Demand Forecasting

Predict MasterSKU level order quantity for each standard dealers group by week or month

- Data: 4 years data; 8000 STD ; 60+ MasterSKU; 1800,000 data
- Step 0: data cleansing
  - Date
  - Dealers
  - MasterSKU
  - Quantity
- Step 1: Analyze STD & HYB
  - Select stable STD & HYB dealers with 3+ years order history
- Step 2: Segment stable STD & HYB dealers
  - Use features like order frequency, order quantity, region, order quantity fluctuations to segment dealers into different class (k-means clustering).
- Step 3: Analyze masterSKU
  - Select mature products with long order history (18 masterSKUs)
  - Time series: white noise testing (Hypothesis test: there is no information contained in the time series)
- Step 4: Modeling
  - Time series:
    - Before modeling:
      - Stationary analysis (Dickey-Fuller test)
        - Yes: Redo analysis after differencing steps
      - Autocorrelation/ partial-autocorrelation plot
    - Modeling:
      - ARIMA(p, d, q) and Seasonal ARIMA(p, d, q) x (P, D, Q)s
      - Test all combination of order value p,d,q and P,D,Q,s in ARIMA and SARIMA
      - Select the combination with lowest AIC & BIC value
      - Do one year prediction and compare them with actual data
        - Root mean square error (can't compare results from different data set)
        - Mean absolute percentage error (easy to compare results from other data set; if testing data contain 0, mape will be infinity)
        - Symmetric mean absolute percentage error
      - Set Naive model as a baseline and compare the result from Fbprophet
        - [describe naive model?]
    - Difficulties:
      - Some masterSKU's data is not enough (less than 40)
        - Combine several similar product together to model

        - Predict weekly
    - Supervised machine learning
        - Feature engineering w/ training data
            - Multidimensional customer clustering (KMeans)
                - Overall order quantity
                    - Sum(order quantity)
                - Normalized average order quantity
                    - Overall / # of days being a customer
                - Product diversity
                    - # of unique masterSKU * # of product category each dealer has ever purchased
                - Quantity variance
                    - var(order quantity)
                - Reorder cadence
                    - mean(reorder interval in days)
                - Patient rate
                    - Patient = requested_date - order_date > 0
                    - # of patient / # of orders
                - New product purchase rate
            - Price drop %
            - New product binary flag (2 mo cutoff)
        - Heteroskedasticity:
            - Bartlett's test p = 0.0
            - plot OLS residuals
        - Features: month, masterSKU, product category, customer cluster, new product flag, price drop %
        - Dependent variable: each customer group's monthly sum order quantity per masterSKU
        - Least square regression
            - OLS
            - Polynomial OLS (d=2 to 4)
            - WLS
        - KNN regressor
        - Random forest regressor
        - Model selection: cross validated RMSE and AIC; output best SMAPE

Digital Marketing

Exploratory analysis

- Overall product order quantity & popularity
- Trend and seasonality (autocorrelation plot)
- Color

- Customer geographic info

Analysis of customer behavior in websites

- Correlation between actions in product detail page and purchase
- What home pages are most effective for purchasing or video finish
- Conversion rate by traffic source and type
- Most performed actions on website