

## Integrantes del equipo:

- Marco Daniel Coronado Rodríguez
- Areli Jiménez Morales
- Diego Ramírez Hernández
- Miguel Ángel Vizcaya

## Problemática:

Nuestro cliente AXA ha solicitado una solución tecnológica para analizar los siniestros que ha registrado desde el 2015 a marzo del 2021, cada siniestro tiene una tipificación y un tipo de vehículo involucrado. La información a partir del mes de Abril deberá ser cargada de manera automática en la solución, este archive incremental tendrá el mismo formato que los anteriores.

## Solución.

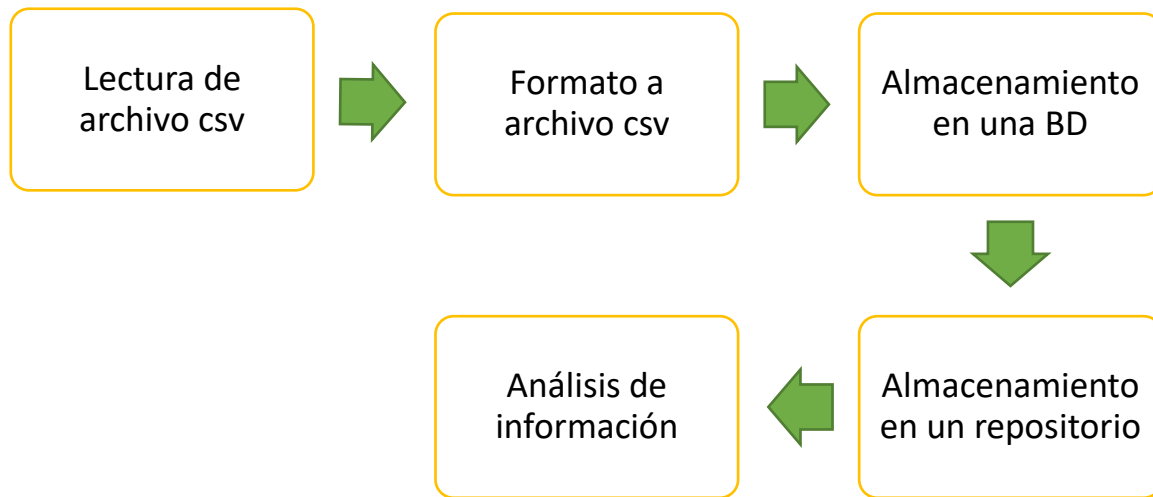
Se cuenta con información que tiene recopilada el negocio; sin embargo dicho conjunto de datos no tiene un formato que mantenga un orden dentro del negocio. Es por ello que everis se ha dado a la tarea de organizar la información con el fin de que la empresa AXA puede hacer un buen análisis de ella para la toma de decisiones.

Toda la información de la empresa está almacenada en archivos con extensión csv, los cuales tienen un formato como el que se muestra a continuación:

