Geo-matched Test For the Effectiveness of Google Performance Max Campaign

Business Problem & Executive Summary

ACME Manufacturing seeks to evaluate the effectiveness of its marketing campaigns, particularly the Google Performance Max campaign, on a national scale. To achieve this, our analytics consulting group has designed a geo-holdout experiment targeting specific states while ensuring statistical integrity and actionable insights. Our approach involved several key steps:

- **Treatment Group Selection:** We carefully selected four treatment markets—Tennessee, Missouri, Montana, and New Mexico—based on factors such as representativeness, statistical significance, and proportion of national sales.
- **Synthetic Control Method:** we constructed a virtual control group that mirrors the treatment group in the pre-treatment period statistically. This approach allowed for accurate comparisons and minimized resource requirements compared to traditional geo-match trials.
- **Treatment Effect Calculation:** By comparing actual sales data from the treatment period with predicted sales from the synthetic control group, we estimated the average causal impact of the Google Performance Max campaign. The analysis revealed a significant reduction in sales when the campaign is held out.
- **Population Causal Effect:** Extending our analysis to the national level, we fitted a linear regression model to estimate the impact of the campaign on national sales. The results showed a substantial decrease in national sales over the treatment period.
- **Recommendation:** Based on our findings, we recommend carefully evaluating the cost of the Google Performance Max campaign against the estimated losses. The result of having assumptions about the real world leads us to recommend having this campaign.

The reason why we choose the treatment group

Firstly, we choose to **remove Delaware** from our analysis since it is relatively small in terms of area compared to other states, indicating that people who live in the state are likely to commute to neighboring states for employment opportunities. Selecting Delaware as the treatment group for our hold-out experiment could pose a risk of violating the consistency assumption due to cross-border exposure to marketing campaigns. Residents of Delaware may be exposed to campaigns in adjacent states, potentially leading to **data leakage** and compromising the integrity of our experimental results.

Secondly, we want to choose the smallest group of treatment markets that can represent national response behavior.Given that our hold-out experiment will continue one month, and that forecasting at an aggregate level tends to be more stable and accurate than at a detailed level, we plan to adopt **a four-week rolling window** approach to restructure our data. This method will involve using **the average values**, with each row representing the mean sales for each state over a cumulative four-week period. Following this reorganization, we will conduct a linear regression analysis, employing sales figures from all states as predictors to estimate the overall national average sales amount.

Thirdly, after developing the model, we must prioritize markets that exhibit significant p-values. In our initial model analysis, we identified ‘New Mexico’, ‘Montana’, and ‘Missouri’ as having p-values less than 0.1, suggesting that these states are highly relevant predictors. Subsequently, we must identify the fourth target market by sequentially incorporating each state into the model. This process focuses on the **adjusted R-squared** value to ensure that the chosen treatment group can effectively explain the variability in national sales level. In models where 'Kentucky' or 'Tennessee' serve as the fourth predictors, both exhibit the highest adjusted R-squared values. Additionally, the p-values for all predictors in these models are below 0.05, indicating statistically significant contributions. Given our need to conserve resources and mitigate risks associated with the hold-out experiment, it is unnecessary to incorporate a fifth state, as doing so would not result in an improvement in our adjusted R-squared value.

Last but not least, to ensure that the treatment effects from these markets provide meaningful insights for national-level decision-making, it is crucial that the proportion of the sum of the selected states' values to the national sales amount is not too small. We have calculated the proportions for the selected treatment groups as shown in the table below. Given that the group comprising 'New Mexico', 'Montana', 'Missouri', and 'Tennessee' displays the highest proportion, we have selected this group as our final choice.

Combination	Proportion
Missouri, Montana, New Mexico, Tennessee	4.52
Missouri, Montana, New Mexico, Kentucky	3.44
Missouri, Montana, New Mexico, Mississippi	3.02
Missouri, Montana, New Mexico, Arkansas	3.01
Missouri, Montana, New Mexico, Kansas	2.99

Synthetic Control

After choosing these four states as the treatment group—Tennessee, Missouri, Montana, and New Mexico—we have identified these states as the treatment market for assessing the campaign's effectiveness. These markets, due to their inclusion in the treatment group, will not be considered for control comparisons. The exclusion of these states, along with the other 5 states deemed unavailable for control due to concurrent marketing activities, requires careful selection of control markets to ensure an accurate comparison.

Our analysis begins with the data preparation process, where we remove non-numeric information and irrelevant columns, such as 'Week', from our dataset. This allows us to focus on the numerical order data, which is critical for our analysis. We then proceed to construct a synthetic control group by selecting geographies from the pre-treatment data, ensuring that the treatment geographies and any unavailable states are excluded from this group.

Utilizing linear regression, we fit a model using the control group's order data as the independent variable and the aggregated pre-treatment sales from the treatment markets as the dependent variable. This model aims at predicting what the sales in the treatment geographies would have looked like in the absence of the marketing campaign, thus creating a baseline for comparison.

Several reasons led to our decision to choose the synthetic control method over geo-match trials for evaluating marketing campaign effectiveness. Firstly, synthetic control is more cost-effective as it utilizes existing aggregate data and reduces the need for extensive new data collection, thereby lowering overall costs. Secondly, it offers greater operational feasibility by simplifying the implementation process; geo-match trials require meticulous pairing and are operationally demanding. In contrast, synthetic control constructs a virtual control group that mirrors the treatment group statistically. Lastly, synthetic control provides greater flexibility for conducting multiple experiments simultaneously without the need for unique setups for each trial, allowing for scalable and rapid testing across different campaigns. These advantages make synthetic control a preferable method for resource conservation, operational simplicity, and analytical robustness in dynamic testing environments.

Treatment Effect Calculation

To assess the causal effect of our geo-marketing campaigns, we opted to create synthetic control groups for evaluation. The rationale for using synthetic control groups is twofold:

1. During the selection of the treatment geos, we chose the smallest geographical combinations that represent the nation as a whole, instead of selecting experimental and control groups based on factors such as size and value. This necessitates the use of synthetic control groups for subsequent evaluations.
2. Utilizing synthetic control groups does not interfere with states that are not part of the experiment, allowing for multiple experiments or strategies to be implemented in those states.

In the second phase, we employed Lasso Cross-Validation to determine the best alpha value. This method allows Lasso to automatically filter out states with features set to zero, ensuring that our control group does not suffer from overfitting. As a result, the states forming the synthetic control included “Florida”, “North Carolina”, “Ohio”, “Oregon”, “Pennsylvania”, and “Texas”.

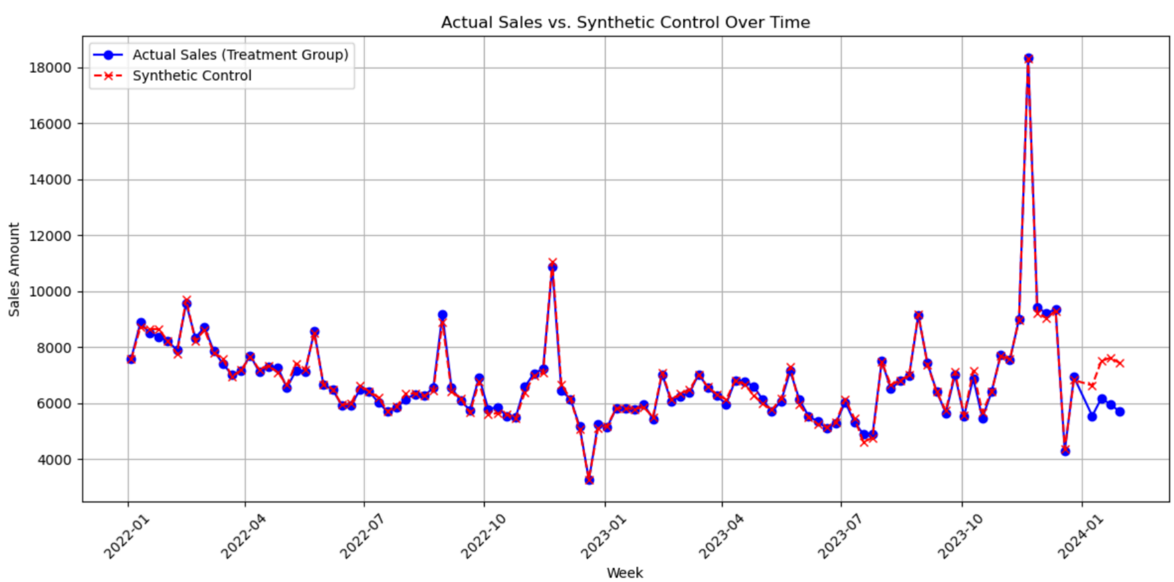
Considering we have a total of 42 states in our control group and only 140 weeks of data, domain knowledge suggests it is best to keep the number of states in the control group below

14. The Lasso Cross-Validation automatically selected 6 states, which fit well within the desired standards.

Next, we used a Linear Regression model to generate data for the synthetic control group, specifically in two steps:

1. The first step involved using the pre-treatment period data. We summed the order numbers from the four treatment geos and six selected control geos on a weekly basis to serve as Y and X, respectively, in training our model.
2. The second step entailed using X data from both the pre-treatment and post-treatment periods for predictions, thus avoiding data leakage.

Through this process, the post-treatment data from the control group served as our counterfactual results. This allowed us to compare with the treatment groups to determine the causal effect of the campaign.To ensure the scientific integrity of constructing the synthetic control group, we also conducted a double T-test comparing the treatment and control groups post-treatment. The obtained p-value was 0.002, indicating a statistically significant difference and confirming a distinct causal effect.



By comparing the actual sales data from the treatment period with the synthetic control's predicted sales, we can estimate the average causal impact of the Google Performance Max marketing campaign for all treatment geos during the treatment period is -1349.73 per week (Since this is a holdout experiment, it is logical for the causal effect to be negative). This analysis not only quantifies the campaign's effectiveness but also provides insights into how ACME Manufacturing can optimize its marketing efforts.

Population Causal Effect

However, recall that till this point the treatment is applied to the 4 selected regions, while we aim to understand how it influences national sales overall in order to decide whether to adopt Google Performance Max or not. To infer the national effect from the treatment effect, we fit a linear regression model where the response variable (Y_national) represents the total national sales during the pre-treatment period. The predictor variable (X_treatment) is the sum of sales in the treatment geos during the same period

```
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const      -5464.6932   1532.389     -3.566      0.001   -8504.179   -2425.207
x1            22.9402      0.218    105.047      0.000      22.507      23.373
```

Fitting an OLS model, we obtain estimates of the coefficients representing the intercept and slope of the relationship between sales in treatment geos and national sales. By applying this fitted model to the calculated treatment causal effects of the 4 treatment periods, we can extend our understanding of how the treatment might impact sales at the national level – if Google Performance Max were not implemented nationwide in 4 treatment weeks, the estimates for our change (loss) in sales will be -$30,403.0173362, -$36,034.60498656, -$43,747.05057861, -$45,177.10310293, with the tendency to loss more over time.

Recommendation

Assuming the cost of the Google Performance Max campaign is significant but manageable, we need to consider whether it justifies the potential reduction in losses which have been previously estimated at around $155,361.77 over four weeks. This comparison is crucial in determining the viability of the campaign.

If the cost of running the Google Performance Max campaign for four weeks is assumed to be less than or equal to the potential losses of $155,361.77 in a month, ACME should consider implementing the campaign. The reason behind this is straightforward: if the campaign costs less than the losses it prevents, it results in a net positive impact on ACME's finances.

Leveraging our domain expertise, we can draw several assumptions and make recommendations accordingly. Our analysis indicates potential losses of $155,361.77 per month if the campaign is not implemented. Given that each customer spends an average of $400 annually, visiting ACME weekly, we estimate that the campaign could attract approximately 2975 new customers. Based on the data, the Google Performance Max campaign exhibits an average Cost-Per-Click (CPC) of $0.6, a Conversion Rate (CR) of 2.04%, and a Customer Acquisition Cost (CAC) of $29.4. Calculating the combined costs of total click cost and total customer acquisition cost yields $145,559.1, which is substantially lower than the potential losses. Therefore, under these assumptions, we recommend that ACME engage in the Google Performance Max campaign to mitigate these losses and potentially increase profitability.

Furthermore, the campaign not only has the potential to mitigate these losses but also enhances customer engagement and increases sales through optimized ad placements across Google's advertising networks. This could lead to additional indirect benefits, such as improved brand recognition and customer loyalty, which are not immediately quantifiable but are valuable over the long term.