**Actividad 1**



**Actividad 1**

Jimmy Saúl Yar Carvajal

Master Universitario en Desarrollo de Aplicaciones y Servicios Web,

Universidad Internacional de Valencia

Ing. David Monzó Ferrer

Ing. Javier Oliver Moll

Octubre, 2023

1. **Introducción**

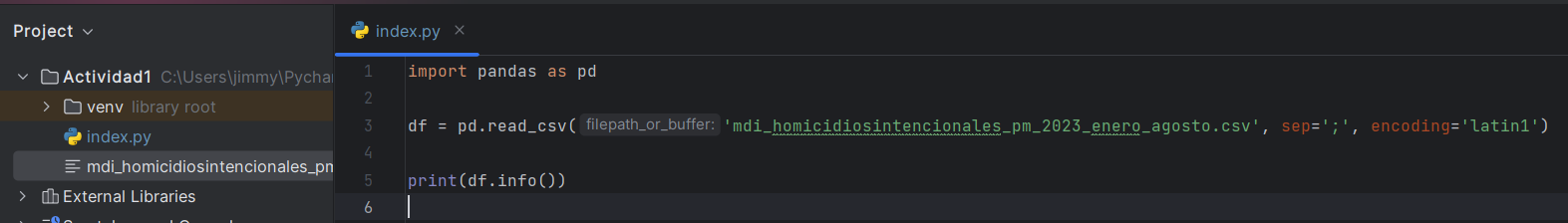
En la elaboración del panel de control, seleccione los datos abiertos de Ecuador. En ellos, descubrí una serie de información relevante para analizar las provincias con mayor incidencia de homicidios intencionales durante el periodo de enero a agosto de 2023. Esto permitirá identificar las zonas de mayor riesgo, brindando a los turistas la posibilidad de evitar estos lugares donde la seguridad ciudadana presenta desafíos. Asimismo, se facilitará la visualización de las diversas causas de homicidio, los lugares donde ocurrieron, así como detalles sobre la nacionalidad y el género de las víctimas.

1. **Procesos ETL**
   1. **Fuente de datos**

La fuente de datos de la que se obtuvo es del sitio oficial de [Datos abiertos Ecuador](https://www.datosabiertos.gob.ec/dataset/homicidios-intencionales/resource/cb8f704e-2b27-4d7f-9431-d40c4e27fa48), los datos a los cuales se van a realizar el proceso de extraer, transformar y cargar, son los registros de datos del ministerio de gobierno de los homicidios desde enero a septiembre.

* 1. **Ingesta**

Primero lo que hacemos es descargarnos el archivo CSV del sitio oficial de [Datos abiertos Ecuador](https://www.datosabiertos.gob.ec/dataset/homicidios-intencionales/resource/cb8f704e-2b27-4d7f-9431-d40c4e27fa48), con la ayuda de un script en Python vamos a leer el archivo como se observa en la imagen 1.

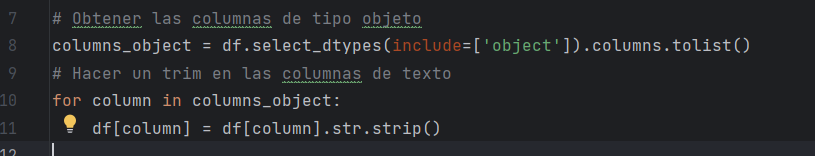


Con el comando df.info() podemos visualizar en la imagen 2 la información en general del dataframe.