| A  | B           | C            | D                    | E       | F          | G            | H            | I                                  | J                   | K                 | L      | M      | N            | O            | P   | Q              |
|----|-------------|--------------|----------------------|---------|------------|--------------|--------------|------------------------------------|---------------------|-------------------|--------|--------|--------------|--------------|-----|----------------|
| 1  | the_geom    | edad_lesiona | the_geom_vrsiniestro | latitud | longitud   | codigo_posta | calle        | colonia                            | causa_siniest       | tipo_vehiculo     | color  | modelo | nivel_dano_v | punto_impact | ano | mes            |
| 2  | 0101000020E | 0            | 01010000201          | 2318763 |            | 0            | PDT          | \N                                 | COLISION Y/C Auto   | BLANCO            |        | 2008   | \N           |              |     | 2015 ENERO     |
| 3  | 0101000020E | 0            | 01010000201          | 2348028 | 17.5575051 | -99.5039141  | 0            | Abasolo                            | CENTRO              | COLISION Y/C Auto | BLANCO | 2011   | Bajo         | Frontal      |     | 2015 ENERO     |
| 4  | 0101000020E | 0            | 01010000201          | 2377039 | 16.928057  | -99.8024118  | 0            | CARRETERA 9 Ampliada               | COLISION Y/C Auto   | ROJO              | 0      | \N     |              |              |     | 2015 FEBRERO   |
| 5  | 0101000020E | 0            | 01010000201          | 2416847 | 16.8378716 | -99.8470545  | 0            | Ejército Nacional                  | COLISION Y/C Auto   | BLANCO            | 2012   | Bajo   |              | Cofre        |     | 2015 MARZO     |
| 6  | 0101000020E | 0            | 01010000201          | 2469176 | 17.2226698 | -100.630163  | 40900        | San Bartolo Centro                 | COLISION Y/C Auto   | ROJO              | 0      | \N     |              |              |     | 2015 ABRIL     |
| 7  | 0101000020E | 0            | 01010000201          | 2477686 |            |              | 0            | CUAUHTEMOC PROGRESO                | COLISION Y/C Auto   | VINO              | 2010   | Bajo   |              | Trasero      |     | 2015 ABRIL     |
| 8  | 0101000020E | 0            | 01010000201          | 2487269 |            |              | 0            | \N                                 | COLISION Y/C \N     | \N                | \N     | \N     | \N           | \N           |     | 2015 ABRIL     |
| 9  | 0101000020E | 0            | 01010000201          | 2498715 | 17.6525181 | -101.544271  | 40880        | Los Hujes                          | COLISION Y/C Auto   | blanco            |        | 2014   | Bajo         |              |     | 2015 ABRIL     |
| 10 | 0101000020E | 0            | 01010000201          | 2523510 |            |              | 0            | RUBI                               | COLISION Y/C Auto   | GRIS              |        | 2001   | Bajo         |              |     | 2015 MAYO      |
| 11 | 0101000020E | 0            | 01010000201          | 2576933 | 17.6397308 | -101.559088  | 40890        | Avenida 5 de Mayo                  | COLISION Y/C Auto   | BLANCO            | 2012   | \N     |              |              |     | 2015 MAYO      |
| 12 |             | 0            | 2664409              | 0       | 0          | 0            | ***AJUSTADO  | COLISION Y/C Camión                | BLANCO              | 2003              | Bajo   |        | Cofre        |              |     | 2015 JULIO     |
| 13 |             | 0            | 2688591              | 0       | 0          | 0            | Minerva      | COLISION Y/C Auto                  | blanco              | 2011              | \N     |        |              |              |     | 2015 JULIO     |
| 14 |             | 25           | 2728051              | 0       | 0          | 0            | ***PROV. 167 | COLISION Y/C Auto                  | ROJO                | 2002              | Bajo   |        | Frontal      |              |     | 2015 AGOSTO    |
| 15 | 0101000020E | 0            | 01010000201          | 2734920 | 16.8710351 | -99.9026716  | 39560        | CERRADA DE Morelos                 | COLISION Y/C Auto   | BLANCO            | 0      | \N     |              |              |     | 2015 AGOSTO    |
| 16 | 0101000020E | 0            | 01010000201          | 2909906 | 16.8830396 | -99.948917   | 0            | Calzada Piedra                     | COLISION Y/C Auto   | PLATA             | 0      | \N     |              |              |     | 2015 NOVIEMBRE |
| 17 | 0101000020E | 0            | 01010000201          | 2393302 |            |              | 0            | PDT                                | PDT                 | COLISION Y/C Auto | AZUL   | 2013   | \N           |              |     | 2015 FEBRERO   |
| 18 | 0101000020E | 0            | 01010000201          | 2313273 |            |              | 0            | FILIBERTO CH                       | COLISION Y/C Camión | BLANCO            | 2007   | Alto   |              |              |     | 2015 ENERO     |
| 19 | 0101000020E | 0            | 01010000201          | 2321543 | 23.3456346 | -102.992996  | 0            | CARR. Fresnillo                    | COLISION Y/C Auto   | AZUL              | 2007   | Bajo   |              |              |     | 2015 ENERO     |
| 20 | 0101000020E | 0            | 01010000201          | 2330648 | 22.7691084 | -102.613635  | 0            | Antigua Carretera de San Francisco | COLISION Y/C Camión | AZUL              | 2004   | Bajo   |              |              |     | 2015 ENERO     |
| 21 | 0101000020E | 0            | 01010000201          | 2364033 |            |              | 0            | AJUSTA IVAN                        | COLISION Y/C Camión | BLANCO            | 1985   | \N     |              |              |     | 2015 FEBRERO   |
| 22 | 0101000020E | 0            | 01010000201          | 2455312 |            |              | 0            | JONATHAN IS                        | COLISION Y/C Camión | BLANCO            | 2012   | \N     |              |              |     | 2015 MARZO     |
| 23 | 0101000020E | 0            | 01010000201          | 2478124 | 22.7721979 | -102.572084  | 98000        | AVENIDA Gue Centro                 | COLISION Y/C Auto   | ROJO              | 2014   | Bajo   |              |              |     | 2015 ABRIL     |
| 24 | 0101000020E | 3            | 01010000201          | 2487761 | 21.8610507 | -102.301265  | 20235        | Av. Las Americas El Dorado 1a      | COLISION Y/C Auto   | TINTO             | 2014   | Bajo   |              | Trasero      |     | 2015 ABRIL     |
| 25 | 0101000020E | 0            | 01010000201          | 2512283 |            |              | 0            | ***PROV 1510 OLIVARES              | COLISION Y/C Auto   | ORO               | 2001   | \N     |              |              |     | 2015 ABRIL     |
| 26 | 0101000020E | 0            | 01010000201          | 2519251 |            |              | 0            | PDT                                | PDT                 | COLISION Y/C Auto | BLANCO | 2012   | \N           |              |     | 2015 ABRIL     |
| 27 | 0101000020E | 0            | 01010000201          | 2536564 |            |              | 0            | 20 DE NOVIEMBRE                    | COLISION Y/C Auto   | BLANCO            | 2014   | Bajo   |              | Trasero      |     | 2015 MAYO      |
| 28 | 0101000020E | 0            | 01010000201          | 2685159 | 28.6465406 | -106.125227  | 31217        | Mirador                            | COLISION Y/C Auto   | GRIS              | 2005   | \N     |              |              |     | 2015 JULIO     |
| 29 | 0101000020E | 0            | 01010000201          | 2558239 | 21.8584338 | -102.308039  | 20285        | Aguascaliente Versalles 2a         | COLISION Y/C Auto   | BLANCO            | 0      | \N     |              |              |     | 2015 MAYO      |
| 30 | 0101000020E | 0            | 01010000201          | 2619517 | 23.634501  | -102.552784  | 0            | \N                                 | \N                  | \N                | \N     | \N     | \N           | \N           |     | 2015 JUNIO     |

