Capstone Project

ElecKart Market Mix Modelling

by

Sohini Chaudhuri

Business Analysis

ElecKart is an e-commerce firm specialising in electronic products. Over the last one year, they had spent a significant amount of money in marketing. They also offered big-ticket promotions.

They are about to create a marketing budget for the next year which includes spending on commercials, online campaigns, and pricing & promotion strategies.

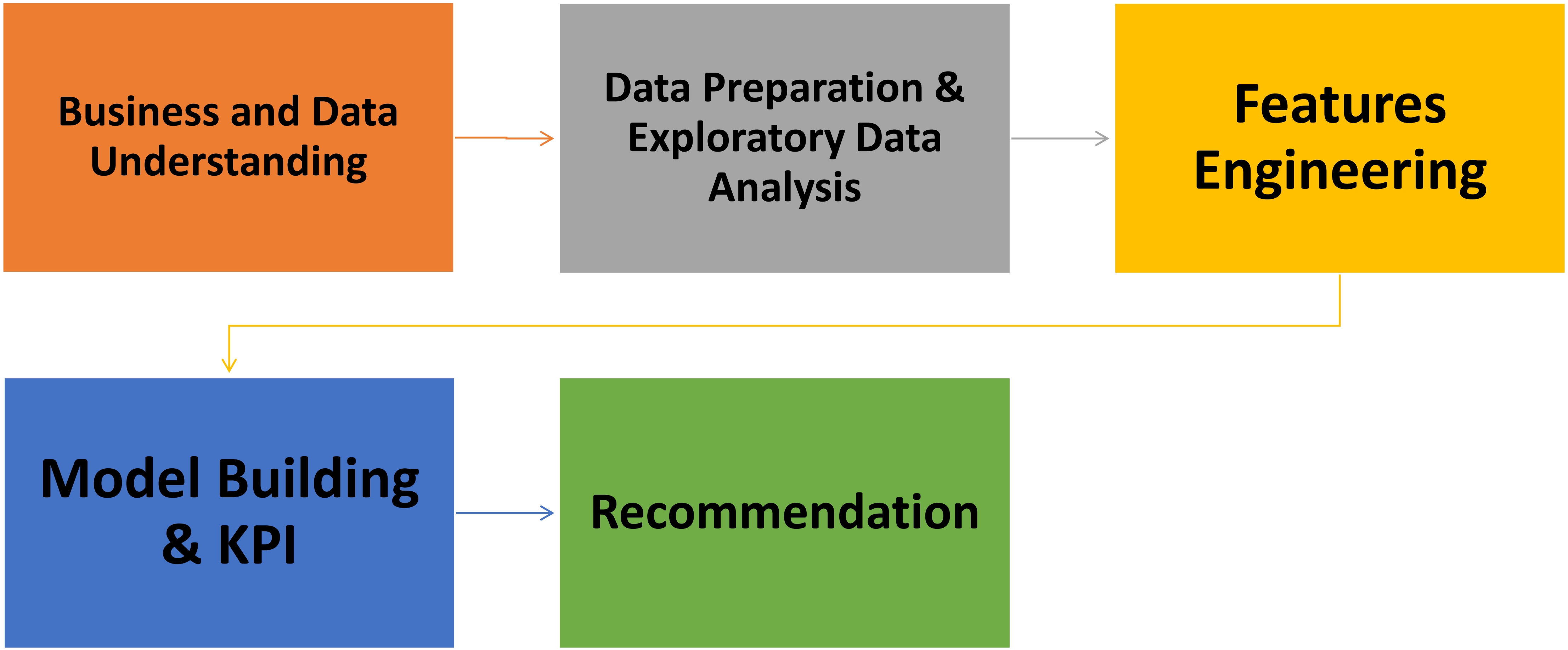
Objective

The aim is to develop a market mix model to observe the actual impact of different marketing variables over the last year.

Basically needs to optimize the marketing levers to improve the revenue response.

The data or variables needs to be consider for analysis :Products Sales data, Media Investment, NPS Score, Special Sale days [Holidays]

Process done on the Data Set



Processing

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 last 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

1
2

- Gather E-commerce Domain knowledge
- Understand the given dataset and the questions that can be answered

3
4

- Cleaning and preparing dataset
- Creating 3 categorical dataset namely “CameraAccessory”, “GamingAccessory”, “HomeAudio”

5
6

- Derived variable creation and grouping by weeks
- Exploratory Data Analysis

7
8

- Uni and Bi variate Analysis
- Multi variate Analysis

9
10

- Using different modelling techniques on the three dataset created
- Choose gthe best model for each segment

Basic understanding of given data

product

- Delivery Days and SLA, Units Sold
- Categories/Sub-categories
- Vertical
- Procurement SLA
- Item Type = LUXURY/Mass-Market

promotion

- Marketing Channel Investment
- Customer Sentiments
- Adstock
- Discounts

price

- Gmv
- Product mrp

place

- Pin-code
- Order Payment type
- Week of the year/Seasonality
- Holixday/Events;isHoliday

Data Cleaning and Preparation of consumer Electronic

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 –

- Splitting the event name and date; also converting the date in proper format.
- Creating event start and end date.
- Producing a dataframe – having all dates possible within the timeframe of analysis and corresponding event names (if any).
- Merging this with consumer data – produced earlier

Data Cleaning Preparation of Media data continued….

3. Monthly NPS Score –

- Transposing the columns into rows.
- Cleaning the naming issues wrt months.
- Populating the same monthly scores to each day of the month.
- Convert it to weekly basis.
- Merging this with consumer data – produced earlier.

4. Media Investment –

- Distributing the monthly investing data for each channel into daily investment – proportionate to “days in that month”.
- Converting it on weekly basis.
- Extracting category-wise investment – proportionately to gvm of each of the 3 category wrt the total gvm.
- Creating 3 category-wise dataframes - Merging this with consumer data – produced earlier.
- 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 .file

Local Factors

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

```
##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)
```

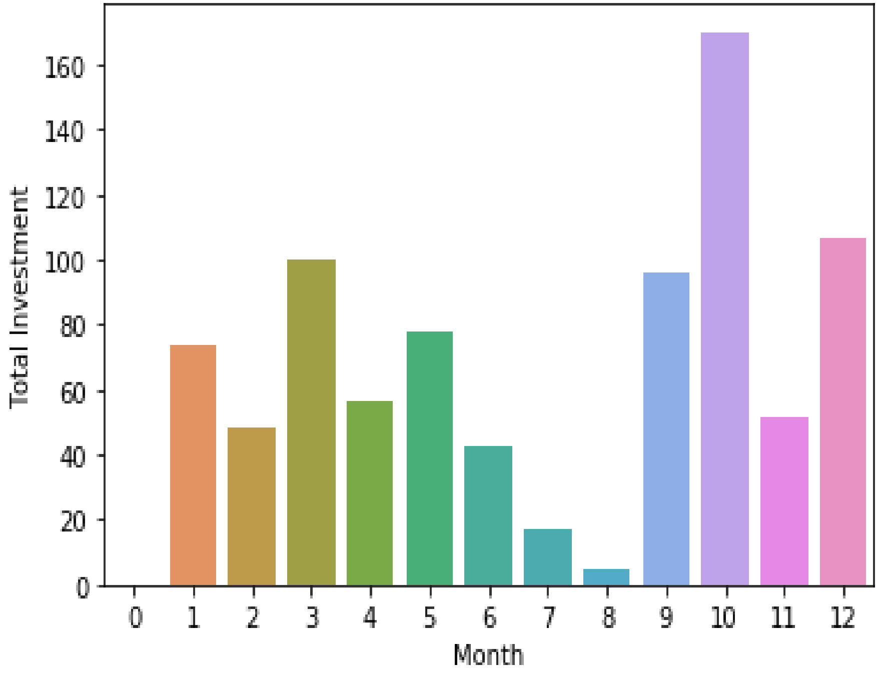
```
(datetime.date(2016, 1, 1), "New Year's Day")
(datetime.date(2016, 2, 15), 'Family Day')
(datetime.date(2016, 3, 25), 'Good Friday')
(datetime.date(2016, 5, 23), 'Victoria Day')
(datetime.date(2016, 7, 1), 'Canada Day')
(datetime.date(2016, 8, 1), 'Civic Holiday')
(datetime.date(2016, 9, 5), 'Labour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

```
#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)
```

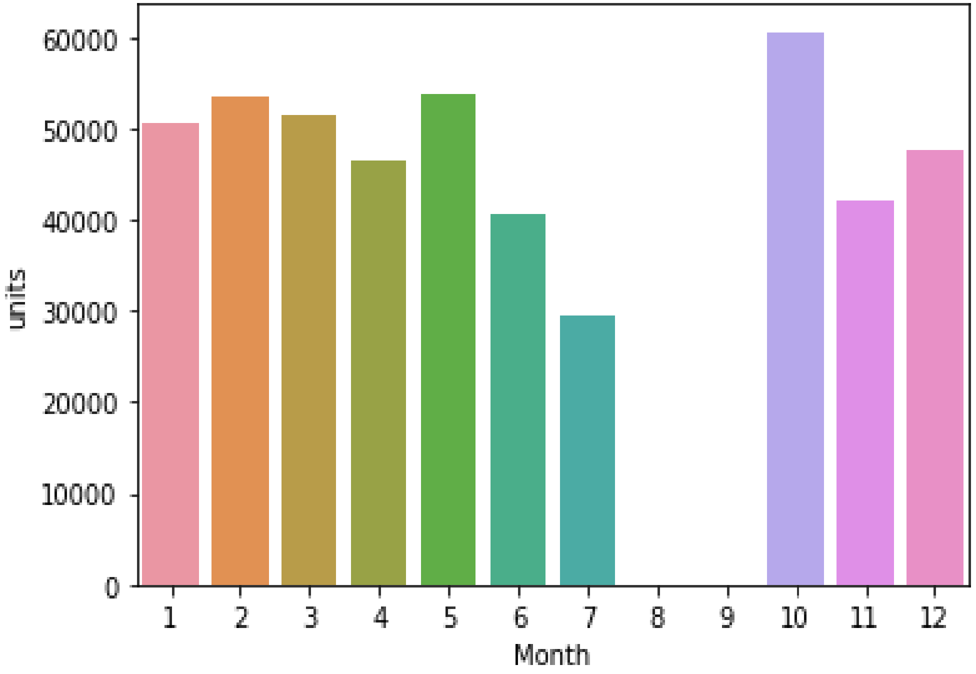
```
(datetime.date(2015, 1, 1), "New Year's Day")
(datetime.date(2015, 2, 16), 'Family Day')
(datetime.date(2015, 4, 3), 'Good Friday')
(datetime.date(2015, 5, 18), 'Victoria Day')
(datetime.date(2015, 7, 1), 'Canada Day')
(datetime.date(2015, 8, 3), 'Civic Holiday')
(datetime.date(2015, 9, 7), 'Labour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

