

Assignment2

Ben Wolin

2024-09-06

Introduction

The purpose of this assignment is to connect to an SQL database to retrieve a data set containing review scores for 6 movies. Then, to store this data into an R dataframe.

Install Libraries

```
library(DBI)
library(RMySQL)
library(RODBC)
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
```

```
library("dbplyr")
```

```
##
## Attaching package: 'dbplyr'

## The following objects are masked from 'package:dplyr':
##
##   ident, sql
```

Step 1: Create and store data set in the MySQL Database

This can be done by running the SQL script at this link: <https://github.com/bwolin99/TestRepo/blob/main/Assignment2/movies.sql>

Step 2: Retriving the data from SQL

Here we connect to the local MySQL statbase using the dbConnect function. Then we load the “movies” table into an R dataframe.

```
fin = dbConnect(RMySQL::MySQL(),
                dbname= 'movies',
                host= '127.0.0.1',
                port=3306,
                user='root',
                password='Bdwbdw993728')

pull = dbSendQuery(fin, "select * from movies")

movies <- fetch(pull)

movies
```

```
##   reviewer saltburn poor_things long_legs twisters fallguy
## 1   Peter      4         NA         NA         1         2
## 2   Sarah      5         4         1         3         2
## 3   Ryan       2         5         4         3        NA
## 4   Paul      NA         1         NA         5         5
## 5   Mickey    NA         5         2         2        NA
```

Step 3: Create an approach for missing data

In this part we will replace all of the null values in this data. This will be done by replacing all null values with something that will indicate the reviewer has not yet seen it.

```
movies[is.na(movies)] <- 'Not Viewed'
movies
```

```
##   reviewer saltburn poor_things long_legs twisters fallguy
## 1   Peter      4 Not Viewed Not Viewed         1         2
## 2   Sarah      5         4         1         3         2
## 3   Ryan       2         5         4         3 Not Viewed
## 4   Paul Not Viewed         1 Not Viewed         5         5
## 5   Mickey Not Viewed         5         2         2 Not Viewed
```