparation:

Total we have 6 brands namely – Amberville, Mudramark, Woodsworth, Bohemiana, Casacraft and Mudra. We have prepared separate csv file for 3 brands(Amberville, Mudramark and Woodsworth)for further analysis and didn't analyzed rest brands since it doesn't contain any records.

- The CSV file consists of 2 column – Week and Sales column which are taken from the actual dataset.
- Did exploratory analysis to check the trend and which model to be used.
- Testing for stationary using Ad-Fuller test - gives 5 different values (ADF Test statistic, p-value, Lags used, Number of observations used).
- Here, we did a hypothesis - H0 - it is not stationary (null hypothesis) and #H1 - It is stationary
- If the p- value is less than 0.05, then we can reject the null hypothesis. so it is stationary.
- So while doing brand wise forecasting – Amberville shows low p value so we didn't do any transformation and moved to model building.

Model Building

General purpose of ARIMA model are -
- Visualize the time series
- Make the time series data stationary
- Plot the partial correlation and autocorrelation charts
- Construct the ARIMA and Seasonal ARIMA based on the data trends.
- Use the model to make prediction

Identification of an AR model is done with the PACF.

p- AR Model lags

d - Differencing

q- moving average lags

For an AR model, the theoretical PACF "Shuts off" past the order of the model. The phrase shutsoff means that in theory the partial correlation are equal to 0 beyond that point.

Arima is implemented only when the data is not seasonal , here in Amberville brand we have applied ARIMA.

- For other two brands, ARIMA model didn't performed well and used log transformation to get p value <0.05 and then used SARIMAX for time series forecasting with univariate data containing trends and seasonality.

Accuracy:

To find accuracy, we calculated mean squared error (MSE) and mean absolute error (MAE) for forecasted data and actual data. The main objective was to minimize the error which is proved in the code.

Conclusion

We have filtered out brand which gave value to the company and forecasted and conclude that Amberville brand gives value whole year where as Mudramark and Woodsworth

ABC Classification

- We have taken all the SKU from special price and calculated its count, unit cost and total cost.
- Here, we took the maximum cost of that SKU as a unit cost.
- Total cost is the multiplication of count and unit cost of each SKU. Then, we arranged in descending order for ABC classification.
- We calculated the total valuation of the SKU's by adding total cost and found class A item which consists of 80% of total value and class B items which consists of 15% and class C which consists of 5% of total value.
- Accordingly, we get the class A, Class B, Class C SKU's respectively.

Conclusion:

Here, we can conclude that 12% of class A product values 80% of the total market and 8% of class B product gives value of 15% to the total value. Whereas, only 4% of class C product gives value to the whole market.

Github Code file link:

https://github.com/prithabda/prithabda_Brand wiseForecasting

Additional Assumption and finding

Assumptions before t test on week 1 data:

Discount is given on Wednesday of every week (special price). So for the SKUs for which there is no value in the special price, we are assuming that no discount was given on those products in week 1 as those SKUs were not available on the website on Wednesday.

Analysis was done on week 1 only. All such SKUs where discount was given on Wednesday were found out and their corresponding sales quantities in W1 were grouped together.

Similarly week 1 sales quantities of SKUs for which no discount was given were grouped together. To see if there is an impact of this special pricing on the quantity sold in a particular week, we need to find if there is a significant difference between the means of these two groups.

For this a t-test was done with 95% confidence interval with a sample size of 100.

H0 : There is no significant difference

HA : There is a significant difference.

As the p-value that was obtained (0.0) is less than alpha significance which is 0.05, the null hypothesis is rejected and thus I conclude that there is a significant impact of special pricing on the sales quantity.

Github T-Test Code file link:

https://github.com/prithabda/prithabda_BrandwiseForecasting