Case Study: Bellbeat’s selling fitness through fashion

This case study is part of the Google Data Analytics Capstone. A fictionalized company needs data driven insights to design a marketing strategy which converts a user of casual membership into a annual membership holder. This analysis is performed by following the steps of the data analysis process: **Ask, Prepare, Process, Analyze, Share, and Act.**

# **Scenario**

I am a junior data analyst working on the marketing analyst team at Bellabeat, a high-tech manufacturer of health-focused products for women. Bellabeat is a successful small company, but has the potential to become a larger player in the global smart device market. Urška Sršen, cofounder and Chief Creative Officer, believes that analyzing smart device fitness data could help unlock new growth opportunities for the company. I have been tasked to analyze smart device data to gain insight into how consumers are using their smart devices. The insights I discover will then help guide the marketing strategy for the company.

# **Background**

Urška Sršen and Sando Mur founded Bellabeat, a high-tech company that manufactures health-focused smart products. Sršen used her background as an artist to develop beautifully designed technology that informs and inspires women around the world. Collecting data on activity, sleep, stress, and reproductive health has allowed Bellabeat to empower women with knowledge about their own health and habits. Since it was founded in 2013, Bellabeat has grown rapidly and quickly positioned itself as a tech-driven wellness company for women.

Sršen knows that an analysis of Bellabeat’s available consumer data would reveal more opportunities for growth. She has asked the marketing analytics team to focus on a Bellabeat product and analyze smart device usage data in order to gain insight into how people are already using their smart devices. Then, using this information, she would like high-level recommendations for how these trends can inform Bellabeat marketing strategy

Data Analytics Process

Step 1: Asking questions

The primary focus of the project assigned is to analyze smart device usage data in order to gain insight into how consumers use non-Bellabeat smart devices. Using these insights, a list of recommendations must be presented to devise a comprehensive marketing strategy to being in new customers. To further simplify the analysis process, scope of work and stakeholder expectations has been defined.

SCOPE OF WORK:

* Acquire fitness data of any one category of non-Bellabeat smart device which the company has a product of.
* Identify usage patterns and trends through the fitness data
* Find how to apply these trends to Bellabeat customers.
* Generate strategies to market the Bellabeat product using data-driven analysis.

STAKEHOLDERS:

1. Urška Sršen, Bellabeat’s cofounder and Chief Creative Officer
2. Sando Mur, Bellabeat’s cofounder and a key member of the Bellabeat executive team
3. Marketing analytics team

Three questions will guide the future marketing program:

1. What are some trends in smart device usage?
2. How could these trends apply to Bellabeat customers?
3. How could these trends help influence Bellabeat marketing strategy?

Step 2: Collection and Preparation of Data

**Data Source**

* The Dataset is made available through Mobius and is on Kaggle. It containing personal fitness tracker records from thirty Fitbit users. The data is [licensed](https://creativecommons.org/publicdomain/zero/1.0/) and pre-validated for reliability and fairness. It is a public dataset.
* The downloaded data contained 2 zip files with varying quantities of data. Data is zip extracted and stored securely.
* Data download [URL](https://www.kaggle.com/datasets/arashnic/fitbit).

**Data features**

* The data is in long format with some files featuring long converted to wide format data. This will be ignored for analysis. The data is grouped by user-id and is pre-sorted date-time wise.
* The data is from 12 march 2016 to 12 may 2016. It is outdated and insights gained may not be as effective.
* There is no demographic info like gender, age, location etc. This could affect the quality of insights as Bellabeat focuses particular on women and fashion market segment.
* The dataset includes information on step count, sleep, activity, calories etc. and is measured minute, hourly and daily bases.

Step 3: Processing and Data transformation

Tool used: Rstudio

The decision to utilize R stems from its capability to seamlessly integrate visualization and analysis tasks through concise code, offering a time-effective and efficient alternative to traditional spreadsheet software

Before starting with data cleaning, some basic inspection of each file:

* There are 2 distinct zip file with different number of files in them. First file contains data from 12 March to 11 April while second from 12 April to 12 May.
* First file has 11 items in it whereas 18 items are present in second one. It appears that second file has all the same files from first plus 1 file converted to wide format, heartrate per second data, and daily data of steps, calories, activity and intensity data.
* Minutes and hours file provide detailed information on particular activity but daily files summarize this information which is better for analysis.
* Weight logs lack consistency and have only 11 entries. This data is unreliable, inconsistent and cannot be correlated with rest of the dataset.

Through preliminary analysis of files available, it is decided that of all the data only the following will be used for analysis:

1. Daily Sleep from 12 April to 12 May

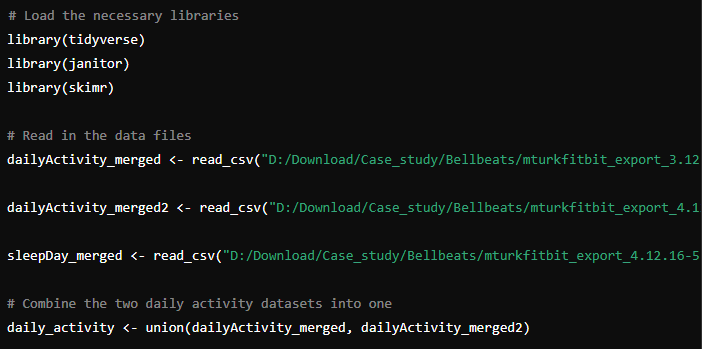
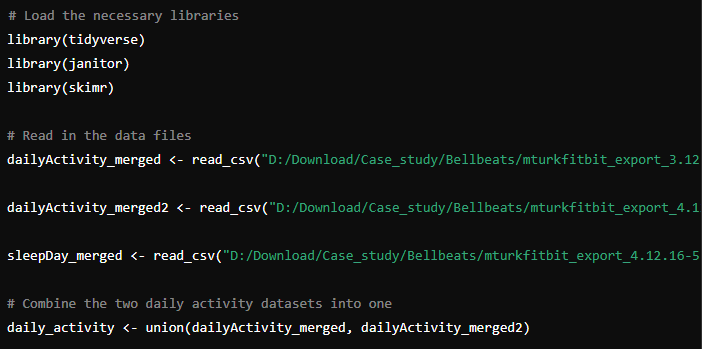
Few limitation of the data:

* No consistency between datasets.
* Data is outdated
* Small sample size which leads to sampling bias.

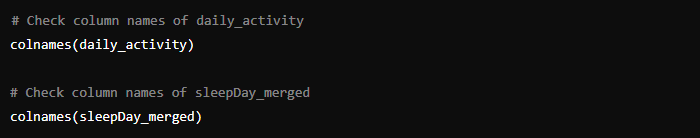
**Data Cleaning**

Data cleaning is a process that involves transforming raw data into a consistent format that can be easily analyzed.

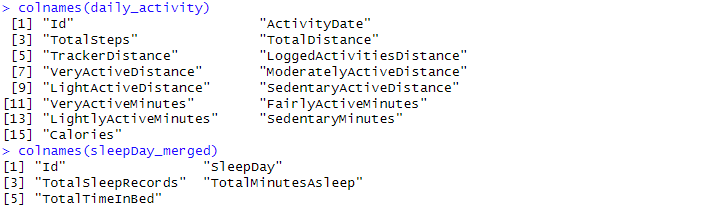
1. Let’s start with first importing data:



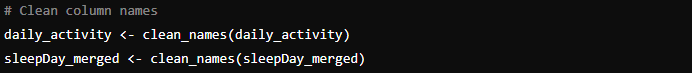
1. Now that datasets have been imported, it is crucial to check column names using colname() function to see if everything was imported correctly.



We get the following output:



Some names have uppercase and no spaces. This will affect the analysis process. Let’s fix that:

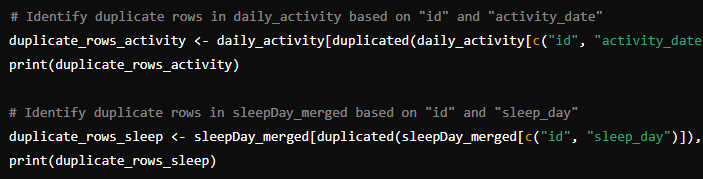


1. After verifying and fixing column names, let’s start finding any NA records. The following R code is run:

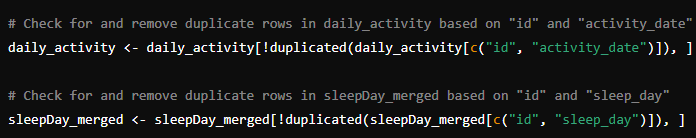


The output generated shows no irregularities in dataset and states that no NA record found.

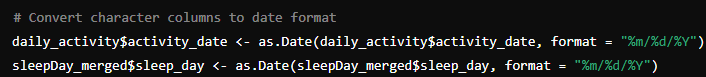
1. Now to make sure analysis results natural duplicates need to be removed. This code will let us identify duplicate records:



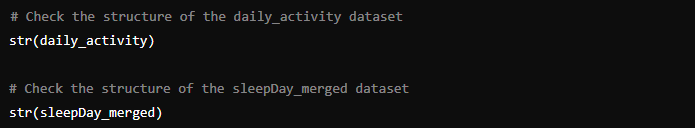
It appears there are duplicates present in both datasets. To remove them this code was run:



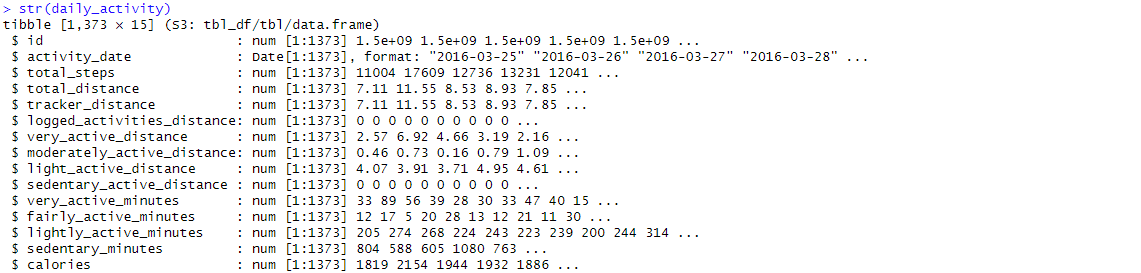
1. Structure of the dataframe is checked using str() function. It appers that **date** is in char format and needs to be converted to a usable date format. The following code can fix that:



Verify all results with a final check.



The final output after data cleaning is

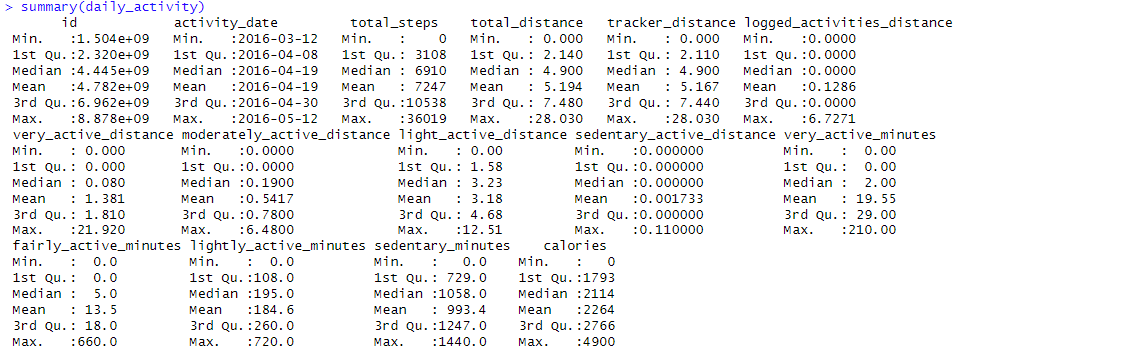


The data is verified to be cleaned and transformed likewise now ready for Visualization.

STEP 4: Data Analysis and Sharing insights through visualisation

Tool used: RStudio

A summary of the cleaned dataset is first produced



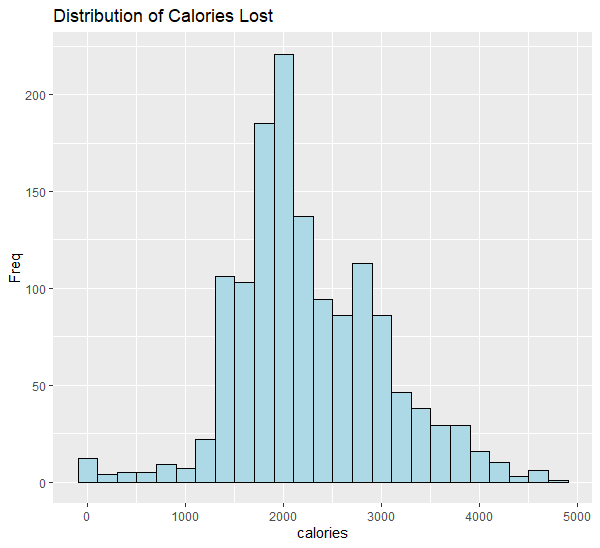
Overall, The **daily\_activity** dataset contains information about various physical activities and related metrics from 33 different users. The activity dates span from March 12 to May 12, 2016. In terms of total steps taken by users, the values range from a minimum of 0 to a maximum of 36,019, with a mean of 7,247 steps. The total distance covered by users varies from 0.000 to 28.030 Km, averaging 5.194 Km. The tracker distance show a similar metrics as total distance with a tiny variance. The dataset indicates that most users did not log additional activities, as the logged activities distance has a mean of 0.1286 Km, and the max of only 6.7271 Km. The very active distance ranges from 0.000 to 21.920 Km, with an average of 1.381 Km, while the moderately active distance ranges from 0.000 to 6.480 Km, averaging 0.5417 Km. The light active distance ranges from 0.00 to 12.51 Km, with a mean of 3.72 Km. Sedentary active distance is almost negligible, with a mean of 0.001733 Km. In terms of activity minutes, very active minutes range from 0.00 to 210.00 minutes, with an average of 19.55 minutes. Fairly active minutes range from 0.0 to 660.0 minutes, averaging 13.6 minutes. Lightly active minutes range from 0.0 to 720.0 minutes, with a mean of 192.8 minutes. Sedentary minutes vary from 0 to 1440 minutes, with a mean of 991.2 minutes. Finally, the calories burned range from 0 to 4900, with an average of 2264 calories.

**Data Analysis**:

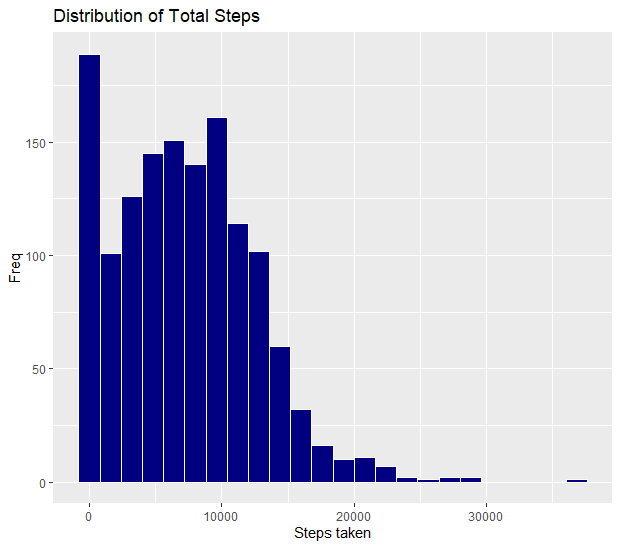
Products sold by Bellabeats try to attract women who wish to understand their daily lives or want to control their health without compromising on their fashion. Looking at this, we need to identify how a product similar to one sold by Bellabeat gets used.

Let’s first establish the frequency of crucial metrics from the data. A histogram can visualize this efficiently.

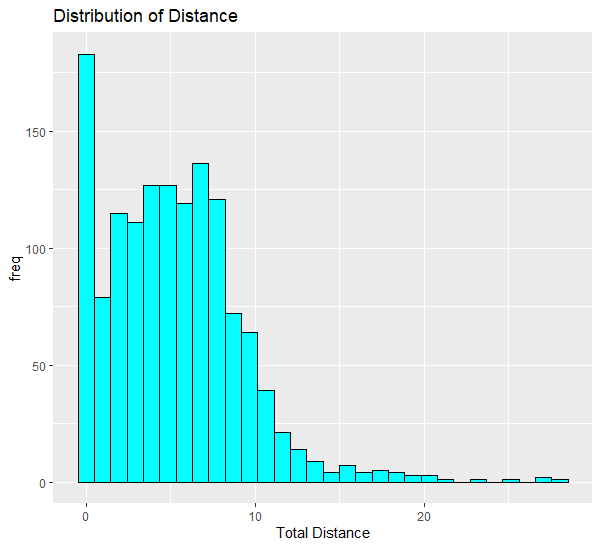
1. Distribution of metrics in Daily Activity data



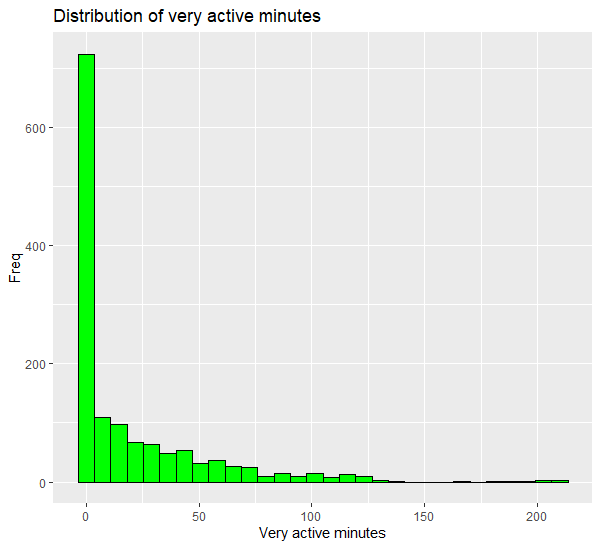
A normal distribution is observed where around 2000 calories burned is the most frequent value.



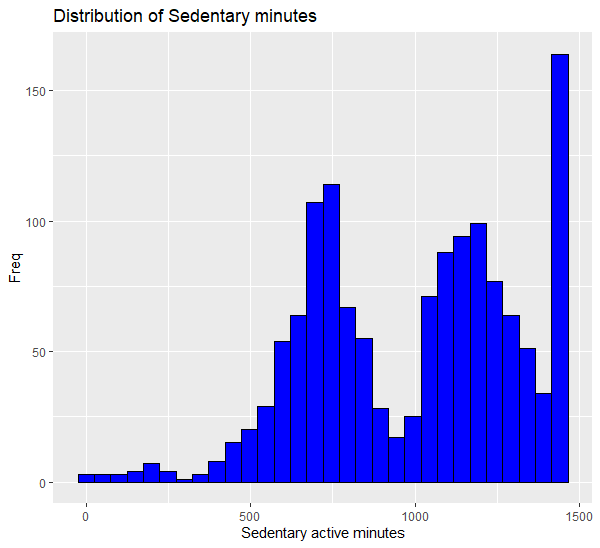
Most frequent values in total steps column appear to be between 0 and 10000.



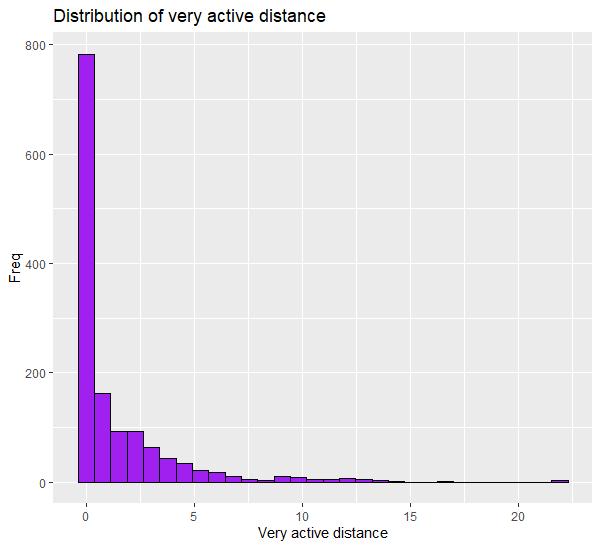
Similar distribution to Total steps graph. This indicates a correlation between them.



Large quantities of Very Active minutes values seem to be 0. The graph also shows drop in frequency as minutes increase.

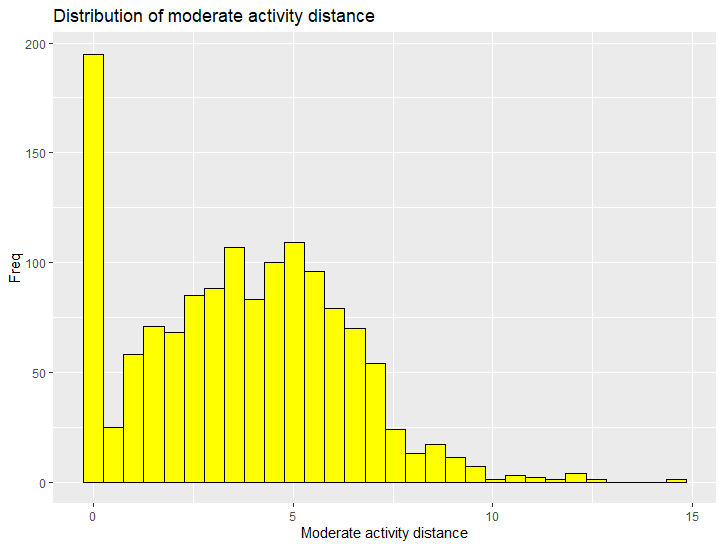


The frequency of the minutes in sedentary minutes is more towards higher values, signifying the product is for majority used for immobile activities.

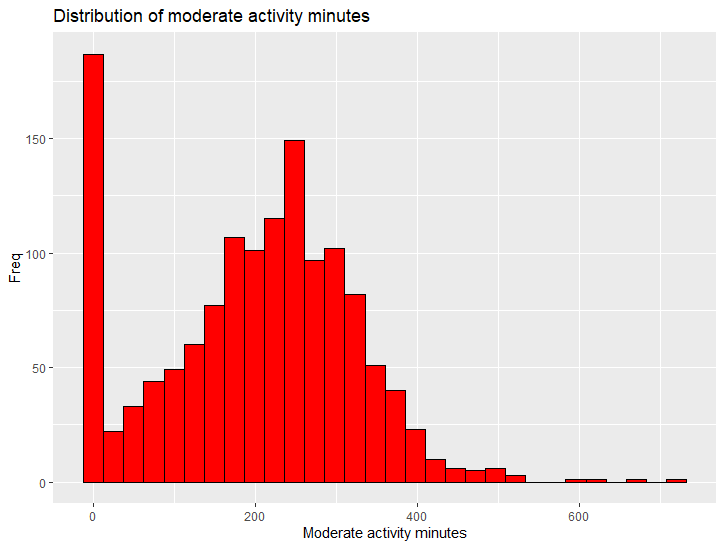


As the very-active distance increases, the frequency of it decreases in the data.

For simplification of analysis, moderately active distance and light active distance is combined into moderate-activity distance. Similarly, fairly active minutes and lightly active minutes is combined moderate-activity minutes. 



Higher frequency of around 5 km mark but the highest is 0 km distance.



Zero Minutes has the highest frequency but after this a normal distribution is observed.

Summary:

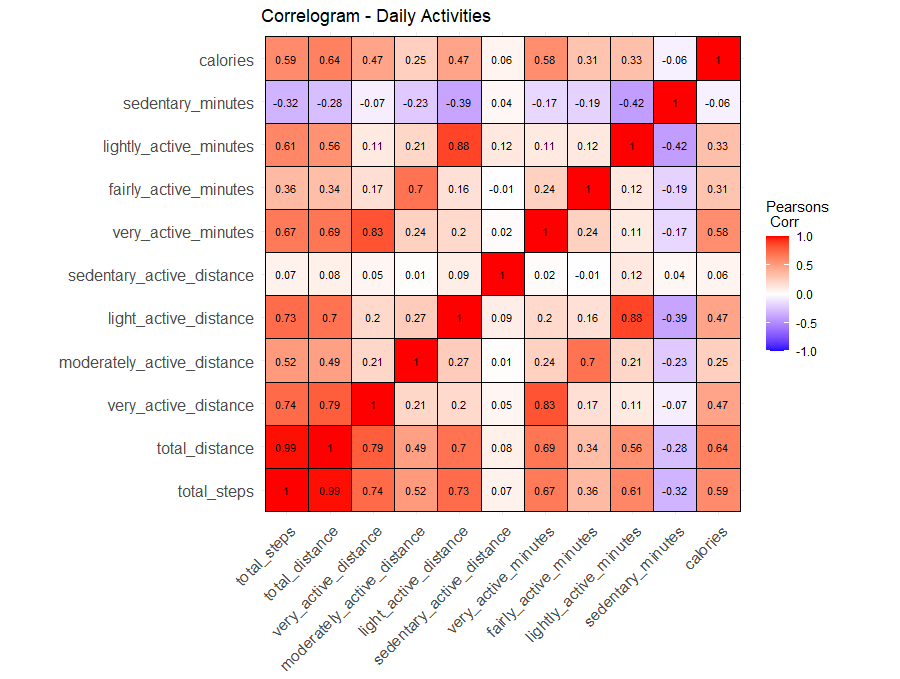
It is observed that sedentary minutes is higher for all participants, very active distance is relatively high

And total steps and total distance seem to compliment each other. For sleep data, both time in bed and time asleep mirror each other in terms of distribution of values. Total calories burned mimic the distance traveled and steps taken graph.

Following insights are obtained:

* Fitbit appears to be used to measure steps taken, distance travelled, and minutes spend being active or sedentary.
* Majority of participants spend their time doing sedentary and moderately active actions.
* Time spent in bed and time asleep mimic each other. It is observed that all participants also prefer using their product to measure sleep.

A correlation analysis can help identify which metrics complement one another.



Summary:

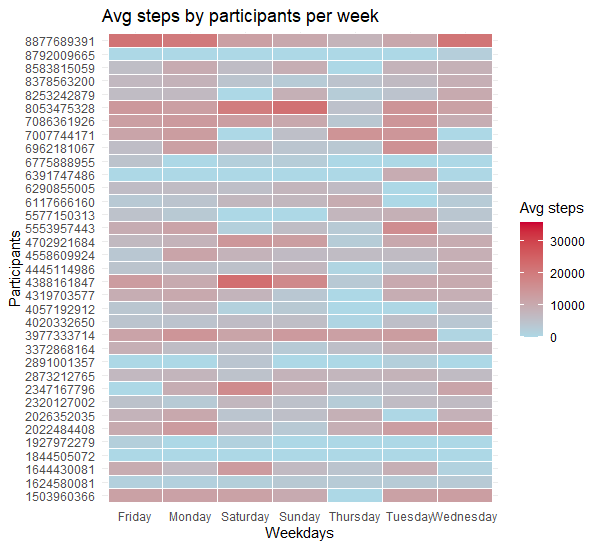
* **Total Steps and Total Distance**:
  + There is a perfect positive correlation (1.0) between total steps and total distance, indicating that as total steps increase, total distance also increases proportionally.
* **Calories**:
  + Calories burned are strongly positively correlated with total steps (0.59), total distance (0.64), very active minutes (0.67), and light active distance (0.73). This suggests that higher activity levels and greater distances covered are associated with more calories burned.
* **Very Active Minutes**:
  + Very active minutes have high positive correlations with very active distance (0.83) and total steps (0.67), indicating that more intense activity leads to longer distances and more steps.
* **Lightly Active Minutes**:
  + Lightly active minutes are highly correlated with light active distance (0.88), showing that more time spent in light activity translates to covering more light active distance.
* **Sedentary Minutes**:
  + Sedentary minutes show a negative correlation with total steps (-0.32), total distance (-0.28), and calories burned (-0.06), suggesting that more sedentary time is associated with lower levels of physical activity.
* **Fairly Active Minutes**:
  + Fairly active minutes have moderate positive correlations with light active distance (0.21) and very active minutes (0.36), indicating some overlap in these activity types.
* **Distances**:
  + Very active distance is strongly correlated with total distance (0.79) and total steps (0.74), indicating that more intense activities contribute significantly to the total distance and step count.
  + Light active distance and moderately active distance have a moderate positive correlation (0.49).

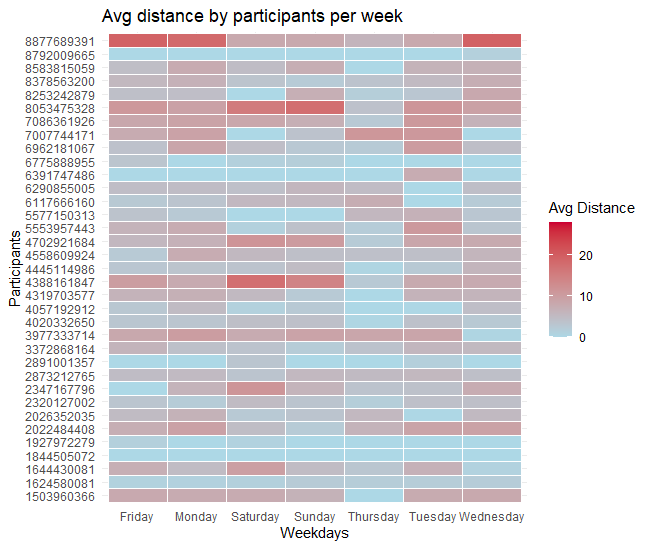
Overall, the correlogram highlights the strong positive relationships between more intense physical activities (very active minutes and distances) and metrics such as total steps, total distance, and calories burned. Conversely, more sedentary time is negatively associated with these activity metrics

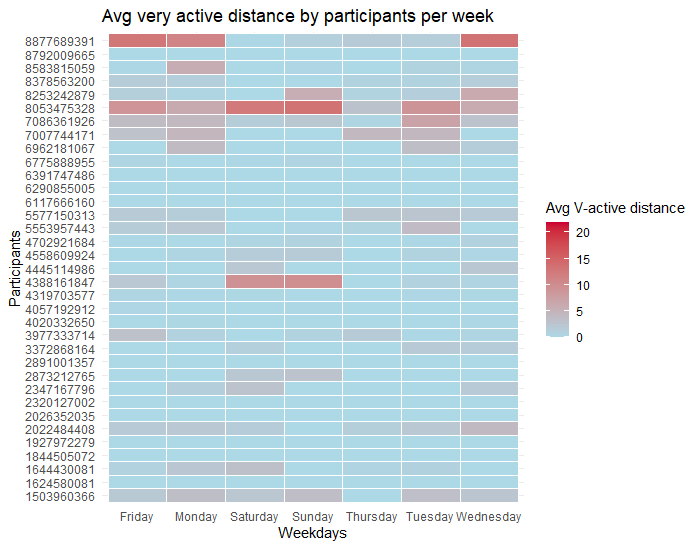
1. Further Exploration using additional variables.

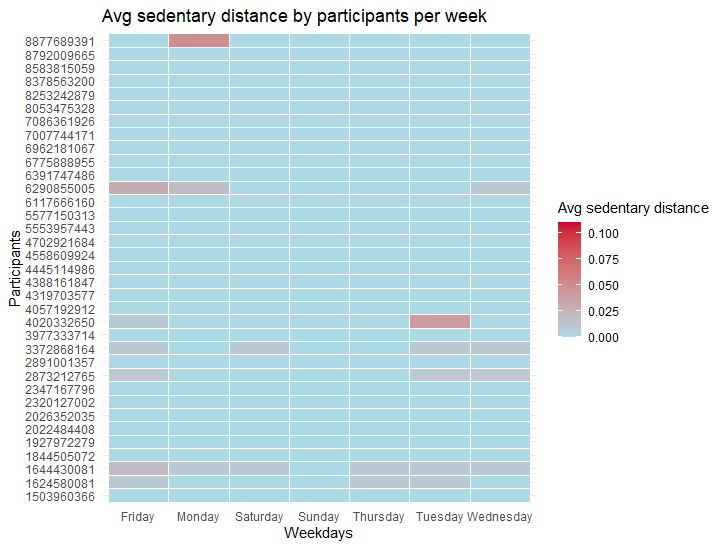
Analyzing characteristics of user behavior:

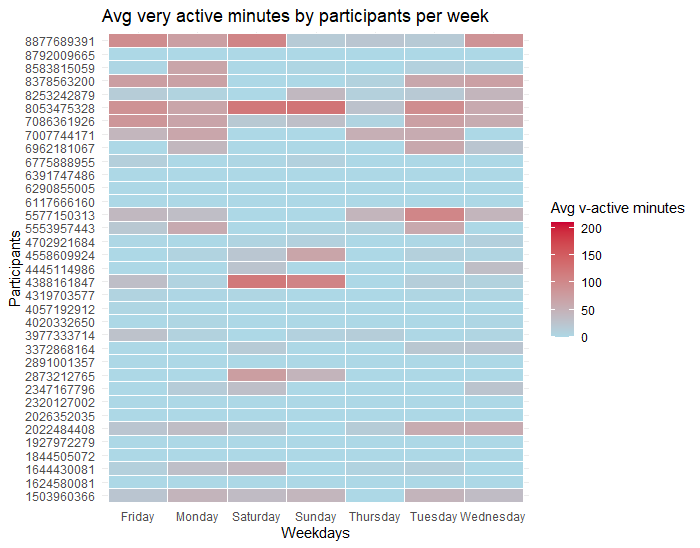
Using tile graph to visualize mean of all metrics in daily activity dataset per participant on weekly basis.

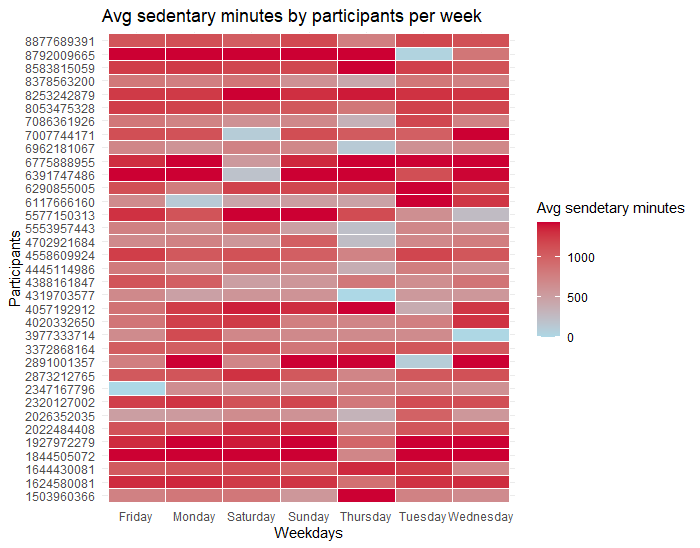


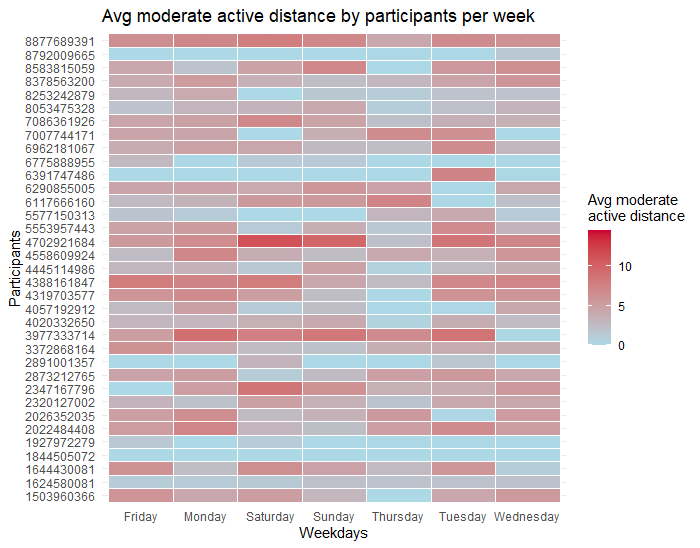


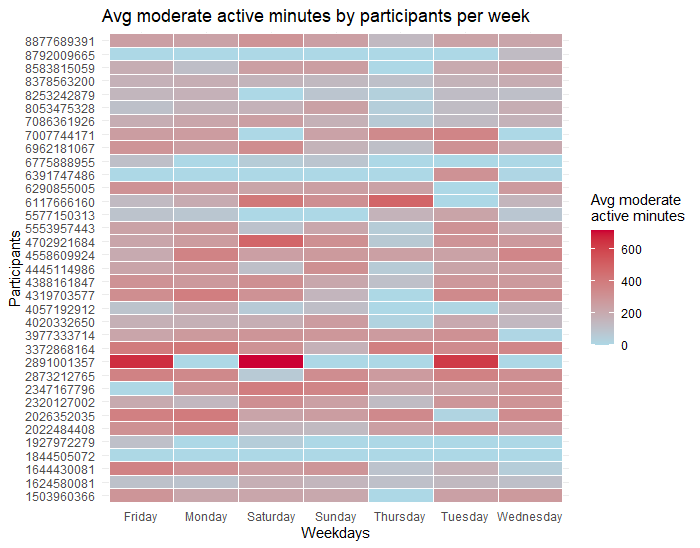


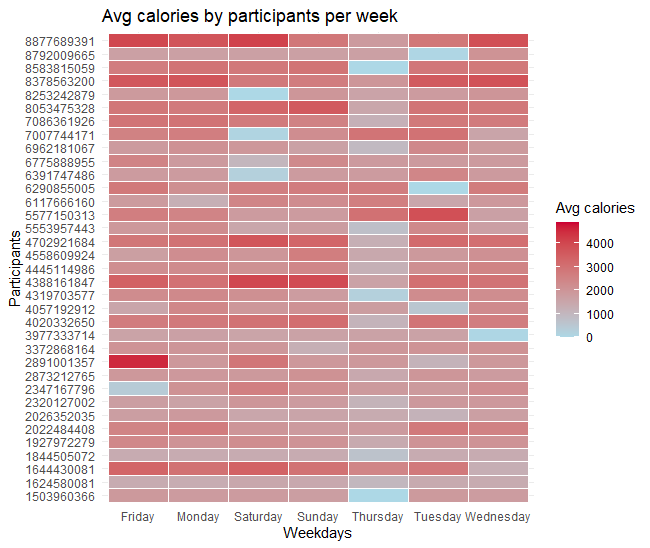












Summary:

1. Calorie Burn: Participants show varied calorie burn throughout the week. Some participants burn more calories consistently across all days, while others have specific days (e.g., Friday, Monday) with higher or lower calorie burns. Sundays and Tuesdays have slightly lower average calorie burns for many participants.
2. Moderate Active Minutes: There is a mix of moderate activity minutes across the week. Certain participants exhibit high moderate active minutes on specific days like Sunday and Thursday, while others show lower activity. Fridays and Mondays generally show lower moderate active minutes for several participants.
3. Moderate Active Distance: The patterns for moderate active distance are similar to moderate active minutes, with some participants showing higher distances on particular days like Sunday and Thursday. Mondays and Fridays tend to have lower moderate active distances for several participants.
4. Very Active Minutes: Very active minutes are sporadic and vary significantly across participants. Some participants have notable spikes in very active minutes on specific days such as Tuesday, Sunday, and Thursday. A large portion of the participants have low very active minutes throughout the week.
5. Sedentary Distance: Sedentary distance remains quite low across all participants and days. There are minimal instances of sedentary distance, indicating that participants are generally either active or stationary without much in-between movement.
6. Very Active Distance: Similar to very active minutes, very active distances show sporadic spikes. Specific days such as Sunday and Thursday have higher very active distances for some participants, suggesting increased activity levels.
7. Average distance and steps: It appear that both graphs portray similar visuals. Thursdays have the least distance and steps taken for all participants while Saturday and Sunday has the most.

STEP 5: Act

After identifying user behavior when using Fitbit. we can develop targeted marketing strategies to attract new customers and upgrade existing customer into membership programs. Though the dataset had some limitations like no demographic info, outdated data, missing key records leading to inconsistency etc., which affect the efficacy of insights generated, strategies will still be very effective.

**Insights from Observations**

1. **Calorie Burn and Activity Correlation**:
   * Higher activity levels, including increased steps, active minutes, and active distances, are strongly correlated with higher calorie burn.
   * Users tend to have varying levels of activity throughout the week, with specific days showing lower or higher activity levels.
2. **Moderate and Very Active Patterns**:
   * Moderate and very active minutes and distances show clear patterns, with some days being significantly more active than others. For example, Sunday and Thursday are more active for some users, while Monday and Friday are generally less active.
   * Very active minutes and distances are sporadic but show significant spikes on certain days, suggesting intense activity sessions.
3. **Sedentary Behavior**:
   * Sedentary distance remains low, indicating that users are either actively moving or stationary with minimal in-between movement.
   * There is a negative correlation between sedentary time and other activity metrics, suggesting that increased sedentary time is associated with lower physical activity levels.
4. **Total Distance and Steps**:
   * There is a perfect positive correlation between total steps and total distance, indicating a consistent relationship between the two metrics.
   * Thursdays show the least distance and steps taken for all participants, while weekends (Saturday and Sunday) see the most.

The insights suggest a story for each fitbit users. The analysis provides trends limited to Daily activity of each participant of the survey, hence the marketing strategies are for products which can provide services related to daily activity.

**Bellabeat App**

Provides comprehensive health data tracking capabilities when using Bellabeat products and if a **Bellabeat membership** holder it will highlight how users can monitor and improve various aspects of their health including activity, sleep, stress, menstrual cycle, and mindfulness. Following feature of app must be included in marketing:

* **Personalized Health Insights**: Promote the app’s ability to provide personalized health insights based on the user’s data. Highlight success stories of users who have improved their health metrics using the app.
* **Cross-Promotion**: Encourage users to connect the app with Bellabeat’s smart wellness products for a seamless and integrated health tracking experience.
* **Weekly and Daily Challenges**: Introduce weekly challenges within the app to motivate users to be more active on their less active days, such as Mondays and Fridays.

**Leaf Wellness Tracker**

The Leaf tracker can help users monitor their activity levels and identify days with lower activity to improve overall wellness. To sell more Leaf Wellness tracked it is necessary to pursue fashion-oriented market alongside fitness market. Following are the recommendation for marketing:

* **Versatility and Style**: Emphasize the Leaf’s versatile design, which can be worn as a bracelet, necklace, or clip, making it a stylish accessory that fits any outfit.
* **Activity Reminders**: Highlight the tracker’s ability to remind users to be more active on their low-activity days, promoting better overall health.
* **Success Stories**: Share testimonials from users who have used the Leaf tracker to improve their daily activity and health on app. Promote sharing these stats to other social-media platforms

**Time Wellness Watch**

The Time watch provides a classic look combined with smart technology to track activity, sleep, and stress, helping users maintain a balanced lifestyle. Though no analysis has been done for sleep activity, following recommendation, based on sedentary data, can still applicable as marketing strategy

**Marketing Strategy**:

* **Classic Elegance with Smart Functionality**: Market the Time watch as a stylish accessory that doesn’t compromise on functionality. Emphasize its ability to track important health metrics while looking elegant.
* **Sleep and Stress Management**: Promote features that help users manage their sleep and stress, crucial for overall well-being. Share tips on how the watch can aid in improving sleep quality and reducing stress.
* **Daily Insights**: Highlight how the watch, in combination with the Bellabeat app, provides daily insights into the user’s wellness, helping them make informed health decisions.

**Bellabeat Membership**

The membership offers personalized guidance on various aspects of health and wellness, which can be tailored to individual user needs based on their activity patterns.

* **Seamless integration with All Bellbeat Products**: Emphasize the 24/7 access to personalized guidance on nutrition, activity, sleep, health, and mindfulness. Show how this personalized approach can help users achieve their health goals.
* **Exclusive Content**: Promote the exclusive content and benefits available to members, such as tailored workout plans, nutrition advice, and mindfulness exercises.
* **Community Engagement**: Foster a sense of community among members by organizing virtual events, webinars, and challenges that encourage active participation and engagement.
* **Trial Memberships**: Offer trial memberships to allow potential users to experience the benefits firsthand, increasing the likelihood of conversion to paid memberships.