XYZ

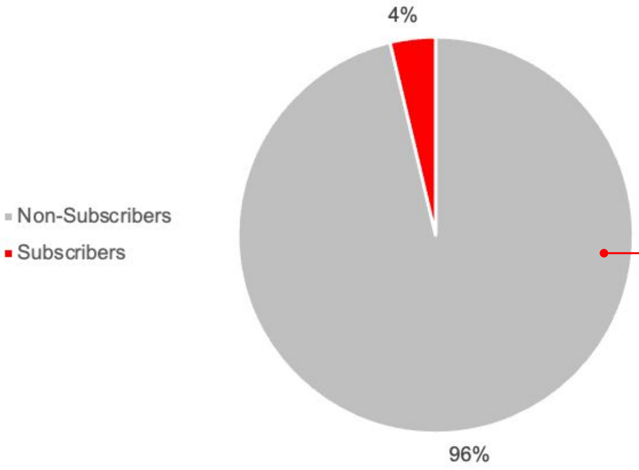
Predictive Analytics For Premium Subscribers

XYZ Data Analyst Team

Lily, Yoon, Shreya, Ujjawal, Harshithan

Purpose of the analytics

"How to build a **predictive model** to identify **which free users are most likely to convert to premium subscribers** in the next campaign"



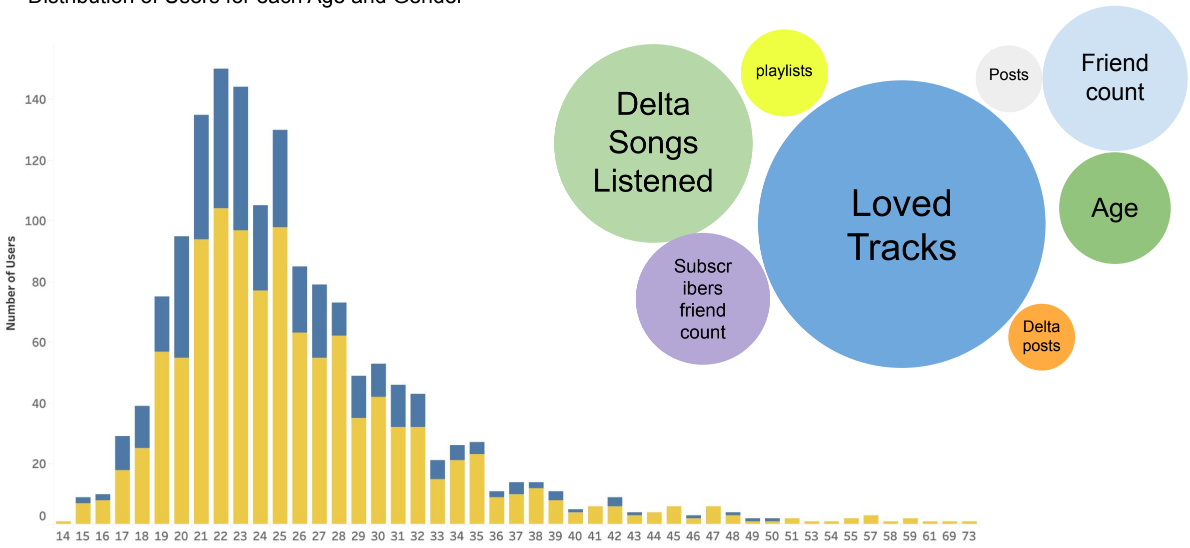
[Previous Campaign Result]

PREMIUM SUBSCRIPTION

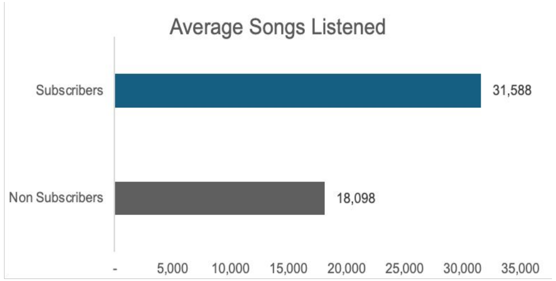
Target Audience - What does our adopters look like?

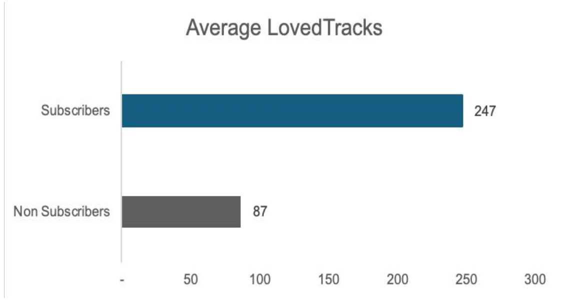
To predict, we analyze 25 user attributes including age, gender, number of friends, listening habits, and engagement.

Distribution of Users for each Age and Gender



Distribution of Users for each Age. Color shows details about Male (0 or 1). The data is filtered on Adopter, which keeps 1.





Analytics Results : AUC 0.7718

What is AUC

We can adopt our model, since it gives as a **AUC** value of **0.7718.**

AUC is a metric that helps measure a model's ability to distinguish between two classes. **AUC reflects how well the model can rank potential adopters above non-adopters**.

An AUC of 0.7718 suggests that our model has a good ability to discriminate between users who are likely to convert to premium subscribers and those who are not. Specifically, there is a **77.18%** chance that a randomly selected user who will become a subscriber will have a higher predicted probability of conversion than a randomly selected user who will not convert.

Why AUC

AUC is especially **useful in imbalanced datasets** like ours, where accuracy alone might be misleading. Furthermore, as AUC shows the model's ranking ability, **it allows us to targeted marketing efforts to focus on the highest probability users.**

Analytics Results

	Predicted Non-Adopter(0)	Predicted Adopter(1)
Actual Non-Adopter(0)	4,228	1,783
Actual Adopter(1)	52	168

Confusion Matrix
It shows how well our predictive model identifies users who will subscribe to the premium service (adopters) versus those who won't (non-adopters)

Predicted Adopters (0): 4,228 users are correctly classified as non adopters, meaning marketing resources were not spent on them unnecessarily. **52** represents missed opportunities—users who were likely to subscribe but were not identified by the model.

Predicted Adopters (1): 1,783 are users that the model incorrectly identified as likely to subscribe but did not convert. **168** are users accurately identified as likely to adopt the premium service.

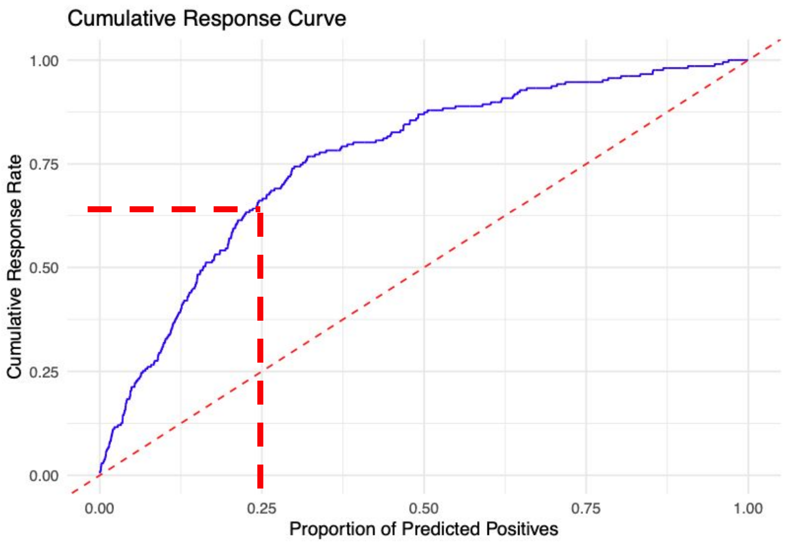
The cost savings from accurately identifying non-adopters outweigh the resources spent on incorrect predictions, and the number of **correctly identified adopters** is greater than the missed opportunities.

