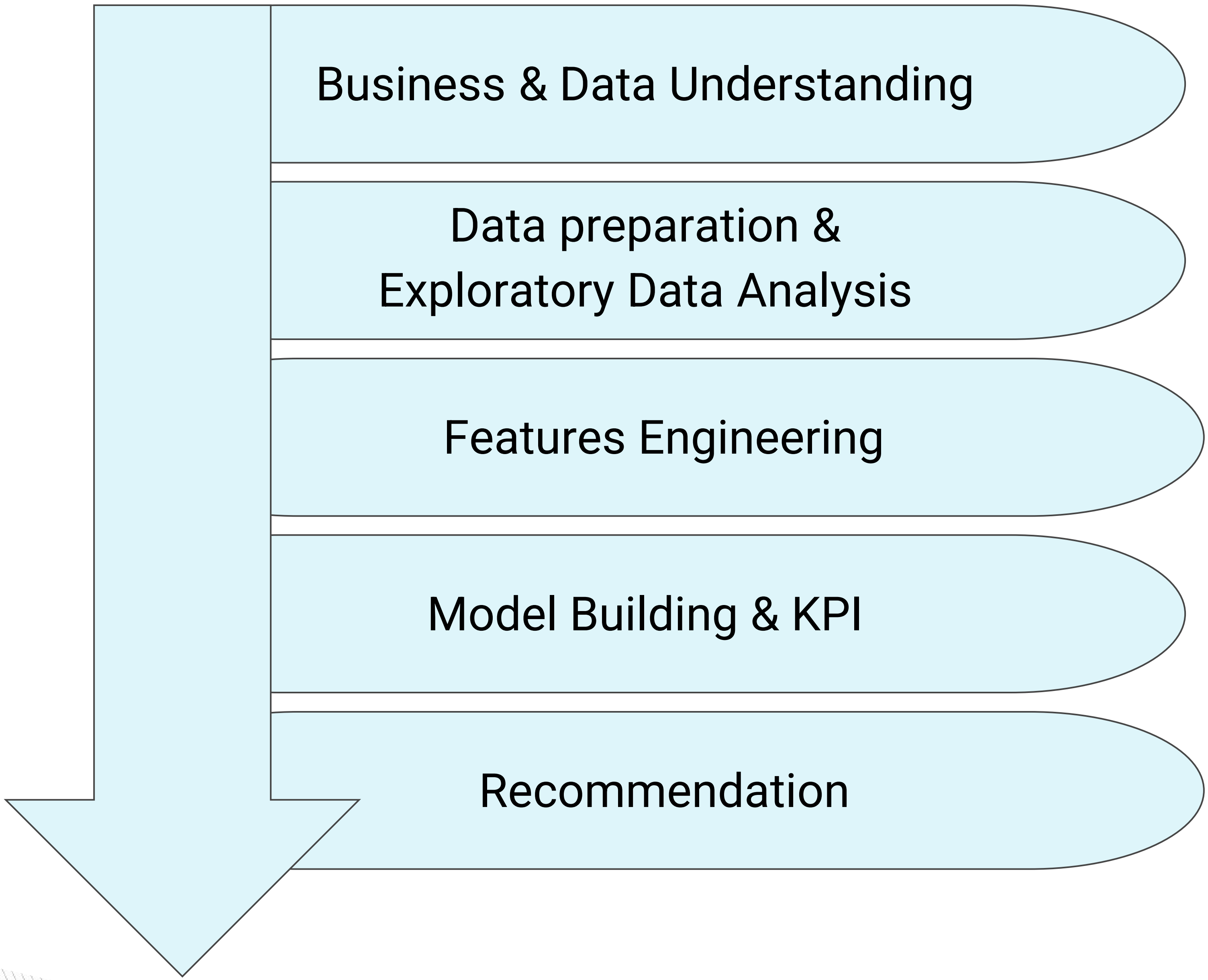
ElecKart

Market Mix Modelling

Capstone Project

Group Members:
AJAY KUMAR YEDLA
AKANSHA PARASHAR
PHANI BHUSHANA RAO K


Business & Data Understanding
Data preparation &
Exploratory Data Analysis
Features Engineering
Model Building & KPI
Recommendation

Objective

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

Our Understanding of the Scope

ElecKart is an e-commerce firm dealing mainly with electronic products

1. Gather e-commerce domain knowledge
2. Understand the given dataset and questions that can be answered
3. Cleaning and prepare the data; with 3 category-wise data sets creation
4. Derived variable creation and grouping by weeks
5. EDA
6. Uni and bi -variate analysis
7. Multi-variate analysis; using different modeling techniques
8. Choose the best model for each segment

Basic understanding of given data

Product

- # of units sold
- Delivery days and SLA
- Categories/sub categories
- Vertical
- Procurement SLA
- Item type = Luxury/Mass-Market

Promotion

- Marketing Channel Investments
- Customer sentiment
- Discounts
- Adstock

Price

- gmv
- Product mrp

Place

- Pin-code
- Order Payment Type
- Week of the year – seasonality
- Holiday / Events; isHoliday

Data Cleaning and Preparation of consumer Electronics

- ConsumerElectronics

1. Removing all the rows having NA and duplicate values.
2. Checking unique values
3. Converting date column’s data type to DATE
4. Filtering out data which does not fall within the timelines of this analysis – 1st July 2015 – 30th June 2016.
5. Creating weeks from the ‘order_id’ data.
6. Converting order_id and order_item_id into proper numeric format - from scientific notation.
7. Removing rows with negative product MRP; gmv and units.
8. Removing rows where (product_mrp*unit) < gmv.
9. Removing rows with negative deliverybdays and deliverycdays; assuming “\N” means no delay.
10. Rarely SLA/procurement SLA for any delivery will be more than 2 months (60 days); hence filtering out these value .
11. Computing discount % for each transaction.
12. Computing gvm/unit.
13. Computing ItemType – categorizing items into Luxury (priced more the 80 %tile) and Mass Market.
14. Removing Columns which will not be used in analysis.
15. Storing the total gmv proportion for each of the 3 categories wrt the total gvm for all items
16. Filtering and keeping only the 3 required categories.

Data Cleaning Preparation of Media data

- Media data and other information.xlsx
  1. Loading all the 4 spreadsheets
  2. Special Sale Calendar –
     1. Splitting the event name and date; also converting the date in proper format.
     2. Creating event start and end date.
     3. Producing a dataframe – having all dates possible within the timeframe of analysis and corresponding event names (if any).
     4. Merging this with consumer data – produced earlier.
  3. Monthly NPS Score –
     1. Transposing the columns into rows.
     2. Cleaning the naming issues wrt months.
     3. Populating the same monthly scores to each day of the month.
     4. Convert it to weekly basis.
     5. Merging this with consumer data – produced earlier.
  4. Media Investment –
     1. Distributing the monthly investing data for each channel into daily investment – proportionate to "days in that month".
     2. Converting it on weekly basis.
     3. Extracting category-wise investment – proportionately to gvm of each of the 3 category wrt the total gvm.
     4. Creating 3 category-wise dataframes - Merging this with consumer data – produced earlier.
     5. Converging CR(10000000) to Total Investment, TV, Digital, Sponsorship, Content-Marketing, Online-Marketing, Affiliates, SEM, Radio, Other attributes.
  5. Any item having gmv/unit more at 80%tile is assumed to be Luxury else Mass-market.

Please note: At the end of this analysis we are producing 3 category wise clean .csv .files

Local Factors

Local factors impacts: Ontario local holidays are considered along with other holidays for said FY(July-15 to June-16).

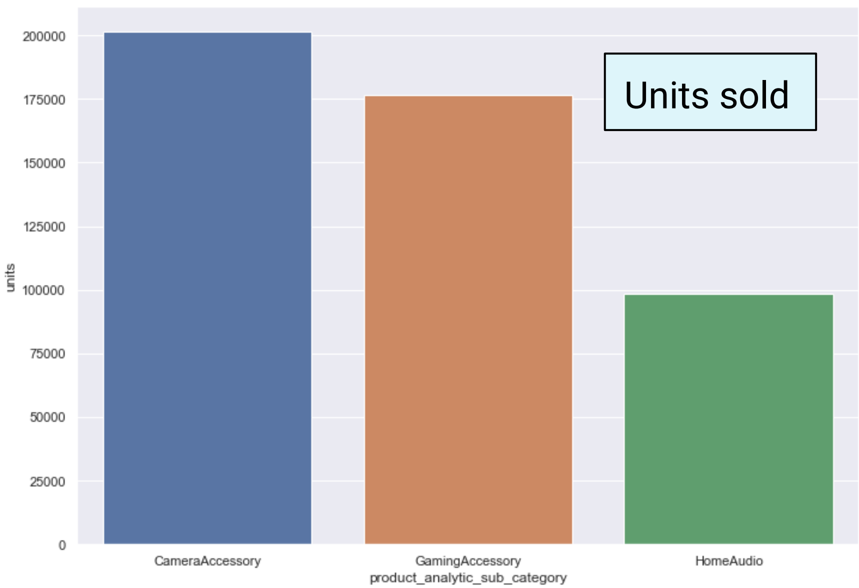
```python
##Local holiday list of state Ontario,Canada for 2016
import holidays
from datetime import date

CA_holidays = holidays.Canada(years=2016, state="ON")

for holiday in CA_holidays.items():
    print(holiday)
```

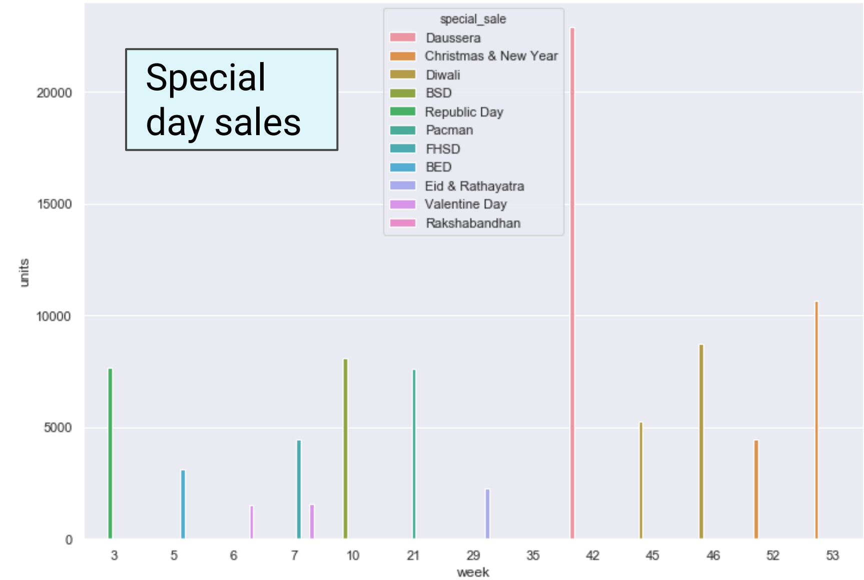
```
(datetime.date(2016, 1, 1), "New Year's Day")
(datetime.date(2016, 2, 15), 'Family Day')
(datetime.date(2016, 3, 25), 'Good Friday')
(datetime.date(2016, 5, 23), 'Victoria Day')
(datetime.date(2016, 7, 1), 'Canada Day')
(datetime.date(2016, 8, 1), 'Civic Holiday')
(datetime.date(2016, 9, 5), 'Labour Day')
(datetime.date(2016, 10, 10), 'Thanksgiving')
(datetime.date(2016, 12, 25), 'Christmas Day')
(datetime.date(2016, 12, 26), 'Christmas Day (Observed)')
(datetime.date(2016, 12, 27), 'Boxing Day (Observed)')
```

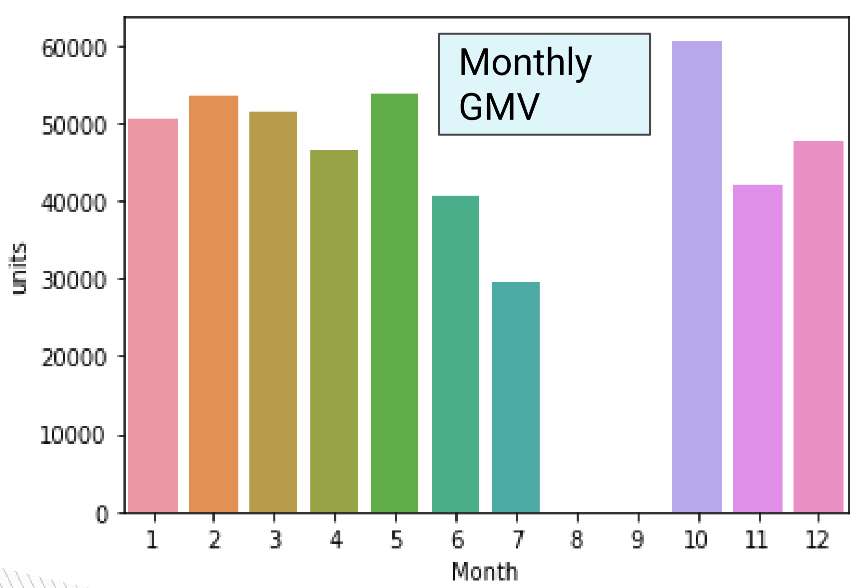
```python
#Local holiday list of state Ontario,Canada for 2015
import holidays
from datetime import date

CA_holidays = holidays.Canada(years=2015, state="ON")

for holiday in CA_holidays.items():
    print(holiday)
```

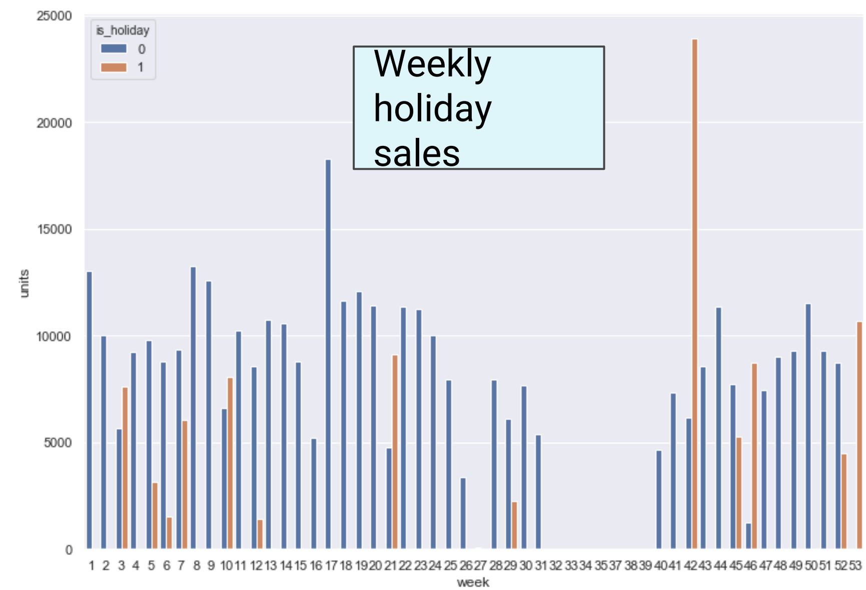
```
(datetime.date(2015, 1, 1), "New Year's Day")
(datetime.date(2015, 2, 16), 'Family Day')
(datetime.date(2015, 4, 3), 'Good Friday')
(datetime.date(2015, 5, 18), 'Victoria Day')
(datetime.date(2015, 7, 1), 'Canada Day')
(datetime.date(2015, 8, 3), 'Civic Holiday')
(datetime.date(2015, 9, 7), 'Labour Day')
(datetime.date(2015, 10, 12), 'Thanksgiving')
(datetime.date(2015, 12, 25), 'Christmas Day')
(datetime.date(2015, 12, 28), 'Boxing Day (Observed)')
```

Bar charts and Graphs of master dataframe

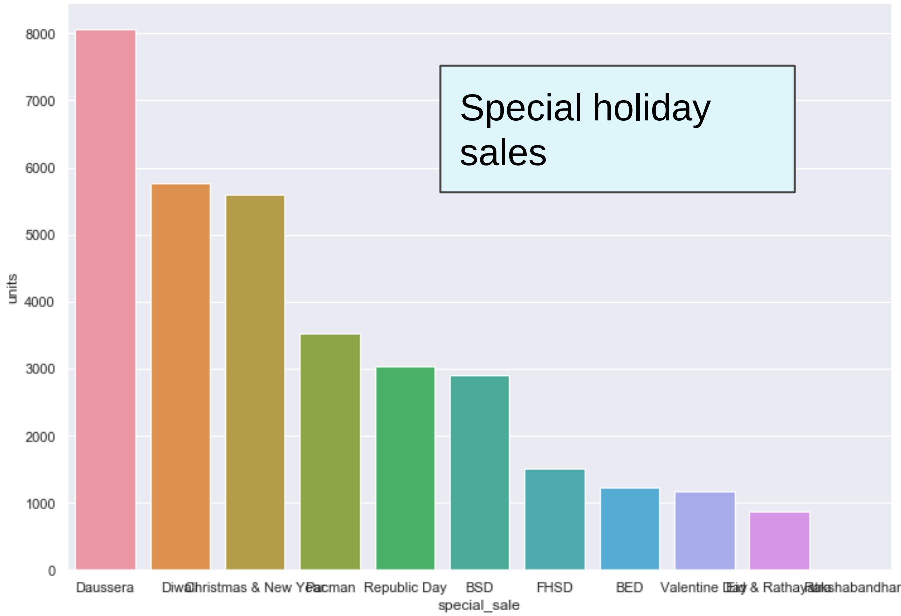


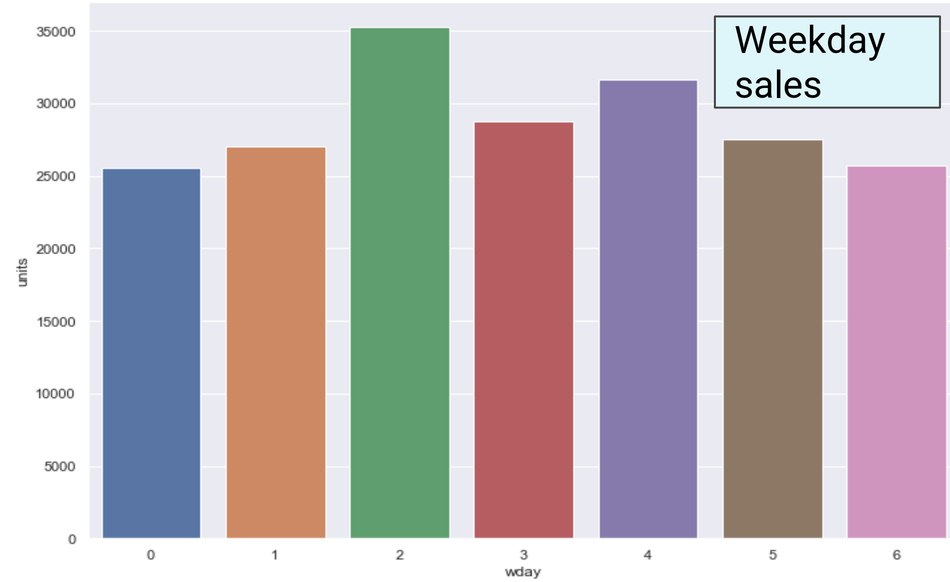


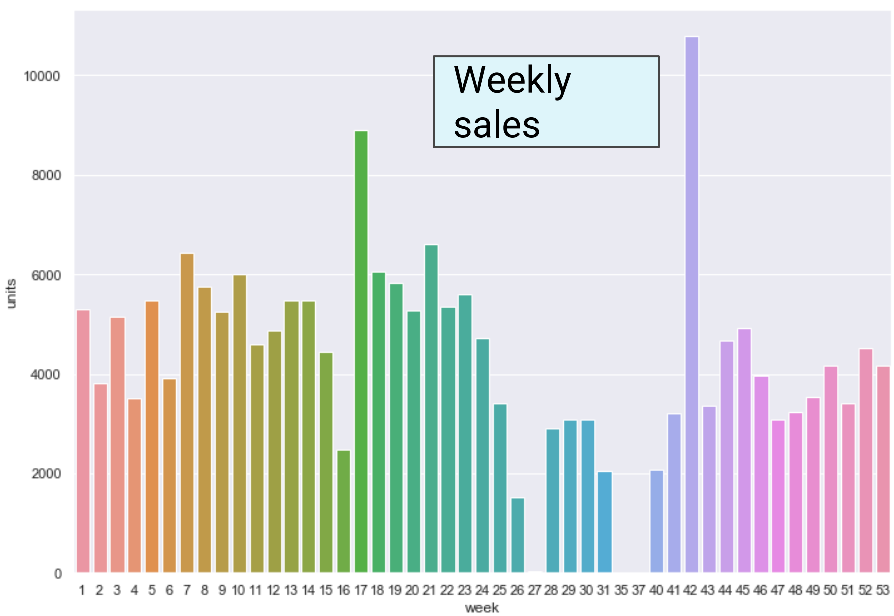


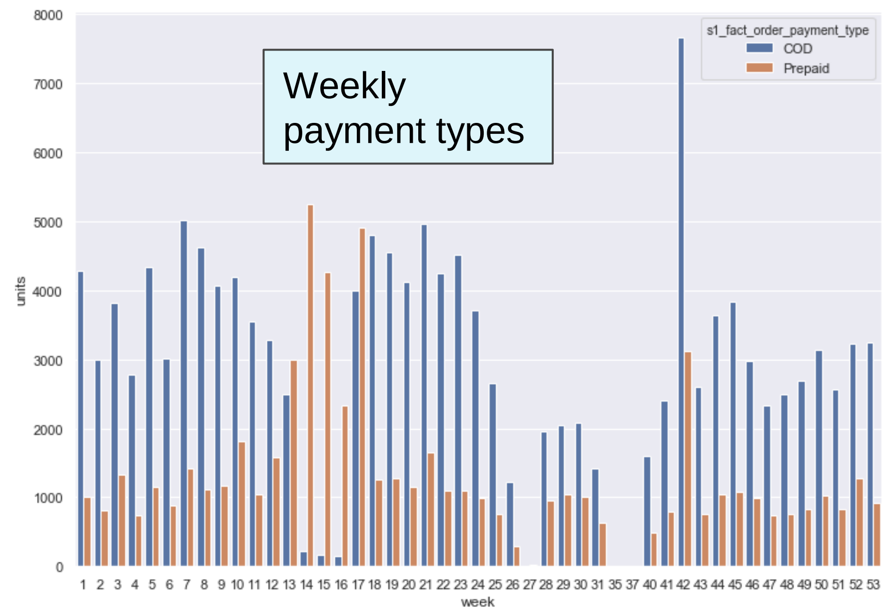


EDA (Camera Accessory)

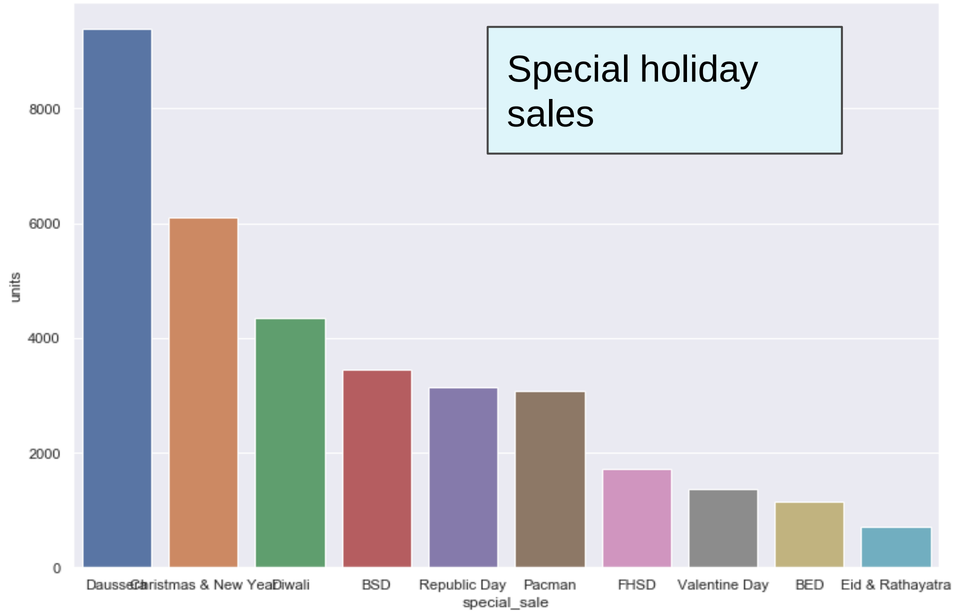


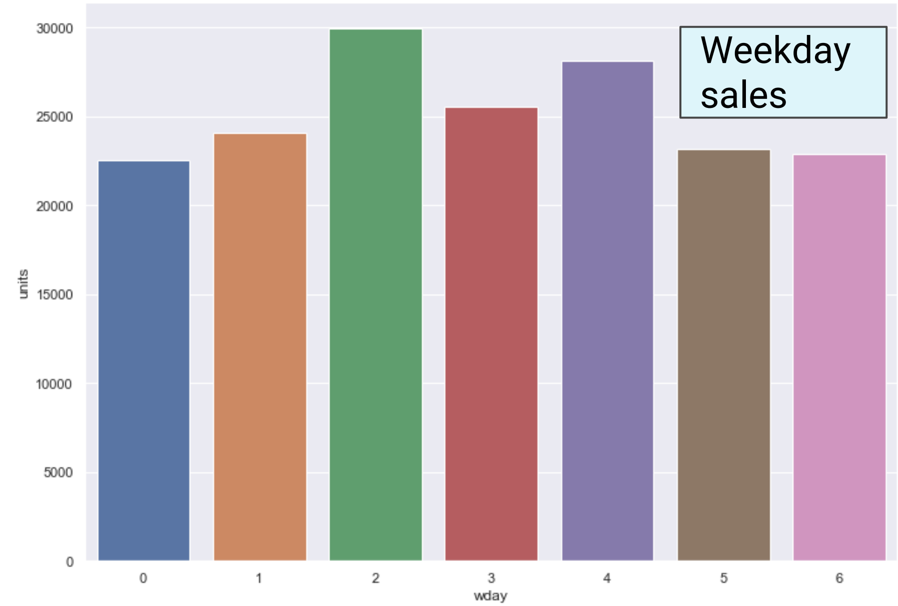


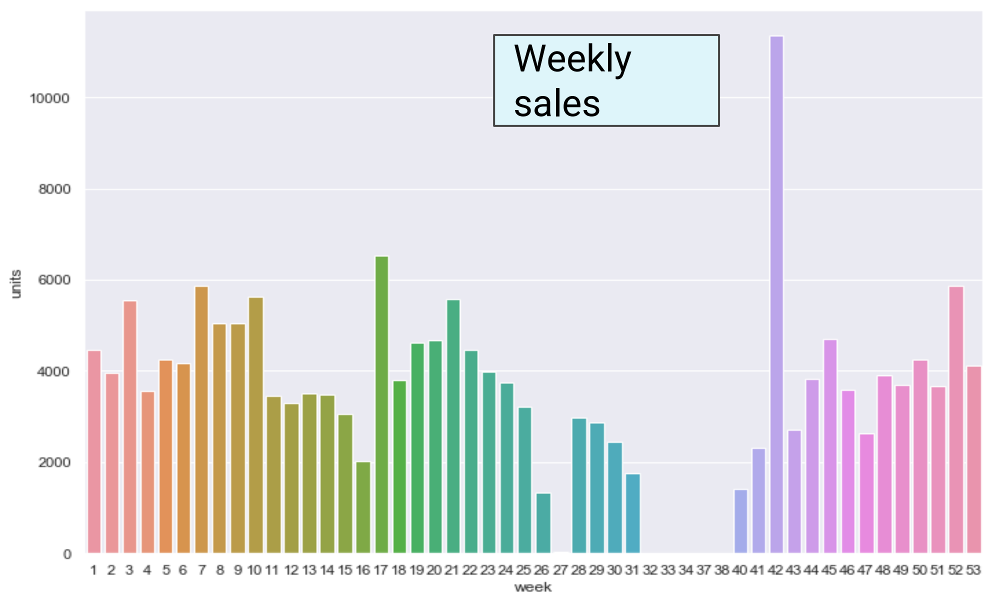


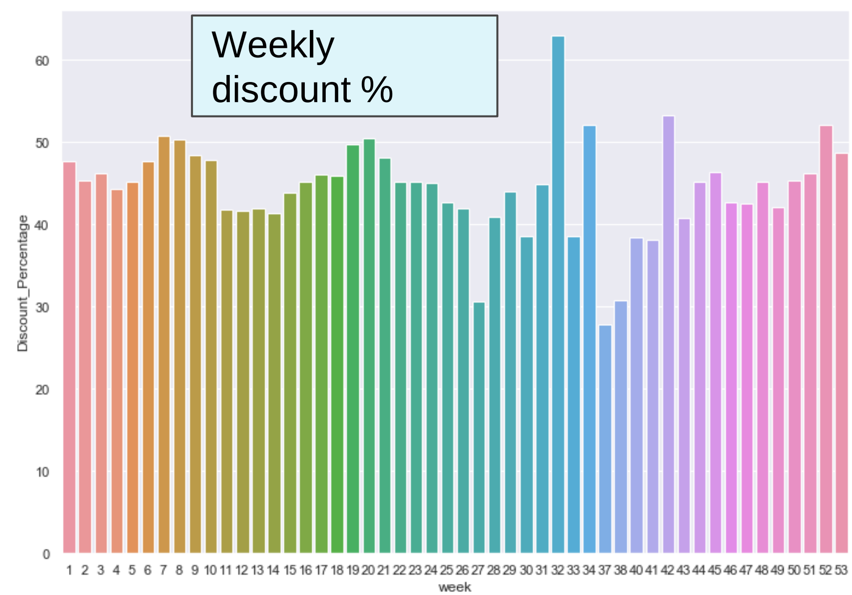


EDA (Gaming Accessory)

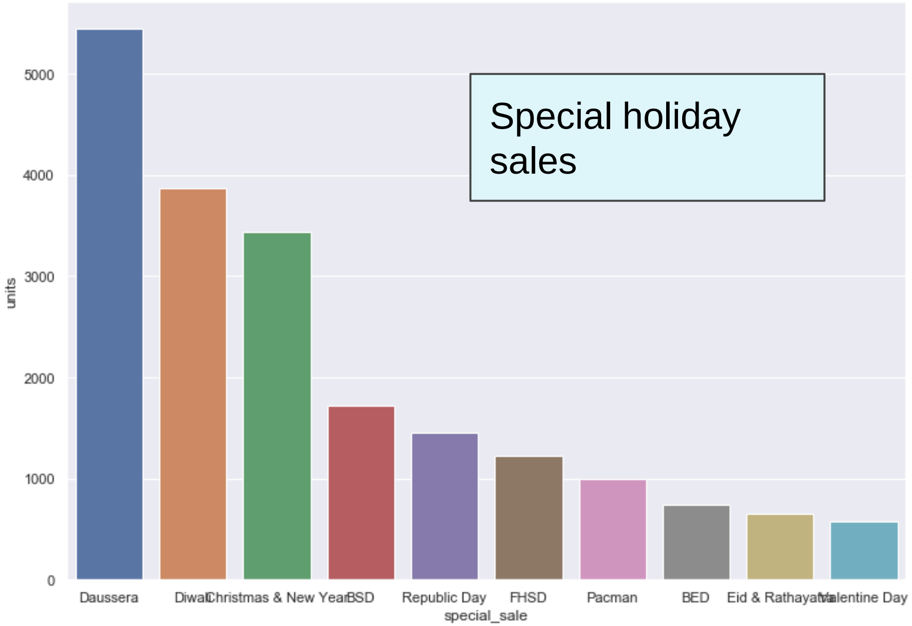


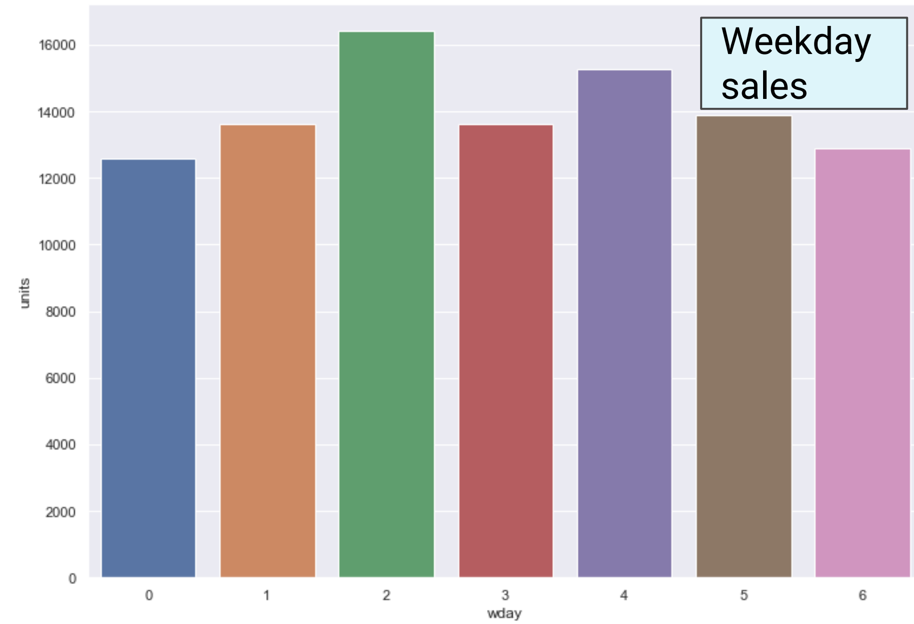


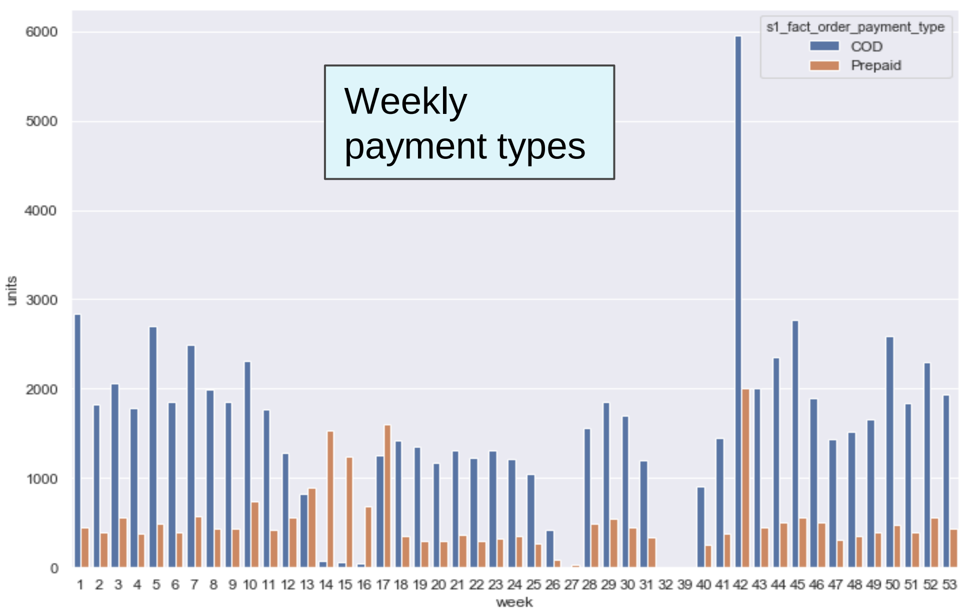


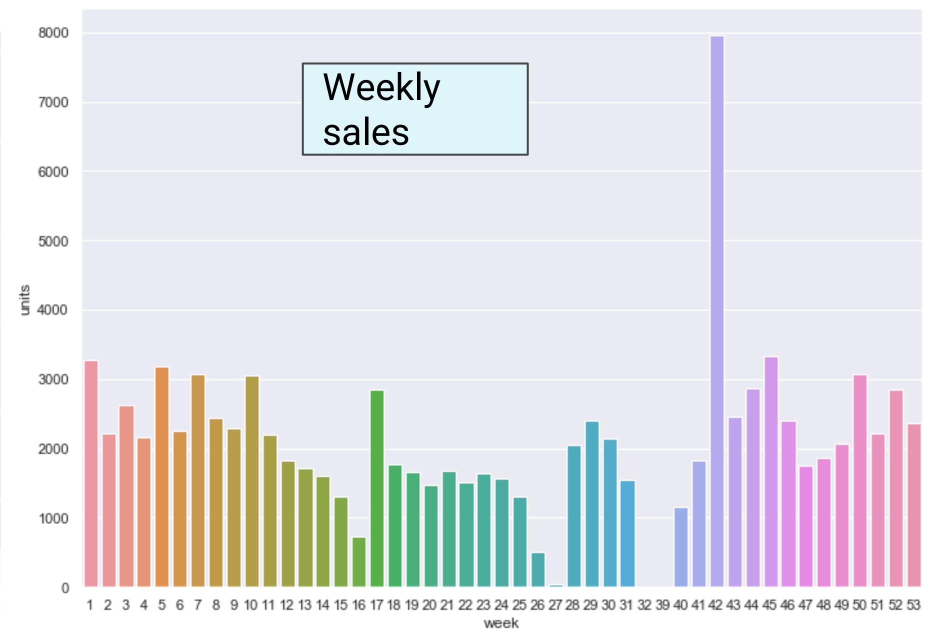


EDA (Home Audio)

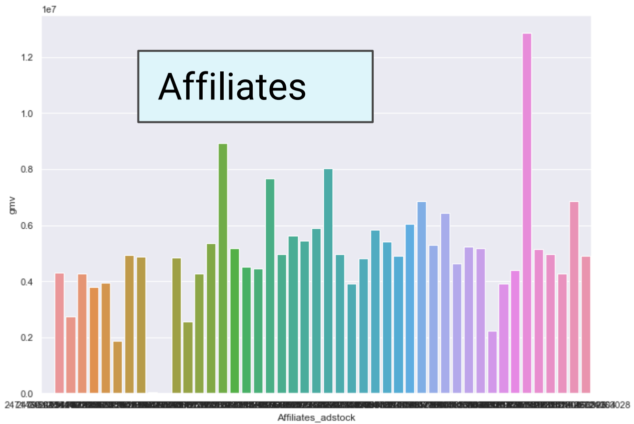


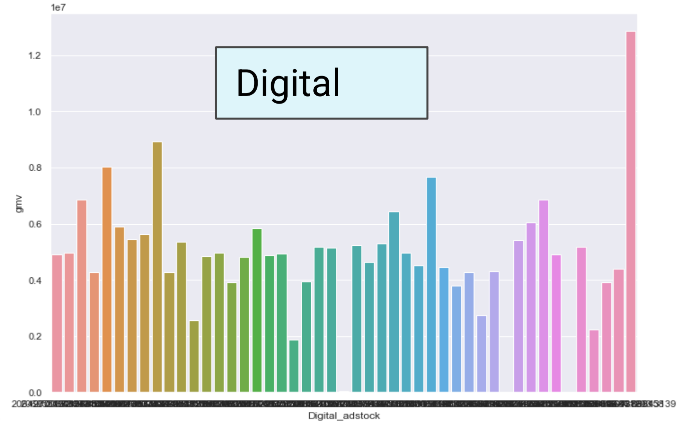


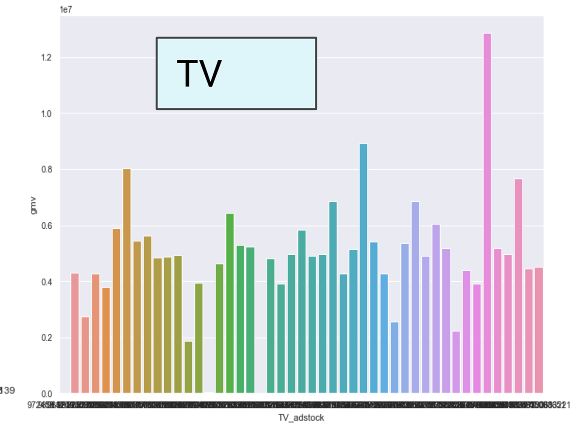




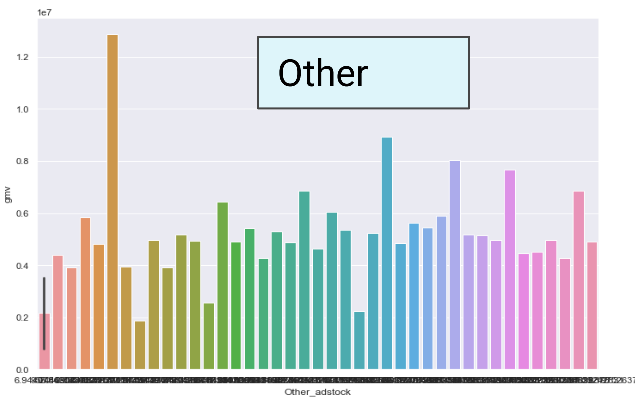
Adstock

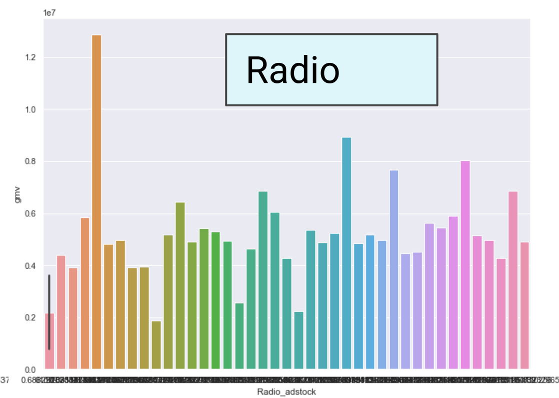

Affiliates
gmv
Affiliates_adstock

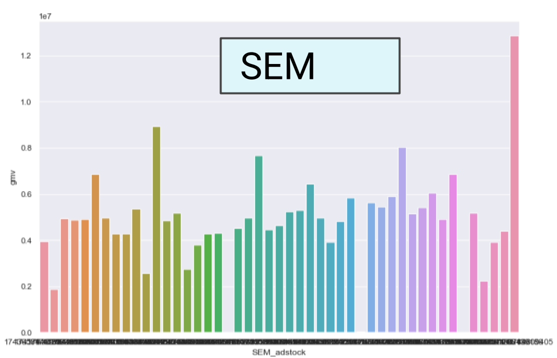

Digital
gmv
Digital_adstock

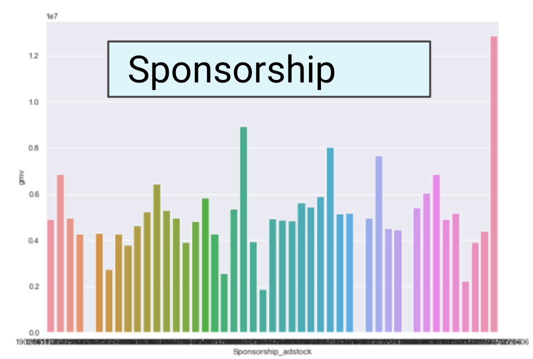

TV
gmv
TV_adstock


Online Marketing
gmv
Online Marketing_adstock


Other
gmv
Other_adstock


Radio
gmv
Radio_adstock


SEM
gmv
SEM_adstock


Sponsorship
gmv
Sponsorship_adstock

Derived KPIs & Modeling

List of derived KPIs and advance KPIs is as follows:

KPIs	Advance KPIs
• Discount Percentage	• Ad-stock of 3 categories
• GMV per unit	• Moving average of last 3 weeks (gmv per unit, DP)
• Total GMV	• Lag variables (gmv per unit, DP) for 3 weeks
• Average GMV	• Promotion Type
• Units	• Holiday Week
• Delivery Status	• Delivery Status
• Item Type	
• Delivery on Time	

Model Dashboard

r-square prediction values (without cross validation)

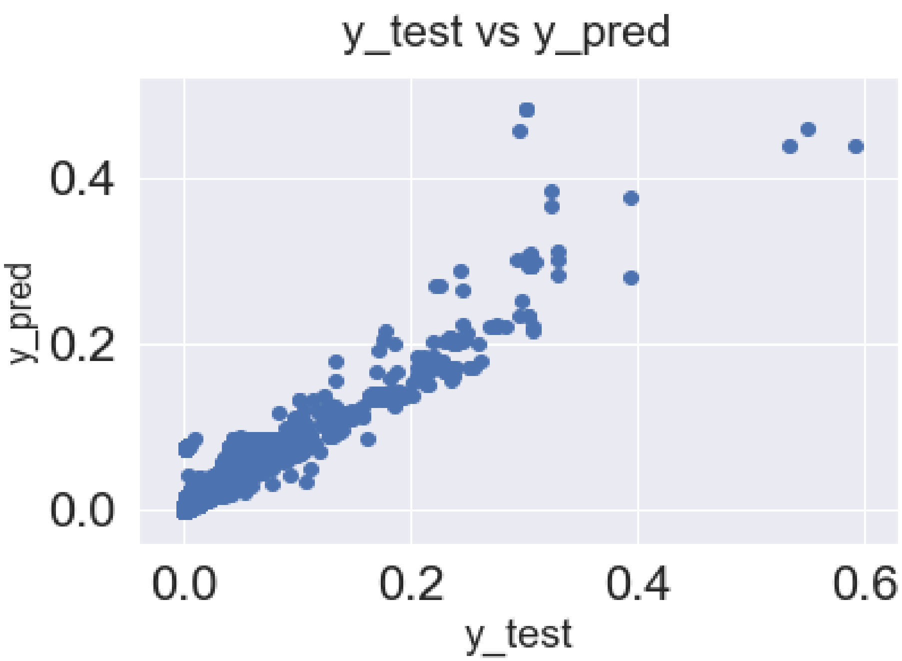
Model/Category	Camera Accessory	Gaming Accessory	Home Audio
Linear	0.89	0.86	0.78
Multiplicative	0.92	0.70	0.86
Koyck	0.86	0.87	0.75
Distributed Lag	0.86	0.86	0.76

r-square prediction values (with cross validation)

Model/Category	Camera Accessory	Gaming Accessory	Home Audio
Linear	0.91	0.89	0.80
Multiplicative	0.88	0.72	0.92
Koyck	0.91	0.89	0.84
Distributed Lag	0.91	0.89	0.84

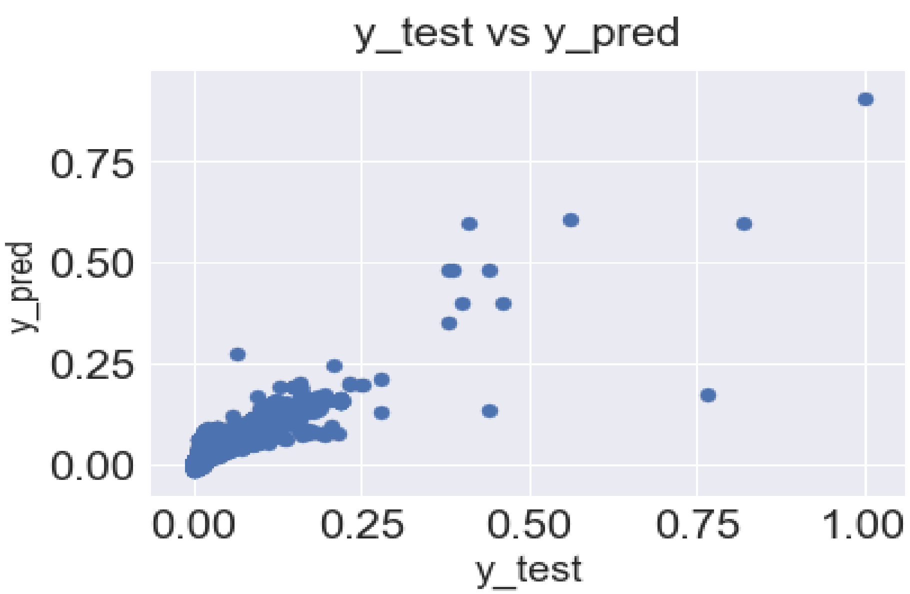
Camera Accessories – Recommendations

- After analyzing 4 different models, its observed that **Koyck** model is best suited for **Camera Accessories**.
- Its has least mean square error (0.007) and have high result in Cross - Validation (0.91) as compare to other three models.
- **product_mrp** is the strongest variable having a good impact on the GMV.
- Other common factor which affect the model is delivery_on_time, Content Marketing etc.



Gaming Accessories – Recommendations

- After analyzing 3 different models, its observed that **Koyck** model is best suited for **Gaming Accessories**.
- For Gaming Accessories, RMSE is 0.0071 for **Koyck** which is the least mean square error as compared to others.
- Also **product_mrp** is the strongest variable having a good impact on the GMV.
- Other common factor which affect the model is delivery_on_time, Content Marketing etc



Home Audio – Recommendations

- After building and analyzing 3 different models, it is observed that **Multiplicative model** is best suited for the category of **Home Audio.**
- R2 scores of **Koyck and Distributed lag models** are almost the same but better results are achieved after **performing cross validation**.
- **product_mrp** is the strongest variable having a good impact on the GMV.
- **gmv_lag_1_per** and **gmv_lag_2_per** are also proving to be good for GMV.
- **GMV** increases with a little bit of tweaking in the **product_mrp.**