Bird Species Observation Analysis

Problem Description

The project aims to analyse 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 behaviour. 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.

Data Understanding

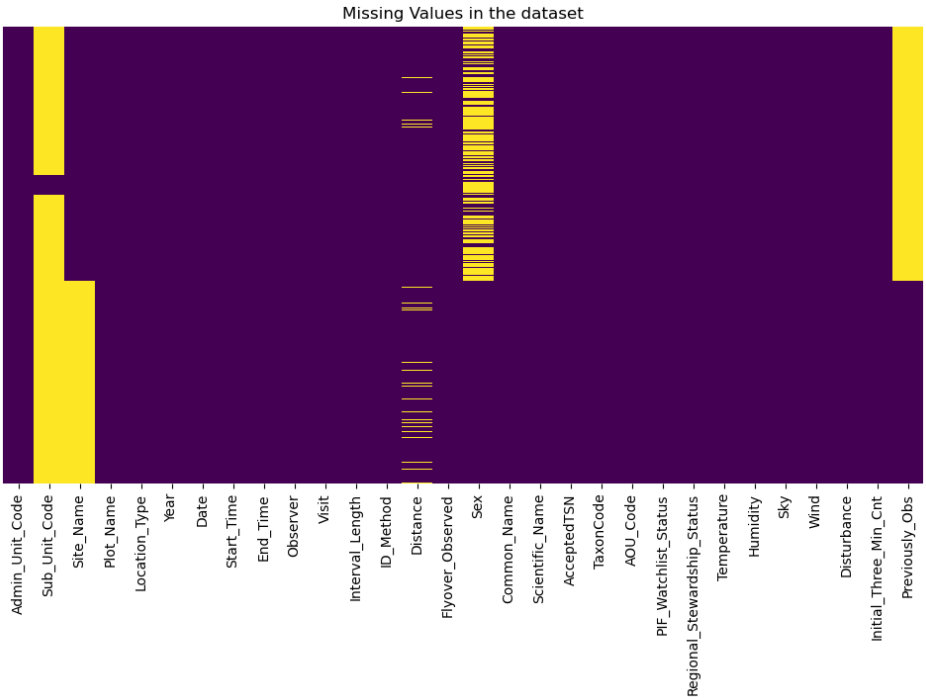
This analysis utilizes two datasets: 'Bird_Monitoring_Data_FOREST.XLSX' and 'Bird_Monitoring_Data_GRASSLAND.XLSX'. These datasets represent bird observation records collected from two distinct habitats: forests and grasslands for the months of May, June, and July 2018. The data was initially stored in multiple sheets per file, corresponding to different administrative units, and was subsequently flattened and combined into a single comprehensive dataset for analysis. The combined dataset has 17077 rows and 30 columns.

Data Characteristics and Limitations:

The combined dataset is inherently unbalanced, with varying numbers of records across habitats, months, and locations. The data coverage is limited to a three-month window in 2018, which constrains the ability to analyse long-term trends or seasonal patterns beyond this period. Additionally, the datasets were collected via standardized bird monitoring protocols, but the number of visits and effort might not be uniform across all observers, sites or habitats, potentially biasing the observations. Certain data fields contain missing or inconsistent entries. Despite these limitations, the dataset provides a valuable snapshot of bird diversity and distribution across forest and grassland habitats during this period.

Data Cleaning

In this analysis, initial duplicate detection revealed 1,705 duplicate records within the dataset. These duplicates were addressed using appropriate methods to ensure data integrity. Following this, a comprehensive missing value analysis was performed across all columns. It was observed that certain fields, such as Sub_Unit_Code and Site_Name, had significant missing entries, with Sub_Unit_Code missing in 14,650 rows and Site_Name missing in 6,826 rows. Other columns, including ID_Method, Distance, and Sex, also contained sporadic missing values.



To prepare the dataset for analysis, several columns were deemed redundant and removed: Sub_Unit_Code, TaxonCode, Previously_Obs, AcceptedTSN, and Site_Name. Missing values in categorical variables such as Sex, ID_Method, and Distance were imputed with default categories like 'Undetermined' or 'Unknown'. Numeric columns were handled accordingly, with AcceptedTSN filled with 'Unknown'. After these cleaning steps, the dataset was saved for subsequent analysis.

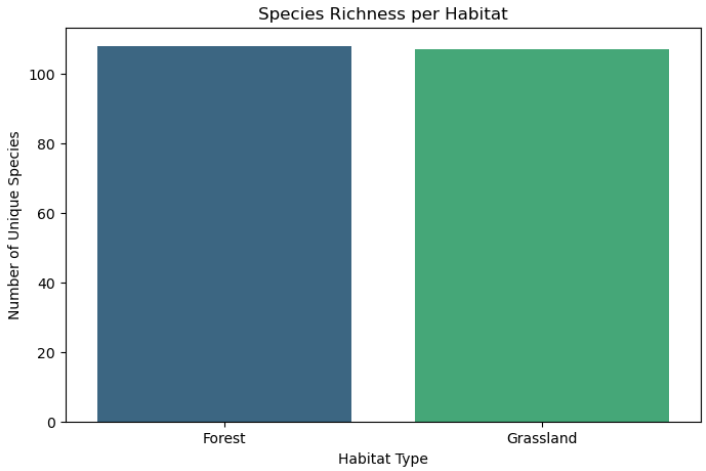
Exploratory Data Analysis

The initial phase of exploration began with examining the dataset's features by identifying the unique values present in each variable. This process provided a clear understanding of the distribution and diversity of the data, as well as helped in detecting any inconsistencies or anomalies.

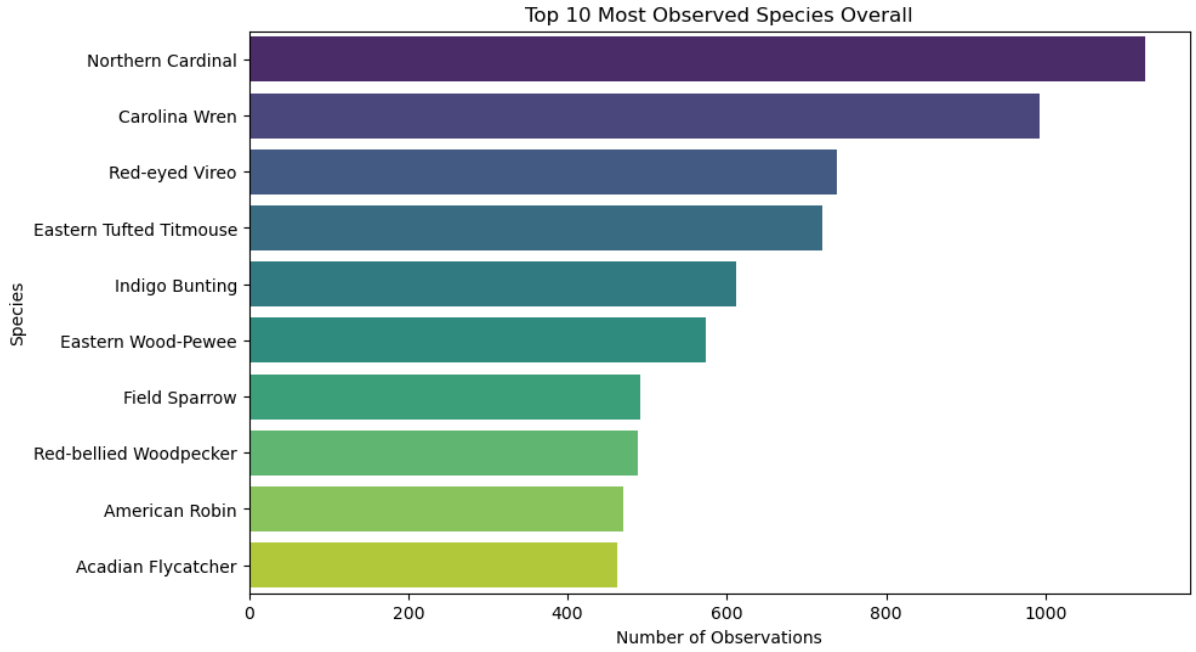
The EDA was divided into different analysis phases.

Spatial and Species Analysis

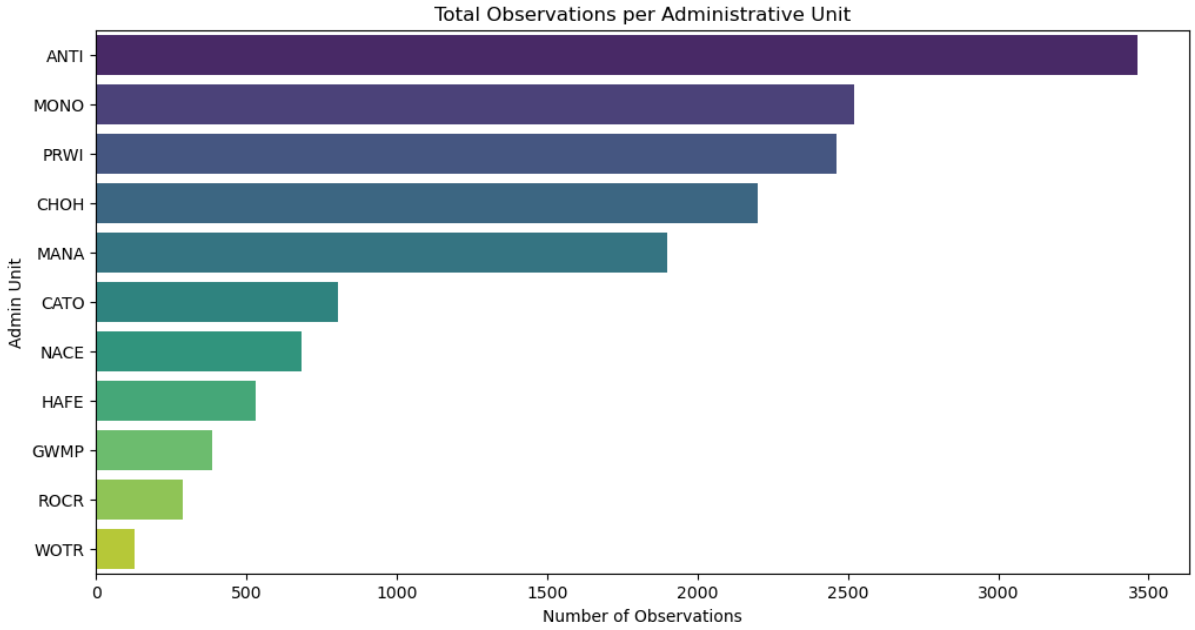
The analysis of spatial and species distribution revealed that observations were made across two primary habitats: forests and grasslands, each with a similar number of unique sites, totalling 108 and 107 respectively, indicating a balanced sampling effort across both environments.

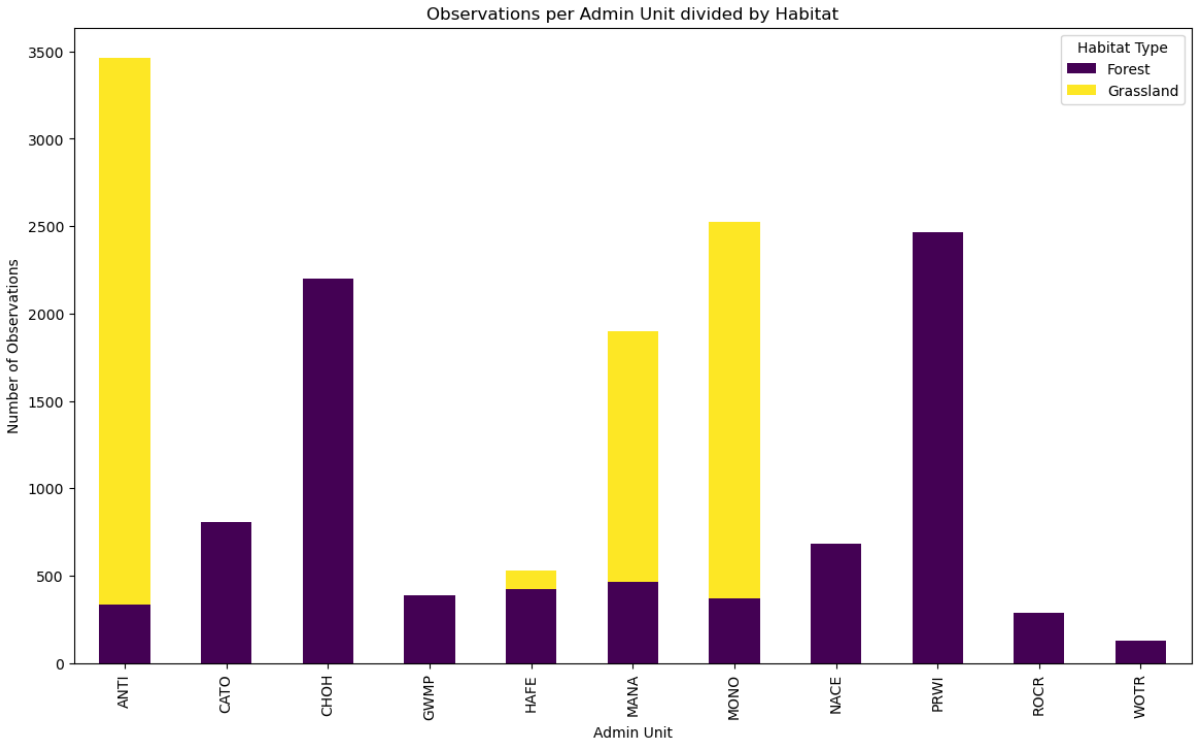


Overall, a total of 126 distinct bird species were recorded within the dataset. The most commonly observed species included the Northern Cardinal, with 1,125 sightings, followed by the Carolina Wren with 993, and the Red-eyed Vireo with 738 sightings. The top 10 species overall provide insight into the prevalent avian diversity in the surveyed areas.

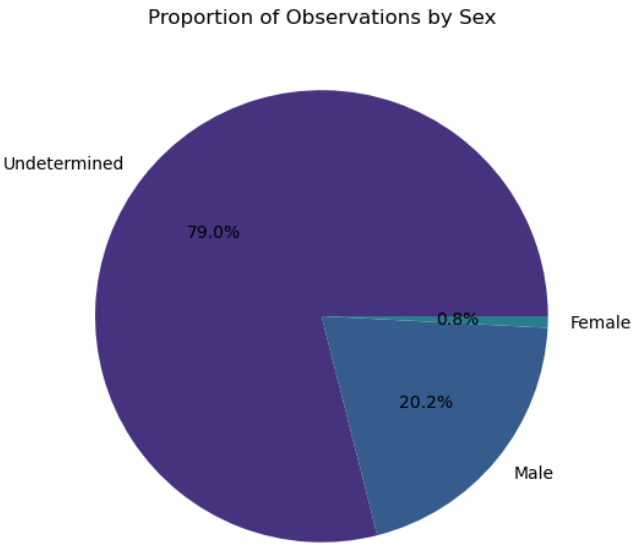


Examining species richness across administrative units, it was observed that the units MONO, ANTI, and MANA led with the highest diversity, recording 100, 81, and 81 species respectively. Other regions such as CHOH and NACE followed closely, with 80 and 66 species each. In contrast, units like GWM and CATO exhibited comparatively lower species counts, with 49 and 46 species respectively.





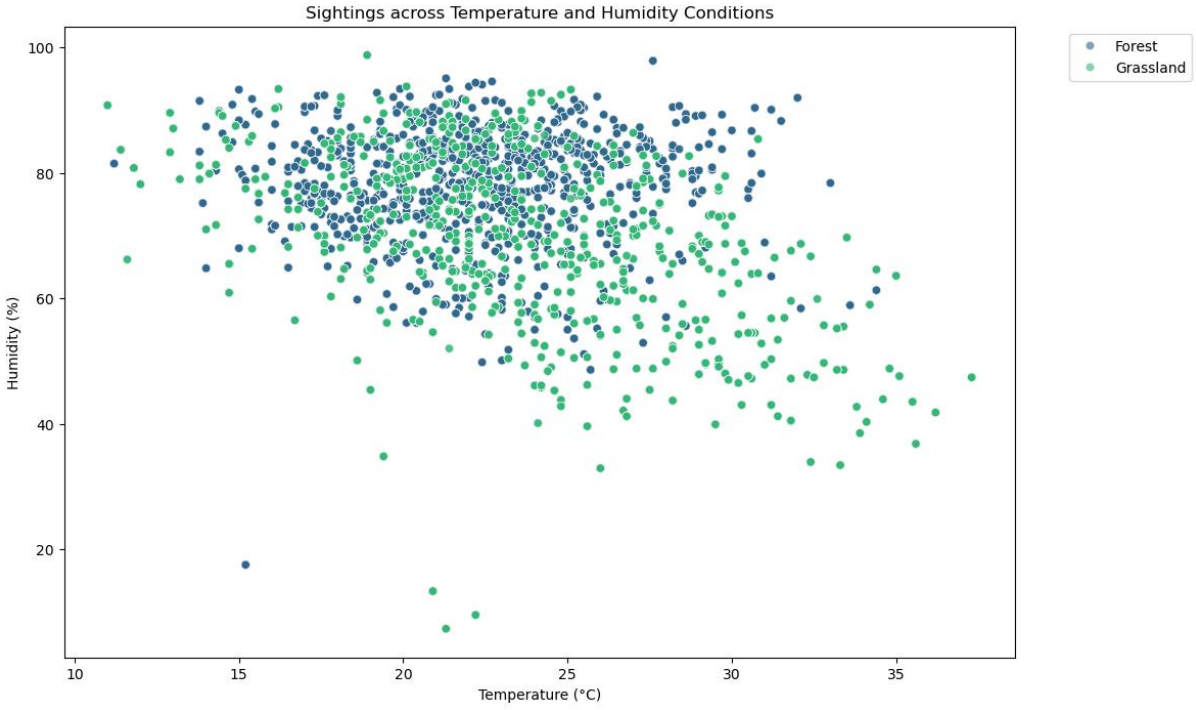
Behavioural insights from the data also indicated a significant gap in sex identification; approximately 79% of observations lacked information on the sex of the birds, marked as “undetermined.” Among the observations where sex was identified, males accounted for a mere 20.2%, suggesting a potential detection bias or inconsistency in data recording practices.

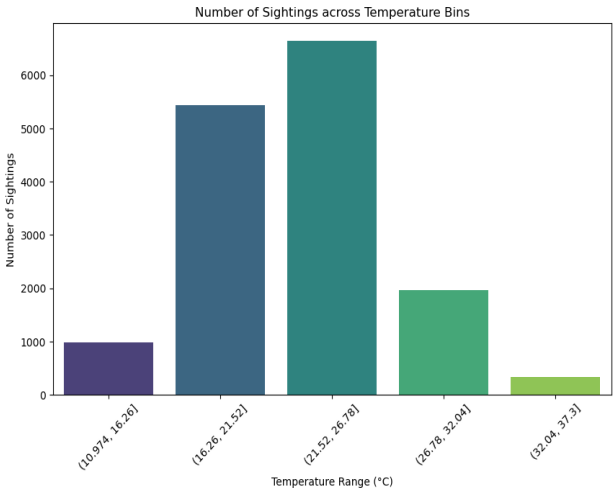


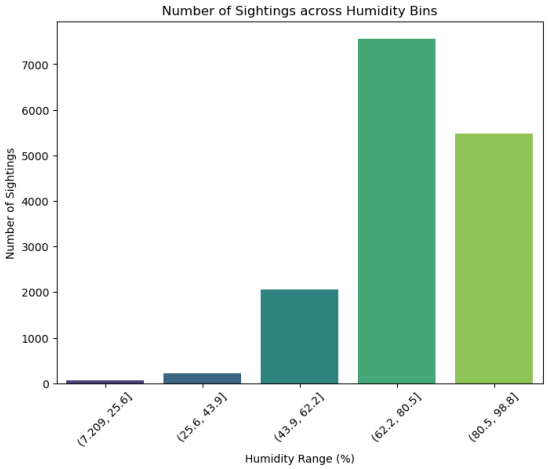
Environmental impact Analysis

Understanding environmental conditions is crucial in interpreting bird activity and observation patterns. This analysis delves into the effects of temperature, humidity, sky conditions, and wind on bird sightings and species diversity. It also examines how these factors influence flyover patterns, identification methods, and responses to disturbances.

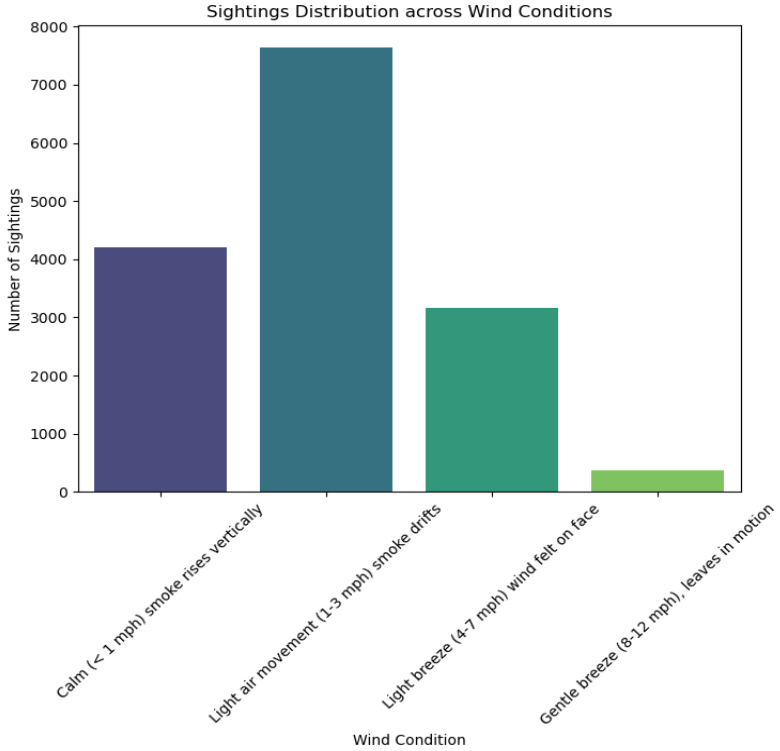
Most bird sightings tend to occur within moderate temperature ranges of 20 to 25 degrees Celsius, and humidity levels between 65 to 80 percent. This suggests that birds are most active or visible under these favourable weather conditions, possibly due to optimal thermoregulation and foraging opportunities.



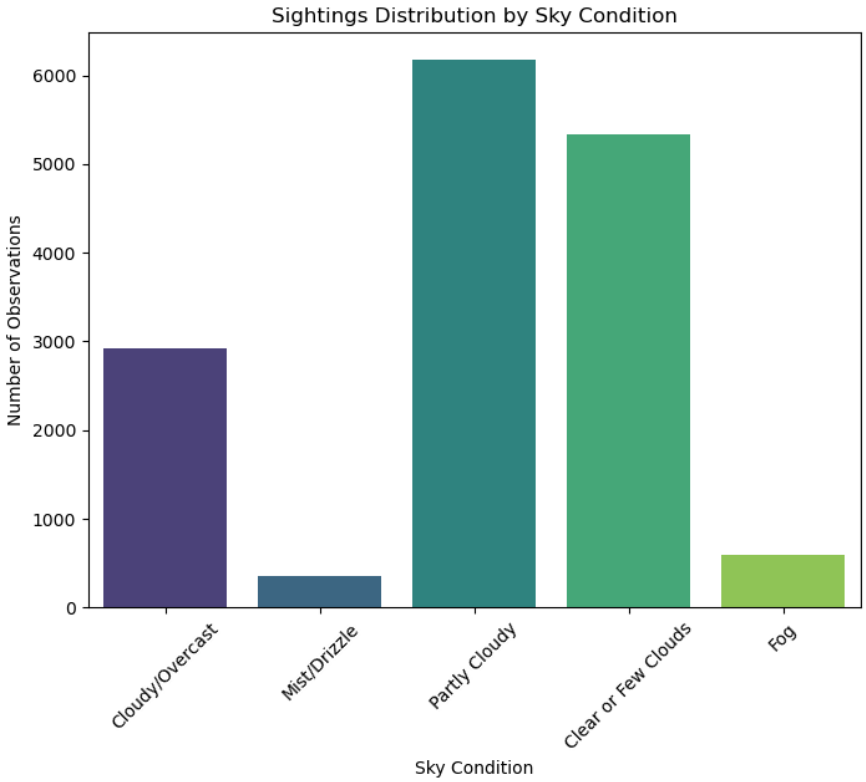




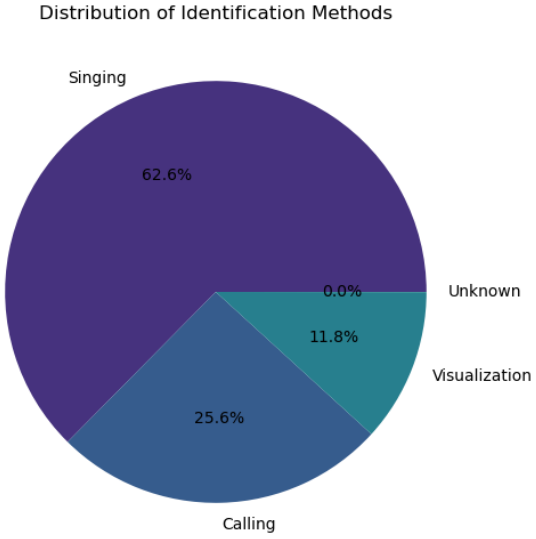
The distribution of sightings is heavily influenced by wind conditions, with the highest number of observations recorded when wind speeds are light, specifically with gentle air movements of 1-3 mph. Such calm or mildly breezy conditions appear to facilitate better detection, both visually and acoustically, and are conducive to bird flight and activity. Conversely, stronger winds (8-12 mph) see notably fewer sightings, probably because increased wind turbulence hampers sound transmission and bird flight stability, reducing observer success.



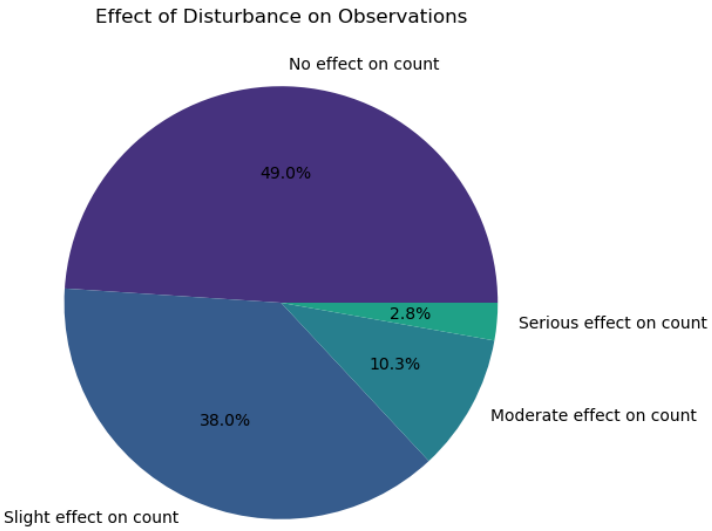
Sightings significantly decline during foggy or misty conditions, which likely obstruct visibility and auditory cues, making bird detection more challenging.



