DOCUMENTO REPORTE DE LOS CASOS DE ALERTA PSICOSOCIAL REGISTRADOS Y ATENDIDOS BASADO EN LOS REPORTES DE SISTEMA SATURNO Y EN LOS REGISTROS DE ATENCIÓN.

CONVENIO ESPECIAL BID SUSCRITO ENTRE EL MINISTERIO DE EDUCACIÓN NACIONAL Y EL TECNOLÓGICO DE ANTIOQUIA.

PRESENTADO POR:

DIEGO LEÓN ZAPATA RUIZ

VIGILADA MINEDUCACIÓN

DIRECCIÓN DE BIENESTAR UNIVERSITARIO

CAMPUS VITAL

MEDELLÍN

2025

Objetivo	Actividad	Indicador	Entregables	Avance
Objetivo Específico 2. Implementar mecanismos de prevención y detección temprana de riesgos psicosociales y trastornos mentales, mediante el diseño y ejecución de rutas de atención integrales, interdisciplinarias y culturalmente pertinentes, que contribuyan a la disminución de barreras para el acceso a servicios de apoyo psicoemocional. .	Actividad 1. Desarrollo de un sistema de alerta temprana para salud mental en integración con la plataforma SATURNO	Porcentaje de casos de riesgo psicosocial identificados y gestionados efectivamente en el marco del protocolo de atención	Documento técnico de diseño del módulo	60%
			Informe de implementación y puesta en marcha	40.31%

Introducción

El proyecto “Estrategia de bienestar y salud mental para la adaptación y la permanencia estudiantil” tiene como propósito fortalecer el bienestar psicoemocional y la salud mental de los estudiantes de educación superior mediante una estrategia institucional integral, alineada con las acciones del CONPES 4122 y el Contrato de Préstamo No. 5850/OC-C. En este marco, el reporte de casos de alerta psicosocial registrados y atendidos constituye un insumo clave para comprender las necesidades reales de la población estudiantil, orientar la focalización de intervenciones y evaluar la efectividad de los mecanismos de detección temprana y respuesta institucional.

El cruce de la base de estudiantes mapeados por el sistema SATURNO mediante el formulario institucional con el reporte de “Atendidos” permite:

- Identificar brechas entre el tamizaje y la atención efectiva, precisando cuántos estudiantes en situación potencialmente vulnerable logran acceder a apoyo psicosocial oportuno.
- Priorizar acciones en grupos y contextos específicos (por ejemplo, por programa académico, género o características sociodemográficas) donde se concentran los casos de atención, promoviendo entornos protectores e inclusivos.
- Retroalimentar y calibrar el instrumento de tamizaje y el índice de riesgo psicosocial, de modo que la detección temprana se corresponda con la realidad de los casos que efectivamente demandan atención.
- Fortalecer la gobernanza participativa con información transparente y accionable para las unidades responsables (bienestar universitario, programas académicos, decanaturas y colectivos estudiantiles).

En síntesis, el análisis comparativo entre estudiantes tamizados y estudiantes con atención psicosocial registrada no solo ofrece una línea base para el monitoreo del riesgo y la permanencia, sino que habilita decisiones de política interna más precisas, equitativas y orientadas a resultados, a

VIGILADA MINEDUCACIÓN

fin de prevenir factores de riesgo psicosocial, favorecer la inclusión de poblaciones diversas y consolidar circuitos de atención integrales y oportunos.

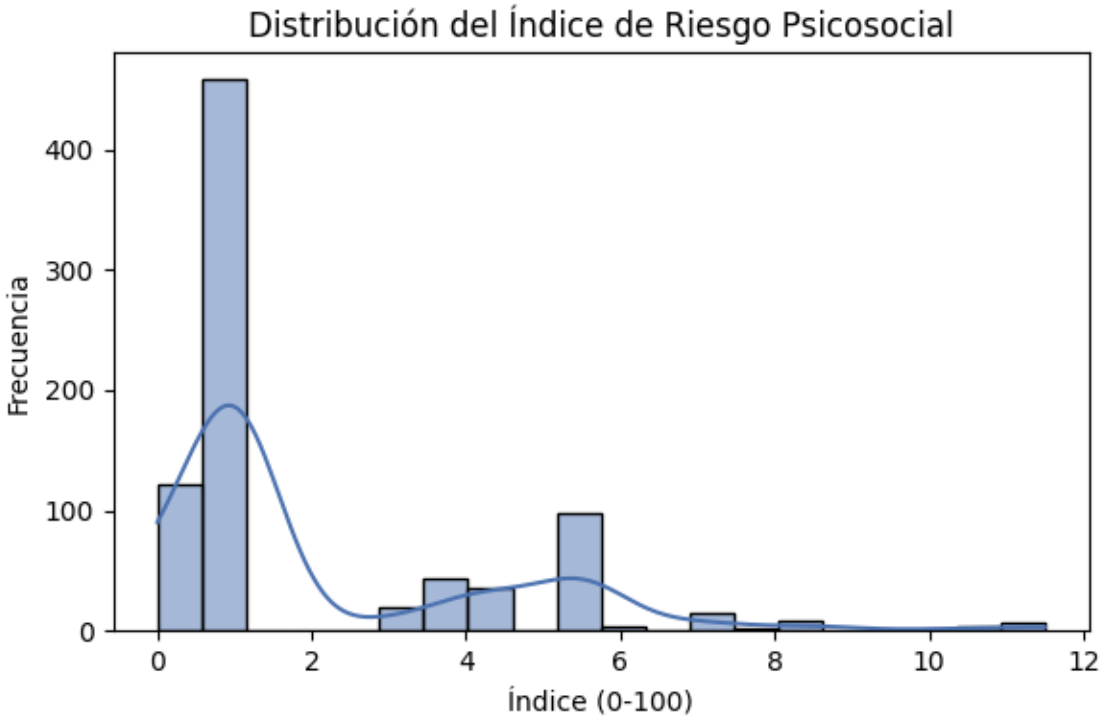
ENTREGABLE 1: Documento técnico de diseño del módulo (Ver Anexo 1.)

Informe Metodológico: Estrategia de bienestar y salud mental para la adaptación y la permanencia estudiantil

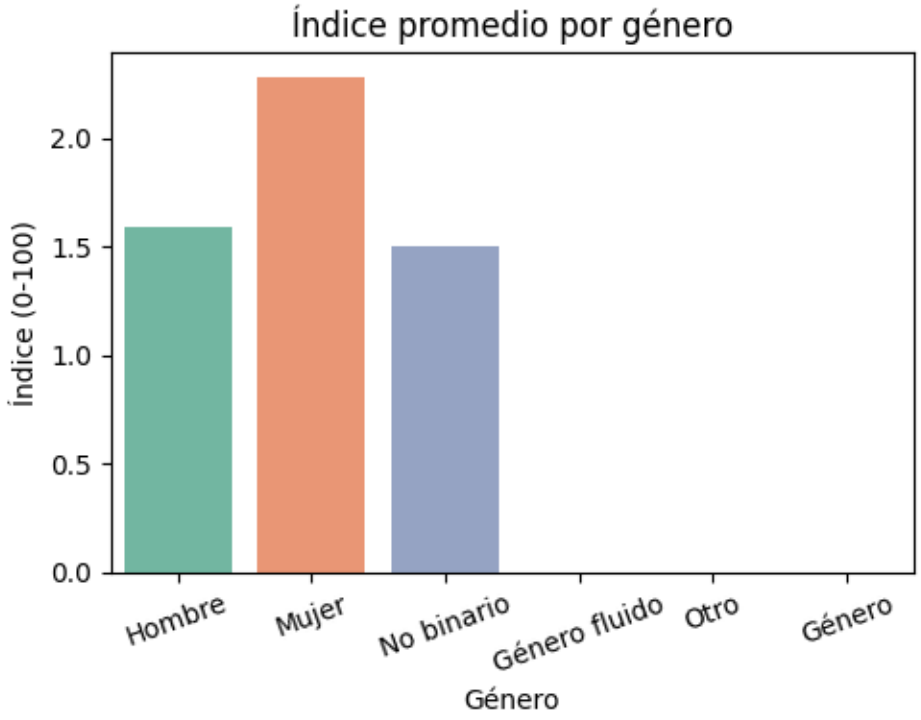
Resumen ejecutivo

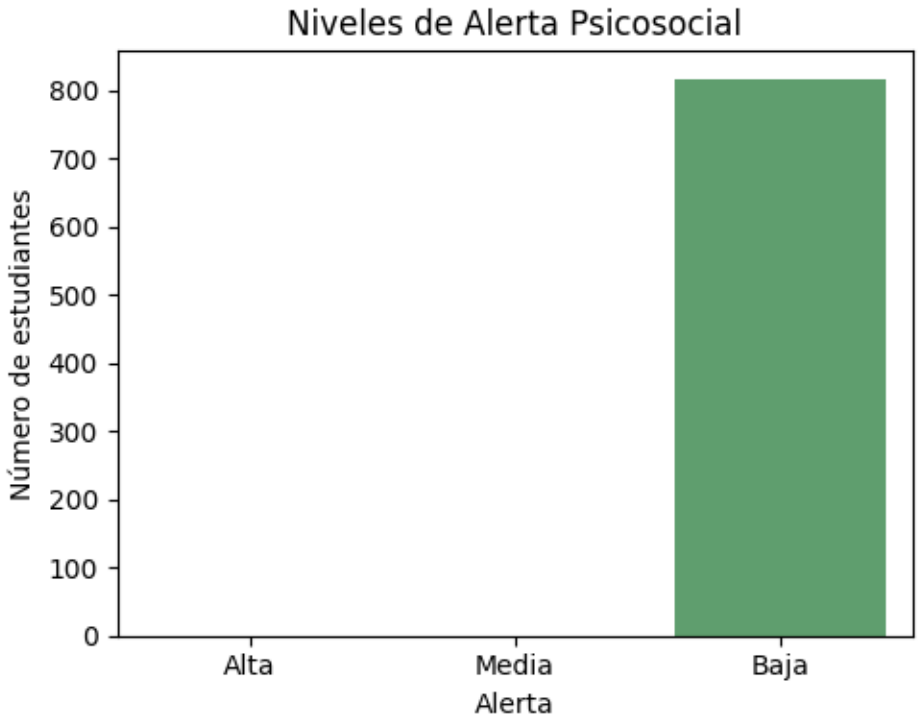
- Periodo de análisis: desde el 18 de septiembre hasta el 29 de octubre de 2025
- Total de encuestados: 817
- Total de mapeados por SATURNO: 397
- Estudiantes registrados y atendidos en reporte externo (cruce por documento): 28
- Tasa de atención sobre los Mapeados: 3.43%
- Índice compuesto de riesgo psicosocial (0–100), agregado al CSV: generado en el archivo “datos_enriquecidos_con_indice.csv” con campos clave (documento, género, edad, facultad/programa, índice, alerta, atendido).

VIGILADA MINEDUCACIÓN



- Distribución de alertas: en esta versión todos los casos quedaron clasificados como “Baja”. Esto sugiere que el instrumento o la columna “Total de puntos” no contiene variación útil o no representa riesgo; ver recomendaciones. **Ver Anexo 2.**





VIGILADA MINEDUCACIÓN
IQNET CERTIFIED MANAGEMENT SYSTEM
GOBERNACIÓN DE ANTIOQUIA
República de Colombia

Metodología

1) Preparación y homologación de datos

- Se cargaron los archivos “Datos.xlsx” (encuesta) y “Atendidos.xlsx” (reporte de casos atendidos).
- Estandarización de nombres de columnas y normalización de campos clave mediante coincidencia por subcadenas en español:
    - Documento de identidad: columnas con “documento”.
    - Nombre: “nombre completo” o “nombre”.
    - Género: columnas con “género” o “genero”.
    - Edad o rango de edad: columnas con “rango de edad” o “edad”.
    - Etnia: columnas con “etnia”.
    - Facultad/Programa: columnas con “facultad”, “programa”.
    - Puntuación global: columna con “Total de puntos” (si existe y es numérica).
- Limpieza del documento de identidad:
    - Extracción de dígitos y eliminación de ceros a la izquierda para un campo estándar documento_std.
- Homologación de género:
    - Mapeo robusto desde la variable textual de la encuesta a categorías estándar: Mujer, Hombre, No binario, Otro (con manejo tolerante de variantes “muj…”, “hom…/masc…”, “no bin…”, “otro”).

2) Cruce con casos atendidos

- En “Atendidos.xlsx” se estandarizó la cédula a documento_std y se renombró el sexo a sexo_atencion.
- Se marcó atendido = True en los registros de la encuesta cuyo documento_std aparece en el reporte de atendidos.

3) Construcción del índice compuesto de riesgo psicosocial (0–100)

En ausencia de una definición explícita del puntaje del instrumento, se construyó un índice compuesto con componentes y ponderaciones heurísticas (transparentes/reproducibles):

- Componente 1: Puntaje normalizado del instrumento
    - Si existe “Total de puntos” y es numérico: normalización min–max a [0,1].
    - Si no, alternativa: suma de columnas “Puntos: …” (si existen), convertidas a numérico y normalizadas min–max.

VIGILADA MINEDUCACIÓN

- Componente 2: Vulnerabilidad demográfica (proxy)
    - Rango de edad: menor de 18 (+0.20), mayor de 45/60 (+0.10) si el texto sugiere grupo etario mayor.
    - Etnia: minorías (afro, indígena, ROM, raizal, palenquero, etc.) (+0.15) como proxy de barreras estructurales.
    - Estado civil: separado/viudo (+0.10).
    - Ocupación: trabajador/cuidador (+0.10) por carga adicional.
    - Acotado a [0,1].
- Componente 3: Vulnerabilidad de género (proxy)
    - Mujer (+0.10), No binario (+0.15). Hombre y otros: 0. Ajuste conservador basado en evidencia sobre exposición a riesgos y barreras.
- Agregación
    - Ponderaciones por defecto (si está disponible el puntaje del instrumento): 0.6 (puntaje), 0.3 (demo), 0.1 (género). Si faltan componentes, se renormalizan los pesos de los presentes.
    - Escala final a 0–100, con cortes de alerta: Baja (<50), Media (50–69.99), Alta (≥70).

Notas metodológicas:

- El enfoque es conservador y transparente; puede afinarse con validación psicométrica del instrumento, análisis de ítems, y calibración de tramos de riesgo con datos de casos atendidos.
- Todas las transformaciones dejaron trazabilidad en el CSV resultante.

4) Análisis exploratorio de datos (EDA)

- Conteos y proporciones por alertas, género, facultad/programa.
- Tendencias del índice (distribución, sesgos).
- Comparativas entre “atendidos” y “no atendidos”.

5) Análisis de correlaciones

- Matriz de correlaciones con variables numéricas (incluidas “Puntos: …” convertidas a numérico, cuando existen).
- Selección de las 15 variables con mayor correlación absoluta con el índice para el mapa de calor.

VIGILADA MINEDUCACIÓN

Resultados clave

A. Resumen cuantitativo

- Total de estudiantes: 817.
- Atendidos según cruce de documentos: 28 (3.43%).
- Distribución de alertas:
    - Baja: 817
    - Media: 0
    - Alta: 0
- Promedios de índice por género (promedios sobre una escala 0–100):
    - Mujer: 2.28
    - Hombre: 1.59
    - No binario: 1.50
    - Género fluido, Otro: cercanos a 0 en esta base.
- Promedio por facultad (muestra que los promedios están muy bajos, ~1.7–2.5):
    - Ingeniería: ~2.25
    - Educación y Ciencias Sociales: ~1.90
    - Derecho y Ciencias Forenses: ~1.76
    - Ciencias Administrativas y Económicas: ~1.74
    - Estudiante/Otro: ~1.75–2.52

Interpretación: Los valores del índice son inusualmente bajos de forma generalizada, y todos caen en “Baja”. Esto sugiere que:

- El campo de “Total de puntos” no está capturando riesgo (p. ej., puede ser un puntaje de satisfacción, o una escala con rango muy estrecho), o
- Las columnas de puntaje no están en formato numérico utilizable (o están ausentes/constantes), o
- El instrumento aplicado no tiene variación suficiente en esta muestra.

Esto no invalida el pipeline, pero sí indica la necesidad de:

- Confirmar el significado de “Total de puntos” y su escala (rango esperado, si mayor valor implica más riesgo o lo contrario).

VIGILADA MINEDUCACIÓN

- Validar si existen variables de ítems tipo Likert por cada dimensión del riesgo psicosocial. Si están como texto, deben mapearse a números (por ejemplo, Nunca=0… Siempre=4).
- Asegurar que las columnas “Puntos: …” estén rellenas y sean numéricas.

B. Correlaciones

- La matriz de correlaciones numéricas disponibles arrojó muy poca señal. La “top” correlación con el índice fue de magnitud baja:
    - Correlación con Id: 0.05 (espuria, refleja orden de carga, no contenido).
- Esto refuerza la hipótesis de falta de variabilidad o desalineación entre las variables disponibles y la construcción del índice.

C. Comparativo “tamizados” vs. “atendidos”

- Tasa de atención entre encuestados: 3.43%.
- Conteo de alertas por condición de atención: al no haber Media/Alta, la segmentación muestra que los atendidos también quedaron en “Baja”.
- Implicación: el índice actual no discrimina a quienes ya están siendo atendidos. Se requiere calibración con datos de verdad-terreno (casos atendidos) para ajustar pesos y/o transformar ítems.

VIGILADA MINEDUCACIÓN

Identificación de casos de alerta psicosocial

- Con el índice actual, no se identificaron casos “Media” o “Alta”.
- Sin embargo, esto probablemente es un falso negativo por la naturaleza del “Total de puntos” o la ausencia de ítems numéricos.
- Recomendación inmediata:
    - Revisar diccionario de variables del instrumento para:
        - Mapear respuestas tipo Likert a escala numérica uniforme (0 a 4/5).
        - Determinar si “Total de puntos” crece con el riesgo o con el bienestar. Si es bienestar, el índice debe invertir el sentido (p. ej., riesgo = 1 – normalizado (total_bienestar)).
    - Con el mapeo correcto, reestimar el índice y reestablecer los cortes para identificar alertas medias y altas.

Planeación de las siguientes acciones de análisis

1. Validación del instrumento y escala:

- Confirmar para cada dimensión del riesgo psicosocial:
    - Las opciones de respuesta y su orden de riesgo.
    - Si existen ponderaciones institucionales específicas (dimensiones con mayor peso).
- Si “Total de puntos” es un puntaje de bienestar, invertir su contribución al riesgo.

2. Ingeniería de variables:

- Convertir todas las columnas “Puntos: …” a numéricas y estandarizar su dirección (mayor = más riesgo).
- Si las respuestas están en texto (“Nunca”, “A veces”, etc.), crear un diccionario de mapeo.
- Calcular subíndices por dimensión (estrés, ansiedad, apoyo social, violencia, consumo, ideación, etc.) para interpretabilidad.

3. Calibración con verdad-terreno:

- Usar la etiqueta “atendido” como proxy de riesgo alto histórico:
    - Ajustar pesos del índice para maximizar discriminación (p. ej., mediante regresión logística o AUC ROC).
    - Establecer umbrales de alerta que maximicen sensibilidad a casos atendidos, manteniendo precisión razonable.

4. Reporte operativo:

- Generar listados priorizados de seguimiento:
    - Top N por índice.
    - Casos con subíndices críticos (ej., ideación o violencia).
    - Brechas de atención: alta alerta y “no atendido”.

5. Gobernanza y ética:

- Reforzar consentimiento informado y principios de uso responsable.
- Minimizar sesgos por demografía; las variables demográficas deben verse como factores de barrera/acceso, no como “culpabilización”.
- Asegurar medidas de protección de datos.

VIGILADA MINEDUCACIÓN

ENTREGABLE 2: Informe de implementación y puesta en marcha

Descripción del Sistema de Alertas Tempranas Universitarias para la Retención desde el Nivel Cero - SATURNO

Se encuentra inmerso en nuestro aplicativo académico Campus y se denomina “Saturno” (Sistema de Alertas Tempranas Universitarias para la Retención desde el Nivel Cero).

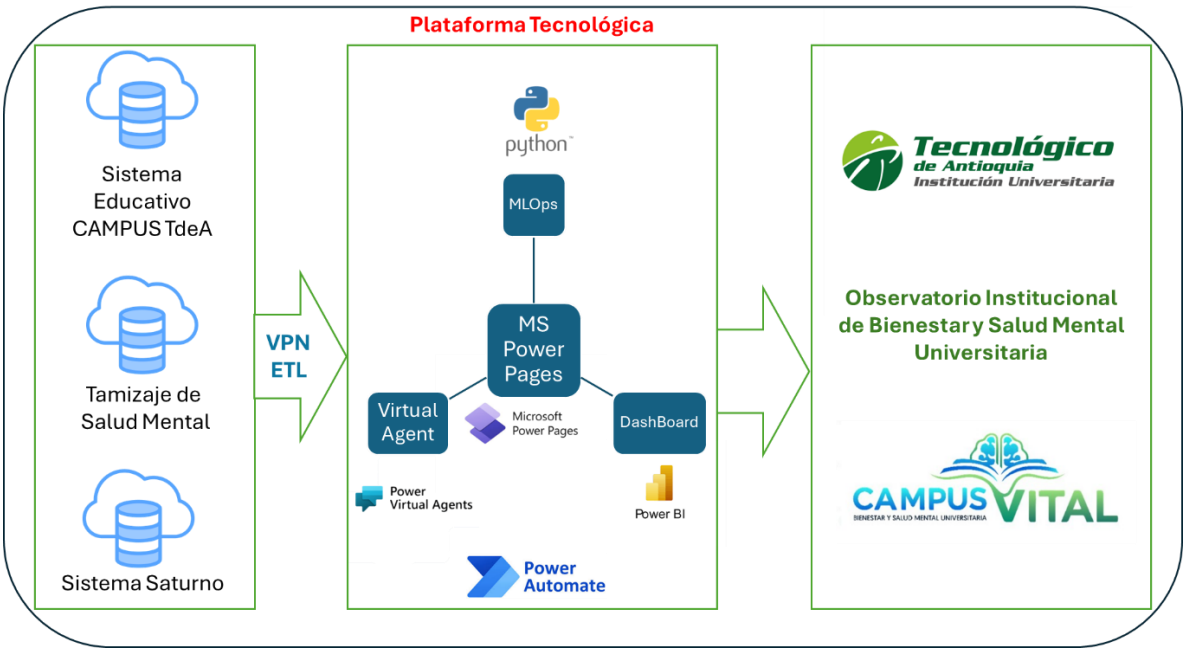
Parte de la Caracterización, proceso en el cual se recopila la información del estudiante a través del formulario de inscripción.

Esta información sumada al desarrollo académico almacenado en el aplicativo se hace visible para las diferentes instancias a través de tableros con datos que apuntan a informar situaciones de atención primaria.

Ítem	Tablero	Alerta	Proceso
1	Inscripciones	Estudiantes sin pago de inscripción	Validar desinterés o abandono de la intención. Contacto individual. Ofrecer orientación.
2		Cantidad de estudiantes inscritos por programa	Identificar inscritos vs cupos.
3	Matriculados	Estudiantes en estado prematriculado	Verificar motivos de no pago o documentación pendiente. Información sobre temas financieros para matricula.
4		Indicadores de abandono	Identificar las posibles causas.
5		Estudiantes por cohorte	Verificar si exceden el tiempo previsto para terminar el programa.
6		Edad del estudiante	Identificarlos por programa. Comunicación con padres o acudientes.
7	Estado asignaturas	Asignaturas pendientes por estudiante	Confirmar cantidad, nivel, causas.
8	Rendimiento académico	Porcentaje calificado	Validar el avance con respecto al calendario académico.
9		Nota acumulada	Verificar las notas inferiores a 3.0
10		Rango de rendimiento	Verificación del rango. Básico y Bajo.
11		Horas de inasistencia	Relacionar con el máximo exigido por el reglamento

VIGILADA MINEDUCACIÓN

			previo a la Reprobación por inasistencia.
12	Bienestar – Acude	Estado estudiante en el Acude	Validar desinterés o abandono de la intención. Contacto individual. Ofrecer orientación.
13	Poblaciones	Detalle de estudiantes por Población (10)	Validar si efectivamente aplica a una población. Promover la participación en los grupos específicos. Adaptación del curriculum. Ajustes razonables. Apoyo de intérpretes.
14		Cruce de diferentes poblaciones por estudiante	Verificar si un estudiante pertenece a varias poblaciones.



Fase 0. Gobernanza, ética y cumplimiento

- Comité del Observatorio de Bienestar y Salud Mental (subcomité del gobierno de datos): define visión, alcance, políticas y prioriza analíticas.
- Lineamientos institucionales actualizados:
  - Política de bienestar y salud mental centrada en derechos, enfoque diferencial e interseccional.
  - Protocolo de tamizaje, consentimiento y derivación; lenguaje responsable para comunicación con estudiantes.

VIGILADA MINEDUCACIÓN

- Lineamientos de analítica responsable y uso de IA: propósito, minimización, explicabilidad, revisión humana obligatoria para alertas críticas.
- Actualización de documentos de seguridad: clasificación de información, retención y destrucción, respuesta a incidentes.

Fase 1. Descubrimiento y diseño funcional (ampliada)

- Diseño técnico de solución a la medida:
  - Requisitos no funcionales: disponibilidad (>99.5%), RPO/RTO, capacidad para 22.000 estudiantes, accesibilidad (WCAG 2.1), multicanal.
  - Catálogo de integraciones: VPN, APIs del Sistema Campus, tamizajes, LMS si aplica, y servicios de Power Platform.
  - Modelo de permisos por rol y dominios de acceso (principio de mínimo privilegio).
- Diseño de la arquitectura de información:
  - Taxonomías y ontologías: dominios de bienestar (ansiedad, depresión, consumo), eventos académicos, intervenciones, casos.
  - Modelo de metadatos: definiciones, linaje, calidad, estandarización de escalas de tamizaje.
  - Mapa de ciclo de vida del dato: captura → procesamiento → analítica → alerta → intervención → retroalimentación.
- Entregables:
  - Documento de diseño técnico a la medida (DDS).
  - Arquitectura de información y diccionario de datos.
  - Prototipos de dashboards y del portal (Power Pages) con guías de estilo.

Fase 2. Integración de datos y canal seguro

- Se mantiene lo definido, agregando:
  - Acuerdos de Intercambio de Datos (Data Sharing Agreements) y Contratos de Datos (Data Contracts).
  - Procesos de seudonimización y ambientes segregados (dev/test/prod).
- Entregables:
  - Pipelines ETL/ELT versionados, reglas de calidad implementadas.

VIGILADA MINEDUCACIÓN

- Modelo semántico de Power BI con RLS documentado.

Fase 3. Ciencia de datos y MLOps

- Se mantiene lo definido, agregando:
  - Especificación de interpretabilidad operativa para equipos clínicos (reportes SHAP por caso, narrativa explicativa por señal).
  - Tablero de fairness: métricas por subgrupo y alertas de posible sesgo.
- Entregables:
  - API de scoring con SLA y playbook de reentrenamiento.
  - Informe técnico de validación y calibración por cohortes.

Fase 4. Orquestación y automatización

- Se mantiene, agregando:
  - Motor de reglas configurable por bienestar (tablas parametrizables en Dataverse/SQL).
  - Bitácora de decisiones con evidencia (qué señal/modelo activó la alerta).
- Entregables:
  - Catálogo de reglas con versionado y controles de cambio.

Fase 5. Plataforma tecnológica del Observatorio

- Objetivo: diseñar y desplegar la "Plataforma del Observatorio de Bienestar y Salud Mental" como capa institucional de inteligencia y gestión del conocimiento.
- Componentes:
  1. Repositorio institucional de conocimiento:
     - Base de conocimientos en Power Pages/SharePoint/Dataverse: protocolos, guías, bibliografía, mejores prácticas, dashboards públicos agregados.
     - Versionado de documentos, control de acceso y flujo de aprobación.
  2. Módulo analítico y de vigilancia:
     - Power BI con paneles por facultad, tendencias, mapas de calor, evaluación de intervenciones, indicadores CONPES.
     - Panel público sin datos sensibles; panel autenticado por roles.
  3. Módulo de casos e intervenciones:

VIGILADA MINEDUCACIÓN

- Gestión de casos, SLAs, trazabilidad, integración con Power Automate y calendario institucional.

