



Anexo 1. Documento Técnico del Conjunto de Datos

1. Información General

- Título del Dataset: [Nombre representativo del conjunto de datos]
- Autores / Equipo de Trabajo: [Nombres, cargos, unidades académicas, correos institucionales]
- Fecha de Publicación: [dd/mm/aaaa]
- Repositorio de Publicación: [URL al repositorio institucional o externo]
- Identificador Persistente (DOI): [Si aplica]
- Licencia de Uso: [Creative Commons CC-BY / CC0 / otra especificar]

2. Descripción General del Dataset

- Objetivo del Conjunto de Datos: Explicar brevemente para qué fue creado y qué fenómeno representa.
- Ámbito Temático: Ej. salud pública, educación, agricultura de precisión, etc.
- Cobertura Geográfica: Municipios, departamentos, países incluidos.
- Cobertura Temporal: Rango de fechas de los datos recolectados.
- Tamaño del Dataset: Número de registros, número de variables, tamaño en MB/GB.
- Idioma principal: Español / Inglés / Ambos

3. Origen y Trazabilidad de los Datos

- Fuente de los Datos Originales: Nombre del proyecto, sistema de información, encuesta, archivo institucional, etc.
- Condiciones de Acceso o Uso Original: Licencias, permisos éticos, convenios, etc.
- Transformaciones realizadas al dataset original: Explicar si se consolidaron fuentes, integraron bases, reformatearon campos, etc.

4. Proceso de Curación y Transformación

- Limpieza de datos realizada: Tratamiento de nulos, outliers, errores tipográficos, formatos inconsistentes, etc.
- Estandarización de variables: Normalización de nombres, unificación de unidades, etc.
- Anotación Semántica (si aplica): Uso de vocabularios controlados, ontologías, codificación estándar.



- Enriquecimiento del Dataset: Nuevas variables calculadas, fusión con otras fuentes, derivación de categorías.

5. Diccionario de Datos (Estructura del Dataset)

- Adjuntar tabla con: Nombre de Variable, Tipo de Dato, Descripción, Valores posibles / Unidad, Notas.
- Adjuntar como anexo completo en formato CSV o XLSX si supera 20 variables.

6. Cumplimiento de Principios FAIR

- Findable: Se asignó un DOI y se registró en repositorio visible públicamente.
- Accessible: Se permite descarga directa bajo licencia abierta.
- Interoperable: Formato CSV; variables anotadas con vocabularios estándar.
- Reusable: Documentación completa; se incluye licencia y limitaciones conocidas.

7. Consideraciones Éticas y Legales

- Anonimización: Indicar si los datos fueron desidentificados. Métodos aplicados.
- Consentimiento informado: Especificar si fue requerido y cómo se documentó.
- Restricciones éticas: Ej. uso restringido a ciertos fines, datos sensibles.

8. Análisis Exploratorio del Dataset

- Resumen estadístico básico: Promedios, medianas, desviaciones estándar, frecuencias, etc.
- Visualizaciones incluidas: Gráficos de barras, histogramas, mapas, diagramas de caja, etc.
- Hallazgos preliminares: Observaciones relevantes sobre la estructura o distribución de los datos.

9. Limitaciones del Dataset

- Cobertura limitada o sesgos conocidos
- Datos faltantes o incompletos en ciertas variables
- Condiciones técnicas de captura que puedan afectar interpretación

10. Usos Potenciales del Dataset

- Escenarios de entrenamiento de IA: Ej. modelos predictivos de abandono escolar, clasificación de cultivos, etc.



**Universidad
de Santander**
UDES



- Casos de uso futuros sugeridos: Estudios epidemiológicos, modelos de planificación, dashboards municipales, etc.
- Recomendaciones para su reutilización: Buenas prácticas o advertencias para futuros investigadores.