Case: Freemium

Team L: Sherine George George, Sherry Li, Wendy Kuo, Yuchen Zhu, Enrique Garcia

Date: February 12th, 2024

Format: Written Memo

Attn: Lisa Peschke

Having been assigned the task to analyze the data compiled by Eaton Jenner and his team, we have set out to find actionable patterns that can help High Note **convert more free users to paid premium subscriptions.** The starting point of this analysis is to understand the status quo for High Note:

- High Note enjoys the short-term advantages of being an early entry to the market.
- There is a need to balance the benefits of having an ad-subsidized free account versus the profitability multiple that paid subscriptions represent for the company.
- The effectiveness of marketing campaigns to boost paid subscriptions have stagnated, and new strategies are required to convert existing free-users to paid-users.

With these considerations in mind, our intention for this analysis has been to prioritize *social features* over traditional *demographics and account/site activity* features, to better predict what are the strings that pull users to convert from free-accounts to paid premium-subscriptions. From a data-exploration and selection perspective, this means establishing the variable of *Adopter* as the target *dependent-variable* in our models, followed by an analysis of different independent variables that can help us infer the best strategies that can help increase that adoption conversion. In this memo that you have requested, we will discuss the model we selected for this analysis and the reasons behind that selection, we will then describe how we executed this model and the different parameters we considered, we will highlight our findings, and we conclude by proposing a set of recommendations on how you can implement these findings into actionable strategies to convert free-users into paid premium-subscriptions.

Model Selection

In selecting the model to be used for our analysis, we shortlisted the Decision Tree (DT) and the Logistic Regression (LR) models as both models are interpretable in the sense that we can distinguish the impact of the decision variables. Both models have been cross-validated with a pruned DT model and a stepwise LR model. Comparing the two models, we see that DT has a lower accuracy and precision of 88.0% and 24.2% respectively, as compared to LR's 91.6% and 29.7%. The full breakdown of the confusion matrix can be seen in Exhibit 1. However, DT exhibits better trade-off between True Positive Rate and False Positive Rate, as can be seen in the comparison of the area under their ROC Plot (AUC), which is attached in Exhibit 2. However, we note that the above metrics for the two models do not differ greatly, showing that both are equally justifiable statistically. Despite that, we decide to choose LR as our preferred model because it is more interpretable than DT, for we are able to determine the specific impact of different decision variables on the probability of adoption, which will aid us greatly in selecting relevant features for target marketing purposes.

Beyond LR, we also decided to consider adopting K-Means Clustering to group our customers into different segmentations. This can better allow us to understand the demographics and social patterns of different groups, and target improving features to improve the conversion of a specific prototypical customer segment. From the R-squared graph in Exhibit 3, we are able to observe the 'honey stick' effect especially at a k-value of 8, which signals that 8 clusters should be determined.

Results - Logistic Regression

From the logistic regression, we noticed that all variables are statistically significant at a 1% significance level. Zooming on the specific social factors, we observe that the number of subscriber friends serve as the most obvious predictor of premium subscription, with an additional subscriber friend increasing adoption rate by 5.22%. This is followed by the number of playlists, where one additional playlist increases the probability by 2.60%. Number of loved tracks and songs listened to are relatively insignificant, as the former requires 100 songs to increase probability by 2.13%, and the latter requires 10000 songs to increase probability by 2%. More interestingly, the number of friends has a slight negative relationship with adoption rate, as 10 more friends decreases such probability by 1.14%. A full list of such data can be found in Exhibit 4.

Results - K-Means Clustering

The analysis of High Note's user data across eight clusters highlights key insights for increasing premium subscription rates. From Exhibit 5, we can see that we would love to focus on clusters 1,3, and 6 as they have the relatively larger coefficients in more than 3 social factors, meaning that any adjustments in these factors can more significantly increase conversion rate. Specifically, for cluster 1, 3, and 6, the social aspect of subscriber friend count stands strong, with an additional subscriber increasing rate of conversion by about 16%. Moreover, they are also heavily influenced by the number of playlists, and friends (albeit a negative relationship for friends), with an increase in one playlist or friend increasing probability of conversion by about 6% and -8% respectively. Lastly, cluster 1 and 3 are also heavily influenced by the number of loved tracks, with an increase in loved tracks increasing conversion rate by roughly 8%.

Recommendations and Implementations

Based on the LR analysis, our strategy will prioritize the three highest positive predictors of premium subscription: 1) the number of subscriber friends, 2) the quantity of playlists, and 3) the number of loved tracks. Additionally, drawing insights from the Clustering, we'll focus on targeting three cluster groups: cluster 1, 3, and 6 due to their high social factor coefficient and adoption rate.

Through the integration of LR and K-Means Clustering analyses, we've structured our marketing plan into four distinct aspects: product, price, place, and promotion, collectively known as the 4Ps. Each aspect will revolve around catering to premium users and leveraging the identified predictors and clusters to enhance our approach.

