ECOMMERCE CAPSTONE PROJECT FINAL-SUBMISSION

Group Members:

1. Varsha Venkapally
2. Ashish Kumar Korukonda
3. Sarthak Dey
4. Richa Malik

AGENDA

- ElecKart is an e-commerce firm specialising 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.
- As part of the marketing team, we will develop a market mix model to observe the actual impact of different marketing variables over the last year. Using our understanding of the model, we will recommend the optimal budget allocation for different marketing levers for the next year.

PROBLEM SOLVING METHODOLOGY

Exploratory Data Analysis

- Importing datasets
- Data Cleaning
- Data Preparation

- Univariate Analysis
- Bivariate Analysis
- Segmented Univariate Analysis

Feature Engineering

- Creating new features that work well with our model.
- Improving or removing the features if needed.

M odelling

- Scaling of continuous variables.
- Creating dummy variables for Categorical variables

- Applying various modelling techniques such as Linear Model , Multiplicative Model, Koyck Model and Distributive Lag Model.
- Identify the significant variables that impact the marketing budget.

- Examine the built model by model evaluation methodologies.
- Predict the final outcome and explain the budget allocation for the next year.

EXPLORATORY DATA ANALYSIS (EDA)

We would be looking into the dataset provided and perform EDA methodology to get insights from the data and find out the factors that effect the marketing budget and derive the significant variables behind it.

Below are the steps we will follow as part of EDA:

Data Sourcing

↓

Data Cleaning

↓

Univariate Analysis

↓

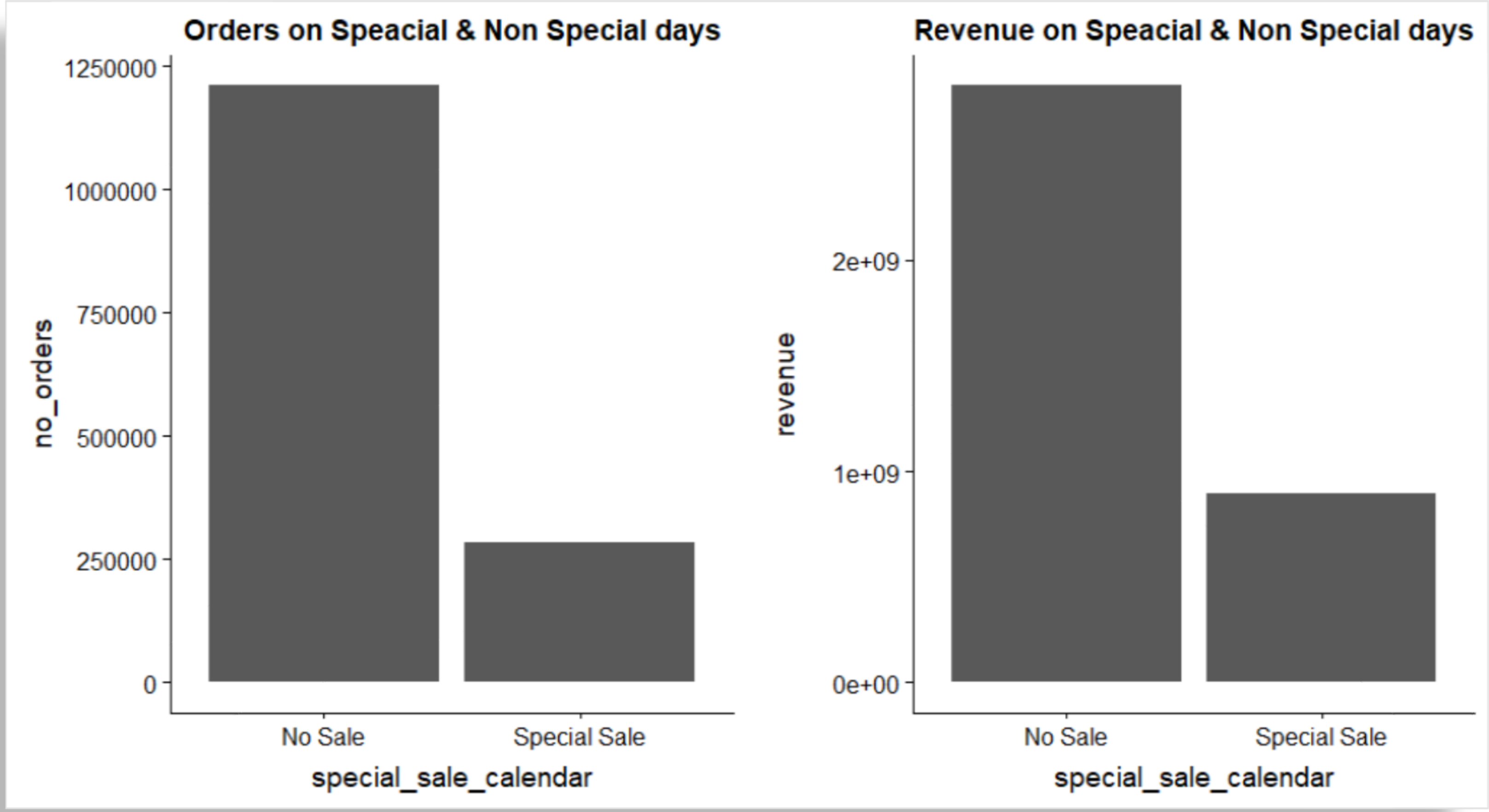
Bivariate Analysis

↓

Derived Metrics

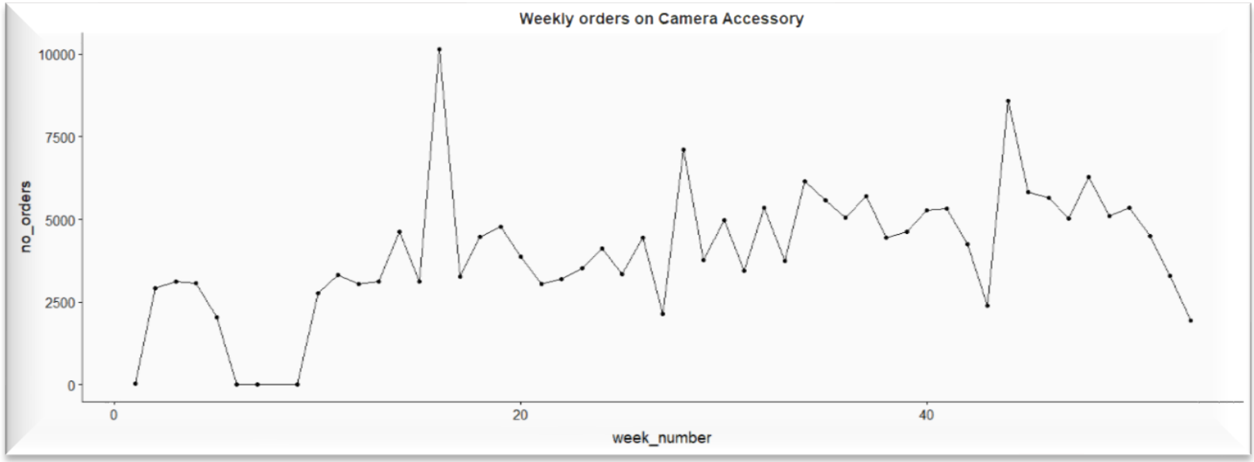
DATA CLEANING AND DATA PREPARATION

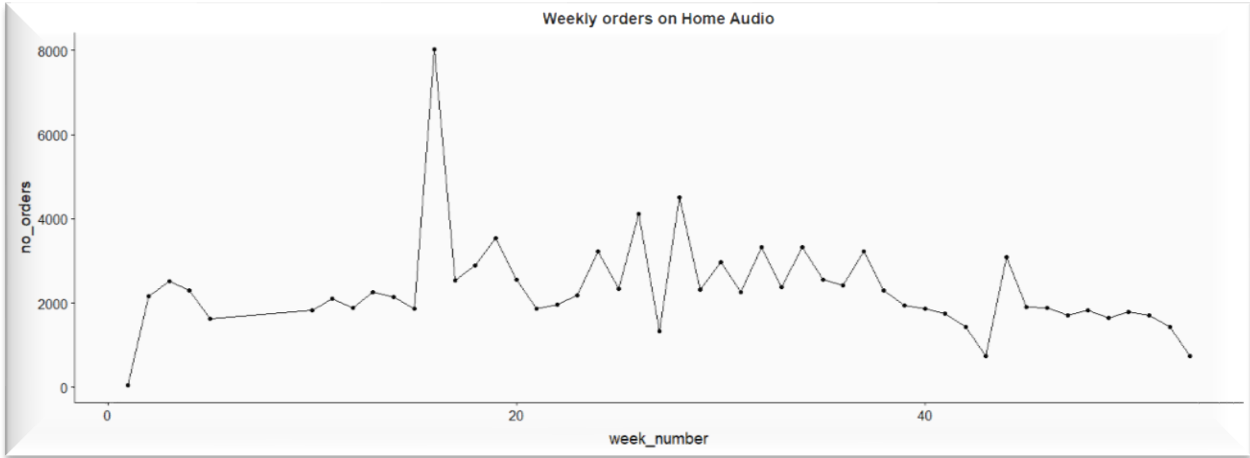
- The datasets given:
  - Consumer electronics – data is at daily level
  - Media Investments and Other information – data is at monthly level
- The above datasets are aggregated to weekly level.
- We performed below activities to achieve clean data:
  - Handling missing values
  - Removing duplicated data
  - Standardising the values
  - Filtering the unwanted data and fixing the invalid values
  - Treating outliers.
- After the data cleaning is done, we have prepared the data for analysis and below are some of the steps:
  - Deriving new variables
  - Merge all the datasets to a single data frame
  - Check for any anomalies in the merged data.

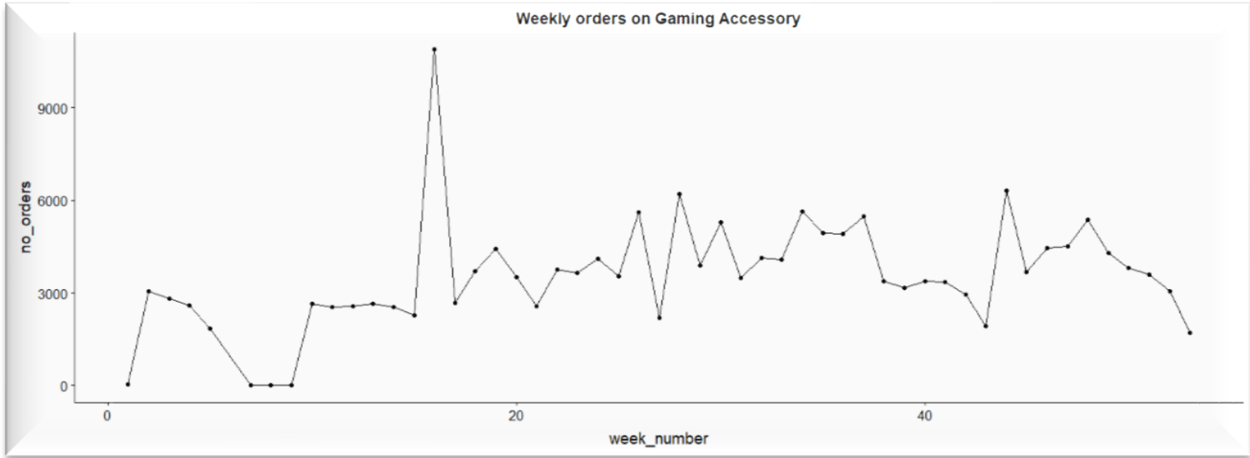


EDA-Orders & Revenue on Special and Non Special Days

- Plot 1 refers to the number of orders on Special & Non Special days.
- Plot 2 refers to the Revenue obtained on Special & Non Special days.
- We can see that the number of orders are more on Regular days and also most of the revenue is obtained on Regular days.
- In order to keep order level high and increase the revenue on Special days , we can increase the media investments and thereby increasing the customers inflow.

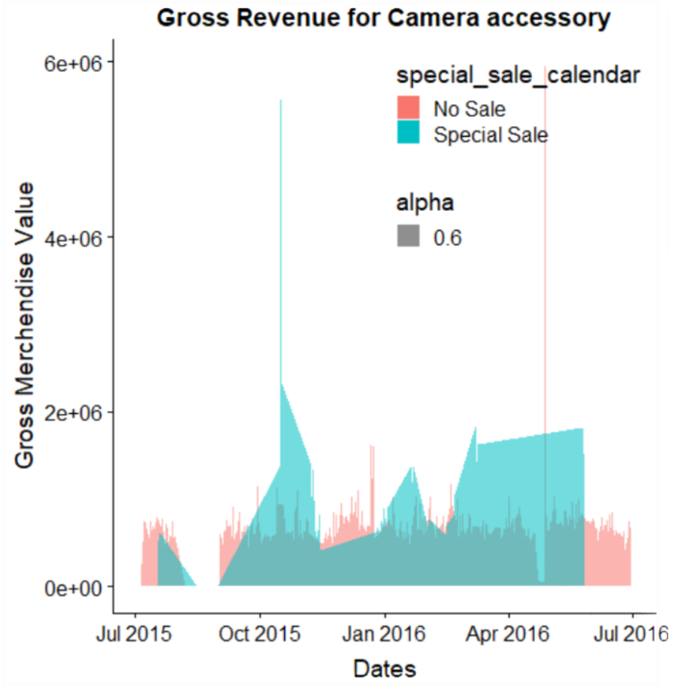


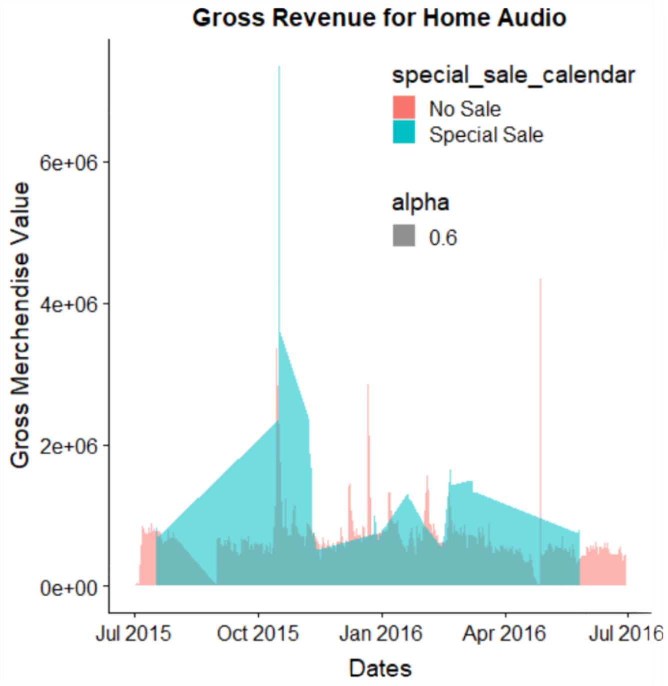


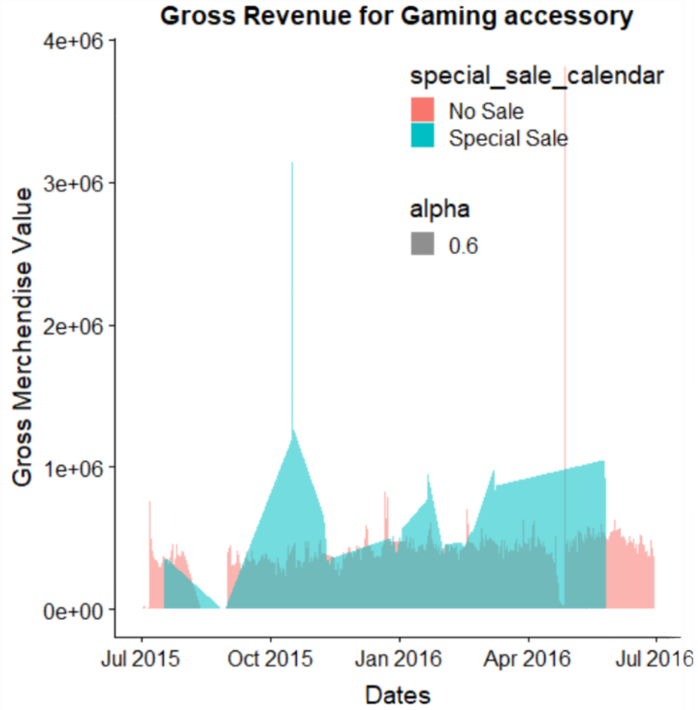


EDA-WEEKLY ORDERS FOR THREE PRODUCT CATEGORIES

- Graphs are plotted for three product categories,
  - Camera Accessory
  - Home Audio
  - Gaming Accessory
- We can see that during the week 17-18 we have highest orders.
- Week 17-18 are the weeks on Special sale (Dussehra).

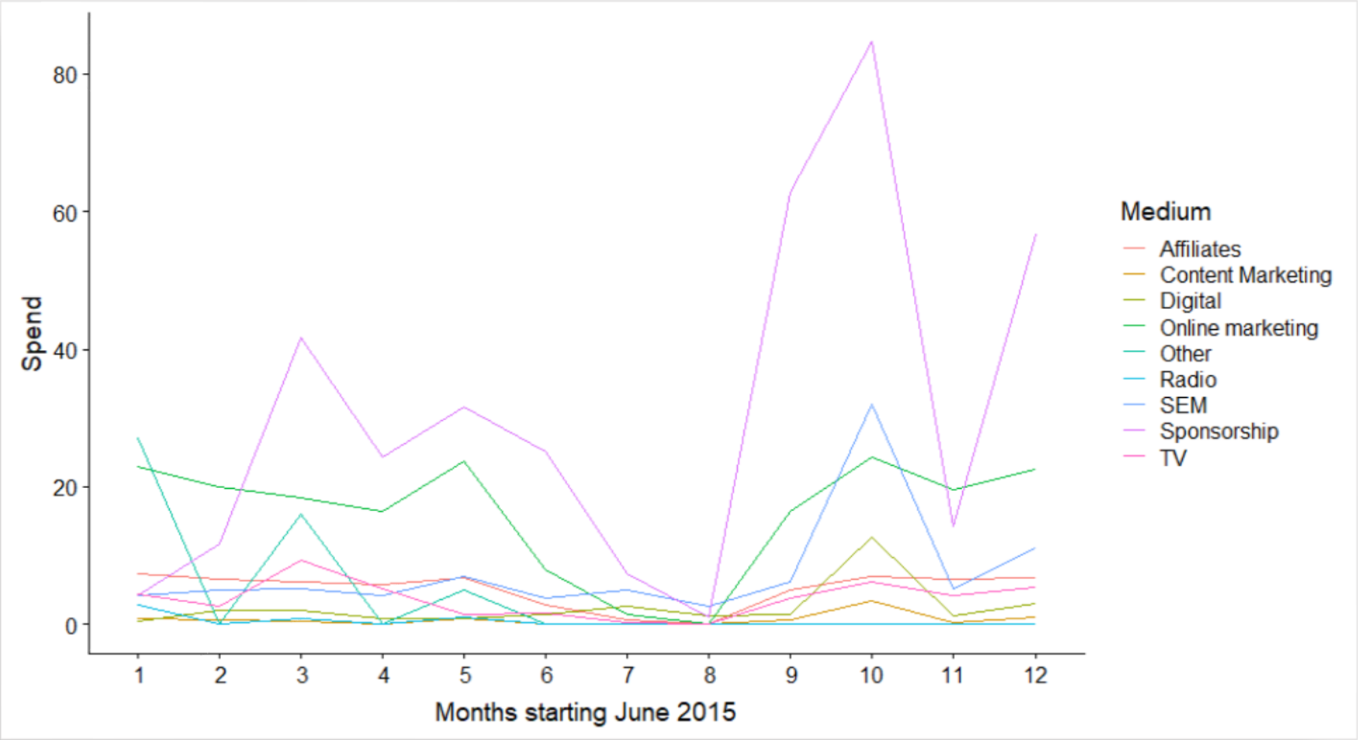






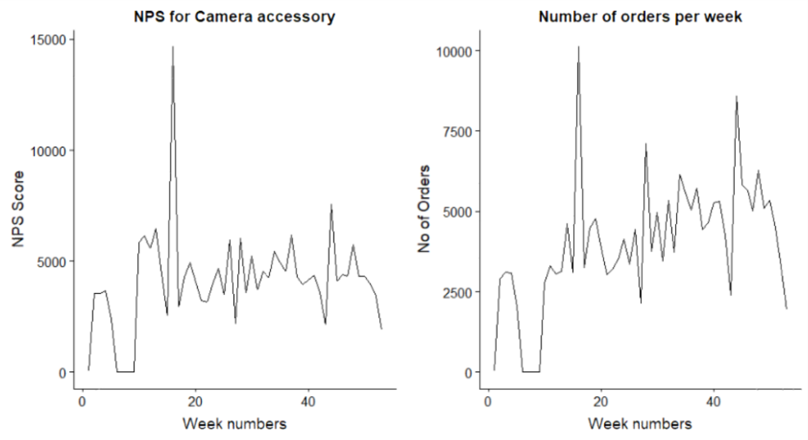
EDA-VISUALIZE GROSS REVENUE ON SPECIAL AND NON SPECIAL DAYS

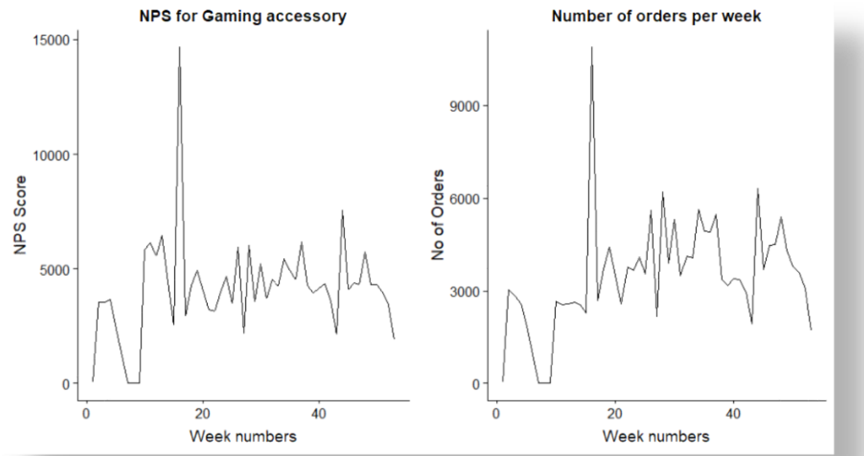
- Visualization of GMV across three product categories
  - Camera Accessory
  - Home Audio
  - Gaming Accessory
- As we can see, the Revenue is higher on Special Sale days such as Diwali and Christmas when compared to Regular days.
- Therefore, we can safely assume that increase in marketing budget during Special Sale days will have high returns in revenue.

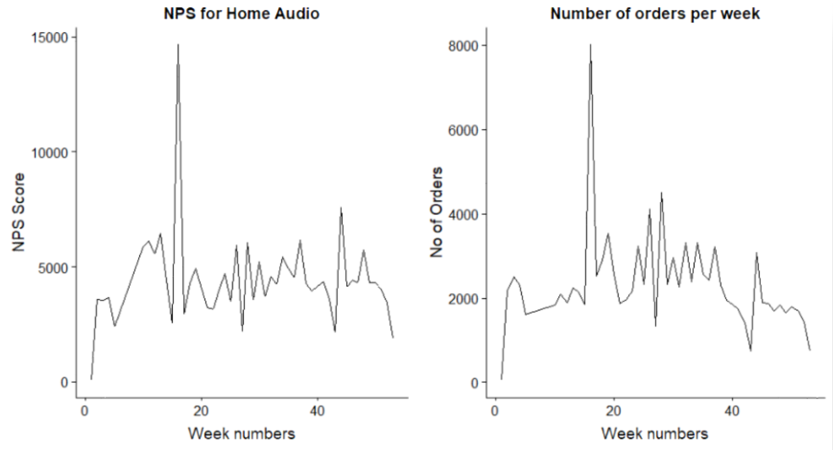


EDA-MARKETING SPENDS ACROSS THE MONTHS

- Marketing spends starting from July'15 to June'16.
- There are 9 types of mediums for Marketing Spends,
  - Affiliates
  - Content Marketing
  - Digital
  - Online Marketing
  - Other
  - Radio
  - SEM
  - Sponsorship
  - TV
- As we can see high spends occur via Sponsorship and lesser spends occur via Radio.
- Most of the peaks occur during seasonal months(i.e. Month of Sept-Oct'15) .







EDA-CUSTOMER SATISFACTION VS ORDERS PER WEEK

- Graphs are plotted for three product categories.
- Plot 1 refers to the Customer Satisfaction (Net Promotor Score) on weekly scale.
- Plot 2 refers to the number of orders placed per week.
- As we can see that there is a high correlation between Customer Satisfaction and number of orders per week. Thereby having a strong customer loyalty.

FEATURE ENGINEERING

- As part of Feature Engineering we have derived few columns based on business understanding.
  - List Price - Revenue per order divided by number of units per order
  - Special Sale Calendar - if it's a Special Day Sale or a Regular Sale
  - Special Sale Calendar Day - If it's a Special Day Sale then what is the Day (Christmas, Dussehra).
  - Promotional Offer/Discount - How much is the discount offered.
  - Product Pricing - If the product ordered is a Mass/ Premium/ Medium Priced product.
  - Ad stock for each Media Investment - We have derived the ad stocks of different Media Investments at a ad stock rate 0.5 and by using below formula.

$$A_t = X_t + \text{adstock rate} * A_{t-1}.$$

MODELLING - I

- Before we get into modelling below are the steps performed:
  1. Split the main dataset into three product categories.
     - Gaming Accessory
     - Camera Accessory
     - Home Audio
  2. Create dummy variables for all the categorical columns in the dataset.
  3. Scale all the numerical variables in order to reduce the sensitivity to magnitude.
  4. Remove unwanted columns which are not useful for modelling, such as columns with one level of values (ex :Product Super Category).
  5. Independent variable here is : GMV - Gross Merchandise Value or Revenue
  6. Split the dataset into test and train using the below statement.
     - sample.split(df$gmv, SplitRatio = 0.7)
  7. Using VIF(multi-collinearity) and P-values from the models we tune the variables to get the significant variables.

MODELLING - II

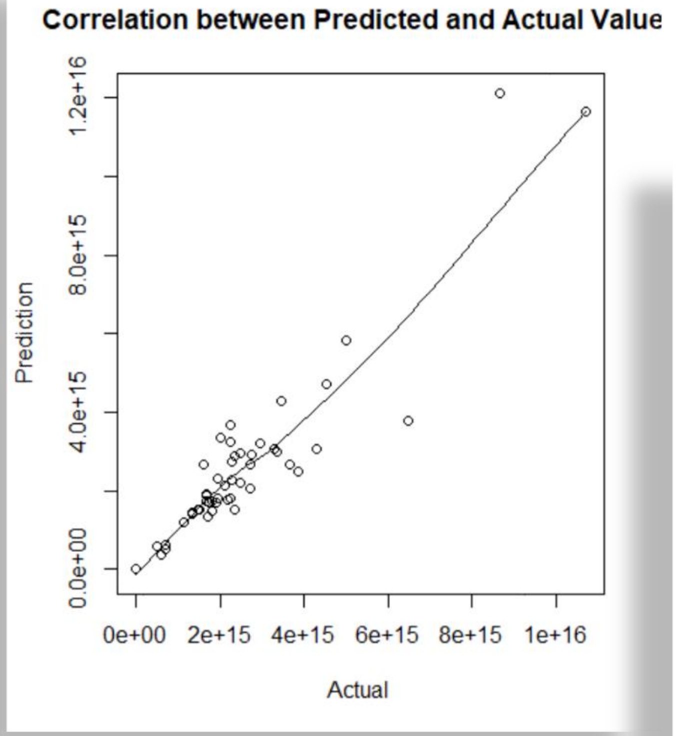
- Below are the list of KPI's used for modelling,
  1. Product Information: Units, Sub Category, Product Vertical.
  2. Product Delivery Information: deliverybdays ,deliverycdays, sla, procurement sla.
  3. Media Spends: Tv, Radio, Digital, online spends etc.
  4. Media spends converted into Ad stocks.
  5. Pricing Information: GMV, MRP, discount, list price.
  6. NPS Score
  7. Week Number
  8. Special Sale Information
  9. Lag variables were derived using MRP, discount, adstock variables
- Below are the model evaluation techniques performed:
  - Compare test and train R-squared values
  - Performed Cross validation on entire dataset (5 folds)
  - Compare the mean square error to get the best fit.

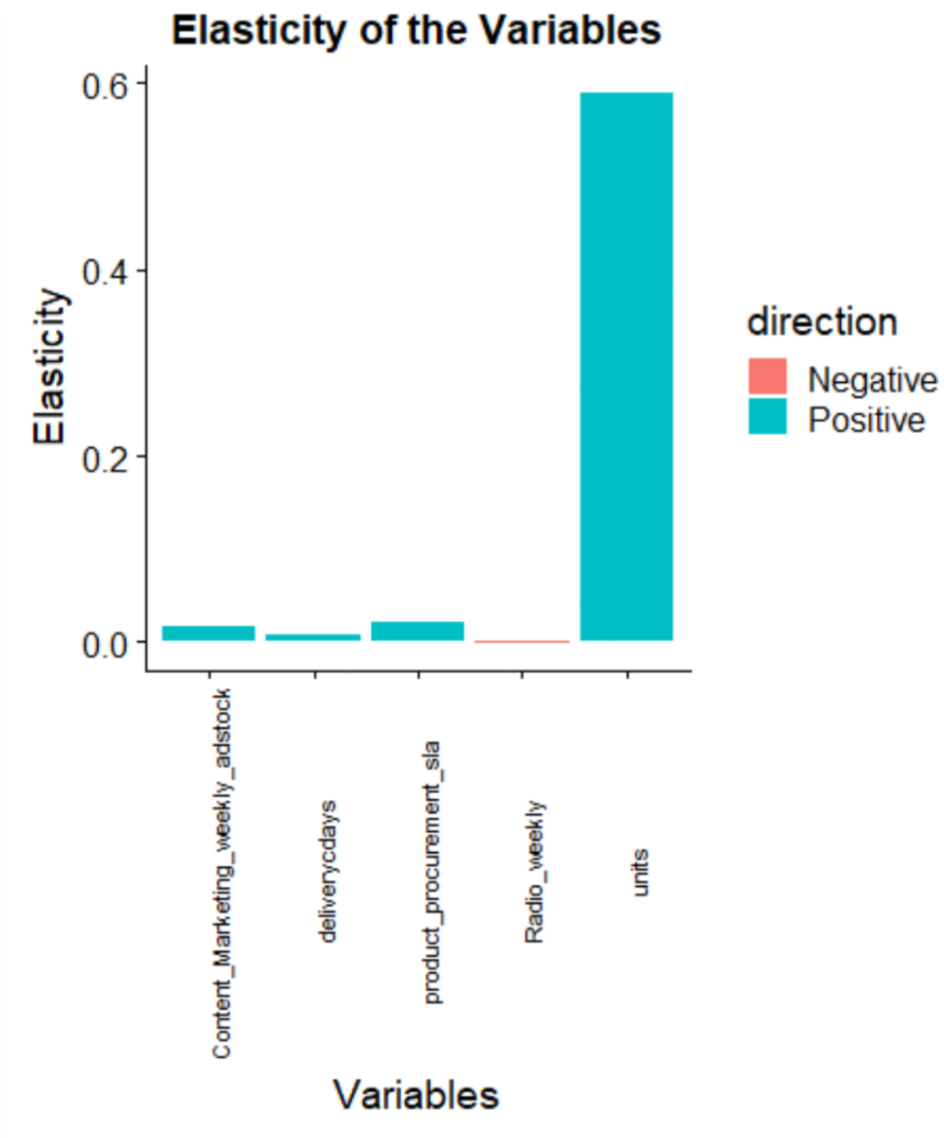
MODELLING – CAMERA ACCESSORY

- Models performed on Camera accessory:

Models	Train : R^2	Test : R^2	CV : Mean square error	Significant variables
Linear	0.888	0.88254	2.27E+11	Content_Marketing_weekly ,units , deliverycdays, product_procurement_sla
Mutiplicative	0.9851	0.969	0.0124	Radio_weekly,Content_Marketing_weekly_adstock ,units, deliverycdays,product_procurement_sla
Kyock	0.909	0.905	1.83E+11	Radio_weekly,NPS_Score_weekly,TV_weekly_adstock+units ,deliverycdays
Distributed lag	0.926	0.851	1.89E+11	NPS_Score_weekly,deliverycdays ,gmv_1
Distributed lag + Multiplicative	0.945	0.9	0.0508	NPS_Score_weekly ,deliverycdays ,gmv_1

- Final model is selected based on the results from model evaluation.
  - Test and Train R-squared are better compared to other models.
  - From the CV performed the mean square error is the least among other models





CAMERA ACCESSORY- FINAL MODEL

- Final model is selected as Multiplicative Model because of the following reasons,
    - Lower Mean Square Error
    - Good Correlation between Test and Train Values.
- Correlation between Predicted and Actual values is 0.984.
- Calculation of elasticity is done on below significant variables.

Variable	Elasticity	direction
Radio_weekly	-0.000354	Negative
Content_Marketing_weekly_adstock	0.017342	Positive
units	0.589291	Positive
deliverycdays	0.006875	Positive
product_procurement_sla	0.020286	Positive

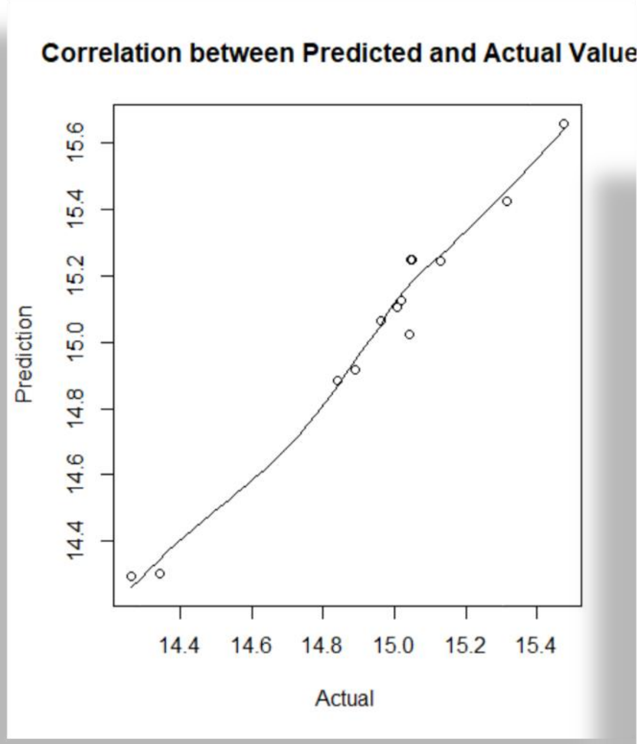
MODELLING – GAMING ACCESSORY

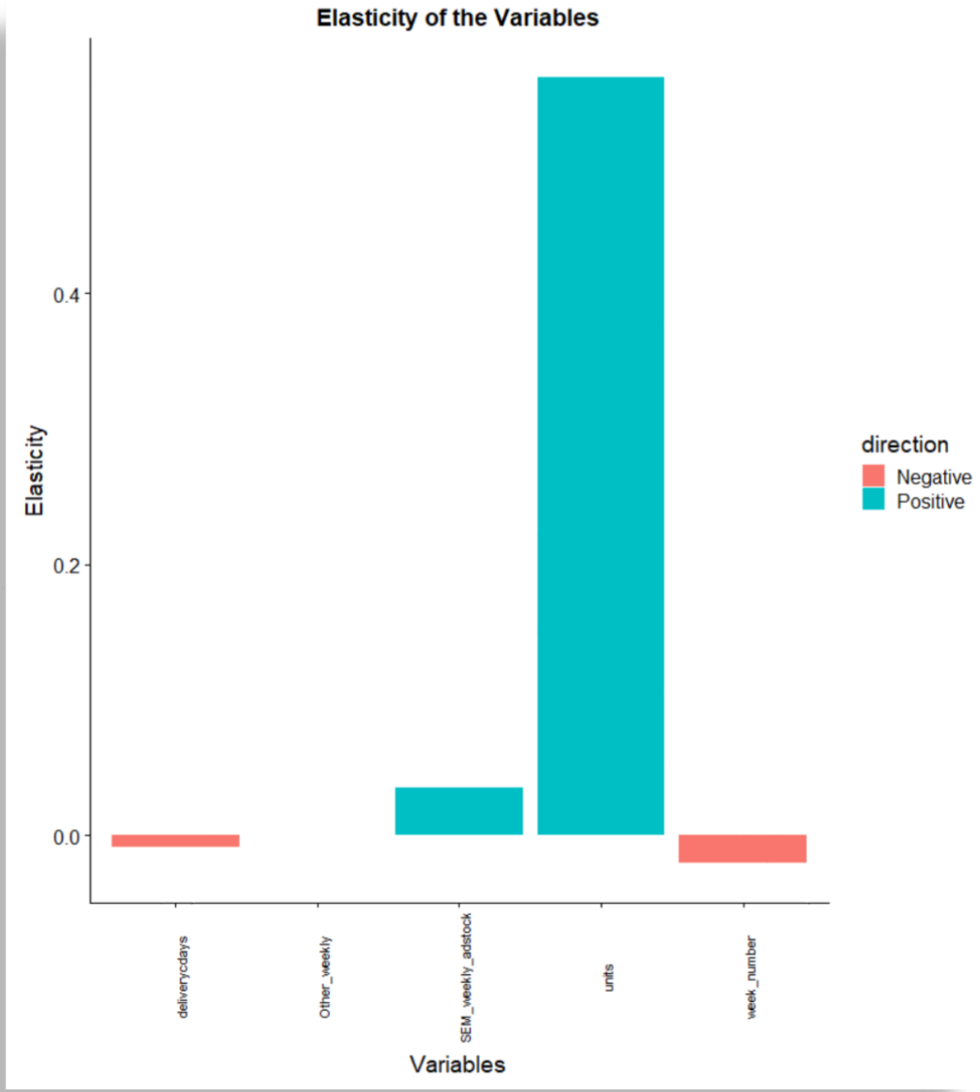
- Below are the models performed on Gaming accessory:

Model	Train : R2	Test : R2	CV : Mean square error	Signicant variables
Linear	0.972	0.939	1.37E+10	week_number , Affiliates_weekly , Radio_weekly , Sponsorship_weekly_adstock , units
Mutiplicative	0.994	0.977	0.00536	week_number,Other_weekly ,SEM_weekly_adstock ,units , deliverycdays
Kyock	0.966	0.939	1.58E+10	week_number,Radio_weekly ,Digital_weekly_adstock, Online_marketing_weekly_adstock ,units, product_procurement_sla
Distributed lag	0.972	0.935	1.31E+10	week_number,Affiliates_weekly,Radio_weekly, units , Sponsorship_weekly_adstock_2
Distributed lag + Multiplicative	0.989	0.979	2.93E+10	Online_marketing_weekly_adstock,units,deliverycdays

- Final model is selected based on the results from model evaluation.
  - Test and Train R-squared are better compared to other models.
  - From the CV performed the mean square error is the least among other models

GAMING ACCESSORY- FINAL MODEL





- Final model is selected as Multiplicative Model because of the following reasons,
  - Lower Mean Square Error
  - Good Correlation between Test and Train Values.
- Correlation between Predicted and Actual values is 0.989.
- Calculation of elasticity is done on below significant variables.

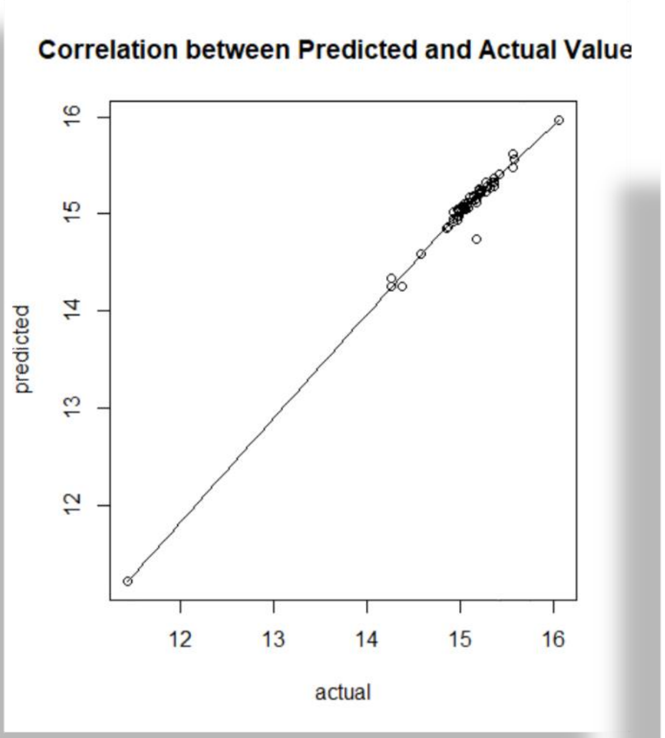
Variables	Elasticity	direction
week_number	-0.020829	Negative
Other_weekly	0.000113	Positive
SEM_weekly_adstock	0.035264	Positive
units	0.558939	Positive
deliverycdays	-0.009175	Negative

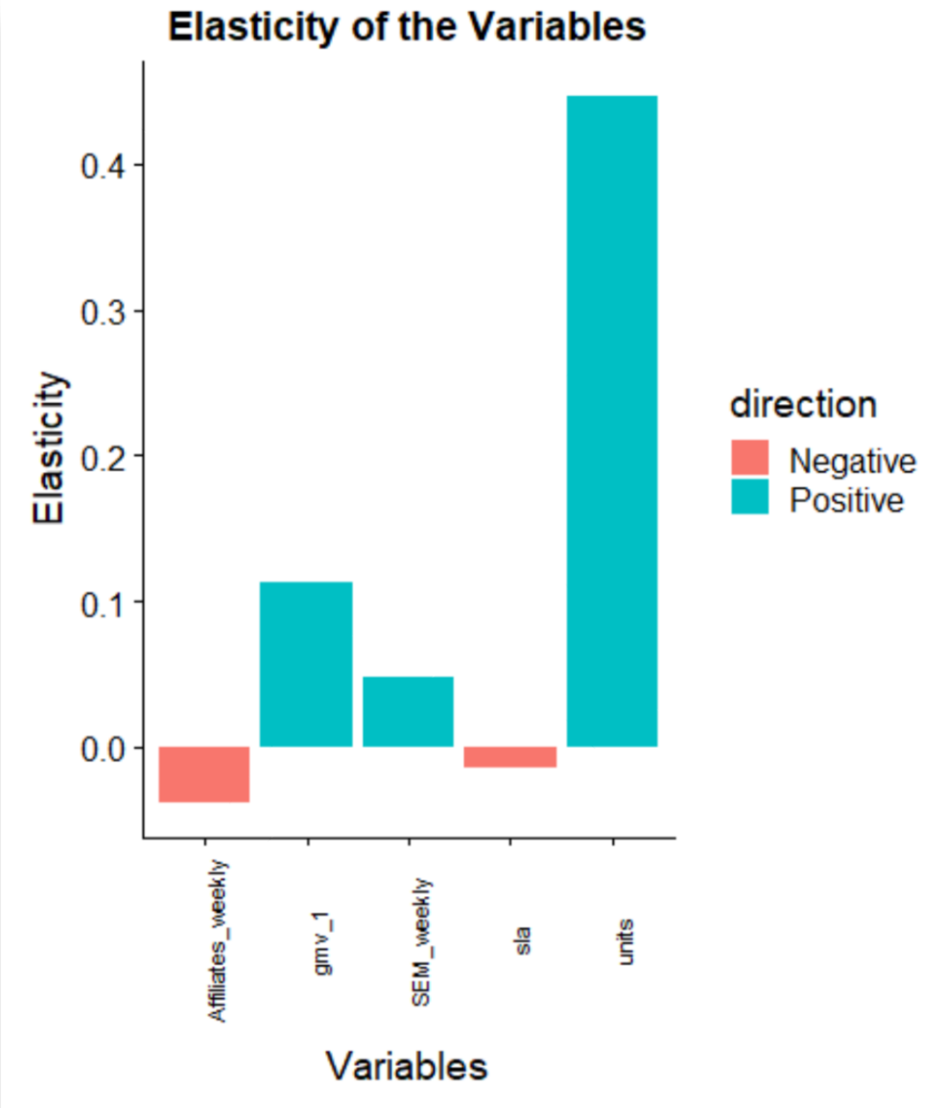
MODELLING – HOME AUDIO

- Below are the models performed on Gaming accessory:

Model	Train : R2	Test : R2	CV : Mean square error	Signicant variables
Linear	0.985	0.987	3.15E+10	Content_Marketing_weekly, units
Mutiplicative	0.994	0.987	0.00305	Content_Marketing_weekly, units
Kyock	0.979	0.985	2.92E+10	SEM_weekly, units,lag_gmv
Distributed lag	0.982	0.988	2.31E+10	SEM_weekly, Radio_weekly_adstock, units, sla
Distributed lag + Multiplicative	0.971	0.988	2.99E-03	Affiliates_weekly, SEM_weekly, units, sla, gmv_1

- Final model is selected based on the results from model evaluation.
  - Test and Train R-squared are better compared to other models.
  - From the CV performed the mean square error is the least among other models





HOME AUDIO-FINAL MODEL

- Final model is selected as Multiplicative +Distributed Lag Model because of the following reasons,
  - Lower Mean Square Error
  - Good Correlation between Test and Train Values.
- Correlation between Predicted and Actual values is 0.994.
- Calculation of elasticity is done on below significant variables.

Variables	Elasticity	direction
Affiliates_weekly	-0.0376	Negative
SEM_weekly	0.0476	Positive
units	0.4461	Positive
sla	-0.0147	Negative
gmv_1	0.1135	Positive

RECOMMENDATIONS

- Based on the significant variables from the final models below are the recommendations to optimise the marketing budget allocation.
- Home Audio –
  - There should be reduction in rewards or loyalty programs to reduce the Affiliates Marketing budget as there is a negative impact on the Revenue.
  - There should be increased allocation of budget in Search Engine Marketing by performing few digital marketing strategies to increase the visibility of the website in search engines results pages.
- Gaming Accessory-
  - There should be increased allocation of budget in Search Engine Marketing by performing few digital marketing strategies to increase the visibility of the website in search engines results pages.
  - Increasing budget allocation in various miscellaneous marketing strategies such as Pay-per click marketing, optimising the marketing efforts on mobile devices and email marketing will help the growth in revenue and sales of the company.
- Camera Accessory-
  - Although Radio Marketing is the best reach to customers but primary drawback is the people listening to it are often engaged in other activities therefore we don't normally get the same level of attention with radio ad as you might get through other media. Therefore , reduction in Radio marketing budget will positively impact the revenue.
  - There should be a dedicated team to create innovative videos of ads and sharing them on various social media or blogs. These ads should target the indented audience, taking these points into consideration an increased budget in Content Marketing will help the growth in revenue or sales.