**ASK**

What problems am I trying to solve?

* Trying to reveal opportunities for growth in Bellabeat marketing strategy by analyzing consumer data of non Bellabeat smart devices and applying these insights of how people are already using their smart devices to one of Bellabeats products.

How can your insights drive business decisions?

* The insight gained from the analysis of non Bellabeat smart device will be applied to one of Bellabeats products with trend reccommendations that could inform Bellabeat Marketing strategy in becoming a larger player in smart device market.

Key Business Task?

* Apply insights gained from smart device data to guide marketing strategy for the company.

Key stakeholders?

* Urska Srsen, Cofounder and CEO
* Sando Mur, Mathematician and Cofounder
* Bellabeat Marketing analytics team.

**PREPARE**

Where is data stored?

* On my pc

How is data organized?

* Data is organized in the wide format

Does data ROCCC?

* There are no issues with bias or credibility of data. Data ROCCCs

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

* The data is rated 10.00 on usability on Kaggle. It is 100% complete, credible and compatible.

How did you verify the data’s integrity?

* Kaggle highlighted that the data was submitted under public domain and all rights have been waived worldwide under copyright law.
* Kaggle also noted that data is updated yearly.

How does it help you answer your question?

* It provides the detailed breakdown of fitbit usage by individuals in a span of 30 days.

Are there any problems with the data?

* Yes. The data description detailed 30 individuals whose data usages were recorded but I see 33 individuals.
* The heart rate data is too large for excel or google sheets. Tried using Bigquery, still wouldn’t open. Bigquery complained about stamps. R opened but didn’t display all data.

How is the data organized?

* Some of the data are organized in the wide format while some are organized in the long format.

**PROCESS**

What tools are you choosing and why?

* I am using Excel and R for data manipulation because some of the data are small data and the large data can’t be opened by Bigquery.

Have you ensured your data’s integrity?

* Yes, I have.

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

* I have checked for duplicate data, whitespaces and null values.
* I have also checked for data errors and if data matches.

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

* I now know about all the quirks of the data. I haven’t made any deletion though because I don’t know what data to delete but all extraneous variables have been noted.

**ANALYZE**

* **For the daily activity merged data, I first listed out the number of unique ids in the first column using =UNIQUE(A2:A941)**
* I then counted the number of unique ids listed out in the R column with =COUNTA(R2:R34) which detailed 33 ids unlike the 30 ids stated in the metadata.
* I then checked for duplicate data and white spaces with the data tab and data cleanup options.
* I then compared total distance covered to total tracker distance to see if equal with =D2=E2. They weren’t all equal. Out of 940 observations, 925 were equal and 15 wasn’t.
* I also merged ids with each number of days.
* I also counted the number of days in a month observed for each id with =COUNTIF(A2:A,A2).
* I then found the day maximum calories was burned by each id with =VLOOKUP(MAX(O2:O32),O2:P1000,2,FALSE)
* I also found the exact number of calories burnty on that day with =VLOOKUP(MAX(O2:O32),O2:P1000,1,FALSE)
* I also found the day maximun steps were taken by each id with =VLOOKUP(MAX(C2:C32),C2:P1000,14,FALSE)
* I also found the exact number of steps taken on that day with =VLOOKUP(MAX(C2:C32),C2:P1000,1,FALSE)
* I also compared total distance to tracker distance with =D2=E2
* I also counted how many had equal tracker and total distances with =COUNTIF(X2:X941,TRUE)
* I also counted how many had unequal tracker and total distances with =COUNTIF(X2:X941,FALSE)
* **For the daily intensity merged data, I first listed out the number of unique ids in the first column using =UNIQUE(A2:A941)**
* I then counted the number of unique ids listed out in the R column with =COUNTA(R2:R34) which detailed 33 ids unlike the 30 ids stated in the metadata.
* I then checked for duplicate data and white spaces with the data tab and data cleanup options.
* I also counted the number of days in a month observed for each id with =COUNTIF(A2:A,A2).
* I then created a column for minutes in a day with =60\*24
* I then created the percentage sedentary minutes per day column for each id with =C2/O2.
* I then created the percentage sedentary minutes per month column with =AVERAGE(P2:P32).
* I then did the same for light active minutes per day for each id with =D2/O2
* I also created the percentage light active minutes per month for each id with =AVERAGE(R2:R32)
* I then created the daily minutes minus daily active minutes column with =1440-(C2+D2+E2+F2) which is 1440 (minutes per day) – sedentary active minutes, light active minutes, fairly active minutes and very active minutes.
* I then did the average inactive fitbit minutes per month per user with =AVERAGE(T2:T32)
* I then converted the average inactive fitbit minutes per month per user to hours and minutes with =U9/60/24.
* I then checked if the tracker distance is greater than the total distance covered with =IF(H9>W9,"YES","NO")
* I then checked how many were greater than the total distance and how many were less with =COUNTIF(X2:X941,"NO") and =COUNTIF(X2:X941,"YES")
* I then merged the daily sleep data with the daily intensity data and counted the number of ids monitored with =UNIQUE(AC2:AC414). I then discovered only 24 ids were accounted for.
* I also found the ids with no sleep data with =FILTER(M3:M34,ISNA(MATCH(M3:M34,AH2:AH25,0)))
* I then found the id with the highest number of sleep minutes with =SUM(AF2:AF26)
* I then found percentage sleep per sedentary minutes per day for the id with the most sleep with =AF368/C820
* I then found found percentage sleep per sedentary minutes per month for the id with the most sleep with =AVERAGE(AL2,AL3,AL4,AL5,AL7,AL8,AL9,AL10,AL11,AL12,AL13,AL14,AL15,AL16,AL17,AL18,AL19,AL20,AL21,AL22,AL23,AL24,AL25,AL26,AL27,AL28,AL29,AL30,AL31) because there some somedays with sleep minutes more than sedentary minutes.
* **I then merged the daily and hourly calories data together into a sheet and then added up the hourly for each id with =SUM(G2:G25)**
* I then compared the hourly to the daily with =C2=I2
* I then applied to the same steps above to the steps data.
* **I then merged the daily and hourly steps data.**
* I then calculated my own total monthly steps taken with =SUM(G2:G25)
* I then compared my calculated monthly steps per id to the one presented by the data to see if they are equal with =C2=I2
* I then checked how many were equal with =COUNTIF(J2:J31,TRUE)
* I then also counted how many were not equal with =COUNTIF(J2:J31,FALSE)

**SUMMARY :**

Fitbit users are mostly sedentary users who spend 69.39% of the total minutes in a month doing sedentary activities. They also spend 60.88% of their sedentary minutes sleeping.