

# **CINEMETRICS: MODELADO PREDICTIVO DE CALIFICACIONES CINEMATOGRAFICAS DE MACHINE LEARNING**



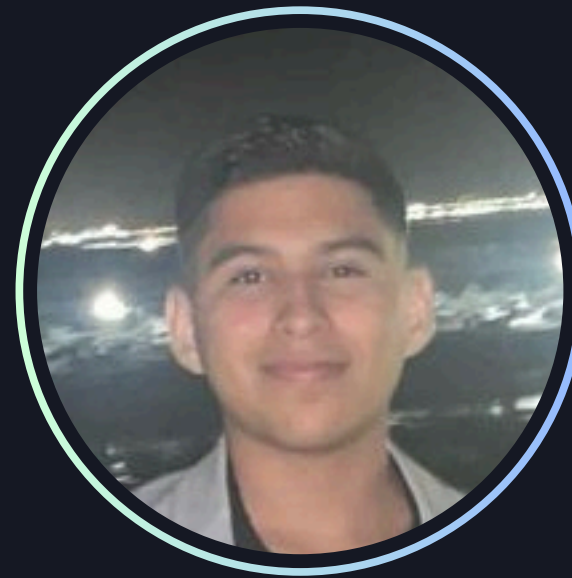
# INTEGRANTES



Andrés Flores  
8-1025-1254



Rashell Vidal  
8-1028-643



Aldahir Aguilar  
8-1029-1115



Diego Garcia  
8-1034-95



Luis Torné  
8-1032-1466

# Preguntas Personales



## ¿Qué le motivó a nivel personal a desarrollar el estudio predictivo?

Lo que más me motivó fue la posibilidad de trabajar en algo real usando machine learning y hacer un proyecto que podría usarse en otros estudios o para hacer predicciones verdaderas.

## ¿Cómo fue su experiencia buscando y seleccionando la data?

Fue complicado. Al inicio dependíamos mucho de encontrar fuentes que se adaptaran lo mejor posible a lo que estábamos buscando.

## ¿Qué lecciones aprendió al hacer el análisis descriptivo?

Aprendí que es necesario separar las variables más importantes y limpiar bien los datos y explorar la distribución de las variables para detectar valores atípicos o patrones.

# Preguntas Personales



## ¿Qué pasos realizó para hacer el análisis de los datos tanto descriptivo como predictivo?

Se limpió la base de datos, se analizaron variables clave como género y edad, y se usaron gráficos para explorar la distribución y correlación. Luego, se codificaron las variables, se dividieron los datos y se probaron varios modelos de regresión. Random Forest fue el más preciso, aunque presentó algo de sobreajuste.

## ¿Qué descubrió de la investigación?

Descubrí que es importante evaluar tanto la precisión ( $R^2$ ) como el error (MAE) para elegir el modelo más equilibrado y confiable.

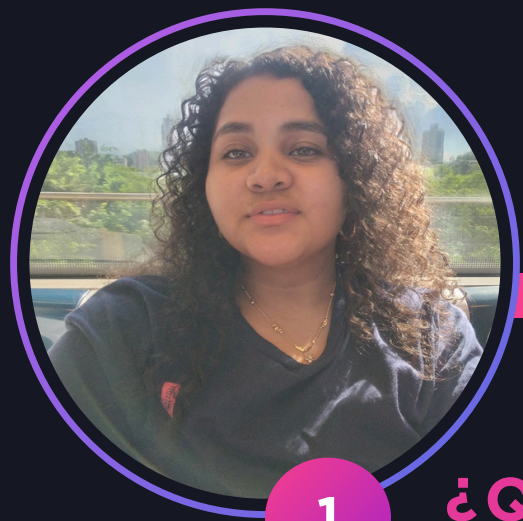
## Si fuese a hacer este estudio de nuevo, ¿qué haría mejor?

Gestionar mejor la búsqueda de una base de datos porque si nos tomó más tiempo de lo esperado ir recopilando todo de cero lo que nos afectó un poco a la hora de seguir desarrollando lo demás.

