PEDRO HERRERA DIAZ

MATERIA: INTRODUCCION A LA CIENCIA DE DATOS

NOMBRE DEL PROFESOR

JAIME ALEJANDRO ROMERO SIERRA

20/10/2025

https://github.com/user-attachments/files/23013583/ObesityDataSet_raw_and_data_sinthetic.csv

Esta base de datos contiene información sobre la obesidad

En esta base de datos encontramos diferentes columnas, como lo son genero, edad, altura, peso, antecedentes, familiares con sobre peso.

GENERO: masculino o femenino

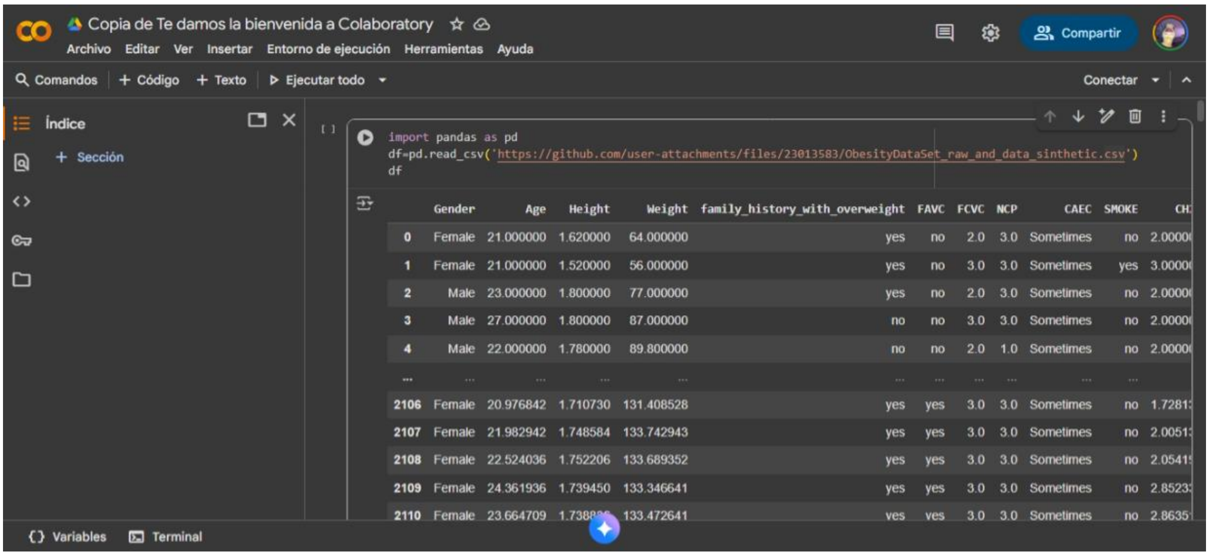
Edad:

Altura:

Peso:

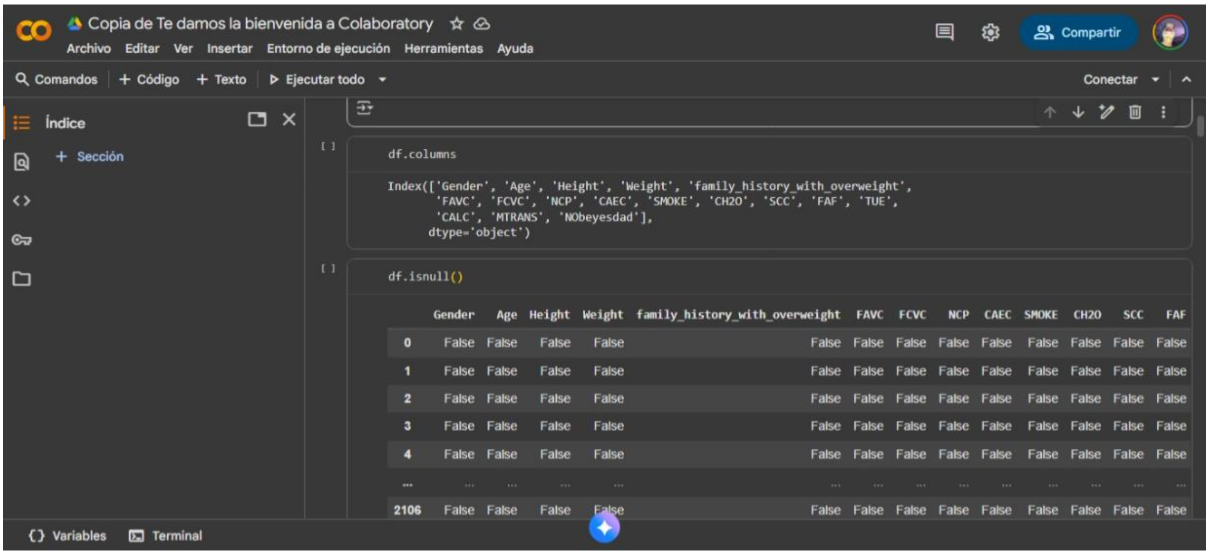
Antecedentes:

Familiares con sobre peso:


Copia de Te damos la bienvenida a Colaboratory
Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda
Compartir
Comandos + Código + Texto ▷ Ejecutar todo
Conectar
Índice
+ Sección
Variables
Terminal

```
import pandas as pd
df=pd.read_csv('https://github.com/user-attachments/files/23013583/ObesityDataSet_raw_and_data_sinthetic.csv')
df
```

	Gender	Age	Height	Weight	family_history_with_overweight	FAVC	FCVC	NCP	CAEC	SMOKE	CH
0	Female	21.000000	1.620000	64.000000	yes	no	2.0	3.0	Sometimes	no	2.0000
1	Female	21.000000	1.520000	56.000000	yes	no	3.0	3.0	Sometimes	yes	3.0000
2	Male	23.000000	1.800000	77.000000	yes	no	2.0	3.0	Sometimes	no	2.0000
3	Male	27.000000	1.800000	87.000000	no	no	3.0	3.0	Sometimes	no	2.0000
4	Male	22.000000	1.780000	89.800000	no	no	2.0	1.0	Sometimes	no	2.0000
...	...	...	...	...	...	...	...	...	...	...	
2106	Female	20.976842	1.710730	131.408528	yes	yes	3.0	3.0	Sometimes	no	1.7281
2107	Female	21.982942	1.748584	133.742943	yes	yes	3.0	3.0	Sometimes	no	2.0051
2108	Female	22.524036	1.752206	133.689352	yes	yes	3.0	3.0	Sometimes	no	2.0541
2109	Female	24.361936	1.739450	133.346641	yes	yes	3.0	3.0	Sometimes	no	2.8523
2110	Female	23.664709	1.7388	133.472641	yes	yes	3.0	3.0	Sometimes	no	2.8635

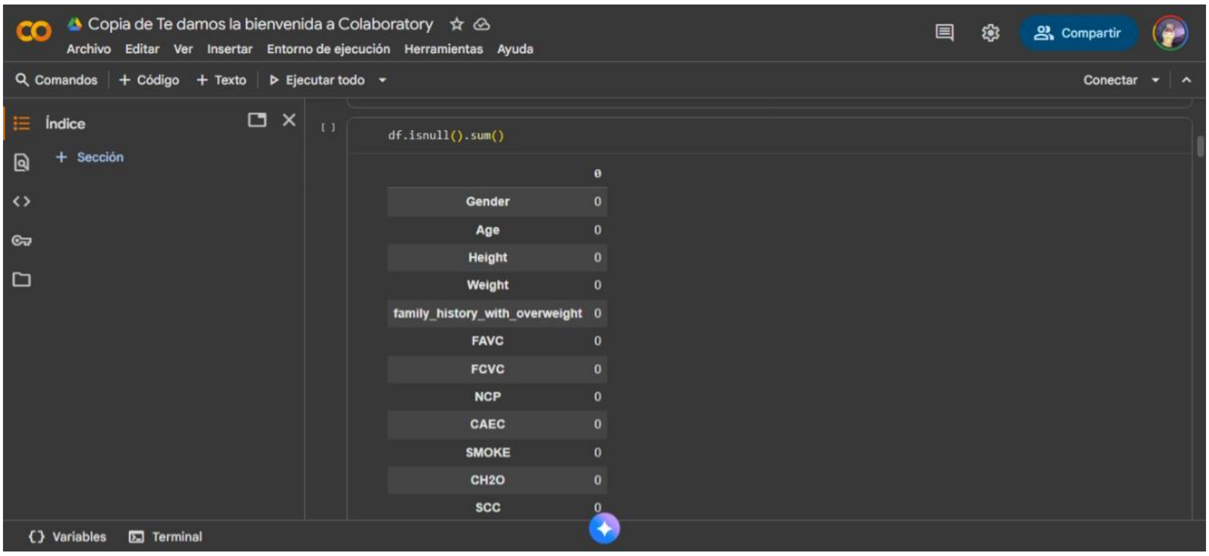

Copia de Te damos la bienvenida a Colaboratory
Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda
Compartir
Comandos + Código + Texto ▷ Ejecutar todo
Conectar
Índice
+ Sección
Variables
Terminal

```
df.columns
```

```
Index(['Gender', 'Age', 'Height', 'Weight', 'family_history_with_overweight',
       'FAVC', 'FCVC', 'NCP', 'CAEC', 'SMOKE', 'CH2O', 'SCC', 'FAF', 'TUE',
       'CALC', 'MTRANS', 'NObeyesdad'],
      dtype='object')
```

