

Generos de Netflix

Abrahám Soto y Odalis Rodriguez

2024-07-05

INTRODUCCIÓN

A través de este análisis estadístico se busca determinar el estudio de los géneros de películas, series, documentales y mucho más correspondientes a todas las entregas de la plataforma de streaming Netflix, desde el año 2014 al año 2021, siendo un total de 583 títulos.

Analizaremos los géneros según sus puntajes los cuales son establecido por la audiencia de `imdb_score` con un rango de 0 a 10 puntos. Obteniendo así cuales son los mejores géneros de la plataforma Netflix y cuántos títulos de diferentes calidades hay según cada uno.

Análisis de data

Al ver nuestra base de datos podemos encontrar las variables “títulos”, “generos” y “puntajes” con los cuales vamos a trabajar. Cada puntaje es correspondiente a cada título, los generos varían puesto que son 112 géneros en total, comenzaremos agrupando estos generos en nuevos grupos asociandolos para minimizar el total de generos, esto volverá nuestro trabajo más fácil.

Objetivo 1 Analizaremos la variable “Generos” para designar nuestras primeras operaciones, al ser una variable de tipo carácter la transformaremos a factor.

Creando el factor “Genero” podemos conocer el total de generos existentes:

```
table(Genero)
```

```
## Genero
##           Action           Action-adventure
##           7               1
##       Action-thriller       Action comedy
##           3               5
##       Action thriller       Action/Comedy
##           1               1
##       Action/Science fiction   Adventure
##           1               2
##       Adventure-romance       Adventure/Comedy
##           1               1
##       Aftershow / Interview   Animated musical comedy
##           6               1
##           Animation           Animation / Comedy
##           5               1
##       Animation / Musicial   Animation / Science Fiction
##           1               1
##       Animation / Short Animation/Christmas/Comedy/Adventure
##           4               1
##       Animation/Comedy/Adventure   Animation/Musical/Adventure
##           1               1
##       Animation/Superhero         Anime / Short
##           1               1
##           Anime/Fantasy           Anime/Science fiction
##           1               2
##       Anthology/Dark comedy       Biographical/Comedy
##           1               1
##           Biopic                 Black comedy
##           9               2
```

##	Christian musical	Christmas comedy
##	1	1
##	Christmas musical	Christmas/Fantasy/Adventure/Comedy
##	1	1
##	Comedy	Comedy-drama
##	49	14
##	Comedy / Musical	Comedy horror
##	2	1
##	Comedy mystery	Comedy/Fantasy/Family
##	1	1
##	Comedy/Horror	Coming-of-age comedy-drama
##	1	1
##	Concert Film	Crime drama
##	6	11
##	Crime thriller	Dance comedy
##	1	1
##	Dark comedy	Documentary
##	2	159
##	Drama	Drama-Comedy
##	77	1
##	Drama / Short	Drama/Horror
##	1	1
##	Family	Family film
##	2	2
##	Family/Christmas musical	Family/Comedy-drama
##	1	1
##	Fantasy	Heist
##	1	1
##	Heist film/Thriller	Hidden-camera prank comedy
##	1	1
##	Historical-epic	Historical drama
##	1	1
##	Horror	Horror-thriller
##	9	2
##	Horror comedy	Horror thriller
##	1	3
##	Horror/Crime drama	Making-of
##	1	2
##	Mentalism special	Mockumentary
##	1	2
##	Musical	Musical / Short
##	2	1
##	Musical comedy	Musical/Western/Fantasy
##	2	1
##	Mystery	One-man show
##	2	3
##	Political thriller	Psychological horror
##	1	1
##	Psychological thriller	Psychological thriller drama
##	4	1
##	Romance	Romance drama
##	6	1
##	Romantic comedy	Romantic comedy-drama
##	39	1

##	Romantic comedy/Holiday	Romantic drama
##	1	5
##	Romantic teen drama	Romantic teenage drama
##	1	1
##	Romantic thriller	Satire
##	1	2
##	Science fiction	Science fiction adventure
##	4	1
##	Science fiction thriller	Science fiction/Action
##	1	1
##	Science fiction/Drama	Science fiction/Mystery
##	3	1
##	Science fiction/Thriller	Sports-drama
##	4	3
##	Sports film	Spy thriller
##	1	2
##	Stop Motion	Superhero
##	1	2
##	Superhero-Comedy	Superhero/Action
##	1	1
##	Supernatural drama	Teen comedy-drama
##	1	1
##	Teen comedy horror	Thriller
##	1	33
##	Urban fantasy	Variety show
##	1	4
##	Variety Show	War
##	1	2
##	War-Comedy	War drama
##	1	2
##	Western	Zombie/Heist
##	3	1

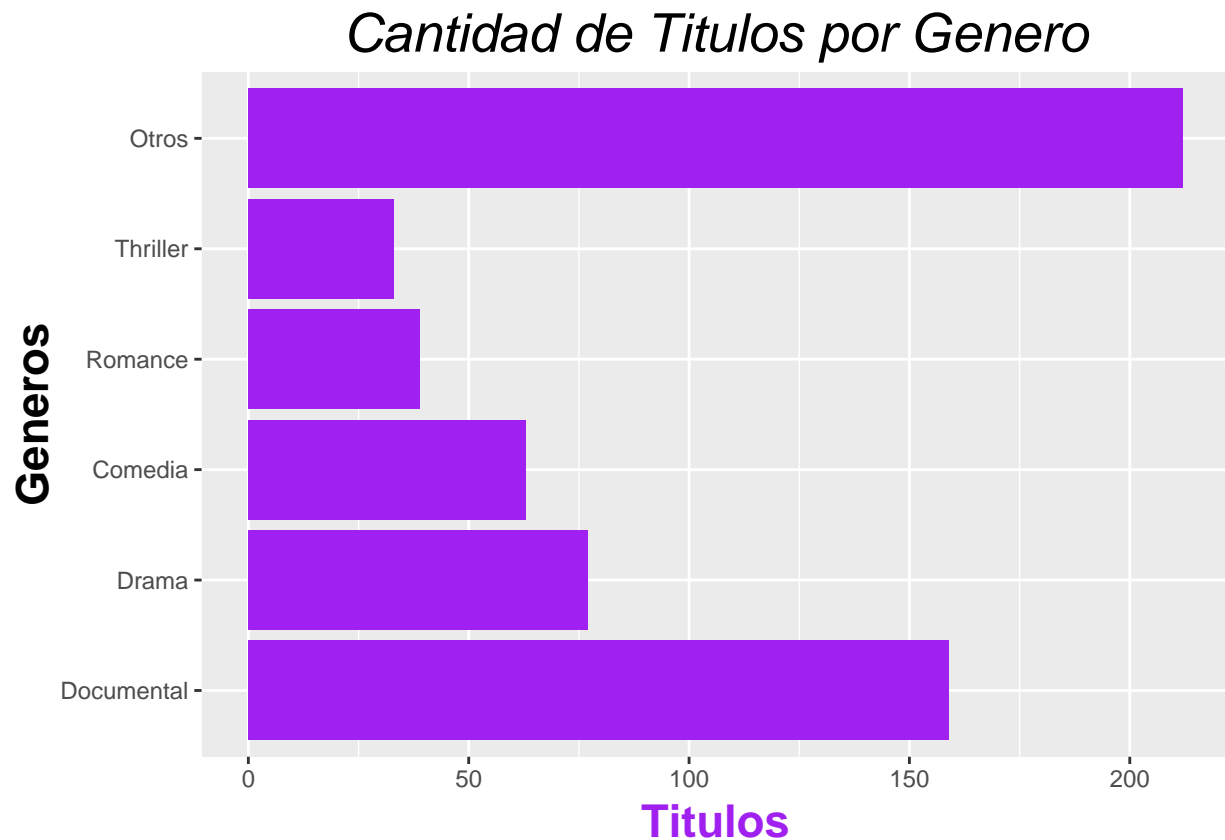
Podemos apreciar que existe un total de 114 géneros, una cantidad sumamente grande de trabajar, vamos a reducirlo a 6 géneros agrupandolos según sus características.

```
table(Generos)
```

##	Generos					
##	Documental	Drama	Comedia	Romance	Thriller	Otros
##	159	77	63	39	33	212

Gracias a nuestra nueva tabla de frecuencias acumuladas, podemos vizualizar a través de un gráfico nuestras magnitudes:

GENEROS



Objetivo 2

Evaluaremos la variable puntaje de `imdb_score` para analizar los puntajes correspondientes a los generos, agrupandolos en los 6 tipos de generos que creamos

```
table(Netflix.G6$imdb_score)
```

```
##
## 2.5 2.6 3.2 3.4 3.5 3.7 3.9 4.1 4.2 4.3 4.4 4.5 4.6 4.7 4.8 4.9 5 5.1 5.2 5.3
## 1 2 1 1 1 2 1 4 2 3 6 4 8 6 7 4 5 6 19 10
## 5.4 5.5 5.6 5.7 5.8 5.9 6 6.1 6.2 6.3 6.4 6.5 6.6 6.7 6.8 6.9 7 7.1 7.2 7.3
## 13 19 15 20 30 16 13 24 18 30 28 26 18 25 24 19 19 28 20 21
## 7.4 7.5 7.6 7.7 7.8 7.9 8 8.1 8.2 8.3 8.4 8.5 8.6 9
## 12 10 10 8 3 4 1 3 5 2 3 1 1 1
```

Podemos observar una tabla de frecuencias para los generos, iniciando de su puntuación más baja 2.5 hasta la más alta 9. Vamos a agrupar los puntajes correspondientes a cada género:

```
knitr::opts_chunk$set(echo = TRUE)
rangos <- c(0, 5, 7, 10)

Netflix.G6 <- Netflix.G6%>%mutate(Calidad=cut(imdb_score, breaks = rangos,
                                             labels = c("Baja" , "Media" , "Alta")))
```

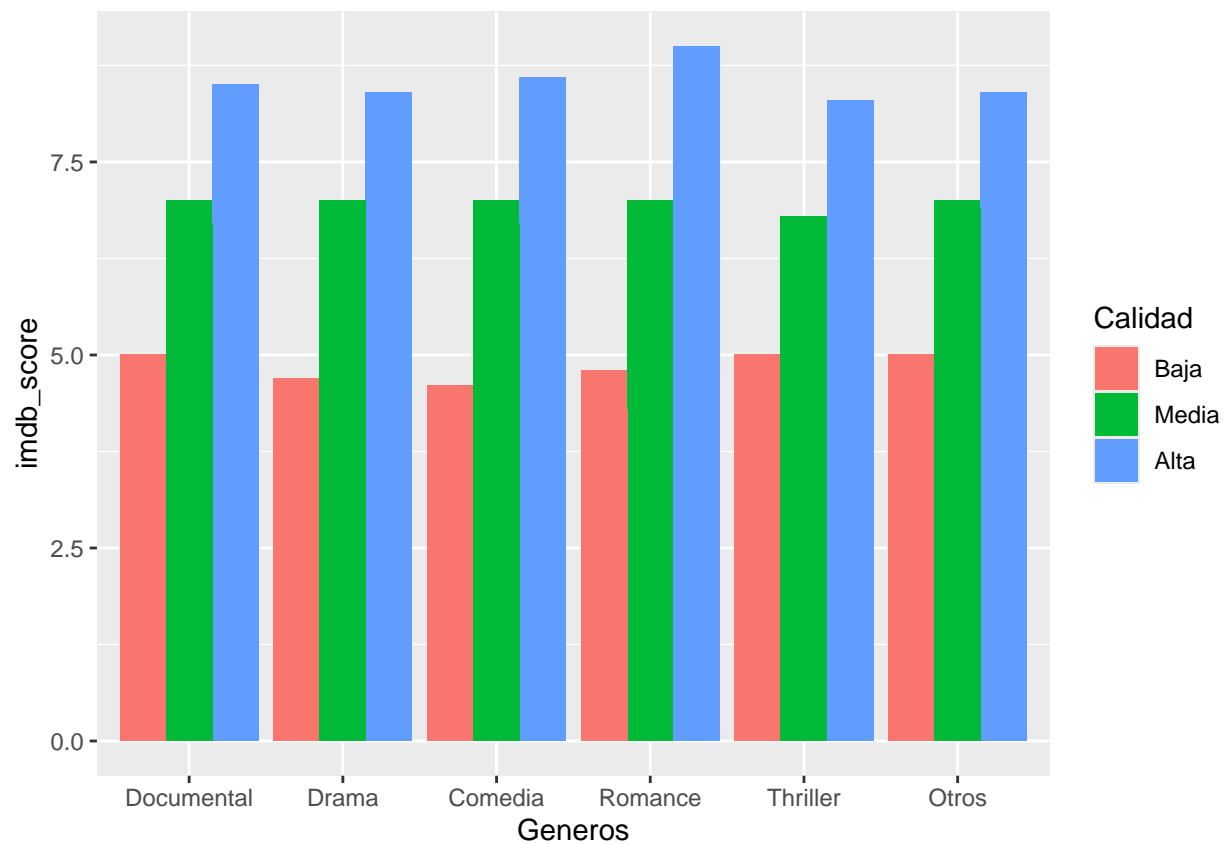
Observemos una tabla de nuestra nueva variable creada a través de otras dos:

```
table(Netflix.G6$Calidad)
```

```
##  
##  Baja Media  Alta  
##   58   392   133
```

Estelizaremos nuestra tabla en un gráfico

C



Observamos que los generos son casi homogéneos entre si

Objetivo 3

A través de una matriz vamos a evaluar la correlación entre los generos

```
table(Generos, Netflix.G6$Calidad)
```

```
##  
## Generos      Baja Media Alta
```

```
## Documental 17 113 29
## Drama 6 53 18
## Comedia 5 41 17
## Romance 4 23 12
## Thriller 4 17 12
## Otros 22 145 45
```

```
knitr::opts_chunk$set(echo = TRUE)
```

```
Datos <- matrix(c(ComediaP, DocumentalP, DramaP, RomanceP, ThrillerP, OtrosP),
                 nrow = 3,
                 ncol = 6)

colnames(Datos) <- c("Comedia" , "Documental" , "Drama" , "Romance", "Thriller", "Otros")
rownames(Datos) <- c("Buena" , "Regular" , "Mala")
```

Datos

```
##          Comedia Documental Drama Romance Thriller Otros
## Buena      17          29    18      12      12     45
## Regular    41         113    53      23      17    145
## Mala       5          17     6       4       4     22
```

```
prop.table(Datos)
```

```
##          Comedia Documental      Drama      Romance      Thriller      Otros
## Buena  0.029159520 0.04974271 0.03087479 0.020583190 0.020583190 0.07718696
## Regular 0.070325901 0.19382504 0.09090909 0.039451115 0.029159520 0.24871355
## Mala   0.008576329 0.02915952 0.01029160 0.006861063 0.006861063 0.03773585
```

Podemos encontrar las frecuencias reltivas en nuestra tabla de datos, en la cual podemos interpretar en porcentajes cualquiera de las celdas de nuestra tabla.

```
prop.table(Datos,1)
```

```
##          Comedia Documental      Drama      Romance      Thriller      Otros
## Buena  0.1278195  0.2180451 0.1353383 0.09022556 0.09022556 0.3383459
## Regular 0.1045918  0.2882653 0.1352041 0.05867347 0.04336735 0.3698980
## Mala   0.0862069  0.2931034 0.1034483 0.06896552 0.06896552 0.3793103
```

Evaluamos las frecuencias relativas de nuestra tabla condicionada por filas

```
prop.table(Datos, 2)
```

```
##          Comedia Documental      Drama      Romance      Thriller      Otros
## Buena  0.26984127  0.1823899 0.23376623 0.3076923 0.3636364 0.2122642
## Regular 0.65079365  0.7106918 0.68831169 0.5897436 0.5151515 0.6839623
## Mala   0.07936508  0.1069182 0.07792208 0.1025641 0.1212121 0.1037736
```

Evaluamos las frecuencias relativas de nuestra tabla condicionada por columnas

Correlación entre Generos

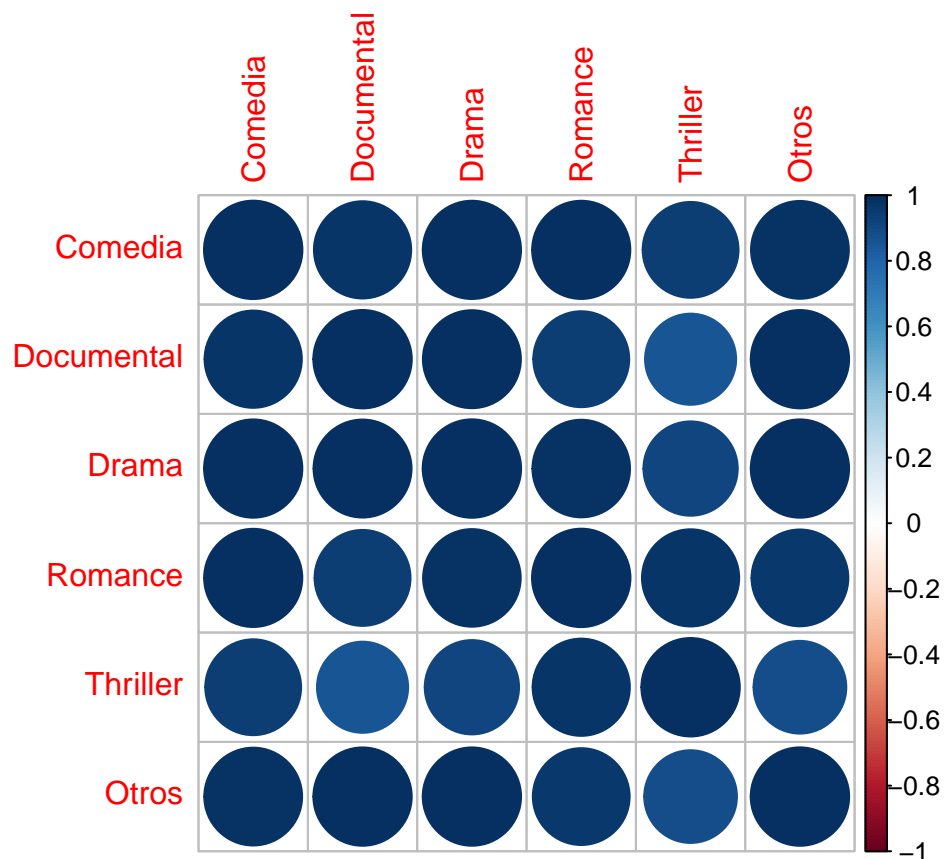
Evaluamos la correlación de cada genero con otro

```
Correlacion_Generos <- cor(Datos, method = "pearson")  
Correlacion_Generos
```

```
##           Comedia Documental      Drama   Romance Thriller      Otros  
## Comedia    1.0000000  0.9762210  0.9963701  0.9950821  0.9484206  0.9877483  
## Documental 0.9762210  1.0000000  0.9911311  0.9499475  0.8571465  0.9980900  
## Drama      0.9963701  0.9911311  1.0000000  0.9830379  0.9179913  0.9974474  
## Romance    0.9950821  0.9499475  0.9830379  1.0000000  0.9751578  0.9674328  
## Thriller   0.9484206  0.8571465  0.9179913  0.9751578  1.0000000  0.8873290  
## Otros      0.9877483  0.9980900  0.9974474  0.9674328  0.8873290  1.0000000
```

Evaluamos nuestros resultados, cada genero se correlaciona con otro, evaluando así cada combinación. Vamos a estetizarlo para una mejor observación:

```
corrplot(Correlacion_Generos)
```



Resultados

Con nuestro analisis hemos determinado los siguientes resultados:

1. El genero con mayor cantidad de titulos de “Buena Calidad” es Otros, con un total de 45 titulos y los generos con menor cantidad de titulos de “Buena Calidad” pertenece a Romance y Thriller con 12 titulos cada uno.
2. El genero con mayor cantidad de títulos de “Baja Calidad” pertenece a Otros con un total de 22 titulos y el genero con menor cantidad de titulos de “Baja Calidad” pertenecen a Romance y Thriller con 4 titulos cada uno.
3. La cantidad de titulos de cada genero ordenados desde el mayor al menor sería: Otros, Documental, Drama, Comedia, Romance, Thriller.
4. La menor correlación entre generos es Documental y Thriller con 0.8571465

Conclusión

Ahora sabemos la calidad de los géneros de la plataforma de streaming Netflix, la proxima vez que querramos ver algún titulo en nuestro día, podemos evaluar a través del analisis elaborado, cuales serían nuestras mejores opciones de Generos según sus porcentajes de calidad.