Big Picture

---

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?

Deliverables

1. Clear summary of the business task
2. A description of sources used
3. Documentation of any cleaning or manipulation of data
4. A summary of the analysis
5. Supporting visualizations and findings
6. Top high-level content recommendations based on the analysis

---

Ask

Guiding Questions

- What is the problem you are trying to solve?
- How can your insights drive business decisions?

Key Tasks

1. Identify the business task
2. Consider key stakeholders

Deliverable

- [x] ~~A clear statement of the business task~~

What is the problem you are trying to solve?

Bellabeat stakeholders are looking to refine their marketing strategy for growth. They want to know:

- Trends in smart device usage
- How the trends could apply to Bellabeat customers
- How the trends could inform Bellabeat marketing strategy

How can your insights drive business decisions?

The analysis outcomes will be used to influence the marketing strategy for the company. This is an important business decision and has a direct effect on company success.

Identify the business task

Analyze smart device usage from an external source, apply the insights derived to Bellabeat and their marketing strategy. Provide high-level recommendations to Bellabeat stakeholders regarding how they might leverage the trends in their marketing strategy.

Consider key stakeholders

Audience

- CCO/cofounder
- Mathematician/cofounder
- Fellow marketing analysts

What do they already know?

All involved should be aware of our current marketing strategy and current sales numbers.

What do they need to know?

The trends observed in external smart device usage and how those trends can be applied to Bellabeat customers and how the trends can inform marketing strategy decisions.

How can you best communicate what they need to know?

There will be a presentation. Visualizations will be important but given the level of involvement of all parties, it could get more detailed if needed.

Business Task

Derive insights from external smart device usage data and identify how they can be applied to Bellabeat. Provide recommendations regarding how these insights can inform Bellabeat

marketing strategy.

Prepare

Guiding Questions

- Where is your data stored?
- How is the data organized? Is it in long or wide format?
- Are there issues with bias or credibility in this data? Does your data ROCCC?
- How are you addressing licensing, privacy, security, and accessibility?
- How did you verify the data's integrity?
- How does it help you answer your question?
- Are there any problems with the data?

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

Deliverable

- [x] ~~A description of all data sources used~~

***I want to find an additional dataset as well*

Where is the data stored?

The dataset is available on Kaggle or Zenodo where it was published.

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

The data is organized into discrete categories. Each sheet has a specific list of information and most sheets have a minute, hour and day version.
It is stored in wide format.

Are there issues with bias or credibility in this data?

In terms of bias, it is hard to evaluate without knowing their eligibility criteria. Additionally, there is no information regarding participant age, gender, etc.
It is likely that it is biased towards people that are active.
The credibility is somewhat low, it came from a third party survey but the data is taken directly from the user's trackers.

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

The data is licensed for public domain.
Privacy has been handled by the dataset publisher, no personal identifying information was included in the dataset.
Due to the wide availability of the data and lack of PII, there is no need to take additional security steps.
The data is widely accessible.

How did you verify the data's integrity?

- Looked at number of users and data collection period - looking for users who missed days
- Look for N/A's
- Checked formatting consistency (ie. dates not in date format) - checked the datatypes of each column and reformatted the dates

How does it help answer your question?

The activity dataset tells me about the users activity during the survey period which will give me information about how they are using their smart device. Also, the weight log and sleep log will give me some additional information as to the frequency of use of these additional services.

Are there any problems with the data?

It is a very small subset of people. Also since it was voluntarily given, there are many users without records for several days of the survey period.
Some issue with the clarity surrounding "TotalDistance", "TrackerDistance" and "LoggedActivitiesDistance". By virtue of its name, total distance seems like it should be tracker distance + logged activities distance but its not. In every single instance of logged activities distance being greater than 0, the total distance is not a sum of the other distance columns. However, there are also several instances where total distance is not equal to tracker distance and these coincide with records having logged activities distances greater than 0. I am not quite sure what to make of these discrepancies and there is not enough documentation to gather insights as to whether these matter or not.
I feel that another dataset would provide more value.

Additional Data

I found a 2017 study about user experience with wearable activity trackers.
I will take the data presented in the paper to augment the case study.
Since I only have access to the findings presented in the paper and not the raw data, it does not require analysis but may provide some insights that could be valuable to Bellabeat.

