



Análisis Exploratorio de Datos (EDA) en Netflix

Miguel Angel Silva Uria



Contexto y Objetivos

Contexto:

- Netflix es una de las plataformas líderes en contenido digital.
- Se requiere comprender las tendencias de producción para mejorar la toma de decisiones.

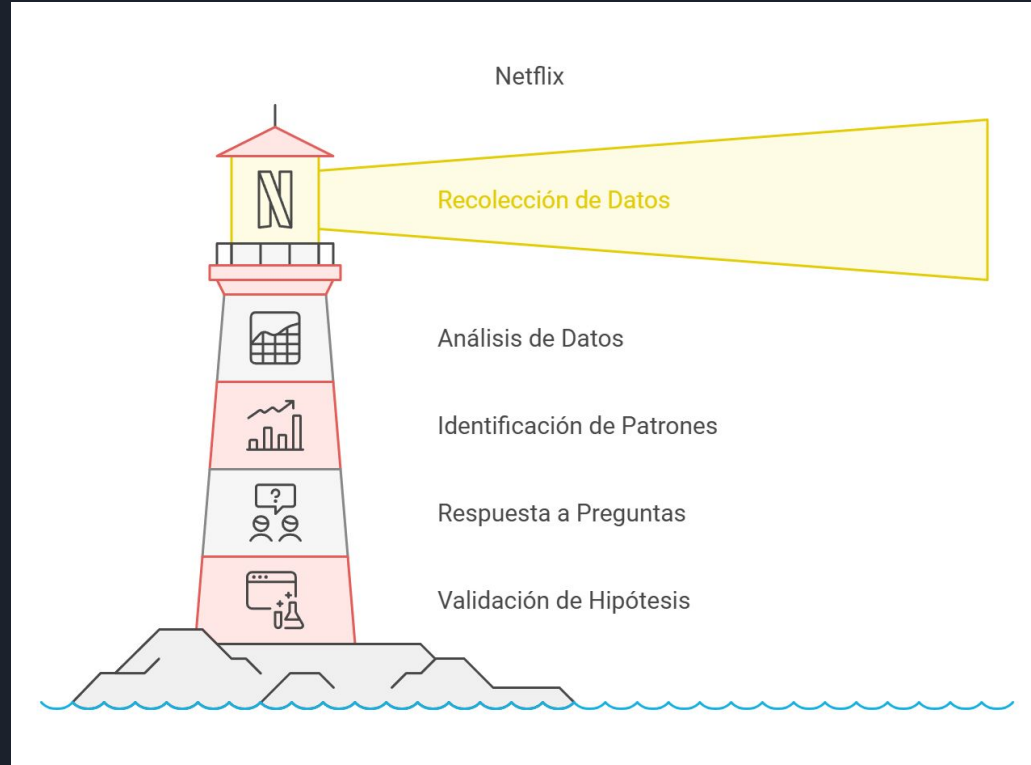
Objetivo del Análisis:

- Explorar y comparar patrones en las producciones de Netflix.

Hipótesis:

- Los géneros más populares en Movies son distintos de los de TV Shows.
- La mayoría de las producciones recientes (2010-2023) tienen clasificaciones orientadas al público joven (PG-13, TV-14).

Pasos del Análisis



Hipótesis y Preguntas

Hipótesis Principal:

- Los géneros más populares en películas son distintos de los géneros más populares en programas de TV.

Hipótesis Secundaria:

- La mayoría de las producciones recientes (2010-2023) tienen clasificaciones orientadas al público joven (PG-13, TV-14).

Preguntas a responder:

- ¿Qué géneros son más frecuentes en películas comparado con programas de TV?
- ¿Cómo ha variado la producción de contenido a lo largo de los años?
- ¿Qué clasificaciones por edades son más comunes y cómo se distribuyen por tipo?
- ¿Cuáles son los países que más producen contenido en Netflix?
- ¿Existe una diferencia notable en la duración promedio entre géneros?





Metodología

Datos:

- Fuente: Dataset de Netflix Movies and TV Shows.
- Variables clave: Type, Genre, Rating, Release Year, Duration, Country.

Herramientas:

- Python (pandas, numpy, seaborn, matplotlib).

Procesos:

- Limpieza y transformación de datos.
- Análisis Univariante, Bivariante y Multivariante.
- Visualización de resultados.
- Contraste de hipótesis y conclusiones.

Base Original Y Base Limpia

Title	Type	Genre	Release Year	Rating	Duration	Country
Title 1	TV Show	Comedy	1955	PG	3 Seasons	Japan
Title 2	TV Show	Horror	2020	G	3 Seasons	India
Title 3	TV Show	Action	1966	TV-PG	140 min	United States
Title 4	Movie	Thriller	2011	PG-13	3 Seasons	Canada
Title 5	TV Show	Romance	1959	TV-14	172 min	India
Title 6	Movie	Action	2007	PG-13	3 Seasons	Japan
Title 7	Movie	Romance	1977	TV-14	68 min	United States
Title 8	Movie	Comedy	1971	TV-PG	104 min	Japan
Title 9	Movie	Drama	2000	PG-13	2 Seasons	Canada
Title 10	Movie	Thriller	1975	TV-MA	1 Seasons	India

Title	Type	Genre	Release Year	Rating	Duration	Country
Title 1	TV Show	Comedy	1955	PG	1800	Japan
Title 2	TV Show	Horror	2020	G	1800	India
Title 3	TV Show	Action	1966	TV-PG	140	United States
Title 4	Movie	Thriller	2011	PG-13	141	Canada
Title 5	TV Show	Romance	1959	TV-14	172	India
Title 6	Movie	Action	2007	PG-13	141	Japan
Title 7	Movie	Romance	1977	TV-14	68	United States
Title 8	Movie	Comedy	1971	TV-PG	104	Japan
Title 9	Movie	Drama	2000	PG-13	141	Canada
Title 10	Movie	Thriller	1975	TV-MA	141	India

```
Title: 3000 valores únicos
Type: 2 valores únicos
Genre: 8 valores únicos
Release Year: 74 valores únicos
Rating: 7 valores únicos
Duration: 124 valores únicos
Country: 8 valores únicos
```

Valores nulos por columna:

```
Title      0
Type       0
Genre      0
Release Year 0
Rating     0
Duration   0
Country    0
dtype: int64
```

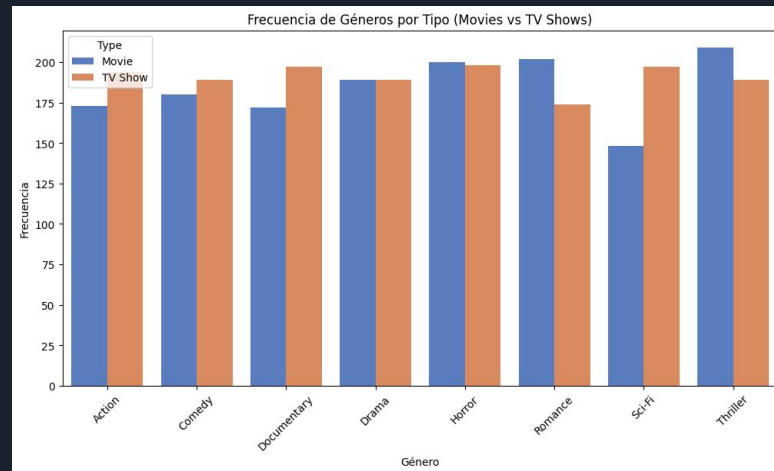
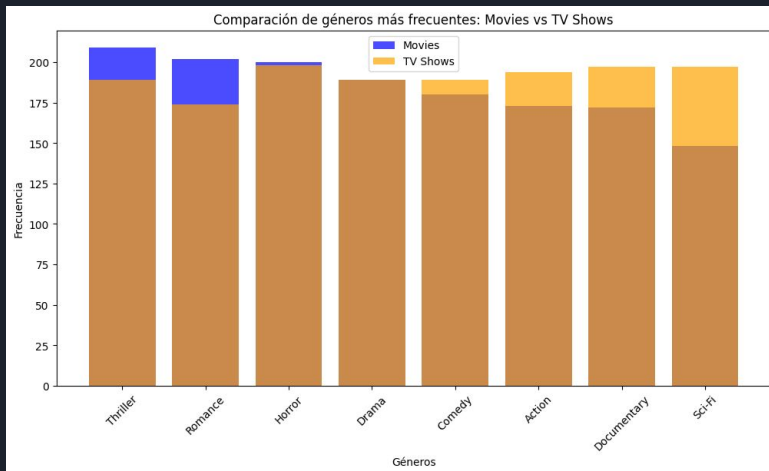
Primeras filas después de la conversión:

	Title	Type	Genre	Release Year	Rating	Duration	Country
0	Title 1	TV Show	Comedy	1955	PG	1800	Japan
1	Title 2	TV Show	Horror	2020	G	1800	India
2	Title 3	TV Show	Action	1966	TV-PG	140	United States
3	Title 4	Movie	Thriller	2011	PG-13	141	Canada
4	Title 5	TV Show	Romance	1959	TV-14	172	India

Resumen estadístico de la columna 'Duration':

```
count    3000.000000
mean      388.305333
std       520.405214
min        60.000000
25%       118.000000
50%       141.000000
75%       179.000000
max       1800.000000
Name: Duration, dtype: float64
```

Hipótesis Principal: Comparación de Géneros más Frecuentes entre Movies y TV Shows.

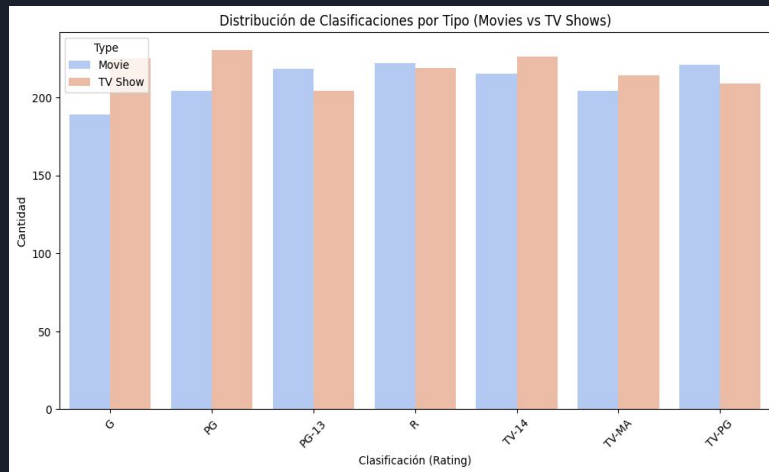
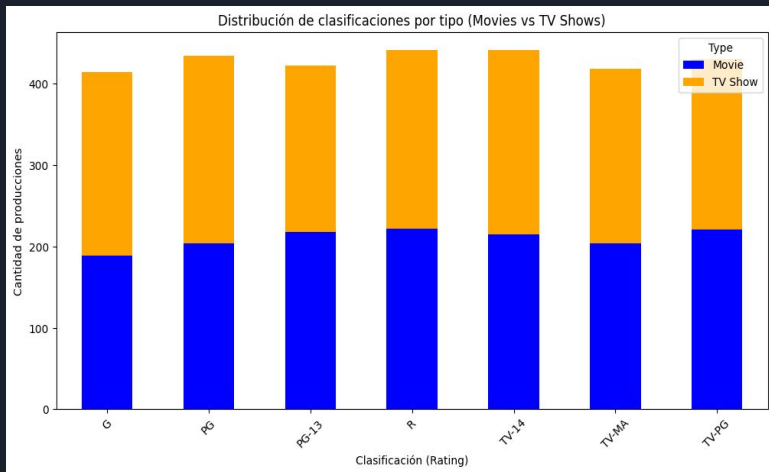


Insight Clave:

- Movies: Los géneros más frecuentes son Thriller y Romance.
- TV Shows: Los géneros más frecuentes son Sci-Fi y Horror.

Conclusión: Se confirma que los géneros más populares son distintos entre Movies y TV Shows.

Hipótesis Secundaria: Distribución de Clasificaciones en Producciones Recientes (2010-2023).

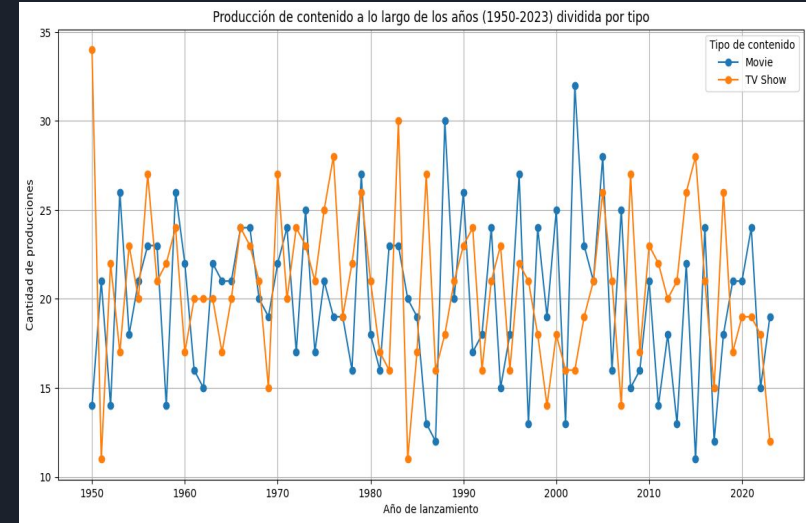
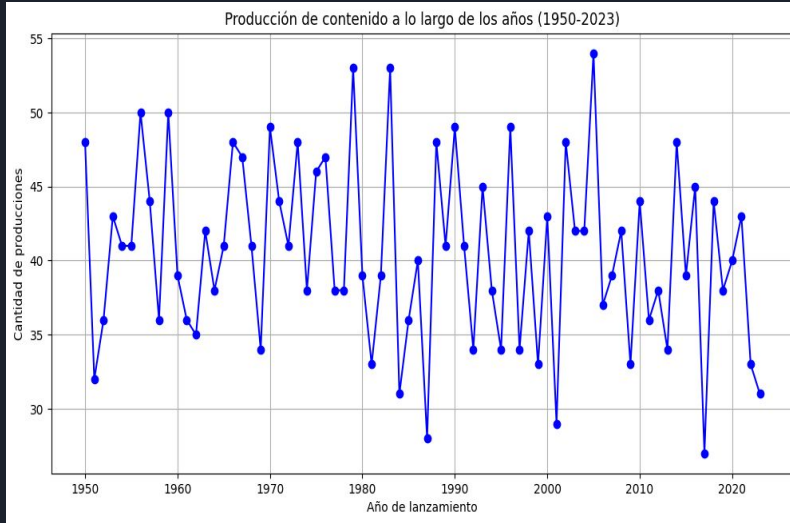


Insight Clave:

- Las clasificaciones TV-14, R, PG y TV-PG dominan en las producciones recientes.
- Movies tienen más títulos clasificados como PG-13 y R.
- TV Shows presentan una ligera inclinación hacia G, PG y TV-14.

Conclusión: La hipótesis secundaria se parcialmente confirma. Las clasificaciones orientadas al público joven (PG-13 y TV-14) son frecuentes, pero R también es común, mostrando una variedad de orientaciones de público.

Producción de Contenido a lo Largo del Tiempo

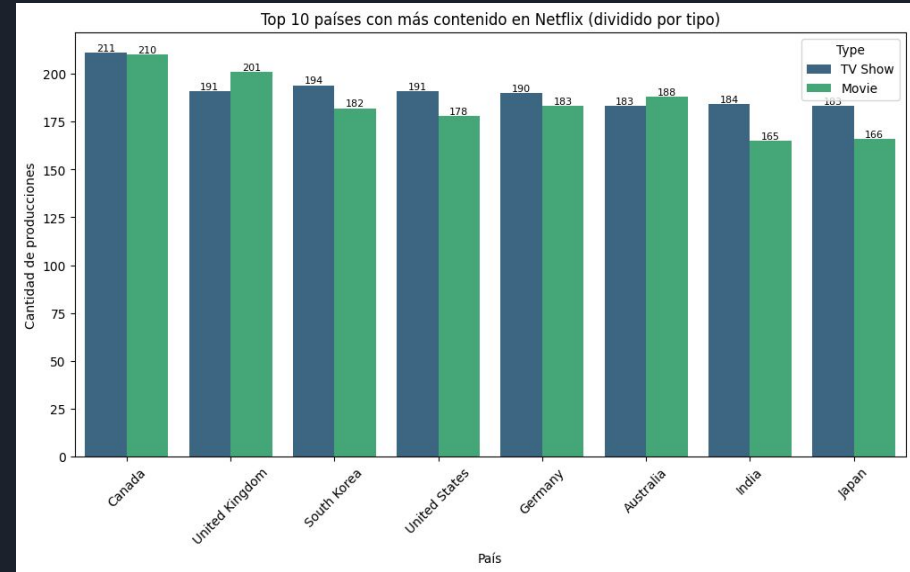
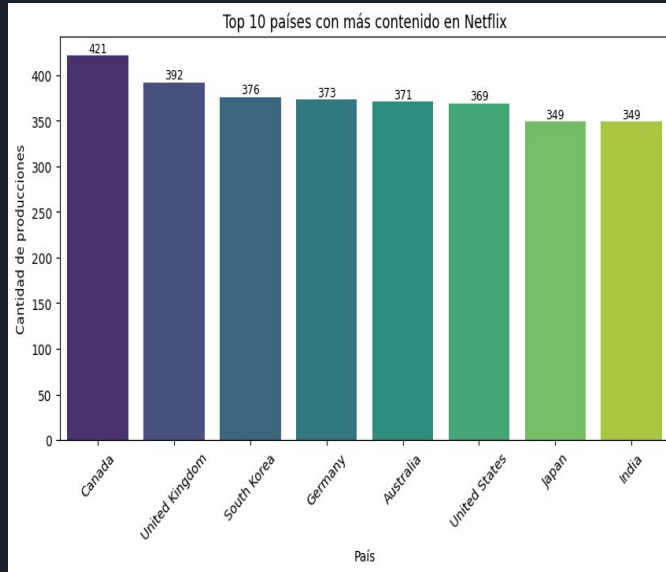


Insight Clave:

- La producción alcanzó picos notables en los años 2005, 1979 y 1983.
- Décadas como la de 1950 y 1980 también muestran alta producción constante.
- En años recientes (2015-2023), la producción no muestra picos equivalentes y tiende a disminuir.

Conclusión: La producción tiene picos en años específicos, como 2005, y luego muestra fluctuaciones. Luego sigue con baja producción.

Producción por País

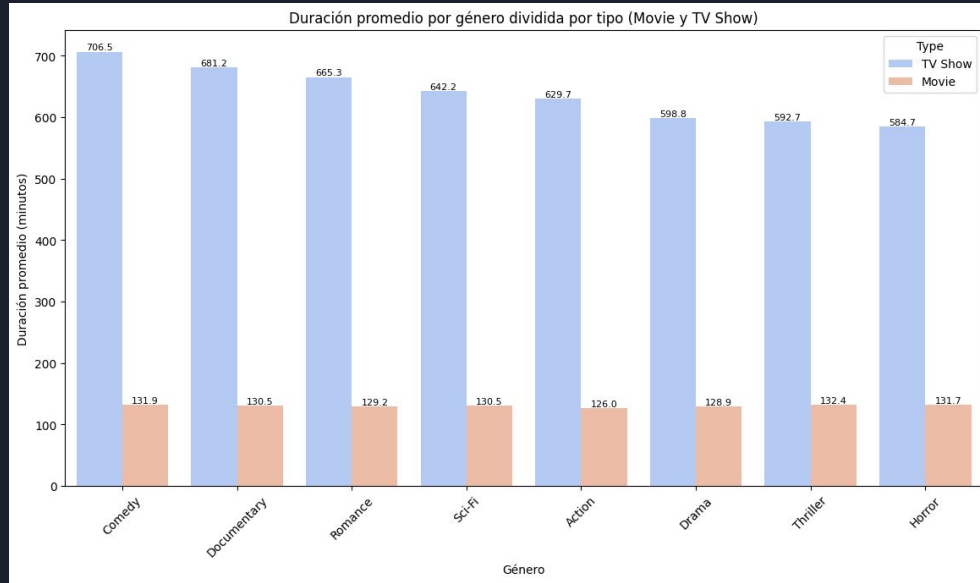


Insight Clave:

- Canadá lidera con 421 producciones.
- Le siguen Reino Unido (392) y Corea del Sur (376).
- Países como India y Japón muestran cifras similares (349 títulos cada uno).

Conclusión: Netflix tiene fuerte presencia de películas y shows en mercados anglosajones y asiáticos.

Duración Promedio por Género

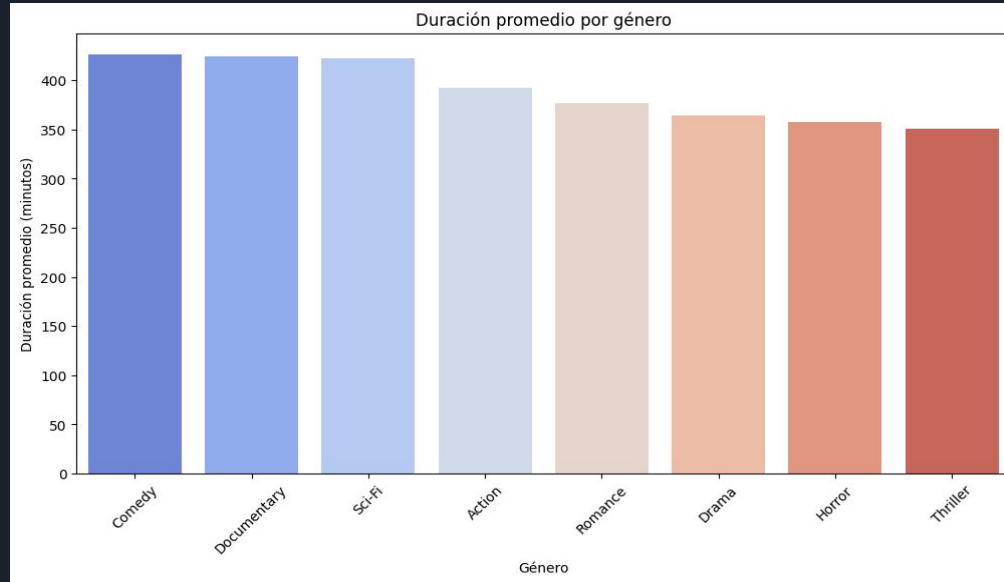


Insight Clave:

- Los TV Shows tienen una duración promedio significativamente mayor en todos los géneros (~600-700 minutos).
- Las Movies mantienen una duración estable alrededor de 130 minutos.

Conclusión: Existe una clara diferencia en la duración promedio entre Movies y TV Shows.

Duración Promedio por Género

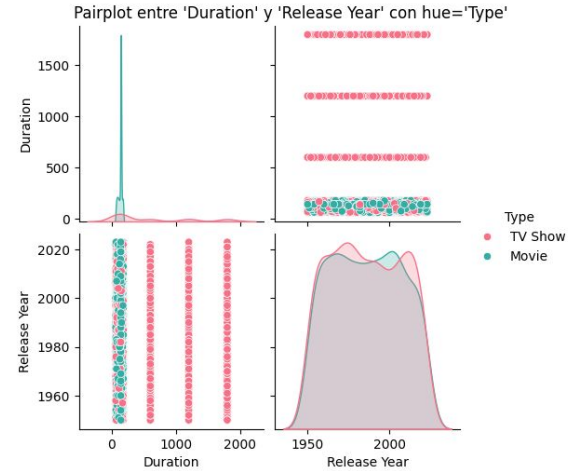
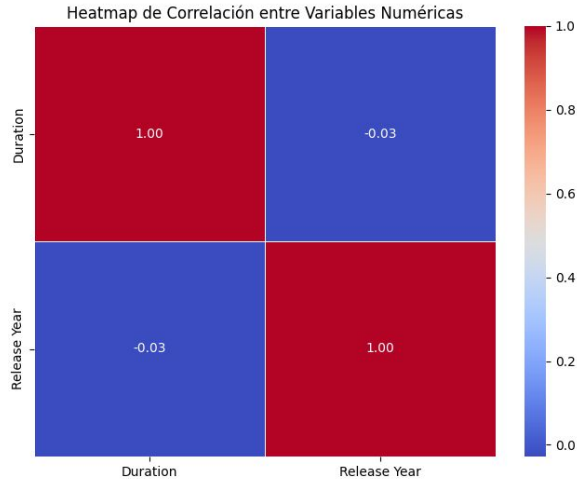


Insight Clave:

- Los géneros con mayor duración promedio son Comedy, Documentary y Sci-Fi, con duraciones superiores a 420 minutos.
- Los géneros Romance, Horror y Thriller tienen duraciones más bajas, cercanas a los 350 minutos.

Conclusión: Las producciones más largas corresponden a géneros que suelen tener múltiples episodios extensos (e.g., Documentaries, Sci-Fi).

Multivariante - Correlación y Pairplot



Heatmap de Correlación:

- La correlación entre Release Year y Duration es casi nula (valor cercano a -0.03).

Pairplot:

- La duración de los TV Shows tiende a ser mucho mayor que la de los Movies en todos los años.
- No hay un patrón claro de incremento de duración con el tiempo.

Conclusión: No existe una relación significativa entre el año de lanzamiento y la duración del contenido.

Resumen de Resultados y Contraste con Hipótesis

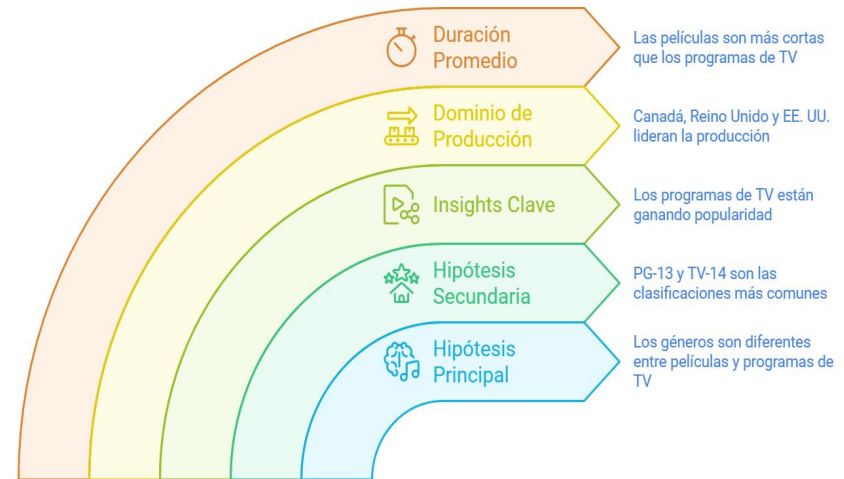
Resumen General:

- Hipótesis Principal: Confirmada – Los géneros más frecuentes son distintos entre Movies y TV Shows.
- Hipótesis Secundaria: Confirmada – PG-13 y TV-14 predominan en producciones recientes (2010-2023).

Insights Clave:

- TV Shows están ganando terreno en años recientes.
- La producción está dominada por Canadá, Reino Unido y Estados Unidos.
- Duración promedio → Movies (130 min), TV Shows (~600-700 min).

Resumen de Tendencias de Medios





Plan de Acción y Recomendaciones

Plan de Acción:

- Fomentar contenido orientado a géneros Sci-Fi y Horror para TV Shows.
- Enfocar en clasificaciones PG-13 y TV-14 para atraer al público joven.
- Explorar la colaboración con países líderes en producción (Canadá, Reino Unido, EE. UU.).

Productivización:

- Crear un dashboard interactivo para monitorear la evolución de contenido por tipo, género y país.

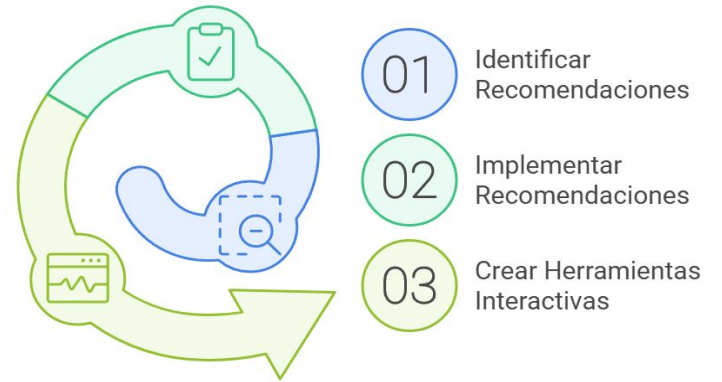
Comunicación:

- Compartir los insights con equipos de negocio y contenido.
- Realizar una presentación interna para alinear estrategias.

Comprensión de las Tendencias de Entretenimiento



Implementación de Estrategia de Contenido y Herramientas de Tendencia



Gracias

Miguel Angel Silva Uria

