

CINEMETRICS: MODELADO PREDICTIVO DE CALIFICACIONES CINEMATOGRAFICAS DE MACHINE LEARNING



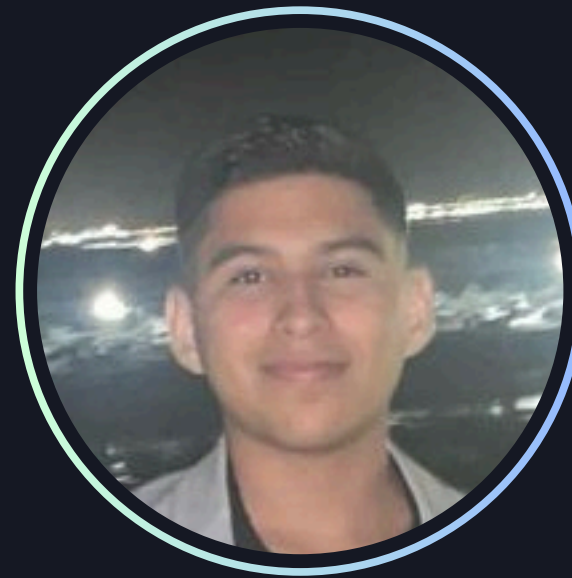
INTEGRANTES



Andrés Flores
8-1025-1254



Rashell Vidal
8-1028-643



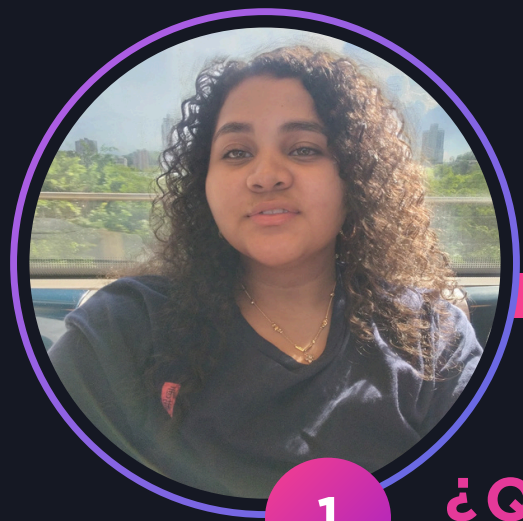
Aldahir Aguilar
8-1029-1115



Diego Garcia
8-1034-95



Luis Torné
8-1032-1466



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 te 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?

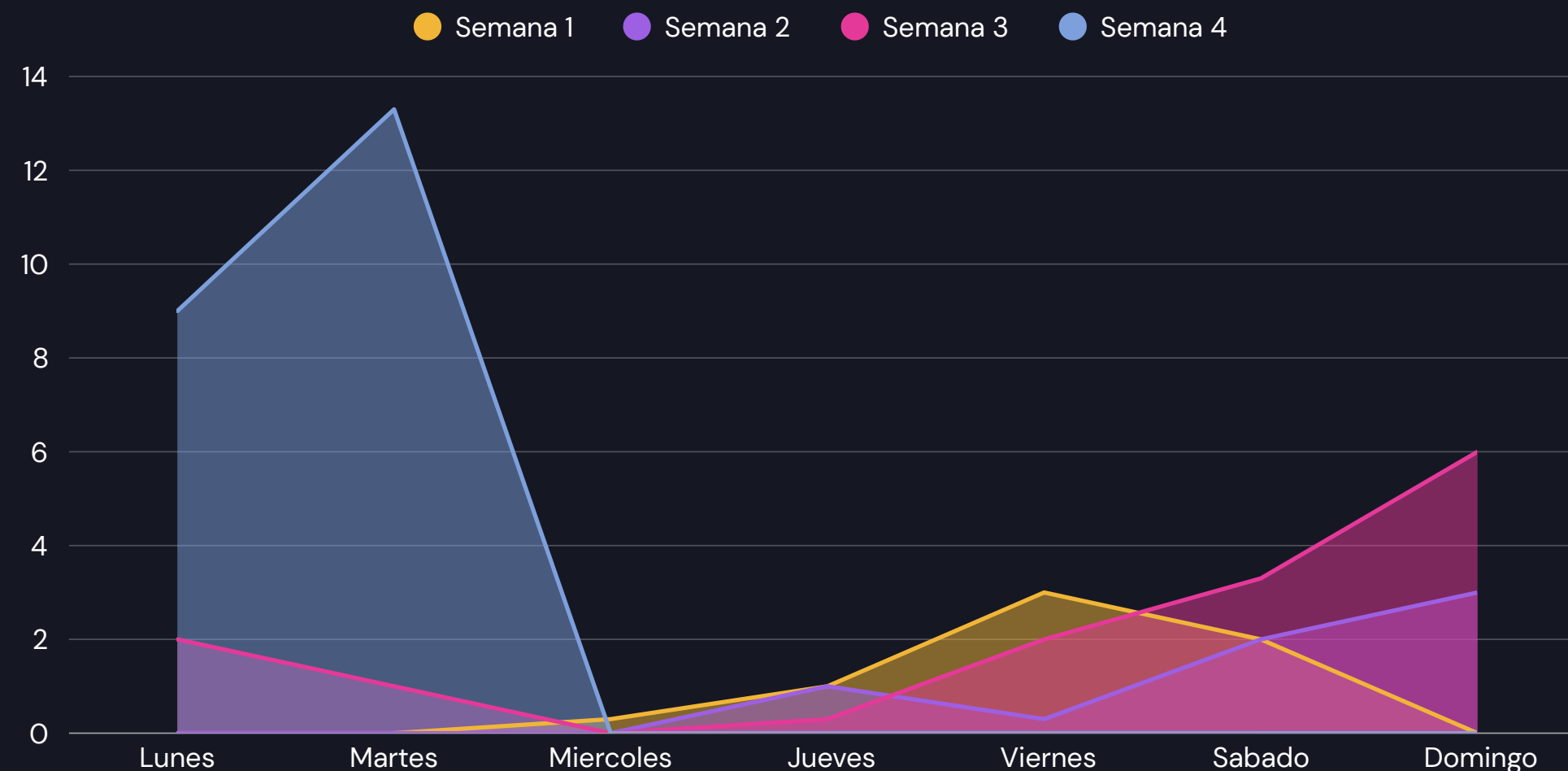
Desde el inicio, planearía con más detalle las variables a recolectar. También automatizaría mejor el proceso de limpieza de datos y me apoyaría más en herramientas estadísticas desde el principio, como Scikit-learn para la predicción o Power BI para las visualizaciones. Eso habría optimizado tiempos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





Muchas gracias

