Arushi Jain, Anant Vashistha, Jennifer Whaley

EXECUTIVE SUMMARY

THE PROBLEM

High Note is an online music and social company that allows anyone to sign up and listen to music for free. Due to the increasing popularity of music streaming platforms over the years, consumers have a multitude of options right at their fingertips. This has left High Note struggling with a stagnant user base and not-so-impressive revenues. Even though paid subscribers are 24 times more profitable than free users, offering a freemium option allows for brand awareness and the sale of advertising space. *(Case, pg 8)* To combat this problem, High Note and Lisa Peschke, director of Marketing needs to create a strong digital marketing strategy to attract new premium users by converting their current freemium user base. To create this strategy, the team needed to dive into existing user data to understand the target listeners and aspects of the platform that are important to attract new users.

RESEARCH FINDINGS (HIGH NOTE DATA SET)

The dataset contains 107,213 observations out of which 100,000 are non-adopters and the remaining 7,213 (6.72%) users are adopters. Demographically, the two groups show stark differences - adopters are slightly older (26 vs. 24), more often male (53% vs. 39%), generally international - not US, UK, Germany (51% vs. 39%), and have been on High Note ~2 months longer (41.5 months vs. 39.4 months). Adopters are also highly social as compared to non-adopters, with more shouts (73 vs. 17), posts (17 vs. 3), friend count (28 vs. 11), and premium friend count (1.25 vs. 0.27). Lastly, adopters are more active on High Note, with more songs listened (26k vs. 12k), loved tracks (226 vs. 67), and playlists (1.15 vs. 0.49). *(Exhibit 1 & 3)*

For the users that converted to premium, just before converting, they had a higher friend count (1.5 vs. 0.4), listened to more songs (1500 vs. 520), posted more (1.3 vs. 0.06), and liked more tracks (12.3 vs. 2.6). Just after converting they increased their social behavior and engagement even further. *(Exhibit 2 & 4)*

We performed extensive analysis to identify what "causes" our users to "adopt". We started with correlation analysis, logistic regression, and CART models but found them mostly inconclusive regarding predicting and causing adoption. *(Exhibit 5, 6, 7)*. We then leveraged a causal inference technique with maximum propensity matching to deal with the biases arising from observational data. The findings showed that the following factors primarily "drove" user adoption:

- 1 new premium subscriber friend = 1.085 times increase in odds of adoption
- 1 new post = 1.125 times increase in odds of adoption within next 3 months
- 1 new playlist = 1.18 times increase in odds of adoption within next 3 months *(Exhibit 8)*

RESEARCH FINDINGS (EXTERNAL DATA)

Once we understood High Note users and what led to the adoption, we turned to look at digital marketing practices in the online music streaming industry. We compiled research on top competitors in the space, including Spotify, Apple Music, Amazon, and YouTube Music. Among these competitors, we found that Instagram was the heaviest-used platform for social digital ads. We also found that display ads were used over video ads, and there was an emphasis on desktop over mobile. The music streaming digital ad spend varied from $180k to $24M. *(Exhibit 9)*

Among our competitors, Spotify has seen the most success in the growth of users and conversion rates. To attract users, Spotify invests in advertisements that rely on pop culture references, song lyrics, and humor to attract younger listeners. Spotify also invests in campaigns such as "#SpotifyWrapped," which allows users to post their most listened-to songs and artists from the year. This creates a fear of missing out among users on social media that are not on Spotify and competition among users that increases user engagement. In 2020, this campaign increased app downloads by 21% in one week. The process of new sign-ups is also made easy for quick conversion, with a clean home page, promotion of freemium first, and pulling personal information from Facebook. This reduction of user friction to register with one click leads to a 20% increase in signups. On the premium user side, as of Q1 2020, Spotify had a conversion rate of 46.2%. This is largely driven by Spotify's strategy to create a slightly unfriendly experience for freemium users. This includes audio advertisements between songs that are often repetitive to disrupt the user experience. There are also restrictions on the skip button and only 30 days of offline listening. *(Source: Spotify)*

Arushi Jain, Anant Vashistha, Jennifer Whaley

RECOMMENDATIONS

As mentioned above, the goal of High Note's digital marketing strategy is to attract new premium users and increase user engagement. Our team recommends the following to achieve our objectives:

1. ***Incentivize premium subscribers to be more social with their network to drive premium user growth***
   High Note needs to incentivize their total 37k premium subscribers to discuss their subscription perks and swag with their social network. A study by the University of Minnesota found that peer influence causes more than 50% increase in the odds of buying the product due to the influence of an adopting friend (Do Friends Make You Pay? by Bapna & Umyarov). Facebook found that users who saw a social message of friends doing a trend were 0.39% more likely to do the trend than users who received an informational message with no social proof *(Hype Machine, Ch. 7 – voting)*. These proof points along with our findings above that premium subscriber friend count is one of the three most effective variables of driving adoption for High Note users, as well as the Spotify example of "#SpotifyWrapped" shows utilizing friends and the "fear of missing out" is an effective way to drive premium user growth.
   To incentivize premium users to share their experiences, High Note should introduce a **referral program** to create a viral impact. Drawing on Uber's referral program, one of the most successful in the world, when a new premium user joins with a friend's referral code both the listeners will earn a bonus. It is important to offer the benefit to the new user as well; otherwise, findings have shown that premium users feel uncomfortable spamming their friends and getting a reward for it. This strategy will be effective as friends know their network's preferences and can route invitations effectively, creating a "local network effect" *(Hype Machine, Ch. 8)*. Another strategy is to incentivize premium users to **share their "Currently Listening"** on their personal Instagram, Whatsapp or Facebook stories which will earn them loyalty points in return. These points can then later be redeemed for discounts on concert tickets and merchandise from artists.

2. ***Create more features and campaigns within the platform to increase user engagement***
   Our analysis of the High Note data set showed that posting and creating playlists are highly impactful in converting to premium within the next 3 months. We also found that overall adopters are more likely to be social and active on the platform. This means we have to push our freemium users to become more active and engaged, which will lead to a longer customer lifetime as well as a higher percentage of conversion.
   To accomplish this, we will create a feature that **suggests new friends to users based on their music preferences**. The more friends users have, the more they will be incentivized to post and read other's posts. We would also start a campaign called **"#QuestionOfTheDay"** in which High Note poses questions such as "What is the most underrated song of the year?" This will give users a reason to post and interact with one another. For playlists, we can start a program in which the most listened-to and shared playlists generated by users will be shared with the whole platform. This will create **competition among users to create popular playlists**. Another feature would be **playlist recommendation,** where High Note can recommend automatically generated playlists to users based on their music preferences, similar to Netflix recommendations.

3. ***Increase the lift of net new subscribers through digital advertising***
   As stated above, Instagram ads are by far the heaviest-used **social digital advertising** tool among music competitors. We also know that display ad are used much more than video ads. *(Source)* Creating true **lift**, the extent to which an ad changes behavior is important to increase new users *(Hype Machine, Ch. 6)*. Disregarding the N/A data in our set, we know that the average High Note user is male, 22-30 years old, and not from the US, UK, or Germany. We can leverage **search advertising** by bidding within the marketplace of digital inventory to identify users that match our typical demographics from their cookie data. We would also try **retargeting** ads, as only 2% of visitors convert on the first visit, and the click-through rate for display ads is 0.07%. However, display ads with retargeting have a 0.7% click-through rate *(S2 Slides)*.

CONCLUSION

In conclusion, Lisa Peschke needs to implement these three recommendations to ensure growth in premium membership and to keep listeners engaged. Implementation should take approximately six months for these initiatives and campaigns to be built out, marketed, and executed. Once proven successful, High Note can continue to invest in building a long-term digital marketing plan to ensure success amongst competitors.

APPENDIX

EXPLORATORY DATA ANALYSIS

Exhibit 1: Cumulative Variables, Adopters vs Non-Adopters features comparison

	Adopters					Non-Adopters				
Variables	**Mean**	**Std**	**Min**	**Median**	**Max**	**Mean**	**Std**	**Min**	**Median**	**Max**
Age	26.30	7.24	8.00	25.00	78.00	24.22	6.78	8.00	23.00	79.00
Male	0.72	0.45	0.00	1.00	1.00	0.62	0.49	0.00	1.00	1.00
Friend Count	28.38	93.42	0.00	9.00	5089.00	11.07	42.90	0.00	3.00	4957.00
Avg. Friend Age	25.85	5.59	12.00	24.71	70.00	24.52	5.73	8.00	23.16	79.00
Avg. Friend Male	0.65	0.28	0.00	0.67	1.00	0.63	0.36	0.00	0.67	1.00
Friend Country	5.38	8.05	0.00	2.00	136.00	2.60	4.63	0.00	1.00	129.00
Subscriber Friend	1.25	4.60	0.00	0.00	287.00	0.27	1.79	0.00	0.00	309.00
Songs Listened	25959.55	40439	0.00	13018.00	1000000	11919.30	23437.23	0.00	3023.00	1000000
Loved Tracks	226.13	674.96	0.00	83.00	44005.00	67.06	228.10	0.00	7.00	12861.00
Posts	16.72	247.75	0.00	0.00	15185.00	2.84	70.89	0.00	0.00	12309.00
Playlists	1.15	22.97	0.00	1.00	1943.00	0.49	1.52	0.00	0.00	261.00
Shouts	73.45	915.28	0.00	3.00	65872.00	17.14	116.57	0.00	2.00	8694.00
Tenure	41.51	19.76	0.00	40.00	111.00	39.41	19.24	0.00	38.00	111.00
Good Country	0.32	0.46	0.00	0.00	1.00	0.37	0.48	0.00	0.00	1.00

As per the above table, adopters category has 72% males against 62% in non-adopters category. It can be inferred that all the social related variables except good country are significantly higher for the adopters' category as compared to the non-adopters category.

Exhibit 2: Delta1 Variables, Adopters vs Non-Adopters Delta 1 features comparison

	Adopters					Non-Adopters				
Variable	**Mean**	**Std**	**Min**	**Median**	**Max**	**Mean**	**Std**	**Min**	**Median**	**Max**
Friend Cnt	1.48	10.36	-486	0.00	418.00	0.40	4.78	-351.00	0.00	521.00
Avg Friend Age	0.20	0.80	-13.67	0.18	12.67	0.23	0.70	-50.00	0.13	19.50
Avg Friend Male	0.00	0.07	-0.67	0.00	1.00	0.00	0.05	-1.00	0.00	1.00
Friend Country	0.21	1.22	-40.00	0.00	25.00	0.05	0.63	-52.00	0.00	30.00
Subscriber Friend	-0.01	1.05	-18.00	0.00	43.00	-0.01	0.41	-23.00	0.00	17.00
Songs Listened	1519.09	3403.56	-93584	428.00	64958.00	520.08	1995.01	-91198.0	0.00	129096
Loved Tracks	12.34	50.47	-1040	0.00	1143.00	2.61	21.08	-532.00	0.00	1338.00
Posts	1.30	49.28	-1.00	0.00	3391.00	0.06	4.60	-6.00	0.00	1185.00
Playlists	0.02	0.28	-8.00	0.00	9.00	0.00	0.38	-6.00	0.00	116.00
Shouts	5.05	132.85	-591	0.00	10022.00	0.48	13.36	-953.00	0.00	2459.00
Good Country	0.00	0.06	-1.00	0.00	1.00	0.00	0.03	-1.00	0.00	1.00

Delta1 variables can best represent the behaviour of users before adoption. The features such as friend country, songs listened, loved tracks, posts and shouts have significantly higher mean values for adopters than non-adopters. It further substantiates that highly active users have more propensity for conversion to the premium category.

Exhibit 3: Demographic Variables Distribution

The user's persona information, such as age, gender, and country, is specifically used by the organization to understand its target audience and design a marketing strategy. Data reveals that customers of the age group "22-30", Male and from the non-good country are more susceptible to purchasing a premium plan.

Adopter vs Non-Adopters age buckets comparison

Category vs. Age Buckets	0-18	18-22	22-30	30+	NA
Non-Adopters	4%	16%	23%	8%	48%
Adopters	2%	14%	32%	14%	38%
Overall	4%	16%	24%	9%	47%

Adopters vs Non-Adopters gender comparison

Category vs. Gender	Female	Male	NA
Non-Adopters	24%	39%	37%
Adopters	21%	53%	26%
Overall	24%	40%	36%

Adopters vs Non-Adopters good country comparison

Category vs. Good Country	0	1	NA
Non-Adopters	39%	23%	37%
Adopters	51%	23%	26%
Overall	40%	23%	37%

