Measuring the Effectiveness of WhatsApp Ads for High-end Streetwear

WhatsApp Ads Experiment for Six Squad

Fahad Alsubaie, Madhumitha Sriram, Olimpia Borgohain ,Raiymbek Ordabayev, Rohit Devanaboina

Introduction

We conducted an experiment on the thriving sneaker and streetwear company 'Six Squad', based in Almaty, Kazakhstan. Founded in December 2022, the company has rapidly expanded, operating two brick-and-mortar stores alongside an online presence. With approximately $1 million in revenue till date, 'Six Squad' provided us access to their sales data, customer loyalty program information, and their WhatsApp bot for our research. Our study aimed to evaluate the effectiveness of marketing messages delivered through the WhatsApp bot and determine if allocating resources to this channel is worthwhile compared to other marketing strategies. Additionally, we sought to discern any disparities in consumer response between standard and 'nudge' messages. Through this experiment, we aimed to provide valuable insights for 'Six Squad' and similar businesses looking to optimize their marketing efforts.

Experiment Design

- **Objective**

The objective of our experiment was to evaluate the efficiency and return on investment (ROI) of using WhatsApp as an advertising platform for promoting two new collections of Air Force 1 Nick sneakers to the members of the Six Squad sneaker store's loyalty program. The introduction of these products presented an opportunity to gauge how digital advertisements could influence the purchasing decisions of 1,549 members enrolled in the loyalty program. To accomplish this, a structured experimental framework was established, leveraging randomization to assign members into distinct groups, including a control group, allowing us to discover the differential impacts derived from various advertising contents. The methodology of our experiment will be explained in the next section.

- **Unit of Randomization**

We randomized at the loyalty member level, using Member-ID. This personalized approach not only ensured the relevance and precision of our advertising efforts but also allowed for a highly targeted analysis, enabling us to measure the direct impact of our interventions on each member's engagement and purchasing behavior.

- **Blocking Randomization Process**

The block randomization method was intentionally implemented, leveraging the historical purchase data of members, such as their past transactions and net sales, to ensure a balanced allocation of members across both treatments and control groups. This detailed approach significantly minimized potential biases and confirmed that pretreatment characteristics were uniform across all groups.

Treatment Arms Design
The experimental design for our study was structured to include three distinct groups to facilitate a comprehensive analysis of the effects of WhatsApp advertising on the promotion of the two new collections of Nike Air Force 1 07s, new in stock at Six Squad's stores. Here's a detailed overview of each group within the experimental framework (**ads attached in appendix**):

- **Control Group**: Participants in this group did not receive any promotional message, serving as a baseline to evaluate the effect of the WhatsApp advertisements.
- **Treatment Arm 1** (Status Quo Ad): This group received Six Squad's traditional advertising strategy, a straightforward WhatsApp message that described the promoted product with attached image.
- **Treatment Arm 2** (Nudge-Based Ad): This group was exposed to a nudge-based advertisement strategy, which incorporated several psychological initiations: scarcity, by emphasizing limited stock; personalization, by addressing the customer by name in the message header; a visual hook in the form of a video showcasing the products. This approach was designed to leverage various motivational drivers, offering a more compelling call to action than the standard advertisement.

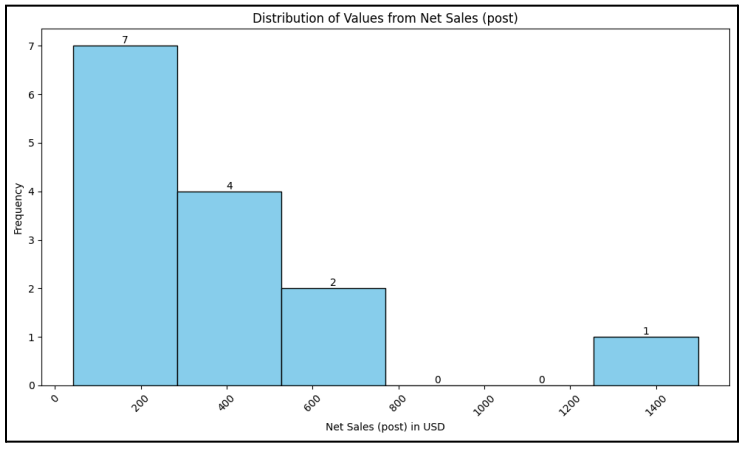
Data Collection:
Data collection was central to our experiment, focusing on metrics that could directly reflect the impact of our experiment. We tracked ad click-through rates using unique tracking links (vk.cc) assigned for the treatment arms, capturing any engagement. Sales data were obtained from Six Squad's transaction records, enabling the analysis of both the conversion rate for the advertised product and the broader impact on average revenue per member. Purchase propensity, defined as the percentage of members making at least one purchase post-ad exposure, was also assessed to determine the overall effectiveness of our advertising efforts.

