ml1days

December 18, 2024

[ ]:
```
pandas
```

[104]:
```
import pandas as pd
```

[106]:
```
df=pd.read_csv(r"C:\Users\hp\Downloads\Titanic-Dataset.csv")
```

[70]:
```
df
```

[70]:

	Country	Region	Year	\
0	Afghanistan	Asia	2007.5	
1	Albania	Rest of Europe	2007.5	
2	Algeria	Africa	2007.5	
3	Angola	Africa	2007.5	
4	Antigua and Barbuda	Central America and Caribbean	2007.5	
..	…	…	…	
174	Venezuela, RB	South America	2007.5	
175	Vietnam	Asia	2007.5	
176	Yemen, Rep.	Middle East	2007.5	
177	Zambia	Africa	2007.5	
178	Zimbabwe	Africa	2007.5	

	Infant_deaths	Under_five_deaths	Adult_mortality	Alcohol_consumption	\
0	71.08125	98.61250	265.804969	0.016125	
1	15.25625	17.14375	83.132969	4.696875	
2	26.75625	31.19375	113.439281	0.400625	
3	88.76875	144.16250	297.844063	4.935625	
4	9.47500	11.51875	142.478813	7.755000	
..	…	…	…	…	
174	15.85000	18.50000	151.056156	7.458125	
175	19.50000	24.47500	133.054406	1.742500	
176	52.44375	69.57500	239.653938	0.047000	
177	60.58125	99.20625	452.761219	2.903125	
178	50.16875	85.32500	572.974312	2.900625	

	Hepatitis_B	Measles	BMI	Polio	Diphtheria	Incidents_HIV	\
0	64.5625	24.3750	22.46250	55.3750	55.1250	0.022500	
1	98.0000	95.9375	25.85625	98.1250	98.0625	0.025625	
2	88.3125	93.2500	24.86875	91.7500	91.8750	0.021875	

```
3        68.8125  64.0000  22.51875  35.7500     55.5625        1.303750
4        98.2500  75.4375  25.85000  96.9375     98.3125        0.125000
..           …        …        …         …           …              …
174      72.4375  83.0000  26.49375  79.1875     75.2500        0.400000
175      87.4375  65.0000  20.76250  94.9375     91.7500        0.196875
176      60.1875  95.0000  23.02500  71.6250     72.6250        0.025000
177      82.8750  39.1875  22.00000  82.9375     83.8125        6.868125
178      79.5625  64.0000  23.54375  79.5625     79.1250        8.006250

     GDP_per_capita  Population_mln  Thinness_ten_nineteen_years  \
0          408.5625       27.450625                      16.58125
1         3071.1250        2.969375                       1.61875
2         3745.1250       34.820625                       6.09375
3         2647.8125       21.623750                       6.19375
4        14678.7500        0.085000                       3.42500
..              …              …                             …
174       9294.8750       27.392500                       1.65000
175       1807.8750       86.032500                      14.92500
176       2406.8125       21.730625                      13.83125
177       1059.8750       12.862500                       6.88125
178       1290.6250       12.538750                       7.01250

     Thinness_five_nine_years  Schooling  Economy_status  Life_expectancy
0                    15.58125    2.90000             1.0         59.65625
1                     1.70000    9.24375             1.0         75.95000
2                     5.97500    6.99375             1.0         73.78750
3                     6.66875    4.60625             1.0         52.82500
4                     3.37500    9.01875             1.0         75.35000
..                        …          …               …              …
174                   1.56250    8.23750             1.0         72.79375
175                  15.62500    6.90000             1.0         74.33125
176                  13.75000    2.22500             1.0         64.16250
177                   6.76250    6.43750             1.0         52.33125
178                   6.98750    7.21250             1.0         48.66250

[179 rows x 20 columns]
```

[ ]:
```
df.describe()
```

[72]:
```
df.describe(include='all')
```

[72]:
```
            Country  Region    Year  Infant_deaths  Under_five_deaths  \
count           179     179   179.0     179.000000         179.000000
unique          179       9     NaN            NaN                NaN
top     Afghanistan  Africa     NaN            NaN                NaN
freq              1      51     NaN            NaN                NaN
mean            NaN     NaN  2007.5      30.363792          42.938268
```

std	NaN	NaN	0.0	26.725485	42.916952
min	NaN	NaN	2007.5	2.381250	3.000000
25%	NaN	NaN	2007.5	8.159375	9.775000
50%	NaN	NaN	2007.5	19.368750	23.137500
75%	NaN	NaN	2007.5	48.959375	68.321875
max	NaN	NaN	2007.5	115.718750	178.725000

	Adult_mortality	Alcohol_consumption	Hepatitis_B	Measles	\
count	179.000000	179.000000	179.000000	179.000000	
unique	NaN	NaN	NaN	NaN	
top	NaN	NaN	NaN	NaN	
freq	NaN	NaN	NaN	NaN	
mean	192.251775	4.820882	84.292598	77.344972	
std	111.659044	3.914554	13.820223	17.315208	
min	57.710313	0.000025	30.687500	16.250000	
25%	107.046906	1.317813	78.218750	64.000000	
50%	164.432406	4.209375	88.000000	83.000000	
75%	247.523922	7.843438	94.375000	92.250000	
max	572.974312	15.100000	98.875000	99.000000	

	BMI	Polio	Diphtheria	Incidents_HIV	GDP_per_capita	\
count	179.000000	179.000000	179.000000	179.000000	179.000000	
unique	NaN	NaN	NaN	NaN	NaN	
top	NaN	NaN	NaN	NaN	NaN	
freq	NaN	NaN	NaN	NaN	NaN	
mean	25.032926	86.499651	86.271648	0.894288	11540.924930	
std	2.165490	13.581153	13.931532	2.311895	16893.054182	
min	20.212500	35.750000	31.312500	0.010000	263.937500	
25%	23.225000	80.531250	80.812500	0.080000	1409.906250	
50%	25.650000	92.375000	92.062500	0.164375	4402.625000	
75%	26.425000	96.062500	95.781250	0.516250	12037.781250	
max	31.687500	98.937500	99.000000	18.164375	102972.687500	

	Population_mln	Thinness_ten_nineteen_years	Thinness_five_nine_years	\
count	179.000000	179.000000	179.000000	
unique	NaN	NaN	NaN	
top	NaN	NaN	NaN	
freq	NaN	NaN	NaN	
mean	36.675915	4.865852	4.899825	
std	136.655286	4.111094	4.195663	
min	0.085000	0.100000	0.100000	
25%	2.108125	1.756250	1.731250	
50%	7.660625	3.556250	3.718750	
75%	22.745313	7.165625	7.056250	
max	1321.239375	27.100000	27.943750	

	Schooling	Economy_status	Life_expectancy

```
count    179.000000      179.000000       179.000000
unique          NaN             NaN              NaN
top             NaN             NaN              NaN
freq            NaN             NaN              NaN
mean       7.632123        0.793296        68.856075
std        3.126912        0.406077         9.197699
min        1.337500        0.000000        45.606250
25%        4.946875        1.000000        62.303125
50%        7.831250        1.000000        71.506250
75%       10.365625        1.000000        74.937500
max       13.268750        1.000000        82.456250
```

[74]:
```
df.describe(include='object')
```

[74]:
```
            Country  Region
count           179     179
unique          179       9
top     Afghanistan  Africa
freq              1      51
```

[76]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 179 entries, 0 to 178
Data columns (total 20 columns):
 #   Column                         Non-Null Count  Dtype
---  ------                         --------------  -----
 0   Country                        179 non-null    object
 1   Region                         179 non-null    object
 2   Year                           179 non-null    float64
 3   Infant_deaths                  179 non-null    float64
 4   Under_five_deaths              179 non-null    float64
 5   Adult_mortality                179 non-null    float64
 6   Alcohol_consumption            179 non-null    float64
 7   Hepatitis_B                    179 non-null    float64
 8   Measles                        179 non-null    float64
 9   BMI                            179 non-null    float64
 10  Polio                          179 non-null    float64
 11  Diphtheria                     179 non-null    float64
 12  Incidents_HIV                  179 non-null    float64
 13  GDP_per_capita                 179 non-null    float64
 14  Population_mln                 179 non-null    float64
 15  Thinness_ten_nineteen_years    179 non-null    float64
 16  Thinness_five_nine_years       179 non-null    float64
 17  Schooling                      179 non-null    float64
 18  Economy_status                 179 non-null    float64
 19  Life_expectancy                179 non-null    float64
dtypes: float64(18), object(2)
```

```
memory usage: 28.1+ KB
```

[78]:
```
df.head()
```

[78]:

	Country	Region	Year	Infant_deaths	Under_five_deaths	Adult_mortality	Alcohol_consumption	Hepatitis_B	Measles	BMI	Polio	Diphtheria	Incidents_HIV	GDP_per_capita	Population_mln	Thinness_ten_nineteen_years	Thinness_five_nine_years	Schooling	Economy_status	Life_expectancy
0	Afghanistan	Asia	2007.5	71.08125	98.61250	265.804969	0.016125	64.5625	24.3750	22.46250	55.3750	55.1250	0.022500	408.5625	27.450625	16.58125	15.58125	2.90000	1.0	59.65625
1	Albania	Rest of Europe	2007.5	15.25625	17.14375	83.132969	4.696875	98.0000	95.9375	25.85625	98.1250	98.0625	0.025625	3071.1250	2.969375	1.61875	1.70000	9.24375	1.0	75.95000
2	Algeria	Africa	2007.5	26.75625	31.19375	113.439281	0.400625	88.3125	93.2500	24.86875	91.7500	91.8750	0.021875	3745.1250	34.820625	6.09375	5.97500	6.99375	1.0	73.78750
3	Angola	Africa	2007.5	88.76875	144.16250	297.844063	4.935625	68.8125	64.0000	22.51875	35.7500	55.5625	1.303750	2647.8125	21.623750	6.19375	6.66875	4.60625	1.0	52.82500
4	Antigua and Barbuda	Central America and Caribbean	2007.5	9.47500	11.51875	142.478813	7.755000	98.2500	75.4375	25.85000	96.9375	98.3125	0.125000	14678.7500	0.085000	3.42500	3.37500	9.01875	1.0	75.35000

[80]:
```
df.columns
```

[80]:
```
Index(['Country', 'Region', 'Year', 'Infant_deaths', 'Under_five_deaths',
       'Adult_mortality', 'Alcohol_consumption', 'Hepatitis_B', 'Measles',
       'BMI', 'Polio', 'Diphtheria', 'Incidents_HIV', 'GDP_per_capita',
       'Population_mln', 'Thinness_ten_nineteen_years',
       'Thinness_five_nine_years', 'Schooling', 'Economy_status',
       'Life_expectancy'],
      dtype='object')
```

[82]:
```
#view first 5 columns
columns=df.iloc[0:5]
columns
```

[82]:

	Country	Region	Year	Infant_deaths	Under_five_deaths	Adult_mortality	Alcohol_consumption	Hepatitis_B	Measles	BMI	Polio	Diphtheria	Incidents_HIV	GDP_per_capita	Population_mln	Thinness_ten_nineteen_years	Thinness_five_nine_years	Schooling	Economy_status	Life_expectancy
0	Afghanistan	Asia	2007.5	71.08125	98.61250	265.804969	0.016125	64.5625	24.3750	22.46250	55.3750	55.1250	0.022500	408.5625	27.450625	16.58125	15.58125	2.90000	1.0	59.65625
1	Albania	Rest of Europe	2007.5	15.25625	17.14375	83.132969	4.696875	98.0000	95.9375	25.85625	98.1250	98.0625	0.025625	3071.1250	2.969375	1.61875	1.70000	9.24375	1.0	75.95000
2	Algeria	Africa	2007.5	26.75625	31.19375	113.439281	0.400625	88.3125	93.2500	24.86875	91.7500	91.8750	0.021875	3745.1250	34.820625	6.09375	5.97500	6.99375	1.0	73.78750
3	Angola	Africa	2007.5	88.76875	144.16250	297.844063	4.935625	68.8125	64.0000	22.51875	35.7500	55.5625	1.303750	2647.8125	21.623750	6.19375	6.66875	4.60625	1.0	52.82500
4	Antigua and Barbuda	Central America and Caribbean	2007.5	9.47500	11.51875	142.478813	7.755000	98.2500	75.4375	25.85000	96.9375	98.3125	0.125000	14678.7500	0.085000	3.42500	3.37500	9.01875	1.0	75.35000

[84]:
```
df.isnull()
```

[84]:

	Country	Region	Year	Infant_deaths	Under_five_deaths
0	False	False	False	False	False
1	False	False	False	False	False
2	False	False	False	False	False
3	False	False	False	False	False
4	False	False	False	False	False

```
..       …     …    …            …                 …
174    False   False  False          False             False
175    False   False  False          False             False
176    False   False  False          False             False
177    False   False  False          False             False
178    False   False  False          False             False

     Adult_mortality  Alcohol_consumption  Hepatitis_B  Measles    BMI  Polio  \
0              False                False        False    False  False  False
1              False                False        False    False  False  False
2              False                False        False    False  False  False
3              False                False        False    False  False  False
4              False                False        False    False  False  False
..               …                    …          …      …    …    …
174            False                False        False    False  False  False
175            False                False        False    False  False  False
176            False                False        False    False  False  False
177            False                False        False    False  False  False
178            False                False        False    False  False  False

     Diphtheria  Incidents_HIV  GDP_per_capita  Population_mln  \
0         False          False           False           False
1         False          False           False           False
2         False          False           False           False
3         False          False           False           False
4         False          False           False           False
..          …            …             …             …
174       False          False           False           False
175       False          False           False           False
176       False          False           False           False
177       False          False           False           False
178       False          False           False           False

     Thinness_ten_nineteen_years  Thinness_five_nine_years  Schooling  \
0                          False                     False      False
1                          False                     False      False
2                          False                     False      False
3                          False                     False      False
4                          False                     False      False
..                           …                       …        …
174                        False                     False      False
175                        False                     False      False
176                        False                     False      False
177                        False                     False      False
178                        False                     False      False

     Economy_status  Life_expectancy
```

```
0             False            False
1             False            False
2             False            False
3             False            False
4             False            False
..              …                …
174           False            False
175           False            False
176           False            False
177           False            False
178           False            False

[179 rows x 20 columns]
```

[86]:
```
df.isnull().sum()
```

[86]:
```
Country                        0
Region                         0
Year                           0
Infant_deaths                  0
Under_five_deaths              0
Adult_mortality                0
Alcohol_consumption            0
Hepatitis_B                    0
Measles                        0
BMI                            0
Polio                          0
Diphtheria                     0
Incidents_HIV                  0
GDP_per_capita                 0
Population_mln                 0
Thinness_ten_nineteen_years    0
Thinness_five_nine_years       0
Schooling                      0
Economy_status                 0
Life_expectancy                0
dtype: int64
```

[88]:
```
#Clean the Data
df.dropna(inplace=True)
```

[90]:
```
df.shape
```

[90]:
```
(179, 20)
```

[92]:
```
# previewing dataset
df.sample(5)
```

[92]:

	Country	Region	Year	Infant_deaths	Under_five_deaths	\
5	Argentina	South America	2007.5	13.85000	15.50625	
52	Eritrea	Africa	2007.5	43.28750	62.59375	
132	Romania	European Union	2007.5	12.90625	15.23125	
61	Georgia	Rest of Europe	2007.5	18.06250	20.38125	
176	Yemen, Rep.	Middle East	2007.5	52.44375	69.57500	

	Adult_mortality	Alcohol_consumption	Hepatitis_B	Measles	BMI	\
5	127.912656	7.959375	85.3125	84.6875	26.85000	
52	314.237875	0.996875	91.8750	88.0000	20.43125	
132	153.233812	10.583750	96.8750	92.8750	25.71875	
61	170.071969	5.524375	77.2500	75.7500	26.58750	
176	239.653938	0.047000	60.1875	95.0000	23.02500	

	Polio	Diphtheria	Incidents_HIV	GDP_per_capita	Population_mln	\
5	92.2500	92.375	0.145625	12169.1250	39.913125	
52	93.5625	92.375	0.333750	1267.5000	2.759375	
132	94.7500	95.000	0.043125	7011.0625	20.861875	
61	86.2500	88.000	0.091250	2780.5625	3.856250	
176	71.6250	72.625	0.025000	2406.8125	21.730625	

	Thinness_ten_nineteen_years	Thinness_five_nine_years	Schooling	\
5	1.07500	0.95000	9.36875	
52	8.08125	8.01875	3.79375	
132	3.16875	3.50625	10.37500	
61	2.75000	2.89375	12.18125	
176	13.83125	13.75000	2.22500	

	Economy_status	Life_expectancy
5	1.0	74.86250
52	1.0	60.28125
132	0.0	72.88125
61	1.0	71.15625
176	1.0	64.16250

[94]:
```
df.dtypes
```

[94]:
```
Country                    object
Region                     object
Year                      float64
Infant_deaths             float64
Under_five_deaths         float64
Adult_mortality           float64
Alcohol_consumption       float64
Hepatitis_B               float64
Measles                   float64
BMI                       float64
```

```
Polio                          float64
Diphtheria                     float64
Incidents_HIV                  float64
GDP_per_capita                 float64
Population_mln                 float64
Thinness_ten_nineteen_years    float64
Thinness_five_nine_years       float64
Schooling                      float64
Economy_status                 float64
Life_expectancy                float64
dtype: object
```

[96]:
```
# view first row
row=df.iloc[0]
row
```

[96]:
```
Country                        Afghanistan
Region                                Asia
Year                                2007.5
Infant_deaths                     71.08125
Under_five_deaths                  98.6125
Adult_mortality                  265.804969
Alcohol_consumption              0.016125
Hepatitis_B                       64.5625
Measles                            24.375
BMI                               22.4625
Polio                              55.375
Diphtheria                         55.125
Incidents_HIV                      0.0225
GDP_per_capita                   408.5625
Population_mln                  27.450625
Thinness_ten_nineteen_years       16.58125
Thinness_five_nine_years          15.58125
Schooling                              2.9
Economy_status                         1.0
Life_expectancy                   59.65625
Name: 0, dtype: object
```

[98]:
```
#view first 5 row
row=df.iloc[0:5]
row
```

[98]:

	Country	Region	Year	Infant_deaths	\
0	Afghanistan	Asia	2007.5	71.08125	
1	Albania	Rest of Europe	2007.5	15.25625	
2	Algeria	Africa	2007.5	26.75625	
3	Angola	Africa	2007.5	88.76875	

```
4  Antigua and Barbuda  Central America and Caribbean  2007.5        9.47500

   Under_five_deaths  Adult_mortality  Alcohol_consumption  Hepatitis_B  \
0           98.61250       265.804969             0.016125      64.5625
1           17.14375        83.132969             4.696875      98.0000
2           31.19375       113.439281             0.400625      88.3125
3          144.16250       297.844063             4.935625      68.8125
4           11.51875       142.478813             7.755000      98.2500

   Measles       BMI    Polio  Diphtheria  Incidents_HIV  GDP_per_capita  \
0  24.3750  22.46250  55.3750     55.1250       0.022500        408.5625
1  95.9375  25.85625  98.1250     98.0625       0.025625       3071.1250
2  93.2500  24.86875  91.7500     91.8750       0.021875       3745.1250
3  64.0000  22.51875  35.7500     55.5625       1.303750       2647.8125
4  75.4375  25.85000  96.9375     98.3125       0.125000      14678.7500

   Population_mln  Thinness_ten_nineteen_years  Thinness_five_nine_years  \
0       27.450625                     16.58125                  15.58125
1        2.969375                      1.61875                   1.70000
2       34.820625                      6.09375                   5.97500
3       21.623750                      6.19375                   6.66875
4        0.085000                      3.42500                   3.37500

   Schooling  Economy_status  Life_expectancy
0    2.90000             1.0         59.65625
1    9.24375             1.0         75.95000
2    6.99375             1.0         73.78750
3    4.60625             1.0         52.82500
4    9.01875             1.0         75.35000
```

[108]:
```
df['Fare'].value_counts()
```

[108]:
```
Fare
8.0500     43
13.0000    42
7.8958     38
7.7500     34
26.0000    31
           ..
35.0000     1
28.5000     1
6.2375      1
14.0000     1
10.5167     1
Name: count, Length: 248, dtype: int64
```

[102]:
```
df['Fare'].unique()
```

```
---------------------------------------------------------------------------
KeyError                                  Traceback (most recent call last)
File ~\anaconda3\Lib\site-packages\pandas\core\indexes\base.py:3805, in Index.
 ↪get_loc(self, key)
   3804 try:
-> 3805     return self._engine.get_loc(casted_key)
   3806 except KeyError as err:

File index.pyx:167, in pandas._libs.index.IndexEngine.get_loc()

File index.pyx:196, in pandas._libs.index.IndexEngine.get_loc()

File pandas\\_libs\\hashtable_class_helper.pxi:7081, in pandas._libs.hashtable.
 ↪PyObjectHashTable.get_item()

File pandas\\_libs\\hashtable_class_helper.pxi:7089, in pandas._libs.hashtable.
 ↪PyObjectHashTable.get_item()

KeyError: 'Fare'

The above exception was the direct cause of the following exception:

KeyError                                  Traceback (most recent call last)
Cell In[102], line 1
----> 1 df['Fare'].unique()

File ~\anaconda3\Lib\site-packages\pandas\core\frame.py:4102, in DataFrame.
 ↪__getitem__(self, key)
   4100 if self.columns.nlevels > 1:
   4101     return self._getitem_multilevel(key)
-> 4102 indexer = self.columns.get_loc(key)
   4103 if is_integer(indexer):
   4104     indexer = [indexer]

File ~\anaconda3\Lib\site-packages\pandas\core\indexes\base.py:3812, in Index.
 ↪get_loc(self, key)
   3807     if isinstance(casted_key, slice) or (
   3808         isinstance(casted_key, abc.Iterable)
   3809         and any(isinstance(x, slice) for x in casted_key)
   3810     ):
   3811         raise InvalidIndexError(key)
-> 3812     raise KeyError(key) from err
   3813 except TypeError:
   3814     # If we have a listlike key, _check_indexing_error will raise
   3815     #  InvalidIndexError. Otherwise we fall through and re-raise
   3816     #  the TypeError.
   3817     self._check_indexing_error(key)
```

```
KeyError: 'Fare'
```

[182]:
```
Fare=df['Fare']
```

[110]:
```
df['Fare'].nunique()
```

[110]:
```
248
```

[112]:
```
df['Pclass'].nunique()
```

[112]:
```
3
```

[114]:
```
df['Pclass'].unique()
```

[114]:
```
array([3, 1, 2], dtype=int64)
```

[116]:
```
df['Pclass'].value_counts()
```

[116]:
```
Pclass
3    491
1    216
2    184
Name: count, dtype: int64
```

[118]:
```
df['Pclass'].value_counts(normalize=True)
```

[118]:
```
Pclass
3    0.551066
1    0.242424
2    0.206510
Name: proportion, dtype: float64
```

[120]:
```
df['Sex'].unique()
```

[120]:
```
array(['male', 'female'], dtype=object)
```

[122]:
```
df['Sex'].nunique()
```

[122]:
```
2
```

[124]:
```
df['Sex'].value_counts()
```

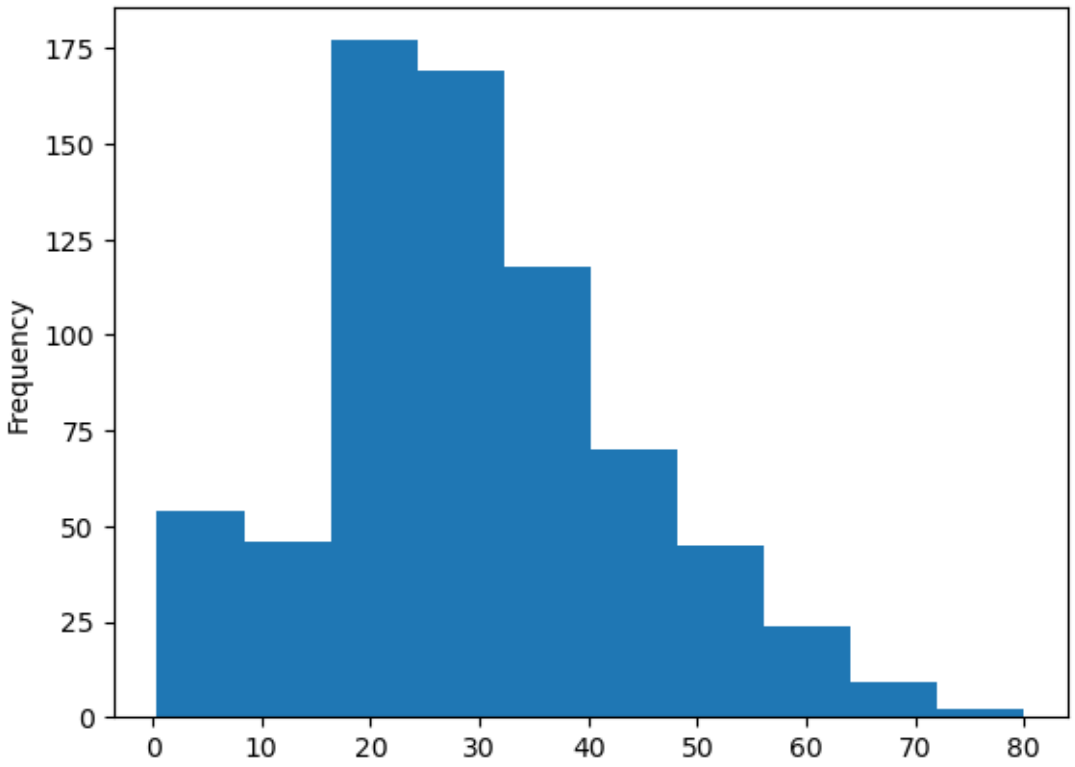
[124]:
```
Sex
male      577
female    314
Name: count, dtype: int64
```

[126]:
```
df['Age'].value_counts()
```

[126]:
```
Age
24.00    30
22.00    27
18.00    26
19.00    25
28.00    25
         ..
36.50     1
55.50     1
0.92      1
23.50     1
74.00     1
Name: count, Length: 88, dtype: int64
```

[128]:
```
df['Age'].plot.hist()
```

[128]:
```
<Axes: ylabel='Frequency'>
```



[130]:
```
df['Cabin'].unique()
```

[130]:
```
array([nan, 'C85', 'C123', 'E46', 'G6', 'C103', 'D56', 'A6',
       'C23 C25 C27', 'B78', 'D33', 'B30', 'C52', 'B28', 'C83', 'F33',
       'F G73', 'E31', 'A5', 'D10 D12', 'D26', 'C110', 'B58 B60', 'E101',
       'F E69', 'D47', 'B86', 'F2', 'C2', 'E33', 'B19', 'A7', 'C49', 'F4',
       'A32', 'B4', 'B80', 'A31', 'D36', 'D15', 'C93', 'C78', 'D35',
       'C87', 'B77', 'E67', 'B94', 'C125', 'C99', 'C118', 'D7', 'A19',
       'B49', 'D', 'C22 C26', 'C106', 'C65', 'E36', 'C54',
       'B57 B59 B63 B66', 'C7', 'E34', 'C32', 'B18', 'C124', 'C91', 'E40',
       'T', 'C128', 'D37', 'B35', 'E50', 'C82', 'B96 B98', 'E10', 'E44',
       'A34', 'C104', 'C111', 'C92', 'E38', 'D21', 'E12', 'E63', 'A14',
       'B37', 'C30', 'D20', 'B79', 'E25', 'D46', 'B73', 'C95', 'B38',
       'B39', 'B22', 'C86', 'C70', 'A16', 'C101', 'C68', 'A10', 'E68',
       'B41', 'A20', 'D19', 'D50', 'D9', 'A23', 'B50', 'A26', 'D48',
       'E58', 'C126', 'B71', 'B51 B53 B55', 'D49', 'B5', 'B20', 'F G63',
       'C62 C64', 'E24', 'C90', 'C45', 'E8', 'B101', 'D45', 'C46', 'D30',
       'E121', 'D11', 'E77', 'F38', 'B3', 'D6', 'B82 B84', 'D17', 'A36',
       'B102', 'B69', 'E49', 'C47', 'D28', 'E17', 'A24', 'C50', 'B42',
       'C148'], dtype=object)
```

[132]:
```
df['Cabin'].nunique()
```

[132]:
```
147
```

[134]:
```
df['Cabin'].isnull().sum()
```

[134]:
```
687
```

[136]:
```
df['Cabin'].dropna()
```

[136]:
```
1             C85
3            C123
6             E46
10             G6
11            C103
           ...
871           D35
872    B51 B53 B55
879           C50
887           B42
889          C148
Name: Cabin, Length: 204, dtype: object
```

[138]:
```
df['Cabin'].dropna().isnull().sum()
```

[138]:
```
0
```

[140]:
```
df['Cabin'].fillna(0).isnull().sum()
```

[140]:
```
0
```

[142]:
```
df['Cabin'].fillna(0).unique()
```

[142]:
```
array([0, 'C85', 'C123', 'E46', 'G6', 'C103', 'D56', 'A6', 'C23 C25 C27',
       'B78', 'D33', 'B30', 'C52', 'B28', 'C83', 'F33', 'F G73', 'E31',
       'A5', 'D10 D12', 'D26', 'C110', 'B58 B60', 'E101', 'F E69', 'D47',
       'B86', 'F2', 'C2', 'E33', 'B19', 'A7', 'C49', 'F4', 'A32', 'B4',
       'B80', 'A31', 'D36', 'D15', 'C93', 'C78', 'D35', 'C87', 'B77',
       'E67', 'B94', 'C125', 'C99', 'C118', 'D7', 'A19', 'B49', 'D',
       'C22 C26', 'C106', 'C65', 'E36', 'C54', 'B57 B59 B63 B66', 'C7',
       'E34', 'C32', 'B18', 'C124', 'C91', 'E40', 'T', 'C128', 'D37',
       'B35', 'E50', 'C82', 'B96 B98', 'E10', 'E44', 'A34', 'C104',
       'C111', 'C92', 'E38', 'D21', 'E12', 'E63', 'A14', 'B37', 'C30',
       'D20', 'B79', 'E25', 'D46', 'B73', 'C95', 'B38', 'B39', 'B22',
       'C86', 'C70', 'A16', 'C101', 'C68', 'A10', 'E68', 'B41', 'A20',
       'D19', 'D50', 'D9', 'A23', 'B50', 'A26', 'D48', 'E58', 'C126',
       'B71', 'B51 B53 B55', 'D49', 'B5', 'B20', 'F G63', 'C62 C64',
       'E24', 'C90', 'C45', 'E8', 'B101', 'D45', 'C46', 'D30', 'E121',
       'D11', 'E77', 'F38', 'B3', 'D6', 'B82 B84', 'D17', 'A36', 'B102',
       'B69', 'E49', 'C47', 'D28', 'E17', 'A24', 'C50', 'B42', 'C148'],
      dtype=object)
```

[144]:
```
df.Cabin
```

[144]:
```
0       NaN
1       C85
2       NaN
3      C123
4       NaN
       ...
886     NaN
887     B42
888     NaN
889    C148
890     NaN
Name: Cabin, Length: 891, dtype: object
```

[148]:
```
df['Cabin'].fillna('deck',inplace=True)
```

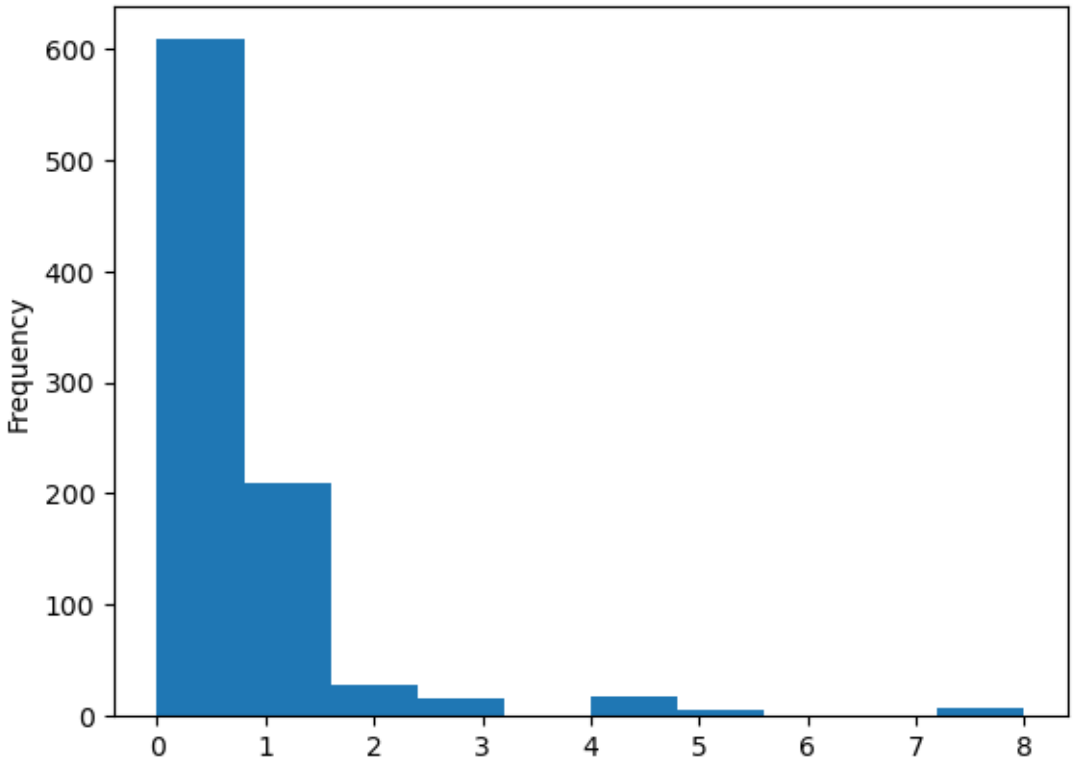
[150]:
```
df
```

[150]:
```
     PassengerId  Survived  Pclass  \
0              1         0       3
1              2         1       1
2              3         1       3
3              4         1       1
4              5         0       3
```

```
..           …       …     …
886          887         0       2
887          888         1       1
888          889         0       3
889          890         1       1
890          891         0       3

                                                  Name     Sex   Age  SibSp  \
0                              Braund, Mr. Owen Harris    male  22.0      1
1    Cumings, Mrs. John Bradley (Florence Briggs Th…  female  38.0      1
2                               Heikkinen, Miss. Laina  female  26.0      0
3         Futrelle, Mrs. Jacques Heath (Lily May Peel)  female  35.0      1
4                             Allen, Mr. William Henry    male  35.0      0
..                                                 …     …   …    …
886                             Montvila, Rev. Juozas    male  27.0      0
887                      Graham, Miss. Margaret Edith  female  19.0      0
888          Johnston, Miss. Catherine Helen "Carrie"  female   NaN      1
889                             Behr, Mr. Karl Howell    male  26.0      0
890                               Dooley, Mr. Patrick    male  32.0      0

     Parch            Ticket     Fare Cabin Embarked
0        0         A/5 21171   7.2500  deck        S
1        0          PC 17599  71.2833   C85        C
2        0  STON/O2. 3101282   7.9250  deck        S
3        0            113803  53.1000  C123        S
4        0            373450   8.0500  deck        S
..     …               …       …   …      …
886      0            211536  13.0000  deck        S
887      0            112053  30.0000   B42        S
888      2        W./C. 6607  23.4500  deck        S
889      0            111369  30.0000  C148        C
890      0            370376   7.7500  deck        Q

[891 rows x 12 columns]
```

[152]:
```
df['SibSp'].plot.hist()
```

[152]:
```
<Axes: ylabel='Frequency'>
```



[154]:
```
df.Age.skew()
```

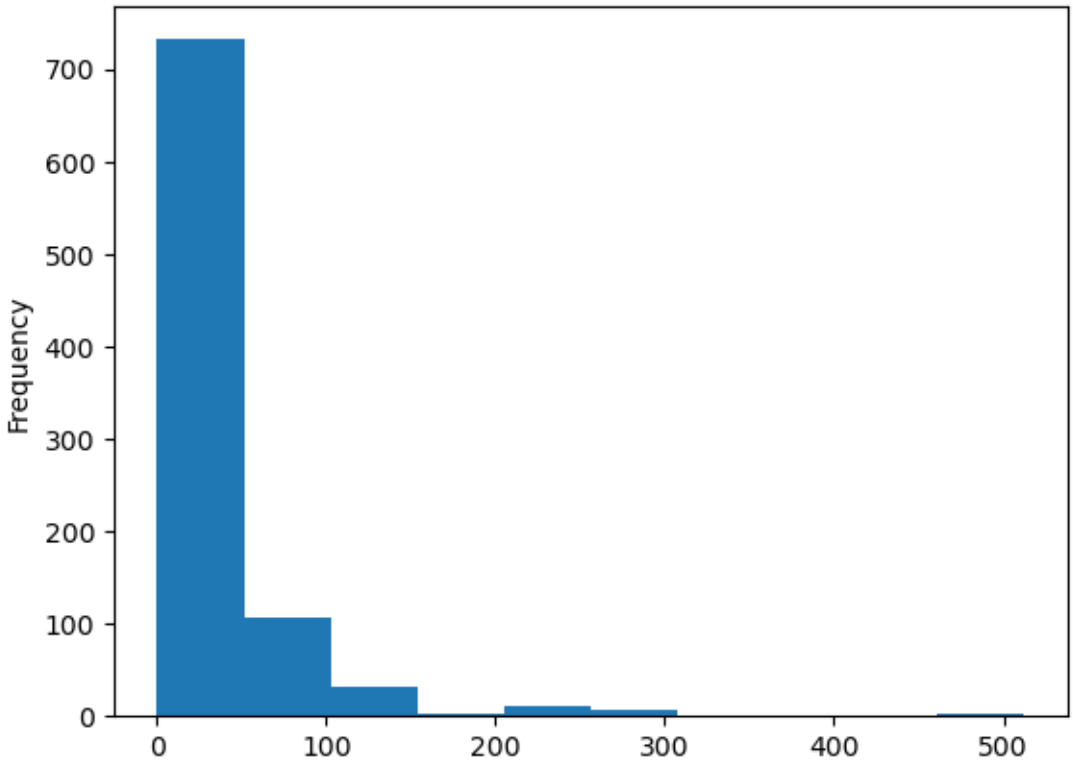
[154]: 0.38910778230082704

[156]:
```
df.SibSp.skew()
```

[156]: 3.6953517271630565

[158]:
```
df['Fare'].plot.hist()
```

[158]: <Axes: ylabel='Frequency'>



[160]:
```
df.Fare.skew()
```

[160]:
```
4.787316519674893
```

[162]:
```
df['Fare'].mean()
```

[162]:
```
32.204207968574636
```

[164]:
```
df['Age'].mean()
```

[164]:
```
29.69911764705882
```

[166]:
```
df.describe()
```

[166]:

	PassengerId	Survived	Pclass	Age	SibSp	\
count	891.000000	891.000000	891.000000	714.000000	891.000000	
mean	446.000000	0.383838	2.308642	29.699118	0.523008	
std	257.353842	0.486592	0.836071	14.526497	1.102743	
min	1.000000	0.000000	1.000000	0.420000	0.000000	
25%	223.500000	0.000000	2.000000	20.125000	0.000000	
50%	446.000000	0.000000	3.000000	28.000000	0.000000	
75%	668.500000	1.000000	3.000000	38.000000	1.000000	

```
max     891.000000    1.000000    3.000000   80.000000    8.000000
```

```
            Parch        Fare
count  891.000000  891.000000
mean     0.381594   32.204208
std      0.806057   49.693429
min      0.000000    0.000000
25%      0.000000    7.910400
50%      0.000000   14.454200
75%      0.000000   31.000000
max      6.000000  512.329200
```

[168]:
```
df.Age.mode()
```

[168]:
```
0    24.0
Name: Age, dtype: float64
```

[170]:
```
df.Fare.mode()
```

[170]:
```
0    8.05
Name: Fare, dtype: float64
```

[172]:
```
df.Fare.median()
```

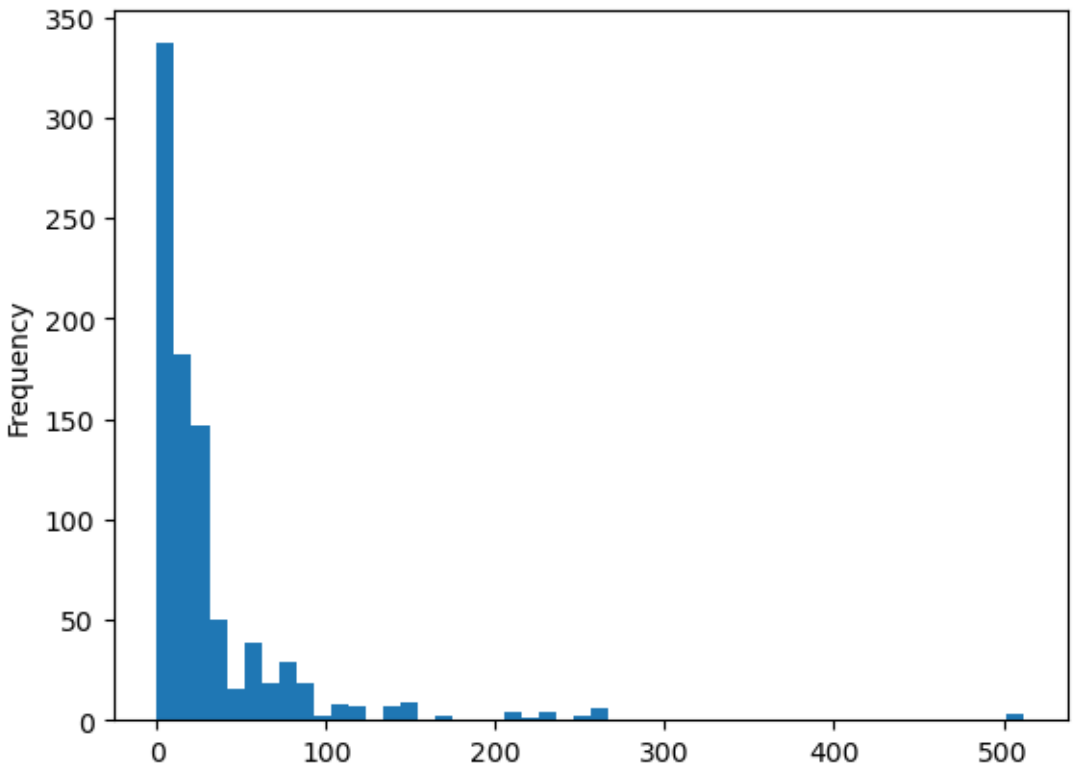
[172]:
```
14.4542
```

[174]:
```
df.Age.median()
```

[174]:
```
28.0
```

[176]:
```
df['Fare'].plot.hist(bins=50)
```

[176]:
```
<Axes: ylabel='Frequency'>
```



[178]:
```
df.loc[2:5]
```

[178]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	deck	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	deck	S
5	6	0	3	Moran, Mr. James	male	NaN	0	0	330877	8.4583	deck	Q

[180]:
```
df.iloc[2:5]
```

[180]:
```
   PassengerId  Survived  Pclass  \
2            3         1       3
3            4         1       1
4            5         0       3

                                                Name     Sex   Age  SibSp  Parch  \
2                             Heikkinen, Miss. Laina  female  26.0      0      0
3  Futrelle, Mrs. Jacques Heath (Lily May Peel)  female  35.0      1      0
4                           Allen, Mr. William Henry    male  35.0      0      0

             Ticket    Fare Cabin Embarked
2  STON/O2. 3101282   7.925  deck        S
3            113803  53.100  C123        S
4            373450   8.050  deck        S
```

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

data set 2

[9]:
```
import pandas as pd
```

[15]:
```
df=pd.read_csv(r"C:\Users\hp\Downloads\Life-Expectancy-Data-Averaged.csv")
df
```

[15]:
```
                 Country                          Region    Year  \
0            Afghanistan                            Asia  2007.5
1                Albania                  Rest of Europe  2007.5
2                Algeria                          Africa  2007.5
3                 Angola                          Africa  2007.5
4    Antigua and Barbuda  Central America and Caribbean  2007.5
..                   …                               …     …
174        Venezuela, RB                   South America  2007.5
175              Vietnam                            Asia  2007.5
176          Yemen, Rep.                     Middle East  2007.5
```

```
177              Zambia                          Africa  2007.5
178            Zimbabwe                          Africa  2007.5

     Infant_deaths  Under_five_deaths  Adult_mortality  Alcohol_consumption  \
0         71.08125           98.61250       265.804969              0.016125
1         15.25625           17.14375        83.132969              4.696875
2         26.75625           31.19375       113.439281              0.400625
3         88.76875          144.16250       297.844063              4.935625
4          9.47500           11.51875       142.478813              7.755000
..             …                  …              …                     …
174       15.85000           18.50000       151.056156              7.458125
175       19.50000           24.47500       133.054406              1.742500
176       52.44375           69.57500       239.653938              0.047000
177       60.58125           99.20625       452.761219              2.903125
178       50.16875           85.32500       572.974312              2.900625

     Hepatitis_B  Measles       BMI    Polio  Diphtheria  Incidents_HIV  \
0        64.5625  24.3750  22.46250  55.3750     55.1250       0.022500
1        98.0000  95.9375  25.85625  98.1250     98.0625       0.025625
2        88.3125  93.2500  24.86875  91.7500     91.8750       0.021875
3        68.8125  64.0000  22.51875  35.7500     55.5625       1.303750
4        98.2500  75.4375  25.85000  96.9375     98.3125       0.125000
..           …      …        …       …          …             …
174      72.4375  83.0000  26.49375  79.1875     75.2500       0.400000
175      87.4375  65.0000  20.76250  94.9375     91.7500       0.196875
176      60.1875  95.0000  23.02500  71.6250     72.6250       0.025000
177      82.8750  39.1875  22.00000  82.9375     83.8125       6.868125
178      79.5625  64.0000  23.54375  79.5625     79.1250       8.006250

     GDP_per_capita  Population_mln  Thinness_ten_nineteen_years  \
0          408.5625       27.450625                      16.58125
1         3071.1250        2.969375                       1.61875
2         3745.1250       34.820625                       6.09375
3         2647.8125       21.623750                       6.19375
4        14678.7500        0.085000                       3.42500
..              …               …                             …
174       9294.8750       27.392500                       1.65000
175       1807.8750       86.032500                      14.92500
176       2406.8125       21.730625                      13.83125
177       1059.8750       12.862500                       6.88125
178       1290.6250       12.538750                       7.01250

     Thinness_five_nine_years  Schooling  Economy_status  Life_expectancy
0                    15.58125    2.90000             1.0         59.65625
1                     1.70000    9.24375             1.0         75.95000
2                     5.97500    6.99375             1.0         73.78750
3                     6.66875    4.60625             1.0         52.82500
```

```
4                    3.37500    9.01875             1.0         75.35000
..                       …          …               …              …
174                  1.56250    8.23750             1.0         72.79375
175                 15.62500    6.90000             1.0         74.33125
176                 13.75000    2.22500             1.0         64.16250
177                  6.76250    6.43750             1.0         52.33125
178                  6.98750    7.21250             1.0         48.66250

[179 rows x 20 columns]
```

[17]:
```
df.describe()
```

[17]:
```
          Year  Infant_deaths  Under_five_deaths  Adult_mortality  \
count   179.0     179.000000         179.000000       179.000000
mean   2007.5      30.363792          42.938268       192.251775
std       0.0      26.725485          42.916952       111.659044
min    2007.5       2.381250           3.000000        57.710313
25%    2007.5       8.159375           9.775000       107.046906
50%    2007.5      19.368750          23.137500       164.432406
75%    2007.5      48.959375          68.321875       247.523922
max    2007.5     115.718750         178.725000       572.974312

       Alcohol_consumption  Hepatitis_B     Measles         BMI       Polio  \
count           179.000000   179.000000  179.000000  179.000000  179.000000
mean              4.820882    84.292598   77.344972   25.032926   86.499651
std               3.914554    13.820223   17.315208    2.165490   13.581153
min               0.000025    30.687500   16.250000   20.212500   35.750000
25%               1.317813    78.218750   64.000000   23.225000   80.531250
50%               4.209375    88.000000   83.000000   25.650000   92.375000
75%               7.843438    94.375000   92.250000   26.425000   96.062500
max              15.100000    98.875000   99.000000   31.687500   98.937500

       Diphtheria  Incidents_HIV  GDP_per_capita  Population_mln  \
count  179.000000     179.000000      179.000000      179.000000
mean    86.271648       0.894288    11540.924930       36.675915
std     13.931532       2.311895    16893.054182      136.655286
min     31.312500       0.010000      263.937500        0.085000
25%     80.812500       0.080000     1409.906250        2.108125
50%     92.062500       0.164375     4402.625000        7.660625
75%     95.781250       0.516250    12037.781250       22.745313
max     99.000000      18.164375   102972.687500     1321.239375

       Thinness_ten_nineteen_years  Thinness_five_nine_years   Schooling  \
count                   179.000000                179.000000  179.000000
mean                      4.865852                  4.899825    7.632123
std                       4.111094                  4.195663    3.126912
min                       0.100000                  0.100000    1.337500
```

```
25%                    1.756250                 1.731250    4.946875
50%                    3.556250                 3.718750    7.831250
75%                    7.165625                 7.056250   10.365625
max                   27.100000                27.943750   13.268750

       Economy_status  Life_expectancy
count      179.000000       179.000000
mean         0.793296        68.856075
std          0.406077         9.197699
min          0.000000        45.606250
25%          1.000000        62.303125
50%          1.000000        71.506250
75%          1.000000        74.937500
max          1.000000        82.456250
```

[19]:
```
df.describe(include='all')
```

[19]:
```
            Country  Region    Year  Infant_deaths  Under_five_deaths  \
count           179     179   179.0     179.000000         179.000000
unique          179       9     NaN            NaN                NaN
top     Afghanistan  Africa     NaN            NaN                NaN
freq              1      51     NaN            NaN                NaN
mean            NaN     NaN  2007.5      30.363792          42.938268
std             NaN     NaN     0.0      26.725485          42.916952
min             NaN     NaN  2007.5       2.381250           3.000000
25%             NaN     NaN  2007.5       8.159375           9.775000
50%             NaN     NaN  2007.5      19.368750          23.137500
75%             NaN     NaN  2007.5      48.959375          68.321875
max             NaN     NaN  2007.5     115.718750         178.725000

        Adult_mortality  Alcohol_consumption  Hepatitis_B     Measles  \
count        179.000000           179.000000   179.000000  179.000000
unique              NaN                  NaN          NaN         NaN
top                 NaN                  NaN          NaN         NaN
freq                NaN                  NaN          NaN         NaN
mean         192.251775             4.820882    84.292598   77.344972
std          111.659044             3.914554    13.820223   17.315208
min           57.710313             0.000025    30.687500   16.250000
25%          107.046906             1.317813    78.218750   64.000000
50%          164.432406             4.209375    88.000000   83.000000
75%          247.523922             7.843438    94.375000   92.250000
max          572.974312            15.100000    98.875000   99.000000

                BMI       Polio  Diphtheria  Incidents_HIV  GDP_per_capita  \
count   179.000000  179.000000  179.000000     179.000000      179.000000
unique         NaN         NaN         NaN            NaN             NaN
top            NaN         NaN         NaN            NaN             NaN
```

```
freq           NaN         NaN         NaN           NaN             NaN
mean     25.032926   86.499651   86.271648      0.894288    11540.924930
std       2.165490   13.581153   13.931532      2.311895    16893.054182
min      20.212500   35.750000   31.312500      0.010000      263.937500
25%      23.225000   80.531250   80.812500      0.080000     1409.906250
50%      25.650000   92.375000   92.062500      0.164375     4402.625000
75%      26.425000   96.062500   95.781250      0.516250    12037.781250
max      31.687500   98.937500   99.000000     18.164375   102972.687500

        Population_mln  Thinness_ten_nineteen_years  Thinness_five_nine_years  \
count       179.000000                   179.000000                179.000000
unique             NaN                          NaN                       NaN
top                NaN                          NaN                       NaN
freq               NaN                          NaN                       NaN
mean         36.675915                     4.865852                  4.899825
std         136.655286                     4.111094                  4.195663
min           0.085000                     0.100000                  0.100000
25%           2.108125                     1.756250                  1.731250
50%           7.660625                     3.556250                  3.718750
75%          22.745313                     7.165625                  7.056250
max        1321.239375                    27.100000                 27.943750

         Schooling  Economy_status  Life_expectancy
count   179.000000      179.000000       179.000000
unique         NaN             NaN              NaN
top            NaN             NaN              NaN
freq           NaN             NaN              NaN
mean      7.632123        0.793296        68.856075
std       3.126912        0.406077         9.197699
min       1.337500        0.000000        45.606250
25%       4.946875        1.000000        62.303125
50%       7.831250        1.000000        71.506250
75%      10.365625        1.000000        74.937500
max      13.268750        1.000000        82.456250
```

[21]:
```
df.describe(include='object')
```

[21]:
```
            Country  Region
count           179     179
unique          179       9
top     Afghanistan  Africa
freq              1      51
```

[23]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 179 entries, 0 to 178
Data columns (total 20 columns):
```

```
 #   Column                       Non-Null Count  Dtype
---  ------                       --------------  -----
 0   Country                      179 non-null    object
 1   Region                       179 non-null    object
 2   Year                         179 non-null    float64
 3   Infant_deaths                179 non-null    float64
 4   Under_five_deaths            179 non-null    float64
 5   Adult_mortality              179 non-null    float64
 6   Alcohol_consumption          179 non-null    float64
 7   Hepatitis_B                  179 non-null    float64
 8   Measles                      179 non-null    float64
 9   BMI                          179 non-null    float64
 10  Polio                        179 non-null    float64
 11  Diphtheria                   179 non-null    float64
 12  Incidents_HIV                179 non-null    float64
 13  GDP_per_capita               179 non-null    float64
 14  Population_mln               179 non-null    float64
 15  Thinness_ten_nineteen_years  179 non-null    float64
 16  Thinness_five_nine_years     179 non-null    float64
 17  Schooling                    179 non-null    float64
 18  Economy_status               179 non-null    float64
 19  Life_expectancy              179 non-null    float64
dtypes: float64(18), object(2)
memory usage: 28.1+ KB
```

[25]:
```
df.head()
```

[25]:

	Country	Region	Year	Infant_deaths	Under_five_deaths	Adult_mortality	Alcohol_consumption	Hepatitis_B	Measles	BMI	Polio	Diphtheria	Incidents_HIV	GDP_per_capita
0	Afghanistan	Asia	2007.5	71.08125	98.61250	265.804969	0.016125	64.5625	24.3750	22.46250	55.3750	55.1250	0.022500	408.5625
1	Albania	Rest of Europe	2007.5	15.25625	17.14375	83.132969	4.696875	98.0000	95.9375	25.85625	98.1250	98.0625	0.025625	3071.1250
2	Algeria	Africa	2007.5	26.75625	31.19375	113.439281	0.400625	88.3125	93.2500	24.86875	91.7500	91.8750	0.021875	3745.1250
3	Angola	Africa	2007.5	88.76875	144.16250	297.844063	4.935625	68.8125	64.0000	22.51875	35.7500	55.5625	1.303750	2647.8125
4	Antigua and Barbuda	Central America and Caribbean	2007.5	9.47500	11.51875	142.478813	7.755000	98.2500	75.4375	25.85000	96.9375	98.3125	0.125000	14678.7500

```
   Population_mln  Thinness_ten_nineteen_years  Thinness_five_nine_years  \
0       27.450625                     16.58125                  15.58125
1        2.969375                      1.61875                   1.70000
2       34.820625                      6.09375                   5.97500
3       21.623750                      6.19375                   6.66875
4        0.085000                      3.42500                   3.37500

   Schooling  Economy_status  Life_expectancy
0    2.90000             1.0         59.65625
1    9.24375             1.0         75.95000
2    6.99375             1.0         73.78750
3    4.60625             1.0         52.82500
4    9.01875             1.0         75.35000
```

[27]:
```
df.columns
```

[27]:
```
Index(['Country', 'Region', 'Year', 'Infant_deaths', 'Under_five_deaths',
       'Adult_mortality', 'Alcohol_consumption', 'Hepatitis_B', 'Measles',
       'BMI', 'Polio', 'Diphtheria', 'Incidents_HIV', 'GDP_per_capita',
       'Population_mln', 'Thinness_ten_nineteen_years',
       'Thinness_five_nine_years', 'Schooling', 'Economy_status',
       'Life_expectancy'],
      dtype='object')
```

[29]:
```
#view first 5 columns
columns=df.iloc[0:5]
columns
```

[29]:
```
                Country                          Region    Year  Infant_deaths  \
0           Afghanistan                            Asia  2007.5       71.08125
1               Albania                  Rest of Europe  2007.5       15.25625
2               Algeria                          Africa  2007.5       26.75625
3                Angola                          Africa  2007.5       88.76875
4  Antigua and Barbuda  Central America and Caribbean  2007.5        9.47500

   Under_five_deaths  Adult_mortality  Alcohol_consumption  Hepatitis_B  \
0           98.61250       265.804969             0.016125      64.5625
1           17.14375        83.132969             4.696875      98.0000
2           31.19375       113.439281             0.400625      88.3125
3          144.16250       297.844063             4.935625      68.8125
4           11.51875       142.478813             7.755000      98.2500

   Measles       BMI    Polio  Diphtheria  Incidents_HIV  GDP_per_capita  \
0  24.3750  22.46250  55.3750     55.1250       0.022500        408.5625
1  95.9375  25.85625  98.1250     98.0625       0.025625       3071.1250
2  93.2500  24.86875  91.7500     91.8750       0.021875       3745.1250
```

```
3  64.0000  22.51875  35.7500     55.5625       1.303750       2647.8125
4  75.4375  25.85000  96.9375     98.3125       0.125000      14678.7500

   Population_mln  Thinness_ten_nineteen_years  Thinness_five_nine_years  \
0       27.450625                     16.58125                  15.58125
1        2.969375                      1.61875                   1.70000
2       34.820625                      6.09375                   5.97500
3       21.623750                      6.19375                   6.66875
4        0.085000                      3.42500                   3.37500

   Schooling  Economy_status  Life_expectancy
0    2.90000             1.0         59.65625
1    9.24375             1.0         75.95000
2    6.99375             1.0         73.78750
3    4.60625             1.0         52.82500
4    9.01875             1.0         75.35000
```

[31]:
```
df.isnull()
```

[31]:
```
     Country  Region   Year  Infant_deaths  Under_five_deaths  \
0      False   False  False          False              False
1      False   False  False          False              False
2      False   False  False          False              False
3      False   False  False          False              False
4      False   False  False          False              False
..       …     …    …             …                 …
174    False   False  False          False              False
175    False   False  False          False              False
176    False   False  False          False              False
177    False   False  False          False              False
178    False   False  False          False              False

     Adult_mortality  Alcohol_consumption  Hepatitis_B  Measles    BMI  Polio  \
0              False                False        False    False  False  False
1              False                False        False    False  False  False
2              False                False        False    False  False  False
3              False                False        False    False  False  False
4              False                False        False    False  False  False
..               …                  …          …      …    …    …
174            False                False        False    False  False  False
175            False                False        False    False  False  False
176            False                False        False    False  False  False
177            False                False        False    False  False  False
178            False                False        False    False  False  False

     Diphtheria  Incidents_HIV  GDP_per_capita  Population_mln  \
0         False          False           False           False
```

```
1         False          False           False           False
2         False          False           False           False
3         False          False           False           False
4         False          False           False           False
..          …              …               …               …
174       False          False           False           False
175       False          False           False           False
176       False          False           False           False
177       False          False           False           False
178       False          False           False           False

     Thinness_ten_nineteen_years  Thinness_five_nine_years  Schooling  \
0                          False                     False      False
1                          False                     False      False
2                          False                     False      False
3                          False                     False      False
4                          False                     False      False
..                           …                         …          …
174                        False                     False      False
175                        False                     False      False
176                        False                     False      False
177                        False                     False      False
178                        False                     False      False

     Economy_status  Life_expectancy
0             False            False
1             False            False
2             False            False
3             False            False
4             False            False
..              …                …
174           False            False
175           False            False
176           False            False
177           False            False
178           False            False

[179 rows x 20 columns]
```

[33]:
```
df.isnull().sum()
```

[33]:
```
Country                       0
Region                        0
Year                          0
Infant_deaths                 0
Under_five_deaths             0
Adult_mortality               0
```

```
Alcohol_consumption              0
Hepatitis_B                      0
Measles                          0
BMI                              0
Polio                            0
Diphtheria                       0
Incidents_HIV                    0
GDP_per_capita                   0
Population_mln                   0
Thinness_ten_nineteen_years      0
Thinness_five_nine_years         0
Schooling                        0
Economy_status                   0
Life_expectancy                  0
dtype: int64
```

[35]:

```
#Clean the Data
df.dropna(inplace=True)
```

[37]:

```
df.shape
```

[37]:

```
(179, 20)
```

[39]:

```
# previewing dataset
df.sample(5)
```

[39]:

```
           Country          Region    Year  Infant_deaths  Under_five_deaths  \
146        Somalia          Africa  2007.5       98.25625          161.26875
68   Guinea-Bissau          Africa  2007.5       81.41875          130.56875
18          Bhutan            Asia  2007.5       40.34375           51.53750
116    New Zealand         Oceania  2007.5        5.30000            6.38750
62         Germany  European Union  2007.5        3.73125            4.51875

     Adult_mortality  Alcohol_consumption  Hepatitis_B  Measles       BMI  \
146       333.185125             0.000025      42.0000  64.0000  22.62500
68        291.873906             3.105625      84.0000  64.0000  22.79375
18        243.186219             0.481875      94.1250  82.0625  23.02500
116        74.433875             9.145000      90.4375  86.0000  27.39375
62         81.360344            11.647500      87.5625  89.3125  26.18750

       Polio  Diphtheria  Incidents_HIV  GDP_per_capita  Population_mln  \
146  40.5000     37.7500       0.130000        263.9375       11.271250
68   74.1875     73.1250       2.608125        576.5000        1.443125
18   94.4375     93.8750       0.181250       1882.6250        0.665000
116  89.1250     90.8750       0.031875      35167.5000        4.225625
62   94.5625     94.9375       0.061250      37390.8125       81.814375
```

```
     Thinness_ten_nineteen_years  Thinness_five_nine_years  Schooling  \
146                      7.26875                   7.08125    2.10000
68                       7.50000                   7.38125    2.50000
18                      17.21875                  17.92500    2.44375
116                      0.31250                   0.30000   11.85625
62                       1.11875                   1.10625   13.26875

     Economy_status  Life_expectancy
146             1.0         53.24375
68              1.0         53.52500
18              1.0         66.45625
116             0.0         80.18750
62              0.0         79.47500
```

[41]:
```
df.dtypes
```

[41]:
```
Country                          object
Region                           object
Year                            float64
Infant_deaths                   float64
Under_five_deaths               float64
Adult_mortality                 float64
Alcohol_consumption             float64
Hepatitis_B                     float64
Measles                         float64
BMI                             float64
Polio                           float64
Diphtheria                      float64
Incidents_HIV                   float64
GDP_per_capita                  float64
Population_mln                  float64
Thinness_ten_nineteen_years     float64
Thinness_five_nine_years        float64
Schooling                       float64
Economy_status                  float64
Life_expectancy                 float64
dtype: object
```

[43]:
```
# view first row
row=df.iloc[0]
row
```

[43]:
```
Country                     Afghanistan
Region                             Asia
Year                             2007.5
Infant_deaths                   71.08125
Under_five_deaths                98.6125
```

```
Adult_mortality                 265.804969
Alcohol_consumption               0.016125
Hepatitis_B                       64.5625
Measles                           24.375
BMI                               22.4625
Polio                             55.375
Diphtheria                        55.125
Incidents_HIV                     0.0225
GDP_per_capita                   408.5625
Population_mln                  27.450625
Thinness_ten_nineteen_years       16.58125
Thinness_five_nine_years          15.58125
Schooling                              2.9
Economy_status                         1.0
Life_expectancy                   59.65625
Name: 0, dtype: object
```

[45]:
```
#view first 5 row
row=df.iloc[0:5]
row
```

[45]:

	Country	Region	Year	Infant_deaths	Under_five_deaths	Adult_mortality	Alcohol_consumption	Hepatitis_B	Measles	BMI	Polio	Diphtheria	Incidents_HIV	GDP_per_capita	Population_mln	Thinness_ten_nineteen_years	Thinness_five_nine_years
0	Afghanistan	Asia	2007.5	71.08125	98.61250	265.804969	0.016125	64.5625	24.3750	22.46250	55.3750	55.1250	0.022500	408.5625	27.450625	16.58125	15.58125
1	Albania	Rest of Europe	2007.5	15.25625	17.14375	83.132969	4.696875	98.0000	95.9375	25.85625	98.1250	98.0625	0.025625	3071.1250	2.969375	1.61875	1.70000
2	Algeria	Africa	2007.5	26.75625	31.19375	113.439281	0.400625	88.3125	93.2500	24.86875	91.7500	91.8750	0.021875	3745.1250	34.820625	6.09375	5.97500
3	Angola	Africa	2007.5	88.76875	144.16250	297.844063	4.935625	68.8125	64.0000	22.51875	35.7500	55.5625	1.303750	2647.8125	21.623750	6.19375	6.66875
4	Antigua and Barbuda	Central America and Caribbean	2007.5	9.47500	11.51875	142.478813	7.755000	98.2500	75.4375	25.85000	96.9375	98.3125	0.125000	14678.7500			

```
4        0.085000                      3.42500                    3.37500

   Schooling  Economy_status  Life_expectancy
0    2.90000             1.0         59.65625
1    9.24375             1.0         75.95000
2    6.99375             1.0         73.78750
3    4.60625             1.0         52.82500
4    9.01875             1.0         75.35000
```

[47]:
```
df['Measles'].value_counts()
```

[47]:
```
Measles
64.0000    30
83.0000    12
65.0000    10
86.0000     5
98.1875     2
           ..
93.3750     1
88.0000     1
83.0625     1
88.9375     1
39.1875     1
Name: count, Length: 115, dtype: int64
```

