**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.

**Integrantes**:

* David Nieto
* David Gómez
* Vanessa Valencia
* Juan Diego Solórzano

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 Dataset (NLID)
* 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 | 82 |
| Genre | String | Género de la película | 82 | 22 |
| Release Year | Int | Año de lanzamiento | 82 | 11 |
| Average Rating | float | Calificación promedio | 82 | 37 |
| Number of Reviews | Int | Numero de reseñas | 82 | 34 |
| Review Highlights | String | Aspecto destacado | 82 | 82 |
| Minute of Life-Changing Insight | String | Minuto de revelación que cambia tu vida | 82 | 82 |
| How Discovered | String | Como la descubriste | 82 | 13 |
| Meaningful Advice Taken | String | Lecciones aprendidas | 82 | 81 |
| Suggested to Friends/Family Y/N %) | String | Recomendación de amigos y familia | 82 | 45 |

1. **Calidad de los Datos**

* ¿Hay valores faltantes en alguna columna? ¿Cuántos? = 0
* ¿Hay columnas con valores repetidos o inconsistentes? = 0
* ¿Existen datos fuera de rango o errores evidentes? = 0
* ¿Conoce la última fecha de actualización de los datos? Hace 20 días

1. **Análisis Cuantitativo Básico**

Para columnas numéricas:

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

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Variable | Mínimo | Máximo | Media | Mediana | Desviación |
| Año de lanzamiento | 2006 | 2021 | 2018.3 | 2019 | 2.35 |
| Calificación promedio | 4 | 9 | 7.4 | 8 | 1.3 |
| Número de reseñas | 8000 | 65000 | 23146 | 19000 | 12718 |

Para columnas categóricas:

**Categoría 1: Título de la película**

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

**Categoría 2: Género**

* Frecuencia de cada categoría =

|  |  |
| --- | --- |
| **Género** | **Frecuencia** |
| Drama | 24 |
| Comedy | 8 |
| Thriller | 7 |
| Sci-Fi | 6 |
| Horror | 6 |
| Romance | 5 |
| Animation | 4 |
| Action | 4 |
| Documentary | 2 |
| War/Drama | 2 |
| Western | 2 |
| Historical Drama | 2 |
| Thriller/Drama | 1 |
| Adventure | 1 |
| Comedy/Family | 1 |
| Mystery | 1 |
| Comedy/Crime | 1 |
| Crime/Drama | 1 |
| Comedy/Drama | 1 |
| Romance/Drama | 1 |
| Horror/Comedy | 1 |
| Satire | 1 |
| **Total general** | **82** |

* Número de categorías únicas = 22

**Categoría 3: Número de reseñas**

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

**Categoría 4: Número de reseñas**

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

**Categoría 5: Aspecto destacado**

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

**Categoría 6: Minuto de revelación que cambia tu vida**

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

**Categoría 7: Como la descubriste**

* Frecuencia de cada categoría. =

|  |  |
| --- | --- |
| **Cómo la descubriste** | **Frecuencia** |
| Netflix original | 48 |
| Netflix recommendation | 13 |
| Friend suggested | 5 |
| Film festival | 4 |
| Social media | 3 |
| Film critic | 2 |
| Podcast mention | 1 |
| Oscar buzz | 1 |
| Stephen King adaptation | 1 |
| Film class | 1 |
| Word of mouth | 1 |
| Critic review | 1 |
| Book adaptation | 1 |
| **Total general** | **82** |

* Número de categorías únicas = 13

**Categoría 8: Lecciones aprendidas**

* Frecuencia de cada categoría. =

|  |  |
| --- | --- |
| **Lecciones aprendidas** | **Frecuencia** |
| Sacrificio demandado por el arte | 2 |
| Resto | 1 |

* Número de categorías únicas = 81

**Categoría 9: Recomendación amigos y familia**

* Número de categorías únicas = 45

1. **Conclusiones Iniciales del Grupo**

* ¿Qué información relevante puede extraerse de esta base de datos?

Podemos extraer el género, como conocimos la película, año de lanzamiento, calificación promedio y numero de reseñas.

* ¿Qué problemas de calidad o estructura podrían afectar un análisis más profundo?

En algunos campos que deberían ser enteros o decimales se complica porque se convierte en cadena de texto.

* ¿Qué preguntas podría ayudar a responder esta base de datos?
  + Porque medio se está difundiendo la película.
  + Cuál es el género que mas consumido.
* ¿Qué mejoras o limpiezas serían necesarias?
  + Normalizar aquellos campos que se combina texto y número

1. **Entrega**

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