1. **Descripción del problema**

La predicción de las cantidades que se han de vender en un lugar ya sea para un determinado producto o para un conjunto de productos similares, ha atraído la atención de las diferentes comercializadoras pues esta nueva herramienta permite la disminución de costos en sus inventarios al optimizar las cantidades que se tienen guardadas o han de perecer. Por esto se pretende realizar una predicción sobre las ventas de diferentes *familias* de productos en la comercializadora Favorita teniendo información de las ventas de los productos en las diferentes tiendas del país.

1. **Data**

El [dataset](https://www.kaggle.com/competitions/store-sales-time-series-forecasting/data) que se ha de utilizar proviene de una competencia de la plataforma *kaggle* la cual consta de 5 archivos los cuales guardan información de:

* Holiday\_events: Este archivo contiene información de los días de festividad en el país de Ecuador, por cada día proporciona el lugar en que se celebra, la fecha, el tipo y si es local o regional
* Oil: Acá se muestra información del precio de la gasolina en el país en las diferentes fechas
* Stores: Contiene la información de las tiendas, la ciudad, el estado, el tipo
* Train: Datos de entrenamiento del modelo donde se guardan las ventas, el tipo de producto, la fecha y si se encontraba en promoción
* Test: Datos de prueba del modelo, contiene las mismas columnas de *train* excepto por las vetas
* Transactions: transacciones realizadas en las tiendas por fechas

1. **Métrica**

La métrica de evaluación principal para el modelo será el error logarítmico medio cuadrático (RMSLE), el cual se calcula mediante la siguiente expresión:

Donde y corresponden a los valores reales y las predicciones respectivamente

1. Para una predicción como estas en una empresa con tantos puntos de venta, donde los inventarios de cada sucursal son considerablemente grandes, se espera que el nivel de predicción sea bastante alto y mas que todo en esos productos de fácil perecimiento. Se pretende con un modelo como este reducir el lo que mas se pueda ese costo que genera la mercancía guardada, valga la redundancia, reduciendo la cantidad de inventario y además reducir los costos de perdida que producen los productos que son más perecederos.