

Proyecto1

Juan Luis Solórzano (carnet: 201598)

Mica (carnet: XXXXXXXX)

2025-01-20

1. (3 puntos) Haga una exploración rápida de sus datos, para eso haga un resumen de su conjunto de datos.

1. (3 puntos) Haga una exploración rápida de sus datos, para eso haga un resumen de su conjunto de datos.

```
pelis$castMenAmount <- as.numeric(pelis$castMenAmount)
```

```
## Warning: NAs introducidos por coerción
```

```
pelis$castWomenAmount <- as.numeric(pelis$castWomenAmount)
```

```
## Warning: NAs introducidos por coerción
```

```
summary(pelis)
```

```
##          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
## 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  castMenAmount
## Min.   : 0.000                  Min.   :    0    Min.   :    0    Min.   :    0
## 1st Qu.: 1.000                  1st Qu.:   13    1st Qu.:    3    1st Qu.:    7
## Median : 1.000                  Median :   21    Median :    6    Median :   12
## Mean   : 1.751                  Mean   :  2148    Mean   :  7004    Mean   : 16557
## 3rd Qu.: 2.000                  3rd Qu.:   36    3rd Qu.:   10    3rd Qu.:   20
## Max.   :155.000                Max.   :919590    Max.   :922162    Max.   :922017
##                                     NA's   :37      NA's   :162

```

2. (5 puntos) Diga el tipo de cada una de las variables (cualitativa ordinal o nominal, cuantitativa continua, cuantitativa discreta)

```

##                                     tipo
## id                                cualitativa nominal
## budget                           cuantitativa discreta
## genres                            cualitativa nominal
## homePage                          cualitativa nominal
## productionCompany                  cualitativa nominal
## productionCompanyCountry           cualitativa nominal
## productionCountry                  cualitativa nominal
## revenue                           cuantitativa discreta
## runtime                           cuantitativa discreta
## video                             cualitativa nominal
## director                          cualitativa nominal
## actors                            cualitativa nominal
## actorsPopularity                   cuantitativa continua
## actorsCharacter                    cualitativa nominal
## originalTitle                      cualitativa nominal
## title                             cualitativa nominal
## originalLanguage                   cualitativa nominal
## popularity                         cuantitativa continua
## releaseDate                       cualitativa ordinal
## voteAvg                           cuantitativa continua
## voteCount                         cuantitativa discreta

```

```
## genresAmount          cuantitativa discreta
## productionCoAmount    cuantitativa discreta
## productionCountriesAmount cuantitativa discreta
## actorsAmount          cuantitativa discreta
## castWomenAmount       cuantitativa discreta
## castMenAmount         cuantitativa discreta
```

3. (6 puntos) Investigue si las variables cuantitativas siguen una distribución normal y haga una tabla de frecuencias de las variables cualitativas. Explique todos los resultados.

```
##                               Variable P_value
## voteAvg                      voteAvg 3.7e-24
## voteCount                    voteCount 3.7e-24
## genresAmount                 genresAmount 3.7e-24
## productionCoAmount           productionCoAmount 3.7e-24
## productionCountriesAmount    productionCountriesAmount 3.7e-24
## actorsAmount                 actorsAmount 3.7e-24
## castWomenAmount              castWomenAmount 3.7e-24
## castMenAmount                castMenAmount 3.7e-24
```

4. Responda las siguientes preguntas:

4.1. (3 puntos) ¿Cuáles son las 10 películas que contaron con más presupuesto?

Las 10 películas con mayor presupuesto fueron:

```
##                               originalTitle    budget
## 717  Pirates of the Caribbean: On Stranger Tides 380000000
## 4711                               Avengers: Age of Ultron 365000000
## 5953                               Avengers: Endgame 356000000
## 164   Pirates of the Caribbean: At World's End 300000000
## 4954                               Justice League 300000000
## 5954                               Avengers: Infinity War 300000000
## 608   Superman Returns 270000000
## 3792                               Tangled 260000000
## 7135                               The Lion King 260000000
## 281   Spider-Man 3 258000000
```

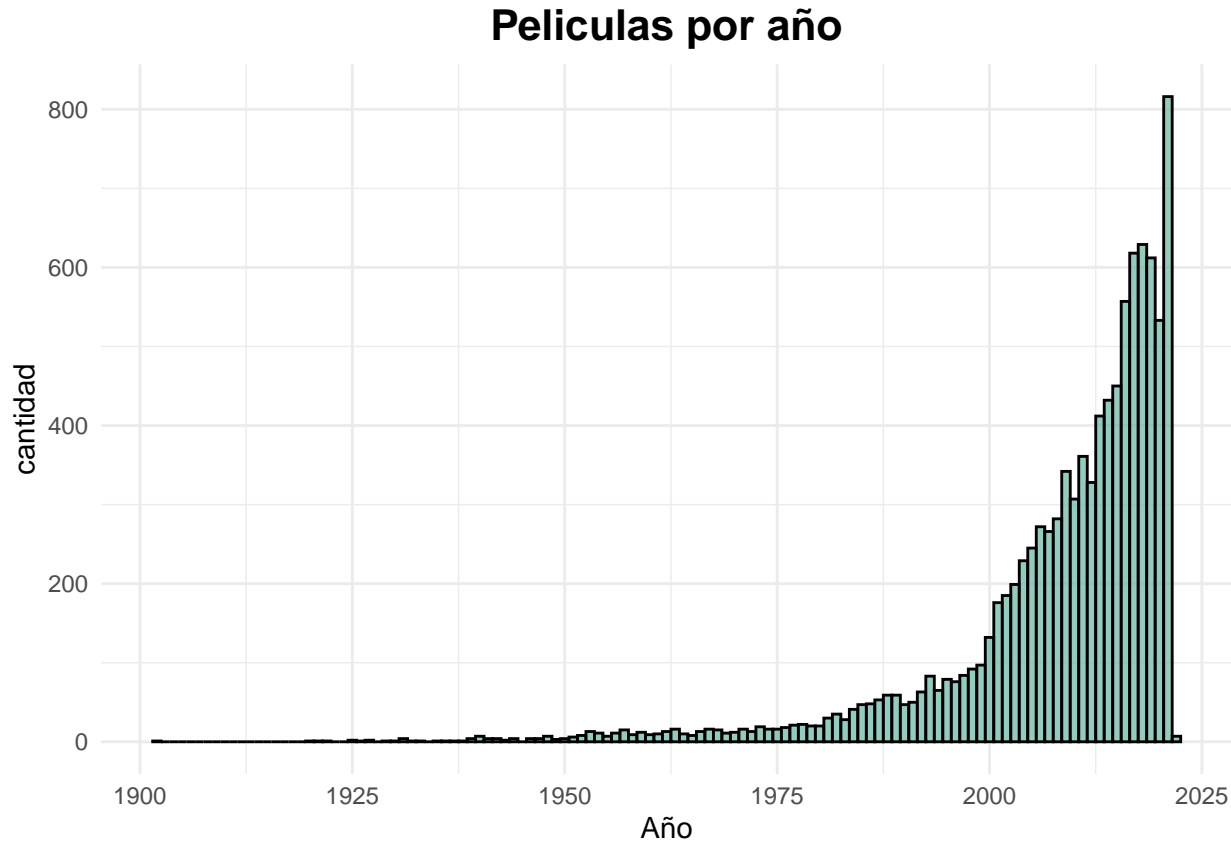
4.2. (3 puntos) ¿Cuáles son las 10 películas que más ingresos tuvieron?

4.3. (3 puntos) ¿Cuál es la película que más votos tuvo?

La película que más votos tuvo fue Inception.

4.4. (3 puntos) ¿Cuál es la peor película de acuerdo a los votos de todos los usuarios?

4.5. (8 puntos) ¿Cuántas películas se hicieron en cada año? ¿En qué año se hicieron más películas? Haga un gráfico de barras.



El año que se rodaron más películas fue el 2021

4.6. (9 puntos) ¿Cuál es el género principal de las 20 películas más recientes? ¿Cuál es el género principal que predomina en el conjunto de datos? Representéelo usando un gráfico. ¿A qué género principal pertenecen las películas más largas?

4.7. (8 puntos) ¿Las películas de qué género principal obtuvieron mayores ganancias?

4.8. (3 puntos) ¿La cantidad de actores influye en los ingresos de las películas? ¿Se han hecho películas con más actores en los últimos años?

4.9. (3 puntos) ¿Es posible que la cantidad de hombres y mujeres en el reparto influya en la popularidad y los ingresos de las películas?

4.10. (8 puntos) ¿Quiénes son los directores que hicieron las 20 películas mejor calificadas?

4.11. (8 puntos) ¿Cómo se correlacionan los presupuestos con los ingresos? ¿Los altos presupuestos significan altos ingresos? Haga los gráficos que necesite, histograma, diagrama de dispersión.

4.12. (5 puntos) ¿Se asocian ciertos meses de lanzamiento con mejores ingresos?

4.13. (6 puntos) ¿En qué meses se han visto los lanzamientos con mejores ingresos? ¿Cuántas películas, en promedio, se han lanzado por mes?

e(datos\$mes)

- 4.14. (7 puntos) ¿Cómo se correlacionan las calificaciones con el éxito comercial?
- 4.15. (5 puntos) ¿Qué estrategias de marketing, como videos promocionales o páginas oficiales, generan mejores resultados?
- 4.16. (4 puntos) ¿La popularidad del elenco está directamente correlacionada con el éxito de taquilla?
5. (¡10 puntos extras!) Genere usted otras seis preguntas que le parezcan interesantes porque le permitan realizar otras exploraciones y respóndalas. No puede repetir ninguna de las instrucciones anteriores.