Lab 5 reviews

Abdul Azim P. Bansara

2024-03-27

```r
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
```

```r
library(stringr)
```

import

```r
imp_review <- read.csv("allproductrev.csv")

reviews <- imp_review
```

reviewer

```r
revs <- reviews$reviewer

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

#reviews

```r
rev_reviews <- reviews$reviews

#tolower the character
reviews$reviews <- tolower(reviews$reviews)

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

#removing none english language (if there is)
reviews$reviews <- gsub("[^A-Za-z ]", "",  reviews$reviews)
```

#Date

```r
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

```r
rev_ratings <- reviews$ratings

#filtering the chars
reviews$ratings <- gsub("\\s*out\\s*of\\s*5\\s*stars", "", reviews$ratings)

#filter the integer
reviews$ratings <- gsub("[^0-9.]", "", reviews$ratings)

#change it to integer
reviews$ratings <- as.integer(reviews$ratings)
```

```r
head(reviews)
```

```
##   X category          name      reviewer
## 1 1  T-shirt Fruit of the Loom          John P
## 2 2  T-shirt Fruit of the Loom Chris Morrison
## 3 3  T-shirt Fruit of the Loom          John P
## 4 4  T-shirt Fruit of the Loom           Pinky
## 5 5  T-shirt Fruit of the Loom         Stephen
## 6 6  T-shirt Fruit of the Loom          Tomasz
##
## 1
## 2 bought these for my partner who likes to wear a tshirt under his shirt these fit well and theyre ex
## 3
## 4
## 5
## 6                                                                                never fail for prove
##            verified date.of.review ratings
## 1 Verified Purchase     2024-01-04       4
## 2 Verified Purchase     2023-04-02       5
## 3 Verified Purchase     2024-01-04       3
## 4 Verified Purchase     2023-02-09       5
## 5 Verified Purchase     2023-09-28       5
## 6 Verified Purchase     2023-03-31       5
```