Prepare Stage

- Where is your data stored?
    - Csv files
- How is the data organized? Is it in long or wide format?
    - Some wide (labeled wide or merged) and some long (labeled narrow)
- Are there issues with bias or credibility in this data? Does your data ROCCC?
    - This set of data does not appear biased as it is user generated data and wasn't taken from a survey. As of creating this
- How are you addressing licensing, privacy, security, and accessibility?
    - All data collected is associated to a user ID by this ID is not publicly linked to a person therefore it is virtually anonymous data. Converting the files to xlsx makes it easier to work with the data but after the fact, files will be reconverted back to csv to make more accessible to other users wanting access to the edited data.
- How did you verify the data's integrity?
    -
- How does it help you answer your question?
    -
- Are there any problems with the data?
    - There are a large amount of null values in the data which can be because of various reasons so they cannot be immediately deemed a 0 value as there are zeros in the data indicating unregistered data.

What are we looking for in the data?

- Trends in smart device usage
- How can these trends apply to Bellabeat customers
- How could these trends help influence Bellabeat marketing strategy

Trends in smart device usage:

- A large majority of people prefer to do light activity over mid to high level activities
    - Average Very active minutes: 21.16
    - Average fairly active minutes: 13.56
    - Average lightly active minutes: 192.81
    - Average sedentary minutes: 991.21
- The distances of these show the same. There is no unit indicators so it is unclear what these distances represent but given that the company is based in the United States, a fair assessment would be that the distances are recorded in miles.
    - Average very active distance: 1.50
    - Average moderately distance: 0.57
    - Average lightly active distance: 3.34

    - Average sedentary active distance: 0.0016
- A baseline tactic solely based on this would be to put more emphasis in catering to the people who don't necessarily want to feel like they have to be training for competition for the smart device to be useful but want to go for a walk around the block and have a personal goal to set out for with a user-friendly interface to reach said goal.
- When looking at the total steps taken per day from April 12th to May 9th, there is a strange instance of customers deciding the stop using their smart device throughout the month of May to count their steps.
- While a couple users recorded their weight day by day, many who even chose to record their weight only did so a few times. With this, there is no definite conclusion that can be drawn by the data alone, a couple reasons that may be used as a guide to one or more conclusions:
    - It is difficult operating the device to record the data
    - Customers aren't buying the device to monitor weight loss/gain
    - Other reasons (WIP)
- A column of data titles "LoggedActivitiesDistance" is largely zero values. This can indicate that customers would rather not log this data however, as it is a large amount of users not logging the data, there is a possibility that this section serves an unclear purpose meaning customers are unable to differentiate this input with the other inputs. The few users to inputted data (manually or was automatically inputted) were all sub five distance units.
    - A solution to this would be to consider implementing short (one sentence) descriptions of the data types. This would be useful for manual and automatic data recordings as it allows the user to fully understand what the smart device is recording, potentially why it is being recorded and various other factors.
- The smart device can also track sleep time and time you spent in bed. Looking over this data, a couple users track their sleep everyday however most people tracked less days leading to an average percent days tracked (out of 31 total days) to be 55%. From this alone it cant be concluded that marketing practices need to be put in place for users tracking their sleep since reasons for not tracking their sleep can stem from a multitude of reasons.
    - A marketing / business practice that could be implemented solely based on the numbers would be to, a small part of a commercial could show how you can be notified to activate sleep tracking or automatic sleep tracking so long as the smart device has battery life similar to apps like DuoLingo.
- Summary:
    - Light activity over heavy activity preferred
        - Units used in active distance is unclear
    - Strange lack in total steps recorded in month of May
    - LoggedActivitiesDistance has no clear explanation and data is mostly zeros
    - Sleep tracking trends not fully understood.

Changelog

Since the data itself was not as messy as it could have been, there was not much that needed to be done by way of changes to get meaningful data out of the spreadsheet. As an example, to create the pie graphs to see what type of intensity users prefer when doing activities, the only manipulation of data that needed to be done was to create a pivot table then select what data was relevant to what I was looking for. Other methods such as using R or SQL could have been used but, in this case, it seemed like it would overcomplicate an otherwise straightforward set of changes.