QMB: 6304
Midterm Project: The Advertising Campaign Analysis

September 27, 2023

Background: A prominent travel agency has conducted an advertising campaign. Two datasets - 'Abandoned.csv' (ABD) and 'Reservation.csv' (RES) - provide insights into customer interactions and outcomes from this campaign.
Objective: Determine the statistical success of a retargeting campaign by matching and analyzing the datasets.

Introduction

'Abandoned.csv' contains data about customers who engaged but didn't purchase a vacation package. Notice the potential missing data and duplicates. These customers were divided into test and control groups for a retargeting campaign. 'Reservation.csv' documents customers who eventually purchased vacation packages.
Task: Establish whether the retargeting campaign was statistically effective.

1 Business Justification

1. Explain why retargeting customers who initially didn't buy a package makes business sense.
2. Analyze the test/control division. Does it seem well-executed?
3. Compute summary statistics for the test variable, segmenting by available State data.

2 Data Alignment

4. From your examination of both files, propose potential data keys to match customers.
5. Detail your procedure to identify customers in:
    - Treatment group who purchased.
    - Treatment group who didn't purchase.
    - Control group who purchased.
    - Control group who didn't purchase.
6. Are there unmatchable records? If yes, provide examples and exclude them from the analysis.
7. Provide a cross-tabulation of outcomes for treatment and control groups.
8. Replicate the cross-tabulation for five randomly chosen states, detailing your selections.

3 Data Refinement

9. Generate a cleaned dataset with columns: Customer ID — Test Group — Outcome — State Available — Email Available. Each row should correspond to a matched customer from the datasets. *(Ensure you attach this cleaned dataset upon submission.)*

4 Statistical Assessment

10. Execute a linear regression for the formula: Outcome = α + β * Test Group + error. Share the results.
11. Justify that this regression is statistically comparable to an ANOVA/t-test.
12. Debate the appropriateness of the regression model in making causal claims about the retargeting campaign's efficacy.
13. Integrate State and Email dummies into the regression. Also consider interactions with the treatment group. Compare these results to the previous regression and provide insights.

5 Reflections

14. Reflect on the project:
    - Would you modify the experiment design if given a chance?
    - Could alternative paths be taken with better-quality data?
    - Are there actionable business implications from this analysis?
15. Self-assessment: Rate your effort (0-100) and anticipated performance. Elaborate if needed, mentioning any collaborations.

Note: This is an individual project. While discussions with peers are encouraged, mention all individuals you collaborated with. Ensure your final submission includes your Rscript, analyses, and the refined dataset.