# Sales Prediction

The hackthon is conducted to test Machine learning skills of the students.

Following are the instructions:

1. The hackathon contains 2 datasets historical\_train.csv, historical \_test.csv

2. Perform some data preprocessing tasks such as visualizations, data standardization, build machine learning models on train.

3.  What is the best model to use that will provide us with a probability estimate of a sale for each SKU?

3. Predict the model performance on test dataset.

4. Evaluation metric: Accuracy, Recall, Precision on test (should be more than 80%)

**About Data**

The file contains historical sales data (identified with the column titled File\_Type). The historical data shows sales for the past 6 months. The binary target (1 = sale, 0 = no sale in past six months) is likely the primary target that should drive the analysis

* SKU\_number: This is the unique identifier for each product.
* Order: Just a sequential counter. Can be ignored.
* SoldFlag: 1 = sold in past 6 mos. 0 = Not sold
* MarketingType = Two categories of how we market the product. This should probably be ignored, or better yet, each type should be considered independently.
* New*Release*Flag = Any product that has had a future release (i.e., Release Number > 1)

.