Health and Sleep

September 28, 2024

Site Address: https://www.kaggle.com/datasets/hanaksoy/health-and-sleep-statistics

[2]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from scipy import stats
from sklearn.linear_model import LinearRegression
from sklearn.metrics import r2_score
```

[4]:
```
import pandas as pd

# Specify the correct file path to your CSV
file_path = "/Users/hassanali/Desktop/USD/AAI-550/Datasets/Original Dataset/
 ↪Health_Sleep_Statistics.csv"

# Read the CSV file
df = pd.read_csv(file_path)

# Display the first few rows of the dataframe to verify it loaded correctly
df.head()
```

[4]:

	User ID	Age	Gender	Sleep Quality	Bedtime	Wake-up Time	Daily Steps	Calories Burned	Physical Activity Level	Dietary Habits	Sleep Disorders	Medication Usage
0	1	25	f	8	23:00	06:30	8000	2500	medium	healthy	no	no
1	2	34	m	7	00:30	07:00	5000	2200	low	unhealthy	yes	
2	3	29	f	9	22:45	06:45	9000	2700	high	healthy	no	
3	4	41	m	5	01:00	06:30	4000	2100	low	unhealthy	yes	
4	5	22	f	8	23:30	07:00	10000	2800	high	medium	no	

```
1              yes
2               no
3               no
4               no
```

[5]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 100 entries, 0 to 99
Data columns (total 12 columns):
 #   Column                   Non-Null Count  Dtype 
---  ------                   --------------  ----- 
 0   User ID                  100 non-null    int64 
 1   Age                      100 non-null    int64 
 2   Gender                   100 non-null    object
 3   Sleep Quality            100 non-null    int64 
 4   Bedtime                  100 non-null    object
 5   Wake-up Time             100 non-null    object
 6   Daily Steps              100 non-null    int64 
 7   Calories Burned          100 non-null    int64 
 8   Physical Activity Level  100 non-null    object
 9   Dietary Habits           100 non-null    object
 10  Sleep Disorders          100 non-null    object
 11  Medication Usage         100 non-null    object
dtypes: int64(5), object(7)
memory usage: 9.5+ KB
```

[6]:
```
df.describe()
```

[6]:

	User ID	Age	Sleep Quality	Daily Steps	Calories Burned
count	100.000000	100.000000	100.000000	100.000000	100.00000
mean	50.500000	36.010000	7.000000	6830.000000	2421.00000
std	29.011492	8.454865	1.734964	2498.706736	281.06759
min	1.000000	22.000000	4.000000	3000.000000	2000.00000
25%	25.750000	28.750000	5.750000	4750.000000	2175.00000
50%	50.500000	35.000000	7.500000	6750.000000	2400.00000
75%	75.250000	44.000000	8.250000	9000.000000	2700.00000
max	100.000000	50.000000	9.000000	11000.000000	2900.00000

[7]:
```
# Check for missing values in each column
df.isnull().sum()
```

[7]:
```
User ID                  0
Age                      0
Gender                   0
Sleep Quality            0
Bedtime                  0
Wake-up Time             0
```

```
Daily Steps                0
Calories Burned            0
Physical Activity Level    0
Dietary Habits             0
Sleep Disorders            0
Medication Usage           0
dtype: int64
```

[8]:
```python
import pandas as pd

# Load the dataset
file_path = "/Users/hassanali/Desktop/USD/AAI-550/Datasets/Original Dataset/
 ↪Health_Sleep_Statistics.csv"
df = pd.read_csv(file_path)

# Get basic info about the dataset
rows, columns = df.shape
print(f"Total records (rows): {rows}")
print(f"Total columns: {columns}")

# Get summary information
df.info()

# Get summary statistics
df.describe()

# Check for missing values
missing_values = df.isnull().sum()
print("Missing values in each column:\n", missing_values)
```

```
Total records (rows): 100
Total columns: 12
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 100 entries, 0 to 99
Data columns (total 12 columns):
 #   Column                   Non-Null Count  Dtype
---  ------                   --------------  -----
 0   User ID                  100 non-null    int64
 1   Age                      100 non-null    int64
 2   Gender                   100 non-null    object
 3   Sleep Quality            100 non-null    int64
 4   Bedtime                  100 non-null    object
 5   Wake-up Time             100 non-null    object
 6   Daily Steps              100 non-null    int64
 7   Calories Burned          100 non-null    int64
 8   Physical Activity Level  100 non-null    object
 9   Dietary Habits           100 non-null    object
 10  Sleep Disorders          100 non-null    object
```

```
 11  Medication Usage         100 non-null    object
dtypes: int64(5), object(7)
memory usage: 9.5+ KB
Missing values in each column:
 User ID                    0
Age                         0
Gender                      0
Sleep Quality               0
Bedtime                     0
Wake-up Time                0
Daily Steps                 0
Calories Burned             0
Physical Activity Level     0
Dietary Habits              0
Sleep Disorders             0
Medication Usage            0
dtype: int64
```

[10]:
```
# Display the last 5 rows of the dataset
df.tail()
```

[10]:

	User ID	Age	Gender	Sleep Quality	Bedtime	Wake-up Time	Daily Steps	Calories Burned	Physical Activity Level	Dietary Habits	Sleep Disorders	Medication Usage
95	96	43	m	7	00:45	07:15	6500	2400	medium	medium	no	no
96	97	33	f	8	23:15	06:15	8500	2600	high	medium	no	no
97	98	46	m	4	01:30	07:00	3000	2000	low	unhealthy	yes	yes
98	99	25	f	9	22:15	06:45	9500	2700	high	healthy	no	no
99	100	41	m	6	00:30	07:00	5000	2200	medium	unhealthy	no	no