The paper only involved participants in Australia and heavily skewed toward female participants (~71%). As Bellabeats target audience is women this should not be a concern.
The paper was published in BioMed Central Public Health and is Open Access.

Business Task

There are 2 sets of data that will be used in this case study.
The first, is the FitBit Fitness Tracker Data collected by Furberg, et al. via online survey between April 12, 2016 and May 12, 2016. The data consists of activity data collect by the users' FitBit's, records of users' weights recorded in the app during the survey period and sleep logs recorded by the users' FitBit's during the survey period.
The second, is data collected by Maher, et al. as part of a study regarding users' experience of wearable activity trackers. The study was conducted in Australia and data was collected via an online survey. The survey was not brand specific and included 6 named brands and a catch-all "other" category. I will use tables and data from the paper to supplement my analysis.

Process

Guiding questions

- What tools are you choosing and why?
- Have you ensured your data's integrity?
- What steps have you taken to ensure that your data is clean?
- How can you verify that your data is clean and ready to analyze?
- Have you documented your cleaning process so you can review and share those results?

Key tasks

1. Check the data for errors.
2. Choose your tools.
3. Transform the data so you can work with it effectively.

4. Document the cleaning process.

Deliverable

- [x] ~~Documentation of any cleaning or manipulation of data~~

What tools are you choosing and why?

I will be using a Kaggle Notebook with Python. It's what I'm most comfortable with and I think I'll end up doing multiple case studies anyways so I'll do this one this way and try SQL/R another time.

Have you ensured your data's integrity?

I have.

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

I have checked the column value formatting and looked for null values. I am also aware of "incomplete" data such as people missing days throughout the survey period.

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

Yes.

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

Yes. They are documented in my Notebook.

Analyze

Guiding questions

- How should you organize your data to perform analysis on it?
- Has your data been properly formatted?
- What surprises did you discover in the data?
- What trends or relationships did you find in the data?
- How will these insights help answer your business questions?

Key tasks

1. Aggregate your data so it's useful and accessible.
2. Organize and format your data.
3. Perform calculations.
4. Identify trends and relationships.

Deliverable

- [x] ~~A summary of your analysis~~

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

Currently, the data is organized into 3 dataframes. This should be fine for the analysis.
I will perform any necessary manipulations like grouping and pivoting as I go.

Has your data been properly formatted?

Yes. The dates have been reformatted to datetime and for the weight log, I removed the time so avoid multiples of a single date.

What surprises did you discover in the data?

I hadn't expected any duplicates in the datasets but found 3 in the sleep log which I almost missed. I was also surprised to find that the total distance was not the sum of the tracker distance and logged activities distance.
Outside of cleaning surprises, I was not expecting Tuesdays to be one of the active days of the week for users. Also, the users were more sedentary that I expected. 24% of users walked less than 5,000 steps a day on average and 48.5% of them were not getting 30 minutes of exercise a day on average. 39.4% were not burning more than 2,000 calories a day.

What trends or relationships did you find in the data?

- Only 21% of users have an average total steps of 10,000 or more
- The percent of users between 50-75% of the goal and users reaching more than 75% of the goal are the same, 27%
- The majority of users, 57.6%, met the the 10,000 step goal 25% of the time or less. That is less than 8 days out of 31
- Only 12% reached the goal 75% of the time or more
- The majority of users, 51.5%, have an average exercise minutes of 30 minutes or more
- 39.4% of users meet their 30 min daily exercise goal 25% of of the time or less
- 54.5% of users did not meet the 30 min per day, 5 times per week goal for any of the 5 weeks

- 60.6% of users are burning more than 2000 calories on average
- The vast majority of users' total distance is made up of light active distance, 93.9%
- Tuesdays and Saturdays are the most active days of the week
- Only 72% of users logged their sleep at least once during the study
- Of those 24, 45.8% logged their sleep more than 75% of the time, more than 23 times
- 92% of users were not getting 8 hours of sleep on average, however, 66.7% of them were getting 5H and 40M or more
- Only 24% of users logged their weight at least once during the study
- Of those 8, the majority (62.5%) only logged their weight once or twice
- For the 6 users that recorded their weight more than once, 4 of them lost some weight, one remained the same and one gained some weight
- The most notable was a user who lost almost 4 pounds over the course of the study, they also had the highest average steps and average calories amongst the 6 users
- There is a moderate to strong relationship between sedentary minutes and time spent asleep, more sleep means fewer sedentary minutes OR alternatively, more sedentary minutes means less sleep ie. poorer sleep quality
- Weak to no correlation between sleep and steps, or total distance
- Weak to moderate between BMI and lightly active minutes
- Strong relationship between active minutes (very and fairly active minutes) and calories burned. Also with total steps and calories burned. Both are logical
- Some weak to moderate correlation between BMI and lightly active minutes and sedentary minutes, indicating users with a higher BMI have higher lower activity/inactive minutes

How will these insights help answer your business questions?

- The activity data tells the story of the users' lifestyles, yes they are relatively active but not to the point of extremes. Most are not meeting daily step goals or activity goals. Fitness trackers aren't just for fitness fanatics, they're for regular people who maybe want to be more mindful of their habits
- The weight data shows that many people are not using this feature. That tells us that weight loss is likely not a main priority for these users
- The relationship between sleep and sedentary minutes shows that more sleep = more energy to be active or more activity = better (more) sleep. Goes both ways, as neither one is independent. Can use this whichever way will be best for the marketing strategy
- The sleep data shows that most users are not meeting sleep goals either but, the majority of users were within 30% of the goal. Combined with the relationship between sedentary minutes and sleep, there can be a push for more and better sleep and how Bellabeat products can support that goal

---

Share

Guiding questions

- Were you able to answer the business questions?
- What story does your data tell?
- How do your findings relate to your original question?
- Who is your audience? What is the best way to communicate with them?
- Can data visualization help you share your findings?
- Is your presentation accessible to your audience?

Key tasks

1. Determine the best way to share your findings.
2. Create effective data visualizations.
3. Present your findings.
4. Ensure your work is accessible.

Deliverable

- [x] ~~Supporting visualizations and key findings~~

Were you able to answer the business questions?

Yes, I identified some key findings that can be used in our marketing strategy.

What story does your data tell?

It tells us the lifestyles of the users and how activity affects their wellbeing. With the relationship between sleep and sedentary minutes, we can see that when they are not active enough it has a negative impact on their sleep.

How do your findings relate to your original question?

Knowing the lifestyles of the users is important because it allows us to better align our marketing strategy with their needs and wants. Also tells us how products like ours are used day-to-day.

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

The cofounders and fellow marketing analysts. The best way to communicate with them would likely be a visualization dashboard and a walk-through presentation. The presentation would

allow me to showcase my findings and demonstrate my thought process. The dashboard provides a clear look at all of the key findings in one place and the audience can explore it on their own after the presentation as well. This may provide additional insights after other perspectives are heard.

Can data visualization help you share your findings?

Always. In some cases numbers will do the trick but a good visualization will tell the story much faster and much better.

Is your presentation accessible to your audience?

Yes. Tooltips have been added to several areas to provide further clarification. The dashboard was made to be very approachable but the audience of mathematicians and analysts would be able to appreciate it as well.

Act

Guiding questions

- What is your final conclusion based on your analysis?
- How could your team and business apply your insights?
- What next steps would you or your stakeholders take based on your findings?
- Is there additional data you could use to expand on your findings?

Key tasks

1. Create your portfolio.
2. Add your case study.
3. Practice presenting your case study to a friend or family member.

Deliverable

- [x] ~~Your top high-level insights based on your analysis~~

What is your final conclusion based on your analysis? / How could your team and business apply your insights?

My final conclusion is that the marketing strategy should focus on how Bellabeat products can help users improve their sleep through activity tracking but also through the mindfulness

tracking, which is unique to Bellabeat.
Also, given complaints about battery and fitness tracker aesthetics, these are good points to push as well since Bellabeat Time has the advantage on both counts.

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

Generate a new marketing campaign.

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

If we could create our own dataset, either through surveys or user data collection, we could obtain more accurate and "personalized" data. Alternatively, we could find additional data from users with competing devices and find the gaps in those product's offerings.