

Homework 2

Sam Reeves

2/10/2021

Description

This is an analysis of Sean Connin's film survey. Metadata is taken from imdb, metacritic, and wikipedia.

Import libraries, and download the data.

Check for any null values.

```
library(DBI)
library(rvest)
library(tidyverse)

data607film <- read.csv("https://raw.githubusercontent.com/TheWerefriend/data607/master/homework2/data607film.csv")
filmMeta <- read.csv("https://raw.githubusercontent.com/TheWerefriend/data607/master/homework2/filmMeta.csv")
alias <- c("theQueensGambit", "emilyInParis", "lucifer", "theUmbrellaAcademy",
           "moneyHeist", "darkDesire", "friends", "theCrown", "ratched", "dark")
colnames(data607film) <- c("timestamp", alias, "genreFavorites", "genreDislike",
                          "weeklyNetflixHours", "recommendation")
filmMeta <- cbind(alias, filmMeta)
anyNA(data607film)
```

```
## [1] FALSE
```

```
anyNA(filmMeta)
```

```
## [1] TRUE
```

Remove null values and check again:

```
data607film <- na.omit(data607film)
filmMeta <- na.omit(filmMeta)
anyNA(data607film)
```

```
## [1] FALSE
```

```
anyNA(filmMeta)
```

```
## [1] FALSE
```

Start a SQLite database and write the tables to it.

```
con <- dbConnect(RSQLite::SQLite(), ":memory")
dbWriteTable(con, "film", data607film, overwrite=TRUE)
```

```
dbWriteTable(con, "meta", filmMeta, overwrite=TRUE)
```

Pull the data from the tables, close the database connection.

```
film_table <- data.frame(dbGetQuery(con, "SELECT * FROM film"))
meta_table <- data.frame(dbGetQuery(con, "SELECT * FROM meta"))
dbDisconnect(con)
```

Focus on the heavy watchers, people who watch at least the median amount of weekly netflix. Remove the columns without metadata.

```
heavy_watchers <- film_table %>%
  filter(weeklyNetflixHours >= median(weeklyNetflixHours)) %>%
  arrange(desc(weeklyNetflixHours))

heavy_watchers <- within(heavy_watchers, rm(moneyHeist, darkDesire))
```

Order the opinion responses.

```
values = c(0:5)
opinions = c("No opinion - I haven't seen it",
             "Poor", "Fair", "Average", "Good",
             "Excellent")
names(values) = opinions
heavy_opinions <- heavy_watchers[,2:9]
```

Are the number of votes on IMDB correlated with the IMDB rating?

```
cor(meta_table$imdbVotes,
    meta_table$imdbRating)
```

```
## [1] 0.6702575
```

Are the Metacritic scores correlated with the Metacritic user rating?

```
cor(meta_table$metacriticScore,
    meta_table$metacriticRating*10)
```

```
## [1] 0.1250453
```

Get number of responses, average response, then a score out of 10 for each.

```
responses <- c()
average <- c()
out_of_ten <- c()

# values[heavy_opinions[i]]
# for opinion levels

for (i in heavy_opinions) {
  tally = 0
```

```

for (v in i) {
  if (v != "No opinion - I haven't seen it") {
    tally <- tally + 1 }
}
responses <- append(responses, tally)
average <- append(average,
                  sum(as.numeric(values[i]))/tally)
out_of_ten <- append(out_of_ten, tail(average, 1)*2)
}

```

Consolidate data

```

data <- t(meta_table) %>%
  rbind(responses) %>%
  rbind(average) %>%
  rbind(out_of_ten)

# First row to column names
colnames(data) <- data[1,]
data <- data[-1,]

```

Nobody watched The Crown.

```

heavy_opinions <- heavy_opinions[, -6]
data <- data[, -6]

rows <- rownames(data)[-c(1,6)]

```

Are the class averages close to the web ratings?

```

cor(as.numeric(data['average',]), as.numeric(data['imdbRating',]))

## [1] 0.5553707
cor(as.numeric(data['average',]), as.numeric(data['metacriticScore',]))

## [1] -0.1232937

```

Are class response rates correlated with the number of IMDB ratings?

```

cor(as.numeric(data['responses',]), as.numeric(data['imdbVotes',]))

## [1] 0.5818079

```

What are the average differences between class scores and metacritic ratings on a 1-10 scale?

```

meta_corr <- c()
count = 1
for (i in heavy_opinions) {
  tally <- 0
  differences <- c()

```

```

for (v in i) {
  if (v != "No opinion - I haven't seen it") {
    tally <- tally + 1
    diff <- abs(as.numeric(data["metacriticRating", count]) - as.numeric(values[v])*2)
    differences <- c(differences, diff)
  }
}
meta_corr <- c(meta_corr, mean(differences))
count <- count + 1
}

meta_corr

## [1] 3.750000 1.433333 1.700000 0.700000 3.000000 1.400000 2.400000

```

Conclusions

Everybody likes Friends. Our class “I’ve seen it” levels are somewhat correlated with the total numbers of votes sent to IMDB. IMDB ratings are indicative of IMDB popularity. Our class agrees a bit with IMDB ratings, and both disagree a bit with Metacritic. Class opinions were most divergent with Metacritic over Friends and The Queen’s Gambit, and the least divergent on The Umbrella Academy.

People in our sample who saw The Umbrella Academy have weird taste.