

CONTENIDO DE PLATAFORMAS DE STREAMING

GRUPO 19.

Angel Villarreal Barroso - Sala 1

Marcela Valverde Estrada - Sala 1

Gisela Belen Figueroa - Sala 2

Gaston Carranza- Sala 1

Maria Belen Rodriguez - Sala 2

Maria Gabriela Rodriguez - Sala 2

INTRODUCCIÓN:

El objetivo del presente reporte es presentar los resultados de un análisis de datos utilizando el lenguaje de programación de Python. En este análisis, se exploró un conjunto de datos compuesto por los dataframes de las plataformas: HBO, Netflix, Amazon Prime y Disney Plus y se aplicaron diversas técnicas de análisis para obtener la información más relevante.

A continuación se detallan los pasos realizados, resultados obtenidos y las conclusiones alcanzadas durante el proceso.

OBJETIVO:

El objetivo del proyecto es colaborar con la identificación de las tendencias de mercado local, preferencias de contenido, oportunidades de expansión y segmentos de audiencia, a fin de respaldar la toma de decisiones estratégicas, como adquisiciones de nuevos contenidos con nuevas características.

ALCANCE:

Se trabajó sobre las tablas de todas las plataformas, a excepción de la tabla de HBO que se adecua para el trabajo sobre una de las dos tablas que componen el dataset, quedando de manera relegada, por una cuestión de tiempo, incorporar los datos de actores y directores que se encuentran en otra tabla. La incorporación y el trabajo de unir ambas tablas forman parte de un Anexo (ANEXO Trabajo sobre las tablas de HBO para dar formato) que se incorpora al trabajo.

METODOLOGÍA:

La metodología utilizada en este trabajo de análisis de datos de plataformas de streaming involucró la recopilación y procesamiento de datos, exploración y análisis de patrones de consumo, así como la visualización de resultados utilizando el lenguaje de programación Python, considerando librerías como Pandas, NumPy, Matplotlib y Seaborn, como así también GitHub como plataforma de colaboración de desarrollo.

Recopilación de datos:

Se obtuvo un conjunto de datos obtenidos desde Kaggle que contenía información sobre el contenido de las plataformas de Netflix, Amazon Prime y Disney Plus, además le sumamos la plataforma de HBO utilizando la misma fuente para enriquecer aún más nuestro análisis. Siendo que HBO su data frame data del año 2022 y las demás de 2021, es decir que este contenido no está actualizado a 2023.

Procesamiento de los datos:

Se realizó la limpieza inicial de los datos a través de diferentes reportes, observar valores nulos y completar registros, así como también se llevaron a cabo las transformaciones necesarias, como la conversión de tipos de datos a los de formato fechas.

Exploración de datos:

Se llevó a cabo un análisis exploratorio de los datos para obtener una comprensión inicial del conjunto de datos. Esto incluyó la visualización de distribuciones, la identificación de relaciones entre variables y el cálculo de estadísticas descriptivas.

Análisis específico:

Se aplicaron técnicas como clasificación, agrupamiento, filtrado, medidas de posición y dispersión entre otras, dependiendo del tipo de datos.

Visualización de resultados:

Se realizaron gráficos y visualizaciones para presentar los hallazgos del análisis de datos. Se utilizaron las bibliotecas de matplotlib y seaborn.

Control de versiones y colaboración:

Se utilizó GitHub como herramienta de control de versiones y colaboración en el proyecto de análisis de datos, se organizaron los archivos y se gestionaron los cambios utilizando ramas, en wikis se registraron las meets realizadas donde participaron todos los colaboradores del proyecto.

RESULTADOS:

A continuación, se presentan los principales resultados.

En relación a la limpieza de los datos y ordenamiento, se realizó la identificación de valores únicos, donde por ejemplo, se pudo identificar que la fuente de datos

adicional referida a HBO, tenía una tipificación diferente respecto a la denominación del país origen de la película respecto al resto de la plataformas, por lo que fue necesaria la unificación de estas descripciones, así como la clasificación entre película o serie (MOVIE / SHOW vs Movie / TV Show) por mencionar otro ejemplo. Para ello fue necesaria la carga en dataframe auxiliar de una tabla de códigos y descripción de países para luego realizar un merge con la tabla final, previo a la transformación del código de país en cuanto a su formato. Adicionalmente fue requerido el renombre y orden de los campos a fin de que el resultado final quede entendible para el análisis.

En el análisis más profundo se realizó un filtrado para observar por país la cantidad de contenido que poseían entre todas las plataformas dividiéndolas en series y países. También se encontraron aquellos contenidos de origen más antiguos y aquellos desarrollados en la contemporaneidad, siendo 'The Prince of Magicians' de 1901 la película más antigua y 'Goyo: En Letra de Otro' de 2022 la película más 'nueva' dentro del dataframe.

En cuanto a la obtención del promedio de duración del contenido tanto para series y películas, encontramos que para películas es del 78.17 minutos pero en las series es de 0.0, quiere decir que este campo para series no hay datos (solo está en número cero en su contenido).

Se realizaron los cálculos de moda y mediana, como así también aquellos que nos interesaba aplicar fórmulas para ver las estadísticas. Pudiendo observar que hay muchos valores nulos o unknown, siendo que la información no está completa. Además, se exploró las temporadas de series que tiene cada plataforma, siendo HBO la única que no tiene esta variable.

En relación a las visualizaciones se trató de hacer el recorrido desde lo general a lo particular, es decir, mostrar un análisis comparativo entre las plataformas para observar la cantidad de contenido que poseían las plataformas, como se divide tal contenido (clasificación), aquellos países que poseen más contenidos en tales plataformas.

Y luego se pasa a un análisis particular teniendo en cuenta a Argentina para visualizar el tipo de contenido, como se distribuye el mismo, aquellas plataformas que ofrecen más contenido para Argentina, y aquellos directores que más se repetían dentro de este análisis particular.

Además, se realizó un filtrado para que el usuario coloque el año de búsqueda para encontrar el contenido disponible tanto para series y películas.

Como resultado de estas visualizaciones y filtrado, podemos observar que:

- Contenido más popular: en términos generales el contenido que más se repite en las plataformas son las películas en un 75.2%.

- En cuanto al rating o calificación de cada contenido, en películas los contenidos para el público de TV-MA (no apto para menores de 14 años) es el que más se repite. Como así también 13+ y R (para mayores de 18 años). Para las series o TV Shows, las series de TV-MA es el contenido que más se repite, como así también TV-PG (Los padres o tutores pueden considerarla inapropiada para niños más pequeños).
- El año 2020 fue el año con mayor cantidad de contenido.
- En comparación entre plataformas, Amazon Prime fue la plataforma que mayor contenido tuvo en el año 2020, seguido por Netflix que su top de contenidos fue en el 2019. Pero todas coinciden que tienen una baja de contenido para el año 2021.
- La plataforma que tiene más contenido en películas en Amazon Prime pero en series, la que supera es Netflix. Aquella que menos contenidos tiene en películas y series es Disney Plus.
- Y las películas y series de origen o producidas en Estados Unidos son las que más aparecen en las plataformas seguidas por la India y Reino Unido.
- En relación a las visualizaciones y filtrado para Argentina, se pudo observar que Amazon Prime tiene el mayor contenido de Argentina con un 40.6%, seguido por Netflix en un 37%. Además, en la clasificación del contenido, el que más aparece es aquel denominado como TV-MA tanto para series y películas, seguido de TV-14 en series y TV-Y (Programas dirigidos a un público muy joven, incluidos niños de 2 a 6 años).
- En cuanto a las búsquedas, se pudo observar que la cantidad de series que Argentina posee son 25 y películas 43. Mostrando a continuación el contenido en las tablas.
Se buscó el primer y último registro para luego hacer una comparación de la evolución del contenido. El primer registro data de 1985 perteneciente a la película "Esperando a la carroza" y el último registro del año 2021 perteneciente a la serie "El reino" (The Kingdom).
- El año 2020 fue el que tuvo mayor contenido en las plataformas para Argentina, teniendo una baja en 2021.

El análisis conjunto de las visualizaciones en las plataformas de streaming nos permitió obtener una visión global del contenido de las mismas.

Amazon Prime se posicionó como la plataforma con el mayor número de contenido y una amplia variedad de contenidos populares.

CONCLUSIONES:

En conclusión, el análisis de datos realizado utilizando el lenguaje de programación Python y diversas técnicas de análisis permitió obtener información relevante sobre las plataformas de streaming HBO, Netflix, Amazon Prime y Disney Plus.

El objetivo del proyecto era identificar tendencias de mercado, preferencias de contenido, oportunidades de expansión y segmentos de audiencia para respaldar la toma de decisiones estratégicas.

Se recopilaron los datos de estas plataformas y se llevó a cabo el procesamiento de los mismos, incluyendo la limpieza, la transformación y la unificación de los registros.

A través del análisis exploratorio de los datos, se obtuvo una comprensión inicial del conjunto de datos, identificando distribuciones, relaciones entre variables y estadísticas descriptivas.

Se aplicaron diversas técnicas de análisis, como regresión, clasificación y agrupamiento, según las preguntas de investigación planteadas.

Se utilizaron gráficos y visualizaciones para presentar los hallazgos del análisis de datos, destacando comparativas entre plataformas, la cantidad y distribución de contenido, así como los países de origen más frecuentes.

Amazon Prime se destacó como la plataforma con mayor cantidad de contenido y una amplia variedad de contenidos populares. Las series originales fueron un factor importante en la popularidad de todas las plataformas.

En el caso de Argentina, se observó que Amazon Prime y Netflix son las plataformas con mayor presencia de contenido, y se identificaron las clasificaciones más frecuentes para series y películas. También se realizó un análisis de la evolución del contenido a lo largo de los años, con una mayor oferta en 2020 y una disminución en 2021.

En resumen, este análisis de datos proporcionó información valiosa sobre las plataformas de streaming y sus contenidos, permitiendo identificar patrones, preferencias y oportunidades que pueden respaldar la toma de decisiones estratégicas en el mercado de streaming.