Module 18 Lab:
Sales Forecasting via RAPIDS Linear Regression

OBJECTIVE

Time-series forecasting is to build models to fit on time-series streaming data and use them to predict future observations. The goal of this lab is to implement two forecasting models via standard linear regression and RAPIDS linear regression, respectively, for sales forecasting, and compare their performance.

PREREQUISITES

Install necessary Python packages below.

- **cuML (https://github.com/rapidsai/cuml)** is a suite of libraries that implement machine learning algorithms and mathematical primitives functions that share compatible APIs with other RAPIDS projects.
- **sklearn** is contains a lot of efficient tools for machine learning and statistical modeling including classification, regression, clustering and dimensionality reduction.

INSTRUCTIONS

- Config Google Colab environment to install all required packages
- Download the training data from here (https://www.kaggle.com/c/demand-forecasting-kernels-only/data?select=train.csv) as the whole dataset
- Building time-series samples on the whole dataset for training and testing of the model
  - Set the time window (**tw**) as 29
  - Obtain features including historical item, sales, store in terms of **tw** and use current sales as the target to build time-series samples
- Split the time-series samples into training and testing datasets with the ratio 0.4
- Train a forecasting model with traditional regression model from sklearn on the training data and record the training time

- Train another forecasting model with regression model from cuML on the training data and record the training time
- Complete the forecasting on testing data with these two models and record the forecasting performance such as MAPE
- Compare the training time and the forecasting performance