REVEALING SUSTAINABLE GROWTH FOR FIT BIT: A DATA-DRIVEN MARKETING APPROACH BASED ON K-MEANS CLUSTERING AND COLLABORATIVE FILTERING

Akansha Akansha and Stuart So

University of Exeter Business School, Rennes Drive, Exeter, United Kingdom

ABSTRACT

Understanding the user segment is highly significant in the age of a highly competitive wearable Fitness Technology market. In this study, we leveraged a comprehensive dataset containing information on user interactions, activity logs and device usage records. For effective segmentation of the users, K-Means clustering was employed. The unsupervised Machine Learning algorithm helped us group the clusters of consumers based on their similarity in the usage of the device, activity levels and engagement patterns. The collaborative Filtering technique refines product recommendations by identifying user preferences based on past patterns. The analysis aims to uncover distinct user segments and provide insights into user behaviours and lifestyles to enhance Fitbit's Market Performance and improve user engagement, customer satisfaction and brand loyalty leading to higher customer retention. The findings of an extensive analysis conducted on Fitbit User data using K-Means Clustering and Collaborative filtering techniques are presented. To achieve sustainable growth in the highly competitive smart wearables market, Fitbit can improve its user experience by addressing the diverse needs of different user segments.

KEYWORDS

Fitbit, Segmentation, K-Means, Collaborative Filtering, Personalisation, Wearable Fitness

1. INTRODUCTION

In the rapidly evolving landscape of contemporary business and marketing, a pivotal convergence occurs - the fusion of sustainability and customer-centricity [1]. As global awareness of environmental concerns grows, and consumer preferences shift, organisations face the dual challenge of elevating their market performance while aligning their practices with sustainable principles. This dynamic convergence of objectives has prompted an in-depth exploration of strategies that transcend traditional norms, ushering in prospects for sustainable economic growth.

This business undertaking can be directed to a voyage into sustainable growth, spotlighting the esteemed health and wellness technology brand Fitbit which is a market leader in wearable technology and related products [2]. Our research conducts an Outside-In Exploration of Fitbit's

David C. Wyld et al. (Eds): ICCSEA, NLAI, SCAI, CSIA, IBCOM, SEMIT, NECO, SPPR, MLDS -2023
pp. 93-108, 2023. CS & IT - CSCP 2023 DOI: 10.5121/csit.2023.132408

Customer-Centric Marketing Approach, which dives into the intricate interplay between customer-centricity and sustainable marketing strategies using cluster analysis and collaborative filtering techniques. Fitbit, renowned for its innovative wearable devices and health-focused technologies, is an example of how companies can adroitly integrate customer insights into their marketing blueprints [3].

By meticulously observing Fitbit's practices, we seek to unveil how the interweaving of these factors could give rise to a comprehensive marketing approach that amplifies brand growth and nurtures sustainability [4]. The study underscores the paramount significance of customer-centric marketing, functioning as the linchpin around sustainable growth. The detailed analysis aims to subject Fitbit's marketing to a customer-centric approach and uncover how to empower the company to nurture sustainable growth while upholding its commitment to environmental stewardship and customer well-being.

In the following sections, this paper will follow a methodical trajectory, from the foundational tenets of sustainable marketing, immersing into the specifics of Fitbit's customer-centric methodology and culminating in pragmatic insights drawn from user data. We will delve deeper into theoretical frameworks, methodological considerations, and findings highlighting the symbiotic relationship between customer-centricity and marketing. At its core, this report aspires to enrich the scholarly discourse on sustainable marketing and furnish actionable insights to practitioners who aspire to harness avenues of growth while navigating the marketing domain, guided by insights from user data.

2. OVERVIEW OF THE WEAR ABLE FITNESS TECHNOLOGY INDUSTRY AND FITBIT

2.1. The Wearable Fitness Technology Industry

The world of wearable fitness tech is like a fast-paced race where high-tech meets our health. It's a place where we can find a variety of gadgets, from simple step counters to super-smart watches that keep track of every move, our sleep, and even our heart rate. With the growing concern towards staying healthy, the market for fitness gadgets is constantly growing. Various brands are trying to develop new technology to win people over [5].

Additionally, the emergence of the COVID-19 pandemic has further impacted the demand for wearable technology within the healthcare sector, as these devices can promptly detect indicators of infection. Furthermore, the increasing disposable income, growing popularity of such devices, market availability of intelligent gadgets, and various other factors are anticipated to play a significant role in the growth of the wearable technology industry [6]. Recent trends in the wearable technology market indicate promising possibilities for expansion. Moreover, the revenue in the wearables market will likely reach USD17.834billionbythe close of 2021,witha significant portion of this revenue originating from the Chinese market [7].

2.2. Fitbit

Established in 2007 by James Park and Eric Friedman, Fitbit assumed a pivotal role in shaping the landscape of wearable fitness technology. The company's establishment corresponded with the industry's early expansion phases, positioning Fitbit as an innovative frontrunner. In its initial stages, Fitbit concentrated on designing wireless activity trackers geared towards aiding individuals in tracking their fitness progress. The inaugural offering, known as the Fitbit Tracker, gained notable recognition, serving as a cornerstone for the subsequent growth of the brand. The popularity of these products was regarded as a positive influence in promoting physical health and cultivating behaviour that enhances the quality of life, with an emphasis on physical activity [8].

2.2.1. Fit Bit Financial Performance

Fitbit's main mission is to contribute to better health worldwide by continuously monitoring and improvingpeople'shealthperformance.AsreportedthroughthewayofBusinessWire,FitbitInc.

