Market Mix Modelling

Objectives

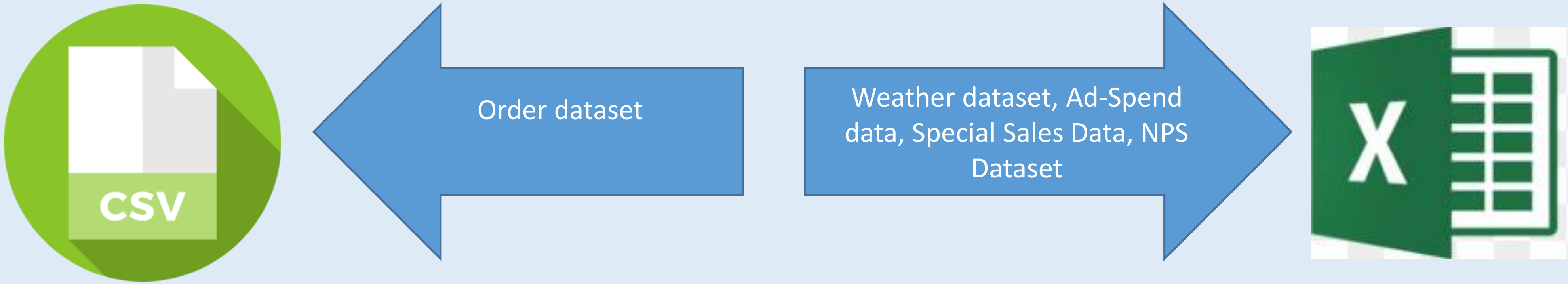
- To develop a market mix model to observe the actual impact of different marketing variables over the last year.
- To recommend the optimal budget allocation for different marketing levers for the following year.
- To understand other factors/ KPIs having impact on sales.
- All these tasks to be done for 3 product sub-categories: Camera Accessory, Game Accessory and Home Audio.

Solution Approach

Data Understanding & Preparation → EDA & Data Visualization → Feature Engineering → Model Building → Future Scope

Data Understanding & Preparation

- Read through all the datasets provided in different files.
- Handled nulls and assigned appropriate datatypes after reading different datasets.
- Order & Weather Datasets were filtered for appropriate date range.
- Imputed missing values.
- Dropped unnecessary columns.
- Outlier Treatment was done.
- Identified other data quality issues and handled them appropriately.



EDA & Data Visualization

- Filtered order dataset for product subcategories = CameraAccessory, HomeAudio, GamingAccessory.
- Performed univariate and bi-variate analysis on all columns of the ConsumerElectronics dataset.
- Filtered rows on basis of some business rules:
    1. Removed the columns having high \N values.
    2. Removed duplicates on basis of combination of order_id and order_item_id.
    3. Set GMV = MRP* Units for the records having GMV higher than MRP * Units.
    4. Dropped records for GMV = 0.
    5. Dropped records for MRP = 0.
    6. Set all the SLA columns with values < 0 to 0.
    7. Outlier Treatment : Took values only within range of (mean – 3* std, mean + 3 * std) for all variables.
- Read ad-spend data from other file.
- Converted monthly ad-spends to daily ad-spends and eventually into weekly ad-spends. Converted given ad-spends in crores to units.
- Read weather dataset. Joined 2 files for different years and filtered for appropriate date range.
- Handled data quality issues of weather dataset and did EDA for all columns.
- Read special sales dataset & NPS dataset from the given excel file.

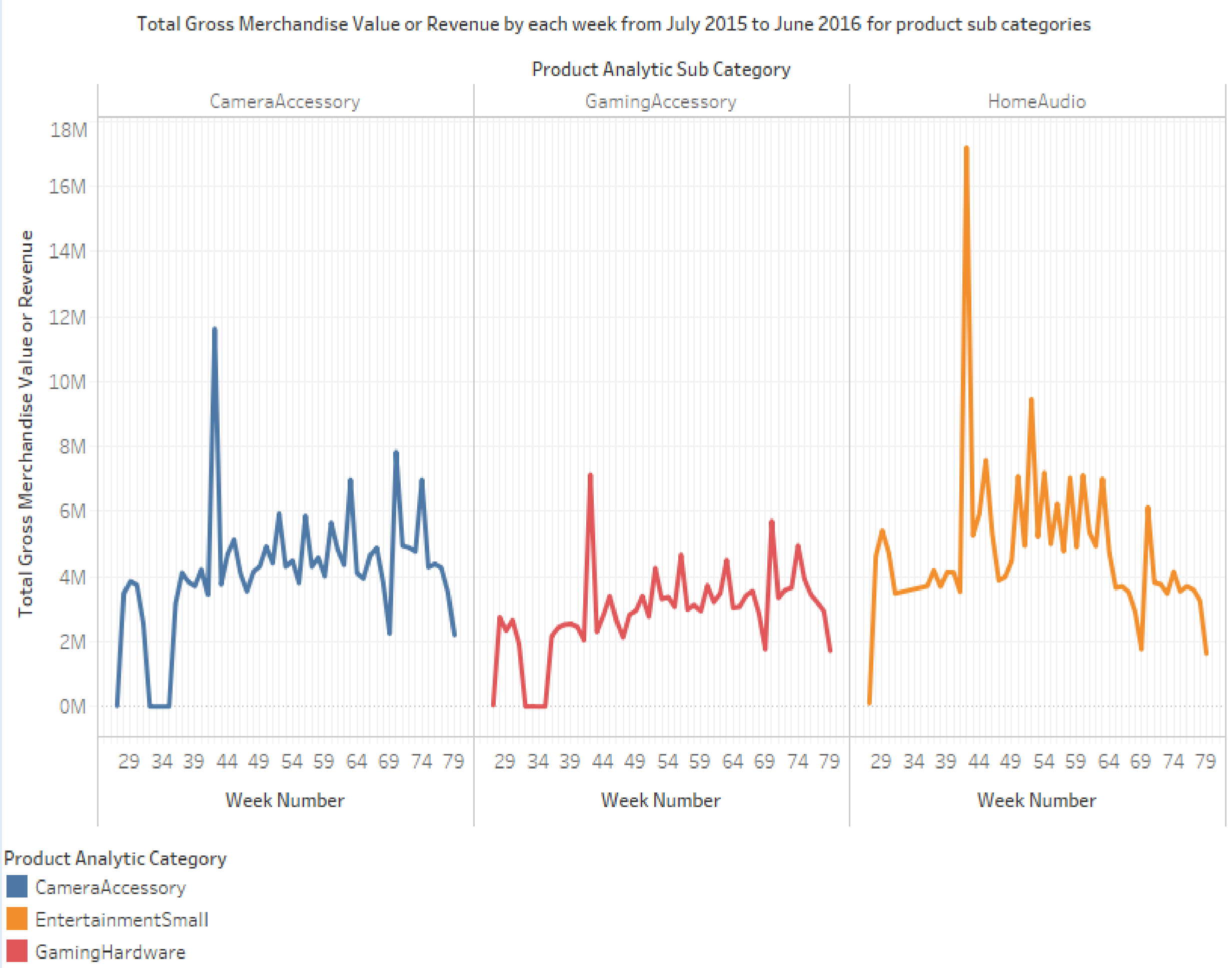
Feature Engineering

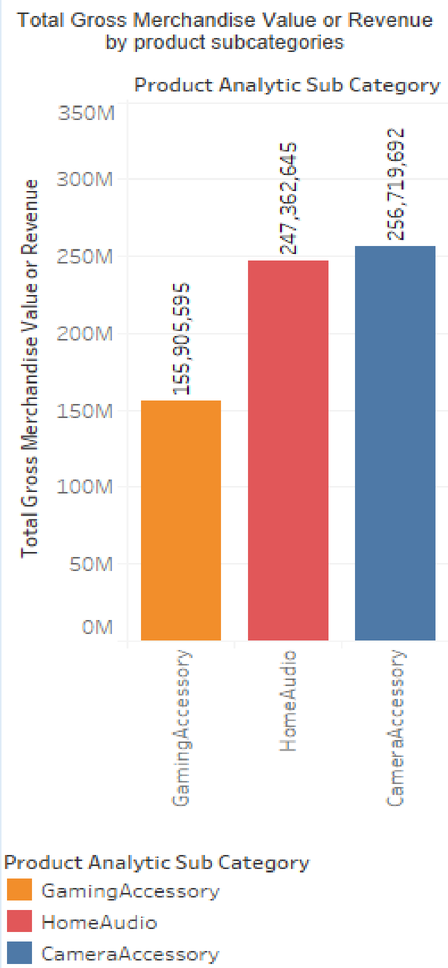
- Created weeks from order date in the order dataset.
- Created weeks in weather dataset and ad-spend datasets and aggregated both to weekly level.
- We took the week number to be continuous. Therefore dataset consist of week 27-79.
- Aggregated all the datasets over weeks and joined on weeks. We divided the datasets in 3 for different sub-categories.
- Handled multi-collinearity before creating the models.

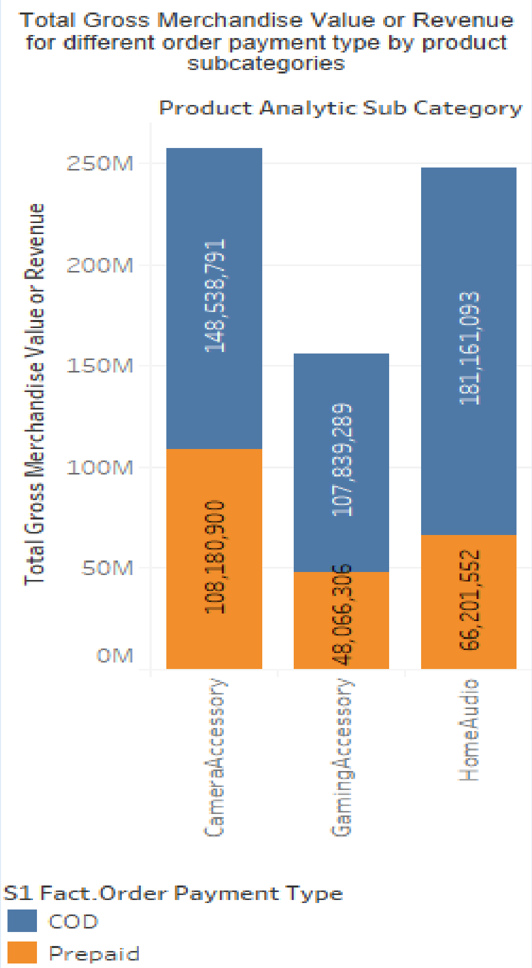
Basic KPIs
Dummy Variable for Prepaid/ COD and took fraction for prepaid to COD Ratio
List Price = GMV/ Units
Discount = (MRP – List Price)/ MRP
Flags: pay_day, special_sales
Weekly NPS, Weekly Ad-Spends, Weekly Mean Temperature, Weekly Mean Rainfall

Advanced KPIs
Ad-Stock Variables for each Marketing Category
Lag Variables for 1 and 2 weeks: GMV, Product MRP
Moving Averages for 2 and 3 weeks: List Price, Discount

