PROYECTO FINAL DE MASTER - COVID

Entregable #1

Participantes:

* Patricia Peña Torres
* Amaia Miranda Ulloa
* Fabián Ascheri Aguerre
* José Chavarría Montero
* Juan Carlos Valcuende Aláez



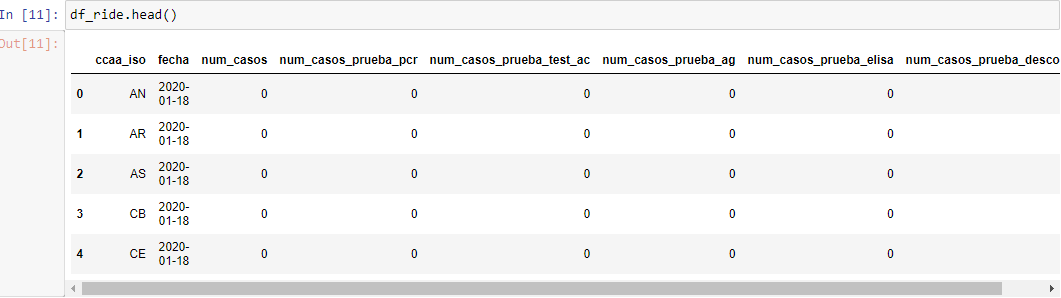
OBJETIVO: A continuación, se presenta el resumen de los ejercicios realizados por el equipo de trabajo durante la semana, que incluyen la propuesta de valor y el post-motorola.

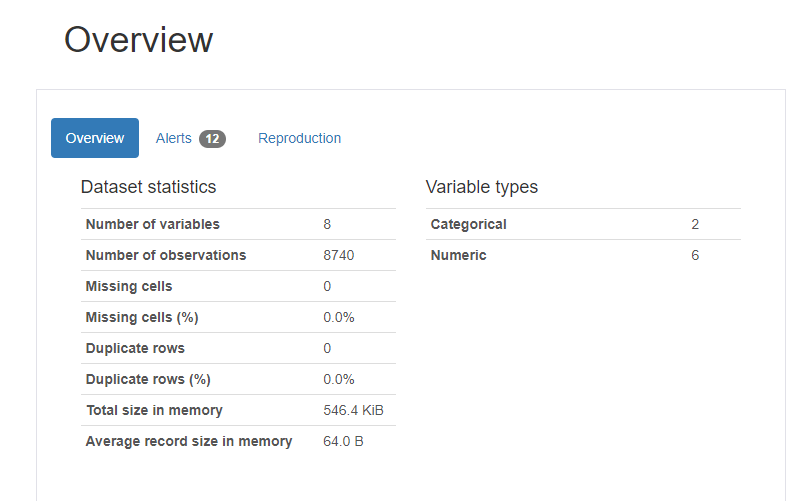
# **Introducción**

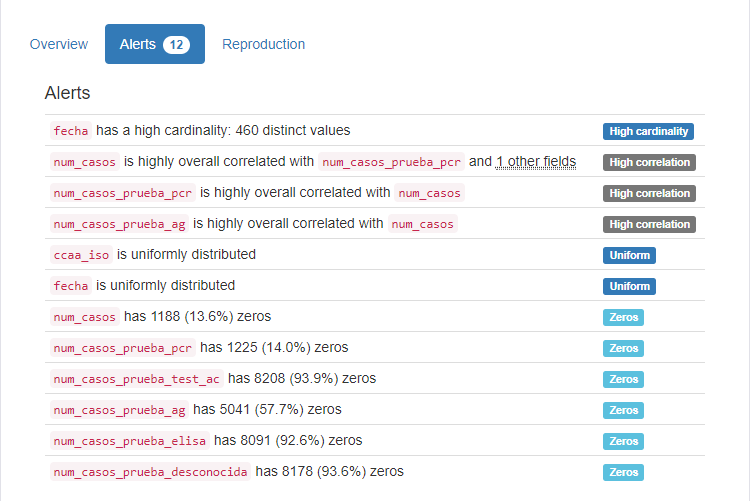
Necesidades de Nuestro Cliente, Objetivos del TFM, etc.

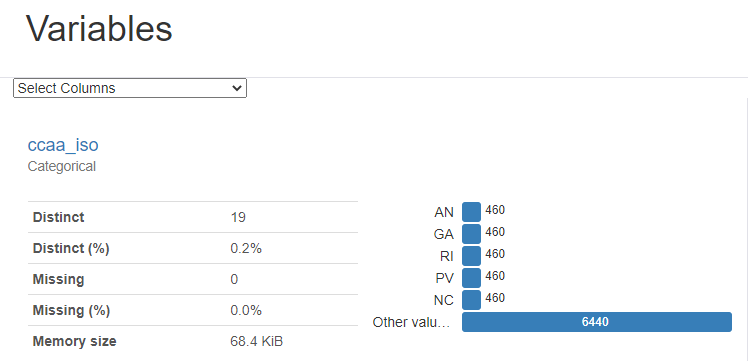
# **Perfilado de Datos // COVID ESPAÑA**

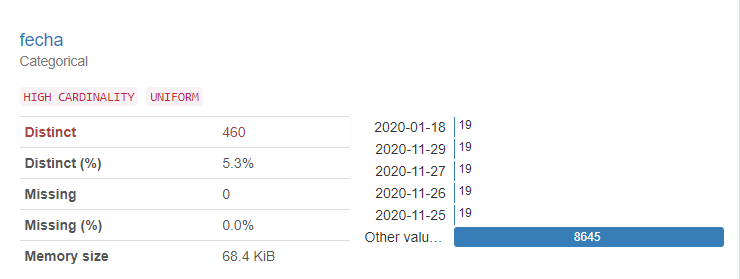
1. **casos\_diag\_ccaadecl (***Número de casos por técnica diagnóstica y CCAA (de declaración)*

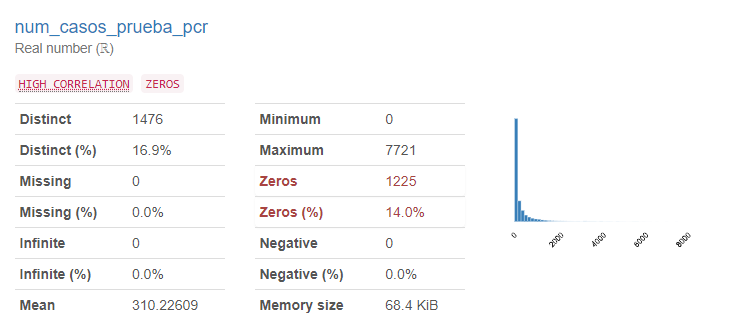


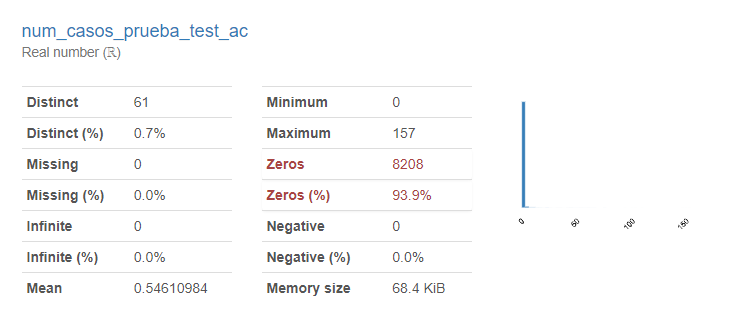


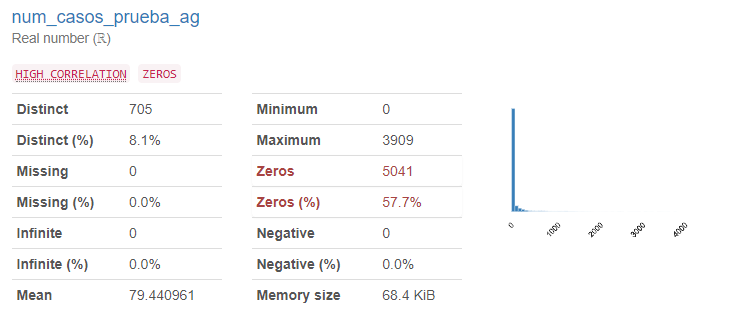


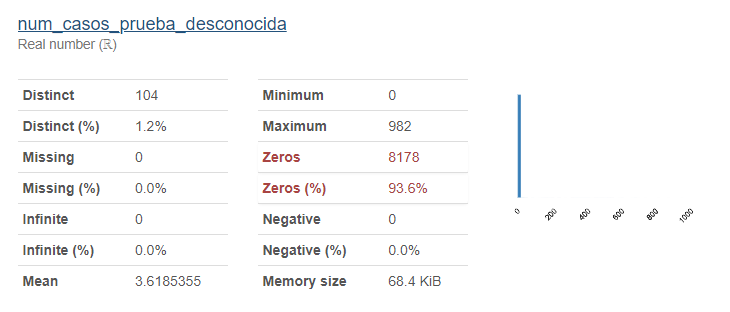








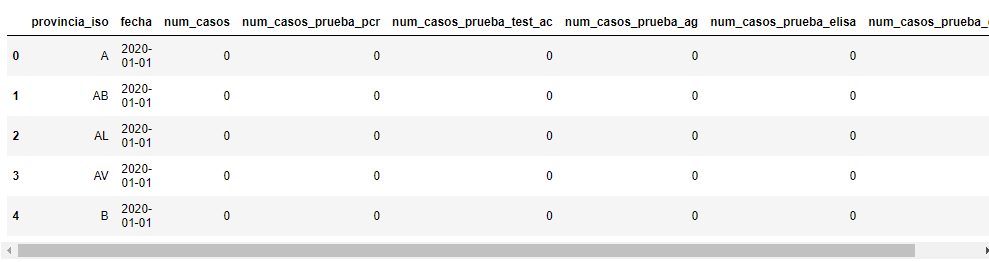


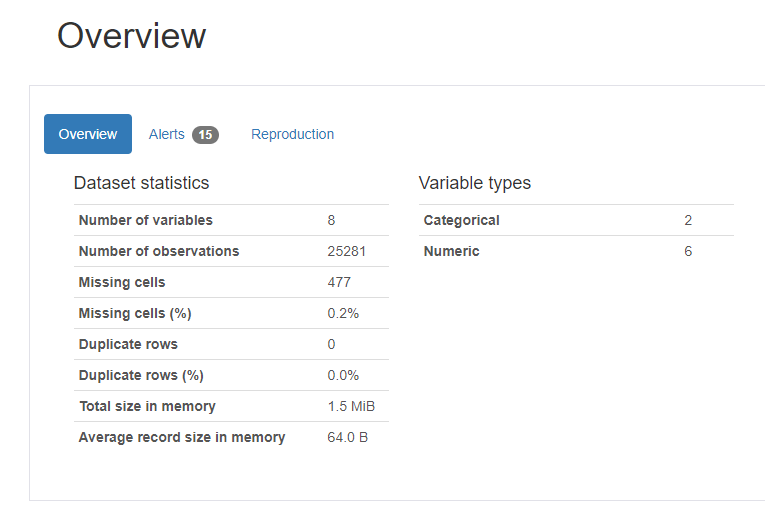


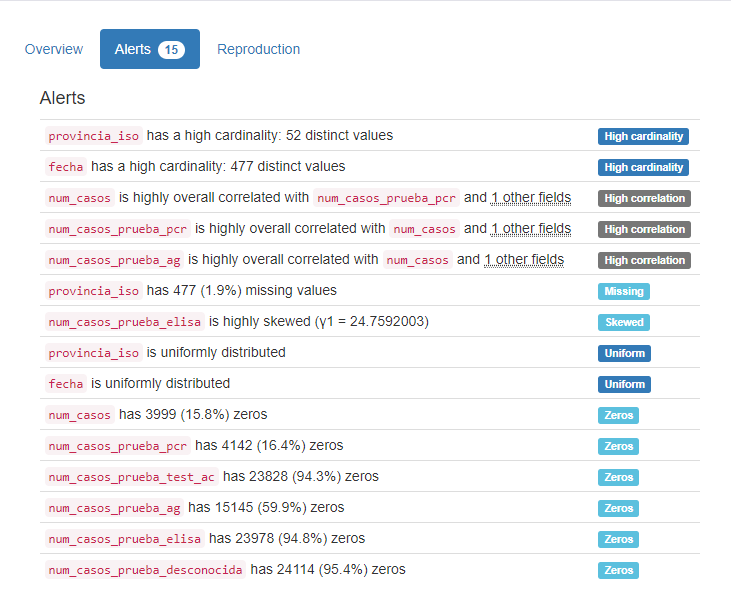
Explicación del resultado

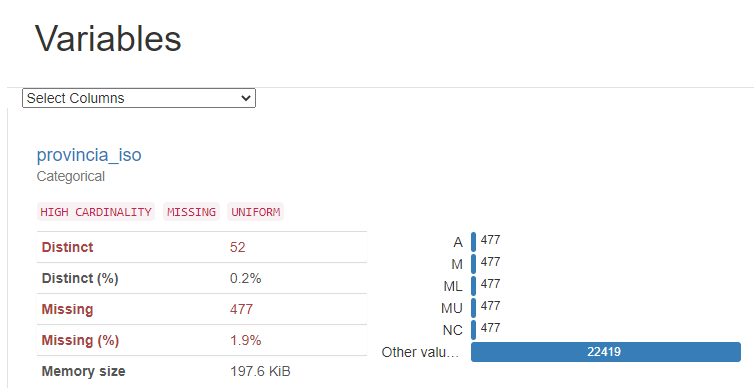
Observamos que el fichero consta de 8 variables, de las cuales 2 son categóricas y 6 numéricas. No se observan duplicados ni valores nulos. Así mismo, hay 3 variables que están altamente correlacionadas como son los *num\_casos*, num\_casos\_prueba\_pcr y num\_casos\_pruega\_ag. Observamos un valor superior al 90% de celdas con valores 0 en las variables num\_casos\_prueba\_test\_ac, num\_casos\_prueba\_elisa y num\_casos prueba\_elisa.

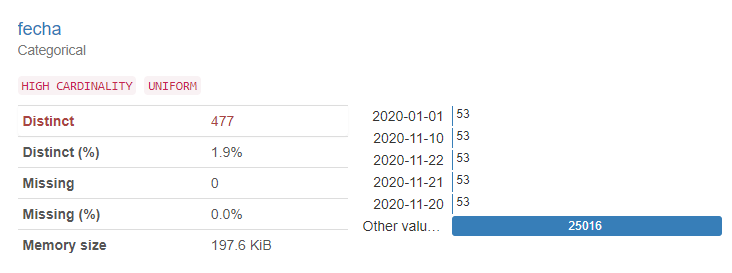
1. **casos\_tecnica\_provincia (***Número de casos por técnica diagnóstica y CCAA (de residencia)*

