D214: Data Analytics Graduate Capstone Task 3

Justin Huynh

Student ID: 012229514

M.S. Data Analytics

A. Executive Summary and Implications

The primary objective of this study was to analyze and predict customer conversion rates in digital marketing campaigns from our chosen dataset, focusing on the impact of different marketing channels. The hypothesis was that certain channels would significantly influence conversion rates more than others.. Specifically, we tested how much each marketing channel (i.e., Social Media, PPC, Email) affected conversion rates.

- The null hypothesis (H_0): There is no significant difference in the conversion rates across different marketing channels.
- The alternate hypothesis (H_1): There is a significant difference in the conversion rates across different marketing channels.

Our goal was to develop a predictive model to identify which customers are more likely to convert, which helps the organization optimize marketing strategies.

We began the analysis by performing Exploratory Data Analysis (EDA) to examine the distributions of key variables such as ad spend, click-through rate, and conversion rate. We then conducted a One-Way ANOVA to test for significant differences in conversion rates across marketing channels. Since the ANOVA tests revealed no significant differences, we proceeded to build a logistic regression model to predict customer conversions based on features like ad spend and click-through rate. To address class imbalance, we applied SMOTE (Synthetic Minority Over-sampling Technique) and tuned the model using GridSearchCV to optimize regularization strength.

Here are the outline of the findings from our analysis:

- ANOVA Results: No statistically significant differences were found in conversion rates across different marketing channels, so we fail to reject the null hypothesis.

- Logistic Regression: After applying SMOTE and tuning the model, we achieved an accuracy of 62%, with balanced precision and recall for both conversions and non-conversions.

- The model's ability to predict both conversion outcomes (class 0 and class 1) was balanced, which indicates that customer conversions are driven by individual factors rather than group-level characteristics (i.e. age or channel).

The primary limitation of the logistic regression model is its assumption of linearity, which may not fully capture the complex, non-linear relationships inherent in customer conversion behavior. Also, while SMOTE addressed class imbalance, it introduced synthetic data that could potentially oversimplify the problem. More advanced models like Random Forest or Gradient Boosting may be better suited for capturing non-linear interactions, although they may be more complex in terms of interpretability.

Based on the findings, we recommend implementing the logistic regression model for predicting customer conversions. The model can be used to prioritize marketing efforts toward customers with a higher likelihood of converting based on individual characteristics such as ad spend and engagement metrics. This data-driven approach should replace broad assumptions about specific marketing channels.

One of the expected benefits of the study include optimization of marketing spend. By targeting customers who are more likely to convert, the organization can improve conversion rates while reducing unnecessary ad spend. Another benefit is

balanced performance. With balanced precision and recall for both conversions and non-conversions, the model ensures fair targeting without overemphasizing one class over the other. The last benefit can be an expected ROI on the organization's ad spend. Even with a moderate accuracy of 62%, the model can lead to a more efficient allocation of resources and potentially increase the overall conversion rate by 5-10%, depending on implementation.

B. Presentation: Organization and Professionalism

The presentation will be provided in the Panopto video submitted with this task.

B1. Presentation: Content

The PowerPoint presentation will be labeled 'WGU_D214_Task3_Justin_huynh'.

C: Sources

1. "Digital Marketing Analytics Dataset" Retrieved from

   https://www.kaggle.com/datasets/arpit2712/digital-marketing-company