Corporación Favorita Grocery Sales Forecasting - Team 3

Can you accurately predict sales for a large grocery chain?

Description

Brick-and-mortar grocery stores are always in a delicate dance with purchasing and sales forecasting. Predict a little over, and grocers are stuck with overstocked, perishable goods. Guess a little under, and popular items quickly sell out, leaving money on the table and customers fuming.

The problem becomes more complex as retailers add new locations with unique needs, new products, ever transitioning seasonal tastes, and unpredictable product marketing. Corporación Favorita, a large Ecuadorian-based grocery retailer, knows this all too well. They operate hundreds of supermarkets, with over 200,000 different products on their shelves.

Corporación Favorita has challenged the Kaggle community to build a model that more accurately forecasts product sales. They currently rely on subjective forecasting methods with very little data to back them up and very little automation to execute plans. They're excited to see how machine learning could better ensure they please customers by having just enough of the right products at the right time.

Data Description

In this competition, you will be predicting the unit sales for thousands of items sold at different Favorita stores located in Ecuador. The training data includes dates, store and item information, whether that item was being promoted, as well as the unit sales. Additional files include supplementary information that may be useful in building your models.
File Descriptions and Data Field Information

train.csv
• Training data, which includes the target unit_sales by date, store_nbr, and item_nbr and a unique id to label rows.
• The target unit_sales can be integer (e.g., a bag of chips) or float (e.g., 1.5 kg of cheese).
• Negative values of unit_sales represent returns of that particular item.
• The onpromotion column tells whether that item_nbr was on promotion for a specified date and store_nbr.
• Approximately 16% of the onpromotion values in this file are NaN.

- **NOTE**: The training data does not include rows for items that had zero unit_sales for a store/date combination. There is no information as to whether or not the item was in stock for the store on the date, and teams will need to decide the best way to handle that situation. Also, there are a small number of items seen in the training data that aren't seen in the test data.

test.csv
- Test data, with the date, store_nbr, item_nbr combinations that are to be predicted, along with the onpromotion information.
- **NOTE**: The test data has a small number of items that are not contained in the training data. Part of the exercise will be to predict a new item sales based on similar products..
- The public / private leaderboard split is based on time. All items in the public split are also included in the private split.

sample_submission.csv
- A sample submission file in the correct format.
- It is highly recommend you zip your submission file before uploading!

stores.csv
- Store metadata, including city, state, type, and cluster.
- cluster is a grouping of similar stores.

items.csv
- Item metadata, including family, class, and perishable.
- NOTE: Items marked as perishable have a score weight of 1.25; otherwise, the weight is 1.0.

transactions.csv
- The count of sales transactions for each date, store_nbr combination. Only included for the training data timeframe.

oil.csv
- Daily oil price. Includes values during both the train and test data timeframe. (Ecuador is an oil-dependent country and it's economical health is highly vulnerable to shocks in oil prices.)

holidays_events.csv
- Holidays and Events, with metadata
- **NOTE**: Pay special attention to the transferred column. A holiday that is transferred officially falls on that calendar day, but was moved to another date by the government. A transferred day is more like a normal day than a holiday. To find the day that it was actually celebrated, look for the corresponding row where type is Transfer. For example, the holiday Independencia de Guayaquil was transferred from 2012-10-09 to 2012-10-12, which means it was celebrated on

2012-10-12. Days that are type Bridgeare extra days that are added to a holiday (e.g., to extend the break across a long weekend). These are frequently made up by the type Work Day which is a day not normally scheduled for work (e.g., Saturday) that is meant to payback the Bridge.
- Additional holidays are days added a regular calendar holiday, for example, as typically happens around Christmas (making Christmas Eve a holiday).

Additional Notes
- Wages in the public sector are paid every two weeks on the 15 th and on the last day of the month. Supermarket sales could be affected by this.
- A magnitude 7.8 earthquake struck Ecuador on April 16, 2016. People rallied in relief efforts donating water and other first need products which greatly affected supermarket sales for several weeks after the earthquake.