DATA607_HW2

Dilip Ganesan

02/09/2017

DATA 607 Home Work 2 - R and SQL.

Introduction.

In this Home Work, I have tried to connect to MYSQL database, get records from table and have created data.frames

SQL Tables: Have created three tables for this home work namely REVIEWER, MOVIE_NAMES and REIVIEW_MOVIE_RATINGS Wanted to try Relational database model, so created three tables rather than one.

Approach consist of manipulating the data set in two ways.

1. *Using the SQL query joins*

2. *Using the R programming functions to achieve resulset data.frame*

Step 1: SQL Connection

```
#Connecting to MySQL database using dbConnect. Password is not masked for home work purpose.
mydb = dbConnect(MySQL(), user='root', password='mysql@123', dbname='DATA_607', host='localhost')
dbListTables(mydb)
```

```
## [1] "movie_names"         "review_movie_rating" "reviewer"
```

Step 2: Fetching records from tables.

```
# Now trying to get the 3 table data as individual data.frames.
reviewer <- dbGetQuery(mydb, "select * from reviewer")

movie_names <- dbGetQuery(mydb, "select * from movie_names")

ratings <- dbGetQuery(mydb, "select * from review_movie_rating")
```

Step 3: Checking how data got populated in data frames..

```
head(reviewer)
```

```
##   reviewer_id reviewer
## 1           1     KYLE
## 2           2   DUUBAR
## 3           3      JAI
## 4           4     JAAN
## 5           5    KELLY
## 6           6  GEORGIA
```

```r
head(movie_names)
```

```
##   movie_id              movie_names
## 1        1 The Shawshank Redemption
## 2        2             Harry Potter
## 3        3               The Matrix
## 4        4              Home Alone
## 5        5          The Godfather
## 6        6                 Titanic
```

```r
head(ratings)
```

```
##   reviewer_id movie_id ratings
## 1           1        1       5
## 2           1        2       5
## 3           1        3       4
## 4           1        4       3
## 5           1        5       2
## 6           1        6       3
```

Step 4: Using the SQL query to attain resultant data frame.

```r
review_movie_rating= dbGetQuery(mydb, "SELECT A.REVIEWER, B.MOVIE_NAMES, C.RATINGS FROM REVIEWER A, MOVI

dim(review_movie_rating)
```

```
## [1] 36  3
```

```r
# So the resultset contains 36 rows and 3 variables

# Cleaning out outstanding database connection.
dbDisconnect(mydb)
```

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

Step 5: Printing Resultant Data Frame.

```r
htmlTable(review_movie_rating, caption = '2017 SPRING CUNY MSDA CLASS MOVIE REVIEW RATINGS')
```

2017 SPRING CUNY MSDA CLASS MOVIE REVIEW RATINGS

REVIEWER

MOVIE_NAMES

RATINGS

1

DUUBAR

The Shawshank Redemption

5

2

DUUBAR

Harry Potter

5

3

DUUBAR

The Matrix

4

4

DUUBAR

Home Alone

3

5

DUUBAR

The Godfather

5

6

DUUBAR

Titanic

3

7

GEORGIA

The Shawshank Redemption

5

8

GEORGIA

Harry Potter

2

9

GEORGIA

The Matrix

3

10

GEORGIA

Home Alone

5

11

GEORGIA

The Godfather

1

12

GEORGIA

Titanic

5

13

JAAN

The Shawshank Redemption

5

14

JAAN

Harry Potter

4

15

JAAN

The Matrix

3

16

JAAN

Home Alone

5

17

JAAN

The Godfather

4

18

JAAN

Titanic

5

19

JAI

The Shawshank Redemption

5

20

JAI

Harry Potter

4

21

JAI

The Matrix

3

22

JAI

Home Alone

3

23

JAI

The Godfather

2

24

JAI

Titanic

1

25

KELLY

The Shawshank Redemption

5

26

KELLY

Harry Potter

2

27

KELLY

The Matrix

1

28

KELLY

Home Alone

3

29

KELLY

The Godfather

2

30

KELLY

Titanic

1

31

KYLE

The Shawshank Redemption

5

32

KYLE

Harry Potter

5

33

KYLE

The Matrix

4

34

KYLE

Home Alone

3

35

KYLE

The Godfather

2

36

KYLE

Titanic

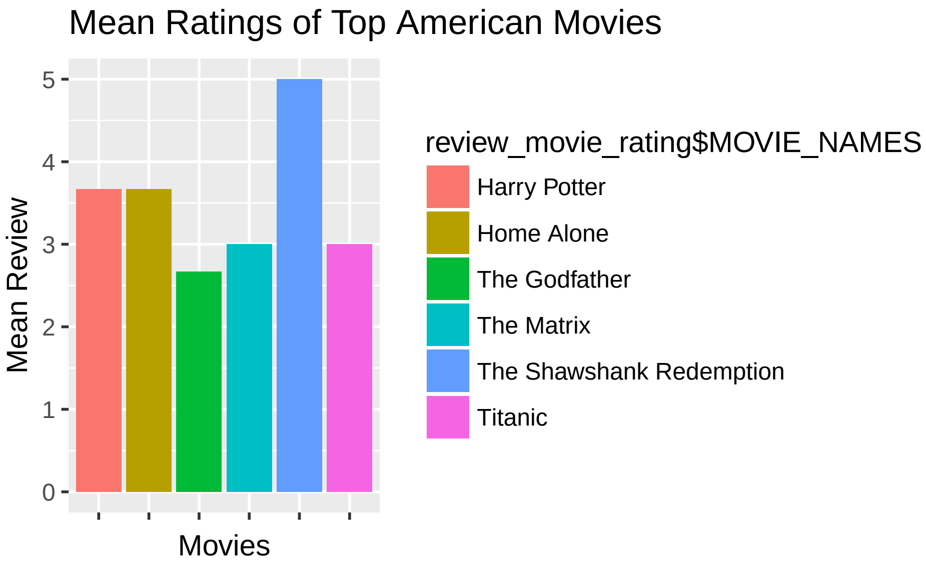
3

Step 6: Using R Programming functions to create Resultant Data Frame.

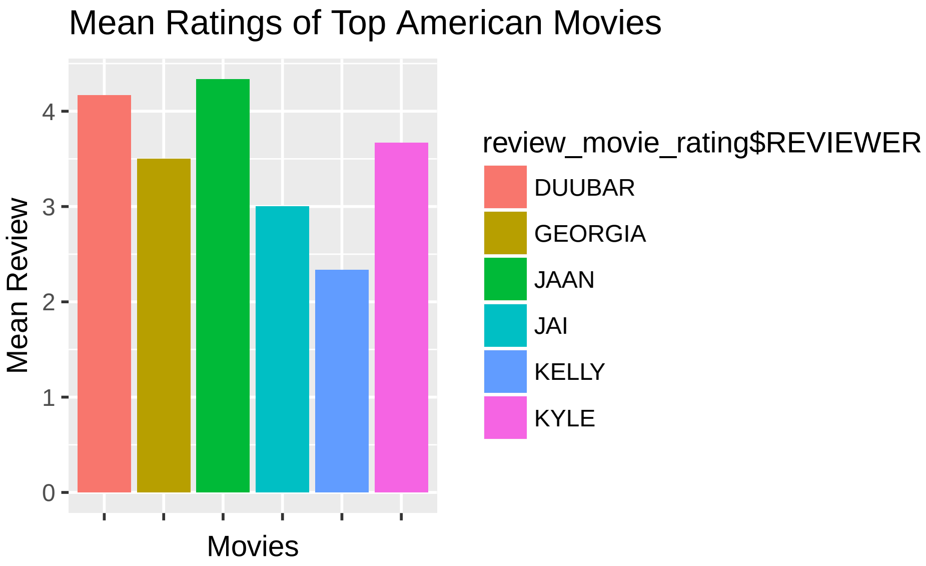
```
# Used the merge function to merge the data set.

dt<-merge(reviewer,merge(movie_names,ratings,by="movie_id"), by="reviewer_id")
finaldata= subset(dt,select = c(2,4,5))
```

Step 7: Using ggPlot to create a small plot..

```
# Plot shows the Mean Ratings across the movies.
ggplot(review_movie_rating) + geom_bar(aes(review_movie_rating$MOVIE_NAMES, review_movie_rating$RATINGS
```



```
ggplot(review_movie_rating) + geom_bar(aes(review_movie_rating$REVIEWER , review_movie_rating$RATINGS,
```



- File creation date: 2017-02-12
- R version 3.3.2 (2016-10-31)
- R version (short form): 3.3.2
- `mosaic` package version: 0.14.4
- Additional session information