Proyecto SQL: Películas y directores de cine

Objetivos del análisis:

23

24

ADD CONSTRAINT FK_Peliculas_Directores

FOREIGN KEY (ID_Director) 25 REFERENCES Directores(ID_Director);

El objetivo del análisis es descubrir tendencias en las películas, como cuáles son los géneros predominantes en el cine, tanto en cantidad como en calidad (popularidad y calificaciones promedio), analizar los presupuestos y las ganancias para ver las rentabilidades, y por otro lado, analizar a los/las directores/as en busca de descubrir cuáles son los más exitosos en lo que hacen, su tendencia a realizar cierto género y la regularidad con la que realizan películas.

El conjunto de datos utilizado es de uso público:

https://www.kaggle.com/datasets/nayanack/movies-and-directors-dataset-forfilm-analytics.

Previo al análisis realicé una limpieza y estandarización en Excel, así como la búsqueda de información faltante en algunos registros sobre el sexo de algunos/as directores, y también agregué una columna con el género principal de cada película a los efectos de ampliar en análisis.

Una vez cargados los datos en SQL Server procedí a su normalización, mediante la creación de claves primarias y foráneas, a modo de poder realizar consultas de tipo JOIN para un mejor análisis:

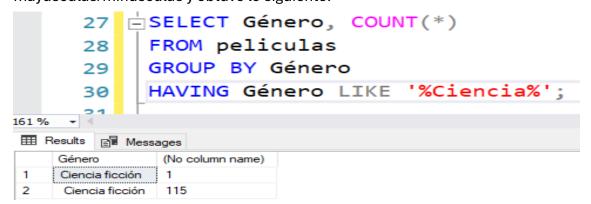
```
1 ⊡--Primero modifico la columna "ID_Director" de la tabla Directores para que no permita valores nulos,
 --para luego transformarla en primary key:
3 — ALTER TABLE Directores
4 | ALTER COLUMN ID_Director INT NOT NULL;
5
   -- Ahora transformo la columna "ID_Director" en primary key:
6
7 ALTER TABLE Directores
8  ADD CONSTRAINT PK_Director PRIMARY KEY (ID_Director);
9
10 🚊--Ahora realizo lo mismo en la tabla "Películas" con la columna "ID_Peliculas" para transformarla en primary key,
   --y también modifico "ID_Director" como clave foránea, para terminar de normalizar la base de datos.
11
ALTER COLUMN ID_Pelicula INT NOT NULL;
13
14
16
   ADD CONSTRAINT PK_Pelicula PRIMARY KEY (ID_Pelicula);
17
18 🗀 ALTER TABLE Peliculas
19
   ALTER COLUMN ID_Director INT;
20
21
   -- Creo la Foreign Key:
```

Una vez normalizada la base de datos, procedí al análisis:

Al comenzar me di cuenta de una inconsistencia en los datos, y es que, al seleccionar todas las películas de un determinado género, no obtenía todos los resultados:

```
28 SELECT * FROM peliculas
         29
                 WHERE Género = 'Ciencia ficción';
         30
177 %
Results Ressages
     ID_Pelicula
               Presupuesto
                          Puntuacion_Popularidad
                                            Ingresos
                                                         Titulo
                                                                           Género
                                                                                       Calificacion_Promedio
                                                                                                        ID Director
                                                                                                                  Año_Estreno
                                                                                                                             Mes_Estreno
    46564
               10500000.00 56
                                             792910554,00 E.T. the Extra-Terrestrial Ciencia ficción
                                                                                       73
                                                                                                        4799
                                                                                                                  1982
                                                                                                                             Abril
```

Utilicé una sentencia con la cláusula "LIKE" para que me mostrara todas las variaciones de Ciencia ficción que podrían estar almacenadas en la base de datos, incluyendo aquellas con espacios adicionales o diferencias en las mayúsculas/minúsculas y obtuve lo siguiente:

