Malaria

November 29, 2024

[1]:
```
import os
cwd=os.getcwd()
print(cwd)
```

```
C:\Users\user
```

[3]:
```
import pandas as pd
import matplotlib.pyplot as plt
```

Load the dataset into a Pandas DataFrame.

[5]:
```
df=pd.read_csv("malaria_indicators_ken (1).csv")
print(df)
```

```
                GHO (CODE)                                     GHO (DISPLAY)  \
0        MALARIA_MICR_TEST  Number of malaria suspects examined by microscopy
1    MALARIA_EST_MORTALITY  Estimated malaria mortality rate (per 100 000 …
2      MALARIA_TOTAL_CASES  Total number of  malaria cases (presumed + con…
3          MALARIA_PF_INDIG   Number of indigenous P. falciparum malaria cases
4          MALARIA_PF_INDIG   Number of indigenous P. falciparum malaria cases
..                     …                                                 …
177      MALARIA_MICR_TEST  Number of malaria suspects examined by microscopy
178          MALARIA_INDIG                Number of indigenous malaria cases
179       MALARIA_SUSPECTS                 Number of suspected malaria cases
180          MALARIA_INDIG                Number of indigenous malaria cases
181        MALARIA_RDT_POS  Number of malaria positive cases by rapid diag…

                                              GHO (URL)  YEAR (DISPLAY)  \
0    https://www.who.int/data/gho/data/indicators/i…            2013
1    https://www.who.int/data/gho/data/indicators/i…            2012
2    https://www.who.int/data/gho/data/indicators/i…            2014
3    https://www.who.int/data/gho/data/indicators/i…            2019
4    https://www.who.int/data/gho/data/indicators/i…            2010
..                                                …              …
177  https://www.who.int/data/gho/data/indicators/i…            2012
178  https://www.who.int/data/gho/data/indicators/i…            2011
179  https://www.who.int/data/gho/data/indicators/i…            2021
180  https://www.who.int/data/gho/data/indicators/i…            2012
181  https://www.who.int/data/gho/data/indicators/i…            2018
```

```
     STARTYEAR  ENDYEAR REGION (CODE) REGION (DISPLAY) COUNTRY (CODE)  \
0         2013     2013           AFR           Africa            KEN
1         2012     2012           AFR           Africa            KEN
2         2014     2014           AFR           Africa            KEN
3         2019     2019           AFR           Africa            KEN
4         2010     2010           AFR           Africa            KEN
..         …        …             …                …              …
177       2012     2012           AFR           Africa            KEN
178       2011     2011           AFR           Africa            KEN
179       2021     2021           AFR           Africa            KEN
180       2012     2012           AFR           Africa            KEN
181       2018     2018           AFR           Africa            KEN

    COUNTRY (DISPLAY)  DIMENSION (TYPE)  DIMENSION (CODE)  DIMENSION (NAME)  \
0               Kenya               NaN               NaN               NaN
1               Kenya               NaN               NaN               NaN
2               Kenya               NaN               NaN               NaN
3               Kenya               NaN               NaN               NaN
4               Kenya               NaN               NaN               NaN
..                …                 …                 …                 …
177             Kenya               NaN               NaN               NaN
178             Kenya               NaN               NaN               NaN
179             Kenya               NaN               NaN               NaN
180             Kenya               NaN               NaN               NaN
181             Kenya               NaN               NaN               NaN

       Population                  Value       Low       High      County
0    6.606885e+06              6 606 885       NaN        NaN     Nairobi
1    2.295441e+01  22.95 [22.22-237.80]  22.22029  237.80007     Mombasa
2    9.698529e+06              9 698 529       NaN        NaN      Kisumu
3    5.019389e+06              5 019 389       NaN        NaN      Nakuru
4    8.985310e+05                898 531       NaN        NaN     Eldoret
..            …                     …       …          …         …
177  4.836617e+06              4 836 617       NaN        NaN     Turkana
178  1.002805e+06              1 002 805       NaN        NaN  West Pokot
179  1.374802e+07             13 748 015       NaN        NaN     Samburu
180  1.453471e+06              1 453 471       NaN        NaN    Laikipia
181  1.490143e+06              1 490 143       NaN        NaN   Nyandarua

[182 rows x 18 columns]
```

Checking for and handling missing values (drop or fill)

[9]:
```
df.isnull().sum()
```

[9]:
```
GHO (CODE)              0
GHO (DISPLAY)           0
```

```
GHO (URL)             0
YEAR (DISPLAY)        0
STARTYEAR             0
ENDYEAR               0
REGION (CODE)         0
REGION (DISPLAY)      0
COUNTRY (CODE)        0
COUNTRY (DISPLAY)     0
DIMENSION (TYPE)    182
DIMENSION (CODE)    182
DIMENSION (NAME)    182
Population            0
Value                 0
Low                 137
High                137
County                0
dtype: int64
```

