```
In [1]: import pandas as pd
        import numpy as np
        %ls
```

```
A3.ipynb                                        'Class3- pt2 (1).ipynb'
'Access to electricity (% of population).csv'   'Class3- pt2.ipynb'
'Access to electricity (% of population).xlsx'   data/
Act2-Binder.csv                                  datasets/
Act2-Binder.xls                                  newdata/
antropometria-dataset-2.csv                      README.md
API_AG.LND.FRST.ZS_DS2_en_csv_v2_3469441.csv     requirements.txt
'Canada (2).ipynb'                              'Spotify (2).ipynb'
Canada.ipynb                                    'Untitled1 (1).ipynb'
Class3.ipynb                                     Untitled.ipynb
```

```
In [2]: df = pd.read_csv('antropometria-dataset-2.csv')
        df.head(5)
```

Out[2]:

	folio	intp	entidad	desc_ent	sexo	edad	meses	peso	ropa	talla	...	hpresion
0	210295	2	21	PUEBLA	2	38	8	73.70	2	146.4	...	16:30
1	101655	3	10	DURANGO	2	11	11	35.65	2	145.1	...	15:25
2	10287	6	1	AGUASCALIENTES	2	18	8	54.80	1	162.0	...	9:40
3	91526	4	9	DISTRITO FEDERAL	1	10	8	33.40	2	146.5	...	7:40
4	210939	3	21	PUEBLA	2	19	3	97.95	2	161.0	...	6:00

5 rows × 26 columns

```
In [3]: df.columns
```

```
Out[3]: Index(['folio', 'intp', 'entidad', 'desc_ent', 'sexo', 'edad', 'meses', 'peso',
               'ropa', 'talla', 'emb', 'temb', 'cintura', 'cadera', 'sistol',
               'diastol', 'hpresion', 'tbrazo', 'htension', 'PrimaryLast', 'code_upm',
               'est_dis', 'est_urb', 'est_marg', 'pondef', 'est_var'],
              dtype='object')
```

```
In [20]: pd.pivot_table(df,
                        index = ['desc_ent'],
                        values = 'peso',
                        columns = 'sexo',
                        margins = True
                        )
```

Out[20]:

sexo	1	2	All
desc_ent			
AGUASCALIENTES	69.985381	63.254292	66.357786
BAJA CALIFORNIA	71.391818	65.893667	68.467270
BAJA CALIFORNIA SUR	75.561015	68.168959	71.809198

sexo	1	2	All
desc_ent			
CAMPECHE	68.622637	62.920493	65.418475
CHIAPAS	58.365222	57.096111	57.613775
CHIHUAHUA	68.892642	64.815295	66.656005
COAHUILA DE ZARAGOZA	68.873094	67.467348	68.154142
COLIMA	70.998878	66.978393	68.743648
DISTRITO FEDERAL	69.460661	61.654128	65.119273
DURANGO	68.846451	64.400031	66.354922
GUANAJUATO	65.008000	63.100995	63.936390
GUERRERO	62.611437	58.899850	60.466315
HIDALGO	61.747958	58.972554	59.995917
JALISCO	67.076482	66.418111	66.697270
MEXICO	63.140314	58.887882	60.680850
MICHOACAN DE OCAMPO	66.374612	63.784311	64.900000
MORELOS	65.741943	62.195391	63.541295
NAYARIT	68.822143	64.890564	66.541827
NUEVO LEON	70.168106	64.823014	67.239853
OAXACA	60.810332	57.450557	58.800067
PUEBLA	61.866875	58.820175	60.102996
QUERETARO	66.751233	60.860793	63.547342
QUINTANA ROO	65.149528	63.538000	64.238904
SAN LUIS POTOSI	62.614907	58.979163	60.576912
SINALOA	71.122911	65.565234	68.051272
SONORA	70.880406	67.024857	68.849473
TABASCO	69.255801	64.190187	66.554701
TAMAULIPAS	72.060601	64.866851	68.210911
TLAXCALA	64.784437	60.201164	62.085398
VERACRUZ DE IGNACIO DE LA LLAVE	65.324094	60.565646	62.546117
YUCATAN	65.717541	60.537515	62.706315
ZACATECAS	66.096483	62.757788	64.238524
All	67.131228	62.700503	64.638707

```
In [61]:   pd.pivot_table(df,
                   index = ['desc_ent'],
                   values = 'peso',
                   columns = 'sexo',
```

```
                aggfunc = ['median'],
                margins = True
            )
```

Out[61]:

	median		
sexo	**1**	**2**	**All**
desc_ent			
AGUASCALIENTES	69.9500	63.7000	66.6000
BAJA CALIFORNIA	70.6500	64.2000	67.9500
BAJA CALIFORNIA SUR	76.6500	68.5500	71.2500
CAMPECHE	67.3500	62.3750	64.3500
CHIAPAS	59.2000	56.0250	57.2000
CHIHUAHUA	69.3000	64.6250	66.3000
COAHUILA DE ZARAGOZA	68.3000	67.3000	67.7000
COLIMA	71.7000	66.0000	68.9000
DISTRITO FEDERAL	69.0000	61.4500	64.0250
DURANGO	70.3500	62.5000	65.7000
GUANAJUATO	65.0625	60.3250	62.0625
GUERRERO	60.5000	57.2000	57.9500
HIDALGO	63.3000	57.9000	59.7000
JALISCO	65.5875	63.3500	64.2000
MEXICO	62.4500	56.9500	59.5500
MICHOACAN DE OCAMPO	67.6250	63.4000	65.8000
MORELOS	65.2000	61.4500	62.5750
NAYARIT	69.0500	63.7000	66.6250
NUEVO LEON	69.4000	65.1000	66.7000
OAXACA	59.5500	56.8000	57.6000
PUEBLA	61.8750	59.5500	60.3500
QUERETARO	65.2750	59.4500	62.4000
QUINTANA ROO	65.9250	60.8500	63.2750
SAN LUIS POTOSI	62.2000	57.4500	60.1000
SINALOA	71.1000	64.3000	67.1500
SONORA	72.3000	66.6500	68.5500
TABASCO	69.7500	62.6000	65.6500
TAMAULIPAS	72.7500	63.3500	67.5000
TLAXCALA	64.2750	59.8750	62.0250
VERACRUZ DE IGNACIO DE LA LLAVE	64.3000	59.4725	61.1000

	median		
sexo	**1**	**2**	**All**
desc_ent			
YUCATAN	66.0000	59.5750	62.1500
ZACATECAS	65.3000	61.6250	63.7500
All	66.7500	61.4500	63.7000

In [ ]:

In [51]:
```python
df.groupby(['sexo', 'peso']).size()
```

Out[51]:
```
sexo   peso
1      20.550     1
       22.750     1
       22.825     1
       23.000     1
       23.075     1
                 ..
2      151.550    1
       154.100    1
       161.450    1
       165.000    1
       168.600    1
Length: 4207, dtype: int64
```

In [50]:
```python
df['peso'].describe()
```

Out[50]:
```
count    18640.000000
mean        64.638707
std         17.660325
min         18.000000
25%         52.700000
50%         63.700000
75%         75.400000
max        170.600000
Name: peso, dtype: float64
```

In [ ]: