CELL2CELL

Customer Connectivity

A Data-Driven Approach to Understanding Customer Churn

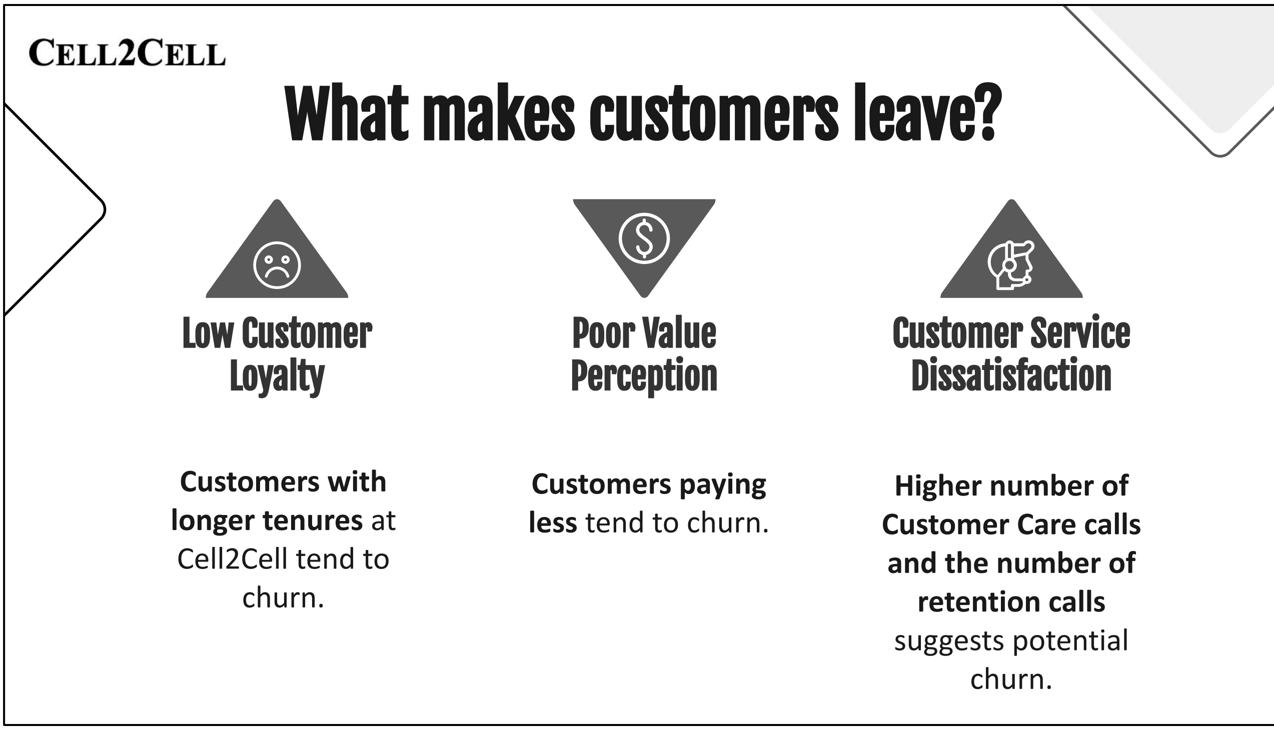
Cece Zhang, Kelly Hong, Clarissa Gunawan,
Yukta Butala, Sylvanne Braganza

Sylvanne Braganza (sbraganz)
Yukta Butala (ybutala)
Clarissa Gunawan (cgunawan)
Kelly Hong (sewonh)
Cece Zhang (xiyuzhan)

Our current approach is reactive in nature. When a customer wants to churn, we scramble to offer incentives for them to stay.

Our solution is to develop a proactive churn management program, which will be recommended to the Customer Base Management (CBM) Group. We will do this by analyzing customer data to identify key factors that lead to customer churn.

High-risk customers, identified through the churn model, will be targeted with personalized incentives such as discounts, upgrades, or loyalty rewards, demonstrating Cell2Cell's commitment to customer satisfaction and loyalty. By implementing a proactive retention program based on these insights, Cell2Cell aims to preemptively address customer concerns and mitigate churn, ultimately fostering stronger customer relationships and enhancing long-term profitability in the competitive wireless market.



1. **Low Customer Loyalty**

   Customers with longer tenure may indicate higher churn likelihood compared to those with a shorter tenure. Based on current customer data, customers who churn have an average tenure of 19.05 months, while those who did not have an average of 18.61.

   This could indicate changes in customers' needs or better alternatives from competitors.

2. **Poor Value Perception**

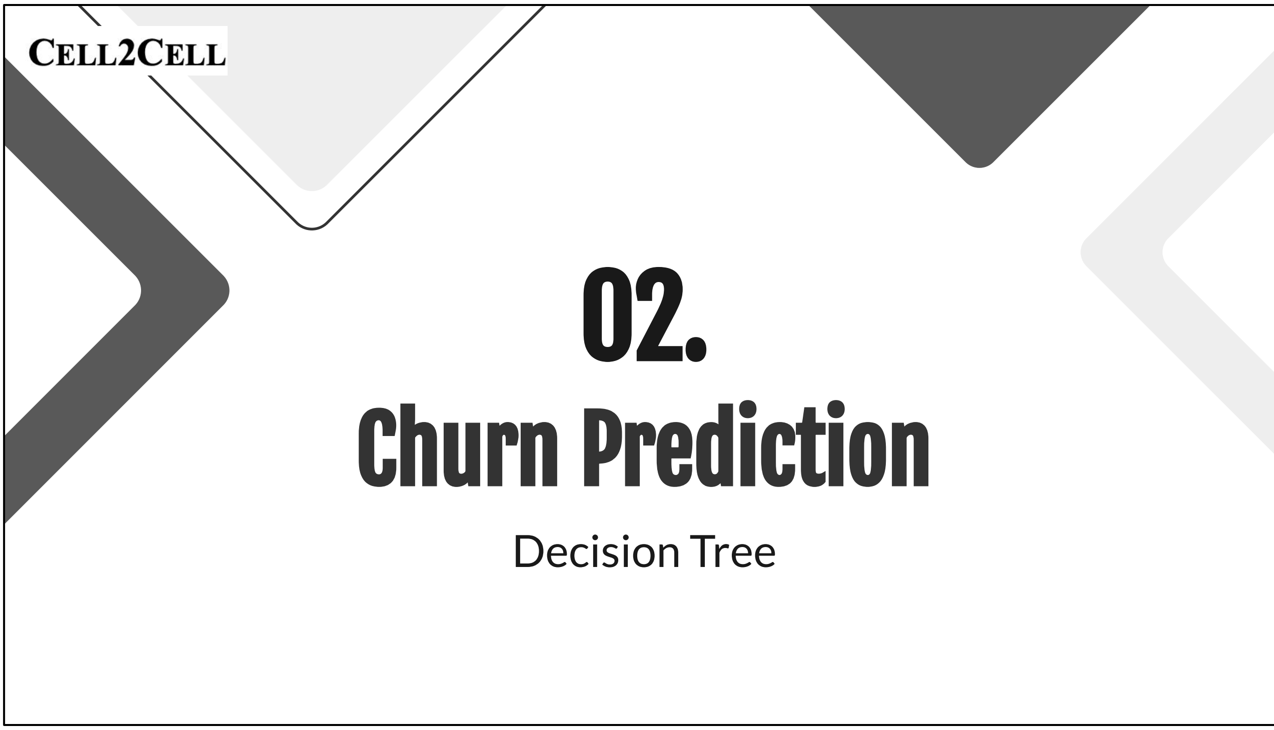
   Interestingly, a lower average monthly recurring charge may correlate with higher churn rates. Based on current customer data, customers who churn have an average recurring charge of $44.72, while those who did not have a mean of $47.82.

   This does not necessarily infer that we need to lower our price. Instead, we need to communicate our value and show customers how our offerings meet customers' needs.

3. **Service Dissatisfaction**

   A high number of customer care calls and previous calls made to the retention team may correlate with higher churn rates. According to current customer data, those who churns made an average of 0.06 customer service calls, while those who did not made an average of 0.03. Increased interaction with customer care may indicate dissatisfaction or unresolved issues, while a higher retention call could indicate that the customer already has expressed dissatisfaction or considered leaving.

This highlights the importance of a proactive intervention. We need to identify and address the root causes of customer concerns through personalized support.



