**Análisis de Base de Datos**

**Objetivo:**  
Cada grupo deberá seleccionar una base de datos, realizar un análisis exploratorio básico, comprender su estructura y contenido, y presentar un informe con los hallazgos más relevantes.

Leidy Agudelo, Jenny Soto, Carlos Valencia

Cada grupo debe realizar las siguientes tareas con la base de datos seleccionada:

1. **Selección de la Base de Datos**

* Nombre de la base de datos: Netflix life impact
* Fuente (URL, repositorio, etc.): https://www.kaggle.com/datasets/towhid121/netflix-life-impact-dataset-nlid
* Formato del archivo (.csv, .xlsx.): csv

1. **Estructura de la Base de Datos**

* Número total de filas (registros): 82
* Número total de columnas (atributos o variables): 10

1. **Descripción de las Columnas**

Por cada columna se debe especificar:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Nombre de la columna** | **Tipo de dato** | **Descripción** | **Cantidad de datos** | **Valores únicos (si aplica)** |
| Movie Title | String | Nombre de la película | 82 |  |
| Genre | String | Genero | 82 |  |
| Release Year | Float | Año de lanzamiento | 82 |  |
| Average Rating | Int | Calificación promedio | 82 |  |
| Number of Reviews | Float | Numero de vistas | 82 |  |
| Review Highlights | String | Reseñas destacadas | 82 |  |
| Minute of Life-Changing Insight | String | Un minuto de revelación que cambia la vida | 82 |  |
| How Discovered | String | Como la descubriste | 82 |  |
| Meaningful Advice Taken | String | Momento significativo | 82 |  |
| Suggested to Friends/Family (Y/N %) | String | Sugerido a amigos y familiares | 82 |  |

1. **Calidad de los Datos**

* ¿Hay valores faltantes en alguna columna? ¿Cuántos? R/ No

¿Hay columnas con valores repetidos o inconsistentes? R/ Si, en la columna “Minute of Life-Changing Insight” no se entiende el texto “â€“”

* ¿Existen datos fuera de rango o errores evidentes? Si, â€“
* ¿Conoce la última fecha de actualización de los datos? 08-04-2025

1. **Análisis Cuantitativo Básico**

Para columnas numéricas:

* Mínimo, máximo, media, mediana, desviación estándar.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Titulo** | **Mínimo** | **Máximo** | **Media** | **Mediana** | **Desviación** |
| Release Year | 2006 | 2021 | 2018 | 2019 | 2,35138 |
| Average Rating | 4 | 9,4 | 8 | 7,8 | 1,09045 |
| Number of Reviews | 8000 | 65000 | 23146 | 19000 | 12718,45399 |

Para columnas categóricas:

* Frecuencia de cada categoría. 1
* Número de categorías únicas. 82

1. **Conclusiones Iniciales del Grupo**

* ¿Qué información relevante puede extraerse de esta base de datos? R/ Se puede saber las películas más vistas, los géneros más vistos, las películas o géneros más recomendados,
* ¿Qué problemas de calidad o estructura podrían afectar un análisis más profundo?

Datos no entendibles como 1:05:22 â€“ Cobbâ€™s totem scene

* ¿Qué preguntas podría ayudar a responder esta base de datos? Que se busca con recopilar esta información: Mejores o peores películas acordes a la calificación, cantidad de vistas, géneros más vistos entre otros.
* ¿Qué mejoras o limpiezas serían necesarias? Minute of Life-Changing Insight, no me parece necesario para la base, ya que es algo más personal su percepción.

1. **Entrega**

Cada grupo debe entregar este documento diligenciado y, si es posible, el archivo de base de datos con sus observaciones o anotaciones.