

Proyecto 1

Gerardo Pineda 22880, Daniel Rayo

2025-01-28

```
##
## Adjuntando el paquete: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

## [1] "LC_COLLATE=es_ES.UTF-8;LC_CTYPE=es_ES.UTF-8;LC_MONETARY=es_ES.UTF-8;LC_NUMERIC=C;LC_TIME=es_ES."
```

Analisis Exploratorio

1. Resumen del conjunto de datos

Haga una exploración rápida de sus datos, para eso haga un resumen de su 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
## 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
##
##
##

```

2. Tipo de cada variable

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

	Variable	Tipo
id	id	integer
budget	budget	integer
genres	genres	character
homePage	homePage	character
productionCompany	productionCompany	character
productionCompanyCountry	productionCompanyCountry	character
productionCountry	productionCountry	character
revenue	revenue	numeric
runtime	runtime	integer

	Variable	Tipo
video	video	logical
director	director	character
actors	actors	character
actorsPopularity	actorsPopularity	character
actorsCharacter	actorsCharacter	character
originalTitle	originalTitle	character
title	title	character
originalLanguage	originalLanguage	character
popularity	popularity	numeric
releaseDate	releaseDate	character
voteAvg	voteAvg	numeric
voteCount	voteCount	integer
genresAmount	genresAmount	integer
productionCoAmount	productionCoAmount	integer
productionCountriesAmount	productionCountriesAmount	integer
actorsAmount	actorsAmount	integer
castWomenAmount	castWomenAmount	character
castMenAmount	castMenAmount	character

3. Analisis de las variables del dataset

Cuantitativas Usando la prueba de Shapiro-will en las variables cuantitativas utilizando un 0.05 de confianza, podemos verificar que todas las variables fallan la prueba, teniendo valores que se pasan del limite de tolerancia de 0.05.

##	esNormal	P_Value
## popularity	No	2.05651644805339e-92
## budget	No	2.00602703821519e-77
## revenue	No	4.55438621378301e-84
## runtime	No	1.71154317470535e-59
## voteCount	No	9.02096038620648e-79
## genresAmount	No	1.86071308417848e-49
## productionCoAmount	No	1.50733998814389e-62
## productionCountriesAmount	No	9.97076416462889e-91
## actorsAmount	No	4.42686884772083e-95
## castWomenAmount	No	6.56146830789997e-94
## castMenAmount	No	5.57916567989461e-92

Cualitativas Viendod los analisis de frecuencia se puede ver a simple vista que las peliculas de Drama son las mas c6munes, la mayoría fueron producidas por Paramount Studios. El director más frecuente es Steven Spielberg y el idioma predominante es el inglés.

```
##
##
## Table: Frequency Table for genres
##
## |genres          | Frequency|
## |:-----|:-----:|
## |Drama           |        521|
## |Comedy          |        440|
```

```
## |Horror          |          230|
## |Drama&#124;Romance |          211|
## |Horror&#124;Thriller |          205|
## |Comedy&#124;Romance |          201|
```

```
##
```

```
##
```

```
## Table: Frequency Table for homePage
```

```
##
```

```
## |homePage          | Frequency|
## |:-----:|-----:|
## |http://www.naruto-movie.com/ |          6|
## |http://www.thehungergames.movie/ |          4|
## |http://initiald-movie.com/ |          3|
## |http://kizumonogatari-usa.com/ |          3|
## |http://www.americanreunionmovie.com/ |          3|
## |http://www.kungfupanda.com |          3|
```

```
##
```

```
##
```

```
## Table: Frequency Table for productionCompany
```

```
##
```

```
## |productionCompany | Frequency|
## |:-----:|-----:|
## | |          457|
## |Paramount |          55|
## |Universal Pictures |          50|
## |Warner Bros. Pictures |          37|
## |Toei Animation |          34|
## |DreamWorks Animation |          33|
```

```
##
```

```
##
```

```
## Table: Frequency Table for productionCompanyCountry
```

```
##
```

```
## |productionCompanyCountry | Frequency|
## |:-----:|-----:|
## | |          1032|
## |US |          997|
## |US&#124;US |          792|
## |US&#124;US&#124;US |          452|
## |&#124;US |          292|
## |US&#124; |          265|
```

```
##
```

```
##
```

```
## Table: Frequency Table for productionCountry
```

```
##
```

```
## |productionCountry          | Frequency|
## |:-----:|-----:|
## |United States of America |          4971|
## |Japan |          613|
## |United Kingdom&#124;United States of America |          339|
## |United Kingdom |          294|
## | |          233|
## |Canada&#124;United States of America |          223|
```