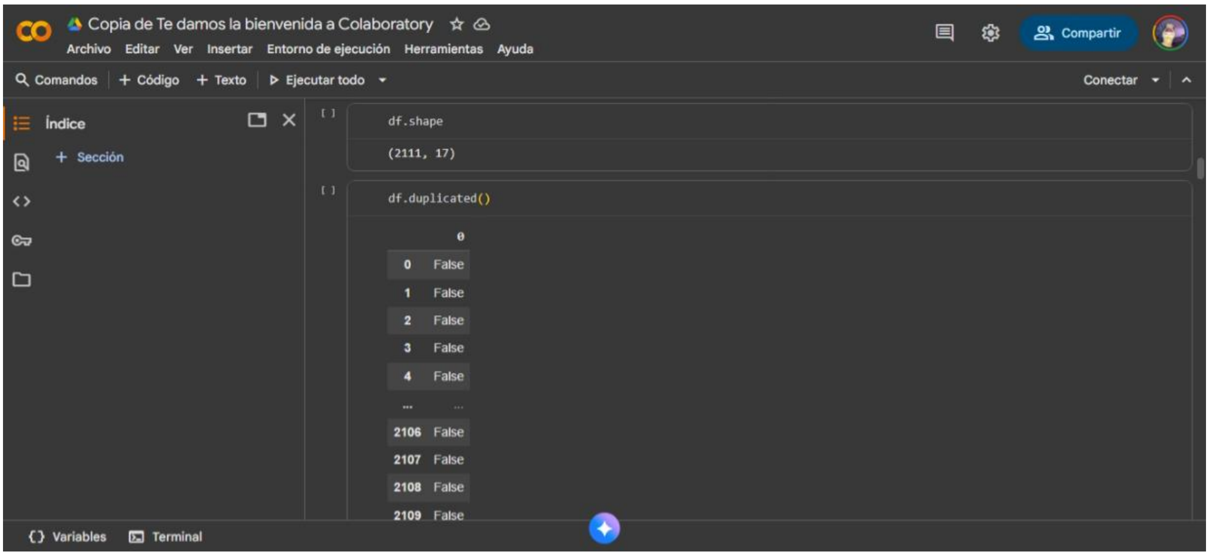
```
df.isnull()
```

	Gender	Age	Height	Weight	family_history_with_overweight	FAVC	FCVC	NCP	CAEC	SMOKE	CH2O	SCC	FAF
0	False	False	False	False	False	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False	False	False	False	False
...	...	...	...	...	...	...	...	...	...	...	...	...	...
2106	False	False	False	False	False	False	False	False	False	False	False	False	False


Copia de Te damos la bienvenida a Colaboratory
Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda
Compartir
Comandos + Código + Texto ▷ Ejecutar todo
Conectar
Índice
+ Sección
Variables
Terminal

```
df.isnull().sum()
```

	0
Gender	0
Age	0
Height	0
Weight	0
family_history_with_overweight	0
FAVC	0
FCVC	0
NCP	0
CAEC	0
SMOKE	0
CH2O	0
SCC	0

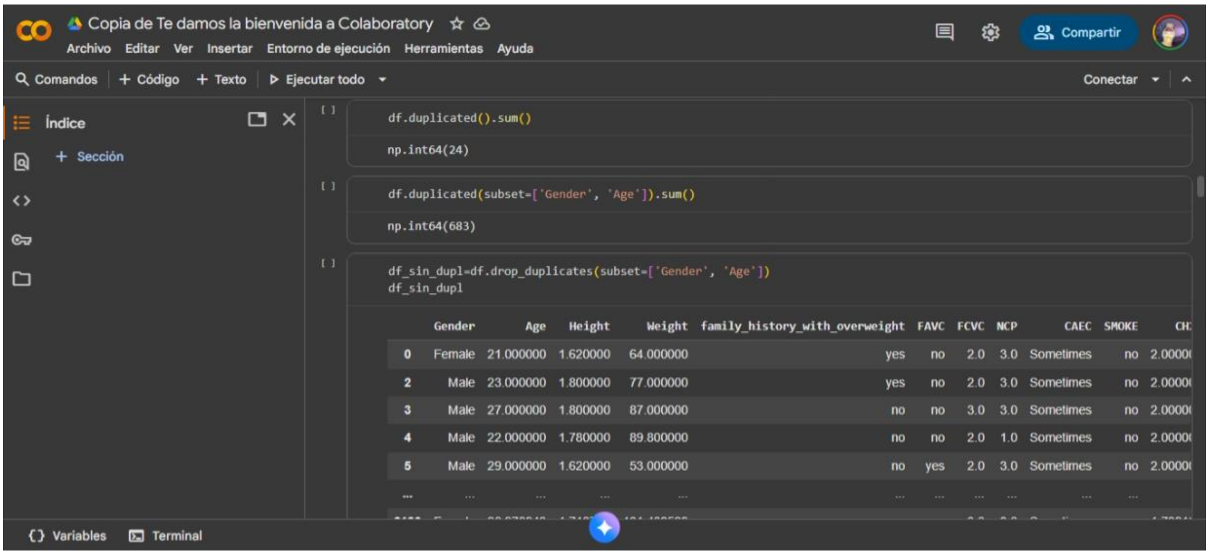

Copia de Te damos la bienvenida a Colaboratory
Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda
Compartir
Comandos + Código + Texto ▷ Ejecutar todo
Conectar
Índice
+ Sección
Variables
Terminal

```
df.shape
```

```
(2111, 17)
```

```
df.duplicated()
```

	0
0	False
1	False
2	False
3	False
4	False
...	...
2106	False
2107	False
2108	False
2109	False


Copia de Te damos la bienvenida a Colaboratory
Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda
Compartir
Comandos + Código + Texto ▷ Ejecutar todo
Conectar
Índice
+ Sección
Variables
Terminal

```
df.duplicated().sum()
```

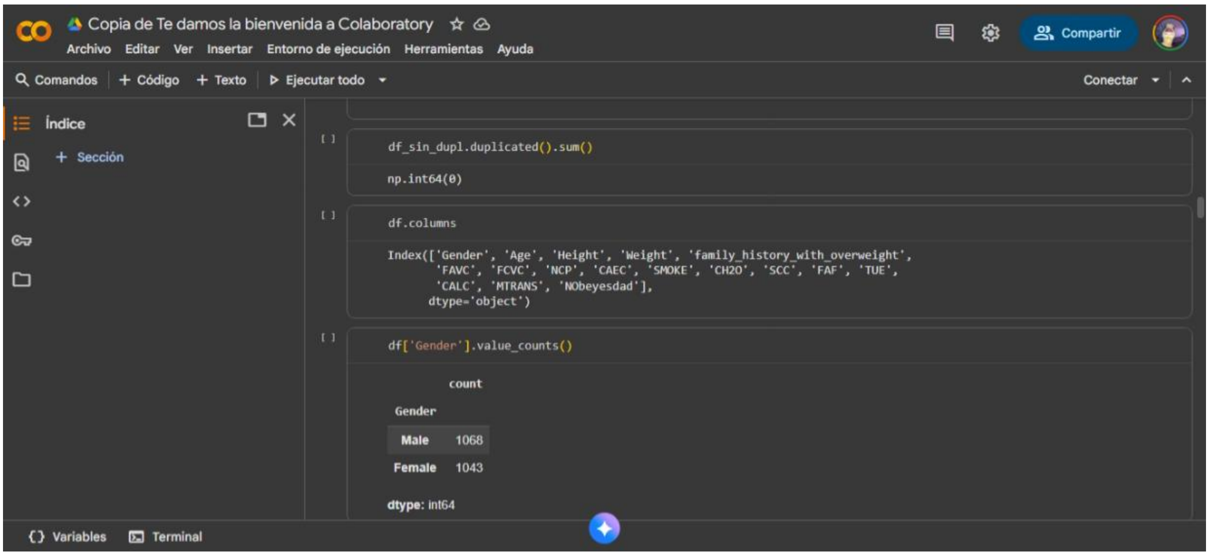
```
np.int64(24)
```

```
df.duplicated(subset=['Gender', 'Age']).sum()
```

```
np.int64(683)
```

```
df_sin_dupl=df.drop_duplicates(subset=['Gender', 'Age'])
df_sin_dupl
```

	Gender	Age	Height	Weight	family_history_with_overweight	FAVC	FCVC	NCP	CAEC	SMOKE	CH
0	Female	21.000000	1.620000	64.000000	yes	no	2.0	3.0	Sometimes	no	2.0000
2	Male	23.000000	1.800000	77.000000	yes	no	2.0	3.0	Sometimes	no	2.0000
3	Male	27.000000	1.800000	87.000000	no	no	3.0	3.0	Sometimes	no	2.0000
4	Male	22.000000	1.780000	89.800000	no	no	2.0	1.0	Sometimes	no	2.0000
5	Male	29.000000	1.620000	53.000000	no	yes	2.0	3.0	Sometimes	no	2.0000
...	...	...	...	...	...	...	...	...	...	...	


Copia de Te damos la bienvenida a Colaboratory
Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda
Compartir
Comandos + Código + Texto ▷ Ejecutar todo
Conectar
Índice
+ Sección
Variables
Terminal

```
df_sin_dupl.duplicated().sum()
```

```
np.int64(0)
```

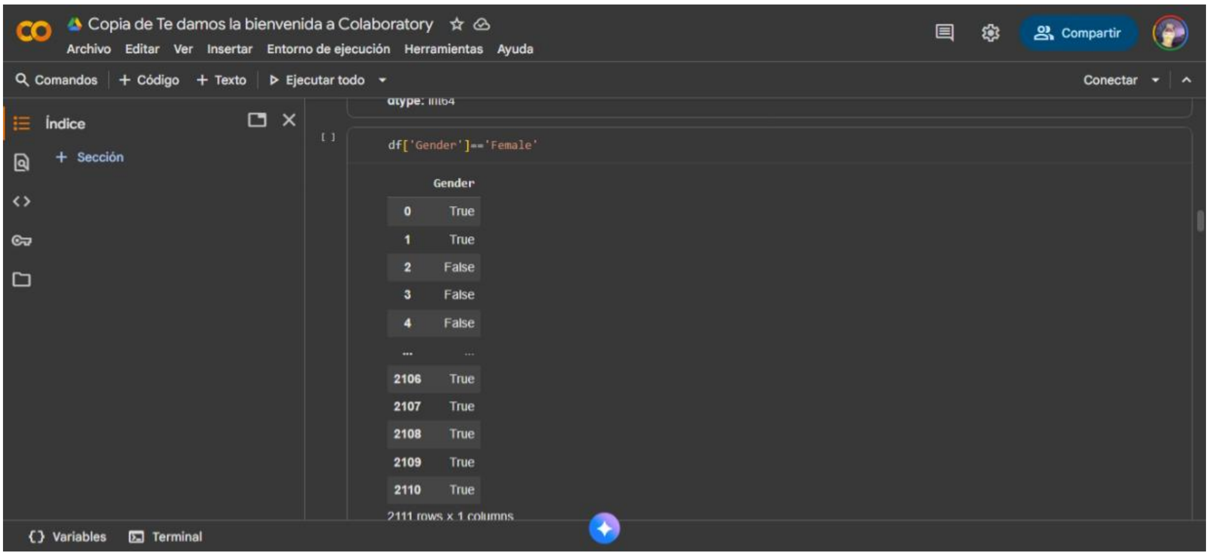
```
df.columns
```

```
Index(['Gender', 'Age', 'Height', 'Weight', 'family_history_with_overweight',
       'FAVC', 'FCVC', 'NCP', 'CAEC', 'SMOKE', 'CH2O', 'SCC', 'FAF', 'TUE',
       'CALC', 'MTRANS', 'NObeyesdad'],
      dtype='object')
```

```
df['Gender'].value_counts()
```

	count
Gender	
Male	1068
Female	1043

dtype: int64

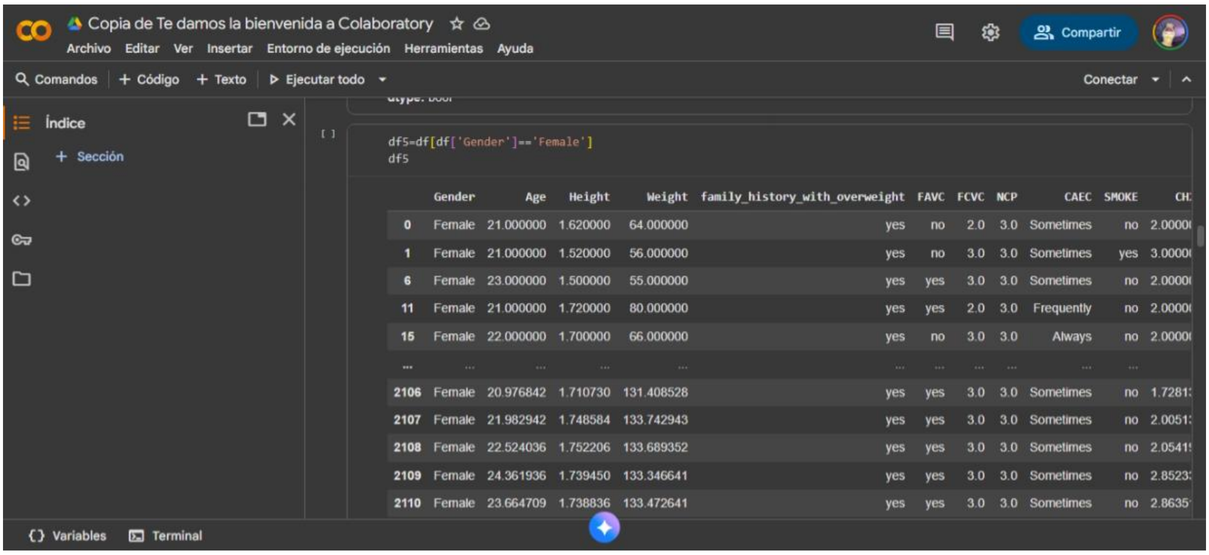

Copia de Te damos la bienvenida a Colaboratory
Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda
Compartir
Comandos + Código + Texto ▷ Ejecutar todo
Conectar
Índice
+ Sección
Variables
Terminal

```
dtype: int64
```

```
df['Gender']=='Female'
```

	Gender
0	True
1	True
2	False
3	False
4	False
...	...
2106	True
2107	True
2108	True
2109	True
2110	True

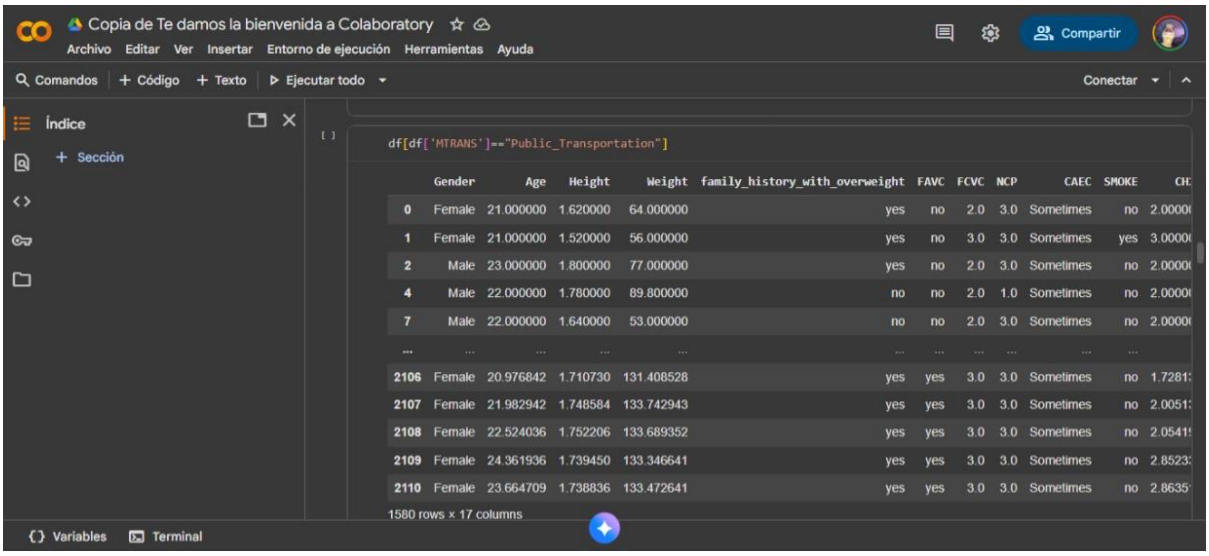
2111 rows × 1 columns


Copia de Te damos la bienvenida a Colaboratory
Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda
Compartir
Comandos + Código + Texto ▷ Ejecutar todo
Conectar
Índice
+ Sección
Variables
Terminal

```
dtype: bool
```

```
df5=df[df['Gender']=='Female']
df5
```

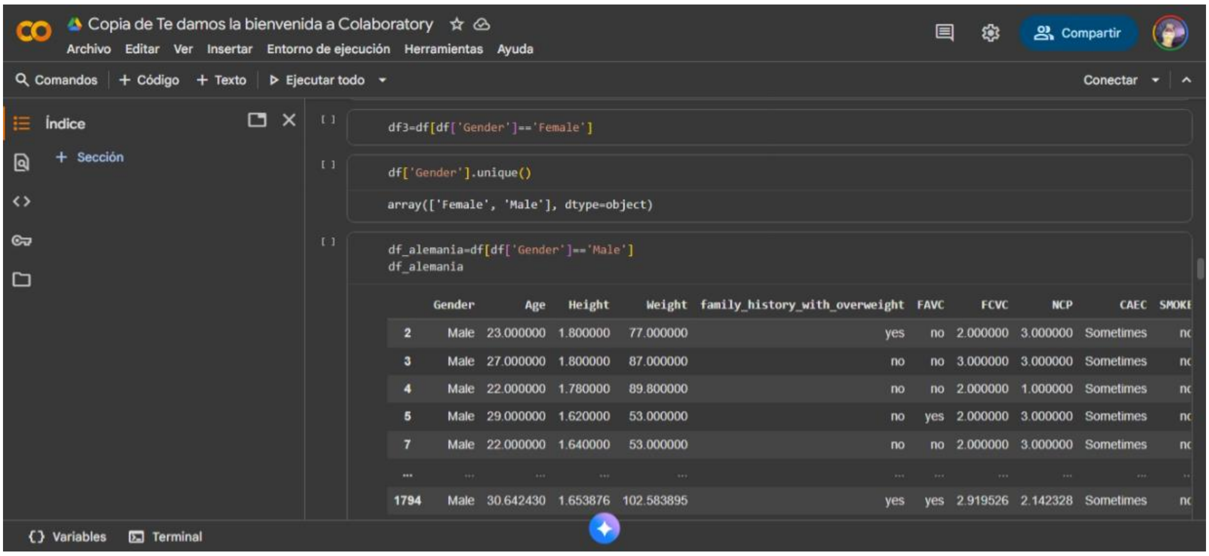
	Gender	Age	Height	Weight	family_history_with_overweight	FAVC	FCVC	NCP	CAEC	SMOKE	CH
0	Female	21.000000	1.620000	64.000000	yes	no	2.0	3.0	Sometimes	no	2.0000
1	Female	21.000000	1.520000	56.000000	yes	no	3.0	3.0	Sometimes	yes	3.0000
6	Female	23.000000	1.500000	55.000000	yes	yes	3.0	3.0	Sometimes	no	2.0000
11	Female	21.000000	1.720000	80.000000	yes	yes	2.0	3.0	Frequently	no	2.0000
15	Female	22.000000	1.700000	66.000000	yes	no	3.0	3.0	Always	no	2.0000
...	...	...	...	...	...	...	...	...	...	...	
2106	Female	20.976842	1.710730	131.408528	yes	yes	3.0	3.0	Sometimes	no	1.7281
2107	Female	21.982942	1.748584	133.742943	yes	yes	3.0	3.0	Sometimes	no	2.0051
2108	Female	22.524036	1.752206	133.689352	yes	yes	3.0	3.0	Sometimes	no	2.0541
2109	Female	24.361936	1.739450	133.346641	yes	yes	3.0	3.0	Sometimes	no	2.8523
2110	Female	23.664709	1.738836	133.472641	yes	yes	3.0	3.0	Sometimes	no	2.8635


Copia de Te damos la bienvenida a Colaboratory
Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda
Compartir
Comandos + Código + Texto ▷ Ejecutar todo
Conectar
Índice
+ Sección
Variables
Terminal

```
df[df['MTRANS']=="Public_Transportation"]
```

	Gender	Age	Height	Weight	family_history_with_overweight	FAVC	FCVC	NCP	CAEC	SMOKE	CH
0	Female	21.000000	1.620000	64.000000	yes	no	2.0	3.0	Sometimes	no	2.0000
1	Female	21.000000	1.520000	56.000000	yes	no	3.0	3.0	Sometimes	yes	3.0000
2	Male	23.000000	1.800000	77.000000	yes	no	2.0	3.0	Sometimes	no	2.0000
4	Male	22.000000	1.780000	89.800000	no	no	2.0	1.0	Sometimes	no	2.0000
7	Male	22.000000	1.640000	53.000000	no	no	2.0	3.0	Sometimes	no	2.0000
...	...	...	...	...	...	...	...	...	...	...	
2106	Female	20.976842	1.710730	131.408528	yes	yes	3.0	3.0	Sometimes	no	1.7281
2107	Female	21.982942	1.748584	133.742943	yes	yes	3.0	3.0	Sometimes	no	2.0051
2108	Female	22.524036	1.752206	133.689352	yes	yes	3.0	3.0	Sometimes	no	2.0541
2109	Female	24.361936	1.739450	133.346641	yes	yes	3.0	3.0	Sometimes	no	2.8523
2110	Female	23.664709	1.738836	133.472641	yes	yes	3.0	3.0	Sometimes	no	2.8635

1580 rows × 17 columns


Copia de Te damos la bienvenida a Colaboratory
Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda
Compartir
Comandos + Código + Texto ▷ Ejecutar todo
Conectar
Índice
+ Sección
Variables Terminal

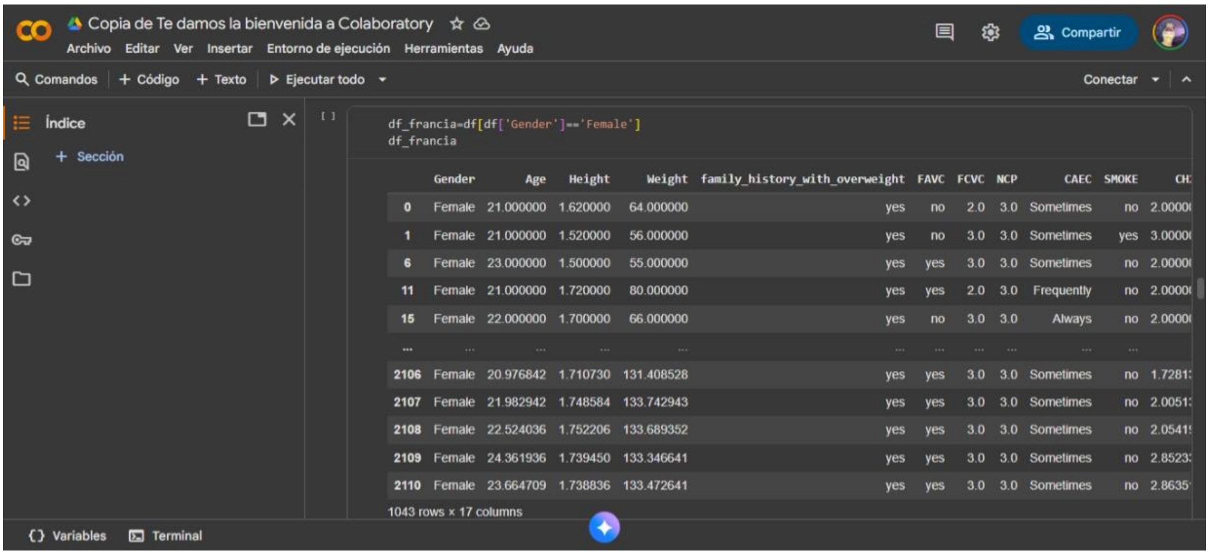
```
df3=df[df['Gender']=='Female']
```

```
df['Gender'].unique()
```

```
array(['Female', 'Male'], dtype=object)
```

```
df_alemania=df[df['Gender']=='Male']
df_alemania
```

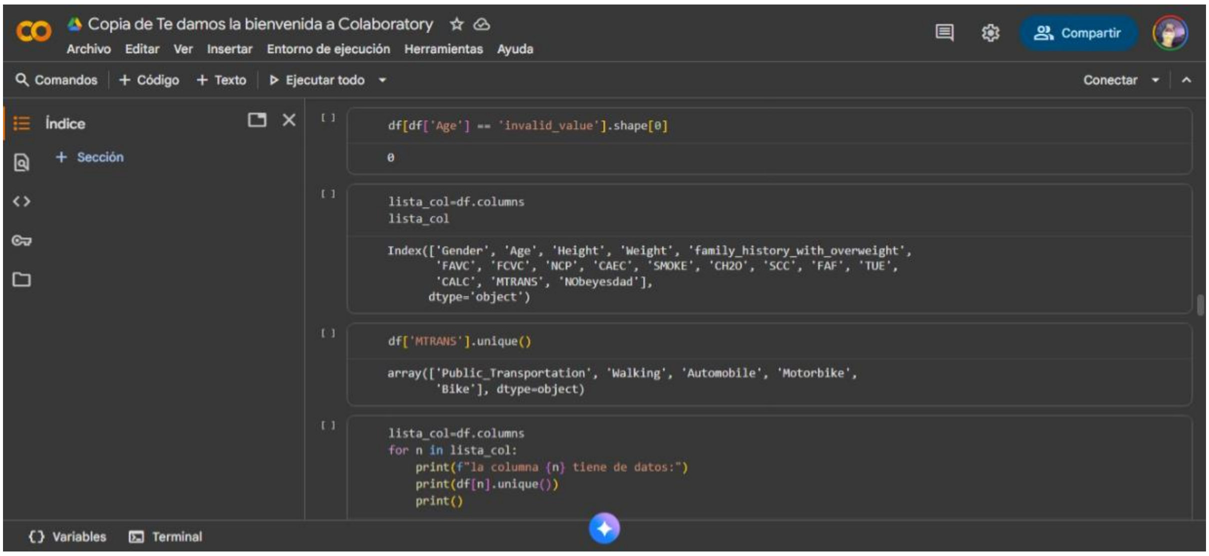
	Gender	Age	Height	Weight	family_history_with_overweight	FAVC	FCVC	NCP	CAEC	SMOKE
2	Male	23.000000	1.800000	77.000000	yes	no	2.000000	3.000000	Sometimes	no
3	Male	27.000000	1.800000	87.000000	no	no	3.000000	3.000000	Sometimes	no
4	Male	22.000000	1.780000	89.800000	no	no	2.000000	1.000000	Sometimes	no
5	Male	29.000000	1.620000	53.000000	no	yes	2.000000	3.000000	Sometimes	no
7	Male	22.000000	1.640000	53.000000	no	no	2.000000	3.000000	Sometimes	no
...	...	...	...	...	...	...	...	...	...	...
1794	Male	30.642430	1.653876	102.583895	yes	yes	2.919526	2.142328	Sometimes	no


Copia de Te damos la bienvenida a Colaboratory
Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda
Compartir
Comandos + Código + Texto ▷ Ejecutar todo
Conectar
Índice
+ Sección
Variables Terminal

```
df_francia=df[df['Gender']=='Female']
df_francia
```

	Gender	Age	Height	Weight	family_history_with_overweight	FAVC	FCVC	NCP	CAEC	SMOKE	CH
0	Female	21.000000	1.620000	64.000000	yes	no	2.0	3.0	Sometimes	no	2.0000
1	Female	21.000000	1.520000	56.000000	yes	no	3.0	3.0	Sometimes	yes	3.0000
6	Female	23.000000	1.500000	55.000000	yes	yes	3.0	3.0	Sometimes	no	2.0000
11	Female	21.000000	1.720000	80.000000	yes	yes	2.0	3.0	Frequently	no	2.0000
15	Female	22.000000	1.700000	66.000000	yes	no	3.0	3.0	Always	no	2.0000
...	...	...	...	...	...	...	...	...	...	...	
2106	Female	20.976842	1.710730	131.408528	yes	yes	3.0	3.0	Sometimes	no	1.7281
2107	Female	21.982942	1.748584	133.742943	yes	yes	3.0	3.0	Sometimes	no	2.0051
2108	Female	22.524036	1.752206	133.689352	yes	yes	3.0	3.0	Sometimes	no	2.0541
2109	Female	24.361936	1.739450	133.346641	yes	yes	3.0	3.0	Sometimes	no	2.8523
2110	Female	23.664709	1.738836	133.472641	yes	yes	3.0	3.0	Sometimes	no	2.8635

1043 rows × 17 columns


Copia de Te damos la bienvenida a Colaboratory
Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda
Compartir
Comandos + Código + Texto ▷ Ejecutar todo
Conectar
Índice
+ Sección
Variables Terminal

```
df[df['Age'] == 'invalid_value'].shape[0]
```

```
0
```

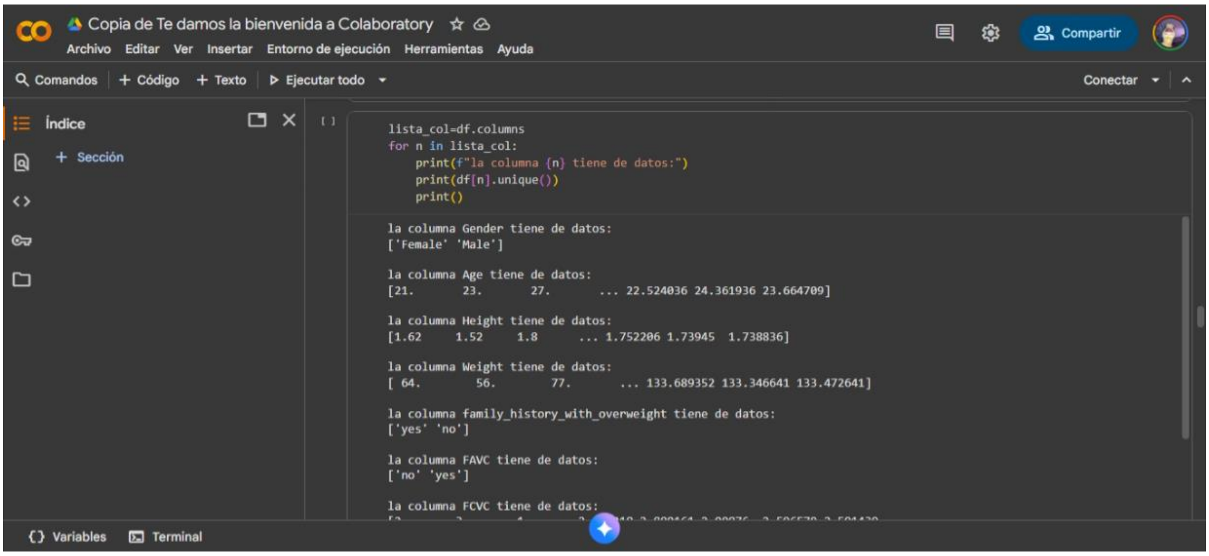
```
lista_col=df.columns
lista_col
```

```
Index(['Gender', 'Age', 'Height', 'Weight', 'family_history_with_overweight',
       'FAVC', 'FCVC', 'NCP', 'CAEC', 'SMOKE', 'CH2O', 'SCC', 'FAF', 'TUE',
       'CALC', 'MTRANS', 'NObeyesdad'],
      dtype='object')
```

```
df['MTRANS'].unique()
```

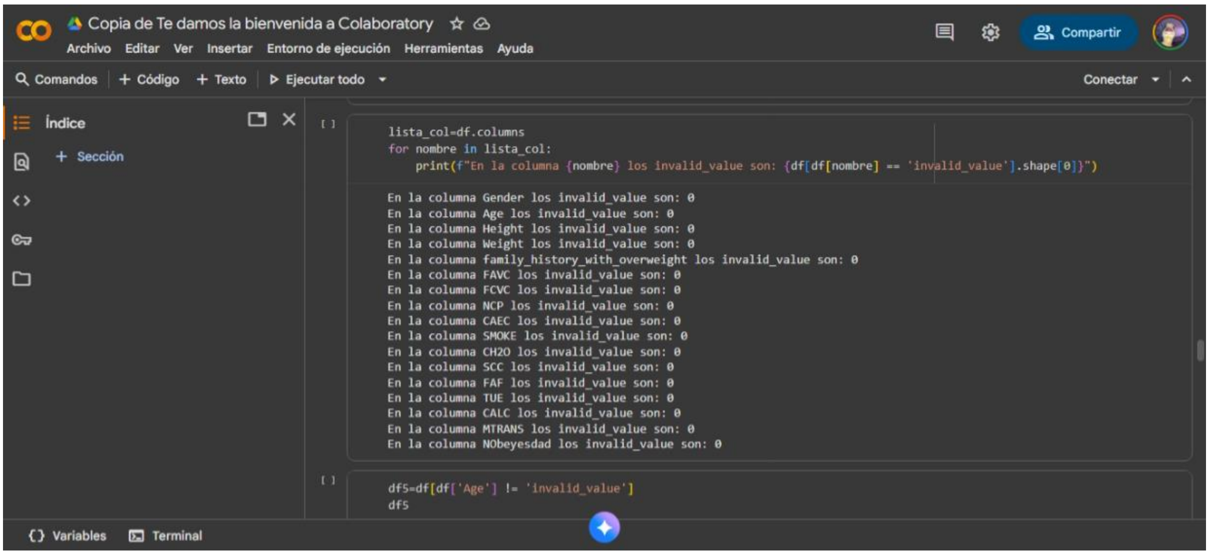
```
array(['Public_Transportation', 'Walking', 'Automobile', 'Motorbike',
       'Bike'], dtype=object)
```

```
lista_col=df.columns
for n in lista_col:
    print(f"la columna {n} tiene de datos:")
    print(df[n].unique())
    print()
```


Copia de Te damos la bienvenida a Colaboratory
Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda
Compartir
Comandos + Código + Texto ▷ Ejecutar todo
Conectar
Índice
+ Sección
Variables
Terminal

```
lista_col=df.columns
for n in lista_col:
    print(f"la columna {n} tiene de datos:")
    print(df[n].unique())
    print()
```

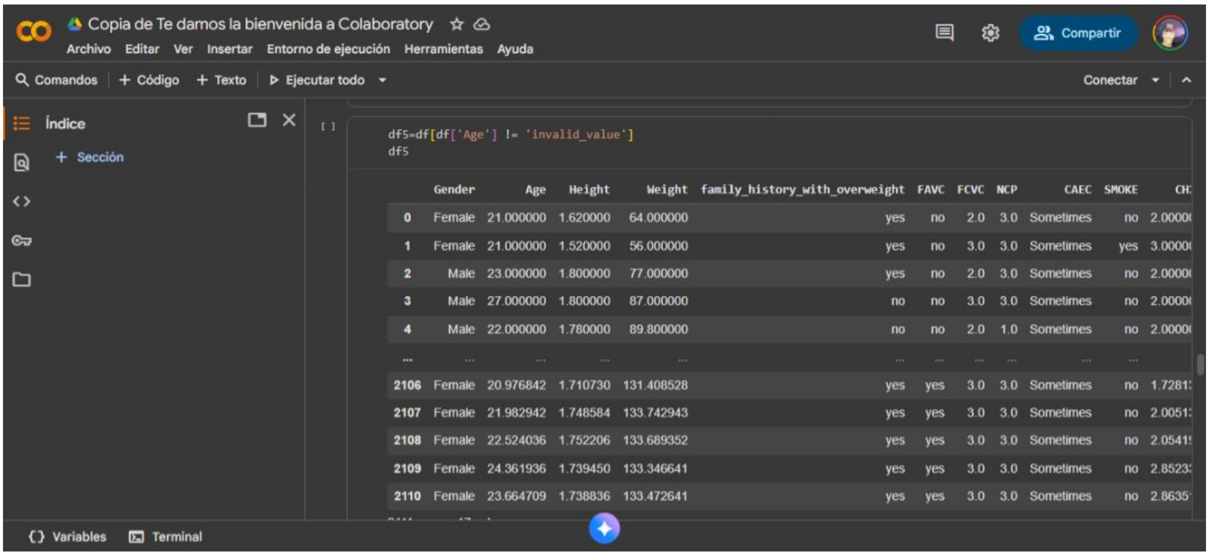
```
la columna Gender tiene de datos:
['Female' 'Male']

la columna Age tiene de datos:
[21.        23.        27.        ... 22.524036 24.361936 23.664709]

la columna Height tiene de datos:
[1.62      1.52      1.8       ... 1.752206 1.73945  1.738836]

la columna Weight tiene de datos:
[ 64.         56.         77.        ... 133.689352 133.346641 133.472641]

la columna family_history_with_overweight tiene de datos:
['yes' 'no']

la columna FAVC tiene de datos:
['no' 'yes']

la columna FCVC tiene de datos:
```


Copia de Te damos la bienvenida a Colaboratory
Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda
Compartir
Comandos + Código + Texto ▷ Ejecutar todo
Conectar
Índice
+ Sección
Variables
Terminal

```
lista_col=df.columns
for nombre in lista_col:
    print(f"En la columna {nombre} los invalid_value son: {df[df[nombre] == 'invalid_value'].shape[0]}")
```

