Intro to Data Science - HW 10


```
# Enter your name here: Sidali Mohamed
```

Attribution statement: (choose only one and delete the rest)

```
# 1. I did this homework by myself, with help from the book and the professor.
```

Association mining is a versatile ML technique that can be used to solve a wide range of data-related problems. Although it is commonly associated with identifying relationships between various products in customer shopping data, it has numerous other potential applications.

In this particular homework assignment, we will revisit the movie data from last week and use association mining techniques to uncover patterns that may be indicative of a film's potential **to become a box office success**.

Part 1: Explore Data Set

A. To create a dataframe called **movies_df** using the **rio package**, we will need to read the contents of the following URL into R. We can accomplish this by following the same procedure we used last week: https://data-science-intro.s3.us-east-2.amazonaws.com/movies.xlsx

```
#install.packages('rio')
library(rio)

movies_df = import("https://data-science-intro.s3.us-east-2.amazonaws.com/movies.xlsx")
```

B. Make sure there are **1,374** rows and **15** columns in **movies_df**.

```
dim(movies_df)
```

```
## [1] 1374   15
```

C. Let's focus on some critical variables present within the dataset. These variables will be converted to **factor variables**. Execute the provided command and be sure to provide a comment that explains the conversion process for each numeric variable:

```
movies_short_df <- data.frame(belongs_to_collection=as.factor(movies_df$belongs_to_collection),
                              budget_low=as.factor(movies_df$budget<median(movies_df$budget)),
                              homepage=as.factor(movies_df$homepage),
                              original_language_en=as.factor(movies_df$original_language_en),
                              popularity_low=as.factor(movies_df$popularity<mean(movies_df$popularity)),
                              production_companies=as.factor(movies_df$production_companies),
                              runtime_long=as.factor(movies_df$runtime>median(movies_df$runtime)),
                              tagline=as.factor(movies_df$tagline),
                              success=as.factor(movies_df$success))
```

D. Use the **table( )** command to calculate the number of successful movies in the **success** variable.

```
MS <- table(movies_df$success)
```

E. To represent the results of problem D as percentages, we can use the **prop.table( )** command. We will need to pass the output of the **table( )** command as an argument to the **prop.table( )** command:

```
is.numeric(movies_df$success)
```

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

```
MS <- prop.table(movies_df$success)
MS
```

```
##    [1] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##    [7] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##   [13] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##   [19] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##   [25] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##   [31] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##   [37] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##   [43] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##   [49] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##   [55] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##   [61] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##   [67] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##   [73] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##   [79] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##   [85] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##   [91] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##   [97] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [103] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [109] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [115] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [121] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [127] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [133] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [139] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [145] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [151] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [157] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [163] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [169] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [175] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [181] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [187] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [193] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [199] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [205] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [211] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [217] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [223] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [229] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [235] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [241] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [247] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [253] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [259] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [265] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [271] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [277] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [283] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [289] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [295] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [301] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [307] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [313] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [319] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [325] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [331] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [337] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [343] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [349] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [355] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [361] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [367] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [373] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [379] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [385] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [391] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [397] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [403] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [409] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [415] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [421] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [427] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [433] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [439] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [445] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [451] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [457] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [463] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [469] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [475] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [481] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [487] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [493] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [499] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [505] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [511] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [517] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [523] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [529] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [535] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [541] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [547] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [553] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [559] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [565] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [571] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [577] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [583] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [589] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [595] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [601] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [607] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [613] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [619] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [625] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [631] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [637] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [643] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [649] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [655] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##  [661] 0.00000000 0.00000000 0.00000000 0.00000000 0.00146647 0.00146647
##  [667] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
##  [673] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
##  [679] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
##  [685] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
##  [691] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
##  [697] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
##  [703] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
##  [709] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
##  [715] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
##  [721] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
##  [727] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
##  [733] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
##  [739] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
##  [745] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
##  [751] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
##  [757] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
##  [763] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
##  [769] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
##  [775] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
##  [781] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
##  [787] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
##  [793] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
##  [799] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
##  [805] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
##  [811] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
##  [817] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
##  [823] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
##  [829] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
##  [835] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
##  [841] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
##  [847] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
##  [853] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
##  [859] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
##  [865] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
##  [871] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
##  [877] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
##  [883] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
##  [889] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
##  [895] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
##  [901] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
##  [907] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
##  [913] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
##  [919] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
##  [925] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
##  [931] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
##  [937] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
##  [943] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
##  [949] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
##  [955] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
##  [961] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
##  [967] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
##  [973] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
##  [979] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
##  [985] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
##  [991] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
##  [997] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
## [1003] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
## [1009] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
## [1015] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
## [1021] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
## [1027] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
## [1033] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
## [1039] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
## [1045] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
## [1051] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
## [1057] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
## [1063] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
## [1069] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
## [1075] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
## [1081] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
## [1087] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
## [1093] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
## [1099] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
## [1105] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
## [1111] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
## [1117] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
## [1123] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
## [1129] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
## [1135] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
## [1141] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
## [1147] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
## [1153] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
## [1159] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
## [1165] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
## [1171] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
## [1177] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
## [1183] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
## [1189] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
## [1195] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
## [1201] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
## [1207] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
## [1213] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
## [1219] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
## [1225] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
## [1231] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
## [1237] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
## [1243] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
## [1249] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
## [1255] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
## [1261] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
## [1267] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
## [1273] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
## [1279] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
## [1285] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
## [1291] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
## [1297] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
## [1303] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
## [1309] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
## [1315] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
## [1321] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
## [1327] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
## [1333] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
## [1339] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
## [1345] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
## [1351] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
## [1357] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
## [1363] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
## [1369] 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647 0.00146647
```

F. Use the same technique to obtain percentages for the **runtime_long** and **budget_low** variables and briefly comment on the output of each command:

```
runtime_long <- prop.table(movies_df$runtime)
```

```
budget_low <- prop.table(movies_df$budget)
```

Part 2: Converting the Data Frame into a Transactions Matrix

A. Install and library two packages: **arules** and **arulesViz**.

```
#install.packages("arules")
#install.packages("arulesViz")
#library(arules)
#library(arulesViz)
```

B. Convert the **movies_short_df** dataframe into a **sparse transactions matrix** called **movies_X**.

```
movies_X <- summary(movies_short_df)
```

C. Use the **itemFrequency( )** and **itemFrequencyPlot( )** commands to explore the contents of **movies_X**. What do you see?

```
#itemFrequency(movies_X)
```

```
#itemFrequencyPlot(movies_X)
```

```
# I am getting an error that itemFrequencyPlot and itemFrequency are not found
```

D. This is a fairly large dataset; write code to explore only the first 10 observations of the **movies_X** transaction matrix using **inspect()**:

E. Explain the difference between **movies_short_df** and **movies_X** in a block comment:

Part 3: Use arules to discover patterns

To provide context for the upcoming task, it's important to note that:

Support is defined as the proportion of times a particular set of items occurs concerning the entire dataset.

Confidence is the proportion of times the consequent occurs when the antecedent is present.

A. We will use the **apriori()** function to create a set of rules that have a **support greater than 0.05** and a **confidence level greater than 0.65**. Our goal is to predict the features associated with movies that were successful at the box office.

Hint: It is necessary to **specify the right-hand side rule (rhs)**.

B. After generating the set of rules that fulfill the conditions mentioned above, please examine the output using the **inspect()** or **inspectDT()** functions. From here, describe any two rules that the algorithm found.

C. Based on your analysis of the data in this homework, briefly explain which variables appear to be associated with box office success: