1. Measuring success of campaigns is the key for our job. Would you consider following campaign a success? Explain your finding.

KPI	Sent Email (Target)	Not Sent Email (Control)
Volume	593	98
Bet on any Sport during Promo Period	410	54
Bet on World Cup during Promo Period	378	49
Deposited during promo period	368	51
Stakes of those betting on any sport	£1,730,100	£303,164
Gross Win of those betting on any sport	£127,796	£42,539
Free bets awarded to those betting on any sport	£18,770	£2,910
Stakes of those betting on World Cup Matches	£274,261	£25,159
Gross Win of those betting on World Cup matches	£56,132	£8,106
Free bets awarded to those betting on World Cup matches	£5,309	£621

Email Opens	132
Email Clicks	12
Email Unsubscribes	2

This seemed to be an Email campaign where there were several variables to analyse in order to gauge the customer behaviour in both Any Sport and World Cup competition. Hence, there was a need to create 2 pots (Target v Control) to the whole population to compare their performance, so the former pot (86% size) was sent an email, whereas the latter pot (14%) was not.

Firstly, the Bet variable can derive the following results:

Any Sport			
Group	**Volume**	**Bet**	**Betting Rate (%)**
Target	593	410	69.1
Control	98	54	55.1

World Cup			
Group	**Volume**	**Bet**	**Betting Rate (%)**
Target	593	378	63.7
Control	98	49	50

So then, it can be clearly seen that the Betting Rate for both pots are different, leaning for the Target groups. But the key question here is if these variable differences (Target v Control) are statistically significant or not.

To solve that, a Chi-Squared Test can be applied to determine whether the campaign was successful or not in terms of betting results. The purpose of this test is to determine if a statistical difference between observed data and expected data is due to chance, or if it is due to a relationship between the pots. Subsequently, two hypotheses need to be tested:

- Null: There is no significant difference in Betting Rate obtained by the two pots.
- Alternative: there is a significant difference in Betting Rate obtained by the two pots.

The tables with the results are as follows:

Any Sport	**Bet**	**No Bet**
Target	410	183
Control	54	44

World Cup	**Bet**	**No Bet**
Target	378	215
Control	49	49

Then, the test is formulated as follows:

$$\chi^2 = \sum_{\forall i,j} \frac{(O_{i,j} - E_{i,j})^2}{E_{i,j}}$$

Where

$$O_{i,j}\text{: The actual value of cell row } i\text{, column } j$$
$$E_{i,j}\text{: The expected value of cell row } i\text{, column } j$$

And the expected values need to be computed with the following formula:

$$E_{i,j} = \frac{\text{total values in row } i \times \text{total values in column } j}{\text{total values in whole table}}$$

Furthermore, after the test statistics are computed, it's necessary to compare it with the table value and this is:

$$\chi^2_k(1-\alpha)$$

Where k and alpha are the degree of freedom and the predefined significance level, respectively. The alpha can be set to 0.05 (5%) throughout the analysis. If the test statistics are greater than the above table, the null hypothesis can be confidently rejected, so then the conclusion would be that there would be significant difference between both pots in Betting.

Calculations can be seen in the Appendix [1].

It's necessary to find the critical value in the lower-tail critical values table, that is the random variable value when the chi-square distribution with 1 degree of freedom (Formula: (R-1)x(C-1) where R stands for Rows and C stands for Columns, so then (2-1)x(2-1) = 1) hits the probability of (1-0.05) = 0.95. The critical value found in the table is 0.004 and the p-values (obtained from applying both test statistics values (7.51 for Any Sport and 6.73 for World Cup) to other tables) are <0.001 for Any Sport and <0.005 for World Cup. Thus, the null hypothesis for both Any Sport and World Cup disciplines at 95% probability is rejected, there is enough evidence to state that there is a significant difference between betting rates calculated in the two test pots.

Alternatively, this test can also be carried out in Power BI (a Business Intelligence and Data Visualization tool) by entering the data, building up the model with the data relationships and then creating few measures to give the test result.

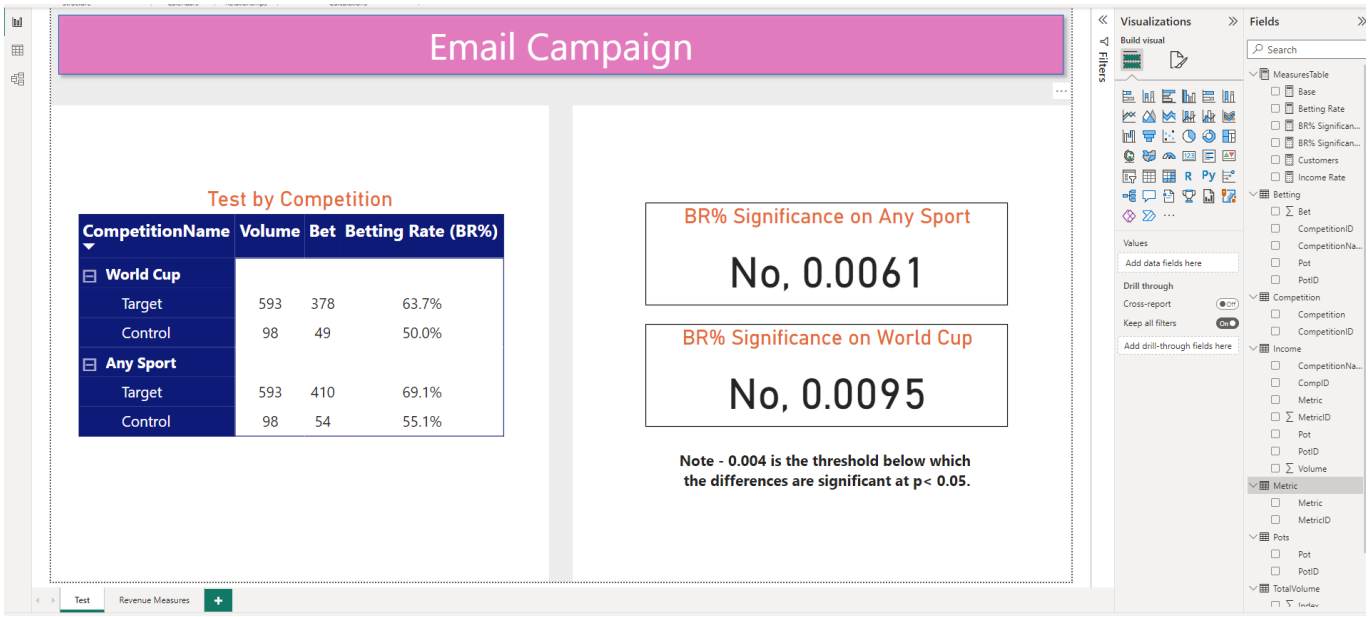

Figure 1:Test

The Figure 1 above shows the test results which demonstrate that the Betting Rate (BT) for both disciplines are below 0.004 at 95% probability so then, the differences between the Target and Control groups are statistically significant.

Both the model relationship and the BR% Significance Value measure can be seen in the Appendix [2].

Regarding the other variables given in the exercise (Stakes, Gross Win and Free Bets), as these were provided with their final amount (and not given case by case for every bet or player), it's not then possible to work out the sample Variance, so then, not possible to calculate a Test to determine whether or not the differences between their means were statistically significant. However, if the Variances had been given, a Welch's T-Test could have been performed.

Anyway, in the Figure 2 below can be seen the relative means (all of them divided by the number of Bets) of those variables mentioned before in both Target and Control pots.

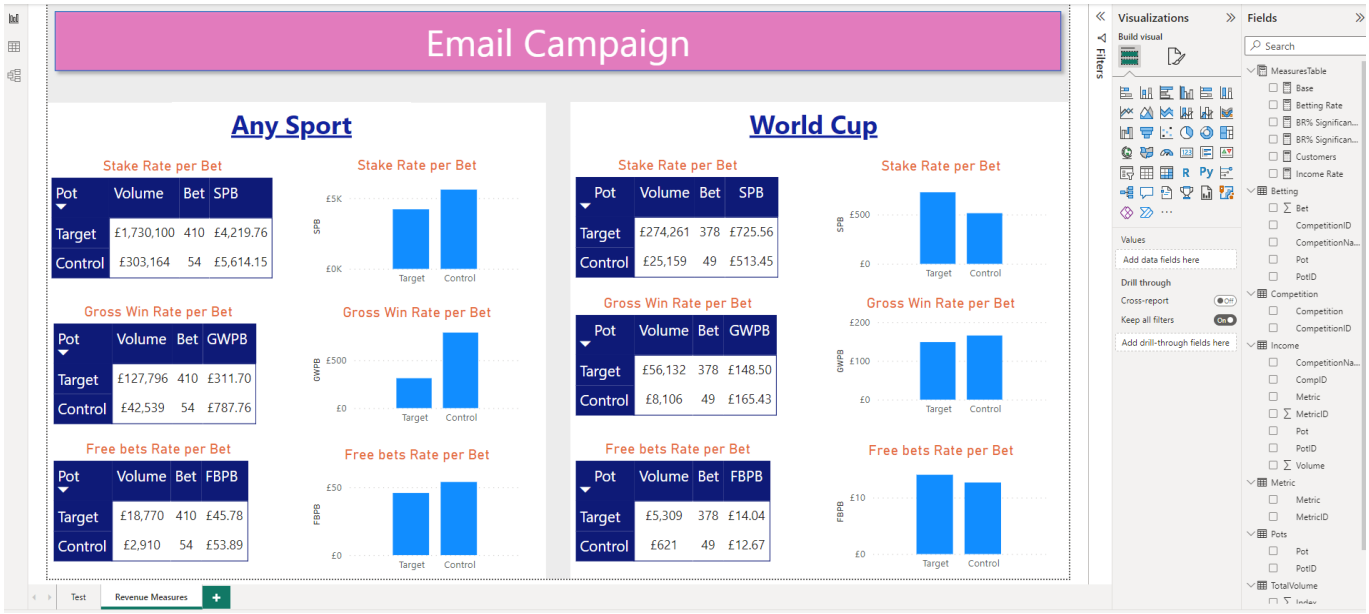

Figure 2: Revenue Measures

Figure 2 depicts that despite of the fact that the Target group had total amount of revenue (called Volume in the tables), this was because the pot was bigger than the Control one. Nevertheless, the bar charts show that the amount of money received per Bet was bigger for Control for all three variables in Any Sport section, whereas in World Cup just Target pot overcame the Control in Stakes (also, similarly for Free bets, but nothing substantial).

Finally, other 3 measures to take into consideration are Open Rate (email opens / Email sent), Click Through Rate (email Clicks / Email opens) and Opt out Rate (email unsubscribes / email sent). Thus, OR is 19.4%, CTR is 9.1% and OOR is 0.3%. These figures would need to be compared with other similar campaigns or other times to be able to determine if they are performing well or not. According to mailchimp.com, the average OR for a gambling industry is 21.62%, the average CTR is 3.3% and the average OOR is 0.15%, so then CTR is the one that is standing out with positive outcome.

To conclude, the Betting Rate Test results for both Any Sport and World Cup disciplines for this email campaign is giving enough confidence to the company to state that the strategy was right. However, other income variables (Stakes, Gross Win and Free Bets) are pointing to the opposite direction, with the result of not sending an email drove better response, with the exception for Stakes in World Cup (but this was not possible to test whether this was significant). Moreover, email measures like OR or OOR were in line with the average proportions, and CTR seemed to be performing better.

2. You have below summary table of emails sent over past 7 week together with statistics about opened emails. What chart would you chose here to give you stakeholder visual trend representation of open rate? Please draw that chart!

Reporting Week	Sent	Opened
202110	57,040	41,592
202111	80,075	52,720
202112	57,435	39,615
202113	79,046	51,159
202114	49,808	48,332
202115	26,845	22,015
202116	27,551	23,369

- What week you would you say was the best for email campaigns and why?

I created the table below (Figure 3) from the table I was given Using Power BI. I adjusted the "Reporting Week" column to get both date columns (Year and Week) and also worked out the Open Rate (Opened/Sent * 100)

Year	Week	Sent	Opened	OpenRate
2021	10	57,040	41,592	72.92%
2021	11	80,075	52,720	65.84%
2021	12	57,435	39,615	68.97%
2021	13	79,046	51,159	64.72%
2021	14	49,808	48,332	97.04%
2021	15	26,845	22,015	82.01%
2021	16	27,551	23,369	84.82%

Figure 3: Email metrics by week

Once having this table, I then plotted the data into a line chart for a time series using in the X axis the Week variable and in the Y axis the Open Rate variable to give the Figure 4 output.

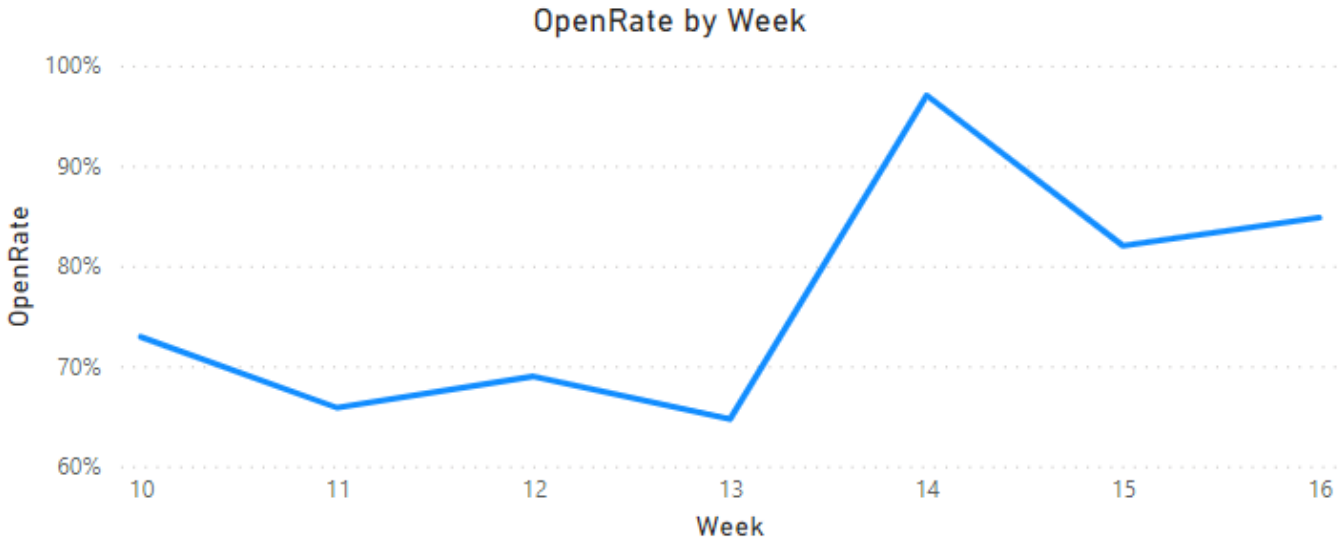

Figure 4: Open Rate Time Series by week

The Figure 4 shows that the week 14 in 2021 hit a summit, being the week with the highest Open Rate (97.04%), whereas the week before (week 13) was the weakest in the period in terms of Open Rate, with a 64.72%. The reasoning behind could be that after a few weeks having stable email open rates, an external factor like an important sport event came into play so players were keener to make bets and then the Open Rate rocketed. Having research, I found out that important Basketball

and Football (Champions League Semi-finals) competitions took place in that week 14 in 2021. Moreover, it could also be the case that some internal promotions were made and then players got more engaged.

APPENDIX

[1]

Any Sport:

E1,1 = $((410 + 183) \times (410 + 54))/(410 + 183 + 54 + 44)$ = 398.2

E1,2 = $((410 + 183) \times (183 + 44))/(410 + 183 + 54 + 44)$ = 194.8

E2,1 = $((410 + 54) \times (54 + 44))/(410 + 183 + 54 + 44)$ = 65.8

E2,2 = $((44 + 54) \times (183 + 44))/(410 + 183 + 54 + 44)$ = 32.2

X^2= ((410-398.2) ^^{2}) /398.2+ ((183-194.8) ^^{2}) /193.2 + ((54-65.8) ^^{2}) /65.8+ ((44-32.2) ^^{2}) /32.2= 7.51

World Cup:

E1,1 = $((378 + 215) \times (378 + 49))/(378 + 215 + 49 + 49)$ = 366.4

E1,2 = $((378 + 215) \times (215 + 49))/(378 + 215 + 49 + 49)$ = 226.6

E2,1 = $((378 + 49) \times (49 + 49))/(378 + 215 + 49 + 49)$ = 60.6

E2,2 = $((49 + 49) \times (215 + 49))/(378 + 215 + 49 + 49)$ = 37.4

X^2= ((378-366.4) ^^{2}) /366.4+ ((215-226.6) ^^{2}) /226.6+ ((49-60.6) ^^{2}) /60.6+ ((49-37.4) ^^{2}) /37.4= 6.73

[2]

Model relationship:

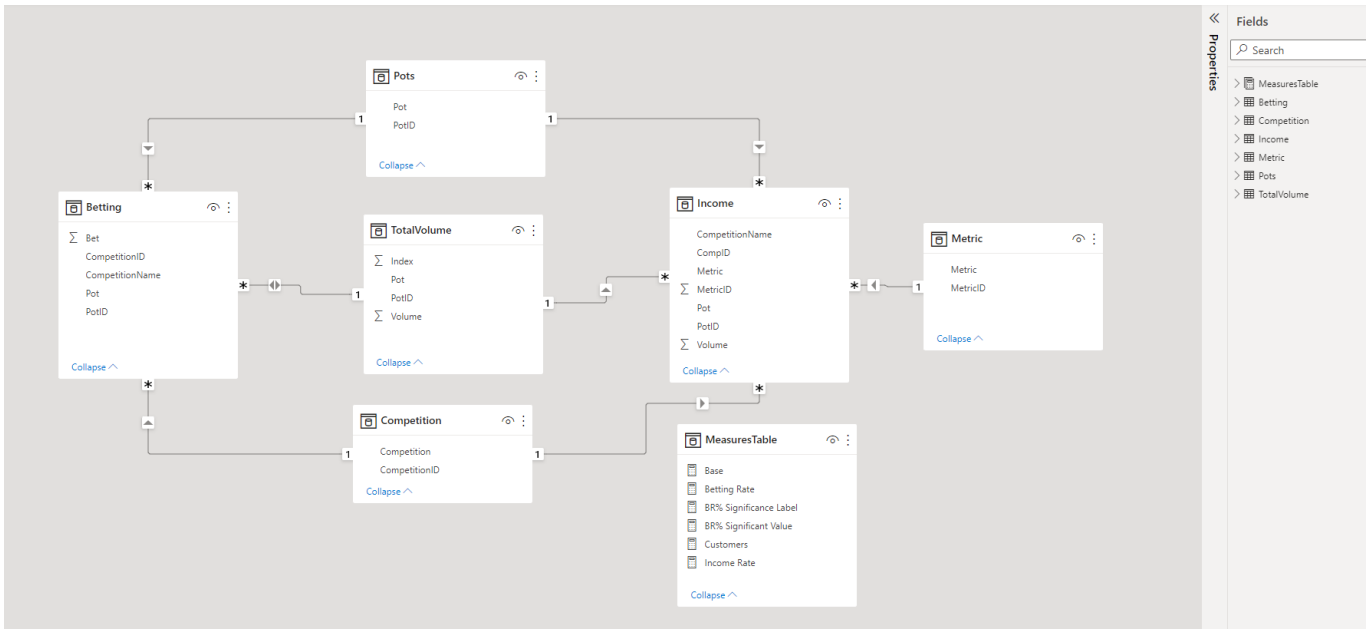

Figure 5: Power BI Campaign Model Relationship

BR% Significance Value DAX Formula:

Figure 6: BR% Significance Value DAX