**DICCIONARIO DE VARIABLES BD PREDICCION DE CAÑA BD\_IPSA\_1940**

|  |  |
| --- | --- |
| **VARIABLE** | **DESCRIPCION** |
| **NOME** | Nombre del registro, que puede corresponder al nombre de la finca, campo o unidad de análisis |
| **FAZ** | Código o identificador numérico de la hacienda o finca |
| **TAL** | identificador de una subunidad o lote |
| **TIPOCORTE** | Tipo de corte realizado, lo que puede indicar la modalidad o método de cosecha |
| **VARIEDAD** | Variedad del cultivo, importante para evaluar diferencias de rendimiento y calidad entre tipos. |
| **MADURADA** | Indicador relacionado con la maduración del cultivo. |
| **PRODUCTO** |  |
| **DOSISMAD** | Dosis aplicada de fertilizantes o agroquímicos |
| **SEMSMAD** | Dosis aplicada de fertilizantes o agroquímicos. |
| **EDAD** | Edad del cultivo o de la planta en el momento del registro (puede expresarse en días, semanas o meses). |
| **CORTES** | Número de cortes o cosechas realizadas en el periodo evaluado |
| **ME** |  |
| **VEJEZ** | Indicador relacionado con la “edad” o “antigüedad” del cultivo, que puede influir en la calidad o el rendimiento. |
| **SACAROSA** | Porcentaje o cantidad de sacarosa (azúcar) presente en el producto, fundamental en cultivos azucareros. |
| **MES** | Mes en el que se realizó el registro o la cosecha |
| **PERIODO** | Periodo o año del registro, útil para análisis de tendencias temporales |
| **TCH** | Toneladas de caña por hectárea. |
| **LLUVIAS** | Cantidad de lluvias registradas en el periodo (milímetros). |
| **GRUPO TENENCIA** | Código o categoría que clasifica la forma de tenencia del terreno. |
| **PCT \_DIATREA** | Porcentaje de infestación por diatrea (una plaga o condición específica), que puede afectar la salud del cultivo |

**BD\_IPSA\_1940**

**1- Verificar y Manejar Valores Nulos**: Revisa el porcentaje de datos faltantes en cada columna. Para columnas con altos porcentajes de nulos (por ejemplo, algunas variables de insumos o climatológicas), considera si es más adecuado imputar valores, agruparlos en una categoría “Desconocido” o eliminarlas si no aportan información valiosa.

**2 - Creación de Variables Derivadas:** Puedes generar nuevas columnas que integren información de dos o más variables existentes.

**DICCIONARIO DE VARIABLES BD PREDICCION DE CAÑA HISTORICO\_SUERTES**

|  |  |
| --- | --- |
| **VARIABLE** | **DESCRIPCION** |
| **PERÍODO** | Identificador del periodo o año de registro |
| **HACIENDA** | Código o identificador de la hacienda (finca o plantación). |
| **NOMBRE** | Nombre o denominación asignada a la unidad de análisis (puede ser el nombre del campo, plantación o cultivo). Ejemplo de uso: Identificar de forma descriptiva cada registro. |
| **ZONA** | Región geográfica o zona donde se ubica la hacienda o cultivo |
| **TENENCIA** | Indica la forma de tenencia de la tierra (propiedad, arrendamiento, etc.). |
| **SUERTE** | Identificador de la suerte o lote específico dentro de la hacienda. |
| **SUELO** | Tipo de suelo presente en la parcela (arenoso, arcilloso, etc.). |
| **AREA NETA** | Área útil o neta destinada al cultivo (en hectáreas o metros cuadrados). |
| **DIST KM** | Distancia en kilómetros (posiblemente desde la hacienda hasta una planta procesadora u otro punto de interés). |
| **VARIEDAD** | Variedad del cultivo (por ejemplo, tipo de caña de azúcar). Comparación del rendimiento y calidad según la variedad. |
| **COD.ESTADO # Y COD.ESTADO** | Códigos que representan el estado o situación del cultivo, ya sea numérico o en formato texto. |
| **F.SIEMBRA** | Fecha de siembra del cultivo. |
| **ULT.RIEGO** | Fecha en la que se realizó el último riego |
| **EDAD ULT COS** | Edad del cultivo en el momento de la última cosecha (puede expresarse en días o semanas). |
| **DESTINO 1=SEMILLA** | Indicador del destino del cultivo; por ejemplo, si se destinó a semilla u otro uso. |
| **COD. T.CULTIVO Y CULTIVO** | Código y nombre del cultivo plantado. |
| **FEC.MADUR**. | Toneladas de caña por hectárea. |
| **PRODUCTO** | Producto obtenido del cultivo (por ejemplo, jugo, azúcar, etc.). |
| **DOSIS MADURANTE** | Dosis de insumo (fertilizante, agroquímico) aplicado durante el periodo de maduración. |
| **SEMANAS MAD.** | Número de semanas que transcurrieron hasta alcanzar la madurez. |
| **TONULTCORTE** | Toneladas obtenidas en el último corte o cosecha |
| **TCH Y TCHM** | Indicadores del rendimiento, por ejemplo, toneladas de caña por hectárea (TCH). |
| **TON.AZUCAR** | Toneladas de azúcar extraídas, indicador clave en cultivos de caña. |
| **RDTO** | Rendimiento general del cultivo (puede ser en términos de cantidad o calidad). |
| **TAH Y TAHM** |  |

**HISTORICO\_SUERTES**

Al trabajar con este dataset, se recomienda:

1. **Verificar y Manejar Valores Nulos:**

Revisa el porcentaje de datos faltantes en cada columna. Para columnas con altos porcentajes de nulos (por ejemplo, algunas variables de insumos o climatológicas), considera si es más adecuado imputar valores, agruparlos en una categoría “Desconocido” o eliminarlas si no aportan información valiosa.

1. **Creación de Variables Derivadas:**

Puedes generar nuevas columnas que integren información de dos o más variables existentes.