Grocery Store Product Forecasting Challenge

Can you forecast the number of products sold per store per week for a local grocery store?

# Competition description

## *About the problem*

Sales forecasting is the foundation of a business’s financial story. Once you have your sales forecast you can create profit and loss statements, cash flow statements and balance sheets, thus helping you set goals for your company. Proper forecasting also ensures you have the right stock at all times and leads to less wasted stock.

Having the skill to create a sales forecast will help you manage anything from a small business up to a large company, where you need to inform investors about your forecasts for a months aquarter or a year.

The objective of this challenge is to create a model to forecast the number of products purchased per week per store over the next eight weeks, for grocery stores located in different areas in the same country. The solution to this challenge can be used by small chain stores to know how much stock to order per week and per month.

# Evaluation

Available error metrics:

* RMSE

# Data

This is anonymised real data. The data looks at 54 different stores in the same country and 33 different products.

The train set contains transaction information for 3 years and 6 months. You are tasked with forecasting the next 8 weeks for the same stores and same products.

**NB: logp1 (log(x+1)) transformation was applied to the label on the testing set, hence we are asking you to apply the logp1 transformation on your predictions before making any submission.**

## *Description of the data*

* Train.csv : historical data containes the daily transactions for 3 years and 6 months in 54 different stores for 33 differnette products family
* Test.csv : contains the IDs of upcoming 8 weeks to forecast for the same stroes and products family as in the Training set
* Holidays.csv : contains the holidays date
* Dates.csv : contains the date with the associated date features (eg : day week of the year , day of the week , month …..)
* Stores.csv: contains the stores location and other information
* SampleSubmission.csv: example of the a submission file

## *Main Variable definitions*

* Target : the total sales for a product category at a particular store at a given date
* Stores\_id: the unique store id
* Category\_id : the unique Product category id
* Date : date in numerical representation
* Onpromotion : gives the total number of items in a Product category that were being promoted at a store at a given date
* Nbr\_of\_transactions : the total number of transactions = happened at a store at a given date
* **year\_weekofyear: the combination of the year and the week of the year,(year\_weekofyear = year\*100+week\_of\_year )**
* ID : the unique udetfier for each row in the testing set : year\_week\_{**year\_weekofyear**}\_{store\_id}\_{Category\_id}

## **Files available for download:**

## **Train.csv** - contains the target. This is the dataset that you will use to train your model.

## **Test.csv-** This is the dataset on which you will apply your model to.

## **SampleSubmission.csv -** shows the submission format for this competition, with the ‘ID’ column mirroring that of Test.csv and the ‘target’ column containing your predictions. The order of the rows does not matter, but the names of the ‘ID’ must be correct.

## **FYI: logp1 (log(x+1)) transformation was applied to the label on the testing set,**

## **hence we are asking you to apply the logp1 transformation on your predictions before making any submission.**

## 

# 