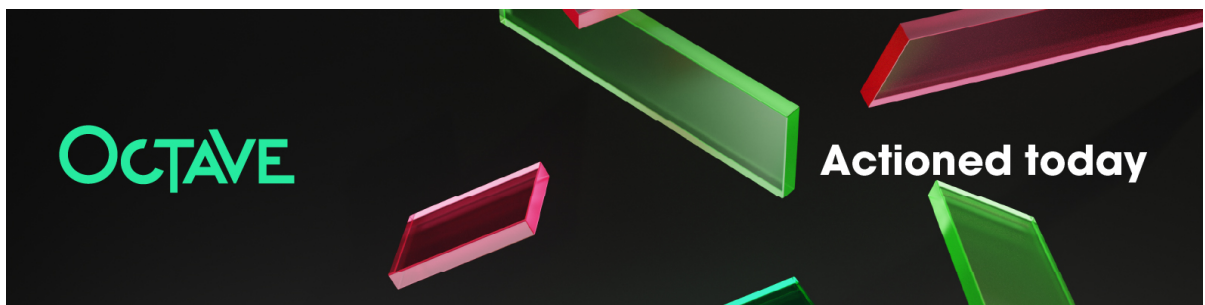


DataStorm 3.0

Semi Final - Case Crack

OCTAVE &
Rotaract Club of University of Moratuwa,
APRIL 4, 2022



1 Sales Demand Forecasting

1.1 Business Problem

On the previous round of Data Storm 3.0, you provided an analytics solution to Retail Chain A to solve their sales demand estimation problem. Since you successfully provided an analytical solution during the previous round, the retailer now expects your support to further refine the previous analytical solution to predict the sales demand.

In the process of predicting the sales demand, at times you would predict a higher sale than the actual sales and you would predict a lower sale than the actual sale. Instances where the predicted sale is higher than the actual sale is known as over forecasting while the instances where the predicted sale is lower than the actual sales is known as the under forecasting.

Also, there are certain items which are on promotion at the store. These items would have a discounted value from their actual selling price during a particular period which would entice customers to purchase these items at an attractive price.

Currently the major concern for the retailer is the under forecasting of items as it would lead to a situation where the retailer would potentially encounter a loss of sales as adequate amount of stock is not available at the store when it should have been. The objective would be to minimize the under forecast error as much as possible while maintaining a lower overall error.

1.2 Data Sources

You are provided with the following data source files to develop a data analytics solution.

1. **Training Data Set** – Historical item sales data set consists of 197 items from 4 different categories with daily sales in quantity units collected through transactions for 5 months.
(*train_data.csv*)
 - (a) CategoryCode: The items are classified into 4 different categories; this column refers to its code.
 - (b) ItemCode: An identifier for each item.

- (c) DateID: The date for the respective sales value. The time period starts from 1st October 2021 to 17th February 2022.
 - (d) DailySales: Sales value in quantity units for the respective item on the given date.
2. **Promotion Data Set** – The data set consists of promotional details related to items. Note that the promotions run on weekly basis (**Friday - Thursday**). (*promotion_data.csv*)
- (a) ItemCode: An identifier for each item.
 - (b) PromotionStartDate: The date the promotion started
 - (c) PromotionEndDate: The date the promotion ended
 - (d) DiscountValue: The amount of discount
 - (e) DiscountType: The type of discount (Percentage/Amount)
 - (f) SellingPrice: The selling price of the item while not in promotion
3. **Validation Data Set** – Similar to training dataset, for specific items with weekly sales in quantity units for the 4 weeks from 18th February to 17th March 2022 (*validation_data.csv*)
- (a) CategoryCode: The items are classified into 4 different categories; this column refers to its code.
 - (b) ItemCode: An identifier for each item
 - (c) Week: The time period starts from 18th February to 17th March (Friday to Thursday). The weeks are as follows:
 - i. w1: 18th February 2022 – 24th February 2022
 - ii. w2: 25th February 2022 – 3rd March 2022
 - iii. w3: 4th March 2022 – 10th March 2022
 - iv. w4: 11th March 2022 – 17th March 2022
 - (d) WeeklySales: Sales value in quantity units for the respective item for the given week.

-
- (e) OnPromo: Flag suggesting whether the item was on promo (1: item on promo, 0: item not on promo)
4. **Testing dataset** – You are required to test your solution using *test_data.csv*. Final weekly sales for the next 4 weeks should be submitted in a CSV format with columns.
- (a) CategoryCode: The items are classified into 4 different categories; this column refers to its code.
- (b) ItemCode: An identifier for each item.
- (c) Week: The time period starts from 18th February to 17th March (Friday to Thursday). The weeks are as follows:
- i. w1: 18th February 2022 – 24th February 2022
 - ii. w2: 25th February 2022 – 3rd March 2022
 - iii. w3: 4th March 2022 – 10th March 2022
 - iv. w4: 11th March 2022 – 17th March 2022
- (d) PredictedSales: You are required to predict the sales value in this column, in quantity units.
- (e) OnPromo: Flag suggesting whether the item was on promo (1: item on promo, 0: item not on promo)

1.3 Deliverables & Evaluation Metrics

In this competition, you are required to submit the following.

1. **Analytical Solution** – In this competition, you are required to create model(s) to forecast the demand for:
 - (a) Items which are on promotion
 - (b) Items which are not on promotion

You must submit a CSV file, which contains the item code (ItemCode), and the predicted sales quantity for each week in the below format. –

CategoryCode	ItemCode	Week	OnPromo	PredictedSales
category_2	4379	w1	1	XXX
category_2	4379	w2	0	XXX
category_2	4379	w3	0	XXX
category_2	4379	w4	1	XXX
category_4	111459	w1	1	XXX
...

Predicted Sales - use the Predicted Sales from the model and the outcome of the descriptive analysis to adjust the prediction.

You can use the Total Mean Absolute Percentage Error (Total MAPE) for evaluation. Total MAPE is given by;

$$\text{Total MAPE} = \frac{\text{Sum}(|\text{Predicted Sales (I,W)} - \text{Actual Sales(I,W)}|)}{\text{Sum}(\text{Actual Sales(I,W)})}$$

X_U = Under Forecast Sales(I, W)

X_A = Actual Sales Under Forecast(I, W)

Under Forecast Error = $\sum |X_U - X_A|$

$$\text{MAPE (Under Forecast)} = \frac{\text{Under Forecast Error}}{\text{Sum}(\text{Actual Sales(I,W)})}$$

where I is for each item, and W is for each week.

Descriptive Analysis – Furthermore, present a descriptive study to reduce the under forecast error while maintaining the overall error at a lower level– **50 Points**

2. **Technical Report** – Report of your solution with clearly defined steps, features, feature engineering steps, modelling approaches, evaluation metrics, all the necessary plots/figures and interesting business findings that you can derive from this analysis – **50 Points**.

- (a) Explain the modelling approach for forecasting of items that are on promotion as well as the forecasting of items which are not on promotion.
- (b) Present a descriptive approach to assess the under forecasting error, clearly explain the approach in detail with your findings
- (c) Based on the results you achieved, list interventions that the management team can take as next steps