Drop the three columns with title DIMENSION

Identify the columns

[13]:
```python
columns_to_drop=df.isnull().sum().nlargest(3).index
```

drop the columns

[16]:
```python
df=df.drop(columns=columns_to_drop)
print(df)
```

```
                GHO (CODE)                                    GHO (DISPLAY)  \
0        MALARIA_MICR_TEST  Number of malaria suspects examined by microscopy
1    MALARIA_EST_MORTALITY  Estimated malaria mortality rate (per 100 000 …
2      MALARIA_TOTAL_CASES  Total number of  malaria cases (presumed + con…
3         MALARIA_PF_INDIG   Number of indigenous P. falciparum malaria cases
4         MALARIA_PF_INDIG   Number of indigenous P. falciparum malaria cases
..                     …                                                  …
177      MALARIA_MICR_TEST  Number of malaria suspects examined by microscopy
178          MALARIA_INDIG                Number of indigenous malaria cases
179       MALARIA_SUSPECTS                 Number of suspected malaria cases
180          MALARIA_INDIG                Number of indigenous malaria cases
181        MALARIA_RDT_POS  Number of malaria positive cases by rapid diag…

                                             GHO (URL)  YEAR (DISPLAY)  \
0    https://www.who.int/data/gho/data/indicators/i…            2013
1    https://www.who.int/data/gho/data/indicators/i…            2012
2    https://www.who.int/data/gho/data/indicators/i…            2014
3    https://www.who.int/data/gho/data/indicators/i…            2019
4    https://www.who.int/data/gho/data/indicators/i…            2010
..                                                 …              …
```

```
177  https://www.who.int/data/gho/data/indicators/i…           2012
178  https://www.who.int/data/gho/data/indicators/i…           2011
179  https://www.who.int/data/gho/data/indicators/i…           2021
180  https://www.who.int/data/gho/data/indicators/i…           2012
181  https://www.who.int/data/gho/data/indicators/i…           2018

     STARTYEAR  ENDYEAR REGION (CODE) REGION (DISPLAY) COUNTRY (CODE)  \
0         2013     2013           AFR           Africa            KEN
1         2012     2012           AFR           Africa            KEN
2         2014     2014           AFR           Africa            KEN
3         2019     2019           AFR           Africa            KEN
4         2010     2010           AFR           Africa            KEN
..         …        …             …                …              …
177       2012     2012           AFR           Africa            KEN
178       2011     2011           AFR           Africa            KEN
179       2021     2021           AFR           Africa            KEN
180       2012     2012           AFR           Africa            KEN
181       2018     2018           AFR           Africa            KEN

    COUNTRY (DISPLAY)    Population                 Value       Low  \
0               Kenya  6.606885e+06             6 606 885       NaN
1               Kenya  2.295441e+01  22.95 [22.22-237.80]  22.22029
2               Kenya  9.698529e+06             9 698 529       NaN
3               Kenya  5.019389e+06             5 019 389       NaN
4               Kenya  8.985310e+05               898 531       NaN
..                …             …                     …       …
177             Kenya  4.836617e+06             4 836 617       NaN
178             Kenya  1.002805e+06             1 002 805       NaN
179             Kenya  1.374802e+07            13 748 015       NaN
180             Kenya  1.453471e+06             1 453 471       NaN
181             Kenya  1.490143e+06             1 490 143       NaN

          High      County
0          NaN     Nairobi
1    237.80007     Mombasa
2          NaN      Kisumu
3          NaN      Nakuru
4          NaN     Eldoret
..         …          …
177        NaN     Turkana
178        NaN  West Pokot
179        NaN     Samburu
180        NaN    Laikipia
181        NaN   Nyandarua

[182 rows x 15 columns]
```

Fill missing values

specify the columns to fill

[20]:
```python
columns_to_fill=['Low','High']
```

[22]:
```python
df[columns_to_fill]=df[columns_to_fill].fillna(0)
print(df)
```

```
                GHO (CODE)                                      GHO (DISPLAY)  \
0         MALARIA_MICR_TEST  Number of malaria suspects examined by microscopy
1     MALARIA_EST_MORTALITY  Estimated malaria mortality rate (per 100 000 …
2       MALARIA_TOTAL_CASES  Total number of  malaria cases (presumed + con…
3          MALARIA_PF_INDIG   Number of indigenous P. falciparum malaria cases
4          MALARIA_PF_INDIG   Number of indigenous P. falciparum malaria cases
..                      …                                                  …
177       MALARIA_MICR_TEST  Number of malaria suspects examined by microscopy
178           MALARIA_INDIG                 Number of indigenous malaria cases
179        MALARIA_SUSPECTS                  Number of suspected malaria cases
180           MALARIA_INDIG                 Number of indigenous malaria cases
181         MALARIA_RDT_POS  Number of malaria positive cases by rapid diag…

                                              GHO (URL)  YEAR (DISPLAY)  \
0    https://www.who.int/data/gho/data/indicators/i…            2013
1    https://www.who.int/data/gho/data/indicators/i…            2012
2    https://www.who.int/data/gho/data/indicators/i…            2014
3    https://www.who.int/data/gho/data/indicators/i…            2019
4    https://www.who.int/data/gho/data/indicators/i…            2010
..                                                 …             …
177  https://www.who.int/data/gho/data/indicators/i…            2012
178  https://www.who.int/data/gho/data/indicators/i…            2011
179  https://www.who.int/data/gho/data/indicators/i…            2021
180  https://www.who.int/data/gho/data/indicators/i…            2012
181  https://www.who.int/data/gho/data/indicators/i…            2018

     STARTYEAR  ENDYEAR REGION (CODE) REGION (DISPLAY) COUNTRY (CODE)  \
0         2013     2013           AFR           Africa            KEN
1         2012     2012           AFR           Africa            KEN
2         2014     2014           AFR           Africa            KEN
3         2019     2019           AFR           Africa            KEN
4         2010     2010           AFR           Africa            KEN
..           …        …             …                …              …
177       2012     2012           AFR           Africa            KEN
178       2011     2011           AFR           Africa            KEN
179       2021     2021           AFR           Africa            KEN
180       2012     2012           AFR           Africa            KEN
181       2018     2018           AFR           Africa            KEN

    COUNTRY (DISPLAY)    Population                  Value        Low  \
0               Kenya  6.606885e+06              6 606 885    0.00000
```

```
1              Kenya  2.295441e+01  22.95 [22.22-237.80]  22.22029
2              Kenya  9.698529e+06            9 698 529   0.00000
3              Kenya  5.019389e+06            5 019 389   0.00000
4              Kenya  8.985310e+05              898 531   0.00000
..               …          …                     …       …
177            Kenya  4.836617e+06            4 836 617   0.00000
178            Kenya  1.002805e+06            1 002 805   0.00000
179            Kenya  1.374802e+07           13 748 015   0.00000
180            Kenya  1.453471e+06            1 453 471   0.00000
181            Kenya  1.490143e+06            1 490 143   0.00000

          High      County
0      0.00000     Nairobi
1    237.80007     Mombasa
2      0.00000      Kisumu
3      0.00000      Nakuru
4      0.00000     Eldoret
..         …          …
177    0.00000     Turkana
178    0.00000  West Pokot
179    0.00000     Samburu
180    0.00000    Laikipia
181    0.00000   Nyandarua

[182 rows x 15 columns]
```

check data types

[25]:
```
print(df.dtypes)
```

```
GHO (CODE)             object
GHO (DISPLAY)          object
GHO (URL)              object
YEAR (DISPLAY)          int64
STARTYEAR               int64
ENDYEAR                 int64
REGION (CODE)          object
REGION (DISPLAY)       object
COUNTRY (CODE)         object
COUNTRY (DISPLAY)      object
Population            float64
Value                  object
Low                   float64
High                  float64
County                 object
dtype: object
```

Converting columns to appropriate datatypes

We don't have any column's data types that need to be converted. NB: Column 'Value'is considered an object in Pandas since it is stored as text because its values have spaces which are not typically part of numeric data.

Removing any duplicate entries

[30]:
```
print(df.drop_duplicates())
```

```
                GHO (CODE)                                       GHO (DISPLAY)  \
0        MALARIA_MICR_TEST  Number of malaria suspects examined by microscopy
1    MALARIA_EST_MORTALITY  Estimated malaria mortality rate (per 100 000 …
2      MALARIA_TOTAL_CASES  Total number of  malaria cases (presumed + con…
3         MALARIA_PF_INDIG   Number of indigenous P. falciparum malaria cases
4         MALARIA_PF_INDIG   Number of indigenous P. falciparum malaria cases
..                     …                                                  …
177      MALARIA_MICR_TEST  Number of malaria suspects examined by microscopy
178          MALARIA_INDIG                  Number of indigenous malaria cases
179       MALARIA_SUSPECTS                   Number of suspected malaria cases
180          MALARIA_INDIG                  Number of indigenous malaria cases
181        MALARIA_RDT_POS  Number of malaria positive cases by rapid diag…

                                             GHO (URL)  YEAR (DISPLAY)  \
0    https://www.who.int/data/gho/data/indicators/i…            2013
1    https://www.who.int/data/gho/data/indicators/i…            2012
2    https://www.who.int/data/gho/data/indicators/i…            2014
3    https://www.who.int/data/gho/data/indicators/i…            2019
4    https://www.who.int/data/gho/data/indicators/i…            2010
..                                                 …             …
177  https://www.who.int/data/gho/data/indicators/i…            2012
178  https://www.who.int/data/gho/data/indicators/i…            2011
179  https://www.who.int/data/gho/data/indicators/i…            2021
180  https://www.who.int/data/gho/data/indicators/i…            2012
181  https://www.who.int/data/gho/data/indicators/i…            2018

     STARTYEAR  ENDYEAR REGION (CODE) REGION (DISPLAY) COUNTRY (CODE)  \
0         2013     2013           AFR           Africa            KEN
1         2012     2012           AFR           Africa            KEN
2         2014     2014           AFR           Africa            KEN
3         2019     2019           AFR           Africa            KEN
4         2010     2010           AFR           Africa            KEN
..         …        …           …              …              …
177       2012     2012           AFR           Africa            KEN
178       2011     2011           AFR           Africa            KEN
179       2021     2021           AFR           Africa            KEN
180       2012     2012           AFR           Africa            KEN
181       2018     2018           AFR           Africa            KEN

    COUNTRY (DISPLAY)    Population                 Value       Low  \
0               Kenya  6.606885e+06             6 606 885   0.00000
1               Kenya  2.295441e+01  22.95 [22.22-237.80]  22.22029
```

```
2              Kenya  9.698529e+06           9 698 529   0.00000
3              Kenya  5.019389e+06           5 019 389   0.00000
4              Kenya  8.985310e+05             898 531   0.00000
..               …           …                  …       …
177            Kenya  4.836617e+06           4 836 617   0.00000
178            Kenya  1.002805e+06           1 002 805   0.00000
179            Kenya  1.374802e+07          13 748 015   0.00000
180            Kenya  1.453471e+06           1 453 471   0.00000
181            Kenya  1.490143e+06           1 490 143   0.00000

          High      County
0      0.00000     Nairobi
1    237.80007     Mombasa
2      0.00000      Kisumu
3      0.00000      Nakuru
4      0.00000     Eldoret
..         …         …
177    0.00000     Turkana
178    0.00000  West Pokot
179    0.00000     Samburu
180    0.00000    Laikipia
181    0.00000   Nyandarua

[182 rows x 15 columns]
```

Print first five rows of the cleaned dataframe

[33]:
```python
print(df.head())
```

```
              GHO (CODE)                                    GHO (DISPLAY)  \
0      MALARIA_MICR_TEST  Number of malaria suspects examined by microscopy
1  MALARIA_EST_MORTALITY  Estimated malaria mortality rate (per 100 000 …
2    MALARIA_TOTAL_CASES  Total number of  malaria cases (presumed + con…
3        MALARIA_PF_INDIG   Number of indigenous P. falciparum malaria cases
4        MALARIA_PF_INDIG   Number of indigenous P. falciparum malaria cases

                                         GHO (URL)  YEAR (DISPLAY)  \
0  https://www.who.int/data/gho/data/indicators/i…            2013
1  https://www.who.int/data/gho/data/indicators/i…            2012
2  https://www.who.int/data/gho/data/indicators/i…            2014
3  https://www.who.int/data/gho/data/indicators/i…            2019
4  https://www.who.int/data/gho/data/indicators/i…            2010

   STARTYEAR  ENDYEAR REGION (CODE) REGION (DISPLAY) COUNTRY (CODE)  \
0       2013     2013           AFR           Africa            KEN
1       2012     2012           AFR           Africa            KEN
2       2014     2014           AFR           Africa            KEN
3       2019     2019           AFR           Africa            KEN
4       2010     2010           AFR           Africa            KEN
```

```
  COUNTRY (DISPLAY)    Population                   Value       Low        High  \
0             Kenya  6.606885e+06               6 606 885   0.00000     0.00000
1             Kenya  2.295441e+01   22.95 [22.22-237.80]   22.22029   237.80007
2             Kenya  9.698529e+06               9 698 529   0.00000     0.00000
3             Kenya  5.019389e+06               5 019 389   0.00000     0.00000
4             Kenya  8.985310e+05                 898 531   0.00000     0.00000

    County
0  Nairobi
1  Mombasa
2   Kisumu
3   Nakuru
4  Eldoret
```

The dataset contains

Calculate and display the mean, median, and standard deviation for numerical columns.

[37]:
```python
print("Population:",df['Population'].mean())
print("Low:",df['Low'].mean())
print("High:",df['High'].mean())
```

```
Population: 3828362.582783132
Low: 13.067958241758243
High: 51.29541181318682
```

[39]:
```python
print("Population:",df['Population'].median())
print("Low:",df['Low'].median())
print("High:",df['High'].median())
```

```
Population: 2400176.0
Low: 0.0
High: 0.0
```

[41]:
```python
print("Population:",df['Population'].std())
print("Low:",df['Low'].std())
print("High:",df['High'].std())
```

```
Population: 4180841.6279189037
Low: 32.58929040604728
High: 102.98504326048062
```

Identify any correlations between different numerical variables

[44]:
```python
numerical_df=df.select_dtypes(include=['number'])
print(numerical_df)
```

```
     YEAR (DISPLAY)  STARTYEAR  ENDYEAR    Population       Low       High
0              2013       2013     2013  6.606885e+06   0.00000    0.00000
1              2012       2012     2012  2.295441e+01  22.22029  237.80007
```

```
2               2014        2014      2014  9.698529e+06   0.00000    0.00000
3               2019        2019      2019  5.019389e+06   0.00000    0.00000
4               2010        2010      2010  8.985310e+05   0.00000    0.00000
..               …           …      …             …         …          …
177             2012        2012      2012  4.836617e+06   0.00000    0.00000
178             2011        2011      2011  1.002805e+06   0.00000    0.00000
179             2021        2021      2021  1.374802e+07   0.00000    0.00000
180             2012        2012      2012  1.453471e+06   0.00000    0.00000
181             2018        2018      2018  1.490143e+06   0.00000    0.00000

[182 rows x 6 columns]
```

[46]:
```
correlation=numerical_df.corr()
print(correlation)
```

```
                YEAR (DISPLAY)  STARTYEAR   ENDYEAR  Population       Low  \
YEAR (DISPLAY)        1.000000   1.000000  1.000000    0.117370 -0.282631
STARTYEAR             1.000000   1.000000  1.000000    0.117370 -0.282631
ENDYEAR               1.000000   1.000000  1.000000    0.117370 -0.282631
Population            0.117370   0.117370  0.117370    1.000000 -0.369200
Low                  -0.282631  -0.282631 -0.282631   -0.369200  1.000000
High                 -0.258575  -0.258575 -0.258575   -0.458605  0.636598

                    High
YEAR (DISPLAY) -0.258575
STARTYEAR      -0.258575
ENDYEAR        -0.258575
Population     -0.458605
Low             0.636598
High            1.000000
```

Group the data by a categorical variable and calculate the total or average for a numerical variable of interest.

[49]:
```
grouped=df.groupby(['County'])[['Low','High']].agg({'Low':'mean','High':
 ↪'mean'})# Calculate average low and high by county
print(grouped)
```

```
                      Low        High
County
Baringo          0.000000    0.000000
Bomet            5.089082   66.096887
Bungoma         16.401968   78.910862
Busia            0.000000    0.000000
Eldoret          5.403168   61.973887
Elgeyo Marakwet 11.942583  129.830687
Embu             0.000000    0.000000
Garissa         35.010740   60.318833
Homa Bay        16.017723   29.460030
```

```
Isiolo           16.702178   78.781477
Kajiado           5.178460   60.876582
Kakamega          6.074382   64.720232
Kericho           0.000000    0.000000
Kiambu            9.505828  103.539300
Kilifi            0.000000    0.000000
Kirinyaga         0.000000    0.000000
Kisii            61.548810   95.547667
Kisumu            5.423900   58.324667
Kitui             0.000000    0.000000
Kwale             5.868032   62.342267
Laikipia         19.344883   34.044272
Lamu             10.548728   20.381363
Machakos         41.622820  120.262533
Mandera          11.711575   20.969887
Marsabit         11.874883   21.640082
Meru             30.494475  112.964620
Migori           15.674338  169.380683
Mombasa           5.555072   59.450018
Murang'a         12.517883   23.005557
Nairobi          10.407758   23.318395
Nakuru            0.000000    0.000000
Nandi            19.148973   33.723757
Narok             0.000000    0.000000
Nyamira          57.615922  160.631375
Nyandarua        12.532585   23.848960
Nyeri             0.000000    0.000000
Samburu          44.170147   69.523400
Siaya            11.831907  128.118667
Taita Taveta      0.000000    0.000000
Tana River       10.945786   54.860723
Tharaka Nithi     5.271922   61.182140
Trans Nzoia      15.440515   88.499465
Turkana          18.240635   84.925472
Uasin Gishu       0.000000    0.000000
Vihiga            0.000000    0.000000
Wajir            23.199580  108.094873
West Pokot       42.740600   69.519700
```

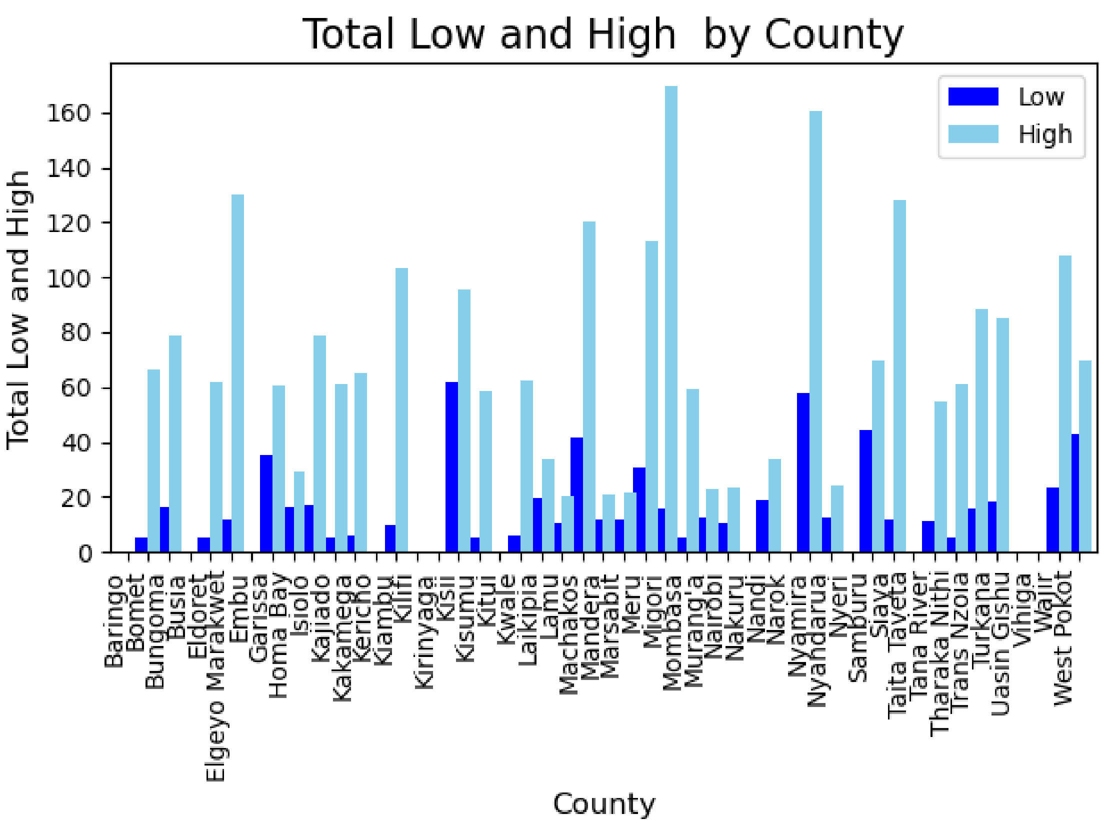
Present the results of the analysis in a clear format (tables or charts)

[52]:
```python
import matplotlib.pyplot as plt

# Create a bar chart and customize.
plt.figure(figsize=(12, 6))
grouped.plot(kind='bar', color=['blue','skyblue'], width=1.2)
plt.title("Total Low and High  by County", fontsize=16)
plt.xlabel("County", fontsize=12)
```

```
plt.ylabel("Total Low and High", fontsize=12)
plt.xticks(rotation=90, fontsize=10, ha='right')# Rotate x-axis labels for better visibility
plt.tight_layout()
plt.show()
```

```
<Figure size 1200x600 with 0 Axes>
```



A bar chart showing the total population distribution across different regions in Kenya.

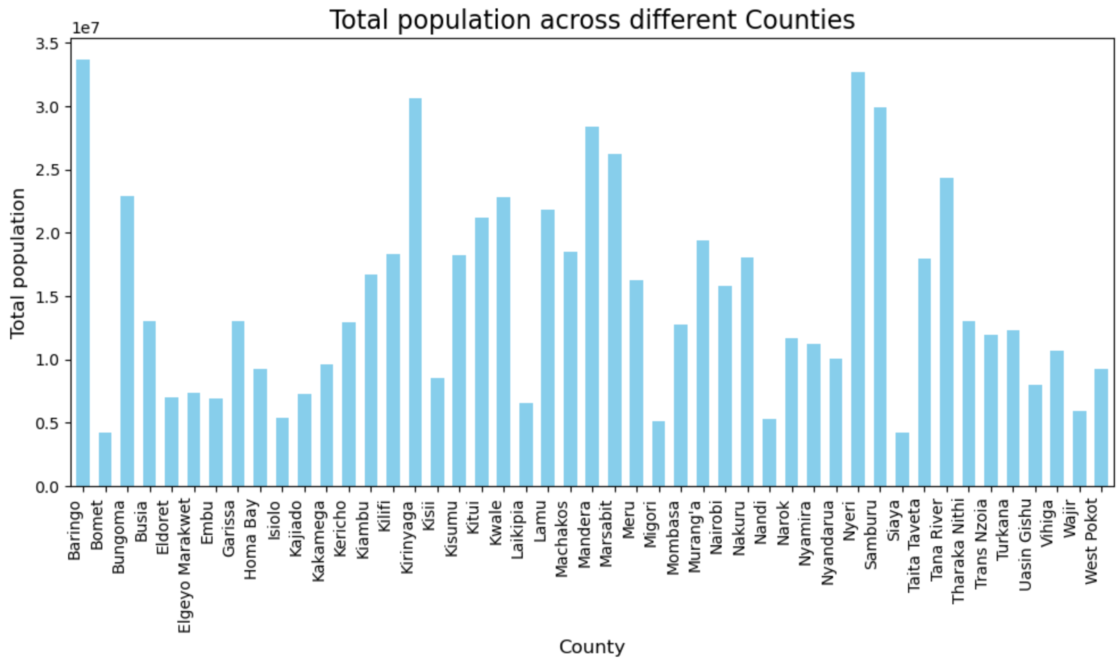
```
[54]: population_across_regions=df.groupby('County')['Population'].sum()# Calculate total population by county
print(population_across_regions)

# Create a bar chart and customize.
plt.figure(figsize=(10, 6))
population_across_regions.plot(kind='bar', color='skyblue', width=0.6)
plt.title("Total population across different Counties", fontsize=16)
plt.xlabel("County", fontsize=12)
plt.ylabel("Total population", fontsize=12)
```

```
plt.xticks(rotation=90, fontsize=10, ha='right')# Rotate x-axis labels for␣
 ↪better visibility
plt.tight_layout()
plt.show()
```

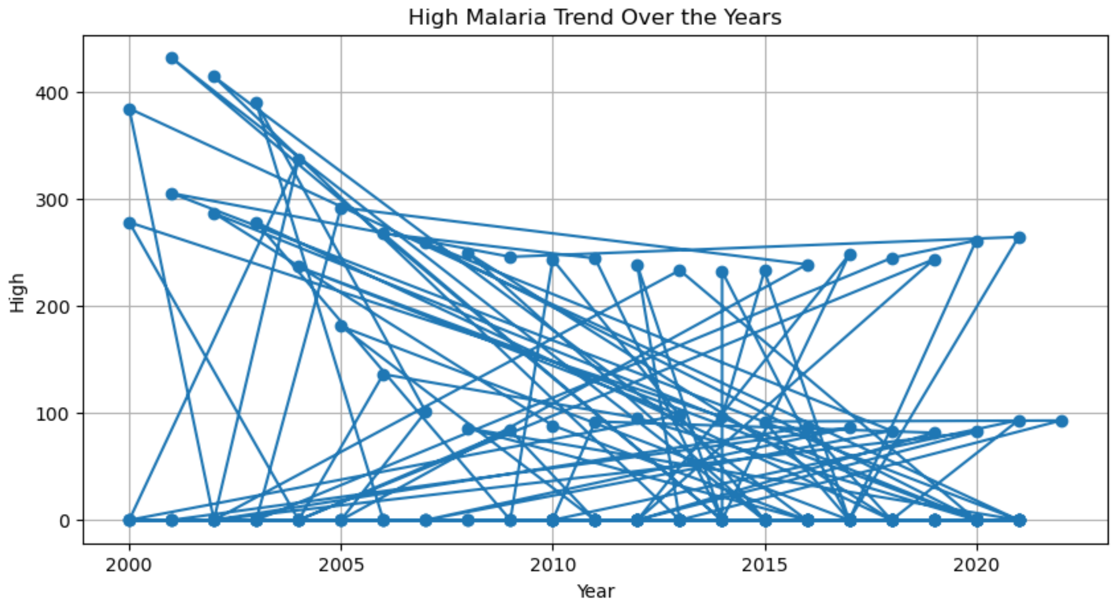
```
County
Baringo            3.368058e+07
Bomet              4.211391e+06
Bungoma            2.286617e+07
Busia              1.306303e+07
Eldoret            6.973674e+06
Elgeyo Marakwet    7.359270e+06
Embu               6.905327e+06
Garissa            1.301012e+07
Homa Bay           9.235986e+06
Isiolo             5.384354e+06
Kajiado            7.313755e+06
Kakamega           9.601507e+06
Kericho            1.295566e+07
Kiambu             1.667882e+07
Kilifi             1.834866e+07
Kirinyaga          3.059670e+07
Kisii              8.529650e+06
Kisumu             1.828047e+07
Kitui              2.124326e+07
Kwale              2.285784e+07
Laikipia           6.576389e+06
Lamu               2.181305e+07
Machakos           1.846720e+07
Mandera            2.841602e+07
Marsabit           2.625882e+07
Meru               1.629045e+07
Migori             5.115127e+06
Mombasa            1.275555e+07
Murang'a           1.936345e+07
Nairobi            1.584817e+07
Nakuru             1.805708e+07
Nandi              5.282305e+06
Narok              1.170378e+07
Nyamira            1.119762e+07
Nyandarua          1.005047e+07
Nyeri              3.272030e+07
Samburu            2.994085e+07
Siaya              4.228224e+06
Taita Taveta       1.801536e+07
Tana River         2.435087e+07
Tharaka Nithi      1.305088e+07
Trans Nzoia        1.198591e+07
```

```
Turkana             1.228158e+07
Uasin Gishu         8.016301e+06
Vihiga              1.070398e+07
Wajir               5.952440e+06
West Pokot          9.223608e+06
Name: Population, dtype: float64
```



A line graph showing the trend of a particular variable over the years.

[57]:
```
# Create a line plot
plt.figure(figsize=(10, 5))
plt.plot(df["YEAR (DISPLAY)"], df["High"], marker='o', linestyle='-')
plt.title("High Malaria Trend Over the Years")
plt.xlabel("Year")
plt.ylabel("High")
plt.grid()
plt.show()
```



Malaria Dataset Analysis Process :

This project analyzes malaria distribution in Kenya using a dataset that includes information about malaria patients population and regional(counties) data over multiple years. The analysis is conducted using Python with libraries such as Pandas, Scipy and Matplotlib. Instructions To run the analysis:

1. Open the Jupyter Notebook `Malaria.ipynb`.
2. Run each cell sequentially to perform data cleaning, analysis, and visualization.

Dependencies:

- pandas
- matplotlib

1. Data Cleaning

1.1 Loading the Dataset into a Pandas DataFrame: Begin by loading the malaria_indicators_ken (1) into a Pandas DataFrame for further analysis using data = pd.read_csv('malaria_indicators_ken (1)').

1.2 Handling Missing Values: Check for any missing values in the dataset and decided to either drop or fill them based on the context of the data. In this process I did both. Dropped three columns with title DIMENSION that are DIMENSION (TYPE), DIMENSION (CODE)and DIMENSION (NAM because the whole columns has no values.E) Filled two columns that is Low and High with 0 where the value was null

1.3 Converting Columns to Appropriate Data Types: Ensure that the relevant columns are converted to the correct data types for proper analysis. In my case the columns are in their appropriate data types.NB: Column ’Value’is considered an object in Pandas since it is stored as text because its values have spaces which are not typically part of numeric data.

1.4 Removing Duplicate Entries: Removed any duplicate entries to ensure the dataset’s integrity

1.5 Output: Displayed the first five rows of the cleaned dataset and summarized the dataset using print(df.head()) and print(data.describe()) for summary.

2. Data Analysis

2.1 Descriptive Statistics: Calculated and displayed the mean, median, and standard deviation for numerical columns to understand the distribution of data. The numerical values are Population,Low and High variables.

2.2 Correlation Analysis: Identified correlations between numerical variables mentioned above to understand relationships.

2.3 Data Grouping: Grouped the data by a categorical variable County and numerical variables Low and High then calculated averages(mean).

2.4 Output: The results are presented clearly using a bar chart plot using Matplotlib.

3. Data Visualization

3.1 Visualization Libraries: I used Matplotlib to create a bar chart and a line graph to display trends.

3.2 Charts created:

3.2.1 Bar Chart: Total Population Distribution Across Counties.

3.2.2 Line Graph: Trend of High Malaria Cases Over the Years.

[ ]:

```
```