Prueba Analytics Expert

En esta prueba contará con dos tablas, una llamada credits y otra titles, sobre series y películas. Estas tablas se relacionan bajo la columna id de ambas.

1. SQL (30%)

En este punto debe escribir la consulta que generé el requerimiento del literal

* 1. Muestre el top 10 de los directores con más películas
  2. Muestre el top 5 del promedio de imdb\_score por actores de shows
  3. En la columna production\_countries algunos registros tienen varias iniciales de varios países, genere una consulta en la que traiga esa columna únicamente con las iniciales del primer país sin corchetes ni comas y que venga acompañado de la columna type y el número de títulos.
  4. Muestre todas las películas de Steven Spielberg como director, con el nombre de la película, release\_year, imdb\_score y tmdb\_score
  5. Muestre el título de las series más antiguas según columna random\_date, junto con random\_date y la columna random date sumándole 5 años y 3 meses en un formato como este ejemplo “Wednesday, March, 2021”

1. Power BI (50%)

Le han encargado generar el mejor dashboard posible para visualizar la información de los títulos en un aplicativo, este debe lograr mostrar de manera efectiva lo siguiente (ignorar random\_date):

* 1. El número de títulos por año y por tipo
  2. Los géneros y tipos de títulos con más votos y mejores scores
  3. Los directores y actores con más votos y mejores scores
  4. Los actores por género con más títulos
  5. Los directores y el tiempo de duración de los títulos

Tenga en cuenta:

* Se evaluará el aspecto general del dashboard
* Para las transformaciones de datos sólo se debe usar Power Query y DAX, no usar R ni Python

1. Python (20%)

En un notebook de Python generar los siguientes puntos:

* 1. Exploración de los datos
  2. Inner Join entre las tablas
  3. Limpieza de los datos
  4. Generar un modelo de regresión lineal que permita predecir el imdb\_score de los títulos
  5. Evaluar el modelo

Deben entregar en un zip los siguientes archivos:  
El primer punto un .txt con las queries, en el segundo un archivo .pbix y en el tercero un notebook .ipynb