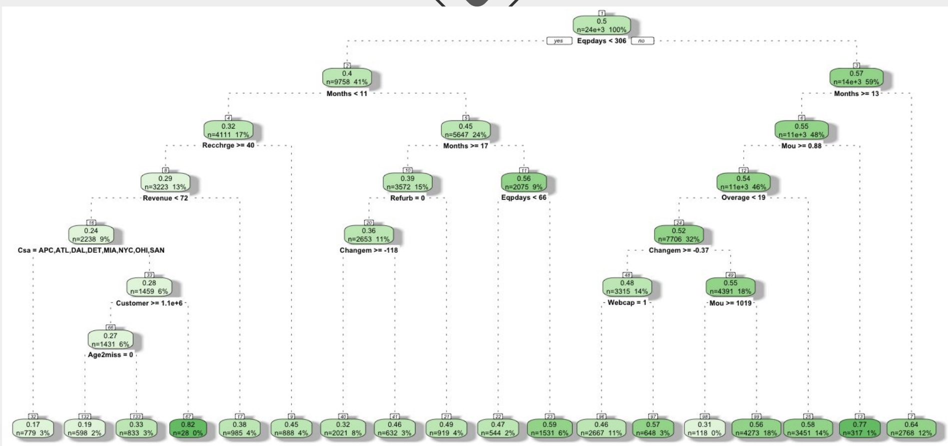
We first constructed a decision tree model to predict churn outcomes. We started by modeling a full decision tree; however to reduce the complexity and thus improve predictive accuracy by the reduction of overfitting, we decided to prune that tree.

Finding the optimal complexity parameter for pruning a decision tree helps strike a balance between model complexity and generalization. By controlling the pruning process, we can prevent overfitting and ensure the tree captures essential patterns in the data without being overly complex.

Plotting the cross-validation relative error against different complexity parameters (cp) allows us to visually identify the point where the relative error stabilizes or starts increasing, indicating the optimal complexity parameter. This helps us select the parameter that minimizes the error while preventing overfitting, optimizing the trade-off between model complexity and generalization performance.

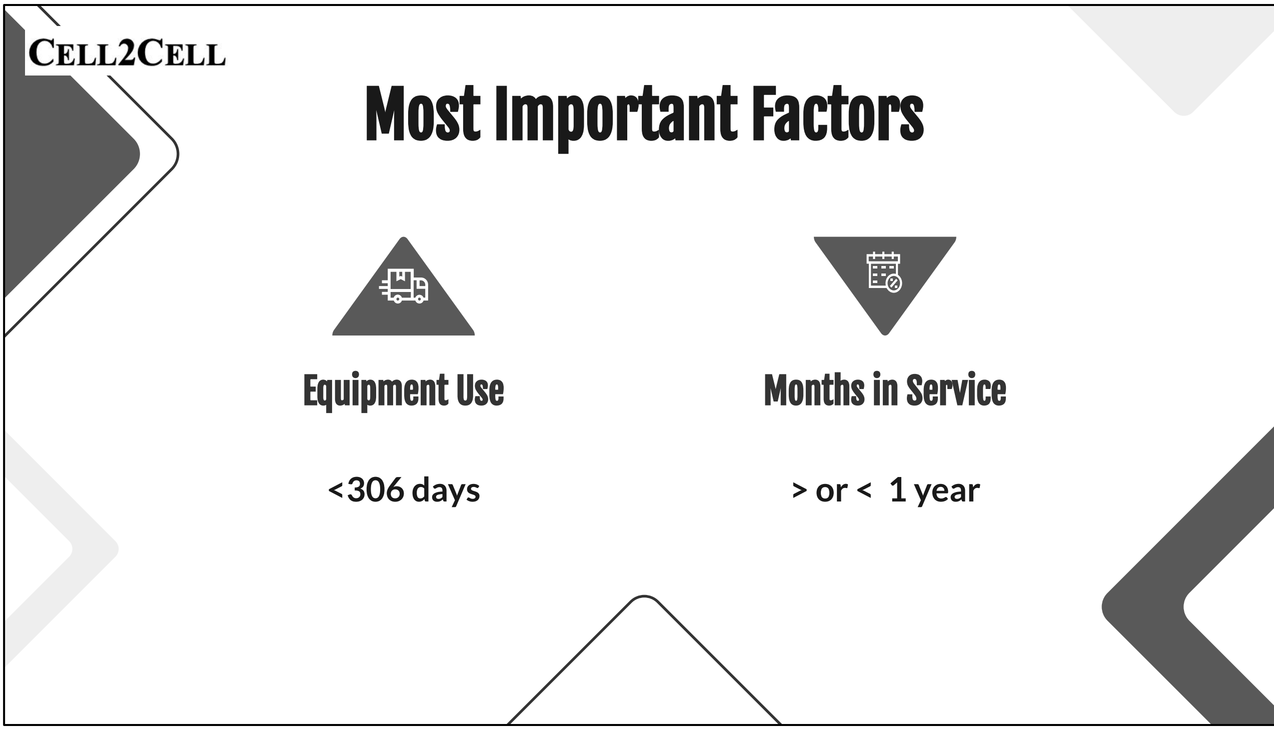
After plotting, we determined the optimal cp to be 0.00090890 and pruned the tree accordingly to arrive at this model.

