EyeOn Supermarket Case Description

Case provided by EyeOn

Introduction

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 Grocery Sales (CFGS), a large Ecuadorian-based grocery retailer, knows this all too well. They operate hundreds of supermarkets, with over 200,000 different product-store combinations.

CFGS has challenged you to help in the forecasting process. They currently rely on subjective forecasting methods with very little data to back them up and very little automation to execute plans. CFGS have collected a large dataset containing detailed point-of-sale data and data on external factors. They're excited to see what insights you bring and what steps you suggest to improve their forecast.

Assignment

We invite you to shed some light on the CFGS case. As you have limited time to work on the case, you'll need to be selective in the insights you want to provide. When presenting your quick insights, particularly think about the background of the challenge:

- Why is it important for CFGS to make their forecast less subjective?
- How can the business context of the case be used in the analysis of the data?

The data and documentation can be downloaded from a dedicated Github repository: https://github.com/jads-nl/execute-eyeon-supermarket. Consider how this data can help you in solving this challenge. How complete is the data? What kind of business dynamics do you recognize in the data? Which patterns do you see in the historical sales (trend, seasonality, events, week patterns, etc.)? What are your hypothesis on the underlying causes for the patterns that you find?

Based on your initial insights, please give your advice to the CFGS sales manager on the period of September to December 2017. Use the data to support your advice and present your insights in a compelling story. We are keen to see not only your results, but also what approach you've taken. Use your creativity to find insights that will raise CFGS's interest and show them the value of their data.

This case intentionally has a lot of open ends. We look forward to seeing how you employ your creativity in tackling this case within limited time. You can use any tool, method or additional data available you can find. Of course, you will need to do the case without help from others.

Data

The data that you have contains historic product sales. There are two versions of the data. In the file *history-per-year.parquet*, you will find data on a granular level: unit sales per item per store per day. The same data is also provided on an aggregated level in the file *history_aggregated.parquet*. Here, you will find the data accumulated to item and week level for the whole country. Additional information about stores, product hierarchy, holidays and even oil prices are provided as supplementary parquet files.

Appendix – Data sources

history-per-year.parquet

Historic point-of-sales data, which includes the actual 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 represents a net return of that particular item.
- The onpromotion column tells whether that item_nbr was on promotion for a specified date and store_nbr.
- Approximately 17% of the onpromotion values in this file are NaN.

NOTE: The 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 a store/date combination, and you will need to decide the best way to handle that situation.

history_aggregated.parquet

- Same as *history-per-year.parquet*, but aggregated to the level of weeks and for the entire country.
- Separate records are included for on promotion sales. You may find both onpromotion sales as well as regular sales in the same week.

stores.csv

- Store metadata, including city, state, type, and cluster.
- Cluster is a grouping of similar stores. In the same way, type is also a grouping of similar stores. The difference between the two groupings is unspecified, and by design of the case.

items.csv

- Item metadata, including family, class, and perishable.

transactions.csv

- The count of sales transactions for each date/store_nbr combination.

oil.csv

- Daily oil price. 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:

- 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 Bridge are 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 day.
- Additional holidays are days added as 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 15th 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.