**BELLABEAT CASE STUDY**

Guiding Questions

1. What is the problem I am trying to solve?

Identify trends in smart device usage and apply these insights to a Bellabeat product to provide recommendations for growth.

1. How can your insights drive business decisions?

Insights from my analysis will show areas of potential growth, where Bellabeat should be focusing its resources.

Key Tasks

1. Identify the business task.

Sršen asks you to analyze smart device usage data 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.

The main goal is to identify trends in smart device usage.

6 steps in Data Analysis:

1 Ask.

What is the business problem.

To identify trends in smart device usage and how it applies to one Bellabeat product. Make recommendations based on insights from analysis.

2 Prepare

* Where is the data stored?

Data is stored locally on my laptop.

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

The data is organized in long format.

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

Yes, there are issues. The data does not ROCCC.

R – Reliability: The data is NOT reliable. It has a sample size of just 30 persons.

O – The data is NOT original. It was generated using Amazon Mechanical Turk.

C- The data is not comprehensive. It makes no mention of demography.

C- The data is not current. It is from 2016

C- The data has not been cited.

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

This data set is CCO: Public Domain.

This means that the person who associated a work with this deed has **dedicated**the work to the public domain by waiving all his or her rights to the work worldwide under copyright law, including all related and neighboring rights, to the extent allowed by law. You can copy, modify, distribute, and perform the work, even for commercial purposes, all without asking permission.

**Data sources used:**

The following CSV files from the dataset are being analyzed for this case study.

* dailyActivity\_merged
* dailyCalories\_merged
* dailyIntensities\_merged
* dailySteps\_merged
* sleepDay\_merged.
* weightLogInfo\_merged
* HourlyCalories\_merged
* HourlyIntensities\_merged

These subsets of data will help me identify trends in smart device usage using averages. The minutes and seconds info are too granular to paint the big picture I want to portray.

3 Process

Daily Data was loaded successfully into MS Excel, Big Query and SQL Server Management Studio and data loaded successfully in all 3 platforms. This helped me test for integrity. I counted **940 rows** across all 3 platforms for “dailyActivity\_merged” table. There were no inconsistencies.

 

However, other data sources like ‘Heartrate in seconds’ had a voluminous set of data over a million rows which got loaded into SSMS. Excel can’t load those many rows.

I chose SSMS (SQL Server Management Studio) for analysis. I find the interface quite simple and intuitive.

Data Cleaning

* Checking for NULLs

Using the ‘is NULL’ function for each column.

* Checking for duplicates

No duplicates found.

The data also shows that not just 30 people (sample size) provided data. 33 unique Ids were found using the below query.



Using the below query, it was also possible to infer the amount of data supplied by each unique Id. Most of the participants provided data for a month (31 days) but 1 person provided data for just 4 days.



**SSMS Analysis**

****

4 Analyze

* Data was organized in tables in SSMS
* Several queries were run on the data to aggregate and make it make more sense.
* Aggregated data was imported into Tableau as extracts and visualizations created to show the trends discovered.
* **Surprise**: I was surprised to see that there were actually 33 rather than 30 participants who logged activity information. Of the 33, only 24 recorded sleep patterns and 8 logged in Weight information.
* **Findings**:
  + Only 21.21% of the participants meet the 10000 steps a day recommendation by the CDC.
  + There is no correlation between activity level and sleep.
  + Activity level and calories burned show a direct relationship.
  + Very few participants were able to record Weight info.
* These insights will help me advise Bellabeat on a particular product that could be remodeled to help address the concerns from the findings.

5.Share

The Analysis was mainly done in SSMS and visualizations in Tableau.

My stakeholders are Bellabeat’s executive team and the Market Analytics team.

I would like to share my findings through a PowerPoint presentation.

6. Act

Main conclusion:

The data provided is not enough to draw actionable insights.

Additional data to support or expand on findings may be found at [Data.gov Home - Data.gov](https://data.gov/)