

# Inconsistencias de la base de prueba de Bastó

- Inexistencia de un diccionario de datos
- La información de cada animal carece de un timestamp, convirtiendo a los campos weight y isPregnant en poco confiables.
- Falta de lecturas de GPS de los días posteriores al 12 de cada mes en todos los animales y durante todo el período de datos recolectados.

PF-Basto > Lotes y campos y vacas.py > data

Código + Markdown Ejecutar todo Borrar todas las salidas Reiniciar Variables Esquema

data

0.0s

	UUID	Fecha	Hora	Latitud	Longitud	dataRowData	dataRowType
0	0004A3080EDF6FA	2022-01-08	00:00:05	-31.475886	-64.193211	('timestamp': '2022-01-08T00:00:05', 'lat': -3...	GPS
1	0004A3080EDF6FA	2022-01-08	00:11:00	-31.475820	-64.193070	('timestamp': '2022-01-08T00:11:06', 'lat': -3...	GPS
2	0004A3080EDF6FA	2022-01-08	00:21:01	-31.475522	-64.193244	('timestamp': '2022-01-08T00:21:17', 'lat': -3...	GPS
3	0004A3080EDF6FA	2022-01-08	00:31:04	-31.475776	-64.193156	('timestamp': '2022-01-08T00:31:47', 'lat': -3...	GPS
4	0004A3080EDF6FA	2022-01-08	01:02:01	-31.475887	-64.193043	('timestamp': '2022-01-08T01:02:12', 'lat': -3...	GPS
...	...	...	...	...	...	...	...
114637	C168C5FEFF1BEBEA	2022-10-03	17:19:05	-31.476150	-64.193200	('timestamp': '2022-10-03T17:19:53', 'lat': -3...	GPS
114638	C168C5FEFF1BEBEA	2022-11-03	21:22:00	-31.476103	-64.193021	('timestamp': '2022-11-03T21:22:02', 'lat': -3...	GPS
114639	C168C5FEFF1BEBEA	2022-11-03	21:32:01	-31.476191	-64.193217	('timestamp': '2022-11-03T21:32:11', 'lat': -3...	GPS
114640	C168C5FEFF1BEBEA	2022-11-03	21:32:01	-31.476198	-64.193234	('timestamp': '2022-11-03T21:32:13', 'lat': -3...	GPS
114641	C168C5FEFF1BEBEA	2022-11-03	21:42:05	-31.476140	-64.193208	('timestamp': '2022-11-03T21:42:52', 'lat': -3...	GPS

114642 rows x 7 columns

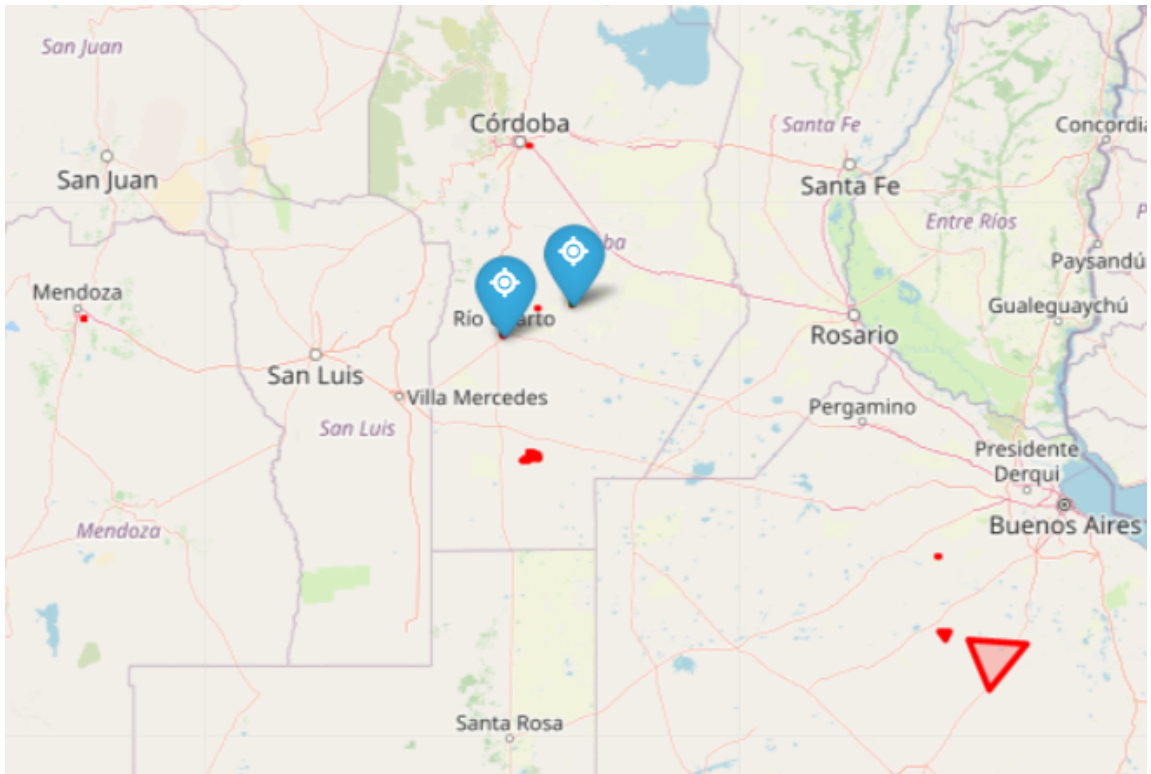
data[data["Fecha"]=="2022-04-14"]

