Case Study Road Map

Ask

Guiding Questions:

What is the problem you are trying to solve?

- (Questions from the task):
    - What are some trends in smart device usage?
    - How could these trends apply to Bellabeat customers?
    - How could these trends help influence Bellabeat marketing strategy?
- How are people already using their non-Bellabeat smart devices, and how can these insights influence Bellabeat marketing strategy?
- What trends are there in smart device usage, and how do these trends apply to Bellabeat customers?

How can your insights drive business decisions?

- Insights into data about how people use non-Bellabeat smart devices can help Bellabeat to focus their marketing strategy on the ways their devices meet the most popular functions of other smart devices.
- It might even lead to further investigation about whether Bellabeat should change their products or add new devices in order to keep up with the most in-demand functions for smart devices today.

Key Tasks:

Identify the business task:

- The business task is to explore current data on how people use their smart devices in order to compare these findings to Bellabeat products. Specifically, Bellabeat wants to compare these findings to one of their own products in order to discover ways that Bellabeat might be able to improve marketing strategy for this product. This will determine the best strategy for Bellabeat marketing in recruiting and maintaining their own customers.

Consider the stakeholders:

- Urška Sršen: Bellabeat's cofounder and Chief Creative Officer
- Sando Mur: Mathematician and Bellabeat's cofounder; key member of the Bellabeat executive team
- Bellabeat marketing analytics team: A team of data analysts responsible for collecting, analyzing, and reporting data that helps guide Bellabeat's marketing strategy. You joined this team six months ago.

Deliverable:

✅ A clear statement of the business task

Prepare

Guiding Questions:

Where is your data stored?

- The linked data is from this Kaggle Notebook, by the user Mobius, but the data originates elsewhere.
- According to the metadata, the original data came from Zenodo. Source: https://zenodo.org/record/53894#.YIBWMOhKiUl

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

- The data is organized in both wide and long formats.
- Note: I will probably want to make the ID the primary key for the data in order to merge and transform data sets into wide data as needed.

Are there issues with bias or credibility in this data? Does your data ROCCC?

- Reliable - This data is somewhat reliable because it is accurate, but I do not believe it to be complete or free from bias. See more detailed notes below:
    - Accurate - I assume this data is accurate because it is presented with the technology from the smart watches. It is objective, rather than subjective, so I believe this data is accurate
    - Incomplete - I do not believe this data is complete in the sense that it only surveys 30 Fitbit users. A larger sample size would be better for having more statistically accurate data. Additionally, I definitely see missing values in each of the spreadsheets, due to missing data from the participants.
    - Bias - I believe this data is slightly biased since it only includes information from Fitbit users. I also don't know any demographic information about the participants, so I don't know if there is a balanced focus group in age, gender, race, and more. It would be less biased if it included information about the sample group, as well as information from other brands of smart watch users. Mostly, there just needs to be a larger sample size.
- Original - This data appears to be original. Details about this data:
    - Here is a description of the data from the source website of Zenodo: "These datasets were generated by respondents to a distributed survey via Amazon Mechanical Turk between 03.12.2016-05.12.2016. Thirty eligible Fitbit users consented to the submission of personal tracker data, including minute-level output for physical activity, heart rate, and sleep monitoring. Individual reports

can be parsed by export session ID (column A) or timestamp (column B). Variation between output represents use of different types of Fitbit trackers and individual tracking behaviors / preferences."
    - Therefore, this data is complete because we know the source, Amazon Mechanical Turk.
- Comprehensive - I do not believe this dataset is comprehensive. Here is why:
    - The dataset contains some important information needed to answer the question or find the solution. However, I don't think the dataset is comprehensive because we just need more data from more smart watch users.
    - The dataset may contain human error, if it is considered human error that the smart watches are not charged, etc. Data about this might actually be helpful, because charge time and length between charge time could be important metrics for marketing to consider as well. However, these metrics are not listed in this dataset.
- Current - This dataset is not exactly current.
    - As seen above, this dataset contains records of 30 Fitbit users from March to May of 2016. As it is 2021, and smart watches have improved in the last 5 years. Additionally, a quick search into the background of Bellabeat shows that the company started in 2014. This data would have been more helpful in the early years of this company, since that is when it was compiled. Now, this company needs newer data, preferably no earlier than 2021. Considering the recent Covid-19 pandemic, smart watch users have probably changed the way they use their watches (i.e. home workouts vs. gym workouts). This means that new metrics or patterns might be needed for gaining the most current insights about smart watch users and their preferences.
- Cited - This dataset is cited. See the links for Kaggle and Zenodo above.
- Conclusion - This data is a good start, but it does not ROCCC. It needs to be analyzed with other larger and more comprehensive and current datasets.

