DATA607 Week 2 Assignment

James Naval

Introduction

For this assignment we were suppose to conduct a survey asking at least 5 people to rate the last 10 movies on a scale of 1-5. For this survey I asked 10 people which consist of some co-workers and friends to rate they most recent movies. After everyone complete the survey I record the result and upload. The results for the survey is available for download from either google.com or github.com using the link below:

Google: https://docs.google.com/spreadsheets/d/13nVSyE9lhiyPPZTH_ HYi9FNYIfYapg3x/edit#gid=45569196

Github: https://raw.githubusercontent.com/jnaval88/DATA607/main/Most_ recent_Movies_Rating.csv.csv

The movies that were using for the survey were: Orion and the Dark, The Beekeeper, Night Swim, Lift, Justice League: Crisis on infinite Earth Part One, The Bricklayer, Badland Hunters, Mean Girls, Role Play and The Tiger's Apprentice. For this assignment I will use the R build in SQL database and store the table into memory.

Github Data Importing into R

```
fileUrl <- getURL('https://raw.githubusercontent.com/jnaval88/DATA607/main/Most_recent_Movies_Rating.csv
Most_recent_Movies_Rating <- read.csv(text = fileUrl)
Most_recent_Movies_Rating
```

```
##    index                                      TITLE RELEASE_YEAR SCORE
## 1      1                          Orion and The Dark    2/2/2024   4.6
## 2      2                               The BeeKeeper   1/12/2024   4.5
## 3      3                                   Night Swim    1/5/2024   2.8
## 4      4                                        Lift   1/12/2024   3.5
## 5      5 Justice League: Crisis on Infinite Earth Part One    1/9/2024   3.3
## 6      6                               The Bricklayer    1/5/2024   3.1
## 7      7                              Badland Hunters   1/26/2024   3.0
## 8      8                                  Mean Girls   1/12/2024   3.8
## 9      9                                   Role Play    1/4/2024   3.9
## 10    10                       The Tiger's Apprentice    2/2/2024   4.0
##    NUMBER_OF_VOTES DURATION       MAIN_GENRE MAIN_PRODUCTION
## 1               10   1h 30m Comedy/Adventure              US
## 2               10   1h 45m  Action/Thriller              US
## 3               10   1h 38m           Horror              US
## 4               10   1h 44m    Comedy/Action              US
```

```
## 5               10   1h 33m        Animation            US
## 6               10   1h 50m   Action/Thriller           US
## 7               10   1h 47m            Action           US
## 8               10      2h     Musical/Comedy           US
## 9               10   1h 41    Action/Thriller           US
## 10              10   1h 24 m  Action/Adventure          US
```

```r
colnames(Most_recent_Movies_Rating)
```

```
## [1] "index"           "TITLE"          "RELEASE_YEAR"    "SCORE"
## [5] "NUMBER_OF_VOTES" "DURATION"       "MAIN_GENRE"      "MAIN_PRODUCTION"
```

Creating SQL Database using Build SQL in R

```r
con <- dbConnect(SQLite(), "memory")
copy_to(con,Most_recent_Movies_Rating)
```

Table into SQL using GetQuery

```r
dbGetQuery(con,'
SELECT *
FROM Most_recent_Movies_Rating
         ')
```

```
##    index                                   TITLE RELEASE_YEAR SCORE
## 1      1                      Orion and The Dark     2/2/2024   4.6
## 2      2                          The BeeKeeper    1/12/2024   4.5
## 3      3                              Night Swim     1/5/2024   2.8
## 4      4                                    Lift    1/12/2024   3.5
## 5      5 Justice League: Crisis on Infinite Earth Part One  1/9/2024   3.3
## 6      6                         The Bricklayer     1/5/2024   3.1
## 7      7                         Badland Hunters    1/26/2024   3.0
## 8      8                              Mean Girls    1/12/2024   3.8
## 9      9                               Role Play     1/4/2024   3.9
## 10    10                    The Tiger's Apprentice     2/2/2024   4.0
##    NUMBER_OF_VOTES DURATION        MAIN_GENRE MAIN_PRODUCTION
## 1               10   1h 30m Comedy/Adventure             US
## 2               10   1h 45m  Action/Thriller             US
## 3               10   1h 38m           Horror             US
## 4               10   1h 44m    Comedy/Action             US
## 5               10   1h 33m        Animation             US
## 6               10   1h 50m  Action/Thriller             US
## 7               10   1h 47m           Action             US
## 8               10      2h    Musical/Comedy             US
## 9               10   1h 41   Action/Thriller             US
## 10              10   1h 24 m  Action/Adventure            US
```

Getting Survey score in Descending Order

```r
dbGetQuery(con, '
SELECT SCORE, COUNT(SCORE)
FROM Most_recent_Movies_Rating
GROUP BY SCORE
ORDER BY COUNT(SCORE) DESC
          ')
```

```
##     SCORE COUNT(SCORE)
## 1    4.6            1
## 2    4.5            1
## 3    4.0            1
## 4    3.9            1
## 5    3.8            1
## 6    3.5            1
## 7    3.3            1
## 8    3.1            1
## 9    3.0            1
## 10   2.8            1
```

Display two Column from SQL using GetQuery

```r
dbGetQuery(con, '
SELECT Most_recent_Movies_Rating.SCORE, Most_recent_Movies_Rating.TITLE
FROM Most_recent_Movies_Rating
JOIN (SELECT SCORE, TITLE as opponent from Most_recent_Movies_Rating) B
ON Most_recent_Movies_Rating.SCORE = b.SCORE
ORDER BY Most_recent_Movies_Rating.SCORE DESC

          ')
```

```
##     SCORE                                        TITLE
## 1    4.6                           Orion and The Dark
## 2    4.5                               The BeeKeeper
## 3    4.0                       The Tiger's Apprentice
## 4    3.9                                    Role Play
## 5    3.8                                   Mean Girls
## 6    3.5                                         Lift
## 7    3.3 Justice League: Crisis on Infinite Earth Part One
## 8    3.1                               The Bricklayer
## 9    3.0                               Badland Hunters
## 10   2.8                                    Night Swim
```

```sql
SELECT Most_recent_Movies_Rating.SCORE, Most_recent_Movies_Rating.TITLE
FROM Most_recent_Movies_Rating
JOIN (SELECT SCORE, TITLE as opponent from Most_recent_Movies_Rating) B
ON Most_recent_Movies_Rating.SCORE = b.SCORE
ORDER BY Most_recent_Movies_Rating.SCORE DESC
```

Creative SQL Query from dplyr

```
Most_recent_Movies_Rating.db = tbl(con, "Most_recent_Movies_Rating")

Most_recent_Movies_Rating_query <- Most_recent_Movies_Rating.db %>%
select(SCORE, TITLE, MAIN_GENRE) %>%
left_join(Most_recent_Movies_Rating.db %>% select(MAIN_GENRE, SCORE = TITLE))

Most_recent_Movies_Rating_query %>% show_query()
```

SQL QUERY from Dplyr

```sql
SELECT
  `Most_recent_Movies_Rating_LHS`.`SCORE` AS `SCORE`,
  `Most_recent_Movies_Rating_LHS`.`TITLE` AS `TITLE`,
  `Most_recent_Movies_Rating_LHS`.`MAIN_GENRE` AS `MAIN_GENRE`
FROM `Most_recent_Movies_Rating` AS `Most_recent_Movies_Rating_LHS`
LEFT JOIN `Most_recent_Movies_Rating` AS `Most_recent_Movies_Rating_RHS`
  ON (
    `Most_recent_Movies_Rating_LHS`.`SCORE` = `Most_recent_Movies_Rating_RHS`.`TITLE` AND
    `Most_recent_Movies_Rating_LHS`.`MAIN_GENRE` = `Most_recent_Movies_Rating_RHS`.`MAIN_GENRE`
  )
```

Table 1: Displaying records 1 - 10

SCORE	TITLE	MAIN_GENRE
4.6	Orion and The Dark	Comedy/Adventure
4.5	The BeeKeeper	Action/Thriller
2.8	Night Swim	Horror
3.5	Lift	Comedy/Action
3.3	Justice League: Crisis on Infinite Earth Part One	Animation
3.1	The Bricklayer	Action/Thriller
3.0	Badland Hunters	Action
3.8	Mean Girls	Musical/Comedy
3.9	Role Play	Action/Thriller
4.0	The Tiger's Apprentice	Action/Adventure

Conclusion

For this lab I took data from a survey that I conducted and inserted it in R build in SQL database. Base on the rating we can that Orion and the Dark have the highest rating followed by the BeeKeeper then The Tiger's Apprentice.