# Prueba Técnica Python

1. Se solicita realizar el siguiente ejercicio (preferiblemente en Python versión 3.7(o superior) o en su defecto 2.7, se entregan dos bases de datos, una en Excel y otra en Texto Plano, para estos ejercicios puede utilizar las librerías que considere necesarias:
   1. Leer la base **ejercicio\_b1.xlsx** Hoja1. De esta base leer todos los campos disponibles en la variable **exceldf**. **Nota:** Se recomienda abrir el Excel y entender los datos.
   2. Eliminar registros (filas) nulos y duplicados de la columna **‘id’**
   3. De la base de Excel queremos conocer: ¿Cuántas compras de cada tipo de pedido ha realizado cada usuario? Las columnas de esta salida deben ser:

|  |  |  |
| --- | --- | --- |
| CEDULA | Tipo de pedido | Cantidad número de pedido |
| Cédula de la persona | Descripción del tipo de pedido realizado (siempre debe ser una palabra) | Conteo de pedidos realizados por cada usuario. |

* 1. Renombrar el campo **cc\_cliente** a **CEDULA**
  2. Nombrar esta base **grouped\_excel\_data**
  3. Leer la base de texto plano llamada **ejercicio\_b3.txt** en la variable **textdf**.
  4. Crear la columna **NOMBRE COMPLETO** en la cual deben de aparecer el nombre concatenado con apellido separado por espacios, de la siguiente manera:

|  |  |  |
| --- | --- | --- |
| NOMBRE | APELLIDO | NOMBRE COMPLETO |
| ANDRES | VALBUENA | Andres Valbuena |

Nótese la primer letra mayúscula en el nombre y el apellido.

* 1. Crear la columna **EDAD** que será la diferencia en años que existe entre **NACIMIENTO**  y la fecha “**25/11/1988**”
  2. Cruzar ambas bases (**grouped\_excel\_data, textdf)** por la columna **CEDULA** dando mayor importancia a **textdf** coincidentes (left join) y guardar los resultados en la variable **outputdf**
  3. Crear un archivo de texto de salida con el resultado del ejercicio **outputdf.**

1. Dada la data del archivo **ejercicio2.json** crear un gráfico (el que mejor le parezca) y entregar los valores estadísticos del dataframe (suma, media, etc).
2. Cree dos dataframes de (100 filas x 10 columnas) cada uno con las siguientes características:
   1. las primeras 3 columnas deben ser categóricas ‘valores de texto’ con las siguientes letras: {‘a’,’b’,’c’,’d’}
   2. Las siguientes 7 columnas valores numéricos enteros valores entre 0 y 100
   3. Se debe crear una función que reciba los dos dataframes e identifique que tan diferente un dataframe del otro (puede ser en porcentaje o valor como prefiera)
   4. Ejecute la función anterior recibiendo los mismos dataframes

Al finalizar esta sección favor guardar el código en un archivo .py como ejercicio1.py, ejercicio2.py y ejercicio3.py respectivamente.