ElecKart Market Mix Modeling Capston Project

Team
Saurabh Dongare
Yogesh Dokwal

Agenda

- **Business & Data Understanding**
- Data preparation & Exploratory Data Analysis
- Model Building
- Recommendation
- Challenges Faced

Business Understanding & Objective – ElectKart Market Mix Modelling

- ElecKart is an e-commerce firm specialising in electronic products. To enhance their revenues they have done significant investment in their marketing efforts, like promotions over last one year. 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 using Market Mix modelling. CRISP-DM framework will be used for modelling purpose.

- To develop a market mix model for 3 product sub-categories - Camera accessory, Gaming accessory and Home Audio to observe the actual impact of different marketing levers over sale of last one year (July 2015 -June 2016) and recommend the optimal budget allocation for different marketing levers for the next year.

Datasets walkthrough

Order Details : Contains daily order level details. 1648824 number of rows (July 2015 and June 2016)

Column Name	Significance
FSN ID	The unique identification of each SKU
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 ID’s under same order ID; Orders are tracked by Order Item ID
GMV	Gross Merchandise Value or Revenue
Units	Number of Units of the specified product Sold
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
Product MRP	Maximum retail price of the product
Product Procurement SLA	Time typically taken to procure the product

Product Details

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

Media Investment Detail

Column Name	Significance
Year	Year
Month	Month
Total Investment	Monthly total ad spent in Crores
TV	Monthly total TV ad spent in Crores
Digital	Monthly digital ad spent in Crores
Sponsorship	Monthly sponsorship spent in Crores
Content Marketing	Monthly Content Marketing spent in Crores
Online Marketing	Monthly Online Marketing Spent in Crores
Affiliates	Monthly Affiliates spent in Crores
SEM	Monthly SEM spent in Crores
Radio	Monthly Radio spent in Crores
Others	Monthly Others Spent in Crores

Yearly Promotional calendar

Column	Significance
Special Sales	Days when there was special holidays or vacation periods. This to be created and analyzed, so that optimization can be suggested for this heavily potentially profitable period for the company.

Monthly customer satisfaction score

Column	Significance
Month	Monthly customer satisfaction score

Data Preparation: Data issues/Data Clean Up

➢ There were total 4904 missing values in entire dataset across all product categories for diffirent columns.

➢ Orders having GMV value of 0
➢ Few records where GMV value was higher than MRP value.
➢ Negative values in deliverybdays and deliverycdays column.
➢ Negative values in customer id and pin code columns.
➢ Negative values in product_procurement_sla column.
➢ Outlier test is checked for all the relevant variables and outliers are detected and dealt with.

Rows(orders) with missing values for GMV, and other columns and rows are deleted (less than 0.5% of entire dataset)

➢ Since there were lots of outliers, they were replaced at appropriate cut off value.
➢ Daily order level data has been aggregated at weekly level for duration between June 2015 to July 2016 for 3

➢ Rows(orders) having GMV value greater than MRP, value of MRP is replaced by their GMV price.
➢ Rows (orders) with 0(MRP) values are deleted, since it is not possible to have product with MRP value as 0.
➢ Negative deliverybdays and deliverycdays are replaced by 0.
➢ Negative product_procurement_values are replaced by 0.

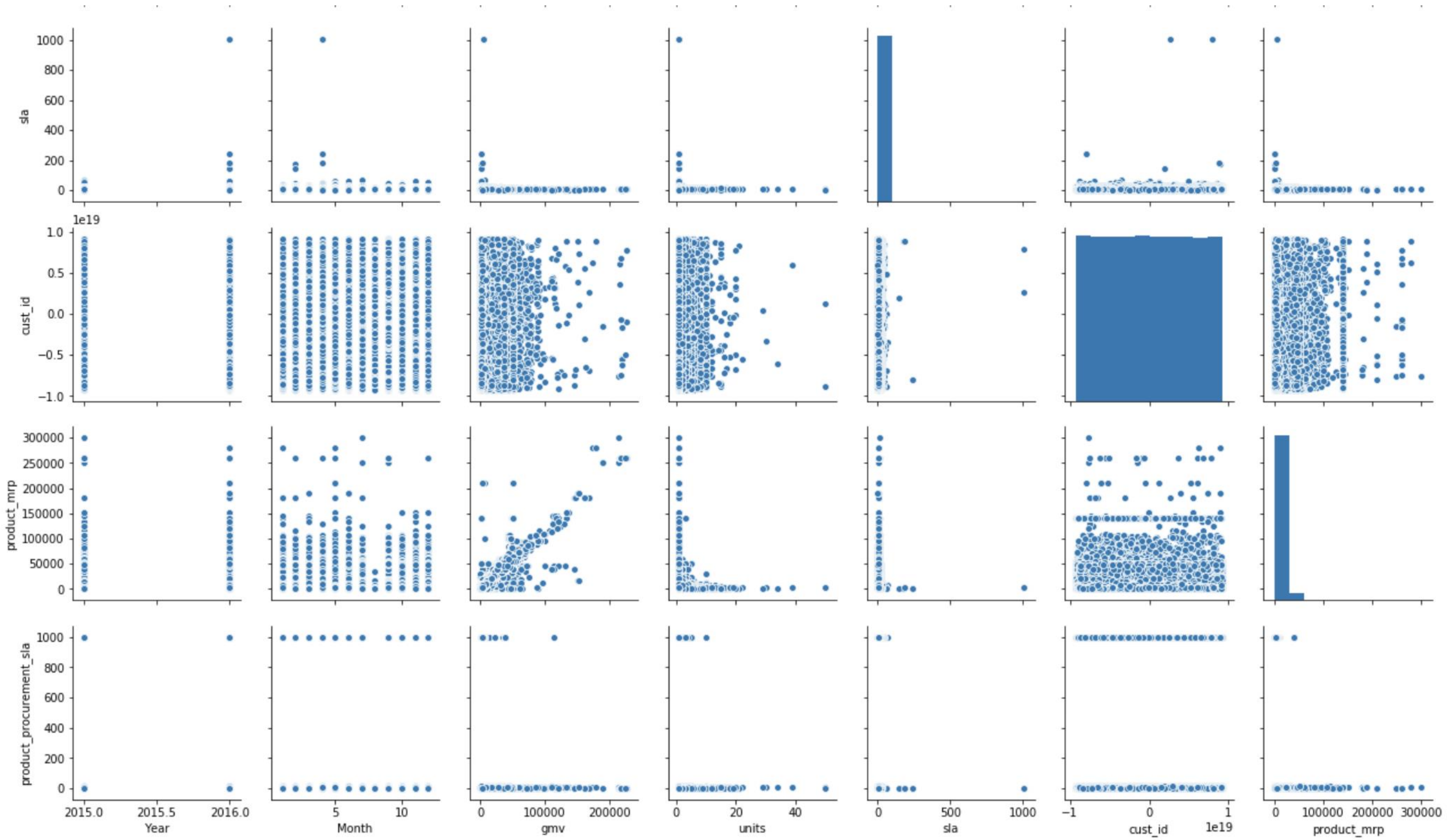
product sub categories - CameraAccessory, Home audio and GamingAccessory.
➢ Monthly level ad spend has been converted into weekly ad spend.
➢ Promotional data has been transformed to weekly level which signifies whether that particular week had any

Promotions are derived from Promotion dates given.

KPI and Variables Selection

- Advertising related KPI
  - Marketing Spend
    - channel wise marketing spend aggregated on weekly level
  - Adstock
    - Creating adstock on weekly level
- Pricing related KPI
  - Base price
    - Mean of MRP of products sold on weekly level
  - Promotional Price
    - Listed price
- Discount and Promotion related KPI
  - Discounts offered
  - Avg mean value
- Product assortment and quality related KPI
  - SLA
  - Order online percentage
- Seasonality and Trend related KPI
  - Climate
  - Special sales days
- Other KPI
  - NPS
  - Stock Index

Pair plots for given data: EDA



Data Understanding : EDA

- The number of customers visiting the website goes up especially during holidays, and sales goes down during summer.
- There have been on time delivery in these periods with online payments.
- The deliveries are delayed during the last 3 months (i.e. April to June) , which need to be addressed.
- There are also subtle spikes during the promotional periods across the year which boosts sales.
- Week 1 is considered as 1st week of July 2015 and week 53 as last week of June 2016.
- Therefore, the amount of sales during the Holiday week is far more than the normal week. This suggests holiday weeks are more profitable.
- 3 different datasets for CameraAccessory, Home audio and GamingAccessory have been filtered.

Models and Results (Home Audio)

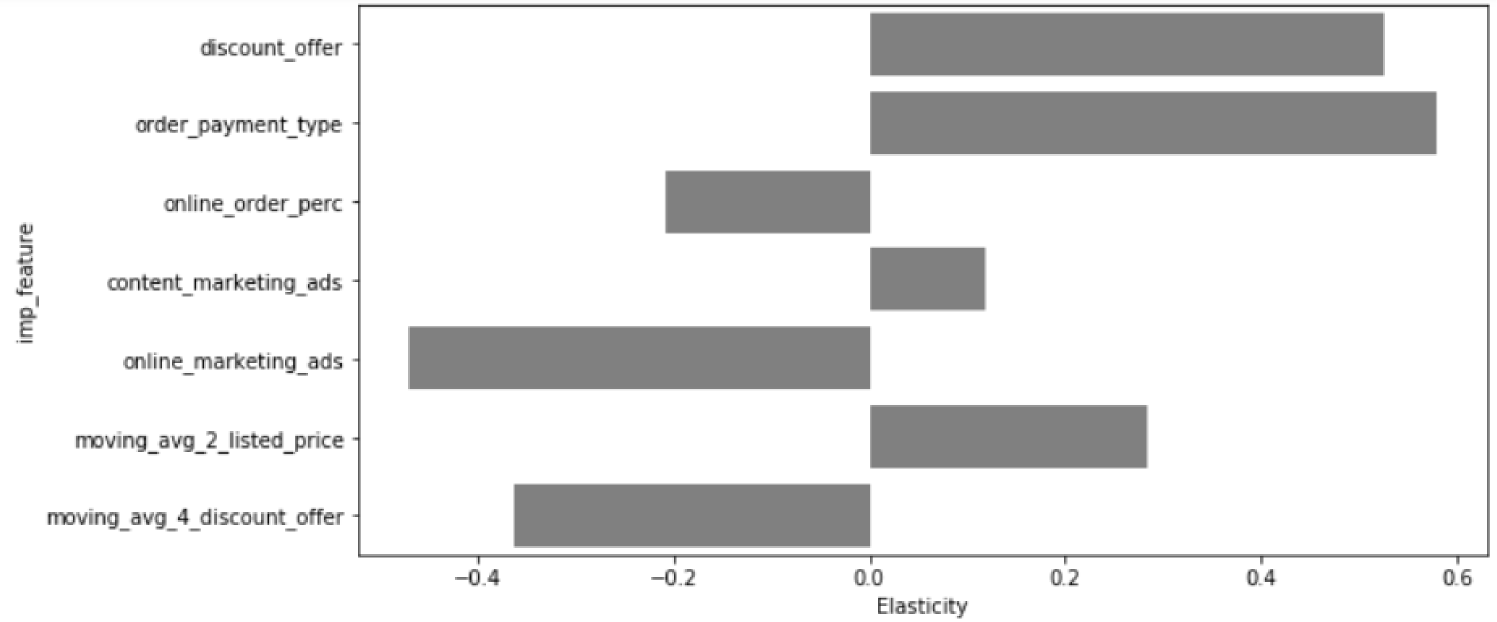
Models	KPIs	Number of features	R. Sq.	MSE value	Adjusted R Square	Business Significance
Linear Model	discount_offer, online_order_perc,content_marketing_ads, online_marketing_ads, moving_avg_2_listed_price, moving_avg_4_discount_offer, order_payment_type	7	0.922	0.0039	0.908	High
Multiplicative Model	discount_offer, sla, order_payment_type, online_order_perc, content_marketing_ads, cool_deg_days	6	0.997	0.00008	0.997	Medium
Kyock Model	discount_offer, order_payment_type, online_order_perc, content_marketing_ads, online_marketing_ads, moving_avg_2_listed_price, moving_avg_4_discount_offer	6	0.922	0.0039	0.908	High
Distributed Lag Model	discount_offer, order_payment_type , online_order_perc , moving_avg_2_listed_price	4	0.857	0.0046	0.844	Low

