Proyecto 1

Gerardo Pineda 22880, Daniel Rayo

2025-01-28

Analisis Exploratorio

Resumen del conjunto de datos

```
##        id              budget               genres            homePage
##  Min.   :     5   Min.   :        0   Length:10000       Length:10000
##  1st Qu.: 12286   1st Qu.:        0   Class :character   Class :character
##  Median :152558   Median :   500000   Mode  :character   Mode  :character
##  Mean   :249877   Mean   : 18551632
##  3rd Qu.:452022   3rd Qu.: 20000000
##  Max.   :922260   Max.   :380000000
##  productionCompany  productionCompanyCountry productionCountry
##  Length:10000       Length:10000             Length:10000
##  Class :character   Class :character         Class :character
##  Mode  :character   Mode  :character         Mode  :character
##
##
##
##     revenue            runtime         video           director
##  Min.   :0.000e+00   Min.   :  0.0   Mode :logical   Length:10000
##  1st Qu.:0.000e+00   1st Qu.: 90.0   FALSE:9430      Class :character
##  Median :1.631e+05   Median :100.0   TRUE :84        Mode  :character
##  Mean   :5.674e+07   Mean   :100.3   NA's :486
##  3rd Qu.:4.480e+07   3rd Qu.:113.0
##  Max.   :2.847e+09   Max.   :750.0
##     actors           actorsPopularity   actorsCharacter    originalTitle
##  Length:10000       Length:10000       Length:10000       Length:10000
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##     title          originalLanguage     popularity         releaseDate
##  Length:10000       Length:10000       Min.   :    4.258   Length:10000
##  Class :character   Class :character   1st Qu.:   14.578   Class :character
##  Mode  :character   Mode  :character   Median :   21.906   Mode  :character
##                                        Mean   :   51.394
##                                        3rd Qu.:   40.654
##                                        Max.   :11474.647
##     voteAvg          voteCount        genresAmount    productionCoAmount
##  Min.   : 1.300   Min.   :     1   Min.   : 0.000   Min.   : 0.000
```

1

```
##   1st Qu.: 5.900   1st Qu.:   120   1st Qu.: 2.000   1st Qu.: 2.000
##   Median : 6.500   Median :   415   Median : 3.000   Median : 3.000
##   Mean   : 6.483   Mean   :  1342   Mean   : 2.596   Mean   : 3.171
##   3rd Qu.: 7.200   3rd Qu.:  1316   3rd Qu.: 3.000   3rd Qu.: 4.000
##   Max.   :10.000   Max.   :30788   Max.   :16.000   Max.   :89.000
##   productionCountriesAmount  actorsAmount    castWomenAmount
##   Min.   :  0.000           Min.   :     0   Length:10000
##   1st Qu.:  1.000           1st Qu.:    13   Class :character
##   Median :  1.000           Median :    21   Mode  :character
##   Mean   :  1.751           Mean   :  2148
##   3rd Qu.:  2.000           3rd Qu.:    36
##   Max.   :155.000           Max.   :919590
##   castMenAmount
##   Length:10000
##   Class :character
##   Mode  :character
##
##
##
```

Tipo de cada variable

```r
str(movies)
```

```
## 'data.frame':    10000 obs. of  27 variables:
##  $ id                      : int  5 6 11 12 13 14 15 16 18 19 ...
##  $ budget                  : int  4000000 21000000 11000000 94000000 55000000 15000000 839727 128000
##  $ genres                  : chr  "Crime|Comedy" "Action|Thriller|Crime" "Adventure|Action|Science
##  $ homePage                : chr  "https://www.miramax.com/movie/four-rooms/" NA "http://www.starwa
##  $ productionCompany       : chr  "Miramax|A Band Apart" "Universal Pictures|Largo Entertainment|JVC
##  $ productionCompanyCountry : chr  "US|US" "US|US|JP" "US|US" "US" ...
##  $ productionCountry       : chr  "United States of America" "Japan|United States of America" "Unite
##  $ revenue                 : num  4.26e+06 1.21e+07 7.75e+08 9.40e+08 6.77e+08 ...
##  $ runtime                 : int  98 110 121 100 142 122 119 141 126 149 ...
##  $ video                   : logi  FALSE FALSE NA NA FALSE FALSE ...
##  $ director                : chr  "Allison Anders|Alexandre Rockwell|Robert Rodriguez|Quentin Taran
##  $ actors                  : chr  "Tim Roth|Jennifer Beals|Antonio Banderas|Valeria Golino|David Pro
##  $ actorsPopularity        : chr  "22.225|23.519|17.816|19.893|9.027|7.147|7.769|6.476|7.906|12.838
##  $ actorsCharacter         : chr  "Ted the Bellhop|Angela|Man|Athena|Siegfried|Jezebel|Diana|Elspet
##  $ originalTitle           : chr  "Four Rooms" "Judgment Night" "Star Wars" "Finding Nemo" ...
##  $ title                   : chr  "Four Rooms" "Judgment Night" "Star Wars" "Finding Nemo" ...
##  $ originalLanguage        : chr  "en" "en" "en" "en" ...
##  $ popularity              : num  20.9 9.6 100 134.4 58.8 ...
##  $ releaseDate             : chr  "1995-12-09" "1993-10-15" "1977-05-25" "2003-05-30" ...
##  $ voteAvg                 : num  5.7 6.5 8.2 7.8 8.5 8 8 7.9 7.5 8.2 ...
##  $ voteCount               : int  2077 223 16598 15928 22045 9951 4253 1335 8726 1963 ...
##  $ genresAmount            : int  2 3 3 2 3 1 2 2 5 2 ...
##  $ productionCoAmount      : int  2 3 2 1 2 2 2 26 2 1 ...
##  $ productionCountriesAmount: int  1 2 1 1 1 1 1 12 1 1 ...
##  $ actorsAmount            : int  25 15 105 24 76 40 152 29 117 24 ...
##  $ castWomenAmount         : chr  "15" "3" "5" "5" ...
##  $ castMenAmount           : chr  "9" "9" "62" "18" ...
```

```
sapply(movies, class)
```

```
##                       id                  budget                    genres
##                "integer"               "integer"               "character"
##                 homePage       productionCompany productionCompanyCountry
##              "character"             "character"               "character"
##        productionCountry                 revenue                   runtime
##              "character"               "numeric"                 "integer"
##                    video                director                    actors
##                "logical"             "character"               "character"
##          actorsPopularity          actorsCharacter             originalTitle
##              "character"             "character"               "character"
##                    title          originalLanguage                popularity
##              "character"             "character"                 "numeric"
##              releaseDate                  voteAvg                 voteCount
##              "character"               "numeric"                 "integer"
##             genresAmount        productionCoAmount productionCountriesAmount
##                "integer"               "integer"                 "integer"
##             actorsAmount          castWomenAmount             castMenAmount
##                "integer"             "character"               "character"
```