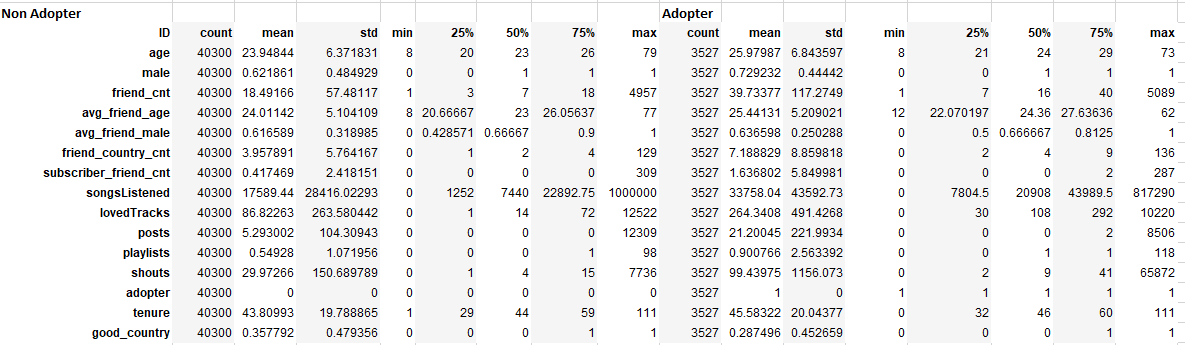
**Final Exam Questions**

1. Summary statistics: Generate **descriptive statistics** for the key variables in the data set, similar to the **table** on the last page of the case. (Note that your table will look different because the data set you are analyzing is different from the one used to generate the table in the case.) Analyze the **differences in the mean values** of the variables, comparing the adopter and non-adapter subsamples. What tentative conclusions can you draw from these comparisons?

**Answer:-**

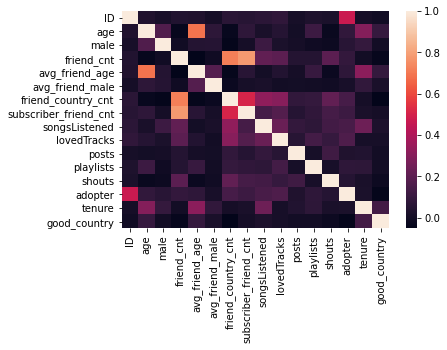
**Adopter**



From the mean values for the adopters and the non-adopters, we can see that a higher proportion of the adopters are male. Also the average friend count is more than double for the adopters. Also, the number of countries from which the friends are is nearly double for adopters compared to non adopters. Also the number of friends who are subscribers is substantially higher for adopters. Also, the songs listened to, the loved tracks, the posts, the playlists and the shouts are all much higher for adopters than for non adopters. From all this we can say that adopters have higher values for content consumption, content organisation and community involvement in line with the paper by Oestreichr-singer and Zalmanson.