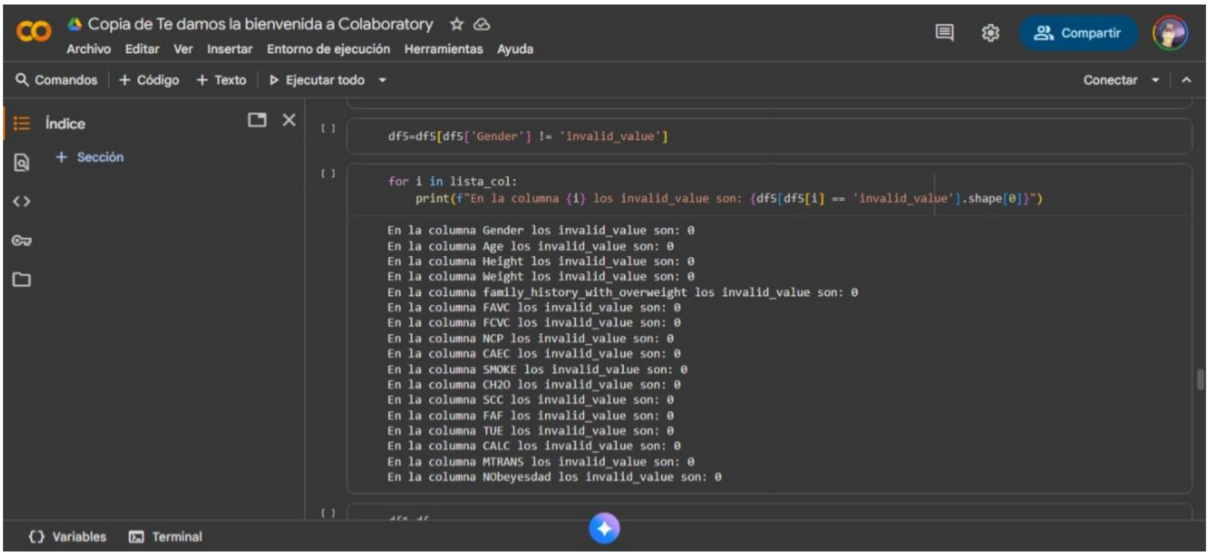
```
En la columna Gender los invalid_value son: 0
En la columna Age los invalid_value son: 0
En la columna Height los invalid_value son: 0
En la columna Weight los invalid_value son: 0
En la columna family_history_with_overweight los invalid_value son: 0
En la columna FAVC los invalid_value son: 0
En la columna FCVC los invalid_value son: 0
En la columna NCP los invalid_value son: 0
En la columna CAEC los invalid_value son: 0
En la columna SMOKE los invalid_value son: 0
En la columna CH2O los invalid_value son: 0
En la columna SCC los invalid_value son: 0
En la columna FAF los invalid_value son: 0
En la columna TUE los invalid_value son: 0
En la columna CALC los invalid_value son: 0
En la columna MTRANS los invalid_value son: 0
En la columna NObeyesdad los invalid_value son: 0
```

```
df5=df[df['Age'] != 'invalid_value']
df5
```


Copia de Te damos la bienvenida a Colaboratory
Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda
Compartir
Comandos + Código + Texto ▷ Ejecutar todo
Conectar
Índice
+ Sección
Variables
Terminal

```
df5=df[df['Age'] != 'invalid_value']
df5
```

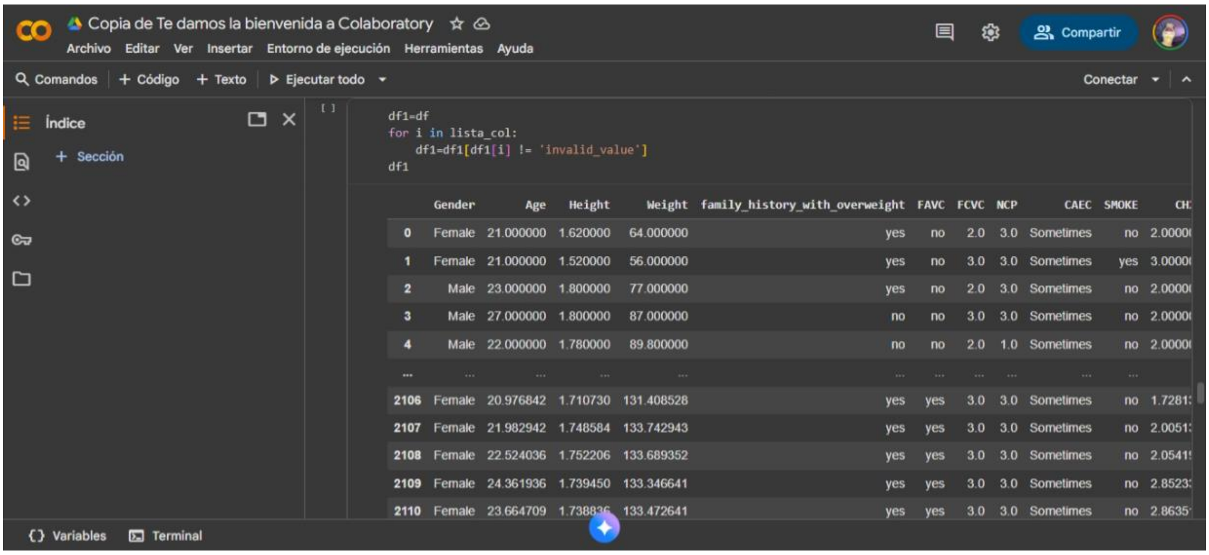
	Gender	Age	Height	Weight	family_history_with_overweight	FAVC	FCVC	NCP	CAEC	SMOKE	CH
0	Female	21.000000	1.620000	64.000000	yes	no	2.0	3.0	Sometimes	no	2.0000
1	Female	21.000000	1.520000	56.000000	yes	no	3.0	3.0	Sometimes	yes	3.0000
2	Male	23.000000	1.800000	77.000000	yes	no	2.0	3.0	Sometimes	no	2.0000
3	Male	27.000000	1.800000	87.000000	no	no	3.0	3.0	Sometimes	no	2.0000
4	Male	22.000000	1.780000	89.800000	no	no	2.0	1.0	Sometimes	no	2.0000
...	...	...	...	...	...	...	...	...	...	...	
2106	Female	20.976842	1.710730	131.408528	yes	yes	3.0	3.0	Sometimes	no	1.7281
2107	Female	21.982942	1.748584	133.742943	yes	yes	3.0	3.0	Sometimes	no	2.0051
2108	Female	22.524036	1.752206	133.689352	yes	yes	3.0	3.0	Sometimes	no	2.0541
2109	Female	24.361936	1.739450	133.346641	yes	yes	3.0	3.0	Sometimes	no	2.8523
2110	Female	23.664709	1.738836	133.472641	yes	yes	3.0	3.0	Sometimes	no	2.8635


Copia de Te damos la bienvenida a Colaboratory
Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda
Compartir
Comandos + Código + Texto Ejecutar todo
Conectar
Índice
+ Sección
Variables Terminal

```
df5=df5[df5['Gender'] != 'invalid_value']
```

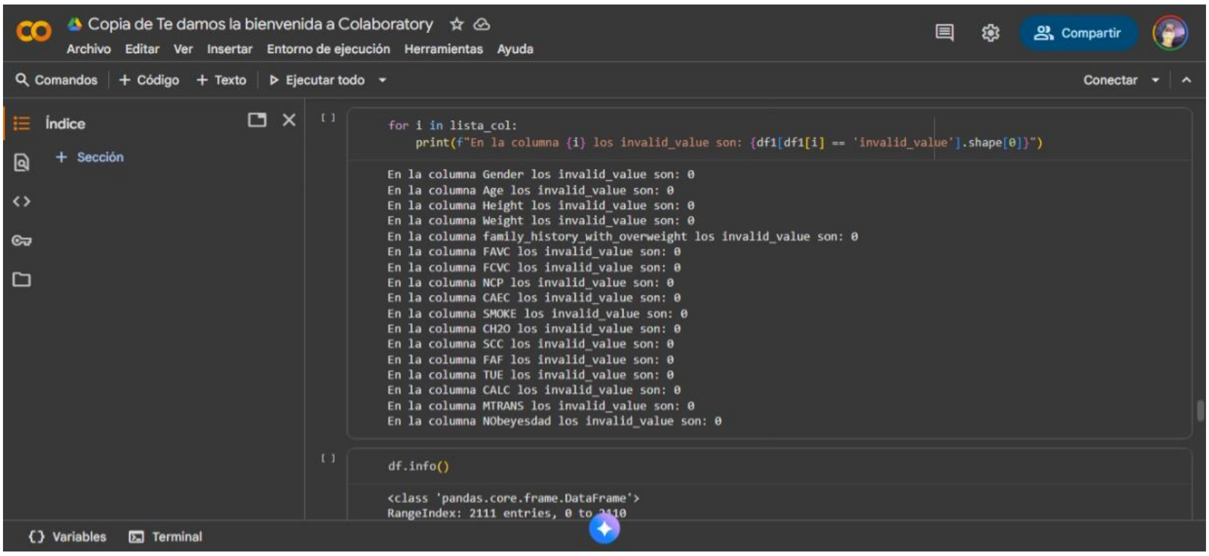
```
for i in lista_col:
    print(f"En la columna {i} los invalid_value son: {df5[df5[i] == 'invalid_value'].shape[0]}")
```

```
En la columna Gender los invalid_value son: 0
En la columna Age los invalid_value son: 0
En la columna Height los invalid_value son: 0
En la columna Weight los invalid_value son: 0
En la columna family_history_with_overweight los invalid_value son: 0
En la columna FAVC los invalid_value son: 0
En la columna FCVC los invalid_value son: 0
En la columna NCP los invalid_value son: 0
En la columna CAEC los invalid_value son: 0
En la columna SMOKE los invalid_value son: 0
En la columna CH2O los invalid_value son: 0
En la columna SCC los invalid_value son: 0
En la columna FAF los invalid_value son: 0
En la columna TUE los invalid_value son: 0
En la columna CALC los invalid_value son: 0
En la columna MTRANS los invalid_value son: 0
En la columna NObeyesdad los invalid_value son: 0
```


Copia de Te damos la bienvenida a Colaboratory
Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda
Compartir
Comandos + Código + Texto Ejecutar todo
Conectar
Índice
+ Sección
Variables Terminal

```
df1=df
for i in lista_col:
    df1=df1[df1[i] != 'invalid_value']
df1
```

	Gender	Age	Height	Weight	family_history_with_overweight	FAVC	FCVC	NCP	CAEC	SMOKE	CH
0	Female	21.000000	1.620000	64.000000	yes	no	2.0	3.0	Sometimes	no	2.0000
1	Female	21.000000	1.520000	56.000000	yes	no	3.0	3.0	Sometimes	yes	3.0000
2	Male	23.000000	1.800000	77.000000	yes	no	2.0	3.0	Sometimes	no	2.0000
3	Male	27.000000	1.800000	87.000000	no	no	3.0	3.0	Sometimes	no	2.0000
4	Male	22.000000	1.780000	89.800000	no	no	2.0	1.0	Sometimes	no	2.0000
...	...	...	...	...	...	...	...	...	...	...	
2106	Female	20.976842	1.710730	131.408528	yes	yes	3.0	3.0	Sometimes	no	1.7281
2107	Female	21.982942	1.748584	133.742943	yes	yes	3.0	3.0	Sometimes	no	2.0051
2108	Female	22.524036	1.752206	133.689352	yes	yes	3.0	3.0	Sometimes	no	2.0541
2109	Female	24.361936	1.739450	133.346641	yes	yes	3.0	3.0	Sometimes	no	2.8523
2110	Female	23.664709	1.738836	133.472641	yes	yes	3.0	3.0	Sometimes	no	2.8635

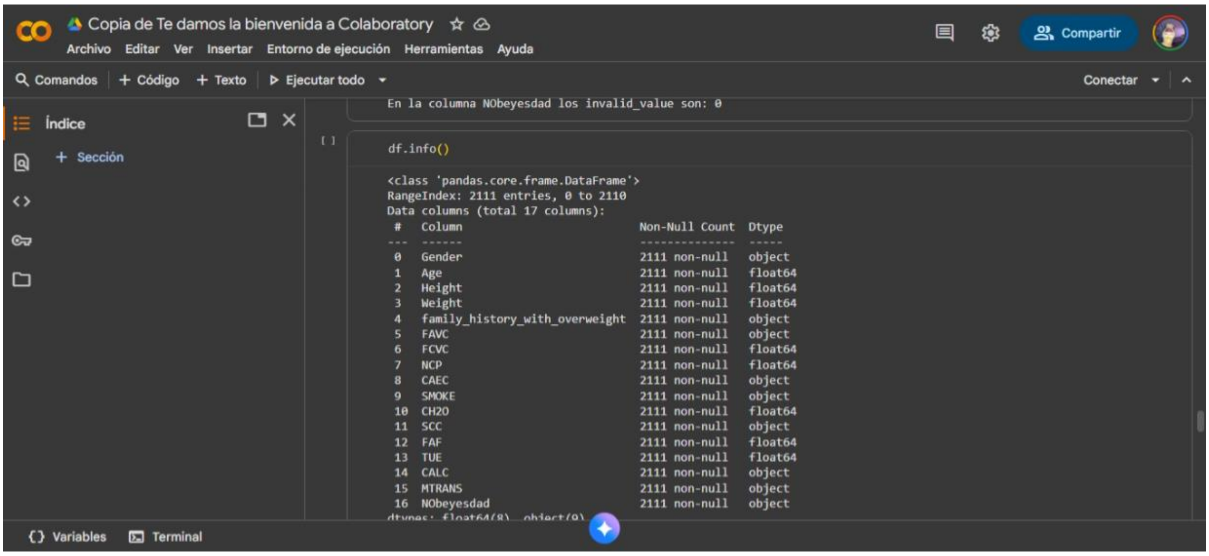

Copia de Te damos la bienvenida a Colaboratory
Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda
Compartir
Comandos + Código + Texto Ejecutar todo
Conectar
Índice
+ Sección
Variables Terminal

```
for i in lista_col:
    print(f"En la columna {i} los invalid_value son: {df1[df1[i] == 'invalid_value'].shape[0]}")
```

```
En la columna Gender los invalid_value son: 0
En la columna Age los invalid_value son: 0
En la columna Height los invalid_value son: 0
En la columna Weight los invalid_value son: 0
En la columna family_history_with_overweight los invalid_value son: 0
En la columna FAVC los invalid_value son: 0
En la columna FCVC los invalid_value son: 0
En la columna NCP los invalid_value son: 0
En la columna CAEC los invalid_value son: 0
En la columna SMOKE los invalid_value son: 0
En la columna CH2O los invalid_value son: 0
En la columna SCC los invalid_value son: 0
En la columna FAF los invalid_value son: 0
En la columna TUE los invalid_value son: 0
En la columna CALC los invalid_value son: 0
En la columna MTRANS los invalid_value son: 0
En la columna NObeyesdad los invalid_value son: 0
```

```
df.info()
```

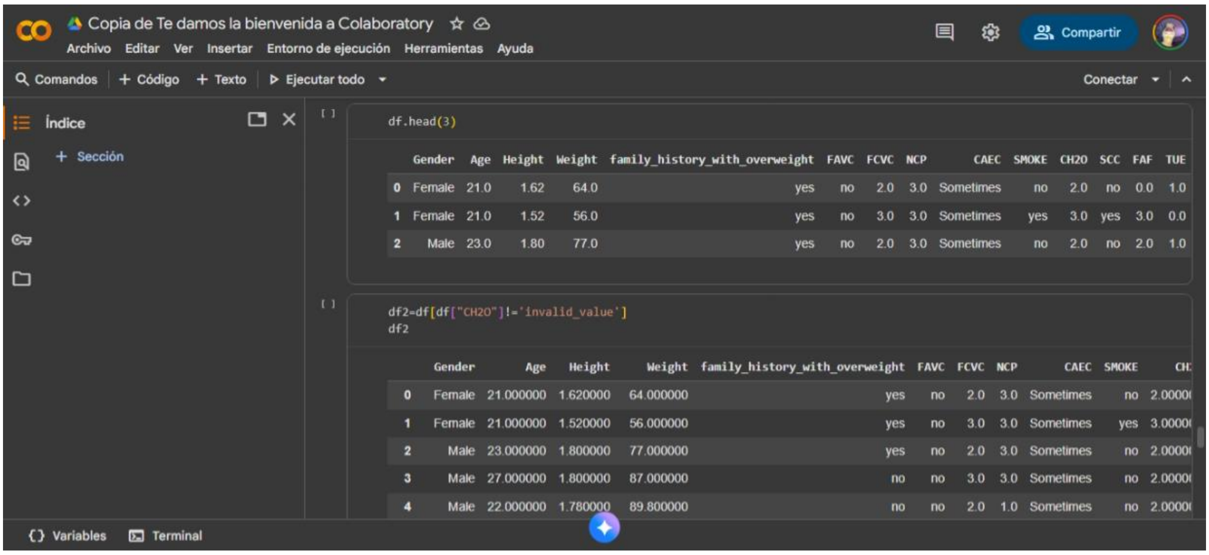
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2111 entries, 0 to 2110
```


Copia de Te damos la bienvenida a Colaboratory
Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda
Compartir
Comandos + Código + Texto Ejecutar todo
Conectar
Índice
+ Sección
Variables Terminal

```
En la columna NObeyesdad los invalid_value son: 0
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2111 entries, 0 to 2110
Data columns (total 17 columns):
 #   Column                          Non-Null Count  Dtype  
---  ------                          --------------  -----  
 0   Gender                          2111 non-null   object 
 1   Age                             2111 non-null   float64
 2   Height                          2111 non-null   float64
 3   Weight                          2111 non-null   float64
 4   family_history_with_overweight  2111 non-null   object 
 5   FAVC                            2111 non-null   object 
 6   FCVC                            2111 non-null   float64
 7   NCP                             2111 non-null   float64
 8   CAEC                            2111 non-null   object 
 9   SMOKE                           2111 non-null   object 
 10  CH2O                            2111 non-null   float64
 11  SCC                             2111 non-null   object 
 12  FAF                             2111 non-null   float64
 13  TUE                             2111 non-null   float64
 14  CALC                            2111 non-null   object 
 15  MTRANS                          2111 non-null   object 
 16  NObeyesdad                      2111 non-null   object 
dtypes: float64(8), object(9)
```

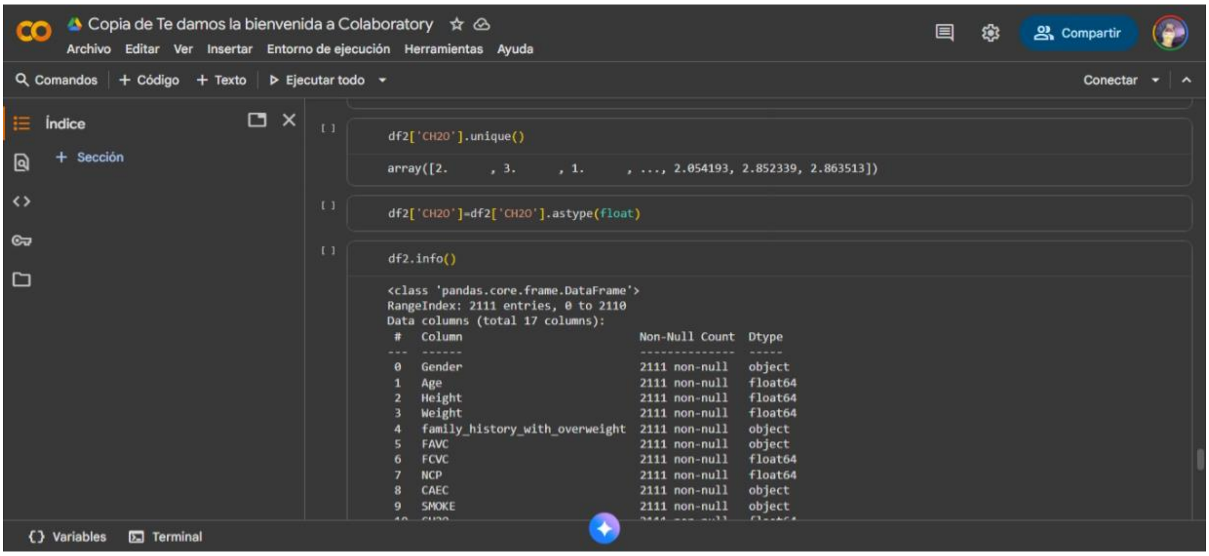

Copia de Te damos la bienvenida a Colaboratory
Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda
Compartir
Comandos + Código + Texto Ejecutar todo
Conectar
Índice
+ Sección
Variables Terminal

```
df.head(3)
```

	Gender	Age	Height	Weight	family_history_with_overweight	FAVC	FCVC	NCP	CAEC	SMOKE	CH2O	SCC	FAF	TUE
0	Female	21.0	1.62	64.0	yes	no	2.0	3.0	Sometimes	no	2.0	no	0.0	1.0
1	Female	21.0	1.52	56.0	yes	no	3.0	3.0	Sometimes	yes	3.0	yes	3.0	0.0
2	Male	23.0	1.80	77.0	yes	no	2.0	3.0	Sometimes	no	2.0	no	2.0	1.0

```
df2=df[df["CH2O"]!='invalid_value']
df2
```

	Gender	Age	Height	Weight	family_history_with_overweight	FAVC	FCVC	NCP	CAEC	SMOKE	CH
0	Female	21.000000	1.620000	64.000000	yes	no	2.0	3.0	Sometimes	no	2.0000
1	Female	21.000000	1.520000	56.000000	yes	no	3.0	3.0	Sometimes	yes	3.0000
2	Male	23.000000	1.800000	77.000000	yes	no	2.0	3.0	Sometimes	no	2.0000
3	Male	27.000000	1.800000	87.000000	no	no	3.0	3.0	Sometimes	no	2.0000
4	Male	22.000000	1.780000	89.800000	no	no	2.0	1.0	Sometimes	no	2.0000


Copia de Te damos la bienvenida a Colaboratory
Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda
Compartir
Comandos + Código + Texto Ejecutar todo
Conectar
Índice
+ Sección
Variables Terminal

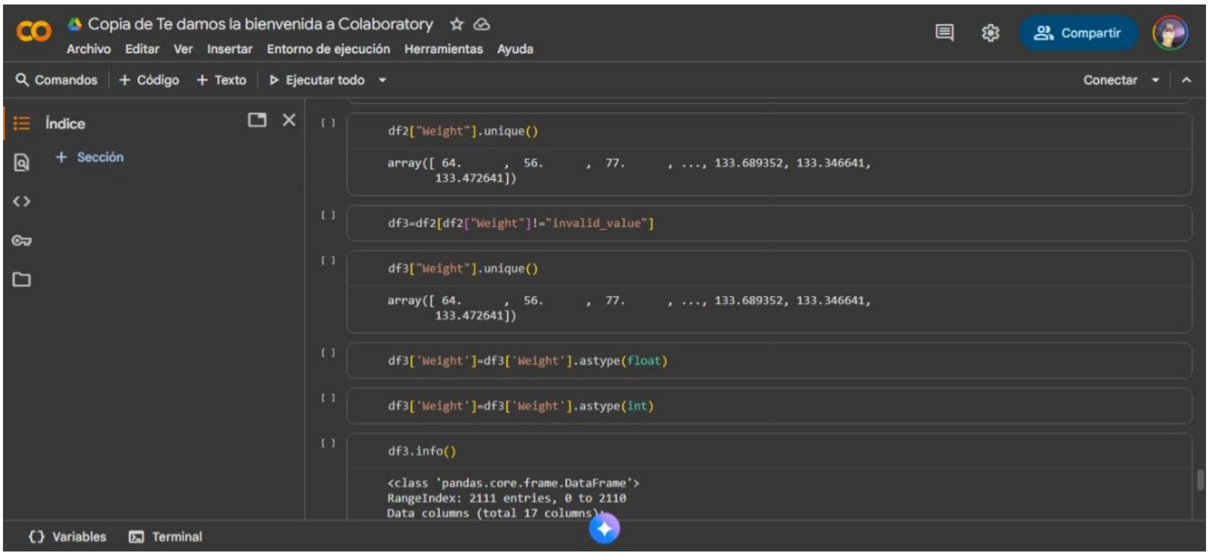
```
df2['CH2O'].unique()
```

```
array([2.      , 3.      , 1.      , ..., 2.054193, 2.852339, 2.863513])
```

```
df2['CH2O']=df2['CH2O'].astype(float)
```

```
df2.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2111 entries, 0 to 2110
Data columns (total 17 columns):
 #   Column                          Non-Null Count  Dtype  
---  ------                          --------------  -----  
 0   Gender                          2111 non-null   object 
 1   Age                             2111 non-null   float64
 2   Height                          2111 non-null   float64
 3   Weight                          2111 non-null   float64
 4   family_history_with_overweight  2111 non-null   object 
 5   FAVC                            2111 non-null   object 
 6   FCVC                            2111 non-null   float64
 7   NCP                             2111 non-null   float64
 8   CAEC                            2111 non-null   object 
 9   SMOKE                           2111 non-null   object 
```



Copia de Te damos la bienvenida a Colaboratory
Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda
Compartir
Comandos + Código + Texto ▷ Ejecutar todo
Conectar
Índice
+ Sección
Variables Terminal

```
df2["Weight"].unique()

array([ 64.      ,  56.      ,  77.      , ..., 133.689352, 133.346641,
       133.472641])
```

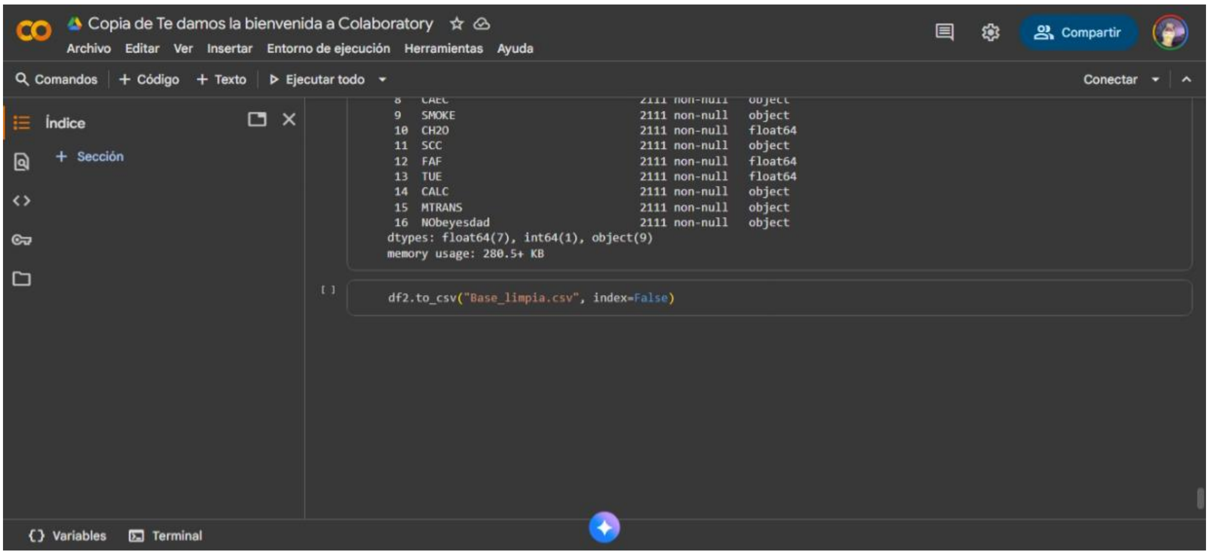
```
df3=df2[df2["Weight"]!="invalid_value"]
```

```
df3["Weight"].unique()

array([ 64.      ,  56.      ,  77.      , ..., 133.689352, 133.346641,
       133.472641])
```

```
df3['Weight']=df3['Weight'].astype(float)
```

```
df3['Weight']=df3['Weight'].astype(int)
```

```
df3.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2111 entries, 0 to 2110
Data columns (total 17 columns):
```



Copia de Te damos la bienvenida a Colaboratory
Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda
Compartir
Comandos + Código + Texto ▷ Ejecutar todo
Conectar
Índice
+ Sección
Variables Terminal

```
 8   CAEC                            2111 non-null   object
 9   SMOKE                           2111 non-null   object
 10  CH2O                            2111 non-null   float64
 11  SCC                             2111 non-null   object
 12  FAF                             2111 non-null   float64
 13  TUE                             2111 non-null   float64
 14  CALC                            2111 non-null   object
 15  MTRANS                          2111 non-null   object
 16  NObeyesdad                      2111 non-null   object
dtypes: float64(7), int64(1), object(9)
memory usage: 280.5+ KB
```

```
df2.to_csv("Base_limpia.csv", index=False)
```

Conclusión

La base de datos no estuvo complicada, ya que en la mayoría esta en una forma donde yo puede trabajar sin ningún problema.