

A business unit responsible for Demand Planning has a history of not meeting their goals when it comes down to Forecast Accuracy due to always increasing product portfolio, volatility and volume of data. They have contacted us to help improve their KPIs. As an input, they have prepared historical volumes. We have also contacted marketing team, that later sent values of various trade spends and marketing expenditures for the past and also plan for future year. We prepared the dataset in following manner:

- *X_train.csv* – product key, date, independent variables (*x1-x112*) – historical expenditures (Marketing team)
- *Y_train.csv* – product key, date, target variable – historical volumes (Demand Planning team)
- *X_test.csv* – same structure as *X_train.csv*, independent variables – plans for future expenditures (Marketing team)
- *Y_test.csv* – template for prediction

(Please note that comma is used as a decimal separator.)

Your task is to support business planning process by creating monthly predictions of Sales (*y*) for 119 products, for the next 12 months. The goal of the business unit is to maximize Forecast Accuracy – they are using WMAPE (*weighted mean absolute percent error*) as an error metric. WMAPE should be interpreted as $\frac{\sum(|y - \hat{y}|)}{\sum(y)} \times 100\%$.

In a 2 week time you should:

- Perform appropriate data transformations & feature engineering/selection
- Build a predictive model(s) that will guarantee best performance on test dataset
- Make prediction on test dataset

As solution please send us:

- Prediction in a format of the *y_test.csv* file by filling column *y*.
- A presentation or a report describing your analysis.
- A commented code in an open-source language (or in SAS 4GL) that was used for the exercise.