



## **Entrega de Avance N°1 del Proyecto**

**Milton Cesar Machado Barreto, 89326**

**José Camilo Rodríguez Ortega, 894995**

**Jhon Kevin Barragán Vásquez, 912409**

**Docente: Harold Enrique Freyte Zarta**

**Curso: Ingeniería de Software**

**NRC: 60-81403**

**Facultad Ingeniería, Programa de Ingeniería de Sistemas**

**Corporación Universitaria Minuto de Dios – UNIMINUTO**

**Vicerrectoría Regional Centro Sur**

**Ciudad Ibagué, Tolima**

Año: 2025

## Contenido

1.	Descripción del Caso de Estudio.....	3
1.1.	Contexto del Caso de Estudio .....	3
1.2.	Objetivos del Sistema .....	4
	• Objetivo General.....	4
	• Objetivos Específicos .....	4
1.3.	Procesos Clave Relacionados con Datos.....	5
1.4.	Diagrama de Flujo de Información .....	7
1.5.	Palabras Clave .....	9
2.	Requerimientos de Datos Identificados.....	11
2.1.	Requerimientos Funcionales .....	11
2.2.	Requerimientos No Funcionales.....	12
3.	Modelo de Datos Conceptual .....	15
3.1.	Diagrama ER – Modelo Conceptual.....	15
3.2.	Descripción de Entidades Principales.....	16
4.	Modelo de Datos Lógico .....	18
4.1.	Diagrama ER – Modelo Lógico.....	18
4.2.	Normalización .....	20
5.	Justificación de Decisiones de Modelado y Análisis de Calidad de Datos .....	21
5.1.	Justificación del Modelado .....	21
5.2.	Análisis de Calidad de Datos .....	22

# 1.Descripción del Caso de Estudio

## 1.1. Contexto del Caso de Estudio

El caso de estudio se fundamenta en el dataset Mental Health in Medical Students from Five Regions of Colombia, publicado en Mendeley Data, que ha tenido cuatro versiones entre noviembre de 2024 y mayo de 2025.

Sitio web: <https://data.mendeley.com/datasets/zj6k5r6xxg/1>

- Historial de Versiones del Dataset Mendeley

Versión	Fecha de publicación	Ultima Fecha de Publicación	Archivos incluidos	Cambios principales
1	24 de Noviembre 2024	25 de noviembre 2024	BASE DATOS MENDELEY DATA FELI.xlsx	Publicación inicial.
2	21 de febrero del 2025	27 de marzo del 2025	Mental Health in medical students.xlsx	Corrección y reorganización.
3	26 de marzo del 2025	27 de marzo del 2025	BD_Depresion.xlsx	Depuración de cuestionarios clínicos.
4	27 de marzo del 2025	01 de mayo del 2025	BD_Depresion.xlsx, datahealth.xlsx, questionnaires.docx, Supplementary Table 3.docx	Consolidación con tablas suplementarias.

Además de la estructura técnica del dataset, el contexto de la organización está directamente ligado a la realidad social y académica de los estudiantes de medicina en Colombia, quienes enfrentan altos niveles de presión académica, carga emocional y estrés. Este grupo poblacional se convierte en un punto crítico para la investigación, ya que el cuidado de su salud mental impacta directamente en su rendimiento académico y en la calidad futura de la atención médica que ofrecerán.

El caso de estudio se conecta con la necesidad de diseñar sistemas tecnológicos de apoyo, capaces de recolectar, organizar y analizar datos de salud mental en contextos universitarios. En este marco, Relaxify surge como una propuesta aplicada que toma la información del dataset y la convierte en la base para una plataforma digital de bienestar que combina investigación académica con innovación tecnológica.

## 1.2. Objetivos del Sistema

- **Objetivo General**

Diseñar e implementar una plataforma virtual denominada Relaxify, que permita gestionar, procesar y analizar datos relacionados con la salud mental y el bienestar de estudiantes universitarios, usando como referencia el dataset de Mendeley Data y aplicando los principios de arquitectura de datos.

