WA-Assignment#2-RestaurantGrades

Course – Web & Social Analytics (ITSM 6209)

Team Members – Rakesh Singh

Professor - Dr. Wenjing Duan

Date – 5-June-2021

**Ans 1:**

To understand the data provided and outcome of variable, my starting point is the table below which represents a sum of all metrics. Difference in Pageview from RG Treatment 1 and then how the pageviews are converted into call and reservations in percentage.

**The table below represents the conversion of Pageviews to calls and reservations**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Treatment | Sum of pageviews | Difference in % (Pageviews) | Sum of calls | Pageview conversion to calls | Sum of reservations | Reservation conversion to calls |
| 0 | 4197794 |  | 340196 | 8.10% | 339604 | 99.83% |
| 1 | 5011908 | 19% | 373885 | 7.46% | 340212 | 90.99% |
| 2 | 4832110 | 15% | 417145 | 8.63% | 416805 | 99.92% |
| Grand Total | 14041812 | NA | 1131226 | NA | 1096621 | NA |

**The table below represents the comparison in Pageviews, Calls and Reservations for Treatment 1 and 2**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Treatment | Sum of pageviews | Difference in % (Pageviews) | Sum of calls | Difference in calls % | Sum of reservations | Difference in % (Reservations) |
| 0 | 4197794 | 0% | 340196 | 0% | 339604 | 0% |
| 1 | 5011908 | 19% | 373885 | 10% | 340212 | 0% |
| 2 | 4832110 | 15% | 417145 | 23% | 416805 | 23% |
| Grand Total | 14041812 |  | 1131226 |  | 1096621 |  |

Although, pageview is the most important variable, however, calls and reservations can be considered and an outcome variable. From the table above it can be easily established that RG Treatment 0 is at bottom and considering that as a baseline I have done a quick comparison between RG 0, 1 and 2. RG Treatment 1 has the maximum pageview and 19% more than treatment 0. Treatment 2 is also close to Treatment 1 and has a better call rate and reservations. Moving on, I am trying to now see how many pageviews are being converted to calls and eventually how many calls are converted to reservations. RG Treatment 0 is able to convert 8.10 % of pageviews to call and has a 99.83% conversion ration when it comes to reservations from calls, this makes RG Treatment 0 better than RG Treatment 1 with respect to conversion rate of Calls and Reservations. Despite of the fact that RG Treatment 1 has the maximum pageviews but only 7.46% of pageviews are converted to calls and then out of these calls only 90.99% customers end-up making a reservation. It is clearly visible that RG Treatment 1 conversion ratio to calls and reservations are worst amongst the competition between RG Treatment 0, 1 and 2. Now let’s analyze RG Treatment 2 – It is best in converting the pageview to calls and then to reservation with a conversion ratio of 8.63% for calls and 99.92% for reservations. This means that every call that lands to RG Treatment 2 for booking is almost converted to a reservation. From the data above it appears that RG Treatment needs to further deep dive and ascertain a reason why the pageview has a low call rate and then why the call conversion to reservation is only 90% compared from RG Treatment 0 and RG Treatment 2 since they are able covert 99% of calls to reservations.

Now, I am proceeding towards the regression analysis and evaluate as to which variable will make the maximum difference.

Note: The dummy variable created for the analysis is as follows, when treatment is 1 Dummy 1 will have a value of 1, when treatment is 2 Dummy 2 will have a value of 1. Also, in restaurant type Chain is converted to 1 and Independent is converted to 0. The analysis has been performed on Python.

|  |  |  |
| --- | --- | --- |
| **Treatment** | **Dummy 1** | **Dummy 2** |
| 0 | 0 | 0 |
| 1 | 1 | 0 |
| 2 | 0 | 1 |

**Target Variable Pageviews**

The 1st variable chosen for the regression is **Pageviews**, please find the analysis done below.

For Pageview R Square is **91.3%**

**Table

Description automatically generated**

**Target Variable Calls**

The 2nd variable chosen for the regression is **Calls**, please find the analysis done below.

For Calls R Square is **58.0%**

Graphical user interface, table

Description automatically generated

**Target Variable Reservations**

The 2nd variable chosen for the regression is **Reservations**, please find the analysis done below.

For Calls R Square is **59.7%**

Graphical user interface, table

Description automatically generated

We can now conclude from the analysis above that the pageviews is the most significant variable with a R square of 91.3%. Hence statically pageview as a target variable is most impactful variable. Additionally, to ensure that we have a statically significant result, the values of coefficient for independent variables was checked along with respective p-values. The p-values were all less than 0.05 (p>0.05) re-iterating their statical significance. We also observed that the coefficients are all positive.

**Ans 2:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Row Labels | Sum of pageviews | Change in Pageviews % | Sum of calls | Change in calls % | Sum of reservations | Change in reservation % |
| 0 | 4197794 |  | 340196 |  | 339604 |  |
| chain | 2398285 | 160277 | 159700 |
| independent | 1799509 | 179919 | 179904 |
| 1 | 5011908 | 19.39% | 373885 | 9.90% | 340212 | 0.18% |
| chain | 2761587 | 15.15% | 176085 | 9.86% | 160404 | 0.44% |
| independent | 2250321 | 25.05% | 197800 | 9.94% | 179808 | -0.05% |
| 2 | 4832110 | 15.11% | 417145 | 22.62% | 416805 | 22.73% |
| chain | 2762284 | 15.18% | 192315 | 19.99% | 192008 | 20.23% |
| independent | 2069826 | 15.02% | 224830 | 24.96% | 224797 | 24.95% |
| Grand Total | 14041812 | NA | 1131226 | NA | 1096621 | NA |

The above table is a pivot of RG and I am trying to establish a pattern to understand the differences between Chain and Independent restaurants. We can determine that the data statically are different for Chain and Independent restaurants. Let’s try to discuss and understand, it is clearly visible that the overall independent restaurant is more significant than the Chain restaurant.

Let me explain this is details treatment wise, Treatment 1 – Independent restaurant pageviews increased by 25% compared to the chain restaurant which is only 15%. Apart from this the difference in other metrics like calls and reservations are negligible. However, independent restaurant has a negative value in reservations but that’s something which isn’t very much significant (-0.05%). So, overall, in Treatment 1 Independent restaurant have managed to run this campaign well.

Treatment 2 – Chain restaurant has slightly high % than Independent restaurant in pageviews, which is not that signification since the different between these two are just (15.18-15.02=0.16%). When we look at other metrics like calls and reservations independent restaurant did well as compared to the Chain restaurant and is clearly ahead by around 5%.

**Recommendation**

From the second analysis we understood that chain-based restaurants naturally get more views, calls and reservations. Even when RestaurantGrades used strategies(both) that affect independent restaurants more to increase their counts, they still stayed behind. Hence, the recommendation would be RestaurantGrades needs to consider using different strategies for chain and independent since they fundamentally are different.

**Ans 3:**

1. First let’s see how both the strategies vary with regards to pageviews, calls and reservations from the pivot table that is discussed in Answer-1.

* Treatment-1 showed more percentage increase in **pageviews** than treatment-2
  + Treatment-1 showed 19% increase in pageviews whereas treatment-2 showed 15% increase.
* However, treatment 2 showed much higher percentage increase in both **calls** and **reservations** than treatment-2.
  + Treatment-1 only showed 10% increase in calls whereas treatment-2 showed a much higher increase of 23%
  + The difference is much higher in case of reservations. Treatment-1 showed almost no increase with just 0.18% whereas treatment-2 showed 23% increase here too.

From the above analysis, it is clear that the treatment-2 i.e the alternate algorithm performs well for calls and reservations which are proved to be the important metrics to consider from our answer-1. Hence the alternate algorithm is better.

1. Let’s prove this with regression analysis (Unlike question 1 we have opted different variables here)

Since we are dealing with count data here, we shouldn’t use the regular Ordinary least square linear regression model, as OLS regression has some assumptions.

* that the outcome variable is distributed from negative infinity to positive infinity whereas our variables start from 0
* Second, the data generation process for count data may be very different from what is required for OLS.

And due to this reason, we need to use a statistical model that models count data, non-negative integers. We use Negative Binomial Regression. The inputs for this model are *treatment* and *restaurant\_type*. The column *treatment* is divided into two dummy columns. treatment-1 and treatment-2. If both are 0 it means it is treatment-0. 1,0 and 0,1 means treatment-1 and treatment-2 respectively. The following table depicts the data.

A screenshot of a computer

Description automatically generated with medium confidence

Starting Regression Analysis

**Pageviews as target variable**

Formula is pageviews ~ treatment1 + treatment2 + restaurant\_type

Text

Description automatically generated

As we can see, both treatment-1 and treatment-2 have a p value < 0.05 which means that they are statistically significant in affecting the pageviews. The coefficients for treatment-1 and treatment-2 are 0.19 and 0.14 respectively. These are very similar. Hence using pageviews we can’t decide which algorithm is better.

**Calls as target variable**

Formula is calls ~ treatment1 + treatment2 + restaurant\_type

Text

Description automatically generated

As we can see, here also both treatment-1 and treatment-2 have a p value < 0.05 which means that they are statistically significant in affecting the calls. The coefficients for treatment-1 and treatment-2 are 0.09 and 0.20 respectively. Since the treatment-2’s coefficient is higher we can say that treatment-2 is better

**Reservations as target variable**

Formula is reservations ~ treatment1 + treatment2 + restaurant\_type

Graphical user interface, text

Description automatically generated

As we can see, here treatment-1 doesn’t even have p value < 0.05. It means that it is statistically insignificant in affecting the number of reservations. Whereas, the p value for treatment-2 is < 0.05 making it significant. This is reflected in the coefficients as well. The coefficients for treatment-1 and treatment-2 are 0 and 0.20 respectively. We can say that treatment-2 is better

Finally, Treatment-2 i.e the alternate algorithm proved to be the best choice to increase the no.of calls and reservations. Even if the current algorithm is better for the no.of pageviews, the alternate algorithm isn’t behind that much. Also, the calls and reservations are the most important metrics. **Hence, RestaurantGrades need to switch to the alternative algorithm for better results.**