

# Resumen Ejecutivo: Análisis de Datos de Fitbit y MINSA

Abel Sánchez

2024-06-29

## Introducción

Este informe sintetiza los análisis realizados sobre dos conjuntos de datos: los recogidos por dispositivos Fitbit y los proporcionados por el MINSA. Estos análisis apuntan a obtener insights que apoyen decisiones informadas y estratégicas en Bellabeat.

## Objetivo:

Documentar el proceso de análisis de datos aplicado al caso de Fitbit para Bellabeat, sirviendo como una guía práctica y educativa para estudiantes y profesionales del análisis de datos.

## Importancia del Análisis de Datos:

Es crucial para tomar decisiones informadas y mejorar estrategias de negocio. Este proyecto analiza datos de dispositivos Fitbit para extraer insights valiosos que pueden beneficiar a Bellabeat.

## Breve Descripción del Proyecto Fitbit:

Se enfoca en analizar datos de actividad física, frecuencia cardíaca y peso registrados por dispositivos Fitbit, con el objetivo de extraer insights y formular recomendaciones estratégicas para Bellabeat.

## Experiencia del Autor

Soy un profesional con más de 15 años de experiencia en contabilidad y finanzas, complementada con una formación en marketing y un máster en finanzas. He trabajado como contador y controller financiero en diversos sectores, el mayor tiempo fue en agencias de publicidad y centrales de medios globales; recientemente he completado la certificación de Google en análisis de datos. Esta mezcla de conocimientos y experiencia me ha permitido abordar el análisis de datos desde una perspectiva tanto técnica como de negocio.

## Problemas Identificados

La socia fundadora, Urška Sršen, estima que con datos sobre la actividad física de los usuarios de los dispositivos tecnológicos se podrían desarrollar nuevas estrategias de marketing para generar nuevas oportunidades de negocio para Bellabeat. La empresa cuenta con dispositivos (Leaf, Time, Spring) y una aplicación que recopilan datos sobre salud, enfocados en mujeres.

## **Estrategias para Resolver los Problemas**

Identificación de Información Clave: Recopilar datos sobre los parámetros medidos por los dispositivos (actividad física, sueño, estrés, ciclo menstrual, hábitos de conciencia plena).

### **Establecimiento de Tendencias:**

Analizar datos históricos para identificar patrones y comportamientos recurrentes entre los usuarios.

### **Análisis Comparativo:**

Se compararan los datos estadísticos de Fitbit con los datos estadísticos del Ministerio de Salud de España, así podremos corroborar si la información que proporciona la pulsera de Fitbit es relevante respecto a los datos de la encuesta sobre la percepción de la salud, la prevalencia de enfermedades y nivel de actividad física.

## **Preguntas para Resolver**

### **Tendencias de Uso:**

¿Cuáles son algunas tendencias de uso de los dispositivos inteligentes? Analizar datos de uso como la frecuencia de uso, horas pico de actividad, características demográficas de los usuarios activos.

### **Aplicación de Tendencias:**

¿Cómo se podrían aplicar estas tendencias a los clientes de Bellabeat?

Personalizar las recomendaciones y estrategias de marketing basadas en los patrones de uso y preferencias de las usuarias.

### **Influencia en Estrategia de Marketing:**

¿Cómo podrían ayudar estas tendencias a influir en la estrategia de marketing de Bellabeat? Desarrollar campañas de marketing dirigidas a segmentos específicos de usuarias, basadas en sus hábitos y necesidades identificadas.

## **Preparación y Ubicación de los Datos**

- 1.- Ubicación de los Datos: Carpeta: data/fitbit/raw
- 2.- Organización de los Datos: Formato Ancho/Largo
- 3.- Credibilidad de los Datos (ROCCC): Confiabilidad: Alta Originalidad: Original Integridad: Integral Actualidad: Actual Citación: Citados correctamente
- 4.- Autorización, Privacidad, Seguridad y Accesibilidad: Autorización: Permiso obtenido Privacidad: Datos anonimizados Seguridad: Encriptación aplicada Accesibilidad: Accesibles para análisis
- 5.- Verificación de la Integridad de los Datos: Datos verificados: Sin errores

- 6.- Relevancia de los Datos para Responder las Preguntas: Datos relevantes: Ayudan a responder las preguntas clave
- 7.- Problemas Identificados en los Datos: Problemas: Datos faltantes, inconsistencias

## **Procesamiento de Datos**

- 1.- Herramientas Elegidas y Por Qué: Se eligió R por su potencia y versatilidad en el procesamiento de datos, y por los paquetes como tidyverse que facilitan la manipulación y análisis de datos.
- 2.- Garantía de Integridad de los Datos: Mediante acceso controlado y técnicas de verificación y limpieza.
- 3.- Pasos para Garantizar que los Datos Están Limpios: Verificación y carga de archivos CSV. Inspección mediante visualización de primeras filas y resúmenes estadísticos. Conversión de columnas de fechas al formato correcto. Filtrado de datos en un rango de fechas específico. Selección de columnas relevantes. Unión de datasets. Identificación y gestión de valores faltantes y atípicos. Guardado de datos limpios en un nuevo archivo CSV.
- 4.- Verificación de que los Datos Están Limpios y Listos para Analizar: Inspección visual y análisis exploratorio. Resúmenes estadísticos y visualización de valores atípicos.
- 5.- Documentación del Proceso de Limpieza: El proceso se documentó en un archivo R Markdown (.Rmd) que contiene el código utilizado para la limpieza de datos, explicaciones y resultados intermedios.

## **Metodología**

Se utilizaron datos de Fitbit y del MINSA, basados en escenarios hipotéticos de septiembre de 2019, para analizar la actividad física y datos de salud de encuestas españolas centradas en mujeres durante 2017. Esto busca validar tendencias y apoyar futuras estrategias de marketing para Bellabeat.

## **Análisis y Resultados**

Datos de Fitbit Período Analizado: Marzo a Mayo de 2016. Análisis Realizados: Análisis estadísticos avanzados como correlaciones y regresiones.

## **Visualización Integrada de Datos**

### **Visualización por Segmentos**

#### **Matriz de Correlación**

- 1.- Correlaciones Moderadas: Entre el número de pasos y el gasto calórico.

### **Análisis Realizados: Análisis estadísticos avanzados como correlaciones.**

Los datos de las pulseras Fitbit fueron analizados para identificar correlaciones entre diferentes parámetros de actividad física y métricas de salud. Se crearon matrices de correlación para tres niveles de actividad: moderada, alta y baja.

- 1.- Actividad Moderada:

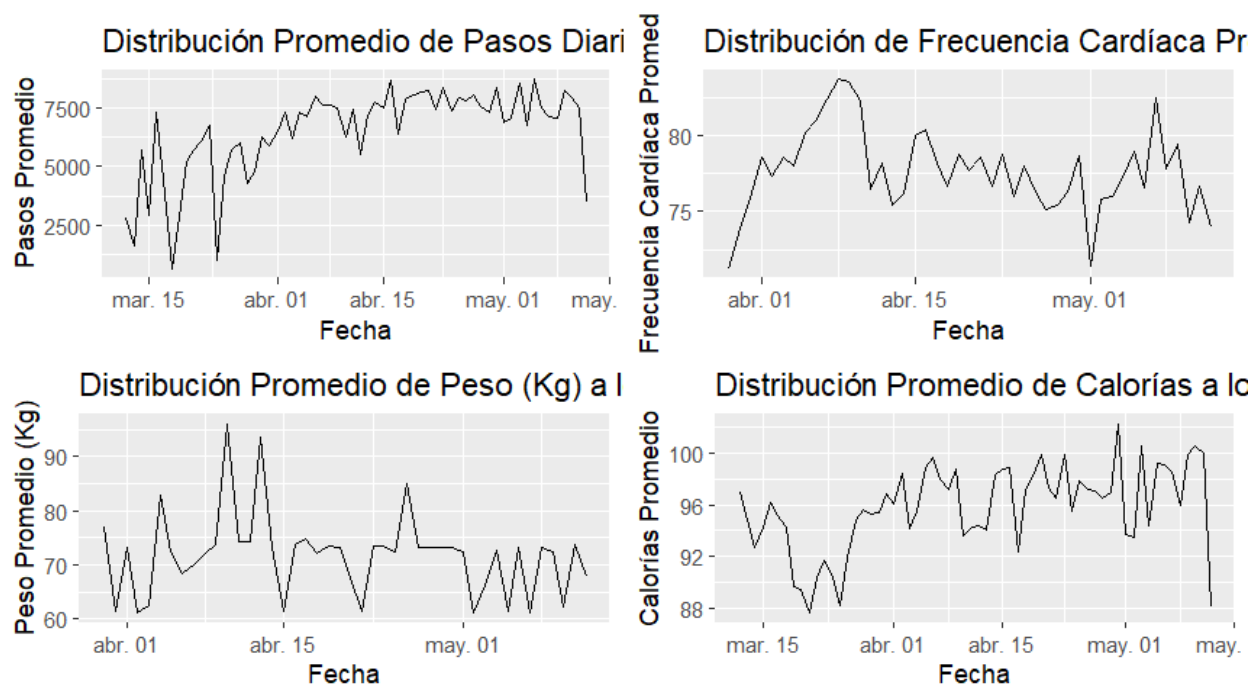


Figure 1: Visualización de los datos en promedios

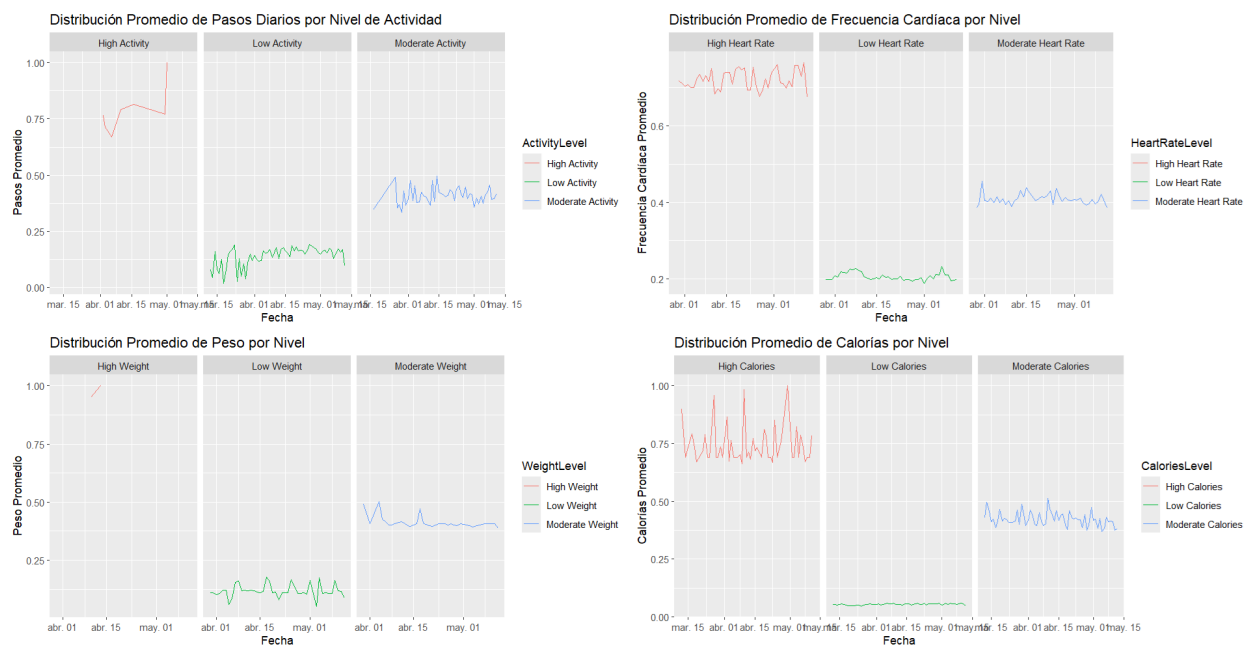


Figure 2: Visualización de los datos por segmentos

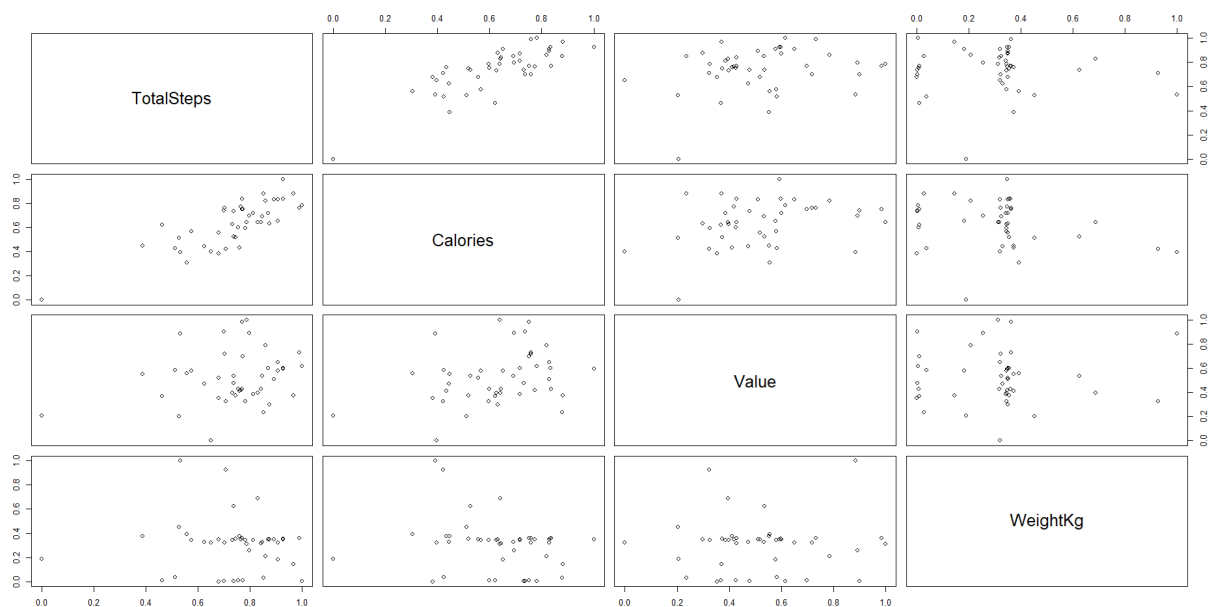


Figure 3: Matriz de correlación

TotalSteps y Calories: Correlación negativa (-0.2808). Esto sugiere que en niveles moderados de actividad, un aumento en los pasos totales podría no estar directamente asociado con un aumento en las calorías quemadas. TotalSteps y WeightKg: Correlación positiva (0.1181). Aunque débil, indica que a medida que aumentan los pasos, puede haber una ligera tendencia a la reducción de peso.

Show  entries

Search:

|            | TotalSteps          | Calories            | Value               | WeightKg            |
|------------|---------------------|---------------------|---------------------|---------------------|
| TotalSteps | 1                   | -0.2807578059135628 | -0.3272243383940743 | 0.1181105465658487  |
| Calories   | -0.2807578059135628 | 1                   | -0.1816402155201073 | -0.3995117926405294 |
| Value      | -0.3272243383940743 | -0.1816402155201073 | 1                   | 0.4526398177400004  |
| WeightKg   | 0.1181105465658487  | -0.3995117926405294 | 0.4526398177400004  | 1                   |

Showing 1 to 4 of 4 entries

Previous
Next

## 2.- Actividad Alta:

TotalSteps y Calories: Correlación positiva fuerte (0.5994). En niveles altos de actividad, hay una fuerte relación directa entre los pasos dados y las calorías quemadas, lo que respalda la idea de que la actividad física intensa contribuye significativamente a la quema de calorías. Calories y WeightKg: Correlación negativa (-0.2897). Esto implica que, con mayor quema de calorías, hay una tendencia a la pérdida de peso.

Show

10

▼

entries

Search:

|            | TotalSteps           | Calories            | Value               | WeightKg             |
|------------|----------------------|---------------------|---------------------|----------------------|
| TotalSteps | 1                    | 0.5994483818623226  | 0.05593782318084402 | -0.09474442365674993 |
| Calories   | 0.5994483818623226   | 1                   | 0.2781054010029978  | -0.2896692985407137  |
| Value      | 0.05593782318084402  | 0.2781054010029978  | 1                   | -0.1525206535498349  |
| WeightKg   | -0.09474442365674993 | -0.2896692985407137 | -0.1525206535498349 | 1                    |

Showing 1 to 4 of 4 entries

Previous

1

Next

Actividad Baja:

Todos los valores son NA, indicando que no hay datos suficientes o válidos para este nivel de actividad.

### Linea de regresión Fitbit

Modelos de Regresión: El número de pasos es un predictor significativo del gasto calórico.

## Resultados de Regresión de Actividades de Fitbit

Se realizaron análisis de regresión para evaluar la relación entre los pasos totales y las calorías quemadas.

1.- Actividad Moderada:

TotalSteps y Calories: La regresión no mostró una relación significativa (p-valor = 0.4643). Esto sugiere que en niveles moderados de actividad, los pasos totales no son un buen predictor de las calorías quemadas.

2.- Actividad Alta:

TotalSteps y Calories: Relación significativa (p-valor = 0.000179). A medida que aumentan los pasos en niveles de alta actividad, se predice un aumento significativo en las calorías quemadas. Este modelo explica el 35.93% de la variabilidad en las calorías quemadas.

3.- Actividad Baja:

Los resultados no son significativos debido a la falta de datos suficientes.





## Datos del MINSA

Análisis Realizados: Percepciones de salud y enfermedades crónicas, utilizando técnicas estadísticas y visuales.

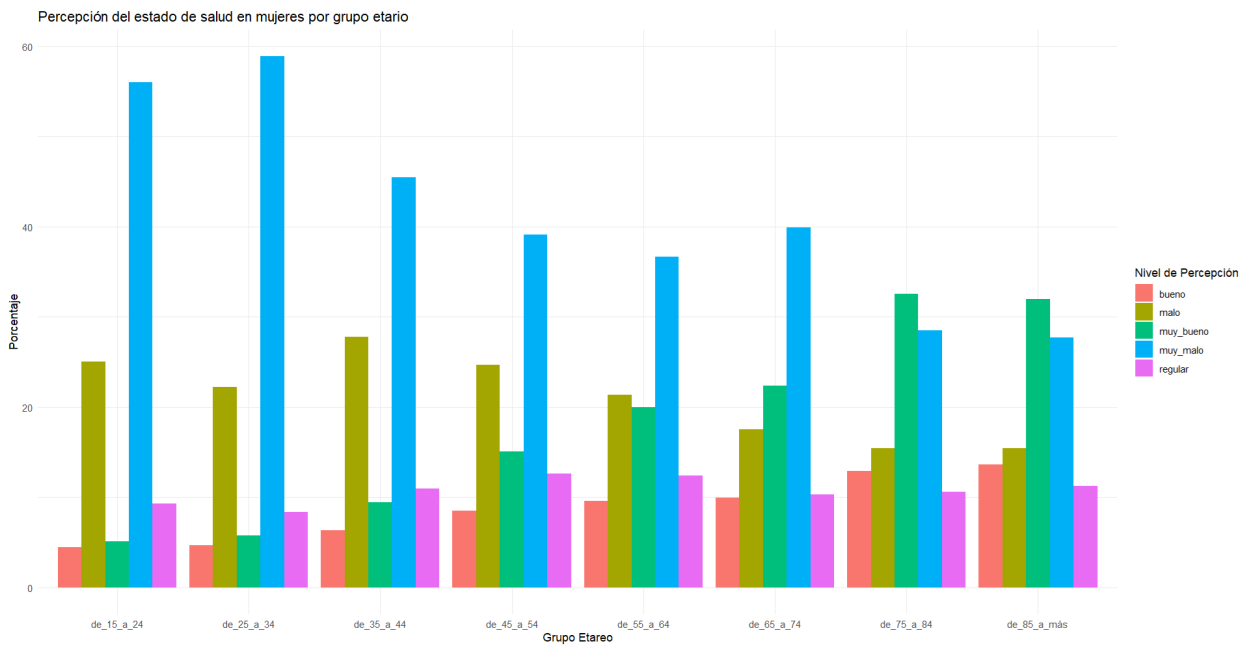
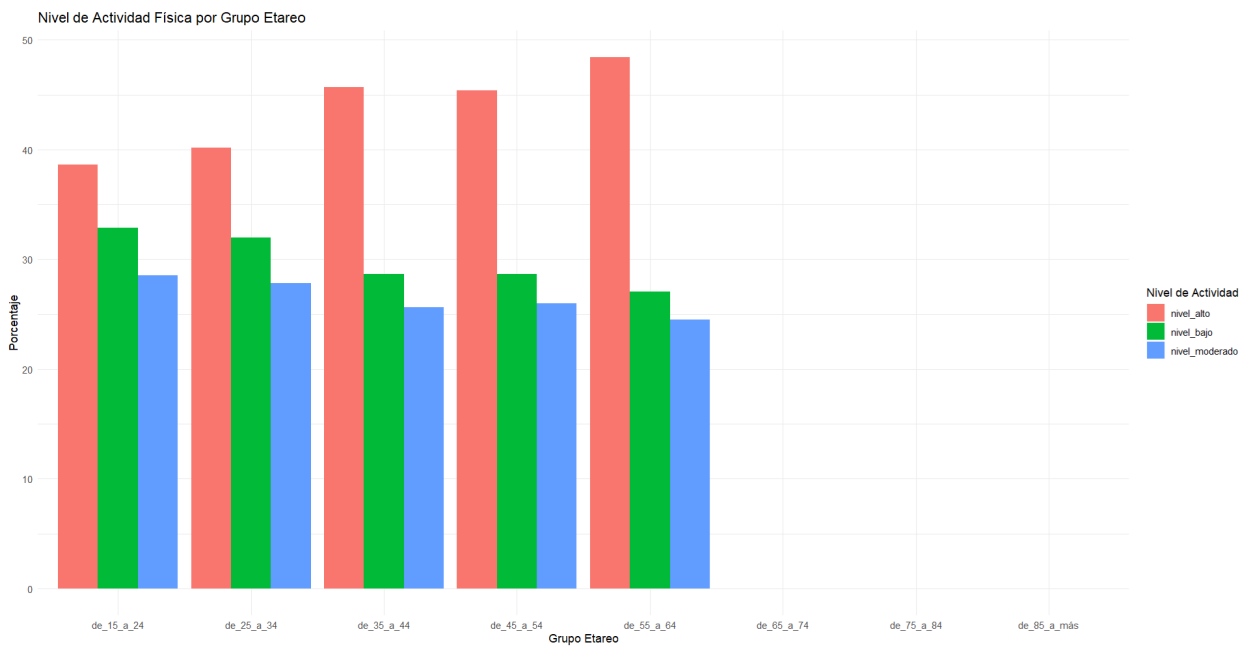


Figure 6: Nivel de percepción



### Análisis de Correlación y Regresión de MINSA Los datos del MINSA de 2017 proporcionan información sobre percepciones de salud y enfermedades crónicas entre las mujeres en España. Se realizó un análisis de correlación y regresión para evaluar la relación entre diferentes factores de salud.

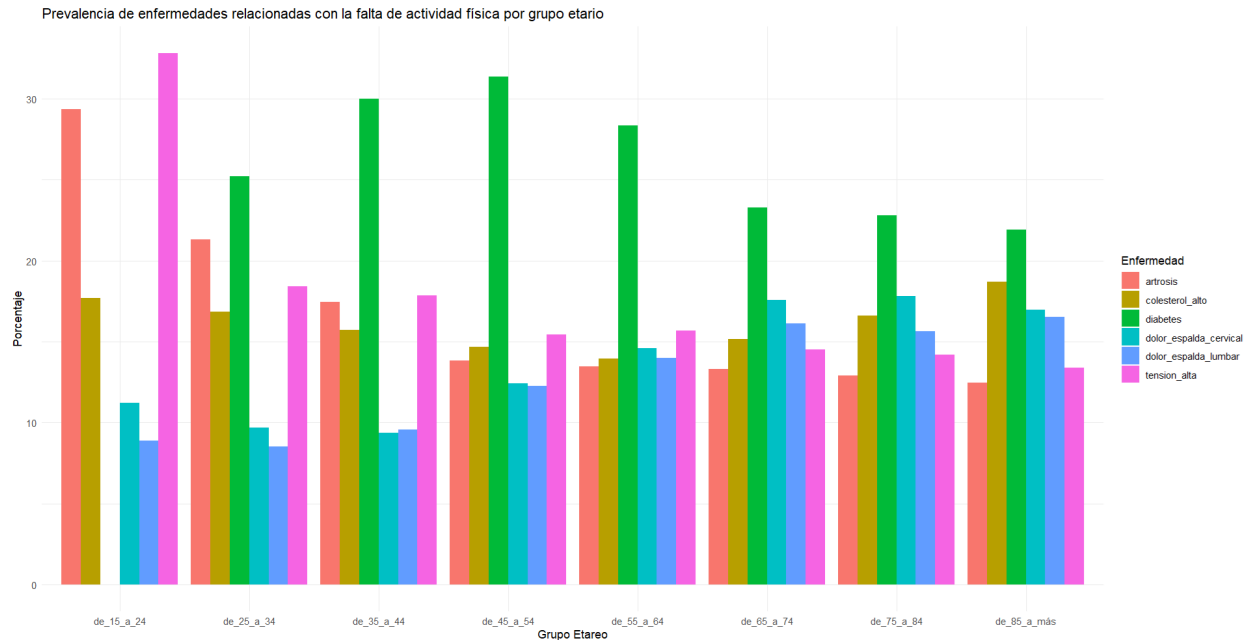


Figure 7: Prevalencia de enfermedades

## Matriz de Correlación de MINSA

La matriz de correlación de MINSA muestra las relaciones entre diversas percepciones de salud y condiciones crónicas.

muy\_bueno y bueno: Fuerte correlación positiva (0.8838), indicando que quienes perciben su salud como muy buena tienden también a evaluarla como buena en otras dimensiones. regular y malo: Fuerte correlación positiva (0.9432), sugiriendo que las percepciones de salud regular están altamente asociadas con evaluaciones negativas de salud. tension\_alta y artrosis: Fuerte correlación positiva (0.9930), indicando una alta prevalencia conjunta de tensión alta y artrosis.

## Resultados de Regresión de MINSA

El modelo de regresión examina la relación entre el nivel de actividad física y la percepción de salud, considerando múltiples factores.

Significancia: El modelo explica el 97.68% de la variabilidad en la percepción de salud basada en el nivel de actividad física y otros factores, aunque los p-valores indican que individualmente, las variables no son estadísticamente significativas, lo que puede deberse a la multicolinealidad o al tamaño reducido de la muestra.

## Conexión entre Fitbit y MINSA

Los datos de Fitbit y MINSA juntos proporcionan una visión comprensiva de cómo la actividad física influye en la salud general de las mujeres. Aunque los datos de Fitbit se centran en la actividad diaria y las métricas físicas inmediatas, los datos de MINSA refuerzan la importancia de la actividad física para la salud a largo plazo.

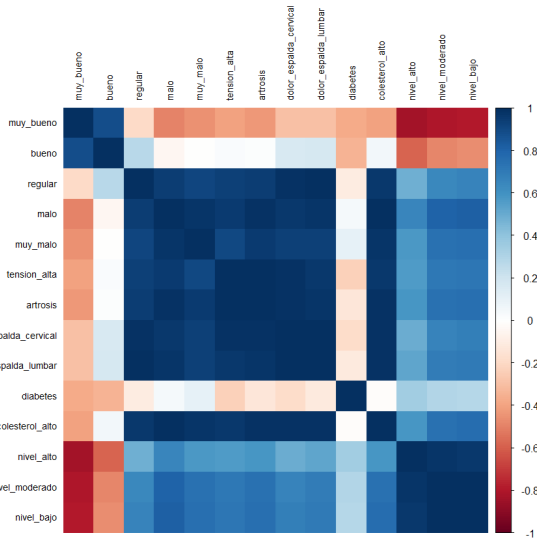


Figure 8: Matriz de correlación

Show 10 entries

|                        | muy_bueno            | bueno                 | regular             | malo                 | muy_malo              | tension_alta        | artrosis             | dolor_espalda_cervical | dolor_espalda_lumbar | diabetes             | colesterol_alto      | nivel_alto          | nivel_moderado      | nivel_bajo          |
|------------------------|----------------------|-----------------------|---------------------|----------------------|-----------------------|---------------------|----------------------|------------------------|----------------------|----------------------|----------------------|---------------------|---------------------|---------------------|
| muy_bueno              | 1                    | 0.88376086744153      | -0.199816534155389  | -0.495285761826112   | -0.4593862500229559   | -0.4013449705269221 | -0.41069742040771474 | -0.2900534256071955    | -0.2970793053413645  | -0.374403617183748   | -0.40385057594871    | -0.8327854221563669 | -0.8056709740031011 | -0.7998047070822708 |
| bueno                  | 0.88376086744153     | 1                     | 0.2739026413542194  | -0.04049415244731389 | -0.008463243080918162 | 0.02579556322322209 | 0.0110758892353109   | 0.1662315650560369     | 0.1707679980871026   | -0.345078567520182   | 0.0531031509639566   | -0.5861626785223673 | -0.4826395112789966 | -0.46828095611158   |
| regular                | -0.199816534155389   | 0.2739026413542194    | 1                   | 0.943242962038603    | 0.9172719256899754    | 0.9373829961874925  | 0.9487227159691878   | 0.981778675796035      | 0.9918524380712886   | -0.1092838926873284  | 0.9880441358737262   | 0.4811482769968756  | 0.6438788470970803  | 0.661655346083656   |
| malo                   | -0.495285761826112   | -0.04049415244731389  | 0.943242962038603   | 1                    | 0.9734889875487146    | 0.9578063166433906  | 0.9823036061324022   | 0.9614018665973559     | 0.9710527789918756   | 0.049482214999437732 | 0.9940739319316134   | 0.6574972200095781  | 0.8020214535983264  | 0.8175414280355777  |
| muy_malo               | -0.4593862500229559  | -0.008463243080918162 | 0.9172719256899754  | 0.9734889875487146   | 1                     | 0.9080659979184536  | 0.9508882255736881   | 0.9353997629138567     | 0.9368150876294418   | 0.106771188768484    | 0.9701279136638142   | 0.5785531783462574  | 0.7402273585439436  | 0.758855408098429   |
| tension_alta           | -0.4013449705269221  | 0.02579556322322209   | 0.9373829961874925  | 0.9578063166433906   | 0.9080659979184536    | 1                   | 0.9929890801453553   | 0.9826458801232935     | 0.969783999156146    | -0.2349419315014436  | 0.9685953053152696   | 0.562069466470813   | 0.7133073948484866  | 0.7289148646214367  |
| artrosis               | -0.41069742040771474 | 0.0110758892353109    | 0.9487227159691878  | 0.9823036061324022   | 0.9508882255736881    | 0.9929890801453553  | 1                    | 0.9864783091447882     | 0.9786966928298035   | -0.33218511111176831 | 0.986885409874361    | 0.5857088178168922  | 0.7411149603313708  | 0.7575915662047451  |
| dolor_espalda_cervical | -0.2900534256071955  | 0.1662315650560369    | 0.981778675796035   | 0.9614018665973559   | 0.9353997629138567    | 0.9826458801232935  | 0.9864783091447882   | 1                      | 0.9936757516857299   | -0.1884857501790949  | 0.9822714356204043   | 0.4976621477481484  | 0.66525865588742    | 0.6835105941438799  |
| dolor_espalda_lumbar   | -0.2970793053413645  | 0.1707679980871026    | 0.9918524380712886  | 0.9710527789918756   | 0.9368150876294418    | 0.969783999156146   | 0.9786966928298035   | 0.9936757516857299     | 1                    | -0.1127850054966388  | 0.9879756658379935   | 0.5290754009196882  | 0.6900261551827538  | 0.7072318767335318  |
| diabetes               | -0.374403617183748   | -0.345078567520182    | -0.1092838926873284 | 0.04948221499943732  | 0.106771188768484     | -0.2349419315014436 | -0.3321851111176831  | -0.1884857501790949    | -0.1127850054966388  | 1                    | -0.01649760340464969 | 0.3470377262212834  | 0.2932539928815224  | 0.2891852918312108  |

Showing 1 to 10 of 14 entries

Previous 1 2 Next

Figure 9: Tabla de matriz de correlación minsa

1.-Correlación entre Actividad y Salud: Ambas fuentes de datos muestran que la actividad física tiene una relación directa con diversas métricas de salud. Los datos de Fitbit sugieren que niveles altos de actividad están fuertemente correlacionados con una mayor quema de calorías, lo que es crucial para la gestión del peso. Los datos de MINSA, por otro lado, muestran que percepciones positivas de salud y menor prevalencia de enfermedades crónicas están asociadas con niveles más altos de actividad física.

2.- Relevancia de las Encuestas de Salud: Los datos de la encuesta de salud del MINSA de 2017 validan la importancia de monitorear y fomentar la actividad física, especialmente entre las mujeres. Las correlaciones encontradas entre la percepción de buena salud y menor prevalencia de condiciones como la tensión alta y la artrosis sugieren que la actividad física puede ser un factor clave para mejorar la salud y reducir el riesgo de enfermedades crónicas.

3.- En conclusión, el análisis integrado de los datos de Fitbit y MINSA subraya la necesidad de promover la actividad física regular como una estrategia fundamental para mejorar la salud de las mujeres. La recopilación continua de datos a través de dispositivos como Fitbit, junto con encuestas de salud, puede proporcionar insights valiosos para el desarrollo de políticas de salud pública y estrategias de marketing de salud.

## **Conclusiones y Recomendaciones**

### **Fitbit**

Conclusión: Relación significativa entre la cantidad de pasos y el gasto calórico, sugiriendo que el seguimiento de la actividad física puede promover la salud.

Recomendación: Bellabeat podría desarrollar campañas para fomentar el uso de dispositivos de seguimiento de actividad física y proporcionar feedback personalizado basado en los datos de actividad de los usuarios.

### **MINSA**

Conclusión: La importancia de monitorear continuamente los indicadores de salud para identificar tendencias y focalizar las intervenciones de salud pública.

Recomendación: Bellabeat podría colaborar con autoridades de salud pública para compartir datos y desarrollar programas conjuntos que aborden las necesidades específicas de la población, mejorando los resultados de salud a través de políticas basadas en datos.

## **Implementación de Estrategias**

Integración de Dispositivos y Datos: Bellabeat podría integrar dispositivos de seguimiento de actividad física con una plataforma que también recoja datos de salud pública, permitiendo un análisis más completo y personalizado. Campañas de Salud y Bienestar: Desarrollar campañas que utilicen los insights obtenidos de los análisis de Fitbit y MINSA para promover hábitos saludables, con recomendaciones personalizadas y programas de ejercicio.

## **Consideraciones y Limitaciones**

## **Uso de Herramientas y Soporte Tecnológico**

Este informe se desarrolló con el apoyo de diversas herramientas y plataformas tecnológicas. Mi experiencia en contabilidad, finanzas, marketing y análisis de datos, complementada con la asistencia de inteligencia artificial (IA), fue crucial para el desarrollo y perfeccionamiento del código utilizado en este análisis.

## Soporte de la Inteligencia Artificial

A lo largo del desarrollo de este proyecto, utilicé una plataforma de IA para obtener asistencia en la escritura y depuración del código en R. Esta herramienta me permitió profundizar en el uso de R, más allá de lo aprendido en la certificación de Google.

## Transparencia y Ética Profesional

A pesar de la asistencia de IA, todos los análisis y conclusiones son del autor. La IA fue una herramienta de apoyo, no un sustituto del juicio profesional.

## Limitaciones del Análisis

Datos Utilizados: Limitaciones inherentes a los datos de Fitbit y MINSA. Alcance del Estudio: Enfocado en datos específicos de actividad física y salud. Herramientas y Técnicas: La interpretación de resultados está sujeta a la calidad de los datos y los modelos utilizados.

## Reflexión Final

Este proyecto ha sido una oportunidad para combinar la experiencia en análisis de datos con el aprendizaje de nuevas herramientas tecnológicas. Compromiso con principios éticos y uso responsable de la tecnología.

```
summary(cars)
```

```
##      speed      dist
##  Min.   : 4.0    Min.   : 2.00
## 1st Qu.:12.0    1st Qu.: 26.00
##  Median :15.0    Median : 36.00
##   Mean  :15.4    Mean   : 42.98
## 3rd Qu.:19.0    3rd Qu.: 56.00
##   Max.  :25.0    Max.    :120.00
```