Name: Muhammad Saad Bin Sagheer   Student Id: 23086746   GitHub: github.com/Saadi222

Components of Healthy Lifestyle

Introduction

A healthy lifestyle is critical for professional success. Understanding sleep patterns and their correlation with health factors is crucial. This analysis examines a dataset consisting of sleep duration, sleep quality, and related variables such as age, occupation, physical activity level, stress level, BMI category, blood pressure, and heart rate. By analysing these factors, we aim to uncover significant relationships between sleep habits and overall health outcomes. This research will provide valuable insights into how sleep impacts various aspects of individual well-being and potentially give recommendations for improving sleep quality and overall health.

## Data Visualisations, Fitting and Clustering

Histogram (Age Distribution)

The histogram shows age distribution of the dataset; with an average age of 42.18, indicating a mostly mature group. The median age of 43 confirmed this, meaning half of the participants were older than 43. The most common age was also 43, suggesting that many participants were around this age, possibly because the dataset consists of employees aged between 27-59 years. The distribution had a slight right skew (0.26), meaning there were slightly more older participants. The negative kurtosis (-0.91) indicated that the age distribution was flatter than a normal curve, meaning the ages were spread more evenly across the group.
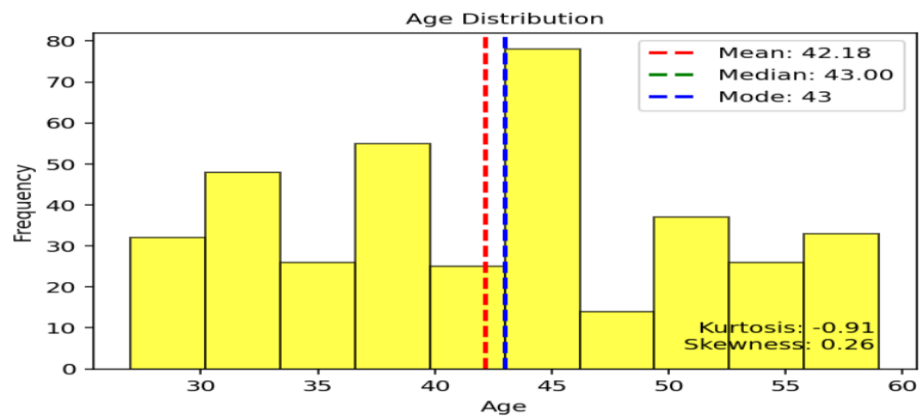

Bar Graph (Count of BMI categories by Occupation)

The analysis of BMI categories across different occupations showed some interesting differences. Most professions had a balanced spread of BMI categories, but nurses had a noticeably higher percentage of individuals in the "Obese" category i.e. 68. On the other hand, doctors had more individuals in the "Normal" BMI category. These results suggest that occupation may affect BMI, possibly due to factors like work-related stress, lifestyle choices, and access to healthy resources. To make stronger conclusions about the link between occupation and BMI, we can consider other related factors.
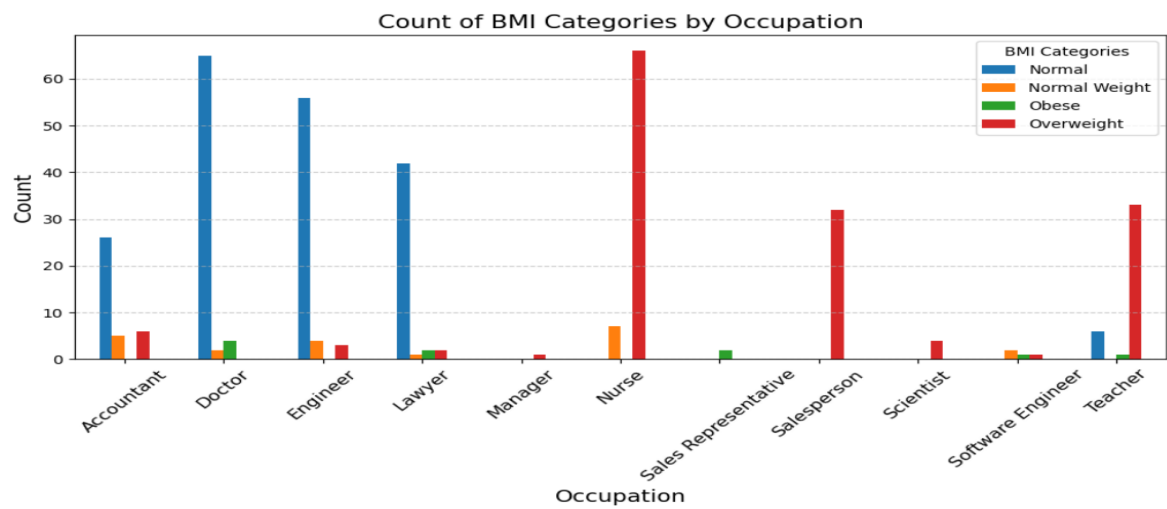
Scatter Plot (Sleep Duration vs Age)

The scatter plot provides an interesting look at the relationship between sleep duration and age. Although no clear linear trend is visible e.g. linearity, but we can see several clusters of data points suggest some interesting patterns. For example, there seems to be a group of individuals in their late 30s and early 40s who report sleeping around 7 hours, possibly indicating an optimal sleep duration for this age group. Additionally, a cluster of some older individuals (50+) appears to have slightly shorter sleep durations compared to other age groups, hinting at potential changes in sleep patterns with age; but also, some individuals of same age have a cluster with the highest sleep duration. This suggest that some other variables such as occupation, may also depend on sleep duration for older people.
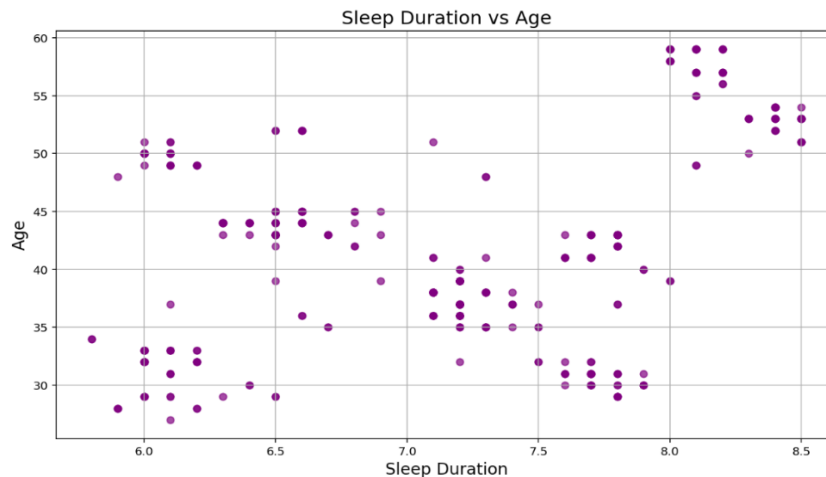

Heatmap of correlations (Kendall)

Heatmap using Kendall method reveals a strong positive correlation between quality of sleep and sleep duration, stress level and heart rate, and physical activity level and daily steps. There are also some strong negative correlations as well. Stress level and quality of sleep has the strongest negative correlation; meaning that the quality of sleep is greatly affected by the stress level. If a person is more stressed, he will not get a quality sleep. Stress level also has a very strong negative correlation with sleep duration, meaning the more stress will result in lesser duration of sleep. Heart Rated and sleep duration, and heart rate and quality of sleep also has a moderate negative correlation. There are very slight or no correlations between other features.
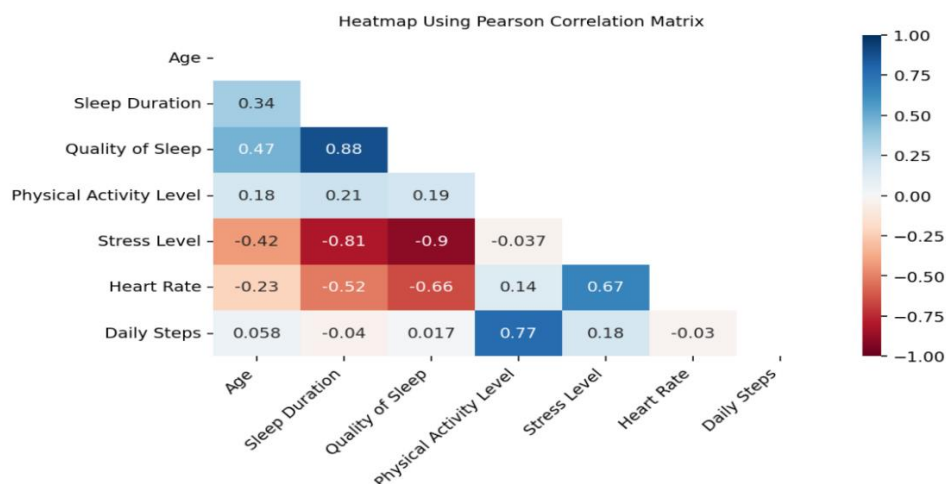

Linear Fitting

Since there is a strong negative correlation between quality of sleep and stress level; they are also inversely proportional to each other. They can be written in the form $y = mx+c$, where m is gradient, and c is the y-intercept. The figure shows the line of best fit with uncertainties and error bars of the points being predicted and includes the equation of the best fit line. Therefore, if we know the quality of the sleep someone is getting, we can predict its stress level.

Linear Fit with predictions
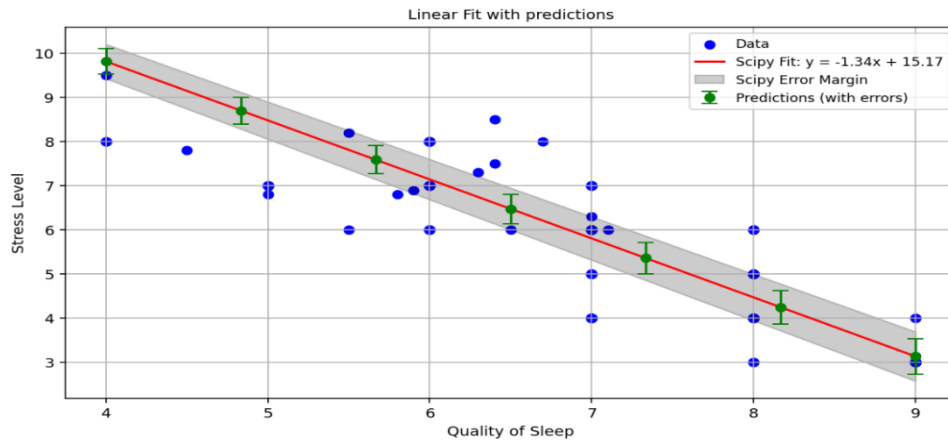
Elbow Method

According to elbow method, 3 to 4 clusters are enough to represent our data, but Silhouette Score is highest for 9 clusters. Since our goal is to clearly distinguish clusters, we will go with 9 clusters (Silhouette Score). We cannot go with 3 – 4 clusters because they are well separated and clearly visible.
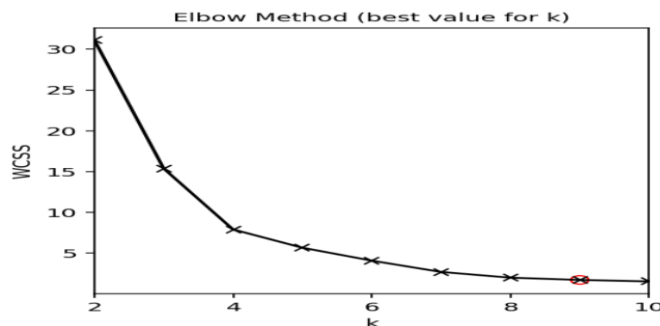
K-Means Clustering

The figure on the left shows the clusters while the figure on the right shows predicted cluster number for some data points. Clusters are clearly distinguishable and well-separated. The predictions of clusters are made at the points ([[25, 8], [43, 6.2], [60, 8], [43,7.7], [25,6]]). The big circle shows the predicted cluster for the data point.