

Project Title	Bird Species Observation Analysis
Skills take away From This Project	Data Cleaning and Preprocessing, Exploratory Data Analysis (EDA), Data Visualization, Geographic Analysis ,Species Analysis, SQL, Streamlit or PowerBI
Domain	Environmental Studies, Biodiversity Conservation, and Ecology

Problem Statement:

The project aims to analyze the distribution and diversity of bird species in two distinct ecosystems: forests and grasslands. By examining bird species observations across these habitats, the goal is to understand how environmental factors, such as vegetation type, climate, and terrain, influence bird populations and their behavior. The study will involve working on the provided observational data of bird species present in both ecosystems, identifying patterns of habitat preference, and assessing the impact of these habitats on bird diversity. The findings can provide valuable insights into habitat conservation, biodiversity management, and the effects of environmental changes on avian communities.

Business Use Cases:

1. **Wildlife Conservation:** Inform decisions on protecting critical bird habitats and enhancing biodiversity conservation efforts.
2. **Land Management:** Optimize land use and habitat restoration strategies by understanding the preferences of different bird species.
3. **Eco-Tourism:** Identify bird-rich areas to develop bird-watching tourism, attracting eco-tourists and boosting local economies.

4. **Sustainable Agriculture:** Support the development of agricultural practices that minimize the impact on bird populations in grasslands and forests.
5. **Policy Support:** Provide data-driven insights to help environmental agencies create effective conservation policies and strategies for vulnerable bird species.
6. **Biodiversity Monitoring:** Track the health and diversity of avian populations, aiding in the monitoring of ecosystem stability.

Approach:

1. Data Cleaning and Preprocessing

- Handle missing data and standardize observational metrics.
- Filter relevant columns for analysis (e.g., species, environmental factors, temporal data).
- Consolidate data from forest and grassland units into comparable formats.

2. Exploratory Data Analysis (EDA)

- Analyze the distribution of species across administrative units and habitat types.
- Study observation frequency by year, month, and season.
- Investigate relationships between environmental conditions (e.g., temperature, humidity) and bird activity.

Types of Analysis- Examples:

1. Temporal Analysis

- **Seasonal Trends:** Analyze the `Date` and `Year` columns to detect patterns in bird sightings across different seasons or years.
- **Observation Time:** Study the `Start_Time` and `End_Time` to determine if specific time windows correlate with higher bird activity.

2. Spatial Analysis

- **Location Insights:** Group data by `Location_Type` (e.g., Grassland) to identify biodiversity hotspots.
 - **Plot-Level Analysis:** Compare observations across different `Plot_Name` to see which plots attract more species or specific kinds of birds.
-

3. Species Analysis

- Diversity Metrics: Count unique species (**Scientific_Name**) observed and their distribution across **Location_Type**.
 - Activity Patterns: Check the **Interval_Length** and **ID_Method** columns to identify the most common activity types (e.g., Singing).
 - Sex Ratio: Analyze the **Sex** column to understand the male-to-female ratio for different species.
-

4. Environmental Conditions

- Weather Correlation: Explore how **Temperature**, **Humidity**, **Sky**, and **Wind** impact observations, such as the number of birds or their distances.
 - Disturbance Effect: Assess the impact of **Disturbance** (e.g., slight effect) on bird sightings.
-

5. Distance and Behavior

- Distance Analysis: Evaluate the **Distance** column to identify species typically observed closer or farther from the observer.
 - Flyover Frequency: Examine the **Flyover_Observed** column to detect trends in bird behavior during observation.
-

6. Observer Trends

- Observer Bias: Analyze data by **Observer** to check if specific individuals report more observations or certain species.
 - Visit Patterns: Evaluate the **Visit** column to see how repeated visits affect species count or diversity.
-

7. Conservation Insights

- Watchlist Trends: Use the **PIF_Watchlist_Status** and **Regional_Stewardship_Status** to identify trends in species that are at risk or require conservation focus.
- AOU Code Patterns: Study the distribution of species based on their **AOU_Code** to correlate with regional or national conservation priorities.

3. Visualization

- Create interactive visualizations using Plotly in Streamlit or Power BI.
- Dynamic scatter plots and bar charts for species distributions.
- Temporal heatmaps for year-wise and month-wise observations.
- Geographic mapping (if location data is available) to highlight high-activity zones.
- Explore specific species or environmental conditions.

4. Other Insights

- Identify high-activity regions and seasons for specific bird species.
- Uncover the influence of environmental factors on species behavior and activity.
- Highlight at-risk species and conservation priorities for targeted efforts.

Expected Results:

By the end of this project, learners should be able to:

1. Derive insights into bird species' temporal and spatial distribution across forest and grassland habitats.
2. Visualize species activity trends using Streamlit and Plotly or Power BI dashboards with interactive features for user-friendly exploration.
3. Provide actionable findings for ecological conservation planning and resource allocation.

Project Evaluation metrics:

- Data Preparation
 - Completeness and accuracy of data cleaning.
 - Clear documentation of preprocessing steps.
- Exploratory Data Analysis
 - Depth and clarity of insights derived.
 - Use of relevant statistical methods.
- Data Visualization
 - Relevance and quality of visualizations.
 - Effective use of charts, maps, and interactivity.
- Business Insights
 - Actionable and stakeholder-focused findings.
 - Addressing key use cases like hotspots and trends.
- Presentation
 - Logical structure and clarity.
 - Effective communication of results.

Technical Tags:

- Data Cleaning
- Data Preprocessing
- Exploratory Data Analysis (EDA)
- Data Visualization- Plotly
- Power BI

Data Set:

Dataset: [Bird Observation DataSet](#)

How to read multiple Excel sheets- Refer this link-

[Read multiple excel sheets in pandas.ipynb](#)

Data Set Explanation:

The dataset contains observational data for bird species recorded across multiple forest sites. It includes detailed columns describing location, observation methods, bird species, and environmental conditions.

1. Admin_Unit_Code: The code for the administrative unit (e.g., "ANTI") where the observation was conducted.
2. Sub_Unit_Code: The sub-unit within the administrative unit for further classification.
3. Site_Name: The name of the specific observation site within the unit.
4. Plot_Name: A unique identifier for the specific plot where observations were recorded.
5. Location_Type: The habitat type of the observation area (e.g., "Forest").
6. Year: The year in which the observation took place.
7. Date: The exact date of the observation.
8. Start_Time: The start time of the observation session.
9. End_Time: The end time of the observation session.
10. Observer: The individual who conducted the observation.
11. Visit: The count of visits made to the same observation site or plot.
12. Interval_Length: The duration of the observation interval (e.g., "0-2.5 min").
13. ID_Method: The method used to identify the species (e.g., "Singing," "Calling," "Visualization").
14. Distance: The distance of the observed species from the observer (e.g., "<= 50 Meters").

15. Flyover_Observed: Indicates whether the bird was observed flying overhead (TRUE/FALSE).
16. Sex: The sex of the observed bird (e.g., Male, Female, Undetermined).
17. Common_Name: The common name of the observed bird species (e.g., "Eastern Towhee").
18. Scientific_Name: The scientific name of the observed bird species (e.g., *Pipilo erythrophthalmus*).
19. AcceptedTSN: The Taxonomic Serial Number for the observed species.
20. NPSTaxonCode: A unique code assigned to the taxon of the species.
21. AOU_Code: The American Ornithological Union code for the species.
22. PIF_Watchlist_Status: Indicates whether the species is on the Partners in Flight Watchlist (e.g., "TRUE" for at-risk species).
23. Regional_Stewardship_Status: Denotes the conservation priority within the region (TRUE/FALSE).
24. Temperature: The temperature recorded at the time of observation (in degrees).
25. Humidity: The humidity percentage recorded at the time of observation.
26. Sky: The sky condition during the observation (e.g., "Cloudy/Overcast").
27. Wind: The wind condition (e.g., "Calm (< 1 mph) smoke rises vertically").
28. Disturbance: Notes any disturbances that could affect the observation (e.g., "No effect on count").
29. Initial_Three_Min_Cnt: The count of the species observed in the first three minutes of the session.

Sheets Information:

The Excel file contains multiple sheets representing different administrative units, with their codes matching the **Admin_Unit_Code** column:

- **ANTI**: Data for the Antietam National Battlefield.
- **CATO**: Data for the Catoclin Mountain Park.
- **CHOH**: Data for the Chesapeake and Ohio Canal National Historical Park.
- **GWMP**: Data for the George Washington Memorial Parkway.
- **HAFE**: Data for Harpers Ferry National Historical Park.
- **MANA**: Data for the Manassas National Battlefield Park.
- **MONO**: Data for the Monocacy National Battlefield.
- **NACE**: Data for the National Capital East Parks.
- **PRWI**: Data for the Prince William Forest Park.
- **ROCR**: Data for the Rock Creek Park.
- **WOTR**: Data for the Wolf Trap National Park for the Performing Arts.

Each sheet contains similar columns but specific data for the respective administrative unit.

Project Deliverables:

Cleaned Dataset:

- Final preprocessed dataset used for analysis.
- Description of cleaning steps (e.g., handling missing values, formatting).

Source Code:

- Python or other scripts used for data cleaning, analysis, and visualization.
- Well-commented code for easy understanding.

Interactive Dashboard:

- Streamlit Application or Power BI dashboard showcasing insights with filters, maps, and charts.

Documentation:

- A concise report explaining the approach, key findings, and actionable insights.
- Include explanations for visualizations, trends, and analyses performed.

Project Guidelines:

Data Handling:

- Clean and preprocess the dataset thoroughly.
- Handle missing values, duplicates, and outliers properly.

Analysis Approach:

- Use a structured process: Cleaning → EDA → Visualization → Insights.

Visualizations:

- **Store data in an SQL database for visualization.**
- Use clear and relevant charts/maps for insights.
- Ensure visualizations are interactive.

Documentation and Reporting:

- Document steps, findings, and insights clearly.
- Use simple, stakeholder-friendly language in reports.

Code Quality:

- Write clean, well-commented, and reusable code.

Presentation:

- Keep the final presentation clear and concise.
- Emphasize key findings and actionable recommendations.

Timeline:

The project must be completed and submitted **within 7 days from the assigned date.**

References:

Project Orientation Recording (Tamil)	EDA Project Session Recordings (TAMIL)(20/12/2024):
Project Orientation Recording (English)	Bird species orientation English
Project Live Evaluation	Project Live Evaluation
EDA Guide	Exploratory Data Analysis (EDA) Guide
Streamlit with Plotly Reference	https://docs.streamlit.io/develop/api-reference/charts/st.plotly_chart
Capstone Explanation Guideline	Capstone Explanation Guideline

GitHub Reference	<u>How to Use GitHub.pptx</u>
PowerBi Session Recording Link	<u>PowerBi by Lokesh</u>
Steps to Create a Free Power BI Work Account	<u>Steps to Create a Free Power BI Work Account</u>

PROJECT DOUBT CLARIFICATION SESSION (PROJECT AND CLASS DOUBTS)

About Session: The Project Doubt Clarification Session is a helpful resource for resolving questions and concerns about projects and class topics. It provides support in understanding project requirements, addressing code issues, and clarifying class concepts. The session aims to enhance comprehension and provide guidance to overcome challenges effectively.

Note: Book the slot at least before 12:00 Pm on the same day

Timing: Monday to Saturday (4:00PM to 5:00PM)

Booking link : <https://forms.gle/XC553oSbMJ2Gcfug9>

LIVE EVALUATION SESSION (CAPSTONE AND FINAL PROJECT)

About Session: The Live Evaluation Session for Capstone and Final Projects allows participants to showcase their projects and receive real-time feedback for improvement. It assesses project quality and provides an opportunity for discussion and evaluation.

Note: This form will Open on Saturday and Sunday Only on Every Week

Timing: Monday-Saturday (5:30PM to 6:30PM)

Booking link : <https://forms.gle/1m2Gsro41fLtZurRA>