



Proyecto Final: ¿Cómo puede hacer una empresa tecnología para el bienestar para tomar decisiones inteligentes?

Autora: Carmen Witsman García

09-10-2024



Tabla de contenidos

Tabla de contenidos

1. Descripción de la tarea empresarial

2. Fuentes de datos utilizadas

3. Preparación y manipulación de datos

4. Análisis de datos y hallazgos clave

5. Recomendaciones para Bellabeat

1. Descripción de la tarea empresarial

En este caso práctico planteado en el Curso de Análisis de Datos de Google, analizaremos datos públicos recogidos por relojes inteligentes de la empresa FitBit,

con el objetivo de **identificar patrones que ayuden a la empresa Bellabeat a mejorar sus dispositivos inteligentes.**

Bellabeat se trata de una empresa comprometida con mejorar la salud y bienestar de las mujeres. Para ello, cuenta con varios dispositivos de seguimiento que se conectan a su propia aplicación, lo que permite a sus clientes hacer un seguimiento de datos relacionados con su salud, como su actividad física o sueño.

En este análisis, responderemos las siguientes cuestiones:

- 1. ¿Cuáles son algunas tendencias de uso de los dispositivos inteligentes?
- 2. ¿Cómo se podrían aplicar estas tendencias a los clientes de Bellabeat?
- 3. ¿Cómo podrían ayudar estas tendencias a influir en la estrategia de marketing de Bellabeat?

2. Fuentes de datos utilizadas

Para extraer los datos de los relojes inteligentes de la empresa Fitbit de distintos usuarios, he extraído en Kaggle varios conjuntos de datos de dominio público. El enlace para acceder a ellos es el siguiente:

<https://www.kaggle.com/datasets/arashnic/fitbit>

Estos *datasets* fueron generados por los participantes de una encuesta distribuida a través de Amazon Mechanical Turk entre el 3 y el 5 de diciembre de 2016. En ellos se encuentra información de seguimiento de 30 usuarios de Fitbit incluidos los datos de salida por minuto de actividad física, frecuencia cardíaca y monitoreo del sueño. Las observaciones abarcan del 12 de abril de 2016 al 12 de mayo de 2016.

Las variables que serán objeto de este análisis son las siguientes:

1 - Conjunto de datos "dailyActivity_merged.csv":

- **Id**: Número de identificación único
- **Calories**: Calorías quemadas
- **SedentaryMinutes**: Minutos en los que no hay actividad de movimiento
- **TotalDistance**: Distancia recorrida en kilómetros

- **TotalTimeInBed**: Tiempo total en minutos pasados en la cama
- **WeightKg**: Peso en kilogramos
- **ActivityHour**: Hora de actividad

Todas estas variables corresponden a las distintas observaciones para cada individuo en un día.

2 - Conjunto de datos "hourlyCalories_merged.csv":

- **ActivityHour**: Fecha y hora de actividad
- **Calories**: Calorías quemadas

Miden cada hora las calorías quemadas por el usuario.

3 - Conjunto de datos "weightLogInfo_merged.csv"

- **BMI**: Índice de Masa Corporal (IMC)

De manera manual o automática se mide el IMC del usuario.

3. Preparación y manipulación de datos

Para tratar los datos he usado las siguientes herramientas:

- **Jupyter Notebook y Python**: He creado un informe en Jupyter Notebook intercalando código en Python con comentarios acerca de los resúmenes numéricos y gráficos, incluyendo información sobre este proyecto.

Puedes consultar el informe en: https://github.com/carmenwi/Google-Data-Analytics-Professional-Certificate/blob/main/Notebook_Informe_Python.ipynb

- **RStudio**: He utilizado el software de R para generar ciertos gráficos que he incluido en el informe en Python. Para crear estos gráficos, he hecho uso de la interfaz gráfica de R Commander y de la librería Dplyr para el filtrado de datos.

Puedes consultar el informe en: https://github.com/carmenwi/Google-Data-Analytics-Professional-Certificate/blob/main/R_informe_markdown.pdf

El tratamiento de datos ha consistido en esta secuencia de pasos:

1 - Importar los datos en Python con la librería pandas

2 - Identificar los valores nulos. Ninguna de nuestras variables presentaba valores nulos.

3 - Cambiar formato de la variable ActivityHour a numérico.

4 - Identificar valores atípicos. En todas las variables el porcentaje de datos atípicos era alrededor del 40-50%.

```
... Atípicos en columna Calories: 11775456 - Porcentaje del total: 47.83131274341743 %  
Atípicos en columna SedentaryMinutes: 11019898 - Porcentaje del total: 44.762273973811304 %  
Atípicos en columna TotalDistance: 11666618 - Porcentaje del total: 47.38921823630296 %  
Atípicos en columna WeightKg: 9142774 - Porcentaje del total: 37.137490262490516 %  
Atípicos en columna ActivityHour: 11267578 - Porcentaje del total: 45.76833773391449 %  
Atípicos en columna TotalTimeInBed: 11887020 - Porcentaje del total: 48.28448012605693 %
```

Figura 1. Output en Python del código para identificar valores atípicos y sus respectivos porcentajes

No manipulé los valores atípicos debido a que se trata de un conjunto de datos con más de 24 millones de observaciones y con datos muy susceptibles a ser atípicos como son las calorías o las distancias.

En su lugar, nos pararemos a observar de cerca cada uno de los casos, ya que tenemos 6 distintos individuos con características diferentes que podemos analizar por separado.

4. Análisis de datos y hallazgos clave

Nota: No se distingue por sexo o edad a los usuarios de la encuesta, y no son una muestra grande representativa de la población, por lo que el análisis puede estar sesgado.

Para comenzar, estudiamos la correlación entre nuestras distintas variables a través de una matriz de correlación representada en un mapa de calor.

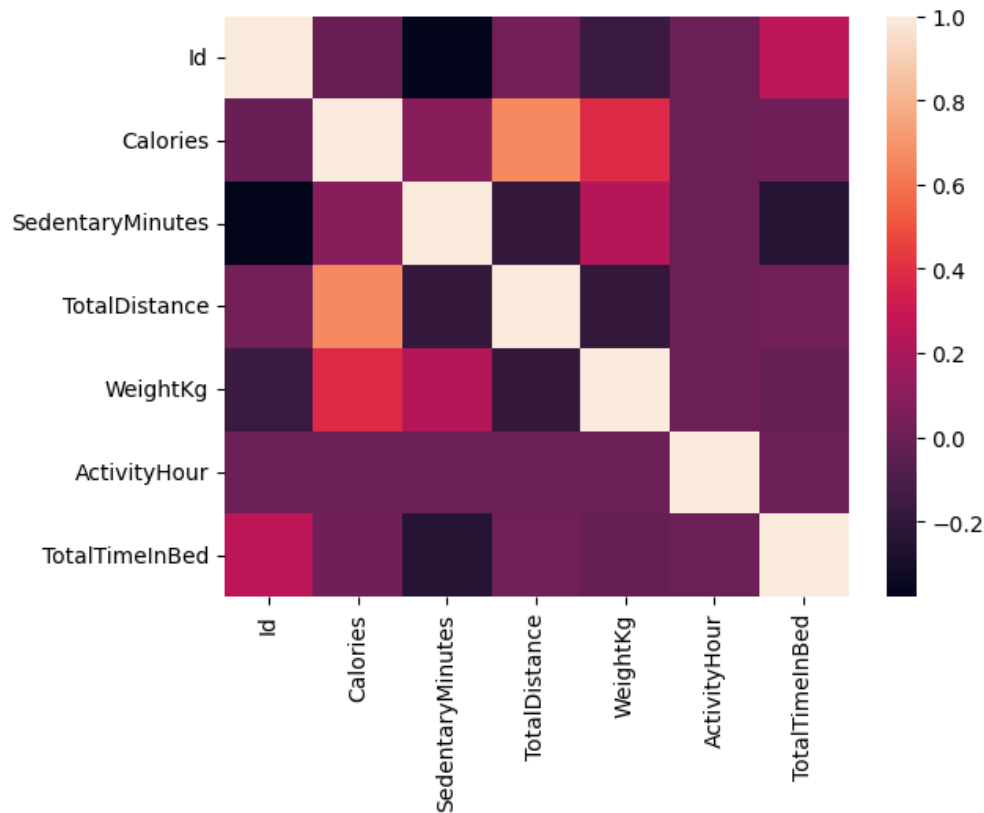


Figura 2. Mapa de calor de matriz de correlaciones generado en Python

No se hallan correlaciones fuertes y claras, por lo que no se puede concluir nada acerca de las correlaciones entre las variables.

Seguimos echando un vistazo general a los distintos usuarios que conforman el *dataset*, y calculamos en una tabla cuál es la media para cada una de las variables.

	Id	Calories	SedentaryMinutes	TotalDistance	WeightKg	ActivityHour	TotalTimeInBed
0	1503960366	1816.419355	848.161290	7.809677	52.599998	11.456067	383.200000
1	1927972279	2172.806452	1317.419355	0.634516	133.500000	11.413043	437.800000
2	4319703577	2037.677419	735.806452	4.892258	72.350002	11.444751	501.961538
3	4558609924	2033.258065	1093.612903	5.080645	69.639999	11.413043	140.000000
4	5577150313	3359.633333	754.433333	6.213333	90.699997	11.398305	460.615385
5	6962181067	1982.032258	662.322581	6.585806	61.553334	11.401639	466.129032

Figura 3. Tabla de medias agrupadas por la variable "Id" en Python

De aquí podemos hacernos varias preguntas:

1 - ¿El usuario que ha pasado más tiempo en la cama es el más sedentario?

