

■ INFORMACIÓN GENERAL DEL DATASET

- Dimensiones: 248,569 filas x 15 columnas
- Memoria utilizada: 39.15 MB
- Tamaño promedio por registro: 0.16 KB

■ COLUMNAS DEL DATASET:

1. CUM | String
2. PRODUCTO | String
3. EXPEDIENTE CUM | Int64
4. ATC | String
5. DESCRIPCIÓN_ATC | String
6. VÍA ADMINISTRACIÓN | String
7. PRINCIPIO ACTIVO | String
8. FORMA FARMACÉUTICA | String
9. CANTIDAD CUM | Float64
10. CANTIDAD | Float64
11. UNIDAD MEDIDA | String
12. VALIDO | Int8
13. ESTADO REGISTRO | String
14. ESTADO CUM | String
15. MUESTRA MÉDICA | String

■ TIPOS DE DATOS:

{'String': 11, 'Int64': 1, 'Float64': 2, 'Int8': 1}

■ ANÁLISIS DE CALIDAD DE DATOS

■ VALORES NULOS POR COLUMNA:

- No hay valores nulos en el dataset

■ CARDINALIDAD POR COLUMNA:

- CUM : 155,341 únicos (62.5%) [ALTA_CARD]
- PRODUCTO : 32,180 únicos (12.9%) [MEDIA_CARD]
- EXPEDIENTE CUM : 34,653 únicos (13.9%) [MEDIA_CARD]
- ATC : 2,419 únicos (1.0%) [BAJA_CARD]
- DESCRIPCIÓN_ATC : 2,149 únicos (0.9%) [BAJA_CARD]
- VÍA ADMINISTRACIÓN : 51 únicos (0.0%) [BAJA_CARD]
- PRINCIPIO ACTIVO : 20,321 únicos (8.2%) [BAJA_CARD]
- FORMA FARMACÉUTICA : 112 únicos (0.0%) [BAJA_CARD]
- CANTIDAD CUM : 683 únicos (0.3%) [BAJA_CARD]
- CANTIDAD : 3,828 únicos (1.5%) [BAJA_CARD]
- UNIDAD MEDIDA : 180 únicos (0.1%) [BAJA_CARD]

■ VALIDO : 2 únicos (0.0%) [BAJA_CARD]
 ■ ESTADO REGISTRO : 14 únicos (0.0%) [BAJA_CARD]
 ■ ESTADO CUM : 2 únicos (0.0%) [BAJA_CARD]
 ■ MUESTRA MÉDICA : 2 únicos (0.0%) [BAJA_CARD]

■ ANÁLISIS DE VALIDEZ DE MEDICAMENTOS

■ Criterios de validez:

■ ESTADO REGISTRO = 'Vigente'
 ■ ESTADO CUM = 'Activo'
 ■ MUESTRA MÉDICA = 'No'

■ DISTRIBUCIÓN DE VALIDEZ:

■ Medicamentos VÁLIDOS: 62,006 (24.9%)
 ■ Medicamentos INVÁLIDOS: 186,563 (75.1%)

■ ANÁLISIS DE MOTIVOS DE INVALIDEZ:

■ Estado Registro ≠ 'Vigente': 149,072
 ■ Estado CUM ≠ 'Activo': 158,891
 ■ Es Muestra Médica: 35,895

■ ANÁLISIS DE SOLAPAMIENTO:

■ Solo por Estado Registro: 7,512
 ■ Solo por Estado CUM: 16,085
 ■ Solo por Muestra Médica: 18,329

■■ ANÁLISIS DE VARIABLES CATEGÓRICAS PRINCIPALES

■ ANÁLISIS DE: PRINCIPIO ACTIVO

■ Total únicos: 20,321
 ■ Valores nulos: 0
 ■ Únicos en válidos: 6,075
 ■ Únicos en inválidos: 18,731
 ■ TOP 5 valores más frecuentes:

1. ACETAMINOFEN	3,572 (1.4%)
2. IBUPROFENO	3,192 (1.3%)
3. HIDROCLOROTIAZIDA	2,808 (1.1%)
4. CLORURO DE SODIO	2,216 (0.9%)
5. FENILEFRINA CLORHIDRATO	2,016 (0.8%)

■ ANÁLISIS DE: ATC

```
=====
■ Total únicos: 2,419
■ Valores nulos: 0
■ Únicos en válidos: 1,568
■ Únicos en inválidos: 2,303
■ TOP 5 valores más frecuentes:
  1. A11AA03          | 15,814 ( 6.4%)
  2. N02BE51          |  8,577 ( 3.5%)
  3. A07CA99          |  5,208 ( 2.1%)
  4. B05BA01          |  3,692 ( 1.5%)
  5. M01AE51          |  3,093 ( 1.2%)
=====
```

■ ANÁLISIS DE: VÍA ADMINISTRACIÓN

```
=====
```

```
■ Total únicos: 51
■ Valores nulos: 0
■ Únicos en válidos: 47
■ Únicos en inválidos: 49
■ TOP 5 valores más frecuentes:
  1. ORAL              | 173,015 ( 69.6%)
  2. INTRAVENOSA       |  18,369 (  7.4%)
  3. TOPICA (EXTERNA)  |  12,274 (  4.9%)
  4. INTRAMUSCULAR     |  10,406 (  4.2%)
  5. PARENTERAL        |   8,486 (  3.4%)
=====
```

■ ANÁLISIS DE: FORMA FARMACÉUTICA

```
=====
```

```
■ Total únicos: 112
■ Valores nulos: 0
■ Únicos en válidos: 104
■ Únicos en inválidos: 112
■ TOP 5 valores más frecuentes:
  1. TABLETA RECUBIERTA |  37,540 ( 15.1%)
  2. TABLETA            |  37,008 ( 14.9%)
  3. SOLUCION INYECTABLE |  27,820 ( 11.2%)
  4. CAPSULA DURA       |  18,483 (  7.4%)
  5. CAPSULA BLANDA      |  14,681 (  5.9%)
=====
```

■ ANÁLISIS DE: UNIDAD MEDIDA

```
=====
```

```
■ Total únicos: 180
■ Valores nulos: 0
```

- Únicos en válidos: 85
- Únicos en inválidos: 160
- TOP 5 valores más frecuentes:

1. mg		181,882 (73.2%)
2. g		42,554 (17.1%)
3. mcg		4,942 (2.0%)
4. IU		4,063 (1.6%)
5. æg		3,227 (1.3%)

■ ANÁLISIS DE: CANTIDAD CUM

- Total únicos: 683
- Valores nulos: 0
- Únicos en válidos: 284
- Únicos en inválidos: 650
- TOP 5 valores más frecuentes:

1. 1.0		30,214 (12.2%)
2. 10.0		24,618 (9.9%)
3. 30.0		22,889 (9.2%)
4. 100.0		14,667 (5.9%)
5. 20.0		14,194 (5.7%)

■ ANÁLISIS DE COBERTURA PARA HOMOLOGACIÓN

- Objetivo: Determinar qué variables permiten mayor cobertura de homologación

- Verificando filtros de validez...
- Medicamentos VÁLIDOS encontrados: 62,006
- Medicamentos INVÁLIDOS encontrados: 186,563
- Total verificado: 248,569

■ COBERTURA - PRINCIPIO ACTIVO:

- Valores únicos en válidos: 6,075
- Valores únicos en inválidos: 18,731
- Valores comunes: 4,485
- Cobertura por valores: 23.9%
- Cobertura por registros: 61.7%

■ COBERTURA - ATC:

- Valores únicos en válidos: 1,568
- Valores únicos en inválidos: 2,303
- Valores comunes: 1,452
- Cobertura por valores: 63.0%
- Cobertura por registros: 92.2%

■ COBERTURA - VÍA ADMINISTRACIÓN:

- Valores únicos en válidos: 47
- Valores únicos en inválidos: 49
- Valores comunes: 45
- Cobertura por valores: 91.8%
- Cobertura por registros: 99.6%

■ COBERTURA - FORMA FARMACÉUTICA:

- Valores únicos en válidos: 104
- Valores únicos en inválidos: 112
- Valores comunes: 104
- Cobertura por valores: 92.9%
- Cobertura por registros: 99.9%

■ ANÁLISIS DE COMBINACIONES DE VARIABLES

■ COMBINACIÓN: PRINCIPIO ACTIVO + VÍA ADMINISTRACIÓN

- Combinaciones únicas en válidos: 7,074
- Combinaciones únicas en inválidos: 21,639
- Combinaciones comunes: 5,114
- Cobertura: 56.9%

■ COMBINACIÓN: ATC + VÍA ADMINISTRACIÓN

- Combinaciones únicas en válidos: 2,119
- Combinaciones únicas en inválidos: 3,741
- Combinaciones comunes: 1,881
- Cobertura: 85.4%

■ COMBINACIÓN: ATC + VÍA ADMINISTRACIÓN + FORMA FARMACÉUTICA

- Combinaciones únicas en válidos: 3,297
- Combinaciones únicas en inválidos: 7,452
- Combinaciones comunes: 2,844
- Cobertura: 68.0%

■ RECOMENDACIONES PARA EL MODELO

■ VARIABLE CON MEJOR COBERTURA INDIVIDUAL:

- FORMA FARMACÉUTICA: 99.9%

■ COMBINACIÓN CON MEJOR COBERTURA:

ATC + VÍA ADMINISTRACIÓN: 85.4%

■ CLASIFICACIÓN DE VARIABLES:

CRÍTICAS ($\geq 80\%$ cobertura): ['ATC', 'VÍA ADMINISTRACIÓN', 'FORMA FARMACÉUTICA']

IMPORTANTES ($\geq 50\%$ cobertura): ['PRINCIPIO ACTIVO']

OPCIONALES ($< 50\%$ cobertura): []

■ ESTRATEGIA RECOMENDADA:

Usar filtro: ATC + VÍA ADMINISTRACIÓN

=====

■ RESUMEN EJECUTIVO DEL EDA

=====

■ DATOS CLAVE:

- Dataset: 248,569 medicamentos, 15 variables
- Medicamentos válidos: 62,006 (24.9%)
- Medicamentos inválidos: 186,563 (75.1%)

■ HALLAZGOS PRINCIPALES:

- Mejor variable individual: FORMA FARMACÉUTICA (99.9% cobertura)
- Mejor combinación: ATC + VÍA ADMINISTRACIÓN (85.4% cobertura)
- Variables críticas identificadas: 3