Market Mix Modelling
For
ElecKart

Approach Document

Team Members:

Arun NR

Sunil Saggar

Prazna Teja

Nilesh Singh

Steps Involved

LEGEND:
Completed So Far
Work in Progress

1. Data Sourcing
2. Data Cleaning
3. EDA
4. Deriving KPIs
5. Generating Weekly Data.
    a. Sales
    b. Media and other Information.
6. Basic Linear Regression Model.
7. Multiplicative Model
8. Koyck Model
9. Distributed Lag Model
10. Combo of Distributed Lag and Multiplicative Model
11. Results.

Data Sourcing

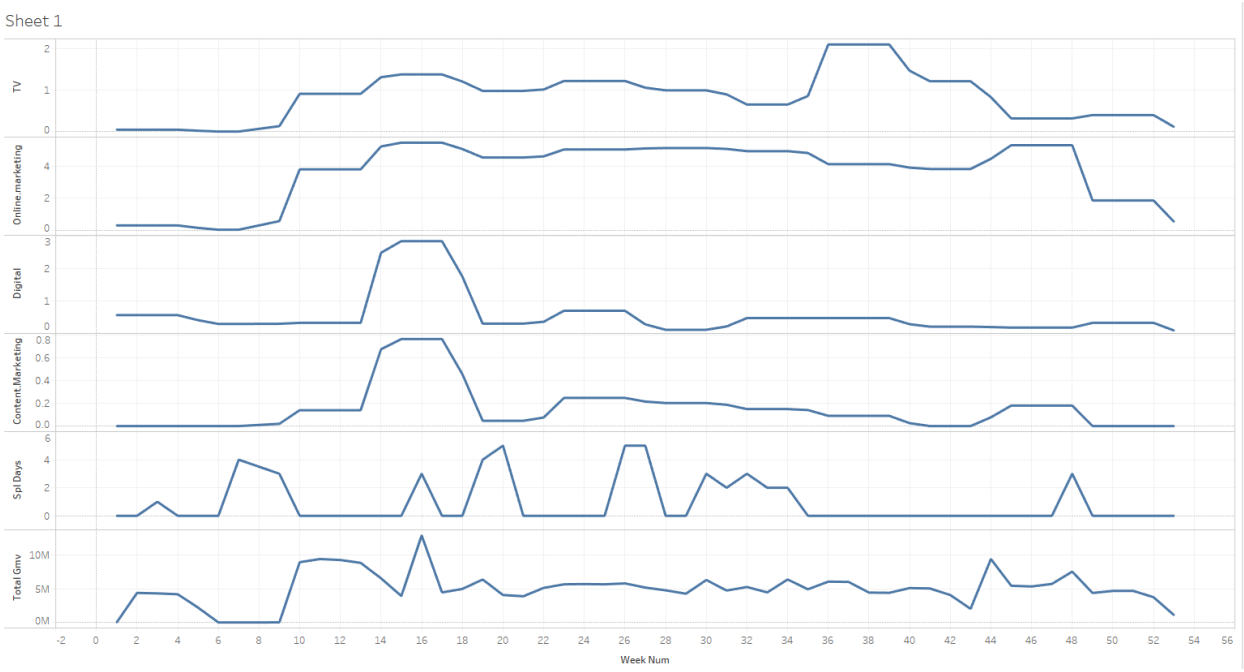
- ConsumerElectronics.csv
    - Contains SKU level transaction details.
- Media data and other information.xlsx
    - Contains 3 tabs
        - Media Investment: Contains investment from different sources of media.
        - Special Sale: Contains dates during when special sale was conducted.
        - Monthly NPS
    - For ease of reading, each of these tabs are manually converted to R readable csv format.

Data Cleaning

- Gmv, product_mrp, product_procurement_sla had null values which was removed.
- Product_procurement_sla had negative values, and was removed.
- Sla and product_procurement_sla had outliers, and was capped accordingly.
- Several rows in gmv & cust_id having NAs were omitted.
- Order date is converted to proper date format.
- Considered only data between 01-Jul-2015 to 30-Jun-2016 as defined in the problem statement.
- Filtered data relevant to only 3 Product_analytic_sub_category as defined in the problem statement.
    - Camera Accessory
    - Audio Accessory
    - Home Audio

EDA

Some of the key EDA Graphs



Insights:

- We can see that GMV peaks are observed where there is significant investment is certain category showing the current effect.
- Also in some cases we can see the lag effect of marketing on gmv.
- On Special Days, the advertisement spends are more and so is the gmv/sales.

