

## **Proyecto Final Introducción a Python para Ciencia de Datos**

Para este proyecto requerirá haber alcanzado la totalidad de las clases del curso Python para Ciencia de Datos nivel Introductorio.

### **Etapas 1. (Búsqueda)**

a) Busca bases de datos que te permitan realizar uniones o combinaciones por medio de sentencias vistas en clase (mínimo 2 bases de datos).

### **Etapas 2 (Carga y exploración)**

- a) Realiza la importación de los paquetes a utilizar en el proyecto
- b) Realiza la carga de las bases de datos a utilizar en el proyecto.
- c) Aplica al menos 5 técnicas de exploración de datos y confirma que tus datos están cargados correctamente.

### **Etapas 3 (Manipulación)**

- a) Selecciona las variables (columnas) con la cual trabajarás tu análisis.
- b) Valida el tipo de dato que contiene las columnas y realiza los métodos de reemplazar o limpiar caracteres especiales tales como "espacios", "símbolo moneda", etc.
- c) Revisa que la columna de fecha tenga el formato correspondiente y utiliza la función de conversión al formato '%Y-%m-%d %H:%M:%S'.
- d) Extraiga de su columna de Fechas, la información referente al Año, Mes y Nombre de Mes.
- e) Realiza la combinación o unión de las bases de datos mediante los métodos de combinación de DataFrames (Merge, concat, etc)

### **Etapas 4 (Limpieza)**

- a) Consulta cuáles son tus valores vacíos (columnas y cantidad)
- b) Reemplazar los valores vacíos de tus datos numéricos por 0 o bien elimina la línea si así te permite realizar un mejor análisis de la información.
- c) Valida los valores duplicados de tu base de datos y considera si es necesario eliminar en caso de encontrar un hallazgo o bien conservar los valores.
- d) Si tienes los campos con resultados finales ejemplo la Venta (utiliza el método de validación a nivel de fila para entender si la operación aritmética es correcta y el resultado expresado en la base está bien)

### **Etapas 4 (Simplificar)**

- a) Realiza 3 agrupaciones en las bases de datos cada una con diferente sentido en la información a brindar (para ellos debe utilizar los métodos de agregación y agrupación vistos en clase).

Ejemplo:

Agrupacion1 = Año, Mes – Venta y Cantidad

Agrupacion2 = Categoría-Subcategoría Venta y Cantidad

Agrupación 3 = Región-Cliente Compras

Ordena cada agrupación según algún campo numérico y resetea los índices.

### **Etapas 5 (Visualización)**

- a) Para cada una de las agrupaciones creadas previamente, realiza gráficos para representar la información y explicar las conclusiones de la información extraída. Se claro y trata de contar una historia con los datos.

Nota:

✓ Deberá realizar la búsqueda de la información en bases de datos en la empresa para la cual labora (ideal) o bien consultar en los enlaces adjuntos (Debe cumplir con requisitos tales como, tener una columna de fecha, información transaccional- venta e información categórica)

<https://dataverse.harvard.edu/>

<https://www.kaggle.com/datasets>

✓ El código deberá estar escrito en un archivo formato ipynb (notebook Python), y deberá de traer de una forma explícita el paso a paso utilizado con sus respectivos print en los resultados obtenidos (indispensable)

✓ Para cada punto solicitado debes incluir un comentario o explicación sobre qué es lo que hace cada método utilizado en su análisis.

*Ejemplo:*

#Pandas es el paquete que nos permite trabajar con dataframes de forma eficiente

```
import pandas as pd
```

# df = pd.read\_csv() #importamos los datos "x" que tratan de "x" tema. Y de esta forma para cada paso realizado.

✓ El proyecto debe ser compartido al correo [reportes@growupcr.com](mailto:reportes@growupcr.com)