

Case Study- Bellabeats

- **Ask**

- Analyze one bellabeat product and answer the following questions:
 - What are some trends in smart device usage?
 - How many calories are burned in each activity type? Are there any correlations to certain activity types to the number of calories burned?
 - Which days of the week are people most active? Which are they least active?
 - Which hours of the day are the most active? Which are they least active?
 - How could these trends apply to Bellabeats customers?
 - How could these trends help influence Bellabeat marketing strategy
- Deliverables - Produce a report with the following:
 - Clear summary of the business task
 - Description of all data sources used
 - Documentation of any cleaning or data manipulation
 - Summary of the analysis
 - Supporting visualizations and key findings
 - High level content recommendations
- Business Task:
 - Discover trends about one of Bellabeats smart devices, then use these trends to recommend marketing strategies.
 - To do this I will follow the process above.
 - Stakeholders;
 - Urška Sršen -Non-technical, Co-founder, Chief creative Officer
 - Sando Mur - Technical, executive and co-founder
 - Scope of Work:
 - The date range used will be that that is provided in the dataset.
 - The user data analyzed will be that that is provided in the dataset.
 - Any other data that is not provided will be analyzed.
 - Since this is a fictional situation, any questions that arise will be answered with the best of my intuition since I cannot ask the stakeholders any questions.

- **Prepare**

- Dataset: **FitBit Fitness Tracker Data**
 - Author: MÖBIUS
 - Link: <https://www.kaggle.com/datasets/arashnic/fitbit>
- The zip file contains 18 csv files. I will be excluding the minute data as I feel it presents too much data that can be tracked hourly. I will also exclude the sleep day and weight info files since I feel there is enough data in the other files to draw conclusions from.

- dailyActivity_merged - contains long data about activity of users, including calories, intensities, and steps
 - dailyCalories_merged - contains long data with calories burned per user
 - dailyIntensities_merged - contains long data about user activity levels
 - dailySteps_merged - contains long data about user number of steps per day
 - hourlyCalories - long data similar to above but hourly instead of daily
 - hourlyIntensities - long data similar to above but hourly instead of daily
 - hourlySteps - long data similar to above but hourly instead of daily
- I will clean the data and store the clean data in Google Spreadsheets. Original copies of the csv files will be kept for reference.
- My Plan will be to use the files mentioned above. I will use only the dailyActivity csv file if I find that it completely represents the other 3 files. If not all must be used.
- **Process**
 - Data Cleanup
 - The data cleanup will be done using Google Sheets but can also be done using Excel.
 - Here are the steps I took to clean the data:
 - 1. Freeze each header row
 - 2. Remove Duplicates
 - 3. Trim whitespace
 - 4. Make sure each column is in the correct format
 - Make sure number columns are in Number, Dates are in Date, Datetime in Datetime, etc.
 - 5. Add in a weekday column using the weekday function in sheets. Then map each weekday number to a string representing the weekday. For example 1 -> Sunday
 - 6. Split the datetime into date and time. Then use the hour function to extract the hour from the time.
 - 7. Create an hour number to represent the hour. This is done for easier reading. For example 1:00:00 is 1 or 1AM.
 - Looking closer at the daily csv files:
 - Next I look closely at the daily csv files. The plan will be to check if the dailyActivity csv file is truly merged correctly and represents the other three completely. If this is the case the daily activity csv file can be used instead of the other 3. This will be done in google sheets but can also be done in Excel.
 - There are 941 rows in each, 940 if we do not include the headers.
 - Using COUNTUNIQUE() we see that there are 33 unique ids in each of these datasets.
 - I then put each of these unique ids into a new sheet and used the countif() function to count the occurrences of each id in each of the daily files.

- From this it can be seen that the occurrences and ids are identical in each of these files.
 - The same type of task can be done on various column values. When we see that the count of each value is identical and that each file contains the same information, we can conclude that the daily activity file contains all of the data from the other daily files.
 - This is a long task that I do fully show in the sheets file. I would be happy to explain the thought process in more detail and how the task could also be done in SQL.
- Note about this task in SQL:
 - An easier way to check if each file contains identical information would be to use SQL.
 - To do this you could use Inner joins to merge each file with the daily activity file.
 - You could also do this through various select statements to check if there are any values that are in one file that are not in the daily activity file.
 - It is important to note that I can present the queries, however I choose to do this task in sheets in order to keep the work together. This task involves more work to check this in sheets, however I believe it is worth having the information in sheets along with the data.
- Results of Process Phase and Data Cleaning
 - It was found that the data in all daily csv files is identical and fully represented in the daily activity file. Therefore the daily activity file will be used in place of the others.
 - All of the actions taken in the data cleaning have been documented, and the data is ready to be analyzed.
- **Analyze**
 - I will be using R for analysis and visualization. Therefore I will export the cleaned csv file and import it into R Studio.
 - For a breakdown of the analysis, please see the R markdown files names:
 - CaseStudy_BellaBeats_Fitness.html
 - CaseStudy_BellaBeats_Fitness.pdf
 - Note: The two files contain the full analysis, I have created both an html and pdf file depending on preferences.
- **Share**
 - The following would be shared with stakeholders:
 - One of the markdown files (depending on preference):
 - CaseStudy_BellaBeats_Fitness.html
 - CaseStudy_BellaBeats_Fitness.pdf
 - These contain the full analysis as well as the visualizations produced. They also contain recommendations

- **Act**

- Here are the final recommendations based on the analysis.
 - BellaBeats should encourage users to participate in very active activities if they are looking to burn more calories.
 - The marketing team should focus their campaigns more heavily on Saturday and Tuesdays and less heavily on Sundays and Thursdays
 - The marketing team should focus marketing during the 17th - 19th hour or (5:00PM - 7:00Pm) and 12th - 14th hour (12:00PM - 2:00PM) as this is when users are most active.