Store Sales – Time Series Forecasting
Project Proposal

1. Problem Statement
How can Corporación Favorita, a large Ecuadorian-based grocery retailer, increase revenue for the second half of August in 2017 through forecasting sales per product family per store per day?

2. Context
The grocery retailer, Corporación Favorita, would like to increase their revenue and please their customers by having just enough of the right products at the right time. Therefore, we need to forecast the store sales.
Predict a little over, and grocers are stuck with overstocked, perishable goods. Guess a little under, and popular items quickly sell out, leading to lost revenue and upset customers. More accurate forecasting could help ensure retailers supply the right, enough products.

3. Criteria for success
(1) Several time series forecasting models will be developed, for all types of products across all the locality dimensions.
(2) The performance of the models will be evaluated using: (a) MAPE (Mean Absolute Percent Error), (b) Distribution of residuals.
(3) The target MAPE is 10%**.**

4. Scope of solution space: Time-permitting, interpretability analyses will be conducted to study the impact of exogenous variables on the target.

5. Constrains: None identified at this point.

6. Stakeholders: Corporación Favorita Sales Manager, Kaggle competition team

7.Data sources
(1) Training data: date (1/1/2013-8/15/2017), store number (1-54), product family (33 types), number of sales, and on-promotion (number of items being promoted)
(2) Test Data: date (8/16/2017-8/31/2017), store number, product family, on-promotion -> Predict sales for the 16 days after the last date in the training data
(3) Store metadata: city, state, type, and cluster (a grouping of similar stores)
(4) Daily oil price: Ecuador is an oil-dependent country, and its economic health is highly vulnerable to shocks in oil prices.
(5) Holidays and Events, with metadata
(6) Additional Notes #1: Wages in the public sector are paid on the 15th and on the last day of the month.
(7) Additional Notes #2: 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.

8. Methods: ARIMA models, other regression algorithms, PROPHET and PMDARIMA.

9. Deliverables: A GitHub repo containing the Jupyter notebooks developed for this project, a written final report, and a presentation slide deck.