Exhibit 4: Adopters: Pre vs Post Engagement Features Comparison

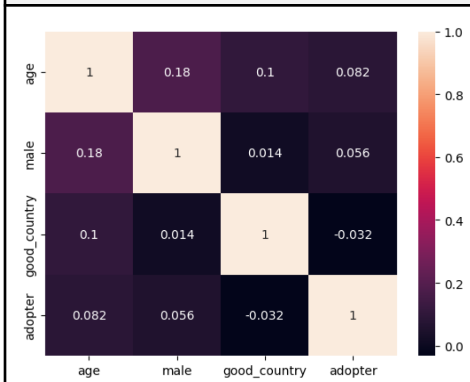
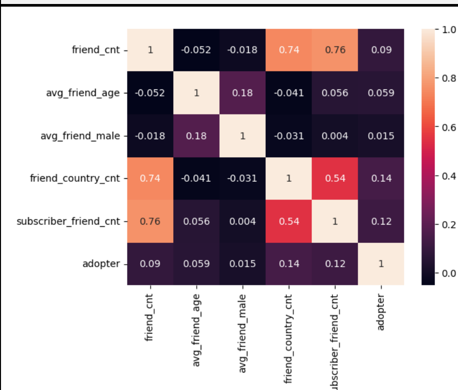
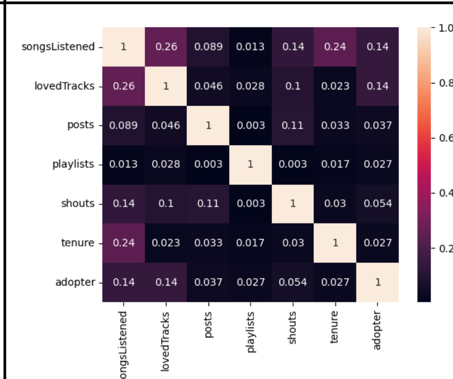
The user pre vs post-adoption behavior can be studied to understand their activity and keep them engaged on a social network. This comparison and effective strategies could help retain the existing user base.

KPI	Friend Cnt	Friend Country Cnt	Subscriber Friend Cnt	Songs Listened	Loved Tracks	Posts	Playlists	Good Country
Delta 1	10,665	1,483	(88)	10,939,002	88,899	9,351	114	(1)
Delta 2	13,288	1,764	(316)	13,956,904	126,970	10,428	(257)	4

The user's friend count, songs listened, loved tracks and posts have significantly increased post-adoption as compared to pre-adoption.

Exhibit 5: Correlation Study for Demographics, Social, and Engagement Features

The correlation study between the demographics, social and engagement features can be leveraged for modeling user behavior, but causal analysis has been carried out further to understand the causation.

Demographics	Social	Engagement
		

The demographics variables including good country, age, and male don't show high correlation with the adopter variable.	The subscriber friend count is highly correlated with the friend count and friend country count variable. But, any social features don't correlate highly with the adopter variable.	Shouts have a high correlation (0.54) with adopter variable whereas the remaining engagement features don't show much correlation among themselves.

PREDICTIVE MODELS TO ASSESS VARIABLE IMPORTANCE

Data Imputation

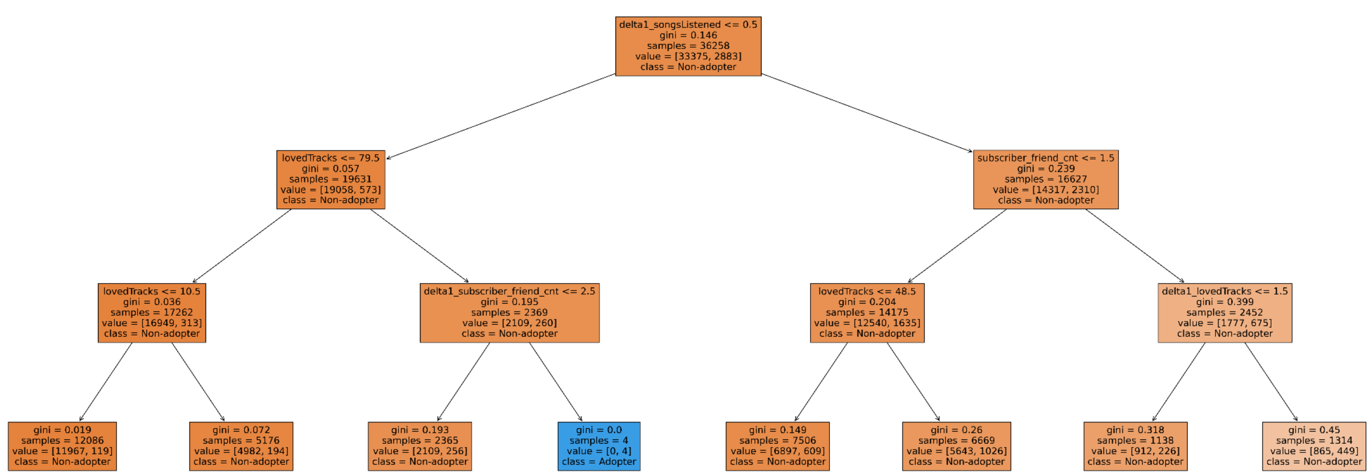
The current dataset contains missing values in some columns (e.g., age, gender, songsListened etc.) which need to be handled. We cannot directly delete rows with missing data because it will lead to huge data loss and removal of other important features. Considering the distribution across features, we replaced all missing values with zero. However, there are better strategies for data imputation like replacing missing values with mean or median finding k-nearest neighbours and use their feature values instead which can be explored in future.

Exhibit 6: Logit

	coef	std err	z	P>|z|	[0.025	0.975]
age	-0.0821	0.002	-34.732	0.000	-0.087	-0.078
male	-0.0008	0.039	-0.021	0.983	-0.077	0.075
good_country	-0.4686	0.042	-11.030	0.000	-0.552	-0.385
friend_country_cnt	-0.0390	0.004	-9.261	0.000	-0.047	-0.031
subscriber_friend_cnt	0.2099	0.014	15.085	0.000	0.183	0.237
tenure	-0.0133	0.001	-11.823	0.000	-0.015	-0.011
songsListened	8.443e-06	6.81e-07	12.397	0.000	7.11e-06	9.78e-06
lovedTracks	0.0007	6.47e-05	11.414	0.000	0.001	0.001
posts	0.0001	0.000	1.302	0.193	-7.41e-05	0.000
playlists	0.1384	0.016	8.472	0.000	0.106	0.170
shouts	4.838e-06	8.59e-05	0.056	0.955	-0.000	0.000
delta1_friend_country_cnt	0.0725	0.022	3.369	0.001	0.030	0.115
delta1_subscriber_friend_cnt	-0.0648	0.030	-2.148	0.032	-0.124	-0.006
delta1_songsListened	8.66e-06	7.65e-06	1.132	0.258	-6.34e-06	2.37e-05
delta1_lovedTracks	0.0002	0.001	0.298	0.765	-0.001	0.001
delta1_posts	-0.0030	0.006	-0.460	0.645	-0.016	0.010
delta1_playlists	0.0568	0.120	0.472	0.637	-0.179	0.293
delta1_shouts	0.0004	0.001	0.620	0.536	-0.001	0.001

We ran logistic regression using various demographic, social, and engagement variables to predict adoption and understand their importance and significance. We observed that good_country and subscriber_friend_cnt are important variables with low p-values. delta1_playlists also seems important; however, its p-value is high, so we cannot make any conclusive remarks about its significance in predicting (let alone causing) adoption.

Exhibit 7: CART



We also ran CART to model the adopter variable. We see those engagement variables such as delta1_songsListened, loved tracks, and the social variable subscriber_friend_cnt appear in the top splits and can be essential precursors to adoption. However, we still don't know if these variables "cause" adoption.

CAUSAL INFERENCE

Exhibit 8

Considering we only have 9 months of observational data for an online social network, we cannot do true causal analysis, which requires randomized control trials. To address this, we use a matching technique called maximum propensity matching to estimate the causal effect of binary treatments (social treatments or engagement treatments) on the outcome (adoption) while controlling the non-treatment variables (covariates or confounders).

Maximum Propensity Matching produces covariate balance (approximately equal distributions of covariates in treatment and control groups), just as in a randomized experiment. Thus, our predictive model (logistic regression) is more robust in predicting the treatment effect of a given variable since our samples are perfectly balanced due to propensity matching.

Steps for Causal Inference on data explained:

1. **Select binarized treatment and covariates:** Here, we select social variables (friend_cnt, friend_country_cnt, and subscriber_friend_cnt) and engagement variables (songs listened, loved tracks, posts, playlists, shouts, tenure) as well as their respective delta1 data as treatment variables one by one. We binarize the treatment variables using certain cut-off thresholds for each treatment. All other variables (demographic, social, engagement, delta1_social, delta1_engagement) are considered confounding variables.
   For example, if we consider **subscriber_friend_cnt** as a treatment variable, we first binarize it by using a value of zero for subscriber_friend_cnt as the threshold for binarized treatment. This means anyone with more than 0 subscriber friends is in the treatment group, and anyone with less than or equal to 0 subscriber friends is in the control group. The variables other than subscriber_friend_cnt are used as confounding factors.
2. **Check Initial Data Imbalance:** Here we check the imbalance in our covariates across the treatment and control groups.
   For subscriber_friend_cnt, we have the following initial data imbalance:

Summary Statistics of Treatment vs Control group before and after matching

Variables	Data imbalance before matching			Data imbalance after matching		
	Means Treatment	**Means Control**	**Std. mean difference**	**Means Treatment**	**Means Control**	**Std. mean difference**
Distance	0.45	0.09	1.07	0.45	0.26	0.56
Age	17.86	11.92	0.45	17.86	18.42	-0.04
Male	0.51	0.37	0.27	0.51	0.53	-0.03
Friend Cnt	46.5	6.12	0.35	46.50	18.34	0.24
Avg Friend Age	25.39	18.62	0.99	25.39	26.52	-0.16
Avg Friend Male	0.64	0.51	0.49	0.64	0.66	-0.08
Friend Country Cnt	8.36	1.79	0.67	8.36	4.56	0.39
Loved Tracks	206.06	54.92	0.25	206.06	128.12	0.13
Songs Listened	29101.55	9972.80	0.45	29101.55	23641	0.13
Posts	17.43	1.34	0.06	17.43	5.16	0.05
Playlists	0.76	0.48	0.08	0.76	0.74	0.004
Shouts	83.28	9.29	0.11	83.28	30.80	0.08
Tenure	44.19	38.70	0.27	44.19	44.92	-0.03
Good Country	0.29	0.22	0.16	0.29	0.32	-0.06
Delta 1 Variables						
Friend Cnt	2.15	0.17	0.16	2.15	0.59	0.12
Avg Friend Age	0.21	0.17	0.05	0.21	0.20	0.01
Avg Friend Male	0.0004	-0.0002	0.01	0.0004	0.0004	0.001
Friend Country Cnt	0.25	0.02	0.17	0.25	0.10	0.11

Subscriber Friends	0.09	-0.03	0.11	0.09	-0.04	0.12
Songs Listened	1487.31	421.19	0.28	1487.31	1092.44	0.10
Loved Tracks	9.53	2.14	0.16	9.53	5.44	0.09
Posts	0.86	0.01	0.02	0.86	0.07	0.02
Playlists	0.01	0.002	0.01	0.01	0.004	0.01
Shouts	3.80	0.22	0.03	3.80	0.93	0.03
Good Country	0.0005	0.0002	0.007	0.29	0.32	-0.06

We can notice severe data imbalance by looking at the Means of the respective covariates in the treatment and the control group as well as the Std. Mean Diff. Additionally, we can see the imbalance in the number of treatment and control units

Number of treatment vs control units before & after matching:

	Before Matching		After Matching	
	Control	Treated	Control	Treated
All	91009	16204	91009	16204
Matched	91009	16204	16204	16204
Unmatched	0	0	74805	0
Discarded	0	0	0	0

3. **Perform Matching:** Here we use Maximum Propensity Matching using Nearest Neighbour Method with a 1:1 ratio and evaluate the quality of the matches by rechecking the data imbalance. After matching, we notice (refer to Table 1.8) that the data imbalance has significantly reduced in the treatment and the control group. Additionally, we can see (refer to Table 1.9) that the number of treatment and control units is now equal.
4. **Run Logistic Model:** Train a glm and check the significance of the treatment variable. If the treatment variable is significant, we can make statements about the true effect of the treatment variable.

Logistic Model output for subscriber friend count treatment variable

Treatment variable	Estimate	Std.Error	Z value	Pr(>z)
Subscriber friend cnt	0.082	0.003	21.73	9.44*e^-105

Here, we see that subscriber_friend_cnt is a significant variable (p-value < 0.05); hence, we can say that a unit increase in subscriber_friend_cnt increases the odds of adoption by $e^{0.082} = 1.085$ times!

