

Se leyó el archivo `technical_test_data_analyst.json` desde Rstudio. El archivo cuenta con 9 columnas, las cuales son `activo`, `fecha_inicial`, `fecha_final`, `latitud`, `longitud`, `valorventa`, `valorarriendo`, `data`, `id`.

A continuación, se comentarán las modificaciones que se realizaron a los datos.

- La columna `activo` fue modificada en sus calores como sigue:

1= "activo"; 0= "no activo"

- Las columnas asociadas a las fechas fueron modificadas de la siguiente forma:

Se le suprimieron los últimos 3 ceros (0), pues generan un ruido en la información y después fueron convertidos a formato de fecha (año-mes-día hora) y por ultimo se le borro la hora pues se le considero irrelevante.

- Se crea la columna `tipo` que caracteriza si el inmueble esta en venta, arriendo o ambas
- Se crea la columna `geodatos` que caracteriza si el inmueble tiene asociados datos validos de longitud y latitud.
- Se extrae la información de la columna `data`:

Se encontró que el mejor separador de la información fue `","` el cual permitió obtener de manera mas optima los nombres de variables y sus datos, los cuales posteriormente se separaron por `":"`, finalmente se creo una tabla con la mayor cantidad de posibles variables y sus valores, la cual quedo registrada en la tabla con nombre `prueba`.

- Se crean las columnas `n_hab` y `n_banos` con la información obtenida de la tabla `prueba`
- Se completan y depuran valores asociados a las columnas `valorarriendo`, `valorventa`, `latitud` y `longitud` obtenidos de la tabla `prueba`.
- Se crea la tabla `informe` el cual permite visualizar el número de datos corregidos y depurados:

Tabla sin depurar			Tabla depurada		
Tipo	geodatos	número	Tipo1	geodatos1	número
arriendo	no tiene	155	arriendo	no tiene	250
arriendo	tiene	257	arriendo	tiene	184
venta	no tiene	213	venta	no tiene	288
venta	tiene	320	venta	tiene	241
venta y arriendo	no tiene	26	venta y arriendo	no tiene	23
venta y arriendo	tiene	14	venta y arriendo	tiene	14
NA	no tiene	63	NA	no tiene	72
NA	tiene	152	NA	tiene	128

En esta tabla podemos ver que se incluyeron 22 valores de arriendo a los datos originales y fueron eliminados 7 datos asociados a los valores venta y arriendo, venta debido a que tenían valores atípicos. Adicionalmente, se eliminaron 176 datos asociados a la longitud y latitud debido a que sus valores no correspondían a lugares geográficos coherentes.