

Laboratorio 11: Pandas II

Se tienen los datos personales de clientes de una tienda comercial en el archivo **clientes.csv** y sus respectivas cantidades de órdenes y monto de consumo en el archivo **consumo.csv**

Estructura del archivo consumos:

id_cliente: Código identificador de cliente
ordenes: Cantidad total de ordenes de pedido
gastado: Monto total gastado por el cliente en dólares

1. Cargue el conjunto de datos **consumo.csv** usando la función `read_csv()` usando el parámetro `sep` teniendo como separador el punto y coma (;), almacenando los datos en un dataframe llamado **consumo**.
2. Usando el dataframe consumo, obtenga los consumos repetidos usando la función `uplicated()`.
3. Luego elimine las filas repetidas del dataframe consumo obtenidas anteriormente usando la función `drop_duplicates(inplace=True)`.
4. Luego establezca como índice a la columna `id_cliente` para el dataframe consumo usando la función `set_index("columna")`.
5. Utilizando los datos sin filas repetidas,
 - a. Obtener los clientes que han realizado igual a 20 órdenes de pedido usando la función `loc`.
 - b. Obtener los clientes que han gastado menos de 600 dólares usando la función `loc`.
 - c. Obtener los clientes que han gastado más de 200 dólares y han realizado menos de 60 órdenes de pedido usando la función `loc`.
 - d. Obtener el promedio del monto gastado por todos los clientes usando la función `mean()`.
 - e. Obtener la varianza del monto gastado por todos los clientes usando la función `var()`.
 - f. obtener la desviación estándar del monto gastado por todos los clientes usando la función `std()`.

Estructura del archivo clientes:

id_cliente: Código identificador de cliente
nombre: Nombre del cliente
apellido: Apellido del cliente
email: Email del cliente
teléfono: Teléfono de contacto del cliente
direccion: Dirección del cliente
sexo: Sexo del cliente
edad: Edad del cliente
registrado: Fecha de registro del cliente
empleo: Ocupación o profesion que posee el cliente
hobbies: Hobbie del cliente
es_casado: Estado civil del cliente

6. Cargue el conjunto de datos **clientes.csv** usando la función `read_csv()` usando el parámetro `sep` teniendo como separador el punto y coma (;), estableciendo como índice a la columna `id_cliente` y almacenando los datos en un dataframe llamado **clientes**.
Nota: para colocar el índice se usa la función `set_index("columna")`.

7. Usando el dataframe `clientes`, Obtener el nombre, apellido y el sexo de aquellas personas que tengan un empleo con valor nulo (`NaN`) usando la función `isna()`.
8. Luego elimine las filas que contengan valores nulos usando la función `dropna(inplace=True)`.
9. Utilizando los datos sin las filas que contienen valores nulos
 - a. Obtener los empleos que tienen los clientes usando la función `unique()`.
 - b. Obtener los hobbies que tienen clientes usando la función `unique()`.
 - c. Obtener el nombre, apellido y empleo de aquellos clientes que tienen como hobby cantar y son casados usando la función `loc`.
 - d. Obtener el nombre, apellido e email de aquellos clientes que son menores de 40 y no son casados usando la función `loc`.
 - e. Obtener la edad máxima de todos hombres que tienen como hobby la costura usando la función `max()`.
 - f. Obtener la edad mínima de todas las mujeres casadas usando la función `min()`.
10. Unir los dataframe ***clientes*** con el de dataframe ***consumo*** usando un ***join externo*** para crear el dataframe ***cliente_consumo*** usando como base el dataframe `clientes`.
Nota: usar la función `join` usando el parametro `how="outer"`.
11. Luego eliminar las filas con valores nulos (`NaN`) creados a partir del join externo usando la función `dropna(inplace=True)`.
12. A partir del dataframe obtenido después de la eliminación de valores nulos,
 - a. Obtener el promedio de lo gastado por aquellos clientes tengan como empleo artista y sean hombres usando la función `loc` y la funcion `mean()`.
 - b. Obtener el promedio de edad de aquellos clientes que tengan como empleo bombero y sean mujeres usando la función `loc` y la funcion `mean()`.
 - c. Obtener la cantidad de clientes por empleo usando la función `value_counts()`.
 - d. Obtener la proporción de clientes que tengan como empleo Camarero.
 - e. Obtener la proporción de clientes que tengan como empleo Docente.