## Marketing Mix Modelling

*Abstract*

Today, marketing is something that every company and organization must implement in its [growth strategy](https://www.cyberclick.net/numericalblogen/how-to-apply-the-get-keep-grow-funnel). Companies are heavily investing in marketing to achieve their goals so as to promote themselves and increase sales of their product or service. These days, marketing is one of the key aspects of businesses.

Effective marketing is vital for enhancing a brand, pushing promotions and differentiating a business from its competition.

Market Mix Modeling (MMM) is a technique which helps in quantifying the impact of several marketing inputs on sales or Market Share. The purpose of using MMM is to understand how much each marketing input contributes to sales, and how much to spend on each marketing input.

i

# Contents

[List of Figures](#_bookmark0) iv

[List of Tables](#_bookmark1) v

1. [Introduction](#_bookmark2) 1
   1. [Outline of Problem Statement](#_bookmark3) 1
   2. [Explaining the problem](#_bookmark4) 2
   3. [Need for Study](#_bookmark5) 2
   4. [Business Opportunity](#_bookmark6) 2
2. [Data Cleaning and Pre-processing](#_bookmark7) 3
   1. [Introduction](#_bookmark8) 3
   2. [Data Cleaning](#_bookmark9) 4
      1. [Removal of Unwanted Variables](#_bookmark10) 4
      2. [Null Value Treatment](#_bookmark12) 5
      3. [Duplicate Value Treatment](#_bookmark16) 7
      4. Checking Data Consistency……………………………………………………………………………..8
3. [Exploratory Data Analysis](#_bookmark17) 9
   1. [Univariate Analysis](#_bookmark18) 9
      1. [Daily](#_bookmark19) Active Users(DAU) 9
      2. [Covid\_index](#_bookmark21) 9
      3. [Brand\_score](#_bookmark23) 10
      4. [Vouchers\_P1](#_bookmark25) 11
      5. [Vouchers\_P2](#_bookmark25) 11
      6. [Meta\_video](#_bookmark29) 12
      7. [Meta\_nonvideo](#_bookmark29) 12
      8. [YT](#_bookmark33) 13
      9. [TikTok](#_bookmark35) 14
      10. [Platform\_1](#_bookmark36) 14
      11. [Platform\_2](#_bookmark36) 15
      12. Own\_digital\_1 16
      13. [Own\_digital\_2](#_bookmark43) 16
      14. [Insta\_all](#_bookmark45) 17

ii

Contents iii

* + 1. [Insta\_own](#_bookmark47) 18
  1. [Outlier Treatment](#_bookmark53) 19
  2. [Bi-Variate Analysis](#_bookmark56) 19
     1. [Correlation plots with Target Variable](#_bookmark57) 19
     2. [Bivariate Plots](#_bookmark60) 19
     3. [Insights from Bivariate Analysis](#_bookmark64) 23
     4. Multivariate Analysis Multivariate Plots………….……………………………………………23
     5. Insights from Multivariate Analysis………………………………………………………………24

1. Modelling
   1. [Feature](#_bookmark78) Transformation 26
   2. Modelling Approach………………………………………………………………………………………………….28
   3. Modelling Steps…………………………………………………………………………………………………….….28
   4. [Model Building](#_bookmark83) 30
   5. [Results](#_bookmark157) 33
2. [Business Recommendations](#_bookmark159) 36
   1. Response Curves…………………………………………………………………………………………………… 36
   2. Simulator…………………………………………………………………………………………………………………. 39
   3. Optimization……………………………………………………………………………………………………………..40

# List of Figures

2.1: KM Data info after Null Treatment .....7

2.2: Spends Data info after Null Treatment…………………………………………………………………………………..7

3.1: Variable: DAU………………………………………………………………………………………………………………………..9

3.2: DAU distribution……………………………………………………………………………………………………………………9

3.3: Variable: Covid\_index………………………………………………………………………………………………………..….9

3.4: Covid\_index distribution……………………………………………………………………………………………………....9

3.5: Variable: Brand\_score………………………………………………………………………………………………………….10

3.6: Brand\_score distribution……………………………………………………………………………………………………..10

3.7: Vouchers\_P1 impressions……………………………………………………………………………………………………11

3.8: Vouchers\_P1 distribution…………………………………………………………………………………………………….11

3.9: Vouchers\_P1 execution……………………………………………………………………………………………………….10

3.10: Variable: Vouchers\_P2………………………………………………………………………………………………………11

3.11: Vouchers\_P2 distribution………………………………………………………………………………………………….11

3.12: Vouchers\_P2 execution…………………………………………………………………………………………………….12

3.13:Meta\_video………………………………………………………………………………………………………………….......12

3.14: Meta\_video distribution……………………………………………………………………………………………………12

3.15: Meta\_nonvideo…………………………………………………………………………………………………………………13

3.16: Meta\_nonvideo distribution……………………………………………………………………………………………..13

3.17: Meta\_nonvideo execution………………………………………………………………………………………………..13

3.18: YT …………………………………………………………………………………………………………………………………….13

3.19: YT distribution…………………………………………………………………………………………………………………..13

3.20: YT execution……………………………………………………………………………………………………………………..13

3.21: TikTok ………………………………………………………………………………………………………………………………14

3.22: TikTok distribution…………………………………………………………………………………………………………….14

3.23: TikTok execution……………………………………………………………………………………………………………….13

3.24: Platform1………………………………………………………………………………………………………………………....14

3.25: Platform1 distribution……………………………………………………………………………………………………... 14

3.26: Platform1 execution………………………………………………………………………………………………………....15

3.27: Platform\_2……………………………………………………………………………………………………………………..…15

3.28: Platform\_2 distribution…………………………………………………………………………………………………..…15

3.29: Platform\_2 execution………………………………………………………………………………………………………..15

3.30: Own\_digital\_1……………………………………………………………………………………………………………………16

3.31: Own\_digital\_1 distribution………………………………………………………………………………………………..16

3.32: Own\_digital\_1 execution …………………………………………………………………………………………………. 16

3.33: Own\_digital\_2……………………………………………………………………………………………………………………16

3.34:Own\_digital\_2\_distribution………………………………………………………………………………………………..16

3.35:Own\_digital\_2\_execution……………………………………………………………………………………………………17

3.36: Insta\_all …………………………………………………………………………………………………………………………..17

3.37: Insta\_all distribution …………………………………………………………………………………………………………17

3.38: Insta\_all execution……………………………………………………………………………………………………………..17

3.39: Insta\_own…………………………………………………………………………………………………………………………. 17

3.40: Insta\_own distribution……………………………………………………………………………………………………….17

3.41: Insta\_own execution………………………………………………………………………………………………………….17

3.42: Bivariate: Vouchers\_P1, Vouchers\_P2………………………………………………………………………………. 20

3.43: Bivariate: meta\_video, meta\_nonvideo……………………………………………………………………………..21

3.44: Bivariate: YT, TikTok…………………………………………………………………………………………………………..21

3.45: Bivariate: Platform1, Platform2……………………………………………………………………………………….…22

3.46: Bivariate: Own\_digital\_1, Own\_digital\_2…………………………………………………………………………...23

3.47: Bivariate: Insta\_all, Insta\_own…………………………………………………………………………………………...23

13.48: Multivariate- Media variables: Average yearly impressions………………………………………………23

3.49: Multivariate: Year wise total sum……………………………………………………………………………………….23

3.50: Multivariate- Cost per keymetric(CPK) ……………………………………………………………………………… 23

3.51: Sigmoid transformation………………………………………………………………………………………..………27

4.1: Modelling Process-1……………………………………………………………………………………………………… 29

4.2: Iterations for new model……………………………………………………………………………………………….. 29

4.3: modelling\_process\_2……………………………………………………………………………………………………… 29

4.4: Modelling: Modelling results-1………………………………………………………………………………………..30

4.5: Model building- Model results-2……………………………………………………………………………………..32

4.6: Modelling: Base vs Incremental contribution…………………………………………………………………..33

4.7: Modelling: Contribution(Base and Incremental) ……………………………………………………………..34

4.8: Modelling: Promo vs Media Contribution in Total Incremental …………………..………………….29

4.9: Modelling: RoI ………………………………………………………………………………………………………………..35

5.1: Response Curves- meta\_video…………………………………………………………………………………………36

5.2: Response Curves- meta\_nonvideo…………………………………………………………………………………..37

5.3: Response Curves- YT………………………………………………………………………………………………………..37

5.4: Response Curves- TikTok………………………………………………………………………………………………….37

5.5: Response Curves- Platform2…………………………………………………………………………………………….38

5.6: Response Curves- Own\_digital\_1……………………………………………………………………………………..38

5.7: Response Curves- Own\_digital\_2……………………………………………………………………………………..38

5.8: Response Curves- Insta\_own……………………………………………………………………………………………39

# List of Tables

2.1: Data Type of Columns………………………………………………………………………………………………..5

2.2: Null Value Treatment - I…………………………………………………………………………………………….6

2.3: Null Value Treatment – II…………………………………………………………………………………………..6

4.1: Modelling: Modelling results-1………………………………………………………………………………..30

4.2: Modelling: Model stats-1………………………………………………………………………………………….31

4.3: Model building- Model results-2………………………………………………………………………………32

4.4: Modelling: Model stats-2………………………………………………………………………………………….32

5.1: Simulator…………………………………………………………………………………………………………………….39

5.2: Optimization results…………………………………………………………………………………………………..40

## Chapter 1

## Introduction

MMM is a technique used to quantify the impact of marketing inputs on sales volume/sales value/active users. The purpose of using MMM is to understand:

1. how much marketing input contribute to sales

2. what is the return on advertising spends

3. how much to spend on each marketing input

Or in other words, finding the right mix of Base and Incremental sales.

Total sales = Base sales + Incremental sales

**Terms**:

**Incremental drivers:** Business outcomes generated by marketing activities like TV and print ads, digital spends, price discounts, promotions, social outreach, etc. This can be further divided into:

ATL(Above-the-Line): main stream advertising/non-targeted (Traditional, digital),

BTL(Below-the-Line): individuals targeted; discounts, instore promotion

**Base drivers:** Base outcome is achieved without any advertisements. It is due to brand equity built over the years, macroeconomic factors (GDP, inflation, unemployment, purchasing power), demography, etc.

### Outline of Problem Statement

An E-commerce company is facing a lot of competition in the current market. To stay and flourish in the market, it is spending heavily on different marketing tactics to increase its reach to the customers. To increase the profits, the company requires to increase its base of active users.

Hence, the company wants to develop a model through which it can study the relation between their different marketing tactics, brand strength and the count of ‘Daily Active Users’. This will in turn help them to enhance marketing effectiveness, maximize the RoI and optimize the current spends on numerous marketing platforms.

We have the following tasks at hand:

* + - We have been assigned to develop a model for this company which explains the relation between the Daily Active Users traffic, Brand Strength and marketing tactics.
    - Provide business recommendations on the marketing tactics to maximize the Return on Investment for the company.

1

### Explaining the problem

* + - Here, we have been provided data collected by the Marketing and Sales arms of the company itself.
    - Data describes the spends by the company on various digital marketing platforms on a daily basis to create impressions among the users of those platforms.
    - Other non-marketing metrics covering the brand image/popularity, market lockdown situation owing to covid pandemic are also available.
    - The aim is to provide a model to identify the contribution of different non-marketing and marketing tactics on Daily Active Users(DAU) traffic on the company’s site.
    - Also, to recommend a strategy to maximize the DAU while optimizing the spends on marketing tactics.

