ADVANCED TARGETING ANALYTICS: A LOGISTIC REGRESSION ANALYSIS OF MULTI-CHANNELED CONSUMER PURCHASE BEHAVIOUR

Cross-sectional analysis of user-level interactions quantifying the influence of impression frequency, targeting typologies, engagement metrics, and funnel positioning on purchase conversions. Implements multivariate logistic regression, interaction effects, collinearity diagnostics (VIF), model selection via AIC/BIC, and ROC-AUC validation to optimize campaign strategy and drive data-informed marketing decisions.

AUGUST 8, 2025

PRATIK GANGULI
University Of Auckland

Introduction

This report presents a statistical analysis evaluating the impact of digital marketing targeting strategies on consumer purchase behaviour. Using logistic regression, the study examines how factors such as impression counts, targeting types, and engagement influence purchase likelihood. Multiple models were assessed using metrics like Tjur R^2, AIC, BIC, accuracy, and AUC. The findings provide data-driven insights into which targeting strategies most effectively drive conversions, supporting future campaign planning.

The dataset contains *2,000 observations* and *20 variables*, each reflecting a user's interaction with a digital marketing campaign. Though time-related fields like ***StartTime*** and ***EndTime*** are present, most users *(1,990 out of 2,000)* appear only once, with minimal duplication. This lack of repeated measurements rules out panel data, while the absence of continuous tracking over uniform intervals excludes time series structure. Instead, each row captures a single user's interaction during a specific campaign window *(Dec 1, 2014 – Jan 31, 2015)*, making it a one-time snapshot across different users. Thus, the dataset is best classified as **cross-sectional**, consistent with its non-repetitive, user-level focus within a short time frame.

To examine the distribution of targeting strategies used by Company XYZ, we analysed interaction data from *1,990* unique users, grouped by ***UserID.*** For each user, we aggregated variables using the mean for average time metrics and targeting strategy proportions, and the sum for total time metrics, impression counts and the final funnel stage reached. Each of the eight targeting types represented the fraction of a user's total exposure (values between 0 and 1). Since users could be exposed to multiple strategies, averaging across users provided a balanced view of campaign exposure, with each user's row summing to 1. The results indicate that ***Contextual*** targeting is the most frequently used approach, accounting for *35.7%* of total exposure and reaching *801* unique users. It is followed by *Behavioral (26.6%), Retargeting (22.9%),* and *Prospecting (8.2%)*. In contrast, strategies like *Predictive (0.3%), Lookalike (0.2%),* and *Remarketing (0.5%)* were rarely used, both in terms of exposure share and user count. This distribution suggests that Company XYZ favours broad relevance-based targeting methods over more personalized or data-driven techniques. The relative usage across strategies is shown in *Table 1.1* and visualized in *Figure 1.1*.

To examine the relationship between the number of ad impressions ***(Count)*** and the likelihood of a consumer making a purchase ***(Purchaser)***, a logistic regression model was developed. The model indicates a statistically significant positive association between impressions and purchase probability ***($p < 0.001$)***. The estimated odds ratio of *1.07 (See Table 1.2, **Model_0**)* suggests that each additional impression is associated with a *7%* increase in the odds of a purchase. While the predictor is statistically significant, model performance metrics suggest limited explanatory power. The **Tjur R^2** of *0.033*, representing the difference in average predicted probabilities between purchasers and non-purchasers is extremely low (values above *0.2* are generally considered acceptable), indicating poor model fit. An **AUC** of *0.559 (Table 1.3, **Model_0**)* further confirms weak discriminatory ability. These results imply that while impressions are a relevant predictor of purchase behaviour, they alone do not capture the complexity of consumer decisions, highlighting the need for additional variables to enhance model accuracy and campaign effectiveness.

To assess whether stage-based funnel positioning improves the model's ability to predict purchase behaviour, the variable ***Funnel*** was converted into two binary indicators

Funnel_Upper and ***Funnel_Lower***, and added to the initial logistic regression model *(See Table 1.2, **Model_0**)*. Due to perfect multicollinearity, only one dummy variable *(Funnel_Upper)* was retained in the model, with *Funnel_Lower* omitted automatically as a reference category. The coefficient for *Funnel_Upper* was not statistically significant *(p = 0.630; Table 1.2, **Model_1**)*, suggesting that users in the upper funnel stage are no more or less likely to convert than those in the lower stage when holding impressions constant. Model performance showed minimal improvement. The log-likelihood increased only slightly from *–969.08* (previous model) to *–968.97*, indicating a marginal improvement in model fit. However, this change is practically negligible. The AIC rose slightly from *1942.2* to *1943.9*, and the BIC increased from *1953.4* to *1960.7 (Table 1.3, **Model_0**, **Model_1**)*, both suggesting that the added variable does not sufficiently improve model quality to justify the added complexity. Similarly, Tjur R^2 remained unchanged at *0.033*, and AUC increased trivially from *0.559* to *0.561*, indicating very limited gain in predictive power. Given the statistical insignificance of *Funnel_Upper* and the absence of meaningful improvement in model fit or predictive ability, the inclusion of funnel stage information does not appear beneficial for this regression task. It may be more valuable to explore interaction effects or other behavioural predictors to enhance the model's performance.

To examine how different targeting strategies relate to purchase behaviour, a new variable, ***CampaignCount***, was computed to capture how many distinct campaign types each user was exposed to. Multiple regression models were tested, and **Model 5** was chosen based on stronger AIC, log-likelihood, Tjur R^2, and AUC. To address multicollinearity, since including all targeting dummies alongside *CampaignCount* would introduce perfect linear dependencies, *Behavioral* targeting was excluded and treated as the **baseline category**. **Variance Inflation Factors** *(Table 1.4)* confirmed that multicollinearity in Model 5 remained within acceptable limits (all VIF < 4). The final model *(See Table 1.2, **Model_5**)* included other targeting types, engagement metrics *(e.g., AvgTime_h),* and key interaction terms to capture conditional effects. Coefficients reflect the relative impact of each targeting type on purchase probability compared to *Behavioral* targeting. **Model 5** performed well *(Tjur R^2 = 0.310, AUC = 0.860; Table 1.2, Table 1.3),* indicating strong explanatory and classification performance. ***Contextual***, ***Predictive***, and ***Retargeting*** strategies significantly increased purchase likelihood, while *Prospecting* had a negative effect relative to *Behavioral* targeting. ***CampaignCount*** was also a strong positive predictor, suggesting that exposure to a wider range of strategies boosted conversion odds. Interaction effects revealed that targeting effectiveness varied by user engagement. ***Retargeting*** was more effective for ***upper-funnel*** users *(OR = 2.58, p = 0.011)*, and broader campaign exposure ***(CampaignCount* × *FunnelUpper)*** also boosted early-stage conversions *(OR = 2.04, p < 0.001)*. ***Contextual*** targeting became slightly less effective with increased ***browsing*** time (OR = 0.99, p < 0.001), while ***Prospecting* × *AvgTime_h*** showed a weak positive trend (OR = 1.01, p = 0.082). These findings highlight the value of aligning **targeting strategies** with user **journey stages** and **behavioural signals** to optimise outcomes.

Conclusion :

In conclusion, while *Contextual*, *Behavioral*, and *Retargeting* made up most campaign exposures *(35.7%, 26.6%, 22.9%)*, their impact on purchases differed. Despite lower usage than *Behavioral*, *Retargeting* proved more impactful, indicating an opportunity to reallocate spend toward high-performing strategies. *Retargeting* and *Contextual* strategies significantly outperformed *Behavioral* targeting, while *Prospecting* had a slight negative effect. The positive role of *CampaignCount* highlights the benefit of diverse targeting. Interaction effects showed that campaign effectiveness varies by user engagement *(e.g., time-on-site, number of exposures)* and funnel stage, underscoring the importance of tailoring strategies to the customer journey for improved performance.

References :

- ChatGPT (OpenAI): OpenAI. (2024). ChatGPT (Aug 3 version) [Large language model]. https://chat.openai.com/

- Grammarly: Grammarly Inc. (n.d.). Grammarly [AI writing assistant]. https://www.grammarly.com/

- Abhishek, V., Jerath, K., & Sharma, S. (2022, February 12). The impact of retail media on online marketplaces: Insights from a field experiment (SSRN Scholarly Paper No. 3013468). SSRN. https://ssrn.com/abstract=3013468

Appendix :

