**Project Title: Twitter**

Group #44 : Yalda Amini , Daniel Barjum, Joao Araojo

We already hold a meeting with our assigned TF.

# Problem statement

With spread of social media and their increasing impact on the news the topic of bots and spread of fake news is one of the hot topics now. The goal of this project is to use machine learning techniques which we have learned during this semester to detect the tweets which were generated by Twitter bots from human users.

# Literature Review

Bots have a heavy presence in the social media. “Of all tweeted links to popular websites, 66% are shared by accounts with characteristics common among automated “bots,” rather than human users.”[[1]](#footnote-1). Due to this huge impact of pots on many political, social and economical topics around the world There is considerable literature trying to solve the issue of the fake bot generated contents in social media. Based on Ferrera et all paper in 2018, there are three types of approaches to detect bot generated contents in social media, consisting “(a) methods based on social network; (b) systems based on crowd-sourcing and human computation; (c) algorithms based on predictive features that separate bots from humans “[[2]](#footnote-2). Recently using RNN and long short-term memory (LSTM) networks seems to provide good performance for bot detection.

# Data Resources

Data is mined and collected from the Twitter developer API (using tweepy). Twitter API passes json format data, we fetched them for two groups of user, i) 23 famous bots ii)100 famous verified real accounts. We saved the fetched tweets from both group after cleaning as a dataframe format in a csv file.

# EDA

Based on below preliminary EDA and histograms it seems that below feature can have predicting power.

