RESUMEN

filter1 = df['Country'].isin(['Colombia','Brazi'])

df2 = df[filter1]

filter1 = df['bath']>3

df2 = df[filter1]

df2

df.query('bath > 3 & price>100')

df.query('Salary\_in\_1000 >= 100 & Age < 60 & FT\_Team.str.startswith("S").values')



#todas las finas, y las primeras 6 columnas

df.iloc[:,6]

idx\_filtro\_country = df\_pob['Country'].isin(['Aruba','Colombia','Austria','Andorra'])

idx\_filtro\_pop = df\_pob['pop'].le(1000000)

idx\_filtro\_year = df\_pob['year'].isin(['2015','2016'])

idx\_filtro\_total = idx\_filtro\_country\*idx\_filtro\_pop\*idx\_filtro\_year

idx\_filtro\_total

#cargue y no tome en cuenta las primeras 4 lines

df\_pob = pd.read\_csv('poblacion.csv', skiprows=4)

df\_pob

#Cambie de notacion cientifica a float

pd.options.display.float\_format='{:,.lf}format

#Cambie a String

df['year']=pd.Categorical(df['year'].apply(str))

**#casilla pob a numerico**

df\_pob['pob'] = pd.to\_numeric(df\_pob['pob'])

idx\_filtro = df['Country'].isin(['Aruba','Colombia'])

idx\_filtro

<nos dara true, false, segun esta>

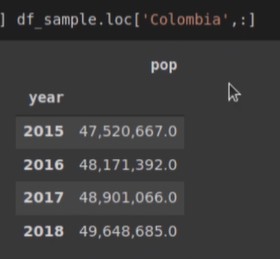
**#Para tener esos paises y sort**

df = df[idx\_filtro]

df = df.set\_index(['Contry','year']).sort\_index()



df.loc['Colombia',:]



df.loc['Colombia',:].loc['2016',:]



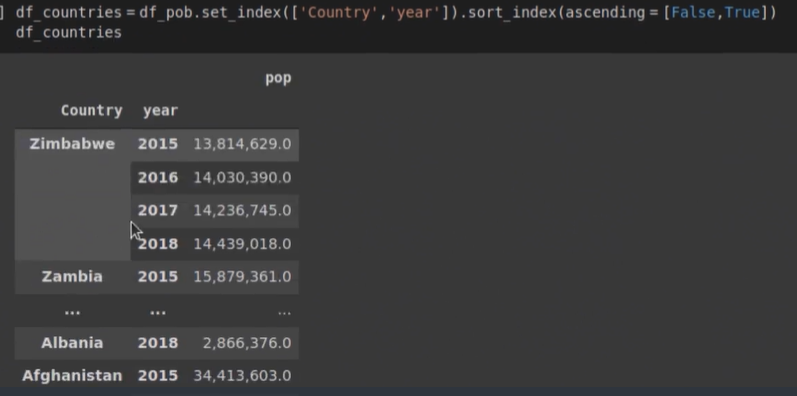
df.xs(['Aruba'])

df.xs(['Aruba','2018'])

df.xs('2018',level='year')



df.set\_index(['Country','year']).sort\_index(ascending=[False,True])



ids = pd.IndexSlice

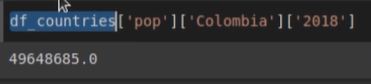
df.loc[ids['Aruba':'Austria','2015':'2017'],:].sort\_index()

df.set\_index([‘Country’,’year’]).sort\_index(ascending-[False,True])

Graphical user interface

Description automatically generated





df.sum(level='year')

Equivalente a

select year, sum(pop) from table group by year



df\_pob['Colombia','2018']

Filtro Si contiene 'or'

df['names'].str.contains('or')

names

0 True

1 False

2 False

3 False

4 True

5 False

= = = = =

#Registros que condicion es 'female' y 'Survice'

# Sex = femane / male Survive 0,1

df2 = df.loc[df.Sex=='female']['Survived']

Nos daria

0

1

0

1

1

0

#Promedio de mujeres que sobrevivieron

# suma de '1' nos da las que sobrevivieron, y len es el total de registros

df2 = sum(df2)/len(df2)

#En nuevo campo poner 'No\_survived' o "Survived' segun valor para categorizar

df['name']= df['Survived'].map({0:'No\_survived', 1:'survived'})