No. Como podemos ver en el resumen numérico, el que más tiempo pasa en la cama de media al día (502 mins, es decir, 8 h 37 mins), pasa de media unas 12h30 sedentarias, lo que es menos que la media general. Es la segunda persona que más se mueve durante el día. El primer puesto lo supera por 1h30 más de movimiento, durmiendo una media de 7h 45mins al día.

De aquí podemos concluir que dormir las 8 horas necesarias para el cuerpo puede impulsar a los usuarios a tener energía para moverse durante el día.

2 - ¿Los usuarios queman calorías andando/corriendo o haciendo ejercicios estáticos?

El usuario que quema más calorías (+3.000 cal/día de media) recorre unos 6km al día de media, ocupando el tercer puesto del usuario que más distancia recorre. Esto podría indicar que combina caminar con ejercicio físico como pesas, flexiones o pilates. El segundo usuario que más calorías quema recorre unos 5km de media al día.

Esto nos puede llevar a pensar que los usuarios que poseen el reloj inteligente concilian el andar fuera de casa con ejercicio físico ya sea dentro de casa o en el gimnasio.

3 - ¿Cuál es la tendencia de quema de calorías según el peso corporal del usuario?

En el caso de la persona que presenta más peso, es la que menos se mueve (solo 2h al día, en total 63 metros de media). Para este caso analizaremos más adelante

el índice de masa corporal para saber el estado de su salud, y así tomarlo en cuenta para los informes.

En los demás casos, la relación de peso-calorías se mantiene equilibrada, los que pesan entre 50 y 70 kilogramos queman alrededor de 2000 calorías diarias. Sin embargo, hay otro caso atípico, el usuario de 90 kilogramos de peso quema unas 3.500 calorías diarias de media.

Este caso atípico corresponde al de una persona con un ligero sobrepeso, lo que puede indicar que este grupo de personas utilice el reloj inteligente para reconducir su peso a un número saludable.

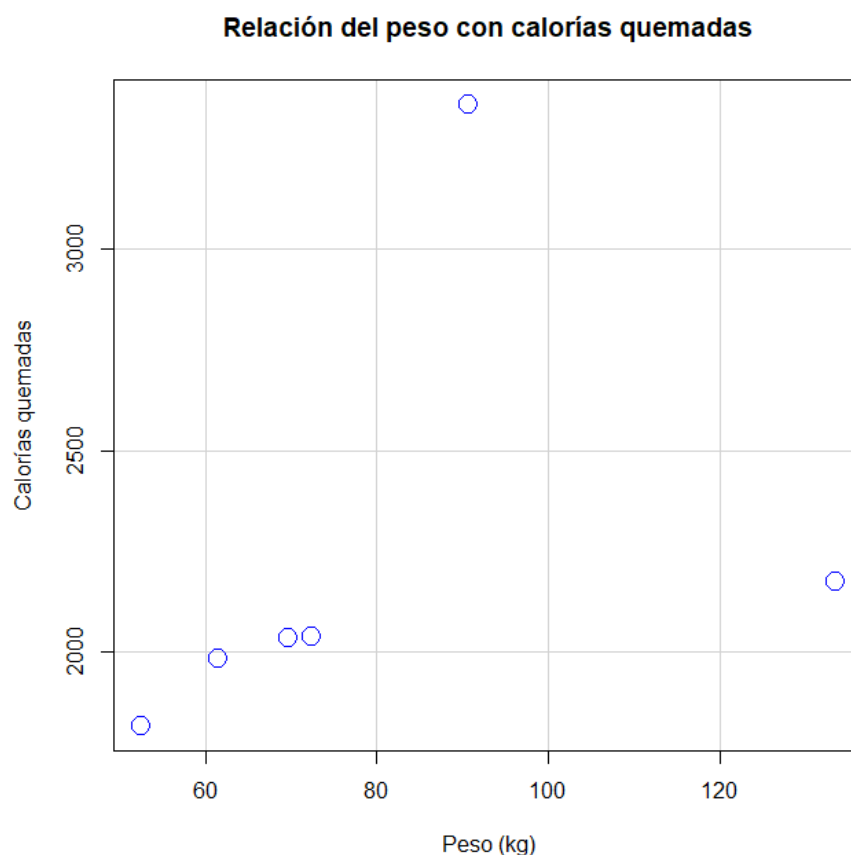


Figura 4. Diagrama de dispersión de las variables WeightKg y Calories

La persona con sobrepeso peligrosamente perjudicial para la salud no se mueve ni quema las calorías suficientes, mientras

que la persona que pasa de estar saludable a tener un ligero sobrepeso se esfuerza en recuperar su estado saludable.

4 - ¿Cuál es la hora más activa de los usuarios?

Elaborando en R un histograma a partir del conjunto de datos de "hourlyCalories_merged.csv", vemos que la hora más activa de los usuarios es entre las 12am y las 1am, pero se mantienen en movimiento de manera uniforme durante el día, en general. El 50% de los datos muestra que se levantan a las 5am y terminan su día a las 5pm.

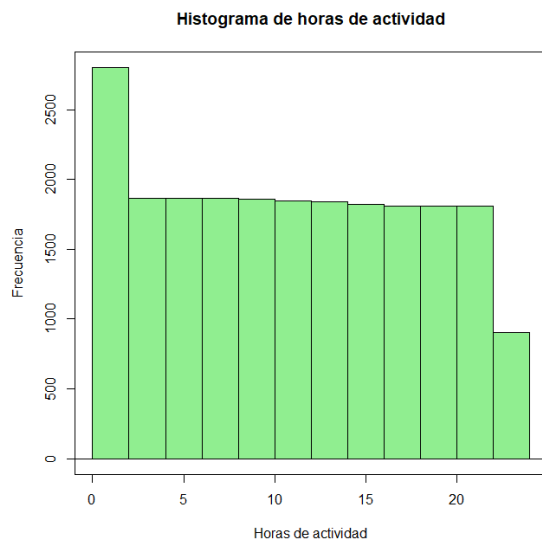


Figura 5. Histograma de variable ActivityHour

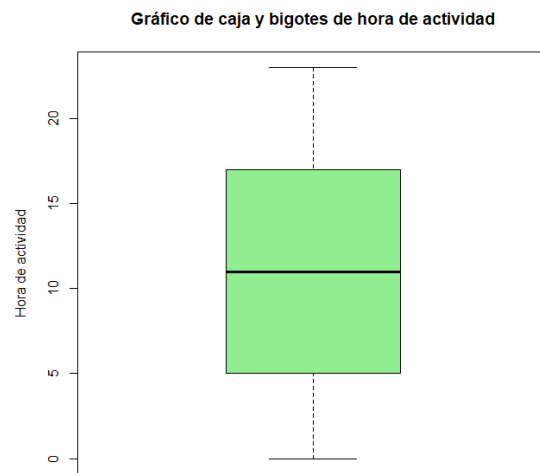


Figura 6. Boxplot de variable ActivityHour

¿Qué sucede si miramos en el caso general de las horas de actividad en las que queman más de 200 calorías? La distribución cambia. Veamos el histograma:

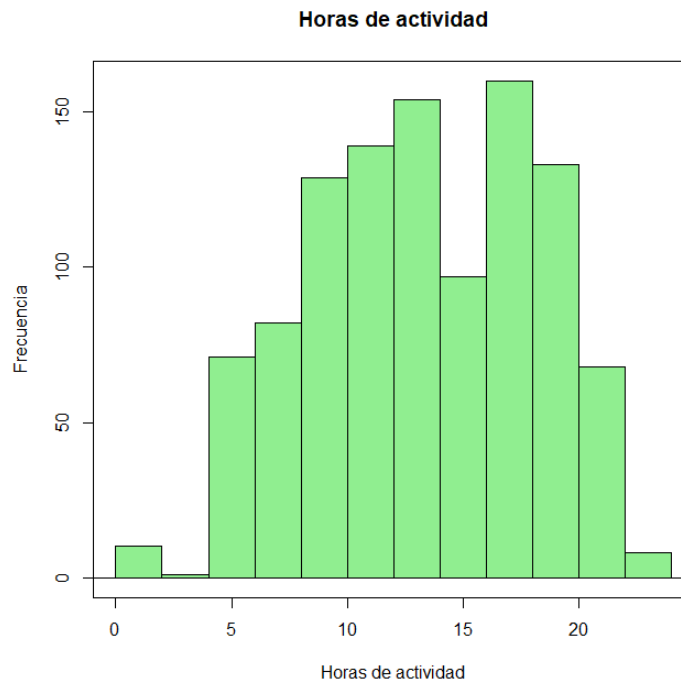


Figura 7. Histograma de variable ActivityHour con filtrado ≥ 200 de la variable Calories

En este caso, los usuarios se mueven de manera algo más moderada entre las 4pm y las 6pm. Aun así, generalmente se mueven de forma moderada durante todo el día, especialmente desde las 6am hasta las 8pm.