Results and Reviews

Kyock Model – is best fit for Home appliances

- Discounted offers have high positive elasticity and also discounts has major contribution in driving the sales.
- Content Marketing ads are major factor – as it has positive elasticity with sales, more investing is needed.
- Moving Average discount offer shows that if discount directly proportional to sales.

- Suggestion – More focus should be given on attractive offers and discounts, and give the customers more benefits while making online payments. Online marketing needs a budget cut, as it doesn't make any impact on sales.



	imp_feature	coef	Elasticity
0	discount_offer	0.428290	0.525
1	order_payment_type	0.622708	0.578
2	online_order_perc	-0.359379	-0.207
3	content_marketing_ads	0.182600	0.119
4	online_marketing_ads	-0.211377	-0.469
5	moving_avg_2_listed_price	0.133699	0.284
6	moving_avg_4_discount_offer	-0.253363	-0.362

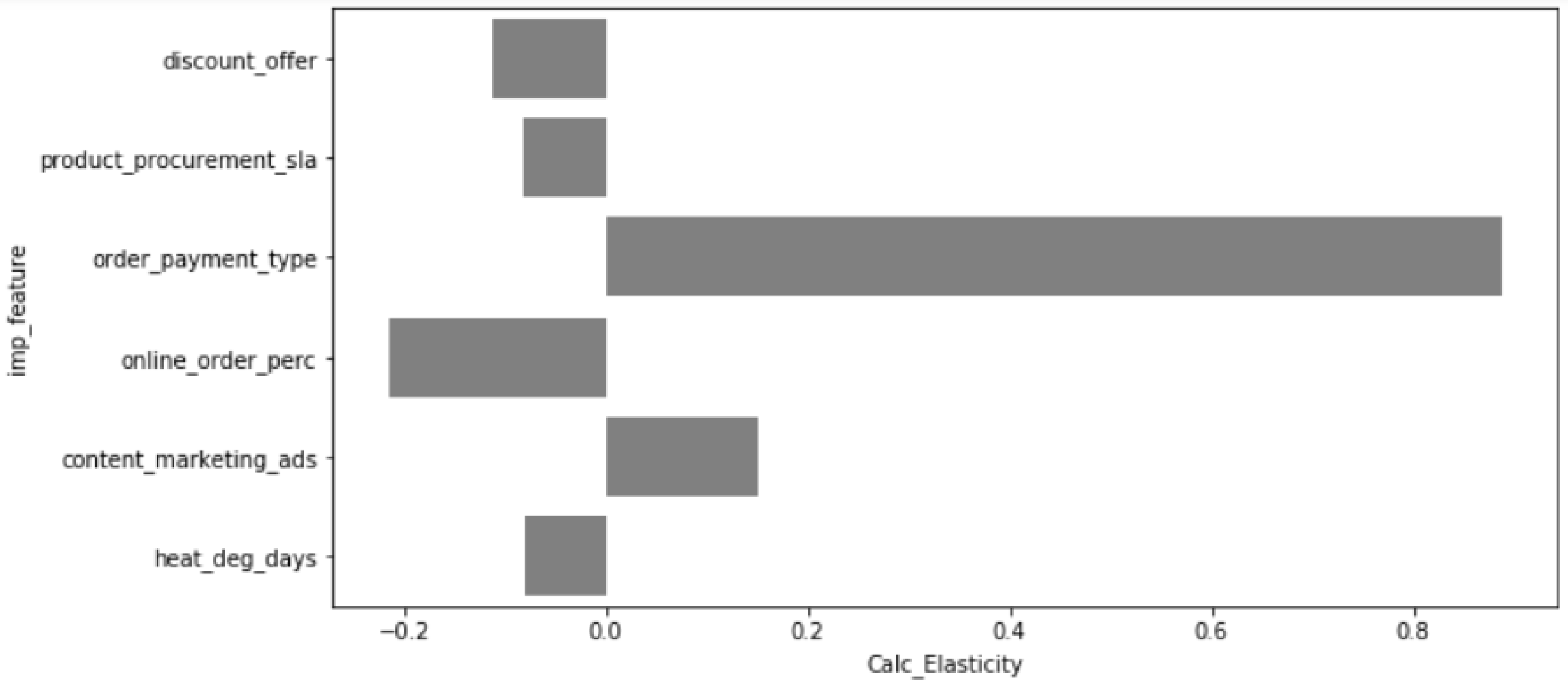
Models and Results (Camera)

Models	KPIs	Number of features	R. Sq.	MSE value	Adjusted R Square	Business Significance
Linear Model	Sla, product_procurement_sla, online_order_perc, online_marketing_ads, moving_avg_2_listed_price, moving_avg_4_listed_price, moving_avg_4_discount_offer, order_payment_type	8	0.882	0.0103	0.860	Low
Multiplicative Model	discount_offer, product_procurement_sla, order_payment_type, online_order_perc, content_marketing_ads, heat_deg_days	6	0.985	0.0035	0.983	High
Kyock Model	Sla, product_procurement_sla, online_order_perc, online_marketing_ads, moving_avg_2_listed_price, moving_avg_4_listed_price, moving_avg_4_discount_offer, order_payment_type	8	0.882	0.0103	0.860	Medium
Distributed Lag Model	product_procurement_sla, online_order_perc, online_marketing_ads, moving_avg_4_listed_price, snow_on_grnd_cm, moving_avg_4_discount_offer, order_payment_type, discount_offer_lag1, discount_offer_lag2, stock_index_lag3	10	0.803	0.0240	0.715	Medium

Results and Reviews

Multiplicative Model – this model is suitable for Camera Accessories

- In this model discounts has negative elasticity on sales
- But the content marketing has a positive elasticity or relation, making it a good attribute for investing more money.
- Summers would be hard time for sales revenue generations, so this would be a better time to give away discounts



	imp_feature	coef	Calc_Elasticity
0	discount_offer	-0.280683	-0.112
1	product_procurement_sla	-0.105832	-0.081
2	order_payment_type	1.008709	0.888
3	online_order_perc	-0.270395	-0.215
4	content_marketing_ads	0.170804	0.151
5	heat_deg_days	-0.089064	-0.080

- Suggestion-

Sales will be increased for camera accessories if we make sure content marketing and discounts are controlled accordingly during seasons.

Models and Results (Gaming)

Models	KPIs	Number of features	R. Sq.	MSE value	Adjusted R Square	Business Significance
Linear Model	Sla, online_order_perc, NPS, pay_day, sponsorship_ads, moving_avg_2_listed_price, moving_avg_2_discount_offer, order_payment_type	8	0.801	0.013	0.763	Medium
Multiplicative Model	discount_offer, sla, order_payment_type, online_order_perc, content_marketing_ads, cool_deg_days, heat_deg_days, snow_on_grnd_cm	7	0.981	0.0024	0.977	Low
Kyock Model	Sla, online_order_perc, NPS, pay_day, sponsorship_ads, moving_avg_2_listed_price, moving_avg_2_discount_offer, order_payment_type	8	0.801	0.0135	0.763	Medium
Distributed Lag Model	Sla, order_payment_type, online_order_perc, online_marketing_ads, pay_day, moving_avg_4_listed_price, moving_avg_2_discount_offer, pay_day_lag3	8	0.733	0.020	0.682	High

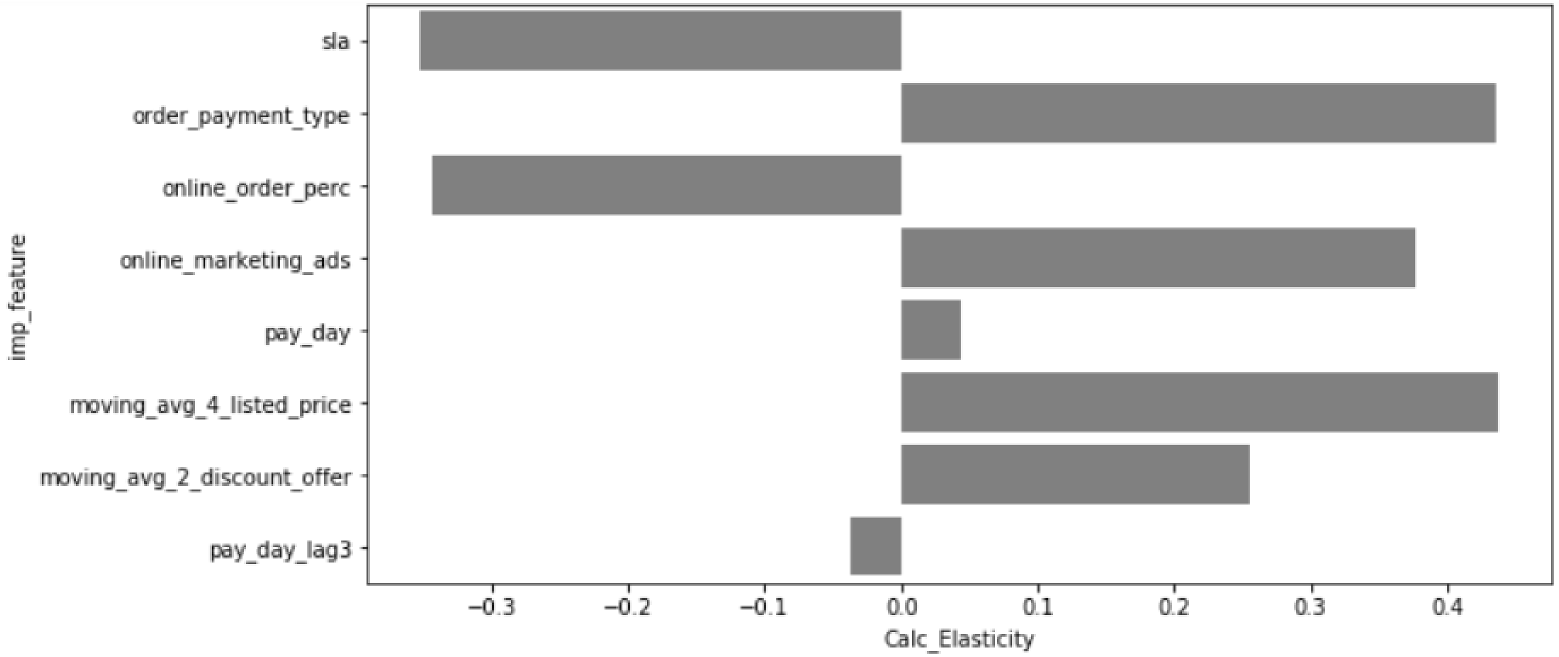
Results and Reviews

Distributed Lag Model – is the key for Gaming Accessories.

Suggestions and observations -

- Budget for online marketing needs uplift.
- Paydays needs to be targeted, marketing to be done around those days.
- Discount seems to be directly related to sales.
- Also the listed price with discount may result in good online sales, where the payment is done online too.

Rather than spending too much money expect on non paydays, more expenditure should focus on discounts.



	imp_feature	coef	Calc_Elasticity
0	sla	-0.185320	-0.352
1	order_payment_type	0.703145	0.436
2	online_order_perc	-0.523202	-0.344
3	online_marketing_ads	0.227643	0.376
4	pay_day	0.050461	0.044
5	moving_avg_4_listed_price	0.247129	0.437
6	moving_avg_2_discount_offer	0.289904	0.256
7	pay_day_lag3	-0.033651	-0.037

Thank You