Se aprecia que los campos no tiene un formato estandarizado. Por lo cual se procede a darle un formato. Para ello se utiliza el lenguaje de programación Python.

Los pasos a seguir en la solución fueron los siguientes



a) Lectura de archivo csv.

Se toma la información recolectada de la empresa en bruto para poder apreciarla y darse un panorama amplio del archivo.

b) Formato a archivo csv.

Una vez que se tiene un panorama amplio del negocio, se procede a depurar los datos. Se eliminan caracteres inconsistentes en los campos de información, tales como:

- Espacios en blanco
- Valores 0
- Lectura incorrecta de acentos ortográficos.
- Diagonales

Posteriormente se ordena la información y se colocan los caracteres correctos apareciendo en el siguiente enlace (este también se muestra en el repositorio pero por cuestiones de tamaño, no se pudo subir tal cual el archivo):

<https://drive.google.com/file/d/1tguwWKLJ7lkWfzocrx8hISQDXkGAh8C7/view?usp=sharing>

c) Almacenamiento en una BD.

Una vez limpiados y esquematizados estos archivos, se proceden a alojar en una base de datos manejando un lenguaje SQL.

d) Almacenamiento en un repositorio.

Se procede a crear un repositorio en <https://github.com>. En este, se alojarán los archivos anteriormente mencionados.

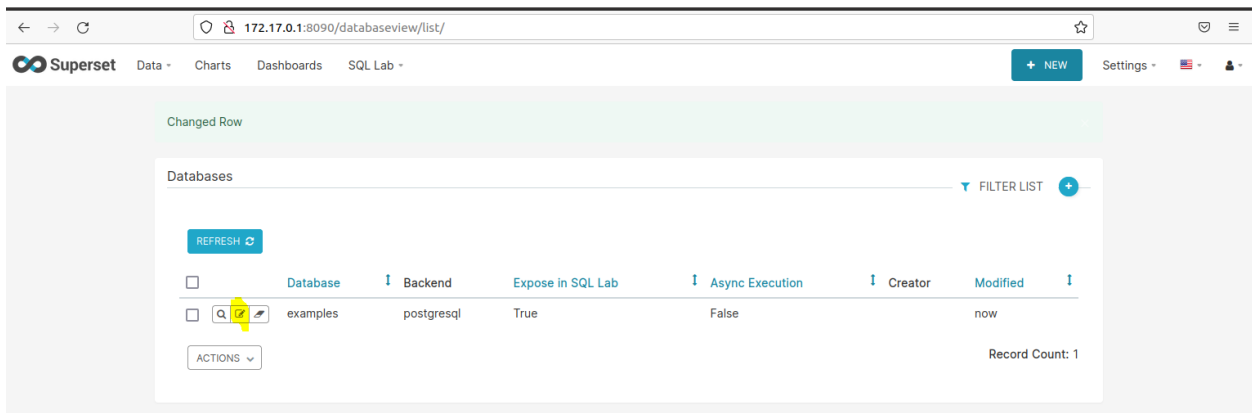
e) Análisis de información.

Esta petición se lleva con la finalidad de que AXA pueda tener un buen análisis de su información para la toma de decisiones.

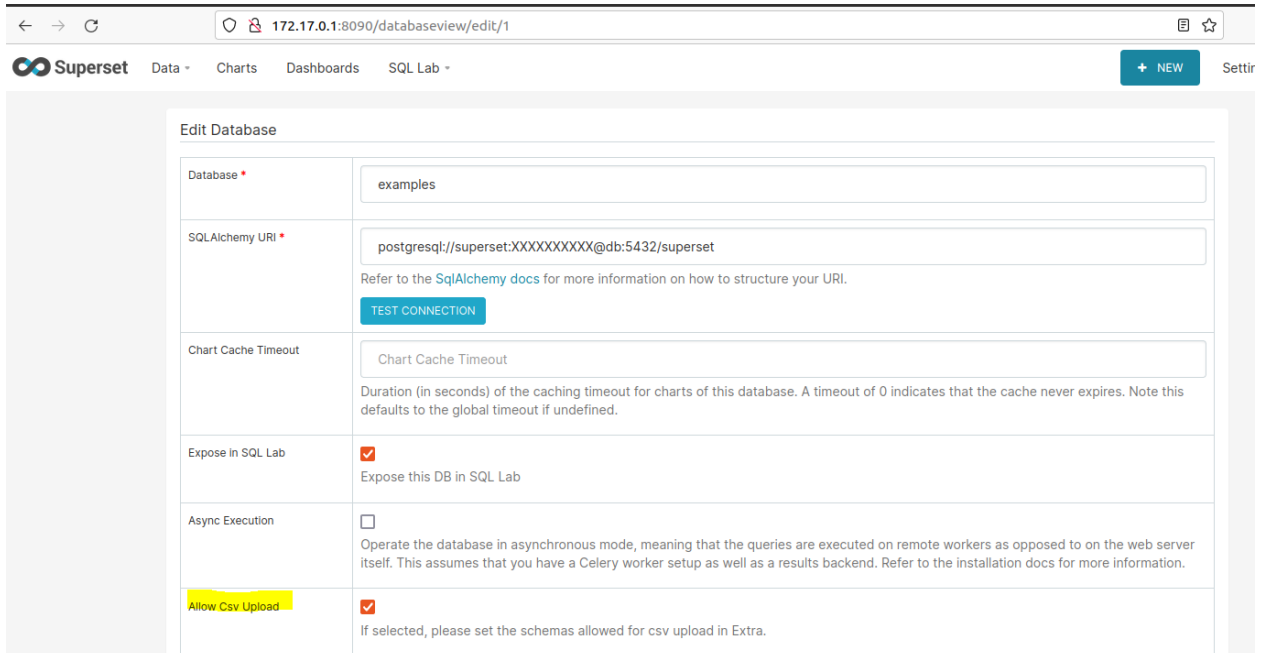
Una vez concluida la tarea, se comparten los resultados con el cliente esperando su completa satisfacción y el equipo de everis queda atento al mantenimiento y futuras peticiones.

## Instrucciones

1. Descargar el repositorio mediante [https://github.com/MVizcayaR/ProyectoAXA\\_Hands\\_on.git](https://github.com/MVizcayaR/ProyectoAXA_Hands_on.git) en un sistema operativo Linux
2. Ingresar a la carpeta de superset mediante la terminal y ejecutar las siguientes instrucciones:
  - `sudo apt update`
  - `sudo apt upgrade`
  - `sudo apt install vim`
  - `sudo apt install docker.io`
  - `sudo apt install docker-compose`
  - `git checkout 0.38`
  - `sudo docker-compose up`
3. Una vez que se haya montado Docker compose, se podrá acceder al entorno Superset de Apache mediante un explorador de internet con la liga <https://{Coloque su ip aquí}:8090>
4. Modifique la database "example" para que pueda ingresar archivos .csv como se muestra en las siguientes figuras:

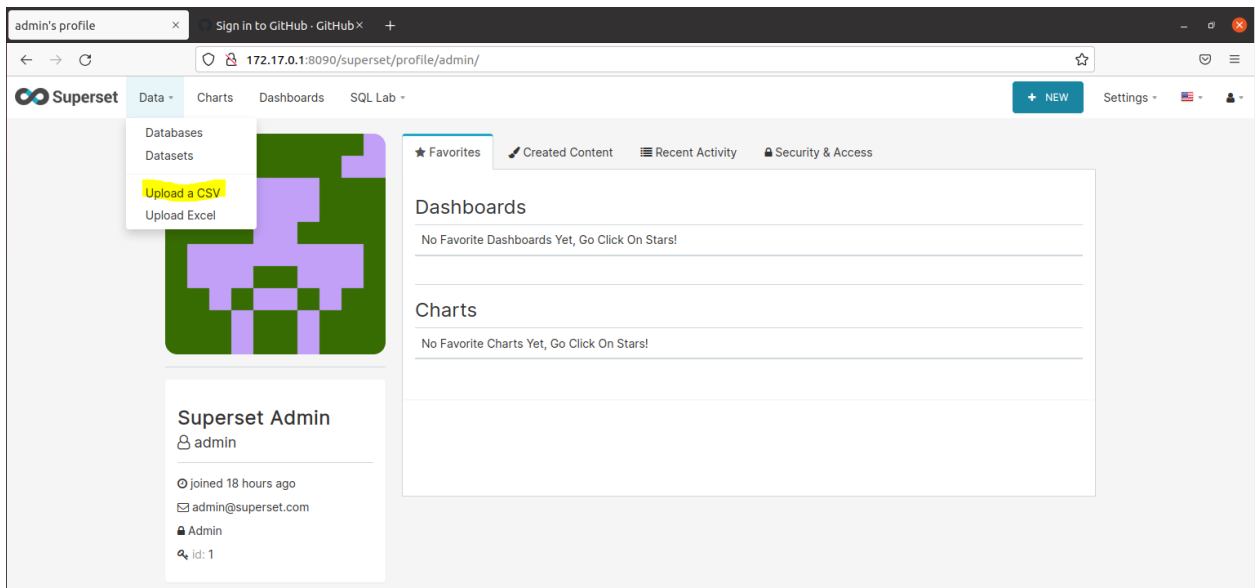


## Seleccionar editar base de datos



Marcarla casilla “allow csv upload”

- Posteriormente se carga el archivo “Incidentes 2020-j.csv” como se muestra en las siguientes figuras:



En la sección Data, dar click en “Upload a csv”

| CSV to Database configuration |   |
|-------------------------------|---|
| Table Name *                  | <input type="text" value="YourTable"/><br>Name of table to be created from csv data.  |
| CSV File *                    | <input type="button" value="Examinar..."/> Ningún archivo seleccionado.<br>Select a CSV file to be uploaded to a database.                                    |
| Database                      | <input type="text" value="examples"/>   |
| Schema                        | <input type="text" value="Schema"/><br>Specify a schema (if database flavor supports this).   |
| Delimiter *                   | <input type="text" value=","/><br>Delimiter used by CSV file (for whitespace use \s+).  |
| Table Exists *                | <input type="button" value="Fail"/><br>If table exists do one of the following: Fail (do nothing), Replace (drop and recreate table) or Append (insert data). |
| Header Row                    | <input type="text" value="0"/><br>Row containing the headers to use as column names (0 is first line of data). Leave empty if there is no header row.         |
| Index Column                  | <input type="text" value="Index Column"/><br>Column to use as the row labels of the dataframe. Leave empty if no index column.                                |

Posteriormente llenar el campo de nombre de la tabla, adjuntar el archivo y dar click en “save”

- Una vez tenemos cargada a base de datos, podemos empezar a generar figuras y consultas mediante las pestañas Charts para generar gráficas, SQL Lab para

realizar consultas y también se puede crear un Dashboard para presentar las

