

Ejercicios de curación y limpieza de datos

Ejercicio 6:

- a. Carga el fichero SAFI_ejercicio6_openrefine.csv para realizar el ejercicio 6.
- b. Transforma las columnas “no_members”, “yrs_liv”, y “buildings_in_compound” de texto a numérico.
- c. Crear facets de tipo numérico de estas columnas y observa que información se puede extraer de los mismos.
- d. Responde: ¿qué sucede y como afectan al facet los valores no numéricos de estas columnas? ¿Y con los NULL? ¿Y con los vacíos? ¿Se gestionan de diferentes formas? Prueba introduciendo este tipo de errores manualmente en columnas de tipo numérico.
- e. ¿Qué tipo de errores podemos identificar y corregir con el facet numérico ¿Cómo?
- f. Deshaz los cambios con “Undo” y vuelve a tener las columnas como tipo texto. Crea ahora un facet de tipo texto sobre estas columnas. ¿Se pueden identificar los errores a corregir con la misma facilidad que con un facet numérico?

M1968 - Ciclo de vida de los datos

Máster en Data Science – Universidad de Cantabria Curso 2018-2019

Diego García Saiz