High Note

Internal Memo

Subject: Recommending how to keep High Note users engaged so they do not wander off

To: Lisa Peschke, Marketing Director

From:

Sylvanne Braganza (sbraganz)

Yukta Butala (ybutala)

Clarissa Gunawan (cgunawan)

Kelly Hong (sewonh)

Cece Zhang (xiyuzhan)

Analytical Marketing

February 12, 2024

Dear Lisa,

RE: Recommendations on how to keep High Note users engaged so they do not wander off

This memo highlights the importance of High Note's (HN) user engagement and proposes strategies to keep them as long-term adopters. To ensure HN's maintain user retention in the platform, we recommend focusing on the following promotions and service improvements:

- **Offer free 3-month premium plans** to showcase its ad-free and offline music listening features. This will let users experience the full value proposition before committing to a subscription.
- **Offer yearly discounts on premium plans** to encourage users to subscribe for a longer period.
- **Friend referral programs for premium plans**, and offer rewards (i.e., free premium months to both the referrer and the friend), to encourage engagement and paid user growth.
- **Improving song recommendations**, by incorporating songs from friends' playlists. This can be done by implementing a "Friends' Top Songs" recommendation.
- **Expand the song selection**, including more songs in the "long tail" and from international artists.
- **Encourage social engagement on HighNote's platform,** by introducing features like a virtual "song-party", where users can sync with their network circle

Part 1: Analyzing current user behaviors

To validate the effectiveness of the retention strategies, below is a comparative analysis of adopters versus non-adopters, in terms of the demographics, behaviors, and network circles (Appendix 1):

Demographics	**Age range:** Adopters are generally older (mean: 26.20) than non-adopters (mean: 25.51). **Gender:** In general, more males use the platform, but a higher proportion of males are adopters (0.72) compared to non-adopters (0.62).
User behavior	**Songs listened:** Adopters listen to ~2 times more songs compared to non-adopters. **Playlists:** Adopters have ~3 times more playlists compared to non-adopters. **Loved Tracks:** Adopters have ~3 times more loved tracks compared to non-adopters. **Posts:** Adopters create ~5 times more posts compared to non-adopters. **Shouts:** Adopters have ~5 times more shouts compared to non-adopters.
Network circle	**Friend count:** Adopters have ~2.5 times more friends who are around the same age. **Subscriber friend count:** Adopters have ~5 times more friends who are adopters. **Friend country count:** Adopters have more friends from different countries (mean: 5.34 countries) compared to non-adopters (mean: 2.62)

The demographics suggest that HN appeals to slightly older male audiences. A high engagement rate shows that users are music lovers. Additionally, adopters' high social activity highlights strong network effects that HN can use to leverage a referral program and enhance its platforms' social features.

Part 2: Understanding value for users

To better understand the factors influencing user premium service adoption and what attracts our users, we considered two predictive models: decision tree and logistic regression. These models were chosen for their effectiveness in analyzing binary outcomes, specifically whether users adopt a paid subscription or not. Our decision to explore these models is informed by insights from our exploratory data analysis, aiming to utilize available data for informed decisions on engaging existing users. Decision trees excel in uncovering complex relationships in the data, while logistic regression offers a straightforward approach to modeling user adoption patterns. Both models provide complementary insights, allowing us to explore various aspects of customer behavior and inform strategic decisions regarding subscription offerings and user engagement on the platform. Ultimately, we built (Appendix 2) and assessed the models to find which best suits our objectives.

Our findings from the decision tree model (Appendix 2.1) revealed several key features that emerged as significant contributors to user adoption patterns on High Note. Notably, metrics such as "lovedTracks," "subscriber_friend_cnt," "songsListened," and "ave_friend_age" were pivotal in predicting whether users would adopt paid subscriptions.

In our analysis of the data using logistic regression (Appendix 2.2), we identified several key factors that influence the likelihood of adoption. The most influential factors in predicting adoption rates of the paid streaming service are the number of friends who subscribe and have playlists, which increase adoption odds by approximately 48% and 24.02%, respectively. Being male is also a significant predictor, increasing adoption odds by approximately 56.75%. Conversely, the interaction between being from a "good_country" and having playlists slightly reduces adoption odds by approximately 16.76%, while being from the US, UK, or Germany decreases adoption odds by approximately 22.15% compared to other countries.

Though both models uncovered insights into engaging our users, and modeling both helped us to gain a better understanding of our users, we decided to base our further analyses and recommendations on our logistic regression model for several reasons.

First, based on the characteristics of our data, we initially thought logistic regression could be considered as a more suitable choice. It offers advantages in handling continuous data, such as age and tenure, preserving the information content in the data. In contrast, decision trees can result in information loss and reduced predictive accuracy when dealing with continuous data. Logistic regression also tends to perform better on data with noises or irrelevant features. Conversely, decision trees are susceptible to potentially poor generalization performance on unseen data, especially when the data contains noise.

As we further continued our analysis, we confirmed our initial intuition. We extensively compared the performance of the two models and focused on understanding how well they could predict which customers might adopt our premium plan. We used a metric called the Area Under the Curve (AUC) to compare the performance of the models (Appendix 3.1 and 3.2). From our analysis, we found that the logistic regression model performed slightly better than the decision tree model.

Additionally, upon reviewing the confusion matrices of both models (Appendix 3.3), which helps us understand how well the models classify customers into different categories, we observed that the logistic regression model had approximately 9% higher accuracy than the decision tree model.

Our primary goal is to accurately identify potential premium customers to target the right customers and keep them engaged. To achieve this, when evaluating our model, we prioritized predicting the proportion of actual adopters among the predicted adopters over the proportion of predictions of adopting the premium plan among the actual adopters. This distinction is crucial as it enables us to allocate marketing resources more effectively, targeting those customers most likely to convert to our premium plan while minimizing wasteful spending on those less likely to do so.

Additionally, the logistic regression model effectively identified 97% of customers who remain on the free plan, which was 11% higher than that of the decision tree model. Moreover, it outperformed the decision tree model in terms of lift value, which shows a superior predictive capability regarding the adoption rate.

By recognizing the significance of the features shown in our logistic regression model, we can tailor our engagement strategies to encourage users to increase their interactions with the platform, thus reducing the likelihood of them wandering off. Additionally, understanding the impact of demographics, such as age, gender, and country of origin, allows us to personalize our promotional efforts and service improvements.

Part 3: Focusing recommendations

We conducted a clustering analysis utilizing the available data on user characteristics for a more targeted approach to understanding user behavior and preferences and crafting recommendations.

We began by fitting a logistic regression model to the available data, aiming to predict a user's likelihood of subscribing to premium based on a set of predictor variables. This model provided valuable insights into the factors influencing user behavior and identified significant predictors. Following the logistic regression analysis, we constructed a model matrix containing the predictor variables used in the logistic regression model. This matrix served as the basis for the subsequent clustering analysis (Appendix 4.1). Utilizing the model matrix, we applied k-means clustering to group users into distinct segments based on their characteristics or behaviors (Appendix 4.2). To ensure that the clustering analysis prioritized variables with the most significant impact on user behavior, we weighted the data using the logistic regression coefficients. This weighting scheme emphasized variables that had a stronger association with the outcome of interest, thereby enhancing the clustering analysis's sensitivity to relevant features (Appendix 4.3).

We identified 5 distinct customer clusters from the analysis: Engaged Veterans, Passive Patrons, Discerning Listener, Casual Explorers, and Energetic Trendsetters (see Appendix 4.4 for detailed profiles).

Observed from the results of the logistic regression, **male, subscriber_friend_cnt, playlists, and age** are a few of the most significant factors in predicting a user's likelihood of subscribing to a premium membership. Therefore, the promotional offer will be the most effective when targeted to clusters 1, 3, and 5 for their moderate to high engagement on the platform and relatively higher age.

We also considered that we make a lot more money from paid users than free users, 24 times more. Therefore, in addition to keeping our existing users engaged, we want to convert as many as possible into paid users.

Assuming an adoption rate of 10% for a generic free 3-month subscription promotion offer, and 20% with our suggested targeted recommendations, we have calculated the expected yearly profit generated in the table below:

	Projected revenue (yearly)	Cost	Projected profit (yearly)
10% adoption rate	10%*($3*3,317*12) = 11,941.2	$1*3,317*3 = 9,951	$1,990.2
20% adoption rate (with recommendations)	20%*($3*3,317*12) = 23.882.4	$1*3,317*3 = 9,951	$13,931.4

For cluster 4, the casual explorers, they are the biggest group among the 5 clusters as observed from our clustering analysis, yet the least active. In order to keep them from wandering off, we are engaging them with our personalized strategies and, eventually, turning them into paying premium subscribers (Appendix 4.5).

Part 4: Summary for customer retention and maximum profit

Our analysis method employs a blend of qualitative and quantitative methods, including clustering and predictive modeling, to understand user behaviors on HighNote. Our data-driven insights leverage advanced statistical techniques and machine learning algorithms, prioritizing the logistic regression model for its accuracy. Through clear projections of revenue impact and targeted strategies tailored to specific user clusters, we offer tangible evidence of the potential for maximizing user retention and conversion, solidifying High Note's competitiveness in the music streaming market.

Our model highlights subscriber friends, playlists, and being male as important parameters. As we saw in the demographics for adopters, these were the most important parameters as well. In assessing High Note's strategic approach to its freemium music streaming model, it becomes evident that clear differentiation among target clusters is imperative for effective marketing. Additionally, aligning conversion rate expectations with industry standards, ranging from 10% to 20%, is crucial for optimizing revenue generation and user acquisition. Presently, High Note has a 2.9% premium user base, demonstrating a competitive standing but with room for growth within its relatively modest user pool.

To expand its premium market share and drive profits, High Note must strategically balance between enriching its free subscription offerings, such as by introducing offline streaming features and enhancing recommendation systems, to attract and retain a larger user base, while also focusing efforts on converting a higher percentage of users into premium subscribers. By pursuing these dual strategies, High Note can improve its competitive position in the rapidly growing music streaming market while maximizing revenue potential.

Appendices

Appendix 1

Below are the descriptive statistics of different features used to assess the characteristics of non-adopters (group: 0) and adopters (group: 1)

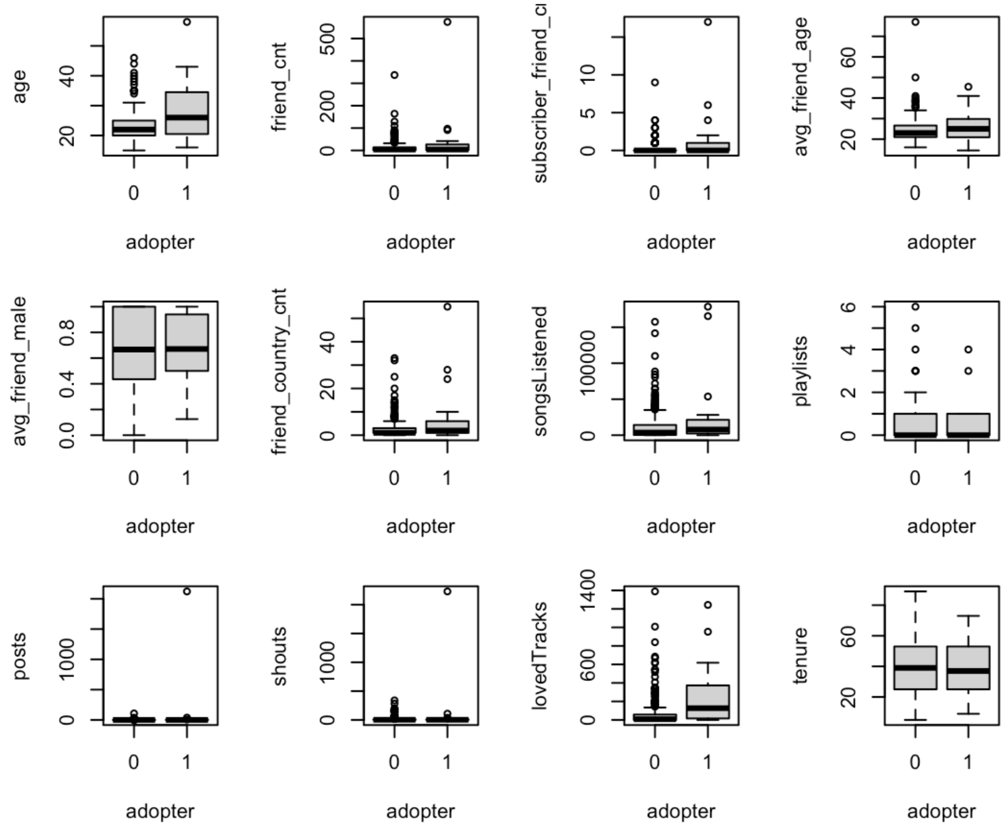
group: 0

	vars	n	mean	sd	min	max	range	se
age	1	60140	24.29	4.88	8	79	71	0.02
age_Missing	2	60140	0.48	0.50	0	1	1	0.00
male	3	60140	0.62	0.39	0	1	1	0.00
male_Missing	4	60140	0.37	0.48	0	1	1	0.00
friend_cnt	5	60140	11.16	41.53	0	3921	3921	0.17
subscriber_friend_cnt	6	60140	0.27	1.89	0	309	309	0.01
avg_friend_age	7	60140	24.52	5.10	8	79	71	0.02
avg_friend_age_Missing	8	60140	0.20	0.40	0	1	1	0.00
avg_friend_male	9	60140	0.63	0.33	0	1	1	0.00
avg_friend_male_Missing	10	60140	0.16	0.36	0	1	1	0.00
friend_country_cnt	11	60140	2.62	4.67	0	119	119	0.02
songsListened	12	60140	12019.41	23629.51	0	1000000	1000000	96.35
playlists	13	60140	0.49	1.54	0	261	261	0.01
posts	14	60140	2.63	47.14	0	5644	5644	0.19
shouts	15	60140	17.58	118.85	0	8694	8694	0.48
shouts_Missing	16	60140	0.02	0.13	0	1	1	0.00
lovedTracks	17	60140	67.53	229.04	0	12522	12522	0.93
tenure	18	60140	39.39	19.27	0	108	108	0.08
good_country	19	60140	0.37	0.38	0	1	1	0.00
good_country_Missing	20	60140	0.37	0.48	0	1	1	0.00

