Impact of Offering Tech Support and/or a Discount Alongside Software Sales

1. Summary

- A software company, having run two sales campaigns, offering tech support and a discount, is interested in learning the impact of both sales initiatives on their revenue and is seeking strategic guidance.
- The incremental impact of offering tech support and/or a discount was determined by uplift modeling.
- The sales campaigns brought in $14 M in revenue, but optimizing which customers receive tech support and a discount can bring in an additional $ 10 M.
- Both treatments, offering tech support and a discount, increase expected revenue for almost all customers with tech support increasing expected revenue by $7000 per customer and discount increasing expected revenue by $5700 per customer.
- Not all customers are equal. The sales team should focus on larger companies with an annual IT spend greater than $37946 per year or annual revenue of $155667.
- Offering tech support and a discount does increase the chance of closing customers. The sales team can offer greater discounts if needed.

2. Introduction

A. Problem Statement

A software company wants to increase their revenue by offering tech support and/or a discount for their products. Having sold software with tech support and a discount, the company now wishes to evaluate the success or failure of both sales initiatives and to use the lessons to guide the efforts of the sales team.

B. Goal

The goal of this project is to evaluate the impact of offering tech support and/or a discount alongside software sales on the company's yearly revenue and to then provide strategic guidance for the company's sales team.

3. Dataset

The data is simulated and comes from Kaggle. The data contains 2000 customer data including columns for the yearly revenue attained from the customer, whether the customer received tech support, whether the customer received a discount, and 8 descriptors. Of those 8 descriptors are 4 binary flags indicating whether the customer has global offices, is a major corporation, is a small/medium corporation, or is a commercial business. The remaining 4 descriptors are continuous and include the customer's annual IT spend, employee count, PC count, and size given by their yearly revenue. The data was stratified by tech support and discount with approximately 500 customers receiving neither, 500 receiving tech support, 500 receiving a discount, and 500 receiving both tech support and a discount.

4. Data Preparation

A. Continuous Descriptors are Exponentially Distributed

Preparing the data for analysis and modeling involved transforming IT spend, employee count, PC count, and size by their occurrence. All four continuous descriptors are exponentially distributed (see Figure 1). To facilitate fitting, each was transformed according to their probability. However, to avoid any data leakage, each distribution was generated using a training set and then used to transform the test set.

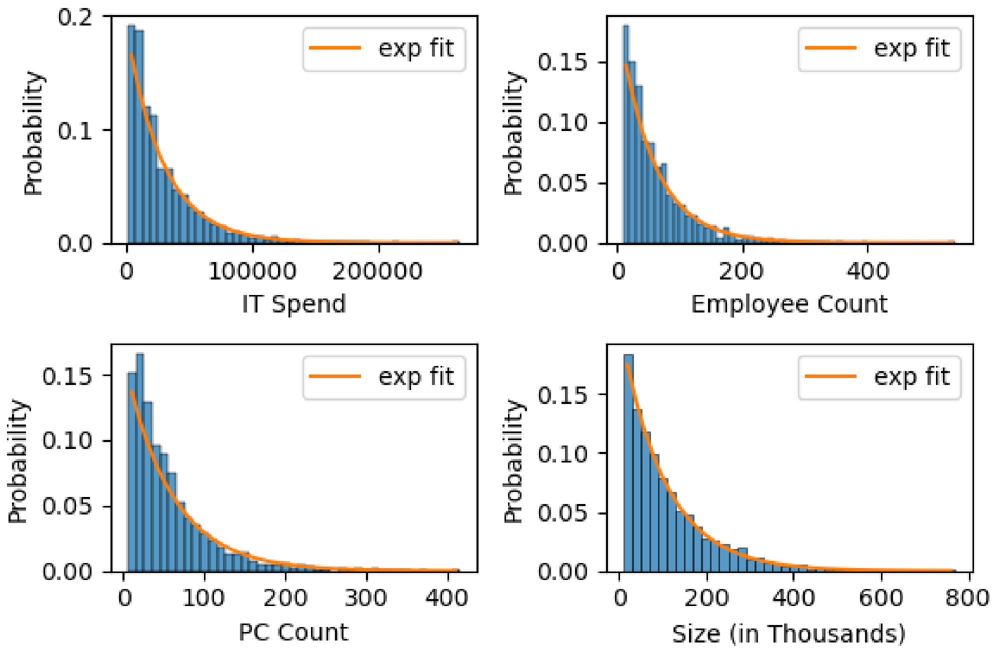

Figure 1: The four continuous customer descriptors, IT Spend, Employee Count, PC Count, and Size determined by annual revenue, are all exponentially distributed with little to no outliers. Preprocessing of these features will involve transforming according to their respective exponential distributions, effectively linearizing the data.

5. Exploratory Data Analysis

A. Majority of Revenue Comes from Larger Customers

With IT Spend, Employee Count, PC Count, and Size all being exponentially distributed, it is important to identify any correlations with Revenue. By plotting each continuous descriptor against Revenue (shown in Figure 2), it is apparent that both IT Spend and Size are positively correlated with Revenue, while PC Count and Employee Count are weakly correlated with Revenue. With IT Spend and Size being exponentially distributed, the company's annual revenue for all accounts may be driven by just a few customers.

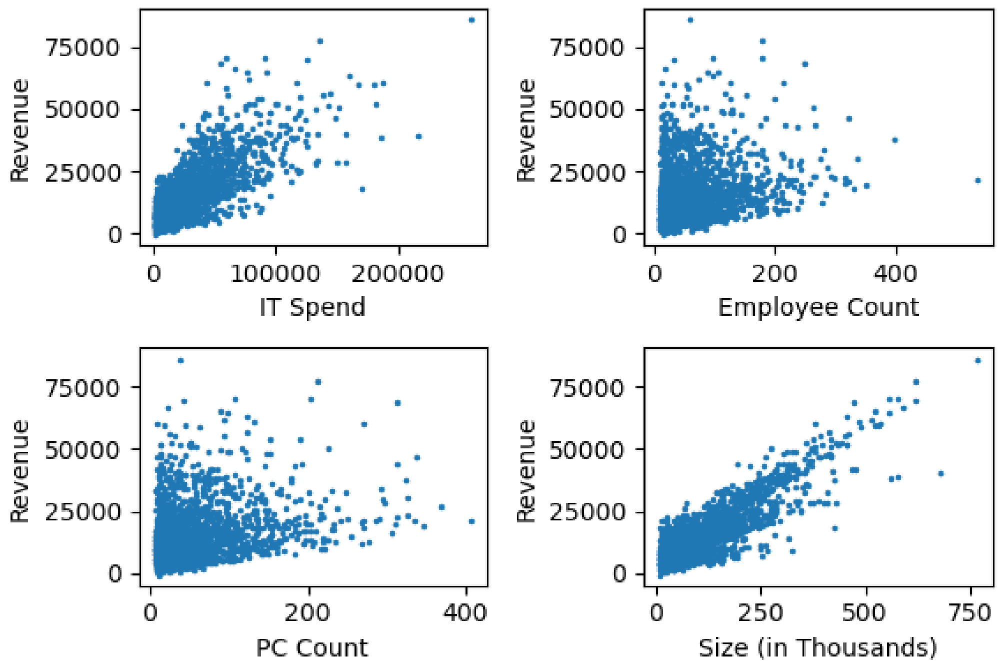

Figure 2: IT Spend, Employee Count, PC Count, and Size plotted against Revenue for each customer. Revenue positively correlates with both IT Spend and Size and minimally correlates with Employee Count and PC Count. Since Revenue correlates with IT Spend and Size, which are exponentially distributed, removing customers with extremely large IT Spend or Size may be detrimental to fitting.

Figure 3 shows that, indeed, Revenue is driven by a few customers. The top 25% of customers contribute 50% of the company's annual revenue while the bottom 25% contribute only 8% of the company's revenue. At the same time, the average top 25% is also in the 88th percentile of IT Spend with $59312 spent on IT and 88th percentile of Size with an annual revenue of $236471. Thus, most of the company's revenue is driven by larger customers, and it may be possible to increase revenue by targeting such customers.

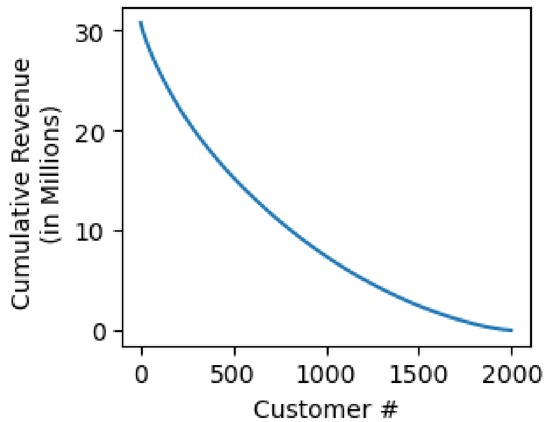

Figure 3: Cumulative revenue plotted against customer number where customers are ordered by decreasing revenue. A minority of customers contribute to most of the revenue with the top 25% of the customers accounting for 50% of the revenue and, conversely, the bottom 25% accounting for only 8% of the revenue.

6. Uplift Modeling

The set of customers is highly diverse with there being 4 binary descriptors and 4 exponentially distributed continuous descriptors. Machine learning methods are best suited for connecting expected revenue to customer profile. At the same time, the goal of the project is to quantify the effect of two sales campaigns, offering tech support and offering a discount, which makes uplift modeling – a set of machine learning and causal inference techniques used to model the incremental impact of an action on a customer's outcome – the most appropriate framework for addressing the company's inquiry.

The impact of offering tech support and a discount on revenue were approximated using two approaches: the S-Learner approach which uses a single machine learning model, and the T-Learner approach which uses multiple machine learning models, one for each treatment variable. The S-Learner can sometimes underestimate the treatment effect while the T-Learner can overestimate the treatment effect. By analyzing the results of the two approaches, we can get a better sense of the impact of offering tech support and a discount.

Prior to training, the 4 continuous descriptors were transformed according to their respective distributions. However, to avoid data leakage, distributions of IT Spend, Employee Count, PC Count, and Size were generated from the training set and then used to transform the test set. With the feature space now containing only binary and linearly distributed features, only linear regression and random forest models were tested.

A. S-Learner

a. Linear Regressor

Training was performed using scikit-learn's ElasticNet model. Unlike T-learner, treatment assignment is also passed into the machine learning model. After optimizing hyperparameters, the best linear regression model was with alpha=0.1 and l1_ratio=0.9. Results are summarized below in Table 1.

b. Random Forest

Training was performed using scikit-learn's RandomForestRegressor model. Again, treatment assignment was passed into the machine learning model. After optimizing hyperparameters, the best random forest model was with n_estimators=125, criterion='absolute_error', max_samples=0.5, max_features=0.7. Results are summarized below in Table 1. Between linear regression and random forest regression, the S-learner using random forest regression generated less errors. S-learner with linear regression likely contained too much bias.

B. T-Learner

a. Linear Regressor

With two binary treatments, there are four treatment outcomes with those being no tech support and no discount, tech support with no discount, no tech support with discount, and tech support and discount. Under the T-learner approach, four machine learning models were trained for each treatment outcome. For all treatment outcomes, ElasticNet(alpha=0.025, l1_ratio-0.975) resulted in the least errors. The results are summarized in Table 1 below.

b. Random Forest

Again, four machine learning models were trained, one for each treatment outcome. For the four treatment outcomes, no tech support and no discount, tech support with no discount, no tech support with discount, and tech support and discount, 175, 125, 150, and 175 estimators were used, respectively. Criterion='absolute_error', max_samples=0.5, max_features=0.5 were used for all treatment outcomes. The results are summarized below in Table 1. Comparing the T-learner with linear regression and with random forest regression, linear regression performed better with less absolute percent error and mean percent error. Random forest regression likely contained too much variance.

Comparing S-learner with random forest regression and T-learner with linear regression, the T-learner approach was more accurate with less absolute percent error, mean percent error, and median percent error.

		Mean Abs % Error	Mean % Error	Median % Error
S-Learner	**Linear Regressor**	21.10 (32.48)	-2.38 (38.66)	1.14
	Random Forest	13.58 (26.88)	4.26 (29.81)	0.37
T-Learner	**Linear Regressor**	4.8 (7)	1.68 (14.09)	0
	Random Forest	11.7 (18.8)	3.3 (22.1)	0

Table 1: Mean and standard deviation (in parentheses) of absolute percent error and percent error of linear regression and random forest models under both the S-learner and T-learner approaches. T-learner with linear regression minimizes the mean absolute percent error, mean percent error, and median percent error.

C. Cumulative Gain

With machine learning models, both S-learner with random forest regression T-learner with linear regression, in hand, it is possible to approximate counterfactuals and to quantify the expected revenue should a customer receive tech support or not or whether a customer should be offered a discount or not. This information is plotted as cumulative gain where gain for each customer is the difference in expected revenue with treatment and without treatment.

Although the T-learner with linear regression performed better than the S-learner with random forest regression in predicting the revenue (see Table 1), both approaches generate qualitative similar cumulative gain plots (Figure 4), and thus qualitatively similar results. Both approaches suggest that all customers be offered tech support with an average gain of $7000 per customer. Both approaches also suggest that 98% of customers be offered discount with an average gain of $5700 per customer. Furthermore, a discount can be offered to only half the customers and still net over 80% of the expected gains.

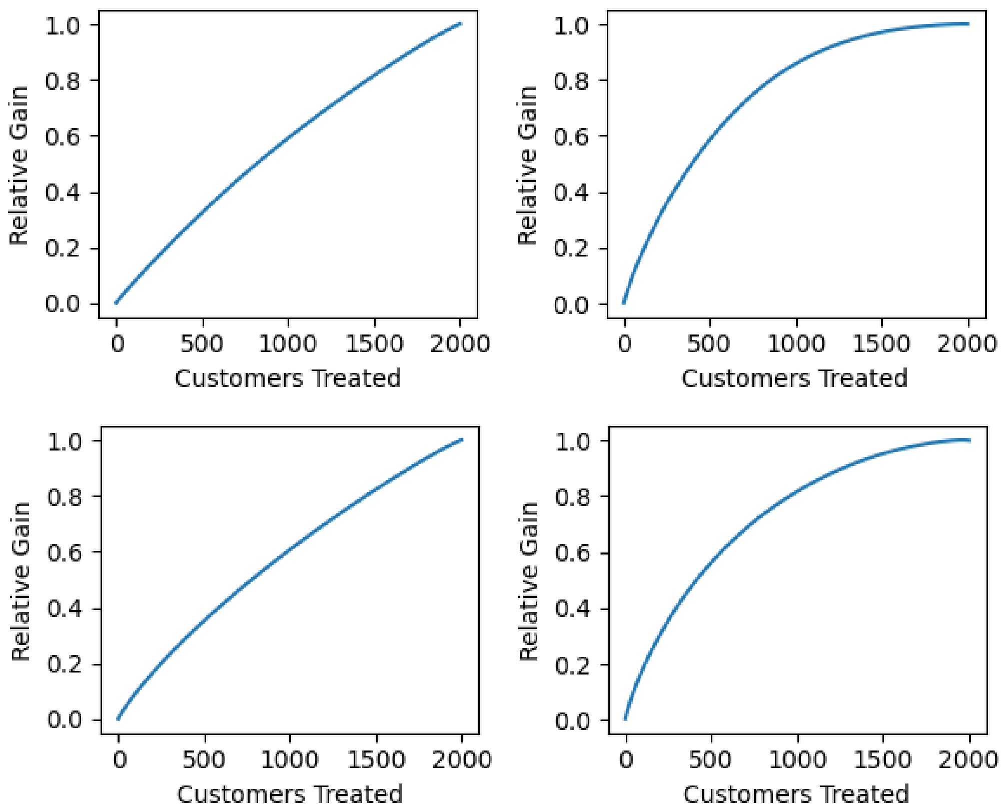

Figure 4: Cumulative gain for S-learner with random forest regression (top) and T-learner with linear regression (bottom) and for offering tech support (left) and discount (right). Both the S-learner and T-learner predict qualitatively similar gains for both tech support and discount with the T-learner predicting slightly higher total gain. For tech support, the S-learner predicts a total gain of $13.8 M while the T-learner predicts $14.5 M. For discount, the S-learner predicts a total gain of $10.3 M while the T-learner predicts $11.5 M. Both the S-learner and T-learner suggest that all customers be offered tech support with an average gain of $7000 per customer. Both the S-learner and T-learner also suggest that over 98% of customers be offered a discount with an average gain of $5700 per

customer. Both the S-learner and T-learner also indicate that offering a discount to only half the customers will net over 80% of the expected gains.

7. Conclusion

A software company, having run two sales campaigns, offering tech support and a discount, is interested in the incremental impact of each campaign on their revenue. With 2000 customer data including binary descriptors such as whether the customer is a global company, a major company, a small/medium company, or operates in the commercial sector and continuous descriptors including the customer's IT spend, employee count, PC count, and size, the incremental impact of each sales treatment (tech support and discount) was best approximated using uplift modeling.

Both S-learner and T-learners were trained. While the T-learner with linear regressors was better at learning the data, both the S-learner and T-learner generated similar results. Both learners suggest that tech support be offered to all customers with an expected gain of $7000 per customer and both learners also support offering a discount to 98% of customers with an expected gain of $5700 per customer. At the same time, while gain is uniformly distributed across customers for tech support, gain is not uniformly distributed across customers for discount. In fact, over 80% of the expected gains from offering a discount could be acquired by offering a discount to only half the customers.

Analysis of customer profiles also indicate that only a few customers contribute to most of the company's revenue with the top 25% of customers contributing half of the company's revenue while the bottom 25% of customers contribute to only 8% of the company's revenue. The top 25% of customers are also larger customers with, on average, in the 88th percentile for IT spend and size. To increase the company's overall revenue, it is suggested that the company target larger companies with an annual IT spend greater than $37946 per year or an annual revenue of $155667. Offering tech support and a discount does increase the chance of closing customers and the sales team should feel free to offer both to close larger customers. Data also suggests that greater discounts may be offered if needed.