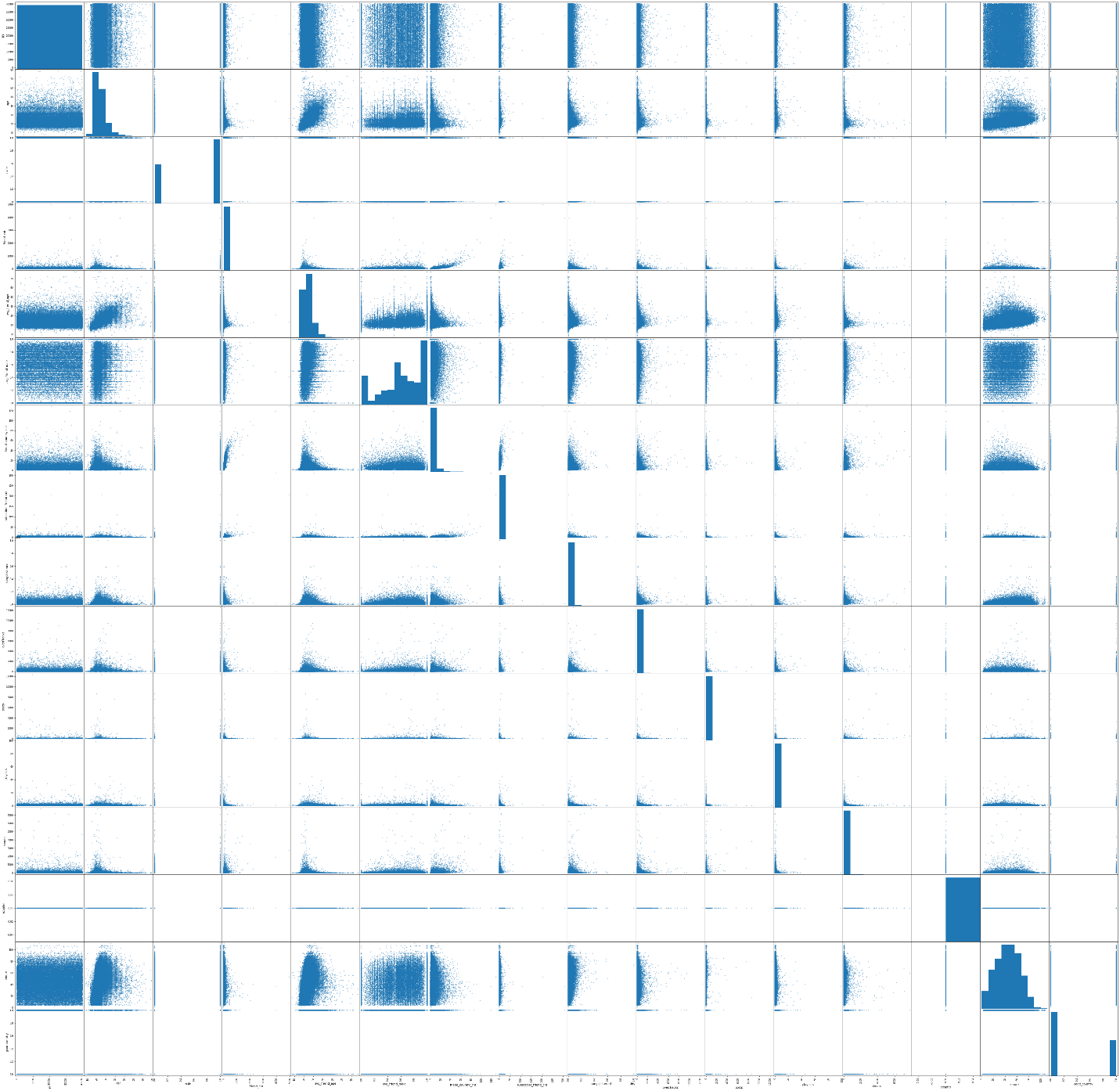
1. Data Visualization: Generate a set of charts (e.g., scatter plots, box plots, etc) to help **visualize** how adopters and non-adopters (of the premium subscription service) differ from each other in terms of (i) demographics, (ii) peer influence, and (iii) user engagement. What can you conclude from your charts?

