Data Analytics
Project Presentation

FitBit Fitness Analysis

Abhinav Gunti (PES1UG20CS008)
Abhishek Singhi (PES1UG20CS011)
Aniket Acharya (PES1UG20CS052)

Abstract and Scope

Fitbits have become very sought after worldwide since the general public has become more aware regarding tracking their health. Similarly, they are also popular amongst fitness experts, physiologists and data scientists for the purpose of data collection. The Fitbit has been designed to provide as intricate of details of the user down to a given minute, making outliers in data easy to view and finding patterns within the given data.

The scope it entails is as follows:

The reason the Fitbit is in such demand is due to its easy-to-use interface (regarding the general public) and the vast insights that can be taken from a user’s Fitbit data history (regarding academic experts). The study done below is for providing the user recommendations for their daily activity based on their various tracked health stats and fitness statistics.

Implementation Details

Front-End:

- The front end of this project has been created by Streamlit which is an open source app framework in Python language.

- Streamlit is an open-source app framework for Machine Learning and Data Science teams. Create beautiful web apps in minutes.

- Each individual component with respect to our UI has been made selectively.

Back-End:

- The back end of our project takes use of python through thorough .ipynb and .py files performing data analysis, inferences and predictions through our models and datasets utilized.

- Back-End interacts with prediction models for heart disease to give outputs.

Implementation Details

ML Models Used:

- **K-Nearest Neighbours (KNN):**

The k-nearest neighbors algorithm, is a non-parametric, supervised learning classifier, which uses proximity to make classifications or predictions about the grouping of an individual data point. Heart disease prediction is performed using this model.

- **Multiple Linear Regression (MLR):**

The multiple linear regression (MLR) model assumes that in addition to the p independent x-variables, a response variable y is measured, which can be explained as a linear combination of the x-variables.

The sleep score prediction depends on multiple independent variables like REM Sleep and Minutes Asleep.

Project Demonstration

fitbit

Fitbit Fitness Analysis

About the data | Data Exploration using Visualizations | Cleaning and Processing | Analysis and Visualizations | Conclu

About the data

This Kaggle data set contains personal fitness tracker data from thirty fitbit users. This dataset generated by respondents to a distributed survey via Amazon Mechanical Turk between 03.12.2016-05.12.2016. Thirty eligible Fitbit users consented to the submission of personal tracker data, including minute-level output for physical activity, heart rate, and sleep monitoring. Variation between output represents use of different types of Fitbit trackers and individual tracking behaviors / preferences. It includes information about daily, hourly and minute activity, steps, and heart rate that can be used to explore users' habits.

Data is stored in multiple CSV files, according to the frequency and type of the collected information.

Major Case study goals

Project Demonstration

Fitbit Fitness Analysis

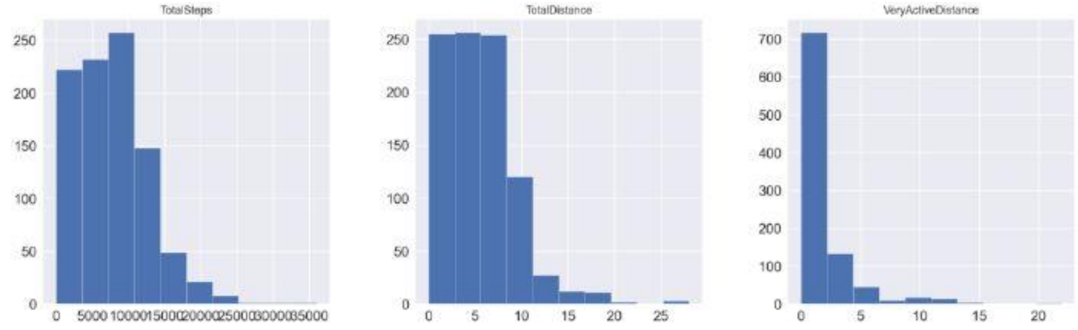
About the data | Data Exploration using Visualizations | Cleaning and Processing | Analysis and Visualizations | Conclu

Data Exploration using Visualizations

Data analysis using graphs and charts prove to be very useful to look at your data. We shall be using Matplotlib and Seaborn for our analysis.

Histograms

Let's start by using histograms to see if the data follows a particular kind of distribution. Since we have a lot of features, let's extract a subset for our analysis.



Project Demonstration

Fitbit Fitness Analysis

About the data | Data Exploration using Visualizations | Cleaning and Processing | Analysis and Visualizations | Conclu

Analysis and Visualizations

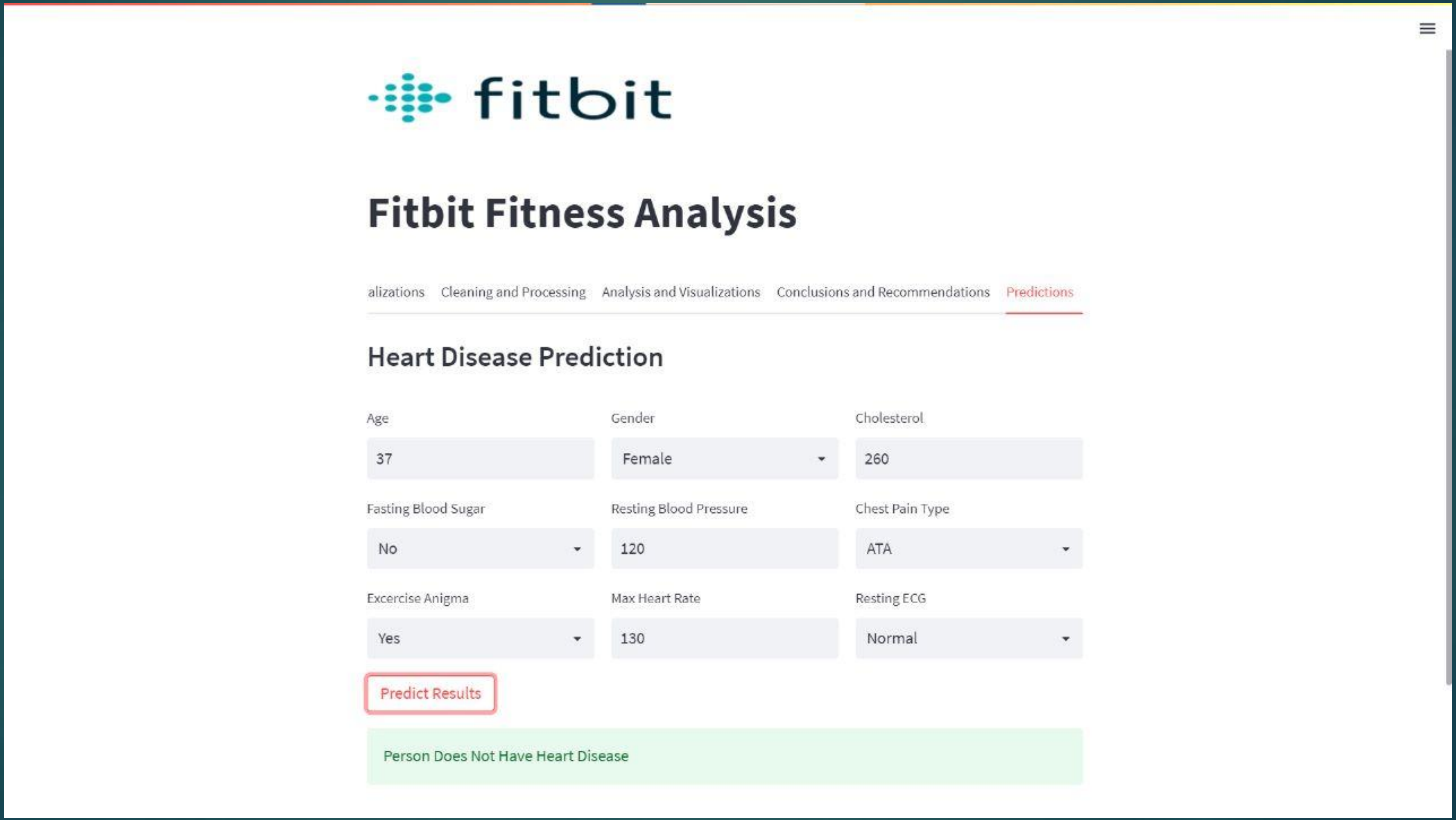
Active and wear time

Users active time and wear time can be calculated using minutes spent on each activity type.

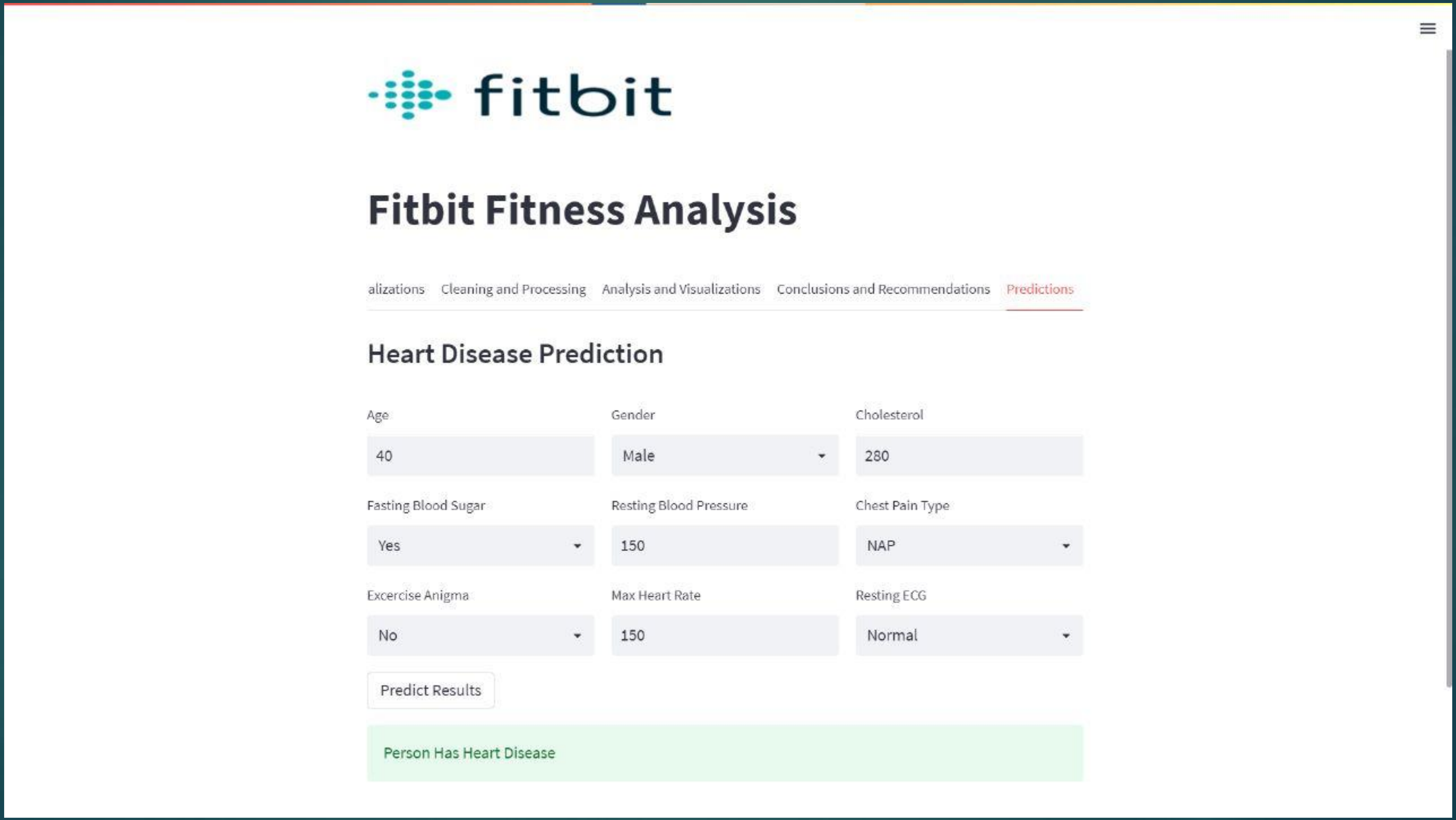
Group the daily data by user Id to get average numbers in each observation category:

	Id	TotalSteps	TotalDistance	TrackerDistance	LoggedActivitiesDistance	VeryActiveDistance
0	1503960366	12,116.7419	7.8097	7.8097	0.0000	2.8584
1	1624580081	5,743.9032	3.9148	3.9148	0.0000	0.9394
2	1644430081	7,282.9667	5.2953	5.2953	0.0000	0.7300
3	1844505072	2,580.0645	1.7061	1.7061	0.0000	0.0084
4	1927972279	916.1290	0.6345	0.6345	0.0000	0.0958
5	2022484408	11,370.6452	8.0842	8.0842	0.0000	2.4216
6	2026352035	5,566.8710	3.4548	3.4548	0.0000	0.0061
7	2320127002	4,716.8710	3.1877	3.1877	0.0000	0.1068

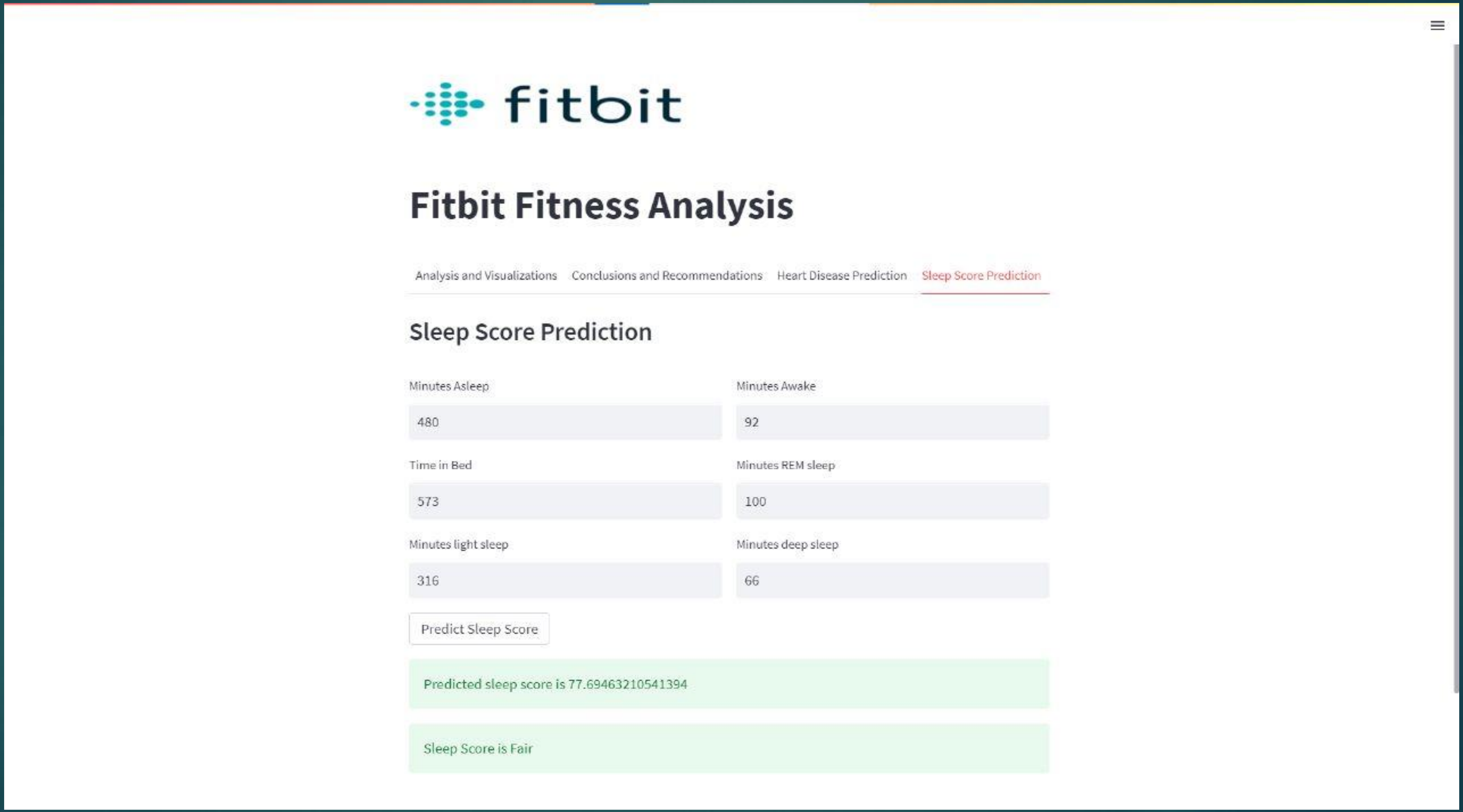
Project Demonstration



Project Demonstration



Project Demonstration



Results

With this project, we were able to analyse data presented to us across various datasets and provide conclusive inferences and visualizations to the various parameters of fitness we have considered from the FitBit dataset.

We also were able to use prediction models like KNN and MLR to make predictions on heart rate conditions and sleep score.

Conclusion

The use of Fitbit devices in interventions has the potential to promote healthy lifestyles in terms of physical activity and weight. Fitbit devices may be useful to health professionals for patient monitoring and support.

The objectives of this analysis are

(1) to assess the effectiveness of interventions that incorporate a Fitbit device for healthy lifestyle outcomes (eg: steps, moderate-to-vigorous physical activity, and weight)

(2) to identify which additional intervention components or study characteristics are the most effective at improving healthy lifestyle outcomes.

Future Scope

