

Student Loan Twitter Data

2022-12-10

Contents

Extract Age of the Twitter Accounts	1
Research Questions:	4
A) Is the engagement more polarized when they are engaging/associating with conservative sources (Fox) or liberal sources (CNN, MSNBC, NPR, NYTimes)? . . .	4
Polarization regarding topic	4
Conservative opinion	5
LIBERAL ENGAGEMENT	6
Other EDA Notes	7
E) Are there non-US-based accounts engaging in the student loan conversation.	7
Liberal opinion	8
Controlled opinion	8
Some multi colored bar plot here	9
visual ideas, neutral get 3 bars, which is averaged out..	9
B) Is there consistency in opinion across tweets that include the 3 groups in their conversation?	9
C) Is there consistency in polarization across tweets that include the 3 groups in their conversation?	9
D) Does one page get more engagement over the other?	9
F) What concentrations of replies are by authors who engage with both new sources?	9

This script's input data uses 5 annotators. Preliminary EDA on the variables are in `explore.Rmd`.

```
tweets = read.csv("data/master_annotated_2.csv")
```

Extract Age of the Twitter Accounts

```

today <- as.Date("2022-12-10")

# FORMATTING AGE
tweets <- tweets %>% mutate(age_dow = substr(created_at, 1, 3)
                             , age_month_day = substr(created_at, 5, 10)
                             , age_time = substr(created_at, 12, 19)
                             , age_yr = substr(created_at, 26, 30),
                             , age_ymd = as.Date(paste0(age_month_day, age_yr), format = "%b %d %Y"),
                             , account_age = today - age_ymd)

# FORMATTING LIBERAL VS CONSERVATIVE
tweets <- tweets %>% mutate(politics = ifelse(experiment_group %in% c('cnn', 'msnbc', 'npr', 'nytimes')
                                              ifelse(experiment_group == 'usedgov', 'controlled', 'conservative'))

print(paste("min age of acct (days): ", min(tweets$account_age)))

```

```
## [1] "min age of acct (days): 7"
```

```
print(paste("max age of acct (days): ", max(tweets$account_age)))
```

```
## [1] "max age of acct (days): 5259"
```

```
print(paste("mean age of acct (days): ", mean(tweets$account_age)))
```

```
## [1] "mean age of acct (days): 1186.82905982906"
```

```
print(paste("median age of acct (days): ", median(tweets$account_age)))
```

```
## [1] "median age of acct (days): 270.5"
```

The youngest `account_age` is 7 days, and the oldest account is 14 years (5259 days).

How many accounts are younger than 2 weeks?

4% (19) of the accounts are younger than two weeks old, and all of them address @foxnews.

```
tweets %>% filter(account_age < 14) %>% group_by(experiment_group) %>% summarize(younger_than_two_weeks = sum(account_age < 14))
```

```
## # A tibble: 1 x 2
##   experiment_group younger_than_two_weeks
##   <chr>                <int>
## 1 foxnews                19
```

What are these 19 accounts sentiment towards student loan forgiveness plan?

10 have neutral support, 5 supports and 4 are against. We will look into the tweets to verify further.

```
tweets %>% filter(account_age < 14) %>% group_by(experiment_group, opinion_label) %>% summarize(younger_than_two_weeks = sum(account_age < 14))
```

```
## 'summarise()' has grouped output by 'experiment_group'. You can override using
## the '.groups' argument.
```

```
## # A tibble: 3 x 3
## # Groups:   experiment_group [1]
##   experiment_group opinion_label younger_than_two_weeks
##   <chr>           <chr>                <int>
## 1 foxnews        AGAINST student loan forgiveness      4
## 2 foxnews        FOR student loan forgiveness          5
## 3 foxnews        NEUTRAL support                      10
```

Given the low agreement score from the annotation, it's expected that the tweets aren't clear cut for/against/neutral on student loan.

```
tweets %>% filter(account_age < 14) %>% select(text, opinion_label) %>% arrange(opinion_label)
```

```
##
## 1
## 2
## 3
## 4 @FoxNews He is a murderer! My family fought in Afghanistan and Biden turn
## 5 @FoxNews It's not debt forgiveness! It is #debtredistribution! And coll
## 6 @FoxNews All lefties when running for
## 7
## 8
## 9 @FoxNews This is no different than the str
## 10
## 11 @FoxNews These are exactly the items of data this man should have everyday. With our country is s
## 12 @FoxNews See... again they are lying. Biden is just proposing the student l
## 13
## 14 @FoxNews Rev. Warnock can't run on dem policy or his voting record so he
## 15
## 16 @FoxNews S
## 17 @FoxNews Crap. He needs a national platform for his pres
## 18 @FoxNews He will promise like Biden did with student loans knowing it will never happen. But i
## 19 @FoxNews Democrats are making EVI
##
##   opinion_label
## 1 AGAINST student loan forgiveness
## 2 AGAINST student loan forgiveness
## 3 AGAINST student loan forgiveness
## 4 AGAINST student loan forgiveness
## 5 FOR student loan forgiveness
## 6 FOR student loan forgiveness
## 7 FOR student loan forgiveness
## 8 FOR student loan forgiveness
## 9 FOR student loan forgiveness
## 10 NEUTRAL support
## 11 NEUTRAL support
## 12 NEUTRAL support
## 13 NEUTRAL support
## 14 NEUTRAL support
## 15 NEUTRAL support
## 16 NEUTRAL support
```

```
## 17          NEUTRAL support
## 18          NEUTRAL support
## 19          NEUTRAL support
```

How many friends/followers do these 19 accounts have?

Given that they're so new to Twitter, we are unsure if they are actual accounts engaging in discourse on Twitter or bots.

How about accounts 30 days old? 44 accounts are 30 days or younger. 39 of them addressed foxnews and 5 replies to liberal sources.

Research Questions:

A) Is the engagement more polarized when they are engaging/associating with conservative sources (Fox) or liberal sources (CNN, MSNBC, NPR, NYTimes)?

Polarization regarding topic

```
# install.packages("reshape2")
library(reshape2)
c <- tweets %>% group_by(opinion_label, ego_involvement_label) %>% summarize(n = n())
```

```
## 'summarise()' has grouped output by 'opinion_label'. You can override using the
## '.groups' argument.
```

```
c
```

```
## # A tibble: 15 x 3
## # Groups:   opinion_label [4]
##   opinion_label          ego_involvement_label      n
##   <chr>                <chr>                <int>
## 1 AGAINST student loan forgiveness cannot judge importance    1
## 2 AGAINST student loan forgiveness Not important at all      5
## 3 AGAINST student loan forgiveness Somewhat important      58
## 4 AGAINST student loan forgiveness Very important         46
## 5 cannot judge support      Not important at all      2
## 6 cannot judge support      Somewhat important      13
## 7 cannot judge support      Very important           7
## 8 FOR student loan forgiveness cannot judge importance    1
## 9 FOR student loan forgiveness Not important at all      6
## 10 FOR student loan forgiveness Somewhat important     62
## 11 FOR student loan forgiveness Very important         54
## 12 NEUTRAL support          cannot judge importance    4
## 13 NEUTRAL support          Not important at all      8
## 14 NEUTRAL support          Somewhat important     125
## 15 NEUTRAL support          Very important         76
```

```
p <- dcast(c , opinion_label~ego_involvement_label)
```

```
## Using n as value column: use value.var to override.
```

```
p <- p[c(1, 4, 3, 2),]
p
```

```
##              opinion_label cannot judge importance Not important at all
## 1 AGAINST student loan forgiveness                1                5
## 4              NEUTRAL support                    4                8
## 3      FOR student loan forgiveness                1                6
## 2              cannot judge support              NA                2
##  Somewhat important Very important
## 1                58                46
## 4                125               76
## 3                62                54
## 2                13                 7
```

The topic of student loan: *BASELINE* - Very Important against : 46 - Somewhat important against : 58 - Not at all important against : 5 - Neutral : 213 - Not at all important for : 6 - Somewhat important for : 62 - Very important for : 54

On aggregate, the population discussion student loan is not polarized. There are more neutrals (213) than support and opposition, and as we go towards the extreme ends, the totals decrease (58→ 46 support, 62→ 54 opposition)

We will use the information above as our base. Next, we focus in on the liberal and conservative sources.

```
tweets %>% group_by(politics) %>% summarize(n = n(), pct = n/nrow(tweets))
```

```
## # A tibble: 3 x 3
##   politics      n    pct
##   <chr>      <int> <dbl>
## 1 conservative  408 0.872
## 2 controlled     6 0.0128
## 3 liberal      54 0.115
```

For context, in our 468 data points on student loan tweets, conservative accounts has over 7 times the engagement than liberal accounts. 87% address the conservative source (@FoxNews) and 12% are from our 4 liberal sources. In short, when it comes to Student Loan Forgiveness, more tweeters engage with the conservative media. (With this knowledge, potential interventions of the conversation could on interventions with accounts engaging with conservative news.)

Conservative opinion

```
c <- tweets %>% filter(politics == "conservative") %>%
  group_by(opinion_label, ego_involvement_label) %>% summarize(n = n())
```

```
## 'summarise()' has grouped output by 'opinion_label'. You can override using the
## '.groups' argument.
```

```
p <- dcast(c , opinion_label~ego_involvement_label)
```

```
## Using n as value column: use value.var to override.
```

```
p <- p[c(1, 4, 3, 2),]
p
```

```
##              opinion_label cannot judge importance Not important at all
## 1 AGAINST student loan forgiveness                1                5
## 4              NEUTRAL support                    4                8
## 3      FOR student loan forgiveness                1                5
## 2              cannot judge support                NA                2
## Somewhat important Very important
## 1                55                41
## 4                109               63
## 3                52                45
## 2                12                 5
```

CONSERVATIVE ENGAGEMENT - Very Important against : 41 - Somewhat important against : 55
 - Not at all important against : 5 - Neutral : 180 - Not at all important for : 5 - Somewhat important for :
 52 - Very important for : 45

Similar to the polarization of the topic overall, within conservative engagement, majority lays within the “neutral”, and as we move out towards the two ends, opposition decreases from 55 to 41 and support decreases from 52 to 45.

LIBERAL ENGAGEMENT

```
c <- tweets %>% filter(politics == "liberal") %>%
  group_by(opinion_label, ego_involvement_label) %>% summarize(n = n())
```

```
## 'summarise()' has grouped output by 'opinion_label'. You can override using the
## '.groups' argument.
```

```
p <- dcast(c , opinion_label~ego_involvement_label)
```

```
## Using n as value column: use value.var to override.
```

```
p <- p[c(1, 4, 3, 2),]
p
```

```
##              opinion_label Not important at all Somewhat important
## 1 AGAINST student loan forgiveness                NA                3
## 4              NEUTRAL support                    NA               15
## 3      FOR student loan forgiveness                1                8
## 2              cannot judge support                NA                1
## Very important
## 1                5
## 4                10
## 3                9
## 2                2
```

LIBERAL ENGAGEMENT - Very Important against : 5 - Somewhat important against : 3 - Not at all
 important against : N/A - Neutral : 25 - Not at all important for : 1 - Somewhat important for : 8 - Very
 important for : 9

Other EDA Notes

```
c <- tweets %>% filter(politics == 'conservative')
c <- c %>% group_by(opinion_label) %>% summarize(n = n(), pct = n/nrow(c))
c
```

```
## # A tibble: 4 x 3
##   opinion_label          n    pct
##   <chr>          <int> <dbl>
## 1 AGAINST student loan forgiveness  102 0.25
## 2 cannot judge support             19 0.0466
## 3 FOR student loan forgiveness    103 0.252
## 4 NEUTRAL support                184 0.451
```

Based on the annotations, 45% are neutral in support for Student Loan Forgiveness plan. This dataset holds ambiguous tweets as well as those where opinion are only extractable if the annotators understand the social context of the sarcastic Ambiguous tweets on this subject does populate the dataset. Here are some examples: “@FoxNews So that’s why they can’t afford their student loans.”, “@FoxNews I don’t believe there’s going to be any student loan forgiveness.”, “@FoxNews My mortgage identified as a student loan.”

Those engaging with FoxNews are both equally in support and against the policy according to the annotators.

Democrats are more polarized

Polarizing: neutral vs. non-neutral; low ego involvement vs. high ego involvement. The more people care a lot about a answer, and if within that population, the both answers (yes and no).

If total population -

E) Are there non-US-based accounts engaging in the student loan conversation.

- Does not support. 66% of the tweets do not have a location tag. 10% of the 66% are meaningless locations (“where ever there’s oil”, “My house”, “in your head, MIA, etc.”). One account is under “Durban, South Africa”. However, it is obvious that the location variable is unreliable.

```
(tweets %>% mutate(has_profile_loc = ifelse(nchar(location) == 0, 0, 1)) ) %>% group_by(has_profile_loc)
```

```
## # A tibble: 2 x 2
##   has_profile_loc agg_profile_loc
##   <dbl>          <int>
## 1         0         312
## 2         1         156
```

```
location <- tweets %>% filter((nchar(location) > 0)) %>% select(location) # 156
# data.frame(table(location))

# 16 non-sense
# the world , 1
# where ever there's oil , 1
# Podunk Missery, 1
# MIA, 1
# my house, 1
```

```
# in your head, 1
# By By 1
# Earth (3 different versions), 5 people
# God's Earth Country/ God's Green Earth, 2
# cucina povera, 1
# Frenchmens Bayou, 1

# Durban, South Africa
```

Liberal opinion

```
l <- tweets %>% filter(politics == 'liberal')
l <- l %>% group_by(opinion_label) %>% summarize(n = n(), pct = n/nrow(l))
l
```

```
## # A tibble: 4 x 3
##   opinion_label          n    pct
##   <chr>             <int> <dbl>
## 1 AGAINST student loan forgiveness    8 0.148
## 2 cannot judge support                3 0.0556
## 3 FOR student loan forgiveness       18 0.333
## 4 NEUTRAL support                   25 0.463
```

Liberal sources have 46% in neutral support, very close to conservative sources (45%). Generally, Republicans are against the forgiveness plan, and Democrats are either for it or the topic isn't important in their voting decision. Our hypothesis is that tweets associated with Democratic news sources would have more polarization in support; we would see more neutral or lack of support. However, against our hypothesis, Democratic sites have less polarizing opinion. 33% are in support of student loan forgiveness and only 15% are against the plan.

Controlled opinion

```
nu <- tweets %>% filter(politics == 'controlled')
nu <- nu %>% group_by(opinion_label) %>% summarize(n = n(), pct = n/nrow(nu))
nu
```

```
## # A tibble: 2 x 3
##   opinion_label          n    pct
##   <chr>             <int> <dbl>
## 1 FOR student loan forgiveness    2 0.333
## 2 NEUTRAL support                4 0.667
```

Department of Education, our neutral source, has 6 tweets total where 4 of which are neutral, and none of them are against student loan forgiveness. This makes sense since those engaging with the Dept of Edu tend to request more information on when they can expect information. Some of the tweets are:

- “@usedgov why are my student loans not transferring from Nelnet to Mohela? I got approved for PSLF months ago and haven't heard a word since July. Can someone please do their job??”