Achieved total earnings of $188 million in 2020. Hence, it is observed that the overall possessions owned by the company decreased from $1,515,547 thousand to $1,368,086 thousand in 2019, primarily because the amount of cash and easily accessible funds had lowered. Fitbit reported that the company's net earnings have decreased over time due to a drop in gross profit and increased operating expenses. The company's net operational income was $(320,711) thousand in 2019. Additionally, Fitbit Inc.'s overall financial health could have been more robust in 2020, asin dicated by its negative cash flow from operations, which amounted to $(80) million.

In summary, Fitbit Inc. is facing challenges in various financial aspects, and the company needs to implement stringent measures to enhance its operational and financial standing [9].

2.2.2. Product Marketing Strategy

The promotion strategies include techniques to improve brand awareness, sales and loyalty. Including:

Direct Marketing: Utilising direct mail, telemarketing, and personalized email.
In-Store Promotion: Incorporating flash sales, discounts, and loyalty points.
Social-Media Marketing: Utilising plat forms like Face book and Instagram to engage with the user-base.
Integrated-Marketing Communication: Ensuring constant messaging throughout all channels for coherent brand communication.

Whilst other brands offer equally competitive products and services, Fitbit’s combination of brand recognition and continuous innovation has contributed to its appearance and popularity among consumers. Innovation is vital to sustain competition in the smart wearable industry, and Fitbit’s ability to consistently offer fresh designs is appealing to consumers. Consumer preferences in the smart wearable market are shaped mainly by the presence of cutting-edge features. A targeted marketing strategy can help a brand stand out amidst fierce competition. A targeted approach based on data-driven insights is pertinent in the competitive smart wear ables market. Fitbit’s ability to harness the power of data-driven insights can create more accurate and sustainable marketing strategies.

We intend to address both facets within our approach, unveiling a flexible, two-stage methodology that dynamically crafts behaviorally coherent segments and subsequently steers the selection of target marketing strategies. Our methodology is rooted in the user data of the customers using the devices constantly throughout their day. The foundational principle of behavioral coherence involves an exhaustive exploration of interconnected purchases across diverse categories on a segment level, involving the precision-targeted engagement of several thousand customers.

In this context, the strategic focus on **data-driven insights** emerges as a valuable approach. The user-generated data can mainly guide marketing strategies by providing a deeper understanding of consumer behaviour and preferences. By delving into user behaviour and preferences, Fitbit can fine-tune its marketing strategies, ensuring they resonate with specific target audiences and improvising the products to meet particular categories of consumers.

3. LITERATURE REVIEW

3.1. Market Segmentation

Market Segmentation is a fundamental step involving a diverse market into smaller, more homogeneous segments based on shared characteristics, behaviours or needs [10]. In recent times, the landscape of market segmentation has been reshaped by the emergence of big data and advancements in data analytics, enabling businesses to craft precise and impactful strategies. This categorisation hinges on shared characteristics, encompassing everyday needs, interests, lifestyles, or analogous demographic traits.

The primary objective of Segmentation revolves around identifying segments with the highest potential for profit generation or future growth. These chosen segments receive special attention, earning the designation of "target markets." The methods for segmenting a market are diverse. In scenarios involving business-to-business (B2B) dynamics, the market might be dissected into distinct business types or geographical regions. Conversely, business-to-consumer (B2C) settings often employ Segmentation based on demographic factors like lifestyle, behaviour, or socioeconomic status.

User data offers priceless insights into consumer behaviour, preferences, and demographics [11]. This level of understanding is aimed at businesses aiming to customise their products, services, and marketing campaigns for specific audiences. The user-generated data from different sources can significantly enhance segmentation accuracy, allowing businesses to differentiate subtle distinctions within the market segments and create highly targeted marketing initiatives resonating with the consumers. Market segmentation is crucial for a company's marketing strategy to be effective. This is because traditional class patterns no longer exist, and consumers have more disposable income. Companies can develop their products in the right direction by dividing consumers into manageable segments based on their needs. Targeting all consumers would be unnecessary and costly, so understanding the consumer segment regarding age, values, purchase behaviour, and attitudes is essential for success [12].

3.2. Data-Driven Personalisation and Underlying Approaches

The defining theme of the contemporary market is personalisation. Businesses harness user data to deliver tailored experiences, content and recommendations [13]. This amplifies customer satisfaction, catering to brand loyalty. [14] delved into the concept of predictive personalisation; by leveraging historical user data and modern machine learning algorithms, businesses can proactively fore castuser preferences to cater to individual needs. While using user data in market segmentation offers a substantial advantage, it raises ethical concerns simultaneously.

The future of market segmentation based on user data will soon be induced with automation and Artificial Intelligence (AI). AI-driven algorithms can swiftly analyse vast datasets in real-time, enabling adaptive and dynamic segmentation strategies [15].

Furthermore, there are possibilities for integrating offline and online data sources. Merging the user data from physical and digital touch points will provide a more holistic understanding of consumer behaviour [16].

3.2.1. K-Means Clustering

K-Means clustering is an unsupervised machine-learning algorithm for data analysis and segmentation [17]. It is designed to divide a large dataset into small groups or clusters, where data points within each cluster are more similar than those in other clusters [18].

Algorithm 1 k-means algorithm

```
Specify the number k of clusters to assign.
Randomly initialize k centroids.
repeat
   expectation: Assign each point to its closest centroid.
   maximization: Compute the new centroid (mean) of each cluster.
until The centroid positions do not change.
```

Figure1 K-Means Algorithm

As shown in **Figure 1**, the process starts by selecting a user-defined number of clusters, ‘K’. The algorithm assigns each data point to the nearest cluster centre, often represented by the mean of the data points in that cluster. It refines the cluster assignments until convergence, minimizing the sum of squared distances between data points and their respective cluster centres [19].

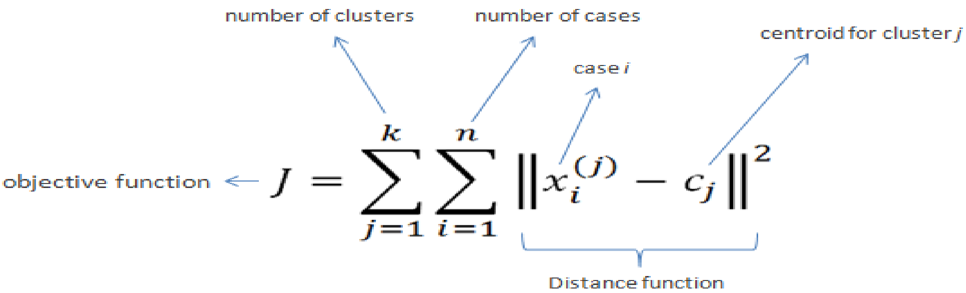

Figure2 Mathematical Model of K-Means Clustering

The mathematical model of K-Means Clustering is expressed in **Figure 2**. K- Means clustering as various applications across a variety of domains. In marketing, the K-means helps in customer segmentation by group in g customers with similar behaviors or preferences[20]. In image processing, it helps in image compression and object recognition [21]. Additionally, it is utilised in biological analysis and clustering [22].

Despite the versatility, K-Means clustering has limitations, such as sensitivity to initial placements of cluster centres and the assumption that all clusters are spherical and equal in size. Researchers have developed an improved version of K-Means to address these limitations [23].

3.2.2. Collaborative Filtering : Are Commendation System Technique

Collaborative Filtering technique is widely used in recommendation systems aimed towards providing personalised suggestions to users [24]. The technique relies on the ideas of users who had similar preferences in the past to have similar preferences in the future. The core principle of this technique is to analyse user interactions and feedback on items, such as ratings or purchase history, to provide recommendations.

Two main approaches to Collaborative Filtering are:

1. **User-Based Collaborative Filtering:** Identifies users with similar preferences based on their historical interactions and recommends items favoured by consumers with the same profiles.
2. **Item-Based Collaborative Filtering:** Items are compared based on user feedback to find similarities; products that are highly rated or interacted with by consumers who have shown interest in the same items are recommended.

Collaborative Filtering has proven its efficiency in various domains, including e-commerce, content streaming and social media [25]. Platforms like Amazon, Netflix and Spotify use this technique to power their recommendation systems offering users personalised products, movies or music suggestions.

Despite the efficacy, Collaborative Filtering faces challenges such as the cold start problem [26] when it deals with new users or items for which the data inter actionis limited. In order to address these limitations, hybrid recommendation systems combine Collaborative Filtering technique with other techniques, such as content-based filtering or matrix factorization [27].

4. METHODOLOGY

The methodology of the study comprises the following major components of a comprehensive research process where the research activity one comes after another which are grouped into the following four major stages within the research process (**Figure 3**):

4.1. Research Process

4.1.1. Data Collection

Considering the factors, we might require addressing our research, the dataset was found on the open-source data science platform Kaggle. It predominantly is a User-Data of the consumers using the products of Fitbit. The entire dataset consists of 18 files, with data related to the users' daily activities, sleep schedule, weight information, heart rate, calories burnt, number of steps walked on a daily basis and their intensity of day-to-day activity. The datasets are secondary data obtained from the data science portal Kaggle.

Although, keeping in mind that the dataset is available via an open-source platform and not by the brand itself does raise the question of its transparency. Kaggle underscore the commitment to safeguarding the privacy and anonymity of individuals whose data is included in the datasets.

In this particular dataset, there is no mention of the names, contact details of the consumers and the privacy of the individuals have been protected to the best extent complying with the GDPR Regulations set in 2018. Kaggle takes the responsibility of ensuring that the data is handled in an ethical and responsible manner, avoiding any misuse that could lead to harm, discrimination and privacy violations. It also promotes transparency and accountability in research conducted using the dataset. Peer review and collaboration is encouraged to maintain a high level of accountability.

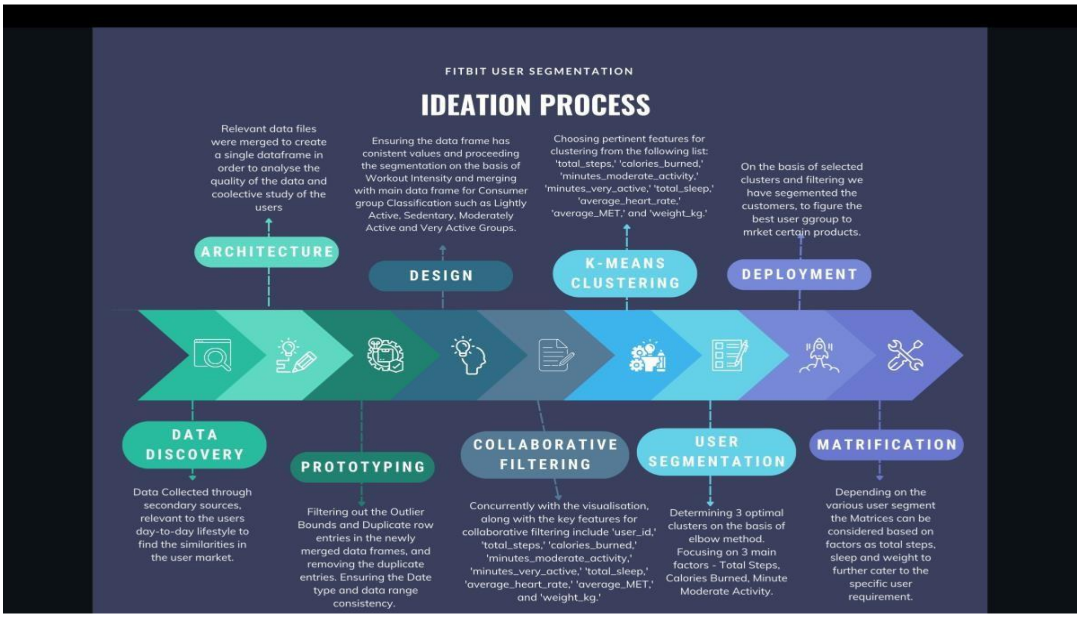

Figure3 Research Process

4.1.2. Data Retrieval

However, out of the 18 datasets, we have narrowed it down to utilising 8 datasets, as the minute-by-minute details of the user would provide us with a detailed insight into their day-to-day lifestyle, making it relevant to our research. Namely:

(a) **Daily Activity**: mentioning the record date of data collection, total steps walked or run throughout the day, and total distance covered. Active minutes and sedentary minutes along with calories.

(b) **Minute Calories Narrow:** indicating the specific minute the calories are expended is recorded.

(c) **Minute Intensities Narrow: a** record of the minutes along with the level of intensity or effort during that particular minute, providing information on how vigorous the activity was during that particular minute.

(d) **Minute Steps Narrow:** representing the number of steps taken during that specific minute of the particular activity in that minute.

(e) **Minute METs Narrow:** record of the Metabolic Equivalent of Tasks, denoting the specific minute it was performed at, measuring the energy expenditure compared to resting.

(f) **Sleep Day:** captures the specific day when the sleep data was recorded and indicates the total number of sleep records captured for that particular day. Total minutes the user was asleep during that day and the total minutes they spent in bed; giving an overview of the time spent in resting position.

(g) **heartrate_seconds:**covers the actual heart rate every five seconds and specifie show fast the heart was beating in that particular moment.

(h) **Weight Log Info:** is a collective record of the weight of the users in kilograms(kg) and pounds(lbs), also including the fat percentage and BMI (Body Mass Index) ((Mooney et al., 2013)), and whether the weight information was entered manually by the user.

The details of the data retrieval process by a data funnel approach are shown in **Figure4** below.

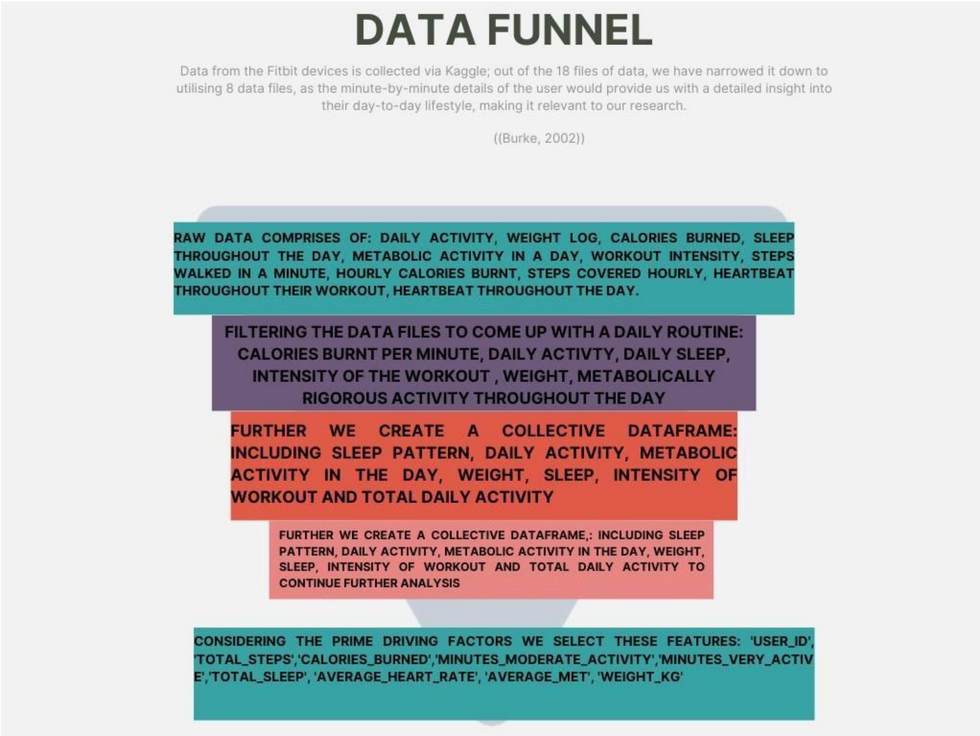

Figure4Adata funnel approach for the data retrieval process

4.1.3. Data Exploration

Using the function, **'general_ info (df)**' for an overview of our data, we get a concise summary of the data frame which offers us an insight in to the contents of the data such as the number of rows, columns and the fundamental details about data-types and null-values. It allows us to swiftly grasp the basic characteristics of the data frame as an initial step towards data exploration.

The second function, '**outlier_ bounds (df, col)'**, is used to identify potential outliers within the specific column. To determine this, we have computed and exhibited the first and third Quartiles (Q1 and Q3) for the specific columns. Using this we calculate the Interquartile Range (IQR), which will help us determine the outliers. It delineates the lower and upper bounds for the potential outliers, through IQR. It is valuable for pinpointing data anomalies and extremes

The third function, '**duplicate _ index _search (df, col1, col2**)', is to find out the duplicate entries in the data frame by focusing on two main columns if in case the values match. After detecting the duplicate row of entries, it provides the corresponding indexes (row numbers). The function proves useful in flagging the duplicate records and facilitate the assessment of data in terms of quality.

The fourth function, '**fitness_device_usage(df, TimePeriod, TotalTime, TimePeriod_name,**

variable_name)', is to analyse the usage of the fitness tracker among distinct consumer groups. By computing the average values of designated activities for each consumer group within the dataset. It also calculates the average percentage of this duration over the span of a month.

After figuring out the duplicated entries and null values in our dataframe, nowwe proceed ahead to check the Consistency of our data.

(i) **ID Consistency Check:** For each dataframe (**df_act, df_cals, df_ints, df_steps, df_sleep, df_heart, df_met, df_weight)**, the main purpose is to ensure there are no repetitions or inconsistent user IDs withing each dataset. For the sake of consistency we also rename the 'Fairly Active Minutes' to 'Moderately Active Minutes'.

(j) **Date Range Consistency Check:** To ensure the regularity in the datasets' time daterange to validate the data's consistency, especially in time-series data. Ensuring that the data covers an expected and coherent time period, with no gaps or irregularities in the range of dates.

To assure the quality of data, the consistency check is a part of the pre-processing.

4.1.4. Data Engineering

As most of the variables are interrelated in our datasets such as the calories burnt on average are dependent on the intensity of the workout. The Metabolic Equivalent of Tasks are responsible for the Calorie expenditure as well, which also in cludes the number of steps covered on a daily basis. To avoid switching between multiple datasets to conclude on the same variables, we chose the joined data frame for the holistic ally comprehensive approach of analysis. We join our data sets to create a singular data frame to avoid the overlap of the same variables.

We then move on to assess the consistency between the intensity of the Steps and the average speed in the data frame '**df_act**', by calculating the average speed for all levels of steps walked based on intensity: **Very Active, Moderately Active and Lightly Active**. Applying conditions to identify inconsistency and check if the average speed of the 'Very Active' workouts is lower than the 'Moderately Active' and if the average speed of 'Moderately Active' workouts is lower than the 'Lightly Active' workouts.

We then move on to calculate the percentage of inconsistent data by dividing the number of rows in **'inconsistent_df_act'** by the total number of rows in '**df_act**' and then multiplying it by 100. This calculated percentage represents the extent to which workout in tensity and average speed are inconsistent in the dataset. Setting the intensity level of workout categories of Sedentary, lightly Active, Moderately Active and Very Active users, respectively, to 0,1,2 and 3. We calculate the average minimum and maximum Steps covered.

Customer Segmentation on the basis of number of steps covered: As per (*Counting Your Steps*,n. d.),10000 steps is the daily recommended target for an adult to live a healthy life .Hence, we have segmented the customers' data in '**df _ act',** as per the following ranges:

Sedentary<=5000 steps Lightly Active: 5001 to 8500 steps

Moderately Active: 8501 to 12500 steps Very Active >= 12500 steps

Furthermore, in the analysis, we segmented the customers based on their sleep, as well as sleep, along with their daily activity intensity and the time of day they were the most active. We have also analysed and visualised the average time it might take for the customers to fall asleep.

4.2. Collaborative Filtering and Algorithms

Collaborative filtering is a technique that can help you find relevant and personalized recommendations based on the preferences or feedback of many users. For example, if you want to watch a movie, a collaborative filtering system can suggest movies that are similar to the ones you have liked before, or movies that other users with similar tastes have liked.

There are different methods and challenges for implementing collaborative filtering, such as how to measure the similarity between users and items, how to deal with sparse or missing data, how to handle new users or items, and how to scale up the system for large datasets. One common approach is to use matrix factorization, which is a way of finding latent features or embeddings that represent the characteristics of users and items. These embeddings can be learned automatically from the data, without relying on hand-engineered features. The embeddings can then be used to compute the similarity or the predicted rating between a user and an item

4.2.1. The Algorithms

Collaborative filtering algorithms are methods that can help you find relevant and personalised recommendations based on the preferences or feedback of many users. There are different types of collaborative filtering algorithms, such as:

(a) **Memory-based algorithms [28]**: These algorithms use the entire feedback matrix to compute the similarity between users or items, and then use these similarities to predict the ratings or preferences of a target user. For example, user-based collaborative filtering finds users who have similar ratings to the target user, and then recommends items that these similar users have liked. Item-based collaborative filtering finds items that have similar ratings to the target item and then recommends items that are similar to the target item. Memory-based algorithms are simple and intuitive, but they can suffer from scalability and sparsity issues.

(b) **Model-based algorithms [29]**: These algorithms use the feedback matrix to learn a model that can capture the latent factors or features that explain the ratingsor preferences of users and items. For example, matrix factorization is a popular model-based algorithm that learns low-dimensional embeddings or vectors for users and items, such that the dot product of these vectors can approximate the rating or preference of a user-item pair. Model-based algorithms can handle spar sity and scalability better than memory-based algorithms, but they can be more complex and prone to over fitting.

(c) **Hybrid algorithms [30]**: These algorithms combine the advantages of memory-based and model-based algorithms or use additional information or techniques to improve the performance of collaborative filtering. For example, some hybrid algorithms can use content-based features or metadata to augment the feedback matrix or use ensemble methods or deep learning models to integrate multiple collaborative filtering models. Hybrid algorithms can achieve higher accuracy and diversity than single algorithms, but they can also be more difficult to implement and interpret.

4.2.2. Top-K Algorithms for User-Based Collaborative Filtering

As mentioned in Section 3, user-based Collaborative Filtering is a technique that uses the ratings or preferences of similar users to recommend items to a target user. It is a memory-based algorithm that uses the similarity between users or items to make recommendations. The algorithm works by first computing the similarity between all pairs of users or items and then selecting the Top-k most similar users or items to make recommendations [31]. More specifically, Top-k algorithms that applied to user-based collaborative filtering were explored where Top-k refers to an algorithm that can efficiently find the top-k items or candidates for a given user or query, based on some criteria or score [31].

There are several Python packages available for collaborative filtering. One of the most popular libraries is Surprise [32]. It is a Python scikit building and analysing recommender systems that deal with explicit rating data. It provides various ready-to-use algorithms and evaluation metrics to help you build your recommender system. To choose the Top-k algorithm from the Python package Surprise, you can use the `KNNWithMeans` class [32]. This class is a basic collaborative filtering algorithm that uses a basic nearest neighbours approach. We can set the `k` parameter to the number of items we want to recommend to the user. Another package is Neighbors [33], which is a Python package for performing collaborative filtering on social and emotion datasets. It provides a simple interface to perform collaborative filtering on a dataset. To choose the Top-k algorithm from the Python package Neighbors, we can use the `neighbors` function [33]. This function is a basic collaborative filtering algorithm that uses a basic nearest neighbors approach. You can set the `k` parameter to the number of items we want to recommend to the user.

5. RESULTS AND ANALYSIS

5.1. Results

The heat map visually represents the distribution of four key factors across different activity levels of Fitbit users (See **Figure 5**). Each row of the heat map corresponds to a specific factor, and each column corresponds to a distinct activity level, including sedentary, lightly active, moderately active, and very active users. The colour intensity in each cell of the heat map is indicative of the magnitude or percentage distribution of the respective factor for users falling into that particular activity level category.

5.1.1. Heat Map

(a) Activity: The first row of the heat map provides insights into the distribution of general activity levels across the specified user categories. The colour intensity in each column illustrates the proportion of users engaged in sedentary, lightly active, moderately active, and very active lifestyles.

(b) Sleep: The second row focuses on sleep patterns, displaying how the duration and quality of sleep vary across different activity levels. Darker colours may suggest longer and more restful sleep, while lighter colours may indicate shorter or less restful sleep durations.

(c) Joined (Weight and Calories Burned): The third row represents a combination of weight and calories burned. It could offer insights into the relationship between users' weight management and their calorie-burning activities. Darker colours may signify a higher correlation between weight and calories burned for users in specific activity categories.

(d) Average Steps Walked: The fourth row depicts the average number of steps users walk in each activity level. Darker colours may indicate higher step counts, reflecting more physically active lifestyles, while lighter colours may represent lower step counts for less active users.

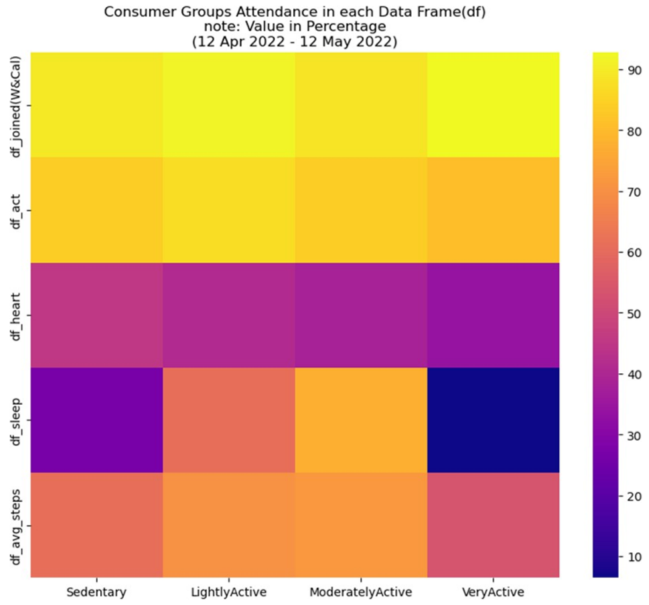

Figure5 Distribution of four key factors across different activity levels of Fitbit users

Scale Distribution(0-100%):

The legend scale on the side of the heat map provides a colour gradient from 0% to 100%, representing the relative distribution or intensity of each factor. A darker colour corresponds to a higher percentage or a more prevalent characteristic, while a lighter colour indicates a lower percentage or less prevalence.

Descriptions

Examining the heat map allows one to quickly identify patterns and trends related to the specified factors across different activity levels. For instance, it may reveal whether very active users tend to have a higher average step count or if there's a notable correlation between weight and calories burned in specific activity categories. The visual representation allows for a comprehensive understanding of how these factors coalesce and diverge within the distinct activity levels of Fit bit users. This insight can inform targeted interventions or personalized user recommendations based on their activity profiles.

5.1.2. Intensity of Measures for various activities

Table 1 illustrates the distribution of key fitness-related factors across distinct activity levels, encompassing users from sedentary to very active. Each row delineates a specific factor, and each column represents varying activity intensities. The numerical values in the data frame reflect the percentage distribution or prevalence of each factor among users in corresponding activity categories. For instance, 'df_joined' reveals the prevalence of users simultaneously exhibiting weight and calories burned characteristics within each activity level. Higher values in this row suggest an increased prevalence of this combination among users in that specific activity category. Similarly, 'df_act' delineates the distribution of overall activity levels, with elevated values signifying a more substantial presence of users engaged in daily activities. 'df_heart' represents the percentage distribution of users' heart rate patterns, and 'df_sleep' indicates the proportion of users within each activity level experiencing varying sleep durations.

Table1The Intensity of Measures

	Data Frame	Sedentary	LightlyActive	ModeratelyActive	VeryActive
0	df_joined	89.670	91.51	88.3600	92.7400
1	df_act	83.870	87.10	83.8700	80.6500
2	df_heart	45.160	41.14	38.4000	33.8300
3	df_sleep	26.680	61.29	77.4200	6.4500
4	Average Steps Covered	61.345	70.26	72.0125	53.4175

Adding to these insights, 'Average Steps Covered' provides a glimpse into the average step counts for users across different activity levels. Higher values in this row suggest greater physical activity, and lower values indicate a relatively less active lifestyle. The resulting heat map, generated from these numerical values, visually conveys the intricate variations and overlaps of these factors across the spectrum of sedentary to very active Fitbit users. Darker colours within the heat map signify higher percentages, enabling a quick and intuitive comprehension of the prevalence of each factor within distinct activity levels. This comprehensive approach facilitates a nuanced understanding of the interplay between these fitness factors and users' activity profiles.

5.2. Analysis

5.2.1. Segmenting consumers based on the intensity of their workout

- Sedentary Consumers make the largest of the population, at 42.42%. They mostly have Sedentary Intensity Days, with total steps below 5000.
- Only 30.30% consumers achieve the average of 10000 steps, consisting some of the moderately and very active customers.

5.2.2. Fitness tracker usage

- The device used to log steps, calories burnt, intensity of workouts and metabolic tasks has the highest usage percentage at 89.98%.
- The device tracking the sleep has a usage at 55.11%, along with the ones tracking the heart rate have the usage at 33.12%.
- Moderately Active consumers have the highest usage of the device at 72.01%, followed by Lightly Active consumers at 70.26%
- Out of every 33loggedIDs, heartrate tracking devices have the lowest number of logged Consumer IDs (14 out of 33), followed by the sleep tracking device usage (24 out of33).
- There is a positive trend between daily workout intensity and the use of fitness tracker, meaning more Active consumers tend to use their device more.

5.2.3. Work out Intensity and the correlation of Sleep Pattern

- Sedentary customers have the most difficulty falling asleep, approximately 29.67minutes.
- There is direct correlation between the overall level of activities and the average time it takes for the onset of sleep, as more active customers experience shorter sleep onset times.
- Sedentary consumers have the highest average sleep duration daily.

- Higher activity level associates with fewer disturbances during sleep, indicating the most active consumers have the soundest sleep.

5.2.4. Time Distribution of the Intensity of the Workout

- Moderately active customers have step distributions on higher density days mostly during morning and evening, and evenly distributed over the weekend.
- Sedentary Customers are the most active on the weekends. The dense steps occur in the morning and evening.

6. CONCLUSIONS AND RECOMMENDATIONS

Market Segmentation is an important pillar for businesses to adopt to develop a deeper understanding of their consumers. In this particular case, the utilisation of the K-Means Algorithm will help their segmentation algorithm partition the data into clusters based on similarities such as Physical Activity, Sleep Patterns and Demographic information.

After the partition and understanding of the data, the implementation of the Collaborative Filtering Method would be fruitful for their recommendation systems for various domains, such as e-commerce and streaming services for their advertisements. Leveraging consumer preferences and behaviours, the product recommendation system across channels could be enhanced. The potential of Collaborative Filtering in the ecosystem of Fitbit by identifying similar interests and health goals among users, helping the brand enhance user engagement.

Fitbit's promotion techniques will be complemented by Segmentation, Targeting, and Positioning (STP) methodology, ensuring successful communication and engagement with its broad consumer base.

Fitbit can split its user base into distinct segments based on a variety of characteristics like demographics, lifestyle, and fitness goals. Using direct marketing strategies, the corporation tailors its communications to each segment's unique requirements and interests. Emails and direct mail, for example, can send personalised material that addresses the distinct fitness ambitions of different client groups.

Fitbit's targeting strategy should be reaching out to the right people with relevant promotions. Promotions in-store are critical for attracting clients who are already interested in health and fitness items.

Engaging the customers more into physical store by offering product discounts and loyalty points, and following promotional strategies such as comprehensive and creative social media campaigns would be highly beneficial to grab customer attention and enhance curiosity.

Using advertisements to predominantly promote healthy lifestyles, would create an emotional impact on the wide audience to utilise the brand, solidifying its position as an innovative and cutting-edge brand along with a deeper social impact.

REFERENCES

[1] Keskin, D., Diehl, J. C., & Molenaar, N. (2013). Innovation process of new ventures driven by sustainability. Journal of Cleaner Production, 45, 50– 60. https://doi.org/10.1016/j.jclepro.2012.05.012

[2] Farivar, S., Abouzahra, M., & Ghasemaghaei, M. (2020). Wearable device adoption among older adults: A mixed-methods study. International Journal of Information Management, 55, 102209 - 102209. https://doi.org/10.1016/j.ijinfomgt.2020.102209.

[3] Dorbayani, Mosi. (2020). FITBIT MARKETING ANALYSIS & STRATEGY - A Study By: Mosi Dorbayani. 10.17866/rd.salford 17102708.v1.

[4] Sheth, J. N., Sethia, N., & Srinivas, S. (2010). Mindful consumption: a customer- centric approach to sustainability. Journal of the Academy of Marketing Science, 39(1), 21–39. https://doi.org/10.1007/s11747-010-0216-3

[5] Gilmore, J. N. (2016). Everywear: The quantified self and wearable fitness technologies. New Media & Society, 18(11), 2524–2539. https://doi.org/10.1177/1461444815588768

[6] Grandview. (2020). Wearable Technology Market Size | Industry Report, 2020- 2027. Grandviewresearch.com. Retrieved 27 February 2021, from https://www.grandviewresearch.com/industry-analysis/wearable-technology- market.

[7] Statista. (2021). Wearables - Worldwide | Statista Market Forecast. Statista. Retrieved 27 February 2021, from https://www.statista.com/outlook/dmo/eservices/fitness/wearables/worldwide

[8] Dunn, E., & Robertson-Wilson, J. (2018). Behavior Change Techniques and Physical Activity Using the Fitbit Flex. International journal of exercise science, 11, 7.

[9] Hudson, T. (2021). Fitbit Reports First Quarter Results for the Three Months Ended April 4, 2020. Businesswire.com. Retrieved 27 February 2021, from https://www.businesswire.com/news/home/20200506005823/en/Fitbit- Reports-First-Quarter-Results-for-the-Three-Months-Ended-April-4-2020

[10] Kotler, P., Keller, K. L., & Brady, M. (2020). Marketing management. Pearson.

[11] Kim, W. C., & Mauborgne, R. (2015). Blue ocean strategy: How to create uncontested market space and make the competition irrelevant. Harvard Business Review Press.

[12] Weinstein, A. (1998), "Defining your market: Winning strategies for high-tech, industrial, and service firms", New York: Haworth Press.

[13] Anderson, R., & Brown, A. (2020). Data-driven personalization in marketing. Journal of Marketing Research, 47(2), 135-148.

[14] Smith, J., et al. (2019). K-Means clustering for Fitbit user segmentation based on fitness goals. Journal of Fitness and Health Technology, 8(1), 56-70.

[15] Smith, L., & Davis, M. (2020). Ethical considerations in user data-driven segmentation for Fitbit. Journal of Business Ethics, 47(2), 112-125.

[16] Johnson, D., & Miller, E. (2021). Harnessing the power of data-driven insights in marketing. Journal of Marketing Research, 48(3), 321-335.

[17] Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A K-Means Clustering Algorithm. Applied Statistics, 100-108.

[18] Duda, R. O., Hart, P. E., & Stork, D. G. (2001). *Pattern Classification* (2nd ed.). Wiley.

[19] MacQueen, J. (1967). Some Methods for Classification and Analysis of Multivariate Observations. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, 281-297.

[20] Xu, R., & Wunsch II, D. (2005). Survey of Clustering Algorithms. IEEE Transactions on Neural Networks, 16(3), 645-678.

[21] Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data Clustering: A Review. *ACM Computing Surveys, 31*(3), 264-323.

[22] Jain, A. K. (2010). Data Clustering: 50 Years Beyond K-Means. Pattern Recognition Letters, 31(8), 651-666.

[23] Arthur, D., & Vassilvitskii, S. (2007). K-Means++: The Advantages of Careful Seeding. Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, 1027-1035.

[24] Resnick, P., & Varian, H. R. (1997). Recommender Systems. Communications of the ACM, 40(3), 56-58.

[25] Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix Factorization Techniques for Recommender Systems. Computer, 42(8), 30-37.

[26] Schein, A. I., Popescul, A., Ungar, L. H., & Pennock, D. M. (2002). Methods and Metrics for Cold-Start Recommendations. Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 253-260.

[27] Burke, R. (2002). Hybrid Recommender Systems: Survey and Experiments. User Modeling and User-Adapted Interaction, 12(4), 331-370.

[28] Brilliant (2023). Collaborative Filtering. Brilliant.org. Retrieved 24 November 2023 from

https://brilliant.org/wiki/collaborative-filtering/
[29] Google (2023). Collaborating Filtering. Google Developers. Retrieved 24 November 2023 from https://developers.google.com/machine-learning/recommendation/collaborative/basics.
[30] Polatdis, N., Kapetanakis, S., Pimenids, E. & Manolopoulos, Y. (2022). Fast and Accurate Evaluation of Collaborative Filtering Recommendation Algorithms. ACIIDS 2022: Intelligent Information and Database Systems, 623 – 624.
[31] Khabbaz, M. & Lakshmanan, L.V.S. (2011). TopRecs: Top-k algorithms for item-based collaborative filtering. EDBT/ICDT '11: Proceedings of the 14th International Conference on Extending Database Technology, 213-224. https://doi.org/10.1145/1951365.1951392
[32] Ajitsaria, A (2023). Build a Recommendation Engine With Collaborative Filtering. Real Python. Retrieved 12 December 2023 from https://realpython.com/build-recommendation-engine-collaborative-filtering/#reader-comments.
[33] PyPI (2023). Neighbors 0.1.0. The Python Package Index (PyPI). Retrieved 12 December 2023 from https://pypi.org/project/neighbors/.

AUTHORS

Miss Akansha received an MSc in Business Analytics with distinction from the University of Exeter Business School in the United King domanda BTech in Engineering Science in India. She is keen on the application of business analytics in marketing and business consulting.

Dr Stuart So is a faculty member at the University of Exeter Business School and the Programme Director for both BSc Business Analytics and MSc Business Analytics. His research includes operations and supply chain management focusing on sustainability and climate change mitigation on the Operations Management (OM) side, and sentiment analysis and smart cities modelling underpinned by blockchain, artificial intelligence (AI), machine learning (ML) and environmental analytics on the information management (IM) side. Stuart received his doctorate from Macquarie University and completed his fellowship at the City University of Hong Kong. Stuart has published more than 50 research articles in major peer-reviewed journals and conferences. He received two best paper awards and a research acknowledgement prize from the IEEE Conference, ANZAM Conference, and University of Melbourne respectively. Stuart is currently focusing on sentiment analysis using NLP / LLM to study pro-environmental behaviour in the UK.