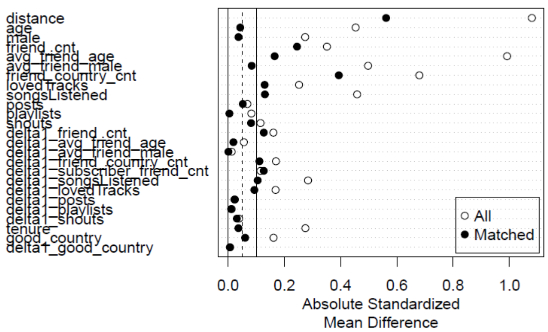
The following table shows the results for each of the treatment variables and highlights the top 3 treatment variables with maximum effect on user adoption:

Causal inference summary statistics

Treatment Type	Treatment Name	Treatment Threshold	p-value	$e^{coefficient}$
Social Variables	friend_cnt	3	2.992847e-22	1.019564
	friend_country_cnt	1	1.882267e-57	1.038259
	subscriber_friend_cnt	0	9.442021e-105	1.085469
Engagement Variables	songsListened	3432	2.754403e-117	1.042902
	lovedTracks	9	0	1.068937
	posts	0	2.247645e-19	1.032733
	shouts	2	0.5971493	0.9989928
	playlists	0	8.82894e-40	1.023703
	tenure	38	0.006223239	0.9955386
Delta1 Social Variables	delta1_friend_cnt	1	1.231898e-45	1.062581
	delta1_friend_country_cnt	1	3.910221e-07	1.033628
	delta1_subscriber_friend_cnt	1	3.613929e-08	1.04887

Delta1 Engagement Variables	delta1_songsListened	1	7.917911e-249	1.07342
	delta1_lovedTracks	1	8.216486e-100	1.080404
	delta1_posts	1	1.243034e-06	1.125028
	delta1_shouts	1	4.39724e-16	1.047275
	delta1_playlists	1	4.513691e-12	1.180587

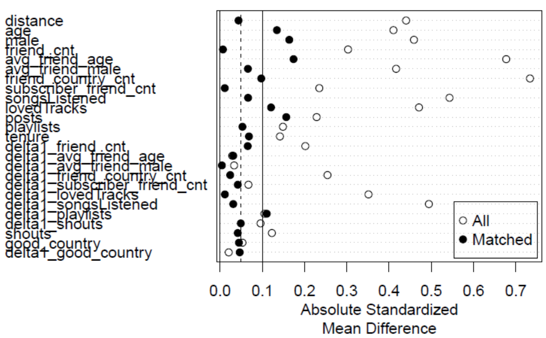
We see that odds of adoption increase just in the immediate period after increased engagement with the app (posts and playlists). We also see that there exists a social influence factor, as well as subscriber users, influencing their non-subscriber friends to adopt.



Final data imbalance after matching for our top 3 treatment variables:

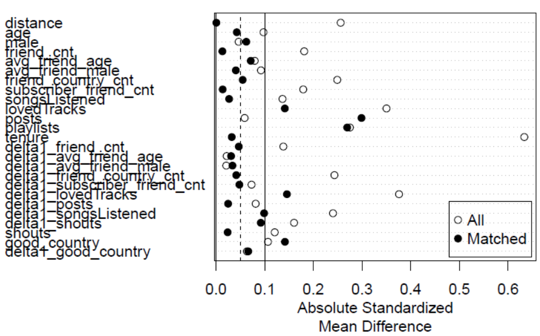
Love Plot for subscriber_friend_cnt:
Most of the variables except male, friend_cnt, and friend_country_cnt are near the acceptable standardized mean difference value of 0.1.

Hence, the data is balanced, and we can say that 1 additional subscriber_friend_cnt can increase the odds of adoption by 1.08 times.



Love Plot for delta1_posts
Most variables except age, male, avg_friend_age, and loved tracks are near the acceptable standardized mean difference value of 0.1.

Hence, the data is balanced, and we can say that 1 additional delta1_post can increase the odds of adoption by 1.12 times.



Love Plot for delta1_playlists
Most variables except loved tracks, posts, playlists, shouts, and delta1_lovedTracks are near the acceptable standardized mean difference value of 0.1.
Hence, the data is balanced, and we can say that 1 additional delta1_post can increase the odds of adoption by 1.12 times.

EXTERNAL RESEARCH

Exhibit 9: Online Music Digital Marketing *(source: The Music Streaming Wars)*

	Subscribers	**Spend on Digital Ads**	**High Spend Categories**
Spotify	172m	$23.8m	59% Instagram, 98% of TV on Hulu
Apple Music	79m	$180k (Total Apple spend $48.2m)	100% budget on display ads
Amazon	68m	$17.7m (Total Amazon spend $582m)	74% Instagram
YouTube Music	42m	$232k (Oct - Jan)	99% Instagram (heavy LA)