

# Market Mix Modelling for Elec Kart

## SUBMISSION

1. Amit Gupta
2. Raghavendra D

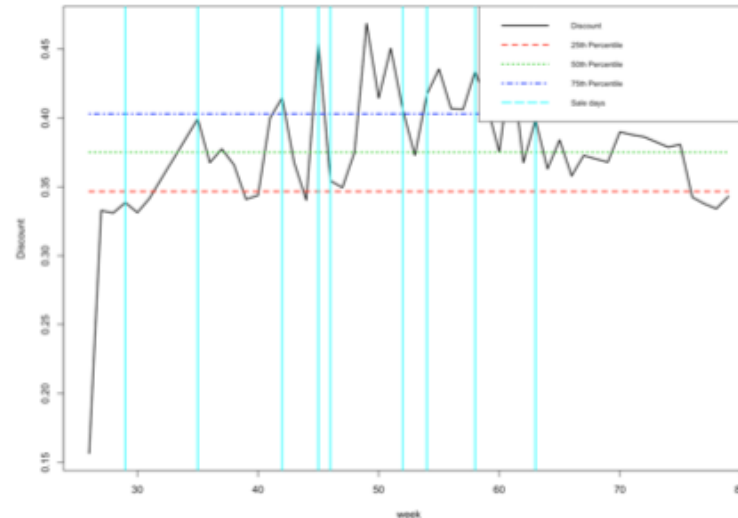
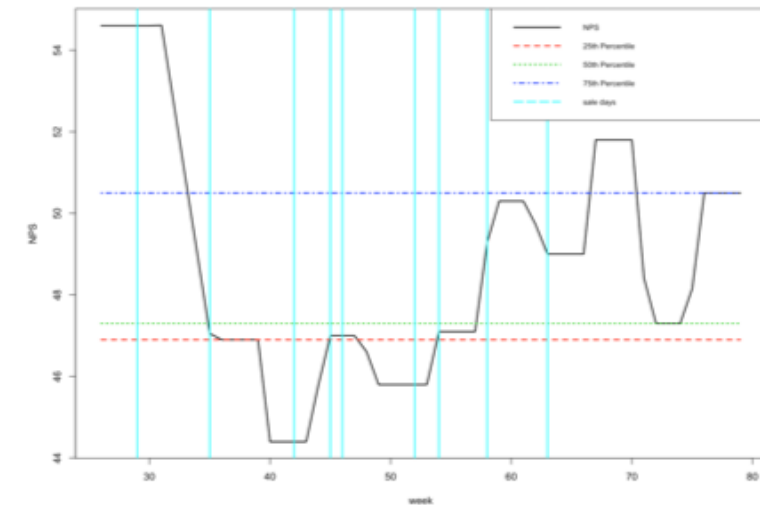
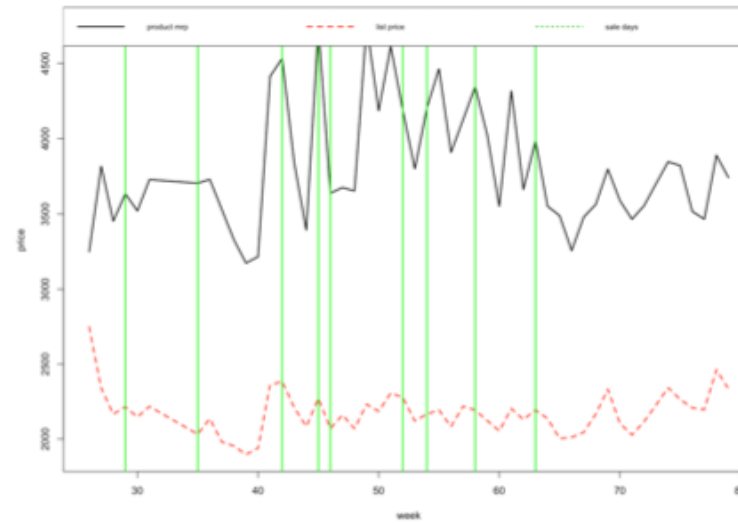
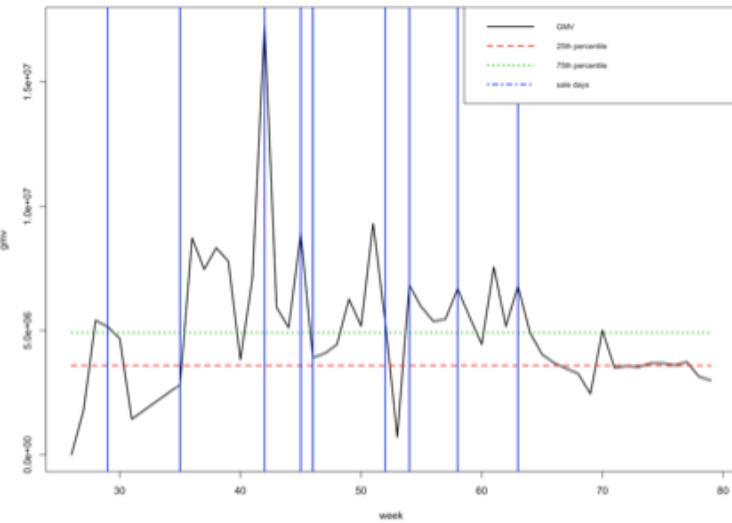
- ElecKart is an e-commerce firm specializing in electronic products. Over the last one year, they had spent a significant amount of money in marketing. Occasionally, they had also offered big-ticket promotions (similar to the Big Billion Day). They are about to create a marketing budget for the next year which includes spending on commercials, online campaigns, and pricing & promotion strategies. The CFO feels that the money spent over last 12 months on marketing was not sufficiently impactful, and, that they can either cut on the budget or reallocate it optimally across marketing levers to improve the revenue response.

## **What is Expected out of this analysis -**

- You have to create market mix models for three product sub-categories -
  - camera accessory,
  - home audio and
  - gaming accessory.
- Also, the models have to be built at a weekly level.
- Choose the best ones for each of the three product sub-categories and explain them through presentation.

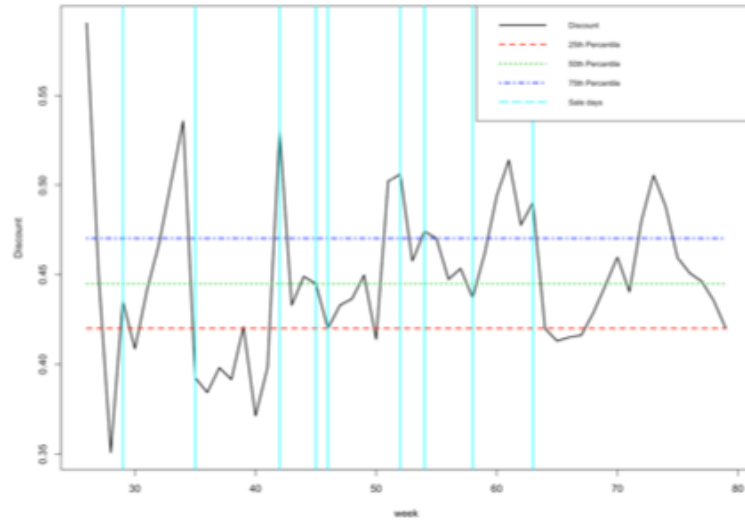
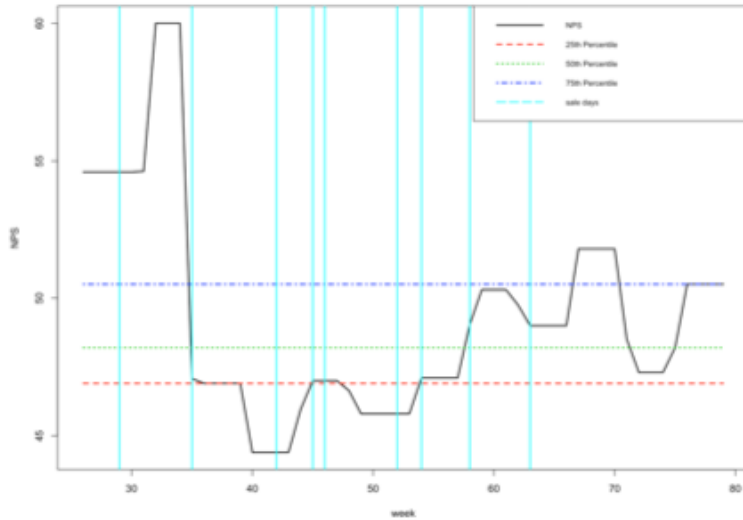
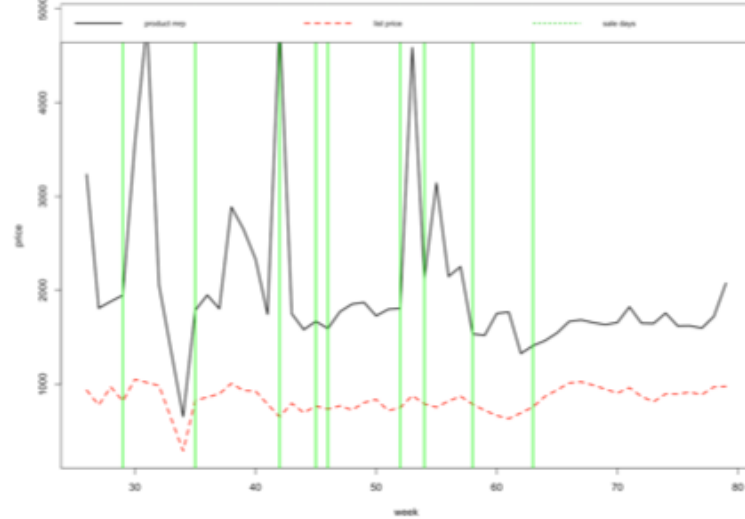
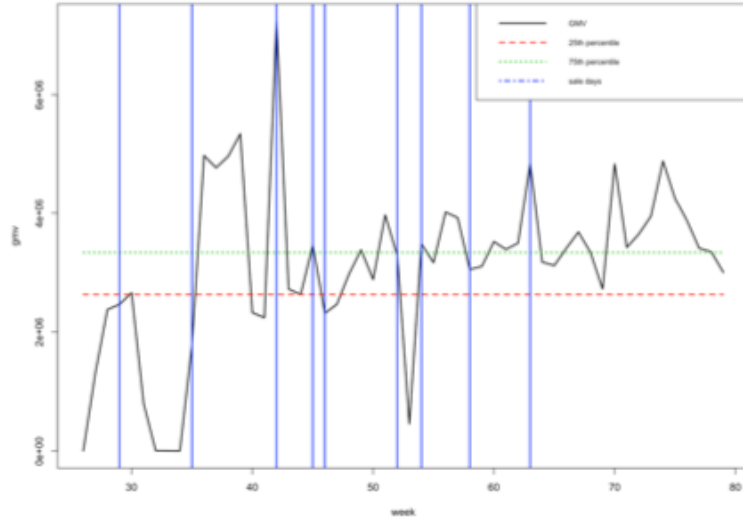
- FSN ID: The unique identification of each SKU
- Order Date: Date on which the order was placed
- Order ID: The unique identification number of each order
- Order item ID: Suppose you order 2 different products under the same order, it generates 2 different order Item IDs under the same order
- ID orders are tracked by the Order Item ID.
- GMV: Gross Merchandise Value or Revenue
- Units: Number of units of the specific product sold
- Order payment type: How the order was paid – prepaid or cash on delivery
- SLA: Number of days it typically takes to deliver the product
- Cust id: Unique identification of a customer
- Product MRP: Maximum retail price of the product
- Product procurement SLA: Time typically taken to procure the product
- Apart from this, the following information is also available:
- Monthly spends on various advertising channels
- Days when there was any special sale
- Monthly NPS score – this may work as a proxy to ‘voice of customer’

- The dataset has the following properties:
  - 1633642 observations with 25 variables under different sub categories.
  - The following columns have missing values and since the number is low (29.74%) they can be imputed from the dataset.
    - Cust\_id has 4904 null values
    - Pincode has 4904 null values
    - gmv has 4904 null values
  - The following attributes have negative values (in low numbers) which had to be treated by Convert them to factors and then convert back as integers to make them proper.
    - Pincode
    - Order\_id
    - Order\_item\_id
    - Cust\_id
  - The data for December 2015 and January 2016 were observed to be missing.
  - Since the analysis has to be done on the sub-categories - camera accessory, home audio and gaming accessory the data can be filtered out of the dataset.
  - Additionally only the data for one year i.e. from 1-Jul-2015 to 30-Jun-2016 has to be considered the data can be further filtered out as well.
  - Additional attributes
  - Univariate Analysis was performed on Year, Month, Week, Day
  - Bivariate Analysis



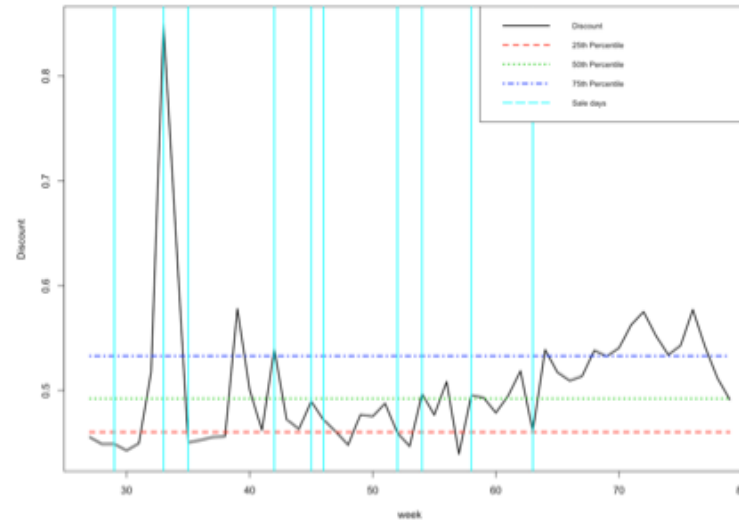
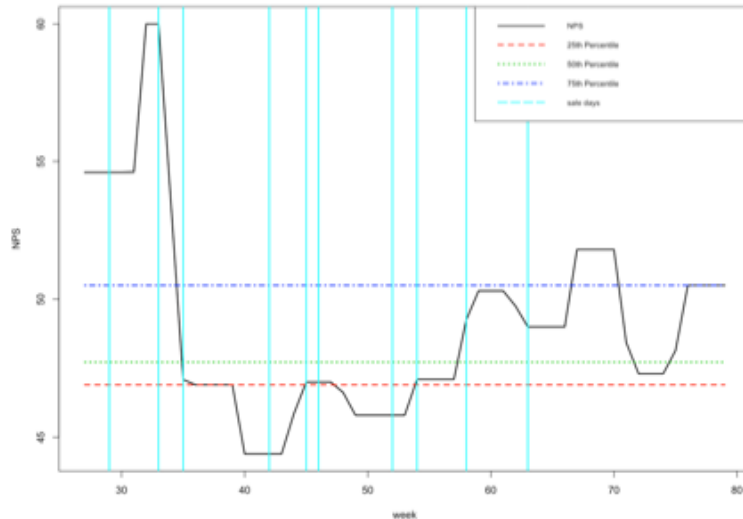
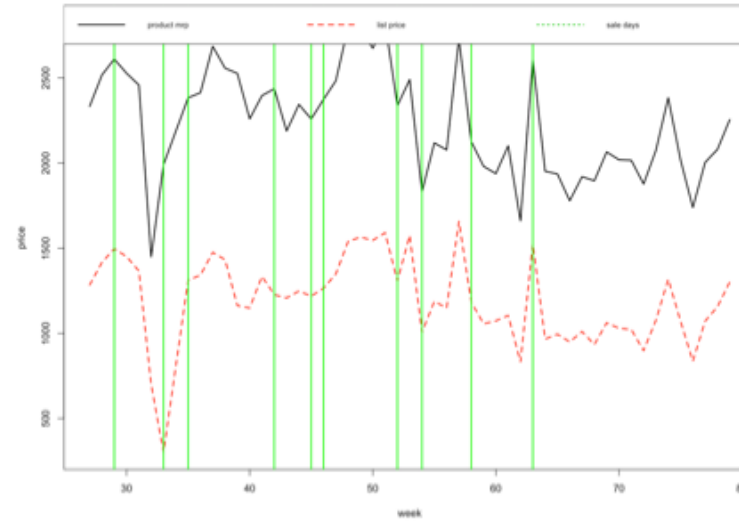
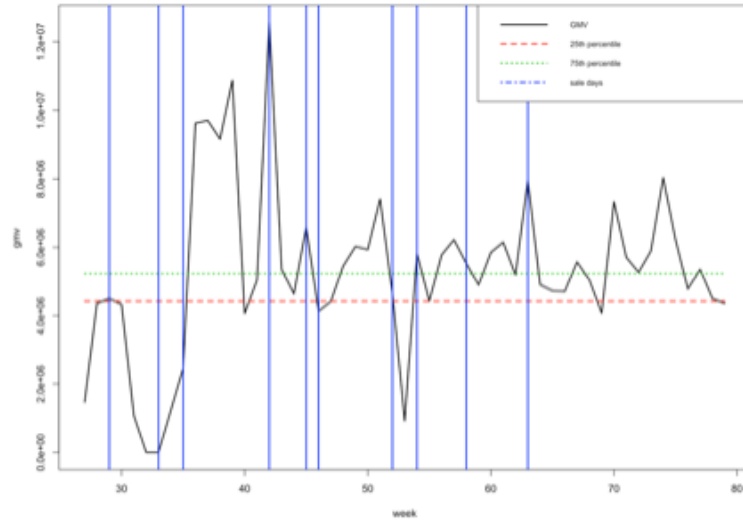
The following are the important observations from the EDA carried out for Home Audio.

