SALES IMPACT OF INTERACTION BETWEEN ADVERTISING CHANNELS OR MARKET MIX MODEL

Department of Mathematics
IIT Madras

Name of the Project Guides
Prof. Dr. S. Sundar
Dr. Sri Vallabha Deevi(Tiger Analytics)

Name - Shreya Karmakar

Roll No. - MA18M023

INDIAN INSTITUTE OF TECHNOLOGY MADRAS
DEPARTMENT OF MATHEMATICS

NAME: Shreya Karmakar

Roll No: MA18M023

PROGRAMME: M.Sc. /M.Tech.

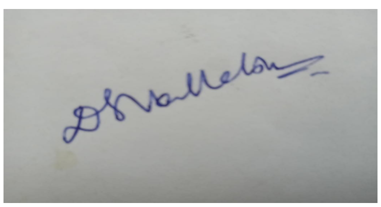
DATE OF JOINING: 28.07.2018

GUIDE (S): Guide Prof. Dr. S. Sundar and Co-guide Dr. Sri Vallabha Deevi

Certificate

This is to certify that the report **"Sales Impact of Interaction Between Advertising Channels"** , submitted by Shreya Karmakar (MA18M023), in partial fulfillment of requirement for the award of the degree of Master of Technology in Industrial Mathematics and Scientific Computing, Indian Institute of Technology Madras, is the record of the work done by her during the academic year 2019-2020 in the Department of Mathematics, IIT Madras, India, under my supervision.

Prof. Dr. S. Sundar
Department of Mathematics
IIT Madras

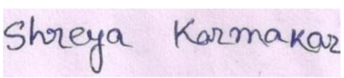


Dr. Sri Vallabha Deevi
Tiger Analytics
Chennai, India

Acknowledgement

I would like to thank **Prof. Dr. S. Sundar** and **Dr. Sri Vallabha Deevi** for their constant help and support, without which the completion of my project for this academic year would be impossible. They have always been there to guide me throughout the course and have patiently solved all the queries associated with the project.

Lastly, I would like to thank the **Department of Mathematics, IIT Madras** and **Tiger Analytics** for providing me the opportunity to work in this exciting field.



Shreya Karmakar
MA18M023
Department of Mathematics
IIT Madras

Abstract

Advertising effectiveness and Return on Investment (ROI) are typically measured through econometric models that measure the impact of varying levels of advertising Gross Ratings Points on sales or on purchase decision and choice. TV advertising has both dynamic and diminishing returns effects on sales, different models capture these dynamic and nonlinear effects differently. This paper focuses on reviewing the econometric rationale behind the popularized Marketing Mix model that allows the inclusion of lagged and non-linear effects in linear models based on aggregate data.

Abbreviations

MMM	Marketing Mix Model
GRP	Gross Rating Points

Table of Contents

Chapter 1 **8**

Introduction 8

1.1 Research and Evaluation 8

Chapter 2 **10**

Basics of advertising and market mix 10

2.1 Terminologies 10

Chapter 3 **12**

Exploratory Data Analysis (EDA) 12

Adstock or Retention 12

Mathematical Transformations – Curve transformations 13

Normalization 14

Chapter 4 **15**

Modeling 15

4.1 Additive Models 15

4.2 Multiplicative models 16

Chapter 5 **19**

Regression Techniques 19

5.1 Linear Regression 19

5.2 Ridge Regression or L2 Norm Regression 19

5.3 Lasso Regression or L1 Norm Regression 20

5.4 Elastic Net 20

Chapter 6 **21**

Model Evaluation Approaches 21

6.1 R2-score 21

6.2 Mean Absolute Percentage Error (MAPE) 22

Chapter 7 **22**

Data Preprocessing 22

7.1 Data Visualization 22

7.2 Adstock or Retention 29

7.3 S-curves and Smoothing curves 30

7.4: Normalization 33

Chapter 8 **34**

Market Mix Models 34

8.1 Additive Models 34

8.1.1 Additive model - I 34

8.1.2 Additive model - II 39

8.2 Multiplicative Models 41

8.2.1 Semi-log Model 41

8.2.2 Log-log model 44

Chapter 10 **48**

Final Results 48

Conclusion 48
Appendix 50

Chapter 1

Introduction

Sales impact of interaction between advertising channels or commonly known as Market Mix Model (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 such as T.V, radio, newspaper, magazine, internet, poster etc.

MMM uses regression techniques and the analysis performed through Regression are further used for extracting key information/insights. Usually MMM is widely used to find *Incremental Sales* (for definition see Terminology part). MMM model breaks down business metrics to differentiate between contributions from marketing and promotional activities (incremental drivers) vs. other (base) drivers. It is often used to optimize advertising mix and promotional tactics with respect to sales revenue or profit.

The techniques were developed by econometricians and were first applied to consumer packaged goods, since manufacturers of those goods had access to accurate data on sales and marketing support. Improved availability of data, massively greater computing power, and the pressure to measure and optimize marketing spend has driven the explosion in popularity of MMM as a marketing tool. In recent times MMM has found acceptance as a trustworthy marketing tool among the major consumer marketing companies. Often in the digital media context, MMM is referred to as attribution modeling.

1.1 Research and Evaluation

Marketing mix modeling, or econometric modeling, provides a tangible situation in which to explore IMC (**Integrated Marketing Communication : "a planning process designed to assure that all brand contacts received by a customer or prospect for a product, service, or organization are relevant to that person and consistent over time."), and the evolving overlap between public relations and marketing, because it pulls statistical information from all communication activities, including public relations, and leads to the strategic coordination of both public relations and marketing activities. One of the underlying issues, however, is pinpointing what to measure. Kitchen (2000) and Schultz (1996) argued that measurement should be based on the effect initiatives have on a company rather than on traditional message evaluation or attitude research. "For the most part, marketing and communication measurement still suffers from an attempt to measure 'outputs,' that is, what is sent out, not 'outcomes' or what impact the marketing or communication activity or investment had" (Schultz & Kitchen, 2000, p. 19). Similarly, other reports and papers have called for a shift to gauging PR effectiveness by measuring outcomes. Weiner and Bender (2006) argued that with marketing mix decisions linked to sales, "it won't be long before media relations campaigns are planned on the basis of what drives sales rather than what drives ink" (p. 46). In this way, one emerging "outcome" area of public relations measurement is tying activities to sales through an evaluation process known as marketing mix modeling. Through market mix modeling, or econometric modeling, companies use sales and marketing data to "evaluate the contribution each element of a marketing program makes to improve sales or share" (Hughes, 2002, p. S4). Marketing mix modeling "uses regression-based techniques to estimate the impact marketing activities are having on sales, and then builds forecasts for future sets of promotional campaigns" (DemandGen, n.d.). Models are built with rich, granular, historical data for regression analysis to correlate factors affecting sales (Doyle, 2004). In short, marketing mix modeling is used to figure out "which part of the ad budget is being wasted, what their optimal spending level is, and what minimum marketing exposure levels should be. Marketing mix modeling demonstrates what is working and what is not (Nardone, n.d.), and modeling variables used run the gamut from pricing decisions and packaging considerations to the weather, economy, and seasonality (Doyle, 2004). Marketing mix modeling is primarily a time-series based evaluation used to predict future performance (Frances, 2005; (Marketing Management Analytics, n.d.), enabling companies to answer "what if" questions about consumer response to brand situations. Frances (2005) argued that good modeling requires consistent data and estimation of data, good forecasting record. It also requires incorporation of consumer response and expectations, controllable variables (i.e. price, advertising spend, etc.) and uncontrollable variables (i.e. seasonality, day of the week, household size, etc.). There is an increasing interest by marketing decision-makers to evaluate their marketing communications using econometric modeling (Hughes, 2002), and marketing mix modeling is becoming a larger part of marketing budgets at companies like Proctor & Gamble (Neff, 2007). In fact, Proctor & Gamble is "uniquely situated to benefit from today's marketing model because it uses its operating efficiency and unmatched clout as the world's biggest marketer to massively outspend its rivals and meticulously measure everything it does, so it knows what is working" (Neff, 2007, p. 1).

Reference: 'Research and Evaluation: Marketing Mix Modeling': https://www.researchgate.net/profile/Brian_Smith68/publication/239591008_Representing_PR_in_the_Marketing_Mix_A_Study_on_Public_Relations_Variables_in_Marketing_Mix_Modeling/links/562948f308ae518e347ca5a2/Representing-PR-in-the-Marketing-Mix-A-Study-on-Public-Relations-Variables-in-Marketing-Mix-Modeling.pdf

** Reference: https://online.purdue.edu/blog/communication/what-is-integrated-marketing-communication-imc

This report is divided into two parts: ‘PART-A’ consists of theoretical explanation of MMM and ‘PART-B’ consists of practical results, some codes etc., which have been done using Python.

PART - A

Chapter 2

Basics of advertising and market mix

2.1 Terminologies

When a person watches or listens in to an advertisement, it stays in his/her memory for a certain time.

- *2.1.1 Impression* – The exposure of a person to one advertisement on a marketing channel. For example, a person being shown a print advertisement in a newspaper.
- *2.1.2 Retention* – Impressions from a certain time period do not vanish after that time period, as people tend to remember advertisements in products that they are interested in. Hence, some impressions are carried over to the next time period from the previous one. This fraction is called retention.
- *2.1.3 Incremental sales*: Sales that can be attributed to marketing activities like TV and print ads, digital spends, price discounts, promotions, social outreach, etc.
- *2.1.4 Base sales:* Base outcome is achieved without any advertisements. It is due to brand equity built over the years. Base outcomes are usually fixed unless there are any economic or environmental changes.
- *2.1.5 Other drivers:* They are a sub-component of baseline factors and are measured as the brand value accumulated over a certain time period due to long-term impact of marketing activities.

A dataset of MMM is based on information collected from different markets. We can apply some methods to get the insights of the data with respect to market sales.

2.2 Approaches:

Two important approaches of MMM are :

a) Additive model and b) Multiplicative model.

These two approaches along with normalization of data gives four models, that will be explained in the report.

Two models are formed under Additive model:

- Normalized (curve transformed (added retention))
- Curve transformed (normalized (added retention))

These two models show percentages total sales, each of Incremental sales and base sale is. Also they verify each other.

Two models are formed under Multiplicative model:

- Semi-logarithmic model or Semi-log model:
    - The coefficients of the independent variables in the model can be interpreted as % change in business outcome (sales) to unit change in the independent variables.
    - Each independent variable in the model works on top of what has been already achieved by other drivers. Hence, they are closer to real-time scenarios.
- Logarithmic model or Log-log model:
    - In Log-Log models, the coefficients are interpreted as % change in business outcome (sales) in response to 1% change in independent variable, like channel impressions.

The main difference between Log-Linear and Log-Log models lies in the interpretation of response coefficients.

Chapter 3

Exploratory Data Analysis (EDA)

Usually data for MMM is listed as weekly basis, i.e. weekly sales, weekly Magazine impression, weekly TV GRP etc. for a one or more year of duration of time. To capture seasonality effects, it is essential to have at least two or more years of weekly data. When we collect raw data from the real world it contains many interruptions like social relationship of customer and seller, unavailability of product due to some unexpected reasons, causing small errors in data. Also due to lack of information the data set may contain null data or missing data.

To overcome these kinds of errors first examine if there are missing values. Since we have week-based data we can't afford row cancellation for missing data(it means there will be a gap of at least one week). Missing values of independent variables can be replaced by the mean of the respective variables. But for a particular variable, if the number of missing entries is a large percentage of it then we should omit that variable in modelling.

Next as a part of EDA a line plot or a scatter plot between each independent variable and total sales can be useful to us. In most cases these raw data plots have absolutely no pattern.

Adstock or Retention

The impact of past advertisement on the current time period is represented as adstock as carry over effect. Since the data is weekly, a small impact of the previous week will be present in the current week impressions. Therefore we need to add some impact of the previous week to our present week impact. This is called Adstock. After adding retention to a variable, entries look like:

$x_j^{(i)} := x_j^{(i)} + \text{retention} \times x_{j-1}^{(i)}$, where $x_j^{(i)}$ is the i-th variable with the j-th week.

Mathematical Transformations – Curve transformations

To capture effects of saturation and critical levels of advertisement required for an impact on sales, weekly data is transformed using logistic or S curves. They are used to capture the linear and non-linear impact of predictor variables.

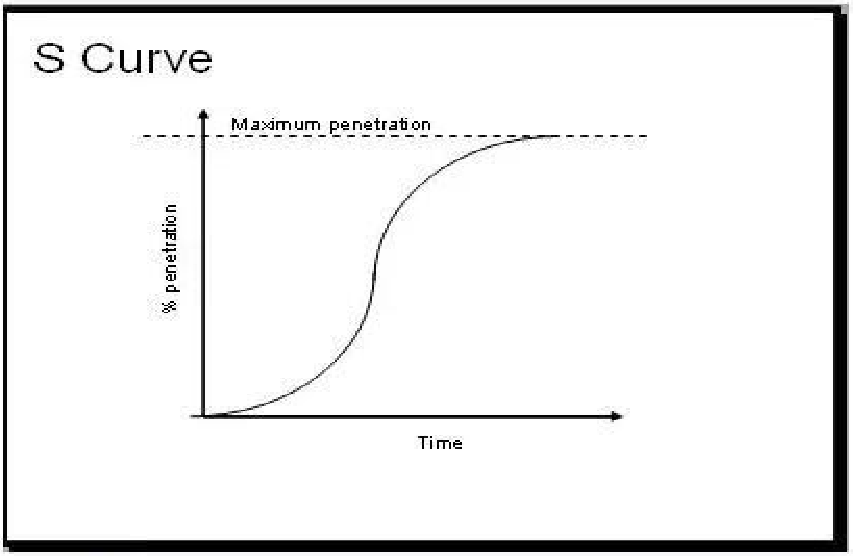


Figure 1 : S-Curve

Reference : https://www.quora.com/What-is-an-S-curve

In the S-shaped response function, sales exhibit increasing returns to scale at low levels and diminishing returns at high levels of marketing effort. This is plausible for advertising which at low levels gets drowned by the noise, and hits an upper limit at very high levels. Besides advertising, the S-shape response function is also used for modeling the effect of shelf space on sales in store.

The S-shape captures the notions of *threshold* and *saturation*. Below the threshold, marketing effort has no impact on sales, and above saturation, there is no further increase in sales. Above or below these bounds, consumers become insensitive to the marketing stimuli. If it truly reflects the response of advertising to sales, the S-Shaped response function has implications on how advertising should be performed : drip versus burst or pulse. An advertising *burst* is a heavy dose of advertising over a short interval. It would ensure that advertising levels cross threshold levels. In contrast drip or continuous advertising is a much lighter weight of advertising spread over a much longer time frame. If thresholds exist, marketers should use less *drip* advertising and more burst or *pulse* advertising ; an approach that falls between burst and drip, with advertising going on and off air over the weeks.

While conceptually appealing, there is not much empirical evidence to support the existence of an S-shape response to advertising effort. It is hard to prove or disprove considering that historical data tends to lie well within these theoretical upper and lower bounds, i.e. if they exist. The *logistic* model depicted below takes a functional form that conforms to the S-shape.

$$\ln \frac{S - S_0}{S^0 - S} = \alpha + \beta \ \ln(X)$$

where $0 < S_0 < S^0$ and S_0 is the intercept and the threshold level, and S^0 is the saturation level.

The elasticity of demand, for a variable with an S-shaped response with sales, follows an inverted bell-shape, starting at 0 at threshold level to a maximum, and back to 0 at the saturation level.

Let us take an example of TV GRP. Increasing TV advertisement will increase sales for a certain extent of time only, since advertisement will create awareness among customers to a certain time. Once the saturation point is reached the advertisement shows less impact on sales and does not create any further incremental awareness among customers since they have already become aware of the product.

*Reference : https://www.ashokcharan.com/Marketing-Analytics/~mx-mmm-sales-response-function.php

Normalization

After adding carry over effects and having S-curve and smoothing transformations different variables containing numerical values are of different ranges. Building a model using them will lead to very small or very large coefficients. To build a robust model, we need to preprocess the data. Normalization is a way to bring all data into a similar range, to estimate coefficients better. The following approaches are usually employed for normalization.

- **Standardization or Z-score normalization**: In this normalization approach variables are rescaled in such a way that their distribution becomes a Standard Normal one i.e. N(0,1). Z-scores of the samples are calculated as

  $$x\_new = \frac{x_{old} - \mu}{\sigma}$$

  where μ is mean and σ is standard deviation

  Standardizing makes the variable centred at 0 with standard deviation 1.

- **Min-Max Scaling:** An alternative method of normalization. In this approach data is scaled from 0 to 1, giving smaller standard deviation which can suppress the effect of outliers. Rescaling computed as:

  $$x\_new = \frac{x_{old} - x_{min}}{x_{max} - x_{min}}$$

- **Simple Feature Scaling:** This method rescales a variable from 0 to 1. Rescaling formula:

$$x\_new = \frac{x_{old}}{x_{max}}$$

After the data is preprocessed using these techniques, it can be used to build a MMM model.

Chapter 4

Modeling

While much of the literature has been devoted to various ways of handling the lags, there have been very few papers that have looked at how an incorrect model specification can affect the estimation of advertising carryover effects. Since real world data is unlikely to adhere to a specific functional form, using any particular function to model the data may lead to misspecification problems. Besides, most of the time media variables do not show much variation within the data. Years of experience enables marketing managers to more or less optimize their media levels and we only observe advertising levels that vary within a very small range. Therefore, for that kind of advertising data, several models can be a good fit. And one can use a number of different functions to model that data.

Reference for above paragraph: https://www.lexjansen.com/nesug/nesug12/sa/sa16.pdf

- **4.1 Additive Models**

The dependent variable 'Sales' is a continuous value; hence a linear additive regression model can be applied. Linear regression can be applied when the dependent variables are continuous and the relationship between the dependent and independent variables is assumed to be linear. The relationship can be defined using the equation:

$$y = \beta_0 + \beta_1 x^{(1)} + \beta_2 x^{(2)} + \ldots\ldots\ldots\ldots + \beta_n x^{(n)} + \varepsilon$$

Here 'y' is the dependent variable, to be estimated, $x^{(i)}$ are the independent variables and ε is the error term. β_i (i=0, 1…n) are the regression coefficients. The difference between the observed outcome y and the predicted outcome $\hat{y}$ is known as a prediction error. Regression analysis is mainly used for:

- Estimating a variable based on other related variables - i.e dependent variable is estimated by one (simple regression) or more independent variables (Multiple regression).

As a result of additive model we calculate Base sales and Incremental sales as:
Base sales = β_0

Incremental Sales = $\sum_{i=1}^{n} \beta_i x^{(i)}$

Hence *percentage of incremental sales :* $\frac{\sum_{i=1}^{n} \beta_i x^{(i)}}{\beta_0 + \sum_{i=1}^{n} \beta_i x^{(i)}} \times 100\%$

i.e. let us suppose a company stops its advertisements on different channels. Then, the sales are like to be β_0. Due to advertisements, an additional sales of $\sum_{i=1}^{n} \beta_i x^{(i)}$ happened. Usually incremental sales ranges in 0-30% and base sales range in 70-100%. However, this method does not perform well on large amounts of data as it is sensitive to outliers, multicollinearity and cross-correlation.

