Group Case Study – E-Commerce

By – Saransh Jain and Tanya Chaudhary

Objective

- ElecKart is an e-commerce firm based out of Ontario, Canada specializing 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.
- Being a part of the marketing team working on budget optimization. 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.

Index

- Data Understanding
- Data Preparation
- Exploratory Data Analysis – I
- Exploratory Data Analysis – II
- Modelling

1. Data Understanding

Importing libraries and csv file

```
# import python libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import statsmodels.api as sm
import sklearn
import squarify
from datetime import date, datetime

from statsmodels.stats.outliers_influence import variance_inflation_factor
from sklearn.linear_model import LinearRegression
from sklearn.preprocessing import StandardScaler, MinMaxScaler
from sklearn.feature_selection import RFE
from sklearn.model_selection import cross_val_score
from sklearn.model_selection import train_test_split
from sklearn.metrics import r2_score
```

```
elec_data.head()
```

	fsn_id	order_date	Year	Month	order_id	order_item_id	gmv	units	deliverybdays	deliverycdays	s1_fact.order_payment_type	sla	
0	ACCCX3S58G7B5F6P	2015-10-17 15:11:54	2015	10	3.419301e+15	3.419301e+15	6400.0	1	NaN	NaN	COD	5	-1
1	ACCCX3S58G7B5F6P	2015-10-19 10:07:22	2015	10	1.420831e+15	1.420831e+15	6900.0	1	NaN	NaN	COD	7	-8
2	ACCCX3S5AHMF55FV	2015-10-20 15:45:56	2015	10	2.421913e+15	2.421913e+15	1990.0	1	NaN	NaN	COD	10	-1
3	ACCCX3S5AHMF55FV	2015-10-14 12:05:15	2015	10	4.416592e+15	4.416592e+15	1690.0	1	NaN	NaN	Prepaid	4	-7
4	ACCCX3S5AHMF55FV	2015-10-17 21:25:03	2015	10	4.419525e+15	4.419525e+15	1618.0	1	NaN	NaN	Prepaid	6	2

2. Data Preparation

- **Removing duplicate** rows where order_date, order_id, order_item_id and units are same

```
# duplicate rows removal from the elec_data
elec_data = elec_data.drop_duplicates(subset = ["order_date", "order_id", "order_item_id" ,"units"], keep = 'first', inplace = Fa
```

- Shape of the data frame after dropping duplicate rows

```
# checking the shape of the dataframe
elec_data.shape
```

```
(1536296, 20)
```

- Converting **order_date** to date-time formate

```
# converting order_date column to datetime format
elec_data['order_datetime'] = pd.to_datetime(elec_data['order_date'], format='%Y-%m-%d %H:%M:%S')
```

- Start date and End date of the transactions are:

```
# checking the starting & ending order dates in the provided data
print("Start date of transaction: ",min(elec_data.order_date))
print("End date of transaction: ", max(elec_data.order_date))
```

```
Start date of transaction:  2015-05-19
End date of transaction:  2016-07-25
```

As mentioned in the problem statement, we have to use data from July 2015 to June 2016. So, we'll filter out the rest of the data.

Creating new column **order_week** which will contain the information about the week number when the order is placed

```
# creating new column i.e. order_week which would contain week number when order is placed
elec_data['order_week'] = elec_data['order_datetime'].dt.week
```

Checking **percentage of null values** present in the data

```
deliverybdays                    78.29
deliverycdays                    78.29
product_analytic_vertical         0.38
pincode                           0.26
cust_id                           0.26
gmv                               0.26
order_week                        0.00
order_date                        0.00
Year                              0.00
Month                             0.00
order_id                          0.00
order_item_id                     0.00
units                             0.00
s1_fact.order_payment_type        0.00
order_datetime                    0.00
sla                               0.00
product_analytic_super_category   0.00
product_analytic_category         0.00
product_analytic_sub_category     0.00
product_mrp                       0.00
product_procurement_sla           0.00
fsn_id                            0.00
dtype: float64
```

- **Removing deliverybdays and deliverycdays** as they contain more than 70% of null values
- **Dropping rows which are null in product_analytic_vertical**

Again checking the null values in the data frame

```
gmv                                 0.26
order_week                          0.00
order_datetime                      0.00
order_date                          0.00
Year                                0.00
Month                               0.00
order_id                            0.00
order_item_id                       0.00
units                               0.00
s1_fact.order_payment_type          0.00
sla                                 0.00
cust_id                             0.00
pincode                             0.00
product_analytic_super_category     0.00
product_analytic_category           0.00
product_analytic_sub_category       0.00
product_analytic_vertical           0.00
product_mrp                         0.00
product_procurement_sla             0.00
fsn_id                              0.00
dtype: float64
```

There are **1265** transaction where **GMV** is zero, we'll drop them because:

- GMV is our **target variable**
- **Imputing with 1 won't be beneficial**
- Model will **mislead** if we try imputing using MRP and units

There are **34606** record where GMV is **greater than** (MRP * Number of units sold), therefore **removing them** as well

Adding **new logic** to the data frame:

```python
# Function to change order_week for the months of the year 2016 to make them in continuation with the
# week number of December 2015 for analysis purpose.
def mapOrderWeek(order_week, year):
    if((year == 2016) & (order_week <= 26)):
         return order_week+53
    else:
        return order_week
```

```python
# Mapping order_week according to the new logic
elec_data['order_week'] = elec_data.apply(lambda x: mapOrderWeek(x['order_week'],x['Year']), axis = 1)
```

Checking unique values in **order_week** column

```
array([27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43,
       44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60,
       61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77,
       78, 79])
```

Also, there is no any product present in the data where **MRP** is zero

Checking **outliers** if any

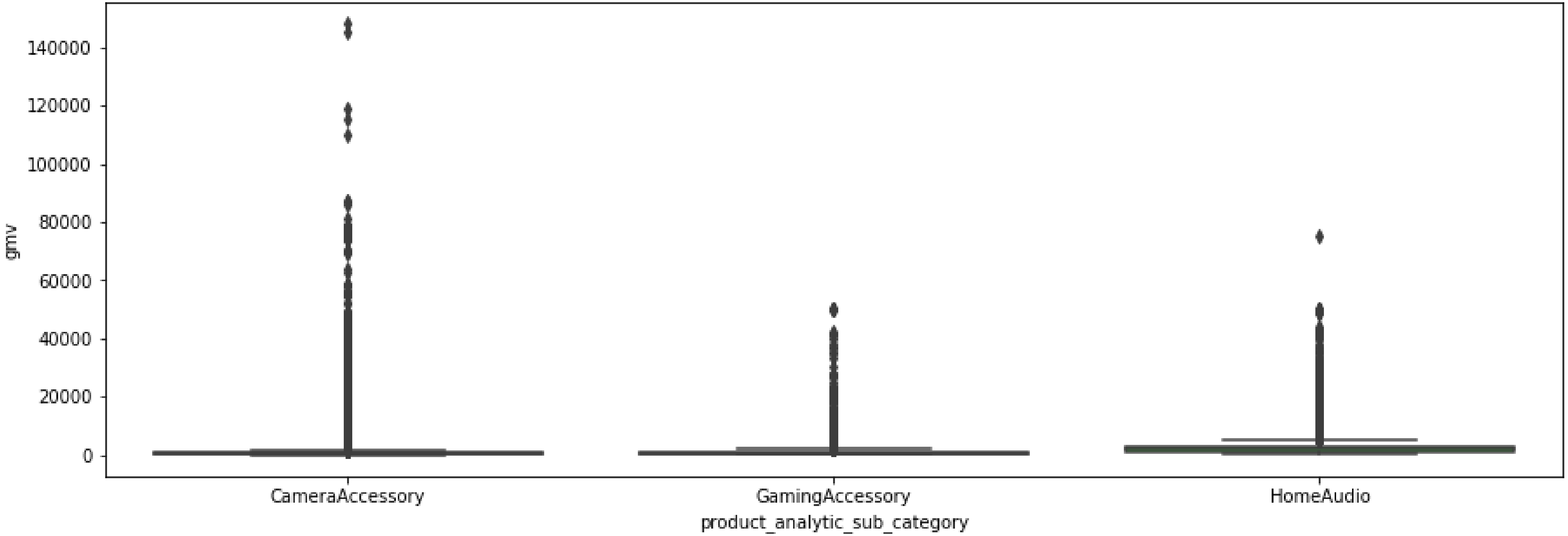
	Year	Month	gmv	units	sla	product_mrp	product_procurement_sla	order_week
count	1.490019e+06	1.490019e+06	1.490019e+06	1.490019e+06	1.490019e+06	1.490019e+06	1.490019e+06	1.490019e+06
mean	2.015556e+03	6.411632e+00	2.495538e+03	1.021619e+00	5.683975e+00	4.214486e+03	5.269346e+00	5.556354e+01
std	4.968865e-01	3.690315e+00	5.677618e+03	2.504660e-01	2.998176e+00	8.686505e+03	5.209425e+01	1.354709e+01
min	2.015000e+03	1.000000e+00	1.000000e+01	1.000000e+00	0.000000e+00	4.900000e+01	-1.000000e+00	2.700000e+01
25%	2.015000e+03	3.000000e+00	3.390000e+02	1.000000e+00	4.000000e+00	8.000000e+02	1.000000e+00	4.400000e+01
50%	2.016000e+03	6.000000e+00	7.500000e+02	1.000000e+00	6.000000e+00	1.599000e+03	2.000000e+00	5.600000e+01
75%	2.016000e+03	1.000000e+01	1.999000e+03	1.000000e+00	7.000000e+00	3.499000e+03	3.000000e+00	6.700000e+01
max	2.016000e+03	1.200000e+01	2.269470e+05	5.000000e+01	1.006000e+03	2.999990e+05	1.000000e+03	7.900000e+01

- There are **0.0058%** of rows present where **SLA > 30 days**, therefore removing those rows

- There are **4.2%** of negative **product_procurement_sla** and **1.9%** of negative product_procurement_sla for the categories – **GamingAccessory, CameraAccessory and HomeAudio,** therefore removing negative product_procurement_sla

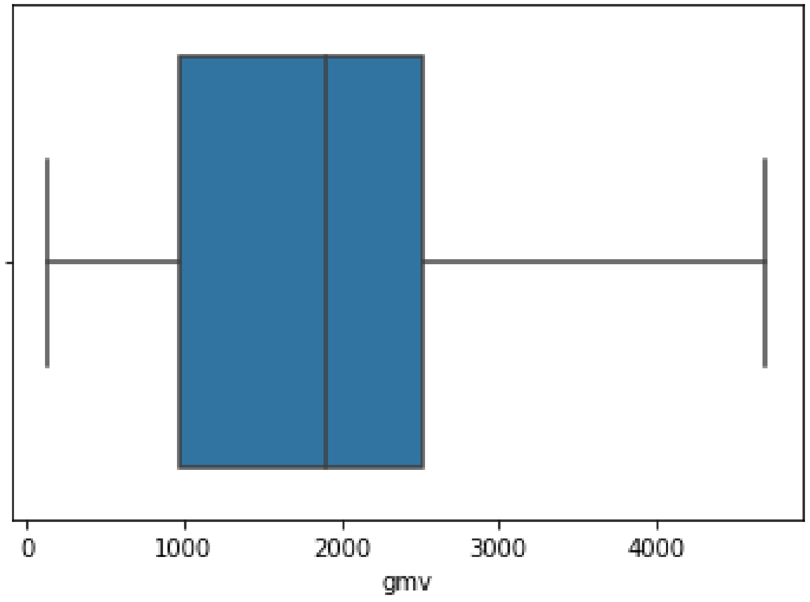
- The number of rows for **product_procurement_sla > 30 days** and **product_procurement_sla > 15 days** is same i.e., **4065,** so we are removing the rows having **product_procurement_sla > 15 days**

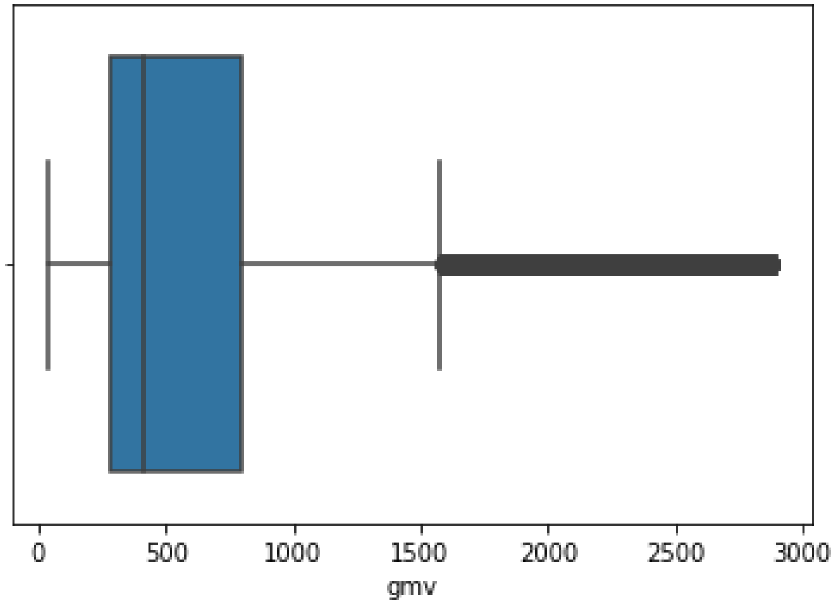
Checking outliers in **GMV** for the three categories :

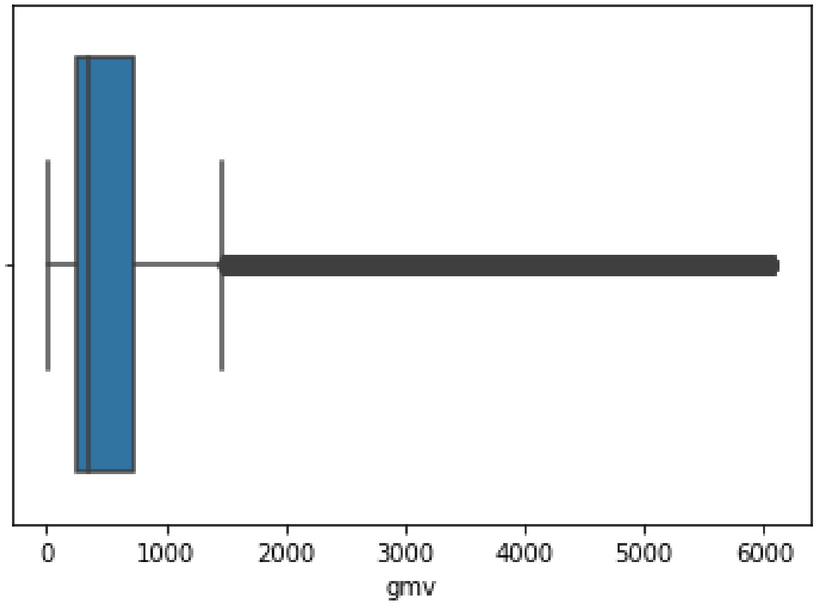


There are many outliers present in **GMV,** therefore treating them

Removing the outliers from the **HomeAudio, GamingAccessory and CameraAccessory** beyond 0.95 quartile range





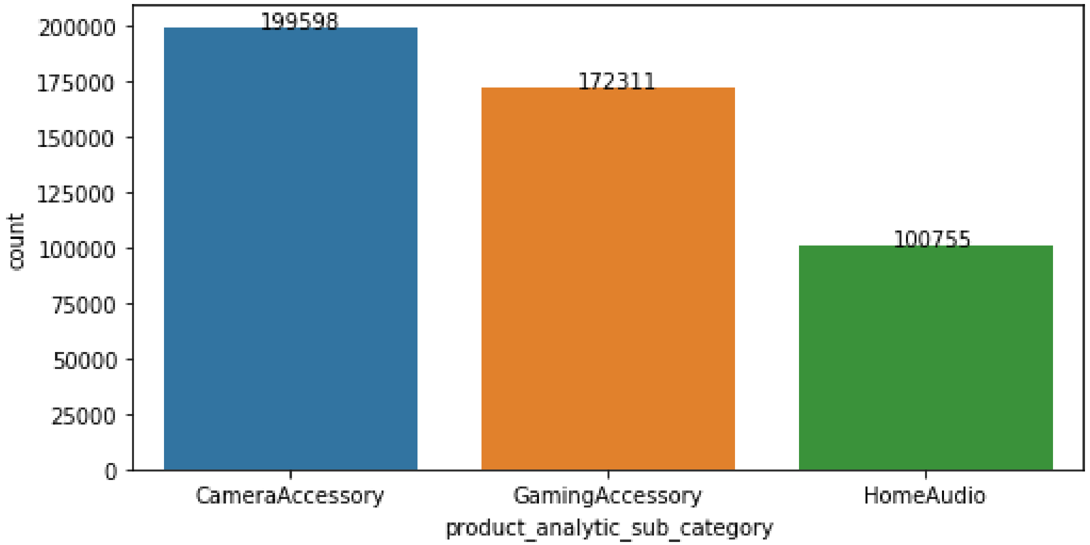


Dropping the **unnecessary columns** such as 'fsn_id', 'order_item_id', 'pincode', 'product_analytic_super_category‘ and 'product_analytic_category'

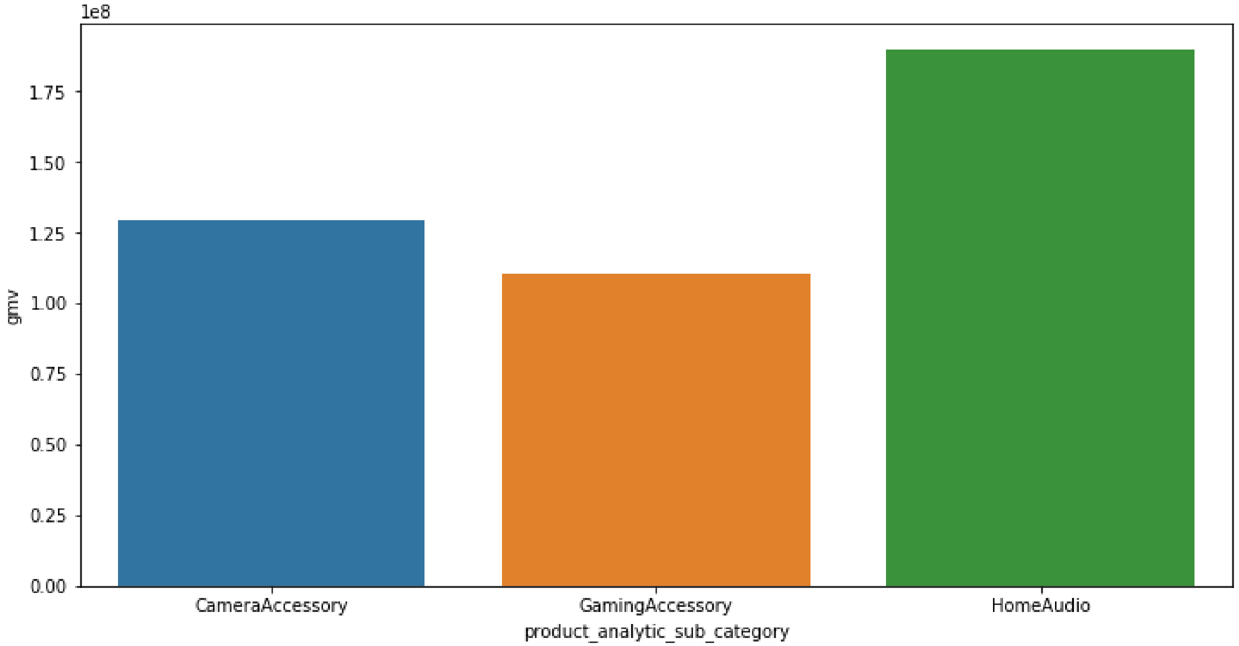
Changing datatypes of ‘Year’, ‘Month’ and ‘order_week’ to integer type

3. Exploratory Data Analysis - I

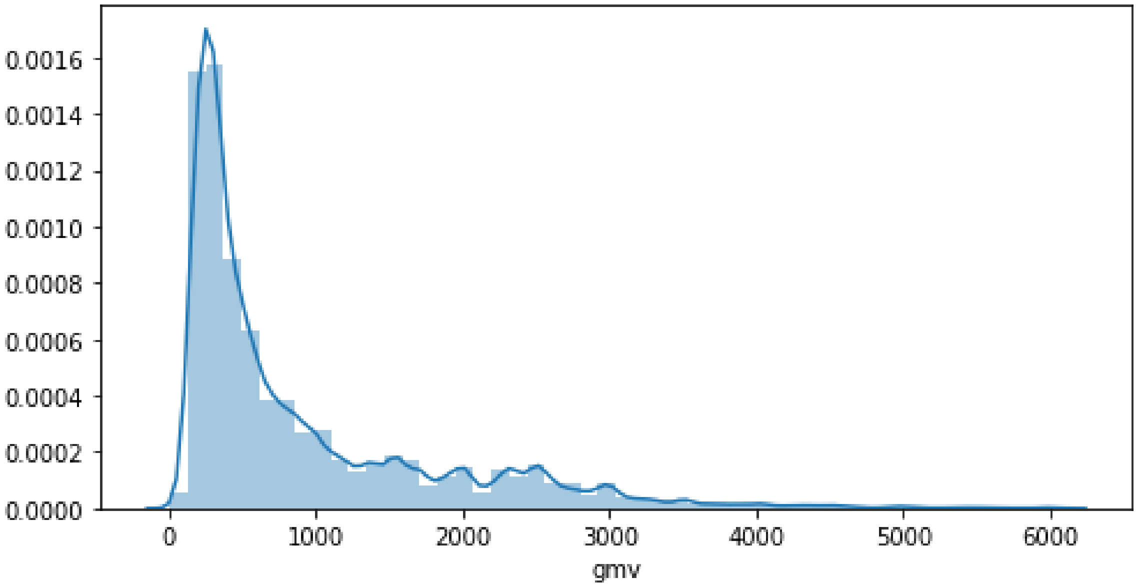
Total **count** of the three categories : **CameraAccessory, GamingAccessory and HomeAudio**



Total **GMV count** for each of the three category

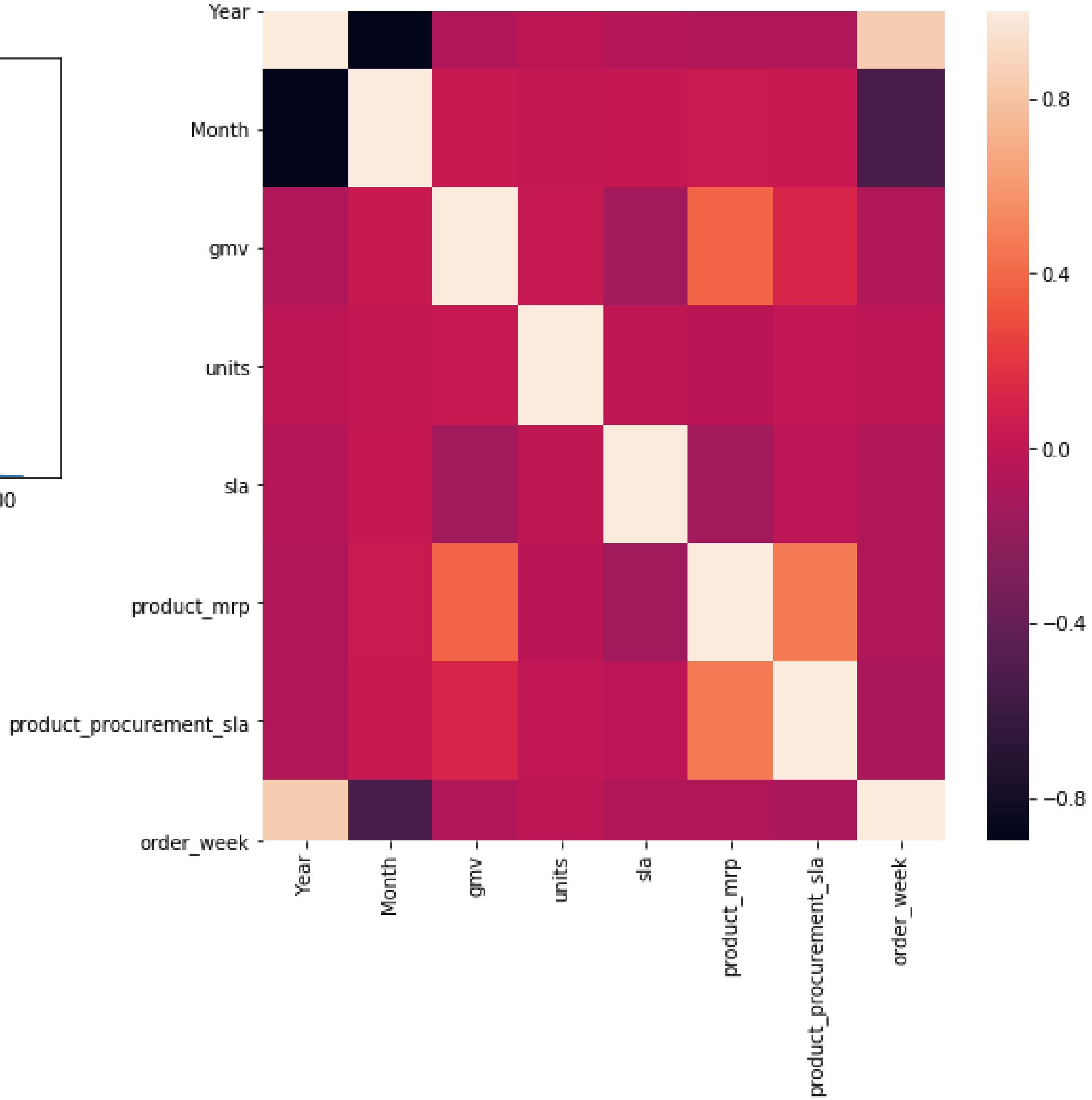


Frequency of **maximum transaction range**

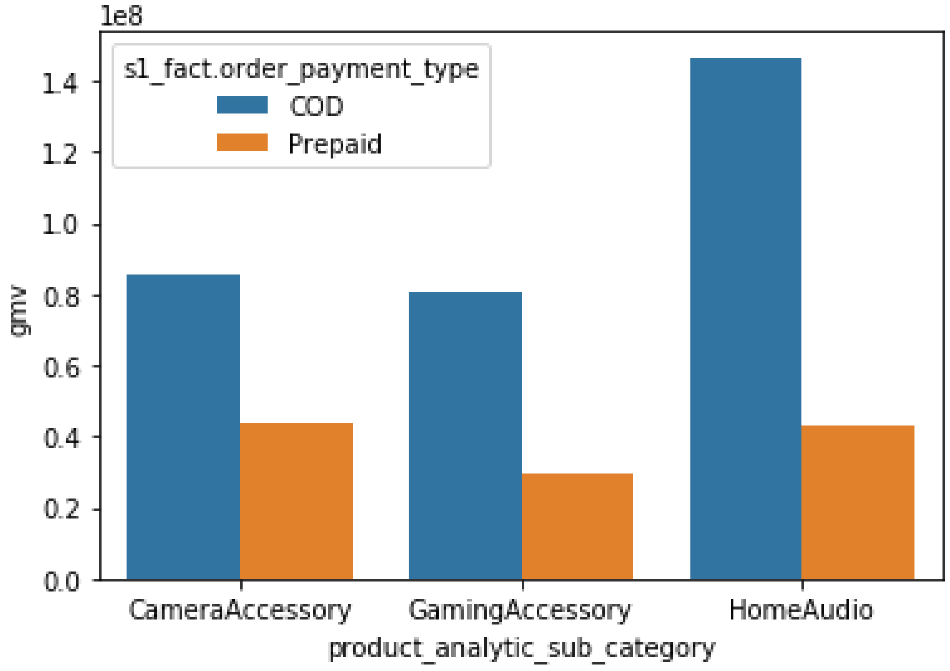


Most of the products lies in the range 0 to 1000

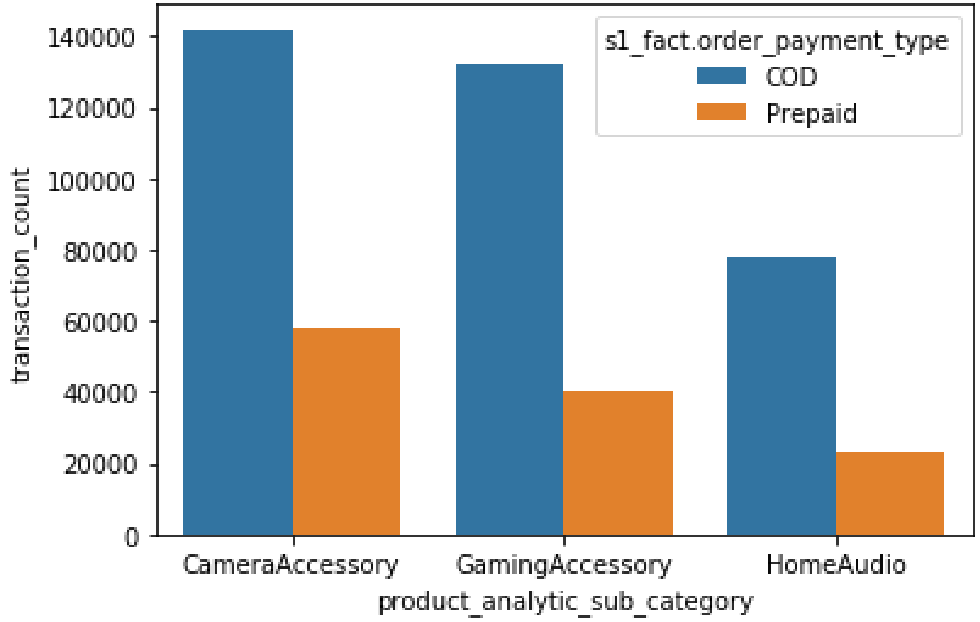
Correlation between different variables



Total **GMV versus Payment Type**



Transaction total versus Payment type



For all the 3 categories, maximum orders are by COD

For COD orders, maximum revenue : HomeAudio > CameraAccessory > GamingAccessory

For Prepaid orders, maximum revenue : CameraAccessory > HomeAudio > GamingAccessory

Creating KPIs for Holidays and Special days

```
# Function to map dataframe containing date, week of a given dictionary, for example: holidays, special_days
def createDataFrameFromDictionary(listOfValues, desiredColumnName):
    df = pd.DataFrame(listOfValues)
    df[desiredColumnName] = pd.to_datetime(df[desiredColumnName]).dt.date
    df["total_"+desiredColumnName] = 1
    df['order_week'] = pd.to_datetime(df[desiredColumnName]).dt.week
    df['year'] = pd.to_datetime(df[desiredColumnName]).dt.year
    df['order_week'] = df.apply(lambda x: mapOrderWeek(x['order_week'],x['year']), axis = 1)
    df = df.groupby('order_week').sum()
    df["is_"+desiredColumnName] = 1
    df.drop('year', axis = 1, inplace = True)
    return df
```

```
#  Taking the special sale days from the data provided
special_sale_days = {'special_sale_day':["2015-07-18","2015-07-19","2015-08-15",
                     "2015-08-16","2015-08-17","2015-08-28",
                     "2015-08-29","2015-08-30","2015-10-15",
                     "2015-10-16","2015-10-17","2015-11-07","2015-11-08","2015-11-09","2015-11-10",
                     "2015-10-11","2015-10-12","2015-11-13","2015-11-14","2015-12-25","2015-12-26",
                     "2015-12-27","2015-12-28","2015-12-29","2015-12-30","2016-01-01","2016-01-02",
                     "2016-01-03","2016-01-20","2016-01-21","2016-01-22","2016-02-01","2016-02-02",
                     "2016-02-20","2016-02-21","2016-02-14","2016-02-15","2016-03-07","2016-03-08",
                     "2016-03-09","2016-05-25","2016-05-26","2016-05-27"]}
special_sale_df = createDataFrameFromDictionary(special_sale_days,'special_sale_day')
```

```
special_sale_df.shape
```

```
(16, 2)
```

Creating **KPIs such as payment_mode_indicator, selling_price, discounts** for the three data frame

```
def createKPIs(df):
#### KPI 1: Payment mode indicator
    df['payment_mode_indicator'] = df['s1_fact.order_payment_type'].apply(lambda x: 0 if x == "COD" else 1)

#### KPI 2: Selling price of each product in every transaction
    df['selling_price'] = df['gmv']/df['units']

#### KPI 3: Discount offered on the products
    df['discount'] = np.round((df['product_mrp']-df['selling_price'])/df['product_mrp'],2)

#### KPI 4: Percent of prepaid orders
    weekly_aggregated_data = df
    weekly_aggregated_data['order_count'] = 1
    weekly_aggregated_data = weekly_aggregated_data.groupby('order_week').sum()
    weekly_aggregated_data['percentage_prepaid_transactions'] = np.round((weekly_aggregated_data.payment_mode_indicator/weekly_ag
    weekly_aggregated_data = weekly_aggregated_data.reset_index()[['order_week','percentage_prepaid_transactions']]
    df = pd.merge(df,weekly_aggregated_data, how = 'inner', on = 'order_week')
    return df
```

```
# Adding the KPIs to the dataframes of each category
GamingAccessory_data_KPIs = createKPIs(GamingAccessory_data)
CameraAccessory_data_KPIs = createKPIs(CameraAccessory_data)
HomeAudio_data_KPIs = createKPIs(HomeAudio_data)
```

```
print(HomeAudio_data_KPIs.shape)
print(GamingAccessory_data_KPIs.shape)
print(CameraAccessory_data_KPIs.shape)
```

```
(100755, 19)
(172311, 19)
(199598, 19)
```

Creating **KPIs for Prepaid/COD transactions for Visualizations**

```python
# Creating KPIs for Prepaid Count, COD Count & Percent Online Orders for GamingAccessory_data_KPIs
cod_cnt = GamingAccessory_data_KPIs.loc[
    GamingAccessory_data_KPIs['s1_fact.order_payment_type'] == "COD"
].groupby('order_week')['s1_fact.order_payment_type'].count().reset_index().rename(
    columns={'s1_fact.order_payment_type':'COD_count'}
)


prepaid_cnt = GamingAccessory_data_KPIs.loc[
    GamingAccessory_data_KPIs['s1_fact.order_payment_type'] == "Prepaid"
].groupby('order_week')['s1_fact.order_payment_type'].count().reset_index().rename(
    columns={'s1_fact.order_payment_type':'Prepaid_count'}
)


online_orders_GA = pd.merge(
    cod_cnt,
    prepaid_cnt,
    on = "order_week",
    how = "outer"
).fillna(0)

online_orders_GA["pct_online_transactions"] = (
    online_orders_GA["Prepaid_count"]/
    (online_orders_GA["Prepaid_count"]+online_orders_GA["COD_count"])
).round(2)
```

Including **KPIs of NPS and Stock Index**

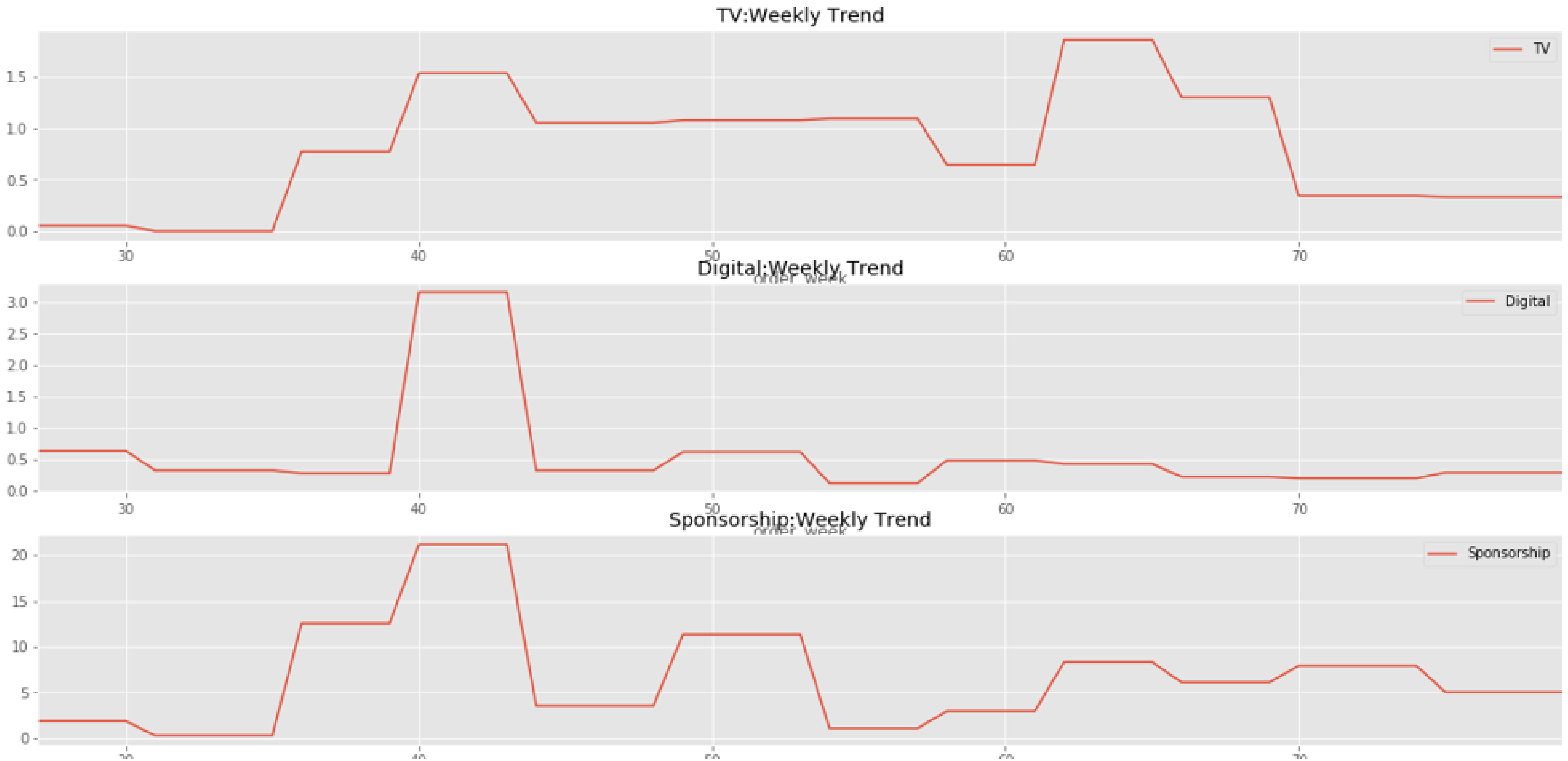
	NPS	Stock Index	Year	Month
0	54.599588	1177	2015	7
1	59.987101	1206	2015	8
2	46.925419	1101	2015	9
3	44.398389	1210	2015	10
4	47.000000	1233	2015	11
5	45.800000	1038	2015	12
6	47.093031	1052	2016	1
7	50.327406	1222	2016	2
8	49.020550	1015	2016	3
9	51.827605	1242	2016	4
10	47.306951	1228	2016	5
11	50.516687	1194	2016	6

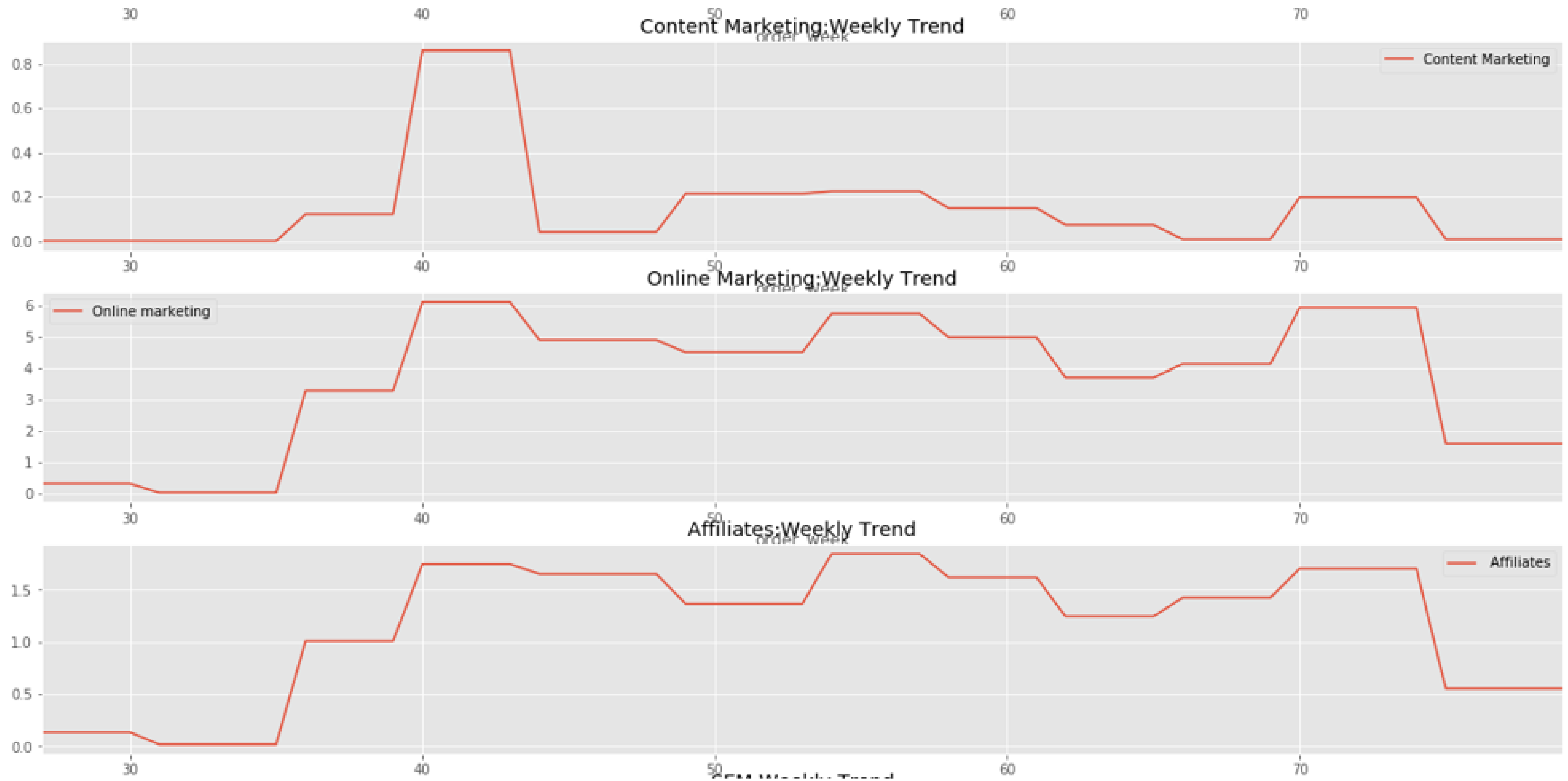
Including **KPIs of Media Investment Data**

```
# reading the media data from the provided file
media_investment = pd.read_excel("Media data and other information.xlsx", sheet_name = 'Media Investment', header = 2)
media_investment
```

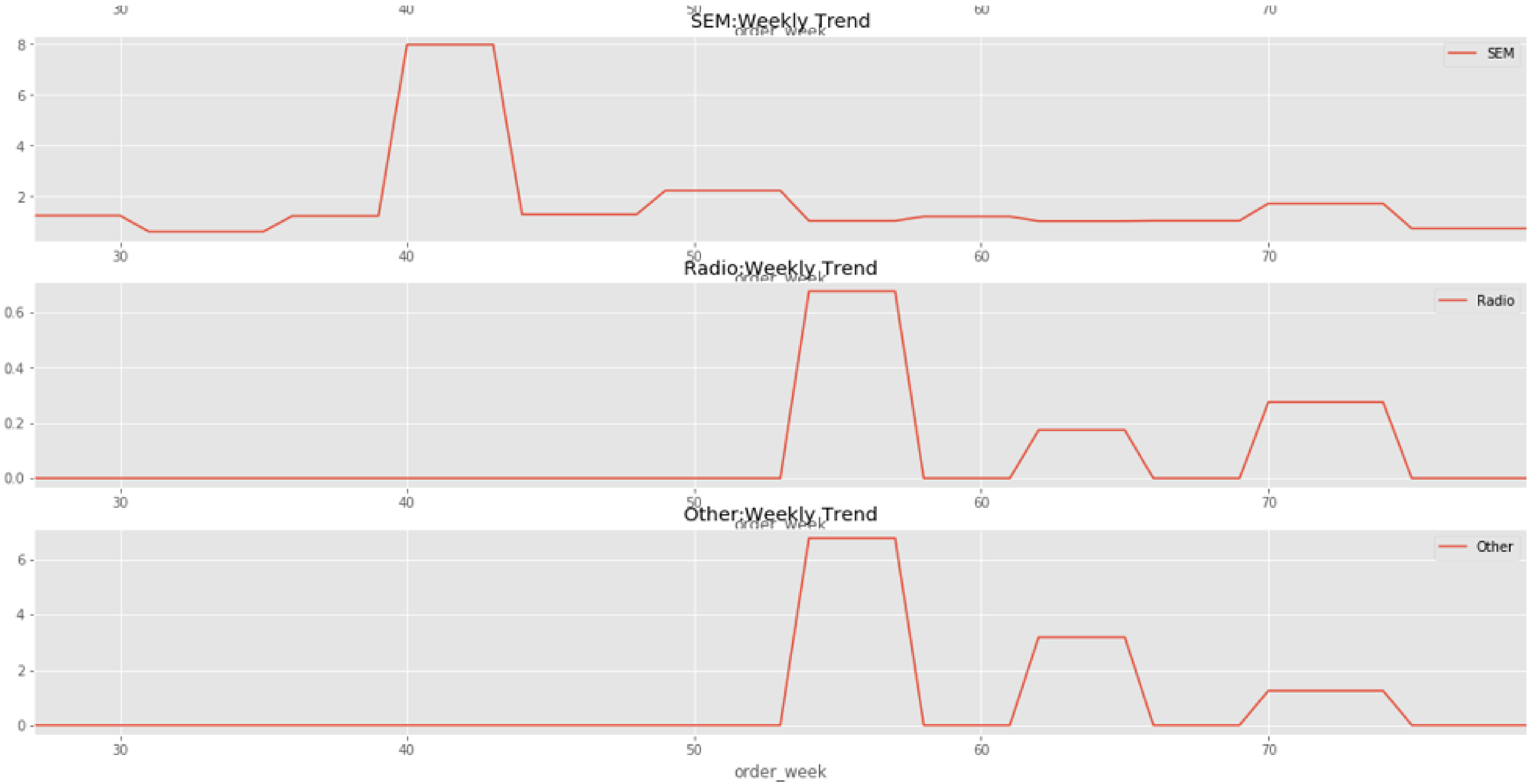
	Unnamed: 0	Year	Month	Total Investment	TV	Digital	Sponsorship	Content Marketing	Online marketing	Affiliates	SEM	Radio	Other
0	NaN	2015.0	7.0	17.061775	0.215330	2.533014	7.414270	0.000933	1.327278	0.547254	5.023697	NaN	NaN
1	NaN	2015.0	8.0	5.064306	0.006438	1.278074	1.063332	0.000006	0.129244	0.073684	2.513528	NaN	NaN
2	NaN	2015.0	9.0	96.254380	3.879504	1.356528	62.787651	0.610292	16.379990	5.038266	6.202149	NaN	NaN
3	NaN	2015.0	10.0	170.156297	6.144711	12.622480	84.672532	3.444075	24.371778	6.973711	31.927011	NaN	NaN
4	NaN	2015.0	11.0	51.216220	4.220630	1.275469	14.172116	0.168633	19.561574	6.595767	5.222032	NaN	NaN
5	NaN	2015.0	12.0	106.745312	5.397502	3.063360	56.705419	1.067307	22.503756	6.826938	11.181030	NaN	NaN

Analyzing **weekly spend** on different channels





Content Marketing:Weekly Trend
Content Marketing
Online Marketing:Weekly Trend
Online marketing
Affiliates:Weekly Trend
Affiliates



The maximum investments occurred between week 39-45 for most channels, Maximum investments was done through Sponsorship media

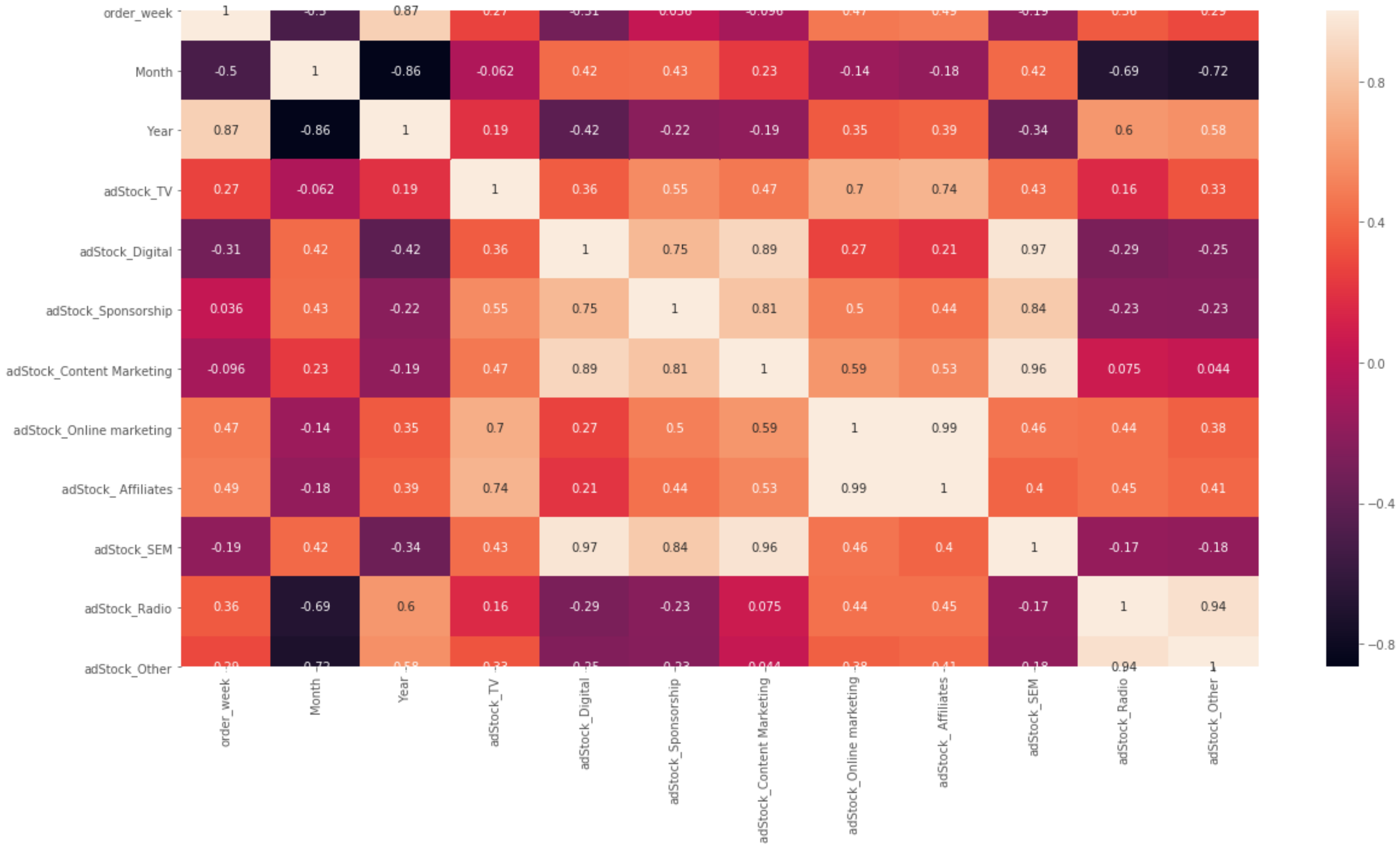
Creating **Adstocks** at the adstock_rate and merging with media investment data

	adStock_TV	adStock_Digital	adStock_Sponsorship	adStock_Content Marketing	adStock_Online marketing	adStock_ Affiliates	adStock_SEM	adStock_Radio	order_week	adStock_Other
0	0.053833	0.633253	1.853567	0.000233	0.331819	0.136813	1.255924	0.0	27	0.0
1	0.080749	0.949880	2.780351	0.000350	0.497729	0.205220	1.883887	0.0	28	0.0
2	0.094207	1.108193	3.243743	0.000408	0.580684	0.239424	2.197868	0.0	29	0.0
3	0.100936	1.187350	3.475439	0.000437	0.622161	0.256525	2.354858	0.0	30	0.0
4	0.052078	0.913194	2.003552	0.000220	0.343392	0.146684	1.805811	0.0	31	0.0

Correlation between adstocks

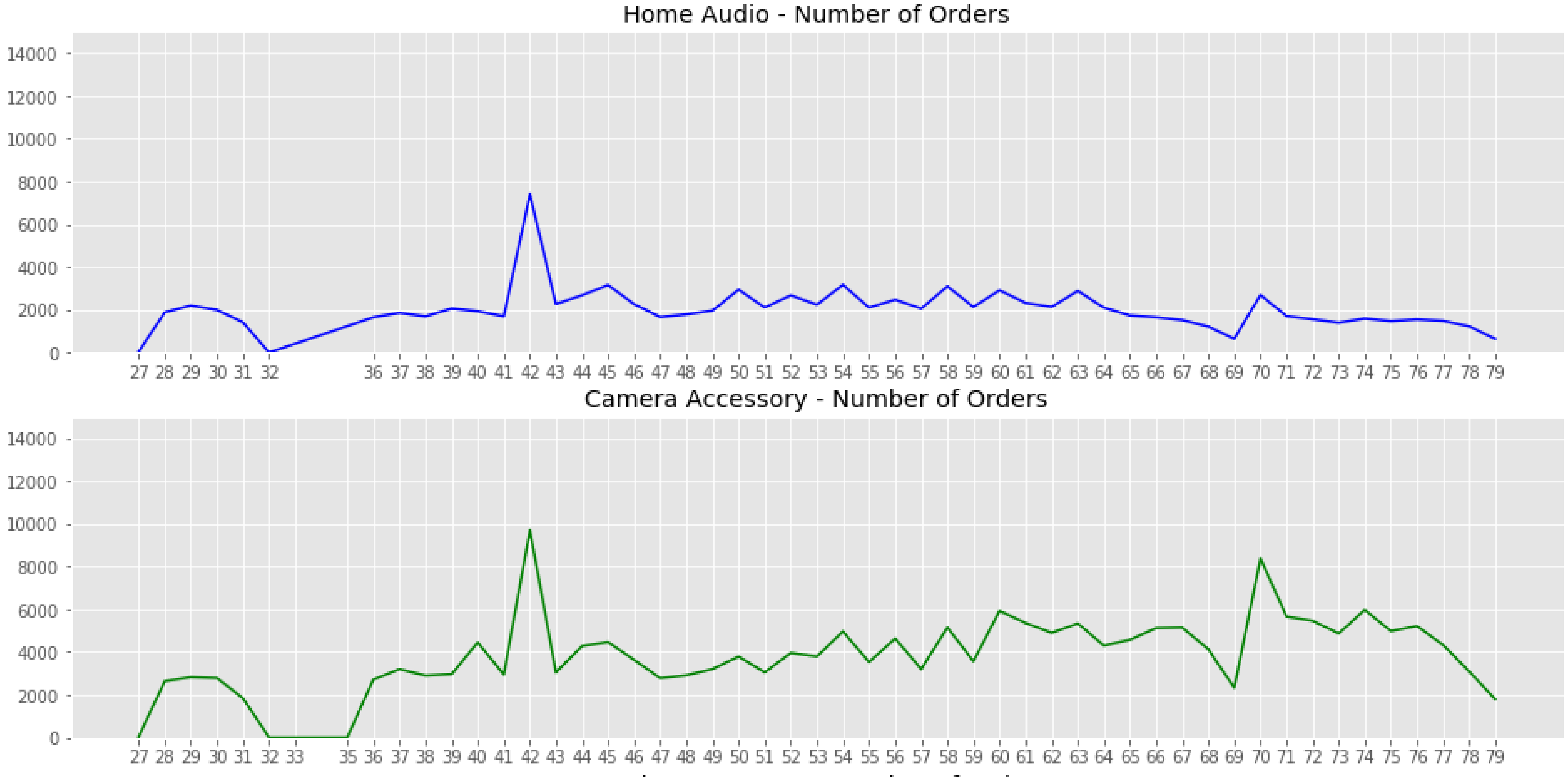
There are high correlation between:

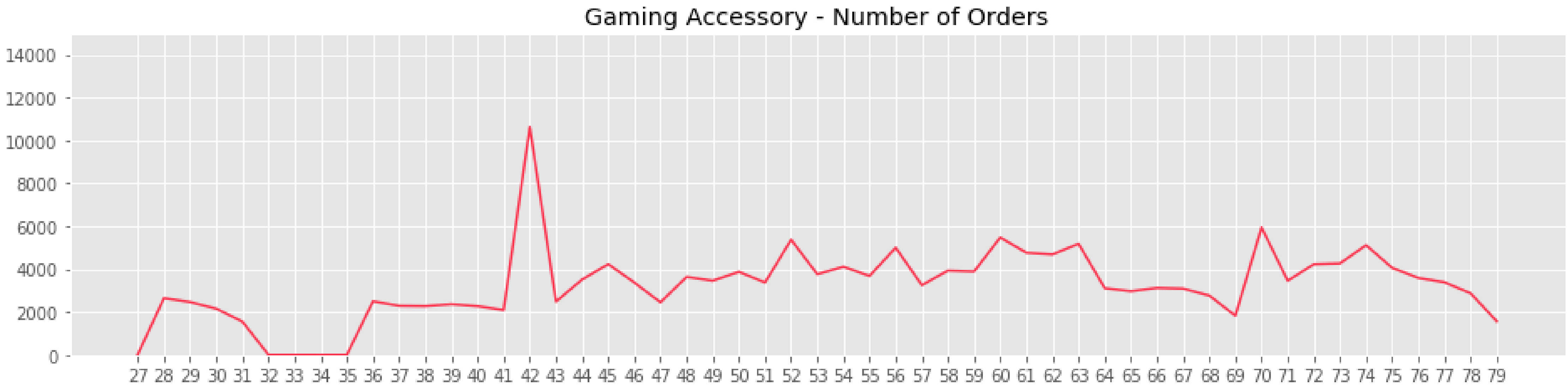
- order_week & month, year & month, order_week & year
- adStock_Affiliates & adStock_Online marketing
- adStock_Other & adStock_Radio
- adStock_Digital & adStock_SEM
- adStock_Content Marketing & adStock_SEM
- Month & adstock_radio, Month & adstock_others
- adStock_TV & adStock_affiliates
- adStock_Sponsorship & adStock_Content Marketing



4. Exploratory Data Analysis - II

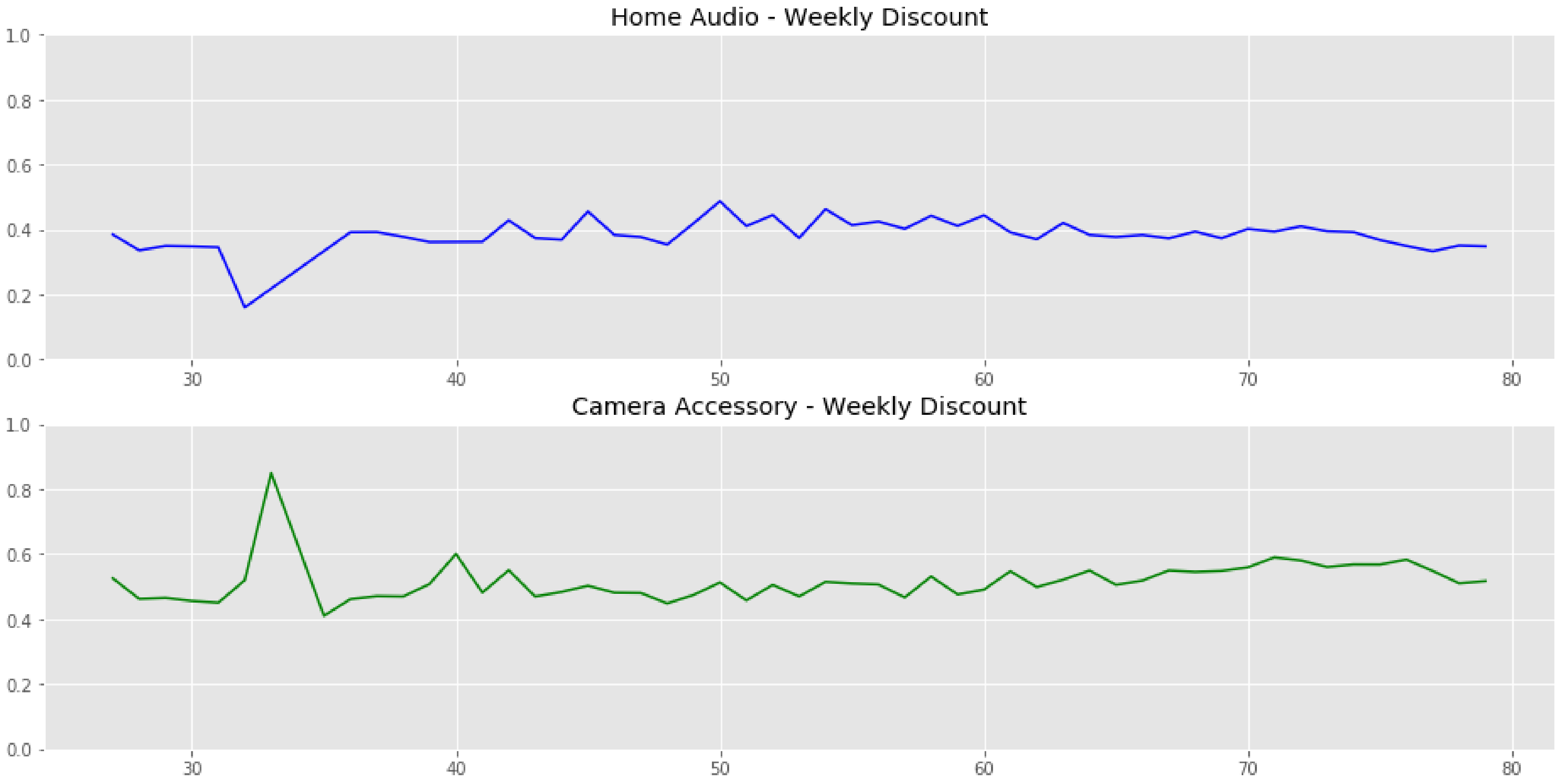
Number of order versus week

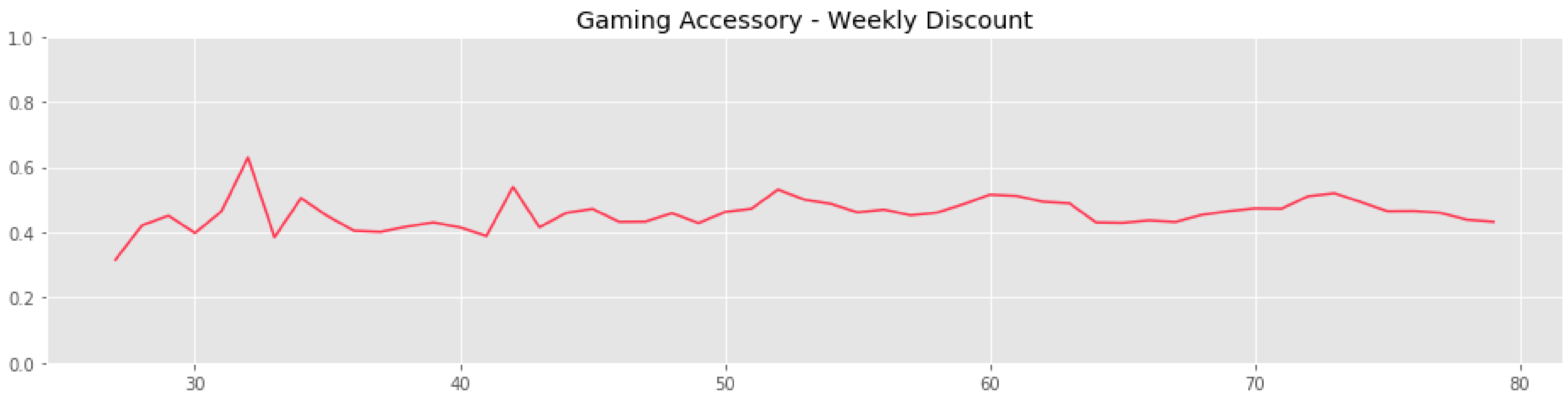




Highest orders - Week 42 for all categories. Gaming having highest orders

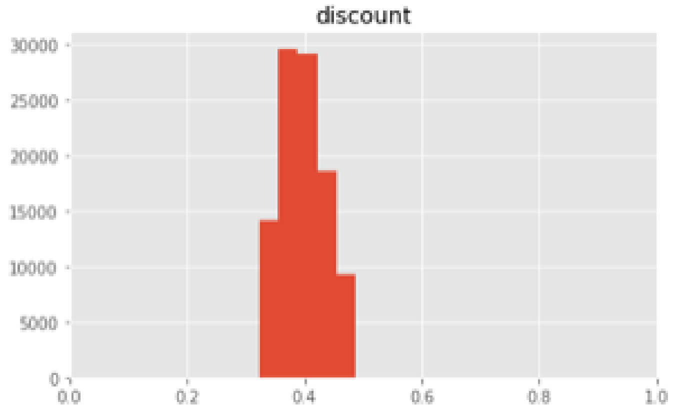
Discount Analysis

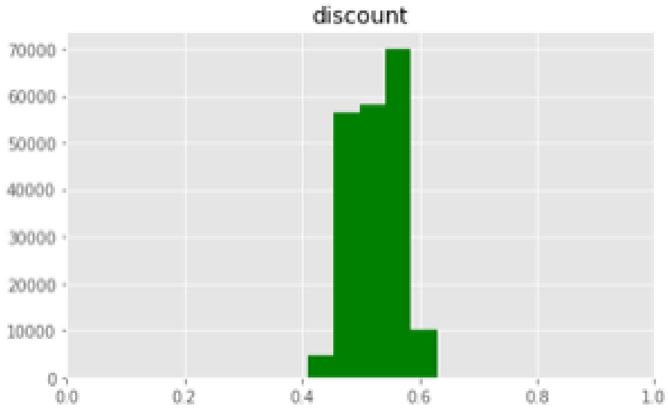


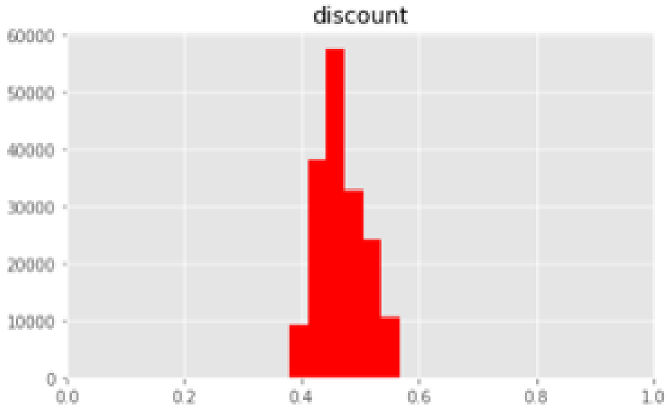


Highest Discounts for Camera Accessory

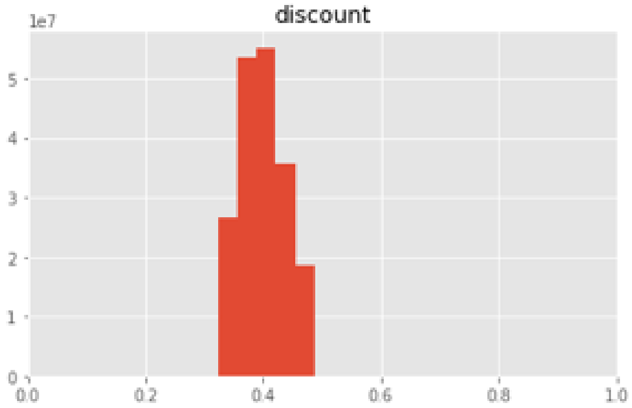
Discount Analysis : Hist Plot of Discount vs No of tranactions

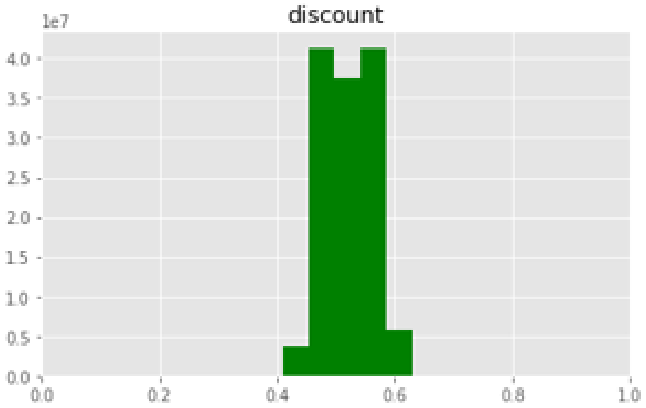


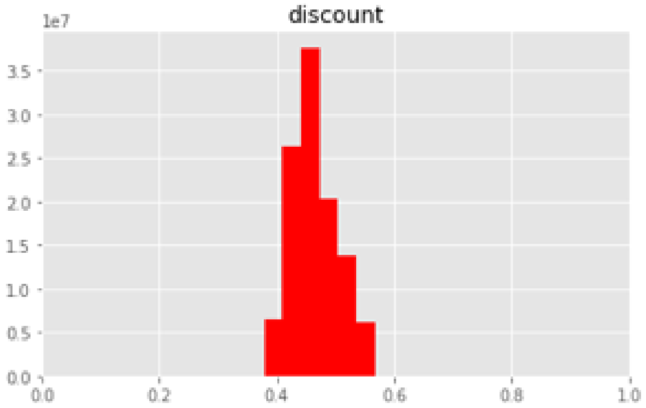




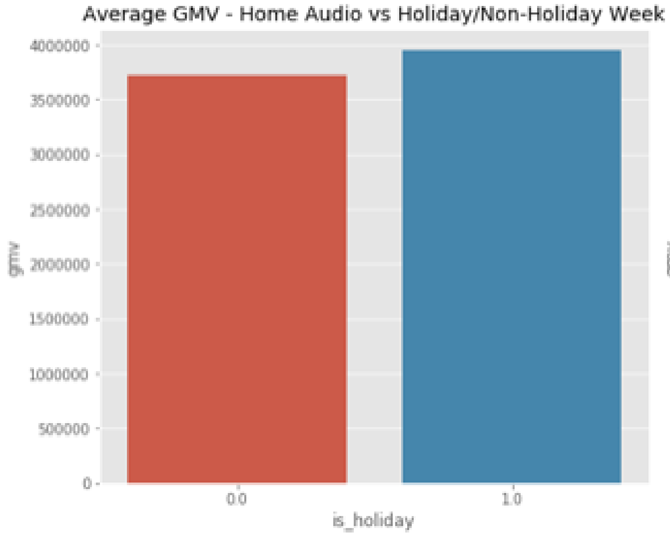
Discount Analysis: Hist Plot of Discount vs GMV