• Followers Count

• Friends Count

• Favorites Count

• Retweet Count

Figure :Human generated tweets

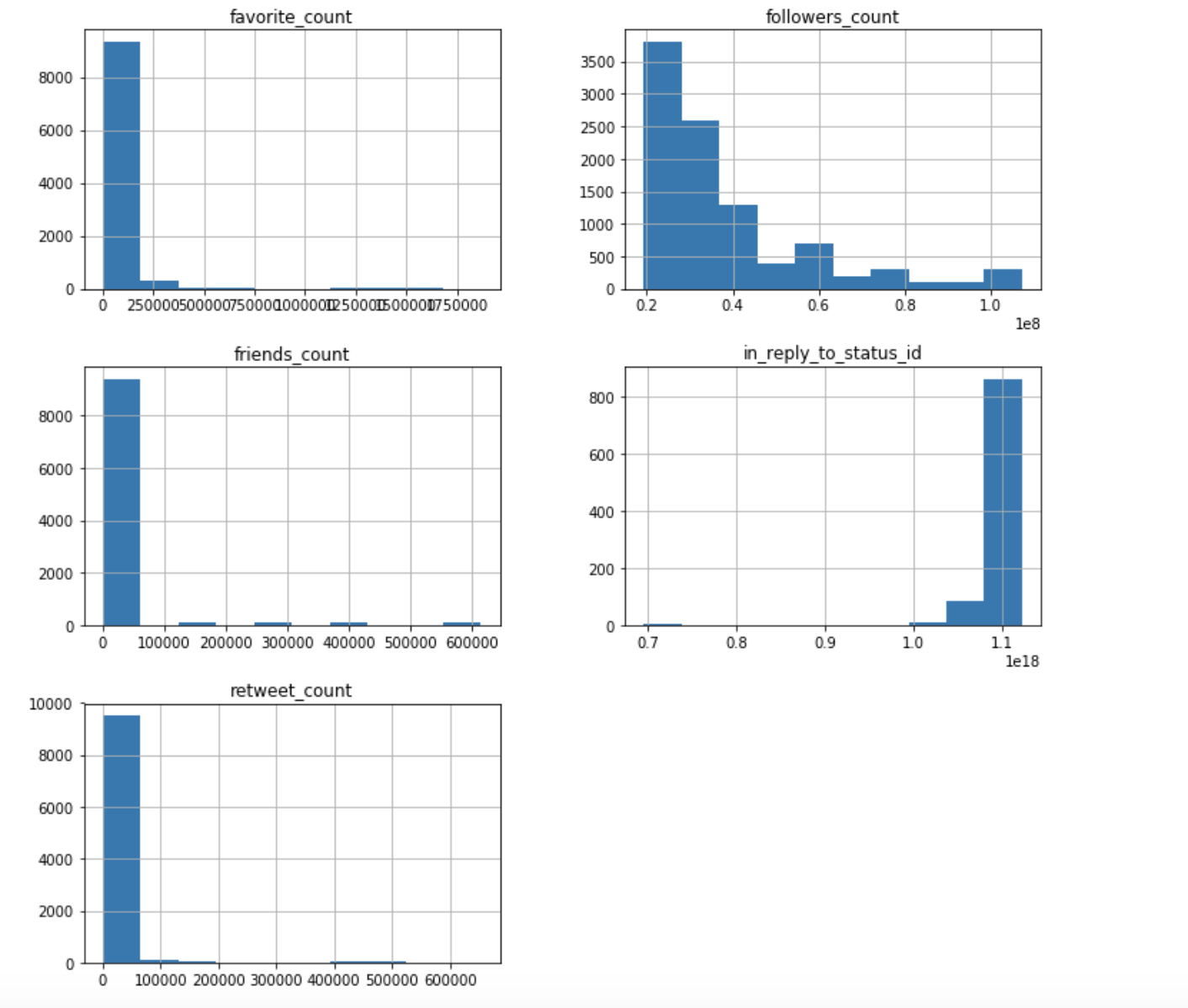


Figure : Bot\_generated

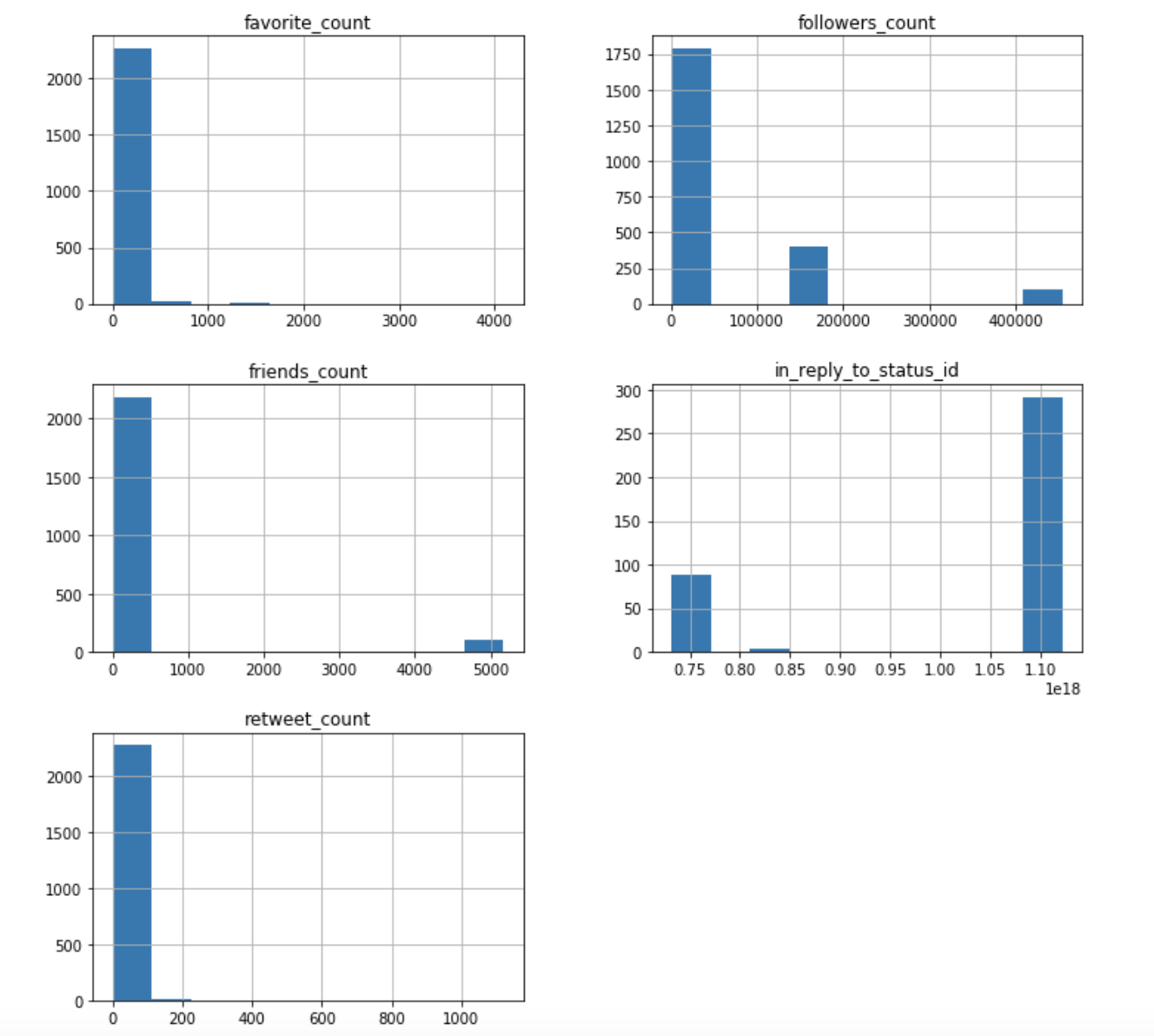
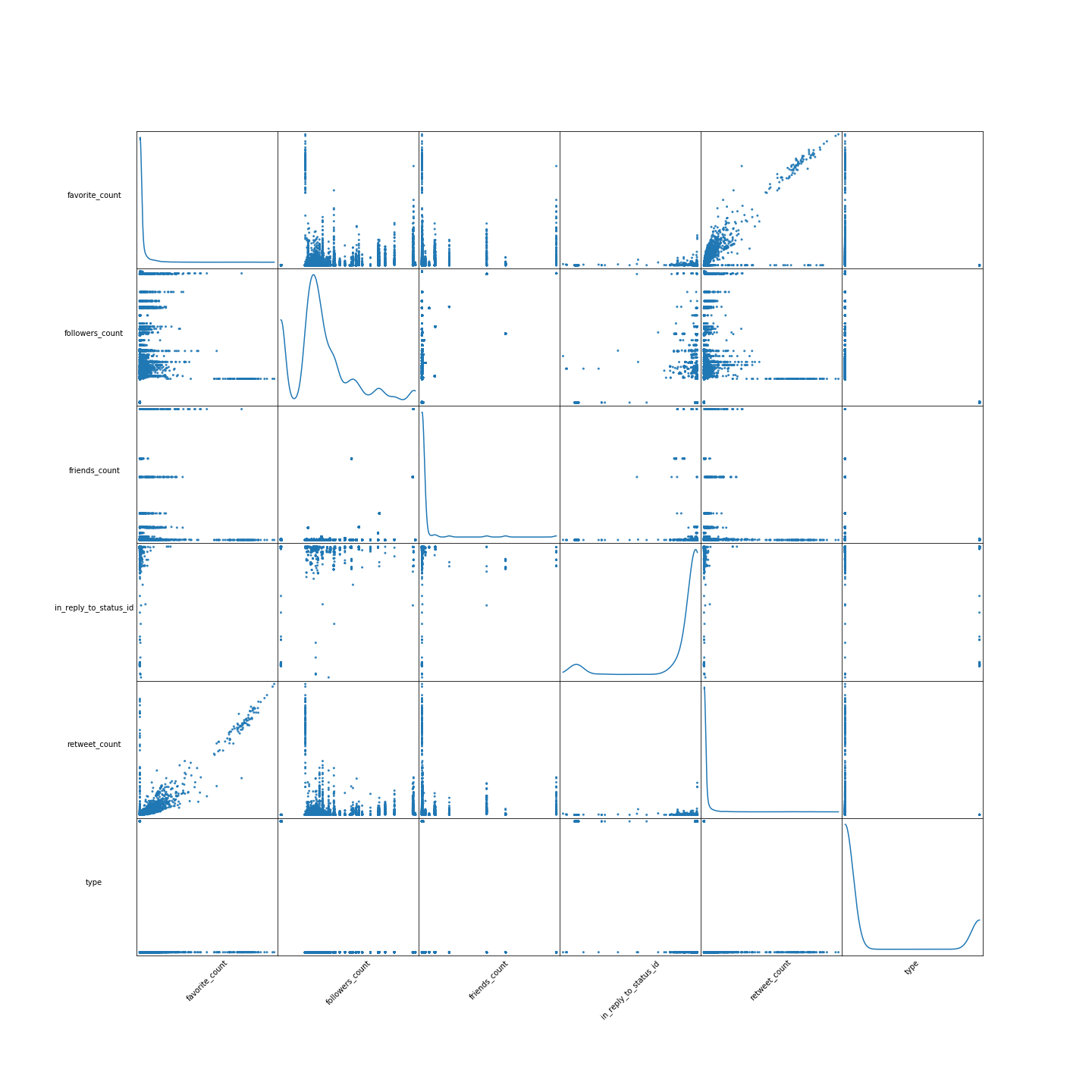


Figure : Scatter Matrix, Type1 = bot, type0=human generated****

# Created DataFrame

# 

# References

Tweepy Python Library

Twitter’s developer resources: developer.twitter.com

Kudugunta, Sneha, and Emilio Ferrara. “Deep Neural Networks for Bot Detection.” *Information Sciences* 467 (October 2018): 312–22. https://doi.org/10.1016/j.ins.2018.08.019.

“Twitter Bots: An Analysis of the Links Automated Accounts Share | Pew Research Center,” April 9, 2018. https://www.pewinternet.org/2018/04/09/bots-in-the-twittersphere/.

1. “Twitter Bots.” [↑](#footnote-ref-1)
2. Kudugunta and Ferrara, “Deep Neural Networks for Bot Detection.” [↑](#footnote-ref-2)