ECOMMERCE CAPSTONE PROJECT
PGDDS-C08

ElecKart Ecommerce Firm

Prepared by:
Bindhu B
Neha Singh
Rahul Mishra
Shubha Agrawal

BUSINESS UNDERSTANDING

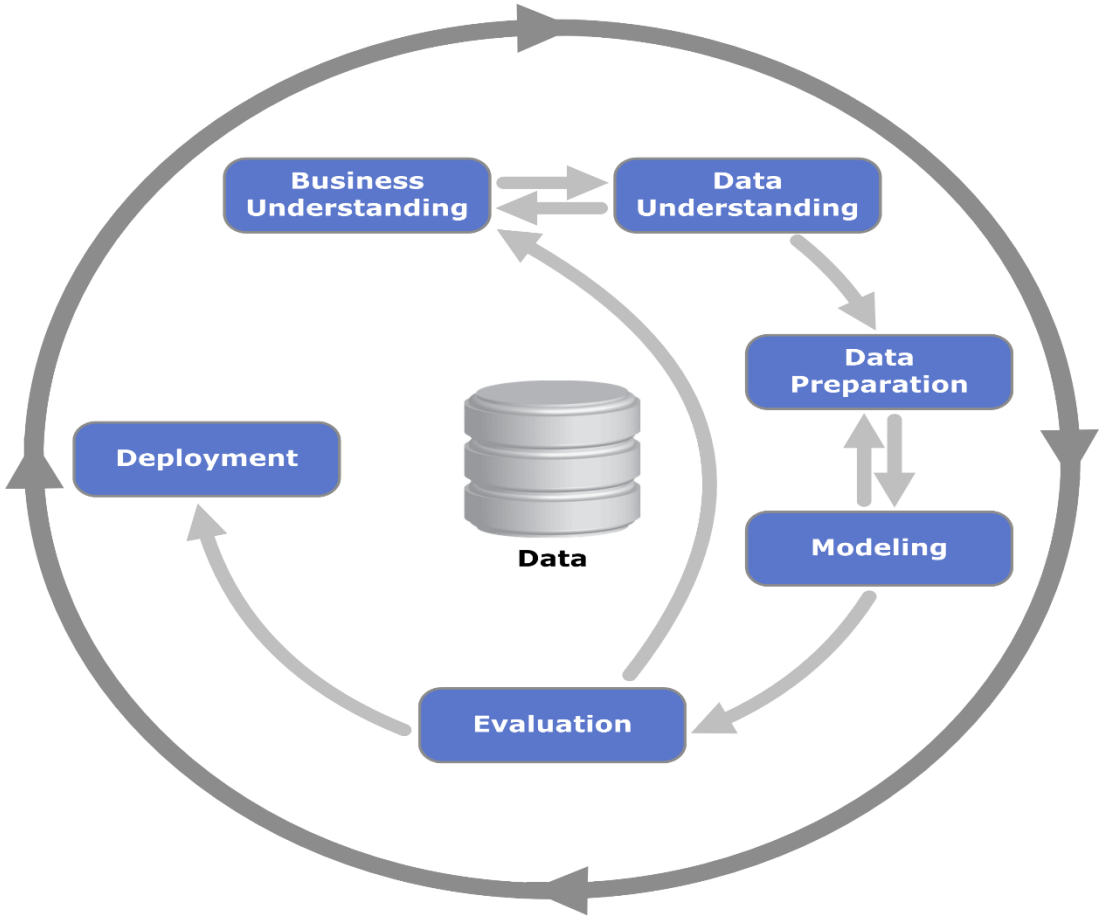
ElecKart is an e-commerce firm based out of Ontario, Canada specializing in electronic products. Over the last one year, they had spent a significant amount of money on marketing. They are about to create a marketing budget for the next year, which includes spending on commercials, online campaigns, and pricing & promotion strategies. We need to help ElecKart by churning the budget for last 12 months on marketing and identify the precise variables affecting budget allocation by predictive modelling.

BUSINESS OBJECTIVE:

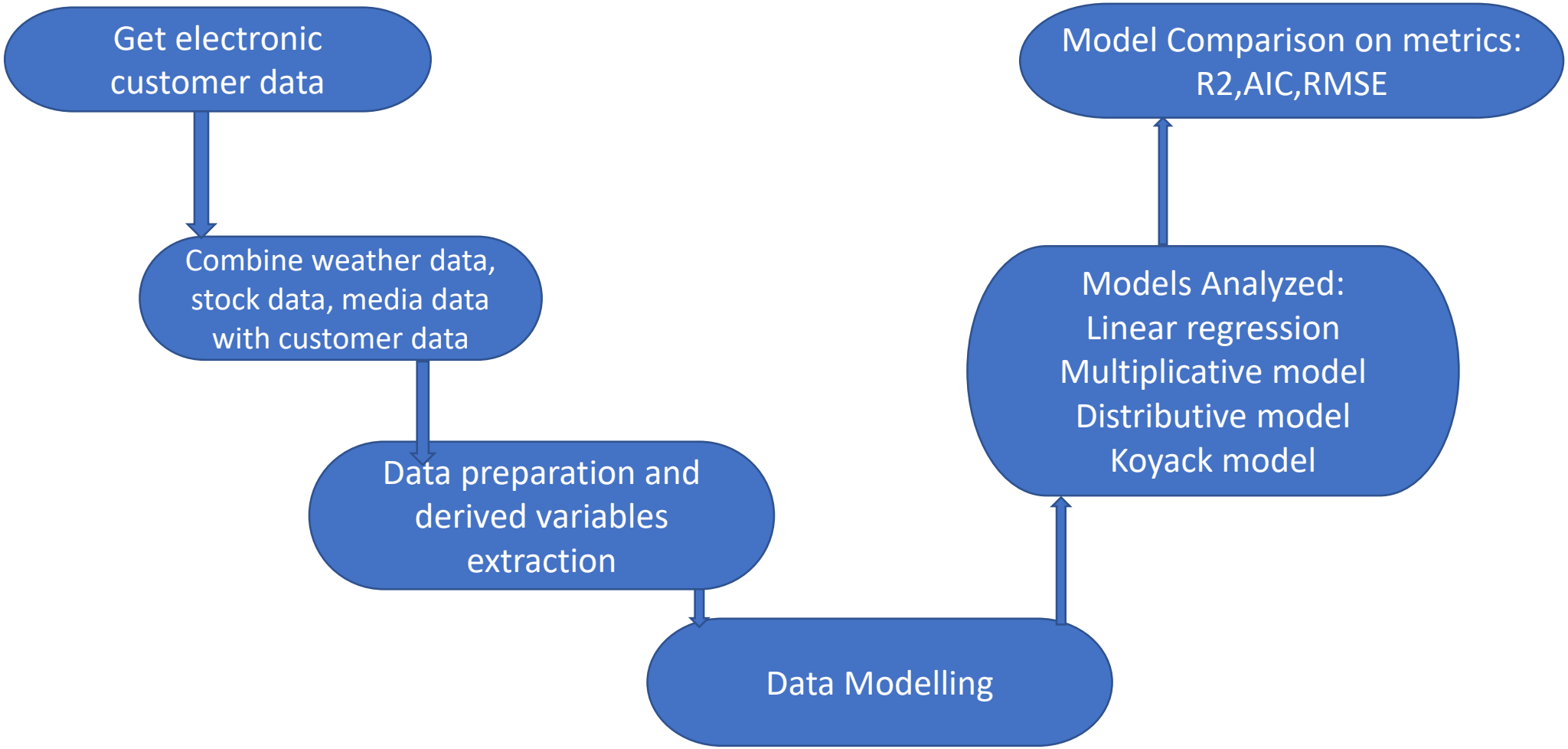
1.The aim of the analysis is to determine the driving factors affecting the gmv (Gross market value) for product categories: Camera, gaming and Home Audio

2.It is a case of typical Market mixed modelling (MMM) where we use multivariate regressions on Sales and marketing time series data to estimate marketing mix on sales and then forecast the future of it.

BUSINESS METHODOLOGY



PROJECT APPROACH



DATA UNDERSTANDING AND PREPARATION:

Data Given :

- **Customer Electronics Data**:
  This contains the product categories with there sold out units, revenue generated ,FSN ID count.
- **Media and communication data**: This contains the monthly spend on each marketing channel with its yearly percentage
- **Weather data**: It contains the attributes like total rain, total snow and weather impacting variables on the sales of the products
- **Stock Index data**: It contain the stock index of the company on monthly basis.
- New derived feature “Payday” is created which takes the value of 1 on dates 1 and 15 and on all other days

New derived features:

- “Special sale” is created which is made 1 on dates of special sales and on all other days
- “Week of the Year” is created which gives the week corresponding to the date
- “Payday” is created which takes the value of 1 on dates 1 and 15 and on all other days

After the data cleaning and deriving new column variables the final matrix structure contains 53 weeks data on all 3 products is as follows:
Camera : 49 x 23 ,Gaming : 49 x 23 , Home audio : 46 x 21

DATA ENGINEERING STEPS

Data was provided from July 2015 to June 2016. Following steps were followed under KPI engineering:

- **Data cleaning** :Null values, duplicates and data cleaning is performed. Features having more than 70 % null values are dropped

- **Dropped columns** : Columns with low variance or constant values – like units , deliverybdays, deliverycdays fsn_id,order_date,product_analytic_category, product_analytic_super_category, product_analytic_vertical,day, just_date, product_mrp and order_item_id are dropped from the data files

- **Dummy creation** is performed on categorical variable s1_fact_order_payment_type

- **Week level conversion** :All data is given to us in month level ,which is categorized in weeks for weekly analysis of the sales between 2015-2016 data.

- **Feature extraction**: From the provided variables ,features are derived such as Special sale, Week of the Year and Payday

- **Corelation matrix :** Heatmap is drawn of all the features to analyse the multicollinearity among them .There finalized graphs after dropping highly collinear graphs are attached in further slides for understanding.

FINAL ENGINEERED DERIVED LIST OF KPI's

After combining all the data files together, following KPI's are considered for the analysis.

#	KPI	Description
1	special sale	whether week is a sale week
2	pay day	whether week has pay day
3	pin code	count of distinct pincodes
4	customer ID	count of distinct customers
5	SLA	average days to deliver a product
6	Prepaid	Total number of prepaid delivery
7	COD	Total number of COD delivery
8	gmv	Revenue
9	week	week number
10	Total	Total spend on various advertising channels
11	TV	Total spend on TV
12	Digital	Total spend on digital
13	Sponsorship	Total spend on sponsorship
14	ContentM	Total spend on ContentM
15	OnlineM	Total spend on OnlineM
16	Affiliates	Total spend on Affiliates
17	SEM	Total spend on SEM
18	Radio	Total spend on Radio
19	Other	Total spend on Other
20	Mean Temp	Average temperature
21	Total Rain	Total amount of rain
22	NPS	NPS score
23	Stock Index	Stock index of the company
24	Total Snow	Total snow

Approach

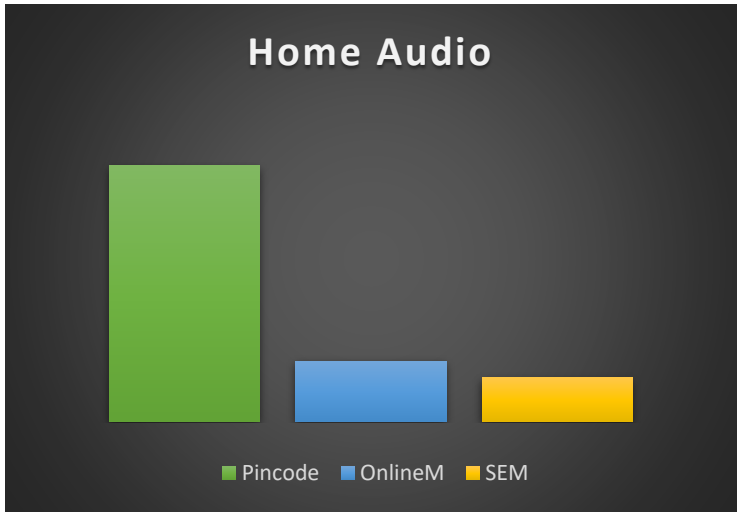
- To develop a market mix model to observe the actual impact of different marketing variables over the years – 2015 and 2016.
- With the given data spread across 4 different csv files- merge relevant columns from the files in to one csv file
- To recommend the optimal budget allocation for different marketing levers for the next year – we are building four regression models – Linear , Multiplicative , One week Lag and Two week lag and Kyock Models
- GMV – Gross Market Value is our target variable (Y-axis) and rest of the features in the dataframe are dependent variables(X-axis)
- We then compare the AIC(Akaike Information Criterion) ,R-Squared value , RMSE(Root Mean Square Error) and number of features in the final model for each category of products- Camera, Home Audio and Gaming to decide which model to go for which category.

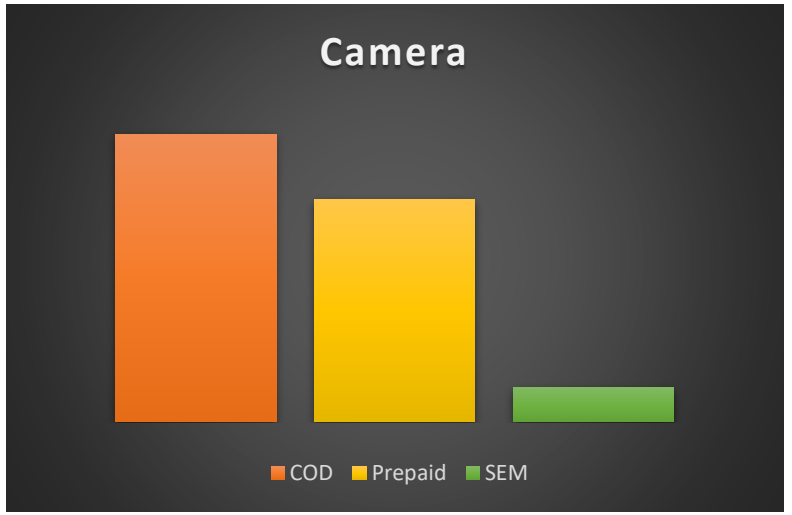
Results

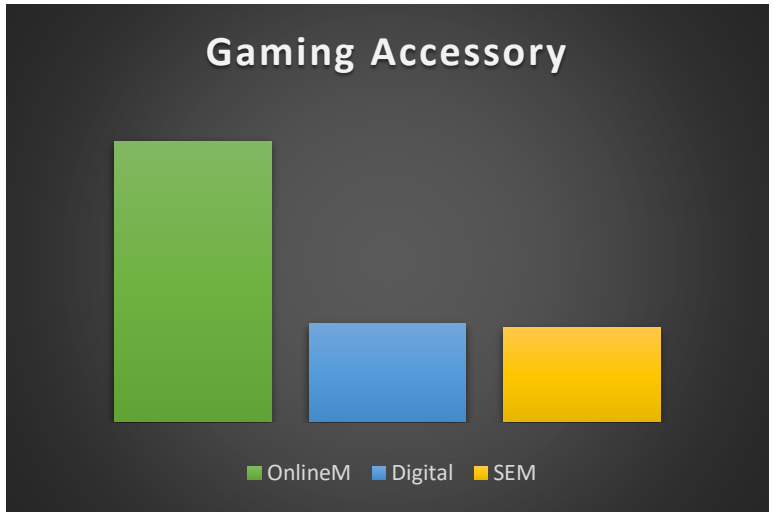
- **Camera**
  - **Linear Model** is best because AIC is minimum, RMSE is lowest and Adjusted R square is highest
- **Gaming Accessory**
  - Multiplicative model reports lowest AIC and highest R square. However, there are no media features. Therefore, second best model – **Kyock Model** with lowest RMSE is selected
- **Home Audio**
  - **Multiplicative model** is the best because AIC is minimum

Sub Category	Linear Model			Multiplicative Model			One week Lag			Two week Lag			Kyock Model		
	AIC	RMSE	Adj R	AIC	RMSE	Adj R	AIC	RMSE	Adj R	AIC	RMSE	Adj R	AIC	RMSE	Adj R
Camera	-169.00	0.03	0.99	-57.19	0.23	0.61	-145.90	0.23	0.97	-138.00	0.05	0.98	-128.00	0.05	0.96
Gaming	-110.00	0.05	0.95	-166.00	0.24	0.99	-43.00	0.24	0.63	-109.00	0.04	0.97	-129.00	0.04	0.97
Home Audio	-130.70	0.06	0.98	-166.30	0.09	0.99	-12.22	0.28	0.12	-110.00	0.08	0.98	-137.00	0.04	0.98

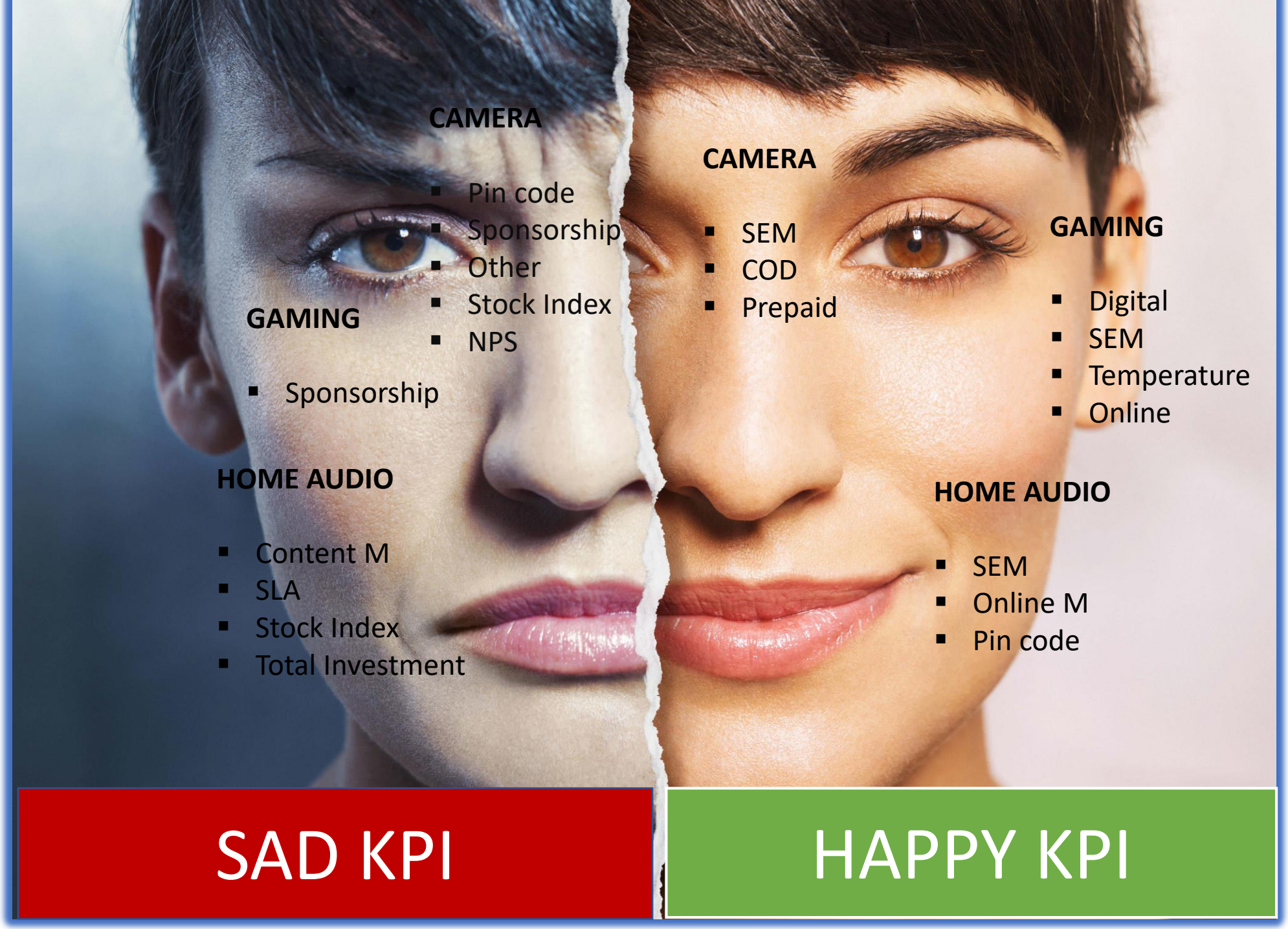
Top 3 Important KPIs per sub category







- Home Audio: Location is paying the most important role in sale of Home Audio followed by Online Marketing and SEM
- Camera: Both Cash on delivery and Prepaid are positive drivers which is interesting. But, COD is a clear winner. SEM (Search Engine Marketing) is a big driver for camera.
- Gaming Accessory: Sales can be given a push by focusing on online medium e.g. online, digital and search engine



Recommendations

1. **Future**

Focus must be on Search Engine Marketing as it is present as an important driver in all three sub categories. For the year 2015- 2016, investment on SEM was only 11%.Online Marketing is doing better for all three sub categories comparing to TV, Radio and Affiliates.

2. **Negative influencers**

Sponsorship is affecting sales negatively for both camera and gaming accessory. The average investment in Sponsorship for the year 2015-16 was 30.4 INR cr. This was 43% of total investment. Stock Index negatively impacts sales of camera and home audio. Net Promoter Score is negatively impacting camera

3. **Biggest Irony**

Location is pushing sales of home audio and pulling customers away from camera.

4. **Reallocation of Investment**

For the past year 44% of investment was towards negative influencers such as Sponsorship and Content Marketing. The investment must be directed towards online medium

Financial Implications

CAMERA :

In Linear Model, SEM is the only significant driver found. Using this model, business can easily focus on one and most important driver. Also business can pull back from investing in negative drivers. An increase of 1 unit in SEM will increase the revenue(gmv) by 0.17 unit. However, an increase of 1 unit in Sponsorship will decrease camera sale by 0.14 unit.

And an increase of 1 unit in Other investment will decrease camera sale by 0.13 unit.

HOME AUDIO :

In Multiplicative Model, SEM and Online Marketing are the two important drivers for sales apart from location (identified from pin code). An increase of 1 unit in SEM will increase the revenue(gmv) by 0.17 unit. An increase of 1 unit in Online Marketing will increase the revenue(gmv) by 0.23 unit.

Also, content Marketing does not work for home audio. An increase of 1 unit in Content Marketing will decrease the revenue(gmv) by 0.11 unit.

Financial Implications

GAMING ACCESSORY :

Kyock Model is chosen and the model provides a clear vision. People who buy gaming accessory cannot be swayed by Radio, TV or Sponsorship type. Focus must be online. Digital, SEM and Online Marketing works for sales in gaming accessory. An increase of 1 unit in SEM will increase the revenue(gmv) by 0.25 unit. An increase of 1 unit in Digital will increase the revenue(gmv) by 0.25 unit. An increase of 1 unit in Online Marketing will increase the revenue(gmv) by 0.72 unit.