



Proyecto Integrador

Integrantes:

Mateo Sebastián Martínez Velásquez

Eric Antonio Alvarado Ontaneda

Asignatura:

Base de datos avanzada

Docente:

Ing. Juan Carlos Morocho

Periodo académico

Abr/2023 - Ago/2023

Introducción

El análisis de datos sobre información de acceso y condiciones de vivienda en Ecuador es un proceso de examinar y evaluar los datos relacionados con la disponibilidad, calidad y materiales de la vivienda en el país. Este tipo de análisis proporciona una comprensión más profunda de los patrones y las disparidades existentes en el acceso a la vivienda, así como las condiciones en las que viven las personas y las fuentes de energía que las mismas poseen.

La vivienda es un aspecto fundamental en la vida de las personas, ya que proporciona un lugar seguro y digno para vivir. El acceso a una vivienda adecuada tiene un impacto directo en la calidad de vida de las personas y está estrechamente relacionado con otros factores socioeconómicos, como la educación, la salud y el bienestar general.

El análisis de datos sobre acceso y condiciones de vivienda en Ecuador es importante por varias razones. En primer lugar, permite identificar las necesidades y demandas de vivienda en diferentes áreas geográficas del país. Esto es crucial para la planificación y el desarrollo de políticas públicas efectivas en materia de vivienda, ya que los datos ayudan a comprender los problemas existentes y a diseñar soluciones adecuadas.

Además, el análisis de datos sobre vivienda también permite identificar desigualdades y brechas en el acceso a una vivienda adecuada dentro de las diferentes ciudades del Ecuador, todo esto basándonos en fuentes confiables y representativas de la población ecuatoriana como lo son los datos del INEC (Instituto Nacional de Estadística y Censos).

Repositorio Github

<https://github.com/eaalvarado8/BDAvanzada-ProyectoIntegrador>

Datos base

Los datos base utilizados en este análisis abarcan una amplia gama de variables relacionadas con el acceso y las condiciones de vivienda en Ecuador. Estos datos proporcionan información detallada sobre diversos aspectos de las viviendas y los hogares, lo que nos permite obtener una imagen completa de la situación habitacional en el país.

En grandes medidas las variables incluidas en los datos base se encuentran:

Área: Esta variable indica si los datos provienen de áreas urbanas o rurales, lo cual es importante para comprender las diferencias en el acceso y las condiciones de vivienda en diferentes entornos geográficos.

Ciudad y conglomerado: Estas variables identifican la ciudad y el conglomerado al que pertenece cada vivienda, lo que permite realizar análisis específicos según la ubicación geográfica.

Identificadores de vivienda y hogar: Estos identificadores únicos permiten distinguir cada vivienda y hogar en el conjunto de datos, lo que facilita el seguimiento y análisis individualizado.

Características de la vivienda y del hogar: Se recopila información detallada sobre la vivienda y el hogar, como el tipo de vivienda, el material utilizado en el techo, las paredes y el piso, el número de cuartos y dormitorios, y si la vivienda cuenta con espacio para cocinar.

Condiciones sanitarias: Las variables relacionadas con las instalaciones sanitarias y el suministro de agua proporcionan información sobre el acceso a servicios higiénicos, la fuente de agua utilizada en la vivienda y si la vivienda cuenta con servicios de ducha.

Tenencia de la vivienda: Se registran datos sobre la forma de tenencia de la vivienda, ya sea propia, alquilada u otra forma de tenencia, así como el valor mensual del arriendo, si aplica.

Energía y combustible: Se recopila información sobre el tipo de alumbrado utilizado en la vivienda y las frecuencias de abastecimiento y gastos mensuales en combustible para vehículos y

diferentes tipos de combustibles, como super, extra, diesel, Ecopais, electricidad y gas.

Periodo: Estas variables son importantes para saber en qué momento se llevó a cabo la encuesta y mediante esto poder realizar diferentes análisis.

Datos complementarios

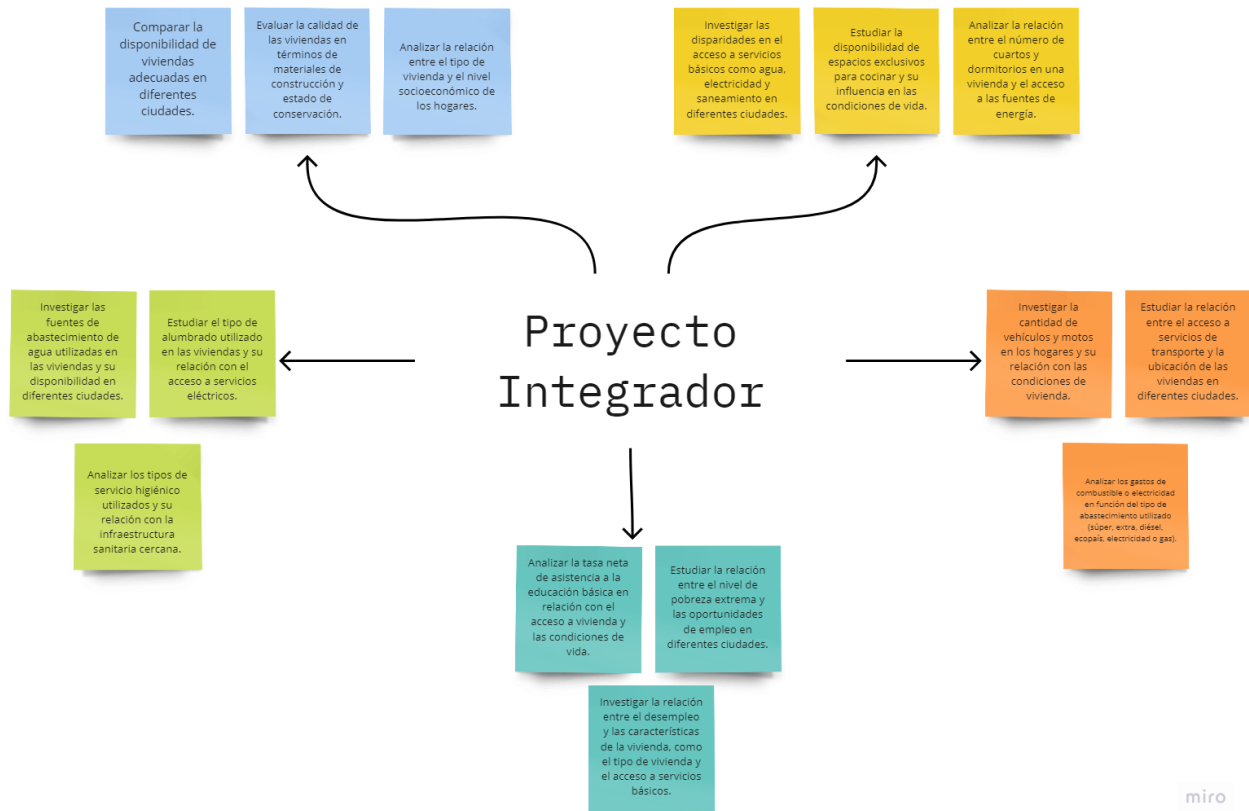
Indicador de desempleo: Este indicador muestra la proporción de personas desempleadas en relación con la población económicamente activa. Es un dato importante para comprender la situación laboral y el nivel de empleo en la población estudiada. El desempleo puede tener un impacto significativo en las condiciones de vida y el acceso a recursos económicos de los hogares.

Tasa neta de asistencias Educación General Básica: Esta variable representa la proporción de niños en edad escolar que asisten a la educación general básica en relación con el total de la población en esa edad. La educación es un factor clave en el desarrollo individual y social, y la tasa de asistencia escolar puede proporcionar información sobre el acceso a la educación y el nivel de desarrollo educativo de la población.

Indicador Pobreza extrema por ingresos: Este indicador refleja la proporción de la población que vive en situación de pobreza extrema, medida a través de la falta de ingresos suficientes para cubrir las necesidades básicas. La pobreza extrema por ingresos es un indicador importante para evaluar el nivel de bienestar económico y social de la población, y puede ayudar a identificar áreas con mayores necesidades y desigualdades socioeconómicas.

Posibles análisis para realizar

En base a la información que se tiene por parte de las encuestas y la información obtenida, se puede evidenciar los siguientes casos en los que se debe tomar énfasis en la siguiente información:



Algunas otras opciones de consultas que se pueden realizar

- Ciudades donde gran parte de su población posee viviendas con el peor tipo de material.
- Ciudades donde gran parte de su población posee viviendas con el mejor tipo de material.
- El material más popular con el que está echas las viviendas
- El combustible más usado en las viviendas del Ecuador
- Tipo de material de las viviendas en las ciudades con mayor tasa de empleabilidad

Depuración y Análisis de datos:

Análisis de datos: Al analizar los datos no sé tubo mayores inconvenientes a parte de que al abrirlas algunas veces con Excel y al no importar algunos datos del csv con texto, excel los ponía en cierto forma que luego no servía para un análisis certero, es por esto que se tubo que volver a descargar dichos datos y volverlos a analizar directamente con el el bloc de notas y con libre office.

Primeramente se realizo un analisis de las columans que tenian nombres clave como vi01, entre otras similares para poder ponerle nombres más descriptivos, para esto se realizo un analisis del diccionario de datos.

Variable	NuevoNombre	Descripción	Opciones de respuesta
vi01	ViaAccesoPrincipalVivienda	Vía de acceso principal a la vivienda	1: Carretera, calle pavimentada;2: Empedrado;3: Lastrado, calle de tierra;4: Sendero;5: Río, mar;6: Otro
vi02	TipoVivienda	Tipo de vivienda	1: Casa o villa;2: Departamento;3: Cuartos en casa de inquilinato;4: Mediagua;5: Rancho, covacha;6: Choza;7: Otra
vi03a	MaterialTechoCubierta	Material del techo o cubierta	1: Hormigón (losa, cemento);2: Fibrocemento, asbesto (eternit, eurolit);3: Zinc, Aluminio;4: Teja;5: Palma, paja u hoja;6: Otro Material
vi03b	EstadoTecho	Estado del techo	1: Bueno;2: Regular;3: Malo
vi04a	MaterialPiso	Material del piso	1: Duela, parquet, tablón tratado o piso flotante;2: Cerámica, baldosa, vinil o porcelanato;3: Mármol o marmetón;4: Ladrillo o cemento;5: Tabla / tablón no tratado;6: Caña;7: Tierra;8: Otro Material
vi04b	EstadoPiso	Estado del piso	1: Bueno;2: Regular;3: Malo
vi05a	MaterialParedes	Material de las paredes	1: Hormigón/Ladrillo o Bloque;2: Asbesto/Cemento (Fibrolit);3: Adobe o Tapia;4: Madera;5: Caña revestida o bahareque;6: Caña no revestida o estera;7: Otra Material
vi05b	EstadoParedes	Estado de las paredes	1: Bueno;2: Regular;3: Malo
vi06	NumCuartos	Número de cuartos	No informa (99)
vi07	NumDormitorios	Número de dormitorios	No informa (99)
vi07a	NumCuartosNegocio	Número de cuartos para negocio	No informa (99)
vi07b	EspacioCocinar	Dispone este hogar de un cuarto o espacio exclu	1: Si;2: No
vi08	MaterialCocina	Material con que cocinan	1: Gas;2: Leña, carbón;3: Electricidad;4: Otro
vi09	TipoServicioHigienico	Tipo de servicio higiénico	1: Excusado y alcantarillado;2: Excusado y pozo séptico;3: Excusado y pozo ciego;4: Letrina;5: No tiene
vi09a	MiembrosSinServicioHigienico	Puesto que mencionó que NO TIENE Servicio Hig	1: Descarga directa al mar, río, lago o quebrada;2: Van al monte, campo, bota la basura en paquete;3: Usan
vi09b	InstalacionSanitariaPrincipal	Qué tipo de instalación sanitaria cercana y/o pr	1: Excusado y alcantarillado;2: Excusado y pozo séptico;3: Excusado y pozo ciego;4: Letrina
vi10	FuenteAgua	De dónde obtiene el agua	1: Red pública;2: Pila o llave pública;3: Otra fuente por tubería;4: Carro repartidor, triciclo;5: Pozo;6: Río, vertiente, acequia;7: Otro
vi101	MedidorAgua	Tiene medidor de agua	1: Si;2: No
vi102	AguaJuntaAgua	El agua que obtiene es de la junta de agua	1: Si;2: No
vi10a	TipoAguaVivienda	El agua que recibe la vivienda es	1: Por tubería dentro de la vivienda;2: Por tubería fuera de la vivienda pero en el lote;3: Por tubería fuera de la vivienda, lote o terreno;4
vi11	ServicioDucha	El servicio de ducha es	1: Exclusivo del hogar;2: Compartido con otros hogares;3: No tiene
vi12	TipoAlumbrado	Tipo de alumbrado	1: Empresa eléctrica pública;2: Planta eléctrica privada;3: Vela, candil, mechero, gas;4: Ninguno