Through this investigation, we aimed to uncover actionable insights to guide future marketing efforts, optimizing the ROI of WhatsApp advertising campaigns for loyalty program members.

Results
Now, let us see the results of our experiment. The primary focus of our results are basic regression tests (with covariates), and heterogeneous treatment effects. For the fixed effects models, please take a look at our code. But first, let us explore our dataset.

Exploratory Data Analysis

We can clearly see that there is an outlier in terms of net sales (cart value). This individual, who belongs to the Control group, made a ~$1500 purchase, almost 4x which the $400 average among all purchases. This will definitely skew the results towards the control group. However, we will not exclude them from our analysis since we have no valid reason to do so.



Proportions Test:

	Control (vs Rest)	T1 - Standard Ad (vs Rest)	T2 - Nudge Ad (vs Rest)	Randomization was successful
Proportions Z-Test (P-values)	1	1	1	

The result of the proportions z-tests, with a p-value of 1.0, indicate that there is no statistical evidence to reject the null hypothesis, which means that the randomization was done properly. They have been divided into 3 equal arms as intended (516 obs vs 516 obs vs 517 obs). This is due to the fact that we used GupShup, a service that automates Whatsapp Ads launch, ensuring all members are treated as intended.

Balance Checks:

P-values	Control		Treatment 1		Treatment 2	
	Treatment 1	Treatment 2	Control	Treatment 2	Control	Treatment 1
Past Transactions	0.513	0.773	0.513	0.720	0.773	0.720
Past Sales	0.737	0.680	0.737	0.972	0.680	0.972

The t-test results for all comparisons between the arms by "Total Transactions (past)" and "Net Sales (past)" show no statistically significant differences in means. This means our blocking was successful, and that all three arms have similar distributions w.r.t. past purchase behavior.

Regression Analysis

Click Rate:

Click Rate - Standard Ad vs Nudge Ad

	Dependent variable: click
	Click Rate
Intercept	0.04***
	(0.02 , 0.06)
Treatment Effect	0.01
	(-0.01 , 0.04)
Observations	1033
Note:	*p<0.1; **p<0.05; ***p<0.01

The Standard Ad had a ~4% click rate, while the Nudge Ad had a ~5.5% click rate. However, this difference in means was not significant at the 0.05 threshold. Despite this, there is evidence that the nudge ad more likely than not has a positive effect on the CTR. A slightly larger experiment would likely generate significant results.

Conversions:

Treatment Effect - Conversions

	Dependent variable: Q("Total Conversions (post)")	
	Control vs Standard Ad	Standard vs Nudge
	(1)	(2)
Intercept	0.00***	0.00***
	(0.00)	(0.00)
Treatment Effect	0.00***	0.00***
	(0.00)	(0.00)
Observations	1032	1033
Note:		*p<0.1; **p<0.05; ***p<0.01

Surprisingly, there were **zero conversions** across all arms (conversion meaning sale of advertised product). Without any variance, we cannot use Conversion as an outcome to test treatment effects. So we drop it from the rest of the analysis.

Rest of the Outcomes:

We measured the estimated treatment effects of the rest of the outcomes in conjunction, using Past Transactions and Past Net Sales as covariates, to account for past purchasing behavior.

Treatment Effects - Control vs Status Quo Ad

	Total Transactions	Quantity	Net Sales	Purchase Propensity
	(1)	(2)	(3)	(4)
Treatment Effect	0.00	-0.01	-494.68	-0.00
	(0.01)	(0.02)	(1484.17)	(0.00)
Past Transactions	0.01	-0.00	-370.68	0.01
	(0.01)	(0.02)	(870.86)	(0.01)
Past Net Sales	0.00	0.00	0.01	0.00
	(0.00)	(0.00)	(0.01)	(0.00)
Observations	1032	1032	1032	1032
Note:				*p<0.1; **p<0.05; ***p<0.01

For Control vs Standard Ad, we did not measure any significant effects, even with the covariates. We detected a negative effect for Net Sales (approx. -1 USD), due to the outlier transaction previously mentioned in the EDA section. If we exclude that member from the regression, the effect turns positive, but is still not significant. Please note that if we take a look at the 95% CF for the effects, we see that the true effects lie close to 0 for most outcomes, indicating the Ads are not very effective at driving purchases. The power for these effects ranged from 5% to 11%.

Treatment Effects - Status Quo Ad vs Nudge Ad

	Total Transactions	Quantity	Net Sales	Purchase Propensity
	(1)	(2)	(3)	(4)
Treatment Effect	0.01	0.00	795.38	0.01
	(0.01)	(0.02)	(1076.92)	(0.01)
Past Transactions	0.01	-0.01	-518.79	0.01
	(0.01)	(0.02)	(786.08)	(0.01)
Past Net Sales	-0.00	0.00	0.01	-0.00
	(0.00)	(0.00)	(0.01)	(0.00)
Observations	1033	1033	1033	1033
Note:				*p<0.1; **p<0.05; ***p<0.01

For Standard Ad vs Nudge Ad, we did not measure any significant effects, even with the covariates. We detected a negative effect for Net Sales (approx. -1 USD), due to the outlier transaction previously mentioned in the EDA section. If we exclude that member from the regression, the effect turns positive, but is still not significant. Please note that if we take a look at the 95% CF for the effects, we see that the true effects lie close to 0 for most outcomes, indicating the Ads are not very effective at driving purchases. The power for these effects ranged from 5% to 25%.

Heterogeneous Treatment Effects

We also measured how the treatment effects varied by the age of members (Generation) and tenure within the loyalty program (Days Since Group)

Regression Results : Interaction Effect of Generation and Treatment

	Sales - C vs T1	Sales - T1 vs T2	Quantity - C vs T1	Quantity - T1 vs T2	Propensity - C vs T1	Propensity - T1 vs T2
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	2213.22	-357.89	0.03	0.01	0.01	-0.01
	(2804.56)	(2851.81)	(0.04)	(0.05)	(0.01)	(0.02)
Treatment Effect (GenX)	7508.52	1382.36	0.14	0.03	0.02	0.01
	(5307.55)	(1178.15)	(0.11)	(0.02)	(0.02)	(0.01)
Treatment Effect (GenZ)	-806.43	632.14	-0.01	-0.00	-0.00	0.01
	(1800.95)	(1116.84)	(0.02)	(0.02)	(0.01)	(0.01)
Treatment Effect (Millenials)	-1150.77	980.91	-0.02	0.01	-0.00	0.01
	(1680.27)	(1364.56)	(0.02)	(0.02)	(0.00)	(0.01)
Observations	981	974	981	974	981	974
Note:						*p<0.1; **p<0.05; ***p<0.01

A notable difference emerges when comparing sales between Treatment Arm 1 and Treatment Arm 2. Specifically, we observe contrasting effects across generational cohorts. Millennials and GenZ demonstrate a negative response of ~1150 KZT and -800 KZ, indicating lower receptivity to WhatsApp Business. GenX exhibits a high positive effect of ~7500 KZT. This higher positive

effect is consistent across other metrics, such as quantity bought and purchase propensity, underscoring the importance of tailoring marketing strategies to different demographic segments.

Regression Results : Interaction Effect of Days since Registration and Treatment Group