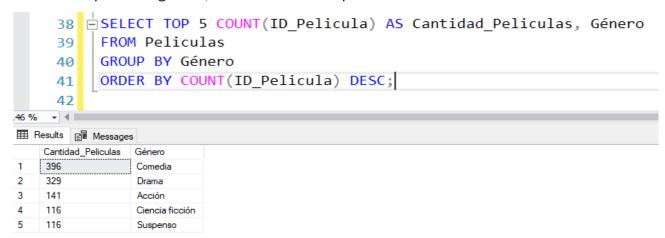


En conclusión, tenía valores con espacios adicionales, lo que provocaba problemas a la hora de realizar consultas, por lo que actualicé los valores de todos los géneros mediante las funciones "LTRIM" Y "RTRIM", las cuales recortan los espacios en blanco a la izquierda y a la derecha de la cadena de caracteres:

```
29 SET Género = LTRIM(RTRIM(Género));
31
32
Messages
(1465 rows affected)

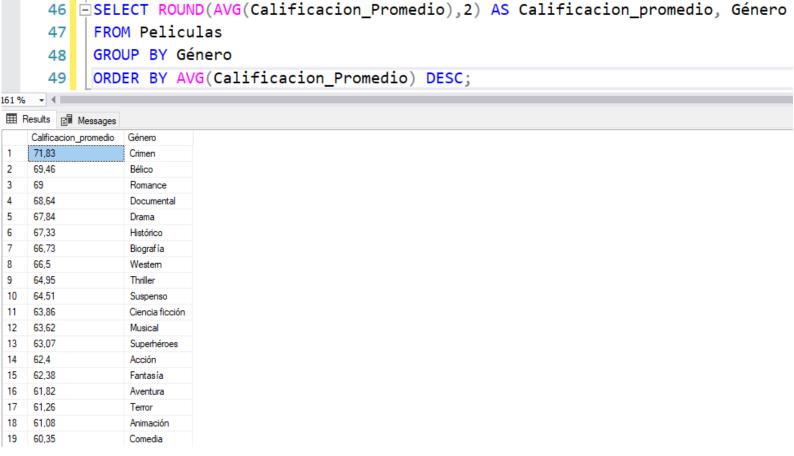
Completion time: 2024-10-21T14:46:52.2530872-03:00
```

Siguiendo con el análisis, cree una consulta que me mostrara cuántas películas había por cada género, seleccionando las primeras cinco con más cantidad:



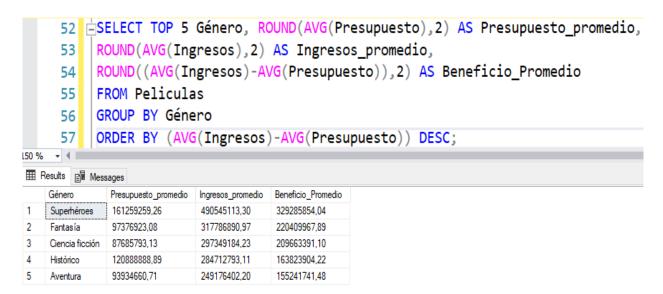
Podemos observar que las películas de comedia predominan en cantidad, seguidas por drama, acción, ciencia ficción y suspenso.

Seleccioné los géneros y la calificación promedio por cada uno, para ver si coinciden los géneros más realizados con las mejores calificaciones y ver si puede ser un factor por el cual se eligen producir los mismos:



En principio se puede ver que no son los géneros más realizados los que tienen mejor calificación promedio.

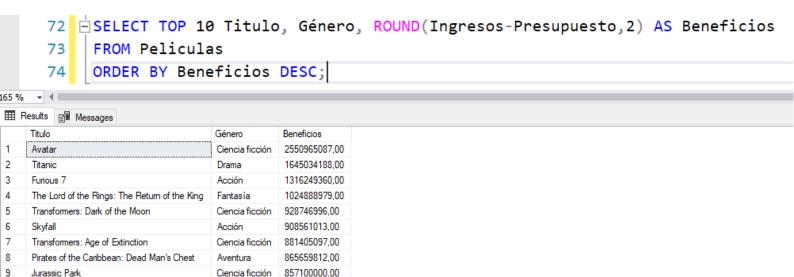
Por último, para terminar con el análisis de los géneros, cree una consulta que me permitió ver el presupuesto y las ganancias promedio, para luego analizar el beneficio y descubrir cuáles son los cinco géneros con más retorno:



En este caso vemos un resultado que coincide con lo que sucede en la actualidad: la tendencia de las películas de superhéroes, que tienen mucho éxito en taquilla.

Pasando de lo macro a lo micro, ahora toca analizar un poco las películas:

Para comenzar realicé una consulta simple para visualizar las diez películas que más beneficio obtuvieron:



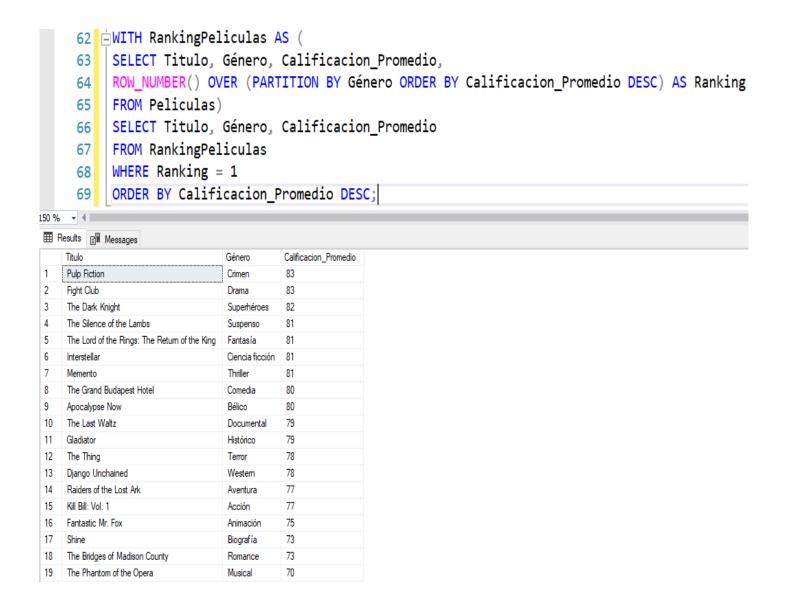
10

Harry Potter and the Philosopher's Stone

Fantasía

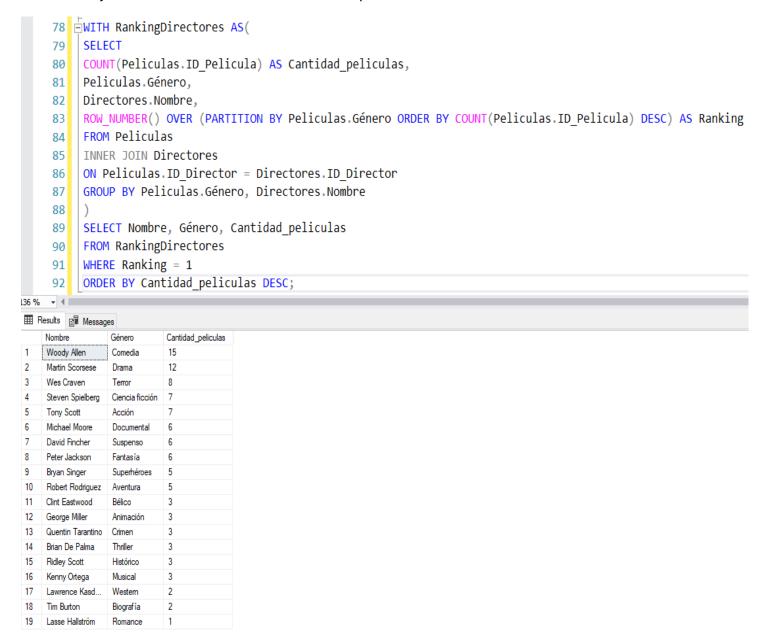
851475550.00

También cree una expresión de tabla común en la que utilicé la función "ROW_NUMBER" para asignar un numero de fila a cada género, ordenando las películas por la calificación promedio de mayor a menor, lo que le da a cada película dentro de los géneros un ranking basado en su calificación. Luego basé la consulta principal en esta tabla, seleccionando solo las películas con el ranking 1, es decir, la película con la calificación más alta de cada género:



Siguiendo con la última parte del análisis, los directores:

Comencé por investigar cuáles son los directores que más películas realizaron en cada género, para eso volví a utilizar una expresión de tabla común, con un inner join dentro a modo de unir la tabla de películas con la de directores:



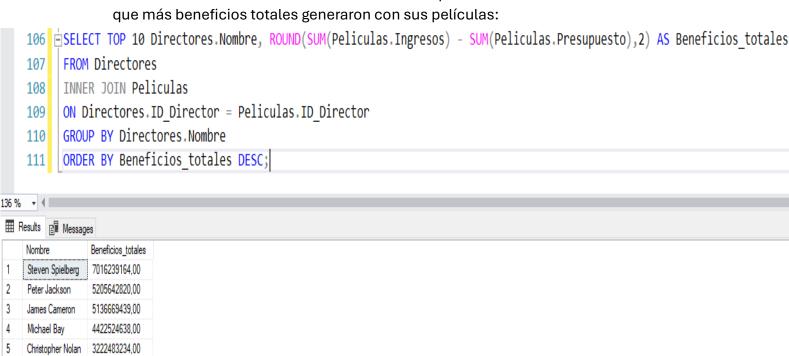
Mediante estos resultados podemos observar las preferencias de los directores a la hora de realizar ciertos géneros.

Siguiendo con los directores cree una subconsulta que me devuelve la calificación promedio de las películas, y la utilicé para filtrar y obtener aquellos directores que tienen una calificación promedio por encima de la media.

En la imagen muestro los primeros 10 resultados para no utilizar toda la hoja:

```
96 SELECT Directores Nombre, ROUND(AVG(Peliculas Calificacion Promedio), 2) AS Calificacion Promedio
      97
            FROM Directores
           INNER JOIN Peliculas
      98
           ON Directores.ID Director = Peliculas.ID Director
      99
            GROUP BY Directores Nombre
     100
           HAVING AVG(Peliculas.Calificacion promedio) > (SELECT AVG(Calificacion Promedio)
     101
            FROM Peliculas)
    102
           ORDER BY Calificacion Promedio DESC;
     103
136 %
Results Messages
    Nombre
                      Calificacion_Promedio
    Christopher Nolan
    Quentin Tarantino
                      77,75
3
    David Lynch
                      74.4
                      74.14
    Wes Anderson
    David Fincher
                      73,4
                      73,33
    Peter Jackson
    James Cameron
                      73,29
    Martin Scorsese
                      73
    Alejandro González Iñámitu 72,33
9
    Paul Thomas Anderson
                       72,17
```

Para finalizar con el análisis cree una consulta que me devuelve a los 10 directores que más beneficios totales generaron con sus películas:



5

8

Chris Columbus

Tim Burton

Sam Raimi

3098631503,00

2340418241,00

2208951462,00

Robert Zemeckis 2600622002,00

Francis Lawrence 2349457182.00