High Note: an analysis on conversion from Free to Premium

Adopter:

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
ID	1	3527	42064.00	1018.30	42064.00	42064.00	1307.65	40301	43827	3526	0.00	-1.20	17.15
age	2	3527	25.98	6.84	24.00	25.05	4.45	8	73	65	1.68	4.39	0.12
male	3	3527	0.73	0.44	1.00	0.79	0.00	0	1	1	-1.03	-0.94	0.01
friend_cnt	4	3527	39.73	117.27	16.00	23.69	17.79	1	5089	5088	26.04	1013.79	1.97
avg_friend_age	5	3527	25.44	5.21	24.36	24.83	3.91	12	62	50	1.68	5.05	0.09
avg_friend_male	6	3527	0.64	0.25	0.67	0.65	0.25	0	1	1	-0.54	-0.05	0.00
friend_country_cnt	7	3527	7.19	8.86	4.00	5.36	4.45	0	136	136	3.61	24.53	0.15
subscriber_friend_cnt	8	3527	1.64	5.85	0.00	0.84	0.00	0	287	287	34.05	1609.52	0.10
songsListened	9	3527	33758.04	43592.73	20908.00	25811.69	23276.82	0	817290	817290	4.71	46.64	734.03
lovedTracks	10	3527	264.34	491.43	108.00	161.68	140.85	0	10220	10220	6.52	80.96	8.27
posts	11	3527	21.20	221.99	0.00	1.44	0.00	0	8506	8506	26.52	852.38	3.74
playlists	12	3527	0.90	2.56	1.00	0.59	1.48	0	118	118	28.84	1244.31	0.04
shouts	13	3527	99.44	1156.07	9.00	23.89	11.86	0	65872	65872	52.52	2969.09	19.47
adopter	14	3527	1.00	0.00	1.00	1.00	0.00	1	1	0	NaN	NaN	0.00
tenure	15	3527	45.58	20.04	46.00	45.60	20.76	0	111	111	0.02	-0.62	0.34
good_country	16	3527	0.29	0.45	0.00	0.23	0.00	0	1	1	0.94	-1.12	0.01

Non-adopter:

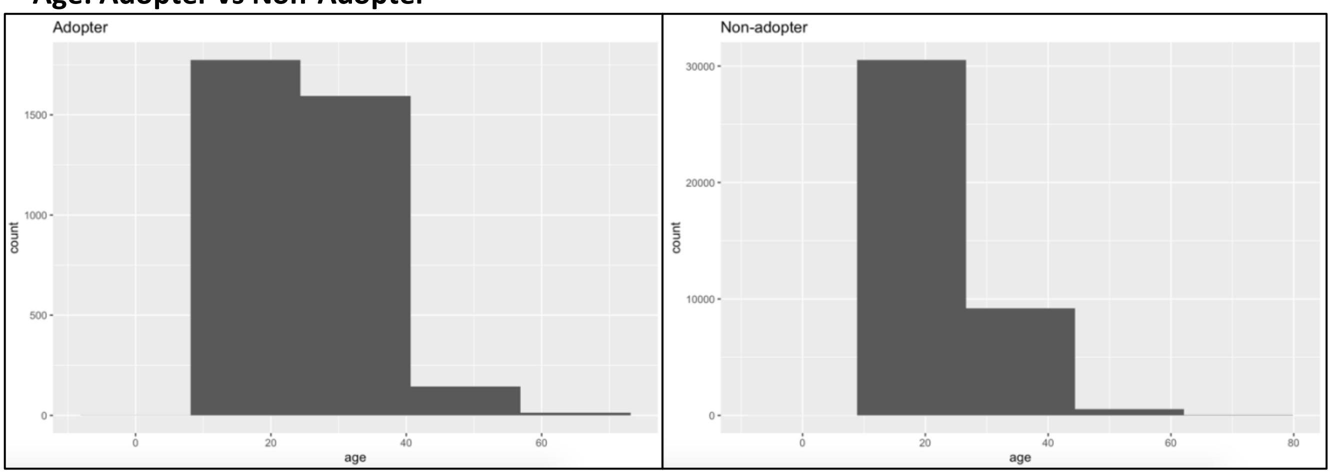
	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
ID	1	40300	20150.50	11633.75	20150.50	20150.50	14937.19	1	40300	40299	0.00	-1.20	57.95
age	2	40300	23.95	6.37	23.00	23.09	4.45	8	79	71	1.97	6.80	0.03
male	3	40300	0.62	0.48	1.00	0.65	0.00	0	1	1	-0.50	-1.75	0.00
friend_cnt	4	40300	18.49	57.48	7.00	10.28	7.41	1	4957	4956	32.67	2087.42	0.29
avg_friend_age	5	40300	24.01	5.10	23.00	23.40	3.95	8	77	69	1.84	7.15	0.03
avg_friend_male	6	40300	0.62	0.32	0.67	0.65	0.35	0	1	1	-0.52	-0.72	0.00
friend_country_cnt	7	40300	3.96	5.76	2.00	2.66	1.48	0	129	129	4.74	38.29	0.03
subscriber_friend_cnt	8	40300	0.42	2.42	0.00	0.13	0.00	0	309	309	72.19	8024.62	0.01
songsListened	9	40300	17589.44	28416.02	7440.00	11817.64	10576.87	0	1000000	1000000	6.05	105.85	141.55
lovedTracks	10	40300	86.82	263.58	14.00	36.35	20.76	0	12522	12522	13.12	335.93	1.31
posts	11	40300	5.29	104.31	0.00	0.23	0.00	0	12309	12309	73.92	7005.34	0.52
playlists	12	40300	0.55	1.07	0.00	0.45	0.00	0	98	98	28.21	1945.28	0.01
shouts	13	40300	29.97	150.69	4.00	8.84	4.45	0	7736	7736	22.53	779.12	0.75
adopter	14	40300	0.00	0.00	0.00	0.00	0.00	0	0	0	NaN	NaN	0.00
tenure	15	40300	43.81	19.79	44.00	43.72	22.24	1	111	110	0.05	-0.70	0.10
good_country	16	40300	0.36	0.48	0.00	0.32	0.00	0	1	1	0.59	-1.65	0.00

Mean values of variables:

1. Age: Those who have chosen to convert from free accounts to premium are on average older than those who remain in the free subscription.
2. Gender: Premium users have a higher proportion of male users than female, i.e. more males converted from a free to premium account.
3. Friend count: The average number of friends an ‘adopter’ has is significantly higher (39) than the number of friends a non-adopter has (18). This may indicate that premium users are more actively using the features of High Note.
4. Average Friend Age & Average male friends: The average age of friends in both subscription models appear to be similar, as does the average proportion of male friends.
5. Friends from other countries: Those in premium subscriptions have friends from a higher variation of countries. This may again be due to the large number of friends they have compared to free users due to more activity and higher usage of High Note.

