A logo with text on it

Description automatically generated

**National College of Ireland**

**Project Submission Sheet**

|  |  |  |  |
| --- | --- | --- | --- |
| **Student Name:** | Brandon Craig D'Souza  Dhruv Sharma  Leslie Rebeca Monroy Ochoa | | |
| **Student ID:** | x23100125  x22228268  x23169761 | | |
| **Programme:** | MSc in Data Analytics | **Year:** | 2023-2024 |
| **Module:** | Domain Application of Predictive Analytics | | |
| **Lecturer:** | Vikas Sahni | | |
| **Submission Due Date:** | 23/02/2024 | | |
| **Project Title:** | Predicting Women's Wellness with Fitbit Insights for Bellabeat | | |
| **Word Count:** | 2643 | | |

**I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.**

**ALL internet material must be referenced in the references section. Students are encouraged to use the Harvard Referencing Standard supplied by the Library. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action. Students may be required to undergo a viva (oral examination) if there is suspicion about the validity of their submitted work.**

|  |  |
| --- | --- |
| **Signature:** | Brandon Dsouza, Dhruv Sharma, Leslie Monroy |
| **Date:** | 23/02/2024 |

**PLEASE READ THE FOLLOWING INSTRUCTIONS:**

1. Please attach a completed copy of this sheet to each project (including multiple copies).

2. Projects should be submitted to your Programme Coordinator.

3. **You must ensure that you retain a HARD COPY of ALL projects**, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer. Please do not bind projects or place in covers unless specifically requested.

4. You must ensure that all projects are submitted to your Programme Coordinator on or before the required submission date. **Late submissions will incur penalties.**

5. All projects must be submitted and passed in order to successfully complete the year. **Any project/assignment not submitted will be marked as a fail.**

|  |  |
| --- | --- |
| **Office Use Only** | |
| Signature: |  |
| Date: |  |
| Penalty Applied (if applicable): |  |

Predicting Women's Wellness with Fitbit Insights for Bellabeat

Brandon Craig D'Souza  
National College of Ireland  
MSc in Data Analytics  
Dublin, Ireland  
[x23100125@student.ncirl.ie](mailto:x23100125@student.ncirl.ie)

Dhruv Sharma   
National College of Ireland  
MSc in Data AnalyticsDublin, Ireland  
[x22228268@student.ncirl.ie](mailto:x22228268@student.ncirl.ie)

Leslie Rebeca Monroy Ochoa  
National College of IrelandMSc in Data AnalyticsDublin, Ireland  
[x23169761@student.ncirl.ie](mailto:x23169761@student.ncirl.ie)

*Abstract*—Wearable technology have gained significant popularity in recent years that includes fitness trackers designed to monitor various aspects of physical activity and health, providing individuals with a great insight of data regarding their daily activity, heart rate, and sleep patterns. These trackers include several functionalities that address the requirements and preferences of each user while focusing on boosting general health and wellness.

By using data collected by fitbit devices, this paper aims on the advantages of fitness trackers for women and making various predictions on it. The authors make use of the data and study the accuracy of the predictions of the fitness trackers and how they are used to improve users health through the analysis of the datasets. These datasets include activity levels, heart rate, sleep patterns and various information collected from approximately 30 fitbit users.

Keywords—fitness trackers, predictions, machine learning.

# Introduction

Fitness trackers have become a very useful wearable accessory in past few years, providing people with a huge insights into their daily activity and general health. Fitness bands and smart watches are being used in an increasing number of studies in science as methods for measuring daily physical activity and sleep habits of people to improve their way of life and health [1 - 3]. S. Bathia showed how well Decision Trees, Random Forests, K-Nearest Neighbour Classifiers, and XGBoost Classifiers were used to predict physical activity based on data collected from Fitbit and Apple Watch [4]. In terms of classification accuracy for activity recognition, dynamic classifiers like RNN work way better than static classifiers such as Naive Bayes, and Random Forest algorithms concluded by I. Khokhlov, L. Reznik and R. Bhaskar in their research [5].

By exploring eighteen datasets having various parameters, including activity levels, heart rate variability, and sleep quality, our main goal is to provide insights and advantages of using this data to improve the daily health and well-being of women.

Through this research, the authors seek to provide useful information on the use of fitness trackers as tools for improving women's health, offering recommendations for improving daily routine and using this technology to support women in achieving their fitness and wellness goals.

## The datasets

The dataset used for this research is obtained from Kaggle and has data from fitness trackers from thirty fitbit users that are gathered by bellabeat. The Fitbit users who met the requirements gave permission to have their personal tracker data recorded, including continuous minute output for heart rate, physical activity, and tracking sleep cycles. It contains various information on steps taken, heart rate, and daily activities that are going be used to investigate and make predictions on user behaviour.

Bellabeat's main goal was to investigate the usage patterns of one of their flagship products, utilizing this dataset to extract huge insights into how individuals are currently engaging with their smart devices and wearables.

By examining the data, the authors aim to discover existing trends and user behaviours, understand the patterns that highlight the preferences, habits, and needs of Bellabeat consumer base. Through precise analysis, the authors can identify key usage patterns, peak activity times, and underlying sleep trends among the users. Moreover, the authors can identify correlations between physical activity levels, heart rate variations, sleep quality and understanding the complex interaction between these features.

# goals

The main goal of this project provides crucial steps with the purpose of utilising the obtained data to improve user experience and help in strategic decision-making for Bellabeat considering the following:

## Advanced Predictive Analysis:

The main goal is to create complex predictive models that can accurately predict user behaviour by utilising data that Fitbit users shared. Through the analysis of complex patterns in the data, which include measurements like step count, dynamic heart rate changes, and physical activity levels, the authors can provide useful information about users' daily habits, preferences, and future health plans .

## Comprehensive User Engagement Analysis:

The research aims to understand the complex factors that influence users motivation and engagement with fitness trackers and going beyond just prediction. By carefully examining usage trends the critical elements affecting long-term Bellabeat product adoption can be identified. This extensive analysis will cover things like characteristics of social interaction and personalised feedback mechanisms to reveal what really connects with users and motivates them to their fitness path.

## Strategic Recommendations and Marketing Insights:

The author’s final objective is to provide Bellabeat with perfect recommendations to strengthen their marketing strategy and improve their product offerings. The recommendations will be supported by data from the prediction analysis. The project aims to help Bellabeat to better target their marketing, improve the functionality of their products, and maximize their distribution channels by breaking down data patterns into precise and easy instructions. The project recommendations are for improving the usability of the user interface, implementing creative motivations to increase user involvement and focusing marketing campaigns on particular population which will be supported by concrete evidence and designed to have the greatest possible impact.

# Ethical Concerns

## Ensure user consent and data privacy:

Findings demonstrate how these gadgets enable private information to be obtained by third parties from the precise position of the device, potentially compromising users' privacy. Through a Bluetooth connection with mobile apps that push and draw data from cloud servers, user data is being pulled [6].

Collecting sensitive health data from Fitbit users raises questions about informed consent, data security, and potential misuse of personal information. Our primary focus when handling this data from revolved around safeguarding the privacy of users, particularly concerning sensitive information like their name, age, gender, and geographical location. The authors prioritise verifying that the dataset has been anonymized and that Fitbit users have provided the necessary consent for the usage of their data.

## Bias mitigation

Large-scale population sleep studies frequently utilise self-reported diaries and questionnaires to measure the length and quality of sleep. But it could be viewed as subjective and biased [7].

This concern involves implementing strategies to address any biases present in the dataset that could impact the precision and impartiality of predictions. This includes identifying and addressing any potential sources of bias to ensure that predictions are as accurate and fair as possible.

## Transparent communication

It is required to maintain open lines of communication with all parties involved in this project, explaining its goals and anticipated results in an honest and open manner. This guarantees that all those impacted by the project are aware of its objectives, possible effects, and associated risks, enabling them to actively engage in the process and make well-informed decisions. The project's ethical basis may be compromised by unclear and opaque communication, which can cause miscommunication, mistrust, and eventually harm to stakeholders.

# Business Value

The main goal is to improve Bellabeat's marketing tactics and strengthen its competitive advantage in the industry by utilising data-driven machine learning approaches. The aim is to enhance Bellabeat's approach to reaching and engaging with its target audience while staying ahead of competitors focusing on the following areas:

## Users Insights

By examining the usage patterns and behaviours of Bellabeat's smart device users, Bellabeat can gain valuable insights into their customers' preferences, habits, and needs. Using this data, Bellabeat can target their audience's needs by customising their marketing strategies, product features, and customer care programmes.

## Product Enhancement

Understanding how users use Bellabeat's smart devices can provide valuable feedback for product development and improvement. The company can prioritise features and functions that are most important to its customers by detecting key usage patterns and trends. This will improve the user experience and increase customer satisfaction.