Luego de esto se procedió analizar los catálogos respecto al diccionario de los datos cifrados

ViaAccesoPrincipalVivienda		EstadoTecho	
ID	Opción	ID	Opción
	1 Carretera, calle pavimentada		1 Bueno
	2 Empedrado		2 Regular
	3 Lastrado, calle de tierra		3 Malo
	4 Sendero		
	5 Río, mar		
	6 Otro		
		MaterialPiso	
ID	Opción	ID	Opción
	1 Casa o villa		1 Duela, parquet, tablón tratado o piso flotante
	2 Departamento		2 Cerámica, baldosa, vinil o porcelanato
	3 Cuartos en casa de inquilinato		3 Mármol o marmetón
	4 Mediagua		4 Ladrillo o cemento
	5 Rancho, covacha		5 Tabla / tablón no tratado
	6 Choza		6 Caña
	7 Otra		7 Tierra
			8 Otro Material

MaterialPiso	
ID	Opción
1	Duela, parquet, tablón tratado o piso flotante
2	Cerámica, baldosa, vinil o porcelanato
3	Mármol o marmetón
4	Ladrillo o cemento
5	Tabla / tablón no tratado
6	Caña
7	Tierra
8	Otro Material

EstadoPiso	
ID	Opción
1	Bueno
2	Regular
3	Malo

MaterialCocina	
ID	Opción
1	Gas
2	Leña, carbón
3	Electricidad
4	Otro

TipoServicioHigienico	
ID	Opción
1	Excusado y alcantarillado
2	Excusado y pozo séptico
3	Excusado y pozo ciego
4	Letrina
5	No tiene

MiembrosSinServicioHigienico	
ID	Opción
1	Descarga directa al mar, río, lago o quebrada
2	Van al monte, campo, bota la basura en paquete
3	Usan una instalación sanitaria cercana y/o prestada

InstalacionSanitariaPrincipal	
ID	Opción
1	Excusado y alcantarillado
2	Excusado y pozo séptico
3	Excusado y pozo ciego
4	Letrina

FuenteAgua	
ID	Opción
1	Red pública
2	Pila o llave pública
3	Otra fuente por tubería
4	Carro repartidor, triciclo
5	Pozo
6	Río, vertiente, acequia

MaterialParedes	
ID	Opción
1	Hormigón/Ladrillo o Bloque
2	Asbesto/Cemento (Fibrolit)
3	Adobe o Tapia
4	Madera
5	Caña revestida o bahareque
6	Caña no revestida o estera
7	Otra Material

EstadoParedes	
ID	Opción
1	Bueno
2	Regular
3	Malo

EliminacionBasura	
ID	Opción
1	Contratan el servicio
2	Servicio municipal
3	Botan a la calle, quebrada, río
4	La queman, entierran
5	Otra

FormaTenenciaVivienda	
ID	Opción
1	En arriendo
2	Anticresis y/o arriendo
3	Propia y la está pagando
4	Propia y totalmente pagada
5	Cedida
6	Recibida por servicios
7	Otra

ServicioDucha	
ID	Opción
1	Exclusivo del hogar
2	Compartido con otros hogares
3	No tiene

TipoAlumbrado	
ID	Opción
1	Empresa eléctrica pública
2	Planta eléctrica privada
3	Vela, candil, mechero, gas
4	Ninguno

De igual manera se analizo los datos y se vio de estos cuales podian ser representados solamente como un valor booleano para no generar un catalogo con estos datos

NombreVariable	Opciones de respuesta
EspacioCocinar	1: Sí; 2: No
MedidorAgua	1: Sí; 2: No
AguaJuntaAgua	1: Sí; 2: No
ServicioDucha	1: Sí; 2: No
IncluyePagoAguaArriendo	1: Sí; 2: No
IncluyePagoLuzArriendo	1: Sí; 2: No
RelacionParentescoPropietario	1: Sí; 2: No
TieneVehiculos	1: Sí; 2: No
TieneMotos	1: Sí; 2: No
AbastecimientoFrecuenciaSuper	1: Sí; 2: No
AbastecimientoFrecuenciaExtra	1: Sí; 2: No
AbastecimientoFrecuenciaDiesel	1: Sí; 2: No
AbastecimientoFrecuenciaEcopais	1: Sí; 2: No
AbastecimientoFrecuenciaElectricidad	1: Sí; 2: No
AbastecimientoFrecuenciaGas	1: Sí; 2: No

En base al analisis pudimos concluir que los hogares le pertenecen a una vivienda, siendo esta una relacion 1:m y que a su vez hay algunas columnas las cuales o se asocian directamente a la vivienda o al hogar, siendo la mayoria una relacion de 1:1, para realizar este análisis cambiamos los valores de las variables cifradas como vi01 para facilitar la creación de tablas.

vi01	vi02	vi03a	vi03b	vi04a	vi04b	vi05a	vi05b	vi06	vi07	
ViaAccesoPrin	TipoVivienda	MaterialTech	EstadoTecho	MaterialPiso	EstadoPiso	MaterialPared	EstadoPared	NumCuartos	NumDormitor	
vi102	vi10a	vi11	vi12	vi13	vi14	vi141	vi142	vi143	vi144	vi145
AguaJuntaAgu	TipoAguaVivie	ServicioDucha	TipoAlumbrac	EliminacionBa	FormaTenenci	ValorArriendo	IncluyePagoA	IncluyePagoLu	RelacionParen	TieneVehiculos

Concluimos que la unica tabla de m:m que existe es la de el combustible que usa el hogar, pues dentro de combustible en los datos se clafica en super, extra, diesel, ecopais pero tambien cuenta como combustible gas y electricida por lo concluimos que un hogar puedes tener diferentes fuentes de combustible y un combustible le puede pertenecer a varios hogares.

Con respecto a los catálogos de los atributos que no estaban cifrados en el diccionario pudimos analizar los siguientes.

	periodo
ID	valor
202301	ene-23
202302	feb-23
202303	mar-23

area	
ID	valor
1	Urbana
2	Rural

vivienda_num	
ID	numVivienda
10	Vivienda Diez (reemplazo)
1	Vivienda Uno
2	Vivienda Dos
3	Vivienda Tres
4	Vivienda Cuatro
5	Vivienda Cinco
6	Vivienda Seis
7	Vivienda Siete
8	Vivienda Ocho (reemplazo)
9	Vivienda Nueve (reemplazo)

	combustible
ID	combustible
1	Super
2	Extra
3	Diesel
4	Ecopais
5	Electricidad
6	Gas

	ciudad
ID	nombre
10150	Cuenca
10151	Baños
10152	Cumbe
10153	Chaucha
10154	Checa (Jidcay)
10155	Chiquintad
10156	Llacao
10157	Molleturo
10158	Nulti
10159	Octavio Cordero Palacios
10160	Paccha

Depuración: En la depuración de los datos, se encontraron algunos valores correspondientes al catálogo de Ciudades, donde habían códigos que no se encontraban en fuentes de datos como el INEC, para esto se realizó una indagación donde se encontraron más valores correspondientes a estos códigos:

100300	COTACACHI
100350	COTACACHI
130400	EL CARMEN
130450	EL CARMEN
190352	Nuevo Paraiso

Los valores presentados también eran usados en las tablas de hogares/viviendas, por lo que al crear la tabla se presentó un error referencial, puesto que en el primer catálogo de ciudades implementados, estos códigos no se encontraban presentes, la siguiente caputra presenta el lugar donde estaban siendo usados esos datos.

	area	ciudad	conglomerado	panelm	vivienda	hogar	vi01	vi02	vi03a	vi03b	vi04a	vi04b
13182	2	100300	204	34	8	1	3	1	3	2	4	
13183	2	100300	603	25	1	1	3	1	3	2	4	
13184	2	100300	603	25	2	1	3	1	2	3	4	
13185	2	100300	603	25	3	1	3	1	3	2	4	
13186	2	100300	603	25	4	1	3	1	3	2	4	
13187	2	100300	603	25	5	1	3	1	3	2	4	

	area	ciudad	conglomerado	panelm	vivienda	hogar	vi01	vi02	vi03a	vi03b	vi04a	vi04b
13196	1	100350	1102	37	1	1	3	1	1	1	2	
13197	1	100350	1102	37	2	1	1	1	2	1	4	
13198	1	100350	1102	37	3	1	1	2	1	2	2	
13199	1	100350	1102	37	4	1	1	1	4	2	4	
13200	1	100350	1102	37	5	1	1	1	1	2	2	
13201	1	100350	1102	37	6	1	1	1	2	2	4	
13202	1	100350	1102	37	7	1	1	1	1	1	2	
13203	1	100350	1601	31	1	1	1	1	1	1	2	
13204	1	100350	1601	31	2	1	1	2	1	1	2	
13205	1	100350	1601	31	3	1	1	1	2	1	2	
13206	1	100350	1601	31	5	1	1	1	1	1	1	

	area	ciudad	conglomerado	panelm	vivienda	hogar	vi01	vi02	vi03a	vi03b	vi04a	vi04b
15399	1	130450	1504	31	1	1	3	2	1	2	2	
15400	1	130450	1504	31	2	1	1	2	3	1	4	
15401	1	130450	1504	31	3	1	3	2	3	2	4	
15402	1	130450	1504	31	4	1	1	1	3	2	4	
15403	1	130450	1504	31	5	1	1	1	3	1	4	
15404	1	130450	1504	31	6	1	3	1	3	2	4	
15405	1	130450	1504	31	7	1	1	2	1	1	2	
15406	1	130450	2902	34	1	1	1	1	3	2	4	
15407	1	130450	2902	34	2	1	1	1	3	2	2	
15408	1	130450	2902	34	3	1	1	2	3	3	4	

Como se pudo evidenciar se estaba usando esos códigos de ciudad, en la implementación de la

solución automática por código se pudo evidenciar errores referenciales, ya que al intentar encontrar esos códigos para dichas ciudades no era capaz de encontrarlos, una vez solucionados esos errores, un código de ciudad que estaba siendo usado en los datos de las tablas no se encontró por ningún lado, por lo que se llegó a la resolución de reemplazar ese código por “No informa” en el código Python:

999999

No informa

```
# Inserta los datos en la tabla "Vivienda"
cursor.execute('INSERT IGNORE INTO vivienda (idVivienda,conglomerado,area,numero,ciudad_idCiudad,periodo_idPeriodo,ViaAccesoPrincipalVivienda_idViaAccesoPrincipalVivienda)
              (fila['id_vivienda'],
               None if fila['conglomerado'] == '' else fila['conglomerado'],
               'u' if fila['area'] == '1' else 'r',
               None if fila['vivienda'] == '' else fila['vivienda'],
               '999999' if fila['ciudad'] == '230200' else fila['ciudad'],
               None if fila['periodo'] == '' else fila['periodo'],
               None if fila['ViaAccesoPrincipalVivienda'] == '' else fila['ViaAccesoPrincipalVivienda'],
               None if fila['TipoVivienda'] == '' else fila['TipoVivienda'],
               )
        )
```

25584	1	230200	1701	41	1	1	3	1	1	1	2	1
25585	1	230200	1701	41	2	1	1	1	3	3	4	2
25586	1	230200	1701	41	3	1	1	1	3	2	4	2
25587	1	230200	1701	41	4	1	3	1	3	1	2	1
25588	1	230200	1701	41	7	1	3	1	3	2	2	1
25589	1	230200	1701	41	8	1	3	1	3	1	4	1
25590	1	230200	1701	41	9	1	3	1	3	2	4	2
25591	1	230200	2001	27	1	1	3	1	3	2	5	2
25592	1	230200	2001	27	2	1	1	1	3	2	2	1

