Analysing the Effectiveness of Personalized v/s Non-Personalized Advertising Messages in Influencing Consumers' Purchasing Decisions

- **Naman Gandhi (A006)**

Abstract:

This Study dives into the changes in the behaviour of the consumer when Personalised Advertising Message are sent in context of Marketing. With the Advancement in the Digital Marketing Channels, the marketers are open to a wide variety of channels which include Personalised and Non-Personalised Advertising Messages. There has been a debate all around regarding which form is more effective when compared to the other. We will be using Logistic Regression to find out the impact of personalised message while taking into account some other important variable of our Research which are Demographic Characteristics, Brand Attitudes, Purchase Intentions, Prior Brand Loyalty, etc. A dummy dataset with 100 datapoints of each variable has been used for the analysis. The results will provide us with insights regarding the medium of Advertising Messages that are to be used varying from Customer to Customer and Product to Product.

Introduction:

Advertising plays a very important role in the success of any Marketing Campaign in the Present time. These Campaign can be spread across various Domains covering Various Sectors, Industries and Products. Therefore, it becomes very Crucial for companies to track the Marketing Strategies which they deploy for various Products. These Strategies aim to give them information about the influence a Particular Marketing Campaign has on the Consumers in terms of their Attitude towards the product, their Perception and viewpoint about the product or strategy, etc which in turn help them to make their Purchasing Decisions (Gupta, 2021). In the recent Developments & Advancements which have taken place in the Marketing & Analytical Sector, it has become very easy for the companies to track down their mistakes regarding a particular Campaign and Improve their strategies based on the Information obtained. Furthermore, through these advancements, the doors to personalised advertising have been opened for the marketers (Cho, 2019). The markets can obtain personalised information about the consumers through the data algorithms available online and provide them with personalised advertisements which the consumer is willing to look. This information is very crucial for these marketers as they have access to the individual customers based on their preferences, behaviours, and demographics as well. But having access to this information is just 40% of the job done. The marketers now need to develop an effective Personalised Advertising Strategy which can maximise its output based on the individual Customer information available. There are still a lot of debates in the industry regarding the approach that is taken in order to provide better results and understand the underlying factors that contribute to its success and effectiveness. This Research paper provides an in-depth analysis of the factors that contribute to the success of any Campaign and the Approach that is taken in order to achieve the same. The Primary objective of this research is to assess and compare the Effectiveness that Personalised and Non-Personalised Advertising Messages have on the Purchasing Decision of the Consumer. There are a lot of statistical tests out there in the domain which can help us to arrive at the conclusion. We will be using Logistic Regression for out Research. Logistic Regression is appropriate because our Dependent Variable here is Binary which in this case is <u>Message Type</u>. Along with that the interpretability in terms of <u>Odds Ratio</u> and <u>Confusion Matrix</u> makes it easy and helps identify the impact of individual independent variables on the dependent variables (Elmeguid SMA, 2018). Logistic Regression can include both Categorical and Continuous Independent variables in its Analysis and does not require any Assumption of Linearity (Radojevic T, 2018). The Logistic Regression Model interpretation of coefficients helps to identify which

Advertising Messages (Personalised V/S Non-Personalised) and other independent factors have an impact which is significant on the Consumer Behaviour.

Data Explanation:

The data used for the analysis is a sample data with 100 datapoints each of the Variables that we have considered for our Research.

The Variables are as follows:

1. **Message Type:** It is the Dependent Variable & describes the type of Advertising Message which can be tailored to various Individual Consumers. These are Categorical in nature with two types namely Personalised Messages which are based on the Preferences, behaviour, demographics, etc. of the Consumers & Non-Personalised Messages which are generic and are targeted to a larger Audience.
2. **Gender:** Independent Variable which is Categorical in Nature involved in the Research to Identify which Message Type is better for Each Gender. Two Types – Male or Female.
3. **Age:** Continuous Independent Variable specifying the Age of the Consumer. This will help Marketers determine how individuals will respond to advertising messages. Considered Age of Consumers between 18 to 65.
4. **Income:** Continuous Independent Variable specifying the Purchasing Capacity of the Consumer. Consumers with Average Income of 50,000 are considered for the Research.
5. **Brand Attitude:** It Measures the Consumers' overall understanding of the brand that is advertised after the Advertising Message has been sent. This is Measured on the Scale of 0 to 10 with 10 being the Positive Attitude towards the Brand.
6. **Purchase Intention:** It informs us whether t the Consumer is Going to purchase the product or service in the Future. This is measured on the Scale of Low-Moderate-High with High showing the interest of the consumer to buy the product or service in Future.
7. **Actual Purchase:** It tells us whether the Consumer is Going to buy the advertised Product or Service after the Advertising Message has been Sent. This is Expressed in Terms of YES or NO.
8. **Product Category:** This Variable informs us about the product category that is being advertised. The effectiveness of Advertising varies between various goods and Services. Product Categories Considered are – Electronics, Apparel & Beauty.
9. **Prior Brand Loyalty:** Previous Experiences and Loyalty towards the brand that is advertised also influences the Effectiveness of the Advertising Messages. This is Measured in terms of Low-Moderate-High with High being the Most Loyal.

Overview & Summary of the Data has been Provided in the table Below:

Gender	**Age**	**Income**	**Brand Attitude**	**Purchase Intention**	**Actual Purchase**	**Message Type**	**Product Category**	**Prior Brand Loyalty**
Female:57	Min:18	Min: 21437	Min :1	High: 33	No:54	Non – Personalised:54	Length:100	High: 31
Male:43	Median:40	Median: 48847	Median:6	Moderate:34	Yes:46	Personalised:46	Class: character	Moderate:36
	Mean:38.24	Mean:49794	Mean:5.37	Low:33			Mode: character	Low:33
	Max:64	Max:87017	Max:10					

Procedure:

Logistic Regression is a data Analysis Technique which is used to find the Relationship between two data factors taking into use mathematics. It takes into account this relationship between two data factors to predict the values of one of the factors which is based on the other one. It takes into consideration Categorical variables that have outcomes like "YES" or "No." The Dependent Variable in Logistic Regression which is Denoted by Y, can take only two Possible Values and the same can be Categorical or Continuous which result from the occurring of any event. The independent variables can be Categorical or Continuous but there is no limit to the number of values they can take. It is used on a wide range as it overcomes the Limitations of the Linear Regression Model and Generalised Linear Models which assumes the Linear Relationship between the Dependent and Independent Variables.

It is based on a logistic function given by (Ghosh J, 2018):

$$f(z) = \frac{e^z}{1+e^z} = \frac{1}{1+e^{-z}}$$

Where e – Euler number and z – Value of the Explanatory variable.

The above formula can be written in various different forms depending upon the calculated value (Wang HY, 2018). The Probability form of the same is given as:

$$P(Y = z_1, z_2, \ldots, z_k) = \frac{e^{\beta_0} + \sum_{i=1}^{k} \beta_i z_i}{1 + e^{\beta_0} + \sum_{i=1}^{k} \beta_i z_i}$$

Where Bi, i=0, 1,…,k are logistic regression coefficients and Zi, i=1,2,…,k are independent variables that can comprise of both qualitative and quantitative values.

This Formula can also be looked upon from the Point of View of Odds Ratio which defines the Probability of occurrence of an Event divided by the Probability of the Non-Occurrence of the same Event. Mathematically, it can be Represented as:

$$\frac{P(Y = Z)}{1 - P(Y = Z)} = e^{\beta_0} + \sum_{i=1}^{k} \beta_i z_i$$

The Logistic Regression Equation when there is only One Independent Variable is represented as:

$$P(Y = Z) = \frac{e^{\beta_0 + \beta_1 z_1}}{1 + e^{\beta_0 + \beta_1 z_1}}$$

Taking Log Function on both the sides, we get the final form as:

$$logit\ P(Y = Z) = ln \frac{P(Y = Z)}{1 - P(Y = Z)} = \beta_0 + \beta_1 Z_i$$

But the only requirement of the Logistic Regression that must be satisfied is test sample size condition. This should be equal to n > 10(K+1) where k – No of Predictors.

Interpretation:

Wald's Test:

It is used to check the statistical significance of individual predicted variables (which is equivalent to t-test in the MLR model). The null hypothesis states that the variable is insignificant and alternative hypothesis states that the variable is significant.

The test was conducted to check whether Gender was a significant variable which influences the choice of Message Type. The p-value obtained is 0.0075 < 0.05, we Reject H0 concluding that Gender is a significant variable.

```
Wald test:
----------

Chi-squared test:
X2 = 7.1, df = 1, P(> X2) = 0.0075
```

Similar tests have been run for other variables which resulted in Concluding that Prior Brand Loyalty is another significant variable which influences the choice of Message Type.