Product: In alignment with our strategic plan, we are prioritizing social interaction engagement to enhance the user experience. Our recommendation involves increasing premium users' interaction by introducing short-term challenges intermittently. For example, users could be prompted to share their favorite song from a specific genre in their loved tracks or participate in a 7-day challenge to curate a playlist of a particular genre in the post section where premium users can interact with others. Additionally, we are focusing on premium features to enrich the platform further. Two key features include enabling shared playlists among premium users and playing the same track queue on different devices simultaneously. This functionality allows users to collaborate on playlists and interact with their premium friends, with each user contributing songs to create a collaborative music experience.

Price: Considering our primary target clusters (1 and 6), which largely consist of young individuals ranging from their early 20s to 30s, we're excited to introduce our new friends and family plans for premium users. Currently, each premium user pays $3 per month. However, our new offering brings an exciting change. With our friends and family plan, designed to enhance affordability and inclusivity, the total cost for all users within the plan is just $10 when you have four people enrolled. This represents significant savings and makes premium access more accessible for everyone involved.

Place: In today's digital age, nearly everyone, regardless of age, owns a cell phone. That's why we're focusing our advertising efforts on the top three social media platforms: Facebook, YouTube, and

Instagram. Whether you're a teenager or a senior citizen, our ads will be accessible to the audience. With smooth licensing processes in the US, UK, and Germany and the high coefficient in variable “Good_County” as shown in Exhibit 4, we're prioritizing the inclusion of more music from these countries. As outlined in the Berne Convention, which is adhered to by the US and EU countries, authors and artists are granted copyright protection for a longer duration. This aligns with our commitment to respecting and protecting the intellectual property rights of creators in the music industry; therefore, we aim to target more users from the US and EU.

Promotion: We're introducing a new monetary reward system as part of our efforts to enhance user engagement and expand our premium user base. With this system, users who invite others to join our premium service will receive one month of premium access for free, up to a maximum of five times. In addition to this initiative, we're preparing customized campaigns tailored to different clusters of customers, based on insights from our clustering analysis. These campaigns will feature user-specific designs crafted through the use of algorithms to attract and engage each target group effectively. For example, we'll have advertisements targeting young adults with popular artists in their age bracket. Additionally, we'll highlight our friends and family plans in advertisements to target price-sensitive individuals.

—---------------- —---------------- —---------------- —---------------- —---------------- —----------------

Appendix

Exhibit 1: Confusion Matrices

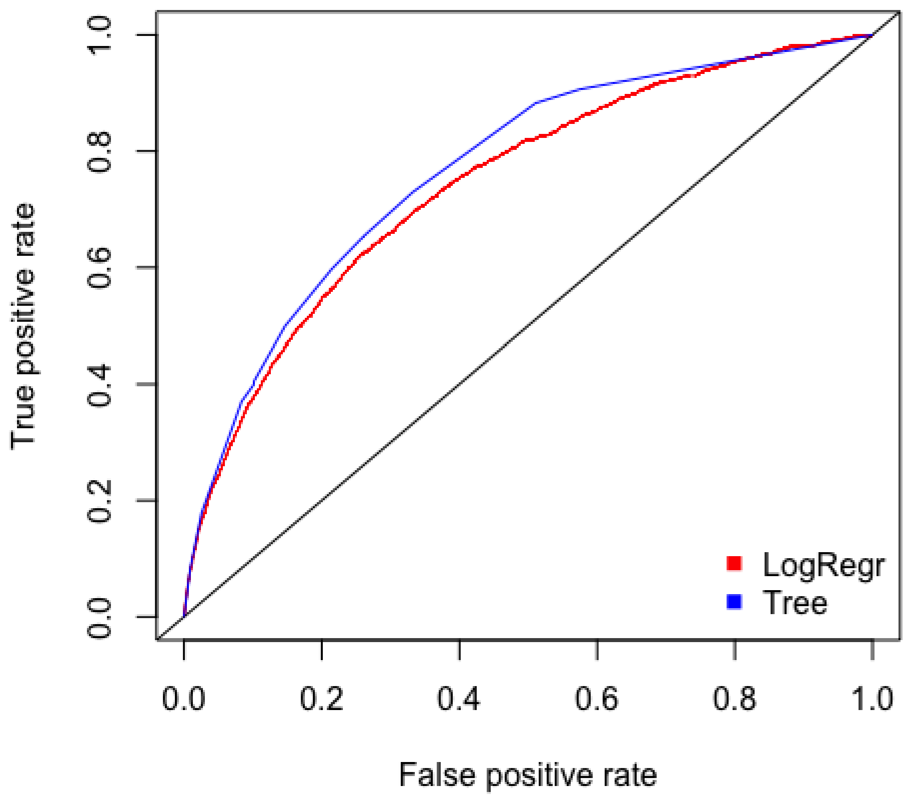
Confusion Matrix_LR		True Class	
		0	**1**
Predicted Class	**0**	29010	1764
	1	928	392

Accuracy	Precision	Recall	Specificity
91.61%	29.69%	18.18%	96.90%

Confusion Matrix_DT		True Class	
		0	**1**
Predicted Class	**0**	27445	1359
	1	2493	797

Accuracy	Precision	Recall	Specificity
87.99%	24.22%	36.97%	91.67%

Exhibit 2: ROC Curve Comparison



Area Under Curve (AUC):

- Logistic Regression: 0.74364
- Decision Tree: 0.76779

Exhibit 3: R-Squared for K-Means

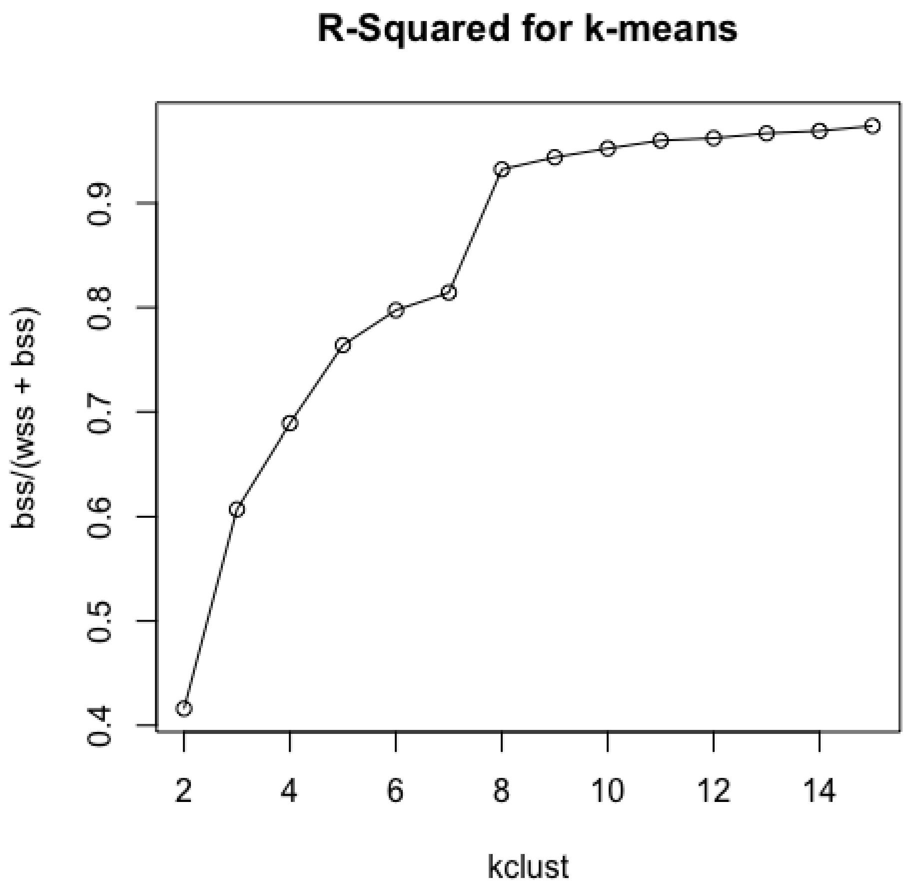


Exhibit 4: Results of Logistic Regression Analysis

Variables	Coefficients	Percentage Change/ unit
Intercept	-0.480	-
LovedTracks	8.51x10^-4	0.0213%
SongsListened	8.86x10^-6	0.0002%
Subscriber_FriendCount	0.209	5.2209%
Age	-0.0353	-0.8819%
Male	2.16x10^-4	0.0054%
Avg_Friend_Age	-0.081	-0.6871%
Good_Country	-0.028	2.6000%
Playlists	0.104	-2.0305%
Friend_Count	-0.004	-0.1114%

Exhibit 5: Results of K Means Cluster Analysis (LR Coefficients & Adoption Rate by clusters)

Cluster	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
LovedTracks	0.26305	0.35275	0.28848	0.36520	-0.04477	0.35095	0.30487	0.26409
SongsListened	0.19862	0.19104	0.26742	0.19386	0.08931	0.15733	0.29764	0.24736
Subscriber_Friend_Count	0.18360	0.68992	0.26447	0.66536	0.34999	0.4387	0.68983	0.16521
Age	0.04044	0.12265	0.01423	0.02346	0.01386	0.03970	0.03741	0.08405
Male	0.09211	-0.07367	0.18196	0.19829	-0.16962	0.18412	-0.06234	0.05547
Good_Country	-0.10175	-0.04153	-0.20436	-0.14193	-0.16196	-0.25881	-0.08283	0.02369
Playlists	0.24670	0.12959	0.17683	0.25677	0.28167	0.15110	0.11957	0.15609
Avg_Friend_Age	0.00037	-0.08482	0.11445	0.03933	-0.28029	0.18153	0.03900	-0.01490
Friend_Count	-0.12843	-0.39356	-0.11965	-0.27258	-0.11521	-0.09997	-0.28801	-0.17003
Adoption_Rate	**5.2%**	**35.3%**	**6.1%**	**19.9%**	**38.3%**	**9.3%**	**27.8%**	**5.7%**

**How to read table:*

- *Top 3 absolute values for coefficients are highlighted in green or blue, with green meaning social factors, and blue meaning demographic factors*
- *Groups with high adoption rates highlighted in yellow to distinguish patterns between factors and adoption rates.*