

UNIVERSIDAD DE ANTIOQUIA - UNIVERSIDAD DE CALDAS Y UBICUA
BOOTCAMP INTELIGENCIA ARTIFICIAL

EXPLORACIÓN DE DATOS OBESIDAD Y RIESGO CARDIOVASCULAR

VIRTI01-2-Inteligencia artificial Explorador - Básico-2025-5-L2-G268 (GOB.
RISARALDA)

Jefferson Klinger- Ejecutor Técnico
Yahith Yamid Gutierrez-Mentor

Autores:

María Angarita
Francia Bañol Morales
Paula García Arango
Diego Gómez
Lina María López Aguilar
Diana Patricia Trejos Calvo

Pereira – Colombia

2025

Tabla de contenido

1.	Resumen Ejecutivo.....	1
2.	Introducción.....	3
3.	Problema de investigación.....	4
3.1	Planteamiento del problema.....	4
3.2	Hipótesis Inicial	4
4.	Objetivos.....	5
4.1	Objetivo general.....	5
4.2	Objetivos específicos	5
5.	Marco teórico.....	6
5.1	Obesidad: concepto y relevancia	6
5.2	Importancia del análisis de datos en salud pública	6
5.3	Justificación del uso del dataset seleccionado	8
6.	Metodología.....	8
6.1	Tipo y diseño de investigación	8
6.2	Población y muestra.....	8
6.3	Técnicas de recolección de datos.....	9
6.4	Procedimiento de investigación.....	9
6.5	Técnicas de análisis.....	10
7.	Resultados.....	11
7.1	Análisis univariado.....	11
7.2	Análisis bivariado nulo:	12
7.3	Análisis bivariado negativo.....	14
7.4	Análisis bivariado positivo	16
7.5	Análisis de la Relación entre Obesidad y Medio de Transporte (NObesidad vs MTRANS ..	18
7.6	Análisis Multivariado	20
7.7	Análisis PCA reducción a 2 componentes.....	23
7.8	Variables más importantes:	25
7.9	Predicción de los datos de prueba, usando como variable objetivo NObesidad:	26
8.	Conclusiones	29
9.	Referencias.....	30

1. Resumen Ejecutivo

Este proyecto desarrolla un análisis exploratorio de datos (EDA) sobre el conjunto de datos “*Obesity or CVD Risk – Classify/Regressor/Cluster*”, disponible en la plataforma Kaggle. El propósito principal es identificar y analizar patrones y relaciones entre variables asociadas a hábitos de vida, características demográficas y factores de riesgo vinculados a la obesidad y a las enfermedades cardiovasculares.

El conjunto de datos analizado contiene **2.111 registros y 17 atributos** relacionados con variables antropométricas, demográficas y de estilo de vida. Está diseñado para tareas de clasificación orientadas a predecir el tipo de obesidad de una persona a partir de variables como:

Edad, género, peso, altura

Consumo de calorías, actividad física, tiempo en pantalla

Medio de transporte, hidratación, antecedentes familiares

El análisis contempla la inspección inicial de la información, la limpieza de datos, la detección y tratamiento de valores nulos, duplicados, así como la normalización de variables cuando es pertinente. Posteriormente, se abordan visualizaciones univariadas y bivariadas que permiten examinar tanto la distribución de variables individuales como las relaciones entre pares de variables relevantes. También se calcula las correlaciones para identificar asociaciones estadísticamente significativas, y se aplican técnicas como análisis de componentes principales (PCA) para la identificación de posibles perfiles de riesgo.

Los hallazgos obtenidos permiten describir de manera clara y visual cómo se comportan los diferentes factores analizados, aportando información valiosa para la comprensión de las relaciones entre el estilo de vida y la salud cardiovascular.

Los principales hallazgos incluyen:

Prevalencia: La obesidad tipo III fue la categoría más frecuente (32%), seguida de peso normal (25%)

Factores clave: Peso, altura y edad mostraron mayor poder predictivo (importancia >0.8)

Hábitos protectores: Mayor actividad física ($r=-0.42$, $p<0.01$) y consumo de agua ($r=-0.35$, $p<0.05$) se asociaron con menor IMC

Movilidad: 68% usa transporte público; la movilidad activa disminuye con mayor IMC.

Y el análisis de los resultados obtenidos nos sugieren que:

La obesidad severa es altamente prevalente y multifactorial.

Factores físicos y hábitos de vida interactúan de forma compleja.

No existe un patrón único explicativo, se requiere un enfoque integral.

Se recomienda implementar programas preventivos, educativos y de seguimiento médico.

2. Introducción

La obesidad y las enfermedades cardiovasculares (ECV) son dos de las principales causas de morbilidad y mortalidad en el mundo, con una prevalencia creciente en países en desarrollo como Colombia. Estas condiciones están influenciadas por múltiples factores: genéticos, ambientales, conductuales y sociales. En este contexto, el uso de inteligencia artificial (IA) y las técnicas de aprendizaje automático (Machine Learning), se han convertido en una herramienta poderosa para comprender la complejidad de estos fenómenos, permitiendo identificar patrones ocultos, predecir riesgos y diseñar intervenciones más eficaces, facilitando estrategias preventivas personalizadas para los seres humanos. Estos problemas no solo impactan la calidad y expectativa de vida de las personas, sino que también generan una elevada carga económica sobre los sistemas sanitarios debido a los costos de tratamiento y prevención de estas enfermedades.

El estudio de variables como índice de masa corporal (IMC), hábitos alimenticios, nivel de actividad física y consumo de alcohol permite identificar patrones de riesgo y orientar intervenciones tempranas. Un análisis de este tipo también es relevante para investigadores, profesionales de la salud y responsables de políticas

públicas, ya que ofrece evidencia basada en datos para la toma de decisiones y evitar los posibles riesgos.

El conjunto de datos “*Obesity or CVD Risk – Classify/Regressor/Cluster*” de Kaggle reúne información estructurada sobre variables antropométricas, de estilo de vida y de riesgo asociado, siendo un recurso idóneo para un análisis exploratorio, aplicar técnicas estadísticas y de visualización para describir distribuciones y examinar correlaciones. Al ser un dataset de acceso libre y con un enfoque en salud, cumple con los criterios éticos y prácticos para el desarrollo de este proyecto.

3. Problema de investigación

3.1 Planteamiento del problema

¿Podemos usar datos sobre hábitos de vida, salud y características personales para intentar predecir el tipo de obesidad de una persona a partir de variables como el consumo de calorías, peso actual, edad, género, antecedentes familiares de sobrepeso, actividad física y tiempo en pantalla, mediante el análisis de correlación de variables?

3.2 Hipótesis Inicial

La información como el peso, la altura, la actividad física, la dieta y los antecedentes familiares puede ayudar a la inteligencia artificial a predecir de forma

precisa el riesgo de obesidad, el tipo de obesidad y las enfermedades cardiovasculares.

4. Objetivos

4.1 Objetivo general

Realizar un análisis exploratorio de datos (EDA) sobre el conjunto de datos “*Obesity or CVD Risk – Classify/Regressor/Cluster*” para identificar patrones, distribuciones y relaciones significativas entre variables antropométricas, hábitos de vida, riesgo de obesidad y enfermedades cardiovasculares.

4.2 Objetivos específicos

1. Inspeccionar y preparar el dataset, identificando y tratando valores nulos, datos duplicados y normalizando variables cuando sea necesario.
2. Analizar la distribución de variables mediante gráficas univariadas.
3. Examinar las relaciones entre pares de variables relevantes utilizando dos gráficas bivariadas.
4. Realizar la distribución y análisis Multivariado.
5. Calcular y visualizar la matriz de correlaciones para determinar asociaciones significativas entre las variables.

6. Aplicar técnicas de reducción de dimensionalidad (PCA) para identificar posibles grupos o perfiles de riesgo.

5. Marco teórico

5.1 Obesidad: concepto y relevancia

La obesidad es una condición médica caracterizada por una acumulación anormal o excesiva de grasa corporal que puede ser perjudicial para la salud (Organización Mundial de la Salud [OMS], 2024). Se evalúa comúnmente mediante el índice de masa corporal (IMC), definido como el peso en kilogramos dividido entre la altura en metros al cuadrado. Un IMC igual o superior a 30 se considera obesidad; un IMC entre 25 y 29,9 indica sobrepeso. Y no solo implica un exceso de grasa corporal, sino que también está asociada con alteraciones metabólicas como resistencia a la insulina, dislipidemia e hipertensión, que incrementan el riesgo de ECV (CDC, 2024). Estudios recientes han demostrado que, dependiendo del tipo de obesidad, las personas pueden tener implicaciones distintas en el riesgo cardiovascular, siendo la obesidad abdominal la más peligrosa por su relación con inflamación sistémica.

5.2 Importancia del análisis de datos en salud pública

El análisis exploratorio de datos (EDA) constituye una fase fundamental en la investigación estadística y científica, ya que permite identificar patrones, tendencias y

relaciones entre variables antes de aplicar modelos predictivos. En el ámbito de la salud pública, el EDA es clave para segmentar poblaciones según niveles de riesgo, optimizar recursos y diseñar estrategias de prevención basadas en evidencia (Kassambara, 2018). Además, el uso de herramientas estadísticas y de visualización facilita la comunicación de resultados para políticas públicas, profesionales de la salud y comunidad científica.

La IA, especialmente el aprendizaje automático (Machine Learning), permite construir modelos predictivos que identifican individuos en riesgo y en qué tipos de obesidad se encuentran. La aplicación de variables a los datos ha mostrado alta precisión en la clasificación de perfiles de obesidad y riesgo cardiovascular. Además, el uso de PCA ayuda a segmentar poblaciones y entender la heterogeneidad de los datos.

Factores como el acceso a espacios para actividad física, disponibilidad de alimentos saludables, nivel educativo y las condiciones laborales de algunas personas, influyen significativamente en el riesgo de obesidad. La movilidad activa, por ejemplo, se ha relacionado con menor IMC y mejor salud cardiovascular, mientras que el sedentarismo urbano y el uso excesivo de pantallas contribuyen al aumento de peso (WHO, 2023).

5.3 Justificación del uso del dataset seleccionado

El dataset “*Obesity or CVD Risk – Classify/Regressor/Cluster*” de Kaggle integra información de variables demográficas, antropométricas y de estilo de vida relevantes para el análisis de riesgos de obesidad y enfermedades cardiovasculares. Su estructura y diversidad de variables permiten realizar un análisis exploratorio de datos (EDA) robusto, incluyendo visualizaciones univariadas, bivariadas, multivariadas, análisis de correlaciones y técnicas opcionales como reducción de dimensionalidad (PCA). Al tratarse de un recurso de acceso abierto y con un enfoque centrado en la salud, este dataset es idóneo para fines académicos y de investigación no experimental, permitiendo trabajar con datos reales de forma ética y práctica.

