

*Streaming inteligente:*  
Eligiendo la mejor plataforma para  
ver películas

**Integrantes:**

Juan Levipil  
Adrian Huizi  
Andrea Riquelme  
Catalina Díaz

## Índice:

1. Contexto y motivación.....	3
2. Objetivo.....	4
3. Datos.....	5-9
4. Pregunta de investigación.....	10
5. Diseño tentativo.....	11

## Contexto y motivación:

Nuestra principal motivación surge a raíz de un gusto en común: Ver películas. Hemos notado que se suele gastar grandes cantidades de dinero en mantener diferentes suscripciones de streaming, ya que actualmente existen muchas plataformas (quizás demasiadas) para ver largometrajes. Debido a lo anterior, queremos identificar la mejor plataforma basándonos en su relación calidad/precio, teniendo en cuenta el costo mensual de cada plataforma, la calidad del catálogo y las restricciones de su uso. La audiencia objetivo está compuesta por usuarios de las plataformas de streaming y personas que estén considerando el adquirir la suscripción de alguna.

Además, el crecimiento acelerado del mercado del streaming en los últimos años ha transformado la forma en que las personas consumen contenido audiovisual. Según diversos estudios, la mayoría de los hogares mantiene más de una suscripción activa, lo que genera un gasto mensual considerable que muchas veces no se evalúa en relación al valor real que ofrecen estas plataformas. Este fenómeno se intensifica cuando se observa que gran parte de los catálogos presentan títulos repetidos, lo que afecta la percepción de calidad y variedad para los usuarios.

En este contexto, nuestro proyecto cobra relevancia porque busca entregar herramientas que faciliten una toma de decisiones más informada. La elección de una plataforma de streaming no solo implica un tema de entretenimiento, sino también de economía doméstica y de maximización del uso del tiempo de ocio. De esta manera, nuestro análisis pretende aportar claridad en un mercado competitivo y saturado, entregando una comparación objetiva que pueda ser utilizada por personas con distintos intereses: desde cinéfilos que buscan variedad y calidad en el catálogo, hasta familias que desean optimizar sus gastos sin perder acceso a contenido atractivo.

## **Objetivo:**

Esperamos resolver la pregunta sobre cuál es la mejor plataforma de streaming (si es que la hay), para encontrar la mejor opción en cuanto a películas y el consumo de los mismos.

Con este estudio, podemos obtener el conocimiento necesario para poder optar por una mejor experiencia para los usuarios en plataformas de streaming.

La audiencia objetivo de este proyecto es el público general que consume películas.

De manera más específica, nuestro objetivo consiste en generar un análisis comparativo que permita evaluar aspectos como el costo mensual de cada servicio, la amplitud y diversidad de su catálogo y la calidad de las producciones disponibles. Este enfoque nos permitirá no solo determinar cuál plataforma ofrece la mejor relación calidad precio, sino también comprender las diferencias y ventajas que presenta cada una frente a la competencia.

A nivel práctico, los resultados de este trabajo pueden convertirse en una guía para orientar la decisión de consumo en un mercado cada vez más competitivo y fragmentado. Además, el análisis entregará recursos relevantes para identificar tendencias en la industria del entretenimiento digital, lo que podría ser útil no solo para usuarios individuales, sino también para familias, grupos de amigos o incluso instituciones que buscan optimizar su inversión en servicios de streaming.

En este sentido, nuestro proyecto no solo busca dar respuesta a una pregunta puntual, sino también ofrecer información accionable que contribuya a un consumo más consciente y eficiente, tanto en términos económicos como de experiencia de usuario.

## Datos:

### 1) API de Watchmode

Esta API tiene un registro sobre distintos títulos, ya sea serie o película, y la disponibilidad de estos en las distintas plataformas de streaming. Cuenta con distintos endpoints para poder buscar individualmente títulos, los detalles de estos, actores que participen en estos y demás, pero principalmente estaremos utilizando el endpoint de List Titles, el cual nos devolverá una lista de títulos que pueden ser filtrados según las características que se pidan, ya sea el tipo de este (si es una serie o película), la disponibilidad en cierta región o plataforma, por géneros y fechas. Los datos que entrega vienen en tipo JSON con los siguientes datos:

- **id:** Dato de tipo int, sirve como identificador único de cada película de la API
- **title:** Dato de tipo string, es el título de la película
- **year:** Dato de tipo int, indica el año en el que se estrenó la película
- **imdb\_id:** Dato de tipo string, es el identificador de la película en IMDb
- **tmdb\_id:** Dato de tipo string, es el identificador de la película en TMDb
- **type:** Dato categórico, es el tipo (película, serie, etc.) que tiene en la API
- **tmdb\_type:** Dato categórico, es el tipo que tiene en TMDb

	id	title	year	imdb_id	tmdb_id	tmdb_type	type
0	1700766	The Thursday Murder Club	2025	tt12001534	744653	movie	movie
1	11020618	Unknown Number: The High School Catfish	2025	tt37674426	1517772	movie	movie
2	1860796	Love Untangled	2025	tt33549683	1355666	movie	movie
3	11007564	Fantasy Football Ruined Our Lives	2025	tt37038661	1500691	movie	movie
4	1985183	Planet Single: Greek Adventure	2025	tt37727295	1468789	movie	movie
5	1804101	Happy Gilmore 2	2025	tt31868189	1263256	movie	movie
6	1749638	The Wild Robot	2024	tt29623480	1184918	movie	movie
7	1691853	Carry-On	2024	tt21382296	1005331	movie	movie
8	1596405	Glass Onion: A Knives Out Mystery	2022	tt11564570	661374	movie	movie
9	1690558	Wallace & Gromit: Vengeance Most Fowl	2024	tt17163970	929204	movie	movie

Por ejemplo, la imagen muestra la lista de películas de Netflix obtenidas por la API.

## 2) IMDb Non-Commercial Datasets

Consisten en archivos de tipo .tsv (Valores separados por tabulaciones), que contienen distintos datos de la página IMDb, a continuación el listado de los archivos y su contenido:

### title.basics.tsv

Contiene datos básicos sobre los títulos (películas, series, etc.) registrados en IMDb, contiene las siguientes columnas:

- tconst: Dato de tipo string que sirve como identificador único para los títulos.
- titleType: Categórico, indica el tipo de el título (película, serie, etc.).
- primaryTitle: String, indica el nombre más popular de la película.
- originalTitle: String, indica el nombre original de la película.
- isAdult: Bool, indica si el título es o no para adultos.
- startYear: Int, año en el que se estrenó, en series es el año en que empezó.
- endYear: Int, año en el que terminó la serie, es nulo para los demás tipos.
- runtimeMinutes: Int, duración del título en minutos.
- genres: Array de strings, contiene hasta 3 géneros asociados al título.

	tconst	titleType	primaryTitle	originalTitle	isAdult	startYear	endYear	runtimeMinutes	genres
8	tt0000009	movie	Miss Jerry	Miss Jerry	0	1894	\N	45	Romance
144	tt0000147	movie	The Corbett-Fitzsimmons Fight	The Corbett-Fitzsimmons Fight	0	1897	\N	100	Documentary,News,Sport
498	tt0000502	movie	Bohemios	Bohemios	0	1905	\N	100	\N
570	tt0000574	movie	The Story of the Kelly Gang	The Story of the Kelly Gang	0	1906	\N	70	Action,Adventure,Biography
587	tt0000591	movie	The Prodigal Son	L'enfant prodigue	0	1907	\N	90	Drama
610	tt0000615	movie	Robbery Under Arms	Robbery Under Arms	0	1907	\N	\N	Drama
625	tt0000630	movie	Hamlet	Amleto	0	1908	\N	\N	Drama
668	tt0000675	movie	Don Quijote	Don Quijote	0	1908	\N	\N	Drama
672	tt0000679	movie	The Fairylogue and Radio-Plays	The Fairylogue and Radio-Plays	0	1908	\N	120	Adventure,Fantasy
828	tt0000838	movie	A Cultura do Cacau	A Cultura do Cacau	0	1909	\N	\N	\N

Fuente: IMDb (Non-Commercial Datasets)

## name.basics.tsv

Contiene datos sobre personas relacionadas a los títulos, contiene las siguientes columnas:

- **nconst**: String, sirve como identificador único para las personas.
- **primaryName**: String, nombre con el que frecuentemente recibe crédito la persona.
- **birthYear**: Int, año de nacimiento de la persona.
- **deathYear**: Int, año de fallecimiento de la persona, es nulo si no aplica.
- **primaryProfession**: Array de strings, hasta 3 profesiones por las que es más conocida la persona.
- **knownForTitles**: Array de strings, títulos por los que es conocida la persona.

	<b>nconst</b>	<b>primaryName</b>	<b>birthYear</b>	<b>deathYear</b>	<b>primaryProfession</b>	<b>knownForTitles</b>
0	nm0000001	Fred Astaire	1899	1987	actor,miscellaneous,producer	tt0050419,tt0072308,tt0027125,tt0025164
1	nm0000002	Lauren Bacall	1924	2014	actress,soundtrack,archive_footage	tt0037382,tt0075213,tt0038355,tt0117057
2	nm0000003	Brigitte Bardot	1934	\N	actress,music_department,producer	tt0057345,tt0049189,tt0056404,tt0054452
3	nm0000004	John Belushi	1949	1982	actor,writer,music_department	tt0072562,tt0077975,tt0080455,tt0078723
4	nm0000005	Ingmar Bergman	1918	2007	writer,director,actor	tt0050986,tt0069467,tt0050976,tt0083922
5	nm0000006	Ingrid Bergman	1915	1982	actress,producer,soundtrack	tt0034583,tt0038109,tt0036855,tt0038787
6	nm0000007	Humphrey Bogart	1899	1957	actor,producer,miscellaneous	tt0034583,tt0043265,tt0033870,tt0037382
7	nm0000008	Marlon Brando	1924	2004	actor,director,writer	tt0078788,tt0068646,tt0047296,tt0044081
8	nm0000009	Richard Burton	1925	1984	actor,producer,director	tt0061184,tt0087803,tt0059749,tt0057877
9	nm0000010	James Cagney	1899	1986	actor,director,producer	tt0029870,tt0031867,tt0042041,tt0034236

Fuente: IMDb (Non-Commercial Datasets)

## title.ratings.tv

Contiene los rankings y cantidad de votos que ha recibido cada título, contiene las siguientes columnas:

- tconst: String, sirve como identificador único de los títulos.
- averageRating: Float, puntuación promediada de los votos.
- numVotes: Int, cantidad de votos que ha recibido el título.

## title.crew.tsv

Contiene los identificadores de los directores y escritores de cada título, contiene las siguientes columnas:

- tconst: String, sirve como identificador único de los títulos.
- directors: Array de strings, identificadores de los directores del título.
- writers: Array de strings, identificadores de los escritores del título.

	tconst	averageRating	numVotes		tconst	directors	writers
0	tt0000001	5.7	2178	0	tt0000001	nm0005690	\N
1	tt0000002	5.5	299	1	tt0000002	nm0721526	\N
2	tt0000003	6.4	2243	2	tt0000003	nm0721526	nm0721526
3	tt0000004	5.2	193	3	tt0000004	nm0721526	\N
4	tt0000005	6.2	2986	4	tt0000005	nm0005690	\N
5	tt0000006	5.0	218	5	tt0000006	nm0005690	\N
6	tt0000007	5.3	927	6	tt0000007	nm0005690,nm0374658	\N
7	tt0000008	5.3	2335	7	tt0000008	nm0005690	\N
8	tt0000009	5.3	228	8	tt0000009	nm0085156	nm0085156
9	tt0000010	6.8	8067	9	tt0000010	nm0525910	\N

Fuente: IMDb (Non-Commercial Datasets)



## title.principals.tsv

Relaciona los títulos con las personas, mostrando que rol tiene cada persona en los títulos, contiene las siguientes columnas:

- tconst: String, sirve como identificador único de los títulos.
- ordering: Int, sirve para diferenciar las filas, enumerando las relaciones entre personas y una misma película
- nconst: String, sirve como identificador único de las personas
- category: String, la categoría de trabajo que hace la persona

	tconst	ordering	nconst	category	job	characters
0	tt0000001	1	nm1588970	self	\N	["Self"]
1	tt0000001	2	nm0005690	director	\N	\N
2	tt0000001	3	nm0005690	producer	producer	\N
3	tt0000001	4	nm0374658	cinematographer	director of photography	\N
4	tt0000002	1	nm0721526	director	\N	\N
5	tt0000002	2	nm1335271	composer	\N	\N
6	tt0000003	1	nm0721526	director	\N	\N
7	tt0000003	2	nm0721526	writer	\N	\N
8	tt0000003	3	nm1770680	producer	producer	\N
9	tt0000003	4	nm0721526	producer	producer	\N

Fuente: IMDb (Non-Commercial Datasets)

### Nota sobre el uso de los datos:

Los datos utilizados en este proyecto provienen de **Watchmode API** y de los **IMDb Non-Commercial Datasets**. El uso de los datos de IMDb se realiza bajo las condiciones de sus *Noncommercial Datasets*, disponibles en [IMDb Non-Commercial Datasets](#), exclusivamente para fines académicos y de investigación no comercial.

