**GOOGLE CAPSTONE : CASE STUDY**

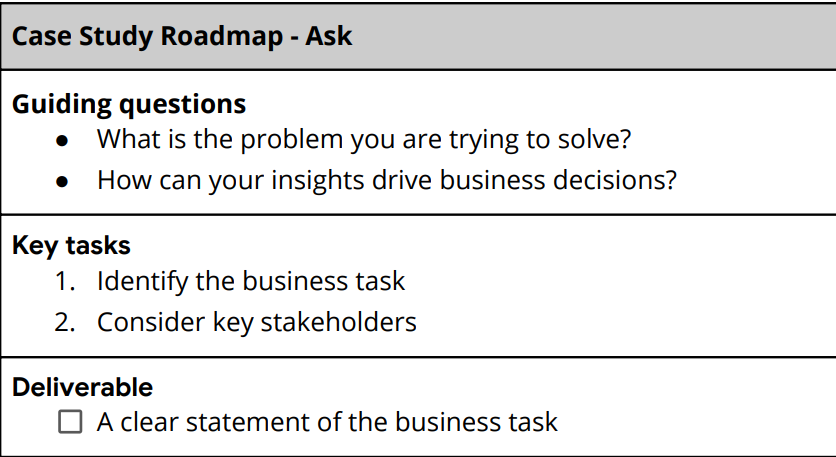
**Ask**

Sršen asks you to analyse smart device usage data in order to gain insight into how consumers use non-Bellabeat smart devices. She then wants you to select one Bellabeat product to apply these insights to in your presentation. These questions will guide your analysis:

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?

****

Guiding Question Answers:

* Need to improve marketing strategy
* It can provide recommendations and improvement steps in the right direction through data-analysis hence guiding data-driven decision making

Key Stakeholders:

* Urška Sršen: Bellabeat’s cofounder and Chief Creative Officer
* Sando Mur: Mathematician and Bellabeat’s cofounder; key member of the Bellabeat executive team

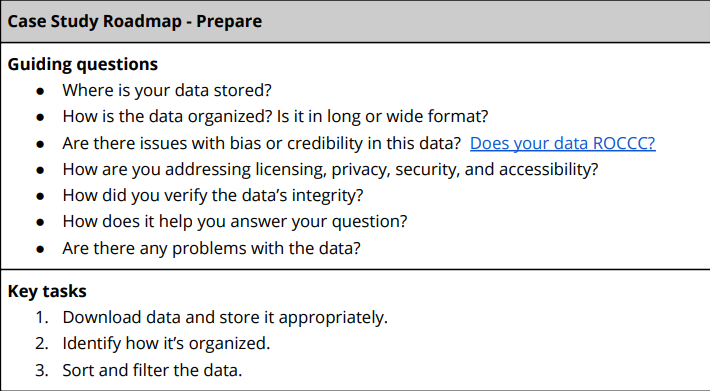
**Business Task:** Identifying potential growth opportunities and improvements in BellaBeat’s marketing strategy based on smart device usage.

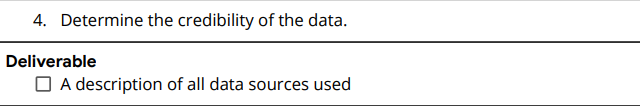
**Prepare**

Sršen encourages you to use public data that explores smart device users’ daily habits. She points you to a specific data set:

FitBit Fitness Tracker Data (CC0: Public Domain, dataset made available through Mobius): This Kaggle data set contains personal fitness tracker from thirty fitbit users. Thirty eligible Fitbit users consented to the submission of personal tracker data, including minute-level output for physical activity, heart rate, and sleep monitoring. It includes information about daily activity, steps, and heart rate that can be used to explore users’ habits.

Sršen tells you that this data set might have some limitations, and encourages you to consider adding another data to help address those limitations as you begin to work more with this data.

****

****

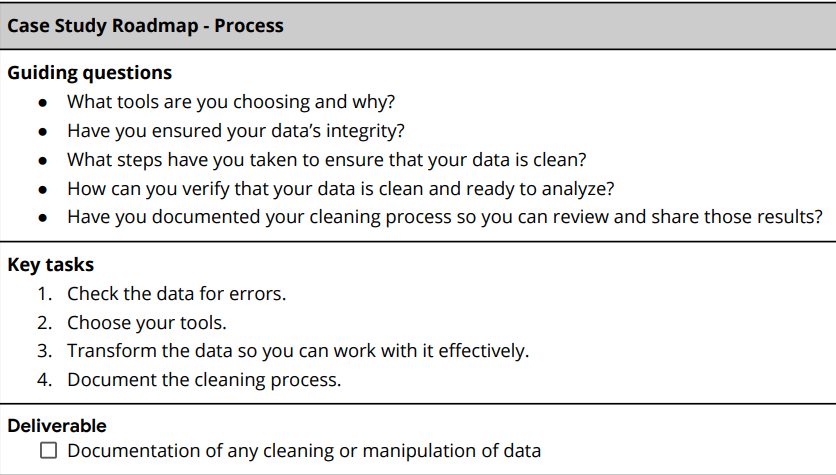
Guiding Question Answers:

* It is an open-source dataset available on Kaggle and is stored on device offline as well as Google Sheets.
* Data is organised in long format.

Data Credibility and Integrity:

Due to the limitation of size (30 users) and not having any demographic information we could encounter a sampling bias. We are not sure if the sample is representative of the population as a whole.

**Process**

****

**Prepare and Process data steps :**

Downloaded the datasets and uploaded them to google sheets for cleaning and manipulation. Then uploaded the datasets to Microsoft SQL Serever for querying.

1. DailyActivity

* First, I changed the format of data to make it uniform throughout the dataset.
* I then changed the format of number to till two decimal points for the following columns -

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| TotalDistance | TrackerDistance | LoggedActivitiesDistance | VeryActiveDistance | ModeratelyActiveDistance | LightActiveDistance | SedentaryActiveDistance |

* The data is provided for each user between 4/12/2016 – 5/12/2016.
* Deleted extra rows from the table to counter null error.

2. HourlyCalories

* I formatted the activity hour column to date time type.
* Deleted extra rows from the table to counter null error.

3. SleepDay

* Inconsistent data. Data for every user for every date between 4/12/2016 – 5/12/2016 is not present.
* Changed the format from date time to simple date.

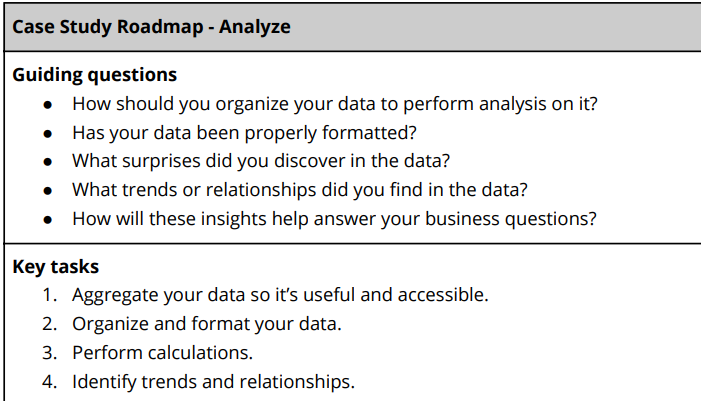
4. HourlySteps

* Changed the date time format.
* Created a pivot table to see which user has walked the most steps.
* Deleted extra rows from the table to counter null error.

**Analyze**

Deliverable:

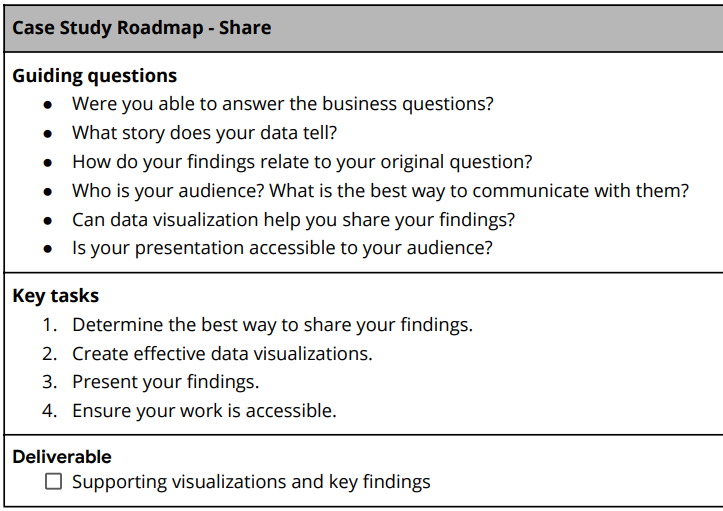
A summary of your analysis

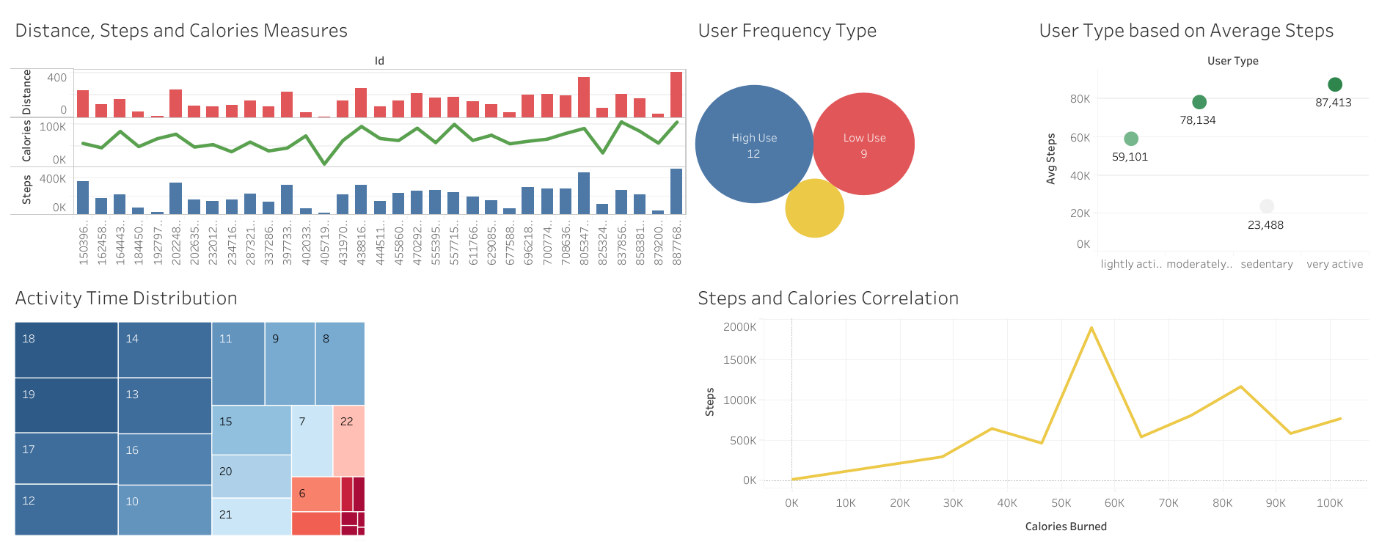
****

* Determined Tracker Accuracy by subtracting the tracker distance measured from the actual total distance. By calculations we find that tracker inaccuracy is present for two users but is highest for user id 7007744171.
* Correlated calories burned and steps taken.
* Categorized users based on how often the log into their device. – High Use, Moderate Use and Low Use.
* Categorized users based on average steps taken. – Very Active, Moderately Active, Lightly Active and Sedentary
* Filtered the data for average steps taken during the hour.
* Found out total steps, distance, and calories per user.

**Share**

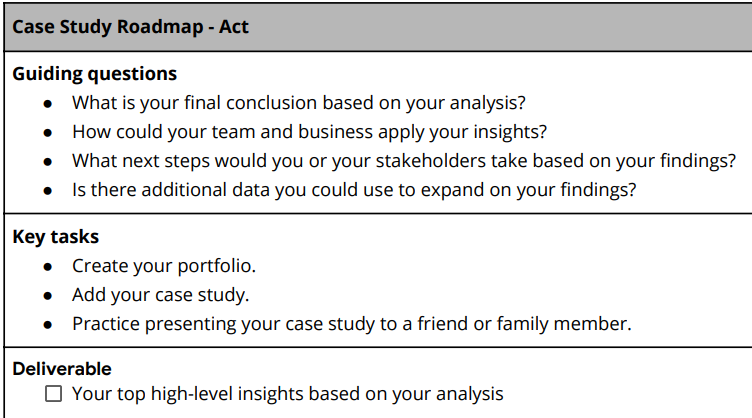
Used Tableau for visualising the insights generated from the Analyze Phase.

****





**Act**

****

Ideas and Recommendations –

* For tracker inaccuracy, BellaBeat should look into the faulty device for the user.
* By analysing I found that the time between 5 P.M. to 7 P.M. has the highest number of average steps taken by the users. The marketing team can add an additional notification during the mentioned time period as a reminder.
* Can also add recommendations and tips for low-calorie meals.
* To improve sleep time, BellaBeat can recommend reducing sedentary time.
* Average total steps per day are 7638 which a little bit less for having health benefits for according to the CDC research. Bellabeat can encourage people to take at least 8 000 explaining the benefits for their health.
* Reward System – Based on steps, for a sense of accomplishment and motivating users to reach their target.
* Daily push notifications.