

*Análisis de Datos sobre el Seguimiento de Ejercicio de Miembros en un Gimnasio para Mejorar el Rendimiento y la Experiencia del Usuario*

**Análisis de Datos sobre el Seguimiento de Ejercicio de Miembros en un Gimnasio para  
Mejorar el Rendimiento y la Experiencia del Usuario**

Integrantes

Andrey Sánchez Zúñiga

David Tencio

Colegio Universitario de Cartago

Minería de Datos BD -123 Grupo 01

Profesor

Oswaldo González Chávez

Abril, 2025

## **CONTENIDO**

CAPÍTULO I. PLANTEAMIENTO .....	3
Introducción .....	3
Justificación .....	5
Pregunta de investigación .....	7
Objetivo general.....	8
Objetivos específicos .....	8
MARCO TEÓRICO.....	9
MARCO METODOLÓGICO.....	10
Diseño de la Investigación .....	11
Población y Muestra .....	11
Técnicas e Instrumentos de Recolección de Datos .....	12
Procedimientos Metodológicos.....	12
Anexos .....	13
Conclusiones .....	18

## **CAPÍTULO I. PLANTEAMIENTO**

### **Introducción**

Este documento tiene como objetivo principal llevar a cabo un análisis exhaustivo de los datos de los miembros de un gimnasio con el fin de identificar patrones, tendencias y relaciones significativas que puedan ser utilizadas para tomar decisiones estratégicas bien fundamentadas. En particular, el análisis se enfoca en desentrañar el comportamiento de los usuarios en relación con sus rutinas de ejercicio, así como sus preferencias y motivaciones en cuanto a los tipos de entrenamientos que eligen. Además, busca explorar cómo diversas variables, tanto individuales como contextuales, influyen en los resultados que los miembros experimentan a lo largo de su participación en el gimnasio. Al comprender estos aspectos, se podrá optimizar la operación del gimnasio, personalizar la oferta de servicios y, en última instancia, mejorar la experiencia general de los miembros.

El conjunto de datos utilizado en este análisis proviene de una base de datos pública disponible en Kaggle, que incluye información detallada sobre los miembros del gimnasio, como datos demográficos, métricas de salud y bienestar, tipos de entrenamientos realizados y los resultados obtenidos de cada sesión de ejercicio. Este conjunto de datos proporciona una visión integral del comportamiento de los usuarios, lo que permite identificar patrones de uso que no siempre son evidentes a simple vista. La recopilación de información acerca de la frecuencia y duración de las actividades, la intensidad de los entrenamientos y el progreso en los objetivos personales de los miembros es clave para entender cómo estos factores se interrelacionan y afectan los resultados a

## *Análisis de Datos sobre el Seguimiento de Ejercicio de Miembros en un Gimnasio para Mejorar el Rendimiento y la Experiencia del Usuario*

largo plazo. Este análisis no solo sirve para obtener una visión clara de los comportamientos actuales, sino también para prever tendencias y patrones que pueden ser utilizados para mejorar la oferta de servicios del gimnasio en el futuro.

El análisis se adentrará en la identificación de correlaciones entre las diferentes variables recogidas, tales como la relación entre la frecuencia de ejercicio y la mejora en los indicadores de salud, o cómo las características demográficas (como la edad, el género o el nivel de condición física) impactan en las preferencias de los usuarios por ciertos tipos de entrenamientos. Asimismo, se explorarán otras variables que pueden influir en el compromiso de los miembros con el gimnasio, como la motivación, las metas personales, y la interacción con los entrenadores. La identificación de estas correlaciones permitirá personalizar de manera más efectiva los programas de entrenamiento, adaptándolos a las necesidades y características individuales de los miembros. Esto no solo mejorará los resultados de los usuarios, sino que también aumentará su satisfacción y los motivará a seguir asistiendo al gimnasio de manera constante.

Otro aspecto fundamental de este análisis será la optimización de los servicios ofrecidos por el gimnasio. Al obtener una comprensión más profunda de los intereses y necesidades de los miembros, se podrán ajustar los programas de entrenamiento para maximizar la efectividad de las sesiones. Esto incluye la posibilidad de diseñar nuevos tipos de entrenamientos, introducir herramientas adicionales para el seguimiento de progreso o incluso personalizar los horarios y la accesibilidad de las clases. Además, se podrá ofrecer un nivel de atención más personalizado, basándose en las preferencias individuales de los usuarios, lo que contribuirá a una experiencia más satisfactoria y a la fidelización de los miembros. Esta personalización no solo mejorará los

resultados de los usuarios, sino que también proporcionará una ventaja competitiva significativa para el gimnasio frente a otros establecimientos de la misma industria.

Finalmente, los resultados derivados de este análisis permitirán identificar nuevas oportunidades de negocio que puedan aprovecharse para seguir creciendo y diversificando los servicios del gimnasio. Por ejemplo, podrían surgir propuestas para introducir programas de entrenamiento exclusivos, crear alianzas con marcas de productos de salud o bienestar, o desarrollar estrategias de marketing más segmentadas y eficaces. Además, la información obtenida sobre el comportamiento de los usuarios podría contribuir a una mejor toma de decisiones en cuanto a la gestión de recursos, la planificación de eventos especiales o la optimización de las estrategias de retención de clientes. En resumen, este análisis no solo tiene el propósito de mejorar el rendimiento de los miembros del gimnasio, sino también de generar un impacto positivo en la rentabilidad y sostenibilidad del negocio a largo plazo.

## **Justificación**

El análisis de datos sobre el comportamiento de los miembros de un gimnasio es fundamental en el contexto actual de la industria del bienestar y la salud, en el que la personalización y la eficiencia en la oferta de servicios son factores clave para la competitividad y el éxito de cualquier establecimiento. En un mercado cada vez más saturado y enfocado en la experiencia del cliente, es esencial comprender en profundidad las necesidades, preferencias y patrones de los usuarios para poder adaptar los servicios de manera efectiva. Los gimnasios que logran ofrecer programas de entrenamiento personalizados y optimizados de acuerdo a los datos de

los miembros tienen una ventaja significativa, ya que logran no solo mejorar los resultados de los usuarios, sino también fomentar la fidelidad y el compromiso a largo plazo.

El análisis de los datos de ejercicio y comportamiento de los miembros permite a los gimnasios identificar patrones de uso que pueden no ser evidentes a simple vista. Por ejemplo, al correlacionar variables demográficas como la edad, el género o el nivel de condición física con los tipos de entrenamiento más frecuentes, se pueden descubrir tendencias clave que ayudan a ajustar la oferta de servicios. Además, se puede identificar la relación entre la frecuencia e intensidad del ejercicio y las mejoras en los indicadores de salud, lo que permite personalizar los programas de entrenamiento y maximizar los beneficios para los miembros. Esto no solo contribuye a la satisfacción del cliente, sino que también mejora la eficiencia operativa del gimnasio al garantizar que los recursos se utilicen de manera óptima.

Asimismo, la justificación de este análisis radica en la necesidad de adaptarse a un entorno de negocio en constante evolución. Con el aumento de la digitalización y el acceso a tecnologías avanzadas, los gimnasios tienen ahora la oportunidad de recopilar y analizar grandes cantidades de datos sobre sus usuarios. Este proceso no solo optimiza la experiencia de los miembros, sino que también abre la puerta a nuevas oportunidades de negocio, como la creación de servicios adicionales, programas de entrenamiento exclusivos o la implementación de estrategias de marketing más segmentadas. A través de este análisis, los gimnasios pueden obtener una ventaja competitiva basada en datos precisos y actualizados, lo que les permite ofrecer servicios más alineados con las expectativas y necesidades de los usuarios, asegurando así su crecimiento y sostenibilidad a largo plazo.

En resumen, la justificación de este análisis de datos no solo radica en la mejora de la experiencia del usuario y la personalización de los programas de entrenamiento, sino también en la capacidad de transformar los datos en herramientas valiosas para la toma de decisiones estratégicas. Este enfoque basado en datos permitirá a los gimnasios adaptarse a los cambios en las preferencias de los usuarios, optimizar sus operaciones y, al mismo tiempo, identificar nuevas oportunidades para expandir su negocio en un mercado altamente competitivo.

### **Pregunta de investigación**

**Pregunta de Investigación 1:** ¿Existe una relación significativa entre la duración del entrenamiento y las calorías quemadas por los miembros del gimnasio?

**Hipótesis 1:** La duración del entrenamiento está positivamente correlacionada con las calorías quemadas, siendo los entrenamientos más largos aquellos que generan un mayor gasto calórico.

¿El tipo de ejercicio realizado por los miembros del gimnasio está relacionado con la frecuencia con la que asisten al gimnasio y su consistencia en los entrenamientos?

**Hipótesis 2:** Los miembros que participan en entrenamientos grupales o de alto impacto tienen una mayor frecuencia de asistencia y consistencia en sus entrenamientos en comparación con aquellos que optan por ejercicios de bajo impacto o entrenamientos individuales.

¿Existe una relación significativa entre el Índice de Masa Corporal (IMC) de los miembros y la frecuencia de sus entrenamientos en el gimnasio?

**Hipótesis 3:** Los miembros con un IMC más alto tienden a asistir con mayor frecuencia al gimnasio en comparación con aquellos con un IMC más bajo, debido a su mayor motivación para mejorar su salud y reducir su peso.

## **Objetivo general**

Analizar los datos de los miembros de un gimnasio para identificar patrones, tendencias y correlaciones clave, que permitan optimizar las operaciones, mejorar la experiencia del usuario, y aplicar insights estratégicos para el crecimiento y la sostenibilidad del negocio, de manera similar a cómo se gestionan las expansiones y operaciones de restaurantes en el sector gastronómico.

## **Objetivos específicos**

Analizar la relación entre la duración del entrenamiento, el tipo de ejercicio realizado, las características demográficas de los miembros y las calorías quemadas, con el fin de personalizar las rutinas de ejercicio y optimizar los resultados de salud de los usuarios.

Evaluar el impacto de las motivaciones personales, las características del gimnasio (entrenadores, equipos y clases grupales) y la frecuencia de asistencia al gimnasio sobre la retención de los miembros, con el fin de mejorar la oferta de servicios y estrategias de fidelización.



Generar recomendaciones basadas en los datos analizados para mejorar la gestión operativa del gimnasio, optimizando recursos, planificación de horarios y estructura de clases, para garantizar una operación más eficiente y rentable.

## **MARCO TEÓRICO**

El auge del análisis de datos en el ámbito de la salud y el bienestar ha transformado la manera en que se gestionan los servicios en gimnasios y centros deportivos. El análisis de grandes volúmenes de información, conocido como Big Data, permite a las organizaciones entender mejor el comportamiento de sus usuarios y tomar decisiones estratégicas basadas en evidencia. Este enfoque se ha vuelto esencial en un contexto donde la personalización de servicios y la optimización de recursos son clave para la fidelización y la eficiencia operativa.

Uno de los pilares en este tipo de análisis es la aplicación de **estadística descriptiva** y **estadística inferencial** para identificar patrones y relaciones entre variables. Herramientas como la media, mediana, moda y desviación estándar permiten comprender la distribución de las variables observadas, mientras que técnicas como la regresión lineal y el análisis de correlación permiten examinar cómo se relacionan entre sí variables como la duración de los entrenamientos y las calorías quemadas.

En el contexto del entrenamiento físico, variables fisiológicas como el Índice de Masa Corporal (IMC), el ritmo cardíaco promedio (Avg\_BPM) y el porcentaje de grasa corporal son indicadores críticos de salud que pueden verse modificados por la frecuencia, intensidad y tipo de ejercicio realizado. Diversos estudios respaldan que existe una correlación positiva entre la duración del entrenamiento y el gasto calórico, así como entre la intensidad del ejercicio y la mejora en la condición física general del individuo.

Además, aspectos motivacionales y sociales, como la participación en entrenamientos grupales o el acompañamiento de entrenadores personales, también inciden directamente en la frecuencia de

asistencia y en la consistencia del entrenamiento. Según teorías del comportamiento como la Teoría de la Autodeterminación (Deci y Ryan, 2000), los individuos que encuentran motivación intrínseca en su práctica deportiva tienden a mantenerla en el tiempo, especialmente cuando perciben progresos medibles o logran conectar socialmente con el entorno del gimnasio.

La implementación de modelos de minería de datos como el clustering (agrupamiento) permite segmentar a los usuarios en grupos con características similares, lo que posibilita diseñar programas de entrenamiento personalizados, ajustados a variables demográficas, nivel de experiencia o metas específicas. De igual forma, el análisis de valores atípicos (outliers) puede señalar casos que requieren atención especial, como individuos con obesidad severa o que presentan un rendimiento significativamente diferente al promedio.

La personalización del entrenamiento mediante el análisis de datos no solo mejora los resultados individuales, sino que también representa una ventaja competitiva para los gimnasios, ya que permite la planificación estratégica de clases, la optimización de recursos y la mejora continua del servicio. Tal como se ha planteado en el presente trabajo, el conocimiento generado a través del análisis de datos facilita la toma de decisiones basadas en evidencia, promueve la eficiencia operativa y genera oportunidades de crecimiento y diversificación en un mercado altamente competitivo.

## **MARCO METODOLÓGICO**

La presente investigación se desarrolla bajo un enfoque cuantitativo, que permite medir y analizar numéricamente el comportamiento de los miembros de un gimnasio a partir de variables específicas extraídas de una base de datos. Este enfoque posibilita una observación objetiva y sistemática de la realidad mediante la recolección y el análisis estadístico de datos, con el propósito

de identificar patrones de comportamiento, relaciones significativas entre variables y proponer estrategias de mejora fundamentadas en evidencia empírica.

Asimismo, se enmarca dentro del tipo de estudio descriptivo-correlacional, ya que no solo se describen las características de los usuarios y sus hábitos de ejercicio, sino que también se analizan las posibles relaciones entre variables como la duración del entrenamiento, las calorías quemadas, el tipo de ejercicio, la frecuencia de asistencia y el Índice de Masa Corporal (IMC). El objetivo no es establecer causalidad, sino encontrar correlaciones y asociaciones que puedan ser utilizadas para mejorar la toma de decisiones estratégicas en la gestión de gimnasios.

### **Diseño de la Investigación**

El diseño de la investigación es de tipo no experimental y transversal. No se manipulan las variables ni se realiza intervención directa sobre los sujetos de estudio, sino que se observa el comportamiento de los mismos a partir de un conjunto de datos ya existente. El diseño es transversal porque el análisis se lleva a cabo en un único momento del tiempo, utilizando una base de datos recolectada previamente, lo que permite obtener una "fotografía" del comportamiento de los miembros del gimnasio en relación con sus rutinas y resultados.

### **Población y Muestra**

La población objetivo está compuesta por usuarios de gimnasios que han sido registrados en una base de datos bajo diferentes criterios de salud, comportamiento y rendimiento. Para esta investigación se trabajó con una muestra no probabilística compuesta por 973 observaciones disponibles en el dataset titulado “gym\_members\_exercise\_tracking”, obtenido de la plataforma Kaggle. Esta muestra incluye un total de 15 variables, tanto numéricas como categóricas, que

proporcionan información detallada sobre el perfil demográfico, fisiológico y conductual de los miembros del gimnasio.

Entre las variables consideradas se encuentran: edad, género, peso, estatura, frecuencia cardíaca (máxima, promedio y en reposo), duración de la sesión de entrenamiento, calorías quemadas, tipo de ejercicio, porcentaje de grasa corporal, consumo de agua diaria, frecuencia de entrenamiento semanal, nivel de experiencia e Índice de Masa Corporal (BMI).

### **Técnicas e Instrumentos de Recolección de Datos**

La recolección de datos fue realizada a través de fuentes secundarias, específicamente de un conjunto de datos estructurados y públicos disponibles en línea, lo cual facilita su reutilización para fines analíticos y académicos. Al tratarse de un dataset ya existente, no fue necesaria la implementación de encuestas, entrevistas o formularios. La base de datos se descargó en formato CSV, lo que permitió su procesamiento mediante herramientas de análisis estadístico y científico de datos como Excel, R y Python (con bibliotecas como pandas, matplotlib, seaborn y scikit-learn).

### **Procedimientos Metodológicos**

El análisis de los datos se llevó a cabo a través de una serie de etapas estructuradas, detalladas a continuación:

1. Limpieza de Datos: Se revisó la base de datos para eliminar registros incompletos, duplicados o inconsistentes. Posteriormente, se recodificaron variables categóricas en valores numéricos para facilitar su análisis. Por ejemplo, la variable “Gender” se transformó de texto a valores numéricos: Female=1, Male=2. Similarmente, “Workout\_Type” fue transformado en un rango de valores (Cardio=1, HIIT=2, Strength=3, Yoga=4).
2. Análisis estadístico descriptivo: Se calcularon medidas de tendencia central (media, mediana, moda) y de dispersión (rango, cuartiles, mínimo, máximo) para entender la

distribución de cada variable. Esto permitió observar, por ejemplo, que la edad promedio de los usuarios es de 38 años, y que la duración promedio de una sesión de entrenamiento es de aproximadamente 1 hora y 15 minutos.

3. Identificación de valores atípicos: Se utilizaron gráficos tipo Box Plot para identificar valores extremos en variables clave como el IMC y las calorías quemadas. Estos valores son importantes para reconocer perfiles de usuarios con características inusuales que podrían requerir programas de entrenamiento especiales o atención médica personalizada.
4. Análisis de correlación: Se realizaron análisis de correlación (coeficiente de Pearson) entre variables como duración del entrenamiento y calorías quemadas, y entre el IMC y la frecuencia de entrenamiento. Los resultados permitieron confirmar la hipótesis de que existe una relación positiva entre la duración del entrenamiento y el gasto calórico.
5. Regresión lineal simple, utilizada para evaluar la relación entre duración del entrenamiento (variable independiente) y calorías quemadas (variable dependiente). El modelo evidenció una pendiente positiva, lo cual indica que, en promedio, a mayor duración de entrenamiento, mayor es el número de calorías quemadas.
6. Clustering (agrupamiento), técnica no supervisada utilizada para segmentar a los usuarios en grupos con características similares. Esta clasificación puede ser útil para definir perfiles de entrenamiento y adaptar los servicios a las necesidades de cada grupo.

## **Anexos**

Este anexo presenta las estadísticas descriptivas fundamentales para el conjunto de datos gym\_cuc. Se incluyen los valores de media, mediana, moda, mínimo, máximo, primer cuartil (Q1) y tercer cuartil (Q3) para cada una de las 15 columnas del dataset. Estas estadísticas permiten entender mejor la distribución y características clave de los datos de los miembros del gimnasio.

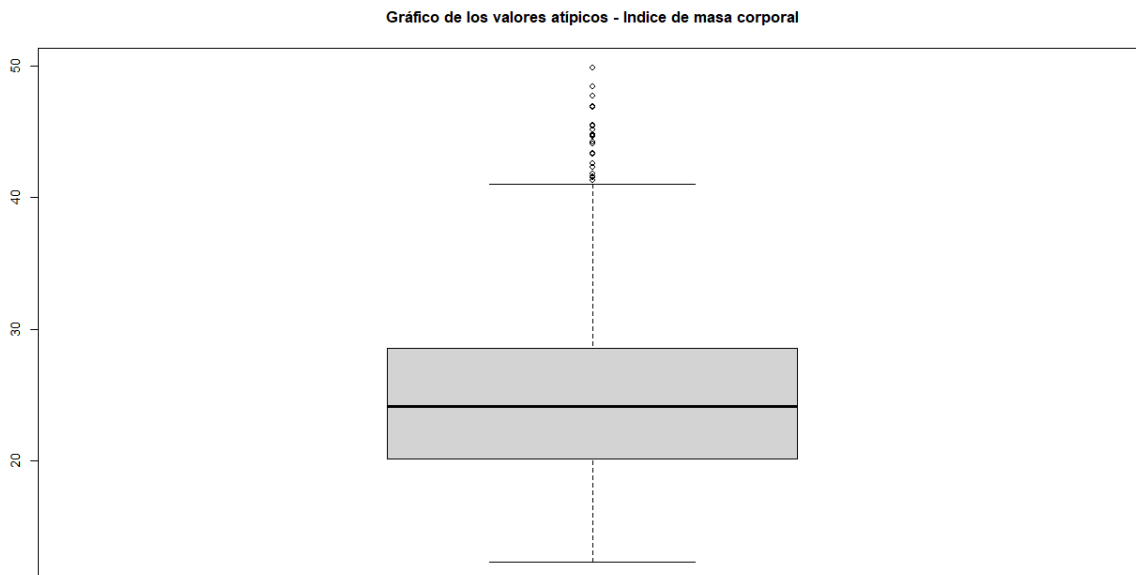
## *Análisis de Datos sobre el Seguimiento de Ejercicio de Miembros en un Gimnasio para Mejorar el Rendimiento y la Experiencia del Usuario*

```
> summary(gym_cuc)
      Age      Gender  Weight..kg.  Height..m.  Max_BPM  Avg_BPM  Resting_BPM  Session_Duration..hours.
Min.   :18.00  Min.   :1.000  Min.   : 40.00  Min.   :1.500  Min.   :160.0  Min.   :120.0  Min.   :50.00  Min.   :0.500
1st Qu.:28.00  1st Qu.:1.000  1st Qu.: 58.10  1st Qu.:1.620  1st Qu.:170.0  1st Qu.:131.0  1st Qu.:56.00  1st Qu.:1.040
Median :40.00  Median :2.000  Median : 70.00  Median :1.710  Median :180.0  Median :143.0  Median :62.00  Median :1.260
Mean   :38.68  Mean   :1.525  Mean   : 73.85  Mean   :1.723  Mean   :179.9  Mean   :143.8  Mean   :62.22  Mean   :1.256
3rd Qu.:49.00  3rd Qu.:2.000  3rd Qu.: 86.00  3rd Qu.:1.800  3rd Qu.:190.0  3rd Qu.:156.0  3rd Qu.:68.00  3rd Qu.:1.460
Max.   :59.00  Max.   :2.000  Max.   :129.90  Max.   :2.000  Max.   :199.0  Max.   :169.0  Max.   :74.00  Max.   :2.000

Calories_Burned  Workout_Type  Fat_Percentage  Water_Intake..liters.  Workout_Frequency..days.week.  Experience_Level  BMI
Min.   : 303.0  Min.   :1.000  Min.   :10.00  Min.   :1.500  Min.   : 2.000  Min.   :1.00  Min.   :12.32
1st Qu.: 720.0  1st Qu.:1.000  1st Qu.:21.30  1st Qu.:2.200  1st Qu.:3.000  1st Qu.:1.00  1st Qu.:20.11
Median : 893.0  Median :3.000  Median :26.20  Median :2.600  Median :3.000  Median :2.00  Median :24.16
Mean   : 905.4  Mean   :2.494  Mean   :24.98  Mean   :2.627  Mean   :3.322  Mean   :1.81  Mean   :24.91
3rd Qu.:1076.0  3rd Qu.:3.000  3rd Qu.:29.30  3rd Qu.:3.100  3rd Qu.:4.000  3rd Qu.:2.00  3rd Qu.:28.56
Max.   :1783.0  Max.   :4.000  Max.   :35.00  Max.   :3.700  Max.   :5.000  Max.   :3.00  Max.   :49.84
>
```

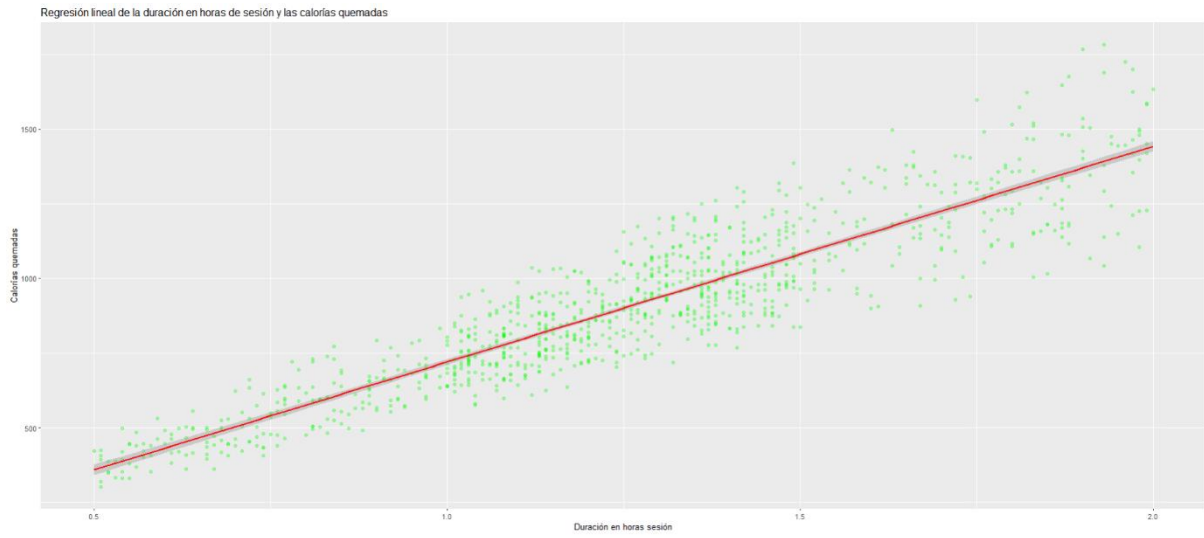
### Estadísticas Básicas del Conjunto de Datos

Este anexo presenta el análisis de los valores atípicos detectados en las observaciones de las calorías quemadas en las sesiones de entrenamiento de los miembros del gimnasio. A través del uso de gráficos de caja (boxplots) y medidas descriptivas, se ha identificado que un grupo de observaciones excede significativamente el rango intercuartílico, indicando la presencia de valores atípicos.

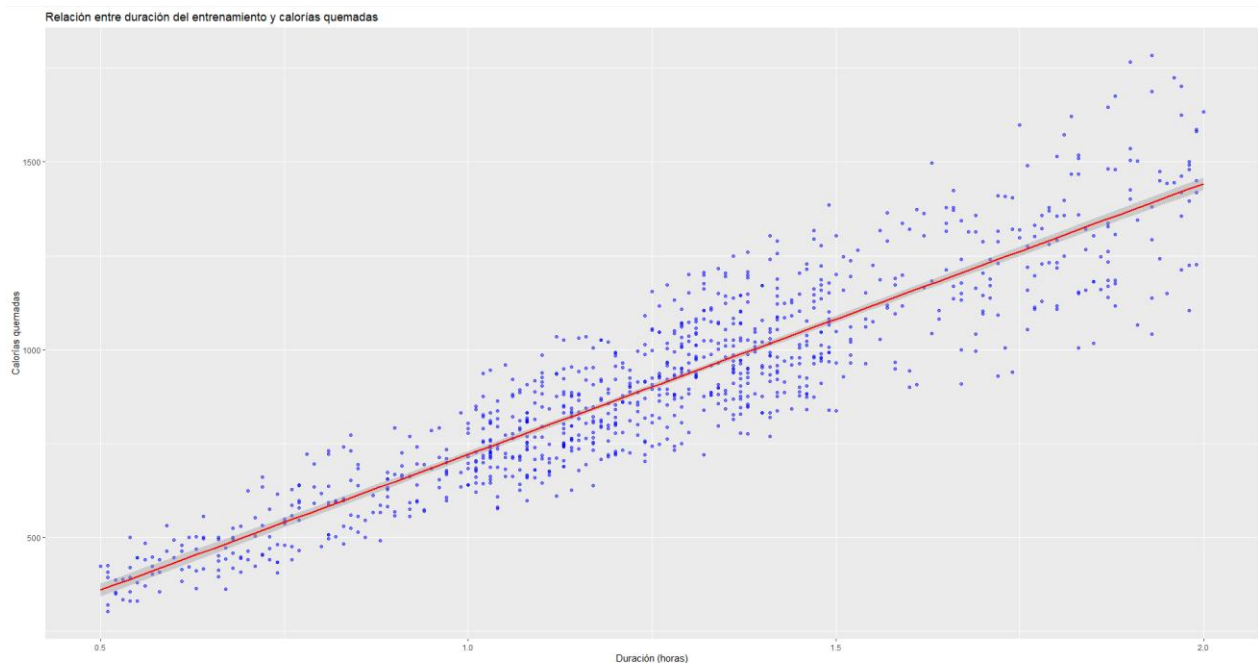


### Análisis de Valores Atípicos en las Calorías Quemadas

*Análisis de Datos sobre el Seguimiento de Ejercicio de Miembros en un Gimnasio para Mejorar el Rendimiento y la Experiencia del Usuario*

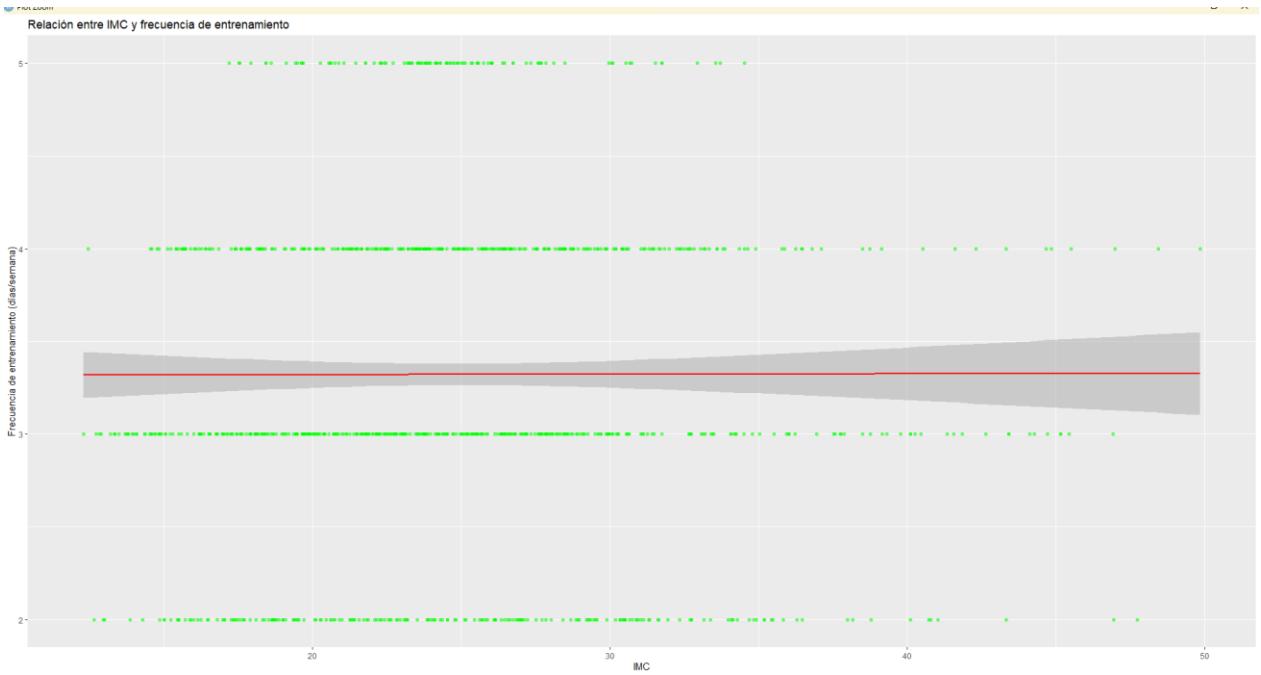


Algoritmo de Regresión lineal

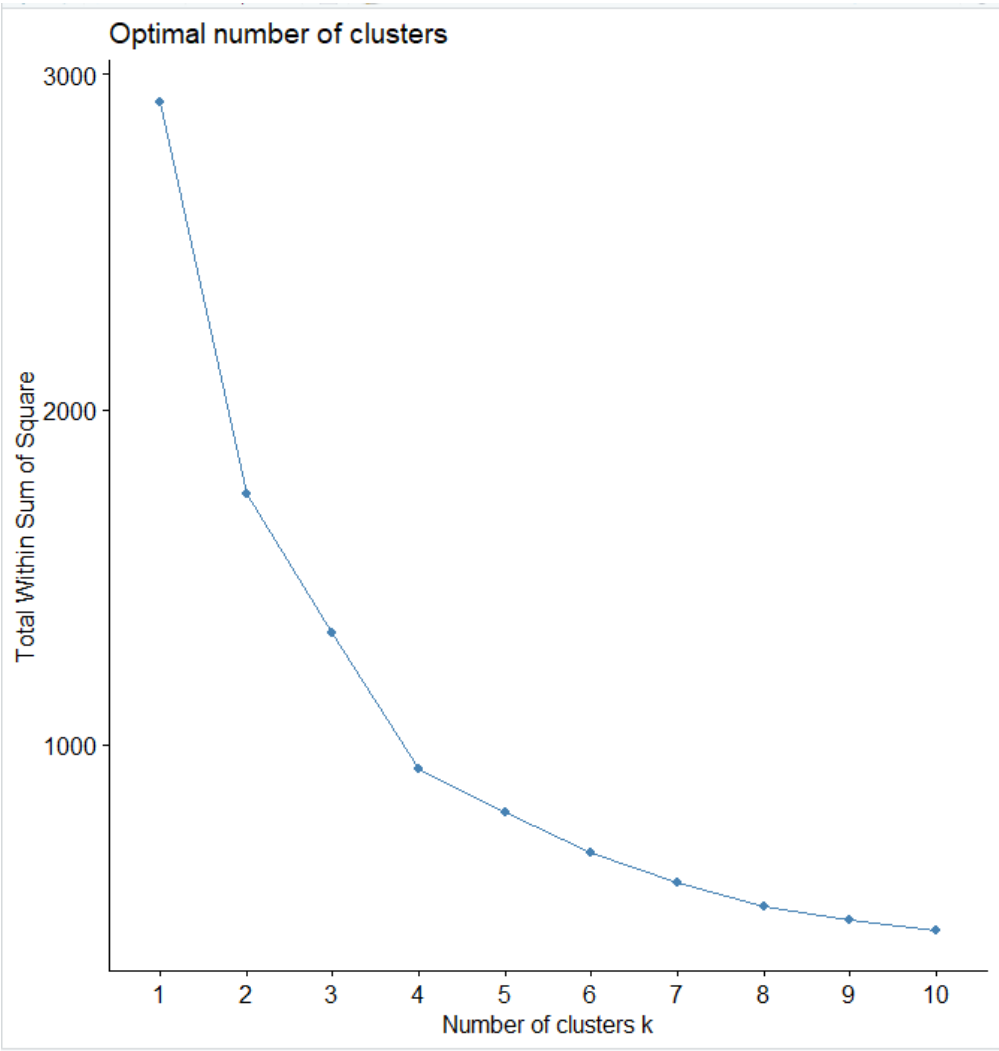


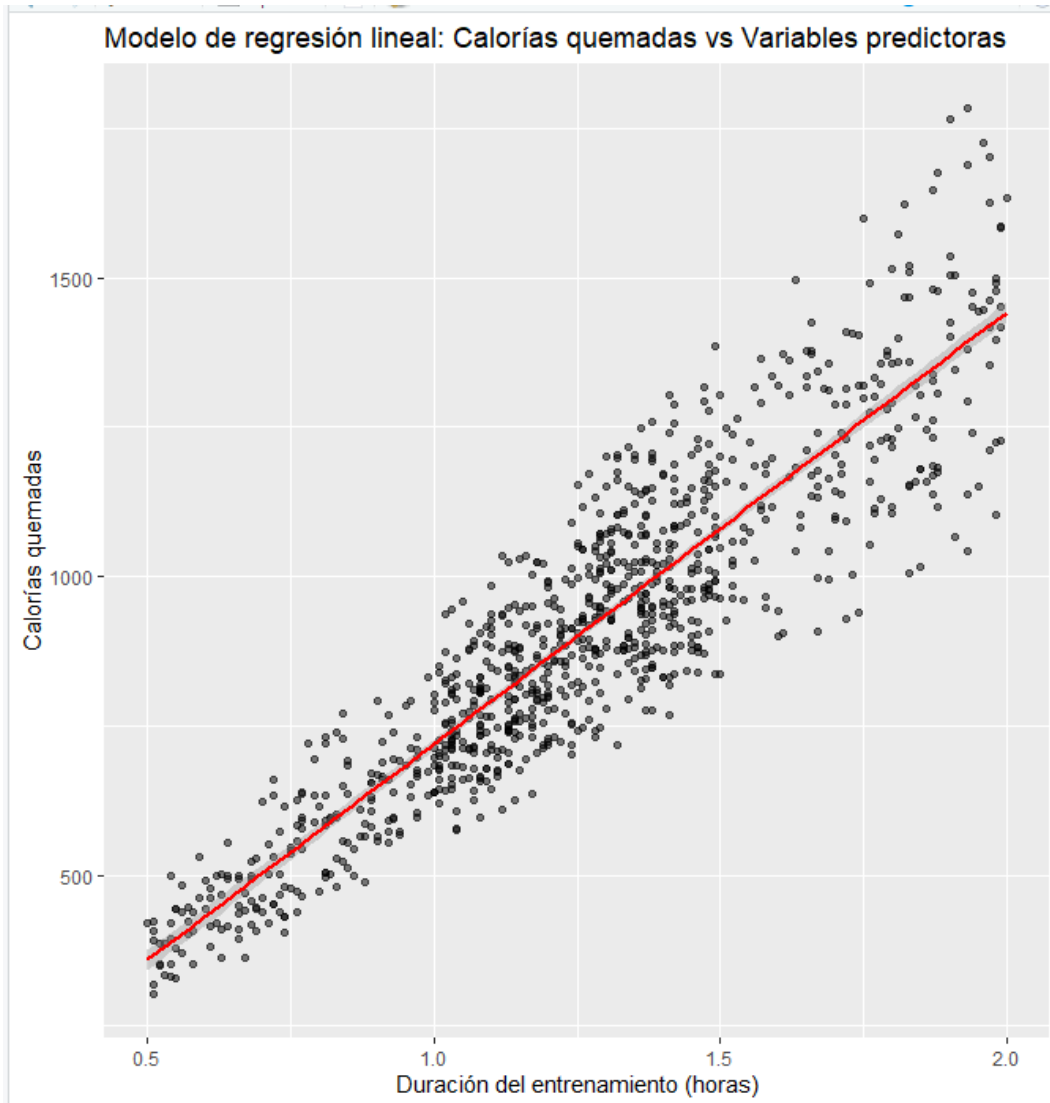
Algoritmo de Regresión lineal 2

*Análisis de Datos sobre el Seguimiento de Ejercicio de Miembros en un Gimnasio para Mejorar el Rendimiento y la Experiencia del Usuario*









### **Conclusiones**

El presente análisis permitió demostrar cómo el uso de técnicas de análisis de datos puede ofrecer una comprensión profunda y significativa del comportamiento de los miembros de un gimnasio, aportando evidencia valiosa para la toma de decisiones estratégicas en la gestión de estos espacios. A través de un enfoque cuantitativo y descriptivo-correlacional, se analizaron variables clave como la duración de las sesiones, las calorías quemadas, el tipo de ejercicio realizado, el IMC, la

frecuencia de entrenamiento, entre otras, lo cual permitió identificar relaciones estadísticas relevantes y patrones de comportamiento entre los usuarios.

Uno de los principales hallazgos fue la existencia de una correlación positiva entre la duración del entrenamiento y la cantidad de calorías quemadas, lo que valida empíricamente la hipótesis de que entrenamientos más largos tienden a generar mayor gasto calórico. Asimismo, se identificó que los miembros que optan por ejercicios grupales o de mayor intensidad suelen mostrar mayor constancia y frecuencia de asistencia, mientras que aquellos con un IMC más alto tienden a asistir con mayor regularidad, motivados por la mejora de su salud física.

La aplicación de técnicas de regresión lineal y clustering permitió no solo verificar hipótesis, sino también segmentar a los usuarios en grupos con características similares, lo que ofrece oportunidades concretas para la personalización de servicios, diseño de rutinas más efectivas y estrategias de fidelización adaptadas a cada perfil de usuario. Además, la detección de valores atípicos sugiere la posibilidad de implementar programas especializados para ciertos grupos que presentan condiciones fuera del rango promedio.

En definitiva, este estudio evidencia cómo el análisis de datos puede convertirse en una herramienta poderosa para transformar la operación de un gimnasio, permitiendo optimizar recursos, mejorar la experiencia del cliente y diseñar servicios más alineados con las necesidades reales de los usuarios. A partir de la información obtenida, es posible no solo mejorar los resultados individuales, sino también fortalecer la competitividad y sostenibilidad del negocio en un mercado altamente dinámico y centrado en la experiencia del usuario.

*Análisis de Datos sobre el Seguimiento de Ejercicio de Miembros en un Gimnasio para Mejorar el Rendimiento y la Experiencia del Usuario*