How are you addressing licensing, privacy, security, and accessibility?

- Licensing - The license for this dataset is "Creative Commons Attribution 4.0 International." Below, I will include the terms for sharing this dataset under this license, as well as how I will comply with those terms:
    - "If You Share the Licensed Material (including in modified form), You must
        - Retain the following if it is supplied by the Licensor with the Licensed Material:
            - identification of the creator(s) of the Licensed Material and any others designated to receive attribution, in any reasonable manner requested by the Licensor (including by pseudonym if designated);
            - a copyright notice;
            - a notice that refers to this Public License;
            - a notice that refers to the disclaimer of warranties;
            - a URI or hyperlink to the Licensed Material to the extent reasonably practicable;

        - indicate if You modified the Licensed Material and retain an indication of any previous modifications; and
        - indicate the Licensed Material is licensed under this Public License, and include the text of, or the URI or hyperlink to, this Public License."
    - I will comply with the above quoted text from the licensing terms by using this citation in my report, that is directly from Zenodo:
        - Furberg, R., Brinton, J., Keating, M., & Ortiz, A. (2016). Crowd-sourced Fitbit datasets 03.12.2016-05.12.2016 [Data set]. Zenodo. http://doi.org/10.5281/zenodo.53894
- Privacy - The privacy of the Fitbit users is already secured through the use of IDs in the dataset rather than names. There is no other identifying information about the smart watch users.
- Security - I will be the only one accessing the data - I do not currently have a team to work with on this dataset. I will only share the chunks of code and visualizations I work through that are necessary to the understanding of the project in a Markdown file.
- Accessibility - My conclusions will be accessible through my final presentation in a Slide Deck format, as well as a Markdown file. I will archive the data after completing my analysis.

How did you verify the data's integrity?

- I verified the data's integrity by locating the source of information, identifying the background for how and when the data was collected, and discovering what kind and how much data was in the dataset.

How does it help you answer your question?

- All of this research into the data helps me answer the business task by identifying the strengths and weaknesses to my dataset. Ultimately, it has led to me realizing that I will also need to compare this dataset to a larger and more current dataset about smart watch users in order to confirm that my insights are conclusive.

Are there any problems with the data?

- Here is a summary of the ROCCC question above, which outlines the major problems with the data:
    - The sample size of 31 Fitbit users is too small, and may not be inclusive of all ages, genders, or ethnicities.
    - The data only includes information about Fitbit users, so this excludes all other smart watch users who have other brands of smart watches.
    - The data is not current, as it was collected in 2016.

Key Tasks:

✅ Download data and store it appropriately.
    ✅ Download original Fitbit dataset
    ✅ Upload to Google drive

        - ✅ Create Google Sheets files for each to preview, if possible
        - ✅ Rename files for consistency
        - ✅ Organize into separate folders: CSV and Sheets
        - ✅ Upload CSVs into MySQL database (Goorme IDE) - Not helpful!
        - ✅ Upload CSVs into BigQuery - Also not helpful!
        - ✅ Upload CSVs into RStudio

- ✅ Identify how it's organized.

- ✅ Sort and filter the data.
    - ✅ Review the tables through "View" in RStudio
    - ✅ Determine tables most useful for analysis

- ✅ Determine the credibility of the data.

Deliverable:

- ✅ A description of all data sources used

Process

Guiding Questions

What tools are you choosing and why?

- I have chosen to use R to process, analyze, and start to visualize the data because the data was too large for Spreadsheets
- I might use spreadsheets for some of the smaller sets of data. I will also export some of my smaller subsets of data from R into spreadsheets in order to make visualizations easier.

Have you ensured your data's integrity?

- I have ensured the data's integrity to the best of my ability. I will make copies of the data and use temporary tables whenever possible in order not to disrupt the original data.

What steps have you taken to ensure that your data is clean?

- See the link to changelog below.

How can you verify that your data is clean and ready to analyze?

- I have done everything that I have learned to verify that my data is cleaned and prepped for the next phase of the process.

Have you documented your cleaning process so you can review and share those results?

- This is the form I fill out for my continuous changelog.
- Here, I can download the responses as a spreadsheet to see all the changes.

Key Tasks:

✅ Check the data for errors.

✅ Choose your tools.

✅ Transform the data so you can work with it effectively.

✅ Document the cleaning process.

Deliverable:

✅ Documentation of any cleaning or manipulation of data

Analyze:

Guiding questions

How should you organize your data to perform analysis on it?

- First, I will need to identify the most important data sets that I will need to use for analysis. Each data set has good information, but not all of it will be relevant or helpful for this analysis. I will look at the data to figure out how many users are actually using each kind of data points collected, and then decide which tables will be the most useful for further analysis. Basically, I will query each of the tables for basic info and calculations, create some exploratory graphs, and then choose which tables to use for deeper analysis.

Has your data been properly formatted?

- Yes, my data has been cleaned and processed through the process phase. If I need to fix any issues, or sort and filter more, I will do so as problems arise in this analyze phase.

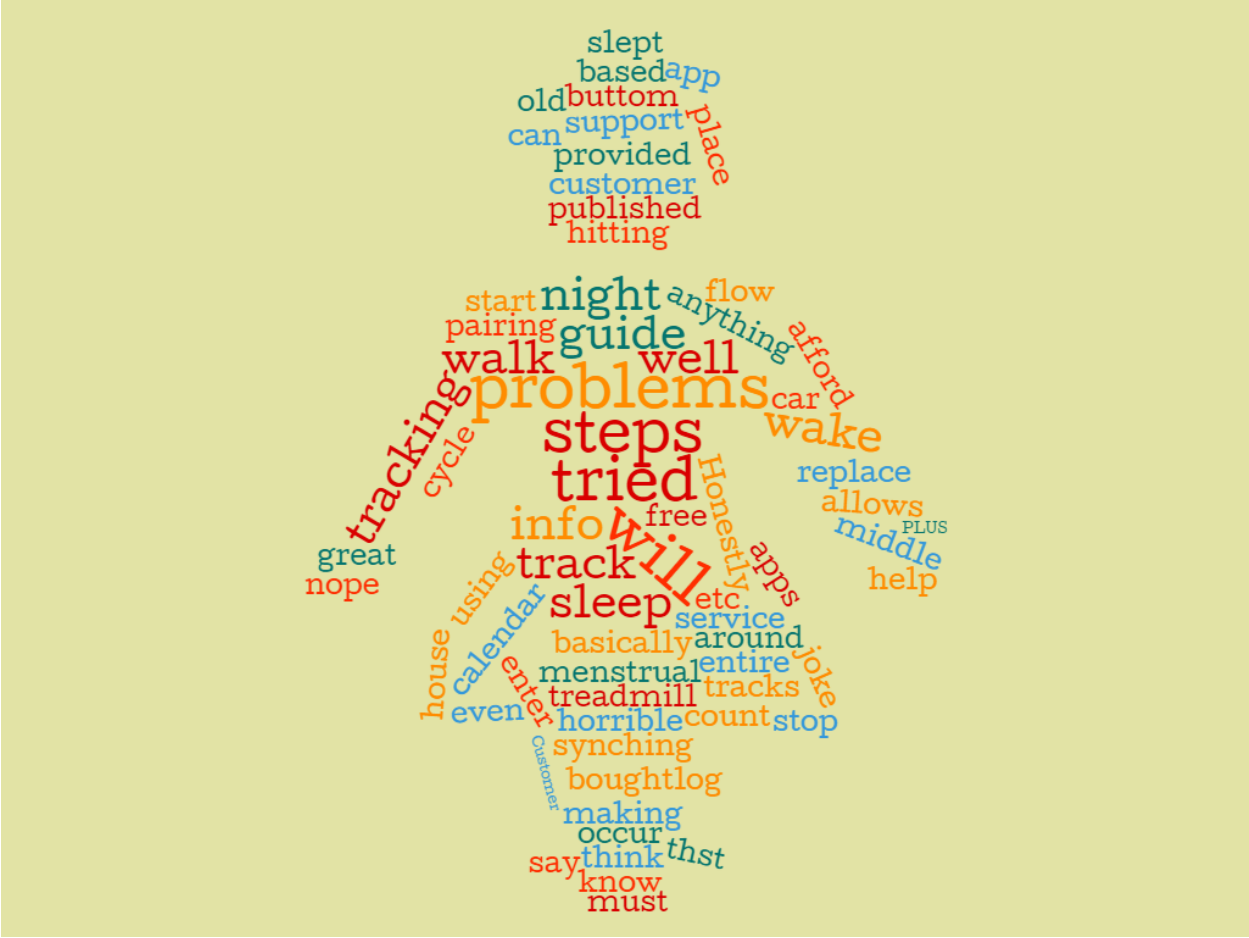
What surprises did you discover in the data?

- I was surprised to see that there are fewer distinct users for both the “sleep” and “weight” datasets. My assumptions here are that fewer users are wearing their watches at night due to the need for charging, and they are likely using other apps or devices like smart scales to track weight.

- I was also surprised to find that there are some outliers that affect the mean and median of the data on sleep records.
- Finally, I was surprised that there wasn't any data at all to track the exact times for falling asleep and waking up - this would be important data for further research.

What trends or relationships did you find in the data?

- I found, unsurprisingly, that as the number of daily steps increases, the number of calories burned also increases.
- I found that only 48% of the participants used the sleep tracker in their watches in order to track sleep records regularly.
- Additionally, I researched Bellabeat products in order to get a better idea of the performance of what they have to offer. I found:
    - Bellabeat products use batteries, track cycles, and offer fashionable choices. All of these might be good marketing to show how Bellabeat is different than its competitors, because…
    - Online reviews of Bellabeat on Amazon give the product less than 3.5 stars. Many of the complaints are about accuracy. See the word cloud below. I copied and pasted all of the 2 star and below reviews to visualize what the biggest issues were.



How will these insights help answer your business questions?

- Of the 25% that did not regularly track sleep records, the patterns of tracking indicate that charging might be happening at night. There may be a need for longer time between charges, or in the case of Bellabeat's products, the use of batteries instead.
- These insights suggest that Bellabeat might want to focus marketing on step and calorie counters, as well as the features they have that other products don't have, such as batteries and cycle tracking.

Key Tasks

✅ Aggregate your data so it's useful and accessible.

✅ Organize and format your data.

✅ Perform calculations.

✅ Identify trends and relationships.

Deliverable

✅ A summary of your analysis

- See this Kaggle of my analysis in R. I have included notes about the visualizations in the code.
- Next, I will create better visualizations in Tableau.

Share

Guiding questions

Were you able to answer the business questions?

- Yes, but I feel that my answer is limited because I would like to have data about Bellabeat products from Bellabeat directly, instead of through reviews on Amazon. Additionally, I think a larger dataset would help to confirm or alter my conclusions.

What story does your data tell?

- See my presentation

How do your findings relate to your original question?

- I compared my insights about how Fitbit users actually use their devices to a Bellabeat product, *Time*, and offered suggestions for marketing strategy about what to focus on when developing their next marketing campaign.

Who is your audience? What is the best way to communicate with them?

- The audience includes the following:
    - Urška Sršen: Bellabeat's cofounder and Chief Creative Officer
    - Sando Mur: Mathematician and Bellabeat's cofounder; key member of the Bellabeat executive team
    - Bellabeat marketing analytics team: A team of data analysts responsible for collecting, analyzing, and reporting data that helps guide Bellabeat's marketing strategy
- The best way to communicate with them will be through compelling and informative visualizations. I will present these visualizations through a slide deck.

Can data visualization help you share your findings?

- Yes, see the link to the raw visualizations on Tableau.

Is your presentation accessible to your audience?

- Yes, see the link to my presentation.

Key tasks

✅ Determine the best way to share your findings.

✅ Create effective data visualizations.

✅ Present your findings.

✅ Ensure your work is accessible.

Deliverable

✅ Supporting visualizations and key findings
    - Visualizations were created through Tableau
    - Presentation created through Google Slides (or Nearpod)

Act

Guiding Questions

What is your final conclusion based on your analysis?

- First, Bellabeat should review the accuracy of the their products in tracking step counts to see if there is room for improvement.
- Next, marketing strategy should focus on what Time does well:
    - It uses a battery, so there is no need for charging overnight. This means that sleep tracking can be more useful.
    - Bellabeat includes menstrual tracking through the connected app, which is a feature that is not included in many other smart watches.
- After issues with accuracy are addressed, marketing should add another focus to their strategy: Emphasize the relationship between taking more steps and burning more calories. By using Time, consumers can improve their health by becoming more aware of their own health habits.

How could your team and business apply your insights?

- See above

What next steps would you or your stakeholders take based on your findings?

- See above

Is there additional data you could use to expand on your findings?

- Yes, I would definitely want to see more data about other smart watch users in order to have a larger sample size to work with. Additionally, I would like a more recent data set, as this Fitbit data set was from 2016.

Key Tasks

✅ Create your portfolio.

✅ Add your case study.

✅ Practice presenting your case study to a friend or family member.

Deliverable

✅ Your top high-level insights based on your analysis