# Data607_assignment2

## karmaGyatso

## 2022-09-09

```
library(RMySQL)
```

```
## Loading required package: DBI
```

```
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
```

## Introduction

I gathered a list of six recent movie names and asked 5 friends and neigbor if they have watched any of them and rate it.

In the MYSQL database, I created 4 tables which are GENRE, MOVIES, PERSON AND REVIEWS.

In the code below, I have established a connection to MySQL database and it requires to enter a password because I didn't wanted to reveal it.

```
mysqlconnection = dbConnect(RMySQL::MySQL(),
                            dbname='movies',
                            host='127.0.0.1',
                            port=3306,
                            user='root',
                            password=rstudioapi::askForPassword("database password"))
```

## MOVIES TABLE

In the MOVIES table, there are list of 6 movie titles with their genre. The FOREIGN KEY is GENRE_ID in the MOVIES table.

## GENRE TABLE

Genre consist of 7 types and ID column in it, which is FOREIGN KEY in MOVIES table

## PERSON TABLE

Person table consist of name of people who volunteered to rate the movies. ID is FOREIGN KEY in Reviews table.

## REVIEWS TABLE

In the table, we have column id, name_id FOREIGN KEY REFERENCES PERSON(ID), movie_id FOREIGN KEY REFERENCES MOVIES(ID), and review

```
dbListTables(mysqlconnection)
```

```
## [1] "GENRE"   "MOVIES"  "PERSON"  "REVIEWS"
```

### Getting data from the Database

I used join condition to join all the tables and get one complete data set and fetch it and store in the data.frame

```
result <- dbSendQuery(mysqlconnection, "SELECT M.MOVIE_TITLE AS MOVIE_TITLE, A.GENRE_TITLE AS GENRE, C.
JOIN GENRE A ON A.ID = M.GENRE_ID
JOIN REVIEWS B ON B.MOVIE_ID = M.ID
JOIN PERSON C ON C.ID = B.NAME_ID;")
```

```
data<-fetch(result)
print(data)
```

```
##                          MOVIE_TITLE     GENRE     NAME RATING
## 1         Thor: Love and Thunder    ACTION     JOHN      4
## 2         Thor: Love and Thunder    ACTION ANGELICA      3
## 3         Thor: Love and Thunder    ACTION     JUAN     NA
## 4         Thor: Love and Thunder    ACTION    CHRIS      4
## 5         Thor: Love and Thunder    ACTION     DANA      3
## 6                    PINOCCHIO     DRAMA     JOHN     NA
## 7                    PINOCCHIO     DRAMA ANGELICA      4
## 8                    PINOCCHIO     DRAMA     JUAN      4
## 9                    PINOCCHIO     DRAMA    CHRIS      3
## 10                   PINOCCHIO     DRAMA     DANA      4
## 11                       BEAST ADVENTURE     JOHN      3
## 12                       BEAST ADVENTURE ANGELICA     NA
## 13                       BEAST ADVENTURE     JUAN      3
## 14                       BEAST ADVENTURE    CHRIS     NA
## 15                       BEAST ADVENTURE     DANA      3
## 16                   SAMARITAN    ACTION     JOHN      4
## 17                   SAMARITAN    ACTION ANGELICA     NA
## 18                   SAMARITAN    ACTION     JUAN     NA
```

```
## 19                                SAMARITAN    ACTION    CHRIS     3
## 20                                SAMARITAN    ACTION     DANA     3
## 21                                  ME-TIME    COMEDY     JOHN     2
## 22                                  ME-TIME    COMEDY ANGELICA     2
## 23                                  ME-TIME    COMEDY     JUAN    NA
## 24                                  ME-TIME    COMEDY    CHRIS    NA
## 25                                  ME-TIME    COMEDY     DANA     2
## 26 Fullmetal Alchemist the Revenge of Scar ADVENTURE     JOHN    NA
## 27 Fullmetal Alchemist the Revenge of Scar ADVENTURE ANGELICA    NA
## 28 Fullmetal Alchemist the Revenge of Scar ADVENTURE     JUAN     3
## 29 Fullmetal Alchemist the Revenge of Scar ADVENTURE    CHRIS    NA
## 30 Fullmetal Alchemist the Revenge of Scar ADVENTURE     DANA    NA
```

## Most loved

Among the six movies, most of them loved PINOCCHIO, which falls under drama GENRE. The rating is 3.75. The dataset is arranged in DESC order.

```r
data %>%
    filter(!is.na(RATING) ) %>%
      group_by(MOVIE_TITLE) %>%
        summarise(avg_rating = mean(RATING)) %>%
          arrange(desc(avg_rating))
```

```
## # A tibble: 6 x 2
##   MOVIE_TITLE                             avg_rating
##   <chr>                                        <dbl>
## 1 PINOCCHIO                                     3.75
## 2 Thor: Love and Thunder                        3.5
## 3 SAMARITAN                                     3.33
## 4 BEAST                                         3
## 5 Fullmetal Alchemist the Revenge of Scar       3
## 6 ME-TIME                                       2
```

## Most Watched

Even though PINOCCHIO is most loved, it competes with THOR: LOVE AND THUNDER in most watched catagory.

```r
data %>%
    filter(!is.na(RATING) ) %>%
      group_by(MOVIE_TITLE) %>%
        count(MOVIE_TITLE, sort=TRUE)
```

```
## # A tibble: 6 x 2
## # Groups:   MOVIE_TITLE [6]
##   MOVIE_TITLE                     n
##   <chr>                       <int>
## 1 PINOCCHIO                       4
## 2 Thor: Love and Thunder          4
## 3 BEAST                           3
```

```
## 4 ME-TIME                                    3
## 5 SAMARITAN                                   3
## 6 Fullmetal Alchemist the Revenge of Scar     1
```

## Most watched GENRE

The most watched GENRE is action. Followed by Adventure.

```
most_watched_genre <- data %>%
    filter(!is.na(RATING) ) %>%
      group_by(GENRE) %>%
        count(GENRE, sort=TRUE)
most_watched_genre
```

```
## # A tibble: 4 x 2
## # Groups:   GENRE [4]
##   GENRE          n
##   <chr>      <int>
## 1 ACTION         7
## 2 ADVENTURE      4
## 3 DRAMA          4
## 4 COMEDY         3
```