**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.

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

* Daniela Utima
* Paula Alejandro Lock
* Leonardo Gonzalez
* Yeisson Torres.

1. **Selección de la Base de Datos**

* Nombre de la base de datos**: BD\_anuario\_2024**
* Fuente (URL, repositorio, etc.): **Google -**
* Formato del archivo (.csv, .xlsx.): .xlsx

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

* Número total de filas (registros): **9697**
* Número total de columnas (atributos o variables): **20**

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)** |
| NÚM. | | | | | Numérico | Ordenado | 9696 | 9696 |
| AÑO | | | | | Numérico | Año De Los Datos | 9696 | 1 |
| GÉNERO | | | | | Texto | Genera La Población | 9696 | 2 |
| GRADO | | | | | Numérico | Grado De Escolaridad | 9696 | 15 |
| GRUPO | | | | | Numérico | Clase | 9201 | 27 |
| FECHA DE NACIMIENTO | | | | | Fecha | Fecha De Nacimiento | 9696 | 3916 |
| EDAD | | | | | Numérico | Antigüedad Del Humano | 9696 | 47 |
| NACIONALIDAD | | | | | Texto | Nacionalidad Del Individuo | 9696 | 7 |
| ESTADO DE ORÍGEN | | | | | Texto | Lugar De Nacimiento | 9696 | 33 |
| NIVEL ANTERIOR | | | | | Texto | Grado De Escolaridad Anterior | 9696 | 6 |
| NIVEL ACADÉMICO | | | | | Texto | Nivel De Profesionalidad | 9696 | 3 |
| NIVEL ACTUAL | | | | | Texto | Nivel Académico Actual | 9696 | 5 |
| UNIDAD ACADÉMICA | | | | | Texto | Facultad | 9696 | 22 |
| SEDE | | | | | Texto | Sede | 9696 | 10 |
| SEDE EDO | | | | | Texto | Ciudad | 9696 | 10 |
| PROGRAMA EDUCATIVO | | | | | Texto | Carrera Profesional | 9696 | 59 |
| LENGUA | | | | | Texto | Idioma Adicional Al Castellano | 1829 | 50 |
| AFRODESCENDIENTE | | | | | Texto | Raza | 749 | 1 |
| DISCAPACIDAD | | | | | Texto | Discapacidad | 354 | 1 |
| TIPO DISCAPACIDAD | | | | | Texto | Tipo Discapacidad | 354 | 5 |
|  |  |  |  |  |
|  |  |  |  |  |

1. **Calidad de los Datos**

* ¿Hay valores faltantes en alguna columna? ¿Cuántos? **4 columnas con valores faltantes**
* ¿Hay columnas con valores repetidos o inconsistentes? **Si, repetidos si, inconsistentes no.**
* ¿Existen datos fuera de rango o errores evidentes? **No. Todos los datos son coherentes.**
* ¿Conoce la última fecha de actualización de los datos? **2024**

1. **Análisis Cuantitativo Básico**

Para columnas numéricas:

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

**GRADO**

**Mínimo = 1, máximo = 18, media 5.04, mediana = 5, desviación estándar =3.01**

Para columnas categóricas:

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

|  |  |
| --- | --- |
| **ESTADO DE ORIGEN.** | **CANTIDAD** |
| AGUASCALIENTES | 44 |
| BAJA CALIFORNIA | 69 |
| BAJA CALIFORNIA SUR | 9 |
| CAMPECHE | 72 |
| CHIAPAS | 480 |
| CHIHUAHUA | 78 |
| CIUDAD DE MÉXICO | 157 |
| COAHUILA | 43 |
| COLIMA | 22 |
| DURANGO | 285 |
| EXTRANJERO | 22 |
| GUANAJUATO | 201 |
| GUERRERO | 458 |
| HIDALGO | 452 |
| JALISCO | 84 |
| MÉXICO | 2579 |
| MICHOACÁN | 148 |
| MORELOS | 209 |
| NAYARIT | 71 |
| NUEVO LEÓN | 9 |
| OAXACA | 1721 |
| PUEBLA | 1025 |
| QUERÉTARO | 114 |
| QUINTANA ROO | 34 |
| SAN LUIS POTOSÍ | 142 |
| SINALOA | 29 |
| SONORA | 15 |
| TABASCO | 57 |
| TAMAULIPAS | 57 |
| TLAXCALA | 159 |
| VERACRUZ | 738 |
| YUCATÁN | 40 |
| ZACATECAS | 73 |
| **Total general** | **9696** |

1. **Conclusiones Iniciales del Grupo**

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

**Demográficos (genero, edad, ciudad de residencia, lugar de origen)**

**Académicos (nivel de estudio profesional)**

**Ubicación Geográfica**

* ¿Qué problemas de calidad o estructura podrían afectar un análisis más profundo?
  + **Datos en blanco**
  + **Fechas erradas**
  + **Longitudes extensas de algunos campos**
  + **Tildes y caracteres especiales**
* ¿Qué preguntas podría ayudar a responder esta base de datos?
  + **Qué genero prevalece, cuales son los rangos de edad, cuales son las principales ciudades de residencia, lugar de origen predominante**
  + **Qué nivel Académicos existen por edad**
  + **Ubicación Geográfica por edad, por estudio, entre otros**
* ¿Qué mejoras o limpiezas serían necesarias?
  + **Si, dado que ayudarían a un análisis más preciso y ágil para la interpretación de los datos tanto cualitativos como cuantitativos.**

1. **Entrega**

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