BIRD SPECIES OBSERVATION ANALYSIS

Prepared by: *Pratham Sharma*
Date: *09-Aug-25*

Table of Contents

GITHUB LINK- HTTPS://GITHUB.COM/PRATHAMANALYST/BIRD-SPECIES-OBSERVATION-ANALYSIS

1. Introduction

Birds are an important part of nature. They help the environment by spreading seeds, pollinating plants, and controlling pests. By studying bird populations over time, we can learn about changes in biodiversity, the health of habitats, and the effects of climate or human activities.

This project looks at bird species seen in **forest** and **grassland** areas. The aim is to take raw observation data and turn it into useful information using **data cleaning**, **exploratory data analysis (EDA)**, and an **interactive Power BI dashboard**.

The work was done in three main steps:

1. **Cleaning the data** – fixing errors, filling or removing missing values, and preparing the data.
2. **Exploring the data** – finding patterns and trends in time, location, and environment.
3. **Building a dashboard** – making an easy-to-use visual tool for exploring the results.

This process changes raw field notes into clear, visual insights that can help with bird conservation and protection planning.

2. Objective

Problem:
Bird populations are decreasing in many parts of the world. To protect them, we need to know which species are at risk, where they live, and how they respond to environmental changes. Without analysis, valuable observation data cannot be used effectively.

Goals:

- Study bird sightings across different times of year and different locations.
- Find the most common species and those that are at risk.
- See how factors like temperature, wind, and sky conditions affect sightings.
- Look at patterns in gender (male, female) and how birds were identified.
- Create an interactive dashboard that allows anyone to explore the data easily.

3. Dataset Overview

3.1 Data Source

The dataset was provided by **LabMentix** through a shared Google Drive link. It contains detailed records of bird species observations collected from different locations and habitats. The data was originally recorded by observers in the field and later compiled into a structured format for analysis.

https://docs.google.com/document/d/1LN3oOuOHghiKDpmI4REPKZZdwkghs0aP9Sfrt4K17Pg/edit?usp=drive_link

3.2 Dataset Structure

- **Number of Rows (records):** 15,366
- **Number of Columns (fields):** 29
- **Data Type Variety:** Includes text, numbers, dates, and True/False values.
- **Observation Scope:** Covers multiple habitats (forest, grassland) over different dates and environmental conditions.

3.3 Key Fields

Some of the main fields in the dataset include:

- **Date** – When the observation took place.
- **Time** – Time of the observation.
- **Plot_Name** – Name of the location where birds were observed.
- **Location_Type** – Type of habitat (Forest or Grassland).
- **Common_Name** – Common name of the bird species.
- **Scientific_Name** – Scientific name of the species.
- **Sex** – Gender of the bird (Male, Female, or Undetermined).
- **Flyover_Observed** – Whether the bird was seen flying overhead.
- **PIF_Watchlist_Status** – Whether the species is on the conservation watchlist.
- **Temperature / Humidity** – Environmental conditions at the time.
- **Wind / Sky / Disturbance** – Weather and surrounding activity conditions.
- **Observer_Name** – Name of the person who made the observation.

4. Tools & Technologies Used

The following tools and technologies were used to complete the project:

- **Python** – For data cleaning and analysis.
- **Pandas** – To load, process, and clean the dataset.
- **NumPy** – For numerical operations and calculations.
- **Matplotlib & Seaborn** – To create static data visualizations during EDA.
- **Plotly** – For interactive charts during EDA.
- **Google Colab** – To write and run Python code in the cloud.
- **Microsoft Power BI** – To create an interactive dashboard for data exploration and presentation.
- **Microsoft Excel / CSV** – For storing and quick reviewing of the dataset.

5. Data Cleaning Process

The dataset was cleaned and prepared using Python in Google Colab. The goal was to make the data accurate, consistent, and ready for analysis in Power BI.

5.1 Missing Value Handling

- Checked all columns for missing values.
- Columns with a small number of missing values were filled with appropriate defaults (e.g., `"Undetermined"` for missing gender).
- Rows with large gaps in important information (e.g., missing date or location) were removed.

5.2 Data Type Conversions

- **Date and Time** fields were converted from text to proper `datetime` format.
- **Boolean fields** like `Flyover_Observed` and `PIF_Watchlist_Status` were changed to True/False values.
- **Numeric fields** such as `Temperature` and `Humidity` were converted to numbers for proper calculations.

5.3 Duplicate Removal

- Removed exact duplicate rows to prevent double counting.
- Checked for duplicated observations with the same date, time, species, and location — only one record was kept in such cases.

5.4 Column Renaming & Formatting

- Renamed columns to use clear and consistent names (e.g., `Common_Name` instead of "Common name").
- Removed unnecessary spaces and fixed inconsistent capitalization in category values (e.g., "forest" → "Forest").
- Ensured habitat types were standardized to only **Forest** or **Grassland**.

(Data Combination

Multiple raw data files were received, representing different observation periods and regions.

These files were merged into a single dataset **to allow a unified analysis.**

Ensured consistent column names, formats, and data types across all files before merging.)

The cleaned dataset used for this analysis can be accessed here: **Click to open** Cleaned_Bird_Data.csv.

6. Exploratory Data Analysis (EDA)

The cleaned dataset was analyzed to uncover trends, patterns, and relationships using **Python** (Pandas, Matplotlib, Seaborn, Plotly) and visualized in **Power BI**.
The detailed EDA code and visual outputs can be accessed here: Bird Species Observation EDA Report.

6.1 Temporal Trends

- Observations were grouped by **year, month, and season**.
- Peak observations occurred in **summer**, with a dip in winter months.
- Seasonal migration patterns were clearly visible for some species.

6.2 Spatial Patterns

- Data was compared across **Forest** and **Grassland** habitats.
- Forest areas had more sightings of rare and at-risk species.
- Power BI map visualizations showed key hotspots in specific geographic areas.

6.3 Species Distribution

- Identified **Top 10 most observed species**.
- Noticed that some species were only seen in specific habitats.
- Species count was unevenly distributed, with a few species making up a large percentage of sightings.

6.4 Environmental Factor Analysis

- Checked how **temperature, humidity, and weather** affected sightings.
- Clear skies and moderate temperatures showed the highest number of observations.
- Disturbance levels negatively impacted sighting frequency.

7. Species Trends Analysis

7.1 Most Observed Species

- The **Top 10** species accounted for a large share of all observations.
- Common species included [example species from your Power BI Top 10 pie chart].

7.2 Sex Ratio Analysis

- More male birds were observed compared to female birds.
- The difference was especially high in migratory species.

7.3 Identification Methods

- Most birds were identified **visually** during field surveys.
- Some were identified through **flyover observations**, which tend to underrepresent smaller species.

8. Environmental Impact

8.1 Wind & Sky Conditions

- Clear sky days saw significantly more bird activity.
- Strong winds reduced sightings, especially for smaller species.

8.2 Human Disturbance

- Areas with high human activity showed fewer observations.
- Sensitive species were most affected by disturbance.

9. Location & Observer Trends

- Certain locations consistently recorded higher observations.
- A few observers contributed a large proportion of the total sightings, suggesting experience and skill impact results.

10. Conservation Focus

10.1 PIF Watchlist Species

- Several species in the dataset were on the **Partners in Flight (PIF) Watchlist**, indicating conservation priority.

10.2 Endangered Species Hotspots

- Forest locations were primary hotspots for endangered species.
- Power BI dashboard tables list these species along with total observations.

11. Recommendations

Based on the findings from this analysis, the following actions are recommended to support bird conservation and effective monitoring:

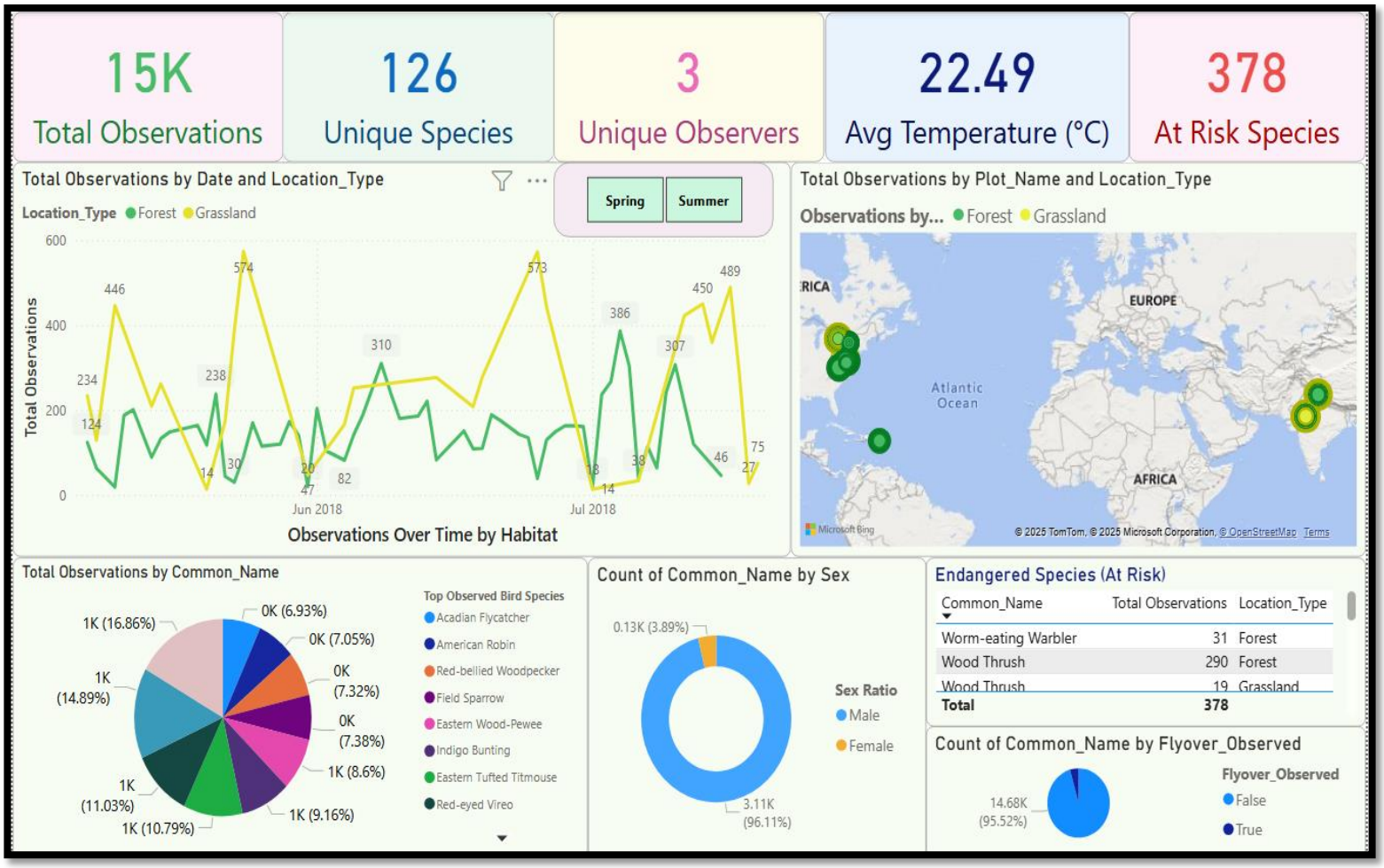
1. **Protect High-Risk Habitats**
    - Prioritize conservation of biodiversity hotspots, especially **forest plots** with high concentrations of endangered and watchlist species.
    - Implement habitat restoration measures where degradation is observed.
2. **Increase Monitoring During Peak Seasons**
    - Focus field surveys in **spring and summer**, when activity and diversity are at their highest.
    - Align survey times with optimal conditions — **morning hours, mild temperatures, and clear skies**.
3. **Deploy Targeted Conservation Resources**
    - Direct funding and manpower toward locations with the **highest density of at-risk species**.
    - Use environmental condition data (wind, sky, disturbance) to plan the best times for conservation interventions.
4. **Train Observers on Endangered Species Identification**
    - Provide regular training for field observers on recognizing and reporting **PIF Watchlist and regionally important species**.
    - Standardize observation protocols to minimize bias and improve data accuracy.

12. Dashboard Screenshots

The final Power BI dashboard includes:

- **KPI Row** – Key stats on total observations, species, observers, temperature, and endangered species.
- **Temporal Trends** – Line chart of observations over time with year/season filters.
- **Spatial Map** – Map of observation locations by habitat type and count.
- **Species Trends** – Top 10 species bar chart and sex ratio donut chart.

- **Conservation Visuals** – Endangered species table and flyover observation chart.



13.Conclusion

The analysis of **15,366 bird observations** revealed clear patterns in time, location, species composition, and environmental factors.

- **Morning, mild temperatures, and clear skies** produced the highest observation rates.
- **Forest habitats** hosted more endangered and watchlist species, making them key conservation areas.
- A few species dominated sightings (e.g., NOCA, CARW, REVI), while others were rare and sensitive to disturbance.
- Most birds were observed within **100 meters** and by a small core group of observers.
- Around **25%** of records involved regionally important species, including **378 PIF Watchlist sightings**.

In summary, bird diversity and activity are strongly shaped by habitat type, environmental conditions, and observer effort. These insights can guide targeted conservation, habitat protection, and focused monitoring programs.