

Bird Species Observation Analysis — Final Report

1. Business Objective

The goal of this project was to analyze bird observation data from Antietam National Battlefield to:

- Identify when and where bird activity is highest,
- Understand which habitats and species are most critical for conservation, and
- Determine how environmental factors and survey methods affect data quality, so that park managers and conservation teams can optimize survey schedules, allocate resources effectively, and prioritize habitats and species of high ecological value.

2. Data Preparation

Datasets Used:

- Bird_Monitoring_Data_FOREST.XLSX
- Bird_Monitoring_Data_GRASSLAND.XLSX

Steps Performed:

1. Data Loading & Exploration (Step 1)

- Checked column meanings, formats, and missing values.
- Found duplicates (458 in grassland data) and inconsistent formats in numeric columns.

2. Data Cleaning & Preprocessing (Step 2)

- Merged forest and grassland data into a single standardized DataFrame.
- Fixed date/time formats, ensured numeric types for temperature, humidity, distance.
- Handled missing values (filled or marked as "Unknown"), removed duplicates.
- Saved cleaned dataset as Cleaned_Bird_Data.csv.

3. Exploratory Data Analysis (Step 3–4)

Key Findings:

- Temporal Patterns:

- Data available only for 2018 (3,327 records).
- Bird activity peaked in summer (2,081 sightings) vs. spring (1,246 sightings).

- Spatial Patterns:

- All records from Antietam National Battlefield (ANTI).
- 81 unique species recorded.
- Grasslands hosted 78 species vs. forests with 46 species.
- Certain plots (ANTI0105, ANTI0009, ANTI0077) were biodiversity hotspots.

- Species & Observer Trends:

- A few species (*Ammodramus savannarum*, *Agelaius phoeniceus*) dominate sightings.
- Three observers contributed ~100% of records, with varying effort levels.
- Sex bias toward males (detected mostly through singing).

- Environmental Insights:

- Bird sightings peaked at 20–25 °C and 60–70% humidity.
- Disturbance and poor weather reduced observations.

- Conservation Priorities:

- 2 species on PIF Watchlist: *Hylocichla mustelina*, *Setophaga discolor*.
- 15 stewardship species needing habitat protection.

4. Data Visualization (Step 5)

Tools Used:

- Plotly & Seaborn: Created 18 detailed charts.
- Streamlit: Built an interactive dashboard with filters, a species search tool, and CSV download options.

Visualizations Created (18):

1. Temporal heatmap (sightings by month/year)
2. Top 15 species (bar chart)
3. Sightings by habitat type (pie chart)
4. Sex distribution (pie chart)
5. Scatter plot — temperature vs. humidity vs. sightings
6. Sightings by park (bar chart)
7. Monthly sightings trend (line chart)
8. Observation method distribution (bar chart)
9. Flyover vs. non-flyover (pie chart)
10. Disturbance impact (bar chart)
11. Top observers (horizontal bar chart)
12. Top plots by species diversity (horizontal bar chart)
13. Sightings by temperature (line chart)
14. Sightings by humidity (line chart)
15. Species count by conservation status (bar chart)
16. Observation interval length (bar chart)
17. Distance distribution (histogram)
18. Correlation heatmap & pair plot (numerical variables)

5. Recommendations to Achieve Business Objectives

1. Optimize survey timing — conduct more surveys during summer and mild weather for maximum data quality.
2. Focus on high-diversity plots — allocate conservation resources to hotspots (e.g., ANTI0105, ANTI0009).
3. Prioritize grassland habitats — higher species diversity than forests.
4. Improve observer training — reduce male detection bias and enhance sex identification accuracy.
5. Protect watchlist & stewardship species — targeted monitoring and habitat protection.
6. Leverage environmental data — predict bird activity to schedule surveys efficiently.

6. Conclusion

This project successfully combined forest and grassland bird datasets, cleaned and standardized the data, performed in-depth exploratory analysis, and built an interactive dashboard to visualize trends.

The findings revealed strong seasonal patterns, habitat-specific diversity, observer biases, and environmental factors affecting bird sightings. These insights enable better resource allocation, improved survey protocols, and focused conservation actions on high-value habitats and species.

The Streamlit dashboard allows stakeholders to dynamically explore species information, temporal activity, environmental correlations, and download filtered data — ensuring the results remain actionable and scalable for future monitoring.