

# INFORME NETFLIX

**Data Analyst: Posse Gonzalo**

Este trabajo arranco con una exploración de los datos en el cual pude descubrir cosas bastante importantes:

- El dataset esta compuesto por 8809 registros y están divididos en 26 columnas, de estas 26 columnas hay 14 que son inservibles y estaban vacías, con valores NaN, en todos los registros. Lo que hice fue eliminar estas columnas y quedarme unicamente con 12 columnas.
- De estas 12 columnas hay 1 que es de tipo numérico, integer, y las demas eran cadenas de texto, string/object. Pero habia un campo especial, date\_added, que contiene informacion sobre la fecha exacta en que fue añadido esa pelicula o serie a nuestra plataforma de Netflix. Se realizo una pequeña transformación sobre este campo para poder trabajarlo como dato de tipo Fecha más adelante.
- Existen muchos valores vacíos en nuestro conjunto de datos, para la columna Director tenemos casi el 30 % de los datos vacíos, es mucha data perdida que deberiamos de poder recuperar si tenemos acceso a fuentes externas de los datos. En las columnas cast y country teniamos un 9 %, en cada una, de valores vacíos que fue rellenado con el valor “Sin Data” para especificar que esos registros no podiamos conseguir los datos.
  - Para solucionar este problema de los datos vacíos, le propongo a la empresa, brindar acceso hacia todas las fuentes de datos, APIs, Bases de Datos Históricas, etc. Todo lo que tenga el area de datos para poder trabajarlo y hacer una limpieza correcta de los datos. Quiero hacer hincapié en que esto va a beneficiar a la empresa a futuro, ya que, de estos datos limpios salen los futuros reportes y analisis que son consumidos para tomar decisiones de inversión.
- Una vez finalizado el trabajo de limpieza del conjunto de datos inicial, se exporto la base de datos hacia Google BigQuery, el motor que utiliza la empresa, para poder trabajar en varias consultas, como por ejemplo
  - Si quisiéramos saber cuales son los principales países productores de titulos, en nuestra plataforma. Con una consulta básica podemos ver este ranking

Fila	country	total_titles
1	United States	3691
2	India	1046
3	United Kingdom	806
4	Canada	445
5	France	393
6	Japan	318
7	South Korea	232
8	Spain	232
9	Germany	226
10	Mexico	169

Donde vemos que Estados Unidos es el principal productor de Peliculas y series.

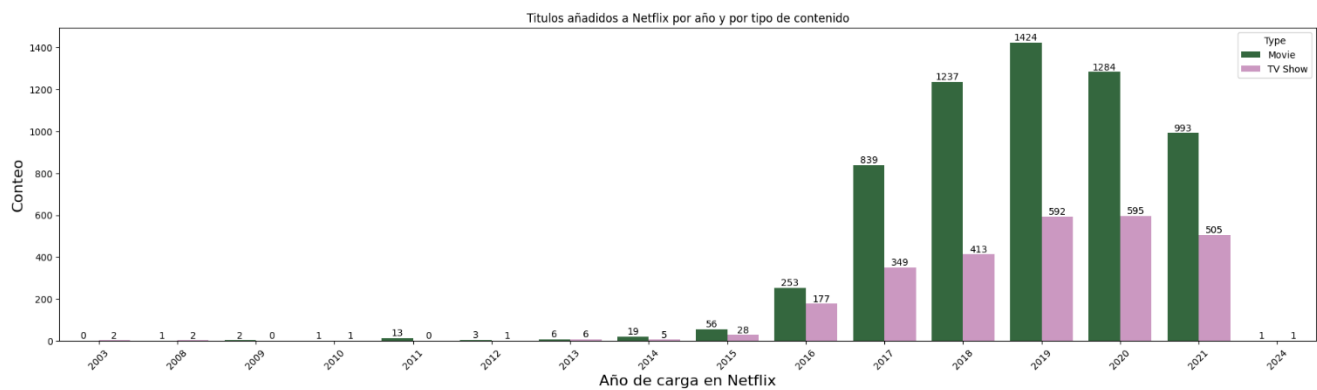
Si tuviéramos los datos de nuestros clientes, podríamos hacer un analisis y saber si estamos alineados entre nuestros principales productores y consumidores. Dada la densidad demográfica del mundo, podemos deducir que si, ya que Estados Unidos e India son nuestros principales productores y también son de los países que tienen mas población, por lo cual se puede inferir que pueden ser nuestros principales consumidores.

También, dentro de la base de datos, podemos buscar cuales son los géneros que mas films tenemos dentro de nuestra plataforma. Los resultados nos dicen que este es el TOP 5:

1. International Movies
2. Dramas
3. Comedias
4. Action & Adventure
5. Documentaries

Me gustaria destacar que en 6 puesto se encuentra Dramas, uno de mis géneros preferidos. Las tendencias actuales del mundo del cine estan llevando a los films mas éxitos a un genero de estilo “psicológico” o de “Suspenso” en donde la trama de la pelicula o serie toma unos giros inesperados en el desarrollo, es una gran recomendación empezar a producir o invertir en traer ese estilo hacia nuestra plataforma, ya que la competencia estuvo creciendo este ultimo tiempo en cuanto a plataformas de streaming.

En la misma linea de las consultas anteriores, también pudimos saber que progresión tuvo la empresa a la hora de subir contenido a la plataforma, en 2019 fue donde mas contenido subimos con un total de 2016 titulos, entre peliculas y series. Y aquí tenemos un grafico por si queremos verlo mejor



Este grafico se encuentra dentro del archivo EDA en donde fui realizando todas estas transformaciones.

Siguiendo con los hallazgos, tenemos muchos registros de títulos viejos pero los mas populares son luego de los 90's. Podriamos realizar una actualizacion masiva en donde pongamos un tope de año, por ejemplo 1980 como minimo, todos los films que sean anteriores a esa fecha eliminarlos o empezar a hacerles un seguimiento exhaustivo para ver que tan consumidos son. Deberiamos invertir en producir nuevas series o peliculas en vez de comprar derechos de títulos viejos.

Y ya para ir finalizando, también en el analisis exploratorio pudimos ver que la mayoría de las peliculas tienen una duración entre 70 y 120 minutos, el promedio esta en 90 minutos. Con esto podemos recomendar que no se carguen peliculas con una duración mayor a 150 minutos, ya que los clientes no van a consumirlas, estar casi 3 horas viendo una pelicula, para algunos, es una perdida de tiempo y van a preferir consumir estos títulos de menor duración.

Por el lado de las series, el promedio es de 1 a 2 temporada de duración, también tenemos que tener en consideración la cantidad de capitulos que contiene cada temporada, esto no es algo que podamos saber con los datos actuales, y el tiempo que demora cada capítulo, así que hacer una estimación en temporadas para saber el tiempo que le va a consumir a nuestros clientes no es algo que podamos estimar por ahora.

Acompañado a este informe va un archivo EDA de tipo Jupyter Notebook en el cual se realizaron transformaciones y hallazgos. También va un informe complementario en Looker Studio en donde estan los graficos interactivos sobre nuestros datos.

Muchas gracias por leer ;