

## Ejercicios de curación y limpieza de datos

### Ejercicio 1:

- a. Carga el fichero SAFI\_ejercicios1y2\_openrefine.csv para el ejercicio 1.
- b. Buscar errores en el campo “province” y corrígelos de la forma más adecuada, usando facets.
- c. Realiza la misma operación en el campo “other\_buildings”. ¿Se pueden corregir los errores fácilmente? ¿Por qué?
- d. En el caso del campo “district”, ¿se pueden corregir fácilmente? ¿Cómo lo harías?
- e. Y en el caso del campo “parents\_liv”, ¿podríamos corregir todos los errores sin problema? ¿Cómo?
- f. ¿Podríamos también corregir errores, usando los facets basado en texto, en campos de tipo numérico? Prueba con el campo “years\_farm”.
- g. Y en el caso de campos que almacenan listas de valores, ¿podríamos usar facets basadas en texto para corregir los errores? Prueba con el campo “month\_no\_water”.
- h. ¿Y en el caso de campos de tipo fecha? Prueba con el campo “interview\_date”.

### Ejercicio 2:

Usando el mismo fichero SAFI\_ejercicios1y2\_openrefine.csv, trata de realizar el mismo ejercicio hecho en clase sobre el campo “village” usando una herramienta para hojas de cálculo como Microsoft Excel o similar. Intenta buscar, encontrar y corregir los errores de ese campo. ¿Cómo lo harías?