

# Project 1\_Group30 Framework

## Problem & Data

Walmart Inc wants to keep up with changes in demands. The data is provided and contains weekly sales between 2010/09/05 to 2012/09/07 with 10 given features.

## Predict & Output

After predicting, we combine the Store, Dept, and the Date tags and output the submission csv file with MAE=2253.32 on kaggle and MAE=1665.50, accuracy=0.97 by self-testing.

## Data Preprocessing

We mainly use **DataFrame** due to its numerous well-developed functions. We modified the dataset with the following aspects:

- Merge training and testing data respectively
- Replace NaN by 0
- Convert Dtype "Object" and "string" into "int" and "Float"

## ML Model

We chose **XGBRegressor** in Xgboost because it had eye-catching performances in lately machine learning competitions. We also tested some parameters and finally decided:

n\_estimators = 600,  
max\_depth = 8,  
learning\_rate = 0.45

