

**Corporación Favorita Grocery Sales Forecasting**

Raksha Kaverappa

Emily Strong

First Draft: November 22, 2017

Second Draft: December 2, 2017

Video: <https://www.dropbox.com/s/6yrlvrwqdf21cor/Team4ProjectProposal.mp4?dl=0>

# **Overview**

Successful forecasting for a grocery store faces the usual balance of satisfying customer demand while not wasting shelf space on excess stock, but has the additional challenge of the risk of overstocking perishable items. For a grocery chain, any model or workflow must also incorporate differences between individual store locations.

Corporación Favorita is an Ecuadorian grocery chain with over 100 stores carrying over 200,000 products. They currently do not use machine learning to predict their sales and are hosting a Kaggle competition to obtain models to improve their forecasting. Because of the importance of oil in the Ecuadorian economy, and the effects of the strength of the economy on consumer habits, oil prices are of particular relevance to any forecast model they might use.

# **Goals**

To forecast unit sales of products by individual store for Corporación Favorita for inventory management, including forecasting new items.

Things we need to ask:

1. What impact do holidays have on the sale of items?
2. How does promotion impact the sale of a product?
3. How can we predict the sales for a new item?

We will offer a web-application to let store employees identify quantities to restock, and quantities to order for new products. We will also identify in our report and final presentation what features are the best predictors of these quantities.

# **Use Cases**

Corporación Favorita corporate employees and managers of individual stores would use this tool to:

1. Improve efficiency in placing orders to restock items.
2. Predict how much of a new product to order.

# **Data**

Data sets are from the Kaggle competition: <https://www.kaggle.com/c/favorita-grocery-sales-forecasting/data>. No data dictionary is provided but the Data page has detailed descriptions for each file.

* Train set of unit sales by date, store, and item ID with flag for whether the item had a promotion
* Test set of dates, stores, and item IDs with flags for promotion
* Store metadata: ID, city, state, type, cluster
* Item metadata: ID, family (e.g. deli, grocery), class and perishable flag
* Total transaction counts for each store by date
* Holiday events
* Oil prices by date

# **Process Outline**

1. Data preprocessing
   1. Join data sets
   2. Handle missing values and outliers
2. Exploratory Data Analysis
3. Build forecasting models with clustering
   1. Unit sales per item and store
      * Feature selection
      * Compare KMeans and manual clustering
      * Models: linear regression, random forests, neural networks, KNN
      * Metrics: MAE, MAPE
   2. Total transaction counts
      * Compare KMeans and manual clustering
      * Models: ARIMA, SARIMA, neural networks
      * Metrics: bias, MAE, MAPE
   3. If oil price is selected as a feature in 3A, we will also develop a time series forecasting model for it
4. Deploy Rest APIs of best model for each store and item cluster on Azure ML Studio
5. Build web application for interacting with APIs
   1. Workflow for predicting individual product sales by store and date

# **Milestones**

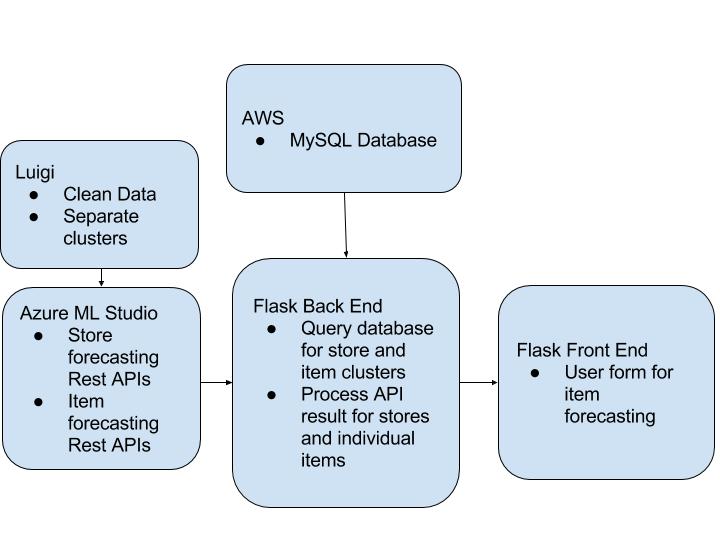
|  |  |
| --- | --- |
| **Days** | **Deliverable** |
| 1-3 | Data wrangling and EDA |
| 4-7 | Model building and selection |
| 8 | API deployment |
| 9-11 | Web application |
| 12-13 | Catch-up for any deliverables that take longer than expected |
| 14-15 | Documentation and presentation |

# **Personas**

1. Inventory manager of individual store
2. Corporate purchasing analyst

# **Deployment Details**

1. Language: Python
2. Database: MySQL
3. Pipeline: Luigi
4. Cloud Tools/Platforms: Microsoft Azure Machine Learning Studio, AWS EC2
5. Web Framework: Flask



# **User Experience Design**

# 

# **References**

Image of groceries from WikiMedia Commons <https://commons.wikimedia.org/wiki/File:Vegetables_in_supermarket_-_DSC04975-001.JPG>

Kaggle Competition: <https://www.kaggle.com/c/favorita-grocery-sales-forecasting>