In [1]:
```
import json
import numpy as np
import pandas as pd
pd.set_option('display.max_columns', None)
```

In [2]:
```
import pandas as pd
df_country= pd.read_csv("Dataset/country_names_area.csv", sep=';')
df_country = df_country.drop(["country_code"], axis=1)

df_asfr= pd.read_csv("Dataset/age_specific_fertility_rates.csv", sep=';')
df_asfr = df_asfr.drop(["country_code"], axis=1)

df_asfr= pd.read_csv("Dataset/age_specific_fertility_rates.csv", sep=';')
df_asfr = df_asfr.drop(["country_code"], axis=1)

df_mypas= pd.read_csv("Dataset/midyear_population_age_sex.csv", sep=';')
df_mypas = df_mypas.drop(["country_code"], axis=1)

df_mle= pd.read_csv("Dataset/mortality_life_expectancy.csv", sep=';')
df_mle = df_mle.drop(["country_code"], axis=1)

df_bdgr = pd.read_csv("Dataset/birth_death_growth_rates.csv", sep=';')
df_bdgr = df_bdgr.drop(["country_code"], axis=1)

df_myp = pd.read_csv("Dataset/midyear_population.csv", sep=';')
df_myp = df_myp.drop(["country_code"], axis=1)
```

In [3]:
```
df_myp5= pd.read_csv("Dataset/midyear_population_5yr_age_sex.csv", sep=';')
df_myp5 = df_myp5.drop(["country_code"], axis=1)

list_flag=[]
for j in range(len(df_myp5)):
      list1=[df_myp5.iloc[j].total_flag,df_myp5.iloc[j].starting_age,df_myp5.i
loc[j].age_group_indicator,df_myp5.iloc[j].ending_age,df_myp5.iloc[j].midyear_
population,df_myp5.iloc[j].midyear_population_male,df_myp5.iloc[j].midyear_pop
ulation_female]
      list_flag.append(list1)
df_myp5['flag']=list_flag
df_myp5.drop(['total_flag','starting_age','age_group_indicator','ending_age',
'midyear_population','midyear_population_male','midyear_population_female'],ax
is=1,inplace=True)
df_myp5 = df_myp5.groupby(['country_name','year'])['flag'].apply(list).reset_i
ndex(name='flag')
```

In [ ]:
```
#On fait le choix de négliger cette table qui est une permutation de la table
 midyear_population_age_sex et d'une autre table que nous n'avons pas
#df_mypac= pd.read_csv("midyear_population_age_country_code.csv", sep=';')
#df_mypac = df_mypac.drop(['country_code','permutation_role','age','populatio
n'], axis=1)
#df_mypac
```

In [4]:
```
df_all=df_myp.merge(df_myp5, on=['country_name','year'],how="outer")
df_all=df_all.merge(df_mypas, on=['country_name','year'],how="outer")
df_all=df_all.merge(df_bdgr, on=['country_name','year'],how="outer")
df_all=df_all.merge(df_asfr,on=['country_name','year'],how="outer")
df_all=df_all.merge(df_country,on=['country_name'],how="outer")
df_all=df_all.merge(df_mle,on=['country_name','year'],how="outer")
df_all.to_csv('finalmerge.csv',index=False)
```

In [6]:
```
df_all=df_all.fillna("null",axis=0)
```

In [7]:

```
list_fm= []
for j in range(len(df_all)):
    if(df_all.iloc[j].sex=='Female'):
        list1=[df_all.iloc[j].sex,df_all.iloc[j].infant_mortality_female,df_al
l.iloc[j].life_expectancy_female,df_all.iloc[j].mortality_rate_under5_female,d
f_all.iloc[j].mortality_rate_1to4_female,df_all.iloc[j].max_age,df_all.iloc[j]
.population_age_0,df_all.iloc[j].population_age_1,df_all.iloc[j].population_ag
e_2,df_all.iloc[j].population_age_3,df_all.iloc[j].population_age_4,df_all.ilo
c[j].population_age_5,df_all.iloc[j].population_age_6,df_all.iloc[j].populatio
n_age_7,df_all.iloc[j].population_age_8,df_all.iloc[j].population_age_9,df_all
.iloc[j].population_age_10,df_all.iloc[j].population_age_11,df_all.iloc[j].pop
ulation_age_12,df_all.iloc[j].population_age_13,df_all.iloc[j].population_age_
14,df_all.iloc[j].population_age_15,df_all.iloc[j].population_age_16,df_all.il
oc[j].population_age_17,df_all.iloc[j].population_age_18,df_all.iloc[j].popula
tion_age_19,df_all.iloc[j].population_age_20,df_all.iloc[j].population_age_21,
df_all.iloc[j].population_age_22,df_all.iloc[j].population_age_23,df_all.iloc[
j].population_age_24,df_all.iloc[j].population_age_25,df_all.iloc[j].populatio
n_age_26,df_all.iloc[j].population_age_27,df_all.iloc[j].population_age_28,df_
all.iloc[j].population_age_29,df_all.iloc[j].population_age_30,df_all.iloc[j].
population_age_31,df_all.iloc[j].population_age_32,df_all.iloc[j].population_a
ge_33,df_all.iloc[j].population_age_34,df_all.iloc[j].population_age_35,df_all
.iloc[j].population_age_36,df_all.iloc[j].population_age_37,df_all.iloc[j].pop
ulation_age_38,df_all.iloc[j].population_age_39,df_all.iloc[j].population_age_
40,df_all.iloc[j].population_age_41,df_all.iloc[j].population_age_42,df_all.il
oc[j].population_age_43,df_all.iloc[j].population_age_44,df_all.iloc[j].popula
tion_age_45,df_all.iloc[j].population_age_46,df_all.iloc[j].population_age_47,
df_all.iloc[j].population_age_48,df_all.iloc[j].population_age_49,df_all.iloc[
j].population_age_50,df_all.iloc[j].population_age_51,df_all.iloc[j].populatio
n_age_52,df_all.iloc[j].population_age_53,df_all.iloc[j].population_age_54,df_
all.iloc[j].population_age_55,df_all.iloc[j].population_age_56,df_all.iloc[j].
population_age_57,df_all.iloc[j].population_age_58,df_all.iloc[j].population_a
ge_59,df_all.iloc[j].population_age_60,df_all.iloc[j].population_age_61,df_all
.iloc[j].population_age_62,df_all.iloc[j].population_age_63,df_all.iloc[j].pop
ulation_age_64,df_all.iloc[j].population_age_65,df_all.iloc[j].population_age_
66,df_all.iloc[j].population_age_67,df_all.iloc[j].population_age_68,df_all.il
oc[j].population_age_69,df_all.iloc[j].population_age_70,df_all.iloc[j].popula
tion_age_71,df_all.iloc[j].population_age_72,df_all.iloc[j].population_age_73,
df_all.iloc[j].population_age_74,df_all.iloc[j].population_age_75,df_all.iloc[
j].population_age_76,df_all.iloc[j].population_age_77,df_all.iloc[j].populatio
n_age_78,df_all.iloc[j].population_age_79,df_all.iloc[j].population_age_80,df_
all.iloc[j].population_age_81,df_all.iloc[j].population_age_82,df_all.iloc[j].
population_age_83,df_all.iloc[j].population_age_84,df_all.iloc[j].population_a
ge_85,df_all.iloc[j].population_age_86,df_all.iloc[j].population_age_87,df_all
.iloc[j].population_age_88,df_all.iloc[j].population_age_89,df_all.iloc[j].pop
ulation_age_90,df_all.iloc[j].population_age_91,df_all.iloc[j].population_age_
92,df_all.iloc[j].population_age_93,df_all.iloc[j].population_age_94,df_all.il
oc[j].population_age_95,df_all.iloc[j].population_age_96,df_all.iloc[j].popula
tion_age_97,df_all.iloc[j].population_age_98,df_all.iloc[j].population_age_99,
df_all.iloc[j].population_age_100]
    elif(df_all.iloc[j].sex=='Male'):
        list1=[df_all.iloc[j].sex,df_all.iloc[j].infant_mortality_male,df_all.
iloc[j].life_expectancy_male,df_all.iloc[j].mortality_rate_under5_male,df_all.
iloc[j].mortality_rate_1to4_male,df_all.iloc[j].max_age,df_all.iloc[j].populat
ion_age_0,df_all.iloc[j].population_age_1,df_all.iloc[j].population_age_2,df_a
ll.iloc[j].population_age_3,df_all.iloc[j].population_age_4,df_all.iloc[j].pop
ulation_age_5,df_all.iloc[j].population_age_6,df_all.iloc[j].population_age_7,
```

```
df_all.iloc[j].population_age_8,df_all.iloc[j].population_age_9,df_all.iloc[j]
.population_age_10,df_all.iloc[j].population_age_11,df_all.iloc[j].population_
age_12,df_all.iloc[j].population_age_13,df_all.iloc[j].population_age_14,df_al
l.iloc[j].population_age_15,df_all.iloc[j].population_age_16,df_all.iloc[j].po
pulation_age_17,df_all.iloc[j].population_age_18,df_all.iloc[j].population_age
_19,df_all.iloc[j].population_age_20,df_all.iloc[j].population_age_21,df_all.i
loc[j].population_age_22,df_all.iloc[j].population_age_23,df_all.iloc[j].popul
ation_age_24,df_all.iloc[j].population_age_25,df_all.iloc[j].population_age_26
,df_all.iloc[j].population_age_27,df_all.iloc[j].population_age_28,df_all.iloc
[j].population_age_29,df_all.iloc[j].population_age_30,df_all.iloc[j].populati
on_age_31,df_all.iloc[j].population_age_32,df_all.iloc[j].population_age_33,df
_all.iloc[j].population_age_34,df_all.iloc[j].population_age_35,df_all.iloc[j]
.population_age_36,df_all.iloc[j].population_age_37,df_all.iloc[j].population_
age_38,df_all.iloc[j].population_age_39,df_all.iloc[j].population_age_40,df_al
l.iloc[j].population_age_41,df_all.iloc[j].population_age_42,df_all.iloc[j].po
pulation_age_43,df_all.iloc[j].population_age_44,df_all.iloc[j].population_age
_45,df_all.iloc[j].population_age_46,df_all.iloc[j].population_age_47,df_all.i
loc[j].population_age_48,df_all.iloc[j].population_age_49,df_all.iloc[j].popul
ation_age_50,df_all.iloc[j].population_age_51,df_all.iloc[j].population_age_52
,df_all.iloc[j].population_age_53,df_all.iloc[j].population_age_54,df_all.iloc
[j].population_age_55,df_all.iloc[j].population_age_56,df_all.iloc[j].populati
on_age_57,df_all.iloc[j].population_age_58,df_all.iloc[j].population_age_59,df
_all.iloc[j].population_age_60,df_all.iloc[j].population_age_61,df_all.iloc[j]
.population_age_62,df_all.iloc[j].population_age_63,df_all.iloc[j].population_
age_64,df_all.iloc[j].population_age_65,df_all.iloc[j].population_age_66,df_al
l.iloc[j].population_age_67,df_all.iloc[j].population_age_68,df_all.iloc[j].po
pulation_age_69,df_all.iloc[j].population_age_70,df_all.iloc[j].population_age
_71,df_all.iloc[j].population_age_72,df_all.iloc[j].population_age_73,df_all.i
loc[j].population_age_74,df_all.iloc[j].population_age_75,df_all.iloc[j].popul
ation_age_76,df_all.iloc[j].population_age_77,df_all.iloc[j].population_age_78
,df_all.iloc[j].population_age_79,df_all.iloc[j].population_age_80,df_all.iloc
[j].population_age_81,df_all.iloc[j].population_age_82,df_all.iloc[j].populati
on_age_83,df_all.iloc[j].population_age_84,df_all.iloc[j].population_age_85,df
_all.iloc[j].population_age_86,df_all.iloc[j].population_age_87,df_all.iloc[j]
.population_age_88,df_all.iloc[j].population_age_89,df_all.iloc[j].population_
age_90,df_all.iloc[j].population_age_91,df_all.iloc[j].population_age_92,df_al
l.iloc[j].population_age_93,df_all.iloc[j].population_age_94,df_all.iloc[j].po
pulation_age_95,df_all.iloc[j].population_age_96,df_all.iloc[j].population_age
_97,df_all.iloc[j].population_age_98,df_all.iloc[j].population_age_99,df_all.i
loc[j].population_age_100]
    elif(df_all.iloc[j].sex=='null'):
        list1="null"

    list_fm.append(list1)
#df_all['sexe']=list_fm

#df_all.drop(['sex','infant_mortality_male','infant_mortality_female','life_ex
pectancy_male','life_expectancy_female','mortality_rate_under5_female','mortal
ity_rate_under5_male','mortality_rate_1to4_male','mortality_rate_1to4_femal
e','max_age','population_age_0','population_age_1','population_age_2','populat
ion_age_3','population_age_4','population_age_5','population_age_6','populatio
n_age_7','population_age_8','population_age_9','population_age_10','population
_age_11','population_age_12','population_age_13','population_age_14','populati
on_age_15','population_age_16','population_age_17','population_age_18','popula
tion_age_19','population_age_20','population_age_21','population_age_22','popu
lation_age_23','population_age_24','population_age_25','population_age_26','po
pulation_age_27','population_age_28','population_age_29','population_age_3
```

```
0','population_age_31','population_age_32','population_age_33','population_age
_34','population_age_35','population_age_36','population_age_37','population_a
ge_38','population_age_39','population_age_40','population_age_41','population
_age_42','population_age_43','population_age_44','population_age_45','populati
on_age_46','population_age_47','population_age_48','population_age_49','popula
tion_age_50','population_age_51','population_age_52','population_age_53','popu
lation_age_54','population_age_55','population_age_56','population_age_57','po
pulation_age_58','population_age_59','population_age_60','population_age_6
1','population_age_62','population_age_63','population_age_64','population_age
_65','population_age_66','population_age_67','population_age_68','population_a
ge_69','population_age_70','population_age_71','population_age_72','population
_age_73','population_age_74','population_age_75','population_age_76','populati
on_age_77','population_age_78','population_age_79','population_age_80','popula
tion_age_81','population_age_82','population_age_83','population_age_84','popu
lation_age_85','population_age_86','population_age_87','population_age_88','po
pulation_age_89','population_age_90','population_age_91','population_age_9
2','population_age_93','population_age_94','population_age_95','population_age
_96','population_age_97','population_age_98','population_age_99','population_a
ge_100'],axis=1,inplace=True)

#df_all=df_all.groupby(['country_name','year','infant_mortality','life_expecta
ncy','mortality_rate_under5','mortality_rate_1to4','midyear_population','ferti
lity_rate_15_19','fertility_rate_20_24','fertility_rate_25_29','fertility_rate
_30_34','fertility_rate_35_39','fertility_rate_40_44','fertility_rate_45_4
9','total_fertility_rate','gross_reproduction_rate','sex_ratio_at_birth','coun
try_area','crude_birth_rate','crude_death_rate','net_migration','rate_natural_
increase','growth_rate'])['sexe'].apply(list).reset_index()
#df_all=df_all.merge(df_myp5,on=['country_name','year'],how="outer")
#df_all
```

