

Assignment 1 (IDS 572)

Due: Oct 4th

Case Study: Targeting Customers for TravelPlus's RoadsidePlus Add-On

Introduction

On a chilly November morning, Joi Megenti, the Vice President of Marketing Analytics at TravelPlus Insurance, flipped through the latest campaign performance report for the company's newest product, **RoadsidePlus**. The add-on offered expanded roadside assistance benefits to existing travel insurance customers.

The numbers bothered her. Despite months of effort, the overall response rate for campaigns promoting RoadsidePlus hovered around 8 percent. While better than the industry's average for add-on insurance campaigns, it still meant that 92 percent of outreach efforts yielded nothing. TravelPlus was spending heavily on emails, agent time, and call center follow-ups — with much of it wasted.

Joi's mandate was clear: find a smarter way to target customers, one that would deliver more sales with less marketing spend.

TravelPlus and RoadsidePlus

TravelPlus is one of the top five providers of travel-related insurance products in North America. Its customer base was large, stable, and diverse, ranging from families booking vacations to retirees investing in long-term travel coverage.

In early 2023, the company launched RoadsidePlus, marketed as a "peace-of-mind add-on" offering enhanced roadside assistance for accidents and breakdowns. Management believed RoadsidePlus had strong potential, particularly for customers who already trusted TravelPlus with core insurance needs. But add-on products were notoriously difficult to sell. Unlike core policies, which customers often actively sought out, add-ons required **push marketing**: proactive emails, outbound calls, and agent cross-sell at renewal.

The Campaign Design

Joi's team had designed the RoadsidePlus campaigns using a **test-and-learn approach** familiar in direct marketing:

- Each month, TravelPlus identified an **eligible pool** of existing customers (those with active policies and no suppression rules such as recent complaints or regulatory restrictions).
- From this pool, the majority were **contacted** via solicitation campaigns — mostly email, supported by follow-up calls from agents.
- A **10 percent random holdout group** was deliberately excluded from contact.

The purpose of the holdout was simple but critical: it provided a **baseline purchase rate** for customers who were not solicited. This allowed Joi's team to separate the effect of marketing from purchases that would have occurred anyway.

To her surprise, the holdout group was not entirely inert. Some holdout customers did purchase RoadsidePlus — usually when they visited the TravelPlus website, renewed policies with agents, or encountered general advertising. Their **natural response rate** was low (around 2 percent), but non-zero. This was the baseline against which the 8 percent contacted response rate had to be judged.

The Data

The campaigns generated a rich dataset. For each customer, Joi's team had access to:

- **Demographics:** age, income, marital status, homeownership, household size, region.
- **Customer relationship variables:** tenure with TravelPlus, acquisition channel, type of existing policy.
- **Engagement behavior:** website visits, promotional email opens and clicks, online purchases.
- **Risk and claims history:** claims filed, incidents in the last three years, history of late payments, prior add-on purchases.
- **Spending patterns:** external credit bureau data estimating spend in categories such as travel, electronics, sports, and healthcare.
- **Composite scores:** internally engineered indices summarizing engagement, overall customer value, risk, and upsell potential.
- **Lifestyle data:** enrichment purchased from a marketing data provider, including streaming hours, pet ownership, commute distance, dining-out frequency, and app downloads.

Some of the lifestyle variables clearly tied to insurance purchasing. Others were more speculative. As Joi reflected, "Marketing insisted that lifestyle signals like dining-out frequency might matter. My gut tells me they're mostly noise."

TravelPlus's analytics team regularly builds composite indices to summarize multiple signals into a single measure. These scores are shared with marketing and product teams as part of standard reporting.

- **Score_Engagement** – A weighted index combining website visits, email opens, email clicks, and app activity. Higher values reflect more frequent customer engagement with TravelPlus touchpoints.
- **Score_Value** – A summary measure of financial value to the company, incorporating policy premium levels, tenure, cross-sell purchases, and customer profitability.
- **Score_Risk** – An index based on past claims, incident history, and late payments, designed to flag higher-risk customers.
- **Score_Upsell** – A proprietary score developed by marketing analysts to estimate a customer's **likelihood of purchasing add-on products** such as RoadsidePlus. It uses prior add-on purchases, engagement patterns, and demographic indicators.

Composite scores are common in practice. They can be useful shortcuts, but they can also introduce redundancy or noise if they are poorly constructed. Joi wondered whether these scores add real predictive value beyond the raw features they summarize: "Should we trust the scores marketing already built, or is our model finding stronger signals elsewhere?"

The **target variable Purchase** was defined as:

- **Yes** if a contacted customer purchased RoadsidePlus within 30 days of solicitation.
- **No** if they did not.

Holdout customers were not used for modeling but were scored alongside contacted customers to measure **incremental lift**: the difference in purchase rates between contacted and holdout customers within each predicted propensity decile.

The Analytics Challenge

Joi now faced her analytics team. "We've got everything here," she said, gesturing to the dataset projected on the screen. "We know who we contacted, who we held out, and who purchased. The raw response rate is 8 percent for contacted customers, 2 percent for holdout. The question is: how do we go from this to a smart, scalable targeting strategy?"

The challenge has three parts:

1. **Exploration.** Which customer attributes are truly predictive of RoadsidePlus adoption? Are demographics meaningful, or are behavioral and engagement variables stronger? Do the lifestyle data add value or act as noise?
2. **Modeling.** Can a data-driven model outperform the marketing team's blunt rules? Which methods — logistic regression, k-NN, decision trees, random forests — offer the best mix of interpretability and accuracy?
3. **Business evaluation.** Beyond regular metrics (accuracy, precision, recall, AUC), what matters is evaluation based on how a model would actually guide marketing decisions. A key metric is the *lift calculation by decile* (or by proportion of customers targeted), which shows how concentrated responses are among the highest-scoring customers. From a business perspective, lift helps evaluate how effectively a model can be implemented in practice, since campaigns often target only a fixed percentage of customers.

The next layer in evaluation is *incremental lift*: will solicitation actually change outcomes compared to what would have happened naturally? This is where the holdout group provides a useful baseline. If the top decile of contacted customers respond at 24% while the same decile of the holdout responds at 4%, then marketing's incremental effect in that decile is 20 percentage points.

Decision Point

Joi leaned back. "Here's the trade-off. Contacting more customers increases reach but wastes spend. Contacting fewer customers raises precision but risks missing buyers. What I need is not just a model, but a **business recommendation**: If we only have budget to target the top X percent of customers, which segment should it be? And why?"

Assignment Questions

1. **Exploratory Data Analysis**
 - (a) What is the overall purchase rate?
 - (b) Explore predictor relationships with Purchase. How do you examine this for numeric vs categorical variables? Which variables seem strongest? Which look like noise?
For numeric variables: compare distributions for purchasers vs. non-purchasers (e.g., boxplots, density plots, group means).
For categorical variables: create cross-tabulations and purchase rates by category; test differences with chi-square.
 - (c) Is Joi's gut feel on lifestyle variables correct?
 - (d) Examine Composite Scores
 - (i) Are they predictive of Purchase?
[For each composite score, create a simple plot (e.g., boxplot) of the score by Purchase. Do purchasers have higher median or average scores than non-purchasers?]
 - (ii) Assess whether these composite scores provide *incremental value beyond their underlying components* (e.g., does Score_Engagement add much beyond email/web activity measures? Does Score_Upsell add signal beyond prior add-ons and engagement?).
Based on your findings, which scores seem useful for prediction, and which may be largely redundant or noisy?

2. Data preparation

You may need to standardize numeric predictors as needed for models that require it – explain what you do, why. Encode categorical variables appropriately – explain what you do and why. Partition the data into training and test sets.

3. Model development and evaluation

(a) Develop logistic regression, k-NN, decision tree (CART, C5.0), random forest and gbm models.

(b) Evaluate each model using confusion matrix and related metrics, and AUC.

(c) Which variables play an important role in the predictions from different models? Tale your best model for each method, and compare and discuss variable importance.

For each method, clearly explain how you investigate the effect of different parameters and determine the ‘best’ values. Consider performance, overfit.

Describe how you handle imbalanced data for the different models.

4. Lift & Incremental Value

(a) TravelPlus needs to understand how these models can guide real marketing decisions.

(i) Explain why lift is a useful way to evaluate predictive models in such marketing contexts.

(ii) Produce lift tables by decile (for the contacted group), for the best models in each method (from step 3).

Compare performance - which models most effectively concentrate purchasers in the top-ranked segments?

(b) Suppose TravelPlus can afford to contact only the top 10%, 20%, or 30% of customers ranked by your model.

Evaluate the trade-offs between these decile levels. Which decile cutoffs appear most attractive, and why?

How would you decide which decile cutoff is best? What factors would influence your recommendation (e.g., marketing costs, incremental revenues, diminishing returns as more customers are targeted)?

(c) Incorporating Incremental Lift

The dataset you have includes only contacted customers, and does not include the holdout set which was not contacted. So, you cannot compute incremental lift precisely (i.e., the additional response caused by marketing vs. customers who would have purchased anyway).

Assume you also have a holdout/control group (customers not contacted).

How would you measure the incremental value of your model by targeting depth (e.g., by decile of predicted score)? Clearly describe the steps for this.

5. Recommendation

Now take the role of Joi, VP of Marketing Analytics. Justify your recommendations with both **statistical evidence** and **business reasoning** (ROI, budget, scalability, strategic priorities). How would you communicate these findings to the CMO in a way that balances technical rigor with business clarity?

(a) Make a clear recommendation: *Should TravelPlus target the top 10%, 20%, or 30% of customers?*

(b) Which model would you choose? Give your rationale for this.

Variables

Category	Variable	Description
Target Variable	Purchase	Binary outcome: whether the customer purchased the RoadsidePlus add-on within 30 days of being solicited (Yes/No).
Demographics	Age	Customer's age in years.
	Income	Estimated annual household income (USD).
	Region	Geographic region (North, South, East, West).
	MaritalStatus	Marital status (Single, Married, Divorced, Widowed).
	HomeOwner	Indicator of whether the customer owns their home (Yes/No).
	HasKids	Indicator of whether there are children in the household (Yes/No).
Customer Relationship	TenureMonths	Number of months the customer has held an active TravelPlus policy.
	Channel	Acquisition channel when the customer joined (Online, Agent, Branch).
	PolicyType	Current policy type: Basic, Standard, or Plus.
Engagement Behavior	WebVisits	Number of visits to the TravelPlus website in the last 6 months.
	OnlinePurchases	Number of online transactions with TravelPlus (excluding RoadsidePlus).
	EmailOpens	Number of promotional emails opened in the last 6 months.
	EmailClicks	Number of promotional email links clicked in the last 6 months.
Risk & History	ClaimsPastYear	Number of insurance claims filed in the past 12 months.
	Incidents3Y	Number of incidents (accidents, losses, etc.) reported in the past 3 years.
	PriorAddons	Number of other add-on products the customer has purchased in the past.
	LatePayments	Indicator of whether the customer has a history of late premium payments (Yes/No).
Spending Patterns	Spend_Travel	Annual spending on travel-related purchases (e.g., flights, hotels).
	Spend_Sports	Annual spending on sports/outdoor goods and services.
	Spend_Electronics	Annual spending on consumer electronics.
	Spend_Fashion	Annual spending on fashion and luxury goods.
	Spend_Health	Annual spending on healthcare and wellness.
Composite Scores	Score_Engagement	Weighted score combining digital engagement (web visits, email opens/clicks).
	Score_Value	Weighted score summarizing overall financial value of the customer (premiums, tenure, add-ons).

Category	Variable	Description
Lifestyle Variables	Score_Risk	Weighted score summarizing claims, incidents, and payment history.
	Score_Upsell	Proprietary score estimating likelihood of purchasing add-on products.
	StreamingHours	Average weekly hours spent using streaming services.
	PetOwnership	Indicator of whether the household owns a pet (Yes/No).
	CommuteDistance	Estimated daily commuting distance in miles.
	DiningOutFreq	Average number of times the household dines out per month.
	AppDownloads	Number of mobile applications downloaded in the past year.