## Marketing Optimisation and Personalised Recommendations

Bellabeat can improve its marketing efforts by optimising its data on peak activity periods, sleep trends, and the relationships between physical activity levels and sleep quality. This may involve modifying the timing and content of marketing campaigns to resonate with users during certain moments.

By utilising Machine Learning techniques, Bellabeat can offer customers personalised recommendations based on their usage habits and behaviours. This could improve the value proposition of Bellabeat's smart devices by offering personalised exercise regimens, sleep optimisation techniques, or product recommendations that match personal interests and objectives.

# Preliminary Visualisations

## Total Steps x Calories

A purple dots in the air

Description automatically generated with medium confidence

Figure 1. Total Steps x Calories

The scatter plot above shows strong correlation between the Calories and the Total of Steps taken, meaning that an increase in step count corresponds to an increase in caloric expenditure. Staying active is essential for preserving good health, and a person's daily step count has a big influence on that.

## Most Active Time of Day

A graph of purple bars

Description automatically generated

Figure 2. Most Active Time of Day

The bar plot provides valuable insights into the users' activity patterns throughout the day. The data shows that users exhibit an initial increase in activity between 6 and 8 am, that shows most of the people are committed to early morning exercising. The level of activity remains relatively stable during the daytime hours. However, a steep rise occurs between 5 and 7 pm. This suggests a trend where users engage more in physical activities after work or school hours.

## User Distribution x Activity Level

The following visualization “User distribution with respect to Activity level” is created from the activity level taking into consideration the percentage of people doing it. From the graph it is observed that the highest activity is Sedentary with Highly active being the least.

A graph with different colored squares

Description automatically generated

Figure 3. User Distribution x Activity Level

## Relation of Sleep Time x Activity Level

To get “The relation with sleep time and activity level” the graph takes the “Avg\_steps” and “Avg\_Very\_Active\_Mins” against the “Avg\_sleep\_time”. This gives a useful insight of how sleep time affects the activity level and steps taken during a day. From the graphs it can be observed that the more the people sleep, the more activity they tend to do over the day.

A graph of a graph of activity

Description automatically generated with medium confidence

Figure 4. Relation of Sleep Time x Activity Level

## User x Activity Level

For “Each users activity level”, the “Total\_calories”, ”Total\_steps”,”Total\_lightly\_active\_mins”,”Total\_very\_active\_mins”, “Total\_fairly\_active\_mins” of each individual user(ID) is taken. This gives a huge insight into how different users perform individually.

A graph of a bar chart

Description automatically generated with medium confidence

Figure 5. User x Activity Level

## Most active days of the week

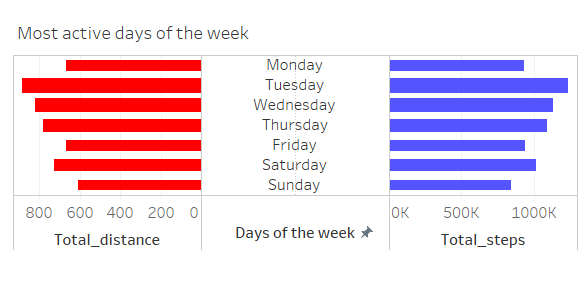


Figure 6. Most active days of the week

The graph “most active days of the week” uses the “Total\_distance” and “Total Steps” against the “Days of the week” to come up with the graph where there are insights on the activity take place during the week. From the above butterfly chart, it can be observed that the activity on Monday is moderate and most of the activity is done on Tuesdays with it decreasing gradually over the week.

## Average heart rate per hour

A graph with purple lines

Description automatically generated

Figure 7. Average heart rate per hour

The line plot that shows the average heart rate each hour provides useful information about users' cardiac activities all day long. Surprisingly, the graph shows a clear pattern: heart rates bottom out about midnight, especially between 3:30 and 4:00 AM. This trough corresponds with the normal nightly drop in heart rate during deep sleep stages, indicating a period of peaceful sleep for users. In contrast, the highest average heart rates are found in the afternoon and evening, suggesting a spike in cardiac activity during this period. This considerable increase in heart rate in the later hours of the day suggests that a significant proportion of users exercise or partake in other physically demanding activities during this time. Using this knowledge to inform our marketing approach, the project may highlight features like personalized exercise suggestions, peak activity warnings during high-intensity hours, and real-time heart rate monitoring while working out. This projects aims to improve users' workout experiences and build a stronger relationship with the brand by matching our fitness products to users' exercise routines. This could lead to a rise in sales among fitness enthusiasts who are looking for more sophisticated heart rate tracking features.

# Techniques

Top of Form

Numerous data science methods will be used in our investigation to improve predicting abilities and glean valuable insights. Predictive modeling can be applied, which makes use of machine learning algorithms to predict user involvement and behavior. Furthermore, by classifying users according to comparable activity patterns, clustering analysis offers a more sophisticated comprehension of user segments. To enable targeted enhancements, a feature importance analysis will be carried out to determine the primary elements driving user engagement. Last but not least, time series analysis will reveal patterns and trends in second-, minute-, and daily-level data, enabling a thorough investigation of temporal dynamics within the dataset. Together, these methods create a strong analytical framework that allows for the extraction of useful information from Fitbit user data.

The use of a variety of analytical tools is expected to produce important new understandings of user behavior and preferences. In particular, correlation analysis will clarify the complex relationship between heart rate and daily steps, providing insightful data for improved products. By segmenting consumers, clustering analysis will help by providing tailored strategies that cater to different age groups. Bellabeat will be guided by feature importance analysis when determining and ranking elements that are essential for improving user engagement. Furthermore, temporal patterns will be revealed via time series analysis, improving the comprehension of users' long-term routines.

Bellabeat can utilize the insights obtained from predictive modeling, customized wellness programs, and motivational notifications to further its goal of data-driven empowerment. Personalized wellness programs, which may be tailored to each user's needs, can serve as a central marketing feature to improve the user experience in general. Users looking for motivation and consistent engagement may be drawn in by highlighting motivational alerts as a unique selling feature. These alerts were found to be influential based on correlation study. One direct result of clustering analysis is targeted user segmentation, which is useful for customizing marketing messaging to particular user groups and making sure campaigns speak to a range of needs.

Our feature importance analysis is in line with the creation of a marketing story that highlights data-driven empowerment and illustrates how Fitbit data insights support women in their quest for well-being. Time series analysis can be used to inform continuous engagement techniques that can be used to maintain user engagement throughout time, with a focus on peak activity hours. Lastly, Bellabeat may go beyond being a product and become an instructor in the wellness market by using educational content based on Fitbit data interpretation to promote Bellabeat as a source of insightful health information. This all-encompassing strategy, which combines data-driven insights with focused tactics, is well-positioned to produce an engaging story that will boost Bellabeat's sales growth and engagement eventually improving women's wellness and empowerment.

# References

1. Z. Liang and M. A. Chapa-Martell, "Combining Resampling and Machine Learning to Improve Sleep-Wake Detection of Fitbit Wristbands," 2019 IEEE International Conference on Healthcare Informatics (ICHI), Xi'an, China, 2019, pp. 1-3, doi: 10.1109/ICHI.2019.8904753.
2. Z. Liang and M. A. C. Martell, "Combining Numerical and Visual Approaches in Validating Sleep Data Quality of Consumer Wearable Wristbands," 2019 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops), Kyoto, Japan, 2019, pp. 777-782, doi: 10.1109/PERCOMW.2019.8730805.
3. Z. Liang, M. A. C. Martell and T. Nishimura, "A Personalized Approach for Detecting Unusual Sleep from Time Series Sleep-Tracking Data," 2016 IEEE International Conference on Healthcare Informatics (ICHI), Chicago, IL, USA, 2016, pp. 18-23, doi: 10.1109/ICHI.2016.99.
4. S. Bathia, “Activity Identification using Supervised Machine Learning on Wearable Activity Tracker Data”, 2021 Asian Conference on Innovation in Technology, doi: 10.1109/ASIANCON51346.2021.9544525.
5. I. Khokhlov, L. Reznik and R. Bhaskar, “The Machine Learning Models for Activity Recognition Applications with Wearable Sensors”, 2019, 2019 18th IEEE International Conference On Machine Learning And Applications, DOI: 10.1109/ICMLA.2019.00072.
6. C. Zhang, H. Shahriar and A. B. M. K. Riad, "Security and Privacy Analysis of Wearable Health Device," 2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC), Madrid, Spain, 2020, pp. 1767-1772, doi: 10.1109/COMPSAC48688.2020.00044.
7. D. S. Lauderdale, K. L. Knutson, L. L. Yan, K. Liu, and Paul J. Rathouz, “Sleep duration: how well do self-reports reflect objective measures? The CARDIA Sleep Study” 2008, Epidemiology, vol. 19, no. 6, pp. 838-845.