Dentro de la información de la documentación otorgada, se eliminó la siguiente información por el poco uso o irrelevancia de esta:

	TieneVehi	NumVehic	TieneMoto
	2		2
	2		2
	2		2
	2		2
	2		2
	2		2
	2		2
	2		2
	1	1	2
	2		2
	2		2
	2		2
	1	1	1
	2		2
	2		2
	2		2
	2		2

	D	E
ado	panelm	vivienda
201	26	1
201	26	2
201	26	3
201	26	4
201	26	5
201	26	6

estrato	fexp	upm	id_vivienda	id_hogar	periodo	mes
2713	30,1109079661847	1.015E+10	0101500002010260102	01015000020102601102	202302	2
2713	30,1109079661847	1.015E+10	0101500002010260202	01015000020102602102	202302	2
2713	30,1109079661847	1.015E+10	0101500002010260302	01015000020102603102	202302	2
2713	30,1109079661847	1.015E+10	0101500002010260402	01015000020102604102	202302	2
2713	30,1109079661847	1.015E+10	0101500002010260502	01015000020102605102	202302	2
2713	30,1109079661847	1.015E+10	0101500002010260602	01015000020102606102	202302	2
2713	30,1109079661847	1.015E+10	0101500002010260902	01015000020102609102	202302	2
2712	64,008634100334	1.015E+10	0101500003040420103	01015000030404201103	202303	3
2712	64,008634100334	1.015E+10	0101500003040420203	01015000030404202103	202303	3
2712	64,008634100334	1.015E+10	0101500003040420303	01015000030404203103	202303	3
2712	64,008634100334	1.015E+10	0101500003040420403	01015000030404204103	202303	3

La información eliminada se lo realizó en base a lo siguiente:

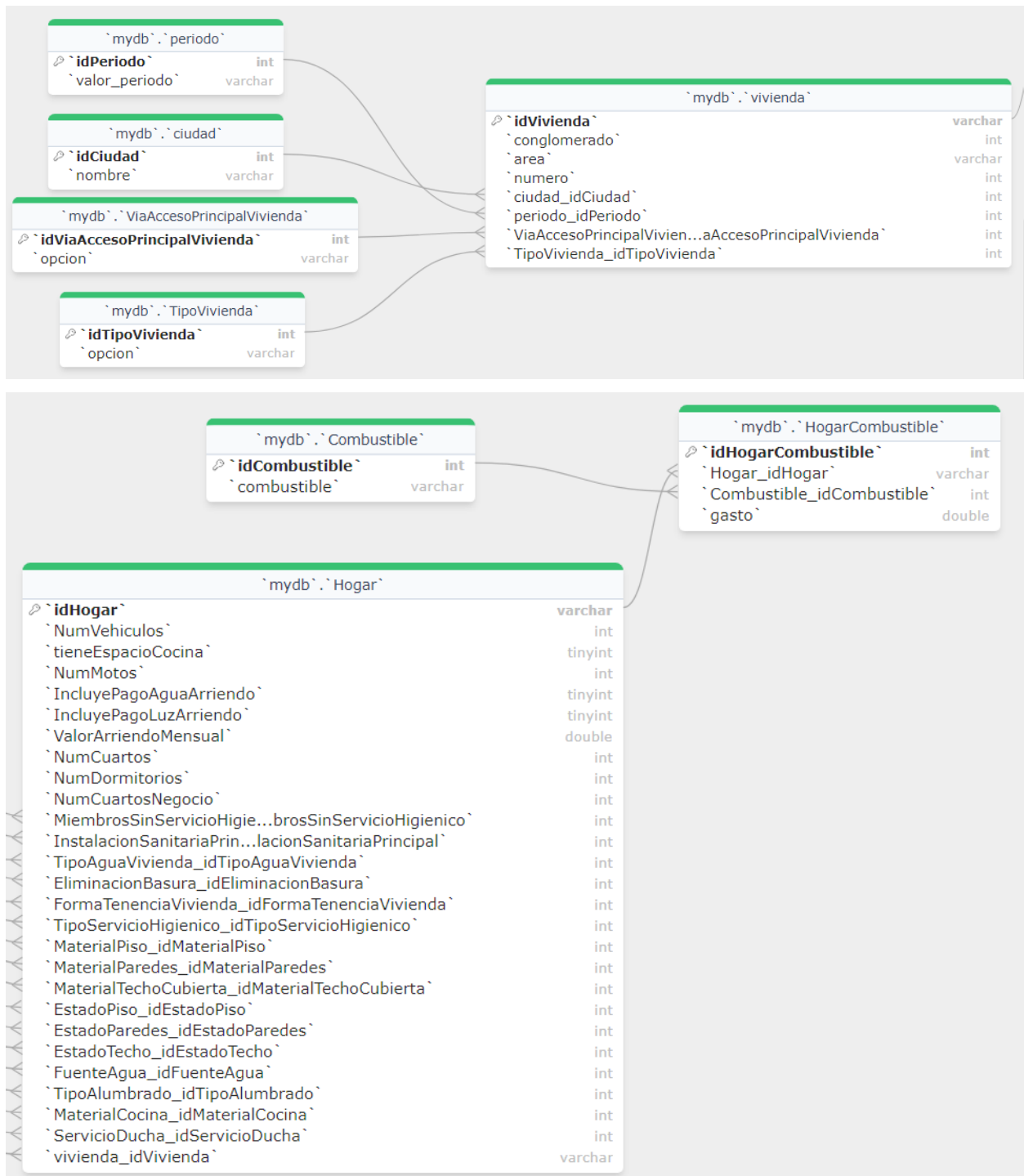
estrato	Es un parámetro que es de uso interno del INEC
upm	Nos indica unidad primaria de muestreo, con lo cual no realizaremos ningún análisis
panelm	Es un grupo de cuatro viviendas (tres originales y una de reemplazo) que tienen características similares, cada uno de los cuales permanece en la muestra por 4 trimestres en dos años consecutivos, por lo que no lo consideramos relevante dentro de los análisis a realizar.
mes	Datos del diccionario innecesarios al venir integrados en el periodo
TieneVehiculos	Solamente nos informa si el hogar tiene vehículos por lo que se lo elimino al poder deducirlo con la columna de NumVehiculos
TieneMotos	Solamente nos informa si el hogar tiene motos por lo que se lo elimino al poder deducirlo con la columna de NumMotos

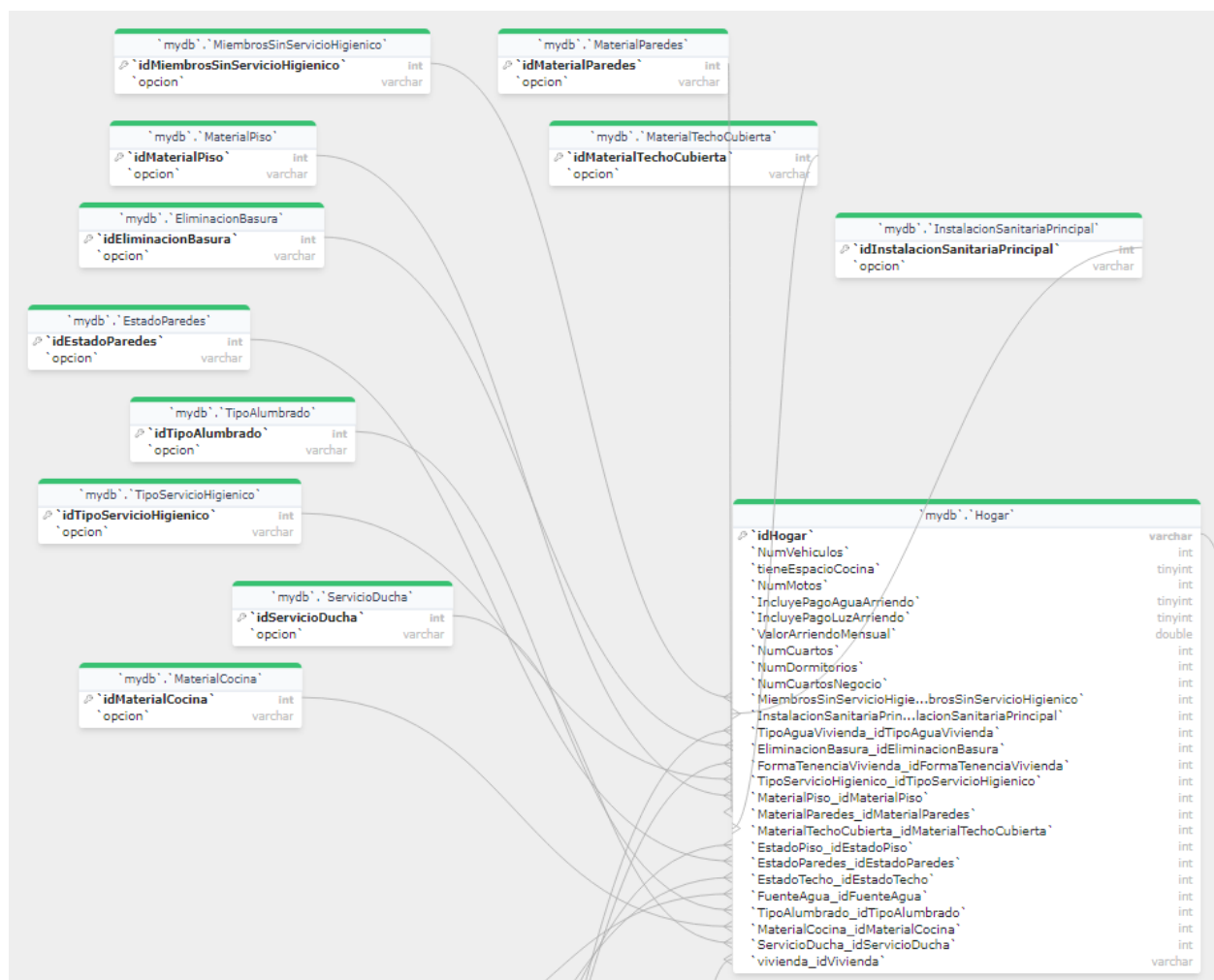
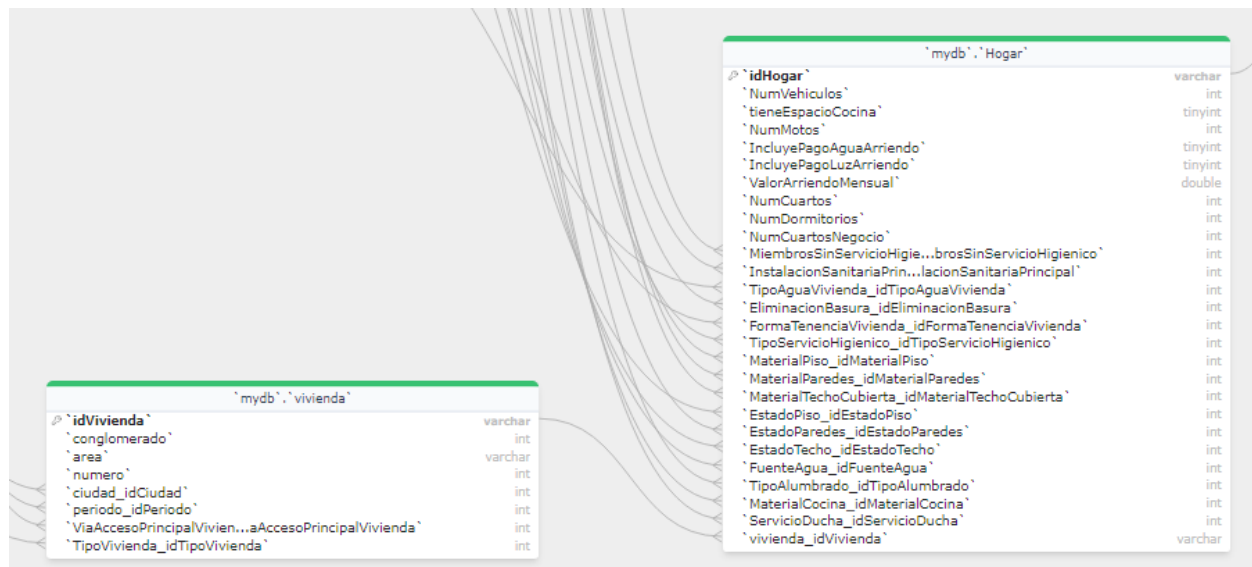
Tablas correspondientes a la información adicional añadida al caso de estudio:

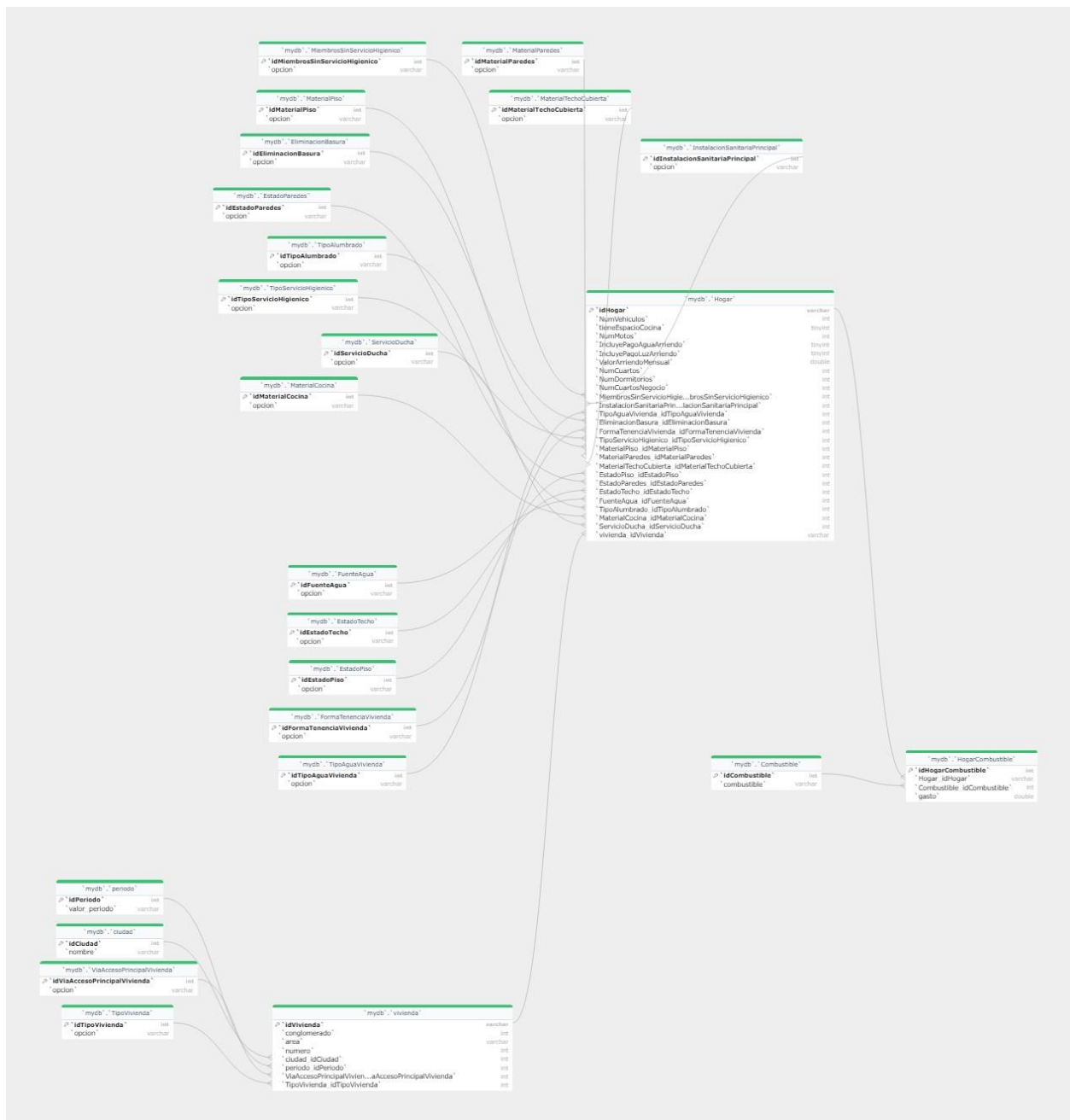
Provincias	Indicador de desempleo 2022	Tasa neta de asistencias Educación General Básica 2022	Indicador Pobreza extrema por ingresos 2022
Azuay	3.30%	96.0%	3.0%
Bolívar	0.90%	95.6%	10.1%
Cañar	3.70%	94.5%	5.2%
Carchi	5.90%	96.1%	14.6%
Cotopaxi	1.70%	97.0%	10.3%
Chimborazo	2.00%	94.0%	13.4%
El Oro	6.00%	94.6%	3.6%
Esmeraldas	9.10%	93.3%	25.3%
Guayas	3.60%	94.0%	5.1%
Imbabura	7.10%	95.2%	11.8%
Loja	3.10%	94.4%	12.5%
Los Ríos	1.80%	92.5%	6.9%
Manabí	2.80%	95.2%	7.8%
Morona Santiago	1.30%	95.9%	49.0%
Napo	1.50%	92.9%	44.5%
Pastaza	1.60%	91.9%	44.5%
Pichincha	8.50%	95.8%	5.7%
Tungurahua	2.20%	95.1%	6.3%
Zamora Chinchipe	3.00%	95.3%	14.2%
Galápagos	5.60%	96.3%	
Sucumbíos	4.50%	97.6%	36.6%
Orellana	1.30%	94.3%	37.7%
Santo Domingo	2.30%	94.4%	6.2%
Santa Elena	2.70%	94.8%	6.5%

Diseño conceptual

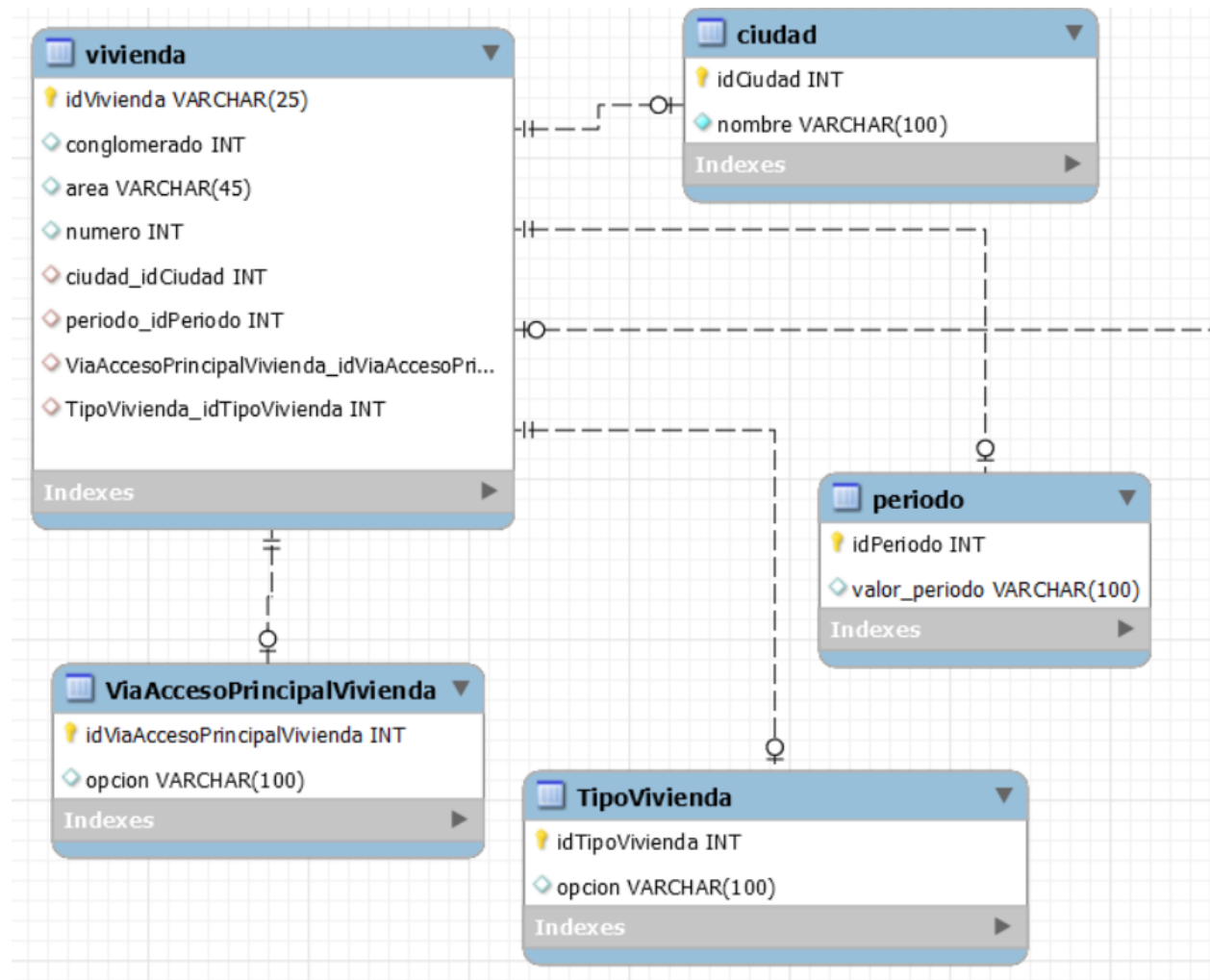
Se procedió a realizar el diagrama conceptual

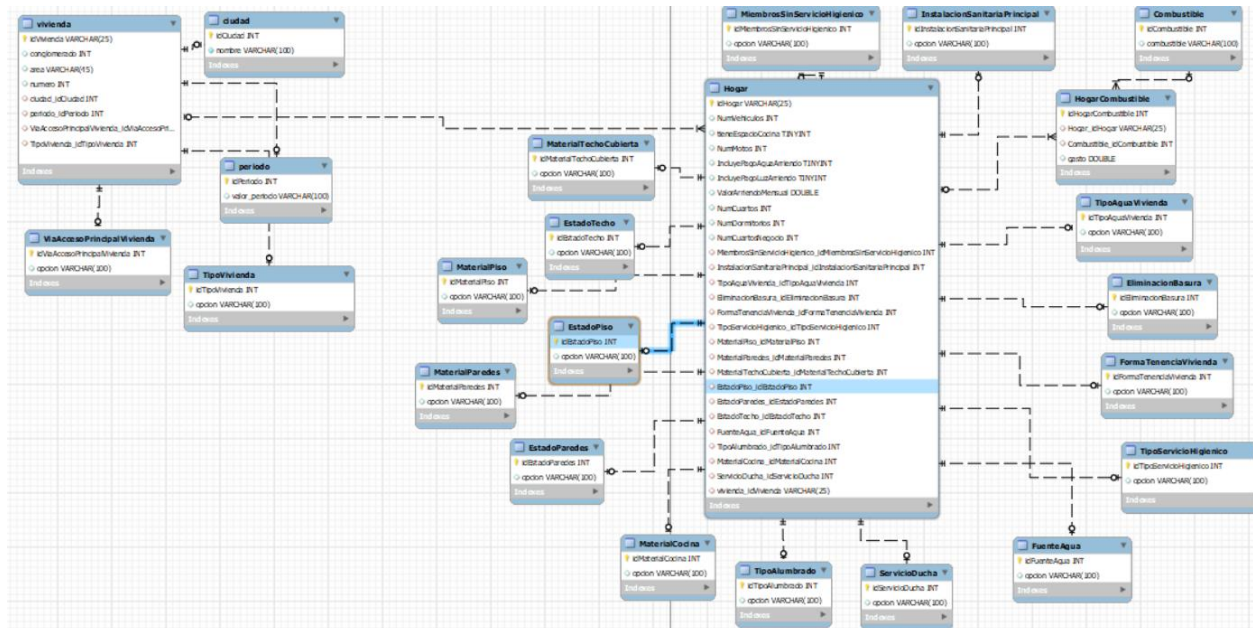
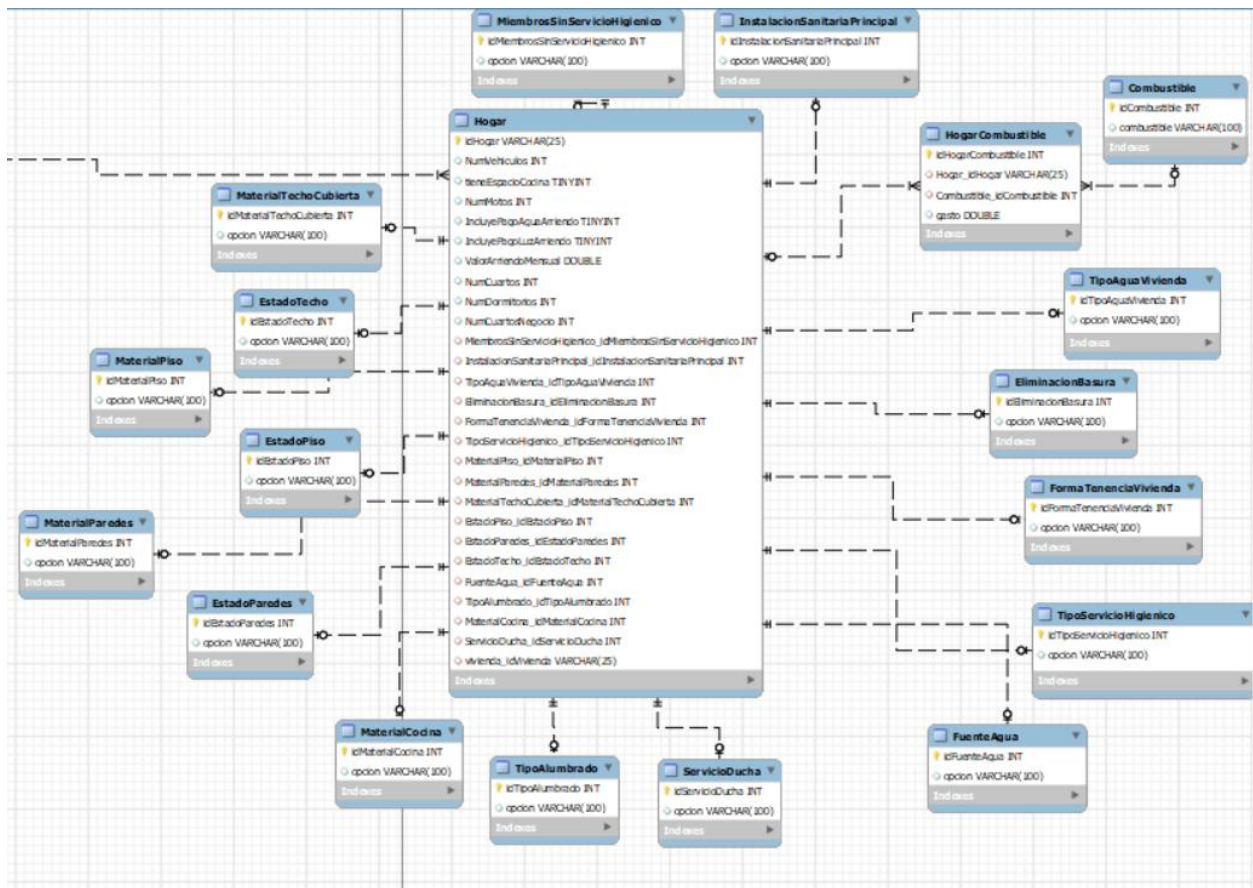


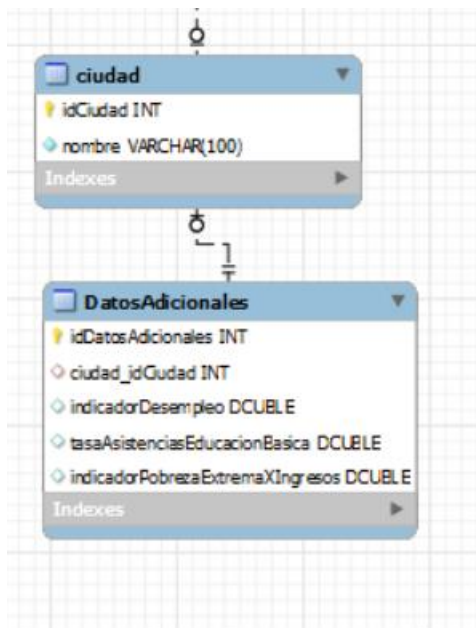




Diseño lógico relacional

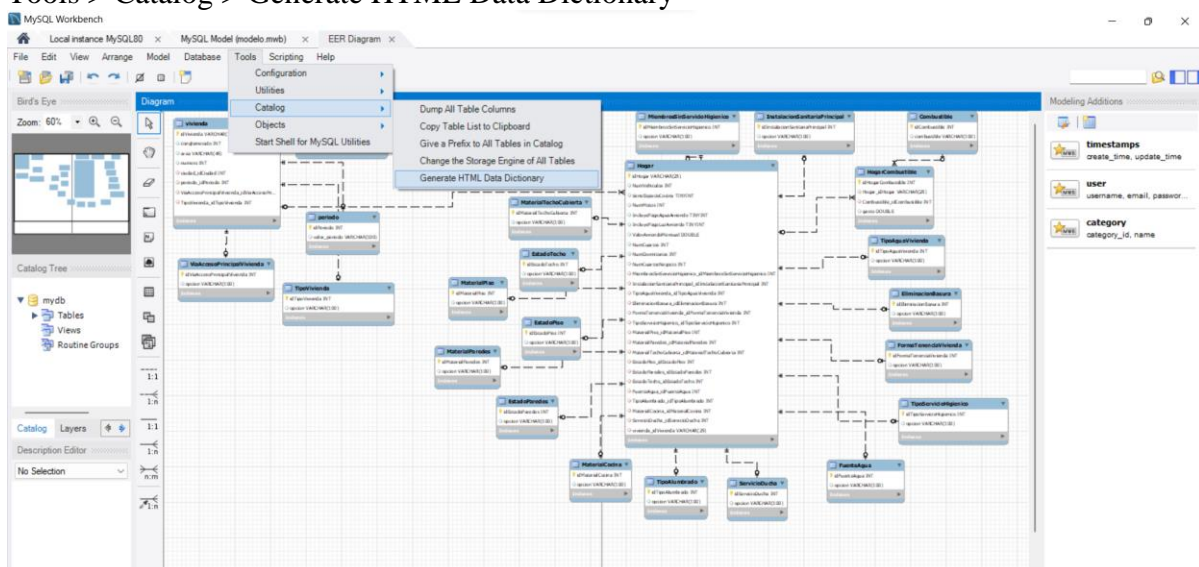




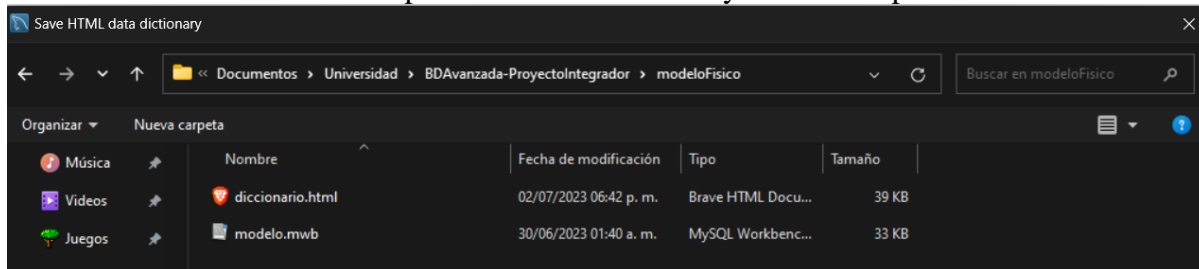


Creación del diccionario

Para poder generar los diccionarios en base al modelo de entidad relación solamente debemos instalar el plugin necesario, luego de eso ingresamos a nuestro modelo vamos a la opción de Tools > Catalog > Generate HTML Data Dictionary



Seleccionamos la ruta deseada para tener el diccionario y finalmente podremos visualizarlo



- | | | | | | | | |
|---------------------------------|---------------------------------|---------------------------------------|------------------------------|--|------------------------|-------------------------|--------------------------|
| TipoAguTivienda | TipoAguInmbrado | TipoServicioHigiénico | TipoVivienda | ViaAccesoPrincipalVivienda | ciudad | período | vivienda |
|---------------------------------|---------------------------------|---------------------------------------|------------------------------|--|------------------------|-------------------------|--------------------------|

Combustible												
Column name	Data Type	PK	NN	UQ	BIN	UN	ZF	AI	Default	Comment		
idCombustible	INT	✓	✓							Identificador del combustible		
combustible	VARCHAR(100)									Tipo de combustible		

EliminacionBasura												
Column name	Data Type	PK	NN	UQ	BIN	UN	ZF	AI	Default	Comment		
idEliminacionBasura	INT	✓	✓							Identificador del tipo de eliminación de basura		
opcion	VARCHAR(100)									Opciones sobre el tipo de eliminación de basura		

EstadoParedes												
Column name	Data Type	PK	NN	UQ	BIN	UN	ZF	AI	Default	Comment		
idEstadoParedes	INT	✓	✓							Identificador del estado de las paredes		
opcion	VARCHAR(100)									Opciones del estado de las paredes		

EstadoPiso												
Column name	Data Type	PK	NN	UQ	BIN	UN	ZF	AI	Default	Comment		
idEstadoPiso	INT	✓	✓							Identificador del estado del piso		
opcion	VARCHAR(100)									Identificador de las opciones de estado del piso		

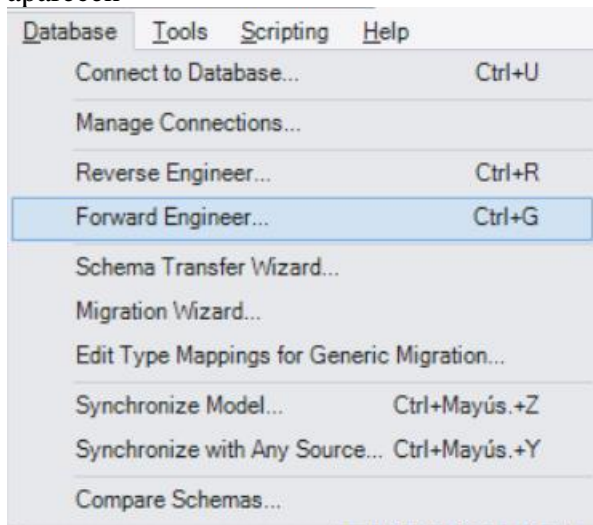
EstadoTecho												
Column name	Data Type	PK	NN	UQ	BIN	UN	ZF	AI	Default	Comment		
idEstadoTecho	INT	✓	✓							Identificador del estado del techo		
opcion	VARCHAR(100)									Identificador de las opciones de estado del techo		

La ruta del diccionario dentro del repositorio es la siguiente

<https://github.com/eaalvarado8/BDAvanzada-ProyectoIntegrador/blob/main/modeloFisico/diccionario.html>

Implementación y carga

Para generar el dll de la base de datos se lo hizo mediante el modelo entidad relación que se modelo en Mysql Workbench, este creo todas las tablas, solamente se tubo que dar clic en el apartado de Database > Forward Engineer y seguido darle a siguiente en las ventanas que aparecen



Forward Engineer to Database

Connection Options

Options

Select Objects

Review SQL Script

Commit Progress

Set Parameters for Connecting to a DBMS

Stored Connection: Local instance MySQL80

Select from saved connection settings

Connection Method: Standard (TCP/IP)

Method to use to connect to the RDBMS

Parameters

SSL

Advanced

Hostname: localhost

Port: 3306

Name or IP address of the server host - and TCP/IP port.

Username: root

Name of the user to connect with.

Password: Store in Vault ... Clear

The user's password. Will be requested later if it's not set.

Default Schema:

The schema to use as default schema. Leave blank to select it later.

Back

Next

Cancel

Forward Engineer to Database

Connection Options

Options

Select Objects

Review SQL Script

Commit Progress

Set Options for Database to be Created

Tables

☐ Skip creation of FOREIGN KEYS

☐ Skip creation of FK indexes as well

☐ Generate separate CREATE INDEX statements

☐ Generate INSERT statements for tables

☐ Disable FK checks for INSERTs

Other Objects

☐ Don't create view placeholder tables

☐ Do not create users. Only create privileges (GRANTS)

Code Generation

☐ DROP objects before each CREATE object

☐ Generate DROP SCHEMA

☐ Omit schema qualifier in object names

☐ Generate USE statements

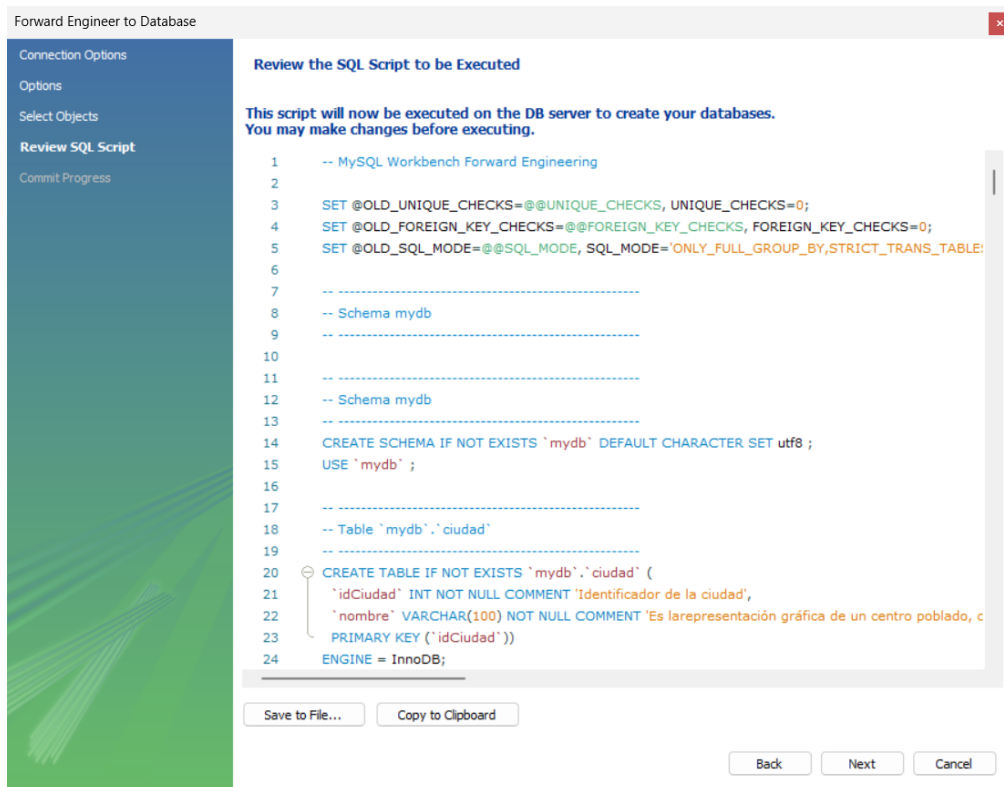
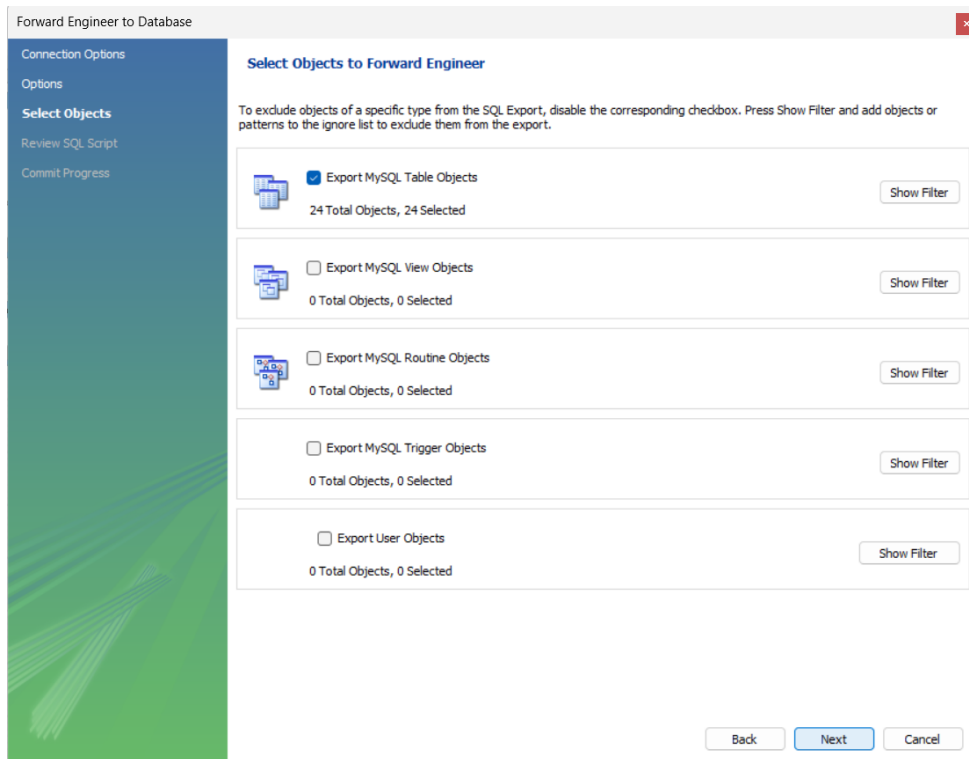
☐ Add SHOW WARNINGS after every DDL statement

☒ Include model attached scripts

Back

Next

Cancel



[Connection Options](#)[Options](#)[Select Objects](#)[Review SQL Script](#)**[Commit Progress](#)**

Forward Engineering Progress

The following tasks will now be executed. Please monitor the execution.
Press Show Logs to see the execution logs.

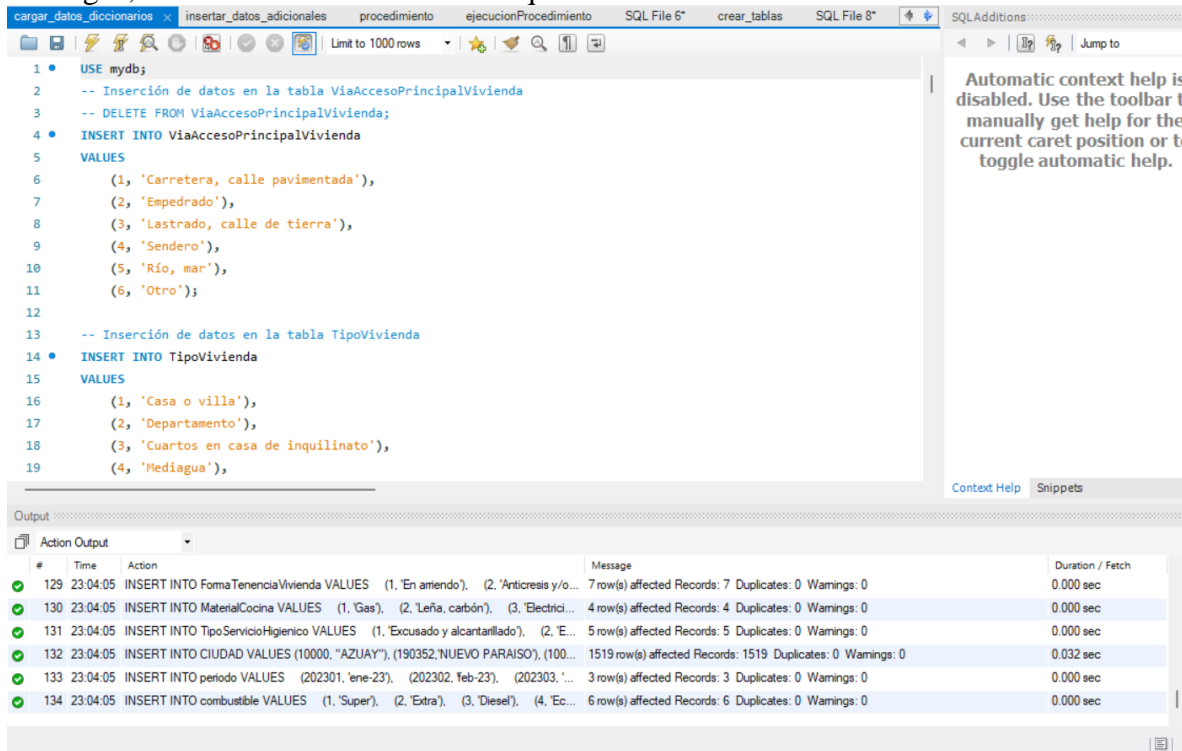
- ☒ Connect to DBMS
- ☒ Execute Forward Engineered Script
- ☒ Read Back Changes Made by Server
- ☒ Save Synchronization State

Forward Engineer Finished Successfully

[Show Logs](#)[Back](#)[Close](#)[Cancel](#)

Implementación y carga

Como la base ya se creo en base al modelo entidad-relación procedimos a insertar daros en los catalogos, todo esto en base a las tablas que clasificamos anteriormente



Para insertar el resto de los datos analizamos que la mejor opción era optar por hacerlo mediante Python, leyendo el csv por lo que implementamos un script para poder realizarlo, el script se basa en 4 partes

1. La conexión con la base de datos

```
insertarDatos > insertar_datos.py > ...
1 import csv
2 import mysql.connector
3
4 conexion = mysql.connector.connect(
5     host="localhost",
6     user="root",
7     password="root",
8     database="mydb"
9 )
```


2. Lectura del archivo csv

```
datos_csv = []

with open('datos.csv', 'r') as archivo_csv:
    lector_csv = csv.DictReader(archivo_csv)

    # Almacena los datos del archivo CSV en una lista
    datos_csv = list(lector_csv)
```

3. Ingresar los datos de vivienda

```
print("Ingresando datos de viviendas")
# Itera sobre los datos almacenados en la lista
for fila in datos_csv:
    try:
        # Inserta los datos en la tabla "Vivienda"
        cursor.execute('INSERT IGNORE INTO vivienda (idVivienda,conglomerado,area,numero,ciudad_idCiudad,periodo_idPeriodo,ViaAccesoPrincipalVivienda,
        (fila['id_vivienda'],
        None if fila['conglomerado'] == '' else fila['conglomerado'],
        'u' if fila['area'] == '1' else 'r',
        None if fila['vivienda'] == '' else fila['vivienda'],
        '999999' if fila['ciudad'] == '230200' else fila['ciudad'],
        None if fila['periodo'] == '' else fila['periodo'],
        None if fila['ViaAccesoPrincipalVivienda'] == '' else fila['ViaAccesoPrincipalVivienda'],
        None if fila['TipoVivienda'] == '' else fila['TipoVivienda'],
        ))
    except mysql.connector.IntegrityError as e:
        # Imprime el valor de "id_vivienda" que está causando el error
        print(f'Error de integridad referencial. Valor de id_vivienda: {fila["id_vivienda"]} ')
    conexion.commit()
print("Ingresando datos de hogares")
```

4. Ingresar los datos de hogares

```
# Itera nuevamente sobre los datos almacenados en la lista
for fila in datos_csv:
    try:
        # Inserta los datos en la tabla "Hogar"
        cursor.execute('INSERT INTO Hogar (idHogar,NumVehiculos,tieneEspacioCocina,NumMotos,IncluyePagoAguaArriendo,IncluyePagoLuzArriendo,ValorArri
        (fila['id_hogar'],
        None if fila['NumVehiculos'] == '' else fila['NumVehiculos'],
        True if fila['EspacioCocinar'] == '1' else False,
        None if fila['NumMotos'] == '' else fila['NumMotos'],
        True if fila['IncluyePagoAguaArriendo'] == '1' else False,
        True if fila['IncluyePagoLuzArriendo'] == '1' else False,
        None if fila['ValorArriendoMensual'] == '' else fila['ValorArriendoMensual'],
        None if fila['NumCuartos'] == '' else fila['NumCuartos'],
        None if fila['NumDormitorios'] == '' else fila['NumDormitorios'],
        None if fila['NumCuartosNegocio'] == '' else fila['NumCuartosNegocio'],
        None if fila['MiembroSinServicioHigienico'] == '' else fila['MiembroSinServicioHigienico'],
        None if fila['InstalacionSanitariaPrincipal'] == '' else fila['InstalacionSanitariaPrincipal'],
        None if fila['TipoAguaVivienda'] == '' else fila['TipoAguaVivienda'],
        None if fila['EliminacionBasura'] == '' else fila['EliminacionBasura'],
        None if fila['FormaTenenciaVivienda'] == '' else fila['FormaTenenciaVivienda'],
        None if fila['TipoServicioHigienico'] == '' else fila['TipoServicioHigienico'],
        None if fila['MaterialPiso'] == '' else fila['MaterialPiso'],
        None if fila['MaterialParedes'] == '' else fila['MaterialParedes'],
        None if fila['MaterialTechoCubierta'] == '' else fila['MaterialTechoCubierta'],
        None if fila['EstadoPiso'] == '' else fila['EstadoPiso'],
        None if fila['EstadoParedes'] == '' else fila['EstadoParedes'],
        None if fila['EstadoTecho'] == '' else fila['EstadoTecho'],
        None if fila['FuenteAgua'] == '' else fila['FuenteAgua'],
        None if fila['TipoAlumbrado'] == '' else fila['TipoAlumbrado'],
        None if fila['MaterialCocina'] == '' else fila['MaterialCocina'],
        None if fila['ServicioDucha'] == '' else fila['ServicioDucha'],
```

```

insertarDatos > insertar_datos.py > ...
67         None if fila['EstadoParedes'] == '' else fila['EstadoParedes'],
68         None if fila['EstadoTecho'] == '' else fila['EstadoTecho'],
69         None if fila['FuenteAgua'] == '' else fila['FuenteAgua'],
70         None if fila['TipoAlumbrado'] == '' else fila['TipoAlumbrado'],
71         None if fila['MaterialCocina'] == '' else fila['MaterialCocina'],
72         None if fila['ServicioDucha'] == '' else fila['ServicioDucha'],
73         fila['id_vivienda']
74     ))
75     conexion.commit()
76     hogar_idHogar = fila['id_hogar']
77
78     # Verifica el tipo de combustible y realiza la inserción en la tabla "HogarCombustible"
79     if fila['AbastecimientoFrecuenciaSuper'] == '1':
80         combustible_idcombustible = 1
81         gasto = fila['GastoCombustibleSuper']
82
83         cursor.execute('INSERT INTO HogarCombustible (hogar_idHogar, combustible_idcombustible, gasto) VALUES (%s, %s, %s)', (hogar_idHogar, com
84
85     elif fila['AbastecimientoFrecuenciaExtra'] == '1':
86         combustible_idcombustible = 2
87         gasto = fila['GastoCombustibleExtra']
88
89         cursor.execute('INSERT INTO HogarCombustible (hogar_idHogar, combustible_idcombustible, gasto) VALUES (%s, %s, %s)', (hogar_idHogar, com
90
91     elif fila['AbastecimientoFrecuenciaDiesel'] == '1':
92         combustible_idcombustible = 3
93         gasto = fila['GastoCombustibleDiesel']
94
95         cursor.execute('INSERT INTO HogarCombustible (hogar_idHogar, combustible_idcombustible, gasto) VALUES (%s, %s, %s)', (hogar_idHogar, com
96
97     elif fila['AbastecimientoFrecuenciaEcopais'] == '1':
98         combustible_idcombustible = 4

```

```

91     elif fila['AbastecimientoFrecuenciaDiesel'] == '1':
92         combustible_idcombustible = 3
93         gasto = fila['GastoCombustibleDiesel']
94
95         cursor.execute('INSERT INTO HogarCombustible (hogar_idHogar, combustible_idcombustible, gasto) VALUES (%s, %s, %s)', (hogar_idHogar, com
96
97     elif fila['AbastecimientoFrecuenciaEcopais'] == '1':
98         combustible_idcombustible = 4
99         gasto = fila['GastoCombustibleEcopais']
100
101         cursor.execute('INSERT INTO HogarCombustible (hogar_idHogar, combustible_idcombustible, gasto) VALUES (%s, %s, %s)', (hogar_idHogar, com
102
103     elif fila['AbastecimientoFrecuenciaElectricidad'] == '1':
104         combustible_idcombustible = 5
105         gasto = fila['GastoCombustibleElectricidad']
106
107         cursor.execute('INSERT INTO HogarCombustible (hogar_idHogar, combustible_idcombustible, gasto) VALUES (%s, %s, %s)', (hogar_idHogar, com
108
109     elif fila['AbastecimientoFrecuenciaGas'] == '1':
110         combustible_idcombustible = 6
111         gasto = fila['GastoCombustibleGas']
112
113         cursor.execute('INSERT INTO HogarCombustible (hogar_idHogar, combustible_idcombustible, gasto) VALUES (%s, %s, %s)', (hogar_idHogar, com
114
115     except mysql.connector.IntegrityError as e:
116         # Imprime el valor de "id_vivienda" que está causando el error
117         print(f'Error de integridad referencial. Valor de id_vivienda: {fila["id_vivienda"]} - Valor de hogar {fila["id_hogar"]}')
118         continue
119     conexion.commit()
120     conexion.close()

```

De manera resumida lo que hace el script es iterar sobre los datos del csv y buscar las columnas que le pertenecen a cada tabla, en el caso de hogares hace un proceso adicional pues también inserta datos en la tabla HogarCombustible que es la relación m:m que existe.

Algo adicional a recalcar es que cuando se insertan los datos de vivienda se lo hace con un INSERT IGNORE puesto que solo necesitamos el primer dato de vivienda que exista con un id único, además de que en cada proceso capturamos las excepciones en el caso de que estas se produzcan.

```

> & C:/Users/Usuario/AppData/Local/Programs/Python/Python311/python.exe c:/Users
Ingresando datos de viviendas
Ingresando datos de hogares

```

Los enlaces son los siguientes

Script de creación de las tablas: https://github.com/eaalvarado8/BDAvanzada-ProyectoIntegrador/blob/main/crearTablas/crear_tablas.sql

Script de subir datos de los catálogos: https://github.com/eaalvarado8/BDAvanzada-ProyectoIntegrador/blob/main/insertarDatos/cargar_datos_diccionarios.sql

Script cargar datos adicionales: https://github.com/eaalvarado8/BDAvanzada-ProyectoIntegrador/blob/main/insertarDatos/insertar_datos_adicionales.sql

Script insertar todos los datos del csv: https://github.com/eaalvarado8/BDAvanzada-ProyectoIntegrador/blob/main/insertarDatos/insertar_datos.py

Descarga en formato CSV

Se procedió a crear el procedimiento almacenado para poder generar los csv

```
USE mydb;
DELIMITER //
CREATE PROCEDURE exportar_csv(
    IN nombre_tabla VARCHAR(100),
    IN ruta VARCHAR(255)
)
BEGIN
    DECLARE nombre_archivo VARCHAR(255);

    -- Concatenar la ruta y el nombre de la tabla para formar el nombre de archivo
    SET nombre_archivo := CONCAT(ruta, '/', nombre_tabla, '.csv');

    SET @query := CONCAT(
        'SELECT * INTO OUTFILE \'', nombre_archivo, '\',
        'FIELDS TERMINATED BY \',\' ',
        'ENCLOSED BY '\"\' ',
        'LINES TERMINATED BY \'\\n\' ',
        'FROM ', nombre_tabla
    );

    PREPARE stmt FROM @query;
    EXECUTE stmt;
    DEALLOCATE PREPARE stmt;
END //

DELIMITER ;
```

Luego e probó la ejecución de este, pero se tuvo algunos errores como el de Error code 1290 the mysql server is running with the --secure-file-priv option

3 20:06:48 CALL exportar_csv('ciudad', 'C:\Users\Usuario\Documents\Universidad\BDAvanzada-Pr... Error Code: 1290. The MySQL server is running with the --secure-file-priv option so it cann... 0.015 sec

El cual se daba debido a que MySQL tiene una configuración que solamente permite guardar en cierta ubicación específica por temas de seguridad por lo que se procedió a comentar dicha configuración

```
166 # those performed by the LOAD DATA and SELECT ... INTO OUTFILE statements and the
167 # LOAD_FILE() function. These operations are permitted only to users who have the FILE privilege.
168 # secure-file-priv="C:/ProgramData/MySQL/MySQL Server 8.0/Uploads"
169
```

```
165 # This variable is used to limit the effect of data import and export operations, such as
166 # those performed by the LOAD DATA and SELECT ... INTO OUTFILE statements and the
167 # LOAD_FILE() function. These operations are permitted only to users who have the FILE privilege.
168 # secure-file-priv="C:/ProgramData/MySQL/MySQL Server 8.0/Uploads"
169
```

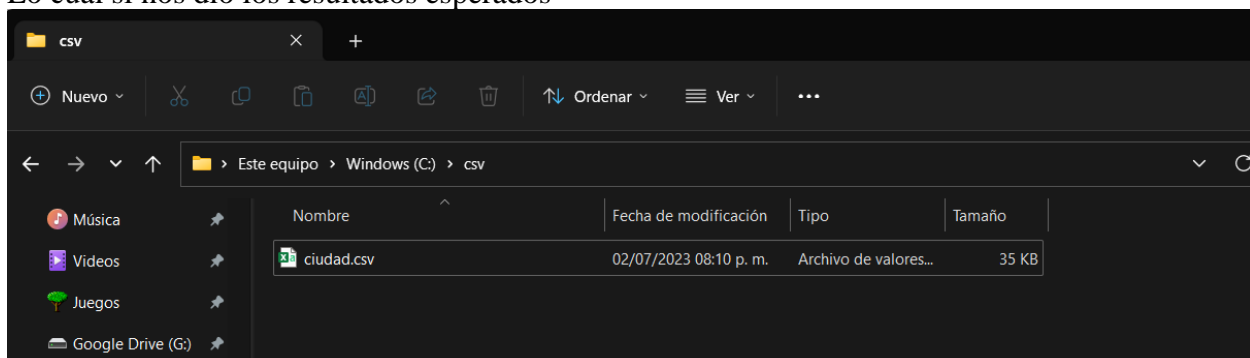
Pero de igual manera daba error por lo que se intentó reemplazar la ubicación con la ubicación objetivo

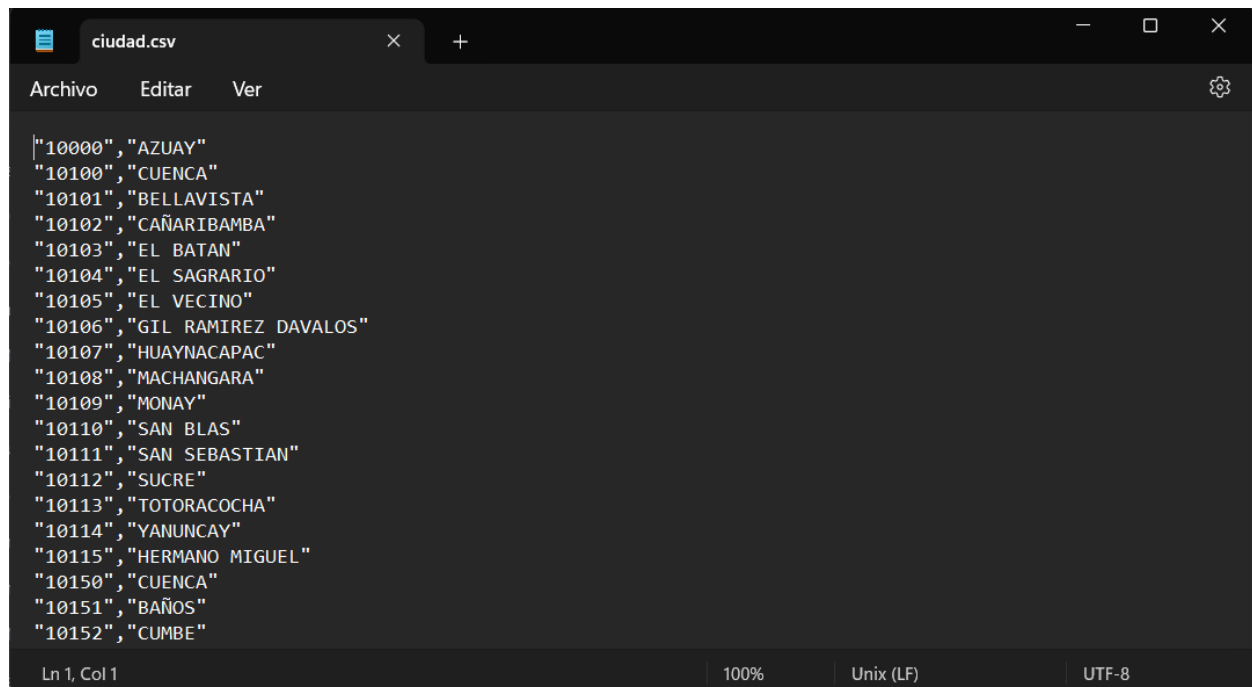
```
# This variable is used to limit the effect of data import and export operations, such as
# those performed by the LOAD DATA and SELECT ... INTO OUTFILE statements and the
# LOAD_FILE() function. These operations are permitted only to users who have the FILE privilege.
secure-file-priv="C:\Users\Usuario\Documents\Universidad\BDAvanzada-ProyectoIntegrador\finales_csv"
```

Pero de igual manera daba errores, por lo que finalmente se optó por dejar una ubicación simple para probar que se creen los archivo csv

```
# This variable is used to limit the effect of data import and export operations, such as
# those performed by the LOAD DATA and SELECT ... INTO OUTFILE statements and the
# LOAD_FILE() function. These operations are permitted only to users who have the FILE privilege.
secure-file-priv="C:/csv"
```

Lo cual si nos dio los resultados esperados

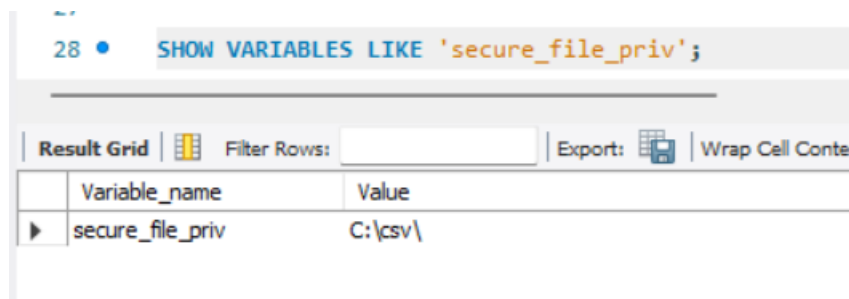
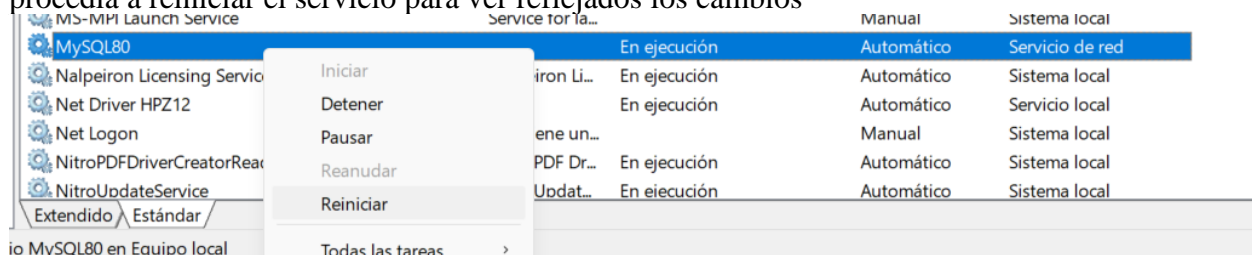




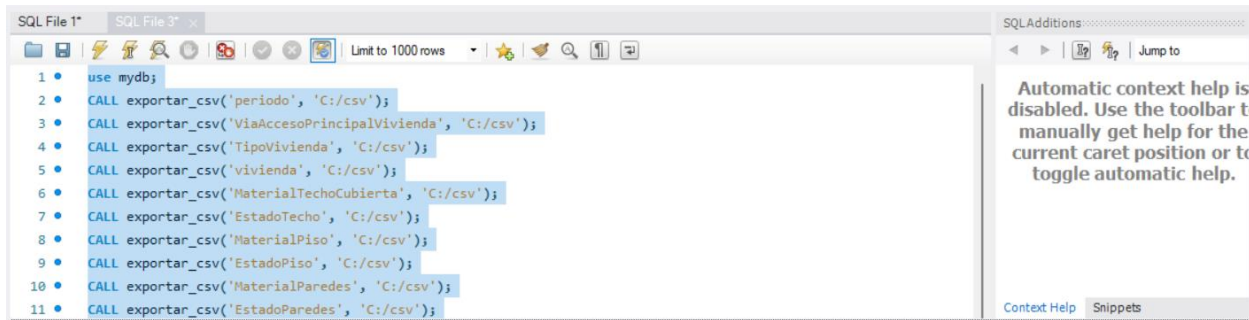
```
"10000", "AZUAY"
"10100", "CUENCA"
"10101", "BELLAVISTA"
"10102", "CAÑARIBAMBA"
"10103", "EL BATAN"
"10104", "EL SAGRARIO"
"10105", "EL VECINO"
"10106", "GIL RAMIREZ DAVALOS"
"10107", "HUAYNACAPAC"
"10108", "MACHANGARA"
"10109", "MONAY"
"10110", "SAN BLAS"
"10111", "SAN SEBASTIAN"
"10112", "SUCRE"
"10113", "TOTORACOCOA"
"10114", "YANUNCAY"
"10115", "HERMANO MIGUEL"
"10150", "CUENCA"
"10151", "BAÑOS"
"10152", "CUMBE"
```

Por lo que se optó por generar los csv en esta ruta y luego ubicarlos dentro del proyecto de github

Cada recalcar que cada vez que se realizaba un cambio en el archivo de configuración my.ini se procedía a reiniciar el servicio para ver reflejados los cambios



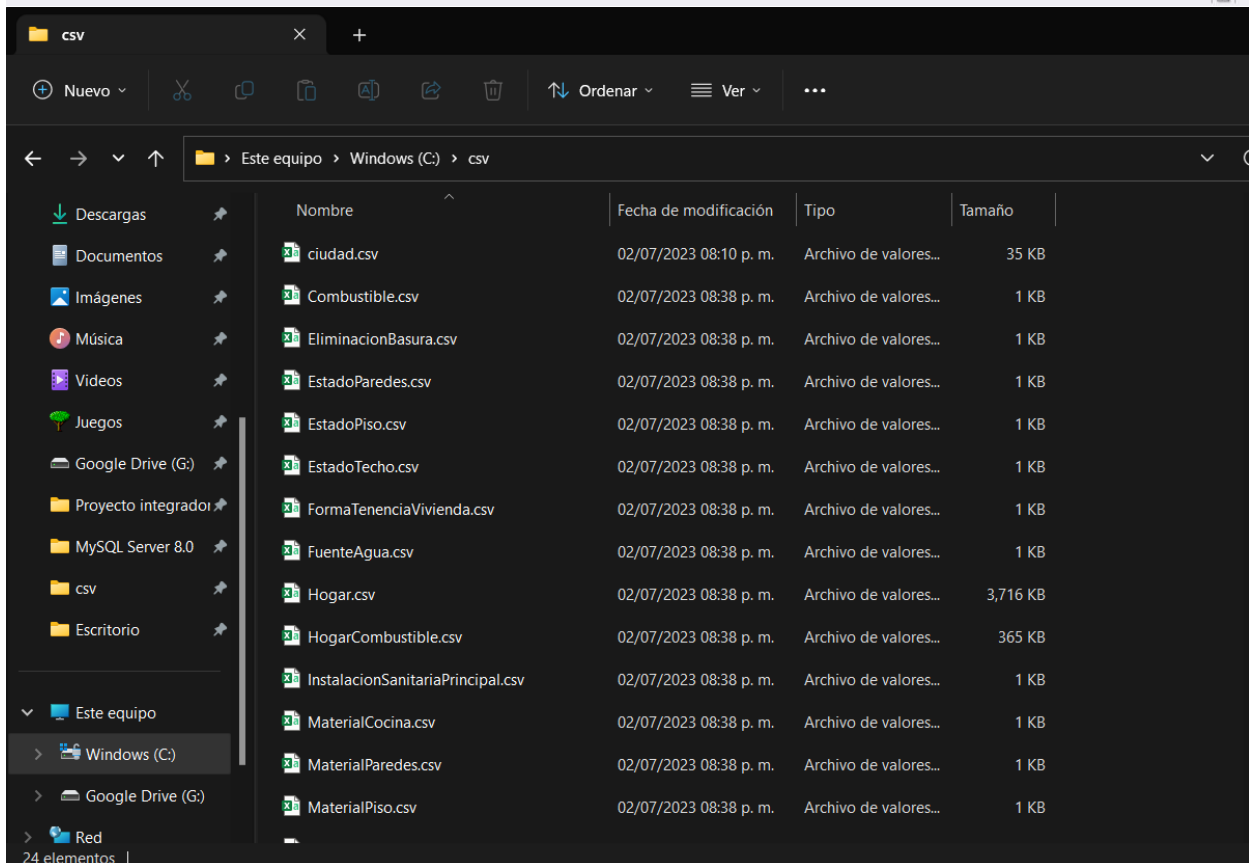
El resultado final fue el siguiente



Output

Action Output

#	Time	Action	Message	Duration / Fetch
24	20:38:16	CALL exportar_csv('MaterialParedes', 'C:/csv')	0 row(s) affected	0.000 sec
25	20:38:16	CALL exportar_csv('EstadoParedes', 'C:/csv')	0 row(s) affected	0.015 sec
26	20:38:16	CALL exportar_csv('MiembrosSinServicioHigienico', 'C:/csv')	0 row(s) affected	0.000 sec
27	20:38:16	CALL exportar_csv('InstalacionSanitariaPrincipal', 'C:/csv')	0 row(s) affected	0.000 sec
28	20:38:16	CALL exportar_csv('TipoAguaVivienda', 'C:/csv')	0 row(s) affected	0.016 sec
29	20:38:16	CALL exportar_csv('EliminacionBasura', 'C:/csv')	0 row(s) affected	0.000 sec
30	20:38:16	CALL exportar_csv('FormaTenenciaVivienda', 'C:/csv')	0 row(s) affected	0.000 sec
31	20:38:16	CALL exportar_csv('TipoServicioHigienico', 'C:/csv')	0 row(s) affected	0.000 sec
32	20:38:16	CALL exportar_csv('FuenteAgua', 'C:/csv')	0 row(s) affected	0.000 sec
33	20:38:16	CALL exportar_csv('TipoAlumbrado', 'C:/csv')	0 row(s) affected	0.015 sec
34	20:38:16	CALL exportar_csv('MaterialCocina', 'C:/csv')	0 row(s) affected	0.000 sec
35	20:38:16	CALL exportar_csv('ServicioDucha', 'C:/csv')	0 row(s) affected	0.000 sec
36	20:38:16	CALL exportar_csv('Hogar', 'C:/csv')	0 row(s) affected	0.094 sec
37	20:38:16	CALL exportar_csv('Combustible', 'C:/csv')	0 row(s) affected	0.016 sec
38	20:38:16	CALL exportar_csv('HogarCombustible', 'C:/csv')	0 row(s) affected	0.015 sec



Enlace de la creación del procedimiento: https://github.com/eaalvarado8/BDAvanzada-ProyectoIntegrador/blob/main/procedimiento_csv/procedimiento.sql

Enlace ejecución del procedimiento: https://github.com/eaalvarado8/BDAvanzada-ProyectoIntegrador/blob/main/procedimiento_csv/ejecucionProcedimiento.sql

Bibliografía

INEC. (2022). *ENEMDU-2022*. Instituto Nacional de Estadística Y Censos.

<https://www.ecuadorencifras.gob.ec/enemdu-2022/>

INEC. (2013). *Ecuador - Encuesta Nacional de Empleo, Desempleo y Subempleo - Diciembre 2011, RONDA XXXIV-12-2011 - Buscar*. Inec.gob.ec.

<https://anda.inec.gob.ec/anda/index.php/catalog/269/search?vk=panelm>