

NOMBRE DE LA INSTITUCIÓN

PROYECTO DE ANÁLISIS DE DATOS

Análisis de datos de Filmaffinity dataset (Spanish)

Presentado por:

Laureano D. Rojas C.

Supervisor:

Carlos Cáceres

Fecha Inicio:

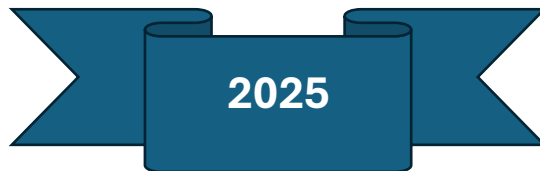
19-02-2025

Fecha Entrega:

28-02-2025

Descripción breve del proyecto:

Conjunto de datos que contiene información de 119.003 películas estrenadas en el periodo 1900-2020



✅ Problema a resolver:

Analizar qué país tiene más películas y cortometrajes, los cinco años en los que se estrenaron más películas e identificar cuál es el género con más películas. Este análisis nos va a permitir tomar una decisión de cual género es el que más o mejor promedio de calificación o Nota obtuvo y poder tomar la decisión de más o menos que película podemos filmar

❓ Preguntas clave:

- ✓ ¿Qué país tiene más filmes estrenados?
- ✓ Ver los cinco años en los que se estrenaron más películas.
- ✓ Identificar cual es el género con más filmes.
- ✓ Determinar cual es el género con mayor promedio de calificación o nota o valoración.
- ✓ Determinar cual tipo de filme es mejor.

⚡ Contexto y objetivo:

La razón por la cual se va a hacer el análisis es para determinar qué género de filme puede tener mejor aceptación de público y así hacer un filme que represente dicho género.

2. Recolección de Datos

📁 Fuentes de datos:

Los datos provienen de la web de Kaggle:

<https://www.kaggle.com/datasets/gan2gan/filmaffinity-dataset-spanish>

🌐 Descripción de los datos:

Este dataset está compuesto por una sola tabla la cual a su vez tiene las siguientes tablas:

1. Título: Define el título del filme.
 2. Año: Es el año en el que se estrenó el filme.
 3. País: Es de donde proviene el filme.
 4. Dirección: Es el nombre de quien está a cargo de dirigir el filme.
 5. Reparto: Son los Actores que participan en el rodaje.
 6. Nota: Son las valoraciones de los filmes.
 7. Tipo filme: Es el tipo al que pertenece el filme, como puede ser cortometraje, película, etc.
 8. Género: Son los géneros a los que pertenece cada filme, como puede ser Acción, Drama, etc.
-

3. Limpieza y Preparación de Datos



Procesos aplicados:

- Eliminación de duplicados ✓ / ✗
- Manejo de valores nulos (indicar estrategia: eliminar, imputar, etc.) ✓ / ✗
- Conversión de tipos de datos ✓ / ✗
- Normalización y transformaciones ✓ / ✗



Variables seleccionadas:

Las columnas que vamos a usar son:

- Título
- País
- Año
- Nota
- Tipo Filme
- Género

Vamos a usar título para obtener el nombre del filme.

Vamos a usar país para ver de donde proviene el filme.

Vamos a usar año para ver el año de estreno del filme.

Vamos a usar nota para que según esto podamos ver cual fue la calificación del filme.

Vamos a usar género porque necesitamos ver a qué género pertenecen los tipos de filme.

4. Exploración y Análisis Descriptivo

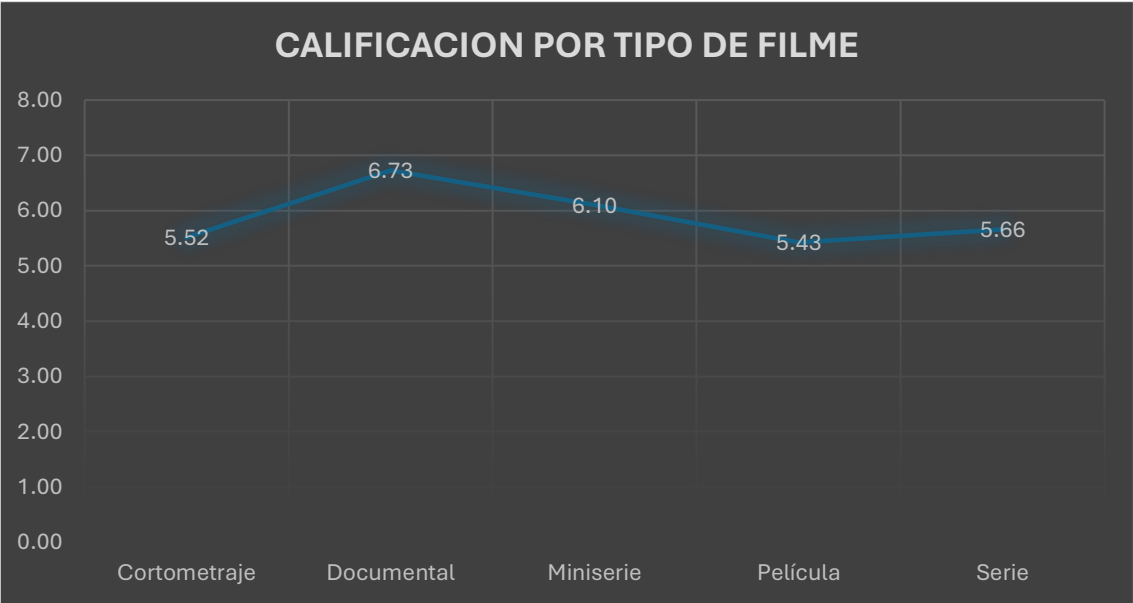
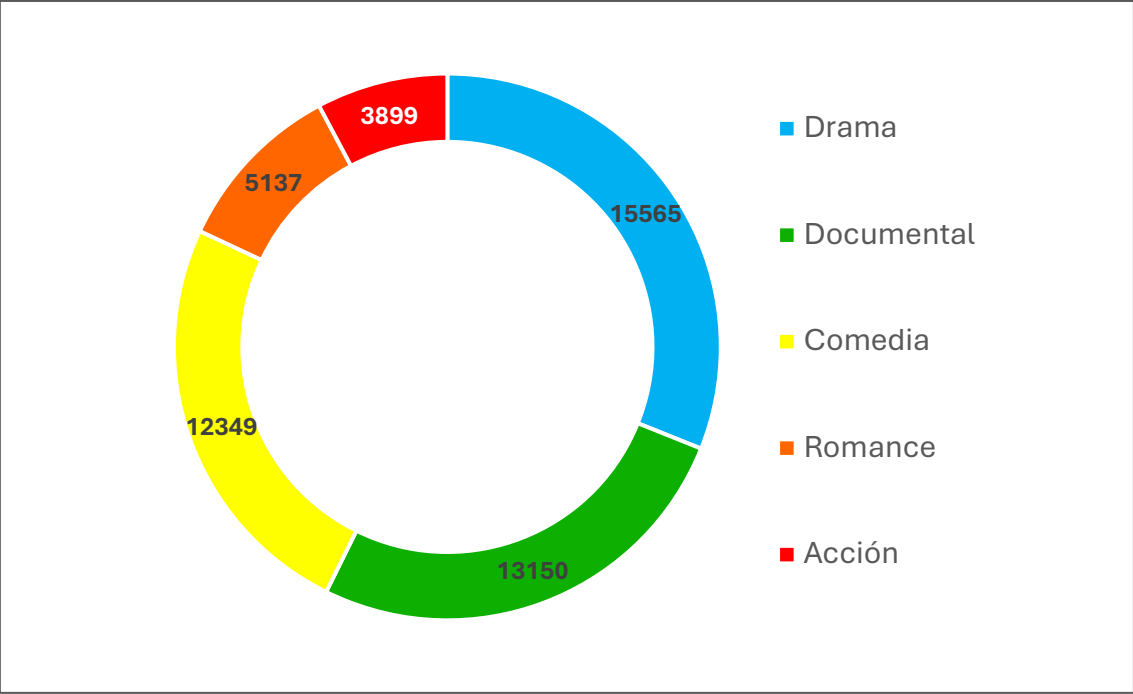


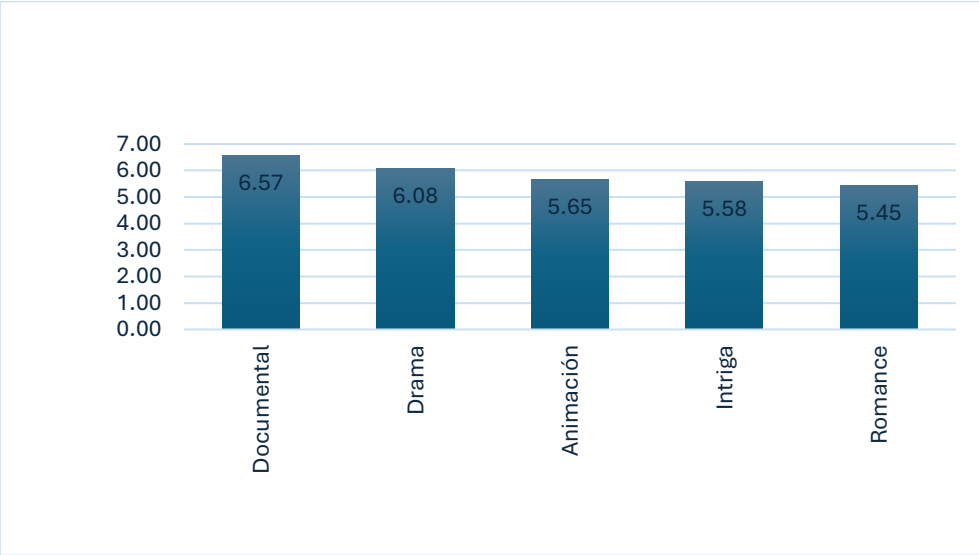
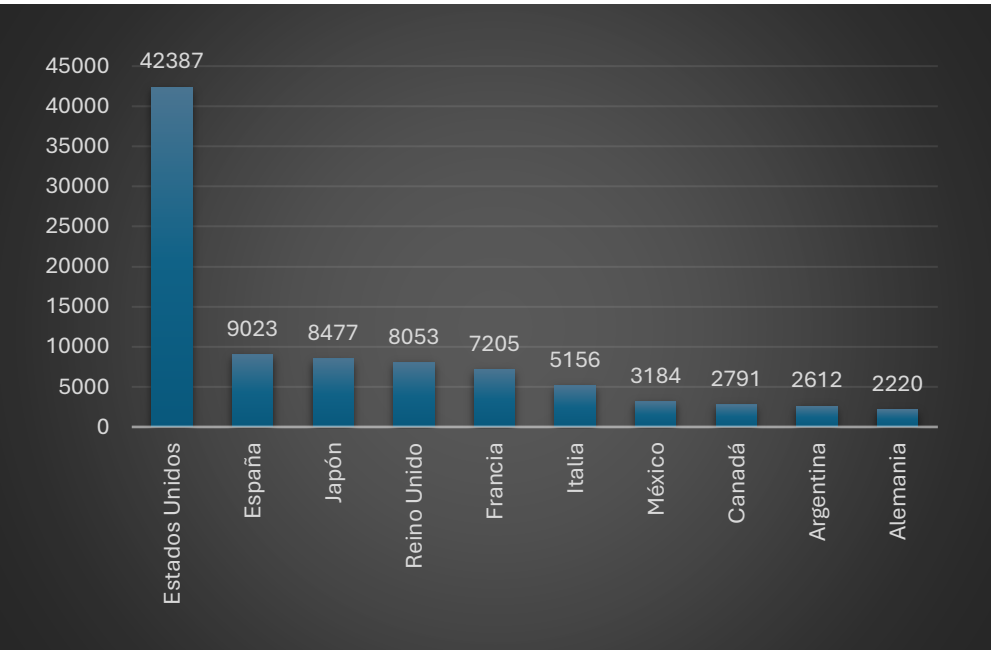