- There is a clear indication of gmv increasing during the sale days.
- First 2-3 sales were a bit down due to customers unaware of the sale happening and the deals given.
- The graphs also strongly indicate hysteresis and pantry loading effect after each sale.
- The NPS was good initially due to less gmv in sales. Once the sales picked up NPS went down.
- This also indicates that the company took corrective actions to increase NPS subsequently during the sales day.



The following are the important observations from the EDA carried out for Gaming Accessory

- There is a clear indication of gmv increasing during the sale days.
- First 2-3 sales were a bit down due to customers unaware of the sale happening and the deals given.
- The graphs also strongly indicate hysteresis and pantry loading effect after each sale.
- The NPS was good initially due to less gmv in sales. Once the sales picked up NPS went down.
- This also indicates that the company took corrective actions to increase NPS subsequently during the sales day.



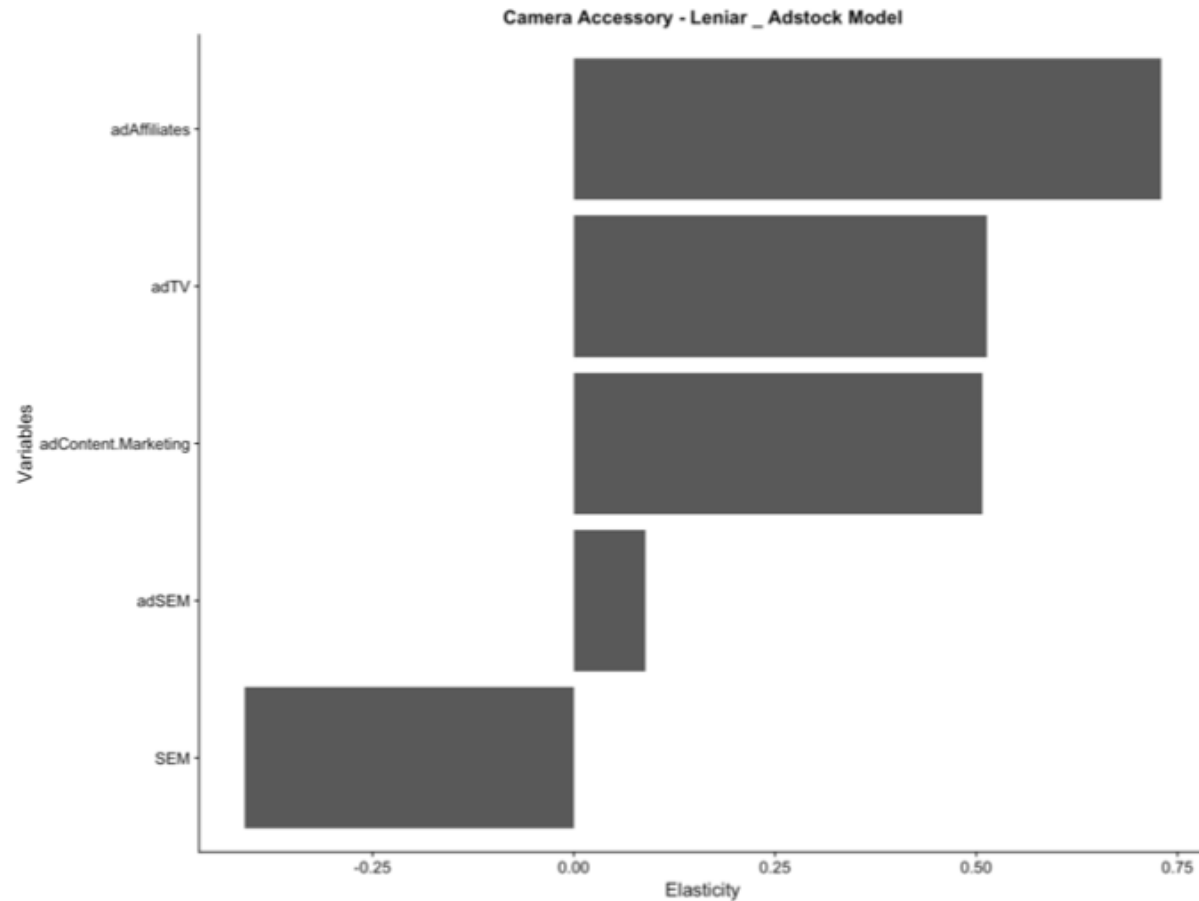
The following are the important observations from the EDA carried out for Camera Accessory

- There is a clear indication of gmv increasing during the sale days.
- First 2-3 sales were a bit down due to customers unaware of the sale happening and the deals given.
- The graphs also strongly indicate hysteresis and pantry loading effect after each sale.
- The NPS was good initially due to less gmv in sales. Once the sales picked up NPS went down.
- This also indicates that the company took corrective actions to increase NPS subsequently during the sales day.

Model Name	Driving KPI's	Adjusted R <sup>2</sup>
Linear	product_mrp , markdown ,product_procurement_sla , prepaid_percentage , NPS_Score ,Digital	0.3435
Linear + AdStock	NPS_Score ,SEM , adTV ,adContent.Marketing , adAffiliates ,adSEM	0.4295
Multiplicative	list_price ,product_mrp markdown ,product_procurement_sla , Affiliates	0.793
Multiplicative + Adstock	TV,Online.marketing ,adOnline.marketing	0.7806
Koyck	product_mrp, Digital , Sponsorship , Affiliates , SEM , Radio, Other	0.3968
Koyck +Adstock	Online.marketing, Affiliates,Radio, Other ,Sponsorship	0.2504
Distributed Lag	Affiliates, shelf_price_inflation_1 ,shelf_price_inflation_2 ,shelf_price_inflation_4 , markdown_inflation2 ,count_promotion_days_in_week.1	0.5399
Multiplicative + Distributed Lag	product_mrp,TV , markdown_inflation1, gmv.1	0.817

- The selection of the best performing model is based on the following criteria.
  - Presence of critical KPI's driving the business.
  - Adjusted R<sup>2</sup> and the results of k cross validation on the model built. (K=10)
  - Elasticity analysis of the KPI's in the model.
- Based on the above selection criteria we will recommend using **Linear + Adstock model** over Multiplicative + Distributed Lag & Multiplicative models due to over dependency on gmv, product mrp, procurement sla.
- Linear + AdStock model shows clear dependency on content marketing showing the quality of the accessory, carry over effect of advertisement and SEM. NPS score indicates the reliability of buying the product.





Base on the elasticity analysis of the Leniar + Ad Stock model the following are the recommendations.

Increase:

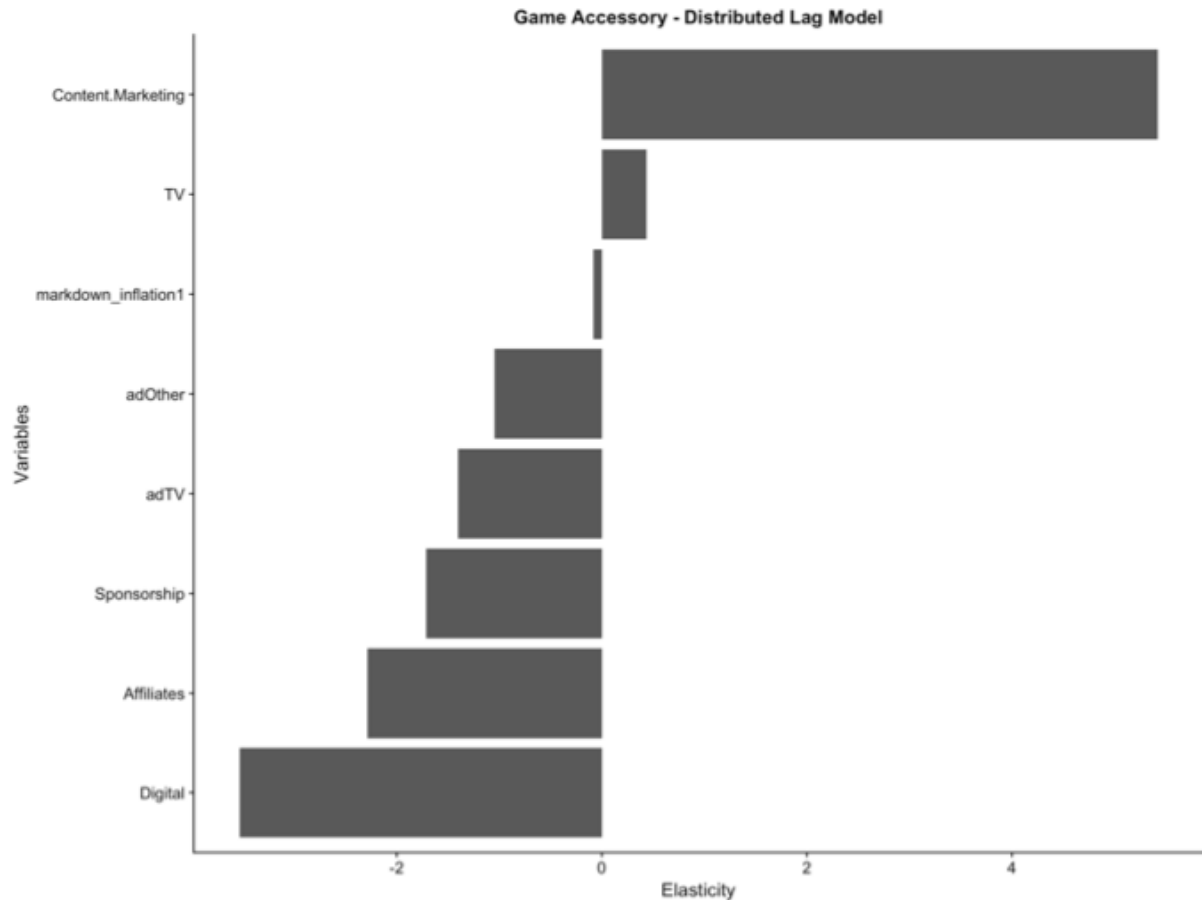
- Affiliation with great products.
- Increase Ad Spending on TV.
- Increase Ad Spending on creating great content for advertising.

Relook:

- At SEM spending as the ROI is less.

Model Name	Driving KPI's	Adjusted R <sup>2</sup>
Linear	TV , Digital, Sponsorship, Content.Marketing ,Online.marketing , Other	0.2792
Linear + AdStock	Content.Marketing ,adTV ,adContent.Marketing ,adOnline.marketing ,adSEM	0.3134
Multiplicative	product_mrp ,Digital, Online.marketing , SEM	0.6405
Multiplicative + Adstock	Sla,adAffiliates	0.6169
Koyck	TV , Digital , Sponsorship , Content.Marketing, Affiliates , Other	0.3473
Koyck +Adstock	Content.Marketing , adTV , adContent.Marketing, adOnline.marketing, adSEM , adOther	0.4172
Distributed Lag	TV , Digital, Sponsorship, Content.Marketing, Affiliates , adTV , adOther, markdown_inflation1	0.6281
Multiplicative + Distributed Lag	product_mrp ,Content.Marketing , Affiliates, Radio, list_price.1, markdown.1	0.6998

- The selection of the best performing model is based on the following criteria.
  - Presence of critical KPI's driving the business.
  - Adjusted R<sup>2</sup> and the results of k cross validation on the model built. (K=10)
  - Elasticity analysis of the KPI's in the model.
- Based on the above selection criteria we will recommend using **Distributed Lag model** over Multiplicative + Distributed Lag & Multiplicative models due to over dependency on gmv, product mrp, procurement sla.
- Distributed lag model shows clear dependency on TV, Digital, Sponsorship to any sports event to promote the products. Additionally it also is dependent on periodic price markdowns.



Base on the elasticity analysis of the Leniar + Ad Stock model the following are the recommendations.

Increase:

- Increase Ad Spending on creating innovative content for advertising that can increase gmV.
- Increase Ad Spending on TV only during sales as it is showing very less carry over effect.

Relook:

- At SEM spending as the ROI is less.

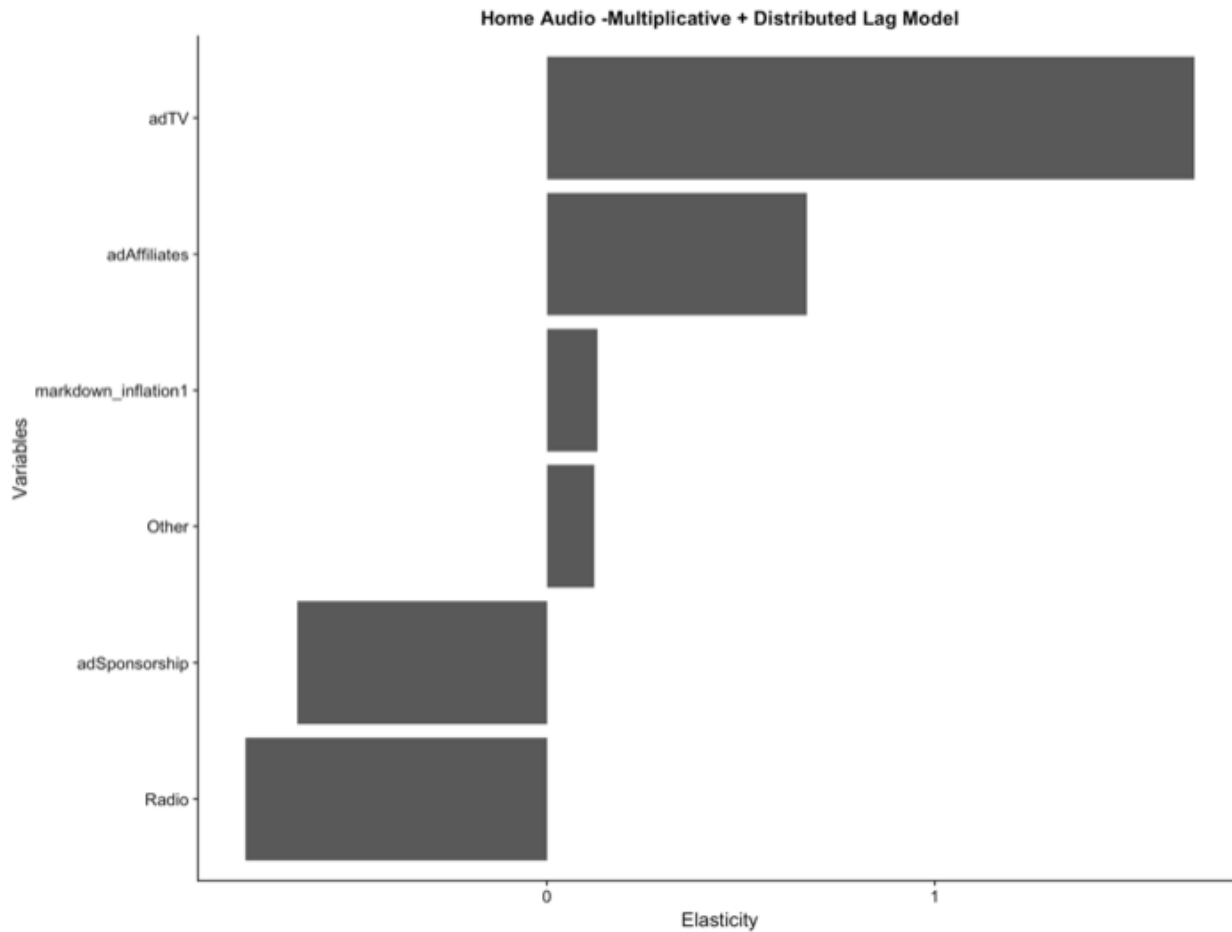
Decrease:

- Spending on Digital, Affiliates and Sponsorship as they clearly show a negative trend.

Overall Game Accessory is a category that is dependent on current effect.

Model Name	Driving KPI's	Adjusted R <sup>2</sup>
Linear	product_mrp , sla , Sponsorship, Affiliates, Other	0.3447
Linear + AdStock	product_mrp, sla ,NPS_Score ,Digital ,SEM ,adTV , adContent.Marketing,,adAffiliates ,adSEM	0.67
Multiplicative	sla , markdown, Online.marketing , Affiliates	0.3116
Multiplicative + Adstock	sla , markdown, Online.marketing, adTV , adSponsorship, adOnline.marketing, adAffiliates	0.4075
Koyck	product_mrp , sla , Sponsorship	0.3835
Koyck +Adstock	product_mrp , prepaid_percentage, Digital,adAffiliates , adOther, lag_gmv	0.2475
Distributed Lag	Content.Marketing , Affiliates , Radio, Other, adTV , adSEM, count_promotion_days_in_week.1, gmv.2	0.424
Multiplicative + Distributed Lag	Radio ,Other, adTV , adSponsorship , adAffiliates, markdown_inflation1	0.421

- The selection of the best performing model is based on the following criteria.
  - Presence of critical KPI's driving the business.
  - Adjusted R<sup>2</sup> and the results of k cross validation on the model built. (K=10)
  - Elasticity analysis of the KPI's in the model.
- Based on the above selection criteria we will recommend using **Multiplicative + Distributed Lag model** over other models due to the following reasons.
  - Adjusted R<sup>2</sup>
  - Critical business KPI's, carry over effect of TV advertisements, sponsorship, affiliates that can drive more sales through people absorbing it from home and aspiring for better home audio experience.



Base on the elasticity analysis of the Multiplicative + Distributed Lag model the following are the recommendations.

Increase:

- Increase spending on Affiliates.
- Look at giving frequent markdowns to increase sales.

Decrease:

- Decrease spending on the sponsorship
- Re-evaluate marketing spends on Radio.