Summary of the Model:

Having proved the Significance of our Variables, we now move ahead with the interpretation of the variables that are considered for the Research. Based on the Output obtained, two variables Gender (0.00754) and Prior Brand Loyalty (0.02755) proved to be significant variables meaning that these variables influence the Dependent Variable-Message Type whether the customer wants a Personalised or Non-Personalised Advertising Message. The P-values of both the variables are less than 0.05 leading us to Rejecting H0 with 95% Confidence Level.

```
glm(formula = Message_Type ~ Gender + Age + Income + Brand_Attitude + 
    Purchase_Intention + Actual_Purchase + Product_Category + 
    Prior_Brand_Loyalty, family = "binomial", data = train_data)

Deviance Residuals: 
    Min       1Q   Median       3Q      Max  
-2.0201  -1.0245   0.4645   0.8961   1.8683  

Coefficients:
                            Estimate Std. Error z value Pr(>|z|)   
(Intercept)               -2.058e+00  1.596e+00  -1.290  0.19719   
GenderMale                 1.670e+00  6.249e-01   2.672  0.00754 **
Age                       -2.446e-02  2.383e-02  -1.027  0.30461   
Income                    -7.066e-06  2.007e-05  -0.352  0.72479   
Brand_Attitude             1.684e-01  1.043e-01   1.615  0.10624   
Purchase_IntentionLow      5.870e-01  6.769e-01   0.867  0.38590   
Purchase_IntentionModerate 1.051e+00  7.124e-01   1.475  0.14032   
Actual_PurchaseYes         5.621e-01  6.103e-01   0.921  0.35704   
Product_CategoryBeauty     3.939e-01  6.588e-01   0.598  0.54986   
Product_CategoryElectronics 8.528e-01  7.764e-01   1.099  0.27198   
Prior_Brand_LoyaltyLow     6.440e-01  7.089e-01   0.908  0.36361   
Prior_Brand_LoyaltyModerate 1.705e+00  7.736e-01   2.204  0.02755 * 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 95.607  on 69  degrees of freedom
Residual deviance: 78.638  on 58  degrees of freedom
AIC: 102.64
```

Hosmer and Lemeshow Test (H-L Test):

Hosmer and Lemeshow Test (H-L Test) is used to assess the Goodness of Fit of Logistic Regression Model. By Goodness of Fit it means how well the Logistic Regression model predicts the observed Probabilities. Based on the P-value obtained (0.2393), we can say that we have insufficient evidence to Reject of H0 meaning that the Observed Probabilities match the Predicted Probabilities across all levels of predicted risks with 95% Confidence Level.

```
        Hosmer and Lemeshow goodness of fit (GOF) test

data:  model1$y, fitted(model1)
X-squared = 10.381, df = 8, p-value = 0.2393
```

Classification Table:

	Actual Responses	
Predicted Responses	No	Yes
No	15	7
Yes	3	5

Accuracy of the Model:

$$\frac{TP + TN}{TP + TN + FP + FN} = 0.667$$

The model has 67% accuracy rate. We can observe that 20 responses are correctly classified by the model, where 5 are True Positives and 15 are True negative.

Were as the 10 responses are false predictions, where 3 were predicted to have a positive response but they did not respond positively. It had falsely predicted that there is a negative response but responded positively.

The notions of sensitivity and specificity are related to the cut-off point. (MS, 2003). Sensitivity is the ability of a model to detect units with a distinguished feature, determines the number of correctly predicted cases in a set of all observed occurrences (Cheng LS, 2018) (Mason C, 2018).

$$\boldsymbol{Sensitivity} = \frac{TP}{TP + FN} = 0.625$$

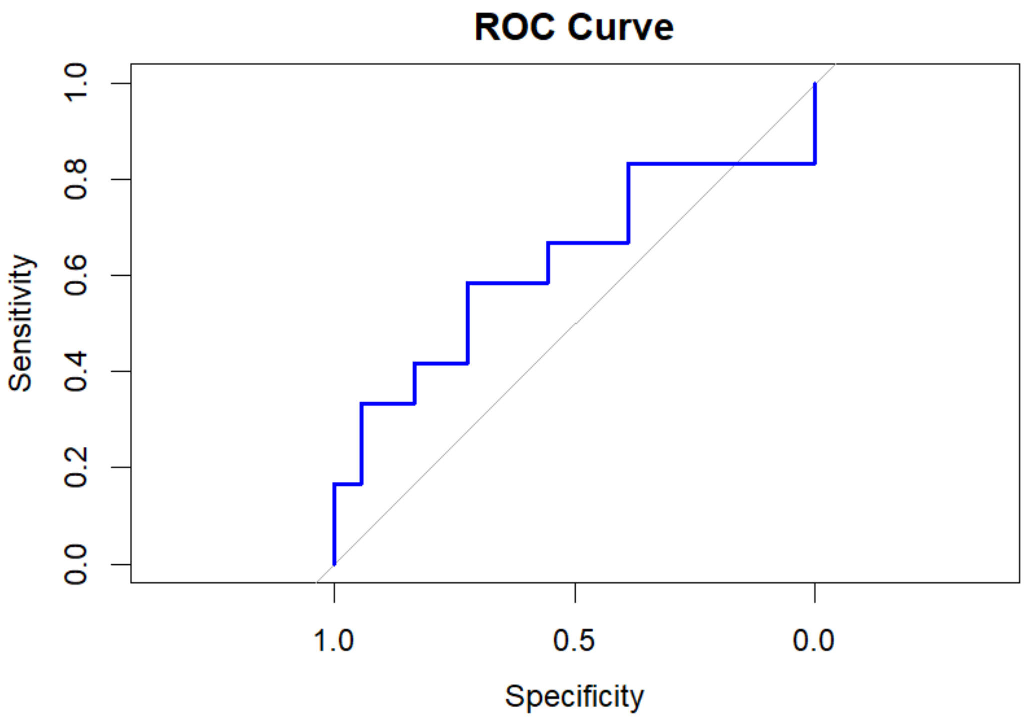
Sensitivity tells percentage of correctly classifying from the positive classes. That means 62.5% of the time the response is correctly predicted from the positive class.

$$\boldsymbol{Specificity} = \frac{TN}{TN + FP} = 0.6818$$

Specificity tells percentage of correctly classifying from the Negative classes. That means 68.18% of the time the response is correctly predicted their negative respond towards the Message Type from the negative class.

ROC Curve:

The relationship between model's specificity and sensitivity is presented by the ROC curve (Chen WJ, 2018) (Kalil AC, 2010). The Placement of the ROC Curve on the graph will signify the Accuracy of the Model. In Order to have a Sensitivity of 1 which implies the Higher Accuracy of the Model, our ROC Curve is supposed to be placed in the Upper Left Corner of the Graph where False Positive Rate = 0 & AUC=1. If True Positive Rate = False Positive Rate, then the co-ordinates of x & y axis will be in the Ratio of 1:1 implying that the graph is drawn on the 45 ° diagonal (y = x) of the ROC curve (AUC = 0.5). In order to be considered Acceptable, the AUC value should be greater than 0.5.



We can observe that AUC > 0.5, concluding that the model can be accepted. In Marketing industry AUC value of 0.625 suggests that the LR model have a discriminative power in distinguishing the customers who positively respond to the Message Type and those who don't.

```
Call:
roc.default(response = test_data$Message_Type, predictor = predictions)

Data: predictions in 18 controls (test_data$Message_Type Non-Personalized)
< 12 cases (test_data$Message_Type Personalized).
Area under the curve: 0.625
```

Conclusion:

In Conclusion, from this research we are able to provide valueable insghts regardiung the effectiveness of different Advertising Message Type that influence the Consumer Decision making behaviour. By applying the Logistic Regression Model, we have showcased the significant role that Advertising Message Type plays alongside Demographic Characters, Brand Attitudes, Purchase Intensions, etc. in shaping the responses of our Consumers.

We will be using our insights to contribute to the ongoing disclosure to optimize Advertising Startegies in this Digital Era by tailoring Personalised Messages to Target Audience. This will help the

Marketers navigate themselves better through the digital space effectively. They will be able to have a better understanding about the consumer behavior which remains a crucial driving tool for the success of the Marketing Campaign.

<u>**References:**</u>

Chen WJ, Y. X. (2018). Spatial prediction of landslide susceptibility using data mining-based kernel logistic regression, naive Bayes and RBF Network models for the Long County area (China). *Bulletin of Engineering Geology and the Environment.*

Cheng LS, T. W. (2018). Combination of High Sensitivity C-Reactive Protein and Homocysteine Predicts the Post-Stroke Depression in Patients with Ischemic Stroke. *Molneurobiol*.

Cho, Y. &. (2019). The effectiveness of personalized advertising: Investigating the role of consumer traits and privacy concerns. *Journal of Advertising*.

Elmeguid SMA, R. M. (2018). Customer satisfaction in sharing economy the case of ridesharing service in Alexandria, Egypt. *The Business and Management Review*.

Ghosh J, L. Y. (2018). On the Use of Cauchy Prior Distributions for Bayesian Logistic Regression. *Bayesian Anal*.

Gupta, S. &. (2021). The impact of personalized advertising on consumer response: A meta-analysis. *International Journal of Research in Marketing*.

Kalil AC, M. J. (2010). Recommendations for the assessment and reporting of multivariable logistic regression in transplantation literature, American journal of transplantation. *official journal of the American Society of Transplantation and the American Society of Transplant Surgeons.*

Mason C, T. J. (2018). Predicting Engineering Student Attrition Risk Using a Probabilistic Neural Network and Comparing Results with a Backpropagation Neural Network and Logistic Regression. *Logistic Regression*.

MS, P. (2003). The Statistical Evaluation of Medical Tests for Classification and Prediction. *Oxford University Press*.

Radojevic T, S. N. (2018). The effects of traveling for business on customer satisfaction with hotel services. *Tourism Management*.

Wang HY, Z. R. (2018). Optimal Subsampling for Large Sample Logistic Regression. *Journal of the American Statistical Association.*