

# Manejo de Datos

## Proyecto final

Profesor: Ricardo López

Fecha de entrega: 13 de Diciembre

### 1. Introducción

El Censo de Población y Vivienda es un ejercicio que realiza el Instituto Nacional de Estadística y Geografía (INEGI) cada 10 años, en el cual un conjunto de entrevistadores recorren las viviendas del país con el fin de obtener información sobre estas tal como conteo de población y algunas características demográficas. La información que se recoge se puede consultar de manera publica en el sitio oficial del INEGI ([inegi.org.mx](https://www.inegi.org.mx)). El último censo realizado fue en el año 2020, tomaremos el conjunto de datos abiertos disponibles en la sección de "Sistema de consulta" (<https://www.inegi.org.mx/datosabiertos/>) para obtener datos estadísticos y crear una pequeña base de datos.

### 2. Problemática

El problema general es leer los datos de `INE_DISTRITO_2020.csv` para obtener algunos datos estadísticos y agregarlos a una pequeña base de datos. La descripción de los encabezados del primer archivo los encontraras en la columna *Nemónico* del archivo `Descriptor_indicadores_Distrito_2020.csv`.

La primer tarea a realizar es leer los datos del archivo `INE_DISTRITO_2020.csv` y responder las siguientes preguntas:

1. ¿Cuál es la población total de cada entidad?
2. Por cada entidad ¿Cuántos distritos contienen población indígena?
3. ¿Cuál es la población total de 0 a 2 años de cada entidad?
4. ¿Cuál es la población total de 3 años y más que hablan alguna lengua indígena de cada entidad?
5. ¿Cuál es el promedio de la población de 3 a 14 años que no asiste a la escuela?

Cada pregunta se debe resolver programando las funciones de Python agregadas en la plantilla `proyectoMD.py`.

Todas las funciones deben regresar como resultado un diccionario de Python, por ejemplo para la pregunta 1 el diccionario tendrá como llaves los nombres de las entidades y como valor la población total calculada. Por ejemplo :

```
{ 'Aguascalientes ': 1425607, 'Baja California ': 3769020, 'Baja California Sur ': 798447, ..., 'Zacatecas ': 1622138 }
```

**Nota IMPORTANTE:** Si las funciones no regresan el diccionario con el formato pedido se considerara que el ejercicio está mal y se evaluara con cero dicha función.

La segunda tarea consiste en agregar datos del archivo excel a una base de datos. El esquema de la base de datos es muy sencillo, solo constara de una tabla con cinco atributos o columnas.

El esquema de la table es el siguiente:

```
entidad(id, nombre, poblacion_total, poblacion_total_fem,
        poblacion_total_masc)
```

Usando el manejador de bases de datos *SQLite3* y la herramienta gráfica *DB Browser for SQLite* crea una base de datos con el nombre: **resumen\_inegi** y crea la tabla con el esquema ya mencionado.

Para continuar implementa la función **agrega\_datos** para que haga lo siguiente:

- Leer el archivo **INE\_DISTRITO\_2020.csv**
- Obtener los totales de población, totales de población femenina y totales de población masculina para cada entidad
- Agregar a la base de datos que creaste anteriormente los totales calculados para cada entidad a la tabla **entidades**.

Por último debes programar la función **consulta\_datos** que debe recibir el identificador de una entidad y regresar los datos almacenados en la tabla **entidades** correspondientes.

Esta función debe regresar un diccionario con el siguiente formato:

```
{
'Aguascalientes': [1425607, 728924, 696683],
'Baja California': [3769020, 1868431, 1900589],
'Baja California Sur': [798447, 392568, 405879],
...,
'Zacatecas': [1622138, 831080, 791058]
}
```

Donde el primer valor de la lista corresponde a la población total, el segundo valor a la población total femenina y el tercero a la población total masculina.

### 3. Lineamientos de entrega

- Utiliza la plantilla **proyectoMD.py** proporcionada, puedes agregar las funciones que consideres necesarias para resolver los ejercicios.
- Todos los ejercicios deben estar en el archivo **proyectoMD.py** si se agregan archivos extra se penalizará con un punto menos.
- Agraga un archivo **readme.txt** con los siguientes datos de cada integrante de tu equipo:
  - Nombre completo.
  - Número de cuenta.
  - Correo electrónico.

- El proyecto se entregará por medio de la plataforma Google Classroom. La entrega de prácticas y tareas tendrá hora límite hasta las 11:59 p.m del día de entrega.

**Importante:** En esta ocasión no habrá prórrogas para la entrega ni se recibirán proyectos entregados con retraso.

- Si se sospecha de algún tipo de copia se entrevistará a los alumnos implicados con el fin de que expliquen o justifiquen sus respuestas, de determinar que hubo una copia se evaluará con CERO a los implicados en ésta práctica.