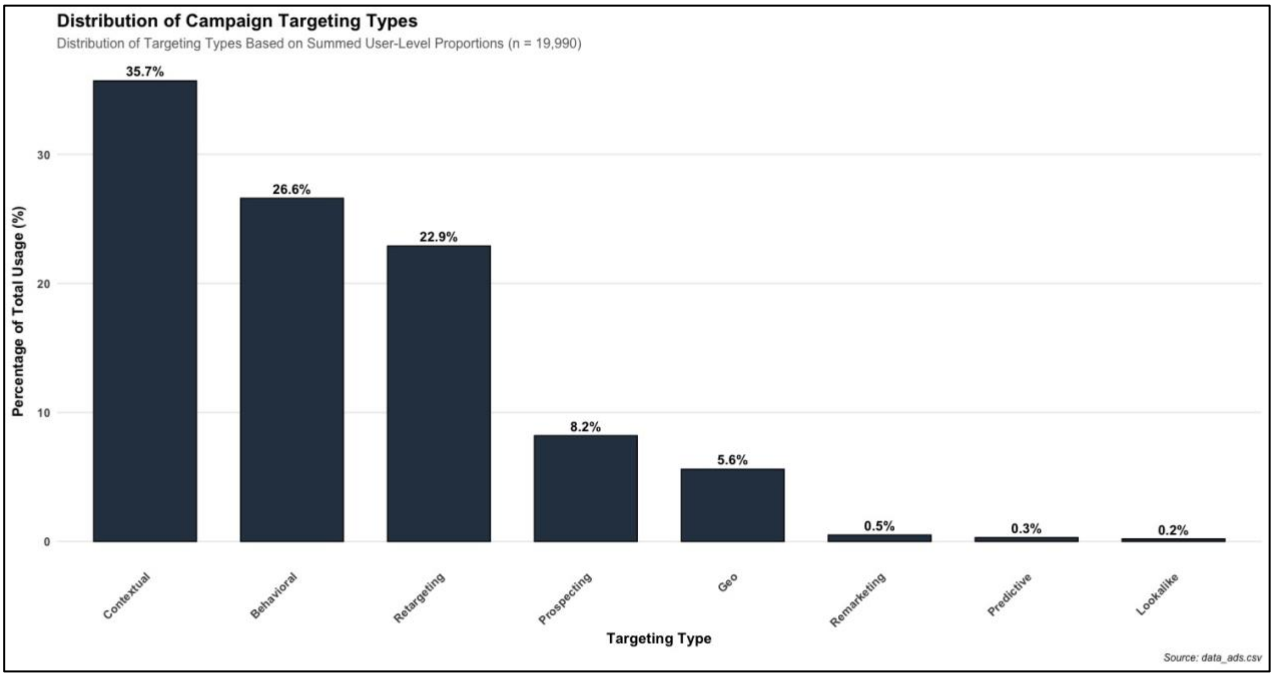

Fig 1.1 : Description of Campaign Targeting Strategies Based on Summed User-Level Proportions Used by Company XYZ

TargetingType	TotalUsage	Percentage	UserCount
Contextual	710.721736	35.7	801
Behavioral	529.220614	26.6	685
Retargeting	455.655012	22.9	545
Prospecting	162.737555	8.2	246
Geo	112.208919	5.6	137
Remarketing	10.361286	0.5	20
Predictive	5.894877	0.3	13
Lookalike	3.200000	0.2	8

Table 1.1 : Distribution of Campaign Exposure Across Targeting Strategies

Note: **TotalUsage represents the cumulative sum of each targeting type's fractional exposure across all 1,990 unique users. **Percentage** denotes the share of each strategy relative to the total exposure. **UserCount** indicates how many unique users were exposed to that particular strategy. Since users could encounter multiple strategies, exposures are fractional (values between 0 and 1) and user counts are not mutually exclusive.*

Logistic Regression Models: Targeting Types and Purchase Probability

	Model_0				**Model_1**				**Model_2**				**Model_3**				**Model_4**				**Model_5**			
Predictors	*Odds Ratios*	*std. Error*	*Statistic*	*p*	*Odds Ratios*	*std. Error*	*Statistic*	*p*	*Odds Ratios*	*std. Error*	*Statistic*	*p*	*Odds Ratios*	*std. Error*	*Statistic*	*p*	*Odds Ratios*	*std. Error*	*Statistic*	*p*	*Odds Ratios*	*std. Error*	*Statistic*	*p*
(Intercept)	0.18	0.01	-23.22	**<0.001**	0.18	0.02	-18.10	**<0.001**	0.02	0.01	-13.98	**<0.001**	0.03	0.01	-13.18	**<0.001**	0.03	0.01	-13.06	**<0.001**	0.04	0.01	-12.78	**<0.001**
Count	1.07	0.01	7.12	**<0.001**	1.07	0.01	7.09	**<0.001**	0.98	0.01	-1.58	0.113	0.98	0.01	-1.85	0.065	0.98	0.01	-1.79	0.073				
Funnel Upper					1.06	0.12	0.48	0.630																
Contextual									1.77	0.38	2.65	**0.008**	2.08	0.44	3.43	**0.001**	2.09	0.44	3.46	**0.001**	3.43	0.81	5.26	**<0.001**
Geo									1.19	0.50	0.41	0.679	1.22	0.50	0.49	0.624	1.25	0.51	0.55	0.582	1.12	0.45	0.28	0.778
Lookalike									14.52	30.80	1.26	0.207	22.20	47.05	1.46	0.144	24.18	51.26	1.50	0.133	15.83	31.66	1.38	0.167
Predictive									14.99	15.99	2.54	**0.011**	15.11	16.34	2.51	**0.012**					13.77	14.67	2.46	**0.014**
Prospecting									0.48	0.23	-1.56	0.120	0.44	0.22	-1.67	0.095	0.45	0.27	-1.32	0.188	0.16	0.13	-2.33	**0.020**
Remarketing									0.07	0.16	-1.22	0.222	0.02	0.05	-1.50	0.135	0.02	0.05	-1.48	0.138	0.02	0.06	-1.40	0.161
Retargeting									10.36	2.23	10.89	**<0.001**	9.73	2.06	10.73	**<0.001**	9.75	2.07	10.74	**<0.001**	7.45	1.65	9.06	**<0.001**
AvgTime h									0.99	0.00	-11.99	**<0.001**	0.99	0.00	-11.85	**<0.001**	0.99	0.00	-11.83	**<0.001**	0.99	0.00	-7.29	**<0.001**
Funnel [Upper]									3.53	0.59	7.55	**<0.001**												
CampaignCount									4.84	0.78	9.83	**<0.001**	3.85	0.64	8.12	**<0.001**	3.76	0.63	7.91	**<0.001**	3.17	0.50	7.34	**<0.001**
CampaignCount × FunnelUpper													2.19	0.34	5.01	**<0.001**	2.21	0.35	5.05	**<0.001**	2.04	0.30	4.87	**<0.001**
Count × FunnelUpper													1.05	0.03	1.62	0.106	1.05	0.03	1.60	0.110				
Count × Prospecting																	1.00	0.08	-0.05	0.956				
Count × Predictive																	4.81	2.71	2.79	**0.005**				
Retargeting × FunnelUpper																					2.58	0.97	2.53	**0.011**
Contextual × AvgTime h																					0.99	0.00	-3.91	**<0.001**
Prospecting × AvgTime h																					1.01	0.00	1.74	0.082
Observations	1990				1990				1990				1990				1990				1990			
R^2 Tjur	0.033				0.033				0.296				0.297				0.300				0.310			
AIC	1942.167				1943.935				1455.773				1451.618				1449.338				1423.342			
log-Likelihood	-969.083				-968.967				-715.887				-712.809				-710.669				-697.671			

Table 1.2 : Logistic Regression Models Predicting Purchase Probability Based on Targeting Strategy and User Behavior

Note: This table summarizes the results of five logistic regression models examining the relationship between various predictors, such as ad impressions (Count), targeting strategies, campaign diversity (CampaignCount), and interaction terms, with the binary outcome of purchase (Purchaser). Odds ratios (OR), standard errors (SE), and p-values are reported. Model performance metrics (Tjur R^2, AIC, BIC, log-likelihood) are included to assess model fit and predictive power. Significant predictors at* p *< 0.05 are highlighted, with* ***Model 5* *demonstrating the strongest explanatory performance (Tjur R^2 = 0.310; AUC = 0.860).*

Logistic Regression Model Comparison				
Model	AIC	BIC	Accuracy	AUC
Model_0	1,942.200	1,953.400	0.801	0.559
Model_1	1,943.900	1,960.700	0.801	0.561
Model_2	1,455.800	1,522.900	0.836	0.852
Model_3	1,451.600	1,524.400	0.834	0.853
Model_4	1,449.300	1,527.700	0.837	0.854
Model_5	1,423.300	1,501.700	0.846	0.860

Table 1.3: Comparison of Logistic Regression Model Performance Metrics

Variance Inflation Factor Table (VIF) for `Model_5`		
Variable	**VIF Score**	**Multicollinearity Risk**
Prospecting:AvgTime_h	3.720	Low
Prospecting	3.359	Low
Contextual	2.333	Low
AvgTime_h	2.322	Low
Retargeting	2.264	Low
CampaignCount:Funnel	2.092	Low
Contextual:AvgTime_h	1.803	Low
Retargeting:Funnel	1.690	Low
CampaignCount	1.380	Low
Geo	1.248	Low
Remarketing	1.053	Low
Lookalike	1.048	Low
Predictive	1.029	Low

Table 1.4: Variance Inflation Factor (VIF) Scores for Model_5 Variables