Lifestyle_Wellbeing

September 28, 2024

Site: https://www.kaggle.com/datasets/ydalat/lifestyle-and-wellbeing-data

[1]:
```
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
```

[3]:
```
# Specify the correct file path to your CSV
file_path = "/Users/hassanali/Desktop/USD/AAI-550/Datasets/Original Dataset/
 ↪Wellbeing_and_lifestyle_data_Kaggle.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()
```

[3]:

	Timestamp	FRUITS_VEGGIES	DAILY_STRESS	PLACES_VISITED	CORE_CIRCLE	\
0	7/7/15	3	2	2	5	
1	7/7/15	2	3	4	3	
2	7/7/15	2	3	3	4	
3	7/7/15	3	3	10	3	
4	7/7/15	5	1	3	3	

	SUPPORTING_OTHERS	SOCIAL_NETWORK	ACHIEVEMENT	DONATION	BMI_RANGE	…	\
0	0	5	2	0	1	…	
1	8	10	5	2	2	…	
2	4	10	3	2	2	…	
3	10	7	2	5	2	…	
4	10	4	2	4	2	…	

	SLEEP_HOURS	LOST_VACATION	DAILY_SHOUTING	SUFFICIENT_INCOME	\
0	7	5	5	1	
1	8	2	2	2	
2	8	10	2	2	

```
3            5              7              5                 1
4            7              0              0                 2

   PERSONAL_AWARDS  TIME_FOR_PASSION  WEEKLY_MEDITATION         AGE  GENDER  \
0                4                 0                  5    36 to 50  Female
1                3                 2                  6    36 to 50  Female
2                4                 8                  3    36 to 50  Female
3                5                 2                  0  51 or more  Female
4                8                 1                  5  51 or more  Female

   WORK_LIFE_BALANCE_SCORE
0                    609.5
1                    655.6
2                    631.6
3                    622.7
4                    663.9

[5 rows x 24 columns]
```

[4]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 15972 entries, 0 to 15971
Data columns (total 24 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   Timestamp               15972 non-null  object
 1   FRUITS_VEGGIES          15972 non-null  int64
 2   DAILY_STRESS            15972 non-null  object
 3   PLACES_VISITED          15972 non-null  int64
 4   CORE_CIRCLE             15972 non-null  int64
 5   SUPPORTING_OTHERS       15972 non-null  int64
 6   SOCIAL_NETWORK          15972 non-null  int64
 7   ACHIEVEMENT             15972 non-null  int64
 8   DONATION                15972 non-null  int64
 9   BMI_RANGE               15972 non-null  int64
 10  TODO_COMPLETED          15972 non-null  int64
 11  FLOW                    15972 non-null  int64
 12  DAILY_STEPS             15972 non-null  int64
 13  LIVE_VISION             15972 non-null  int64
 14  SLEEP_HOURS             15972 non-null  int64
 15  LOST_VACATION           15972 non-null  int64
 16  DAILY_SHOUTING          15972 non-null  int64
 17  SUFFICIENT_INCOME       15972 non-null  int64
 18  PERSONAL_AWARDS         15972 non-null  int64
 19  TIME_FOR_PASSION        15972 non-null  int64
 20  WEEKLY_MEDITATION       15972 non-null  int64
 21  AGE                     15972 non-null  object
```

```
 22  GENDER                    15972 non-null  object
 23  WORK_LIFE_BALANCE_SCORE   15972 non-null  float64
dtypes: float64(1), int64(19), object(4)
memory usage: 2.9+ MB
```

[10]:
```
df.describe()
```

[10]:

	FRUITS_VEGGIES	PLACES_VISITED	CORE_CIRCLE	SUPPORTING_OTHERS	\
count	15972.000000	15972.000000	15972.000000	15972.000000	
mean	2.922677	5.232970	5.508077	5.616454	
std	1.442694	3.311912	2.840334	3.242021	
min	0.000000	0.000000	0.000000	0.000000	
25%	2.000000	2.000000	3.000000	3.000000	
50%	3.000000	5.000000	5.000000	5.000000	
75%	4.000000	8.000000	8.000000	10.000000	
max	5.000000	10.000000	10.000000	10.000000	

	SOCIAL_NETWORK	ACHIEVEMENT	DONATION	BMI_RANGE	\
count	15972.000000	15972.000000	15972.000000	15972.000000	
mean	6.474267	4.000751	2.715314	1.410656	
std	3.086672	2.755837	1.851586	0.491968	
min	0.000000	0.000000	0.000000	1.000000	
25%	4.000000	2.000000	1.000000	1.000000	
50%	6.000000	3.000000	3.000000	1.000000	
75%	10.000000	6.000000	5.000000	2.000000	
max	10.000000	10.000000	5.000000	2.000000	

	TODO_COMPLETED	FLOW	DAILY_STEPS	LIVE_VISION	SLEEP_HOURS	\
count	15972.000000	15972.000000	15972.000000	15972.000000	15972.000000	
mean	5.745993	3.194778	5.703606	3.752129	7.042888	
std	2.624097	2.357518	2.891013	3.230987	1.199044	
min	0.000000	0.000000	1.000000	0.000000	1.000000	
25%	4.000000	1.000000	3.000000	1.000000	6.000000	
50%	6.000000	3.000000	5.000000	3.000000	7.000000	
75%	8.000000	5.000000	8.000000	5.000000	8.000000	
max	10.000000	10.000000	10.000000	10.000000	10.000000	

	LOST_VACATION	DAILY_SHOUTING	SUFFICIENT_INCOME	PERSONAL_AWARDS	\
count	15972.000000	15972.000000	15972.000000	15972.000000	
mean	2.898886	2.930879	1.728963	5.711558	
std	3.692180	2.676301	0.444509	3.089630	
min	0.000000	0.000000	1.000000	0.000000	
25%	0.000000	1.000000	1.000000	3.000000	
50%	0.000000	2.000000	2.000000	5.000000	
75%	5.000000	4.000000	2.000000	9.000000	
max	10.000000	10.000000	2.000000	10.000000	

```
       TIME_FOR_PASSION  WEEKLY_MEDITATION  WORK_LIFE_BALANCE_SCORE
count      15972.000000       15972.000000             15972.000000
mean           3.326572           6.233346               666.751503
std            2.729293           3.016571                45.019868
min            0.000000           0.000000               480.000000
25%            1.000000           4.000000               636.000000
50%            3.000000           7.000000               667.700000
75%            5.000000          10.000000               698.500000
max           10.000000          10.000000               820.200000
```

[11]:
```
# Check for missing values in each column
df.isnull().sum()
```

[11]:
```
Timestamp                 0
FRUITS_VEGGIES            0
DAILY_STRESS              0
PLACES_VISITED            0
CORE_CIRCLE               0
SUPPORTING_OTHERS         0
SOCIAL_NETWORK            0
ACHIEVEMENT               0
DONATION                  0
BMI_RANGE                 0
TODO_COMPLETED            0
FLOW                      0
DAILY_STEPS               0
LIVE_VISION               0
SLEEP_HOURS               0
LOST_VACATION             0
DAILY_SHOUTING            0
SUFFICIENT_INCOME         0
PERSONAL_AWARDS           0
TIME_FOR_PASSION          0
WEEKLY_MEDITATION         0
AGE                       0
GENDER                    0
WORK_LIFE_BALANCE_SCORE   0
dtype: int64
```

[13]:
```
# Load the dataset
file_path = "/Users/hassanali/Desktop/USD/AAI-550/Datasets/Original Dataset/
 ↪Wellbeing_and_lifestyle_data_Kaggle.csv"
df = pd.read_csv(file_path)

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

```python
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): 15972
Total columns: 24
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 15972 entries, 0 to 15971
Data columns (total 24 columns):
 #   Column                   Non-Null Count  Dtype
---  ------                   --------------  -----
 0   Timestamp                15972 non-null  object
 1   FRUITS_VEGGIES           15972 non-null  int64
 2   DAILY_STRESS             15972 non-null  object
 3   PLACES_VISITED           15972 non-null  int64
 4   CORE_CIRCLE              15972 non-null  int64
 5   SUPPORTING_OTHERS        15972 non-null  int64
 6   SOCIAL_NETWORK           15972 non-null  int64
 7   ACHIEVEMENT              15972 non-null  int64
 8   DONATION                 15972 non-null  int64
 9   BMI_RANGE                15972 non-null  int64
 10  TODO_COMPLETED           15972 non-null  int64
 11  FLOW                     15972 non-null  int64
 12  DAILY_STEPS              15972 non-null  int64
 13  LIVE_VISION              15972 non-null  int64
 14  SLEEP_HOURS              15972 non-null  int64
 15  LOST_VACATION            15972 non-null  int64
 16  DAILY_SHOUTING           15972 non-null  int64
 17  SUFFICIENT_INCOME        15972 non-null  int64
 18  PERSONAL_AWARDS          15972 non-null  int64
 19  TIME_FOR_PASSION         15972 non-null  int64
 20  WEEKLY_MEDITATION        15972 non-null  int64
 21  AGE                      15972 non-null  object
 22  GENDER                   15972 non-null  object
 23  WORK_LIFE_BALANCE_SCORE  15972 non-null  float64
dtypes: float64(1), int64(19), object(4)
memory usage: 2.9+ MB
Missing values in each column:
 Timestamp                  0
FRUITS_VEGGIES              0
```

```
DAILY_STRESS                0
PLACES_VISITED              0
CORE_CIRCLE                 0
SUPPORTING_OTHERS           0
SOCIAL_NETWORK              0
ACHIEVEMENT                 0
DONATION                    0
BMI_RANGE                   0
TODO_COMPLETED              0
FLOW                        0
DAILY_STEPS                 0
LIVE_VISION                 0
SLEEP_HOURS                 0
LOST_VACATION               0
DAILY_SHOUTING              0
SUFFICIENT_INCOME           0
PERSONAL_AWARDS             0
TIME_FOR_PASSION            0
WEEKLY_MEDITATION           0
AGE                         0
GENDER                      0
WORK_LIFE_BALANCE_SCORE     0
dtype: int64
```

[14]:
```
# Display the last 5 rows of the dataset
df.tail()
```

[14]:

	Timestamp	FRUITS_VEGGIES	DAILY_STRESS	PLACES_VISITED	CORE_CIRCLE	\
15967	3/14/21 5:42	3	3	0	4	
15968	3/14/21 6:30	3	3	6	8	
15969	3/14/21 8:35	4	3	0	10	
15970	3/14/21 8:43	1	1	10	8	
15971	3/14/21 9:03	5	4	0	2	

	SUPPORTING_OTHERS	SOCIAL_NETWORK	ACHIEVEMENT	DONATION	BMI_RANGE	\
15967	0	10	0	4	2	
15968	7	4	6	3	1	
15969	10	8	6	5	1	
15970	2	7	3	2	1	
15971	10	10	5	1	2	

	…	SLEEP_HOURS	LOST_VACATION	DAILY_SHOUTING	SUFFICIENT_INCOME	\
15967	…	7	0	1	1	
15968	…	6	0	0	2	
15969	…	7	0	1	2	
15970	…	8	7	2	2	
15971	…	8	5	2	2	

```
       PERSONAL_AWARDS  TIME_FOR_PASSION  WEEKLY_MEDITATION         AGE  \
15967                5                 2                  5  51 or more
15968               10                 5                  8    21 to 35
15969               10                 1                 10    21 to 35
15970                1                 6                  8    21 to 35
15971                1                 8                  4    21 to 35

       GENDER  WORK_LIFE_BALANCE_SCORE
15967  Female                   644.5
15968  Female                   714.9
15969    Male                   716.6
15970  Female                   682.0
15971  Female                   651.4

[5 rows x 24 columns]
```