# Assignment 2

# Store Sales - Time Series Forecasting

# Due Sunday End of Week 11

# This Assessments accounts for 15% of total marks.

Use machine learning to predict grocery sales

## Dataset Description

In this project, you will predict sales for the 5 product families (POULTRY, MEATS, LIQUOR, PREPARED FOODS, FROZEN FOODS)sold at Favorita stores located in Ecuador. The training data includes dates, store and product information, whether that item was being promoted, as well as the sales numbers.

## File Descriptions and Data Field Information

## train.csv

* The training data, comprising time series of features **store\_nbr**, **family**, and **onpromotion** as well as the target **sales**.
* **store\_nbr** identifies the store at which the products are sold.
* **family** identifies the type of product sold.
* **sales** gives the total sales for a product family at a particular store at a given date. Fractional values are possible since products can be sold in fractional units (1.5 kg of cheese, for instance, as opposed to 1 bag of chips).
* **onpromotion** gives the total number of items in a product family that were being promoted at a store at a given date.

## Additional Notes

* Wages in the public sector are paid every two weeks on the 15 th and on the last day of the month. Supermarket sales could be affected by this.
* A magnitude 7.8 earthquake struck Ecuador on April 16, 2016. People rallied in relief efforts donating water and other first need products which greatly affected supermarket sales for several weeks after the earthquake.

# Project requirement:

* This is a project for groups up to 5 members.
* The goal is to predict the sales for 15 days.
* You need to conduct a multivariant analysis and predict the sales for each item for the next 15 days.
* You can aggregate all stores for each day.
* Submit one CSV file containing the same attributes as training except for each of 5 family products there should be a column with the predicted prices.
* Submit a report and explain what you have done, include relevant graphs and charts.
* You need to perform any data pre-processing (normalization, transformation, smoothing, etc) necessary.
* You need to deal with and report for any case of Handling anomalies and NA.
* You need to do some exploratory Data Analysis
* Your report should answer the following questions:
  + Is there a Structural Breaks on data?
  + Is there a trend in data?
  + Is there Seasonal Effect in data?
  + Is there any relationship including a causal relationship in the data?
* For prediction fit a Holt-Winters model, a regression model and an ARIMA.
* Which one is better predictor?
* Evaluation and analysis your results.