Bird Species Observation Analysis: Project Output

1. Data Cleaning & Preprocessing

The initial phase of the project involved meticulous data cleaning and preprocessing to ensure reliability and consistency across all records. Missing values were addressed by imputing data where feasible; in cases where critical columns such as species, habitat type, or observation metadata were incomplete, the entries were removed. This approach helped maintain the integrity of key analyses. Standardisation efforts included harmonising time and date formats, as well as categorical variables—such as habitat types (“Forest” and “Grassland”) and observer IDs. Furthermore, data originating from various administrative units (across multiple Excel sheets) were consolidated into a unified structure, enabling robust comparative analysis across locations.

2. Exploratory Data Analysis (EDA) & Insights

2.1 Temporal Analysis

A detailed temporal analysis revealed that bird species counts were significantly higher during summer (approximately 11,000) compared to spring (around 5,000), suggesting either increased bird activity or greater detectability during the summer months. Diversity indices, such as Shannon and Simpson, generally peaked mid-week and declined towards weekends, a trend that may reflect variations in observer effort or changing environmental conditions.

Bird activity, as measured by species richness, was found to be greater early in the morning, tapering off by late morning. This pattern underscores the importance of timing in bird observation efforts.

2.2 Spatial Analysis

Analysis by location type indicated that forests recorded more total bird observations (8,500) than grasslands (6,800), although both habitat types made significant contributions to overall species richness. Biodiversity was found to be distributed across multiple administrative units, with certain units such as "ANTI" and "CHOH" identified as hotspots for particular bird species.

At the plot level, some sites consistently attracted a greater diversity of species, providing valuable information for prioritising monitoring or conservation efforts.

2.3 Species Analysis

A total of 126 unique bird species were recorded, with 113 species contributing to the overall species richness metric. The most commonly observed species included the Northern Cardinal, Carolina Wren, and Red-eyed Vireo. Among the species, eight were identified as at-risk and placed on a “watchlist”.

Most observations were made through singing and calling, highlighting these as the dominant bird detection methods. The initial three-minute count was especially significant, contributing a large proportion of the total observations and demonstrating the importance of rapid survey efforts.

Regarding sex ratio, a considerable portion of birds observed (about 45%) had undetermined sex. Among those with known sex, males were observed more frequently than females.

2.4 Environmental Correlation

Species richness demonstrated a positive correlation with humidity, with more birds observed as humidity exceeded 50%. The average temperature during observation periods was approximately 22.5°C. Most observations were conducted under clear or partly cloudy conditions, with fewer records during fog, mist, or drizzle. Disturbance levels were typically minimal or classified as 'no effect', indicating relatively undisturbed observation settings.

2.5 Distance and Behaviour

The majority of birds were observed within close proximity (≤50 metres), particularly in forested areas. Only about 4.5% of birds were recorded as “flyover”, with most birds seen perched or stationary during the observation periods.

2.6 Observer Patterns

Analysis revealed observer bias, as some individuals contributed disproportionately to the dataset. This observer effect was accounted for in subsequent analyses. Observers tended to conduct more visits and record a higher number of species during the spring. Additionally, seasonality in visit numbers was aligned with the higher species counts registered in the summer months.

3. Conservation Insights

3.1 Watchlist Species

Eight bird species were identified as conservation priorities and flagged for focused attention. Their distribution and frequency were mapped, revealing that watchlist

species were present in both forest and grassland habitats, thereby indicating the need for targeted monitoring in both environments. Analysis using AOU codes further helped map specific species distributions to relevant regions and parks for more effective conservation action.

4. Visualisations

Interactive dashboards were developed using Power BI, facilitating comprehensive exploration of the dataset. Key visualisations included filterable bar charts for observation counts by habitat and observer, donut charts depicting sex ratios, sky conditions, and flyover frequency, and line charts illustrating diversity indices across dates, days, and seasons. Histograms visualised the relationships between species richness and both humidity and start time. User interactivity was enabled through filters for species, habitat, observer, administrative unit, date, and environmental variables, allowing stakeholders to conduct in-depth analyses tailored to their interests.

5. Key Findings & Actionable Insights

- Higher Bird Diversity in Forests: Forested areas were identified as crucial for bird conservation due to higher species counts and diversity metrics.
- Species Seasonality: Bird richness and activity peaked in summer, providing guidance for optimal timing of eco-tourism and monitoring efforts.
- Environmental Influences: Greater bird observations were associated with increased humidity and clear skies, informing monitoring protocols and resource allocation.
- Sex and Method Data Gaps: Improved documentation practices could enhance understanding of population dynamics and behavioural patterns.
- Priority Hotspots Identified: Certain plots and parks with consistently high observation rates were highlighted for conservation or eco-tourism development.
- At-Risk Species Monitoring: Real-time alerts or focused surveys during peak periods and in preferred habitats can support conservation efforts for watchlist species.

6. Business Use Cases Addressed

- Wildlife Conservation: The data-driven approach enabled identification of high-priority habitats and optimal seasons for conservation action.
- Land Management: Insights into habitat preferences informed targeted restoration strategies.

- Eco-Tourism: Identification of richness peaks and site hotspots provided opportunities for marketing to bird-watching enthusiasts.
- Policy & Monitoring: Robust metrics and trend analyses supported evidence-based policy-making and ongoing ecosystem health monitoring.

7. Limitations of the Bird Observation Data

7.1 Missing and Undetermined Values

A significant portion of the sex data for observed birds (over 45%) is marked as "Undetermined". This limitation restricts demographic analyses and impedes a thorough assessment of population structure or the detection of sex-based trends. Additionally, some columns contain missing entries or partially filled environmental variables, which can introduce bias or uncertainty into correlational studies.

7.2 Observer Bias

The dataset reflects observer bias, as a small number of individuals contributed a large proportion of the observations. Variations in individual skill, experience, and detection ability may impact both species counts and identification accuracy. Repeated visits by the same observers can amplify this effect, potentially skewing measures of abundance and diversity.

7.3 Detection Methods

Observation methods such as "Singing" and "Calling" dominate the records, potentially underrepresenting less vocal or less active species during survey periods. Furthermore, inconsistencies in detection ranges or effort across sessions were not fully controlled or standardised.

7.4 Habitat Representation

Despite representation of both forests and grasslands, the dataset is skewed towards forest observations, which could limit direct comparisons and introduce sample imbalance. Some sites, plots, or administrative units were sampled more intensively than others, affecting the spatial representativeness of the findings.

7.5 Temporal Gaps

Observation efforts were concentrated in spring and summer, resulting in fewer records for other seasons and the potential omission of key migratory or behavioural periods. Observation timing was also biased towards early mornings, possibly missing species active later in the day.

7.6 Disturbance and Environmental Data

Reported disturbance was consistently low ("no effect"), which could be due to underreporting and might mask the true impacts of human or environmental disruption on bird observations. Environmental variables like wind and humidity were broadly categorised, possibly missing more nuanced site-specific conditions.

7.7 Spatial Limitations

The scope of the study was confined to select administrative units or parks within a specific region, which limits the generalisability of results to other geographic areas. Additionally, greater observation intensity in certain plots or sites may have biased the identification of biodiversity hotspots.

7.8 Conservation Status & Rarity

Low counts of watchlist or rare species may either reflect their true scarcity or be due to detection difficulties, resulting in uncertainty regarding the conservation status of these species in the study area.

These limitations should be carefully considered when interpreting results, drawing management or policy conclusions, or applying the findings to broader conservation strategies.

8. Skills Demonstrated

The project showcased a range of analytical and technical skills, including data cleaning, exploratory data analysis, Power BI dashboard development, and both geographic and statistical analysis. The communication of actionable insights was tailored to suit the requirements of environmental and ecological studies.