Monthly Total Investment

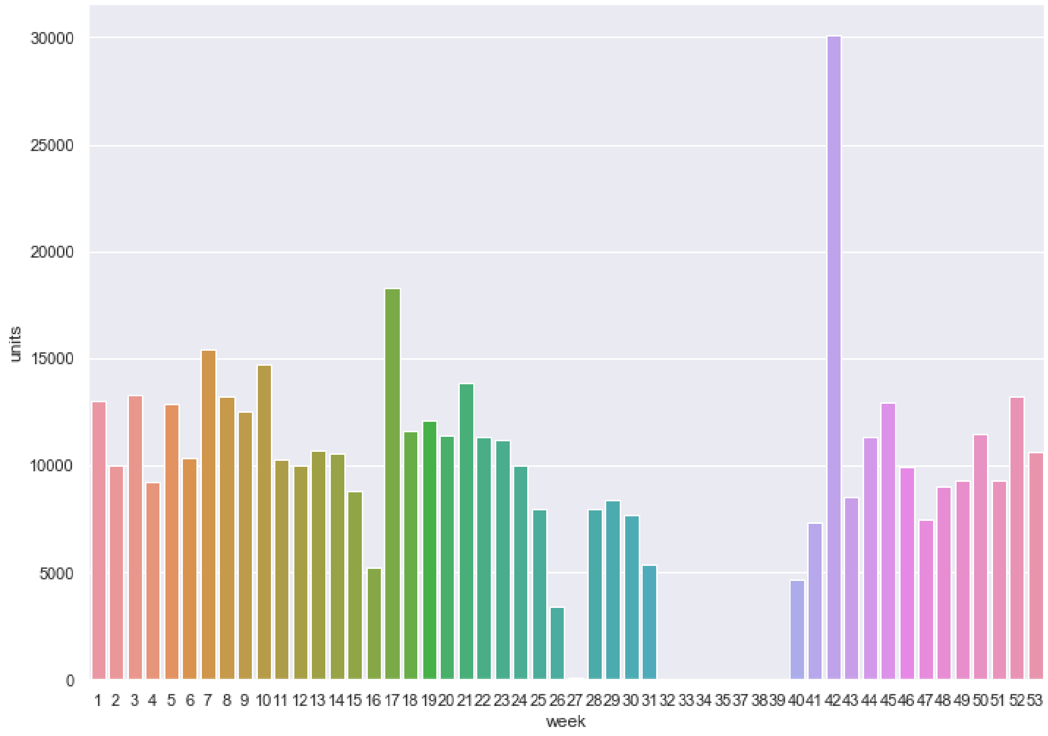


Monthly Units sold

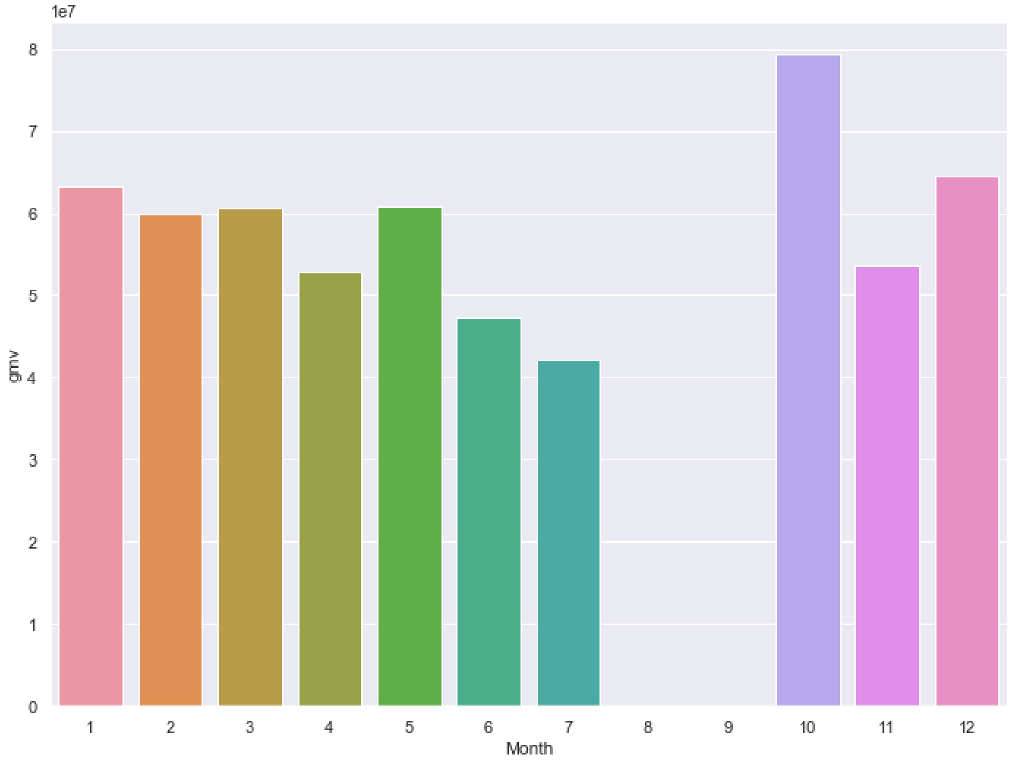


Bar Charts and Graphs of Master Dataframe

Weekly Unit Sold

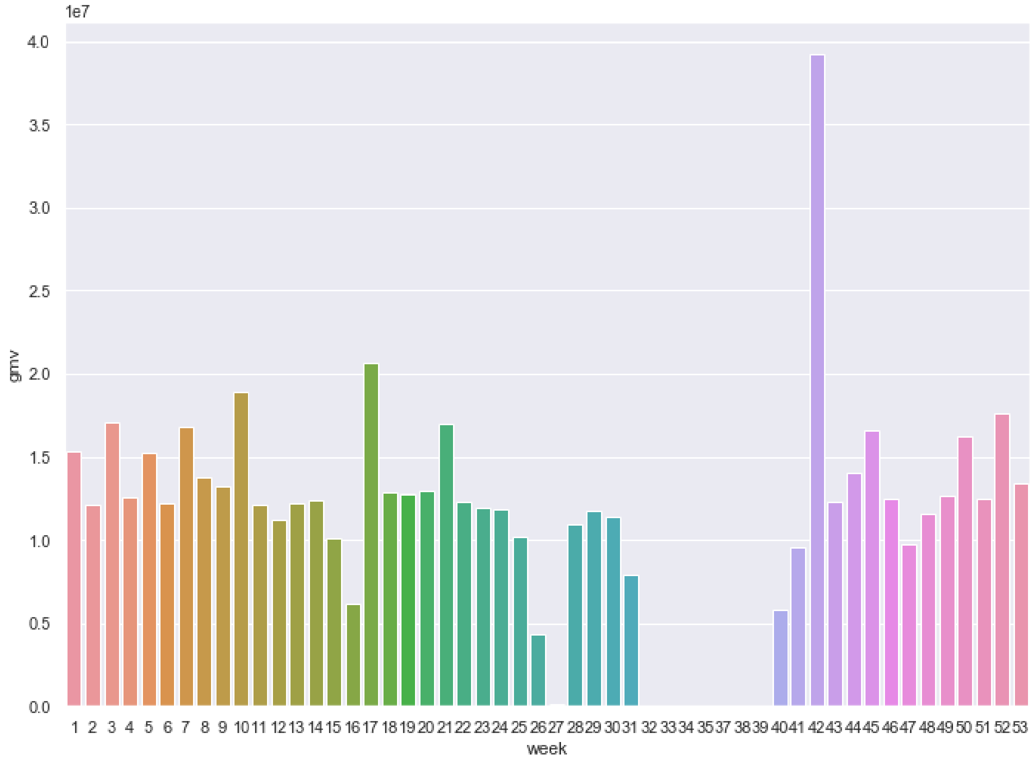


Monthly gmv

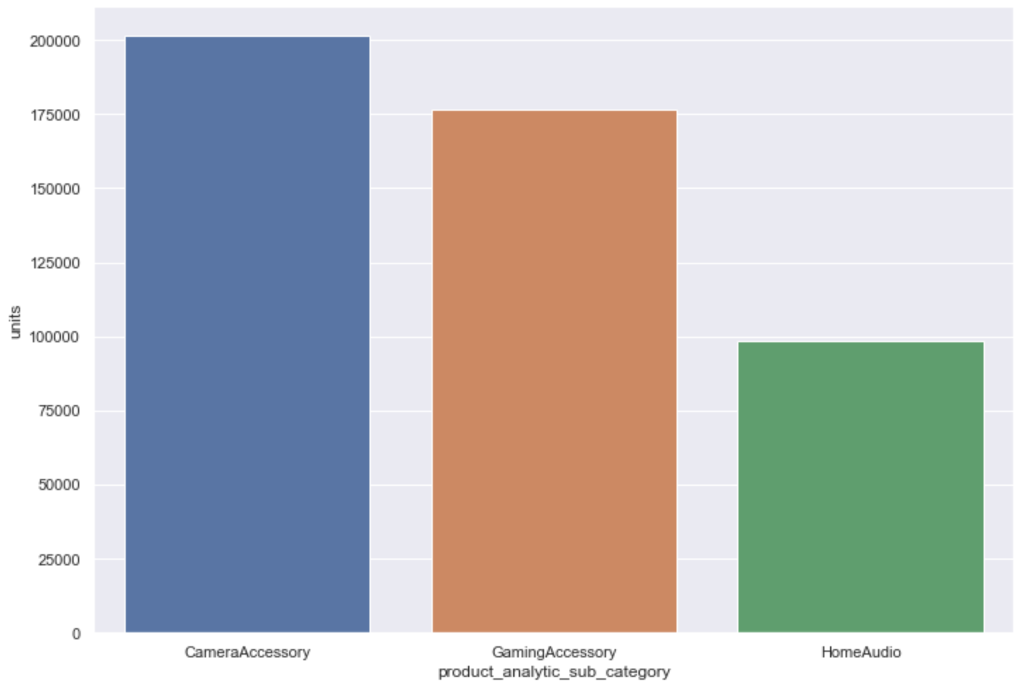


Bar Charts and Graphs of the Master Dataframe

Weekly gmv

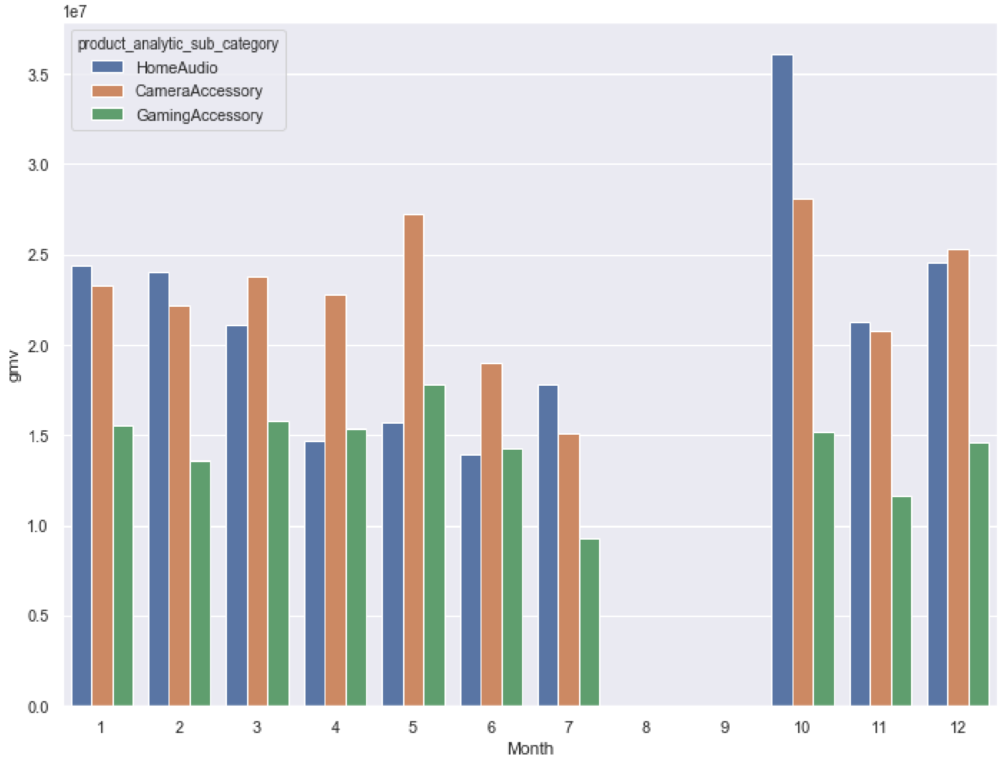


Product units sold

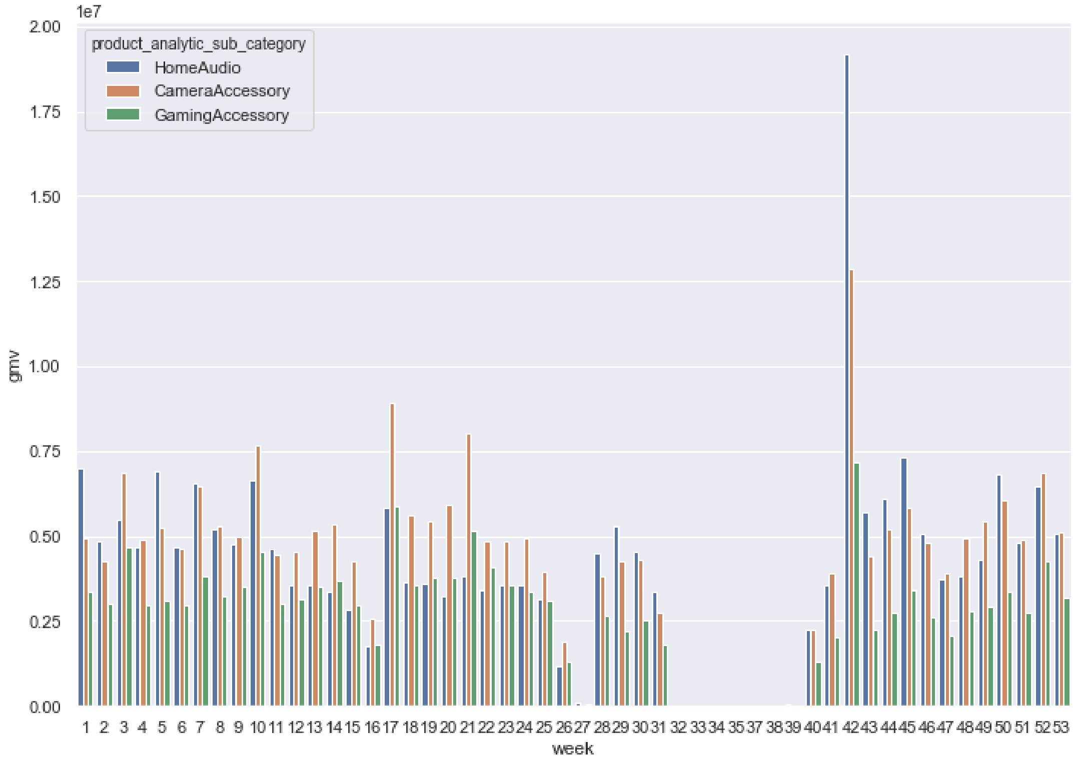


Bar Charts and Graphs of the Master Dataframe

Monthly Product gmv

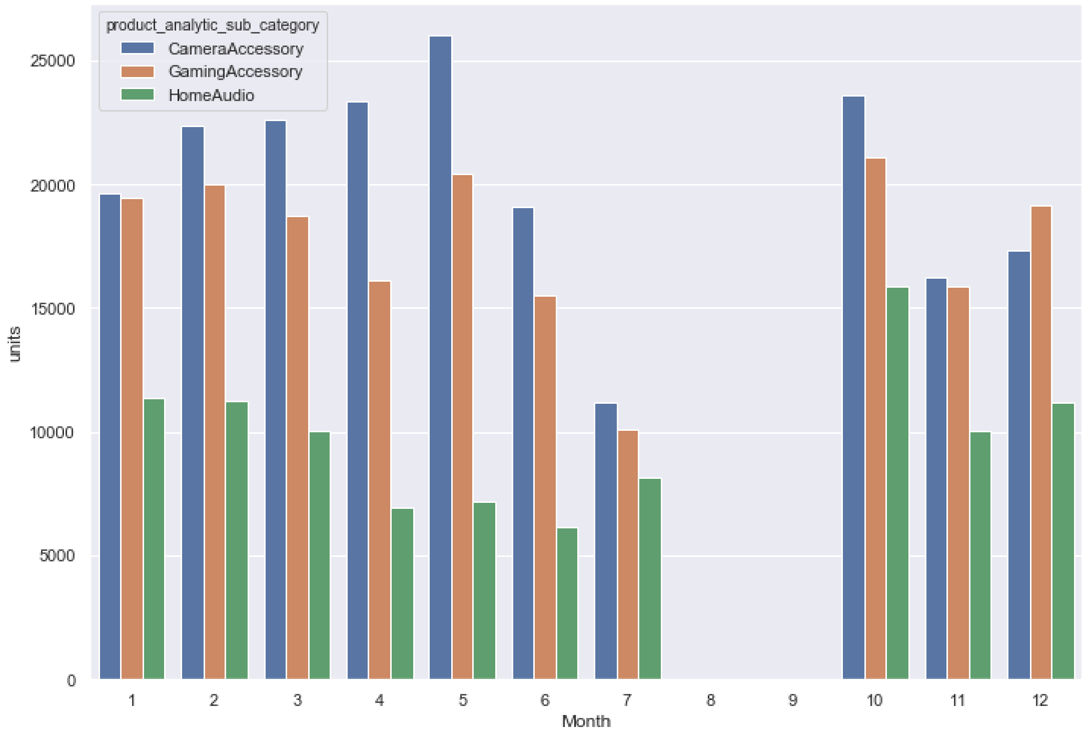


Weekly Product gmv

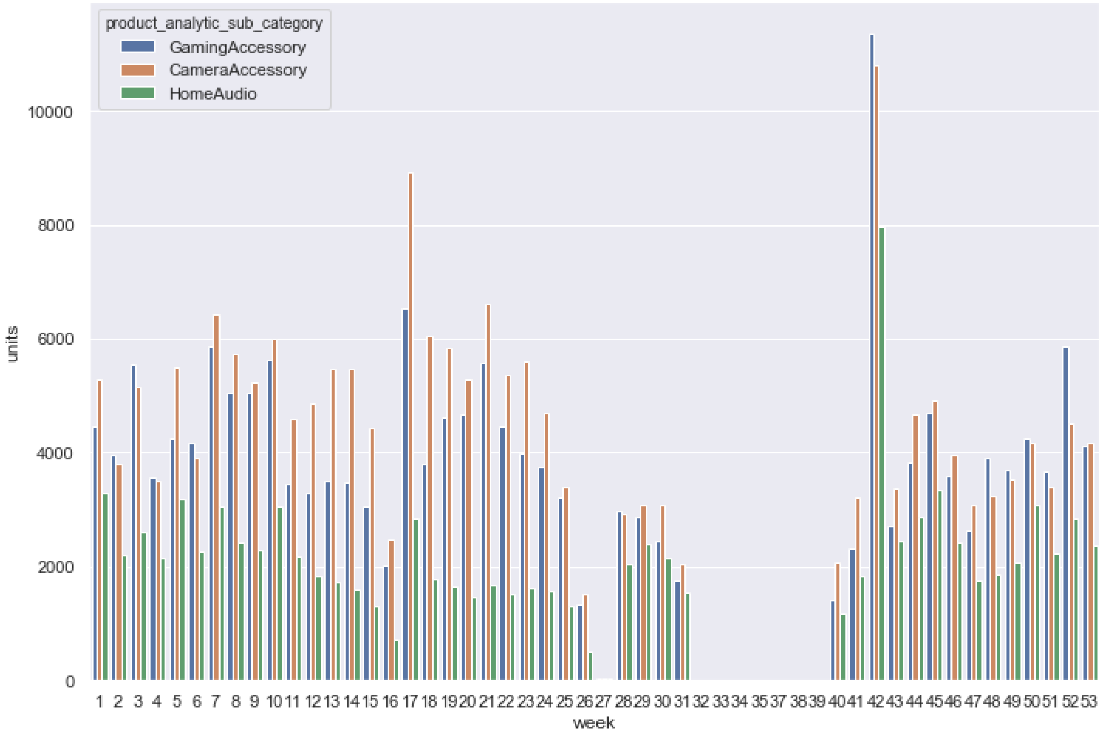


Bar Charts and Graphs of the master dataframe

Monthly Product Unit Sold

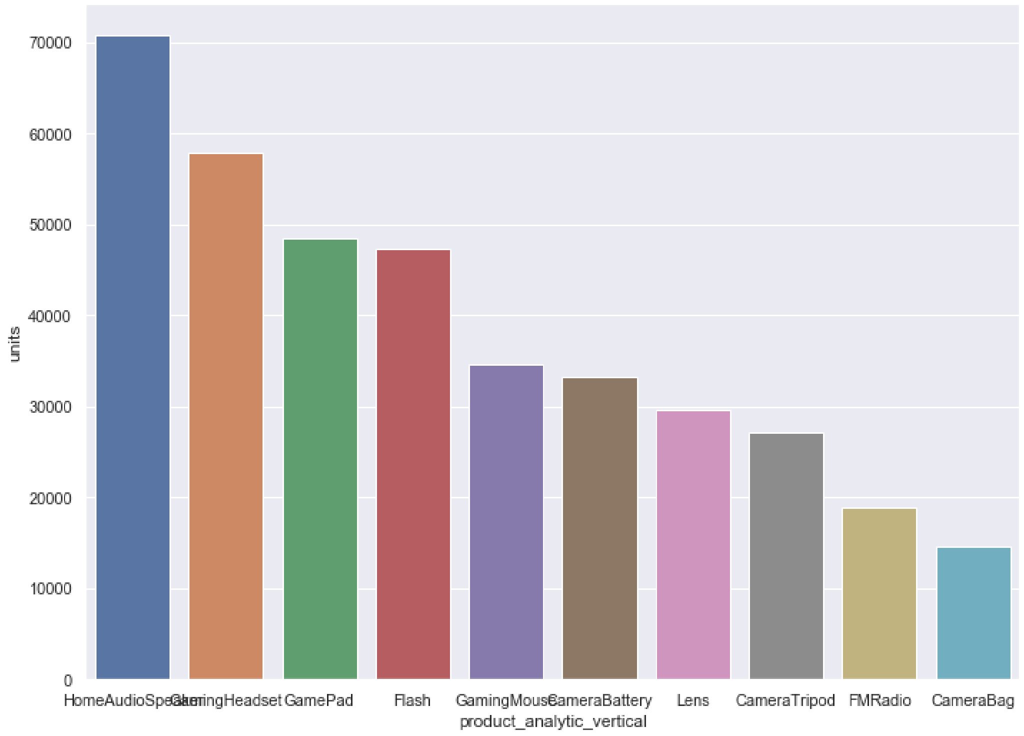


Weekly Product Data Sold

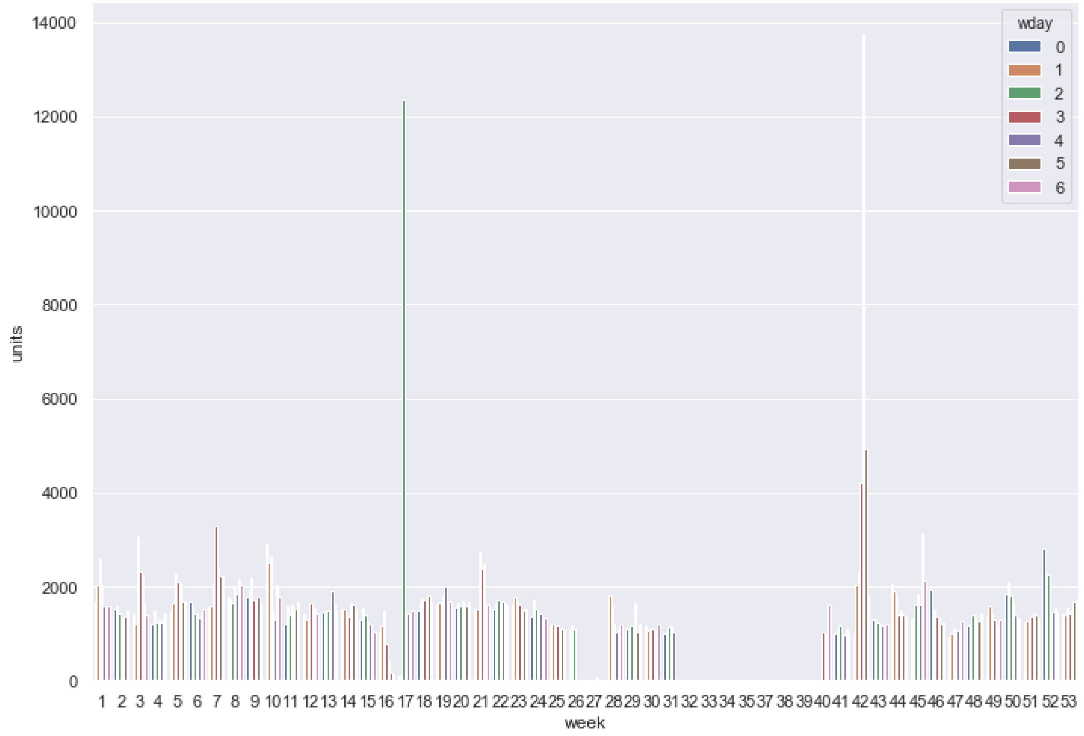


Bar Charts and Graphs of the Master Dataframe

Product Vertical Unit Sold

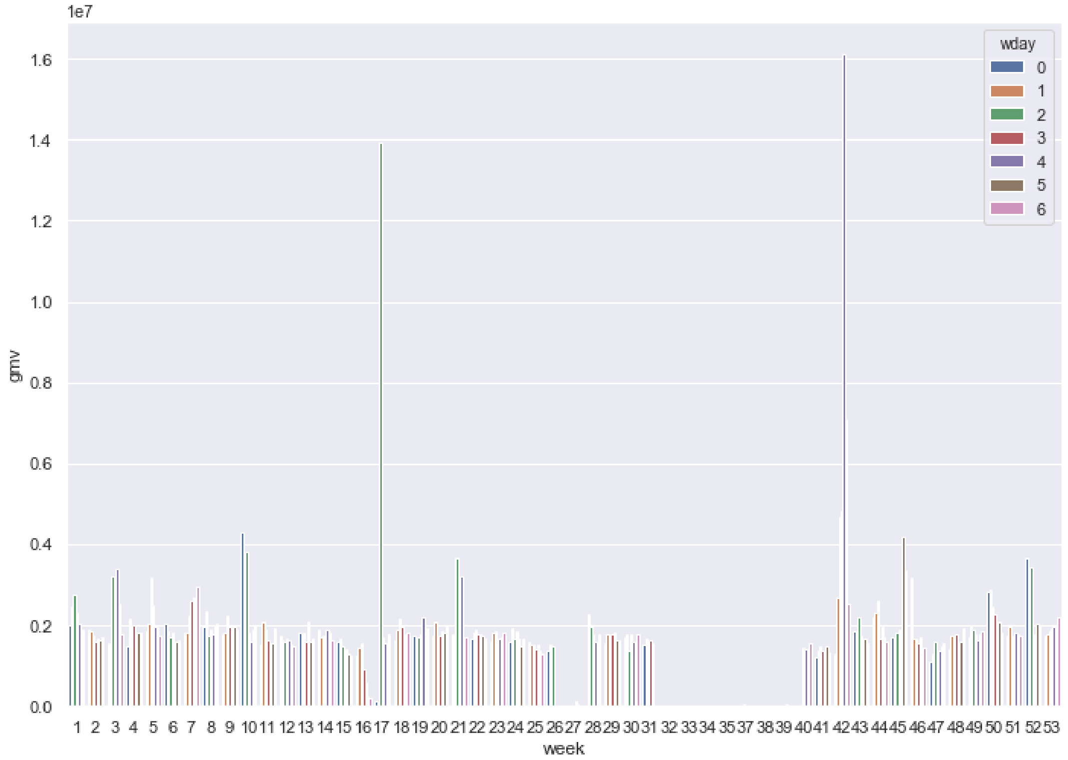


Weekday Unit Sold

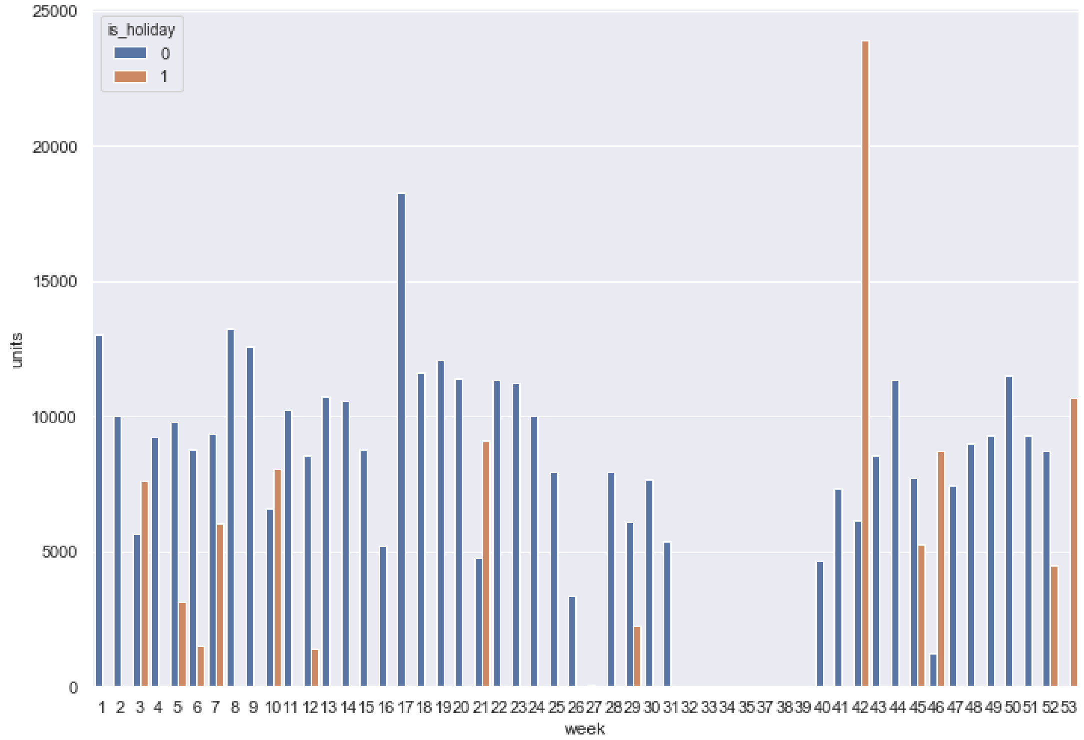


Bar Charts and Graphs of the Master Dataframe

Weekday gmv

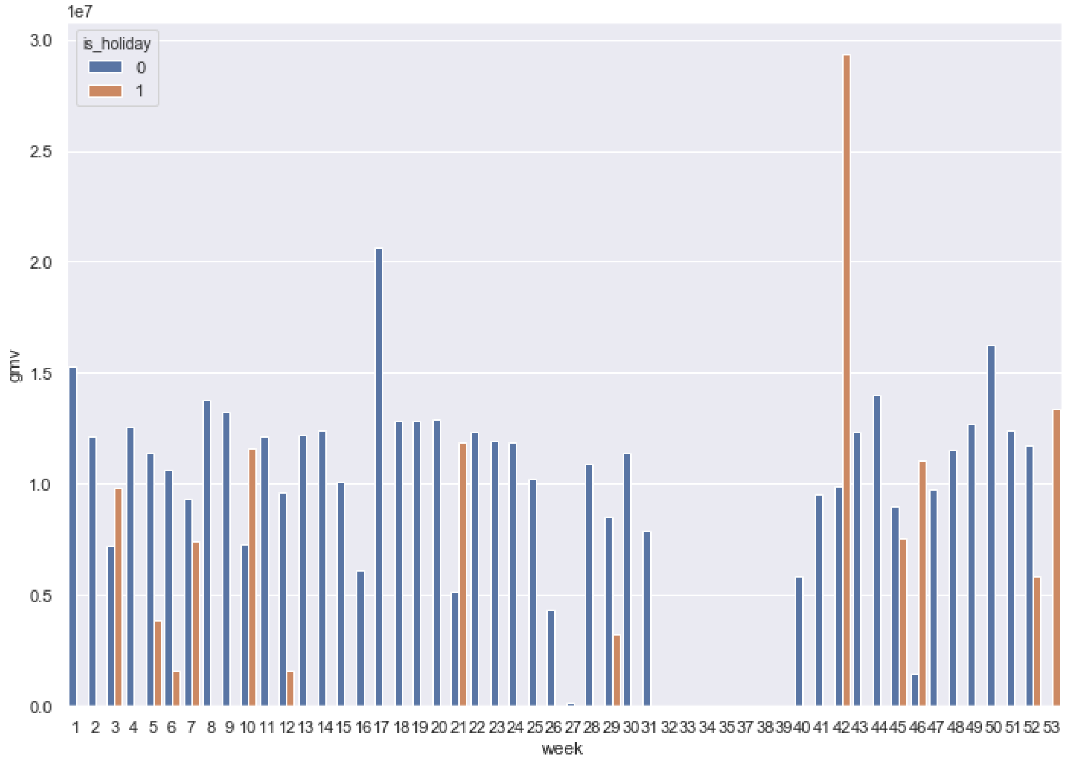


Holiday Unit Sold

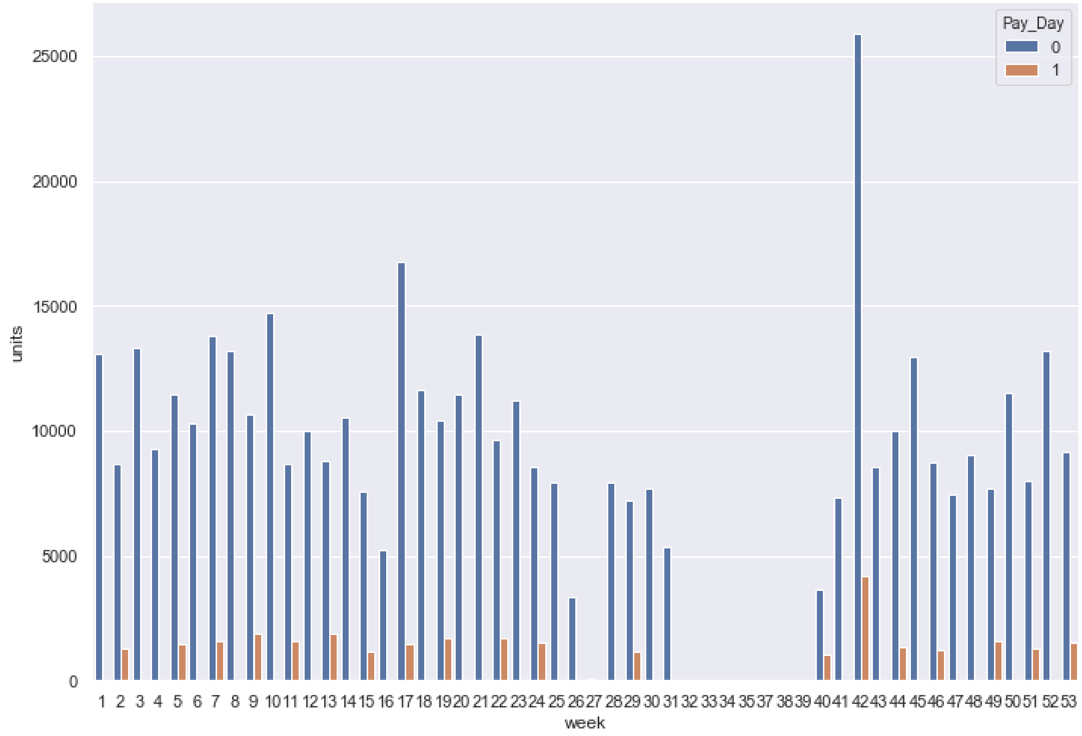


Bar Charts and Graphs of the Master Dataframe

Holiday gmv

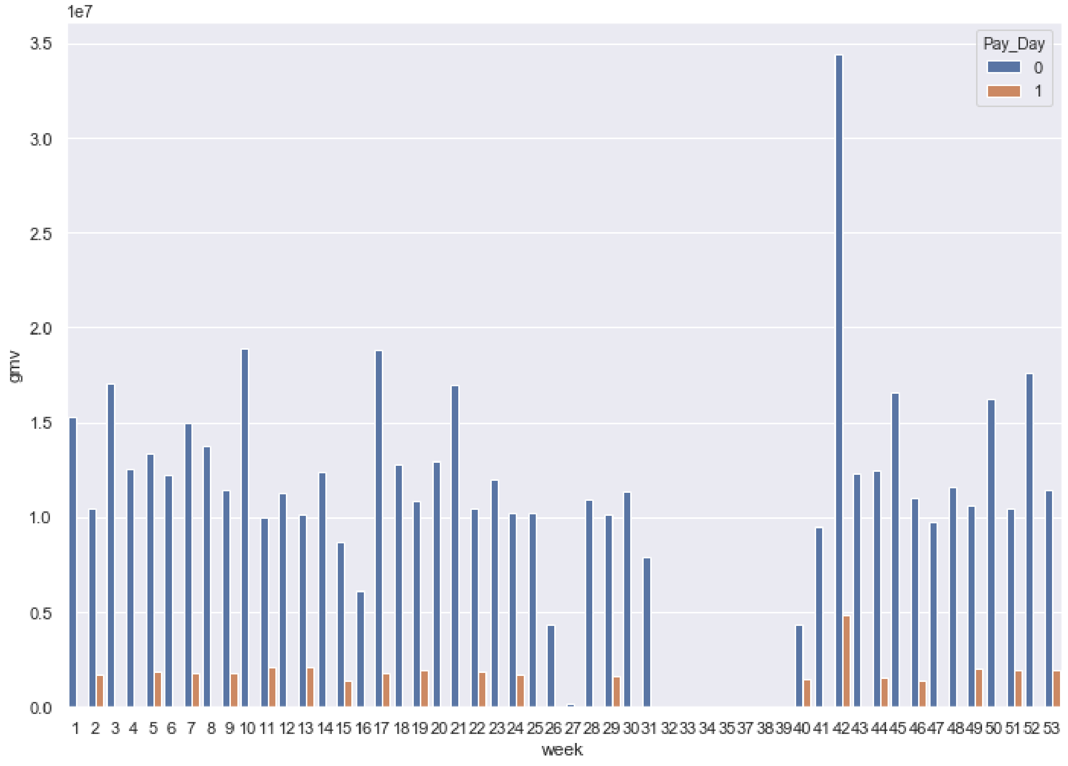


Payday Units Sold

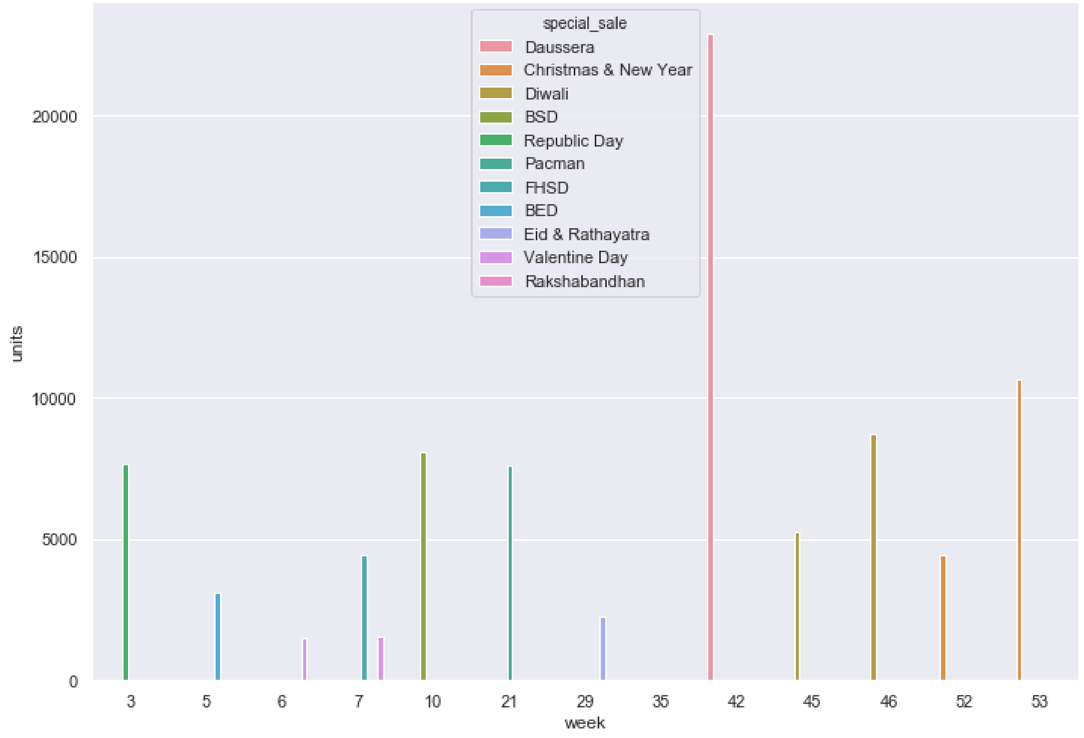


Bar Charts and Graphs of the Master Dataframe

Payday gmv

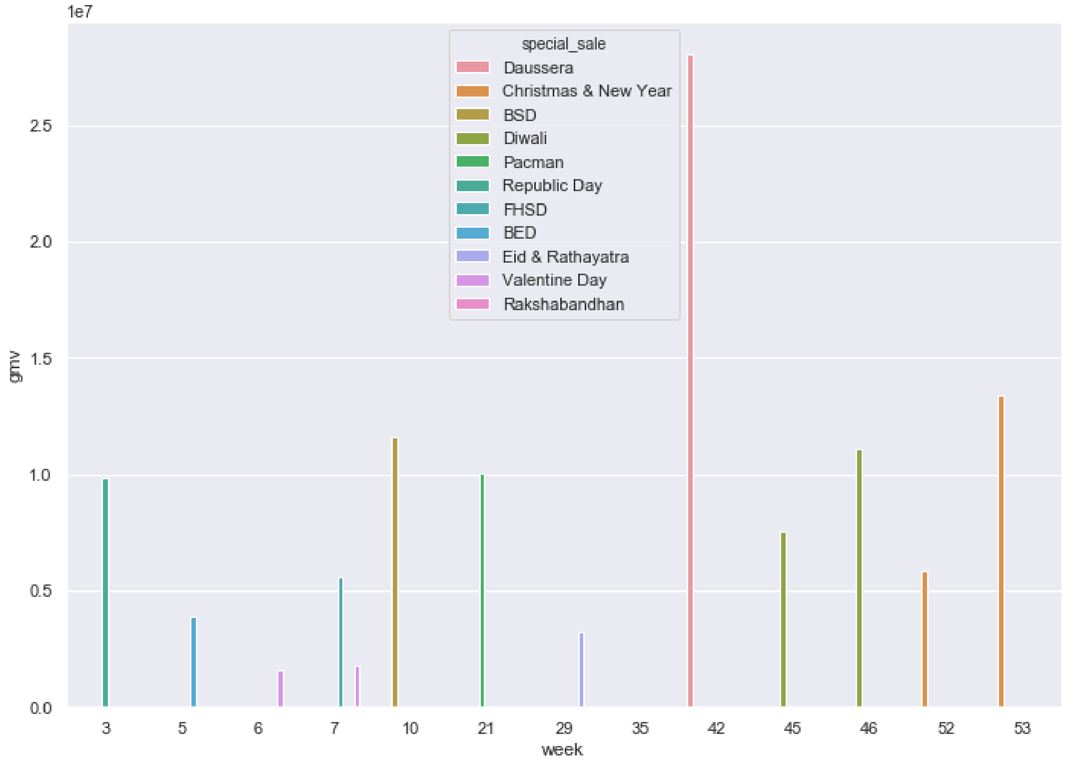


Special Sale Day Unit Sold

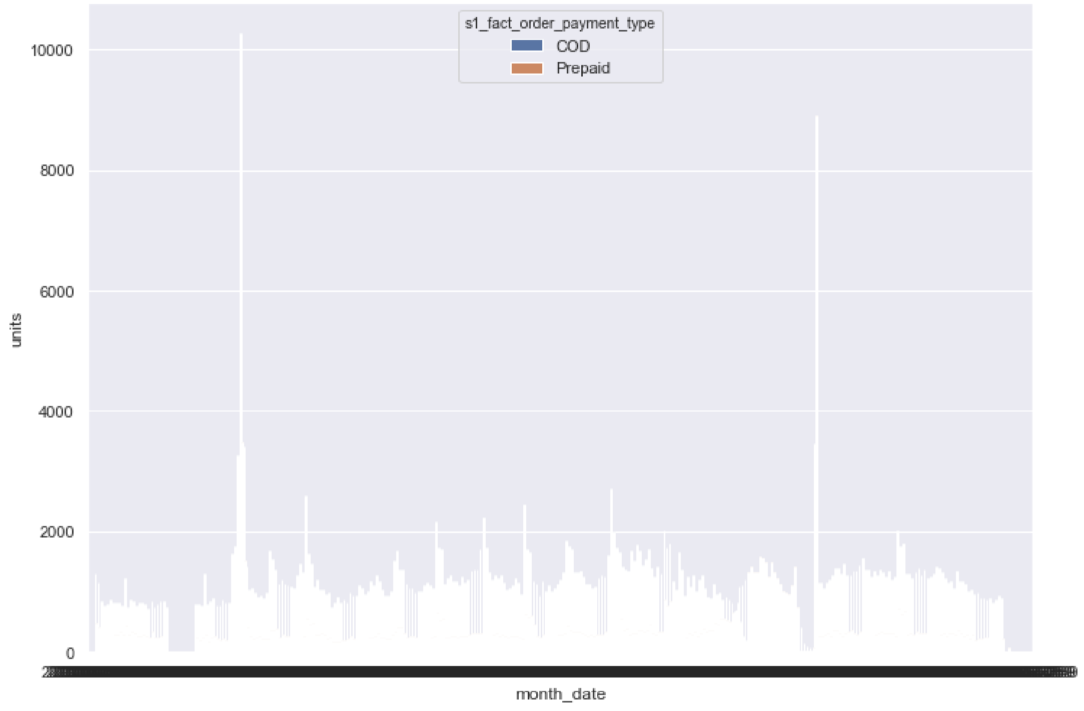


Bar Charts and Graphs of the Master Dataframe

Special Sale gmv

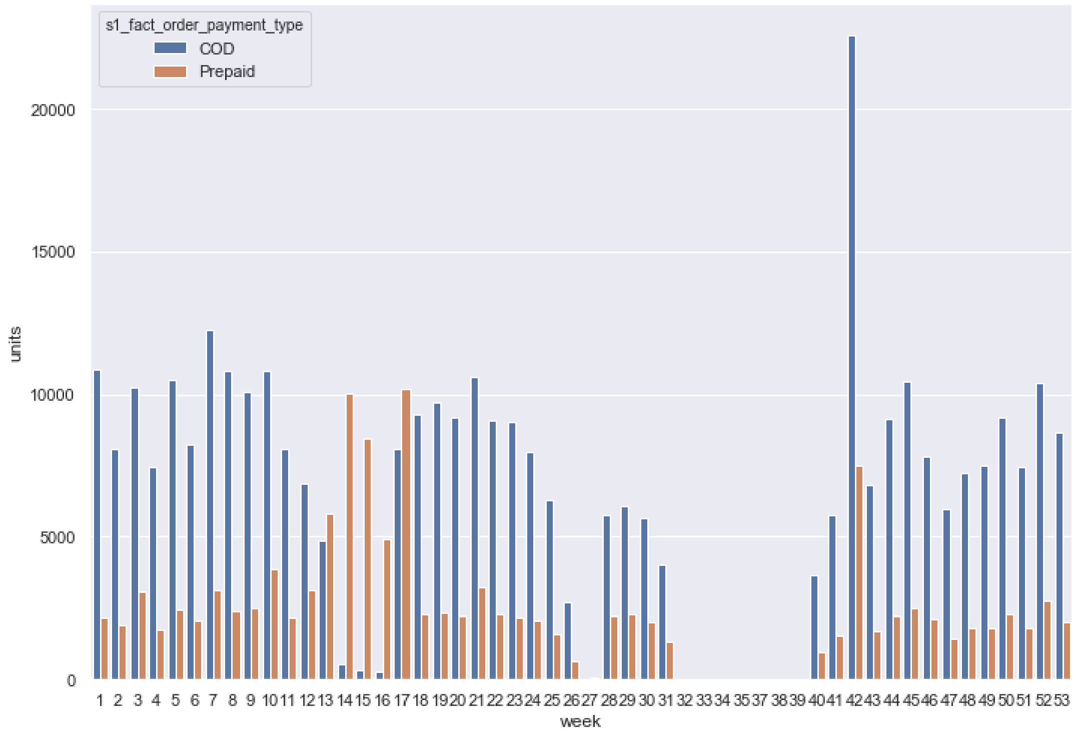


Monthly Unit Sold by Payment Type

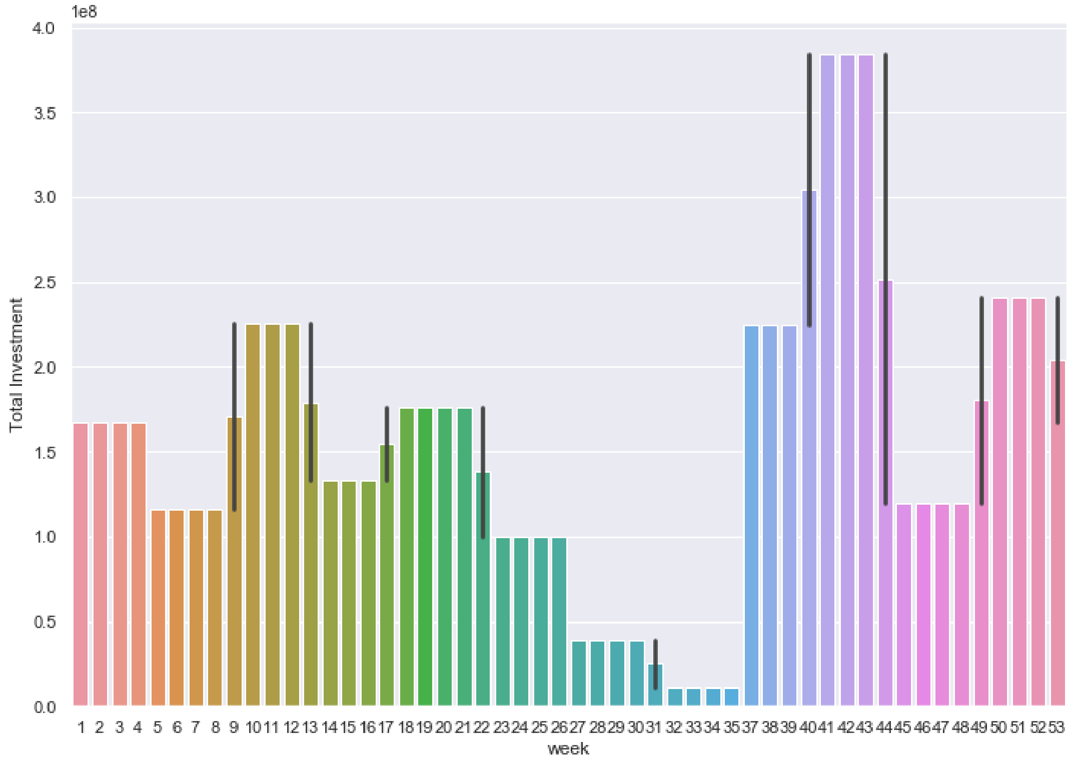


Bar Charts and Graphs of the Master Dataframe

Weekly Unit Sold by Payment Type

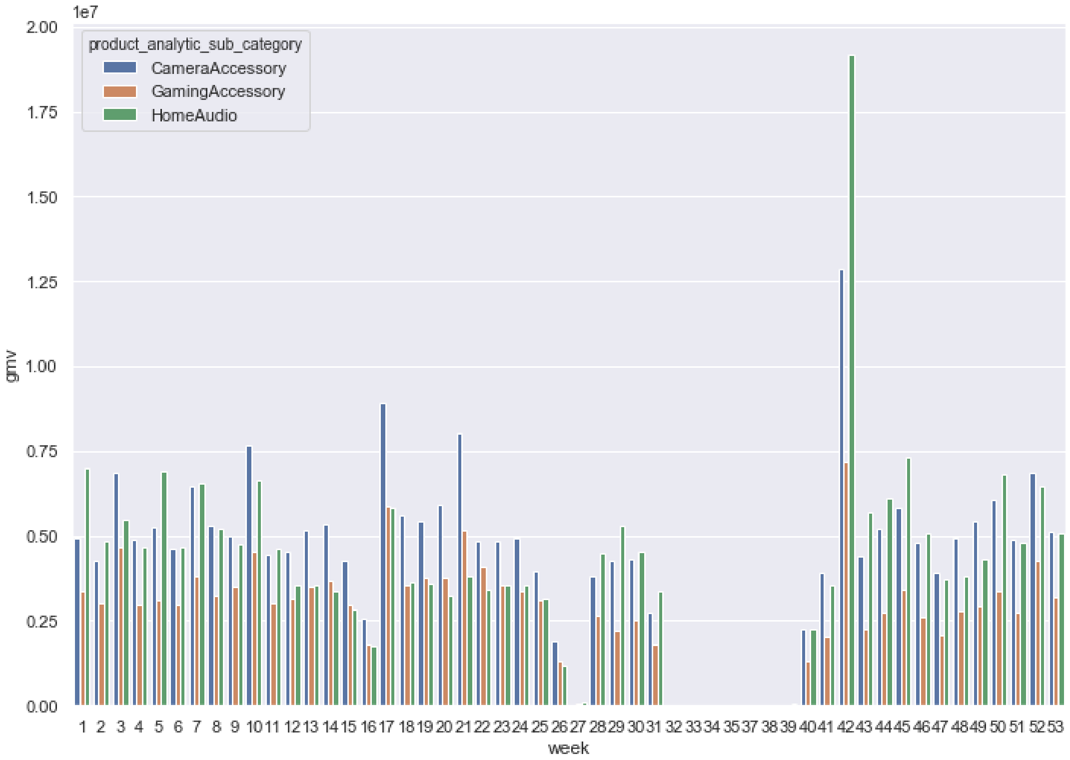


Weekly Total Investment

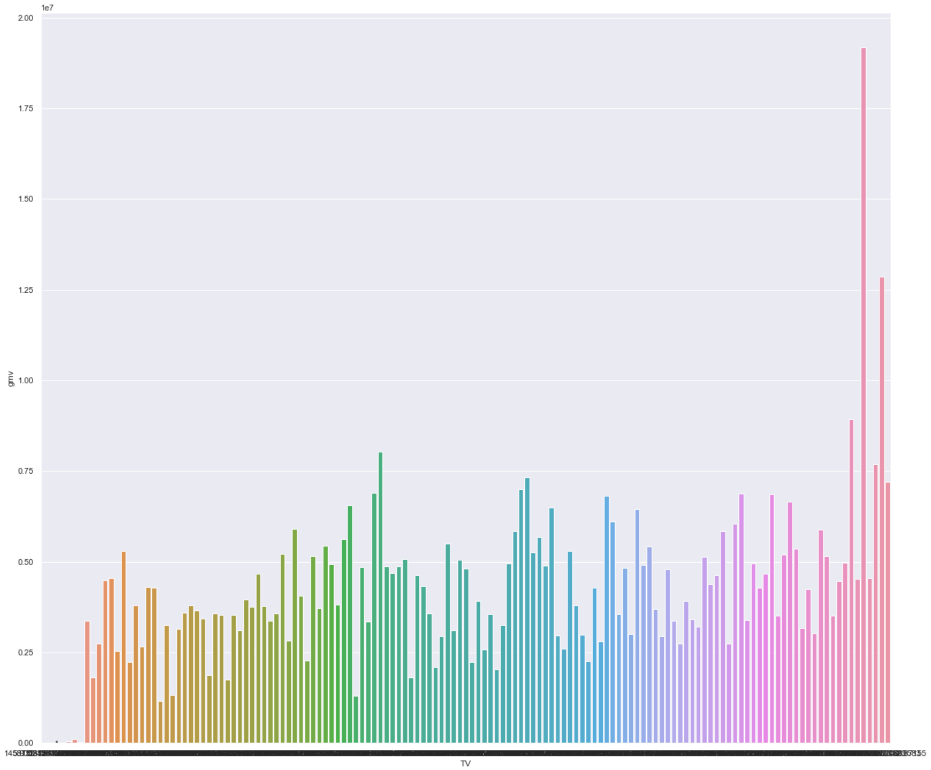


Bar Charts and Graphs of the Master Dataframe

Weekly Ad spent (Week)

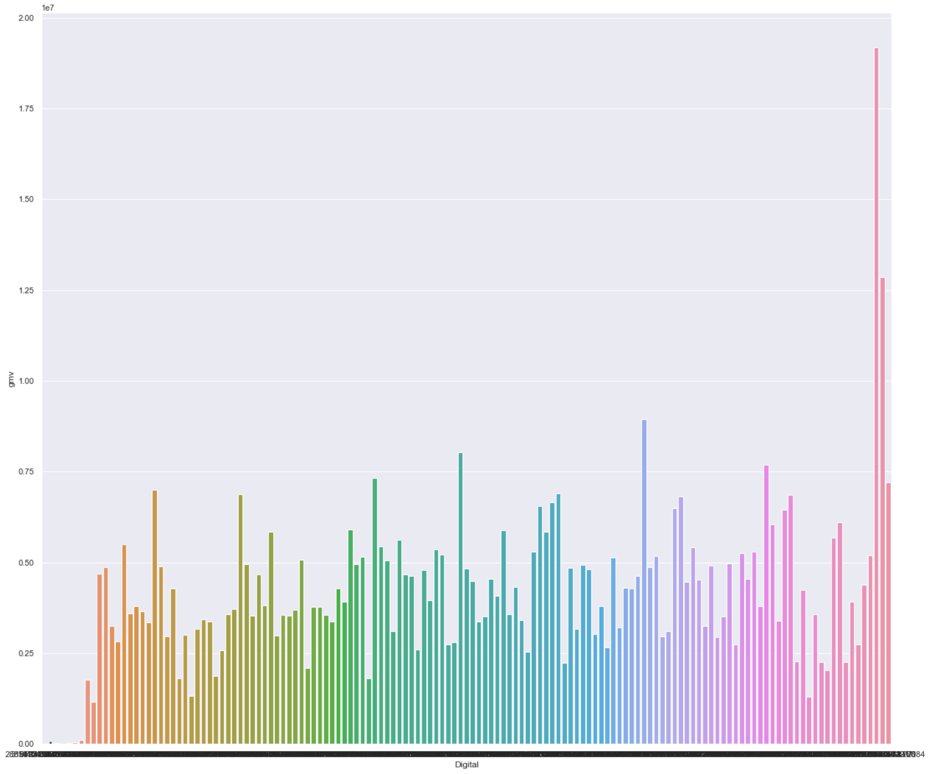


Weekly Ad spent (TV)

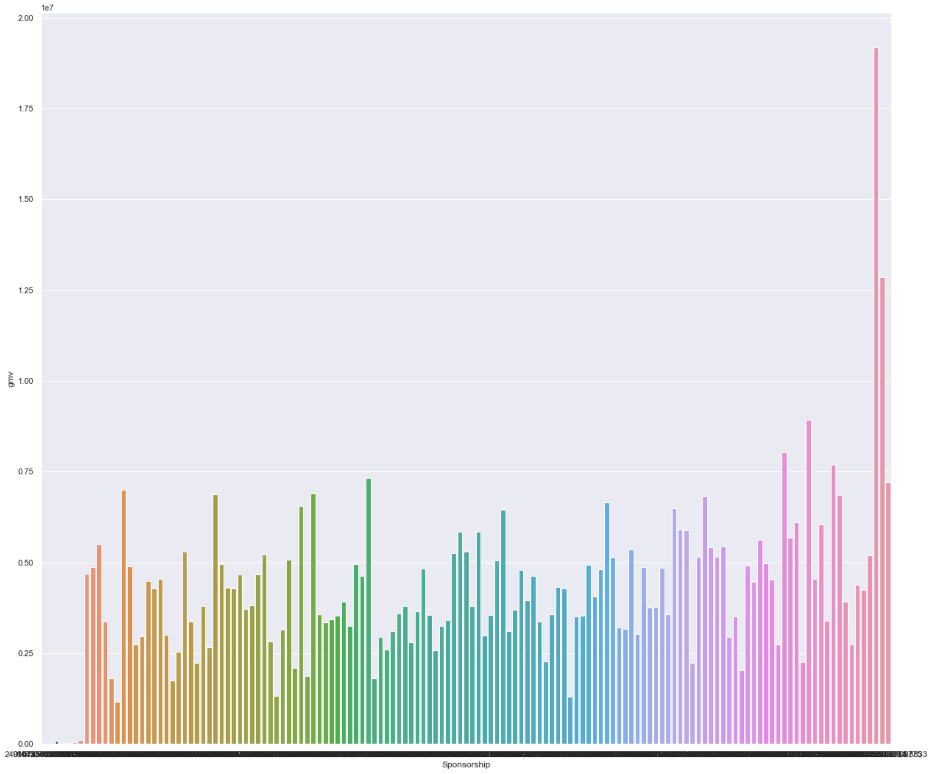


Bar Charts and Graphs of the Master Dataframe

Weekly Ad Spent (Digital)

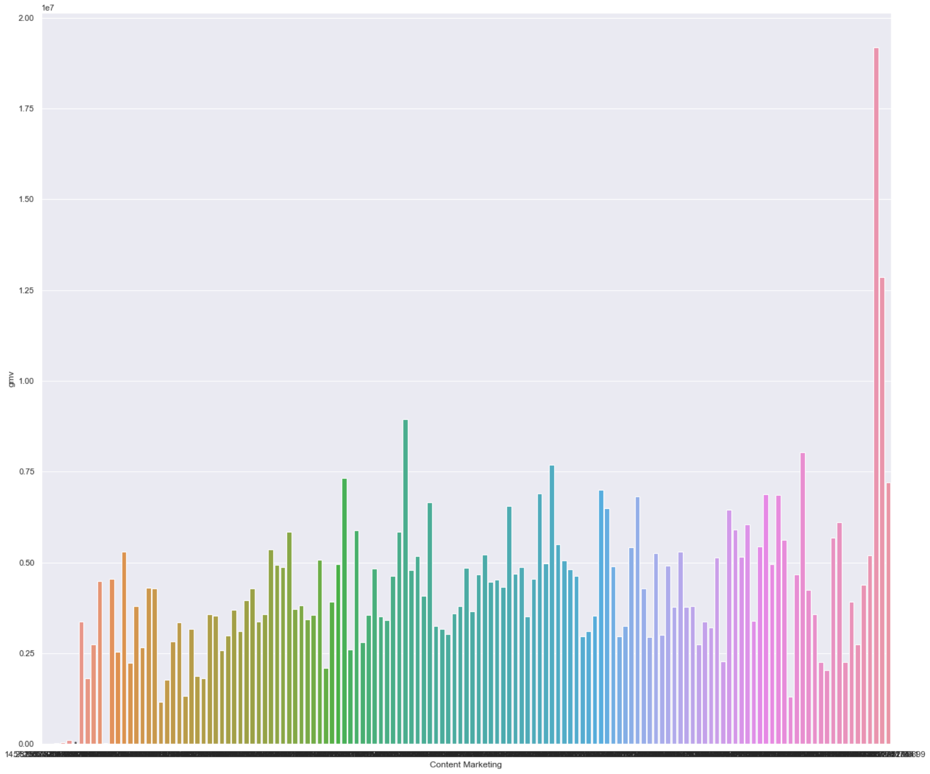


Weekly Ad Spent (Sponsorship)

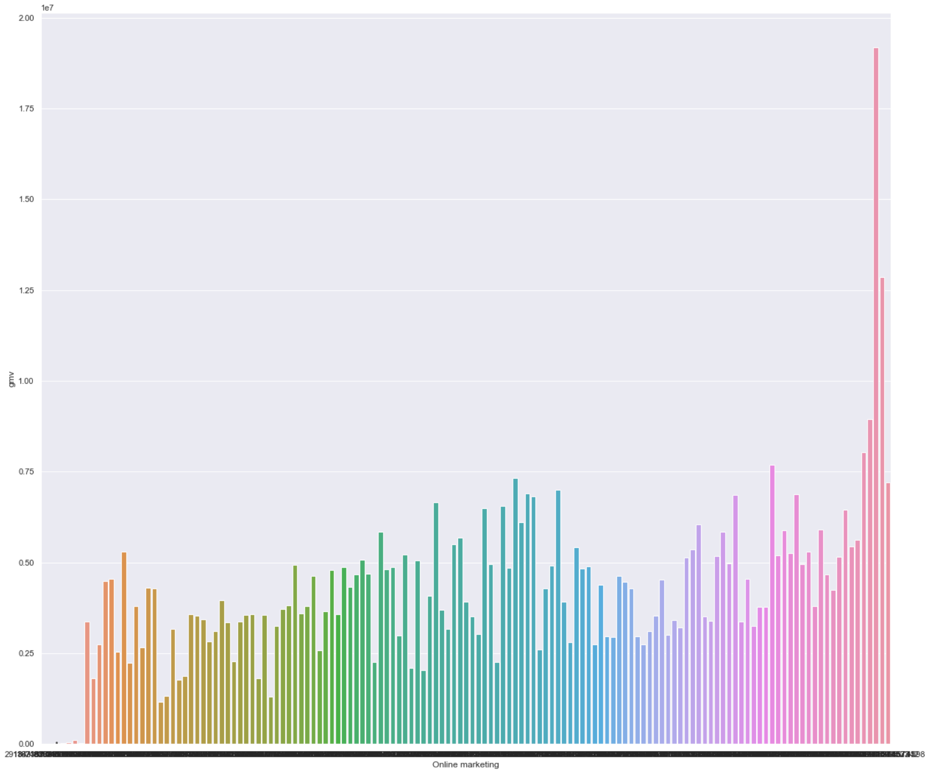


Bar Charts and Graphs of the Master Dataframe

Weekly Ad Spent (Content Marketing)

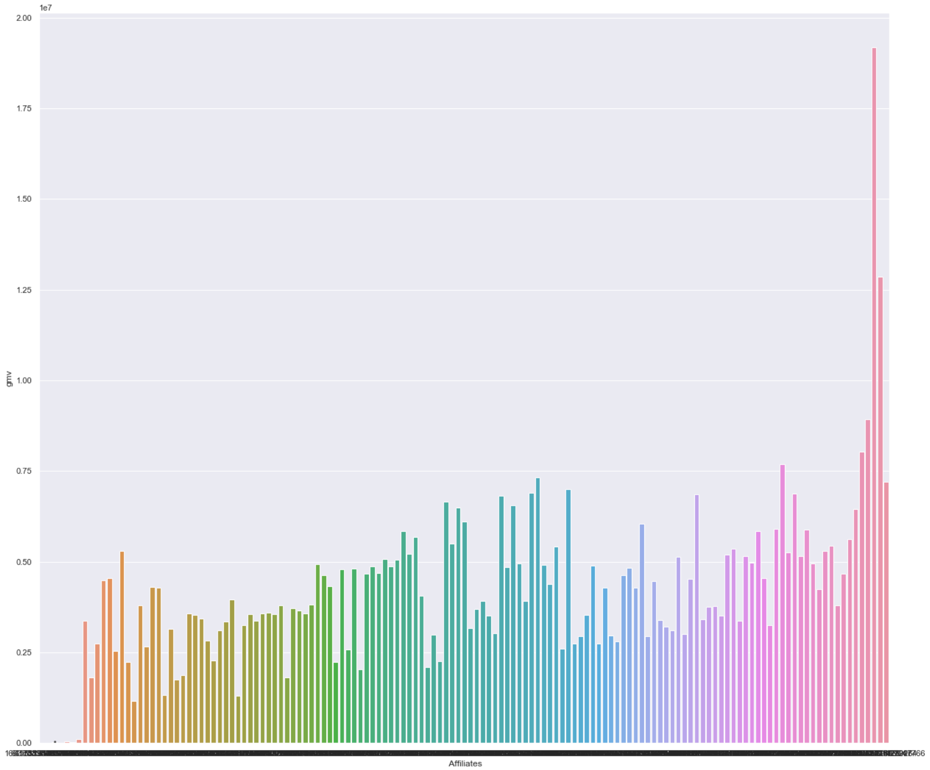


Weekly Ad Spent (Online marketing)

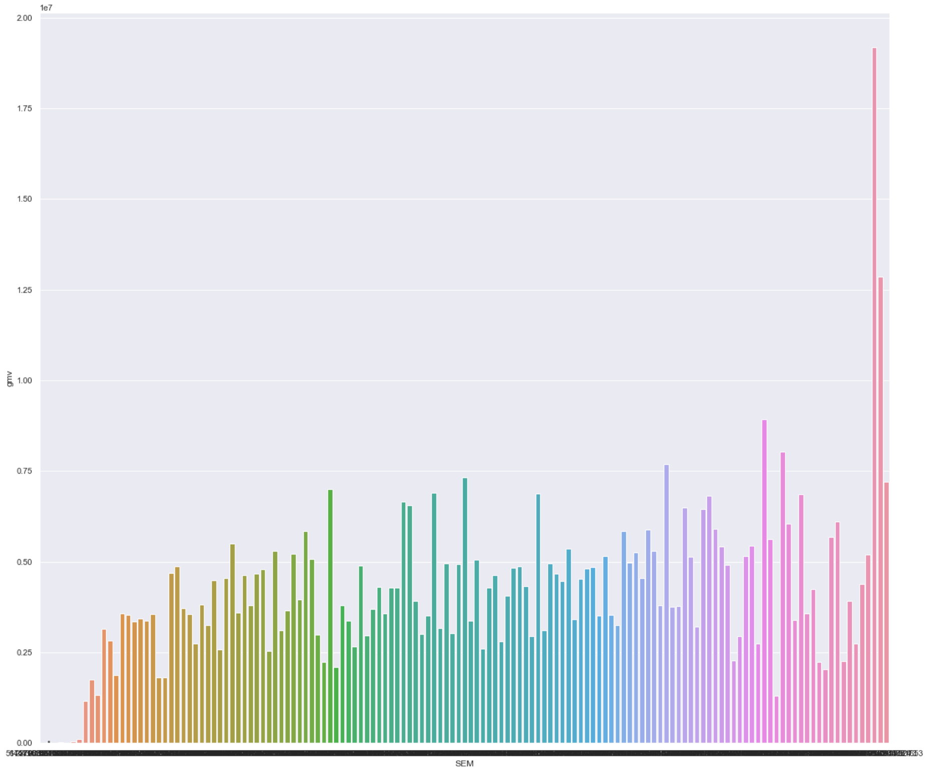


Bar Charts and Graphs of the Master Dataframe

Weekly Ad Spent (Affiliates)

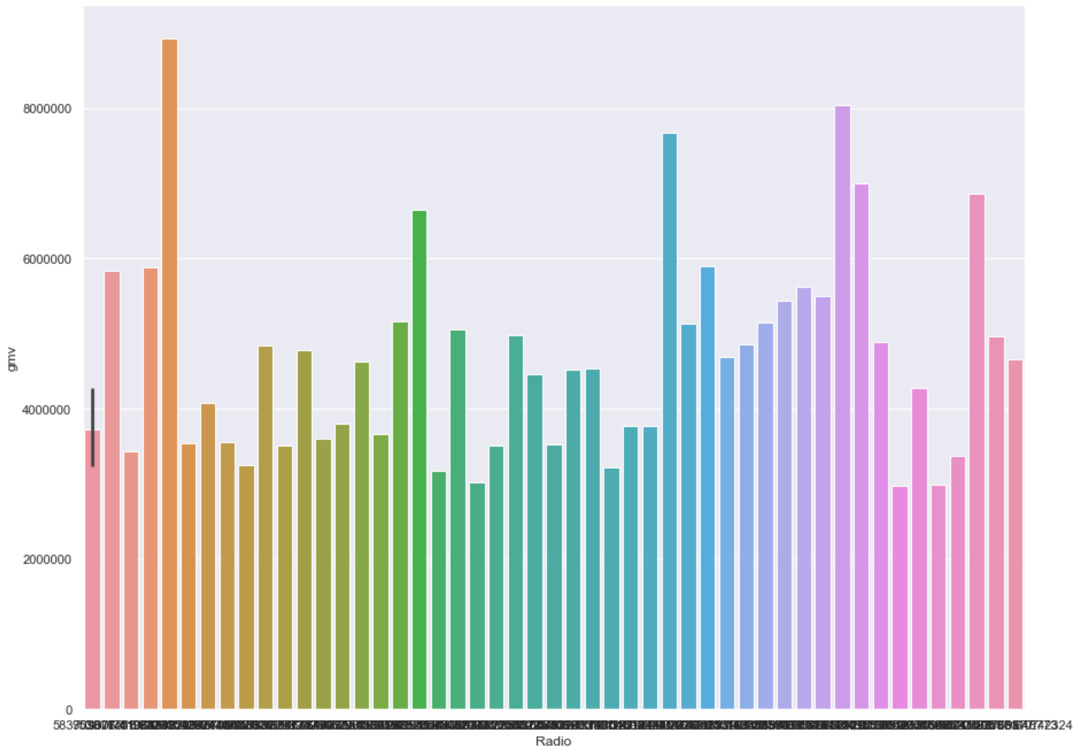


Weekly Ad Spent (SEM)

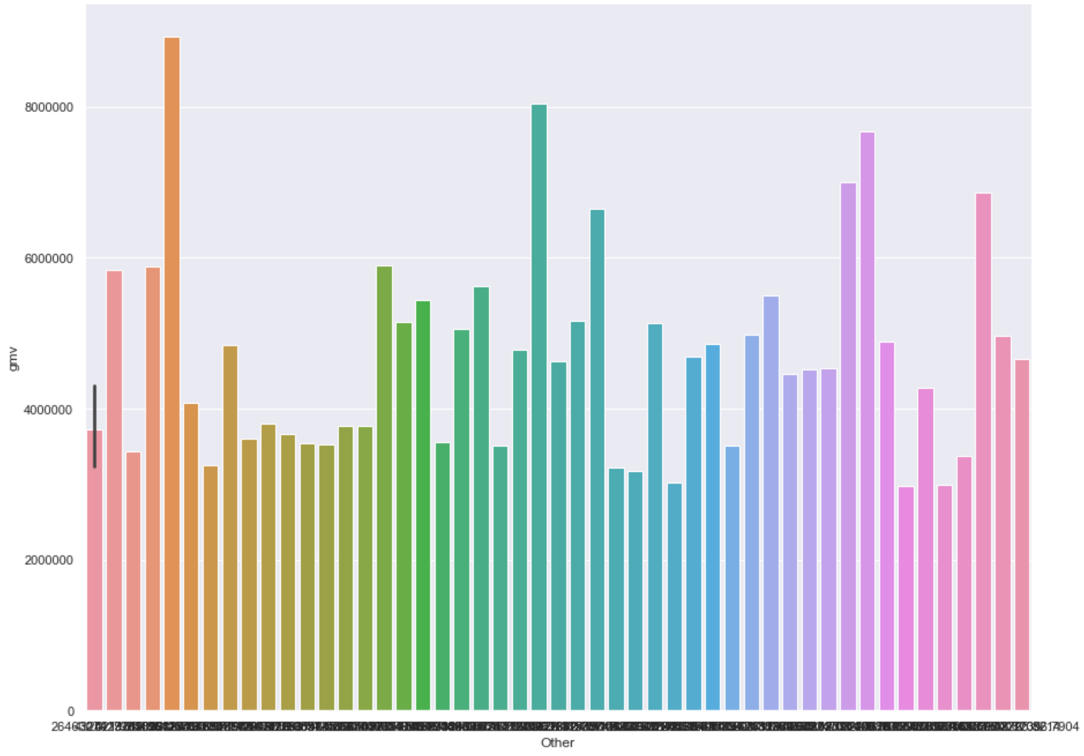


Bar Charts and Graphs of the Master Dataframe

Weekly Ad Spent (Radio)

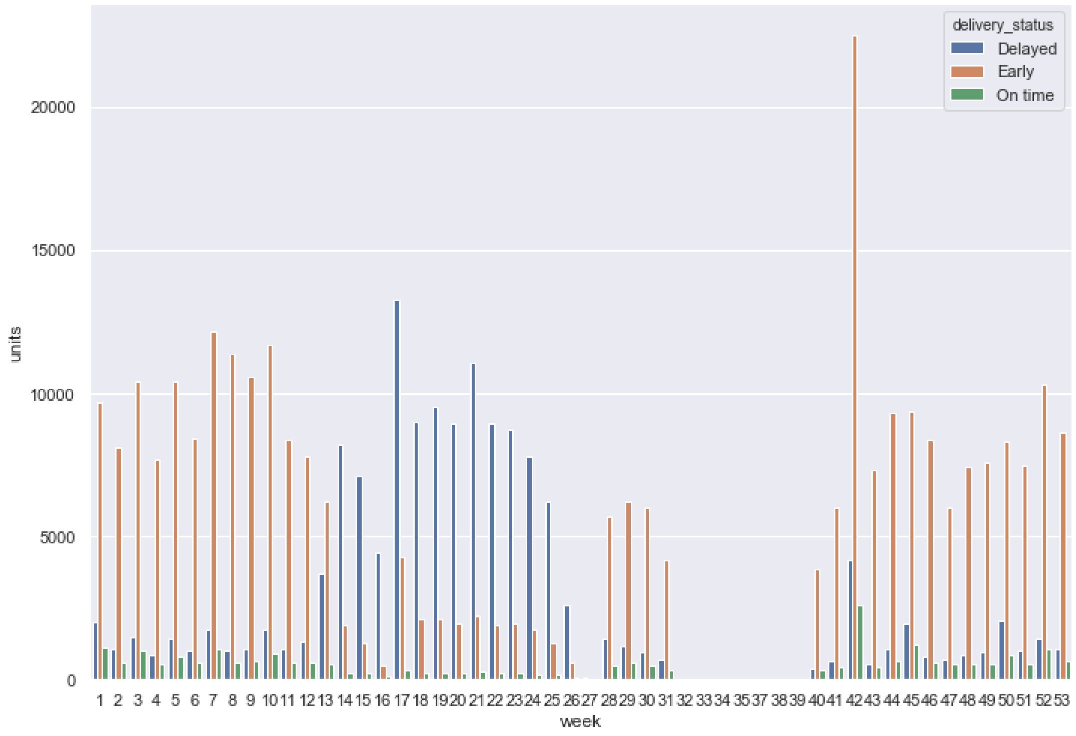


Weekly Ad Spent (Other)

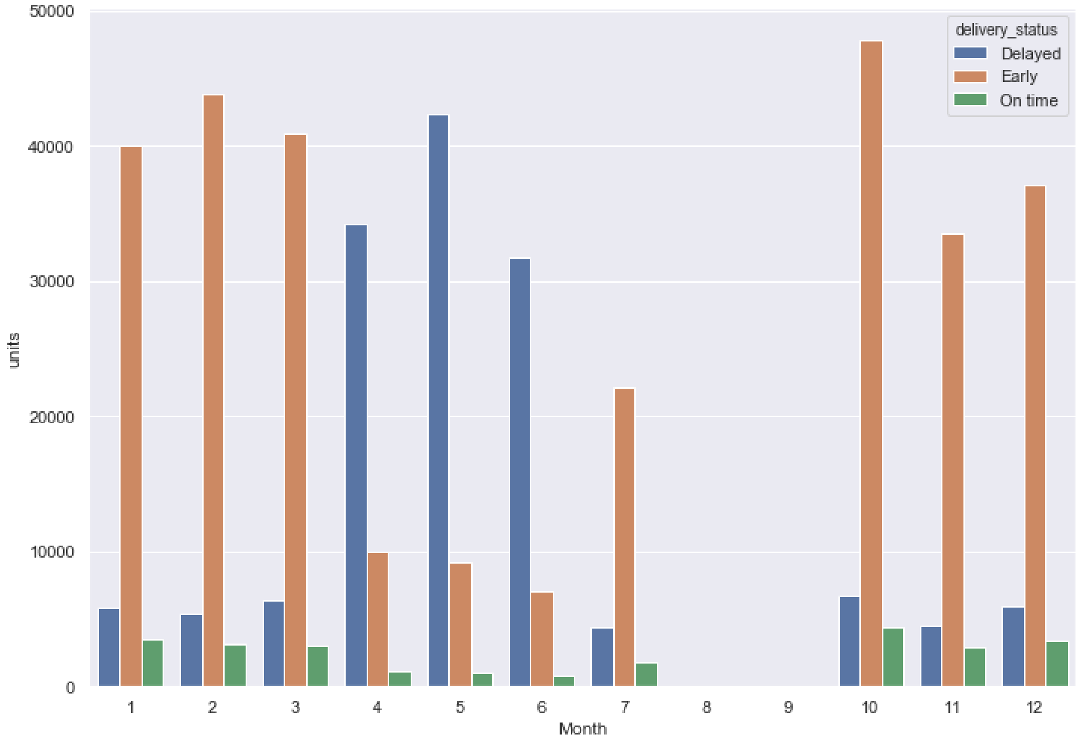


Bar charts and Graphs of the Master Dataframe

Order Status

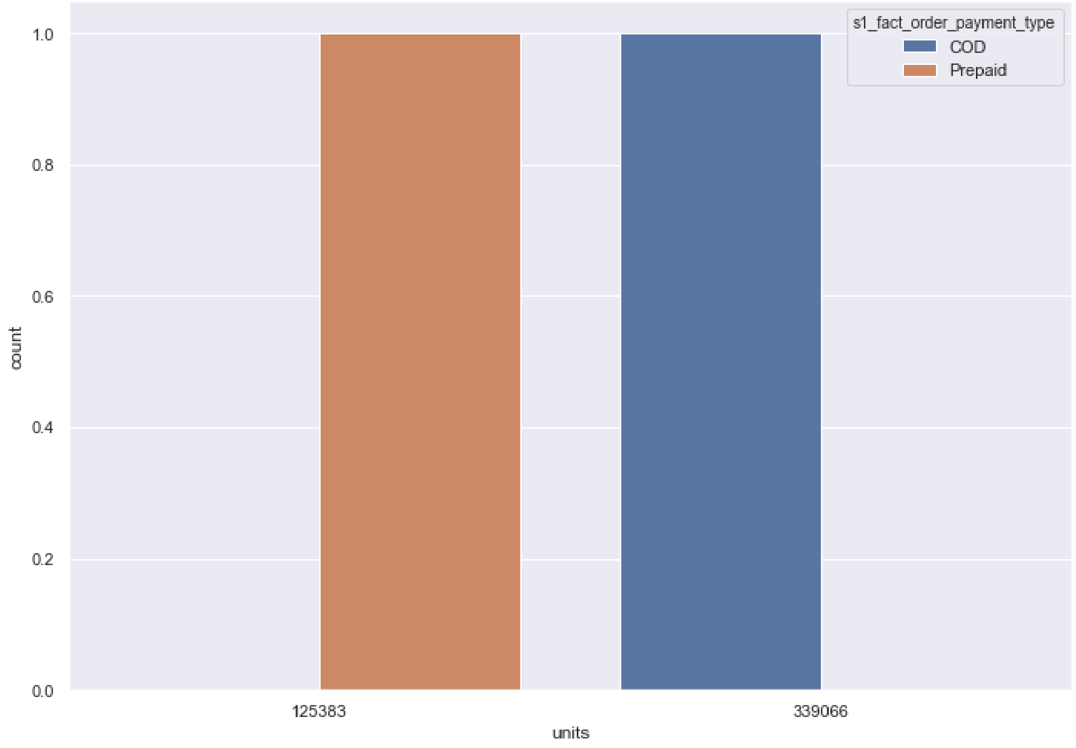


Order Status Monthly

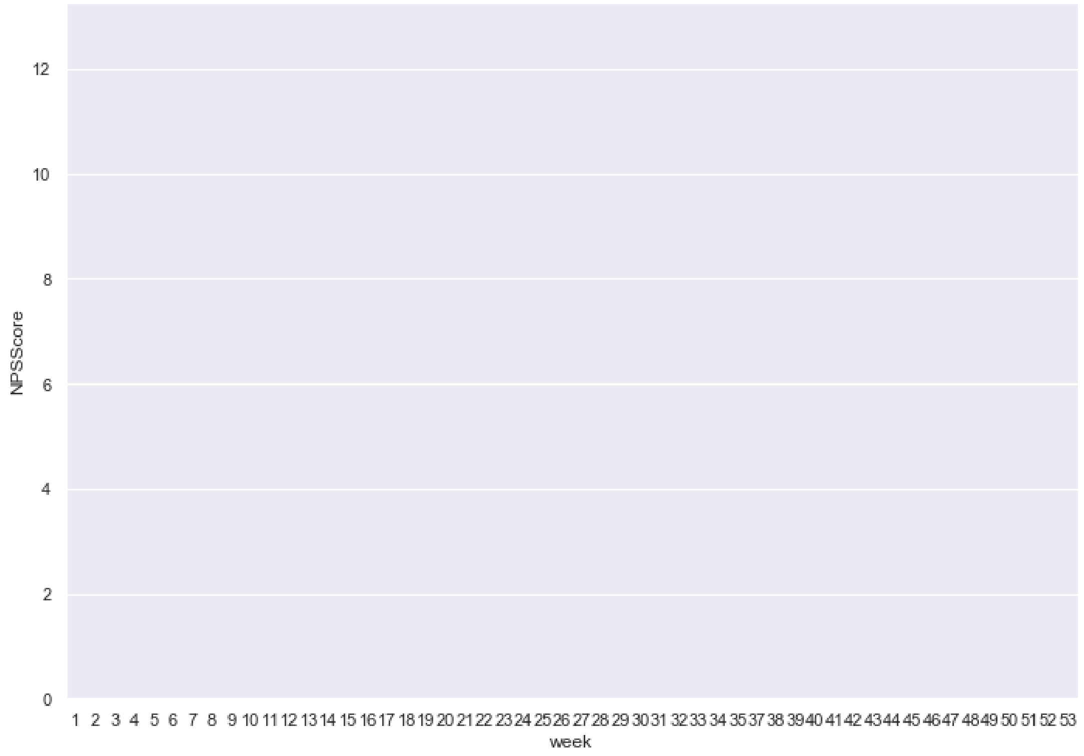


Bar Charts and Graphs of the Master Dataframe

Payment Type

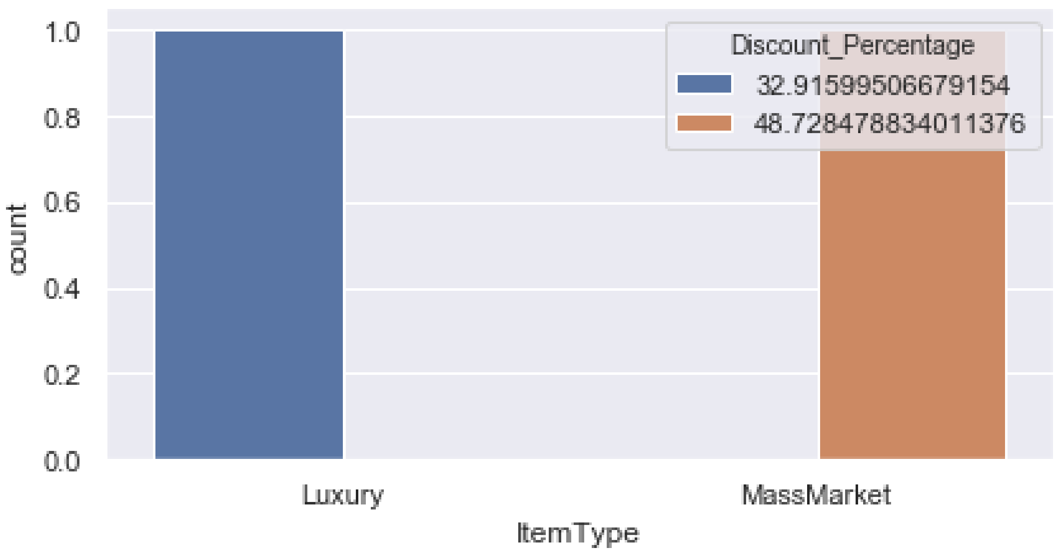


NPS Score

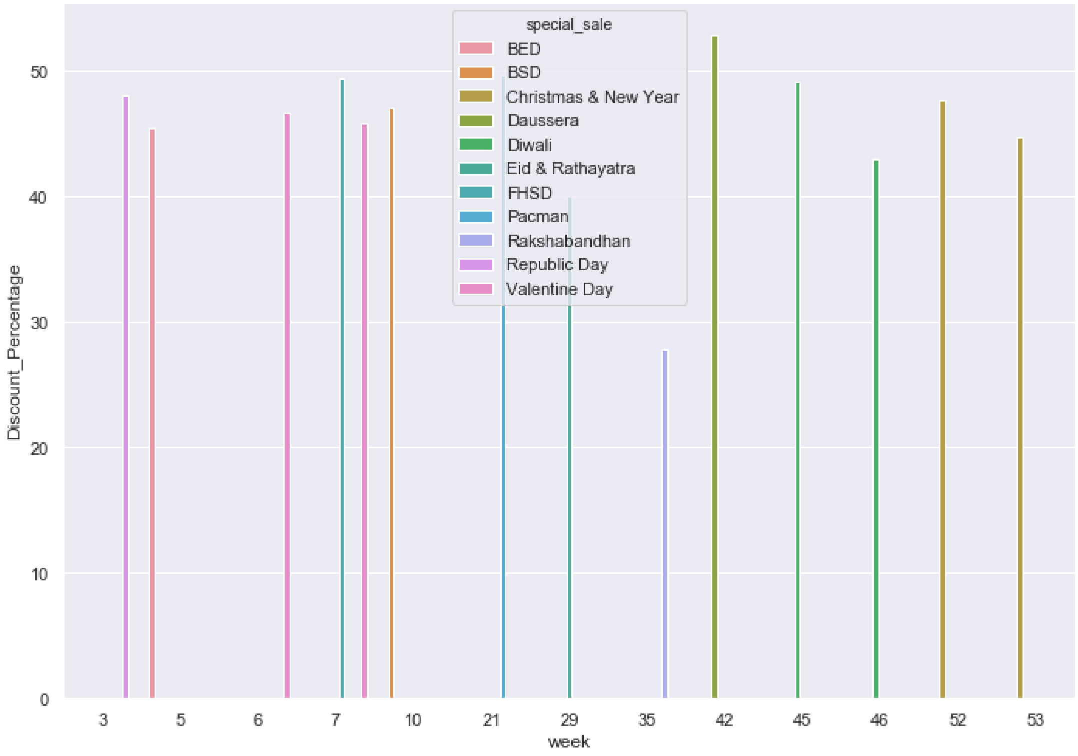


Bar Charts and Graphs of the Master Dataframe

Item Type

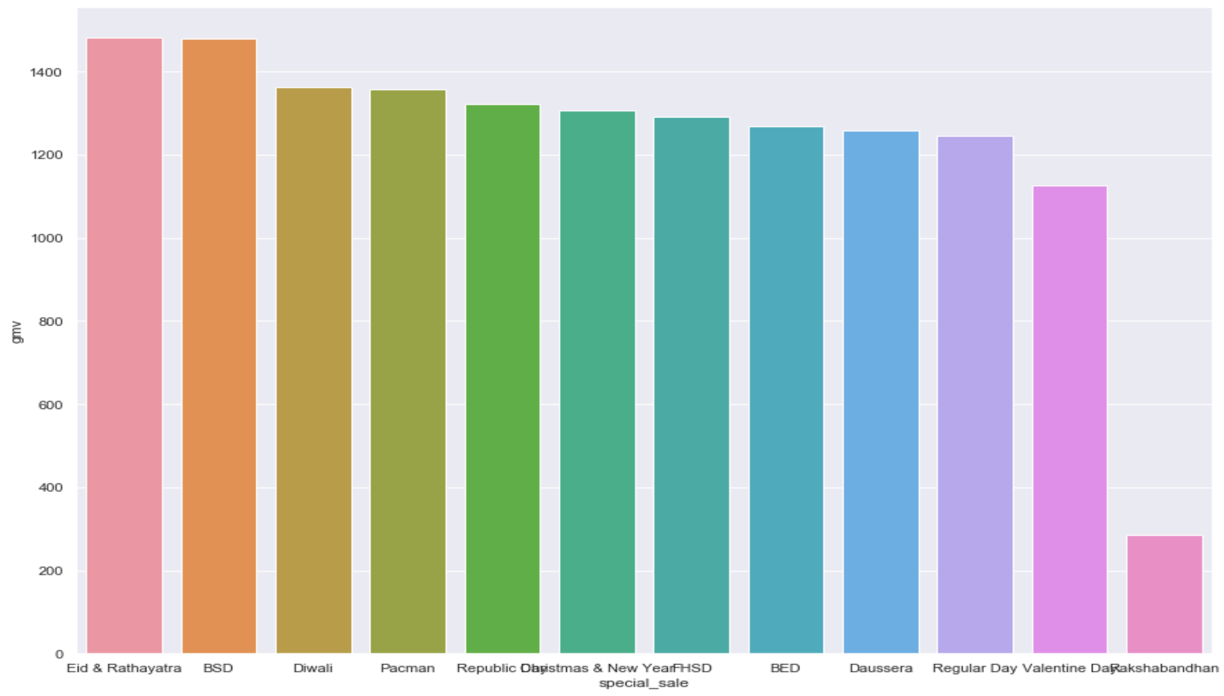


Weekly Discount Percentage



Bar Charts and Graphs of the Master Dataframe

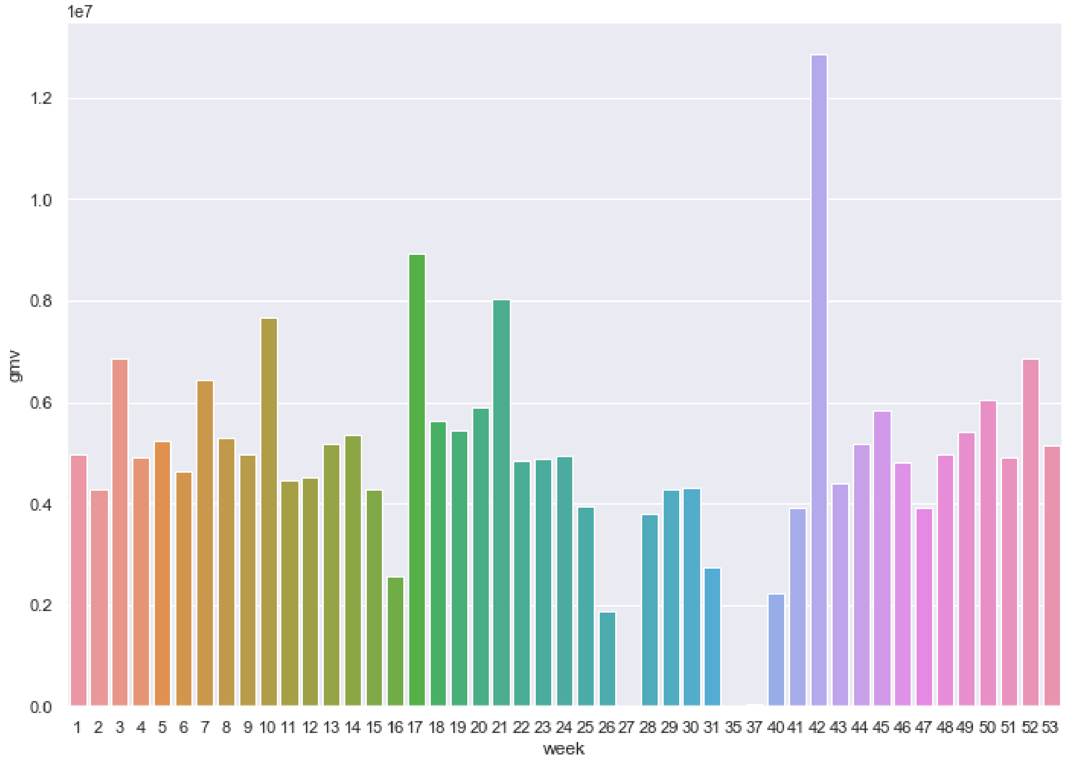
Special Sale gmv



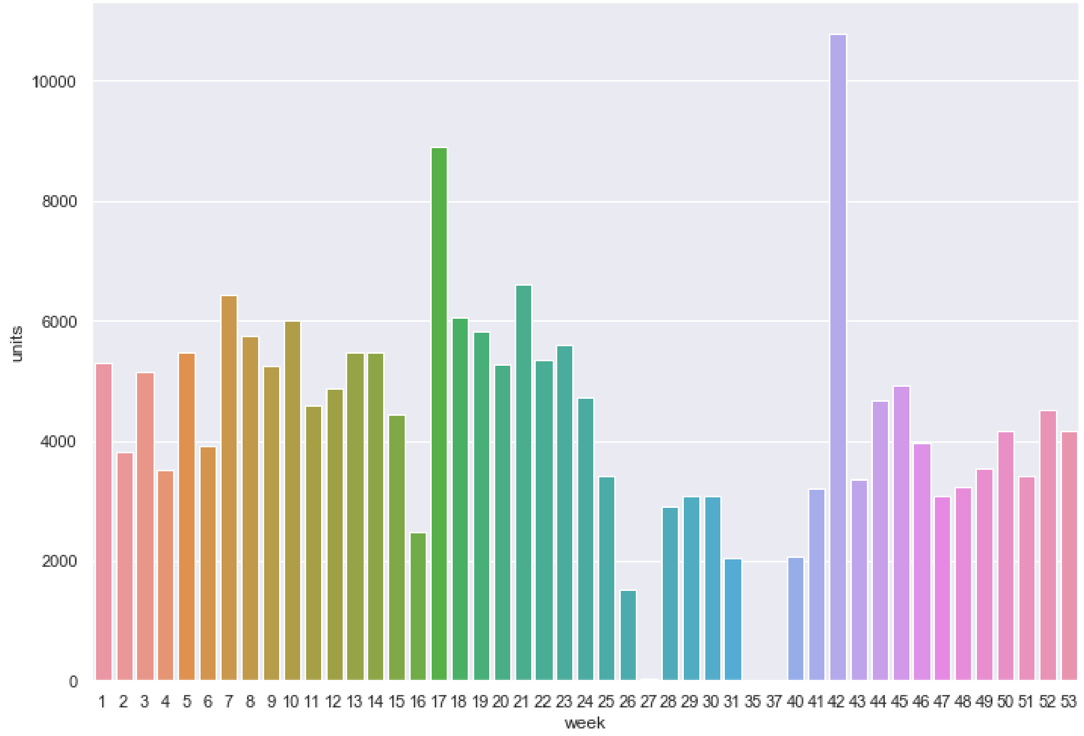
Exploratory Data Analysis based on sub-categories

Camera Accessory

Weekly gmv



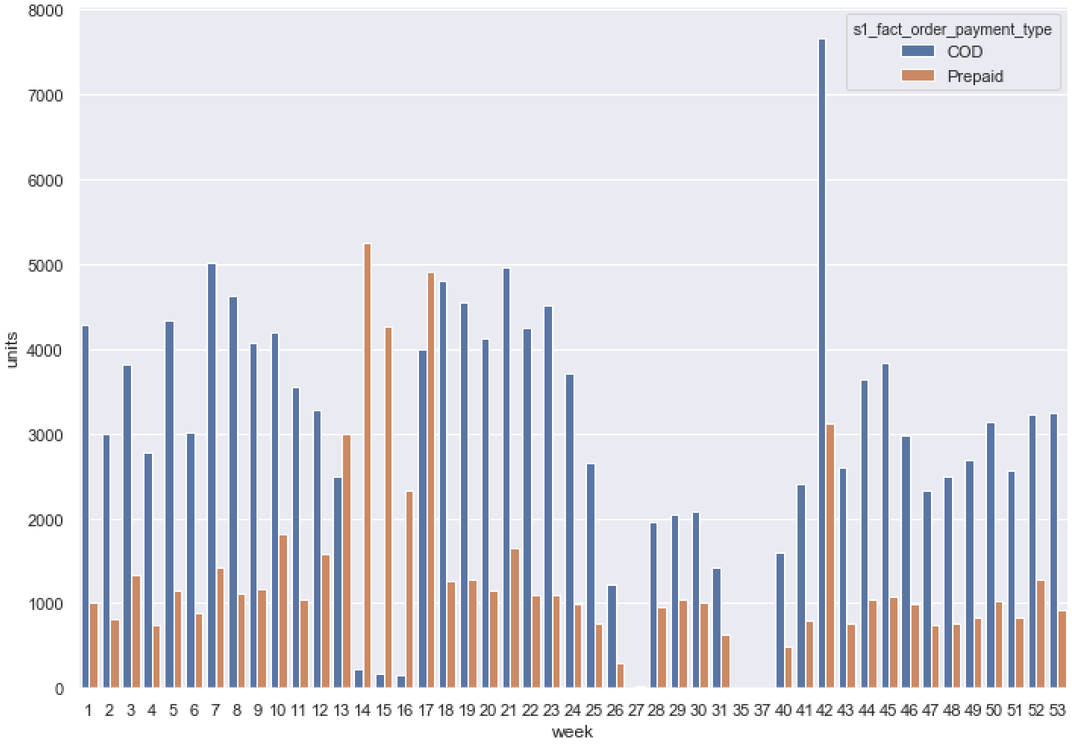
Weekly Units



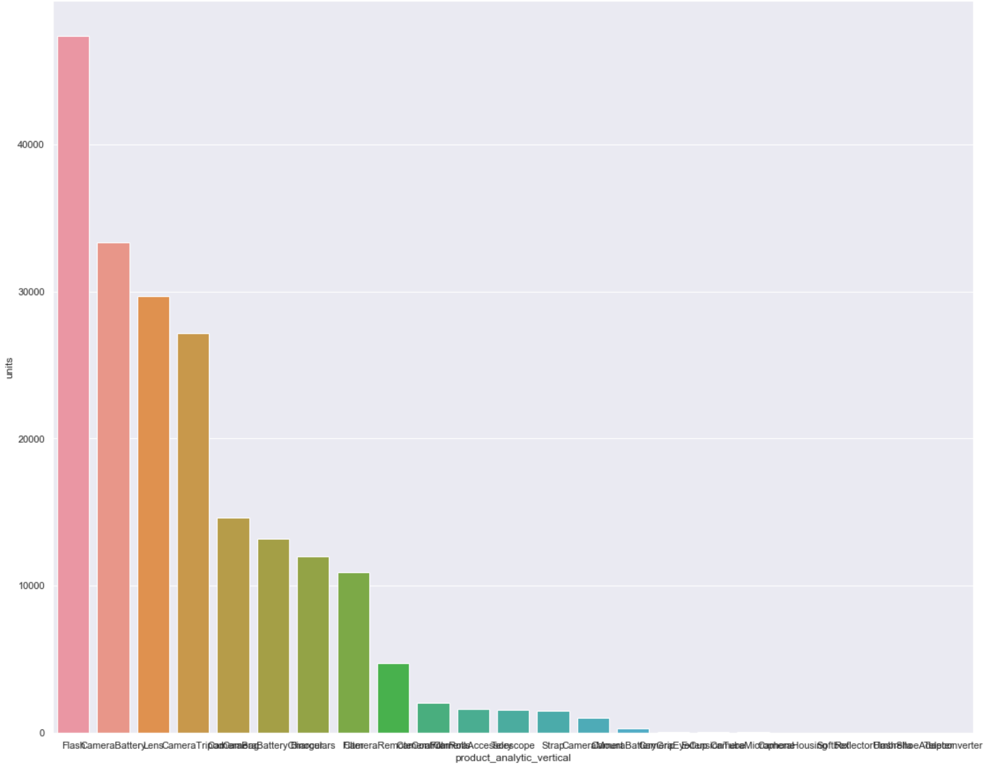
Exploratory Data Analysis based on sub-categories

Camera Accessory

Weekly Units sold by Payment Type



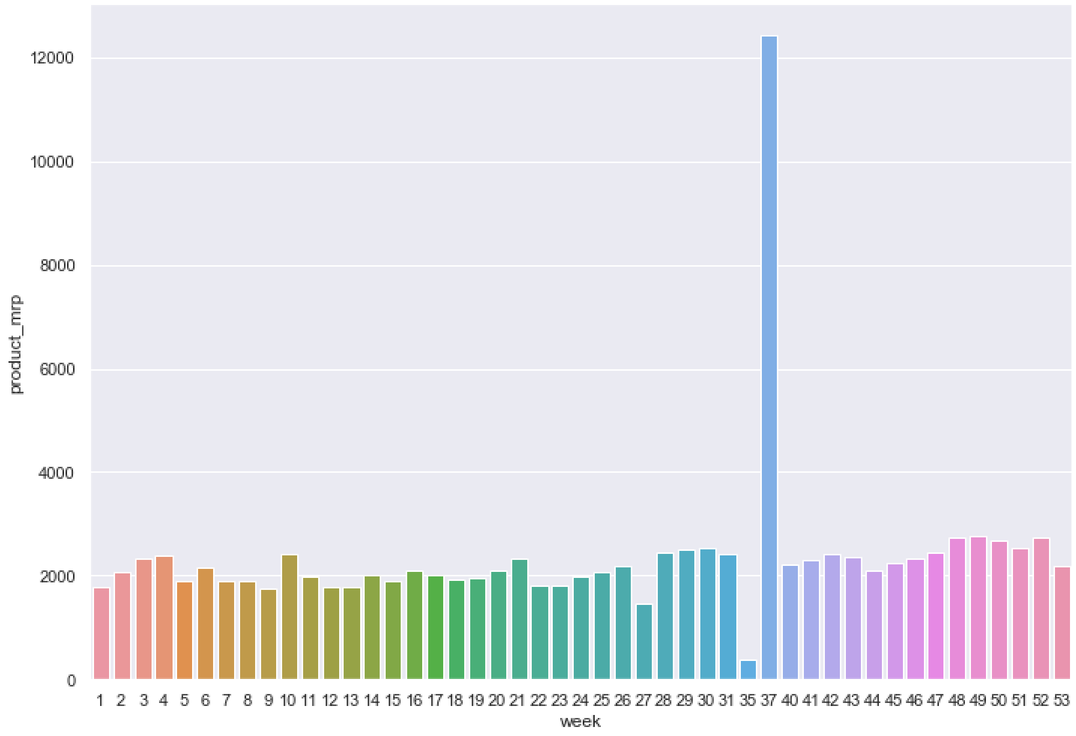
Product Vertical Units sold



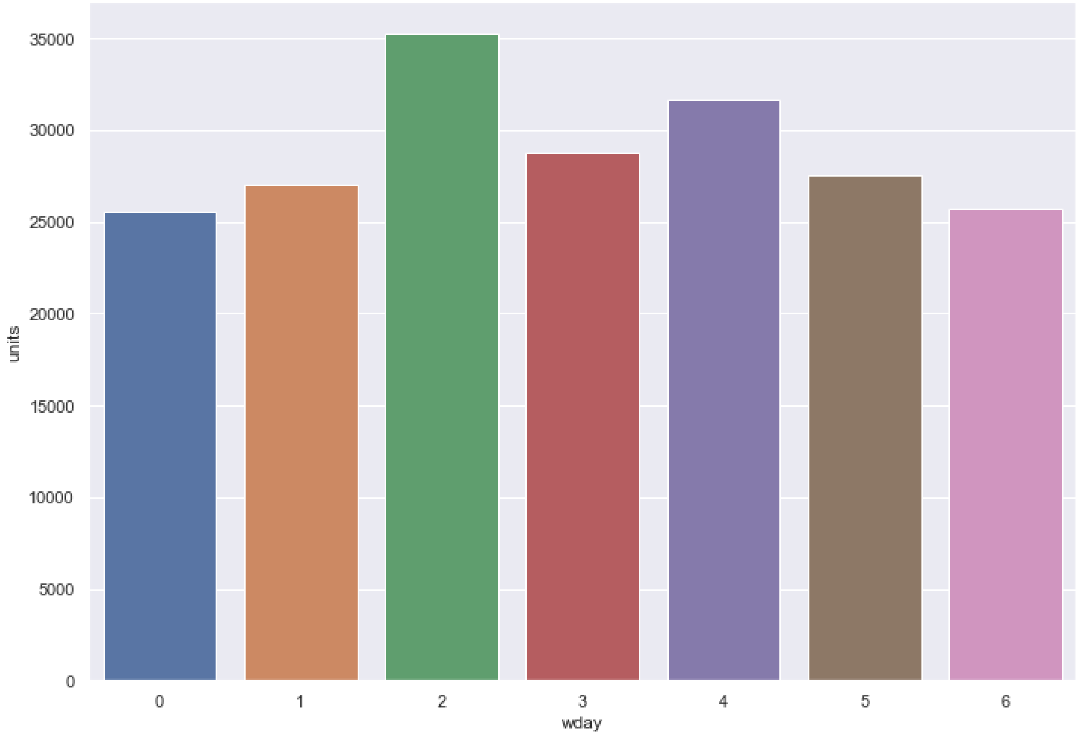
Exploratory Data Analysis based on sub-categories

Camera Accessory

Weekly Product Mrp



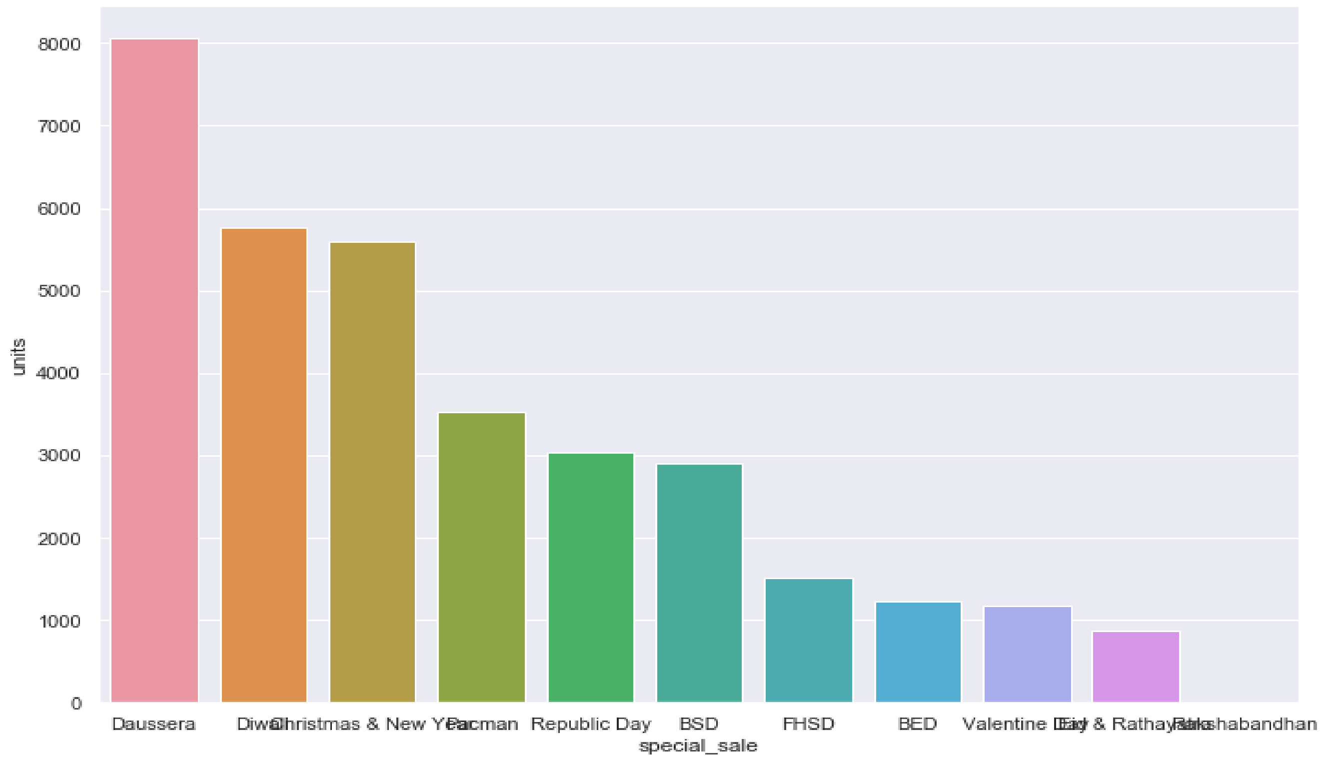
Weekly Units Sold



Exploratory Data Analysis based on sub-categories

Camera Accessory

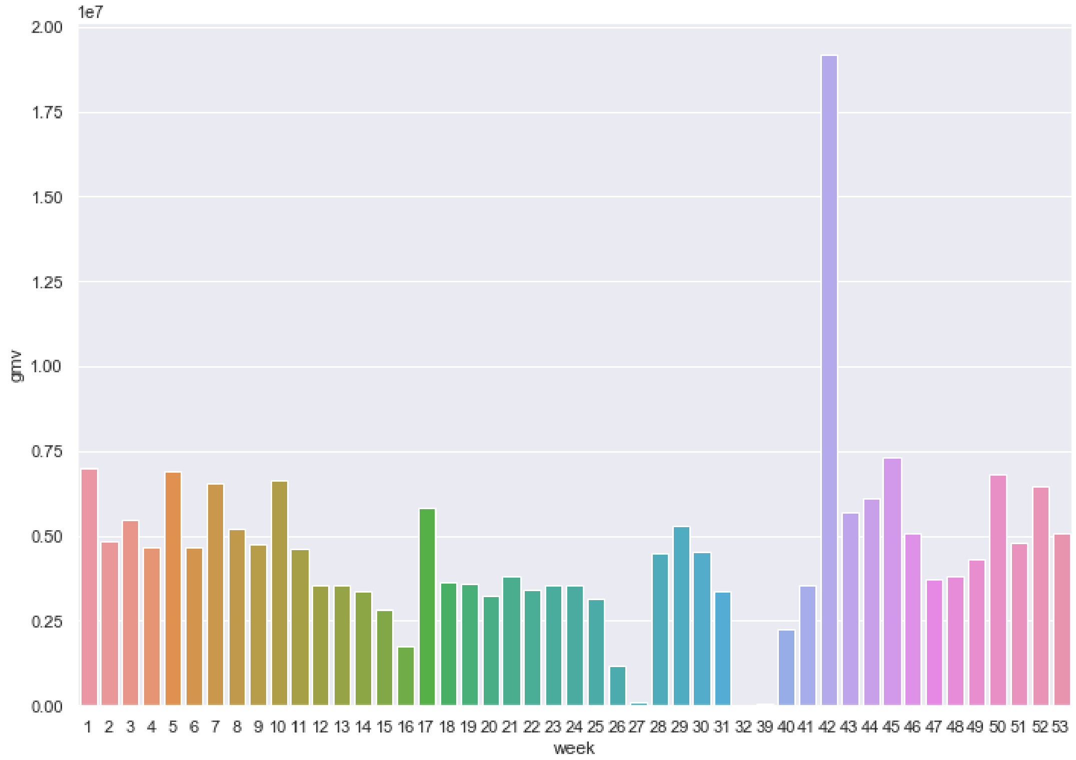
Special Sale Day Units Sold



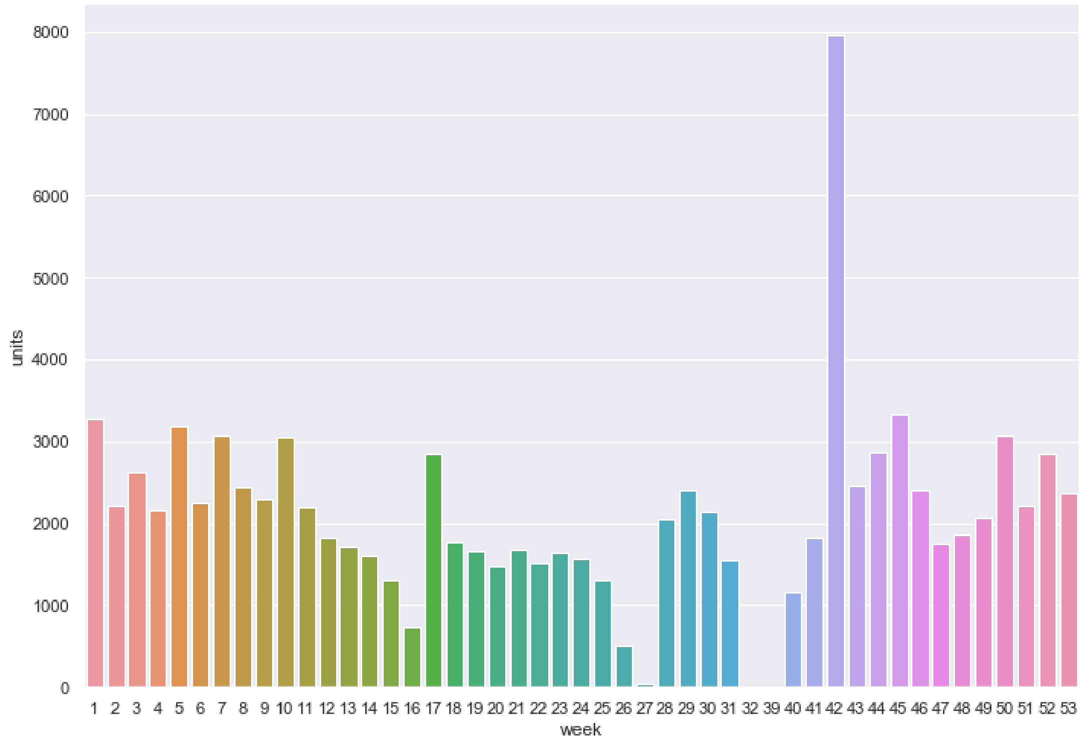
Exploratory Data Analysis based on sub-categories

Home Audio

Weekly gmv



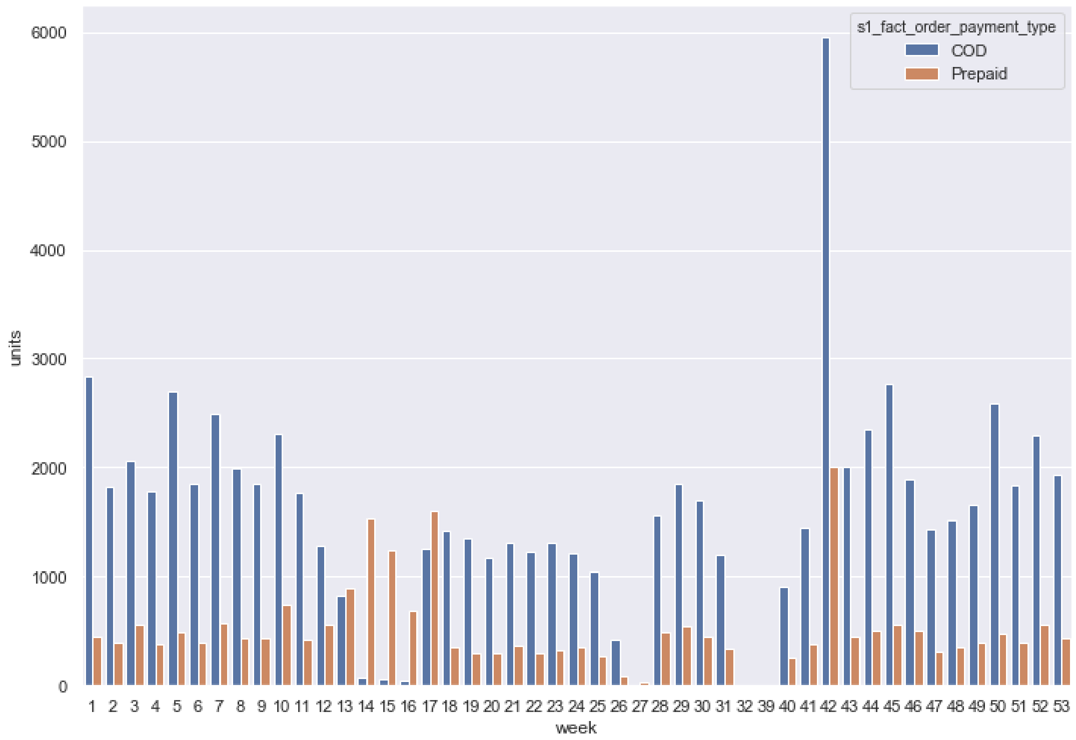
Weekly Units



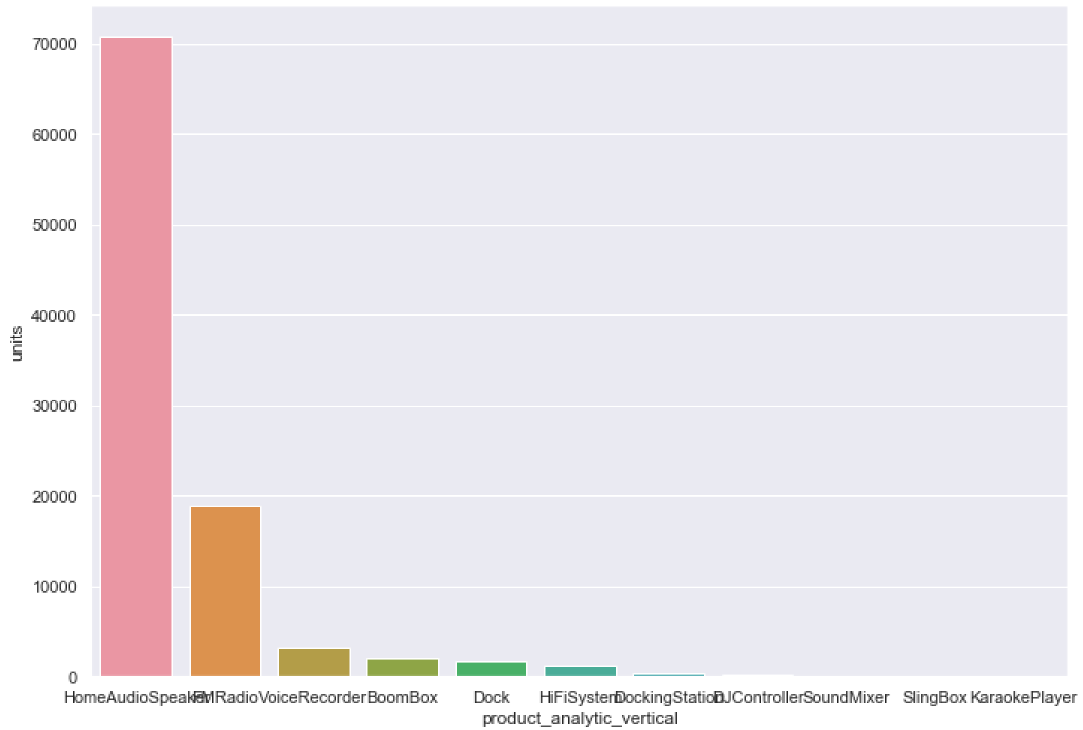
Exploratory Data Analysis based on sub-categories

Home Audio

Weekly Units Sold by Payment Type



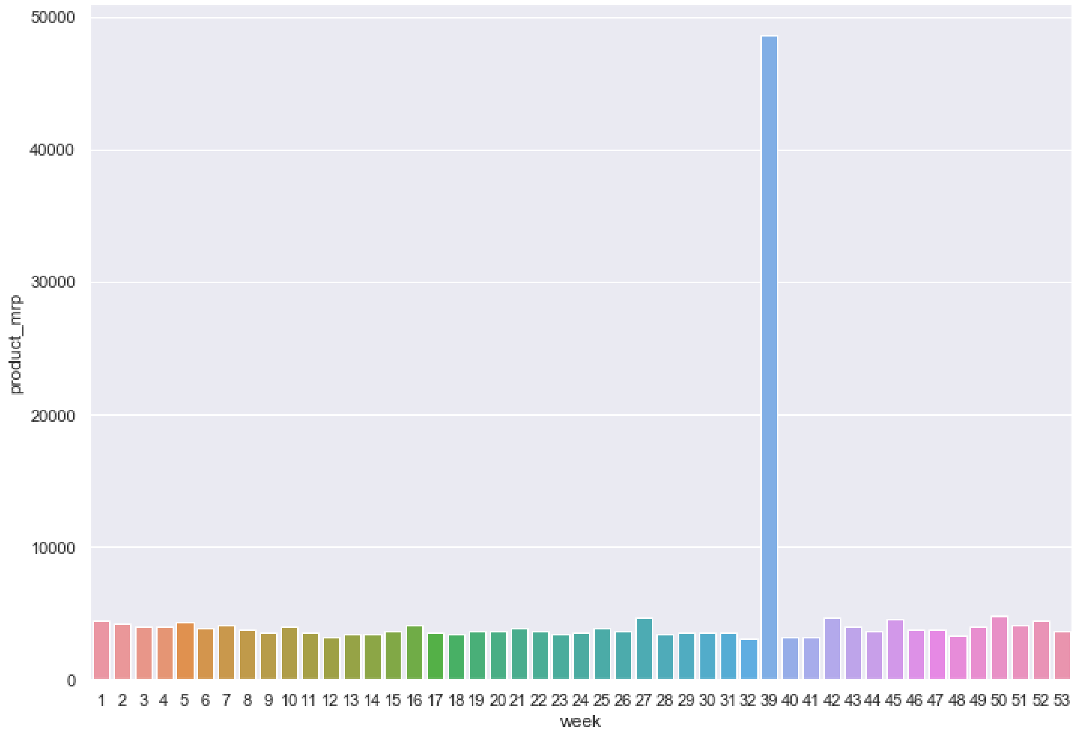
Product Vertical Units Sold



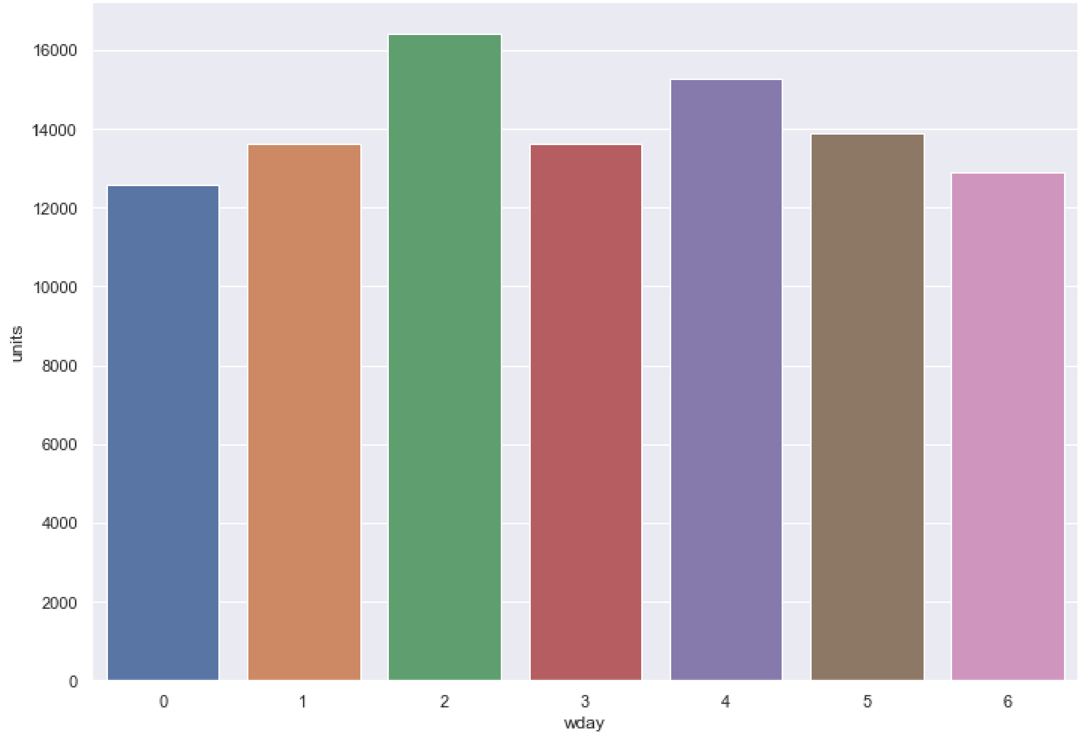
Exploratory Data Analysis based on sub-categories

Home Audio

Weekly Product mrp



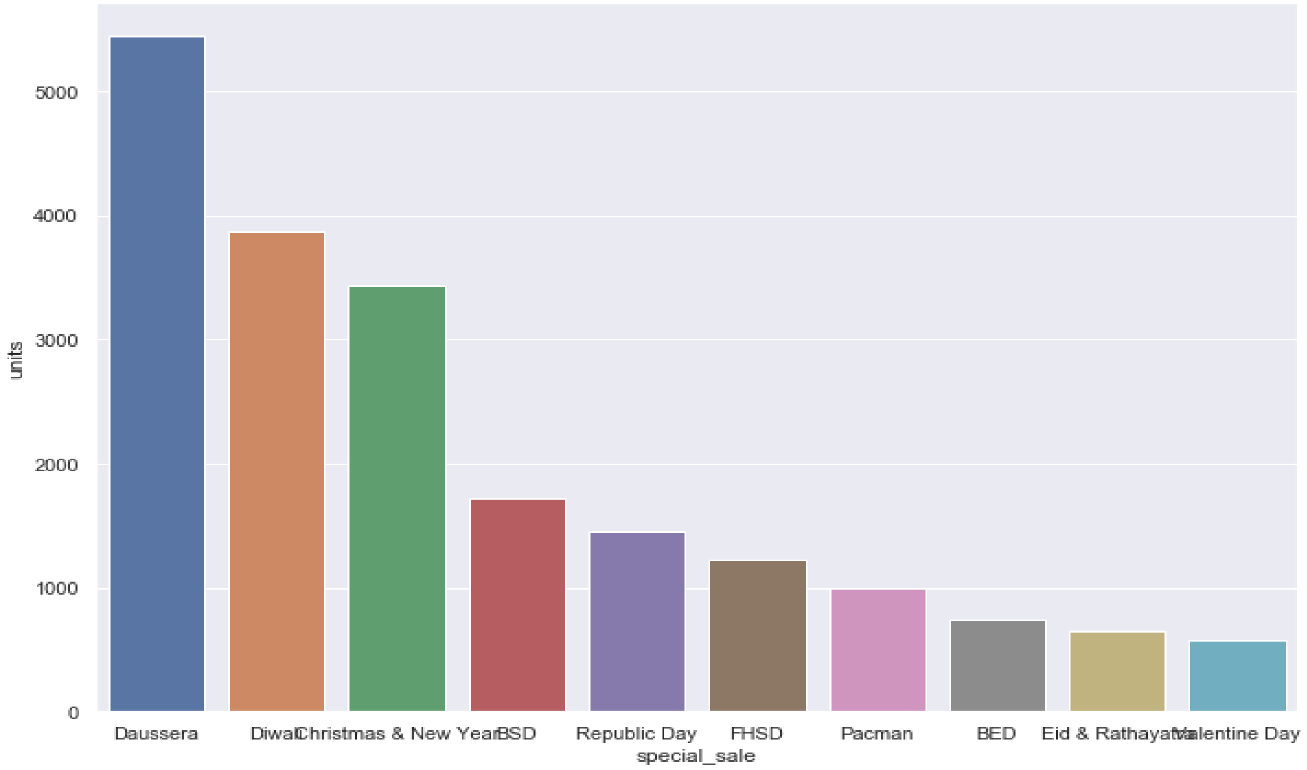
Weekday Units Sold



Exploratory Data Analysis based on sub-categories

Home Audio

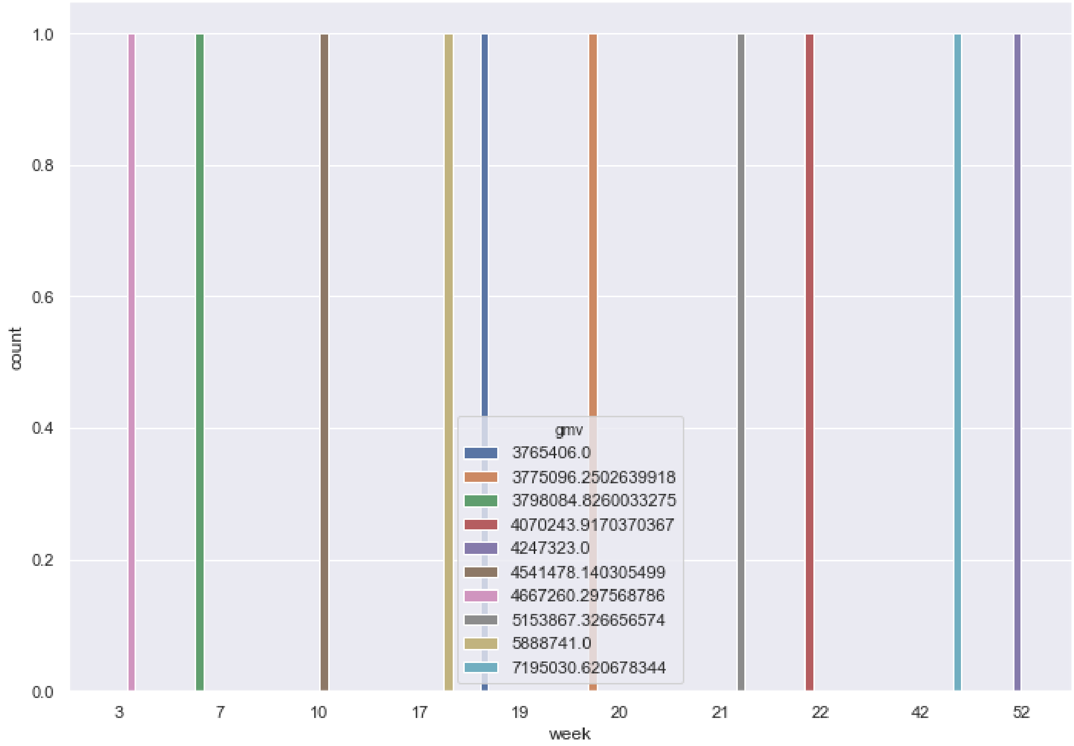
Special Sale Day Units Sold



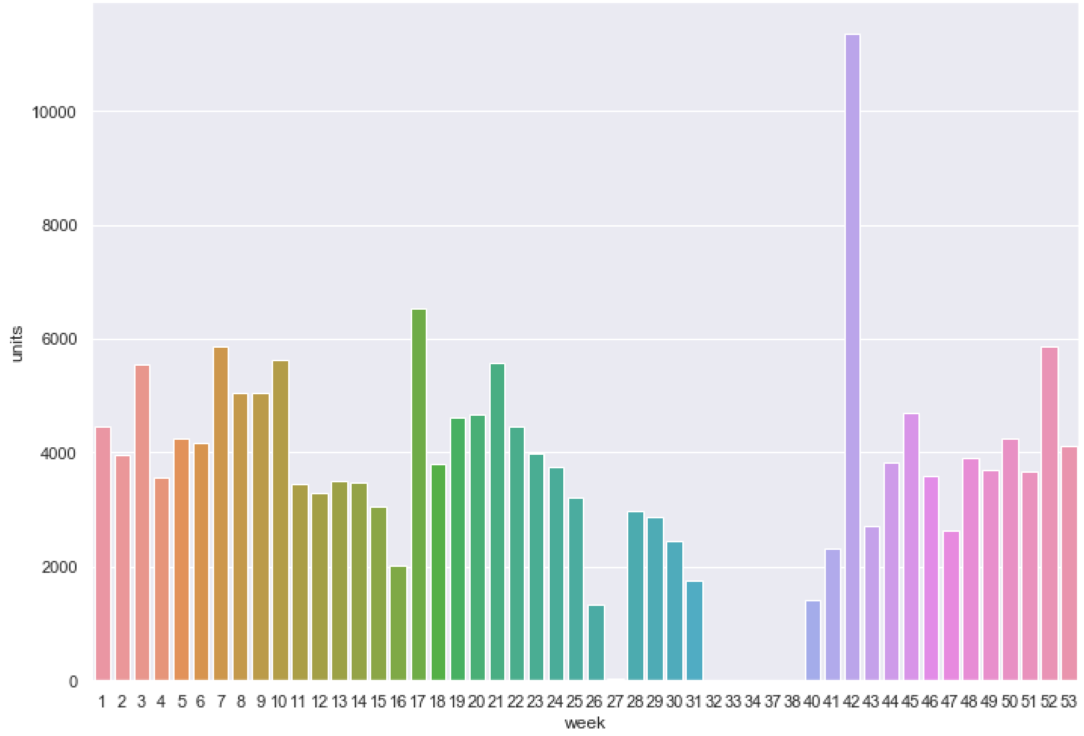
Exploratory Data Analysis based on sub-categories

Gaming Accessory

Weekly gmv

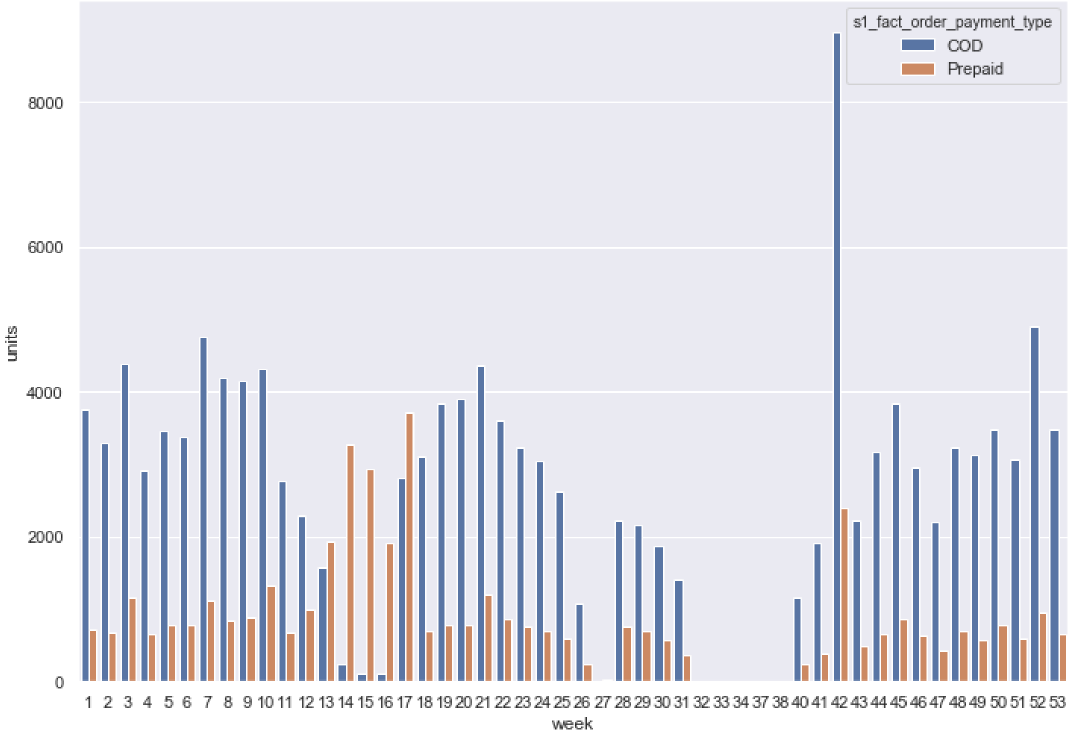


Weekly Units Sold

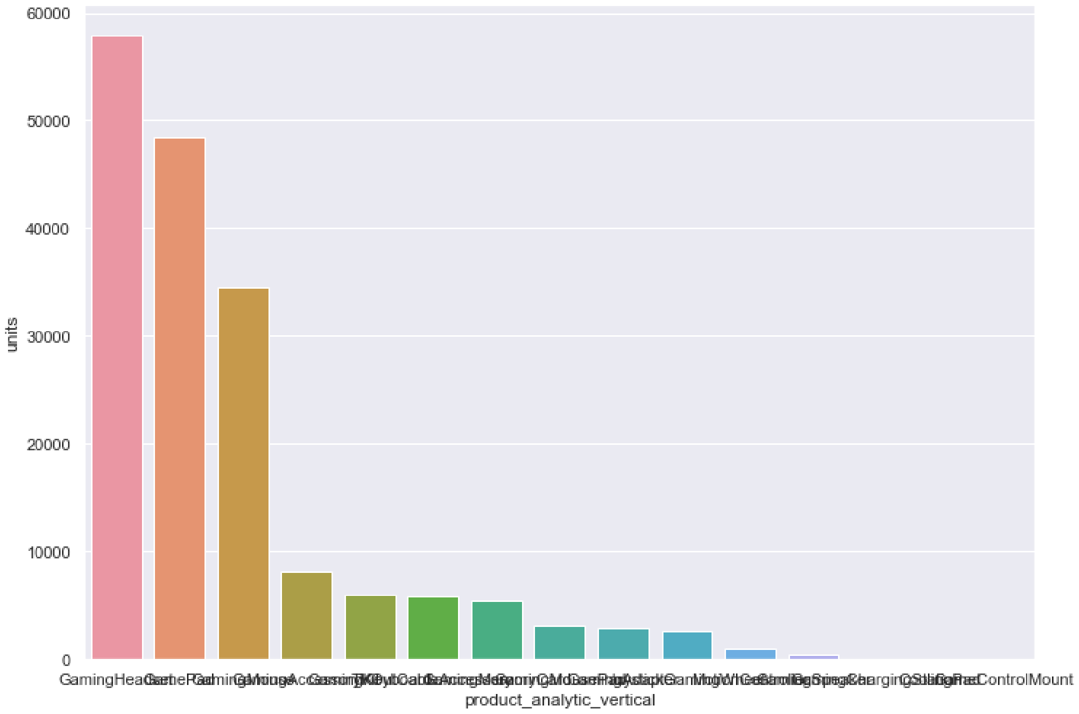


Exploratory Data Analysis based on sub-categories
Gaming Accessory

Weekly Units Sold by Payment Type



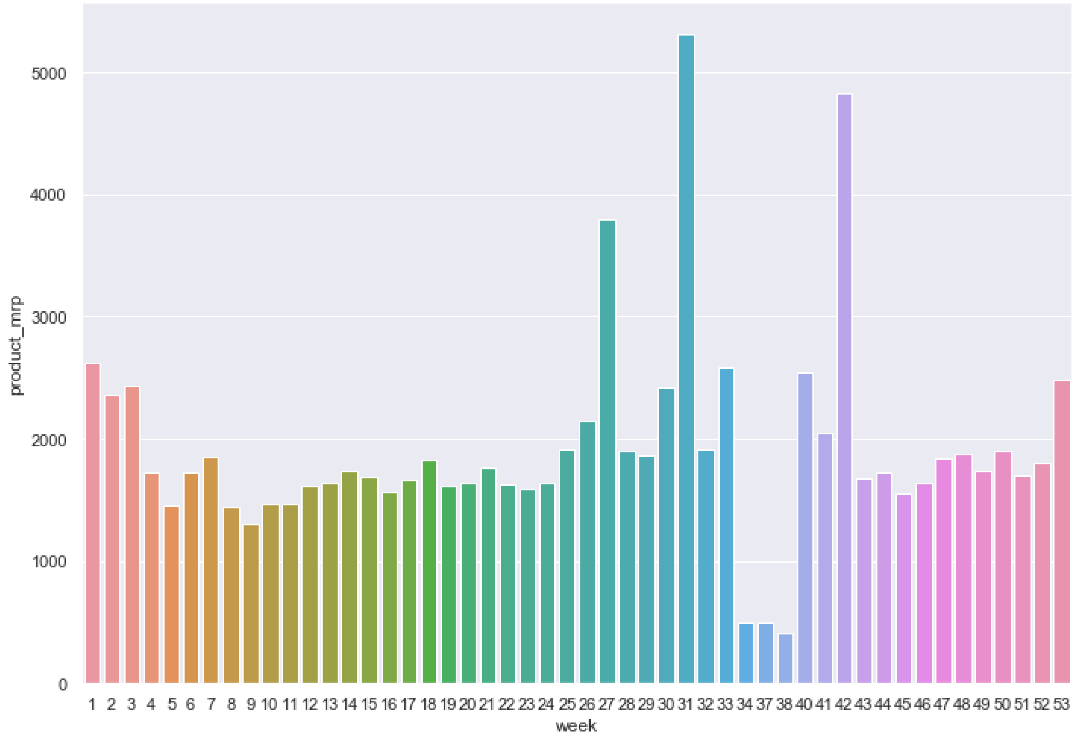
Product Vertical Units Sold



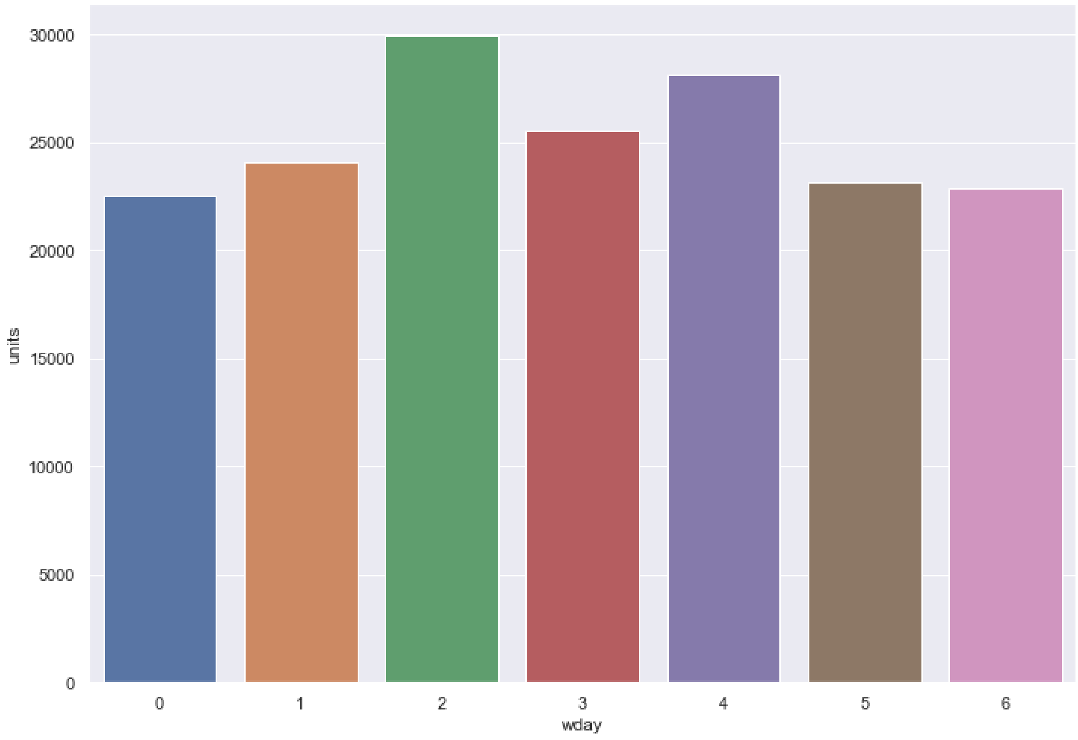
Exploratory Data Analysis based on sub-categories

Gaming Accessory

Weekly Product Mrp



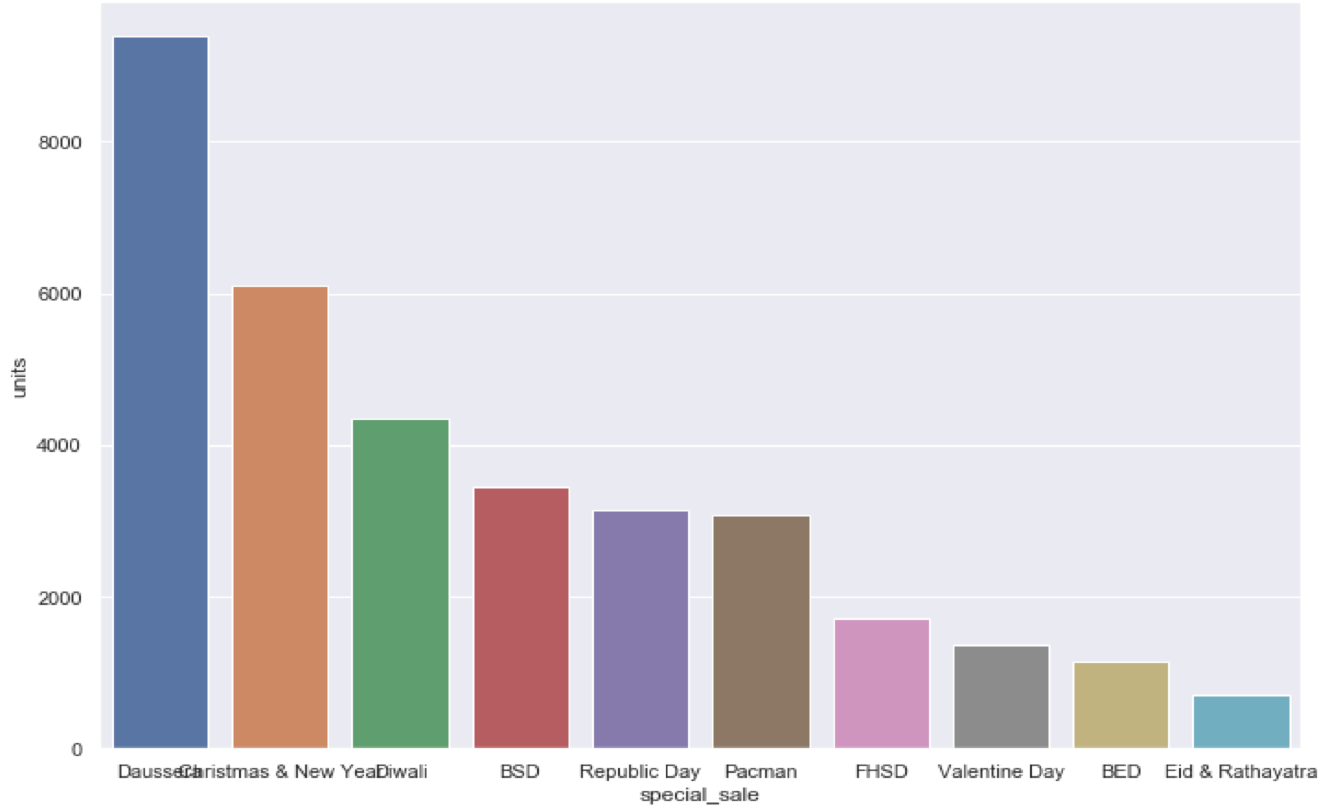
Weekday Units Sold



Exploratory Data Analysis based on sub-categories

Gaming Accessory

Special Sale Day Units Sold



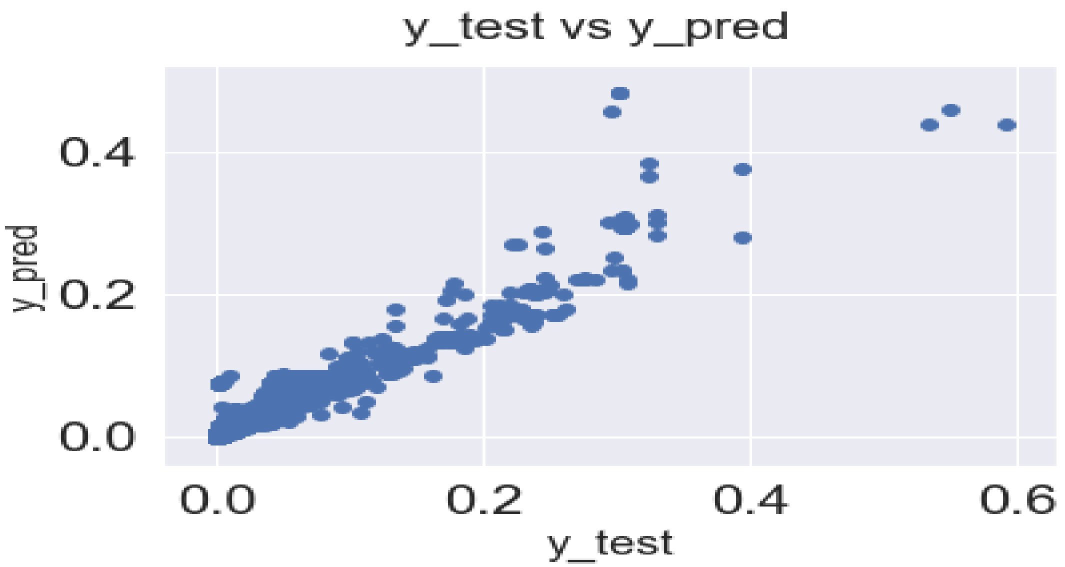
Derived KPIs & Modeling

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	

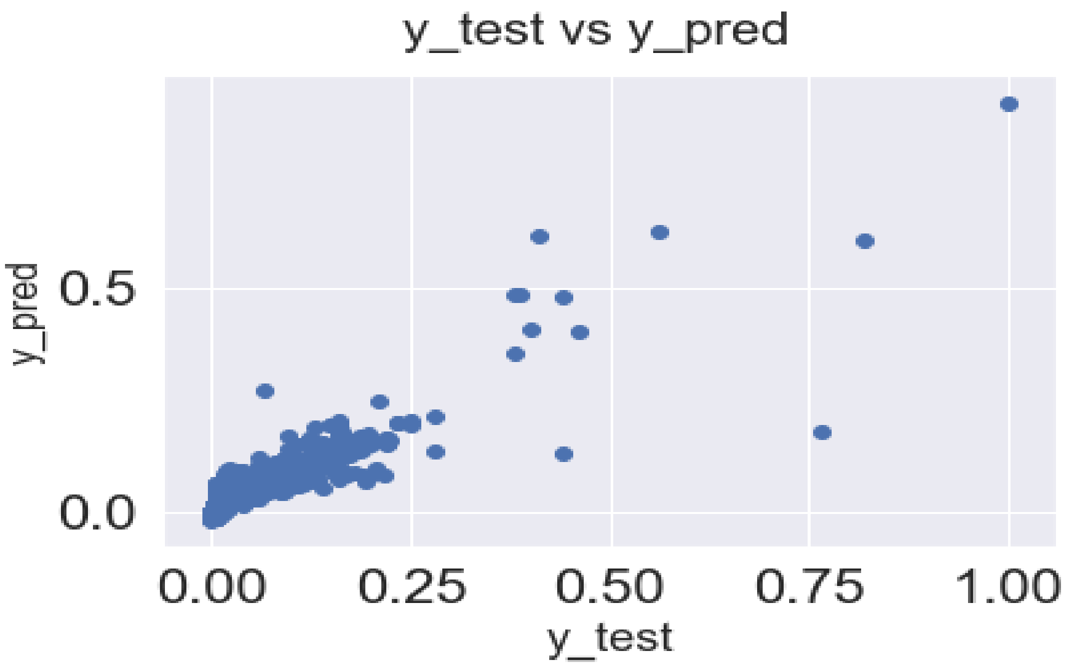
Camera Accessory – Recommendations

- After analyzing 4 different models, its observed that **Koyck** model is best suited for **Camera Accessory**.
- Its has least mean square error (0.007) and have high result in Cross -Validation (0.86) 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 Accessory – 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.**