

# StoreProphet - Time Series Forecasting

Efraín García Valencia  
CC.1001370984

## 1. Descripción del problema predictivo:

Se usarán pronósticos de series de tiempo para construir un modelo capaz de predecir en mayor medida la *cantidad de unidades vendidas* de diferentes productos en las tiendas de la compañía *Corporación Favorita*, a partir de características como familia a la que pertenece el producto (Higiene, Comida congelada, repostería, etc), ventas unitarias del producto y promociones.

## 2. Dataset a utilizar:

Para este proyecto se trabajará con el Dataset de Kaggle de la competición "[Store Sales - Time Series Forecasting](#)" que tiene 3 millones de datos (Id's) agrupados por fecha (01/01/13 hasta 14/08/17) y número de tienda (1 - 54) y que tiene columnas como familia, ventas, en\_promoción, además, se tienen otros archivos que contienen datos relevantes para el desarrollo del proyecto y que pueden ser tomados como columna, por ejemplo, información de eventos festivos con sus fechas, tipo de evento y localización, al igual que la variación del valor del petróleo a través del mismo periodo que los datos.

## 3. Métricas de desempeño:

Como métrica de desempeño para ML se usará la **RMSLE (Root Mean Squared Logarithmic Error)** propuesta en la misma competición, como métrica de negocio podría usarse el porcentaje de mejora en la rentabilidad de "*Corporación Favorita*" gracias a la disminución de costos relacionada a una minimización de la pérdida de alimentos perecederos por culpa del sobreabastecimiento y un aumento de las ganancias al maximizar el ingreso por productos que se encuentran desabastecidos.

## 4. Criterio de desempeño:

Debido a la alta tasa a la que se genera nueva información que debe ser introducida en el modelo para mantenerlo actualizado y los costes de desarrollo y mantenimiento, se espera que para que sea beneficiosa la implementación del modelo la rentabilidad de la empresa debe aumentar al menos un 10%.