**Answer:-**

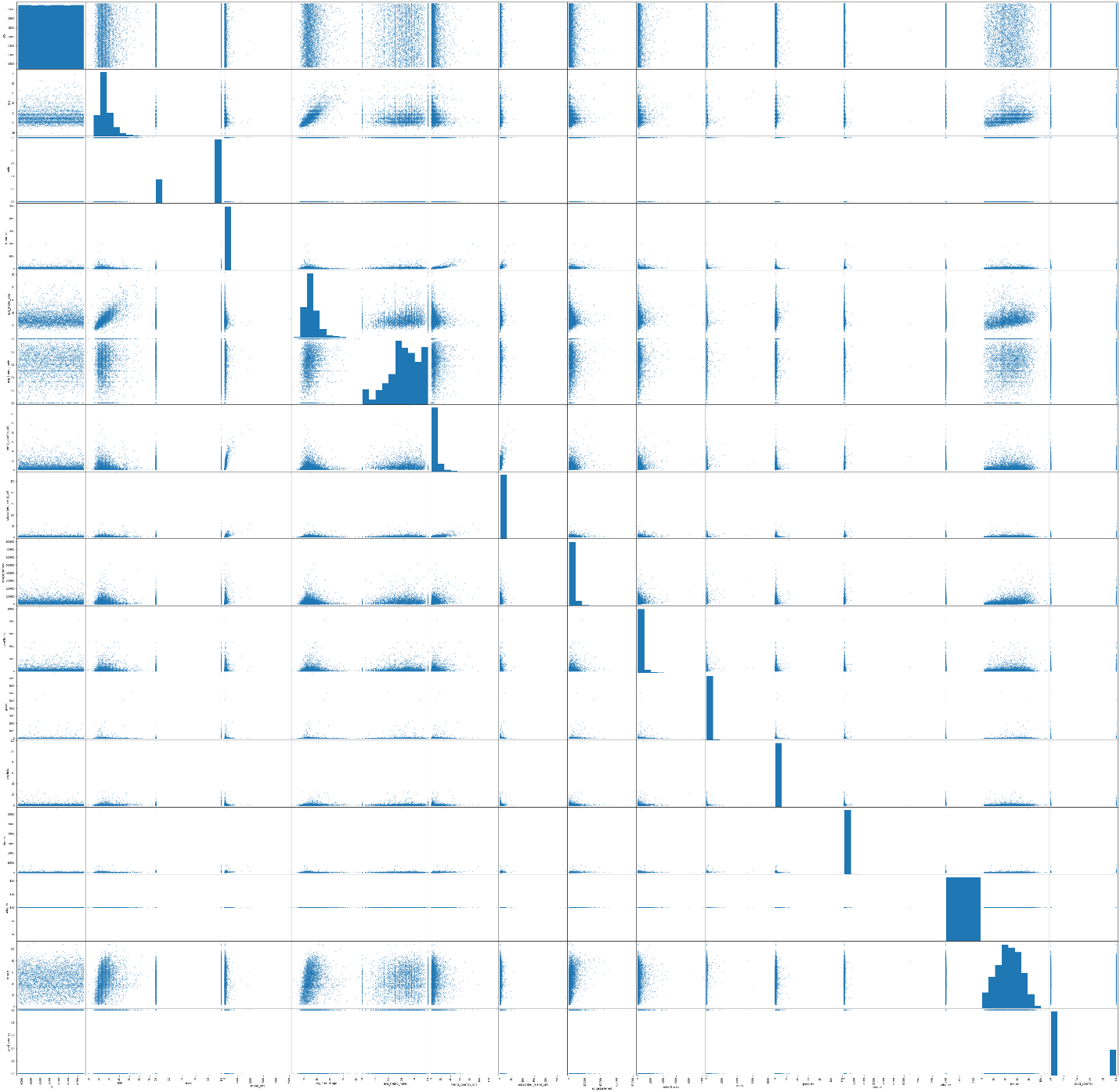
****

We can create scatter matrices for both adopters and non-adopters.

**Non-Adopters:-**

****

**Adopters:-**

****

1. **Demographics :-**

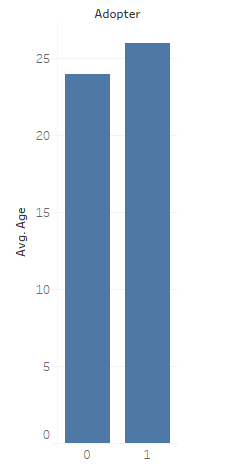
The demographics columns are :-

a) age

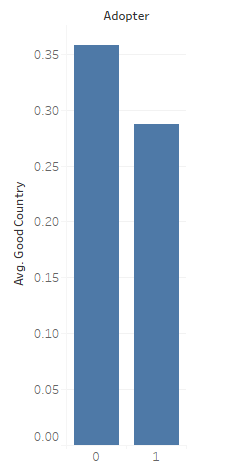
b) male

c) good\_country

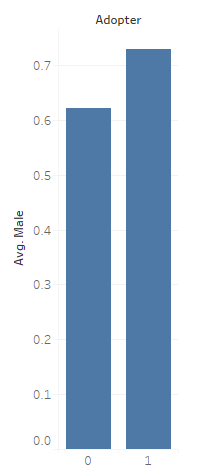
**Age**

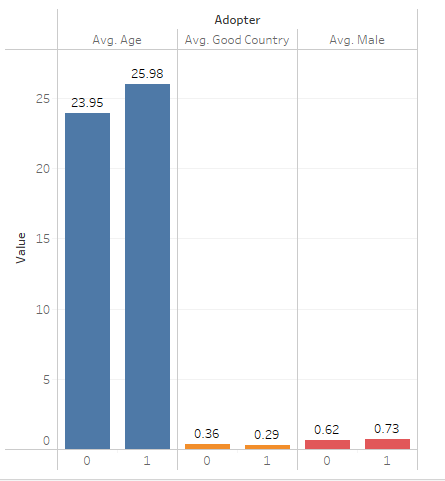
****

**Good country**

****

**Male**

****

****

From these charts we can see that when it comes to demographics, that adopters in general are slightly older, have a slightly higher probability to come from a non-good country and have a slightly higher probability of being male. The differences are not very significant though.

1. **Peer Influence :-**

The peer influence columns are :-

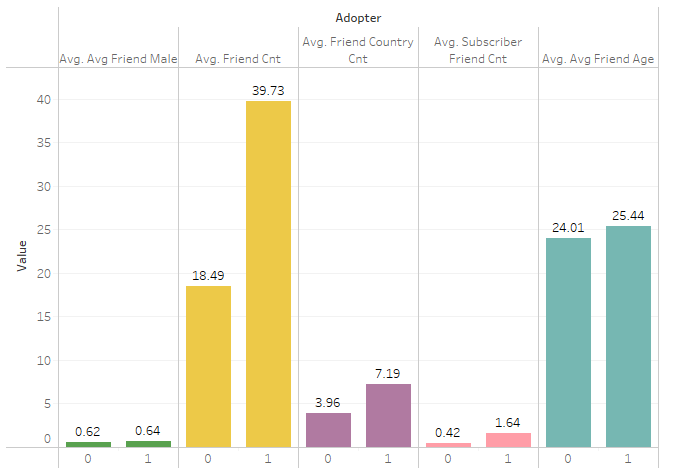
a) friend\_count

b) avg\_friend\_age

c) avg\_friend\_male

d) friend\_country\_cnt

e) subscriber\_friend\_cnt

****

From the above diagram, we can see that the average of the friend count and the number friend countries is more than twice for the adopter than that of the non-adopters. Also the average subscriber friend count is higher for the adopters compared to the non-adopters.

From this, it is apparent that the adopters have more peer influence than non-adopters and are able to influence more people and also people in a larger geographic distribution than non-adopters.