- “@usedgov Wait times for customer service @MOHELA are currently 4 HOURS. This organization does not have the capacity to provide federal student loan service. Please investigate.”

```
tweets %>% filter(politics == 'controlled') %>% select(text)
```

```
##
## 1
## 2
## 3 @usedgov @usesgov @potus #studentloanforgiveness #ppp loans it should be brought before
## 4 @usedgov @SecCardona @usesgov @potus #studentloanforgiveness #ppp loans it should be brought before
## 5
## 6
```

Some multi colored bar plot here

visual ideas, neutral get 3 bars, which is averaged out..

B) Is there consistency in opinion across tweets that include the 3 groups in their conversation?

C) Is there consistency in polarization across tweets that include the 3 groups in their conversation?

D) Does one page get more engagement over the other?

- Yes, Fox News dominates with 87% of the tweets addressing them.

F) What concentrations of replies are by authors who engage with both new sources?

- Only 9 tweeters posted twice in this dataset. Only one engages with both liberal and conservative sources (RogerWPetersen1). From his tweets, this account appears to not support the plan while the annotators believe that his opinion is neutral. We learn that only about 0.2% would “cross-tweet” about this topic, although his opinion is consistent.

```
dup_sn <- tweets[duplicated(tweets$screen_name) ,] %>% select( screen_name)
(tweets %>% filter(screen_name %in% as.vector(dup_sn$screen_name))) %>% select(screen_name, politics, ex
```

```
##      screen_name      politics experiment_group
## 1    DahlmanCarl conservative          foxnews
## 2    DahlmanCarl conservative          foxnews
## 3     fabulosi_t controlled           usedgov
## 4     fabulosi_t controlled           usedgov
## 5 jackSpa81774793    liberal           nytimes
## 6 jackSpa81774793    liberal             cnn
## 7   johnbutler410 conservative          foxnews
## 8   johnbutler410 conservative          foxnews
## 9 michael_favreau    liberal           msnbc
```

```
## 10 michael_favreau      liberal      msnbc
## 11 PCopposition conservative foxnews
## 12 PCopposition conservative foxnews
## 13 RogerWPetersen1      liberal      msnbc
## 14 RogerWPetersen1 conservative foxnews
## 15 thomaslew13 conservative foxnews
## 16 thomaslew13 conservative foxnews
## 17 Vincent08571545      liberal      cnn
## 18 Vincent08571545      liberal      cnn
```

```
tweets %>% filter(screen_name == 'RogerWPetersen1') %>% select(screen_name, text, opinion_label)
```

```
##      screen_name
## 1 RogerWPetersen1
## 2 RogerWPetersen1
##
## 1 @MSNBC @MaddowBlog Nothing simple about it this Student loan forgiveness is UNCONSTITUTIONAL.
## 2 @FoxNews I heard that Student Debt forgiveness can not be passed as it is Un-Constitutional
##      opinion_label
## 1 NEUTRAL support
## 2 NEUTRAL support
```

##Brainstorm notes for analysis: A) Is the engagement opinions more polarized when they are engaging/associating with conservative sources or liberal sources CNN?

- The opinion on student loan forgiveness is more polarized on liberal sources than on conservative sources.

D) Does one page get more engagement over the other?

- Yes. Although our data only has 1 conservative source, 87% of the tweets are engaging with conservative sources. That is over 7 times the amount of foot traffic of the 4 liberal sources combined. With this knowledge, potential interventions of the conversation could focus on accounts that tweet @FoxNews.

E) Are there non-US-based accounts engaging in the student loan conversation.

- Does not support. 66% of the tweets do not have a location tag. 10% of the 66% are meaningless locations (“where ever there’s oil”, “My house”, “in your head, MIA, etc.”). One account is under “Durban, South Africa”. However, it is obvious that the location variable is unreliable.

F) What concentrations of replies are by authors who engage with both news sources?

- 9 tweeters posted twice in this dataset, and one engages with both liberal and conservative sources (RogerWPetersen1). We learn that only about 0.2% would “cross-tweet” about this topic, although his opinion is consistent. This tells us that individuals discussing student loan forgiveness on Twitter in less of an intellectual back and forth discourse. In other words, many of these posts are “to tell” rather than “to listen”.