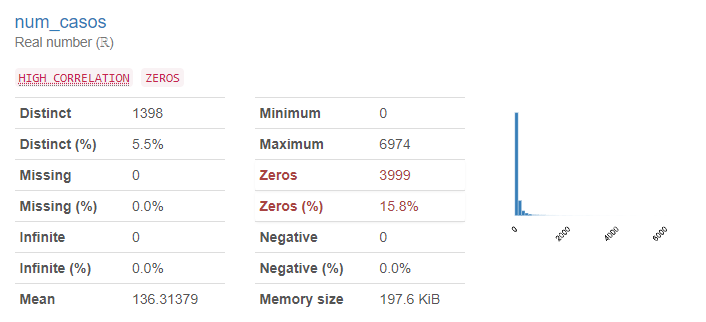


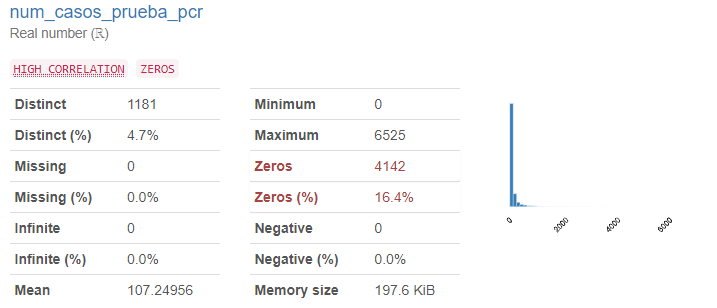


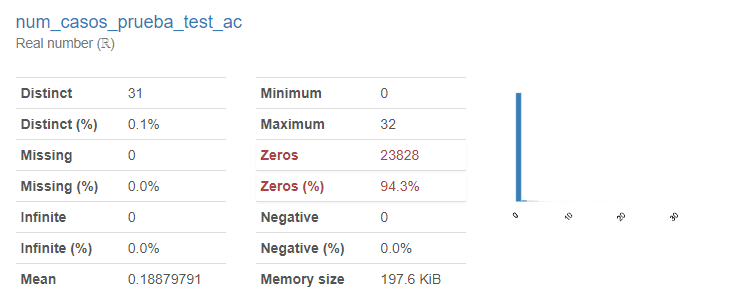


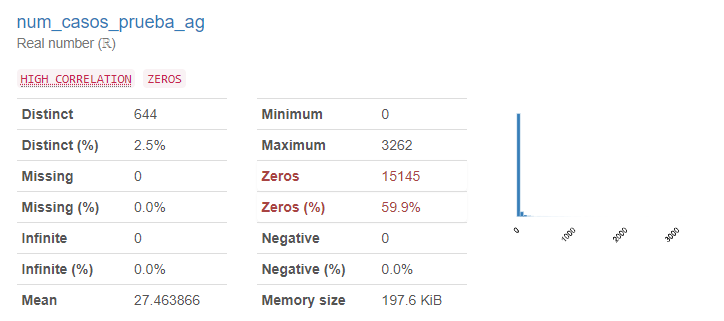


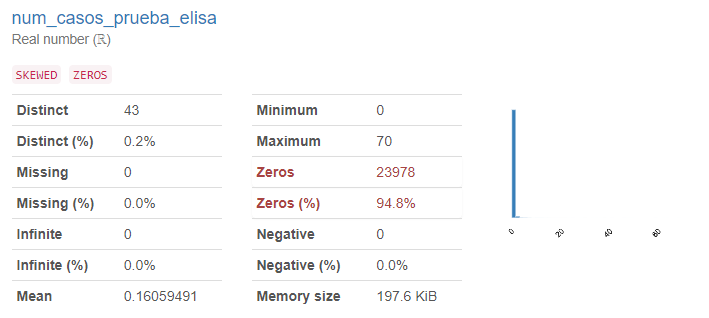


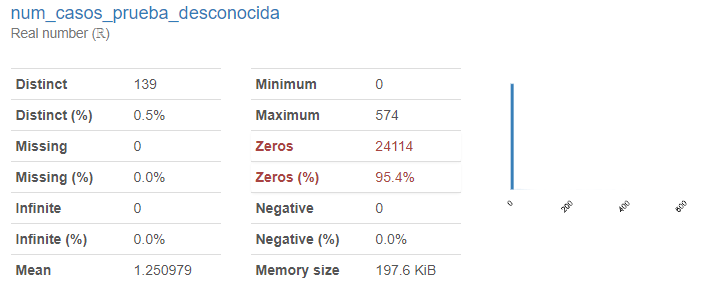










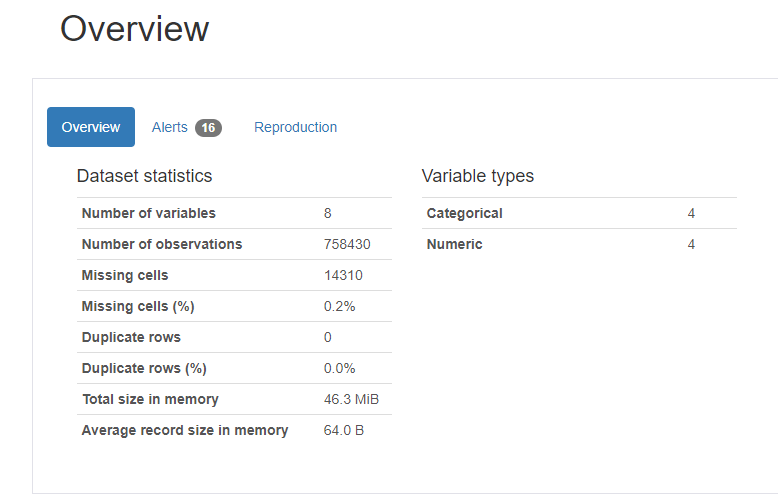


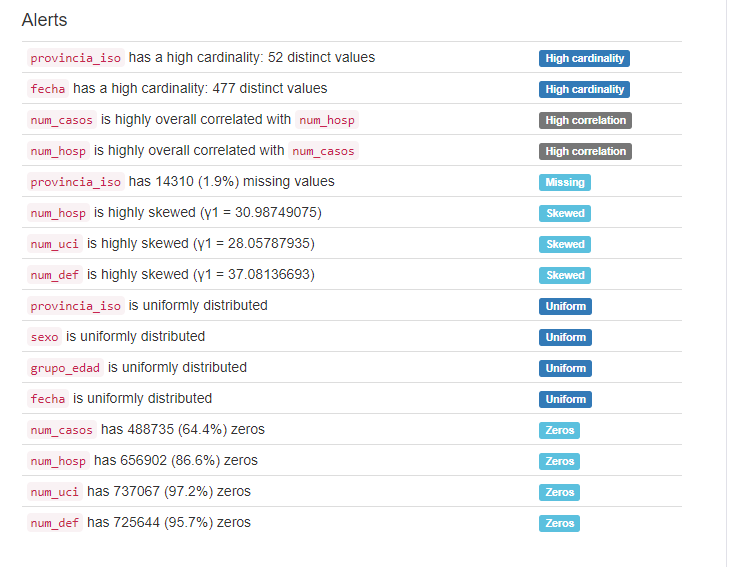
Explicación del resultado

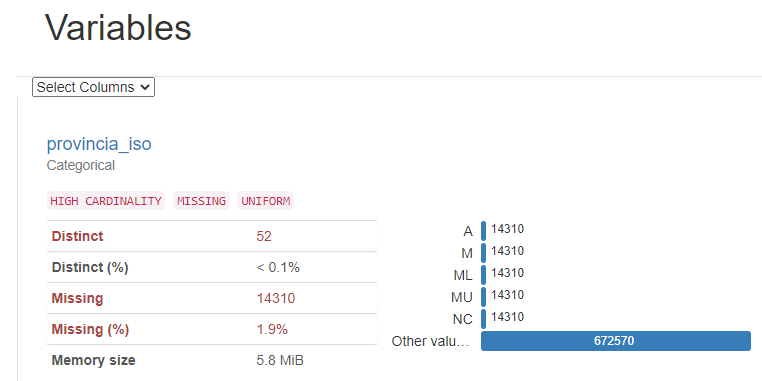
Observamos que el fichero consta de 8 variables, de las cuales 2 son categóricas y 6 numéricas. No se observan duplicados y solamente el 0,2% de valores nulos. Así mismo, hay 3 variables que están altamente correlacionadas como son los *num\_casos*, num\_casos\_prueba\_pcr y num\_casos\_pruega\_ag. Observamos un valor superior al 90% de celdas con valores 0 en las variables num\_casos\_prueba\_test\_ac, num\_casos\_prueba\_elisa y num\_casos prueba\_elisa.

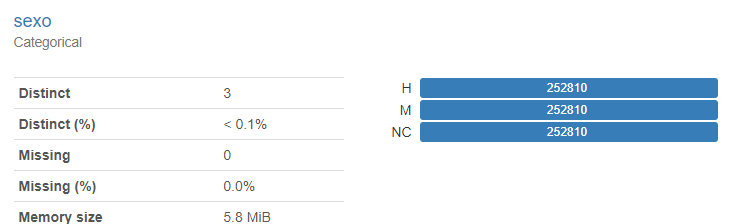
1. **casos\_ hosp\_uci\_def\_sexo\_edad\_prov\_res:** (*Número de hospitalizaciones, número de ingresos en UCI, número de defunciones por sexo, edad y provincia de residencia)*

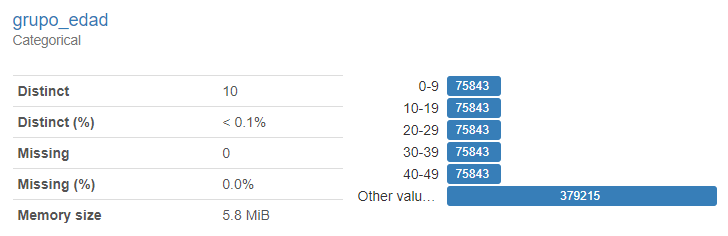


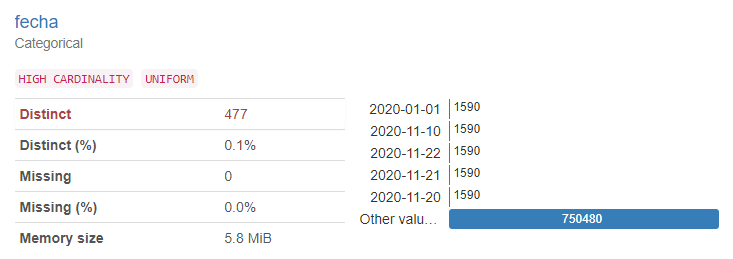


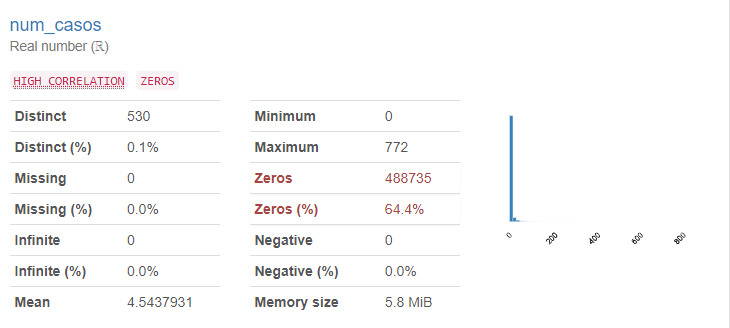


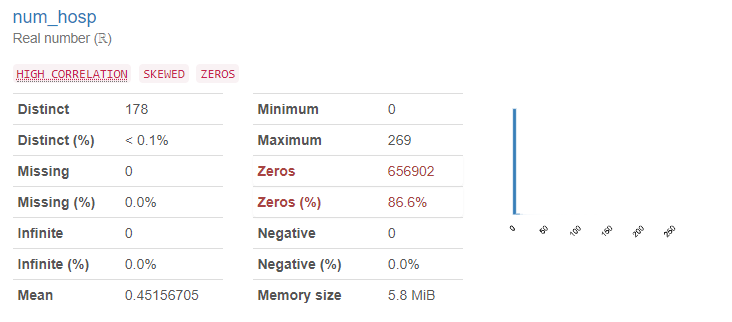


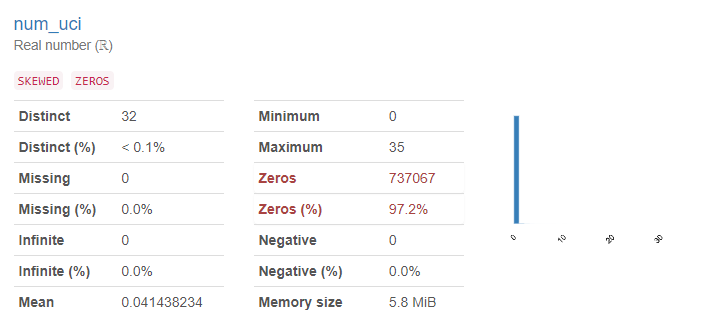


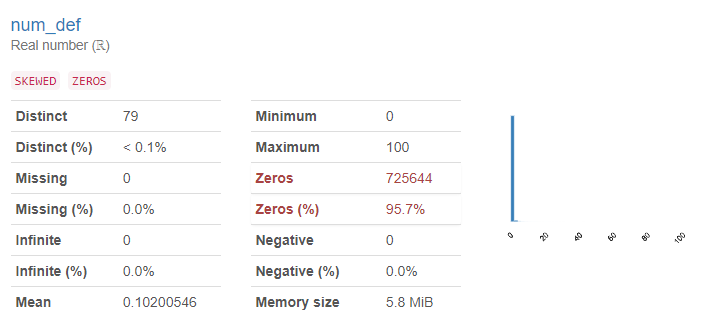








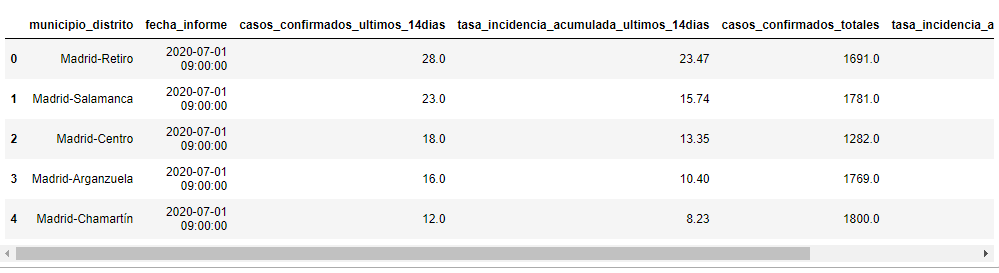


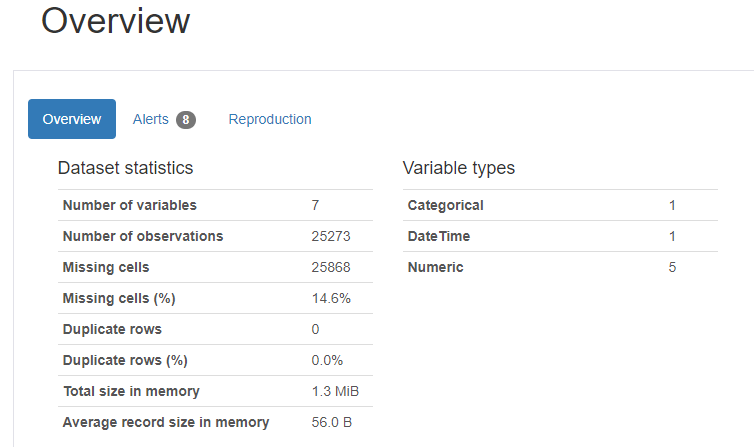


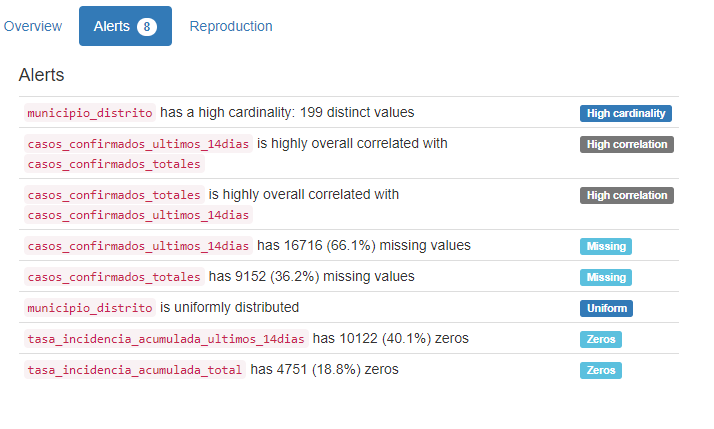
Explicación del resultado

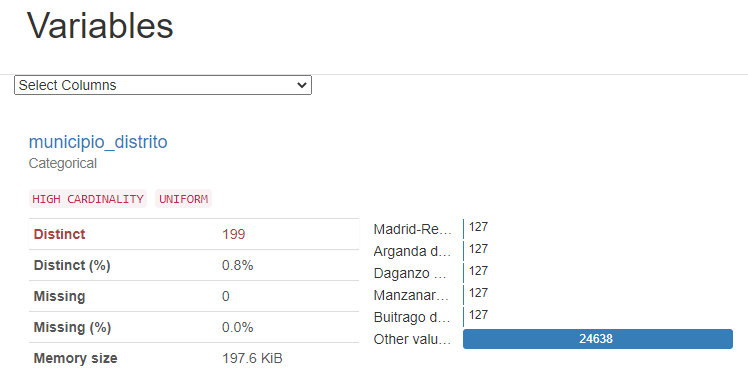
Observamos que el fichero consta de 8 variables, de las cuales 4 son categóricas y 4 numéricas. No se observan duplicados y solamente el 0,2% de valores nulos. Así mismo, hay 2 variables que están altamente correlacionadas como son los *num\_casos y el num\_hosp*. Observamos un valor superior al 90% de celdas con valores 0 en las variables num\_uci y num\_def.

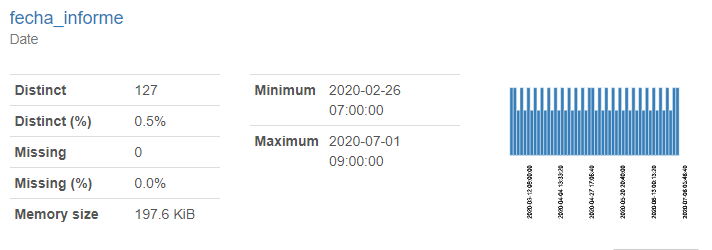
1. **covid19\_tia\_muni\_y\_distritos:** (*Número de casos confirmados, tasa de incidencia acum en los últimos 14 días y global de la Provincia de Madrid y sus Municipios/Distritos)*

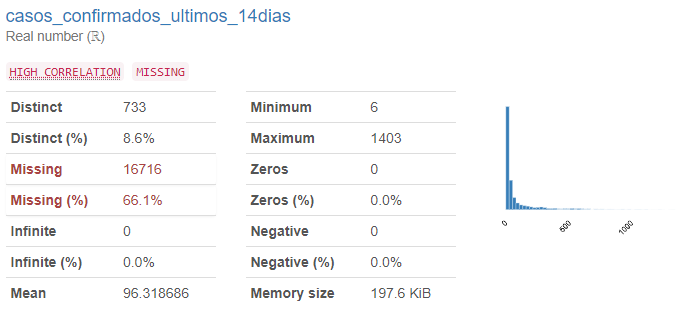


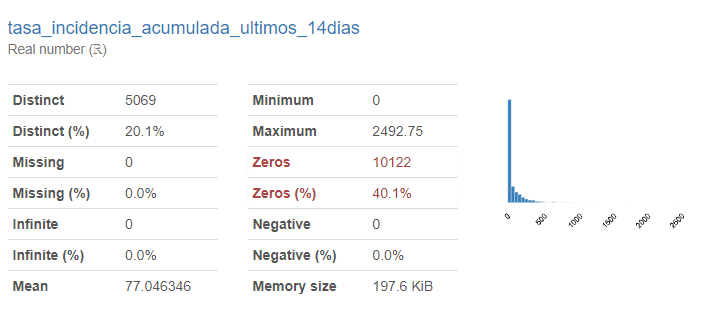


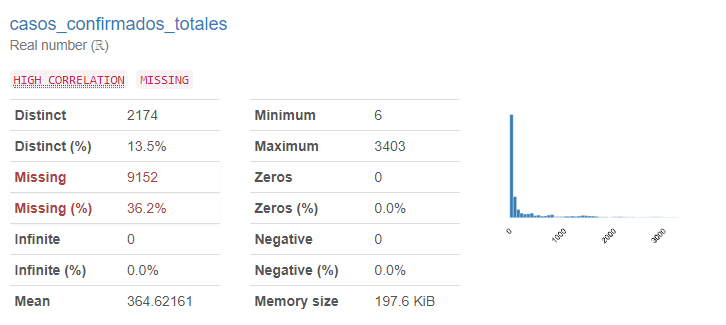


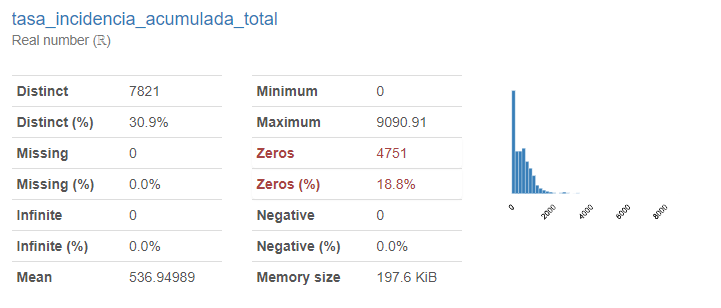


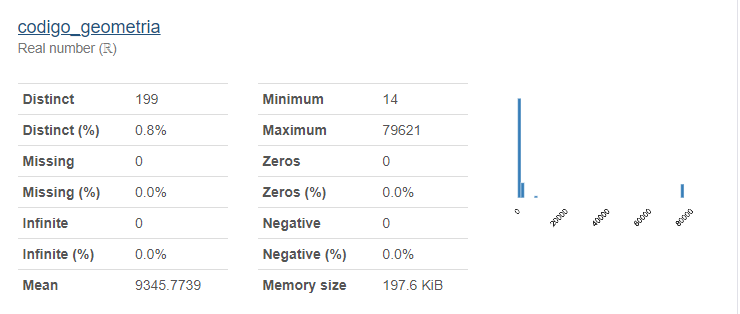








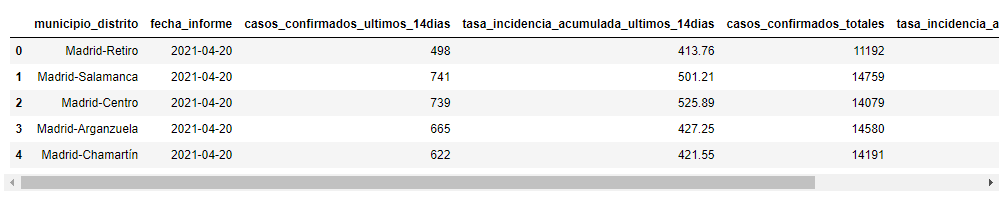


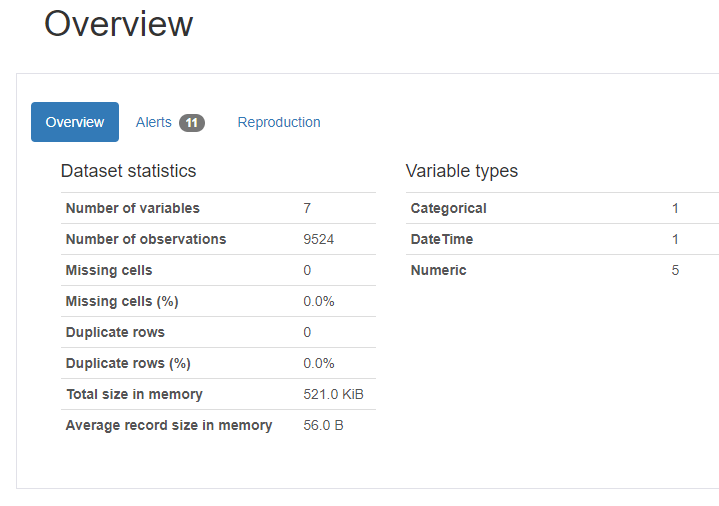


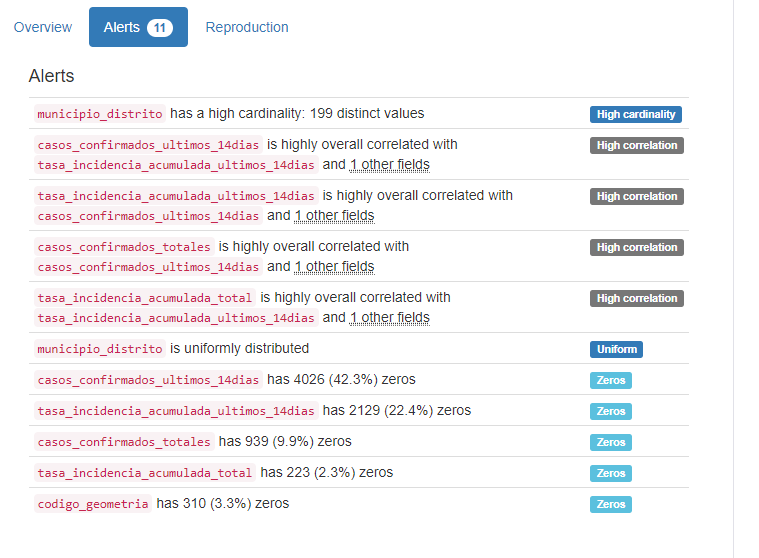
Explicación del resultado

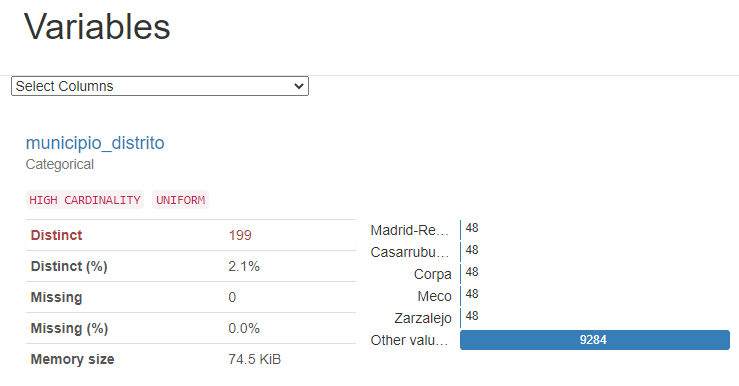
Observamos que el fichero consta de 7 variables, de las cuales 5 són numéricas, 1 categórica y 1 formato fecha. No se observan duplicados y hay un 14,6% de valores nulos. Así mismo, hay 2 variables que están altamente correlacionadas como son los *el num\_casos\_confirmados\_ultimos\_14dias y el num\_casos\_confirmados\_totales*.

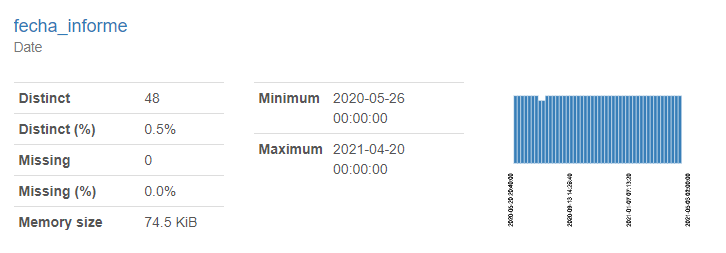
1. **covid19\_tia\_muni\_y\_distritos\_s:** (*Número de casos confirmados, tasa de incidencia acum en los últimos 14 días y global de la Provincia de Madrid y sus Municipios/Distritos) Continuación del fichero anterior.*



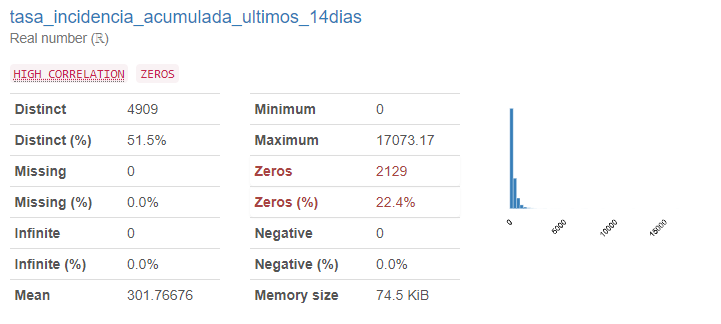


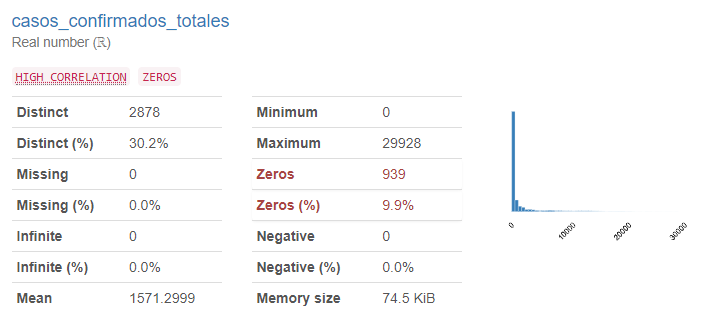


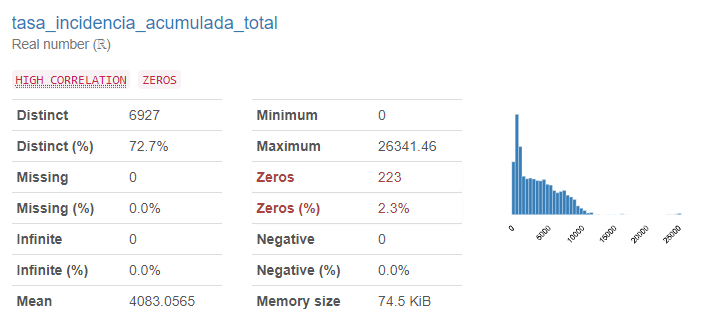


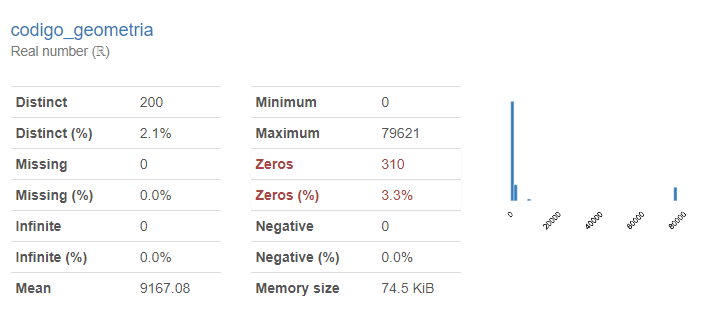












Explicación del resultado

Observamos que el fichero consta de 7 variables, de las cuales 5 són numéricas, 1 categórica y 1 formato fecha. No se observan duplicados y no hay ningún valor nulo. Así mismo, hay 4 variables que están altamente correlacionadas como son los *el num\_casos\_confirmados\_ultimos\_14dias con tasa\_incidencia\_acumulada\_ultimos\_14dias, num\_casos\_confirmados\_totales* y tasa\_incidencia\_acumulada\_total.

Explicación del resultado

* **Nombre del Archivo**
* Ejemplo de Datos (head)
* Resumen General (overview)
* Alertas (Alerts)
* Variables (de todas las variables que sean relevantes, no más de 10)

Explicación del resultado

# **Estructura de Base de Datos**

INCLUIR DIAGRAMA DE ESTRUCTURA RELACIONAL (DIAGRAMA ER)