Spark_Icesi_Jorge_Lizarazo

March 16, 2024

1 Entrega Parcial Procesamiento en la Nube ”

Jorge Lizarazo & Leandro León Arce

15/03/2024 Proyecto creado como nota parcial en la materia de Procesamiento en la Nube de la Universidad Icesi Evaluada por el Porfesor Daniel Amariles

1.1 Introduction

Utilizando los datos recopilados a través de redes de niebla entre los años 2015 y 2019 en el Cerro Montezuma, área adyacente al Parque Nacional Natural Tatamá, llevamos a cabo un proceso de extracción de información. Este proceso implicó el análisis de varios conjuntos de datos raster, incluyendo aquellos de WorldClim 30seg y Hansen GFC (2022 - treecover200), junto con su recorte mediante un archivo shapefile de tipo polígono con un buffer alrededor del área de reserva (disponible en el enlace de GitHub proporcionado). Posteriormente, fusionamos dos bases de datos: una con información de campo y otra con datos extraídos de los raster. Realizamos una limpieza exhaustiva de los datos para homogeneizar nombres de especies, eliminar información superflua y estandarizar formatos. Utilizando análisis de componentes principales (PCA), determinamos el número mínimo de componentes necesario para representar el 90% de los datos. Luego, mediante el algoritmo de agrupamiento K-means, identificamos qué especies se agrupaban en qué comunidades, dividiendo así el conjunto en tres comunidades distintas. Finalmente, entrenamos un modelo de XGBoost para predecir a qué comunidad pertenecerían las especies a partir de los datos extraídos de los raster. Trasladamos este modelo a un escenario geográfico y generamos distribuciones o regiones ideales para cada comunidad de aves en el área del PNN Tatamá y su zona circundante.

[1]:
```
import findspark
findspark.init()
findspark.find()
```

[1]:
```
'C:\\Users\\ASUS\\anaconda3\\envs\\pyspark-en\\Lib\\site-packages\\pyspark'
```

[2]:
```
## Solo para ver la cosa general con pandas
import pandas as pd
#df = pd.read_csv('Data_ColombiaGradientes_JFCO.csv', sep=',')
#df.head(2)
```

[2]:
```
   Entrada Station        Date  Year month  Day ID Station  net  N:  …  \
0      621   Tatama  25-Feb-14  2014   Feb   25  1    TA01  NaN NaN  …
1      622   Tatama  25-Feb-14  2014   Feb   25  2    TA01  NaN NaN  …
```

```
   Photos  Blood  Parasite F-Chamber M. Feather  \
0      no    NaN       NaN       NaN       Y
1      no    NaN       NaN       NaN       Y

                                          Notes_Bird Num_Colect replica  \
0              Muestras de plumas R4 y 2P, Marca R5        NaN     NaN
1  Muestras 2pecho y 7 rectrices; primarias 1 y 2…        NaN     NaN

  Anillador Notes_General
0       NaN           NaN
1       NaN           NaN

[2 rows x 70 columns]
```

[3]:
```
## Solo general para pillar si son los documentos necesarios
#dj = pd.read_csv('final_df_raster_data.csv', sep=',')
#dj.head(2)
```

[3]:
```
  Estacion   Latitud   Longitud  \
0     TA01  5.223194 -76.080167
1      TA2  5.223056 -76.078000

   clipped_Hansen_GFC-2022-v1.10_lossyear_10N_080W  \
0                                                0
1                                                0

   clipped_Hansen_GFC-2022-v1.10_treecover2000_10N_080W  \
0                                                 45
1                                                 95

   resampled_clipped_wc2.1_30s_bio_1  resampled_clipped_wc2.1_30s_bio_10  \
0                          19.033333                           19.383333
1                          19.033333                           19.383333

   resampled_clipped_wc2.1_30s_bio_11  resampled_clipped_wc2.1_30s_bio_12  \
0                                18.8                              2965.0
1                                18.8                              2965.0

   resampled_clipped_wc2.1_30s_bio_13  …  \
0                               416.0  …
1                               416.0  …

   resampled_clipped_wc2.1_30s_bio_19  resampled_clipped_wc2.1_30s_bio_2  \
0                               970.0                           7.816667
1                               970.0                           7.816667
```

```
   resampled_clipped_wc2.1_30s_bio_3  resampled_clipped_wc2.1_30s_bio_4  \
0                           90.89148                          23.580935
1                           90.89148                          23.580935

   resampled_clipped_wc2.1_30s_bio_5  resampled_clipped_wc2.1_30s_bio_6  \
0                               23.3                               14.7
1                               23.3                               14.7

   resampled_clipped_wc2.1_30s_bio_7  resampled_clipped_wc2.1_30s_bio_8  \
0                           8.599999                               18.8
1                           8.599999                               18.8

   resampled_clipped_wc2.1_30s_bio_9  resampled_clipped_wc2.1_30s_elev
0                          18.983334                              1493
1                          18.983334                              1493

[2 rows x 25 columns]
```

[4]:
```
import pyspark
from pyspark.sql import SparkSession
from pyspark.sql.functions import explode, split, coalesce, concat
from pyspark.sql.functions import when, count, col, regexp_replace, lit
```

[5]:
```
spark = SparkSession\
    .builder \
    .appName("Proyecto_Gradientes_Tatama") \
    .config("spark.driver.memory", "4g") \
    .getOrCreate()
```

[229]:
```
import requests
# URL del archivo CSV en GitHub
csv_url = 'https://raw.githubusercontent.com/jorgelizarazo94/Cloud_procesing/
 ↪6487afb8aa687ce23512c4962e79a432f08cc32d/Data/final_df_raster_data.csv'

# Utiliza requests para descargar el archivo CSV
r = requests.get(csv_url)
with open('final_df_raster_data.csv', 'wb') as f:
    f.write(r.content)



csv_url2 = 'https://github.com/jorgelizarazo94/Cloud_procesing/blob/
 ↪6487afb8aa687ce23512c4962e79a432f08cc32d/Data/Data_ColombiaGradientes_JFCO.
 ↪csv'

# Utiliza requests para descargar el archivo CSV
r = requests.get(csv_url2)
```

```
with open('Data_ColombiaGradientes_JFCO.csv', 'wb') as f:
    f.write(r.content)
```

[230]:
```
dj_spark = spark.read.option('header', 'true').csv('final_df_raster_data.csv',
                                                   header=True,␣
 ↪inferSchema=True)
dj_spark = dj_spark.withColumnRenamed("Estacion", "Station")
#dj_spark = dj_spark.withColumnRenamed("Elevacion", "Elevation")

dj_spark.show(20)
```

```
+-------+-----------------+-----------------+--------------------------------
-------------+-----------------------------------------------------+-----------
--------------------+-----------------------------------+----------------------
----------+-----------------------------------+--------------------------------
-+-----------------------------------+-----------------------------------+-------
-------------------------+-----------------------------------+-----------------
---------------+-----------------------------------+---------------------------
-----+-----------------------------------+-----------------------------------+------
-------------------------+-----------------------------------+-----------------
--------------+-----------------------------------+---------------------------
---+-----------------------------------+
|Station|           Latitud|          Longitud|clipped_Hansen_GFC-2022-
v1.10_lossyear_10N_080W|clipped_Hansen_GFC-2022-
v1.10_treecover2000_10N_080W|resampled_clipped_wc2.1_30s_bio_1|resampled_clipped
_wc2.1_30s_bio_10|resampled_clipped_wc2.1_30s_bio_11|resampled_clipped_wc2.1_30s
_bio_12|resampled_clipped_wc2.1_30s_bio_13|resampled_clipped_wc2.1_30s_bio_14|re
sampled_clipped_wc2.1_30s_bio_15|resampled_clipped_wc2.1_30s_bio_16|resampled_cl
ipped_wc2.1_30s_bio_17|resampled_clipped_wc2.1_30s_bio_18|resampled_clipped_wc2.
1_30s_bio_19|resampled_clipped_wc2.1_30s_bio_2|resampled_clipped_wc2.1_30s_bio_3
|resampled_clipped_wc2.1_30s_bio_4|resampled_clipped_wc2.1_30s_bio_5|resampled_c
lipped_wc2.1_30s_bio_6|resampled_clipped_wc2.1_30s_bio_7|resampled_clipped_wc2.1
_30s_bio_8|resampled_clipped_wc2.1_30s_bio_9|resampled_clipped_wc2.1_30s_elev|
+-------+-----------------+-----------------+--------------------------------
-------------+-----------------------------------------------------+-----------
--------------------+-----------------------------------+----------------------
----------+-----------------------------------+--------------------------------
-+-----------------------------------+-----------------------------------+-------
-------------------------+-----------------------------------+-----------------
---------------+-----------------------------------+---------------------------
-----+-----------------------------------+-----------------------------------+------
-------------------------+-----------------------------------+-----------------
--------------+-----------------------------------+---------------------------
---+-----------------------------------+
|   TA01| 5.223194444444444|-76.08016666666666|
0|                                                  45|
```

19.033333| 19.383333| 18.8|
2965.0| 416.0| 152.0|
34.28378| 970.0| 532.0|
875.0| 970.0| 7.816667|
90.89148| 23.580935| 23.3|
14.7| 8.599999| 18.8|
18.983334| 1493|
| TA2| 5.223055555555556| -76.078|
0| 95|
19.033333| 19.383333| 18.8|
2965.0| 416.0| 152.0|
34.28378| 970.0| 532.0|
875.0| 970.0| 7.816667|
90.89148| 23.580935| 23.3|
14.7| 8.599999| 18.8|
18.983334| 1493|
| TA3| 5.219972222222222|-76.07927777777778|
0| 92|
19.033333| 19.383333| 18.8|
2965.0| 416.0| 152.0|
34.28378| 970.0| 532.0|
875.0| 970.0| 7.816667|
90.89148| 23.580935| 23.3|
14.7| 8.599999| 18.8|
18.983334| 1493|
| TA4|5.2322500000000005|-76.09191666666666|
0| 95|
18.7875| 19.116667| 18.55|
2605.0| 379.0| 94.0|
35.614826| 862.0| 458.0|
765.0| 862.0| 7.741667|
92.162704| 22.373373| 22.9|
14.5| 8.4| 18.55|
18.9| 1544|
| TA5| 5.240222222222222|-76.09208333333333|
0| 90|
18.754168| 19.083334| 18.5|
3087.0| 414.0| 163.0|
26.5564| 921.0| 637.0|
853.0| 885.0| 7.725|
91.96429| 23.204458| 22.9|
14.5| 8.4| 18.5|
18.666666| 1568|
| TA6| 5.244694444444445| -76.09375|
0| 90|
18.3375| 18.633333| 18.066666|
3066.0| 418.0| 162.0|
31.792627| 992.0| 555.0|

```
875.0|                              992.0|                          7.758333|
92.361115|                         22.676077|                             22.5|
14.1|                               8.4|                        18.066666|
18.366667|                            1667|
|    TA7|           5.24625| -76.0961111111111|
0|                                                    90|
18.3375|                         18.633333|                          18.066666|
3066.0|                            418.0|                             162.0|
31.792627|                            992.0|                              555.0|
875.0|                              992.0|                          7.758333|
92.361115|                         22.676077|                             22.5|
14.1|                               8.4|                        18.066666|
18.366667|                            1667|
|    TA8| 5.238166666666666|-76.08394444444444|
0|                                                    92|
19.8875|                             20.25|                             19.6|
2871.0|                            363.0|                             128.0|
31.151634|                            892.0|                              474.0|
868.0|                              892.0|                            8.425|
91.57609|                          25.77216|                             24.5|
15.3|                               9.2|                             19.6|
20.05|                             1377|
|    TA9| 5.247638888888889|-76.09861111111111|
0|                                                    90|
18.3375|                         18.633333|                          18.066666|
3066.0|                            418.0|                             162.0|
31.792627|                            992.0|                              555.0|
875.0|                              992.0|                          7.758333|
92.361115|                         22.676077|                             22.5|
14.1|                               8.4|                        18.066666|
18.366667|                            1667|
|   TA10| 5.249138888888889|-76.14944444444444|
0|                                                    95|
16.729166|                         16.966667|                          16.449999|
2485.0|                             373.0|                             123.0|
37.646206|                            848.0|                              397.0|
685.0|                              814.0|                         7.2416663|
89.4033|                         21.474976|                             20.8|
12.7|                            8.099999|                        16.466667|
16.816666|                            1938|
|   TA11| 5.248777777777778|          -76.1275|
0|                                                    95|
15.220833|                         15.400001|                          14.966666|
2248.0|                             309.0|                              82.0|
38.576023|                            771.0|                              316.0|
649.0|                              771.0|                         7.1749997|
89.68751|                         19.593414|                             19.3|
11.3|                            7.999999|                        14.966666|
```

```
15.316667|                                2228|
|   TA12| 5.251166666666666|-76.10408333333332|
0|                                                                90|
16.629166|                          16.866667|                          16.366667|
2140.0|                            289.0|                             83.0|
37.20048|                            745.0|                             323.0|
602.0|                           745.0|                          7.408333|
90.345535|                         20.830303|                             20.8|
12.6|                         8.199999|                       16.366667|
16.733334|                                1984|
|   TA13| 5.253777777777778|-76.10866666666666|
0|                                                                99|
14.575|                            14.75|                          14.333333|
1992.0|                            260.0|                             82.0|
36.895878|                             667.0|                             308.0|
595.0|                           667.0|                        7.2166667|
90.20832|                         19.128756|                             18.7|
10.7|                         8.000001|                       14.333333|
14.666666|                                2360|
|   TA14| 5.257611111111111|-76.11013888888888|
0|                                                                90|
14.575|                            14.75|                          14.333333|
1992.0|                            260.0|                             82.0|
36.895878|                             667.0|                             308.0|
595.0|                           667.0|                        7.2166667|
90.20832|                         19.128756|                             18.7|
10.7|                         8.000001|                       14.333333|
14.666666|                                2360|
|   TA15| 5.256888888888889| -76.1121111111111|
0|                                                                90|
14.575|                            14.75|                          14.333333|
1992.0|                            260.0|                             82.0|
36.895878|                             667.0|                             308.0|
595.0|                           667.0|                        7.2166667|
90.20832|                         19.128756|                             18.7|
10.7|                         8.000001|                       14.333333|
14.666666|                                2360|
|   TA16| 5.256194444444445|-76.11355555555555|
0|                                                                90|
14.575|                            14.75|                          14.333333|
1992.0|                            260.0|                             82.0|
36.895878|                             667.0|                             308.0|
595.0|                           667.0|                        7.2166667|
90.20832|                         19.128756|                             18.7|
10.7|                         8.000001|                       14.333333|
14.666666|                                2360|
| TA01_1|  5.22388542745007|-76.07804003964577|
0|                                                                95|
```

```
19.033333|                          19.383333|                          18.8|
2965.0|                           416.0|                         152.0|
34.28378|                            970.0|                           532.0|
875.0|                           970.0|                       7.816667|
90.89148|                         23.580935|                          23.3|
14.7|                         8.599999|                        18.8|
18.983334|                            1493|
| TA01_2| 5.224508772224742|-76.07835764967228|
0|                                                  95|
19.033333|                          19.383333|                          18.8|
2965.0|                           416.0|                         152.0|
34.28378|                            970.0|                           532.0|
875.0|                           970.0|                       7.816667|
90.89148|                         23.580935|                          23.3|
14.7|                         8.599999|                        18.8|
18.983334|                            1493|
| TA01_3|  5.22500346143882|-76.07885233888635|
0|                                                  95|
19.033333|                          19.383333|                          18.8|
2965.0|                           416.0|                         152.0|
34.28378|                            970.0|                           532.0|
875.0|                           970.0|                       7.816667|
90.89148|                         23.580935|                          23.3|
14.7|                         8.599999|                        18.8|
18.983334|                            1493|
| TA01_4| 5.225321071465324|-76.07947568366103|
0|                                                  95|
19.620832|                          19.966667|                         19.366667|
2351.0|                           306.0|                           88.0|
37.38887|                            784.0|                           344.0|
720.0|                           725.0|                        8.224999|
91.388885|                         24.069817|                          24.1|
15.1|                           9.0|                      19.816668|
19.766666|                            1420|
+-------+------------------+------------------+--------------------------------
-------------+-----------------------------------------------------+-----------
---------------------+-----------------------------------+----------------------
-----------+-----------------------------------+--------------------------------
-+---------------------------------+---------------------------------+--------
-------------------------+-----------------------------------+------------------
---------------+-----------------------------------+---------------------------
-----+---------------------------------+---------------------------------+------
-------------------------+-----------------------------------+------------------
--------------+---------------------------------+---------------------------
---+---------------------------------+
only showing top 20 rows
```

[7]:
```
# Split the 'Station' column based on '_'
split_col = split(dj_spark['Station'], '_')
dj_spark = dj_spark.withColumn('Station_new', split_col.getItem(0)) \
                   .withColumn('Station_sample',
                               when(split_col.getItem(1).isNull(),
                                    lit('first'))
                               .otherwise(concat(lit('S'), split_col.
 ↪getItem(1))))

# Show the result
dj_spark.show(20)
```

```
+-------+-----------------+-----------------+--------------------------------
-------------+-----------------------------------------------------+-----------
--------------------+-----------------------------------+----------------------
----------+-----------------------------------+--------------------------------
-+-----------------------------------+-----------------------------------+-------
--------------------------+-----------------------------------+-----------------
---------------+-----------------------------------+---------------------------
-----+-----------------------------------+-----------------------------------+------
--------------------------+-----------------------------------+-----------------
---------------+-----------------------------------+---------------------------
---+-----------------------------------+----------+--------------+
|Station|           Latitud|          Longitud|clipped_Hansen_GFC-2022-
v1.10_lossyear_10N_080W|clipped_Hansen_GFC-2022-
v1.10_treecover2000_10N_080W|resampled_clipped_wc2.1_30s_bio_1|resampled_clipped
_wc2.1_30s_bio_10|resampled_clipped_wc2.1_30s_bio_11|resampled_clipped_wc2.1_30s
_bio_12|resampled_clipped_wc2.1_30s_bio_13|resampled_clipped_wc2.1_30s_bio_14|re
sampled_clipped_wc2.1_30s_bio_15|resampled_clipped_wc2.1_30s_bio_16|resampled_cl
ipped_wc2.1_30s_bio_17|resampled_clipped_wc2.1_30s_bio_18|resampled_clipped_wc2.
1_30s_bio_19|resampled_clipped_wc2.1_30s_bio_2|resampled_clipped_wc2.1_30s_bio_3
|resampled_clipped_wc2.1_30s_bio_4|resampled_clipped_wc2.1_30s_bio_5|resampled_c
lipped_wc2.1_30s_bio_6|resampled_clipped_wc2.1_30s_bio_7|resampled_clipped_wc2.1
_30s_bio_8|resampled_clipped_wc2.1_30s_bio_9|resampled_clipped_wc2.1_30s_elev|St
ation_new|Station_sample|
+-------+-----------------+-----------------+--------------------------------
-------------+-----------------------------------------------------+-----------
--------------------+-----------------------------------+----------------------
----------+-----------------------------------+--------------------------------
-+-----------------------------------+-----------------------------------+-------
--------------------------+-----------------------------------+-----------------
---------------+-----------------------------------+---------------------------
-----+-----------------------------------+-----------------------------------+------
--------------------------+-----------------------------------+-----------------
---------------+-----------------------------------+---------------------------
---+-----------------------------------+----------+--------------+
|   TA01| 5.223194444444444|-76.08016666666666|
```

```
0|                                                  45|
19.033333|                          19.383333|                                18.8|
2965.0|                           416.0|                             152.0|
34.28378|                            970.0|                              532.0|
875.0|                           970.0|                          7.816667|
90.89148|                         23.580935|                               23.3|
14.7|                         8.599999|                            18.8|
18.983334|                              1493|       TA01|         first|
|    TA2| 5.223055555555556|           -76.078|
0|                                                  95|
19.033333|                          19.383333|                                18.8|
2965.0|                           416.0|                             152.0|
34.28378|                            970.0|                              532.0|
875.0|                           970.0|                          7.816667|
90.89148|                         23.580935|                               23.3|
14.7|                         8.599999|                            18.8|
18.983334|                              1493|       TA2|         first|
|    TA3| 5.219972222222222|-76.07927777777778|
0|                                                  92|
19.033333|                          19.383333|                                18.8|
2965.0|                           416.0|                             152.0|
34.28378|                            970.0|                              532.0|
875.0|                           970.0|                          7.816667|
90.89148|                         23.580935|                               23.3|
14.7|                         8.599999|                            18.8|
18.983334|                              1493|       TA3|         first|
|    TA4|5.2322500000000005|-76.09191666666666|
0|                                                  95|
18.7875|                         19.116667|                               18.55|
2605.0|                            379.0|                               94.0|
35.614826|                             862.0|                              458.0|
765.0|                           862.0|                          7.741667|
92.162704|                         22.373373|                               22.9|
14.5|                            8.4|                             18.55|
18.9|                           1544|        TA4|         first|
|    TA5| 5.240222222222222|-76.09208333333333|
0|                                                  90|
18.754168|                          19.083334|                                18.5|
3087.0|                           414.0|                             163.0|
26.5564|                            921.0|                              637.0|
853.0|                           885.0|                            7.725|
91.96429|                         23.204458|                               22.9|
14.5|                            8.4|                            18.5|
18.666666|                             1568|       TA5|         first|
|    TA6| 5.244694444444445|         -76.09375|
0|                                                  90|
18.3375|                          18.633333|                            18.066666|
3066.0|                            418.0|                               162.0|
```

```
31.792627|                              992.0|                              555.0|
875.0|                             992.0|                            7.758333|
92.361115|                           22.676077|                              22.5|
14.1|                              8.4|                           18.066666|
18.366667|                              1667|        TA6|         first|
|    TA7|           5.24625| -76.0961111111111|
0|                                                 90|
18.3375|                           18.633333|                              18.066666|
3066.0|                              418.0|                                 162.0|
31.792627|                              992.0|                              555.0|
875.0|                             992.0|                            7.758333|
92.361115|                           22.676077|                              22.5|
14.1|                              8.4|                           18.066666|
18.366667|                              1667|        TA7|         first|
|    TA8| 5.238166666666666|-76.08394444444444|
0|                                                 92|
19.8875|                             20.25|                                 19.6|
2871.0|                             363.0|                                128.0|
31.151634|                              892.0|                              474.0|
868.0|                             892.0|                               8.425|
91.57609|                           25.77216|                               24.5|
15.3|                              9.2|                              19.6|
20.05|                             1377|        TA8|         first|
|    TA9| 5.247638888888889|-76.09861111111111|
0|                                                 90|
18.3375|                           18.633333|                              18.066666|
3066.0|                              418.0|                                 162.0|
31.792627|                              992.0|                              555.0|
875.0|                             992.0|                            7.758333|
92.361115|                           22.676077|                              22.5|
14.1|                              8.4|                           18.066666|
18.366667|                              1667|        TA9|         first|
|   TA10| 5.249138888888889|-76.14944444444444|
0|                                                 95|
16.729166|                           16.966667|                              16.449999|
2485.0|                             373.0|                                123.0|
37.646206|                              848.0|                              397.0|
685.0|                             814.0|                             7.2416663|
89.4033|                          21.474976|                                20.8|
12.7|                            8.099999|                           16.466667|
16.816666|                              1938|       TA10|         first|
|   TA11| 5.248777777777778|          -76.1275|
0|                                                 95|
15.220833|                           15.400001|                              14.966666|
2248.0|                             309.0|                                 82.0|
38.576023|                              771.0|                              316.0|
649.0|                             771.0|                             7.1749997|
89.68751|                           19.593414|                                19.3|
```

```
11.3|                          7.999999|                          14.966666|
15.316667|                              2228|       TA11|         first|
|   TA12| 5.251166666666666|-76.10408333333332|
0|                                                    90|
16.629166|                          16.866667|                          16.366667|
2140.0|                              289.0|                              83.0|
37.20048|                              745.0|                              323.0|
602.0|                             745.0|                            7.408333|
90.345535|                          20.830303|                              20.8|
12.6|                          8.199999|                          16.366667|
16.733334|                              1984|       TA12|         first|
|   TA13| 5.253777777777778|-76.10866666666666|
0|                                                    99|
14.575|                              14.75|                           14.333333|
1992.0|                              260.0|                              82.0|
36.895878|                              667.0|                              308.0|
595.0|                             667.0|                           7.2166667|
90.20832|                          19.128756|                              18.7|
10.7|                          8.000001|                          14.333333|
14.666666|                              2360|       TA13|         first|
|   TA14| 5.257611111111111|-76.11013888888888|
0|                                                    90|
14.575|                              14.75|                           14.333333|
1992.0|                              260.0|                              82.0|
36.895878|                              667.0|                              308.0|
595.0|                             667.0|                           7.2166667|
90.20832|                          19.128756|                              18.7|
10.7|                          8.000001|                          14.333333|
14.666666|                              2360|       TA14|         first|
|   TA15| 5.256888888888889| -76.1121111111111|
0|                                                    90|
14.575|                              14.75|                           14.333333|
1992.0|                              260.0|                              82.0|
36.895878|                              667.0|                              308.0|
595.0|                             667.0|                           7.2166667|
90.20832|                          19.128756|                              18.7|
10.7|                          8.000001|                          14.333333|
14.666666|                              2360|       TA15|         first|
|   TA16| 5.256194444444445|-76.11355555555555|
0|                                                    90|
14.575|                              14.75|                           14.333333|
1992.0|                              260.0|                              82.0|
36.895878|                              667.0|                              308.0|
595.0|                             667.0|                           7.2166667|
90.20832|                          19.128756|                              18.7|
10.7|                          8.000001|                          14.333333|
14.666666|                              2360|       TA16|         first|
| TA01_1|  5.22388542745007|-76.07804003964577|
```

```
0|                                                  95|
19.033333|                          19.383333|                              18.8|
2965.0|                           416.0|                           152.0|
34.28378|                             970.0|                             532.0|
875.0|                           970.0|                         7.816667|
90.89148|                         23.580935|                              23.3|
14.7|                        8.599999|                           18.8|
18.983334|                             1493|       TA01|            S1|
| TA01_2| 5.224508772224742|-76.07835764967228|
0|                                                  95|
19.033333|                          19.383333|                              18.8|
2965.0|                           416.0|                           152.0|
34.28378|                             970.0|                             532.0|
875.0|                           970.0|                         7.816667|
90.89148|                         23.580935|                              23.3|
14.7|                        8.599999|                           18.8|
18.983334|                             1493|       TA01|            S2|
| TA01_3|  5.22500346143882|-76.07885233888635|
0|                                                  95|
19.033333|                          19.383333|                              18.8|
2965.0|                           416.0|                           152.0|
34.28378|                             970.0|                             532.0|
875.0|                           970.0|                         7.816667|
90.89148|                         23.580935|                              23.3|
14.7|                        8.599999|                           18.8|
18.983334|                             1493|       TA01|            S3|
| TA01_4| 5.225321071465324|-76.07947568366103|
0|                                                  95|
19.620832|                          19.966667|                          19.366667|
2351.0|                           306.0|                            88.0|
37.38887|                             784.0|                             344.0|
720.0|                           725.0|                         8.224999|
91.388885|                         24.069817|                              24.1|
15.1|                           9.0|                       19.816668|
19.766666|                             1420|       TA01|            S4|
+-------+-----------------+-----------------+--------------------------------
-------------+-----------------------------------------------------+-----------
---------------------+-----------------------------------+----------------------
-----------+---------------------------------+--------------------------------
-+---------------------------------+---------------------------------+--------
-------------------------+---------------------------------+-----------------
---------------+---------------------------------+--------------------------
-----+---------------------------------+---------------------------------+------
-------------------------+---------------------------------+-----------------
--------------+---------------------------------+--------------------------
---+-------------------------------+----------+-------------+
only showing top 20 rows
```

[8]:
```
dj_spark = dj_spark.drop('Station').withColumnRenamed('Station_new', 'Station')

dj_spark.show(20, False)
```

```
+-----------------+-----------------+-----------------------------------------
------+-------------------------------------------------------+-------------------
------------+-----------------------------------+-------------------------------
---+-----------------------------------+-----------------------------------+------
--------------------------+-----------------------------------+---------------
-----------------+-----------------------------------+-----------------------
--------+-----------------------------------+-----------------------------------+--
-------------------------------+-----------------------------------+-------------
-----------------+-----------------------------------+-----------------------
-------+-----------------------------------+-----------------------------------+----
--------------------------+-------+--------------+
|Latitud           |Longitud          |clipped_Hansen_GFC-2022-
v1.10_lossyear_10N_080W|clipped_Hansen_GFC-2022-
v1.10_treecover2000_10N_080W|resampled_clipped_wc2.1_30s_bio_1|resampled_clipped
_wc2.1_30s_bio_10|resampled_clipped_wc2.1_30s_bio_11|resampled_clipped_wc2.1_30s
_bio_12|resampled_clipped_wc2.1_30s_bio_13|resampled_clipped_wc2.1_30s_bio_14|re
sampled_clipped_wc2.1_30s_bio_15|resampled_clipped_wc2.1_30s_bio_16|resampled_cl
ipped_wc2.1_30s_bio_17|resampled_clipped_wc2.1_30s_bio_18|resampled_clipped_wc2.
1_30s_bio_19|resampled_clipped_wc2.1_30s_bio_2|resampled_clipped_wc2.1_30s_bio_3
|resampled_clipped_wc2.1_30s_bio_4|resampled_clipped_wc2.1_30s_bio_5|resampled_c
lipped_wc2.1_30s_bio_6|resampled_clipped_wc2.1_30s_bio_7|resampled_clipped_wc2.1
_30s_bio_8|resampled_clipped_wc2.1_30s_bio_9|resampled_clipped_wc2.1_30s_elev|St
ation|Station_sample|
+-----------------+-----------------+-----------------------------------------
------+-------------------------------------------------------+-------------------
------------+-----------------------------------+-------------------------------
---+-----------------------------------+-----------------------------------+------
--------------------------+-----------------------------------+---------------
-----------------+-----------------------------------+-----------------------
--------+-----------------------------------+-----------------------------------+--
-------------------------------+-----------------------------------+-------------
-----------------+-----------------------------------+-----------------------
-------+-----------------------------------+-----------------------------------+----
--------------------------+-------+--------------+
|5.223194444444444 |-76.08016666666666|0
|45                                                    |19.033333
|19.383333                          |18.8                               |2965.0
|416.0                              |152.0                              |34.28378
|970.0                              |532.0                              |875.0
|970.0                              |7.816667                           |90.89148
|23.580935                          |23.3                               |14.7
|8.599999                           |18.8                               |18.983334
|1493                               |TA01   |first         |
```

```
|5.223055555555556 |-76.078           |0
|95                                                |19.033333
|19.383333                          |18.8                                |2965.0
|416.0                              |152.0                               |34.28378
|970.0                              |532.0                               |875.0
|970.0                              |7.816667                           |90.89148
|23.580935                          |23.3                               |14.7
|8.599999                           |18.8                               |18.983334
|1493                               |TA2    |first         |
|5.219972222222222 |-76.07927777777778|0
|92                                                |19.033333
|19.383333                          |18.8                                |2965.0
|416.0                              |152.0                               |34.28378
|970.0                              |532.0                               |875.0
|970.0                              |7.816667                           |90.89148
|23.580935                          |23.3                               |14.7
|8.599999                           |18.8                               |18.983334
|1493                               |TA3    |first         |
|5.2322500000000005|-76.09191666666666|0
|95                                                |18.7875
|19.116667                          |18.55                               |2605.0
|379.0                              |94.0                                |35.614826
|862.0                              |458.0                               |765.0
|862.0                              |7.741667                           |92.162704
|22.373373                          |22.9                               |14.5
|8.4                                |18.55                              |18.9
|1544                               |TA4    |first         |
|5.240222222222222 |-76.09208333333333|0
|90                                                |18.754168
|19.083334                          |18.5                                |3087.0
|414.0                              |163.0                               |26.5564
|921.0                              |637.0                               |853.0
|885.0                              |7.725                              |91.96429
|23.204458                          |22.9                               |14.5
|8.4                                |18.5                               |18.666666
|1568                               |TA5    |first         |
|5.244694444444445 |-76.09375         |0
|90                                                |18.3375
|18.633333                          |18.066666                           |3066.0
|418.0                              |162.0                               |31.792627
|992.0                              |555.0                               |875.0
|992.0                              |7.758333                           |92.361115
|22.676077                          |22.5                               |14.1
|8.4                                |18.066666                          |18.366667
|1667                               |TA6    |first         |
|5.24625           |-76.0961111111111 |0
|90                                                |18.3375
|18.633333                          |18.066666                           |3066.0
```

```
|418.0                              |162.0                              |31.792627
|992.0                              |555.0                              |875.0
|992.0                              |7.758333                           |92.361115
|22.676077                          |22.5                               |14.1
|8.4                                |18.066666                          |18.366667
|1667                               |TA7    |first         |
|5.238166666666666 |-76.08394444444444|0
|92                                                     |19.8875
|20.25                              |19.6                               |2871.0
|363.0                              |128.0                              |31.151634
|892.0                              |474.0                              |868.0
|892.0                              |8.425                              |91.57609
|25.77216                           |24.5                               |15.3
|9.2                                |19.6                               |20.05
|1377                               |TA8    |first         |
|5.247638888888889 |-76.09861111111111|0
|90                                                     |18.3375
|18.633333                          |18.066666                          |3066.0
|418.0                              |162.0                              |31.792627
|992.0                              |555.0                              |875.0
|992.0                              |7.758333                           |92.361115
|22.676077                          |22.5                               |14.1
|8.4                                |18.066666                          |18.366667
|1667                               |TA9    |first         |
|5.249138888888889 |-76.14944444444444|0
|95                                                     |16.729166
|16.966667                          |16.449999                          |2485.0
|373.0                              |123.0                              |37.646206
|848.0                              |397.0                              |685.0
|814.0                              |7.2416663                          |89.4033
|21.474976                          |20.8                               |12.7
|8.099999                           |16.466667                          |16.816666
|1938                               |TA10   |first         |
|5.248777777777778 |-76.1275           |0
|95                                                     |15.220833
|15.400001                          |14.966666                          |2248.0
|309.0                              |82.0                               |38.576023
|771.0                              |316.0                              |649.0
|771.0                              |7.1749997                          |89.68751
|19.593414                          |19.3                               |11.3
|7.999999                           |14.966666                          |15.316667
|2228                               |TA11   |first         |
|5.251166666666666 |-76.10408333333332|0
|90                                                     |16.629166
|16.866667                          |16.366667                          |2140.0
|289.0                              |83.0                               |37.20048
|745.0                              |323.0                              |602.0
|745.0                              |7.408333                           |90.345535
```

```
|20.830303                        |20.8                             |12.6
|8.199999                         |16.366667                        |16.733334
|1984                             |TA12   |first         |
|5.253777777777778 |-76.10866666666666|0
|99                                                    |14.575
|14.75                             |14.333333                        |1992.0
|260.0                             |82.0                             |36.895878
|667.0                             |308.0                            |595.0
|667.0                             |7.2166667                        |90.20832
|19.128756                         |18.7                             |10.7
|8.000001                          |14.333333                        |14.666666
|2360                              |TA13   |first         |
|5.257611111111111 |-76.11013888888888|0
|90                                                    |14.575
|14.75                             |14.333333                        |1992.0
|260.0                             |82.0                             |36.895878
|667.0                             |308.0                            |595.0
|667.0                             |7.2166667                        |90.20832
|19.128756                         |18.7                             |10.7
|8.000001                          |14.333333                        |14.666666
|2360                              |TA14   |first         |
|5.256888888888889 |-76.1121111111111 |0
|90                                                    |14.575
|14.75                             |14.333333                        |1992.0
|260.0                             |82.0                             |36.895878
|667.0                             |308.0                            |595.0
|667.0                             |7.2166667                        |90.20832
|19.128756                         |18.7                             |10.7
|8.000001                          |14.333333                        |14.666666
|2360                              |TA15   |first         |
|5.256194444444445 |-76.11355555555555|0
|90                                                    |14.575
|14.75                             |14.333333                        |1992.0
|260.0                             |82.0                             |36.895878
|667.0                             |308.0                            |595.0
|667.0                             |7.2166667                        |90.20832
|19.128756                         |18.7                             |10.7
|8.000001                          |14.333333                        |14.666666
|2360                              |TA16   |first         |
|5.22388542745007  |-76.07804003964577|0
|95                                                    |19.033333
|19.383333                         |18.8                             |2965.0
|416.0                             |152.0                            |34.28378
|970.0                             |532.0                            |875.0
|970.0                             |7.816667                         |90.89148
|23.580935                         |23.3                             |14.7
|8.599999                          |18.8                             |18.983334
|1493                              |TA01   |S1            |
```

```
|5.224508772224742 |-76.07835764967228|0
|95                                                    |19.033333
|19.383333                          |18.8                                |2965.0
|416.0                              |152.0                               |34.28378
|970.0                              |532.0                               |875.0
|970.0                              |7.816667                            |90.89148
|23.580935                          |23.3                                |14.7
|8.599999                           |18.8                                |18.983334
|1493                               |TA01   |S2            |
|5.22500346143882  |-76.07885233888635|0
|95                                                    |19.033333
|19.383333                          |18.8                                |2965.0
|416.0                              |152.0                               |34.28378
|970.0                              |532.0                               |875.0
|970.0                              |7.816667                            |90.89148
|23.580935                          |23.3                                |14.7
|8.599999                           |18.8                                |18.983334
|1493                               |TA01   |S3            |
|5.225321071465324 |-76.07947568366103|0
|95                                                    |19.620832
|19.966667                          |19.366667                           |2351.0
|306.0                              |88.0                                |37.38887
|784.0                              |344.0                               |720.0
|725.0                              |8.224999                            |91.388885
|24.069817                          |24.1                                |15.1
|9.0                                |19.816668                           |19.766666
|1420                               |TA01   |S4            |
+-------------------+------------------+-----------------------------------------
------+-----------------------------------------------------+--------------------
------------+-----------------------------------+------------------------------
---+-----------------------------------+-----------------------------------+------
--------------------------+-----------------------------------+----------------
-----------------+-----------------------------------+-------------------------
--------+-----------------------------------+---------------------------------+--
--------------------------------+---------------------------------+--------------
-----------------+-----------------------------------+-------------------------
-------+-----------------------------------+---------------------------------+----
--------------------------+-------+--------------+
only showing top 20 rows
```

[9]:
```
df_spark = spark.read.option('header', 'true').
 ↪csv('Data_ColombiaGradientes_JFCO.csv',
                                                  header=True,␣
 ↪inferSchema=True)
```

[ ]:
```
```

[10]:
```
from IPython.core.display import HTML
display(HTML("<style>pre { white-space: pre !important; }</style>"))
```

```
<IPython.core.display.HTML object>
```

[11]:
```
df_spark.show(5)
```

```
+-------+--------+---------+----+-----+---+---+-------+----+----+----+---------
+--------+----+---------+-----------+-------------+-----------+-----------+---
----------------+-------------+--------------------+-----+---------+-----+-----
-------+-------------+-------------+-------------+-----------+------+-----+---
---------+--------------+---------------+------------------+---------+---+------
--------+------+---+-----------+-----------+---+---------+---------------+------
---+-----------+-------------+---------+---------+-----------+-----------+----
--------+------------+-----+-------+--------+------+-----------+------+-----+--
------+------------+-------+--------------------+----------+-------+---------+--
-----------+
|Entrada|Station |     Date|Year|month|Day| ID|Station| net|  N:|  W:|Elev.
(m)|altitude|time|Band_Code|Color_Anillo|        Order|      Family|      Genus|
Species|code _Species|         common_name|Recap|new_recap|
Mass|Culmen_Total|Culmen_Exposed|PW bill_width|PH bill_depth|Culmen_gapes|Tarsus
|Halux|Nail_1finger|feet_extension|Central_rectrix|External_Rectrices|Wing_cord|
P-S|age_Historical|   Sex|Rep|Pro_Cloacal|Brood_Patch|Fat|Condition|Pectoral_Mus
cle|Body_Molt|FLIGHT_MOLT|Fheather_wear|Molt_tail|Wear_tail|  molt_notes| Molt_
Limit|   Cycle_WRP|  Code_cycle|Skull|How_sex|How_aged|Status|
Iris_Color|Photos|Blood|Parasite|F-Chamber M.|Feather|
Notes_Bird|Num_Colect|replica|Anillador|Notes_General|
+-------+--------+---------+----+-----+---+---+-------+----+----+----+---------
+--------+----+---------+-----------+-------------+-----------+-----------+---
----------------+-------------+--------------------+-----+---------+-----+-----
-------+-------------+-------------+-------------+-----------+------+-----+---
---------+--------------+---------------+------------------+---------+---+------
--------+------+---+-----------+-----------+---+---------+---------------+------
---+-----------+-------------+---------+---------+-----------+-----------+----
--------+------------+-----+-------+--------+------+-----------+------+-----+--
------+------------+-------+--------------------+----------+-------+---------+--
-----------+
|    621|  Tatama|25-Feb-14|2014|  Feb| 25|  1|   TA01|null|null|null|     1400|
null|  NA|        x|           x|  Apodiformes|
Trochilidae|Phaethornis|Phaethornis syrma…|         TBHE|Tawny-bellied Hermit|
NA|     null| 6.32|          40|          39.9|          3.9|            3|
NA|    NA|   NA|          NA|            NA|             70|                NA|
64| NA|          ADULT|   UNK| NA|          NA|          NA| NA|     null|
2|        7|          S|   desgastada|        0|     null|
p1s|Check_photos|Check_photos|Check_photos|   NA|     NA|      NA|
NA|Check_photos|    no| null|      NA|        null|      Y|Muestras de pluma…|
null|   null|     null|         null|
|    622|  Tatama|25-Feb-14|2014|  Feb| 25|  2|   TA01|null|null|null|     1400|
```

```
null|  NA|     C616|           x|Passeriformes|  Tyrannidae|
Mionectes|Mionectes striati…|         STNF|Streak-Nekced Fly…|   NA|
null|   NA|         9.5|          8.2|          4.5|          3.7|          NA|
16.2|   NA|          NA|           NA|             48|               NA|
66| NA|           UNK|   UNK| NA|         NA|         NA|  M|     null|
3|        0|          0|   desgastada|        0|
null|Check_photos|Check_photos|Check_photos|Check_photos|   NA|     NA|      NA|
NA|Check_photos|    no| null|      NA|        null|      Y|Muestras 2pecho y…|
null|   null|     null|         null|
|    623|  Tatama|25-Feb-14|2014|  Feb| 25|  3|   TA01|null|null|null|     1400|
null|  NA|     A813|           x|Passeriformes|Fringillidae|   Euphonia|Euphonia
xanthoga…|         OBEU|Orange-bellied Eu…|   NA|     null|12.85|
8.7|          6.6|          7.8|          5.5|          NA|  16.1|   NA|
NA|           NA|             32|               NA|       59| NA|
ADULT|Female| NA|         NA|         NA|  A|     null|             4|
3|          0|   intermedio|        0|
null|Check_photos|Check_photos|Check_photos|Check_photos|   NA|     NA|      NA|
NA|Check_photos|    si| null|      NA|        null|      Y|Muestras de pluma…|
null|   null|     null|         null|
|    624|  Tatama|25-Feb-14|2014|  Feb| 25|  4|   TA01|null|null|null|     1400|
null|  NA|     B825|           x|Passeriformes|  Tyrannidae|  Zimmerius|
Zimmerius chrysops|         GFTY|Golden-faced Tyra…|   NA|     null| 8.43|
6.8|          5.2|          3.9|          3.4|          NA|  14.7|   NA|
NA|           NA|             40|               NA|       48| NA|
ADULT|   UNK| NA|         NA|         NA|  F|     null|             4|
Muchas|          0|   intermedio|        0|
null|Check_photos|Check_photos|Check_photos|Check_photos|   NA|     NA|      NA|
NA|Check_photos|    si| null|      NA|        null|      Y|Muestras de pluma…|
null|   null|     null|         null|
|    625|  Tatama|25-Feb-14|2014|  Feb| 25|  5|   TA01|null|null|null|     1400|
null|  NA|     A814|           x|Passeriformes|  Tyrannidae|  Mionectes|
Mionectes olivaceus|         OSTF|Olive-striped Fly…|   NA|     null|13.28|
13.1|          9.6|          5.3|          3.9|          NA|  16.1|   NA|
NA|           NA|             45|               NA|       60| NA|
INMATURE|   UNK| NA|         NA|         NA| NA|     null|             4|
0|          0|   desgastada|        0|
null|Check_photos|Check_photos|Check_photos|Check_photos|   NA|     NA|      NA|
NA|Check_photos|    si| null|       Y|        null|      Y|Muestras de pluma…|
null|   null|     null|         null|
+-------+--------+---------+----+-----+---+---+-------+----+----+----+---------
+--------+----+---------+------------+-------------+------------+-----------+---
----------------+-------------+--------------------+-----+---------+-----+-----
-------+--------------+-------------+-------------+------------+------+-----+---
---------+--------------+---------------+------------------+---------+---+------
--------+------+---+-----------+-----------+---+---------+---------------+------
---+-----------+-------------+---------+---------+------------+------------+----
--------+------------+-----+-------+--------+------+------------+------+-----+--
------+------------+-------+--------------------+----------+-------+---------+--
```

```
-----------+
only showing top 5 rows
```

[12]:
```
df_spark = df_spark.drop("N:", "W:")

df_spar = df_spark.join(dj_spark, ["Station"], "left")
#df_spar.printSchema()
df_spa = df_spar.drop("Localidad", "Elev. (m)", "net", "Notes_General",
                      "Anillador", "replica", "Num_Colect", "Notes_Bird",␣
 ↪"Blood", "Photos",
                      "Feather", "F-Chamber M.", "Parasite", "Iris_Color",␣
 ↪"Status", "How_aged",
                      "How_sex", "Skull", "Code_cycle", "Cycle_WRP", "Molt_␣
 ↪Limit", "molt_notes",
                      "Wear_tail", "Molt_tail", "Fheather_wear", "FLIGHT_MOLT",␣
 ↪"Body_Molt",
                       "Pectoral_Muscle", "Condition", "Fat", "Station ",␣
 ↪"P-S", "Sex", "age_Historical",
                     "Rep", "Pro_Cloacal", "Brood_Patch", "Pro_Cloacal",␣
 ↪"External_Rectrices",
                     "new_recap", "common_name",␣
 ↪"Color_Anillo","time","altitude", "Recap","Band_Code")
df_spa.show(20)
```

```
+-------+-------+---------+----+-----+---+---+----------+----------+----------
-+--------------------+------------+----+-----------+-------------+----------
---+-------------+------------+------+-----+-----------+--------------+--------
-------+---------+-----------------+-----------------+-----------------------
----------------------+-------------------------------------------------------+----
-----------------------------+-----------------------------------+--------------
------------------+-----------------------------------+-----------------------
---------+-----------------------------------+
+---------------------------------+---------------------------------+---------
------------------------+---------------------------------+-------------------
-------------+---------------------------------+-----------------------------
--+---------------------------------+---------------------------------+---------
------------------------+---------------------------------+--------------------
-----------+---------------------------------+--------------+
|Station|Entrada|     Date|Year|month|Day| ID|      Order|     Family|
Genus|             Species|code _Species|Mass|Culmen_Total|Culmen_Exposed|PW
bill_width|PH bill_depth|Culmen_gapes|Tarsus|Halux|Nail_1finger|feet_extension|C
entral_rectrix|Wing_cord|          Latitud|          Longitud|clipped_Hansen_GFC
-2022-v1.10_lossyear_10N_080W|clipped_Hansen_GFC-2022-
v1.10_treecover2000_10N_080W|resampled_clipped_wc2.1_30s_bio_1|resampled_clipped
_wc2.1_30s_bio_10|resampled_clipped_wc2.1_30s_bio_11|resampled_clipped_wc2.1_30s
_bio_12|resampled_clipped_wc2.1_30s_bio_13|resampled_clipped_wc2.1_30s_bio_14|re
```

```
sampled_clipped_wc2.1_30s_bio_15|resampled_clipped_wc2.1_30s_bio_16|resampled_cl
ipped_wc2.1_30s_bio_17|resampled_clipped_wc2.1_30s_bio_18|resampled_clipped_wc2.
1_30s_bio_19|resampled_clipped_wc2.1_30s_bio_2|resampled_clipped_wc2.1_30s_bio_3
|resampled_clipped_wc2.1_30s_bio_4|resampled_clipped_wc2.1_30s_bio_5|resampled_c
lipped_wc2.1_30s_bio_6|resampled_clipped_wc2.1_30s_bio_7|resampled_clipped_wc2.1
_30s_bio_8|resampled_clipped_wc2.1_30s_bio_9|resampled_clipped_wc2.1_30s_elev|St
ation_sample|
+-------+-------+---------+----+-----+---+---+----------+----------+---------
-+--------------------+-------------+----+-----------+-------------+---------
---+-------------+------------+------+-----+-----------+-------------+--------
-------+---------+-----------------+------------------+------------------------
----------------------+-------------------------------------------------------+----
-----------------------------+-----------------------------------+--------------
------------------+-----------------------------------+------------------------
---------+-----------------------------------+--------------------------------
+-----------------------------------+-----------------------------------+---------
-------------------------+-----------------------------------+------------------
-------------+-----------------------------------+------------------------------
--+-----------------------------------+-----------------------------------+---------
-----------------------+-----------------------------------+--------------------
-----------+-----------------------------------+--------------+
|   TA01|    621|25-Feb-14|2014|  Feb| 25|
1|Apodiformes|Trochilidae|Phaethornis|Phaethornis syrma…|         TBHE|6.32|
40|          39.9|          3.9|            3|          NA|    NA|   NA|
NA|            NA|             70|
64|5.223194444444444|-76.07793059868916|
0|                                                  95|
19.033333|                          19.383333|                              18.8|
2965.0|                            416.0|                             152.0|
34.28378|                              970.0|                               532.0|
875.0|                             970.0|                           7.816667|
90.89148|                         23.580935|                               23.3|
14.7|                          8.599999|                             18.8|
18.983334|                               1493|          S20|
|   TA01|    621|25-Feb-14|2014|  Feb| 25|
1|Apodiformes|Trochilidae|Phaethornis|Phaethornis syrma…|         TBHE|6.32|
40|          39.9|          3.9|            3|          NA|    NA|   NA|
NA|            NA|             70|
64|5.222503461438818|-76.07804003964577|
0|                                                  95|
19.033333|                          19.383333|                              18.8|
2965.0|                            416.0|                             152.0|
34.28378|                              970.0|                               532.0|
875.0|                             970.0|                           7.816667|
90.89148|                         23.580935|                               23.3|
14.7|                          8.599999|                             18.8|
18.983334|                               1493|          S19|
|   TA01|    621|25-Feb-14|2014|  Feb| 25|
```

```
1|Apodiformes|Trochilidae|Phaethornis|Phaethornis syrma…|         TBHE|6.32|
40|          39.9|          3.9|            3|          NA|    NA|   NA|
NA|            NA|             70|
64|5.221880116664146|-76.07835764967228|
0|                                                  95|
19.033333|                          19.383333|                              18.8|
2965.0|                            416.0|                             152.0|
34.28378|                            970.0|                              532.0|
875.0|                           970.0|                          7.816667|
90.89148|                        23.580935|                              23.3|
14.7|                         8.599999|                            18.8|
18.983334|                             1493|           S18|
|   TA01|    621|25-Feb-14|2014|  Feb| 25|
1|Apodiformes|Trochilidae|Phaethornis|Phaethornis syrma…|         TBHE|6.32|
40|          39.9|          3.9|            3|          NA|    NA|   NA|
NA|            NA|             70|
64|5.221385427450069|-76.07885233888635|
0|                                                  90|
19.033333|                          19.383333|                              18.8|
2965.0|                            416.0|                             152.0|
34.28378|                            970.0|                              532.0|
875.0|                           970.0|                          7.816667|
90.89148|                        23.580935|                              23.3|
14.7|                         8.599999|                            18.8|
18.983334|                             1493|           S17|
|   TA01|    621|25-Feb-14|2014|  Feb| 25|
1|Apodiformes|Trochilidae|Phaethornis|Phaethornis syrma…|         TBHE|6.32|
40|          39.9|          3.9|            3|          NA|    NA|   NA|
NA|            NA|             70|
64|5.221067817423564|-76.07947568366103|
0|                                                  90|
19.033333|                          19.383333|                              18.8|
2965.0|                            416.0|                             152.0|
34.28378|                            970.0|                              532.0|
875.0|                           970.0|                          7.816667|
90.89148|                        23.580935|                              23.3|
14.7|                         8.599999|                            18.8|
18.983334|                             1493|           S16|
|   TA01|    621|25-Feb-14|2014|  Feb| 25|
1|Apodiformes|Trochilidae|Phaethornis|Phaethornis syrma…|         TBHE|6.32|
40|          39.9|          3.9|            3|          NA|    NA|   NA|
NA|            NA|             70|
64|5.220958376466945|-76.08016666666666|
0|                                                  95|
19.033333|                          19.383333|                              18.8|
2965.0|                            416.0|                             152.0|
34.28378|                            970.0|                              532.0|
875.0|                           970.0|                          7.816667|
```

```
90.89148|                        23.580935|                              23.3|
14.7|                          8.599999|                                18.8|
18.983334|                              1493|           S15|
|   TA01|    621|25-Feb-14|2014|  Feb| 25|
1|Apodiformes|Trochilidae|Phaethornis|Phaethornis syrma…|         TBHE|6.32|
40|          39.9|          3.9|            3|          NA|    NA|   NA|
NA|            NA|             70|
64|5.221067817423564|-76.08085764967228|
0|                                                    90|
19.033333|                           19.383333|                              18.8|
2965.0|                             416.0|                             152.0|
34.28378|                              970.0|                              532.0|
875.0|                             970.0|                           7.816667|
90.89148|                        23.580935|                              23.3|
14.7|                          8.599999|                                18.8|
18.983334|                              1493|           S14|
|   TA01|    621|25-Feb-14|2014|  Feb| 25|
1|Apodiformes|Trochilidae|Phaethornis|Phaethornis syrma…|         TBHE|6.32|
40|          39.9|          3.9|            3|          NA|    NA|   NA|
NA|            NA|             70|
64|5.221385427450069|-76.08148099444696|
0|                                                   10|
19.033333|                           19.383333|                              18.8|
2965.0|                             416.0|                             152.0|
34.28378|                              970.0|                              532.0|
875.0|                             970.0|                           7.816667|
90.89148|                        23.580935|                              23.3|
14.7|                          8.599999|                                18.8|
18.983334|                              1493|           S13|
|   TA01|    621|25-Feb-14|2014|  Feb| 25|
1|Apodiformes|Trochilidae|Phaethornis|Phaethornis syrma…|         TBHE|6.32|
40|          39.9|          3.9|            3|          NA|    NA|   NA|
NA|            NA|             70|
64|5.221880116664146|-76.08197568366103|
0|                                                    0|
19.033333|                           19.383333|                              18.8|
2965.0|                             416.0|                             152.0|
34.28378|                              970.0|                              532.0|
875.0|                             970.0|                           7.816667|
90.89148|                        23.580935|                              23.3|
14.7|                          8.599999|                                18.8|
18.983334|                              1493|           S12|
|   TA01|    621|25-Feb-14|2014|  Feb| 25|
1|Apodiformes|Trochilidae|Phaethornis|Phaethornis syrma…|         TBHE|6.32|
40|          39.9|          3.9|            3|          NA|    NA|   NA|
NA|            NA|             70|
64|5.222503461438818|-76.08229329368754|
0|                                                   90|
```

```
19.033333|                          19.383333|                              18.8|
2965.0|                            416.0|                             152.0|
34.28378|                            970.0|                               532.0|
875.0|                             970.0|                           7.816667|
90.89148|                          23.580935|                               23.3|
14.7|                           8.599999|                             18.8|
18.983334|                               1493|           S11|
|   TA01|    621|25-Feb-14|2014|  Feb| 25|
1|Apodiformes|Trochilidae|Phaethornis|Phaethornis syrma…|         TBHE|6.32|
40|          39.9|          3.9|            3|          NA|    NA|   NA|
NA|            NA|             70|
64|5.223194444444444|-76.08240273464415|
0|                                                  90|
19.033333|                          19.383333|                              18.8|
2965.0|                            416.0|                             152.0|
34.28378|                            970.0|                               532.0|
875.0|                             970.0|                           7.816667|
90.89148|                          23.580935|                               23.3|
14.7|                           8.599999|                             18.8|
18.983334|                               1493|           S10|
|   TA01|    621|25-Feb-14|2014|  Feb| 25|
1|Apodiformes|Trochilidae|Phaethornis|Phaethornis syrma…|         TBHE|6.32|
40|          39.9|          3.9|            3|          NA|    NA|   NA|
NA|            NA|             70|       64|
5.22388542745007|-76.08229329368754|
0|                                                   6|
19.033333|                          19.383333|                              18.8|
2965.0|                            416.0|                             152.0|
34.28378|                            970.0|                               532.0|
875.0|                             970.0|                           7.816667|
90.89148|                          23.580935|                               23.3|
14.7|                           8.599999|                             18.8|
18.983334|                               1493|            S9|
|   TA01|    621|25-Feb-14|2014|  Feb| 25|
1|Apodiformes|Trochilidae|Phaethornis|Phaethornis syrma…|         TBHE|6.32|
40|          39.9|          3.9|            3|          NA|    NA|   NA|
NA|            NA|             70|
64|5.224508772224742|-76.08197568366103|
0|                                                  37|
19.033333|                          19.383333|                              18.8|
2965.0|                            416.0|                             152.0|
34.28378|                            970.0|                               532.0|
875.0|                             970.0|                           7.816667|
90.89148|                          23.580935|                               23.3|
14.7|                           8.599999|                             18.8|
18.983334|                               1493|            S8|
|   TA01|    621|25-Feb-14|2014|  Feb| 25|
1|Apodiformes|Trochilidae|Phaethornis|Phaethornis syrma…|         TBHE|6.32|
```

```
40|          39.9|          3.9|            3|          NA|    NA|   NA|
NA|            NA|             70|       64|
5.22500346143882|-76.08148099444696|
0|                                                    90|
19.033333|                          19.383333|                              18.8|
2965.0|                           416.0|                            152.0|
34.28378|                            970.0|                             532.0|
875.0|                           970.0|                          7.816667|
90.89148|                         23.580935|                              23.3|
14.7|                        8.599999|                            18.8|
18.983334|                             1493|            S7|
|   TA01|    621|25-Feb-14|2014|  Feb| 25|
1|Apodiformes|Trochilidae|Phaethornis|Phaethornis syrma…|         TBHE|6.32|
40|          39.9|          3.9|            3|          NA|    NA|   NA|
NA|            NA|             70|
64|5.225321071465324|-76.08085764967228|
0|                                                    95|
19.620832|                          19.966667|                           19.366667|
2351.0|                           306.0|                             88.0|
37.38887|                            784.0|                             344.0|
720.0|                           725.0|                          8.224999|
91.388885|                         24.069817|                              24.1|
15.1|                           9.0|                         19.816668|
19.766666|                             1420|            S6|
|   TA01|    621|25-Feb-14|2014|  Feb| 25|
1|Apodiformes|Trochilidae|Phaethornis|Phaethornis syrma…|         TBHE|6.32|
40|          39.9|          3.9|            3|          NA|    NA|   NA|
NA|            NA|             70|
64|5.225430512421943|-76.08016666666666|
0|                                                    95|
19.620832|                          19.966667|                           19.366667|
2351.0|                           306.0|                             88.0|
37.38887|                            784.0|                             344.0|
720.0|                           725.0|                          8.224999|
91.388885|                         24.069817|                              24.1|
15.1|                           9.0|                         19.816668|
19.766666|                             1420|            S5|
|   TA01|    621|25-Feb-14|2014|  Feb| 25|
1|Apodiformes|Trochilidae|Phaethornis|Phaethornis syrma…|         TBHE|6.32|
40|          39.9|          3.9|            3|          NA|    NA|   NA|
NA|            NA|             70|
64|5.225321071465324|-76.07947568366103|
0|                                                    95|
19.620832|                          19.966667|                           19.366667|
2351.0|                           306.0|                             88.0|
37.38887|                            784.0|                             344.0|
720.0|                           725.0|                          8.224999|
91.388885|                         24.069817|                              24.1|
```

```
15.1|                              9.0|                         19.816668|
19.766666|                            1420|            S4|
|   TA01|    621|25-Feb-14|2014|  Feb| 25|
1|Apodiformes|Trochilidae|Phaethornis|Phaethornis syrma…|         TBHE|6.32|
40|          39.9|          3.9|            3|          NA|    NA|   NA|
NA|            NA|             70|       64|
5.22500346143882|-76.07885233888635|
0|                                                  95|
19.033333|                          19.383333|                               18.8|
2965.0|                             416.0|                            152.0|
34.28378|                             970.0|                             532.0|
875.0|                             970.0|                          7.816667|
90.89148|                         23.580935|                               23.3|
14.7|                         8.599999|                             18.8|
18.983334|                             1493|            S3|
|   TA01|    621|25-Feb-14|2014|  Feb| 25|
1|Apodiformes|Trochilidae|Phaethornis|Phaethornis syrma…|         TBHE|6.32|
40|          39.9|          3.9|            3|          NA|    NA|   NA|
NA|            NA|             70|
64|5.224508772224742|-76.07835764967228|
0|                                                  95|
19.033333|                          19.383333|                               18.8|
2965.0|                             416.0|                            152.0|
34.28378|                             970.0|                             532.0|
875.0|                             970.0|                          7.816667|
90.89148|                         23.580935|                               23.3|
14.7|                         8.599999|                             18.8|
18.983334|                             1493|            S2|
|   TA01|    621|25-Feb-14|2014|  Feb| 25|
1|Apodiformes|Trochilidae|Phaethornis|Phaethornis syrma…|         TBHE|6.32|
40|          39.9|          3.9|            3|          NA|    NA|   NA|
NA|            NA|             70|       64|
5.22388542745007|-76.07804003964577|
0|                                                  95|
19.033333|                          19.383333|                               18.8|
2965.0|                             416.0|                            152.0|
34.28378|                             970.0|                             532.0|
875.0|                             970.0|                          7.816667|
90.89148|                         23.580935|                               23.3|
14.7|                         8.599999|                             18.8|
18.983334|                             1493|            S1|
+-------+-------+---------+----+-----+---+---+-----------+-----------+----------
-+--------------------+-------------+----+------------+--------------+----------
---+-------------+------------+------+-----+------------+--------------+--------
-------+---------+-----------------+------------------+------------------------
----------------------+-----------------------------------------------------+----
----------------------------+-----------------------------------+--------------
------------------+-----------------------------------+------------------------
```

```
---------+-----------------------------------+-----------------------------------
+-----------------------------------+-----------------------------------+---------
------------------------+-----------------------------------+------------------
-------------+-----------------------------------+-------------------------------
--+-----------------------------------+-----------------------------------+---------
------------------------+-----------------------------------+--------------------
------------+-----------------------------------+--------------+
only showing top 20 rows
```

[ ]:

[13]:
```
df_spa = df_spa \
    .withColumnRenamed("clipped_Hansen_GFC-2022-v1.10_lossyear_10N_080W",
                       "lossyear") \
    .withColumnRenamed("clipped_Hansen_GFC-2022-v1.10_treecover2000_10N_080W",
                       "treecover") \
    .withColumnRenamed("resampled_clipped_wc2.1_30s_bio_1", "bio_1") \
    .withColumnRenamed("resampled_clipped_wc2.1_30s_bio_2", "bio_2")  \
    .withColumnRenamed("resampled_clipped_wc2.1_30s_bio_3", "bio_3")  \
    .withColumnRenamed("resampled_clipped_wc2.1_30s_bio_4", "bio_4")  \
    .withColumnRenamed("resampled_clipped_wc2.1_30s_bio_5", "bio_5")  \
    .withColumnRenamed("resampled_clipped_wc2.1_30s_bio_6", "bio_6")  \
    .withColumnRenamed("resampled_clipped_wc2.1_30s_bio_7", "bio_7")  \
    .withColumnRenamed("resampled_clipped_wc2.1_30s_bio_8", "bio_8")  \
    .withColumnRenamed("resampled_clipped_wc2.1_30s_bio_9", "bio_9")  \
    .withColumnRenamed("resampled_clipped_wc2.1_30s_bio_10", "bio_10") \
    .withColumnRenamed("resampled_clipped_wc2.1_30s_bio_11", "bio_11")  \
    .withColumnRenamed("resampled_clipped_wc2.1_30s_bio_12", "bio_12")  \
    .withColumnRenamed("resampled_clipped_wc2.1_30s_bio_13", "bio_13")  \
    .withColumnRenamed("resampled_clipped_wc2.1_30s_bio_14", "bio_14")  \
    .withColumnRenamed("resampled_clipped_wc2.1_30s_bio_15", "bio_15")  \
    .withColumnRenamed("resampled_clipped_wc2.1_30s_bio_16", "bio_16")  \
    .withColumnRenamed("resampled_clipped_wc2.1_30s_bio_17", "bio_17")  \
    .withColumnRenamed("resampled_clipped_wc2.1_30s_bio_18", "bio_18")  \
    .withColumnRenamed("resampled_clipped_wc2.1_30s_bio_19", "bio_19")  \
    .withColumnRenamed("resampled_clipped_wc2.1_30s_elev", "elev")

# Mostrar el esquema del DataFrame modificado para verificar los cambios
df_spa.printSchema()
```

```
root
 |-- Station: string (nullable = true)
 |-- Entrada: integer (nullable = true)
 |-- Date: string (nullable = true)
 |-- Year: integer (nullable = true)
 |-- month: string (nullable = true)
 |-- Day: integer (nullable = true)
```

```
 |-- ID: string (nullable = true)
 |-- Order: string (nullable = true)
 |-- Family: string (nullable = true)
 |-- Genus: string (nullable = true)
 |-- Species: string (nullable = true)
 |-- code _Species: string (nullable = true)
 |-- Mass: string (nullable = true)
 |-- Culmen_Total: string (nullable = true)
 |-- Culmen_Exposed: string (nullable = true)
 |-- PW bill_width: string (nullable = true)
 |-- PH bill_depth: string (nullable = true)
 |-- Culmen_gapes: string (nullable = true)
 |-- Tarsus: string (nullable = true)
 |-- Halux: string (nullable = true)
 |-- Nail_1finger: string (nullable = true)
 |-- feet_extension: string (nullable = true)
 |-- Central_rectrix: string (nullable = true)
 |-- Wing_cord: string (nullable = true)
 |-- Latitud: double (nullable = true)
 |-- Longitud: double (nullable = true)
 |-- lossyear: integer (nullable = true)
 |-- treecover: integer (nullable = true)
 |-- bio_1: double (nullable = true)
 |-- bio_10: double (nullable = true)
 |-- bio_11: double (nullable = true)
 |-- bio_12: double (nullable = true)
 |-- bio_13: double (nullable = true)
 |-- bio_14: double (nullable = true)
 |-- bio_15: double (nullable = true)
 |-- bio_16: double (nullable = true)
 |-- bio_17: double (nullable = true)
 |-- bio_18: double (nullable = true)
 |-- bio_19: double (nullable = true)
 |-- bio_2: double (nullable = true)
 |-- bio_3: double (nullable = true)
 |-- bio_4: double (nullable = true)
 |-- bio_5: double (nullable = true)
 |-- bio_6: double (nullable = true)
 |-- bio_7: double (nullable = true)
 |-- bio_8: double (nullable = true)
 |-- bio_9: double (nullable = true)
 |-- elev: integer (nullable = true)
 |-- Station_sample: string (nullable = true)
```

[14]:
```
df_spa.show(2)
```

```
+-------+-------+---------+----+-----+---+---+-----------+-----------+----------
```

```
-+-------------------+------------+----+-----------+-------------+---------
---+-------------+-----------+------+-----+-----------+-------------+--------
-------+---------+----------------+-----------------+--------+---------+------
---+---------+------+------+------+------+--------+------+------+------+------+-
-------+--------+---------+-----+-----+--------+-----+---------+----+----------
---+
|Station|Entrada|     Date|Year|month|Day| ID|      Order|     Family|
Genus|             Species|code _Species|Mass|Culmen_Total|Culmen_Exposed|PW
bill_width|PH bill_depth|Culmen_gapes|Tarsus|Halux|Nail_1finger|feet_extension|C
entral_rectrix|Wing_cord|          Latitud|
Longitud|lossyear|treecover|    bio_1|   bio_10|bio_11|bio_12|bio_13|bio_14|
bio_15|bio_16|bio_17|bio_18|bio_19|   bio_2|   bio_3|    bio_4|bio_5|bio_6|
bio_7|bio_8|    bio_9|elev|Station_sample|
+-------+-------+---------+----+-----+---+---+-----------+-----------+----------
-+-------------------+------------+----+-----------+-------------+---------
---+-------------+-----------+------+-----+-----------+-------------+--------
-------+---------+----------------+-----------------+--------+---------+------
---+---------+------+------+------+------+--------+------+------+------+------+-
-------+--------+---------+-----+-----+--------+-----+---------+----+----------
---+
|   TA01|    621|25-Feb-14|2014|  Feb| 25|
1|Apodiformes|Trochilidae|Phaethornis|Phaethornis syrma…|         TBHE|6.32|
40|          39.9|          3.9|            3|          NA|    NA|   NA|
NA|            NA|             70|
64|5.223194444444444|-76.07793059868916|       0|       95|19.033333|19.383333|
18.8|2965.0| 416.0| 152.0|34.28378| 970.0| 532.0| 875.0|
970.0|7.816667|90.89148|23.580935| 23.3| 14.7|8.599999| 18.8|18.983334|1493|
S20|
|   TA01|    621|25-Feb-14|2014|  Feb| 25|
1|Apodiformes|Trochilidae|Phaethornis|Phaethornis syrma…|         TBHE|6.32|
40|          39.9|          3.9|            3|          NA|    NA|   NA|
NA|            NA|             70|
64|5.222503461438818|-76.07804003964577|       0|       95|19.033333|19.383333|
18.8|2965.0| 416.0| 152.0|34.28378| 970.0| 532.0| 875.0|
970.0|7.816667|90.89148|23.580935| 23.3| 14.7|8.599999| 18.8|18.983334|1493|
S19|
+-------+-------+---------+----+-----+---+---+-----------+-----------+----------
-+-------------------+------------+----+-----------+-------------+---------
---+-------------+-----------+------+-----+-----------+-------------+--------
-------+---------+----------------+-----------------+--------+---------+------
---+---------+------+------+------+------+--------+------+------+------+------+-
-------+--------+---------+-----+-----+--------+-----+---------+----+----------
---+
only showing top 2 rows
```

[15]:
```
from pyspark.sql.types import StringType
for column in df_spa.columns:
    if isinstance(df_spa.schema[column].dataType, StringType):
        df_spa = df_spa.withColumn(column, when(col(column) == "NA", None).
 ↪otherwise(col(column)))
```

[16]:
```
# Contar los valores NA en la columna Latitud
na_count = df_spa.agg(
    count(when(col("elev").isNull(), True))
).collect()[0][0]

print("valores NA en --:", na_count)
```

```
valores NA en --: 0
```

[ ]:
```

```

[17]:
```
na_counts = df_spa.select([count(when(col(c).isNull(),
                                      c)).alias(c) for c in df_spa.columns]).
 ↪collect()[0].asDict()

# Imprime el conteo de valores NA por columna
for column, na_count in na_counts.items():
    print(f"Número de valores NA en la columna {column}: {na_count}")
```

```
Número de valores NA en la columna Station: 0
Número de valores NA en la columna Entrada: 0
Número de valores NA en la columna Date: 0
Número de valores NA en la columna Year: 0
Número de valores NA en la columna month: 0
Número de valores NA en la columna Day: 0
Número de valores NA en la columna ID: 0
Número de valores NA en la columna Order: 21
Número de valores NA en la columna Family: 84
Número de valores NA en la columna Genus: 63
Número de valores NA en la columna Species: 0
Número de valores NA en la columna code _Species: 2793
Número de valores NA en la columna Mass: 10164
Número de valores NA en la columna Culmen_Total: 10038
Número de valores NA en la columna Culmen_Exposed: 36267
Número de valores NA en la columna PW bill_width: 10710
Número de valores NA en la columna PH bill_depth: 35889
Número de valores NA en la columna Culmen_gapes: 49476
Número de valores NA en la columna Tarsus: 21987
Número de valores NA en la columna Halux: 49539
Número de valores NA en la columna Nail_1finger: 49665
Número de valores NA en la columna feet_extension: 49854
```

```
Número de valores NA en la columna Central_rectrix: 13125
Número de valores NA en la columna Wing_cord: 10563
Número de valores NA en la columna Latitud: 0
Número de valores NA en la columna Longitud: 0
Número de valores NA en la columna lossyear: 0
Número de valores NA en la columna treecover: 0
Número de valores NA en la columna bio_1: 0
Número de valores NA en la columna bio_10: 0
Número de valores NA en la columna bio_11: 0
Número de valores NA en la columna bio_12: 0
Número de valores NA en la columna bio_13: 0
Número de valores NA en la columna bio_14: 0
Número de valores NA en la columna bio_15: 0
Número de valores NA en la columna bio_16: 0
Número de valores NA en la columna bio_17: 0
Número de valores NA en la columna bio_18: 0
Número de valores NA en la columna bio_19: 0
Número de valores NA en la columna bio_2: 0
Número de valores NA en la columna bio_3: 0
Número de valores NA en la columna bio_4: 0
Número de valores NA en la columna bio_5: 0
Número de valores NA en la columna bio_6: 0
Número de valores NA en la columna bio_7: 0
Número de valores NA en la columna bio_8: 0
Número de valores NA en la columna bio_9: 0
Número de valores NA en la columna elev: 0
Número de valores NA en la columna Station_sample: 0
```

[18]:
```
#### Posiblemente algunas especies esten mal escritas o duplicadas por error en␣
 ↪typing then...
#unique_species = df_spa.select("Species").distinct().collect()
#print(unique_species)
```

[19]:
```
#unique_species = [row['Species'] for row in unique_species]
```

[20]:
```
#import csv
#with open('unique_speciesk.csv', 'w', newline='', encoding='utf-8') as file:
#    writer = csv.writer(file)
#    writer.writerow(['Species'])  # escribir el encabezado
#    for species in unique_species:
#        writer.writerow([species])
```

[18]:
```
#############################################################################################
specie_colum = spark.read.option('header', 'true').csv('updated_species.csv',
                                                       header=True,␣
 ↪inferSchema=True)
```

[19]:
```
df_spa_updated = df_spa.join(specie_colum, df_spa["Species"] ==␣
 ↪specie_colum["Species1"], "left")
df_spa_updated = df_spa_updated.withColumn("Species",␣
 ↪coalesce(col("Species_new"), col("Species")))
df_sp = df_spa_updated.drop("Species_new", "Species1")
df_sp = df_sp.filter(~(col("Species").isin(["Buscar", "buscar"])))
```

[20]:
```
df_sp.show(2)
```

```
+-------+-------+---------+----+-----+---+---+----------+----------+----------
-+--------------------+-------------+----+------------+--------------+----------
---+-------------+------------+------+-----+------------+--------------+--------
-------+---------+-----------------+------------------+--------+---------+------
---+---------+------+------+------+------+--------+------+------+------+------+-
-------+--------+---------+-----+-----+--------+-----+---------+----+-----------
---+
|Station|Entrada|     Date|Year|month|Day| ID|      Order|     Family|
Genus|             Species|code _Species|Mass|Culmen_Total|Culmen_Exposed|PW
bill_width|PH bill_depth|Culmen_gapes|Tarsus|Halux|Nail_1finger|feet_extension|C
entral_rectrix|Wing_cord|          Latitud|
Longitud|lossyear|treecover|    bio_1|   bio_10|bio_11|bio_12|bio_13|bio_14|
bio_15|bio_16|bio_17|bio_18|bio_19|   bio_2|   bio_3|    bio_4|bio_5|bio_6|
bio_7|bio_8|    bio_9|elev|Station_sample|
+-------+-------+---------+----+-----+---+---+----------+----------+----------
-+--------------------+-------------+----+------------+--------------+----------
---+-------------+------------+------+-----+------------+--------------+--------
-------+---------+-----------------+------------------+--------+---------+------
---+---------+------+------+------+------+--------+------+------+------+------+-
-------+--------+---------+-----+-----+--------+-----+---------+----+-----------
---+
|   TA01|    621|25-Feb-14|2014|  Feb| 25|
1|Apodiformes|Trochilidae|Phaethornis|Phaethornis syrma…|         TBHE|6.32|
40|          39.9|          3.9|            3|        null|  null| null|
null|          null|             70|
64|5.223194444444444|-76.07793059868916|       0|       95|19.033333|19.383333|
18.8|2965.0| 416.0| 152.0|34.28378| 970.0| 532.0| 875.0|
970.0|7.816667|90.89148|23.580935| 23.3| 14.7|8.599999| 18.8|18.983334|1493|
S20|
|   TA01|    621|25-Feb-14|2014|  Feb| 25|
1|Apodiformes|Trochilidae|Phaethornis|Phaethornis syrma…|         TBHE|6.32|
40|          39.9|          3.9|            3|        null|  null| null|
null|          null|             70|
64|5.222503461438818|-76.07804003964577|       0|       95|19.033333|19.383333|
18.8|2965.0| 416.0| 152.0|34.28378| 970.0| 532.0| 875.0|
970.0|7.816667|90.89148|23.580935| 23.3| 14.7|8.599999| 18.8|18.983334|1493|
S19|
+-------+-------+---------+----+-----+---+---+----------+----------+----------
```

```
-+--------------------+------------+----+-----------+-------------+---------
---+------------+------------+------+-----+-----------+-------------+--------
-------+---------+-----------------+-----------------+--------+---------+------
---+---------+------+------+------+------+--------+------+------+------+------+-
-------+--------+---------+-----+-----+--------+-----+---------+----+----------
---+
only showing top 2 rows
```

[21]:
```
## Sera mejor eliminar los espacios y solo usar _ para genero y epiteto␣
 ↪especifico
df_sp = df_sp.withColumn("Species", regexp_replace(col("Species"), " ", "_"))
#df_sp.show(3)
```