Analytics Results

- **Efficient Targeting**: **By targeting 25% of the users, our model can capture about 65% of the actual premium subscribers.** We can ensure marketing resources are spent on users with the highest likelihood of conversion, which boosts the effectiveness of the marketing campaign

- **Cost Savings**: Since the model outperforms random targeting (as shown by the blue curve above the red line), We can **reduce marketing costs** by targeting fewer users while still driving significant conversion rates.

- **Diminishing Returns**: After targeting around 40% of users, the model's effectiveness declines, meaning further targeting adds fewer subscribers. Focusing on the top users identified by the model will **maximize impact without wasting resources** on less likely converters.



How can our model help our business?

- Concentrate resources on users most likely to convert,

  Reduce the cost on users unlikely to convert

- Improves the overall conversion rate for the marketing campaign

- Focus on features of our models in communication and marketing campaigns to better

  resonate with users who are most likely to subscribe.

Features of Our Model

Our model highlights several top features—: **loved tracks, delta song listened, delta loved tracks, song listened, subscriber friend count, shouts**—as primary factors in predicting whether users will subscribe to the premium service after the marketing campaign.

Loved tracks, Song listened:

The volume of music that users have liked and listened are key factors influencing conversion to premium users

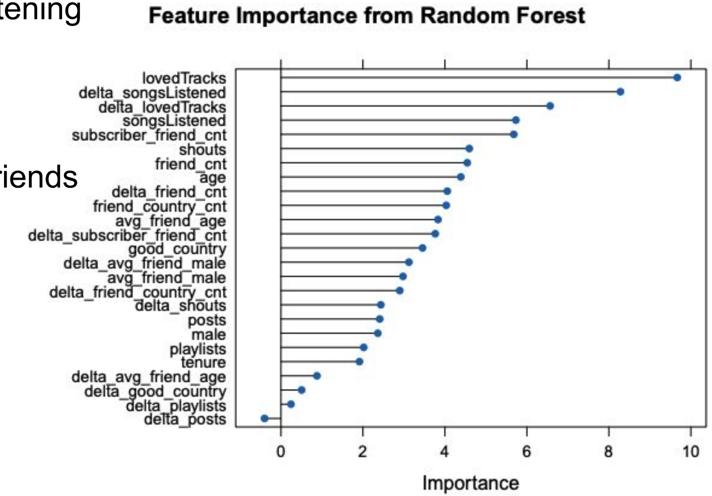
Delta song listened, Delta loved tracks:

During the marketing campaign, users who have shown a change in song-listening activity and interest in exploring new music are more likely to convert.

Subscriber friend count, Shouts:

Potential premium subscribers are influenced by their premium-subscribing friends and social engagement on the platform

Corresponding to these important features, we can emphasize our premium services such as personalized tracks recommendations and social interaction features to attract potential subscribers.



Conclusion and Next Step

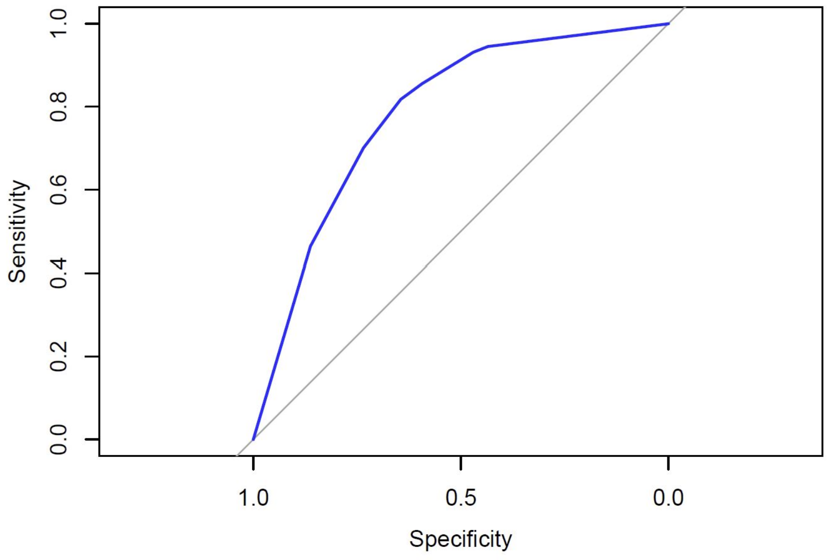
We can target more potential premium subscribers and optimize marketing costs using this predictive model in the next marketing campaigns.

Next Steps: We recommend testing the model in the next marketing campaign and monitor the adoption rates. By collecting more adopters data, we can further refine the predictive model to improve our targeting strategies.

End of Document

Analytics Results

- **Specificity** tell us about the proportion of non-adopters incorrectly predicted as adopters.
- **Sensitivity** tell us about the proportion of actual adopters that are correctly identified by the model.
- **The ROC curve** in blue indicates the trade-off between the true positive rate and the false positive rate at various threshold levels.
- The closer the curve is to the top-left corner, the better the model is at distinguishing between adopters and non-adopters. In this case, the ROC curve is above the diagonal line and bends towards the top-left, showing that the decision tree model is better than random guessing. However, it doesn't fully hug the top-left corner, indicating that while the model is somewhat effective, it could be improved further.



The benefit of our solutions

Our model achieves an AUC of 0.7599, meaning it has about a **76%** chance of **correctly identifying a potential premium subscriber** when comparing one adopter and one non-adopter.

This demonstrates that our model can distinguish between users likely to convert and those who are not, correctly ranking a potential subscriber higher than a non-subscriber approximately 76 out of 100 times in similar cases.
This level of accuracy suggests that the model is significantly more effective than random targeting, which would succeed only 50 out of 100 times.

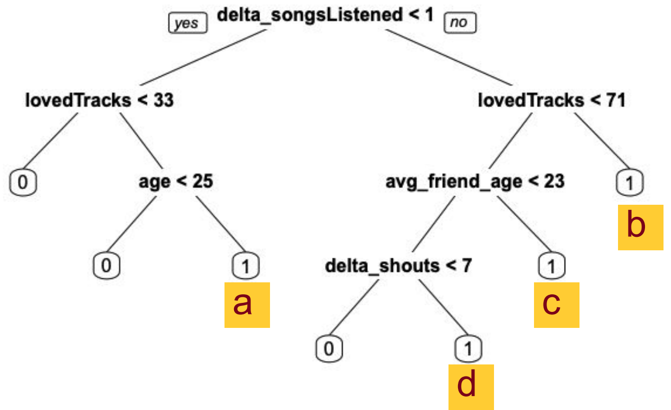
Features of Our Model

Our model highlights several features—: **delta song listened, loved tracks, age, average friend age, delta shouts**—as primary factors influencing a user's decision to subscribe to the premium service.

- Potential premium subscribers:
  Users who didn't increase their song-listening activity (delta song listened < 1):
  a. with higher loved Tracks (>= 33), and older age (>= 25) are more likely to convert.
  Users who increased their song-listening activity (delta song listened >= 1):
  b. with interest in discovering different music (loved Tracks >= 71)
  c. with lower loved tracks (< 71), and older friends (average age >= 23)
  d. with less loved tracks(< 71), younger friend (average age < 23), and more social engagement (delta shouts >= 7)

XYZ can emphasize its premium features such as personalized tracks recommendations, social interaction features to attract the potential subscribers.

a



Cost Sensitive Evaluation

- **XYZ's goal is to prioritize marketing costs to potential adopters,** which means targeting users unlikely to convert costs more to XYZ company, comparing to miss potential subscribers. Thus, we want to minimize the cost of misclassifying non-adopter("0") to adopter("1")
- Let's assume the average marketing spend per user is $10
- Let's assume the monthly premium subscription fee is $10, and a user remains subscribed for an average of 12 months

	last campaign (random targeting)	next campaign (our predictive model)
cost of misclassifying each user	4.4	4.5

Cost Sensitive Evaluation (Appendix)

	last campaign (random targeting)	next campaign (predictive model)
FP cost	10	10
FN cost	120	120
adopters %	0.5	
non adopter %	0.5	
FP	4416	2213
FN	89	47
total users	12462	6171
FP* FP cost	44160	22130
FN* FN cost	10680	5640
total misclassification cost	54840	27770
Cost per users (total misclassification cost / total users)	4.400577756	4.500081024