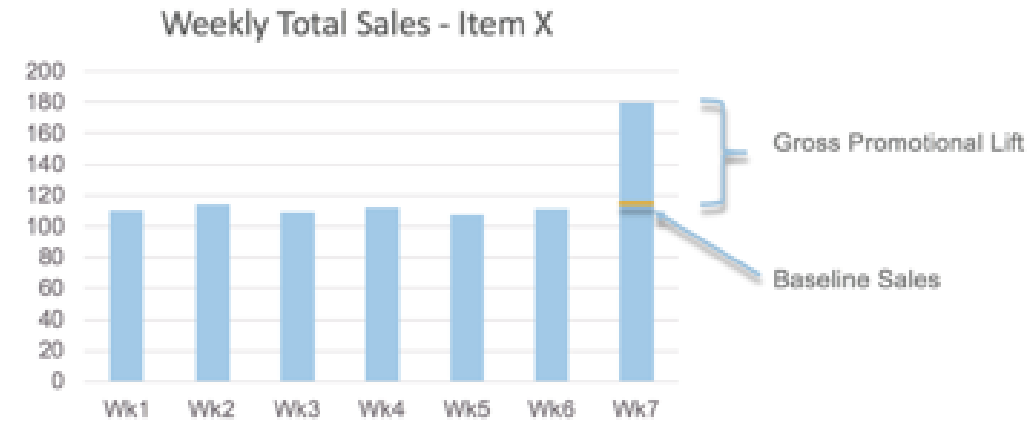
# Baseline Model

For MVP in Q4 2019, Only Store TPRs are considered to measure Promo Effectiveness




Total Sales = Baseline Sales + Gross Promotional Lift

True Lift = Gross Promotional Lift – Promo Stacking – Pull-forward – Cannibalization + Halo – Advertising Expenses

- **Baseline model is an Ensemble model which has**
  - State space model is used for baseline
    - It takes care of organic growth, seasonality, inflation, previous promotion
    - **Kalman Filter** is the State space model that is used
  - Interpolation techniques:
    - Spline method is the interpolation method that is used
  - Moving Average
    - 3 week Weighted moving average is used
- Recent 10 weeks sales data has been given more weight

- 2.5 years of data is being used for computing baseline
- In the promo score card Summary view
  - Success Metric is color coded: Green (all 3 metrics are positive), Red (2 or 3 of metrics are negative), Yellow (1 metric is negative)
  - Sales Lift, Margin Lift and Units Lift are the 3 metrics that are used
- Lift = (Actual - Baseline)/Baseline * 100

- There are two data sources – one is from POS data for transaction sales and another is Promotion data from Main frame (ADV, Promolink)
- Basket Impact is at week level wherever the item was part of
- Assortment -> sales lift and margin lift is with respect to last year sales data
- Customer type is DIY and Pros

**Inputs**:

- Input to Baseline Model will be the Transaction tables (Sale) and Promotion data
  - Initially the sale data will be at item-location-day level for items that are on promotion
  - Note: On a given week, 2000-6000 items are on promotion. out of this 90% of the promotions are TPR
- Data science team will aggregate data at item-week level
  - Aggregation of data happens by considering sales data of all the promo items at chain level (chain level can be set of stores or full chain i.e. 1700 stores)
  - Promotion type, Average sale value for the promo item historically is also considered as an input

**How the algorithm works**:

- There are some challenges with respect to data here
  - Promo planning week (Thursday to Wednesday is different from fiscal week (Sat to Sunday - Standard 4-5-4 retail fiscal week). so comparison of promotion event planning with fiscal week sales maybe difficult
  - Promo activation at store could be late and promotion can run only for say 2 days

        - Promo could be run only in some stores and not full chain
- The Input data is cleaned by removing outliers, noise and then fed into the data model
- Kalman filter, Spline method and 3 week weighted average models are used here.
- The output of state space model is compared with output of other models like Spline model and 3 week average model to check whether output is correct or not. whichever result is closer to sales is considered as baseline. Usually Median of these baseline numbers computed using different methods is taken as baseline
- For weeks that are on promotion in previous weeks, baseline is computed and stored in the tables by the model. So when a new promotion is run, input used to compute lift and baseline is baseline sales of previous promotions
- Since we have 2.5 years of data, the model takes care of
    - Organic growth by the trend line drawn on the sales data.
    - Seasonality by crest and troughs of the weekly data

**Outputs**:

- The sale during a week is divided into baseline sales and Lift

**Data validation proofs:**

- A non promo week is taken and based on previous promotion lift, a sales lift is added. Then baseline is computed to see if it matches the actual sales during a non promo week

**Assumptions:**

- There is no negative lift because of promotion except for luxury items (ex: iphone - sales may decrease if item is put on promotion). At Lowe's there are no such luxury items that we are selling

**Questions**:

1. In the sales data that is fed into the model, even return sales is also considered. this may lead to -ve lift in sales. return can be because of product quality issues or any other reasons. Promotion effectiveness should consider only the normal sales data. Why is return sales computed