CAPSTONE PROJECT : ELECKART MODEL MIX MODELLING

MID –TERM SUBMISSION

Submitted By:

VIKUL AGGARWAL

ANKIT SAHU

BUSINESS OBJECTIVE & DATA UNDERSTANDING

- Eleckart, e-commerce company based out of Canada, needs insight to optimally reallocate the budget optimally across the marketing levers.

- Analysis of Data from July 2015 to June 2016.

- Product categories – Camera Accessory, Game Accessory and Home Audio

- Sub – Goals:
  - Performance driven
  - ROI of marketing spend – Impact Analysis
  - Optimization of Marketing spend.

PROBLEM SOLVING METHODOLOGY

- The methodology for analysis is in accordance with the CRISP DM Framework.

- Business Understanding – Help the CFO decide the marketing budget on different levers.

- Data Understanding – Understanding the monthly spend on advertisements, online campaigns, days of special sale, Stock Index of the company, Monthly NPS score

- Data Preparation – Remove discrepancies in data to build effective models.

- Modeling – Build various Market Mix models with the help of data prepared. Find the best model on basis of factors like Adjusted R-Square, MSE, etc.

- Evaluation - Using the best model, evaluate the important variables that effect the ROI and make recommendations.

- Deployment – Final Phase of the cycle. Use the model in real time for business.

DATA PREPARATION
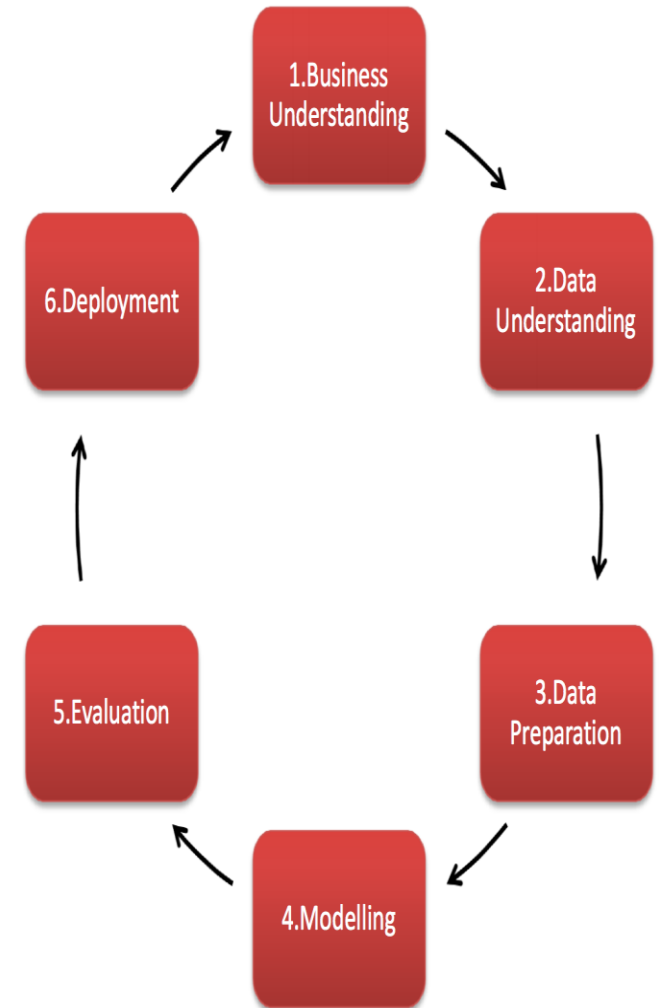
- Found and treated **leading white spaces** in Orders data GMV Column

- Handled **negative values** in columns like product_procurement_sla, deliverybdays and deliverycdays

- Handling **extreme large values** for product_procurement_sla.

- **"\N" values** in deliverybdays and deliverycdays were imputed with 0.

- Incorrect GMV values (gmv>product_mrp*units) have been treated by imputing the MRP with GMV/unit

- **99283 rows were duplicate**. Dropped them

- Some 'id' columns which did not bring significance to the analysis were dropped.

- Selected data of the time period **July 2015 – June 2016** to conform to the need of the problem statement.

- **Binary encoding** for categorical variables with two levels.

- **One hot encoding** for categorical variables with more than 2 level.

- Column with singlt unique value has been dropped.

- Three separate dataframes for three product sub-categories : Camera Accessory, Home Audio and Gaming Accessory.

- Convert daily order data to Weekly Level by aggregation of the numeric variable based on the week number of the year.

- Merge of order dataset with other secondary dataframes.

- Split of data into train and test for all the three product sub-categories.

FEATURE ENGINEERING

- List Price → GMV/Units (Gross Merchandise Value/Units)

- Discount → MRP - List Price

- Discount_Percentage → (Discount / List Price) * 100

- Mov. Average of List Price, Discount and MRP

- payDayWeek : If Payday falls in that week, payDayWeek = 1 else 0.

- holidayWeek : If any holiday in that week, holidayWeek = 1 else 0.

- ProductType – Luxury / Mass-market

- Exponential Moving Average for 8 weeks for Advertising Media Channels.

- AdStock values for all Advertising Media.

- paymentModeIndicator : Created from original feature for the algorithm to understand.