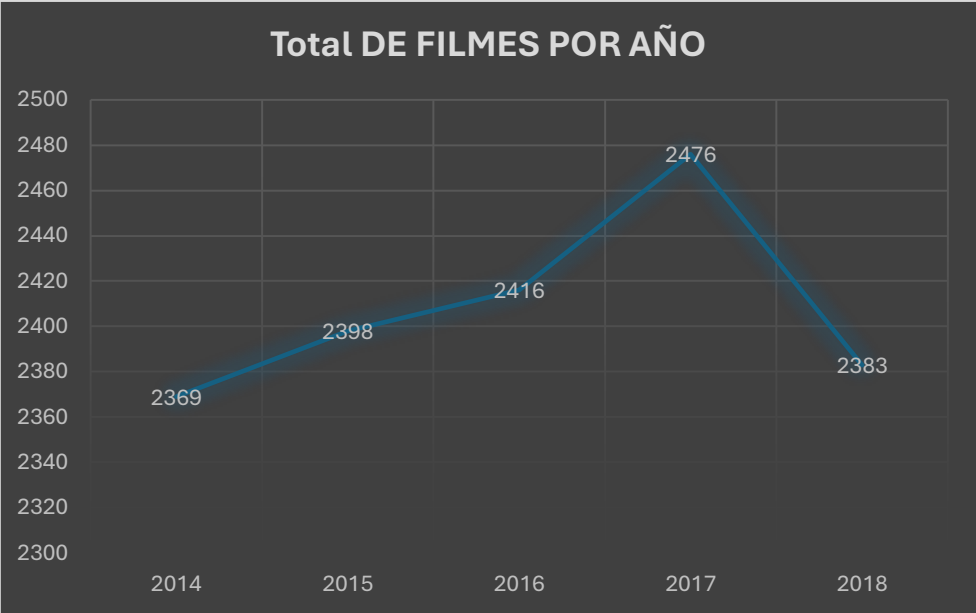
Estadísticas básicas:

- Promedio, mediana, moda
- Máximos y mínimos
- Distribuciones de los datos











Visualizaciones clave:





5. Modelado y Análisis Inferencial

Métodos utilizados:

- Pruebas de hipótesis  / 
- Regresión  / 
- Machine Learning (indicar técnicas si aplica)  / 
- Segmentación o clustering  / 

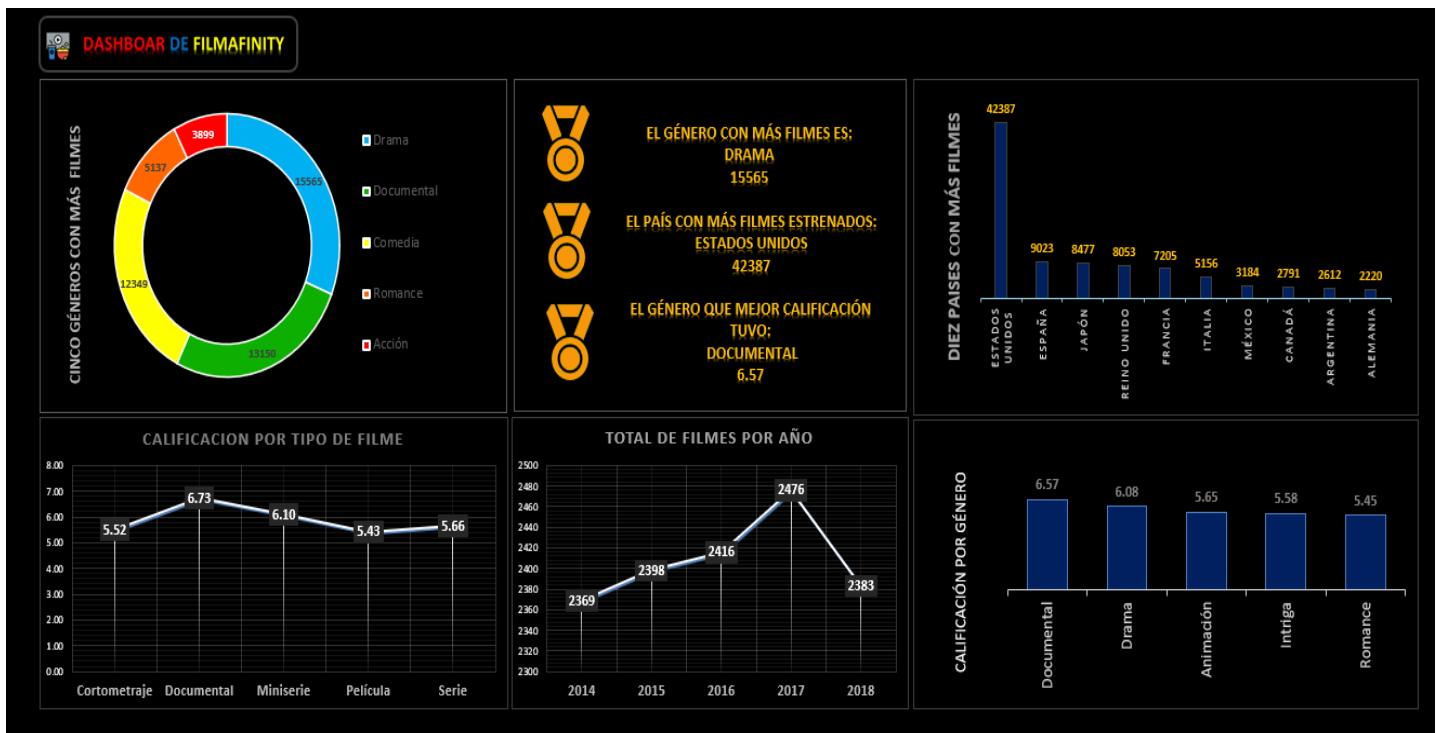
6. Interpretación y Comunicación de Resultados

Principales hallazgos:

Las conclusiones que sacamos con el análisis son las siguientes:

- El tipo de filme mejor calificado fue el Documental, con una puntuación promedio de 6.73.
- El país con más filmes estrenados es Estados Unidos, con un total de 42387 filmes estrenados.
- Los cinco años en los que se estrenaron más filmes fueron, 2018, 2017, 2016, 2015 y 2014.
- El género con más filmes fue

Visualizaciones finales:



7. Toma de Decisiones y Acción

Recomendaciones basadas en datos:

Según el análisis hecho basado en los datos, podemos concluir que, un Documental sería una buena elección ya que este recibió mejor promedio de puntuación en cuanto a tipos de filme que debemos hacer

Impacto esperado:

El impacto que esperamos es que al momento de estrenar un filme del tipo documental, tengamos la misma aceptación que con los filmes del tipo Documental de los datos analizados.


8. Monitoreo y Mejora Continua

Métricas de seguimiento:

Debemos estrenar el filme y obtener los resultados o el impacto esperado

Actualizaciones futuras:

Debemos de recolectar los datos obtenidos después de estrenar el filme para así ir actualizando nuestros datos e ideass.

 **Este documento te ayudará a estructurar tu proyecto paso a paso. Completa cada sección con la información de tu análisis y avanza de manera organizada.**