250401_ Google Data Analytics Capstone: How can a wellness company play it smart? - Bella’s smartdevice

Ask

Goal

The goal of this analysis is to examine one of Bella’s products to gain insights into consumer usage patterns of Bella’s smart devices. These insights will serve as a foundation for refining the company’s marketing strategy.

Business Task

Analyzing smart device fitness data can reveal trends and patterns in user behavior, helping Bellabeat identify new growth opportunities and optimize marketing efforts.

Key Stakeholders

- **Urška Sršen** – Co-founder and Chief Creative Officer of Bellabeat.
- **Sando Mur** – Mathematician, Co-founder, and key executive team member at Bellabeat.
- **Bellabeat Marketing Analytics Team** – A team responsible for data collection, analysis, and reporting to guide Bellabeat’s marketing strategies. As a junior data analyst who joined the team six months ago, your role is to contribute insights that align with Bellabeat’s mission and business goals.

Deliverable

- Clearly define how consumers use Bellabeat smart devices based on available data.
- Identify key usage trends and behavioral patterns.
- Develop data-driven recommendations for a marketing campaign tailored to user habits and preferences.

Prepare

1. **Data storing**

    - The data is sourced from the **FitBit Fitness Tracker Data** (CC0: Public Domain), a dataset available through Kaggle's Mobius repository.

    - It contains personal fitness tracker data collected from **thirty Fitbit users**.

2. **Data organizing**

    - The dataset includes **minute-level** physical activity data, heart rate, and sleep monitoring details.

    - It provides insights into **daily activity levels, step counts, and heart rate trends**, allowing for an in-depth analysis of user habits.

3. **Data format**

    - The data follows a **long format**, where multiple observations per individual are recorded over time.

4. **Data credibility**
    - The data is **ROCCC-compliant** (Reliable, Original, Comprehensive, Current, and Cited).

    - However, it is important to acknowledge potential limitations due to the small sample size (**30 users**) and self-reported nature of some variables.
    - The dataset is **open-source** and publicly available under the **CC0 Public Domain license**, ensuring compliance with ethical and data privacy considerations.
    - The dataset was examined for missing values, duplicates, and inconsistencies to ensure accuracy and completeness before analysis.

5. **Research question and data relationship**

    - It provides valuable insights into user behavior, physical activity patterns, and overall device usage, which can inform marketing strategies.

6. **Others**
    - The dataset has limitations, including a **small sample size, lack of demographic diversity, and potential reporting inconsistencies**.

- **Sršen** has suggested considering **additional datasets** to enhance the robustness of the analysis.

Process

Key Tasks

1. **Check the data for errors** – Identified inconsistencies, missing values, and duplicates.

2. **Choose your tools** – **R** was selected for data processing and analysis due to its **efficiency in handling large datasets**, **comprehensive statistical functions**, and **powerful data visualization capabilities**.

3. **Transform the data** – Reformatted columns, standardized units, and ensured data compatibility for analysis.

4. **Document the cleaning process** – Logged each step for reproducibility and transparency.

Analyze

- The data was structured in **multiple datasets** (daily activity, daily calories, sleep, and weight). Since participants were not equal across datasets, **data merging** was performed where necessary, ensuring compatibility for cross-analysis.
- The **number of participants differed** across datasets:
  - **Daily activity** and **calories data** had **940** unique records.
  - **Sleep data** had only **410 unique records**, indicating fewer participants tracked sleep.
  - **Weight data** had just **67 records**, suggesting a limited number of users recorded their weight.
- **Activity Levels:**
  - **Very Active Distance** was the most preferred, followed by **Light Active Distance**.
  - **Moderately Active Distance** had the least engagement, implying users either exercised intensely or engaged in light activities but did not balance

moderate activity.

- **Minutes Spent in Different Activity Levels:**
  - **Sedentary Minutes** had the **highest recorded values**, meaning users spent a significant portion of their time inactive.
  - **Lightly Active Minutes** ranked **second**, showing that users engaged in low-intensity activities.
  - **Very Active Minutes** were higher than **Fairly Active Minutes**, suggesting users preferred more intense workouts.
- **Calories Burned vs. Activity Levels:**
  - High calorie expenditure was observed **with Light Active and Sedentary Minutes**.
- **Sleep Patterns vs. Activity Levels:**
  - Users with **Moderate and Light Active Distances** were more likely to have **400-600 minutes of sleep**.
  - Those with **Very Active Distances** showed **fewer occurrences of 400-600 minutes of sleep**.

To answer the business questions

- The trends identified in **activity levels, calorie burn, and sleep habits** will help shape **marketing strategies** for Bellabeat's smart devices by:
  - **Targeting users based on their engagement levels**—sedentary users might need encouragement through motivational content, while very active users could benefit from **performance-driven features**.
  - **Enhancing sleep tracking features**, since a **significant portion of users track sleep**, but fewer users record weight.
  - **Promoting moderate activity programs**, as engagement in moderate activity is relatively low.

Share

- **Were you able to answer the business questions?**
  Yes, based on the data analysis, it is evident that people are concerned about how their daily activities impact their calorie expenditure and sleep duration.

- **What story does your data tell?**
  Maintaining **moderate activity levels** consistently can help individuals burn a significant amount of calories while also ensuring sufficient sleep.

- **How do your findings relate to your original question?**
  The insights from the data analysis allow for the identification of the right target audience for the marketing strategy.

- **Who is your target audience?**
  Individuals who aim to burn a high number of calories and maintain quality sleep. These insights can be leveraged for targeted advertisements and e-commerce strategies.

- **What is the best way to communicate with them?**
  Using **internet pop-ups, YouTube ads, and storytelling** in marketing campaigns that feature active individuals demonstrating how smart devices can help them optimize calorie expenditure and sleep quality.

- **Can data visualization help in sharing findings?**
  Yes, statistical visualizations provide strong evidence to support the insights and make them more comprehensible.

- **Is your presentation accessible to your audience?**
  Since the graphs only include essential elements (e.g., X and Y axes), they are designed to be simple and easy to understand.

Act

- **What is your final conclusion based on your analysis?**
  Smart devices can effectively track calorie expenditure and sleep duration, helping users assess whether their training regimen is suitable for their fitness goals.

- **How could your team and business apply your insights?**
  Develop a marketing campaign that highlights how smart devices support fitness tracking and overall well-being.

- **What next steps would you or your stakeholders take based on your findings?**
  Launch a targeted marketing campaign emphasizing the benefits of smart devices for monitoring daily activity and sleep quality.

- **Is there additional data you could use to expand on your findings?**
  Yes, incorporating data on diet, heart rate, and stress levels could provide a more comprehensive understanding of fitness and wellness trends.