E-Commerce Capstone Project

Market Mix Modelling

Presented by:

Dhruv Vashisht

Pallavi RV

Subhradeep Ray

Shilpi Agrawal

Objective & Data

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 aim is to develop a market mix model to observe the impact of marketing spending over the last year. On the basis of this understanding, recommendations are to be made regarding the optimal budget allocation for different marketing levers for the next year.

We have 5 datasets:

1. **Consumer Electronics** comprising of details of products purchased such as order date, product category, payment type, Gross merchandise value, etc.
2. **Media Investment** comprising of money spent on a monthly basis on different media outlets
3. **Monthly NPS score** comprising of monthly customer feedback scores
4. **Special Sale Calendar** comprising of all holiday dates during the year
5. **Product List** comprising of all products available at ElecKart

Data Preparation and Engineered KPIs

The following steps were taken to prepare the data for modeling:

1. Data cleaning is done for each of these data sets.
2. NPS monthly scores and investments on different media are converted to weekly scores.
3. The main Electronics data is cleaned by removing duplicate entries, converting few important NA to 0 or mean and also removing few rows having majority NA values.
4. The week column is also added to the data as it is supposed to be done at week level and all tables are merged.

The following new KPIs were derived:

1. **Spec/Specday**: Spec = 1 for special day. Specday contains list of Special Days on which a sale is offered
2. **Adstock**: Every marketing investment column has an adstock effect associated with it, which is represented by the adstock columns. The value for adstock rate is taken as 0.5
3. **Payment mode**: If payment is made by COD or prepaid
4. **List Price**: Price of product obtained by dividing gmv by units sold
5. **Promo Offer**: Discount on price on account of sale or otherwise
6. **Prepaid order percentage** : Denotes the orders that are prepaid as a percentage of total orders.

EDA

Exploratory Data Analysis was conducted to gain insights into the variables and observe patterns on a weekly basis.

The following plots were observed:

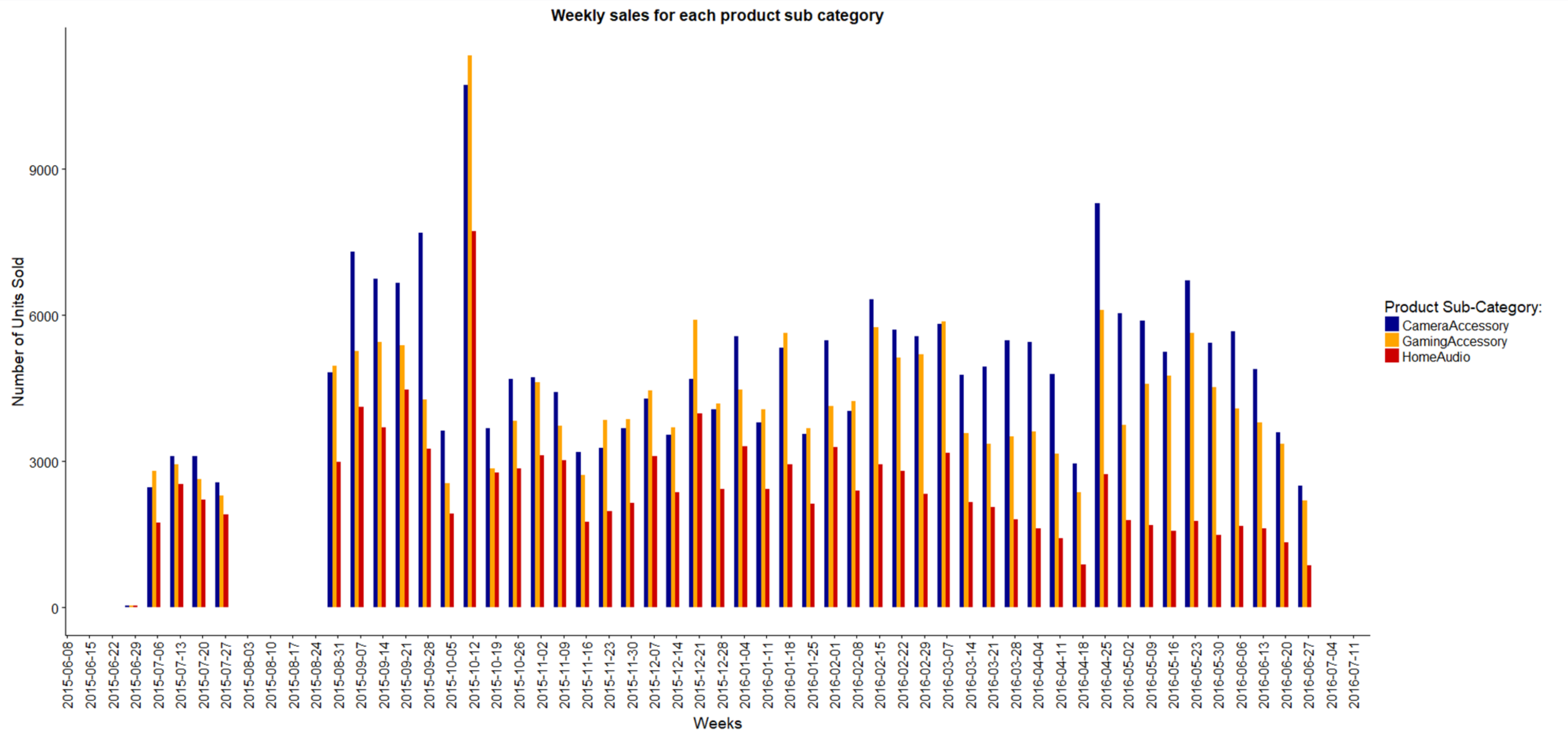
1. Weekly sales for each product sub category (Camera Accessory, Home Audio, Gaming Accessory)
2. Weekly Product Units sold by different payment types (COD vs Prepaid)
3. Weekly GMV for each product subcategory
4. GMV for each sale day
5. GMV on sale day vs non sale day
6. GMV vs total investment on marketing