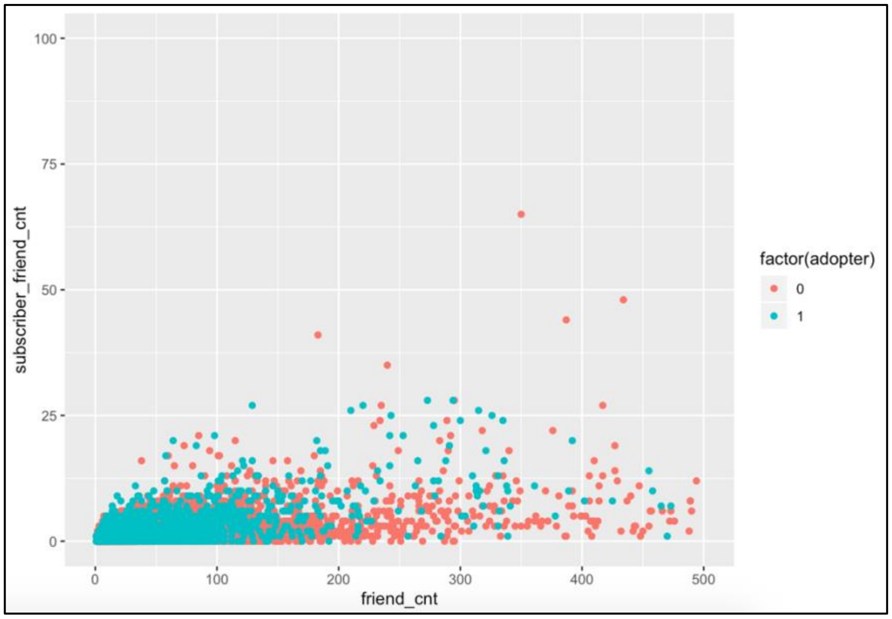
6. Subscriber Friend Count: A person who has subscribed to a premium subscription has, on average, more number of friends who have premium accounts. This implies that the likelihood of a person converting from a free account to a premium account is higher if the person has more friends that already have a premium account.
7. Songs Listened: The premium users listened to almost double the number of songs, on average, compared to free users. This may be because premium users are able to listen to songs on more devices and this added flexibility makes it easier for them to listen to songs on the platform.
8. All interaction with the platform such as *Loved Tracks, posts, playlists* and *shouts* are all higher in number among 'adopters'. The median value of *posts* is 0.00 for both Adopters and Non-Adopters – indicating that most people do not post anything on the site.

Age: Adopter vs Non-Adopter



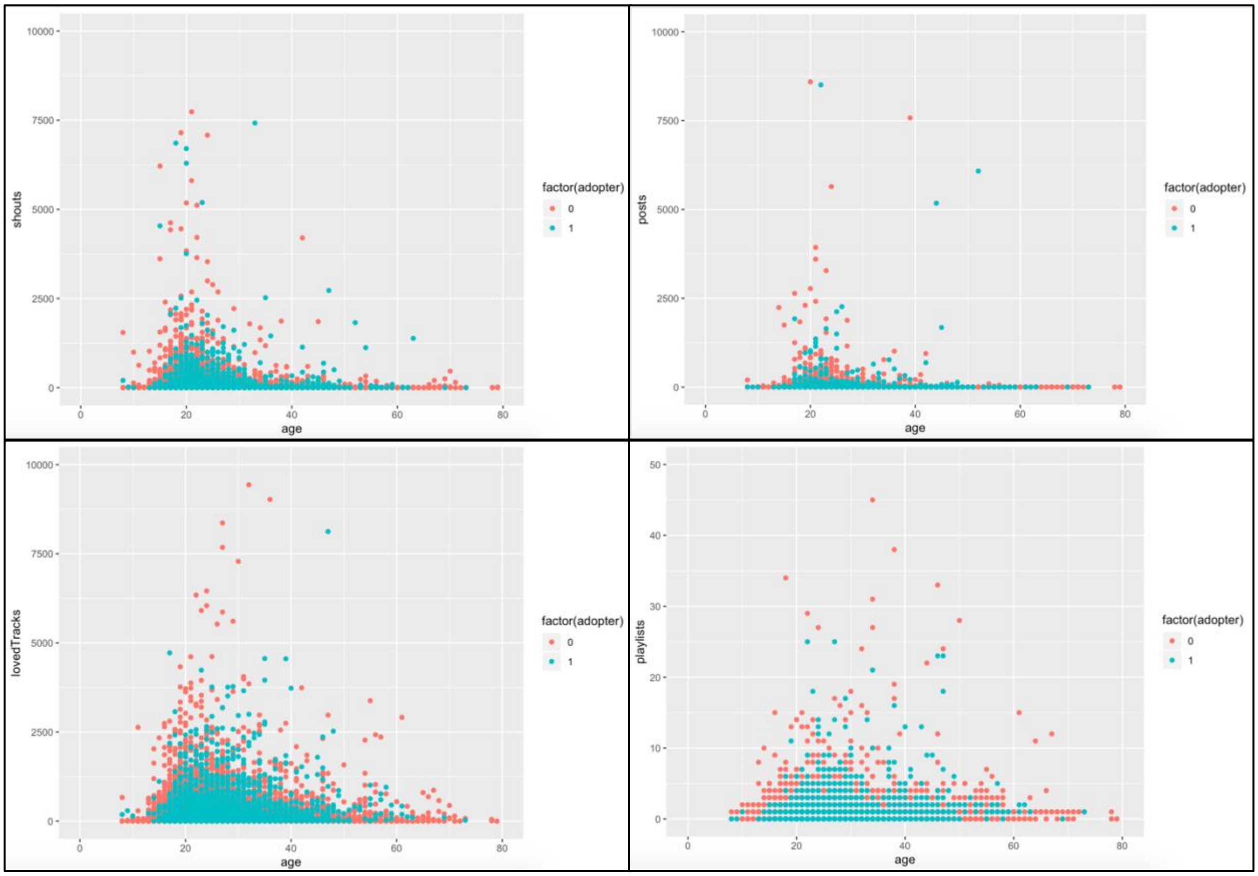
The histograms above show the distribution of age among adopters and non-adopters. The percentage of people in the 25-40 age range is significantly higher in adopters. This could imply that older and hence more financially stable people in general prefer to opt into the subscription compared to younger people.

Peer influence: How many friends subscribe to premium?



For those that haven not subscribed to the premium model – as the number of friends they have increases, the number of subscriber friends also increases, as expected. However, for those that are already premium members, the same trend follows between 200 and 300 friends, but any increase in the number of friends after this does not necessarily increase on the number of subscriber friends – in general, it is decreasing with increase in the number of friends as seen in the graph above. There is no observed relationship between friend count and subscriber friend count for adopters below 150 friends.

Engagement with the site in relation to age

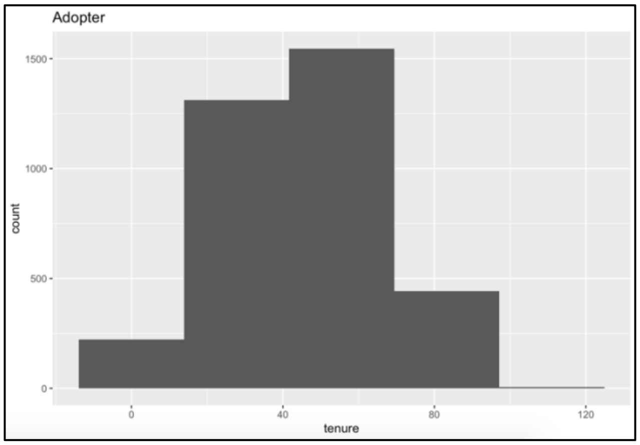


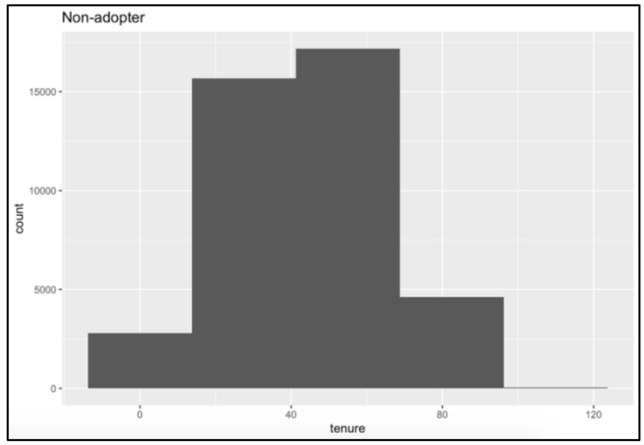
The scatter plots above show the increase in interaction with the music platform (*shouts, posts, lovedTracks* and *playlists*), in relation to age. As seen from the graph, the number of *shouts* and *lovedTracks* tends to increase sharply in the 15-25 years age range. More people in this age range have shouts and have indicated that they like a song in this age group. These interactions decrease as people get older. This trend is also visible in the number of posts people make – it tends to increase in the late teens to 25 years age range, and then begins to decrease again.

However slightly older people create playlists compared to the other above interaction – there is an increase in the number of playlists created for people between the ages of 25 and 35.

Affect of tenure on Adopter/Non-adopter

The amount of a time a person has had an account on HighNote does not change significantly between adopters and non-adopters, as seen from the graphs below. However, the percentage of people in the 20-40 months range is slightly less for Adopters – which could imply that people are more likely to subscribe once they have been using the site as a free user for some time.





Affect of Tenure on the number of posts

The graph below shows a plot of how the number of posts changes as people stay on the site for longer. Those that have had an account on the site just for a few months are not posting a lot, as seen from the graph. However, with an increase in the number of years, the number of posts people make does not necessarily increase. As seen from the graph, even when people have been on the site for over 2 years, the majority of people continue to not post anything. Therefore, the number of posts people make are not related to the amount of time a person has had an account on High Note – there are other factors that cause people to post, such as their friends posting, and whether they have a premium subscription or not.

User age

The relationship between the age of the users and the average age of their friends is shown below. The relationship is linear – showing that people tend to have friends who are around the same age. Those who are in the premium model in particular have more friends closer to their own age than those who have not yet subscribed.

```
       Welch Two Sample t-test

data:  adopter by sub_friends
t = -30.961, df = 11815, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.1330281 -0.1171869
sample estimates:
mean in group 0 mean in group 1
     0.05243501      0.17754250
```

The initial t-test without propensity-score matching showing a p-value less than 1%. Therefore, we can say from this initial test that it is more likely that people will upgrade to the premium membership if they are in group 1, i.e. if they have friends who have subscribed to the premium model – and this difference is statistically significant. However propensity score matching will need to be performed to remove any pre-treatment bias.

Covariate Means

	sub_friends	age	male	friend_cnt	avg_friend_age	friend_country_cnt	songsListened
	<fct>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	0	23.7	0.629	10.4	23.8	2.73	14602.
2	1	25.4	0.636	54.0	25.4	9.39	33736.

	sub_friends	lovedTracks	posts	playlists	shouts	tenure	good_country
	<fct>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	0	65.2	2.54	0.529	16.4	43.2	0.355
2	1	225.	20.5	0.744	102.	46.5	0.343

In the above tables, *sub*_friends = 0 represents the people that do not have any friends that subscribe to the premium model of HighNote, *sub*_friends = 1 represents the group that have friends subscribing to the premium mode. Those people using HighNote that have no friends who have subscribed to the premium model are younger on average. They have lesser number of friends, and have less friends from other countries. On average, they listen to less than half the number of songs people who have premium subscriber friends listen to. Their interactions with the site (*lovedTracks, posts, playlists* and *shouts*) is on average considerably less.
On average, 63.6% of friends are male for those that have friends that are on the premium subscription, while those that are not have 62.9% of their friends being male. The percentage of people that are from within the US, UK and Germany are fairly similar between the two groups (0 and 1) at 35.5% percent and 34.3% respectively.
Interestingly, the results between the groups of people that have not have premium subscription friends and the ones that two are similar to the results between those who have not subscribed to a premium account and those that have – indicating an overlap in each of these respective groups.

Propensity Score Matching

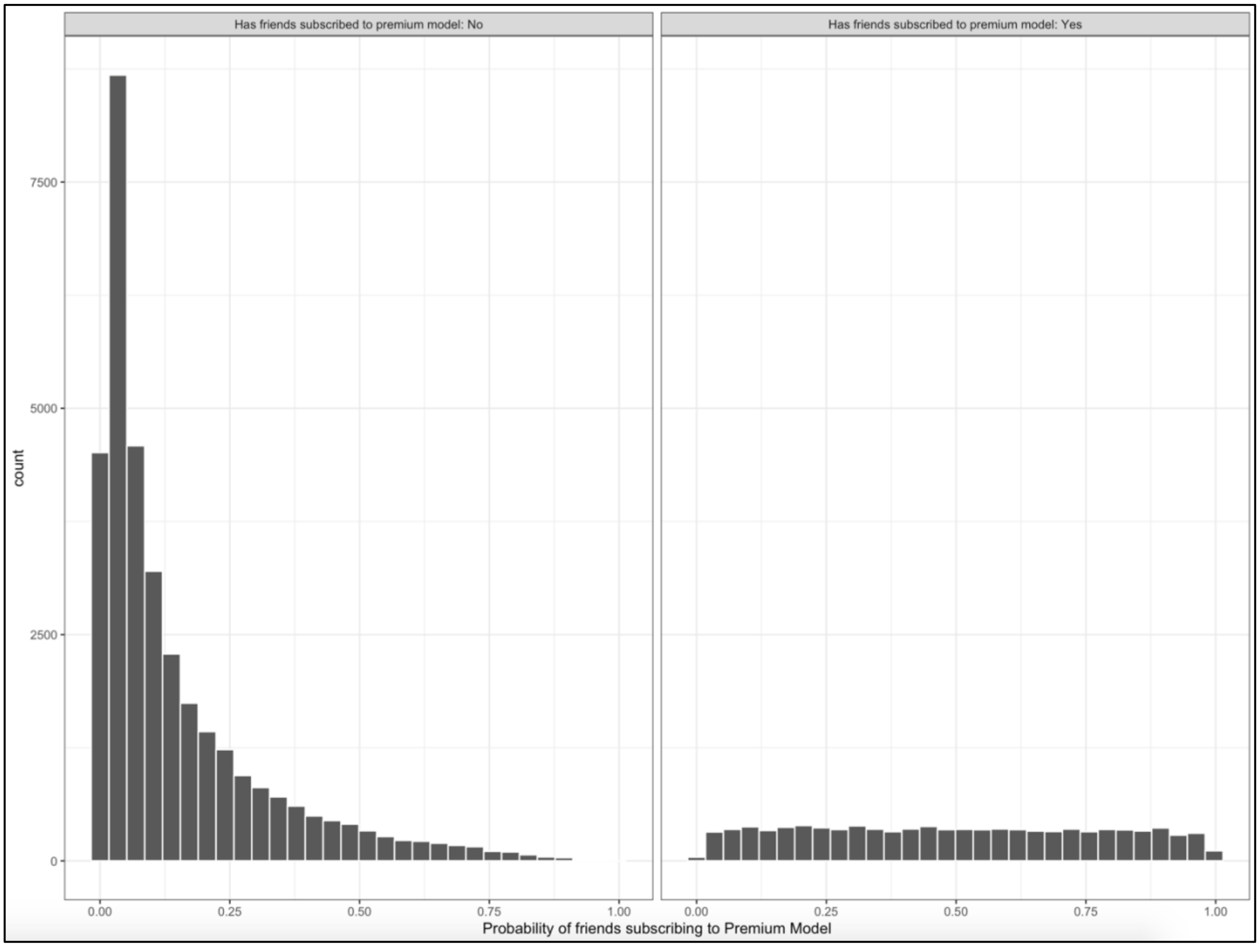
For the purpose of propensity score matching, the variable *avg_friend_male* was removed as I considered it not to have a significant effect on the treatment variable. All other variables were included. The following variables as shown below were log-transformed in order to remove skew and decrease their large range compared to other variables.

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.5982  -0.5715  -0.3055  -0.1316   3.2607

Coefficients:
                            Estimate Std. Error z value Pr(>|z|)
(Intercept)               -8.7160155  0.1206610 -72.236  < 2e-16 ***
age                        0.0246554  0.0031286   7.881 3.25e-15 ***
male                       0.0722676  0.0310923   2.324 0.020110 *
log(friend_cnt + 1)        1.0630919  0.0269597  39.433  < 2e-16 ***
avg_friend_age             0.1088275  0.0038975  27.922  < 2e-16 ***
log(friend_country_cnt + 1)  0.5656527  0.0318013  17.787  < 2e-16 ***
log(songsListened + 1)     0.0548836  0.0091634   5.989 2.11e-09 ***
log(lovedTracks + 1)       0.0864843  0.0079453  10.885  < 2e-16 ***
log(posts + 1)             0.0846196  0.0147911   5.721 1.06e-08 ***
log(playlists + 0.1)      -0.0588226  0.0116923  -5.031 4.88e-07 ***
log(shouts + 1)           -0.0466168  0.0135767  -3.434 0.000596 ***
tenure                    -0.0061654  0.0008633  -7.142 9.23e-13 ***
good_country               0.0592316  0.0303166   1.954 0.050728 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 46640  on 43826  degrees of freedom
Residual deviance: 31761  on 43814  degrees of freedom
AIC: 31787
```

A logistic regression is generated with the categorical variable *sub_friends* as the dependent variable as shown above. A sample of the propensity scores generated are displayed below, and the QQ plots for the differences before and after matching are shown in Appendix B.

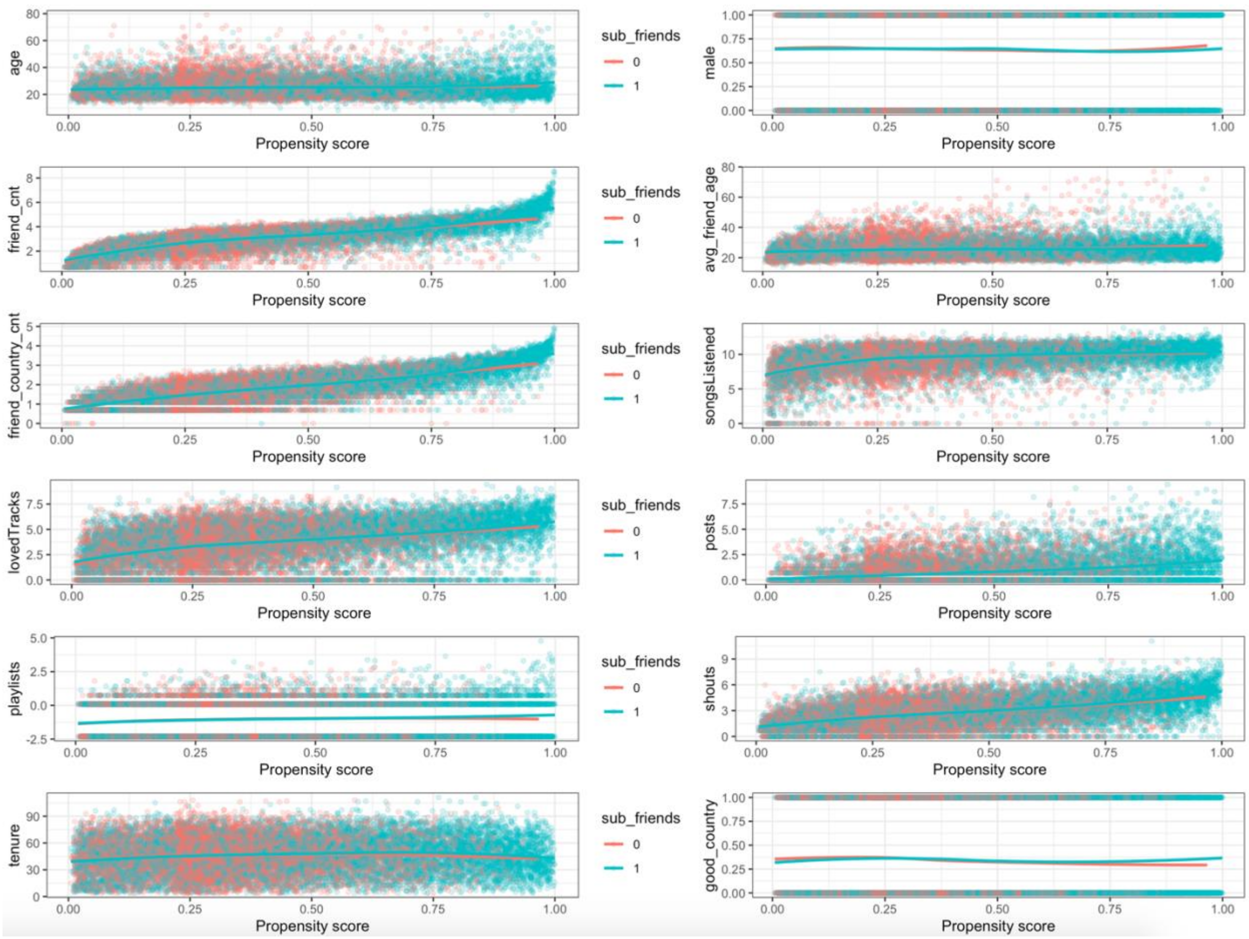
	pr_score	sub_friends
1	0.08002209	0
2	0.04323555	0
3	0.02006080	0
4	0.42613087	1
5	0.63563072	0
6	0.14685328	0

Propensity Score Histogram



The histogram of the propensity scores is as shown above for both the "control" (left) and the "treatment" (right). As expected, there is a higher frequency of **lower** probabilities of subscribing to a premium model for the control group (people that have no friends who subscribe to the premium model), while the probabilities of subscribing for the treatment group is more spread out over the values.

Covariate Balance



In the graphs above, the green plot corresponds to those that don't have any friends that subscribe, and the red line corresponds to the people that have friends that subscribe. As seen from the graphs above, the plots for the *sub_friends* = 0 and *sub_friends* = 1 are very close for certain variables. This implies that the propensity scores were matched well between the control and the treatment group for these variables. For each value in the treatment (*sub_friends* = 1), we were able to find a value in the control (*sub_friends* = 0) which had a very similar propensity score.

When the means between individual variables were tested for significance (shown in appendix A), the difference in means between many variables were still significant, even after the matching. The difference in means were insignificant for some variables, in particular *good_country, tenure* and *male*.

When the t-test is run again on the matched data (shown below), we see that there is still a difference in the average "adopter" values between the two groups. The mean of group 1 (treatment group) is the same, however the mean of the control group is higher than it was in the original t-test. After propensity score matching, we find that the average is 0.09.

Therefore, the number of people that switch to premium subscription when they have friends that also have a premium subscription is more, compared to the number of people that have no friends who have a premium account, that make the switch.
Therefore, after propensity score matching, although there is there is still a significant difference in the means between some of the individual variables that the matching was run on, there is a smaller difference between the means in the two groups.

```
        Welch Two Sample t-test

data:  adopter by sub_friends
t = -17.373, df = 18345, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.09368881 -0.07469152
sample estimates:
mean in group 0 mean in group 1
     0.09335234      0.17754250
```

Logistic regression on the dependent variable *adopter* run with **all** the variables gives the following results:

```
Call:
glm(formula = adopter ~ age + male + friend_cnt + avg_friend_male +
    friend_country_cnt + songsListened + lovedTracks + posts +
    playlists + shouts + tenure + good_country + sub_friends,
    family = binomial(), data = df_c)
```

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.9071  -0.4302  -0.2845  -0.1805   3.2541

Coefficients:
                    Estimate Std. Error z value Pr(>|z|)
(Intercept)        -6.375165   0.146776 -43.435  < 2e-16 ***
age                 0.038022   0.002817  13.498  < 2e-16 ***
male1               0.327316   0.042660   7.673 1.69e-14 ***
friend_cnt          0.086622   0.033951   2.551  0.01073 *
avg_friend_male     0.180736   0.071730   2.520  0.01175 *
friend_country_cnt  0.015183   0.043479   0.349  0.72694
songsListened       0.201243   0.014408  13.968  < 2e-16 ***
lovedTracks         0.302434   0.011476  26.353  < 2e-16 ***
posts               0.132943   0.017394   7.643 2.12e-14 ***
playlists           0.046524   0.015036   3.094  0.00197 **
shouts             -0.137204   0.017342  -7.912 2.54e-15 ***
tenure             -0.006455   0.001106  -5.836 5.34e-09 ***
good_country1      -0.441341   0.041445 -10.649  < 2e-16 ***
sub_friends1        0.677264   0.045103  15.016  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 24537  on 43826  degrees of freedom
Residual deviance: 20875  on 43813  degrees of freedom
AIC: 20903
```

Based on the above results, the variables *friend_country_count* is not significant in predicting whether a free user will subscribe to a premium account or not.

The reduced logistic regression results is shown below. As seen, there is an improvement in the AIC value indicating a better fit. It is also a much better fit compared to the model that was generated before the propensity score matching, which had an AIC of 31787.

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.9049  -0.4301  -0.2845  -0.1804   3.2538

Coefficients:
                 Estimate Std. Error z value Pr(>|z|)
(Intercept)     -6.373335   0.146705 -43.443  < 2e-16 ***
age              0.038132   0.002799  13.622  < 2e-16 ***
male1            0.326341   0.042567   7.667 1.77e-14 ***
friend_cnt       0.094327   0.025794   3.657 0.000255 ***
avg_friend_male  0.180351   0.071744   2.514 0.011943 *
songsListened    0.200967   0.014390  13.966  < 2e-16 ***
lovedTracks      0.302775   0.011436  26.476  < 2e-16 ***
posts            0.133249   0.017371   7.671 1.71e-14 ***
playlists        0.046735   0.015024   3.111 0.001867 **
shouts          -0.136638   0.017262  -7.915 2.46e-15 ***
tenure          -0.006482   0.001103  -5.874 4.25e-09 ***
good_country1   -0.441711   0.041434 -10.661  < 2e-16 ***
sub_friends1     0.679318   0.044710  15.194  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 24537  on 43826  degrees of freedom
Residual deviance: 20875  on 43814  degrees of freedom
AIC: 20901
```

The ANOVA table that compares the above two regression equations on their fit shows the p-value to be significant. Therefore, the fitting of the reduced model is not worse than the original model that had all the variables.

```
Analysis of Deviance Table

Model 1: adopter ~ age + male + friend_cnt + avg_friend_male + friend_country_cnt +
    songsListened + lovedTracks + posts + playlists + shouts +
    tenure + good_country + sub_friends
Model 2: adopter ~ age + male + friend_cnt + avg_friend_male + songsListened +
    lovedTracks + posts + playlists + shouts + tenure + good_country +
    sub_friends
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1     43813      20875
2     43814      20875 -1 -0.12202   0.7269
```

Odds Ratio Interpretation

The odds ratio of *age, male, avg_male_friend, tenure, good_country* and *sub_friends=1* are as shown below.

With every year increase in the age of a person, the odds of them subscribing to the premium model only increases by 3.8%, however attracting a male user would increase the odds of a premium subscription by 38.6%.

With every increase in the number of friends that a person has that is already a premium member, the odds of that person also converting to premium membership increase by **97.3%**.

The odds ratio for *tenure* is less than one – implying that for every month a user stays on the site, the odds of them converting to a premium membership actually decreases by 0.62%.

If a person is from US, UK or Germany, the odds of conversion decreases by 36.4%.

```
> exp(fit_partial$coefficients[2])
     age
1.038868
> exp(fit_partial$coefficients[3])
   male1
1.385888
> exp(fit_partial$coefficients[5])
avg_friend_male
       1.197637
> exp(fit_partial$coefficients[11])
   tenure
0.9935393
> exp(fit_partial$coefficients[12])
good_country1
    0.6429353
> exp(fit_partial$coefficients[13])
sub_friends1
    1.972532
```

Analysing the odds ratio for the log transformed variables:

songsListened: 1.0021
lovedTracks: 1.003
posts: 1.0013
shouts: 0.9992
friend_cnt: 1.00094

From the above ratios, we can conclude that a 1% increase in the number of songs increases the odds of a person being in a premium membership by 0.21%, while an increased interaction with the site by 1% through 'liking' songs and writing posts increases the odds of a premium membership by 0.31% and 0.13% respectively. Having more friends has a very marginal increase in converting to a premium membership – it depends more on the *type* of friend a person has (premium vs free user). Conversely, more number of shouts implies that the odds of being in a premium membership decreases. The above factors may not be causal – for example, more interaction with the site may not lead to increased chance of premium membership, however, having a premium subscription may increase the amount of engagement a person has with HighNote.

According to the analysis, the number of friends a person has, and the number of different countries they are from is not significant at all in predicting whether a person will switch to a premium subscription. The most important variable in making this prediction is *sub_friends* – which is a binary variable that indicates whether a person has friends that subscribe or not. A person with a free account is most likely to upgrade to a premium account if their friends also have premium accounts.

The factor that decreases the odds of conversion significantly is *good_country* – i.e. if the person is residing in the US, UK or Germany. HighNote would gain a lot by targeting people that live in other countries – an increased effort in globalisation may pay off.

There is a significant difference in the account type (free vs premium) between those with zero subscriber friends, and those with at least one, as confirmed by the t-test after propensity score matching.

Appendix A

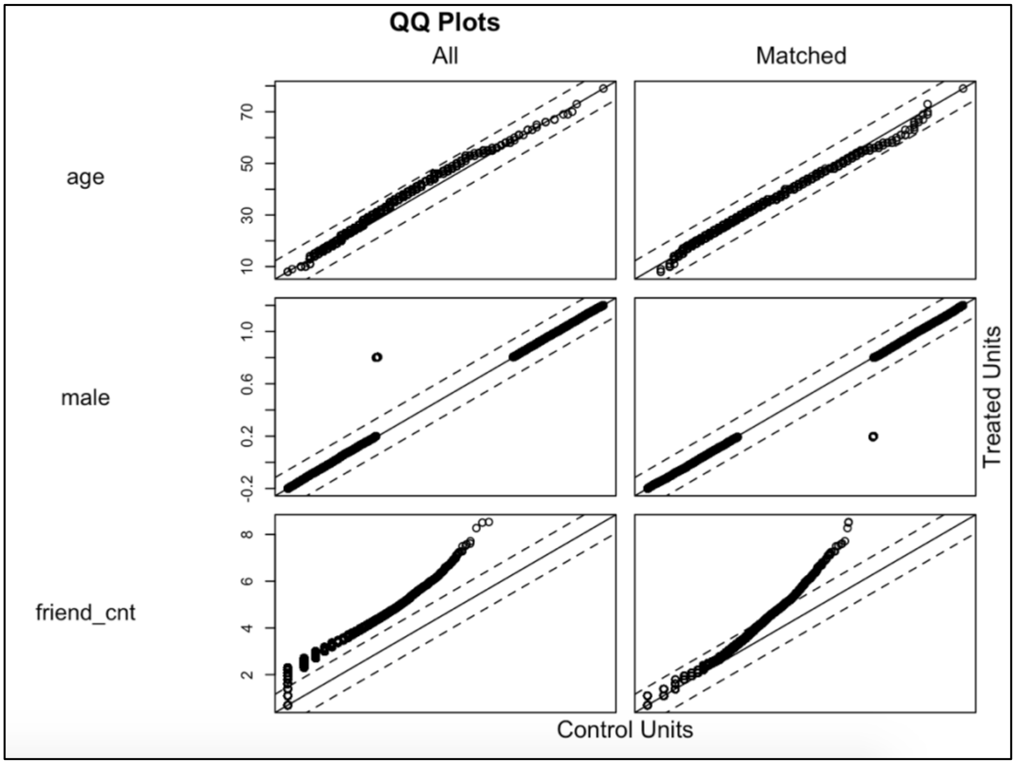
```
[[1]]