[22]:
```
#### Solo para rectificar que los cambios estan bien y ya nos vamos con estas␣
 ↪especies...
unique_species = df_sp.select("Species").distinct().collect()
print(unique_species)
```

```
[Row(Species='Phylloscartes_ophthalmicus'), Row(Species='Habia_cristata'),
Row(Species='Picumnus_cinnamomeus'), Row(Species='Myrmotherulia_schisticolor'),
Row(Species='Haplophaedia_aureliae'), Row(Species='Tangara_labradorides'),
Row(Species='Creurgops_verticalis'), Row(Species='Troglodytes_aedon'),
Row(Species='Myadestes_ralloides'), Row(Species='Thripadectes_holostictus'),
Row(Species='Aulacorhynchus_prasinus'), Row(Species='Catharus_minimus'),
Row(Species='Clibanornis_rubiginosus'), Row(Species='Xenops_MInutus'),
Row(Species='Myrmotherula_schiticolor'), Row(Species='Veniliornis_dignus'),
Row(Species='Aglaiocercus_coelestis'), Row(Species='Oreothraupis_arremonops'),
Row(Species='Thripadectes_ignobilis'), Row(Species='Chlorospingus_semifuscus'),
Row(Species='Psarocolius_wagleri'), Row(Species='Myiothlypis_fulvicauda'),
Row(Species='Xiphorhynchus_triangularis'), Row(Species='Arremon_castaneiceps'),
Row(Species='Platyrinchus_coronatus'), Row(Species='Catamblyrhynchus_diadema'),
Row(Species='Formicarius_rufipectus'), Row(Species='Momotus_aequatorialis'),
Row(Species='Veniliornis_affinis'), Row(Species='Xenops_rutilans'),
Row(Species='Philydor_rufum'), Row(Species='Margarornis_stellatus'),
Row(Species='Pseudocolaptes_lawrencii'), Row(Species='Eubucco_bourcierii'),
Row(Species='Cyphorhinus_thoracicus'), Row(Species='Lophotriccus_pileatus'),
Row(Species='Coeligena_coeligena'), Row(Species='Sporophila_luctuosa'),
Row(Species='Dysithamnus_occidentalis'), Row(Species='Synallaxis_brachyura'),
Row(Species='Eriocnemis_vestita'), Row(Species='Margarornis_squamiger'),
Row(Species='Piranga_rubra'), Row(Species='Rupicola_peruvianus'),
Row(Species='Thripadectes_flammulatus'),
Row(Species='Basileuterus_tristriatus'),
Row(Species='Cantorchilus_nigricapillus'), Row(Species='Heliangelus_exortis'),
Row(Species='Thraupis_cyanocephala'), Row(Species='Troglodytes_solstitialis'),
Row(Species='Phaetornis_syrmatophorus'),
Row(Species='Microbates_cinereiventris'), Row(Species='Phaethornis_guy'),
Row(Species='Drymotoxeres_pucheranii'), Row(Species='Heliodoxa_imperatrix'),
```

```
Row(Species='Sporophila_sp.'), Row(Species='Synallaxis_unirufa'),
Row(Species='Mionectes_striaticollis'), Row(Species='Cinnycerthia_unirufa'),
Row(Species='Pachyramphus_versicolor'), Row(Species='Phyllomyias_cinereiceps'),
Row(Species='Malacoptila_mystacalis'), Row(Species='Atlapetes_albinucha'),
Row(Species='Ramphocelus_flammigerus'), Row(Species='Zentrygon_frenata'),
Row(Species='Xiphorhynchus_erythropygius'), Row(Species='Scytalopus_vicinior'),
Row(Species='Myiarchus_tuberculifer'), Row(Species='Atlapetes_tricolor'),
Row(Species='Stilpnia_cyanicollis'), Row(Species='Tangara_arthus'),
Row(Species='Doryfera_ludovicae'), Row(Species='Cyanolyca_pulchra'),
Row(Species='Dromococcyx_pavoninus'), Row(Species='Cercomacroides_parkeri'),
Row(Species='Grallaricula_flavirostris'), Row(Species='Ochthoeca_diadema'),
Row(Species='Haplospiza_rustica'), Row(Species='Henicorhina_leucosticta'),
Row(Species='Myiotriccus_ornatus'), Row(Species='Trogon_collaris'),
Row(Species='Campephilus_haematogaster'),
Row(Species='Onychorhynchus_coronatus'),
Row(Species='Chlorothraupis_stolzmanni'),
Row(Species='Myiothlypis_chrysogaster'),
Row(Species='Chlorostilbon_mellisugus'), Row(Species='Thamnophilus_unicolor'),
Row(Species='Chlorospingus_canigularis'), Row(Species='Myioborus_ornatus'),
Row(Species='Grallaria_flavotincta'), Row(Species='Arremon_aurantiirostris'),
Row(Species='Rhynchocyclus_brevirostris'), Row(Species='Uropsalis_lyra'),
Row(Species='Tangara_icterocephala'), Row(Species='Boissonneaua_jardini'),
Row(Species='Nephelomyias\xa0pulcher'),
Row(Species='Lepidocolaptes_lacrymiger'),
Row(Species='Hemitriccus_granadensis'),
Row(Species='Phyllomyias_nigrocapillus'), Row(Species='Diglossa_caerulescens'),
Row(Species='Automolus_ochrolaemus'), Row(Species='Premnoplex_brunnescens'),
Row(Species='Phaethornis_longirostris'), Row(Species='Saltator_atripennis'),
Row(Species='Leptopogon_superciliaris'), Row(Species='Amazilia_franciae'),
Row(Species='Zimmerius_chrysops'), Row(Species='Synallaxis_azarae'),
Row(Species='Urochroa_bougueri'), Row(Species='Anabacerthia_variegaticeps'),
Row(Species='Anisognathus_somptuosus'), Row(Species='Pyrrhomyias_cinnamomeus'),
Row(Species='Pseudocolaptes_boissonneautii'), Row(Species='Serpophaga_cinerea'),
Row(Species='Myiodynastes_chrysocephalus'), Row(Species='Nephelomyias_pulcher'),
Row(Species='Glaucidium_jardinii'), Row(Species='Zonotrichia_capensis'),
Row(Species='Cyclarhis_nigrirostris'), Row(Species='Coeligena_wilsoni'),
Row(Species='Micromonacha_lanceolata'),
Row(Species='Aulacorhynchus_haematopygus'),
Row(Species='Pseudotriccus_ruficeps'), Row(Species='Pipreola_jucunda'),
Row(Species='Dendrocincla_tyrannina'), Row(Species='Chlorornis_riefferii'),
Row(Species='Myrmotherula_schisticolor'), Row(Species='Syndactyla_subalaris'),
Row(Species='Campylorhamphus_pusillus'), Row(Species='Anisognathus_notabilis'),
Row(Species='Phaethornis_syrmatophorus'), Row(Species='Turdus_leucops'),
Row(Species='Microcerculus_marginatus'), Row(Species='Diglossa_gloriosissima'),
Row(Species='Lafresnaya_lafresnayi'), Row(Species='Diglossa_albilatera'),
Row(Species='Mitrospingus_cassinii'), Row(Species='Sporophila_nigricollis'),
Row(Species='Euphonia_xanthogaster'), Row(Species='Henicorhina_leucophrys'),
Row(Species='Mionectes_olivaceus'), Row(Species='Pipreola_riefferii'),
```

```
Row(Species='Ochthoeca_cinnamomeiventris'), Row(Species='Scytalopus_sp.'),
Row(Species='Diglossa_cyanea'), Row(Species='Anisognathus_lacrymosus'),
Row(Species='Arremon_atricapillus'), Row(Species='Tangara_xanthocephala'),
Row(Species='Poecilotriccus_ruficeps'),
Row(Species='Rhynchocyclus_fulvipectus'), Row(Species='Ocreatus_underwoodii'),
Row(Species='Trogon_personatus'), Row(Species='Malacoptila_panamensis'),
Row(Species='Pseudotriccus_pelzelni'), Row(Species='Henicorhina_negreti'),
Row(Species='Piaya_cayana'), Row(Species='Phaethornis_striigularis'),
Row(Species='Machaeropterus_striolatus'),
Row(Species='Anthracothorax_nigricollis'), Row(Species='Thalurania_colombica'),
Row(Species='Stilpnia_heinei'), Row(Species='Turdus_serranus'),
Row(Species='Manacus_manacus'), Row(Species='Myioborus_miniatus'),
Row(Species='Hafferia_zeledoni'), Row(Species='Ramphomicron_microrhynchum'),
Row(Species='Dendrocincla_fuliginosa'),
Row(Species='Chlorospingus_flavigularis'), Row(Species='Schistes_albogularis'),
Row(Species='Islerothraupis_luctuosa'), Row(Species='Aglaiocercus_kingii'),
Row(Species='Masius_chrysopterus'), Row(Species='Myiobius_villosus'),
Row(Species='Atlapetes_latinuchus'), Row(Species='Dysithamnus_mentalis'),
Row(Species='Sayornis_nigricans'), Row(Species='Machaeropterus_deliciosus'),
Row(Species='Bangsia_aureocincta'), Row(Species='Cyanoloxia_cyanoides'),
Row(Species='Chlorochrysa_phoenicotis'), Row(Species='Nothocercus_bonapartei'),
Row(Species='Turdus_flavipes'), Row(Species='Chlorophonia_pyrrhophrys'),
Row(Species='Bangsia_melanochlamys'), Row(Species='Sporophila_funerea'),
Row(Species='Phaethornis_yaruqui'), Row(Species='Iridosornis_porphyrocephalus'),
Row(Species='Xenops_minutus'), Row(Species='Pygochelidon_cyanoleuca'),
Row(Species='Sclerurus_mexicanus'), Row(Species='Diglossa_indigotica'),
Row(Species='Pipreola_arcuata'), Row(Species='Setophaga_fusca'),
Row(Species='Tangara_gyrola'), Row(Species='Colaptes_rubiginosus'),
Row(Species='Premnornis_guttuliger'), Row(Species='Chaetocercus_mulsant'),
Row(Species='Eutoxeres_aquila'), Row(Species='Turdus_ignobilis'),
Row(Species='Entomodestes_coracinus'), Row(Species='Hafferia_immaculata'),
Row(Species='Heliodoxa_jacula'), Row(Species='Turdus_fuscater'),
Row(Species='Drymophila_striaticeps'), Row(Species='Tachyphonus_luctuosus'),
Row(Species='Kleinothraupis_atropileus'),
Row(Species='Sporophila_crassirostris'), Row(Species='Pheugopedius_spadix'),
Row(Species='Boissonneaua_flavescens'), Row(Species='Ixothraupis_rufigula'),
Row(Species='Platyrinchus_mystaceus'), Row(Species='Arremon_brunneinucha'),
Row(Species='Myiophobus_flavicans'), Row(Species='Semnornis_ramphastinus'),
Row(Species='Cinnycerthia_olivascens'), Row(Species='Thripadectes_virgaticeps'),
Row(Species='Snowornis_cryptolophus'), Row(Species='Coeligena_torquata'),
Row(Species='Empidonax_sp.'), Row(Species='Glaucidium_nubicola'),
Row(Species='Micrastur_ruficollis'), Row(Species='Dryobates_fumigatus'),
Row(Species='Saltator_maximus'), Row(Species='Catharus_ustulatus'),
Row(Species='Conopophaga_castaneiceps'), Row(Species='Tangara_nigroviridis'),
Row(Species='Scytalopus_spillmanni'), Row(Species='Glyphorynchus_spirurus')]
```

[23]:
```
df_sp.show(2)
```

```
+-------+-------+---------+----+-----+---+---+----------+----------+----------
-+--------------------+-------------+----+------------+-------------+----------
---+-------------+------------+------+-----+------------+-------------+--------
-------+---------+-----------------+------------------+--------+---------+------
---+---------+------+------+------+------+--------+------+------+------+------+-
-------+--------+---------+-----+-----+--------+-----+---------+----+-----------
---+
|Station|Entrada|     Date|Year|month|Day| ID|      Order|     Family|
Genus|             Species|code _Species|Mass|Culmen_Total|Culmen_Exposed|PW
bill_width|PH bill_depth|Culmen_gapes|Tarsus|Halux|Nail_1finger|feet_extension|C
entral_rectrix|Wing_cord|          Latitud|
Longitud|lossyear|treecover|    bio_1|   bio_10|bio_11|bio_12|bio_13|bio_14|
bio_15|bio_16|bio_17|bio_18|bio_19|   bio_2|   bio_3|    bio_4|bio_5|bio_6|
bio_7|bio_8|    bio_9|elev|Station_sample|
+-------+-------+---------+----+-----+---+---+----------+----------+----------
-+--------------------+-------------+----+------------+-------------+----------
---+-------------+------------+------+-----+------------+-------------+--------
-------+---------+-----------------+------------------+--------+---------+------
---+---------+------+------+------+------+--------+------+------+------+------+-
-------+--------+---------+-----+-----+--------+-----+---------+----+-----------
---+
|   TA01|    621|25-Feb-14|2014|  Feb| 25|
1|Apodiformes|Trochilidae|Phaethornis|Phaethornis_syrma…|         TBHE|6.32|
40|          39.9|          3.9|            3|        null|  null| null|
null|          null|             70|
64|5.223194444444444|-76.07793059868916|       0|       95|19.033333|19.383333|
18.8|2965.0| 416.0| 152.0|34.28378| 970.0| 532.0| 875.0|
970.0|7.816667|90.89148|23.580935| 23.3| 14.7|8.599999| 18.8|18.983334|1493|
S20|
|   TA01|    621|25-Feb-14|2014|  Feb| 25|
1|Apodiformes|Trochilidae|Phaethornis|Phaethornis_syrma…|         TBHE|6.32|
40|          39.9|          3.9|            3|        null|  null| null|
null|          null|             70|
64|5.222503461438818|-76.07804003964577|       0|       95|19.033333|19.383333|
18.8|2965.0| 416.0| 152.0|34.28378| 970.0| 532.0| 875.0|
970.0|7.816667|90.89148|23.580935| 23.3| 14.7|8.599999| 18.8|18.983334|1493|
S19|
+-------+-------+---------+----+-----+---+---+----------+----------+----------
-+--------------------+-------------+----+------------+-------------+----------
---+-------------+------------+------+-----+------------+-------------+--------
-------+---------+-----------------+------------------+--------+---------+------
---+---------+------+------+------+------+--------+------+------+------+------+-
-------+--------+---------+-----+-----+--------+-----+---------+----+-----------
---+
only showing top 2 rows
```

[24]:
```
df_sp = df_sp.drop("Wing_cord", "code _Species", 'Central_rectrix', "PW␣
↪bill_width",
                   'feet_extension', 'Nail_1finger', 'Halux', 'Tarsus',␣
↪'Culmen_gapes' ,
                    'PH bill_depth','Culmen_Exposed',  'Culmen_Total', 'Mass')
```

[25]:
```
from pyspark.sql.functions import monotonically_increasing_id
df_sp = df_sp.withColumn("id", monotonically_increasing_id())

df_sp.show(2)
```

```
+-------+-------+---------+----+-----+---+---+----------+----------+----------
-+--------------------+------------+-----------------+-----------------+------
--+---------+---------+---------+------+------+------+------+--------+------+---
---+------+------+--------+--------+---------+-----+-----+--------+-----+-------
--+----+--------------+
|Station|Entrada|     Date|Year|month|Day| id|      Order|     Family|
Genus|             Species|PW bill_width|          Latitud|
Longitud|lossyear|treecover|    bio_1|   bio_10|bio_11|bio_12|bio_13|bio_14|
bio_15|bio_16|bio_17|bio_18|bio_19|   bio_2|   bio_3|    bio_4|bio_5|bio_6|
bio_7|bio_8|    bio_9|elev|Station_sample|
+-------+-------+---------+----+-----+---+---+----------+----------+----------
-+--------------------+------------+-----------------+-----------------+------
--+---------+---------+---------+------+------+------+------+--------+------+---
---+------+------+--------+--------+---------+-----+-----+--------+-----+-------
--+----+--------------+
|   TA01|    621|25-Feb-14|2014|  Feb| 25|
0|Apodiformes|Trochilidae|Phaethornis|Phaethornis_syrma…|
3.9|5.223194444444444|-76.07793059868916|       0|       95|19.033333|19.383333|
18.8|2965.0| 416.0| 152.0|34.28378| 970.0| 532.0| 875.0|
970.0|7.816667|90.89148|23.580935| 23.3| 14.7|8.599999| 18.8|18.983334|1493|
S20|
|   TA01|    621|25-Feb-14|2014|  Feb| 25|
1|Apodiformes|Trochilidae|Phaethornis|Phaethornis_syrma…|
3.9|5.222503461438818|-76.07804003964577|       0|       95|19.033333|19.383333|
18.8|2965.0| 416.0| 152.0|34.28378| 970.0| 532.0| 875.0|
970.0|7.816667|90.89148|23.580935| 23.3| 14.7|8.599999| 18.8|18.983334|1493|
S19|
+-------+-------+---------+----+-----+---+---+----------+----------+----------
-+--------------------+------------+-----------------+-----------------+------
--+---------+---------+---------+------+------+------+------+--------+------+---
---+------+------+--------+--------+---------+-----+-----+--------+-----+-------
--+----+--------------+
only showing top 2 rows
```

[26]:
```
from pyspark.ml.feature import VectorAssembler, StandardScaler
from pyspark.ml.clustering import KMeans
from pyspark.ml.feature import PCA
from pyspark.ml.linalg import Vectors
import numpy as np
import matplotlib.pyplot as plt
```

[27]:
```
columns_for_pca = ["treecover", "bio_1", "bio_10", "bio_11", "bio_12",␣
 ↪"bio_13", "bio_14", "bio_15",
                   "bio_16", "bio_17", "bio_18", "bio_19", "bio_2", "bio_3",␣
 ↪"bio_4", "bio_5", "bio_6", "bio_7",
                   "bio_8", "bio_9", "elev"]
```

[28]:
```
assembler = VectorAssembler(inputCols=columns_for_pca, outputCol="features")
df_vector = assembler.transform(df_sp)
```

[29]:
```
scaler = StandardScaler(inputCol="features", outputCol="scaledFeatures",␣
 ↪withStd=True, withMean=False)
scalerModel = scaler.fit(df_vector)
df_scaled = scalerModel.transform(df_vector)
df_scaled.show(6)
```

```
+-------+-------+---------+----+-----+---+---+-----------+-----------+----------
-+--------------------+-------------+-----------------+------------------+------
--+---------+---------+---------+------+------+------+------+--------+------+---
---+------+------+--------+--------+---------+-----+-----+--------+-----+-------
--+----+--------------+--------------------+--------------------+
|Station|Entrada|     Date|Year|month|Day| id|      Order|     Family|
Genus|             Species|PW bill_width|          Latitud|
Longitud|lossyear|treecover|    bio_1|   bio_10|bio_11|bio_12|bio_13|bio_14|
bio_15|bio_16|bio_17|bio_18|bio_19|   bio_2|   bio_3|    bio_4|bio_5|bio_6|
bio_7|bio_8|    bio_9|elev|Station_sample|            features|
scaledFeatures|
+-------+-------+---------+----+-----+---+---+-----------+-----------+----------
-+--------------------+-------------+-----------------+------------------+------
--+---------+---------+---------+------+------+------+------+--------+------+---
---+------+------+--------+--------+---------+-----+-----+--------+-----+-------
--+----+--------------+--------------------+--------------------+
|   TA01|    621|25-Feb-14|2014|  Feb| 25|
0|Apodiformes|Trochilidae|Phaethornis|Phaethornis_syrma…|
3.9|5.223194444444444|-76.07793059868916|       0|       95|19.033333|19.383333|
18.8|2965.0| 416.0| 152.0|34.28378| 970.0| 532.0| 875.0|
970.0|7.816667|90.89148|23.580935| 23.3| 14.7|8.599999| 18.8|18.983334|1493|
S20|[95.0,19.033333,1…|[5.28145580923873…|
|   TA01|    621|25-Feb-14|2014|  Feb| 25|
1|Apodiformes|Trochilidae|Phaethornis|Phaethornis_syrma…|
3.9|5.222503461438818|-76.07804003964577|       0|       95|19.033333|19.383333|
```

```
18.8|2965.0| 416.0| 152.0|34.28378| 970.0| 532.0| 875.0|
970.0|7.816667|90.89148|23.580935| 23.3| 14.7|8.599999| 18.8|18.983334|1493|
S19|[95.0,19.033333,1…|[5.28145580923873…|
|   TA01|    621|25-Feb-14|2014|  Feb| 25|
2|Apodiformes|Trochilidae|Phaethornis|Phaethornis_syrma…|
3.9|5.221880116664146|-76.07835764967228|       0|       95|19.033333|19.383333|
18.8|2965.0| 416.0| 152.0|34.28378| 970.0| 532.0| 875.0|
970.0|7.816667|90.89148|23.580935| 23.3| 14.7|8.599999| 18.8|18.983334|1493|
S18|[95.0,19.033333,1…|[5.28145580923873…|
|   TA01|    621|25-Feb-14|2014|  Feb| 25|
3|Apodiformes|Trochilidae|Phaethornis|Phaethornis_syrma…|
3.9|5.221385427450069|-76.07885233888635|       0|       90|19.033333|19.383333|
18.8|2965.0| 416.0| 152.0|34.28378| 970.0| 532.0| 875.0|
970.0|7.816667|90.89148|23.580935| 23.3| 14.7|8.599999| 18.8|18.983334|1493|
S17|[90.0,19.033333,1…|[5.00348445085774…|
|   TA01|    621|25-Feb-14|2014|  Feb| 25|
4|Apodiformes|Trochilidae|Phaethornis|Phaethornis_syrma…|
3.9|5.221067817423564|-76.07947568366103|       0|       90|19.033333|19.383333|
18.8|2965.0| 416.0| 152.0|34.28378| 970.0| 532.0| 875.0|
970.0|7.816667|90.89148|23.580935| 23.3| 14.7|8.599999| 18.8|18.983334|1493|
S16|[90.0,19.033333,1…|[5.00348445085774…|
|   TA01|    621|25-Feb-14|2014|  Feb| 25|
5|Apodiformes|Trochilidae|Phaethornis|Phaethornis_syrma…|
3.9|5.220958376466945|-76.08016666666666|       0|       95|19.033333|19.383333|
18.8|2965.0| 416.0| 152.0|34.28378| 970.0| 532.0| 875.0|
970.0|7.816667|90.89148|23.580935| 23.3| 14.7|8.599999| 18.8|18.983334|1493|
S15|[95.0,19.033333,1…|[5.28145580923873…|
+-------+-------+---------+----+-----+---+---+-----------+-----------+----------
-+--------------------+-------------+-----------------+-----------------+------
--+---------+---------+---------+------+------+------+------+--------+------+---
---+------+------+--------+--------+---------+-----+-----+--------+-----+-------
--+----+-------------+--------------------+--------------------+
only showing top 6 rows
```

[30]:
```
n_components = len(columns_for_pca)
```

[31]:
```
pca = PCA(k=n_components, inputCol="scaledFeatures", outputCol="pcaFeatures")
pcaModel = pca.fit(df_scaled)
df_pca = pcaModel.transform(df_scaled)
```

[32]:
```
# Varianza explicada por cada componente principal
explainedVariance = pcaModel.explainedVariance
# Varianza acumulada
cumulativeVariance = explainedVariance.cumsum()
```

[33]:
```
print("Varianza explicada por cada componente principal:")
print(explainedVariance)
```

```
Varianza explicada por cada componente principal:
[0.7432952779327794,0.12998221241187907,0.04817708788095776,0.04164101832457,0.0
18502906414520206,0.011051574919055781,0.0031070658744734177,0.00163710023176608
5,0.0010650491820818266,0.0007299359706796621,0.0004374224961990203,0.0002534773
5953235457,7.251391294561552e-05,3.474961683847048e-05,1.0703268831694875e-
05,1.7847320080810641e-06,9.278156778362384e-08,2.4064112238068714e-
08,2.625124902246856e-09,4.978739172667878e-14,2.6864249417919408e-14]
```

[34]:
```
print("\nVarianza acumulada:")
print(cumulativeVariance)
```

```
Varianza acumulada:
[0.74329528 0.87327749 0.92145458 0.9630956  0.9815985  0.99265008
 0.99575714 0.99739424 0.99845929 0.99918923 0.99962665 0.99988013
 0.99995264 0.99998739 0.9999981  0.99999988 0.99999997 1.
 1.         1.         1.        ]
```

[36]:
```
optimal_k = np.argmax(cumulativeVariance >= 0.90) + 1  # +1 porque los índices␣
↪en Python empiezan en 0
print("\nC90% de varianza explicada:", optimal_k)
```

```
C90% de varianza explicada: 3
```

[47]:
```
#############################################################################################333
df_pca.show(2)
```

```
+-------+-------+---------+----+-----+---+---+----------+----------+----------
-+--------------------+-------------+-----------------+-----------------+------
--+---------+---------+---------+------+------+------+------+--------+------+---
---+------+------+--------+--------+---------+-----+-----+--------+-----+-------
--+----+--------------+--------------------+--------------------+--------------
-----+
|Station|Entrada|     Date|Year|month|Day| id|      Order|     Family|
Genus|             Species|PW bill_width|          Latitud|
Longitud|lossyear|treecover|    bio_1|   bio_10|bio_11|bio_12|bio_13|bio_14|
bio_15|bio_16|bio_17|bio_18|bio_19|   bio_2|   bio_3|    bio_4|bio_5|bio_6|
bio_7|bio_8|    bio_9|elev|Station_sample|            features|
scaledFeatures|         pcaFeatures|
+-------+-------+---------+----+-----+---+---+----------+----------+----------
-+--------------------+-------------+-----------------+-----------------+------
--+---------+---------+---------+------+------+------+------+--------+------+---
---+------+------+--------+--------+---------+-----+-----+--------+-----+-------
--+----+--------------+--------------------+--------------------+--------------
```

```
-----+
|   TA01|    621|25-Feb-14|2014|  Feb| 25|
0|Apodiformes|Trochilidae|Phaethornis|Phaethornis_syrma…|
3.9|5.223194444444444|-76.07793059868916|       0|       95|19.033333|19.383333|
18.8|2965.0| 416.0| 152.0|34.28378| 970.0| 532.0| 875.0|
970.0|7.816667|90.89148|23.580935| 23.3| 14.7|8.599999| 18.8|18.983334|1493|
S20|[95.0,19.033333,1…|[5.28145580923873…|[56.0384615290574…|
|   TA01|    621|25-Feb-14|2014|  Feb| 25|
1|Apodiformes|Trochilidae|Phaethornis|Phaethornis_syrma…|
3.9|5.222503461438818|-76.07804003964577|       0|       95|19.033333|19.383333|
18.8|2965.0| 416.0| 152.0|34.28378| 970.0| 532.0| 875.0|
970.0|7.816667|90.89148|23.580935| 23.3| 14.7|8.599999| 18.8|18.983334|1493|
S19|[95.0,19.033333,1…|[5.28145580923873…|[56.0384615290574…|
+-------+-------+---------+----+-----+---+---+----------+----------+----------
-+--------------------+------------+-----------------+-----------------+------
--+---------+---------+---------+------+------+------+------+--------+------+---
---+------+------+--------+--------+---------+-----+-----+--------+-----+-------
--+----+--------------+--------------------+--------------------+--------------
-----+
only showing top 2 rows
```

[ ]:
```
```

[48]:
```
# Ensamblar solo las columnas que se utilizarán para PCA
assembler = VectorAssembler(inputCols=columns_for_pca, outputCol="features")
df_vector = assembler.transform(df_sp)
```

[49]:
```
scaler = StandardScaler(inputCol="features", outputCol="scaledFeatures",␣
↪withStd=True, withMean=False)
scalerModel = scaler.fit(df_vector)
df_scaled = scalerModel.transform(df_vector)
```

[50]:
```
pca = PCA(k=3, inputCol="scaledFeatures", outputCol="pcaFeatures")
pcaModel = pca.fit(df_scaled)
df_pca = pcaModel.transform(df_scaled)
```

[55]:
```
df_pca.select("pcaFeatures").show(truncate=False)
```

```
+-----------------------------------------------------------+
|pcaFeatures                                                |
+-----------------------------------------------------------+
|[56.038461529057415,-11.052202954735705,21.779740548918824]|
|[56.038461529057415,-11.052202954735705,21.779740548918824]|
|[56.038461529057415,-11.052202954735705,21.779740548918824]|
|[56.037949245249855,-11.078376771435948,21.522740004571258]|
|[56.037949245249855,-11.078376771435948,21.522740004571258]|
|[56.038461529057415,-11.052202954735705,21.779740548918824]|
```

```
|[56.037949245249855,-11.078376771435948,21.522740004571258]|
|[56.02975270432888,-11.497157838639826,17.410731295010205] |
|[56.02872813671375,-11.54950547204031,16.89673020631507]   |
|[56.037949245249855,-11.078376771435948,21.522740004571258]|
|[56.037949245249855,-11.078376771435948,21.522740004571258]|
|[56.02934287728283,-11.51809689200002,17.20513085953215]   |
|[56.03251903688971,-11.355819228458518,18.798534234487057] |
|[56.037949245249855,-11.078376771435948,21.522740004571258]|
|[54.59309908144201,-15.700881425299293,22.23430700608372]  |
|[54.59309908144201,-15.700881425299293,22.23430700608372]  |
|[54.59309908144201,-15.700881425299293,22.23430700608372]  |
|[56.038461529057415,-11.052202954735705,21.779740548918824]|
|[56.038461529057415,-11.052202954735705,21.779740548918824]|
|[56.038461529057415,-11.052202954735705,21.779740548918824]|
+-----------------------------------------------------------+
only showing top 20 rows
```

[52]:
```python
from pyspark.ml.clustering import KMeans
```

[56]:
```python

```

[60]:
```python
k = 3

# Configura KMeans para que utilice la columna 'pcaFeatures'
kmeans = KMeans().setK(k).setSeed(123).setFeaturesCol("pcaFeatures")

# Ajusta el modelo KMeans utilizando solo las columnas 'id' y 'pcaFeatures'
model = kmeans.fit(df_pca.select('id', 'pcaFeatures'))

# Realiza las predicciones
predictions = model.transform(df_pca.select('id', 'pcaFeatures'))

# Une las predicciones con el DataFrame original para añadir la columna de␣
 ↪predicciones
df_sp_with_predictions = df_sp.join(predictions, 'id')
```

[64]:
```python
from pyspark.ml.evaluation import ClusteringEvaluator

# Suponiendo que 'model' es el modelo KMeans ya ajustado y 'predictions'␣
 ↪contiene las predicciones

# Inercia (coste de entrenamiento)
inertia = model.summary.trainingCost

# Coeficiente de silueta
evaluator = ClusteringEvaluator(featuresCol="pcaFeatures")
```

```
silhouette = evaluator.evaluate(predictions)

print("Coeficiente de silueta:", silhouette)

print("Inercia:", inertia)
```

```
Coeficiente de silueta: 0.7635437048318442
Inercia: 259979.75872146673
```

[66]:
```
# Muestra las primeras filas con las predicciones añadidas
df_sp_with_predictions.show(4)
```

```
+---+-------+-------+---------+----+-----+---+------------+----------+---------
+--------------------+-------------+-----------------+------------------+-------
-+---------+---------+---------+------+------+------+------+--------+------+----
--+------+------+--------+--------+---------+-----+-----+--------+-----+--------
-+----+--------------+--------------------+----------+
| id|Station|Entrada|     Date|Year|month|Day|        Order|    Family|
Genus|             Species|PW bill_width|          Latitud|
Longitud|lossyear|treecover|    bio_1|   bio_10|bio_11|bio_12|bio_13|bio_14|
bio_15|bio_16|bio_17|bio_18|bio_19|   bio_2|   bio_3|    bio_4|bio_5|bio_6|
bio_7|bio_8|    bio_9|elev|Station_sample|         pcaFeatures|prediction|
+---+-------+-------+---------+----+-----+---+------------+----------+---------
+--------------------+-------------+-----------------+------------------+-------
-+---------+---------+---------+------+------+------+------+--------+------+----
--+------+------+--------+--------+---------+-----+-----+--------+-----+--------
-+----+--------------+--------------------+----------+
| 26|   TA01|    622|25-Feb-14|2014|  Feb|
25|Passeriformes|Tyrannidae|Mionectes|Mionectes_striati…|
4.5|5.220958376466945|-76.08016666666666|       0|       95|19.033333|19.383333|
18.8|2965.0| 416.0| 152.0|34.28378| 970.0| 532.0| 875.0|
970.0|7.816667|90.89148|23.580935| 23.3| 14.7|8.599999| 18.8|18.983334|1493|
S15|[56.0384615290574…|         0|
| 29|   TA01|    622|25-Feb-14|2014|  Feb|
25|Passeriformes|Tyrannidae|Mionectes|Mionectes_striati…|
4.5|5.221880116664146|-76.08197568366103|       0|        0|19.033333|19.383333|
18.8|2965.0| 416.0| 152.0|34.28378| 970.0| 532.0| 875.0|
970.0|7.816667|90.89148|23.580935| 23.3| 14.7|8.599999| 18.8|18.983334|1493|
S12|[56.0287281367137…|         0|
|474|   TA01|    643|26-Feb-14|2014|  Feb| 26|Passeriformes|Thraupidae|
Tangara|      Tangara_arthus|          7.8|5.224508772224742|-76.08197568366103|
0|       37|19.033333|19.383333|  18.8|2965.0| 416.0| 152.0|34.28378| 970.0|
532.0| 875.0| 970.0|7.816667|90.89148|23.580935| 23.3| 14.7|8.599999|
18.8|18.983334|1493|            S8|[56.0325190368897…|         0|
|964|   TA01|    667|26-Feb-14|2014|  Feb| 26|Passeriformes|Cotingidae|
Rupicola| Rupicola_peruvianus|         11.7|
5.22388542745007|-76.07804003964577|       0|       95|19.033333|19.383333|
18.8|2965.0| 416.0| 152.0|34.28378| 970.0| 532.0| 875.0|
```

```
970.0|7.816667|90.89148|23.580935| 23.3| 14.7|8.599999| 18.8|18.983334|1493|
S1|[56.0384615290574…|         0|
+---+-------+-------+---------+----+-----+---+------------+---------+--------
+--------------------+------------+-----------------+-----------------+-------
-+---------+---------+---------+------+------+------+------+--------+------+----
--+------+------+--------+--------+---------+-----+-----+--------+-----+--------
-+----+--------------+--------------------+----------+
only showing top 4 rows
```

[67]:
```
df_pandas = df_sp_with_predictions.toPandas()

# Separar características y etiquetas
X = df_pandas[['treecover', 'bio_1', 'bio_10', 'bio_11', 'bio_12', 'bio_13',␣
 ↪'bio_14', 'bio_15', 'bio_16', 'bio_17', 'bio_18', 'bio_19', 'bio_2',␣
 ↪'bio_3', 'bio_4', 'bio_5', 'bio_6', 'bio_7', 'bio_8', 'bio_9', 'elev']]
y = df_pandas['prediction']
```

[107]:
```
X.head()
```

[107]:

	treecover	bio_1	bio_10	bio_11	bio_12	bio_13	bio_14	bio_15	\
0	95	19.033333	19.383333	18.8	2965.0	416.0	152.0	34.28378	
1	0	19.033333	19.383333	18.8	2965.0	416.0	152.0	34.28378	
2	37	19.033333	19.383333	18.8	2965.0	416.0	152.0	34.28378	
3	95	19.033333	19.383333	18.8	2965.0	416.0	152.0	34.28378	
4	95	19.033333	19.383333	18.8	2965.0	416.0	152.0	34.28378	

	bio_16	bio_17	…	bio_19	bio_2	bio_3	bio_4	bio_5	bio_6	\
0	970.0	532.0	…	970.0	7.816667	90.89148	23.580935	23.3	14.7	
1	970.0	532.0	…	970.0	7.816667	90.89148	23.580935	23.3	14.7	
2	970.0	532.0	…	970.0	7.816667	90.89148	23.580935	23.3	14.7	
3	970.0	532.0	…	970.0	7.816667	90.89148	23.580935	23.3	14.7	
4	970.0	532.0	…	970.0	7.816667	90.89148	23.580935	23.3	14.7	

	bio_7	bio_8	bio_9	elev
0	8.599999	18.8	18.983334	1493
1	8.599999	18.8	18.983334	1493
2	8.599999	18.8	18.983334	1493
3	8.599999	18.8	18.983334	1493
4	8.599999	18.8	18.983334	1493

[5 rows x 21 columns]

[91]:
```
import xgboost as xgb
from hyperopt import hp, fmin, tpe, STATUS_OK, Trials
from sklearn.metrics import mean_squared_error
from sklearn.model_selection import train_test_split
```

```
import numpy as np
from hyperopt import fmin, tpe, hp, STATUS_OK, Trials
```

[88]:
```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,␣
↪random_state=42)

# Convertir los conjuntos a DMatrix, que es el formato preferido por XGBoost
dtrain = xgb.DMatrix(X_train, label=y_train)
dtest = xgb.DMatrix(X_test, label=y_test)
```

[92]:
```
from sklearn.metrics import f1_score
```

[93]:
```
def objective(space):
    clf = xgb.train(space, dtrain, num_boost_round=1000, evals=[(dtest,␣
↪"test")], early_stopping_rounds=30, verbose_eval=False)

    preds = clf.predict(dtest)
    # Asegúrate de que 'preds' sean las clases predichas para calcular F1-score
    preds = np.asarray([np.argmax(line) for line in preds])
    f1 = f1_score(y_test, preds, average='weighted')  # 'weighted' considera el␣
↪desbalance de clases

    # Hyperopt minimiza la función objetivo; se retorna 1 - F1 para maximizar␣
↪el F1-score
    return {'loss': 1 - f1, 'status': STATUS_OK}

space = {
    'max_depth': hp.choice('max_depth', np.arange(1, 14, dtype=int)),
    'eta': hp.uniform('eta', 0.01, 0.3),
    'objective': 'multi:softmax',
    'num_class': 3,
    'lambda': hp.uniform('lambda', 1e-8, 1.0),  # Regularización L2
    'alpha': hp.uniform('alpha', 1e-8, 1.0)  # Regularización L1
}

trials = Trials()
best_hyperparams = fmin(fn=objective,
                        space=space,
                        algo=tpe.suggest,
                        max_evals=100,
                        trials=trials)

print("Los mejores hiperparámetros son: ", best_hyperparams)
```

```
100%|      | 100/100 [03:04<00:00,  1.84s/trial, best loss:
0.7035852236355664]
Los mejores hiperparámetros son:  {'alpha': 0.6590523500251971, 'eta':
```

```
0.16275265224049576, 'lambda': 0.0033755891434843924, 'max_depth': 12}
```

[94]:
```
import joblib
```

[95]:
```
final_params = {
    'max_depth': 12,
    'eta': 0.16275265224049576,
    'objective': 'multi:softmax',
    'num_class': 3,
    'lambda': 0.0033755891434843924,
    'alpha': 0.6590523500251971
}
```

[96]:
```
final_bst = xgb.train(final_params, dtrain, num_boost_round=1000)

# Guardar el modelo
joblib.dump(final_bst, "xgboost_aves_model.dat")
```

[96]:
```
['xgboost_aves_model.dat']
```

[147]:
```
import rasterio
import os
from rasterio.enums import Resampling
```

[ ]:
```
##########################################3 Mapping
```

[108]:
```
xgb_model = joblib.load("xgboost_aves_model.dat")
```

[113]:
```
def read_rasters_and_stack(raster_files, folder_path):
    data = []
    for file in raster_files:
        with rasterio.open(os.path.join(folder_path, file)) as src:
            band = src.read(1)  # read only the first band
            data.append(band)
    return np.stack(data, axis=-1)  # stack rasters
```

[114]:
```
nombres_columnas = X.columns.tolist()
nombres_columnas
```

[114]:
```
['treecover',
 'bio_1',
 'bio_10',
 'bio_11',
 'bio_12',
 'bio_13',
 'bio_14',
 'bio_15',
```

```
 'bio_16',
 'bio_17',
 'bio_18',
 'bio_19',
 'bio_2',
 'bio_3',
 'bio_4',
 'bio_5',
 'bio_6',
 'bio_7',
 'bio_8',
 'bio_9',
 'elev']
```

[142]:
```
folder_path = 'resampled_rasters'


raster_files = ['r_aligned_clipped_Hansen_GFC-2022-v1.10_treecover2000_10N_080W.
 ↪tif',
 'resampled_clipped_wc2.1_30s_bio_1.tif',
 'resampled_clipped_wc2.1_30s_bio_10.tif',
 'resampled_clipped_wc2.1_30s_bio_11.tif',
 'resampled_clipped_wc2.1_30s_bio_12.tif',
 'resampled_clipped_wc2.1_30s_bio_13.tif',
 'resampled_clipped_wc2.1_30s_bio_14.tif',
 'resampled_clipped_wc2.1_30s_bio_15.tif',
 'resampled_clipped_wc2.1_30s_bio_16.tif',
 'resampled_clipped_wc2.1_30s_bio_17.tif',
 'resampled_clipped_wc2.1_30s_bio_18.tif',
 'resampled_clipped_wc2.1_30s_bio_19.tif',
 'resampled_clipped_wc2.1_30s_bio_2.tif',
 'resampled_clipped_wc2.1_30s_bio_3.tif',
 'resampled_clipped_wc2.1_30s_bio_4.tif',
 'resampled_clipped_wc2.1_30s_bio_5.tif',
 'resampled_clipped_wc2.1_30s_bio_6.tif',
 'resampled_clipped_wc2.1_30s_bio_7.tif',
 'resampled_clipped_wc2.1_30s_bio_8.tif',
 'resampled_clipped_wc2.1_30s_bio_9.tif',
 'resampled_clipped_wc2.1_30s_elev.tif'
]
```

[144]:
```
# Check shapes of all rasters
for file in raster_files:
    with rasterio.open(os.path.join(folder_path, file)) as src:
        print(file, src.shape)
```

```
r_aligned_clipped_Hansen_GFC-2022-v1.10_treecover2000_10N_080W.tif (1634, 1534)
```

```
resampled_clipped_wc2.1_30s_bio_1.tif (1634, 1534)
resampled_clipped_wc2.1_30s_bio_10.tif (1634, 1534)
resampled_clipped_wc2.1_30s_bio_11.tif (1634, 1534)
resampled_clipped_wc2.1_30s_bio_12.tif (1634, 1534)
resampled_clipped_wc2.1_30s_bio_13.tif (1634, 1534)
resampled_clipped_wc2.1_30s_bio_14.tif (1634, 1534)
resampled_clipped_wc2.1_30s_bio_15.tif (1634, 1534)
resampled_clipped_wc2.1_30s_bio_16.tif (1634, 1534)
resampled_clipped_wc2.1_30s_bio_17.tif (1634, 1534)
resampled_clipped_wc2.1_30s_bio_18.tif (1634, 1534)
resampled_clipped_wc2.1_30s_bio_19.tif (1634, 1534)
resampled_clipped_wc2.1_30s_bio_2.tif (1634, 1534)
resampled_clipped_wc2.1_30s_bio_3.tif (1634, 1534)
resampled_clipped_wc2.1_30s_bio_4.tif (1634, 1534)
resampled_clipped_wc2.1_30s_bio_5.tif (1634, 1534)
resampled_clipped_wc2.1_30s_bio_6.tif (1634, 1534)
resampled_clipped_wc2.1_30s_bio_7.tif (1634, 1534)
resampled_clipped_wc2.1_30s_bio_8.tif (1634, 1534)
resampled_clipped_wc2.1_30s_bio_9.tif (1634, 1534)
resampled_clipped_wc2.1_30s_elev.tif (1634, 1534)
```

[ ]:

[143]:
```
# Read the rasters
raster_stack = read_rasters_and_stack(raster_files, folder_path)
```

[148]:
```
# Reshape the raster stack for prediction
num_features = raster_stack.shape[-1]
raster_stack_reshaped = raster_stack.reshape(-1, num_features)
```

[151]:
```
feature_names = ['treecover', 'bio_1', 'bio_10', 'bio_11', 'bio_12', 'bio_13',␣
 ↪'bio_14', 'bio_15', 'bio_16', 'bio_17', 'bio_18', 'bio_19', 'bio_2',␣
 ↪'bio_3', 'bio_4', 'bio_5', 'bio_6', 'bio_7', 'bio_8', 'bio_9', 'elev']

# Create the DMatrix for prediction, specifying feature names
dtest = xgb.DMatrix(raster_stack_reshaped, feature_names=feature_names)

# Predict using the model
predictions = model.predict(dtest)
```

[152]:
```
predictions_reshaped = predictions.reshape(raster_stack.shape[0], raster_stack.
 ↪shape[1])
```

[153]:
```
with rasterio.open(os.path.join(folder_path, raster_files[0])) as src:
    meta = src.meta.copy()
    meta.update(dtype='uint8', count=1)  # Assuming predictions are categorical
```

```
with rasterio.open('predictions.tif', 'w', **meta) as dst:
    dst.write(predictions_reshaped.astype('uint8'), 1)
```

[ ]:

[167]:
```
########################################################################
import geopandas as gpd
from rasterio.mask import mask
from rasterio.plot import show
```

[205]:
```
shapefile_path = 'Shape/tTatama.shp'  # Ruta al archivo shapefile
raster_path = 'predictions.tif'       # Ruta al archivo raster
# Leer el shapefile usando Geopandas
shapes = gpd.read_file(shapefile_path)
```

[ ]:

[ ]:

[215]:

```
Valores únicos en el raster: [0 1 2]
```

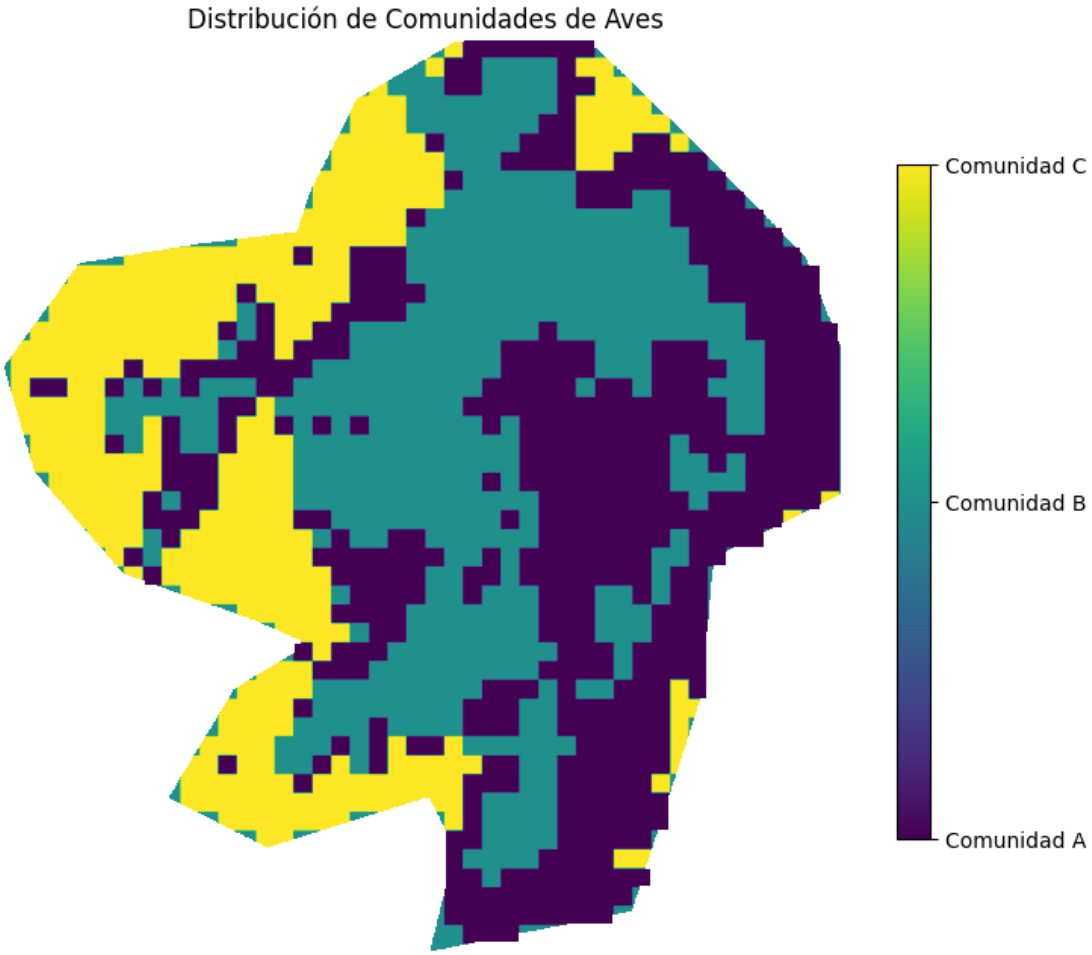
[216]:

[ ]:

[226]:
```
import geopandas as gpd

import rasterio.mask
import matplotlib.colors
```

[225]:
```
nodata_value = -9999

shapes = gpd.read_file('Shape/tTatama.shp')

# Leer el raster de predicciones usando Rasterio
with rasterio.open('predictions.tif') as src:
    # Leer la primera banda del raster como un MaskedArray
    raster_data = src.read(1, masked=True)

    # Crea una máscara para el raster basada en el shapefile, donde todo fuera␣
 ↪de la geometría sea True
    raster_mask, out_transform = rasterio.mask.mask(src, shapes.geometry,␣
 ↪invert=True)
    # Combina la máscara de la geometría con la máscara existente en los datos
    raster_data.mask |= raster_mask[0]
```

```
# Asignar el colormap manualmente
cmap = plt.cm.viridis  # o cualquier otro colormap que prefieras
cmap.set_bad('white')  # Asignar el color blanco para valores nodata/máscara

# Crear la figura y los ejes para el mapa
fig, ax = plt.subplots(figsize=(8, 8))
img = ax.imshow(raster_data, cmap=cmap)

# Agregar la barra de colores con etiquetas para las comunidades
colorbar = fig.colorbar(img, ax=ax, fraction=0.036, pad=0.04, ticks=[0, 1, 2])
colorbar.ax.set_yticklabels(['Comunidad A', 'Comunidad B', 'Comunidad C'])

# Esconder los ejes
ax.axis('off')

# Agregar título
ax.set_title('Distribución de Comunidades de Aves')

# Mostrar el gráfico
plt.show()
```


Distribución de Comunidades de Aves
Comunidad C
Comunidad B
Comunidad A

[ ]: