**ElecKart Market Mix Modelling**

**Problem Statement:**

* Identifying the Profitable channels for marketing & media investments
* Identifying the important factors having positive impact on sales of three different product categories – Camera Accessory, Game Accessory & Home Audio

**Data Source:**

* Orders data from June-2015 to June-2016 in Order level
* Products details
* Investment in various media heads, INR Cr. In Month level
* NPS, Stock index data in Month Level
* Special Sale Calendar
* Holidays in Ontario State (<https://en.wikipedia.org/wiki/Public_holidays_in_Canada>)

**Approach:**

* Importing Data
* Data Cleansing
* Feature Engineering
  + Creating New Features
  + Converting all datasets into Weekly Level
  + Creating separate datasets for Cam accessories, home audio, game accessories
* Exploratory Data Analysis
* Model building
  + Linear Model
  + Time Series Model (Multiplicative Model)
  + Logarithmic Model

**Present Status:**

Basic Linear Model on Cam accessories dataset, home audio and game accessories

**Future Roadmap:**

* Creating Multiplicative Model
* Creating Logarithmic Model
* Validation of Models
* Selection of best model to predicting Sales

**Assumptions:**

* Media Investment in a month will be uniformly distributed across all the weeks.

Sum of Investment in Weeks should be equal to Monthly Investment and It is same through the Week.

* Net Promoter’s Score provided Monthly It has been assumed that this is same through- out the Month and this will change in a week when same week has Two months data

**Importing Data:**

* Data has been imported from Consumer Electronics file to get orders data
* Media Investment Data, Product list , Special Sale Dates and Monthly NPS Score imported.
* Special Sale Calendar and Holiday Lists are calculated.

**Data Cleansing:**

* In orders data checks have been done to remove data other than June 2015-July 2016
* gmv values with leading spaces have been removed
* Duplicate rows have been removed.
* Orders with product mrp zero has been removed
* Orders having gmv value
* After creating KPI list price(gmv/units) Records that having list\_price greater than product\_mrp has been removed

**Feature Engineering:**

**Creating New Features:**

* List price is created by dividing gmv by units
* Discount Created by substraction of product\_mrp and list price
* % Discount by dividing discount and list price
* Holiday flag created If holiday indicated by 1 else 0
* Special Sale flag created if order date equals to special sale date indicated by 1 else 0
* Pay Date flag created If order date is in between 1 to 15 will show 1 else 0
* As dataset is from July 2015 to June 15 Weekly data should be from Week 27th of 2015 and Week 26th of year 2016.This has been converted as Week 27th to Week 79th Data

**Converting all datasets into Weekly Level:**

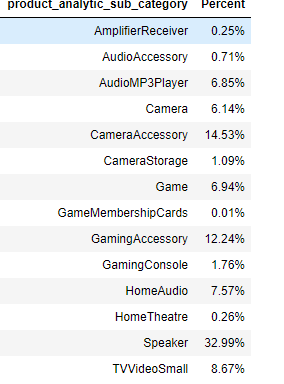
* In Orders data Week can be created from order date and data
* Media Investment Data and NPS data provided has been joined created by taking weighted average on dates as multiplying No of days in week and divided by No. of Days in Month created Week level Media Investment and NPS
* For Media Investments Sum is the aggregator and for NPS and Stock Index average is the aggregator
* After Creating the Week level KPIs new Datasets created in weekly Level

**Creating separate datasets for Cam accessories, home audio, game Accessories:**

* Orders data has been converted to three different datasets based on product analytic sub categories
* Data has been grouped to Week as list price, discount as sum and discount has been created by dividing discount with list price, COD, Prepaid orders as count, sla, product\_procurement sla as Average, No. of holidays, Special sales, Payment date in that particular Week
* Media Investment and NPS data has been not consists of product analytic sub category

So, product list data has been used.

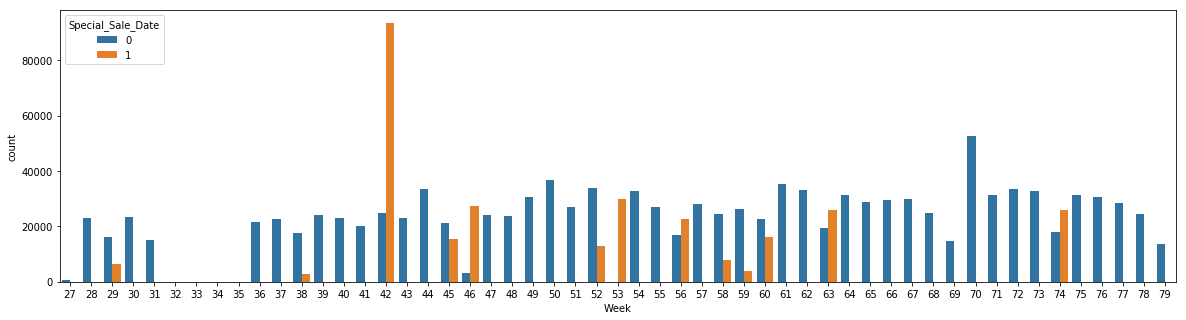
* Each product Frequency % has been aggregated for respective sub categories



* For Cam accessories, Investment data has been calculated by multiplying the data with 14.53%, for Home Audio with 7.57% and for Game Accessories with 12.24%
* After creating Investment data, NPS data in Separate datasets these will be joined with orders data as Week is common Field.

**Exploratory Data Analysis**

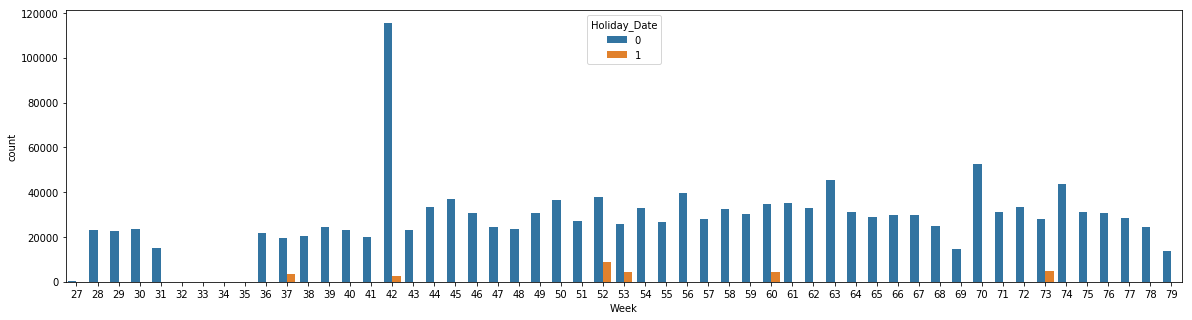
* In EDA it has been observed No sales are there from Week 32 to Week 36
* In Week 42 It has been Observed More orders has been ordered in between from October 12th to October 18th and 15th-17th Special sales was going.
* Big Discount Sale in Week 42, Week 63 and Week 74 It has been Observed More Orders

****

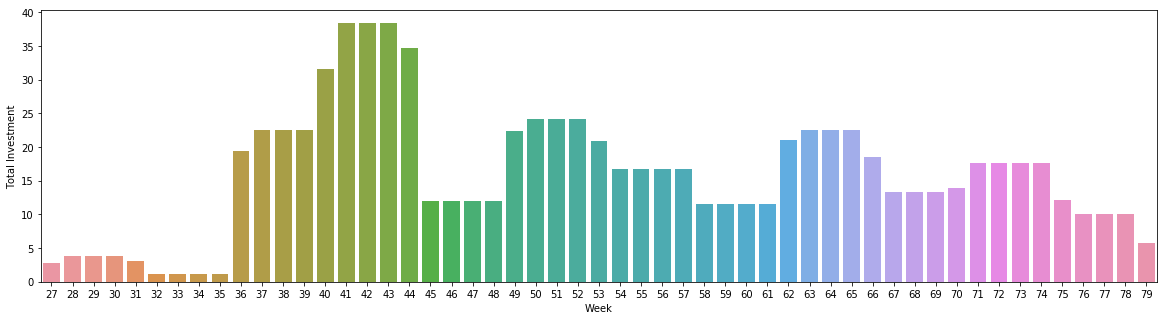
**Week 42:**



* More No. of orders has been ordered in Weeks doesn’t having any Public Holidays

****

Most of the ad spends during Week 36 to Week 44

****

Out of 846.45 Crores ,31% amount spent in this Period

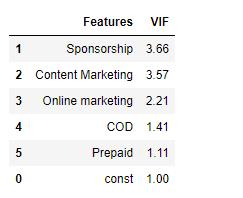
* Most of the Investments went through Sponsorship channels
* Less ads have been given by Content Marketing and Radio
* ElecKart got good net promoter score during Week 31 to Week 35 ranging of 57.29% to 59.99
* Drastic Drop has beens Observed of 6.53% from Week 35th to Week 36th and Week 36th to Week 37th.
* Detailed EDA is available in python File.

**Model Building:**

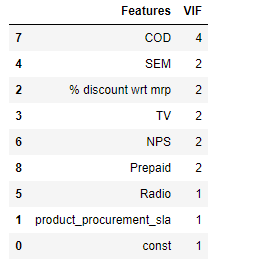
* Multiple Linear Regression is applied on the Three Datasets where List price as predictor
* As Dataset is with less records Type-II error observed which is causing High P Values on few variables
* Planning to Regularize the model by applying Ridge.
* Results of Basic Linear Models after RFE and VIF

|  |  |  |  |
| --- | --- | --- | --- |
| Dataset | Cam Accessories | Home Audio | Gaming Accessories |
| R-Square | 0.860 | 0.992 | 0.956 |
| R2\_Score | 0.838 | 0.975 | 0.908 |

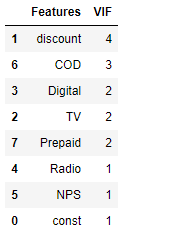
* **Variables in Cam accessories Model:**



* **Variables in Home audio Model:**



* **Variables in Game accessories Model:**



As Feature Engineering is done with multiple iterations there will be a chance of adding/Removing of Variables here. COD and Prepaid or No. of orders.