In [8]:
```
df_all['sexe']=list_fm
```

In [12]:
```
df_all=df_all.merge(df_myp5,on=['country_name','year'],how="outer")
```

In [13]:
```
df_all.infant_mortality = df_all.infant_mortality.map(lambda x: str(x).replace
(",","."))
df_all.life_expectancy = df_all.life_expectancy.map(lambda x: str(x).replace(
",","."))
df_all.mortality_rate_under5 = df_all.mortality_rate_under5.map(lambda x: str(
x).replace(",","."))
df_all.mortality_rate_1to4 = df_all.mortality_rate_1to4.map(lambda x: str(x).r
eplace(",","."))
df_all.fertility_rate_15_19 = df_all.fertility_rate_15_19.map(lambda x: str(x)
.replace(",","."))
df_all.fertility_rate_20_24 = df_all.fertility_rate_20_24.map(lambda x: str(x)
.replace(",","."))
df_all.fertility_rate_25_29 = df_all.fertility_rate_25_29.map(lambda x: str(x)
.replace(",","."))
df_all.fertility_rate_30_34 = df_all.fertility_rate_30_34.map(lambda x: str(x)
.replace(",","."))
df_all.fertility_rate_35_39 = df_all.fertility_rate_35_39.map(lambda x: str(x)
.replace(",","."))
df_all.fertility_rate_40_44 = df_all.fertility_rate_40_44.map(lambda x: str(x)
.replace(",","."))
df_all.fertility_rate_45_49 = df_all.fertility_rate_45_49.map(lambda x: str(x)
.replace(",","."))
df_all.total_fertility_rate = df_all.total_fertility_rate.map(lambda x: str(x)
.replace(",","."))
df_all.gross_reproduction_rate = df_all.gross_reproduction_rate.map(lambda x:
str(x).replace(",","."))
df_all.sex_ratio_at_birth = df_all.sex_ratio_at_birth.map(lambda x: str(x).rep
lace(",","."))
df_all.crude_birth_rate = df_all.crude_birth_rate.map(lambda x: str(x).replace
(",","."))
df_all.crude_death_rate = df_all.crude_death_rate.map(lambda x: str(x).replace
(",","."))
df_all.net_migration = df_all.net_migration.map(lambda x: str(x).replace(",",
"."))
df_all.rate_natural_increase = df_all.rate_natural_increase.map(lambda x: str(
x).replace(",","."))
df_all.growth_rate = df_all.growth_rate.map(lambda x: str(x).replace(",","."))
df_all.country_area = df_all.country_area.map(lambda x: str(x).replace(","."
))
```

In [14]:
```
df_all=df_all.fillna("null",axis=0)
```

Out[14]:

	country_name	year	infant_mortality	life_expectancy	mortality_rate_under5	mortality_rat
0	Afghanistan	1950	null	null	null	
1	Afghanistan	1951	null	null	null	
2	Afghanistan	1952	null	null	null	
3	Afghanistan	1953	null	null	null	
4	Afghanistan	1954	null	null	null	
...	...	...	...	...	...	
23074	Zimbabwe	2046	15.08	65.52	22.66	
23075	Zimbabwe	2047	14.72	65.92	22.12	
23076	Zimbabwe	2048	14.35	66.33	21.55	
23077	Zimbabwe	2049	13.98	66.75	20.98	
23078	Zimbabwe	2050	13.62	67.17	20.41	

23079 rows × 25 columns

In [15]:
```
df_all.to_csv('out3.csv',index=False)
```

In [17]:

```
def flag_dict(i):
    full_flag=[]
    #for i in range(len(df_all)): #if 1 one country one year
        #print(df_all.iloc[i])
    if(df_all.flag[i]=="null"):
        return ("null");

    else :
        for j in range(len(df_all.iloc[i].flag)):
            #print(df_all.iloc[i].flag[j])
            dict1={'total_flag': str(df_all.iloc[i].flag[j][0]), 'starting_ag
e': str(df_all.iloc[i].flag[j][1]), 'age_group_indication': str(df_all.iloc[i]
.flag[j][2]), 'ending_age': str(df_all.iloc[i].flag[j][3]),'midyear_populatio
n': str(df_all.iloc[i].flag[j][4]), 'midyear_population_male': str(df_all.iloc
[i].flag[j][5]),'midyear_population_female': str(df_all.iloc[i].flag[j][6])}
            full_flag.append(dict1)
            """
            x=len(df_all.iloc[i].flag)-1
            if j!=x:
                full_flag=full_flag+', """
    return(full_flag)

def Male_index(i):
    if((df_all.iloc[i].sexe[0][0])=="Male"):
        return '0'
    else: return '1'

def Female_index(i):
    if((df_all.iloc[i].sexe[0][0])=="Female"):
        return '0'
    else: return '1'

def Male_dict(i):
    if(df_all.sexe[i][0]!="null"):
        mix=int(Male_index(i))
        fix=int(Female_index(i))
        dict1={'infant_mortality_male': str(df_all.iloc[i].sexe[mix][1]).repla
ce(",",".") , 'life_expectancy_male': str(df_all.iloc[i].sexe[mix][2]).replace
(",","."), 'mortality_rate_under5_male': str(df_all.iloc[i].sexe[mix][3]).repl
ace(",","."), 'mortality_rate_1to4_male': str(df_all.iloc[i].sexe[mix][4]).rep
lace(",","."),'max_age': str(df_all.iloc[i].sexe[mix][5]).replace(",",".")}
        for j in range(101):
            key='population_age_%s'%(j)
            key_value=str(df_all.iloc[i].sexe[mix][j+5]).replace(",",".")
            dict1[key] = key_value
    else : dict1="null"
    return (dict1)

def Female_dict(i):
    if(df_all.sexe[i][0]!="null"):
        mix=int(Male_index(i))
        fix=int(Female_index(i))
        dict1={'infant_mortality_female': str(df_all.iloc[i].sexe[fix][1]).rep
lace(",",".") , 'life_expectancy_female': str(df_all.iloc[i].sexe[fix][2]).rep
lace(",","."), 'mortality_rate_under5_female': str(df_all.iloc[i].sexe[fix][3
]).replace(",","."), 'mortality_rate_1to4_female': str(df_all.iloc[i].sexe[fix
```

```
][4]).replace(",",".'),'max_age': str(df_all.iloc[i].sexe[fix][5]).replace(","
,".")}
        for j in range(101):
            key='population_age_%s'%(j)
            key_value=str(df_all.iloc[i].sexe[fix][j+5]).replace(",",".")
            dict1[key] = key_value
    else : dict1="null"
    return (dict1)

def creation_json(df_all):
    fh = open("all4.json", "a+")
    all_data=[]
    for i in range(len(df_all)):
        y=str(df_all.iloc[i].year)
        cn=str(df_all.iloc[i].country_name)
        ca=str(df_all.iloc[i].country_area)
        cbr=str(df_all.iloc[i].crude_birth_rate)
        cdr=str(df_all.iloc[i].crude_death_rate)
        nm=str(df_all.iloc[i].net_migration)
        rni=str(df_all.iloc[i].rate_natural_increase)
        gr=str(df_all.iloc[i].growth_rate)
        fr1519=str(df_all.iloc[i].fertility_rate_15_19)
        fr2024=str(df_all.iloc[i].fertility_rate_20_24)
        fr2529=str(df_all.iloc[i].fertility_rate_25_29)
        fr3034=str(df_all.iloc[i].fertility_rate_30_34)
        fr3539=str(df_all.iloc[i].fertility_rate_35_39)
        fr4044=str(df_all.iloc[i].fertility_rate_40_44)
        fr4549=str(df_all.iloc[i].fertility_rate_45_49)
        tfr=str(df_all.iloc[i].total_fertility_rate)
        grr=str(df_all.iloc[i].gross_reproduction_rate)
        srab=str(df_all.iloc[i].sex_ratio_at_birth)
        mpx=str(df_all.iloc[i].midyear_population)
        im=str(df_all.iloc[i].infant_mortality)
        le=str(df_all.iloc[i].life_expectancy)
        mru5=str(df_all.iloc[i].mortality_rate_under5)
        mr1t4=str(df_all.iloc[i].mortality_rate_1to4)

        data={"_id":{"country_name": cn, "year": y},"country_stats":{"country_
area":ca,"crude_birth_rate":cbr,"crude_death_rate":cdr,"net_migration":nm,"rat
e_natural_increase":rni,"growth_rate":gr,"fertility_rate_15_19":fr1519,"fertil
ity_rate_20_24":fr2024,"fertility_rate_25_29":fr2529,"fertility_rate_30_34":fr
3034,"fertility_rate_35_39":fr3539,"fertility_rate_40_44":fr4044,"fertility_ra
te_45_49":fr4549,"total_fertility_rate":tfr,"gross_reproduction_rate":grr,"sex
_ratio_at_birth":srab,"midyear_population":mpx,"life_expectancy":le,"mortality
_rate_under5":mru5,'mortality_rate_1to4':mr1t4, "Flag":flag_dict(i),"Male":Mal
e_dict(i),"Female":Female_dict(i)}}
        all_data.append(data)
    fh.write(json.dumps(all_data))
    fh.close()
```

In [18]:
```
creation_json(df_all)
```

In [ ]:
```

```