

reviews

Carl

2024-03-14

```
library(dplyr)

##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
##
##   filter, lag
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
library(stringr)

#import
reviewsorg <- read.csv("twofivereviews.csv")

reviews <- reviewsorg

#reviewer
reviewerdump <- reviews$reviewer

reviews$reviewer <- gsub("\\p{So}", "", reviews$reviewer, perl = TRUE)

#reviews
reviewscol <- reviews$reviews

#lowering all characters
reviews$reviews <- tolower(reviews$reviews)

#removing Emoji
reviews$reviews <- gsub("\\p{So}", "", reviews$reviews, perl = TRUE)

#removing none english
reviews$reviews <- gsub("[^A-Za-z ]", "", reviews$reviews)

reviewscol <- reviews$reviews

#Date
reviews$date.of.review <- str_extract(reviews$date.of.review, "\\d+\\s[A-Za-z]+\\s\\d+")
```

```

reviews$date.of.review <- as.Date(reviews$date.of.review, format = "%d %b %Y")

#ratings
dumpofrat <- reviews$ratings

#filtering the chars
reviews$ratings <- gsub("\\s*out\\s*of\\s*5\\s*stars", "", reviews$ratings)
#filter inly the intiger
reviews$ratings <- gsub("[^0-9.]", "", reviews$ratings)
#changing it to integer
reviews$ratings <- as.integer(reviews$ratings)

```