
MegaTelco and the Churn Problem (Part 2)

The glass-walled conference room was quieter than usual. Only Dana Cho and Natalie Ferman remained. The analytics team had gone home after a long sprint to finish the churn prediction model, and now Dana had called Natalie back in for one more round of advice before next week's budget meeting with Finance.

"It took longer than I hoped," Dana said, "but we finally have the predictive model. Luis followed your instructions: each instance is a contract two months before expiration, and the model predicts whether the line will be cancelled within three months after that. Rachel Fu wasn't thrilled, though. She wanted to model how long customers keep paying when they don't renew. But, as you warned, our data can't handle that kind of survival analysis. Since we limited the training data to the period after the iPhone deal to keep it representative of today's competitive landscape, we just don't observe customers far enough out to estimate that well."

Natalie listened with quiet satisfaction. When they'd first met, Dana had barely followed half of what Luis said about features and instances. Now the marketing executive was explaining time windows and data limitations like a pro, bridging analytics and business as if she had been doing it for years.

"Good," Natalie said. "Just curious—did you check that most customers who switch actually do so within those three months? If that's our target, we should make sure it reflects real behavior, not just what the data or the KPIs make convenient."

"Luis checked. Roughly ninety percent of cancellations happen in that window. After that, most customers either renew on their own or keep paying month-to-month. That's why we felt three months was the right cutoff."

Natalie jotted a note, then looked up again. "And what about decline in usage or the auto-renew flag—did you manage to include those at decision time?"

Dana sighed. "We tried, but it was too messy. Usage data is aggregated by billing cycle, and the timestamps don't always line up with contract periods. In some cases, the last usage record appears weeks after a contract technically ends. Same with the contract table—lots of inconsistencies. So we went with a simpler setup: customer-level data captured two months before expiration for each line. It was cleaner and easier to process."

Natalie nodded, the word *leakage* flickering back in her mind. In their past meetings, she'd filled a whole page of her notes with that single term—warning the team that what's available at training time may not exist at decision time. Now, hearing Dana talk, she realized how far they'd come. They were finally past the stage of building "demo" models and ready to build systems that could actually be used.

“Makes sense to me,” Natalie said, “So, what’s the plan for deployment? Have you thought about how to turn the predictions into action?”

Dana hesitated, then gave a half-smile. “That’s exactly why I wanted to meet. The model’s done, but we still need to decide two things: how big the renewal-credit budget should be, and how to use the model to assign it. The offer we’re considering is a \$200 credit for customers who renew for one more year after their contract expires. Reaching out through the call center costs about \$5 per customer, even if they say no. And we’re estimating each customer’s value based on their average monthly charge from the last year of the contract. The problem is, I can’t convince Frank to approve a meaningful budget without a clear business case—and we can’t estimate the impact without first testing the offer. It’s... a bit circular.”

Natalie nodded. “The classic chicken-and-egg problem. You need a budget to run the test, but you need the test to justify the budget.”

“Exactly.” Dana turned her laptop toward her. A spreadsheet filled the screen—columns for customer ID, predicted churn probability, and average monthly charge. “Luis suggested ranking customers purely by churn probability. But Frank’s going to ask why we’d give \$200 to someone who only pays sixty a month. I think we should account for both churn risk and revenue—maybe multiply them?”

Natalie studied the sheet for a moment. “That’s a reasonable start. Think of it as expected loss: the chance they’ll leave times what they’re worth if they do. It’s not perfect—you still don’t know how much the offer will actually change their behavior—but it gives you a logical way to prioritize.”

“That’s what I was hoping you’d say. If we can frame it like that, maybe we can convince Frank to increase the budget. But to convince him, I’d need to make an argument for how much those losses can actually be reduced.”

Natalie smiled. “Not exactly. You don’t need to argue how much the offer will reduce losses—only how effective it would have to be to break even. That’s a much easier problem. You’re setting a threshold, not estimating profitability.”

Dana frowned slightly, then nodded as the logic clicked. “So instead of guessing the effect, I show how big it needs to be for the investment to make sense.”

“Yes,” Natalie said. “Frame it that way, and you’re showing what kind of outcome would justify the spend. That’s a conversation Frank will understand.”

Dana hesitated for a moment, then said, “Would you have time to put together that analysis? Just something simple so Luis can see the structure and take it from there for the Finance meeting. I’m just not sure I can translate all this on my own.”

“Of course. I’ll draft something. We should also invest part of the budget in data. If everyone at the top of the ranking gets the offer, we won’t be able to estimate the counterfactual. Let’s keep a control group and include a few lower-ranked customers to check if we’re missing persuadable ones.”

“So it’s like running an A/B test?”

“Essentially, yes,” Natalie said. “But it can be more than that. The data we collect could also help us improve targeting in the future by using it to train a new model that predicts how different customers respond to the offer. That way, we can predict not just who’s likely to churn, but who’s likely to be convinced to stay.”

Dana leaned back, almost speaking to herself. “So we can keep improving the system every time we deploy it. I hadn’t thought of it that way.”

“That’s the idea,” Natalie said with a smile. “Data science is iterative—we should always try to deploy with the door open for what may come next.”

Exhibit A – Data Used for the Churn Prediction Model

These attributes correspond to data two months before contract expiration:

- Gender: Whether the customer is a male or a female.
- SeniorCitizen: Whether the customer is a senior citizen or not.
- Partner: Whether the customer has a partner or not.
- Dependents: Whether the customer has dependents or not.
- tenure: Number of months the customer has been with the company.
- PhoneService: Whether the customer has a phone (landline) service or not.
- MultipleLines: Whether the customer has multiple lines or not.
- InternetService: Customer's internet service provider.
- OnlineSecurity: Whether the customer has online security or not.
- OnlineBackup: Whether the customer has online backup or not.
- DeviceProtection: Whether the customer has device protection or not.
- TechSupport: Whether the customer has premium tech support or not.
- StreamingTV: Whether the customer has streaming TV or not.
- StreamingMovies: Whether the customer has streaming movies or not.
- Contract: The contract term of the customer.
- PaperlessBilling: Whether the customer has paperless billing or not.
- PaymentMethod: The customer's payment method (electronic check, mailed check, Bank transfer (automatic), Credit card (automatic)).
- MonthlyCharges: Average monthly charge (last year).

Target variable:

- Churn: Did the customer cancel the postpaid line associated with this contract within three months after the contract ended?