### Need for Study

* + - The study helps companies identify the effectiveness of different marketing media in order to retain as well as increase the number of active users & sales.
    - For a long term growth perspective, the brand health and sentiment is also to be focused upon. This will help us better understand the contribution of brand strength in the overall sales/active users acquisition.
    - Measuring the impact of lockdown, pandemic and their consequences on online users’ traffic and sales can help companies tackle such situations more effectively.
    - In e-commerce business, staying visible & relevant to the public is very important. Being at the right place with right strategy can create big impact on company’s future performance .

### Business Opportunity

* + - Businesses can focus and prioritize the spends on the best performing channels for future investments.
    - Based on the performance of each media, we may also get to know the relevance of that media as well as can analyze deeper the company’s marketing methodology as to why high spends are not translating into the desired output.
    - At the end it boils down to maximizing the returns with optimal investment. Once we know the optimal mix of spends required for each media tactic, the company can save a lot of money.

## Chapter 2

**Data Cleaning and Pre-processing**

### Introduction

We have been provided data containing various media key-metrics, spends, brand metrics, covid related index for a duration of 2 years, 2020 & 2021 on daily basis.

Dataset-1: Key-metrics data

|  |  |
| --- | --- |
| Date | Depicts the date (01.01.2020 – 31.12.2021) |
| Month | Depicts the month number(1-12) |
| Month\_Year | Depicts Month & Year |
| Week | Week format(01 WK 2020) |
| Year | Year(2020, 2021) |
| DAU | No. of daily active users of the company’s website |
| Covid\_index | A consolidated covid score based on no. of covid cases and intensity of lockdown. It ranges from 0-100. With covid intensity, this also increases |
| Brand\_score | Metric tells about the brand health & popularity. This is based on the surveys conducted by the Survey analytics team of the company |
| Vouchers\_P1 | No. of promo vouchers directly offered on company’s website |
| Vouchers\_P2 | No. of promo vouchers directly offered on media other than company’s website |
| Meta\_video\_imp | Impressions created through video advertisements on Meta platform |
| Meta\_nonvideo\_imp | Impressions created through non-video advertisements on Meta platform |
| YT\_imp | Impressions created on YouTube |
| TikTok\_imp | Impressions created on TikTok |
| Platform1\_imp | Impressions created on a live-streaming platform- Platform1 |
| Platform2\_imp | Impressions created on a live-streaming platform- Platform2 |
| Own\_digital\_1\_imp | Impressions created on company owned digital media platform-1 |
| Own\_digital\_2\_imp | Impressions created on company owned digital media platform-2 |
| Insta\_all\_imp | Impressions created on Instagram using accounts/influencers other than the company’s |
| Insta\_own\_imp | Impressions created on Instagram using company’s account |

We will use the term key-metric to collectively refer No. of counts of vouchers & impressions for other media variables

Dataset-2: Spends data

Similar to the above impressions data, we also have daily spends data for each of the media and promo variables against each key-metric for creating those impressions/redeeming cost of the vouchers.

Before we proceed to build our model, we need to check each column for data quality and need to perform some level of pre-processing to achieve consistency.

### Data Cleaning

#### Removal of Unwanted Variables

We observe that many date columns are redundant. We can remove “Month” and “Month\_Year” columns. Rest of the columns appear to be useful, thus we will retain them.

Data Types of the Variables:

|  |  |
| --- | --- |
| Date | datetime64[ns] |
| Week | object |
| Year | object |
| DAU | float64 |
| Covid\_index | float64 |
| Brand\_score | float64 |
| Vouchers\_P1 | float64 |
| Vouchers\_P2 | float64 |
| Meta\_video\_imp | float64 |
| Meta\_nonvideo\_imp | float64 |
| YT\_imp | float64 |
| TikTok\_imp | float64 |
| Platform1\_imp | float64 |
| Platform2\_imp | float64 |
| Own\_digital\_1\_imp | float64 |
| Own\_digital\_2\_imp | float64 |
| Insta\_all\_imp | float64 |
| Insta\_own\_imp | float64 |

Table 2.1: Data Type of Columns

#### Null Value Treatment

#### It has been observed that key-metrics data for DAU(dependent variable), Vouchers\_P1 is

#### missing for October, 2020 due to data collection issue at their end. One month dependendent var

#### data is a considerable amount to be imputed, so will drop the other columns also for the same period. Now, we are left with 700 days of data.

#### Other than that, a few null values are present in the data.

|  |  |  |  |
| --- | --- | --- | --- |
| Variables | Null Count | Incorrect Values | Null Value Treatment |
| Date | 0 | - | No missing values |
| Week | 0 | - | No missing values |
| Year | 0 | - | No missing values |
| DAU | 0 | - | No missing values |
| Covid\_index | 0 | - | No missing values |

Table 2.2: Null Value Treatment - I

|  |  |  |  |
| --- | --- | --- | --- |
| Brand\_score | 0 | - | No missing values |
| Vouchers\_P1 | 0 | - | No missing values |
| Vouchers\_P2 | 10 | - | Spends for the same time-stamps is available. We can impute missing impressions value using Cost Per Impressions(CPP) of Voucher\_P2. |
| Meta\_video\_imp | 0 | - | No missing values |
| Meta\_nonvideo\_imp | 0 | - | No missing values |
| YT\_imp | 0 | - | No missing values |
| TikTok\_imp | 7 | - | Spends for the same time-stamps is available. We can impute missing impressions value using Cost Per Impressions(CPP) of TikTok. |
| Platform1\_imp | 0 | - | No missing values |
| Platform2\_imp | 0 | - | No missing values |
| Own\_digital\_1\_imp | 0 | - | - |
| Own\_digital\_2\_imp | 0 | - | No missing values |
| Insta\_all\_imp | 0 | - | No missing values |
| Insta\_own\_imp | 0 | - | No missing values |

Table 2.3: Null Value Treatment - II

Above missing values for Vouchers\_P2 and TikTok can be imputed using the

**Cost Per Rating Point/Key-metric(CPP)**. It is the average cost to create single impression/keymetric.

* CPP = Total spends/Total impressions
* Once we have the CPP, we can find the corresponding keymetric as below:
* Missing Impressions for day D = Spends for day D/CPP

We observe the following after Null treatment:

Key-metrics dataset Spends dataset

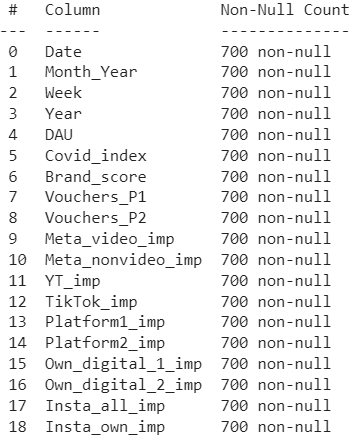
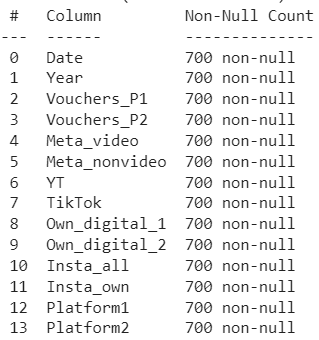


Figure 2.1: KM Data info after Null Treatment Figure 2.2: Spends Data info after Null Treatment

#### Duplicate Value Treatment

Number of Duplicate Rows are zero. Since we don’t have duplicate rows, we will proceed further.

* + 1. **Checking Data consistency**

Since spends are done only to create the keymetrics, it should not happen that spends values are present for rows with zero keymetric values and vice-versa.

It is observed that spends and keymetrics are consistent.

## Chapter 3

**Exploratory Data Analysis**

### Univariate Analysis

#### Daily Active Users

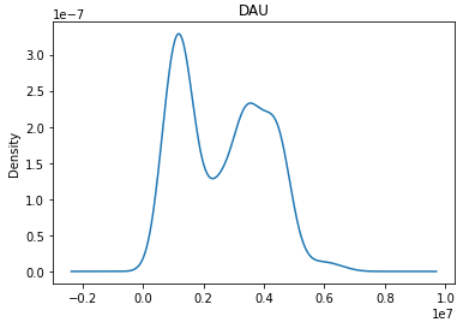
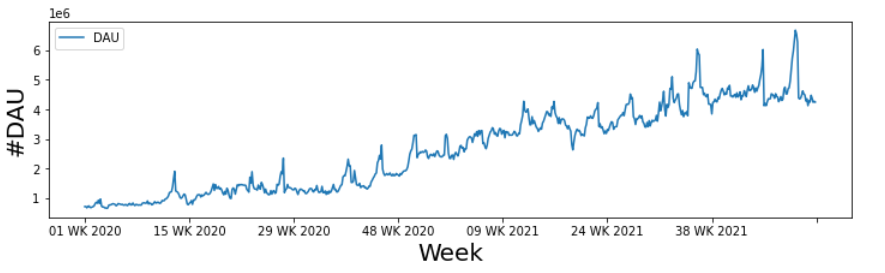
 ****

Figure 3.1: Variable: DAU Figure 3.2: DAU distribution

Insight: We can observe an overall upward trend in #Daily active users. It shows that the company is consistently acquiring users traffic on their website

#DAU varies from 2.6 million to 6.8 million with highest distribution of around 1 mn active users daily.

#### Covid\_index

#### 

#### 

Figure 3.3: Variable: Covid\_index Figure 3.4: Covid\_index distribution

Remarks: We observe that covid\_index ranges right from 0 to as high as 75 with highest presence at 69. High covid index for year 2021 can be a big reason for increase in online sales and traffic as lockdowns significantly stopped offline shopping in malls and markets which diverted people towards online shopping. We will analyze it better as we study the correlation.

#### Brand\_score

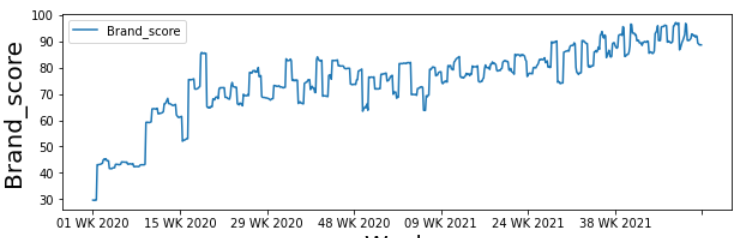
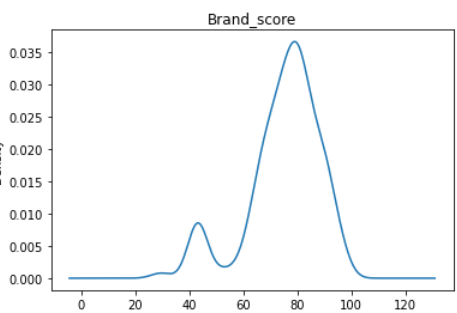
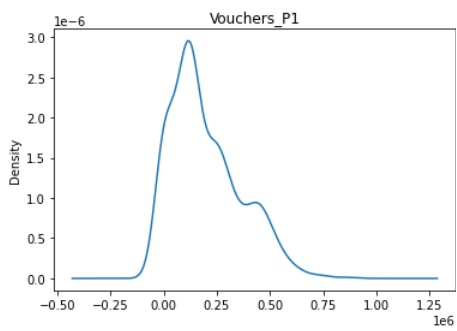
****

Figure 3.5: Variable: Brand\_score Figure 3.6: Brand\_score distribution

Remarks: There is an overall upward trend in brand\_score indicating strengthening of the brand health. It shows brand is gaining more and more popularity and acceptance.

Mean brand\_score is 75 which is quite good. This metric can be a useful metric in driving the DAUs growth.

Let’s plot the marketing variables now.

**3.1.4 Vouchers\_P1**

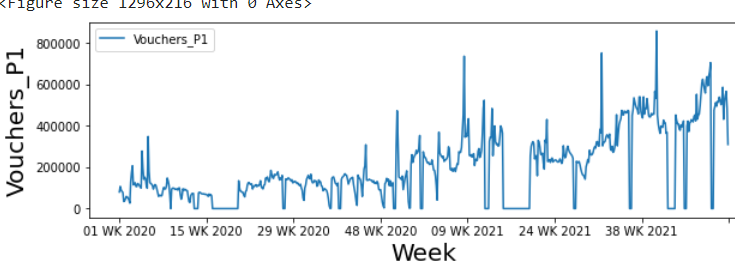


Figure 3.7: Vouchers\_P1 impressions Figure 3.8: Vouchers\_P1 distribution

Variable: Vouchers\_P1

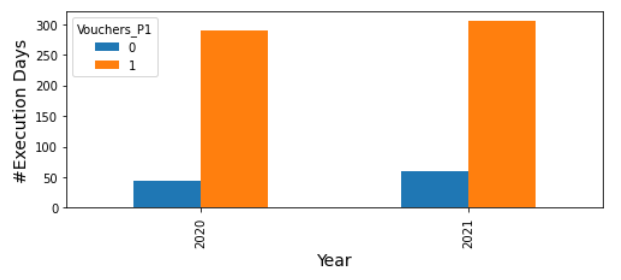


Figure 3.9: Vouchers\_P1 execution

Above plot shows the no. of days with and without execution in the particular tactic.

0: for number of day with no execution or zero values

1: for number of days with non zero execution

Vouchers\_P1 values depict the no. of vouchers offered directly on the company’s website. Company has been offering an average of 0.2mn vouchers daily. There are a very few days in both the years where vouchers offering has been zero as visible in above plot

#### Vouchers\_P2

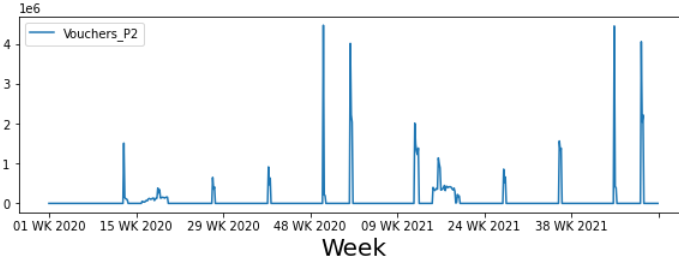
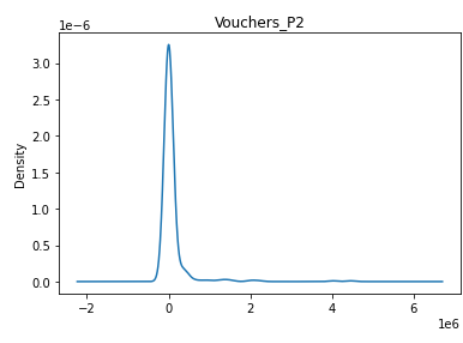
****

Figure 3.10: Variable: Vouchers\_P2 Figure 3.11: Vouchers\_P2 distribution

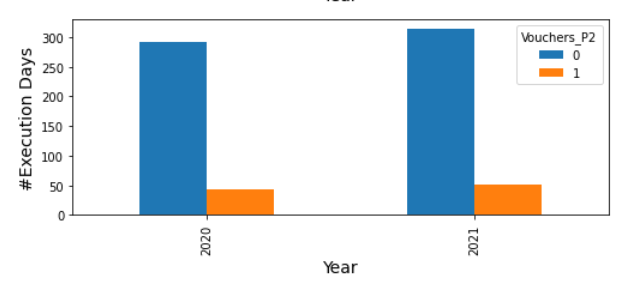


Fig 3.12: Vouchers\_P2 execution

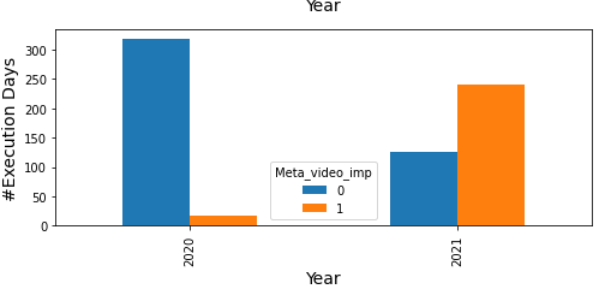
Remarks: In contrast to Vouchers\_P1, company is offering vouchers on very few occasions on platforms other than its own website. But the individual values on these few occasions is very high. This may be the company’s strategy to create massive attraction on festive days or holidays.

#### Meta\_video

#### 

#### Fig 3.12: Meta\_video Fig 3.13: Meta\_video distribution

Figure 3.6: Variable: Account\_user\_count



Remarks: We observe that client has executed for very few days in 2020. Total execution days for whole 2 yrs duration stands at 257 out of 700 days which is still quite low. We observe kind of “burst execution” happening around Week 21- Week 35, 2021 having high intensity for a short duration of time.

#### Meta\_nonvideo

#### 

Figure 3.15: Meta\_nonvideo Figure 3.16: Meta\_nonvideo distribution

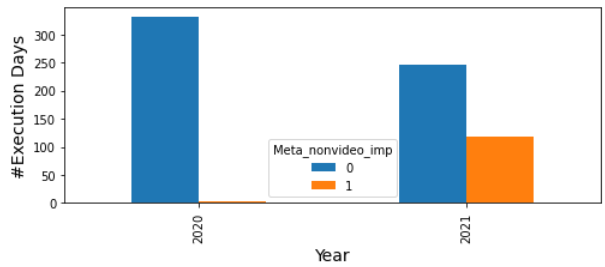


Figure 3.17: Meta\_nonvideo execution

Remarks: We observe that execution in meta\_nonvideo is almost nil in the year 2020. In 2021 also, execution remains less. Total number of execution days remains as low as 122 out of 700.

Most of the executions have happened only between 9th-28th week of 2021.

#### YT

Figure 3.18: YT Figure 3.19: YT distribution

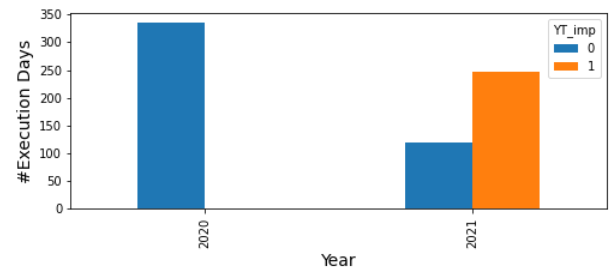


Figure 3.20: YT execution

Remarks: YouTube has no execution in 2020. Though in 2021 company has started investing here.

#### TikTok

Figure 3.21: TikTok Figure 3.22: TikTok distribution

#### 

Figure 3.23: TikTok execution

Execution in TikTok has been minimal with year 2020 having no execution. Company has executed intermittently in 2021 that too in few chunks. It has least days of execution among all media and promo variables. TikTok being a very popular media, it is strange to see less impressions being created there.

#### Platform\_1

Figure 3.24: Platform1 Figure 3.25: Platform1 distribution

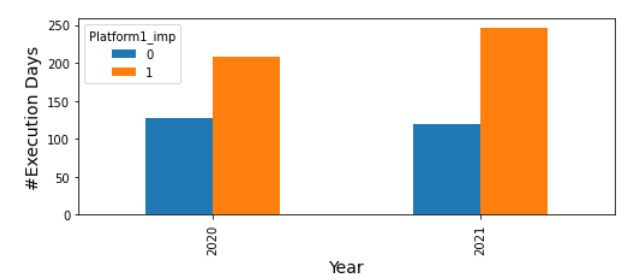


Figure 3.26: Platform\_1 execution

Remarks: We observe most of the executions happening in Platform\_1 are in late 2020s onwards.

No. of days of execution is similar for both the years but the average impressions created in 2021 is much more than year 2020. It shows company is spending regularly in livestreaming

#### Platform\_2

Figure 3.27: Platform\_2 Figure 3.28: Platform\_2 distribution

#### 

Figure 3.29: Platform\_2 execution

Remarks: As compared to previous livestreaming platform(Platform\_1), company is less consistent in terms of execution here as total execution days is just 245 though average impressions is almost similar for both the platforms.

#### Own\_digital\_1

Figure 3.31: Own\_digital\_1.

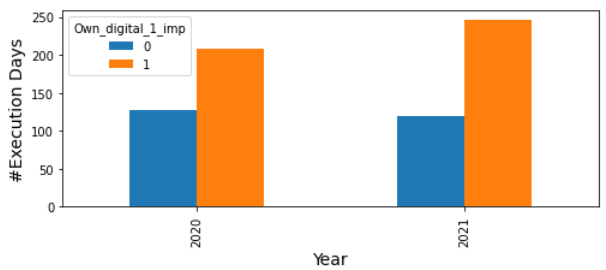
Figure 3.30: Own\_digital\_1 distribution 

Figure 3.32: Own\_digital\_1 execution

Remarks It can be observed that this tactic has second highest days of execution (454 days) after Voucher\_P1 only, both being owed by the company itself

#### Own\_digital\_2

Figure 3.33: Own\_digital\_2 Figure 3.34: Own\_digital\_2 distb

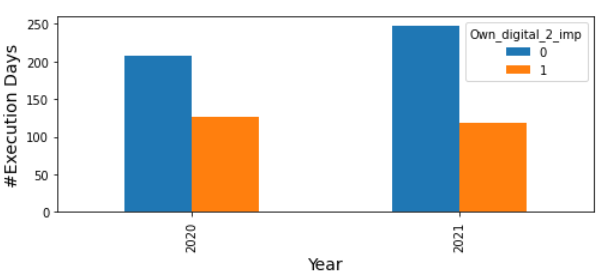


Figure 3.35: Own\_digital\_2 execution

Remarks: Surprisingly, in contrast with other two company owned platforms, this one has less no. of executions. Also, average impressions value is lesser compared to Own\_digital\_1 platform. Maybe company is focusing more on the other owned platforms. We can analyze this more in later sections.

#### Insta\_all

#### 

Figure 3.36: Insta\_all Figure 3.37: Insta\_all disb.

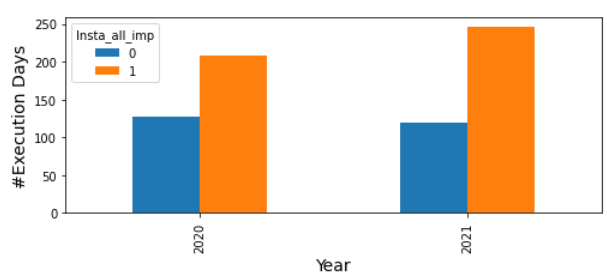


Figure 3.38: Insta\_all execution

Remarks: Insta\_all has highest average impressions as well as highest percentage increase in impressions among all media tactics. No. of execution days is 454. It shows company has very often engaged influencers and other insta channels for marketing. Its effectiveness is to be studied in later sections.

#### Insta\_own

#### 

Figure 3.39: Insta\_own Figure 3.40: Insta\_own distribution

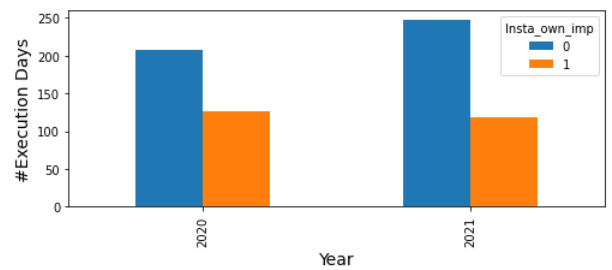


Figure 3.41: Insta\_own execution

Remarks: Unlike Insta\_all, company has not been very active over its own Instagram account in terms of creating impessions. Total Execution days being 245 only out of 700.

### Outlier Treatment

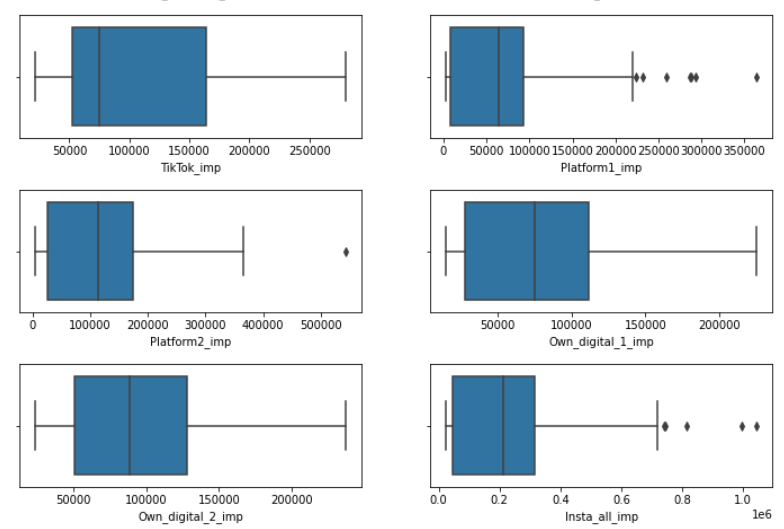
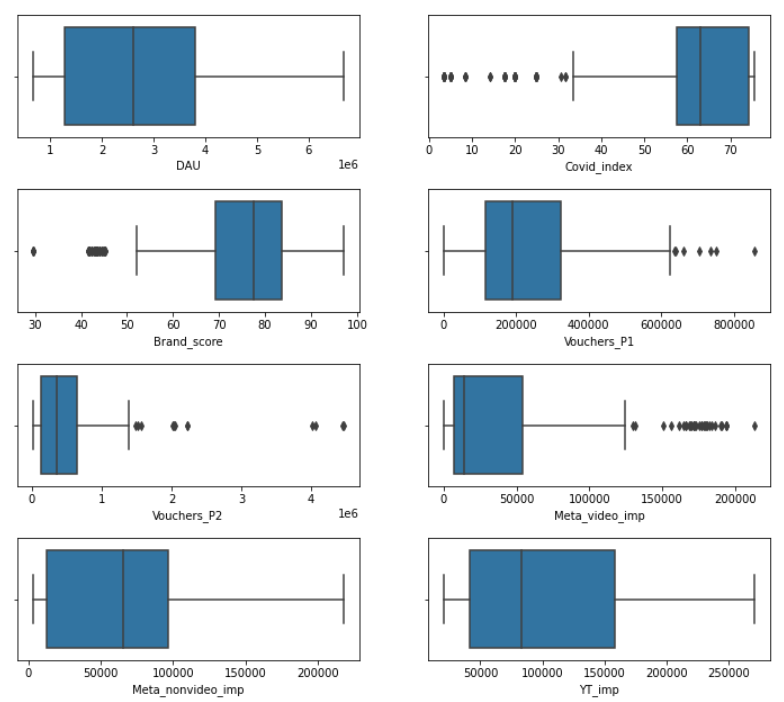


Figure 3.18: Outliers

To study the presence of outliers, we have plotted boxplots. Since, no of days with no execution/impressions are high among all, we have ignored records with zero impressions. Hence, the plots are only explaining the data with non-zero impressions.

Observations:

1. It can be observed that Vouchers\_P1, Vouchers\_P2, meta\_video, Platform\_1, Platform\_2 and Insta\_all have few extreme values However, on going deeper and studying the spends for those few points, values seem to be fine and consistent with the spends. These extreme values in key metrics can be due to aggressive push during festive seasons & holidays.
2. Low values observed in covid\_index are during the starting weeks of 2020 when covid was in its initial phase and lockdowns were not imposed.
3. Low values of brand\_score are also observed. These values were also observed during staring weeks of 2020 after which brand strength increased a lot.

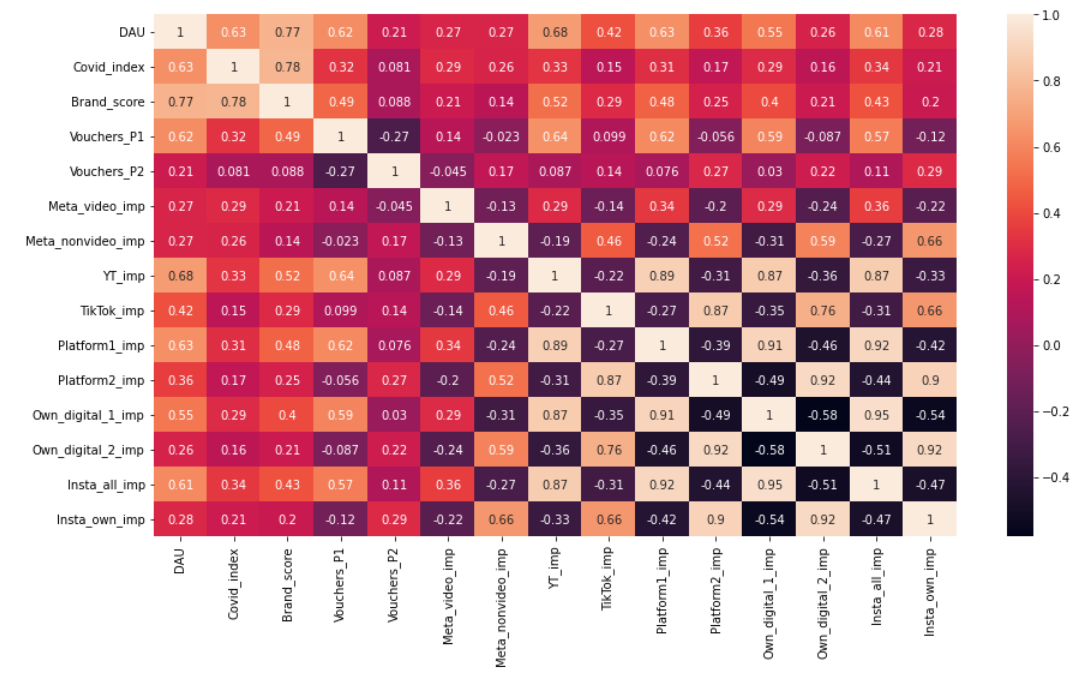
Hence, no outliers are observed in the data.

### Bi-Variate Analysis

#### 3.3.1 Correlation plots with Target Variable

### 

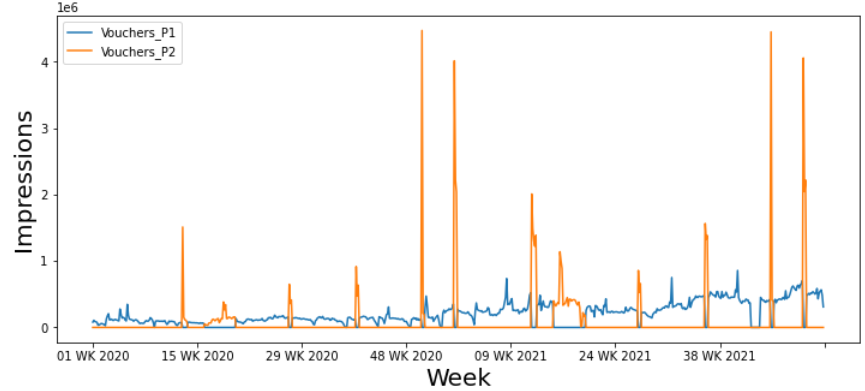
Figure 3.42: Correlation with DAU

Figure 3.43: Correlation Heatmap

Covid\_index, Brand\_score, Vouchers\_P1, Own\_digital\_1 have the highest correlation with the dependent variable. These are highly likely to come in our model.

However, correlation between media and DAU are not of final importance as of now as we will do feature transformation of media variables before modelling in the next chapter.

#### Bivariate Plots

Since, we would expect many combinations for Bivariate analysis, thus we will club the variables and make plots in one go.

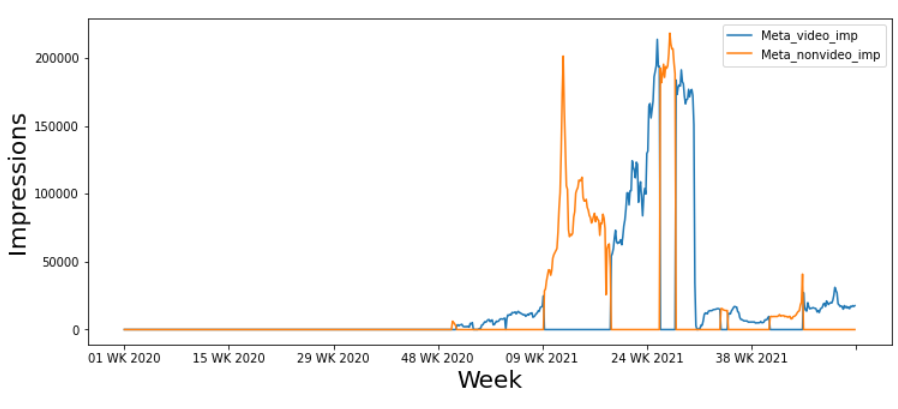
Figure 3.42: Bivariate: Vouchers\_P1, Vouchers\_P2

Figure 3.43: Bivariate: meta\_video, meta\_nonvideo

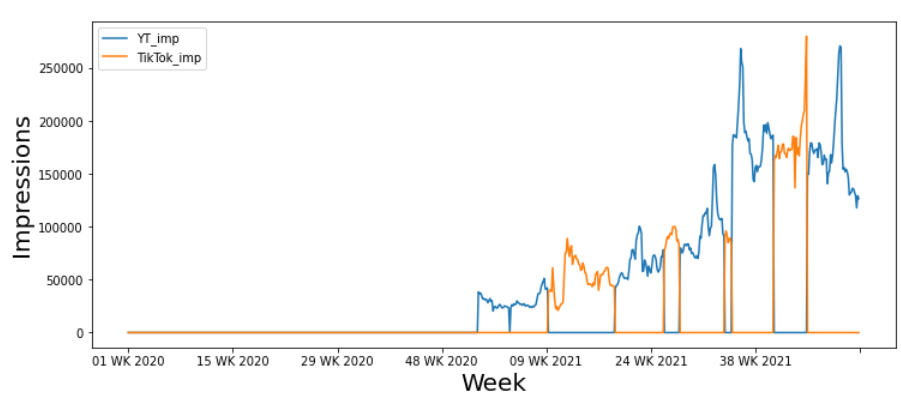
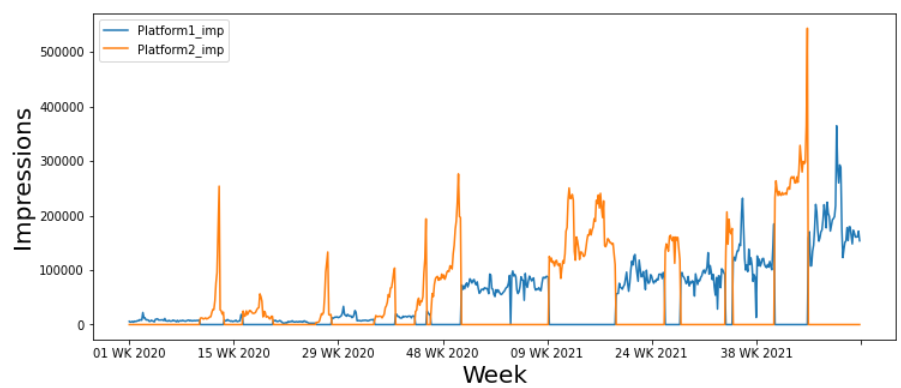
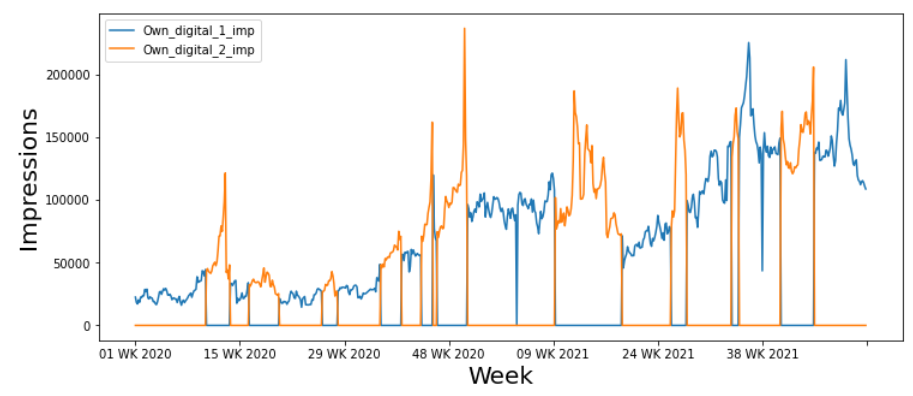


Figure 3.44: Bivariate: YT, TikTok



Figure 3.45: Bivariate: Platform1, Platform2

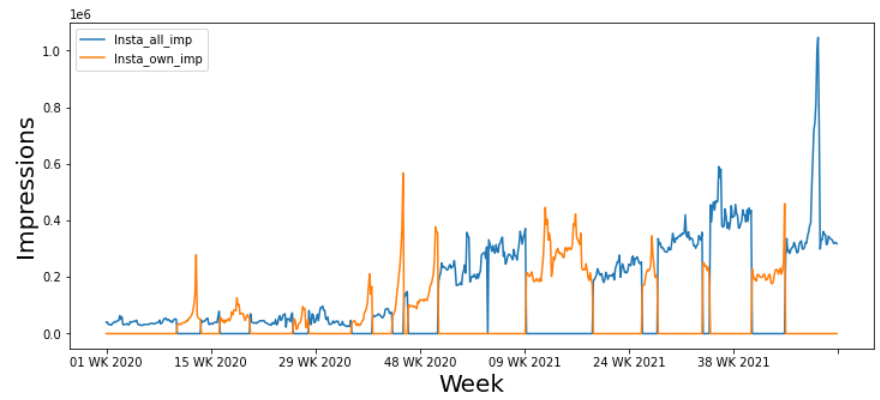
Figure 3.46: Bivariate: Own\_digital\_1, Own\_digital\_2

Figure 3.47: Bivariate: Insta\_all, Insta\_own

When observed carefully by plotting combinations together, it can be noticed that all variables have alternate executions, i.e. only one among each combination is being executed at a particular time.

That implies, at a particular day, company is executing in only one among below pairs:

Voucher\_P1 & Voucher\_P1, meta\_video & meta\_nonvideo, YT & TikTok, Platform1 & Platform2, : Own\_digital\_1 & Own\_digital\_2, Insta\_all & Insta\_own

### Multi-Variate Analysis

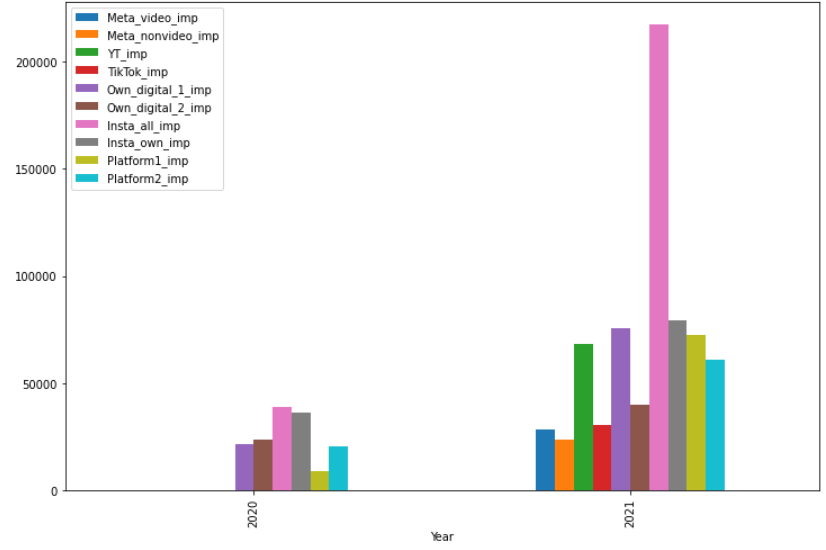
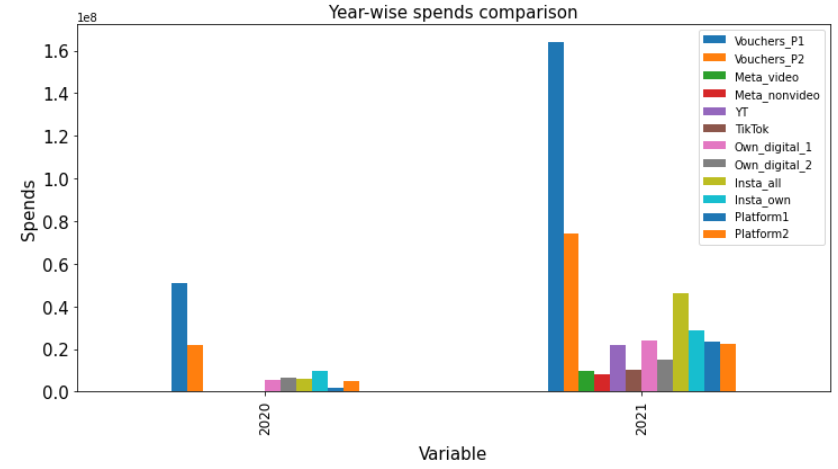


Figure 3.48: Multivariate- Media variables: Average yearly impressions

**Observations:**

* Average impressions of Insta\_all is the highest among all media tactics especially in 2021.
* Overall, average impressions are much higher in 2021 as compared to 2020.
* Few variables like YT, Meta\_nonvideo have TikTok have no presence in 2020.

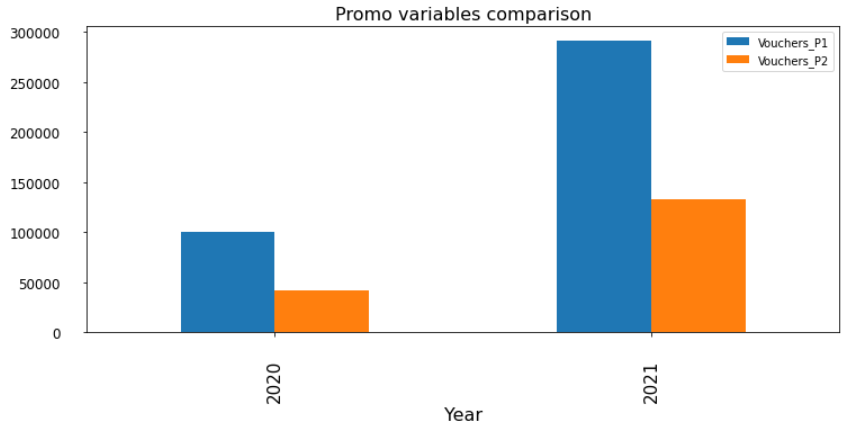
 Vouchers impressions comparison

Figure 3.49: Multivariate: Year wise total sum

**Observations:**

* In both the years, spends on promo variables(Vouchers\_P1, Vochers\_P2) are extremely high as compared to media variables.
* Spendings on advertising has increased a lot from year 2020 to 2021.

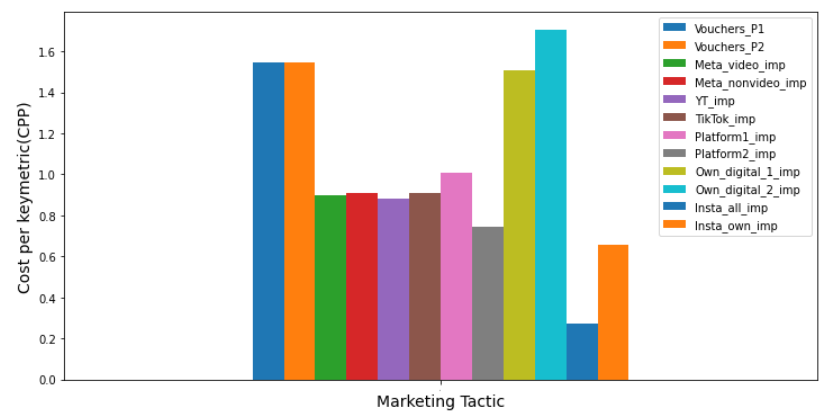


Figure 3.50: Multivariate- Cost per keymetric(CPK)

**CPK = Total spends/ Total keymetric**

It is the average cost to buy/create one key-metric(Impressions, no. of vouchers)

**Observations:**

* CPK for Own\_digital\_2 is the highest among all media as well as promo variables.
* Surprisingly, Own\_digital\_1 & Own\_digital\_2 both being company owned, have high CPK. Generally, the price to advertise in own media channels is much less as compared to other more famous media channels.
* CPK for both voucher variables is almost same, though the spends on Vouchers\_P1 is much high compared to the latter. Higher CPK for promo variables maybe due to the cost involved in compensating discounts vouchers/coupons.
* Lowest CPK is observed for Insta\_all. This means company has utilised Instagram infuencers & channels great cost effectiveness.

.

## Chapter 4

**Modelling**

## 4.1 Feature Transformation

It has been observed that Above the Line(ATL)/Media tactics have non-linear relation with the DAU and sales.

This is so because advertising exhibits several properties like:

1. **Memory effect**: Any impression created among the audience doesn’t end right then and there only but remains in the memory for some time.
2. **Lag effect:** The part of impression or memory carried forward from the advertisement may cause the audience to buy or visit the website later after some time lag. This is said to be the lag effect.