Additionally, the majority of species are identified by their singing, emphasizing that vocal cues are a critical detection tool in avian surveys.

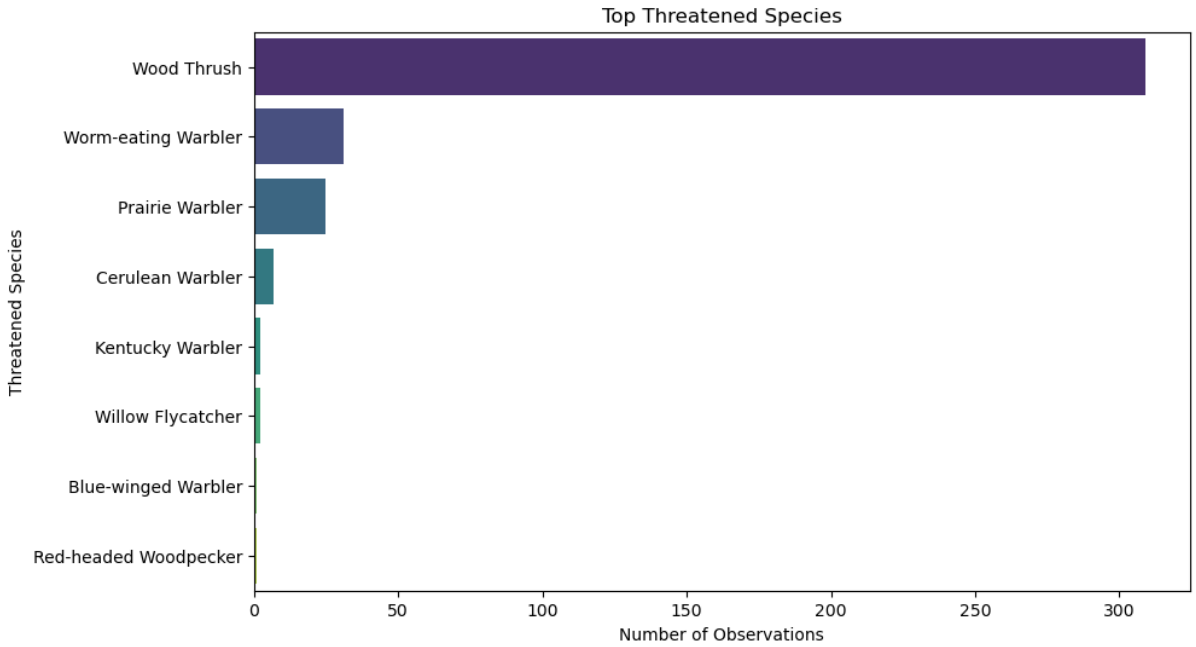


Finally, analysis indicates that disturbances such as nearby human or environmental disruptions generally have minimal impact on bird sightings and behaviour, with a large proportion (approximately 49%) showing no effect. When disturbances do occur, they tend to have only a slight or moderate effect on counts, suggesting that many bird species are resilient to minor environmental disturbances during observation periods.

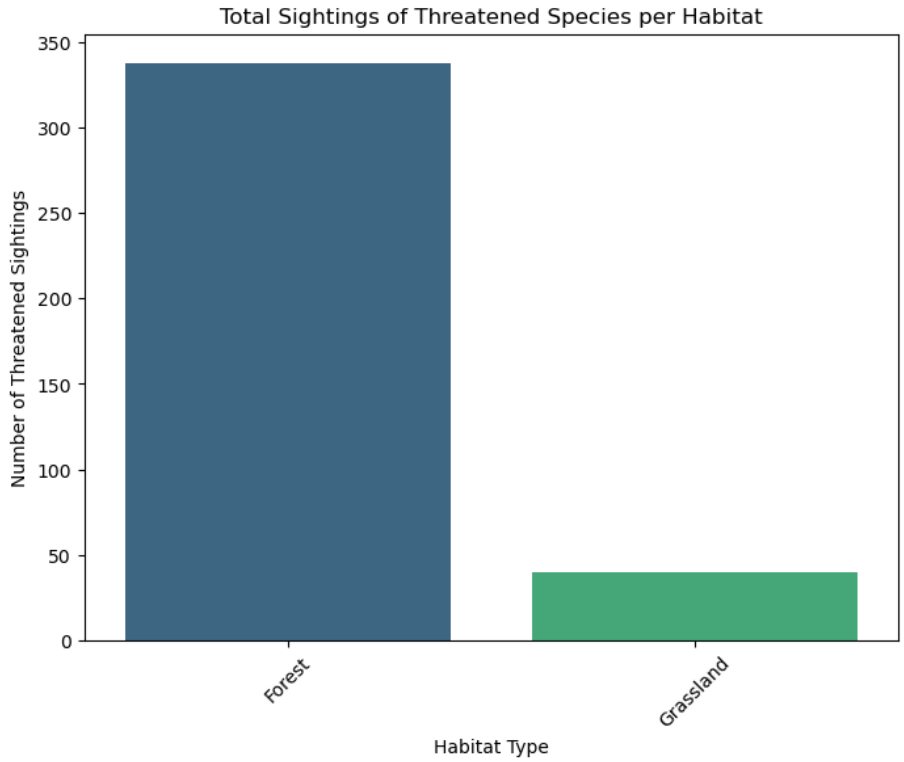


Threatened Species Analysis

The analysis identified a total of eight distinct threatened bird species within the dataset, highlighting a notable conservation concern. Among these, the **Wood Thrush** emerged as the most observed threatened species, with a total of 309 sightings, indicating its considerable presence across the sampled habitats. Other threatened species such as the Worm-eating Warbler, Prairie Warbler, Cerulean Warbler, Kentucky Warbler, Willow Flycatcher, Blue-winged Warbler, and the Red-headed Woodpecker were observed much less frequently, with counts ranging from 1 to 31 sightings.



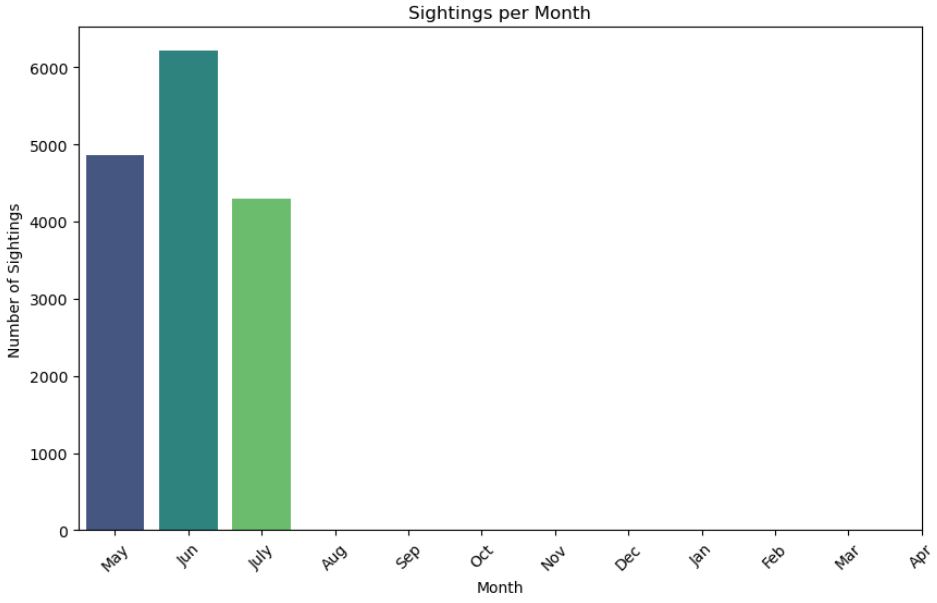
When examining the distribution across different habitat types, it was evident that most threatened species sightings occurred within **forest habitats**, with a total of 338 threatened observations compared to 40 in grasslands. This suggests that forests are critical habitats for threatened bird species, possibly due to greater habitat complexity and resource availability that support these species' survival.



Furthermore, the data reveals that threatened species are relatively rare compared to non-threatened species, highlighting the need for targeted conservation efforts in habitats where these species are most prevalent. Identifying and protecting key forest areas could be vital for conserving these threatened populations, particularly the Wood Thrush, which appears to be the most common among them.

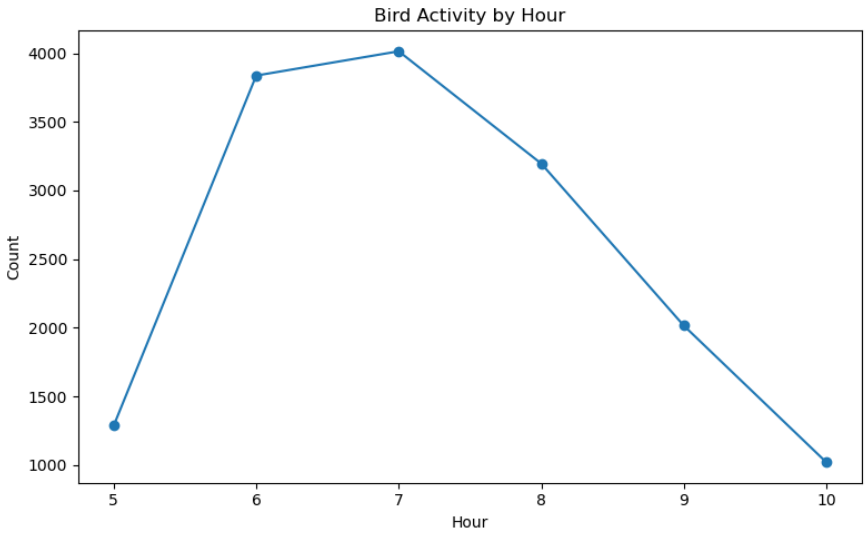
Temporal Pattern Analysis

The dataset encompasses bird observation data exclusively from May, June, and July of 2018.

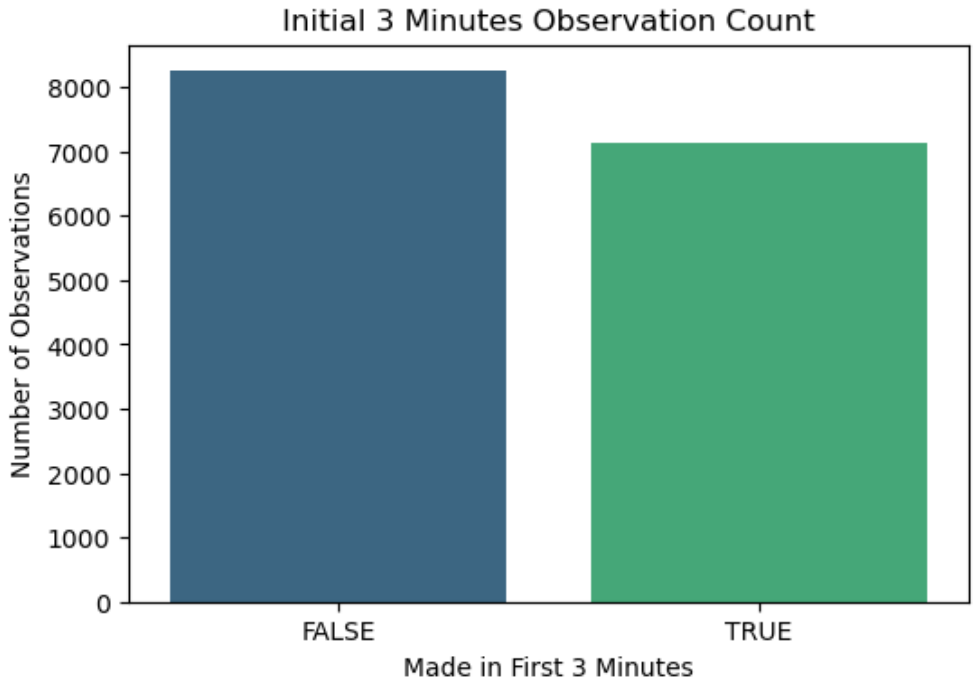


An analysis of hourly activity patterns revealed peaks in bird sightings during early morning hours, specifically between 6:00 and 8:30 AM. The visualized trend clearly shows that the highest number

of observations occurs around 6:00 and 7:00 AM, which corresponds with peak bird activity during dawn, a period when many species vocalize and forage actively. Conversely, sightings decline steadily after 8:30 AM, underscoring the importance of early morning surveys for capturing maximum bird activity.



The analysis of the initial three minutes of observation shows that a significant proportion of bird detections are made during this period, with a higher count of observations made during the first three minutes (TRUE) compared to later. This highlights that the initial moments of a survey often yield the most sightings, likely because birds are most active and vocal during dawn. It also suggests that short, focused surveys during this period can efficiently capture a representative snapshot of bird activity, making field efforts more effective without necessarily losing important data.



Together, these insights underscore the critical importance of early morning surveys, especially in the first few minutes, for effective bird monitoring and research.

Summary of EDA Insights

- The dataset shows an imbalance with uneven records across habitats, months, and locations, along with some missing data points in key fields.
- A total of 126 unique bird species were identified, with the Northern Cardinal, Carolina Wren, and Red-eyed Vireo being the most frequently observed.
- Threatened species such as the Wood Thrush are primarily found in forest habitats, highlighting the importance of forest conservation.
- Bird activity peaks at moderate temperatures between 20-25°C and humidity levels of 65-80%, while sightings decrease during foggy or misty conditions.
- Light wind conditions (1-3 mph) are associated with higher detection rates, especially during bird flyovers.
- Observation activity is highest between 5:55 and 8:30 AM, coinciding with peak vocal activity, and most species are identified through singing.
- The majority of sightings occur during the first three minutes of observation, emphasizing early morning as the most effective time for bird monitoring.

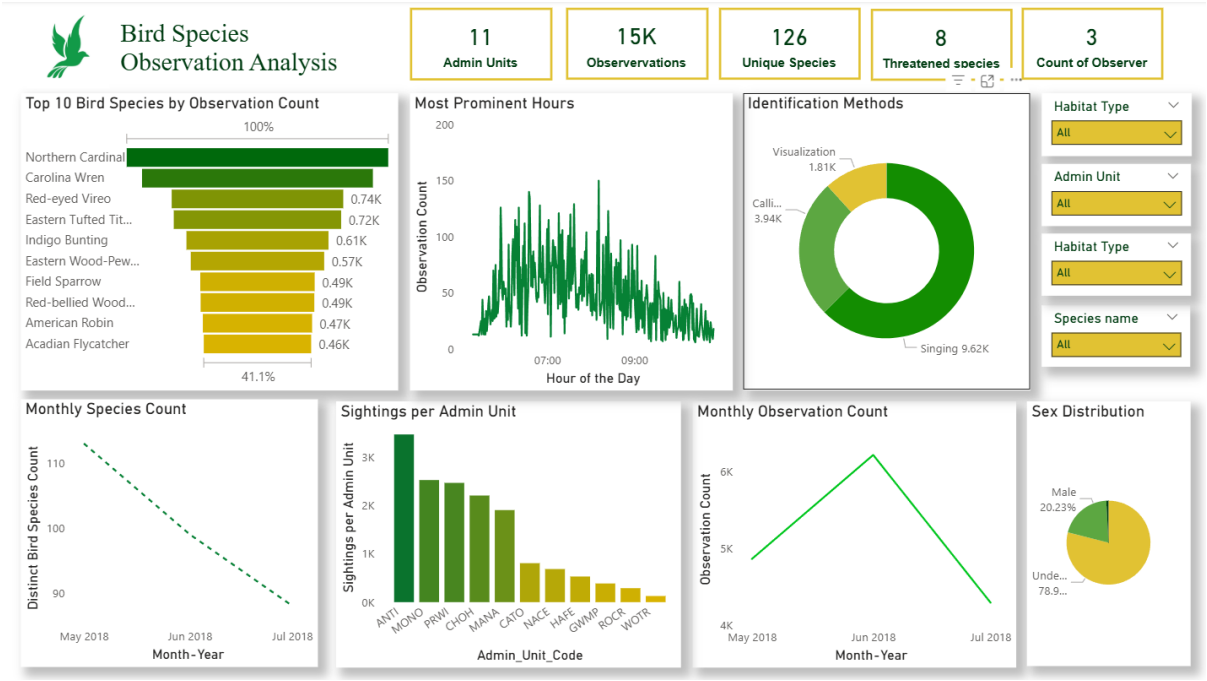
Further analysis was conducted on Power BI, creating various dashboards.

Power BI Dashboard Analysis

Power BI dashboard consists of Four Pages:

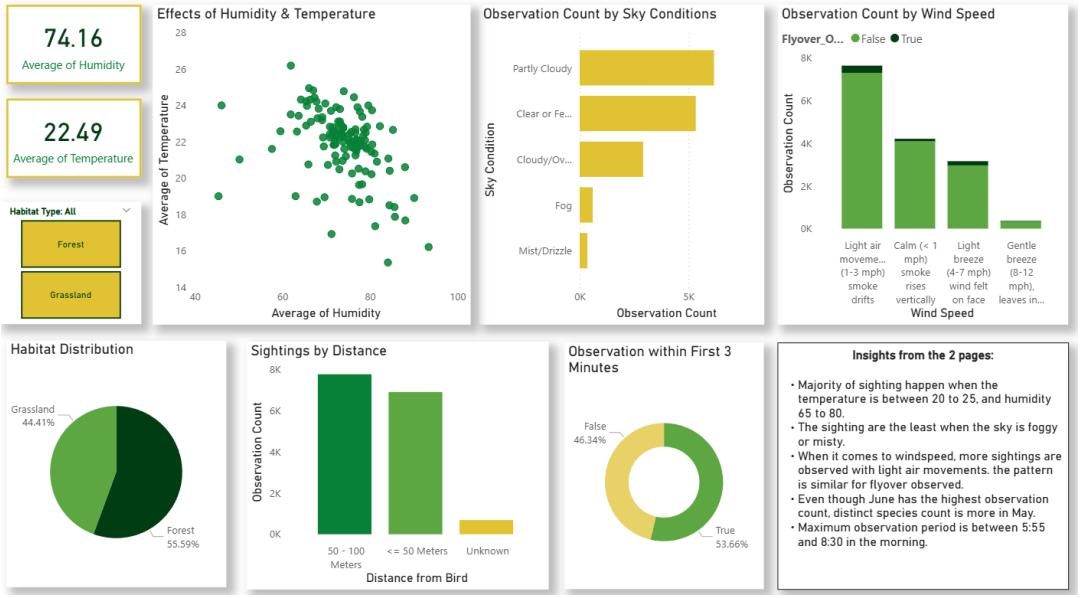
Overview – Temporal, Spatial, and Species Analysis

The dashboard provides the same information we obtained from the EDA. It provides key metrics indicating robust bird activity, with over 15,000 observations spanning 126 species across 11 administrative units. The top observed species include the Northern Cardinal, Carolina Wren, and Red-eyed Vireo, highlighting their prevalence during the survey period. Early morning hours, particularly between 6:00 and 8:30 AM, show peak activity, underscoring the importance of morning surveys for capturing maximum bird detections. Acoustic methods, especially singing, dominate identification, emphasizing their effectiveness and the need for consistent recording protocols. Monthly trends reveal the highest diversity and sightings in May, tapering into June and July, aligning with seasonal migration and breeding periods. Certain admin units such as MONO and CHOH contribute disproportionately to total sightings, reflecting higher bird richness or survey effort in these areas. Notably, about 79% of observations lack sex identification, highlighting an area for improvement in field data collection. Filters available on the dashboard such as habitat type, Admin Unit, Habitat Type, Species Name, allow users to explore specific patterns and customize their view based on research focus or interest.



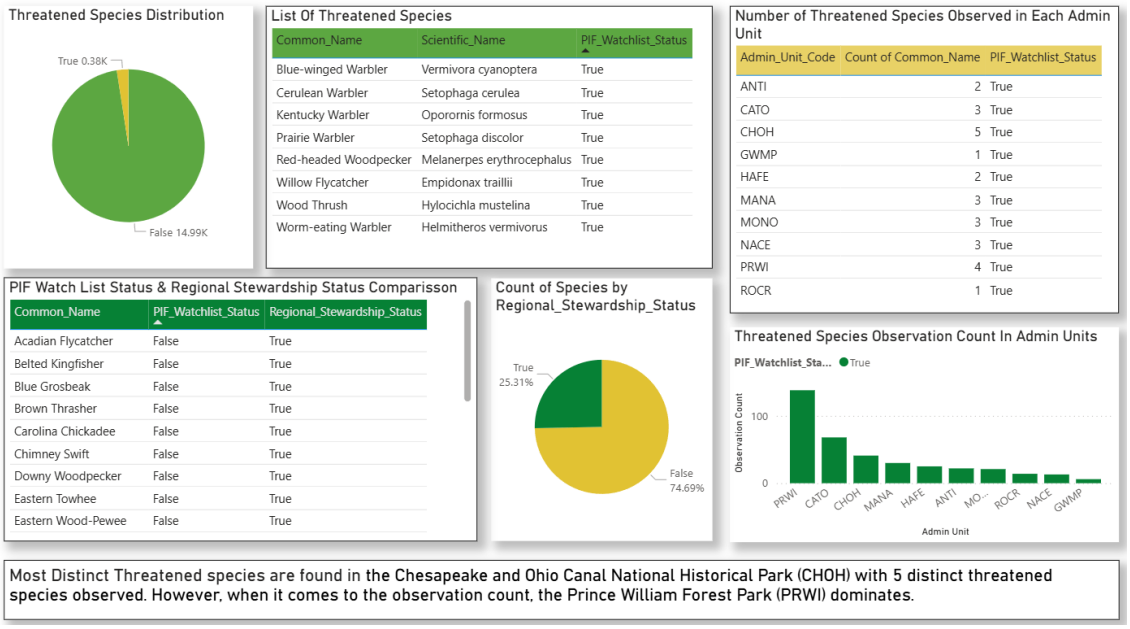
Environmental and Habitat Impact Dashboard

The dashboard provides insights into bird sightings related to environmental conditions and habitat. The average humidity is 74.16, and the average temperature is 22.49, with most sightings occurring within a temperature range of 20 to 25 and a humidity range of 65 to 80. The habitat distribution is fairly even, with 55.59% of sightings in forests and 44.41% in grasslands. Sightings are most frequent at distances of 50-100 meters and under partly cloudy or clear skies, and they are less common in foggy or misty conditions. Light air movements (1-3 mph) correlate with more sightings, a pattern also seen in flyover observations. A majority (53.66%) of observations occur within the first 3 minutes. While June has the highest overall observation count, May shows a greater diversity in species. The most active observation period is between 5:55 AM and 8:30 AM.



Threatened Species Dashboard

The dashboard focuses on threatened bird species, indicating that only a small fraction (0.38K) are on the PIF Watchlist, while a much larger number (14.99K) are not. A list identifies specific threatened species, including the Blue-winged Warbler and Cerulean Warbler, all marked as "True" for PIF Watchlist Status. The count of species by regional stewardship status shows that 25.31% have regional stewardship, while 74.69% do not. A comparison of PIF Watch List Status and Regional Stewardship Status shows several species like the Acadian Flycatcher and Belted Kingfisher are not on the PIF Watchlist ("False") but do have regional stewardship ("True"). The Chesapeake and Ohio Canal National Historical Park (CHOH) has the highest count of distinct threatened species (5), but when considering observation count, the Prince William Forest Park (PRWI) dominates. In terms of threatened species observed in each admin unit, CHOH has 5, while PRWI has 4.



Observer Analysis Dashboard

The dashboard analyses bird observations by different observers across various administrative units. Elizabeth Oswald accounts for 37.49% of the observations, Kimberly Serno for 34.78%, and Brian Swimelar for 27.73%. Each observer has visited a similar number of admin units, with Elizabeth Oswald and Kimberly Serno both visiting 11 and Brian Swimelar visiting 10. The observation counts per visit vary across different admin units (ANTI, CATO, CHOH, GWMP), with the highest counts often occurring during the first visit (Visit 1). Elizabeth Oswald has the most distinct visits, followed by Brian Swimelar and Kimberly Serno. While the number of visits doesn't directly correlate with the observation count, factors like the month of the visit and weather patterns could play a role. Kimberly Serno, despite having fewer visits than Brian, has a higher observation count, potentially due to experience. The location (National Park) significantly impacts the observation count and distinct species count.

Summary of Dashboard Insights

- The dataset exhibits imbalances in data distribution and contains missing values in key fields.
- The survey identified 126 bird species, with the Northern Cardinal, Carolina Wren, and Red-eyed Vireo being the most frequently observed.
- Threatened species, such as the Wood Thrush, are primarily found in forest habitats, emphasizing the importance of forest conservation.
- Optimal bird activity is observed between 20-25°C and 65-80% humidity, while sightings decrease during foggy or misty conditions.
- Light wind conditions (1-3 mph) are associated with higher bird detection rates.
- Peak bird activity occurs between 5:55 and 8:30 AM, and most sightings are made within the first three minutes of observation.
- Species are primarily identified through acoustic methods, particularly by their singing.
- Observer experience and location impact the number of bird sightings recorded.
- A small percentage of species are on the PIF Watchlist, and forests are a critical habitat for these species.

Suggestions based on Insights

For optimizing surveys targeting threatened species specifically, prioritize forest habitats known to harbour species like the Wood Thrush. The ideal environmental conditions for these surveys are temperatures between 20-25°C with humidity levels of 65-80%. It's best to conduct these observations during light wind conditions (1-3 mph) to improve detection rates, but avoid foggy or

misty days. The most effective time for bird monitoring is during the early morning hours, specifically between 5:55 and 8:30 AM, focusing on the first three minutes to capture the most activity. When in the field, emphasize acoustic methods, particularly identifying species by their songs. Additionally, consult regional stewardship lists to understand which species are of concern locally, even if they aren't on the national PIF Watchlist, and tailor survey efforts accordingly. Ensuring these practices are followed will lead to more thorough and accurate bird monitoring.

Conclusion

This analysis of bird species observation data reveals valuable insights into species distribution, habitat preferences, and environmental factors influencing bird activity. The findings highlight the importance of forest habitats for threatened species, the impact of weather conditions on bird sightings, and the need for standardized data collection practices. By implementing targeted conservation efforts, enhancing data analysis, and engaging the community, it is possible to improve bird monitoring and promote effective biodiversity management. Further research and continuous data collection are essential to understanding long-term trends and adapting conservation strategies to protect avian communities in the face of environmental changes.