	Sales - C vs T1	Sales - T1 vs T2	Quantity - C vs T1	Quantity - T1 vs T2	Propensity - C vs T1	Propensity - T1 vs T2
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	2144.38	-445.96	0.03	0.01	0.01	-0.01
	(2809.02)	(2803.51)	(0.04)	(0.05)	(0.01)	(0.02)
Treatment Effect (<100 days)	633.43	1399.95	0.00	0.01	0.00	0.01*
	(1411.13)	(1238.94)	(0.02)	(0.02)	(0.01)	(0.01)
Treatment Effect (100-200d)	-1038.82	782.79	-0.01	0.00	-0.00	0.01
	(1663.22)	(1320.63)	(0.02)	(0.02)	(0.00)	(0.01)
Treatment Effect (200-300 days)	-806.19	516.28	-0.01	0.00	-0.00	0.01
	(1820.27)	(1046.06)	(0.03)	(0.01)	(0.01)	(0.01)
Treatment Effect (Day One Members)	-1262.58	171.52	-0.02	-0.00	-0.01	0.00
	(1654.47)	(1078.28)	(0.02)	(0.02)	(0.01)	(0.01)
Observations	981	974	981	974	981	974
Note:						*p<0.1; **p<0.05; ***p<0.01

In examining the sales data across different registration periods, we find that the impact varies. Notably, those registered for <100 days experienced the only positive increase of ~630 KZT for sales (Control vs Ad). Conversely, the remaining groups have negative effects on sales. This suggests that newer members may be more inclined to make purchases when advertised to.

ROI Analysis

Without significant results, we can only provide ROI ranges based on best and worst case scenarios. Here are the ranges:

ROI Analysis	Estimate	Lower Bound	Upper Bound
Without Outlier	14.68%	-7.21%	36.63%
With Outlier	-12.56%	-70.25%	45.13%

Here, we have calculated the ROI estimates for two scenarios: Without the outlier that skewed results towards the Control group, and with the outlier. We calculate the ROI using the estimate and both bounds to give a range, based on these values;

Cost per ad - ~$0.06 | **Margin (on sales)** = 57% (average margin across products)

Limitations of the Experiment

1. **Power of Experiment**:

With the extremely small effects detected by our experiment in particular, and Ads experiments in general, we need ~20x more members (approx. 30,000) in order to increase the power of our experiment to the 80% threshold. There is not much we can do to address this issue w.r.t. Six Squad, but once

2. **High-end products**:

The average price of Six Squad's products is 160 USD (closer to 450 USD by purchasing power parity). These are high-end products with limited demand. Trying to effect a change in this demand is a big task, one which advertising alone may not be able to achieve.

3. **Duration of Measurement**:

As seen in the outlier skewing the results to the Control group, the 7-day data collection period is not enough to smooth over short term fluctuations in the outcomes. We believe that a 30-day period would be sufficient enough to account for major fluctuations and provide more accurate estimates.

Recommendations

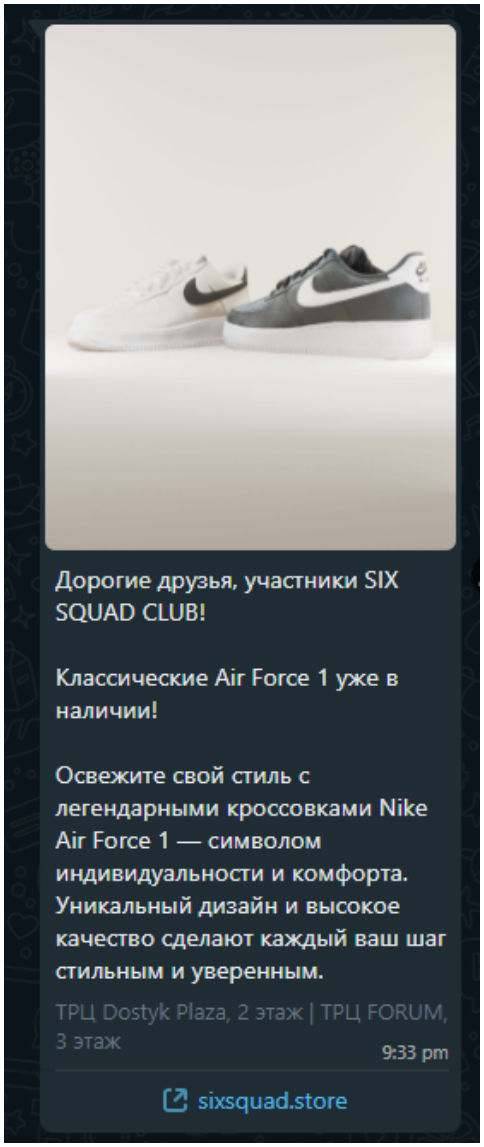
1. **Focus marketing spend to attract new customers, not loyalty members**
   The allocated marketing budget would be more beneficial if it were spent on other marketing channels, such as influencer marketing, Instagram, and others. This is because WhatsApp ads do not effectively reach existing customers, as they rarely visit the store.

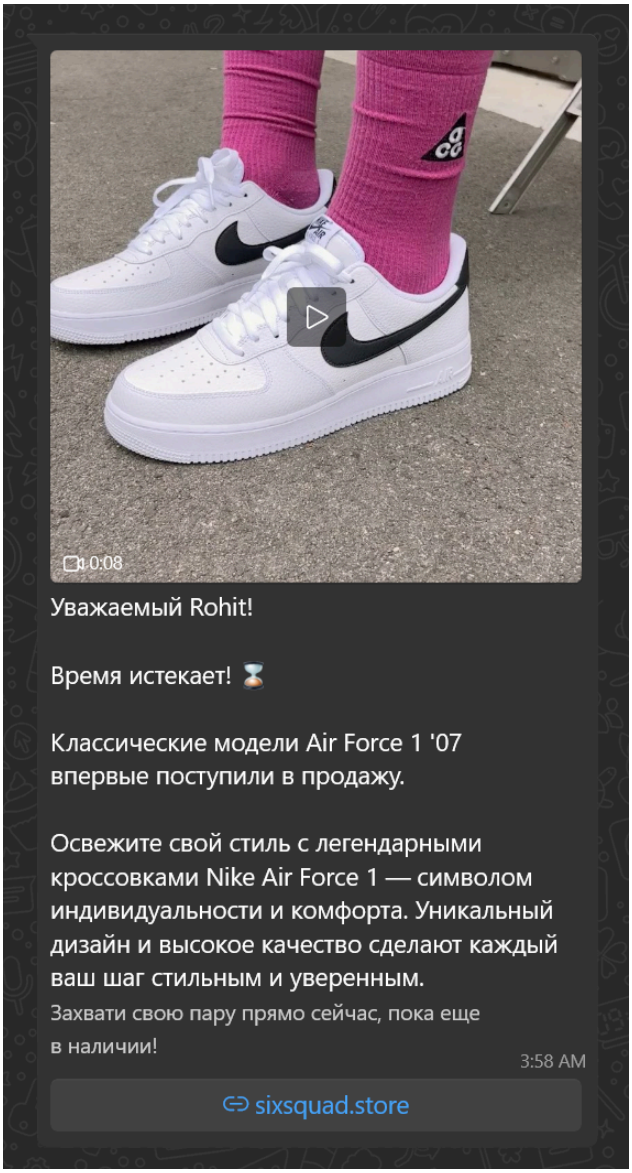
2. **Try personalized ads by customer segment**
   Ads work best with a personalized approach because different types of customers exist based on their purchase history and demographics. By personalizing Ads, we can increase their effectiveness and effect on key metrics.

3. **Keep conducting more experiments in business is very important**
   Experimentation is crucial in every business and is key to success. Many successful businesses conduct experiments to continually improve and outperform their competitors. With Six Squad just entering their second year of business, we strongly recommend they continue to use experiments to guide their decision making.

Appendix

Standard Ad (left) & Nudge Ad (right)





Fixed Effects - Control vs Treatment 1

Fixed Effects - Control vs Treatment 1

	est1	est2	est3	est4
depvar	transactions	Quantity	Summ	propensity
treatment_group	-0.000 (0.003)	-0.006 (0.019)	-556.843 (1417.712)	-0.000 (0.003)
day_of	x	x	x	x
propensity_past	x	x	x	x
R2	0.799	0.305	0.113	0.713
S.E. type	hetero	hetero	hetero	hetero
Observations	1032	1032	1032	1032

Fixed Effects - Treatment 1 vs Treatment 2

Fixed Effects - Treatment 1 vs Treatment 2

	est1	est2	est3	est4
depvar	transactions	Quantity	Summ	propensity
treatment_group	-0.005 (0.003)	-0.017 (0.009)	-533.770 (560.994)	-0.002 (0.002)
day_of	x	x	x	x
propensity_past	x	x	x	x
R2	0.895	0.725	0.674	0.899
S.E. type	hetero	hetero	hetero	hetero
Observations	1033	1033	1033	1033