

ps9utqk0a

July 31, 2023

```
[ ]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
[ ]: df=pd.read_csv("/content/16_Sleep_health_and_lifestyle_dataset.csv")
df
```

```
[ ]:
    Person ID  Gender  Age      Occupation  Sleep Duration \
0           1   Male   27   Software Engineer           6.1
1           2   Male   28           Doctor           6.2
2           3   Male   28           Doctor           6.2
3           4   Male   28  Sales Representative           5.9
4           5   Male   28  Sales Representative           5.9
..          ...   ...   ...                ...         ...
369         370  Female  59           Nurse           8.1
370         371  Female  59           Nurse           8.0
371         372  Female  59           Nurse           8.1
372         373  Female  59           Nurse           8.1
373         374  Female  59           Nurse           8.1

    Quality of Sleep  Physical Activity Level  Stress Level BMI Category \
0                   6                   42           6  Overweight
1                   6                   60           8    Normal
2                   6                   60           8    Normal
3                   4                   30           8    Obese
4                   4                   30           8    Obese
..                  ...                   ...         ...
369                  9                   75           3  Overweight
370                  9                   75           3  Overweight
371                  9                   75           3  Overweight
372                  9                   75           3  Overweight
373                  9                   75           3  Overweight

    Blood Pressure  Heart Rate  Daily Steps  Sleep Disorder
0          126/83         77        4200         None
1          125/80         75       10000         None
```

| | | | | |
|-----|--------|-----|-------|-------------|
| 2 | 125/80 | 75 | 10000 | None |
| 3 | 140/90 | 85 | 3000 | Sleep Apnea |
| 4 | 140/90 | 85 | 3000 | Sleep Apnea |
| .. | ... | ... | ... | ... |
| 369 | 140/95 | 68 | 7000 | Sleep Apnea |
| 370 | 140/95 | 68 | 7000 | Sleep Apnea |
| 371 | 140/95 | 68 | 7000 | Sleep Apnea |
| 372 | 140/95 | 68 | 7000 | Sleep Apnea |
| 373 | 140/95 | 68 | 7000 | Sleep Apnea |

[374 rows x 13 columns]

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

```
[ ]:
  Person ID Gender Age      Occupation Sleep Duration \
0         1  Male  27  Software Engineer          6.1
1         2  Male  28           Doctor          6.2
2         3  Male  28           Doctor          6.2
3         4  Male  28  Sales Representative          5.9
4         5  Male  28  Sales Representative          5.9

  Quality of Sleep Physical Activity Level Stress Level BMI Category \
0                 6                   42           6  Overweight
1                 6                   60           8    Normal
2                 6                   60           8    Normal
3                 4                   30           8     Obese
4                 4                   30           8     Obese

  Blood Pressure Heart Rate Daily Steps Sleep Disorder
0      126/83         77      4200      None
1      125/80         75     10000      None
2      125/80         75     10000      None
3      140/90         85       3000  Sleep Apnea
4      140/90         85       3000  Sleep Apnea
```

1 DATA CLEANING AND DATA PREPROCESSING

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

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 374 entries, 0 to 373
Data columns (total 13 columns):
#   Column              Non-Null Count  Dtype
---  -
0   Person ID           374 non-null   int64
1   Gender              374 non-null   object
2   Age                 374 non-null   int64
```

```

3 Occupation 374 non-null object
4 Sleep Duration 374 non-null float64
5 Quality of Sleep 374 non-null int64
6 Physical Activity Level 374 non-null int64
7 Stress Level 374 non-null int64
8 BMI Category 374 non-null object
9 Blood Pressure 374 non-null object
10 Heart Rate 374 non-null int64
11 Daily Steps 374 non-null int64
12 Sleep Disorder 374 non-null object
dtypes: float64(1), int64(7), object(5)
memory usage: 38.1+ KB

```

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

```

[ ]:
count      Person ID      Age  Sleep Duration  Quality of Sleep \
mean    187.500000    42.184492      7.132086      7.312834
std     108.108742      8.673133      0.795657      1.196956
min       1.000000    27.000000      5.800000      4.000000
25%      94.250000    35.250000      6.400000      6.000000
50%     187.500000    43.000000      7.200000      7.000000
75%     280.750000    50.000000      7.800000      8.000000
max      374.000000    59.000000      8.500000      9.000000

count      Physical Activity Level  Stress Level  Heart Rate  Daily Steps
mean           59.171123      5.385027    70.165775    6816.844920
std           20.830804      1.774526     4.135676    1617.915679
min           30.000000      3.000000    65.000000    3000.000000
25%           45.000000      4.000000    68.000000    5600.000000
50%           60.000000      5.000000    70.000000    7000.000000
75%           75.000000      7.000000    72.000000    8000.000000
max           90.000000      8.000000    86.000000   10000.000000

```

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

```

[ ]: Index(['Person ID', 'Gender', 'Age', 'Occupation', 'Sleep Duration',
        'Quality of Sleep', 'Physical Activity Level', 'Stress Level',
        'BMI Category', 'Blood Pressure', 'Heart Rate', 'Daily Steps',
        'Sleep Disorder'],
        dtype='object')

```

```

[ ]: df1=df.dropna(axis=1)
df1

```

```
[ ]:      Person ID  Gender  Age      Occupation  Sleep Duration \
0          1      Male   27      Software Engineer      6.1
1          2      Male   28              Doctor      6.2
2          3      Male   28              Doctor      6.2
3          4      Male   28  Sales Representative      5.9
4          5      Male   28  Sales Representative      5.9
..      ...      ...      ...      ...      ...
369       370  Female   59              Nurse      8.1
370       371  Female   59              Nurse      8.0
371       372  Female   59              Nurse      8.1
372       373  Female   59              Nurse      8.1
373       374  Female   59              Nurse      8.1

      Quality of Sleep  Physical Activity Level  Stress Level BMI Category \
0          6          42          6  Overweight
1          6          60          8    Normal
2          6          60          8    Normal
3          4          30          8    Obese
4          4          30          8    Obese
..      ...      ...      ...      ...
369          9          75          3  Overweight
370          9          75          3  Overweight
371          9          75          3  Overweight
372          9          75          3  Overweight
373          9          75          3  Overweight

      Blood Pressure  Heart Rate  Daily Steps  Sleep Disorder
0          126/83          77          4200          None
1          125/80          75         10000          None
2          125/80          75         10000          None
3          140/90          85          3000  Sleep Apnea
4          140/90          85          3000  Sleep Apnea
..      ...      ...      ...      ...
369       140/95          68          7000  Sleep Apnea
370       140/95          68          7000  Sleep Apnea
371       140/95          68          7000  Sleep Apnea
372       140/95          68          7000  Sleep Apnea
373       140/95          68          7000  Sleep Apnea

[374 rows x 13 columns]
```

```
[ ]: df1.columns
```

```
[ ]: Index(['Person ID', 'Gender', 'Age', 'Occupation', 'Sleep Duration',
          'Quality of Sleep', 'Physical Activity Level', 'Stress Level',
          'BMI Category', 'Blood Pressure', 'Heart Rate', 'Daily Steps',
          'Sleep Disorder'],
```

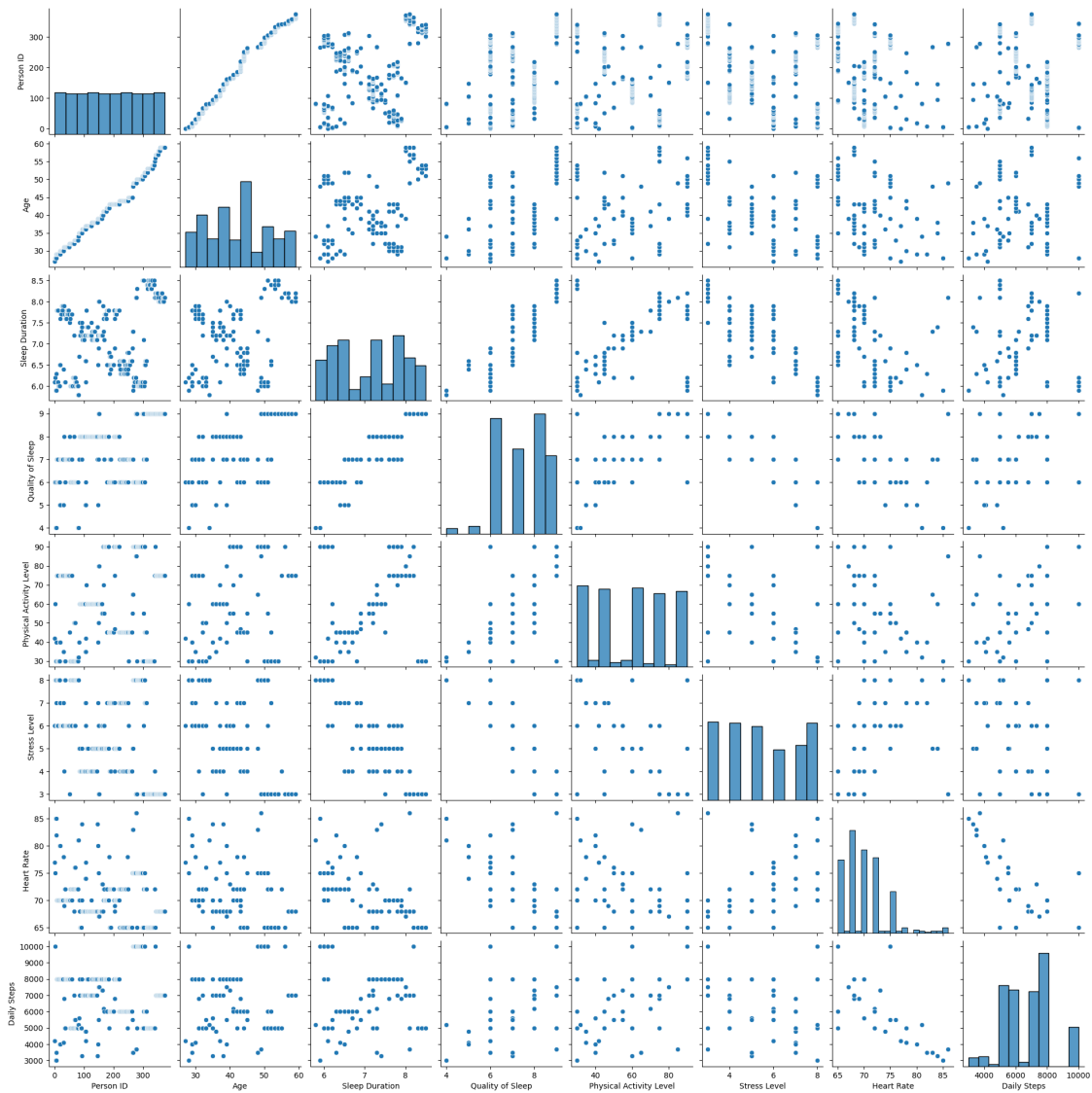
```
dtype='object')
```

```
[ ]: df1=df1[['Person ID', 'Gender', 'Age', 'Occupation', 'Sleep Duration',  
            'Quality of Sleep', 'Physical Activity Level', 'Stress Level',  
            'BMI Category', 'Blood Pressure', 'Heart Rate', 'Daily Steps',  
            'Sleep Disorder']]
```

2 EDA AND VISUALIZATION

```
[ ]: sns.pairplot(df1)
```

```
[ ]: <seaborn.axisgrid.PairGrid at 0x7c9392df1630>
```



```
[ ]: sns.distplot(df1[ 'Daily Steps'])
```

<ipython-input-13-5d148120d475>:1: UserWarning:

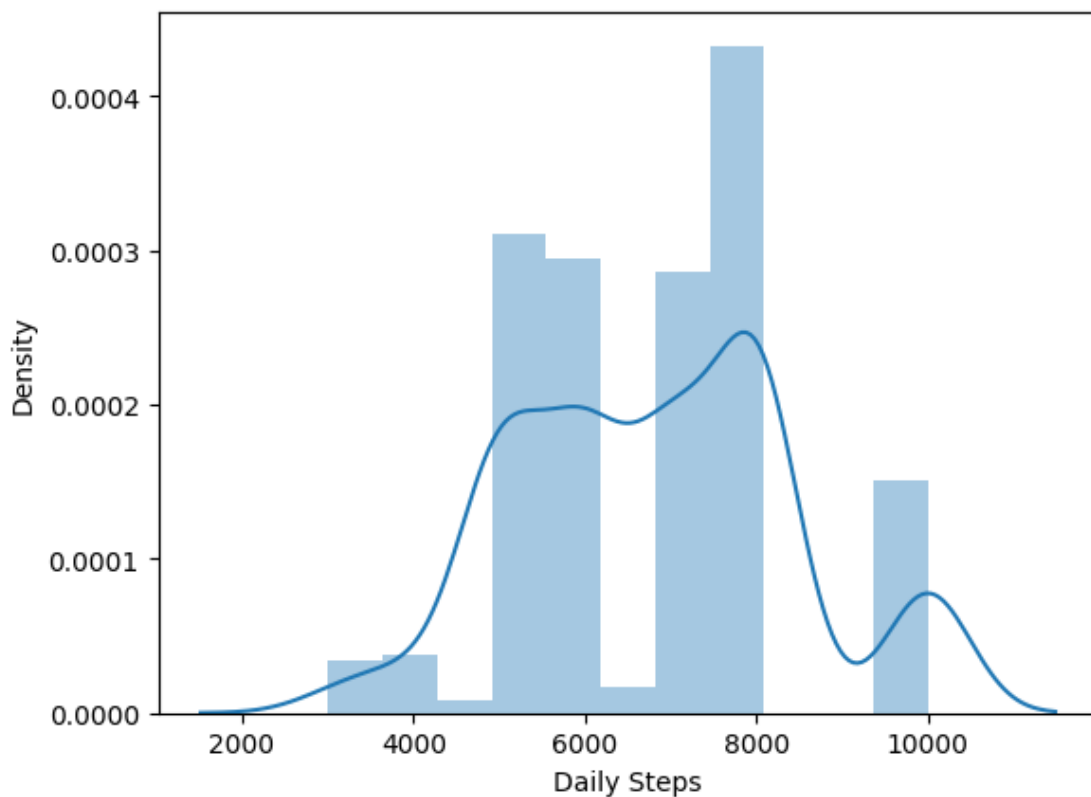
`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see <https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751>

```
sns.distplot(df1[ 'Daily Steps'])
```

```
[ ]: <Axes: xlabel='Daily Steps', ylabel='Density'>
```

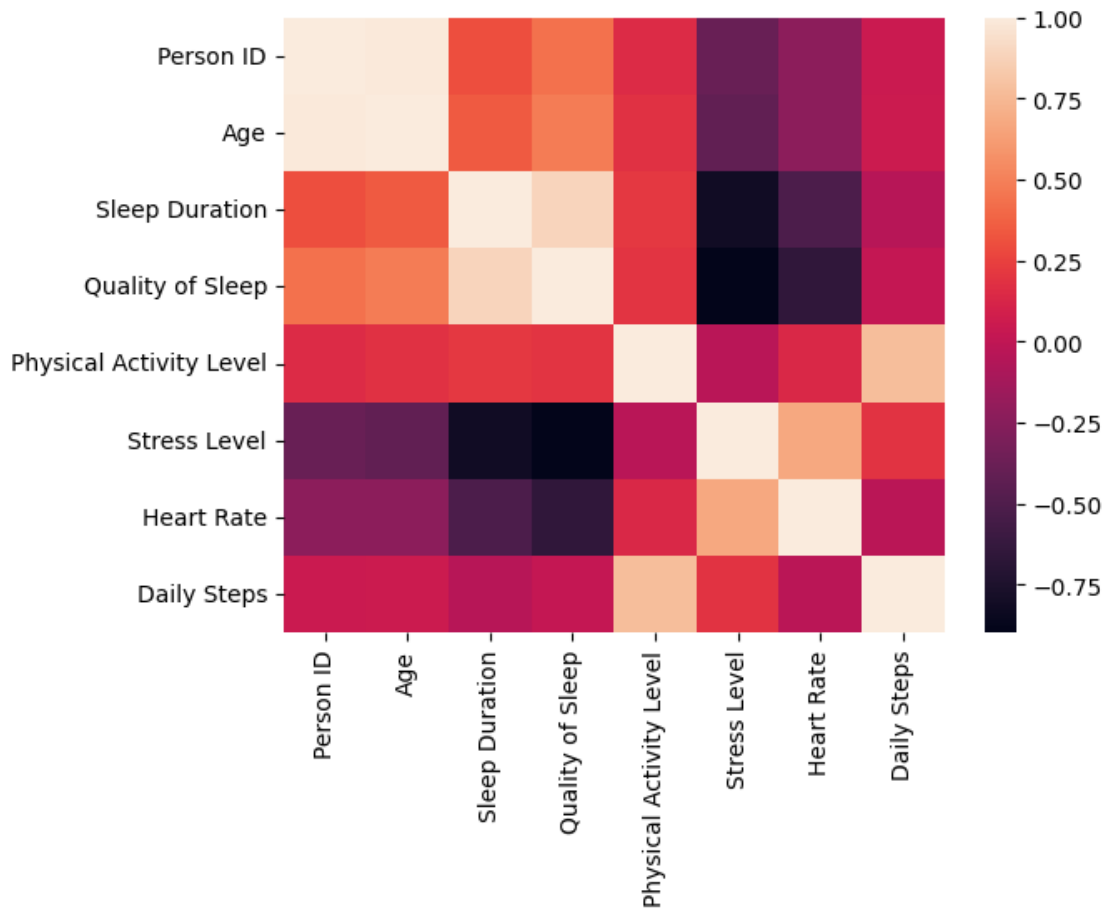


```
[ ]: sns.heatmap(df1.corr())
```

<ipython-input-14-3ed1a1a51dc0>:1: FutureWarning: The default value of numeric_only in DataFrame.corr is deprecated. In a future version, it will default to False. Select only valid columns or specify the value of numeric_only to silence this warning.

```
sns.heatmap(df1.corr())
```

```
[ ]: <Axes: >
```



3 TO TRAIN THE MODEL AND MODEL BUILDING

```
[ ]: x=df[['Person ID', 'Age', 'Sleep Duration',
          'Quality of Sleep', 'Physical Activity Level', 'Stress Level', 'Heart_Rate']]
      y=df['Daily Steps']
```

```
[ ]: from sklearn.model_selection import train_test_split
      x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.3)
```

```
[ ]: from sklearn.linear_model import LinearRegression
      lr = LinearRegression()
      lr.fit(x_train, y_train)
```

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

```
[ ]: lr.intercept_
```

```
[ ]: 12909.224743415758
```

```
[ ]: coeff=pd.DataFrame(lr.coef_,x.columns,columns=['Co-efficient'])  
coeff
```

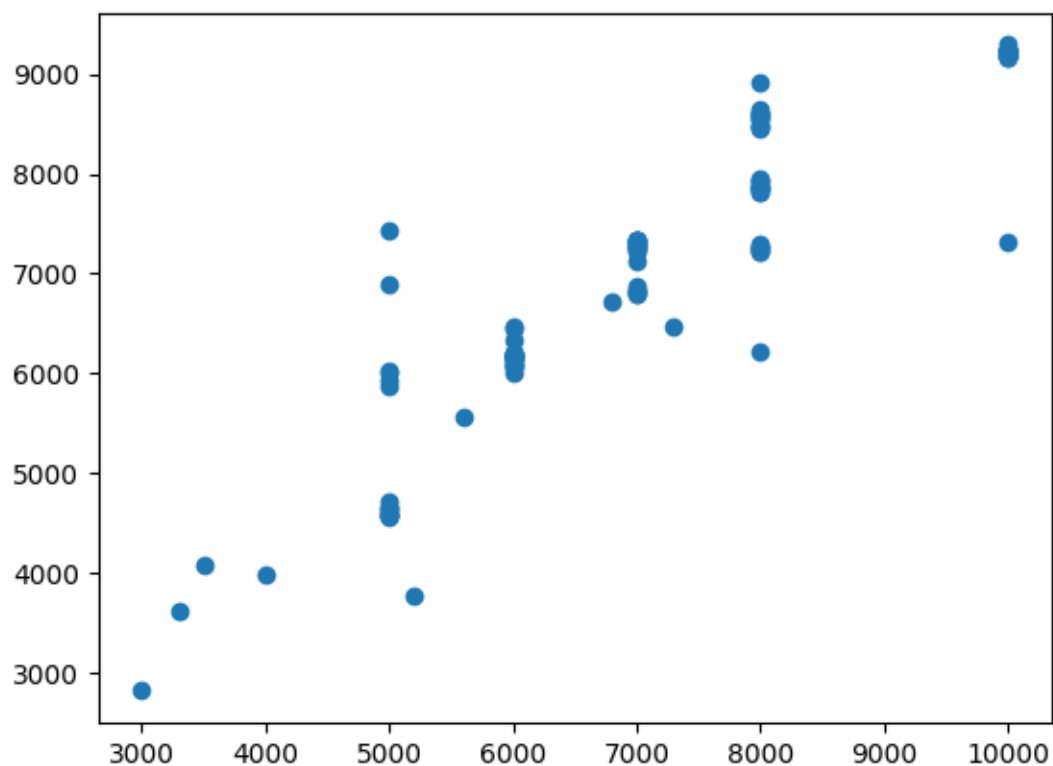
```
[ ]: 

|                         | Co-efficient |
|-------------------------|--------------|
| Person ID               | -5.855784    |
| Age                     | 71.307044    |
| Sleep Duration          | -441.798866  |
| Quality of Sleep        | 329.766415   |
| Physical Activity Level | 65.446192    |
| Stress Level            | 524.938216   |
| Heart Rate              | -199.252235  |


```

```
[ ]: prediction =lr.predict(x_test)  
plt.scatter(y_test,prediction)
```

```
[ ]: <matplotlib.collections.PathCollection at 0x7c9388b1e890>
```



4 ACCURACY

```
[ ]: lr.score(x_test,y_test)
```

```
[ ]: 0.848370364859215
```

```
[ ]: lr.score(x_train,y_train)
```

```
[ ]: 0.775766191963639
```

```
[ ]: from sklearn.linear_model import Ridge,Lasso
```

```
[ ]: rr=Ridge(alpha=10)  
    rr.fit(x_train,y_train)
```

```
[ ]: Ridge(alpha=10)
```

```
[ ]: rr.score(x_test,y_test)
```

```
[ ]: 0.846077109415987
```

```
[ ]: rr.score(x_train,y_train)
```

```
[ ]: 0.7744545191998512
```

```
[ ]: la=Lasso(alpha=10)  
    la.fit(x_train,y_train)
```

```
[ ]: Lasso(alpha=10)
```

```
[ ]: la.score(x_test,y_test)
```

```
[ ]: 0.8468484246428218
```

```
[ ]: la.score(x_train,y_train)
```

```
[ ]: 0.7746253958822812
```

```
[ ]: from sklearn.linear_model import ElasticNet  
    en=ElasticNet()  
    en.fit(x_train,y_train)
```

```
[ ]: ElasticNet()
```

```
[ ]: print(en.coef_)  
    print(en.intercept_)
```

```
[ -3.53866147  39.80394701 -130.23937824 -47.35780527  65.71879417
 321.28574295 -180.22504533]
14080.192276362655
```

```
[ ]: prediction = en.predict(x_test)
prediction
```

```
[ ]: array([8957.56173913, 8961.10040061, 8530.9712209 , 8954.02307766,
 6084.9932824 , 4730.81329552, 7279.163564 , 8054.32054245,
 5947.14512798, 8634.24301547, 8085.5418622 , 8582.92043494,
 8072.2658871 , 8964.63906208, 8937.46047836, 8905.36048781,
 8058.73787471, 8587.58980449, 4723.73597257, 7295.72616329,
 7267.92094606, 7294.44891797, 6777.35115278, 6197.62423341,
 7090.28232746, 6835.75105626, 7054.64367543, 6194.08557194,
 8099.6965081 , 4785.17046296, 4191.47420905, 8923.05379518,
 6088.53194387, 8055.7032878 , 4735.8812396 , 8925.46174858,
 6758.59338034, 8545.1258668 , 6322.35004507, 7283.83293354,
 8958.69244721, 6291.63279987, 6158.44691991, 8095.90580935,
 8936.9564038 , 8950.07325231, 8916.85514303, 8933.92181689,
 4716.65864962, 7294.8474925 , 7309.0021384 , 4729.05595393,
 6277.47815398, 8625.63640992, 4746.24518674, 8965.03763662,
 7061.72099838, 4204.37944547, 6083.00432087, 7088.99102571,
 7044.02769101, 4277.13396604, 7285.11017887, 7101.15034916,
 6157.31621184, 8044.8352661 , 5594.94035692, 7406.18876541,
 6789.33191463, 6136.08424299, 6062.9030601 , 6798.83515891,
 8618.55908697, 7288.02220683, 4757.99187924, 6168.32086057,
 6357.9886971 , 3445.42551003, 6816.52846628, 8548.66452827,
 6825.38710911, 7107.97563483, 8620.08836958, 8622.09774844,
 6819.41585888, 8082.37779696, 6787.9671372 , 5601.45977484,
 8915.97647224, 5666.43292671, 7265.0089181 , 7100.89831188,
 4543.67552804, 6363.92664004, 8534.50988237, 6150.23888889,
 8095.65377207, 8047.2432195 , 6318.8113836 , 6143.16156594,
 6251.98140086, 4717.53732041, 5610.69301391, 8930.38315541,
 7301.92481545, 8568.76578905, 4744.71590413, 8961.9790714 ,
 6121.92959709, 7362.52358362, 5659.35560376, 6100.69762825,
 7302.80348624])
```

```
[ ]: en.score(x_test,y_test)
```

```
[ ]: 0.822335521229122
```

```
[ ]: from sklearn import metrics
print("Mean Absolute Error: ", metrics.mean_absolute_error(y_test,prediction))
print("Mean Squared Error: ", metrics.mean_squared_error(y_test,prediction))
print("Root Mean Squared Error: ", np.sqrt(metrics.
↪mean_squared_error(y_test,prediction)))
```

```
Mean Absolute Error: 526.6588674615243
```

Mean Squared Error: 508794.9471444348
Root Mean Squared Error: 713.2986381204122