UNIVERSITÉ DU LUXEMBOURG

Master in Data Science

Workshop I - Environmental Data Analytics

Winter Semester 2024

Birds Occurrence in Kenya

Instructor:
Mohammad Mahdi Rajabi

Students:
Clara Duchossois
Patrick Silva

Contents

1 Data Collection and Data Cleaning

The data used for this project was provided to us and can be found under this link: **https://www.gbif.org/dataset/d1e8a7af-097a-4ca8-aa91-62d3ff834711**.

This link offers multiple ways to download the data. The recommended way requires an account. To avoid having to create an account, we downloaded using the second option which gives us a .txt file and some .xml files with meta information.

The data is mostly clean. Most columns are not needed for the given tasks. We will focus on the 1st, 2nd, and 4th task. That most columns would not be necessary was not known to us at the beginning, so we studied the data to see if everything was clean.

In the data, the only column with null values is the 'individualCount'. Because this column is necessary, all the rows with a null value in that column are dropped. We also checked the number of unique values in each column and some columns just have 1 unique value. These columns have been dropped to make the data smaller in size and we stored these unique values as variables in our environment.

This last modification was made after the project was finished as we did not notice the issue at first. In the 'order' column, there are some typos. These had to be corrected so that the orders would be complete.

In the different plots, you will see the map of Kenya and its counties in the background. To display these, we downloaded files from **Natural Earth Data** and **GitHub**. The downloaded Kenya files could be interpreted and displayed using the GeoPandas and MatPlotLib libraries.

The columns that will be mainly used are 'decimalLongitude' and 'decimalLatitude'. These two allow us to draw our observations on the map. The other two columns are 'individualCount' and 'order'. The 'individualCount' column tells us how many birds have been sighted at the same time on the same place. The 'order' describes the bird's order.

2 Taxonomic orders distribution across Kenya

In this chapter the distribution of each order in Kenya will be shown. The following plots will all have a map of Kenya with points corresponding to where the birds have been seen. The size of the point corresponds to how many birds have been seen together. To realize this, the following columns were used: decimalLatitude, decimalLongitude, individualCount, order

Looking at the different distributions, we can see a pattern emerge through Kenya. There is a lot of activity on the south-western/central part of Kenya.

There is one order called "p". There is no information about that order or what it stands for.

Additionally, to each order the Moran's I has been calculated to check whether or not that specific order is living in clusters. As a reminder, if that value approaches 1, the birds are supposed to live in clusters. If it is around -1, it means that the birds live in dispersion. There is no order living in dispersion. If the value is near to 0, it means that their location is random. Some orders obtain NaN as value. This means that there are not enough data points to fulfill the calculation properly.

The code calculating the Moran's value fails for one specific order. The order is 'Passeriformes'. A solution would be to maybe use the Pysal library. However, in all of our attempts, the installation of that library failed due to the existence of new version among dependencies. We opted to leave this order out and have it work for all the other orders.

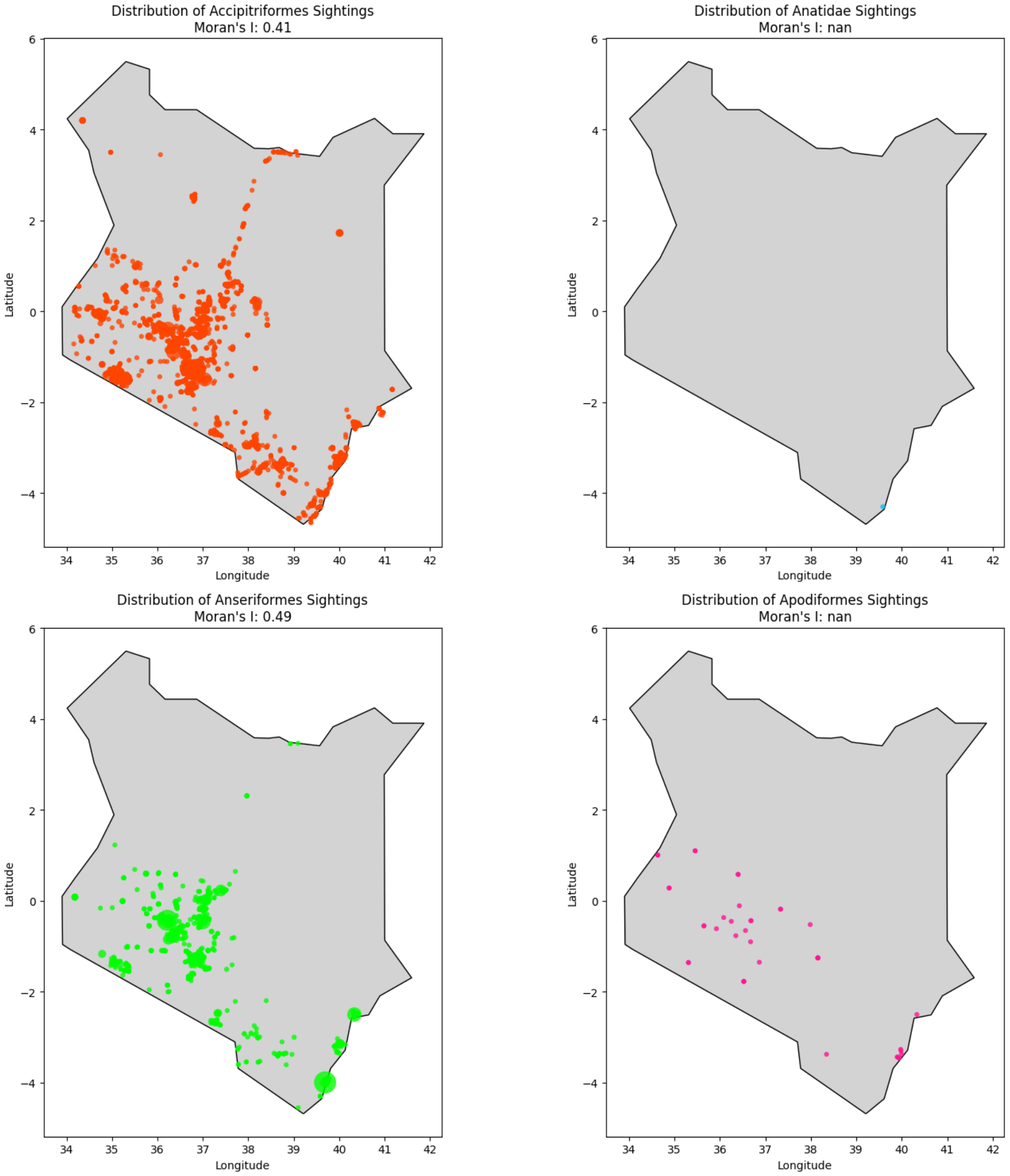

Distribution of Accipitriformes Sightings
Moran's I: 0.41
Distribution of Anatidae Sightings
Moran's I: nan
Distribution of Anseriformes Sightings
Moran's I: 0.49
Distribution of Apodiformes Sightings
Moran's I: nan
Latitude
Longitude
6
4
2
0
−2
−4
34
35
36
37
38
39
40
41
42

Figure 1: Distribution of Accipitriformes, Anatidae, Anseriformes, Apodiformes

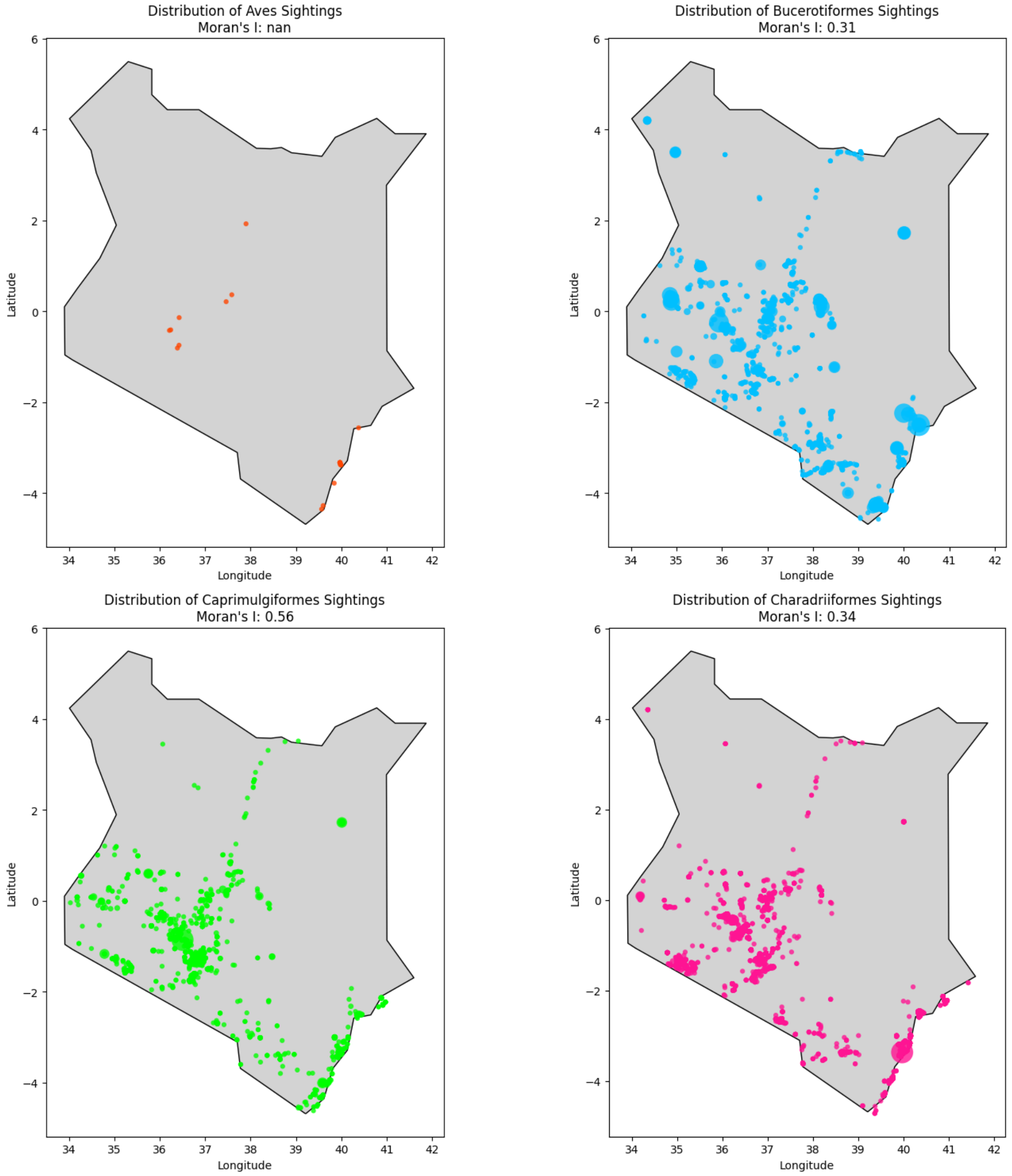

Distribution of Aves Sightings
Moran's I: nan
Distribution of Bucerotiformes Sightings
Moran's I: 0.31
Distribution of Caprimulgiformes Sightings
Moran's I: 0.56
Distribution of Charadriiformes Sightings
Moran's I: 0.34
Latitude
Longitude
6
4
2
0
−2
−4
34
35
36
37
38
39
40
41
42