- **4.2 Multiplicative models**

Additive models imply a constant absolute effect of each additional unit of explanatory variables. They are suitable only if businesses occur in more stable environments and are not affected by interaction among explanatory variables. But in scenarios such as when pricing is zero, the sales will become infinite.
To overcome the limitations inherent in linear models, multiplicative models are often preferred. These models offer a more realistic representation of reality than additive linear models do. In these models independent variables are multiplied together instead of added.

There are two kinds of multiplicative models:

- ***4.2.1 Semi-logarithmic models:*** In the first type of multiplicative models – the exponents of explanatory factors are multiplied.

$$y = \exp(\beta_0).\exp(\beta_1 x^{(1)}).\exp(\beta_2 x^{(2)})\ldots\ldots\ldots\ldots \exp(\beta_n x^{(n)}) . \exp(\varepsilon)$$

Products of exponents can be rewritten as

$$y = \exp(\beta_0 + \beta_1 x^{(1)} + \beta_2 x^{(2)} + \ldots\ldots\ldots\ldots + \beta_n x^{(n)} + \varepsilon)$$

Logarithmic transformation linearizes the model form, which in turn can be estimated as an additive model.

$$\ln(y) = \beta_0 + \beta_1 x^{(1)} + \beta_2 x^{(2)} + \ldots\ldots\ldots\ldots + \beta_n x^{(n)} + \varepsilon$$

The only difference between the additive and semi-logarithmic model is that the latter one has the modeled variable 'sales' expressed as a logarithmic transformation. However, the implications are significant.

The semi-logarithmic model provides benefits in several critical aspects. First, each factor – the explanatory variable in the model – works on top of what has been already achieved by other factors. The response coefficients, interpreted as % change in sales to unit change in explanatory variables,

$$100 \,.\, \beta = \% \; \frac{\Delta Dependent\_Variable}{\Delta Explanatory\_Variable},$$

are closer to real life situations. As brand sales get stronger on the back of improved distribution, or temporarily due to the season or price promotions, advertising uplift will be proportionally stronger compared to the situation, when sales are at lower levels.

Second, the point above implies that the modeled variable is driven by interactions of all factors in the model. In other words, explanatory factors have a synergistic effect on the modeled variable,. In many situations, the variables of marketing mix indeed interact and their simultaneous effect enhances sales more than the sum of the two effects occurring alone, like a consumer being simultaneously exposed to a video advertisement on television as well as a print advertisement in newspaper.
In some cases, the semi-logarithmic model is flexible enough to capture relationships with non-linear shapes.

- ***4.2.2 Logarithmic models:*** In the second type of multiplicative model, explanatory variables are multiplied together.

$$\mathbf{y = exp(}\beta_0\mathbf{)\ .\ (}x^{(1)^{\beta_1}}\mathbf{)\ ....\ (}x^{(n)^{\beta_n}}\mathbf{)\ .\ exp(}\varepsilon\mathbf{)}$$

We can linearize the model through logarithmic transformation.

$$\ln(y) = \beta_0 + \beta_1 \ln(x^{(1)}) + \ldots\ldots\ldots + \beta_n \ln(x^{(n)}) + \varepsilon$$

In this case, all model variables are subject to logarithmic transformation. Benefits of this model are very similar to those of semi-logarithmic model. The main difference lies in the interpretation of response coefficients. While in the semi-logarithmic models β's stand for % change in sales in response to unit change of explanatory variable, here they denote elasticity

$$\beta = \% \frac{\Delta Dependent\_Variable}{\Delta Explanatory\_Variable},$$

i.e. % change in sales in response to 1% change in explanatory variable. This implies constant elasticity of sales to explanatory factors. The constant elasticity feature is sometimes pointed out as a drawback of the logarithmic model. It is up to modeler's considerations if the assumption of constant elasticity is optimal.

In semi-logarithmic models, elasticity cannot be directly estimated but can be calculated from the coefficient as $\beta \cdot x$ for every time period. It increases in absolute value with the explanatory variable.

The following Table summarizes three discussed functional forms. We assume a simple model with one dependent and one explanatory variable, *y* and *x*, respectively, with β as response coefficient.

Table 4.1: MMM functional forms

Model form	Dependent variable	Independent variable	interpretation of β	Marginal effect of Δx
Additive	y	x	$\beta = \frac{\Delta y}{\Delta x}$	β
Semi-logarithmic	ln(y)	x	$100.\,\beta = \% \frac{\Delta y}{\Delta x}$	$y.\,\beta$
Logarithmic	ln(y)	ln(x)	$\beta = \frac{\%\Delta y}{\%\Delta x}$	$\frac{y.\beta}{x}$

Reference http://sd-group.com.au/en/blog/additive-versus-multiplicative-marketing-mix-model

Chapter 5

Regression Techniques

Four types of regressions can be used to fit the data.

- **5.1 Linear Regression**

Multiple linear regression in which the target value is expected to be a linear combination of the features. In mathematical notation, y is the predicted variable and $x^{(i)}$ s are independent ones,

$$y = \beta_0 + \beta_1 x^{(1)} + \beta_2 x^{(2)} + \ldots\ldots\ldots\ldots + \beta_n x^{(n)}$$

Across the module, we designate the vector y=(β_1 , β_2 …… β_n) as coefficients and β_0 as the intercept.

In this model Residual Sum of Square,

$$\text{RSS} = \sum_{j=1}^{m} (y_j - \hat{y}_j)^2$$

where, $\hat{y}$ is the predicted value of y using *linear regression*, which *tries to minimize RSS* .

- **5.2 Ridge Regression or L2 Norm Regression**

When the number of independent variables is high there is a high chance of overfitting. Ridge regression is a way to balance the *bias-variance trade off.* Ridge imposes a penalty on the size of the coefficients. To fit to training data Ridge tries to minimize

$$\text{RSS} = \sum_{j=1}^{m} (y_j - \hat{y}_j)^2 + \alpha \; \| y \|_2^2$$

Where,

- $\| . \|_2$ is l_2-norm
- α is a penalty term. The larger the value of α, the greater the amount of shrinkage and thus the coefficients become more robust to collinearity.

- **5.3 Lasso Regression or L1 Norm Regression**

The lasso is a linear model that estimates sparse coefficients. It is useful in some contexts due to its tendency to prefer solutions with fewer non-zero coefficients, effectively reducing the number of features upon which the given solution is dependent. For this reason Lasso and its variants are fundamental to the field of compressed sensing. Under certain conditions, it can recover the exact set of non-zero coefficients. Mathematically, it consists of a linear model with an added regularization term. The objective function to minimize is:

$$\text{RSS} = \sum_{j=1}^{m} (y_j - \hat{y}_j)^2 + \alpha \ \| y \|_1^2$$

where, $\| \ . \ \|_1$ is l_1-norm.

- **5.4 Elastic Net**

Elastic Net is a linear regression model trained with both l_1 and l_2-norm regularization of the coefficients. This combination allows for learning a sparse model where few of the weights are non-zero like *Lasso*, while still maintaining the regularization properties of *Ridge.* We control the convex *combination of l_1 and l_2* using the l1_ratio(in python) parameter.

This method is useful when there are multiple features which are correlated with one another. Lasso is likely to pick one of these at random, while elastic-net is likely to pick both. A practical advantage of trading-off between Lasso and Ridge is that it allows Elastic-Net to inherit some of Ridge's stability under rotation. The objective function to minimize is in this case

$$\text{RSS} = \sum_{j=1}^{m} \left(y_j - \hat{y}_j\right)^2 + \alpha \ \| y \|_1^2 + (1-\alpha) \ \| y \|_2^2$$

Where, $0 \leq \alpha \leq 1$.

Chapter 6

Model Evaluation Approaches

- **6.1 R^2-score**

R-squared (R^2) is a statistical measure that represents the proportion of the variance for a dependent variable that's explained by an independent variable or variables in a regression model. R-squared explains to what extent the variance of one variable explains the variance of the second variable. R^2 ranges from 0 to 1; value near 0 shows very low accuracy of the model, whereas value near 1 shows the model is quite good and is highly accurate.

- Calculation of R^2

To calculate R^2 first we need to calculate Relative Squared Error (RSE).

Formula of RSE:

$$\text{RSE} = \frac{\textit{unexplained variation}}{\textit{total variation}}$$

$$= \frac{\sum_{l=1}^{n} (y_l - \hat{y}_i)^2}{\sum_{l=1}^{n} (y_l - \overline{y_i})^2}$$

where y_l is given, $\hat{y}_i$ is predicted value of y_l, $\overline{y_i}$ is the mean value of a given y-vector.

$$R^2 = 1 - \text{RSE}$$

$$= 1 - \frac{\sum_{l=1}^{n} (y_l - \hat{y}_i)^2}{\sum_{l=1}^{n} (y_l - \overline{y_i})^2}$$

- R-Squared is a statistical measure of fit that indicates how much variation of a dependent variable is explained by the independent variable(s) in a regression model.

- **6.2 Mean Absolute Percentage Error (MAPE)**

The mean absolute percentage error (MAPE) is a statistical measure of how accurate a forecast system is. It measures this accuracy as a percentage, and can be calculated as the average absolute percent error for each time period minus actual values divided by actual values. The mean absolute percentage error (MAPE) is the most common measure used to forecast error, and works best if there are no extremes to the data.

$$\text{MAPE} = \frac{1}{n} \sum_{l=1}^{n} \left| \frac{y_l - \hat{y}_i}{y_l} \right|$$

R^2 score and MAPE will be used to get the best model and predicted incremental sales from the model.

PART – B

We have Market Mix data, collected from Tiger Analytics.

This part deals with anonymised Advertisement and Sales data in a business context obtained from Tiger Analytics. This is raw data and requires pre-processing before building a model.

