

**MODALIDAD PRESENCIAL**

**ÁREA TÉCNICA**

**NOMBRE DE LA TITULACIÓN**

**INGENIERÍA CIENCIAS DE LA COMPUTACIÓN**

**PROYECTO INTEGRADOR BASES DE DATOS AVANZADA**

**INFORME FINAL**

**INTEGRANTES**: Cristhian Michael Jiménez Jiménez

Jorge Enrique Armijos Cabrera

Roger Lizandro Criollo Largo

**FECHA**: 02/07/2023

**DOCENTE**: Mgtr. Juan Carlos Morocho

**PERIODO ACADÉMICO**

Abril 2023 – Agosto 2023

Contenido

[Introducción 3](#_Toc139213129)

[Objetivo General 4](#_Toc139213130)

[Objetivos Específicos 4](#_Toc139213131)

[Repositorio GitHub 4](#_Toc139213132)

[Datos base 5](#_Toc139213133)

[Datos Complementarios 7](#_Toc139213134)

[Lugares Relevantes 7](#_Toc139213135)

[Diseño Conceptual 8](#_Toc139213136)

[Diseño Lógico relacional 10](#_Toc139213137)

[Diseño Físico 11](#_Toc139213138)

[Implementación y carga 12](#_Toc139213139)

[Script SQL 16](#_Toc139213140)

[https://github.com/michael-jmz/analisis\_acceso\_viviendas\_ecuador/blob/main/script\_corregido.sql 16](#_Toc139213141)

[Descarga en formato CSV 16](#_Toc139213142)

[Conclusiones 17](#_Toc139213143)

[Recomendaciones 17](#_Toc139213144)

[BIBLIOGRAFIA 18](#_Toc139213145)

# Introducción

El Instituto Nacional de Estadística y Censos del Ecuador, en su constante compromiso de proporcionar información confiable y actualizada sobre la situación socioeconómica del país, llevó a cabo la Encuesta Nacional de Empleo Desempleo y Subempleo en el primer trimestre del año 2023. Esta encuesta tiene como objetivo principal recopilar datos exhaustivos sobre el mercado laboral ecuatoriano, pero también proporciona una valiosa oportunidad para investigar las principales características físicas de las viviendas en el país.

El análisis de los datos de la Encuesta Nacional de Empleo Desempleo y Subempleo - I trimestre 2023 relacionado con las características físicas de las viviendas es de vital importancia, ya que ofrece información crucial para comprender las condiciones de vida de la población ecuatoriana. El acceso a una vivienda adecuada es un derecho fundamental y su estudio detallado nos permite evaluar el grado de satisfacción de esta necesidad básica en diferentes regiones y segmentos de la población (INEC, 2023).

Este proyecto de análisis de datos se pretende investigar y examinar las principales características físicas de las viviendas en Ecuador, utilizando los datos recopilados en la encuesta. Entre las variables relevantes se encuentran el tamaño de las viviendas, la calidad de la construcción, la disponibilidad de servicios básicos como agua potable y electricidad, la ubicación geográfica, entre otros aspectos fundamentales.

Al analizar estos datos, también se busca identificar patrones, tendencias y disparidades en cuanto a las condiciones físicas de las viviendas en el país. Este análisis contribuirá a generar información precisa y actualizada que pueda ser utilizada por el gobierno, las organizaciones no gubernamentales y otros actores involucrados en la planificación y ejecución de políticas públicas relacionadas con vivienda y desarrollo urbano (INEC, 2023).

Los resultados obtenidos a partir de este análisis podrán ser utilizados para tomar decisiones informadas en materia de vivienda, identificar áreas de mejora y diseño de intervenciones específicas que permitan promover el acceso a viviendas dignas y mejoradas en todo el territorio ecuatoriano.

# Objetivo General

El propósito de este proyecto es llevar acabo un análisis detallado y desarrollar un esquema de Base de Datos escalable y completamente funcional que gestione información los datos sobre información de acceso y condiciones de vivienda en el Ecuador en el primer trimestre del año 2023 que realizo el INEC

# Objetivos Específicos

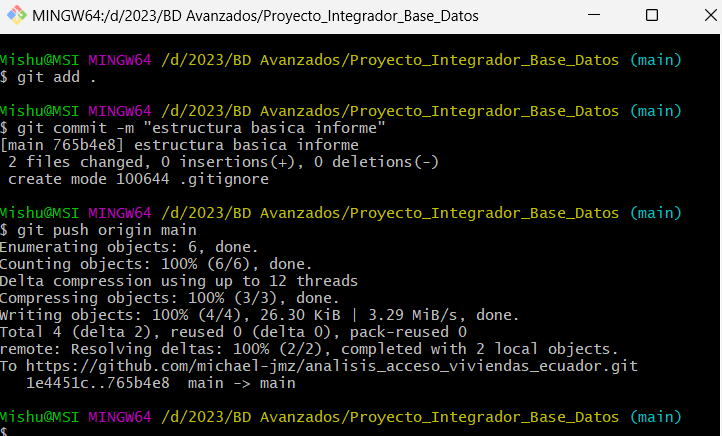
* Análisis sobre la estructura y relación de la data (formato proporcionado .csv) proporcionados por el INEC en el primer trimestre del 2023
* Búsqueda e implementación de nuevos datos que nos pueden ayudar a investigación por el INEC 2023
* Interpretación y limpieza de la data
* Diseño, modelado y relación de la data mediante modelo Entidad-Relación (E-R)
* Implementación de modelo Lógico a partir del modelo E-R
* Aplicación de Formas Normales de Base de Datos al modelo lógico y así reducir redundancia
* Implementación del modelo físico de base de datos.

# Repositorio GitHub

Enlace al repositorio [aquí](https://github.com/michael-jmz/analisis_acceso_viviendas_ecuador.git) https://github.com/michael-jmz/analisis\_acceso\_viviendas\_ecuador

**Figura 1**

*Añadiendo archivos al repositorio de GitHub*



**Nota**: Elaboración propia

# Datos base

Paro una mejor comprensión de la data que se nos proporcionó, explicamos algunos columnas y palabras presentes en este [archivo\_csv](https://github.com/michael-jmz/analisis_acceso_viviendas_ecuador/blob/main/data_inicial_enemdu_vivienda_hogar_2023_I_trimestre.csv) que podrían ser algo confusas al inicio, pero con las definiciones siguiente que nos proporciona el INEC estamos seguros de que podrás tener una idea clara de que estamos hablando:

* **Área**: Indica el área geográfica donde se realizó la encuesta, pudiendo ser urbana (1) o rural (2).
* **Provincia**: tenemos un código de dos dígitos por cada provincia del ecuador de acuerdo con la División Político-Administrativa (DPA),

Imagen que contiene Gráfico

Descripción generada automáticamente

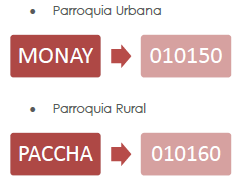
Nota: INEC. (2020). *Codificacion Provincial* [png]. Ecuador en Cifras.

* **Cantones:** Un cantón se identifica con 4 dígitos, 2 primero de provincia y los 2 del cantón respectivita como se muestra a continuación.

Icono

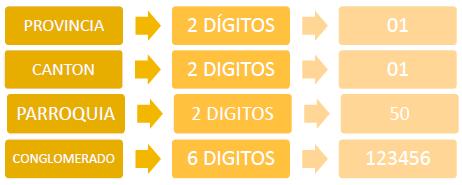
Descripción generada automáticamente

* **Parroquia**: Una parroquia se identifica con 6 dígitos: dos dígitos de provincia, dos dígitos de cantón y dos dígitos de parroquia, de la siguiente manera:



***Importante:*** Todas las parroquias urbanas son codificadas con “50” para este estudio.

* **Ciudad**: Nombre de la ciudad o localidad donde se llevó a cabo la encuesta.
* **Conglomerado**: Identificador del conglomerado de viviendas seleccionado para la muestra de la encuesta. Es una identificación única de 6 dígitos por Unidad Primaria de Muestreo diferenciado por área urbana y rural.



* **Zonas no delimitadas:** Son zonas que tienen pendiente definir sus límites o sus jurisdicciones

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

* **Panelm**: Indicador del panel muestral al que pertenece la vivienda.
* **Vivienda**: Número de identificación de la vivienda encuestada.
* **Hogar**: Número de identificación del hogar dentro de la vivienda.
* **Vi01** a **Vi14**: Estas variables representan una serie de preguntas realizadas durante la encuesta. Cada una de ellas tiene un código asociado y su significado específico puede ser consultado en la documentación de la encuesta. Estas preguntas pueden estar relacionadas con diferentes aspectos, como la composición del hogar, características demográficas, acceso a servicios básicos, entre otros.
* **Estrato**
* : Indicador del estrato socioeconómico al que pertenece la vivienda encuestada.
* **Fexp**: Factor de expansión que permite extrapolar los resultados de la muestra a la población total.
* **Upm**: Unidad primaria de muestreo a la que pertenece la vivienda encuestada.
* **Id\_vivienda:** Identificador único de la vivienda.
* **Id\_hogar:** Identificador único del hogar dentro de la vivienda.
* **Periodo**: Indica el periodo de tiempo en el que se realizó la encuesta.
* **Mes**: Mes correspondiente al periodo de tiempo de la encuesta.

Estas variables proporcionan información detallada sobre las características físicas de las viviendas, la composición de los hogares encuestados y otros aspectos relevantes relacionados con la encuesta. El análisis de estas variables permitirá obtener una visión más completa de la situación socioeconómica y de vivienda en Ecuador durante el primer trimestre del 2023 (INEC, 2020).

# Datos Complementarios

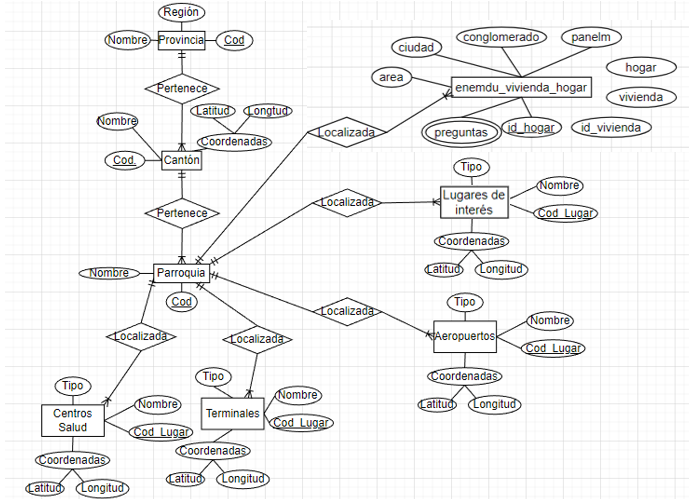
|  |  |  |
| --- | --- | --- |
| **Descripción General** | **Propósito** | **Fuentes de datos** |
| Proyecciones Poblacionales en Ecuador | La disposición de datos sobre la población en cada territorio del ecuador, tanto en hombres como mujeres abre la posibilidad a realizar un cálculo estimado del mercado que dispone dicho alojamiento. | <http://ambar.utpl.edu.ec/dataset/proyecciones-poblacionales-en-ecuador> |
| Lugares Relevantes | La data sobre los lugares de relevancia turística en todo el Ecuador, con su ubicación geográfica y dirección. Se puede obtener una estimación de que tanta cantidad de personas se alojen en dichos establecimientos a partir de las atracciones turísticas del lugar en el que se encuentra. | <http://ambar.utpl.edu.ec/nl/dataset/lugares-turistico> |
| Centros de salud | En este archivo se encuentran las ubicaciones de los diferentes centros de salud. | <http://ambar.utpl.edu.ec/dataset/centros-de-salud-sistema-nacional-de-informacion-sni> |
| Localización Municipios en Ecuador | Con esta información se piensa localizar las diferentes cabeceras cantonales. | <http://ambar.utpl.edu.ec/dataset/localizacionmunicipios-en-ecuador> |
| Cantidad de Población rural. | Esta información nos ayudara para totalizar la cantidad de población rural existe por lugar, e identificar la densidad poblacional. | <https://data.apps.fao.org/catalog/dataset/datos-ecuador/resource/fede540f-8c9a-4872-8b63-643ffdd9e349> |
| Datos generales y relevantes del Ecuador | Esta información sirve para nutrir la información existente y proporcionar de información general al dataset. | <https://knoema.es/atlas/Ecuador/datasets> |
| Datos de la organización territorial del Ecuador. | Esta información nos sirve para poder identificar las provincias, cantones y ciudades, en donde se encuentran las casas encuestadas. | <http://ambar.utpl.edu.ec/dataset/clasificador-geografico-estadistico-en-ecuador> |

# Diseño Conceptual

En el diagrama Entidad-Relación de la **Figura** se puede observar entre las entidades de mayor importancia son: Área, Regiones, Conglomerados, Ciudad, Vivienda además se modela las cuatro entidades más que consideramos podría ser de utilidad para su posterior análisis del acceso y condiciones de vivienda en Ecuador 2023

**Figura 2**

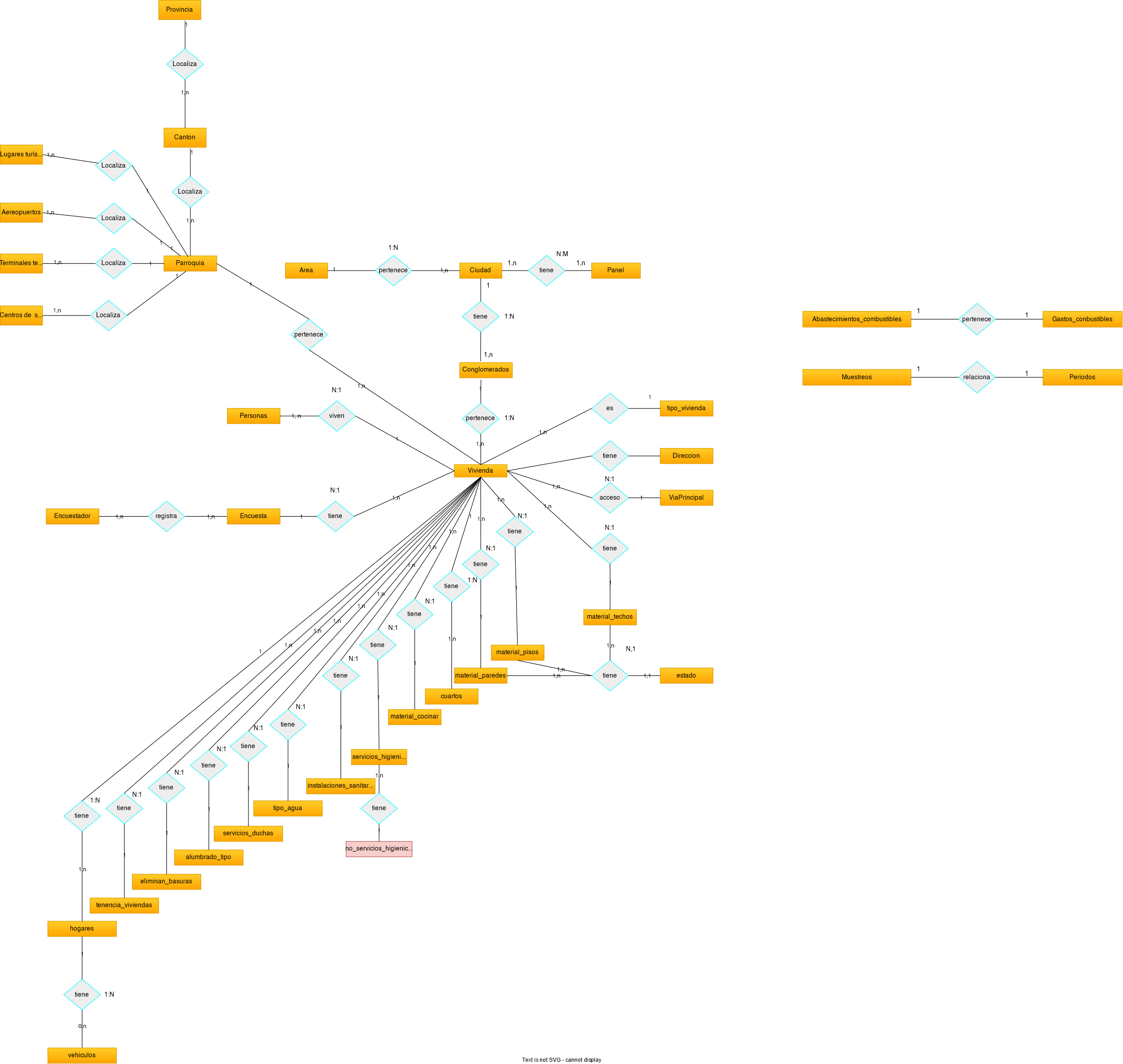
*Modelo conceptual de acceso y condiciones de vivienda*

**Nota:** Elaboracion propia, se puede muestra un porcion del modelo completo por cuestiones de resolucion en el documento, Para visualizar el modelo conceptual clic [aquí](https://drive.google.com/file/d/1Op5cI0pbPgGI2UjWe4vpxqjvkUt6SWVc/view?usp=sharing). <https://drive.google.com/file/d/1Op5cI0pbPgGI2UjWe4vpxqjvkUt6SWVc/view?usp=sharing>

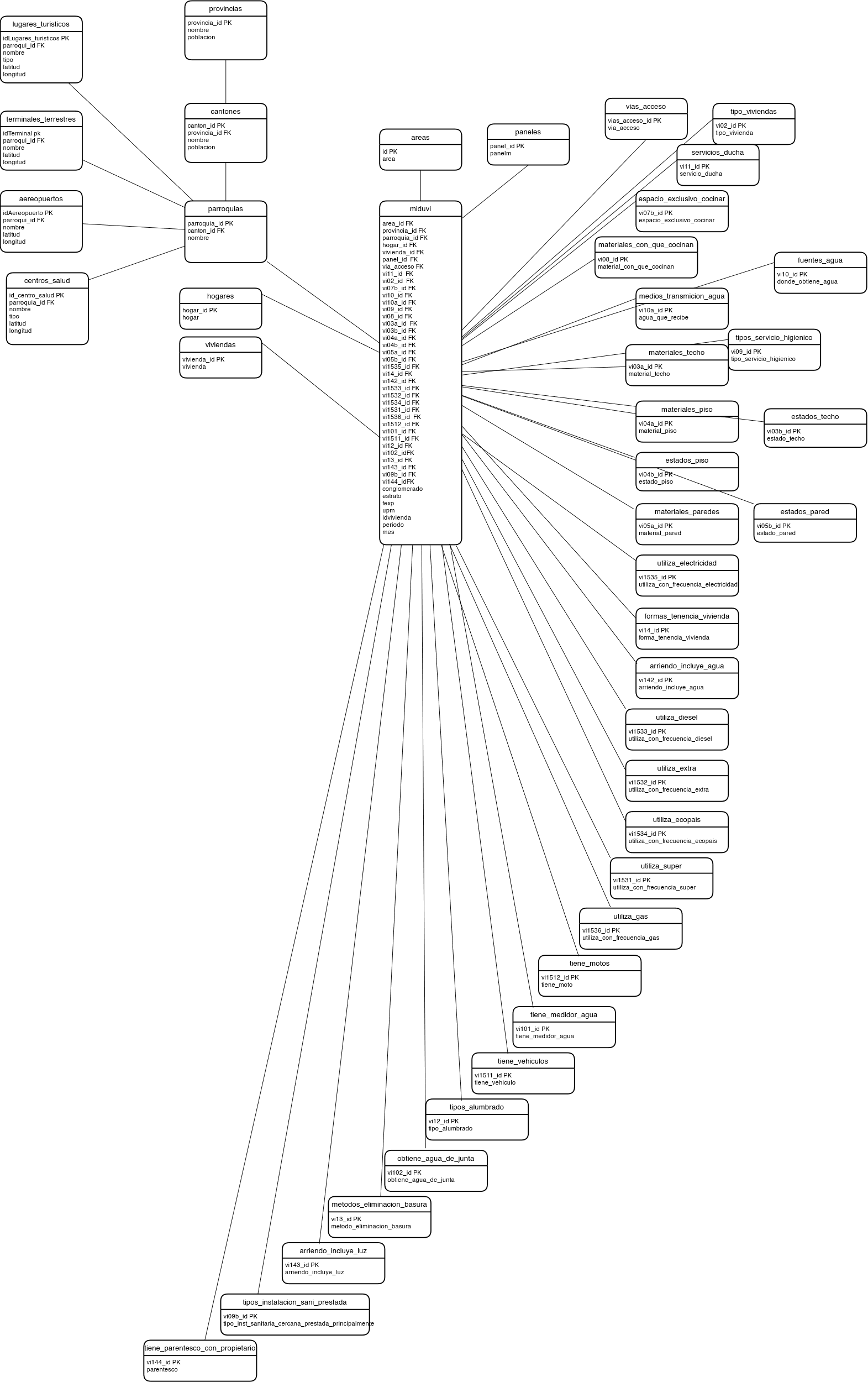
**Figura 3**

*Modelo conceptual en minutara completo*



# Diseño Lógico relacional

Para una mejor visualización del diseño lógico vaya a este enlace:



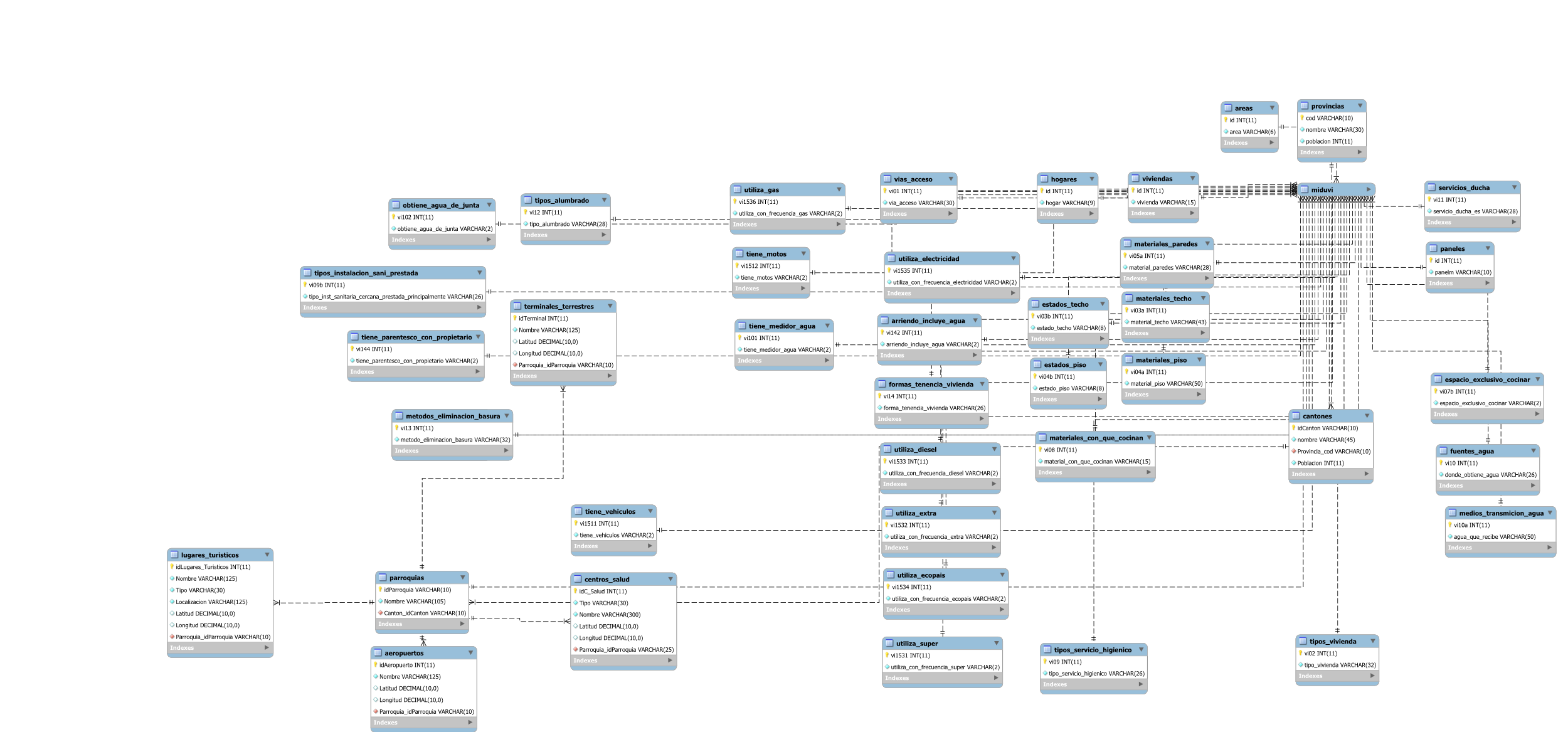
# Diseño Físico

Enlace del diccionario clic [aquí](https://github.com/michael-jmz/analisis_acceso_viviendas_ecuador/blob/main/diccionario_vivienda.html): https://github.com/michael-jmz/analisis\_acceso\_viviendas\_ecuador/blob/main/diccionario\_vivienda.html

Enlace del modelo físico clic [aquí](https://github.com/michael-jmz/analisis_acceso_viviendas_ecuador/blob/main/Digramas/Diagramas%20Modelo%20Fisico.svg) https://github.com/michael-jmz/analisis\_acceso\_viviendas\_ecuador/blob/main/Digramas/Diagramas%20Modelo%20Fisico.svg

**Figura 5**

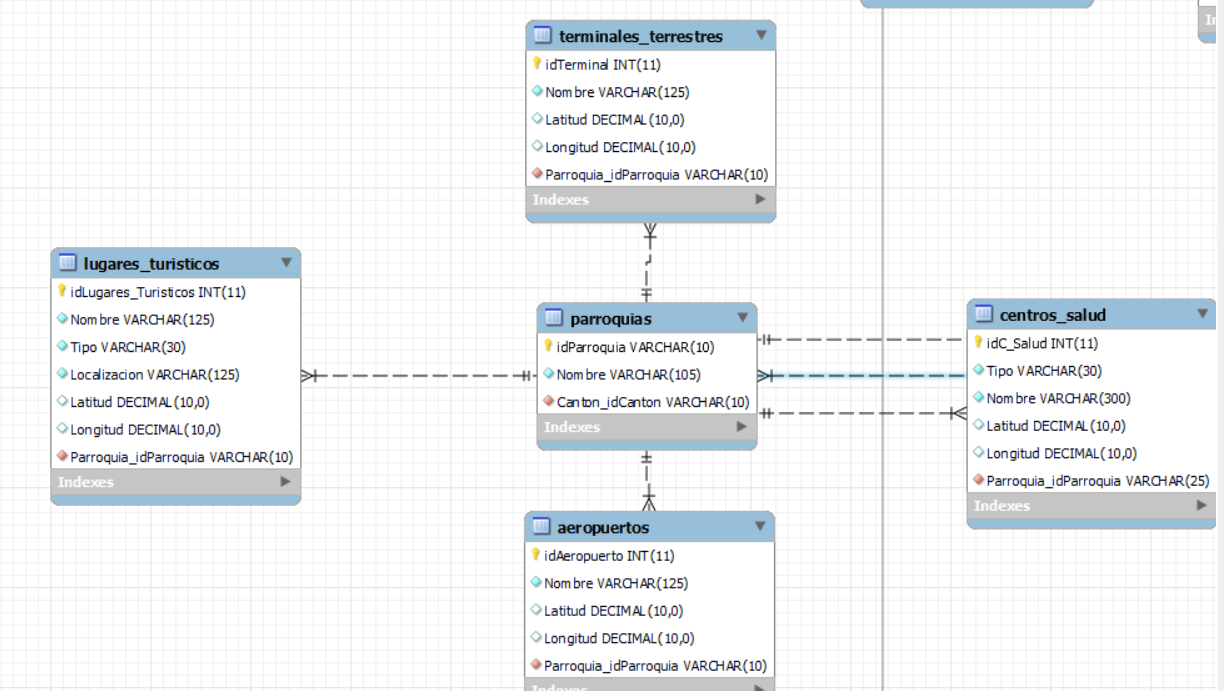
*Modelo general físico para implementación de Base de datos “Acceso y condiciones de vivienda INEC Primer Trimestre”*



**Nota:** Elaboración propia, Imagen Modelo físico he implementación de nueva data de los aeropuertos, terminales terrestres, centros de salud, lugares turísticos y población.

**Figura 6**

*Imagen de tablas nuevas relacionadas con los datos del INEC*

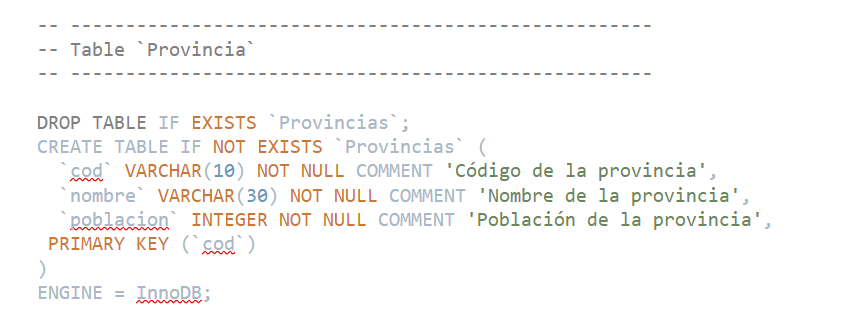


***Nota:*** Elaboración propia, Una ampliación del modelo donde se puede observar la inserción de nuevas tablas como aeropuertos, terminales, centros de salud y lugares turísticos.

# Implementación y carga

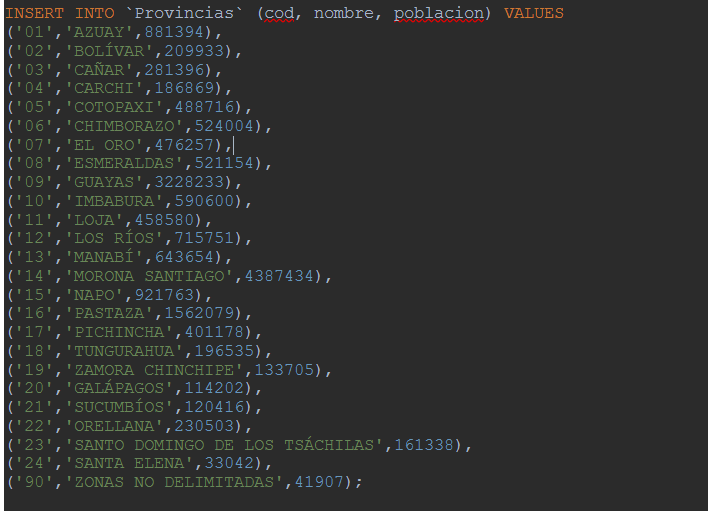
**Figura 7**

*Creación de tabla Provincias*



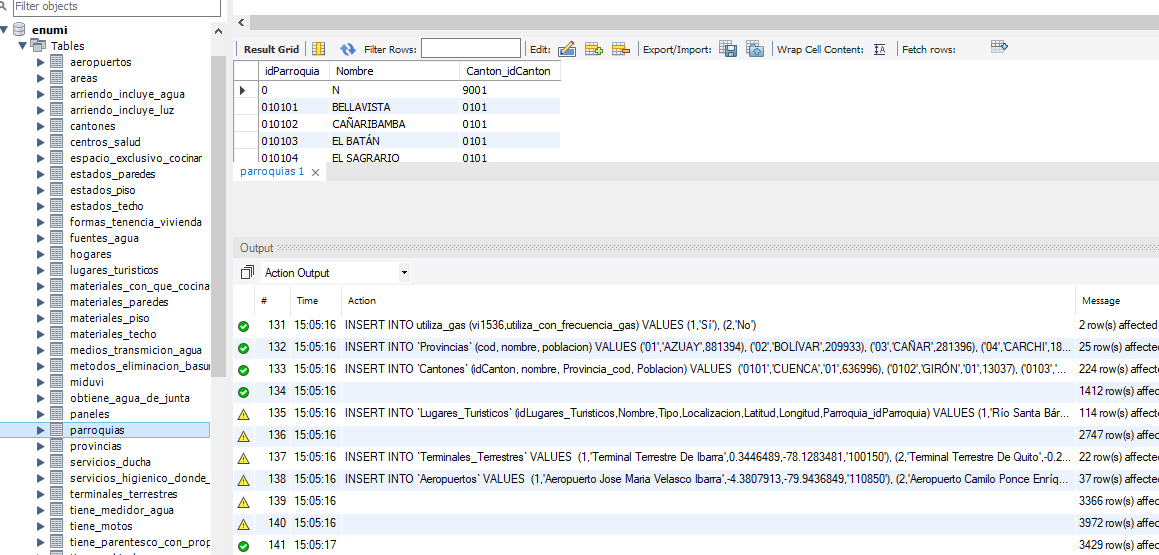
**Figura 8**

*Sentencia sql para insertar Provincias*



**Nota**: Elaboración propia, Se inserta en la tabla de provincias datos como código, nombre y la población.

**Figura 9**



**Nota:** Elaboración propia, En esta imagen podemos observar que se insertaron varios campos en el modelo donde se muestra una consulta de parroquias.

**Figura 10**

*Inserción de catálogos: Área, Paneles y Viviendas*

Texto

Descripción generada automáticamente

**Nota:** Elaboración propia, Se Inserta las tablas correspondientes a varios catálogos en base a la encuesta del INEC.

Figura 10.1

*Ejecución correcta de catálogos.*

  
**Nota:** Elaboración propia, En esta imagen se muestra los campos insertados de las tablas correspondiente a los catálogos de la encuesta del INEC.

Se inserta la tabla ***miduvi*** la que contiene la mayor cantidad de registros como se muestra en la captura algunos gropos de datos insertados exitosamente

INSERT INTO miduvi (area,ciudad,conglomerado,panelm,vivienda,hogar,vi01,vi02,vi03a,vi03b,vi04a,vi04b,vi05a,vi05b,vi06,vi07,vi07a,vi07b,vi08,vi09,vi09a,vi09b,vi10,vi101,vi102,vi10a,vi11,vi12,vi13,vi14,vi141,vi142,vi143,vi144,vi1511,vi1521,vi1512,vi1522,vi1531,vi1541,vi1532,vi1542,vi1533,vi1543,vi1534,vi1544,vi1535,vi1545,vi1536,vi1546,estrato,fexp,upm,id\_vivienda,id\_hogar,periodo,mes) VALUES   
(1,'010150','000201',26,1,1,1,1,2,3,4,2,1,2,4,3,0,1,1,1,null,null,1,null,null,1,1,1,2,4,150,null,null,null,2,null,2,null,null,null,null,null,null,null,null,null,null,null,null,null,2713,30.1109079661847,'010150000201','0101500002010260102','01015000020102601102','202302',2),   
(1,'010150','000201',26,2,1,1,2,6,1,2,1,1,1,3,2,0,1,1,1,null,null,1,null,null,1,1,1,2,1,100,1,1,1,2,null,2,null,null,null,null,null,null,null,null,null,null,null,null,null,2713,30.1109079661847,'010150000201','0101500002010260202','01015000020102602102','202302',2),   
(1,'010150','000201',26,3,1,1,2,1,1,1,1,1,1,3,2,0,1,1,1,null,null,1,null,null,1,1,1,2,1,200,2,2,2,2,null,2,null,null,null,null,null,null,null,null,null,null,null,null,null,2713,30.1109079661847,'010150000201','0101500002010260302','01015000020102603102','202302',2),   
(1,'010150','000201',26,5,1,1,1,4,3,1,3,1,2,6,5,0,1,1,1,null,null,1,null,null,1,1,1,2,4,800,null,null,null,2,null,2,null,null,null,null,null,null,null,null,null,null,null,null,null,2713,30.1109079661847,'010150000201','0101500002010260502','01015000020102605102','202302',2);

**Figura 11**

*inserción de datos en la tabla* ***miduvi***

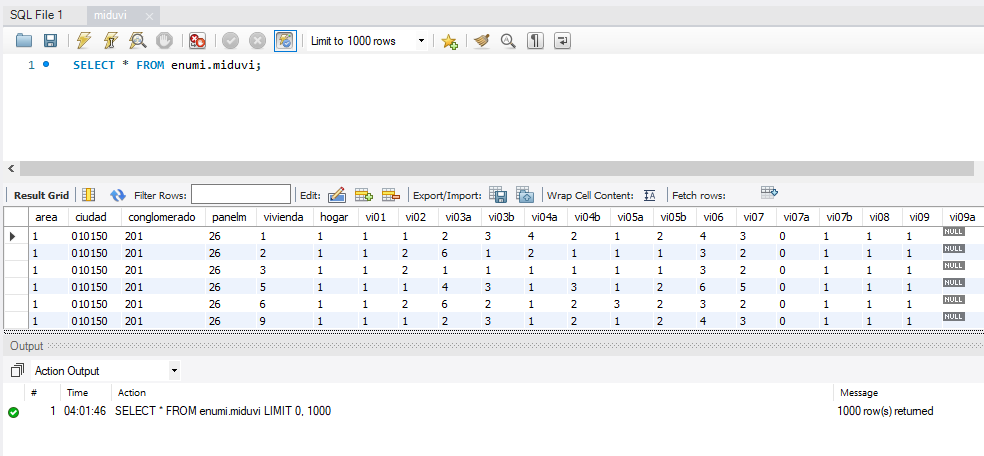
Interfaz de usuario gráfica

Descripción generada automáticamente

**Nota:** Elaboración propia, Al momento de cargar los datos surgió un error que no permitía insertar en la tabla ***miduvi*** debido a que unos id correspondientes a cuidad no eran válidos, pero finalmente se pudo corregir.

**Figura 12**

*Consulta a la tabla* ***miduvi***



**Nota:** Elaboración propia, Al realizar una consulta de selección para verificar que la existencia de datos obtuvimos

# Script SQL

Enlace del script para generar la base de datos clic [aquí](https://github.com/michael-jmz/analisis_acceso_viviendas_ecuador/blob/main/script_corregido.sql)

# <https://github.com/michael-jmz/analisis_acceso_viviendas_ecuador/blob/main/script_corregido.sql>

# Descarga en formato CSV

Código para obtener el csv

1. SELECT m.\*
2. p.Nombre, cn.nombre, cn.Poblacion, po.nombre, po.region, tm.Nombre, tm.Latitud, tm.Longitud, a.Nombre, a.Latitud, a.Longitud,
3. cs.Tipo, cs.Nombre, cs.Latitud, cs.Longitud, l.Nombre, l.Tipo, l.Localizacion, l.Latitud, l.Longitud
4. INTO OUTFILE 'C:\\ProgramData\\MySQL\\MySQL Server 8.0\\Uploads\\encuestas.csv'
5. FIELDS TERMINATED BY ',' OPTIONALLY ENCLOSED BY '"' ESCAPED BY '\\' LINES TERMINATED BY '\r\n'
6. FROM miduvi m
7. LEFT [JOIN](http://www.php.net/join) parroquias p ON p.idParroquia = c.Parroquia\_idParroquia
8. LEFT [JOIN](http://www.php.net/join) cantones cn ON p.Canton\_idCanton = cn.idCanton
9. LEFT [JOIN](http://www.php.net/join) provincias po ON cn.idCanton = po.Cod
10. LEFT [JOIN](http://www.php.net/join) terminales\_terrestres tm ON tm.Parroquia\_idParroquia = p.idParroquia
11. LEFT [JOIN](http://www.php.net/join) aeropuertos a ON a.Parroquia\_idParroquia = p.idParroquia
12. LEFT [JOIN](http://www.php.net/join) centros\_salud cs ON cs.Parroquia\_idParroquia = p.idParroquia
13. LEFT [JOIN](http://www.php.net/join) lugares\_turisticos l ON l.Parroquia\_idParroquia = p.idParroquia;

# Conclusiones

Se logró elaborar y plantear el objetivo propuesto en el proyecto integrador lo cual consiste en la limpieza de datos y la agregación de 5 tipos de variables más para tener una base de datos sin datos basura.

En él se abordaron los aspectos fundamentales para poder cumplir con el proyecto integrador, y poderlo poner a prueba lo plateado.

A través de un análisis se logró identificar las mejoras que podíamos hacer al momento de general las bases de datos para tener una base de datos bien estructurada sin ningún tipo de datos basura.

A modo de cierre, del proyecto integrador podremos afirmar que tenemos que ser más cuidadosos al buscar la data de las 5 variables a agregar y buscar la data en fuentes confiables. De lo contrario, contribuiremos al ingreso de datos basura, al punto de generarle un daño a la base de datos**.**

Recomendaciones**:**

Se recomienda principalmente hacer una limpieza de datos para que no haya datos basura y lo cual mejores el funcionamiento de la base de datos

# BIBLIOGRAFIA

INEC. (2020). *Marco Maestro de Muestreo SIEH*. https://www.ecuadorencifras.gob.ec/documentos/web-inec/Estadisticas\_Sociales/Violencia\_de\_genero\_2019/Marco%20Maestro%20Muestreo%20SIEH.pdf

INEC. (2023). *MANUAL DEL ENCUESTADOR/A ENCUESTA NACIONAL DE EMPLEO, DESEMPLEO Y SUBEMPLEO ENEMDU-ABRIL 2023 COORDINACIÓN GENERAL TÉCNICA DE PRODUCCIÓN ESTADÍSTICA*. https://www.ecuadorencifras.gob.ec/documentos/web-inec/EMPLEO/2023/Abril/202304\_Manual\_del\_encuestador\_ENEMDU\_Abril\_2023.pdf