group: 1

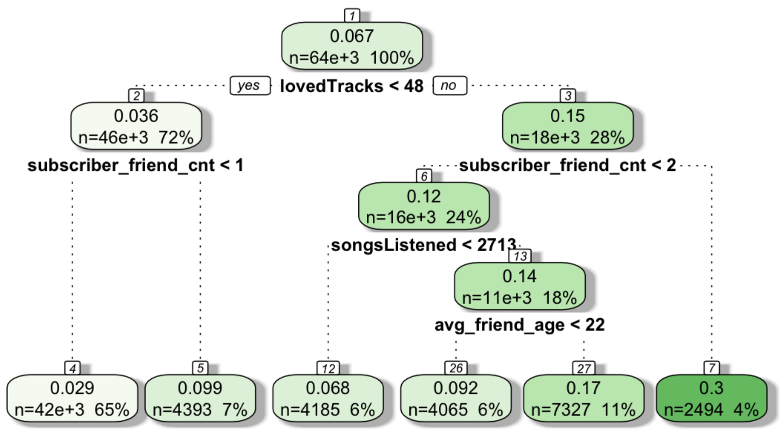
	vars	n	mean	sd	min	max	range	se
age	1	4327	25.51	5.66	8	78	70	0.09
age_Missing	2	4327	0.38	0.49	0	1	1	0.01
male	3	4327	0.70	0.38	0	1	1	0.01
male_Missing	4	4327	0.27	0.45	0	1	1	0.01
friend_cnt	5	4327	28.83	108.10	0	5089	5089	1.64
subscriber_friend_cnt	6	4327	1.28	5.39	0	287	287	0.08
avg_friend_age	7	4327	25.63	5.19	12	70	58	0.08
avg_friend_age_Missing	8	4327	0.16	0.37	0	1	1	0.01
avg_friend_male	9	4327	0.65	0.26	0	1	1	0.00
avg_friend_male_Missing	10	4327	0.13	0.34	0	1	1	0.01
friend_country_cnt	11	4327	5.34	8.11	0	136	136	0.12
songsListened	12	4327	25654.79	41737.56	0	1000000	1000000	634.50
playlists	13	4327	1.34	29.63	0	1943	1943	0.45
posts	14	4327	13.36	126.29	0	5176	5176	1.92
shouts	15	4327	82.06	1138.39	0	65872	65872	17.31
shouts_Missing	16	4327	0.04	0.19	0	1	1	0.00
lovedTracks	17	4327	223.57	798.85	0	44005	44005	12.14

Below is a boxplot to visually compare the quartiles and the distribution of non-adopters (value = 0) and adopters (value = 1).

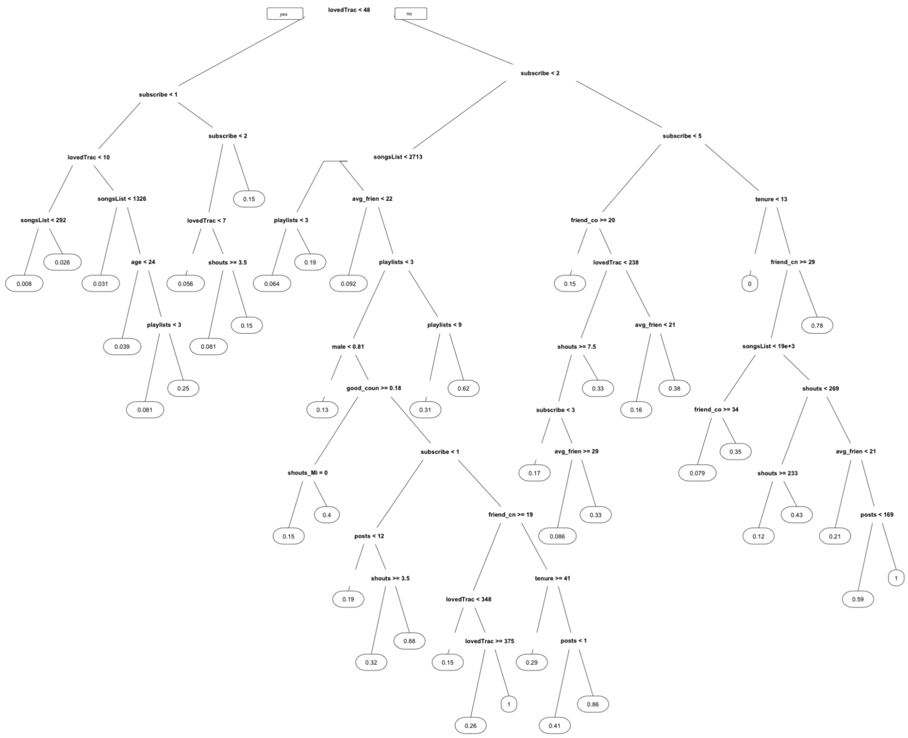


Appendix 2
Building the models:

Appendix 2.1
Decision Tree:



Appendix 2.1.1
Full decision tree:

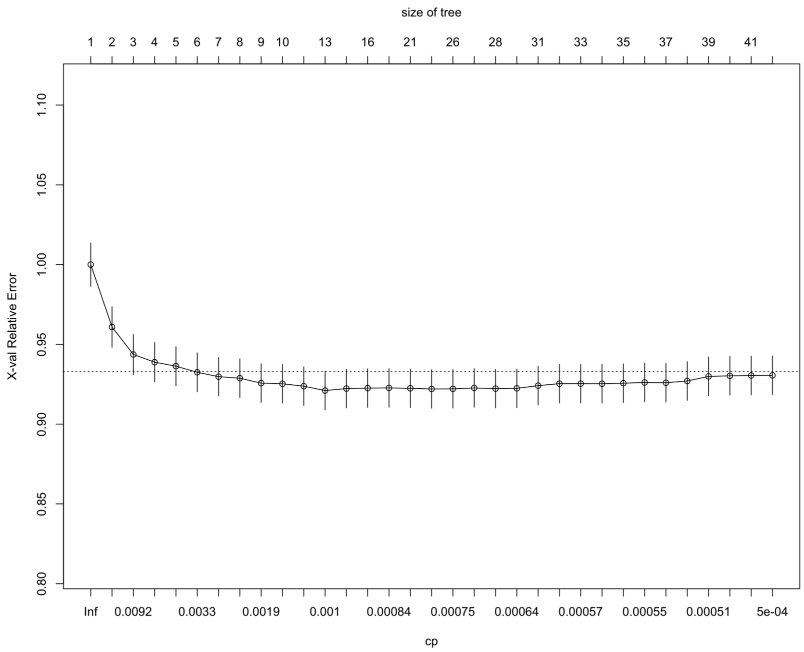


To reduce the complexity and thus improve predictive accuracy by the reduction of overfitting, we must prune this tree.

Appendix 2.1.2

Finding optimal complexity parameter (cp):

Finding the optimal cp 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.

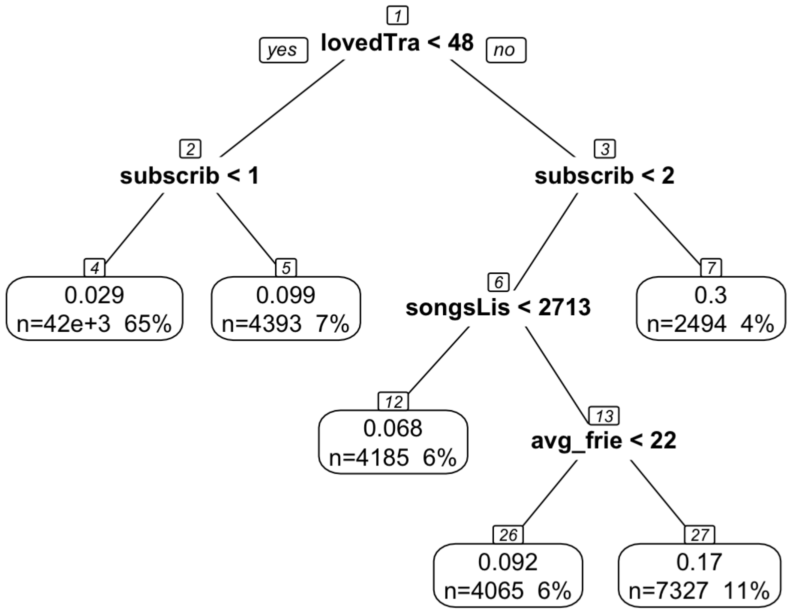


Plotting the cross-validation relative error against different complexity parameters (cp) allows us to visually identify the point where the relative error stabilizes or starts increasing, indicating the optimal complexity parameter. This helps us select the parameter that minimizes the error while preventing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

As we can see from the plot above, the optimal cp for our DT model is 0.0033.

Appendix 2.1.3

Pruned DT using optimal cp:



Appendix 2.1.4

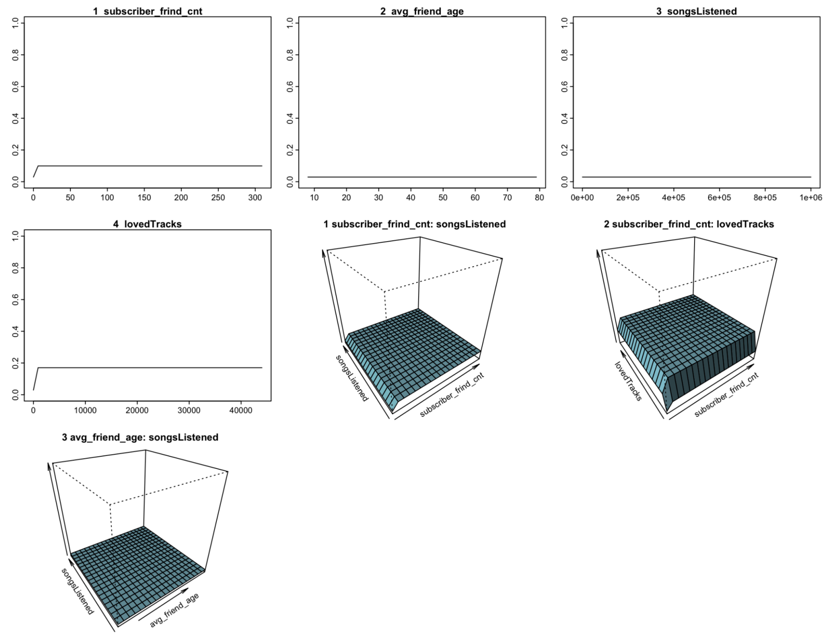
Extracting important features from model:

```
Variable importance
           lovedTracks   subscriber_friend_cnt              friend_cnt      friend_country_cnt
                    39                      25                       9                       9
         songsListened                  shouts          avg_friend_age                     age
                     6                       5                       4                       2
avg_friend_male_Missing  avg_friend_age_Missing
                     1                       1
```

So the important features are: lovedTracks, subscriber_friend_cnt, songsListened, ave_friend_age.

Appendix 2.1.5

Partial dependent graphs (plots of a model's response over a range of predictor values (the model surface)) also show this:



Appendix 2.2

Logistic Regression:

Appendix 2.2.1

Logistic Regression Coefficients:

```
Call:
glm(formula = adopter ~ lovedTracks + songsListened + subscriber_friend_cnt +
    age + male + good_country + playlists + friend_cnt + friend_country_cnt +
    avg_friend_age + subscriber_friend_cnt:age + good_country:playlists +
    subscriber_friend_cnt:playlists + lovedTracks:friend_cnt +
    friend_cnt:friend_country_cnt, family = "binomial", data = rfreemium[trainsample,
    crvarlist])

Coefficients:
                                 Estimate Std. Error z value Pr(>|z|)    
(Intercept)                    -4.556e+00  9.228e-02 -49.366  < 2e-16 ***
lovedTracks                     9.275e-04  5.750e-05  16.130  < 2e-16 ***
songsListened                   6.874e-06  4.936e-07  13.925  < 2e-16 ***
subscriber_friend_cnt           3.921e-01  3.306e-02  11.861  < 2e-16 ***
age                             2.808e-02  3.166e-03   8.868  < 2e-16 ***
male                            4.495e-01  4.440e-02  10.125  < 2e-16 ***
good_country                   -2.504e-01  4.522e-02  -5.536 3.09e-08 ***
playlists                       2.153e-01  1.874e-02  11.489  < 2e-16 ***
friend_cnt                      1.664e-03  7.974e-04   2.086   0.0369 *  
friend_country_cnt              1.837e-02  4.405e-03   4.169 3.05e-05 ***
avg_friend_age                  2.200e-02  3.266e-03   6.736 1.63e-11 ***
subscriber_friend_cnt:age      -7.090e-03  1.003e-03  -7.069 1.56e-12 ***
good_country:playlists         -1.834e-01  1.977e-02  -9.277  < 2e-16 ***
subscriber_friend_cnt:playlists -1.604e-02  1.872e-03  -8.569  < 2e-16 ***
lovedTracks:friend_cnt         -3.828e-06  5.340e-07  -7.169 7.58e-13 ***
friend_cnt:friend_country_cnt  -6.914e-05  8.544e-06  -8.092 5.87e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

The model is assuming an additive relationship between the log-odds of the outcome variable (adopter) and the predictor variables. In other words, for each unit increase in a predictor variable, the log-odds of the outcome variable will increase or decrease by a certain amount, depending on the sign and magnitude of the coefficient for that predictor variable.

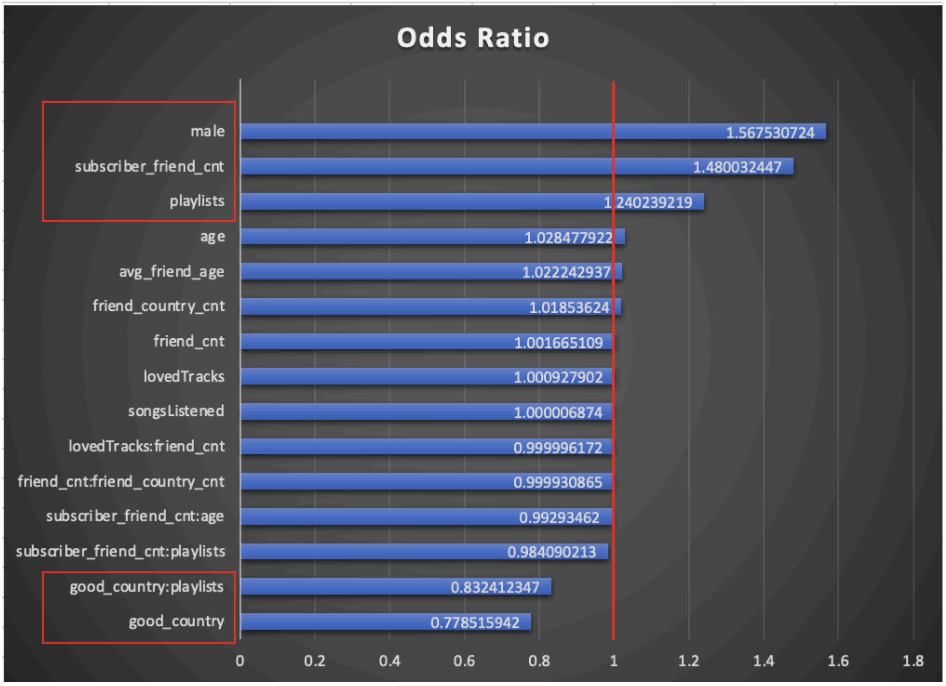
Here is an interpretation of each term in the model:

- Intercept: The intercept is the log-odds of the outcome variable when all of the predictor variables are equal to zero. In this case, the intercept is -4.556, which means that the initial log-odds of someone being an adopter is negative, or less than 1. This suggests that, on average, people are more likely not to be adopters.
- lovedTracks: The coefficient for lovedTracks is 0.0009275, which means that for each additional track that someone loves, the log-odds of them being an adopter increases by 0.0009275. This is a small but statistically significant effect.
- songslistened: The coefficient for songslistened is 0.00006874, which means that for each additional song that someone listens to, the log-odds of them being an adopter increases by 0.00006874. This is another small but statistically significant effect.
- subscriber_friend_cnt: The coefficient for subscriber_friend_cnt is 0.3921, which means that for each additional friend that someone has who is also a subscriber, the log-odds of them being an adopter increases by 0.3921. This is a moderate and statistically significant effect.
- age: The coefficient for age is 0.02808, which means that for each additional year of age, the log-odds of someone being an adopter increases by 0.02808. This is a small but statistically significant effect.
- male: The coefficient for male is 0.4495, which means that males have log-odds of being an adopter that are 0.4495 higher than females. This is a moderate and statistically significant effect.

- good_country: The coefficient for good_country is -0.2504, which means that people who listen to mostly country music have log-odds of being an adopter that are 0.2504 lower than people who do not listen to mostly country music. This is a moderate and statistically significant effect.
- playlists: The coefficient for playlists is 0.2153, which means that for each additional playlist that someone has, the log-odds of them being an adopter increases by 0.2153. This is a moderate and statistically significant effect.
- friend_cnt: The coefficient for friend_cnt is 0.001664, which means that for each additional friend that someone has, the log-odds of them being an adopter increases by 0.001664. This is a small but statistically significant effect.
- friend_country_cnt: The coefficient for friend_country_cnt is 0.01837, which means that for each additional friend that someone has who listens to mostly country music, the log-odds of them being an adopter increases by 0.01837. This is a small but statistically significant effect.
- avg_friend_age: The coefficient for avg_friend_age is 0.022, which means that for each additional year of average age of someone's friends, the log-odds of them being an adopter increases by 0.022. This is a small but statistically significant effect.
- subscriber_friend_cnt:age: The coefficient for the interaction term subscriber_friend_cnt:age is -0.00709, which means that the effect of having friends who are also subscribers is weaker for older people. This is a small but statistically significant interaction effect.
- good_country:playlists: The coefficient for the interaction term good_country:playlists is -0.1834, which means that the negative effect of listening to mostly country music is weaker for people who have more playlists. This is a small but statistically significant interaction effect.
- subscriber_friend_cnt:playlists: The coefficient for the interaction term subscriber_friend_cnt:playlists is -0.01604, which means that the effect of having friends who are also subscribers is weaker for people who have more

Appendix 2.2.2

The coefficients were exponentiated to get odds ratios, which are easier to interpret. An odds ratio greater than 1 indicates an increase in the odds of adopting a paid subscription for a one-unit increase in the feature, while an odds ratio less than 1 indicates a decrease in the odds:

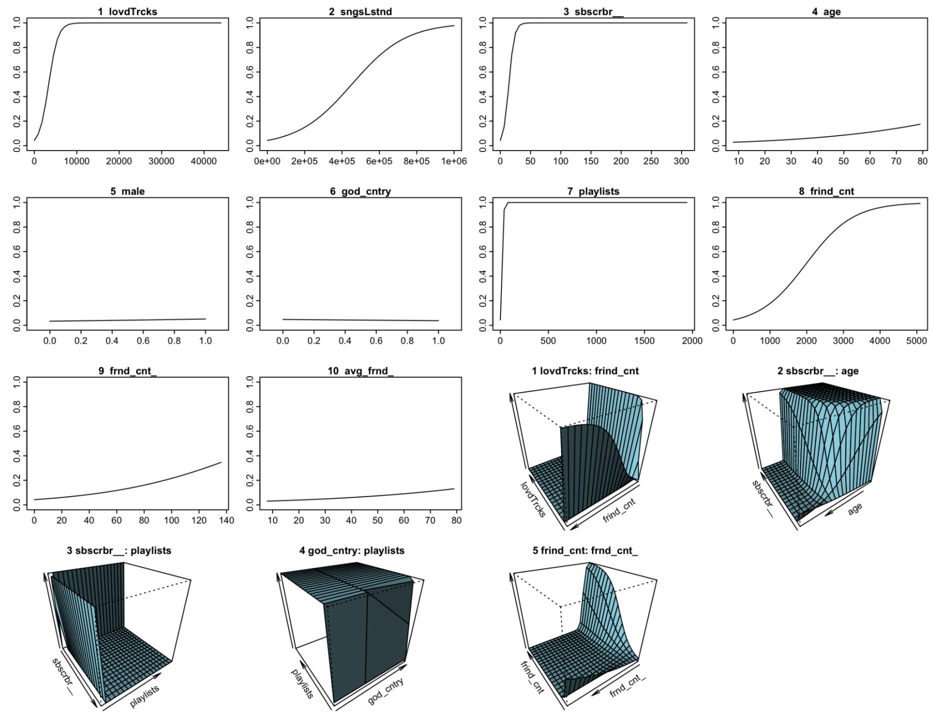


From this we can see that the most impactful features for our current users are:

- Number of friends that subscribe: Each additional subscriber friend increases the odds of adoption by approximately 48%.
- Having playlists: Each playlist is associated with a 24.02% increase in the odds of adoption.
- Being male: Being male increases the odds of adoption by approximately 56.75% compared to being female.
- Age: Each additional year increases the odds of adoption by approximately 2.85%.
- Good country:playlists (interaction): The interaction between being from a "good_country" and having playlists decreases the odds of adoption by approximately 16.76% compared to the baseline.
- Good country (US, UK, or Germany): Being from the US, UK, or Germany decreases the odds of adoption by approximately 22.15% compared to other countries.

Appendix 2.2.3

Partial dependent graphs (plots of a model's response over a range of predictor values (the model surface)) also show this:



Appendix 2.2.3

Logistic regression simulation with a sample user X1:

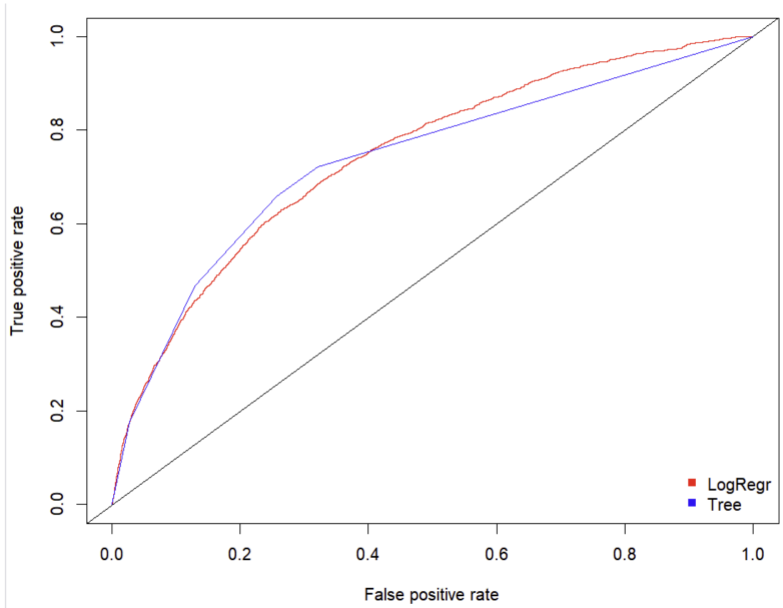
rn	Estimate	Std. Error	z value	Pr(>|z|)	meandata	sddata	userdata	X1		score		userdata - mean	score	rn
(Intercept)	-4.555650582	0.092283348	-49.36590076	0	1	0		1		-4.55565058		0	-2.7920749	(Intercept)
lovedTracks	0.000927472	5.75E-05	16.13012315	1.57E-58	78.0015822	305.4294013		348		0.32276013		269.9984178	0.25041587	# of lovedTracks
songsListened	6.87E-06	4.94E-07	13.92524673	4.45E-44	12934.6135	25483.66318		8414		0.057838328		-4520.613492	-3.11E-02	# of songsListened
subscriber_friend_cnt	0.392064011	0.033055649	11.86072662	1.89E-32	0.3366994	2.31357166		1		0.392064011		0.663300603	0.2600563	# of friends that subscribe
age	0.028079964	0.003166307	8.868364702	7.42E-19	24.3744319	4.942647433		24.3868581		0.684782087		0.012426205	3.49E-04	age
male	0.449501594	0.044397076	10.12457647	4.30E-24	0.62367736	0.38628242		0		0		-0.623677357	-0.280344	male (1 is male, 0 is female)
good_country	-0.25036581	0.045224474	-5.536068997	3.09E-08	0.36692536	0.383664806		1		-0.25036581		0.633074643	-1.59E-01	good_country (1 is US, UK, or Germany)
playlists	0.21530428	0.018740093	11.48896554	1.50E-30	0.5449765	7.822435253		1		0.21530428		0.4550235	0.09796851	# of playlists
friend_cnt	0.001663724	0.000797435	2.086345037	0.03694738	12.3498069	49.11935009		20		0.033274478		7.650193122	1.27E-02	# of friends
friend_country_cnt	0.018366538	0.004404993	4.169481428	3.05E-05	2.80228644	5.022880578		14		0.257131527		11.19771356	0.20566323	# of diff countries friends are from
avg_friend_age	0.021999171	0.003265971	6.735875111	1.63E-11	24.5945253	5.118339921		30.2857143		0.666260609		5.69118903	1.25E-01	avg age of friends
subscriber_friend_cnt:age	-0.007090458	0.001003048	-7.068909697	1.56E-12	8.78296001	82.50322883		24.3868581		-0.172914		15.60389806	-0.1106388	subscriber_friend_cnt:age (interaction)
good_country:playlists	-0.183427352	0.019772516	-9.276884622	1.75E-20	0.19594076	1.339113632		1		-0.18342735		0.804059236	-1.47E-01	good_country:playlists (interaction)
subscriber_friend_cnt:playlists	-0.016037707	0.001871609	-8.568941451	1.04E-17	0.39525649	10.43271842		1		-0.01603771		0.604743512	-0.0096987	subscriber_friend_cnt:playlists (interaction)
lovedTracks:friend_cnt	-3.83E-06	5.34E-07	-7.168607616	7.58E-13	3532.89365	61910.93897		6960		-0.0266452		3427.106349	-1.31E-02	lovedTracks:friend_cnt (interaction)
friend_cnt:friend_country_cnt	-6.91E-05	8.54E-06	-8.092023552	5.87E-16	215.958072	4106.724774		280		-0.01935858		64.04192843	-0.0044277	friend_cnt:friend_country_cnt (interaction)
									score=	-2.59498378		score=	-2.5949838	
									prob=	6.95%		prob=	6.95%	

The bars represent why the user is more likely or less likely to adopt a premium subscription.

Appendix 3

Appendix 3.1

AUC of Decision Tree model and Logistic Regression model:



Appendix 3.2

Actual AUC(Area Under the Curve) in numbers:

```
AUC for decision tree model: 0.7358563
AUC for logistic regression model: 0.7446576
```

Appendix 3.3

Confusion matrix of Decision Tree and Logistic Regression:

[DT]

```
$confmatrix
         trueclass
predclass    0    1
        0 8600  406
        1 1322  324

$accuracy
[1] 0.8377769

$truepos
[1] 0.4438356

$precision
[1] 0.1968408

$trueneg
[1] 0.8667607

$lift
[1] 2.872258
```

[LR]

```
$confmatrix
         trueclass
predclass    0    1
        0 9597  600
        1  325  130

$accuracy
[1] 0.9131618

$truepos
[1] 0.1780822

$precision
[1] 0.2857143

$trueneg
[1] 0.9672445

$lift
[1] 3.175697
```

Appendix 4

Appendix 4.1 Model Matrix

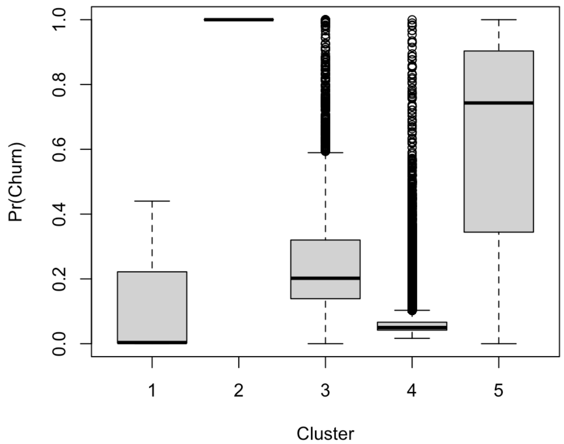
```
  (Intercept) lovedTracks songsListened subscriber_friend_cnt      age      male good_country playlists friend_cnt
1           1         348          8414                     1 24.38686 0.0000000    1.0000000         1         20
2           1           0          1943                     0 24.38686 0.6232952    0.3678186         0          3
3           1         194          9687                     0 22.00000 0.0000000    1.0000000         1          8
4           1          12         26863                     0 31.00000 0.0000000    0.0000000         0          0
5           1           0           187                     0 24.38686 0.6232952    0.3678186         0          1
6           1           0             0                     0 35.00000 0.0000000    0.0000000         0          2
  friend_country_cnt avg_friend_age subscriber_friend_cntXage good_countryXplaylists subscriber_friend_cntXplaylists
1                 14       30.28571                  24.38686                      1                               1
2                  1       30.50000                   0.00000                      0                               0
3                  1       22.57143                   0.00000                      1                               0
4                  0       24.60980                   0.00000                      0                               0
5                  1       24.60980                   0.00000                      0                               0
6                  2       28.00000                   0.00000                      0                               0
  lovedTracksXfriend_cnt friend_cntXfriend_country_cnt
1                   6960                           280
2                      0                             3
3                   1552                             8
4                      0                             0
5                      0                             1
6                      0                             4
```

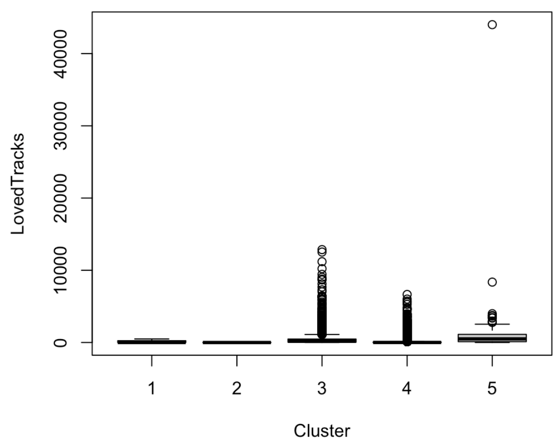
Appendix 4.2 Weighing Scheme

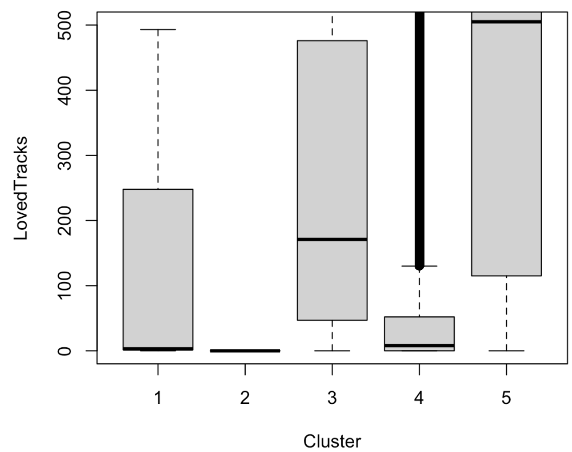
weighing user data with logistic regression coefficients

```
xmodeldata[1,1:4]
          (Intercept)           lovedTracks         songsListened subscriber_friend_cnt
                    1                   348                  8414                     1
parm[1:4]
          (Intercept)           lovedTracks         songsListened subscriber_friend_cnt
        -4.555651e+00          9.274716e-04          6.874058e-06          3.920640e-01
wuserdata[1,1:4]
          (Intercept)           lovedTracks         songsListened subscriber_friend_cnt
          -4.55565058            0.32276013            0.05783833            0.39206401
```

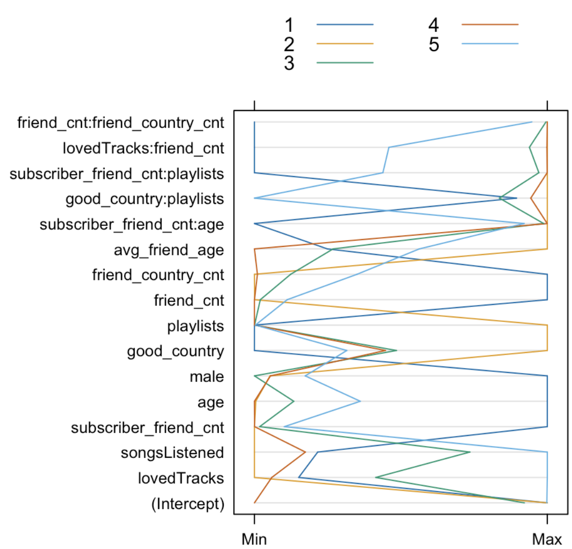
Appendix 4.2.1 Boxplot of Clusters

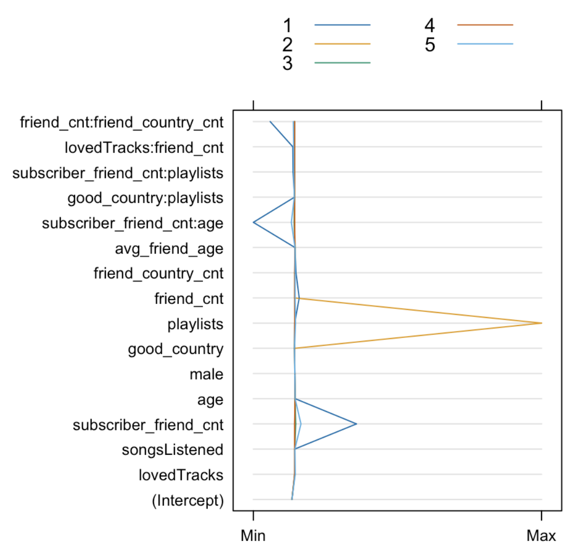






Appendix 4.2.2 Parallel Plot





Appendix 4.2.3 Parallel Plot Summary

1	friend_cnt:friend_country_cnt		low
	lovedTracks:friend_cnt		low
	subscriber_friend_cnt:playlists		low
	good_country:playlists		high
	subscriber_friend_cnt:age		low
	ave_friend_age	average age of the friends	Low to mod
	friend_country_cnt	Number of different countries this user's friends are from	high
	friend_cnt	number of friends	high
	playlists	number of playlists made till the current period	low
	good_country	= 1 if from US, UK or Germany, otherwise rest of the world	low
	male	if 1 then male else female	high
	age		high
	subscriber_friend_cnt	number of friends who are premium subscribers	high
	songsListened	cumulative number of songs listened til the beginning of the current period	Low to mod
	lovedTracks	number of tracks loved at the start of CURRENT	high

2	friend_cnt:friend_country_cnt		high
	lovedTracks:friend_cnt		high
	subscriber_friend_cnt:playlists		high
	good_country:playlists		high
	subscriber_friend_cnt:age		high

	ave_friend_age	average age of the friends	high
	friend_country_cnt	Number of different countries this user's friends are from	low
	friend_cnt	number of friends	low
	playlists	number of playlists made till the current period	high
	good_country	= 1 if from US, UK or Germany, otherwise rest of the world	high
	male	if 1 then male else female	low
	age		low
	subscriber_friend_cnt	number of friends who are premium subscribers	low
	songsListened	cumulative number of songs listened til the beginning of the current period	low
	lovedTracks	number of tracks loved at the start of CURRENT	low

3	friend_cnt:friend_country_cnt		high
	lovedTracks:friend_cnt		high
	subscriber_friend_cnt:playlists		high
	good_country:playlists		High to mod
	subscriber_friend_cnt:age		high
	ave_friend_age	average age of the friends	Low to mod
	friend_country_cnt	Number of different countries this user's friends are from	low
	friend_cnt	number of friends	low
	playlists	number of playlists made till the current period	low

	good_country	= 1 if from US, UK or Germany, otherwise rest of the world	mid
	male	if 1 then male else female	low
	age		Low to mod
	subscriber_friend_cnt	number of friends who are premium subscribers	low
	songsListened	cumulative number of songs listened til the beginning of the current period	High to mod
	lovedTracks	number of tracks loved at the start of CURRENT	mid

4	friend_cnt:friend_country_cnt		high
	lovedTracks:friend_cnt		high
	subscriber_friend_cnt:playlists		high
	good_country:playlists		high
	subscriber_friend_cnt:age		high
	ave_friend_age	average age of the friends	low
	friend_country_cnt	Number of different countries this user's friends are from	low
	friend_cnt	number of friends	low
	playlists	number of playlists made till the current period	low
	good_country	= 1 if from US, UK or Germany, otherwise rest of the world	mid
	male	if 1 then male else female	low
	age		low
	subscriber_friend_cnt	number of friends who are premium subscribers	low

	songsListened	cumulative number of songs listened til the beginning of the current period	Low to mod
	lovedTracks	number of tracks loved at the start of CURRENT	low

5	friend_cnt:friend_country_cnt		high
	lovedTracks:friend_cnt		mid
	subscriber_friend_cnt:playlists		mid
	good_country:playlists		low
	subscriber_friend_cnt:age		high
	ave_friend_age	average age of the friends	mid
	friend_country_cnt	Number of different countries this user's friends are from	Low to mod
	friend_cnt	number of friends	low
	playlists	number of playlists made till the current period	low
	good_country	= 1 if from US, UK or Germany, otherwise rest of the world	Low to mod
	male	if 1 then male else female	Low to mod
	age		Mow to mod
	subscriber_friend_cnt	number of friends who are premium subscribers	low
	songsListened	cumulative number of songs listened til the beginning of the current period	high
	lovedTracks	number of tracks loved at the start of CURRENT	high

Appendix 4.3 Customer Profile

- **Cluster 1: The Engaged Veterans:** Engaged veterans with a high number of friends and a strong sense of loyalty. Predominantly male and older, they are active users and are connected to many premium subscribers.
- **Cluster 2: The Passive Patrons:** Passive users with older friends from diverse international backgrounds and extensive social networks. Predominantly male. They exhibit lower engagement metrics such as songs listened and loved tracks, indicating a more passive usage pattern.
- **Cluster 3: The Discerning Listeners:** Discerning listeners who prioritize quality content. Predominantly female. They enjoy immersive experiences and curated content, exhibiting a discerning approach to content creation and curation, focusing on quality over quantity.
- **Cluster 4: The Casual Explorers:** Casual explorers who enjoy the platform at their own pace, occasionally discovering new content. Predominantly female. With a balanced representation across demographics, they appreciate the platform's diverse offerings but may not be deeply involved.
- **Cluster 5: The Energetic Trendsetters:** Energetic trendsetters who lead the way in discovering and sharing new content. Predominantly female. Active users. Highly engaged across age groups, they embrace the platform's social aspects and contribute to its dynamic community.

Appendix 4.4 Personalized Strategies

1. Create curated playlists based on users' favorite artists, genres, and moods, and prominently feature them on the High Note's homepage.
2. Introduce interactive features such as quizzes, polls, and challenges related to music discovery, where users can earn rewards or unlock exclusive content by participating.
3. Enable collaborative playlist creation, allowing users to invite friends to contribute to playlists and share music recommendations.
4. Partner with popular artists to release exclusive content, such as acoustic sessions, behind-the-scenes footage, or interviews, available only to premium subscribers.
5. Offer early access to new album releases or concert tickets for premium subscribers, creating a sense of exclusivity and value.
6. Launch targeted email campaigns or push notifications to passive users, highlighting personalized recommendations, new releases from favorite artists, or limited-time offers to upgrade to premium.
7. Create social media campaigns featuring user-generated content, encouraging users to share their favorite playlists or music discoveries using branded hashtags.
8. Solicit feedback from passive users through surveys or feedback forms to understand their pain points and preferences, and use this insight to continuously refine and enhance the user experience.

Appendix 5

Appendix 5.1

Since Q4 2015, the number of music streaming subscribers has risen from **68 million** to **616.2 million** in Q2 2022 - an increase of **almost 10x**.