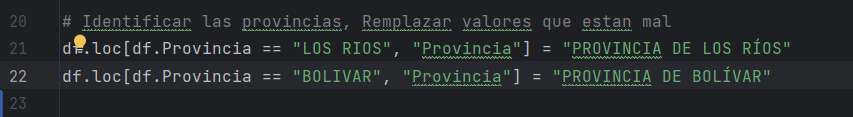


* 1. **Limpieza de datos**

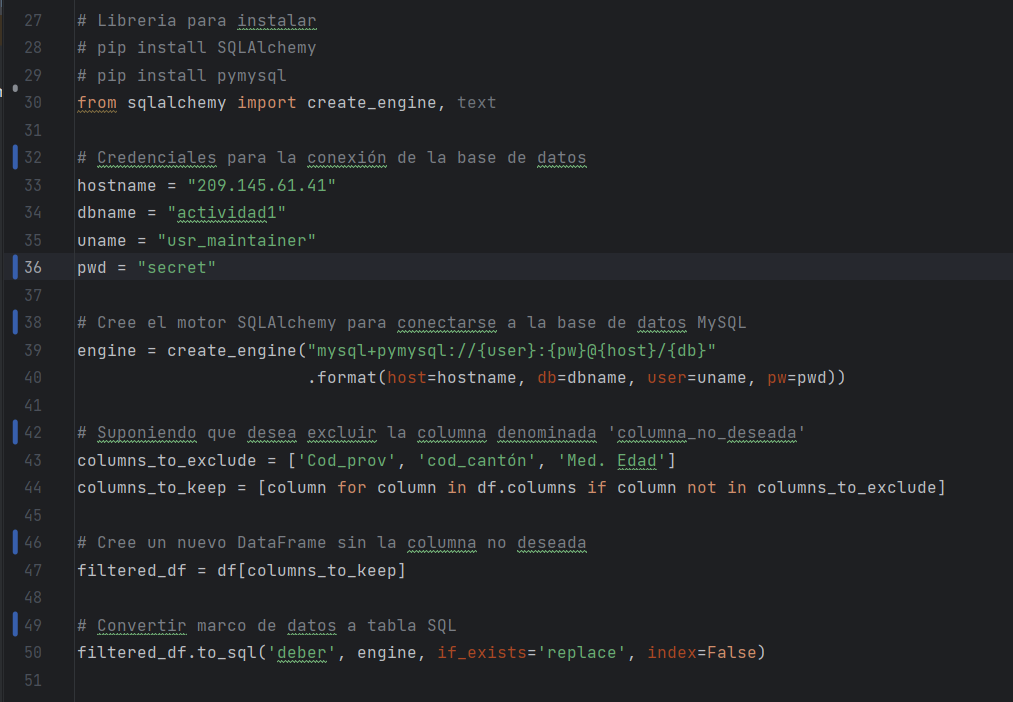
A Continuación, vamos a limpiar los campos de tipo texto que contengan espacios en blanco, como se visualiza en la imagen 3.



Vamos a trasformar los valores de la columna provincia, los valores Los Ríos y Bolívar, por el Provincia de los Ríos y Provincia de Bolívar, como se visualiza en la imagen 4.

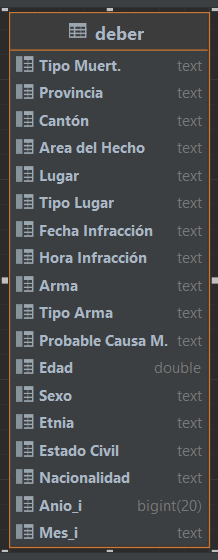


Procederemos a subir el dataframe a una base de datos Mysql, debemos instalar las librerías (pymysql, SQLAlchemy), creamos una cadena de conexión, después excluimos los campos que no queremos cargar a la tabla deber, por último, volvemos a crear un dataframe con las columnas que si consumiremos en Looker Studio y subimos a la base de datos como se visualiza en la imagen 5.



* 1. **Modelo de datos**

Por último, se creó una tabla única denominada deber, que contiene 18 columnas: tipo de muerte, provincia, cantón, área del hecho, lugar, tipo lugar, fecha de la infracción, hora de la infracción, arma, tipo arma, problema causa de muerte, edad, sexo, etnia, estado civil, nacionalidad, año y mes del homicidio. En la imagen 6 se visualiza la estructura de la tabla deber.



1. **Filtros aplicados sobre las variables**

Los filtros que se aplicaron son tres, el primer filtro que se realizo es por edad que parte desde 0 a 91 años, el segundo filtro realizado es por el tipo de sexo que están separados por tres (Mujer, Hombre, No determinado), y el ultimo filtro realizado es por arma que se cometió el homicidio. Adicional se colocó una tarjeta de resultado de las muertes totales. En el primer grafico se insertó un gráfico de pastel con los valores del porcentaje de total de las etnias que fallecieron. El segundo grafico es una tabla mostrando el lugar del homicidio y el total de los fallecidos en los lugares. El tercer grafico es un gráfico de barras con el total de fallecidos por provincias. El ultimo grafico de mapa muestra de que nacionalidad eran los fallecidos. A continuación, se visualiza en la imagen 7 el resultado del Dashboard.



**Link de acceso al cuadro de mando**

<https://lookerstudio.google.com/reporting/fd59524e-9ac6-4edc-b6ee-61ee6242aaa8>

1. **Conclusiones**

* Podemos concluir que la región Costa tiene una alta tasa de homicidios.
* También notamos que existe más muertes de hombres respecto a las mujeres.
* La mayoría de los homicidios donde ocurrieron se produjeron en la vía pública seguida de las viviendas privadas.
* La mayor parte de los fallecidos son de Etnia mestizos y en segundo lugar las personas afro.

1. **Recomendaciones**

* Se recomienda leer la información proporcionada por los sitios web de datos abiertos donde se resume la información detallada de cada uno de los campos.
* Es necesario realizar una limpieza de los datos porque puede haber inconsistencia de los datos, pueden estar mal ingresados o ingresados valores nulos.
* Se recomienda mostrar la información con los gráficos indicados, para una mejor comprensión para que no se requiera mayor explicación, por ejemplo, es recomendable los gráficos de barras al momento de comparar datos