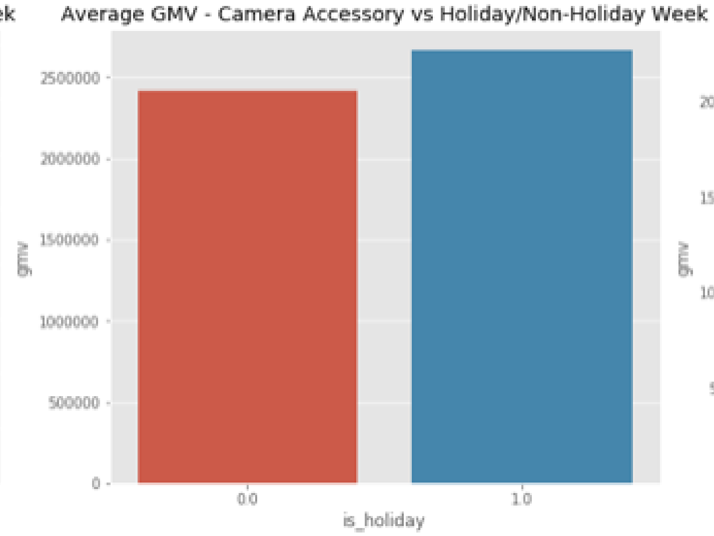


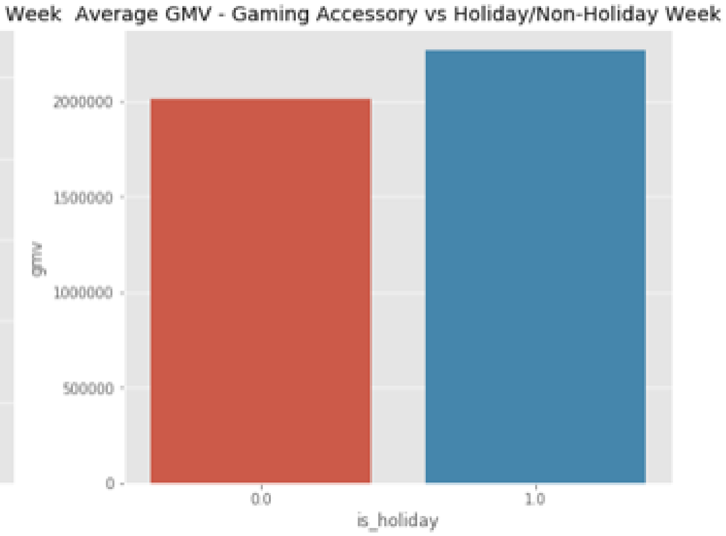




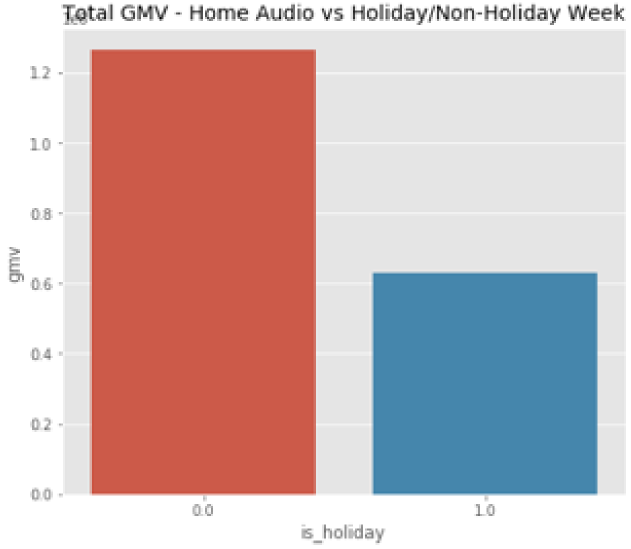
Holiday week vs Average GMV

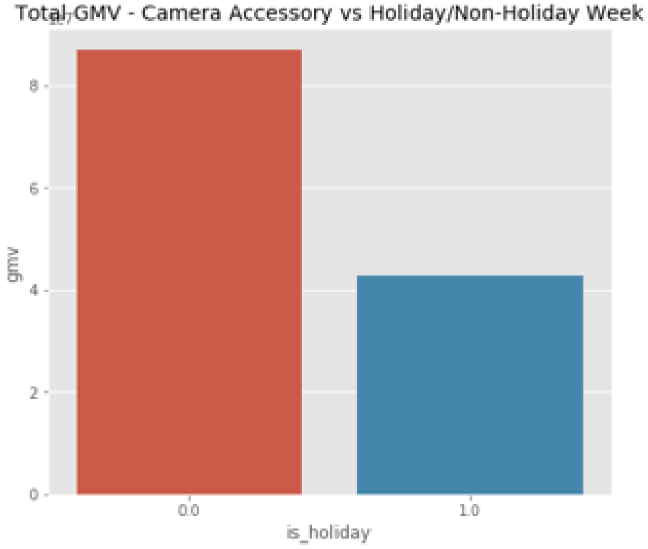


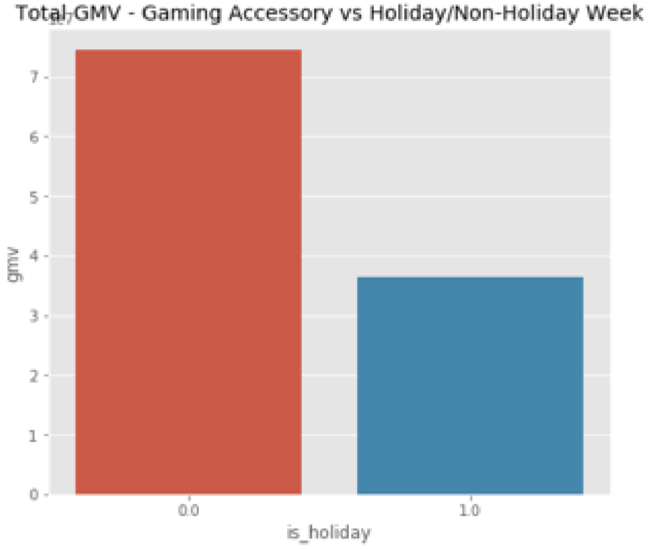




Holiday week vs Total GMV

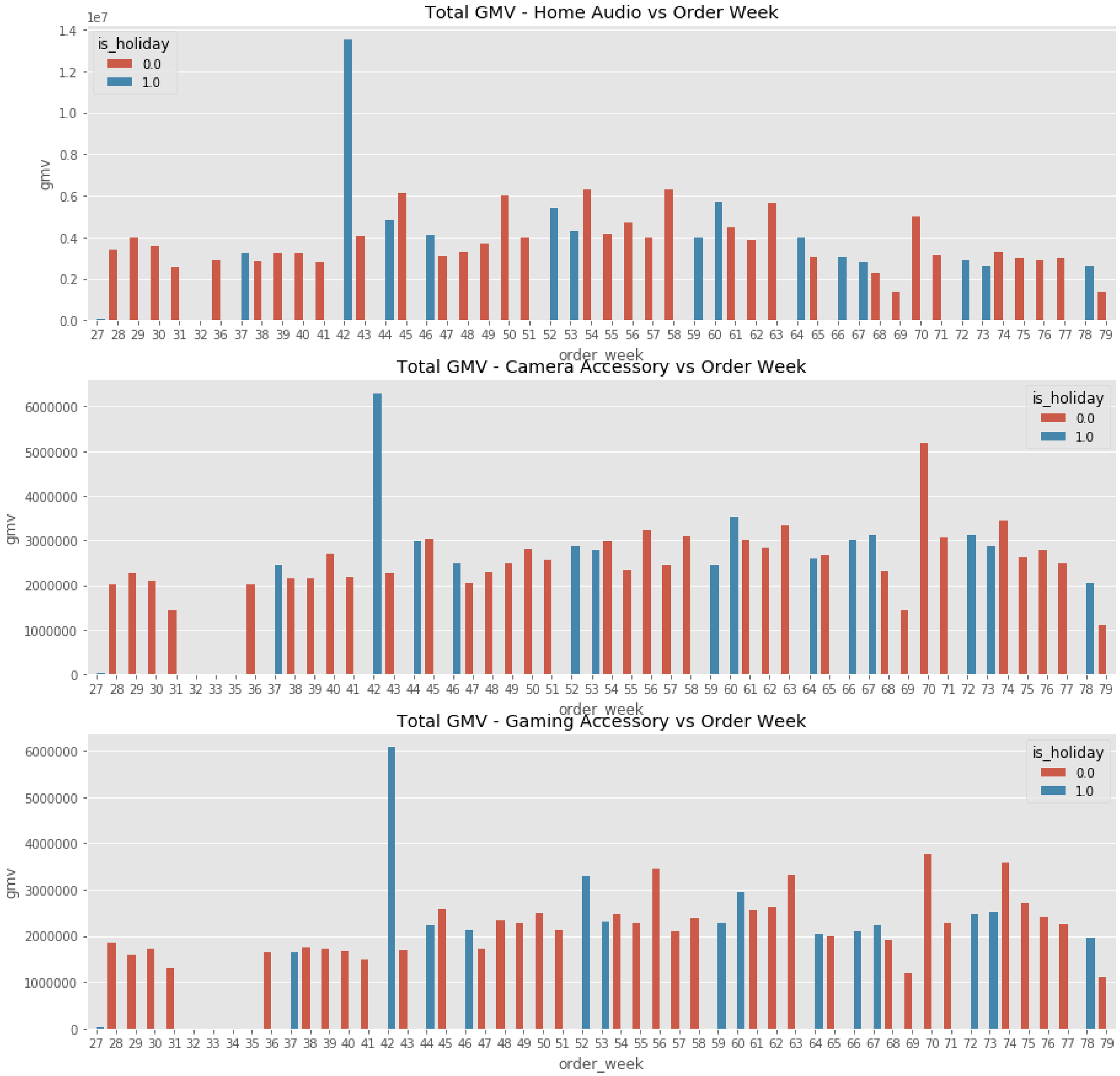




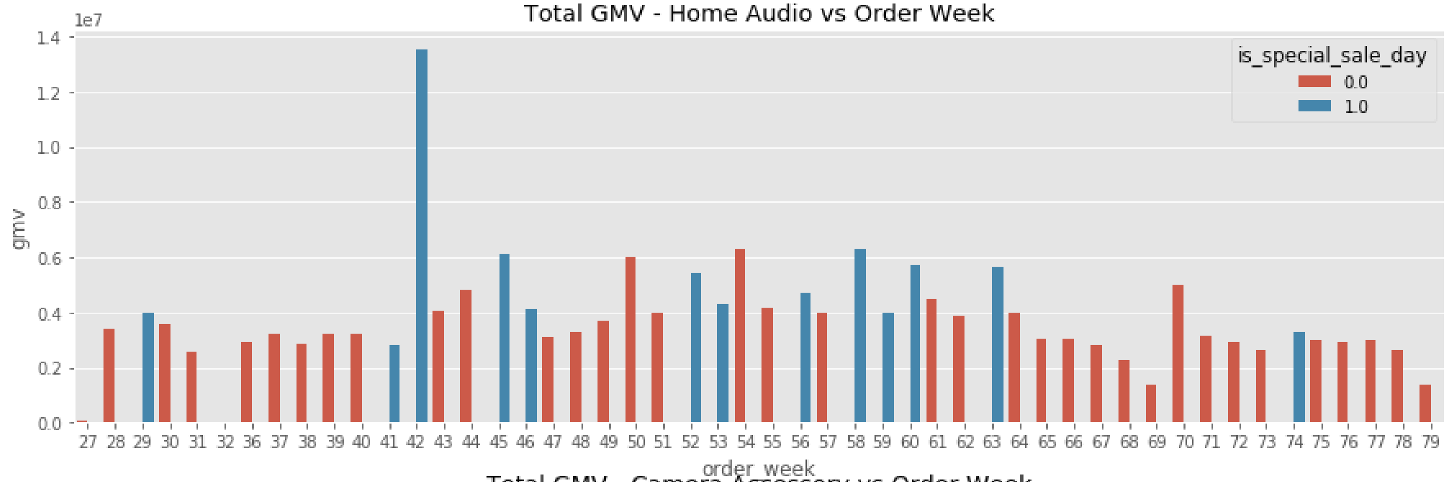


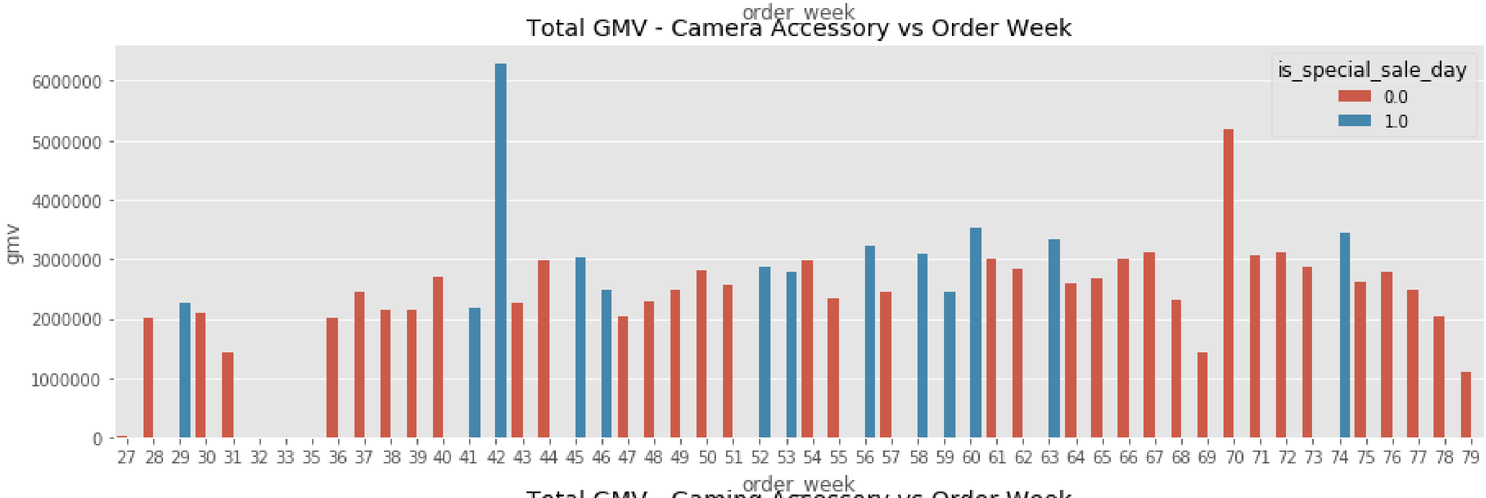
From the above two graphs(mean and sum), overall, total gmv on weeks without holidays is more, however mean_gmv for weeks with holiday is higher

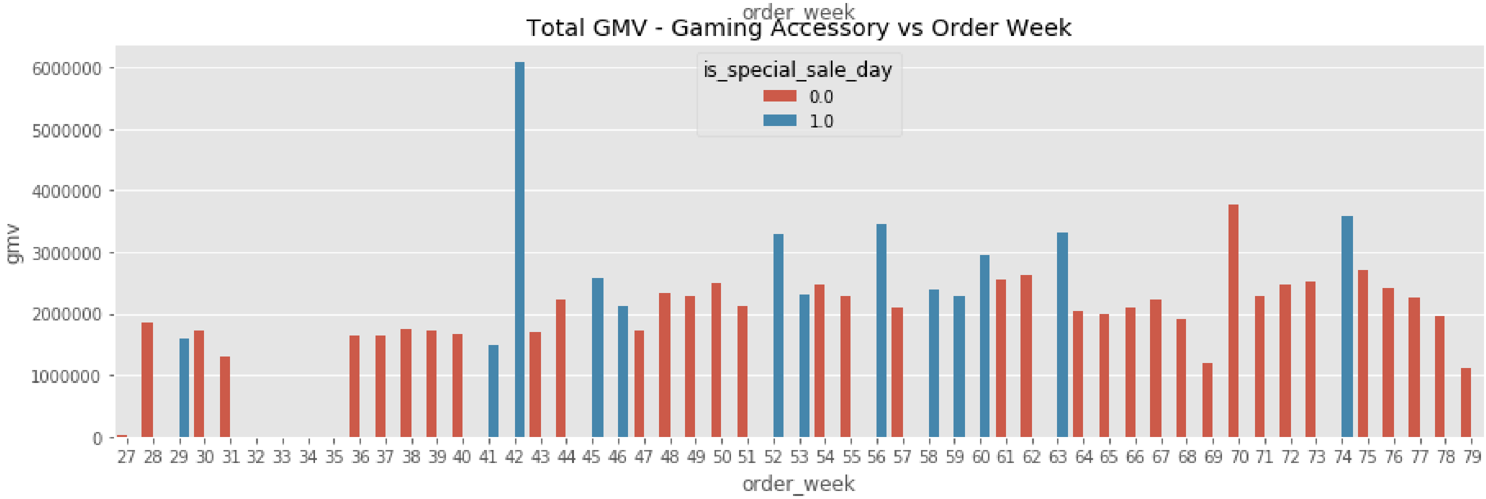
Analyzing the total gmv over the weeks based upon whether the week contained a holiday or not for all the three categories



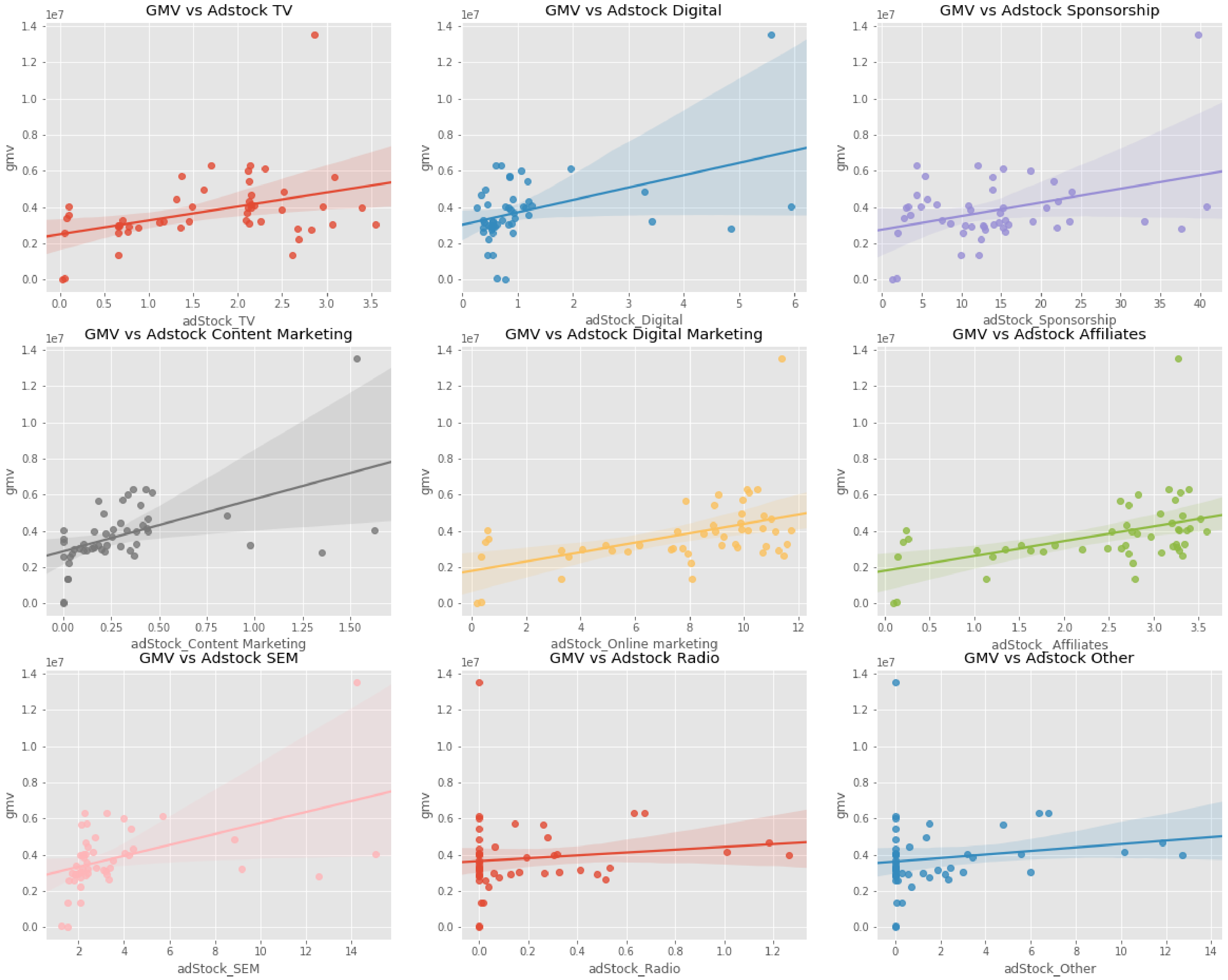
Analyzing the total gmv over the weeks based upon whether the week contained a special sales day or not for all the three categories



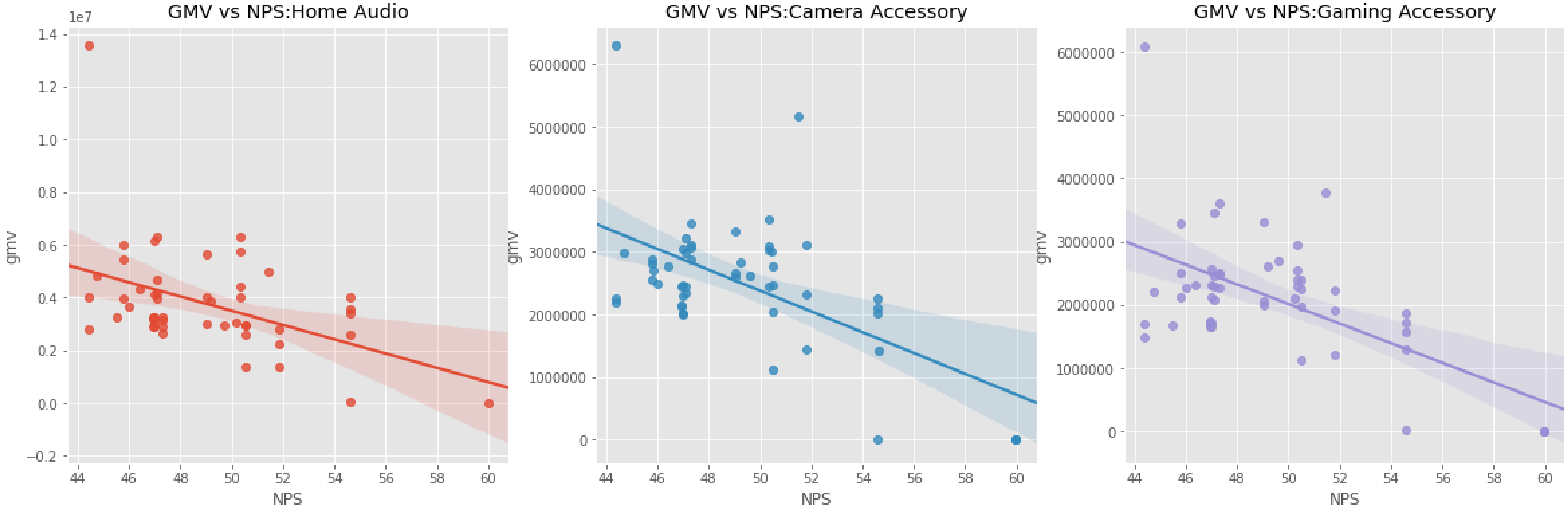




Relationship between GMV and Adstocks of different channels for Home Audio



Relationship between GMV and NPS for all three categories



5. Modelling

Building the Basic Linear Models

1. Gaming Accessory

Train-Test split and scaling of the data

```
# doing the test train split
np.random.seed(0)
df1_train, df1_test = train_test_split(gamingAccessoryDf, train_size = 0.8, test_size = 0.2, random_state = 100)
```

```
# Scaling the features
scaler=MinMaxScaler()

# Apply scaler() to all the columns except the 'yes-no' and 'dummy' variables
num_vars1 = [
    'adStock_TV', 'adStock_Digital', 'adStock_Sponsorship', 'adStock_Content Marketing',
    'adStock_Online marketing', 'adStock_ Affiliates', 'adStock_SEM', 'adStock_Radio', 'adStock_Other',
    'gmv', 'sla', 'product_procurement_sla', 'discount', 'NPS', 'total_holiday', 'total_special_sale_day',
    'total_pay_days', 'pct_online_transactions', 'CoolingPad', 'GameControlMount', 'GamePad', 'GamingAccessoryKit',
    'GamingAdapter', 'GamingChargingStation', 'GamingHeadset', 'GamingKeyboard', 'GamingMemoryCard', 'GamingMouse',
    'GamingMousePad', 'GamingSpeaker', 'JoystickGamingWheel', 'MotionController', 'TVOutCableAccessory'
]
df1_train[num_vars1] = scaler.fit_transform(df1_train[num_vars1])
df1_test[num_vars1] = scaler.fit_transform(df1_test[num_vars1])
```

```
# assiging the values for x_train & y_train
y_train1 = df1_train.pop('gmv')
X_train1 = df1_train
```

OLS regression model with 12 selected features using RFE

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    gmv   R-squared:                       0.988
Model:                            OLS   Adj. R-squared:                  0.983
Method:                 Least Squares   F-statistic:                     195.7
Date:                Sat, 29 Feb 2020   Prob (F-statistic):           2.26e-24
Time:                        13:55:27   Log-Likelihood:                 108.77
No. Observations:                  42   AIC:                            -191.5
Df Residuals:                      29   BIC:                            -168.9
Df Model:                          12
Covariance Type:            nonrobust
=======================================================================================
                          coef    std err          t      P>|t|      [0.025      0.975]
---------------------------------------------------------------------------------------
const                  -0.0635      0.033     -1.951      0.061      -0.130       0.003
adStock_TV              0.1363      0.064      2.129      0.042       0.005       0.267
adStock_Digital         1.0561      0.418      2.525      0.017       0.201       1.911
adStock_Sponsorship     0.1432      0.054      2.632      0.013       0.032       0.254
adStock_SEM            -1.2549      0.477     -2.633      0.013      -2.229      -0.280
adStock_Radio           0.7345      0.305      2.407      0.023       0.110       1.359
adStock_Other          -0.6571      0.297     -2.213      0.035      -1.265      -0.050
GamePad                 0.1891      0.033      5.761      0.000       0.122       0.256
GamingAccessoryKit      0.1665      0.034      4.919      0.000       0.097       0.236
GamingHeadset           0.1751      0.027      6.514      0.000       0.120       0.230
GamingMemoryCard        0.0787      0.032      2.496      0.018       0.014       0.143
GamingMouse             0.4738      0.041     11.621      0.000       0.390       0.557
GamingSpeaker           0.0616      0.027      2.262      0.031       0.006       0.117
==============================================================================
Omnibus:                        0.662   Durbin-Watson:                   1.605
Prob(Omnibus):                  0.718   Jarque-Bera (JB):                0.449
Skew:                           0.250   Prob(JB):                        0.799
Kurtosis:                       2.925   Cond. No.                         338.
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

VIFs for the features

	Features	VIF
4	adStock_SEM	1132.16
2	adStock_Digital	868.31
5	adStock_Radio	519.89
6	adStock_Other	492.67
0	const	93.09
1	adStock_TV	29.53
3	adStock_Sponsorship	16.20
10	GamingMemoryCard	4.87
8	GamingAccessoryKit	4.32
7	GamePad	3.57
9	GamingHeadset	3.45
11	GamingMouse	3.05
12	GamingSpeaker	2.49

Final model after removing the features where VIFs is greater than 5 and p-value is greater than 0.05

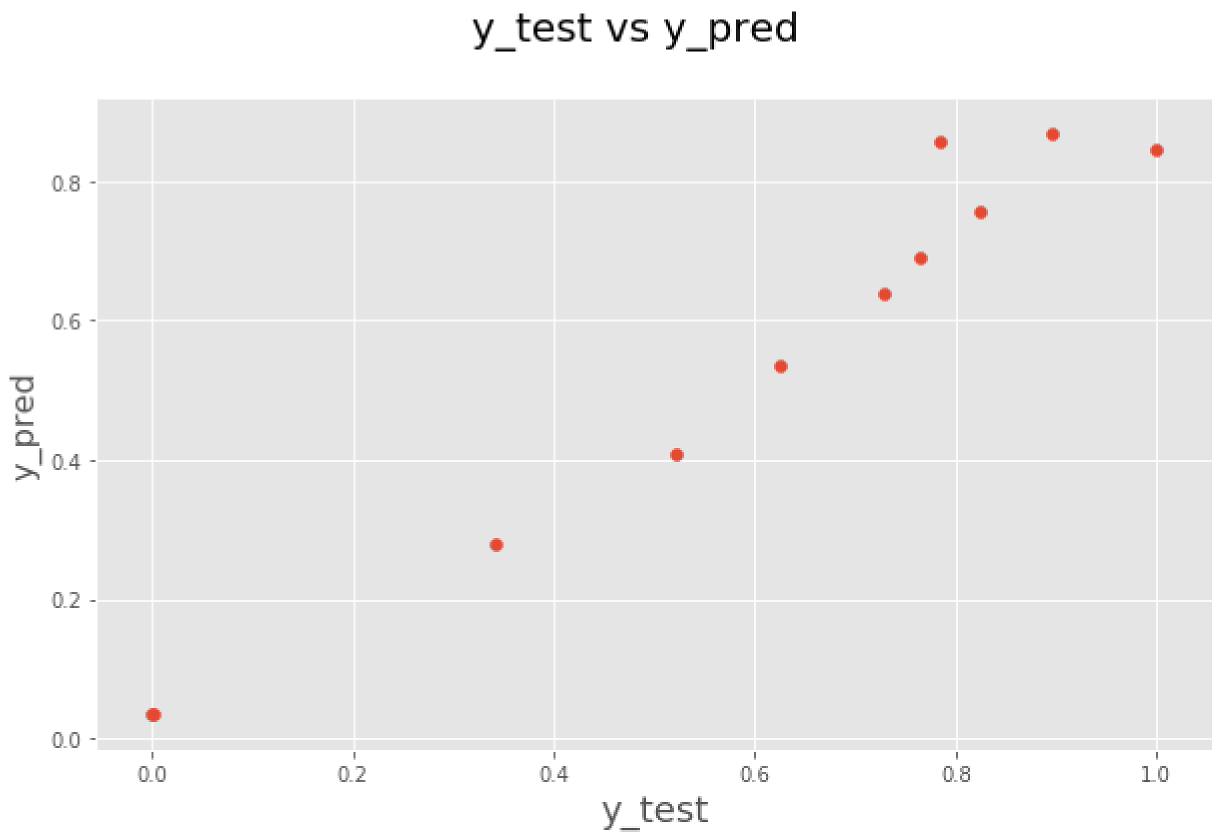
```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    gmv   R-squared:                       0.969
Model:                            OLS   Adj. R-squared:                  0.965
Method:                 Least Squares   F-statistic:                     287.5
Date:                Sat, 29 Feb 2020   Prob (F-statistic):           2.58e-27
Time:                        14:01:01   Log-Likelihood:                 89.059
No. Observations:                  42   AIC:                            -168.1
Df Residuals:                      37   BIC:                            -159.4
Df Model:                           4
Covariance Type:            nonrobust
=================================================================================
                    coef    std err          t      P>|t|      [0.025      0.975]
---------------------------------------------------------------------------------
const             0.0360      0.013      2.703      0.010       0.009       0.063
GamePad           0.3163      0.033      9.608      0.000       0.250       0.383
GamingHeadset     0.2695      0.030      9.056      0.000       0.209       0.330
GamingMouse       0.3969      0.049      8.117      0.000       0.298       0.496
GamingSpeaker     0.0838      0.028      2.985      0.005       0.027       0.141
==============================================================================
Omnibus:                        4.665   Durbin-Watson:                   1.933
Prob(Omnibus):                  0.097   Jarque-Bera (JB):                3.413
Skew:                          -0.538   Prob(JB):                        0.181
Kurtosis:                       3.890   Cond. No.                         13.7
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

	Features	VIF
0	const	7.78
3	GamingMouse	2.19
2	GamingHeadset	2.11
1	GamePad	1.79
4	GamingSpeaker	1.32

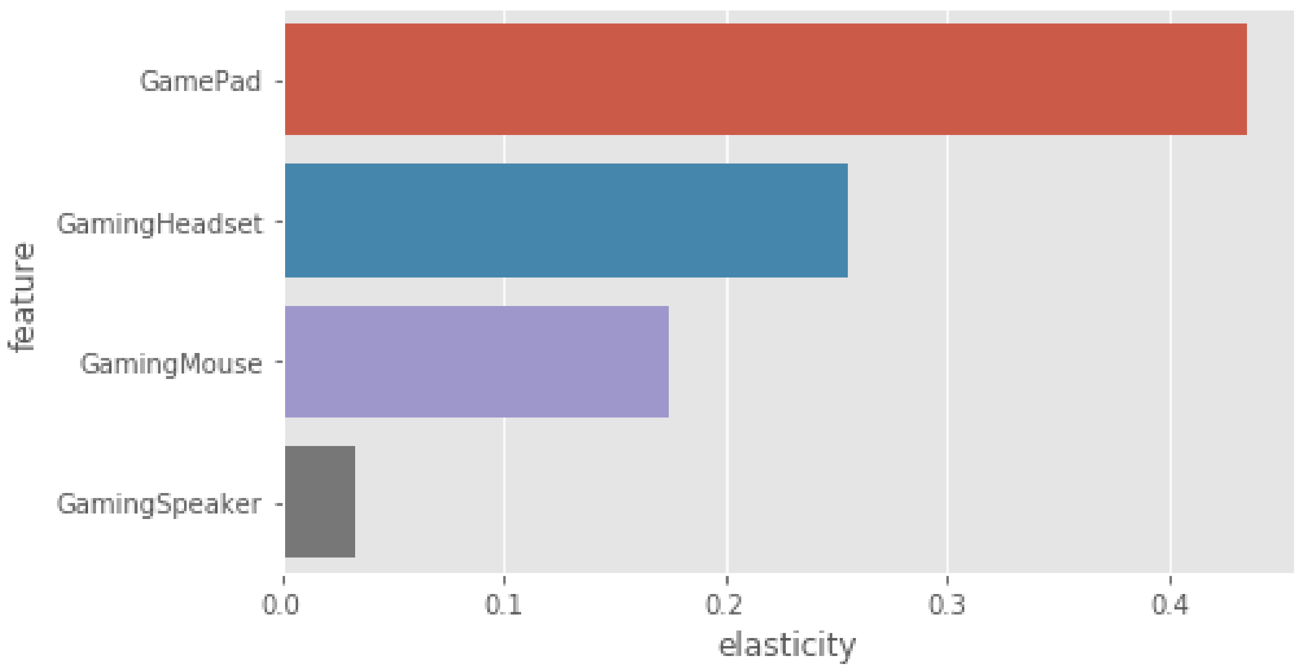
The variance explained by this model is 91% which is good

Model Evaluation

Plotting y_test and y_pred to understand the spread



Elasticity - Gaming Accessory - Basic Linear Model



2. Camera Accessory

Train-Test split and scaling of the data

```
# splitting the data into test and train set
np.random.seed(0)
df_train, df_test = train_test_split(cameraAccessoryDf, train_size = 0.8, test_size = 0.2, random_state = 100)
```

```
scaler = MinMaxScaler()

# Apply scaler() to all the columns except the categorical variables
num_vars = [
    'adStock_TV', 'adStock_Digital','adStock_Sponsorship', 'adStock_Content Marketing', 'adStock_Online marketing',
    'adStock_ Affiliates', 'adStock_SEM', 'adStock_Radio', 'adStock_Other', 'gmv', 'sla', 'product_procurement_sla',
    'discount', 'NPS', 'total_holiday', 'total_special_sale_day', 'total_pay_days', 'pct_online_transactions',
    'Binoculars', 'CameraAccessory', 'CameraBag', 'CameraBattery', 'CameraBatteryCharger', 'CameraBatteryGrip',
    'CameraEyeCup', 'CameraFilmRolls', 'CameraHousing', 'CameraMicrophone', 'CameraMount', 'CameraRemoteControl',
    'CameraTripod', 'ExtensionTube', 'Filter', 'Flash', 'FlashShoeAdapter', 'Lens', 'ReflectorUmbrella',
    'Softbox', 'Strap', 'Telescope'
]

df_train[num_vars] = scaler.fit_transform(df_train[num_vars])
df_test[num_vars] = scaler.fit_transform(df_test[num_vars])
```

```
# assigning the x_train and y_train values
y_train = df_train.pop('gmv')
X_train = df_train
```

OLS regression model with 12 selected features using RFE

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    gmv   R-squared:                       0.988
Model:                            OLS   Adj. R-squared:                  0.982
Method:                 Least Squares   F-statistic:                     186.0
Date:                Sat, 29 Feb 2020   Prob (F-statistic):           2.21e-23
Time:                        20:28:46   Log-Likelihood:                 106.91
No. Observations:                  41   AIC:                            -187.8
Df Residuals:                      28   BIC:                            -165.5
Df Model:                          12
Covariance Type:            nonrobust
=============================================================================================
                                coef    std err          t      P>|t|      [0.025      0.975]
---------------------------------------------------------------------------------------------
const                         0.1494      0.030      4.952      0.000       0.088       0.211
adStock_TV                   -0.0373      0.041     -0.917      0.367      -0.120       0.046
adStock_Content Marketing    -0.2053      0.124     -1.654      0.109      -0.460       0.049
adStock_SEM                   0.1869      0.128      1.462      0.155      -0.075       0.449
adStock_Radio                -0.1134      0.096     -1.180      0.248      -0.310       0.083
adStock_Other                 0.1701      0.100      1.709      0.099      -0.034       0.374
discount                     -0.1716      0.031     -5.478      0.000      -0.236      -0.107
NPS                          -0.1468      0.043     -3.385      0.002      -0.236      -0.058
pct_online_transactions       0.0482      0.022      2.234      0.034       0.004       0.092
CameraBag                     0.4912      0.055      8.962      0.000       0.379       0.603
CameraBattery                 0.0996      0.044      2.245      0.033       0.009       0.190
CameraTripod                  0.3379      0.038      8.970      0.000       0.261       0.415
Flash                         0.2539      0.036      7.109      0.000       0.181       0.327
==============================================================================
Omnibus:                        0.433   Durbin-Watson:                   2.099
Prob(Omnibus):                  0.805   Jarque-Bera (JB):                0.060
Skew:                          -0.075   Prob(JB):                        0.971
Kurtosis:                       3.112   Cond. No.                         84.6
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

VIFs for the features

	Features	VIF
3	adStock_SEM	82.60
0	const	80.14
2	adStock_Content Marketing	78.62
5	adStock_Other	56.81
4	adStock_Radio	52.72
1	adStock_TV	11.21
12	Flash	8.97
7	NPS	8.31
9	CameraBag	6.73
10	CameraBattery	5.63
6	discount	4.64
11	CameraTripod	4.54
8	pct_online_transactions	3.29

Final model after removing the features where VIFs is greater than 5 and p-value is greater than 0.05

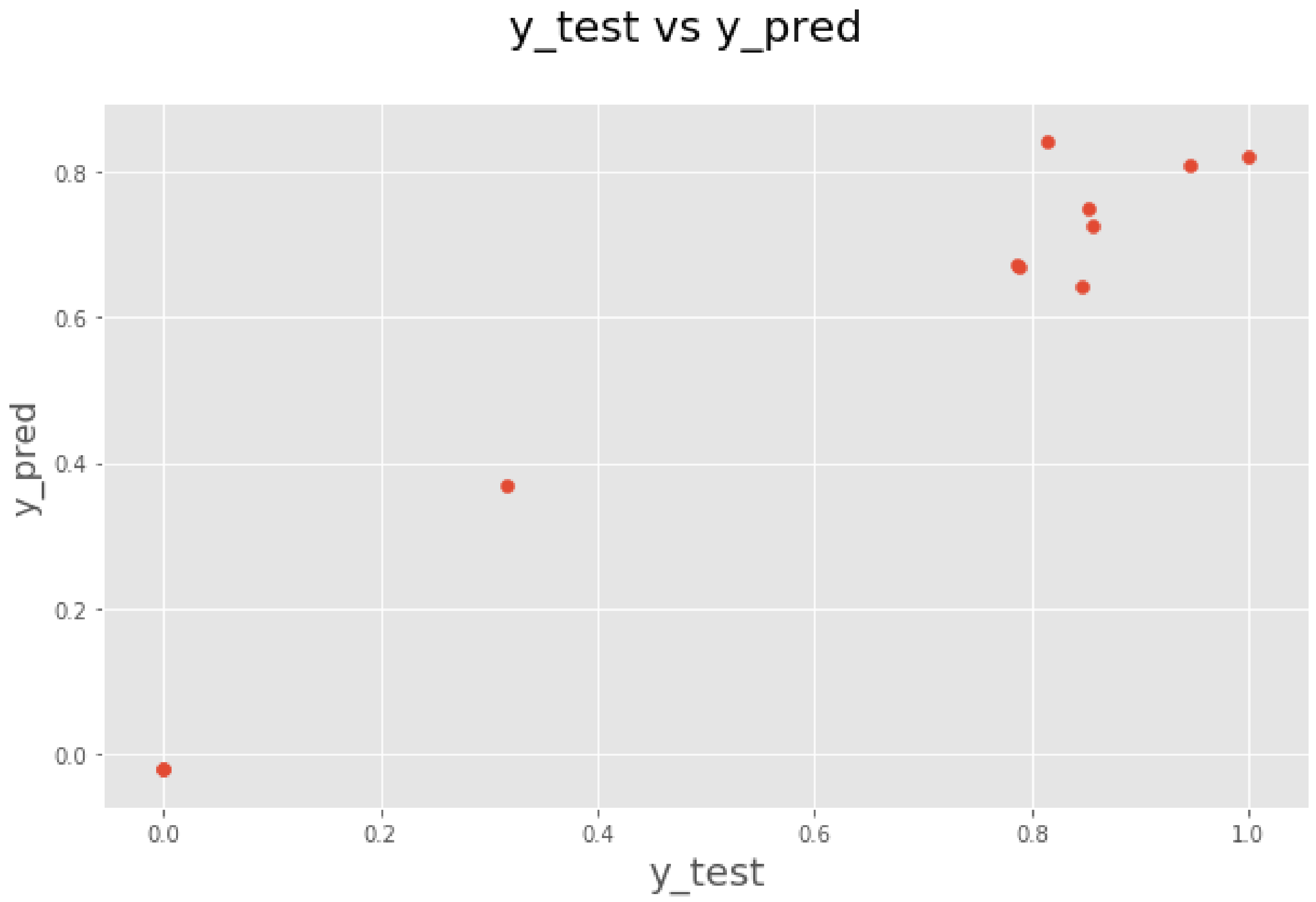
```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    gmv   R-squared:                       0.938
Model:                            OLS   Adj. R-squared:                  0.931
Method:                 Least Squares   F-statistic:                     135.7
Date:                Sat, 29 Feb 2020   Prob (F-statistic):           3.47e-21
Time:                        20:36:17   Log-Likelihood:                 73.838
No. Observations:                  41   AIC:                            -137.7
Df Residuals:                      36   BIC:                            -129.1
Df Model:                           4
Covariance Type:            nonrobust
=================================================================================
                    coef    std err          t      P>|t|      [0.025      0.975]
---------------------------------------------------------------------------------
const             0.0740      0.026      2.893      0.006       0.022       0.126
NPS              -0.0945      0.035     -2.716      0.010      -0.165      -0.024
CameraBag         0.5890      0.072      8.220      0.000       0.444       0.734
CameraBattery     0.2431      0.064      3.804      0.001       0.114       0.373
CameraTripod      0.1232      0.046      2.696      0.011       0.031       0.216
==============================================================================
Omnibus:                       19.330   Durbin-Watson:                   2.035
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               26.807
Skew:                           1.434   Prob(JB):                     1.51e-06
Kurtosis:                       5.734   Cond. No.                         16.7
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

	Features	VIF
0	const	14.73
3	CameraBattery	2.99
2	CameraBag	2.94
4	CameraTripod	1.71
1	NPS	1.37

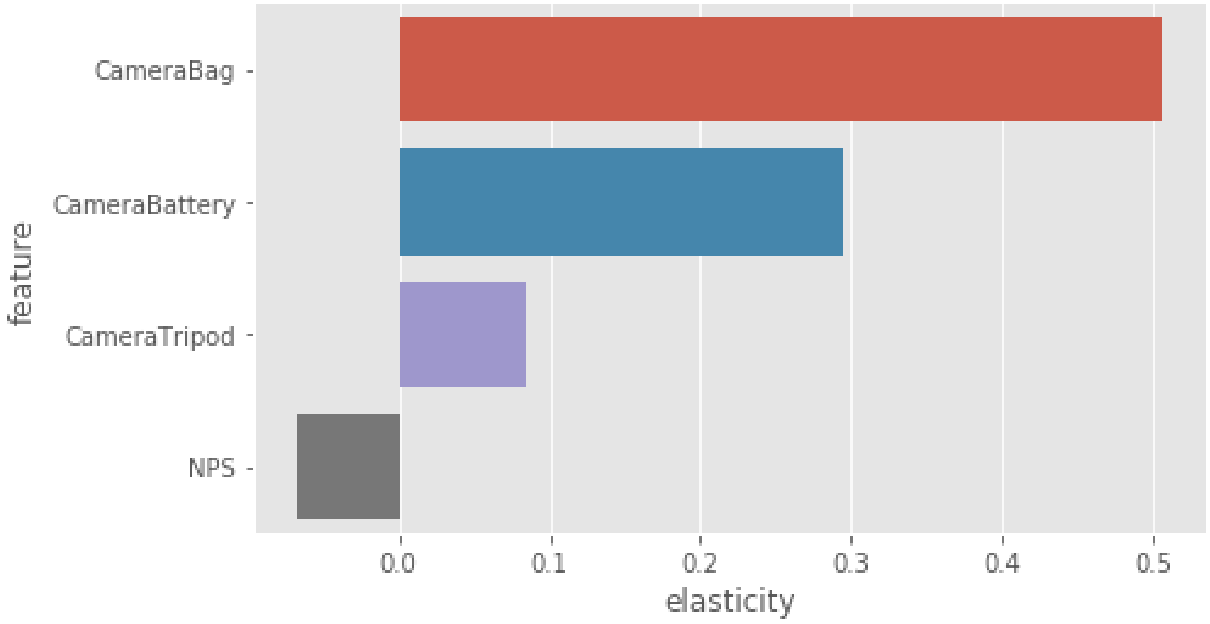
The variance explained by this model is 41%

Model Evaluation

Plotting y_test and y_pred to understand the spread



Elasticity - Camera Accessory



3. Home Audio

Train-Test split and scaling of the data

```
# Doing the test-train split
np.random.seed(0)
df2_train, df2_test = train_test_split(homeAudioDf, train_size = 0.8, test_size = 0.2, random_state = 100)
```

```
scaler = MinMaxScaler()

# Apply scaler() to all the columns except the 'yes-no' and 'dummy' variables
num_vars2 = [
    'adStock_TV', 'adStock_Digital', 'adStock_Sponsorship', 'adStock_Content Marketing', 'adStock_Online marketing',
    'adStock_ Affiliates', 'adStock_SEM', 'adStock_Radio', 'adStock_Other', 'gmv', 'sla', 'product_procurement_sla',
    'discount', 'NPS', 'total_holiday', 'total_special_sale_day', 'total_pay_days', 'pct_online_transactions',
    'BoomBox', 'DJController', 'Dock', 'DockingStation', 'FMRadio', 'HiFiSystem', 'HomeAudioSpeaker', 'SlingBox',
    'SoundMixer', 'VoiceRecorder'
]

df2_train[num_vars2] = scaler.fit_transform(df2_train[num_vars2])
df2_test[num_vars2] = scaler.fit_transform(df2_test[num_vars2])
```

```
# Splitting into X_train & y_train
y_train2 = df2_train.pop('gmv')
X_train2 = df2_train
```

OLS regression model with 12 selected features using RFE

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    gmv   R-squared:                       0.999
Model:                            OLS   Adj. R-squared:                  0.998
Method:                 Least Squares   F-statistic:                     1863.
Date:                Sat, 29 Feb 2020   Prob (F-statistic):           3.97e-36
Time:                        20:49:54   Log-Likelihood:                 155.29
No. Observations:                  40   AIC:                            -284.6
Df Residuals:                      27   BIC:                            -262.6
Df Model:                          12
Covariance Type:            nonrobust
==============================================================================
                            coef    std err          t      P>|t|      [0.025      0.975]
-------------------------------------------------------------------------------------------
const                     0.0325      0.006      5.835      0.000       0.021       0.044
adStock_TV                0.0154      0.020      0.779      0.443      -0.025       0.056
adStock_Digital          -0.1374      0.046     -3.016      0.006      -0.231      -0.044
adStock_Sponsorship      -0.0369      0.026     -1.440      0.161      -0.090       0.016
adStock_Online marketing -0.1144      0.115     -0.993      0.329      -0.351       0.122
adStock_ Affiliates       0.0869      0.112      0.779      0.443      -0.142       0.316
adStock_SEM               0.1855      0.056      3.311      0.003       0.071       0.300
adStock_Other            -0.0165      0.009     -1.897      0.069      -0.034       0.001
pct_online_transactions  -0.0212      0.005     -4.043      0.000      -0.032      -0.010
BoomBox                   0.0276      0.013      2.173      0.039       0.002       0.054
FMRadio                   0.0215      0.015      1.407      0.171      -0.010       0.053
HomeAudioSpeaker          0.9162      0.014     66.221      0.000       0.888       0.945
VoiceRecorder             0.0576      0.008      7.655      0.000       0.042       0.073
==============================================================================
Omnibus:                        2.122   Durbin-Watson:                   1.916
Prob(Omnibus):                  0.346   Jarque-Bera (JB):                1.452
Skew:                          -0.464   Prob(JB):                        0.484
Kurtosis:                       3.108   Cond. No.                         288.
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

VIFs for the features

	Features	VIF
4	adStock_Online marketing	1304.67
5	adStock_ Affiliates	1259.08
6	adStock_SEM	197.28
2	adStock_Digital	131.60
3	adStock_Sponsorship	47.30
0	const	33.78
1	adStock_TV	31.04
9	BoomBox	6.45
10	FMRadio	5.74
7	adStock_Other	5.47
11	HomeAudioSpeaker	4.52
12	VoiceRecorder	2.16
8	pct_online_transactions	1.83

Final model after removing the features where VIFs is greater than 5 and p-value is greater than 0.05

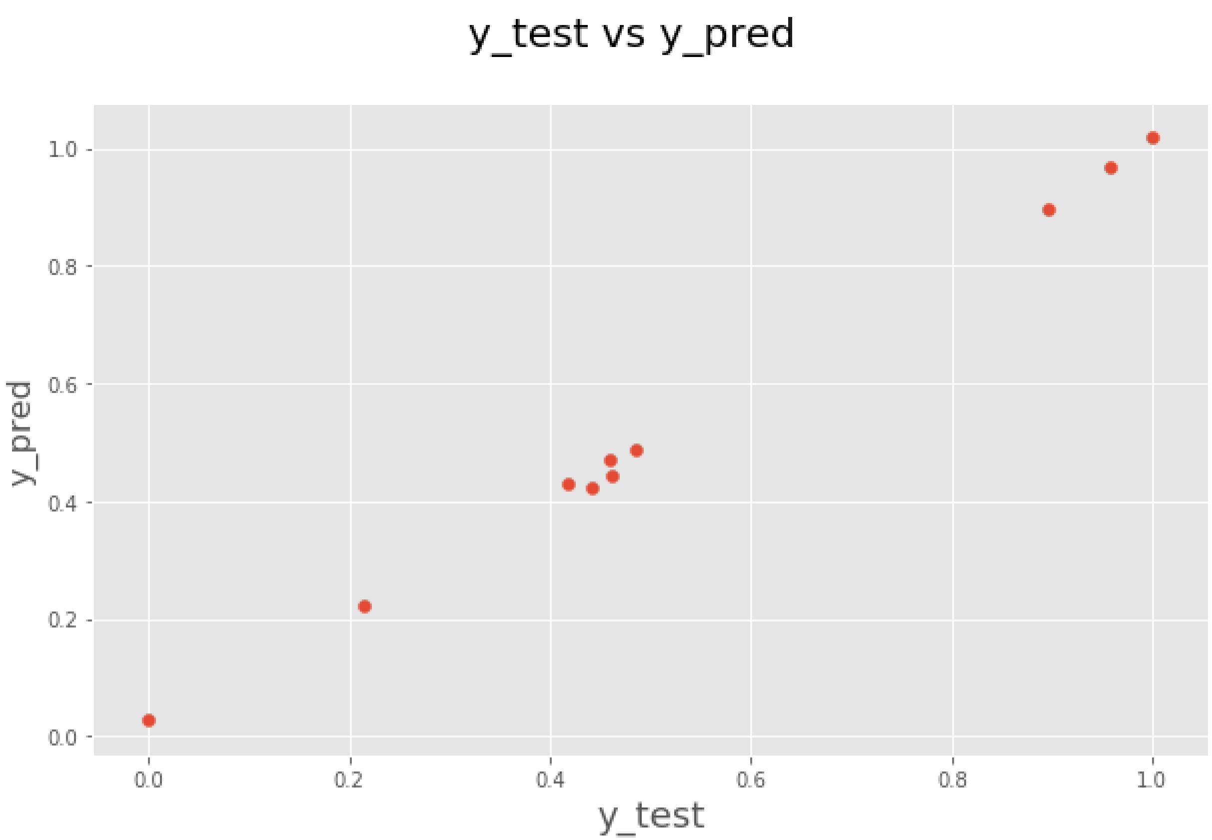
```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    gmv   R-squared:                       0.996
Model:                            OLS   Adj. R-squared:                  0.995
Method:                 Least Squares   F-statistic:                     4162.
Date:                Sat, 29 Feb 2020   Prob (F-statistic):           2.82e-44
Time:                        20:59:30   Log-Likelihood:                 129.30
No. Observations:                  40   AIC:                            -252.6
Df Residuals:                      37   BIC:                            -247.5
Df Model:                           2
Covariance Type:            nonrobust
====================================================================================
                       coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------------
const                0.0271      0.004      6.929      0.000       0.019       0.035
HomeAudioSpeaker     0.9669      0.011     90.332      0.000       0.945       0.989
VoiceRecorder        0.0333      0.008      3.960      0.000       0.016       0.050
==============================================================================
Omnibus:                        3.518   Durbin-Watson:                   1.843
Prob(Omnibus):                  0.172   Jarque-Bera (JB):                2.754
Skew:                          -0.641   Prob(JB):                        0.252
Kurtosis:                       3.089   Cond. No.                         7.58
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

	Features	VIF
0	const	6.23
1	HomeAudioSpeaker	1.01
2	VoiceRecorder	1.01

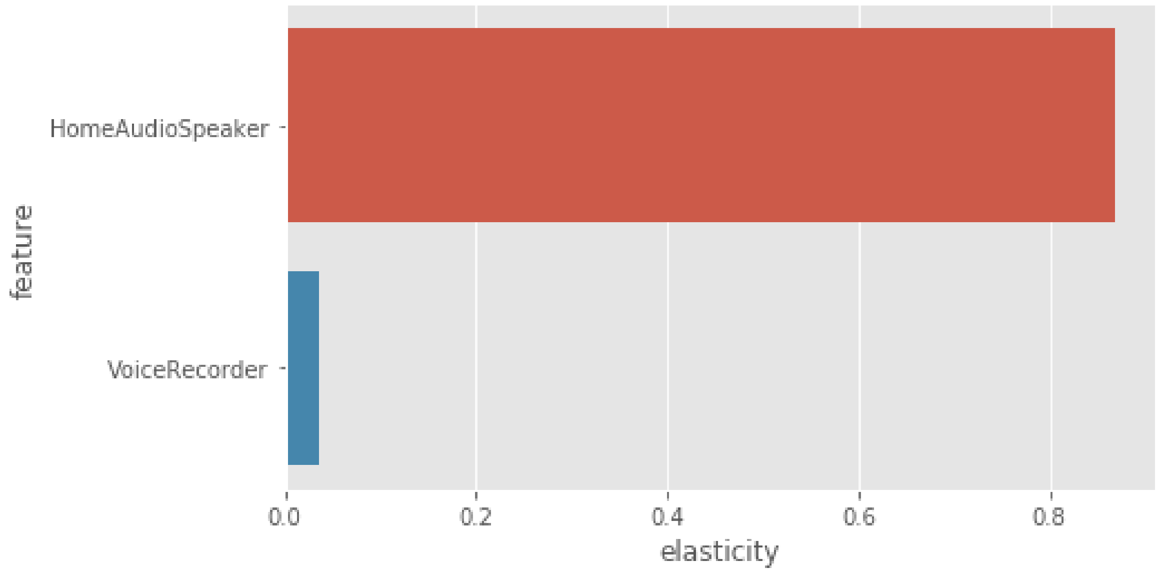
The variance explained by this model is 97%

Model Evaluation

Plotting y_test and y_pred to understand the spread



Elasticity - Home Audio - Basic Linear model



Building Multiplicative Models

I. Gaming Accessory - Multiplicative model

Train-Test split and scaling of the data

```
# Test-train split
df_train, df_test = train_test_split(df, train_size = 0.8, test_size = 0.20, random_state = 100)
```

```
# initialising scaler
scaler = MinMaxScaler()
```

```
# checking the columns
df.columns
```

```
Index(['order_week', 'Month', 'Year', 'adStock_TV', 'adStock_Digital',
       'adStock_Sponsorship', 'adStock_Content Marketing',
       'adStock_Online marketing', 'adStock_ Affiliates', 'adStock_SEM',
       'adStock_Radio', 'adStock_Other', 'gmv', 'sla',
       'product_procurement_sla', 'discount', 'NPS', 'total_holiday',
       'is_holiday', 'total_special_sale_day', 'is_special_sale_day',
       'total_pay_days', 'is_pay_days', 'pct_online_transactions',
       'CoolingPad', 'GameControlMount', 'GamePad', 'GamingAccessoryKit',
       'GamingAdapter', 'GamingChargingStation', 'GamingHeadset',
       'GamingKeyboard', 'GamingMemoryCard', 'GamingMouse', 'GamingMousePad',
       'GamingSpeaker', 'JoystickGamingWheel', 'MotionController',
       'TVOutCableAccessory'],
      dtype='object')
```

OLS regression model with 12 selected features using RFE

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    gmv   R-squared:                       0.999
Model:                            OLS   Adj. R-squared:                  0.999
Method:                 Least Squares   F-statistic:                     2330.
Date:                Sat, 29 Feb 2020   Prob (F-statistic):           7.10e-40
Time:                        22:15:15   Log-Likelihood:                 154.40
No. Observations:                  42   AIC:                            -282.8
Df Residuals:                      29   BIC:                            -260.2
Df Model:                          12
Covariance Type:            nonrobust
============================================================================================
                               coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------------------
const                       29.4770     59.005      0.500      0.621     -91.201     150.155
Year                        -3.9048      7.755     -0.504      0.618     -19.765      11.956
adStock_Online marketing    -0.2195      0.201     -1.093      0.283      -0.630       0.191
adStock_ Affiliates          0.2161      0.201      1.077      0.290      -0.194       0.626
pct_online_transactions     -0.1119      0.053     -2.112      0.043      -0.220      -0.004
GamePad                      0.3442      0.048      7.243      0.000       0.247       0.441
GamingAccessoryKit          -0.1729      0.075     -2.311      0.028      -0.326      -0.020
GamingAdapter               -0.0915      0.086     -1.064      0.296      -0.267       0.084
GamingHeadset                0.5193      0.066      7.928      0.000       0.385       0.653
GamingKeyboard               0.2181      0.105      2.083      0.046       0.004       0.432
GamingMouse                  0.3748      0.068      5.535      0.000       0.236       0.513
JoystickGamingWheel          0.0750      0.042      1.802      0.082      -0.010       0.160
TVOutCableAccessory          0.0695      0.063      1.112      0.275      -0.058       0.197
==============================================================================
Omnibus:                        1.202   Durbin-Watson:                   1.920
Prob(Omnibus):                  0.548   Jarque-Bera (JB):                1.163
Skew:                          -0.277   Prob(JB):                        0.559
Kurtosis:                       2.401   Cond. No.                     4.29e+05
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 4.29e+05. This might indicate that there are
strong multicollinearity or other numerical problems.
```

VIFs for the features

	Features	VIF
0	const	2.689631e+09
3	adStock_ Affiliates	2.235860e+03
2	adStock_Online marketing	2.166350e+03
7	GamingAdapter	3.520000e+02
9	GamingKeyboard	3.409500e+02
6	GamingAccessoryKit	1.881200e+02
12	TVOutCableAccessory	1.863600e+02
8	GamingHeadset	8.307000e+01
4	pct_online_transactions	8.183000e+01
10	GamingMouse	7.885000e+01
11	JoystickGamingWheel	7.558000e+01
5	GamePad	6.858000e+01
1	Year	2.850000e+00

Final model after removing the features where VIFs is greater than 5 and p-value is greater than 0.05

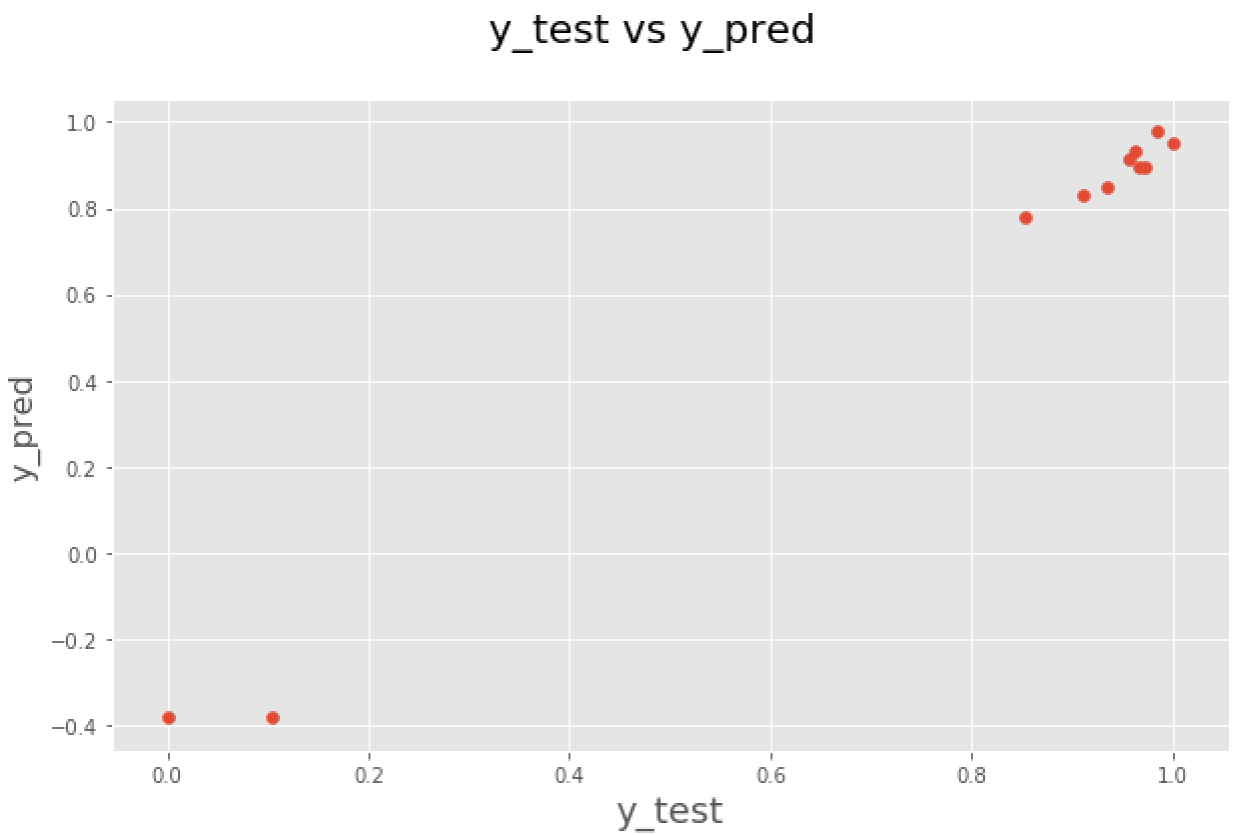
```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    gmv   R-squared:                       0.979
Model:                            OLS   Adj. R-squared:                  0.978
Method:                 Least Squares   F-statistic:                     919.8
Date:                Sat, 29 Feb 2020   Prob (F-statistic):           1.54e-33
Time:                        22:15:22   Log-Likelihood:                 91.452
No. Observations:                  42   AIC:                            -176.9
Df Residuals:                      39   BIC:                            -171.7
Df Model:                           2
Covariance Type:            nonrobust
=================================================================================
                    coef    std err          t      P>|t|      [0.025      0.975]
---------------------------------------------------------------------------------
const            -0.3794      0.029    -13.111      0.000      -0.438      -0.321
GamingHeadset     0.7896      0.034     23.010      0.000       0.720       0.859
GamingMouse       0.5720      0.036     15.710      0.000       0.498       0.646
==============================================================================
Omnibus:                       57.568   Durbin-Watson:                   2.119
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              460.003
Skew:                          -3.187   Prob(JB):                    1.29e-100
Kurtosis:                      17.907   Cond. No.                         16.5
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

	Features	VIF
0	const	43.44
1	GamingHeadset	1.53
2	GamingMouse	1.53

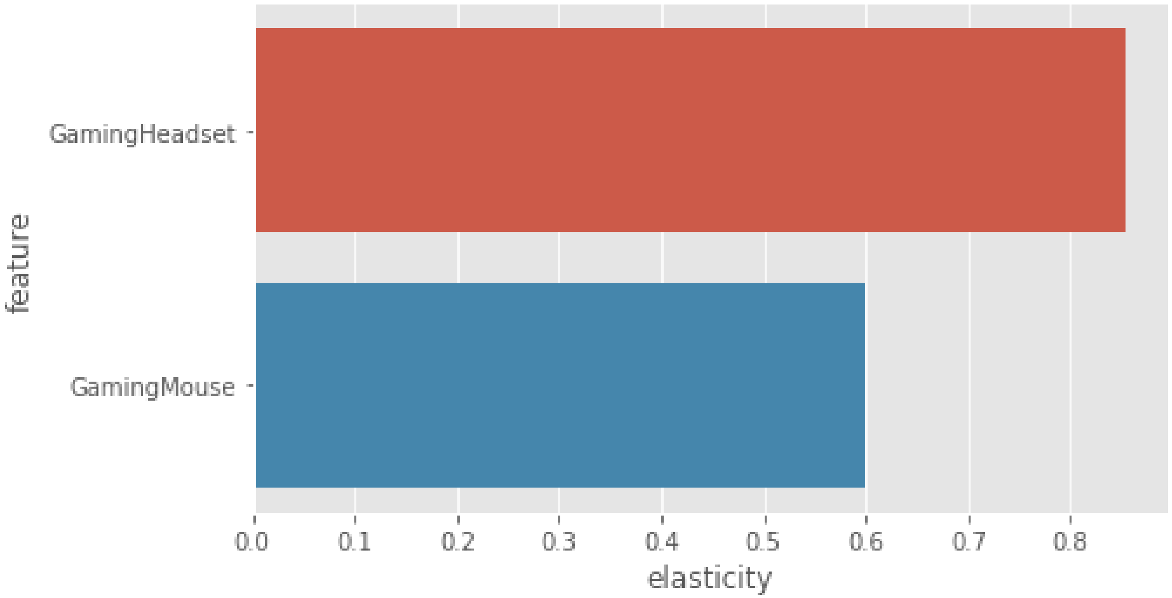
The variance explained by this model is 64%

Model Evaluation

Plotting y_test and y_pred to understand the spread



Elasticity - Gaming Accessory - Basic Linear Model



II. Camera Accessory - Multiplicative model

Train-Test split and scaling of the data

```
# Test-train split
df_train, df_test = train_test_split(df, train_size = 0.8, test_size = 0.2, random_state = 100)
```

```
# scaling
scaler = MinMaxScaler()
```

```
# checking the features in the dataset
df.columns
```

```
Index(['order_week', 'Month', 'Year', 'adStock_TV', 'adStock_Digital',
       'adStock_Sponsorship', 'adStock_Content Marketing',
       'adStock_Online marketing', 'adStock_ Affiliates', 'adStock_SEM',
       'adStock_Radio', 'adStock_Other', 'gmv', 'sla',
       'product_procurement_sla', 'discount', 'NPS', 'total_holiday',
       'is_holiday', 'total_special_sale_day', 'is_special_sale_day',
       'total_pay_days', 'is_pay_days', 'pct_online_transactions',
       'Binoculars', 'CameraAccessory', 'CameraBag', 'CameraBattery',
       'CameraBatteryCharger', 'CameraBatteryGrip', 'CameraEyeCup',
       'CameraFilmRolls', 'CameraHousing', 'CameraMicrophone', 'CameraMount',
       'CameraRemoteControl', 'CameraTripod', 'ExtensionTube', 'Filter',
       'Flash', 'FlashShoeAdapter', 'Lens', 'ReflectorUmbrella', 'Softbox',
       'Strap', 'Telescope'],
      dtype='object')
```

OLS regression model with 12 selected features using RFE

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    gmv   R-squared:                       0.998
Model:                            OLS   Adj. R-squared:                  0.998
Method:                 Least Squares   F-statistic:                     1438.
Date:                Sat, 29 Feb 2020   Prob (F-statistic):           9.86e-36
Time:                        22:20:06   Log-Likelihood:                 146.97
No. Observations:                  41   AIC:                            -267.9
Df Residuals:                      28   BIC:                            -245.7
Df Model:                          12
Covariance Type:            nonrobust
========================================================================================
                           coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------------
const                  821.8660    417.348      1.969      0.059     -33.032    1676.764
Year                  -108.0362     54.850     -1.970      0.059    -220.391       4.319
adStock_Radio            0.0650      0.027      2.402      0.023       0.010       0.121
sla                      0.0922      0.048      1.902      0.068      -0.007       0.191
Binoculars               0.0336      0.139      0.242      0.810      -0.250       0.318
CameraAccessory         -0.1921      0.093     -2.066      0.048      -0.382      -0.002
CameraBatteryCharger     0.3299      0.150      2.203      0.036       0.023       0.637
CameraMount              0.1225      0.064      1.908      0.067      -0.009       0.254
CameraRemoteControl     -0.1237      0.157     -0.788      0.438      -0.445       0.198
CameraTripod             0.5275      0.119      4.426      0.000       0.283       0.772
Flash                    0.0384      0.044      0.867      0.394      -0.052       0.129
Lens                     0.3440      0.040      8.559      0.000       0.262       0.426
Telescope               -0.0206      0.071     -0.288      0.775      -0.167       0.126
==============================================================================
Omnibus:                        1.191   Durbin-Watson:                   2.122
Prob(Omnibus):                  0.551   Jarque-Bera (JB):                0.646
Skew:                          -0.300   Prob(JB):                        0.724
Kurtosis:                       3.131   Cond. No.                     2.71e+06
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 2.71e+06. This might indicate that there are
strong multicollinearity or other numerical problems.
```

VIFs for the features

	Features	VIF
0	const	1.081907e+11
8	CameraRemoteControl	3.468500e+02
6	CameraBatteryCharger	3.250000e+02
4	Binoculars	2.748900e+02
9	CameraTripod	1.930200e+02
12	Telescope	1.346100e+02
5	CameraAccessory	1.242400e+02
1	Year	1.133000e+02
7	CameraMount	1.066400e+02
2	adStock_Radio	8.835000e+01
10	Flash	5.293000e+01
11	Lens	3.080000e+01
3	sla	2.299000e+01

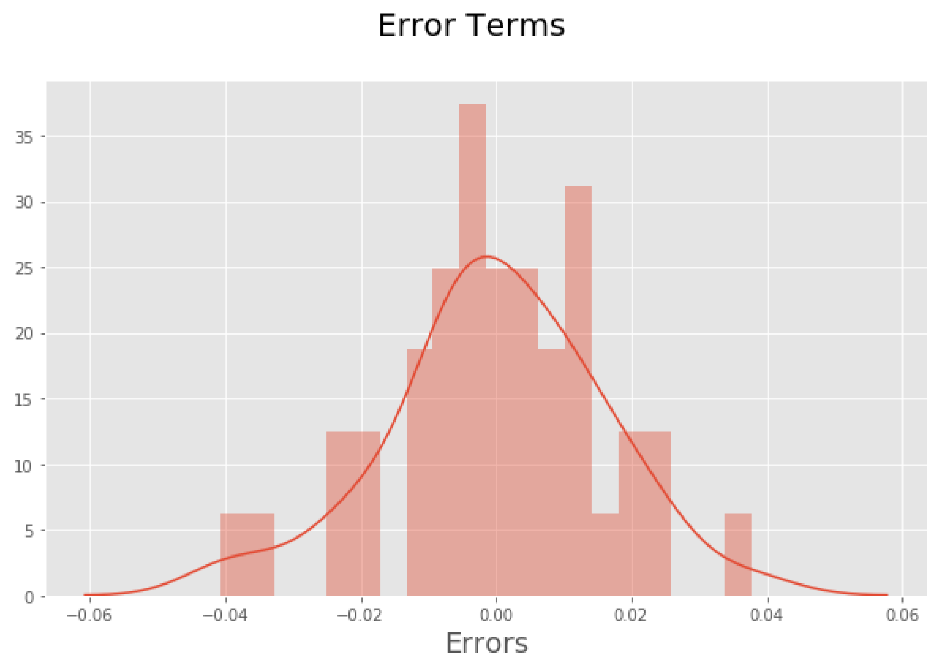
Final model after removing the features where VIFs is greater than 5 and p-value is greater than 0.05

```
                            OLS Regression Results                            
==============================================================================
Dep. Variable:                    gmv   R-squared:                       0.991
Model:                            OLS   Adj. R-squared:                  0.990
Method:                 Least Squares   F-statistic:                     2086.
Date:                Sat, 29 Feb 2020   Prob (F-statistic):           1.43e-39
Time:                        22:34:06   Log-Likelihood:                 111.75
No. Observations:                  41   AIC:                            -217.5
Df Residuals:                      38   BIC:                            -212.4
Df Model:                           2                                         
Covariance Type:            nonrobust                                         
================================================================================
                   coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------
const           -0.0404      0.016     -2.518      0.016      -0.073      -0.008
CameraTripod     0.4775      0.027     17.729      0.000       0.423       0.532
Lens             0.6035      0.023     26.543      0.000       0.557       0.650
==============================================================================
Omnibus:                        1.403   Durbin-Watson:                   1.852
Prob(Omnibus):                  0.496   Jarque-Bera (JB):                0.641
Skew:                          -0.257   Prob(JB):                        0.726
Kurtosis:                       3.335   Cond. No.                         20.8
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

	Features	VIF
0	const	38.96
1	CameraTripod	2.40
2	Lens	2.40

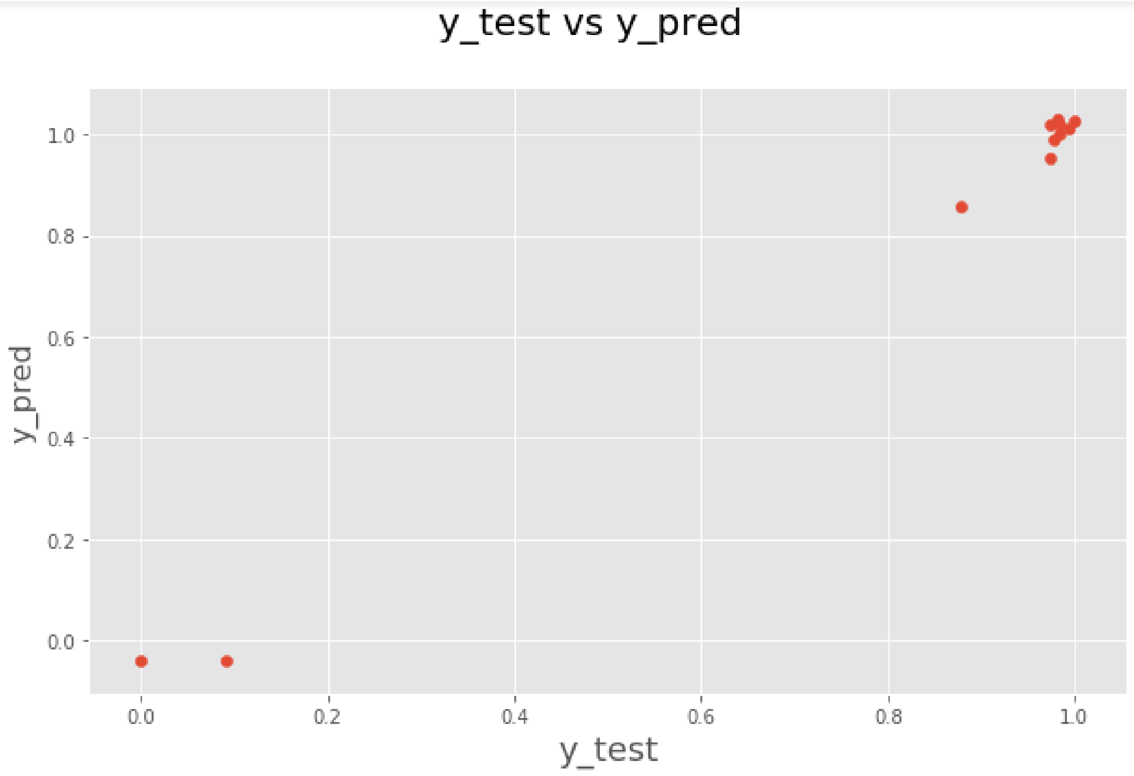
The variance explained by this model is 86%

Residual Analysis

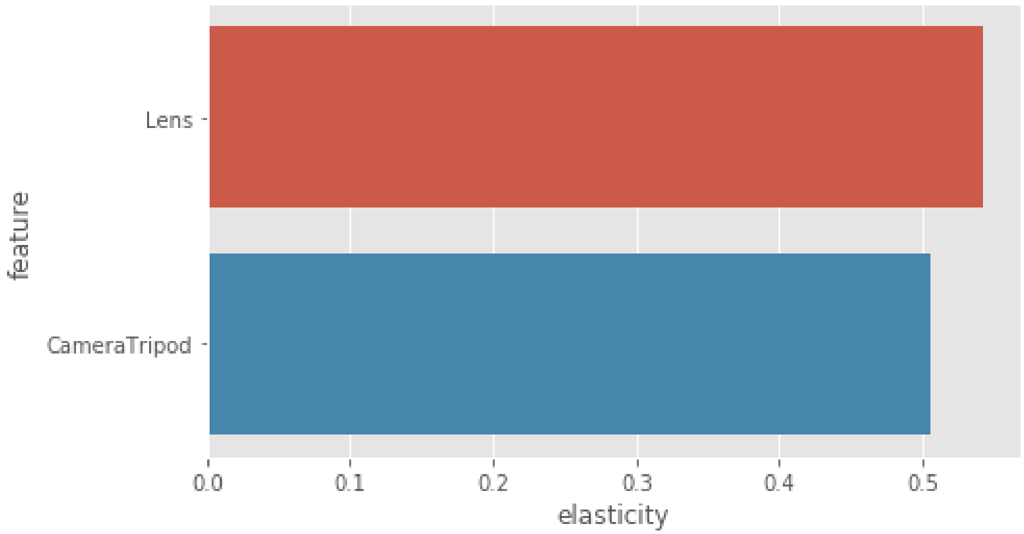


Model Evaluation

Plotting y_test and y_pred to understand the spread



Elasticity Analysis



III. Home Audio - Multiplicative model

Train-Test split and scaling of the data

```
# Test-train split
df_train, df_test = train_test_split(df, train_size = 0.8, test_size = 0.2, random_state = 100)
```

```
scaler = MinMaxScaler()
```

```
# Checking columns
df.columns
```

```
Index(['order_week', 'Month', 'Year', 'adStock_TV', 'adStock_Digital',
       'adStock_Sponsorship', 'adStock_Content Marketing',
       'adStock_Online marketing', 'adStock_ Affiliates', 'adStock_SEM',
       'adStock_Radio', 'adStock_Other', 'gmv', 'sla',
       'product_procurement_sla', 'discount', 'NPS', 'total_holiday',
       'is_holiday', 'total_special_sale_day', 'is_special_sale_day',
       'total_pay_days', 'is_pay_days', 'pct_online_transactions', 'BoomBox',
       'DJController', 'Dock', 'DockingStation', 'FMRadio', 'HiFiSystem',
       'HomeAudioSpeaker', 'SlingBox', 'SoundMixer', 'VoiceRecorder'],
      dtype='object')
```

OLS regression model with 12 selected features using RFE

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    gmv   R-squared:                       0.999
Model:                            OLS   Adj. R-squared:                  0.999
Method:                 Least Squares   F-statistic:                     3093.
Date:                Sat, 29 Feb 2020   Prob (F-statistic):           4.26e-39
Time:                        22:52:49   Log-Likelihood:                 167.80
No. Observations:                  40   AIC:                            -309.6
Df Residuals:                      27   BIC:                            -287.6
Df Model:                          12
Covariance Type:            nonrobust
============================================================================================
                               coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------------------
const                    -1181.7147    295.380     -4.001      0.000   -1787.785    -575.644
Year                       155.3175     38.823      4.001      0.000      75.659     234.976
adStock_TV                   0.0473      0.021      2.218      0.035       0.004       0.091
adStock_Sponsorship         -0.0137      0.011     -1.251      0.222      -0.036       0.009
adStock_Online marketing     0.0426      0.122      0.349      0.730      -0.208       0.293
adStock_ Affiliates         -0.0839      0.127     -0.661      0.515      -0.344       0.177
adStock_Radio                0.1452      0.062      2.360      0.026       0.019       0.271
adStock_Other               -0.2268      0.070     -3.250      0.003      -0.370      -0.084
sla                         -0.0189      0.012     -1.630      0.115      -0.043       0.005
FMRadio                      0.1059      0.028      3.789      0.001       0.049       0.163
HiFiSystem                   0.0280      0.043      0.654      0.519      -0.060       0.116
HomeAudioSpeaker             0.8692      0.018     49.640      0.000       0.833       0.905
VoiceRecorder                0.0171      0.040      0.422      0.676      -0.066       0.100
==============================================================================
Omnibus:                        1.999   Durbin-Watson:                   1.892
Prob(Omnibus):                  0.368   Jarque-Bera (JB):                1.873
Skew:                          -0.455   Prob(JB):                        0.392
Kurtosis:                       2.455   Cond. No.                     3.42e+06
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 3.42e+06. This might indicate that there are
strong multicollinearity or other numerical problems.
```

VIFs for the features

	Features	VIF
0	const	1.771571e+11
5	adStock_ Affiliates	2.599730e+03
4	adStock_Online marketing	2.426360e+03
7	adStock_Other	1.947590e+03
6	adStock_Radio	1.485630e+03
1	Year	1.878700e+02
10	HiFiSystem	8.265000e+01
12	VoiceRecorder	7.282000e+01
2	adStock_TV	6.799000e+01
9	FMRadio	3.254000e+01
3	adStock_Sponsorship	1.628000e+01
11	HomeAudioSpeaker	1.166000e+01
8	sla	9.110000e+00

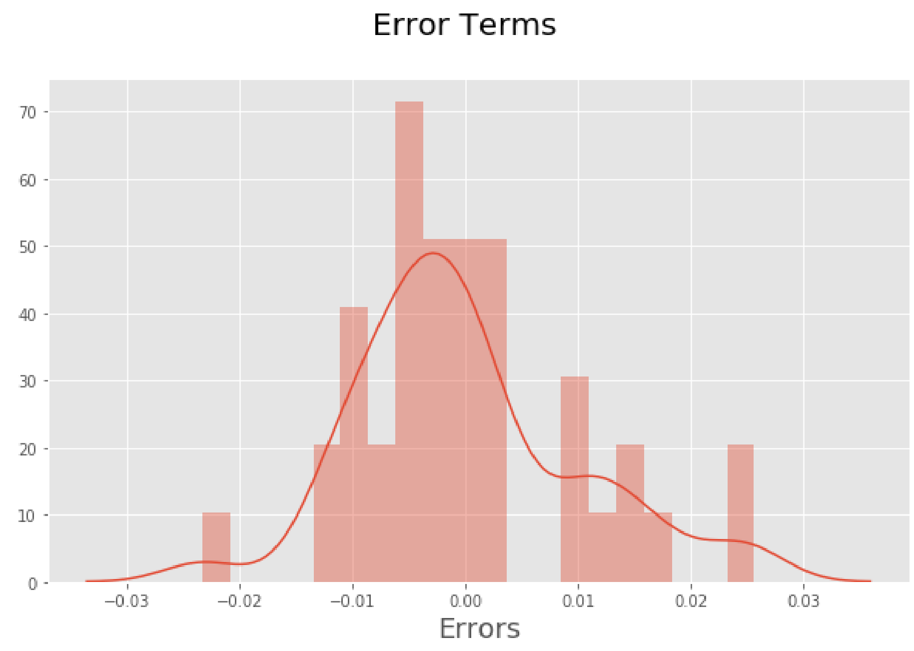
Final model after removing the features where VIFs is greater than 5 and p-value is greater than 0.05

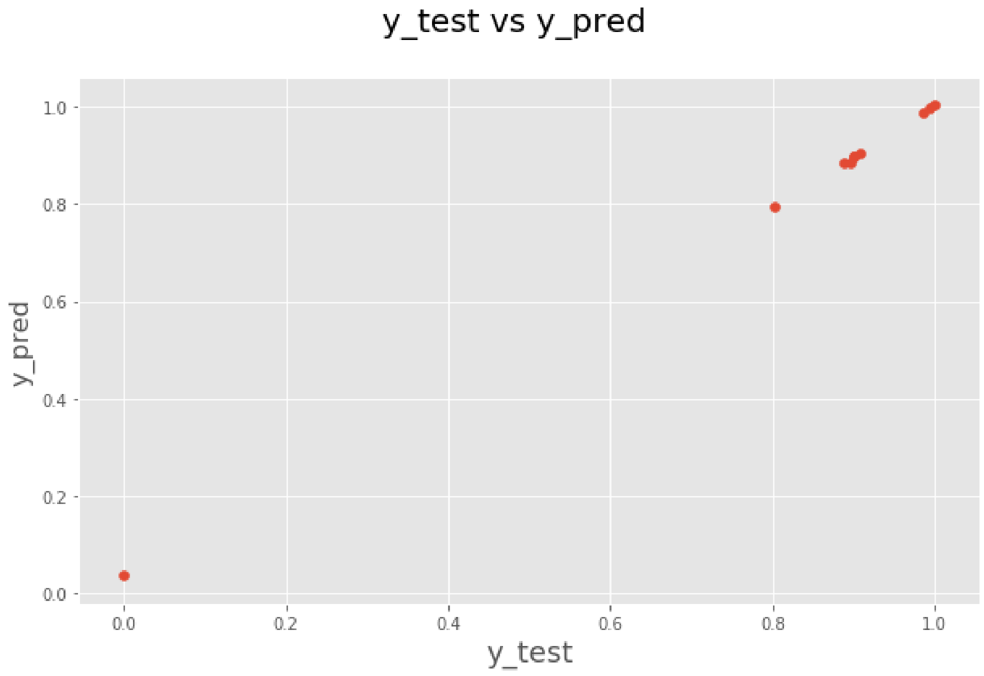
```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    gmv   R-squared:                       0.995
Model:                            OLS   Adj. R-squared:                  0.994
Method:                 Least Squares   F-statistic:                     3352.
Date:                Sat, 29 Feb 2020   Prob (F-statistic):           1.52e-42
Time:                        22:55:51   Log-Likelihood:                 127.37
No. Observations:                  40   AIC:                            -248.7
Df Residuals:                      37   BIC:                            -243.7
Df Model:                           2
Covariance Type:            nonrobust
====================================================================================
                       coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------------
const                0.0233      0.009      2.598      0.013       0.005       0.041
sla                  0.0139      0.011      1.312      0.198      -0.008       0.035
HomeAudioSpeaker     0.9755      0.014     69.093      0.000       0.947       1.004
==============================================================================
Omnibus:                        3.801   Durbin-Watson:                   1.771
Prob(Omnibus):                  0.150   Jarque-Bera (JB):                2.656
Skew:                           0.595   Prob(JB):                        0.265
Kurtosis:                       3.419   Cond. No.                         14.1
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

	Features	VIF
0	const	29.64
1	sla	1.38
2	HomeAudioSpeaker	1.38

The variance explained by this model is 88%

Residual Analysis

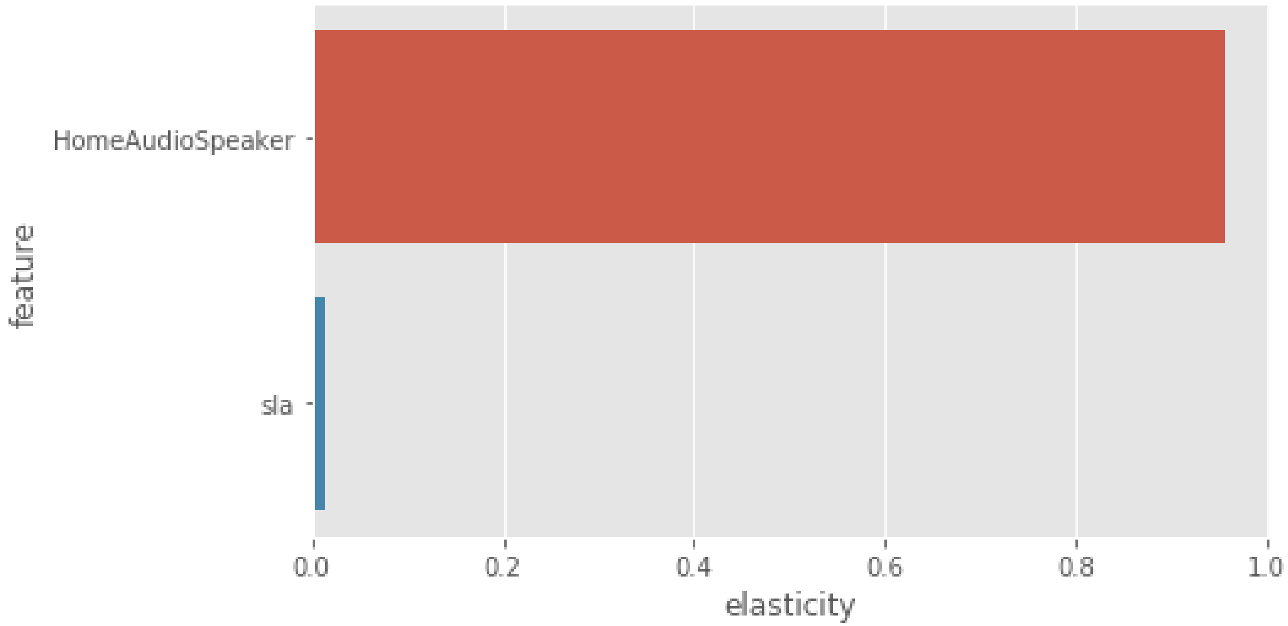




Model Evaluation

Plotting y_test and y_pred to understand the spread

Elasticity Analysis



Building the Kyock's Models

I. Gaming Accessory - Kyock's model

Train-Test split and scaling of the data

```
df_train, df_test = train_test_split(df, train_size = 0.8, test_size = 0.20, random_state = 100)
```

```
scaler = MinMaxScaler()
```

```
df.columns
```

```
Index(['order_week', 'Month', 'Year', 'adStock_TV', 'adStock_Digital',
       'adStock_Sponsorship', 'adStock_Content Marketing',
       'adStock_Online marketing', 'adStock_ Affiliates', 'adStock_SEM',
       'adStock_Radio', 'adStock_Other', 'gmv', 'sla',
       'product_procurement_sla', 'discount', 'NPS', 'total_holiday',
       'is_holiday', 'total_special_sale_day', 'is_special_sale_day',
       'total_pay_days', 'is_pay_days', 'pct_online_transactions',
       'CoolingPad', 'GameControlMount', 'GamePad', 'GamingAccessoryKit',
       'GamingAdapter', 'GamingChargingStation', 'GamingHeadset',
       'GamingKeyboard', 'GamingMemoryCard', 'GamingMouse', 'GamingMousePad',
       'GamingSpeaker', 'JoystickGamingWheel', 'MotionController',
       'TVOutCableAccessory', 'gmv_lag'],
      dtype='object')
```

OLS regression model with 12 selected features using RFE

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    gmv   R-squared:                       0.973
Model:                            OLS   Adj. R-squared:                  0.961
Method:                 Least Squares   F-statistic:                     85.50
Date:                Sat, 29 Feb 2020   Prob (F-statistic):           2.75e-19
Time:                        23:03:52   Log-Likelihood:                 91.701
No. Observations:                  42   AIC:                            -157.4
Df Residuals:                      29   BIC:                            -134.8
Df Model:                          12
Covariance Type:            nonrobust
============================================================================================
                              coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------------------
const                      -0.0246      0.061     -0.401      0.691      -0.150       0.101
adStock_TV                  0.3156      0.152      2.074      0.047       0.004       0.627
adStock_Digital             1.6961      0.677      2.504      0.018       0.311       3.081
adStock_Sponsorship         0.1964      0.085      2.316      0.028       0.023       0.370
adStock_Online marketing   -0.1361      0.112     -1.215      0.234      -0.365       0.093
adStock_SEM                -2.0008      0.764     -2.620      0.014      -3.563      -0.439
adStock_Radio               1.5386      0.562      2.740      0.010       0.390       2.687
adStock_Other              -1.4004      0.551     -2.540      0.017      -2.528      -0.273
NPS                        -0.0811      0.065     -1.240      0.225      -0.215       0.053
GamingAccessoryKit          0.2725      0.043      6.342      0.000       0.185       0.360
GamingHeadset               0.2180      0.060      3.605      0.001       0.094       0.342
GamingMemoryCard            0.1005      0.048      2.095      0.045       0.002       0.199
GamingMouse                 0.5297      0.068      7.771      0.000       0.390       0.669
==============================================================================
Omnibus:                        2.153   Durbin-Watson:                   1.629
Prob(Omnibus):                  0.341   Jarque-Bera (JB):                1.626
Skew:                           0.482   Prob(JB):                        0.444
Kurtosis:                       3.002   Cond. No.                         391.
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

VIFs for the features

	Features	VIF
5	adStock_SEM	1290.35
2	adStock_Digital	1010.65
6	adStock_Radio	781.09
7	adStock_Other	753.52
0	const	146.82
1	adStock_TV	74.03
4	adStock_Online marketing	51.26
3	adStock_Sponsorship	17.48
8	NPS	10.07
10	GamingHeadset	7.75
11	GamingMemoryCard	5.00
12	GamingMouse	3.79
9	GamingAccessoryKit	3.09

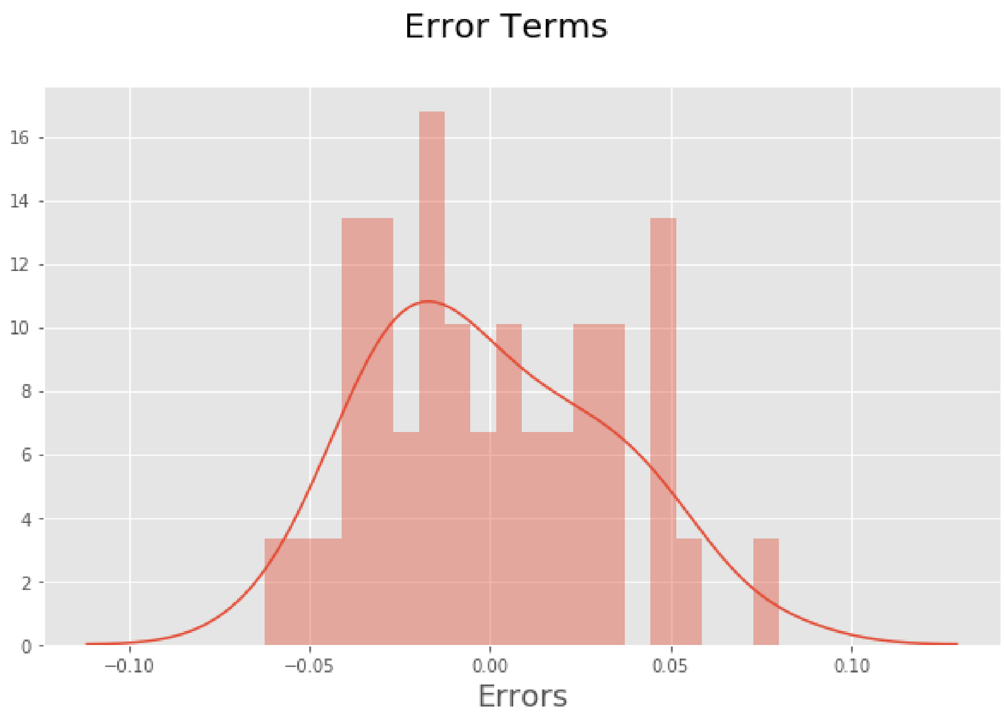
Final model after removing the features where VIFs is greater than 5 and p-value is greater than 0.05

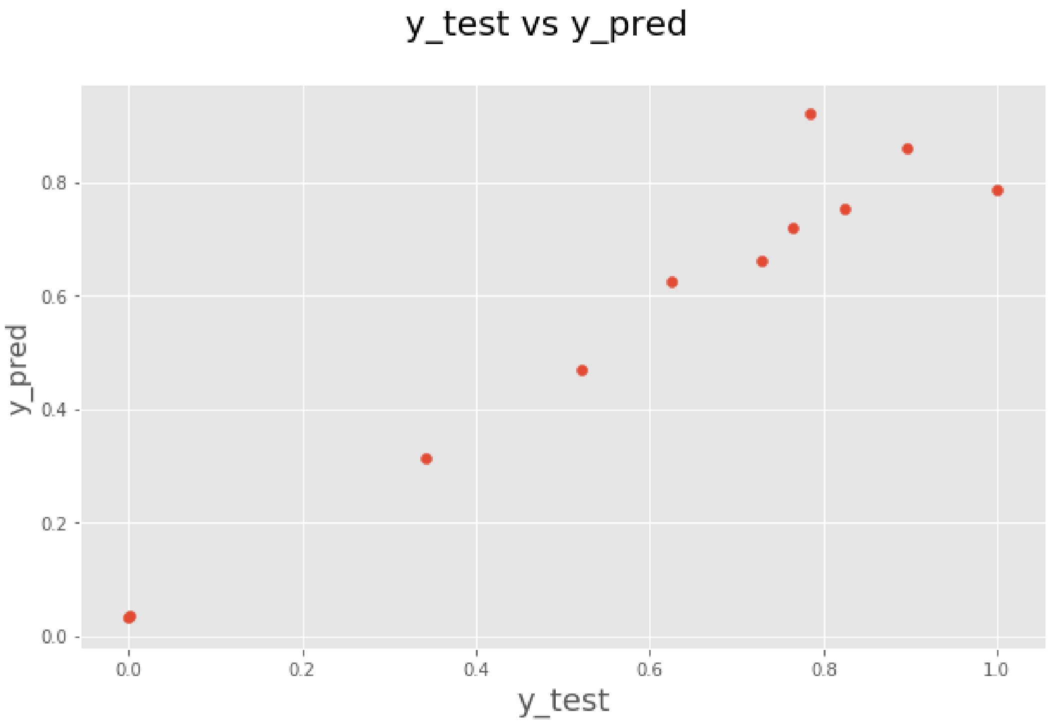
```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    gmv   R-squared:                       0.961
Model:                            OLS   Adj. R-squared:                  0.956
Method:                 Least Squares   F-statistic:                     177.9
Date:                Sat, 29 Feb 2020   Prob (F-statistic):           2.49e-24
Time:                        23:07:32   Log-Likelihood:                 84.413
No. Observations:                  42   AIC:                            -156.8
Df Residuals:                      36   BIC:                            -146.4
Df Model:                           5
Covariance Type:            nonrobust
====================================================================================
                       coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------------
const                0.0345      0.015      2.268      0.029       0.004       0.065
adStock_SEM         -0.0582      0.028     -2.097      0.043      -0.114      -0.002
adStock_Radio        0.0849      0.024      3.517      0.001       0.036       0.134
GamingAccessoryKit   0.3153      0.035      8.934      0.000       0.244       0.387
GamingHeadset        0.1808      0.038      4.784      0.000       0.104       0.257
GamingMouse          0.5597      0.058      9.685      0.000       0.442       0.677
==============================================================================
Omnibus:                        1.756   Durbin-Watson:                   1.602
Prob(Omnibus):                  0.416   Jarque-Bera (JB):                1.620
Skew:                           0.375   Prob(JB):                        0.445
Kurtosis:                       2.397   Cond. No.                         14.4
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

	Features	VIF
0	const	7.94
4	GamingHeadset	2.66
5	GamingMouse	2.39
3	GamingAccessoryKit	1.83
1	adStock_SEM	1.49
2	adStock_Radio	1.27

The variance explained by this model is 91%

Residual Analysis

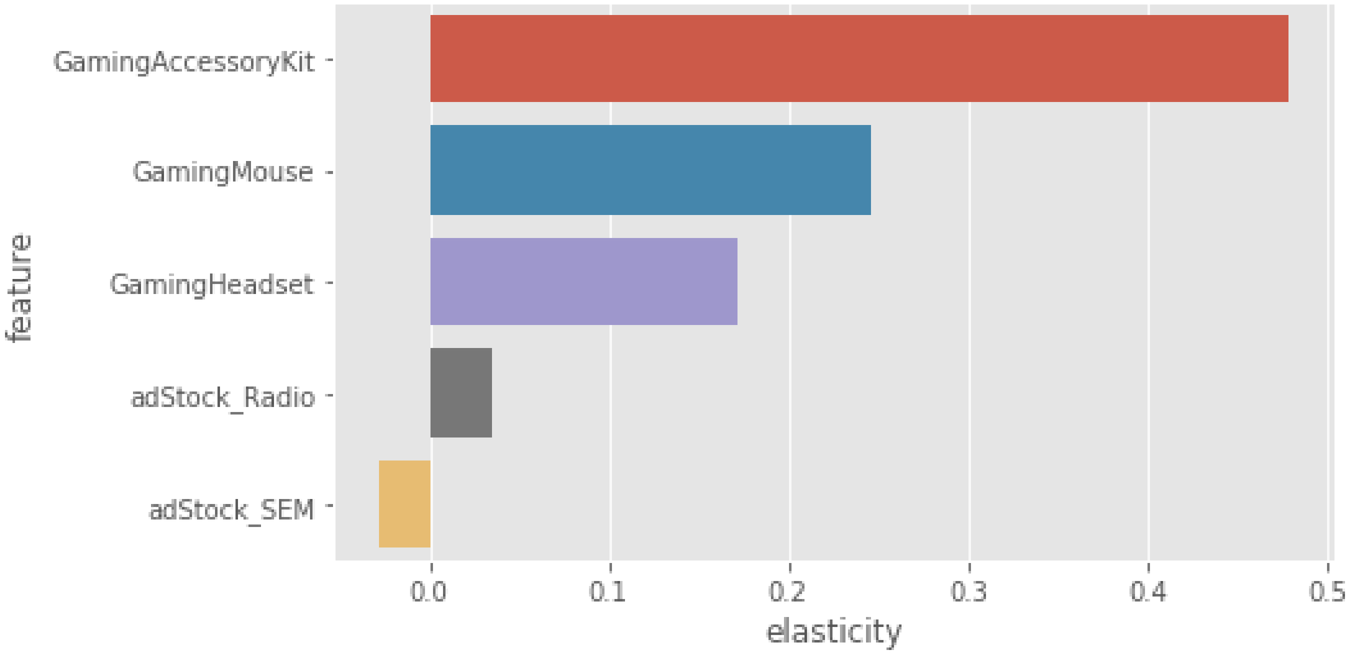




Model Evaluation

Plotting y_test and y_pred to understand the spread

Elasticity - Gaming Accessory



II. Camera Accessory - Kyock's Model

Train-Test split and scaling of the data

```
# Test-train split
df_train, df_test = train_test_split(df, train_size = 0.8, test_size = 0.2, random_state = 100)
```

```
scaler = MinMaxScaler()
```

```
df.columns
```

```
Index(['order_week', 'Month', 'Year', 'adStock_TV', 'adStock_Digital',
       'adStock_Sponsorship', 'adStock_Content Marketing',
       'adStock_Online marketing', 'adStock_ Affiliates', 'adStock_SEM',
       'adStock_Radio', 'adStock_Other', 'gmv', 'sla',
       'product_procurement_sla', 'discount', 'NPS', 'total_holiday',
       'is_holiday', 'total_special_sale_day', 'is_special_sale_day',
       'total_pay_days', 'is_pay_days', 'pct_online_transactions',
       'Binoculars', 'CameraAccessory', 'CameraBag', 'CameraBattery',
       'CameraBatteryCharger', 'CameraBatteryGrip', 'CameraEyeCup',
       'CameraFilmRolls', 'CameraHousing', 'CameraMicrophone', 'CameraMount',
       'CameraRemoteControl', 'CameraTripod', 'ExtensionTube', 'Filter',
       'Flash', 'FlashShoeAdapter', 'Lens', 'ReflectorUmbrella', 'Softbox',
       'Strap', 'Telescope', 'gmv_lag'],
      dtype='object')
```

OLS regression model with 12 selected features using RFE

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    gmv   R-squared:                       0.991
Model:                            OLS   Adj. R-squared:                  0.987
Method:                 Least Squares   F-statistic:                     261.3
Date:                Sat, 29 Feb 2020   Prob (F-statistic):           2.02e-25
Time:                        23:16:03   Log-Likelihood:                 113.81
No. Observations:                  41   AIC:                            -201.6
Df Residuals:                      28   BIC:                            -179.3
Df Model:                          12
Covariance Type:            nonrobust
============================================================================================
                               coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------------------
const                        0.0577      0.021      2.811      0.009       0.016       0.100
adStock_Digital             -0.6313      0.203     -3.107      0.004      -1.048      -0.215
adStock_Online marketing    -0.4264      0.212     -2.012      0.054      -0.861       0.008
adStock_ Affiliates          0.3443      0.178      1.938      0.063      -0.020       0.708
adStock_SEM                  0.7266      0.239      3.035      0.005       0.236       1.217
adStock_Radio               -0.1889      0.079     -2.404      0.023      -0.350      -0.028
adStock_Other                0.2087      0.074      2.820      0.009       0.057       0.360
product_procurement_sla      0.0728      0.019      3.739      0.001       0.033       0.113
discount                    -0.1267      0.025     -4.985      0.000      -0.179      -0.075
CameraBag                    0.4747      0.047     10.107      0.000       0.378       0.571
CameraBattery                0.1206      0.042      2.854      0.008       0.034       0.207
CameraTripod                 0.2917      0.032      9.153      0.000       0.226       0.357
Flash                        0.2481      0.029      8.437      0.000       0.188       0.308
==============================================================================
Omnibus:                        0.815   Durbin-Watson:                   1.798
Prob(Omnibus):                  0.665   Jarque-Bera (JB):                0.885
Skew:                           0.235   Prob(JB):                        0.642
Kurtosis:                       2.455   Cond. No.                         243.
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

VIFs for the features

	Features	VIF
2	adStock_Online marketing	543.46
4	adStock_SEM	405.26
3	adStock_ Affiliates	388.02
1	adStock_Digital	294.86
0	const	51.97
5	adStock_Radio	49.34
6	adStock_Other	43.94
12	Flash	8.51
10	CameraBattery	7.14
9	CameraBag	6.91
11	CameraTripod	4.54
8	discount	4.28
7	product_procurement_sla	1.89

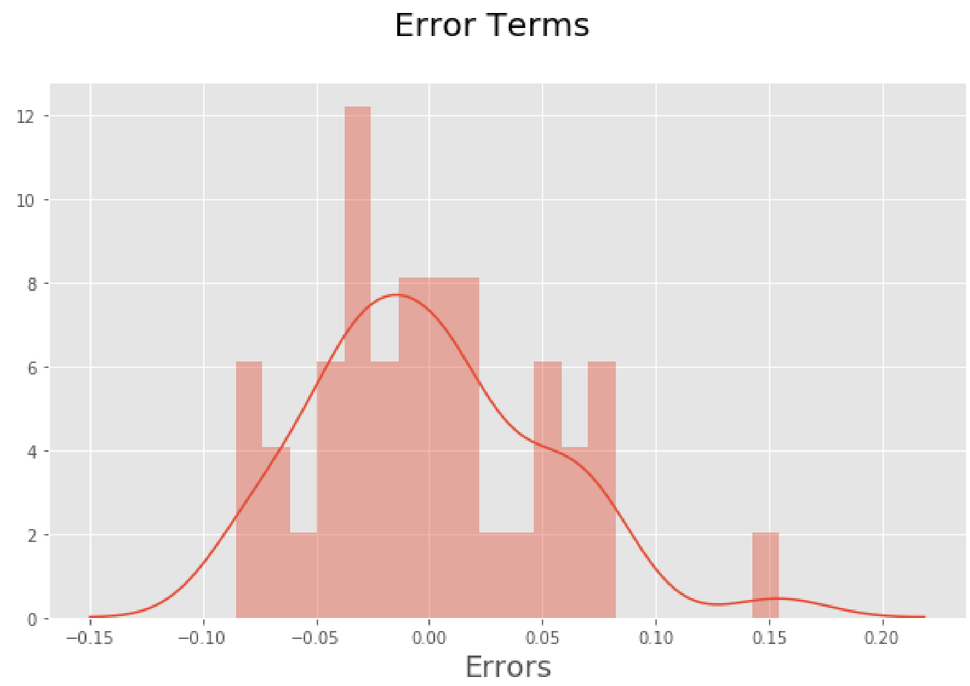
Final model after removing the features where VIFs is greater than 5 and p-value is greater than 0.05

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    gmv   R-squared:                       0.900
Model:                            OLS   Adj. R-squared:                  0.895
Method:                 Least Squares   F-statistic:                     170.6
Date:                Sat, 29 Feb 2020   Prob (F-statistic):           1.04e-19
Time:                        23:25:47   Log-Likelihood:                 64.059
No. Observations:                  41   AIC:                            -122.1
Df Residuals:                      38   BIC:                            -117.0
Df Model:                           2
Covariance Type:            nonrobust
================================================================================
                   coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------
const            0.0680      0.020      3.412      0.002       0.028       0.108
CameraBag        0.8012      0.060     13.400      0.000       0.680       0.922
CameraTripod     0.2100      0.050      4.196      0.000       0.109       0.311
==============================================================================
Omnibus:                        4.190   Durbin-Watson:                   1.866
Prob(Omnibus):                  0.123   Jarque-Bera (JB):                3.044
Skew:                           0.634   Prob(JB):                        0.218
Kurtosis:                       3.415   Cond. No.                         9.16
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

	Features	VIF
0	const	5.86
1	CameraBag	1.34
2	CameraTripod	1.34

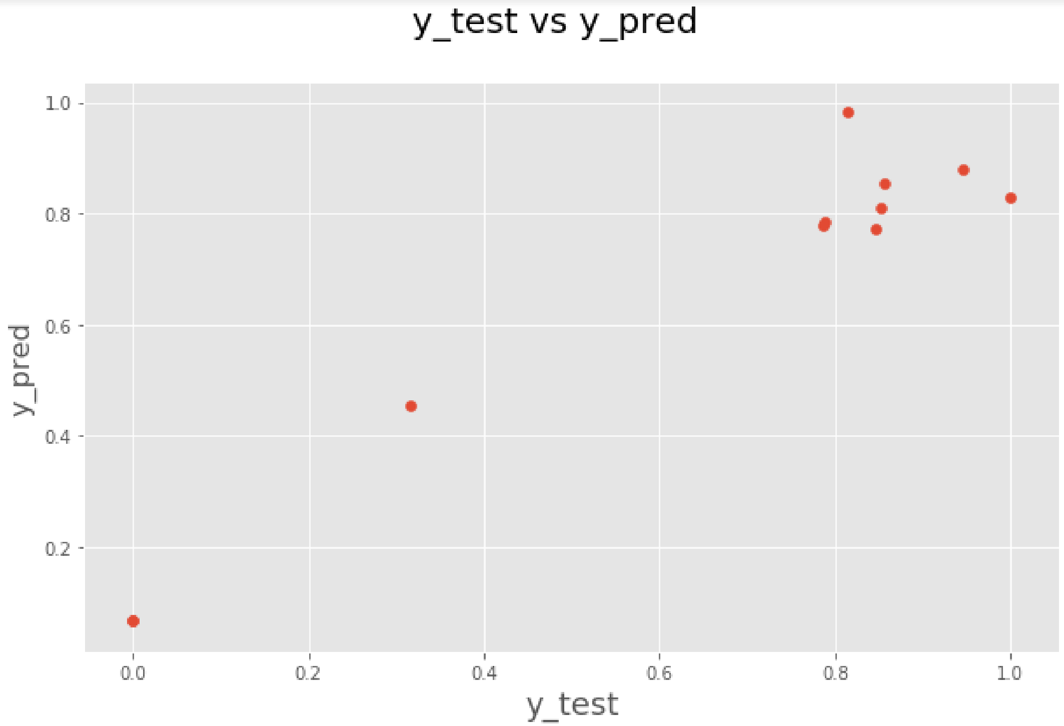
The variance explained by this model is 79%

Residual Analysis

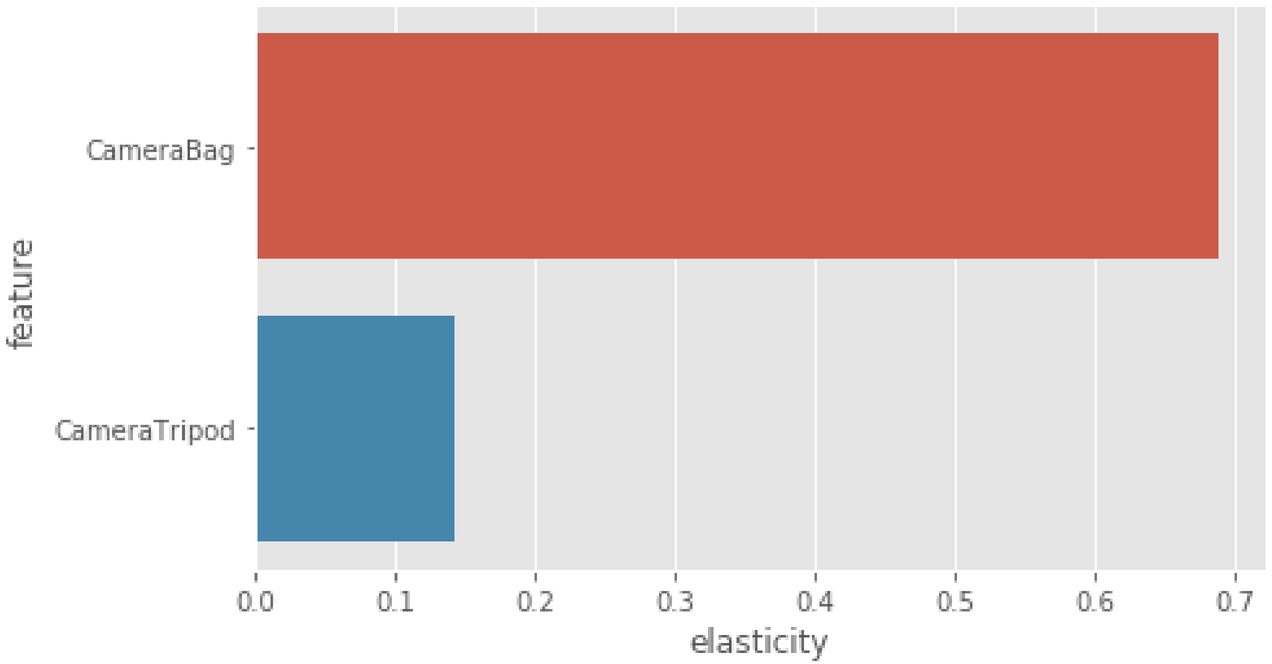


Model Evaluation

Plotting y_test and y_pred to understand the spread



Elasticity



CameraBag
CameraTripod
feature
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
elasticity

III. Home Audio - Kyock's model

Train-Test split and scaling of the data

```
# Test-train split
df_train, df_test = train_test_split(df, train_size = 0.8, test_size = 0.2, random_state = 100)
```

```
scaler = MinMaxScaler()
```

```
# Apply scaler() to all the columns except the 'yes-no' and 'dummy' variables
num_vars = [
    'adStock_TV', 'adStock_Digital',
    'adStock_Sponsorship', 'adStock_Content Marketing',
    'adStock_Online marketing', 'adStock_ Affiliates', 'adStock_SEM',
    'adStock_Radio', 'adStock_Other', 'gmv', 'sla',
    'product_procurement_sla', 'discount', 'NPS', 'total_holiday',
    'total_special_sale_day',
    'total_pay_days', 'pct_online_transactions', 'BoomBox',
    'DJController', 'Dock', 'DockingStation', 'FMRadio', 'HiFiSystem',
    'HomeAudioSpeaker', 'SlingBox', 'SoundMixer', 'VoiceRecorder',
    'gmv_lag'
]

df_train[num_vars] = scaler.fit_transform(df_train[num_vars])
df_test[num_vars] = scaler.fit_transform(df_test[num_vars])
```

```
y_train = df_train.pop('gmv')
X_train = df_train
```

OLS regression model with 12 selected features using RFE

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    gmv   R-squared:                       0.999
Model:                            OLS   Adj. R-squared:                  0.998
Method:                 Least Squares   F-statistic:                     1628.
Date:                Sat, 29 Feb 2020   Prob (F-statistic):           2.43e-35
Time:                        23:30:33   Log-Likelihood:                 152.61
No. Observations:                  40   AIC:                            -279.2
Df Residuals:                      27   BIC:                            -257.3
Df Model:                          12
Covariance Type:            nonrobust
=============================================================================================
                                coef    std err          t      P>|t|      [0.025      0.975]
---------------------------------------------------------------------------------------------
const                         0.0325      0.008      3.902      0.001       0.015       0.050
adStock_TV                    0.0719      0.067      1.070      0.294      -0.066       0.210
adStock_Sponsorship          -0.0136      0.038     -0.361      0.721      -0.091       0.064
adStock_Content Marketing     0.0875      0.108      0.812      0.424      -0.133       0.308
adStock_Online marketing     -0.2897      0.110     -2.625      0.014      -0.516      -0.063
adStock_ Affiliates           0.2151      0.101      2.136      0.042       0.008       0.422
adStock_SEM                  -0.0503      0.092     -0.548      0.588      -0.239       0.138
adStock_Radio                 0.2233      0.127      1.759      0.090      -0.037       0.484
adStock_Other                -0.2434      0.150     -1.626      0.116      -0.551       0.064
pct_online_transactions      -0.0315      0.010     -3.124      0.004      -0.052      -0.011
FMRadio                       0.0331      0.016      2.129      0.042       0.001       0.065
HomeAudioSpeaker              0.9333      0.013     70.486      0.000       0.906       0.961
VoiceRecorder                 0.0455      0.007      6.591      0.000       0.031       0.060
==============================================================================
Omnibus:                        1.952   Durbin-Watson:                   1.830
Prob(Omnibus):                  0.377   Jarque-Bera (JB):                1.333
Skew:                          -0.189   Prob(JB):                        0.513
Kurtosis:                       2.190   Cond. No.                         415.
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

VIFs for the features

	Features	VIF
8	adStock_Other	1423.04
4	adStock_Online marketing	1047.77
7	adStock_Radio	992.97
5	adStock_ Affiliates	898.28
3	adStock_Content Marketing	635.62
6	adStock_SEM	463.45
1	adStock_TV	312.28
2	adStock_Sponsorship	89.04
0	const	65.93
9	pct_online_transactions	5.91
10	FMRadio	5.21
11	HomeAudioSpeaker	3.62
12	VoiceRecorder	1.59

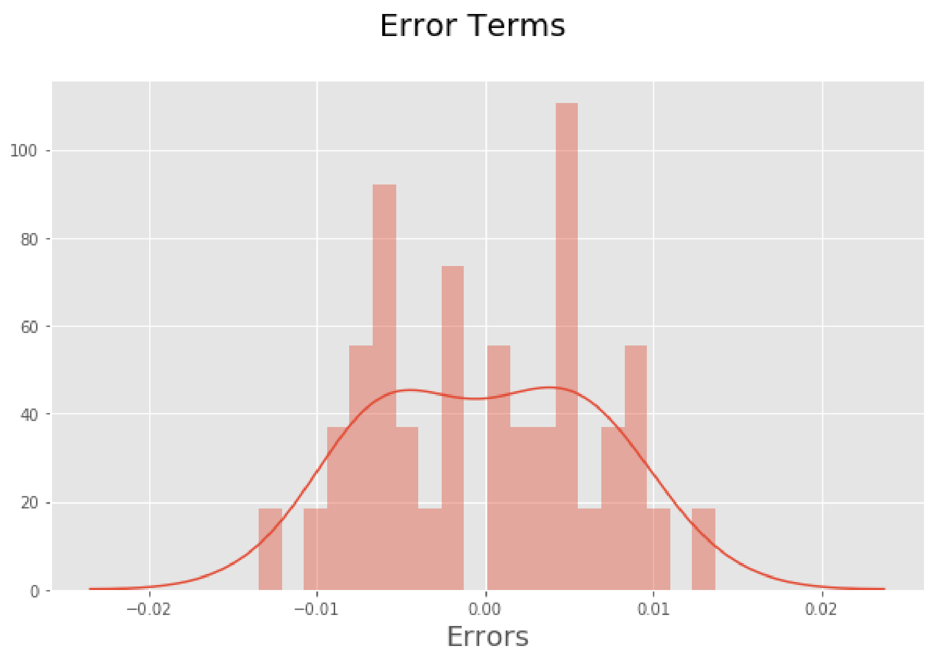
Final model after removing the features where VIFs is greater than 5 and p-value is greater than 0.05

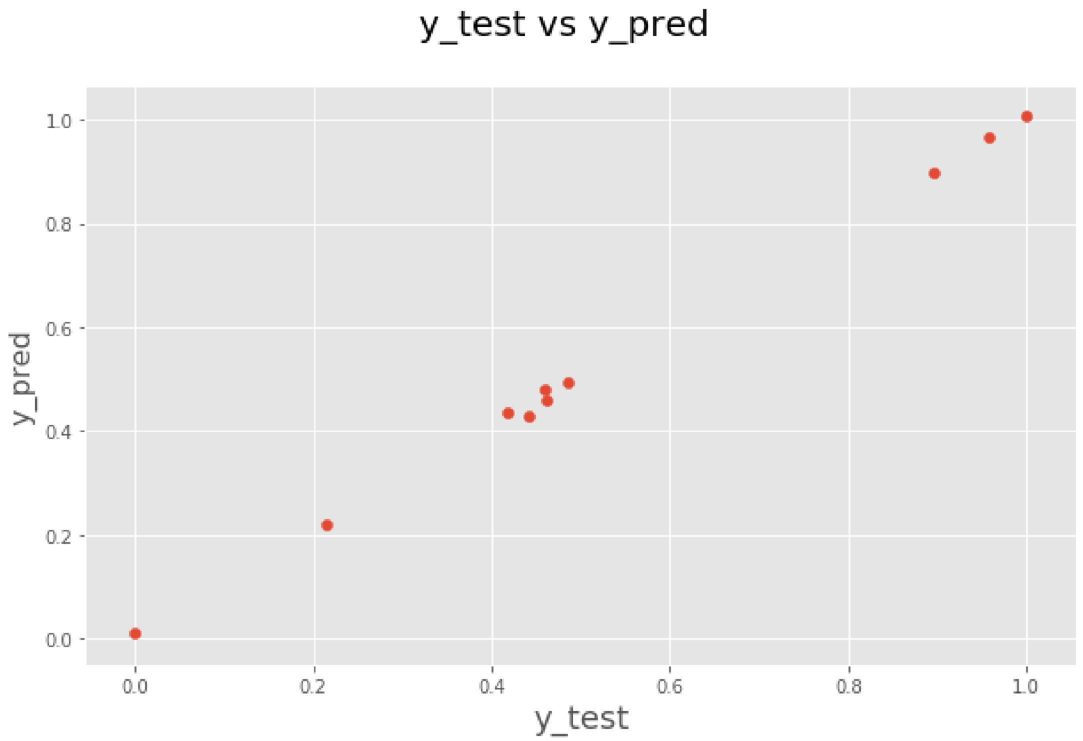
```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    gmv   R-squared:                       0.998
Model:                            OLS   Adj. R-squared:                  0.998
Method:                 Least Squares   F-statistic:                     3303.
Date:                Sat, 29 Feb 2020   Prob (F-statistic):           1.21e-44
Time:                        23:34:46   Log-Likelihood:                 144.65
No. Observations:                  40   AIC:                            -277.3
Df Residuals:                      34   BIC:                            -267.2
Df Model:                           5
Covariance Type:            nonrobust
===========================================================================================
                              coef    std err          t      P>|t|      [0.025      0.975]
-------------------------------------------------------------------------------------------
const                       0.0233      0.004      6.031      0.000       0.015       0.031
adStock_Other              -0.0108      0.005     -2.169      0.037      -0.021      -0.001
pct_online_transactions    -0.0136      0.005     -2.676      0.011      -0.024      -0.003
FMRadio                     0.0459      0.013      3.637      0.001       0.020       0.072
HomeAudioSpeaker            0.9244      0.012     75.309      0.000       0.899       0.949
VoiceRecorder               0.0405      0.006      6.351      0.000       0.028       0.054
==============================================================================
Omnibus:                        3.228   Durbin-Watson:                   1.933
Prob(Omnibus):                  0.199   Jarque-Bera (JB):                1.537
Skew:                           0.020   Prob(JB):                        0.464
Kurtosis:                       2.040   Cond. No.                         17.3
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

	Features	VIF
0	const	11.98
3	FMRadio	2.90
4	HomeAudioSpeaker	2.63
1	adStock_Other	1.34
2	pct_online_transactions	1.27
5	VoiceRecorder	1.15

The variance explained by this model is 99%

Residual Analysis

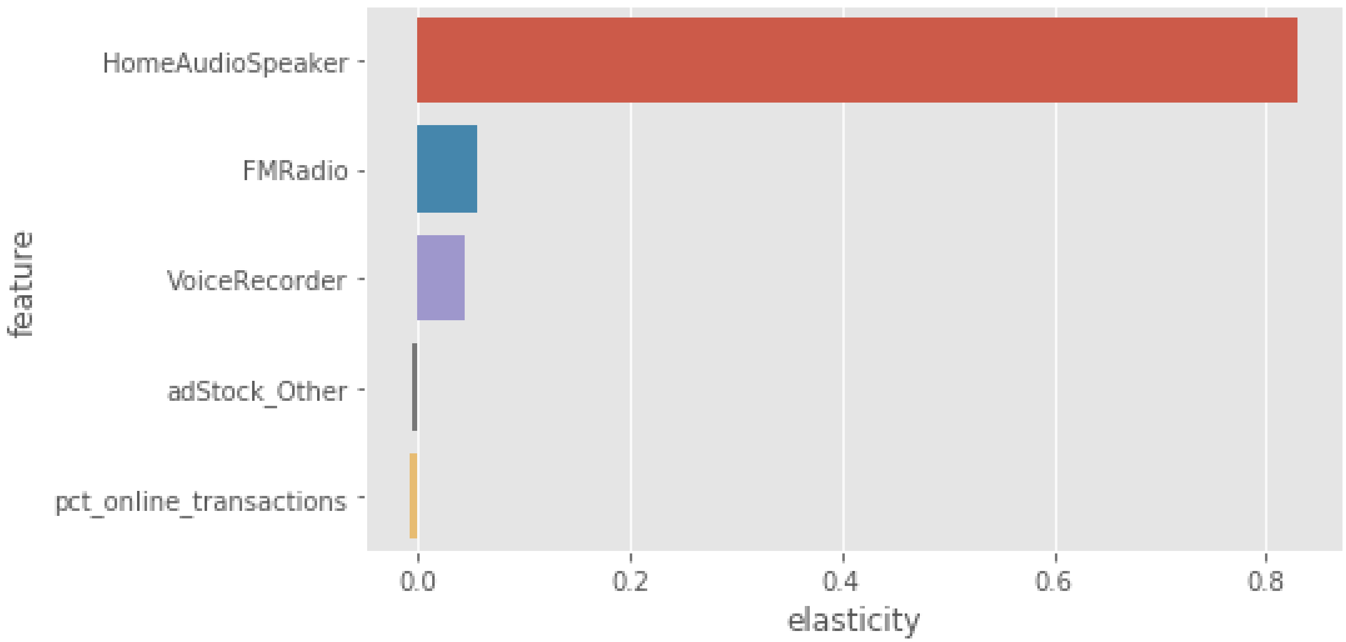




Model Evaluation

Plotting y_test and y_pred to understand the spread

Elasticity



HomeAudioSpeaker
FMRadio
VoiceRecorder
adStock_Other
pct_online_transactions
feature
0.0
0.2
0.4
0.6
0.8
elasticity

Best model for each category

1. **GamingAccessory**
    - **Kyock's Model**
        - Highest Adjusted R-square
        - Low MSE value
        - Features on which company can act upon:
            - ➢ GamingMouse
            - ➢ GamingHeadset
            - ➢ GamingAccessoryKit
            - ➢ adStock_Radio
            - ➢ adStock_SEM
2. **CameraAccessory**
    - **Kyock's Model**
        - Highest Adjusted R-square
        - Low MSE value
        - Features on which company can act upon:
            - ➢ Camerabag
            - ➢ CameraTripod
3. **HomeAudio**
    - **Kyock's Model**
        - Highest Adjusted R-square
        - Low MSE value
        - Features on which company can act upon:
            - ➢ HomeAudioSpeaker
            - ➢ FMRadio
            - ➢ VoiceRecorder
            - ➢ adStock_Other
            - ➢ pct_online_transactions