Las horas clave de actividad pueden estar comprendidas entre las 6am y las 8pm, aunque la franja de datos centrales comprenden desde las 5am hasta las 5pm.

5 - ¿Cuál es el porcentaje de usuarios con sobrepeso?

Para calcular este dato, nos basaremos en el Índice de Masa Corporal (IMC). Lo extraeremos del conjunto de datos "weightLogInfo_merged.csv".

	Id	BMI
0	1503960366	22.650000
1	1927972279	47.540001
2	2873212765	21.570001
3	4319703577	27.415000
4	4558609924	27.214000
5	5577150313	28.000000
6	6962181067	24.028000
7	8877689391	25.487083

Figura 8. Tabla de variable BMI según Id

De los 8 usuarios presentes en el conjunto de datos, vemos que 5 de ellos superan el índice de 25, que es el que marca el inicio de sobrepeso. Eso constituye un 62.5% de los usuarios en este estudio. 1 de estas personas sufre de obesidad mórbida. Nos damos cuenta de que la persona que más se mueve, resulta tener sobrepeso.

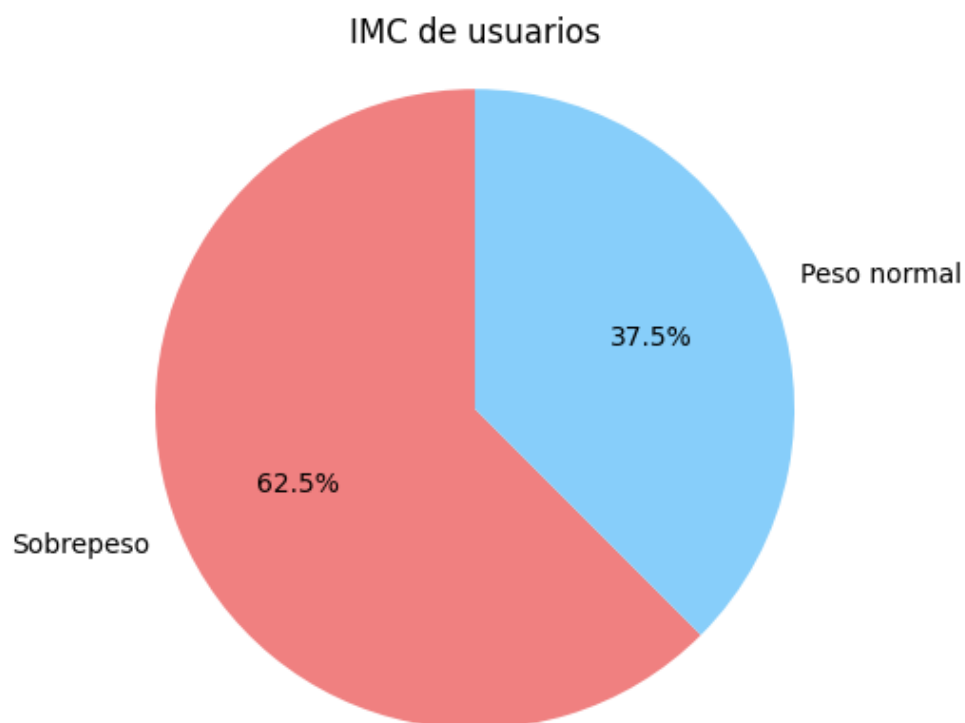


Figura 9. Gráfico de tarta de la variable BMI según filtrado de ≥ 25 y < 25

6 - ¿Qué diferencia hay entre las horas de actividad de las personas con sobrepeso y las personas de peso normal?

Elaboramos otros histogramas para comparar las horas de actividad moderada. En el caso de las personas con sobrepeso, filtramos los casos en los que queman más de 200 calorías en una hora y, en el otro caso, cuando queman más de 100 calorías.

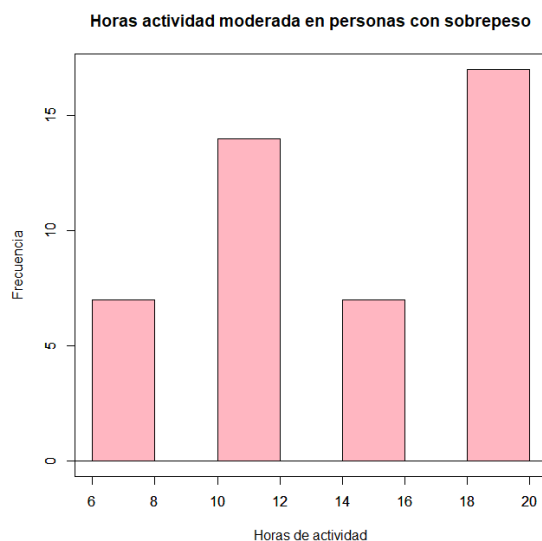


Figura 10. Histograma de variable ActivityHour con filtrado de personas con sobrepeso y ≥ 200 Calories

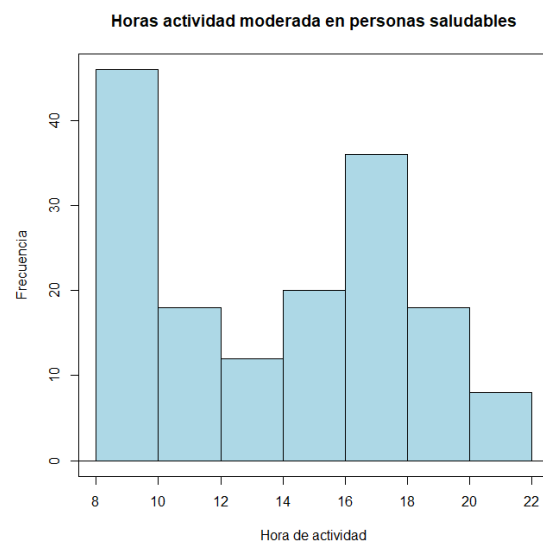


Figura 11. Histograma de variable ActivityHour con filtrado de personas sin sobrepeso y ≥ 100 Calories

Hay una clara diferencia en la continuidad de la actividad física moderada a lo largo del día entre estos dos grupos de personas. Los que se mantienen activos durante todo el día son aquellas personas que no tienen sobrepeso, y su hora más activa es entre las 8am y las 10am, seguida del tramo horario de las 4pm y las 6pm.

Por otro lado, las personas con sobrepeso parecen descansar entre intervalos de movimiento, siendo el momento del día favorito para moverse entre las 6pm y 8pm.

Pasamos a ver la diferencia en las horas de actividad más intensa, en la que consideraremos que para las personas con sobrepeso es cuando queman más de 450 calorías en 1 hora, y 250 calorías en el otro caso.

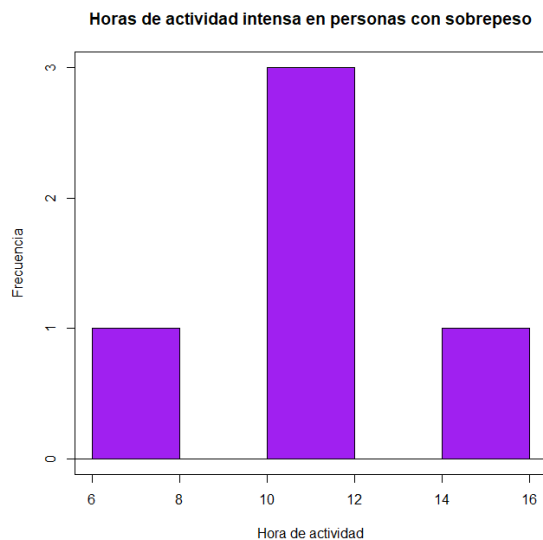


Figura 12. Histograma de variable ActivityHour con filtrado de personas con sobrepeso y \geq 450 Calories

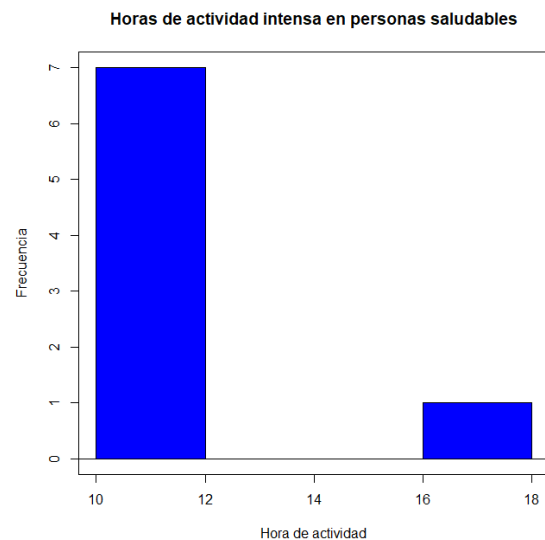


Figura 13. Histograma de variable ActivityHour con filtrado de personas sin sobrepeso y \geq 250 Calories

A pesar de ser más las personas con sobrepeso, son las que menos actividad física intensa hacen. En los pocos casos en los que la hacen, la mayoría es entre las 10 y las 12, así como las personas con peso normal.

7 - Durante el mes de uso del dispositivo, ¿cuántos días el usuario no lo ha llevado consigo?

He analizado para cada uno de los 6 usuarios la tendencia en la quema de calorías diarias durante los 31 días. Para ello, he transformado la variable ActivityDate en variables numéricas que indican con números enteros desde el día 1 hasta el 31, a esa nueva variable la he llamado "Día".

Todas las gráficas presentan altibajos, ninguna sigue un patrón constante. Este es un ejemplo, del usuario con ID 4558609924:

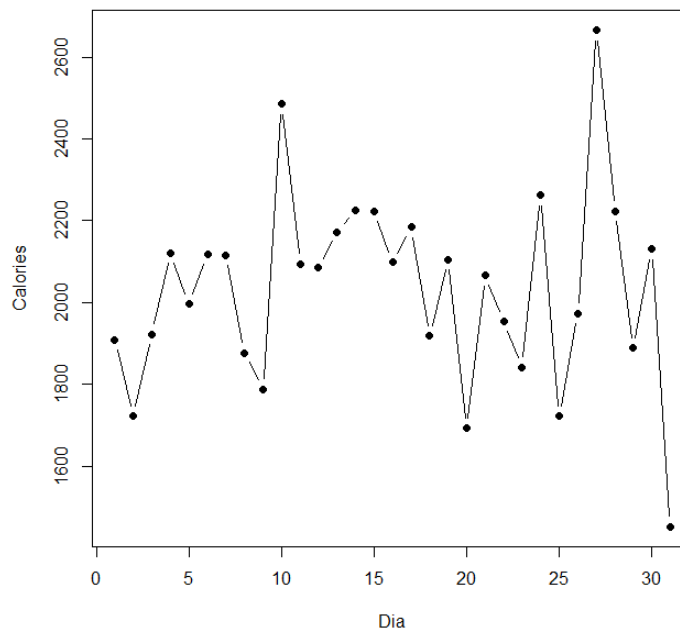


Figura 14. Gráfico de línea de variables Día y Calories

Al mirar los datos más de cerca, me doy cuenta de un punto importante:

En los conjuntos de datos, cuando hay días en los que el usuario no da ningún paso, se introduce la media de calorías que quema diariamente. Esto podría ser para no dar datos falsos o erróneos sobre la actividad del usuario cuando este se quita su dispositivo inteligente.

En el caso de este conjunto de datos, hay una persona que se quitó 1 día el reloj, otra persona que se lo quitó 14 días, y una última que no lo utilizó por 2 días. La persona que se lo quitó por 14 días coincide con la que padece de obesidad mórbida.

Esto podría indicar que las personas que padecen dicha enfermedad se rehúsan más a utilizar este tipo de dispositivos porque tienen hábitos de actividad y movimiento poco saludables.

5. Recomendaciones para Bellabeat

Recapitulamos las cuestiones que se planteaban al inicio de este proyecto:



- 1. ¿Cuáles son algunas tendencias de uso de los dispositivos inteligentes?
- 2. ¿Cómo se podrían aplicar estas tendencias a los clientes de Bellabeat?
- 3. ¿Cómo podrían ayudar estas tendencias a influir en la estrategia de marketing de Bellabeat?

En nuestro previo análisis, vimos mediante gráficos y resúmenes numéricos las distintas tendencias de uso de los dispositivos inteligentes de FitBit. Observamos como estos patrones de uso se distinguían según personas con o sin sobrepeso.

Ahora, ¿cómo aplicamos estos conocimientos a los clientes de Bellabeat?

Aunque los dispositivos de Bellabeat son dirigidos hacia mujeres, no hemos podido distinguir entre sexos en el análisis de Fitbit debido a la ausencia de esa variable. Por lo tanto, elaboraremos las recomendaciones atendiendo a los hallazgos encontrados en el previo análisis.

Sabemos que los clientes de Bellabeat cuentan con una aplicación conectada a sus dispositivos inteligentes donde pueden monitorear sus calorías, pasos, tiempo de sueño, etcétera. En definitiva, todas las variables tomadas en cuenta en el análisis. Por ello, es de vital importancia dicha aplicación a la hora de aplicar las tendencias de Fitbit a los usuarios de Bellabeat.

Teniendo en cuenta todo lo anterior, comenzaremos con las recomendaciones que Bellabeat podría implementar para mejorar sus productos y su estrategia de marketing:

1 Personalización según patrones de actividad y hábitos

- **Segmentación de usuarios:** Los datos sugieren que los usuarios tienen patrones de actividad muy variados, especialmente aquellos con sobrepeso que tienen periodos de actividad menos regulares. Bellabeat podría ofrecer recomendaciones personalizadas según los hábitos de cada usuario, adaptando los consejos según los niveles de actividad diaria, el tiempo de sueño o las calorías quemadas.
- **Notificaciones personalizadas:** Podrían desarrollar notificaciones automáticas que incentiven el movimiento en los usuarios que permanecen sedentarios por largos periodos, o recordatorios para mantener hábitos de sueño saludable para aquellos que no descansan lo suficiente.

2 Enfoque en la importancia del descanso y la salud integral

- **Incentivar buenos hábitos de sueño:** Vimos que dormir alrededor de 8 horas ayuda a mantenerse más activo durante el día. Bellabeat podría destacar el monitoreo del sueño y generar informes semanales o recomendaciones personalizadas sobre cómo mejorar la calidad del descanso.
- **Promociones y campañas sobre la importancia del descanso:** Basado en los datos, podrían hacer hincapié en la idea de que un buen descanso impacta positivamente en la energía y actividad física. Esto puede influir en la estrategia de marketing, fomentando productos como rastreadores de sueño.

3 Desarrollar funciones que alienten a usuarios con sobrepeso a ser más activos

- **Desafíos y recompensas:** A los usuarios con sobrepeso o con escasa actividad física, Bellabeat podría ofrecer desafíos dentro de la app que premien el aumento progresivo de su actividad. Por ejemplo, alcanzar un determinado número de pasos diarios o aumentar la cantidad de calorías quemadas podría traducirse en recompensas o logros virtuales dentro de la aplicación.
- **Incentivar la actividad moderada-intensa:** El análisis reveló que las personas con sobrepeso realizan menos actividad física intensa, pero mantienen algo de actividad moderada. Bellabeat podría lanzar guías o recomendaciones específicas para estos usuarios con sugerencias de ejercicios progresivos que se adapten a su nivel actual y los ayuden a aumentar su actividad de manera sostenible.

4 Adaptar la estrategia de marketing

- **Mensajes de empoderamiento para mujeres:** Como Bellabeat se enfoca en mujeres, la campaña de marketing podría destacar cómo la tecnología les permite tomar control de su salud y bienestar. Se puede llegar a oídos de mujeres mediante redes sociales como Instagram, en las que podrían hacerse publicaciones con frases motivadoras sobre la salud, con una estética femenina y con representación de cuerpos tanto con sobrepeso como con peso normal.
- **Historias de éxito:** Mostrar casos reales o simulados de mujeres que utilizan los dispositivos para mejorar su salud general, destacando no solo la pérdida de peso, sino también el aumento de energía y bienestar.
- **Colaboraciones con marcas de ropa deportiva plus-sized:** Colaborar con ese tipo de marcas daría visibilidad a Bellabeat y sus dispositivos y se centrarían en un público dispuesto a mejorar su salud.

5 Monitoreo de salud basado en el IMC y la actividad física

- **Monitoreo del IMC:** Como el análisis indica que más del 60% de los usuarios tienen sobrepeso, Bellabeat podría integrar funciones adicionales de seguimiento de IMC que alienten a las usuarias a mantener un peso saludable. Al vincular la actividad diaria con las calorías quemadas y los objetivos de IMC, la app podría proporcionar recomendaciones dietéticas y de ejercicio personalizadas para las usuarias con sobrepeso.
- **Alertas preventivas para salud física:** Se podrían incluir alertas o notificaciones si los datos de IMC o el peso corporal cruzan ciertos umbrales. Esto podría alentar a las usuarias a estar más atentas a su salud física.

6 Mejorar el uso continuo del dispositivo

- **Gamificación:** Los usuarios que usan el dispositivo de forma inconsistente (como el usuario con obesidad mórbida) podrían ser alentados a usar el dispositivo con más frecuencia a través de la gamificación. Por ejemplo, mantener una "racha" de días de uso podría desbloquear funciones o logros dentro de la app.
- **Sugerencias cuando no se usa el dispositivo:** Implementar una función que envíe notificaciones cuando el dispositivo no se ha utilizado en varios días, recordando a las usuarias lo útil que es para monitorear su bienestar y ofrecer estadísticas interesantes que motivan a su uso.

7 Integración de rutinas de ejercicios

- **Guías de actividad física personalizada:** Basado en los patrones de quema de calorías que identificamos en el análisis, Bellabeat podría proporcionar sugerencias diarias de ejercicios personalizados según los objetivos de la usuaria. Esto incluiría tanto ejercicios en casa como caminatas o rutinas intensas, adaptados a los datos personales de la clienta.
- **Videos de ejercicios en la app:** La creación de una biblioteca de ejercicios podría ser una característica adicional dentro de la aplicación, destacando rutinas adaptadas para diferentes niveles de fitness.

8 Enfoque en usuarios con IMC saludable

- **Nuevas metas de bienestar:** A las usuarias que ya tienen un IMC saludable, se les podría proponer nuevas metas, como mejorar la calidad del sueño, aumentar la resistencia física, o mantener un equilibrio entre la actividad física moderada e intensa.