4. Módulo de analítica avanzada:
    - Acceso a resultados de modelos, explicaciones y fairness.
    - Laboratorio seguro para evaluación de nuevas hipótesis y features.
5. Integración con el agente conversacional:
    - PVA como "línea de atención" con escalamiento y derivación.

- Entregables:
    - Plataforma del Observatorio operativa.
    - Catálogo de contenidos, flujos de publicación y mantenimiento.

Fase 6. Agente conversacional y línea de atención

- Como definido, con guías de estilo y mensajes normalizados por lineamientos institucionales actualizados.

Fase 7. Articulación con gestión del conocimiento

- Informe de articulación con procesos de gestión del conocimiento:
    - Mapeo de procesos existentes: bienestar, permanencia, orientación psicológica, comunicación institucional, calidad y planeación.
    - Flujo de aprendizajes: desde datos/alertas hacia diseño de intervenciones, retroalimentación a políticas y currículo.
    - Comunidades de práctica: psicólogos, tutores, decanaturas; ritmo de reuniones y tablero de lecciones aprendidas.
    - Repositorio de evidencia: intervenciones, outcomes, evaluaciones cuasi-experimentales (cuando sea posible).
- Entregables:
    - Informe de articulación y plan de acción con responsables y cronograma.
    - Procedimientos normalizados (SOPs) para captura y difusión de aprendizajes.

Fase 8. Estrategia de socialización y sostenibilidad

- Socialización:
    - Plan de comunicaciones: públicos (estudiantes, docentes, directivos), mensajes clave, calendario, canales (portal, email, redes internas), identidad visual.

VIGILADA MINEDUCACIÓN

- Jornadas de inducción y reinducción; microlearning para docentes y personal de apoyo.
- Kit de socialización: infografías, videos cortos, FAQs, manual de uso por rol.

- Sostenibilidad:
  - Modelo operativo y financiero: estimación de OPEX (nube, licencias Power Platform), roles requeridos, plan de continuidad y DRP.
  - Gobierno continuo: comité del Observatorio, indicadores de salud del sistema, auditorías semestrales de ética/fairness.
  - Plan de escalamiento y transferencia: formación de equipos internos, documentación técnica y funcional.
- Entregables:
  - Estrategia de socialización y sostenibilidad con presupuesto y cronograma.
  - Manuales operativos y de usuario.

Fase 9. Piloto, despliegue y operación continua

- Igual que antes, añadiendo:
  - Evaluación de la estrategia de socialización (alcance, comprensión, adopción).
  - Revisión de sostenibilidad post-piloto y ajustes de costos/recursos.
- Entregables:
  - Informe del piloto con lecciones y roadmap de escalamiento.

Informes técnicos periódicos

- Periodicidad sugerida: trimestral (estratégico).
- Contenido típico:
  - Salud operativa: disponibilidad, latencia, fallas, costos (FinOps).
  - Datos y modelos: calidad, drift, rendimiento del modelo, fairness y acciones correctivas.
  - Operación de alertas: volumen por criticidad, tiempos de atención, outcomes.
  - Gestión del conocimiento: nuevas lecciones e intervenciones evaluadas.
  - Riesgos y mitigaciones; cambios en lineamientos o reglas.
- Formatos:

VIGILADA MINEDUCACIÓN

- Reporte ejecutivo (2–4 páginas).
- Anexo técnico detallado (métricas, tablas, metodologías).
- Tableros Power BI vinculados.

- Entregables:
  - Plantilla institucional de informe técnico.

Documento técnico del Observatorio: descripción, alcance y estrategias

- Propuesta de índice:
  1. Resumen ejecutivo y marco CONPES 4122.
  2. Alcance del Observatorio: objetivos, población, servicios, exclusiones.
  3. Lineamientos institucionales y principios éticos.
  4. Arquitectura de información: taxonomías, metadatos, modelos de datos.
  5. Arquitectura tecnológica: red, datos, MLOps, aplicaciones y seguridad.
  6. Metodología analítica: modelos, validación, fairness, explicabilidad.
  7. Operación: roles, procesos, SOPs, motor de reglas y gestión de casos.
  8. Gobernanza y cumplimiento: RACI, auditoría, gestión de riesgos.
  9. Estrategia de socialización y sostenibilidad.
  10. Plan de monitoreo y evaluación: KPIs, SLAs, informes periódicos.
  11. Roadmap de evolución y anexos (diccionario de datos, plantillas).
- Entregable:
  - Documento técnico maestro del Observatorio (versión controlada) listo para aprobación por cuerpos colegiados.

Resumen de entregables

- Diseño técnico a la medida (DDS).
- Lineamientos institucionales actualizados.
- Arquitectura de información y diccionario de datos.
- Plataforma tecnológica del Observatorio operativa.

VIGILADA MINEDUCACIÓN

- Informe de articulación con gestión del conocimiento.
- Estrategia de socialización y sostenibilidad.
- Plantillas y calendario de informes técnicos periódicos.
- Documento técnico maestro del Observatorio con alcance y estrategias.

Anexo 1. Script de Python para el análisis estadístico

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from pathlib import Path

# Load files
path_datos = 'Datos.xlsx'
path_atendidos = 'Atendidos.xlsx'

# Read with pandas (infer first sheet)
df = pd.read_excel(path_datos)
att = pd.read_excel(path_atendidos)

# Standardize column names (strip and lower)
df.columns = [str(c).strip() for c in df.columns]
att.columns = [str(c).strip() for c in att.columns]

# Identify key fields by fuzzy matching common Spanish headers
cols = {c.lower(): c for c in df.columns}

# Try to locate identity/document and demographics by substring
col_doc = next((c for c in df.columns if 'documento' in c.lower()), None)
col_nombre = next((c for c in df.columns if 'nombre completo' in c.lower() or c.lower()== 'nombre'), None)
col_genero = next((c for c in df.columns if 'género' in c.lower() or 'genero' in c.lower()), None)
col_edad = next((c for c in df.columns if 'rango de edad' in c.lower() or 'edad' in c.lower()), None)
col_etnia = next((c for c in df.columns if 'etnia' in c.lower()), None)
col_facultad = next((c for c in df.columns if 'facultad' in c.lower()), None)
col_programa = next((c for c in df.columns if 'programa' in c.lower()), None)
col_total_puntos = next((c for c in df.columns if 'total' in c.lower() and 'punto' in c.lower()), None)

# Build a working copy
work = df.copy()

# Clean documento: keep digits only
if col_doc:
    work['documento_std'] = (work[col_doc].astype(str)
                             .str.extract(r'(\d+)')[0]
                             .str.lstrip('0'))
else:
    work['documento_std'] = np.nan

# Standardize sexo from survey if present
if col_genero:
```

VIGILADA MINEDUCACIÓN

```
    g = work[col_genero].astype(str).str.strip().str.lower()
    work['sexo_encuesta'] = np.select([
        g.str.contains('muj', na=False),
        g.str.contains('hom|masc', na=False),
        g.str.contains('no bin|nb|no-b', na=False),
        g.str.contains('otro', na=False)
    ], ['Mujer', 'Hombre', 'No binario', 'Otro'], default=np.where(g.isna(), np.nan, work[col_genero].astype(str)))
else:
    work['sexo_encuesta'] = np.nan

# Prepare atendidos
att['documento_std'] = att[att.columns[0]].astype(str).str.extract(r'(\d+)')[0].str.lstrip('0')
att['sexo_atencion'] = att[att.columns[1]].astype(str).str.strip()

# Flag atendido by matching documento
work['atendido'] = work['documento_std'].isin(set(att['documento_std']))

# Build psychosocial risk index
# Heuristic: use available Likert-type items or existing total points. If there is a total score, normalize it 0-100.
# Additionally, derive stressors from fields like estado civil, ocupación, etnia (vulnerable), genero, etc. Create components and aggregate.

components = {}

# Component 1: normalized total points (if present)
if col_total_puntos and np.issubdtype(work[col_total_puntos].dropna().infer_objects().dtype, np.number):
    s = work[col_total_puntos].astype(float)
    s_min, s_max = s.min(), s.max()
    if s_max > s_min:
        comp1 = (s - s_min) / (s_max - s_min)
    else:
        comp1 = pd.Series(0.0, index=work.index)
    components['puntaje_normalizado'] = comp1
else:
    # Fallback: count of affirmative risk answers by scanning columns
    # Assuming items with 'Puntos:' columns represent scores
    puntos_cols = [c for c in df.columns if 'Puntos:' in c]
    if puntos_cols:
        s = work[puntos_cols].apply(pd.to_numeric, errors='coerce').sum(axis=1)
        s_min, s_max = s.min(), s.max()
        comp1 = (s - s_min) / (s_max - s_min) if s_max > s_min else pd.Series(0.0, index=work.index)
        components['puntaje_normalizado'] = comp1

# Component 2: demographic vulnerability proxy
vuln = pd.Series(0.0, index=work.index)
# Age range: mark adolescents/older adults if present
if col_edad:
    e = work[col_edad].astype(str).str.lower()
    vuln += np.where(e.str.contains('menor|<\s*18', na=False), 0.2, 0.0)
    vuln += np.where(e.str.contains('>\s*45|mayor|60', na=False), 0.1, 0.0)
# Etnia: minority groups add risk due to potential discrimination/access barriers
if col_etnia:
    et = work[col_etnia].astype(str).str.lower()
    vuln += np.where(et.str.contains('afro|neg|palenq|raizal|ind[ií]gena|rom|gitano', na=False), 0.15, 0.0)
# Estado civil: separados/viudos add slight
col_estado = next((c for c in df.columns if 'estado civil' in c.lower()), None)
```

VIGILADA MINEDUCACIÓN

```
if col_estado:
    es = work[col_estado].astype(str).str.lower()
    vuln += np.where(es.str.contains('separ|viud', na=False), 0.1, 0.0)
# Ocupación: trabajadores y estudiantes con cuidado aumentan carga
col_ocup = next((c for c in df.columns if 'ocupación' in c.lower() or 'ocupacion' in c.lower()), None)
if col_ocup:
    oc = work[col_ocup].astype(str).str.lower()
    vuln += np.where(oc.str.contains('trab|cuidad', na=False), 0.1, 0.0)
# Facultad/Programa not used in score directly
components['vulnerabilidad_demo'] = np.minimum(vuln, 1.0)

# Component 3: gender-based vulnerability (women and non-binary may face higher risk in some contexts)
sex = work['sexo_encuesta'].fillna('')
comp3 = np.where(sex=='Mujer', 0.1, np.where(sex=='No binario', 0.15, 0.0))
components['vulnerabilidad_genero'] = pd.Series(comp3, index=work.index)

# Aggregate index with weights (sum and scale to 0-100)
comp_df = pd.DataFrame(components)
if not comp_df.empty:
    # default weights: 0.6 score, 0.3 demo, 0.1 gender if available
    weights = {
        'puntaje_normalizado': 0.6 if 'puntaje_normalizado' in comp_df else 0.0,
        'vulnerabilidad_demo': 0.3,
        'vulnerabilidad_genero': 0.1
    }
    # Normalize weights to sum to 1 for present components
    w_present = {k: v for k, v in weights.items() if k in comp_df}
    sw = sum(w_present.values()) if w_present else 1.0
    for k in w_present:
        w_present[k] = w_present[k] / sw
    idx = sum(comp_df[k] * w_present[k] for k in w_present)
    risk_index = (idx.clip(0,1) * 100).round(2)
else:
    risk_index = pd.Series(np.nan, index=work.index)

work['indice_riesgo_psicosocial'] = risk_index

# Define alert thresholds
# High: >= 70, Medium: 50-69.99, Low < 50
work['alerta'] = pd.cut(work['indice_riesgo_psicosocial'], bins=[-np.inf, 50, 70, np.inf], labels=['Baja', 'Media', 'Alta'])

# Correlation analysis: select numeric columns, including converted Likert scores
num_df = work.select_dtypes(include=[np.number]).copy()
# Try to coerce 'Puntos:' columns to numeric and include them
puntos_cols = [c for c in df.columns if 'Puntos:' in c]
for c in puntos_cols:
    num_df[c] = pd.to_numeric(work.get(c), errors='coerce')

corr = num_df.corr(numeric_only=True).round(2)

# EDA summaries
n = len(work)
by_alert = work['alerta'].value_counts(dropna=False)
by_sex = work.groupby('sexo_encuesta', dropna=False)['indice_riesgo_psicosocial'].mean().round(2)
by_fac = work.groupby(col_facultad, dropna=False)['indice_riesgo_psicosocial'].mean().round(2) if col_facultad else pd.Series()
```

VIGILADA MINEDUCACIÓN

```
# Comparative analysis with atendidos
matched = work[work['atendido']]
unmatched = work[~work['atendido']]
att_rate_overall = matched.shape[0] / n if n else np.nan
att_by_alert = matched['alerta'].value_counts().reindex(['Alta','Media','Baja']).fillna(0)

# Save outputs: figures and enriched CSV
out_dir = Path('.')

# Plot 1: Distribution of risk index
plt.figure(figsize=(6,4))
sns.histplot(work['indice_riesgo_psicosocial'].dropna(), bins=20, kde=True, color='#4C72B0')
plt.title('Distribución del Índice de Riesgo Psicosocial')
plt.xlabel('Índice (0-100)')
plt.ylabel('Frecuencia')
plt.tight_layout()
plt.show()

# Plot 2: Alert levels count
plt.figure(figsize=(5,4))
order = ['Alta','Media','Baja']
sns.countplot(x='alerta', data=work, order=order, palette=['#C44E52','#DD8452','#55A868'])
plt.title('Niveles de Alerta Psicosocial')
plt.xlabel('Alerta')
plt.ylabel('Número de estudiantes')
plt.tight_layout()
plt.show()

# Plot 3: Mean risk by sexo
if work['sexo_encuesta'].notna().any():
    plt.figure(figsize=(5,4))
    sns.barplot(x='sexo_encuesta', y='indice_riesgo_psicosocial', data=work, estimator=np.nanmean, ci=None,
palette='Set2')
    plt.title('Índice promedio por género')
    plt.xlabel('Género')
    plt.ylabel('Índice (0-100)')
    plt.xticks(rotation=20)
    plt.tight_layout()
    plt.show()

# Plot 4: Heatmap correlations (top 15 vars most correlated with risk)
if not corr.empty:
    target_corr = corr.get('indice_riesgo_psicosocial')
    if target_corr is not None:
        top_vars = target_corr.abs().sort_values(ascending=False).head(15).index.tolist()
        subcorr = corr.loc[top_vars, top_vars]
        plt.figure(figsize=(8,6))
        sns.heatmap(subcorr, annot=True, fmt='.2f', cmap='vlag', center=0)
        plt.title('Matriz de correlaciones (top variables)')
        plt.tight_layout()
        plt.show()
# Plot 5: Atendidos vs no atendidos by alerta
plt.figure(figsize=(6,4))
sns.countplot(x='alerta', hue='atendido', data=work, order=order, palette={True:'#4C72B0', False:'#A1A9B1'})
```

VIGILADA MINEDUCACIÓN

```
plt.title('Alerta por condición de atención')
plt.xlabel('Alerta')
plt.ylabel('Conteo')
plt.tight_layout()
plt.show()

# Export enriched CSV
export_cols = [col for col in [col_doc, col_nombre, col_genero, col_edad, col_facultad, col_programa] if col]
export = work[export_cols + ['documento_std','sexo_encuesta','indice_riesgo_psicosocial','alerta','atendido']].copy()
export.to_csv('datos_enriquecidos_con_indice.csv', index=False)
print('Archivo generado: datos_enriquecidos_con_indice.csv')

# Prepare textual summary
summary = {
    'n_estudiantes': n,
    'atendidos_en_survey': int(matched.shape[0]),
    'tasa_atencion': round(att_rate_overall*100,2) if pd.notna(att_rate_overall) else np.nan,
    'distrib_alerta': by_alert.to_dict(),
    'riesgo_promedio_por_genero': by_sex.dropna().to_dict(),
    'riesgo_promedio_por_facultad': by_fac.dropna().to_dict() if not by_fac.empty else {},
}

print('RESUMEN_JSON_START')
print(summary)
print('RESUMEN_JSON_END')

# Show top correlations with index
if not corr.empty and 'indice_riesgo_psicosocial' in corr.columns:
    tc = corr['indice_riesgo_psicosocial'].drop('indice_riesgo_psicosocial',
errors='ignore').dropna().sort_values(key=np.abs, ascending=False).head(10)
    print('TOP_CORR_WITH_INDEX_START')
    print(tc.to_string())
    print('TOP_CORR_WITH_INDEX_END')
```

VIGILADA MINEDUCACIÓN