Estadísticas en el Mundo del **Entretenimiento**

Proyecto de Ciencia de Datos

Tecnológico de Monterrey
Escuela de Ingeniería y Ciencias | Campus CVA

Analista de Datos

Annette Pamela Ruiz Abreu
A01423595
Estudiante de Ingeniería en Ciencia de Datos y Matemáticas

Problema

- Alta demanda, pero mucha competencia.
- Aumento de producción de películas entre el 2000 y 2020
- El cine pasó de tener un IVA del 8% al 21%.
- Es un hecho que más del 50% de todas las películas de Hollywood no producen ganancia. Son un “flop”.

¿Cómo saber en qué película invertir?

Solución

Crear un modelo de predicción para la calificación y número de votos de una serie o película con base en distintas variables.

IMDb

Internet Movie Database es una base de datos en línea que almacena información relacionada con películas

01 **title.akas**
ID, título, origen, idioma, tipo, atributos, si es el título original

02 **title.basics**
ID, tipo, título, año inicial, año final, géneros, minutos, para adultos

03 **title.crew**
ID, director, escritor

04 **title.principals**
ID, ID persona, catergoría, trabajo, personaje

05 **title.ratings**
ID, calificación promedio, número de votos

06 **name.basics**
ID persona, nombre, año de nacimiento y muerte, películas conocidas, profesion

Bases de **Datos**

Total de **Datos**

Iniciales

Records	8,374,640

8,374,640 videos

Records	11,350,212

11,350,212, personas

Datos usados

Records	1,199,980

1,199,980 videos con calificación y votos

Records	353,000

353,000 personas vivas

Herramienta Tecnológica

Inteligencia artificial y tecnologías de aprendizaje automático para **preparar**, **visualizar**, **supervisar** e **implementar** conjuntos de datos.

Hipótesis 1

Entre más número de votos tenga una película o serie, mejor calificación tendrá. El número de votos (popularidad) y calificación son directamente proporcionales.

La producción de películas ha ido aumentando con los años, así como su popularidad.

Hipótesis 2

Objetivos

Clasificación

Hacer un modelo de clasificación para determinar la popularidad de la película con base en variables de la base de datos.

Análisis Estadístico

Analizar los cambios de popularidad y calificación a lo largo de los años.

Clientes

Productoras de Cine

Cines

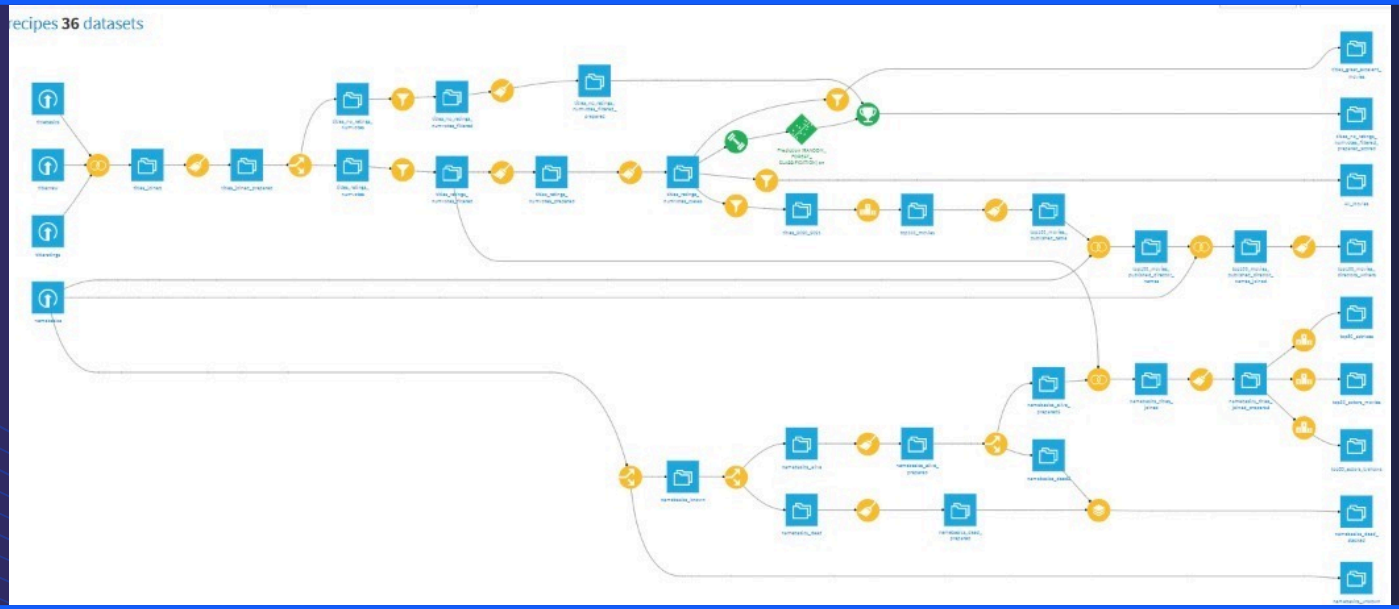
Plataformas de *Streaming*

Cronograma

Fase 1
Entendimiento del Negocio
Fase 3
Preparación de los Datos
Fase 2
Entendimiento de los Datos
Fase 4
Modelación de los Datos

Flow del **Proyecto**

Recetas Utilizadas: Join, Split, Top N, Filter, Prepare, Stack

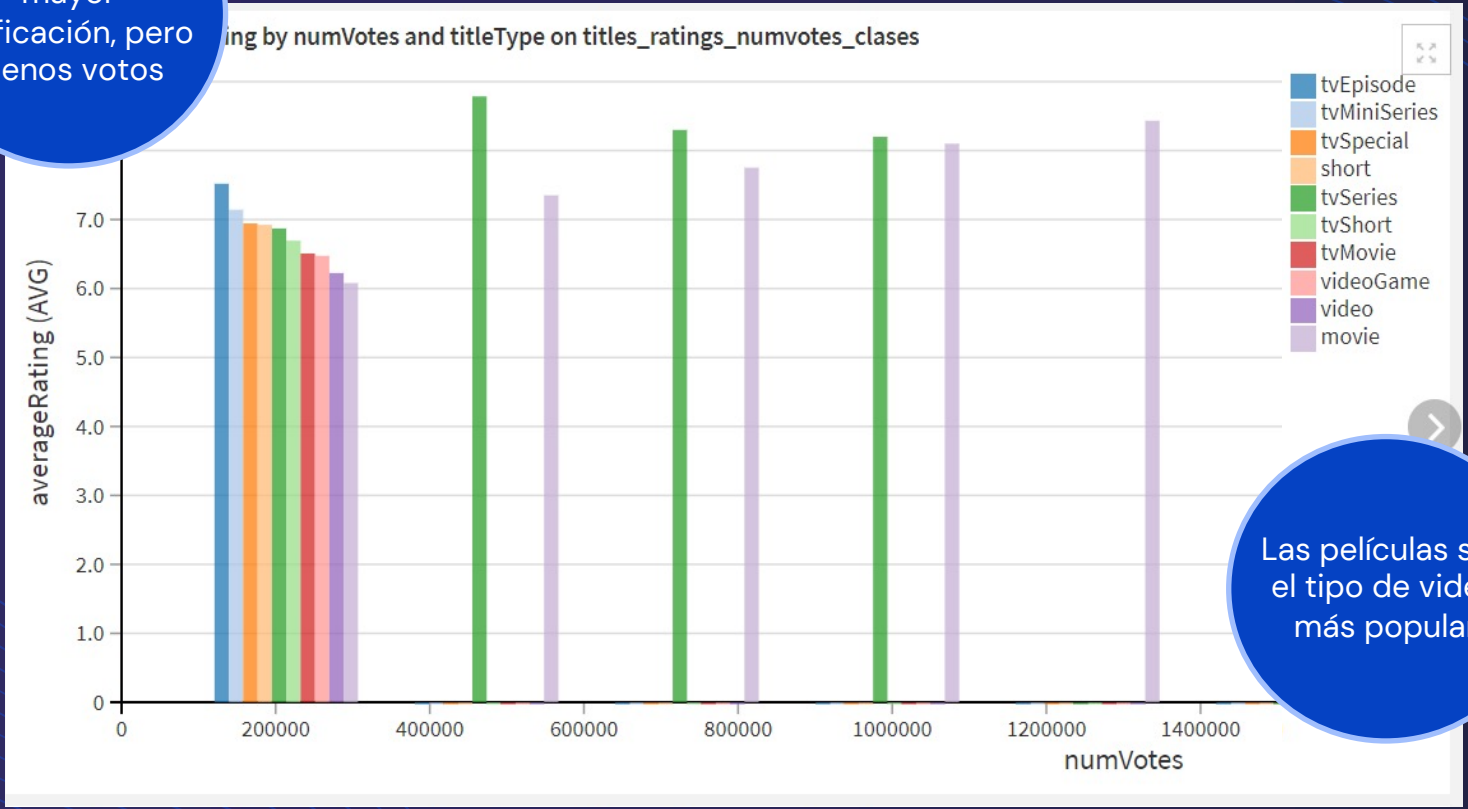


1

Análisis de All Title Types

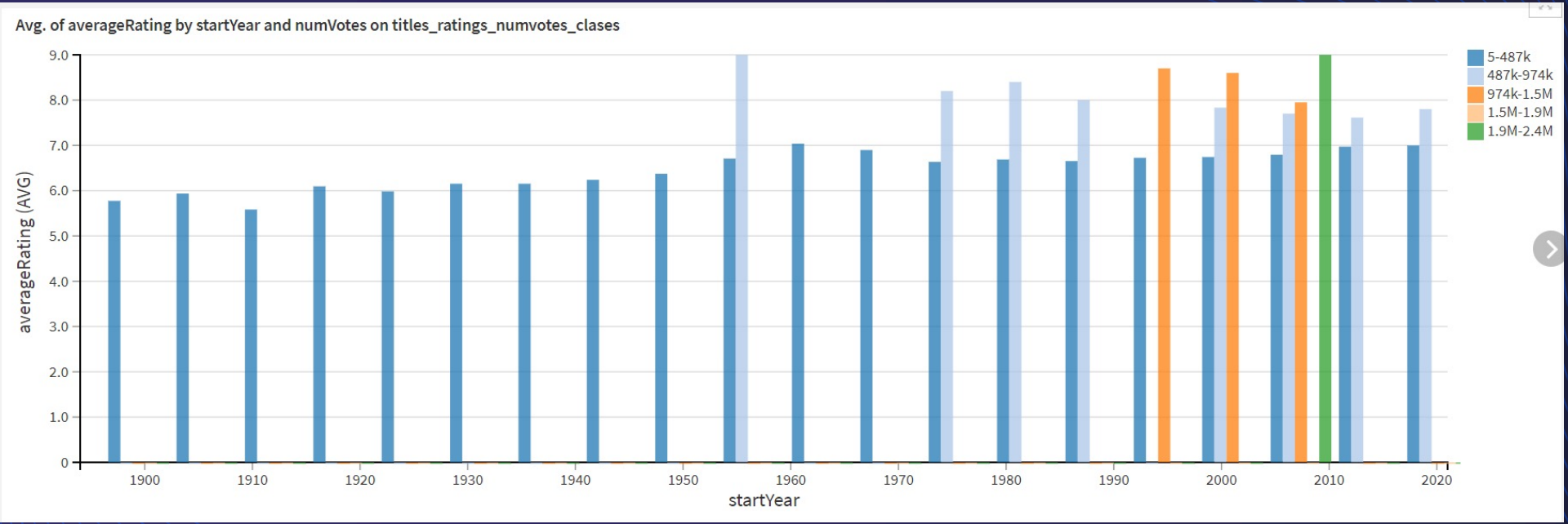
Popularidad por Tipo

Las series tienen mayor calificación, pero menos votos



Las películas son el tipo de video más popular

Cambio de Popularidad por Años



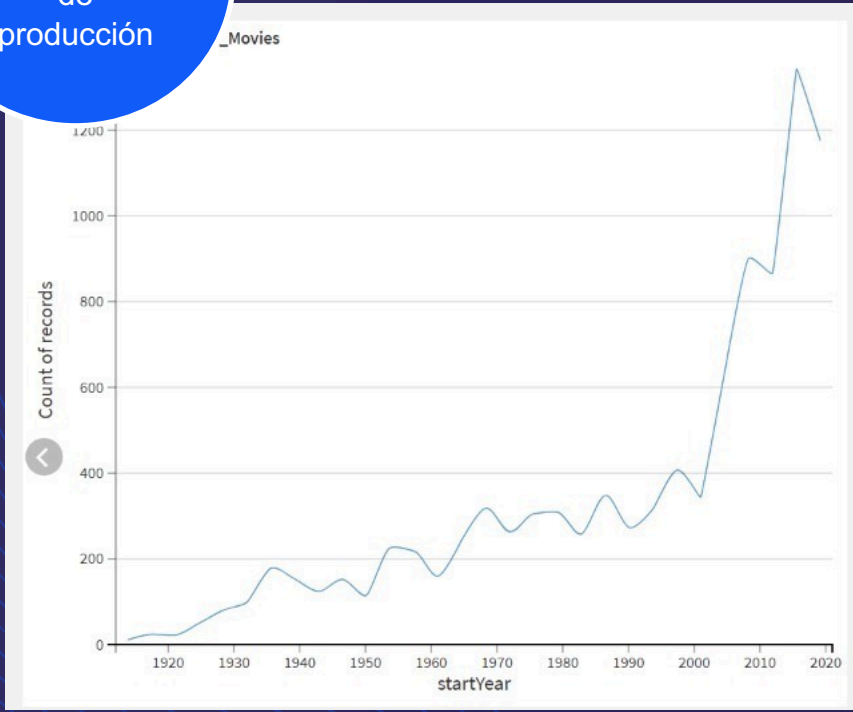
Los videos con más votos tienen mayor promedio calificación.
Los votos y las calificaciones han aumentado con los años.

