Sleep Analysis Project Documentation

Aakash Sharma and Risha Poojary

Objective of the Project

The objective of this project is to analyze various factors affecting sleep quality by examining features such as age, gender, daily steps, calories burned, wake-up time, medication usage, sleep disorders, and others. The goal is to uncover patterns and relationships that can lead to insights on improving sleep quality and understanding the impact of different variables on sleep health.

Import Libraries

To carry out this analysis, the following libraries were imported:

- **pandas**: For data manipulation and analysis.

- **numpy**: For numerical operations.

- **matplotlib.pyplot**: For creating various plots and charts.

- **seaborn**: For statistical data visualization.

- **sklearn.linear_model**: For building linear regression models.

- **warnings**: To suppress warnings during analysis.

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

import warnings

warnings.filterwarnings('ignore')

Data Reading

The dataset used in this analysis was read from a CSV file using the pandas read_csv function. The dataset contains 100 entries and 12 columns, which include:

- **User ID**: An identifier for each user.

- **Age**: The age of the individual.

- **Gender**: Gender of the individual.

- **Sleep Quality**: The quality of sleep on a scale.

- **Bedtime**: The time the individual goes to bed.

- **Wake-up Time**: The time the individual wakes up.

- **Daily Steps**: The number of steps taken by the individual each day.

- **Calories Burned**: The total calories burned by the individual.

- **Physical Activity Level**: A categorization of the individual's physical activity.

- **Dietary Habits**: The individual's dietary habits.

- **Sleep Disorders**: Whether the individual has any sleep disorders.

- **Medication Usage**: Whether the individual uses medication affecting sleep.

df = pd.read_csv(f'/content/Health_Sleep_Statistics.csv')

df.head()

Data Cleaning

The dataset was checked for any issues like missing values or duplicate rows. The following steps were performed:

- **Null Values**: No missing values were found in the dataset.

- **Duplicate Values**: There were no duplicate entries.

- **Dropped 'User ID'**: Since the 'User ID' column did not provide useful insights for analysis, it was dropped.

df.isna().sum() #Check for missing values

df.duplicated().sum() #Check for duplicates

df.drop(['User ID'], axis=1, inplace=True) #Drop 'User ID'

Visualization

Several visualizations were created to understand the relationships between different features and sleep quality:

1. Age vs Sleep Quality

A scatter plot and bar chart were created to observe the relationship between age and sleep quality. It was found that sleep quality decreases as age increases, with a significant reduction in sleep quality between ages 22-50.

df.plot(kind='scatter', x='Age', y='Sleep Quality', s=28, alpha=1, c='purple')

2. Age vs Sleep Quality (Bar Chart)

A bar chart further illustrated how sleep quality changes with age.

sns.barplot(x='Age', y='Sleep Quality', data=df, color='orange')

3. Linear Regression Model

A linear regression model was built to predict sleep quality based on age. The actual and predicted data were compared to show how well the model fits the data.

```
model = LinearRegression()

model.fit(X, Y)

sns.scatterplot(x='Age', y='Sleep Quality', data=df, label='Actual Data', color='red')

plt.plot(X, model.predict(X), color='green', label='Predicted Data')
```

4. Gender vs Sleep Quality

A bar plot was created to compare sleep quality between males and females.

```
sns.barplot(x='Gender', y='Sleep Quality', data=df, color='yellow')
```

5. Wake-up Time vs Sleep Quality

A bar plot showed the relationship between wake-up time and sleep quality.

```
sns.barplot(x='Wake-up Time', y='Sleep Quality', data=df, color='green')
```

6. Sleep Duration Hours vs Sleep Quality

A plot illustrated how sleep duration hours impact sleep quality.

```
sns.barplot(x='Wake-up Time', y='Sleep Quality', data=df, color='purple')
```

7. Sleep Disorders vs Sleep Quality

A box plot helped visualize the distribution of sleep quality by sleep disorders.

```
sns.boxplot(x='Sleep Disorders', y='Sleep Quality', data=df, color='magenta')
```

8. Medication Usage vs Sleep Quality

A bar plot was used to explore the relationship between medication usage and sleep quality.

```
sns.barplot(x='Medication Usage', y='Sleep Quality', data=df, color='grey')
```

9. Correlation Matrix

A heatmap was generated to visualize correlations between numeric variables. Key insights include:

- **Daily Steps** and **Calories Burned** are strongly correlated with **Sleep Quality**.

- **Age** has a negative correlation with **Sleep Quality**.

```
corr = df_numerics.corr()

sns.heatmap(corr, annot=True)
```

Conclusion

From the analysis, the following conclusions were drawn:

- **Age**: Sleep quality decreases as age increases, with a marked decline from ages 22 to 50.

- **Physical Activity**: Both daily steps and calories burned are strongly correlated with improved sleep quality.

- **Gender**: There are differences in sleep quality between males and females.

- **Wake-up Time**: The wake-up time seems to have a moderate impact on sleep quality.

- **Sleep Disorders**: The presence of sleep disorders negatively affects sleep quality.

- **Medication Usage**: Individuals using medication had varied sleep quality based on their usage patterns.

- **Correlations**: The correlation matrix confirmed that physical activity levels, age, and sleep quality are strongly related.