Co-relation Matrix

	total_gmv	total_unit	no_of_orc	total_payi	total_payi	avg_sla	avg_product_mrp	avg_prod	avg_disco	avg_unitP	Total.Inve	TV	Digital	Sponsorsh	Content.Mark	Online.mi	Affiliates	SEM	Radio	Other	spl_days	NPS
total_gmv	1	0.880488	0.884207	0.872011	0.63629	-0.4763	0.186487556	-0.42052	0.378974	-0.02142	0.319775	0.186719	0.17513	0.299371	0.247960052	0.270961	0.252419	0.217805	0.018254	0.014921	0.076053	-0.06302
total_units	0.880488	1	0.999898	0.949981	0.790526	-0.41477	-0.100358399	-0.42204	0.597956	-0.29805	0.268246	0.169052	0.119851	0.24295	0.189886327	0.253707	0.24171	0.159328	0.038131	0.030638	0.030389	-0.03574
no_of_orders	0.884207	0.999898	1	0.950169	0.79043	-0.41861	-0.094659602	-0.42574	0.596115	-0.29347	0.268989	0.170412	0.119512	0.24256	0.190341592	0.256004	0.244129	0.159297	0.039764	0.032166	0.030941	-0.03569
total_payment_cod	0.872011	0.949981	0.950169	1	0.560086	-0.40027	-0.052908772	-0.39029	0.551417	-0.25027	0.294432	0.133224	0.15583	0.253006	0.248871318	0.272221	0.249373	0.193419	0.078328	0.067515	0.072808	-0.07683
total_payment_prepaid	0.63629	0.790526	0.79043	0.560086	1	-0.32592	-0.147592118	-0.3645	0.500649	-0.28812	0.136284	0.191085	0.011399	0.14745	0.016804672	0.145422	0.158759	0.043267	-0.04824	-0.04719	-0.06084	0.05614
avg_sla	-0.4763	-0.41477	-0.41861	-0.40027	-0.32592	1	-0.185649511	0.387229	-0.23208	-0.00632	0.135842	0.061518	0.172563	0.108523	0.178527059	0.084185	0.074209	0.176027	-0.024	-0.01164	0.072026	0.021674
avg_product_mrp	0.186488	-0.10036	-0.09466	-0.05291	-0.14759	-0.18565	1	-0.23427	-0.15669	0.942891	-0.04695	-0.0678	0.020109	-0.04822	-0.01152564	-0.04486	-0.04789	0.010685	-0.01739	-0.02811	-0.02414	-0.00855
avg_product_procurement_sla	-0.42052	-0.42204	-0.42574	-0.39029	-0.3645	0.387229	-0.234271781	1	-0.65002	-0.06686	0.034364	-0.03198	0.050582	0.077995	0.053508998	-0.0346	-0.04975	0.053066	-0.0726	-0.07084	-0.01092	0.003781
avg_discount	0.378974	0.597956	0.596115	0.551417	0.500649	-0.23208	-0.156688422	-0.65002	1	-0.38071	0.124015	0.129138	0.024053	0.073865	0.071667185	0.182602	0.185832	0.051182	0.064463	0.059638	0.041893	-0.04927
avg_unitPrice	-0.02142	-0.29805	-0.29347	-0.25027	-0.28812	-0.00632	0.942890559	-0.06686	-0.38071	1	-0.09522	-0.11153	0.002328	-0.07942	-0.050707566	-0.10925	-0.11198	-0.01668	-0.03338	-0.04476	-0.05459	-0.00039
Total.Investment	0.319775	0.268246	0.268989	0.294432	0.136284	0.135842	-0.046948264	0.034364	0.124015	-0.09522	1	0.690544	0.713662	0.890488	0.861415227	0.725967	0.663002	0.812555	0.075448	0.110198	-0.04677	-0.32872
TV	0.186719	0.169052	0.170412	0.133224	0.191085	0.061518	-0.067804411	-0.03198	0.129138	-0.11153	0.690544	1	0.296469	0.516548	0.360898386	0.599415	0.645226	0.331631	0.15681	0.365241	-0.04913	-0.16686
Digital	0.17513	0.119851	0.119512	0.15583	0.011399	0.172563	0.02010943	0.050582	0.024053	0.002328	0.713662	0.296469	1	0.691608	0.896780477	0.254269	0.176721	0.973154	-0.27985	-0.24241	-0.03235	-0.18148
Sponsorship	0.299371	0.24295	0.24256	0.253006	0.14745	0.108523	-0.048219079	0.077995	0.073865	-0.07942	0.890488	0.516548	0.691608	1	0.722331823	0.441243	0.360297	0.76438	-0.29086	-0.25131	-0.17953	-0.29193
Content.Marketing	0.24796	0.189886	0.190342	0.248871	0.016805	0.178527	-0.01152564	0.053509	0.071667	-0.05071	0.861415	0.360898	0.89678	0.722332	1	0.572432	0.481955	0.955317	0.041129	0.008228	0.064109	-0.31745
Online.marketing	0.270961	0.253707	0.256004	0.272221	0.145422	0.084185	-0.044856312	-0.0346	0.182602	-0.10925	0.725967	0.599415	0.254269	0.441243	0.57243169	1	0.988302	0.41743	0.358514	0.302999	0.184688	-0.22145
Affiliates	0.252419	0.24171	0.244129	0.249373	0.158759	0.074209	-0.047885104	-0.04975	0.185832	-0.11198	0.663002	0.645226	0.176721	0.360297	0.481954584	0.988302	1	0.327973	0.372746	0.344949	0.194232	-0.17819
SEM	0.217805	0.159328	0.159297	0.193419	0.043267	0.176027	0.010684623	0.053066	0.051182	-0.01668	0.812555	0.331631	0.973154	0.76438	0.95531708	0.41743	0.327973	1	-0.19557	-0.20492	-0.00685	-0.25246
Radio	0.018254	0.038131	0.039764	0.078328	-0.04824	-0.024	-0.017392038	-0.0726	0.064463	-0.03338	0.075448	0.15681	-0.27985	-0.29086	0.041129156	0.358514	0.372746	-0.19557	1	0.944553	0.127531	-0.13171
Other	0.014921	0.030638	0.032166	0.067515	-0.04719	-0.01164	-0.028110921	-0.07084	0.059638	-0.04476	0.110198	0.365241	-0.24241	-0.25131	0.008228177	0.302999	0.344949	-0.20492	0.944553	1	0.087798	-0.10744
spl_days	0.076053	0.030389	0.030941	0.072808	-0.06084	0.072026	-0.024137939	-0.01092	0.041893	-0.05459	-0.04677	-0.04913	-0.03235	-0.17953	0.064109114	0.184688	0.194232	-0.00685	0.127531	0.087798	1	0.050779
NPS	-0.06302	-0.03574	-0.03569	-0.07683	0.05614	0.021674	-0.00854933	0.003781	-0.04927	-0.00039	-0.32872	-0.16686	-0.18148	-0.29193	-0.317454568	-0.22145	-0.17819	-0.25246	-0.13171	-0.10744	0.050779	1

Can Observe lot of co-relation between the following:

- Total number of units, no of orders, gmv

- average unit price, average product mrp

- Digital and SEM

- SEM and Content Marketing

This will be useful consideration while eliminating variables during regression.

Deriving other Factors from Data

The following additional factors were derived from the data

- Week Number: Which will be used to aggregate later on.
- Discount Offered: Calculated based on GMV and MRP values.
- Unit Price: Will be the price per unit after discount. Derived from GMV and 'no. of units'.
- No_of_orders: No of orders registered during that particular week.

Generating Weekly Data

Weekly Sales Data

- Weekly sales data is derived from the week numbers associated with each sales record.
- All the information is aggregated by grouping at
    - 'week number'
    - 'product analytic sub category'
- The rest of the data is aggregated and summarized as follows:
    - Count(number of orders)
    - Sum(gmv)
    - Sum(units)
    - Sum(payment_type)
    - Mean(sla)
    - Mean(mrp)
    - Mean(sla)
    - Mean(discount)
    - Mean(unitPrice)

Weekly Media Investment Data

- As input, monthly wise media investment data was received.
- From monthly wise data, we derived weekly wise media investment data.
- Same weekly data will be used for all three product analytic sub category, as marketing is done, irrespective of sub-category.

Weekly Special Sale

- Days during which the special sale happens was provided.
- Special days was mapped to week numbers.
- No. of special days within each week was calculated.

- Same special sale weekly data is used for all three product analytic sub category.

Weekly NPS Data:

- Monthly NPS score was provided.
- From monthly wise data, we derived weekly wise NPS score.
- Same weekly data will be used for all three product analytic sub category

Merging of weekly data:

The following information was merged to form one master table.

- Weekly sales information
- Weekly media spends across different segments
- Weekly wise number of sale days
- Weekly NPS Data.

Resulting in table below:

```
> str(sales_all)
'data.frame':   152 obs. of  24 variables:
 $ week_num                     : num  1 1 1 2 2 2 3 3 3 4 ...
 $ product_analytic_sub_category: Factor w/ 14 levels "AmplifierReceiver",..: 2 5 11 2 5 11 2 5 11 2 ..
 $ total_gmv                    : num  12340 15731 106756 269180 4375483 ...
 $ total_units                  : int  2 26 38 99 3400 2459 88 3194 2423 107 ...
 $ no_of_orders                 : int  2 26 38 95 3252 2434 87 3097 2403 107 ...
 $ total_payment_cod            : num  0 0 1 63 2171 ...
 $ total_payment_prepaid        : num  2 26 37 32 1081 ...
 $ avg_sla                      : num  9 2.23 2.79 8.69 7.2 ...
 $ avg_product_mrp              : num  10568 1412 4642 3667 2412 ...
 $ avg_product_procurement_sla  : num  2.5 2.88 3.61 2.68 2.81 ...
 $ avg_discount                 : num  42.2 49.9 35.6 31 43.8 ...
 $ avg_unitPrice                : num  6170 605 2809 2802 1328 ...
 $ Total.Investment             : num  3.86 3.86 3.86 3.86 3.86 ...
 $ TV                           : num  0.0452 0.0452 0.0452 0.0452 0.0452 ...
 $ Digital                      : num  0.565 0.565 0.565 0.565 0.565 ...
 $ Sponsorship                  : num  1.67 1.67 1.67 1.67 1.67 ...
 $ Content.Marketing            : num  0 0 0 0 0 0 0 0 0 0 ...
 $ Online.marketing             : num  0.294 0.294 0.294 0.294 0.294 ...
 $ Affiliates                   : num  0.113 0.113 0.113 0.113 0.113 ...
 $ SEM                          : num  1.13 1.13 1.13 1.13 1.13 ...
 $ Radio                        : num  0 0 0 0 0 0 0 0 0 0 ...
 $ Other                        : num  0 0 0 0 0 0 0 0 0 0 ...
 $ spl_days                     : num  0 0 0 0 0 0 1 1 1 0 ...
 $ NPS                          : num  12.3 12.3 12.3 12.3 12.3 ...
```

Linear Regression Model Building

- Separate Linear model was built for each product analytic sub category.
- Variable elimination method was followed to identify the key factors in each case.
- Used STEP-AIC function to gain the initial model, then on continued with variable elimination based on VIF and p-values.

Camera Accessory

Final Model:

```
Coefficients:
             Estimate Std. Error t value             Pr(>|t|)
(Intercept)  5031182     240147  20.950 < 0.0000000000000002 ***
TV           -745888     364282  -2.048             0.046215 *
Sponsorship  2035885     425742   4.782            0.0000175 ***
Affiliates   1365160     332808   4.102             0.000161 ***
SEM          -955318     386603  -2.471             0.017153 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1732000 on 47 degrees of freedom
Multiple R-squared:  0.5447,    Adjusted R-squared:  0.5059
F-statistic: 14.06 on 4 and 47 DF,  p-value: 0.000000129

> sort(vif(model_6))
 Affiliates          TV         SEM Sponsorship
   1.883650    2.256774    2.541811    3.082527
```

Testing:

Adjusted R2 of model6 = 0.5059

R2 from prediction = 0.5446743

Difference is 0.0387743

Conclusion: Influencing Factors

- TV
- Sponsorship
- Affiliates
- SEM

Audio Accessory

Final Model

```
Coefficients:
            Estimate Std. Error t value             Pr(>|t|)
(Intercept)   338171      18614  18.168 < 0.0000000000000002 ***
Sponsorship    82853      29440   2.814              0.00717 **
Affiliates     57079      20072   2.844              0.00663 **
SEM           -73730      29237  -2.522              0.01520 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 131600 on 46 degrees of freedom
Multiple R-squared:  0.2932,    Adjusted R-squared:  0.2471
F-statistic:  6.36 on 3 and 46 DF,  p-value: 0.001068

> sort(vif(model_6))
 Affiliates         SEM Sponsorship
   1.139544    2.417876    2.451451
```

Testing

Adjusted R2 of model6 = 0.2471

R2 from prediction = 0.2931744

Difference is 0.046

Conclusion: Influencing Factors

- Sponsorship
- Affiliates
- SEM

Home Audio

Final Model

```
Coefficients:
             Estimate Std. Error t value             Pr(>|t|)
(Intercept)   5029974     297577  16.903 < 0.0000000000000002 ***
avg_discount  1220087     315360   3.869             0.000343 ***
Sponsorship   1284111     310612   4.134             0.000149 ***
spl_days       539935     317001   1.703             0.095268 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2104000 on 46 degrees of freedom
Multiple R-squared:  0.4868,    Adjusted R-squared:  0.4534
F-statistic: 14.55 on 3 and 46 DF,  p-value: 0.0000008499

> sort(vif(model_8))
 Sponsorship avg_discount     spl_days
    1.067734     1.100627     1.112109
```

Testing

Adjusted R2 of model6 = 0.4534

R2 from prediction = 0.4868225

Difference is 0.033

Conclusion: Influencing Factors

- AverageDiscount
- Sponsorship
- spl_days

Future Roadmap.

With linear model, we were able to model the current effect of advertising. We will implement the following models to determine other effects of marketing on revenue.

1. Multiplicative Model
2. Koyck Model
3. Distributed Lag Model
4. Combo of Distributed Lag and Multiplicative Model
5. Results.