

Question

An important goal in retail is to be able to forecast article demand as this enables us to plan our business. This can be done using a basic model that considers the average demand over some weeks in the past. But your product owner wants to know if something better can be done. So she contacts the data engineer who gladly provides a dataset. Now it's your turn to show her what is possible with more advanced models!

The Data Set

The data given consists of weekly sales data for 25 articles in the category of "Processed Foods" for a certain country for approximately two years. The problem is to create a demand forecaster for this category that predicts article-level demand one week in advance. That is, given the features of the current week, the model is able to predict the number of units sold in the following week. You may assume that the unit price, the unit discount and the low stock warning are known for the following week in the current week itself.

Here is a short explanation of the data columns.

- * 'sold_qty_units': the number of units sold. This is the variable that has to be predicted.

- * 'art_no': the article number
- * 'art_name': the article name
- * 'hierarchy_level{}_id': product hierarchy level id
- * 'hierarchy_level{}_desc': product hierarchy level description
- * 'date_of_day': the date at the start of the week
- * 'store_count': the number of stores that offer that article in that week
- * 'unit_price_weekly': the weekly price of one unit of that article for that week (possible discounts already included)
- * 'unit_discount_weekly': the Euro value of weekly discount per unit of that sold
- * 'total_cust_count': the total number of customers buying that article that week
- * 'low_stock_warning': this is binary flag and equals 1 if the stock across a majority of stores is low

The Problem

1. Please provide a brief description of the proposed model along with the features that were used. If you create additional features, please state this clearly.
2. Please explain how you measured the performance of your model.
3. You are free to use any language but we prefer Python/R in association with Jupyter notebooks or some form of markdown document. We suggest that you have some explanations in your notebook/markdown document for your presentation.
4. Please note that we do not expect a full-fledged textbook-perfect solution but that we are interested in seeing how you would tackle such a problem and which decisions you take.