

Proyecto de Modelado Dimensional: Análisis de Películas IMDb

Kenny Berrones

IES Pere Maria Orts

20 de noviembre de 2024

Introducción al Caso de Uso

- IMDb (Internet Movie Database) es una de las bases de datos más completas y populares sobre películas, series y programas de televisión. Contiene información detallada sobre títulos, actores, equipos de producción, calificaciones y mucho más.



Preguntas de Negocio o KPIs

Estas son tres preguntas importantes que el modelo ayudará a responder:

- ① **¿Qué géneros de películas son los más rentables por año?**
 - Analizar la ganancia de cada género a lo largo de los años.
- ② **¿Cuál es el promedio de calificación y votos por país de producción?**
 - Identificar países que producen las películas mejor valoradas y más votadas.
- ③ **¿Qué tendencias de popularidad tienen las películas de cierto genero?**
 - Ver cómo la popularidad de este género ha cambiado con el tiempo.

Descripción de las Fuentes de Datos Encontradas

- **Dataset 1: Top 5000 películas**

- Fuente: Base de datos pública de IMDb (encontrada en Kaggle).
- Tamaño: 5000 filas y 20 columnas.

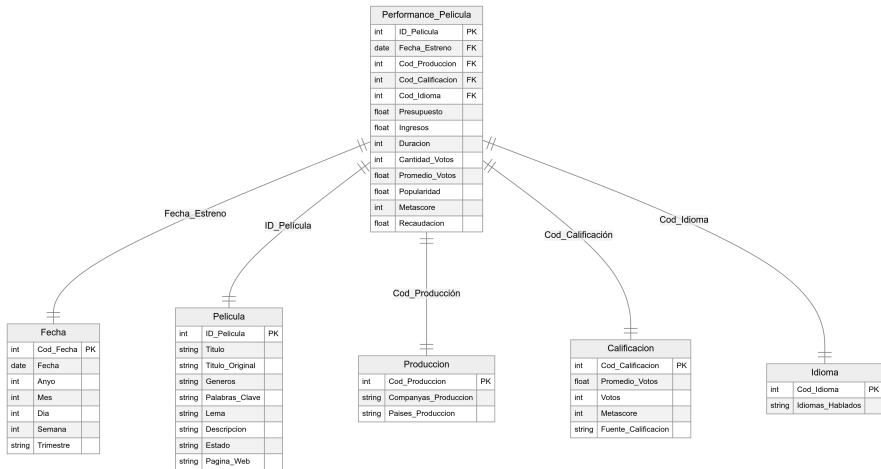
- **Dataset 2: Top 1000 películas**

- Fuente: Base de datos pública de IMDb (encontrada en Kaggle).
- Tamaño: 1000 filas y 8 columnas.

Esquema Conceptual: Tabla de Hechos y Dimensiones

- La **Tabla de Hechos** almacena las métricas clave relacionadas con las películas, como ingresos, presupuesto, duración y popularidad.
- Las **Dimensiones** ayudan a contextualizar estas métricas:
 - ① **Fecha**: Incluye año, mes, día y trimestre.
 - ② **Película**: Contiene título, géneros, palabras clave y descripción.
 - ③ **Producción**: Información sobre compañías productoras y países involucrados.
 - ④ **Calificación**: Datos de calificaciones, votos y metascor.
 - ⑤ **Idioma**: Idiomas hablados en cada película.
 - ⑥ **Popularidad**: Tendencias y rangos de popularidad.

Esquema Dimensional



Posible Ampliación: Dimensión de Actores

- Una mejora futura sería agregar una dimensión dedicada a los actores, que podría incluir:
 - Nombre, género y edad de los actores.
 - Número de películas en las que han participado.
 - Reconocimientos como premios y nominaciones.
- Esto permitiría responder preguntas como:
 - ¿Qué actores son los más rentables o populares?
 - ¿Cómo afecta el elenco al rendimiento financiero y la popularidad de una película?

Conclusión

- Este proyecto proporciona un modelo eficiente para analizar grandes volúmenes de datos de películas de IMDb.
- Responde preguntas de negocio clave que van más allá de las capacidades de los sistemas tradicionales OLTP.
- Sus aplicaciones incluyen análisis de tendencias, éxito financiero y calificaciones de películas.

¡Gracias por su atención! Estoy abierto a preguntas.