

Informe Bellabeat Time

Análisis de actividad física y patrones de sueño

Introducción

Bellabeat es una empresa enfocada en el desarrollo de dispositivos inteligentes orientados al bienestar y la salud de las mujeres. Con el fin de apoyar la toma de decisiones estratégicas relacionadas con el producto Bellabeat Time, se realizó un análisis exploratorio de datos de actividad física y sueño obtenidos a partir de dispositivos inteligentes.

El objetivo de este análisis es identificar patrones de comportamiento asociados a:

- La actividad diaria
- La intensidad del movimiento
- Los hábitos temporales de actividad
- La relación entre actividad física y sueño

Los hallazgos obtenidos buscan aportar información relevante para el diseño de estrategias de producto y marketing basadas en hábitos reales y sostenibles de las usuarias.

Tarea empresarial

Analizar los datos de uso de dispositivos inteligentes de compañías distintas a Bellabeat con el fin de identificar patrones de comportamiento relevantes que permitan orientar decisiones estratégicas de producto y marketing para el reloj Bellabeat Time.

Preguntas de análisis

Para guiar el análisis, se definieron las siguientes preguntas clave:

1. ¿Cuáles son las tendencias más destacadas en el comportamiento diario de los usuarios de dispositivos inteligentes?

2. ¿Cómo pueden aprovecharse estas tendencias para ofrecer soluciones relevantes a las usuarias del reloj Bellabeat Time?
 3. ¿De qué manera estas tendencias pueden influir en la estrategia de marketing de Bellabeat?
-

Fuentes de datos

Dataset principal

- **Nombre:** *Fitbit Fitness Tracker Data*
 - **Plataforma:** Kaggle
 - **Autor:** Mobius
 - **Licencia:** CC0 (Dominio público)
-

Archivos utilizados en este proyecto

Los siguientes archivos fueron seleccionados por su relevancia para el análisis de actividad física y sueño:

- `dailyActivity_merged.csv`
 - `hourlySteps_merged.csv`
 - `minuteSleep_merged.csv`
-

Almacenamiento de los datos

Los datos se almacenaron localmente en la carpeta `bellabeat_project`, en formato **CSV**, lo que permitió su utilización en distintas herramientas de análisis, entre ellas:

- Hojas de cálculo
- SQL

- R

Esta flexibilidad facilitó la exploración, limpieza y análisis de los datos a lo largo del proyecto.

Credibilidad de los datos (criterios ROCCC)

Reliable (Confiables)

Los datos fueron recolectados directamente desde dispositivos Fitbit, lo que garantiza consistencia en la medición de variables como pasos, actividad y sueño.

Original

Se trata de datos originales, aunque no pertenecen directamente a la empresa Bellabeat.

Comprehensive (Integrales)

El conjunto de datos incluye información de **35 usuarios**, lo que permite identificar patrones generales, pero limita la representatividad para inferencias a gran escala.

Current (Actuales)

Los datos corresponden al año **2016**, por lo que no reflejan comportamientos recientes de los usuarios.

Cited (Citados)

La fuente de los datos está debidamente citada y disponible públicamente a través de la plataforma Kaggle.

Conclusión sobre los datos

El conjunto de datos es adecuado para realizar un **análisis exploratorio** y para identificar **tendencias generales** en el uso de dispositivos inteligentes. Sin embargo, presenta limitaciones relacionadas con el **tamaño de la muestra** y el **periodo de tiempo analizado**.

No se identificaron otras fuentes de datos públicos y confiables que permitieran complementar este análisis, ya que la información recopilada por dispositivos inteligentes suele ser de carácter privado y su acceso requiere autorización expresa de los usuarios. Adicionalmente, las empresas que comercializan este tipo de dispositivos no comparten estos datos por razones estratégicas.

Proceso de limpieza y preparación de datos

Archivo **dailyActivity**

Para el archivo **dailyActivity**, se creó una copia de trabajo del dataset original con el fin de realizar las tareas de limpieza y estandarización sin alterar los datos fuente.

En primer lugar, se verificó la estructura del conjunto de datos y se confirmó que la columna **Id** no presenta valores nulos, identificándose un total de 35 usuarios únicos.

Posteriormente, se revisó la existencia de filas duplicadas mediante la herramienta de eliminación de duplicados de Google Sheets, sin encontrarse registros repetidos.

Asimismo, se aplicaron filtros para identificar posibles valores nulos en las variables clave, sin detectarse celdas vacías. Finalmente, se validaron los formatos de los datos, asegurando que la columna de fecha estuviera correctamente estandarizada y que las variables numéricas presentaran valores consistentes.

Archivo **minuteSleep**

En el caso del archivo **minuteSleep**, se llevó a cabo un proceso de revisión y limpieza más exhaustivo debido a su nivel de granularidad. Durante esta etapa, se identificaron y eliminaron 525 filas duplicadas, correspondientes a registros idénticos de sueño a nivel de minuto.

Una vez finalizada la limpieza, el archivo depurado fue cargado en **BigQuery** para su análisis conjunto con los demás conjuntos de datos utilizados en el proyecto.

Herramientas utilizadas y justificación técnica

Inicialmente se evaluó la posibilidad de realizar la importación y el análisis de los datos utilizando MySQL 8.0. No obstante, debido a las restricciones de seguridad de MySQL Workbench para la carga de archivos locales, se optó por emplear BigQuery, una plataforma ampliamente utilizada en entornos profesionales de análisis de datos.

El uso de BigQuery permitió una integración más fluida de los archivos CSV, así como la ejecución eficiente de consultas SQL para la exploración y el análisis de los datos.

Durante la carga de los datos, se definió manualmente el esquema de algunas tablas debido a inconsistencias en el formato de fecha-hora que impedían la autodetección del esquema, además se permitió un número mínimo de registros defectuosos que no impactaron la calidad del análisis.

Este análisis está orientado al equipo de marketing y los CEO de la compañía pero también es comprensible para audiencias no técnicas. Los resultados obtenidos fueron visualizados utilizando **R**, empleando la librería **ggplot2**, con el fin de comunicar de manera clara y efectiva los patrones identificados en los datos.

Sección 1. Análisis del nivel de actividad diaria

Objetivo

Identificar las tendencias generales en el nivel de actividad diaria de los usuarios de dispositivos inteligentes, con el propósito de comprender los patrones de movimiento y sedentarismo que puedan orientar decisiones de marketing para el reloj Bellabeat Time.

Descripción del análisis

Para este análisis se utilizó el conjunto de datos **dailyActivity**, clasificando los días registrados según el nivel de actividad diaria. El total de pasos diarios se empleó como indicador principal de actividad física.

A partir de rangos de pasos predefinidos, los días fueron categorizados en tres niveles:

- Sedentarios
- Moderadamente activos
- Activos

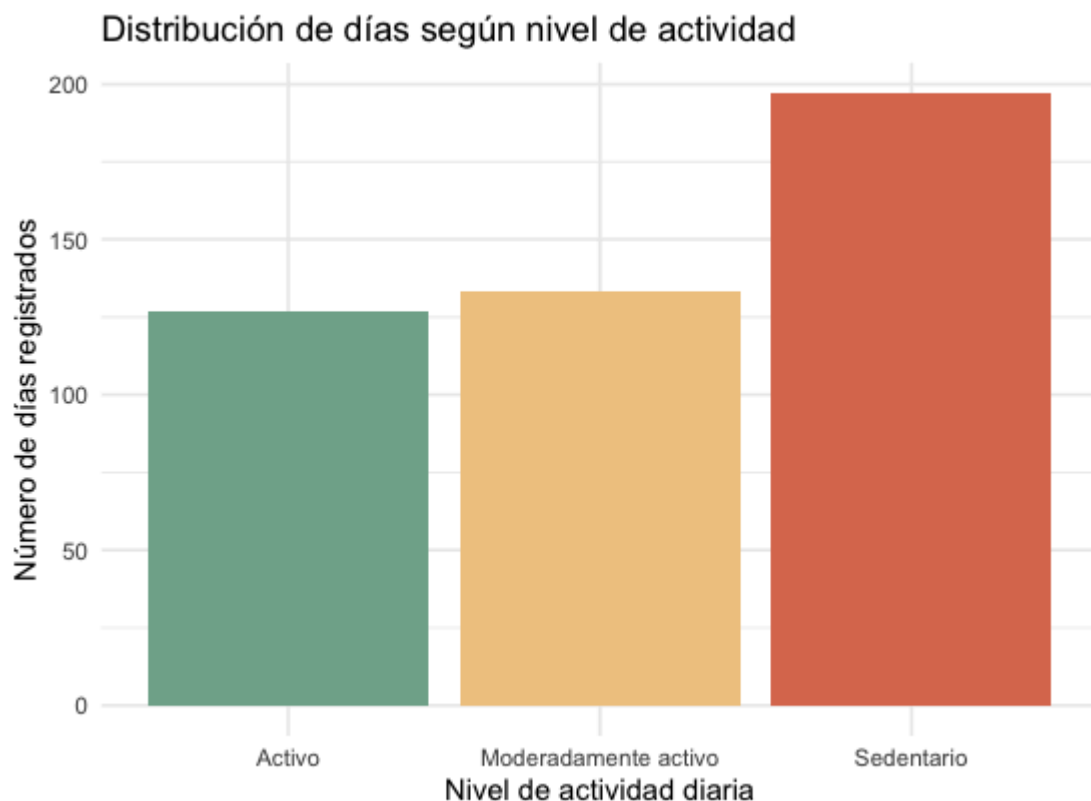
Esta clasificación permitió obtener una visión general del comportamiento diario de los usuarios durante el periodo analizado.

Resultados principales

Los resultados muestran que una proporción significativa de los días se concentra en las categorías moderadamente activo y activo, lo que indica que los usuarios presentan, en general, niveles de movimiento diarios aceptables.

No obstante, también se observa una presencia relevante de días clasificados como sedentarios, lo que evidencia que una parte sustancial de los usuarios alterna periodos de actividad con jornadas de baja movilidad.

Figura 1. Distribución de días según nivel de actividad diaria de los usuarios.



Interpretación para Bellabeat Time

Estos hallazgos evidencian una oportunidad para posicionar el reloj Bellabeat Time como una herramienta orientada a la promoción de hábitos de movimiento progresivos y sostenibles.

Dado que una proporción relevante de los días presenta niveles bajos de actividad, el producto puede enfocarse en acompañar a las usuarias en la transición desde el sedentarismo hacia rutinas más activas. A través de metas personalizadas, recordatorios de actividad y retroalimentación diaria, Bellabeat Time puede incentivar pequeños incrementos en el movimiento cotidiano, facilitando la adopción gradual de niveles de actividad más saludables y alcanzables.

Sección 2. Análisis del sedentarismo

Objetivo

Analizar el nivel de sedentarismo de los usuarios a través del promedio de minutos sedentarios diarios, con el fin de identificar oportunidades para promover hábitos de movimiento más saludables.

Descripción del análisis

Para este análisis se utilizó el conjunto de datos **dailyActivity**, a partir del cual se calculó el promedio de minutos sedentarios diarios registrados por los usuarios.

Adicionalmente, este valor fue comparado con los promedios de minutos correspondientes a los distintos niveles de actividad física:

- Actividad ligera
- Actividad moderada
- Actividad muy activa

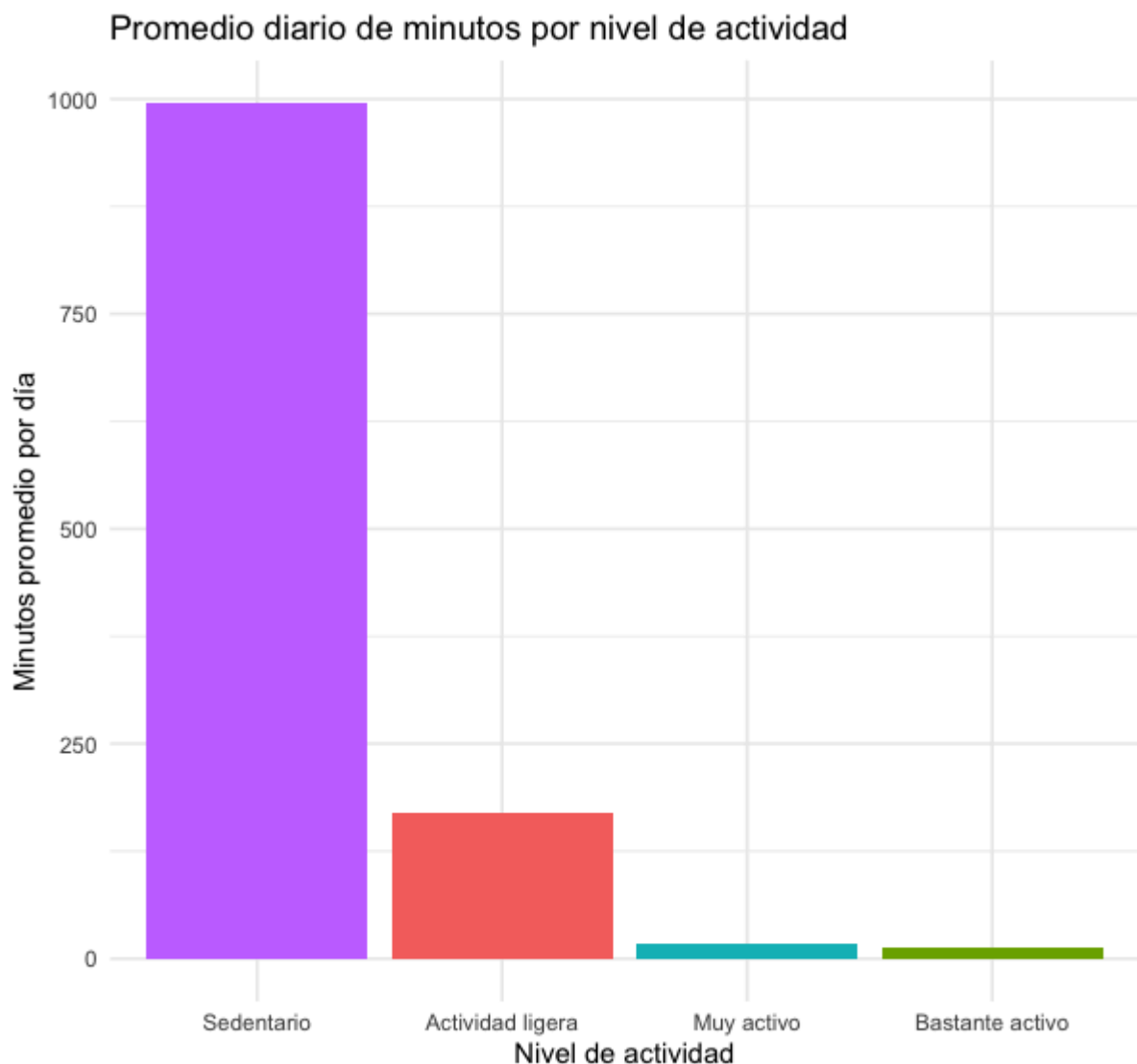
Esta comparación permitió contextualizar el peso relativo del sedentarismo dentro del comportamiento diario de los usuarios.

Resultados principales

Los resultados muestran que los minutos sedentarios representan la mayor proporción del tiempo diario promedio, superando ampliamente el tiempo dedicado a la actividad física, incluso cuando se consideran los minutos de actividad ligera.

Este patrón sugiere que, aunque los usuarios realizan algún nivel de actividad, una parte significativa de su jornada diaria transcurre en estados de inactividad prolongada.

Figura 2. Distribución promedio diaria del tiempo dedicado a cada nivel de actividad.



Interpretación para Bellabeat Time

Los resultados evidencian que el tiempo sedentario representa la mayor parte de la jornada diaria de los usuarios, mientras que la actividad física intensa es limitada y se concentra en períodos cortos. La mayor proporción del movimiento diario corresponde a actividad ligera, lo que sugiere que muchos usuarios realizan desplazamientos cotidianos, pero no alcanzan niveles sostenidos de actividad moderada o vigorosa.

Este patrón de comportamiento representa una oportunidad clara para el reloj Bellabeat Time, que puede posicionarse como un dispositivo orientado al bienestar diario, más que al rendimiento deportivo. A través de recordatorios inteligentes, metas personalizadas y retroalimentación continua, Bellabeat Time puede ayudar a las usuarias a reducir el tiempo sedentario e incorporar más movimiento de forma progresiva y alcanzable, alineándose con un enfoque de salud integral y sostenible.

Sección 3. Análisis de patrones horarios de actividad

Objetivo

Identificar los patrones de actividad física a lo largo del día para comprender en qué momentos los usuarios son más activos y cuándo predomina el sedentarismo, con el fin de apoyar decisiones de diseño y marketing del reloj Bellabeat Time.

Descripción del análisis

Para este bloque se utilizó el conjunto de datos **hourlySteps**, el cual contiene el número de pasos registrados por hora para cada usuario.

A partir de este dataset:

- Se extrajo la hora del día a partir del registro temporal.
- Se calculó el promedio de pasos por hora, lo que permitió identificar tendencias generales de actividad a lo largo de una jornada típica.

El análisis se centró en detectar:

- Horas de mayor actividad
 - Horas de menor actividad
 - Posibles patrones asociados a rutinas diarias
-

Resultados principales

Los resultados muestran que:

- La actividad es muy baja durante la madrugada.
- Se observa un incremento progresivo de la actividad a partir de la mañana.
- Los picos de actividad se concentran en las horas diurnas y en las primeras horas de la noche.
- Posteriormente, la actividad vuelve a disminuir de forma gradual.

Este patrón sugiere la presencia de rutinas diarias asociadas a:

- Desplazamientos
 - Jornada laboral
 - Actividades personales
-

Interpretación para Bellabeat Time

El análisis de los patrones horarios indica que los usuarios presentan franjas específicas de mayor actividad a lo largo del día, mientras que otros períodos permanecen marcadamente inactivos.

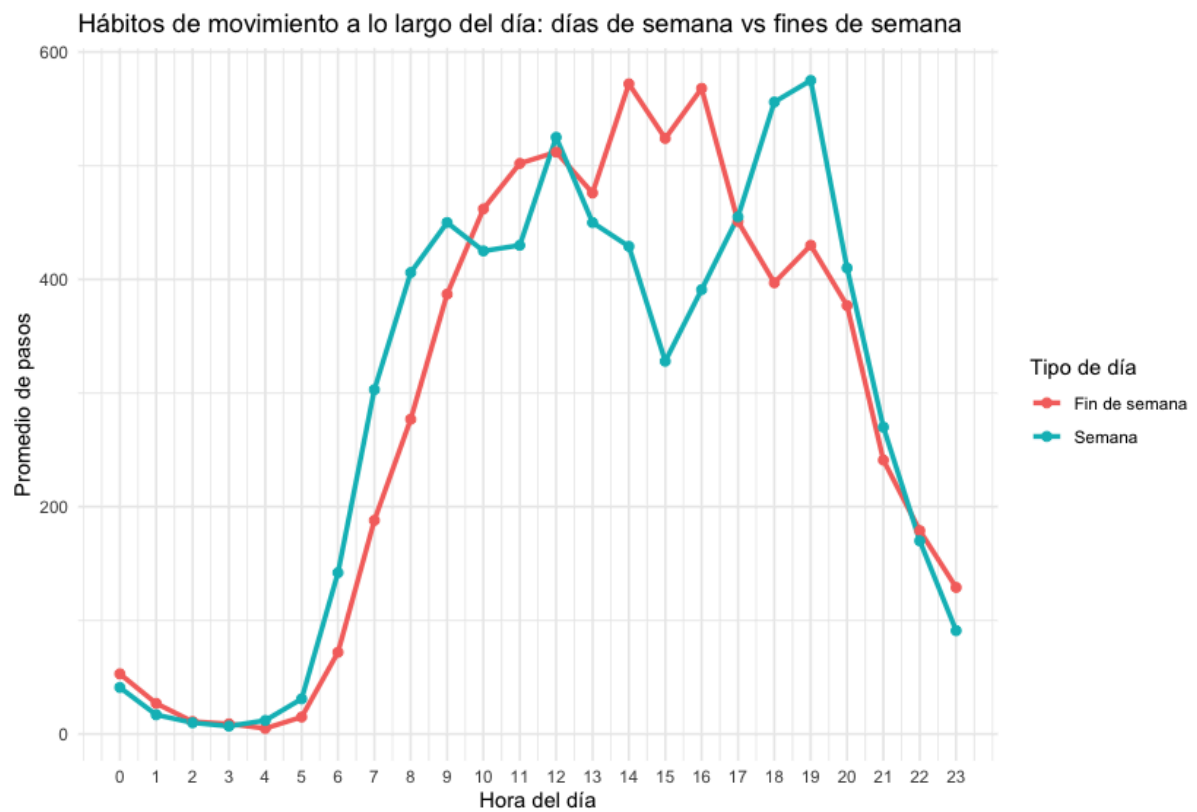
Esta información puede ser utilizada por Bellabeat Time para optimizar el momento en que se envían recordatorios de movimiento o notificaciones motivacionales. Al identificar horas recurrentes de baja actividad, el dispositivo puede ofrecer intervenciones oportunas y personalizadas, fomentando pausas activas y ayudando a reducir el sedentarismo sin interrumpir los momentos en los que las usuarias ya se encuentran naturalmente activas.

Implicación de los patrones diferenciados por tipo de día (semana vs. fin de semana)

La comparación entre días de semana y fines de semana evidencia diferencias claras en los patrones de actividad física. Durante los días de semana, la actividad tiende a concentrarse en franjas horarias específicas asociadas a rutinas estructuradas, mientras que en los fines de semana se observa un comportamiento más flexible, con variaciones tanto en la intensidad como en la distribución de la actividad a lo largo del día.

Estas diferencias sugieren que las usuarias ajustan su nivel de movimiento según el contexto semanal, lo que representa una oportunidad para que Bellabeat Time adapte sus mensajes y recordatorios de actividad en función del tipo de día. Al considerar estas dinámicas, el dispositivo puede ofrecer intervenciones más oportunas y personalizadas, alineadas con los hábitos reales de las usuarias y orientadas a promover un bienestar sostenible.

Figura 3. Distribución promedio de la actividad física a lo largo del día, clasificada por hora.



Sección 4. Relación entre actividad física y patrones de sueño

Objetivo

Analizar la relación entre el nivel de actividad física diaria y la duración del sueño, con el fin de identificar patrones que permitan comprender cómo los hábitos de movimiento se asocian con el descanso de los usuarios.

Descripción del análisis

Para este análisis se integraron los conjuntos de datos `dailyActivity` y `minuteSleep` mediante un JOIN por identificador de usuario (`Id`) y fecha. Esta integración permitió analizar, a nivel diario, la relación entre el número total de pasos y el tiempo total de sueño registrado.

El análisis se centró en:

- Evaluar si mayores niveles de actividad física diaria se asocian con una mayor duración del sueño.
 - Identificar patrones generales en la relación entre actividad y descanso, sin asumir una causalidad directa.
-

Resultados principales

El análisis integrado muestra que la relación entre actividad física y sueño no es lineal ni fuerte. Para confirmar esta observación también se calculó el coeficiente de correlación entre el número de pasos diarios y los minutos totales de sueño, el cual arrojó un valor de $r = -0.18$, correspondiente a una correlación negativa débil.

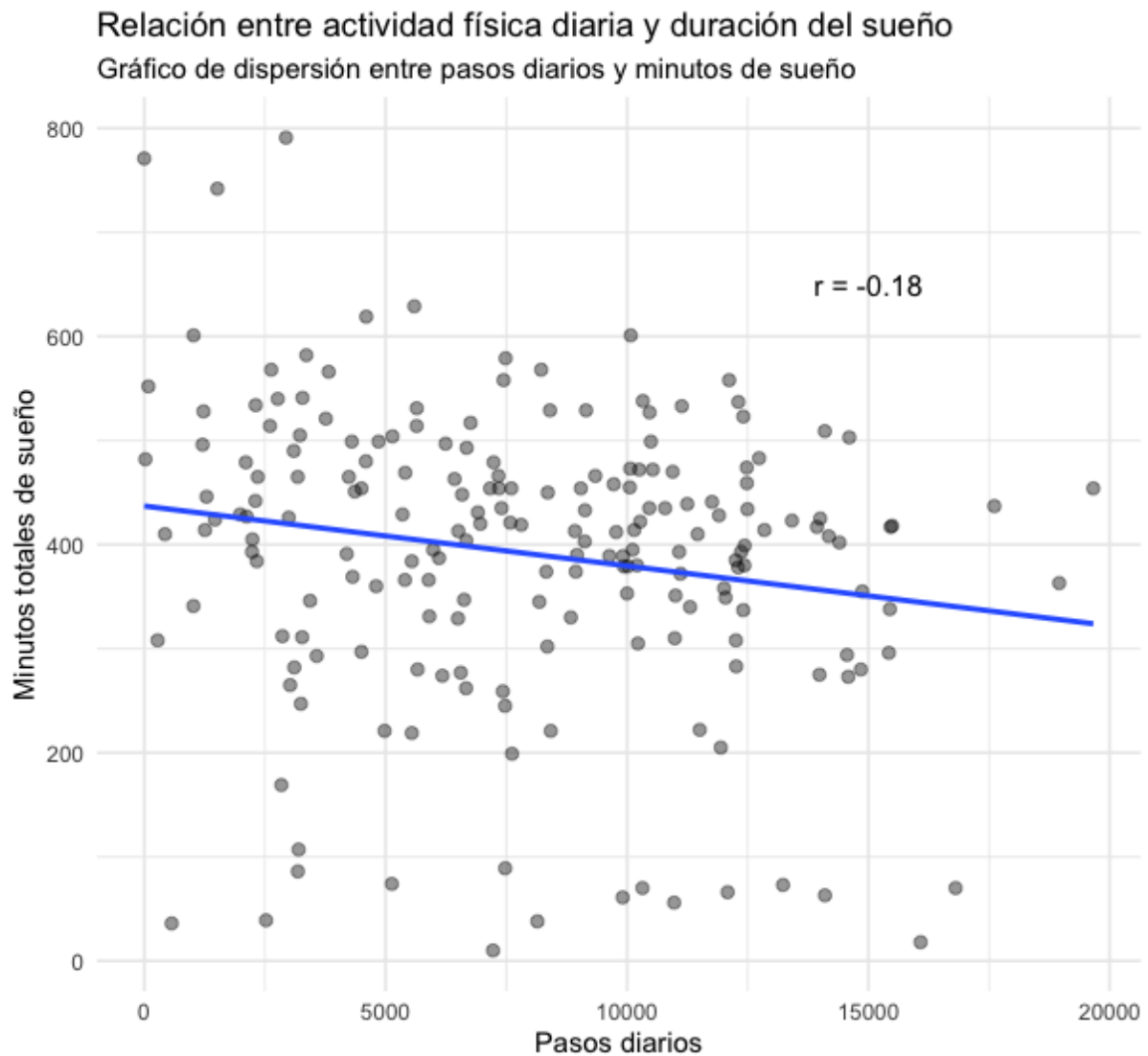
Este resultado indica que no existe una asociación significativa entre una mayor cantidad de pasos diarios y una mayor duración del sueño. En consecuencia, el tiempo de descanso de los usuarios no parece depender directamente del nivel de actividad física medido en pasos, sino que probablemente está influenciado por otros factores relacionados con hábitos, rutinas y condiciones individuales.

Interpretación para Bellabeat Time

La relación observada entre actividad física y sueño resalta la oportunidad de posicionar Bellabeat Time como un dispositivo que monitorea tanto el movimiento como el descanso dentro de una visión integral del bienestar.

Al ofrecer retroalimentación conjunta sobre actividad y sueño, el reloj puede ayudar a las usuarias a comprender mejor cómo sus hábitos diarios se relacionan con su recuperación, promoviendo decisiones más informadas y sostenibles a largo plazo.

Figura 4. Relación entre el número de pasos diarios y la duración total del sueño.



Sección 5. Síntesis de hallazgos y recomendaciones estratégicas

Objetivo

Integrar los principales hallazgos obtenidos en los análisis anteriores para formular **recomendaciones accionables** que apoyen la toma de decisiones de negocio y marketing de **Bellabeat**.

Hallazgos principales

Actividad diaria

Si bien la suma de los niveles de actividad moderada e intensa supera al sedentarismo, al analizar las categorías de forma individual se observa que el **sedentarismo predomina** como el nivel de actividad más frecuente.

Distribución de la intensidad de la actividad

La mayor parte del tiempo activo de las usuarias corresponde a **actividad ligera**, mientras que los minutos dedicados a actividad intensa son considerablemente menores.

Patrones temporales de actividad

Se identifican diferencias claras entre los **días de semana y los fines de semana**. Durante la semana, los patrones de actividad tienden a ser más estructurados y concentrados en horarios específicos, mientras que los fines de semana presentan una mayor variabilidad en la distribución del movimiento.

Relación entre actividad y sueño

No se evidencia una relación fuerte y directa entre el número de pasos diarios y la duración total del sueño, lo que sugiere que el descanso está influenciado por múltiples factores más allá del nivel de actividad física.

Recomendaciones accionables para Bellabeat Time

- Los **mensajes y notificaciones** del dispositivo deben ser **personalizados**, considerando el nivel de actividad y el contexto diario de cada usuaria.
- La **segmentación de usuarias** según su nivel de actividad y el tipo de día (entre semana o fin de semana) resulta clave para ofrecer recomendaciones más relevantes y efectivas.
- La **estrategia de marketing** debe basarse en **hábitos reales y sostenibles**, priorizando el bienestar integral sobre metas de alto rendimiento.

Conclusión

El análisis de los datos de uso de dispositivos inteligentes permitió identificar patrones claros en la actividad diaria, la intensidad del movimiento, los horarios de mayor y menor actividad, y la relación entre actividad física y sueño. Los resultados muestran que, aunque las usuarias realizan actividad física de manera regular, el **sedentarismo y la actividad ligera predominan** dentro de su comportamiento diario, y los niveles de actividad intensa son limitados.

Asimismo, se evidencian **patrones temporales diferenciados** entre días de semana y fines de semana, lo que sugiere que los hábitos de movimiento están fuertemente influenciados por el contexto diario y las rutinas personales. Por otro lado, la relación entre actividad física y sueño resulta débil, indicando que el descanso depende de múltiples factores más allá del número de pasos diarios.

Estos hallazgos refuerzan la oportunidad de posicionar **Bellabeat Time** como un dispositivo enfocado en el **bienestar integral**, capaz de acompañar a las usuarias mediante recomendaciones personalizadas, recordatorios oportunos y una visión equilibrada de actividad y descanso. De esta manera, Bellabeat puede diseñar estrategias de producto y marketing basadas en hábitos reales y sostenibles, alineadas con las necesidades cotidianas de sus usuarias.

Nota metodológica

El presente análisis se desarrolló siguiendo las fases del proceso de análisis de datos propuestas en el programa **Google Data Analytics**, que incluyen: **preguntar, preparar, procesar, analizar, compartir y actuar**.

Los datos utilizados provienen del conjunto **Fitbit Fitness Tracker Data**, disponible públicamente en la plataforma Kaggle. Previo al análisis, se realizó un proceso de limpieza y validación de los datos, que incluyó la verificación de duplicados, valores nulos y formatos de fecha, así como la estandarización de las variables relevantes.

El análisis se llevó a cabo mediante consultas SQL ejecutadas en **BigQuery**, lo que permitió integrar y explorar múltiples conjuntos de datos de manera eficiente. Posteriormente, los resultados fueron visualizados utilizando **R** y la librería **ggplot2**, con el objetivo de comunicar de forma clara los patrones identificados.

Dado el tamaño limitado de la muestra y el periodo temporal de los datos, los resultados deben interpretarse como **hallazgos exploratorios**, orientados a identificar tendencias generales y oportunidades estratégicas, sin asumir relaciones causales ni generalizaciones a poblaciones más amplias.