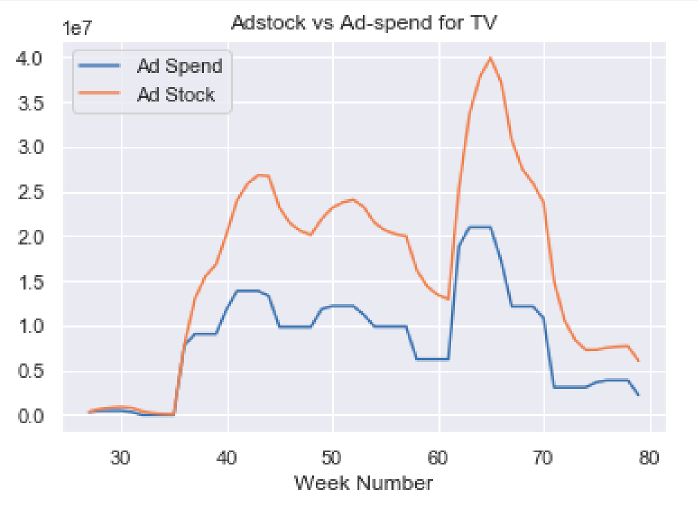
EDA & Data Visualization

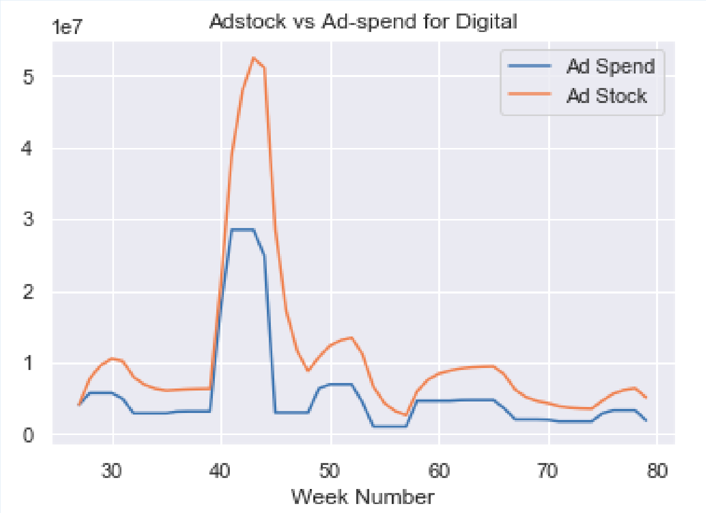


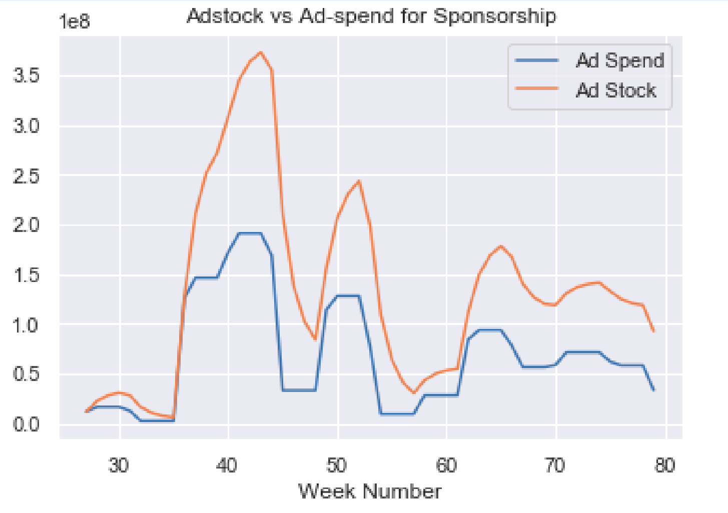


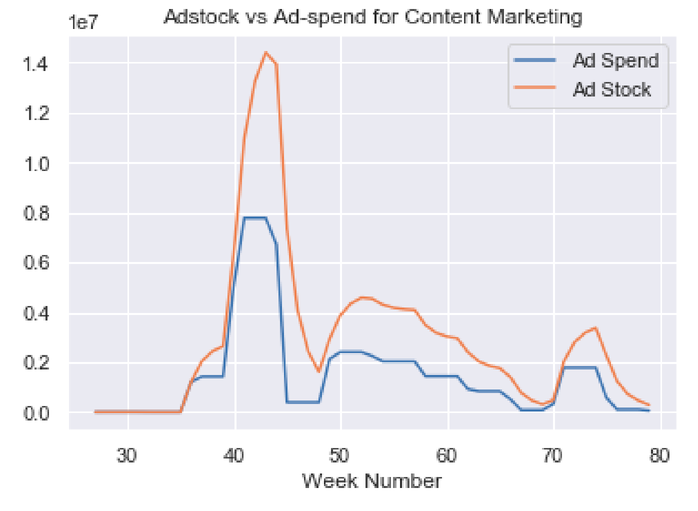


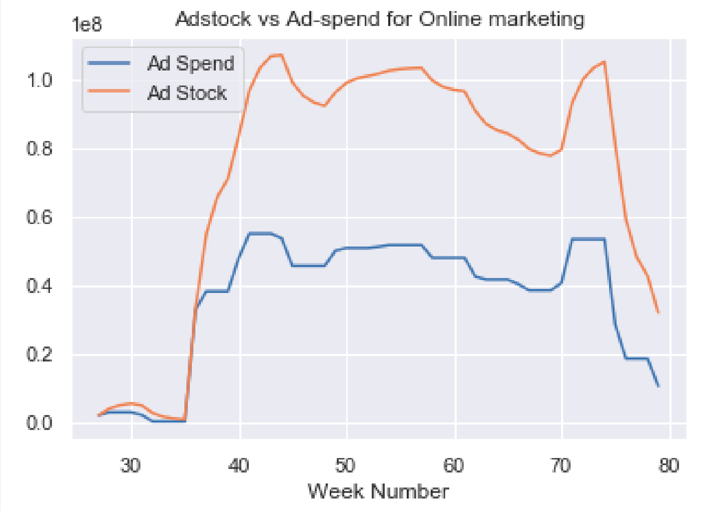
EDA & Data Visualization

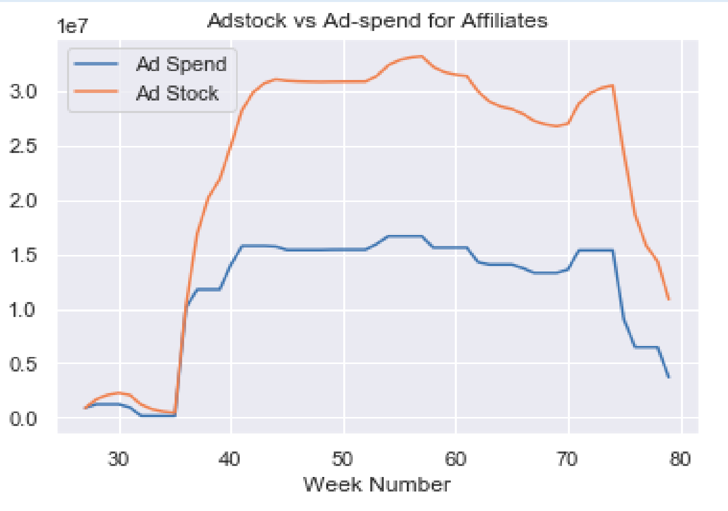




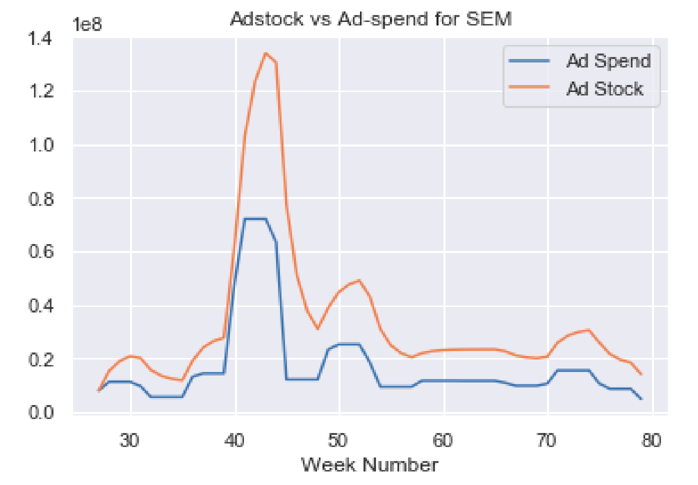


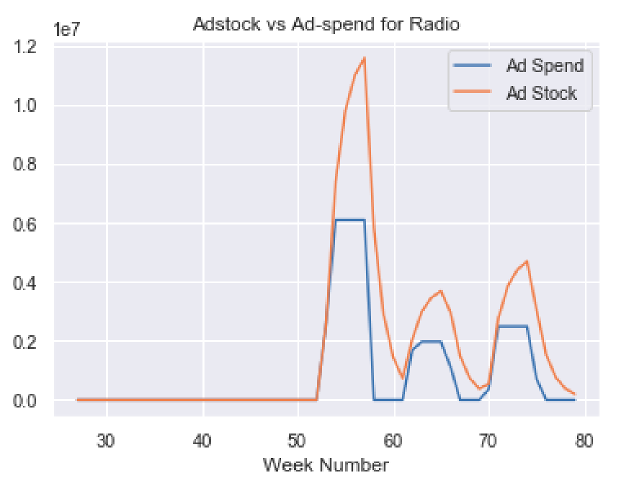


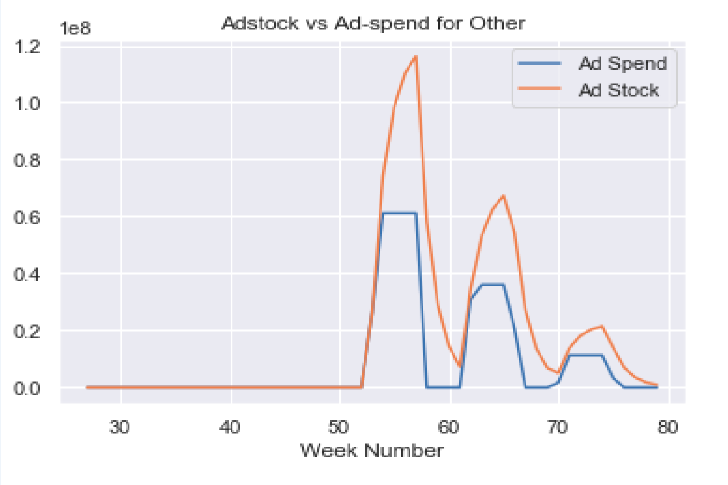




EDA & Data Visualization







- Plotted ad-spends v/s ad-stocks over week for all categories. These are all in crores or 10 crores (scale is indicated on top left).
- Some of the patterns looks quite similar. These hints at strong collinearity among few categories which was validated during model creation.

Model Building

- Created 1 dataset for each product sub category after creating basic features, data preparation, EDA and aggregating all the datasets to a weekly level and joined them over weeks. Final datasets had 1 aggregated row for each week.