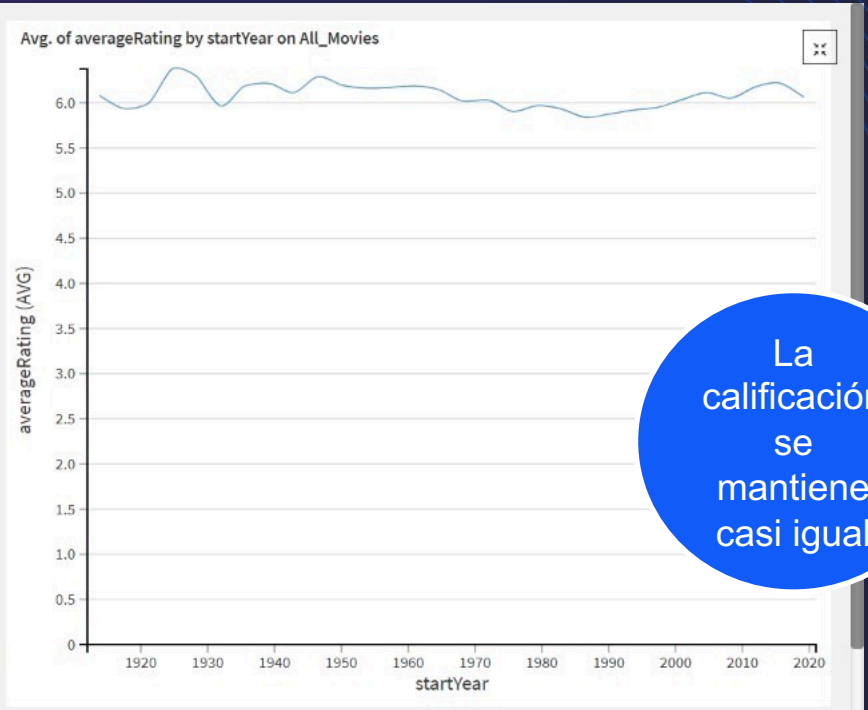
2

All Movies

Producción y Calificación de Películas

2015-2017: Mayor años de producción





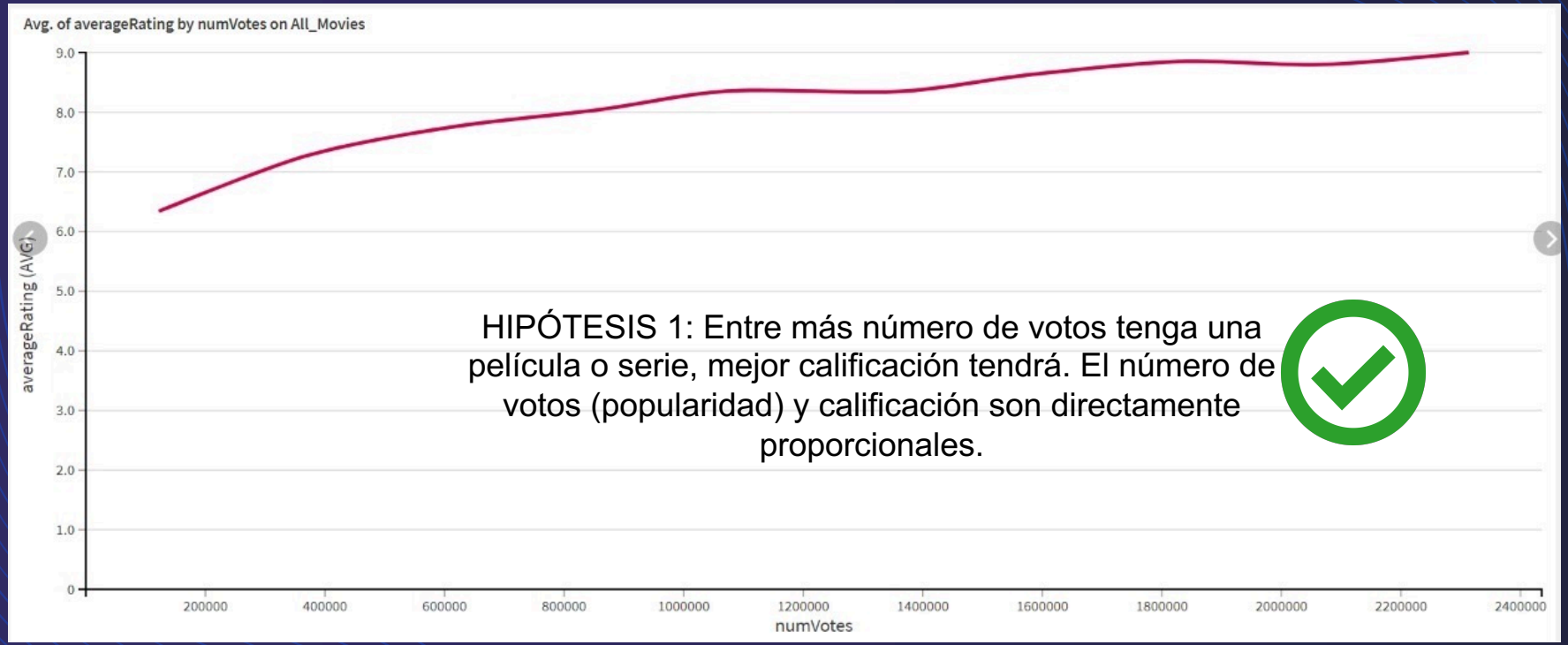
La calificación se mantiene casi igual

Producción y calificación **por Año**

En la gráfica anterior parece que las calificaciones se han mantenido o disminuido un poco, eso es gracias a que entre 1920 y 1970 había muchas menos películas y tenían calificaciones entre 7.5-9, mientras que en entre 1990 y el 2020 hay muchas más películas y con todas las calificaciones.

Hipótesis 2: La producción de películas ha ido aumentando con los años, así como su popularidad.

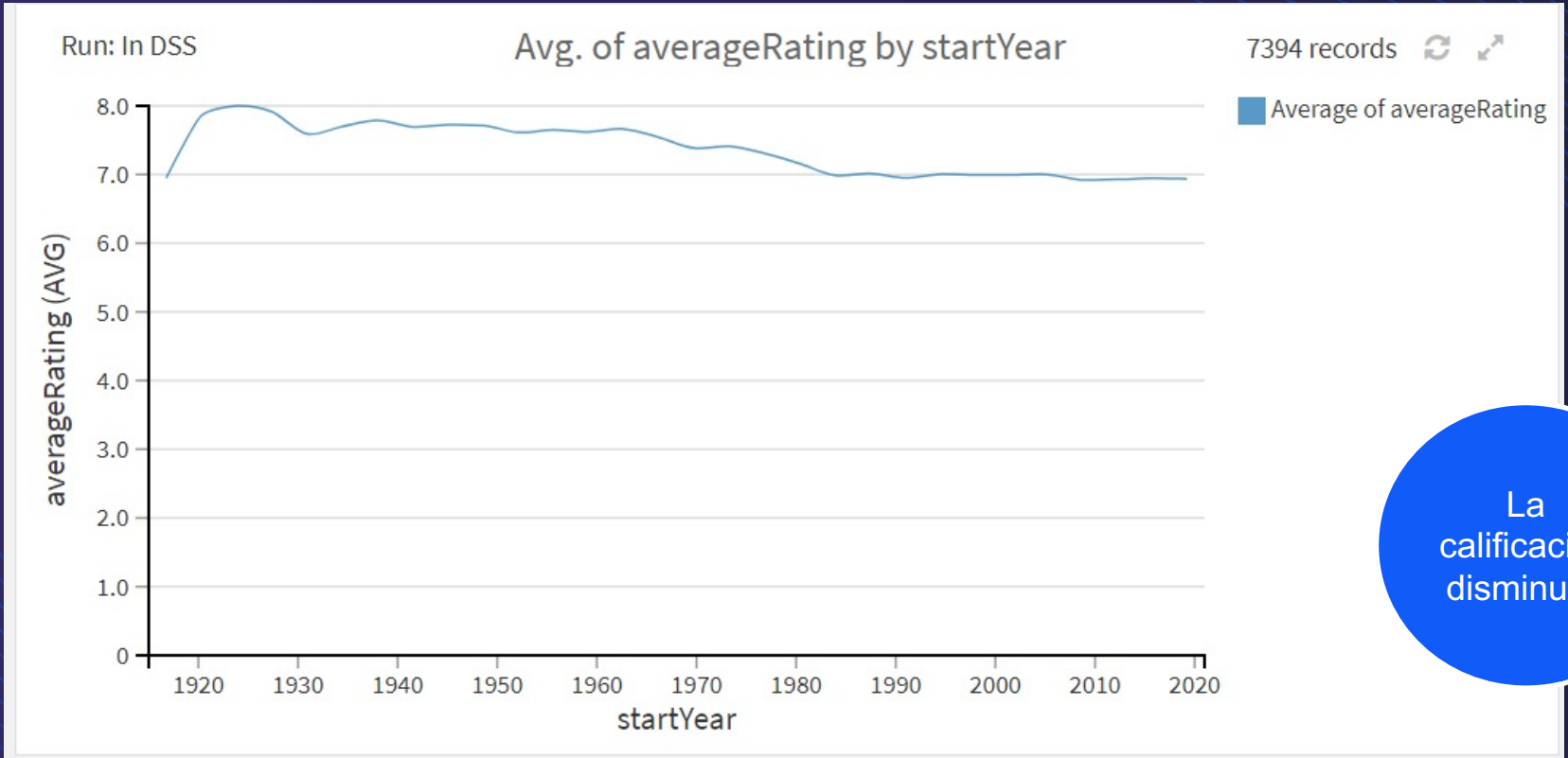
Votos por **Calificación**



HIPÓTESIS 1: Entre más número de votos tenga una película o serie, mejor calificación tendrá. El número de votos (popularidad) y calificación son directamente proporcionales.

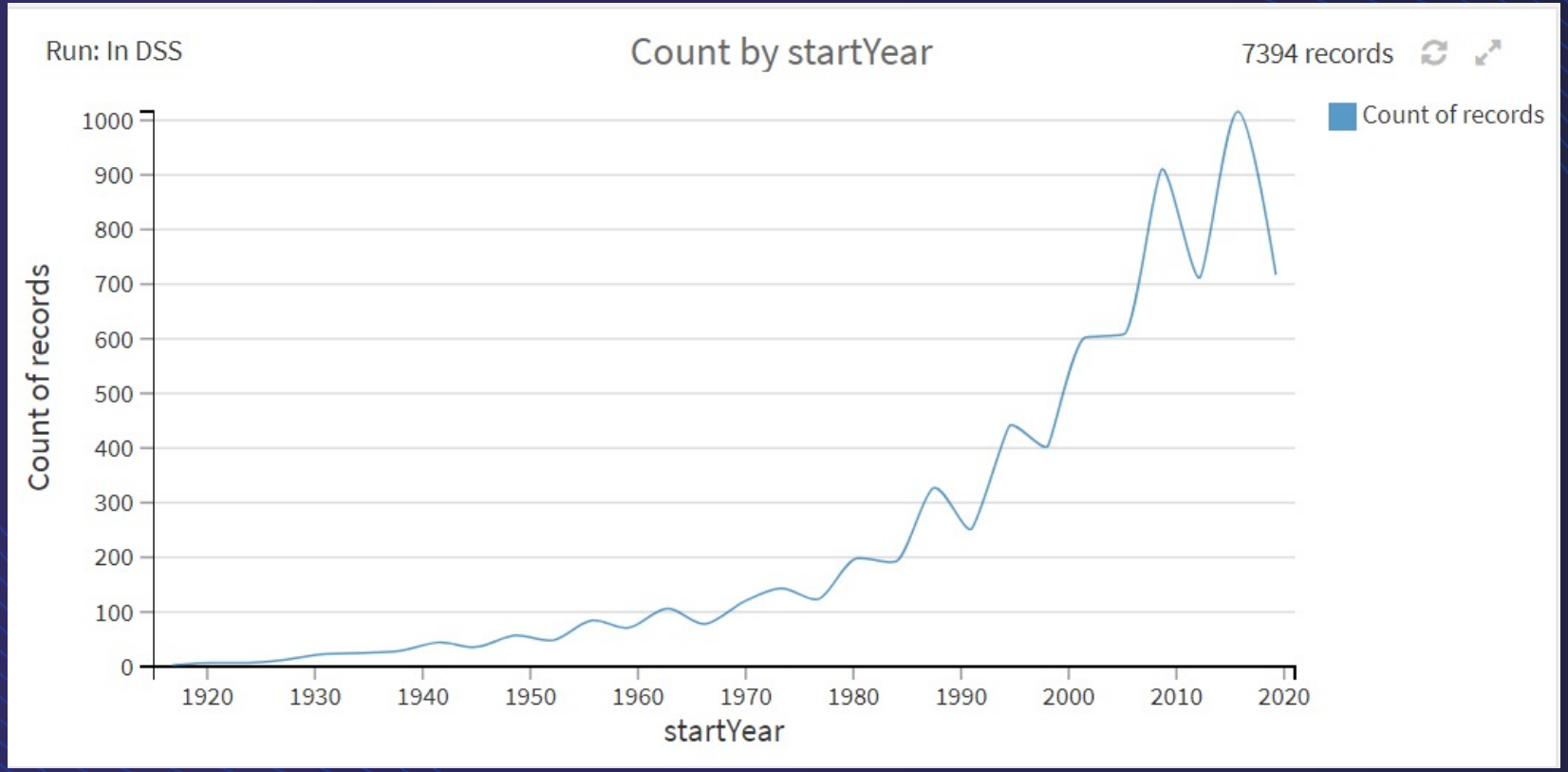
3

All Great Excellent Movies

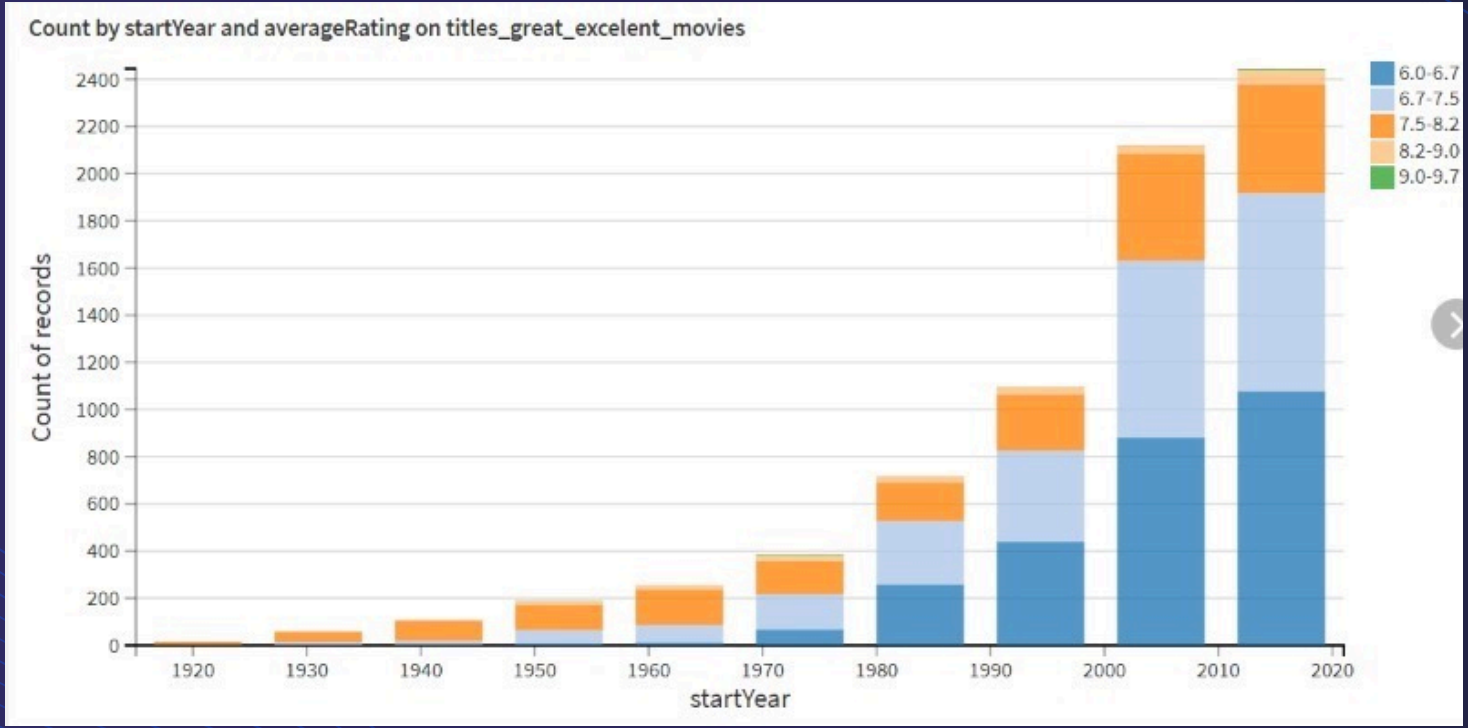


La calificación disminuye

Producción de Películas



Producción y calificación **por Año**

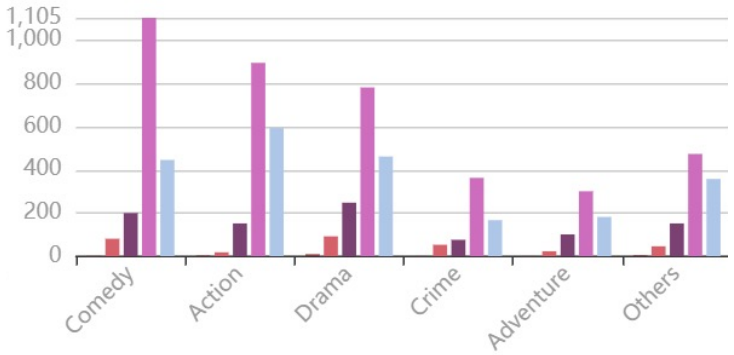


[1900,1928)
[1928, 1956)
[1956, 1984)
[1984, 2012)
[2021, 2040)

Drama
Drama
Drama
Comedia
Acción

Crime	[1900,1928)	1
	[1928,1956)	53
	[1956,1984)	76
	[1984,2012)	363
	[2012,2040)	167
Adventure	[1900,1928)	1
	[1928,1956)	23
	[1956,1984)	101
	[1984,2012)	301
	[2012,2040)	182
Others	[1900,1928)	5
	[1928,1956)	46
	[1956,1984)	152
	[1984,2012)	474
	[2012,2040)	358

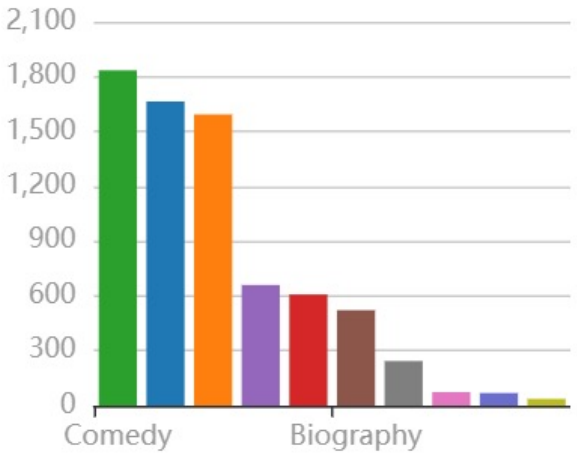
Histogram



Comedy	[1900,1928)	2
	[1928,1956)	81
	[1956,1984)	200
	[1984,2012)	1105
	[2012,2040)	446
Action	[1900,1928)	4
	[1928,1956)	18
	[1956,1984)	152
	[1984,2012)	896
	[2012,2040)	594
Drama	[1900,1928)	9
	[1928,1956)	92
	[1956,1984)	248
	[1984,2012)	782
	[2012,2040)	462

A genres_0

Histogram



Summary stats

N values	7394
N distinct	19
Mode	Comedy
N empty	0

Frequency table

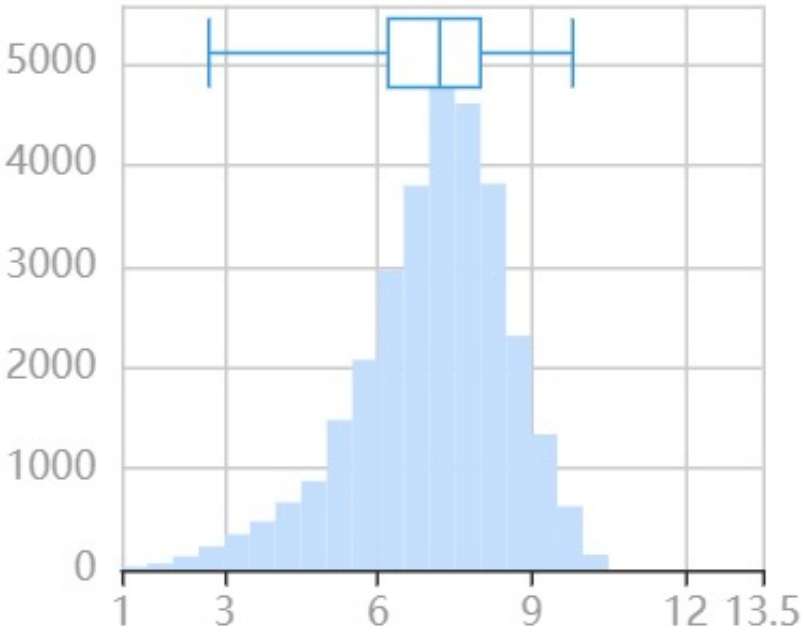
Comedy	25%	1834
Action	23%	1664
Drama	22%	1593
Crime	9%	660
Adventure	8%	608
Biography	7%	522
Horror	3%	244
Documentary	1%	74
Animation	1%	69
Mystery	1%	37
Others	1%	89
N distinct		19

4

2020–2021 All Titles

averageRating

Histogram



Summary stats

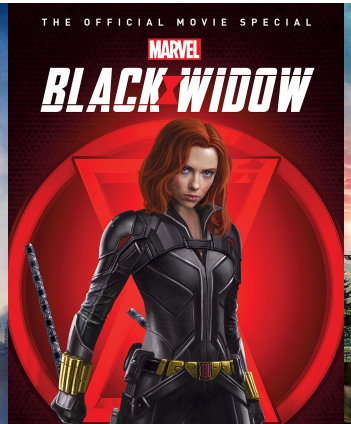
N values	30779
N distinct	91
N finite	30779
Mean	6.9990318074
Median	7.2
Std Dev	1.4765057543
Min	1
Max	10

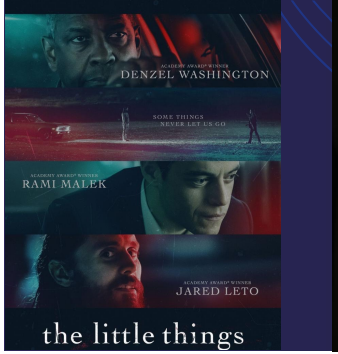
5

2020-2021 Top 100 Movies

primaryTitle	startYear	runtimeMinutes	genres_0	genres_1	genres_2
Soorarai Pottru	2020	153	Drama		
Dara of Jasenovac	2020	130	Drama	War	
Shershaah	2021	135	Action	Biography	Drama
Hamilton	2020	160	Biography	Drama	History
The Father	2020	97	Drama		
Dune	2021	155	Action	Adventure	Drama
Dil Bechara	2020	101	Comedy	Drama	Romance
Zack Snyder's Justice League	2021	242	Action	Adventure	Fantasy
Soul	2020	100	Adventure	Animation	Comedy
Shang-Chi and the Legend of the Ten	2021	132	Action	Adventure	Fantasy
The Trial of the Chicago 7	2020	129	Drama	History	Thriller
Another Round	2020	117	Comedy	Drama	
Master	2021	179	Action	Thriller	
The Mitchells vs the Machines	2021	113	Adventure	Animation	Comedy
The Social Dilemma	2020	94	Documentary	Drama	
No Time to Die	2021	163	Action	Adventure	Thriller
Promising Young Woman	2020	113	Crime	Drama	Mystery
Luca	2021	95	Adventure	Animation	Comedy
Minari	2020	115	Drama		
Judas and the Black Messiah	2021	126	Biography	Drama	History
Tenet	2020	150	Action	Sci-Fi	Thriller
Cruella	2021	134	Adventure	Comedy	Crime
Nobody	2021	92	Action	Crime	Drama
Palm Springs	2020	90	Comedy	Fantasy	Mystery
Nomadland	2020	107	Drama		
Onward	2020	102	Adventure	Animation	Comedy
Raya and the Last Dragon	2021	107	Action	Adventure	Animation
The Suicide Squad	2021	132	Action	Adventure	Comedy

A Quiet Place Part II	2020	97	Drama	Horror	Sci-Fi
Free Guy	2021	115	Action	Adventure	Comedy
The Invisible Man	2020	124	Drama	Horror	Mystery
Wrath of Man	2021	119	Action	Crime	Thriller
The Devil All the Time	2020	138	Crime	Drama	Thriller
The Dig	2021	112	Biography	Drama	History
The White Tiger	2021	125	Crime	Drama	
The King of Staten Island	2020	136	Comedy	Drama	
Love and Monsters	2020	109	Action	Adventure	Comedy
Greyhound	2020	91	Action	Drama	History
Mank	2020	131	Biography	Comedy	Drama
Black Widow	2021	134	Action	Adventure	Sci-Fi
News of the World	2020	118	Action	Adventure	Drama
Boss Level	2020	94	Action	Mystery	Sci-Fi
Extraction	2020	116	Action	Thriller	
The Old Guard	2020	125	Action	Adventure	Fantasy
Run	2020	90	Mystery	Thriller	
Fear Street: Part Two - 1978	2021	109	Drama	Horror	Mystery
The Tomorrow War	2021	138	Action	Adventure	Drama
Enola Holmes	2020	123	Action	Adventure	Crime
Bad Boys for Life	2020	124	Action	Comedy	Crime

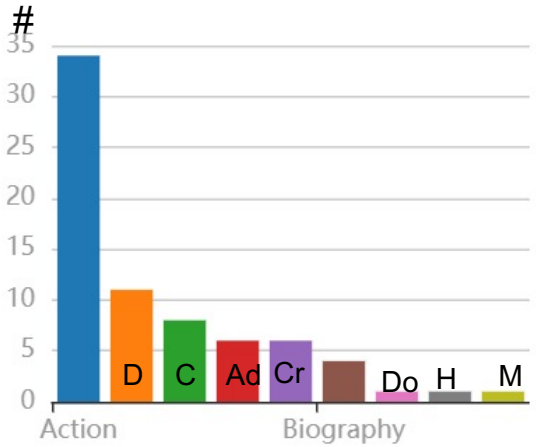




Borat Subsequent Moviefilm	2020	95	Comedy		
Jungle Cruise	2021	127	Action	Adventure	Comedy
I'm Thinking of Ending Things	2020	134	Drama	Thriller	
The Green Knight	2021	130	Adventure	Drama	Fantasy
Sonic the Hedgehog	2020	99	Action	Adventure	Comedy
The Hunt	2020	90	Action	Horror	Thriller
Eurovision Song Contest: The Story of	2020	123	Comedy	Music	
Godzilla vs. Kong	2021	113	Action	Sci-Fi	Thriller
Greenland	2020	119	Action	Drama	Thriller
I Care a Lot	2020	118	Comedy	Crime	Thriller
The Conjuring: The Devil Made Me Dc	2021	112	Horror	Mystery	Thriller
The Little Things	2021	128	Crime	Drama	Mystery
The Guilty	2021	90	Crime	Drama	Thriller
Malignant	2021	111	Crime	Horror	Mystery
Venom: Let There Be Carnage	2021	97	Action	Adventure	Sci-Fi
Spenser Confidential	2020	111	Action	Comedy	Crime
Fear Street: Part One - 1994	2021	107	Drama	Horror	Mystery
Birds of Prey	2020	109	Action	Adventure	Comedy
Mortal Kombat	2021	110	Action	Adventure	Fantasy
Hitman's Wife's Bodyguard	2021	100	Action	Comedy	Crime
Holidate	2020	104	Comedy	Romance	
Project Power	2020	113	Action	Crime	Sci-Fi
Unhinged	2020	90	Action	Thriller	