## Pregunta de investigación:

1. ¿Qué plataforma presenta mayor cantidad de películas?
2. ¿Cuál es la plataforma que tiene mayor variedad de géneros?
3. ¿Cuál es la plataforma que presenta mayor cantidad de películas actuales y cuál la que concentra más películas antiguas en su catálogo?
4. ¿Cuál es la plataforma que concentra la mayor cantidad de películas con mejores calificaciones?
5. ¿Qué actores se repiten más en el catálogo de cada plataforma?
6. ¿Qué directores se repiten más en el catálogo de cada plataforma?
7. ¿Qué plataforma presenta mejor la relación calidad y costo mensual percibida de su catálogo?

Estas preguntas buscan abarcar diferentes dimensiones que permitan analizar en profundidad las características de cada plataforma. Por un lado, se pretende estudiar la diversidad y amplitud de la oferta (cantidad de géneros, películas actuales o antiguas, tamaño total del catálogo) y, por otro, la calidad del contenido disponible (películas mejor calificadas por críticos o usuarios).

Asimismo, incorporar la variable económica del costo mensual frente a calidad del catálogo que permite dar una mirada más global que no solo se centre en la cantidad de películas disponibles, sino también en la conveniencia y el valor percibido por el usuario. De esta manera, estas preguntas no solo ayudarán a identificar patrones y diferencias entre servicios de streaming, sino también a establecer comparaciones objetivas que sean útiles para quienes buscan optimizar tanto su experiencia de entretenimiento como sus gastos.

En conjunto, estas interrogantes constituyen la base de nuestro análisis, ya que conectan directamente la información que obtendremos de los datos con las necesidades de la audiencia objetivo, entregando respuestas prácticas y concretas.

## Diseño tentativo:

Los métodos computacionales que ocuparemos serán:

- Recopilación y organización de datos a través de API: uso de requests para poder realizar la obtención, almacenamiento y manipulación en pandas.
- Integración de fuentes(precios de las plataformas, catálogos y clasificaciones): Haremos la unión mediante join keys, ordenamiento de los formatos y la eliminación de sus duplicados, creación de alias si es que llegan a faltar IDs
- Limpieza y validación de los datos: Proceso de datos faltantes, datos que tengan una desviación significativa a lo que se está buscando o su desviación sea mucha con otros datos a comparar(manejo de outliers), y eliminación de datos que sean nulos o estén en rango inválidos
- Modelado e feature engineering : Creación de las variables comparables entre las plataformas (ej: tamaño de los catálogos, rating de las películas, variedad de géneros, entre otros). A través de la normalización de características combinaremos en un análisis multicriterio las comparaciones entre las plataformas.
- Simulación y eficiencia: Creación de prototipo virtual para la simulación del proyecto que nos ayude a anticipar el comportamiento de esta antes del prototipo final, permitiéndonos optar por alternativas de diseño, optimización de parámetros e identificación de problemas que se nos podrían dar. Implementación de simulaciones de escenarios sobre nuestras variables.

Métodos estadísticos:

- Descriptivos: Por cada plataforma y género calcularemos media, desviación estándar y percentiles, además del rating ponderado por los votos y así poder impedir datos inconsistentes a la hora de ver las valoraciones de las diferentes películas.
- Análisis de datos: Veremos las distintas comparaciones entre precio, calidad, variedad, entre otros a través de la correlación. Se realizará un score multicriterio con métricas normalizadas.
- Interpretación de datos: Conclusiones de distintos análisis que realizaremos en relación sobre la calidad de las distintas plataformas, entre otros. Comprender la implicación de los resultados obtenidos y a base de esto tener una conclusión clara.
- Presentación: Entrega de resultados mediante realización de gráficos/ tablas(matplotlib/plotly) a través de leyendas, notas y sus enlaces respectivos de estos datos, con el fin de mostrar una evidente visualización de comparación entre estas plataformas.