DAB 304 – health care analytics project report

Multilevel Monitoring of Activity and Sleep In Healthy People

Section 2-Group 005

Group Members

Weeramundage Ishani Madushika Piyathilake - 0796108

Julie Kunnuvila Thomas - 0792998

Shashvat Sachdeva -0768857

Sneha Sabu – 0792518

1 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. The analysis is carried out adhering to analytical disciplines on data-oriented decision-making to discover and explore for trends and patterns in the data. It is merely fitting that we observe how important health care is for everyone in every aspect of a person's life. We ought to fully make the most of technology feasible because it is fascinating to witness how it develops through time. To determine whether a person experiences a side effect from physical and psychological factors experiments are done and observations are collected to better use of the person and data is collected through wrist worn divices. This is done to ensure that the medication through observation and analysis is risk-free for consumption and has no adverse effects on anyone.

We researched the concept of health care and found the Sleep Quality of a healthy person to be fascinating because of how individuals have been able to cope with external factors affecting a good quality of sleep which then results in sever medical issues for a very long period. Your health, happiness, and mental function all improve with adequate sleep. Lack of regular, good-quality sleep increases the chance of developing a variety of illnesses and disorders. These include dementia and obesity as well as heart disease and stroke. Having a quality sleep gives you many advantages such as improve concentration and productivity, strengthen your heart, Supports a healthy immune system etc.

Based on the 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**. After the finding correlations, our objective is to derive a sleep quality index followed by machine learning models to predict the sleep quality index for a healthy person. The parameters are listed as follows.

- 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

2 Related Work

The Adult Changes in Thought (ACT) Sleep and Activity Monitoring Project employs wearable devices to investigate potential links between brain health and all behaviours occurring within the 24-hour activity cycle (exercise, sedentary behavior, sleep). Previous research has suggested that getting enough sleep and exercising moderately or vigorously can help reduce the risk of cognitive decline. The ACT Study will build on this research by tracking how people's movement, sitting, and sleeping habits interact from day to day and over long periods of time. This project is notable for its use of a device called activPAL to track everyday activity on a large scale 24 hours a day. The project will also collect objective sleep data with the help of a wrist-worn accelerometer sleep watch (Philips Actiwatch Spectrum).

Since 2016, the experienced project team has been collecting accelerometer data on the ACT Study cohort using the activPAL and another device known as ActiGraph. The project will continue to collect activPAL data and will include more study volunteers and new data collected via the sleep watch in the future. In the future, collaboration with ACT Study cores will allow for the exploration of 24-hour activity cycle data with additional life course, neuropathologic, and imaging data. No other study includes such detailed cognitive, physical, and behavioural data over a 24-hour period.

Finally, the findings of this project will provide much-needed evidence to inform the guidance and approaches for older adults to get enough sleep and exercise, which can promote healthy ageing and lower the risk of cognitive decline and late-life dementia.

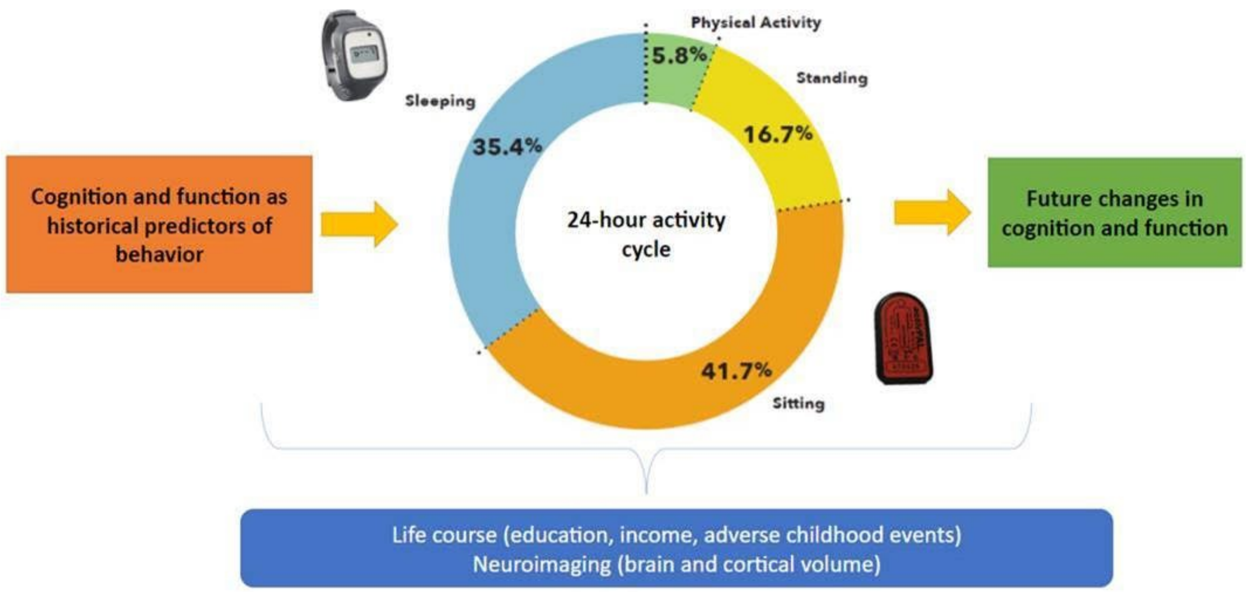


Figure 1: The ACT Sleep and Activity Monitoring Project. Source: https://actagingresearch.org/about/research-program/act-study-research-projects/act-sleep-and-activity-monitoring-project

3 Methods

This section briefly describes the procedures followed, exploratory data analysis, preprocessing techniques, machine learning models, and the assessment metrics we aim to utilize for our data in this part.

3.1 Data Processing

3.1.1 Data Collection

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

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

This data set was obtained from Physionet.Org. 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. The following data in different seven files for each participant has been collected.

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.

3.1.2 Data Preprocessing

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 macro script in Excel**.

Ex: User Info for 22 participants were merged into a single .csv fille

Once the data was merged, they had been loaded to python file to explore the data. For each file,

- Missing Values were checked
- Data Types were checked and Transformed the Date columns to **DateTime** format.
- converted the time columns into seconds.
- New columns were added accordingly
    - **User Info** – Age Groups and Weight Groups columns were added
    - **Sleep data** – Efficiency column was added (This is used to derive the Sleep Quality index)
    - Activity – Replaced Activity id with the activity description.
- Inaccurate data inputs checked
- Data misinterpretation checked

3.2 Importing Libraries, Packages, and Dataset

We imported the necessary libraries and packages that are required to run the code sections in the Python notebook (ipynb file) as follows.

- import os
- import pandas as pd
- import seaborn as sns
- import matplotlib.pyplot as plt
- import numpy as np
- from collections import Counter
- import numpy
- import scipy
- from scipy import stats
- import warnings
- warnings.filterwarnings("ignore")

3.3 Exploration and Analysis

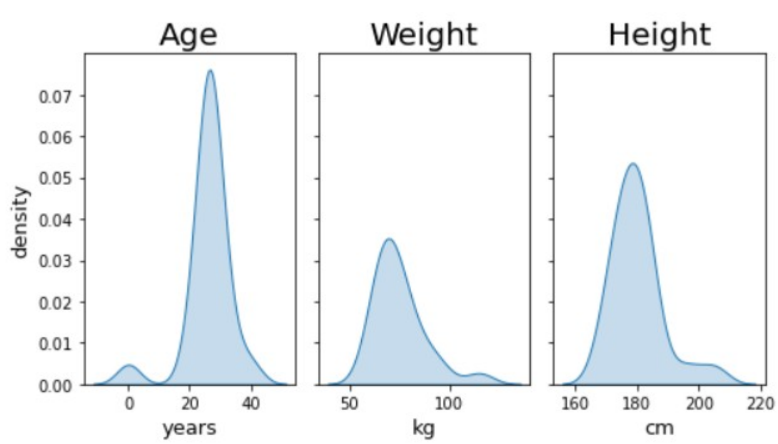
Data Exploration was on the datasets as this is the key aspect of data analysis and model building. Exploring the data helped us understand the data and the correlations that exist between them to come to conclusions and build the model.

3.3.1 User Info

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

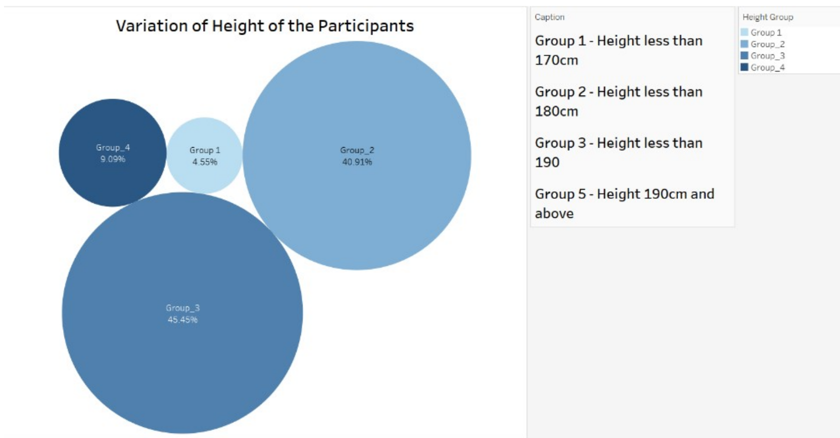
```
df_userinfo.describe().T # Descriptive statistics
```

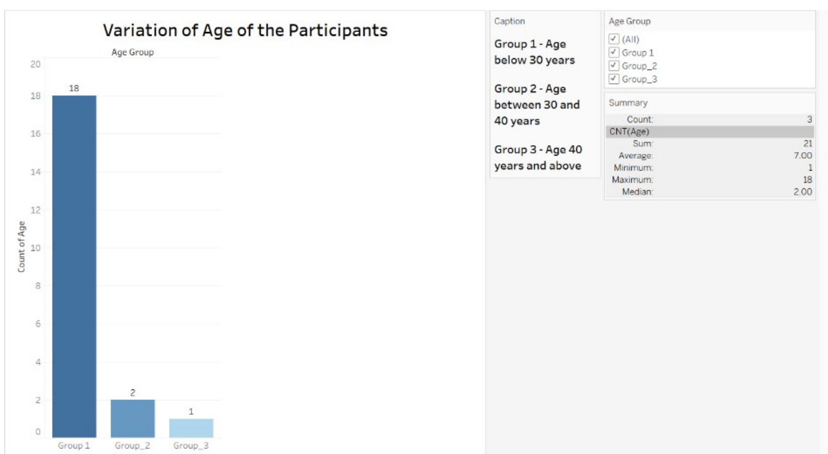
	count	mean	std	min	25%	50%	75%	max
Weight	22.0	75.045455	12.789420	60.0	67.0	70.0	80.00	115.0
Height	22.0	179.909091	8.216760	169.0	175.0	180.0	183.00	205.0
Age	22.0	26.045455	7.121244	0.0	25.0	27.0	27.75	40.0



Graphical distribution of the following aspects was done for the participants using Tableau.

- Age
- Height



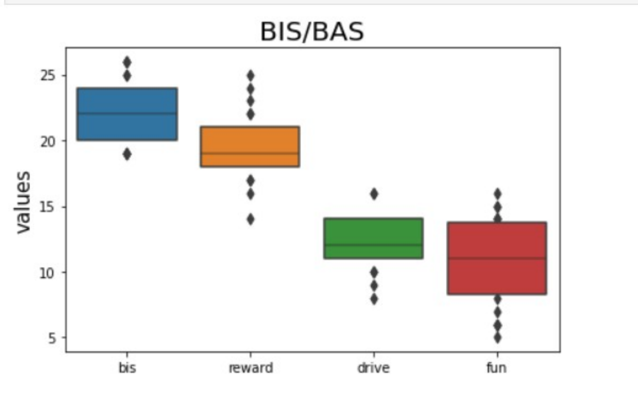


3.3.2 Questionnaire

BIS/BAS: Behavioral avoidance/inhibition index was explored. (BIS/BAS scales are a typical measure of reinforcement sensitivity theory that establish biological roots in personality characteristics, derived from neuropsychological differences)

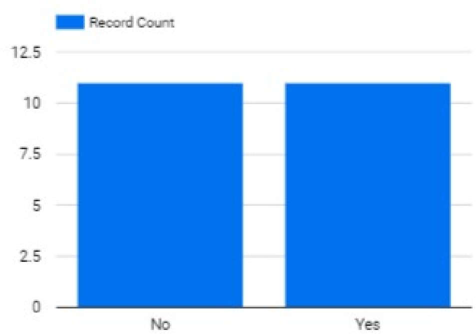
- **Bis facet** reflects subject sensitivity toward aversive events that promote avoidance behaviors.
- **Drive** describes individual persistence and motivational intensity.
- **Reward** corresponds to Reward Responsiveness which indicates a propensity to show a higher degree of positive emotion for goal attainment.

- **Fun** corresponds to Fun-Seeking is related to impulsivity and immediate reward due to sensory stimuli or risky situations

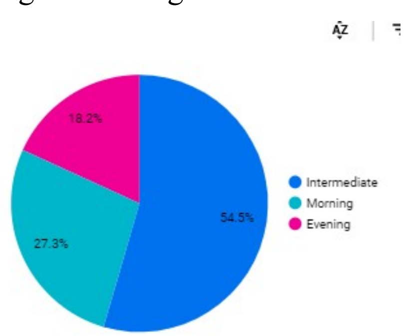


Users' Anxiety, Morningness-Eveningness, and Sleep Quality Questionnaires were explored, and following is the analysis.

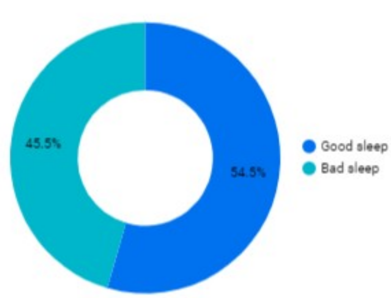
- Half of the Participants answered that they are stressed out, and others are not.



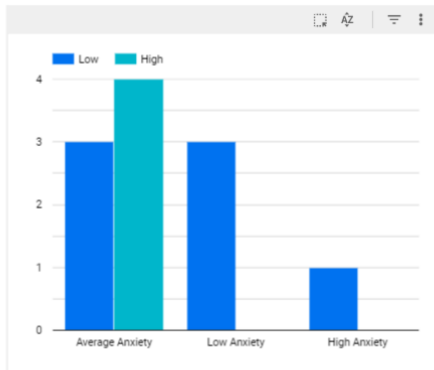
- More than 50% of the participants are in intermediate as they do not have a preference in Morning or Evening.



- More than 50% of the participants are having a Good Sleep as per the Questionnaire.



- People who have average anxiety tend to have more responsiveness toward any activity which stimulates their anxiety



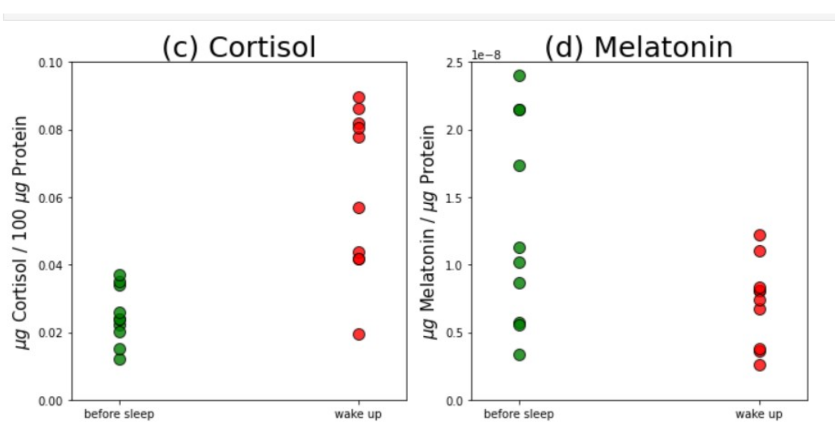
3.3.3 Saliva

Following is the Descriptive statistic of hormone concentration before sleep and at wake-up time.

```
df_Saliva.groupby('SAMPLES').describe().T
```

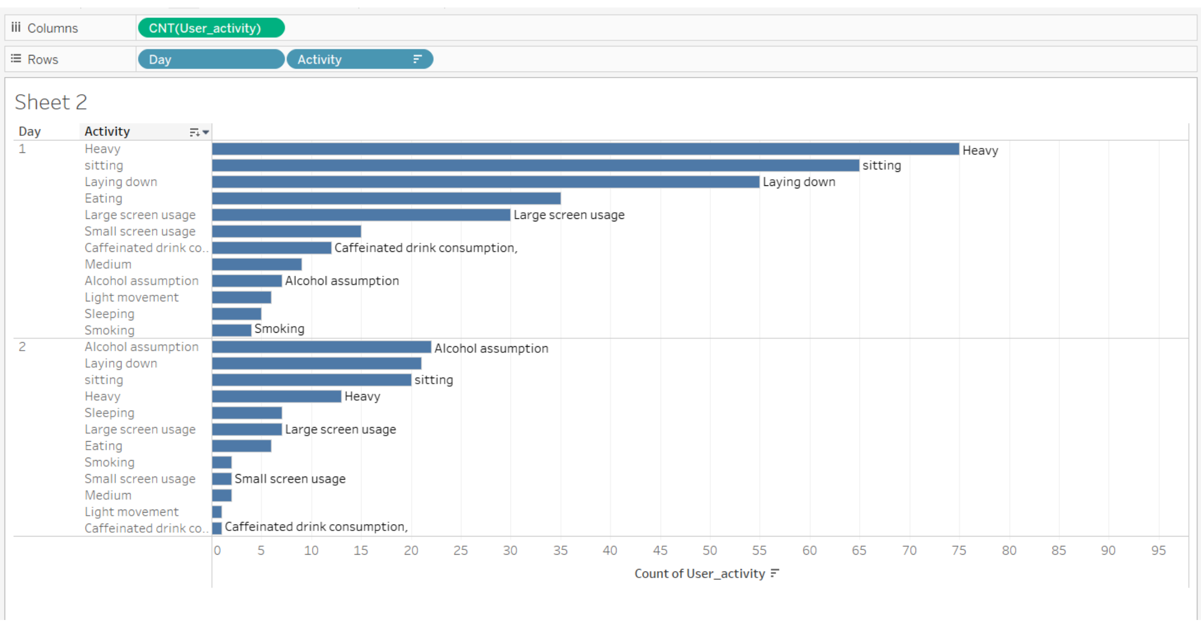
	SAMPLES	before sleep	wake up
Cortisol NORM	count	1.000000e+01	1.000000e+01
	mean	2.494971e-02	6.203210e-02
	std	8.362741e-03	2.431014e-02
	min	1.214184e-02	1.965878e-02
	25%	2.073234e-02	4.224304e-02
	50%	2.377192e-02	6.750675e-02
	75%	3.209440e-02	8.155590e-02
	max	3.704592e-02	8.951608e-02
Melatonin NORM	count	1.000000e+01	1.000000e+01
	mean	1.292000e-08	7.181000e-09
	std	7.570674e-09	3.132643e-09
	min	3.330000e-09	2.570000e-09
	25%	6.450000e-09	4.552500e-09
	50%	1.075000e-08	7.725000e-09
	75%	2.047500e-08	8.285000e-09
	max	2.400000e-08	1.220000e-08

Line plot of Cortisol and Melatonin concentrations were obtained from saliva sample. Difference between "before sleep" and "wake up" hormones concentration done on Python.



3.3.4 Activity

The following is the distribution of list of the activity categories throughout the day of all users done on Tablue.

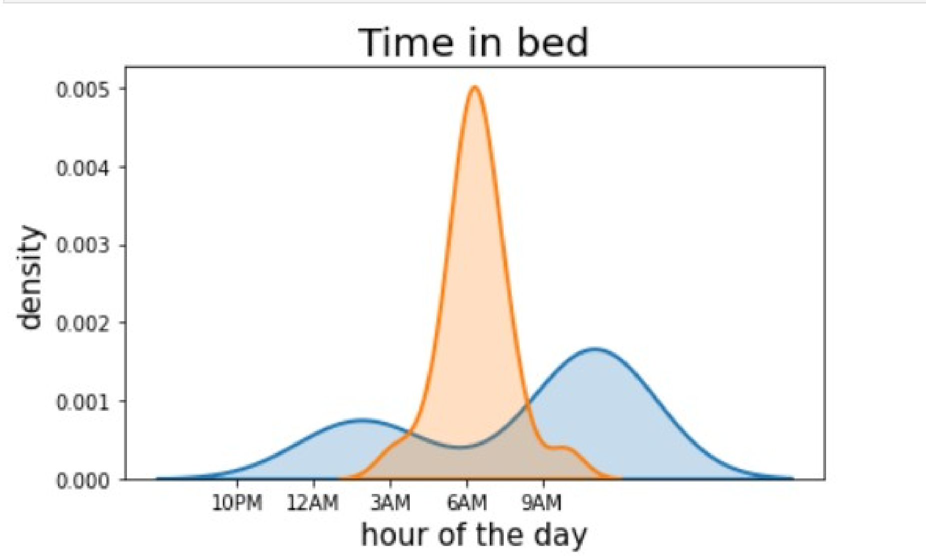


3.3. Sleep

Statistical distribution and Density distribution of Sleep data are as follows.

```
df_sleep.describe().T
```

	count	mean	std	min	25%	50%	75%	max
Participant ID	22.0	11.045455	6.869375	1.000	5.2500	11.0000	16.75000	22.000
In Bed Date	22.0	1.818182	0.394771	1.000	2.0000	2.0000	2.00000	2.000
In Bed Time	22.0	540.818182	293.659932	63.000	179.2500	722.0000	751.50000	777.000
Out Bed Date	22.0	1.818182	0.394771	1.000	2.0000	2.0000	2.00000	2.000
Out Bed Time	22.0	424.227273	88.279282	211.000	395.2500	422.0000	471.00000	660.000
Onset Date	22.0	1.818182	0.394771	1.000	2.0000	2.0000	2.00000	2.000
Onset Time	22.0	509.590909	305.914501	61.000	151.2500	708.5000	751.00000	778.000
Latency	22.0	1.500000	1.711307	0.000	0.0000	0.5000	3.00000	4.000
Efficiency	22.0	83.906818	6.746207	73.490	77.1600	85.2200	89.06000	94.230
Total Minutes in Bed	22.0	374.318182	96.184043	165.000	329.7500	368.5000	431.00000	630.000
Total Sleep Time (TST)	22.0	313.000000	84.309520	144.000	253.5000	326.0000	342.75000	578.000
Wake After Sleep Onset (WASO)	22.0	59.818182	30.504701	17.000	40.0000	52.5000	77.75000	118.000
Number of Awakenings	22.0	19.272727	9.779385	4.000	12.2500	18.5000	21.00000	44.000
Average Awakening Length	22.0	3.555455	2.386114	1.330	2.0750	2.8050	4.24750	12.250
Movement Index	22.0	13.506273	4.379958	6.734	9.1620	13.1900	17.37775	20.669
Fragmentation Index	22.0	10.333455	9.224163	0.000	0.0000	9.7620	15.68800	28.125
Sleep Fragmentation Index	22.0	23.839727	11.494290	6.734	16.1865	22.1165	30.81975	45.526



3.4 Machine Learning Models:

Machine Learning is a field which is used for building a model that can process sample data (Training data), to train the model for most accurate predictions using for test data. For our project we used 4 Machine Learning Models as follows.

- **Random Forest (RF)**
- **Decision Tree Classifier**
- **Linear Regression (LR)**
- **K-Nearest Neighbors (KNN)**

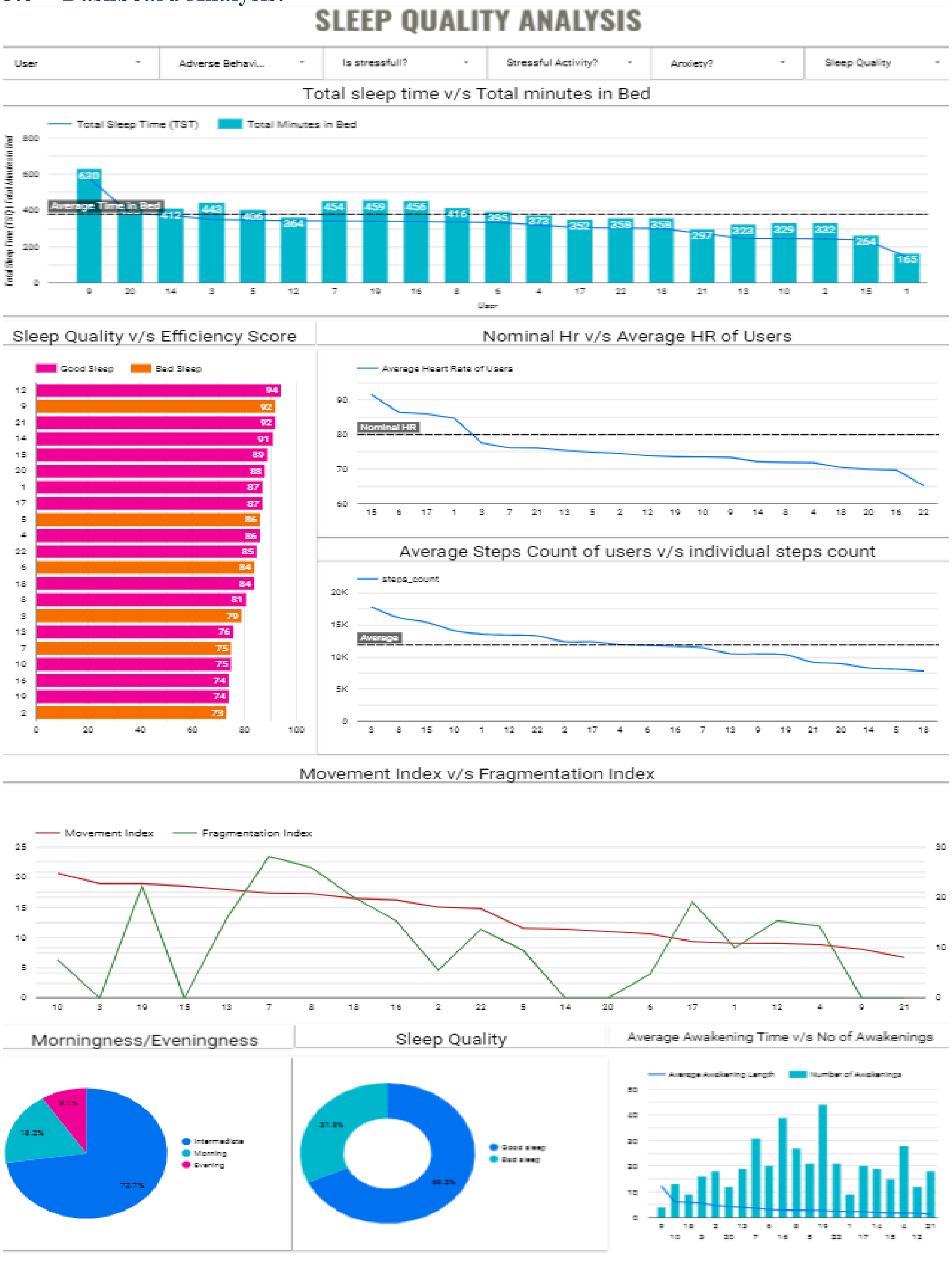
For building the machine learning model we took Adverse behavior, Efficiency, Is stressful? State Anxiety, Total Sleep Time (TST), Total Minutes in Bed, Average Awakening Length columns of the sleep data set to predict the Sleep Quality of a person

3.5 Evaluation Metrics:

These measures must be used since they provide a clearer picture of how each model is operating at its highest level and what needs to be changed to raise the scores later. The measures we used are as follows.

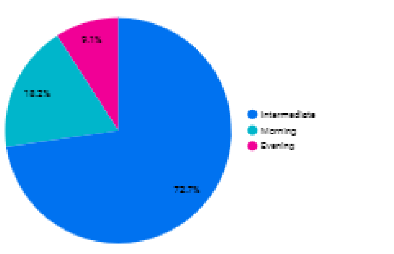
- **F1 Score**
- **Confusion Matrix**
- **Mean Absolute Error**
- **Receiver Operating Characteristics (ROC) Curve**

3.6 Dashboard Analysis:

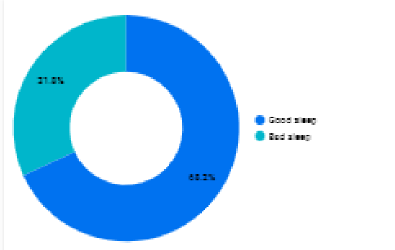


USER QUESTIONNAIRE ANALYSIS

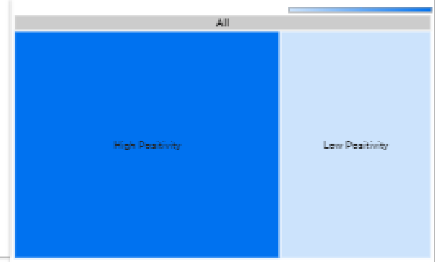
Morningness/Eveningness



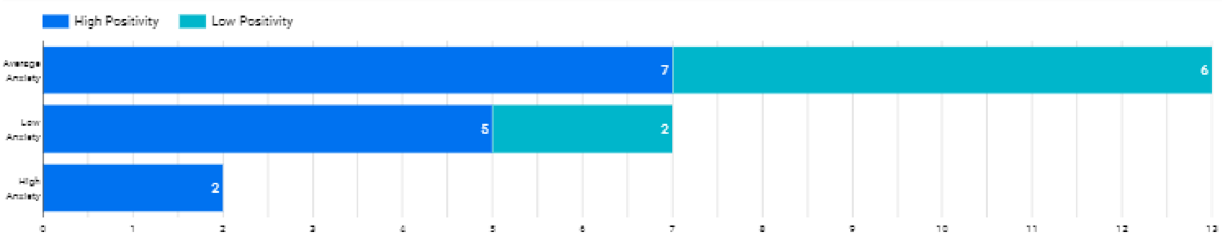
Sleep Quality



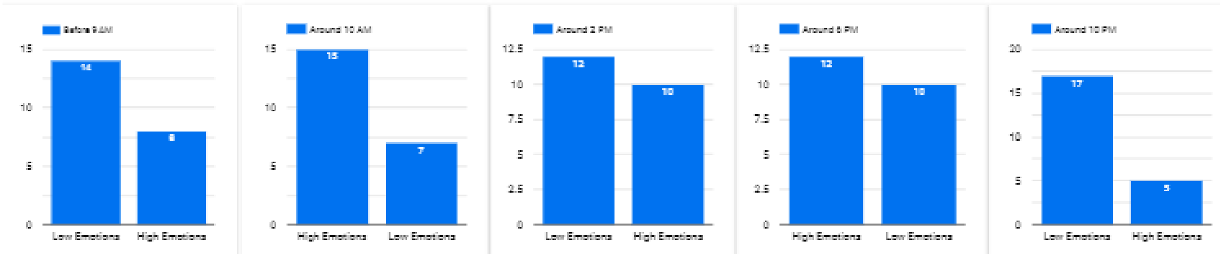
Positive/Negative Attribution



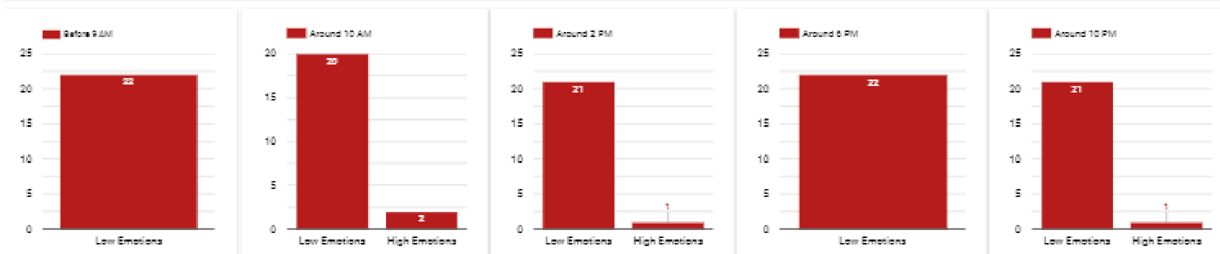
Anxiety Level v/s Avoidance Behavior



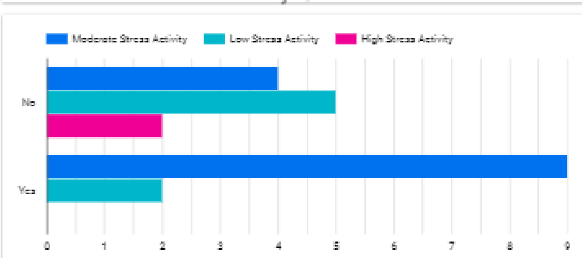
Perceived Emotions towards Positive News



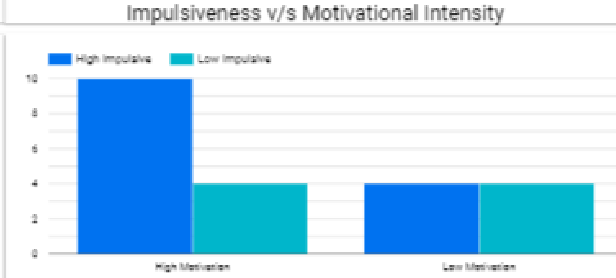
Perceived Emotions towards Negative News



Stressful Activity v/s Stressfulness



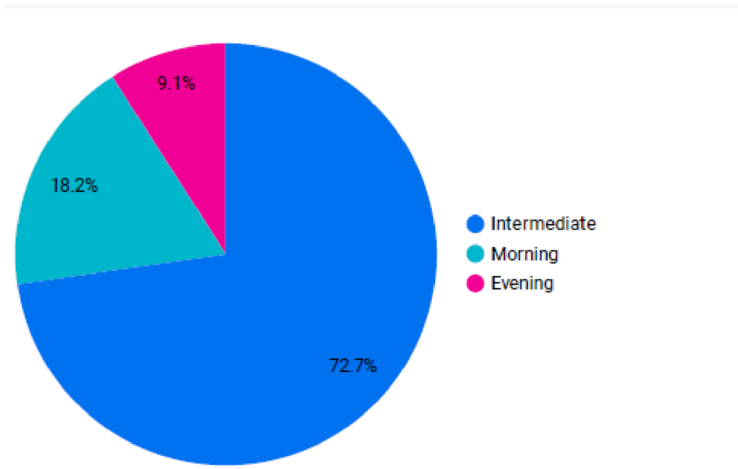
Impulsiveness v/s Motivational Intensity



3.6.1 QUESTIONNAIRE

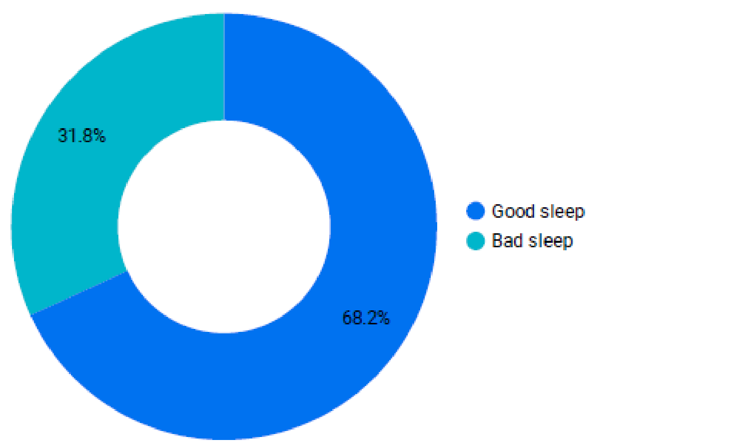
Visual Insights:

1. **Morningness/Eveningness –**



Out of 21 users, only 18.2% of them said that they were Morning People. Moreover, only 9.1% said they were Evening People. The major proportionate of the users were Intermediate.

2. **Sleep Quality**

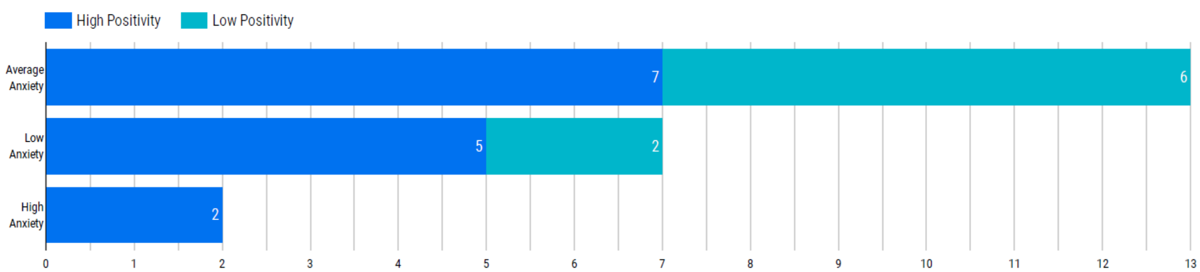


Out of all the users, 68.2% people said they had a good sleep. However, only 31.8 percent of users said that they had a bad sleep.

3. **Positive/Negative Attribution**

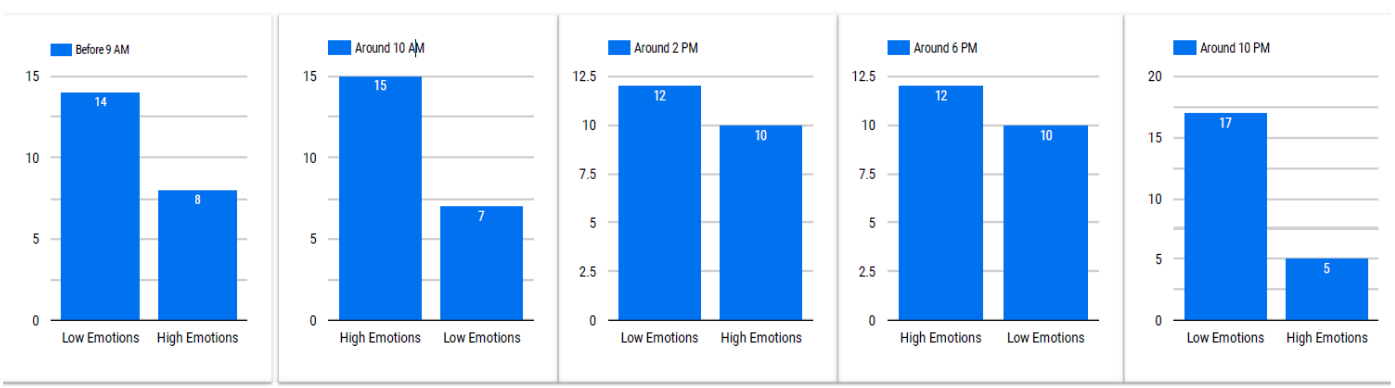
Out of all the users, nearly 60% of users had a High positive attribute throughout the day.

4. **Anxiety Level v/s Avoidance Behavior**



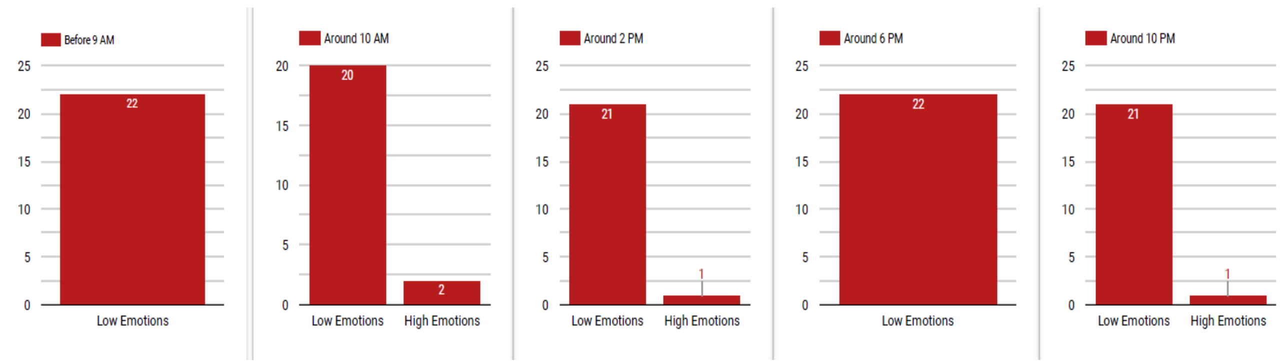
Among the users, a major proportionate of them had the average anxiety level throughout the day. Whereas the least of them had a high anxiety level. Users with high anxiety levels didn't show any positivity towards avoidance behavior. However, users with low anxiety showed comparatively high positivity.

5. **Perceived Emotions towards Positive News:**



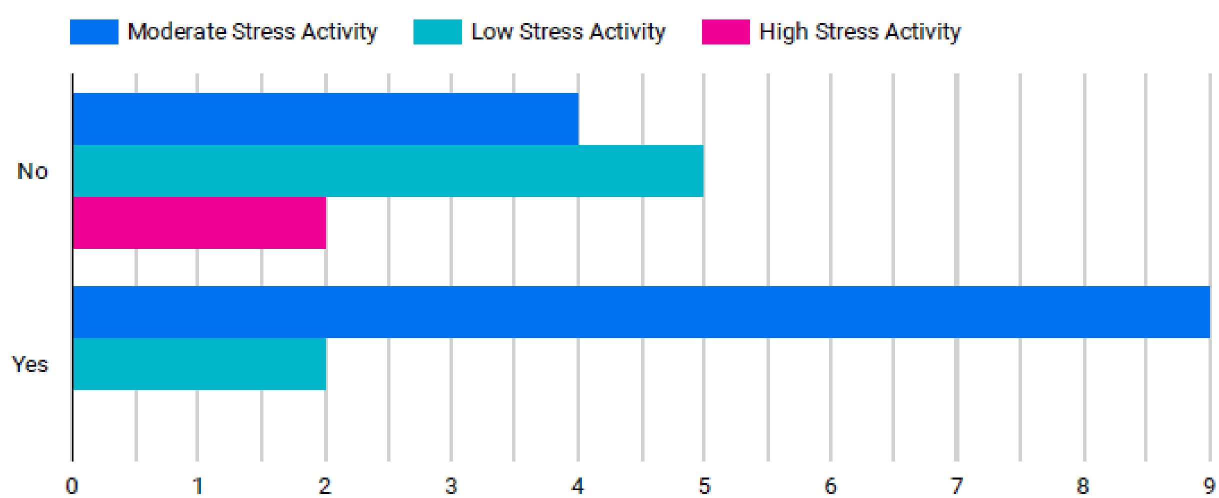
All the users had a slightly similar pattern toward their emotions for any positive news throughout the day. All the users showed high emotions around 10 am & 6 pm and less high emotions at night-time.

6. Perceived Emotions towards Positive News:



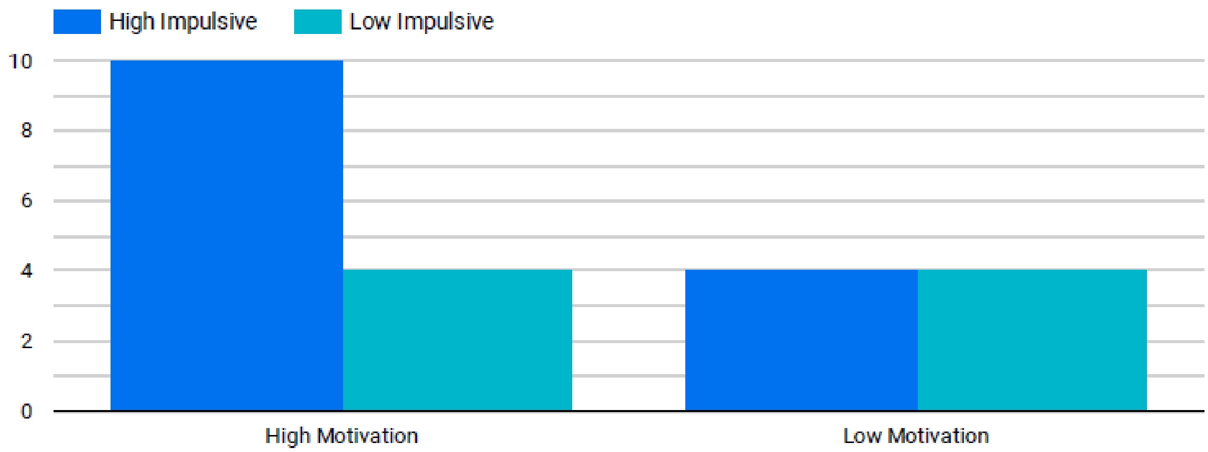
All the users comparatively showed low emotions throughout the day with the least high emotions for any negative news. All the users showed no high emotions before 9 am and around 6 pm. Moreover, they showed the most comparative low emotions at nighttime.

7. Stressful Activity v/s Stressfulness



Users who answered that they were stressed throughout the day had moderately stressful activity and surprisingly no highly stressful activity. However, of users that answered that they didn’t have a stressful day, among them ~20% of them answered they had a high-stress activity.

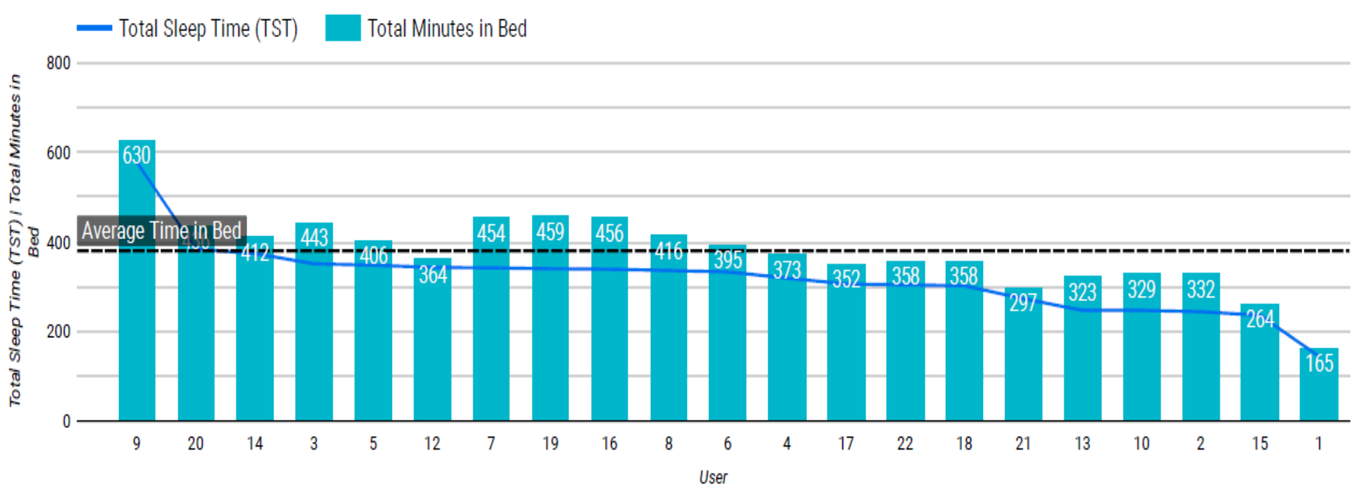
8. Impulsiveness v/s Motivational Intensity



Users who had low motivation throughout the day had an equal proportionate of impulsiveness. However, users with high motivation had, most of them had high impulsiveness.

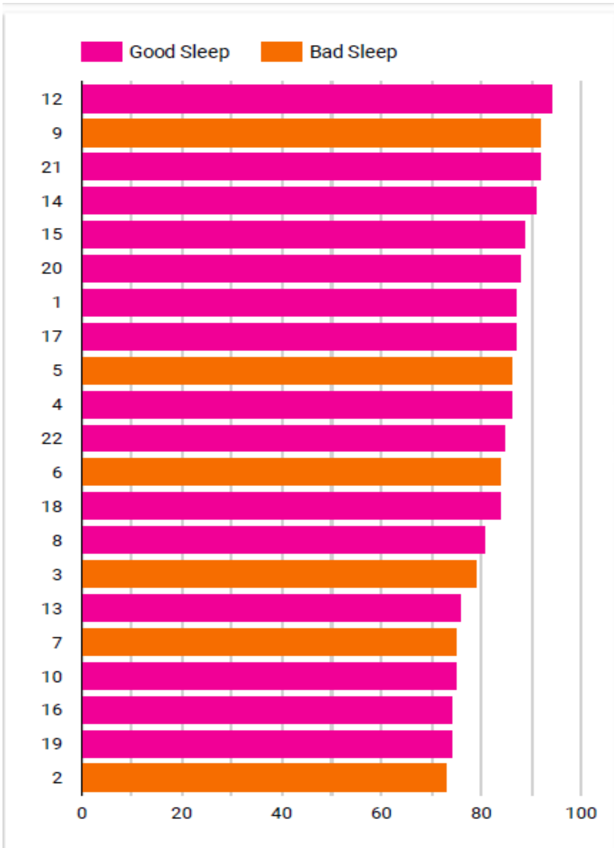
3.6.2 SLEEP ANALYSIS

1. Anxiety Level v/s Avoidance Behavior



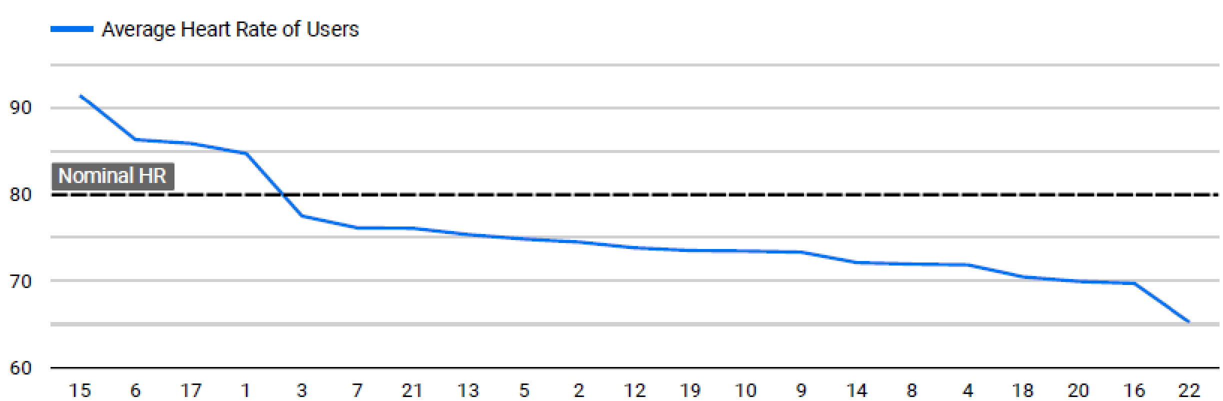
The graph depicts the average time of students in the bed v/s their total time in bed v/s their total sleep time. Less the difference between sleep time and time in bed is considered efficient sleep. User 9 had the most time in bed and efficient sleep and user 1 had the least.

2. **Sleep Quality v/s Efficiency Score**



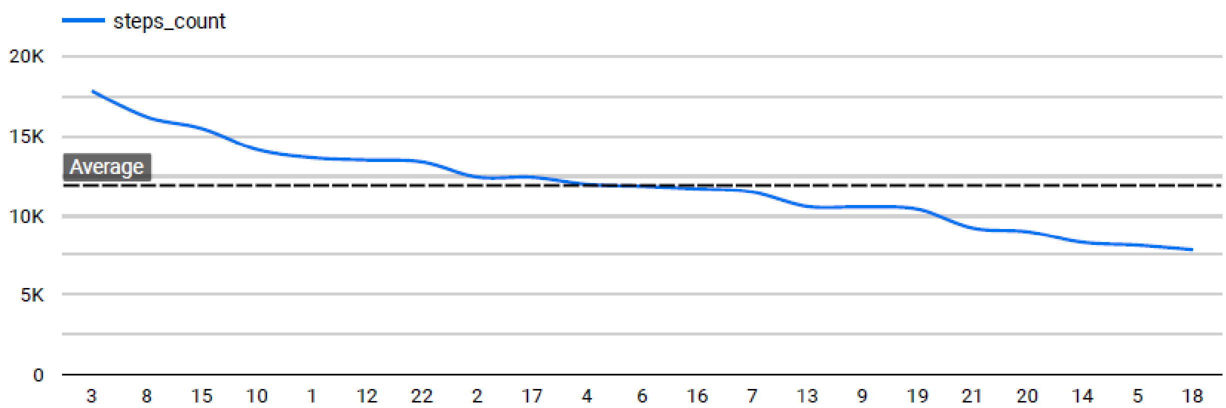
The chart depicts the sleep quality as per the user questionnaire and their efficiency score from the analysis. Based on the responses and data collected from a wearable device, it seems that users 9,5 & 6 have lied about their sleep quality since they contradict the efficiency score and answer to the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

3. **Nominal HR v/s Average HR of Users**



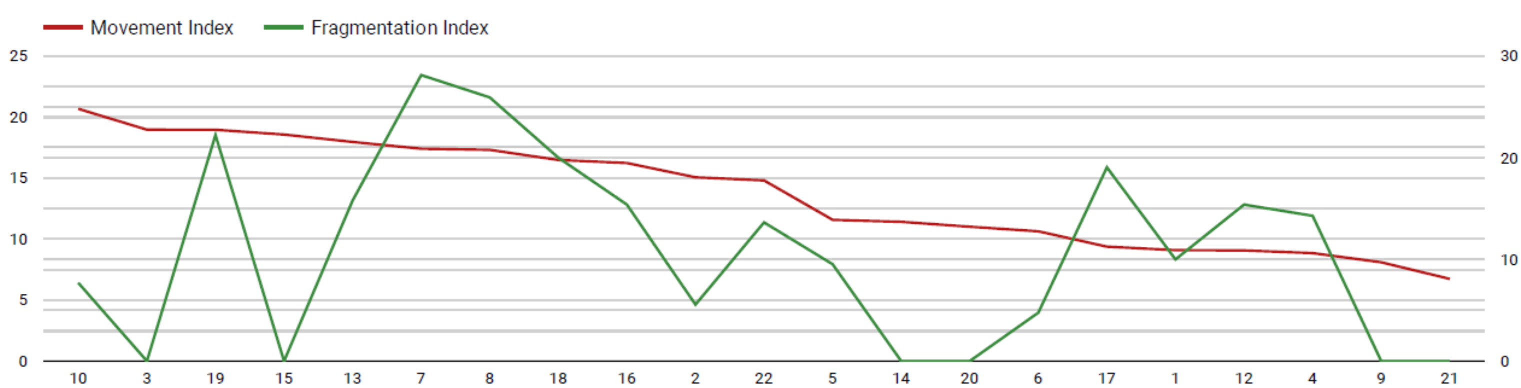
The following chart depicts the Average Heart rate of the users throughout the day and Nominal HR as per Wikipedia.com. Seems that users 1 & 3 had a nominal heart rate throughout the day. User 15 has the highest and user 22 has the least average heart rate throughout the day.

4. **Average Steps Count of user's v/s individual steps count**



The line chart shows the total step count of the users throughout the day v/s the average for all users. User 3 had the greatest number of steps throughout the day compared to user 18 having the least.

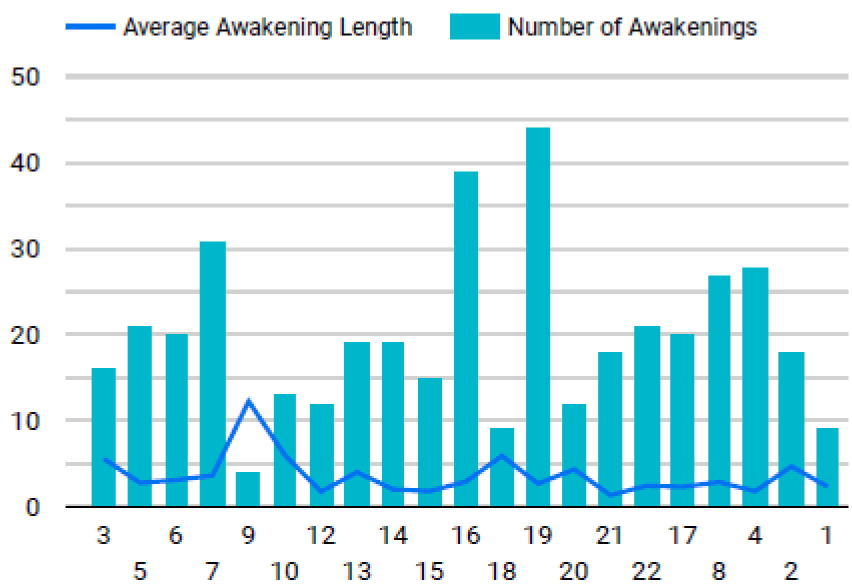
5. **Movement Index v/s Fragmentation Index**



The line graph shows the Movement index which is movement during the night-time when they were awake and the fragmentation index, which implies that no movement was there when they were awake (idle time). User 10 had the greatest number of movements during the awake time compared to user 21.

6. Average Awakening Time v/s No of Awakenings

The graph depicts the number of awakenings of each user (bars) and the average awakening length in time for all users(line). Users 19 & 16 had the greatest number of awakenings throughout the time and user 9 had the least.



4 Results

4.1 Multilevel Monitoring of Activity and Sleep using Data Exploration and Analysis

Analysis	Results
User Info (Anthropometric characteristic)	Statistical Distribution concludes that the average (mean) of the. • Height of the participant – 179 cm • Weight of the participant – 75kg • Age of the participant – 26 years Majority of the Participants belong to height group less than 190cm and minority of the participants belong to less than 170cm. Age analysis resulted such that highest no of participants belong to age group below 30 and lowest no of participants belong to age group above 40 years.
Scores for all the questionnaires	• Most of the people who tend to say they are stressed are morning people and tend to have better sleep and are more of morning people. People who tend to that are not stressed are more intermediate people and proportionately more of them tend to have good sleep than the ones who have bad sleep.

	• People who have average anxiety tend to have more responsiveness toward any activity which stimulates their anxiety. People who are stressed tend to have the least persistence in any unpleasant activity and that stimulates their anxiety level. • Behavioral avoidance/inhibition analysis results that, Overall participants are persistent, motivational intensity, and higher degree of positive emotion for goal attainment.
Activity	• Following Unique Activities where available, Sleep, In Bed, Non-idle sitting, light movement, medium movement, heavy movement, eat, screen time, screen time Caffeine, smoke, alcohol. • User activity gave us an overview of how a user's day was spent with various activities listed above. • Our agenda behind analyzing activities was to identify whether stressful activities affect user's sleep. However, we did not find an evident relation of activity vs sleep.
Sleep	• Quality of sleep is measured with efficiency index, i.e., Total Sleep Time / Total Bedtime • The Average of Efficiency, Total Minutes in Bed, Total Sleep Time, Average Awakening Length are 83.42, 382, 318.58, 3.63 respectively. • The mean value of efficiency tells us that overall the users in observation had a better sleep.

4.2 Multilevel Monitoring of Activity and Sleep using Machine Learning Models

Machine Learning Model	Accuracy Achieved
Random Forest (RF)	**Accuracy: 80%**
Decision Tree Classifier	**Accuracy: 60%**
Linear Regression (LR)	**Accuracy: 23%**
K-Nearest Neighbors (KNN)	**Accuracy: 40%**

5 Discussion

The dashboard was made with the intention of analyzing the questionnaire and contrasting the findings with the information gathered by the wearable devices for 22 users. Additionally, the main goal was to find a link between the characteristics and the data from the questionnaire. However, after visualizing the results, we discovered several outcomes that were in conflict. By contrasting user responses to the questionnaire with the information gathered, the discrepancies were examined. We discovered that certain users' responses to their sleep quality for that day were inconsistent with the efficiency as determined by technology. Furthermore, individuals who engaged in high-stress activities experienced stress throughout the day since their average heart rates were higher than the Nominal Heart Rate, according to Wikipedia. Additionally, individuals that exhibit good unfavorable behavior throughout the day take more steps than the average user. In conclusion, we identified several behaviors that have the potential to affect how well someone sleeps, including general stress levels, how someone reacts to good or bad news, and how likely they are to express more positive emotions while achieving their goals.

6 Conclusion

- Our dataset had limited data for 22 users. We could only perform a simple model analysis to predict sleep quality.
- Random Forest Classifier gave 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 the 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.

7 Contributions

Name	Contribution
Julie K. Thomas	Project work – • Merged all Data 22 Users Data into Single Data File • Created and evaluated various classifier Model • EDA with User Activity Final Report – • Conclusion • Analysis description of various datafiles
Shashvat Sachdeva	Project work – • Creating Exploratory Analysis using python • Creation of a dashboard to visualize data using Looker Studio Final Report – • Finalizing results using dashboard • Creating a storyline using the dashboard.
Weeramundage Ishani Madushika Piyathilake	Project work – • Data Merging User Info and Activity • Cleaning the Data in User Info and Activity and Saliva • Exploring the data for User Info and Activity and Saliva • Create visualizations and finding correlations between the above data • Preparing Sleep and Questionnaire data for modeling • Researching to derive a Sleep Quality Index for the model Final Report – • Introduction

	• Methods • Results: Multilevel Monitoring of Activity and Sleep using Data Exploration and Analysis
Sneha Sabu	Project work – • Presentation Files and Story flow • Preparing Sleep and Questionnaire data for modeling • Creating Exploratory Analysis using python Final Report – • Related Work

8 References

1. *Horne JA, Ostberg O (1976). "A Self Assessment Questionnaire to DetermineMorningness Eveningness in Human Circadian Rhythms". Int J Chronobiology. 4: 97–110. PMID 1027738.*
2. *CD Spielberger, Gorsuch RL, Lushene RE, Vagg PR, Jacobs GA (1983). "Manual for the State-Trait Anxiety Inventory. (Form Y)". Palo Alto, California: Consulting Psychologists Press.*
3. *Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ (1989). "The Pittsburgh Sleep Quality Index: A New Instrument for Psychiatric Practice and Research". Psychiatry Res. 28: 193-213. DOI 10.1016/0165-1781(89)90047-4.*
4. *Carver CS, White TL (1994). "Behavioural inhibition, behavioural activation, and affective responses to impending reward and punishment: The BIS/BAS Scales". J Pers Soc Psychol. 67: 319–333. DOI 10.1037/0022-3514.67.2.319.*
5. *Brantley PL, Waggoner C, Jones GN, Rappaport N (1989). "A Daily Stress Inventory: Development, reliability, and validity". J Behav Med. 10: 61-73. DOI 10.1007/BF00845128.*
6. *Crawford JR (2004). "The Positive and Negative Affect Schedule (PANAS): Construct validity, measurement properties and normative data in a large non-clinical sample". Br J Clin Psychol. 43: 245–265. DOI 10.1348/0144665031752934*
7. *Sleep and Activity Monitoring Project | ACT Study. (n.d.). https://actagingresearch.org/about/research-program/act-study-research-projects/act-sleep-and-activity-monitoring-project*