



CAPSTONE PROJECT

E-COMMERCE

Group Members:

- Vivek Subramanyam
- Shailesh Kumar C



Table of Contents

- 1. Business Understanding
- 2. Current Issues faced by ElecKart
- 3. Optimal Solution
- 4. Understanding the Different Datasets
- 5. Steps for Data Preparation
- 6. Important KPI's
- 7. Exploratory Data Analysis
- 8. Model Building Linear Regression



ElecKart is an E-commerce firm based out of Ontario, Canada specialising in electronic products

Offers big-ticket promotions (similar to the Big Billion Day)

Spent excess amount on Marketing for Commercials, Online Campaigns, Pricing and Promotions

Business Understanding



Current Issues faced by ElecKart

Spent significant amount of money on Marketing

Money Spent over last 12 months was not impactful

Unsuccessful budget reallocation among different marketing channels

Unsuccessful usage of Market Mix Model



Optimal Solution

Develop the right Market Mix Model for Categories: Camera accessory, Gaming accessory and Home Audio

Recommend correct budget allocation for the coming year

Cut costs wherever required

Ensure right Discounts and Promotions are provided during the peak seasons



Understanding the different datasets (Order Data)

FSN ID: The unique identification of each SKU

Order Date: Date on which the order was placed

Order ID: The unique identification number of each order

Order item ID: Suppose you order 2 different products under the same order, it generates 2 different order Item IDs under the same order ID;

orders are tracked by the Order Item ID.

GMV: Gross Merchandise Value or Revenue

Units: Number of units of the specific product sold

Order payment type: How the order was paid – prepaid or cash on

delivery

SLA: Number of days it typically takes to deliver the product

Cust id: Unique identification of a customer

Product MRP: Maximum retail price of the product

Product procurement SLA: Time typically taken to procure the product



Understanding the different datasets (contd. . .)

Media Investment:

Monthly spend on various advertising channels like TV, Digital, Sponsorship, Content Marketing, Online Marketing, Affiliates, SEM, Radio and Other

Special Sales Calendar:

Days when there was any special sale or Promotional Week Sales

Monthly NPS score : (Net Promoter Score):

- Customer Satisfaction Score is calculated each month.
- This score is given by the customer for the products that he has already purchased

Climate Dataset:

- Climate Parameters are provided from 2015-Jul to 2016-June
- Need to check if this data has any impact on Sales

Steps for Data Preparation (Order Data)



- 1. Read the Order Data into a Dataframe
- 2. Analyze the Dataframe by checking various properties
- 3. Create New Calculated Columns (List Price, Discount, Delivery_status)
- 4. Delete the rows with gmv=null & 0, product_mrp is null & 0
- 5. Delete rows where gmv > product_mrp
- 6. Delete the other columns which will not be used in Weekly Data
- 7. if Product_mrp x units >= 'gmv' then include, else exclude the data
- 8. Rename the columns for Better Readability
- 9. Perform the Outlier Analysis for 'sla', 'gmv', 'product_mrp'
- 10. Add the Week Logic from 0 to 52 starting from July-15 to June-16
- 11. Create a Holiday_flag based on the dates provided



Steps for Data Preparation (Order Data contd...)

- 12. Create a New Paydate flag
- 13. Retain 3 Subcategories: HomeAudio, GamingAccessory, CameraAccessory
- 14. Approportionate the GMV value
- 15. Split the Dataframes based on Subcategories (3 Dataframes)
- **16.** Create the Dummies for Categorical Variables
- 17. Convert the Daily Data to weekly based on Week Number
- 18. Create 3 Order Datasets (grouped by Week Number)



Steps for Data Preparation (Media Investment)

1. Read the excel into a Dataframe

2. Parse the Dataframe to pick the necessary rows in the excel

3. Divide the Investment Data by 4 or 5 based on Number of Weeks

4. Explode a Single row into multiple rows based on Week Count

5. Create the Final Media Dataframe



Steps for Data Preparation (NPS & Stock Index Dataset)

1. Read the excel into a Dataframe

2. Parse the Dataframe to pick the necessary rows in the excel

3. Transpose the Dataframe

4. Create New Year and Month Column

5. Add the Week Number Logic for the Dataset

6. Prepare the Final Dataframe

Steps for Data Preparation (Climate Dataset)



- 1. Read the Climate Data excels into 2 Dataframes (2015 and 2016)
- 2. Parse the Dataframe to pick the necessary rows in the excel
- 3. Pick the Data from 2015-July to 2016-June
- 4. Add the Week Number Logic for the Dataset
- **5. Drop the Unnecessary Columns**
- 6. Impute the Nan's with Last Week Data
- 7. Aggregate the Data based on Week Number
- 8. Prepare the Final Dataset



Steps for Data Preparation (Merge all the Datasets)

1. Merge the Climate, Media, NPS data together in a Single Dataframe

2. Create the Master Dataset for the above 3

3. Merge each category data with the above master dataset

4. Use Week number column to join the dataframes

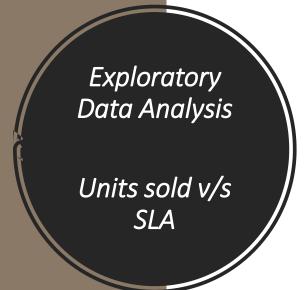
5. Create 3 Different Merged Dataframes for Modelling

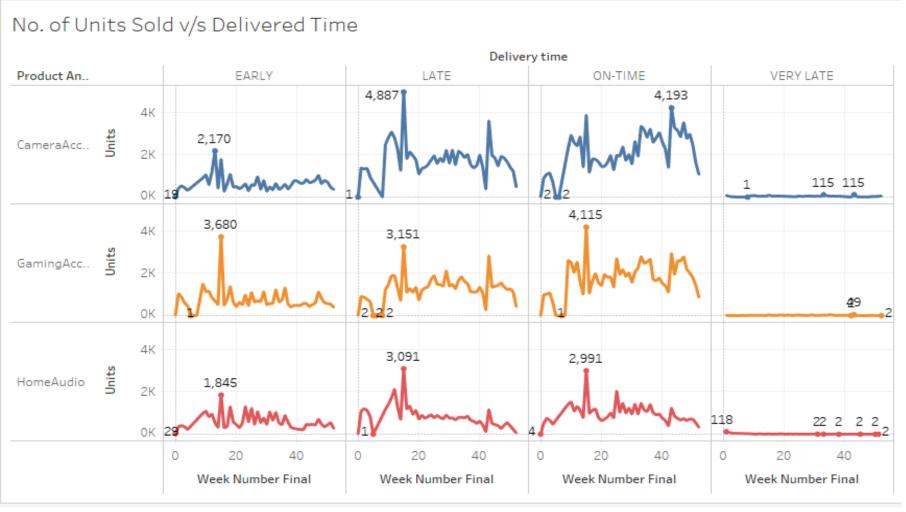
6. Perform EDA on each of the Dataframes

Important KPI's

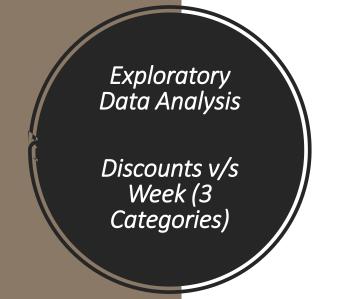


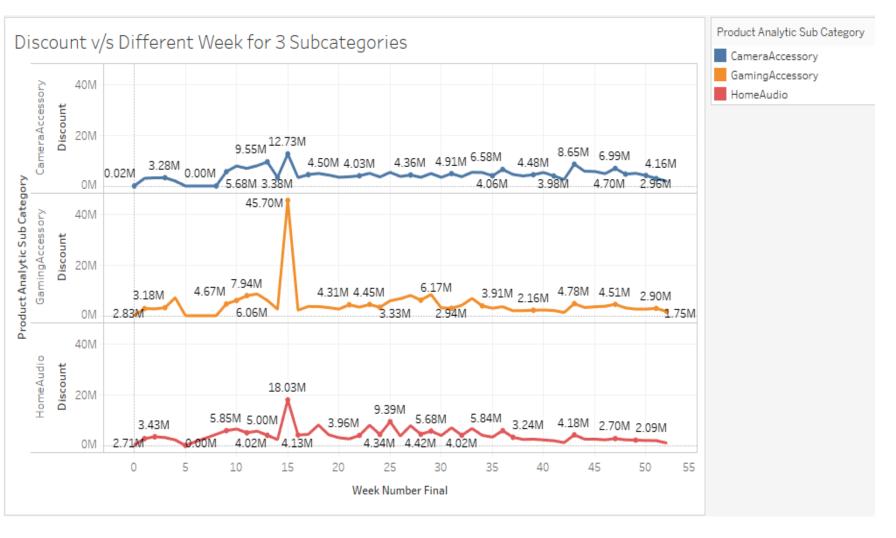
- 1. List Price
- 2. Discount
- 3. Holiday Flag
- 4. Pay Date flag
- 5. AdStock Value
- 6. Delivery Status (Early, On-Time, Late, Very Late)
- 7. Investment Percentage (per Week)
- 8. Average MRP
- 9. Average GMV
- **10.** Average Units



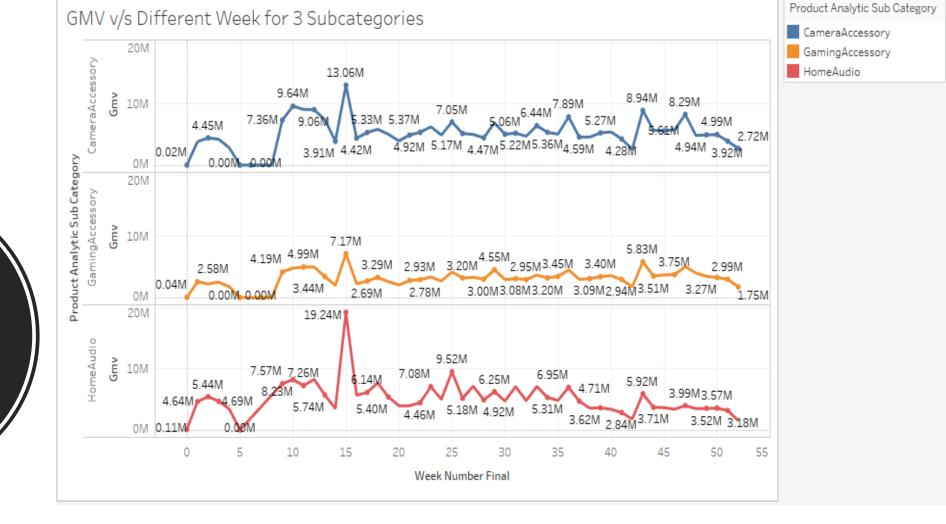


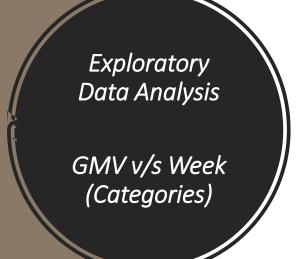
- Across different categories, most of the orders have been delivered 'LATE'
- Most order take 7 to 15 days to deliver
- Need to decrease the SLA lower than to attract more customers



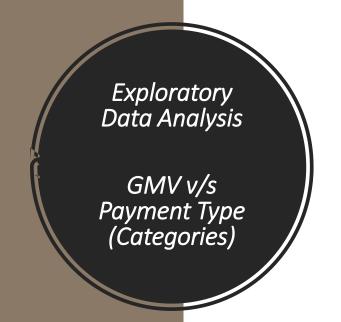


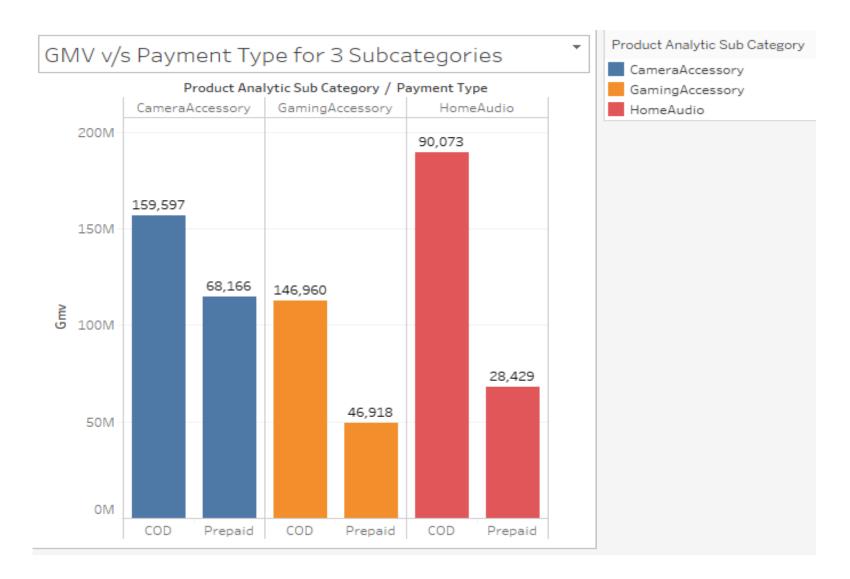
- Heavy discounts are offered in Week 15
- These are the weeks where Promotions are provided and also Holiday Weeks
- Gaming Accessories see a huge discount rise



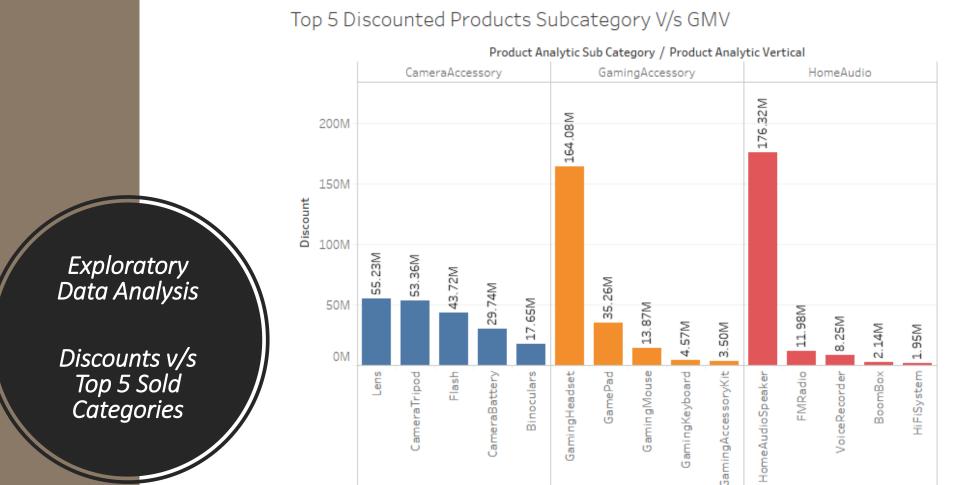


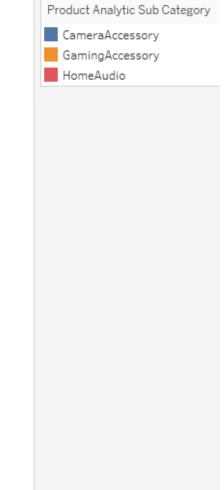
- Heavy Sales numbers are present for Week 15
- These are the weeks where Promotions are provided and also Holiday Weeks
- Home Audio have clear Spike of Sales during Holiday Weeks



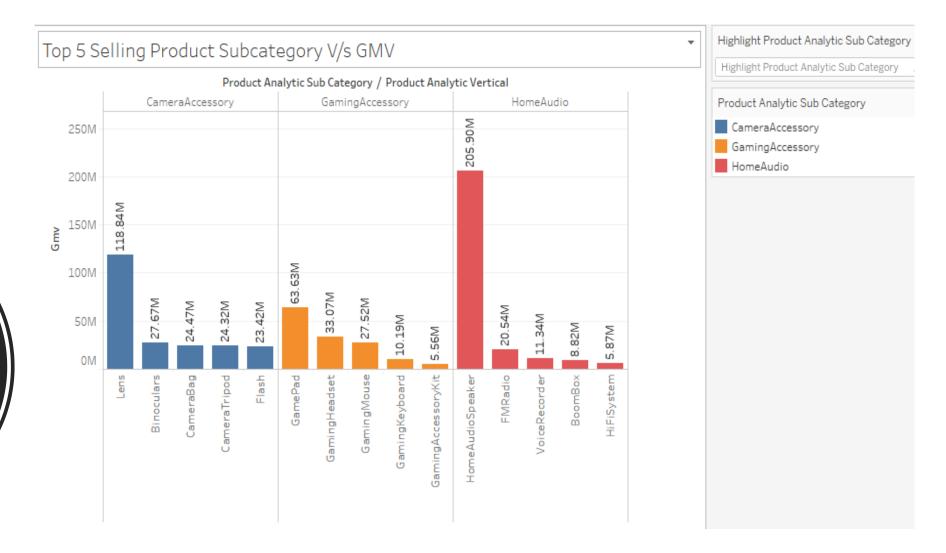


- Cash on Delivery(COD) is most preferred Payment Method.
- All Categories buyers opt for COD.
- Need to address the issue why prepaid is used by less users





- In Camera Accessory, Lens is the most Discounted item
- In Gaming Accessory, Gaming Headset is the most Discounted Item
- In Home Audio, HomeAudioSpeaker is the most Discounted Item
- Need to check why other items have been given less discounts



• In Camera Accessory, Lens is the most sold item

Exploratory

Data Analysis

GMV v/s Top 5

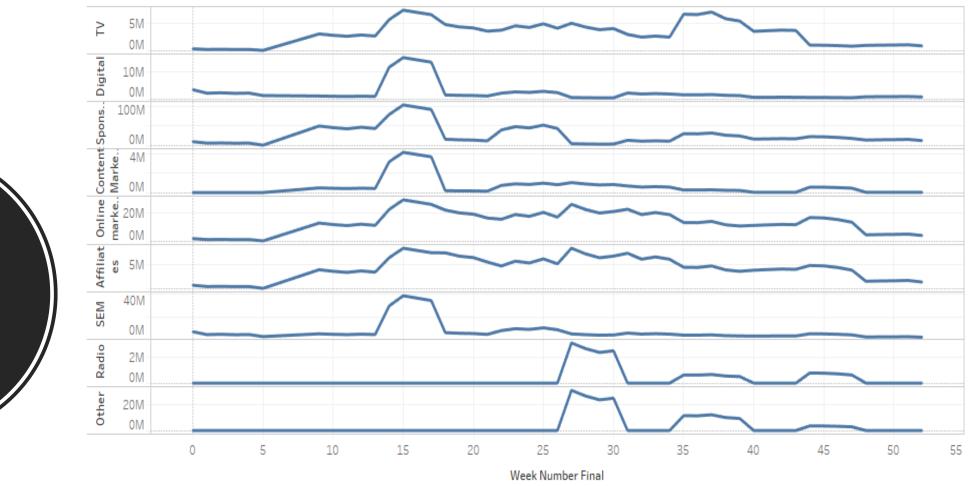
Sold Categories

- In Gaming Accessory, Gamepad is the most Sold Item
- In Home Audio, HomeAudioSpeaker is the most Sold Item
- Need to check why other item sales have been less



Exploratory
Data Analysis
GMV v/s MRP

- MRP takes over GMV in all categories, which shows clear losses incurred
- In Gaming Accessory for week 15, MRP is very high but company has offered lot of discounts which resulted in loss
- In Home Audio for week 15, MRP is very high but company has offered lot of discounts which resulted in loss
- For Home Audio and other categories, there is no Sales between week 5 and 9, need to address that

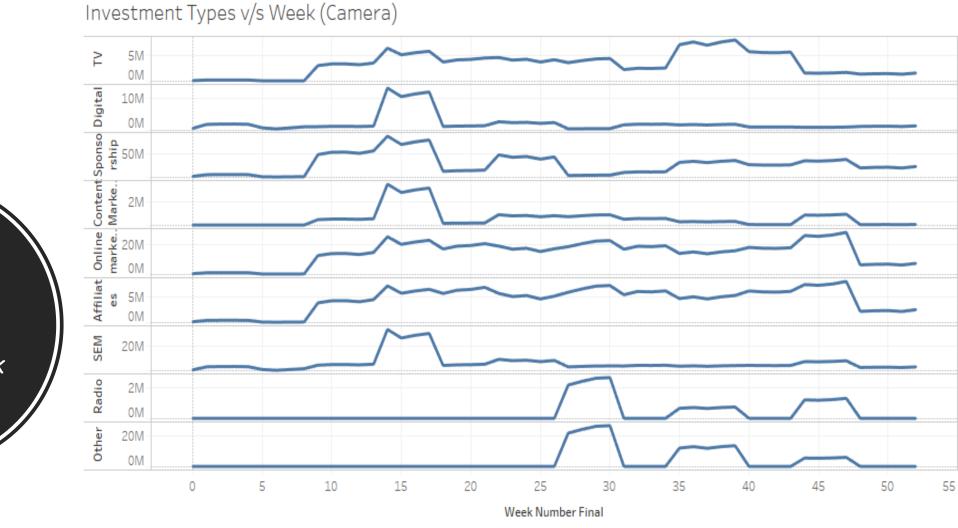


- We can see most of the Investment Spends have been done in week 13 to 15
- This is a Promotional Week where Majority Holidays are Present
- Sponsorship has the highest investment in this sector

Investment Spends v/s Weeks (Home Audio)

Exploratory Data Analysis

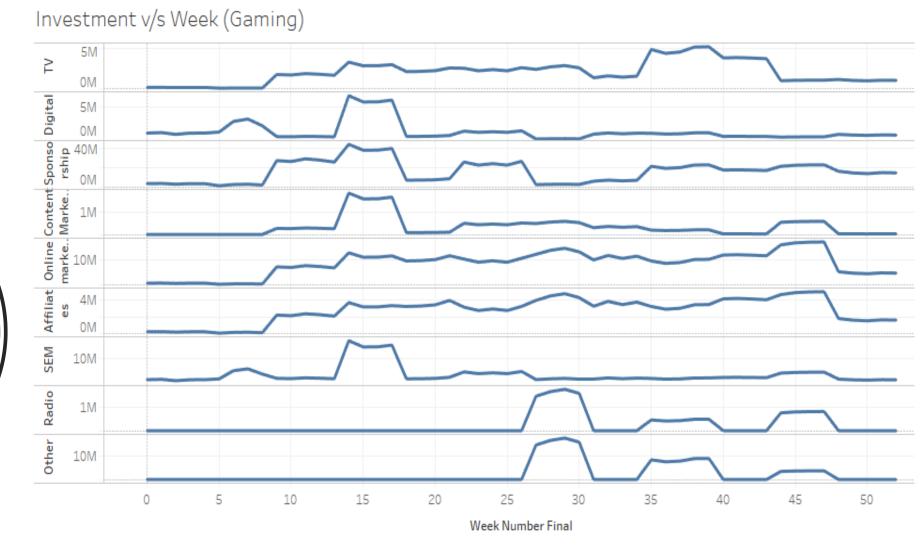
Investment Spend v/s Week (Home Audio)



- We can see most of the Investment Spends have been done in week 13 to 15
- This is a Promotional Week where Majority Holidays are Present
- Sponsorship has the highest investment in this sector

Exploratory Data Analysis

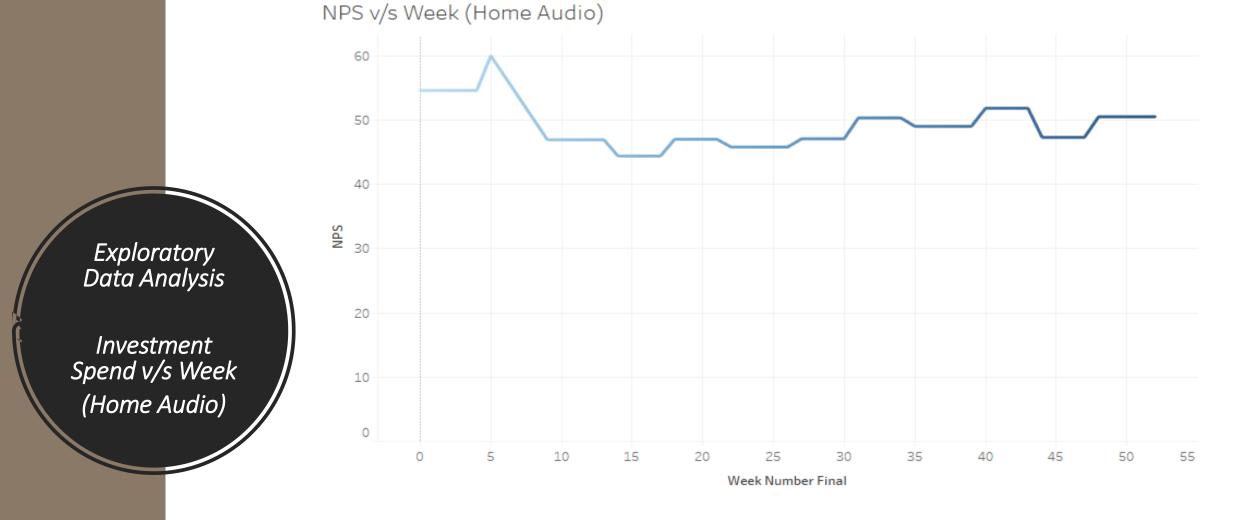
Investment Spend v/s Week (Camera)



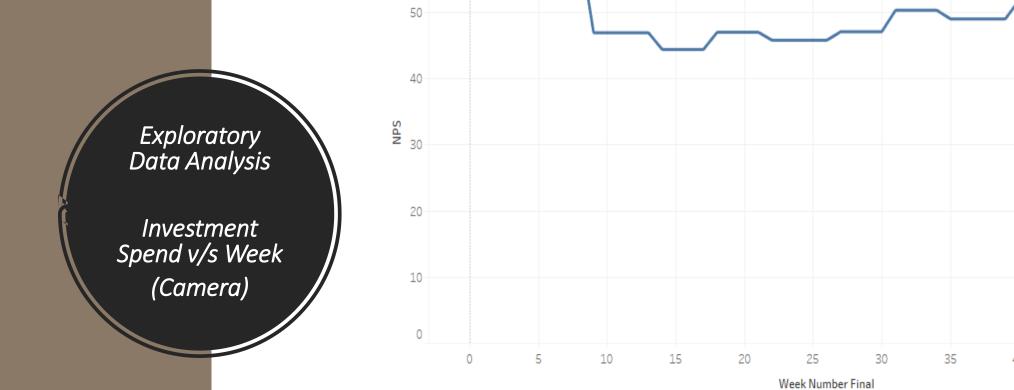
- We can see most of the Investment Spends have been done in week 13 to 16
- This is a Promotional Week where Majority Holidays are Present
- Sponsorship has the highest investment in this sector

Exploratory Data Analysis

Investment Spend v/s Week (Gaming)



- Week Number 5 has the highest NPS Score
- The Score Dips after week 5.
- Need to do the analysis behind the spike and fall



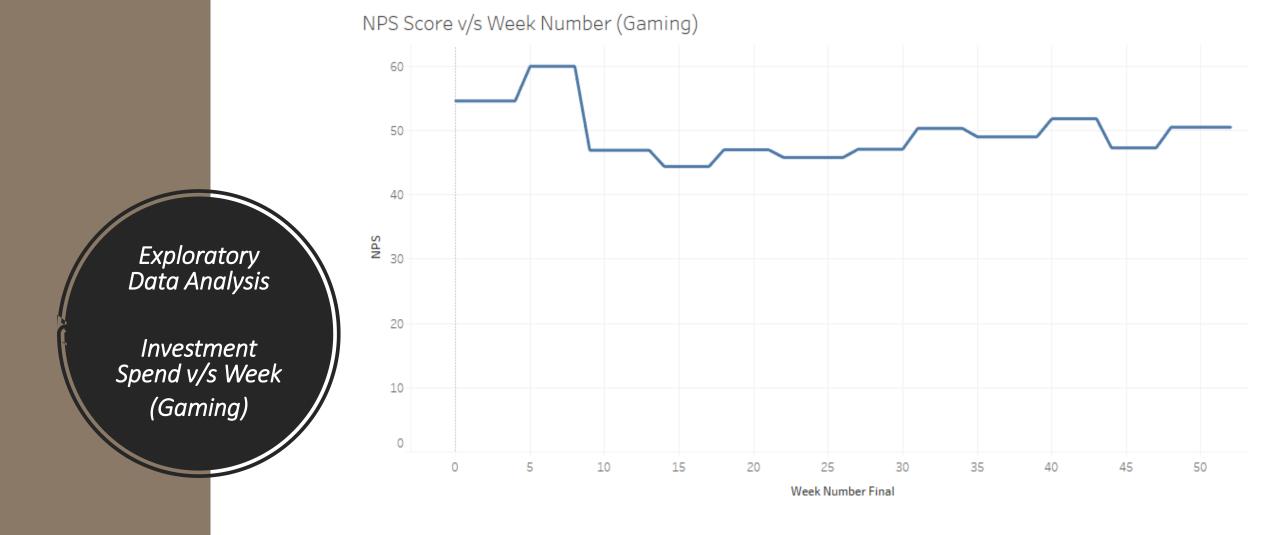
NPS v/s Week Number (Camera)

60

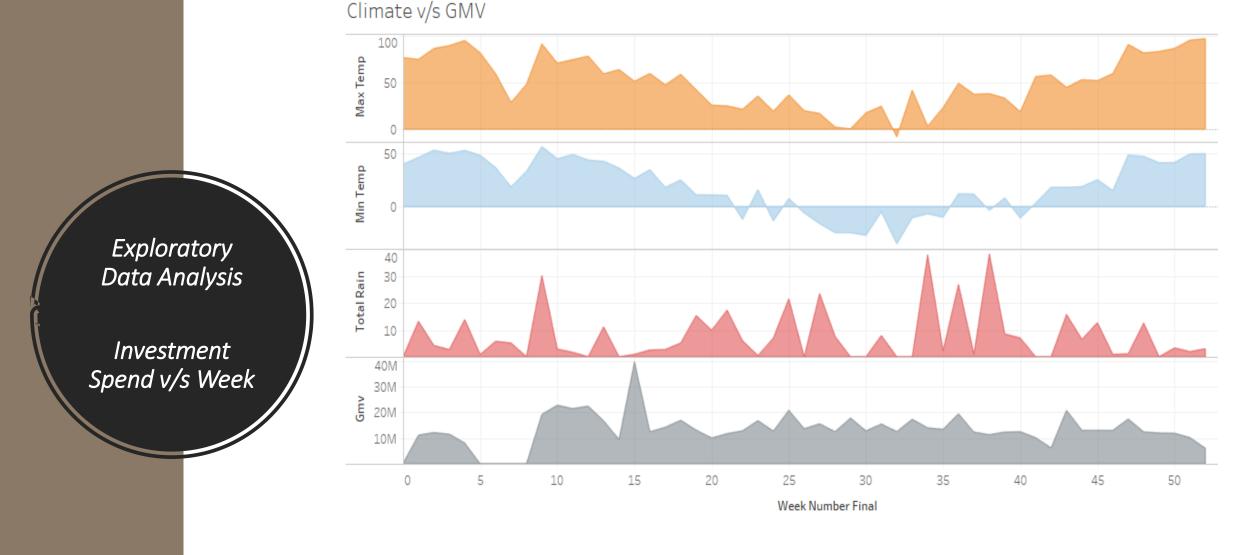
Week Number 5 to 8 has the highest NPS Score

45

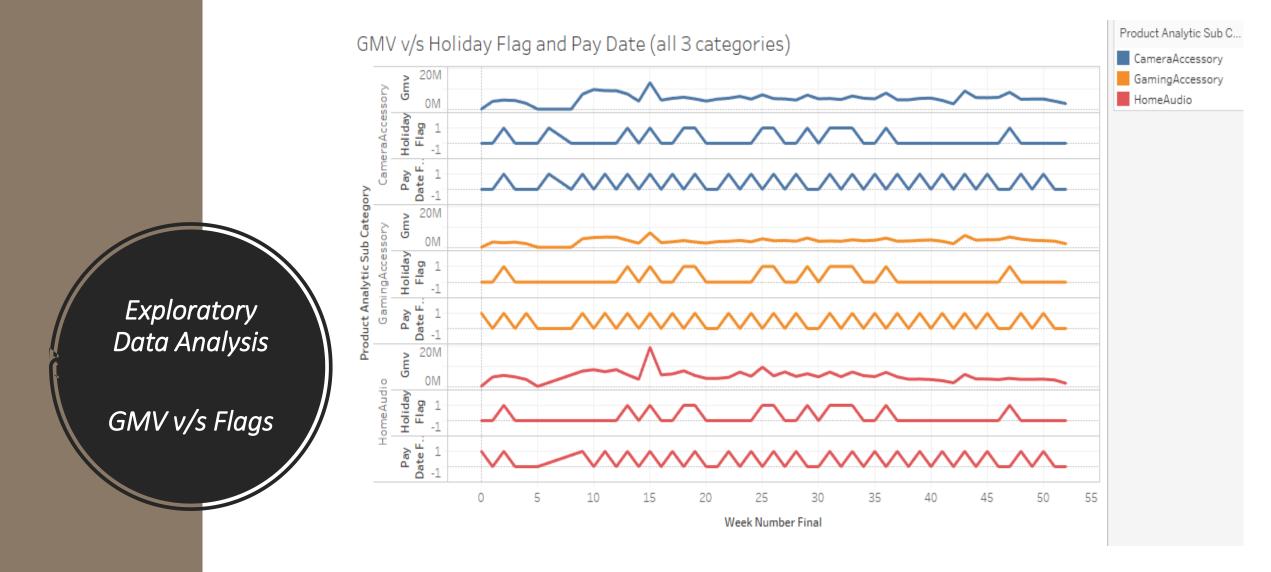
- The Score Dips after week 8.
- Need to do the analysis behind the spike and fall



- Week Number 5 to 8 has the highest NPS Score
- The Score Dips after week 8.
- Need to do the analysis behind the spike and fall



- Less Rainfall may be a factor for More Sales as per week 14,15, conversely more rainfall leads to less sales for all 3 categories combined.
- When Temperature is in a Mid Range (10-15 degrees), we can see more sales happening



- When Holiday Flag is 1, we can see good spike in Sales for 3 all Categories combined.
- When Pay Date is on 1st or 15th, Flag is 1. Hence we see a good spike in Sales
- Hence Holiday Week and Pay Date are very important parameters for good sales



Model Building (Linear Regression)

Home Audio

Important Features:

Online Marketing, Total Investment, Sponsorship, List Price

R-Score Value: 0.904

Inference:

Online Marketing, Total Investment, Sponsorship have a direct relation with GMV

Digital, Delivery_on_time variables have negative co-relation with GMV



Model Building (Linear Regression)

Camera Accessory

Important Features:

Affiliates, Digitial, SEM, Other, Sponsorship, units, list_price

R-Score Value: 0.839

Inference:

Affiliates, Digital, SEM and other have a direct relation with GMV

Online Marketing and Payment_type_COD have a negative correlation with GMV



Model Building (Linear Regression)

Gaming Accessory

Important Features:

Affiliates, SEM, Sponsorship, list price

R-Score Value: 0.89

Inference:

Affiliates, SEM and Sposorship have a direct relation with GMV Payment_type_prepaid shows negative relation

List Price has a direct relationshop with GMV