Customer Analytics Assignment 3

Beiming Zhang

BU.450.760.K1.SP24

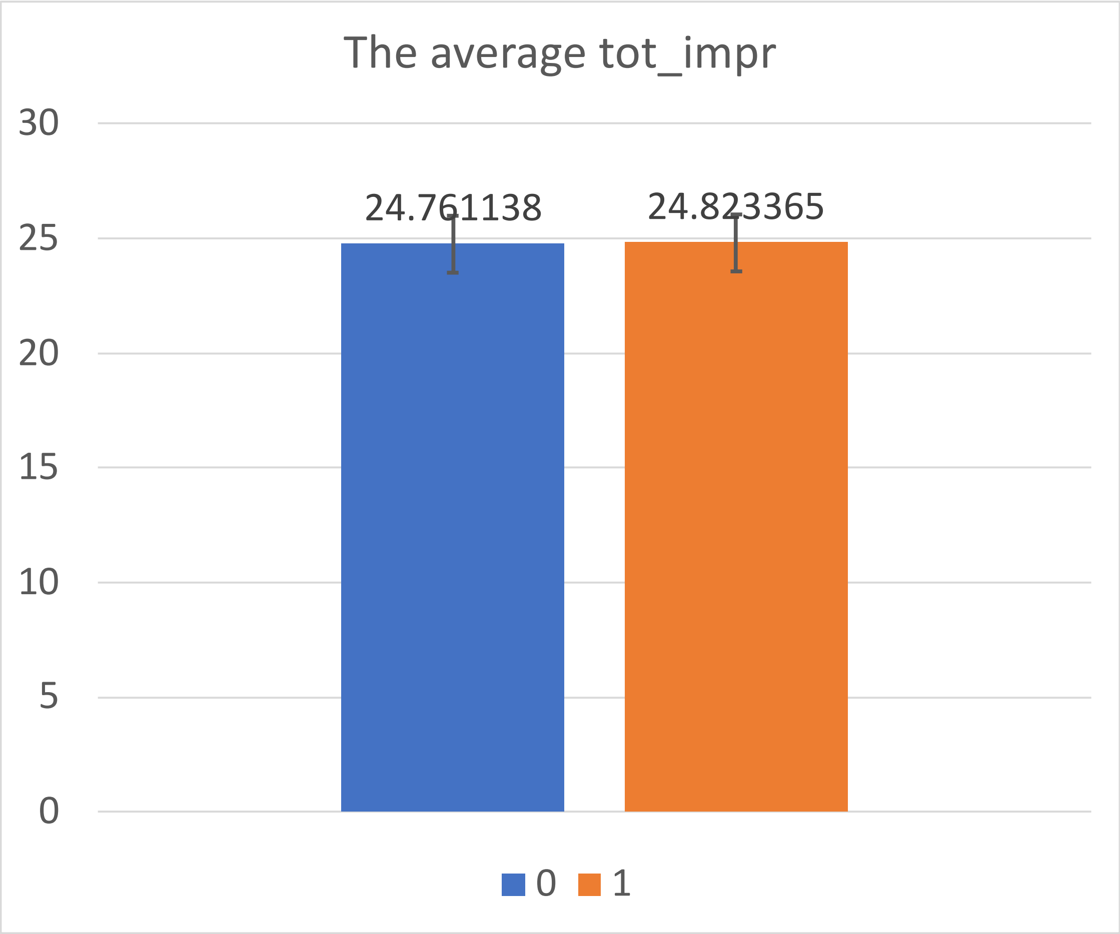
Completely independent assignment.

Question 1(a)

- Excel Formula: *=1-AVERAGE(rocketfuel_data!B:B)*
- The share of users allocated to the control group is 3.999993%, close to 4%.

Question 1(b)

- Using charts and T-tests to verify whether or not the campaign properly randomized consumers into the test and control group.
- The blue bars in the chart represent the control group, while the orange bars represent the test group.
- From the graph, we can observe that the total number of ad impressions per user in both groups are very close, and the 95% error bars are also very similar.



