

SELECCIÓN DE FUENTE DE DATOS

- DENUE 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 (alfanuméricico)-
- 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

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

```

import pandas as pd
df = pd.read_csv('denue_inegiconstrucion.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_inegiconstruciontransformado.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 default 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 | telefono | correolec | www | AM | latitud | longitud | fecha_alsa |
|----|-----------------|-------------------|---------|------------|------------------------|------------------------------------|-------------------------|------------------------|------------|------------|---------------|------------|
| 1 | cve_mun | manzana | cve_loc | localidad | ageb | manzana | | | | | | |
| 2 | 1 Aguascaliente | 1 Aguascaliente | 2371 | 32 | 449192112 | ADMINISTRACIÓNAZ2A2AQUITECTURA.COM | Fijo | WWW.AZ2AQUITECTURA.COM | 21.9264522 | 102.311928 | July,2010 | |
| 3 | 1 Aguascaliente | 1 Aguascaliente | 3663 | 20 | 5569145359 | IMANRQUE@ACI.COM.MX | Fijo | | 21.9276748 | 102.265668 | November,2020 | |
| 4 | 1 Aguascaliente | 1 Aguascaliente | 4359 | 28 | | ACEROCONCORDIHOTMAIL.COM | Fijo | | 21.934422 | 102.303338 | November,2024 | |
| 5 | 1 Aguascaliente | 1 Aguascaliente | 159A | 41 | 4499787052 | ACHER.CONSTRUCCION@GMAIL.COM | Fijo | | 21.851388 | 102.279452 | July,2010 | |
| 6 | 1 Aguascaliente | 1 Aguascaliente | 866 | 64 | 52 | 449914455 | MARU.ASA@YAHOO.COM | Fijo | 21.8620541 | 102.295714 | December,2014 | |
| 7 | 1 Aguascaliente | 1 Aguascaliente | 1975 | 15 | 48 | 449914455 | RODRIGOARROYO@GMAIL.COM | Fijo | 21.8620547 | 102.295714 | December,2014 | |
| 8 | 1 Aguascaliente | 1 Aguascaliente | 2390 | 15 | 52 | 449914455 | AGORA.ARQ@HOTMAIL.COM | Fijo | 21.909909 | 102.315229 | November,2020 | |
| 9 | 1 Aguascaliente | 1 Aguascaliente | 2386 | 1 | | ROD.FRANCO@YAHOO.CO | Fijo | | 21.925688 | 102.303887 | November,2019 | |
| 10 | 1 Aguascaliente | 1 Aguascaliente | 801 | 16 | 4499184672 | | Fijo | | 21.8734395 | 102.282038 | July,2010 | |
| 11 | 1 Aguascaliente | 1 Aguascaliente | 657 | 15 | | | Fijo | | 21.883748 | 102.282488 | November,2024 | |
| 12 | 3 Calvillo | 55 Ojocaliente | 172 | 9 | | | Fijo | | 21.880676 | 102.674642 | December,2014 | |
| 13 | 1 Aguascaliente | 1 Aguascaliente | 820 | 35 | 4491922172 | JAVI_LINO@HOTMAIL.COM | Fijo | | 21.871118 | 102.300383 | July,2010 | |
| 14 | 1 Aguascaliente | 1 Aguascaliente | 3717 | 10 | 4495536065 | | Fijo | | 21.882390 | 102.295388 | November,2024 | |
| 15 | 5 Jesús María-i | 427 Paseos de Ag. | 458 | 9 | 4494279294 | | Fijo | | 21.9648432 | 102.292156 | November,2019 | |
| 16 | 1 Aguascaliente | 1159 Villa Su | 1763 | 800 | | | Fijo | | 21.7302744 | 102.277909 | November,2024 | |
| 17 | 1 Aguascaliente | 1 Aguascaliente | 3858 | 37 | | andadesarrolladora@hotmail | Fijo | | 21.879818 | 102.333905 | November,2020 | |
| 18 | 1 Aguascaliente | 1025 Pocitos | 3952 | 25 | | AIA.MX | Fijo | | 21.915426 | 102.326857 | November,2024 | |
| 19 | 1 Aguascaliente | 1 Aguascaliente | 2102 | 36 | 4498605659 | RICARDO@ARGUS.COM.MX | Fijo | | 21.8599794 | 102.287075 | July,2010 | |
| 20 | 1 Aguascaliente | 1 Aguascaliente | 2511 | 22 | 4491121296 | ARCOSA_COMPRA@GMAIL.COM | Fijo | | 21.915426 | 102.326857 | November,2024 | |
| 21 | 1 Aguascaliente | 1 Aguascaliente | 2559 | 12 | | CONTACTO@WWW.AFRIT.COM | Fijo | | 21.915673 | 102.300762 | May,2021 | |
| 22 | 1 Aguascaliente | 1 Aguascaliente | 3153 | 15 | 4491485966 | | Fijo | | 21.9147748 | 102.282358 | November,2024 | |
| 23 | 1 Aguascaliente | 1 Aguascaliente | 619 | 10 | | ARQEDESDEV@HOTMAIL.COM | Fijo | | 21.882390 | 102.304488 | July,2010 | |
| 24 | 1 Calvillo | 1 Calvillo | 153 | 57 | 4551039735 | | Fijo | | 21.8410741 | 102.726489 | November,2024 | |
| 25 | 5 Jesús María-i | 374 | 24 | | | OLAGUVEI@WWW.AISSAN | Fijo | | 21.952569 | 102.345616 | November,2019 | |
| 26 | 1 Aguascaliente | 1 Aguascaliente | 891A | 54 | 4491737142 | RODRIGOARROYO@YAHOO.COM | Fijo | | 21.873138 | 102.271175 | October,2016 | |
| 27 | 1 Aguascaliente | 1 Aguascaliente | 731 | 11 | | CONFIDORES123@HOTMAIL | Fijo | | 21.876308 | 102.277101 | July,2010 | |
| 28 | 1 Aguascaliente | 1763 | 50 | 4491922172 | SANTOSMARTIN@GMAIL.COM | Fijo | | 21.9144068 | 102.300383 | 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

| | cve_loc | localidad | ageb | manzana | telefono | correolec | www | latitud | longitud | fecha_alsa | |
|----|------------------|-----------------|------|------------|------------------------------------|-------------------------|------------------------|------------|------------|---------------|---------------|
| 1 | 1 Aguascalientes | 2371 | 32 | 20 | 21.9264522 | 102.311928 | July,2010 | | | | |
| 2 | 1 Aguascalientes | 3663 | 20 | 449192112 | ADMINISTRACIÓNAZ2A2AQUITECTURA.COM | Fijo | WWW.AZ2AQUITECTURA.COM | 21.9276748 | 102.265668 | November,2020 | |
| 3 | 1 Aguascalientes | 4359 | 28 | 5569145359 | IMANRQUE@ACI.COM.MX | Fijo | | 21.934422 | 102.303338 | November,2024 | |
| 4 | 1 Aguascalientes | 159A | 41 | 4499787052 | ACHER.CONSTRUCCION@GMAIL.COM | Fijo | | 21.851388 | 102.279452 | July,2010 | |
| 5 | 1 Aguascalientes | 866 | 64 | 52 | 449914455 | MARU.ASA@YAHOO.COM | Fijo | 21.8620541 | 102.295714 | December,2014 | |
| 6 | 1 Aguascalientes | 1975 | 15 | 48 | 449914455 | AGORA.ARQ@HOTMAIL.COM | Fijo | 21.909909 | 102.315229 | November,2020 | |
| 7 | 1 Aguascalientes | 2390 | 15 | 52 | 449914455 | ROD.FRANCO@YAHOO.CO | Fijo | 21.925688 | 102.303887 | November,2019 | |
| 8 | 1 Aguascalientes | 2386 | 1 | 0 | | ROD.FRANCO@YAHOO.COM | Fijo | 21.915426 | 102.326857 | November,2024 | |
| 9 | 1 Aguascalientes | 801 | 16 | 4499184672 | | Fijo | | 21.8734395 | 102.282038 | July,2010 | |
| 10 | 1 Aguascalientes | 657 | 15 | | | Fijo | | 21.883748 | 102.282488 | November,2024 | |
| 11 | 3 Calvillo | 55 Ojocaliente | 172 | 9 | | | Fijo | | 21.880676 | 102.674642 | December,2014 |
| 12 | 1 Aguascaliente | 1 Aguascaliente | 2559 | 12 | | CONTACTO@WWW.AFRIT.COM | Fijo | | 21.9147748 | 102.282358 | November,2024 |
| 13 | 1 Aguascaliente | 1 Aguascaliente | 3153 | 15 | 4491485966 | | Fijo | | 21.915673 | 102.300762 | May,2021 |
| 14 | 1 Aguascaliente | 1 Aguascaliente | 619 | 10 | | ARQEDESDEV@HOTMAIL.COM | Fijo | | 21.882390 | 102.304488 | July,2010 |
| 15 | 1 Calvillo | 1 Calvillo | 153 | 57 | 4551039735 | | Fijo | | 21.8410741 | 102.726489 | November,2024 |
| 16 | 5 Jesús María-i | 374 | 24 | | | OLAGUVEI@WWW.AISSAN | Fijo | | 21.952569 | 102.345616 | November,2019 |
| 17 | 1 Aguascaliente | 1 Aguascaliente | 891A | 54 | 4491737142 | RODRIGOARROYO@YAHOO.COM | Fijo | | 21.873138 | 102.271175 | October,2016 |
| 18 | 1 Aguascaliente | 1763 | 50 | 4491922172 | SANTOSMARTIN@GMAIL.COM | Fijo | | 21.9144068 | 102.300383 | July,2010 | |

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.