EXPLORATORY DATA ANALYSIS

Fig 1. → Finding the number of transactions by each Category

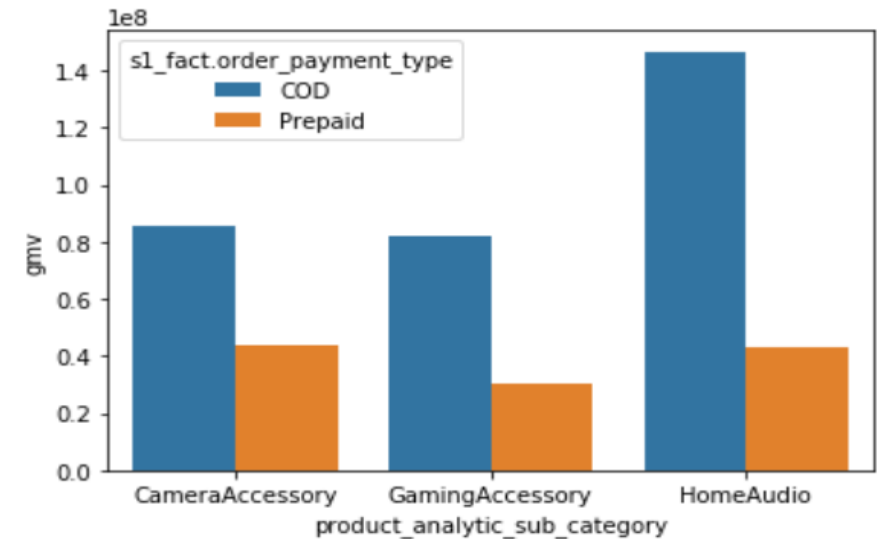Camera Accessory>Gaming Accessory > Home Audio
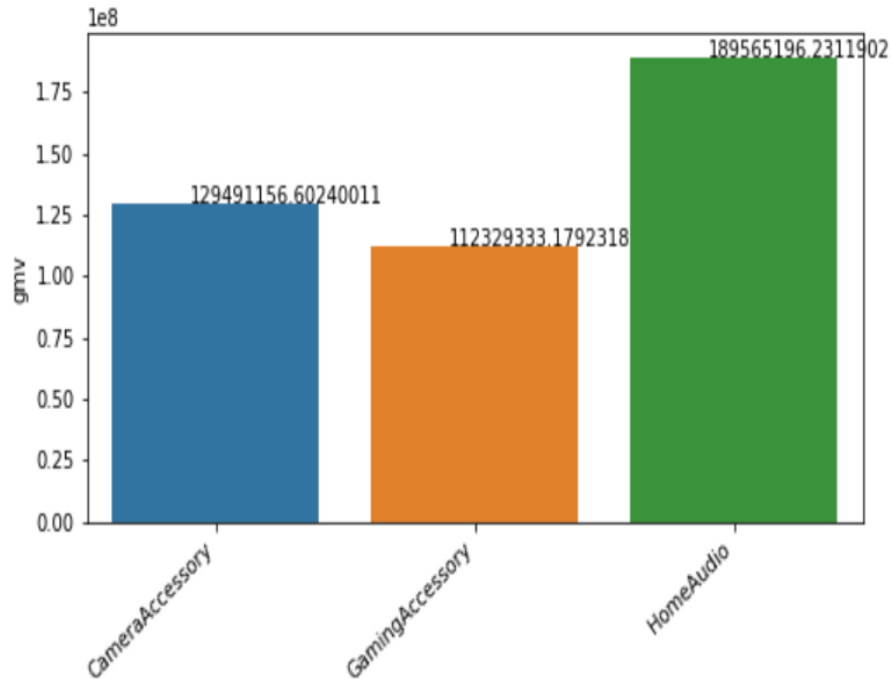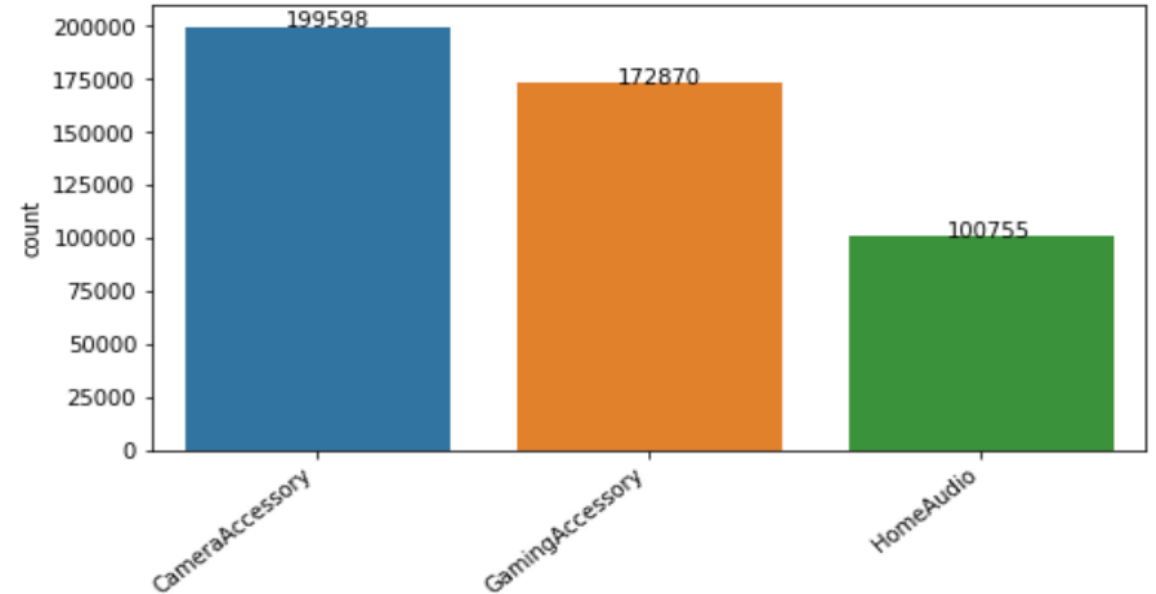
FIG 2. Total GMV by Product subcategory.
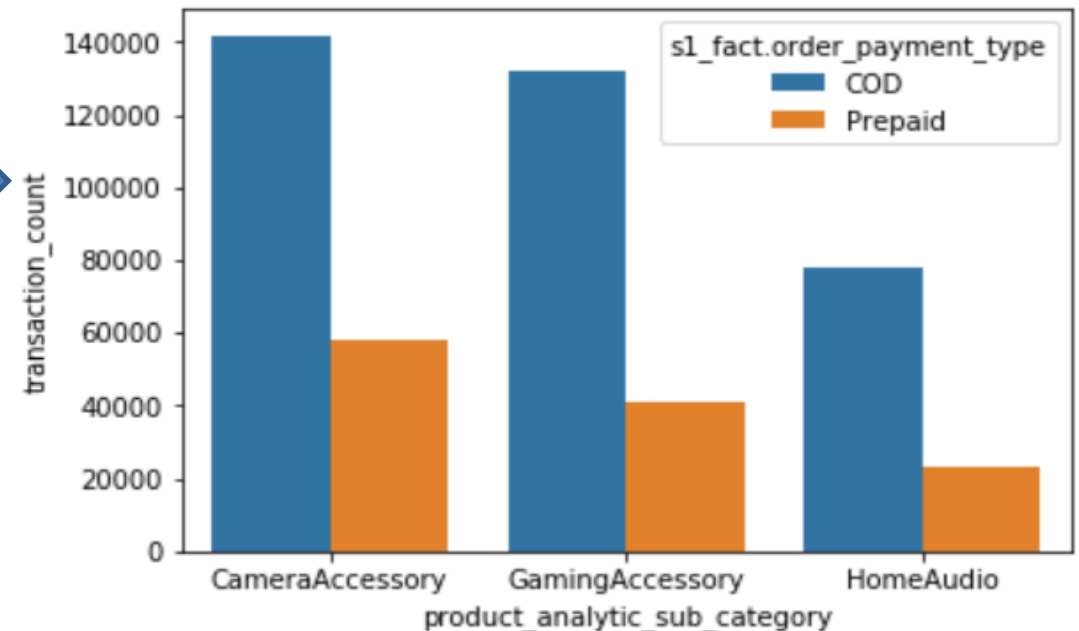Home Audio > Camera Accessory > Gaming Accessory

EDA CONTD.

Total GMV vs Payment Type

For COD – Max revenue from Home Audio
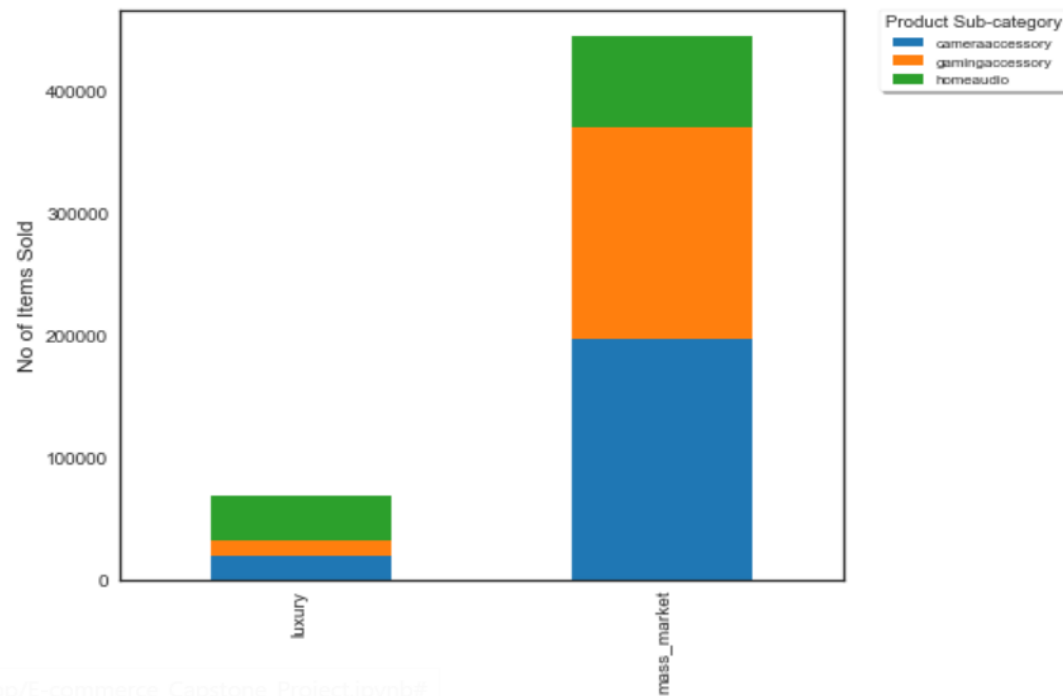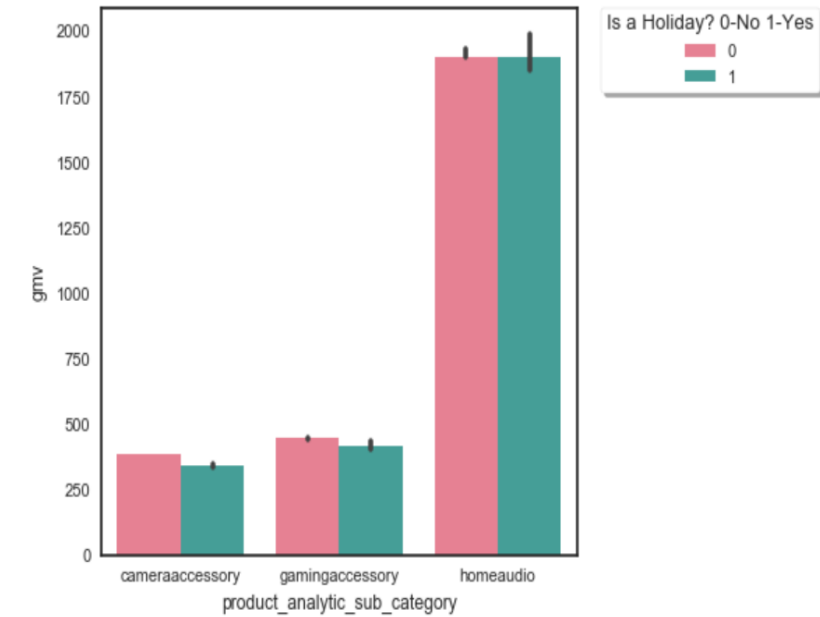For Prepaid - - Max revenue from Camera Accessory

Order Count v/s Payment Type

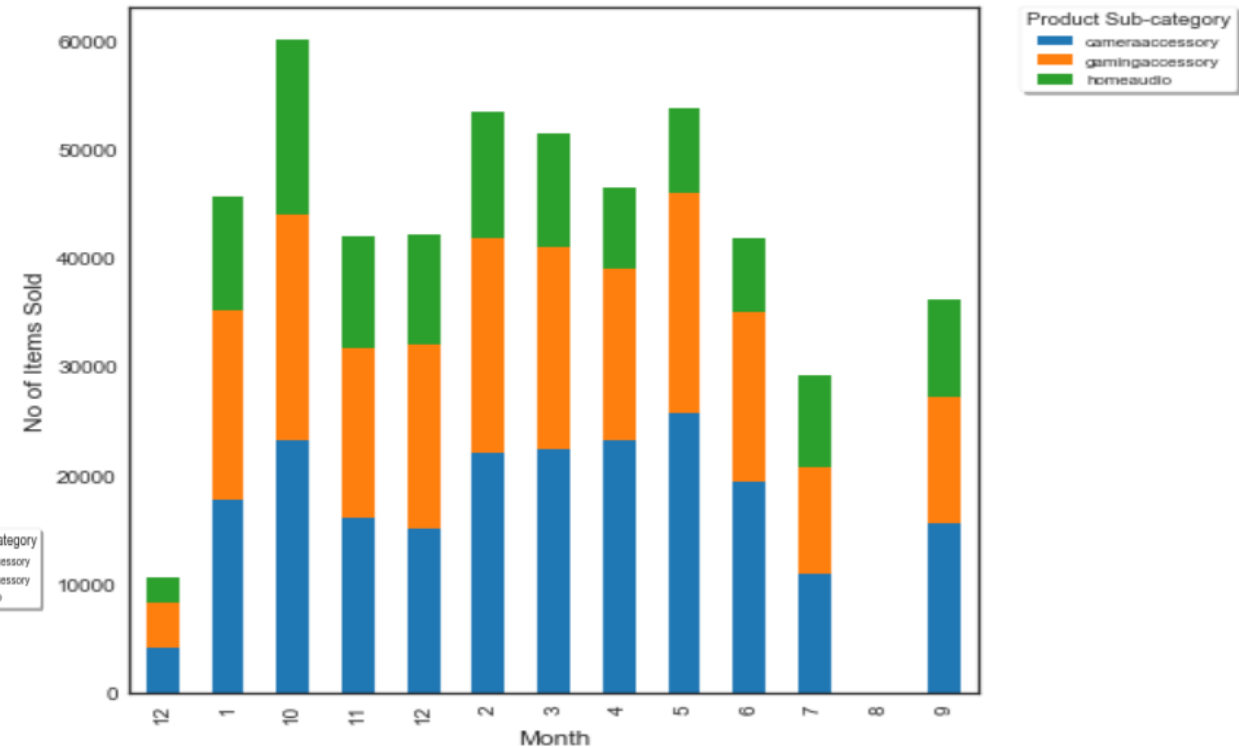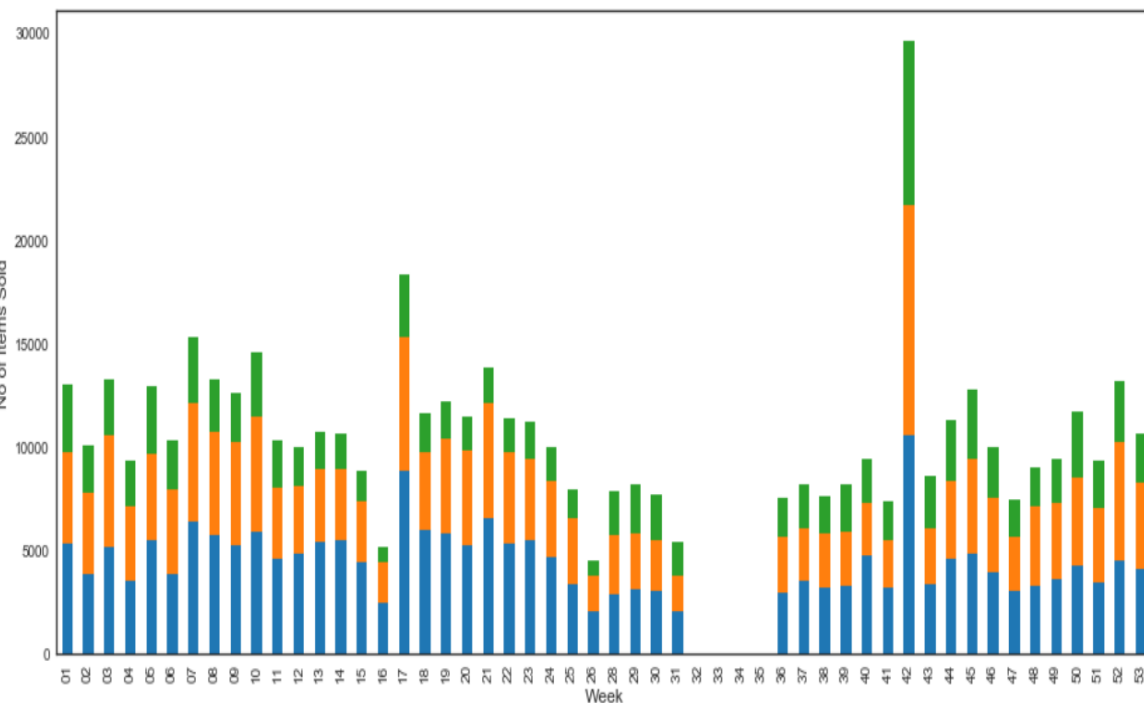Max Orders are through COD mode for all the product subcategories.

Average Revenue on Holidays/non holidays
for three product subcategories

Number of Items(Luxury/Mass-
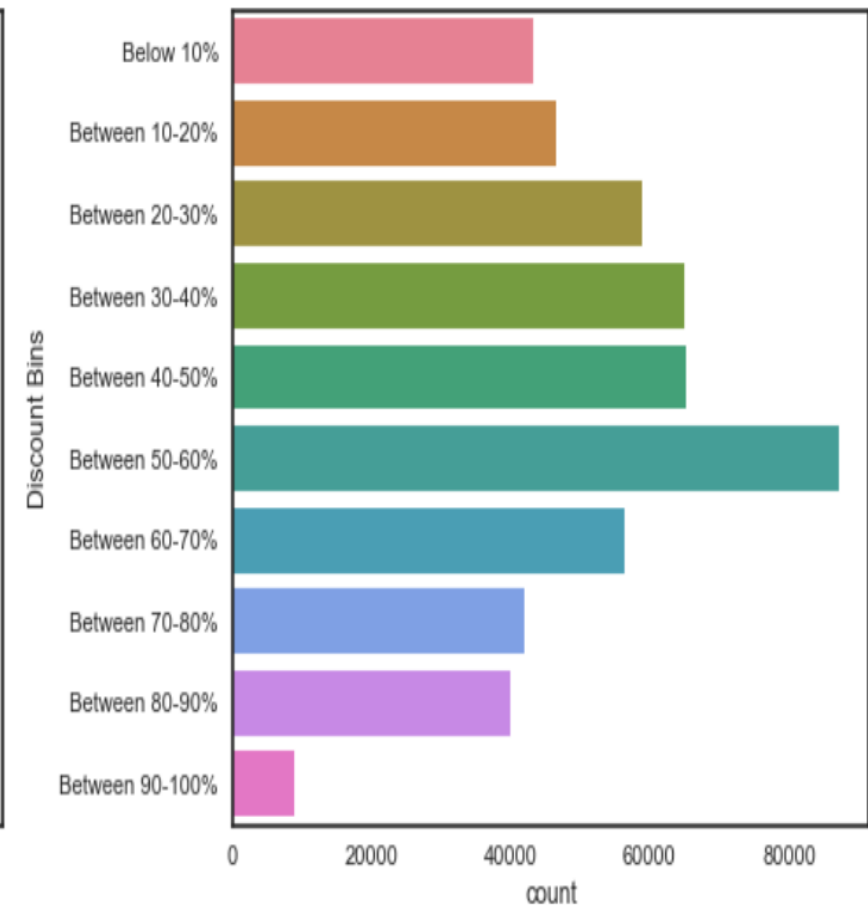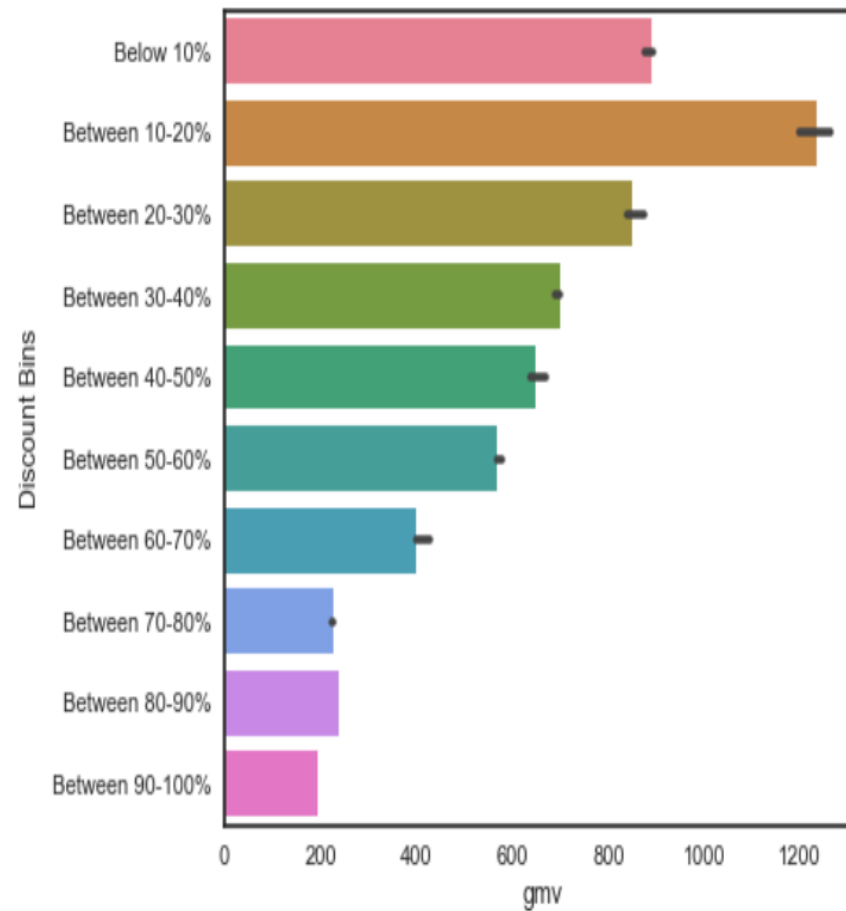market) sold per 3 product sub-
categories

Number of Items sold per subcategory per month
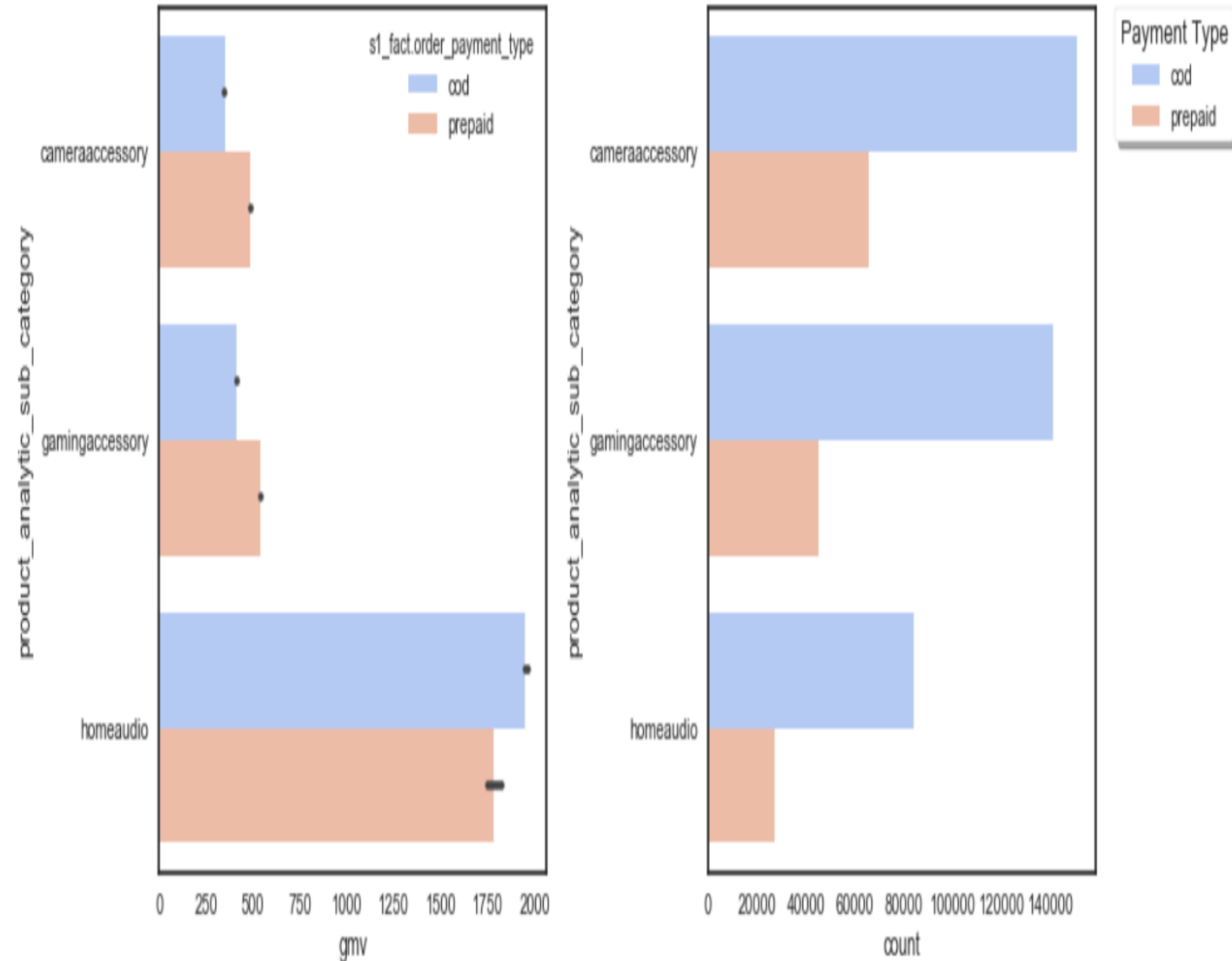
Number of Items sold per subcategory per Week

Variation of sales amount and revenue based on percentage of Discount

Variation of sales amount and revenue based on Payment Types.

Analysis – Median revenue from prepaid accounts is more than from COD products for all categories except Home Audio. This is observed even after the number of products sold is way higher in case of COD for all the sub-categories.
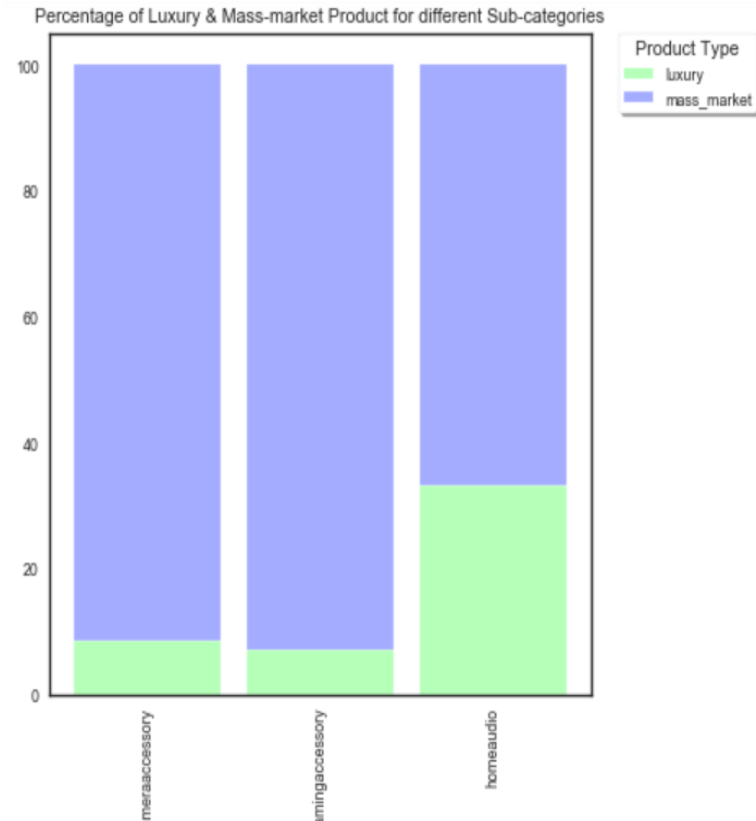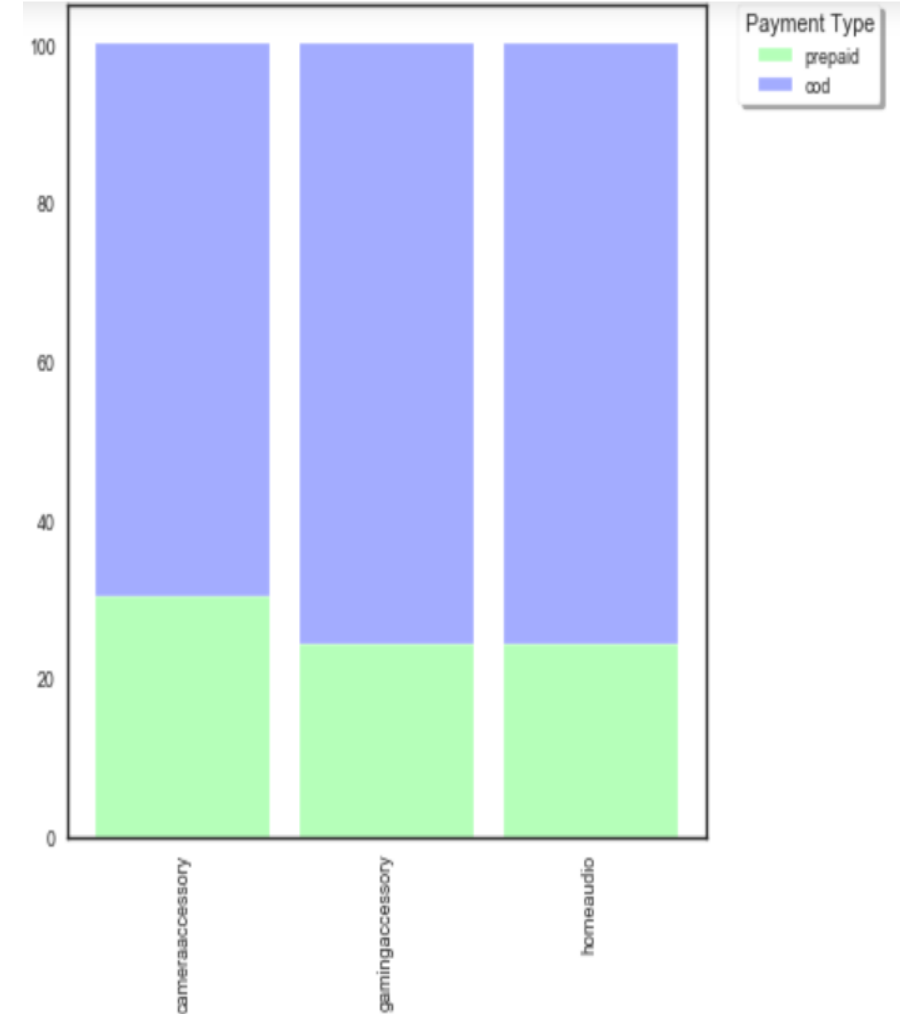
Number Of Items sold at different Discount Percentage.

Analysis – Most of the sale happens when the discount percentage is between 50-60 %.
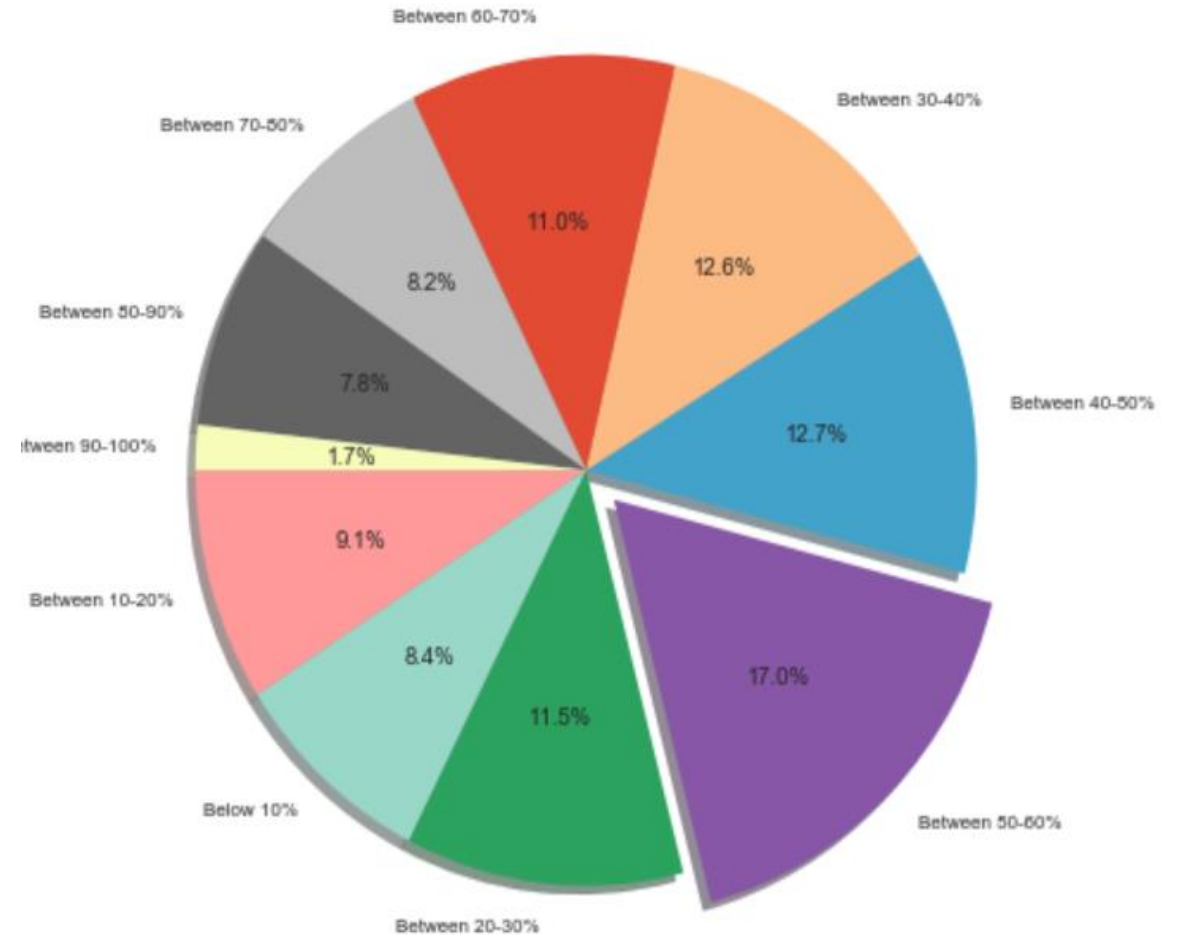
Percentage of COD and Prepaid transaction type for all the sub categories.

Analysis – Percentage of prepaid payments under camera accessory was observed to be slightly higher when compared to other categories.
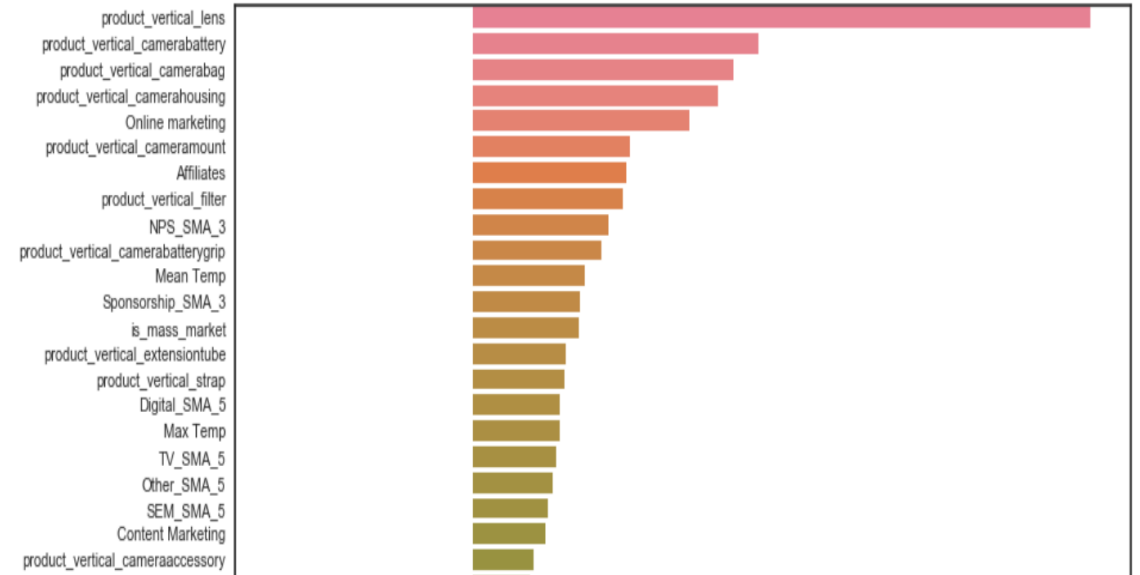

Percentage of Luxury and Mass market products for all the sub categories.

Analysis – Percentage of luxury product in home audio is comparatively higher when compared to others

Percentage of Luxury & Mass-market Product for different Sub-categories

LINEAR REGRESSION MODEL – CAMERA ACCESSORY

R2 SCORE → 0.83
MSE → 0.17

FIVE MOST IMPORTANT VARIABLES
AFFECTING GMV FOR CAMERA ACCESSORY

Features	Coefficients
product_vertical_lens	0.432
product_vertical_camerabattery	0.200
product_vertical_camerabag	0.182
product_vertical_camerahousing	0.171
Online marketing	0.151

LINEAR REGRESSION MODEL – GAMING ACCESSORY

R2 SCORE → 0.51
MSE → 0.49
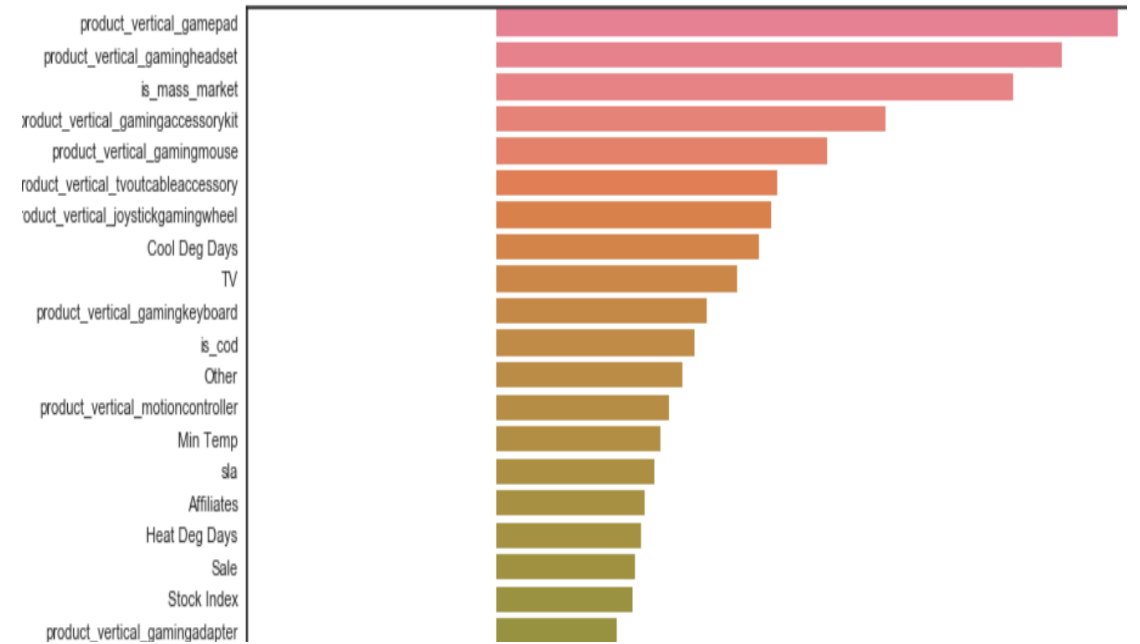
TOP 5 FEATURES AFFECTING THE
REVENUE IN GAMING ACCESSORIES

	Features	Coefficients
9	product_vertical_gamepad	0.201
13	product_vertical_gamingheadset	0.183
7	is_mass_market	0.167
10	product_vertical_gamingaccessorykit	0.126
16	product_vertical_gamingmouse	0.107

LINEAR REGRESSION MODEL – GAMING ACCESSORY

R2 Square → 0.73
MSE → 0.27
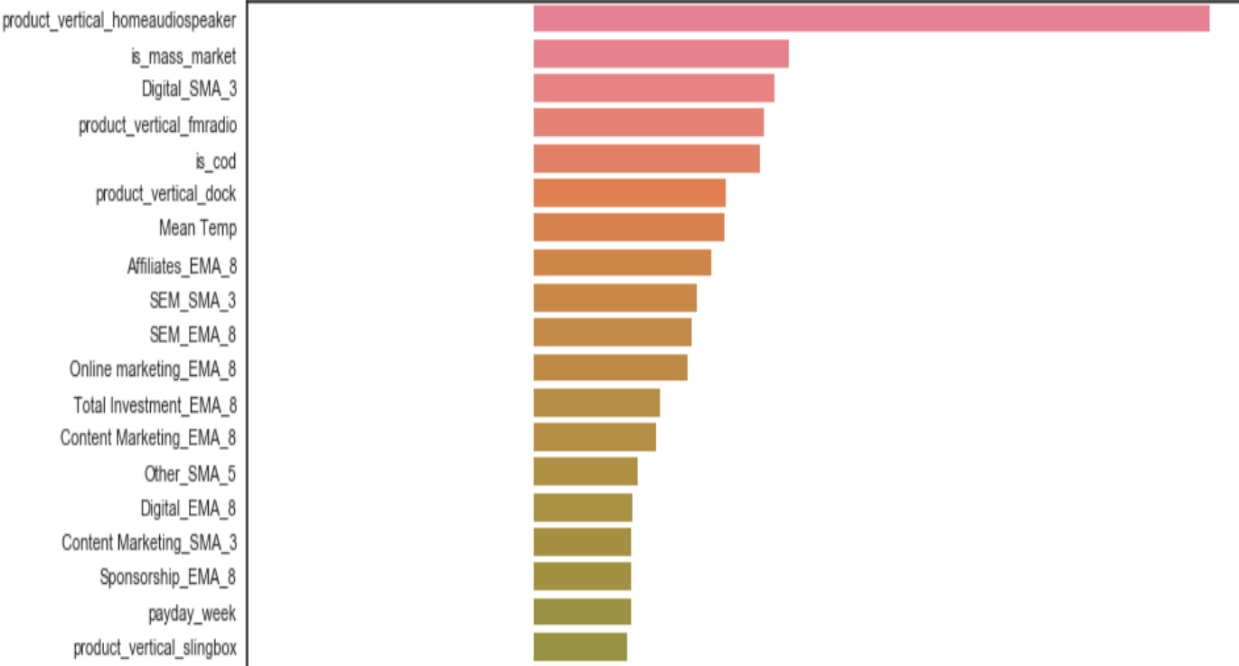
TOP 5 FEATURES FOR REVENUE
GENERATION IN HOME AUDIO

	Features	Coefficients
13	product_vertical_homeaudiospeaker	0.392
7	is_mass_market	0.148
31	Digital_SMA_3	0.139
11	product_vertical_fmradio	0.133
6	is_cod	0.131

FUTURE ROADMAP

- Building various other models like Multiplicative Model, Koyck Model, Distributive Lag Model for all of the three product subcategories.

- The models would then be used to make predictions and recommendations for helping the company to mark their budget accordingly to get the best ROI.

- Presentation of final results would be given in the End Term Capstone Submission.