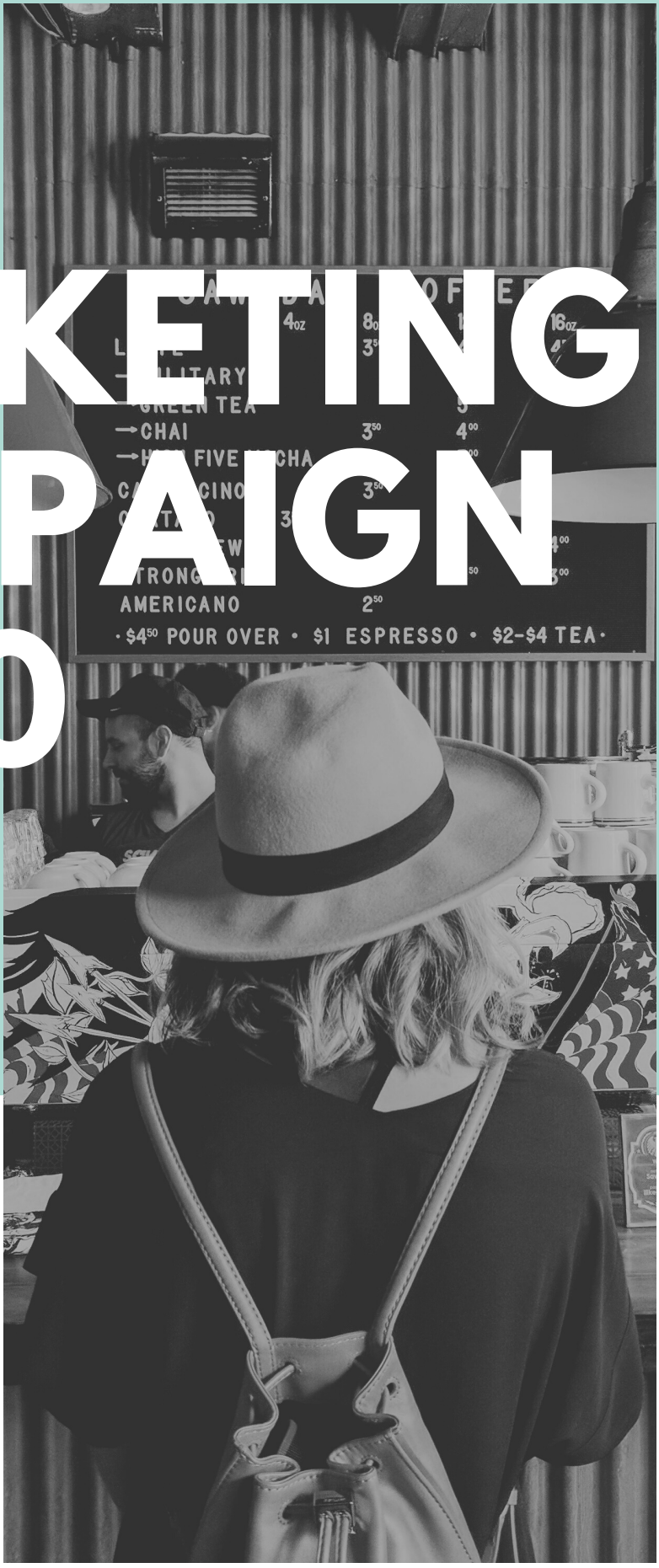
MARKETING CAMPAIGN 2020



Omar's store

INTRODUCTION TO THE DATA ANALYSIS

As told by the Chief Marketing Officer, Omar's store has been struggling with the recent marketing campaigns.

To understand why the campaigns were not functioning, Team 7 developed multiple analysis involving different factors that seemed significant to understand the problem.

The analysis was divided in two parts, Part 1 involves Statistical Tests and Regression and Part 2 involves Data Visualization using multiple graphs.

Team 7

DATA ANALYSTS/REGRESSION&DATA VISUALIZATION

PART 1: STATISTICAL TESTS AND REGRESSION

A)

What factors are significantly related to the number of store purchases?

Regression Statistics	
Multiple R	0.78311231
R Square	0.61326489
Adjusted R Square	0.61100632
Standard Error	2.02760171
Observations	2240

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	13	14511.9224	1116.30172	271.529049	0
Residual	2226	9151.4615	4.11116869		
Total	2239	23663.3839			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	2.95608125	0.22919125	12.8978803	9.2677E-37	2.50663028	3.40553221	2.50663028	3.40553221
Kidhome	-0.64471171	0.1047716	-6.15349703	8.9619E-10	-0.85017198	-0.43925144	-0.85017198	-0.43925144
In meat	0.71781038	0.05251888	13.6676634	6.7556E-41	0.61481927	0.8208015	0.61481927	0.8208015
MntWines	0.00328655	0.00021246	15.4693014	2.3844E-51	0.00286992	0.00370319	0.00286992	0.00370319
MntFruits	0.00591236	0.00144187	4.10047487	4.2711E-05	0.0030848	0.00873991	0.0030848	0.00873991
MntSweetProducts	0.00406523	0.00138377	2.93780158	0.00333945	0.00135162	0.00677883	0.00135162	0.00677883
NumDealsPurchases	0.18593651	0.02656425	6.99950188	3.3868E-12	0.13384321	0.2380298	0.13384321	0.2380298
NumWebPurchases	0.09926058	0.02130005	4.66010972	3.3471E-06	0.05749053	0.14103063	0.05749053	0.14103063
NumCatalogPurchases	-0.16855133	0.02426754	-6.94554739	4.93E-12	-0.2161407	-0.12096195	-0.2161407	-0.12096195
NumWebVisitsMonth	-0.23208851	0.02513063	-9.23528295	5.8221E-20	-0.28137045	-0.18280658	-0.28137045	-0.18280658
AcceptedCmp3	-0.45626973	0.17341827	-2.63103617	0.00857122	-0.7963482	-0.11619126	-0.7963482	-0.11619126
AcceptedCmp5	-0.95028903	0.20342624	-4.67141821	3.1697E-06	-1.34921404	-0.55136402	-1.34921404	-0.55136402
AcceptedCmp2	1.03146344	0.39148148	2.63476943	0.00847782	0.26375641	1.79917047	0.26375641	1.79917047
Response	-0.91180182	0.13499091	-6.75454214	1.8228E-11	-1.17652309	-0.64708055	-1.17652309	-0.64708055

FIGURE 1.

Shows Regression Statistics, Anova Table and Table of relevant variables with respective coefficients and P-values

With a confidence level of 95%, several multiple regression models were done before getting to our final result. The P-value of several variables were too big to take them into consideration since our alpha is only 0.05, so one by one, each variable was being eliminated to run the next model and examine the relevance between variables and the number of store purchases. With an F of 271 and a significance of 0 we can say that the factors are significantly related to the number of store purchases. The adjusted R square is 0.61 which means that 61% of the time we can explain the variation of the dependent variables.

PART 1: STATISTICAL TESTS AND REGRESSION

The final variables that show a P-value less than 0.05 are number of kids home, natural logarithm of meat*, wines, fruits, sweet products, number of deal purchases, number of catalog purchases, number of web visits in a month, campaign 3, campaign 5, campaign 2 and response.

*Meat products had to be transformed due to the fact that the initial scatterplot showed a logarithm pattern and after transforming it, there seem to be more resemblance to a linear regression.

Some coefficients of these variables are negative which means that the relationship to store purchase is negative, affecting the number of store purchases instead of making them increase.
For example, the number of kids' home affect -0.65 for the store purchase, which we can interpret this like the more kids are at home the less a parent wants to spend.

RECOMMENDATIONS TO THE CMO

The analysis shows how deal purchases affects positively to store purchases; more deals should be created to attract more customers.