A genres_0

Histogram

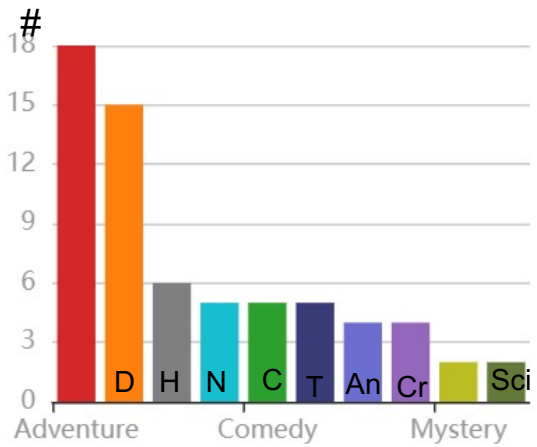


Summary stats

N values	72
N distinct	9
Mode	Action
N empty	0

A genres_1

Histogram

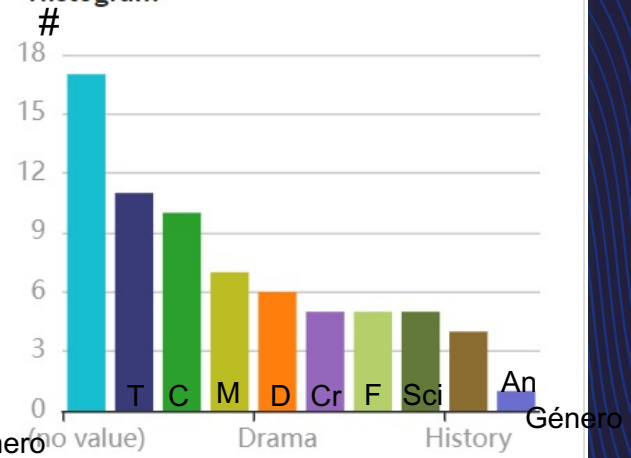


Summary stats

N values	72
N distinct	16
Mode	Adventure
N empty	5

A genres_2

Histogram



Summary stats

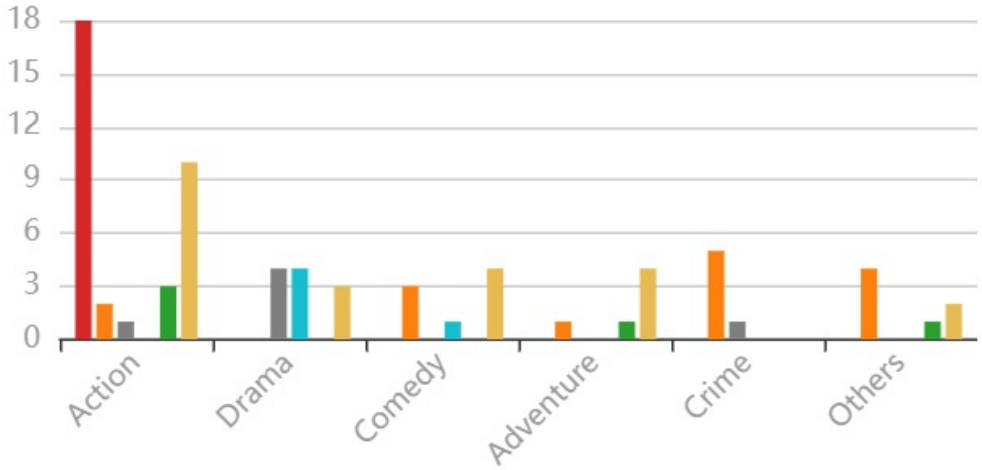
N values	72
N distinct	11
Mode	(no value)
N empty	17

Adventure
Others
Horror
Comedy
No value
Drama

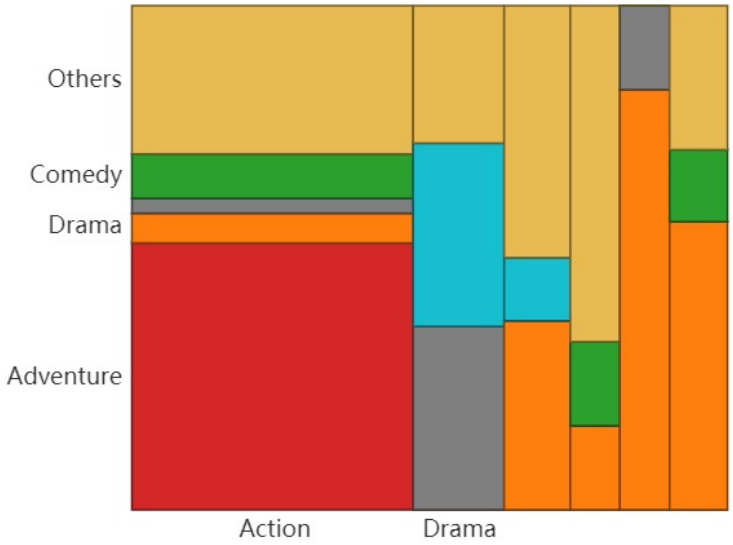
Primer y Segundo Género de las **Top 100** Películas (2020-2021)

A genres_1 by A genres_0

Histogram



Mosaic plot



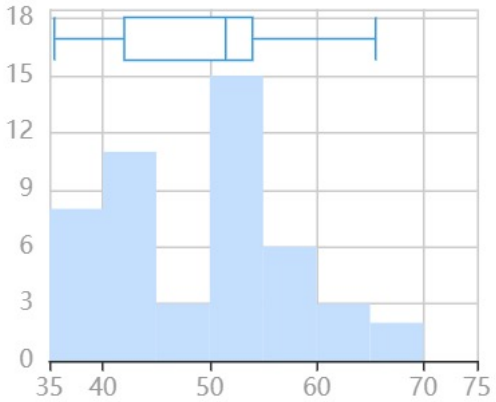
Las películas más populares entre el 2020 y 2021 son de acción y aventura.

Primer y Segundo Género de las **Top 100** Películas (2020-2021)

genres_0	genres_1	Count
Action	Adventure	18
	Drama	2
	Horror	1
	(no value)	0
	Comedy	3
	Others	10
Drama	Adventure	0
	Drama	0
	Horror	4
	(no value)	4
	Comedy	0
	Others	3
Comedy	Adventure	0
	Drama	3
	Horror	0
	(no value)	1
	Comedy	0
	Others	4
Adventure	Adventure	0
	Drama	1
	Horror	0
	(no value)	0
	Comedy	1
	Others	4
Crime	Adventure	0
	Drama	5
	Horror	1
	(no value)	0
	Comedy	0
	Others	0
Others	Adventure	0
	Drama	4
	Horror	0
	(no value)	0
	Comedy	1
	Others	2

director_age

Histogram

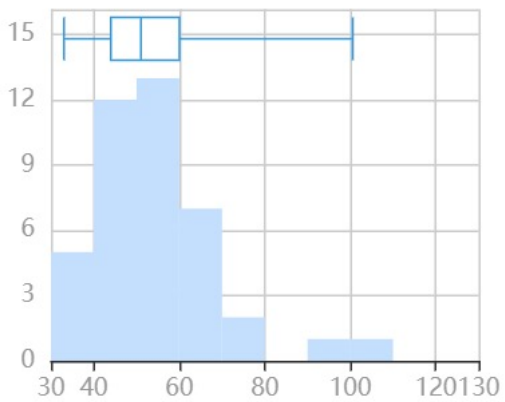


Summary stats

N values	72
N distinct	27
N finite	48
Mean	48.916666667
Median	51.5
Std Dev	8.2715439742
Min	35
Max	66

writer_age

Histogram



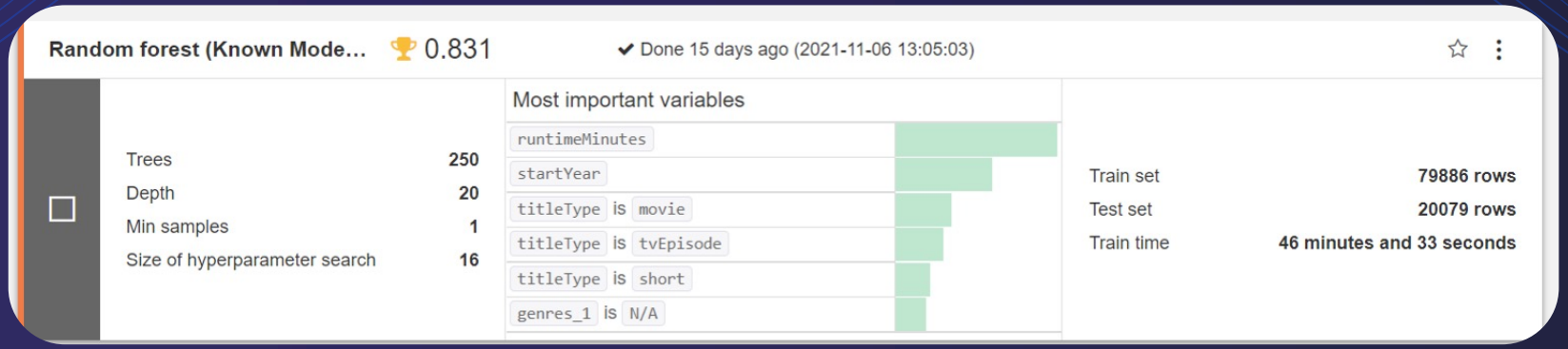
Summary stats

N values	72
N distinct	29
N finite	41
Mean	53.317073171
Median	51
Std Dev	14.770644915
Min	31
Max	107

Edades de los directores y escritores de las **Top 100** películas del 2020 y 2021

Modelos de Clasificación

Modelo 1: **Predecir popularidad**

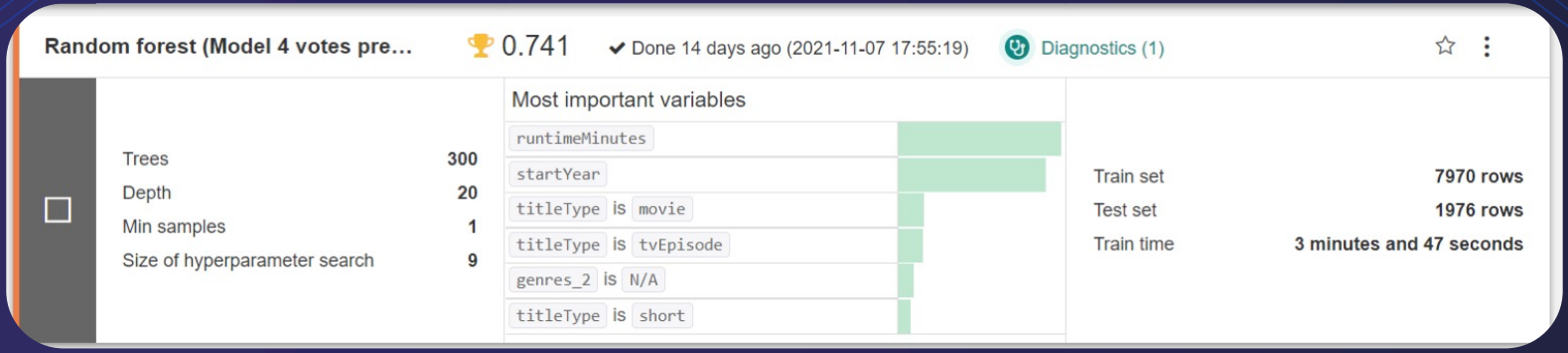


	Predicted			
Actual	unknown	average	known	
unknown	79 %	8 %	13 %	100 %
average	19 %	40 %	41 %	100 %
known	9 %	13 %	78 %	100 %

VARIABLES

- Duración
- Tipo de video
- Género
- Año
- Director

Modelo 2: **Predecir popularidad**

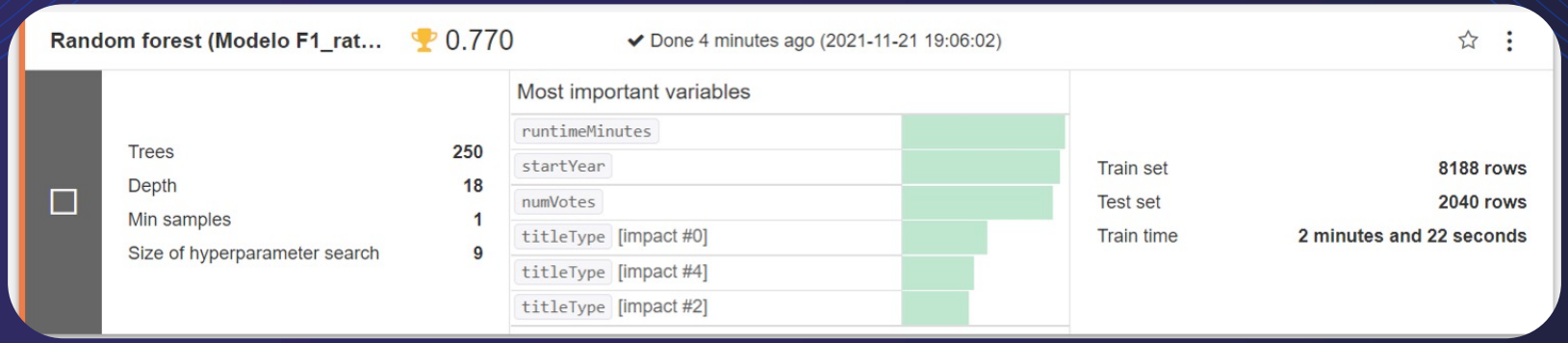


	Predicted							
Actual	0-50	50-100	500-5000	200-500	100-200	10000+	5000-10000	
0-50	51 %	19 %	3 %	8 %	10 %	4 %	6 %	100 %
50-100	26 %	24 %	8 %	13 %	15 %	9 %	4 %	100 %
500-5000	8 %	5 %	27 %	14 %	11 %	18 %	16 %	100 %
200-500	13 %	13 %	15 %	21 %	13 %	14 %	13 %	100 %
100-200	16 %	16 %	12 %	11 %	25 %	11 %	10 %	100 %
10000+	< 1 %	1 %	7 %	2 %	2 %	65 %	22 %	100 %
5000-10000	2 %	4 %	11 %	4 %	5 %	44 %	29 %	100 %

VARIABLES

- Duración
- Tipo de video
- Género
- Año
- Director

Modelo 3: **Predecir calificación**



	Predicted					
Actual	great	good	mediocre	excelent	bad	
great	43 %	13 %	8 %	31 %	6 %	100 %
good	19 %	25 %	25 %	20 %	11 %	100 %
mediocre	5 %	15 %	42 %	14 %	24 %	100 %
excelent	25 %	5 %	4 %	62 %	4 %	100 %
bad	5 %	8 %	25 %	12 %	51 %	100 %

VARIABLES

- Duración
- Votos
- Tipo de video
- Género
- Año
- Director

CONCLUSIONES

Las series tienen mayor calificación, pero menos votos

2015-2017: Mayor años de producción de películas

La calificación de películas se mantiene casi igual con los años

Las películas son el tipo de video más popular

TOP 3 DE PELÍCULAS GÉNEROS
COMEDIA
ACCIÓN DRAMA

CALIFICACIÓN PROMEDIO DE LAS PELÍCULAS DEL 2020 AL 2021: 7

AMBAS HIPÓTESIS SON CORRECTAS

EDAD PROMEDIO DE LOS DIRECTORES Y ESCRITORES: 51

Las películas más populares entre el 2020 y 2021 son de acción y aventura.

Los videos con más votos tienen mayor promedio calificación.
Los votos y las calificaciones han aumentado con los años.

GRACIAS POR SU ATENCIÓN