

## SELECCIÓN DE FUENTE DE DATOS

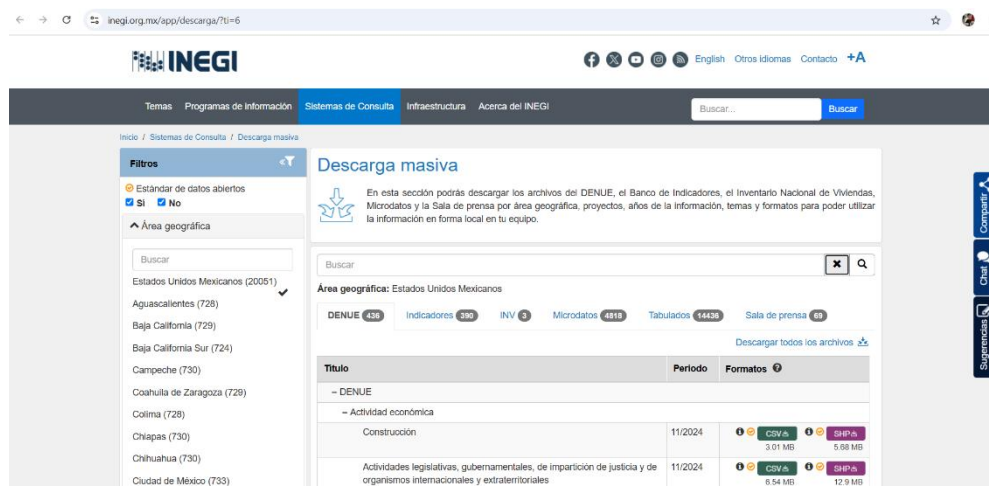
- DENEUE de INEGI como base de datos
- El archivo es un CSV de construcción

## IMPLEMENTACIÓN DE PROCESO ETL

- La extracción de la información se realizó buscando en la base de datos del INEGI
- Se adaptó una transformación en una de las columnas del archivo CSV mediante códigos de Python para cambiar la fecha original (numérico) a un formato (alfanumérico)-
- La carga del archivo CSV modificado se realizó a través de una interfaz gráfica (HeidiSQL) a la base de datos de MariaDB.

## DOCUMENTACIÓN

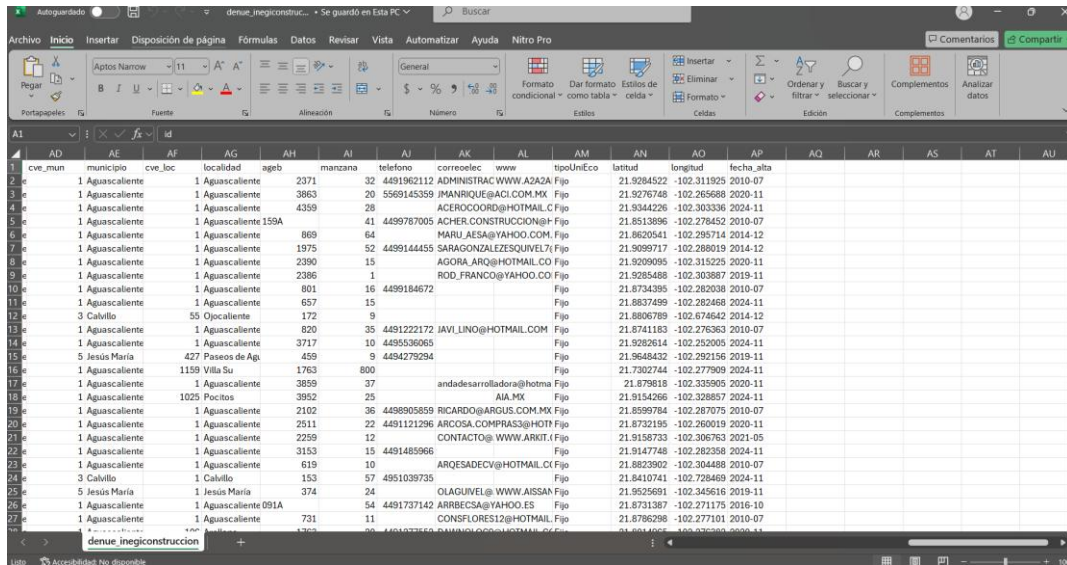
### Extracción en base de datos



Se seleccionó el archivo CSV más cercano a la fecha actual.

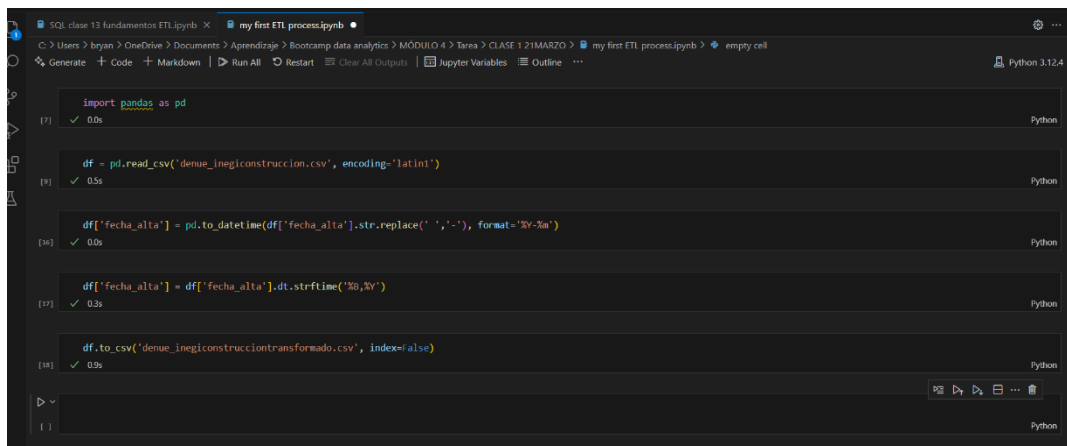
- Una de las principales dificultades en este punto es poder encontrar una plataforma o una base de datos que contenga información y estadísticas confiables.

## Transformación del archivo CSV



AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM	AN	AO	AP	AQ	AR	AS	AT	AU
cve_mun	municipio	cve_loc	localidad	ageb	manzana	telefono	correoelec	www	tipoUnEco	latitud	longitud	fecha_alta					
1	1	1	Agascaliente	2371	32	4491962112	ADMINISTRAC	WWW.A2A2A	Fijo	21.9284522	-102.311925	2010-07					
2	1	1	Agascaliente	3863	20	5569145359	JMANRIQUE@ACI.COM.MX		Fijo	21.9276748	-102.265688	2020-11					
3	1	1	Agascaliente	4359	28		ACEROCORR@HOTMAIL.C		Fijo	21.9344226	-102.303336	2024-11					
4	1	1	Agascaliente	159A	41	4499787005	ACHER.CONSTRUCCION@H		Fijo	21.8513896	-102.278452	2019-07					
5	1	1	Agascaliente	869	64		MARU.AESA@YAHOO.COM		Fijo	21.8620541	-102.295714	2014-12					
6	1	1	Agascaliente	1975	52	4499144455	SARAGONZALEZSQUIVEL7		Fijo	21.9099717	-102.288019	2014-12					
7	1	1	Agascaliente	2390	15		AGORA.ARQ@HOTMAIL.CO		Fijo	21.9209095	-102.315225	2020-11					
8	1	1	Agascaliente	2386	1		ROD.FRANCO@YAHOO.CO		Fijo	21.9285488	-102.303887	2019-11					
9	1	1	Agascaliente	801	16	4499184672			Fijo	21.8734395	-102.262038	2010-07					
10	1	1	Agascaliente	857	15				Fijo	21.8837489	-102.282468	2024-11					
11	3	Calvillo	55 Opascaliente	172	9				Fijo	21.8806789	-102.674642	2014-12					
12	1	1	Agascaliente	820	35	4491222172	JAVI.LINO@HOTMAIL.COM		Fijo	21.8741183	-102.276363	2010-07					
13	1	1	Agascaliente	3717	10	4495536065			Fijo	21.9282614	-102.252005	2024-11					
14	5	Jesús María	427 Paseos de Ags.	459	9	4494279294			Fijo	21.9648432	-102.292156	2019-11					
15	1	1	Agascaliente	1159 Villa Su	1763	800			Fijo	21.7302744	-102.277909	2024-11					
16	1	1	Agascaliente	3859	37		andadesarrolladora@hotma		Fijo	21.879819	-102.235905	2020-11					
17	1	1	Agascaliente	3952	25		AIA.MX		Fijo	21.9154266	-102.328657	2024-11					
18	1	1	Agascaliente	2102	36	4498905859	RICARDO@ARGUS.COM.MX		Fijo	21.8599784	-102.287075	2010-07					
19	1	1	Agascaliente	2511	22	4491121296	ARCOSA.COMPRAS3@HOT		Fijo	21.8732195	-102.260019	2020-11					
20	1	1	Agascaliente	2259	12		CONTACTO@WWW.ARKIT.		Fijo	21.9158733	-102.306763	2021-05					
21	1	1	Agascaliente	3153	15	4491485966			Fijo	21.9147748	-102.282358	2024-11					
22	1	1	Agascaliente	619	10		ARQESADECV@HOTMAIL.C		Fijo	21.8623892	-102.204498	2010-07					
23	3	Calvillo	1 Calvillo	153	57	4951039735			Fijo	21.8410741	-102.728469	2024-11					
24	5	Jesús María	1 Jesús María	374	24		OLAGUIVEL@WWW.AISSAN		Fijo	21.9525691	-102.345616	2019-11					
25	1	1	Agascaliente	091A	54	4491737142	ARRRECSA@YAHOO.ES		Fijo	21.8731387	-102.277175	2016-10					
26	1	1	Agascaliente	731	11		CONSFLORES12@HOTMAIL		Fijo	21.8786298	-102.277101	2010-07					
27	1	1	Agascaliente	1763	35	4491222172	JAVI.LINO@HOTMAIL.COM		Fijo	21.8741183	-102.276363	2010-07					

Se identificó el formato de fecha en la última columna como (%Y-%M)



```
import pandas as pd

df = pd.read_csv('denue_inegiconstruccion.csv', encoding='latin1')

df['fecha_alta'] = pd.to_datetime(df['fecha_alta'].str.replace(' ','-'), format='%Y-%m')

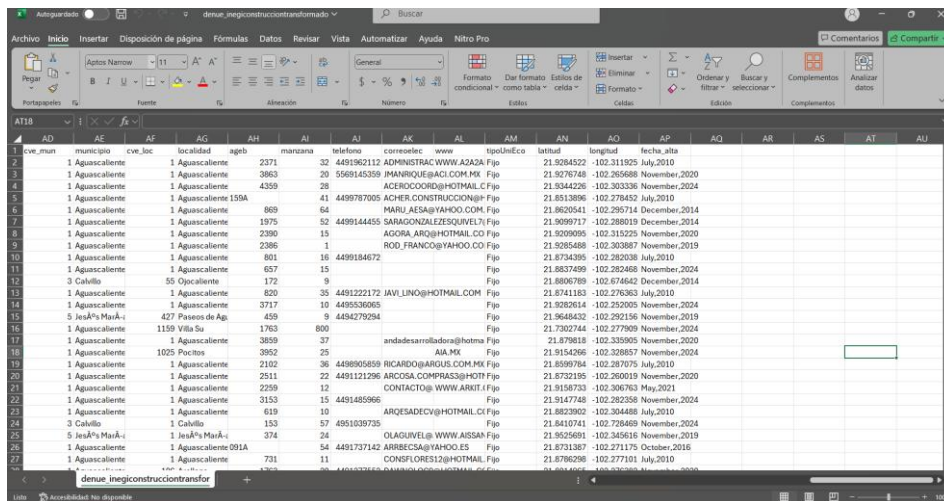
df['fecha_alta'] = df['fecha_alta'].dt.strftime('%B, %Y')

df.to_csv('denue_inegiconstrucciontransformado.csv', index=False)
```

A través de cinco códigos en un jupyter Notebook se logró transformar el formato de la fecha original (%Y-%M) a un formato alfanumérico (%B, %Y).

- Algunas de las complicaciones encontradas fueron la lectura del archivo, pues Python no reconocía un carácter pues por de fault estaba leyendo el archivo con codificación UTF-8 no obstante, se tuvo que especificar con el comando encoding = latin1, para que pudiera reconocer ese carácter.
- Otra complicación fue referente a que una fila de la columna fecha, tenía espacio en lugar de guion medio, y Python marcaba error, entonces se utilizó el comando

.str.replace () para poder explicarle a Python que debía **corregir** esos **espacios vacíos** con guiones medios.



AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM	AN	AO	AP	AQ	AR	AS	AT	AU
cve_mun	municipio	cve_loc	localidad	ageb	municipio	telefono	correo_elec	www	tipoUnico	latitud	longitud	fecha_alta					
1	Aguascaliente	1	Aguascaliente	2371	32	4491962112	ADMINISTRAC	WWW.AZAZA	Fijo	21.9284522	-102.311925	July, 2010					
2	Aguascaliente	1	Aguascaliente	3863	20	5569145359	IMANRIQUE@	ACI.COM.MX	Fijo	21.9276748	-102.265688	November, 2020					
3	Aguascaliente	1	Aguascaliente	4359	28		ACEROCORD@	HOTMAIL.C	Fijo	21.9344226	-102.303336	November, 2024					
4	Aguascaliente	1	Aguascaliente	159A	41	4499787085	ACHER.CON	STRUCO@nig	Fijo	21.8913896	-102.278452	July, 2010					
5	Aguascaliente	1	Aguascaliente	869	64		MARU_AESA	Y@HOO.COM	Fijo	21.8620541	-102.295714	December, 2014					
6	Aguascaliente	1	Aguascaliente	1975	52	4499144455	SARAGONZALE	ZESQUIVEL	Fijo	21.9099717	-102.288019	December, 2014					
7	Aguascaliente	1	Aguascaliente	2390	15		AGORA_ARQ	@HOTMAIL.CO	Fijo	21.9209095	-102.315225	November, 2020					
8	Aguascaliente	1	Aguascaliente	2386	1		ROD_FRANCO	@Y@HOO.CO	Fijo	21.9265468	-102.330387	November, 2019					
9	Aguascaliente	1	Aguascaliente	801	16	4499184672			Fijo	21.8734395	-102.282038	July, 2010					
10	Aguascaliente	1	Aguascaliente	657	15				Fijo	21.8837499	-102.282468	November, 2024					
11	Aguascaliente	1	Aguascaliente	172	9				Fijo	21.8806789	-102.674642	December, 2014					
12	Aguascaliente	1	Aguascaliente	820	35	4491222172	JAVU_LINO	@HOTMAIL.COM	Fijo	21.8741163	-102.276363	July, 2010					
13	Aguascaliente	1	Aguascaliente	3717	10	4495536065			Fijo	21.9282614	-102.252005	November, 2024					
14	Aguascaliente	1	Aguascaliente	459	9	4494279294			Fijo	21.9648432	-102.292156	November, 2019					
15	Aguascaliente	1	Aguascaliente	1150	Villa Su	1763	800		Fijo	21.7302744	-102.277909	November, 2024					
16	Aguascaliente	1	Aguascaliente	3859	37		andadesarrollado	rox@hotmail	Fijo	21.8796118	-102.330905	November, 2020					
17	Aguascaliente	1	Aguascaliente	1025	Pocitos	3952	25		Fijo	21.9154266	-102.328657	November, 2024					
18	Aguascaliente	1	Aguascaliente	2102	36	4498905859	RICARDOD@	AROUS.COM.MX	Fijo	21.8599764	-102.287075	July, 2010					
19	Aguascaliente	1	Aguascaliente	2511	22	4491121296	ARCOSA.COM	PRAC@HOT	Fijo	21.9732195	-102.260019	November, 2020					
20	Aguascaliente	1	Aguascaliente	2259	12		CONTACTO@	WWW.ARKIT	Fijo	21.9158733	-102.306763	May, 2021					
21	Aguascaliente	1	Aguascaliente	3153	15	4491485966			Fijo	21.9147748	-102.282358	November, 2024					
22	Aguascaliente	1	Aguascaliente	619	10		ARQESADECV	@HOTMAIL.C	Fijo	21.8823902	-102.304488	July, 2010					
23	Aguascaliente	1	Aguascaliente	153	57	4951039755			Fijo	21.8410741	-102.728405	November, 2024					
24	Aguascaliente	1	Aguascaliente	374	24		OLACUIVEL@	WWW.AISSA	Fijo	21.9525691	-102.345616	November, 2019					
25	Aguascaliente	1	Aguascaliente	091A	54	4491737142	ARRRECSA@	Y@HOO.ES	Fijo	21.8731387	-102.271175	October, 2016					
26	Aguascaliente	1	Aguascaliente	731	11		CONFLORES	12@hotmail	Fijo	21.8786286	-102.277101	July, 2010					

Al final el archivo se guardó con un formato de fecha modificado y con las transformaciones necesarias.

## Carga de base de datos relacional

HeidiSQL: my\_first\_eti... HeidiSQL 12.10.000

Archivo Editor Buscar Consultas Herramientas Ayuda

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Filtro de base de datos

Filtro de tabla

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...process

Tabla: construccion\_inegi

Datos

Consultar

Unselected

Base de datos: my\_first\_eti...

Finalmente se cargó la nueva base de datos de construcción en MariaDB, utilizando como interfaz a HeidiSQL.

- Una de las cosas importantes al importar un archivo CSV a heidiSQL es que se debe indicar que los valores deben estar separados por una coma (,) y no por punto y coma (;).
- De igual manera, cerciorarse de que la codificación sea UTF-8.

## REFLEXIÓN

Un proceso ETL es de gran importancia para poder tomar decisiones basadas en datos dentro de una empresa, este proceso requiere aprender adecuadamente varios conceptos clave dentro de la codificación, tener una mente atenta a los detalles pues al importar un archivo a Python se pueden encontrar errores o valores faltantes y es importante conocer códigos que permitan una correcta transformación de la información, que incluso resuma lo mejor del archivo y así al cargarlo, se pueda generar una base de datos más confiable.