Intitulé du projet : Analyse des données de santé « Cas du Paludisme en Afrique »

Contenu du Dataset « Africa Malaria » :

L'ensemble de données "Africa Malaria" comprend des données sur les pays africains de 2007 à 2017, avec les caractéristiques suivantes :

- Code de pays ISO-3 unique : Chaque pays est identifié par un code de pays ISO-3, qui est un code standardisé utilisé pour représenter les pays dans les données internationales.
- Latitude et longitude : Pour chaque pays, l'ensemble de données fournit également les coordonnées de latitude et de longitude, qui donnent la position géographique approximative du pays.
- Cas de paludisme signalés : L'ensemble de données comprend des informations sur les cas de paludisme signalés dans chaque pays et chaque année. Ces données peuvent inclure le nombre total de cas, le nombre de cas selon le sexe, l'âge ou d'autres caractéristiques démographiques, ainsi que la gravité des cas signalés.
- Mesures préventives : L'ensemble de données fournit également des données sur les mesures préventives prises pour lutter contre le paludisme dans chaque pays. Cela peut inclure des informations sur les campagnes de sensibilisation, les programmes de distribution de moustiquaires, les traitements médicaux administrés, etc.

1. Importation des bibliothèques pandas, numpy et matplotlib, pour manipuler les données et effectuer des analyses statistiques

In [101]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

2. Chargement du fichier du Dataset dans un dataframe pandas

Lien de téléchargement du Dataset : https://www.kaggle.com/datasets/lydia70/malaria-in-africa (https://www.kaggle.com/datasets/lydia70/malaria-in-africa)

In [102]:

```
df = pd.read_csv("DatasetAfricaMalaria.csv")
```

3. Nettoyage et structuration des données pour l'analyse

3.1. Copie du Dataset

In [103]:

```
dfc = df.copy()
```

3.2. Affichage des 5 premières lignes du Dataset

In [104]:

```
dfc.head()
```

Out[104]:

	Country Name	Year	Country Code	Incidence of malaria (per 1,000 population at risk)	Malaria cases reported	Use of insecticide-treated bed nets (% of under-5 population)	Children with fever receiving antimalarial drugs (% of children under age 5 with fever)	Intermittent preventive treatment (IPT) of malaria in pregnancy (% of pregnant women)	People using safely managed drinking water services (% of population)	People using safely managed drinking water services, rural (% of rural population)	...	Urban population growth (annual %)	People using at least basic drinking water services (% of population)	People using at least basic drinking water services, rural (% of rural population)
0	Algeria	2007	DZA	0.01	26.0	NaN	NaN	NaN	NaN	NaN	...	2.71	91.68	85.83
1	Angola	2007	AGO	286.72	1533485.0	18.0	29.8	1.5	NaN	NaN	...	5.01	47.96	23.77
2	Benin	2007	BEN	480.24	0.0	NaN	NaN	NaN	NaN	NaN	...	4.09	63.78	54.92
3	Botswana	2007	BWA	1.03	390.0	NaN	NaN	NaN	NaN	NaN	...	4.80	78.89	57.60
4	Burkina Faso	2007	BFA	503.80	44246.0	NaN	NaN	NaN	NaN	NaN	...	5.91	52.27	45.13

5 rows × 27 columns

3.3. Affichage des 5 dernières lignes du Dataset

In [105]:

```
dfc.tail()
```

Out[105]:

	Country Name	Year	Country Code	Incidence of malaria (per 1,000 population at risk)	Malaria cases reported	Use of insecticide-treated bed nets (% of under-5 population)	Children with fever receiving antimalarial drugs (% of children under age 5 with fever)	Intermittent preventive treatment (IPT) of malaria in pregnancy (% of pregnant women)	People using safely managed drinking water services (% of population)	People using safely managed drinking water services, rural (% of rural population)	...	Urban population growth (annual %)	People using at least basic drinking water services (% of population)	Pe usin least b drink w servi rural (% r populat
589	Togo	2017	TGO	278.20	1755577.0	69.7	31.1	41.7	NaN	NaN	...	3.79	65.13	4
590	Tunisia	2017	TUN	NaN	NaN	NaN	NaN	NaN	92.66	NaN	...	1.57	96.25	8
591	Uganda	2017	UGA	336.76	11667831.0	NaN	NaN	NaN	7.07	4.46	...	6.25	49.10	4
592	Zambia	2017	ZMB	160.05	5505639.0	NaN	NaN	NaN	NaN	NaN	...	4.21	59.96	4
593	Zimbabwe	2017	ZWE	108.55	467508.0	NaN	NaN	NaN	NaN	NaN	...	1.28	64.05	4

5 rows × 27 columns

3.4. Examen des différentes colonnes et types de données dans le Dataset

In [106]:

```
dfc.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 594 entries, 0 to 593
Data columns (total 27 columns):
 #   Column                                                                  Non-Null Count  Dtype
---  ------                                                                  --------------  -----
 0   Country Name                                                            594 non-null    object
 1   Year                                                                    594 non-null    int64
 2   Country Code                                                            594 non-null    object
 3   Incidence of malaria (per 1,000 population at risk)                     550 non-null    float64
 4   Malaria cases reported                                                  550 non-null    float64
 5   Use of insecticide-treated bed nets (% of under-5 population)           132 non-null    float64
 6   Children with fever receiving antimalarial drugs (% of children under age 5 with fever)  122 non-null  float64
 7   Intermittent preventive treatment (IPT) of malaria in pregnancy (% of pregnant women)  106 non-null  float64
 8   People using safely managed drinking water services (% of population)   99 non-null     float64
 9   People using safely managed drinking water services, rural (% of rural population)  88 non-null  float64
 10  People using safely managed drinking water services, urban (% of urban population)  176 non-null  float64
 11  People using safely managed sanitation services (% of population)       132 non-null    float64
 12  People using safely managed sanitation services, rural (% of rural population)  110 non-null  float64
 13  People using safely managed sanitation services, urban  (% of urban population)  132 non-null  float64
 14  Rural population (% of total population)                                588 non-null    float64
 15  Rural population growth (annual %)                                      588 non-null    float64
 16  Urban population (% of total population)                                588 non-null    float64
 17  Urban population growth (annual %)                                      588 non-null    float64
 18  People using at least basic drinking water services (% of population)  588 non-null    float64
 19  People using at least basic drinking water services, rural (% of rural population)  566 non-null  float64
 20  People using at least basic drinking water services, urban (% of urban population)  566 non-null  float64
 21  People using at least basic sanitation services (% of population)      588 non-null    float64
 22  People using at least basic sanitation services, rural (% of rural population)  566 non-null  float64
 23  People using at least basic sanitation services, urban  (% of urban population)  566 non-null  float64
 24  latitude                                                                594 non-null    float64
 25  longitude                                                               594 non-null    float64
 26  geometry                                                                594 non-null    object
dtypes: float64(23), int64(1), object(3)
memory usage: 125.4+ KB
```

Nombre d'index dans notre Dataset

In [107]:

```
dfc.index
```

Out[107]:

```
RangeIndex(start=0, stop=594, step=1)
```

Conversion du type de données de la colonne "Malaria cases reported" en nombre entier

In [108]:

```
dfc['Malaria cases reported']
```

Out[108]:

```
0            26.0
1       1533485.0
2             0.0
3           390.0
4         44246.0
           ...
589     1755577.0
590           NaN
591    11667831.0
592     5505639.0
593      467508.0
Name: Malaria cases reported, Length: 594, dtype: float64
```

In [109]:

```
dfc['Malaria cases reported'] = dfc['Malaria cases reported'].astype('Int64')
```

In [110]:

```
dfc['Malaria cases reported']
```

Out[110]:

```
0            26
1       1533485
2             0
3           390
4         44246
          ...
589     1755577
590        <NA>
591    11667831
592     5505639
593      467508
Name: Malaria cases reported, Length: 594, dtype: Int64
```

Renommons certaines colonnes de notre Dataset

In [111]:

```
### Listons les noms de colonnes du Dataset avant renommage
list(dfc)
```

Out[111]:

```
['Country Name',
 'Year',
 'Country Code',
 'Incidence of malaria (per 1,000 population at risk)',
 'Malaria cases reported',
 'Use of insecticide-treated bed nets (% of under-5 population)',
 'Children with fever receiving antimalarial drugs (% of children under age 5 with fever)',
 'Intermittent preventive treatment (IPT) of malaria in pregnancy (% of pregnant women)',
 'People using safely managed drinking water services (% of population)',
 'People using safely managed drinking water services, rural (% of rural population)',
 'People using safely managed drinking water services, urban (% of urban population)',
 'People using safely managed sanitation services (% of population)',
 'People using safely managed sanitation services, rural (% of rural population)',
 'People using safely managed sanitation services, urban  (% of urban population)',
 'Rural population (% of total population)',
 'Rural population growth (annual %)',
 'Urban population (% of total population)',
 'Urban population growth (annual %)',
 'People using at least basic drinking water services (% of population)',
 'People using at least basic drinking water services, rural (% of rural population)',
 'People using at least basic drinking water services, urban (% of urban population)',
 'People using at least basic sanitation services (% of population)',
 'People using at least basic sanitation services, rural (% of rural population)',
 'People using at least basic sanitation services, urban  (% of urban population)',
 'latitude',
 'longitude',
 'geometry']
```

Renommage des colonnes du Dataset

In [112]:

```
n  (% of urban population)':'% of urban population using at least basic sanitation services', 'geometry':'Localisation'}, inplace = True)
```

In [113]:

```
list(dfc)
```

Out[113]:

```
['Country Name',
 'Year',
 'Country Code',
 'Incidence of malaria (per 1,000 population at risk)',
 'Malaria cases reported',
 'Use of insecticide-treated bed nets (% of under-5 population)',
 '% of children under age 5 with fever receiving antimalarial drugs',
 '% of pregnant women using Intermittent preventive treatment (IPT) of malaria in pregnancy',
 '% of population using safely managed drinking water services',
 '% of rural population using safely managed drinking water services',
 '% of urban population using safely managed drinking water services',
 '% of population using safely managed sanitation services',
 '% of rural population using safely managed sanitation services',
 '% of urban population using safely managed sanitation services',
 'Rural population (% of total population)',
 'Rural population growth (annual %)',
 'Urban population (% of total population)',
 'Urban population growth (annual %)',
 '% of population using at least basic drinking water services',
 '% of rural population using at least basic drinking water services',
 '% of urban population using at least basic drinking water services',
 ' % of population using at least basic sanitation services',
 '% of rural population using at least basic sanitation services',
 '% of urban population using at least basic sanitation services',
 'latitude',
 'longitude',
 'Localisation']
```

4. Analyse exploratoire des données

4.1. Statistiques descriptives du Dataset

In [114]:

```
df.describe()
```

Out[114]:

	Year	Incidence of malaria (per 1,000 population at risk)	Malaria cases reported	Use of insecticide-treated bed nets (% of under-5 population)	Children with fever receiving antimalarial drugs (% of children under age 5 with fever)	Intermittent preventive treatment (IPT) of malaria in pregnancy (% of pregnant women)	People using safely managed drinking water services (% of population)	People using safely managed drinking water services, rural (% of rural population)	People using safely managed drinking water services, urban (% of urban population)	People using safely managed sanitation services (% of population)	...	Urban population (% of total population)	po (a
count	594.000000	550.000000	5.500000e+02	132.000000	122.000000	106.000000	99.000000	88.000000	176.000000	132.000000	...	588.000000	58
mean	2012.000000	190.087491	1.068330e+06	42.530303	30.201639	15.013958	33.478990	12.470568	51.549545	28.768939	...	43.164116	
std	3.164943	163.054527	2.192802e+06	20.157059	18.903198	12.389166	26.678321	10.078371	24.157416	18.631510	...	18.086118	
min	2007.000000	0.000000	0.000000e+00	1.000000	0.500000	0.000000	5.770000	0.930000	11.200000	6.370000	...	9.860000	
25%	2009.000000	30.857500	2.211750e+03	26.675000	17.275000	5.763285	8.975000	4.185000	34.125000	16.532500	...	28.795000	
50%	2012.000000	174.775000	1.130260e+05	42.900000	29.300000	11.500000	28.390000	10.675000	51.365000	25.410000	...	41.560000	
75%	2015.000000	347.637500	1.154808e+06	56.325000	42.625000	21.850000	43.890000	16.887500	70.747500	35.725000	...	56.945000	
max	2017.000000	585.540000	1.682113e+07	95.500000	76.900000	59.600000	92.660000	39.930000	89.540000	78.120000	...	88.980000	

8 rows × 24 columns

4.2. Analyses statistiques pour identifier les tendances des maladies, les facteurs de risque

*** Tendances annuelles des cas de paludisme, classées par année du nombre de cas le plus élevé de paludisme au nombre de cas le plus faible**

In [115]:

```
som_tendances_annuelles = dfc.groupby("Year")["Malaria cases reported"].sum().sort_values(ascending=False)
```

In [117]:

```
som_tendances_annuelles
```

Out[117]:

```
Year
2017    128146255
2016    123330818
2015     92968526
2014     77028814
2013     47174959
2012     39719221
2011     23002865
2010     22184760
2009     14414616
2007     10102341
2008      9508374
Name: Malaria cases reported, dtype: Int64
```

In [118]:

```
moy_tendances_annuelles = dfc.groupby("Year")["Malaria cases reported"].mean().sort_values(ascending=False)
```

In [119]:

```
moy_tendances_annuelles
```

Out[119]:

```
Year
2017     2562925.1
2016    2466616.36
2015    1859370.52
2014    1540576.28
2013     943499.18
2012     794384.42
2011      460057.3
2010      443695.2
2009     288292.32
2007     202046.82
2008     190167.48
Name: Malaria cases reported, dtype: Float64
```

*** Tendances par pays des cas de paludisme de 2007-2017, classé par pays ayant le nombre de cas de paludisme le plus élevé au nombre de cas le plus faible**

In [120]:

```
som_tendances_pays = dfc.groupby("Country Name")["Malaria cases reported"].sum().sort_values(ascending=False)
```

In [121]:

```
som_tendances_pays
```

Out[121]:

```
Country Name
Congo, Dem. Rep.            77555524
Mozambique                  43984158
Burkina Faso                41601963
Uganda                      41126230
Burundi                     40249714
Nigeria                     40066570
Ghana                       28008309
Tanzania                    24283088
Angola                      24034182
Malawi                      19445640
Niger                       18866709
Zambia                      18619166
Cote d'Ivoire               17978339
Rwanda                      17498217
Kenya                       17353851
Ethiopia                    16266669
Liberia                     12237026
Mali                        12106257
Sierra Leone                11919725
Togo                        11162390
Benin                        8579720
Sudan                        7265570
Chad                         6198112
Madagascar                   4616832
Guinea                       4580444
Cameroon                     4373628
Zimbabwe                     3344410
Senegal                      3283980
South Sudan                  3143581
Central African Republic     2561261
Gambia, The                  1675997
Congo, Rep.                   932292
Guinea-Bissau                 749386
Equatorial Guinea             336825
Eritrea                       306876
Comoros                       289016
Gabon                         264144
Somalia                       254342
Namibia                       120908
Mauritania                    117335
South Africa                   93089
Djibouti                       61718
Sao Tome and Principe          54276
Botswana                        8756
Eswatini                        4531
Cabo Verde                       699
Algeria                           94
Seychelles                         0
Libya                              0
Egypt, Arab Rep.                   0
Tunisia                            0
Morocco                            0
Mauritius                          0
Lesotho                            0
Name: Malaria cases reported, dtype: Int64
```

In [122]:

```
moy_tendances_pays = dfc.groupby("Country Name")["Malaria cases reported"].mean().sort_values(ascending=False)
```

In [123]:

```
moy_tendances_pays
```

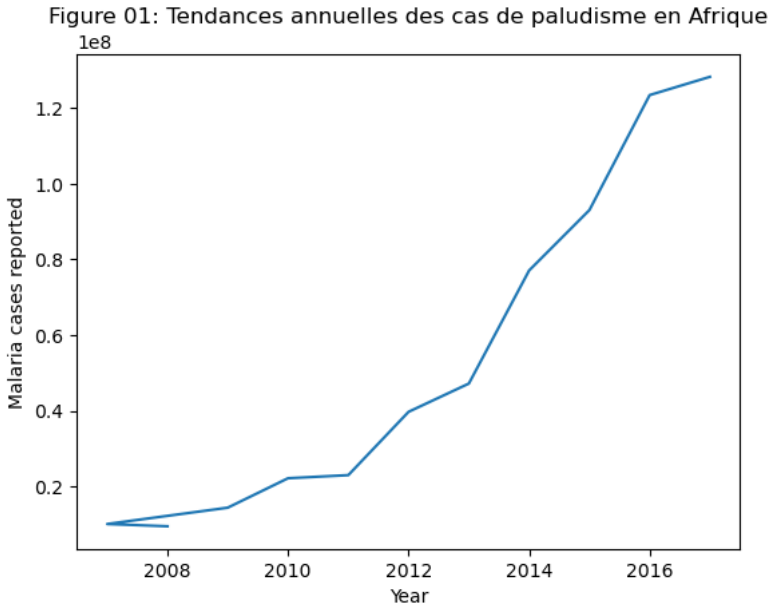
Out[123]:

```
Country Name
Congo, Dem. Rep.            7050502.181818
Mozambique                  3998559.818182
Burkina Faso                3781996.636364
Uganda                      3738748.181818
Burundi                     3659064.909091
Nigeria                     3642415.454545
Ghana                       2546209.909091
Tanzania                    2207553.454545
Angola                      2184925.636364
Malawi                      1767785.454545
Niger                       1715155.363636
Zambia                      1692651.454545
Cote d'Ivoire               1634394.454545
Rwanda                         1590747.0
Kenya                       1577622.818182
Ethiopia                    1478788.090909
Liberia                     1112456.909091
Mali                        1100568.818182
Sierra Leone                1083611.363636
Togo                        1014762.727273
Benin                        779974.545455
Sudan                        660506.363636
Chad                         563464.727273
Madagascar                       419712.0
Guinea                           416404.0
Cameroon                     397602.545455
Zimbabwe                     304037.272727
Senegal                      298543.636364
South Sudan                  285780.090909
Central African Republic     232841.909091
Gambia, The                  152363.363636
Congo, Rep.                   84753.818182
Guinea-Bissau                    68126.0
Equatorial Guinea             30620.454545
Eritrea                       27897.818182
Comoros                       26274.181818
Gabon                         24013.090909
Somalia                          23122.0
Namibia                       10991.636364
Mauritania                    10666.818182
South Africa                   8462.636364
Djibouti                       5610.727273
Sao Tome and Principe          4934.181818
Botswana                           796.0
Eswatini                      411.909091
Cabo Verde                     63.545455
Algeria                         8.545455
Egypt, Arab Rep.                     0.0
Morocco                              0.0
Libya                                0.0
Lesotho                             <NA>
Mauritius                           <NA>
Seychelles                          <NA>
Tunisia                             <NA>
Name: Malaria cases reported, dtype: Float64
```

**** Visualisation des résultats des tendances annuelles***

In [135]:

```
som_tendances_annuelles.plot(kind="line")
plt.xlabel("Year")
plt.ylabel("Malaria cases reported")
plt.title("Figure 01: Tendances annuelles des cas de paludisme en Afrique")
plt.show()
```



Lecture de la Figure 01 : la tendance est que le nombre de cas de paludisme en Afrique n'a fait qu'augmenter d'année en année entre 2007 et 2017

*** Tendances des cas de paludisme en fonction de l'utilisation de moustiquaires imprégnées d'insecticide (% de la population de moins de 5 ans)**

In [131]:

```
tendances_treated_bed_nets = dfc.groupby("Use of insecticide-treated bed nets (% of under-5 population)")["Malaria cases reported"].sum()
```

In [132]:

```
som_tendances_treated_bed_nets
```

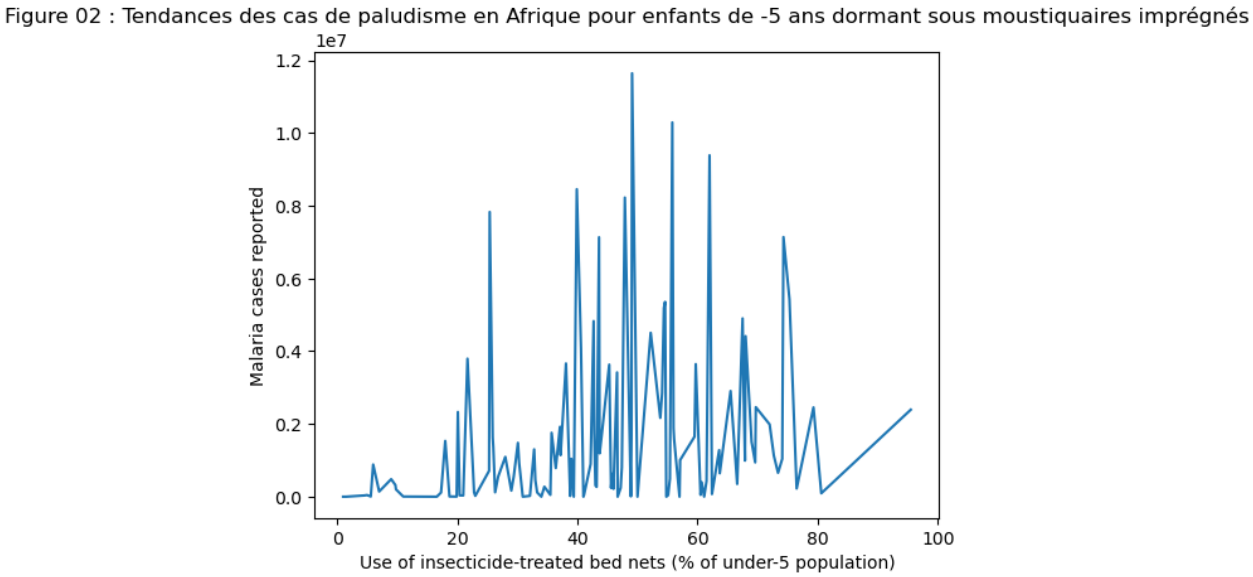
Out[132]:

```
Use of insecticide-treated bed nets (% of under-5 population)
1.0          84
1.5         268
4.5       33405
5.0       44518
5.6        4911
          ...
75.3    5428655
76.5     224498
79.3    2454508
80.6      93431
95.5    2392108
Name: Malaria cases reported, Length: 117, dtype: Int64
```

*** Visualisation de la tendance des cas de paludisme en Afrique pour enfants de -5 ans dormant sous moustiquaires imprégnés**

In [136]:

```
som_tendances_treated_bed_nets.plot(kind="line")
plt.xlabel("Use of insecticide-treated bed nets (% of under-5 population)")
plt.ylabel("Malaria cases reported")
plt.title("Figure 02 : Tendances des cas de paludisme en Afrique pour enfants de -5 ans dormant sous moustiquaires imprégnés")
plt.show()
```



Lecture de la Figure 02 : la tendance est que plus les enfants de moins de 5 ans dorment sous moustiquaires imprégnés, moins ils ont le paludisme.

4.3. Corrélations entre les variables pour analyser les relations entre les facteurs de risque et les cas de paludisme

Definition : La relation statistique entre deux variables est appelée leur corrélation.

In [138]:

```
correlation_matrix = dfc.corr()
```

```
C:\Users\HP ZBOOK\AppData\Local\Temp\ipykernel_28904\2508308341.py:1: FutureWarning: The default value of numeric_only in D
ataFrame.corr is deprecated. In a future version, it will default to False. Select only valid columns or specify the value
of numeric_only to silence this warning.
  correlation_matrix = dfc.corr()
```

In [139]:

```
correlation_matrix
```

Out[139]:

	Year	Incidence of malaria (per 1,000 population at risk)	Malaria cases reported	Use of insecticide-treated bed nets (% of under-5 population)	% of children under age 5 with fever receiving antimalarial drugs	% of pregnant women using Intermittent preventive treatment (IPT) of malaria in pregnancy	% of population using safely managed drinking water services	% of rural population using safely managed drinking water services	% of urban population using safely managed drinking water services	% of population using safely managed sanitation services	...	Urban population (% of total population)	po (ar
Year	1.000000e+00	-0.070605	0.371363	0.531014	-0.043820	0.454461	0.089462	0.171102	0.050082	0.081409	...	0.084292	-(
Incidence of malaria (per 1,000 population at risk)	-7.060502e-02	1.000000	0.288509	0.009608	0.532327	0.154911	-0.501461	-0.459303	-0.678977	-0.613592	...	-0.253803	(
Malaria cases reported	3.713634e-01	0.288509	1.000000	0.271116	0.301719	0.238335	-0.278893	-0.197603	-0.408585	-0.172106	...	-0.219683	(
Use of insecticide-treated bed nets (% of under-5 population)	5.310142e-01	0.009608	0.271116	1.000000	0.079162	0.307991	0.051592	-0.099520	-0.121495	-0.293166	...	-0.136613	(
% of children under age 5 with fever receiving antimalarial drugs	-4.381980e-02	0.532327	0.301719	0.079162	1.000000	0.314459	-0.208050	-0.156504	-0.616924	-0.153175	...	-0.090201	(
% of pregnant women using Intermittent preventive treatment (IPT) of malaria in pregnancy	4.544615e-01	0.154911	0.238335	0.307991	0.314459	1.000000	0.184018	-0.272432	0.123014	-0.201184	...	0.178811	-(
% of population using safely managed drinking water services	8.946214e-02	-0.501461	-0.278893	0.051592	-0.208050	0.184018	1.000000	0.921974	0.946457	0.927503	...	0.822392	-(
% of rural population using safely managed drinking water services	1.711018e-01	-0.459303	-0.197603	-0.099520	-0.156504	-0.272432	0.921974	1.000000	0.814306	0.951713	...	0.695415	-(
% of urban population using safely managed drinking water services	5.008197e-02	-0.678977	-0.408585	-0.121495	-0.616924	0.123014	0.946457	0.814306	1.000000	0.904425	...	0.507907	-(
% of population using safely managed sanitation services	8.140907e-02	-0.613592	-0.172106	-0.293166	-0.153175	-0.201184	0.927503	0.951713	0.904425	1.000000	...	0.427114	-(
% of rural population using safely managed sanitation services	2.196390e-01	-0.205144	0.103046	0.038775	0.042544	-0.231655	0.950821	0.978063	-0.046646	0.650613	...	0.588600	-(
% of urban population using safely managed sanitation services	5.671254e-02	-0.499475	-0.106211	-0.198089	0.218336	-0.112146	0.898606	0.977609	0.947106	0.971281	...	0.406288	-(
Rural population (% of total population)	-8.429034e-02	0.253799	0.219674	0.136593	0.090178	-0.178830	-0.822383	-0.695418	-0.507889	-0.427094	...	-1.000000	(

	Year	Incidence of malaria (per 1,000 population at risk)	Malaria cases reported	Use of insecticide-treated bed nets (% of under-5 population)	% of children under age 5 with fever receiving antimalarial drugs	% of pregnant women using Intermittent preventive treatment (IPT) of malaria in pregnancy	% of population using safely managed drinking water services	% of rural population using safely managed drinking water services	% of urban population using safely managed drinking water services	% of population using safely managed sanitation services	...	Urban population (% of total population)	po (ar
Rural population growth (annual %)	-4.988517e-02	0.411329	0.254232	0.198001	0.289761	-0.043661	-0.826812	-0.826590	-0.729787	-0.368766	...	-0.651653	0
Urban population (% of total population)	8.429226e-02	-0.253803	-0.219683	-0.136613	-0.090201	0.178811	0.822392	0.695415	0.507907	0.427114	...	1.000000	-0
Urban population growth (annual %)	-5.670150e-02	0.350415	0.284905	0.214265	0.363973	-0.012669	-0.856482	-0.723214	-0.654459	-0.618049	...	-0.228287	1
% of population using at least basic drinking water services	1.194164e-01	-0.410028	-0.257250	0.034067	-0.207770	0.241572	0.870408	0.754185	0.735460	0.666937	...	0.631472	-0
% of rural population using at least basic drinking water services	1.138409e-01	-0.352018	-0.199931	0.146298	-0.154126	0.186519	0.774201	0.592671	0.661666	0.649492	...	0.290278	-0
% of urban population using at least basic drinking water services	1.034695e-01	-0.430043	-0.177815	0.004874	-0.394741	0.140941	0.936218	0.821300	0.942218	0.656757	...	0.280251	-0
% of population using at least basic sanitation services	6.887078e-02	-0.512391	-0.202425	-0.042730	-0.461137	-0.190158	0.886794	0.932769	0.742667	0.677943	...	0.518670	-0
% of rural population using at least basic sanitation services	7.023027e-02	-0.431779	-0.128802	0.057677	-0.396981	-0.224373	0.837097	0.867757	0.656613	0.703158	...	0.281079	-0
% of urban population using at least basic sanitation services	6.026910e-02	-0.479482	-0.231405	-0.150451	-0.463726	-0.172149	0.873025	0.925027	0.749673	0.736721	...	0.465425	-0
latitude	-1.513567e-15	[illegible]	[illegible]	[illegible]	[illegible]	0.098717	0.809231	0.761962	-0.309760	0.555364	...	0.227344	0
longitude	-1.679748e-15	-0.287883	0.072659	-0.011117	-0.110809	-0.203676	-0.321583	-0.520107	-0.050205	0.317706	...	-0.305732	-0

24 rows × 24 columns

Analyse des corrélations et identification des facteurs de risque potentiels en examinant les valeurs de corrélation

**** Corrélations entre les cas de paludisme et les variables démographiques***

- Avec Incidence du paludisme (pour 1 000 habitants à risque)

In [140]:

```
corr_cases_population_1 = correlation_matrix["Malaria cases reported"]["Incidence of malaria (per 1,000 population at risk)"]
```

In [141]:

```
corr_cases_population_1
```

Out[141]:

0.28850886831479716

- Avec % d'enfants de moins de 5 ans ayant de la fièvre recevant des médicaments antipaludiques

In [142]:

```
rr_cases_population_2 = correlation_matrix["Malaria cases reported"]["% of children under age 5 with fever receiving antimalarial drugs"]
```

In [143]:

```
corr_cases_population_2
```

Out[143]:

0.30171947845151337

- Avec l'utilisation de moustiquaires imprégnées d'insecticide (% de la population de moins de 5 ans)

In [144]:

```
corr_cases_population_3 = correlation_matrix["Malaria cases reported"]["Use of insecticide-treated bed nets (% of under-5 population)"]
```

In [145]:

```
corr_cases_population_3
```

Out[145]:

```
0.2711160562803437
```

- Avec la Population urbaine (% de la population totale)

In [146]:

```
corr_cases_population_4 = correlation_matrix["Malaria cases reported"]["Urban population growth (annual %)"]
```

In [147]:

```
corr_cases_population_4
```

Out[147]:

```
0.28490501349417036
```

Forte Corrélation entre "% of rural population using safely managed sanitation services" et "% of rural population using safely managed drinking water services"

In [148]:

```
x["% of rural population using safely managed sanitation services"]["% of rural population using safely managed drinking water services"]
```

In [149]:

```
corr_cases_population_5
```

Out[149]:

```
0.9780628308867829
```

Machine Learning : Modélisation des données pour identifier des modèles et des relations importantes dans les données

In [184]:

```
## Utilisons la bibliothèque scikit-learn pour effectuer une régression linéaire et importons le modèle de régression linéaire
from sklearn.linear_model import LinearRegression
## Le module Impute de Sklearn (scikit-learn) permet de nettoyer notre dataset des valeurs manquantes qui le composes.
## SimpleImputer remplace toute valeur manquante par une statistique ou une constante donnée
from sklearn.impute import SimpleImputer
```

In [201]:

```
## X représente des variables indépendantes : dans notre cas "% d'enfants de moins de 5 ans dormant sous moustiquaires imprégnées" et "% d
X = dfc[["Use of insecticide-treated bed nets (% of under-5 population)", "% of pregnant women using Intermittent preventive treatment (IP
## y représente la variable cible : dans notre cas "Year (Année)"
y = dfc["Year"]
```

In [203]:

```
## Utilisons SimpleImputer pour remplacer toute valeur manquante par une statistique ou une constante donnée. Dans notre cas, la moyenne d
imputer = SimpleImputer(strategy='mean', missing_values=np.nan)
## Transformons dans les variables indépendantes contenues dans X, les valeurs manquantes par np.nan
X_imputed = imputer.fit_transform(X)
```

In [206]:

```
## Utilisons SimpleImputer pour remplacer toute valeur manquante par une statistique ou une constante donnée. Dans notre cas, la moyenne d
imputer_y = SimpleImputer(strategy='mean')
## Appelons la méthode `reshape` sur `y.values`, pour remodeler le tableau en un tableau 2D avec une seule colonne,
## ce qui est le format d'entrée attendu pour la méthode `fit_transform` de la classe `SimpleImputer`.
y_imputed = imputer_y.fit_transform(y.values.reshape(-1, 1))
```

In [207]:

```
## Instancions ensuite un objet modèle de régression linéaire en utilisant la classe LinearRegression().
model = LinearRegression()
## À l'aide de la méthode fit(), entraînons le modèle en utilisant les données X et y.
model.fit(X_imputed, y_imputed)
## Utilisons la méthode predict() pour effectuer des prédictions sur les données d'entraînement X et stockez les prédictions
## résultantes dans la variable predictions.
predictions = model.predict(X_imputed)
```

Machine Learning : Evaluation du modèle de Regression Linéaire

In [210]:

```
## Importation de la fonction "mean_squared_error" du module "sklearn.metrics". Cela permettra de calculer l'erreur
## quadratique moyenne ou Mean Squared Error(MSE).
from sklearn.metrics import mean_squared_error

## Calcul de l'erreur quadratique moyenne (MSE) entre les valeurs réelles 'y' et les valeurs prédites 'predictions'
## en utilisant la fonction 'mean_squared_error'.
mse = mean_squared_error(y, predictions)

## Calcul de la racine carrée de l'erreur quadratique moyenne (MSE) en utilisant la fonction `np.sqrt` de numpy.
rmse = np.sqrt(mse)

## Affichage de la valeur du RMSE en utilisant la fonction 'print'
## Habituellement, un score RMSE inférieur à 180 est considéré comme un bon score pour un algorithme qui fonctionne modérément
## ou bien. Dans le cas où la valeur RMSE dépasse 180, nous devons effectuer une sélection de caractéristiques et un réglage
## des paramètres hyper sur les paramètres du modèle.
print("RMSE:", rmse)
```

```
RMSE: 3.0520908649792458
```

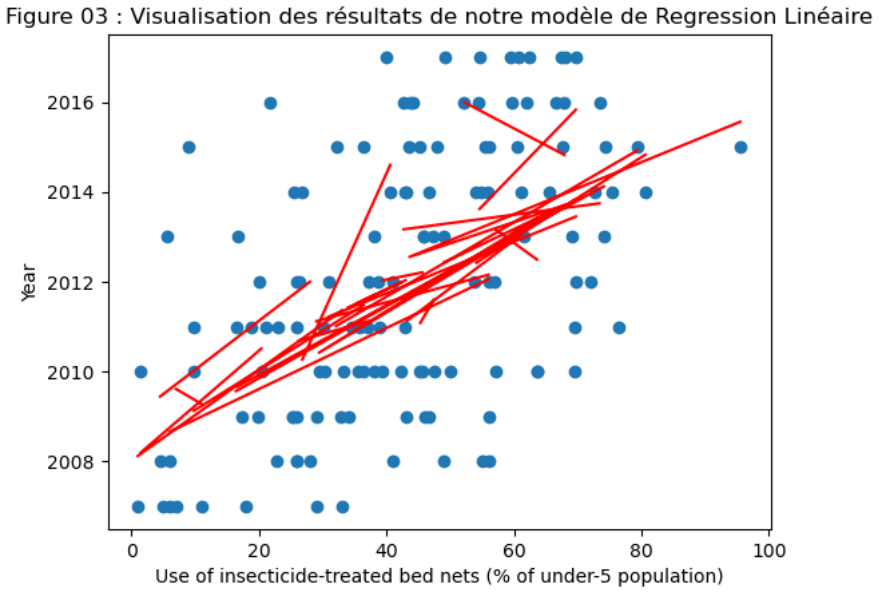
Conclusion RMSE : Le résultat du RMSE est inférieur à 180, donc notre algorithme de Machine Learning fonctionne bien

Machine Learning : Communication des résultats

Visualisation des résultats

In [214]:

```
plt.scatter(X['Use of insecticide-treated bed nets (% of under-5 population)'], y)
plt.plot(X['Use of insecticide-treated bed nets (% of under-5 population)'], predictions, color='red')
plt.xlabel('Use of insecticide-treated bed nets (% of under-5 population)')
plt.ylabel('Year')
plt.title("Figure 03 : Visualisation des résultats de notre modèle de Regression Linéaire")
plt.show()
```



Lecture de la Figure 03 : le taux d'enfants de moins de 5 ans dormant sous moustiquaires imprégnés, continuera à augmenter d'année en année afin de réduire le nombre de cas de paludisme.

In [ ]: