08/12/2022

Multilevel Monitoring of Activity and Sleep in Healthy People

Group 005– Section 2
Group Members

Weeramundage Ishani Madushika Piyathilake - 0796108
Julie Kunnuvila Thomas - 0792998
Shashvat Sachdeva -0768857
Sneha Sabu - 0792518

Introduction

Healthcare analytics refers to the process of analyzing recent as well as historical data with the aim of finding meaningful insights to overcome enhance outreach and deliver optimal healthcare management. Analysis is carried out adhering to analytical disciplines on data-oriented decision making to discover and explore for trends and patterns in the data.

Problem Statement

Based on sleep quality of a healthy person we composed our problem statement. Our primary objective of the project is to monitor and explore the correlations of several aspects of people's everyday life such as **cardiovascular responses, psychological** perceptions (e.g., stress, anxiety, and emotions), **sleep quality, movement information** (e.g., wrist accelerometer data and steps) and **hourly activity descriptions**

- Anthropocentric characteristics of the participant:
- Information about sleep duration and sleep quality of the participant:
- Beat-to-beat interval data:
- Scores for all the questionnaires:
- List of the activity categories throughout the day.
- Accelerometer and inclinometer data recorded throughout the day

Data Collection and Description

Following is the link to the dataset that is used in this project.

https://physionet.org/content/mmash/1.0.0/

The dataset contained research data from 22 participants and the data is stored in different folders and files for the users based on the features.

1. user_info.csv - anthropocentric characteristics of the participant:
2. sleep.csv - information about sleep duration and sleep quality of the participant:
3. RR.csv - beat-to-beat interval data:
4. questionnaire.csv - scores for all the questionnaires:
5. Activity.csv - list of the activity categories throughout the day
6. Actigraph.csv - accelerometer and inclinometer data recorded throughout the day:
7. saliva.csv - clock genes and hormones concentrations in the saliva before going to bed and after waking up.

Data Preprocessing and Cleaning

Since the data is collected and stores in different folders and files for each participant, data merging was done based on the above 7 features to analyze the data more conveniently. This merging was done using the VBS macroscript in Excel.

- We merge the data for all 22 participants into one single csv file for each features.
- We have added Avg_HR, steps_count, Sleep Quality, Stress Activity in last 24 hours, Is stressful?, State Anxiety, ab(adverse behaviour) to sleep data set from questionnaire and activity dataset.
- We checked for missing data and found that user 11 had no data. So ,we dropped that.
- Data Type were checked and Transformed the Date columns to **Date-Time** format.
- converted the time columns into seconds.
- Inaccurate data inputs
- Data misinterpretation

Exploration and Analysis

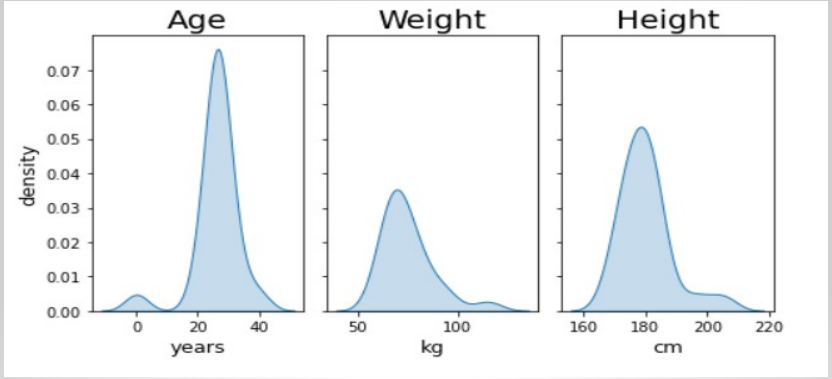
- Data exploration was one of the important step in our project. Exploring the data helped us understand the data and the correlations exists in between them to come to conclusions
- We have explored user_info, questionnaire, sleep dataset which we think are the most important dataset which can give some insights to our project.

User info

Descriptive Statistics and Density plots were created for User Info data as follows.

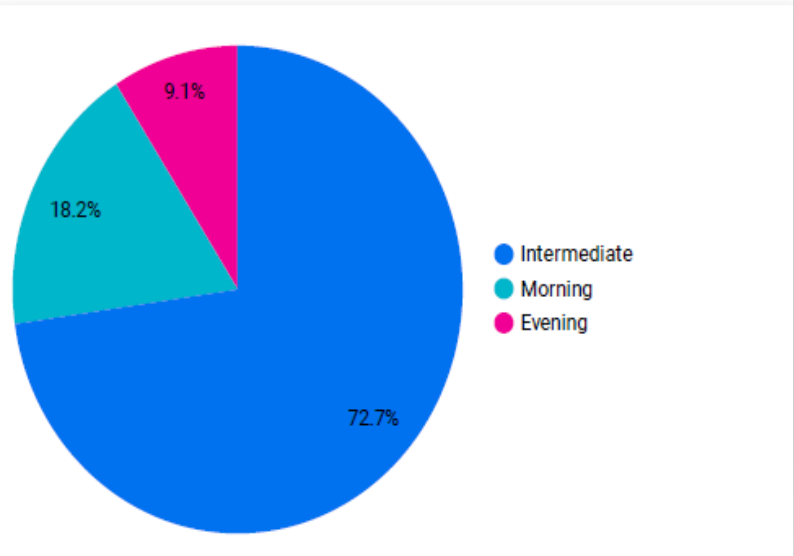
```
df_userinfo.describe().T # Descriptive statistics
```

	count	mean	std	min	25%	50%	75%	max
Weight	22.0	75.045455	12.789420	60.0	67.0	70.0	80.00	115.0
Height	22.0	179.909091	8.216760	169.0	175.0	180.0	183.00	205.0
Age	22.0	26.045455	7.121244	0.0	25.0	27.0	27.75	40.0

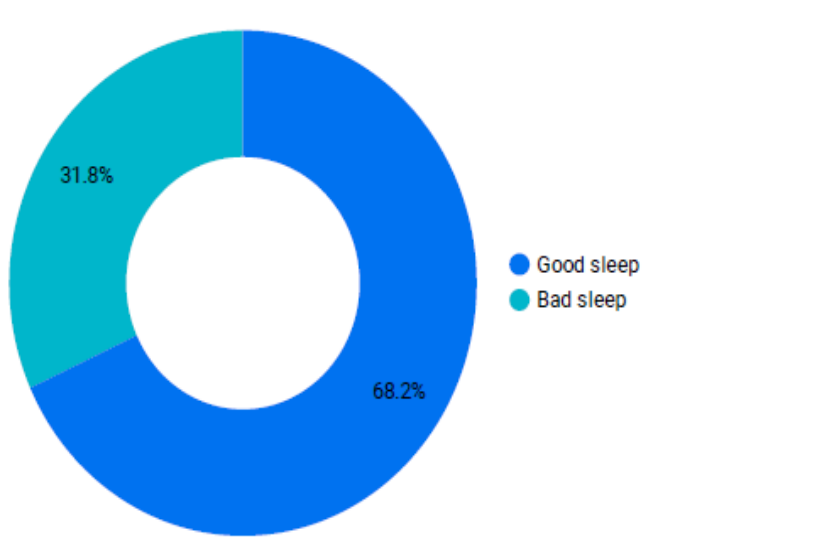


Questionnaire

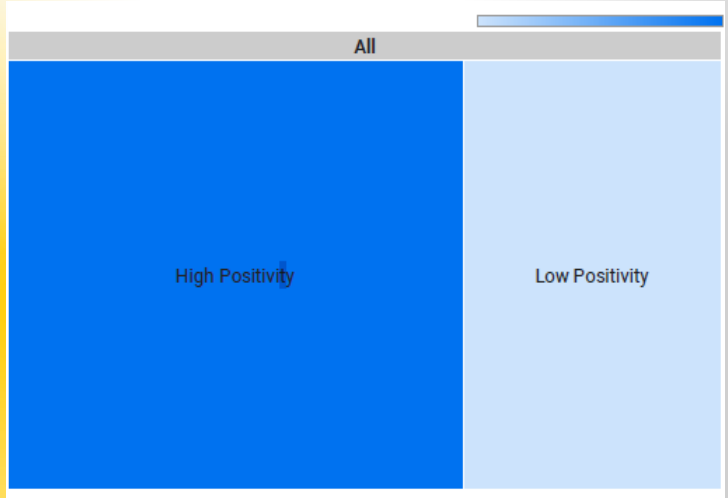
1. **Morningness/Eveningness –**



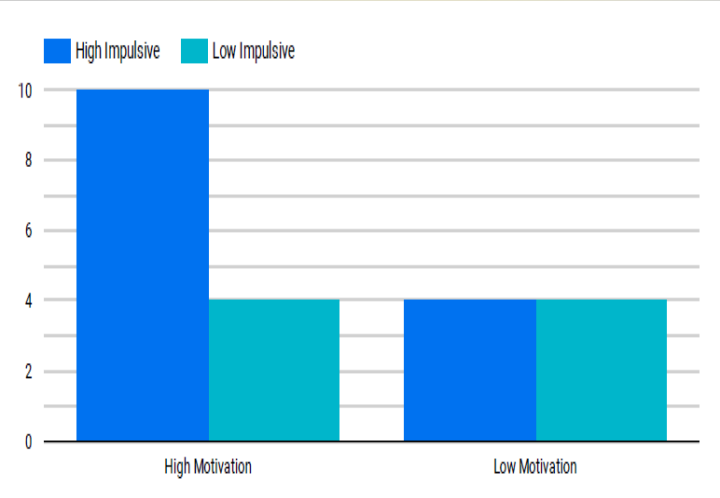
2. **Sleep Quality**



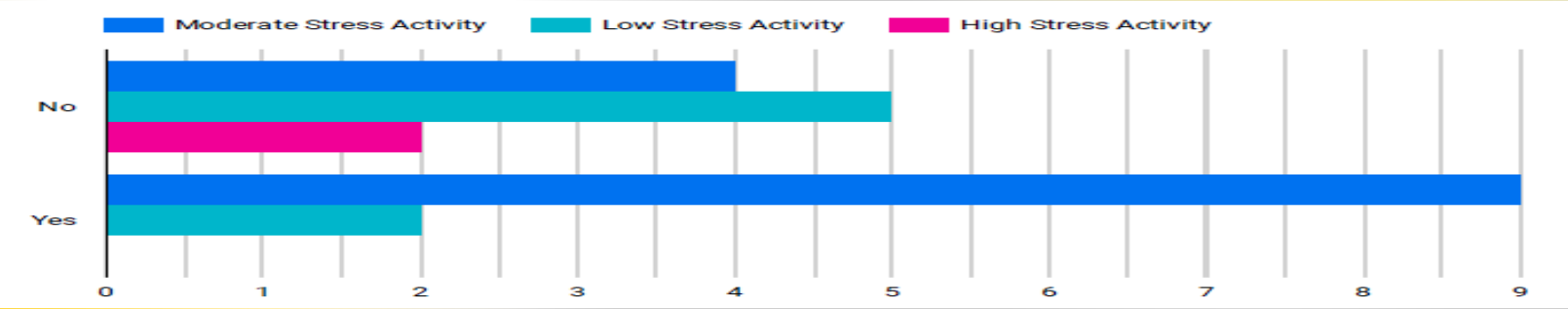
3. Positive/Negative Attribution



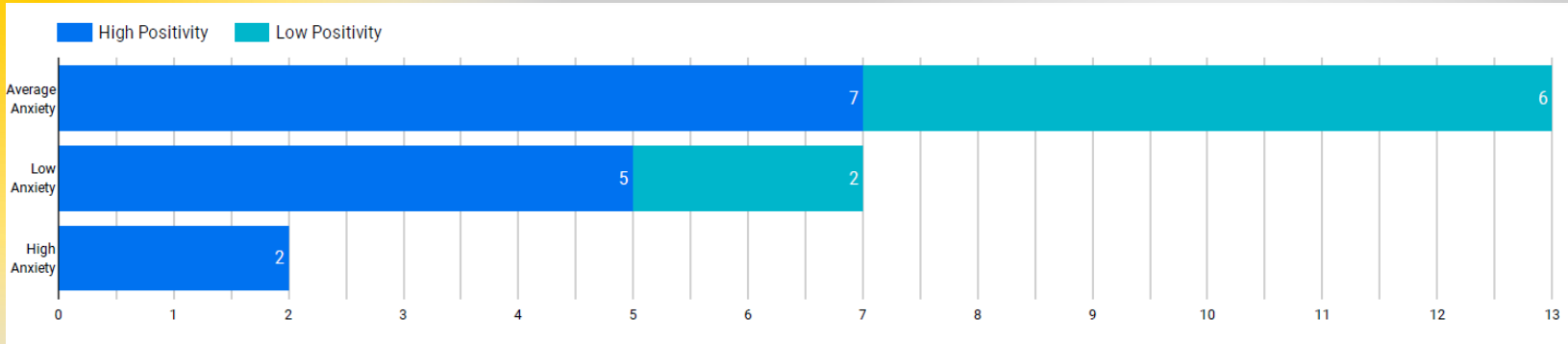
4. Impulsiveness v/s Motivational Intensity



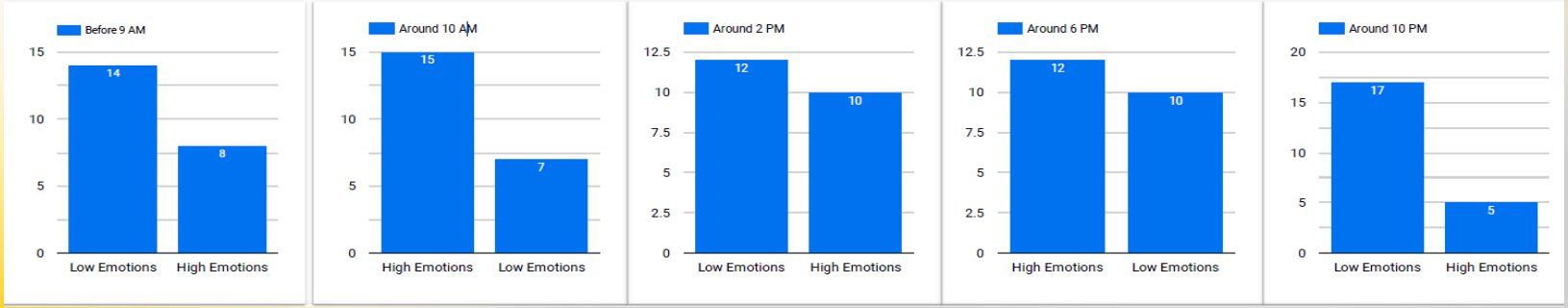
5. Stressful Activity v/s Stressfulness



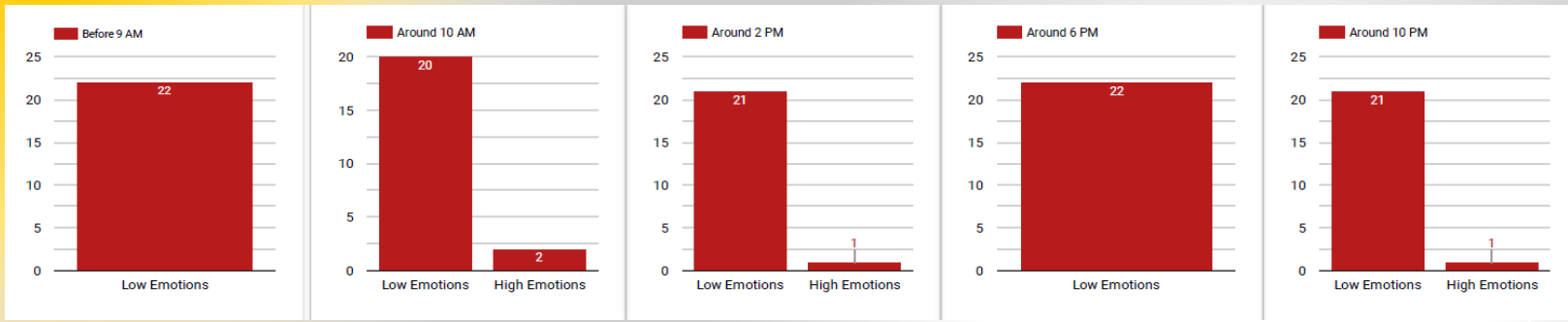
6. Anxiety Level v/s Avoidance Behaviour



7. Perceived Emotions towards Positive News:

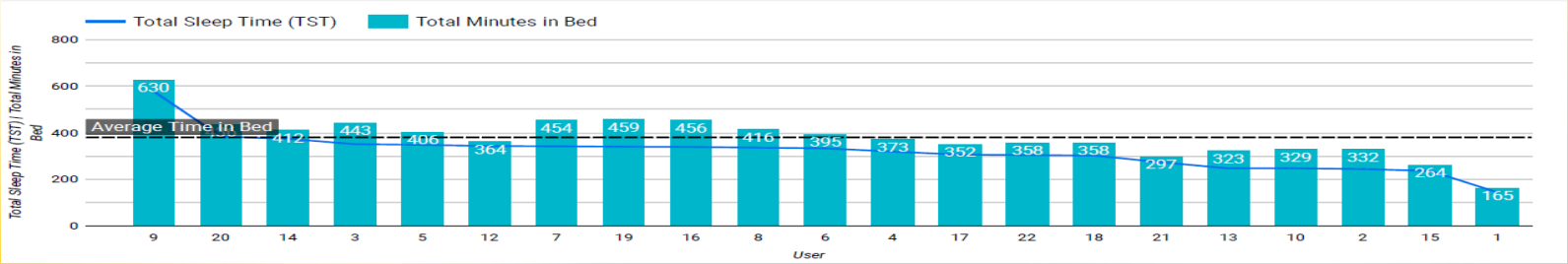


8. Perceived Emotions towards Negative News

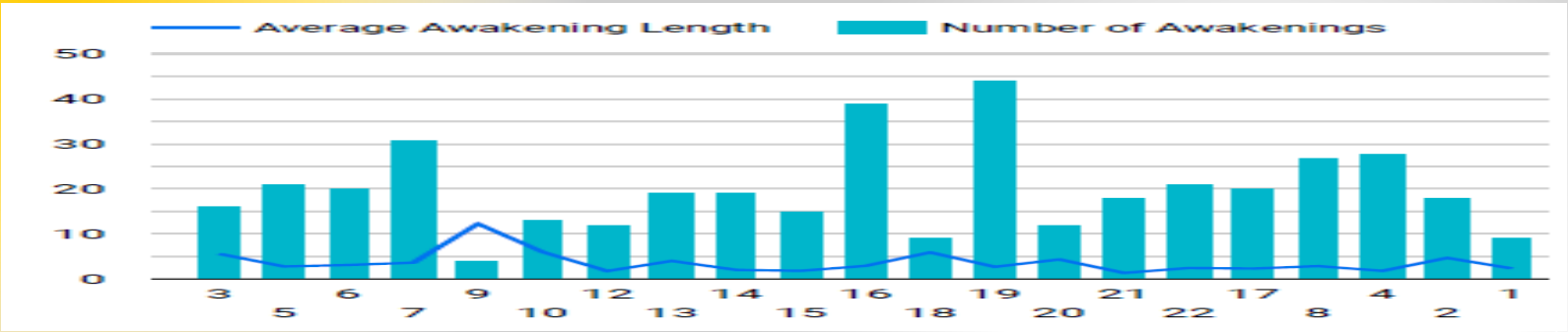


SLEEP ANALYSIS

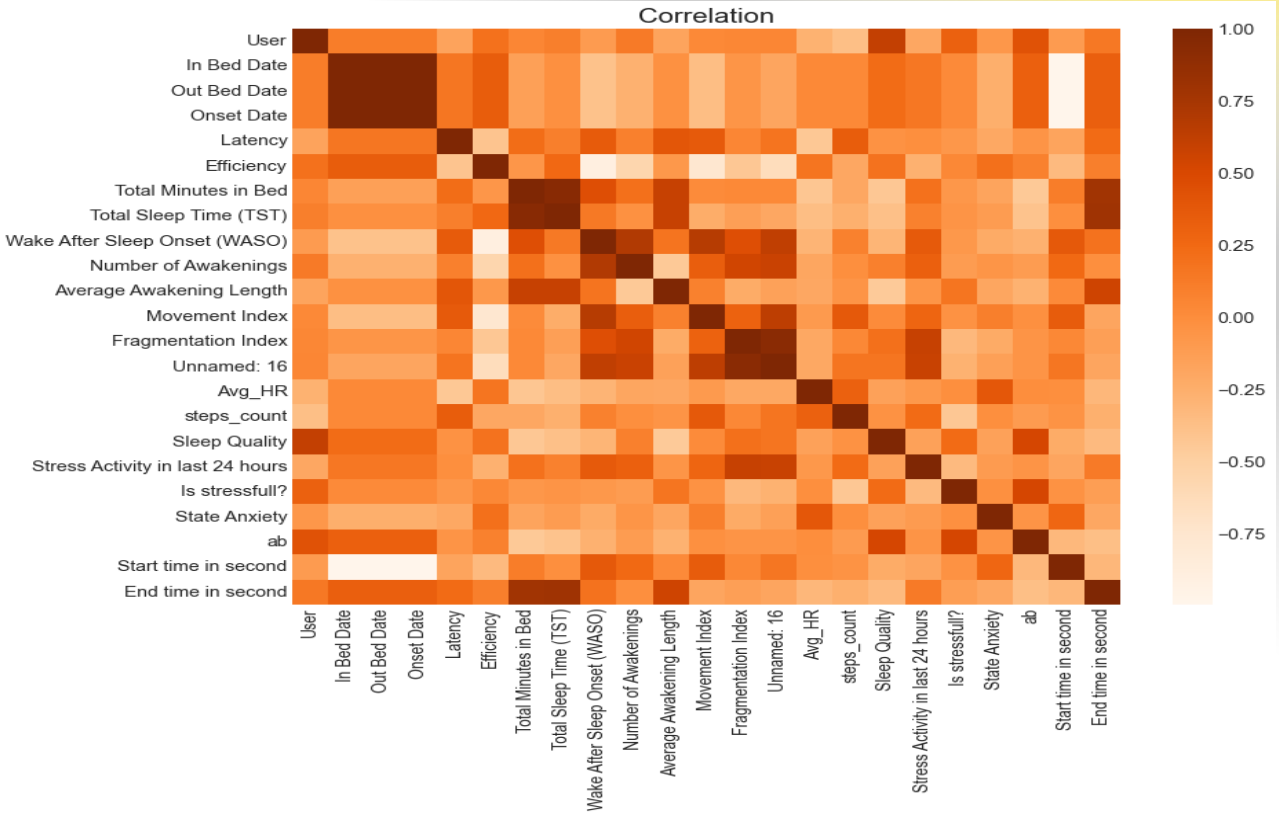
1. Anxiety Level v/s Avoidance Behavior



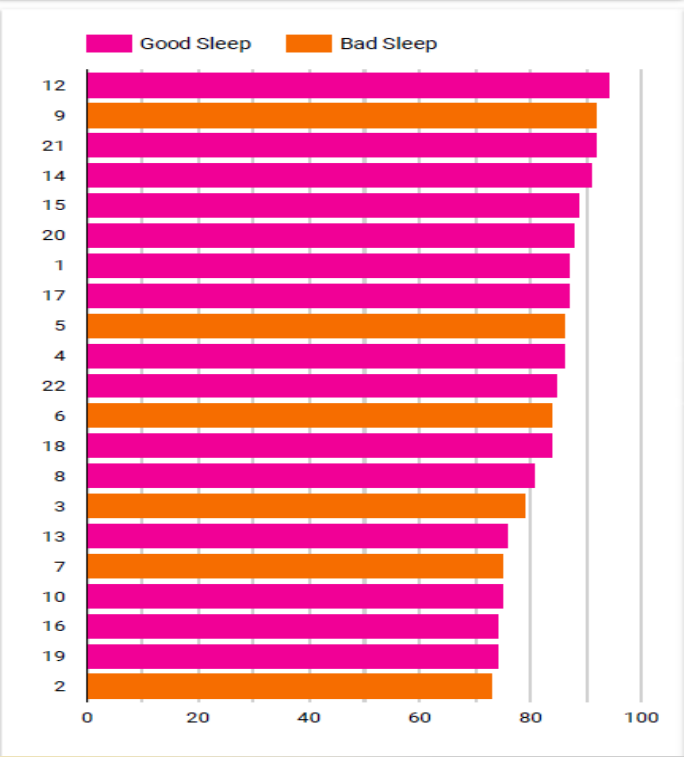
6. Average Awakening Time v/s No of Awakenings



Correlation of elements in Sleep Dataset

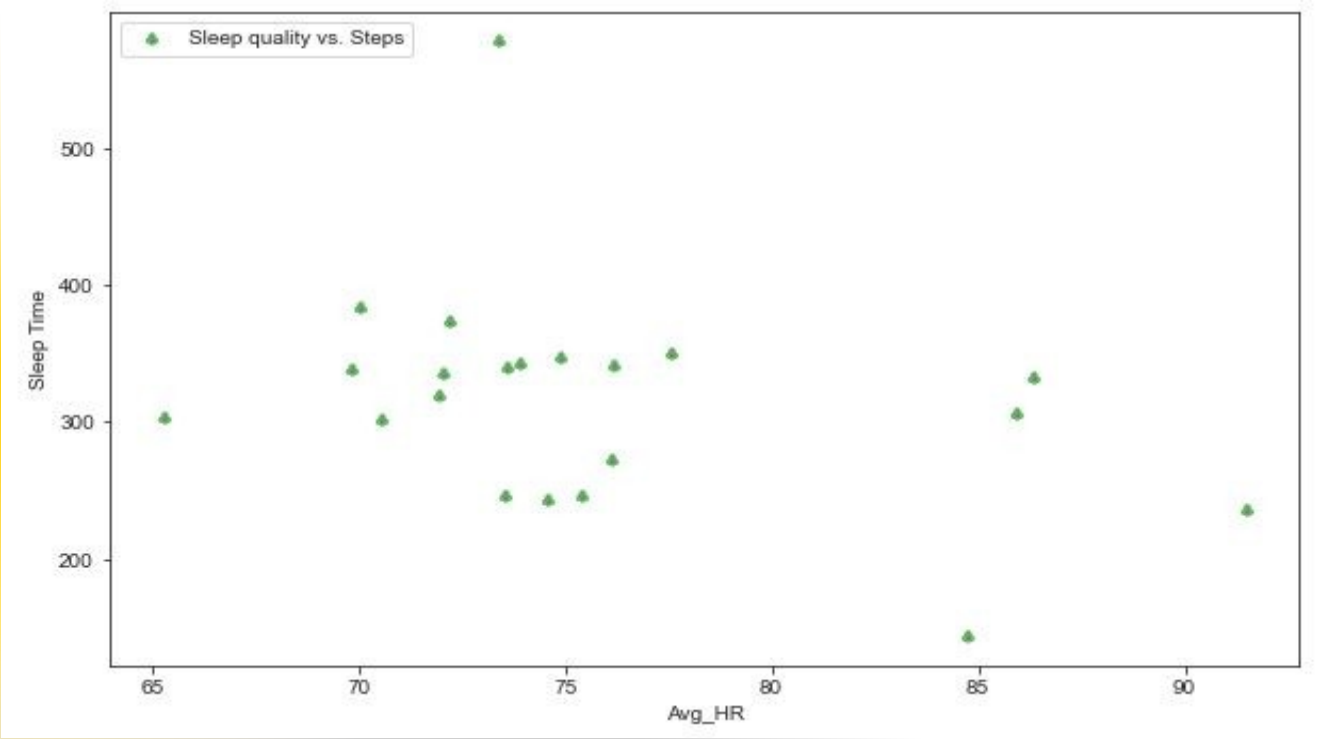


2. Sleep Quality v/s Efficiency Score



The chart depicts the sleep quality as per the user questionnaire and their efficiency score from the analysis. Based on the responses and data collected from a wearable device, it seems that users 9,5 & 6 have lied about their sleep quality since they contradict the efficiency score and answer to the questionnaire.

Sleep Quality vs the Avg HR



Link to the Dashboard

https://datastudio.google.com/u/0/reporting/bfee223f-9c47-4acf-86f8-fc75f215c326/page/vTZ9C

Machine Learning Models

- For building the machine learning model we took Adverse behavior, Efficiency, Is stressful?, State Anxiety, Total Sleep Time (TST), Total Minutes in Bed, Average Awakening Length columns of the sleep data set to predict the **Sleep Quality** of a person.
- We have tried building models with the limited data we have
- We were able to run Linear Regression , Decision Tree classifier, Random Forest classifier, KNN
- Even though random forest classifier gave us an accuracy of 80% its reliability is questionable due the limitations of our dataset quantity.

```
from sklearn.linear_model import LinearRegression
lm = LinearRegression()
model = lm.fit(X_train,y_train)

print(model.predict(X_test))
print(y_test)
```

```
[ ] from sklearn.neighbors import KNeighborsClassifier
    knn = KNeighborsClassifier(n_neighbors=5) # why 5 is because of Elbow method
    knn.fit(X_train,y_train)
    print('test accuracy:', knn.score(X_test,y_test))
```

```
test accuracy: 0.4
```

```
[ ] from sklearn.tree import DecisionTreeClassifier
    dtree = DecisionTreeClassifier()
    dtree.fit(X_train, y_train)
    print('test accuracy:', dtree.score(X_test,y_test))
```

```
test accuracy: 0.6
```

```
[ ] from sklearn.ensemble import RandomForestClassifier
    rfc = RandomForestClassifier(n_estimators=10)
    rfc.fit(X_train,y_train)
    print('test accuracy:', rfc.score(X_test,y_test))
```

```
test accuracy: 0.8
```

Conclusion

- Our dataset had limited data for 22 users. We could only perform simple model analysis to predict the sleep quality.
- Random Forest Classifier gave a better accuracy. However, the train dataset had only 16 records. We suspect overfitting.
- The correlation between the independent variables was very less.
- Decision Tree with various *max_features* were tried out to avoid overfitting.
- Our conclusions are derived from the explorations,
- Sleep quality is measured with efficiency index, i.e., Total Sleep Time / Total Bedtime
- We calculated Ideal Sleep Index, i.e., Total Sleep Time / (8*60) {Ideal Sleep Time}
- In questionnaire data, 68% of the User(s) have mentioned having a better sleep.
- When Total Sleep Time was compared with Ideal Sleep Time, i.e., 8hrs we could see 18/22 User(s) had ideal sleep quality.

Thank You.

“Your future depends on your dreams, so go to sleep.”