ECOMMERCE CAPSTONE PROJECT ELECKART

- PRANAV KANWAR
- SHIVALI DALMIA
- SHIJESH VELAYUDHAN

AGENDA

Business and data understanding

Data preparation and exploratory data analysis (EDA)

Model Building

Presentation of Results (recommendation)

Challenges faced

BUSINESS UNDERSTANDING AND OBJECTIVE

This capstone ecommerce project is based on Market Mix Modeling and we are provided with data for ElecKart an e-commerce firm based out of Ontario, Canada.

Background

ElecKart is an e-commerce firm specialising in electronic products. They have spent a significant amount of money on marketing over the last year. They have also offered big-ticket promotions occasionally. They are about to create a marketing budget for the next year, which includes spending on commercials, online campaigns, and pricing & promotion strategies. They feel 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.

Objective

- Work on budget optimisation assuming we are a part of the marketing team .
- Develop a market mix model to observe the actual impact of different marketing variables over the last year.
- Recommend the optimal budget allocation for different marketing levers for the next year based on the understanding of the model.

DATA UNDERSTANDING

Order level data: The data is from July 2015 to June 2016 and consists the following types of information

Column	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 IDs under the same order ID; orders are tracked by the Order Item ID.
GMV:	Gross Merchandise Value or Revenue
Units:	Number of units of the specific 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

DATA UNDERSTANDING

Media Investment: This contains different marketing spend at monthly level.

Column	Significance
Year	Year
Month	Month
Total Investment	Monthly Total ad spend in CR
TV	Monthly Total TV ad spend in CR
Digital	Monthly Digital ad spend in CR
Sponsorship	Monthly Sponsorship spend in CR
Content Marketing	Monthly Content marketing spend in CR
Online Marketing	Monthly Offline marketing spend in CR
Affiliates	Monthly Affiliates spend in CR
SEM	Monthly SEM spend in CR
Radio	Monthly Radio spend in CR
Other	Monthly Other spend in CR

DATA UNDERSTANDING

Yearly Promotional calendar
Promotion Name along with date of sale

	Sales Calendar
2015	(18-19th July)
	(15-17th Aug)
	(28-30th Aug)
	(17-15th Oct)
	(7-14th Nov)
	(25th Dec'15 - 3rd Jan'16)
2016	(20-22 Jan)
	(1-2 Feb)
	(20-21 Feb)
	(14-15 Feb)
	(7-9 Mar)
	(25-27 May)

Pay days are 1st and 15th of every month.

Monthly customer satisfaction score: It contains month wise customer satisfaction score in percentage.

Column	Significance
Month	Monthly customer satisfaction score

Product Details

Column	Significance
Product Category	Category name
Frequency	Frequency of the products sold
Percent	Percentage w.r.t to total sales

DATA PREPARATION DATA-ISSUES

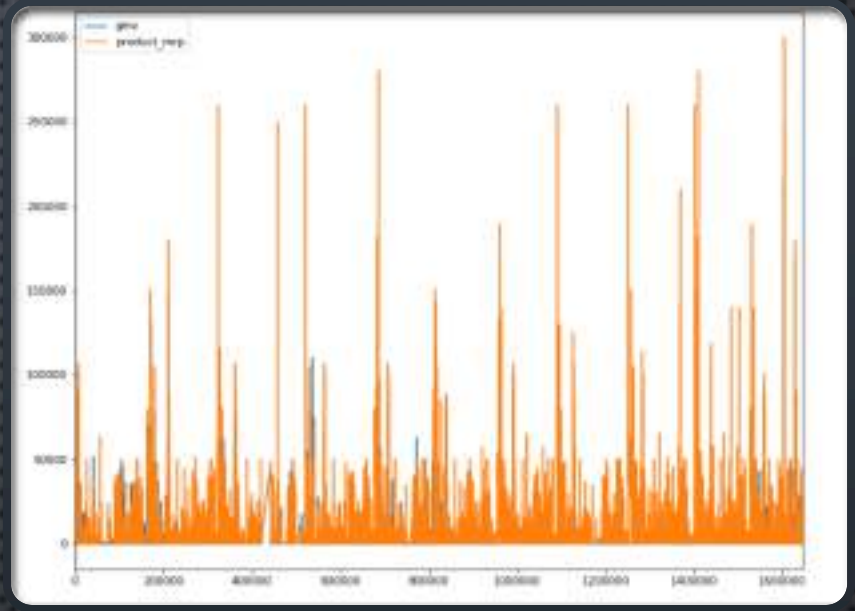
- Missing values in GMV, customer id and pin code column. There were total----- missing values in entire dataset across all product categories.
- Orders having GMV value of 0
- Few records where GMV value was higher than MRP value.
- Negative values in deliverybdays and deliverycdays column.
- Negative 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.

DATA PREPARATION DATA-CLEANUP

- Rows(orders) with missing values for GMV, Customer ID, PIN Code are deleted (Less than 0.5% of entire dataset)
- 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.
- 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 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, this is derived from Promotion dates given.
- Monthly level NPS score has been considered for each week of the month.
- All different datasets are merged together to form a single master file for carrying out modelling.
- Filtering has been done to create 3 different datasets for CameraAccessory, Home audio and GamingAccessory

DATA PREPARATION

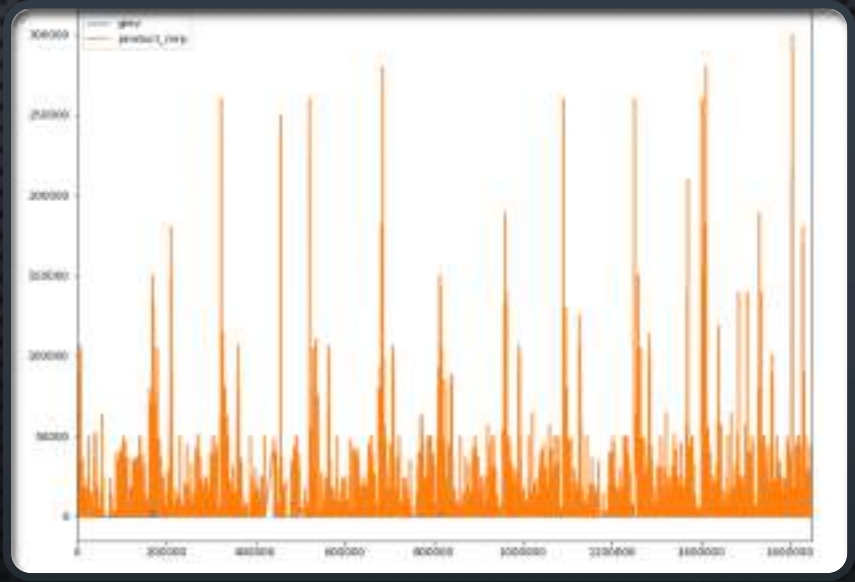
GMV V/S Product MRP



Observation
Some GMV Values are greater than their corresponding MRPs

Action Required
We will replace such MRPs with GMV

AFTER REPLACING SUCH MRPS WITH GMVS

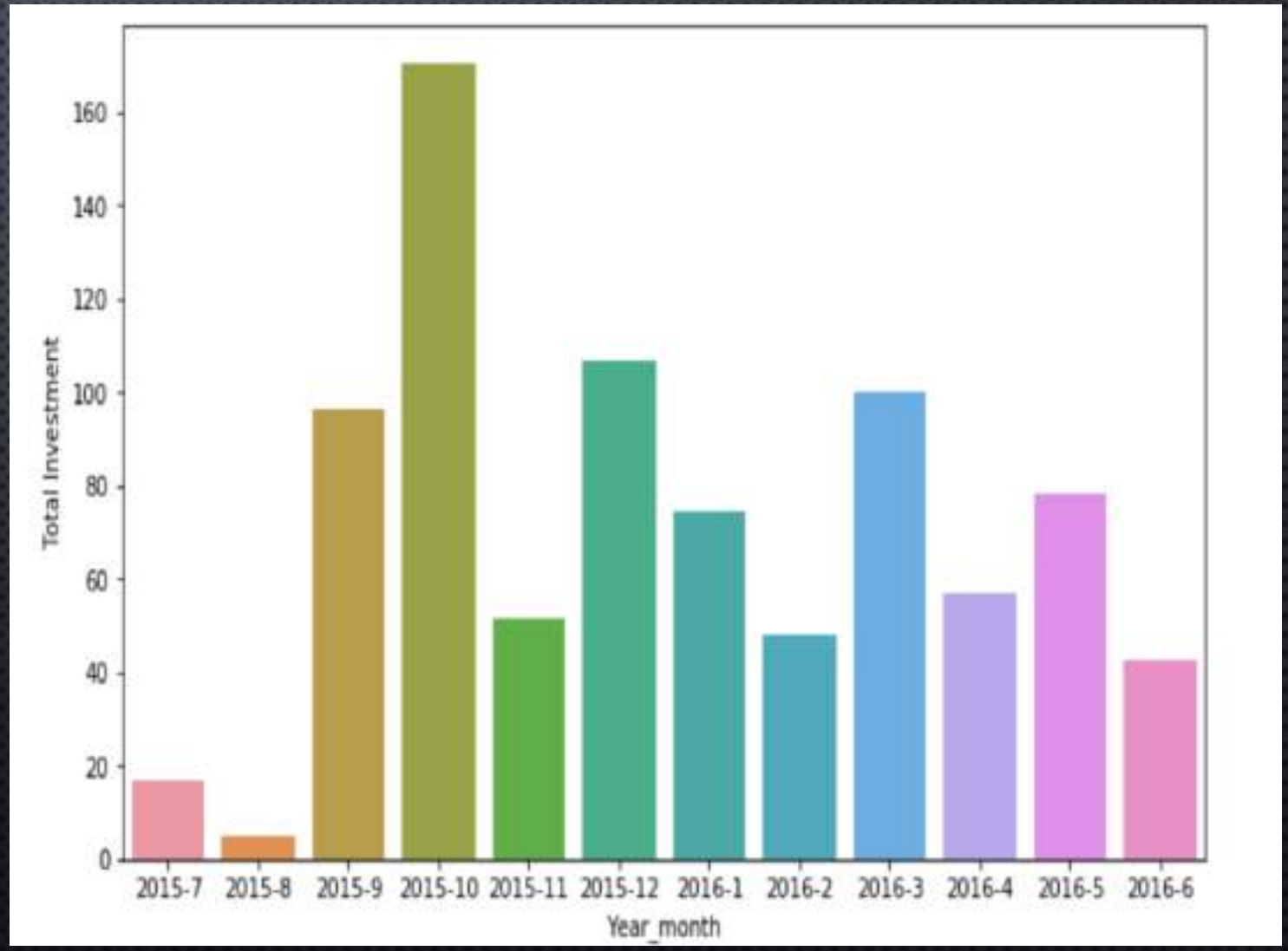


DATA UNDERSTANDING : EDA ANALYSIS

MEDIA INVESTMENT SUMMARY YEAR 2015-2016

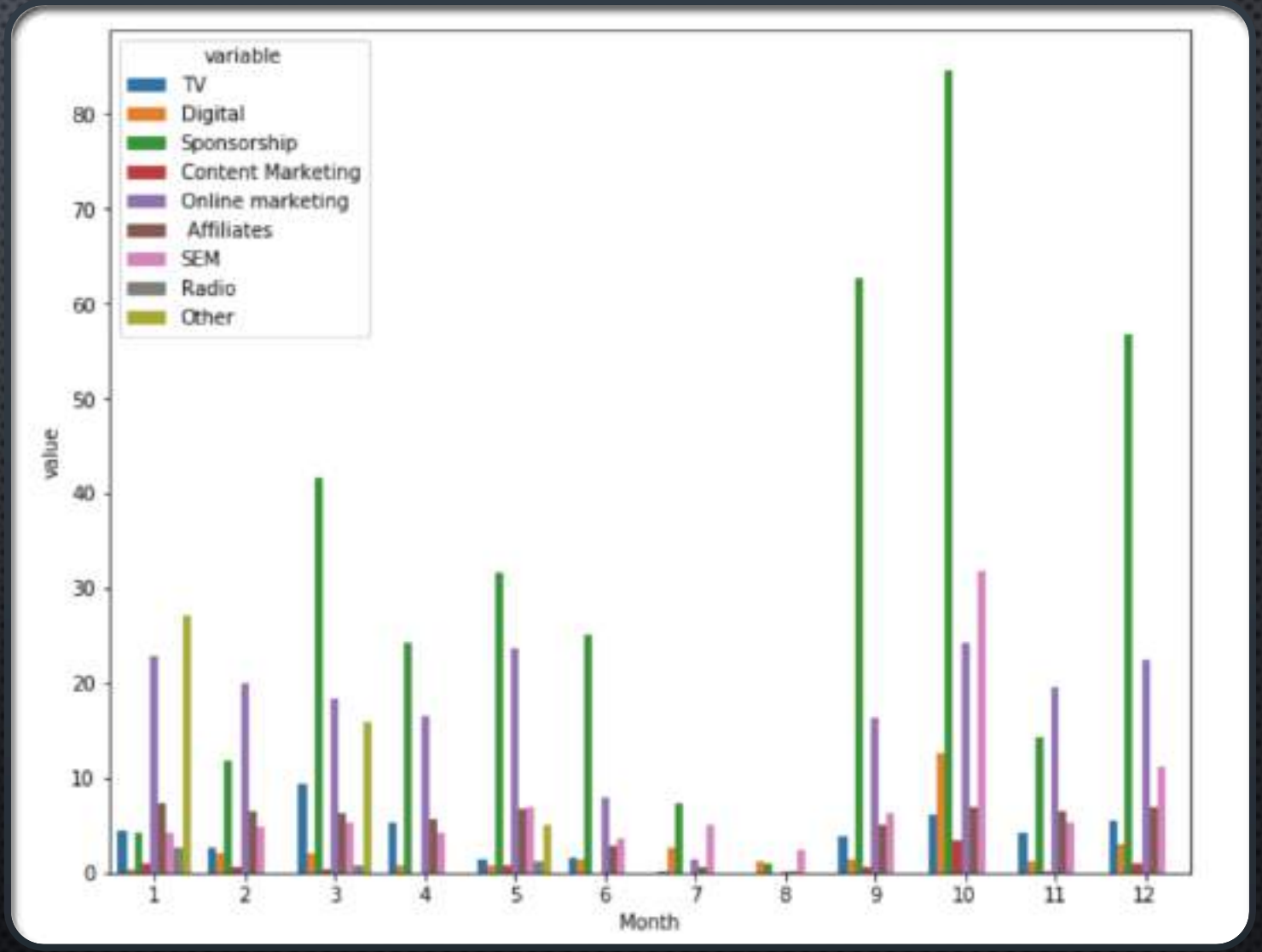
- THE LEAST SPENT IS IN 2015-08 (AUGUST 2015)
- THE HIGH SPENT ARE IN THE MONTHS OF

A. 2015-10 (OCTOBER 2015)
B. 2015-12 (DECEMBER 2015)
C. 2016-03 (MARCH 2016)
D. 2015-09 (SEPTEMBER 2015)



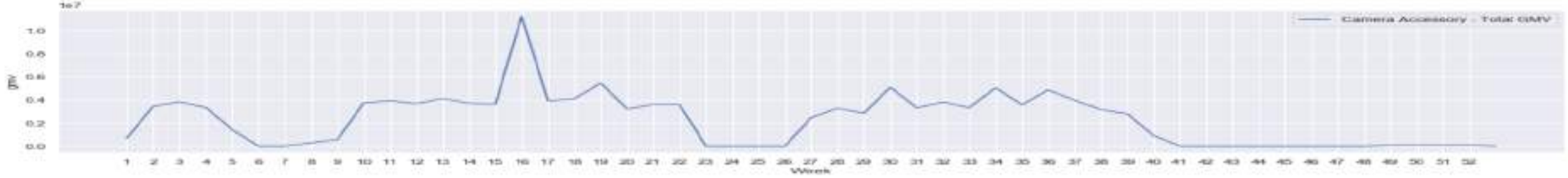
DATA UNDERSTANDING : EDA ANALYSIS

INSIGHTS OF INVESTMENT IN VARIOUS MEDIA HEADS

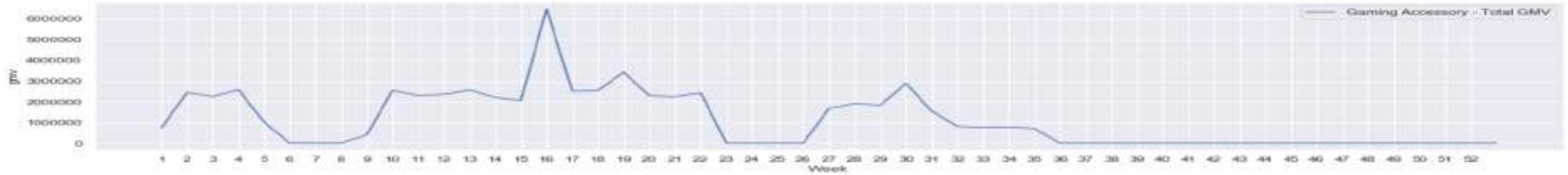


DATA UNDERSTANDING : EDA ANALYSIS

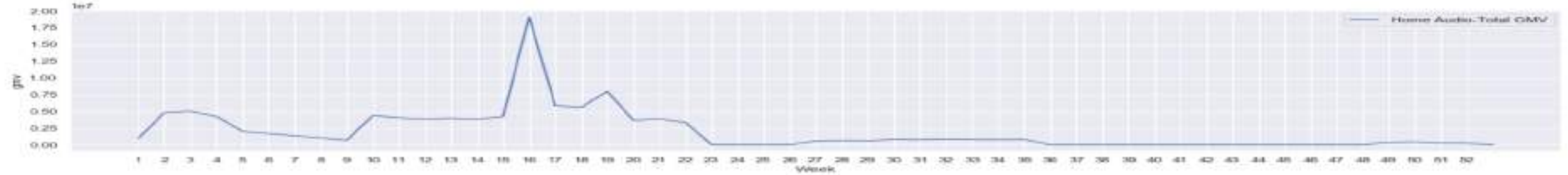
GMV WEEKLY FOR CAMERA ACCESSORIES



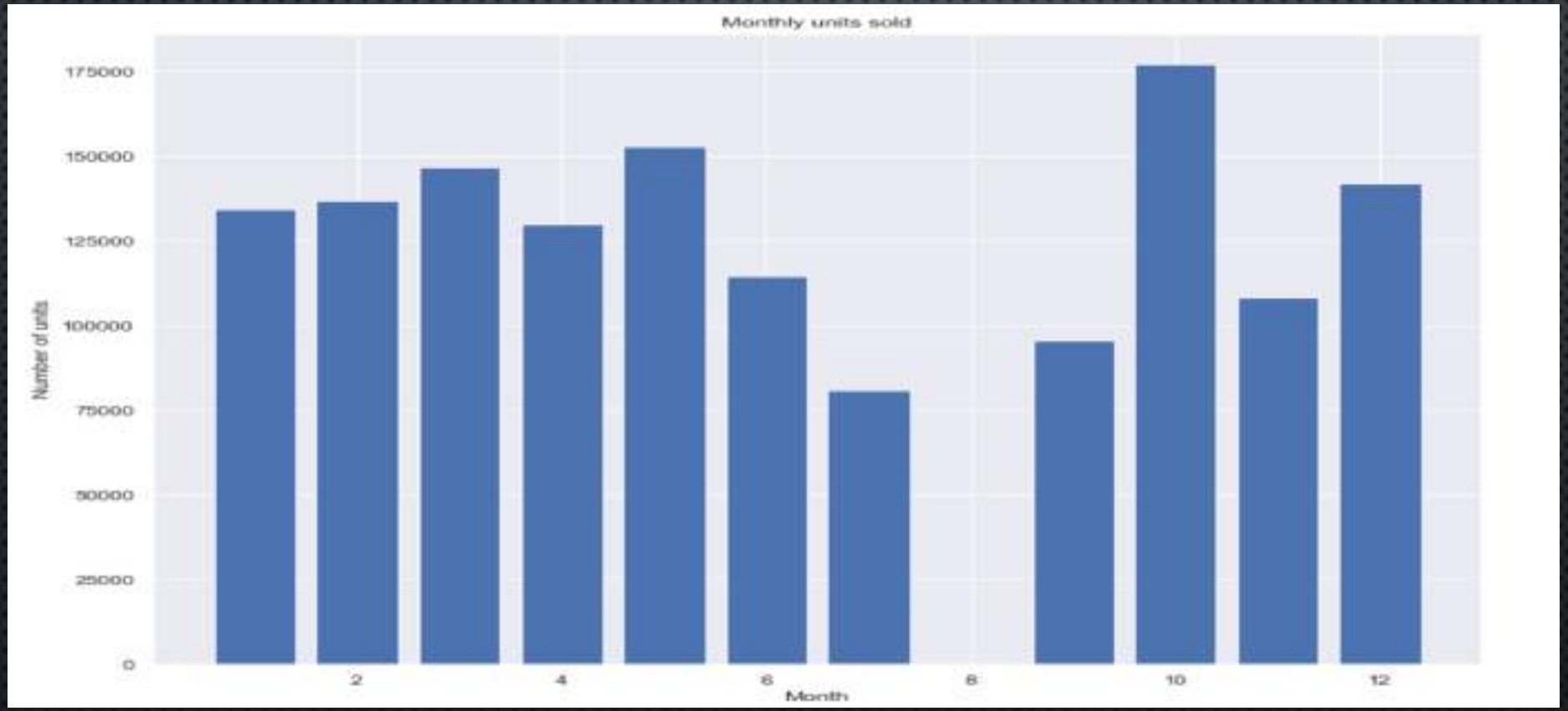
GMV WEEKLY FOR GAMING ACCESSORIES



GMV WEEKLY FOR HOME AUDIO ACCESSORIES



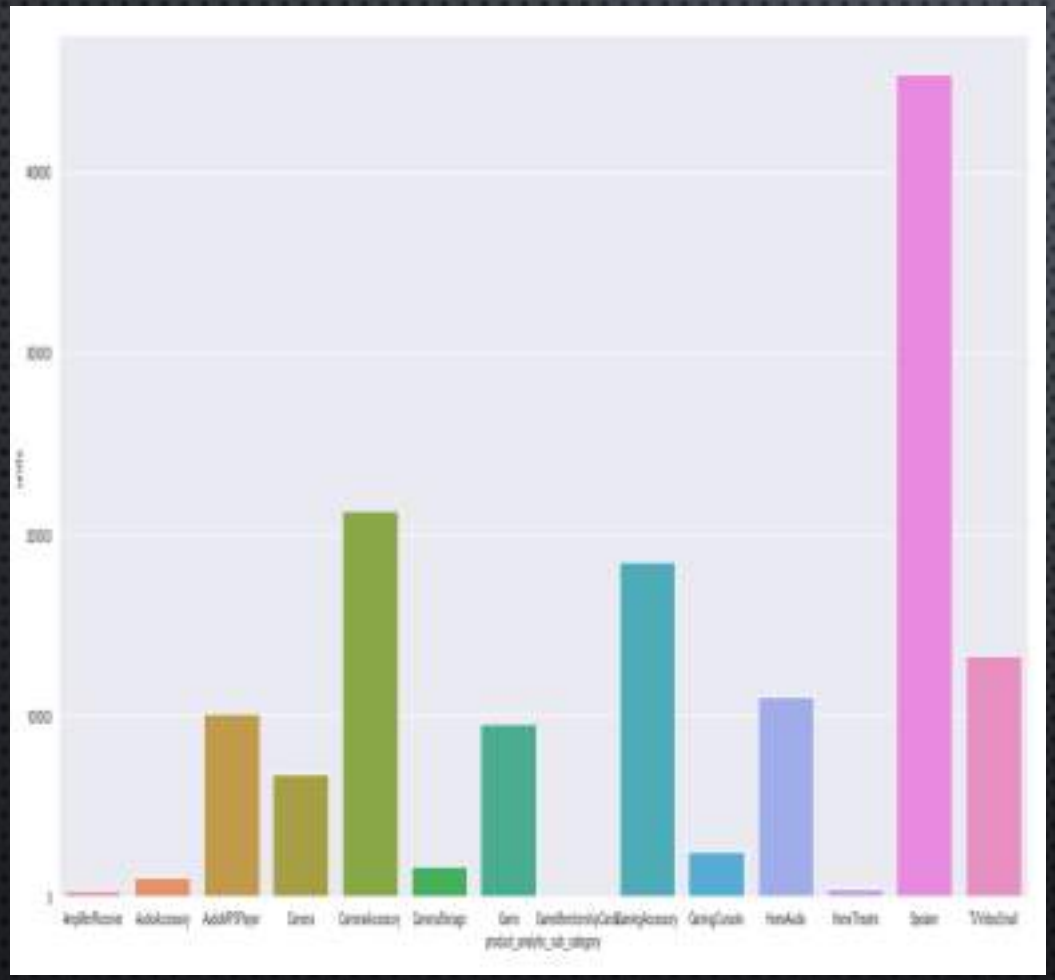
DATA UNDERSTANDING : EDA ANALYSIS



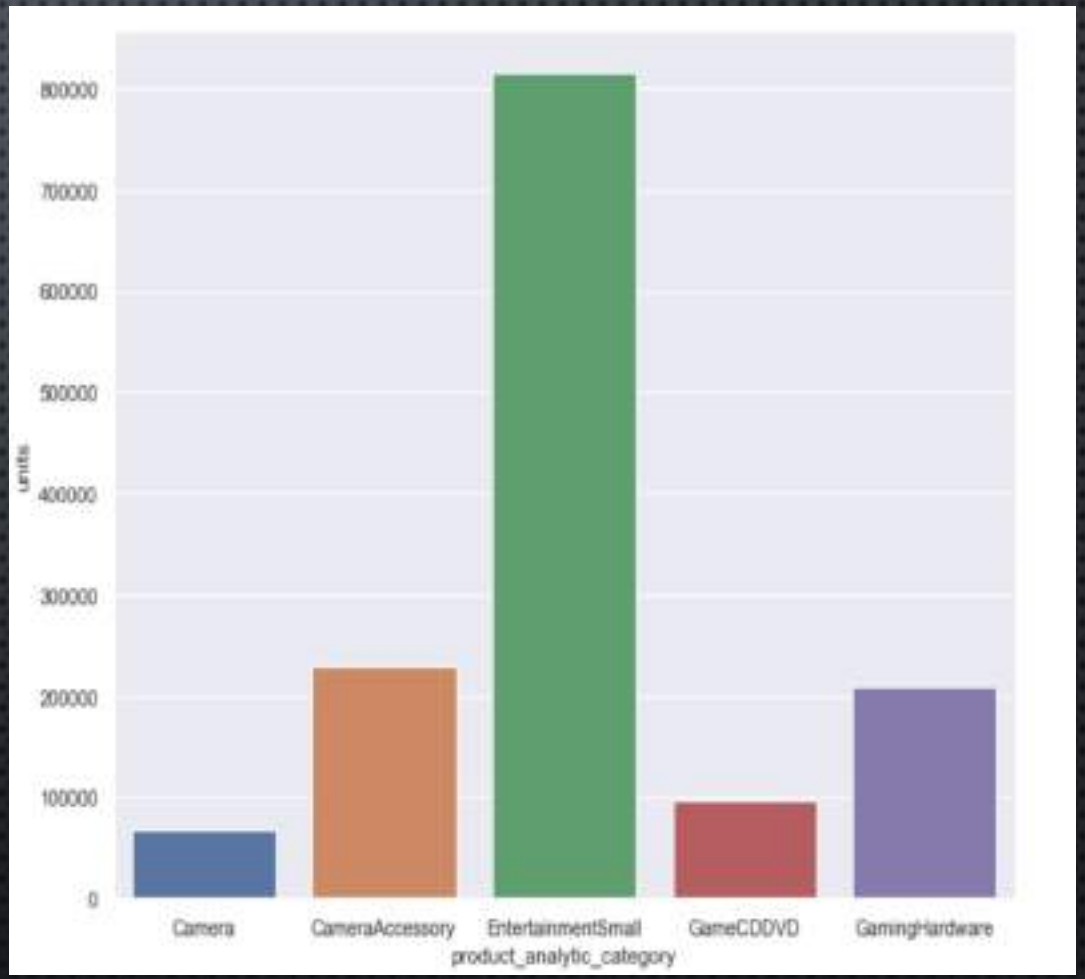
Number of Units v/s Month

October month shows highest number of units sold

DATA UNDERSTANDING : EDA ANALYSIS

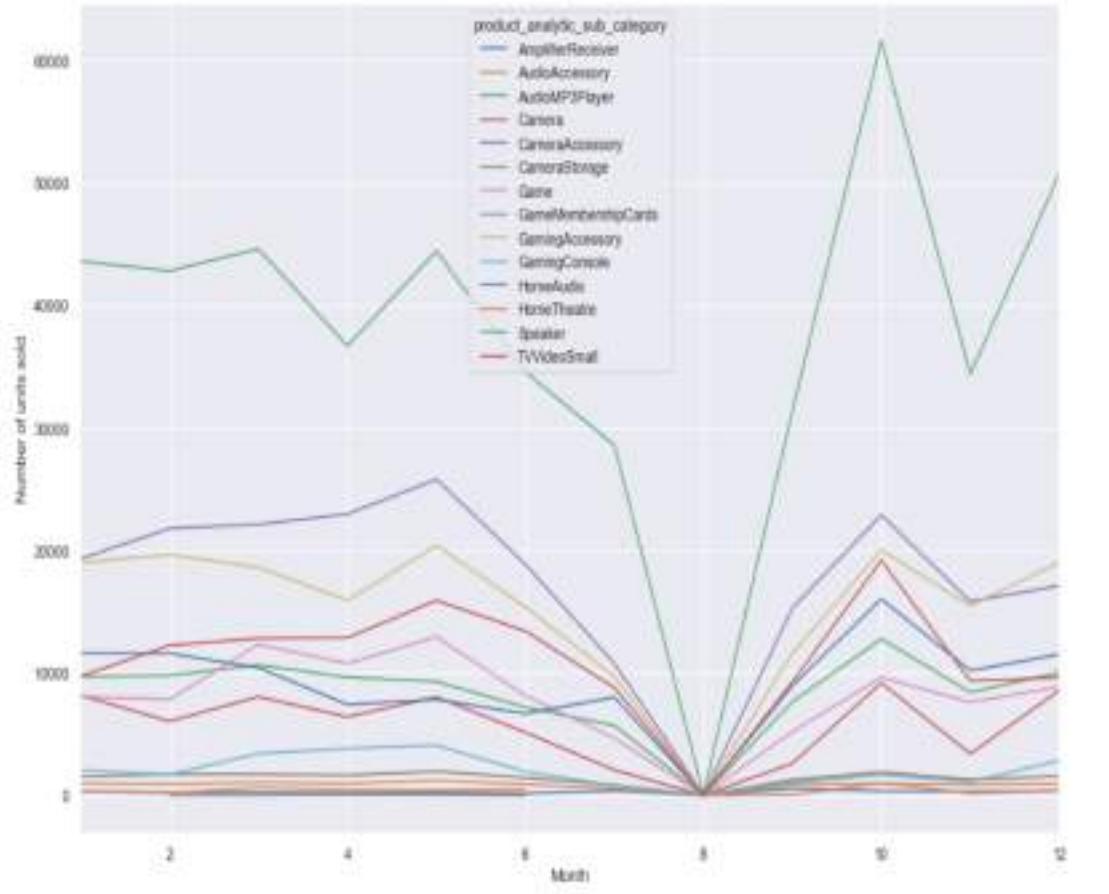


Units v/s product_analytic_sub_category

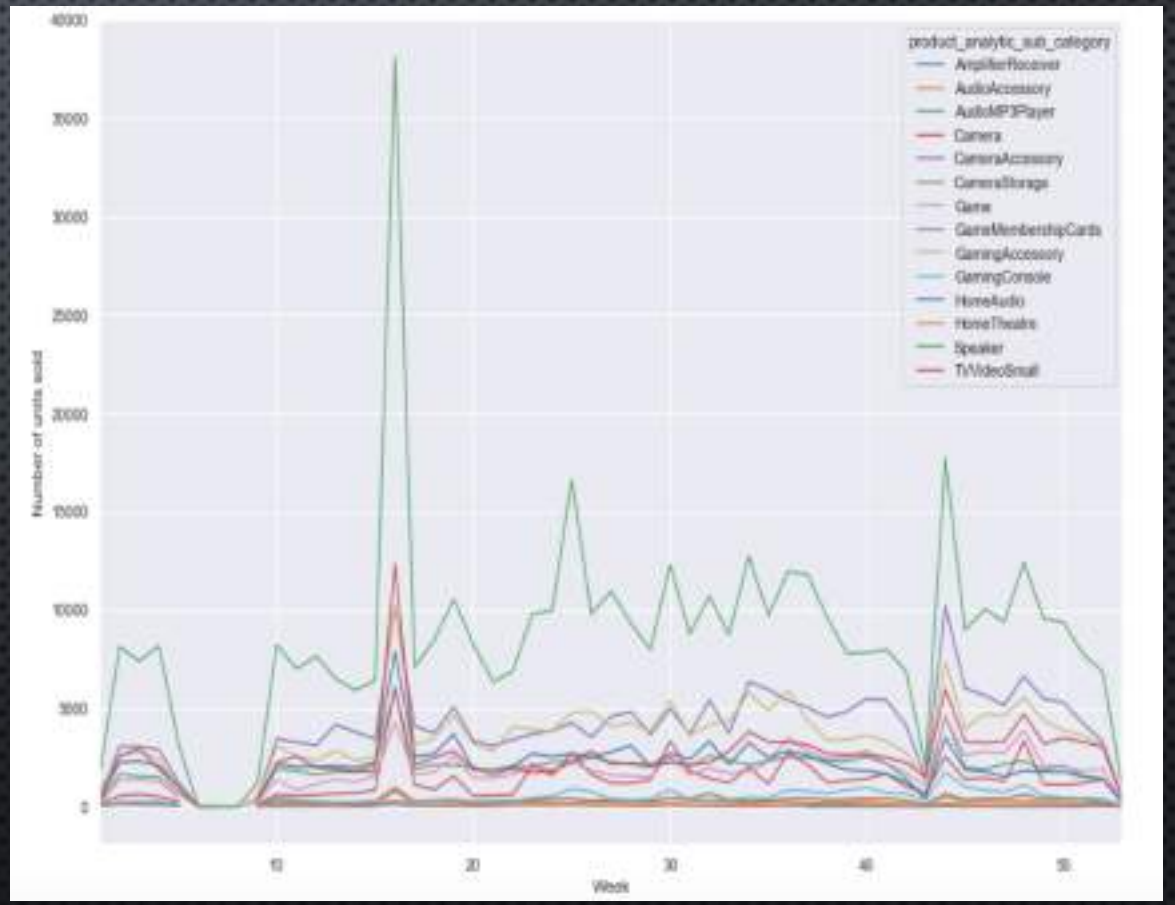


Units v/s product_analytic_category

DATA UNDERSTANDING : EDA ANALYSIS

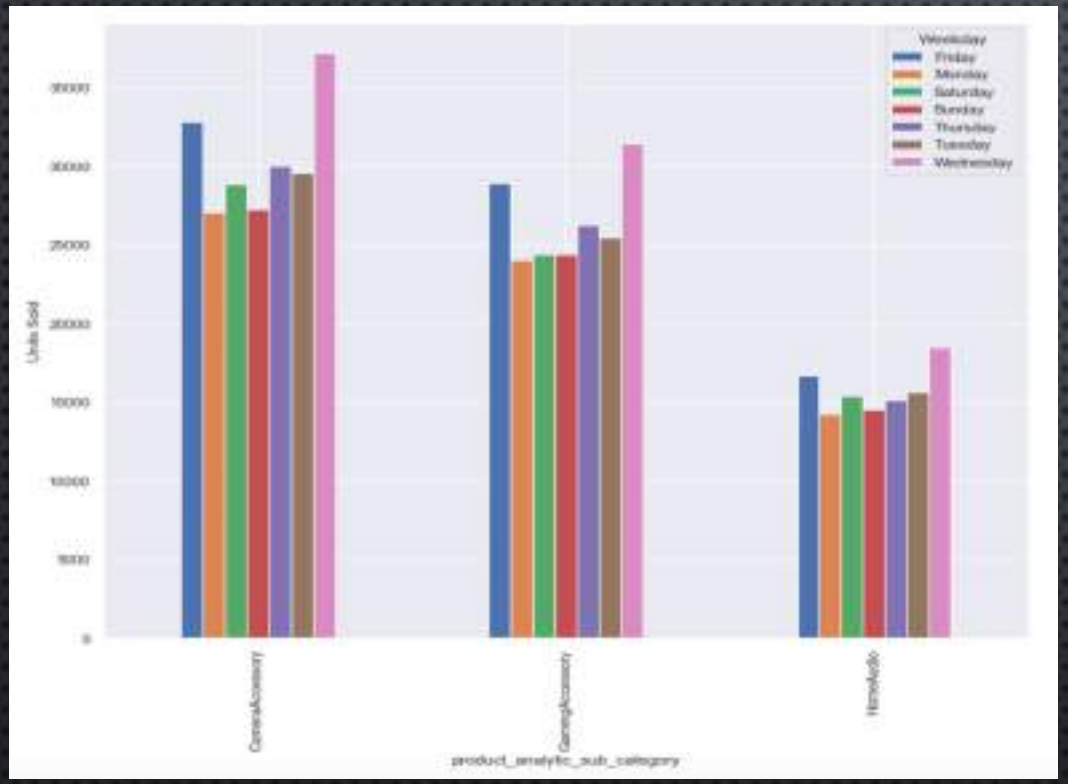


AGGREGATED MONTHLY ORDERS BY PRODUCT SUB-CATEGORY

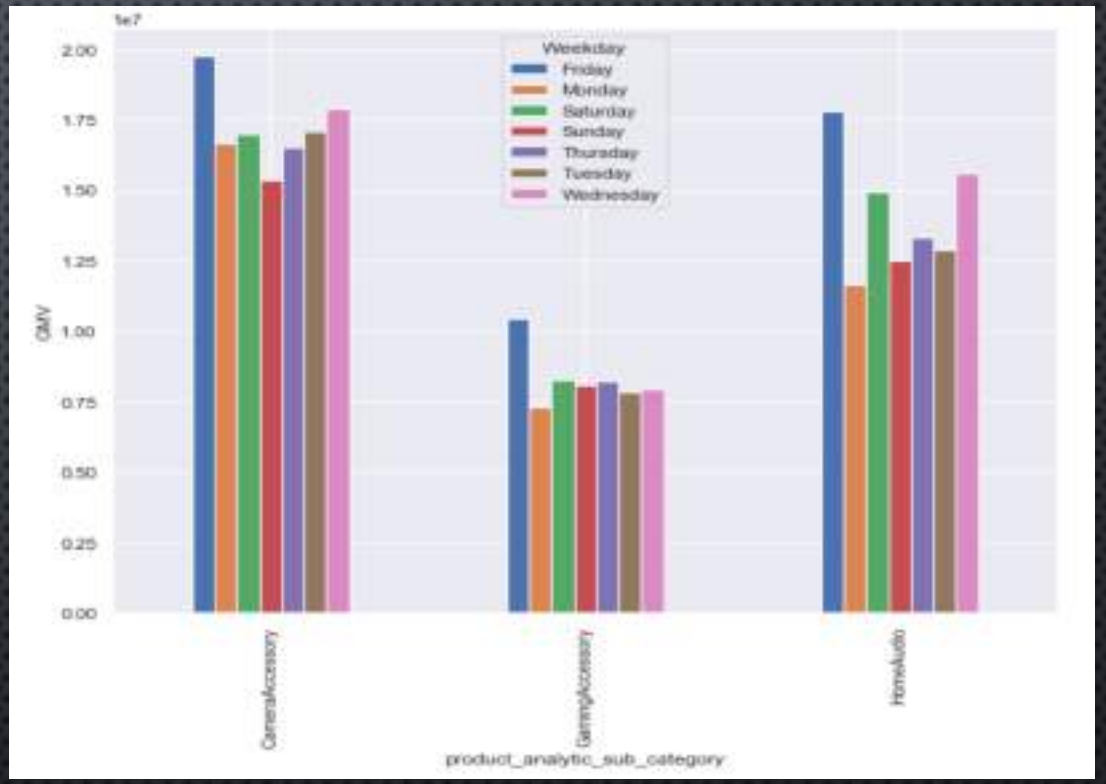


AGGREGATED WEEKLY ORDERS BY PRODUCT SUB-CATEGORY

DATA UNDERSTANDING : EDA ANALYSIS



Units sold v/s product_analytic_sub_category

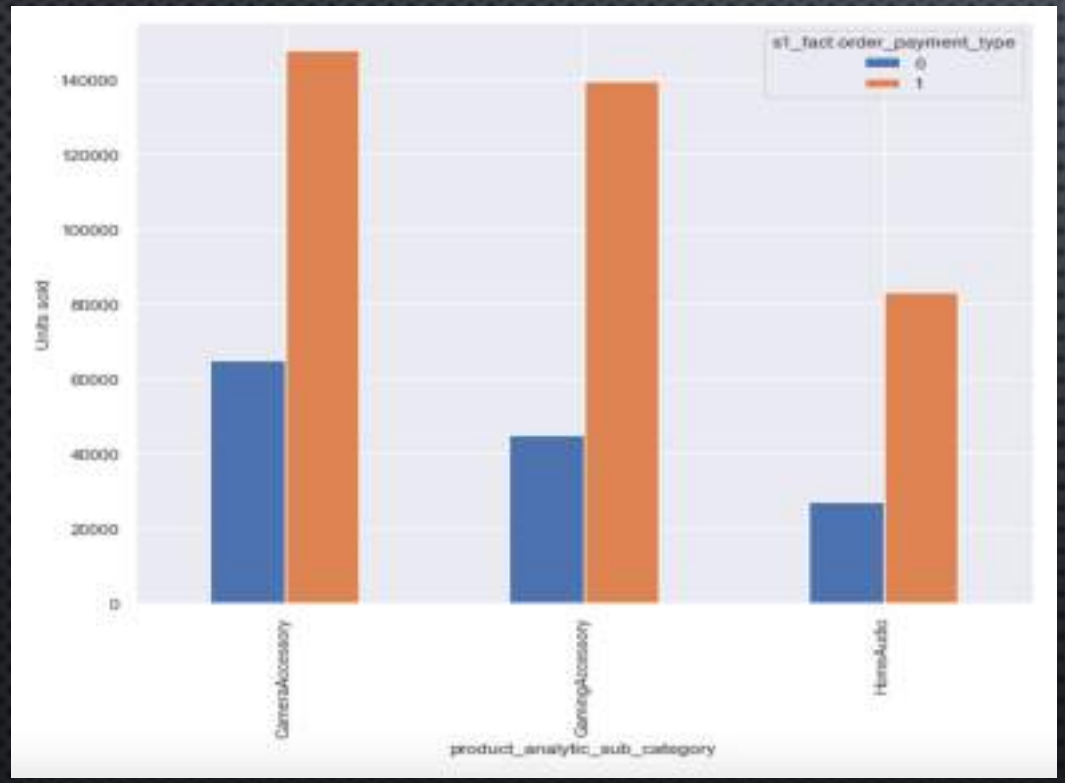


GMV v/s product_analytic_sub_category

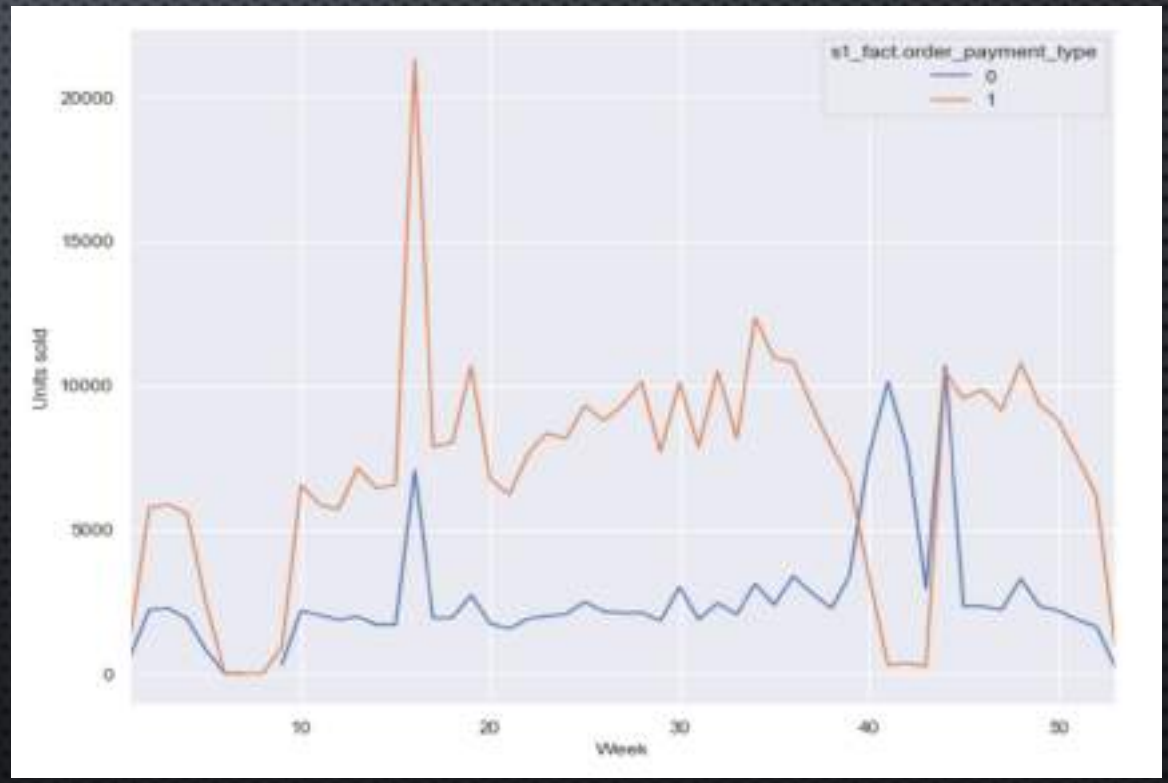
Wednesday and Friday show the highest sales both in terms of 'GMV' and 'Units'

DATA UNDERSTANDING : EDA ANALYSIS

Aggregated orders by different payment types

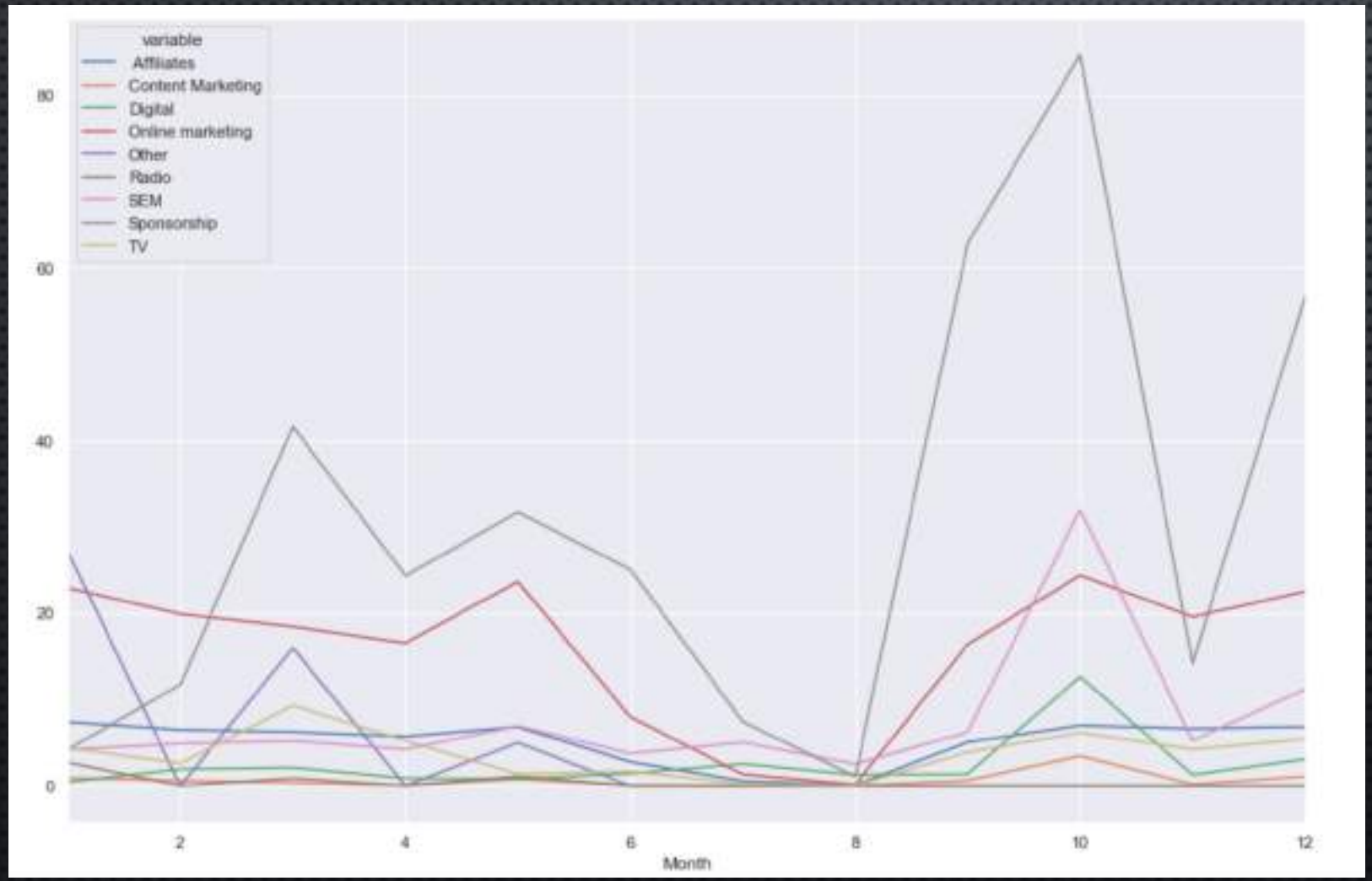


Units sold v/s product_analytic_sub_category



Units sold v/s week

VISUALIZING MEDIA INVESTMENT DATA



Ad Spends are high throughout the year and high during the promotional events especially before Dussehra, Diwali and New Year. This Period also sees the maximum sale of units

DATA UNDERSTANDING : EDA ANALYSIS

MEDIA INVESTMENT CORRELATION

VERY HIGH CORRELATION BETWEEN:

AFFILIATES & ONLINE MARKETING

AFFILIATES & RADIO

RADIO & OTHER

SEM & CONTENT MARKETING

SEM AND DIGITAL

NEGATIVE CORRELATION BETWEEN:

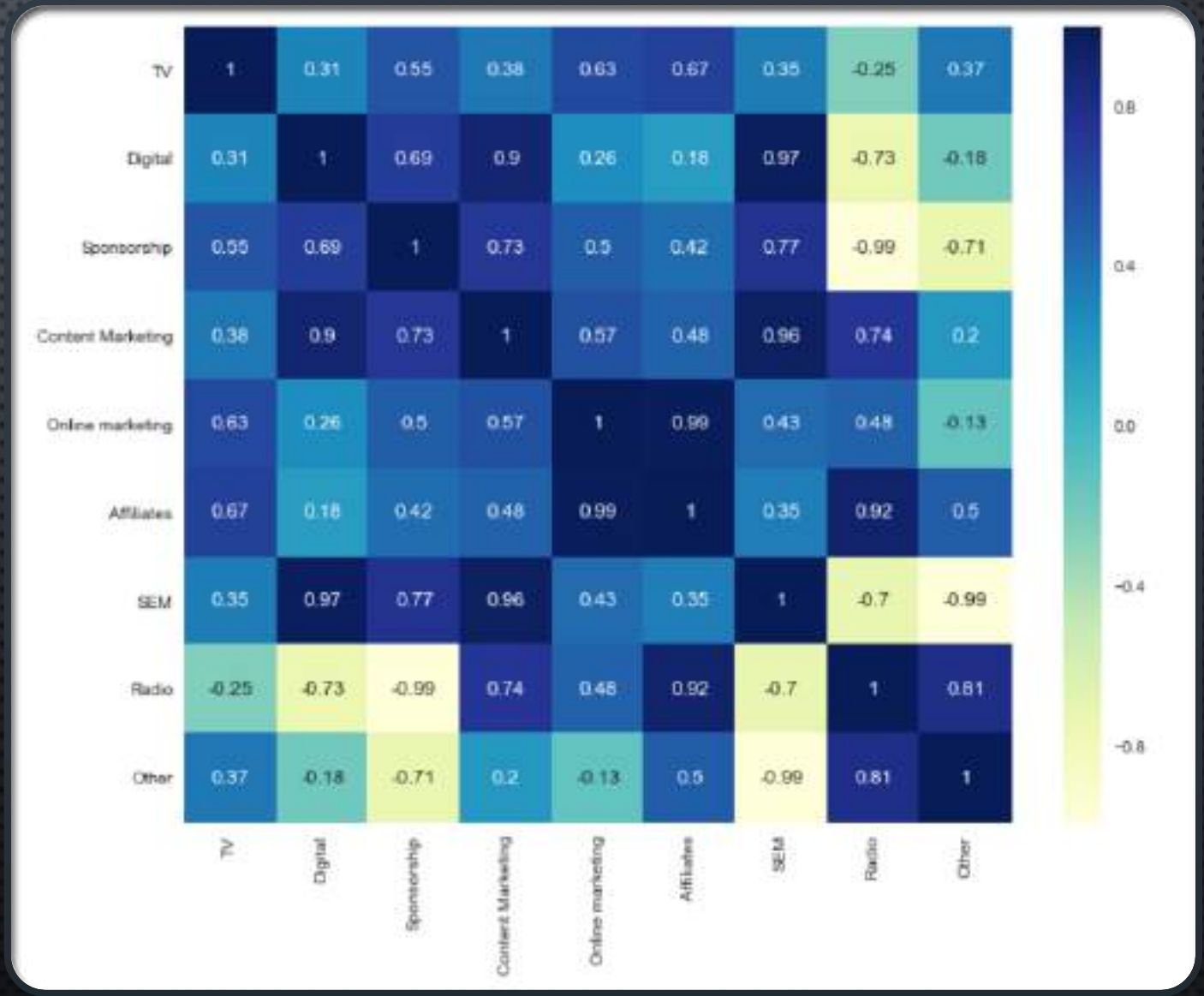
SEM & OTHER

RADIO & SPONSORSHIP

RADIO & SEM

RADIO & TV

OTHER & SPONSORSHIP



DATA PREPARATION DERIVED KPIS

Following variables have been derived by the insights from EDA

1. Average GMV
2. Average MRP
3. List price
4. Lag GMV values for last 3 weeks
5. Lag price values for last 3 weeks
6. Discount over MRP
7. Average number of orders
8. Promotion type
9. Ad stock value
10. Holiday week
11. Delivery statuses(Early, Ontime, Delayed)
12. Ontario Weather Conditions (Min Temp, Max Temp, Rain Flag, Snow Flag)
13. NPS Mean

CAMERA ACCESSORIES – OUTCOME OF 5 MODELS

Model	Significant Variables	Adjusted R Square on Test data set	Residual Mean Squared Error (MSE)
Simple Linear Model	Weekly_investment _Mean+Product Procurement SLA mean+NPS Mean	0.75	0.5
Multiplicative Model	Weekly_investment_tv_mean+product_procurement_sla_mean+tv_adstock	0.70	0.65
Koyck Model	Product_procurement_sla+weekly_investment_radio_mean+sales_flag+unit_sum	0.64	0.45
Distributed Lag Model	Pay_day_flag+Minimum_Temerature+discount_off_percent	0.65	0.32

Best Model = Simple Linear

GAME ACCESSORIES – OUTCOME OF 5 MODELS

Model	Significant Variables	Adjusted R Square on Test data set	Residual Mean Squared Error (MSE)
Simple Linear Model	Weekly_investment_radio mean+product procurement_sla + weekly_investment digital_mean	0.60	0.52
Multiplicative Model	Weekly_investment_affiliate_mean + tv_adstock_mean+nps_mean	0.65	1.24
Koyck Model	Weekly_digital_ad_stock_mean+unit_sum+product_ procurement_sla_mean+weekly_investment_ other_mean	0.77	0.10
Distributed Lag Model	Self_price_inflation+pay_day_flag+gmv_ lag	0.80	0.2

Best Model =Koyck Model

HOME AUDIO – OUTCOME OF 5 MODELS

Model	Significant Variables	Adjusted R Square on Test data set	Residual Mean Squared Error (MSE)
Simple Linear Model	Weekly_investment_digital + weekly_sem_adstock_mean +Sale_flag	61.4	0.11
Multiplicative Model	Weekly_investment_sem_mean+unit_sum+product+procurement_sla_mean	0.68	0.86
Koyck Model	Order_payment_type+weekly_investment_tv_mean+minimum_temperature	0.80	0.11
Distributed Lag Model	Sla_mean+weekly_tv_ad_stock_mean +discount_off_percent	0.80	0.90

Best Model =Koyck Model

Camera Accessories – Recommendations based on Elasticity of KPIs

Weekly investment digital mean spend has positive impact on sales.

One unit of this ad spend will increase the sale by 0.35 units.

Weekly sales has positive impact by a factor of 0.23

Product procurement SLA has the maximum impact with a factor of 1.24

Recommendations:

1) Company should focus on ad spend on digital means.
2) Product procurement should be taken seriously.
3) Better timed weekly sales will have better impact.

GAME ACCESSORIES – RECOMMENDATIONS BASED ON ELASTICITY OF KPIS

Weekly digital ad stock mean has best positive impact on sale

One unit of this ad spend will increase the sale by 0.92 units.

Weekly investment sem has a negative impact with a factor of -0.66

Product procurement sla has the impact with a factor of 0.37

Recommendations:

1) Company should focus on digital ad stocks
2) Sem investments are least effective and don't add up to the revenue.
3) Product procurement has to be taken seriously.

HOME AUDIO – RECOMMENDATIONS BASED ON ELASTICITY OF KPIS

Order payment type impact positively with a factor of 0.38

Weekly investment tv mean impact by a factor of 0.15

Minimum temperature has positive impact of 0.15

Recommendations:

1) More focus should be provided to cash on delivery customer's.
2) Media investment through television channel.
3) More marketing schemes towards colder region might help the sales.

CHALLENGES FACED

- Deciding on number of derived KPI needed which are important as well as logical.
- Arriving at a base dataset because many iterations were performed as the model results were leading to perfectly fitting model.
- To solve this problem, correlation concept was used to remove correlated variables.
- Selection of significant variables which are finally decided by correlation matrix and non zero variance.
- Identifying the functions/ packages to perform ad stock and creating lag variables.
- Limited availability of the sample implementation of Kyock, Lag + Multiplicative models in PYTHON
- Decision on removal of variables from analysis.