Question 1(b)

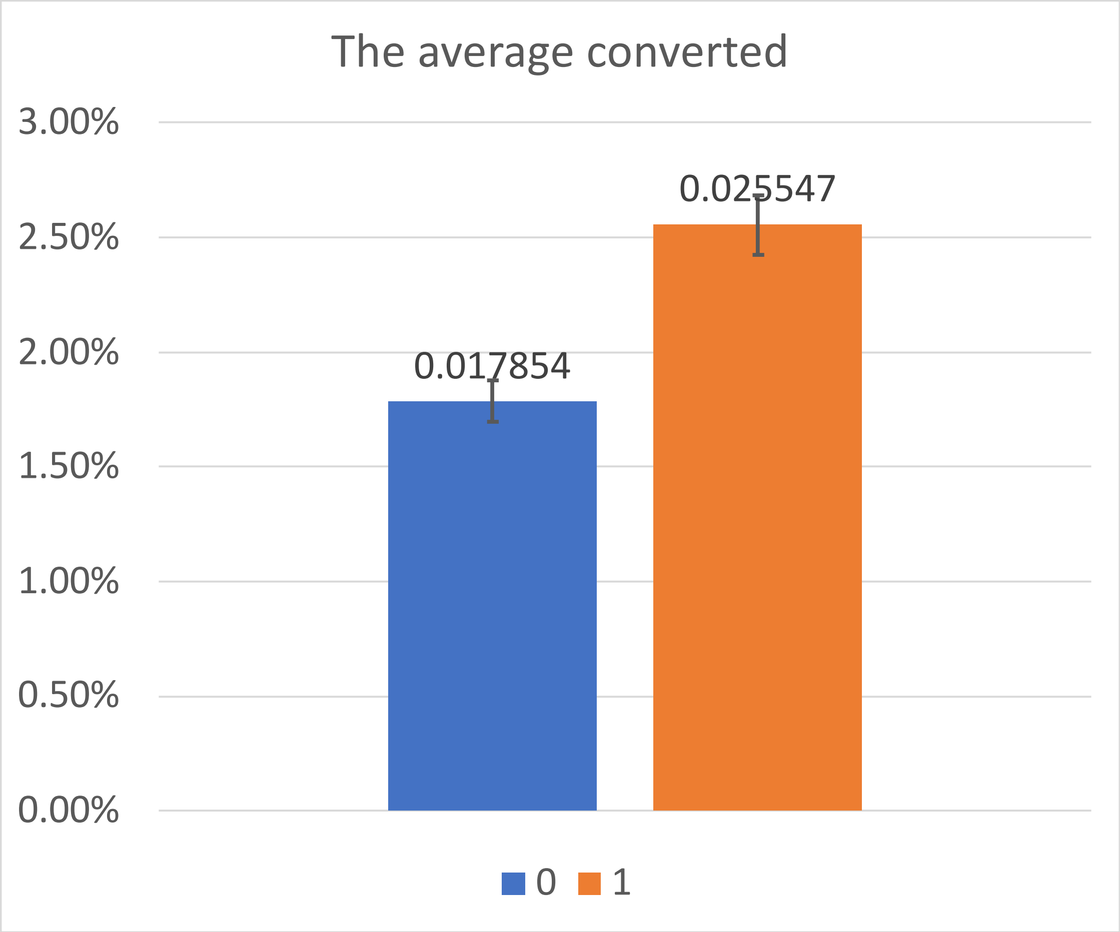
- In the two-tailed T-test section, use Excel formula to output the probability of observing the data or more extreme data under the condition that the null hypothesis H0 (no significant difference between the two datasets) is true.
- Excel Formula: *=T.TEST(FILTER(rocketfuel_data!D:D,rocketfuel_data!B:B=1),FILTER(rocketfuel_data!D:D,rocketfuel_data!B:B=0),2,2)*
- Using the FILTER() function to determine the control group and test group, select two-tailed distribution and two-sample equal variance (homoscedastic) as parameters for the T.TEST() function.
- The p-value is 0.830613, which is much higher than 0.01. Therefore, accept the null hypothesis H0, indicating that there is no significant difference between the two datasets.
- Considering both the charts and the results of the T-test, the random selection of the group is reasonable.

Question 1(c)

- Firstly, random assignment can help ensure that the control group and the test group are similar in all potential confounding variables before the test begins. This can effectively reduce the impact of these variables on the results.
- Secondly, through random assignment, the test results are more likely to be generalized to a wider population, as this method can reduce sample selection bias. This is crucial for the scalability of this business activity.
- Lastly, under the premise of random selection, when there are differences between the control group and the test group, any differences in results between the groups can be more confidently attributed to the controlled variables, rather than to other confounding factors.

Question 2

- Using charts and T-tests to verify whether the campaign effective in increasing conversion rates.
- The blue bars in the chart represent the control group, while the orange bars represent the test group.
- From the graph, we can see that there is a significant difference in the average conversion rates between the two groups, and there is no overlap in the 95% error bars.



Question 2

- In the one-tailed T-test section, use Excel formula to output the probability of observing the data or more extreme data under the condition that the null hypothesis H0 (conversion rate is not higher in the test group) is true.
- Excel Formula:
  =T.TEST(FILTER(rocketfuel_data!C:C,rocketfuel_data!B:B=1),FILTER(rocketfuel_data!C:C,rocketfuel_data!B:B=0),1,2)
- Using the FILTER() function to determine the control group and test group, select one-tailed distribution and two-sample equal variance (homoscedastic) as parameters for the T.TEST() function.
- The p-value is 0.000000, which is much lower than 0.01. Therefore, reject the null hypothesis H0, indicating that conversion rate is higher in the test group.
- Considering both the charts and the results of the T-test, the campaign effective in increasing conversion rates.

Question 3(a-c)

a) The number of people who watched the . Based on "a convert user is worth about $40", we can calculate that the money TaskBella made more by running the ad campaign is $173,719.29(Profit = (conversion rate of exposed users – conversion rate of control group) * number of exposed users * value of converted user).

b) The number of ads impressions are roughly 14.5 million. Based on the statement "an average CPM of $9", we can calculate that the cost of the campaign is $131,374.64.

c) According to a) and b), the ROI from the campaign is 0.32.

Total number of people	588,101.00	=COUNT(rocketfuel_data!A2:A588102)
Number of exposed users	564,577.00	=SUM(rocketfuel_data!B2:B588102)
Number of ads	14,597,182.00	=SUM(rocketfuel_data!D:D)
Cost of ads	$ 131,374.64	=B4/1000*9
Number of purchasers	14,843.00	=SUM(rocketfuel_data!C2:C588102)
Revenue	$ 593,720.00	=B6*40
The money ads make	$ 173,719.29	=B3*(0.0255-0.01789)*40
Naïve ROI	3.52	=B7/B5-1
ROI	0.32	=B8/B5-1

Question 3(d)

- In this scenario, using a control group is worthwhile as it demonstrates the significant effect of the advertisement through the A/B test method without notably impacting the ROI (considering the control group only accounts for 4%).
- Given the sizable sample (see the following page), we also have sufficient evidence to demonstrate the effectiveness of the advertising.
- The control group allows for a more direct observation of the advertisement's effect while controlling the impact of external variables.
- This helps ensure the validity of the test results and enhances the credibility of the conclusions. Based on the conclusions of the A/B test, the advertising scheme can be further promoted.

Question 3(d)

- Calculate the minimum sample size required using a calculator. For α and β, default values of 0.05 and 0.20 are chosen. q1 is 0.96. According to Question 1(a), this is an approximate value representing the proportion of the test group. P0 is 0.0179, and P1 is 0.0255. These values are derived from the results of Question 2.
- The minimum total group size is 79,408, which is much smaller than the total sample size of 588,101.
- Therefore, the control group can be smaller. In this scenario, the minimum control group size is approximately 3,177.

α (two-tailed) =	0.05	Threshold probability for rejecting the null hypothesis. Type I error rate.
β =	0.20	Probability of failing to reject the null hypothesis under the alternative hypothesis. Type II error rate.
q_1 =	0.96	Proportion of subjects that are in Group 1 (exposed)
q_0 =	0.04	Proportion of subjects that are in Group 0 (unexposed); $1-q_1$
P_0 =	0.0179	Risk in Group 0 (baseline risk)

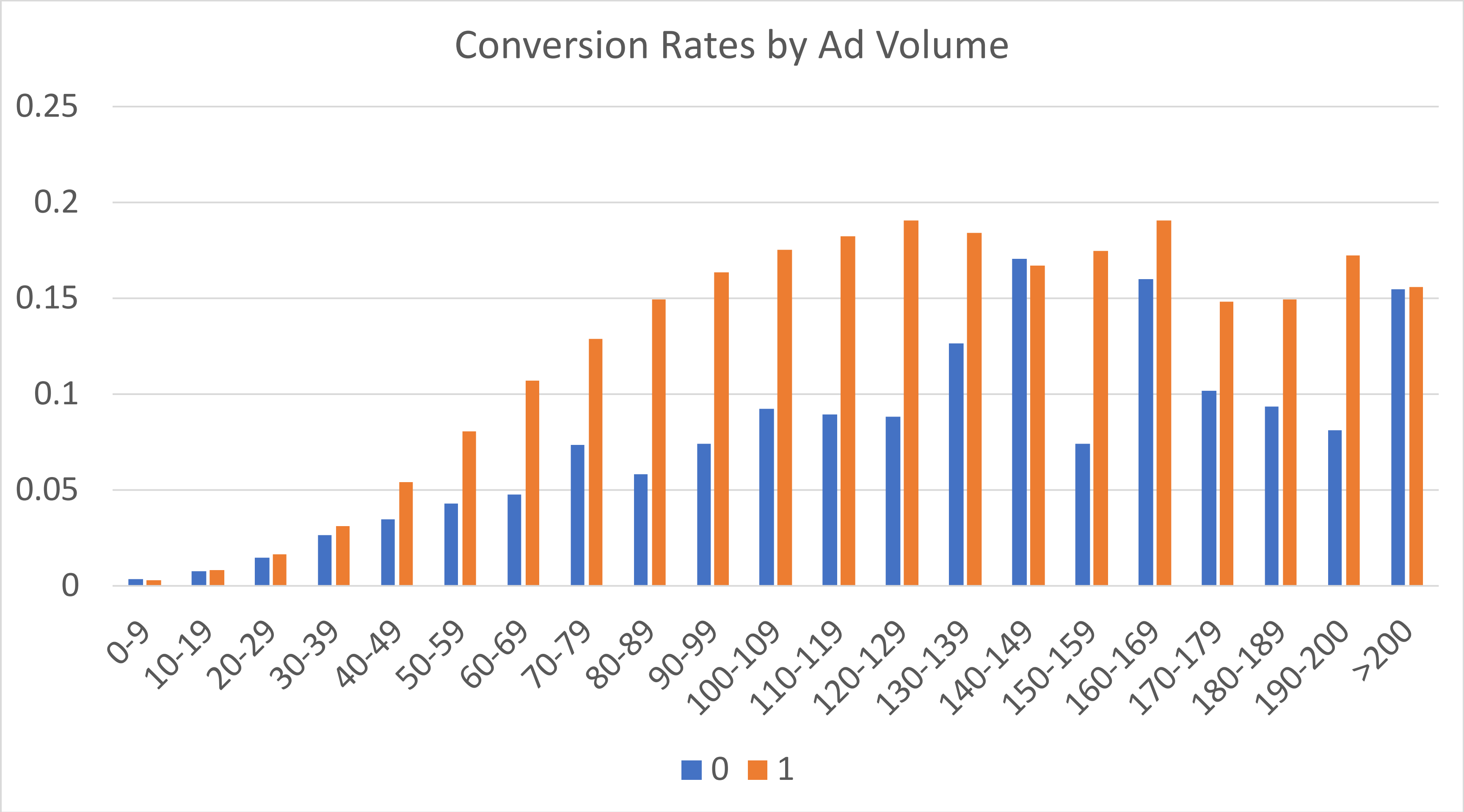
Enter any ONE of the following three parameters (the other two will be calculated automatically):

P_1 =	0.0255	Risk in Group 1 (exposed)
OR =	1.4357	Odds ratio $(P_1/(1-P_1))/(P_0/(1-P_0))$
RR =	1.4246	Risk ratio (P_1 to P_0)

Calculate

The standard normal deviate for α = Z_α = 1.9600
The standard normal deviate for β = Z_β = 0.8416
Pooled proportion = $\bar{P} = (q_1 * P_1) + (q_0 * P_0)$ = 0.0252
$A = Z_\alpha \sqrt{\bar{P}(1-\bar{P})(1/q_1 + 1/q_0)}$ = 1.5675
$B = Z_\beta \sqrt{P_1(1-P_1)(1/q_1) + P_0(1-P_0)(1/q_0)}$ = 0.5741
$C = (P_1-P_0)^2$ = 0.0001
Total group size = $N = (A+B)^2/C$ = 79408
Continuity correction (added to N for Group 0) = $CC = 1/(q_1 * |P_1-P_0|)$ = 137

Question 4



Conversion Rates by Ad Volume
0.25
0.2
0.15
0.1
0.05
0
0-9
10-19
20-29
30-39
40-49
50-59
60-69
70-79
80-89
90-99
100-109
110-119
120-129
130-139
140-149
150-159
160-169
170-179
180-189
190-200
>200
0
1

Question 4

- Construct a bar chart of conversion rates as a function of the number of ads displayed to consumers using a pivot table and group the data in sets of ten. Please refer to the chart on the previous page.
- The blue bars in the chart represent the control group, while the orange bars represent the test group.
- I believe there is NOT an obvious frequency effect to advertising. Although within certain intervals, such as from 40 to 129, the data shows a certain frequency effect, indicated by a gradually increasing trend in the difference in average conversion rates between the control group and the test group, it is not reasonable to assert the existence of a frequency effect overall. Especially after 130, the relationship between the number of advertisements and their effectiveness is not stable and sometimes even shows adverse effects.
- Therefore, it is not reasonable to directly state that a frequency effect to advertising exists.

Thank you

Beiming Zhang

BU.450.760.K1.SP24