These two effects can be captured using Ad-stocking. Ad-stocking refers to adding some part of the previous adstocked impression to the next timestamp raw impressions.

This is done using Half Life and Retention Rate.

Half life tells the time taken for the keymetric to reduce to half of its present value whereas retention rate depicts what percent of current keymetric shall be carried forward to the next time period.

Ad-stocking can be formulated as:

A(t) = X(t) + retention \* A(t-1)

Where,

A(t) = Ad-stocked impressions at t

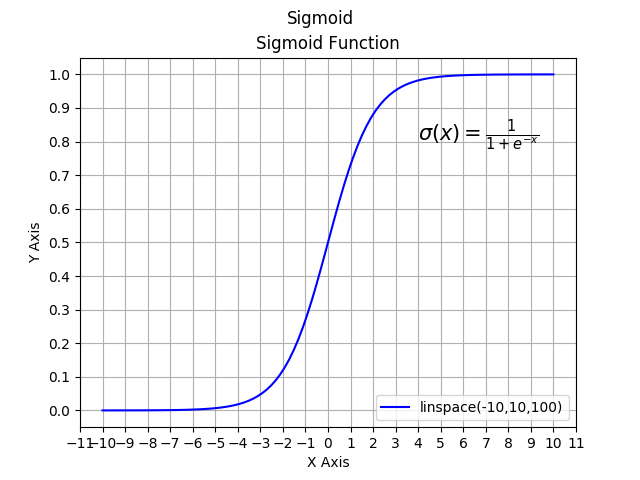
X(t) = New impressions at t

Retention rate = (0.5)^(1/half-life)

For example, for a media having current week impressions of 100k and half life of 2 weeks, half of the impressions, i.e. 50k impressions from that period would remain present in memory even after 2 weeks.

1. **Saturation effect**: As already said, there isn’t a straightforwardly linear relationship between sales and marketing channels. Marketing activity will carry diminishing rates of return as total spend increases beyond a certain point. In other words, we can’t spend an infinite amount of money and expect continued gains. After a certain threshold, every additional dollar generates fewer returns than the previous dollar. This is called the saturation effect.

To capture this, we transform the ad-stocked keymetric using the sigmoid curve as shown below.



Sigmoid is a non-linear function of saturation, steepness and half-life values which together govern the curve characteristic.

Figure 3.51: Sigmoid transformation

Each ATL media is transformed using ad-stocking and saturation to make it more suitable for modelling and capture the overall relation between sales and advertising spends.

Here, for each ATL media, we will transform the raw Impressions using values from below parameter ranges:

Half-life: 3, 5, 7, 10, 15, 21 days

Saturation: 0.1 to 1.25

Steepness: 0.1 to 1.25

We have taken 6\*8\*8 = 384 combinations for each media variable.

For example, meta\_video variable will have combinations like:

meta\_video\_halflife\_7\_saturation\_0.5\_steepness\_0.5

meta\_video\_ halflife\_15\_saturation\_0.9\_steepness\_1 and so on…

In total no. of transformed combinations for 10 media variables = 10 \* 384 = 3840

Total variables for modelling = 3840(media) + 2(promo) + Base variables(Brand\_score, Covid\_index)

= 3844

Now we are ready for modelling.

### 4.2 Modelling Approach

The aim is to come up with a regression model explaining the relation between dependent variable DAU and Base and Incremental variables. For this, we choose Ordinary Least Square Regression.

We input Base & BTL vars as it is but for ATL vars, we do adstocking and saturation transformation.

As we don’t know the best combination of half-life, steepness, saturation for variables, we iterate through the combinations and select the best combination based on the model stats.

The model should be statistically significant with decent MAPE, p-values and R-square values.

We will also try to include most of the variables in the model so as to maximize the spends coverage. Maximizing the spends coverage will help us in making more impactful business recommendations using simulation and optimization.

### 4.3 Modelling Steps

The transformed data has total of 3844 columns and 700 rows.

Out of these, Base (Brand\_score, Covid\_index) and Promo variables (Vouchers\_P1, Vouchers\_P2) will be used as it is with all 340 combinations of each media variables in the modelling process.

Out of 340 combinations of each media variable, one best performing & statistically significant variable shall be included in each step of modelling as shown below:

1. Start with a few variables(base vars and incremental vars)
2. Run Regression algorithm
3. Check model stats:
   * 1. p-values for coefficients
     2. MAPE
     3. R-square/Adjusted R-square
4. Include/Exclude those vars in the model
5. Choose more vars and repeat from step 2

For example: A basic model to start with can be as shown below:

|  |  |
| --- | --- |
|  |  |

Fig 4.1: Modelling Process-1

Suppose we choose to include meta\_video into our basic model. For that, all combinations of meta\_video will be tried with the model and key statistical values shall be checked and the best performing model shall be chosen based on:

* MAPE(Mean Absolute Percentage Error)
* p-values
* R-square
* Spends coverage: as we want to make business recommendations at the end, hence we would like to cover maximum number of variables in our insights and recommendations.



Fig 4.2: Iterations for new model

Suppose a combination “meta\_video\_halflife\_21\_saturation\_0.5\_steepness\_1.0” performs best with the basic model statistically, so will add this to the basic one to get a new model as below.



Fig 4.3: modelling\_process\_2

Similarly, we will add other media variables too and will try to include most of the them.

### 

### 4.4 Model Building

In this section we demonstrate the performance of our chosen regression model.

On following the modelling steps, the best model selected is as follows:

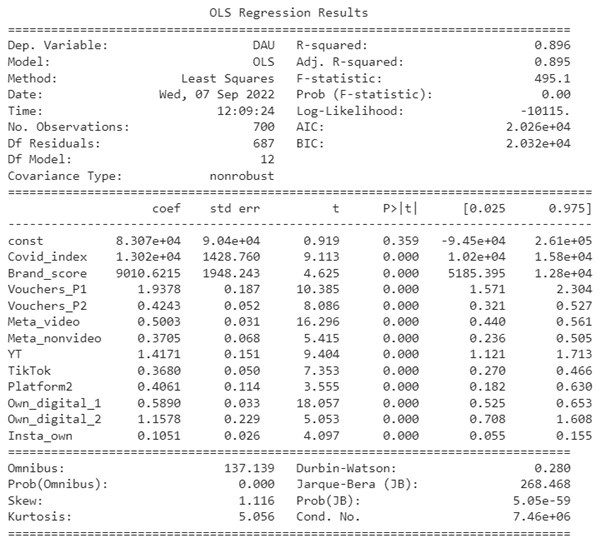


Figure 4.4: Modelling: Modelling results-1

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **S.No.** | **Variable** | **Type** | **Selected combination** | | | **Coefficient** | **Contribution** |
| **Half-life** | **Saturation** | **Steepness** |
| 1 | constant | Base | - | - | - | 83067.2614 | 3.12% |
| 2 | Covid\_index | Base | - | - | - | 13020.6452 | 27.09% |
| 3 | Brand\_score | Base | - | - | - | 9010.62146 | 20.38% |
| 6 | Vouchers\_P1 | Incremental | - | - | - | 1.937774 | 14.5% |
| 7 | Vouchers\_P2 | Incremental | - | - | - | 0.42426 | 1.41% |
| 8 | Meta\_video | Incremental | 15 | 0.6 | 1.25 | 0.500277 | 3.99% |
| 9 | Meta\_nonvideo | Incremental | 15 | 0.7 | 1.25 | 0.370548 | 1.87% |
| 10 | YT | Incremental | 3 | 0.8 | 1.25 | 1.417107 | 2.48% |
| 11 | TikTok | Incremental | 15 | 0.6 | 1.25 | 0.368045 | 1.96% |
| 12 | Platform2 | Incremental | 3 | 0.5 | 1.25 | 0.406076 | 2.18% |
| 13 | Own\_digital\_1 | Incremental | 12 | 0.5 | 1.25 | 0.589046 | 15.61% |
| 14 | Own\_digital\_2 | Incremental | 3 | 0.7 | 0.75 | 1.157778 | 3.51% |
| 15 | Insta\_own | Incremental | 12 | 0.6 | 1.25 | 0.10509 | 1.84% |

Table 4.1: Modelling: Modelling results-1

|  |  |
| --- | --- |
| MAPE | 15.90% |
| R-square | 0.896 |
| Adjusted R2 | 0.895 |
| Max p-value | 35.90% |

Table 4.2: Modelling: Model stats-1

After numerous iterations, we have come up with a model having a total of 12 variables with the regression constant.. Platform1 and Insta\_all could not come in the model.

However, MAPE and max p-value is too high. Let us try to introduce some dummy variables based on records having high residuals in the model. We will be introducing 2 variables:

AHS(Abnormally High values): Based on records having > 1.3\*std. dev. of residuals

ALS(Abnormally Low values): Based on records having < -1.3\*std. dev. of residuals

Let us fit the regression again and see if that works.

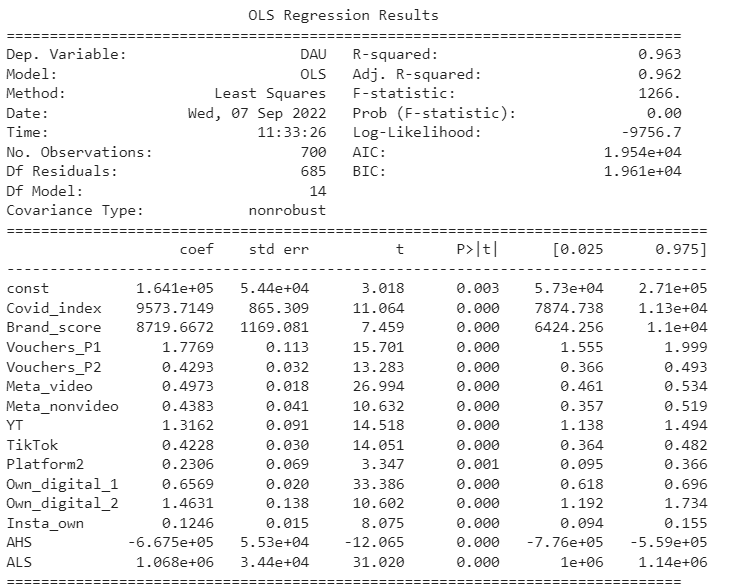


Figure 4.5: Model building- Model results-2

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| S.No. | Variable | Type | Selected combination | | | Coefficient | Contribution |
| Half-life | Saturation | Steepness |
| 1 | constant | Base | - | - | - | 164100 | 6.18% |
| 2 | Covid\_index | Base | - | - | - | 9573.7149 | 19.92% |
| 3 | Brand\_score | Base | - | - | - | 8719.6672 | 19.72% |
| 4 | AHS | Base | - | - | - | -667500 | -1.07% |
| 5 | ALS | Base | - | - | - | 1068000 | 4.48% |
| 6 | Vouchers\_P1 | Incremental | - | - | - | 1.776912 | 13.29% |
| 7 | Vouchers\_P2 | Incremental | - | - | - | 0.4292799 | 1.43% |
| 8 | Meta\_video | Incremental | 15 | 0.6 | 1.25 | 0.4973339 | 3.97% |
| 9 | Meta\_nonvideo | Incremental | 15 | 0.7 | 1.25 | 0.4382608 | 2.21% |
| 10 | YT | Incremental | 3 | 0.8 | 1.25 | 1.316173 | 2.30% |
| 11 | TikTok | Incremental | 15 | 0.6 | 1.25 | 0.4227876 | 2.25% |
| 12 | Platform2 | Incremental | 3 | 0.5 | 1.25 | 0.2305968 | 1.24% |
| 13 | Own\_digital\_1 | Incremental | 12 | 0.5 | 1.25 | 0.6569031 | 17.41% |
| 14 | Own\_digital\_2 | Incremental | 3 | 0.7 | 0.75 | 1.463143 | 4.44% |
| 15 | Insta\_own | Incremental | 12 | 0.6 | 1.25 | 0.1246021 | 2.19% |

Table 4.3: Model building- Model results-2

Model stats:

|  |  |
| --- | --- |
| MAPE | 9.20% |
| R-square | 0.963 |
| Adjusted R2 | 0.962 |
| Max p-value | 0.00% |

Table 4.4: Modelling: Model stats-1

**Inferences:** With the same combination of media variables, we have got a much better model.

* MAPE: MAPE has come down from 15.9% to 9.3% which is a big improvement.
* R-square: It has improved to 0.96 from 0.89.
* P-values: p-value for constant has come down from 36% to 0%, indicating the constant is significant in the model.
* Spends coverage: 86.06%

Let us finalize this model for further process.

**Final results:**

**Regression equation is of the form:**

y = b1.x1 + b2.x2 + b3.x3… + bo + residual

The term coefficient\*Independent variable in the above equation can be referred to as volume/decomposition.

**Contributions:**

As we know**, Total DAU = Incremental DAU + Base DAU**

**Incremental DAU due to var1** = volume of var1 = decomps of v1 = b1.x1 = coefficient1 \* var1

**Total Incremental DAU** = DAU due to incremental variables = vol (v1) + vol (v2) + … + vol(vn)

In above, v1, v2…vn are only incremental vars.

**How much of the total DAU is due to each var?**

Contribution of var1 (%) = (vol of var1)/sum(DAU) \* 100%

DAU(y) = Base DAU + Incremental DAU

Base DAU = DAU(y) - Incremental DAU

Base DAU/volume = Total DAU – Total Incremental DAU

Hence,

**Base DAU = DAU due to base variables**

**= Intercept + residual + base vars volume(Brand, Covid\_index)**

**Uplift volume (%)** = Total vol/Base vol \* 100%

**RoI(Return on Investment)** of var1 = vol of var1/Spends on var1

**Total RoI = {vol(var1) + vol(var2) + vol(var3) + … }/ Total spends**

Let us analyze the above results using plots.

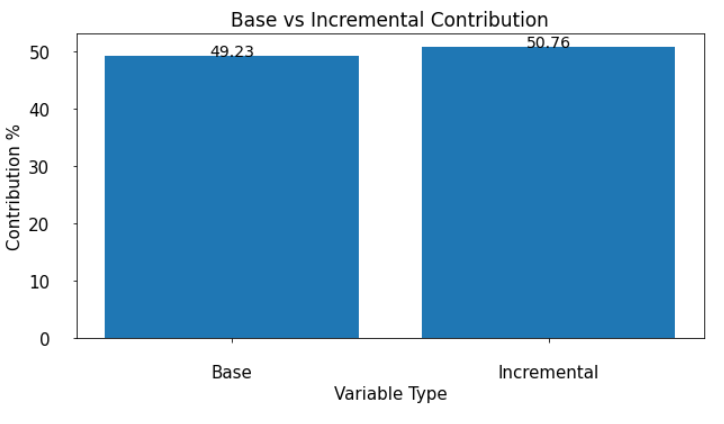


Figure 4.6: Modelling: Base vs Incremental contribution

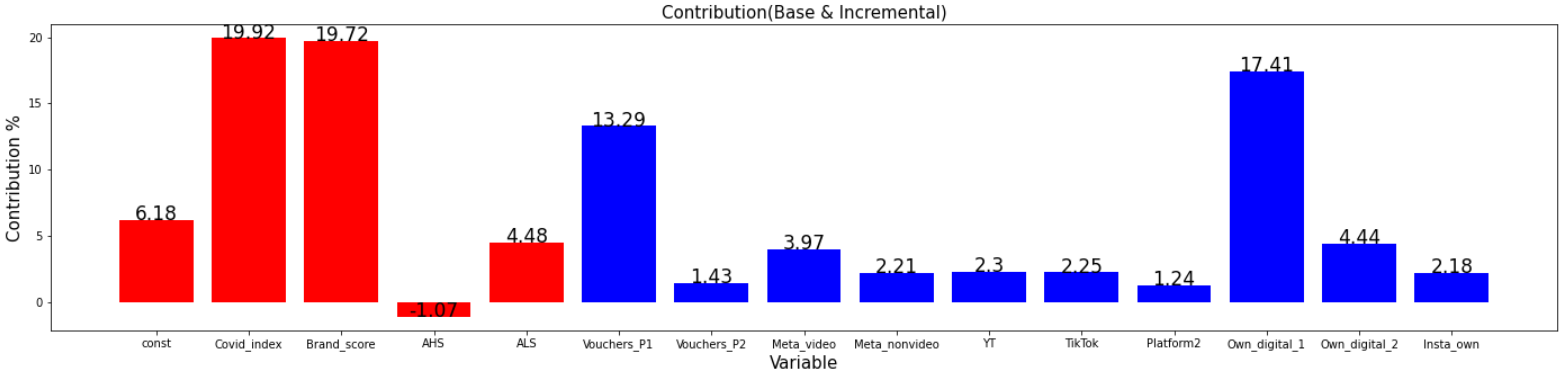


Figure 4.7: Modelling: Conribution(Base and Incremental)

**Observations:**

* Covid\_index has the highest contribution in the model. It signifies that closing of markets due to lockdown had big positive impact on the no. of active users on the website for shopping.
* Brand\_score contribution of almost 20% shows mediocre contribution due to brand. Though, the Brand\_score metric had high values indicating strong positive brand strength, popularity and acceptance among the people, there seems to be a gap between conversions to DAUs due to brand and the brand strength itself. There is a scope of improvement here.
* Own\_digital\_1 platform has very high contribution compared to other media tactics indicating strong and effective digital marketing activities on its platform.
* Promo variable Vouchers\_P1 has significantly high contribution as compared to other promo Vouchers\_P2. Though, the spends on Vouchers\_P1 are also very high. Reason could be the vouchers redeeming cost.

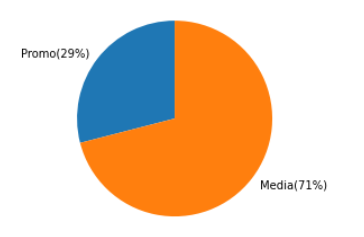


Figure 4.8: Modelling: Promo vs Media Contribution in Total Incremental

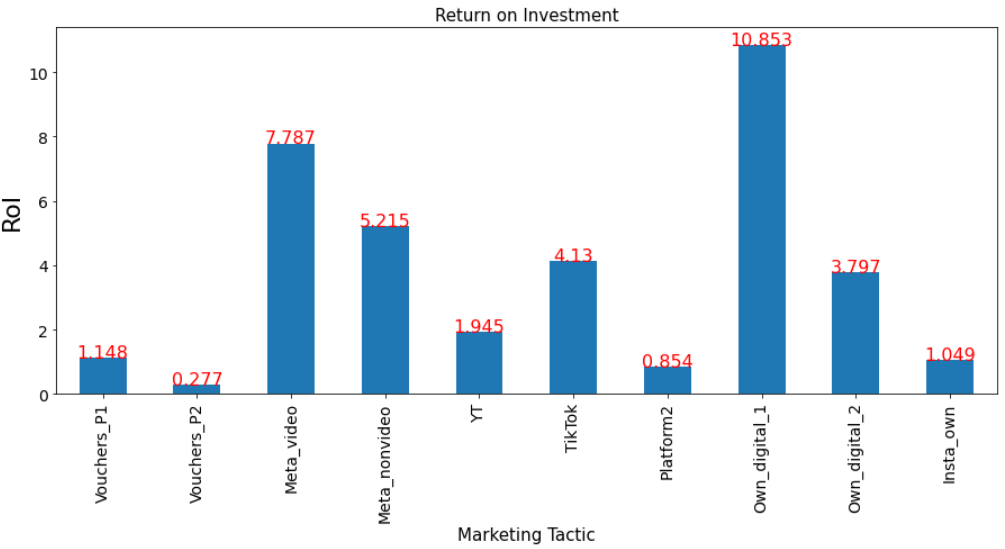


Figure 4.9: Modelling: RoI

Observations:

* RoI on owned media- Own\_digital\_1 is the highest. Own\_digital\_2 also has good RoI.
* RoI on Promo variables are among the least owing to very high spends.
* Meta platform is highly effective in terms of return compared to other media like YT, TikTok, Insta, Platform2

## Chapter 5

**Business Recommendations**

**5.1 Response Curves**

To be able to better interpret our model results and give business recommendations, let us plot the response curves for media variables.

Response curves show the estimated cause and effect relationship between various marketing activities and business performance(Lift here). This helps to forecast, make plans, and optimize the budget allocation.

**Points on the Response curve:**

**Sufficiency Point:** Execution at minimum threshold (Sufficiency). This is the minimum amount of execution required in order to get significant lift.

**Max Marginal:** Execution at max of Marginal Lift. This is the point where rate of change of lift w.r.t. Media Pressure is maximum

**Max Effectiveness:** Execution at Max RoI. At this point the RoI is maximum.

**Saturation Point:** Execution at Saturation. This is the point after which the lift saturates. Meaning increase in lift becomes negligible with ncrease in Media Pressure

**Current Point**: Current Execution by the client

**x-axis of RC:** This is ad-stocked media pressure/impression

**y-axis of RC:** %lift over median of base

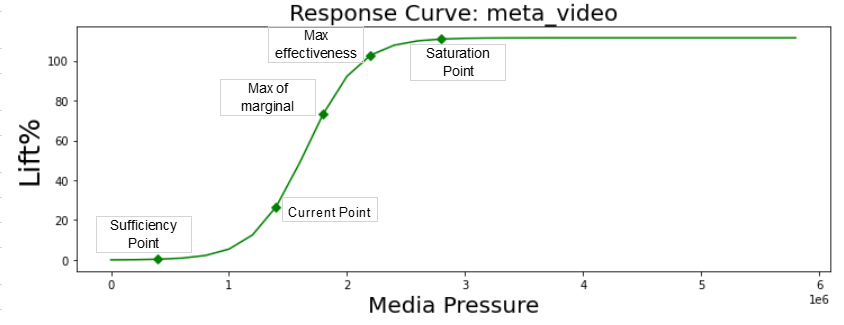


Figure 5.1: Response Curves- meta\_video

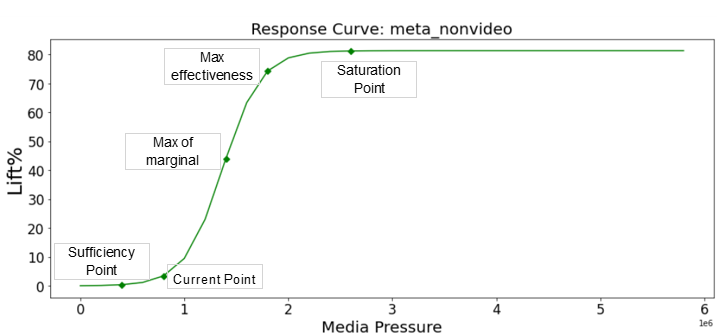


Figure 5.2: Response Curves- meta\_nonvideo

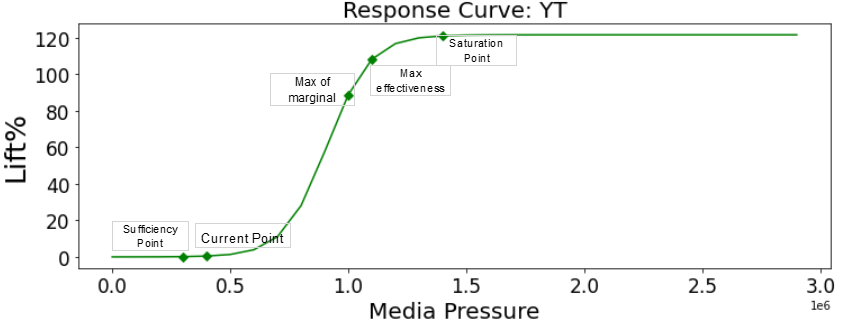


Figure 5.3: Response Curves- YT

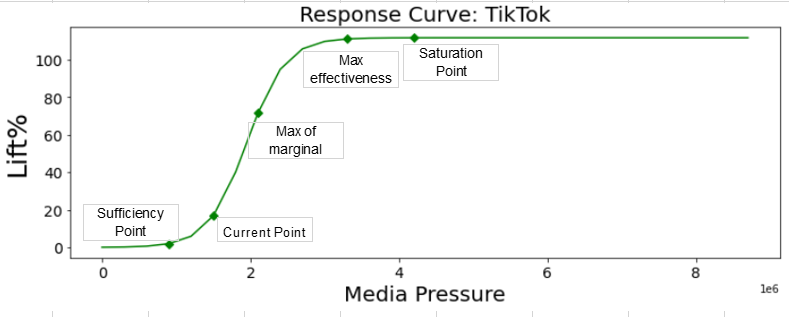


Figure 5.4: Response Curves- TikTok

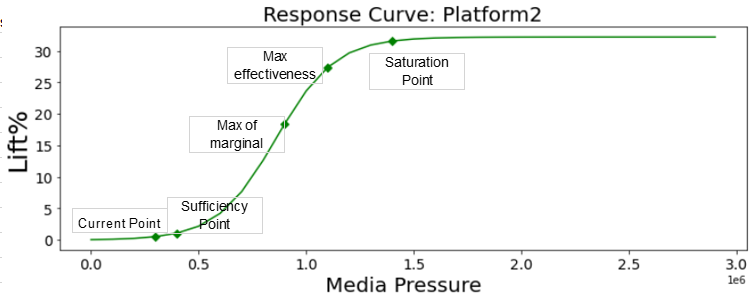


Figure 5.5: Response Curves- Platform2

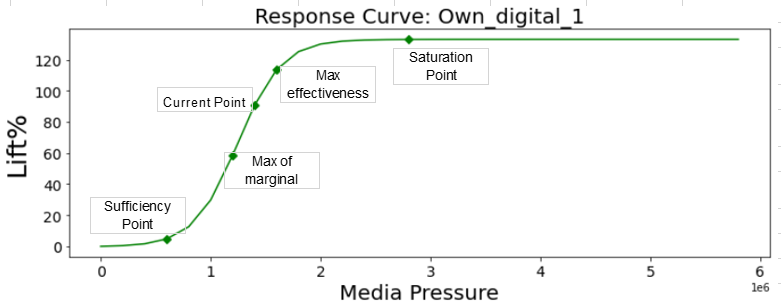


Figure 5.6: Response Curves- Own\_digital\_1

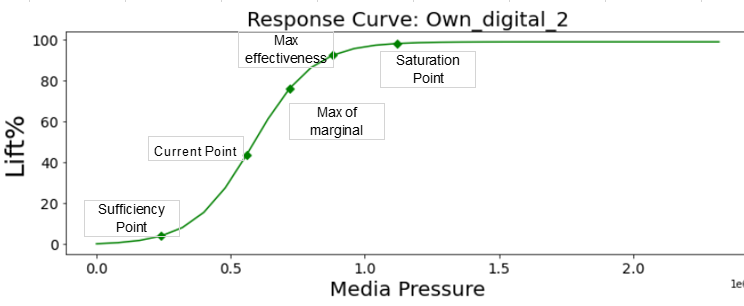


Figure 5.7: Response Curves- Own\_digital\_2

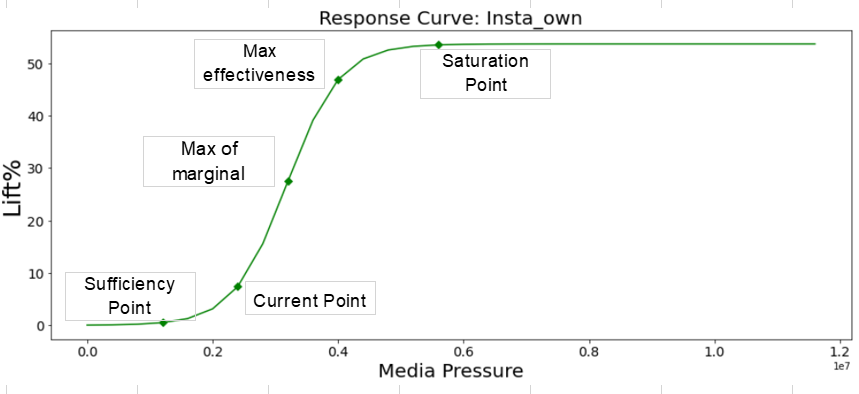


Figure 5.8: Response Curves- Insta\_own

**Observations:**

* The Current Point of meta\_video, meta\_nonvideo, YT, TikTok, Own\_digital\_2, Insta\_own lies between the Sufficiency Point and Max of Marginal. This indicates that the company is falling short of capitalizing full potential from these media. Spends on these should be increased to bring the Current point in the Optimal zone, i.e. between Max of marginal and Max Effectiveness point.
* Current point of Own\_digital\_1 is lying correctly in the Optimal zone.
* Current Point of Platform\_2 is even below than the Sufficiency Point. This means that current spends are not enough to create any significant lift.

## 5.2 Simulator

With the help of simulator, we can play around with the spends to see how the lift and RoI change.

**It answers the ‘What-if’ scenarios.**

**For example:**

What will be the DAU if I spend 10% more in Meta\_video and Meta\_nonvideo?

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **Current spends** | **Change** | **Planned spends** | **Lift current** | **RoI current** | **Lift Planned** | **RoI Planned** |
| **Meta\_video** | 9476996 | 10% | 10424695.6 | 73796842 | 7.786944513 | 83410699.92 | 8.00126 |
| **Meta\_nonvideo** | 7896504 | 10% | 8686154.4 | 41180296 | 5.2150035 | 51389721.55 | 5.91628 |

Table 5.1: Simulator

Like here, we have changed the current spends by 10% to get the Planned spends. These Planned spends are used to find the execution using CPK values. Now this raw execution is ad-stocked and transformed to get the transformed ad-stocked value. This transformed value is then multiplied with coefficient to get the lift. RoI is later calculated using the planned Lift and Planned Spends.

5.3 Optimization

Optimization is a key component in MMM. Optimization helps prioritizing the spends to maximize the RoI. Generally, it can be done in 2 ways:

1. Spend Based(Constant spends): given fixed total spends, what should be the best spends mix to achieve maximum sales.
2. Goal Based(Constant sales): to achieve a particular amount of sales, what should be the optimal spends allocation.

Here, we will try Spend Based optimization.

Hence, we have fixed the total spends and will allow +-30% spends change for each variable.

**Objective Function**: Maximize the lift

**Constraints:**

1. Optimized sum of spends of all variables should be less than or equal to the current spends
2. Optimized Spends of all variables should lie between +- 30% of current spends of variables
3. Spends should be greater than zero

Optimization is done using Non-Linear optimizers.

**Optimization Results:**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **Current spends** | **Optimized spends** | **Change** | **% Spends change** | **Lift current** | **RoI current** | **Lift optimized** | **RoI Optimized** |
| **Vouchers\_P1** | 215225300 | 217236010 | 2010710 | 0.9342% | 246733173.1 | 1.146394839 | 249038241.8 | 1.146394844 |
| **Vouchers\_P2** | 96040980 | 67228686 | -28812294 | -30.0000% | 26599007.92 | 0.276954774 | 20619305.55 | 0.30670398 |
| **Meta\_video** | 9476996 | 12320094 | 2843098 | 30.0000% | 73796842 | 7.786944513 | 100081643 | 8.123448 |
| **Meta\_nonvideo** | 7896504 | 10265455 | 2368951 | 30.0000% | 41180296 | 5.2150035 | 73023070.68 | 7.113476283 |
| **YT** | 22026510 | 28634463 | 6607953 | 30.0000% | 42839409 | 1.944902256 | 74517542.24 | 2.602372611 |
| **TikTok** | 10154550 | 13200915 | 3046365 | 30.0000% | 41941256 | 4.130291938 | 69664849.5 | 5.277274 |
| **Platform2** | 26993350 | 35091355 | 8098005 | 30.0000% | 23046412 | 0.853781098 | 33002461.98 | 0.9404721 |
| **Own\_digital\_1** | 29811210 | 38754572 | 8943362 | 30.0000% | 323545221 | 10.85313951 | 458211641.3 | 11.82342154 |
| **Own\_digital\_2** | 21748830 | 28273479 | 6524649 | 30.0000% | 82579224 | 3.796950181 | 112958203.3 | 3.9952 |
| **Insta\_own** | 38769470 | 27138629 | -11630841 | -30.0000% | 40681072 | 1.04930689 | 28476726.24 | 1.049306 |
| **Total** | 478143700 | 478143700 | 478143700 | 478143658 | 942941913 | 1.972088962 | 1219593686 | 2.55041697 |

Table 5.2: Optimization results

**Observations:**

* + 1. We observe that there is a big scope of improvement in RoI by investing more in overall Media tactics.
    2. Current spending in media tactics is significantly low. Optimization results are aligned with the Response curve results which showed that current investment is even below the minimum threshold for most of the media tactics. RoI will even shoot up as the execution approaches Max marginal and effective point.
    3. Huge investments in Promo variables(Vouchers\_1, Vouchers\_2) has not yielded great results. Also, optimization is indicating for big investment cut for Vouchers\_2 as its highly under performing.
    4. With the help of optimization, by just re-allocating the budgets, Lift can be significantly increased by more than 29%.
    5. Own\_digital\_1, meta\_video, meta\_nonvideo and TikTok are highly promising media tactics in terms of the RoI.
    6. High RoI of company owned Own\_digital\_1 shows that it’s highly effective in acquiring the DAUs in contrast with Own\_digital\_2. Both of these can be further analyzed in contrast to make the latter more effective.
    7. Poor performance of Insta\_own shows less effectiveness of company’s Instagram account in DAUs acquisition. This may be due to less activity or content relevance on Instagram. Reasons could be further analyzed and worked upon.
    8. Optimizer is recommending to reduce spends on Vouchers\_P2 by 30% which is increasing the RoI. Huge investment in Vouchers\_2 has not yielded good results.

THANK YOU!