CELL2CELL



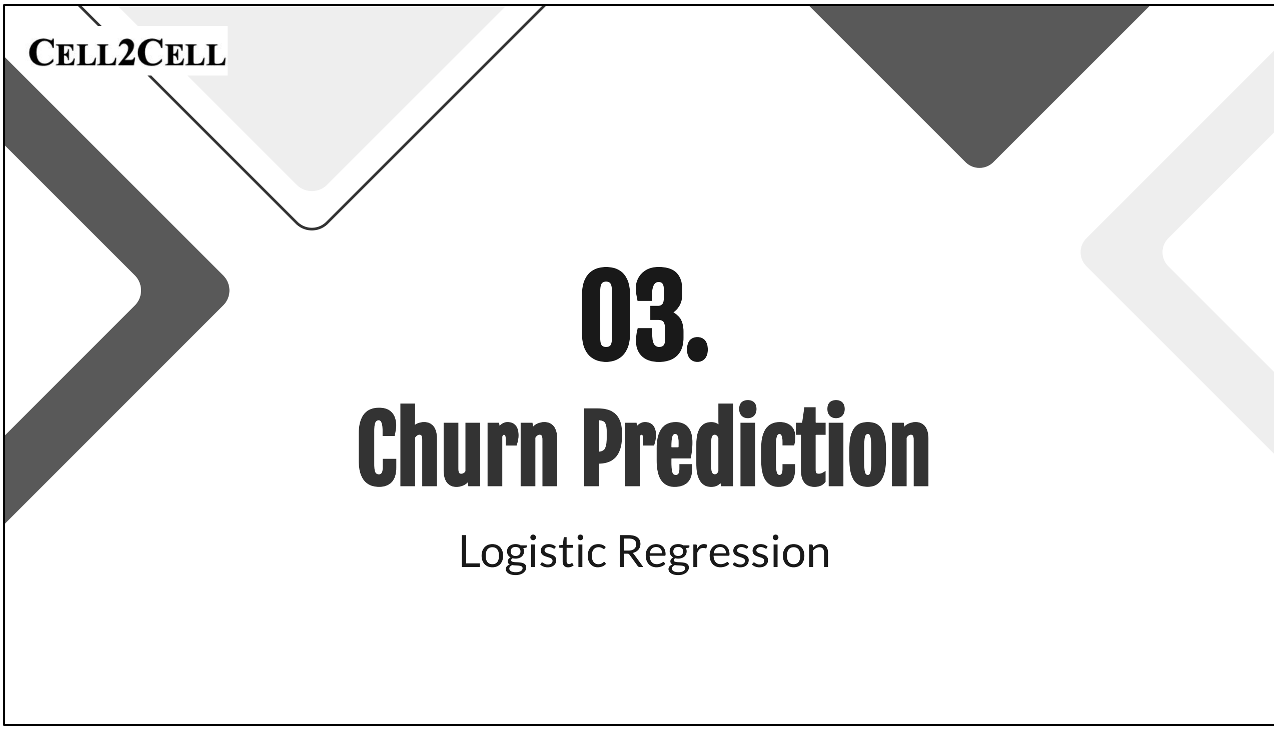
Decision Tree Helps Us Predict Churn Based on Customer Data

As you can see here, in this decision tree, the reasons for churning emerge as the leaves of our tree, with each node representing a node of deliberation and decision. We can see the percent of our data that is influenced by each factor and follow the lines of the tree to make predictions about churn based on these deliberations and decisions.



Overall, from our decision tree model, we can see that

- The most important factor in predicting customer churn is whether they use the equipment for more or less than 306 days.
- The number of months they are in service, with a major dividing line at 12 months, is the 2nd most important factor.



We then constructed a logistic regression model using a stepwise approach to select variables. We started by estimating a null model with only an intercept and then constructs a full model containing all variables of interest. The stepwise regression procedure incrementally adds variables to the model, up to a specified number of steps, based on their contribution to model fit. The resulting model, lrmdl, is then summarized, and its trained parameters are evaluated.

CELL2CELL

LOGISTIC REGRESSION

VARIABLE	PARAMETER ESTIMATE	IMPORTANCE (Pr > (|z|))	MEANING
Eqpdays	0.0026	4.32 E-51 ***	For every extra day that a customer uses current equipment, odds of churning versus not churning go up by 0.26%.
Retcall	0.75	9.89 E-26 ***	Customers who have made a call to the retention team have approximately 2.11 times higher odds of the outcome compared to those who haven't.
Months	-0.0016	0.63	For every extra month in service, odds of churning versus not churning decrease by 0.16%.
Refurb	0.29	1.63 E-13 ***	For every extra handset refurbished, odds of churning versus not churning increase by 34.18%.
Uniqsubs	0.20	4.37 E-15 ***	For every extra unique subscriber, odds of churning versus not churning go up by 21.87%.
Mailres	-0.21	3.20 E-14 ***	Customers who respond to mail offers have approximately 80.97% of the odds of churning versus not churning compared to those who don't respond.
Overage	0.0018	2.04 E-22 ***	For each additional mean overage minute of use, the odds of churning versus not churning increase by approximately 0.18%.
Mou	-2.17 E-5	0.72	For each additional mean monthly minute of use, the odds of churning versus not churning decrease by approximately 0.002%.

To evaluate the results of our logistic regression model, we examined the estimated coefficients, standard errors, z-values, and p-values for each variable included in the model. These metrics provide insights into the significance and impact of each predictor variable on the likelihood of churn. Notably, variables such as "Eqpdays" (number of days of the current equipment), "Retcall" (customer has made a call to the retention team), "Refurb" (handset is refurbished), and "Uniqsubs" (number of unique subscriptions) demonstrate statistically significant associations with churn, as evidenced by their low p-values and non-zero estimates.

Furthermore, we analyzed the odds ratios associated with each predictor variable. The odds ratio represents the change in the odds of churn for a one-unit increase in the predictor variable. For instance, an odds ratio greater than 1 indicates a positive association with churn, while an odds ratio less than 1 suggests a negative association. Variables with odds ratios significantly different from 1 are considered influential predictors of churn.

Additionally, we evaluated interaction terms such as "Eqpdays:Months" and "Months:Mou," which capture the joint effects of two predictor variables on churn. These interactions provide nuanced insights into how the relationship between variables may change under different conditions or contexts.

CELL2CELL

LOGISTIC REGRESSION

VARIABLE	PARAMETER ESTIMATE	IMPORTANCE (Pr > (|z|))	MEANING
Setprcm	-0.25	1.77 E-09 ***	Instances with missing data on handset price have approximately 77.89% of the odds of churning versus not churning compared to instances with available data.
Creditde	-0.24	9.08 E-09 ***	Instances with a low credit rating have approximately 78.47% of the odds of churning versus not churning compared to instances without a low credit rating.
Actvsubs	-0.18	1.35 E-07 ***	For each additional active subscriber, the odds of churning versus not churning decrease by approximately 16.80%.
Roam	0.011	1.77 E-05 ***	For each additional mean number of roaming calls, the odds of churning versus not churning increase by approximately 1.06%.
Changem	-0.00024	2.57 E-06 ***	For each percentage change in minutes of use, the odds of churning versus not churning decrease by approximately 0.024%.
Eqpdays: Months	-3.97 E-05	5.12 E-16 ***	For each unit increase in the product of "Eqpdays" and "Months," the odds of decrease by approximately 0.004%. The joint effect of the number of days of the current equipment and the months in service on the outcome is very close to 1 , indicating minimal impact on the odds of churning versus not churning.
Months: Mou	-1.42 E-05	6.19 E-07 ***	For each unit increase in the product of "Months" and "Mou," the odds of decrease by approximately 0.001%. The joint effect of the number of months in service and the mean monthly minutes of use on the outcome is very close to 1 , indicating minimal impact on the odds of churning versus not churning.

Overall, by carefully analyzing the coefficients, odds ratios, and interaction terms, we gain a comprehensive understanding of the factors driving customer churn in our business. These insights are crucial for informing strategic decisions aimed at mitigating churn and maximizing customer retention.

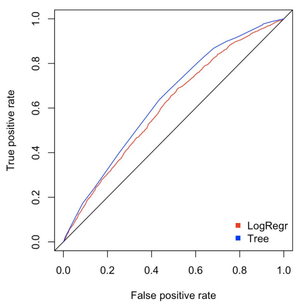
CELL2CELL

04.
The Better Model

CELL2CELL

COMPARING MODELS

AUC



Comparing the AUC scores of the decision tree and logistic regression, **Decision Tree.** seems to perform better

Lift in the 10th Decile

Decision Tree: 1.3245
Logistic Regression: 1.3360

The lift signifies the true churn rate for the group. Basically, how many people in the 10th decile churned divided by the total number of customers in the 10th decile (customers having **0.9 < churn probability < 1**). The higher the lift, the better the model is, so **logistic regression** model seems to be slightly better than the DT.

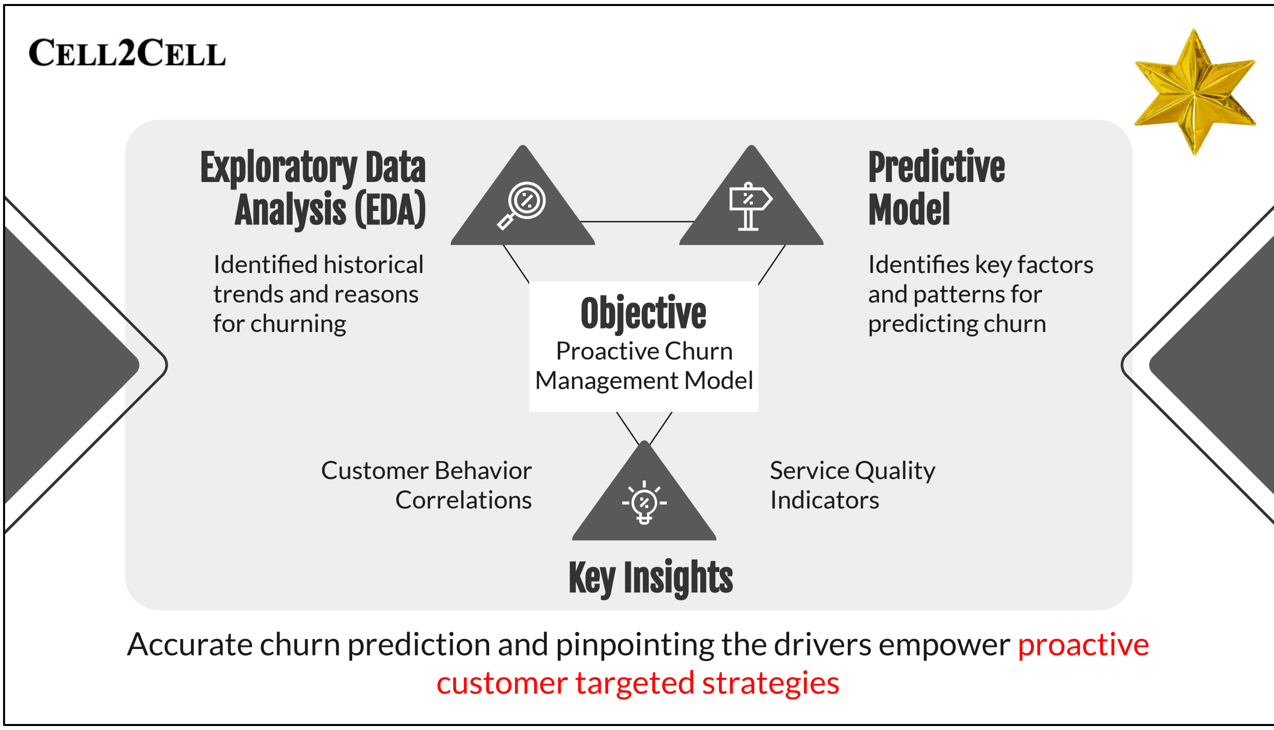
False Negatives

Decision Tree: 1480
Logistic Regression: 1723

False Negatives signify how many customers were predicted to not churn but ended up being churned. We want to minimize the FNs, since we do not want to miss the opportunity of offering promotions to them. FNs are significantly low in **Decision Tree** when compared with logistic regression, it performs better wrt this parameter.

Winner Model: Decision Tree

We compared both, the decision tree model and the logistic regression model with respect to three parameters. First, their Area under the curve for their ROC plots. The decision tree performed better there. Second, the lift in the 10th decile. It signifies how many people in a particular group(10th decile in this case) actually ended up churning. ie the percentage of people who had a predicted probability of churning more than 0.9 / 10. If the lift is more than 1, then the model does give a better return than just selecting the customers at random. The logistic regression has a slightly higher lift than the decision tree model. Thirdly, we looked at the False Negatives. If there's a large number of FNs, we might miss out the opportunity to offer promotions to the customers who are likely to churn. The decision tree has significantly lower number of false negatives than the logistic regression mode. Overall we feel the decision tree gives a better statistical performance than the logistic regression, hence we want to go with it.



Our objective is to anticipate and prevent customer churn through targeted incentives and proactive interventions.

We have created a Predictive Model that identifies the key factors in predicting churn.

Based on these analyses, we have found that:

- Longer tenure and lower value perception correlate with higher churn rates
- Increased customer care calls and retention calls signal dissatisfaction and potential churn
- There are many identifiable factors in predicting churn in our customers, such as equipment usage duration and months in service.

Because of these insights and our predictive capabilities thanks to our churn model, we can make recommendations for our customers to proactively stop churning. **Our predictive model empowers us to anticipate and prevent churn through targeted incentives and proactive interventions, ensuring sustained customer satisfaction and loyalty.**

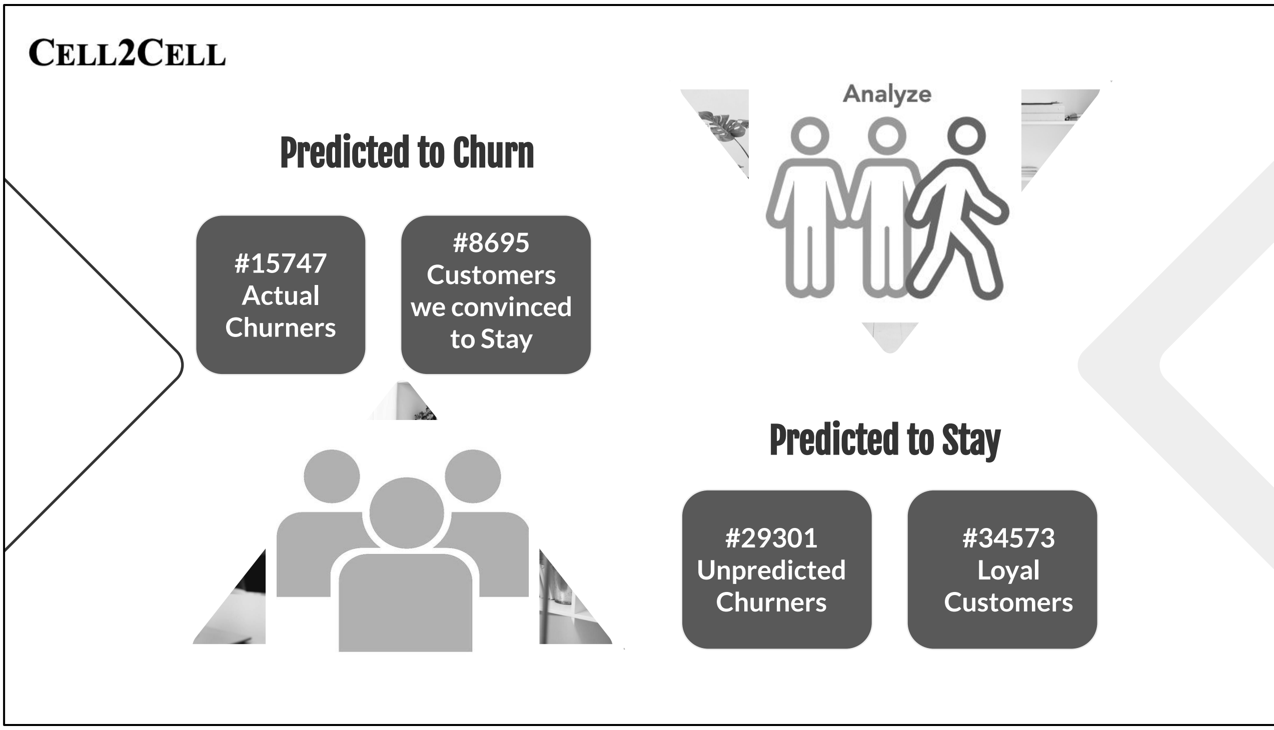
Here are some recommendations for our customers

CELL2CELL

Predicted Results of the four customers

	X15747	X29301	X8695	X34573
Customer	1039199	1073314	1021961	1086325
Revenue	60.325	53.715	5	34.99
Churn	1	1	0	0
pchurn.lr	0.79514	0.276412	0.732123	0.262196
pchurn.tree	0.772871	0.170732	0.772871	0.170732

This is how the prediction for the selected fours users look like (customers on rows #15747, 29301, 8695, 34573).
The first customer on the chart is likely to churn more than 77%, and the third customer is also likely to churn more than 73%.
The second and the third customers are less likely to churn, 17~27% of churn rate respectively.



When we take a closer look, we can categorize them into four categories. First, the one who is predicted to churn, and actually end up churning, second, the one who is predicted to churn but stayed, third, the one who is predicted to stay and stayed, and the one who is predicted to stay but churned.

CELL2CELL

Predicted to Churn

1. **Enhancing Loyalty**
   a. Personalized Promotion Offers - flexible payment options
   b. Priority Customer Service
   c. Free Equipment Upgrade
   d. Bonus data or minutes
2. **Proactive Communication**
3. **Incentives for re-joining**

For individuals predicted to churn, (like customers, rows **# 15747, 8695**), we primary recommendation is to enhance customer loyalty to our company. First, we propose offering personalized promotion offers, such as flexible payment options and customizable plans tailored to their preferences. Additionally, they can receive priority customer service and support, as well as free equipment upgrades if needed. Also offering bonus data or minutes may make customers to be more on their phones, increasing the likelihood of continuing to use our services.

Moreover, to deter them from actually making the decision to churn,proactive communication is crucial. Targeted outreach to assure customers that their concerns are valued and offering solutions to address any issues they may encounter will demonstrate our commitment to their satisfaction.
These strategies will likely to convince customers like customer #8695 to change their minds, by satisfying their needs.

Furthermore, even if they actually lead to churn, we can entice customers to rejoin by offering incentives such as free additional services upon return. This strategy aims to encourage them to reconsider their decision and remain with us.
With this strategy, we can also bring back potential customers like customer #15747 even after they decide to churn.

CELL2CELL

Predicted to Stay

1. **Loyalty Awards**
   a. Bonus data or minutes
   b. Early or bonus access to premium features
   c. Annual discount for staying with us for another year
2. **Feedback Sollicitation**

For people who are predicted remain with us**(customer #29301, 34573)**, we can implement a Loyalty Awards Program to show our appreciation for their loyalty and encourage their continued patronage, fostering strong relationships. As part of this program, we can offer various incentives such as bonus data or minutes, early or bonus access to premium features, and an annual discount for those who remain with us for a year to stay with us after the normal 12 months cutoff.

Also, actively seeking feedback from these customers enables us to understand their needs, preferences, and areas of improvement better. This proactive approach helps us to continuously enhance the customer experience and strengthen customer loyalty to Cell2cell.
This strategy will allow us not only to enhance company loyalty to current satisfied customers like customer #34573, but also to deter potential unexpected customers from churning, like customer #29301.


CELL2CELL
Thank you!
Any Questions?