Hantz_Angrand_HW2

Hantz Angrand

September 9, 2018

Database Creation

First, we create a database named boxoffice that has 3 tables: -movie_goers: ID, first_name, last_name, address, revenue -movie: movie_ID, title, genre -ratings: ID, movie_ID, ratings,comments

we upload data for all the tables.See script in github for sql statements.

Upload libraries

```
library(RMySQL)
```

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

```
library(DBI)
```

Connecting to the database locally

```
condb= dbConnect(MySQL(), user='root', password='Clorox1!', dbname='boxofiice')
```

Query table and create data frame

```
#user table
user<-dbGetQuery(condb, "select * from movie_goers")
```

```
## Warning in .local(conn, statement, ...): Decimal MySQL column 4 imported as
## numeric
```

```
#movies table
movies<-dbGetQuery(condb, "select * from movies")

#ratings table
ratings<-dbGetQuery(condb, "select * from ratings")

#Join ratings and movies
movies_ratings<-dbGetQuery(condb, "select a.movie_ID, a.title, b.ratings from movies a
                           join ratings b on a.movie_ID=b.movie_ID")

#Create data frame
mrating_df<-as.data.frame(movies_ratings)

#get first data from data frame
head(mrating_df)
```

```
##   movie_ID                                 title ratings
## 1   157200                    Money Monster (2016)       2
```

```
## 2   158314    Daniel Tosh: Completely Serious (2007)       3
## 3   158528                        The Shallows (2016)       2
## 4   159755 Popstar: Never Stop Never Stopping (2016)       4
## 5   162672                        Mohenjo Daro (2016)       3
## 6   157296                        Finding Dory (2016)       3
```

Data Manipulation and summary

```
#calculating the highest and the lowest rating

queries<-"select a.title, min(ratings) as lowest, max(ratings) as highest from movies a join ratings b o
hl_ratings<-dbGetQuery(condb,queries)

head(hl_ratings)
```

```
##                                   title lowest highest
## 1 Daniel Tosh: Completely Serious (2007)      1       4
## 2                     Finding Dory (2016)      1       4
## 3                       I Am Wrath (2016)      2       3
## 4                     Mohenjo Daro (2016)      2       4
## 5                    Money Monster (2016)      1       3
## 6                     Mother's Day (2016)      2       4
```

```
#Calculate the average
q<-"select a.title, avg(ratings) as average_ratings from movies a join
ratings b on a.movie_ID=b.movie_ID group by title order by title"

avg_ratings<-dbGetQuery(condb, q)
```

```
## Warning in .local(conn, statement, ...): Decimal MySQL column 1 imported as
## numeric
```

```
head(avg_ratings)
```

```
##                                   title average_ratings
## 1 Daniel Tosh: Completely Serious (2007)          2.4286
## 2                     Finding Dory (2016)          2.5000
## 3                       I Am Wrath (2016)          2.7500
## 4                     Mohenjo Daro (2016)          3.0000
## 5                    Money Monster (2016)          2.2000
## 6                     Mother's Day (2016)          3.0000
```