## ¿Qué descubrió de sí mismo?

Descubrí que A la hora de trabajar en grupo y me toca coordinar prefiero tener como una mano derecha para hacer más llevadero el trabajo





# Mi Jornada y Preguntas Personales

1

## ¿Qué le motivó a nivel personal a desarrollar el estudio predictivo?

En lo personal, desde pequeña he sido bastante fan de las películas de varios géneros, con el pasar del tiempo he notado el cambio en la industria con la llegada y el auge de las distintas plataformas de streaming, y me emociona el hecho de evaluar y estudiar un modelo predictivo que favorezca a el cine, a las industrias, pero sobre todo a la audiencia que es quien hace posible que el mundo del cine se mantenga en pie.

2

## ¿Cómo fue su experiencia buscando y seleccionando la data?

La experiencia fue desafiante. Inicialmente, traté de encontrar bases de datos ya existentes en internet, pero ninguna cumplía con los criterios que tenía en mente. Eso me llevó a investigar cómo recolectar datos por mí mismo, y fue ahí cuando descubrí que podía acceder a TMDb a través de una API. Fue un proceso largo, sobre todo porque el script que escribí tardó muchas horas en completarse, pero fue muy gratificante ver cómo el esfuerzo rendía frutos y generaba una base de datos de más de 40,000 películas.

3

## ¿Qué lecciones aprendió al hacer el análisis descriptivo?

Aprendí que, por más que los datos estén bien estructurados, siempre hay detalles que deben ser corregidos o normalizados, lo que suele ocupar mucho esfuerzo físico, mental y de tiempo si no se automatiza dicho trabajo. También entendí el valor de las visualizaciones para descubrir patrones y relaciones que a simple vista no se perciben, es útil para determinar variables tanto dependientes como independientes y resulta favorable para este tipo de proyecto en el que trabajamos.

## ¿Qué pasos realizó para hacer el análisis de los datos tanto descriptivo como predictivo?

Desde mi experiencia, comencé realizando un análisis descriptivo para familiarizarme con la base de datos. Me dediqué primero a limpiarla y depurarla, enfocándome en identificar las variables más relevantes, como la calificación del actor (actor\_rating), el género y el año de estreno. Para profundizar el análisis, filtré los 50 actores que aparecían con mayor frecuencia y analicé cómo se comportaba la variable objetivo, imdb\_pelicula, gráficos KDE y un mapa de correlación que me ayudó a identificar relaciones significativas entre variables.



## ¿Qué descubrió de la investigación?

En los primeros resultados de las graficas pudimos deducir la gran correlación y dependencia entre las variables imdb\_pelicula, director\_ranking y actor\_ranking, ya que el éxito de imdb\_pelicula hace que las otras dos variables tengan éxito o decaigan de manera significativa. También se observa el genero mas destacable y con un posible éxito entre el publico.

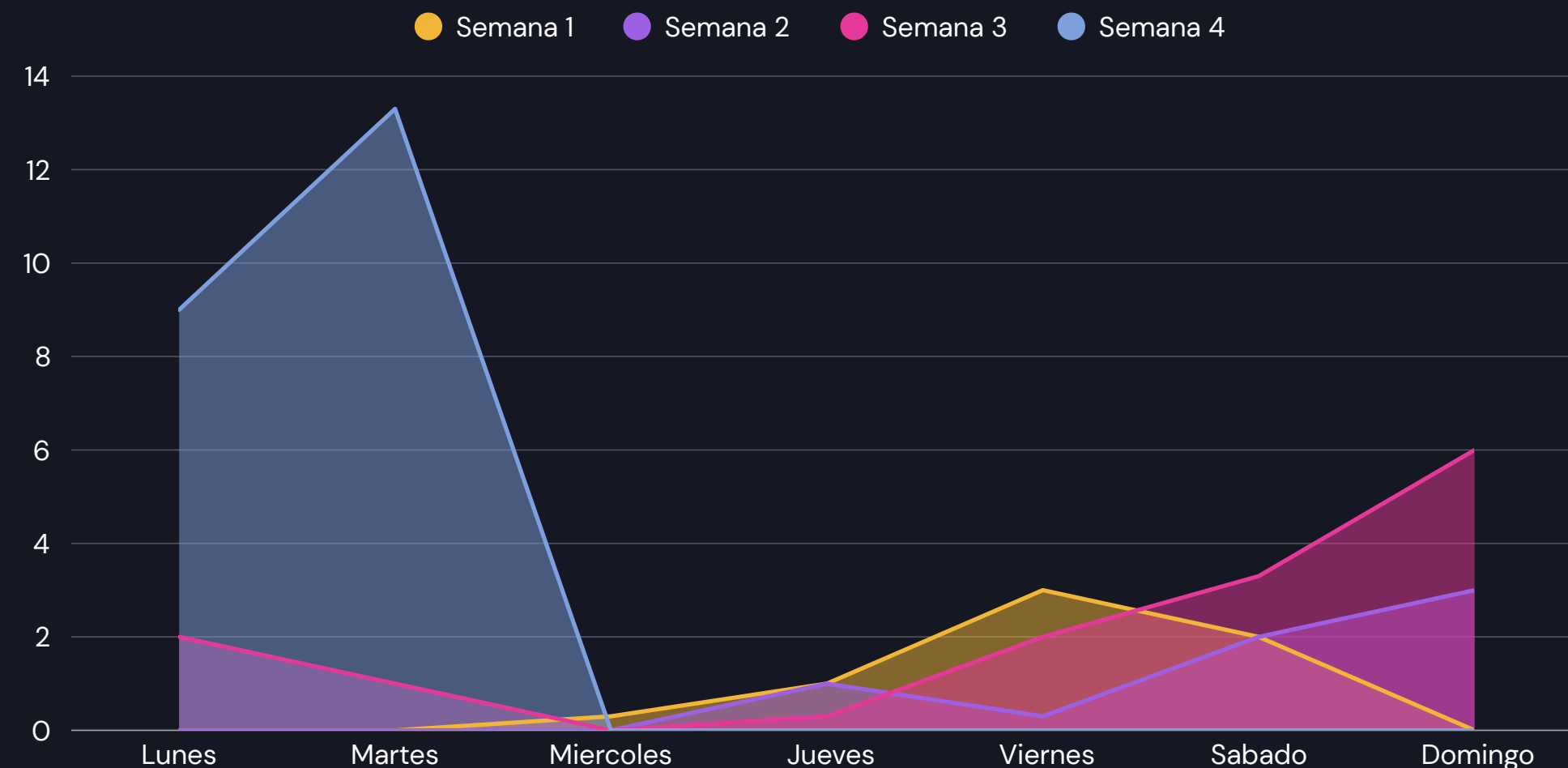
## Si fuese a hacer este estudio de nuevo, ¿qué haría mejor?

Desde el inicio, planearía con más detalle las variables a recolectar. También automatizaría mejor el proceso de limpieza de datos y me apoyaría más en herramientas estadísticas desde el principio, como Scikit-learn para la predicción o Power BI para las visualizaciones. Eso habría optimizado tiempos y mejorado aún más los resultados.

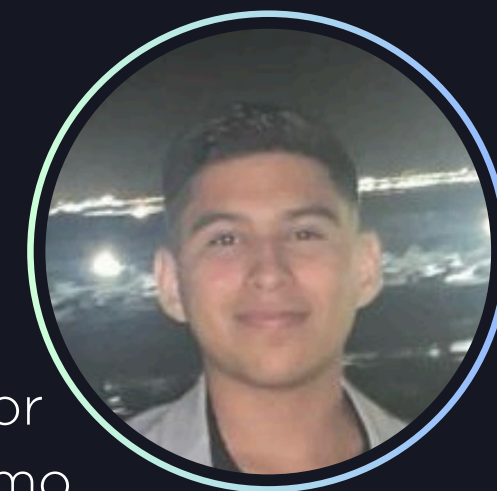
## ¿Qué descubrió de sí mismo?

Lo que aprendí de mí mismo durante este proceso fue mi capacidad para mantener la perseverancia y el compromiso, incluso cuando las tareas se volvieron tediosas o frustrantes. Me di cuenta de que tengo un enfoque detallista y que realmente me esfuerzo por hacer las cosas bien, especialmente al ver cómo pequeños errores en los datos podían afectar significativamente los resultados.

Tiempo Invertido Para el Data Bases



# Preguntas Personales



## **A ¿Qué le motivo a nivel personal a desarrollar el estudio predictivo?**

Siempre me han llamado la atención las películas. Me gusta observar qué hace que unas funcionen mejor que otras y qué factores influyen en eso. Quería ir más allá de solo verlas, así que usé este proyecto como una forma de analizar ese mundo desde otro ángulo, con datos y herramientas que manejo.

## **B ¿Cómo fue su experiencia buscando y seleccionando la data?**

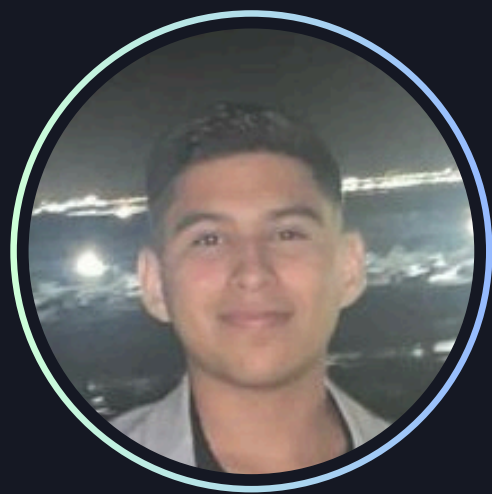
Fue un reto. Al principio fue difícil no encontramos una base que se adaptara, así que hicimos un script en Python con la API de TMDb para automatizar todo. Tuvimos problemas con datos duplicados, así que creamos otro script para corregir eso y generar una base limpia en SQL llamada cinemetrics.

## **C ¿Qué lecciones aprendió al hacer el análisis descriptivo?**

Aprendí que los datos tienen mucho que decir, pero no todo es tan directo. No basta con mostrarlos en una gráfica, hay que entender bien qué significan y si realmente representan algo útil. También entendí que no todas las variables sirven, algunas solo hacen ruido y pueden confundir. Lo importante es saber interpretar bien lo que se está viendo.

## **d ¿Qué pasos realizó para hacer el análisis de los datos tanto descriptivo como predictivo?**

Primero limpié los datos y escogí variables clave como actor\_rating, género y año. Analicé la distribución de la calificación (imdb\_pelicula) y las relaciones entre variables. Luego codifiqué datos, dividí en entrenamiento y prueba, y probé varios modelos. El que mejor funcionó fue Random Forest, aunque mostró un poco de sobreajuste.



# Preguntas Personales

## ¿Qué descubrió de la investigación?

Descubrí que aunque es posible predecir el éxito de una película en base a características como el actor principal, el género o el director, el resultado nunca será exacto. El modelo que utilicé (Random Forest) logró explicar un 61% del comportamiento del público, lo cual es aceptable, pero también mostró que hay factores impredecibles que influyen en la valoración de una película. Esta investigación me permitió entender que el análisis de datos puede acercarnos bastante a la realidad, pero siempre habrá un margen de error, especialmente en fenómenos tan complejos como las opiniones del público.



## Si fuese a hacer este estudio de nuevo, ¿qué haría mejor?

Organizaría mejor el flujo de trabajo. Guardaría versiones intermedias del análisis y documentaría mejor cada paso, para facilitar ajustes y correcciones. También experimentaría con modelos más avanzados y variables externas como presupuesto o taquilla.



## ¿Qué descubrió de sí mismo?

Que me gusta enfrentarme a problemas complicados. Disfruté todo el proceso, incluso cuando me costó avanzar o me equivoqué.







# Mis experiencias con respecto al proyecto

## ¿Qué le motivó a nivel personal a desarrollar el estudio predictivo?

Quise aplicar lo aprendido sobre análisis de datos en un tema que me interesa, y el cine nos pareció una forma más accesible y entretenida de hacerlo, ya que muestra cómo los datos pueden anticipar resultados.

## ¿Cómo fue su experiencia buscando y seleccionando la data?

FComo no encontramos una base de datos adecuada, creamos un script en Python con la API de TMDb para recolectar los datos. Tuvimos que corregir duplicados en SQL y convertir la base a Excel, lo cual fue complicado por el formato de los géneros. Finalmente, exportarla a CSV desde MySQL Workbench facilitó el proceso. Después de varias semanas, logramos construir una base de datos funcional y limpia.

## ¿Qué lecciones aprendió al hacer el análisis descriptivo?

La lección más importante fue entender que un análisis solo es confiable si los datos están limpios y completos. Aprendí que una base mal depurada puede llevar a resultados incorrectos, y también comprendí cómo clasificar las variables y tratarlas según su tipo, ya sea con gráficas, tablas o análisis estadísticos.



## ¿Qué pasos realizó para hacer el análisis de los datos tanto descriptivo como predictivo?

Primero analicé y limpié los datos, seleccionando variables clave y explorando la distribución de IMDb. Luego apliqué distintos modelos de regresión, y Random Forest fue el que mejor desempeño tuvo, aunque con algo de sobreajuste.



# Mis experiencias con respecto al proyecto

## ¿Qué descubrió de la investigación?

En lo análisis que hicimos con respecto a las gráficas, vimos que el modelo identificó al director nominado como el principal predictor del éxito en IMDb, por encima de género, año o actor. Esto puede orientar futuras producciones a priorizar directores reconocidos si el objetivo es lograr mayor impacto crítico.

## Si fuese a hacer este estudio de nuevo, ¿qué haría mejor?

Automatizaría más procesos desde el inicio, ya que tareas como la limpieza de datos fueron manuales y tomaron mucho tiempo. También exploraría otros modelos y usaría herramientas de visualización para presentar mejor los resultados.

## ¿Qué descubrió de sí mismo?

Descubrí que tengo más capacidad de análisis de lo que pensaba. Este proyecto me hizo tomar decisiones y resolver problemas en tiempo real, y me di cuenta de que disfruto enfrentando desafíos tanto técnicos como prácticos.



# Preguntas Personales



## ¿Qué le motivó a nivel personal a desarrollar el estudio predictivo?

Siempre me ha gustado entender el "por qué" detrás de las cosas. Me pareció fascinante usar datos para tratar de anticipar qué factores hacen que una película sea exitosa.



## ¿Qué lecciones aprendió al hacer el análisis descriptivo?

Aprendí que un análisis descriptivo bien hecho es esencial para cualquier estudio predictivo. Es como conocer a fondo a tu sujeto antes de hacer suposiciones sobre él.



## ¿Cómo fue su experiencia buscando y seleccionando la data?

Fue muy educativa. Aunque fue difícil al principio, la experiencia de usar una API real y construir la base desde cero fue algo muy valioso.



## ¿Qué pasos realizó para hacer el análisis de los datos tanto descriptivo como predictivo?

El primer paso fue comprender cada variable: qué representaba y cómo se medía. Luego hice análisis por tipo de variable (nominal, ordinal, continua) y detecté tendencias generales.



### **¿Qué descubrió de la investigación?**

Descubrí que el cine tiene patrones, incluso cuando parece subjetivo. También entendí cómo ciertas características, como el historial del director o la combinación de géneros, pueden influir en el resultado final de forma consistente.



### **¿Si fuese a hacer este estudio de nuevo, ¿qué haría mejor?**

Empezaría por planear el análisis desde el diseño de la base. Pensaría más en cómo se van a usar las variables antes de recolectarlas.



### **¿Qué descubrió de sí mismo?**

Que disfruto más de lo que pensaba trabajar en proyectos largos y complejos. Me gustó ser parte de cada etapa, desde la recolección hasta las conclusiones.





Muchas gracias