Chapter 7

Data Preprocessing

- **7.1 Data Visualization**

The dataset is a record of total weekly sales of a product of a retail company for two years of duration. Also it is listed with weekly impressions of different advertising channels. . The data has twenty-five columns, of which twenty-two are advertising channels. They are:

Table 7 . 1 List of Columns

MarketName
WeekDay
Internet_Advertising_Email_Blast_impressions
Newspaper_ROP_impressions
Magazine_Advertising_impressions
Newspaper_Preprint_impressions
Internet_Adv_Display_impressions
Broadcast_Ads_Radio_impressions
Internet_Adv_Retargeting_impressions
Internet_Adv_Social_impressions
Broadcast_Ads_Digital_Radio_impressions
Broadcast_Ads_TV_impressions
Direct_Mail_Multipage_Customer_impressions
Direct_Mail_Multipage_NCA_impressions
Direct_Mail_Post_Card_Customer_impressions
Direct_Mail_PostCard_NCA_impressions
Direct_Mail_New_HomeOwner_impressions
Newspaper_Preprint_Smart_Source_impressions
Internet_Adv_Search_Non_Brand_impressions
Internet_Adv_Search_Shopping_impressions
Internet_Adv_Search_Brand_impressions
Internet_Adv_Search_Local_impressions
Internet_Adv_Search_overall_impressions
Direct_Mail_impressions
Total_Weekly_Sales

Note here that 'Total_Weekly_Sales' is the target variable. Also 'MarketName' is the name of the market and 'WeekDay' is the day of the week on which accumulated data for a total week is taken. All other columns are weekly channel impressions. To feed the channels to machine learning models, the twenty-two channels are considered and the other three columns are removed from the list of columns.

Now we'll look at the numerical behavior of the channels. Some channels consist of *Null* values. In some channels, the number of null values is a very small percentage of the total time period., In such cases, it is sufficient to replace null values by the mean value of the channel. the table of mean, median, standard deviation, upper quartile, lower quartile etc. of channel impressions and total spend is shown in the table below.

Table 7 . 2 Statistical description of market data

Statistical Measurements	Internet_Advertising_Email_Blast_impressions	Newspaper_ROP_impressions	Magazine_Advertising_impressions	Newspaper_Preprint_impressions
count	390	390	390	390
min	518707.0719094225	442047.65900236304	209.25374116719	119162.11204906972
std	454290.98282526166	299169.47336335876	2800.1841510308595	152783.45076326348
min	1623.2940191486111	,2.4824914234927578	,0.0243811169602566	3.455343369109599e-05
25%	85860.49569714276	107429.43163865496	1.9471006211779458	12855.684339568079
50%	439454.5913776811	452868.44764106383	4.064286438630646	67755.4625327792
75%	846454.5119415075	737699.8240302232	,6.1453689243999	158021.40181682608
max	2053658.3361611385	1168983.1417699691	39152.381174291564	1283790.3868192486

Statistical Measurements	Internet_Adv_Display_impressions	Broadcast_Ads_Radio_impressions	Internet_Adv_Retargeting_impressions	Internet_Adv_Social_impressions
count	390	390	390	390
min	415642.17674757075	21.662215440111712	208773.26570422883	164553.3201565926

std	1355276.6061676543	59.8986212241601	309123.3879262009	127402.49720277259
min	8.119072520329507e-06	3.051055559011233e-05	10.874861011371673	9.788018096201156
25%	0.0038737060954961494	0.0012181271732995469	73618.6780843426	71302.23918577899
50%	0.008160801291812413	0.002569466076218793	137531.54950886694	124573.05526323657
75%	327759.75228983286	7.597580982717423	228154.9802136588	244400.30963898287
max	13729054.008288553	522.7778209691811	3033215.3184844963	826600.6489826076

Statistical Measurements	Broadcast_Ads_Digital_Radio_impressions	Broadcast_Ads_TV_impressions	Direct_Mail_Multipage_Customer_impressions	Direct_Mail_Multipage_NCA_impressions
count	390	390	390	390
min	108106.09898359122	0.3691762140748046	2651.535762513836	12192.320722751547
std	271908.9714060775	0.929157910250965	11440.104935775462	29006.819150838815
min	0.04119587944115605	5.129524062006549e-06	0.005593528637543275	0.0009007097994745179
25%	55.870529977221715	0.003193169619447682	0.2787276209917399	0.2395993177569523
50%	106.26546421443945	0.006102827422754542	0.4982389620589177	0.4645604262044102
75%	160.66568917615092	0.008916668027753476	0.7229339108065547	4685.6005627462155
max	2199607.725894767	4.066434072654647	119796.70084124574	172707.606374359

Statistical Measurements	Direct_Mail_Post_Card_Customer_impressions	Direct_Mail_PostCard_NCA_impressions	Direct_Mail_New_HomeOwner_impressions	Newspaper_Preprint_Smart_Source_impressions
count	390	390	390	390
min	12713.720935723724	19825.538855916366	681.2869034746741	106926.16083336453
std	29874.154705373385	38624.808259694684	1871.6275122565687	270122.22759882745
min	0.00018230503892087662	0.0014618275423707288	0.0001337935454738329	1.402539676197336
25%	0.07450657982007675	0.17516467060714017	0.01149300117553404	115.45289082710131
50%	527.9083587634129	1039.9579960436672	0.023602757330520793	240.52336425898474
75%	8465.093171926308	25714.104432748456	226.426840010747	357.0912247576789
max	253519.92087137673	284935.8176633802	17029.56117036954	1560861.1190612891

Statistical Measurements	Internet_Adv_Search_Non_Brand_impressions	Internet_Adv_Search_Shopping_impressions	Internet_Adv_Search_Brand_impressions	Internet_Adv_Search_Local_impressions
count	390	390	390	390
min	13850.684216735826	752092.7577807643	26238.738413126604	153568.20863780563
std	17641.800813080503	873928.8485060864	21186.120355921204	615157.8009675469

min	0.001479658482233005	10938.336799992087	6236.8401536751535	45.02912972816365
25%	1619.5050710248456	217112.11215706985	13766.75729342775	1137.2869436506899
50%	6953.003538405935	438018.0648655944	20844.21633818508	2328.720521182788
75%	19905.754492719614	901277.1745448677	31563.234105984026	3615.5076883205293
max	122298.00613359234	6133102.718571986	163157.59812018627	5394464.5868054265

Statistical Measurements	Internet_Adv_Search_overall_impressions	Direct_Mail_impressions
count	390	390
min	943629.7191880628	48063.47866504422
std	1253046.5361670954	61289.27566807693
min	25105.573312647106	0.0025324956871989037
25%	239281.2174504768	1463.8749074342072
50%	496619.3439026275	21724.423247304938
75%	962986.6717274907	80153.46241636181
max	8439206.762986982	392885.05104759906

From ‘max’ and ‘min’(and other statistical counts) in the tables we can see that channels are in different numerical ranges. We need to bring them in the same range to have a good regression fit. The plots below show the pattern of channel impressions vs. total_sales, to understand the behaviour.

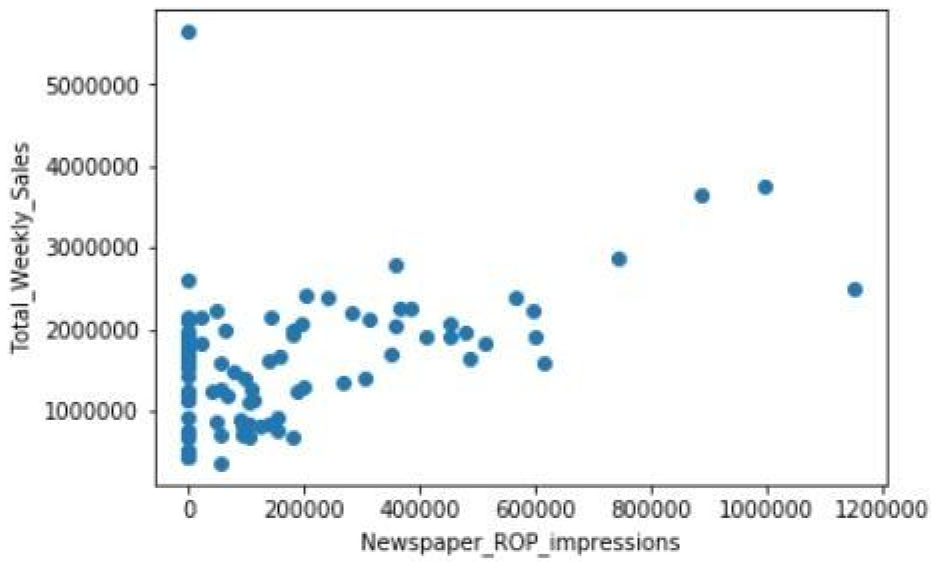

5000000
4000000
3000000
2000000
1000000
Total_Weekly_Sales
0
200000
400000
600000
800000
1000000
1200000
Newspaper_ROP_impressions

Fig - 7.1

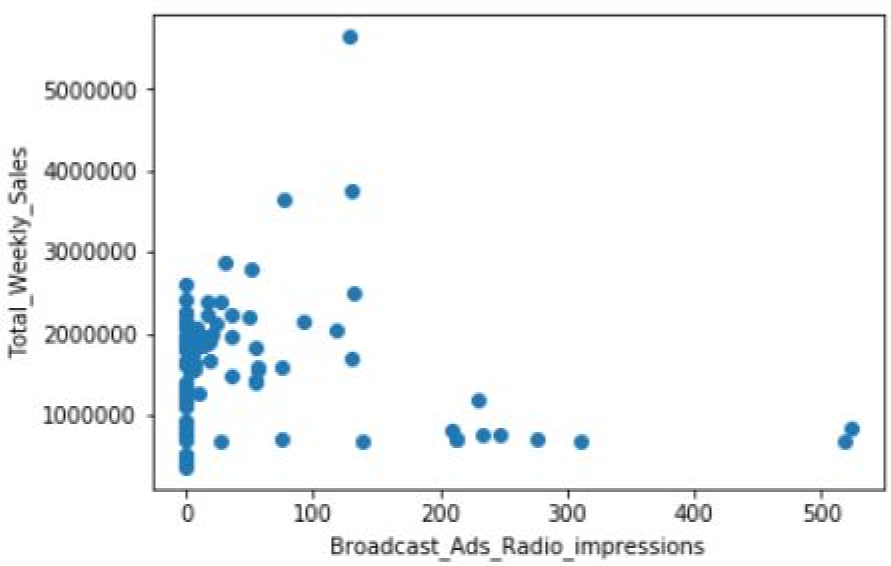

5000000
4000000
3000000
2000000
1000000
Total_Weekly_Sales
0
100
200
300
400
500
Broadcast_Ads_Radio_impressions

Fig – 7.2

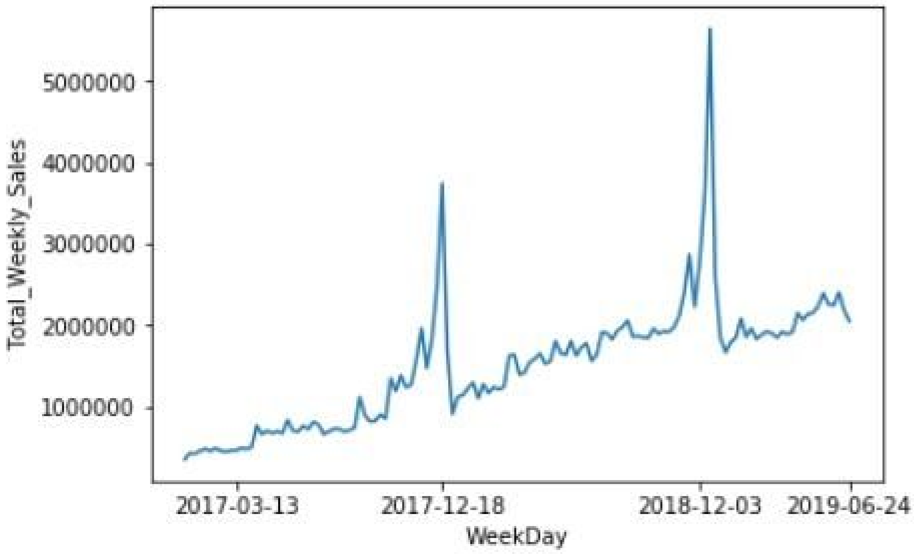


Fig – 7.3

In two years duration plots each of channels vs. total sales graph shows no pattern(here only two are shown). But Weekdays vs. Total_Sales has a pattern, showing that once a year (in December) an amount of high sales takes place.

To remove the irregularity we shall apply *Adstock or Retention* to the channels and then convert them into *S-curves and Smoothing curves*.

- **7.2 Adstock or Retention**

Adstock of the previous week is added to the present week because the impact of previous week's advertisement watched or listened by a customer remains in memory . Usually it remains under 40% . We have twenty-two channel impressions for which ad-stocks have to be estimated. The following methodology is used - for each channel we use four values of retention 10%, 20%, 30%, 40% and find the one which has the largest correlation with total sales, and only that retention is taken into account to feed the model. Here an example of ***20% retention*** addition for a channel is shown by the following chart.

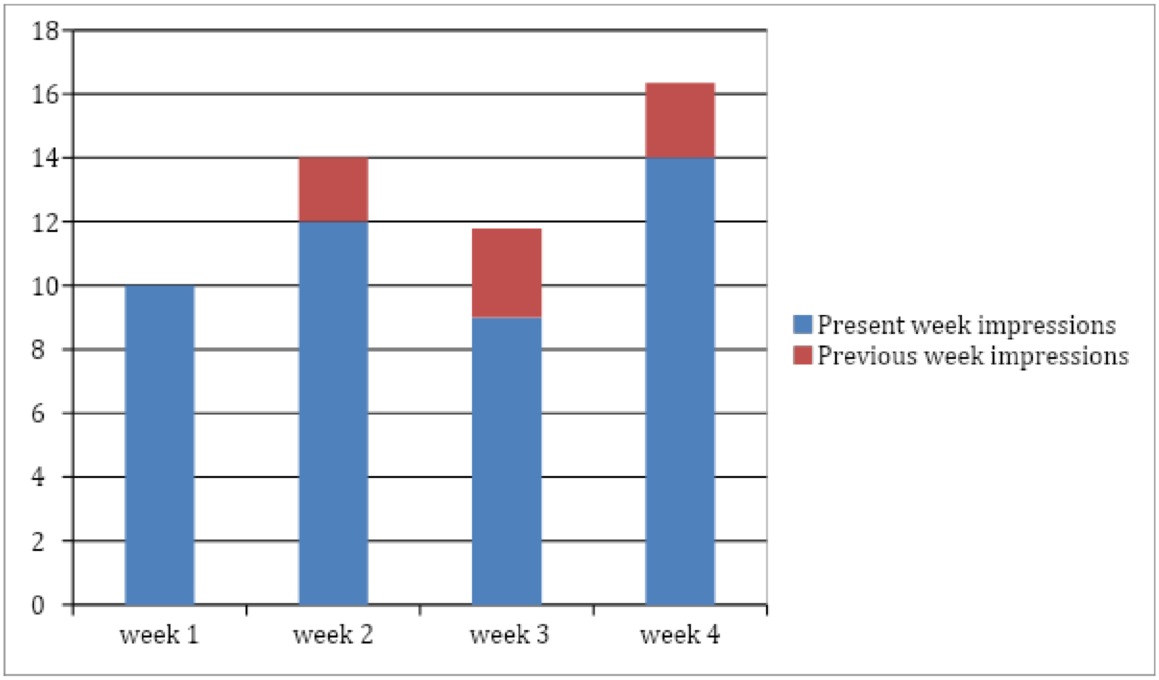


Figure 2 : total impressions = present week impressions + previous week impressions taken as 20%

- **7.3 S-curves and Smoothing curves**

To cut out irregularity of channels five types of curves are used.

- *7.3.1: Gompertz function*

The Gompertz curve or Gompertz function is a type of mathematical model for a time series. It is a sigmoid function which describes growth as being slowest at the start and end of a given time period. The right-hand or future value asymptote of the function is approached much more gradually by the curve than the left-hand or lower valued asymptote. This is in contrast to the simple logistic function in which both asymptotes are approached by the curve symmetrically. It is a special case of the generalized logistic function.

Formula:

$$f(t) = a\ e^{-be^{-ct}}$$

where,

- a is an asymptote, since
- $a\,e^{-be^{-ct}} = a\ e^{0} = a$

- b sets the displacement along the *x*-axis (translates the graph to the left or right). Symmetry is when b =log(2).
- c sets the growth rate
- e is Euler's Number (e = 2.71828...)

- *7.3.2: Logistic transformation*
  A logistic function or logistic curve is a common S-shaped curve (sigmoid curve) with equation

$$f(x) = \frac{L}{1+e^{-k(x-x_0)}}$$

where,

- e = the natural logarithm base (Euler's number)
- x_0 = the x value of the sigmoid's midpoint
- L = the curve's maximum value
- k = the logistic growth rate or steepness of the curve x is in domain of real numbers - ∞ to + ∞ .

- *7.3.3: Chapman Richards transformation*
  Growth functions in general describe the change in size of an individual

Formula:

$$f(x) = x_{max} (1- e^{-kt})^p$$

where,

- x_{max} = the maximum value of variable
- $(1 - e^{-kt})^p$ = a modifier reducing the maximum growth variable to its current state at time t.
- k = an empirical growth parameter scaling the absolute growth rate.
- p= empirical parameter is related to complex structure

- *7.3.4: Logarithmic transformation*

  The log transformation is, arguably, the most popular among the different types of transformations used to transform skewed data to approximately conform to normality. If the original data follows a log-normal distribution or approximately so, then the log-transformed data follows a normal or near normal distribution.

Formula:

$$f(x) = \ln(x)$$

where, ln = normal log

- *7.3.5: Negative Exponential transformation*

A transformation which allows for different saturation levels or learning rates is the negative exponential form, is

$$f(x_j) = 1 - e^{-vA_t} + \lambda f(x_{j-1})$$

Where,

- A_t = advertising at time t
- λ =a smoothing parameter bounded between 0 and 1.
- v= accommodates different saturation levels. The ceiling effect is also accounted for, since it is a bounded function. For values of v close to zero it will represent constant returns or no learning at all.

Following is an example of adding 40% adstock and different curve transformations plots on 'Newspaper_ROP_impressions'. We have manipulated the curves to get comparable ranges for S-curves and Smoothing curves.

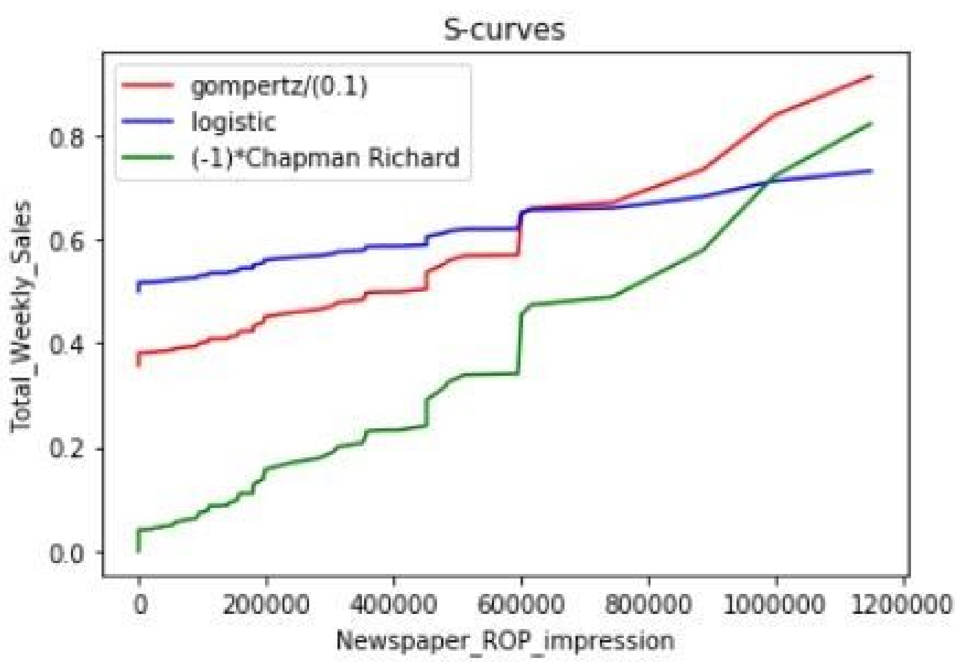


Fig 7.5

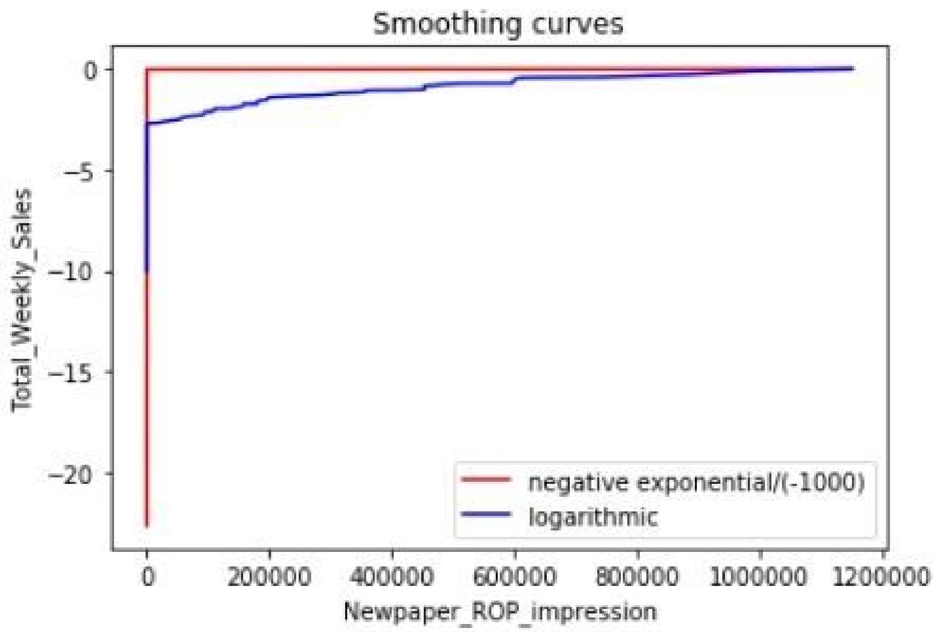


Fig 7.6

- **7.4: Normalization**

  From the above figures it seems that different transformations have different ranges. Normalization brings them to the same range. For the particular data, proper numerical values for all the channels are obtained by the normalization formula

$$x_j^{(i)} := \frac{x_j^{(i)}}{max(x^{(i)})}$$

for i = 1,2 …….130, j=1,2……22

Chapter 8

Market Mix Models

- **8.1 Additive Models**

8.1.1 Additive model - I

The very first additive model is linear regression which defined as:

$$\text{Total\_Weekly-Sales} = \beta_0 + \beta_i \sum_{i=1}^{22} x^{(i)}$$

where $x^{(i)}$'s are channels with modification:

Normalized (curve transformed (added retention))

One can notice for each channel we are talking about four retentions and five curve transformations, from which only one combination is fed to the regression model.

- ***Finding the combination for a channel***

  a ***4*** $\times$ ***5*** **matrix** is created for each channel, whose entries are correlation between modified variable and total weekly sales, i.e. if a_{ij} is an entry of the matrix then,

  a_{ij} = correlation between sales and *(j-curve transformation(i-added retention)) of the channel*

where,	i = 1	if retention = 10%
	2	if retention = 20%
	3	if retention = 30%
	4	if retention =40%
and	j = 0	if Gompertz
	1	if Logistic
	2	if Chapman Richards

3 if Logarithmic
4 if Negative Exponential

then the *highest absolute value* is taken to the consideration. i.e. for a channel if a_{41} is the highest element in the matrix. we take *Gompertz_transformation(40% retention)* modified channel impressions to make a regression model. The following one is an example of the matrix with respect to the channel 'Direct_Mail_impressions'

```
[[0.13834638806457034, 0.13846207526404544, -0.1372573850866555, 0.09335977871837496, -0.0064
45731732282475], [0.15464574448213295, 0.15723700182704786, -0.1531191360682424, 0.1333832699
6949426, -0.057043245465039624], [0.17348996425908045, 0.17808062822681567, -0.17174742555141
248, 0.17242865685543154, -0.12274340979747589], [0.19550913059190692, 0.2017655706765885, -
0.19373005808262367, 0.21262098777138358, -0.195316559416214]]
```

and *logarithmic (40% retention))* modified 'Direct_Mail_impressions' is taken, since the highest absolute value is a_{43}.

- After creating twenty-two 4×5 matrixes, like above discussed manner, list of all modified channels are written in the Table 8.1

Table 8 . 1 Channels with Adstock and Curve transformation

Channel Name	Retenti on	Curve transformation
Internet_Advertising_Email_Blast_impressions	0.4	logarithmic
Newspaper_ROP_impressions	0.4	logistic
Magazine_Advertising_impressions	0.2	logarithmic
Newspaper_Preprint_impressions	0.4	logistic
Internet_Adv_Display_impressions	0.3	richards

Internet_Adv_Retargeting_impressions	0.4	negative exponential
Internet_Adv_Social_impressions	0.4	negative exponential
Broadcast_Ads_Digital_Radio_impressions	0.4	logarithmic
Broadcast_Ads_TV_impressions	0.4	logarithmic
Direct_Mail_Multipage_Customer_impressions	0.4	logistic
Broadcast_Ads_Radio_impressions	0.4	negative exponential
Direct_Mail_Multipage_NCA_impressions	0.4	negative exponential
Direct_Mail_Post_Card_Customer_impressions	0.4	logarithmic
Direct_Mail_PostCard_NCA_impressions	0.4	richards
Direct_Mail_New_HomeOwner_impressions	0.4	logistic
Newspaper_Preprint_Smart_Source_impressions	0.4	logarithmic
Internet_Adv_Search_Non_Brand_impressions	0.4	logistic
Internet_Adv_Search_Shopping_impressions	0.4	logistic
Internet_Adv_Search_Brand_impressions	0.4	logarithmic
Internet_Adv_Search_Local_impressions	0.4	logarithmic
Internet_Adv_Search_overall_impressions	0.4	logarithmic
Direct_Mail_impressions	0.4	logarithmic

Plugging the above modified and normalized channels in

- Linear
- Ridge
- Lasso
- Elastic Net

we can find predicted total weekly sales and consequently incremental sales for all the four models and will take the model with highest R^2-score into consideration as the best predictive model.

- **Results**

8.1.1.1: Linear Regression

For reference of code see Fig - 8.1 in Appendix. Result is shown in Table 8.2

8.1.1.2: Ridge Regression

Ridge regression takes α (alpha) as a penalty parameter. Trial and error for different values of alpha gives the best result.

8.1.1.3: Lasso Regression

Like Ridge α need to be set for Lasso regression. Best result for α-R^2-score combination is shown in the following table.

8.1.1.4: Elastic Net

Elastic Net, a combination Ridge and Lasso, we need to set α as well as ratio for both of above two regressions.

Results of above regressions:

Table 8.2 Regression results for model-I

Regression model	value of α	l1-ratio	R^2-score
Linear	NA	NA	0.8823393331674816
Ridge	0.01	NA	0.879937197874308
Lasso	0.01	NA	0.8824687706356047
Elastic Net	0.00001	0.5	0.8824688674364977

Comparing all R^2-scores, *Elastic Net* can predict total_weekly_sales best for the dataset. Consequently Incremental Sales can be found. Recorded weekly sales vs. predicted weekly sales are shown in the following figure.

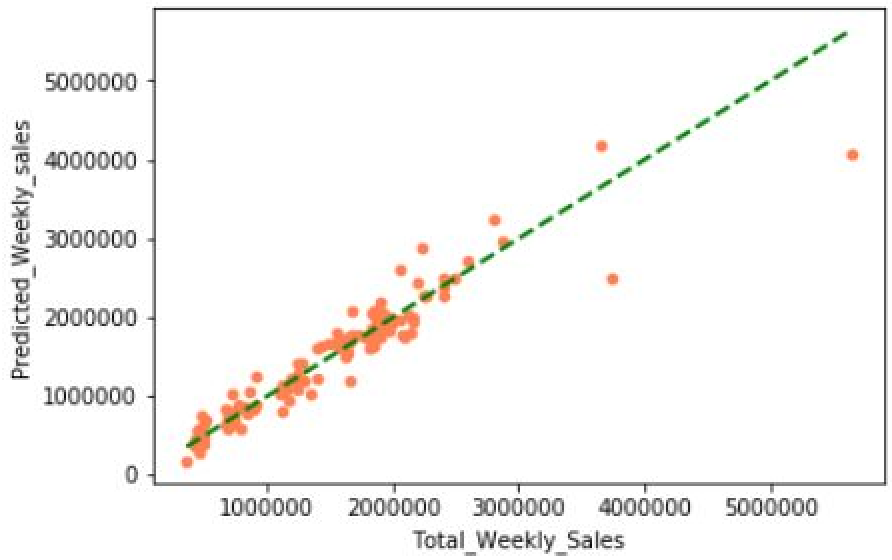


Figure 3 : Elastic Net recorded vs. predicted sales

Predicted Incremental Sales is 17.43% and the result is less than the usually expected maximum value of 30%

Sales impact of advertising channels as percentage of incremental sales is shown in table 8.3 below.

Table 8 . 3 Sales impact of advertising channels

Name of Channels	% of Incremental sales
Magazine_Advertising_impressions	32.43
Direct_Mail_Multipage_Customer_impressions	27.59
Broadcast_Ads_TV_impressions	39.97
Others Channel Impressions	0.01

8.1.2 Additive model - II

Next, to verify above model a second model is formed with a different modification of channels as:

curve transformed(normalized(added retention))

For each channel a correlation matrix is formed like discussed above. And best correlated converted channels with curve and retention is taken. See appendix for code.

Following is the list of all channels with format as shown in the previous model.

Table 8 . 4 Channels with Adstock and Curve transformation

Channel Name	Retention	Curve transformation
Internet_Advertising_Email_Blast_impressions	0.4	gompertz
Newspaper_ROP_impressions	0.2	negative exponential
Magazine_Advertising_impressions	0.4	negative exponential
Newspaper_Preprint_impressions	0.4	logistic
Internet_Adv_Display_impressions	0.1	gompertz
Internet_Adv_Retargeting_impressions	0.4	negative exponential
Internet_Adv_Social_impressions	0.2	logistic
Broadcast_Ads_Digital_Radio_impressions	0.4	logarithmic
Broadcast_Ads_TV_impressions	0.4	logistic
Direct_Mail_Multipage_Customer_impressions	0.4	negative exponential
Broadcast_Ads_Radio_impressions	0.4	logistic
Direct_Mail_Multipage_NCA_impressions	0.1	negative exponential
Direct_Mail_Post_Card_Customer_impressions	0.2	negative exponential

Direct_Mail_PostCard_NCA_impressions	0.4	logistic
Direct_Mail_New_HomeOwner_impressions	0.1	logarithmic
Newspaper_Preprint_Smart_Source_impressions	0.4	logarithmic
Internet_Adv_Search_Non_Brand_impressions	0.1	logistic
Internet_Adv_Search_Shopping_impressions	0.4	richard
Internet_Adv_Search_Brand_impressions	0.1	logistic
Internet_Adv_Search_Local_impressions	0.4	negative exponential
Internet_Adv_Search_overall_impressions	0.4	richard
Direct_Mail_impressions	0.4	logarithmic

Now four types of regressions are formed with the channels.

- **Results**

Best results for the regressions are shown in Table 8.5

Table 8.5 Regression results for model-II

Regression model	value of α	l1-ratio	R^2-score
Linear	NA	NA	0.6702619429624171
Ridge	10^{-5}	NA	0.670252940860252
Lasso	10^{-5}	NA	0.6702404651013721
Elastic Net	10^{-5}	0.5	0.5747921489748316

In The second model, from the above tables, we can see that Linear Regression is giving the best result for the particular data we have.

Here predicted Incremental Sales is 20.08% , which is close to the previous model's result and also is less than 30%. Hence the first and second models are assumed to verify each other.

- **8.2 Multiplicative Models**

8.2.1 Semi-log Model

The additive model does not capture any possible synergies between explanatory variables. If such synergies exist, their parts correlated with individual explanatory variables are allocated to them. However, a proportion of variation may stay unexplained – as part of residuals. Such models suffer from omitted variable bias and autocorrelation of residuals. One way to fix this is to construct interaction variables and estimate synergies through the constructed variables. For the reasons above, semi-logarithmic models became popular in MMM. All it requires is to apply logarithmic transformation on the dependent variable. On the plus side, a response to unit change in the explanatory variable is more flexible. The absolute level of response increases proportionally with the modeled variable.

$$\text{Total\_Weekly\_sales} = \exp(\beta_0).\ \exp(\beta_1 \times \text{Internet\_Advertising\_Email\_Blast\_impressions}).\ \exp(\beta_2 \times \text{Newspaper\_ROP\_impressions})\ldots\ldots\ldots\ \exp(\beta_{22} \times \text{Direct\_Mail\_impressions}).\ \exp(\varepsilon)$$

or, equivalently can be written as

$$\ln(\text{Total\_Weekly\_sales}) = \beta_0 + \beta_1 \times \text{Internet\_Advertising\_Email\_Blast\_impressions} + \beta_2 \times \text{Newspaper\_ROP\_impressions} + \ldots\ldots\ldots + \beta_{22} \times \text{Direct\_Mail\_impressions} + \varepsilon$$

Logarithmic transformation of sales and the modified channel impressions are taken exactly as the very first model. The best results from the four regressions are taken with respect to R^2-score as well as MAPE.

- **Results**

Table 8.6 Regression results for model-III

Regression model	Alpha	l1_ratio	R^2-score	MAPE
Linear	NA	NA	0.7876695740799238	0.1607806365357363
Ridge	10^{-5}	NA	0.7876695741214759	0.16078063655185534
Lasso	10^{-5}	NA	0.7874785658953359	0.16075863505988833
Elastic Net	10^{-5}	0.5	0.7594866347004737	0.17475329756595456

MAPE < 10% is mostly preferable. But 0 - 20% can be considered. Here MAPE ≈ 16%. Also R^2-score is considerable in the model.

Comparing the above tables, Ridge Regression gives the best result here.

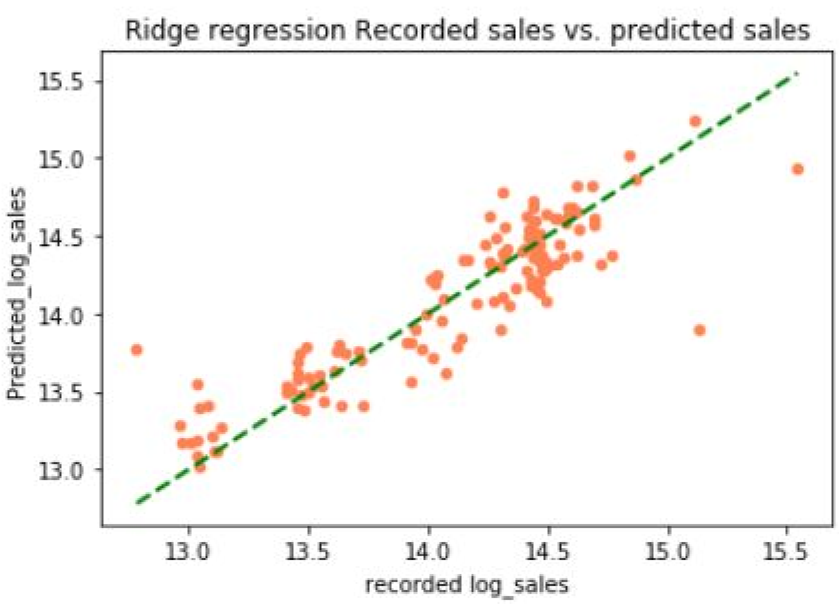


Figure 4 : Recorded vs. predicted log(sales)

As we know coefficients of semi-log model represents percentage change in business outcome (sales) to unit change in the independent variables, *the list of all* *channels and percentage change in Total_Weekly_Sales when the channel impressions are changed in a unit quantity :*

Table 8 . 7 % change for unit change of impressions

Channel Name	% change for unit change of impressions
Internet_Advertising_Email_Blast_impressions	0.24
Newspaper_ROP_impressions	3.15×10^{-7}
Magazine_Advertising_impressions	-1.32×10^{-7}
Newspaper_Preprint_impressions	1.52
Internet_Adv_Display_impressions	0.71
Internet_Adv_Retargeting_impressions	0.03
Internet_Adv_Social_impressions	-0.07
Broadcast_Ads_Digital_Radio_impressions	0.02
Broadcast_Ads_TV_impressions	10.9×10^{5}
Direct_Mail_Multipage_Customer_impressions	3.77×10^{-8}
Broadcast_Ads_Radio_impressions	-407.74
Direct_Mail_Multipage_NCA_impressions	1.51×10^{-8}
Direct_Mail_Post_Card_Customer_impressions	0.02
Direct_Mail_PostCard_NCA_impressions	-8.75
Direct_Mail_New_HomeOwner_impressions	-0.02
Newspaper_Preprint_Smart_Source_impressions	-1.00×10^{-2}
Internet_Adv_Search_Non_Brand_impressions	125.94
Internet_Adv_Search_Shopping_impressions	-0.43
Internet_Adv_Search_Brand_impressions	-123.36

Internet_Adv_Search_Local_impressions	4.08×10^{-5}
Internet_Adv_Search_overall_impressions	0.22
Direct_Mail_impressions	-0.01

Above table is very useful for a retail company. They can increase or decrease expenditure on a particular channel using the table, such that their total spent amount can be optimized ultimately on market to have an increment on sales of their product.

8.2.2 Log-log model

Log-linear or log-log regression is 'theoretically' better than linear regression because of its better physical explanation on the extreme values or boundary conditions of the regressors. In linear regression, if we put distribution equal to zero, we still get a finite sales. This is contrary to common perception which dictates that sales should be zero or infinite if distribution is zero. Log-linear regression takes care of this fact and with any base like variable equal to zero; it gives us a zero or infinite sales (depending upon the sign of associated coefficient). So, if we put $x^{(i)}$ equal to zero, we get sales equal to either 0 or 0^{-1} that is infinity. Whereas, if we put any incremental regressor equal to zero, log-linear regression still gives us a finite result, which is in line with our business perception which says that sales should not reduce to zero if no marketing or incremental activity is done, because there is an existence of base sales. The model is formed as

$$\text{Total\_Weekly\_Sales} = \exp(\beta_0) \,.\, (\mathit{Internet\_Advertising\_Email\_Blast\_impressions})^{\beta_1} \ldots\ldots\ldots\, . \,(\mathit{Newspaper\_ROP\_impressions})^{\beta_{22}} \,.\, \exp(\varepsilon)$$

We can linearize the model through logarithmic transformation.

$$\ln(\text{Total\_Weekly\_Sales}) = \beta_0 + \beta_1 \ln(\mathit{Internet\_Advertising\_Email\_Blast\_impressions}) + \ldots\ldots\ldots + \beta_2 \ln(\mathit{Newspaper\_ROP\_impressions}) + \varepsilon$$

Before predicting sales, the channels have been modified. The fact is that *log transformation itself brings saturation to channel impressions values*, i.e. they don't increase unboundedly. Hence there is no need for normalization for logarithmic model. So,first taking the list of highest correlation of retention added and curve transformed channels, then changing to their log transformations, we have observed that some transformed are giving infinite values. Hence instead of log transformation, I have taken them as *ln(1+x)*. So now finally the model is

$$\mathbf{ln(Total\_Weekly\_Sales)} = \beta_0 + \beta_1 \,\mathbf{ln(1+} \mathit{Internet\_Advertising\_Email\_Blast\_impressions}\mathbf{)} + \ldots\ldots\ldots + \beta_2 \,\mathbf{ln(1+} \mathit{Newspaper\_ROP\_impressions}\mathbf{)} + \varepsilon$$

The result from building a regression model using these is shown in table 8.8 below

- **Results**

Table 8.8 Regression results for model-IV

Regression model	Alpha	l1_ratio	R^2-score	MAPE
Linear	NA	NA	0.955543386754065	0.08164071037399746
Ridge	10^{-3}	NA	0.9554025859063862	0.08142876590295778
Lasso	10^{-3}	NA	0.9541627664612061	0.08113953641787938
Elastic Net	10^{-3}	0.5	0.9548519116289306	0.0812636501869568

For all the models MAPE < 10%, hence the log-log model is a good fit with the data (but also has a probability of overfitting). Here MAPE ≈ 8%. Also R^2-score is impressive for the models. However linear regression is giving the best result. Log-log model gives % change in business outcome (sales) in response to 1%

change in channel impressions. The figure of recorded vs. predicted sales followed by list of channels are shown here.

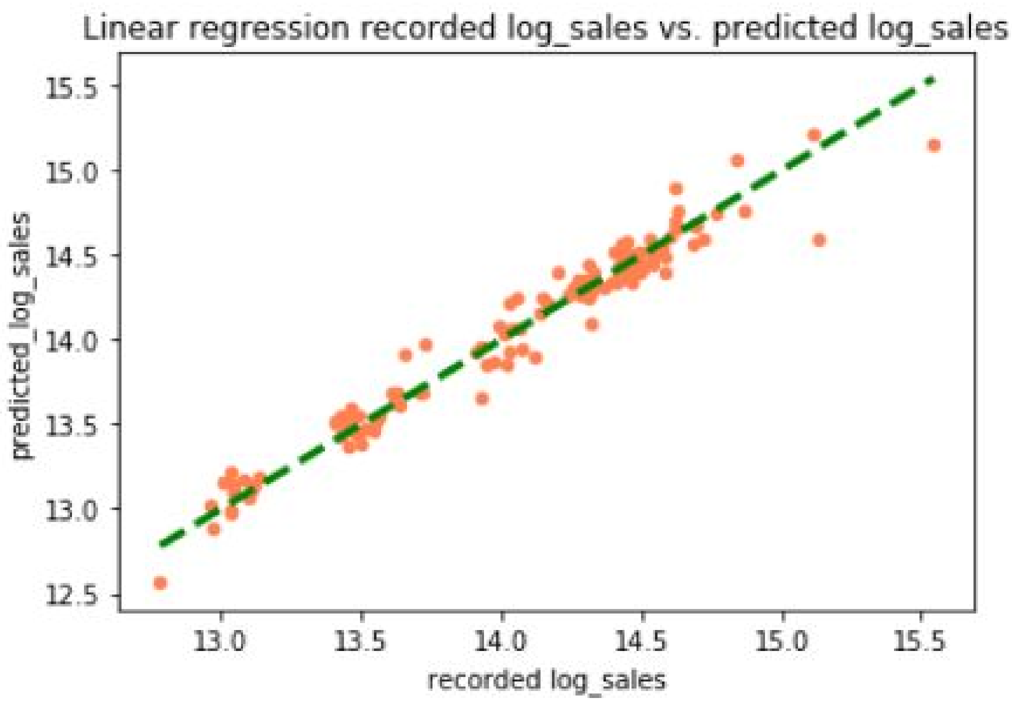


Figure 5 : Recorded vs. predicted log (sales)

Table 8.8: % change for 1% impression change

Channel Name	% change for 1% change of impressions
Internet_Advertising_Email_Blast_impressions	0.33
Newspaper_ROP_impressions	0.03
Magazine_Advertising_impressions	-0.01
Newspaper_Preprint_impressions	0.07
Internet_Adv_Display_impressions	0.01
Internet_Adv_Retargeting_impressions	0.08
Internet_Adv_Social_impressions	-0.02

Broadcast_Ads_Digital_Radio_impressions	2.2×10^{-4}
Broadcast_Ads_TV_impressions	-0.06
Direct_Mail_Multipage_Customer_impressions	0.01
Broadcast_Ads_Radio_impressions	-0.01
Direct_Mail_Multipage_NCA_impressions	-4.14×10^{-3}
Direct_Mail_Post_Card_Customer_impressions	2.11×10^{-3}
Direct_Mail_PostCard_NCA_impressions	0.03
Direct_Mail_New_HomeOwner_impressions	1.74×10^{-3}
Newspaper_Preprint_Smart_Source_impressions	-1.77×10^{-3}
Internet_Adv_Search_Non_Brand_impressions	-0.09
Internet_Adv_Search_Shopping_impressions	-0.12
Internet_Adv_Search_Brand_impressions	0.39
Internet_Adv_Search_Local_impressions	0.04
Internet_Adv_Search_overall_impressions	0.24
Direct_Mail_impressions	-0.05

From the point of view of how 1% change of channel impressions can change sales, table 8.8 is very useful.

Chapter 10

Final Results

Market Mix data can contain various types of channels, such as channel impressions, channel expenditure, some categorical variables etc. But with the data provided by Tiger Analytics in this case, with channel impressions, three important results are very clear.

1. Total sales has two components – Base sales and Incremental sales. Here Incremental sales is 17 – 20% of the total sales. Remaining is Base sales.
2. How the percentage of sales will change for unit changes in channel impressions is shown in *Table 8.7*
3. How the percentage of sales will change for one percent changes in channel impressions is shown in *Table 8. 8*

Conclusion

MMM model is totally dependent on its variables. Properly preprocessed data can make high profit to a company, otherwise can take several years to recover. A model developer should understand the market data and should not blame raw data since they may not be arranged properly. They should also keep in mind a company's profit vs. expenditure to spread awareness for their product in the market. Adstock plays an important role here. When a new product comes on the market, the company should promote it on a large scale. They can use several strategies like discounts, advertisements, posters,attractive offers etc. But as time passes on promotion should be less. Otherwise the company may face a loss, which can be shown by s-curve. It may be possible that after an amount of time the company does not need to even encourage the product. At this time base sales becomes almost 100% and incremental sales drops to 0% of the total sales. I have tried my best to get into the data and form a presentable MMM that can help a company to analyze the market of a product. One can consider the properties and statistical insights of market mix and build a model that can optimize the profit of the company.

References:

- A complete guide to Marketing Mix Modeling.
  URL: https://www.latentview.com/marketing-mix-modeling/

- Wikipedia.
  URL : https://en.wikipedia.org/wiki/Marketing_mix_modeling

- Marketing Mix Model-101.
  URL:
  https://towardsdatascience.com/market-mix-modeling-mmm-101-3d094df976f9

- Multiplicative Marketing Mix Modeling Simplified.

URL:http://learn.fractalanalytics.com/rs/fractalanalytics/images/white%20paper-%20multiplicative%20mmm%20simplified.pdf

Appendix

```
In [25]: n_sfo=[]
for col,chn in zip(df_sfo1.columns,norm_df_sfo1.columns):
    print(col)
    matrix1=transformation_retention_matrix(norm_df_sfo1[chn])
    abc1=np.array(matrix1)
    pos_abc1=abs(abc1)
    index1 = np.where(pos_abc1 == np.amax(pos_abc1))
    #print(index)
    loc1 = list(zip(index1[0], index1[1]))
    #print(loc1[0][1])
    if loc1[0][1]==0:
        print(str(R[loc1[0][0]])+';  gompertz')
        #new_chn=f(int(index1[1]),rec(norm_df_sfo1[chn],R[loc1[0][0]]))
    elif loc1[0][1]==1:
        print(str(R[loc1[0][0]])+';  logistic')
    elif loc1[0][1]==2:
        print(str(R[loc1[0][0]])+';  richards')
    elif loc1[0][1]==3:
        print(str(R[loc1[0][0]])+';  logarithmic')
    elif loc1[0][1]==4:
        print(str(R[loc1[0][0]])+';  negative exponential')
    new_chn=f(int(index1[1]),rec(norm_df_sfo1[chn],R[loc1[0][0]]))
    n_sfo.append(new_chn)
```

Figure 6 : Curve transformation code

```
from sklearn import linear_model
l_reg=linear_model.LinearRegression()
l_reg.fit(d_sfo,sales)
predict_sfo_linear=l_reg.predict(d_sfo)

import sklearn
r2_score_linear=sklearn.metrics.r2_score(predict_sfo_linear,sales)
mape_linear=sklearn.metrics.mean_absolute_error(predict_sfo_linear,sales)
print('r_2 score :'+ str(r2_score_linear))
```

Fig - 8.1