Figure 2: Distribution of Aves, Bucerotiformes, Caprimulgiformes, Charadriformes

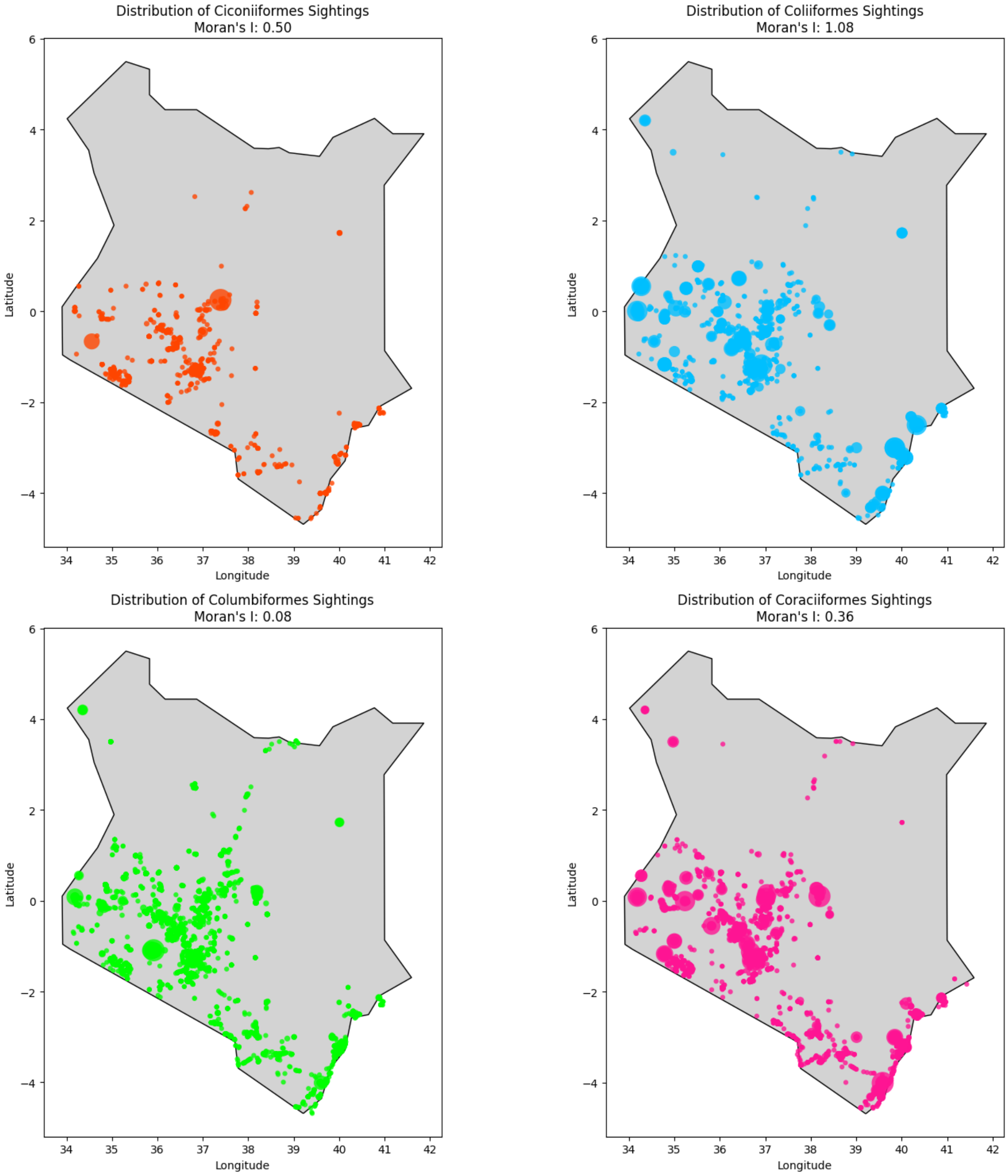

Distribution of Ciconiiformes Sightings
Moran's I: 0.50
Distribution of Coliiformes Sightings
Moran's I: 1.08
Distribution of Columbiformes Sightings
Moran's I: 0.08
Distribution of Coraciiformes Sightings
Moran's I: 0.36
Latitude
Longitude
6
4
2
0
−2
−4
34
35
36
37
38
39
40
41
42

Figure 3: Distribution of Ciconiiformes, Coliiformes, Columbiformes, Coraciiformes

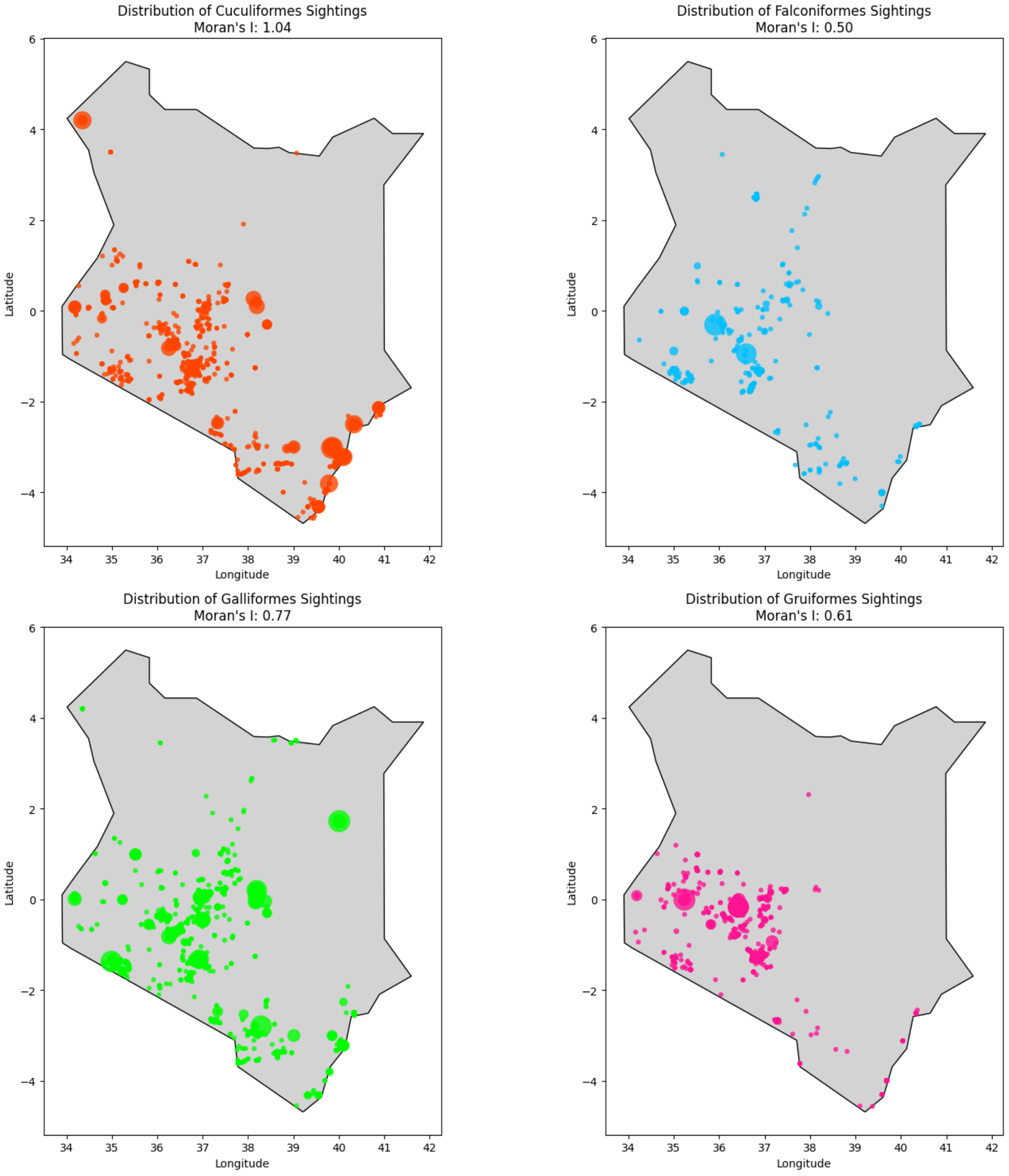

Distribution of Cuculiformes Sightings
Moran's I: 1.04
Distribution of Falconiformes Sightings
Moran's I: 0.50
Distribution of Galliformes Sightings
Moran's I: 0.77
Distribution of Gruiformes Sightings
Moran's I: 0.61
Latitude
Longitude
6
4
2
0
−2
−4
34
35
36
37
38
39
40
41
42

