**FINAL DATA AND DOCUMENTATION**

We have produced two final datasets as CSVs:

1. **BirdDataSet\_1800\_2019 \_All\_Countries**
2. **InfluentialFactorsAffectingBirdPopulation**

The first dataset has various species of birds spotted from year 1800 up until 2019 in different countries. The second dataset however is more restricted to the years 1970 up until 2014 but includes other influential factors that might affect bird population. Both these datasets can be analysed separately to understand and research local bird counts, migratory birds and their patterns over the years. Hence these 2 final datasets are the outcome of this project.

**BIRDDATASET\_1800\_2019\_All\_Countries METADATA**

**DATA INCLUDED:**

This dataset contains the information of bird species and its count over 1800 to 2019 for different countries in the world.

**DATASET FIELDS:**

SPECIES – The character datatype specifies the list of different species spotted over the year for a country

COUNT – This character datatype specifies the list of last spotted bird count; first spotted bird count and highest spotted bird count and X represents the birds that were spotted but the count was unknown. Removing X will misrepresent the data in certain cases will not provide the specific bird that was spotted previously.

LOCATION – This character datatype specifies the list of bird spotted locations over the countries.

DATE – This character datatype specifies the list of dates that the bird was spotted in various countries.

COUNTRY – The character datatype specifies the two-character country codes for different countries as represented by ISO 3166 standard. For example, India is represented as “IN”

ATTRIBUTE – This character datatype specifies the category of birds spotted as First, Last and High

YEAR - This integer datatype specifies the list of year from 1970 to 2014 in which the birds were spotted in a particular country.

**Influential Factors Affecting Bird Population Metadata**

**DATA INCLUDED:**

This dataset includes the records of bird species that are spotted in various countries over the years that is been affected by the population, pollution and temperature of the countries.

**DATASET FIELDS:**

The below list includes all the dataset fields for the influential factors affecting bird population dataset file,

COUNTRY – The character datatype specifies the different country names.

COUNTRY\_CODE – The character datatype specifies the two-character country codes for different countries as represented by ISO 3166 standard. For example, New Zealand is represented as “NZ”.

SPECIES – The character datatype specifies the list of different species spotted over the year for a country

COUNT – This character datatype specifies the list of last spotted bird count; first spotted bird count and highest spotted bird count and X represents the birds that were spotted but the count was unknown. Removing X will misrepresent the data in certain cases will not provide the specific bird that was spotted previously.

LOCATION – This character datatype specifies the list of bird spotted locations over the countries.

DATE – This character datatype specifies the list of dates that the bird was spotted in various countries.

YEAR – This integer datatype specifies the list of year from 1970 to 2014 in which the birds were spotted in a particular country.

ATTRIBUTE – This character datatype specifies the category of birds spotted as First, Last and High

POPULATION\_COUNT – This integer datatype represents the population count of humans in the particular country.

CO2\_EMISSION – This double datatype specifies the CO2 emissions in tonnes.

AVERAGE\_TEMPERATURE – This double datatype specifies the average temperature of a country in Celsius

LATITUDE – This double datatype gives the list of latitudinal coordinates of a country.

LONGITUDE – This double datatype gives the list of longitudinal coordinates of a country.

**SAMPLING EVENT DATA:**

The dataset would be useful for other scientists or users to know the relation of birds’ population over the years for different countries which are affected by factors such as population, pollution and average temperature. This dataset would provide an insight to users to know more about the species and their extinction rate in relation to the above said factors. The variation in population, pollution and average temperature against bird population is obtained from the dataset.

**PROJECT LINK:**

The link to the final dataset related information is provided here,

<https://github.com/aarthy-badarakalimuthu/datawrangling/tree/master/Final%20DataSet>