

## Aplicación de Analítica a un Caso Estudio en Recursos Humanos

**Autores: Laura Herrera, Carlos Vásquez y Yuddy Giraldo**

### *a. Diseño de la solución*

<b>Problema de negocio:</b> Mejorar la experiencia de los usuarios para aumentar la fidelización y la recomendación de nuevos clientes	<b>Problema analítico:</b> Recomendar películas a los clientes según sus preferencias e implementar diferentes tipos de sistemas de recomendación.
<b>Zoom a la solución analítica</b> <p>Se han diseñado cuatro modelos de sistemas de recomendación para sugerir las próximas películas que podrían ser de interés para los usuarios. Estos resultados se integrarán en un producto consumible. Los sistemas de recomendación se actualizan periódicamente para garantizar su precisión y relevancia.</p>	

### **Solución:**

- **Sistema de recomendaciones basadas en popularidad:** Un Top 10 diario de las películas más populares entre todos los usuarios de la plataforma.  
*Diariamente* se toma la información de los usuarios y se actualiza el sistema basado en popularidad, este sistema se mostrará en la pantalla de inicio de la aplicación y se pone una sección de "Películas Populares" en la página principal, mostrando las películas más vistas y mejor valoradas por la comunidad.
- **Sistemas de recomendación de contenido general:** Top 5 - Películas similares, mostrar películas similares a las que el usuario ha visto, basándose en el género.  
Identificar películas dentro del catálogo que comparten géneros o temas similares a las que el usuario ha visto. *Semanalmente*, el sistema actualiza estas recomendaciones para reflejar las últimas interacciones del usuario y las incorporaciones al catálogo de películas. Al final de cada película, se destaca el "Top 5 - Películas similares", ofreciendo al usuario una selección personalizada y diversa que se alinea con sus intereses cinematográficos.
- **Sistemas de recomendación de filtros colaborativos:** Top 10 Recomendaciones de Usuarios Similares. Este sistema se basa en la interacción entre usuarios de la plataforma. Analiza las preferencias y comportamientos de usuarios con gustos similares al del usuario actual para generar recomendaciones personalizadas. quincenalmente, el sistema recopila datos de interacciones entre usuarios y películas, identificando patrones de consumo y afinidades entre diferentes perfiles, y se sugieren películas que otros usuarios con gustos afines han disfrutado y valorado positivamente. En la pantalla de inicio de la aplicación, se presenta una sección de "Recomendaciones de otros usuarios". Este sistema se actualiza

quincenalmente para reflejar las últimas interacciones y garantizar una experiencia de recomendación dinámica y relevante para cada usuario.

- **Sistemas de recomendación basados en el contenido de cada usuario:** "Perfil Personalizado". Este sistema crea un perfil único para cada usuario, analizando sus interacciones pasadas con películas.

En la página principal, se muestra una sección de "Recomendaciones Personalizadas", ofreciendo películas relevantes basadas en el perfil del usuario. Se actualiza mensualmente para adaptarse a las preferencias cambiantes del usuario y mejorar la experiencia de recomendación.

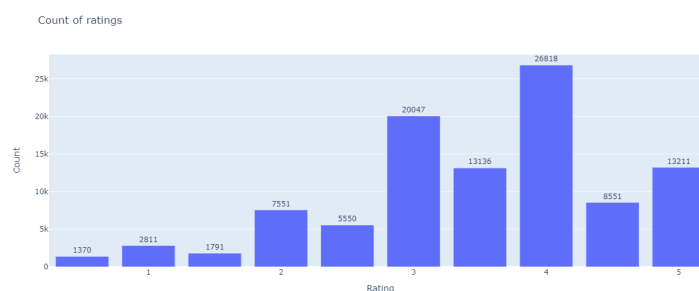
### ***b. Limpieza y transformación***

Se abordan aspectos como la detección y tratamiento de valores nulos, duplicados, así como la conversión de tipos de datos y la preparación de los datos para su análisis posterior. Al verificar la presencia de valores nulos en los conjuntos de datos de películas y calificaciones, no se detectan valores nulos ni registros duplicados en ninguno de los conjuntos de datos. Se crea una nueva base de datos "db\_movies2.db" que incluye datos procesados y una tabla combinada "full\_ratings" con información detallada sobre calificaciones y películas.

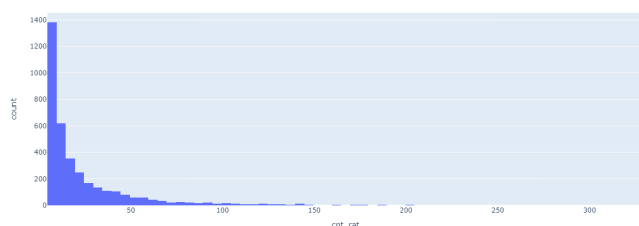
### ***c. Análisis exploratorio***

En el análisis exploratorio, se observó que los usuarios tienden a dar calificaciones altas, siendo 4 la calificación más común, y la menos común 0.5. Además, se encontró una gran variabilidad en la cantidad de calificaciones otorgadas por cada usuario, con algunos usuarios proporcionando un número considerablemente mayor de calificaciones que otros. Para simplificar el análisis, se decidió considerar únicamente los usuarios que han calificado un máximo de 1000 películas.

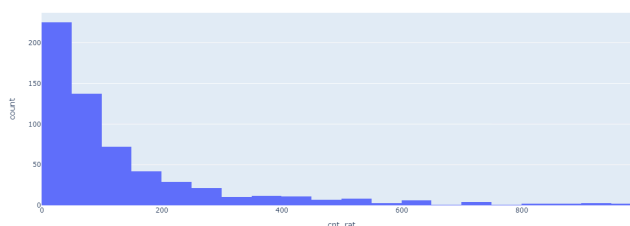
En cuanto a las películas, se encontró que algunas reciben muchas más calificaciones que otras, lo que sugiere una variabilidad significativa en la popularidad de las películas. La mayoría de las películas tienen un número relativamente bajo de calificaciones, con un promedio de aproximadamente 10.37 calificaciones por película. Para reducir el sesgo causado por las películas con un número muy bajo de calificaciones, se decidió tomar en cuenta solo aquellas que tienen más de 5 calificaciones, lo que proporcionará una muestra más representativa de las películas populares en la base de datos.



Histograma frecuencia de número de calificaciones por película



Histograma frecuencia de número de calificaciones por usuario



#### d. Algoritmos y técnicas de modelado

**Sistemas de recomendación por popularidad:** Se seleccionan las películas que han sido vistas por más de 20 personas, se calcula su promedio de calificación y devuelve las 10 películas con el promedio de calificación más alto, ordenadas primero por el promedio de calificación y luego por el número de personas que han visto la película.

*Top 10 de las películas más vistas y con mejores calificaciones.*

movie_title	avg_rat	movie_num
Shawshank Redemption, The (1994)	4.442177	294
Manchurian Candidate, The (1962)	4.434783	23
Raging Bull (1980)	4.403846	26
Hoop Dreams (1994)	4.400000	25
Rosemary's Baby (1968)	4.363636	22
Lawrence of Arabia (1962)	4.319444	36
Cinema Paradiso (Nuovo cinema Paradiso) (1989)	4.307692	26
In the Name of the Father (1993)	4.300000	20
Cool Hand Luke (1967)	4.287234	47
Philadelphia Story, The (1940)	4.282609	23

Se seleccionan las películas más vistas por año de publicación, devolviendo solo la mejor película de cada año, junto con su año de publicación y el número máximo de vistas para ese año.

*Top de las películas más vistas de cada año.*

Año	movie_title	num_vistas
1996	Batman (1989)	83
1997	Independence Day (a.k.a. ID4) (1996)	27
1998	Air Force One (1997)	6
1999	Saving Private Ryan (1998)	15
2000	American Beauty (1999)	27
2001	Star Wars: Episode IV - A New Hope (1977)	13
2002	Silence of the Lambs, The (1991)	11
2003	Lord of the Rings: The Two Towers, The (2002)	14
2004	Austin Powers: The Spy Who Shagged Me (1999)	9
2005	Sixth Sense, The (1999)	14
2006	Lord of the Rings: The Two Towers, The (2002)	17
2007	Fight Club (1999)	12
2008	Pulp Fiction (1994)	16
2009	Forrest Gump (1994)	7
2010	Avatar (2009)	12
2011	Inception (2010)	11
2012	Forrest Gump (1994)	16
2013	Godfather, The (1972)	10
2014	Matrix, The (1999)	8
2015	Matrix, The (1999)	32
2016	Dark Knight, The (2008)	23
2017	Shawshank Redemption, The (1994)	29
2018	Inception (2010)	18