Figure 4: Distribution of Cuculiformes, Falconiformes, Galliformes, Gruiformes

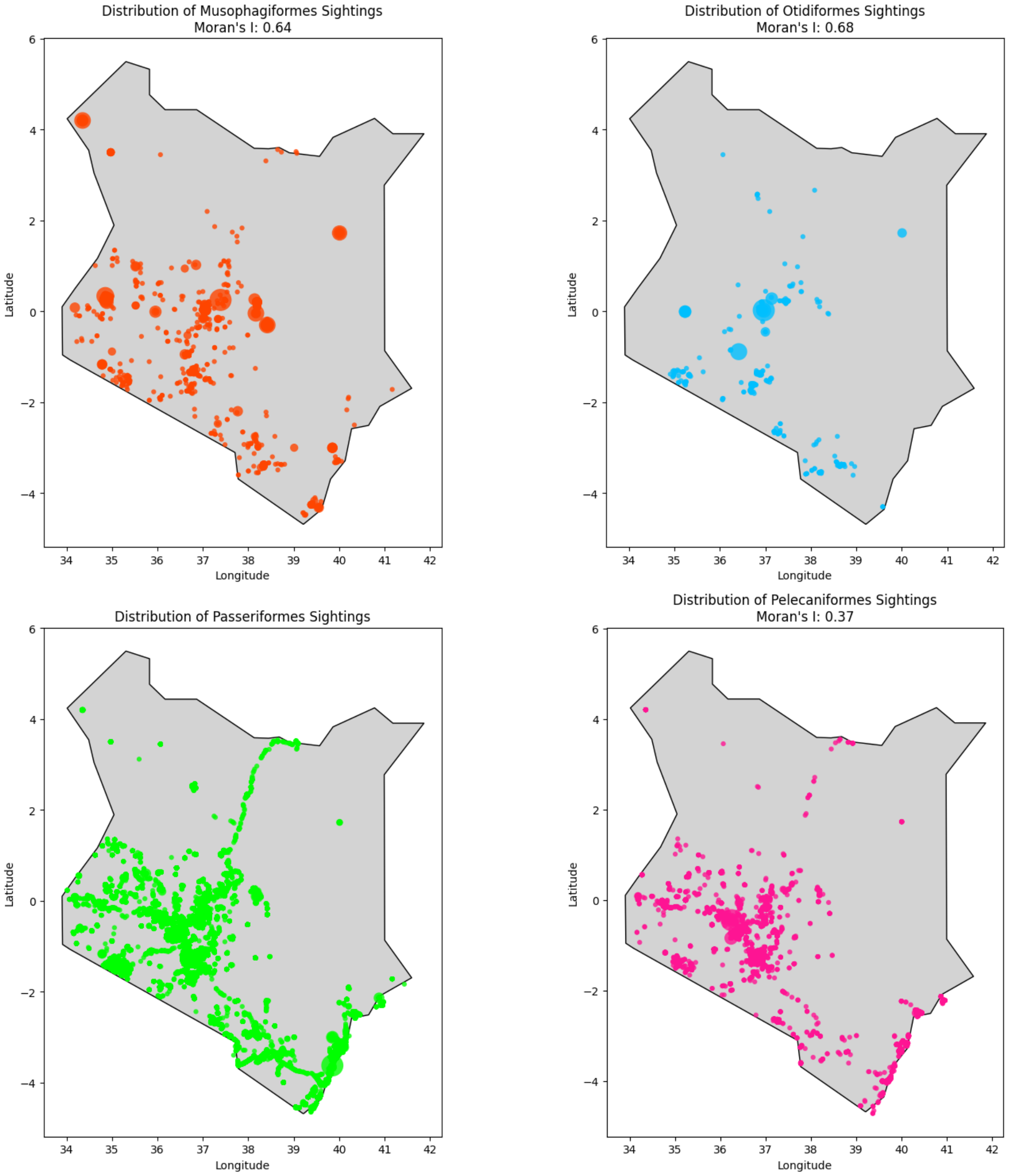

Distribution of Musophagiformes Sightings
Moran's I: 0.64
Distribution of Otidiformes Sightings
Moran's I: 0.68
Distribution of Passeriformes Sightings
Distribution of Pelecaniformes Sightings
Moran's I: 0.37
Latitude
Longitude
6
4
2
0
−2
−4
34
35
36
37
38
39
40
41
42

Figure 5: Distribution of Musophagiformes, Otidiformes, Passeriformes, Pelecaniformes

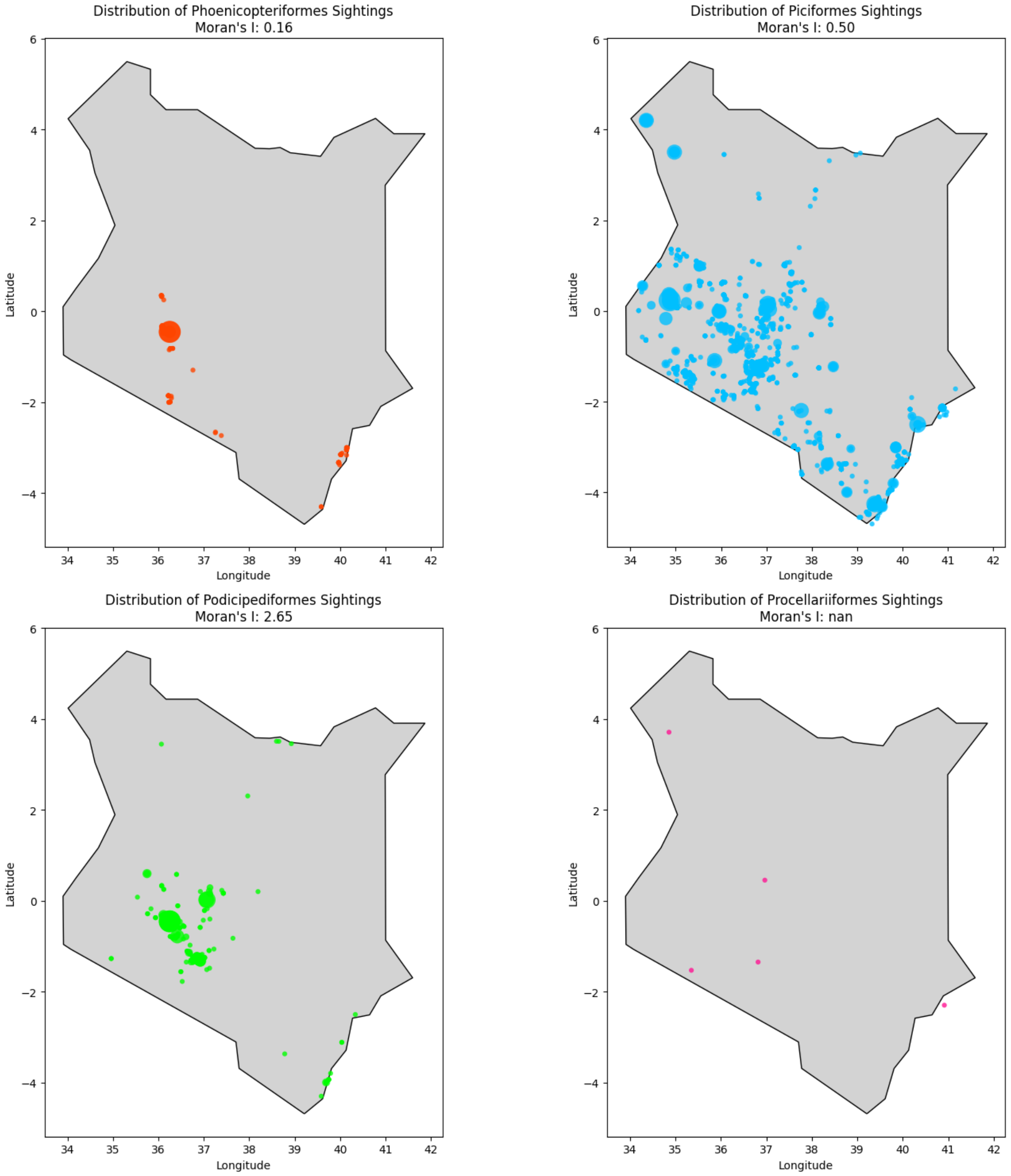
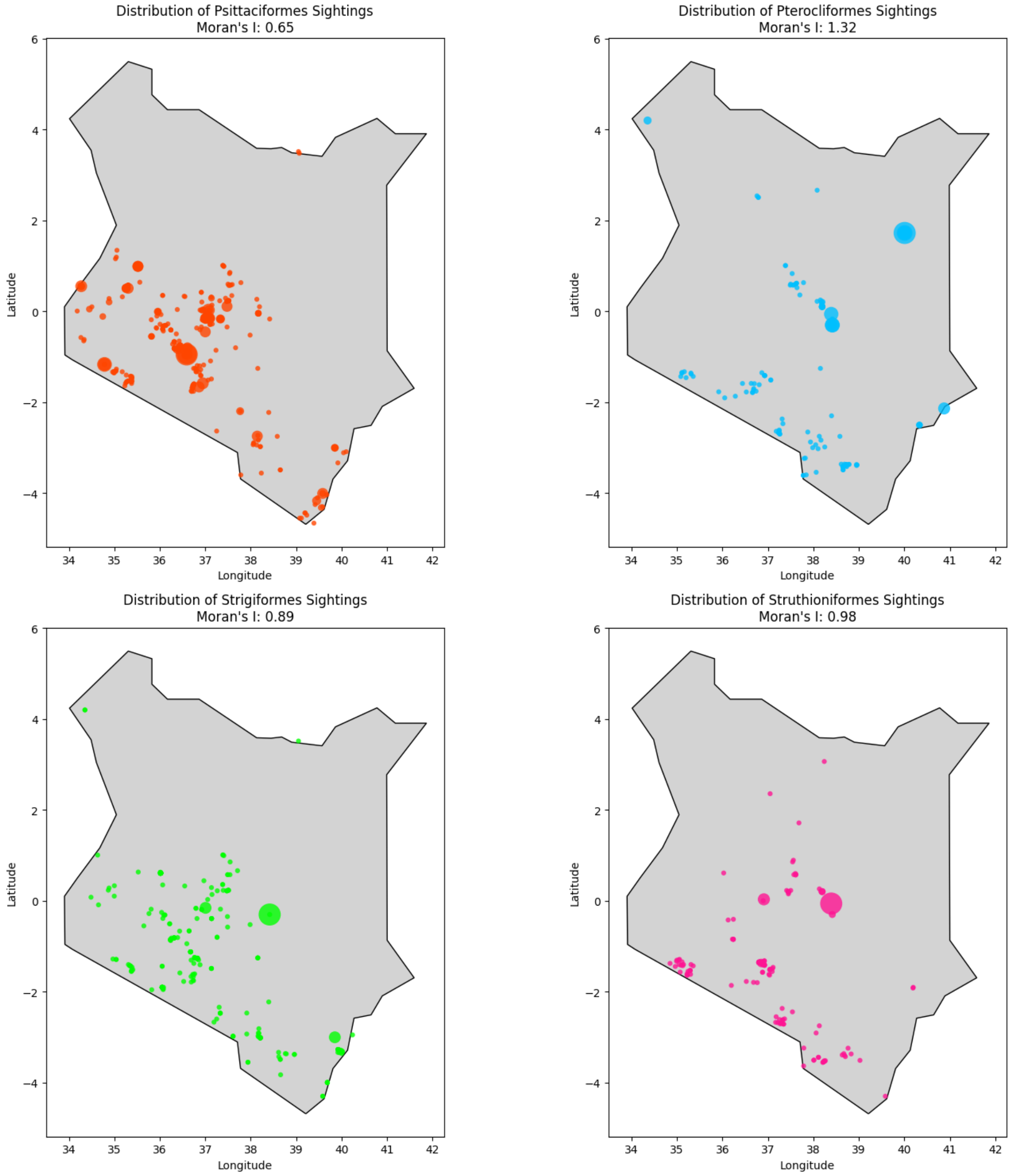

Distribution of Phoenicopteriformes Sightings
Moran's I: 0.16
Distribution of Piciformes Sightings
Moran's I: 0.50
Distribution of Podicipediformes Sightings
Moran's I: 2.65
Distribution of Procellariiformes Sightings
Moran's I: nan
Latitude
Longitude
6
4
2
0
−2
−4
34
35
36
37
38
39
40
41
42

Figure 6: Distribution of Phoenicopteriformes, Piciformes, Podicipediformes, Procellariiformes

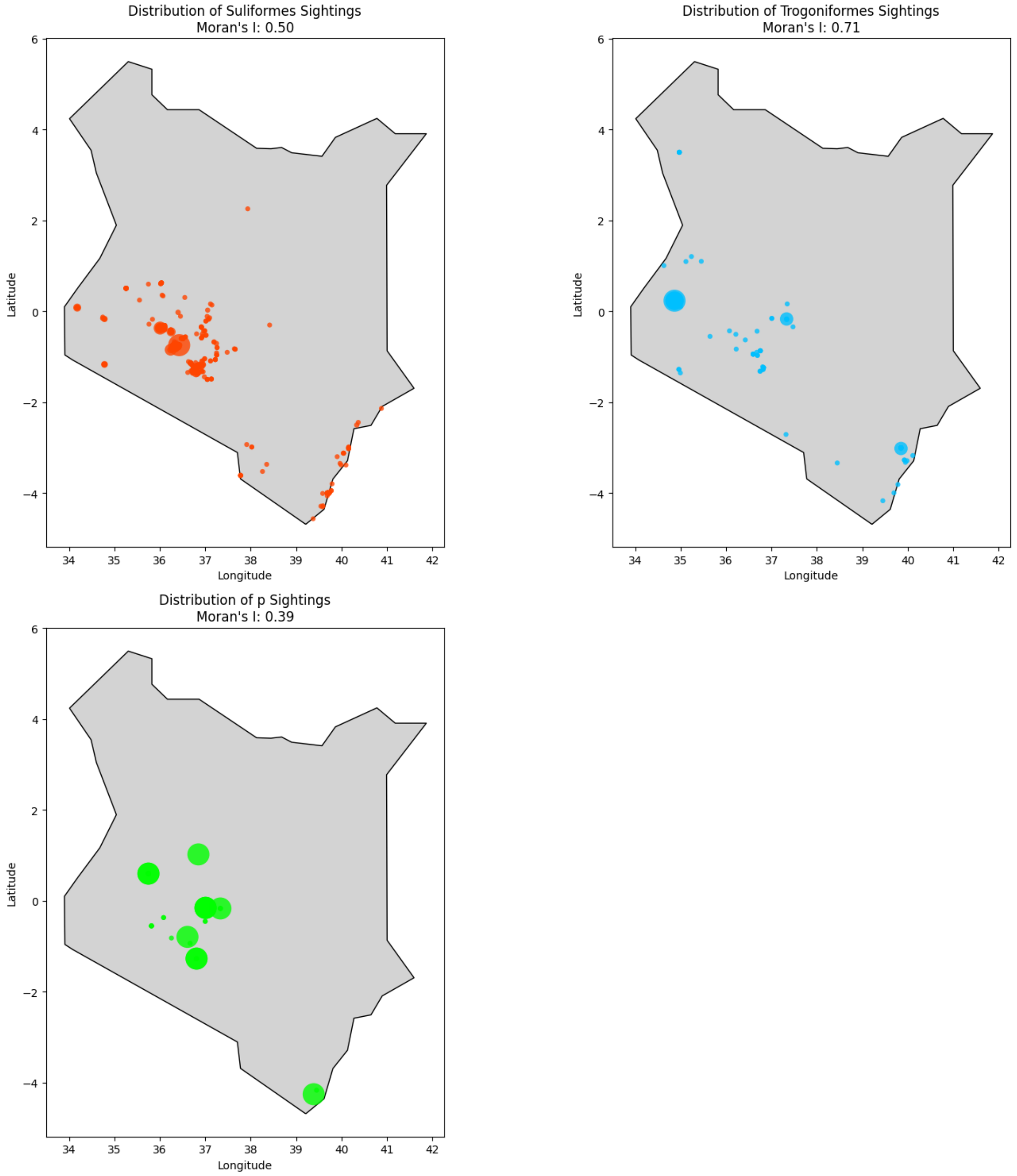

Distribution of Psittaciformes Sightings
Moran's I: 0.65

Distribution of Pterocliformes Sightings
Moran's I: 1.32

Distribution of Strigiformes Sightings
Moran's I: 0.89

Distribution of Struthioniformes Sightings
Moran's I: 0.98

Figure 7: Distribution of Psittaciformes, Pterocliformes, Strigiformes, Struthioniformes


Distribution of Suliformes Sightings
Moran's I: 0.50

Distribution of Trogoniformes Sightings
Moran's I: 0.71

Distribution of p Sightings
Moran's I: 0.39

Figure 8: Distribution of Suliformes, Trogoniformes, p

3 Clusters of birds sightings

For the following density maps, we proceeded the same way as for the plots of the previous chapter. The notion of orders is here no longer important and will be ignored. Instead of a scatter plot, a KDE-map is done. This shows us where most birds were observed. This map is giving us an estimate of the clusters, this means that it's normal that it extends the borders of Kenya (due to having data on the borders of the country). Don't focus too much on what is outside of Kenya. This goes out of the given data.

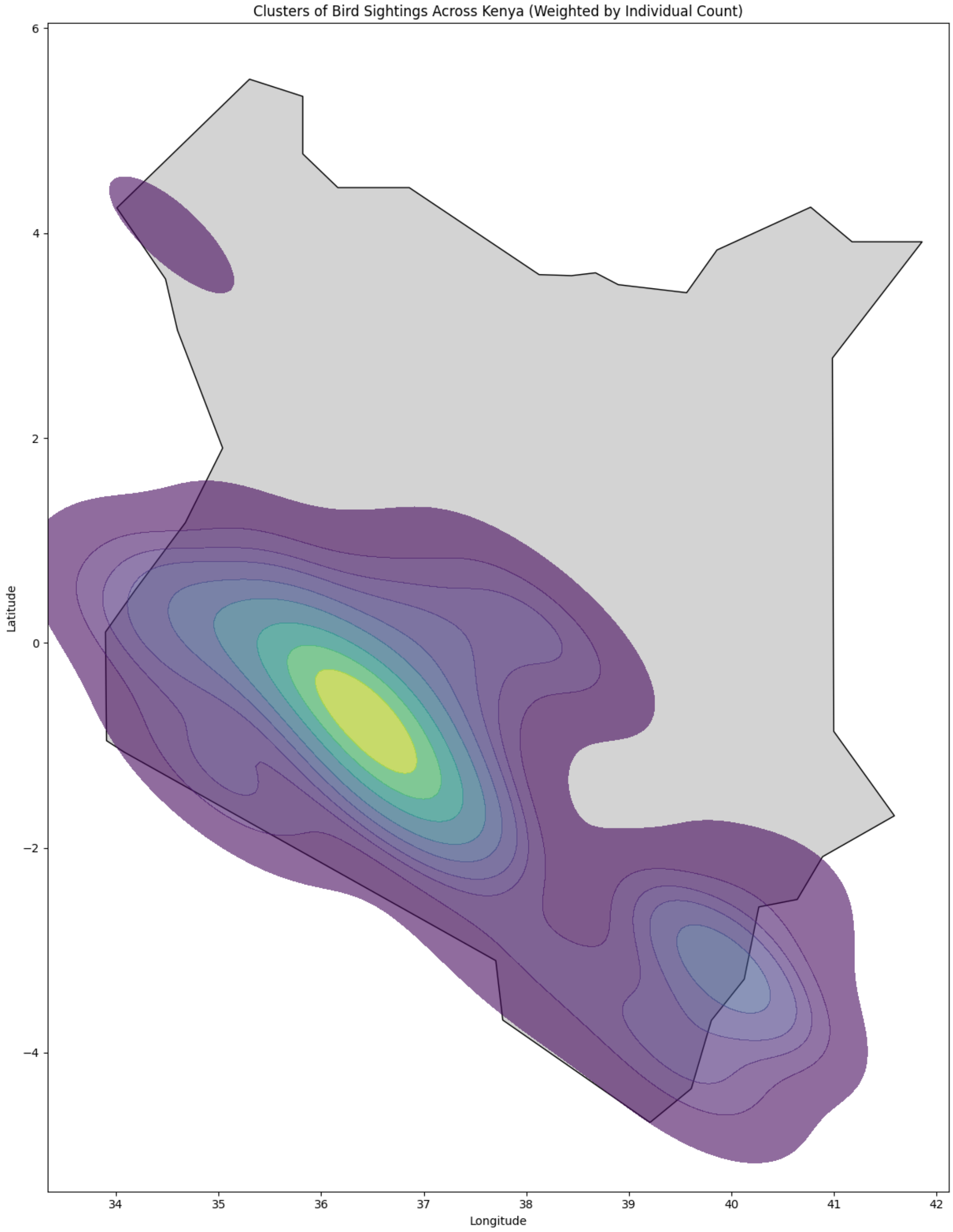

Clusters of Bird Sightings Across Kenya (Weighted by Individual Count)
6
4
2
0
−2
−4
Latitude
34
35
36
37
38
39
40
41
42
Longitude

Figure 9: Density map of bird observations

This second plot exists because in the data there are some rows that assume having counted over 100 birds at the same time. In our understanding, the counting is done by humans but there may very well be some kind of technology in play. There are also extreme values, notably one that counted a total of 7000 birds in one observation. The doubt rose whether or not these extreme values were manipulating the density and hiding other clusters of bird sightings. For this plot, all observations that assume having counted more than 20 birds have been dropped. The result is similar but the area seems to be preciser to what has been shown in the previous chapter.