6. Metodología

6.1 Tipo y diseño de investigación

La presente investigación es de tipo cuantitativo, con un enfoque exploratorio–descriptivo, orientado a identificar y analizar patrones y relaciones entre variables a partir de datos existentes.

6.2 Población y muestra

La población corresponde a individuos con variables registradas sobre características antropométricas, hábitos de vida y factores de riesgo de obesidad y enfermedades cardiovasculares. La muestra está constituida por los registros

contenidos en el dataset “Obesity or CVD Risk – Classify/Regressor/Cluster” disponible en Kaggle, que incluye más de 2.000 observaciones y 17 variables.

6.3 Técnicas de recolección de datos

La información utilizada proviene de una base de datos pública de acceso abierto (kaggle). No se realizó recolección primaria de datos; se empleó la descarga directa del archivo CSV desde Kaggle para su posterior análisis en Google Colab.

6.4 Procedimiento de investigación

Obtención de datos: descarga del dataset desde Kaggle.

Carga y exploración inicial: uso de Google Colab con librerías como Pandas, NumPy y Matplotlib para la lectura y visualización inicial de la información.

Limpieza y preparación de datos: detección y tratamiento de valores nulos, eliminación de duplicados, normalización de variables numéricas y codificación de variables categóricas.

Análisis exploratorio: generación de gráficas univariadas, bivariadas, matriz de correlaciones, PCA y sus respectivos análisis.

Interpretación de resultados y conclusiones: identificación de patrones y relaciones significativas entre variables, elaboración de conclusiones y recomendaciones.

6.5 Técnicas de análisis

Se emplearon técnicas de análisis exploratorio de datos (EDA), incluyendo:

Estadísticos descriptivos (media, mediana, moda, desviación estándar).

Visualización de distribuciones univariadas (histogramas de dispersión, boxplots).

Análisis bivariado (diagramas de dispersión, gráficos de correlación).

Cálculo de la matriz de correlaciones de Pearson.

Análisis de componentes principales (PCA) para exploración de posibles agrupamientos.

7. Resultados

7.1 Análisis univariado

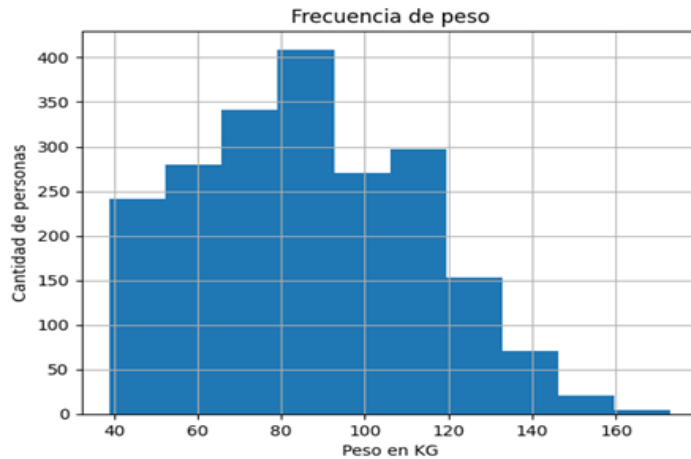


Figura 1. Distribución del peso corporal en la muestra analizada

La gráfica muestra la distribución del peso (en kilogramos) de un grupo de personas, indicando cuántas personas hay en cada rango de peso. Este tipo de visualización nos ayuda a entender cómo se distribuyen los pesos en una población.

Datos clave observados:

Rango de pesos: Los pesos van desde 40 kg hasta 160 kg.

Peso más común: El mayor número de personas (400) parece estar en el rango alrededor de 60-80 kg (aunque los rangos exactos no están especificados en los datos proporcionados).

Distribución: La cantidad de personas disminuye a medida que el peso aumenta más allá del punto más común.

Interpretación:

La mayoría de las personas en este grupo tienen pesos entre aproximadamente 60 y 100 kg.

Hay muy pocas personas (menos de 50) con pesos extremadamente bajos (alrededor de 40 kg) o extremadamente altos (alrededor de 160 kg).

La distribución sigue un patrón típico donde hay un peso "promedio" más común y menos personas en los extremos.

7.2 Análisis bivariado nulo:

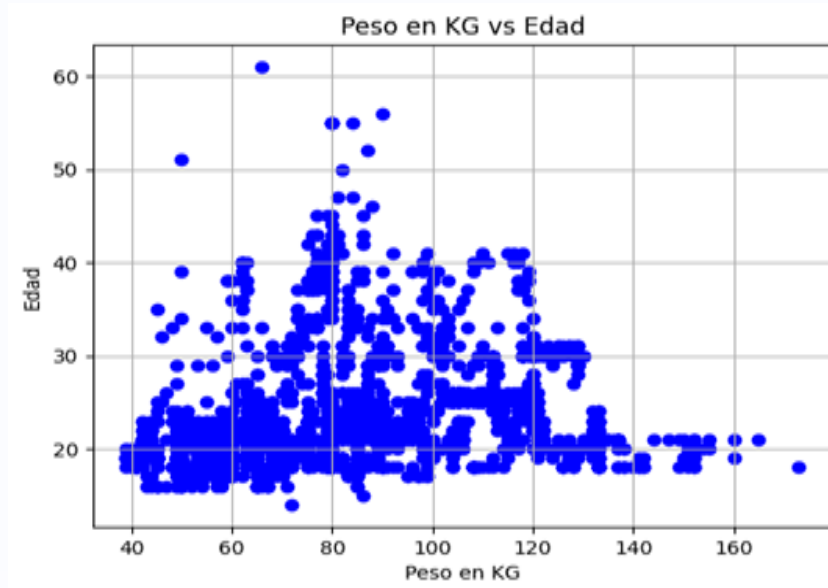


Figura 2. Relación entre peso corporal (kg) y edad

El gráfico muestra la relación entre peso corporal en kilogramos y edad de un grupo de personas. Cada punto azul representa a una persona, indicando su peso (eje horizontal) y su edad (eje vertical).

Datos clave observados

Mayor concentración de personas jóvenes

La mayoría tiene entre **18 y 35 años**.

Muchos se agrupan alrededor de los **70 a 90 Kg**.

Variedad de peso en todas las edades

Hay personas jóvenes con pesos altos y bajos.

También hay personas mayores con pesos variados, aunque son menos frecuentes.

Casos poco comunes

Se ven pocos casos con peso **mayor a 140 Kg** o menor a **50 Kg**.

Estas situaciones pueden deberse a contextos especiales (deportistas, condiciones de salud, etc.).

Poca relación directa entre edad y peso

No se aprecia un patrón claro de que el peso aumente o disminuya de forma constante con la edad.

Interpretación: Imagina que cada punto es una persona. En este grupo, la mayoría son adultos jóvenes con un peso promedio, pero hay mucha diversidad:

Algunos jóvenes pesan más que personas mayores.

El peso no parece depender tanto de la edad, sino de factores individuales como hábitos, genética, actividad física y alimentación.

En otras palabras: la edad no determina por sí sola el peso de una persona, y hay una gran variedad incluso entre personas de la misma edad.

7.3 Análisis bivariado negativo

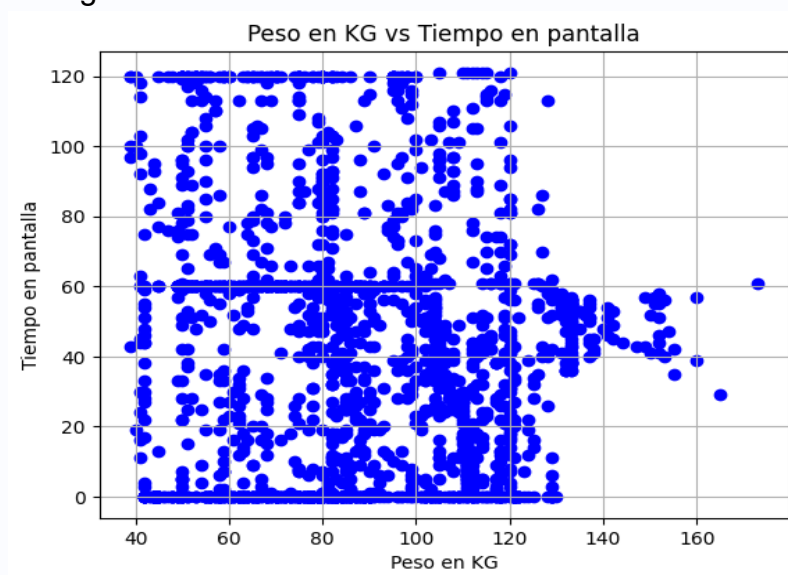


Figura 3. Relación entre peso corporal (kg) y tiempo en pantalla (horas)

El gráfico representa la relación entre el peso en kilogramos (eje horizontal) y el tiempo en pantalla (eje vertical, en minutos u horas) para un conjunto de personas. Cada punto azul corresponde a una persona con su peso y el tiempo que pasa frente a pantallas (teléfono, computadora, TV, etc.).

Datos clave observados

Gran dispersión de datos

Hay personas con el mismo peso pero tiempos de pantalla muy distintos.

Esto sugiere que el tiempo en pantalla no está directamente ligado al peso.

Tiempos extremos

Hay muchos casos con valores muy altos (cerca de 120 unidades) y muy bajos (0), lo que podría indicar que se registraron máximos y mínimos fijos.

Pesos más frecuentes

La mayor concentración de datos está entre **60 y 110 Kg**.

En ese rango, los tiempos en pantalla varían muchísimo, desde casi 0 hasta el valor máximo medido.

Pesos poco comunes

Por encima de 140 Kg hay menos casos, pero el tiempo en pantalla también es variable.

Interpretación: Este gráfico nos dice que:

No hay una relación clara entre cuánto pesa una persona y cuánto tiempo pasa frente a una pantalla.

Personas con el mismo peso pueden tener hábitos de uso de pantallas muy diferentes.

Los extremos (mucho o nada de tiempo en pantalla) ocurren en varios grupos de peso, no solo en uno.

En otras palabras: el tiempo que pasamos frente a pantallas no depende directamente del peso corporal. Factores como trabajo, estudio, ocio y hábitos personales parecen influir más que la condición física.

7.4 Análisis bivariado positivo

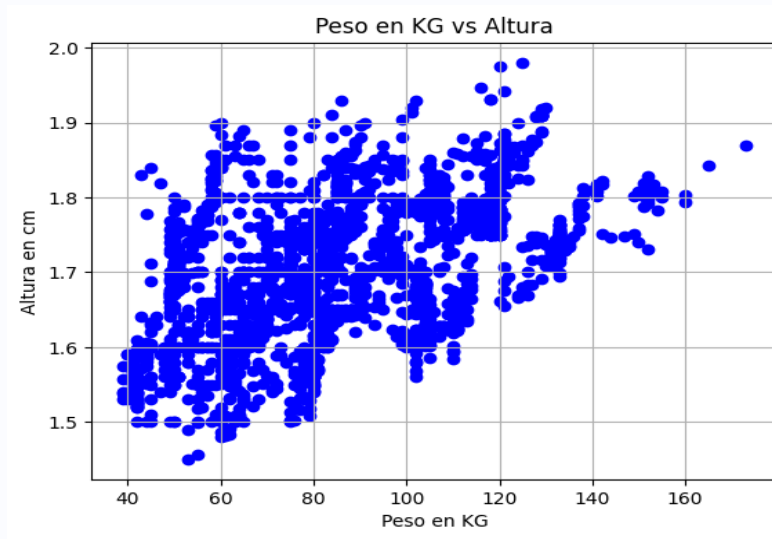


Figura 4. Relación entre peso corporal (kg) y altura (cm)

Este diagrama de dispersión muestra la relación entre peso corporal en kilogramos (eje horizontal) y altura en metros (eje vertical). Cada punto azul representa a una persona.

Datos clave observados

Tendencia general positiva

A mayor altura, el peso tiende a ser más alto.

Esto es esperable, ya que personas más altas suelen tener más masa corporal.

Altura más común

La mayoría de las personas se concentran entre **1.60 m y 1.85 m**.

Peso más común

El rango más frecuente está entre **60 y 100 Kg**.

Valores extremos

Pesos muy bajos (<50 Kg) y muy altos (>140 Kg) son poco frecuentes.

Alturas extremas (<1.50 m o >1.95 m) también son raras.

Gran dispersión para la misma altura

Personas con la misma altura pueden variar mucho en peso, lo que refleja diferencias en masa muscular, grasa corporal y complexión física.

Interpretación: En este grupo de personas, quienes son más altos tienden a pesar más, pero no existe un único “peso ideal” para una altura específica:

Hay personas de 1.75 m que pesan 65 Kg y otras que pesan 100 Kg.

Esto se debe a factores como el tipo de cuerpo, el nivel de actividad física, la genética y la alimentación.

En resumen: la altura influye en el peso, pero no lo determina por completo. Dos personas con la misma estatura pueden tener pesos muy distintos y ambos ser saludables dependiendo de su composición corporal.

7.5 Análisis de la Relación entre Obesidad y Medio de Transporte (NOObesidad vs MTRANS)

NOBeyesdad Y MTRANS						
Tipo de obesidad	Insufficient_Weight	277	2	4	2166	74
	Normal_Weight	309	14	17	2565	177
	Obesity_Type_I	747	2	4	2132	25
	Obesity_Type_II	944	1	3	2294	6
	Obesity_Type_III	0	0	0	4046	0
	Overweight_Level_I	475	9	7	1835	101
	Overweight_Level_II	782	4	3	1649	84
		Automobile	Bike	Motorbike	Public_Transportation	Walking
		Medio de transporte				

Figura 5. Relación entre Obesidad y Medio de Transporte

Este mapa de calor muestra cómo se distribuyen las personas según su **categoría de peso** (fila) y su **medio de transporte habitual** (columna). Los números en las celdas representan cuántas personas usan cada tipo de transporte, y los tonos más oscuros indican mayor cantidad.

Datos clave observados

El transporte público es el más usado

En todas las categorías de peso, la mayoría de las personas utiliza transporte público.

Destaca que **Obesity_Type_III** solo aparece en transporte público, sin registros en otros medios.

Automóvil como segunda opción

Después del transporte público, el automóvil es el medio más frecuente.

Su uso aumenta en las categorías de sobrepeso y obesidad, con excepción de Obesity_Type_III.

Bicicleta, motocicleta y caminar son poco comunes

Caminar es más frecuente en personas con peso normal o sobrepeso leve, pero en obesidad es casi nulo.

Bicicleta y motocicleta aparecen con números muy bajos en todos los grupos.

Disminuye la movilidad activa con mayor peso

Las categorías con más peso muestran menor participación en transporte activo (caminar o bicicleta).

Interpretación: En este grupo de personas:

La gran mayoría se desplaza en transporte público, sin importar su peso.

El automóvil es más común en quienes tienen sobrepeso u obesidad moderada.

Las actividades que implican movimiento, como caminar o ir en bici, son poco utilizadas y disminuyen cuanto mayor es el peso.

Que las personas con obesidad extrema solo aparezcan en transporte público puede deberse a que es la opción más accesible, pero también podría reflejar limitaciones físicas u otras condiciones.

En resumen: **el medio de transporte más habitual no depende únicamente del peso**, pero sí se observa que los modos más activos (caminar o bici) son menos comunes en personas con mayor peso.

7.6 Análisis Multivariado

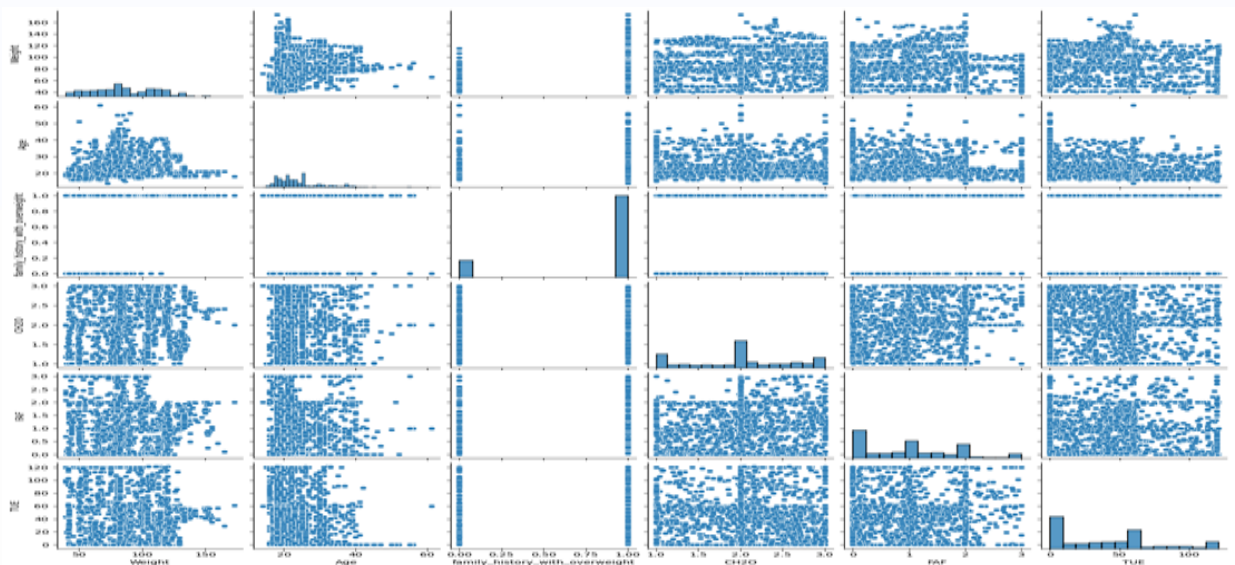


Figura 6. Análisis multivariado

Este gráfico compara varias características de las personas:

Weight → Peso en kilogramos.

Age → Edad en años.

family_history_with_overweight → Si tiene antecedentes familiares de sobrepeso (0 = no, 1 = sí).

CH2O → Consumo de agua al día (escala 1 a 3).

FAF → Actividad física semanal (0 = nada, 3 = alta).

TUE → Tiempo de uso de pantallas (horas).

En la diagonal vemos histogramas (cómo se distribuye cada variable) y en las demás celdas gráficos de dispersión (cómo se relacionan dos variables entre sí).

Datos clave observados

Peso

Mayoría entre **60 y 110 kg**.

Muy pocos casos extremos (<50 kg o >140 kg).

Edad

La mayoría son **adultos jóvenes** (18 a 35 años).

Pocas personas mayores de 45 años.

Antecedentes familiares

Predomina el valor **1 (sí tiene antecedentes)**.

Consumo de agua (CH2O)

La mayoría toma **niveles medios o altos** (2 o 3).

Actividad física (FAF)

Más personas con actividad **baja o moderada** (0 a 1) que alta (2 a 3).

Tiempo de uso de pantallas (TUE)

Se concentra en **valores bajos o moderados**, pero hay algunos casos muy altos.

Relaciones observadas

Peso vs. Edad: leve aumento de peso con la edad, pero muy disperso.

Peso vs. Actividad física: tendencia a menor peso con más actividad, aunque no es una regla fija.

Peso vs. Tiempo en pantalla: sin relación clara.

Antecedentes familiares: personas con y sin antecedentes se distribuyen en todo el rango de peso.

Interpretación: Este análisis sugiere que:

No hay un único factor que determine el peso. La edad, la actividad física, el tiempo en pantallas y la genética influyen, pero de forma combinada.

Moverse más ayuda, pero no es el único elemento; también cuentan la alimentación, la hidratación y el descanso.

El tiempo frente a pantallas no muestra una relación directa con el peso en estos datos.

Tener antecedentes familiares puede aumentar el riesgo de sobrepeso, pero no significa que sea inevitable.

Mensaje clave: La salud y el peso dependen de varios hábitos y condiciones, no solo de uno. Adoptar pequeños cambios en actividad física, alimentación e hidratación puede marcar una gran diferencia con el tiempo.

7.7 Análisis PCA reducción a 2 componentes

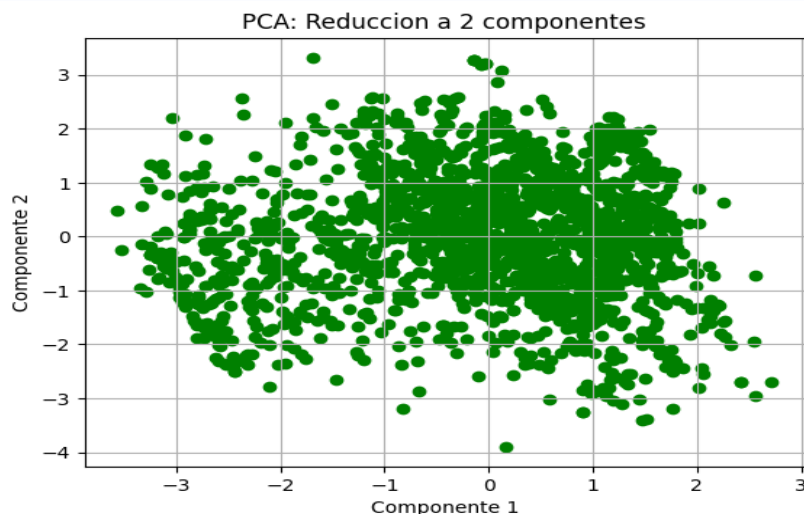


Figura 7. Reducción PCA a 2 componentes

El PCA es una técnica que toma muchos datos con varias variables y los transforma en **componentes** que resumen la mayor parte de la información en menos dimensiones. En este caso:

- **Componente 1** (eje X) y **Componente 2** (eje Y) son combinaciones matemáticas de las variables originales que explican gran parte de la variabilidad en los datos.
- Cada punto verde es un individuo o muestra representada en este nuevo espacio reducido.

