ElecKart
Market Mix Model

Capstone Project Submission By –
Shruti Diwakar
Swapnik Chimalamarri
Tarunay Roy

- **Business and Data Understanding**
- EDA and Feature Engineering
- Model Results
- Business Recommendations
- Appendix

Business Objective

- Optimize the marketing budget allocation across different levers (commercials, online campaigns, and pricing & promotion strategies)
- Maximize revenue response in 3 product categories – Camera Accessories, Game Accessories and Home Audio
- Identify most impactful marketing channels to invest in and get maximum return on investment in terms of revenue response

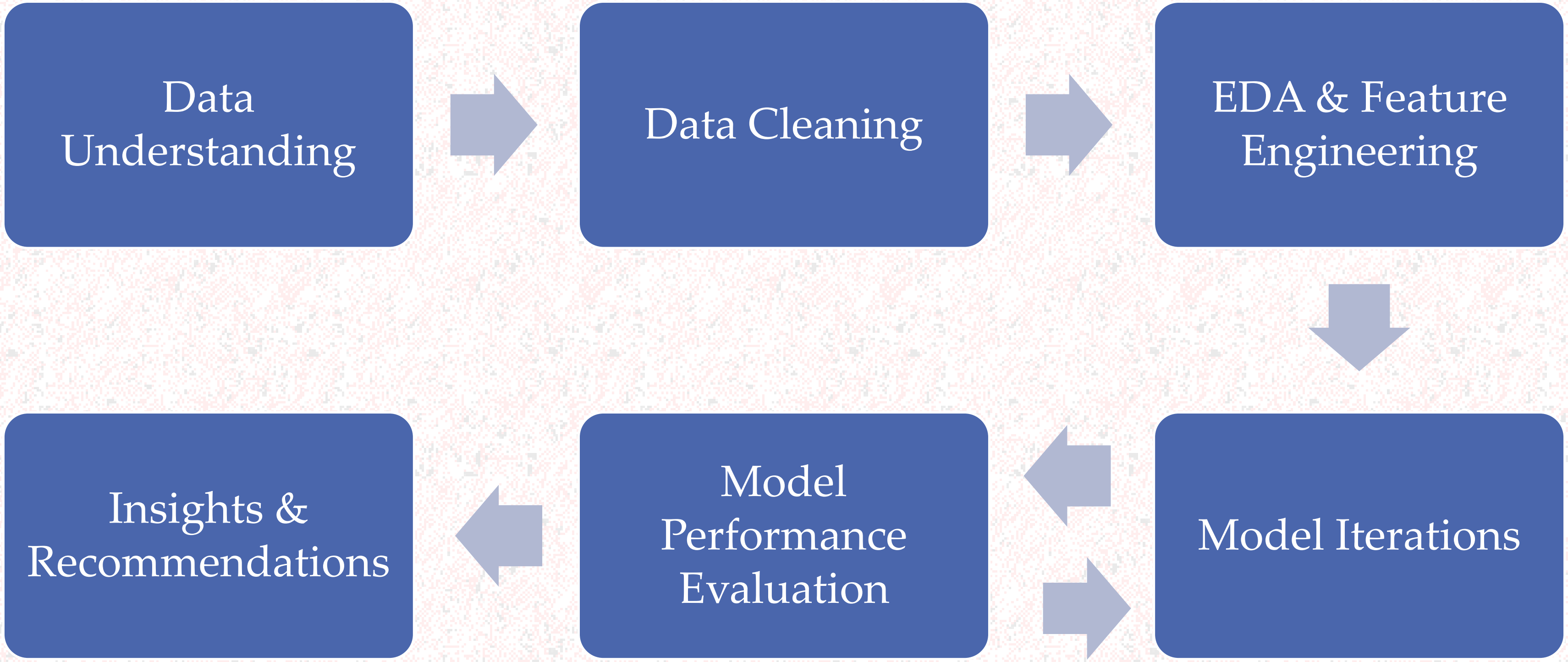
Data Understanding

- Order-Level data has been provided for 1 year, i.e. July 2015 – June 2016
- The monthly spends on various advertising channels are available in Crores
- Special Sale Days on the e-commerce platform have been provided
- The SLA and Procurement SLA at order level gives us the delivery time promised while ordering
- Monthly NPS scores has been provided to understand customer experience and perception of the company

Assumptions

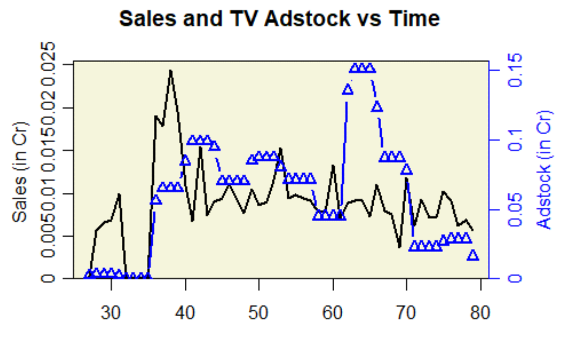
- Media advertising spend is distributed equally among
  - All 14 product categories
  - All weeks in the month
- Days when no sales were made have been ignored while rolling up data to weekly level
- ‘Frequency’ provided in the Media Data file has been understood as the number of times an average household or person is exposed to the schedule among the persons reached in the specific period of time
- ‘Percent’ provided in the Media Data file has been understood as Reach, i.e. the percentage of different homes or persons exposed at least once to an advertising schedule over a specific period of time. It has been assumed to exclude duplication
- These have been used to calculate the Gross Rating Points (GRP)

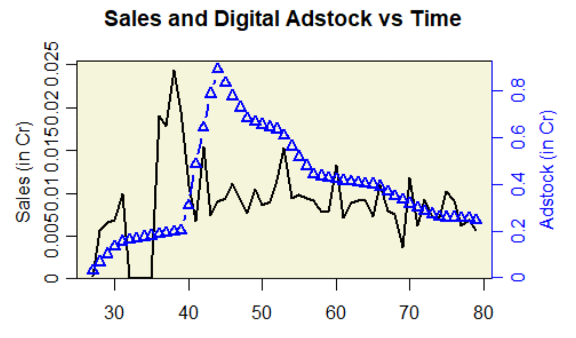
Steps for Solving Business Problem

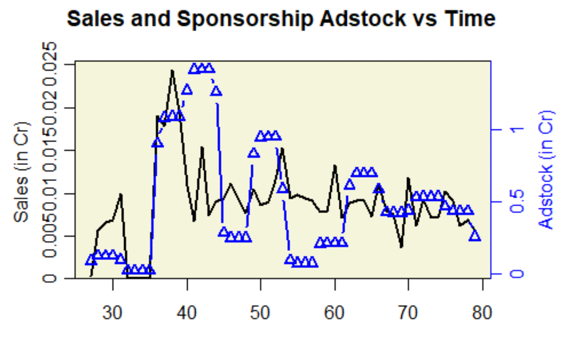


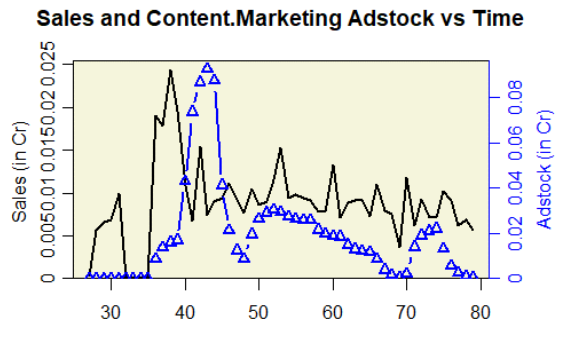
- Business and Data Understanding
- **EDA and Feature Engineering**
- Model Results
- Business Recommendations
- Appendix

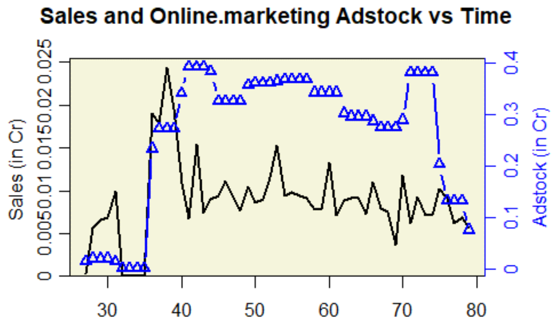
Gaming Accessories: EDA of Ad-stock vs. Time

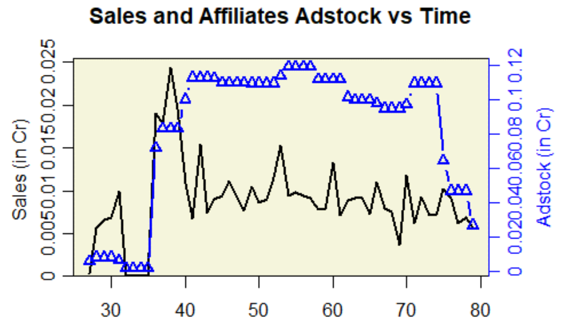


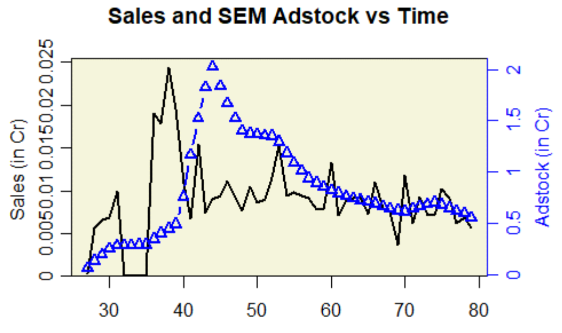










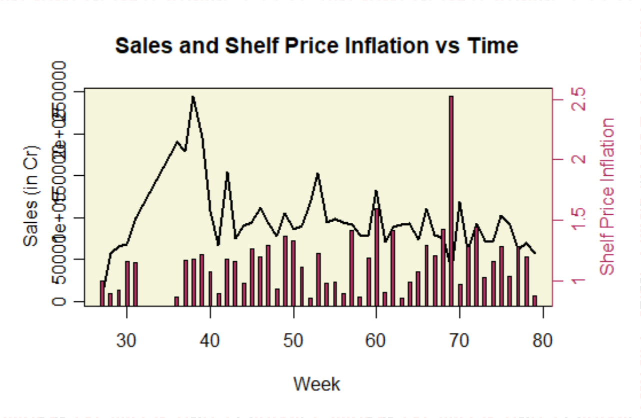


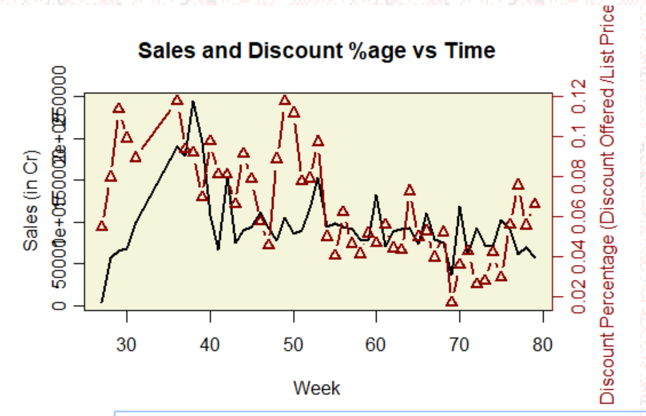
- **Inference**: We see that Sponsorship Ad stock and Online Marketing peaks are corresponded by Sale peaks, which shows that it is effecting the sales the most of all channels.
- The Affiliate Channel has been spent on heavily, but it is not having the desired effects on Gaming Accessory sales

Note: We have added similar visualizations for Home Audio and Camera Accessories in the Appendix.

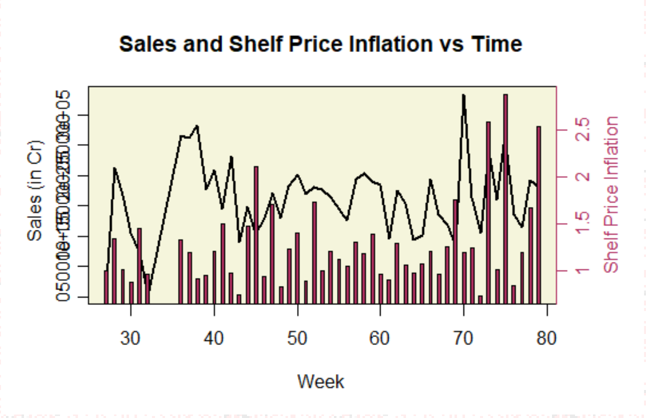
EDA of Shelf Price Inflation, Discount and Sales vs. Time

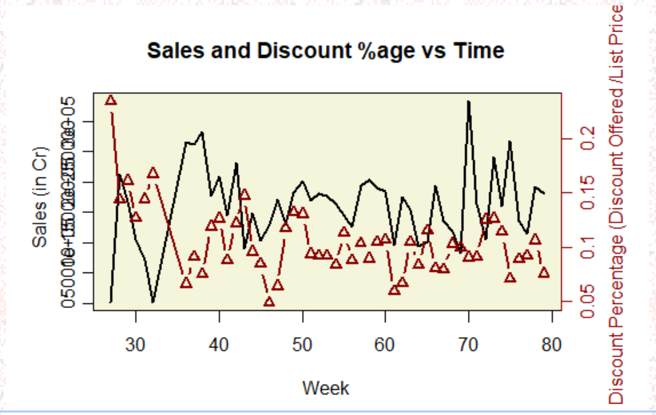
Gaming Accessories



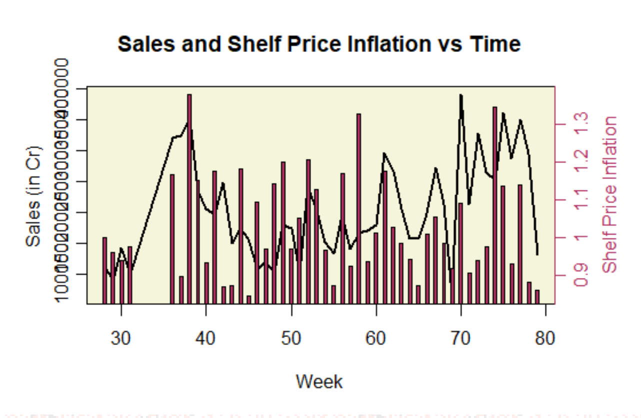


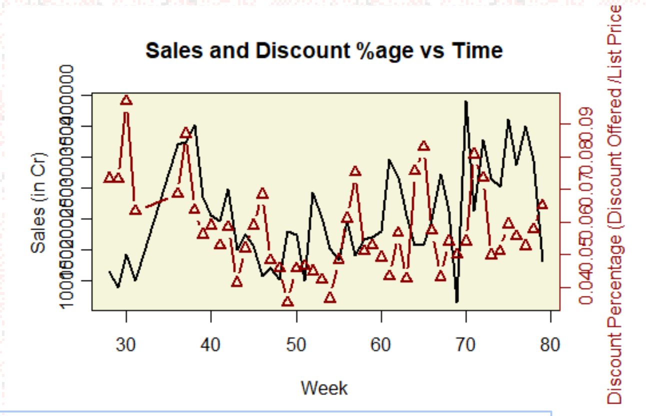
Camera Accessories





Home Audio

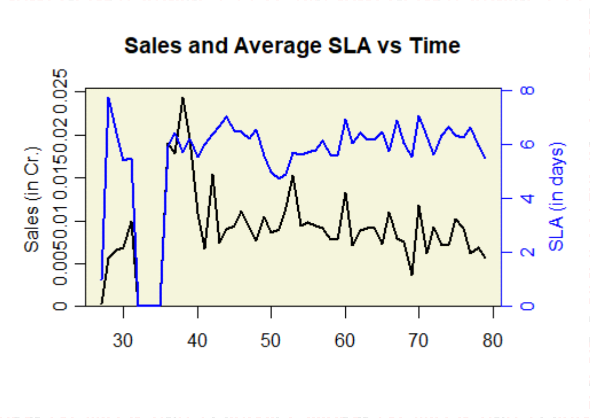


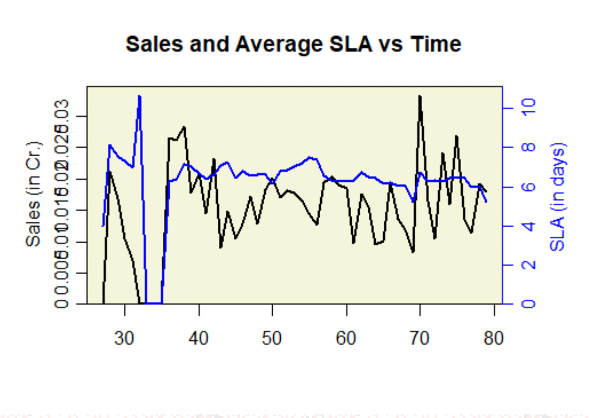


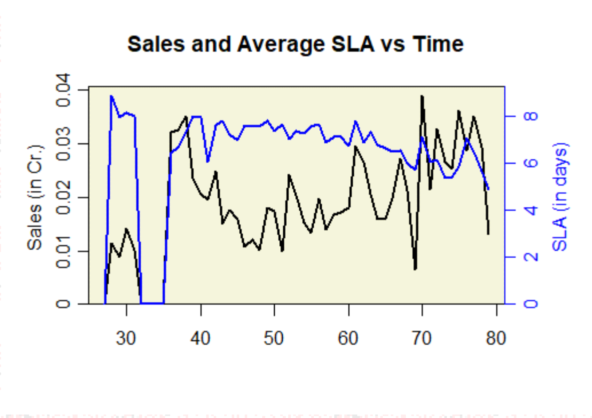
Inference: We see that as Shelf Price Inflation rises, the sales drop. Similarly when discount %age rises the sales increase. This shows the customers shopping on online platforms are sensitive to cost, and like discounts more

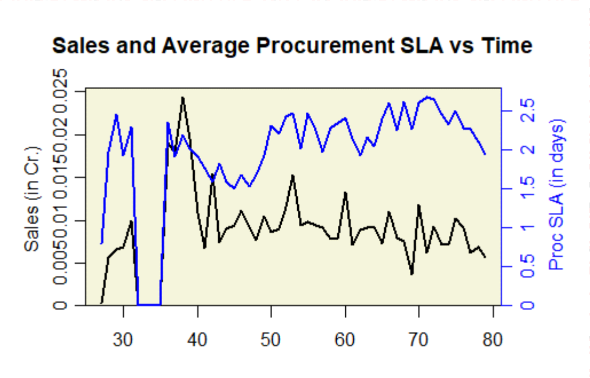
EDA of Sales vs Delivery and Procurement SLA over Time

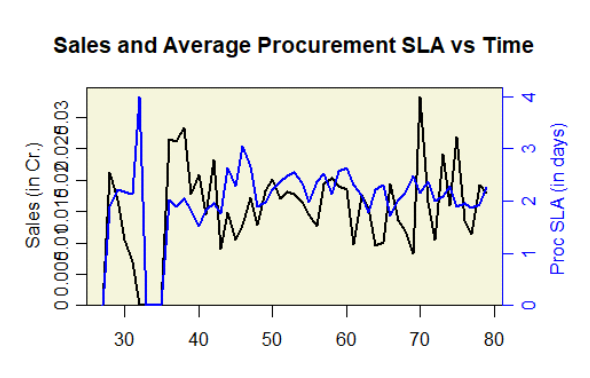
Gaming Accessories	Camera Accessories	Home Audio

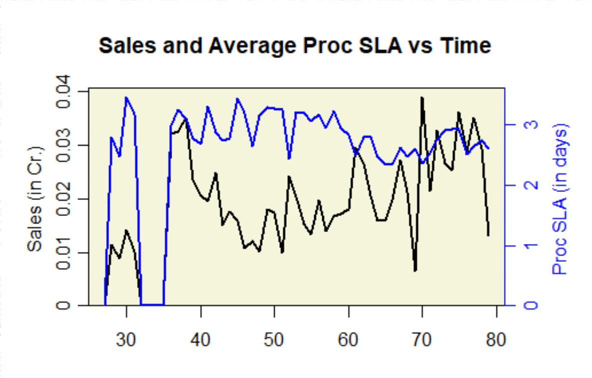












Inference: We see for Camera Accessories and Home Audio categories when procurement SLA is lesser, the sales are higher. Similarly when the Delivery SLA is lesser the sales are higher. However for Gaming Accessories we see that during peak sales, the SLA was high. Gaming accessories customers do not seem to be sensitive to SLA.

Engineered KPIs (weekly level)

Engineered Features	Definitions
TV_Adstock	TV Adstocks
Digital_Adstock	Digital Adstocks
Spons_Adstock	Sponsorship Adstocks
Cont_Adstock	Content Adstocks
Online_Marketing_Adstock	Online Marketing Adstocks
Aff_Adstock	Affiliate Adstocks
SEM_Adstock	SEM Adstocks
shelfprice_mean	Shelf Price Inflation
discounted_perc	% of orders which are discounted
listing_price	Avg. List Price
ma2_shelfprice_mean	Moving Average of past two weeks SPI
ma3_shelfprice_mean	Moving Average of past three weeks SPI
perc_prepaid	% of prepaid orders in a week
perc_sale	%of days in a week having a sale
avg_sla	Avg. delivery SLA
avg_procurement_sla	Avg. procurement SLA
perc_less_than_3dayssla	% of orders less than 3 days SLA
avg_grp	Gross Rating Points
mean_nps	Avg. NPS

Gaming Accessories – Model Results

Key Takeaways –

- We have selected **Multiplicative Model** over the combination of Multiplicative and Distributed Lag Model because the latter has only 3 important variables and hence business can take fewer actions on them

- Other models are ruled out due to low adjusted R squared and/or high MS from Cross Validation

- Adjusted R squared are from models run on training data and the MS figures are from 10-fold Cross Validation on the same data

- Hence, we select the Multiplicative Model to derive business insights for the Gaming Accessories BU

Model Type	Selected Variables	Adj. R Square	10-Fold CV (MS)
Simple Linear Model	spons_adstock + listing_price + perc_prepaid	0.4682	0.599
Multiplicative Model	Cont_Adstock + discounted_perc + listing_price + avg_grp + avg_sla + avg_procurement_sla	0.766	0.348
Distributed Lag Model	TV_Adstock + SEM_Adstock + listing_price + perc_prepaid + perc_sale + `sales-1`	0.454	0.606
Koyck Model	TV_Adstock + Digital_Adstock + listing_price + perc_prepaid + perc_sale + avg_grp + `sales-1` + `sales-3`	0.6016	0.598
Multiplicative Model + Distributive	perc_prepaid + avg_procurement_sla + perc_less_than_3dayssla + mean_nps	0.779	0.13

Camera Accessories – Model Results

Key Takeaways –

- We have selected Multiplicative Model due to the best balance of Adjusted R squared being the highest and Cross Validation MS being the lowest. At the same time, we have a large number of significant variables that the business can act upon
- Other models are ruled out due to low adjusted R squared and/or high MS from Cross Validation
- Adjusted R squared are from models run on training data and the MS figures are from 10-fold Cross Validation on the same data
- Hence, we select the Multiplicative Model to derive business insights for the Gaming Accessories BU

Model Type	Selected Variables	Adj. R Square	10-Fold CV (MS)
Simple Linear Model	Digital_Adstock + Spons_Adstock + shelfprice_mean + discounted_perc + listing_price + avg_procurement_sla + avg_grp	0.779	0.263
Multiplicative Model	TV_Adstock + Spons_Adstock + Cont_Adstock + SEM_Adstock + shelfprice_mean + listing_price + perc_prepaid + avg_sla + avg_grp + mean_nps	0.893	0.285
Distributed Lag Model	Digital_Adstock + Spons_Adstock + listing_price + avg_procurement_sla + avg_grp,data=dist_model_came ra	0.693	0.341
Koyck Model	Spons_Adstock + SEM_Adstock + discounted_perc + listing_price + avg_grp	0.667	0.426
Multiplicative Model + Distributive	avg_procurement_sla + mean_nps + \`sales-3\` + \`listing_price-1\`	0.735	0.555

Home Audio – Model Results

Key Takeaways –

- We have selected Simple Linear Model over the combination of Multiplicative and Distributed Lag Model because the latter has only 3 important variables and hence business can take fewer actions on them. Simple Linear Model is the second best performing model in this case
- Other models are ruled out due to low adjusted R squared and/or high MS from Cross Validation
- Adjusted R squared are from models run on training data and the MS figures are from 10-fold Cross Validation on the same data
- Hence, we select the Multiplicative Model to derive business insights for the Gaming Accessories BU

Model Type	Selected Variables	Adj. R Square	10-Fold CV (MS)
Simple Linear Model	TV_Adstock + Digital_Adstock + Spons_Adstock + Cont_Adstock + Aff_Adstock + listing_price + ma3_shelfprice_mean	0.745	0.299
Multiplicative Model	Digital_Adstock + Spons_Adstock + Cont_Adstock + listing_price	0.589	0.464
Distributed Lag Model	Digital_Adstock + Spons_Adstock + Aff_Adstock + listing_price + ma3_shelfprice_mean	0.676	0.375
Koyck Model	Digital_Adstock + Spons_Adstock + listing_price + ma3_shelfprice_mean	0.704	0.35
Multiplicative Model + Distributive	Digital_Adstock + Spons_Adstock + avg_sla	0.968	0.0366

- Business and Data Understanding
- EDA and Feature Engineering
- Model Results
- **Business Recommendations**
- Appendix

Gaming Accessories: Business Recommendations

- For Gaming Accessory, ElecKart should focus on Content Advertising Channel the most. For every 1% increase in Content Advertising spends, the sales increase by 0.35%

- ElecKart should also focus on increasing the exposure to Ad channels (GRP)

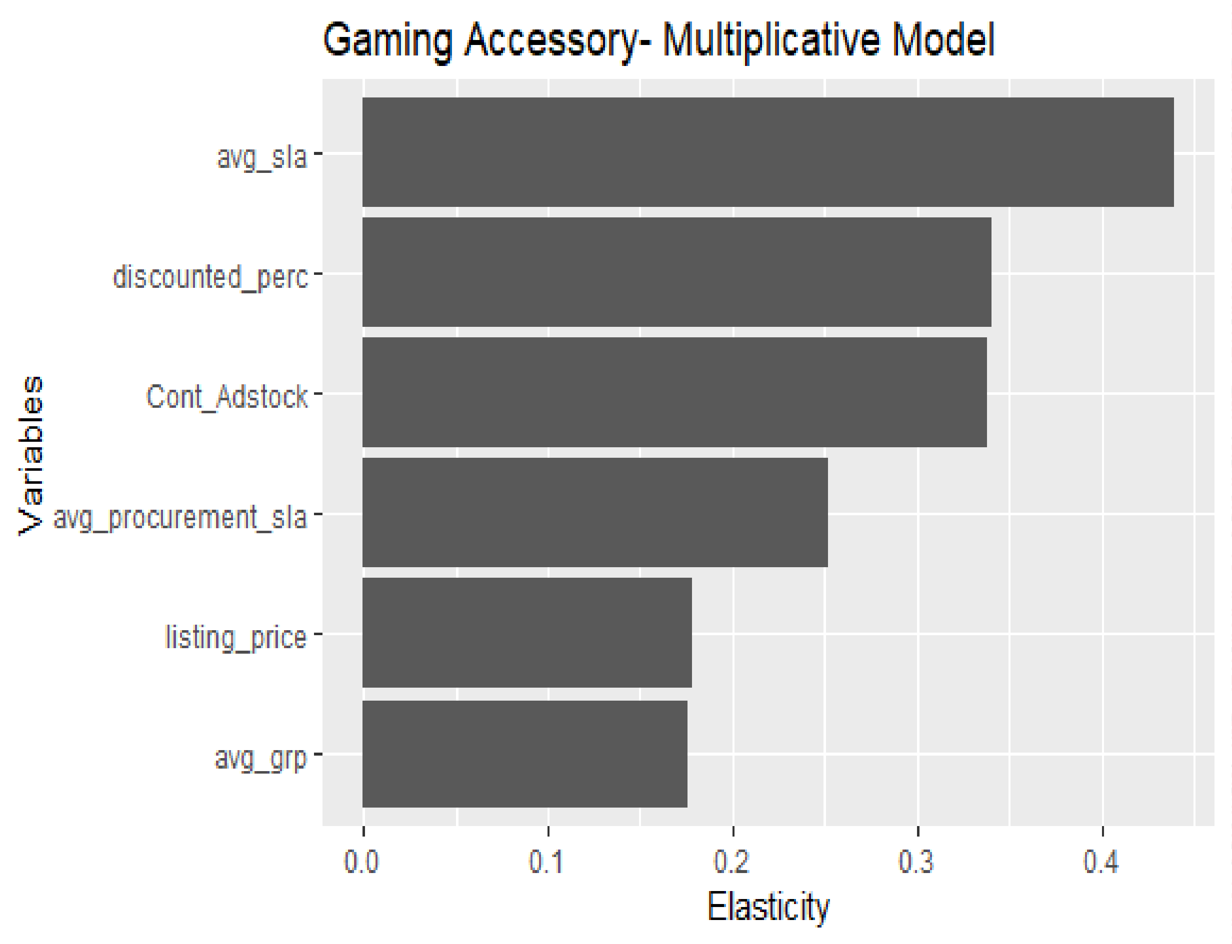
- The category users prefer discounts on high listing price products: ElecKart should focus on increase discounted products in the category

Note -

- Recommendations are based on the elasticity of KPIs shown in the adjoining figure

- Positive elasticity implies that increasing spends for those KPIs will lead to increase in sales

- Conversely, negative elasticity means that increasing spends on those KPIs would have a negative impact on sales and thus spends should be reduced on them



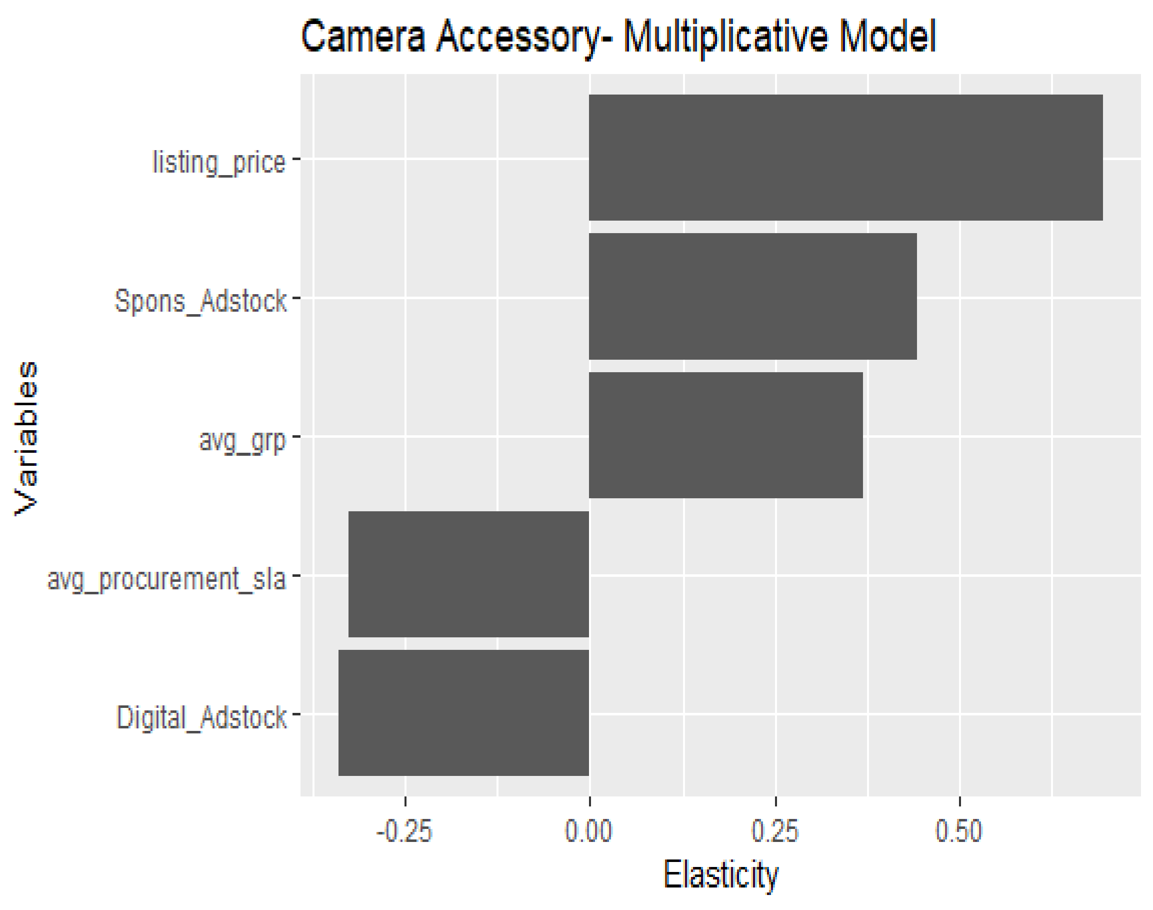
Camera Accessories: Business Recommendations

- Using the multiplicative model results, we can see that for Camera Accessories, ElecKart should invest in Sponsorship Advertising Channel and on the exposure to the channel (GRP)

- For every 1% increase in spends on Sponsorship Adstock, the sales increases by 0.40%

- ElecKart should decrease its spends on Digital Marketing, since it is not helping increase sales

- ElecKart should also focus on decreasing the Procurement SLA, to increase sales.

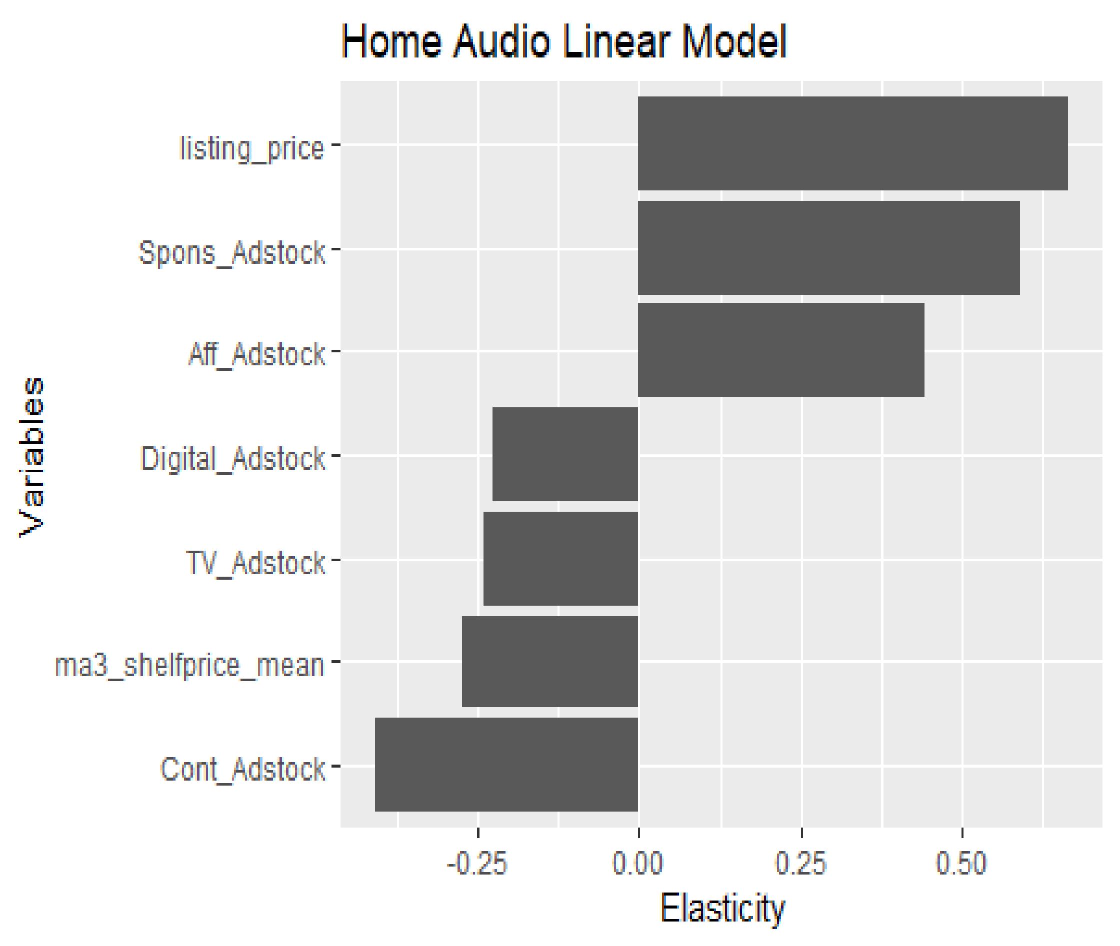


Home Audio: Business Recommendations

- Using the linear model results, we can see that for Home Audio, ElecKart should invest in Sponsorship Advertising Channel and on Affiliate Marketing Channels.

- ElecKart should decrease its spends on Digital Marketing,TV Advertisement and Content Marketing Channel, since it is not helping increase sales

- ElecKart should also focus on decreasing the Shelf Price Inflation in general



- Business and Data Understanding
- EDA and Feature Engineering
- Model Results
- Business Recommendations
- **Appendix**

In-Depth View on Adstocks Calculation (1/2)

Calculation of Custom Adstock Rates

- We have been provided Adstock numbers at a monthly level, across all sub-categories
- We have distributed them evenly across all days where sales have been made and then aggregated them into weeks
- We have then distributed it evenly across all 14 sub-categories
- Next, we have calculated the Adstock rate in R by setting up and optimizing the function Predicted Sales = $\alpha + \beta$ * adstock(Advertising). We minimize the sum of squared errors for the regression formula by changing Adstock rate between 0 and 1.

In-Depth View on Adstocks Calculation (2/2)

Calculation of Adstocks

- Once we obtained the Adstock rate, we used it to calculate the Adstocks for every channel at a weekly level for each of the subcategories in R
- This was done by considering the current week's Adstock and adding it with the previous weeks' Adstocks after applying the corresponding Adstock rate based on the recency
- This ensured that we took into account the memory retention factor of the audience from previous weeks
- It also shows us the effects of the campaigns if nothing was done this week and all we had to rely on was the impact of the previous campaigns

In-Depth View of Shelf Price Inflation calculation

- First, the list price was calculated by dividing GMV/Units
- Next, the list price was added up for every product within each week to bring the total list price at a product + week level
- The Shelf Price inflation was calculated at a product + week level by dividing product list price of week i by product list price of week (i-1)
- Finally, the Shelf Price Inflation is calculated at a sub-category + week level by taking a weighted mean of all the products' Shelf Price Inflation obtained in the above step, using the number of units of each product as weights
- This is the final Shelf Price Inflation obtained for each Sub-Category for each week

Order List Price = GMV/Units *(at an order level)*

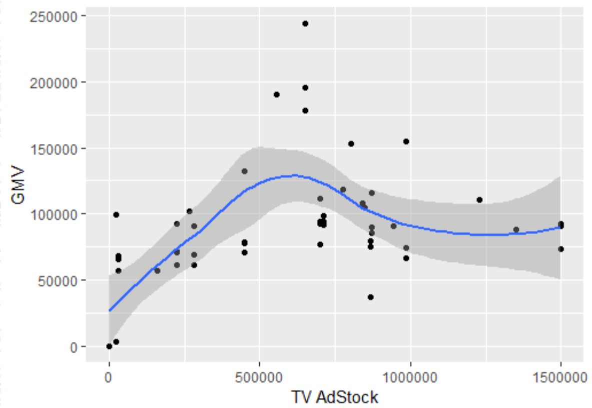
Product List Price = ∑Order List Price *(at a product + week level)*

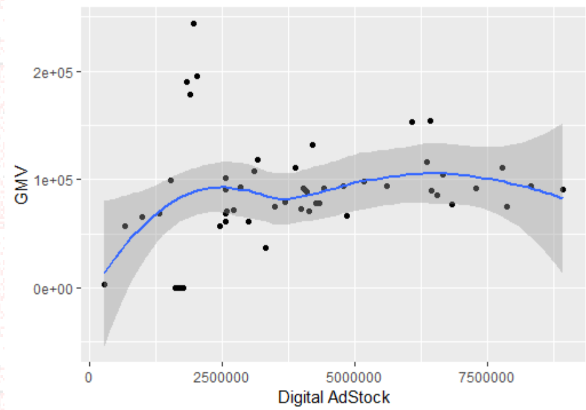
Product Shelf Price Inflation = $(\text{Product List Price})_{\text{week } i} / (\text{Product List Price})_{\text{week } i-1}$ *(at a product + week level)*

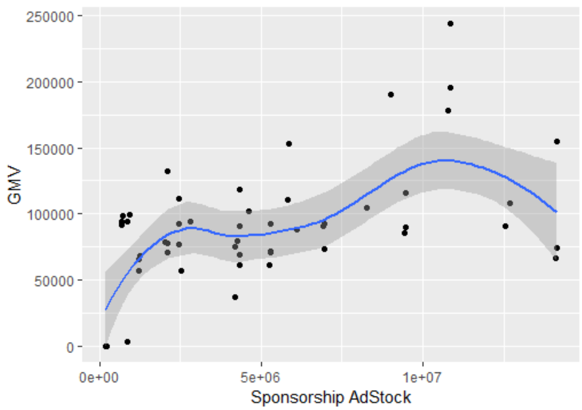
Sub-Category Shelf Price Inflation = ∑(Product Shelf Price Inflation * No. of Units)/ ∑No. of Units

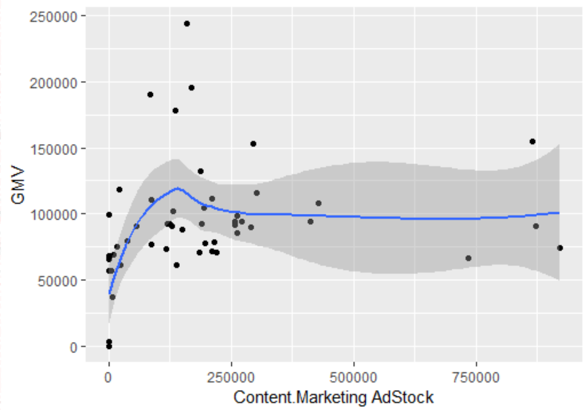
(Sub-Category + Week Level)

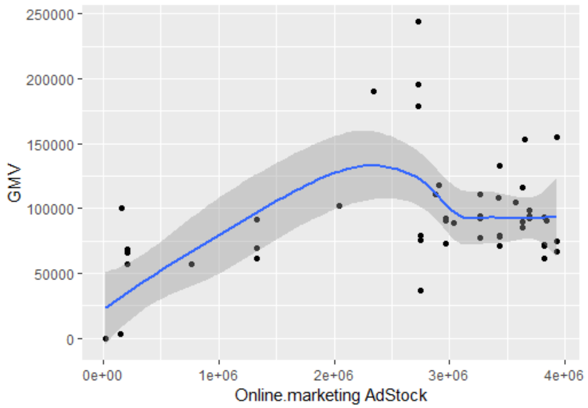
GMV vs. Adstock EDA

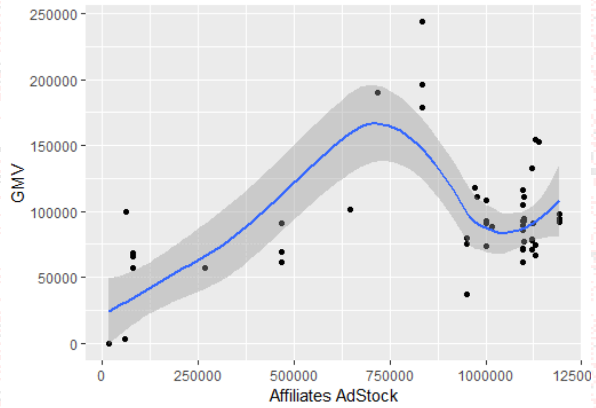


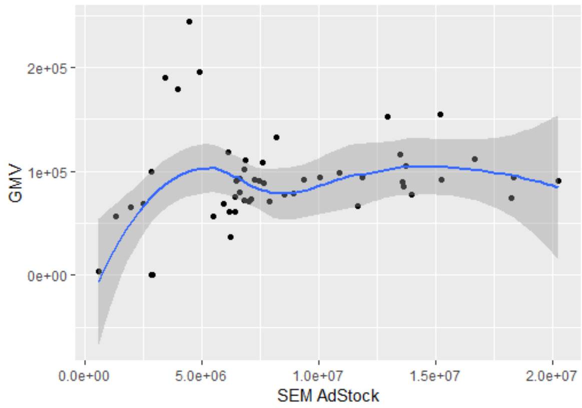










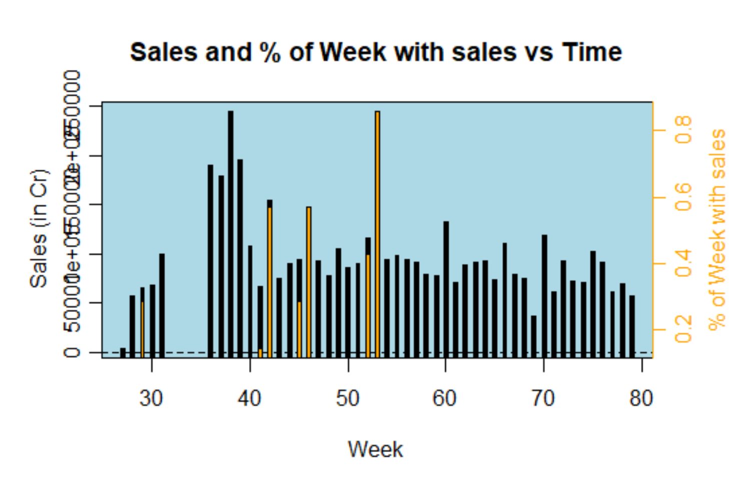


Inference: We see TV adstock, Sponsorship adstock, and Online marketing adstock form a possible correlation. Correlation with other adstocks is fairly flat.

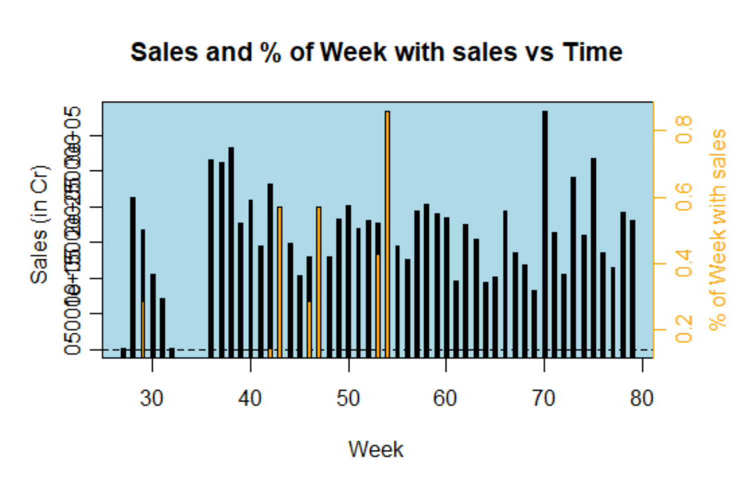
Note: We have presented different visualizations for Gaming Accessories Sub-Category here and added similar visualizations for Home Audio and Camera Accessories in the Appendix.

Sales and %age of week having Sales Events vs. Time EDA

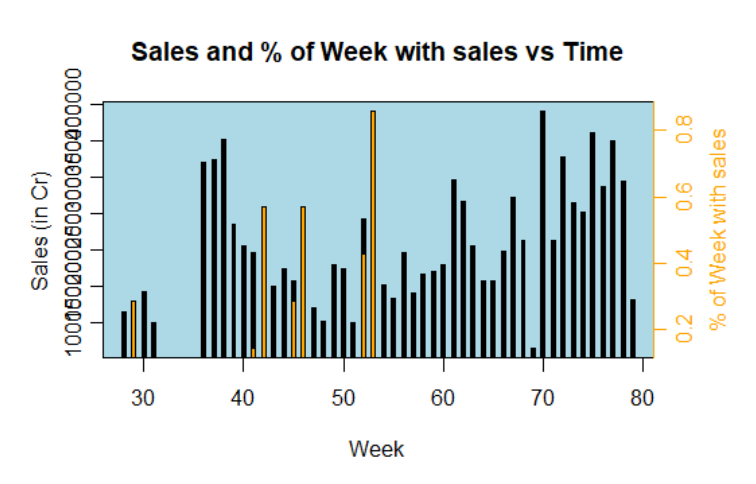
Gaming Accessories



Camera Accessories



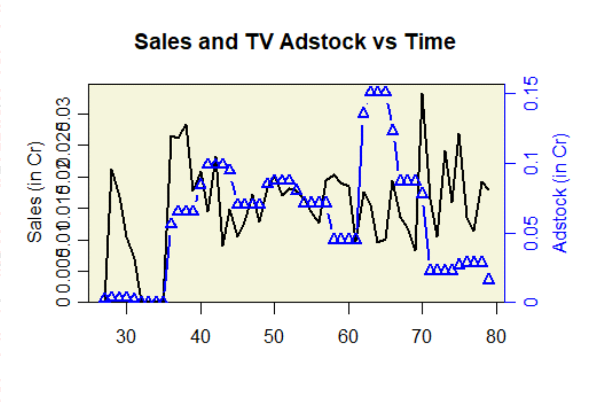
Home Audio

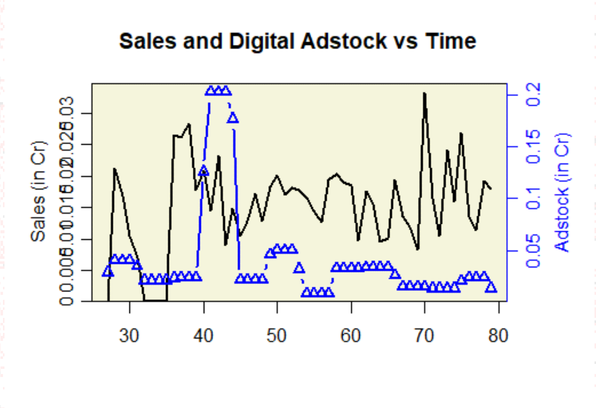


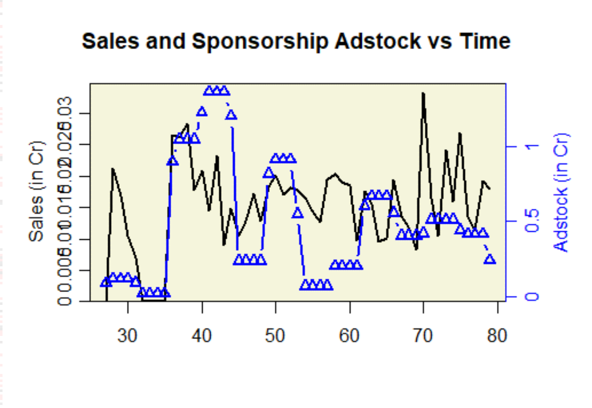
Inference: We can see the Sales peaking during the sale seasons, again re-enforcing that the customers are sensitive to price and like discounts.

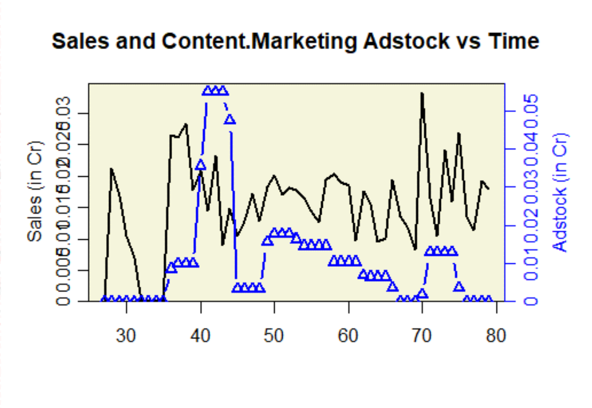
Visualizing Sales and Adstocks vs Time

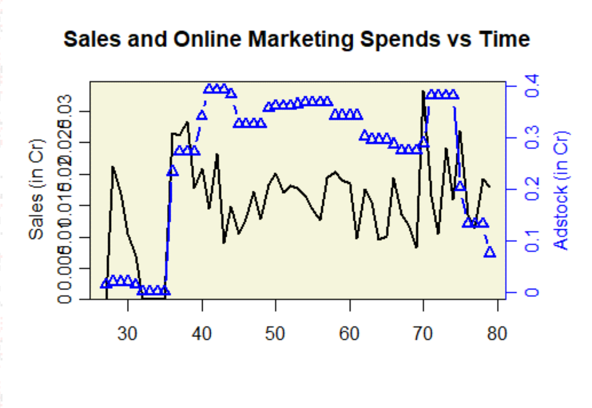
Camera Accessories Visualizations

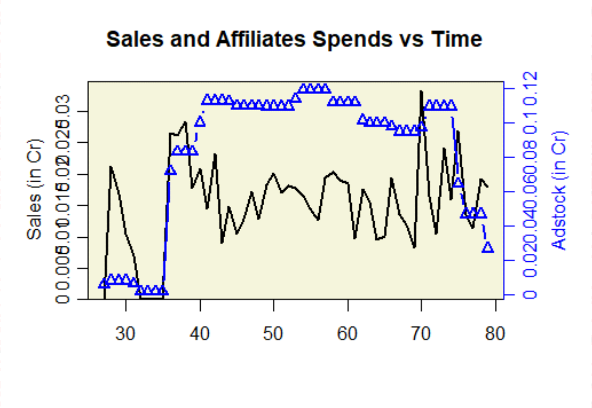


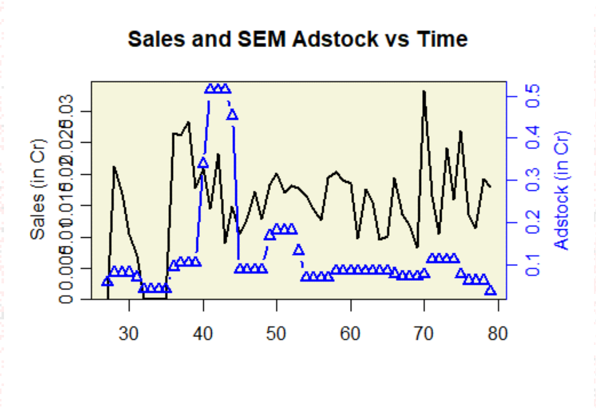






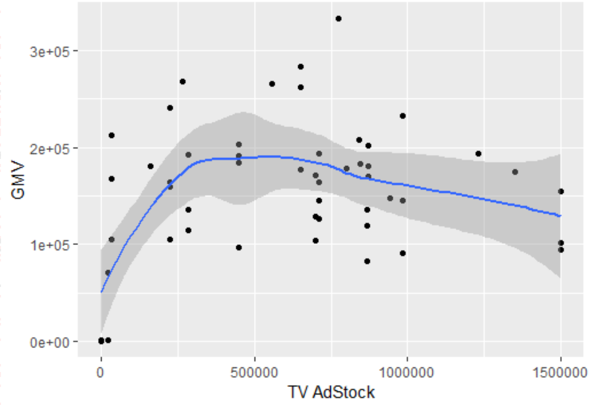


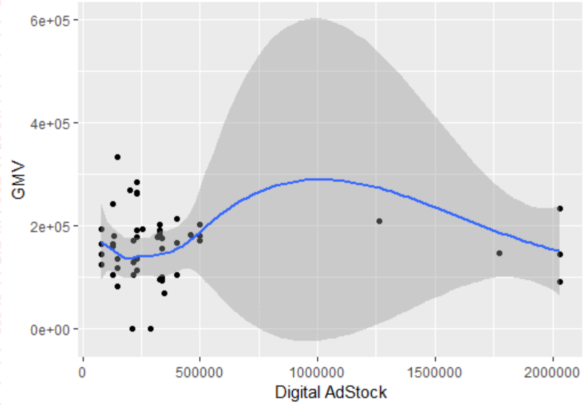


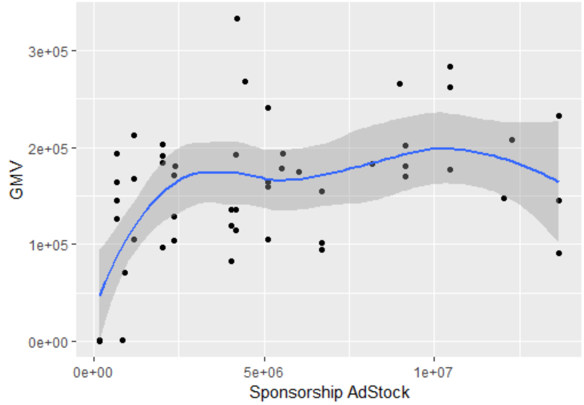


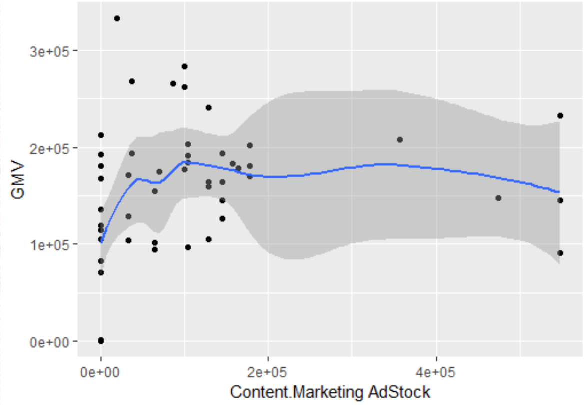
Visualizing GMV vs. Adstock

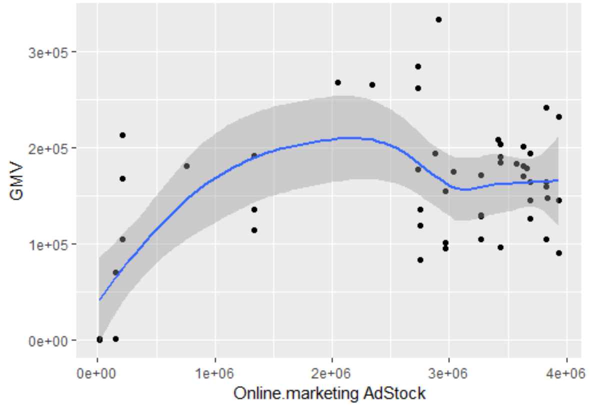
Camera Accessories Visualizations

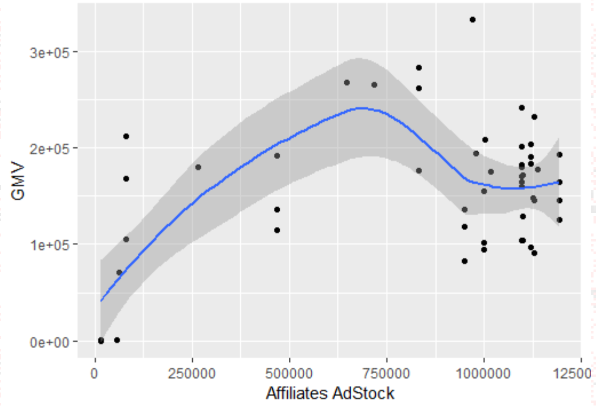


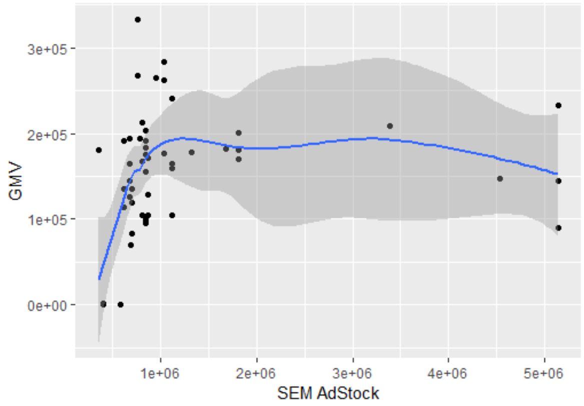






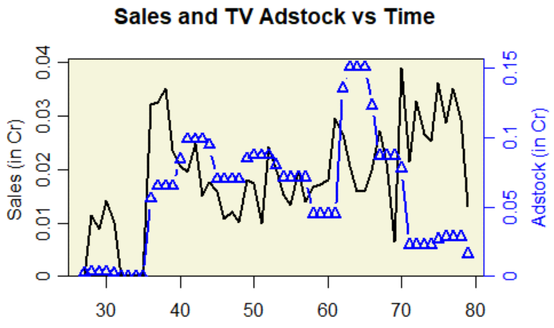


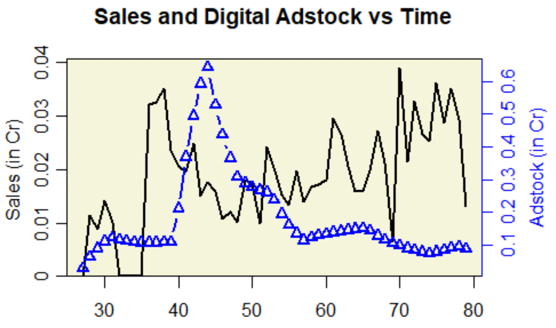


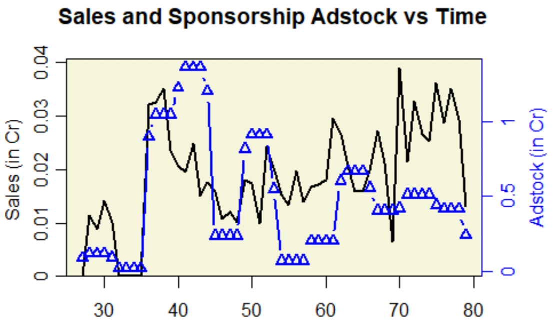


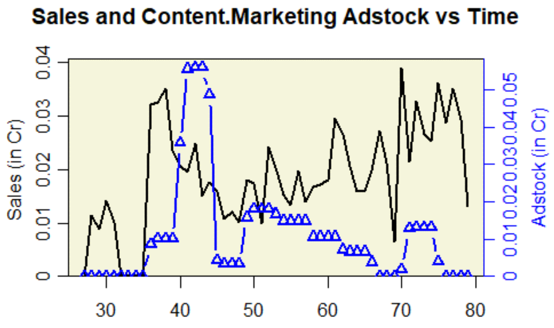
Visualizing Sales and Adstocks vs Time

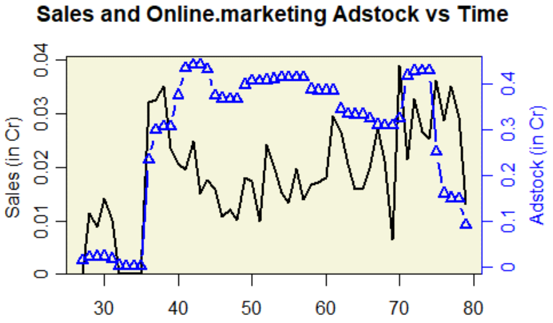
Home Audio Visualizations

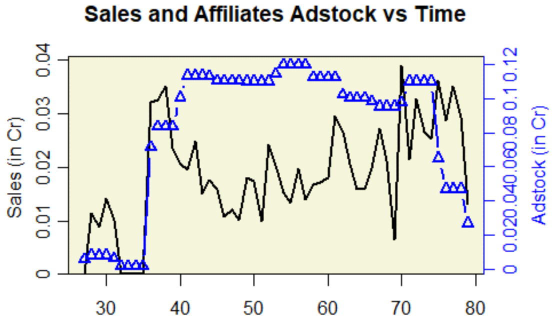


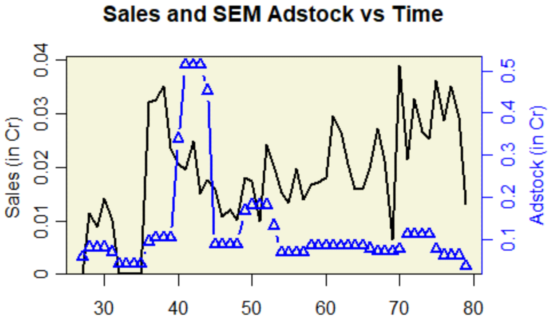








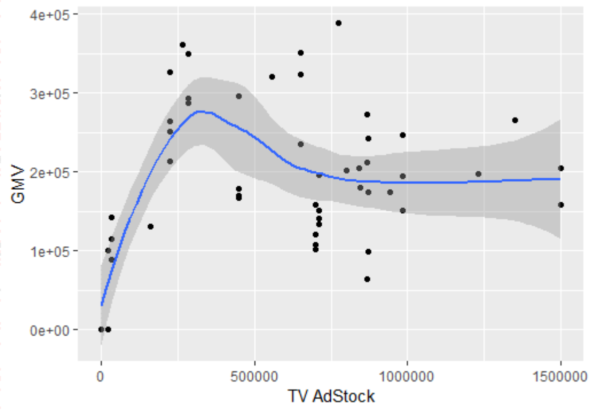


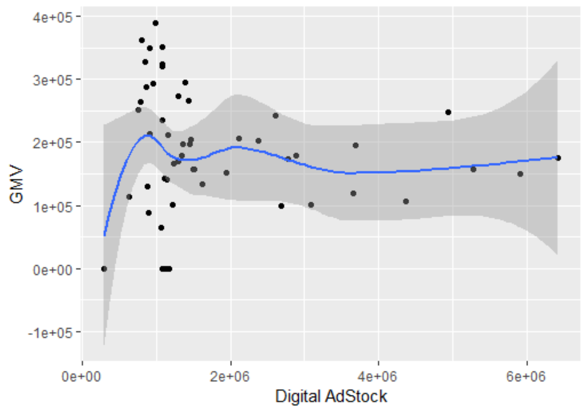


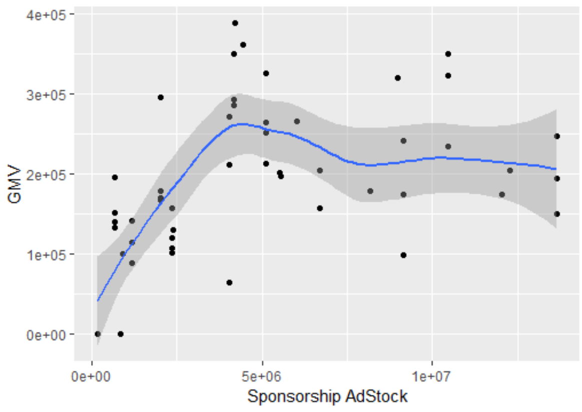
APPENDIX

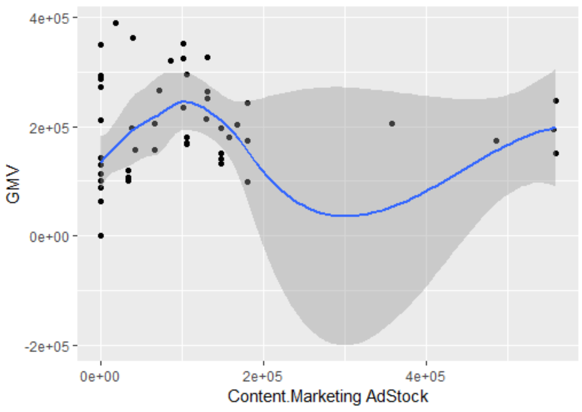
Visualizing GMV vs. Adstock

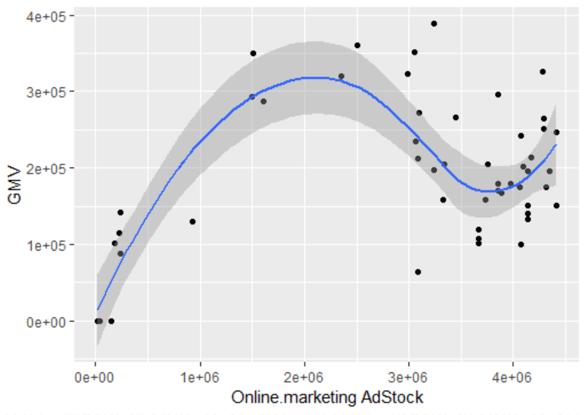
Home Audio Visualizations

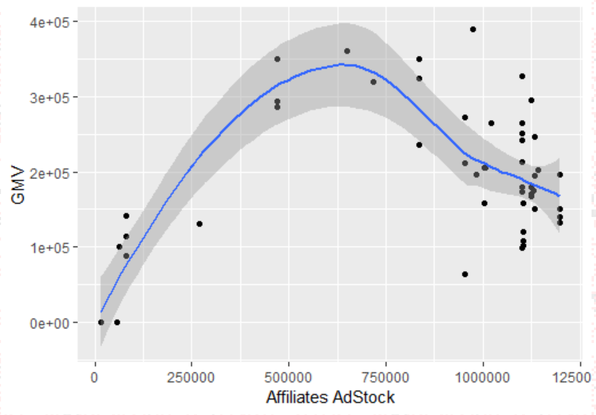


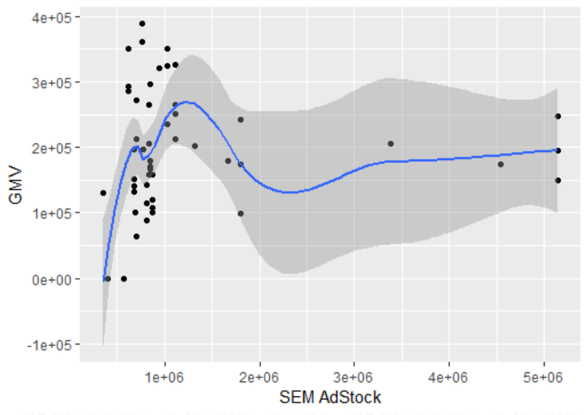




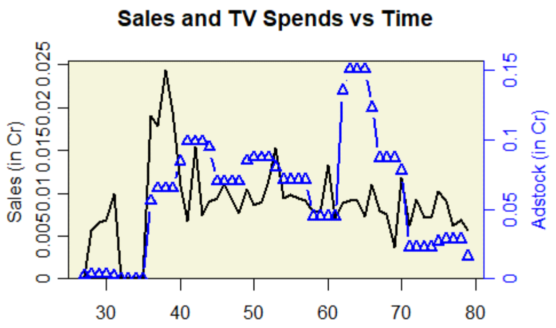


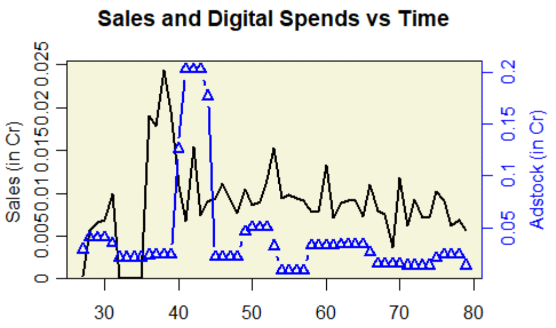


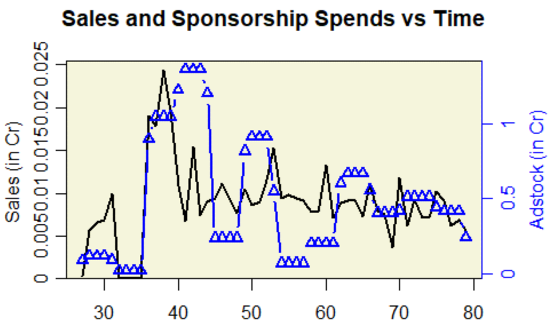


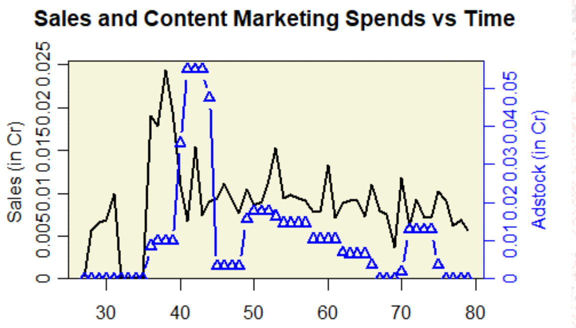


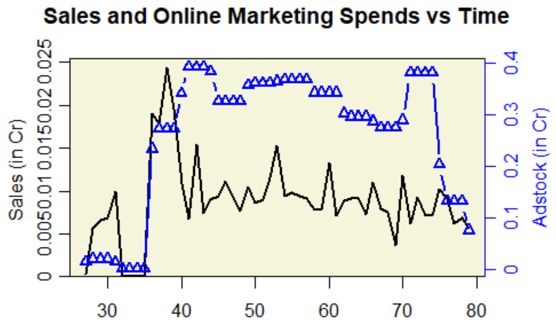
EDA of Advertising Spends and Sales vs Time

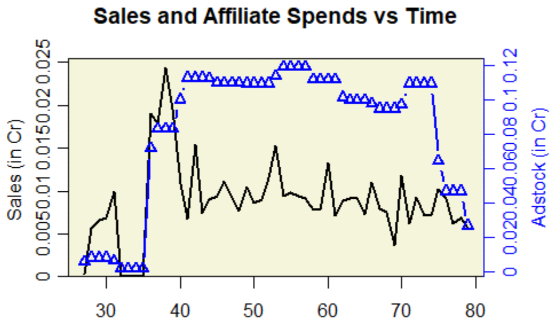


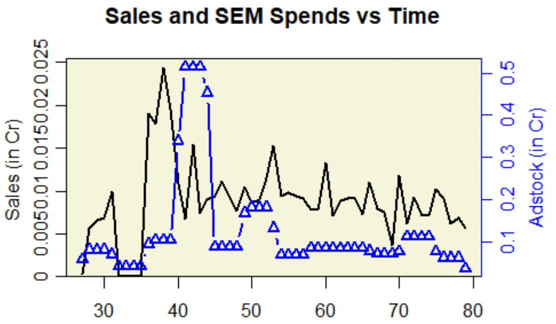












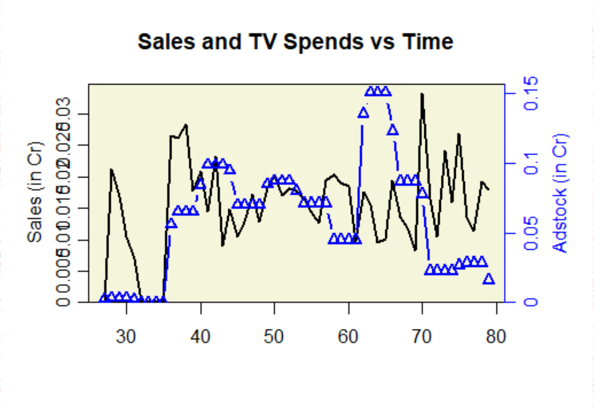
Inference:

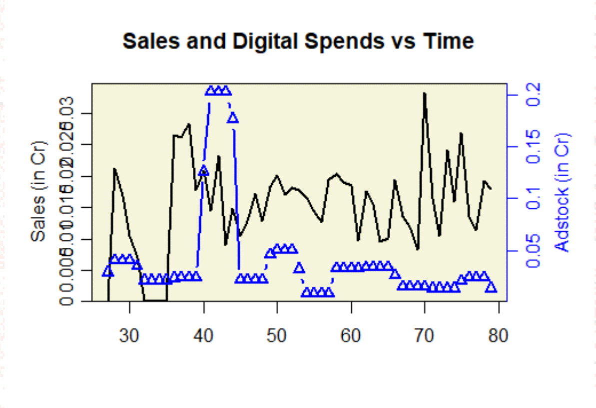
Note: We have presented different visualizations for Gaming Accessories Sub-Category here and added similar visualizations for Home Audio and Camera Accessories in the Appendix.

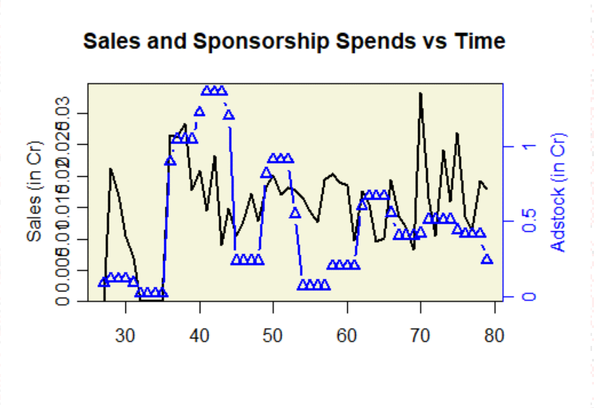
APPENDIX

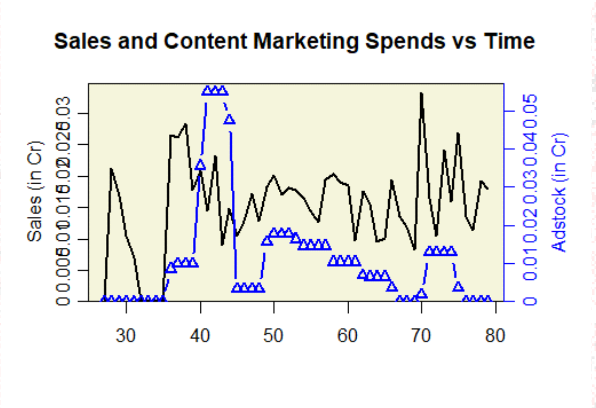
EDA of Advertising Spends and Sales vs Time

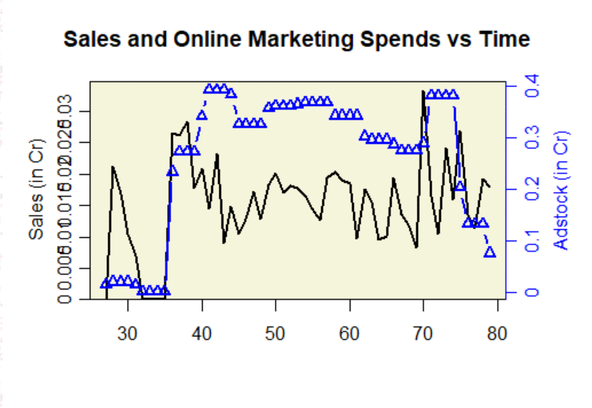
Camera Accessories Visualizations

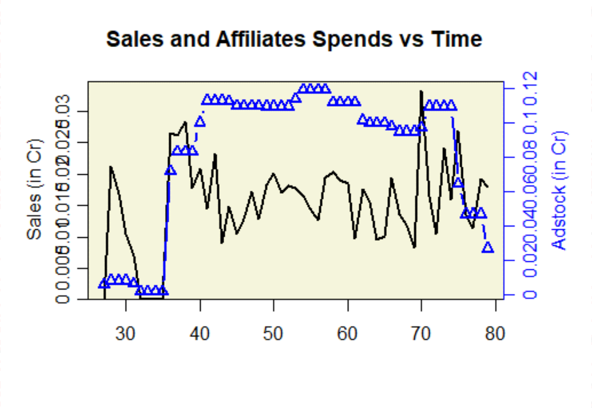


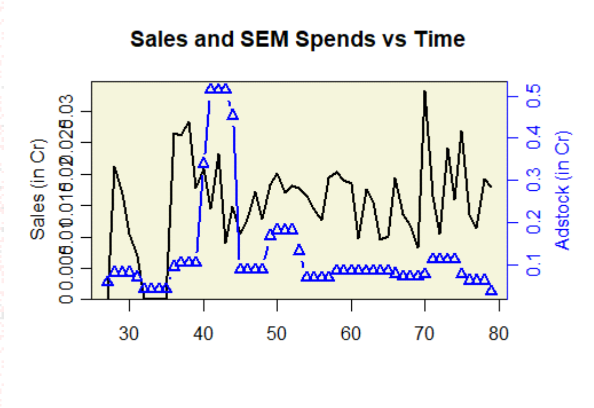










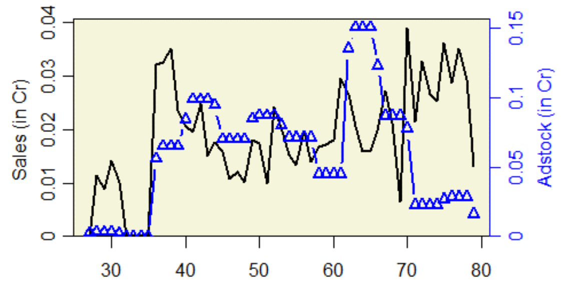


APPENDIX

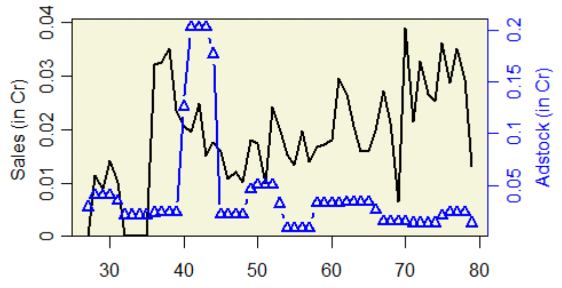
EDA of Advertising Spends and Sales vs Time

Home Audio Visualizations

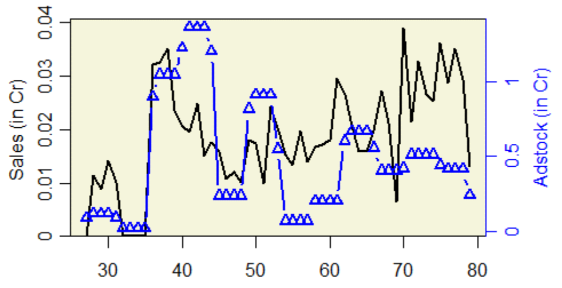
Sales and TV Spends vs Time



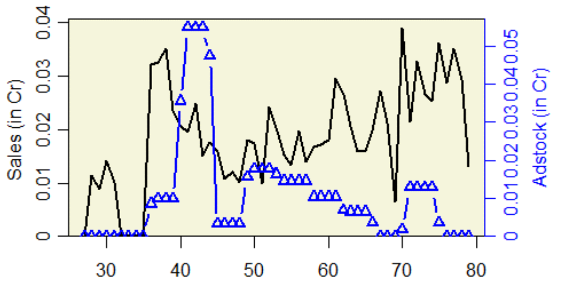
Sales and Digital Spends vs Time



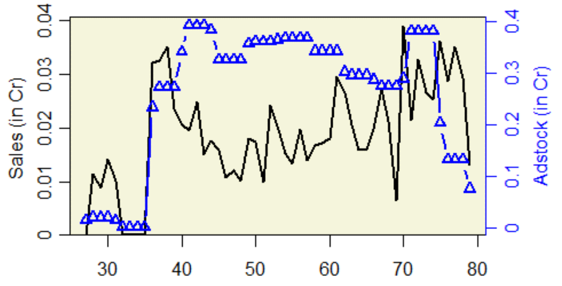
Sales and Sponsorship Spends vs Time



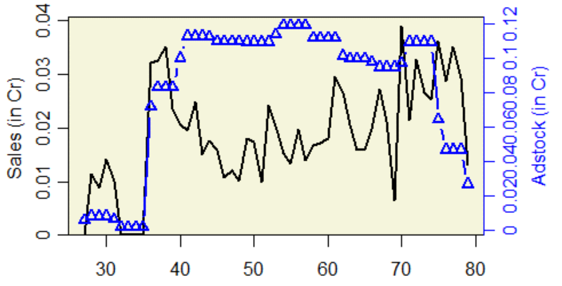
Sales and Content Marketing Spends vs Time



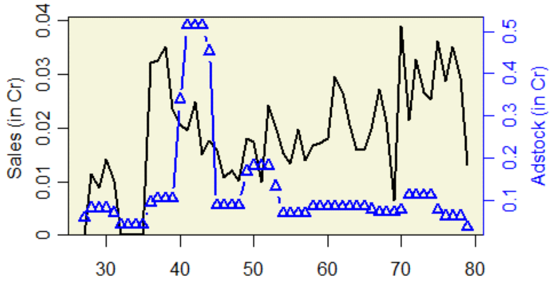
Sales and Online Marketing Spends vs Time



Sales and Affiliates Spends vs Time



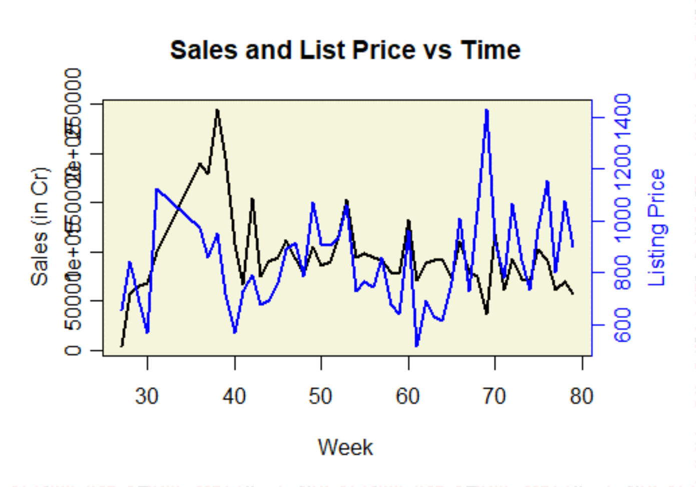
Sales and SEM Spends vs Time



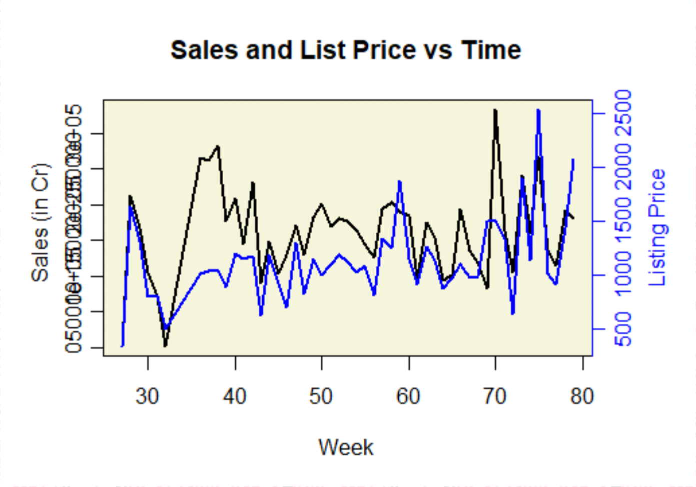
APPENDIX

EDA of Sales and List Price vs. Time

Gaming Accessories



Camera Accessories



Home Audio