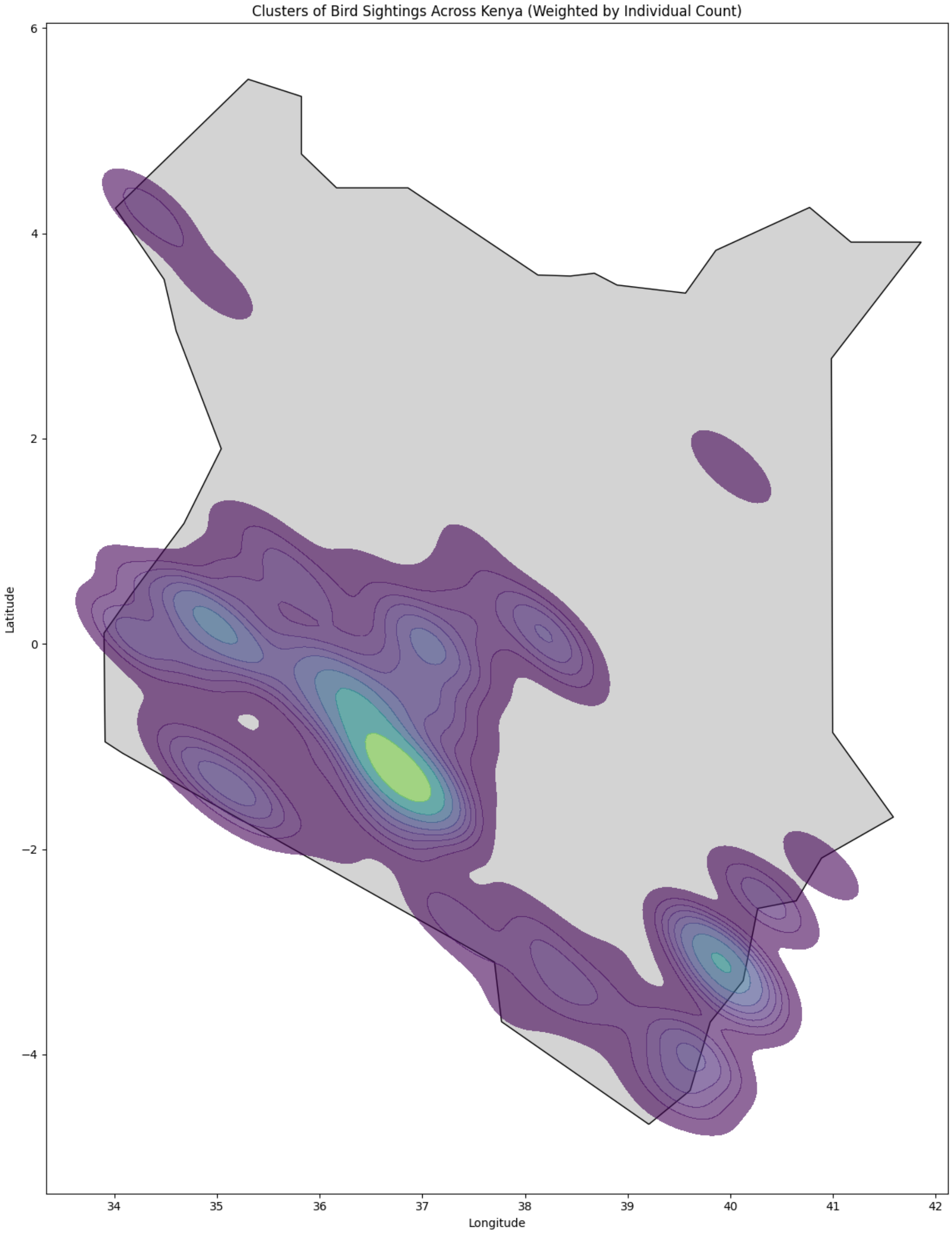

Clusters of Bird Sightings Across Kenya (Weighted by Individual Count)
Latitude
6
4
2
0
−2
−4
34
35
36
37
38
39
40
41
42
Longitude

Figure 10: Density map of bird observations without extreme values

4 Birds species richness map

In order to create a richness map for bird species across Kenya, we chose to divide our map according to the different official counties of the country. This required us to find a way to display them using Python implementation, as well as associate the corresponding county to each observation in our dataframe, based on the referenced coordinates. We managed to find a solution to this problem on the following GitHub repository: `https://github.com/Mondieki/kenya-counties-subcounties/tree/master`. It encloses data structured in JSON files, containing for each Kenyan county their respective code, capital, subcounties, multipolygon and polygon coordinates (in geoJSON format). Nevertheless, we had to make a few adjustments to exploit this data in our Python program.
First, we looped through each JSON file and extracted coordinates while creating a polygon based on them. We then converted this data to a GeoDataFrame, and joined it to the DataFrame previously provided. This way, we were able to identify the county in which each bird sighting happened.
We then obtained the following map:

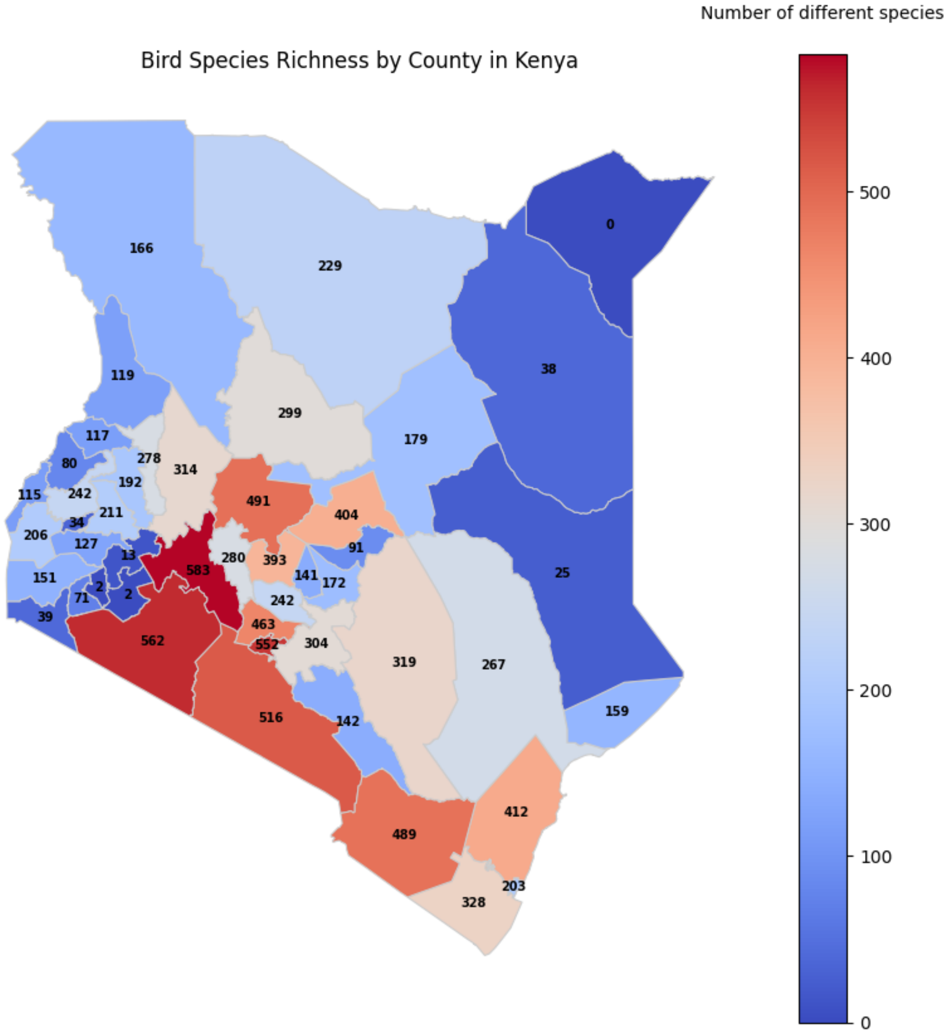


Figure 11: Birds species richness map for Kenya

We can observe that the counties with the highest recorded species richness are located in the center and the south of the country. We can also see that the two regions with the lowest species richness are the north east side and the south west corner of Kenya.

5 Final Discussions

Ultimately, we can see a pattern among the various clusters found during our different studies. Regardless of the quantity studied (Order distributions, birds sightings or the amount of different species in a given area), the maps we obtained show significant clusters in the same regions: the center and the south of the country. This leads us to think that there is a connection between all these locations and the bird biodiversity. We hypothesize that these "clusters of bird biodiversity" are related to the locations of national parks and natural reserves in Kenya.

