**Six phases of a Data analytics project:**

* Ask
* Prepare
* Process
* Analyze
* Share
* Act

**Background**:

You are 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 they have the potential to become a larger player in the global smart device market. Urška Sršen, cofounder and Chief Creative Officer of Bellabeat, believes that analyzing smart device fitness data could help unlock new growth opportunities for the company. You have been asked to focus on one of Bellabeat’s products and analyze smart device data to gain insight into how consumers are using their smart 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.

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 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.

**Ask**

These are few questions that can be answered to gain overview of smart device trends.

1. What are some trends in smart device usage?

Firstly, we will find general trends from the customer usage data of smart devices products. We will find insights from it by data analysis.

1. How could these trends apply to Bellabeat customers?

Bellabeat company manufactures tech products for women to inform them about their health habit daily. It does this by collecting data about activity, stress, sleep, and reproductive health. Through the insights gained from the data analysis on smart device usage data we can recommend Bellabeat on areas to focus for women’s health with design based on the insights.

1. How could these trends help influence Bellabeat marketing strategy?

Bellabeat is using variety of marketing strategy and ad campaigns as well as digital platform to market their products. The trends could help to target particular areas for the marketing.

Roadmap:

1. What is the problem you are trying to solve?

Bellabeat wants to use the insights gathered from analyzing public smart device usage data on physical activity and make marketing strategies that will help design and build products for women?

1. How can your insights drive business decisions?

The insights on daily fitness habits of 30 participants from the data can help determine the loopholes and opportunities on areas to focus on making better features and devices that could help achieve fitness goals for women.

**Limitations of the Data**:

This data does not have any information on demographics of the participants. Their age, location, or their diet which limits to form a relationship between certain activities and how they can be affected with demographics.

Some questions that cannot be answered,

1. Age group insights, which age group is staying highly active or which age group uses the fitness tracker the most.
2. Which age group performs more physical activities on what times during a day.
3. How food intake and hydration play a role to quickly build but the energy levels?
4. Which weight group expends more calories and less calories during what time of the day?
5. Which age group can sleep more?

Another limitation of the data is what kind of activity is being performed by the participant, for example like, running, walking, swimming, jogging, cycling, or any physical activity like sports, etc.

The sample size for this data is only 30 users, which is the least required to form any statistical foundation. In order to provide more accurate findings that can related to a large population the sample size should be increased to at least 100 users.

This data is taken from users using only devices from one brand.

It is difficult to generalize the findings since we don’t know the what age group these users belong to. Whether they are students, adults with busy lifestyle or older people.

**Prepare**

The founder has suggested to use Fitbit fitness tracker data to gain insights on thirty users who consented to provide data about their daily habits of using the fitbit smart device. It will have information on heart rate, sleep activity and physical activity, daily steps and other such habits.

Guiding questions

● Where is your data stored?

The data is stored in Kaggle data repository licensed under public domain.

● How is the data organized? Is it in long or wide format?

Available in both long and wide format

● Are there issues with bias or credibility in this data? Does your data ROCCC? No

● How are you addressing licensing, privacy, security, and accessibility? Its public data

● How did you verify the data’s integrity?

● How does it help you answer your question? Very well

● Are there any problems with the data? Not yet

**Key tasks:**

1. Download data and store it appropriately.

2. Identify how it’s organized.

3. Sort and filter the data.

4. Determine the credibility of the data.

**What we need**:

Data on activity, sleep, stress, and reproductive health

**What current data is providing**:

* Physical activity
* Sleep
* Steps
* Heart rate
* MET

**About the Dataset:**

1. Daily activity merged contains Intensity minutes, steps, calories, and distance
2. Hourly merged contains Calories, Intensity, Step total
3. Minutes merged contains Calories, Intensities, Steps, MET,

Level of stress is associated with HRV values. The HRV can be determined by Root mean square of the all the difference between each interval value of heart beat for an period of 5 min which is nominal for this measurement.

**[1].** Shaffer, F., & Ginsberg, J. P. (2017, September 28). An Overview of Heart Rate Variability Metrics and Norms. In *PubMed Central (PMC)*. https://doi.org/10.3389/fpubh.2017.00258

**[2].** Kim, H. G., Cheon, E. J., Bai, D. S., Lee, Y. H., & Koo, B. H. (2018, February 28). Stress and Heart Rate Variability: A Meta-Analysis and Review of the Literature. In *PubMed Central (PMC)*. <https://doi.org/10.30773/pi.2017.08.17>

**[3]**. Labs, D. (2022, August 15). Heart Rate Variability | The Ultimate Guide to HRV | WHOOP. In *WHOOP*. https://www.whoop.com/thelocker/heart-rate-variability-hrv/

**Process**:

Let’s perform data cleaning on few datasets and check for the following:

* Data consistency
* Data accuracy
* Data correctness

The dailyActivity\_merged.csv file is merged with Calories, intensities, and steps data.

However we will merge hourly calories, intensities and hourly steps data into one file as the number of rows and id’s are exactly the same.

This task we will perform in BigQuery.

Graphical user interface, text, application

Description automatically generated

Now we will focus on checking the data quality on Google Sheets for daily activity and hourly activity data.

**Analyze**:

**Breakdown on how we will analyze the data:**

Business context, Business requirement: Insights on habits of using fitness devices

The data given to us is for month use of fitness device by 30 Females.

So, we will start by breaking down the habits on daily usage basis.

**Daily Habits Insights**

**Sleep:**

1. Average amount of time spent on sleep, on very high intense activities, moderate activities, and light activities.
2. What amount of people do not track sleep?
3. How many percentages of users track sleep?
4. What % of time the device was not used to record any activity?
5. What % of data does not have sleep recorded?
6. What is the average sleep time for all users whose sleep was recorded?
7. Amount of time lying in the bed distributed across intervals of Less than 30 mins, between 30 to 60 mins and more than 60 mins.
8. At what time do users are going to bed based on total Intensity level?

**Steps:**

1. Average number of steps taken daily?
2. What percentage of time is spent on different intensity activities and what are their average values?
3. Segregate the number of users based on categories like highly active, moderately active and light active.