1. **User Engagement :-**

The user engagement columns are :-

a) tenure

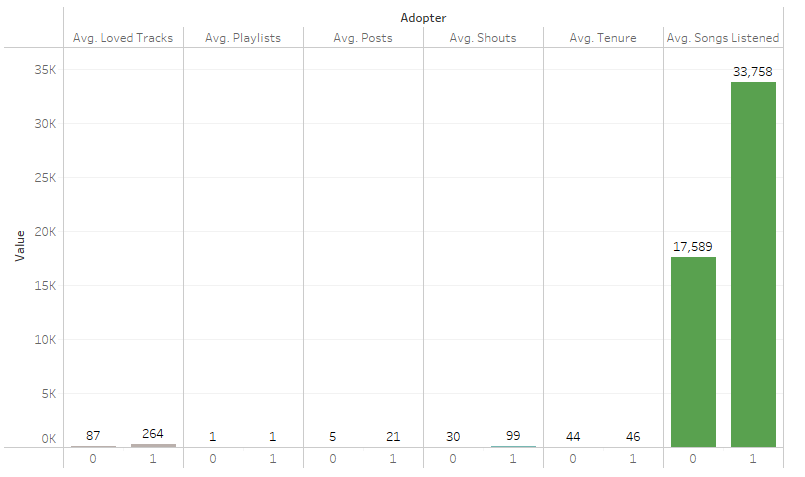
b) shouts

c) playlists

d) posts

e) lovedTracks

f) songsListened

****

From the above diagram we can see that the average score for the loved tracks, posts, shouts and songs listened to is much higher for adopters than for non-adopters. This shows that the user engagement is much higher for the adopters than for non adopters.

1. Propensity Score Matching (PSM): You will **use PSM** to test whether having subscriber friends affects the likelihood of becoming an adopter (i.e., fee customer). For this purpose, the **"treatment" group** will be users that have one or more subscriber friends (subscriber\_friend\_cnt >= 1), while the **"control" group** will include users with **zero subscriber friends**. Use PSM to **first create matched treatment and control samples**, then test whether there is a significant average treatment effect. Provide an interpretation of your results.

**Answer :-**

Call:

matchit(formula = adopter ~ songsListened + lovedTracks + playlists +

posts + shouts + friend\_cnt + have\_subscriber\_friend + age +

male + tenure + good\_country, data = hnd, method = "nearest")

Summary of Balance for All Data:

Means Treated Means Control Std. Mean Diff. Var. Ratio eCDF Mean eCDF Max

distance 0.1422 0.0751 0.5639 3.1659 0.2461 0.3726

songsListened 33758.0405 17589.4415 0.3709 2.3534 0.1917 0.2633

lovedTracks 264.3408 86.8226 0.3612 3.4761 0.1029 0.3676

playlists 0.9008 0.5493 0.1371 5.7184 0.0094 0.0899

posts 21.2005 5.2930 0.0717 4.5293 0.0188 0.1689

shouts 99.4398 29.9727 0.0601 58.8577 0.0407 0.1583

friend\_cnt 39.7338 18.4917 0.1811 4.1625 0.0413 0.2514

have\_subscriber\_friend 0.4945 0.2005 0.5880 . 0.2940 0.2940

age 25.9799 23.9484 0.2968 1.1536 0.0300 0.1583

male0 0.2708 0.3781 -0.2416 . 0.1074 0.1074

male1 0.7292 0.6219 0.2416 . 0.1074 0.1074

tenure 45.5832 43.8099 0.0885 1.0259 0.0164 0.0402

good\_country0 0.7125 0.6422 0.1553 . 0.0703 0.0703

good\_country1 0.2875 0.3578 -0.1553 . 0.0703 0.0703

Summary of Balance for Matched Data:

Means Treated Means Control Std. Mean Diff. Var. Ratio eCDF Mean eCDF Max Std. Pair Dist.