- Create a repeat purchase frequency tool for the customers, there are many ways to increase purchase customer retention, but the most common ones are giving out gift cards, vouchers and sending personalized emails to appreciate their loyalty.
- Loyalty programs and rewarding points are also effective marketing tools to apply to retain customers, you can offer lower costs to clients who assure you to maintain a frequency of purchase of a base volume in a year.

PART 1: STATISTICAL TESTS AND REGRESSION

- Those deals should be focused primarily on meat products. At the moment if the purchase of meat increases by 100% then the store purchases increase by 0.72, creating deals that boost meat purchases will increase store purchases.And then a combination of wine products, fruits and sweet products, this way the effect of this variables can increase store purchase on the long run.

- Also, creating deals for sweet products maybe the coefficient of number of kids' home can turn positive and customers might be more interested in buying sweet products for their kids.

- Another thing that we noticed is that campaign 2 works positively for store purchases increasing the amount by 1. A recommendation is if a customer already accepted campaign 2 then the rest of the campaigns are useless cause it is only making decrease the store purchases.

PART 1: STATISTICAL TESTS AND REGRESSION

B)

Does US fare significantly better than Rest of the World in terms of total purchases?

Regression Statistics	
Multiple R	0.8233582
R Square	0.67791872
Adjusted R Sq	0.67762764
Standard Erro	4.09062284
Observations	2216

ANOVA					
	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	77942.1605	38971.0802	2328.96824	0
Residual	2213	37030.5611	16.7331952		
Total	2215	114972.722			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	6.54677785	0.12475312	52.4778703	0	6.30213243	6.79142326	6.30213243	6.79142326
total purchase	0.00983135	0.00014417	68.192818	0	0.00954863	0.01011408	0.00954863	0.01011408
USA	0.87314139	0.4053796	2.15388586	0.0313568	0.07817718	1.6681056	0.07817718	1.6681056

FIGURE 2.

Shows Regression Statistics, Anova Table and Table of relevant variables with respective coefficients and P-values

Created a new column of total purchases, which is the sum of all purchases made in store, web purchases, catalog purchases, purchases made through deals, as well as, created another column with the sum of the amount spent on fish, meat, fruit products, sweet products and wine products. Then, created a dummy variable for the country with 1 and 0: if the country is US, we'll have the value 1 otherwise we'll have 0. So, the explanatory variable to compare US with RoW in terms of total purchases is a DUMMY VARIABLE for the country and the response will be the TOTAL PURCHASE.

H0: US is not significantly better than RoW

H1: US is significantly better than RoW

PART 1: STATISTICAL TESTS AND REGRESSION

The analysis of the results show that the regression model provides evidence that concludes that the model is statistically significant, since the P-value is 0 which means the significance of F < 0.05 and the null hypothesis can be rejected. Moreover, by analyzing the regression output, the adjusted r-square is 0.67 which means the explanation towards the response can be acceptable.
With this we can conclude that the US is significantly better than the Rest of the World.

C)

Test if people who buy gold in the last 2 years would have spent more in store purchases.

:GoldProds	NumStorePurchases	<GOLD	<STORE	<<	<>	><	>>
12	3	1	1	1	0	0	
2	1	1	1	1	0	0	
4	1	1	1	1	0	0	
8	1	1	1	1	0	0	
1	0	1	1	1	0	0	
1	0	1	1	1	0	0	
1	0	1	1	1	0	0	
1	0	1	1	1	0	0	
3	0	1	1	1	0	0	
207	13	0	0	0	0	0	
148	13	0	0	0	0	0	
58	10	0	0	0	0	0	
168	13	0	0	0	0	0	

FIGURE 3.

Table showing the dummy variables.

Horizontally is gold condition				
Gold	<AVG	>AVG		
<AVG	1065	189	1254	
>AVG	454	508	962	
Store Purch	1519	697	2216	Total
EXPECTATION	<AVG	>AVG		
<AVG	859.5785199	394.4215	1254	
>AVG	659.4214801	302.5785	962	
Store Purch	1519	697	2216	
X^2 test	3.55993E-80			

FIGURE 4.

Tables used for the Chi-square test of independence.

PART 1: STATISTICAL TESTS AND REGRESSION

The proposed null hypothesis is that the amount spent on gold is independent of the number of store purchases. Firstly, there are 4 dummy variables for 4 possible outcomes, which the total count of them are our observed numbers. Then using the ratios (e.g. sum of row/total) determine the values that will go inside the expectation table. Using excel built in Chi-squared test function the result of the p-value is < 0.05 hence we can reject the null hypothesis and conclude that the people that buy gold also spend on store purchases.

D)

Fish has Omega 3 fatty acids, good for brain, accordingly, do "Married PhD candidates" have a significant relation with amount spent on fish?

Regression Statistics	
Multiple R	0.10995065
R Square	0.01208915
Adjusted R	0.01076368
Standard Er	54.3341795
Observatior	2240

ANOVA					
	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	80778.51379	26926.17	9.120704	5.34E-06
Residual	2236	6601126.036	2952.203		
Total	2239	6681904.55			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	42.2227357	1.651807685	25.56153	1.2E-126	38.98349869	45.4619727	38.9834987	45.46197266
Marital_Du	-4.4519023	2.668637844	-1.66823	0.09541	-9.685169186	0.7813645	-9.6851692	0.781364503
PhD	-15.89143	3.322563942	-4.78288	1.84E-06	-22.4070621	-9.3757969	-22.407062	-9.37579691
Interaction	5.40727463	4.479681944	1.207067	0.227534	-3.377495866	14.1920451	-3.3774959	14.19204513

FIGURE 5.

Shows Regression Statistics, Anova Table and Table of relevant variables with respective coefficients and P-values

PART 1: STATISTICAL TESTS AND REGRESSION

With knowledge of interaction variables, the interaction effect was done by multiplying the two relevant explanatory variables which in this case is the dummy variables for marital status and education. Since we are trying to find if married PhD candidates have a significant relationship with the amount spent on fish, we simply gave 'married' a value of 1 and 'PhD' a value of 1, while the rest were given a value of zero. To test our hypothesis we ran our regression with the created interaction and dummy variables, which helps to conclude there is no significant relationship between the interaction variable, marital status, or education status (Full Regression in Excel). Therefore, we can not statistically prove that married PhD candidates are more likely to buy more fish than other candidates.

What other factors are significantly related to amount spent on fish?
Since there was no significant relationship between Married PhD candidates with the amount spent on fish, we wanted to know if there are other factors that could be related. Although we know there is no significance among Married PhD candidates, we wanted to see if candidates of other marital and educational status' were more likely to buy fish.
However, both proved to not be significant as the educational factor had an r-square of only 0.02 and the non married had an r-square of 0.009 which are not significant values.

PART 1: STATISTICAL TESTS AND REGRESSION

We then considered age as a factor, because due to the health benefits associated with fish and the degradation of health as people age, we figured older candidates would be inclined to buy more fish. However, after running our regression only 0.1% of the variation in amount spent on fish can be explained by age, therefore it is not significant.

We also considered income, because people with higher income tend to live a healthier lifestyle and therefore may buy more fish. However, after running our regression we can conclude due to a significance f <0.05, with an adjusted r-square of 19%, income might be a significant factor related to the amount spent on fish.

Regression Statistics	
Multiple R	0.437564
R Square	0.191462
Adjusted R	0.191101
Standard E	49.13266
Observatio	2240

ANOVA					
	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	1279333	1279333	529.9599408	2.00E-105
Residual	2238	5402572	2414.018		
Total	2239	6681905			

	Coefficients	*andard Erro*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-12.346	2.40225	-5.13936	2.99565E-07	-17.05689446	-7.63515	-17.056894	-7.635154
Income	0.000955	4.15E-05	23.02086	2.0032E-105	0.000873372	0.001036	0.0008734	0.001036

FIGURE 6.

Shows Regression Statistics, Anova Table and Table of relevant variables with respective coefficients and P-values

PART 1: STATISTICAL TESTS AND REGRESSION

E)

Other analysis relevant to show.

In part A) some recommendations were made to create deals based on sweet products to boost store purchases made for parents with kids or teens at home. To backup these recommendations a t- test was made to analyze if there is a difference between the amount of kids and teenagers at home with the number of purchases made with deals. The following hypothesis were tested:
Ho: There is no difference with the amount of kids and teenagers with the number of purchases made with deals.
H1: There is a difference with the amount of kids and teenagers at home with the number of purchases made with deals.

	Home	*NumDealsPurchases*
Mean	0.950446429	2.325
Variance	0.565207483	3.73354176
Observations	2240	2240
Pooled Variance	2.149374621	
Hypothesized Mea	0	
df	4478	
t Stat	-31.37723162	
P(T<=t) one-tail	7.9313E-196	
t Critical one-tail	1.645193977	
P(T<=t) two-tail	1.5863E-195	
t Critical two-tail	1.960493887	

FIGURE 7.

T-Test with respective P-value

PART 1: STATISTICAL TESTS AND REGRESSION

With a P-value <0.05, we are 95% confident that we can reject the null hypothesis, concluding that there is a difference with the amount of kids and teens at home and the number of purchases made with deals.
After realizing this, a pivot table was created to show the quantity of kids that created more impact on the purchases. Having one or two kids and teens create a bigger use of spending with deals.
Therefore, we can also conclude statistically that the recommendations made previously will create an impact on purchases if applied correctly.

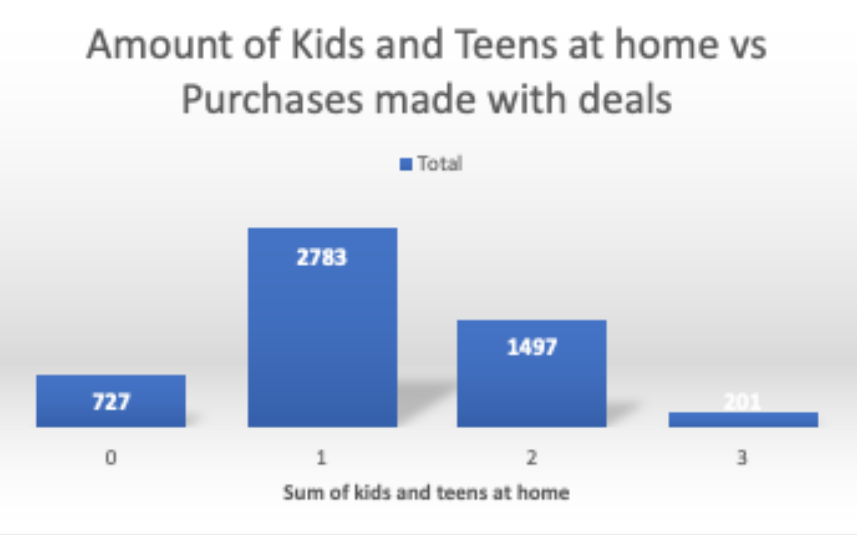


FIGURE 8.

Graph showing the amount of kids and teens at home and purchases made with deals

CONCLUSION

To help Omar's store overcome the struggle with its recent marketing campaigns, we performed a multitude of analyses to figure out what factors contributed to the issue. Overall, we found that the US outperforms other countries in terms of total purchases and people that have bought gold in the past 2 years are more likely to spend more on store purchases. On the other hand, we found that regardless of age, income, marital status, and education status, people will still buy fish which gives precedence that one factor will not buy more than other.

We also found that people respond well to deal purchases, particularly households that have 1 or 2 kids and teens. As stated before, for Omar's store to retain customers we suggest the CMO to implement a purchase frequency tool and a loyalty program that helps lower costs and send personalized emails to customers. We also suggest deals to primarily focus on meats in combination with wine, fruits, and sweet products, which will increase store purchases in the long run.

Furthermore, since households with kids and teens are spending the most, the creation of deals for sweet products will also have a positive effect on store purchases.