A
Market Mix Model
on
Consumer Electronics

Mid-Term Capstone Submission

Specialization:	E-Commerce
Team Members:	Hari Krishna Eluri
	Vidya Kurada
Submission Date:	Monday, Dec 2, 2019

Business Understanding

- EleKart is an **e-Commerce firm based out of Canada specializing in consumer electronics**

- As part of customer reach, **they channel their marketing expenditure through various levers**. Few of the levers include
  - **Media Marketing** (Digital/TV/Print and So on)
  - **Discounts and Promotions**

- The Company wants to strategize their market spent for the next year (July 2016- June 17) by capturing insights from the previous year (July 2015- June 16)

- For that reason, **they plan to study the impact of market spent on revenue generated**

- They want to understand few aspects like
  - Whether to give a **budget rise or cut?**
  - How to **optimize allocation of budget to different marketing levers?**

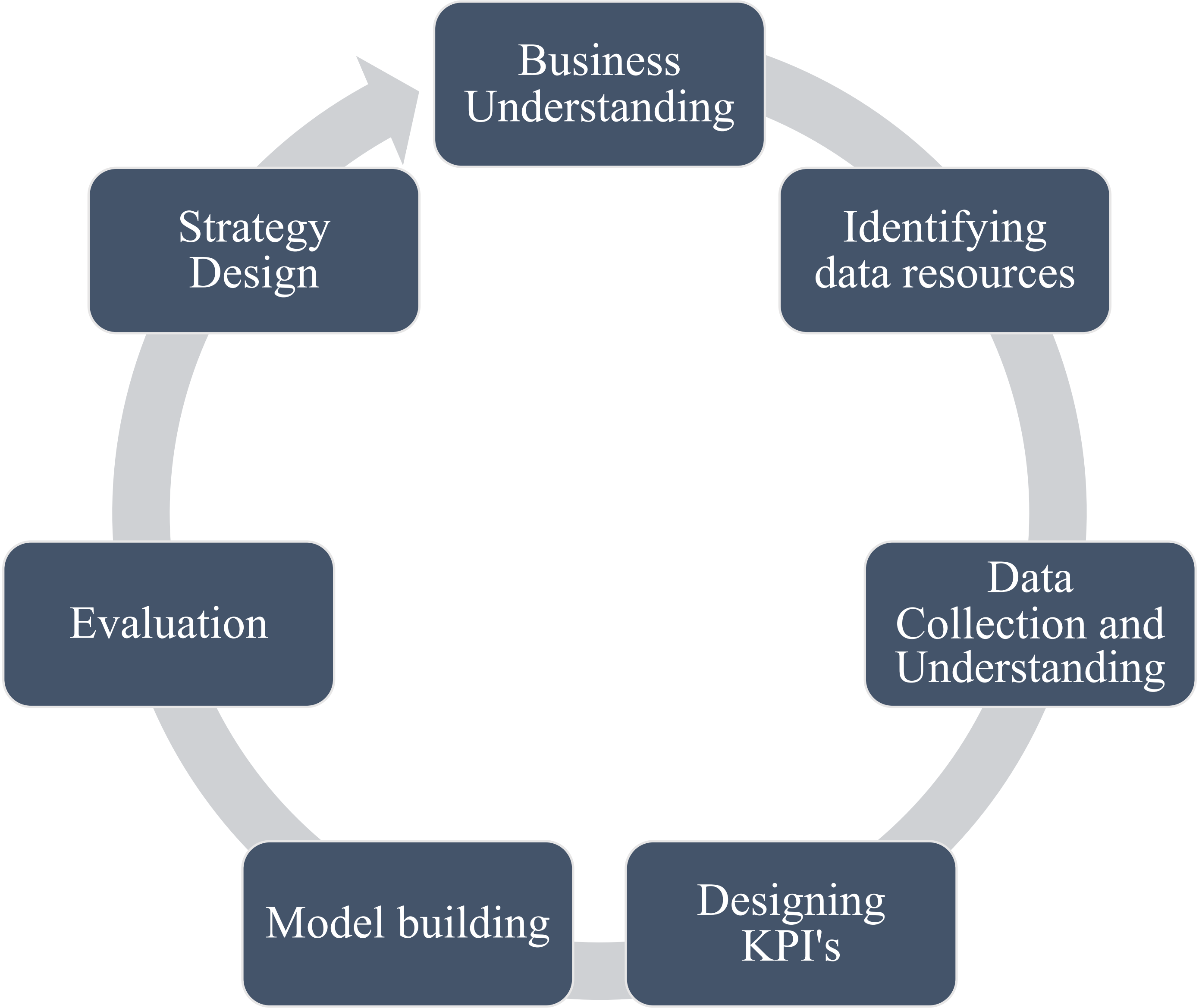
Objective

Develop a **market mix model to observe the actual impact of different marketing variables over the last year** and **recommend the optimal budget allocation for different marketing levers** for the next year.

Available Data and Expectations:

- Data is available from **July 2015 – June 2016**
- We are expected to **make the strategy recommendations for 3 product sub-categories**. They include:
  - ✓ *Camera Accessories*
  - ✓ *Gaming Accessories*
  - ✓ *Home Audio*
- Also, we are expected to **formulate the market mix models at a weekly level**
- The following slide gives a **higher level approach to address in the present project**

Approach: Crisp Analysis Framework



Internal & External Data Resources

- From the approach framework, **problem description (business understanding: what to address) and data are a given** for our current project. So, let us start discussing from data resources.

- For this analysis, it is advised in the problem statement to utilize the following data sources to construct KPI's (Key Performance Index)

- Data Resources include:
  - **Consumer Electronics Revenue Data** : To understand patterns in revenues from order level data
  - **Market Spent** : To understand the impact of different marketing levers on revenues
  - **NPS** (Net Promoter Scores): To understand the impact on revenue from consumer satisfaction
  - **Stock indices** : To understand growth impact on revenues
  - **Weather**: To understand the indirect impact on revenues through delayed deliveries and so on..

- In the following slide, a snapshot of raw schema is provided for quick user understanding

Raw Data & Details (Schema/Meta-Data):

Consumer Electronics Data Set	
Id's: are unique identifiers of products, order and customer details	fsn: Product SKU information Order Id/Order item ID or Cust Id's
Categories: are the headers under which company sells the product	Product Categories/Sub Categories Product Verticals
Service level agreements: are about Delivery or Procurement promises made	Sla and Procurement Sla
GMV & MRP: are revenue generated and MRP of products	gmv and product_mrp
Dates & Important days: are the order dates and days where promotions happened or pay days	Date of Order, Month, Year…. Pay day and Special sale day…..
Mislleneous: Any other information provided	Units sold, Payment Type and So on

Market Spent Investment
TV: Ad's on TV
Digital: Social Media interactions
Sponsorship: Events sponsored at various venues
Content Marketing: Blogs and videos
Online Marketing: Ad's spent online
Affiliates: Redirecting traffic from other websites
SEM: Spend on search engine optimization

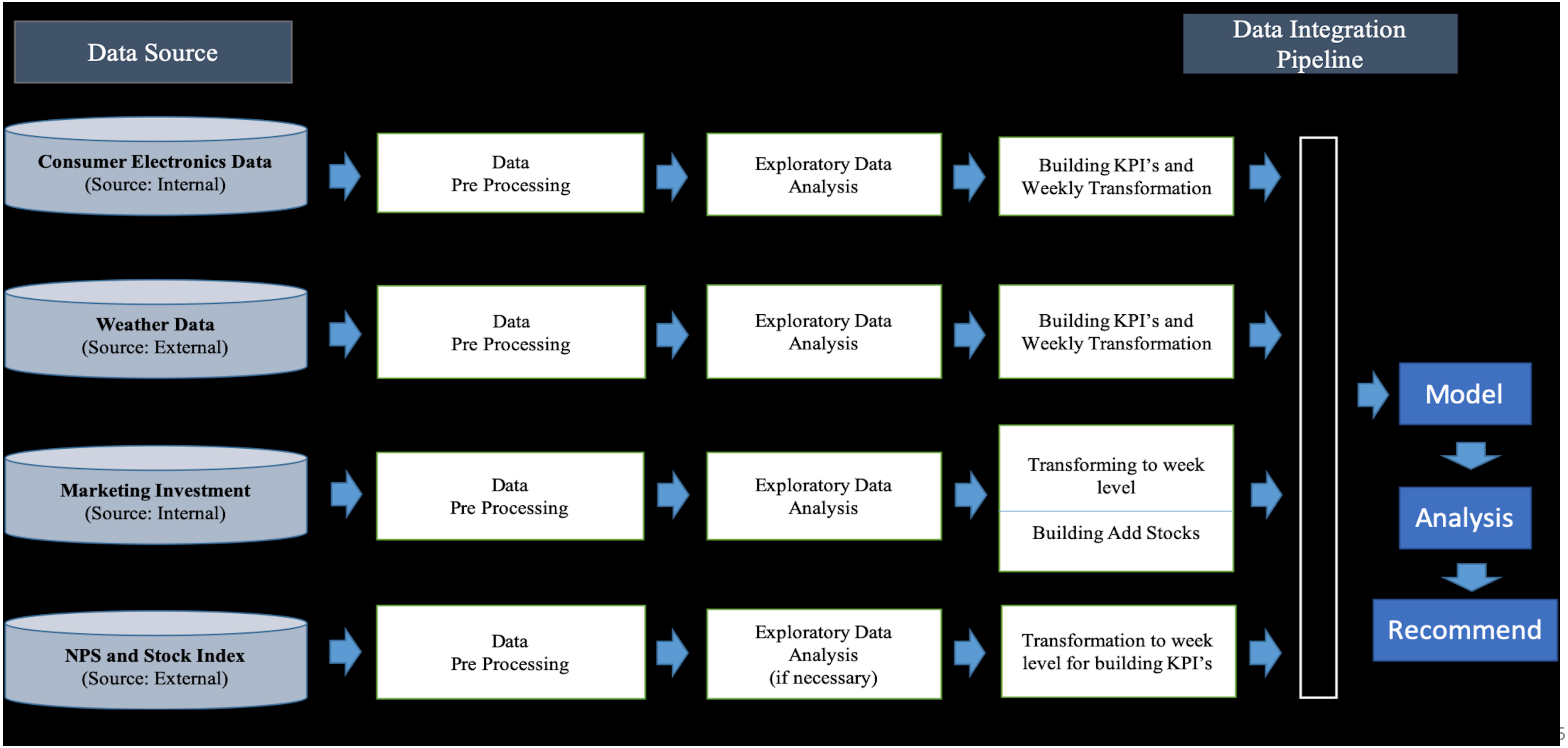
Weather Data
Temperature, Rain, Snow, Precipitation and Likes

Other Data
NPS (Net Promoter Scores): Derived Metrics from user ratings & **Stock Index**

Note:

- *The Schema is only representative.*
- *For Ease of understanding, we did not use the actual names from schema*

Approach Snap Shot



Data Pre-Processing

- In this case study, the data is given **raw as well as at different levels**
- Different **levels meaning daily and monthly**
- **Pre-Processing** happened **simultaneously with feature engineering**
- Few of the **essential steps of data pre processing are:**

Basic Data Pre-Processing

- Basic Data Type Conversions and Date Conversions
- Removing garbage values and Imputing rest of null values
- Treating Outliers

Transforming Data to Weekly Level

- Converting Daily level and Order level Data to Weekly Level (Part of building KPI's too)
- Converting monthly level data to weekly level

Market Investment Distribution:

- A monthly level total market investments and distribution across levers was given
- This is converted to add stocks to extract KPI's including weighted lag spend extraction

Exploratory Data Analysis

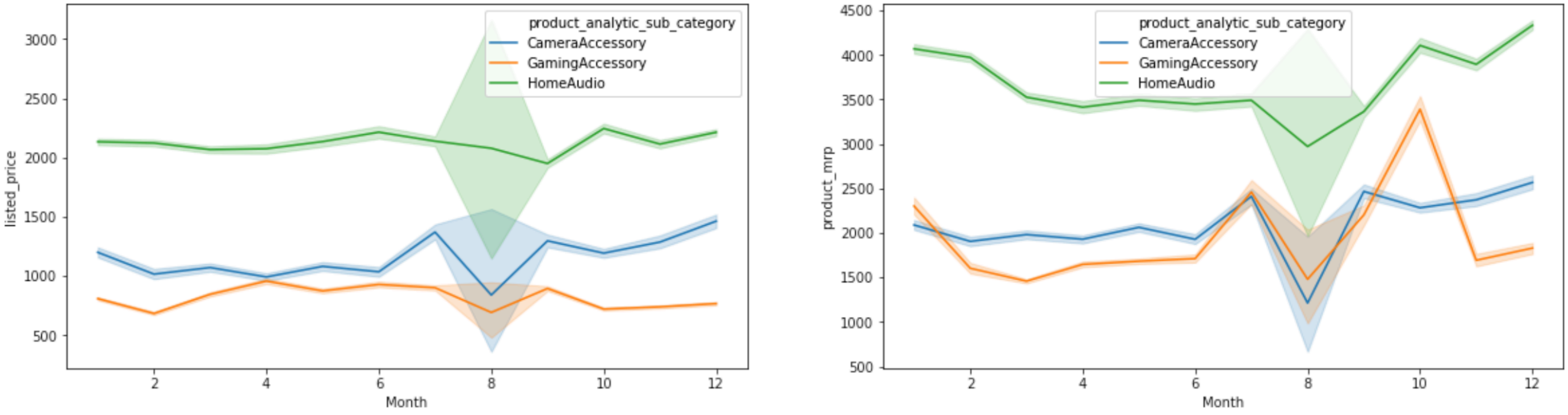
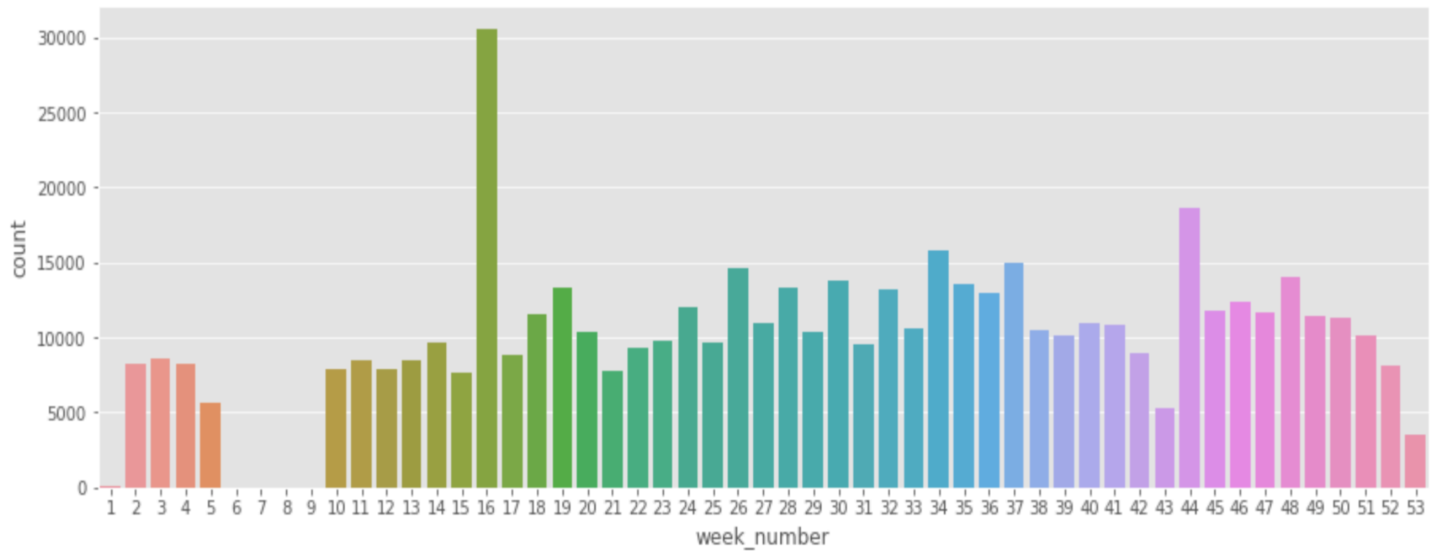


Figure: 1. Listed Price Vs Month, 2. MRP Vs Month

- The graphs give an idea on **listed price vs MRP ranges over months**
- Evidently, **listed price is always less than the MRP** in range
- This indicates us to **formulate a KPI using discounts**

Exploratory Data Analysis

UpGrad



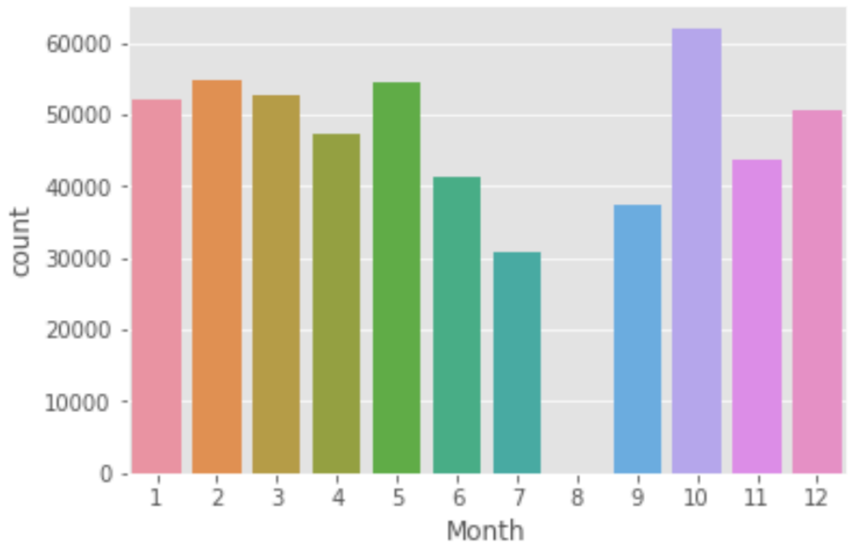
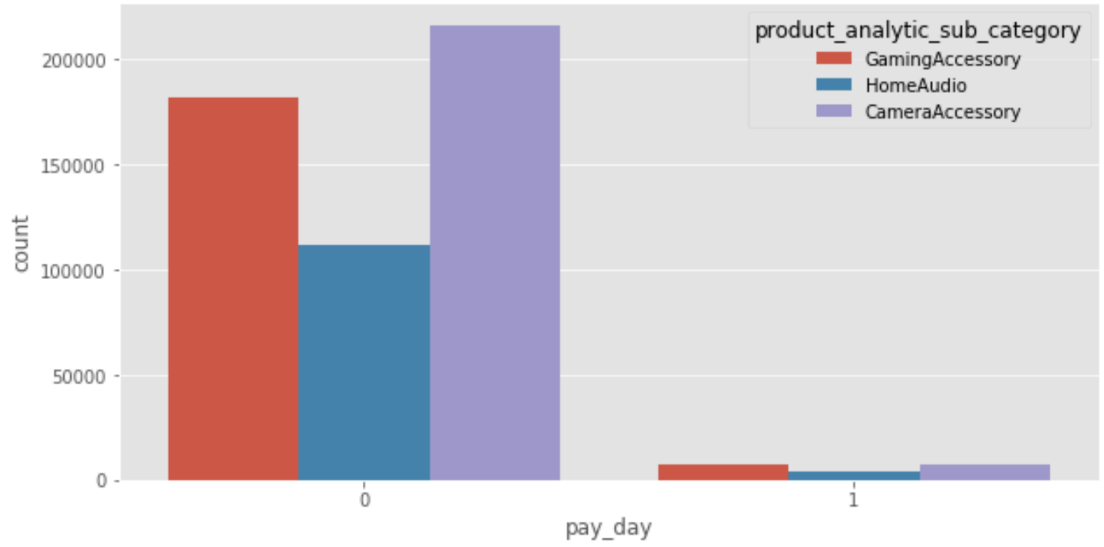


Figure: 3. Week Vs Order Count 4. Month Vs Count

- These graphs primarily indicate a dip in orders in 6-9 weeks or 8th (August) Month
- We observed that even **media investments are low in this month**
- It is interesting to observe **whether it's a cause effect of low campaign – low return** or whether factors like delayed delivery due to seasonal differences had a more natural impact on revenues. **Thus delivery status and weather are considered potential KPI's**

Exploratory Data Analysis



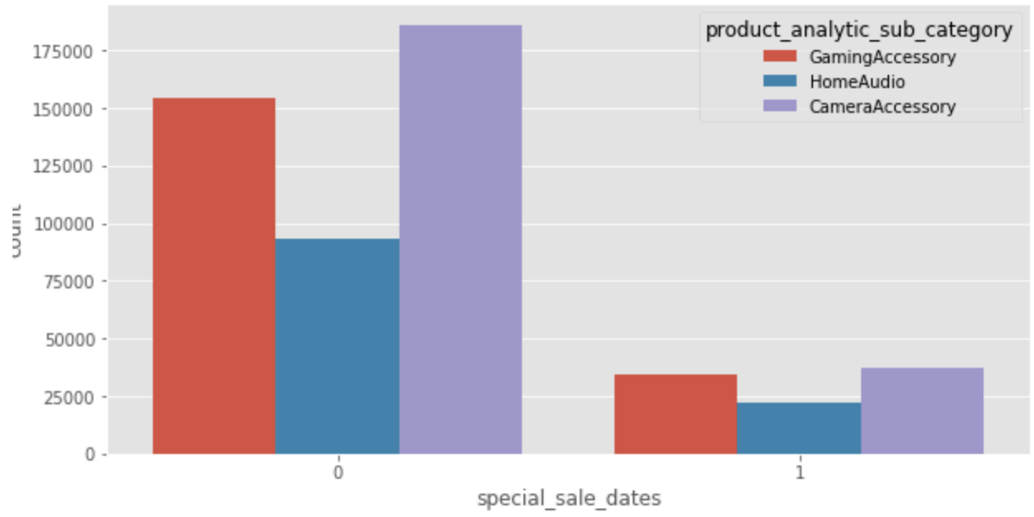


Figure: 5. Pay Days Vs Order Count 6. Special Sale Days Vs Order Count

- These graphs are to be taken with a pinch of salt. While pay days are 2 per month and special days are 15 per year, the orders from other days looks like dominating the revenue.

- But we should see that, even visually we can say pay days (24 per year) looks like sharing at least 5-10% of total revenue of year. Similarly, Special Sale days cater to a range of 10-20%

- Thus confirming quantifying the impact will potentially explain the revenue. So **pay day and special day are added to the KPI bucket**

Exploratory Data Analysis

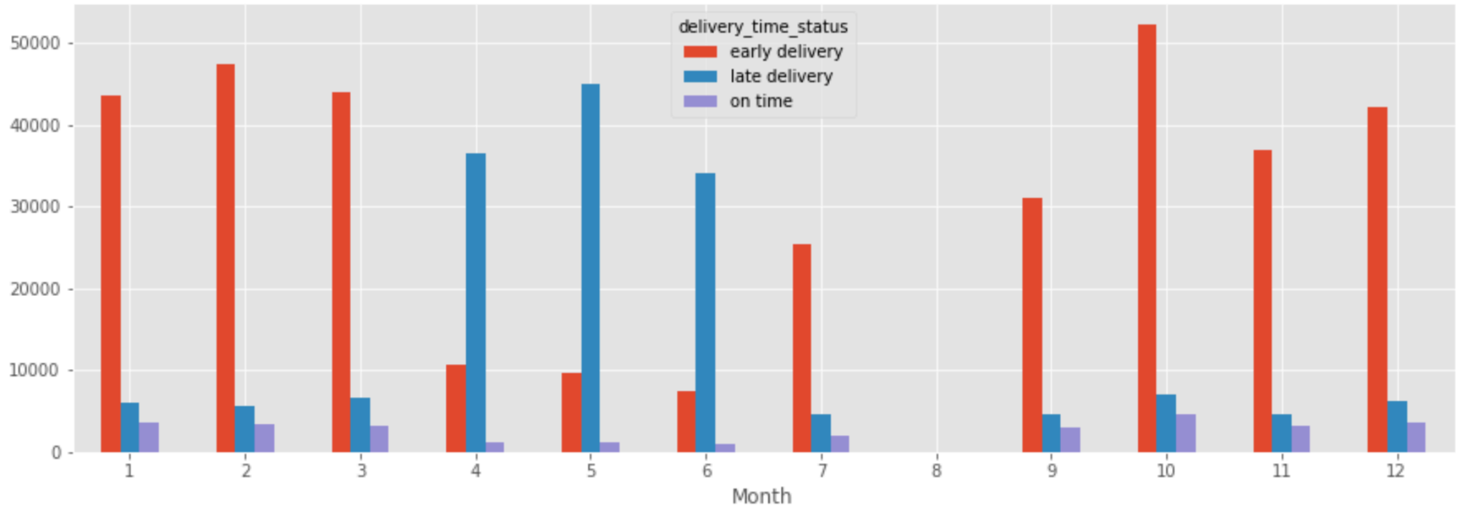


Figure: 7. Month Vs Delivery Status

- This is one plot discovered after formulating a delivery status KPI.
- Previously we had a discussion on 8th month having very less sales and we plan to formulate external influencing KPI's other than market share.
- For this reason, we **formulated delivery status**. It is evident that, **late delivery rate over 4-6 months had a lagged effect on the dip in sales in 8th month**. It became a **customer trust issue**.
- This EDA correspond to understanding data and designing KPI's at order level

Exploratory Data Analysis

- Data provided for weather is an extremely sparse (non continuous) in nature. Thus allowing no scope to perform any weather study.

- For that reason, we ignored **exploratory data analysis on weather**

- Similarly, **NPS and Stock index are constant values provided month wise with little or no change over year.**

- While we **want to see their effect on revenue through models**, those **data are not meant for analysis w.r.t this problem statement**.

- Finally a monthly wise and lever wise market spend information is provided for us. To study the effect, we have gone through extensive add stock computation process to capture lag effect as much as possible.

- We have considered **deterministic approach as per industry standards to compute add stocks.**

- In the following slides, we **added add stock computation methodology too.**

Categorizing Designed KPI's (Feature Engineering)

- As part of process, we designed multiple KPI's at different levels of processing making it indistinguishable from data pre-processing

- Intuitively, we categorized KPI's into 3 categories:

Order Level KPI's

- **Week Number** as Index
- **Discount** on MRP
- **Status KPI's**
  - Payment Type
  - Delivery Status

- **Day indicator KPI's**
  - Special Sale Day
  - Pay Day

Market Spend KPI's

- **Add Stock for Each Media**
  - TV
  - Digital Media
  - Sponsorship
  - Content
  - Online
  - Affiliates
  - SEM

Other KPI's

- **Weather KPI's**
  - Temperature
  - Precipitation
  - Rain
  - Snow

- **NPS**
- **Stock Index**

Note: *Details on assumptions and conversion mechanics of add stock KPI's are discussed in next slide*

Add Stock KPI's Design: Assumptions & Calculations

- **Premise:** Marketing Investment vs Revenue has no direct one on one solution. This week's marketing might have
  - An effect on sales only in upcoming week revenues
  - might have depreciating but existing effect on future sales including

- Thus, add stocks are a terminology for KPI's that indicate the accumulative effect of past and present marketing investment on the revenues

- Typically 2 factors are assumed for computing KPI's
  - **Lag**: Explains **how much time today's campaign lasts in future**. *Ex: 6 weeks or 7 weeks*
  - **Lag Weights**: Explains **Proportion of effect the campaign is still holding**. *Ex. TV add stocks has 60% effect after one week*

- Considering the importance of these add stocks, we have taken industry standard assumptions on these

Add Stock KPI's Design: Assumptions & Calculations

- Assumptions include: Industry Standard lags and Weights suggested in discussions

Media	Lag taken	1	2	3	4	5	6	7	8
TV	5	60.0%	36.0%	21.600%	13.0%	7.8%	0	0	0
Digital	6	43.0%	18.5%	7.951%	3.419%	1.470%	0.632%		
Sponsorship	6	31.0%	9.6%	2.979%	0.924%	0.286%	0.089%		
Content Marketing	8	31.0%	9.6%	2.979%	0.924%	0.286%	0.089%	0.028%	0.009%
Online marketing	4	43.0%	18.5%	7.951%	3.419%	0	0	0	0
Affiliates	4	31.0%	9.6%	2.979%	0.924%	0	0	0	0
SEM	4	42.0%	17.6%	7.409%	3.112%				

- The following steps explain the computation procedures
  1. Take all media heads/levers and their market spends independently
  2. For each of them, **compute the weighted lag features**
  3. Add all the lag influence to the original to compute an add stock

- The weights and lags are as mentioned in the above assumption

- *Ex: For TV: Spend is 10 Cr for 5 weeks, at the end of 5th week, add stock is 20 Cr approx*

Model: Linear Regression

- As a starting point, we **estimated linear regression models for all 3 sub categories**
- We present the **R-Squared scores, Plots and Other details** here

Sub Category	Train R-Score	Test R-Score
Home Audio	0.997	0.962
Camera Accessory	0.966	0.665
Gaming Accessory	0.941	0.634

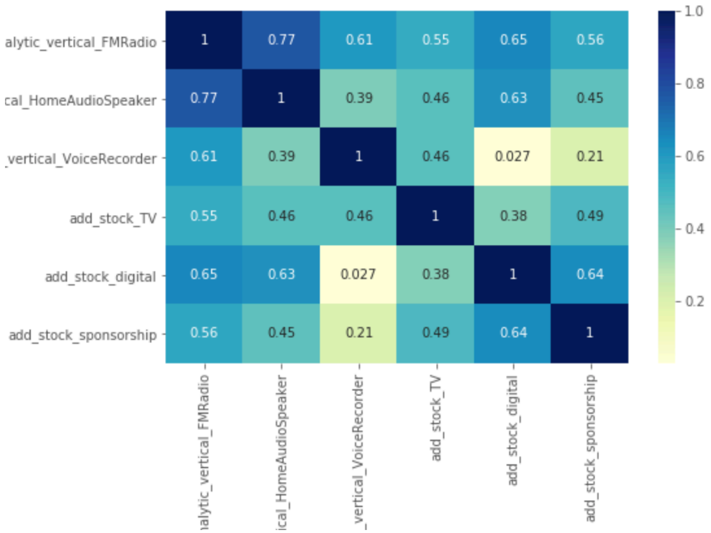
- These are the consolidated results for all the category models
- At one shot we can observe from train, test scores that
  - **Home Audio is a very good fit**
  - **Camera Accessory** is demonstrating an **over fit**
  - **Gaming accessories** also exhibiting **over fit**
- Let us discuss each one of them in detail in follow up slides

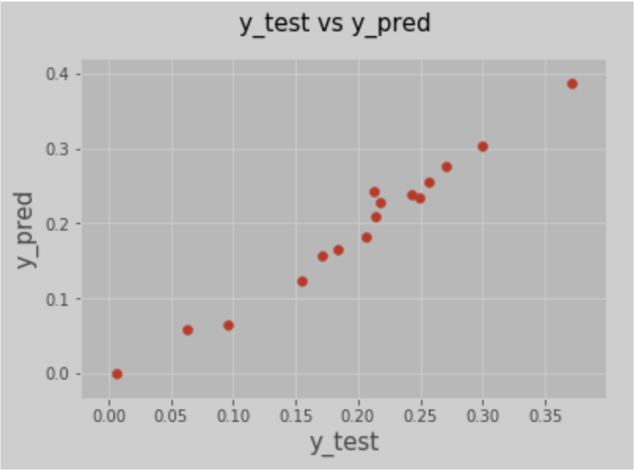
Model: Linear Regression- Home Audio

- As a starting point, we **estimated linear regression models for all 3 sub categories**

- We present the R-Squared scores, Plots and Other details here

Home Audio	Train: 0.997	Test: 0.962





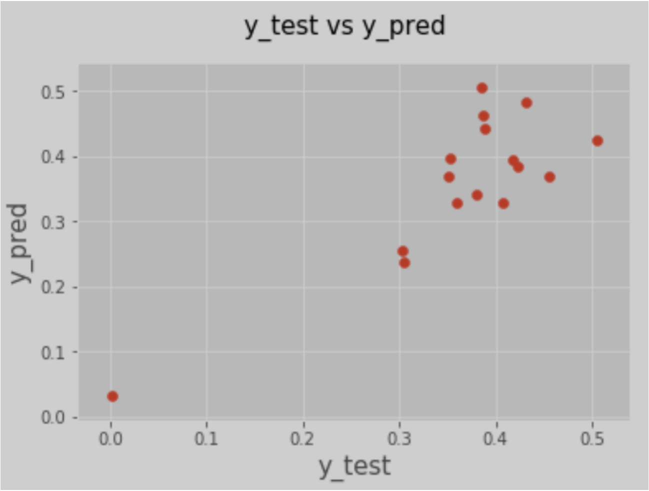
- We can see that, **Home Audio has a very good linear regression model**
- Error terms are clearly white noise – Normally distributed (0, sigma)
- Even multi collinearity effects handled decently
- **Add Stocks features influenced largely in estimations**
- **KPI's** used are enlisted in the last

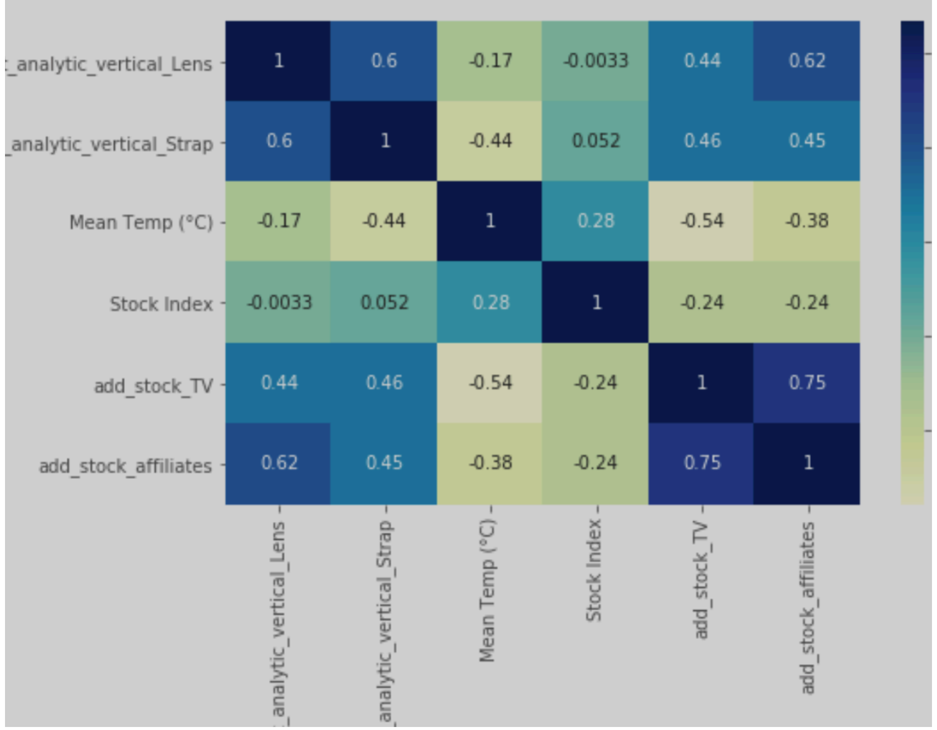
Model: Linear Regression- Camera Accessory

- As a starting point, we **estimated linear regression models for all 3 sub categories**

- We present the R-Squared scores, Plots and Other details here

Camera Accessory	**Train : 0.966**	**Test : 0.665**



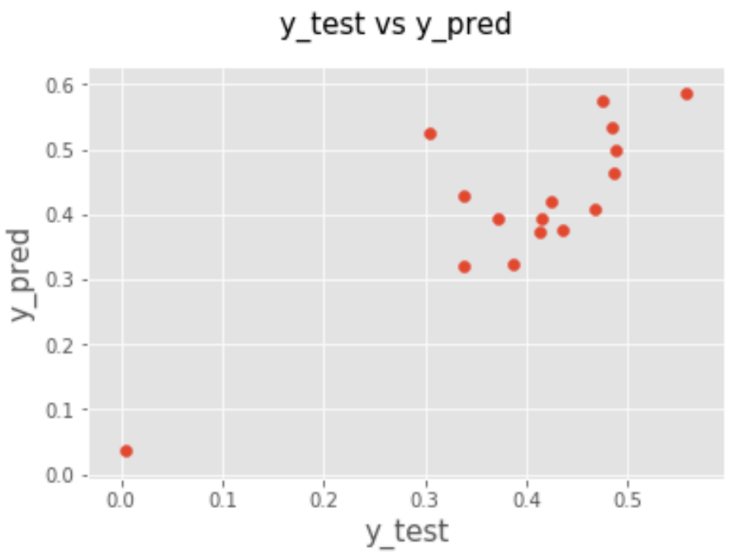


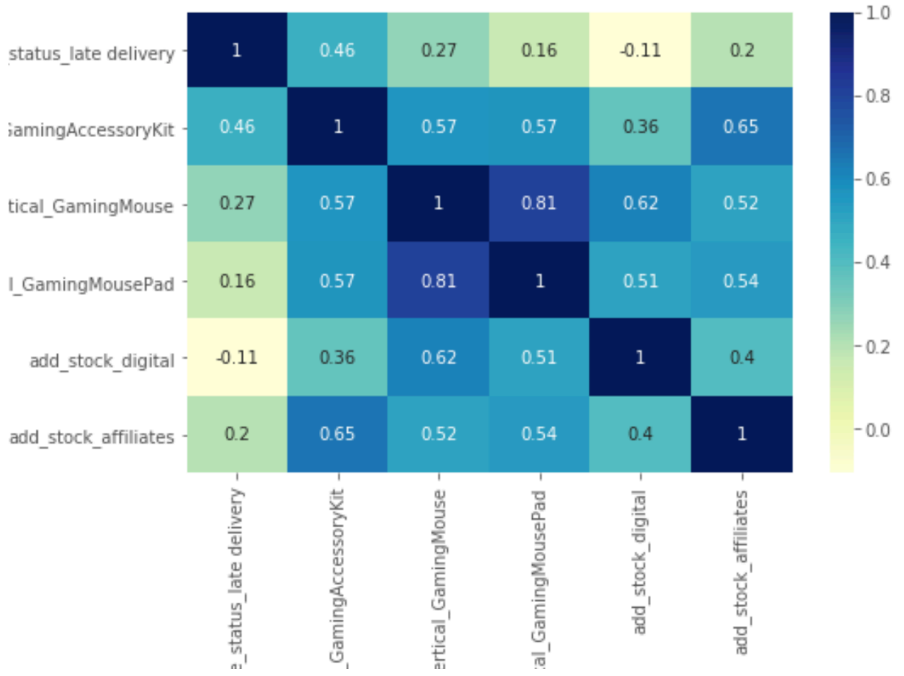
- We can see that, **Camera Accessory** has a very poor test performance stating a over fit model
- In linear regression, more columns to rows ratio results in over fitting
- As a follow up step, we have to estimate other models as well as linear regression with less features to understand the over fit

Model: Linear Regression- Gaming Accessory

- As a starting point, we **estimated linear regression models for all 3 sub categories**
- We present the R-Squared scores, Plots and Other details here

Gaming Accessory	**0.941**	**0.634**





- We can see that, **Gaming Accessory** has a very poor test performance stating a over fit model
- In linear regression, more columns to rows ratio results in over fitting
- As a follow up step, we have to estimate other models as well as linear regression with less features to understand the over fit

Model: Linear Regression Snap Shot

- To consolidate, final KPI's used in all models are

Home Audio KPI's	Camera Accessory KPI's	Gaming Accessory KPI's
FM Radio Vertical (Binary)	Vertical Lens (Binary)	Vertical Mouse (Binary)
Home Audio vertical (Binary)	Vertical Strap (Binary)	Vertical Mousepad (Binary)
TV Add Stock	Temperature (Weather)	Vertical Kit (Binary)
Digital Add Stock	Stock Index	Late Delivery
Sponsorship Add Stock	TV Add Stock	Digital Add Stock
	Affiliates Add Stock	Add Stock Affiliates

- **Add Stocks has high contribution**
- **Repeated binary variables might be a problem for over fit in Camera and Gaming Accessory Categories**

Future Road Map

- In future, we are planning to study other market mix models prevalent in the industry

- The models prevalent in the industry include
  - **Linear Model:** Linear regression is a prevalent one
  - **Multiplicative Model:** Exponential or log models to capture non linearity in data
  - **Koyck Model:** A time series extension of simple linear models having lagged response variables as features
  - **Distributed Lag Model:** It's an extension of koyck with time series models lag features on existing feature list

- We want to target answering the following questions
  - **Whether the existing marketing lever has positive or negative influence on revenues?** (Linear Regression explains this)
  - **Which model can explain maximum variance in the models?** (Extensively work on multiplicative and lag models)
  - **Study the lag contribution on add stocks**

- **Also add few KPI's if necessary along with formulating lag KPI's essential for lag models**