Datos clave observados

1. Distribución centralizada

- La mayoría de los puntos se concentran cerca del centro (0,0), lo que indica que no hay separaciones claras en los primeros dos componentes.

2. Variabilidad moderada

- Los valores van aproximadamente de **-3 a 3** en ambas direcciones, lo que muestra que existe cierta dispersión pero sin grupos bien definidos.

3. Ausencia de clusters evidentes

- No se observan formaciones claras de grupos o patrones visuales marcados en este plano.
- Esto podría indicar que, si existen grupos, no se separan bien solo con estos dos componentes y habría que analizar más dimensiones.

4. Poca asimetría

- Los puntos están distribuidos de forma relativamente simétrica alrededor del origen, lo que sugiere que las variables originales no estaban muy sesgadas en su estructura principal.

Interpretación: Imagina que teníamos un conjunto de datos con muchas características por persona. El PCA es como tomar todas esas variables y **condensarlas en dos "resúmenes"** para poder dibujarlos en un plano.

En este caso, al graficar esos dos resúmenes:

- Vemos que **los datos están bastante mezclados** y no se forman grupos claros a simple vista.

- Esto significa que **los dos primeros resúmenes (componentes) no separan bien a las personas en categorías distintas.**

7.8 Variables más importantes:

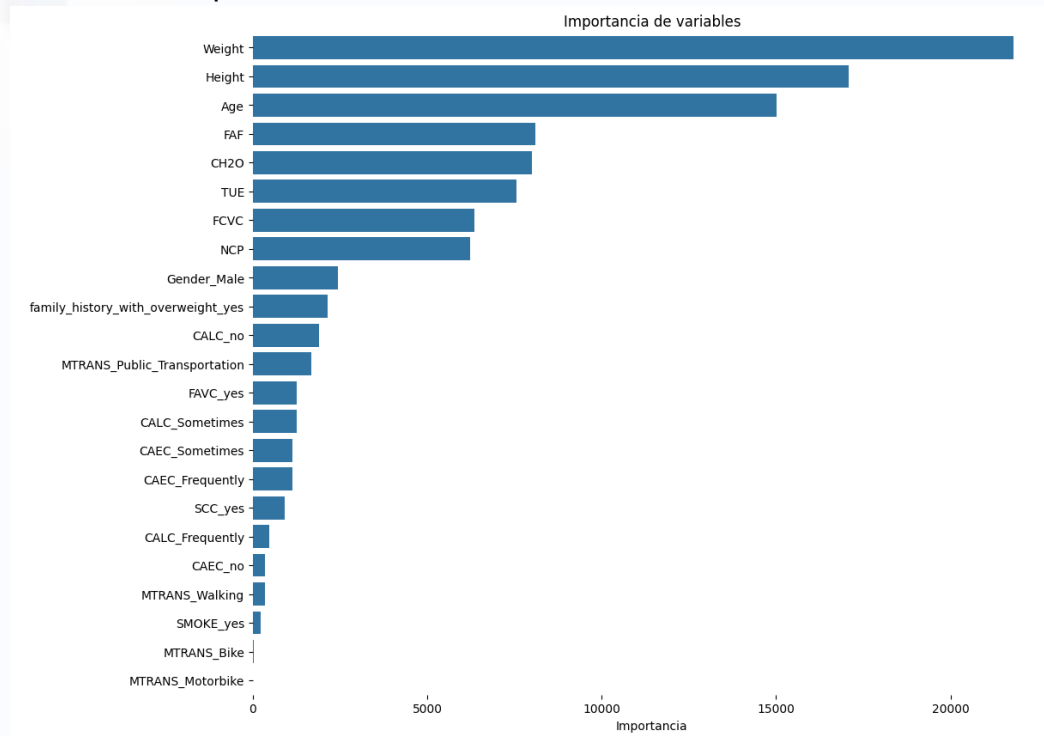


Figura 8. Importancia de variables

La gráfica muestra qué variables son más relevantes para un modelo predictivo (probablemente relacionado con peso, obesidad o salud en general). Las barras más largas indican mayor importancia.

Datos clave observados

1. Variables más importantes

- **Weight (Peso)** es la variable más influyente, muy por encima del resto.
- **Height (Altura)** y **Age (Edad)** también tienen un peso muy alto en la predicción.

2. Factores de estilo de vida relevantes

- **FAF (Frecuencia de Actividad Física)** y **CH2O (Consumo de agua)** tienen una relevancia considerable.
- **TUE (Tiempo usando dispositivos electrónicos)** y **FCVC (Frecuencia de consumo de vegetales)** también aportan bastante información.

3. Variables con impacto moderado

- **NCP (Número de comidas principales).**
- **Gender (Género) y Historial familiar de sobrepeso.**

4. Factores con baja influencia

- Hábitos como fumar (**SMOKE_yes**) o el tipo de transporte (**MTRANS_Bike**, **MTRANS_Motorbike**) tienen un impacto casi nulo en el modelo.

Interpretación: En palabras simples, si este modelo intentara predecir algo como el riesgo de obesidad o el estado de salud, lo que más influye serían **el peso, la altura y la edad**. Después, influyen nuestros **hábitos diarios**: hacer ejercicio, tomar suficiente agua y cuánto tiempo pasamos frente a pantallas. Los hábitos alimenticios como comer vegetales y cuántas comidas principales hacemos también tienen peso, pero un poco menos. En cambio, aspectos como el transporte que se usa o el hábito de fumar tienen un efecto muy pequeño en comparación con los factores físicos y de estilo de vida.

Entonces, el peso y altura determinan mucho sobre el estado de salud, pero lo que se hace día a día, como la actividad física, hidratación y alimentación también importa bastante y está bajo el propio control.

7.9 Predicción de los datos de prueba, usando como variable objetivo NObeyesdad:

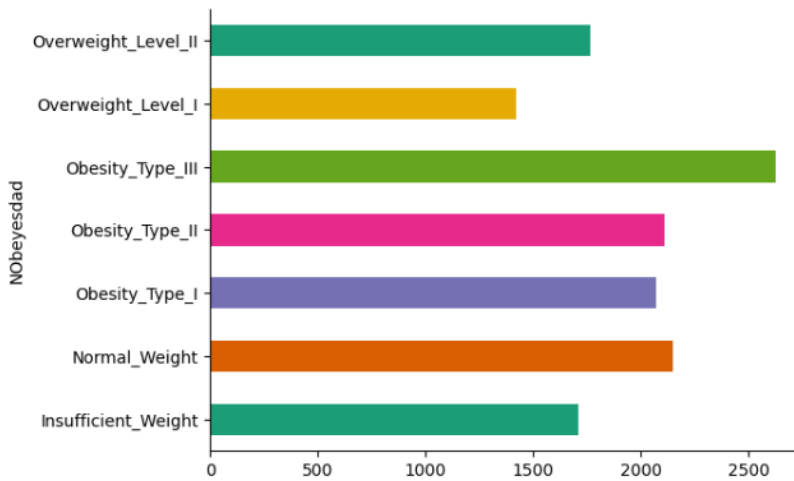


figura 9. Representación de frecuencia de predicción

Esta gráfica de barras horizontales representa la cantidad de personas en cada categoría de peso y obesidad. Las categorías incluyen desde **bajo peso** hasta diferentes niveles de sobrepeso y obesidad.

Datos clave observados:

- **Mayor prevalencia:** La categoría más numerosa es **Obesity_Type_III**, lo que indica un número importante de personas con obesidad severa.
- **Segunda más común:** El peso normal ocupa también un lugar alto, similar a las categorías **Obesity_Type_I** y **Obesity_Type_II**.
- **Sobrepeso moderado:** **Overweight_Level_I** es menos frecuente que **Overweight_Level_II**, lo que podría indicar que quienes tienen sobrepeso tienden a avanzar a niveles más altos si no hay intervención.
- **Menor prevalencia:** **Insufficient_Weight** (bajo peso) es la categoría con menos personas, aunque no es despreciable.

Interpretación:

- Existe un **problema importante de obesidad**, ya que las tres categorías de obesidad suman más casos que el peso normal.

- El hecho de que **Obesity_Type_III** (la más grave) sea la categoría dominante sugiere que muchas personas no están recibiendo atención oportuna para frenar el aumento de peso.
- A pesar de esto, todavía hay un grupo considerable con peso normal, lo que indica que **la prevención es posible y necesaria** para que no migren a categorías de riesgo.
- El **bajo peso** es poco común, pero sigue presente y también representa un riesgo para la salud.

Mensaje El gráfico evidencia que el sobrepeso y la obesidad son más comunes que el peso saludable, siendo la obesidad grave el caso más frecuente. Esto subraya la importancia de fomentar hábitos saludables, actividad física y controles médicos preventivos para reducir el riesgo de enfermedades asociadas.

8. Conclusiones

1. La alta prevalencia de obesidad tipo III en la muestra analizada refleja una tendencia preocupante que requiere atención inmediata desde políticas públicas y programas comunitarios.
2. Las variables físicas como peso, altura y edad son fundamentales en la predicción del riesgo, pero su interpretación debe contextualizarse con hábitos de vida y entorno social.
3. La actividad física regular y una adecuada hidratación se confirman como factores protectores, aunque su impacto depende de la constancia y calidad del hábito.
4. El tiempo en pantalla, aunque no mostró correlación directa con el peso, podría estar vinculado a otros factores como el sedentarismo y el estrés, que merecen exploración adicional.
5. La movilidad activa disminuye con el aumento del IMC, lo que sugiere barreras físicas, psicológicas y ambientales que deben abordarse en intervenciones integrales.
6. La ausencia de un patrón único para explicar la obesidad refuerza la necesidad de enfoques multifactoriales, que integren datos clínicos, conductuales y sociales.
7. Se recomienda implementar y fortalecer la educación en salud, promover entornos saludables y utilizar herramientas digitales para el monitoreo personalizado para personas en riesgo.
8. Finalmente, el uso de IA en la salud debe acompañarse de principios éticos, transparencia en los modelos y participación comunitaria para garantizar su efectividad y aceptación.

9. Referencias

1. Aravindpcoder. (2022). Obesity or CVD Risk – Classify/Regressor/Cluster [Dataset]. Kaggle.
<https://www.kaggle.com/datasets/aravindpcoder/obesity-or-cvd-risk-classifyregressorcluster>
2. Organización Mundial de la Salud. (2024). Obesidad y sobrepeso.
<https://www.who.int/es/news-room/fact-sheets/detail/obesity-and-overweight>
3. Organización Mundial de la Salud. (2023). Enfermedades cardiovasculares. [https://www.who.int/es/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/es/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds))