**CASE STUDY 2**

It has been a rewarding experience working on this case study for Bellabeat and a good process

addressing an important problem – to analyze smart device usage data for valuable insights to

guide the direction of Bellabeat’s marketing strategy.

In analyzing the case study document, I aspired to fully understand the business objectives and

what my analysis was expected to reveal to Bellabeat. I further understand they require specific

insights from a large Fitbit users’ dataset to direct their marketing strategy.

Given that we are dealing with real-world data, Fitbit users’ dataset is vast and provides me with

the perfect opportunity to dive in and analyze it.

One of the main challenges has been the cleaning of the data. The minute-long output of

physical activity, heart rate and sleep data that’s included in the dataset needs to remain

consistent in each row by generating tasks to complete, dealing with encoding errors and natural

missing values. Furthermore, creating a system to group and organize this cheap of data in an

efficient way has been a major learning curve.

A further challenge is to draw conclusions that are substantively interesting, and moreover

useful to Bellabeat’s products and users. Matching up the inferred trends in the Fitbit data to the

product features and user behaviors that are of direct relevance for Bellabeat requires

knowledge of both.

My question is designed to refine my technical approach and to make sure that my analysis

has the impact I’m after:

What advanced data cleaning techniques can be employed to handle large datasets efficiently?

How can I effectively visualize complex data to communicate insights clearly to non-technical

stakeholders?