Aprendizaje Máquina

Maestro:

Andres Garcia Floriano

Alumno: Gerardo Martinez Ayala

6/Septiembre/2024

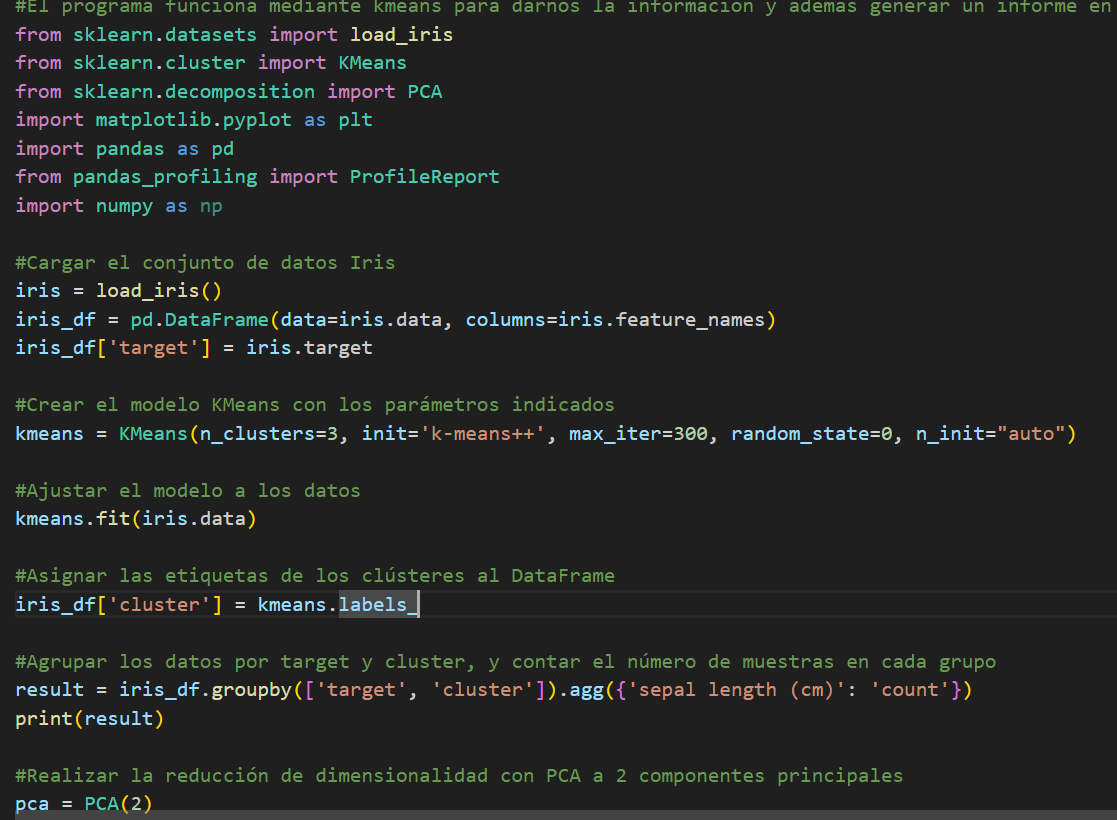
Instituto politécnico Nacional

El siguiente programa utiliza kmeans para el análisis de datos incluyendo también un reporte en HTML del material a utilizar.

Para la realización del ejercicio se empleó un entorno de python utilizando anaconda, en este caso empezaremos por asegurarnos de tener todas las librerías que se ocuparan durante la codificación, en este caso ya se contaban con todas por lo que se dio inicio.

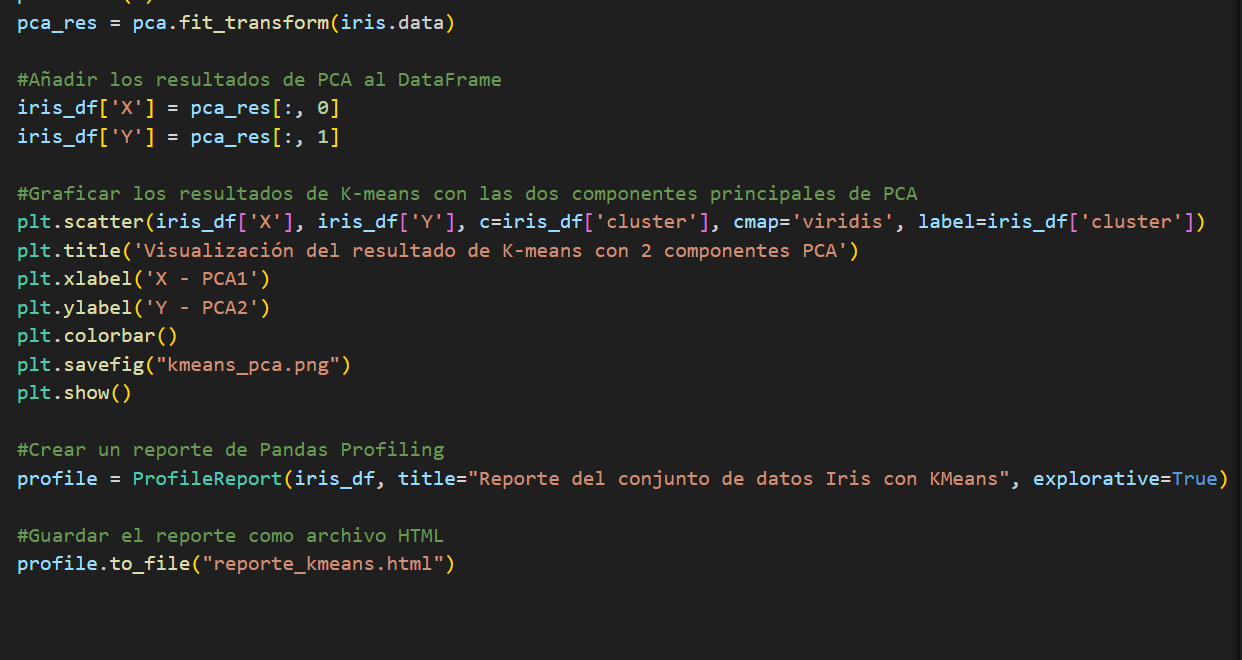
Primero se cargaron los datos de iris que se utilizaran posteriormente, así mismo se ajustaron los parámetros para el kmeans, como puede ser el propio modelo, los datos y asignar etiquetas a los clusters que aparecerán en nuestro proyecto, así mismo no olvidar ajustar el PCA para que los datos sean procesados apropiadamente. El codigo deberia quedar como en la figura 1.0

Figura 1.0: Ajustando los parámetros para los clusters y verificando que la información está puesta en la forma correcta



Una vez con la parte de la información y los clusters habiendo verificado que esté bien programado deberemos continuar con la visualización de los datos, para este cometido se utilizó matplotlib para mostrar todo en una sola imagen, ya con esto solo habría que codificar la parte del informe en HTML y el programa debería estar completo como en la figura 1.1.

Figura 1.1: se muestran los datos como deberían ir.



Finalmente corremos el programa como en la figura 1.3, si todo fue bien codificado Visual Studio nos dará una imagen en formato png y el informe en HTML listo para ser usado.

Figura 1.3: Programa compilado y funcionando

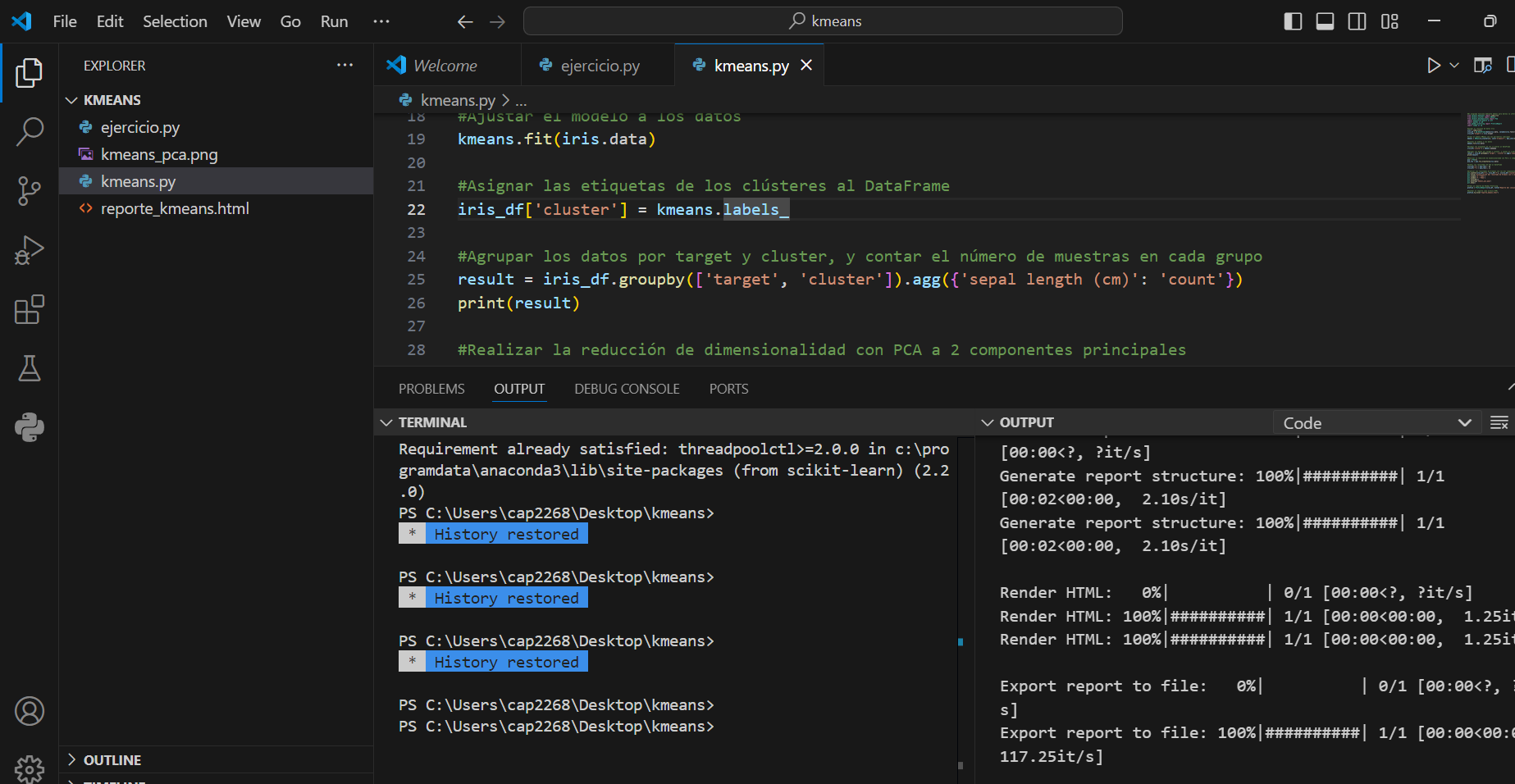
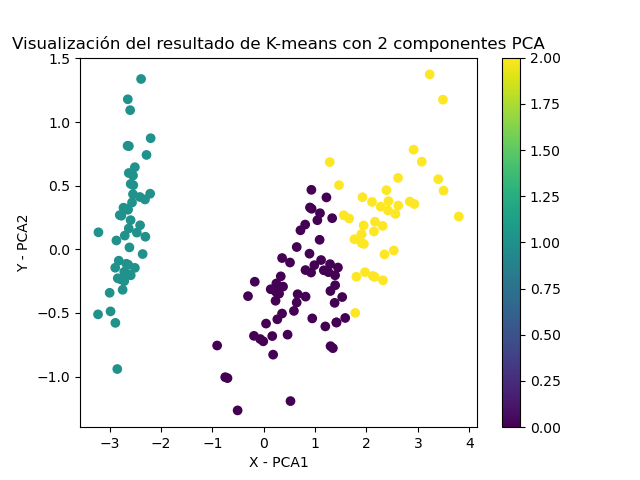
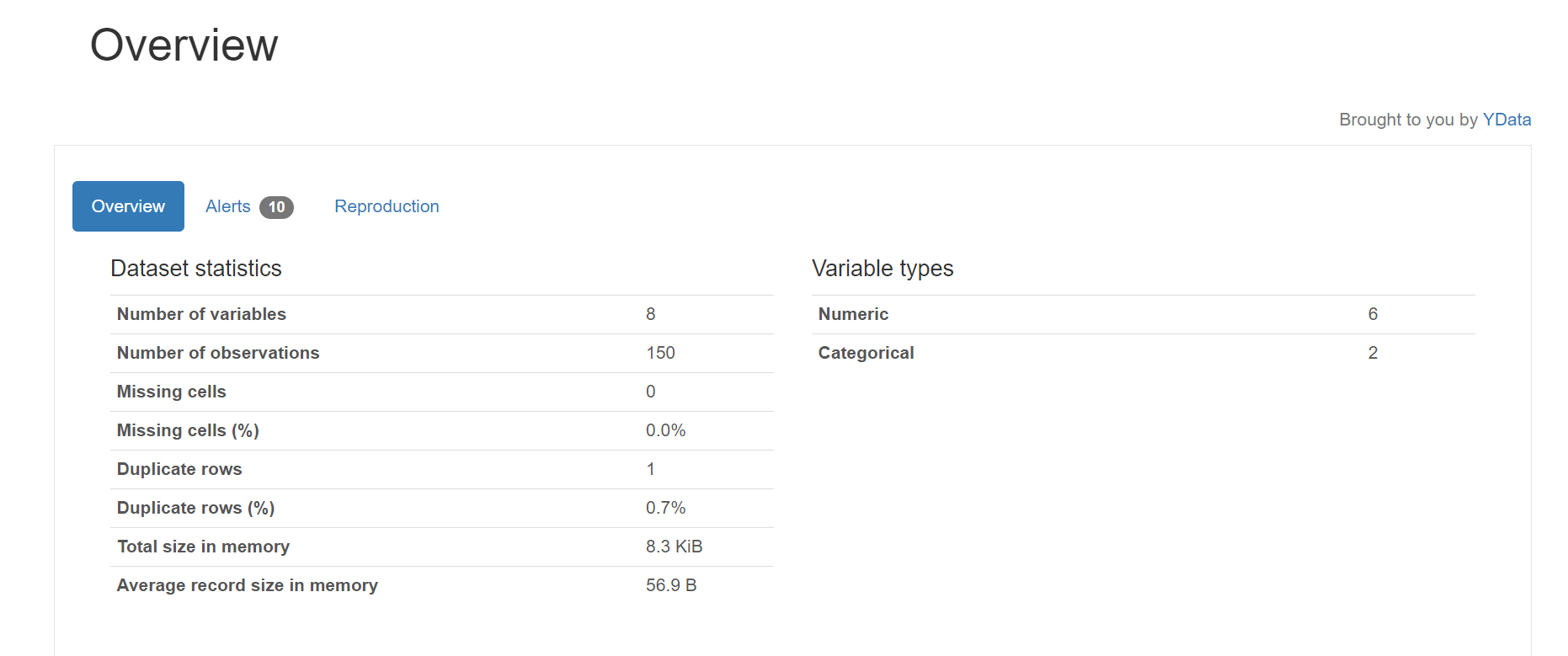
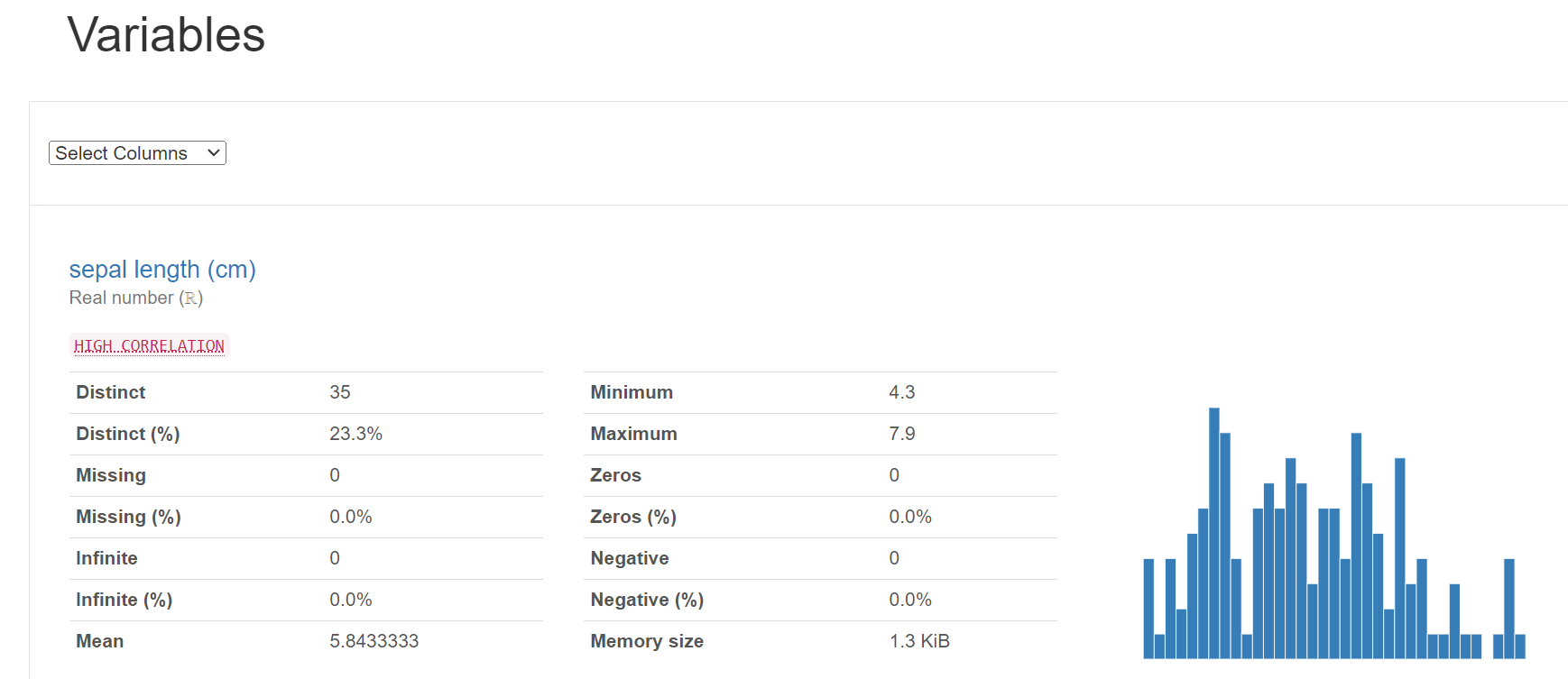
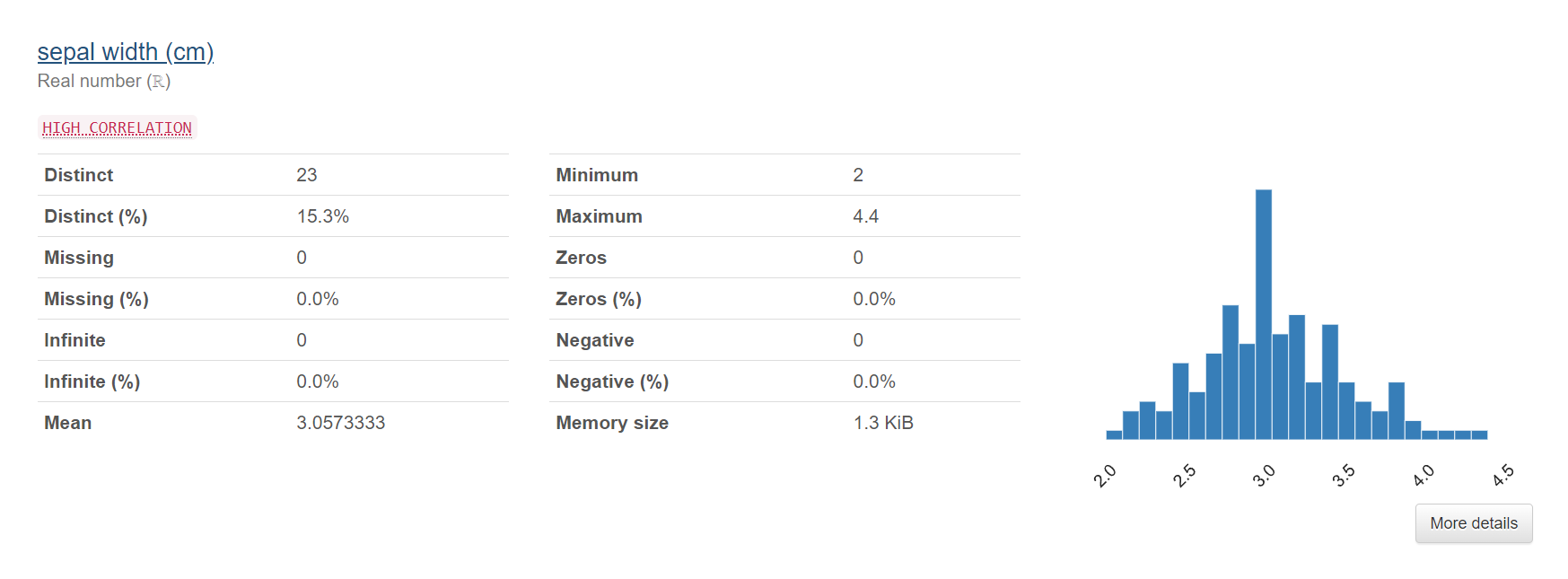


