E-commerce Capstone Project

Market Mix Modelling for ElecKart

By
Vyom Bhatt
Abdul Mukheem

The Problem Statement

Over the last one year, ElecKart has spent a significant amount of money on marketing.

For the coming year, ElecKart needs to understand how to split their marketing budget, which includes spending on commercials, online campaigns, and pricing & promotion strategies. This project will help them understand what were the most impactful marketing levers and KPIs that resulted in incremental revenue in the last year.

As part of the marketing team working on budget optimization, we've developed market mix models to observe the actual impact of the different marketing variables over the last year. We've used this to suggest recommendations for ElecKart on how to most optimally split their marketing budget across the different channels for the upcoming year.

Project Objectives

- Exploratory Data Analysis
    - To understand the data and generate business-impactful insights
- Market Mix Modelling
    - To identify the best performing models across the different product sub-categories
- Recommendations
    - The learnings from EDA & models to help plan the marketing budget of ElecKart for the coming year

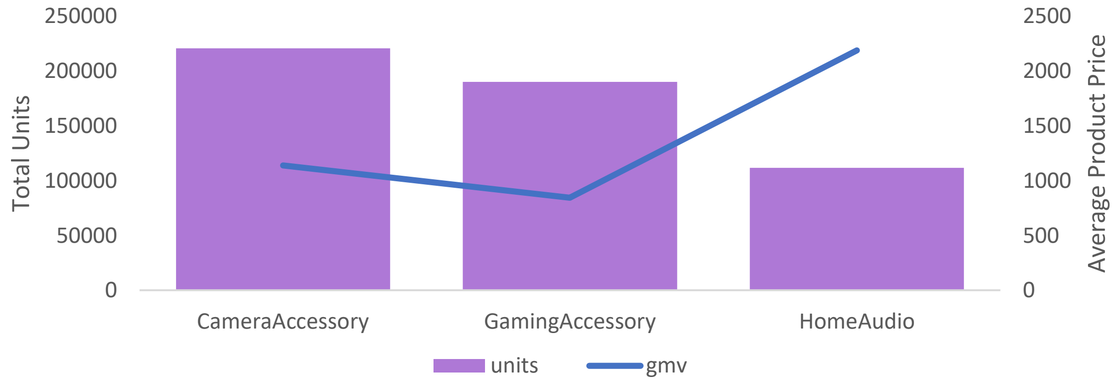
Methodology

1. Reading the different datasets – orders, media data, weather, holidays, product information
2. Cleaning and pre-processing the data
3. Feature Engineering to generate new marketing KPIs
4. Exploring the data to generate business-relevant insights
5. Visualizing the impact of marketing on revenue: current effect, carry-over effect, dynamic effect, shape effect, media effect
6. Splitting the data into 3 sets for each of the following product sub-categories: Camera Accessory, Game Accessory, Home Audio. Aggregate each dataset at a week level.
7. Creating the different variants for MMM to model the KPIs generated:
   - ❑ Additive Model
   - ❑ Multiplicative Model
   - ❑ Koyck Model
   - ❑ Distributed Lag Model
8. Identify the best performing model for each of the product sub-categories
9. Present recommendations on how the upcoming annual marketing budget should be planned

Business – Relevant Insights from EDA

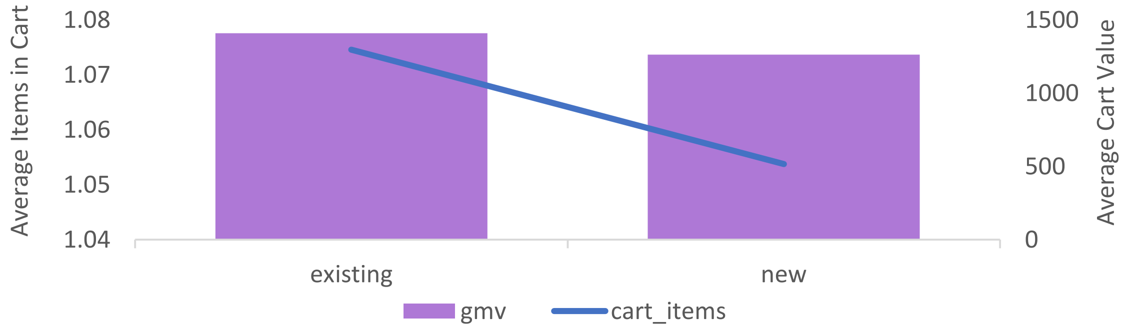
We tried identifying the different trends being depicted in the data to understand the buying pattern of different customers.

