EV Sentiment Analysis SURV 627

University of Maryland, College Park

Sagnik Chakravarty and Namit Shrivastava

Table of contents

List of Figures

List of Tables

The initial task is to gather all the possible sources of data from the internet regarding EV

Reddit

```
# install.packages("RedditExtractoR")
# install.packages("tidytext")
```

Collecting Reddit Data

```
ev_reddit <- find_subreddits("Electric Vehicle")
```

```
parsing URLs on page 1...
parsing URLs on page 2...
```

```
ev_df_reddit <- data.frame(ev_reddit)
ev_df_reddit_clean <- ev_df_reddit %>%
  select(subreddit,
         title,
         description,
         subscribers,
         date_utc)

ev_df_reddit_clean$date_utc <- as.Date(ev_df_reddit_clean$date_utc)
rownames(ev_df_reddit_clean) <- 1:nrow(ev_df_reddit_clean)
str(head(ev_df_reddit_clean, 1))
```

```
'data.frame':   1 obs. of  5 variables:
 $ subreddit  : chr "ChinaEV"
 $ title      : chr "ChinaEV"
 $ description: chr "A place for high quality discussion, news and due diligence on the Chine
 $ subscribers: num 275
 $ date_utc   : Date, format: "2020-03-11"
```

```
pander(dim(ev_df_reddit))
```

138 and *7*

```
pander(dim(ev_df_reddit_clean))
```

138 and *5*

```
ev_df_reddit_clean %>% ggplot(aes(x = subreddit, y = subscribers))+
  geom_bar(stat = 'identity',
           color = 'black',
           fill = 'red')+
  theme(axis.text.x = element_text(angle = 90,
                                   vjust = 0,
                                   hjust = 1,
                                   size = 2))+
  scale_y_continuous(
    breaks = pretty_breaks(n=5))
```

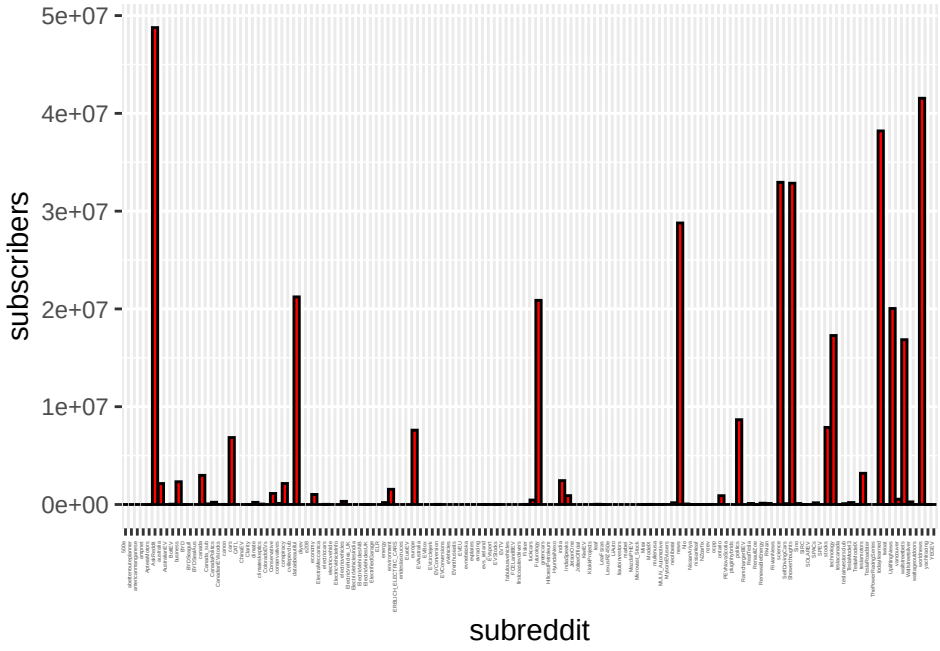

Figure 1: Subscriber count by Subreddits

Now as we can see there are 138 subreddits were we were able to find the keyword = 'Electric Vehicle'. Now working with so many subreddits is not feasible hence we would be narrowing it down to top 5 subreddits. First we would exclude the subreddit based on there descriptions then we would select the top 5 sub reddits

```r
desc_clean <- function(x){
  tk <- tibble(line = 1, text = x)
  tk <- tk %>%
    unnest_tokens(word, text) %>%
    anti_join(stop_words)
  return(tolower(str_c(tk$word, collapse = ' ')))
}

ev_df_reddit_clean <- ev_df_reddit_clean %>%
  rowwise() %>%
  mutate(cleaned_description = desc_clean(description)) %>%
  select(-description) %>%
  ungroup()
```

```r
check_ev <- function(x){
  keywords <- c('electric car',
                'electric vehicle',
                'ev',
                'motor',
                'car',
                'battery',
                'autonomous',
                'tesla'
                )
  pattern <- paste0("\\b(", paste(keywords, collapse = "|"), ")\\b")
  any(str_detect(x, pattern = pattern))
}

ev_df_reddit_clean <- ev_df_reddit_clean %>%
  rowwise()  %>%
  mutate(ev_related = if_else(check_ev(cleaned_description), TRUE, FALSE)) %>%
  ungroup()
```

```r
reddit_df <- ev_df_reddit_clean %>%
  filter(ev_related == TRUE) %>%
  arrange(desc(subscribers))  %>%
  head(n=5)
reddit_df[c(-5,-6)] %>%
  pander(caption = "Subreddits Related to EV")
```

Table 1: Subreddits Related to EV

subreddit	title	subscribers	date_utc
teslamotors	r/TeslaMotors - The original and largest Tesla community!	3191330	2010-09-04
fuckcars	fuck cars	456017	2016-02-25
electricvehicles	Electric Vehicle News and Discussion	320030	2009-04-20
TeslaModel3	Tesla Model 3 Electric Vehicle	199354	2015-01-25
Rivian	Rivian Automotive	103991	2017-01-10

Subreddit to be used for the study

```
reddit_df %>% ggplot(aes(x = subscribers,
                         y = subreddit))+
  geom_bar(stat = 'identity',
           fill = 'black')+
  theme_classic()
```

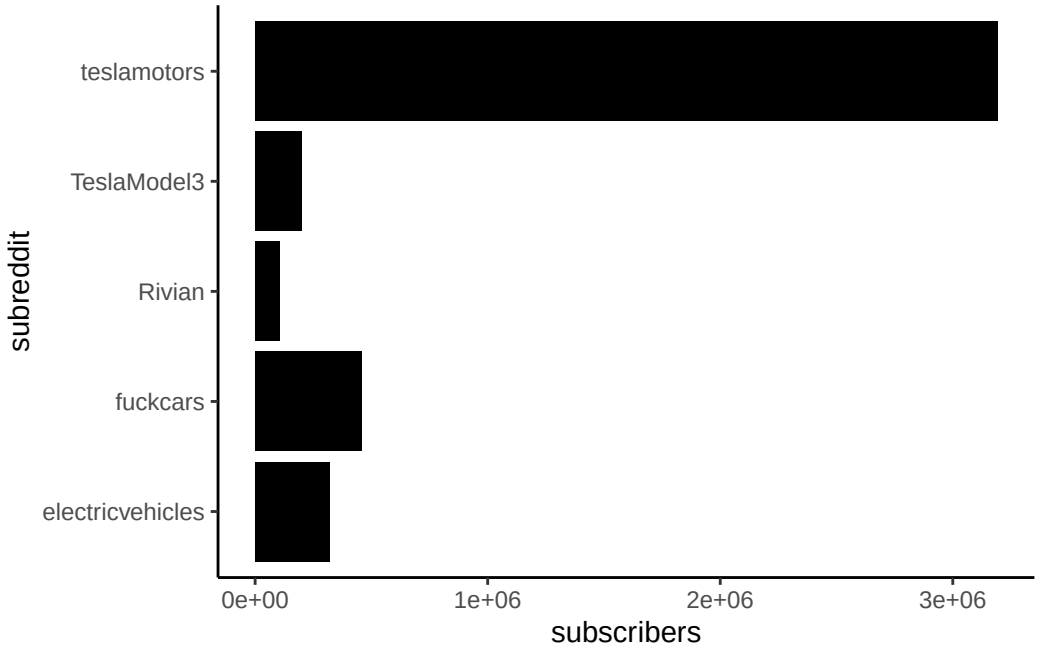

Figure 2: Subreddit to be used for the study

so the subreddits we would be looking at are

r/teslamotors, r/fuckcars, r/Rivian, r/TeslaModel3, r/electricvehicles

```
write.csv(reddit_df, 'workable_reddit_df.csv')
```