UTA Business Analytics Competition
- Hewlett Packard Enterprise Use Case

Name: Bhupesh Kumar Srivastava   Email: bks210000@utdallas.edu
Name: Piyush Raisinghani   Email: bxr200045@utdallas.edu
Name: Jielin Zhang   Email: jxz210026@utdallas.edu
Name: Yixiao Ji   Email: yxj210010@utdallas.edu
Name: Min Shi   Email: min.shi@utdallas.edu

Abstract

In today's business landscape, digital marketing has become a crucial component for most companies. However, due to its extensive reach, a seemingly generous marketing budget may quickly prove inadequate. This project aims to assist Super Server Company's marketing department in campaign planning and optimizing marketing budget allocations to maximize the impact of their digital marketing efforts.

Tableau, Excel, Python, and Jupyter Notebook were used to analyze the dataset in this project. Initially, Tableau was used for exploratory data analysis (EDA), and after dropping some unimportant columns, regression models were run to reach a solution.

The budget allocation process was adjusted to include previously neglected channels and audience types based on performance coefficients derived from models. The optimization of the budget was guided by these coefficients while adhering to the provided restrictions.

Introduction

To maximize the impact of their upcoming campaigns, Super Server Company is seeking to determine the most effective budget allocation for each audience group and digital channel. To inform this decision, historical data on marketing performance is analyzed, utilizing a dataset comprising various metrics generated by previous ad campaigns.

Solutions

Solution Formulation:

Upon visualization of the dataset, it became clear that cost and ad metrics should be the basis of our model. Given that the problem required an analysis of historical metrics to determine expenditure, we opted for a regression-based solution. We utilized visualizations to examine the linear relationship between cost and ad metrics, ultimately selecting linear regression as our modeling technique, given its assumption of a linear relationship between independent and dependent variables.

Linear regression is an intuitive and easily interpretable technique, making it a suitable choice for our analysis. To accommodate the constraints affecting all ad channels and audience types, we incorporated them as indicator variables in our model, providing a granular understanding of budget optimization and enabling better business decisions. By leveraging mathematical modeling and optimization techniques, we can enhance our ability to stay ahead of the curve and make more effective use of our resources, ultimately driving better business outcomes.

Solution Implementation:

Our analysis began with exploratory data analysis (EDA) to gain insights and a better understanding of the data. During this process, we identified certain columns that were irrelevant to our objective and decided to drop them. Additionally, we noticed that some columns had an excessive number of null values, which we addressed by either imputing missing data or dropping the columns altogether.

We calculated performance metrics from ad metrics i.e. Clicks, impressions, etc. This performance was calculated by taking weighted sum of ad metrics and weights. Weights for each ad metric was calculated by this formula:
Weight of clicks = (sum of all metrics – sum of clicks)/sum of all metrics
Where sum of all metrics = sum of clicks + sum of impressions + sum of video completes + ... + sum of form completes
We used this formula to penalize attributes which have numerically higher values and to give significance to attributes which did not have a larger sum.
We then tested several regression models using different dependent and independent variables to derive a performance variable that guided the optimization of our budget across various channels and audience types. We were able to achieve close to 23% R^2 with meaningful and significant coefficients.
These coefficients were used in Excel for optimizing the budget, we identified certain channels and audience types that did not perform well, leading to a decision to eliminate those specific ad channels within those audience types entirely.

<u>Model Evaluation Metrics:</u>
ANOVA analysis for the model

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                   Perf   R-squared:                       0.227
Model:                            OLS   Adj. R-squared:                  0.227
Method:                 Least Squares   F-statistic:                 1.112e+04
Date:                Tue, 14 Mar 2023   Prob (F-statistic):               0.00
Time:                        16:20:53   Log-Likelihood:            -2.6772e+06
No. Observations:              265388   AIC:                         5.354e+06
Df Residuals:                  265380   BIC:                         5.355e+06
Df Model:                           7
Covariance Type:            nonrobust
================================================================================================
                                          coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------------------------
const                                -241.1411     15.494    -15.563      0.000    -271.509    -210.773
$ Spend                               104.5746      0.475    220.143      0.000     103.644     105.506
Ad Digital Channel_Programmatic Display  -51.1644  34.790     -1.471      0.141    -119.352      17.023
Ad Digital Channel_Search            -262.7477     73.313     -3.584      0.000    -406.438    -119.057
Ad Digital Channel_Social              72.7710     45.623      1.595      0.111     -16.649     162.191
Audience Type_Audience Type 1       -1627.7216     82.390    -19.756      0.000   -1789.205   -1466.239
Audience Type_Audience Type 2        -805.9432     34.578    -23.308      0.000    -873.715    -738.172
Audience Type_Audience Type 3        3501.0592     33.570    104.290      0.000    3435.262    3566.856
Audience Type_Audience Type 4          20.3209     34.036      0.597      0.550     -46.390      87.031
Audience Type_Audience Type 5       -1328.8563     33.284    -39.925      0.000   -1394.092   -1263.621
==============================================================================
Omnibus:                   414564.198   Durbin-Watson:                   1.933
Prob(Omnibus):                  0.000   Jarque-Bera (JB):        528100922.301
Skew:                           9.700   Prob(JB):                         0.00
Kurtosis:                     220.673   Cond. No.                     1.87e+16
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The smallest eigenvalue is 9.27e-25. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.
```

Results

Describe the optimal budget plan for each digital driver and audience type.

Conclusions

<u>Results Summarization:</u>

Channel Type	Audience Type	Budget	Performance Coefficients
Programmatic Display	1	$20,000.00	0.021206
Programmatic Display	2	$0.00	0.046803
Programmatic Display	3	$0.00	0.166879
Programmatic Display	4	$440,000.00	0.072494
Programmatic Display	5	$20,000.00	0.030517
Search	1	$0.00	0.014636
Search	2	$20,000.00	0.040233

Search	3	$0.00	0.160309
Search	4	$0.00	0.065923
Search	5	$0.00	0.023947
Social	1	$0.00	0.025037
Social	2	$0.00	0.050634
Social	3	$500,000.00	0.17071
Social	4	$0.00	0.076324
Social	5	$0.00	0.034347

Limitations:

We should keep our average campaign budget to maximize our return on investment. The data set provided did not have enough successful observations with respect to form completion and that proved to be a major limitation for any model that we ran. If we provide majority failures, the model is bound to predict that the future would also produce failed results. Audience types were provided as numbers rather than actual audience groups, which could have led to better insights and better models. Since the budget was varying for every month, we could not capture seasonality effects in our model.

Recommendations for next steps:

1. More successful observations across the dataset, meaning that the dataset should include more observations where ad campaign proved to be successful in order to determine future ad campaigns.
2. Better data quality i.e. more data which does not have 0s and NaN values.
3. Understanding the audience gender can also help in optimizing ad revenue.
4. The previous campaigns shows that we have allocated only 11% expenditure on social platforms from June 2022 through Feb 2023, which may have led to missed opportunities. Spending more ad revenue on social platforms can provide useful since most of the adult population uses it.
5. Previous results show that we have allocated 21% of our budget on audience type 1, but the analysis proves that we need to expand our budget for audience type 3 for better return on investment.
6. Keep a constant budget to observe seasonality monthly.
7. To optimize returns, it is not advisable to exceed a campaign budget by more than $125

References

Sklearn docs, Pandas docs, Python docs, Statsmodel docs, Seaborn docs, Matplotlib docs, Numpy docs