**Sistemas de recomendación de contenido general:** este sistema de recomendación utiliza el algoritmo de vecinos más cercanos (KNN) y variables dummy de género de película para recomendar películas similares a una película dada. La interfaz interactiva permite a los usuarios explorar y descubrir nuevas películas basadas en sus preferencias. Por ejemplo, para la película Toy Story se recomiendan las siguientes películas.

movie_name	Toy Story (1995) ▼
Películas Recomendadas	
0	Antz (1998)
1	Toy Story 2 (1999)
2	Adventures of Rocky and Bullwinkle, The (2000)
3	Emperor's New Groove, The (2000)
4	Monsters, Inc. (2001)

**Sistema de recomendación basado en el contenido de cada usuario:** En este enfoque, primero se seleccionan las películas calificadas por el usuario para calcular su centroide o perfil. Luego, se utiliza un modelo de vecinos más cercanos (KNN) entrenado con las películas no vistas por el usuario. Este modelo encuentra las películas más similares al perfil del usuario y las recomienda como sugerencias.

user_id	1 ▼	
	movie_clean_title	movie_id
1798	Constantine (2005)	31696
4831	Kizumonogatari Part 1: Tekketsu (2016)	162968
6837	Enough (2002)	5387
1162	Third Man, The (1949)	1212
3393	Ninotchka (1939)	936
1453	Where the Heart Is (2000)	3565
3206	Pat Garrett and Billy the Kid (1973)	7889
3805	Crazies, The (a.k.a. Code Name: Trixie) (1973)	6395
1282	Kung Fu Panda (2008)	59784
3088	Dad (1989)	4581

**Sistema de recomendación de filtros colaborativos:** En este apartado, se calcula el desvío de cada calificación de película con respecto al promedio. Luego, se utiliza el algoritmo de vecinos más cercanos (KNNBaseline) para encontrar usuarios con gustos similares y generar predicciones sobre cómo un usuario podría calificar películas no vistas. Estas predicciones se utilizan para recomendar al usuario las 10 películas más relevantes y que podrían interesarle.

	iid	est	movie_id	movie_clean_title	movie_year
0	318	5.215185	318	Shawshank Redemption, The (1994)	1994
1	7991	5.175682	7991	Death Race 2000 (1975)	1975
2	1228	5.087423	1228	Raging Bull (1980)	1980
3	912	5.082546	912	Casablanca (1942)	1942
4	178827	5.063936	178827	Paddington 2 (2017)	2017
5	1276	5.043140	1276	Cool Hand Luke (1967)	1967
6	1225	5.032114	1225	Amadeus (1984)	1984
7	166291	5.022270	166291	A Silent Voice (2016)	2016
8	164917	5.021129	164917	13th (2016)	2016
9	58559	5.012544	58559	Dark Knight, The (2008)	2008

### ***e. Despliegue del modelo***

Esta etapa facilita la disponibilidad de los resultados del Sistema de Recomendación basado en el contenido de cada usuario en varios formatos. Estos formatos incluyen archivos planos, como CSV, que son fácilmente utilizables para análisis adicionales en Python, así como archivos XLSX, que permiten una visualización más intuitiva y amigable de las recomendaciones. Este sistema se actualizará de forma mensual, tanto el entrenamiento de los datos, como las predicciones.

### ***f. Conclusiones***

La implementación de cuatro modelos diferentes de sistemas de recomendación muestra una comprensión profunda de las necesidades y preferencias de los usuarios. Esta diversidad asegura que se abordan diferentes aspectos de la recomendación de películas, desde la popularidad general hasta las preferencias individuales de los usuarios.

La decisión de actualizar periódicamente los sistemas de recomendación demuestra una mentalidad de mejora continua y adaptación a medida que cambian las preferencias de los usuarios y se actualiza el catálogo de películas.