Luis

* En la fila [15]

Revisar, porque capital federal no aparece en el place\_with\_parent de tres str porque esa posición es para las ciudades

* Y revisar la comparación de la fila 15 porque capital federal si apare en place\_name

data["place\_name"].**value\_counts**(), le da 1297 registros

No aparece porque siempre que aparece Capital Federal en place\_name es porque place\_with\_parent tiene 2 str. Se puede comprobar con

data[data['place\_name']=='Capital Federal']['place\_with\_parent\_names'].value\_counts()

* revisar las conclusiones
* los nan de place\_name también está en la fila [12], estaríamos repitiendo.

Si, hay que ordenar esa notebook, no sé bien aún cómo puede ser la narrativa. Me gusta esto de dividir en partes según la cantidad de str en place\_with\_parent\_names pero creo que puede ser ordenado mejor.

Discusion de las tres columnas de ubicación politica:

* Los datos contenidos en country\_name, state\_name y place\_name son los mismos o están contenidos en place\_with\_parent\_names, por lo cual pudimos llegar a la conclusión
* Que los registros de ubicación de los inmuebles están ubicados hasta la ciudad/Partido (en %)
* Seguido de ubicaciones más detallada como es el nombre de la localidad/barrio (en %)
* Existen 4780 (%) registros donde no agregó el de la localidad o barrio (en %): pueden imputarse datos de localidad/barrio en la columna place\_name desde place\_with\_parent\_names
* Y 548 (%) registros que corresponden a un barrio de la localidad, muy específico: podrían reemplazarse por la localidad o no.
* Hay 23 datos nulos para place\_name que pueden imputarse con el dato de partido Tigre
* Capital Federal, el caso más relevante que por su conformación política detalla el barrio únicamente
* Se encontró además que existen registros incompletos sin ciudad/Partido en data\_place\_with.. por lo cual en place\_name solo se detalla la ubicación por provincia o state\_name. Esto nos lleva a usar las columnas de geolocalización para mejorar el dataset (próximo)

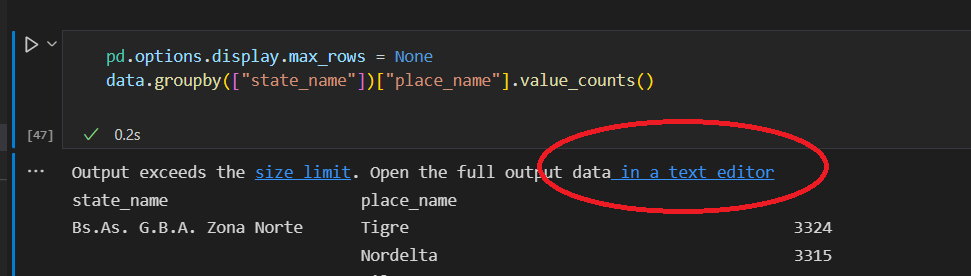
como pregunta/conclusión me queda si pudiéramos organizar la información de place\_name hasta ciudad/partido (tigre) o hasta la localidad (ejm Nordelta) o si la cantidad de datos no es relevante, esto no aplica para cap federal

ver con codigo hay datos de state\_name en place\_name: vi en gba norte usando

**pd**.options.display.max\_rows = None

data.**groupby**(["state\_name"])["place\_name"].**value\_counts**()

Si hacés click en in a text editor te los muestra todos



Yo igual no vi en una mirada rápida muchos casos…

Luis, me ayudas luego con esto, porque me dan diferentes

revisar

data['place\_with\_parent\_names'].apply(lambda x: len(x.split("|"))).value\_counts()

con esto

location\_data['place\_with\_parent\_names'].apply(lambda x: len(x)).value\_counts()

Este segundo cuenta la cantidad de letras del campo. x es cada campo, por ejemplo |Argentina|Capital Federal|Mataderos| y len de x la cantidad de letras (37). Se ve más claro si le sacas el value\_counts para comparar

Para que te dé igual deberías primero splitearlo a un array y ahí len cuenta la cantidad de elementos del array.