EDA - Observations

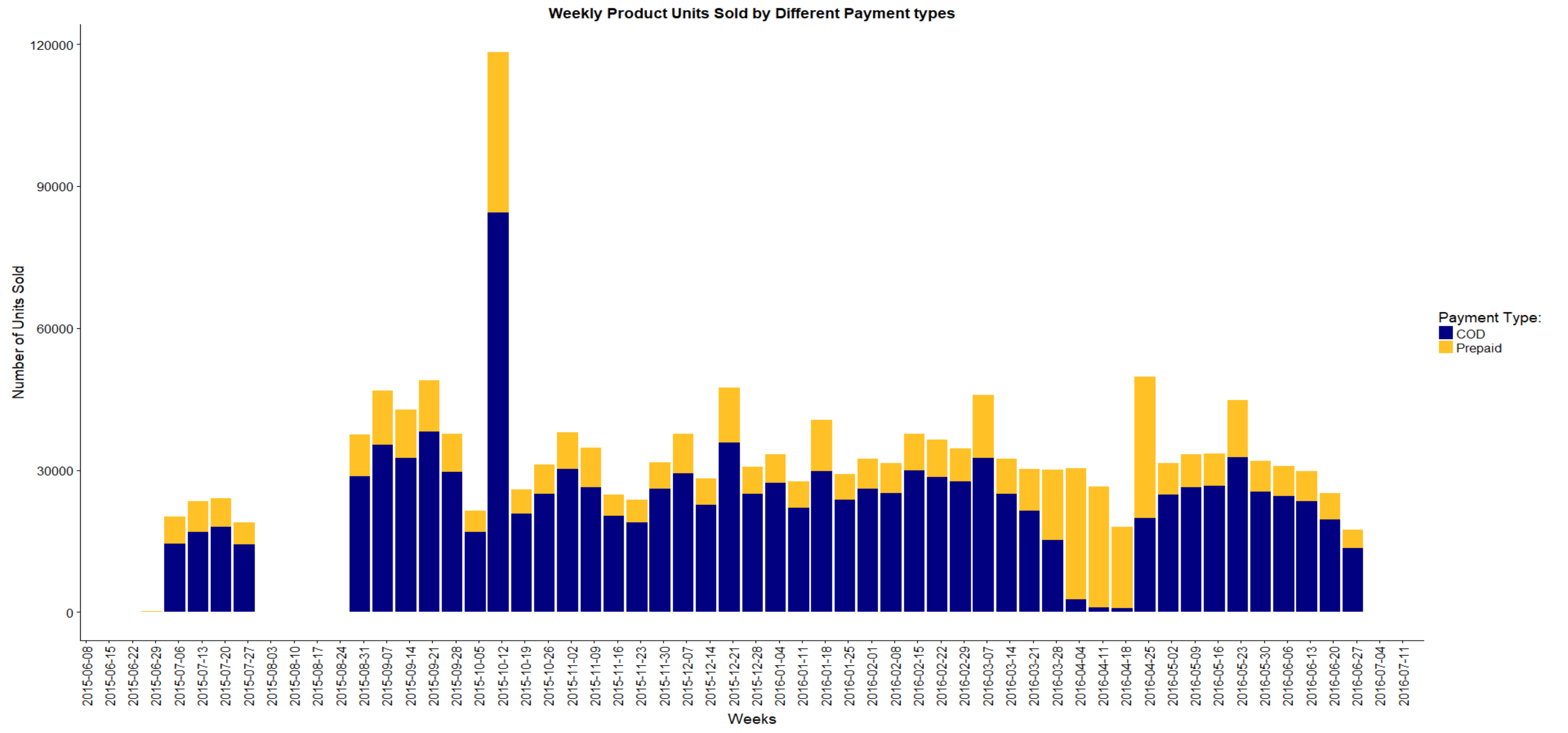
The following observations were made from EDA plots:

1. Camera accessories have been purchased the most throughout the year while home-audio have been purchased the least
2. Most orders are made on COD payment type
3. Home Audio in spite of being purchased the least, generate the most GMV. Gaming accessories generated least GMV
4. Apart from Regular Sale, Big Diwali and Dussehra Sale have generated most revenue
5. Sale days in spite of being very few in number have generated 25% of the GMV. These days are most effective in generating revenue
6. Revenue shoots up when the investment is increased at certain points then again drops down

EDA – Sales during special days



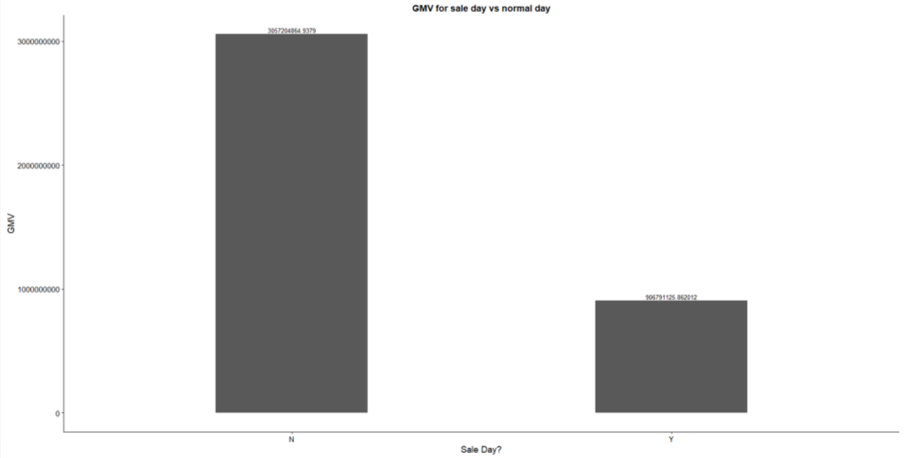
EDA – Sales Prepaid vs COD

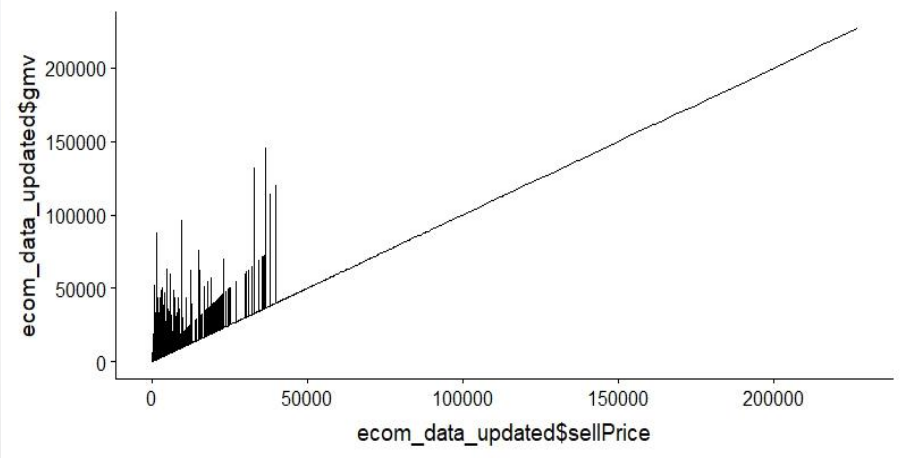


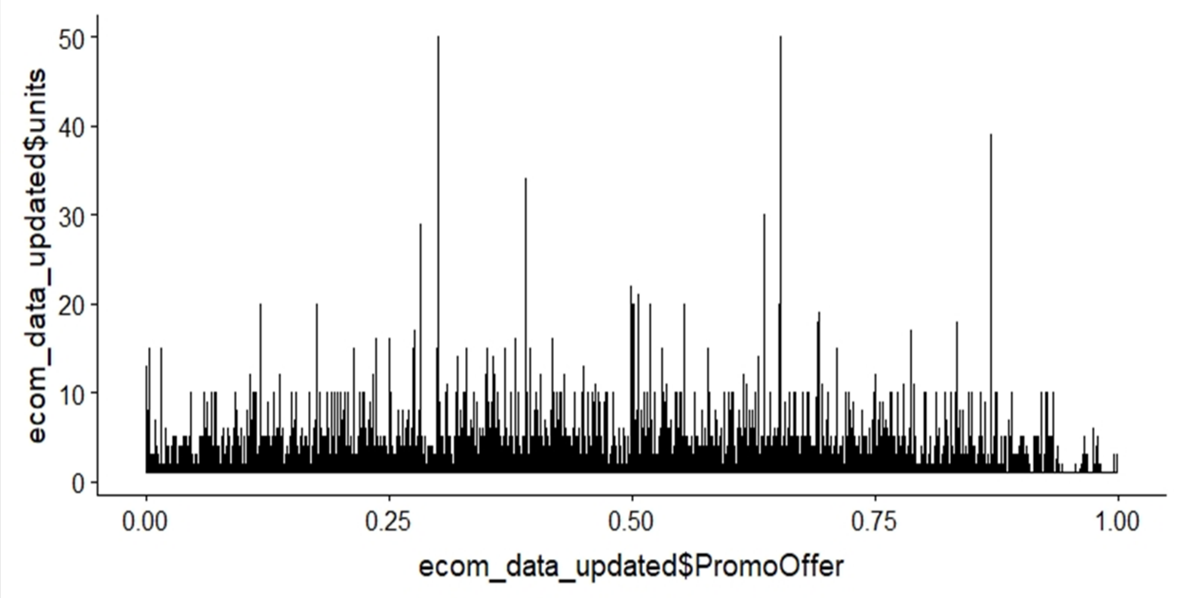
EDA

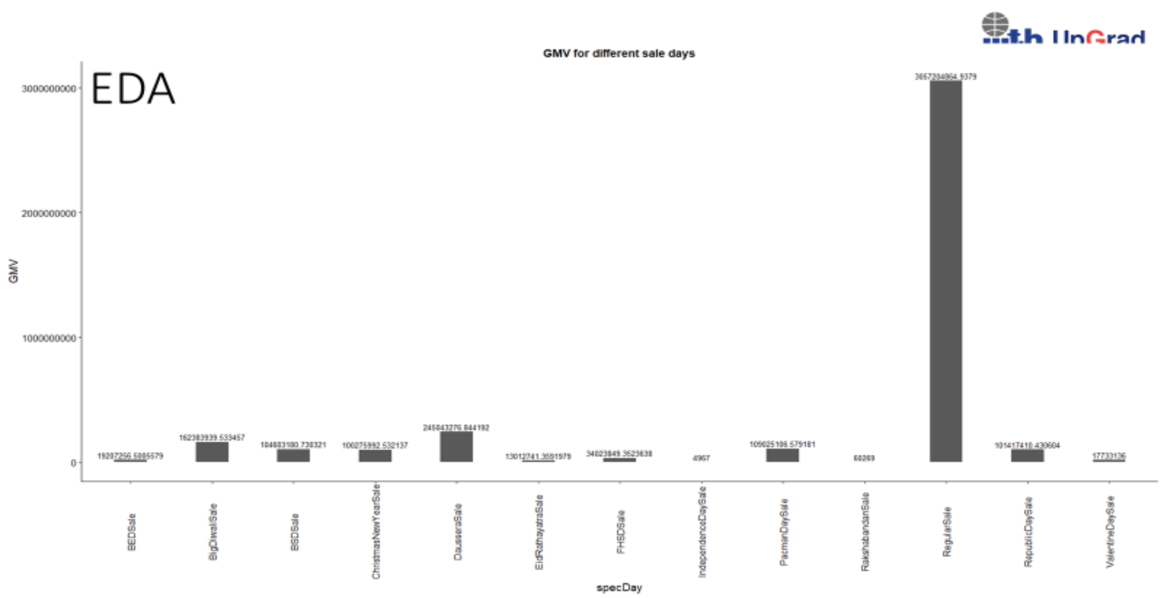
Weekly GMV for each product sub category
GMV
15000000
10000000
5000000
0
Weeks
as.factor(product_analytic_sub_category)
CameraAccessory
GamingAccessory
HomeAudio

EDA Observations









Model Building

For Camera Accessories:

Model	Significant Variables	R-sq value (CV)
Linear	SLA + product_mrp + NPS + prepaid	0.89
Multiplicative	NPS + Spec + Prepaid	0.623
Distributive Lag	units + spec + prepaid + sellPrice + PromoOffer + GMV_lag_2_per	0.99
Koyk	PromoOffer + NPS + TV + Sponsorship + Total_Investment + Online_Marketing	0.89
Distributive Lag - Multiplicative	PromoOffer+ GMV_lag_1_per+ GMV_lag_2_per+ prepaid	0.67

Chosen Model

Model Building

For Home Audio:

Model	Significant Variables	R-sq value (CV)
Linear	sla +Sponsorship + Content_Marketing + SEM +spec + PromoOffer+specDay.xBigDiwaliSale+specDay.xRegularSale	0.96
Multiplicative	spec+Content_Marketing+sla+PromoOffer+Total_Investment+prepaid+prepaidPerOrder	0.95
Distributive Lag	sla+Total_Investment+Online_Marketing+Sponsorship+Other+specDay.xRegularSale+spec	0.98
Koyk	specDay.xRegularSale+spec+sla+Sponsorship+Digital	0.99
Distributive Lag - Multiplicative	sla+ PromoOffer + prepaid + SEM + prepaidPerOrder+Content_Marketing	0.98

Chosen Model

Model Building

For Gaming Accessories:

Model	Significant Variables	R-sq value (CV)
Linear	spec + specDay.xChristmasNewYearSale + PromoOffer + NPS + Online_Marketing + Sponsorship + Digital + prepaid	0.93
Multiplicative	PromoOffer + Digital + SEM + Online_Marketing + Total_Investment + prepaid	0.99
Distributive Lag	spec + specDay.xChristmasNewYearSale + PromoOffer + NPS + Other + Online_Marketing + Sponsorship + Digital + prepaid	0.67
Koyk	spec + PromoOffer + NPS + Other + Online_Marketing + + Digital + prepaid	0.95
Distributive Lag - Multiplicative	Radio + Online_Marketing + TV + spec + prepaid + Sponsorship + NPS + Digital + SEM + prepaidPerOrder	0.43

Chosen Model

Challenges

We encountered a number of challenges while working on the models:

1. Bringing the data down to weekly granularity posed an issue as the investment data was given at the monthly level. This required us to group the days by month and week and divide the data accordingly so that there is no misrepresentation of the figures.
2. Engineering the KPIs with multiple sets of data source was challenging. Deriving the right KPI's needed in depth understanding of the datasets.
3. Having grouped the data by week, we were left with 52 rows of data for each week. The number of features however far exceeded this which led to overfitting. As a result we had to see which of the existing and derived KPIs were actually concerned with the problems statement and remove the irrelevant ones accordingly.
4. Some of the KPIs such as units sold and price are highly correlated to the GMV. As such, although they turned up in every model as a significant variable, we understood that these variables were not as relevant to the investment problem statement. These, too, had to be removed.
5. Understanding the different kinds of models algorithms and how they behaved also posed a challenge.
6. Working as a distributed team with everyone working on different sections was a challenge to help each other in the team.

Recommendations

On the basis of the models chosen, we would like to put forth the following recommendations:

1. **Recommendation - Camera Accessories**

ElecKart company should invest more on online marketing than on sponsorships.

Digital marketing should be given more importance

2. Recommendation - Home Accessories

Eleckart company should have a good logistics process as sla ie number of days to deliver product has strong impact on revenue.

Special day sales is earning more revenue

Company should invest more on Sponsorship and online marketing

3. Recommendation - Gaming Accessories

ElecKart company should invest more on online marketing than on sponsorships.

Offering Promotional discounts would add more value here.

Holiday season/discounts is getting more sales