Universidad Técnica Particular de Loja

Estudiante:

Camilo José López Correa

Docente:

Mgtr. Ángel Eduardo Encalada Encalada

Asignatura:

Base de Datos Avanzada

Fecha:

Viernes 28 de junio 2024

Tema:

Análisis implementación y carga de datos sobre Detenciones

Proyecto integrador Practicum 1.2

Abril - agosto 2024

Tabla de contenido

[1. Introducción 1](#_Toc170556294)

[2. Repositorio GitHub 1](#_Toc170556295)

[3. Datos Base 1](#_Toc170556296)

[4. Datos Complementarios 1](#_Toc170556297)

[5. Posibles Análisis para realizar 1](#_Toc170556298)

[6. Diseño conceptual 1](#_Toc170556299)

[7. Diseño lógico relacional 1](#_Toc170556300)

[8. Implementación y carga 1](#_Toc170556301)

[9. Script SQL 1](#_Toc170556302)

[10. Descarga en formato CSV 1](#_Toc170556303)

[11. Bibliografía 1](#_Toc170556304)

# Introducción

En este informe se presentar información detallada sobre el manejo de los datos proporcionados. Se han buscado e implementado datos adicionales con el objetivo de complementar y enriquecer nuestra base de datos, permitiendo realizar análisis más completos y precisos en el futuro. Para la limpieza y carga de los datos, se ha utilizado Python junto con la biblioteca Pandas, facilitando el proceso y asegurando la calidad de los datos gestionados.

# Repositorio GitHub

Para facilitar la revisión de los datos y evidencia de la implementación del proyecto de a creado un repositorio en la plataforma GitHub, la cual es especializada en programación y bases de datos.

El enlace correspondiente para la redirección a al repositorio es el siguiente:

<https://github.com/CJlopez17/IntegradorCuarto.git>

En se encontrará con dos carpetas: Datos crudos, Datos Limpios.

* Datos Crudos: Contine los datos sin limpiar, con campos que deben eliminarse, valores con tipos de datos incorrecto, entre otros problemas.
* Datos Limpios: Contiene los datos después del proceso de limpieza. Se ha aplicado los scripts **limpiezaDatosComple.py** y **limpieza.py** para depurar estos datos.

Adicional se encontrará con el script que se usó para la carga de los datos.

# Datos Base

Para los datos base, se nos procinto un archivo .xlsx titulado “Base\_de\_datos\_DETENIDOS\_ANIO\_2016\_ABRIL\_2024” que contenía 9 hojas, las cuales llevaban el nombre del año al cual correspondían sus datos.

* Hoja 2016 a la 2021: estas hojas comparten las mismas columnas, que son: Zona, Subzona, Cantón, presunta\_subinfraccion, Fecha de Detención, Sexo, Edad, Nacionalidad.
* Hoja 2022 a la 2024: esta hoja se añadieron columnas adicionales para mejorar la precisión de la información sobre la detención. Las columnas adicionales incluyen: codigo\_iccs, tipo, estado\_civil, estatus\_migratorio, edad, sexo, genero, nacionalidad, autoidentificacion\_etnica, numero\_detenciones, nivel\_de\_instruccion, condicion, movilizacion, tipo\_arma, arma, fecha\_detencion\_aprehension, hora\_detencion\_aprehension, lugar, tipo\_lugar, codigo\_distrito, codigo\_circuito, codigo\_subcircuito, nombre\_zona, nombre\_subzona, nombre\_distrito, nombre\_circuito, nombre\_subcircuito, codigo\_provincia, codigo\_canton, codigo\_parroquia, nombre\_provincia, nombre\_canton, nombre\_parroquia, presunta\_flagrancia, presunta\_infraccion, presunta\_subinfraccion, presunta\_modalidad.

# Datos Complementarios

Para complementar nuestra análisis del índice de criminalidad por Provincia, Cantón y Parroquia, se decidió incorporar datos demográficos de las provincias. Estos datos fueron obtenidos de la página web [www.censoecuador.gob.ec](http://www.censoecuador.gob.ec/), donde se localizaron varios archivos en formatos .csv y .xlsx que aportan información valiosa para completar nuestros datos.

Uno de los archivos más relevantes fue el ENEMDU, que incluye un censo realizado en el año 2023. Este archivo, denominado “2022\_CPV\_NACIONAL\_DENSIDAD\_POBLACIONAL”, resultó ser el más preciso en cuanto a los niveles de población y densidad poblacional. Contiene 3 hojas que detallan el número de habitantes, la superficie en kilómetros cuadrados y la densidad poblacional. Los datos se desglosan de manera progresiva a medida que se avanza en las hojas: la primera hoja presenta datos a nivel de provincia, la segunda a nivel de cantón y la tercera a nivel de parroquia. Por ejemplo, en la primera hoja encontramos: Azuay; 801.609 habitantes; 8.173 km²; densidad de 98 habitantes por km². En la segunda hoja: Azuay; Cuenca; 596.101 habitantes; 3.195 km²; densidad de 187 habitantes por km². Y en la tercera hoja: Azuay; Cuenca; Baños; 21.797 habitantes; 251 km²; densidad de 87 habitantes por km².

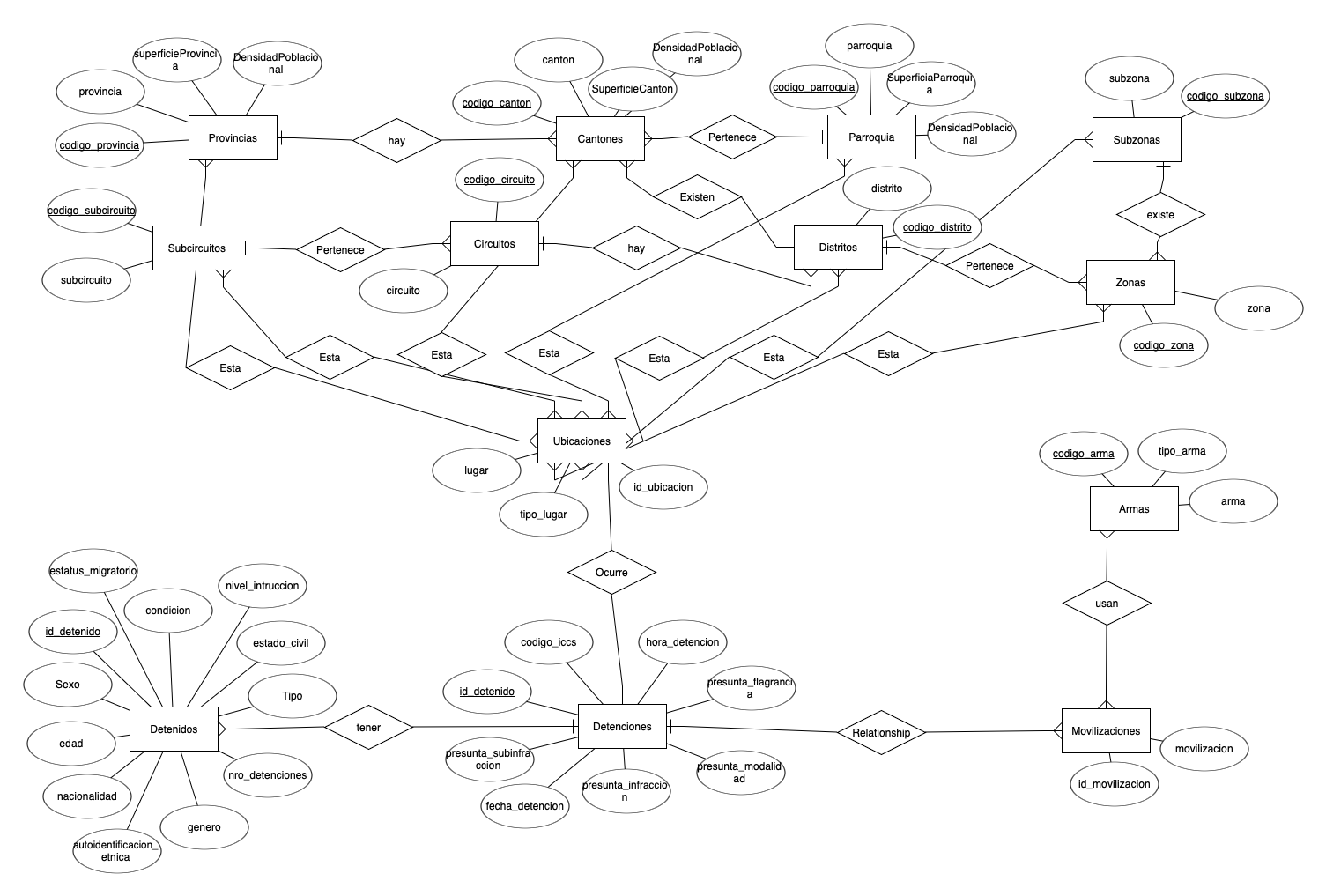
# Posibles Análisis para realizar

A partir de la evaluación de la naturaleza de los datos recolectados, hemos determinado que los análisis óptimos son los siguientes:

* Índice de criminalidad por provincia, cantón, parroquia, zona.
* Número de armas por provincia, cantón y parroquia.
* Tipo de arma más frecuente en cada provincia, cantón y parroquia.
* Tipo de arma más frecuente en las regiones Costa, Sierra, Amazonia y Galápagos.
* Lugares más frecuentes de robos, asesinatos, femicidios, entre otros crímenes.
* Provincia con mayor insistencia de robos.
* Catones con mayor tasa de asesinatos.
* Genero con mayor incidencia en actos violentos.
* Genero más violento desglosados por provincia.
* Nacionalidad extranjera más frecuente involucrada en robos, asesinatos, femicidios y otros crímenes
* Meses del año con mayor frecuencia de crímenes, diferenciados por año.

Estos análisis proporcionan una base comprensiva para entender las tendencias y patrones de criminalidad en diferentes niveles de geografía y demografía. Además, es importante señalar que si las circunstancias lo requieres, se podría realizar análisis adicionales para explorar más profundamente cualquier aspecto específico o emergente que surja de los datos iniciales.

# Diseño conceptual



En la imagen presentada observamos un diseño conceptual relacional, en el que nos encontramos con 13 entidades, cada una con sus respectivos atributos. Las relaciones costa de la siguiente manera:

* Detenido (1: M) Detenciones.
* Detenciones (1: M) Movilizaciones
* Movilizaciones (M:M) Armas
* Detenciones (1: N) Ubicaciones
* Ubicaciones (N: N) Provincias
* Ubicaciones (N: N) Cantones
* Ubicaciones (N: N) Circuitos
* Ubicaciones (N: N) Subcircuitos
* Ubicaciones (N: N) Parroquias
* Ubicaciones (N: N) Distrititos
* Ubicaciones (N: N) Subzonas
* Ubicaciones (N: N) Zonas
* Provincias (1: N) Cantones
* Cantones (1: N) Distrititos
* Distritos (1: N) Circuitos
* Circuitos (1: N) Subcircuito
* Cantones (1: N) Parroquias
* Zonas (1: N) Distrititos
* Zonas (1: N) Subzonas

Todas estas relaciones nos permiten ver un panorama claro acerca de cómo sería una de las formas de relacionar los datos proporcionados y los adicionales.

## Diccionario de datos

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Tipo Entidad | Atributo | Obligatorio | Restricciones | Valor por defecto |
| Detenciones | presunta\_flagrancia | x | Solo los valores: ‘Si’ y ‘No’ | ‘No’ |
| Detenido | sexo | x | Solo valores: ‘Hombre’ y ‘mujer’ |  |
| Detenido | tipo | x | Solo valores: ‘Detenido’ y ‘Aprehendido’ |  |
| Detenciones | Presunta\_modalidad | x |  | ‘No aplica’ |

# Diseño lógico relacional

# Implementación y carga

# Script SQL

# Descarga en formato CSV

# Bibliografía