- Used RFE after feature scaling and handling multi-collinearity for creating each models for 3 categories.
- Created Linear Model and Multiplicative models using basic KPIs.
- Created advanced KPIs on the aggregated data for each categories.
- Created Distributed Lag, Koyck and Distributed Lag Multiplicative models after feature scaling, handling multi-collinearity and doing the RFE.
- Computed training R-Square, Adjusted R-Square and Test RMSE and MSE values after keeping features having low p-value and VIF.

Model Building: CameraAccessory

- R-Square and Adjusted R-Square shown are based training data performance.
- RMSE and MSE are test data scores.
- Linear Model have decent performance but Koyck Model has slightly better performance.
- Distributed Lag model have lowest adj. R-Square and high RMSE and MSE.
- Distributive Lag Multiplicative model have high adjusted R-Square but also have high RMSE and MSE values.
- Multiplicative model has decent performance but none of ad-spend feature.
- Hence we would use Koyck Model.

Model	Variables	R-Square Adj. R-Square	RMSE	MSE
Linear	List Price + Discount + Online Marketing	0.501 0.468	0.565	0.319
Multiplicative	Prepaid Frac + Digital + SLA	0.573 0.545	0.112	0.012
Distributed Lag	SLA + Discount + SLA_lag_2 + List_Price_ma_3	0.45 0.40	0.616	.380
Koyck	List Price + Discount + Online Marketing_adstock	0.520 0.488	0.090	0.008
Distributed Lag Multiplicative	List Price + Discount_ma_2 + SLA	0.795 0.781	0.735	0.540

Model Building: GamingAccessory

- R-Square and Adjusted R-Square shown are based training data performance.
- RMSE and MSE are test data scores.
- Linear Model have decent performance but Distributed Lag Model has slightly better performance.
- Multiplicative model have lowest adj. R-Square and high RMSE and MSE.
- Koyck Model have good performance with only 3 features.
- In this case we would prefer using Distributed Lag Model as it has highest Adj. R-Square and very low RMSE and MSE values. Also, it has marketing ad-stock feature.

Model	Variables	R-Square Adj. R-Square	RMSE	MSE
Linear	List Price + Discount + Online Marketing	0.620 0.595	0.849	0.721
Multiplicative	Prepaid Frac + Product MRP + Product Procurement SLA	0.413 0.375	0.556	0.309
Distributed Lag	SLA_lag_1 + Discount + List_Price_ma_2 + Product MRP_lag_1 + Online Marketing_adstock	0.650 0.610	0.129	0.016
Koyck	List Price + Discount + Online Marketing_adstock	0.613 0.588	0.149	0.122
Distributed Lag Multiplicative	List Price + List Price_ma_2 + Prepaid Frac + Product_MRP_lag_2	0.590 0.553	0.351	0.123

Model Building: HomeAudio

- R-Square and Adjusted R-Square shown are based training data performance.
- RMSE and MSE are test data scores.
- Linear Model have decent performance but Distributed Lag and Koyck Model have slightly better performance.
- Multiplicative model have lowest adj. R-Square and high RMSE and MSE.
- Distributed Lag Multiplicative model has high Adjusted R-Square but RMSE and MSE are also higher than Koyck. This model don't have any marketing ad-stock variable as well.
- Hence, in this case we would prefer Koyck Model as it has high Adjusted R-Square and lowest RMSE and MSE.

Model	Variables	R-Square Adj. R-Square	RMSE	MSE
Linear	Content Marketing + Discount + Online Marketing + Product Procurement SLA	0.665 0.632	0.293	0.086
Multiplicative	List Price + SLA + Discount	0.572 0.542	0.281	0.079
Distributed Lag	Content Marketing_adstock + SLA + List_Price_ma_2 + GMV_lag_1 + Discount	0.682 0.642	0.053	0.002
Koyck	Content Marketing_adstock + SLA + GMV_lag_1 + Discount + Product Procurment SLA	0.700 0.662	0.050	0.002
Distributed Lag Multiplicative	SLA + Product MRP_lag_1 + List_Price_ma_2 + Discount_ma_2	0.715 0.687	0.258	0.066

Conclusions

- Camera Accessory:
  - Koyck models is giving List Price, Discount and Online Marketing Ad-stock as predictor variables all having positive coefficients.
  - Online Marketing spending should be increased. Reduce spending in other marketing channels and those funds can be utilized in giving better discounts.
  - By giving better discounts products list price should also change.

Koyck Model for Camera Accessory

	coef
const	-0.1815
List_Price	0.3428
discount	0.3476
Online marketing_adstock	0.2399

- Gaming Accessory:
  - Distributed Lag Model is giving Discount, Online Marketing Ad-stock, Product MRP Lag for 1 week, SLA lag for 1 week and List Price Moving Average for 2 weeks as predictor features with MRP Lag and SLA Lag having negative coefficient and rest having positive coefficeints.
  - Online Marketing spending should be increased. Reduce spending in other marketing channels and those funds can be utilized in giving better discounts.
  - More focus should be there to reduce the SLAs and MRP of the products and thus creating lower List Prices.

Distrbuted Lag Model for Gaming Accessory

	coef
const	-0.0127
discount	0.5958
Online marketing_adstock	0.3387
product_mrp_lag_1	-0.4189
sla_lag_1	-0.5913
List_Price_ma_2	0.7366

Conclusions

- Home Audio

- Koyck models is giving SLA, Product Procurement SLA, Discount, Content Marketing Ad-stock and GMV Lag as predictor variables.
- Content Marketing spending should be increased. Reduce spending in other marketing channels and those funds can be utilized in giving better discounts and improving SLA and product Procurement SLAs.

Koyck Model for Home Audio

	coef
const	0.1902
sla	0.2246
product_procurement_sla	-0.3385
discount	0.3697
Content Marketing_adstock	0.1861
gmv_lag_1	-0.2264

- Most of the Marketing Channels have high co-relation among themselves.
- More funds should be allocated for Digital Marketing and Content Marketing spending which would generate better ad-stocks and boost sales.
- Funding for other channels should be reduced and that budget should be allocated towards offering discounts and providing better SLAs (both SLA and Product Procurement SLA).
- List Prices of the products and its perception (through moving average and lags) also plays significant role in their sales. So, products should be listed at competitive prices.