Imagen png  


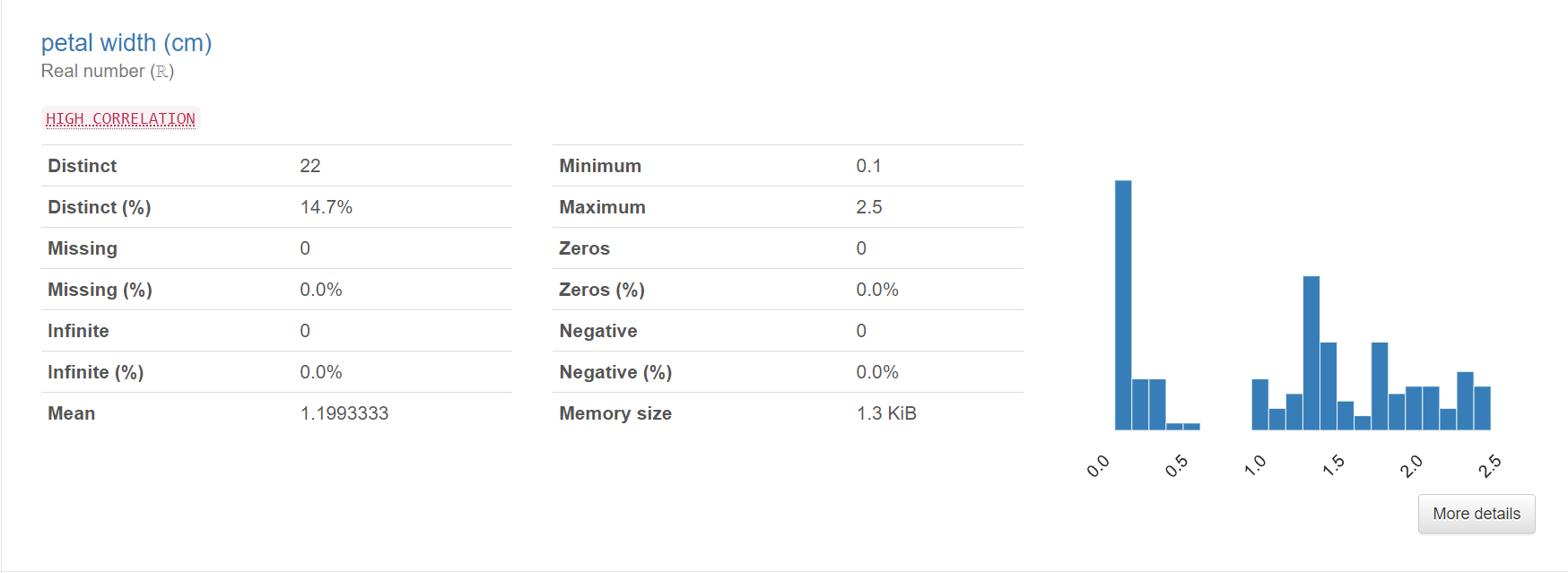
Informe en HTML

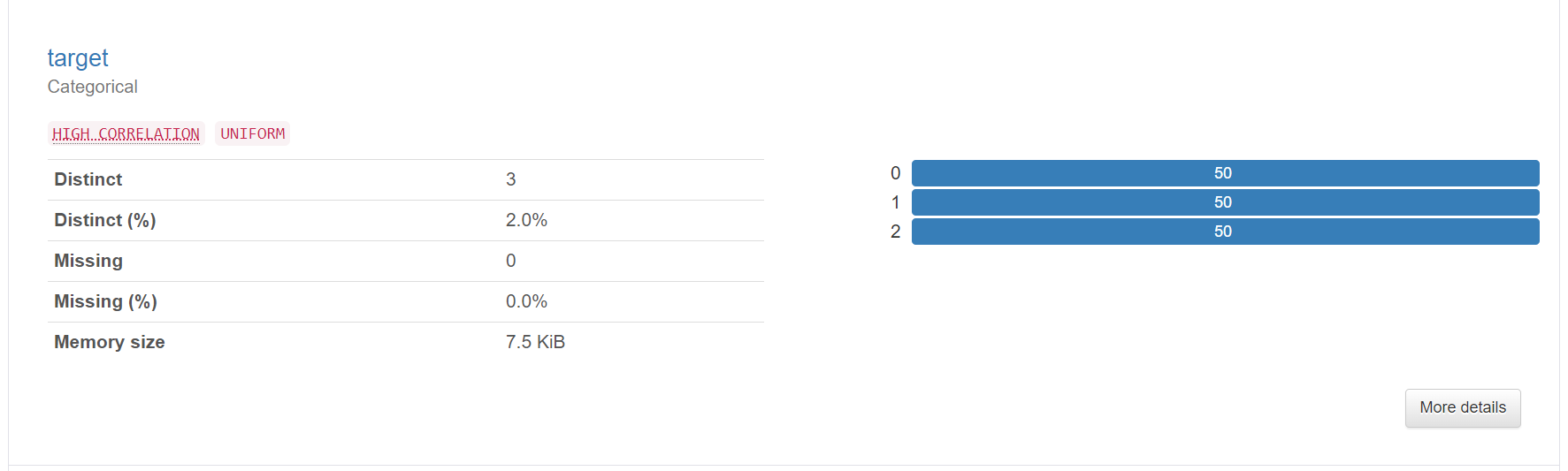
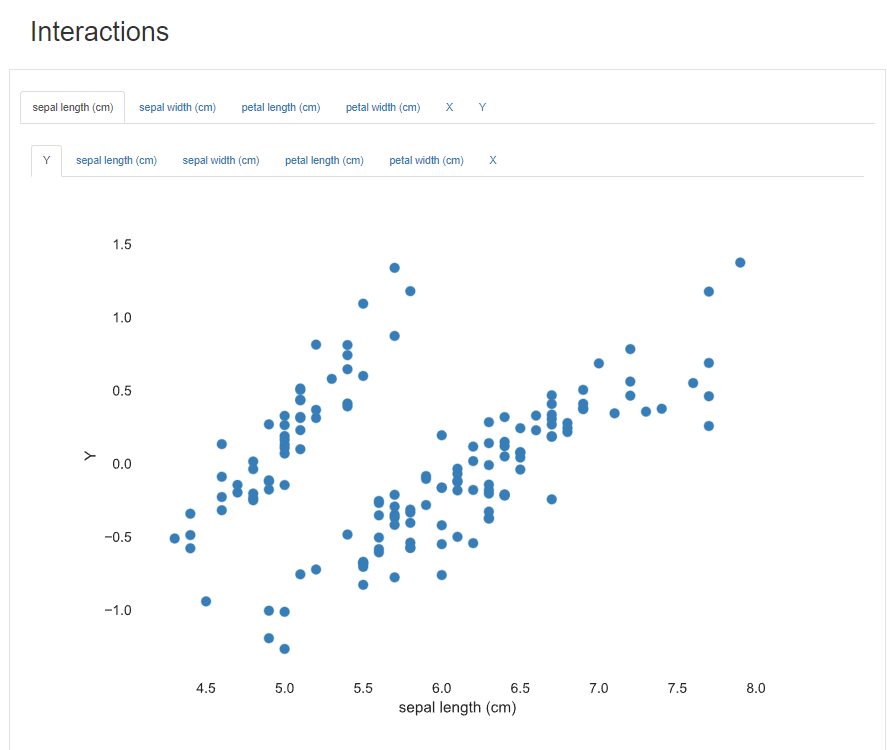
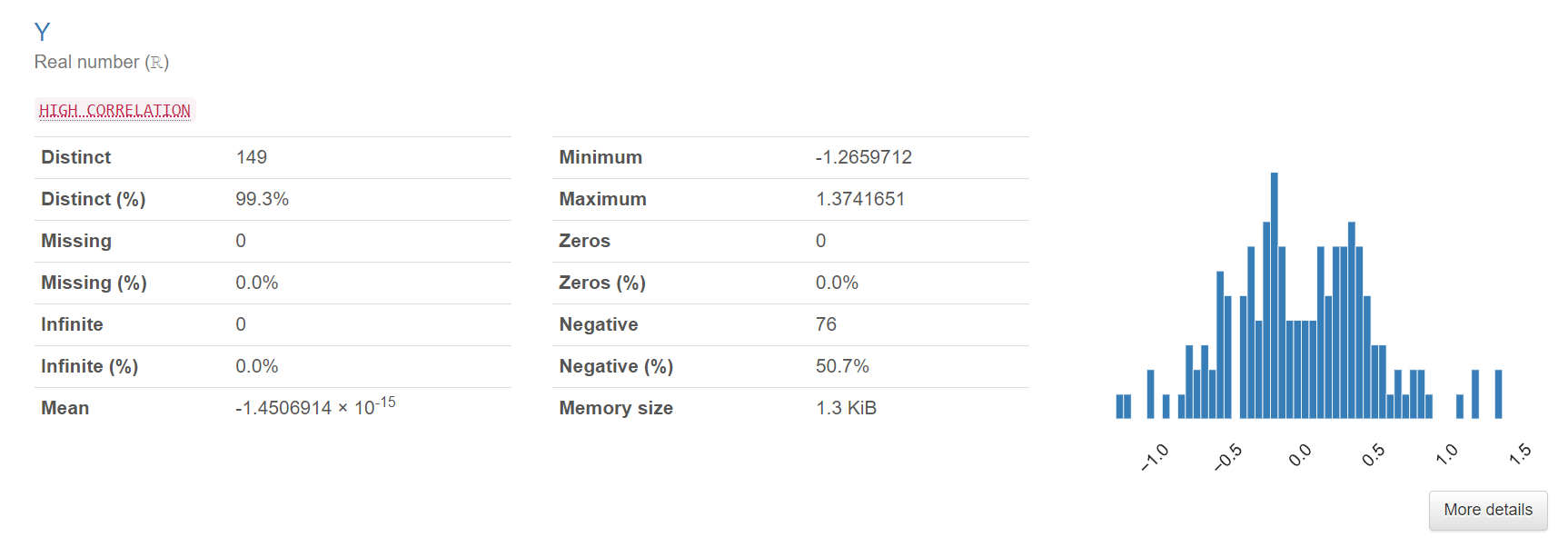
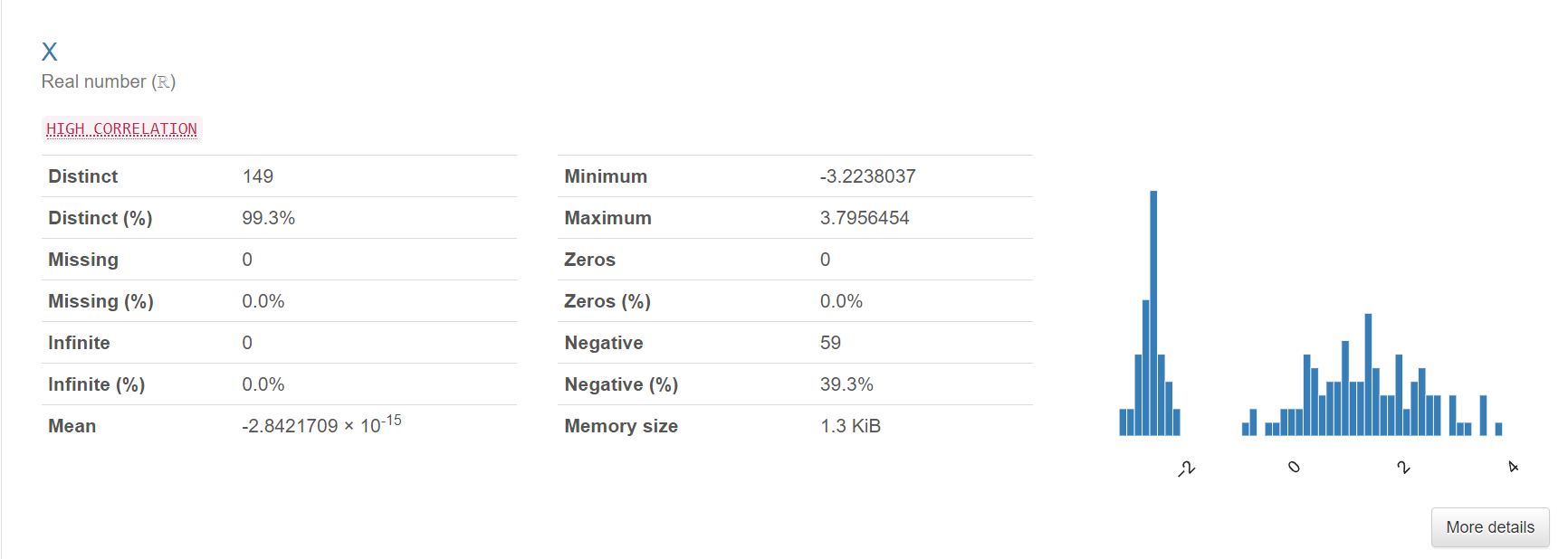
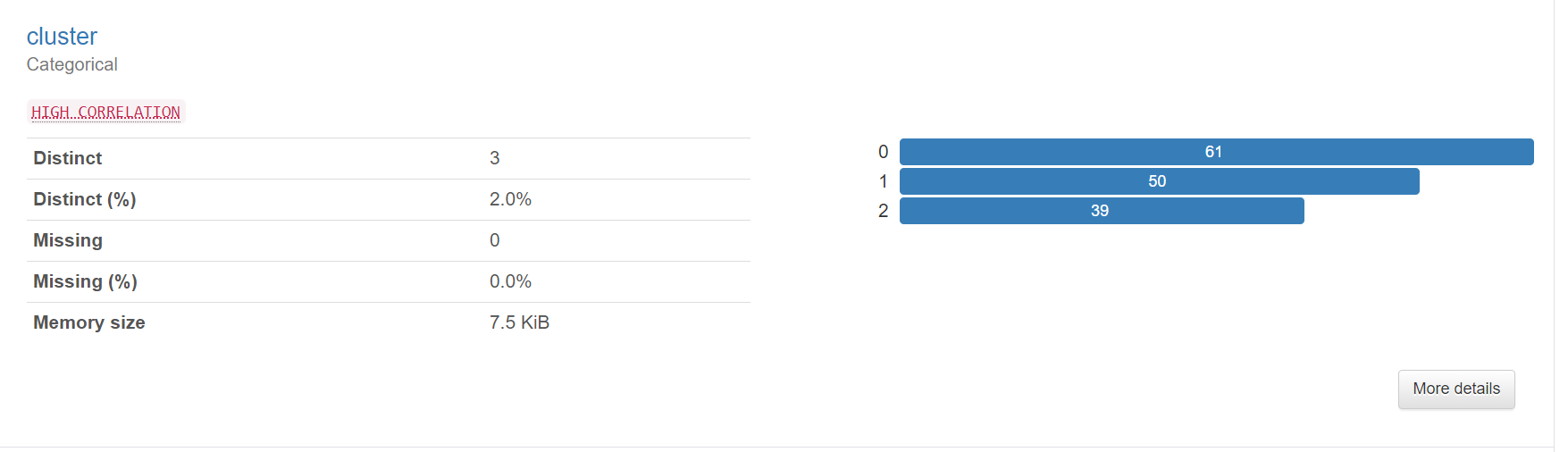


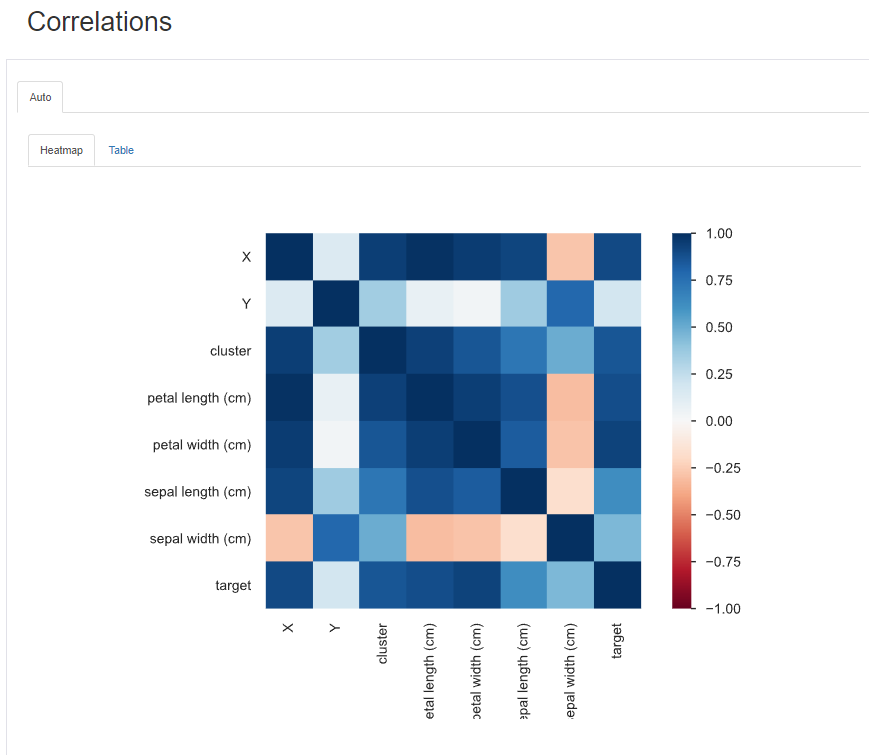


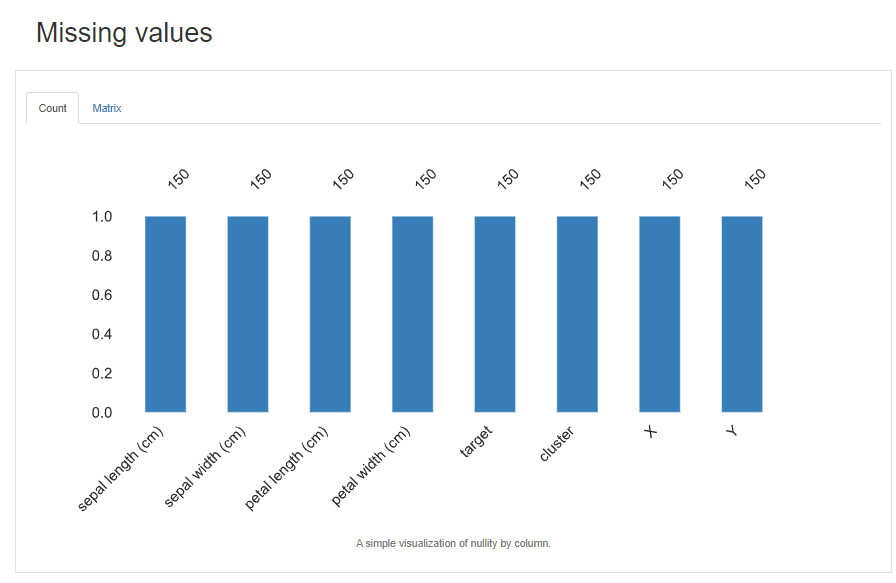


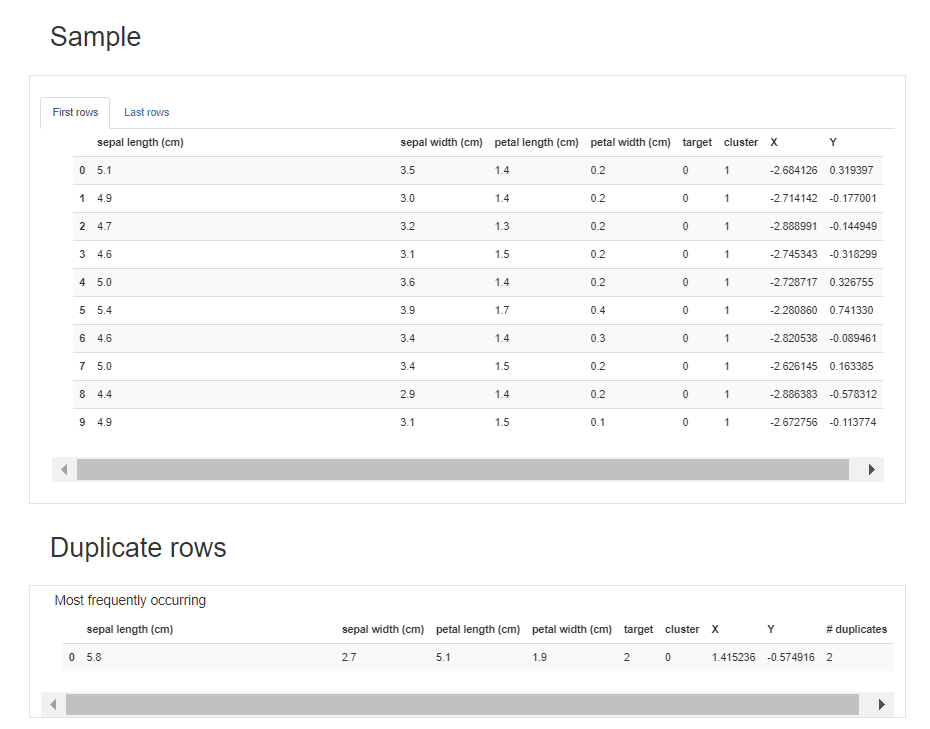




****

****

****

****

Conclusiones

El programa ciertamente es sencillo de aplicar y relativamente corto para lo que puede servir, pero cabe destacar que hay que el número de clusters puede variar dependiendo del tipo de programa y de los datos que vayamos a manejar en un entorno real, sin embargo sigue siendo una excelente opción para los profesionales o estudiantes que necesiten llevar a cabo una tarea similar.