0.0s

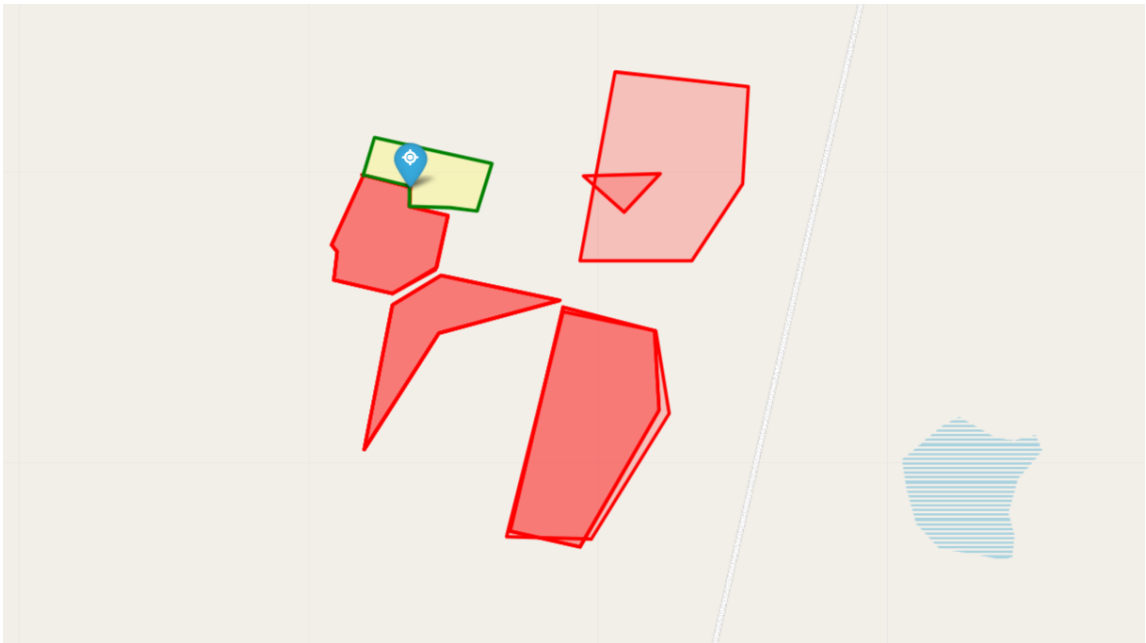
UUID	Fecha	Hora	Latitud	Longitud	dataRowData	dataRowType
------	-------	------	---------	----------	-------------	-------------

IMPORTANTE: En la columna de hechos no hay ninguna lectura de GPS posteriores al día 12 de cada mes. Las lecturas van desde el día primero al 12 inclusive. ata es un dataframe de 14642 lecturas d GPS (luego de eliminads las fechas outliers y filtrada sólo para los dispositivos GPS)

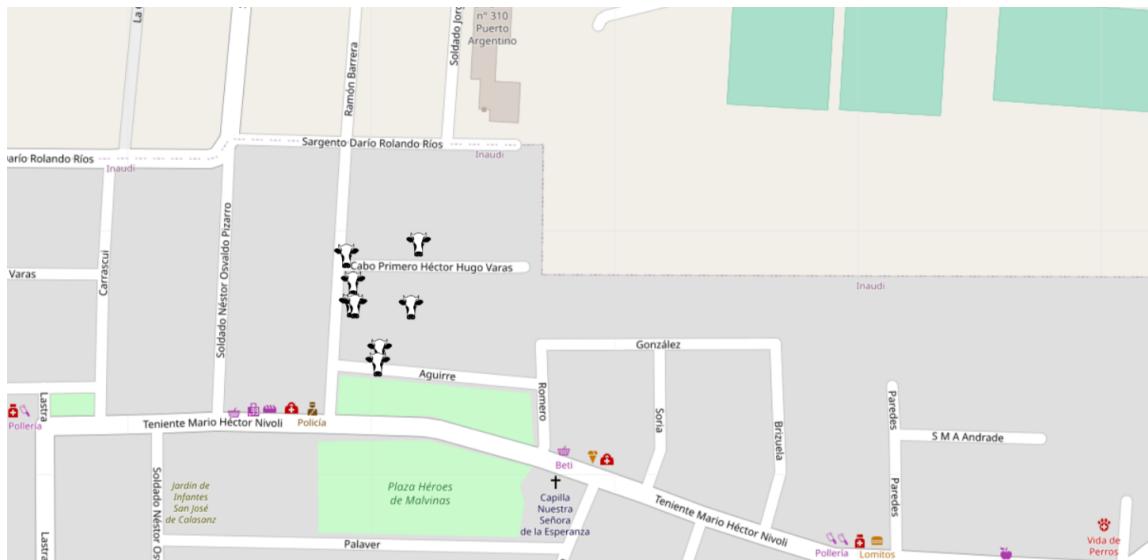
- Inconsistencias con los con la ubicación y límites de los Establecimientos de la colección Settlements. En la gráfica se ve por ejemplo un lote triangular en Prov. de BsAs. que incluye a toda la ciudad de Las Flores, y muy pequeñas parcelas dentro de las ciudades de Córdoba, Río Cuarto, S. Miguel de Tucumán y Mendoza. También hay inconsistencias en el campo centralPoints que indica la ubicación de los establecimientos.



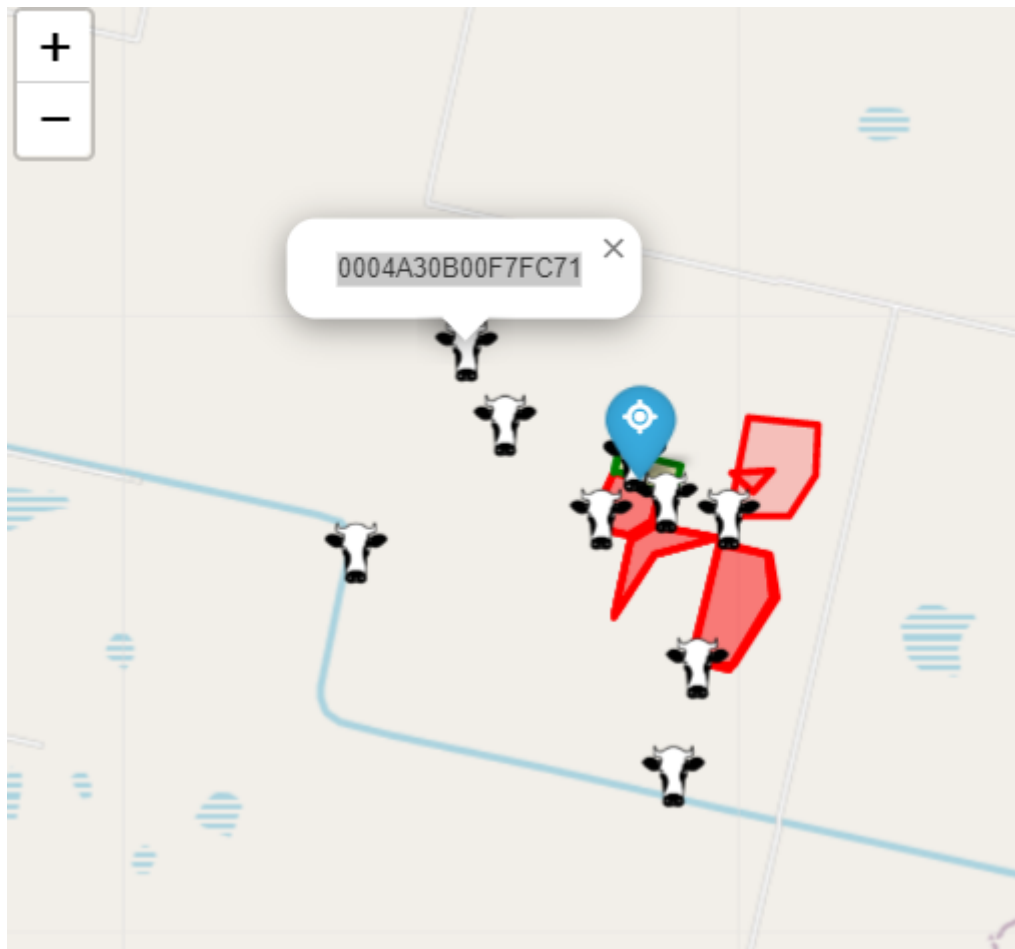
- Se detectaron varias inconsistencias en los datos de la colección Plots en los límites de los lotes en el campo geoPoints. Como ejemplo podemos ver en rojo, los lotes de un establecimiento, y en verde y amarillo los límites del perímetro del mismo establecimiento obtenidos de la colección Settlements



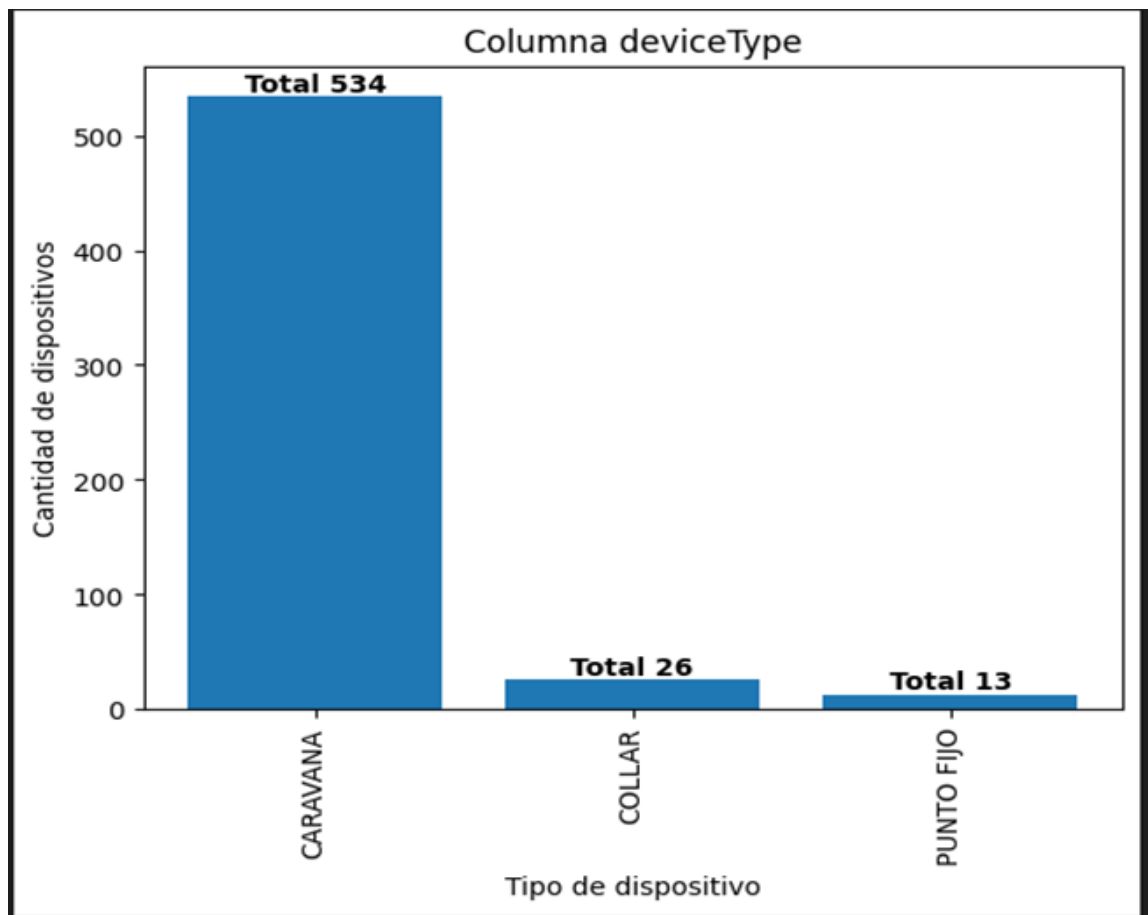
- La posición de algunas vacas con GPS están en lugares aparentemente no aptos para ganadería (en este caso un barrio en los suburbios de la ciudad de Córdoba). De las 30 vacas con GPS, sólo veinte se ubican aparentemente en establecimientos ganaderos, aptos para hacer un seguimiento.



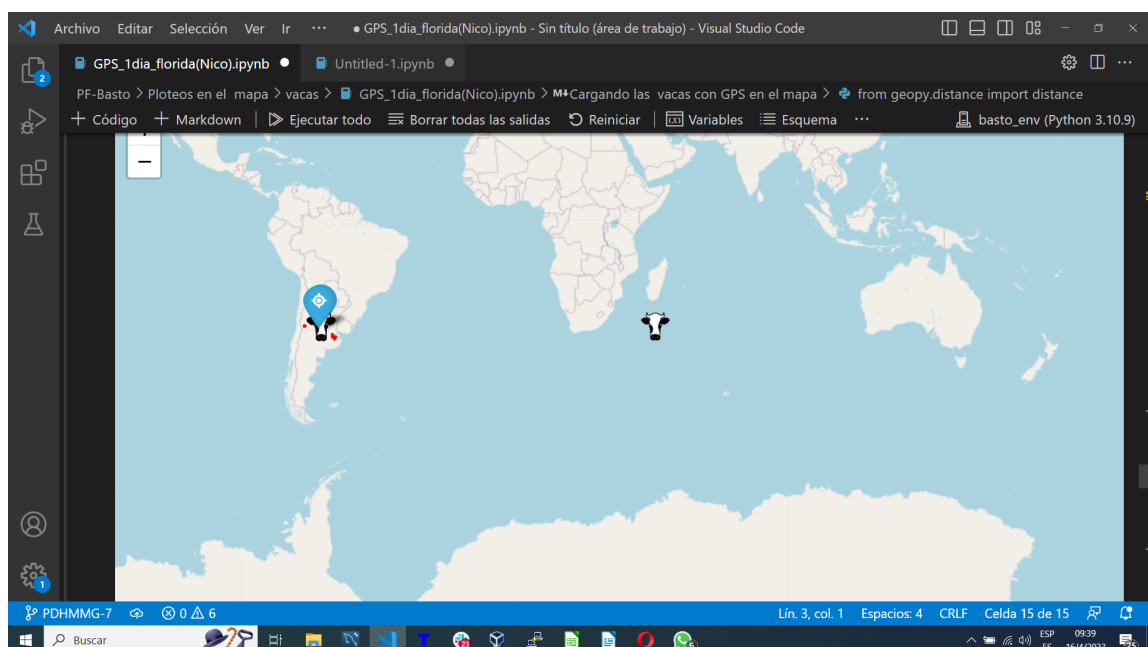
- La posición de los animales con GPS está totalmente afuera de los lotes del establecimiento (GeoPoints de Settlements) e inclusive afuera de los límites del establecimiento



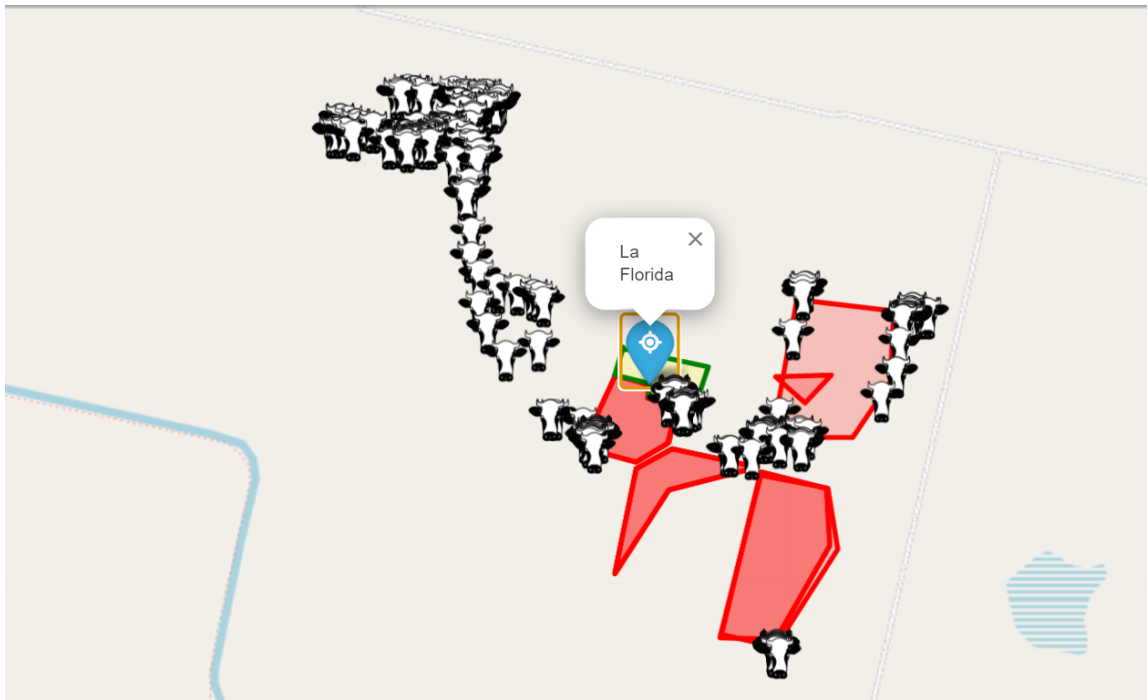
- Inconsistencia entre la cantidad total de animales en la collection "animals" (630), y el número total de dispositivos informado en "devices" en la columna deviceType (total entre collars y caravans: 560). Debería corresponder un dispositivo (ya sea collar o caravana), para cada animal.



Hay outliers en laposición de las vacas reportadas por GPS.Habría que implementar un código que excluya las mediciones que tienen lugar fuera del perímetro del campo.



Graficando un día completos de un único UUID con 313 registros perteneciente al establecimiento La Florida



distancia lineal máxima entre las coordenadas de los puntos más alejados entre si es de 3.

```
# Calcular la distancia máxima y las coordenadas correspondientes
max_distance = 0
max_coords = None

for i in range(len(tuple_list)):
    for j in range(i+1, len(tuple_list)):
        distance_ij = haversine(*tuple_list[i], *tuple_list[j])
        if distance_ij > max_distance:
            max_distance = distance_ij
            max_coords = (tuple_list[i], tuple_list[j])

print("La distancia máxima es", max_distance, "metros, entre las coordenadas", max_coords)
```

La distancia máxima es 3374.385649472407 metros, entre las coordenadas ((-32.864352333333336, -63.6331905), (-32.840919, -63.65614333333333))