Project- Bird Species observation analysis (Report and final summary)

Project Summary

The bird monitoring analysis project aims to understand bird species distribution, behavioral patterns, and environmental dependencies within forest and grassland habitats. The unified and cleaned dataset—derived from two distinct habitat-specific files—provides insights into observation frequency, species diversity, and environmental correlations using exploratory data analysis (EDA) and Power BI visualization tools.

Problem Statement

There is a critical need to study bird activity across varying habitats to assess biodiversity and understand how external variables like time, temperature, and habitat type influence species detection. The main challenge is to merge fragmented datasets and extract reliable patterns that can inform conservation strategies, identify observation biases, and optimize future bird monitoring efforts.

Methodologies and Tools Used

1. Python (Pandas, Seaborn, Matplotlib) for data cleaning and exploratory analysis.
2. Power BI for interactive dashboard visualizations.
3. Excel for initial raw dataset storage and formatting.

Approach

1. Data Preprocessing

The original datasets ('Bird_Monitoring_Data_FOREST.XLSX' and 'Bird_Monitoring_Data_GRASSLAND.XLSX') were appended after adding a new 'Habitat' column to distinguish between forest and grassland observations. This allowed unified analysis across habitats while maintaining contextual clarity.

2. Data Cleaning

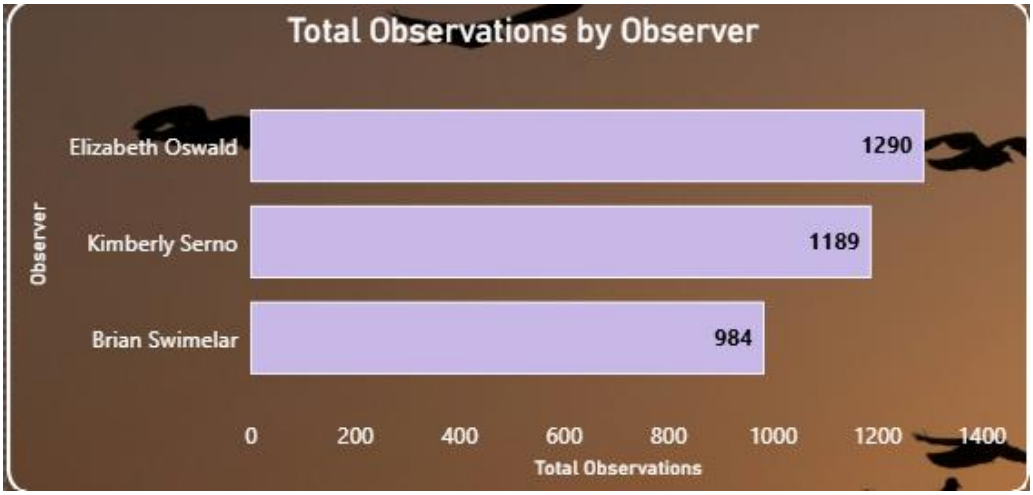
Columns were renamed for consistency, missing values were identified and handled (e.g., 'Unknown' for missing identification methods), and date/time formats were standardized. Duplicate rows were removed, and derived fields such as hour of observation were created.

3. Exploratory Data Analysis (EDA)

EDA revealed temporal observation trends, habitat-specific diversity, and observer activity distributions. Correlation analysis was performed to evaluate environmental influences on bird detection.

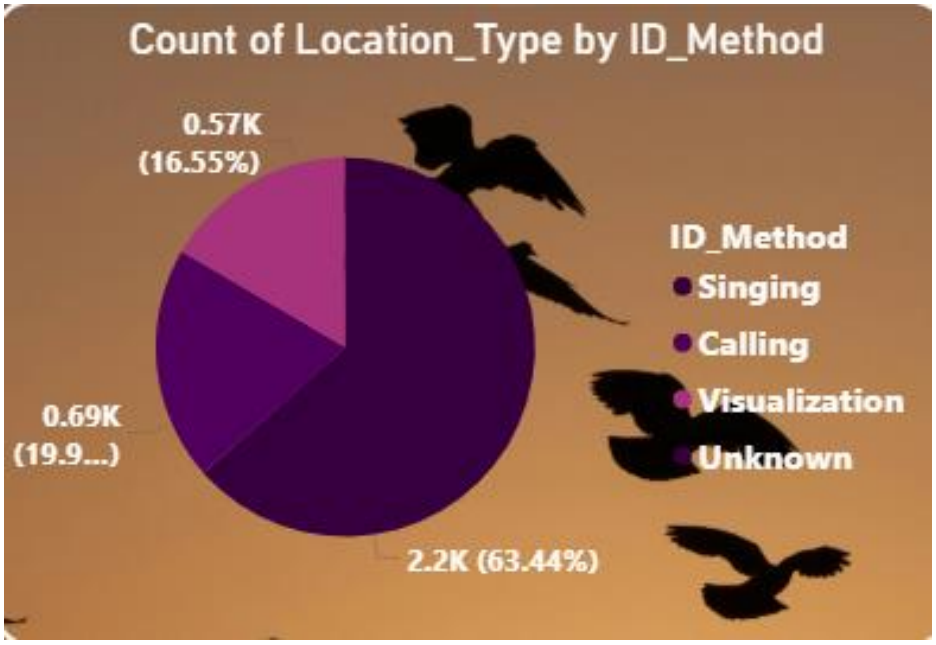
4. Data Visualization

Key visuals from Power BI and Python-based EDA are embedded below:

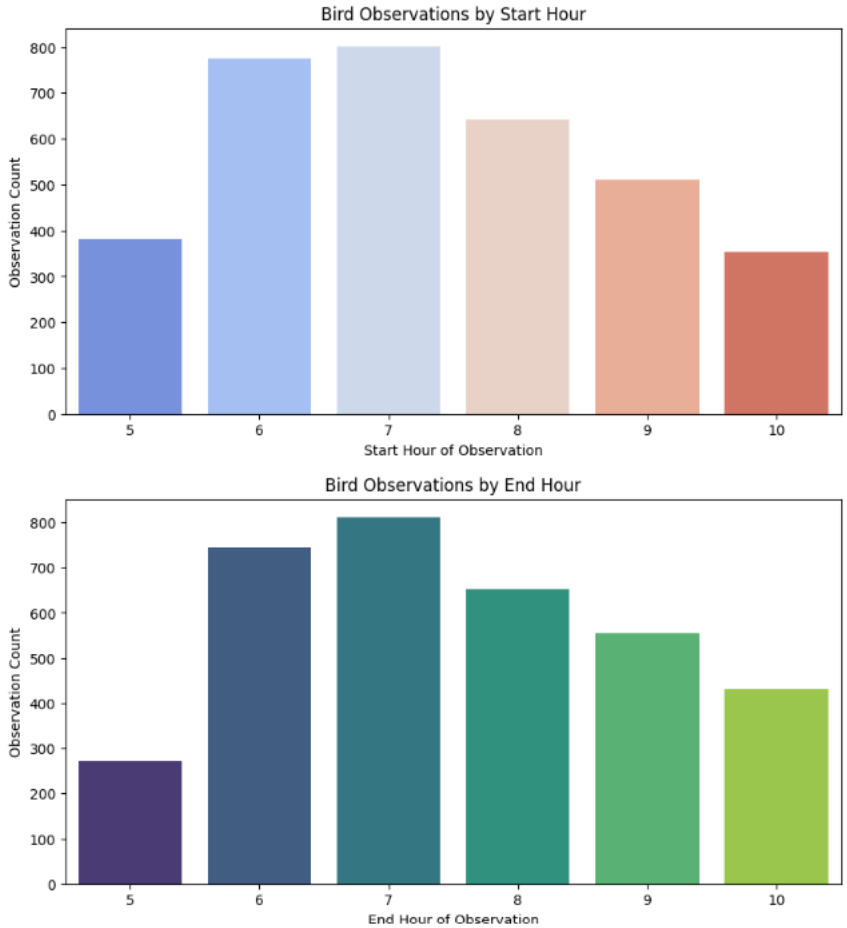


Findings- Elizabeth Oswald recorded the highest number of observations (1290), followed by Kimberly Serno (1189) and Brian Swimelar (984)

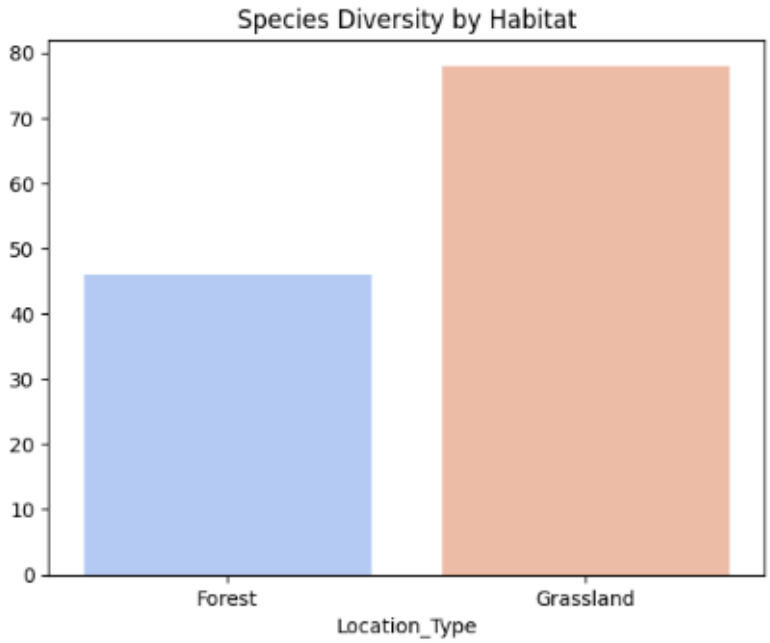
Findings- Spizella pusilla and Cardinalis cardinalis were most frequently observed across all habitats



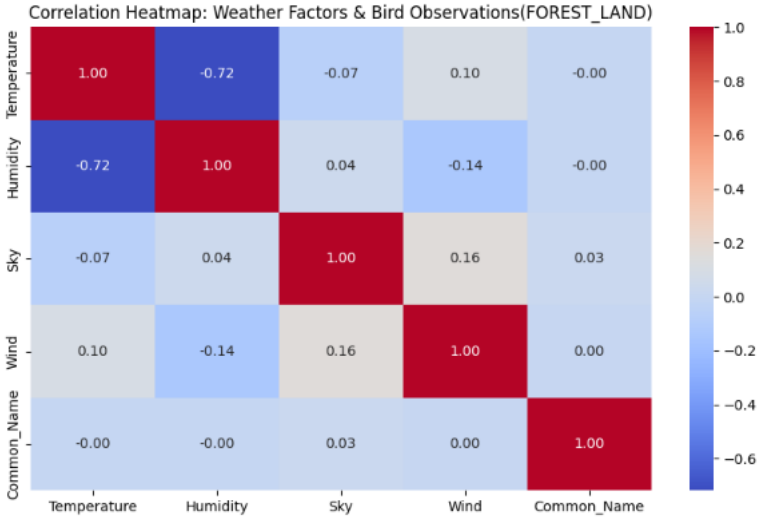
Findings- Most birds were identified through singing, followed by calling and visualization.



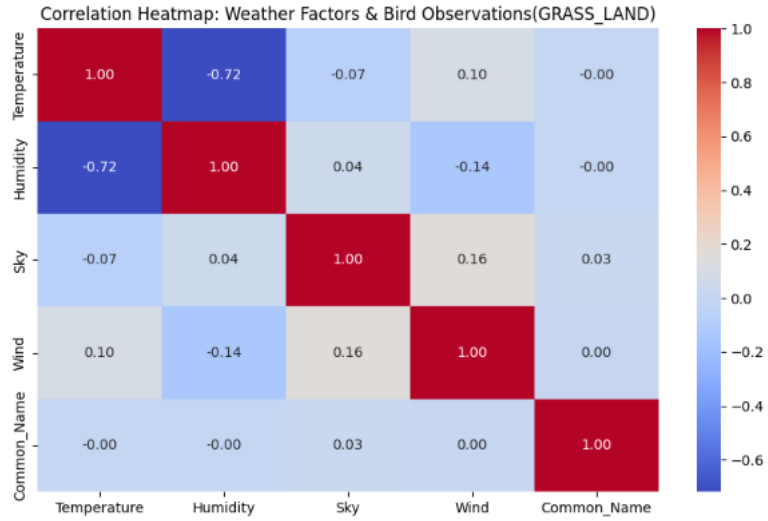
Findings- Peak bird observations were between 6 AM to 8 AM. Activity significantly reduced post 9 AM.



Findings- Grasslands exhibited more species diversity than forest areas.



Findings- Temperature and humidity showed a strong negative correlation (-0.72), while bird counts showed negligible correlation with weather parameters, implying that in forests, bird presence is not strongly influenced by basic weather factors.



Findings- Similar to forest land, temperature and humidity had a negative correlation (-0.72), but no significant direct correlation was found between weather variables and bird counts in grassland either.

Actionable Insights / Recommendations

• Most bird observations occur between 6–8 AM. Future fieldwork should focus on these hours to optimize effort.

• Grassland regions exhibit higher species diversity than forests—conservation efforts could prioritize these zones.
• Visual sightings dominate the ID methods. Training in auditory identification may expand recognition capacity.
• The most frequently observed species are Spizella pusilla and Cardinalis cardinalis—trend analysis of their populations is recommended.
• Elizabeth Oswald contributed the highest number of observations; resource allocation can be optimized based on observer productivity.
• Humidity shows moderate inverse correlation with temperature but limited effect on total observation count.
• Power BI dashboard allows filtering by season, habitat, and observer—valuable for dynamic insights during presentations.
• Forest data shows less species variation—indicating either fewer visits or ecological limitations.
• Improved tracking of observation times (start and end) is necessary for temporal behavior modeling.
• Correlation heatmaps show very weak relationship between weather factors and total bird counts—suggesting robustness of observations across weather conditions.