

Preguntas orientativas

Preguntar

¿Cuál es el problema que intentas resolver?

Encontrar áreas de oportunidad para expandir el mercado de Bellabeat.

¿Cómo tus conocimientos pueden impulsar las decisiones empresariales?

Al observar las tendencias de uso de otros dispositivos inteligentes en el mercado, se hallarán necesidades que los usuarios buscan satisfacer, de esta manera se podrá seleccionar o modificar un dispositivo que tenga el potencial de cumplirlas y mejorar las estrategias de marketing para ese producto.

Preparar

¿Dónde se almacenan tus datos?

Los datos se almacenan en una computadora de manera local en archivos de valores separados por coma.

¿Cómo están organizados los datos? ¿Están en formato largo o ancho?

Los archivos de datos se encuentran organizados en formato ancho. Se guardan en subcarpetas de nombres “**No Seleccionados**” y “**Sucios**”, dependiendo si se van a utilizar para el análisis o no.

¿Hay problemas con el sesgo o la credibilidad de estos datos? ¿Tus datos son confiables, originales, integrales, actuales y citados (ROCCC)?

No hay problemas con el sesgo, originalidad o credibilidad, dado que los datos se recopilaban directamente desde la salida de los aparatos de Fitbit. Son confiables, dado que se aplicaron criterios de elegibilidad a los usuarios que contestaron la encuesta de Amazon Mechanical Turk. En cuanto a la integridad de los datos, se puede decir que están completos en el sentido de no haberse perdido información desde su recolección, sin embargo, podrían realizar insuficientes a causa de la falta de un dispositivo que automatice la gestión del consumo de agua.

Al igual que en el caso de la integridad, los datos no son actuales, sin embargo, se utilizan al ser un caso de ejemplo que requería del uso de los mismos. La fuente de los datos se encuentra citada en el lugar indicado para su descarga, de la siguiente forma:

Furberg, R., Brinton, J., Keating, M., & Ortiz, A. (2016). Crowd-sourced Fitbit datasets 03.12.2016-05.12.2016 [Data set]. Zenodo.
<https://doi.org/10.5281/zenodo.53894>

¿Cómo estás abordando la autorización, la privacidad, la seguridad y la accesibilidad?

Si, dado que los datos tienen una licencia CC0, y por tanto son de dominio público para ser utilizados sin necesidad de solicitar autorización por parte del autor. La privacidad se aborda identificado por una id a cada cliente y nunca por su nombre real. Para el caso de la seguridad debería limitarse el acceso a la carpeta “**Bellabeat**” accediendo a sus propiedades desde el explorador de archivos y después editar los permisos en la pestaña de seguridad, sin embargo, en este proyecto no se realizó por ser un ejercicio de muestra pensado para un público amplio.

Por último, para el caso de la accesibilidad, se facilita la localización de los archivos mediante una estructura de carpetas a partir de la carpeta “**Bellabeat**”, pero también se da lugar a la portabilidad mediante la opción de modificar la ruta en la que se almacene la carpeta antes mencionada.

¿Cómo verificaste la integridad de los datos?

Asegurado que los datos generados mediante dispositivos de otra compañía (Fitbit) se relacionaran directamente con funciones en los productos similares que ofrece Bellabeat, dado esto, los registros pueden indicar tendencias de uso que sean acertadas para usuarios de esta clase de dispositivos. De esta manera, se indicaron las relaciones entre los datos y los productos, y se generó el siguiente bosquejo del proyecto:

Pastel registro de IMC, Registro de IMC y %GC, O Ninguno (Datos: weightLogInfo_merged, Columnas: Id, Fat, BMI)

Diagramas de caja y bigote de gasto calórico (Datos: dailyActivity_merged, Columnas: Id, Calories)

Pastel reporte manual y reporte automático (Datos: weightLogInfo_merged, Columnas: Id, IsManualReport)

Frecuencia de actividad fuerte, moderada y ligera (Datos: dailyActivity_merged, Columnas: Id, VeryActiveMinutes, FairlyActiveMinutes, LightlyActiveMinutes, SedentaryMinutes)

Cantidad de horas de sueño y en cama registradas (Datos: sleepDay_merged, Columnas: Id, TotalMinutesAsleep, TotalTimeInBed)

Mapa de calor registro de pasos (Datos: dailyActivity_merged, Columnas: Id, ActivityDate, TotalSteps)

Frecuencia del ritmo cardiaco (Datos: heartrate_seconds_merged, Columnas: Id, Time, Value)

Sueño continuo (Datos: sleepDay_merged, Columnas: Id, TotalSleepRecords)

Cantidad de reportes por semana (Datos: weightLogInfo_merged; dailyActivity_merged ; sleepDay_merged; heartrate_seconds_merged, Columnas: Id, Date; Id, ActivityDate; Id, SleepDay; Id, Time)

- **Aplicación Bellabeat:** actividad física, sueño, estrés, ciclo menstrual y hábitos de conciencia plena. Se conecta a su línea de productos de bienestar inteligentes.
- **Leaf:** actividad física, el sueño y el estrés.
- **Time:** actividad física, el sueño y el estrés del usuario.
- **Spring:** consumo de agua, niveles de hidratación.
- **Membresía de Bellabeat:** acceso 24/7 a una orientación totalmente personalizada sobre nutrición, actividad física, sueño, salud y belleza y conciencia plena.

¿De qué manera te ayuda a responder tu pregunta?

Brindan datos que se relacionan directa o indirectamente a las cualidades que presentan los dispositivos de Bellabeat, por ende, proporcionan fundamentos para discernir cual es el dispositivo con mayor potencial de interesar a los consumidores cuando se les presente por medio del marketing.

¿Existe algún problema con los datos?

Si, no cubren el total de las funciones de los productos.

Procesar

¿Qué herramientas eliges y por qué?

El lenguaje de programación R, ya que está orientado al manejo de datos y cuenta con una gran variedad de librerías y herramientas que permitirán realizar todo el procedimiento desde la limpieza (librerías de tidyverse) hasta la documentación (R Markdown).

¿Has garantizado la integridad de los datos?

No del todo, en el sentido en que los datos se ciñeron únicamente a los proporcionados por Fitbit. Pero, por parte de mantener el nivel de integridad que tenían, durante todo el procedimiento se observan periódicamente los efectos que los cambios tienen en los datos. Además, solo se guardaron los archivos de los datos sucios y los datos limpios una vez, mientras que los datos redundantes y que no aportarían mucho más al análisis se conservaron en la carpeta **“No Seleccionados”**.

¿Qué pasos seguiste para garantizar que tus datos están limpios?

Primero se eligieron solo los archivos con datos relevantes para el caso de estudio. Después se comprobó que no existieran datos duplicados, en el caso de los registros de sueño se realizó la corrección correspondiente. Además, se seleccionaron solo las columnas a utilizar para cada gráfico, y se realizaron los cambios de formato (en caso de requerirlos) para ser utilizables en la transformación de los datos.

¿Cómo puedes verificar que tus datos están limpios y listos para analizar?

Mediante distintas herramientas de la paquetería de R, algunas como; la función View, para observar los datos en su totalidad; la función str, para comprobar que tipo de dato tiene cada columna; la función filter, para eliminar datos que no corresponden o no son coherentes; y la función duplicated, para evidenciar las filas de datos que se encuentran más de una vez.

¿Documentaste tu proceso de limpieza para poder revisar y compartir estos resultados?

Sí, en el archivo HTML “**Documentación_Bellabeat**” creado con R Markdown.