**BDD – Películas**

**Andrés Cortés – Juan Sebastián Carvajal**

**Modelo relacional**

**Proceso de carga de datos:**

Para la carga masiva de datos utilizamos un código en Python el cual se conecta a una base de datos PostgreSQL, crea una tabla llamada "Pelicula" con ciertas columnas y luego inserta datos desde en archivo de Excel en dicha tabla. De esta forma:

1. Se establece una conexión con una base de datos PostgreSQL que se encuentra en 'LocalHost', utilizando el usuario 'postgres' y la base de datos 'postgres'.

2. Se crea una tabla llamada "Pelicula" con varias columnas como `Id`, `Titulo`, `Fecha\_lanzamiento`, etc. La columna `Id` se define como clave primaria.

3. Se carga un archivo de Excel ubicado en '/home/jsdinge/Proyecto ing datos/Datos\_peliculas (1).xlsx' utilizando la biblioteca Pandas.

4. Itera sobre las filas del conjunto de datos, formatea la fecha y el título, y luego ejecuta consultas SQL de inserción para cada fila en la tabla "Pelicula" creada anteriormente.

5. Se incluye un bloque para manejar excepciones. Si ocurre alguna excepción durante la ejecución del bloque `try`, se imprime la excepción.

6. Se realiza un `commit` para confirmar las inserciones en la base de datos y se cierran el cursor y la conexión a la base de datos.

Este fue el paso a paso tomando de ejemplo la creación de la tabla “Pelicula”, proceso el cual mantiene la misma estructura para la creación de las demás tablas, pero las variables a tener en cuenta son las que cambian.

**Escenarios a evaluar:**

1. Analizar la cantidad de películas estrenadas por género.

Para analizar la cantidad de películas estrenadas por género, se utilizará la tabla "Genero". Esta tabla relaciona el ID de cada película con su género correspondiente. El modelo de datos sugiere que una película puede tener varios géneros, ya que se utiliza una tabla independiente para almacenar esta información.

Esta consulta realiza una unión entre las tablas "Genero" y "Pelicula" basándose en el ID de la película, y luego cuenta la cantidad de películas para cada género. El resultado proporcionará un recuento de películas para cada género presente en la base de datos.

1. Cuales son los lenguajes hablados mas comunes en la películas lanzadas.

La tabla "lenguajes\_hablados" está diseñada para almacenar los lenguajes hablados asociados a cada película, utilizando el ID de la película como clave foránea para establecer la relación con la tabla "info\_especial", que contiene información adicional sobre cada película.

Esta consulta agrupa los registros por lenguaje y cuenta cuántas veces aparece cada lenguaje en la tabla. Luego, ordena los resultados en orden descendente según la cantidad, proporcionando así los lenguajes hablados más comunes en las películas lanzadas. Este enfoque asume que cada registro en "lenguajes\_hablados" representa un lenguaje hablado en una película específica.

1. Cuantas peliculas son consideradas para mayores de edad (+18).

La columna "adulto" de tipo booleano en la tabla "Pelicula" se utiliza para indicar si la película está clasificada como para mayores de edad. Por lo tanto, para obtener la cantidad de películas consideradas para mayores de edad, se puede realizar una consulta SQL simple que cuente los registros en los que la columna "adulto" es verdadera.

Esta consulta cuenta la cantidad de películas en la tabla "Pelicula" donde la columna "adulto" tiene el valor de true, lo que indica que son películas consideradas para mayores de edad. El resultado proporcionará la cantidad total de películas que cumplen con este criterio en la base de datos.

1. Consultar cuantas peliculas fueron estrenadas por año.

En el escenario proporcionado, se han creado varias tablas en una base de datos mediante SQL. Para consultar cuántas películas fueron estrenadas por año, se utilizará la tabla "Pelicula" y su columna "Fecha\_lanzamiento", que almacena la fecha de lanzamiento de cada película.

Esta consulta utiliza la función EXTRACT para obtener el año de la fecha de lanzamiento de cada película. Luego, agrupa los resultados por año y cuenta la cantidad de películas en cada año. El resultado proporcionará un desglose del número de películas estrenadas por año en la base de datos.

1. Cual genero tiene la mayor popularidad.

En el escenario proporcionado, se han creado varias tablas en una base de datos mediante SQL. Para determinar qué género tiene la mayor popularidad, se utilizarán las tablas "Genero" y "Op\_publica", que contienen información sobre el género de cada película y la popularidad asociada a cada película, respectivamente.

Esta consulta realiza una unión entre las tablas "Genero" y "Op\_publica" utilizando el ID de la película como clave foránea. Luego, suma la popularidad para cada género y ordena los resultados en orden descendente según la popularidad total.

**URL de repositorio:**