MARKET MIX MODELLING

FOR ELECKART

(*AN ECOMMERCE COMPANY*)

ECOMMERCE CAPSTONE PROJECT

UPGRAD + IIITB PGDDS C4 (OCTOBER 2017)

SANTANU DEY

TRIDIB CHOWDHURY

MEDHAVI SHRUTI

ANKIT SINGH

CONTENTS

BUSINESS UNDERSTANDING & OBJECTIVE

<u>Business Understanding</u>:

- *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.
- They also offered big-ticket promotions.
- They are about to create a marketing budget for the next year which includes spending on commercials, online campaigns, and pricing & promotion strategies.

<u>Objective</u>:

- Develop a market mix model to observe the actual impact of different marketing variables over the last year.
- Recommend the optimal budget allocation for different marketing levers for the next year.
- Analysis and recommendations should be performed for 3 product categories:
  - *Gaming Accessory*
  - *Camera Accessory*
  - *Home Audio*

DATA UNDERSTANDING

Data available is for the period July 2015 to June 2016 and consists of:

- Order Level data for the above period (***ConsumerElectronics.csv***)
- Monthly advertising spends (Ad Stocks) on different marketing channels (***Final_adstock.csv***)
- Monthly NPS (or brand perception) data (***nps.csv***)
- Holiday List for the above period (***Media data and other information.xlsx***)
- Product Category/Sub-category Details (***Product Details.docx***)

Note: The two files ***Final_adstock.csv*** and ***nps.csv*** (present in submission zip file) are not part of the dataset shared as part of the Course Problem Statement from UpGrad. These two files have been shared separately by the Capstone Mentor assigned by UpGrad.

ANALYSIS APPROACH

- Load the required libraries and the datasets
- ***Data Cleaning*** *and Data Sanity Checks*:
  - Data should be between July 2015 and June 2016
  - Data is converted to weekly level. Jan 2016 should start with week 54, and not week 1
  - Remove free products (MRP equal to 0); Remove NA values
  - If GMV is 0, set it to 1; GMV should not be more than MRP * Units
  - Divide the dataset into 3 product categories – GamingAccessory, CameraAccessory, and HomeAudio
- ***Data Preparation*** *for each of the 3 product categories*:
  - KPI Engineering
  - K-means Clustering: Divide products into three price/market category clusters – mass market, medium market, and premium market
  - Merge NPS, Holiday Data, and AdStock Data with order-level data by week of the year
  - Create Moving Average (2, 3, 4-point) Variables, corresponding Incremental Lifts, and Lag data (1, 2, 3 weeks)
- ***Exploratory Data Analysis*** for each of the 3 product categories for GMV vs each of the 7 AdStock categories

ANALYSIS APPROACH (CONTD.)

- ***Model Building***: Build the following models for each of the 3 product categories with GMV as the target variable:
  - *Basic Linear* Model
  - *Multiplicative* Model
  - *Koyck* Model
  - *Distributed Lag* Model
  - *Multiplicative + Distributed Lag* Model
- Perform ***Cross Validation*** and ***Elasticity Analysis*** for each model
- ***Choose the best model(s)*** for each product category based on Adjusted R-squared and Cross Validation (SSE) values
- Provide marketing ***recommendations*** based on the models chosen

CHALLENGES FACED

- ***Which variables to include or exclude from the first model in each case***? An instinctive approach was taken
  - *Linear Model*: week_year, lag variables, price category variables were removed
  - *Multiplicative Model*: week_year, moving averages, lag variables, holiday frequency were removed
  - *Koyck Model*: week_year, lag variables, price category variables were removed
  - *Distributed Lag Model*: week_year, price category variables were removed
  - *Multiplicative + Distributed Lag Model*: week_year, moving averages, lag variables, holiday frequency, and holiday lag variables were removed
- ***Scaling performed after taking logarithms for multiplicative model(s) was causing errors in the first model. Why***? No scaling was performed in such cases
- ***Which variable to remove for each subsequent model in each case***? A combined approach based on high multicollinearity and low significance was taken for the same
- ***Which model(s) to choose for each product category as the best***? The model(s) having high adjusted R-squared and low Cross Validation (SSE) values was/were chosen

KPI ENGINEERING

The following KPIs (Key Performance Indicators) have been engineered from the available data

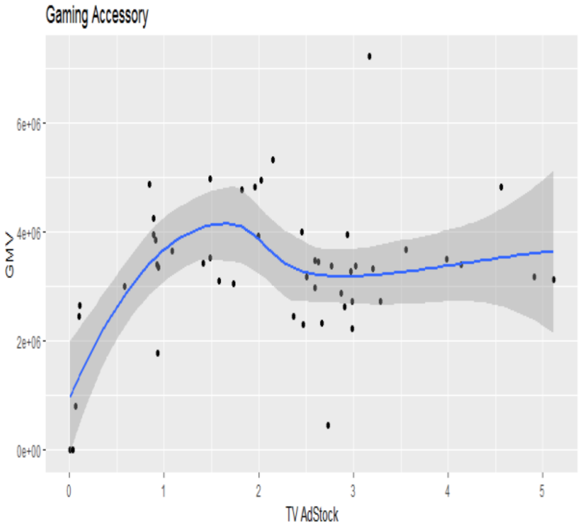
- ***List Price*** for all the products
- ***Promotional Offer*** for all the products
- ***Market Price Category*** (*mass, medium, premium*) based on *MRP, Units*, and *List Price*
- ***Payment Model Indicator***
- ***Percentage Online Order***
- 2, 3, and 4 point **Moving Averages** and corresponding **Incremental Lifts** for *list price* and *promotional offers*
- 1, 2, and 3 weeks' **Lag** data for *list price*, *promotional offer*, *NPS* and *holiday frequency*

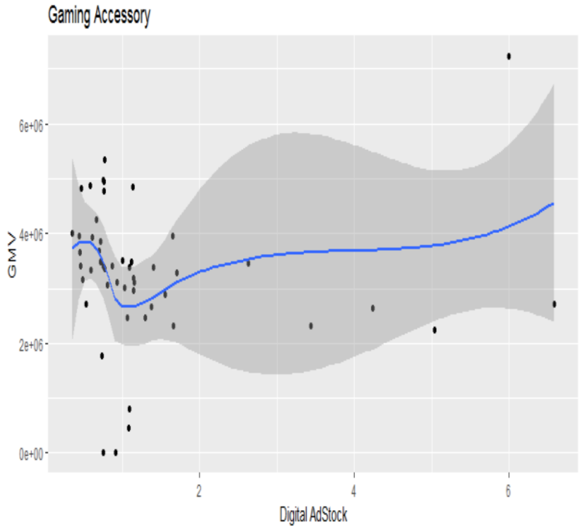
EXPLORATORY DATA ANALYSIS

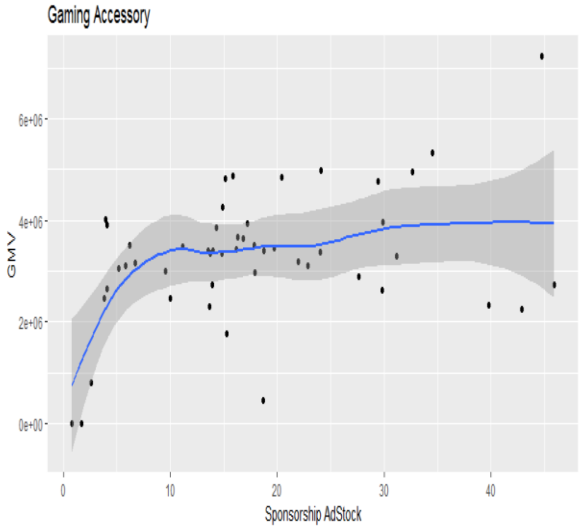
The Exploratory Data Analysis (EDA) has been performed for each Product Category (*Gaming Accessory, Camera Accessory* and *Home Audio*) for the following categories of Marketing AdStocks:

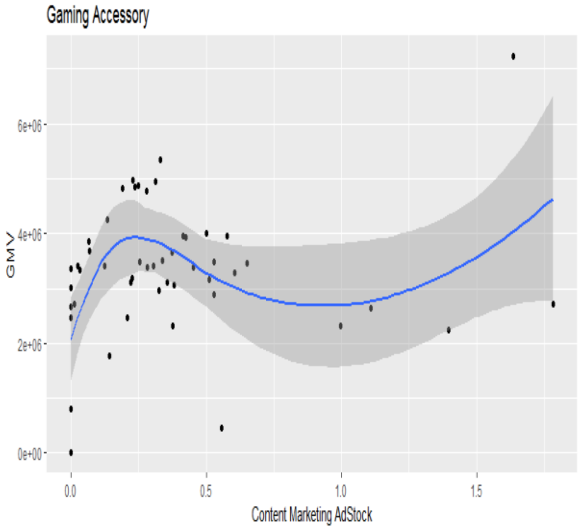
- TV
- Digital
- Sponsorship
- Content Marketing
- Online Marketing
- Affiliates
- SEM (Search Engine Marketing)

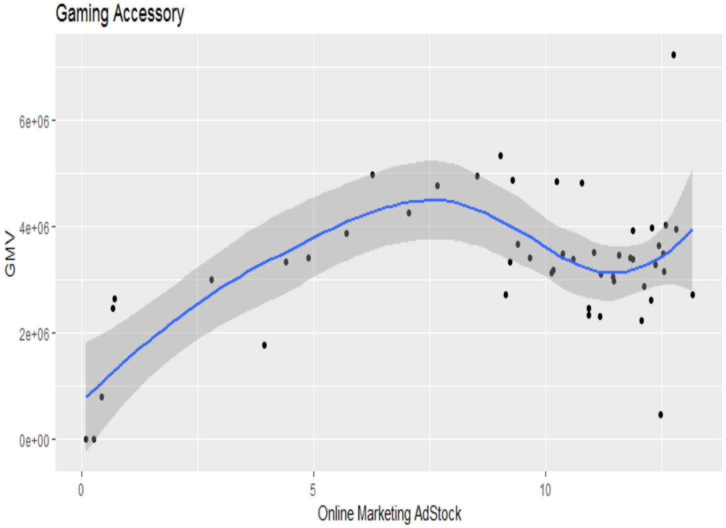
EDA - Gaming Accessory

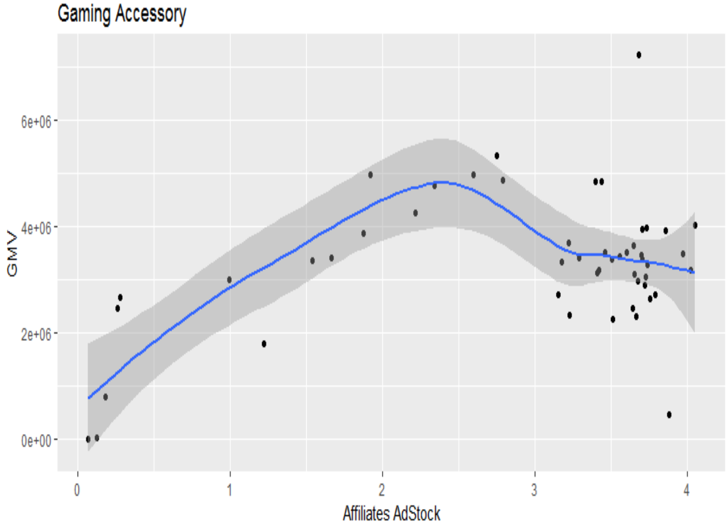


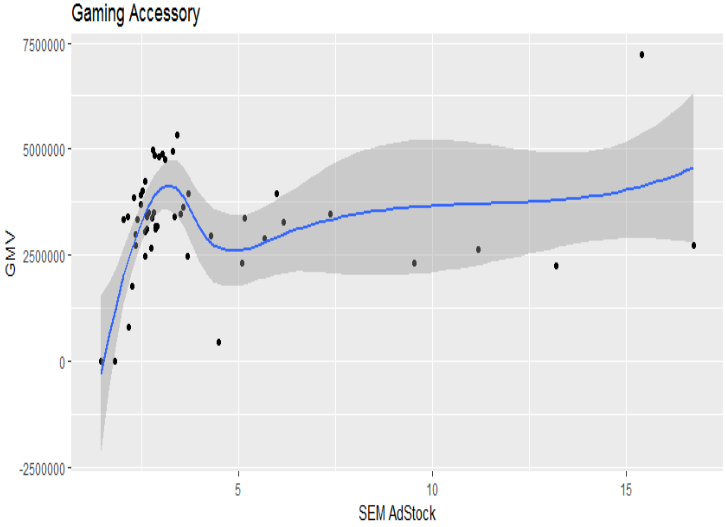








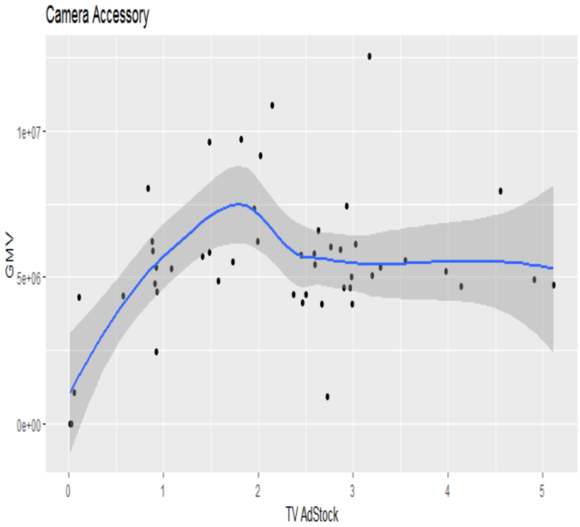


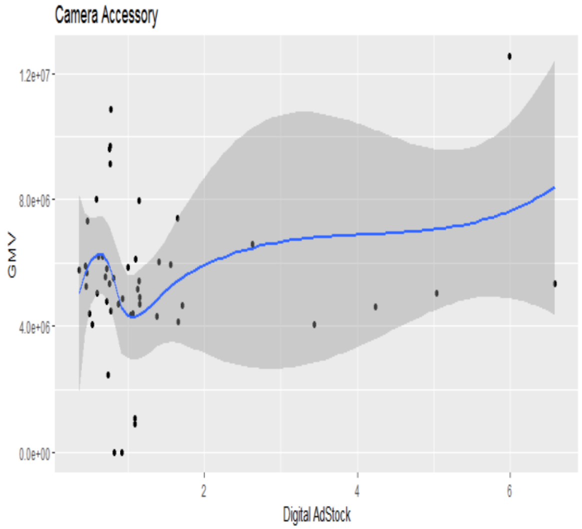


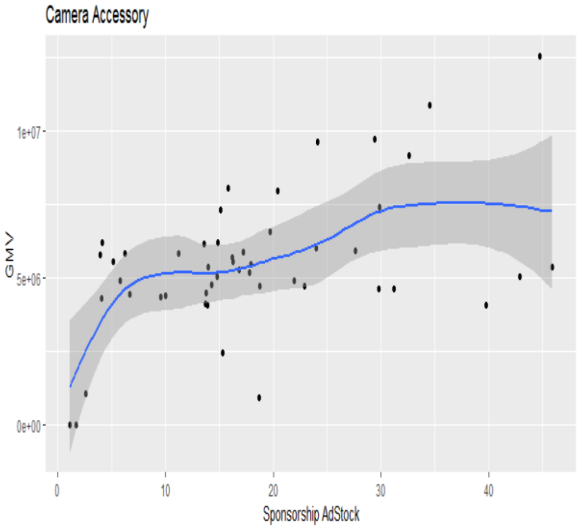
EDA OBSERVATIONS
- *GAMING ACCESSORY*

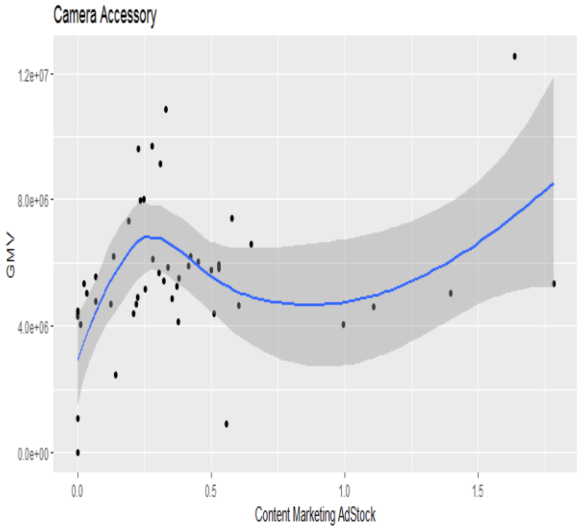
Marketing Channels	Impact on GMV
TV AdStock	GMV peaks at around 30% of TV AdStock
Digital AdStock	GMV reaches an initial peak at pretty low AdStock, then tapers and rises again to reach overall peak for maximum AdStock
Sponsorship AdStock	GMV reaches an initial peal at pretty low AdStock, and keeps on rising at a very low rate and plateaus at around 75% of AdStock
Content Marketing AdStock	GMV reaches an initial peak at pretty low AdStock, then tapers and rises again to reach overall peak for maximum AdStock
Online Marketing AdStock	GMV peaks at around 66% of Online Marketing AdStock
Affiliates AdStock	GMV peaks at around 60% of Affiliates AdStock
SEM AdStock	GMV peaks at a pretty low value of SEM AdStock

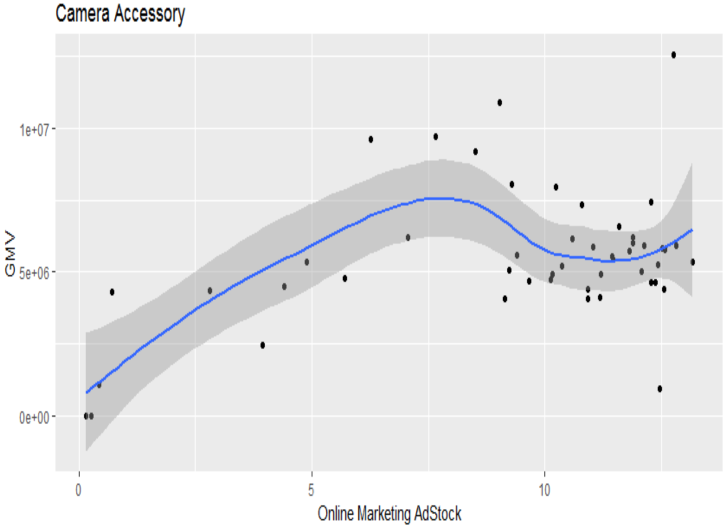
EDA - Camera Accessory

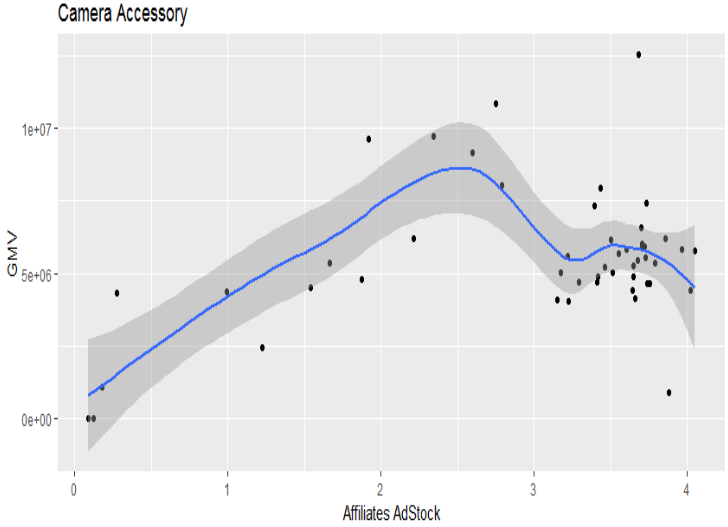


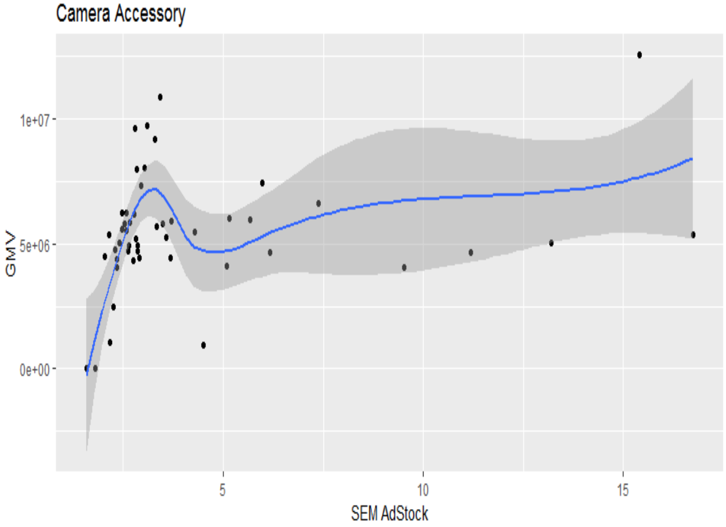








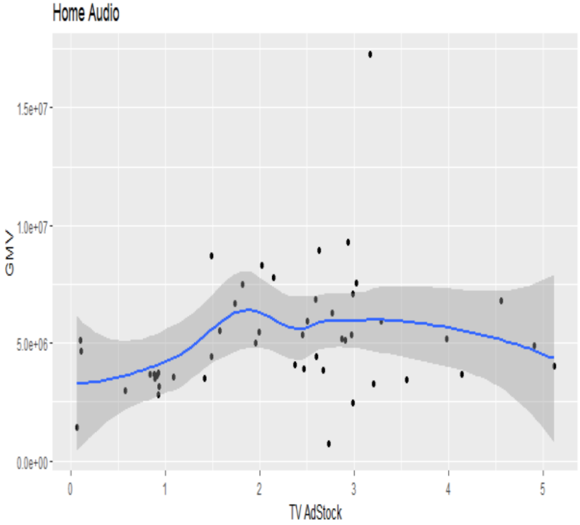


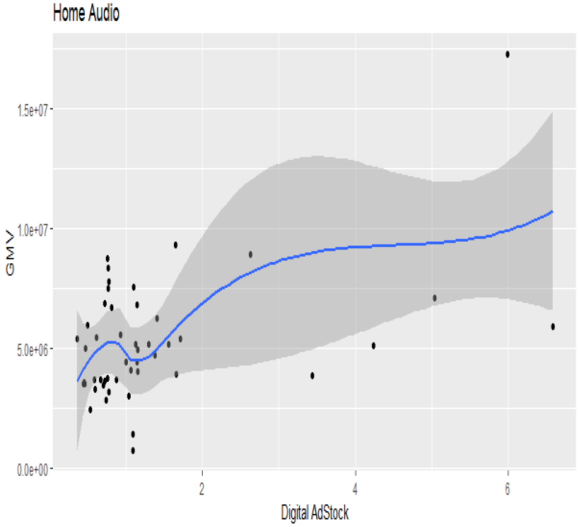


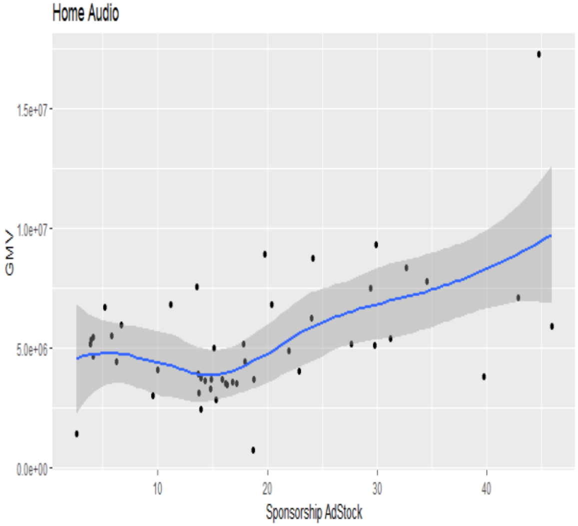
EDA OBSERVATIONS
- *CAMERA ACCESSORY*

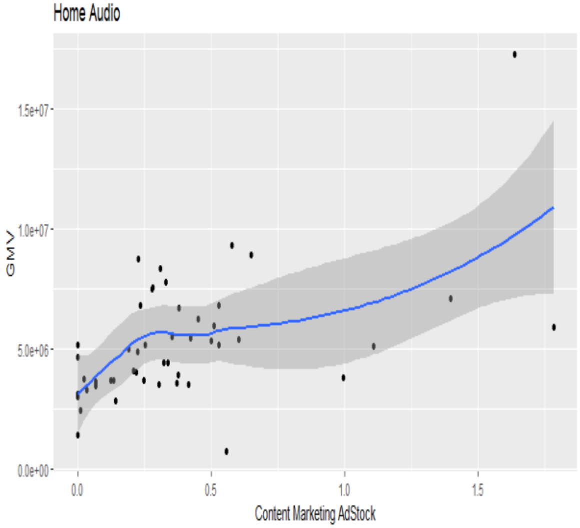
Marketing Channels	Impact on GMV
TV AdStock	GMV peaks at less than 40% of TV AdStock
Digital AdStock	GMV reaches an initial peak at pretty low AdStock, then tapers and rises again to reach overall peak for maximum AdStock
Sponsorship AdStock	GMV reaches an initial peal at pretty low AdStock, and keeps on rising at a very low rate and plateaus at around 75% of AdStock
Content Marketing AdStock	GMV reaches an initial peak at pretty low AdStock, then tapers and rises again to reach overall peak for maximum AdStock
Online Marketing AdStock	GMV peaks at around 75% of Online Marketing AdStock
Affiliates AdStock	GMV peaks at around 60% of Affiliates AdStock
SEM AdStock	GMV reaches an initial peal at pretty low AdStock, and keeps on rising at a very low rate to reach overall peak for maximum AdStock

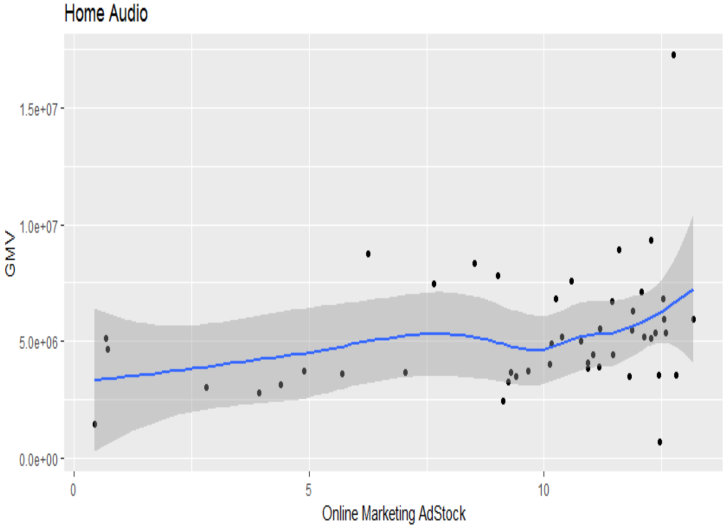
EDA - Home Audio

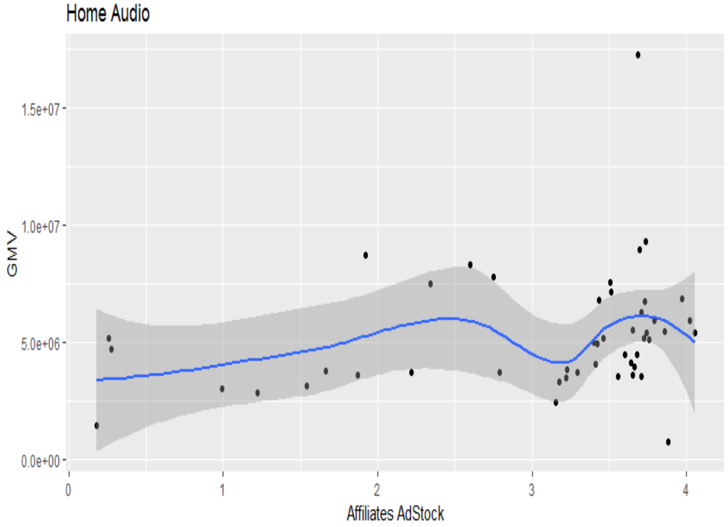


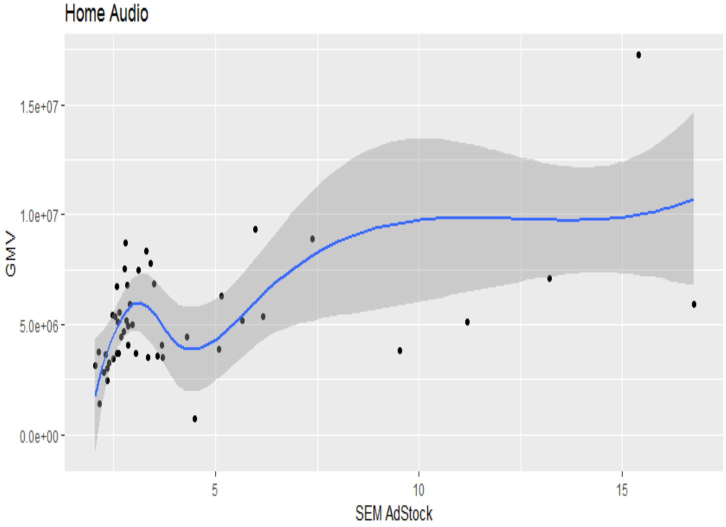












EDA OBSERVATIONS
- *HOME AUDIO*

Marketing Channels	Impact on GMV
TV AdStock	GMV peaks at less than 40% of TV AdStock
Digital AdStock	GMV reaches an initial peak at pretty low AdStock, then tapers and rises again to reach overall peak for maximum AdStock
Sponsorship AdStock	GMV reaches an initial peal at pretty low AdStock, then tapers and rises again to reach overall peak for maximum AdStock
Content Marketing AdStock	GMV reaches an initial peak at pretty low AdStock, and keeps on rising at a very low rate to reach overall peak for maximum AdStock
Online Marketing AdStock	GMV reaches an initial peal at around 58% AdStock, then tapers and rises again to reach overall peak for maximum AdStock
Affiliates AdStock	GMV peaks at around 60% of Affiliates AdStock
SEM AdStock	GMV reaches an initial peal at pretty low AdStock, then tapers and keeps on rising at a very low rate to reach overall peak for maximum AdStock

MODEL BUILDING & SELECTION

MODEL BUILDING & SELECTION
- *GAMING ACCESSORY*

Model	Variables	Adjusted R-squared	Cross Validation (SSE)
Basic Linear Model	NPS + inc_LP_MA2 + inc_PO_MA2	0.385	0.712
Multiplicative Model	ContentMarketing + OnlineMarketing + Affiliates	0.791	0.546
Koyck Model	Procurement_SLA + Sponsorship + ContentMarketing + SEM + inc_LP_MA2 + inc_PO_MA2	0.508	1.039
Distributed Lag Model	MRP + list_price.2 + holiday_freq.1 + gmv.1	0.359	1.206
Multiplicative + Distributed Lag Model	OnlineMarketing + Affiliates + promotional_offer.3	0.827	0.470

- *Koyck* Model and *Distributed Lag* Model are *ruled out* due to high SSE values
- *Basic Linear* Model is *ruled out* due to low adjusted R-squared value
- We select the *Multiplicative* Model and the *combination of Multiplicative and Distributed Lag* models due to high adjusted R-squared and low SSE values

MODEL BUILDING & SELECTION
- *CAMERA ACCESSORY*

Model	Variables	Adjusted R-squared	Cross Validation (SSE)
Basic Linear Model	Sponsorship + inc_LP_MA3 + inc_PO_MA1	0.551	0.512
Multiplicative Model	Units + per_order + OnlineMarketing	0.878	0.506
Koyck Model	Sponsorship + inc_LP_MA3 + inc_PO_MA1	0.548	0.524
Distributed Lag Model	Sponsorship + inc_LP_MA2 + list_price.3 + promotional_offer.1	0.416	0.765
Multiplicative + Distributed Lag Model	Procurement_SLA + SEM + gmv.1	0.876	1.557

- Combination of *Multiplicative and Distributed Lag* Model is *ruled out* due to high SSE value
- *Distributed Lag* Model is *ruled out* due to low adjusted R-squared value
- We select the *Basic Linear* Model, the *Multiplicative* Model and the *Koyck* Model due to high adjusted R-squared and low SSE values

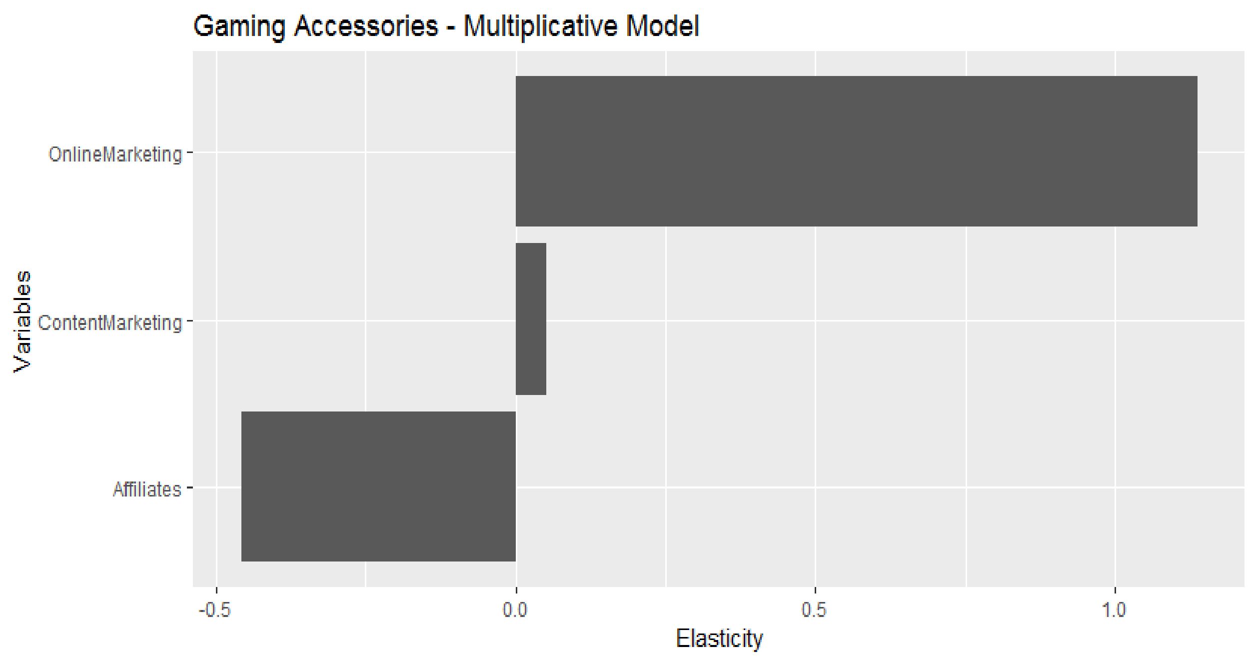
MODEL BUILDING & SELECTION
- *HOME AUDIO*

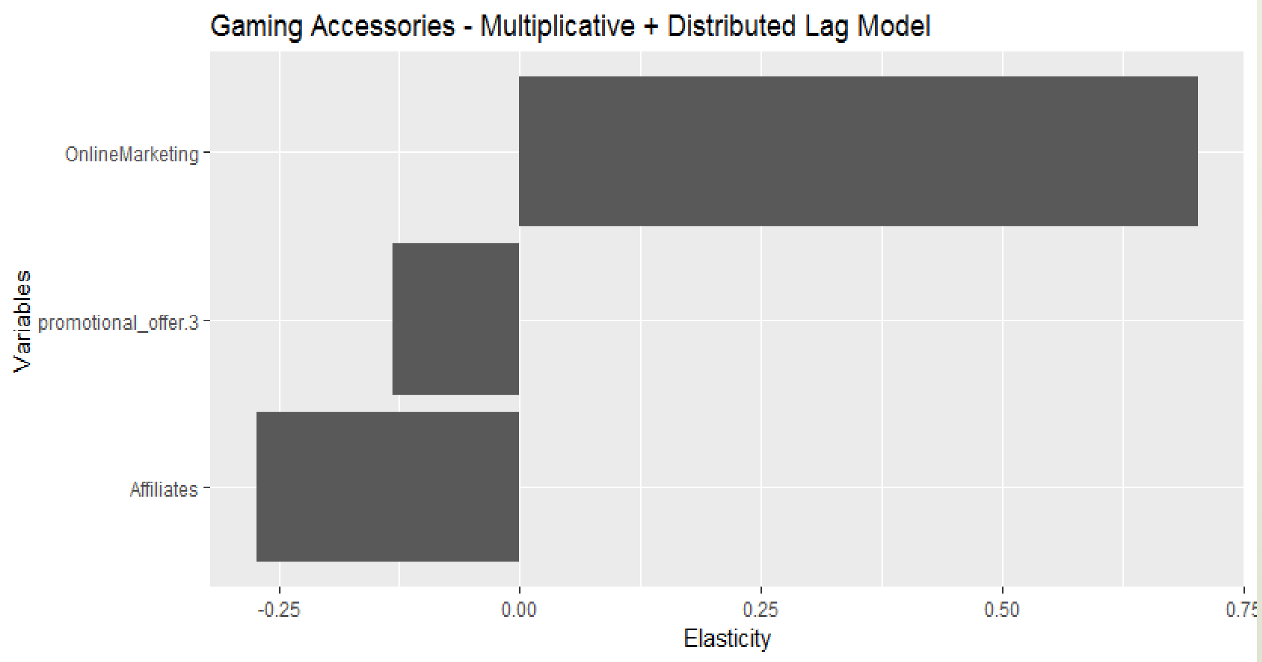
Model	Variables	Adjusted R-squared	Cross Validation (SSE)
Basic Linear Model	ContentMarketing + inc_PO_MA3	0.359	0.821
Multiplicative Model	MRP + TV + OnlineMarketing + Affiliates	0.248	0.224
Koyck Model	ContentMarketing + inc_PO_MA3	0.362	0.806
Distributed Lag Model	inc_LP_MA1 + inc_LP_MA2 + inc_LP_MA3 + inc_PO_MA3 + list_price.2 + list_price.3	0.5	0.815
Multiplicative + Distributed Lag Model	promotional_offer + NPS + TV + OnlineMarketing + Affiliates + NPS.1	0.324	0.182

- *Basic Linear*, *Koyck*, and *Distributed Lag* models are *ruled out* due to high SSE values.
- *Multiplicative* Model is *ruled out* due to low adjusted R-squared value
- We select the *combination of Multiplicative and Distributed Lag* Model due to high adjusted R-squared and low SSE values

RECOMMENDATIONS

Recommendations *Gaming Accessory*





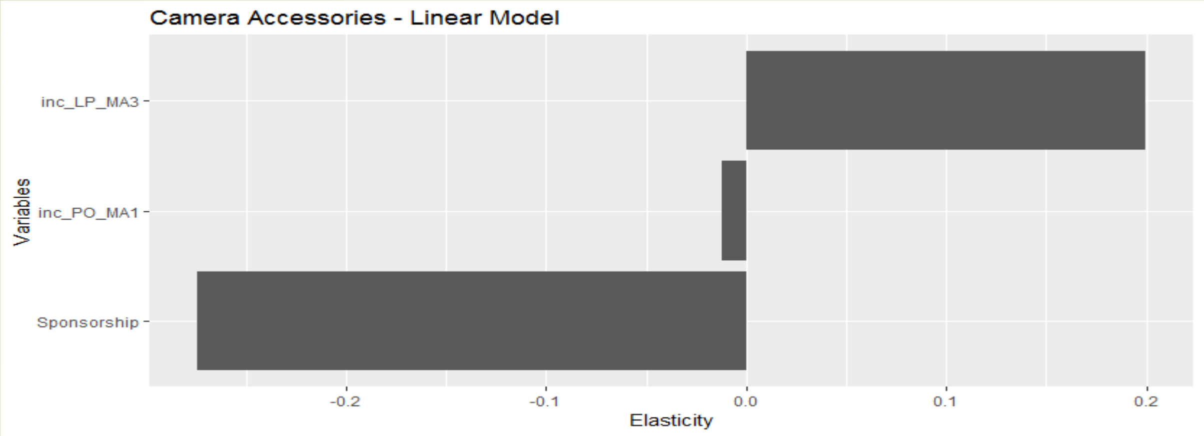
Based on the *Multiplicative* model, ElecKart should:

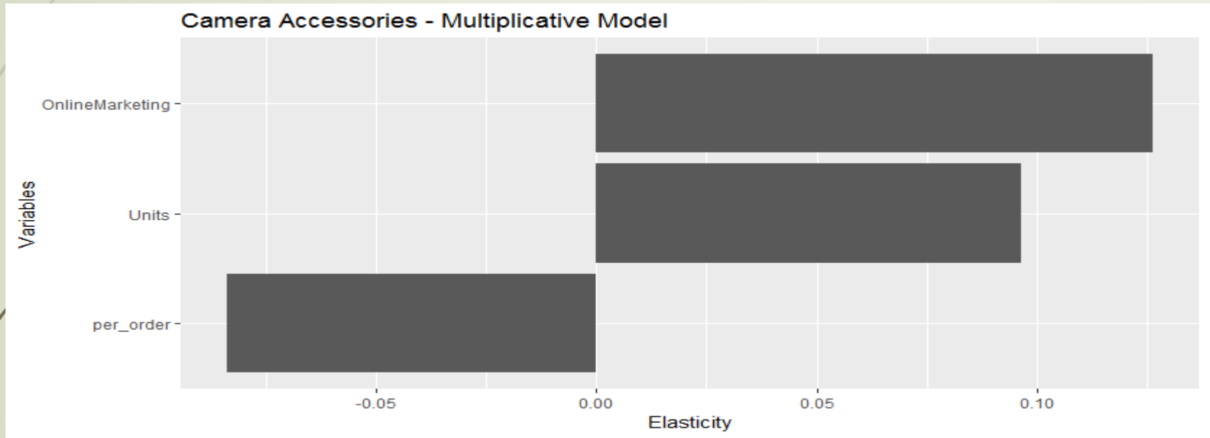
- **increase** its spending on ***Online Marketing*** significantly
- **maintain** or increase by a small measure, its spending on ***Content Marketing***
- **reduce** its spending on ***Affiliates***

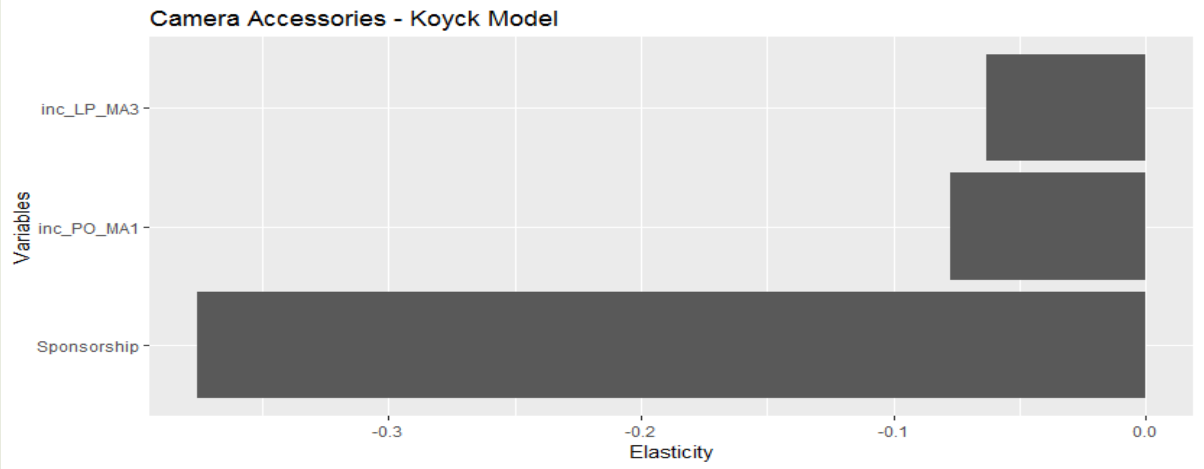
Based on the *Multiplicative + Distributed Lag* model, ElecKart should:

- **increase** its spending on ***Online Marketing*** significantly
- **reduce** its spending on ***Promotional Offers*** and ***Affiliates***

Recommendations *Camera Accessory*







Based on the *Basic Linear* model, ElecKart should:

- **reduce** its ***discounts*** w.r.t. previous week
- **reduce** its spending on ***Sponsorship***

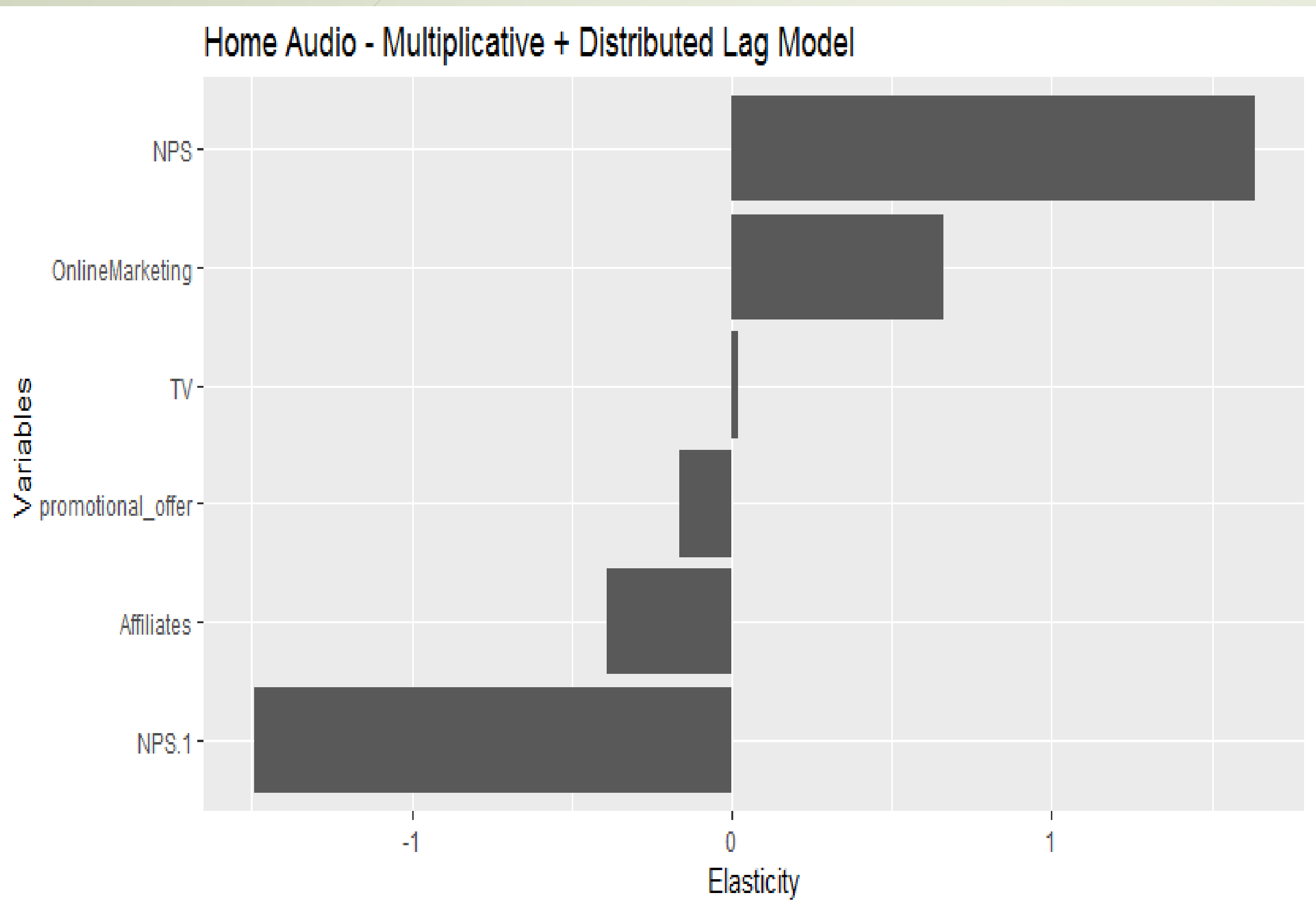
Based on the *Multiplicative* model, ElecKart should:

- **increase** its spending on ***Online Marketing*** significantly
- **increase** the ***number of units***
- **reduce** its preference for ***online payments***

Based on the *Koyck* model, ElecKart should:

- **reduce** its ***discounts*** w.r.t. previous week
- **reduce** its spending on ***Sponsorship***

Recommendations *Home Audio*



Based on the *Multiplicative + Distributed Lag* model, ElecKart should:

- **increase** its focus on improving the ***monthly NPS score***
- **increase** its spending on ***Online Marketing*** significantly
- **maintain** or slightly increase its spending on ***TV advertisements***
- **reduce** its spending on ***Promotional Offers*** and ***Affiliates***

THE END