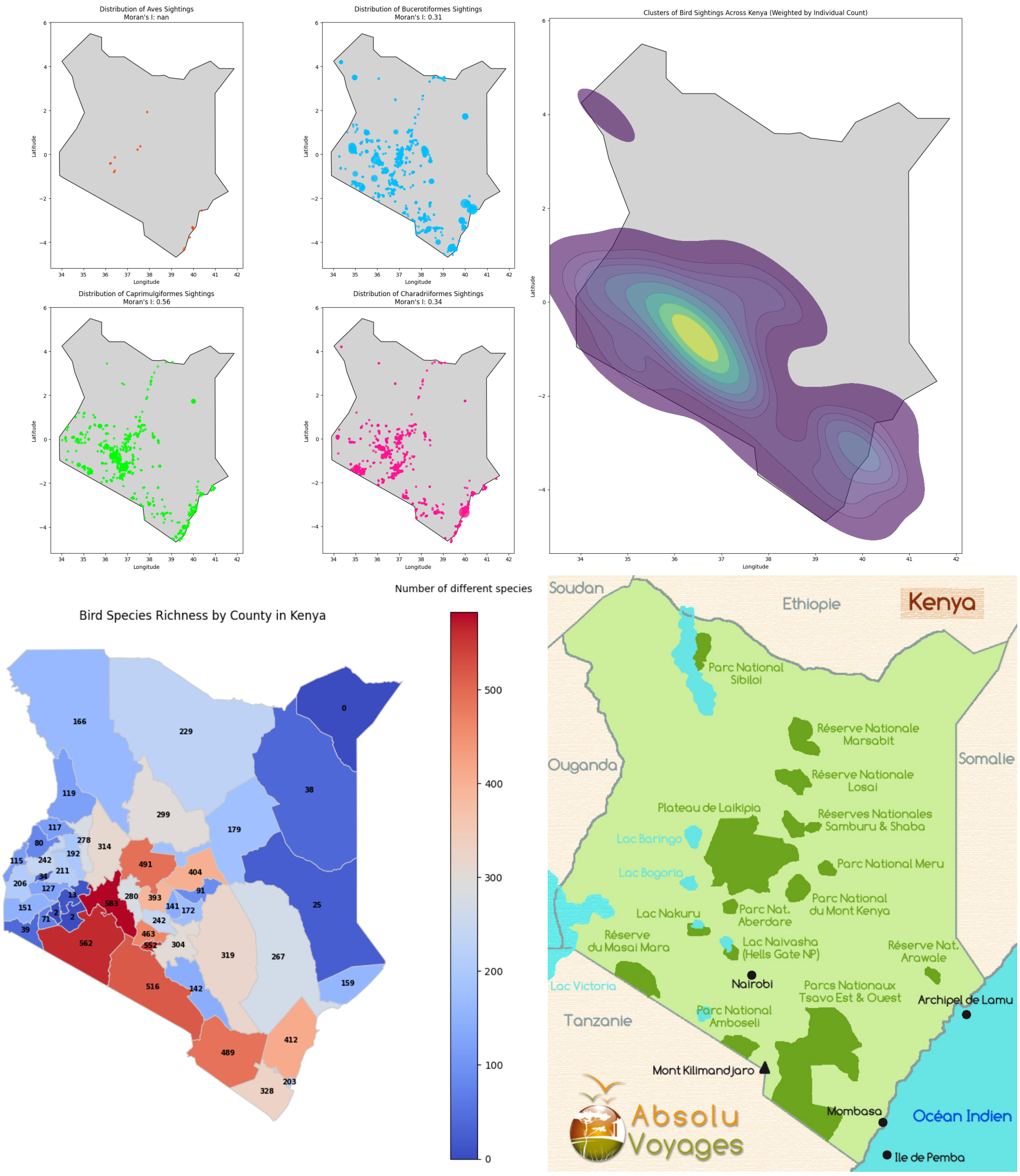

Distribution of Aves Sightings
Moran's I: nan
Distribution of Bucerotiformes Sightings
Moran's I: 0.31
Distribution of Caprimulgiformes Sightings
Moran's I: 0.56
Distribution of Charadriiformes Sightings
Moran's I: 0.34
Latitude
Longitude
Clusters of Bird Sightings Across Kenya (Weighted by Individual Count)
Bird Species Richness by County in Kenya
Number of different species
500
400
300
200
100
0
Soudan
Ethiopie
Kenya
Parc National Sibiloi
Réserve Nationale Marsabit
Ouganda
Réserve Nationale Losai
Somalie
Plateau de Laikipia
Réserves Nationales Samburu & Shaba
Lac Baringo
Parc National Meru
Lac Bogoria
Parc National du Mont Kenya
Lac Nakuru
Parc Nat. Aberdare
Réserve du Masai Mara
Lac Naivasha (Hells Gate NP)
Réserve Nat. Arawale
Lac Victoria
Nairobi
Parcs Nationaux Tsavo Est & Ouest
Archipel de Lamu
Tanzanie
Parc National Amboseli
Mont Kilimandjaro
Absolu Voyages
Mombasa
Océan Indien
Ile de Pemba

Figure 12: Comparison between "bird biodiversity clusters" and position of natural parks (cf. last plot) [1]

Comparing the position of natural parks with the clusters found on our maps, we can see that our hypothesis has a ground to stand on. Nevertheless, we cannot be fully certain of the correctness of our theory, given that we do not have any additional information about the data collection process of our data set. For example, we can imagine that some birds simply have not been observed even though they were present, which means that a lack of observations in a specific region does not correlate a lack of birds there, but maybe there were not enough people on site at the time.

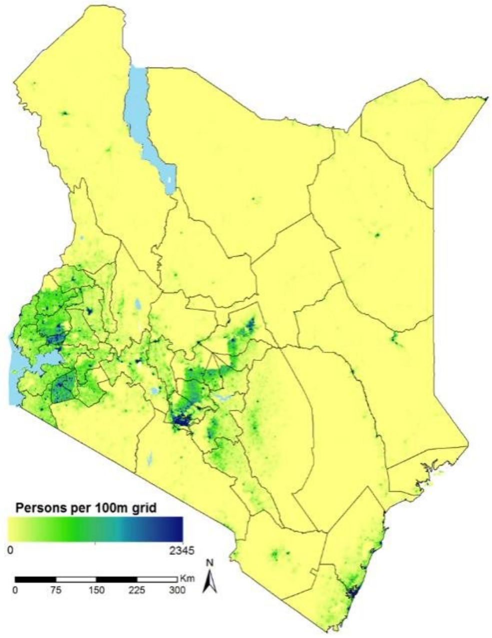


Figure 13: Density of the Kenyan population [2]

Indeed, if we look at the density of population in Kenya, we can see that the north of the country is far less populated that the south, which could potentially explain the lack of bird observations in this region.
However, we can only theorize about these possibilities, since we cannot be certain that people did not travel around the country to specifically collect this data.

References

[1] Absolu Voyages, `http://www.absolu-voyages.com/kenya.htm`.

[2] *The epidemiology and control profile of malaria in Kenya: reviewing the evidence to guide the future vector control*, Abdisalan Noor, Peter M. Macharia, Paul Ouma, Stephen Oloo, Joseph Maina, Ezekiel Gogo, David Kyalo, Lukio Olweny, Caroline Kabaria, Damaris Kinyoki, Robert W Snow, Ngozi Erondu, David Schellenberg, Rebecca Kiptui, Kiambo Njagi, Andrew Wamari, Christine Mbuli, Ahmed Deen Omar, Waqo Ejersa, 2016.