distance 0.1422 0.1421 0.0008 1.0066 0.0000 0.0017 0.0014

songsListened 33758.0405 31945.3462 0.0416 0.7499 0.0571 0.1032 0.7138

lovedTracks 264.3408 217.9940 0.0943 0.7756 0.0380 0.2056 0.5306

playlists 0.9008 0.7335 0.0653 2.9668 0.0042 0.0431 0.3823

posts 21.2005 10.4117 0.0486 3.9761 0.0094 0.0666 0.1326

shouts 99.4398 64.9779 0.0298 18.0992 0.0134 0.0380 0.1136

friend\_cnt 39.7338 37.1480 0.0220 0.9848 0.0086 0.0868 0.3931

have\_subscriber\_friend 0.4945 0.5262 -0.0635 . 0.0318 0.0318 0.2586

age 25.9799 26.2711 -0.0425 0.7188 0.0115 0.0417 1.0301

male0 0.2708 0.2546 0.0364 . 0.0162 0.0162 0.7229

male1 0.7292 0.7454 -0.0364 . 0.0162 0.0162 0.7229

tenure 45.5832 45.2288 0.0177 0.9902 0.0053 0.0247 1.1212

good\_country0 0.7125 0.7295 -0.0376 . 0.0170 0.0170 0.8457

good\_country1 0.2875 0.2705 0.0376 . 0.0170 0.0170 0.8457

Percent Balance Improvement:

Std. Mean Diff. Var. Ratio eCDF Mean eCDF Max

distance 99.9 99.4 100.0 99.5

songsListened 88.8 66.4 70.2 60.8

lovedTracks 73.9 79.6 63.0 44.1

playlists 52.4 37.6 55.3 52.0

posts 32.2 8.6 49.8 60.5

shouts 50.4 28.9 67.1 76.0

friend\_cnt 87.8 98.9 79.2 65.5

have\_subscriber\_friend 89.2 . 89.2 89.2

age 85.7 -131.1 61.6 73.7

male0 84.9 . 84.9 84.9

male1 84.9 . 84.9 84.9

tenure 80.0 61.4 67.5 38.6

good\_country0 75.8 . 75.8 75.8

good\_country1 75.8 . 75.8 75.8

Sample Sizes:

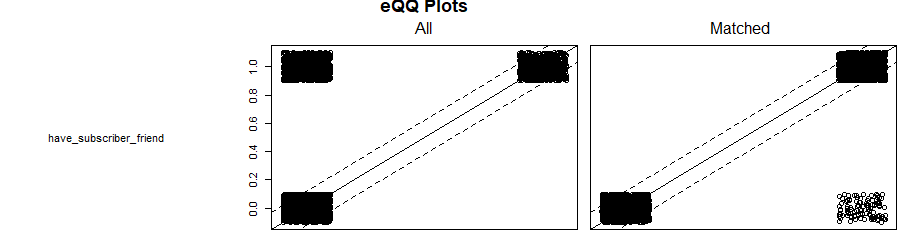
Control Treated

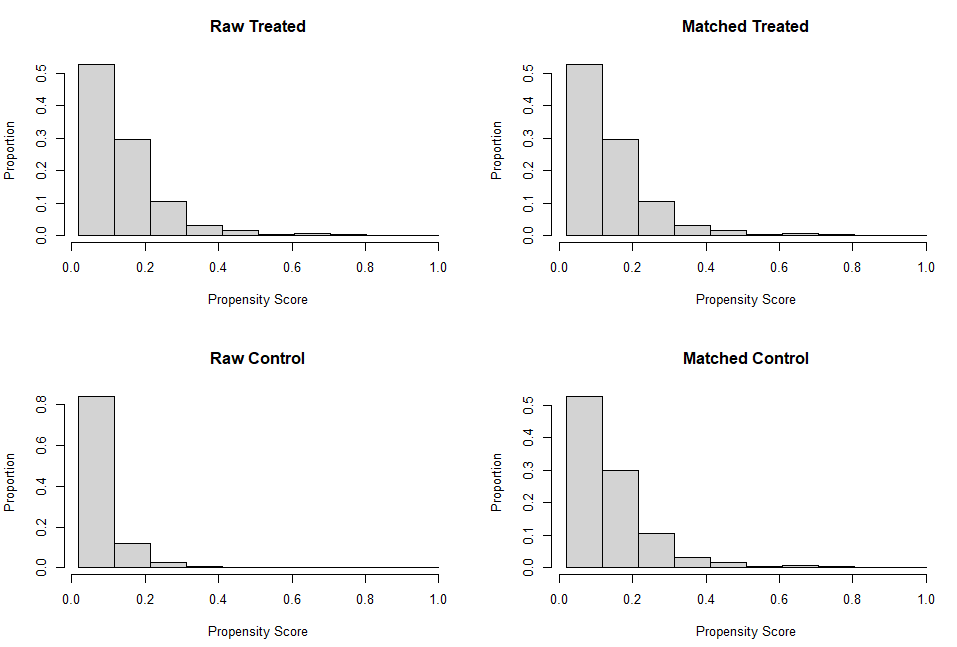
All 40300 3527

Matched 3527 3527

Unmatched 36773 0

Discarded 0 0



****

From the above diagram, we can safely say that there is a significant average treatment effect.

1. Regression Analyses: Now, we will use a **logistic regression** approach to test which variables (including subscriber friends) are significant for explaining the likelihood of becoming an adopter. Use your judgment and visualization results to decide which variables to include in the regression. Estimate the **odds ratios** for the key variables. What can you **conclude** from your results?

**Answer :-**

Call:

glm(formula = adopter ~ songsListened + lovedTracks + playlists +

posts + shouts + friend\_cnt + have\_subscriber\_friend + age +

male + tenure + good\_country, family = "binomial", data = dta\_m)

Deviance Residuals:

Min 1Q Median 3Q Max

-2.4015 -1.1663 -0.4604 1.1757 1.3755

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) 1.843e-01 1.006e-01 1.833 0.06685 .

songsListened 3.376e-07 5.702e-07 0.592 0.55380

lovedTracks 1.606e-04 5.167e-05 3.108 0.00188 \*\*

playlists 5.538e-02 1.720e-02 3.219 0.00128 \*\*

posts 4.516e-04 2.411e-04 1.873 0.06102 .

shouts 9.641e-05 9.090e-05 1.061 0.28890

friend\_cnt 6.183e-05 2.258e-04 0.274 0.78424

have\_subscriber\_friend -2.046e-01 5.076e-02 -4.031 5.56e-05 \*\*\*

age -7.192e-03 3.427e-03 -2.099 0.03583 \*

male1 -9.468e-02 5.553e-02 -1.705 0.08819 .

tenure 9.581e-04 1.294e-03 0.740 0.45910

good\_country1 8.864e-02 5.415e-02 1.637 0.10164

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 9778.9 on 7053 degrees of freedom

Residual deviance: 9718.9 on 7042 degrees of freedom

AIC: 9742.9

Number of Fisher Scoring iterations: 5

The **odds ratios** are:-

have\_subscriber\_friend – e^-0.2046

age – e^( -7.192e-03)

posts – e^( 4.516e-04)

playlists – e^( 5.538e-02)

lovedTracks – e^( 1.606e-04)

We can see that the variables that are significant for explaining the likelihood of becoming an adopter are have\_subscriber\_friend, age, posts, playlists and lovedTracks.

So it looks like the higher the number of subscriber friends, the more likely you are to go premium. This is the only peer influence factor that is significant. Also it looks like the age is important. This is even though the difference in average ages is actually quite less. That is the only demographic factor that is significant. The number of posts, playlists and lovedtracks is also significant. This shows that user engagement is important for determining whether a person will go premium.

1. Takeaways: Discuss some key **takeaways** from your analysis. Specifically, how do your results inform a “free-to-fee” **strategy** for High Note?

**Answer :-** It looks like the most important factors for determining whether a person shifts to the premium level from the free version are having a subscriber friend, age, posts, playlists and loved tracks.

The strategy should be to concentrate on those people who are friends of premium members and who already do a lot of engagement on the platform. Specifically friends of premium members who have a lot of posts and have playlists and who have many loved tracks should be concentrated on. A viral marketing strategy is viable. The age variable is also surprisingly significant. This might be because younger users are more interested and should be targeted.