        Welch Two Sample t-test

data:  df_m[, v] by df_m$sub_friends
t = -4.1876, df = 19618, p-value = 0.00002831
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.6227657 -0.2256513
sample estimates:
mean in group 0 mean in group 1
       24.94900        25.37321


[[2]]

        Welch Two Sample t-test

data:  df_m[, v] by df_m$sub_friends
t = 0.95096, df = 19644, p-value = 0.3416
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.006913802  0.019944445
sample estimates:
mean in group 0 mean in group 1
      0.6427772       0.6362618


[[3]]

        Welch Two Sample t-test

data:  df_m[, v] by df_m$sub_friends
t = -23.204, df = 10431, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -32.97551 -27.83810
sample estimates:
mean in group 0 mean in group 1
       23.61417        54.02097
```

```
[[4]]

        Welch Two Sample t-test

data:  df_m[, v] by df_m$sub_friends
t = -2.1535, df = 19025, p-value = 0.03129
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.33495953 -0.01575252
sample estimates:
mean in group 0 mean in group 1
       25.21507        25.39043


[[5]]

        Welch Two Sample t-test

data:  df_m[, v] by df_m$sub_friends
t = -36.325, df = 13651, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -4.241056 -3.806791
sample estimates:
mean in group 0 mean in group 1
       5.361702        9.385626


[[6]]

        Welch Two Sample t-test

data:  df_m[, v] by df_m$sub_friends
t = -12.853, df = 17594, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -8021.159 -5898.369
sample estimates:
mean in group 0 mean in group 1
       26775.88        33735.64
```

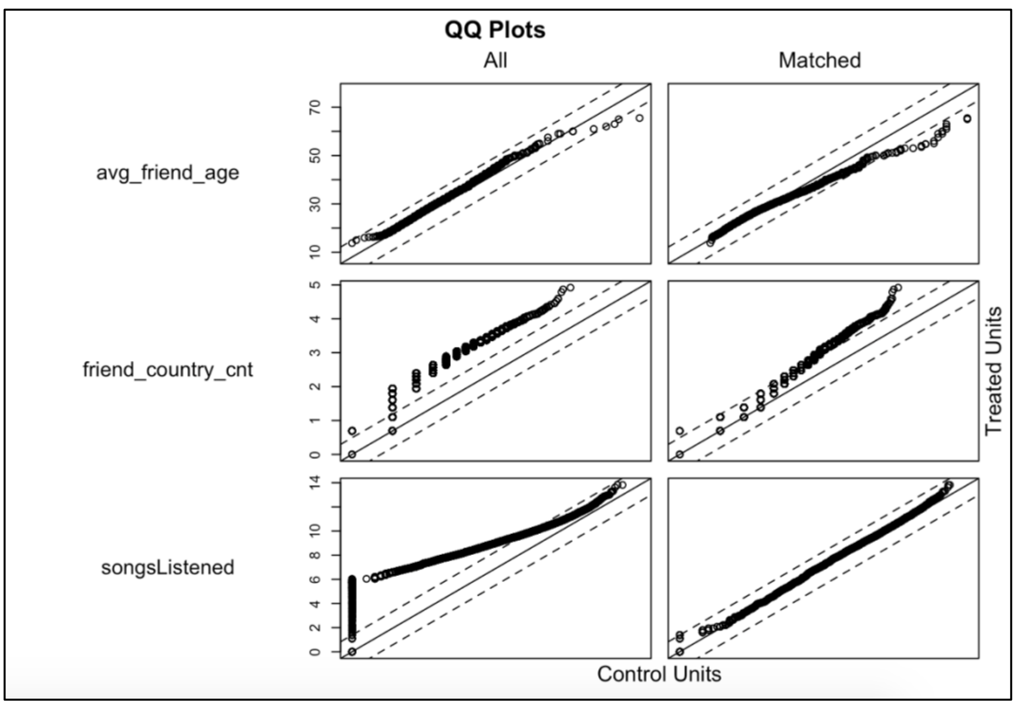
```
[[7]]

        Welch Two Sample t-test

data:  df_m[, v] by df_m$sub_friends
t = -16.215, df = 15293, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -104.34507  -81.83899
sample estimates:
mean in group 0 mean in group 1
       132.2726        225.3647


[[8]]

        Welch Two Sample t-test

data:  df_m[, v] by df_m$sub_friends
t = -5.6898, df = 10924, p-value = 1.304e-08
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -19.139682  -9.331254
sample estimates:
mean in group 0 mean in group 1
       6.287489       20.522956


[[9]]

        Welch Two Sample t-test

data:  df_m[, v] by df_m$sub_friends
t = -5.6359, df = 14051, p-value = 1.775e-08
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.16629585 -0.08047194
sample estimates:
mean in group 0 mean in group 1
      0.6206861       0.7440700
```

```
[[10]]

        Welch Two Sample t-test

data:  df_m[, v] by df_m$sub_friends
t = -8.1719, df = 10515, p-value = 3.383e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -76.92471 -47.16060
sample estimates:
mean in group 0 mean in group 1
       39.77685       101.81951


[[11]]

        Welch Two Sample t-test

data:  df_m[, v] by df_m$sub_friends
t = -1.6432, df = 19632, p-value = 0.1004
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.01194080  0.08900483
sample estimates:
mean in group 0 mean in group 1
       46.08724        46.54871


[[12]]

        Welch Two Sample t-test

data:  df_m[, v] by df_m$sub_friends
t = 0.69007, df = 19644, p-value = 0.4902
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.008618542  0.017984316
sample estimates:
mean in group 0 mean in group 1
      0.3479589       0.3432760
```

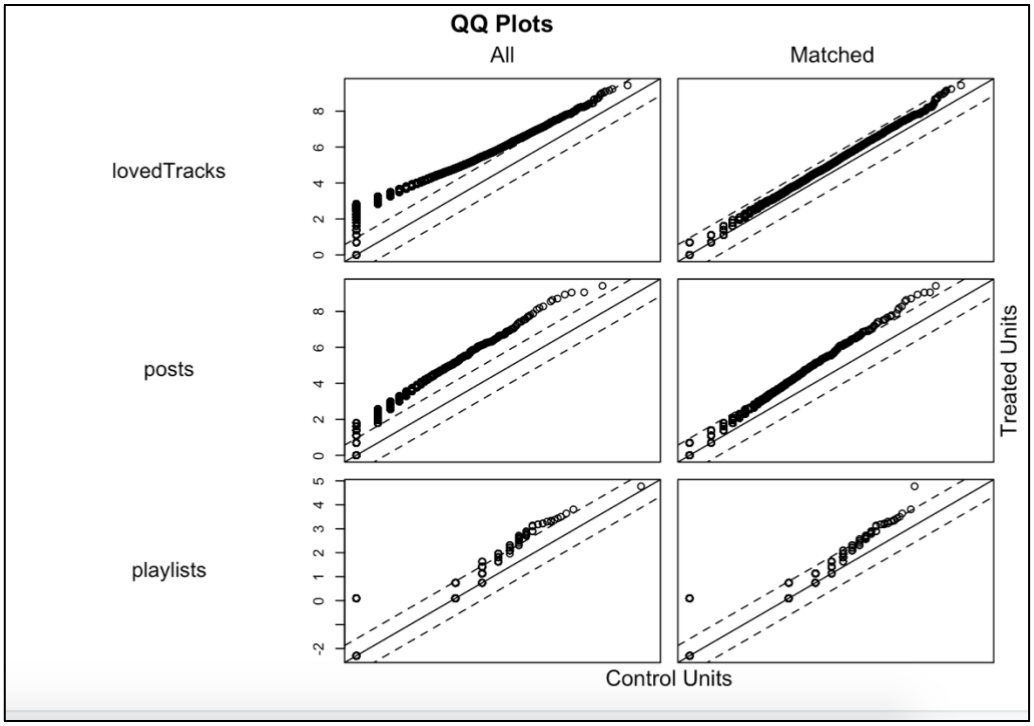
Appendix B

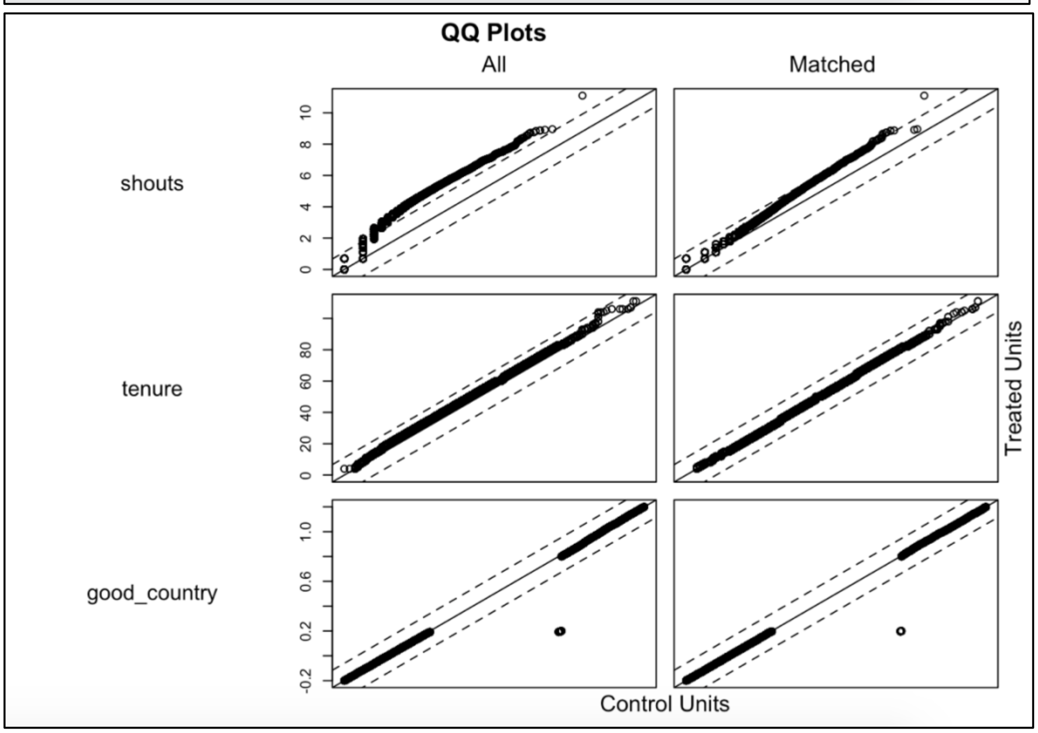


QQ Plots
All
Matched
age
male
friend_cnt
Treated Units
Control Units



QQ Plots
All
Matched
avg_friend_age
friend_country_cnt
songsListened
Treated Units
Control Units


QQ Plots
All
Matched
lovedTracks
posts
playlists
Treated Units
Control Units


QQ Plots
All
Matched
shouts
tenure
good_country
Treated Units
Control Units