model-monitor-project

May 23, 2024

[1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score
from sklearn.linear_model import Ridge
from sklearn.linear_model import Lasso
from sklearn.impute import KNNImputer
from sklearn.preprocessing import MinMaxScaler
from sklearn.model_selection import train_test_split, cross_val_score
```

0.0.1 MLRCARE-Careseeking for children under age 5 with fever-Percentage of children (under age 5) with fever for whom advice or treatment was sought

[2]:
```
pd.set_option("display.max_columns", None)
mlrcare = pd.read_excel("Final_malaria.xlsx", sheet_name = 0)
mlrcare
```

[2]:
```
     ISO     Countries          UNICEF Reporting region  \
0    AFG   Afghanistan                       South Asia
1    AFG   Afghanistan                       South Asia
2    AGO        Angola      Eastern and Southern Africa
3    ALB       Albania  Eastern Europe and Central Asia
4    ALB       Albania  Eastern Europe and Central Asia
..   …             …                              …
202  ZAF  South Africa      Eastern and Southern Africa
203  ZMB        Zambia      Eastern and Southern Africa
204  ZMB        Zambia      Eastern and Southern Africa
205  ZWE      Zimbabwe      Eastern and Southern Africa
206  ZWE      Zimbabwe      Eastern and Southern Africa

    UNICEF Programme Region World Bank Income Group (2022)  Year  \
0                      ROSA                     Low income  2015
1                      ROSA                     Low income  2018
2                     ESARO            Lower middle income  2016
```

```
3                      ECARO           Upper middle income  2009
4                      ECARO           Upper middle income  2018
..                       …                           …    …
202                    ESARO           Upper middle income  2016
203                    ESARO                    Low income  2014
204                    ESARO                    Low income  2019
205                    ESARO           Lower middle income  2014
206                    ESARO           Lower middle income  2015

                      Short Source  \
0                         DHS 2015
1    Afghanistan Health Survey 2018
2                    DHS 2015-2016
3                    DHS 2008-2009
4                    DHS 2017-2018
..                             …
202                       DHS 2016
203                  DHS 2013-2014
204                  DHS 2018-2019
205                      MICS 2014
206                       DHS 2015

                                          Long Source  National  Male  \
0                    Demographic and Health Survey 2015      63.2   NaN
1    Afghanistan Health Survey 2018. Amsterdam: KIT…      62.1  60.9
2               Demographic and Health Survey 2015-2016      50.8  51.0
3               Demographic and Health Survey 2008-2009      71.2  80.9
4               Demographic and Health Survey 2017-2018      59.6  60.8
..                                                 …        …    …
202                  Demographic and Health Survey 2016      68.4  69.0
203             Demographic and Health Survey 2013-2014      74.9  75.9
204             Demographic and Health Survey 2018-2019      77.2  79.4
205              Multiple Indicator Cluster Survey 2014      47.1  46.6
206                  Demographic and Health Survey 2015      49.7  45.9

     Female  Rural  Urban  Poorest  Second  Middle  Fourth  Richest  None  \
0       NaN   62.3   65.9     61.1    62.0    61.1    67.7     64.1   NaN
1      63.2   63.1   59.5     56.6    67.2    59.9    63.5     63.6  60.9
2      50.5   42.6   57.1     36.9    47.3    56.7    59.7     63.3   NaN
3      60.7   69.6   73.2      NaN     NaN     NaN     NaN      NaN   NaN
4      58.3   53.6   66.9     44.3    66.3    63.6     NaN      NaN   NaN
..        …      …     …        …       …       …      …        …     …
202    67.7   63.7   70.8     64.0    67.3    68.6    69.3     73.3   NaN
203    73.9   72.9   79.4     70.1    75.7    75.0    79.6     77.1   NaN
204    74.9   77.7   75.9     73.6    79.4    84.2    75.4     75.7  77.5
205    47.5   48.1   43.6     44.6    48.9    48.3    45.3     49.4   NaN
206    52.9   44.4   60.4     48.1    40.7    45.1    51.1     63.7   NaN
```

```
     Primary  Sec & Higher
0        NaN           NaN
1       66.1           NaN
2        NaN           NaN
3        NaN           NaN
4       54.0          63.9
..       …             …
202      NaN           NaN
203      NaN           NaN
204     75.0          80.6
205      NaN           NaN
206      NaN           NaN

[207 rows x 21 columns]
```

[3]:
```
mlrcare.isnull().sum()
```

[3]:
```
ISO                              0
Countries                        0
UNICEF Reporting region          0
UNICEF Programme Region          0
World Bank Income Group (2022)   0
Year                             0
Short Source                     0
Long Source                      0
National                         1
Male                            20
Female                          20
Rural                           16
Urban                           15
Poorest                         27
Second                          29
Middle                          27
Fourth                          27
Richest                         29
None                           161
Primary                        154
Sec & Higher                   156
dtype: int64
```

[4]:
```
mlrcare.head(2)
```

[4]:
```
   ISO    Countries UNICEF Reporting region UNICEF Programme Region  \
0  AFG  Afghanistan              South Asia                    ROSA
1  AFG  Afghanistan              South Asia                    ROSA
```

```
  World Bank Income Group (2022)  Year                    Short Source  \
0                     Low income  2015                        DHS 2015
1                     Low income  2018  Afghanistan Health Survey 2018

                                        Long Source  National  Male  Female  \
0                Demographic and Health Survey 2015      63.2   NaN     NaN
1  Afghanistan Health Survey 2018. Amsterdam: KIT…      62.1  60.9    63.2

   Rural  Urban  Poorest  Second  Middle  Fourth  Richest  None  Primary  \
0   62.3   65.9     61.1    62.0    61.1    67.7     64.1   NaN      NaN
1   63.1   59.5     56.6    67.2    59.9    63.5     63.6  60.9     66.1

   Sec & Higher
0           NaN
1           NaN
```

Using KNN impute to Handle missing values because each entity has different pattern of values.

An entity is the summary of a country in a year

[5]:
```python
# Columns to impute
columns_to_impute = ['National', 'Male', 'Female', 'Rural', 'Urban', 'Poorest',␣
 ↪'Second', 'Middle', 'Fourth', 'Richest', 'None', 'Primary', 'Sec & Higher']

#Create a KNN Imputer object
imputer = KNNImputer(n_neighbors=5, weights='uniform')

#Fit and transform the data
imputed_data = imputer.fit_transform(mlrcare[columns_to_impute])

# Convert the imputed data back to a pandas DataFrame
imputed_data = pd.DataFrame(imputed_data, columns=columns_to_impute)

# Replace the original columns with the imputed data
mlrcare[columns_to_impute] = imputed_data
```

[6]:
```python
mlrcare.isna().sum()
```

[6]:
```
ISO                               0
Countries                         0
UNICEF Reporting region           0
UNICEF Programme Region           0
World Bank Income Group (2022)    0
Year                              0
Short Source                      0
Long Source                       0
```

```
National                    0
Male                        0
Female                      0
Rural                       0
Urban                       0
Poorest                     0
Second                      0
Middle                      0
Fourth                      0
Richest                     0
None                        0
Primary                     0
Sec & Higher                0
dtype: int64
```

[7]:
```
# checking for duplicates
mlrcare.duplicated().sum()
```

[7]:
```
0
```

[8]:
```
mlrcare.describe()
```

[8]:
```
              Year    National        Male      Female       Rural  \
count   207.000000  207.000000  207.000000  207.000000  207.000000
mean   2015.169082   61.011981   61.520193   60.568406   59.087440
std       3.641110   14.324669   14.283789   14.639060   15.176285
min    2004.000000   22.800000   23.000000   22.500000   19.000000
25%    2013.000000   51.300000   53.100000   51.000000   48.050000
50%    2015.000000   62.100000   62.000000   62.000000   59.600000
75%    2018.000000   71.200000   72.510000   70.900000   70.750000
max    2022.000000   92.900000   92.800000   94.100000   93.200000

            Urban     Poorest      Second      Middle      Fourth     Richest  \
count  207.000000  207.000000  207.000000  207.000000  207.000000  207.000000
mean    65.422512   54.595266   58.930918   60.757681   64.771014   69.262899
std     12.863030   16.964624   16.148627   15.006293   13.713915   12.484468
min     25.800000    6.800000   18.500000   21.100000   25.000000   34.500000
25%     58.170000   43.900000   48.500000   51.070000   56.770000   61.770000
50%     65.800000   56.600000   60.000000   61.200000   65.700000   70.300000
75%     74.450000   66.750000   70.900000   72.700000   73.700000   77.760000
max     93.100000   96.700000   90.800000   95.400000   92.300000   95.100000

             None     Primary  Sec & Higher
count  207.000000  207.000000    207.000000
mean    56.573430   62.927923     67.693527
std     13.936579   12.974423     10.376065
min     23.700000   32.900000     44.000000
```

```
25%     46.230000   55.100000     60.630000
50%     57.000000   63.940000     68.300000
75%     67.320000   73.860000     75.820000
max     87.900000   88.100000     90.700000
```

[9]:
```
mlrcare.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 207 entries, 0 to 206
Data columns (total 21 columns):
 #   Column                         Non-Null Count  Dtype
---  ------                         --------------  -----
 0   ISO                            207 non-null    object
 1   Countries                      207 non-null    object
 2   UNICEF Reporting region        207 non-null    object
 3   UNICEF Programme Region        207 non-null    object
 4   World Bank Income Group (2022) 207 non-null    object
 5   Year                           207 non-null    int64
 6   Short Source                   207 non-null    object
 7   Long Source                    207 non-null    object
 8   National                       207 non-null    float64
 9   Male                           207 non-null    float64
 10  Female                         207 non-null    float64
 11  Rural                          207 non-null    float64
 12  Urban                          207 non-null    float64
 13  Poorest                        207 non-null    float64
 14  Second                         207 non-null    float64
 15  Middle                         207 non-null    float64
 16  Fourth                         207 non-null    float64
 17  Richest                        207 non-null    float64
 18  None                           207 non-null    float64
 19  Primary                        207 non-null    float64
 20  Sec & Higher                   207 non-null    float64
dtypes: float64(13), int64(1), object(7)
memory usage: 34.1+ KB
```

[10]:
```
mlrcare = mlrcare.drop(columns= ["ISO","UNICEF Reporting region","UNICEF␣
↪Programme Region",
                                 "World Bank Income Group (2022)","Long␣
↪Source","Short Source"], axis=1)
```

[11]:
```
mlrcare.sample(2)
```

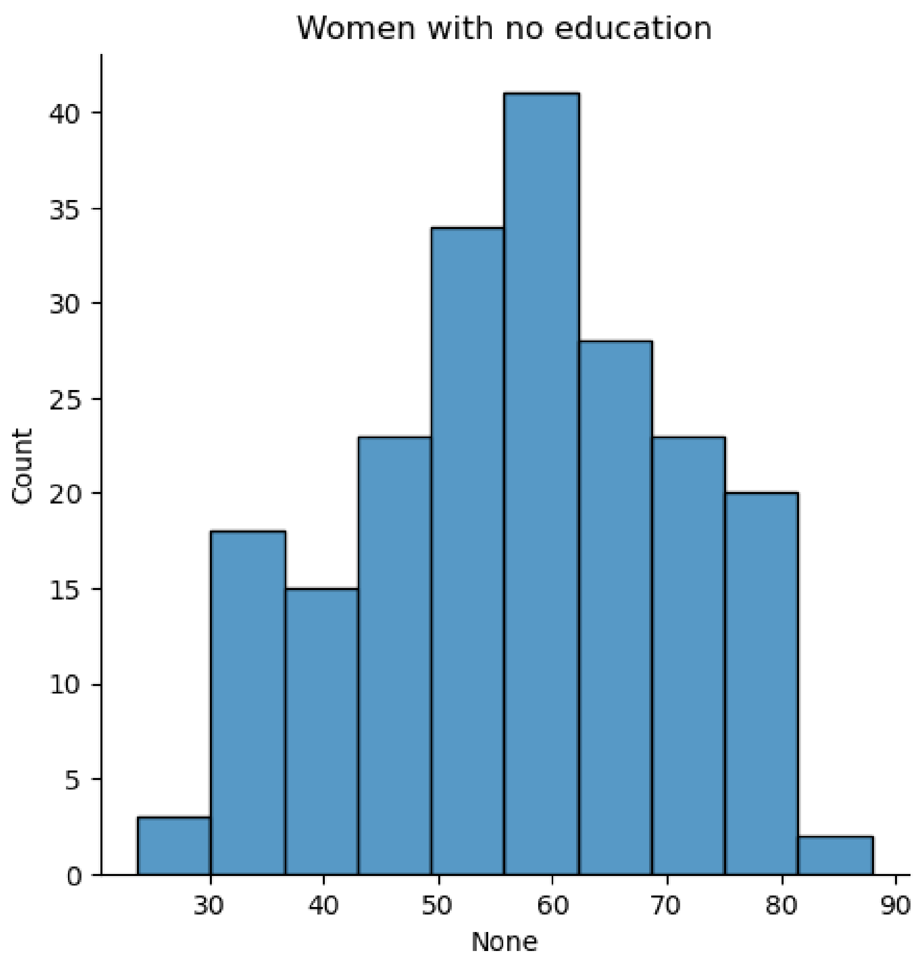
[11]:
```
               Countries  Year  National  Male  Female  Rural  Urban  Poorest  \
41  Dominican Republic  2013      65.1  62.0    68.4   73.3   62.7     71.2
8              Burundi  2010      62.1  62.1    62.1   61.8   66.4     57.1
```

```
     Second  Middle  Fourth  Richest   None  Primary  Sec & Higher
41     64.9    60.3    65.1     60.1  63.38    69.34         67.28
8      63.7    62.7    61.8     66.7  58.64    66.34         67.50
```

[12]:
```
mlrcare["Year"].unique()
```

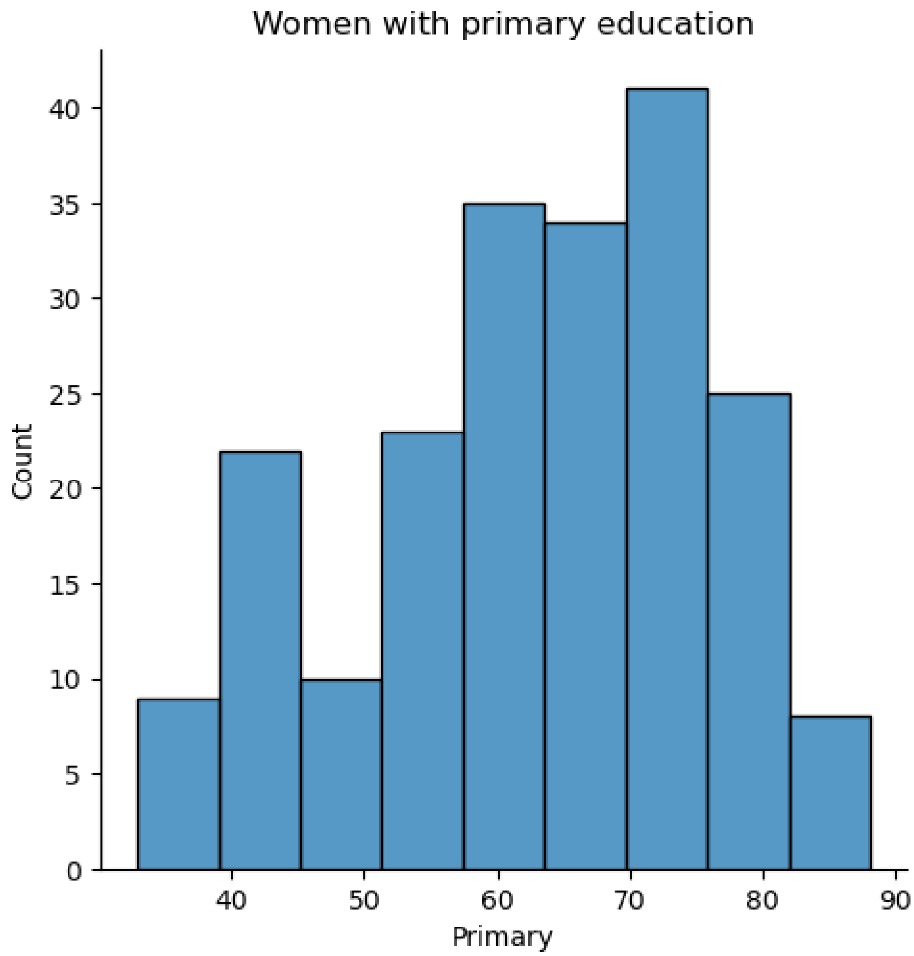
[12]:
```
array([2015, 2018, 2016, 2009, 2020, 2010, 2012, 2017, 2014, 2021, 2011,
       2019, 2007, 2013, 2006, 2022, 2004], dtype=int64)
```

[13]:
```
sns.displot(data=mlrcare,x="None")
plt.title("Women with no education")
plt.show()
```



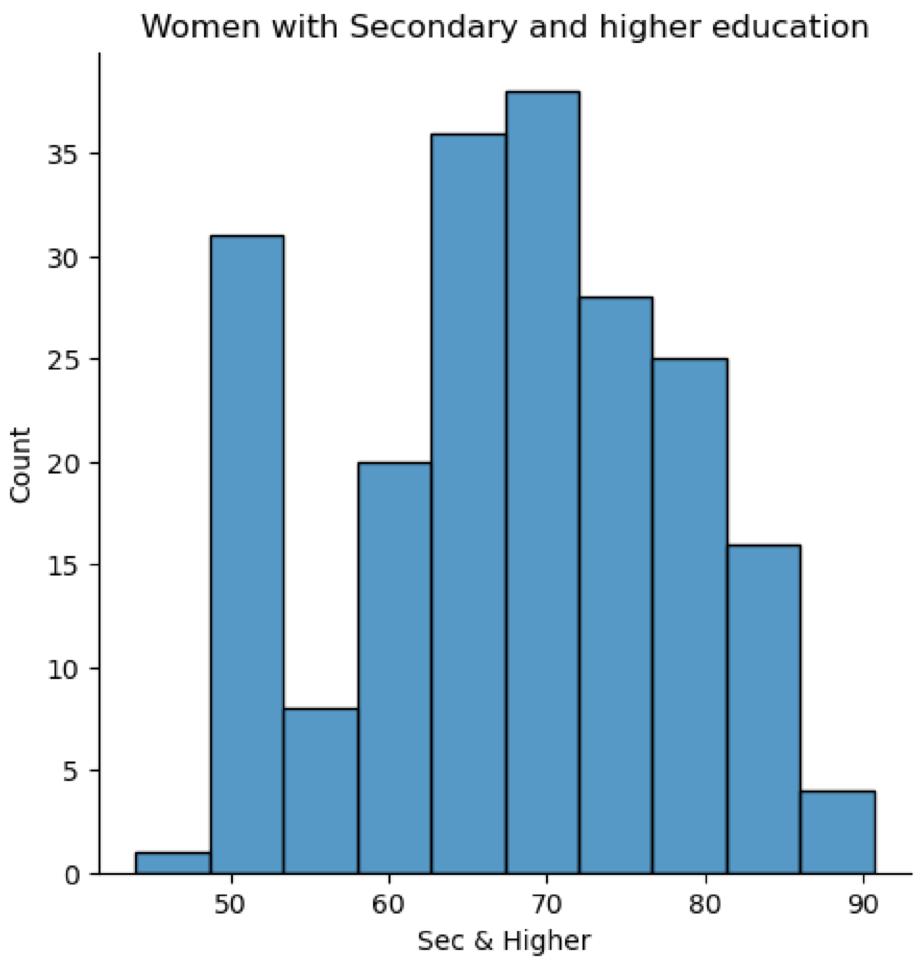
The highest count is seen in the age group of 55-60%, with around 40 Countries. This implies that about 55-60% of children with fever in most countries had their mothers with no education sought advice for their treatment.

[14]:
```
sns.displot(data=mlrcare,x="Primary")
plt.title("Women with primary education")
plt.show()
```



The visual above implies that about 65% of children with fever in most countries had their mothers with primary education sought advice for their treatment.

[15]:
```
sns.displot(data=mlrcare,x="Sec & Higher")
plt.title("Women with Secondary and higher education")
plt.show()
```



The visual above implies that about 70% of children with fever in most countries had their mothers with Secondary or Higher education sought advice for their treatment.

In conclusion, we can state that the more educated the mother is, the more likely she is to seek advice when her child develop a fever.

0.1 MIRDIAG-Malaria Diagnostics Usage-Percentage of febrile children (under age 5) who had a finger or heel stick for malaria testing.

[16]:
```
mlrdiag = pd.read_excel("Final_malaria.xlsx", sheet_name = 1)
mlrdiag
```

[16]:
```
   ISO    Countries      UNICEF Reporting Region UNICEF Programme Region  \
0  AFG  Afghanistan                   South Asia                    ROSA
1  AGO       Angola  Eastern and Southern Africa                   ESARO
2  AGO       Angola  Eastern and Southern Africa                   ESARO
3  BDI      Burundi  Eastern and Southern Africa                   ESARO
```

```
4    BDI      Burundi  Eastern and Southern Africa                    ESARO
..   …          …                           …                      …
147  ZMB       Zambia  Eastern and Southern Africa                    ESARO
148  ZWE     Zimbabwe  Eastern and Southern Africa                    ESARO
149  ZWE     Zimbabwe  Eastern and Southern Africa                    ESARO
150  ZWE     Zimbabwe  Eastern and Southern Africa                    ESARO
151  ZWE     Zimbabwe  Eastern and Southern Africa                    ESARO

    World Bank Income Group (2022)  Year   Short Source  \
0                       Low income  2015       DHS 2015
1              Lower middle income  2011       MIS 2011
2              Lower middle income  2016  DHS 2015-2016
3                       Low income  2010       DHS 2010
4                       Low income  2012       MIS 2012
..                               …   …            …
147                     Low income  2019  DHS 2018-2019
148            Lower middle income  2011  DHS 2010-2011
149            Lower middle income  2014      MICS 2014
150            Lower middle income  2015       DHS 2015
151            Lower middle income  2019      MICS 2019

                                Long Source  National  Male  Female  Rural  \
0         Demographic and Health Survey 2015       7.9   NaN     NaN    8.5
1              Malaria Indicator Survey 2011      26.0  24.8    27.0   16.4
2    Demographic and Health Survey 2015-2016      34.3  35.5    33.0   23.4
3         Demographic and Health Survey 2010      27.0  26.5    27.6   25.6
4              Malaria Indicator Survey 2012      28.3  29.2    27.5   27.9
..                                         …       …    …       …      …
147  Demographic and Health Survey 2018-2019      63.0  64.4    61.4   67.4
148  Demographic and Health Survey 2010-2011       7.0   8.0     7.0    8.0
149   Multiple Indicator Cluster Survey 2014      14.1  13.4    14.9   16.3
150       Demographic and Health Survey 2015      12.7  13.7    11.8   14.7
151   Multiple Indicator Cluster Survey 2019      12.2  13.4    11.1   13.6

     Urban  Poorest  Second  Middle  Fourth  Richest  None  Primary  \
0      6.0      3.9     9.4     9.6    10.5      5.9   NaN      NaN
1     46.0      7.9    12.8    18.6    33.3     41.6   NaN      NaN
2     42.8     19.7    29.2    38.9    44.1     52.7   NaN      NaN
3     47.8     18.7    24.9    27.4    31.8     36.3   NaN      NaN
4     34.7     29.9    24.4    30.6    27.6     30.1   NaN      NaN
..     …       …      …      …      …       …    …        …
147   52.0     67.1    69.9    68.5    47.0     51.5  68.2     65.4
148    5.0      5.0    14.0     5.0     2.0      2.0   NaN      NaN
149    6.6     14.9    17.7    16.5    12.9      5.1   NaN      NaN
150    8.7     16.2    12.5    12.8     9.2     12.9   NaN      NaN
151    8.2     19.0    11.5     9.9    10.5      6.6   NaN      NaN
```

```
     Sec & Higher
0             NaN
1             NaN
2             NaN
3             NaN
4             NaN
..            …
147          57.6
148           NaN
149           NaN
150           NaN
151           NaN

[152 rows x 21 columns]
```

[17]:
```
mlrdiag.isnull().sum()
```

[17]:
```
ISO                               0
Countries                         0
UNICEF Reporting Region           0
UNICEF Programme Region           0
World Bank Income Group (2022)    0
Year                              0
Short Source                      0
Long Source                       0
National                          2
Male                             31
Female                           31
Rural                             5
Urban                             5
Poorest                           7
Second                            8
Middle                            8
Fourth                            9
Richest                          10
None                            119
Primary                         117
Sec & Higher                    116
dtype: int64
```

[18]:
```
mlrdiag.head(2)
```

[18]:
```
   ISO    Countries      UNICEF Reporting Region UNICEF Programme Region  \
0  AFG  Afghanistan                   South Asia                    ROSA
1  AGO       Angola  Eastern and Southern Africa                   ESARO

  World Bank Income Group (2022)  Year Short Source  \
```

```
0                     Low income  2015     DHS 2015
1            Lower middle income  2011     MIS 2011

                         Long Source  National  Male  Female  Rural  Urban  \
0  Demographic and Health Survey 2015       7.9   NaN     NaN    8.5    6.0
1       Malaria Indicator Survey 2011      26.0  24.8    27.0   16.4   46.0

   Poorest  Second  Middle  Fourth  Richest  None  Primary  Sec & Higher
0      3.9     9.4     9.6    10.5      5.9   NaN      NaN           NaN
1      7.9    12.8    18.6    33.3     41.6   NaN      NaN           NaN
```

[19]:

```
# Columns to impute
columns_to_impute = ['National', 'Male', 'Female', 'Rural', 'Urban', 'Poorest',␣
 ↪'Second', 'Middle', 'Fourth', 'Richest', 'None', 'Primary', 'Sec & Higher']

#Create a KNN Imputer object
imputer = KNNImputer(n_neighbors=5, weights='uniform')

#Fit and transform the data
imputed_data = imputer.fit_transform(mlrdiag[columns_to_impute])

# Convert the imputed data back to a pandas DataFrame
imputed_data = pd.DataFrame(imputed_data, columns=columns_to_impute)

# Replace the original columns with the imputed data
mlrdiag[columns_to_impute] = imputed_data
```

[20]:

```
mlrdiag.isna().sum()
```

[20]:

```
ISO                            0
Countries                      0
UNICEF Reporting Region        0
UNICEF Programme Region        0
World Bank Income Group (2022)    0
Year                           0
Short Source                   0
Long Source                    0
National                       0
Male                           0
Female                         0
Rural                          0
Urban                          0
Poorest                        0
Second                         0
Middle                         0
Fourth                         0
Richest                        0
```

```
None                              0
Primary                           0
Sec & Higher                      0
dtype: int64
```

[21]:
```
mlrdiag.duplicated().sum()
```

[21]:
```
0
```

[22]:
```
mlrdiag.head(2)
```

[22]:
```
   ISO    Countries      UNICEF Reporting Region UNICEF Programme Region  \
0  AFG  Afghanistan                   South Asia                    ROSA
1  AGO       Angola  Eastern and Southern Africa                   ESARO

  World Bank Income Group (2022)  Year Short Source  \
0                     Low income  2015     DHS 2015
1            Lower middle income  2011     MIS 2011

                              Long Source  National   Male  Female  Rural  Urban  \
0  Demographic and Health Survey 2015       7.9   7.16    7.58    8.5    6.0
1       Malaria Indicator Survey 2011      26.0  24.80   27.00   16.4   46.0

   Poorest  Second  Middle  Fourth  Richest   None  Primary  Sec & Higher
0      3.9     9.4     9.6    10.5      5.9   7.96     7.80         11.26
1      7.9    12.8    18.6    33.3     41.6  21.60    24.74         31.14
```

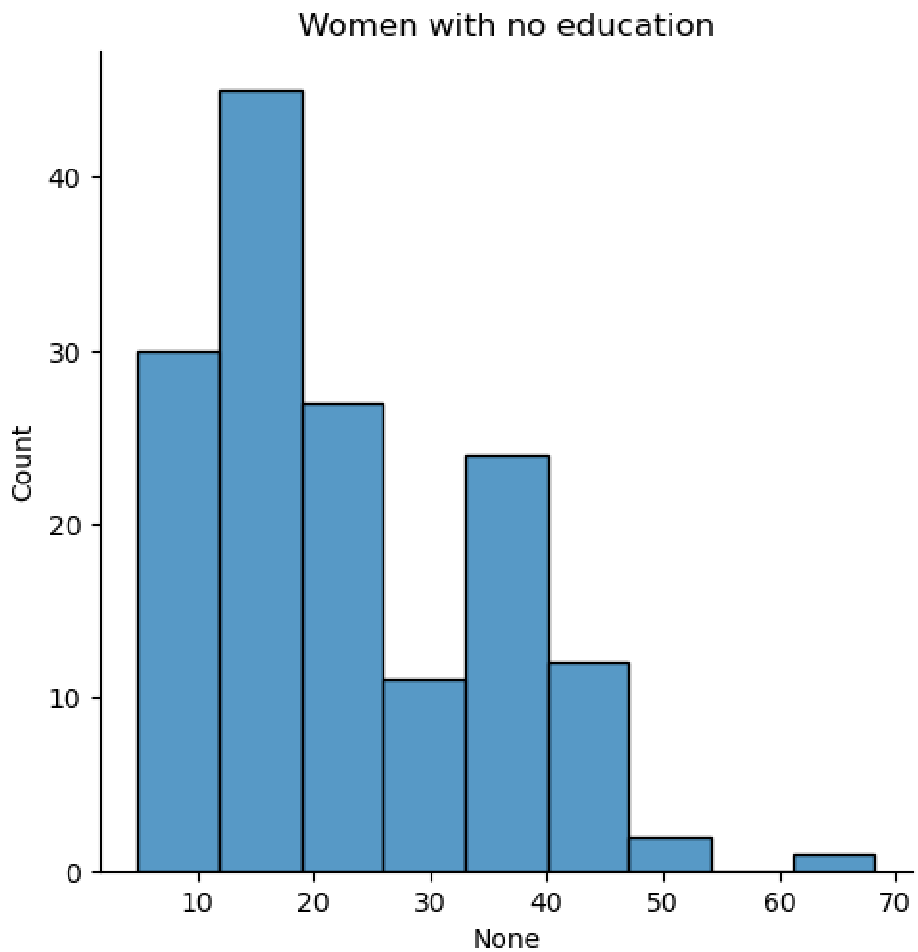
[23]:
```
mlrdiag.head(2)
```

[23]:

```
   ISO    Countries      UNICEF Reporting Region UNICEF Programme Region  \
0  AFG  Afghanistan                   South Asia                    ROSA
1  AGO       Angola  Eastern and Southern Africa                   ESARO

  World Bank Income Group (2022)  Year Short Source  \
0                     Low income  2015     DHS 2015
1            Lower middle income  2011     MIS 2011

                              Long Source  National   Male  Female  Rural  Urban  \
0  Demographic and Health Survey 2015       7.9   7.16    7.58    8.5    6.0
1       Malaria Indicator Survey 2011      26.0  24.80   27.00   16.4   46.0

   Poorest  Second  Middle  Fourth  Richest   None  Primary  Sec & Higher
0      3.9     9.4     9.6    10.5      5.9   7.96     7.80         11.26
1      7.9    12.8    18.6    33.3     41.6  21.60    24.74         31.14
```

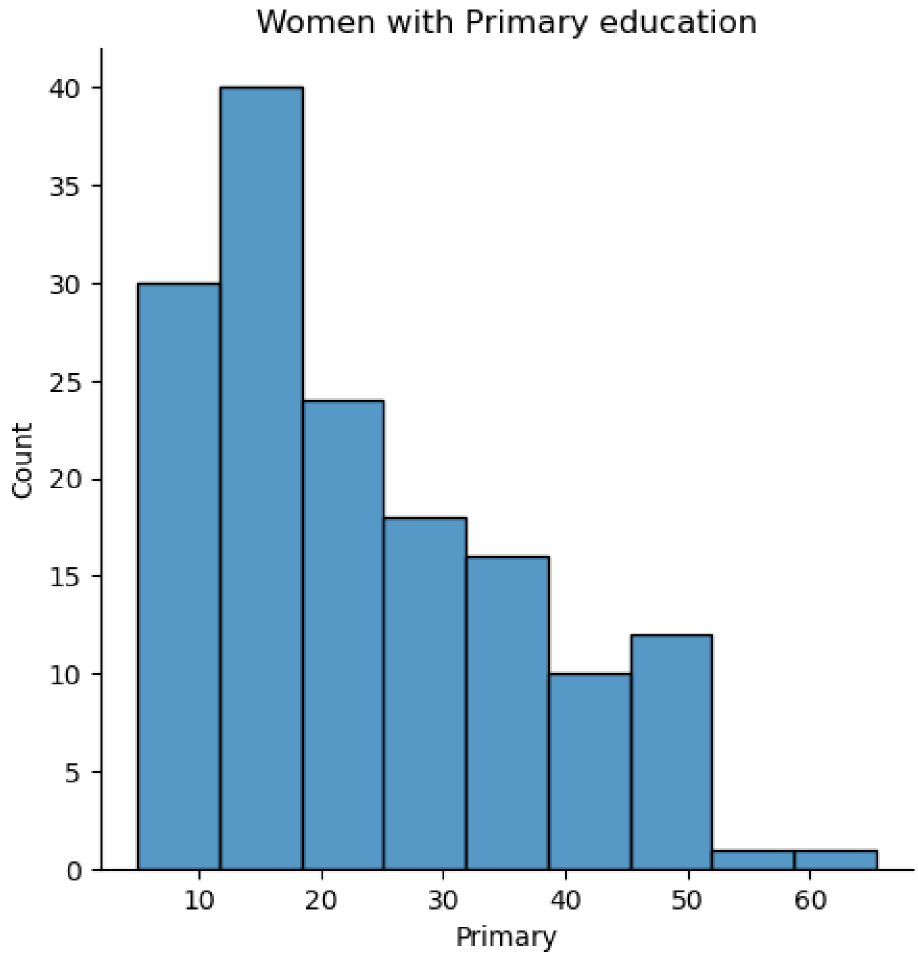

[24]:
```
sns.displot(data=mlrdiag,x="None")
plt.title("Women with no education")
```

```
plt.show()
```



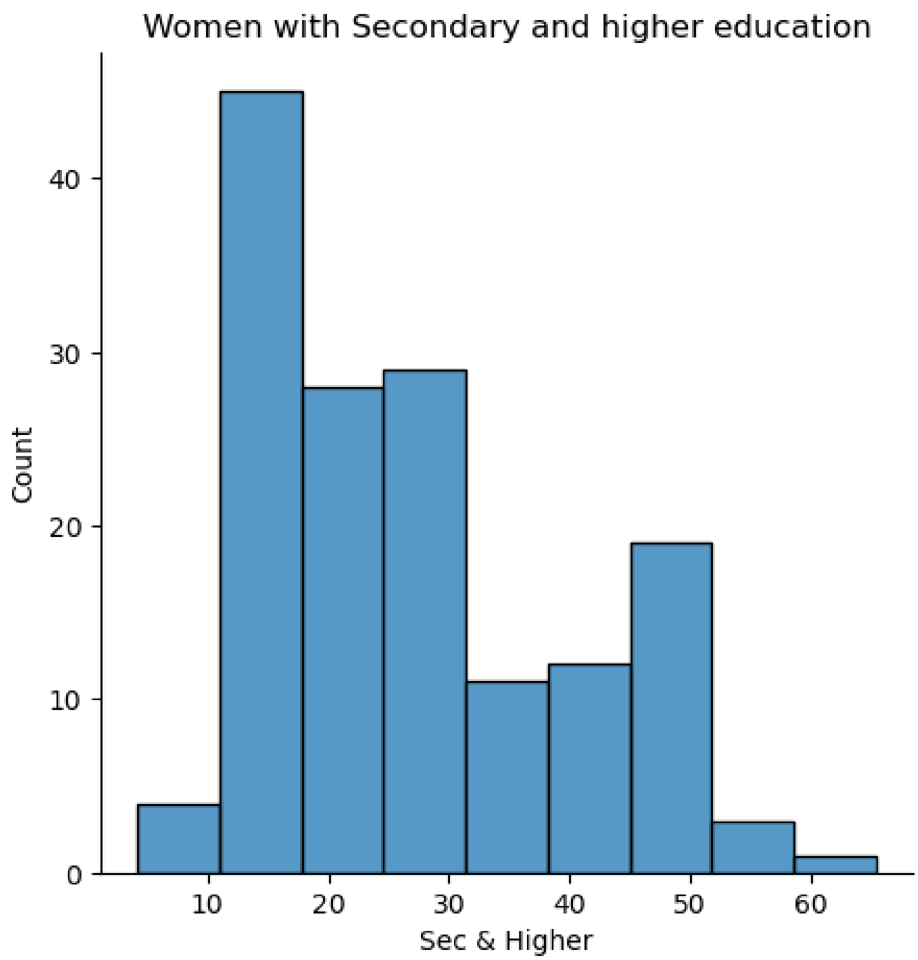
This plot implies that about 12-18% of children who were tested for malaria in most countries had their mothers having no education.

[68]:
```
sns.displot(data=mlrdiag,x="Primary")
plt.title("Women with Primary education")
plt.show()
```



The above plot infers that about 12-18% of children who were tested for malaria in most countries had their mothers having **primary** education.

[69]:
```
sns.displot(data=mlrdiag,x="Sec & Higher")
plt.title("Women with Secondary and higher education")
plt.show()
```



The above plot infers that about 12-17% of children who were tested for malaria in most countries had their mothers having **Secondary and Higher** education.

Our conclusion from this data shows that the education of a mother doesn't directly affect if or not she will bring in her child for malaria diagnosis.

0.2 MLRACT-First-line treatment (ACT) for children under age 5 with fever-Percentage of febrile children (under age 5) receiving ACT (first-line antimalarial drug), among those receiving any antimalarial drugs.

[27]:
```
mlract = pd.read_excel("Final_malaria.xlsx", sheet_name = 2)
mlract.head(2)
```

[27]:
```
   ISO    Countries      UNICEF Reporting Region UNICEF Programme Region  \
0  AFG  Afghanistan                   South Asia                    ROSA
1  AGO       Angola  Eastern and Southern Africa                   ESARO
```

```
  World Bank Income Group (2022)  Year   Short Source  \
0                       Low income  2015       DHS 2015
1              Lower middle income  2007  MIS 2006-2007

                         Long Source  National  Male  Female  Rural  Urban  \
0  Demographic and Health Survey 2015       4.4   5.3     3.6    6.6    0.2
1  Malaria Indicator Survey 2006-2007       5.5   NaN     NaN    NaN    NaN

   Poorest  Second  Middle  Fourth  Richest  None  Primary  Sec & Higher
0      5.9    11.1     2.2     3.9      NaN   NaN      NaN           NaN
1      NaN     NaN     NaN     NaN      NaN   NaN      NaN           NaN
```

[28]:
```
mlract.isnull().sum()
```

[28]:
```
ISO                                0
Countries                          0
UNICEF Reporting Region            0
UNICEF Programme Region            0
World Bank Income Group (2022)     0
Year                               0
Short Source                       0
Long Source                        0
National                           4
Male                              46
Female                            44
Rural                             33
Urban                             41
Poorest                           46
Second                            45
Middle                            51
Fourth                            49
Richest                           52
None                             136
Primary                          137
Sec & Higher                     140
dtype: int64
```

[29]:
```
# Columns to impute
columns_to_impute = ['National', 'Male', 'Female', 'Rural', 'Urban', 'Poorest',␣
 ↪'Second', 'Middle', 'Fourth', 'Richest', 'None', 'Primary', 'Sec & Higher']

#Create a KNN Imputer object
imputer = KNNImputer(n_neighbors=5, weights='uniform')

#Fit and transform the data
imputed_data = imputer.fit_transform(mlract[columns_to_impute])
```

```
# Convert the imputed data back to a pandas DataFrame
imputed_data = pd.DataFrame(imputed_data, columns=columns_to_impute)

# Replace the original columns with the imputed data
mlract[columns_to_impute] = imputed_data
```

[30]:
```
mlract.isnull().sum()
```

[30]:
```
ISO                            0
Countries                      0
UNICEF Reporting Region        0
UNICEF Programme Region        0
World Bank Income Group (2022)    0
Year                           0
Short Source                   0
Long Source                    0
National                       0
Male                           0
Female                         0
Rural                          0
Urban                          0
Poorest                        0
Second                         0
Middle                         0
Fourth                         0
Richest                        0
None                           0
Primary                        0
Sec & Higher                   0
dtype: int64
```

[31]:
```
mlract.head(2)
```

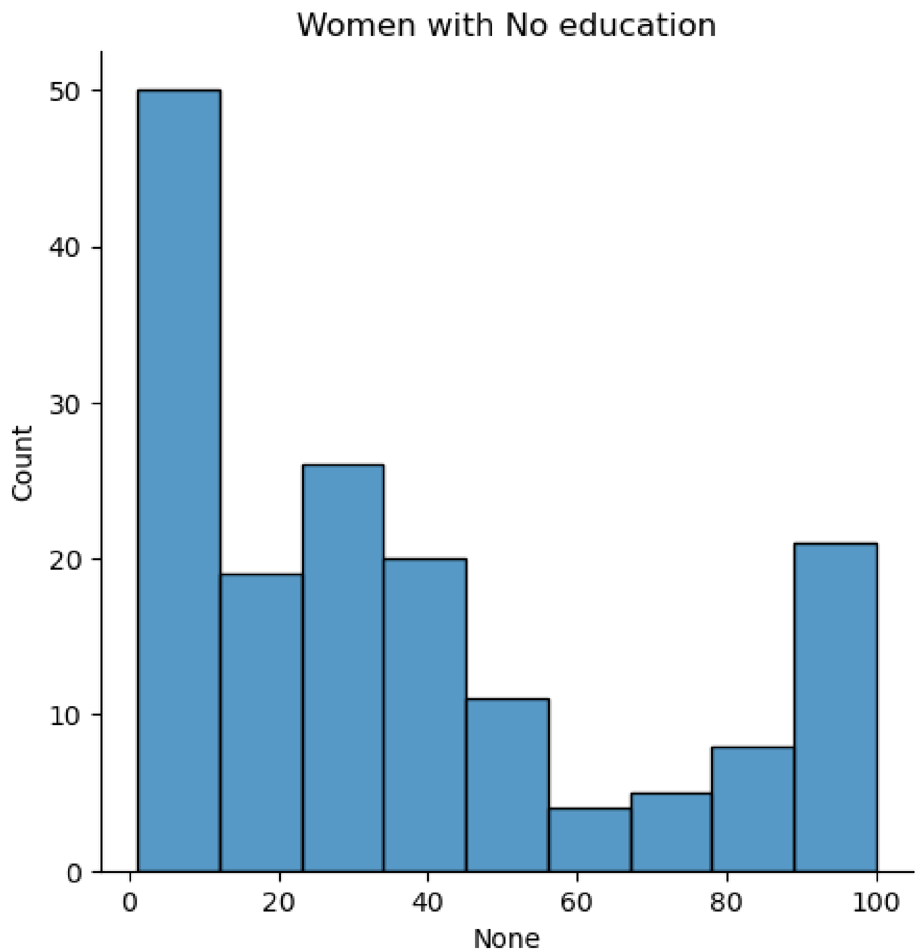
[31]:
```
   ISO    Countries      UNICEF Reporting Region UNICEF Programme Region  \
0  AFG  Afghanistan                   South Asia                    ROSA
1  AGO       Angola  Eastern and Southern Africa                   ESARO

  World Bank Income Group (2022)  Year   Short Source  \
0                     Low income  2015       DHS 2015
1            Lower middle income  2007  MIS 2006-2007

                          Long Source  National  Male  Female  Rural  Urban  \
0  Demographic and Health Survey 2015       4.4  5.30    3.60   6.60   0.20
1  Malaria Indicator Survey 2006-2007       5.5  7.08    4.06   5.58   6.34

   Poorest  Second  Middle  Fourth  Richest  None  Primary  Sec & Higher
0      5.9   11.10    2.20    3.90     5.18  3.82    10.14          7.88
```

```
1      4.6     5.68     8.44     4.82      6.10   3.82     10.14            7.88
```

[32]:
```
mlract.duplicated().sum()
```

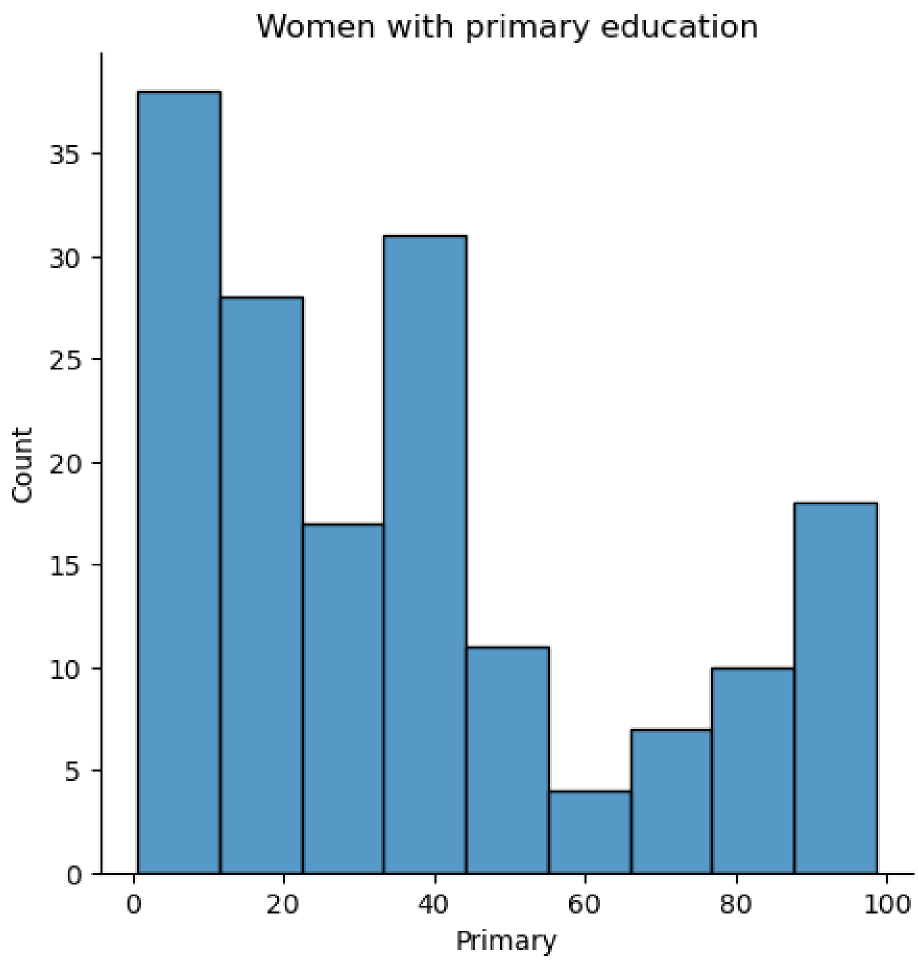
[32]: 0

[70]:
```
sns.displot(data=mlract,x="None")
plt.title("Women with No education")
plt.show()
```



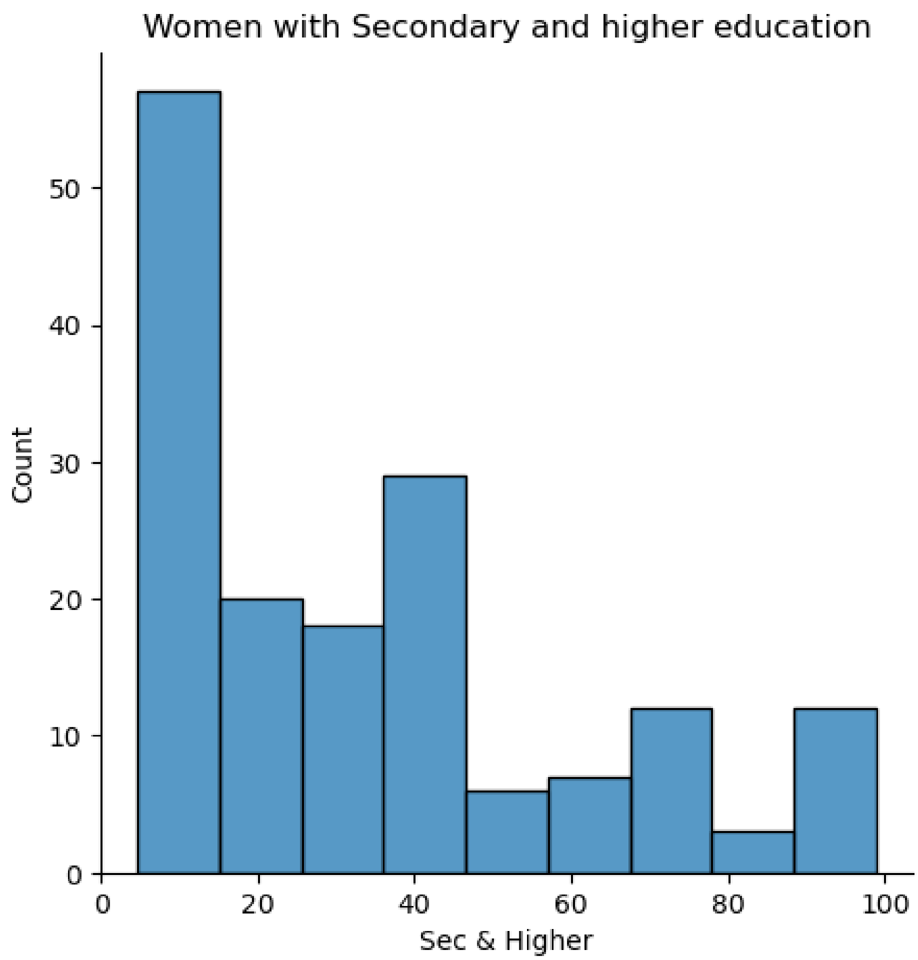
The chart above indicates that about 1-10% of children who recieves antimalaria treatment in most countries had their mothers having no education.

[34]:
```
sns.displot(data=mlract,x="Primary")
plt.title("Women with primary education")
plt.show()
```



The chart above indicates that about 1-15% of children who recieves antimalaria treatment in most countries had their mothers having primary education.

[35]:
```
sns.displot(data=mlract,x="Sec & Higher")
plt.title("Women with Secondary and higher education")
plt.show()
```



The chart above indicates that about 5-15% of children who recieves antimalaria treatment in most countries had their mothers having Secondary and higher education.

In conclusion, this data does not show if the mother's educational level is dependent on whether the child will receive antimalaria treatment. This data shows no clear correlation.

Modelling

[36]:
```
columns_to_keep = ['National', 'None', 'Primary', 'Sec & Higher']

# Select these columns from each DataFrame
mlrcare_selected = mlrcare[columns_to_keep]
mlrdiag_selected = mlrdiag[columns_to_keep]
mlract_selected = mlract[columns_to_keep]
```

Building seperate models for the 3 different dataset (mlrcare, mlrdiag, mlract) because the 3 dataset has same variable names but contain different information

mlrcare model

```
[37]: # building model for the mlrcare dataset
X = mlrcare_selected.drop(columns = ['National'])
y = mlrcare_selected[['National']]
```

```
[38]: x_train, x_test, y_train, y_test = train_test_split(X, y, test_size=0.2,␣
  ↪random_state=42)
```

```
[39]: numerical = ['None', 'Primary', 'Sec & Higher']
```

```
[40]: scaler = MinMaxScaler()
scaler.fit(x_train[numerical])
```

```
[40]: MinMaxScaler()
```

```
[41]: x_train[numerical] = scaler.transform(x_train[numerical])
```

```
[42]: x_train.head()
```

[42]:

	None	Primary	Sec & Higher
86	0.586242	0.573551	0.477516
202	0.664765	0.700362	0.556317
67	0.586242	0.605797	0.549036
82	0.596309	0.587319	0.554604
204	0.902685	0.762681	0.783726

```
[43]: x_test[numerical] = scaler.transform(x_test[numerical])
```

```
[44]: x_test.head()
```

[44]:

	None	Primary	Sec & Higher
161	0.553356	0.530797	0.610707
15	0.553356	0.597101	0.503212
73	0.693960	0.712319	0.697216
96	0.944295	0.904348	0.886938
166	0.490940	0.451087	0.391006

```
[45]: model = LinearRegression()
model.fit(x_train, y_train)
```

```
[45]: LinearRegression()
```

```
[46]: coefficients = model.coef_
print(coefficients)
```

```
[[36.32372328 14.64173702 10.60432609]]
```

From the model, we have 3 different coefficients, which indicates the coeeficients of (None, Primary, Sec & Higher) varaibles respectively. This provides a deeper relationship into the effects of maternal education on the rate of malaria on children.

The model tells us that the None (which is the percentage of children having fever whose mothers with no education sought advice or treatment for) has the highest effects on the overall population of children with fever, followed by the primary variable.

This implies that mothers who have no education contributes most to a higher malaria rates among the children. Which shows that the lower the education of the mothers, the higher the malaria rates among the children.

mlrdiag model

[47]:
```
# building model for the mlrdiag dataset
X = mlrdiag_selected.drop(columns = ['National'])
y = mlrdiag_selected[['National']]
```

[48]:
```
x_train, x_test, y_train, y_test = train_test_split(X, y, test_size=0.2,␣
↪random_state=42)
```

[49]:
```
numerical = ['None', 'Primary', 'Sec & Higher']
```

[50]:
```
scaler = MinMaxScaler()
scaler.fit(x_train[numerical])
```

[50]:
```
MinMaxScaler()
```

[51]:
```
x_train[numerical] = scaler.transform(x_train[numerical])
```

[52]:
```
x_train.head()
```

[52]:

	None	Primary	Sec & Higher
29	0.236449	0.230444	0.255882
22	0.251402	0.268499	0.297712
51	0.084112	0.076110	0.138889
75	0.591121	0.782241	0.740196
11	0.563551	0.490063	0.437255

[53]:
```
x_test[numerical] = scaler.transform(x_test[numerical])
```

[54]:
```
x_test.head()
```

[54]:

	None	Primary	Sec & Higher
68	0.336449	0.340381	0.238562
147	1.481308	1.276956	0.872549
96	0.242991	0.272727	0.230392
82	0.741121	0.742918	0.730719
135	0.658879	0.841438	0.748366

```
[55]: model = LinearRegression()
      model.fit(x_train, y_train)
```

```
[55]: LinearRegression()
```

```
[56]: coefficients = model.coef_
      print(coefficients)
```

```
[[ 8.07904825 40.40812902  4.85946773]]
```

From the model, we have 3 different coefficients, which indicates the coeeficients of (None, Primary, Sec & Higher) varaibles respectively. This provides a deeper relationship into the effects of maternal education on the rate of malaria on children.

The model tells us that the Sec & Higher (which is the percentage of children who were tested for malaria whose mothers has secondary or higher education) has the lowest effects on the overall population of children who were tested for malaria.

This implies that mothers who has secondary or higher education contributes least to a higher malaria rates among the children. Which shows that the higher the education of the mothers, the lesser the malaria rates among the children.

mlract model

```
[57]: # building model for the mlract dataset
      X = mlract_selected.drop(columns = ['National'])
      y = mlract_selected[['National']]
```

```
[58]: x_train, x_test, y_train, y_test = train_test_split(X, y, test_size=0.2,␣
       ↪random_state=42)
```

```
[59]: numerical = ['None', 'Primary', 'Sec & Higher']
```

```
[60]: scaler = MinMaxScaler()
      scaler.fit(x_train[numerical])
```

```
[60]: MinMaxScaler()
```

```
[61]: x_train[numerical] = scaler.transform(x_train[numerical])
```

```
[62]: x_train.head()
```

[62]:

	None	Primary	Sec & Higher
84	0.982811	1.000000	1.000000
2	0.193124	0.217365	0.193631
94	0.967644	0.977528	0.917834
45	0.027503	0.113177	0.033758
42	0.087159	0.146476	0.080467

[63]:
```
x_test[numerical] = scaler.transform(x_test[numerical])
```

[64]:
```
x_test.head()
```

[64]:

	None	Primary	Sec & Higher
135	0.171891	0.193258	0.125265
115	0.176340	0.217365	0.193631
131	0.401011	0.373647	0.348408
55	0.065723	0.064351	0.047771
95	0.460061	0.448417	0.416773

[65]:
```
model = LinearRegression()
model.fit(x_train, y_train)
```

[65]: LinearRegression()

[66]:
```
coefficients = model.coef_
print(coefficients)
```

```
[[50.68254443 20.01555925 26.60376744]]
```

From the model, we have 3 different coefficients, which indicates the coeeficients of (None, Primary, Sec & Higher) varaibles respectively. This provides a deeper relationship into the effects of maternal education on the rate of malaria on children.

The model tells us that the None variable (which is the percentage of children who recieved anti-malaria drugs having mothers with no education) has the highest effects on the overall population of children who were giving anti-malaria drugs.

This implies that mothers who has no education contributes most to a higher malaria rates among the children. Which shows that the lower the education of the mothers, the higher the malaria rates among the children.

[ ]:
```

```