WyngCommerce: Data Science: Exercise 1

Problem Statement

This exercise has been designed to replicate real-world problems that we solve at WyngCommerce. The datasets closely reflect real-world data of retailers in India and the problem statement is very core to our work.

A leading fashion retailer needs to plan their supplies keeping in mind expected demand in the upcoming quarter. They are operating in 3 stores with 1,000+ products - each identified by a 'SKU_Code'. The data provided is historical sales and product availability levels in the period from 01-Jan-2016 to 31-Dec-2017, and the forecast period is 01-Jan-2018 to 31-Mar-2018. The problem statement is to forecast the SKU-wise demand for all the available products in the forecast period.

Data Description

1. Secondary Sales

FileName: WC_DS_Ex1_Sec_Sales.csv
Description: Store-wise SKU-wise daily sales records with quantity, MRP and actual selling price
Columns:

- Store_Code: The code for the store at which sales took place (3 values)
- SKU_Code: SKU stands for Stock-Keeping Unit; One SKU corresponds to one specific product
- Category: This is product category; all products are divided across 4 categories
- Date: Date of the sale to the customer
- Sales_Qty: The number of units sold
- MRP: Maximum Retail Price of the product
- SP: Actual selling price of the product - changes by date & store

2. Primary Sales

FileName: WC_DS_Ex1_Pri_Sales.csv
Description: Store-wise SKU-wise records of stock movements from retailer warehouse (WH) to stores, and vice versa
Columns:

- Store_Code: Same as Secondary Sales
- SKU_Code: Same as Secondary Sales
- Category: Same as Secondary Sales
- Date: This is the date when the products left the WH
- Qty: The no. of units of each product sent to store
- *NOTE:* if the Qty is negative, then that implies products were moved back from store to WH, and in that case, the date is when they were received back in WH

3. Inventory Snapshots

FileName: WC_DS_Ex1_Inv.csv
Description: Snapshots of no. of units of each products held in each store on specific dates (once per month)
Columns:

- Store_Code: Same as Secondary Sales
- SKU_Code: Same as Secondary Sales
- Category: Same as Secondary Sales
- Date: This is the date on which inventory snapshot was taken
- SOH: The no. of units of each product available at the store on that date
- *NOTE:* For each date, typically a store has two inventory readings: opening inventory (at the start of day) and closing inventory (at the end of day); the inventory snapshots here correspond to closing inventory

Retailer Supply Chain

1. This retailer works with 3 stores - they are all located in the same city
2. As products sell in the store, the retailer plans replenishments to be sent to stores from its warehouse - these are termed as primary sales
3. If certain products haven't sold for a long time in the stores, or if they need to make space for new products - then the retailers sends back some products from stores to warehouse. They are termed as 'return-to-warehouse' (part of primary sales data)
4. The time taken for primary sales or return-to-warehouse is assumed to be instantaneous for this exercise i.e. 0 days
5. While in reality, stores have limited space - but for our calculations we can assume the store space to be unlimited
6. The retailer keeps running several promotions or discounts on different products at different points in time; these promotions can vary by stores as well. The degree of promotion can be measured by the discount between actual selling price and MRP
7. The MRP of the product can also vary by time & store e.g. a product ABC maybe selling at Rs. 999 in Oct'16 but the MRP may get changed to Rs. 799 by Feb'17, and it may even increase as well

Challenges

1. **Missing values:** Inventory snapshots are once per month, and the forecasting problem may require creating more snapshots
2. **Inconsistent data:** The price points, inventory levels, etc. may not be entirely consistent e.g. we may see a sale of a particular product on a day where the inventory snapshot maybe showing zero inventory on that day
3. **Outliers:** The sales records or discount levels may have some outliers - which might need some treatment before finalizing predictions
4. **Seasonality:** Some of the products may have some seasonal demand - which needs to be taken care of before finalizing predictions

Answer the following

1. Aggregate the Sales_Qty for each Store-SKU at a month level; detect any Outliers in the Sales_Qty for each Store-SKU combination and apply an outlier treatment on the same. Specify the outlier treatment technique.
2. Estimate the level of promotions (Discount%) for each Category-Store level at a month level - remove any outliers / inconsistencies from this, and specify the technique used; the level of promotions is defined as Discount% = (1 - sum of SP / sum of MRP)
3. Estimate the inventory levels at a weekly level for each Store-SKU by interpolating missing values from data on secondary and primary sales; the following equation holds true in general: (*you can do this for a shorter period of Jan 2017 to Mar 2017*)

   Closing inventory on day [t] = Closing inventory on day [t-1]
   - Secondary (sales - returns) on day [t]
   + Primary (sales - returns) on day [t]

   NOTE:
   a. Secondary sales is the file named “WC_DS_Ex1_Sec_Sales.csv” - and it refers to sales from stores to customers (and returns by customers)
   b. Primary sales is the file name “WC_DS_Ex1_Pri_Sales.csv” - and it refers to stock movements from retailer WH to stores (and returns back to WH)
   c. Returns in both datasets are indicated by negative values in ‘Sales_Qty’ and ‘Qty’ fields respectively
4. The inventory estimations in Question 3 will have data inconsistencies - take any assumption to resolve them and explain that assumption
5. Using the Secondary sales data and inventory series from Question 3, determine average out-of-stock percentage (OOS%) for each Category-Store combination at a monthly level; the OOS % is defined as:

   OOS % = 1 - {Average of no. of unique SKUs in stock each day
   / No. of unique SKUs in stock over the entire month}
   (for each Category-Store combination each month)

(*Again - do this for a short period of Jan 2017 - Mar 2017; for forecasting you can assume that the retailer will experience similar OOS levels in Jan-Mar 2018*)

6. Using the historical secondary sales, inventory series, OOS% levels and promotion levels, determine the demand for each Store-SKU combination at a monthly level for the forecast period; use any forecasting technique that you're comfortable with (you may use multiple techniques)
7. Explain the approach for Question 6 clearly e.g. dividing data into train, validation and test sets, choice of technique used, metric(s) used to evaluate the results
8. If any of the above steps is becoming computationally too expensive or taking too long; you are free to either simplify them or reduce the complexity (e.g. impute weekly inventory positions instead of daily)

Submit the following

1. All the codes used to complete the analysis - they maybe in any programming language of your choice
2. A report (word doc) with answers to all the questions above; feel free to write as long a report as needed, and add visualizations wherever you feel necessary
3. Submit all the codes & report by zipping all the files into a single folder named as "WC_REPORT_FirstName_LastName.zip", and send that back by reply email
4. The time limit for completing this exercise will be specified in the email sent by us
5. In case of any queries, feel free to reply by email and ask - we will respond within 24 hours of receiving the query