- Distribution of transactions across the 3 product sub-categories



We observe that the Home Audio category was the one with the greater average product price among the 3 categories, and the Camera Accessory Products were the most sought after among the 3 categories. Hence, the home audio category had the pricier products among all the 3 categories.

- New v. Existing Customers



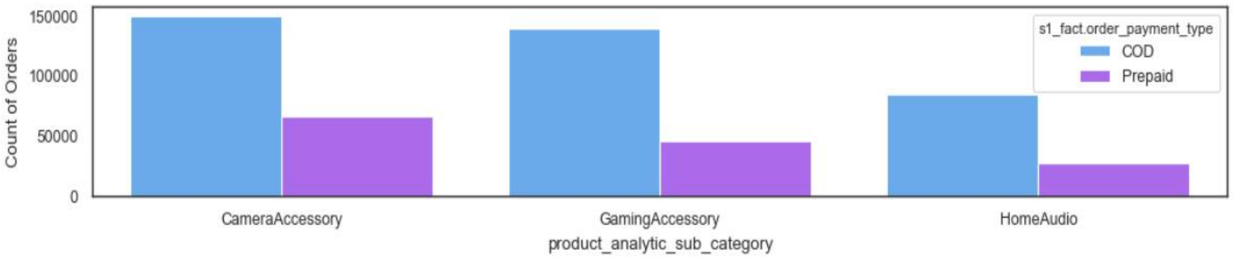
We see that the existing customers are clearly more valuable than the new customers. They tend to have a greater number of items in their cart and tend to have a greater cart value compared to the new customers.

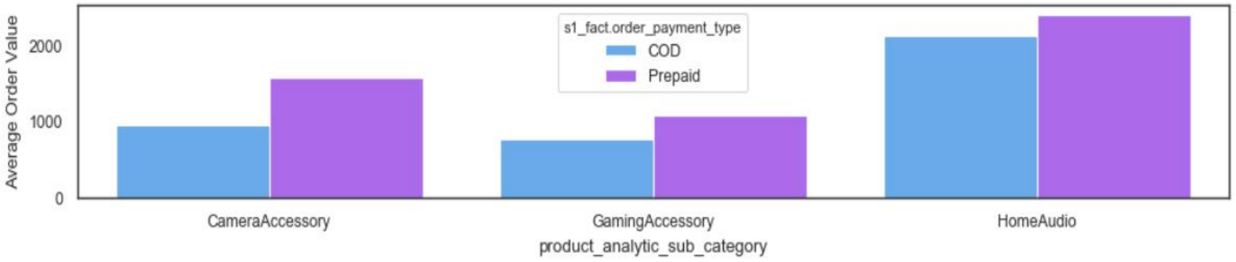
Business – Relevant Insights from EDA

- Highest Revenue generating and most sought after products in each of the category

Product Sub-Category	Highest Revenue Generating Product	Most Purchased Product
Camera Accessory	Lens	Flash
Gaming Accessory	Game Pad	Gaming Headset
Home Audio	Home Audio Speaker	Home Audio Speaker

- Looking at the trends in payment modes across the categories





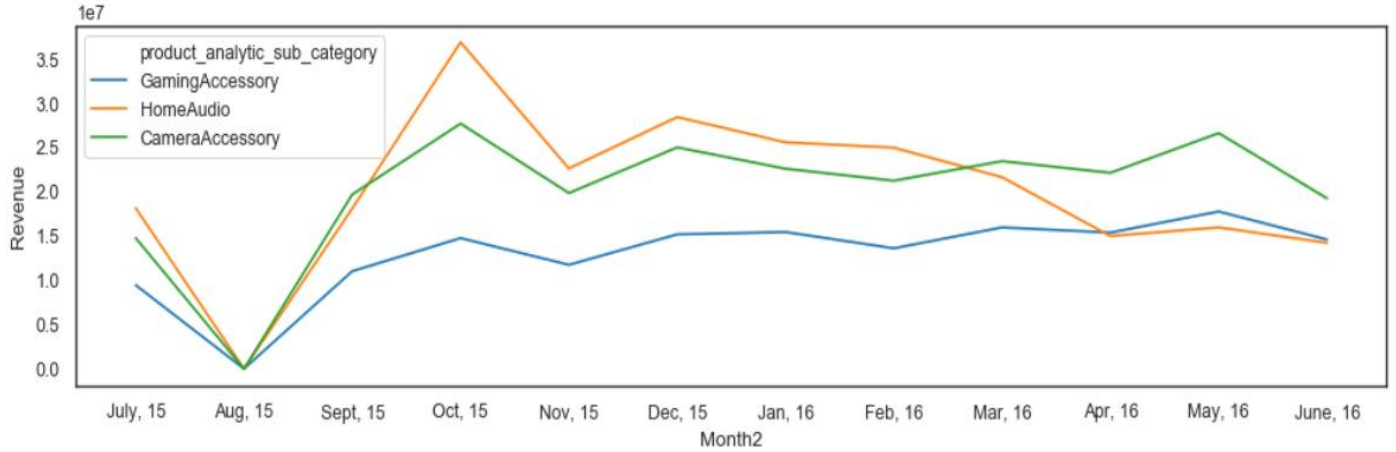
Interestingly, we see that customers mostly prefer to pay via cash on delivery method for all the categories.

However, when we add the cart value to the picture, we see that customers tend to prefer the prepaid method for orders which have a relatively high order value rather than the cash on delivery method

Business – Relevant Insights from EDA

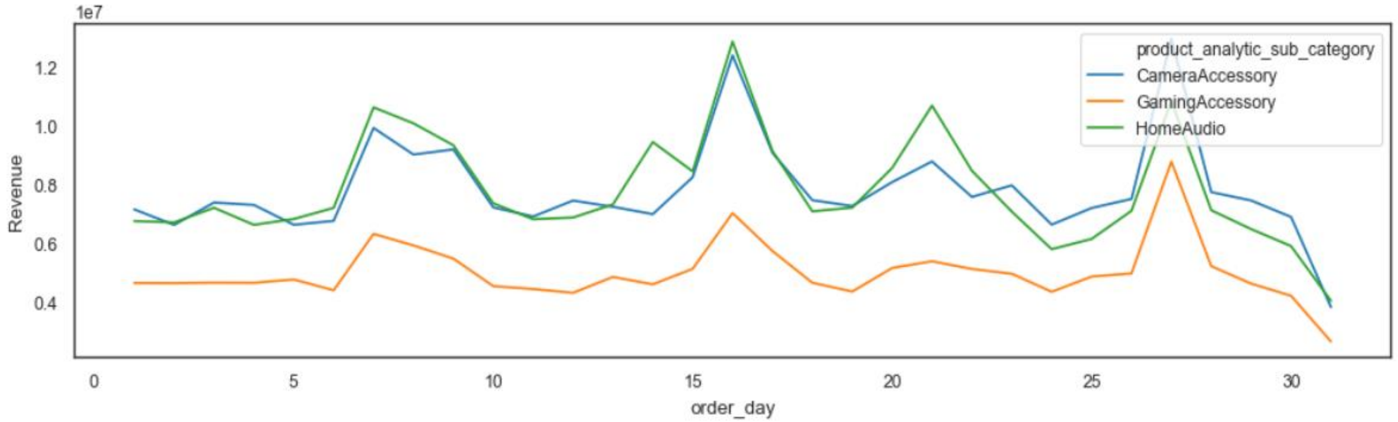
❑ Looking at the seasonal and temporal trends for sales across the year

* Trends by Month of Year



We see that August sees a massive dip in revenue numbers. However, this is followed by a significant increase in sales in the next couple of months. The likely reason for the surge in sales towards the last quarter of the year is because of the shopping & festival seasons around Black Friday, Christmas, Hanukkah, Halloween etc.

* Trends by day of month



We see that there is a spike in revenue at the middle of the month, which could be directly linked to paydays

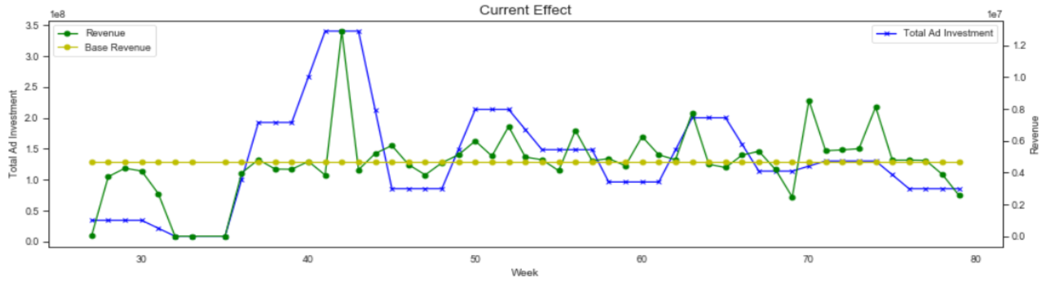
Final List of KPI's used for Modelling

All the datasets were merged and **aggregated at week level**. The final dataset was broadly split into 3 datasets, one for each category. The final list of KPI's that was considered for modelling include:

- Average discount percentage per week
- Average SLA per week
- Average product procurement SLA per week
- Total paydays per week
- Total special sale days per week
- Total holidays per week
- Total COD/Prepaid payments in the week
- Total mass-market/luxury products per week
- Total orders per week
- Total unique customers per week
- Total items bought per week
- Different products within the sub-category bought per week
- Total Investments across every channel per week
- Ad Stock values across each investment per week
- 3-week moving average for investments across every channel
- 5-week moving average for investments across every channel
- NPS Score
- 3-week NPS Score
- 5-week NPS Score
- Stock Index
- 3-week moving average for Stock Index
- 5-week moving average for Stock Index
- Min, Max and Mean Temperatures per week
- Total Rain (mm) per week
- Total Snow (cm) per week
- Aggregated Heat Deg Days (°C) per week
- Aggregated Cool Deg Days (°C) per week
- Total Snow on ground (cm) per week
- 1-Week, 2-Week, 3-Week Lag for each KPI (for the distributed lag model)
- 1-Week Lag for GMV (for the Koyck model)

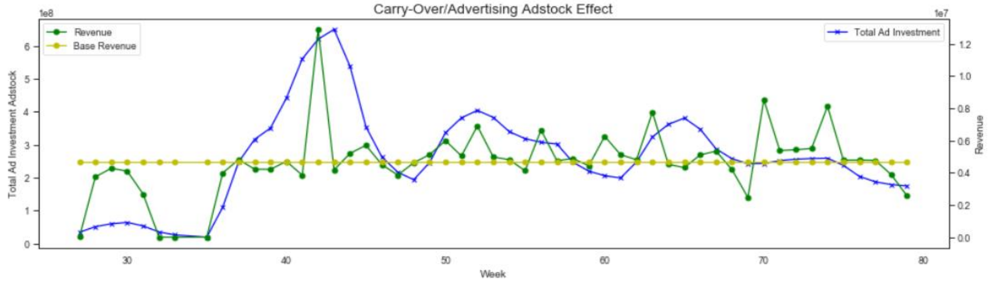
Impact of Advertising on Revenue

Current Effect: effect on revenue at the same time the ad is aired.



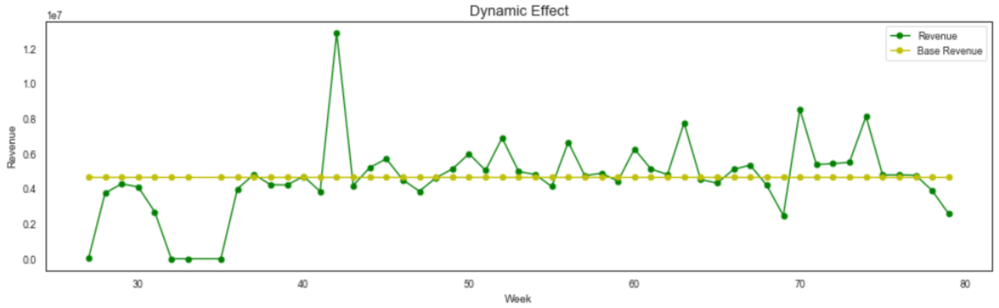
Correlation Coefficient: 0.54

Carry Over Effect: effect on revenue that follows even after the day of the ad exposure.

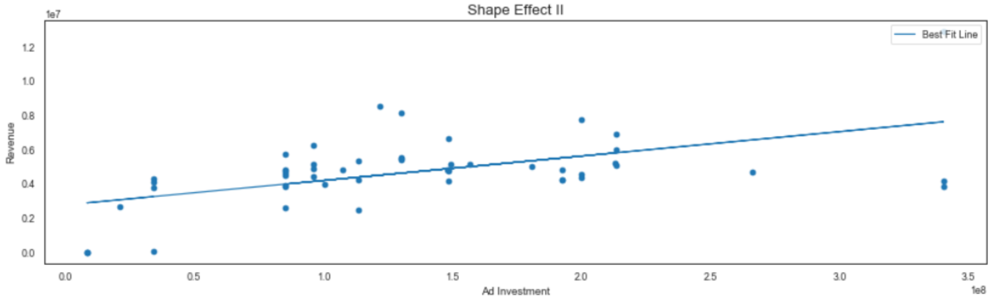


Correlation Coefficient: 0.57

Dynamic Effect: effect on revenue that changes with time. Broadly includes the wear-in effect, wear-out effect and hysterical effect.



Shape Effect: effect on revenue with increasing intensity of advertisement.



Market Mix Models Created for each Category

1. **Linear Additive Model**

   The linear model assumes that all the different KPI's are additive in nature and contribute linearly to the revenue generated by the company. The equation that the linear model follows is:

   $Y = \alpha + \beta 1At + \beta 2Pt + \beta 3Dt + \beta 4Qt + \beta 5Tt + \epsilon$

2. **Multiplicative Model**

   The linear model assumes that all the different KPI's are just additive in nature. However, this may not be entirely true. The multiplicative model assumes that the models have a multiplicative effect on each other. The equation that the model follows is:

   $\ln(Y) = \alpha + \beta 1\ln(At) + \beta 2\ln(Pt) + \beta 3\ln(Dt) + \beta 4\ln(Qt) + \beta 5\ln(Tt) + \epsilon'$

3. **Koyck Model**

   The Koyck model is a special type of distributed lag model that captures the carry over effect of the dependent variable too. The equation that the model follows is:

   $Y = \alpha + \mu Yt\text{-}1 + \beta 1At + \beta 2Pt + \beta 3Dt + \beta 4Qt + \beta 5Tt + \epsilon$

4. **Distributed Lag Model**

   As opposed to the Koyck model, not only is the dependent variable entered in its lagged version, but the independent variables are as well. This is a more generalist model and captures the carry-over effect of all the variables. The equation that the model follows is:

   Yt $= \alpha + \mu 1Yt\text{-}1 + \mu 2Yt\text{-}2 + \mu 3Yt\text{-}3 + ...$

   $+ \beta 1X1t\text{-}1 + \beta 1X1t\text{-}2 + \beta 1X1t\text{-}3 + ...$

   Where: μ, β are coefficients, and the other variables are different marketing KPI’s

Procedure for Creating the Linear Models

- Data Preparation (Depending on the type of model)
  - Log transform for all variables for Multiplicative Model
  - Add lag GMV variable for Koyck Model
  - Add Lags for every variable for the Distributed Lag Model
- Converting all the categorical variables to numerical variables
- Splitting the data into test and train (70-30 split)
- Scaling all the numerical features in the train set using standard scaling
- Next, **we build a vanilla model** considering all the features present first
- We observe that the basic models have very high multi-collinearity among the predictor variables
- We'll opt for **Recursive Feature Elimination** to come down to about a smaller number of variables that explain the highest variance in the data and exhibit the least multicollinearity
- Next, we fine tune the model by eliminating a variable 1 at a time based on the VIF and p-values to end up with the optimal number of features
- Next, we check that the assumptions of linear regression are held by performing residual analysis
- We run the final model on our test dataset to generate the predictions
- The model is then evaluated using the metrics:
  - RMSE
  - R-Squared
  - Normalized RMSE: RMSE ÷ Standard Deviation of the Distribution. The NRMSE is one way in which we can compare different models as this is scale independent
- Finally, in order to achieve the primary objective – to find the most important features/KPI's that impact revenue – we plot the coefficients of the model to judge the importance of the features

KPI Elasticities

We performed linear regression to identify the **elasticities of each KPI** (impact of 1 unit change in a particular KPI on the revenue for that product sub-category). The below charts show the 5 most (positive and negative) elastic KPI's for each category

Camera Accessory Sub-Category

Top 5 Positive Elastic KPIs
Total Luxury Products
Product Analytic Vertical – Lens
Product Analytic Vertical – Camera Housing
Product Analytic Vertical – Camera Mount
Product Analytic Vertical – Binoculars

Top 5 Negative Elastic KPIs
Max Temp (°C)
TV Investment
Product Analytic Vertical – Camera Battery Charger
Stock Index
Mean Discount %

Game Accessory Sub-Category

Top 5 Positive Elastic KPIs
Total Luxury Products
Product Analytic Vertical – Gaming Headset
Total Orders
Total Customers
Total Quantity Purchased

Top 5 Negative Elastic KPIs
Max Temp (°C)
Mean Discount %
Radio – 5-Week Moving Average
Sponsorship – 5-Week Moving Average
Mean Product Procurement SLA

Home Audio Sub-Category

Top 5 Positive Elastic KPIs
Product Analytic Vertical – Home Audio Speaker
Total Customers
Total Quantity Purchased
Total Orders
Product Analytic Vertical – Karaoke Player

Top 5 Negative Elastic KPIs
Online Marketing
Cool Deg Days (°C)
Affiliates
Content Marketing – 5-Week Moving Average
Product Analytic Vertical – FM Radio

Model Evaluation Summary

Summarizing the models created for each of the 3 datasets, along with their evaluation parameters. Highlighted in blue are the best models identified for each category based on the R-squared value (should be high) and the Normalized RMSE value (should be low).

Product Sub-category	Regression Model	R-squared on Test Dataset	Normalized RMSE on Test Dataset
Camera Accessory	Additive	0.96	0.22
	Multiplicative	**0.97**	**0.40**
	Koyck	0.92	0.31
	Distributive Lag Model	0.84	0.45
Game Accessory	**Additive**	**0.98**	**0.18**
	Multiplicative	0.94	0.33
	Koyck	0.92	0.32
	Distributive Lag Model	0.83	0.46
Home Audio	Additive	-0.17	0.79
	Multiplicative	**0.99**	**0.21**
	Koyck	-0.10	0.76
	Distributive Lag Model	0.60	0.48

The Top KPI’s for the Best Models

Let's narrow down on the best models identified within each category and see what are the top KPI's that contribute to the revenue for each category. The KPI's marked in Blue are those that affect the revenue positively, and those in Red are the ones that affect the revenue negatively for their respective product sub-category (they’re ranked according to importance – strength of coefficient)

Product Sub-category	Regression Model	R-squared on Test Dataset	Normalized RMSE on Test Dataset	KPI's Considered (Rank)
Camera Accessory	Multiplicative	0.97	0.40	Product Analytic Vertical - Binoculars (1)
				Product Analytic Vertical - Camera Accessory (2)
				Total Holidays (3)
				Stock Index (4)
				Mean SLA (5)
Game Accessory	Additive	0.98	0.18	Product Analytic Vertical - Game Pad (1)
				Product Analytic Vertical - Gaming Headset (2)
				TV - 3-Week Moving Average (3)
				Product Analytic Vertical - Gaming Mouse (4)
				Min Temp (°C) (5)
				Digital - Ad Stock (6)
Home Audio	Multiplicative	0.99	0.21	Total Customers (1)
				SEM - Ad Stock (2)
				Product Analytic Vertical - Docking Station (3)
				Radio - 3-Week Moving Average (4)
				Sponsorship (5)

Recommendations

Camera Accessory Sub-Category:

- From EDA, we identified Lens as the highest revenue generating product & Flash as the most sought after one. From the model, we see that Binoculars contribute significantly to the revenue. ElecKart **should look to market these products more** in the coming year.
- We see that holidays and average SLA contribute negatively to the revenue. What this means is that **consumers prefer purchasing products in this vertical during non-holidays**. Also, the company should ensure that they **focus on delivering the products in this category with a small SLA** to the consumers.
- **'Mass-market' products are better contributors** to the increased revenue in comparison to the Luxury products.
- The company should **give higher percentage of 'Discounts'** for this category as from the EDA we see that majority of the consumers look forward to use discount coupons for camera accessory products.

Gaming Accessory Sub-Category:

- Company should **promote 'Game Pad', 'Gaming Headset' & 'Gaming Mouse'** as these products are the ones that significantly contribute to the revenue as per our final additive model.
- We observe that there is a positive impact of TV advertisement (3-week Moving Average), and **they should invest a significant amount in TV advertising** in the coming year.
- From our model, the 'Min Temp (°C)' is a significant factor in the revenue for this category. A positive impact of this means that higher min temperatures result in higher revenue, and ElecKart can use this information to **shift advertising and promotional spends to times with higher temperatures**.
- As opposed to TV Ad, the Digital Ad Stock tend to have a negative impact on the revenue. ElecKart can look to **shift some of the ad investments out of digital, into other channels.**
- From the EDA, we see that **'Mass-market' products are better contributors** to the increased revenue in comparison to the Luxury products.

Recommendations

Home Audio Sub-Category:

- ElecKart should **promote 'Docking Stations'** as this product appears to be contributing most significantly to the revenue.
- We observe several Marketing levers coming up as impactful KPI's for this category. From the multiplicative model, we can infer that SEM (Ad Stock) contributes positively to the revenue generated as opposed to Radio (3-Week Moving Average) & Sponsorship, which affect the revenue negatively. ElecKart can look to re-align their marketing budget for next year by **shifting spends to the SEM strategy**.
- The total customers play a significant part in generating revenue for this category according to our model. As part of the company's marketing strategy for next year, they should **look to drive more traffic/footfall to their website/stores** as this would have a positive impact on the revenue.
- The category has products which form the pricier section of the company. From the EDA, we see that the customers prefer paying for high-value products using prepaid modes rather than COD. ElecKart can **use prepaid promotional techniques (such as cashbacks, prepaid payment mode discounts) to encourage more people to purchase these products**.

In General:

- During festive time(e.g. Black Friday, Halloween, Christmas, Hanukkah) there are clear surges in sales. It is evident from the data that this forms the shopping season, and the company should **focus having their sale dates around these festive seasons** to garner more revenue.
- From EDA, we see that there is a clear spike in sales mid-month (potential pay-date). ElecKart should **ensure that they're stocking up their inventory** to prepare for the rise in sales at this time so that they're not missing out on revenue due to lack of inventory.
- From the EDA, we see that existing customers are clearly more valuable to ElecKart than any new customers (both in terms of items purchased and cart value). The company should ensure that they **retain their loyal customers by providing additional benefits** such as loyalty benefits, discounts with every subsequent purchase.
- ElecKart can also **introduce their own wallet** which can be used just to purchase their products - with added discounts. This would be another strategy to prevent existing customers from churning.
- The company should **advertise for different categories of products differently**. From the models we created, we figured that **different categories require different channels for advertising**. The advertising budget for next year should be created keeping in mind the split for each product category.