Problem Statement

ElecKart is an e-commerce firm based out of Ontario, Canada specialising in electronic products. Over the last year, they had spent a significant amount of money on 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 the 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.

Imagine that you are a part of the marketing team working on budget optimisation. You need to develop a market mix model to observe the actual impact of different marketing variables over the last year. Using your understanding of the model, you have to recommend the optimal budget allocation for different marketing levers for the next year.

Data Given

- ConsumerElectronics.csv: This has data for consumer which has around 20 columns SLA: Number of days it typically takes to deliver the product, GMV: Gross Merchandise Value or Revenue, Product MRP: Maximum retail price of the product, SLA: Number of days it typically takes to deliver the product etc.

- Media data and other information.xlsx: So this has four important tabs; Product List, Media Investment, Special Sale Calendar and Monthly NPS score.

    The Product list has the list of the products name, there frequency and percentage. Where as media investment has the investment cost made on various marketing channels such as Tv, Digital, Online Marketing etc from the period of 2015 and 2016. The special sale calendar has the sale date from July 2015 to May 2016 with the pay days as 1st and 15th of every month. On the other Monthly NPS Score has data of NPS and Stock Index from the period of July 2015 to May 2016

- Product Details.docx: This is the document file which has categories and sub categories for gaming, camera and home audio accessories

Steps Involved

Imported the ConsumerElectronics data, Media data , weather data and other necessary datasets
Performing basic EDA steps:
- Identified the null values from the dataset
- Checking the anamolies like
- GMV > product_mrp*units.
- Finding the negative values of deliverybdays, deliverycdays, other columns in dataframe and removing them from the dataframe
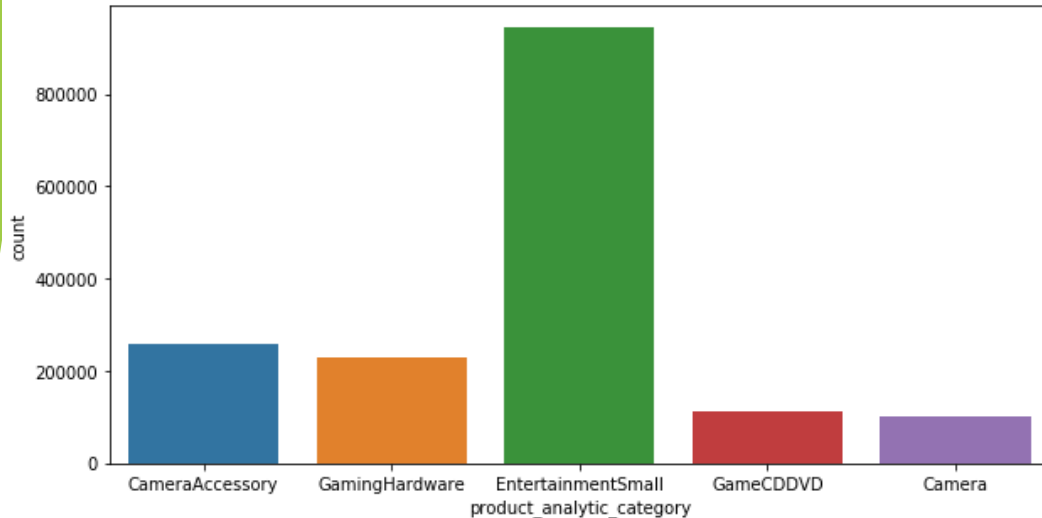

Checking the count of categories and Entertainment has most of the values compare to the other categories.
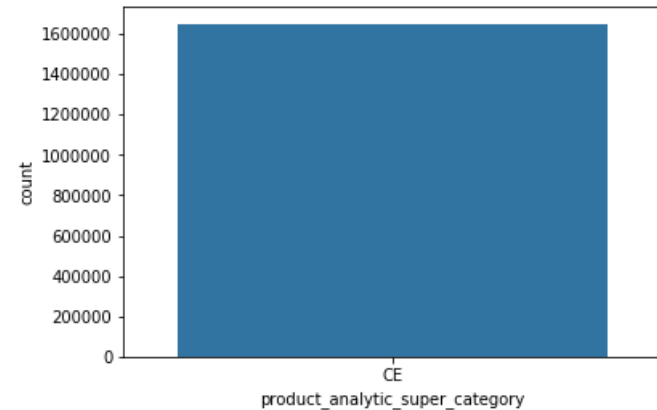
There is only one category in this column, so better to remove as it doesn't provide any useful information to the model

- Filtering data from July 2015 to June 2016 as per the problem statement
- Changing the datatypes to the proper format as per the data
- Removing un-necessary values where most of the values are null and unique category columns like customer_id,fsn_id,etc...
- Finding Correlation among the variables.

- Eventhough the graph is not much clear to present here. But We get to know from the plot that investment is 90% positively correlated to sponsorship and 86% correlated to content marketing. Content Marketing is 90% positively correlated to SEM
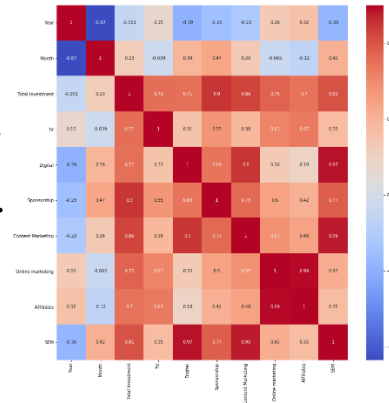
Derived Metrics:

- Derived new columns for pay_day where payday is happening for 1st and 15th of every month.
- Created holiday flay variable to identify the holidays impact to the sales which in turn results in the revenue
- Created Adstock variables using the formula **Adstock (At) = Xt + Adstock rate*At-1**
- Derived the week number and aggregated the data at week level as per the problem statement
- Created new dataframes for Camera Accessories, Gaming Accessories and Home Audio accessories to implement and analyze them separately.

Steps Involved

Scaling:
Implemented Standard Scaler to scale the finalized variables to bring them under same scale.

Train-Test Splitting:
Splitted the dataset in the 70:30 ratio for train and test dataset

Model Building:
- We first implement Linear Regression for each dataframe camera,gaming and homeaudio datasets.
- Linear model assumes an additive relationship between the different KPIs. Hence their impacts are also additive towards the dependent Y variable. This was represented by the equation $Y = \alpha + \beta_1 A_t + \beta_2 P_t + \beta_3 D_t + \beta_4 Q_t + \beta_5 T_t + \epsilon$

- We later implement linear regression by implementing Additve, Multiplicative, Koyck, distributive lag model (additive) and distributive lag model (multiplicative) models separately on the camera,gaming and homeaudio datasets

Summary Results of the Models

So we have implemented Linear Regression by using the following techniques

1. Additive   2. Multiplicative   3. Koyck   4. Distribution Lag Model(Additive)

5. Distribution Lag Model(Multiplicative)

Out of all the implemented models.

Distributed Lag Model multiplicative is giving better results

In all the products

Linear Regression Model	Category of Product	R2	MSE
Additive	Camera Accessories	0.16	86795.12
Multiplicative	Camera Accessories	0.55	0.726
Koyck	Camera Accessories	0.55	0.15
Distributive Lag Model(additive)	Camera Accessories	0.69	0.36
Distributive Lag Model(multiplicative)	Camera Accessories	0.74	0.37
-------	-------	-------	-------
Additive	Gaming Accessories	-159.98	4666678.82
Multiplicative	Gaming Accessories	-0.584	3.182
Koyck	Gaming Accessories	-51.51	30.43
Distributive Lag Model(additive)	Gaming Accessories	-0.96	1.18
Distributive Lag Model(multiplicative)	Gaming Accessories	-1.68	1.01
-------	-------	-------	-------
Additive	Home Audio Accessories	-4.6738	39930.97
Multiplicative	Home Audio Accessories	-3.48	1.77
Koyck	Home Audio Accessories	-9.8	2.7
Distributive Lag Model(additive)	Home Audio Accessories	0.39	0.27
Distributive Lag Model(multiplicative)	Home Audio Accessories	0.48	0.26

Summary Results of the Models

- For Camera Accessories the distributive Lag Model(multiplicative) is giving R2 of 0.74 and MSE of 0.37 this indicates that 74 % of the variance from the data is able to explain by the model with mean squared error of 0.37

Best Equation with Top 5 Products:

Revenue(GMV) = (0.08) * product_mrp + (0.04) * product_vertical_CameraMicrophone +(0.04) * units + (0.03) * Cool Deg Days (°C) + (0.037) * product_vertical_FlashShoeAdapter

- For Gaming Accessories the distributive Lag Model(multiplicative) is giving R2 of -1.68 and MSE of 1.01.

- Actually all the implemented models has been not able to explain the good amount of variance from the gaming accessories dataset. Even though we got the less error but the model is not able to identify the variance properly and here R2 is negative which signifies that the chosen model does not follow the trend of the data, so it fits worse than a horizontal line. It simply means the chosen model (with its constraints) fits the data really poor.

Best Equation with Top 5 Products:

Revenue(GMV) = (0.07) * product_vertical_GamePad +(0.05) *product_mrp + (0.054)*units +(0.04) * holidays_flag +(0.044) * product_vertical_GameControlMount

Summary Results of the Models

- For Home Audio Accessories the distributive Lag Model(multiplicative) is giving R2 of 0.48 and MSE of 0.26. It means 48% of the variance from the data is able to explain by this model with mean squared error of 0.26. Even though the R2 is very less but it is comparably high with other models.

  Best Equation with Top 5 Products:

  Revenue(GMV) = (0.11)* product_vertical_HomeAudioSpeaker + (0.10) * product_mrp + (0.090) * product_vertical_HiFiSystem+(0.078) * product_vertical_Dock+ (0.047) * product_vertical_DJControlle

Recommendations to the ElecKart Firm based on the results

For Camera Accessories:

Company should take product_mrp into consideration. Since 1 unit increase in product_mrp by keeping other variables constant would lead to increase revenue by 0.08 times

Apart from that company should focus on product_vertical_CameraMicrophone , units, Cool Deg Days (°C) and product_vertical_FlashShoeAdapter features to increase revenue

For Gaming Accessories:

Company should take product_vertical_GamePad into consideration. Since 1 unit increase in product_mrp by keeping other variables constant would lead to increase revenue by 0.07 times

Apart from that company should focus on product_mrp ,units ,holidays_flag and product_vertical_GameControlMount to increase revenue

Recommendations to the ElecKart Firm based on the results

For Home Audio Accessories:

Company should take product_vertical_HomeAudioSpeaker into consideration. Since 1 unit increase in product_mrp by keeping other variables constant would lead to increase revenue by 0.11 times

Apart from that company should focus on product_mrp , product_vertical_HiFiSystem, product_vertical_Dock and product_vertical_DJController to increase revenue