- **Objetivos Específicos**

- 1) Recolectar información sociodemográfica, académica y psicológica de estudiantes mediante formularios digitales.
- 2) Procesar y analizar cuestionarios clínicos (PHQ-9, GAD-7, EEP-10, Rosenberg, APGAR, IPAQ) y clasificarlos por nivel de riesgo.
- 3) Generar reportes individuales y grupales sobre el estado de salud mental de la población estudiantil.
- 4) Implementar mecanismos de visualización que apoyen la toma de decisiones en instituciones educativas.
- 5) Diseñar una arquitectura escalable para integrar nuevas fuentes de datos como dispositivos digitales.
- 6) Garantizar seguridad, confidencialidad y calidad de la información almacenada y procesada.

### 1.3. Procesos Clave Relacionados con Datos

- Registro y validación de información sociodemográfica.

Este proceso asegura que los datos básicos de los estudiantes (edad, sexo, semestre, ciudad, nivel socioeconómico, entre otros) se registren correctamente en el sistema. La validación evita inconsistencias y garantiza la calidad inicial de la información.

- Captura digital de cuestionarios clínicos estandarizados.

Incluye la aplicación de instrumentos reconocidos como PHQ-9, GAD-7, EEP-10, Rosenberg, Apgar o IPAQ. Estos cuestionarios son la base de la evaluación psicológica y emocional del caso de estudio y se digitalizan en Relaxify para su recolección eficiente.

- Almacenamiento seguro y clasificación de puntajes.

Una vez aplicados los cuestionarios, los datos deben guardarse en una base de datos estructurada. Además, se clasifican automáticamente en niveles de riesgo (bajo, moderado, alto), lo cual permite un análisis inmediato y la identificación temprana de alertas.

- Integración de datos en un historial centralizado.

Cada estudiante cuenta con un historial que consolida sus encuestas, resultados y actividades de bienestar. Este registro longitudinal facilita el seguimiento de la evolución en el tiempo y la toma de decisiones personalizadas.

- Generación de reportes y análisis en tiempo real.

El sistema produce reportes individuales (para el propio estudiante) y grupales (para docentes o instituciones), con resultados actualizados. Esto convierte los datos en información práctica para la prevención y el acompañamiento académico.

- Exportación de información hacia herramientas estadísticas externas (R, SPSS, Python).

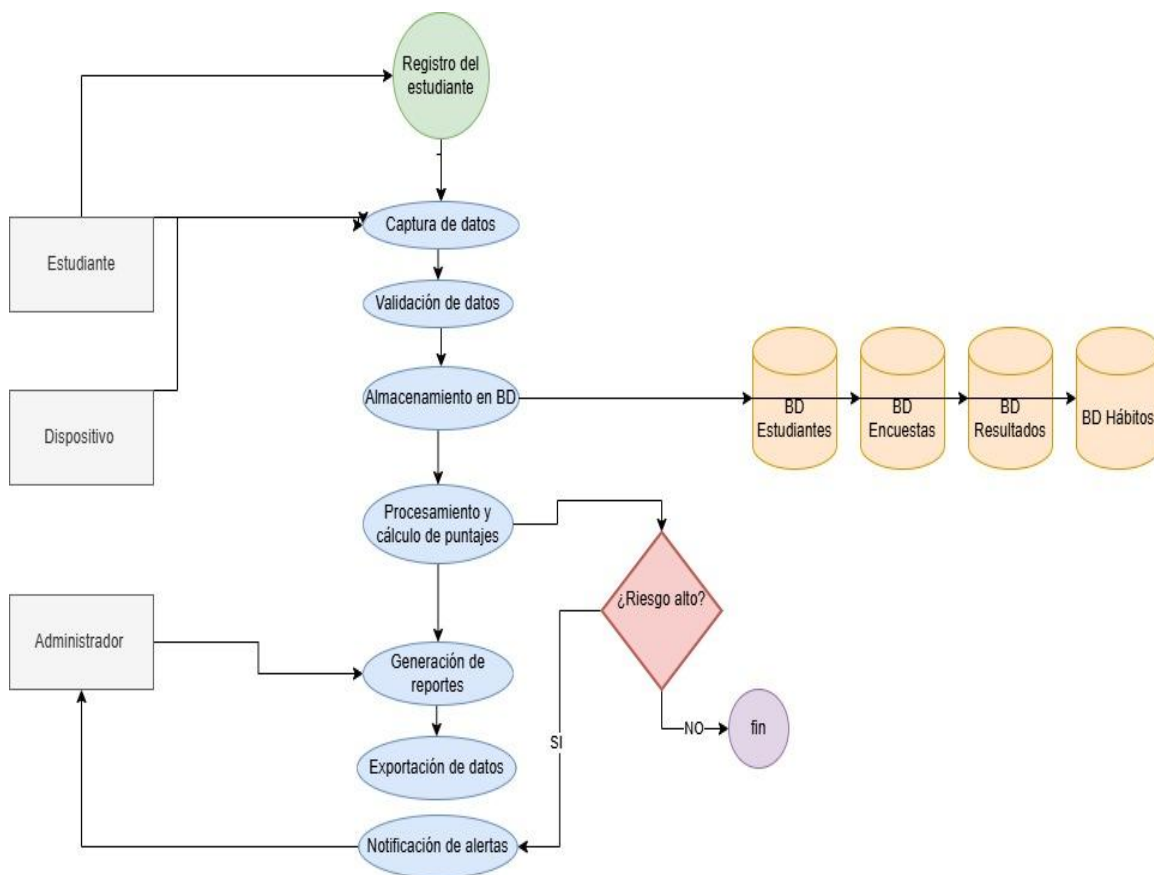
Este proceso permite que la información almacenada en Relaxify pueda ser utilizada en entornos especializados para investigación científica. SPSS se aplica en ciencias sociales y médicas, R ofrece análisis multivariados y gráficos avanzados, y Python permite aplicar minería de datos y machine learning. Esta interoperabilidad fortalece el vínculo entre el sistema y la producción de conocimiento académico.

## 1.4. Diagrama de Flujo de Información

El flujo de información en Relaxify comprende las fases:

- Captura: mediante formularios digitales y dispositivos conectados.
- Almacenamiento: en una base de datos relacional con seguridad.
- Procesamiento: cálculo de puntajes y niveles de riesgo.
- Visualización: reportes para estudiantes, docentes y administradores.

### 1.4.1. Diagrama de flujo del Sistema Relaxify



Este es el diagrama de flujo del sistema, donde se muestran los actores que intervienen, los procesos principales, las bases de datos utilizadas y el flujo lógico desde el inicio hasta el final del sistema:

1) Inicio

El proceso comienza con el registro del estudiante, que es indispensable para poder capturar información de cada usuario.

2) Captura y validación de datos

Después, se realiza la captura de datos a través del estudiante y su dispositivo.

Una vez ingresados, el sistema hace la validación de datos, es decir, revisa que la información sea correcta, completa y coherente.

3) Almacenamiento en bases de datos

Cuando los datos son validados, se procede al almacenamiento en la base de datos.

Aquí vemos que la información se organiza en varias bases: estudiantes, encuestas, resultados y hábitos, lo que permite una gestión más clara de la información.

4) Procesamiento y puntajes

El sistema luego hace el procesamiento y cálculo de puntajes, lo cual permite evaluar la información y determinar posibles riesgos.

5) Decisión de riesgo

Aquí hay un punto clave: el sistema evalúa si existe un riesgo alto.

Si no hay riesgo, el proceso termina.

Si sí hay riesgo, se activa un flujo adicional para generar reportes y notificar al administrador.

6) Generación de reportes y notificación



En caso de riesgo, el sistema genera reportes, permite la exportación de datos y envía notificaciones de alerta al administrador, quien es el encargado de tomar acciones.

#### 7) Cierre del proceso

Finalmente, el sistema concluye su flujo, ya sea porque no hay riesgo o porque se notificó al administrador.

#### 8) Conclusión

En conclusión, este diagrama refleja el ciclo completo del sistema: desde que el estudiante se registra y aporta información, hasta que el administrador recibe alertas y reportes en caso de ser necesario. Es una visión global y ordenada del funcionamiento.

## 1.5. Palabras Clave

- PHQ-9 (Patient Health Questionnaire-9): Cuestionario clínico de 9 ítems utilizado para evaluar la presencia y la severidad de síntomas de depresión.
- GAD-7 (Generalized Anxiety Disorder-7): Escala de 7 preguntas diseñada para detectar y medir la gravedad de la ansiedad generalizada.
- APGAR Familiar: Instrumento que mide la percepción de funcionalidad familiar en cinco áreas: Adaptabilidad, Participación, Gradiente de crecimiento, Afecto y Recursos.

- Escala de Autoestima de Rosenberg: Cuestionario ampliamente usado para medir la autoestima global de una persona.
- EEP-10 (Escala de Estrés Percibido-10): Instrumento de 10 ítems que evalúa la percepción de estrés en el último mes.
- Encuesta sociodemográfica: Cuestionario que recopila datos básicos como edad, sexo, estado civil, ingresos, semestre o lugar de residencia.
- Apoyo social: Percepción del individuo respecto al respaldo que recibe de familiares, amigos o comunidad en situaciones difíciles.
- Hábitos de vida: Información relacionada con el consumo de sustancias (alcohol, tabaco), actividad física, calidad del sueño y rutinas diarias.
- Nivel de riesgo: Clasificación derivada de los cuestionarios aplicados (leve, moderado, severo) que indica la probabilidad de problemas en salud mental.
- Mendeley Data: Plataforma de datos abiertos donde se encuentra alojado el conjunto de información utilizado en este caso de estudio.

## 2. Requerimientos de Datos Identificados

En este proyecto, los requerimientos se identificaron considerando dos niveles:

El proyecto Relaxify, que busca diseñar una plataforma digital de bienestar emocional apoyada en inteligencia artificial y dispositivos inteligentes para la detección temprana de estrés y problemas de salud mental.

El caso de estudio seleccionado, basado en la base de datos de Mendeley sobre estudiantes universitarios en Colombia, utilizado en esta primera entrega como un ejemplo real para aplicar los conceptos de arquitectura de datos.

De esta forma, los requerimientos que se presentan a continuación reflejan tanto las necesidades del proyecto Relaxify como las características del caso de estudio.

### 2.1. Requerimientos Funcionales

- Transaccionales
  - 1) Registro de estudiantes participantes.
  - 2) Captura de respuestas en encuestas y cuestionarios.
  - 3) Almacenamiento inmediato de resultados.
  
- Analíticos
  - 1) Reportes individuales sobre riesgo en salud mental.

- 2) Reportes grupales sobre tendencias y prevalencia.
- 3) Identificación de patrones en la evolución de estudiantes.

- Históricos

- 1) Historial de encuestas realizadas.
- 2) Seguimiento de resultados longitudinales.
- 3) Registro histórico de actividades de bienestar.

## 2.2. Requerimientos No Funcionales

- Seguridad (anonimización y cifrado).

El sistema debe garantizar la confidencialidad de la información mediante la anonimización de datos personales y el cifrado en el almacenamiento y la transmisión.

- Escalabilidad para diferentes universidades.

La arquitectura debe soportar la incorporación de nuevas instituciones educativas sin afectar el rendimiento ni requerir rediseños complejos.

- Disponibilidad en múltiples plataformas (web y móvil).

Relaxify debe ser accesible desde navegadores y dispositivos móviles, facilitando el acceso de los estudiantes y administradores en distintos entornos.

- Rendimiento eficiente en consultas.

El sistema debe responder de manera rápida a las consultas, reportes y análisis, incluso con grandes volúmenes de datos.

- Usabilidad para usuarios sin conocimientos técnicos.

La interfaz debe ser intuitiva y comprensible, permitiendo que estudiantes y personal administrativo interactúen sin requerir formación avanzada.

- Compatibilidad con exportaciones (CSV, Excel, SPSS).

El sistema debe generar exportaciones en formatos estándar que faciliten la interoperabilidad con otras herramientas estadísticas y de análisis.

- Mantenibilidad.

El sistema debe permitir actualizaciones, correcciones de errores y mejoras sin interrumpir el servicio ni comprometer la integridad de los datos.

- Confiabilidad.

Relaxify debe asegurar la consistencia de los resultados, evitando pérdida de información y garantizando que los cálculos de riesgo se mantengan precisos.

- Accesibilidad.

El sistema debe contemplar lineamientos de accesibilidad (ej. compatibilidad con lectores de pantalla) para estudiantes con discapacidades.

- Interoperabilidad.

Debe permitir la integración con otros sistemas académicos o institucionales, asegurando el intercambio de información sin necesidad de reprocesar datos.

### 3. Modelo de Datos Conceptual

El modelo de datos conceptual define la estructura lógica del caso de estudio, resaltando las entidades principales, sus atributos más relevantes y las relaciones de alto nivel entre ellas. Este modelo busca representar cómo los participantes universitarios, dentro del contexto de Relaxify y los cuestionarios aplicados, generan información asociada a su salud mental, hábitos de vida y uso de la aplicación.

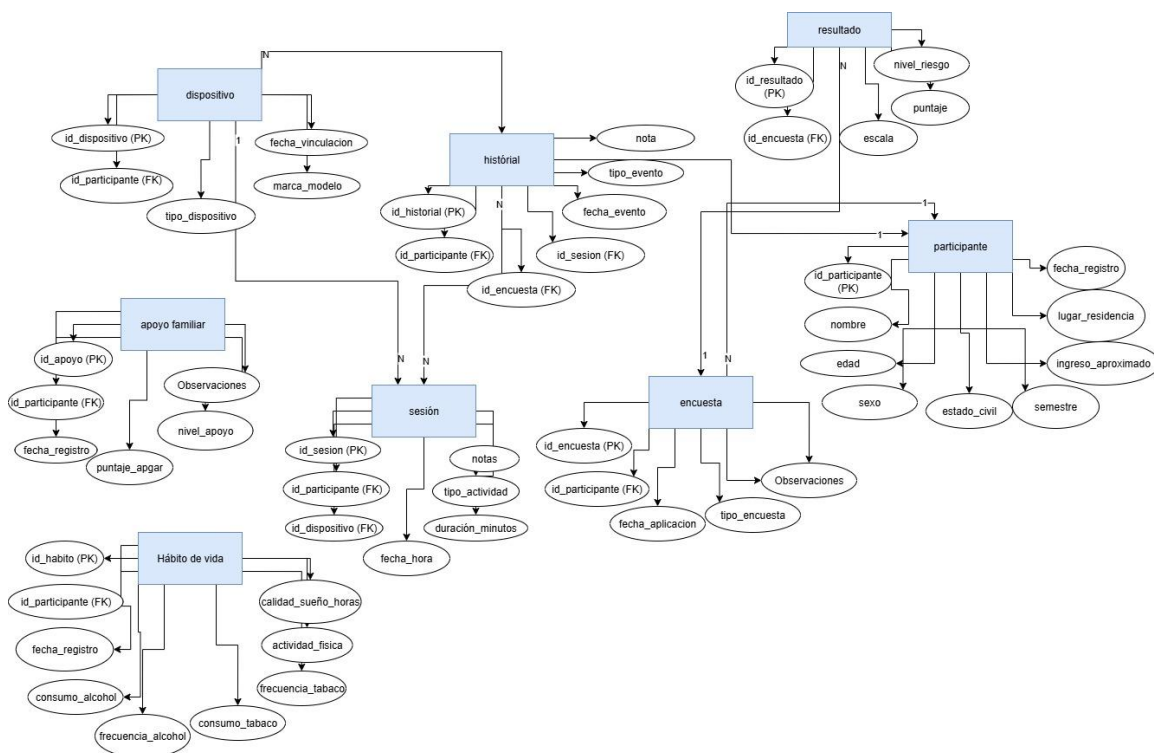
En este nivel, se identifican entidades clave como Participante, Encuesta, Resultado, Hábitos de Vida, Apoyo Familiar, Dispositivo, Sesión e Historial, las cuales reflejan los distintos procesos de recolección y análisis de datos.

Este modelo conceptual garantiza una visión clara del dominio del problema y proporciona la base para el diseño del modelo lógico que se implementará en el sistema.

#### 3.1. Diagrama ER – Modelo Conceptual

El modelo conceptual del sistema Relaxify parte de la identificación de entidades clave relacionadas con los estudiantes universitarios y sus datos sociodemográficos, hábitos de vida y resultados clínicos derivados de cuestionarios estandarizados.

Este diagrama representa las entidades principales y sus relaciones de alto nivel, sin depender aún de una tecnología específica.



### 3.2. Descripción de Entidades Principales

- ❖ **Participante**: almacena la información general y sociodemográfica del estudiante (edad, sexo, semestre, ciudad, etc.).
- ❖ **Encuesta**: contiene la información de los cuestionarios aplicados (PHQ-9, GAD-7, Rosenberg, Apgar, IPAQ, etc.), con fecha y tipo de instrumento.
- ❖ **Resultado**: refleja el puntaje obtenido en cada encuesta y su nivel de riesgo asociado (bajo, moderado, alto).
- ❖ **Hábito de Vida**: integra datos relacionados con actividad física, horas de sueño, consumo de alcohol y otras prácticas cotidianas.

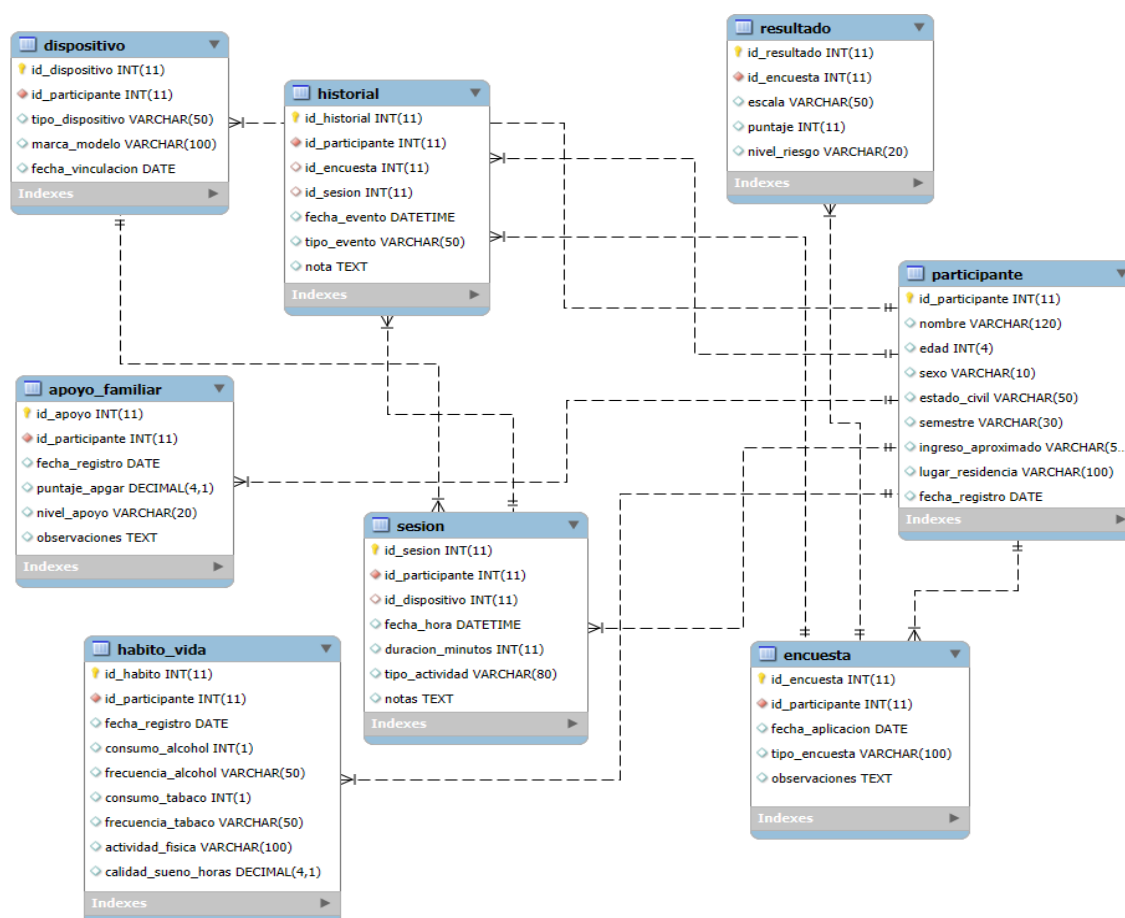


- ❖ Sesión de Bienestar: registra las actividades realizadas en el marco de la intervención (ejercicios de relajación, meditación, terapias).
- ❖ Reporte: agrupa la información en salidas individuales o grupales, útiles para estudiantes y administradores.

## 4. Modelo de Datos Lógico

El modelo de datos lógico del caso de estudio fue construido a partir del modelo conceptual previamente definido y normalizado hasta la Tercera Forma Normal (3FN). Este modelo refleja cómo la información se organiza en tablas, llaves primarias y foráneas, además de las relaciones entre ellas.

### 4.1. Diagrama ER – Modelo Lógico



Las principales decisiones tomadas fueron:

- Participante como entidad central (maestro): Todos los datos de encuestas, resultados, hábitos, sesiones y dispositivos dependen de un participante, lo cual asegura consistencia y control de los registros.
- Encuesta y Resultado como maestro–detalle: Una encuesta puede generar múltiples resultados, dependiendo de la escala aplicada (ejemplo: PHQ-9, GAD-7, APGAR).
- Hábitos de Vida separados en encabezado y detalle: Para evitar atributos multivalorados (como consumo + frecuencia), se normalizó creando una tabla `habito_vida` (encabezado) y `habito_detalle` (detalle), lo cual facilita registrar varios hábitos distintos de un mismo participante.
- Apoyo Familiar independiente: Permite almacenar puntajes y observaciones de escalas como APGAR, garantizando independencia del resto de atributos.
- Dispositivo y Sesión: Un participante puede registrar múltiples dispositivos, y cada dispositivo puede registrar múltiples sesiones de uso en la aplicación Relaxify.
- Historial como tabla de hechos: Centraliza todos los eventos asociados a un participante (encuestas, sesiones, dispositivos), actuando como bitácora unificada.

Este modelo lógico fue implementado en SQL (MySQL Workbench) con la definición de tablas, columnas, tipos de datos, llaves primarias y foráneas, además de restricciones de integridad referencial.

## 4.2. Normalización

El proceso de normalización se aplicó hasta la 3FN (Tercera Forma Normal):

1. Eliminación de redundancias en atributos (ej., datos sociodemográficos separados de resultados).
2. Separación de dependencias parciales (ej., encuesta y resultado en tablas distintas).
3. Aseguramiento de dependencia total de la clave primaria en cada tabla.
4. Esto garantiza integridad de los datos, evita duplicidades y mejora la eficiencia en consultas.

## 5. Justificación de Decisiones de Modelado y Análisis de Calidad de Datos

### 5.1. Justificación del Modelado

Las decisiones de modelado se tomaron considerando:

- ❖ Separación entre participante, encuesta y resultado: permite un historial claro y reduce redundancia.
- ❖ Inclusión de hábitos de vida: se consideró esencial para enriquecer el análisis de bienestar integral.
- ❖ Entidades independientes para reportes: facilitan la exportación y visualización de resultados sin alterar los datos primarios.

Alternativas consideradas:

- ❖ Incluir hábitos de vida directamente en la entidad participante (descartado por redundancia).
- ❖ Generar resultados como atributos de encuesta (descartado por limitación al crecimiento).

## 5.2. Análisis de Calidad de Datos

El dataset de Mendeley Data sobre salud mental en estudiantes universitarios ha tenido cuatro versiones (noviembre 2024 – mayo 2025). En cada una se han corregido errores y mejorado variables, pero aún persisten algunos desafíos que deben considerarse en el diseño de Relaxify.

➤ Principales Desafíos:

- 1) Datos duplicados: estudiantes con múltiples registros en encuestas similares.
  - Estrategia: reglas de validación y llaves compuestas (id\_participante + fecha\_encuesta).
  
- 2) Inconsistencias en respuestas: valores fuera de rango en cuestionarios.
  - Estrategia: restricciones de dominio (ej., puntaje 0–27 en PHQ-9).
  
- 3) Valores faltantes: omisión de respuestas en cuestionarios.
  - Estrategia: imputación con promedios o exclusión en análisis estadísticos.
  
- 4) Formato inconsistente en fechas: distintas formas de registrar fechas (ej. DD/MM/AAAA vs. AAAA-MM-DD).
  - Estrategia: estandarización al formato ISO (AAAA-MM-DD).

➤ Comparación entre versiones

Versión	Fechas de publicación	Observaciones principales	Persisten problemas
---------	-----------------------	---------------------------	---------------------

1	24 de Noviembre hasta el 25 de noviembre de 2024	Primera carga de datos, con varios valores faltantes y duplicados.	Faltantes y duplicados.
2	21 de febrero hasta el 27 de marzo de 2025	Corrección de errores iniciales, mejor codificación en variables.	Formatos de fecha no estandarizados.
3	26 de marzo hasta el 27 de marzo de 2025	Ajustes en escalas de cuestionarios y consolidación de datos.	Duplicados en algunos registros.
4	27 de marzo hasta el 01 mayo de 2025	Versión más completa y depurada, con mayor consistencia en atributos.	Persisten valores faltantes e inconsistencias menores.

Por lo tanto, los problemas de calidad no son exclusivos de la versión 4, sino que han estado presentes desde la versión inicial. Sin embargo, la versión 4 (mayo 2025) es la más estable y se toma como base para el presente proyecto, aplicando estrategias de control y limpieza de datos al integrarlos en la plataforma Relaxify.