To address the lack of physical activity and resulting health issues, a substantial amount of research has been dedicated to tracker-based interventions which may synergize with the growing use of wearable devices by consumers. Among several brands of commercial wearables, Fitbit stands out as one of the most popular commercial wearable activity trackers, with more than 63 million devices sold worldwide in the last 10 years and with an active community.

Fitness trackers are getting more personal, powerful in 2022 and beyond. The near-term future of fitness tracking is about helping you make more sense of health data, especially as companies explore more advanced health metrics.
The medical community is also excited about the potential that future fitness devices hold for detecting more advanced metrics, like changes in glucose levels and the role of wearables in preventive care.

We intend to implement a model to calculate calorie expenditure for different activities for a user for a given amount of time and also looking into implementing a weight calculator which determines weight fluctuations that may occur in a user, having calculated the total calorie expenditure on a given day.

References

Fitbit-Based Interventions for Healthy Lifestyle Outcomes: Systematic Review and Meta-Analysis-Gunther Eysenbach, Reviewed by Marcia Johansson and Hiroyuki Sasai-2020 Oct; 22

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7589007/

https://dl.acm.org/doi/abs/10.1145/3154862.3154897?casa_token=iOEKNoqB0x8AAAAA:5UsPnn5lOWjLseJJmA8KbBm_5M2XgnOJV3rC2zatsDbZkjBInSixEv-z9KCJTl4_dIWI2rsv5xUC6Q

Literature Survey

Paper 1: Is FitBit fit for sleep-tracking?

Year of Publication: 2017

Author: Zilu Liang

Link:

https://dl.acm.org/doi/abs/10.1145/3154862.3154897?casa_token=iOEKNoqB0x8AAAAA:5UsPnn5lOWjLseJJmA8KbBm_5M2XgnOJV3rC2zatsDbZkjBInSixEv-z9KCJTl4_dIWI2rsv5xUC6Q

This paper talks about how it is now easy to track one's sleep through consumer wearable devices like Fitbit from the comfort of one's home. However, compared to clinical measures, the data generated by such consumer devices is limited in its accuracy. The aim of this paper is to explore how users perceive accuracy issues, possible measurement errors and what can be done to address these issues. Through an interview study with Fitbit users, three main sources of errors are found which are: (1) lack of definition of sleep metrics, (2) limitations in underlying data collection and processing mechanisms, and (3) lack of rigor in tracking approach. This paper proposes countermeasures to address these issues, both from the aspect of technological advancement and through engaging end-users more closely with their data.

Literature Survey

Paper 2: Heart Rate Variability

Year of Publication: 1993

Author: Conny M. A. van Ravenswaaij

Link:

https://www.acpjournals.org/doi/abs/10.7326/0003-4819-118-6-199303150-00008

This paper talks about the amount of short- and long-term variability in heart rate that reflects the vagal and sympathetic function of the autonomic nervous system, respectively. Therefore heart rate variability can be used as a monitoring tool in clinical conditions with altered autonomic nervous system function. In postinfarction and diabetic patients, low heart rate variability is associated with an increased risk for sudden cardiac death.
Heart rate variability analysis is easily applicable in adult medicine, but physiologic influences such as age must be considered. The most important application is the surveillance of postinfarction and diabetic patients to prevent sudden cardiac death. With heart rate variability analysis, individual therapy adjustments to achieve the most favorable sympathetic-parasympathetic balance might be possible in the future.

Literature Survey

Paper 3: Daily Associations Between Sleep and Physical Activity

Year of Publication: 2019

Author: Michael P. Mead
Link: https://link.springer.com/article/10.1007/s12529-019-09810-6

This paper talks about Research that has demonstrated a correlational relationship between sleep and physical activity, though this work has been largely cross sectional and fails to demonstrate temporal relationships. The purpose of the study was to test the daily, bidirectional relationships between sleep and physical activity, and whether this varied between weekdays and weekend days.

Mixed linear models revealed that physical activity did not predict subsequent night's sleep. However, on nights when participants had longer than their own average total sleep time, and greater than their own average wake after sleep onset, this predicted less physical activity the following day.

Results further suggest that, in healthy young adults, physical activity may not promote healthier subsequent sleep, but sleep duration and continuity influence physical activity in their own way.

Literature Survey

Paper 4:Measuring the fitness of fitness trackers

Year of Publication: 2017

Author: Chelsea G. Bender

Link:
https://ieeexplore.ieee.org/abstract/document/7894077

Data collected by fitness trackers could play an important role in improving the health and well-being of the individuals who wear them. Many insurance companies even offer monetary rewards to participants who meet certain steps or calorie goals. However, in order for it to be useful, the collected data must be accurate and also reflect real-world performance. While previous studies have compared step counts data in controlled laboratory environments for limited periods of time, few studies have been done to measure performance over longer periods of time, while the subject does real-world activities. There are also few direct comparisons of a range of health indicators on different fitness tracking devices. In this paper, a comparison was deduced between step counts, calories burned, and miles travelled data collected by three pairs of fitness trackers over a 14-day time period in free-living conditions. It indicates that the number of steps reported by different devices worn simultaneously could vary as much as 26%. At the same time, the variations seen in distance travelled, based on the step count, followed the same trends. Little correlation was found between the number of calories burned and the variations seen in the step count across multiple devices. The results demonstrate that the reporting of health indicators, such as calories burned and miles travelled, are heavily dependent on the device itself, as well as the manufacturer's proprietary algorithm to calculate or infer such data. As a result, it is difficult to use such measurements as an accurate predictor of health outcomes, or to develop a consistent criteria to rate the performance of such devices in head-to-head comparisons with advance medical machines.

Thank You!