

Informe de Análisis de Datos de Salud Mental

1. Análisis descriptivo inicial

Dimensiones y estructura: El dataset contiene **21,210** filas y **111** columnas.

Tipos de variables:

- Fechas: Fecha de nacimiento, Fecha de Ingreso, Fecha de Fin Contacto, etc.
- Carácter/Textos: Nombre, Diagnóstico Principal, Servicio, Centro Recodificado, etc.
- Categóricas: Comunidad Autónoma, Categoría, Procedimiento 9-20, POA Diagnóstico 1-20, etc.
- Numéricas: Sexo, Edad, Edad en Ingreso, Estancia Días, Reingreso, Coste APR, etc.

Valores nulos o desconocidos: Algunas columnas presentan valores nulos en todos los registros (ejemplo: CCAA Residencia, Procedimiento Externo 6). Otras columnas clave no presentan nulos (ejemplo: Comunidad Autónoma, Nombre, Fecha de nacimiento, Sexo, Edad en Ingreso).

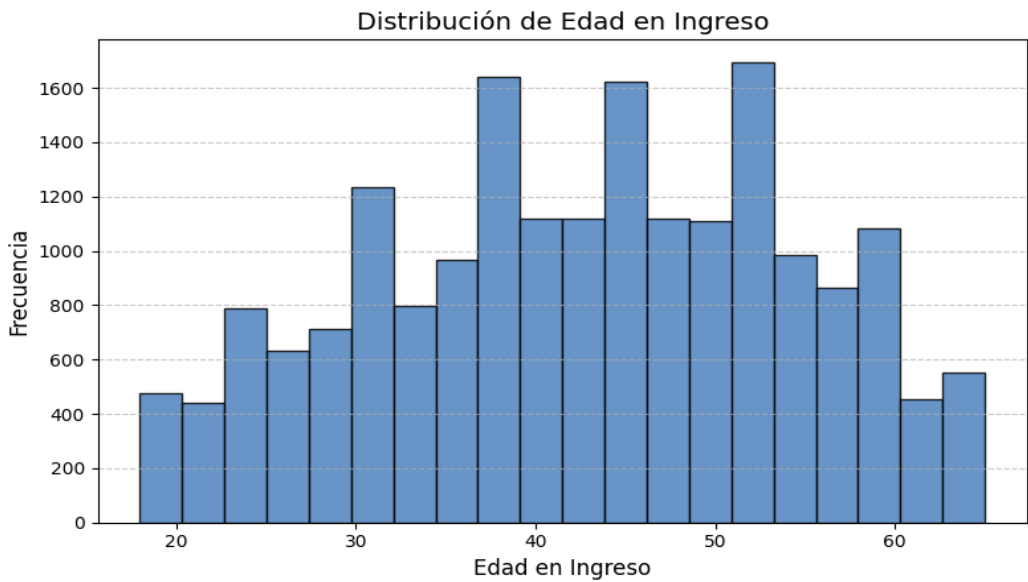
Estadísticos básicos:

Variable	Media	Mínimo	Máximo	Desviación estándar
Edad en Ingreso	43.7	0	96	14.1
Sexo	1.45	1	9	0.56

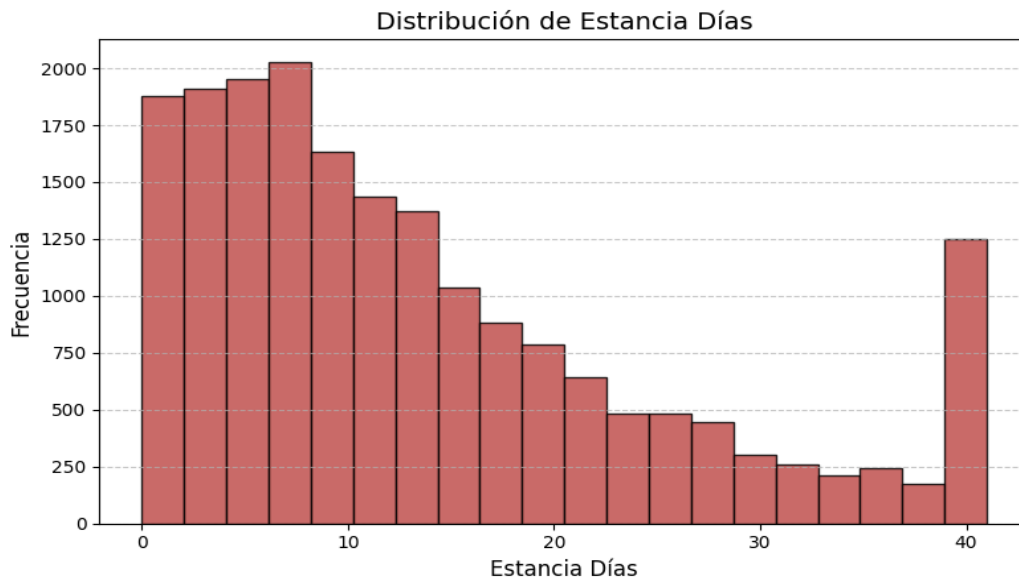
Outliers:

- Estancia Días: 1,229 posibles outliers
- Edad en Ingreso: 149 posibles outliers
- Otros: Coste APR, Riesgo Mortalidad APR, etc.

Gráficas:



Distribución de Edad en Ingreso



Distribución de Estancia Días

2. Ingeniería de características

Creación de nuevas variables:

- Edad agrupada: Agrupar Edad en Ingreso en rangos (ejemplo: 0-18, 19-30, 31-50, 51+)
- Año de ingreso: Extraer el año de la columna Fecha de Ingreso
- Duración estancia categorizada: Categorizar Estancia Días (ejemplo: corta, media, larga)

Transformaciones:

- Normalización y centrado: Aplicar Z-score a Edad en Ingreso y Estancia Días
- Winsorización: Limitar valores extremos en variables continuas

Codificación:

- Sexo binario: 1=Hombre, 2=Mujer → 0/1
- One-hot encoding: Para Comunidad Autónoma, Categoría, etc.

Este informe resume el análisis inicial y las propuestas de ingeniería de variables para el dataset de salud mental. Para detalles completos, consultar el archivo analisis_salud_mental.txt y los scripts de procesamiento.