[49]:
```
df['Measles'].unique()
```

[49]:
```
array([24.375 , 95.9375, 93.25  , 64.    , 75.4375, 84.6875, 94.6875,
       87.25  , 60.9375, 79.5625, 83.    , 98.1875, 66.5625, 84.5625,
       97.75  , 80.9375, 90.9375, 82.0625, 89.1875, 70.    , 94.75  ,
       90.4375, 31.5   , 57.375 , 94.3125, 49.75  , 89.625 , 93.875 ,
       79.25  , 86.5   , 97.6875, 88.0625, 98.0625, 87.125 , 82.    ,
       88.9375, 83.0625, 88.    , 94.125 , 93.375 , 89.8125, 92.    ,
       67.625 , 75.75  , 89.3125, 56.25  , 78.5   , 79.4375, 82.6875,
       99.    , 92.3125, 38.8125, 47.75  , 65.    , 62.0625, 86.    ,
       94.1875, 83.4375, 82.3125, 88.8125, 92.75  , 98.25  , 28.    ,
       47.    , 90.75  , 94.25  , 52.5   , 70.4375, 93.4375, 81.5625,
       72.75  , 84.1875, 92.8125, 95.625 , 95.6875, 68.9375, 94.    ,
       97.    , 91.9375, 36.625 , 83.1875, 66.    , 53.5625, 27.6875,
       95.4375, 85.    , 94.4375, 92.875 , 95.5   , 87.    , 49.    ,
       71.3125, 97.25  , 36.25  , 88.5625, 98.5625, 60.    , 94.8125,
       53.625 , 92.1875, 93.0625, 16.25  , 78.0625, 73.1875, 90.375 ,
       92.625 , 97.625 , 81.4375, 97.8125, 96.1875, 78.9375, 92.5625,
       83.5625, 95.    , 39.1875])
```

[54]:
```
Measles=df['Measles']
```

[56]:
```
df['Measles'].nunique()
```

[56]:
```
115
```

[58]:
```
df['Measles'].unique()
```

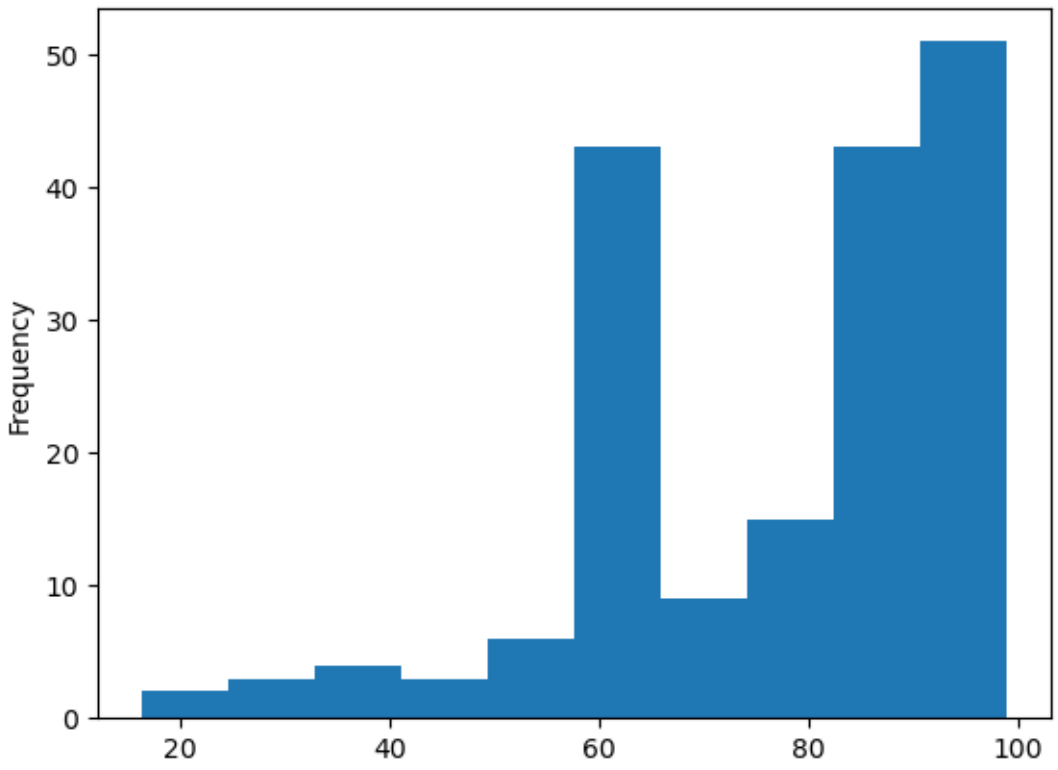
[58]:
```
array([24.375 , 95.9375, 93.25  , 64.    , 75.4375, 84.6875, 94.6875,
       87.25  , 60.9375, 79.5625, 83.    , 98.1875, 66.5625, 84.5625,
       97.75  , 80.9375, 90.9375, 82.0625, 89.1875, 70.    , 94.75  ,
       90.4375, 31.5   , 57.375 , 94.3125, 49.75  , 89.625 , 93.875 ,
       79.25  , 86.5   , 97.6875, 88.0625, 98.0625, 87.125 , 82.    ,
       88.9375, 83.0625, 88.    , 94.125 , 93.375 , 89.8125, 92.    ,
       67.625 , 75.75  , 89.3125, 56.25  , 78.5   , 79.4375, 82.6875,
       99.    , 92.3125, 38.8125, 47.75  , 65.    , 62.0625, 86.    ,
       94.1875, 83.4375, 82.3125, 88.8125, 92.75  , 98.25  , 28.    ,
       47.    , 90.75  , 94.25  , 52.5   , 70.4375, 93.4375, 81.5625,
       72.75  , 84.1875, 92.8125, 95.625 , 95.6875, 68.9375, 94.    ,
       97.    , 91.9375, 36.625 , 83.1875, 66.    , 53.5625, 27.6875,
       95.4375, 85.    , 94.4375, 92.875 , 95.5   , 87.    , 49.    ,
       71.3125, 97.25  , 36.25  , 88.5625, 98.5625, 60.    , 94.8125,
       53.625 , 92.1875, 93.0625, 16.25  , 78.0625, 73.1875, 90.375 ,
       92.625 , 97.625 , 81.4375, 97.8125, 96.1875, 78.9375, 92.5625,
       83.5625, 95.    , 39.1875])
```

[62]:
```
df['Measles'].value_counts(normalize=True)
```

[62]:
```
Measles
64.0000    0.167598
83.0000    0.067039
65.0000    0.055866
86.0000    0.027933
98.1875    0.011173
             …
93.3750    0.005587
88.0000    0.005587
83.0625    0.005587
88.9375    0.005587
39.1875    0.005587
Name: proportion, Length: 115, dtype: float64
```

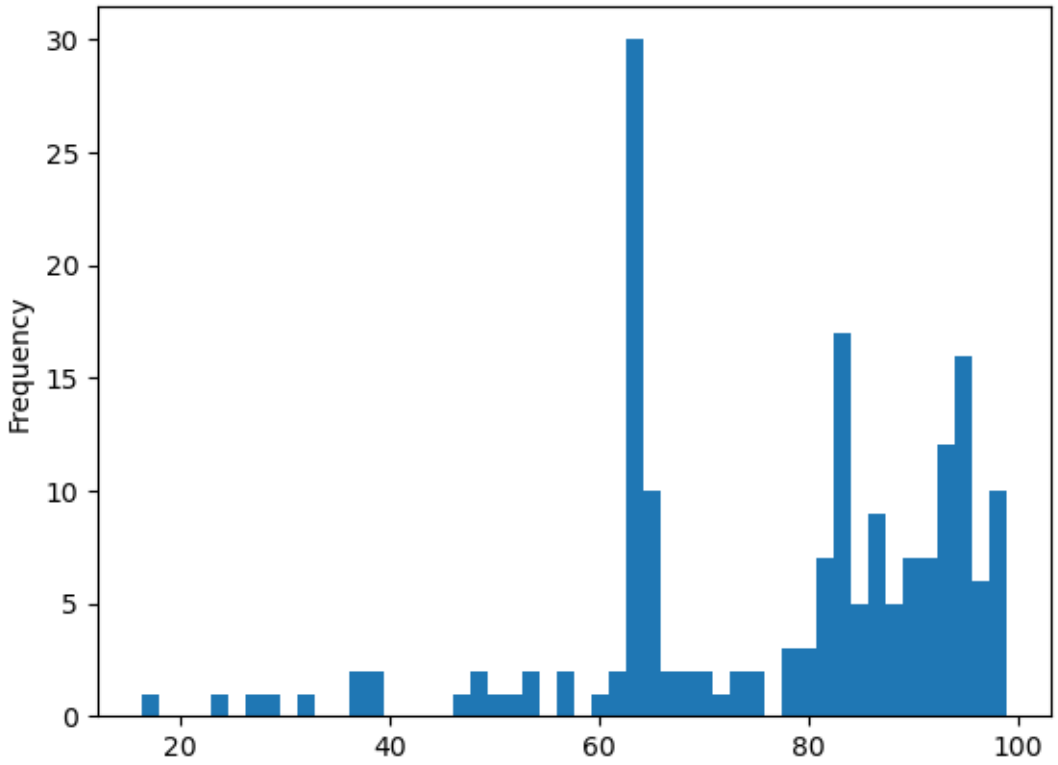
[64]:
```
df['Measles'].plot.hist()
```

[64]:
```
<Axes: ylabel='Frequency'>
```



[66]:
```
df['Measles'].plot.hist(bins=50)
```

[66]: <Axes: ylabel='Frequency'>



[184]:

```
---------------------------------------------------------------------------
AttributeError                            Traceback (most recent call last)
~\AppData\Local\Temp\ipykernel_15340\4185571416.py in ?()
----> 1 df.Year.skew()

~\anaconda3\Lib\site-packages\pandas\core\generic.py in ?(self, name)
   6295             and name not in self._accessors
   6296             and self._info_axis.
 ↪_can_hold_identifiers_and_holds_name(name)
   6297         ):
   6298             return self[name]
-> 6299         return object.__getattribute__(self, name)

AttributeError: 'DataFrame' object has no attribute 'Year'
```

[ ]: