Eleckart Market Mix Modelling

Group Members:

- Sanjeev Ojha
- Rakhee Roy
- Ribhav Shridhar
- Reena Poddar

Business Understanding and Objective

Business Understanding:

ElecKart is an e-commerce firm specializing in electronic products. They had spent a significant amount of money in marketing over last one year. 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. They want to reallocate their budget optimally across different marketing levers to improve the revenue response.

Business Objective:

We need to develop a market mix model for three product sub-categories - camera accessory, home audio and gaming accessory and to observe the actual impact of different marketing variables over the last year (July 2015 to June 2016). Also, we have to recommend the optimal budget allocation for different marketing levers for the next year.

Problem Solving Approach

Steps Involved:

- Business Understanding
- Data Understanding
- Data Cleaning
- Data Preparation
- Derived KPIs
- EDA
- Divide data into three different datasets for each three product sub-categories - camera accessory, home audio and gaming accessory
- Model Building: -All 5 types of models are built for each of the 3 sub-categories including
    i. Linear Model
    ii. Kyock Model
    iii. Distributed Lag Model
    iv. MultiplicativeModel
    v. LagPlusMultiplicativeModel
- Model Evaluation for all 5 types of model built for each 3 sub categories
- Recommendation/Presentation of result

Data Understanding

Business has provided 5 datasets for Analysis which includes:

- Consumer Electronics.csv
- Product List
- Media Investment
- Special Sale Calendar
- Monthly NPS score

Consumer Electronics

This contains daily order level data between 19th May 2015 till 25th July 2016. 1648824records and20 variables are present.

Column Name	Significance
FSN ID	The unique identification of each SKU
Year	Year when order was placed
Month	Month when order was placed
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
DeliveryBdays	Dispatch delay from warehouse
Deliverycdays	Delivery delay from warehouse
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
pincode	pincode
Product_analytic-subcategory	Product subcategory
Product MRP	Maximum retail price of the product
Product procurement SLA	Time typically taken to procure the product

Product List

It contains details related to products.74 different products and its details are mentioned.

Column Name	Significance
Product Name	Name of the product
Frequency	Frequency of product sold
Percent	Percentage with respect to total sales

Media Investment

It contains detail Monthly spends in on various advertising channels /Media heads in INR Cr. Data is from period July 2015 to June 2016. Various Advertising channels are:

i. TV
ii. Digital
iii. Sponsorship
iv. Content Marketing
v. Online Marketing
vi. Affiliates
vii. SEM
viii. Radio
ix. Others

Monthly NPS score

It contains month wise customer satisfaction score in percentage. This may work as a proxy to 'voice of customer'. Data is from period July 2015 to June 2016.

Special Sale Calendar

It contains details of promotion type and date on which promotion was valid.

Data Cleaning

Steps involved are:

1. 4904NA values in GMV, customer id and pin code column were removed.
2. Converted negative values of customer id and pincode to absolute values.
3. Corrected the columns name to proper column names for FSNID,Payment_Mode
4. Negative values or product procurement sla was substituted with zero
5. Dropped columns product_analytic_super_category, product_analytic_category having single value.
6. Substituted negative value for deliverybdays , Deliverycdays with zero
7. Substituted \\N in deliverybdays, Deliverycdays with zero.
8. Removed rows having MRP as 0
9. All the NA values in media_investment file were substituted with zero.
10. Outliers were replaced with appropriate cutoff values.

Data Preparation/Derived KPIs

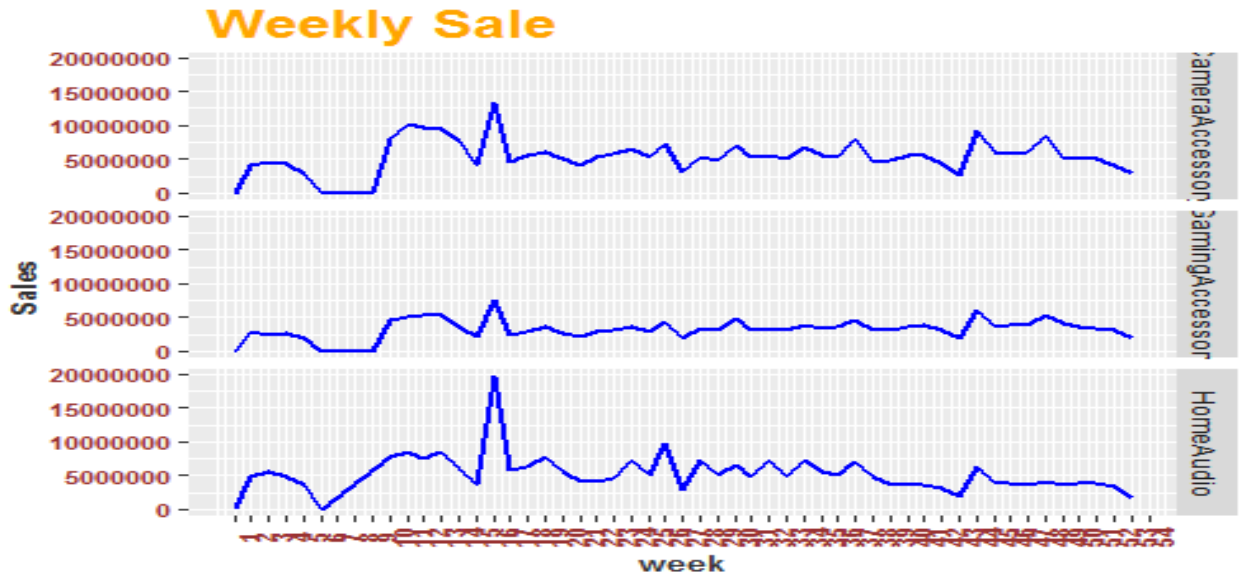
Steps involved are:

1. Filter the data between 1st July 2015 to 30th June 2016
2. Extract week number from order date .Make the week number in sequence starting from 1 to 54 so that first week of July 2015 represent “1” and vice versa.
3. Aggregated daily level data into weekly level data
4. Monthly level media investment data was converted to weekly level investment data and merged with order detail data
5. Promotional level data was converted to weekly level data and merged with main data frame.
6. Monthly level NPS score was converted to weekly level data and merged.

7. All different datasets were merged together to form one master data frame to build our model.
8. Restricted our analysis to 3 sub categories i.e. "CameraAccessory","GamingAccessory","HomeAudio" as per requirement.
9.
10. Following KPIs are derived: -
    i. List price =gmv/units sold
    ii. Discount=(MRP - list price)/MRP
    iii. Revenue Per visitor =gmv/number of unique customer
    iv. Delivery_satus = Delayed , 'On time' , 'Early' based on Delivery_on_time variable whereDelivery_on_time =sla-(deliverybdays+Deliverycdays+product_procurement_sla).
    If Delivery_on_time =0 then On-time
    If Delivery_on_time >0 Early
    If Delivery_on_time <0 Delayed
    v. Holiday_week– 'N' if non-promotional week, 'Y' if promotional week
    vi. Cod_cnt = sum of COD payment mode,Prepaid count = sum of Prpaid payment
    vii. Ontime_delivery_cnt, early_delivery_cnt, delayed_delivery_cnt
    viii. Average GMV
    ix. Average MRP
    x. Average no of orders
    xi. Adstock of each of the 'commercial spends' (Since we don't have TRP/ Impression information, we will assume spend is directly proportional to the impression)
    xii. Lag sale, i.e. the past value of GMV (useful for Koyck and distributed lag models) for last 3 weeks
    xiii. Lag promotion, i.e. the past value of promotions for last three weeks
    xiv. Lag price for last three weeks
    xv. Promotion type based on special sale calendar.

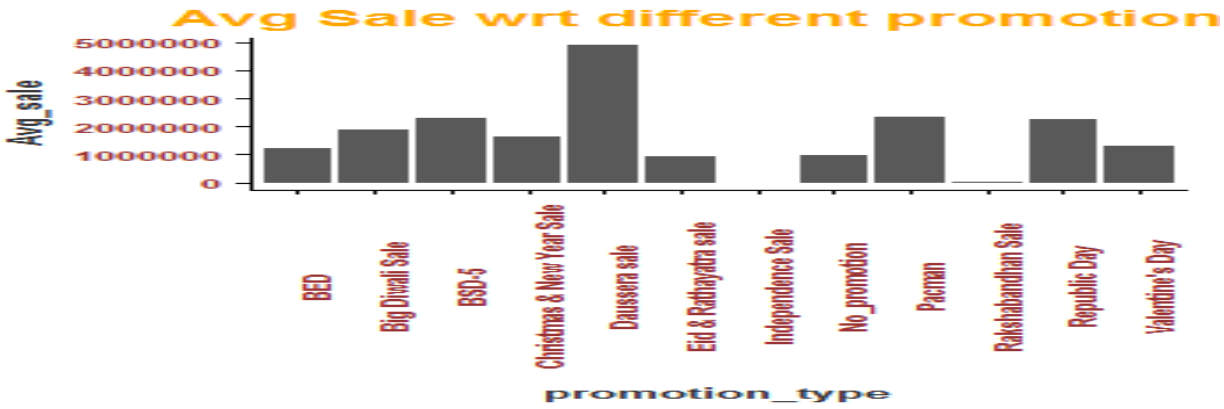
Exploratory Data Analysis

EDA-Weekly Sale



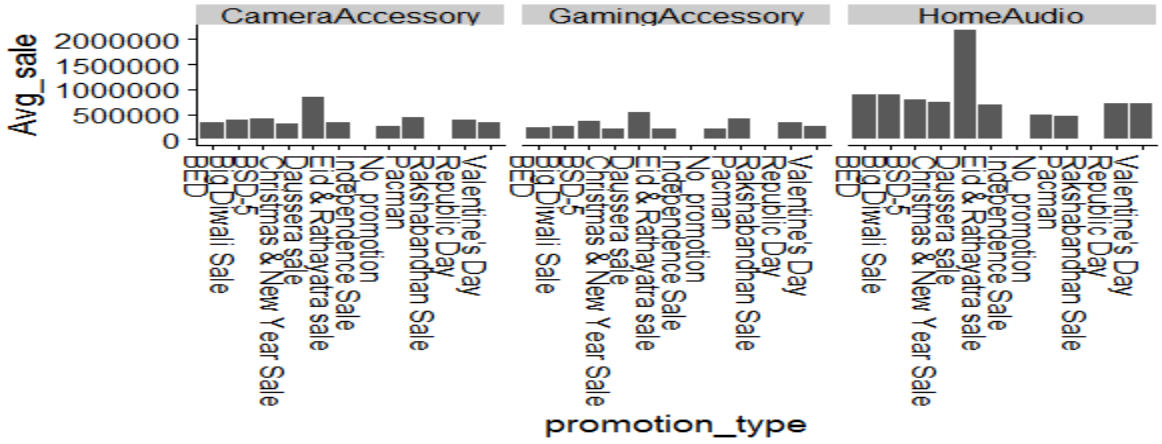
We can observe that there is spike during week 15-16 which dusera week. Across different promotional week we can see that sale is comparatively on higher side for all three subcategories.

EDA- Average Sale wrt different promotional and non promotional weeks



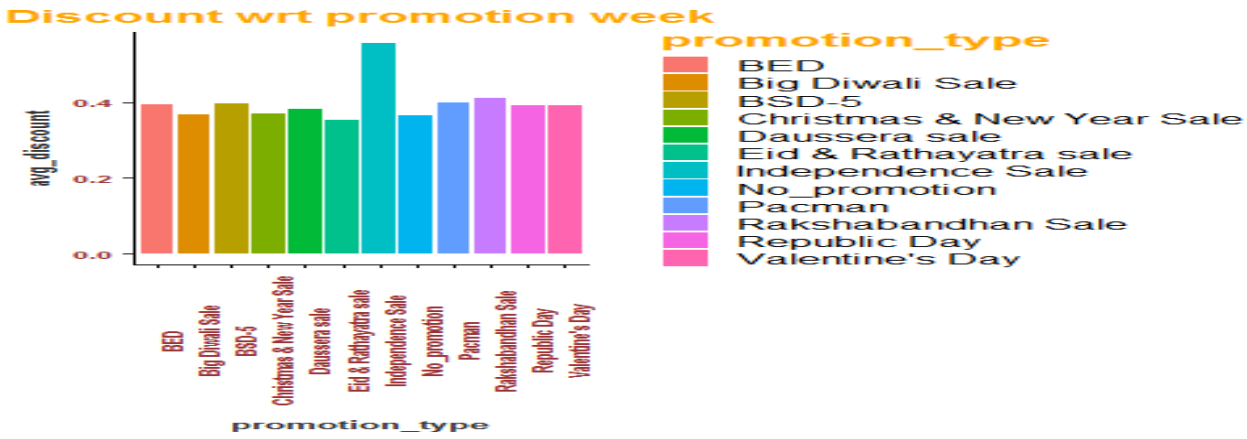
Average Sale during promotional week is far more better than Non-promotional week. This is indicative that promotional week is profitable.

EDA- Avg Sale w.r.t different promotional weeks for 3 sub categories

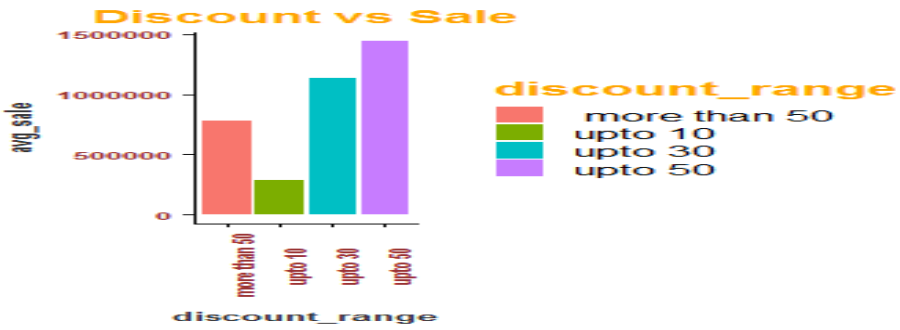


Among different promotional weeks, Dauseera Sale seems to be more profitable .This is common for all three sub categories.

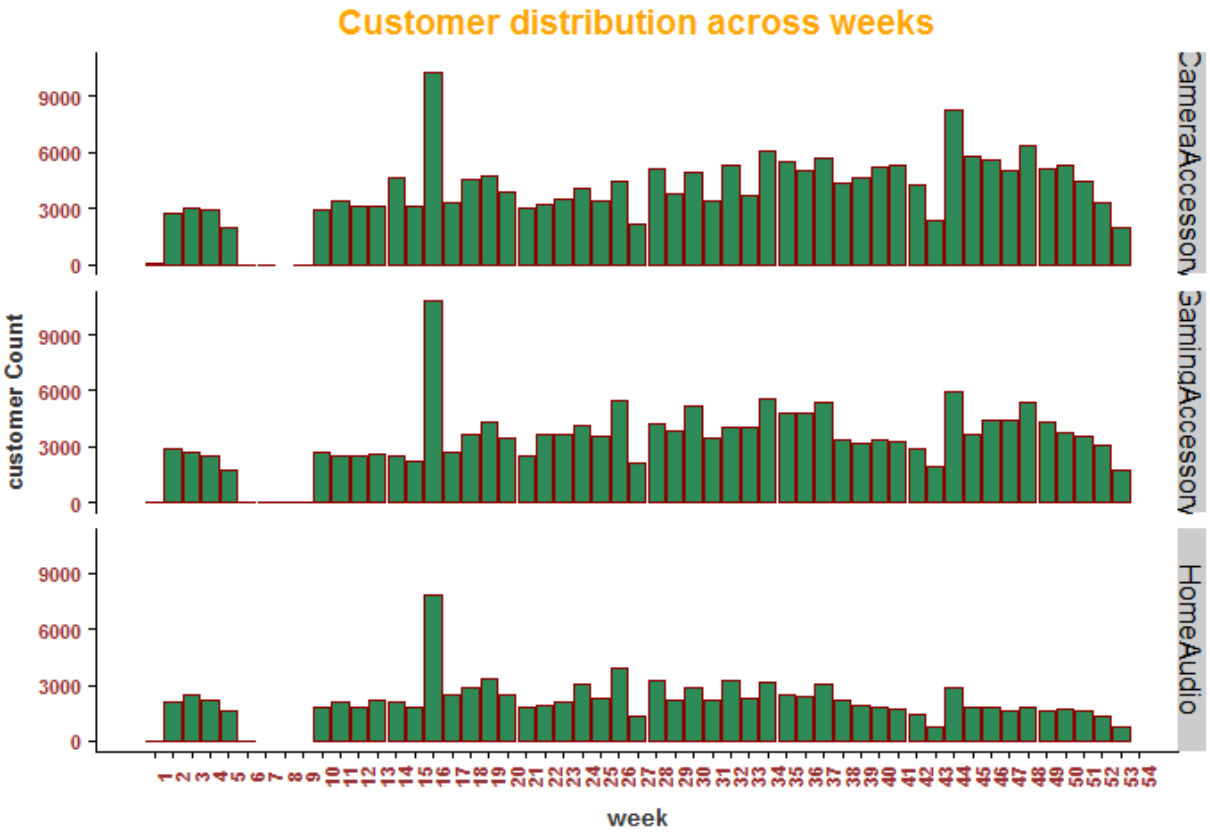
EDA- Avg discount wrt different promotional week



Company is providing heavy discount during promotional week especially during Independence Day sale.

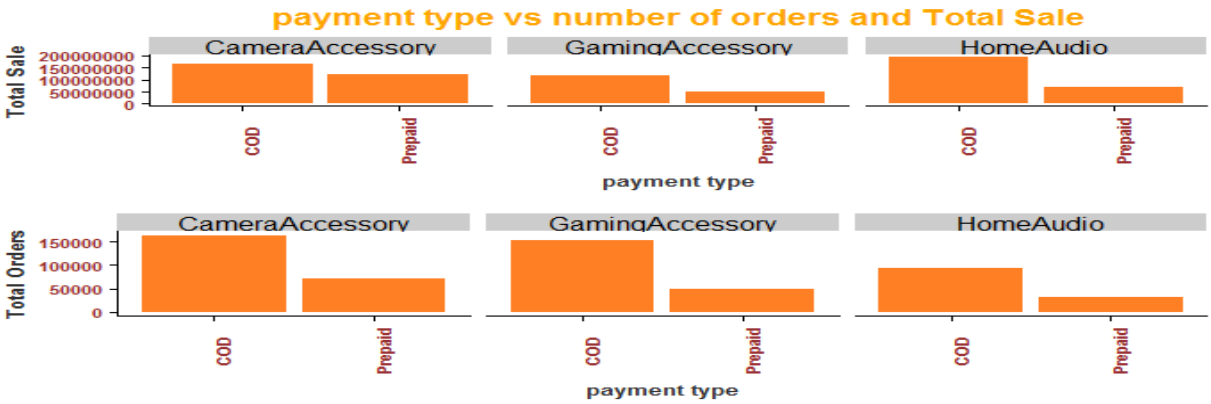


Discount in the range 30-50 percent seems to be profitable as sale increases .But Discount over 50 percent is not effective which should be taken care of.



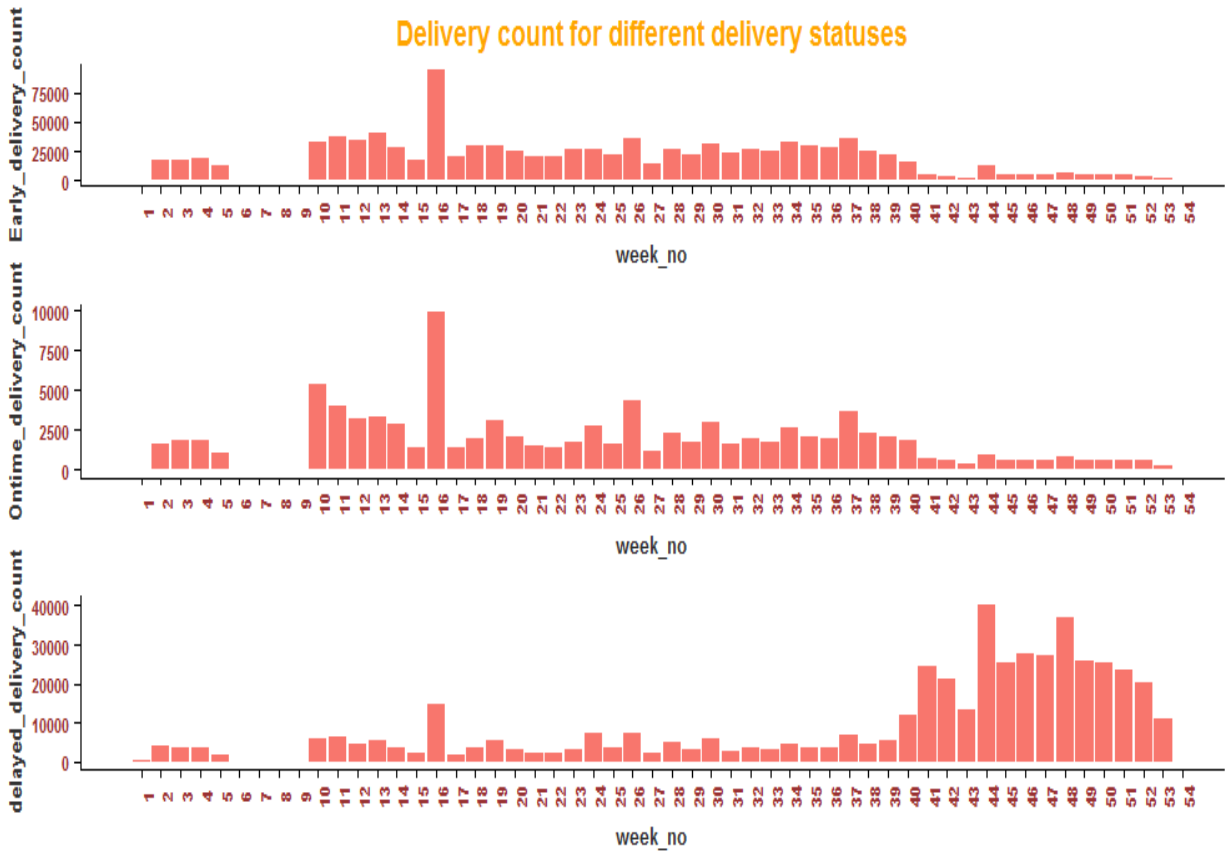
Customers visiting company's website is more on promotional weeks as compared to non-promotional week. Highest count is during Daussera Sale.

EDA -Payment Type vs Sale



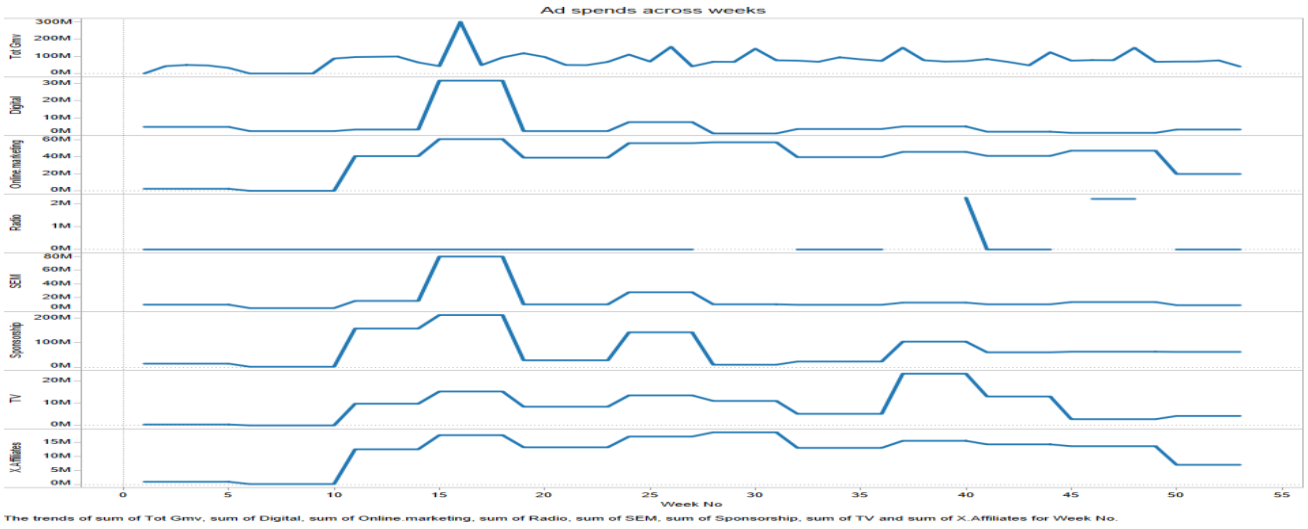
Cash on delivery mode is preferred most and it contributes largely to revenue. Since prepaid payment method is more beneficial to company from both operational and financial point of view, company should deep dive and see how this mode can be improved and liked by most of the customers. May be customers should be shown the benefits of prepaid mode through promotions. Extra offer should be given to attract customers.

EDA - Delivery count for different delivery statuses



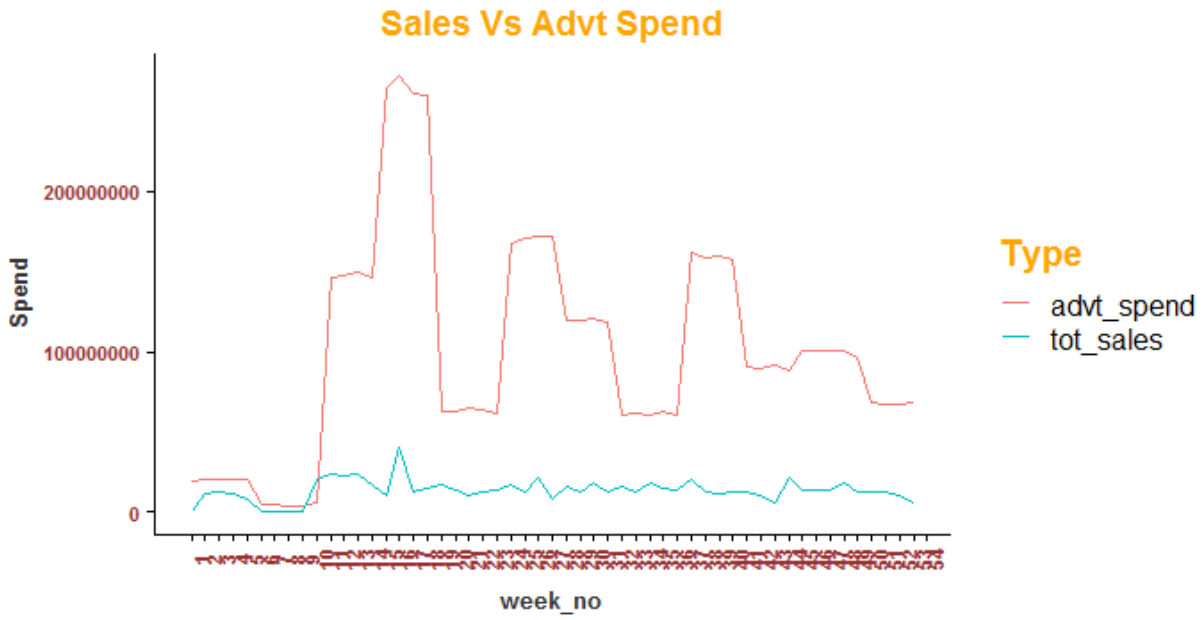
On time and Early deliveries are higher during promotional weeks . Delayed deliveries are seen during last few weeks (period of April till June).This needs to be catered by company.

EDA-Weekly Advertisement Spend



We can observe that weekly advertisement spend are higher during promotional weeks .

Sale Vs Advertisement Spend



Spend on promotion is much higher than revenue generated. It can also be seen during promotional week which should be taken care of.

RESULTS

CAMERA ACCESSORY			
Model	**Significant Variables**	**Adjusted R-Squared Value**	**MSE(cv =10)**
Linear Regression	product_analytic_vertical.xCameraMount+product_analytic_vertical.xTelescope + product_analytic_vertical.xStrap +product_analytic_vertical.xCameraBatteryGrip +product_analytic_vertical.xCameraAccessory + product_analytic_vertical.xCameraRemoteControl +product_analytic_vertical.xCameraFilmRolls + product_analytic_vertical.xFilter + product_analytic_vertical.xCameraEyeCup + product_analytic_vertical.xCameraBatteryCharger + product_analytic_vertical.xCameraMicrophone +product_analytic_vertical.xCameraHousing + product_analytic_vertical.xSoftbox +promotion_type.xRakshabandhan.Sale	0.83	0.236
Multiplicative Model	list_price + X.Affiliates + product_analytic_vertical.xCameraBattery +product_analytic_vertical.xCameraTripod +product_analytic_vertical.xFlash	0.734	0.994
Koyck Model	product_analytic_vertical.xLens+promotion_type.xDaussera.sale +promotion_type.xEid...Rathayatra.sale + X.Affiliates+Digital +list_price	0.731	0.232
Distributed Lag Model	product_analytic_vertical.xLens+promotion_type.xDaussera.sale + Sponsorship	0.72	0.275
Distributed Lag + Mulplicative	product_analytic_vertical.xLens + list_price+ Sponsorship + change_wrt_w1_gmv + change_wrt_w2_gmv +change_wrt_w3_gmv	0.863	0.692

Simple linear Model is best among all five models based on high Adjusted R-Squared value and low MSE value. MSE figures shown in the table above is based on 10 cross fold validations on training data set.

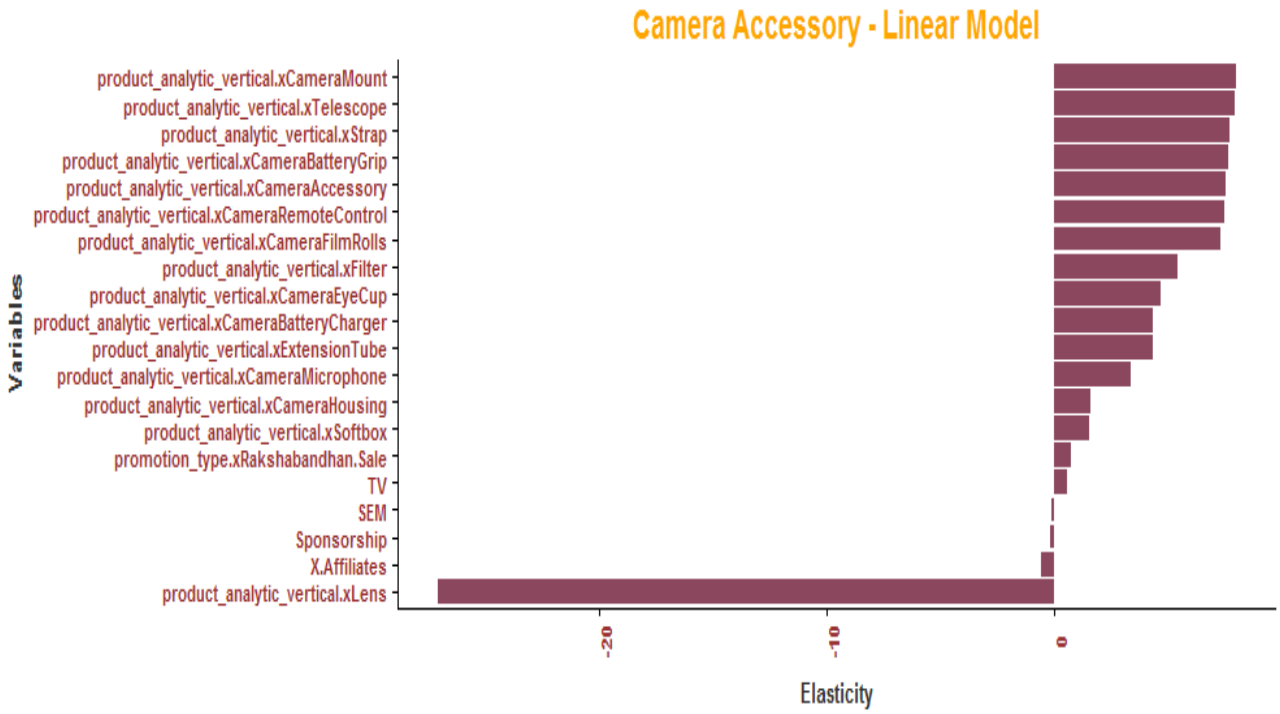
GAMING ACCESSORY			
Model	**Significant Variables**	**Adjusted R-Squared Value**	**MSE(cv= 10)**
Linear Regression	product_analytic_vertical.xGamePad +product_analytic_vertical.xGamingHeadset +product_analytic_vertical.xGamingMouse +product_analytic_vertical.xGamingKeyboard +X.Affiliates	0.6	0.265
Multiplicative Model	list_price + X.Affiliates + product_analytic_vertical.xGamePad +product_analytic_vertical.xGamingHeadset +product_analytic_vertical.xGamingMouse + product_analytic_vertical.xGamingKeyboard +product_analytic_vertical.xGamingAccessoryKit	0.6	0.792
Koyck Model	product_analytic_vertical.xGamingSpeaker +promotion_type.xChristmas...New.Year.Sale + TV +promotion_type.xIndependence.Sale+ promotion_type.xRakshabandhan.Sale + Digital	0.669	0.319
Distributed Lag Model	product_analytic_vertical.xGamePad +promotion_type.xDaussera.sale +product_analytic_vertical.xGamingHeadset +product_analytic_vertical.xGamingMouse +Sponsorship +Online.marketing	0.678	0.333
Distributed Lag + Mulplicative	product_analytic_vertical.xGamePad + product_analytic_vertical.xGamingHeadset +product_analytic_vertical.xGamingMouse + promotion_type.xDaussera.sale +product_analytic_vertical.xGamingKeyboard + promotion_type.xPacman + product_analytic_vertical.xGamingAccessoryKit + list_price + spon_ad_stock +discount_over_mrp + X.Affiliates + change_wrt_w1_gmv + change_wrt_w2_gmv + change_wrt_w3_gmv +price_change_wrt_w2	0.776	0.858

Distributed Lag Model is chosen best among all five models based on high Adjusted R-Squared value and low MSE value. MSE figures shown in the table above is based on 10 cross fold validations on training data set.

HOME AUDIO			
Model	**Significant Variables**	**Adjusted R-Squared Value**	**MSE(cv=10)**
Linear Regression	product_analytic_vertical.xHomeAudioSpeaker +product_analytic_vertical.xFMRadio +Sponsorship	0.81	0.163
Multiplicative Model	list_price + Online.marketing + product_analytic_vertical.xHomeAudioSpeaker +product_analytic_vertical.xFMRadio	0.656	0.903
Koyck Model	discount_over_mrp + product_analytic_vertical.xDock + product_analytic_vertical.xDockingStation + discount_over_mrp +Sponsorship +Digital	0.896	0.182
Distributed Lag Model	product_analytic_vertical.xHomeAudioSpeaker + promotion_type.xDaussera.sale	0.713	1.26
Distributed Lag + Mulplicative	product_analytic_vertical.xHomeAudioSpeaker + product_analytic_vertical.xFMRadio + promotion_type.xDaussera.sale + list_price + X.Affiliates + change_wrt_w2_gmv + change_wrt_w3_gmv + change_wrt_w1_gmv	0.772	0.752

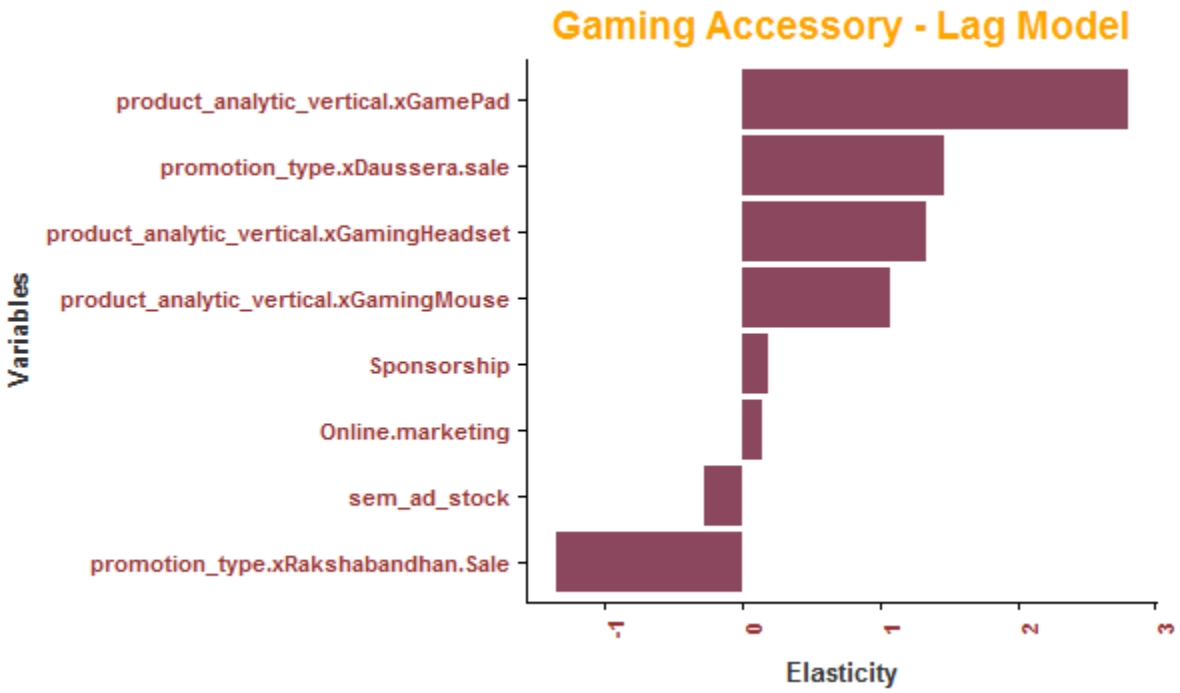
Koyck Model is chosen best among all five models based on high Adjusted R-Squared value and low MSE value. MSE figures shown in the table above is based on 10 cross fold validations on training data set.

Recommendation-Camera Accessories



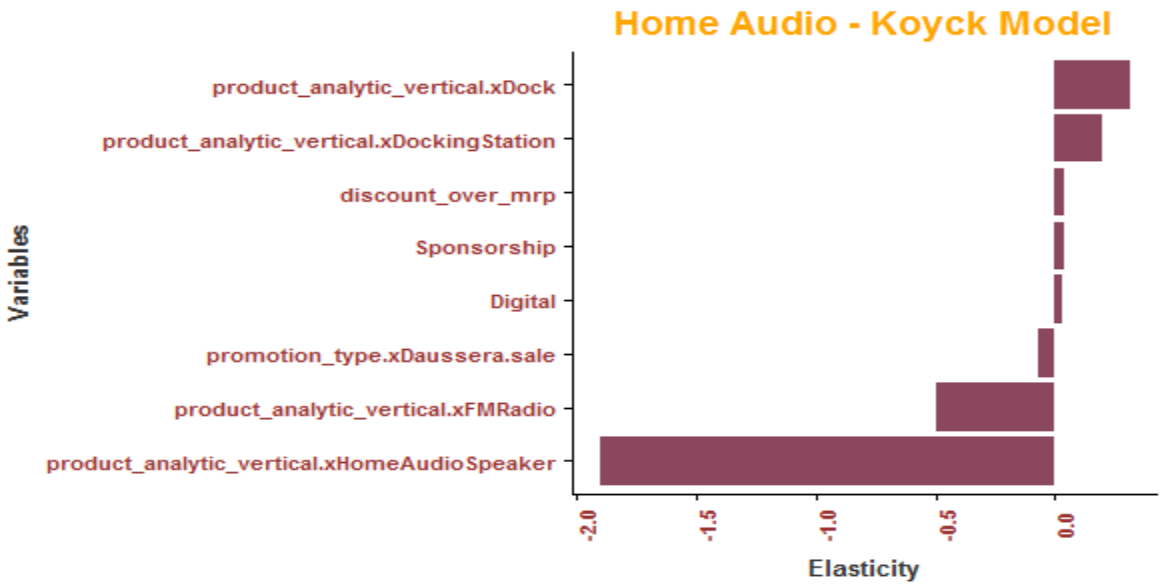
- Above figure represent elasticity of different variables w.r.t overall sales figure. Positive elasticity means increasing the value of KPI will increase the sales
- CameraMount, Telescope, Strap, Camera Battery Grip, Camera Remote Control, CameraFilmRolls are some of the significant variables and they have positive effect on revenue.
- Lens seems to be creating huge loss for company so its sale should be detained. One Unit spend on this product will decrease company's revenue by 27 Units.
- TV promotions seems to create positive impact on sale .On the other hand SEM and sponsorship advertisements are creating negative impact and hence company should stop these kind of promotable channels for camera Accessories and So more resources be allocated in TV promotions.
- Rakshabandhan sale has positive impact and it will increase sale by 0.74 units.

Recommendation-Gaming Accessories



- Above figure represent elasticity of different variables w.r.t overall sales figure. Positive elasticity means increasing the value of KPI will increase the sales
- Gamepad, Gaming Headset, Gaming Mouse are some of the significant products and they have positive effect on revenue.
- Sponsorship and Online Marketing promotions seem to create positive impact on sale. 1 unit spends on Sponsorship and Online Marketing promotions, increases revenue by 0.18 and 0.15 units respectively. So more resources be allocated in sponsorships and Online Marketing Media channels.
- Daussera Sale has very positive impact and it will increase the sale by 1.463 units
- Rakshabandhan sale is not profitable for company and hence this should be taken care by company.

Recommendation-Home Audio



- Above figure represent elasticity of different variables w.r.t overall sales figure. Positive elasticity means increasing the value of KPI will increase the sales
- Docks, Docking Station are some of the significant products and they have positive impact on revenue. Docks increases revenue by 0.31 units and Docking station by 0.195 units.
- Sponsorship and Digital promotions seem to create positive impact on sale. 1 unit spends on Sponsorship and Digital promotions, increases revenue by 0.03 and 0.0274 units respectively. So more resources be allocated in sponsorships and Digital Media channels.
- Daussera Sale has very negative impact and it should be taken care by company.
- FM Radio, Audio Speaker is causing huge loss for company and their sale should be demoted.
- Offering promotional discount is beneficial and so company should look forward to give discounts on important products. I unit discount increases revenue by 0.04 units.