# Problem statement

Bellabeat’s products and analyze sma device data to gain insight into how consumers are using their sma devices. The

insights you discover will then help guide marketing strategy for the company. You will present your analysis to the Bellabeat

executive team along with your high-level recommendations for Bellabeat’s marketing strategy.

# Ask

## Business Task

To identify potential opportunities for growth and provide recommendations for the Bellabeat marketing strategy improvement based on trends in smart device usage.

**Key Stakeholders:**

* Urška Sršen: Bellabeat's cofounder and Chief Creative Officer
* Sando Mur: Mathematician and Bellabeat’s co-founder

**Questions to explore for the analysis:**

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: Prepare

The data being used in this case study can be found here: [FitBit Fitness Tracker Data](https://www.kaggle.com/datasets/arashnic/fitbit) CC0: Public Domain, dataset made available through [Mobius](https://www.kaggle.com/arashnic)

The data is stored and uploaded in R Studio. This Kaggle data set contains personal fitness tracker from thirty fitbit users. Thirty eligible Fitbit users consented to the submission of personal tracker data, including minute-level output for physical activity, heart rate, and sleep monitoring. It includes information about daily activity, steps, and heart rate that can be used to explore users’ habits.

The data set contains 18 CSV files organized in long format. Below is a breakdown of the data using the ROCCC approach:

* **Reliability - LOW:** The data comes from 30 fitbit users who consented to the submission of personal tracker data, including minute-level output for physical activity, heart rate, and sleep monitoring.
* **Original - LOW:** Third party data collect using Amazon Mechanical Turk.
* **Comprehensive - MED:** The dataset contains multiple fields on daily activity intensity, calories used, daily steps taken, daily sleep time and weight record.
* **Current - LOW:** This data is from March 2016 through May 2016. The data is not current, meaning that user habits may have changed over the years.
* **Cited - LOW:** Data was collected from a third party, therefore unknown.

# Step 3: Process

Now we will start cleaning errors from data sets in spreadsheet in excel.

I import daily\_activity\_merged.csv, dailyCalories\_merged.csv, dailyIntensities\_merged.csv, WeightLog.csv, SleepDay\_merged.csv.

Data is now cleaned and we will start investigating the data.

# Step4: Analyze

To summarize the above data, there are 33 participants in the activity, calories, and intensities datasets, 24 in the sleep dataset, and only 8 in the weight dataset. The fact that there are only 8 participants in the weight dataset means that more data would be needed to make a strong recommendation or conclusion.

-No significant change in weight of 8 participants

|  |  |  |  |
| --- | --- | --- | --- |
| **Row Labels** | **Average of WeightKg** | **Max of WeightKg** | **Min of WeightKg** |
| 1503960366 | 52.59999847 | 52.59999847 | 52.59999847 |
| 1927972279 | 133.5 | 133.5 | 133.5 |
| 2873212765 | 57 | 57.29999924 | 56.70000076 |
| 4319703577 | 72.35000229 | 72.40000153 | 72.30000305 |
| 4558609924 | 69.63999939 | 70.30000305 | 69.09999847 |
| 5577150313 | 90.69999695 | 90.69999695 | 90.69999695 |
| 6962181067 | 61.55333379 | 62.5 | 61 |
| 8877689391 | 85.14583429 | 85.80000305 | 84 |
| **Grand Total** | **72.03582137** | **133.5** | **52.59999847** |

Due to less data I cant analyze it more.

## Calories vs Step

|  |
| --- |
| Correlation b/w Calories and Steps |
|  |
| 0.591568086 |
| reuslt is positive more steps they take more they will burn calories. |

## TimeAsLeep vs Time in Bed

|  |
| --- |
| **Corellation b/w TimeAslepp vs Time in Bed** |
|  |
| 0.113975 |
| Result is positive bellabeat should consider scheduling a more consisitency in sleep timing |

From looking at the graph above, we can see there is a negative correlation between SedentaryMinutes and TotalMinutesAsleep. This means that the less active a participant is, the less sleep they tend to get.

cor(merged\_df$TotalMinutesAsleep,merged\_df$SedentaryMinutes)

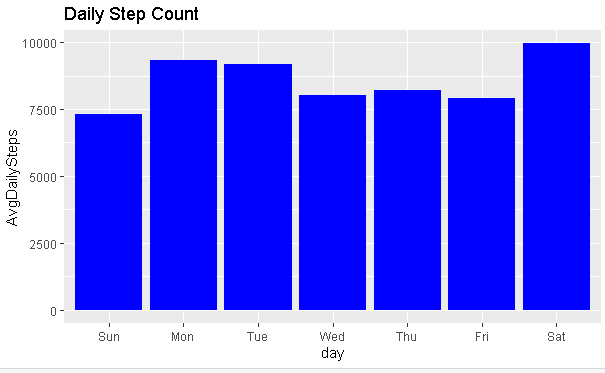
[1] -0.599394

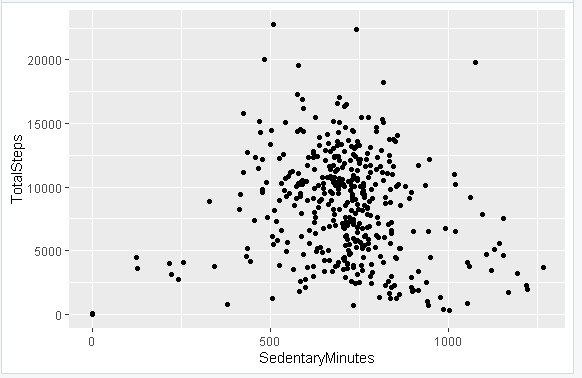
### Observations made from the above summaries:

* Sedentary minutes on average is 16.5 hours
* The average number of steps per day is 7638. The CDC recommends people take 10,000 steps daily.
* The majority of the participants are lightly active.
* The average participant burns 97 calories per hour
* On average, participants sleep for 7 hours.

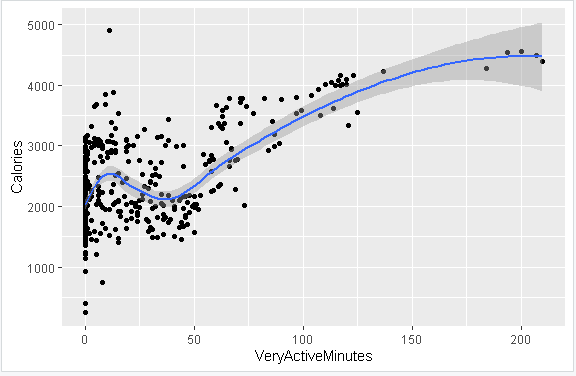
Now I want to look at whether the day of the week affects our activity levels and sleep.



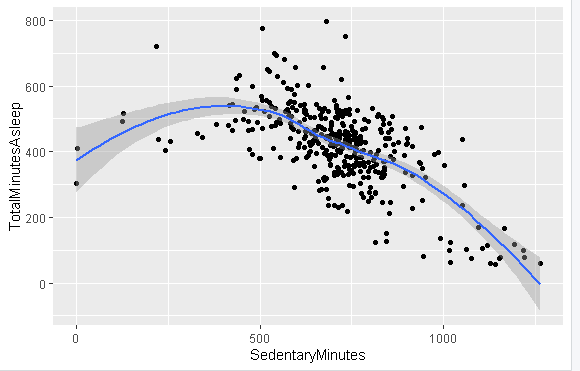




As you can see that mostly users who are 10 hour inactive, their total steps are about 7k-10k.

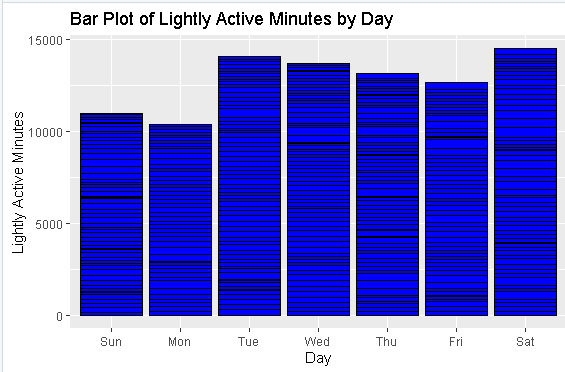


Calories burn more for users having a very active minutes between 60-130 minutes. Avg users are in between 30-50 m,inutes.

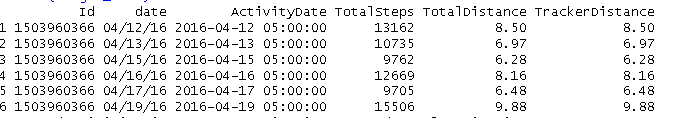


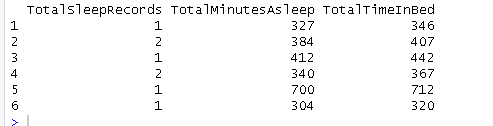
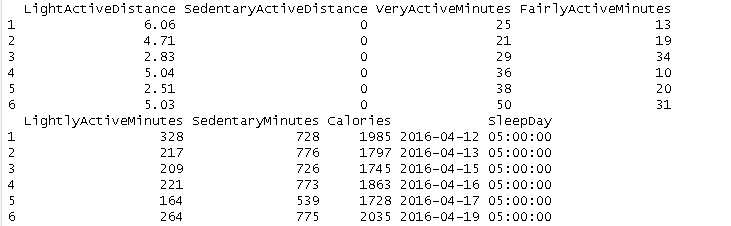
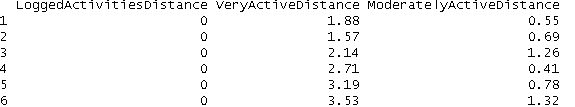
Users who are 14-16 hour inactive takes less sleep. Medium level users are at low risk. Neither too much high inactiveness nor too much low inactiveness.

### Lightly Active Time By Day



Mostly participants are lightly Active they should need a more guidance to stay fairly active to burn calories and for healthy lifestyle.





# **Step 6: Act**

# Conclusion

* Avg Daily Step is between 7k – 9k while 10k steps are necessary for adults. 10k steps are low risk for any damage.
* App should notify the user to stay active. Average sedentary time is 12 hour.
* Mostly participants are lightly Active they should need a more guidance to stay fairly active to burn calories and for healthy lifestyle.
* Saturday is day where huge amount of users stay active. Low number of users are active on Sunday thery need to follow same rule as of Saturday.