```
##
```

```
##
```

Table: Frequency Table for director

##

director	Frequency
----------	-----------

:	:
---	---

	74
--	----

Steven Spielberg	29
------------------	----

Clint Eastwood	28
----------------	----

Ridley Scott	23
--------------	----

Kunihiko Yuyama	19
-----------------	----

Martin Scorsese	19
-----------------	----

##

##

Table: Frequency Table for originalTitle

##

originalTitle

:

Pinocchio

<U+30C9><U+30E9><U+3048><U+3082><U+3093> <U+306E><U+3073><U+592A><U+3068><U+96F2><U+306E><U+738B><U+306E>

Alice in Wonderland

Awake

Carrie

Cinderella

##

##

Table: Frequency Table for title

##

title	Frequency
-------	-----------

:	:
---	---

Cinderella	4
------------	---

Pinocchio	4
-----------	---

Alice in Wonderland	3
---------------------	---

Awake	3
-------	---

Beauty and the Beast	3
----------------------	---

Carrie	3
--------	---

##

##

Table: Frequency Table for originalLanguage

##

originalLanguage	Frequency
------------------	-----------

:	:
---	---

en	7772
----	------

ja	644
----	-----

es	425
----	-----

fr	271
----	-----

ko	167
----	-----

zh	119
----	-----

##

##

Table: Frequency Table for releaseDate

##

releaseDate	Frequency
-------------	-----------

:	:
---	---

2021-02-12	15
------------	----

2021-10-08	15
------------	----

```
## |2021-10-01 |      14|
## |2021-11-12 |      12|
## |2019-05-24 |      11|
## |2019-11-08 |      11|
```

4. Preguntas sobre los datos

4.1 ¿Cuáles son las 10 películas que contaron con más presupuesto?

Las películas que tuvieron un mayor presupuesto son

	id	title	budget
717	1865	Pirates of the Caribbean: On Stranger Tides	380,000,000
4711	99861	Avengers: Age of Ultron	365,000,000
5953	299534	Avengers: Endgame	356,000,000
164	285	Pirates of the Caribbean: At World's End	300,000,000
4954	141052	Justice League	300,000,000
5954	299536	Avengers: Infinity War	300,000,000
608	1452	Superman Returns	270,000,000
3792	38757	Tangled	260,000,000
7135	420818	The Lion King	260,000,000
281	559	Spider-Man 3	258,000,000

4.2. ¿Cuáles son las 10 películas que más ingresos tuvieron?

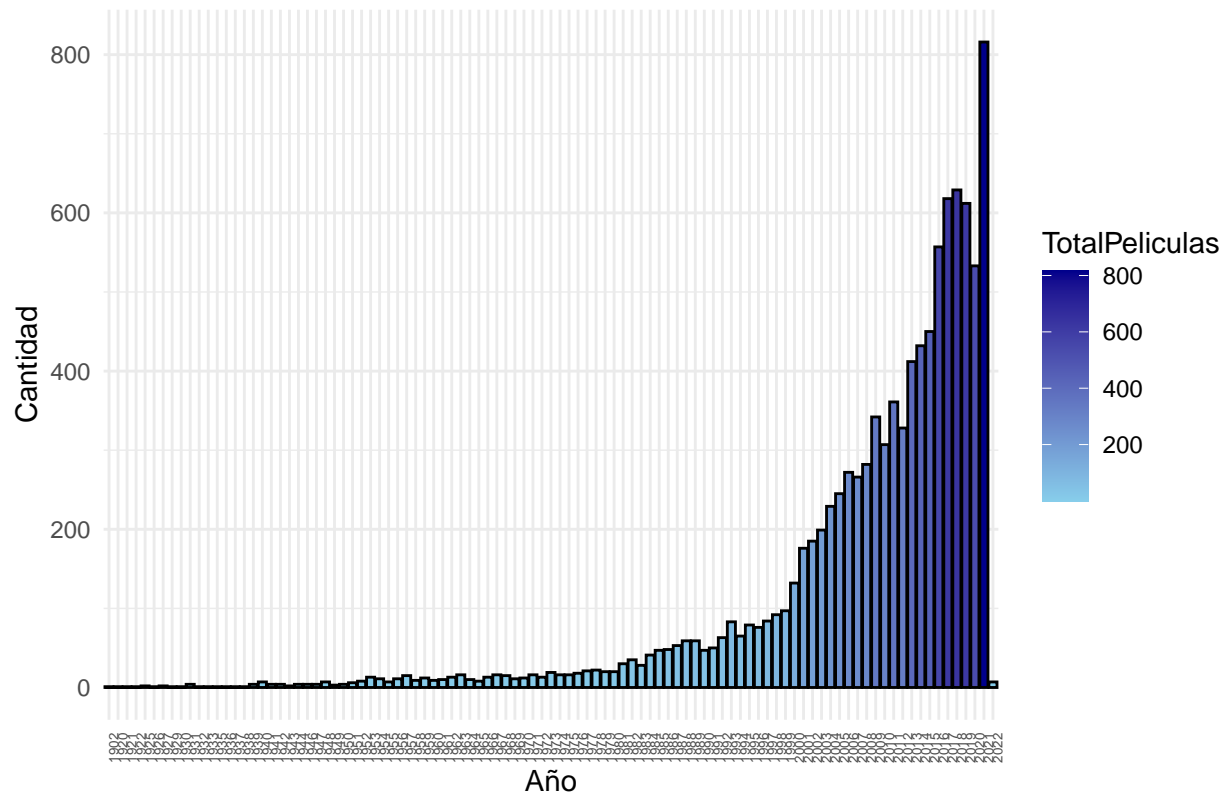
4.3 ¿Cuál es la película que más votos tuvo?

La película que mas tuvo votos fue **Inception** con una cantidad de **30788** votos

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

4.5 ¿Cuántas películas se hicieron en cada año? ¿En qué año se hicieron más películas?

Grafica 1: Peliculas por Año



En el año **** se publicaron un total de **** películas

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

4.7 ¿Las películas de qué género principal obtuvieron mayores ganancias?

Las películas del género **Action|Adventure|Science Fiction** fueron las que presentaron una mayor ganancia teniendo en total una ganancia de **14,832,790,887 \$**

4.8 ¿La cantidad de actores influye en los ingresos de las películas? ¿se han hecho películas con más actores en los últimos años?

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

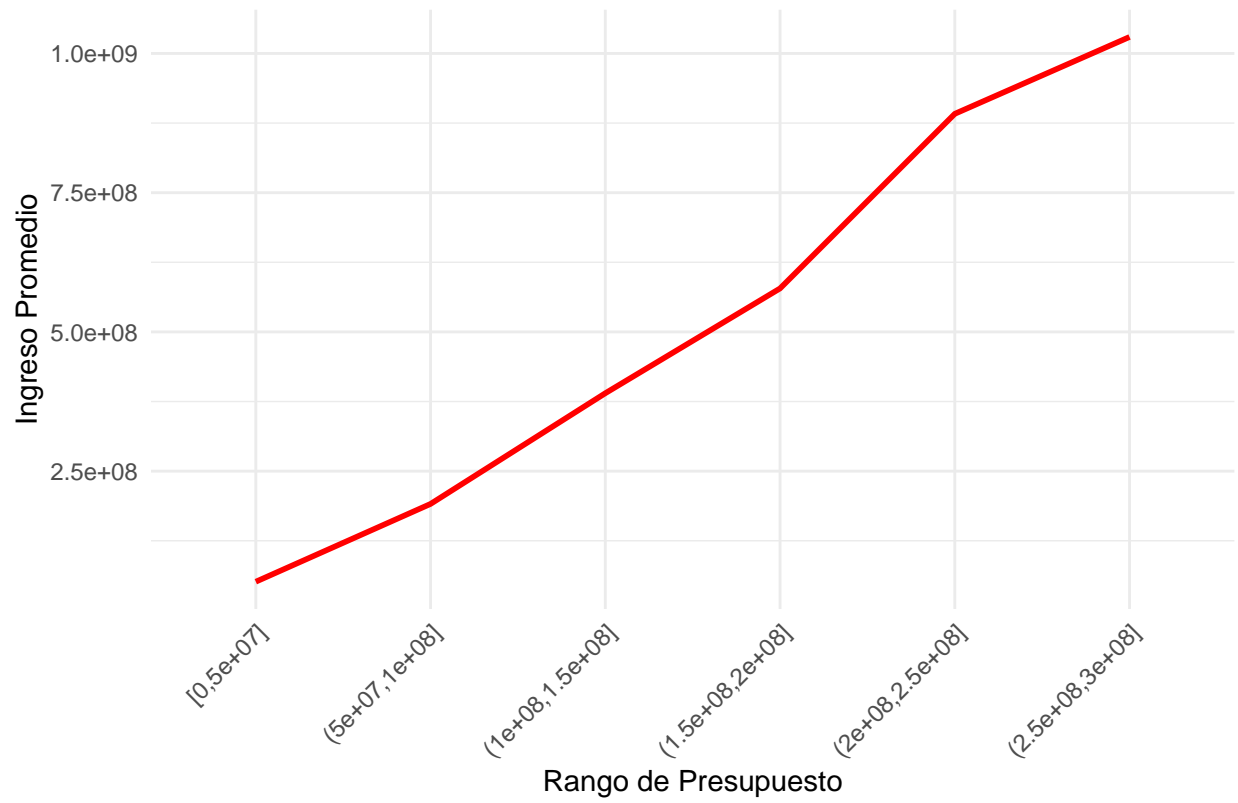
MasHombres	popularity	revenue
FALSE	47.95163	49056978
TRUE	55.12404	65191679

Como se puede evidenciar en la tabla las películas que tienen una mayor cantidad de hombres son más populares y tienen una mayor ganancia

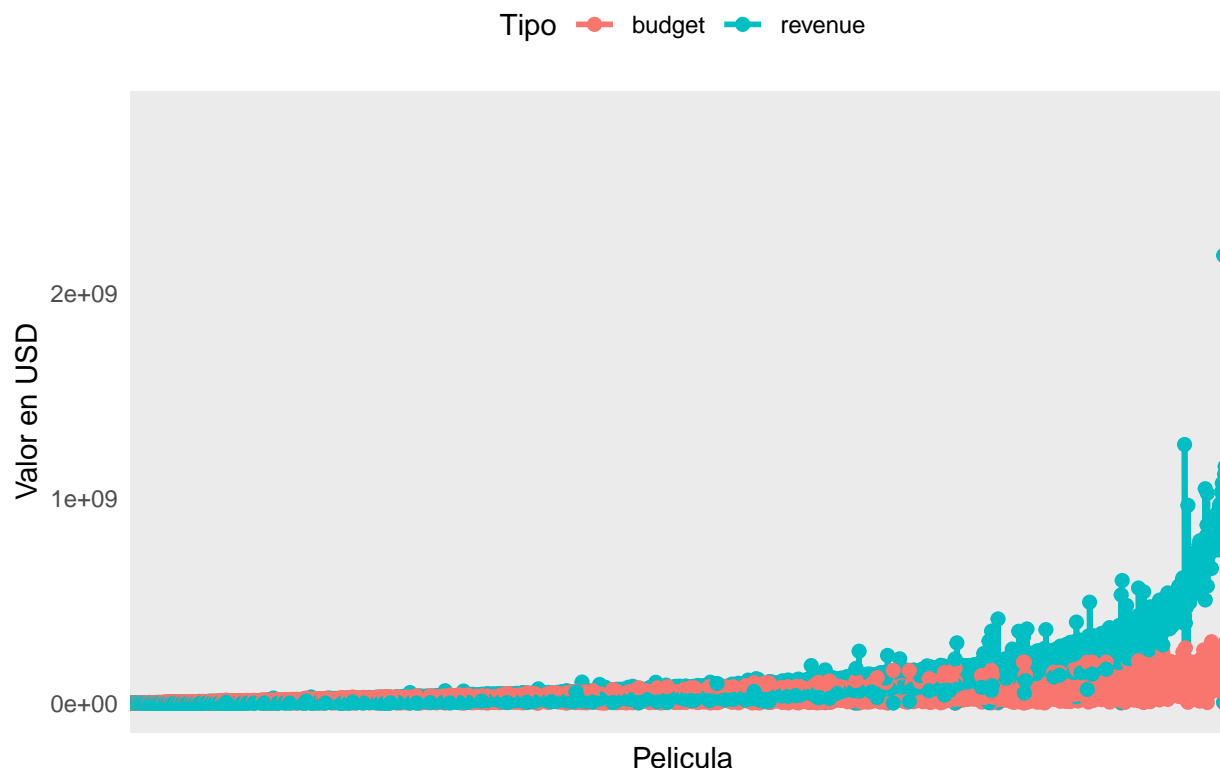
4.10 ¿Quiénes son los directores que hicieron las 20 películas mejor calificadas?

4.11 ¿Cómo se correlacionan los presupuestos con los ingresos? ¿Los altos presupuestos significan altos ingresos?

Grafica 1: Ingresos Promedio por Rango de Presupuesto



Grafica 2: Comparacion de Presupuesto e Ingresos por Pelicula



Como se puede evidenciar apartir de la Grafica 1, cuando el rango de presupuesto es mas elevado este suele precentar una mayor cantidad de ingres. Al igual como se ve en la Grafica 2, que en la mayoria de peliculas, se puede evidenciar que tener un presupuesto elevado significa tener mas ingresos.

4.12 Se asocian ciertos meses de lanzamiento con mejores ingresos?

4.13 ¿En qué meses se han visto los lanzamientos con mejores ingresos? ¿cuantas películas, en promedio, se han lanzado por mes?

En el mes en el que se presentan mejores ingresos es 06 con un total de ingresos promedio de 94,747,108 \$

4.14 ¿Cómo se correlacionan las calificaciones con el éxito comercial?

4.15 ¿Qué estrategias de marketing, como videos promocionales o páginas oficiales, generan mejores resultados?

4.16 ¿La popularidad del elenco está directamente correlacionada con el éxito de taquilla?

Preguntas Extras

- 1.
- 2.
- 3.
- 4.
- 5.
- 6.