



Project Report: Bird Species Observation Analysis

1. Project Overview



Problem Statement

The project focuses on analyzing bird species distribution in forests and grasslands to explore how environmental factors like vegetation, climate, and terrain shape bird populations and behavior. Using observational data, it aims to uncover habitat preferences and assess biodiversity impacts. This research will provide critical insights for habitat conservation strategies and understanding the influence of environmental changes on avian communities.



Business Objective

- Study bird species distribution and diversity across forest and grassland habitats.
- Identify habitat preferences and key environmental influences on bird presence.
- Determine peak observation times to optimize fieldwork and tourism activities.
- Highlight dominant and rare species for targeted conservation efforts.
- Recognize top observers to improve and motivate data collection.
- Provide actionable insights for biodiversity conservation and eco-tourism planning.

2. Methodology



Data Sources

File	Description
Bird_Monitoring_Data_FOREST.XLSX	Core forest birds records
Bird_Monitoring_Data_GRASSLAND.XLSX	Core Grassland birds records

These Excel files 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 Catoctin 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.

Tools Used

- **Python** : Data ingestion, cleaning, EDA.
- **Power BI**: Interactive 3 Paged dashboard.

◆ Phase 1: Data Acquisition & Initial Exploration

- Imported essential libraries like `pandas`, `numpy`, `matplotlib`, `seaborn`, and `missingno`.
- Loaded Excel files with multiple sheets and combined all sheets for both Forest and Grassland datasets into separate DataFrames.
- Displayed initial rows and checked dataset dimensions using `.head()` and `.shape()`.
- Used `.info()`, `.duplicated().sum()`, and `.isnull().sum()` to assess data structure, types, duplicate entries, and missing values.
- Visualized missing data patterns using `missingno`.

◆ Phase 2: Data Understanding & Structuring

- Reviewed column names and data summaries using `.columns`, `.describe()`, and `.nunique()`.
- Documented the role of each variable including observation details, species, location, and environmental conditions.
- Understood sheet-wise administrative unit information to map ecological sites (e.g., ANTI, CATO, etc.).
- Verified categorical diversity and value ranges across key columns for both datasets.

◆ Phase 3: Data Wrangling & Preprocessing

- Merged Forest and Grassland datasets into a unified DataFrame.
- Dropped low-utility column `Sub_Unit_Code`.
- Imputed missing values in:
 - `Site_Name`, `Distance` → "Unknown"
 - `Sex` → "Undetermined"

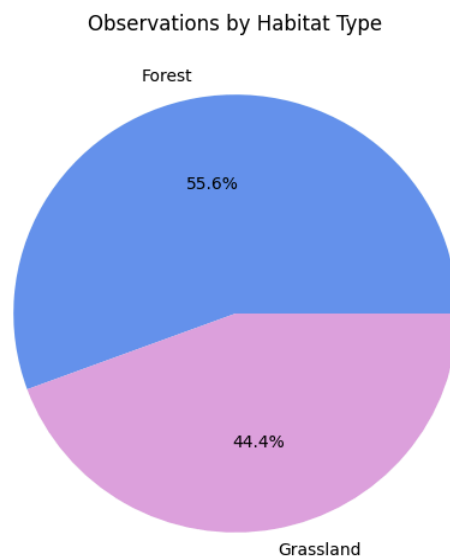
- NPSTaxonCode, TaxonCode → "N/A"
- Previously_Obs → Mode value
- Dropped rows with nulls in essential columns: ID_Method, AcceptedTSN.
- Cleaned Start_Time and End_Time fields: stripped strings, extracted HH:MM:SS, converted to proper time format.
- Created a new column Observation_Hour for time-based analysis.
- Standardized boolean fields (Flyover_Observed, PIF_Watchlist_Status, etc.) to True/False.
- Converted categorical columns to category dtype for efficiency.
- Removed duplicate rows to ensure data integrity.
- Saved final cleaned dataset as Bird_Monitoring_Clean_Merged_dataset.csv.

◆ Phase 4: Exploratory Data Analysis (EDA)

- Habitat Type Distribution
- Sex Distribution
- Top 10 Bird Species
- Distance Distribution
- Top Observers
- Hourly Observations
- Observations Over Time
- Boolean Category Distributions
- Temperature vs Humidity by Distance
- Sex Distribution by Habitat Type
- Top Species vs Environmental Factors
- Observer Activity by Habitat
- Observation Hour by Habitat

3. Key Results & Insights

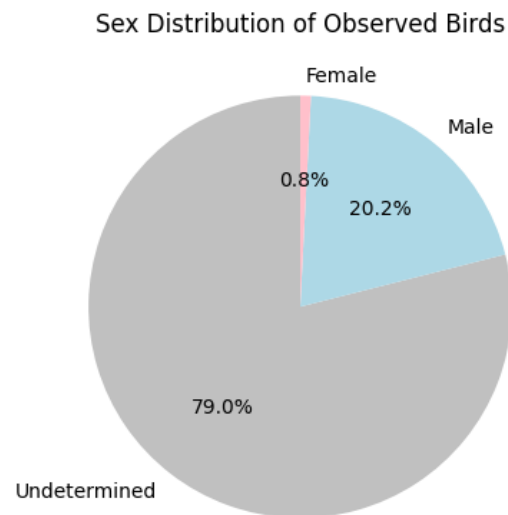
● Chart 1: Habitat Type Distribution



Insights:

The chart reveals that 55.6% of observations occur in forests, while 44.4% are in grasslands.

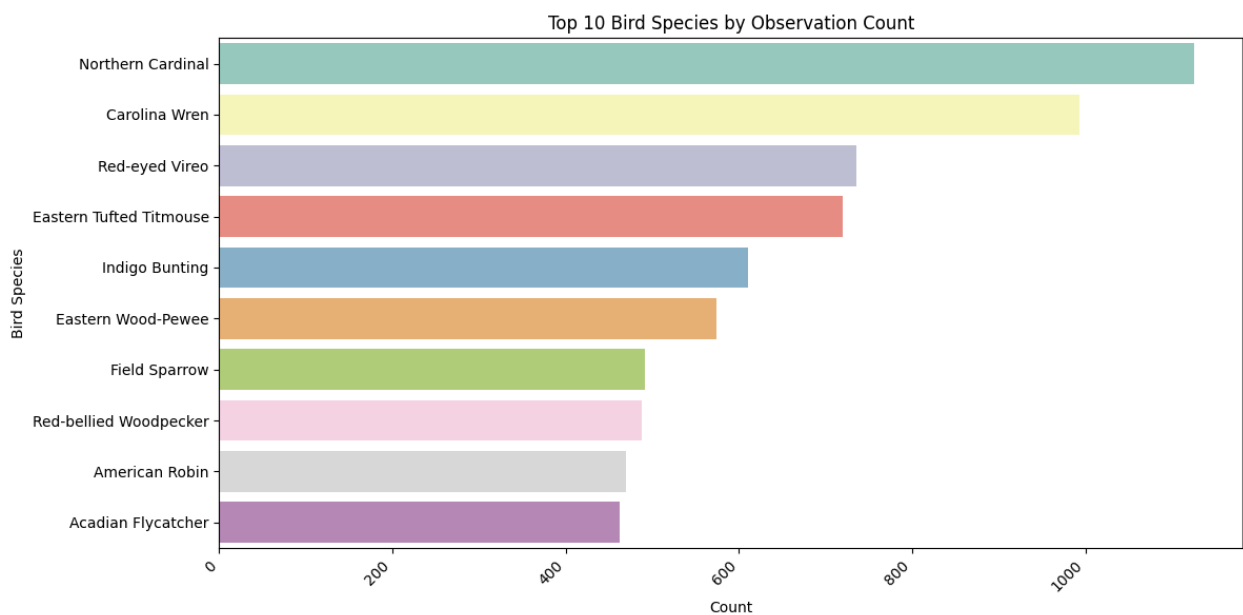
- **Chart 2: Sex Distribution**



Insights:

The data shows that 79% of observations have undetermined sex, 20.2% are male, and only 0.8% are female.

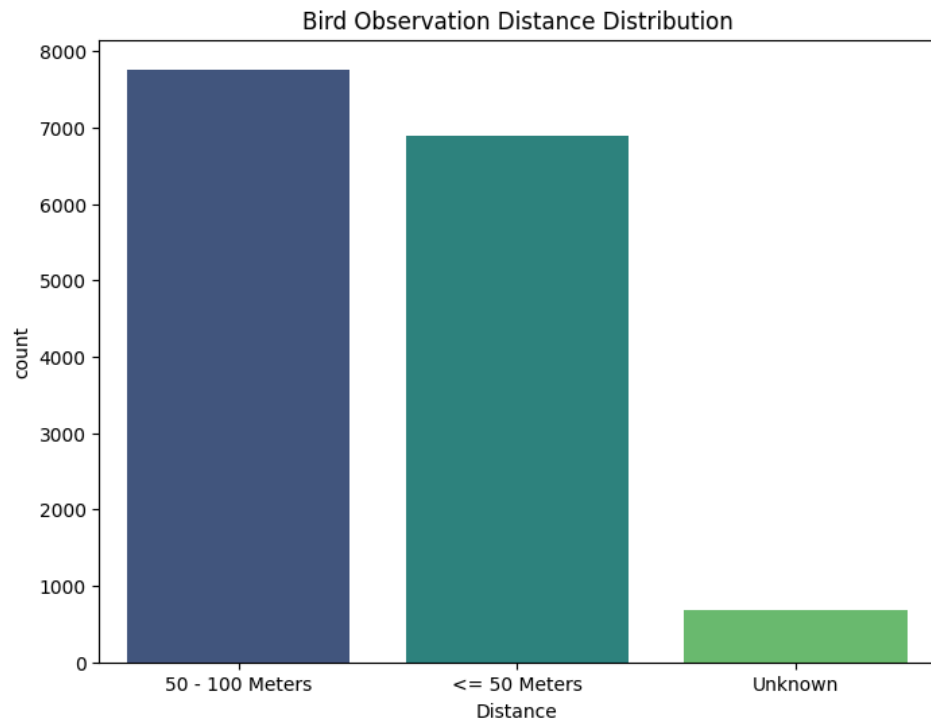
- **Chart 3: Top 10 Bird Species**



Insights:

The Northern Cardinal and Carolina Wren have the highest observation counts, indicating their dominance in the dataset.

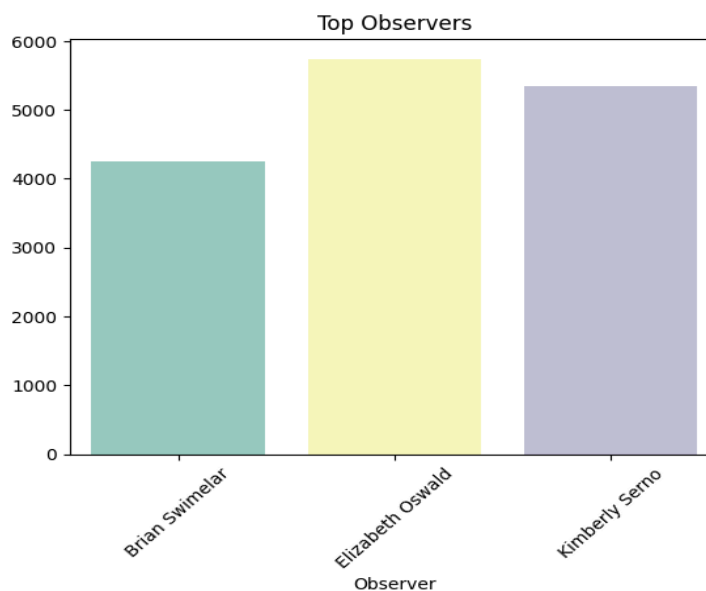
- **Chart 4: Distance Distribution**



Insights:

The insights are that most bird observations occurred at a distance of 50-100 meters, followed by those within 50 meters. A significantly smaller number of observations had an unknown distance.

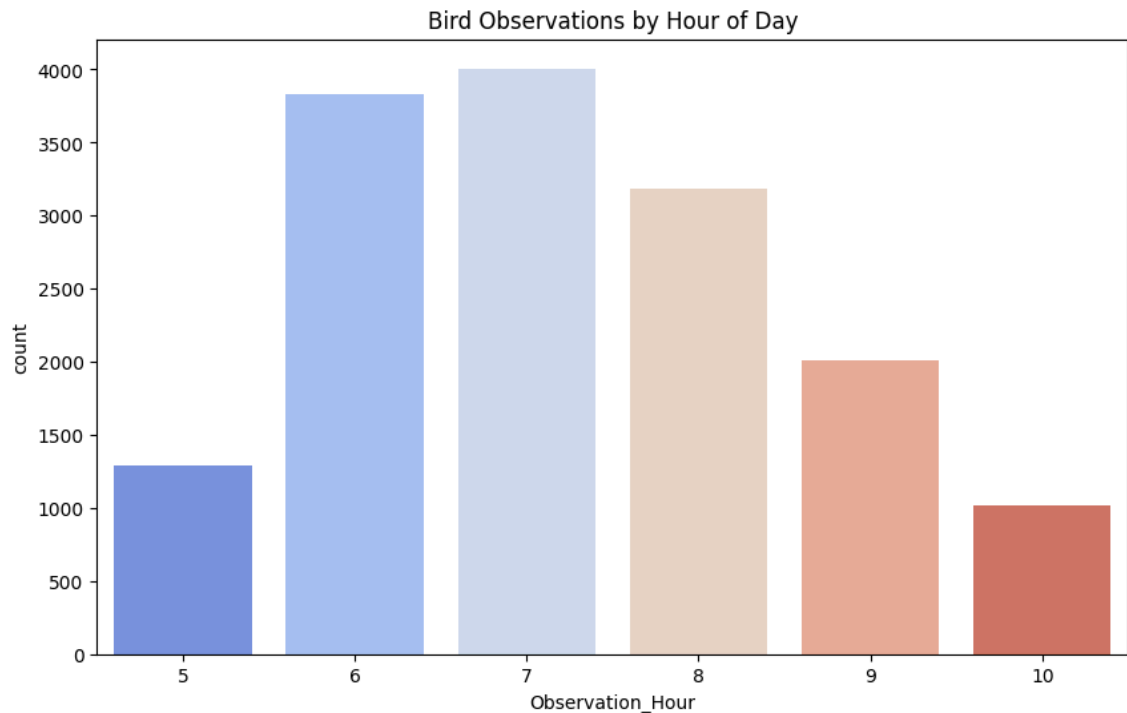
- **Chart 5: Top Observers**



Insights:

The chart reveals that Elizabeth Oswald is the most frequent observer, followed closely by Kimberly Serno, with Brian Swimelar having the lowest number of observations among the top three.

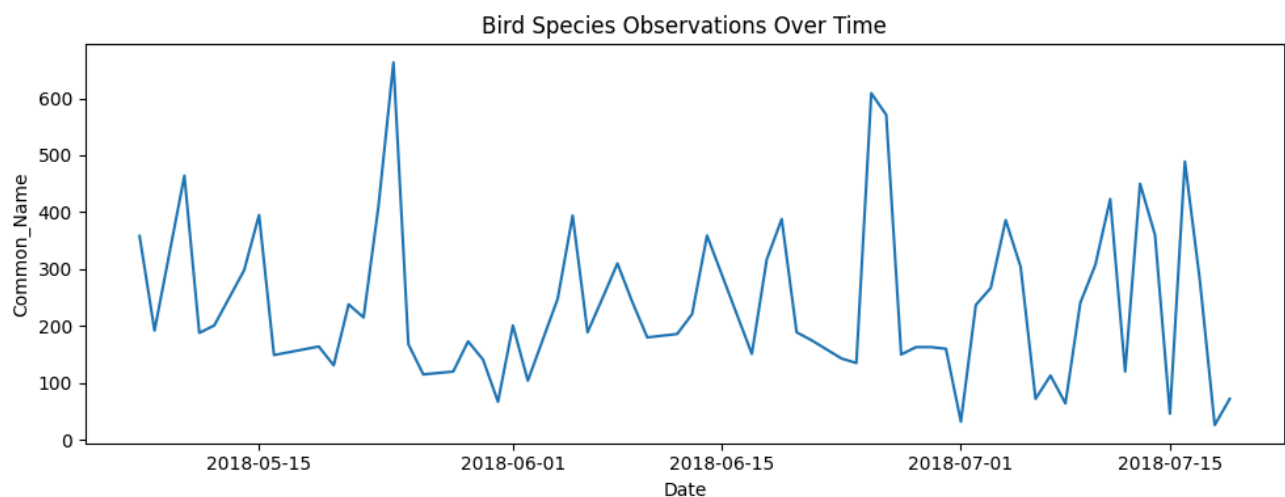
- **Chart 6: Hourly Observations**



Insights:

The insights are that bird observations peak at the 7th Hour, with the next high volume occurring at the 6th hour. The number of observations decreases steadily after the 7th Hour, with the lowest number of observations happening at 10th Hour.

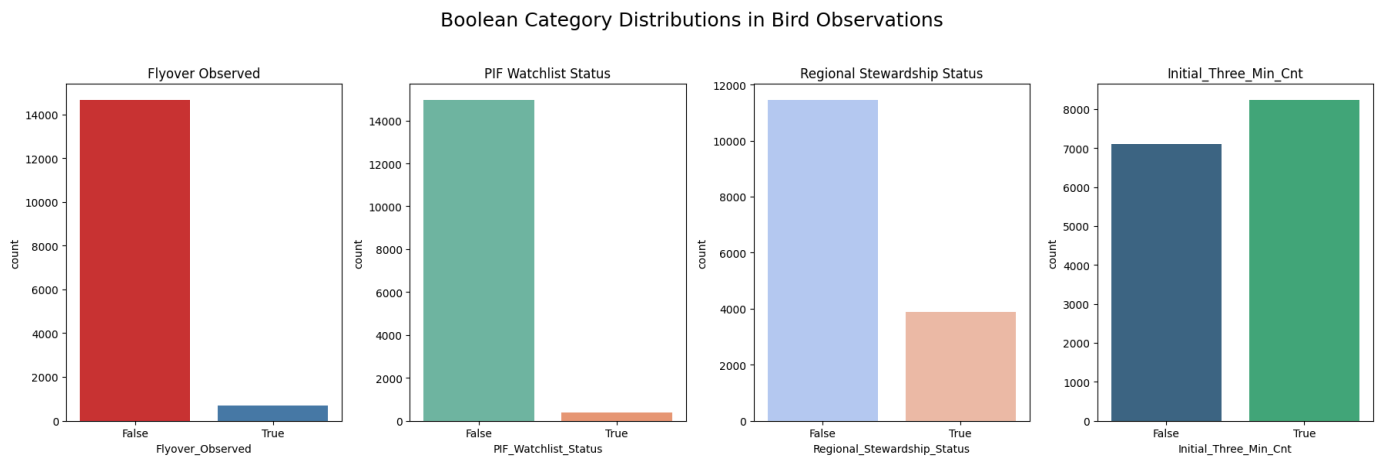
- **Chart 7: Observations Over Time**



Insights:

The chart reveals significant daily fluctuations in the number of bird species observations, with no clear long-term upward or downward trend. There are several noticeable peaks and valleys throughout the observed period.

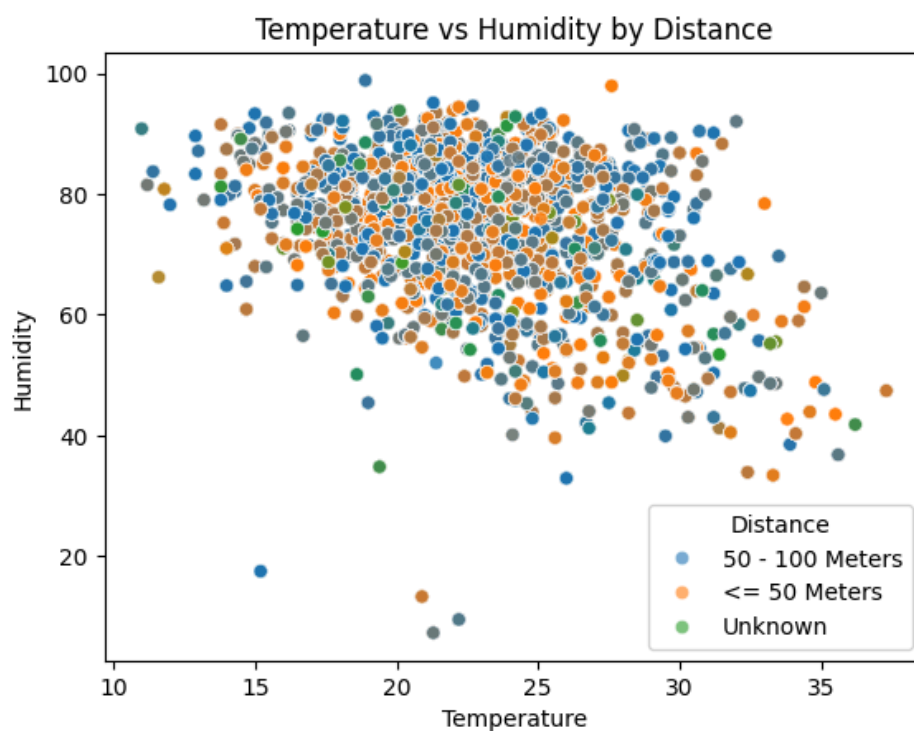
- **Chart 8: Boolean Category Distributions**



Insights:

The insights are that in most observations, the bird was not a flyover, was not on the PIF Watchlist, and was observed for more than three minutes. The Regional Stewardship Status, however, was 'False' in the majority of cases.

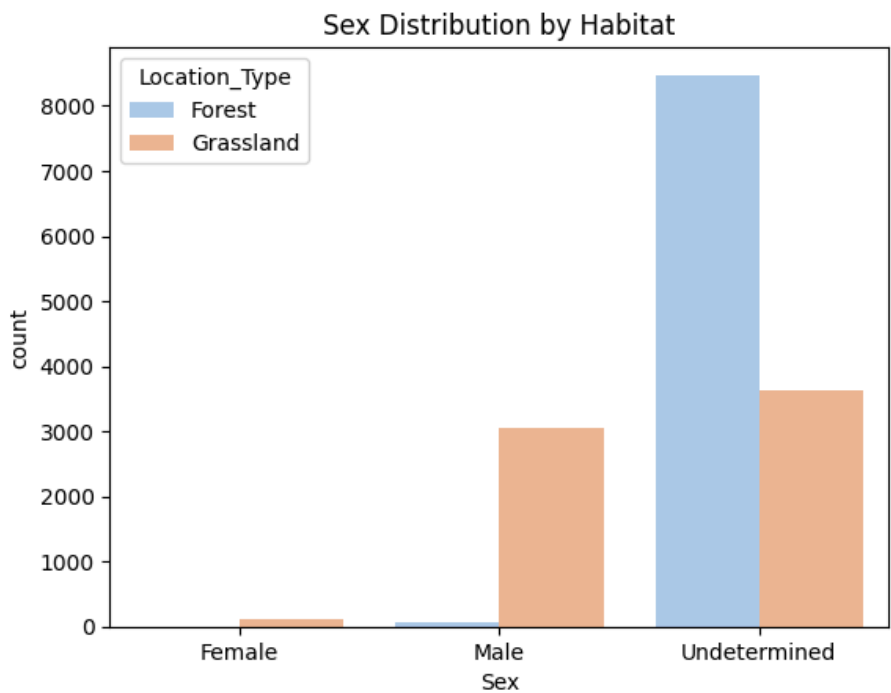
- **Chart 9: Temperature vs Humidity by Distance**



Insights:

The plot shows that most observations cluster in the 18–28 °C temperature range with humidity between 60–90%, regardless of distance. Humidity generally decreases slightly as temperature increases, and no strong separation is observed among the distance categories.

● Chart 10: Sex Distribution by Habitat Type

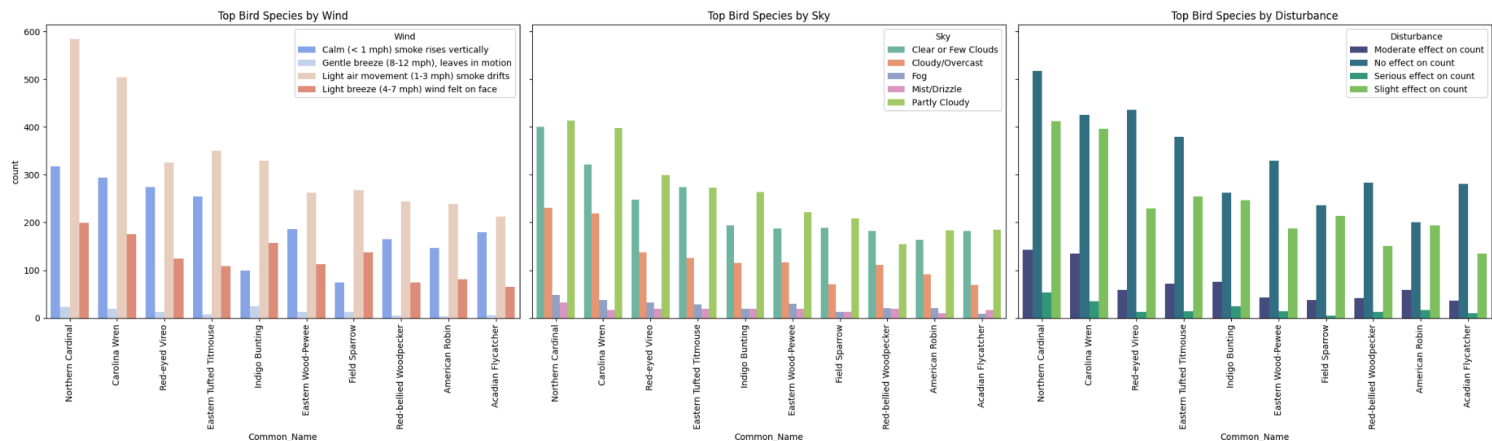


Insights:

The chart reveals that most observations in both habitats have an undetermined sex. A higher number of male birds were observed in grasslands, while a higher number of observations with undetermined sex occurred in forests.

● Chart 11: Top Species vs Environmental Factors

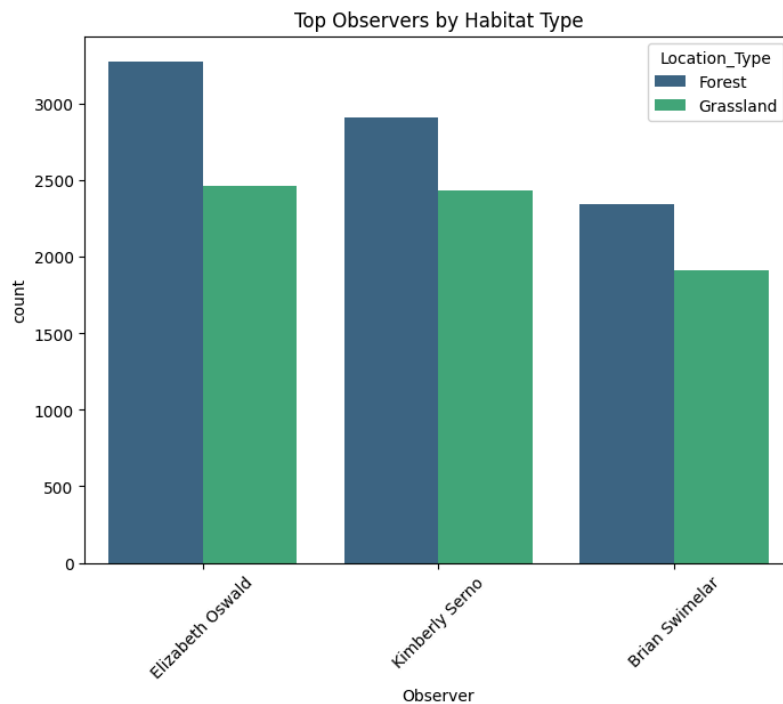
Top Bird Species Distribution Across Environmental Factors



Insights:

The charts show that Northern Cardinal, Carolina Wren, and Red-eyed Vireo are among the most frequently observed species across different wind, sky, and disturbance conditions. Bird counts tend to be highest under light breezes, clear or partly cloudy skies, and when disturbances have no effect on counts.

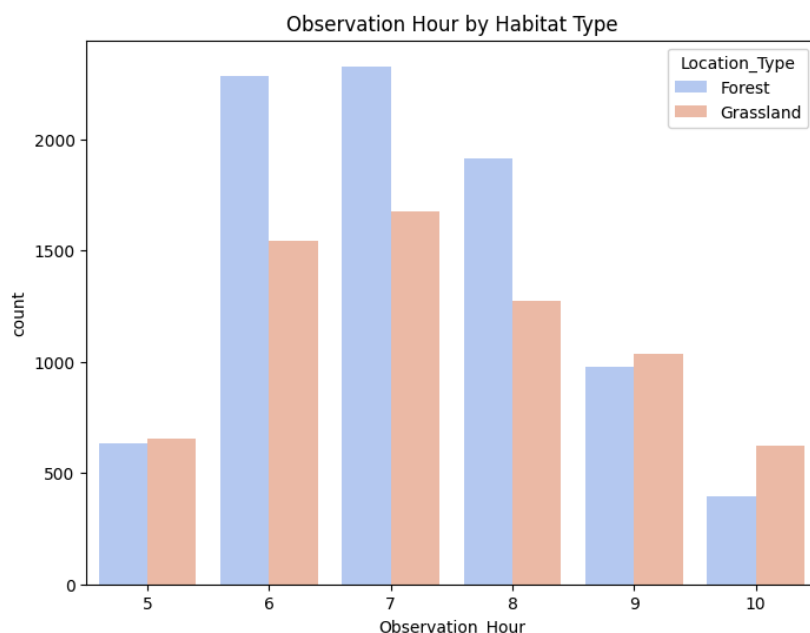
- **Chart 12: Observer Activity by Habitat**



Insights:

The insights are that all three top observers made more observations in forest habitats than in grasslands. Elizabeth Oswald made the most observations overall, with the largest difference between her forest and grassland counts.

- **Chart 13: Observation Hour by Habitat**



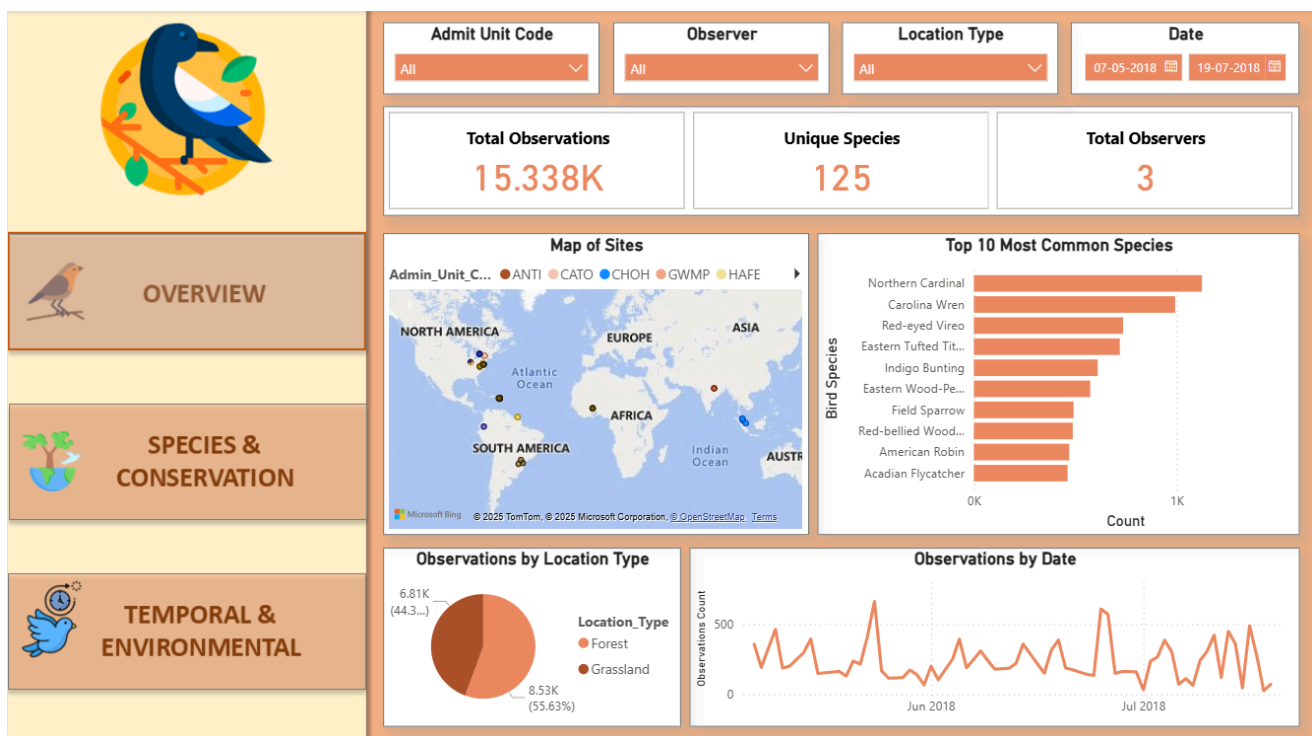
Insights:

The insights are that forest observations peak at 7th Hour. and then decline, while grassland observations are more consistent in the morning and peak slightly later. At the 5th, 9th and 10th Hour, there are more grassland observations than forest ones.

4. Dashboard Walkthrough (Power BI)

This Dashboard analyzes the Bird Species Observation Analysis using Power BI to extract actionable business insights. It includes data transformation to build a comprehensive 3-page dashboard.

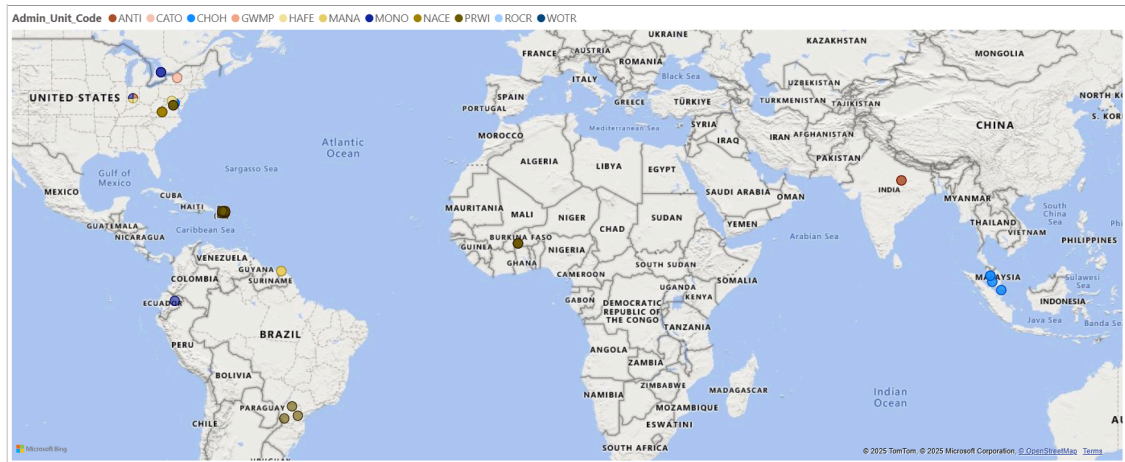
Page 1: Overview



- **KPI Cards:** Total Observations, Unique Species, Total Observers
- **Slicers:** Admit Unit Code, Observer, Location Type, Date

Visuals

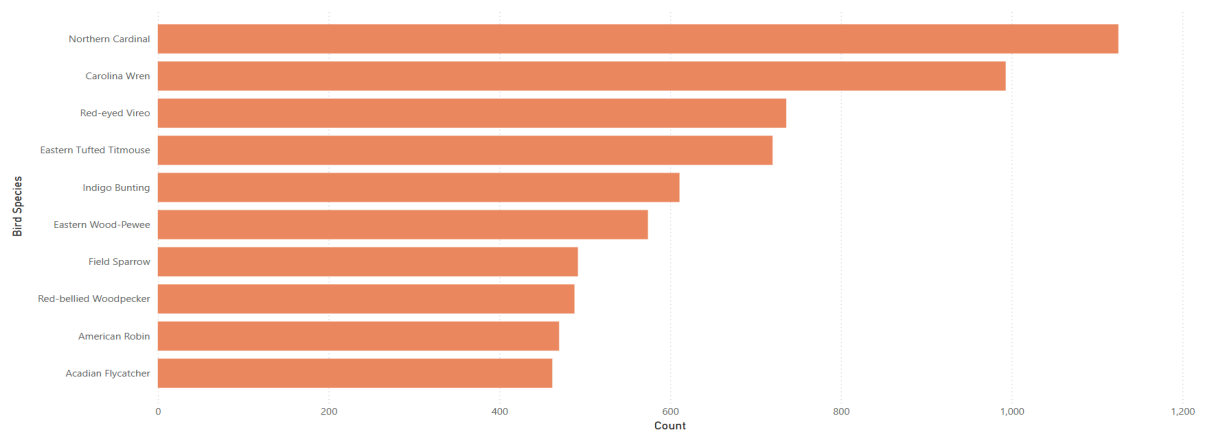
- Map Of Sites



Insights:

The graph displays the geographical locations of various "Admin_Unit_Code" entries across the world. Each code is represented by a specific color, with clusters of these codes appearing in the **United States**, **Haiti**, **Brazil**, **India**, and **Indonesia**.

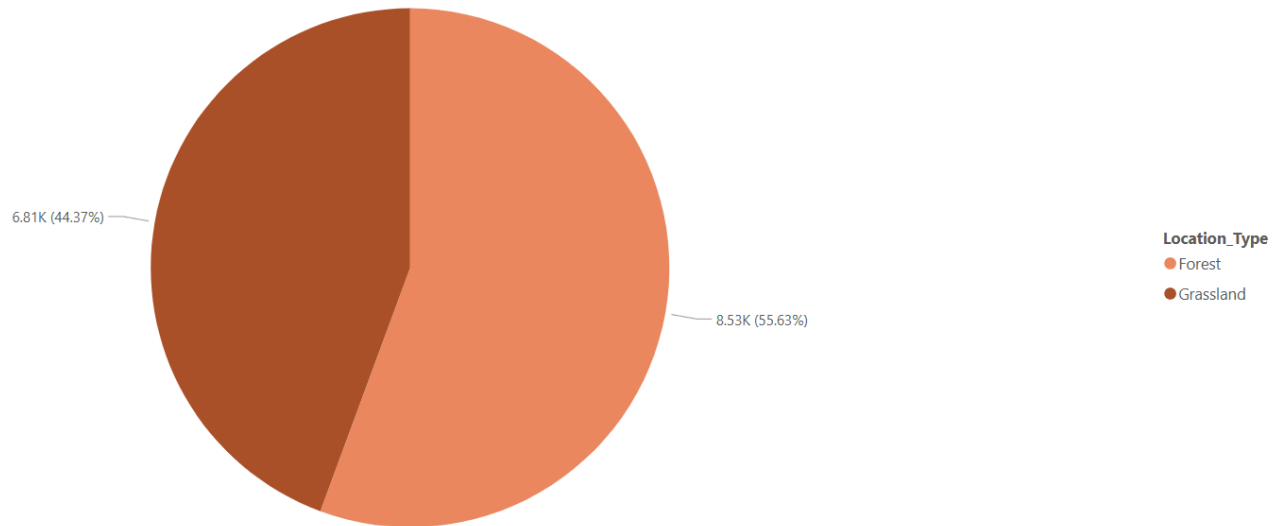
- Top 10 Most Common Species



Insights:

The horizontal bar graph displays the count of different bird species, with the **Northern Cardinal** having the highest count (around 1,150) and the **Acadian Flycatcher** having the lowest (around 450). The chart clearly shows the relative abundance of each species, with the top three being the **Northern Cardinal**, **Carolina Wren**, and **Red-eyed Vireo**.

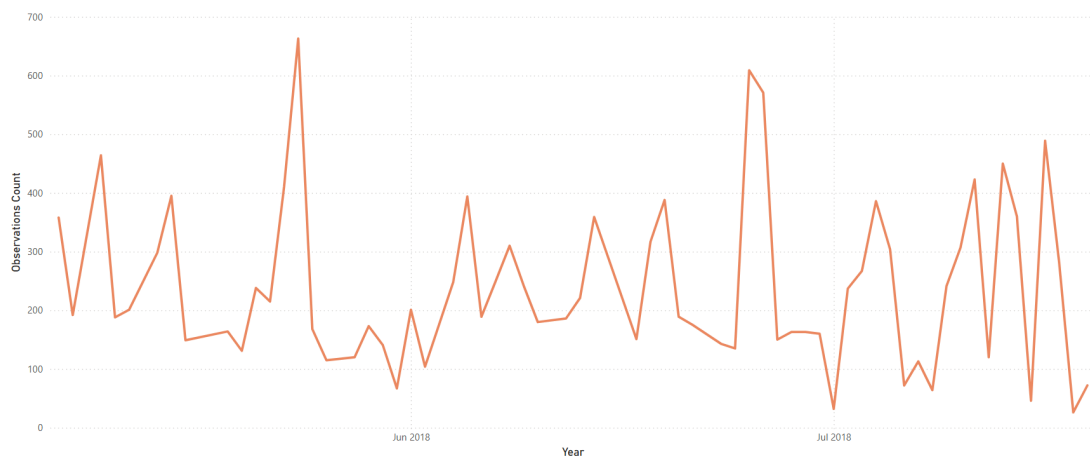
● Observations By Location Type



Insights:

The pie chart illustrates that **Forest** is the predominant location type, accounting for **55.63%** (**8.53K**) of the total. In comparison, **Grassland** makes up the remaining **44.37%** (**6.81K**).

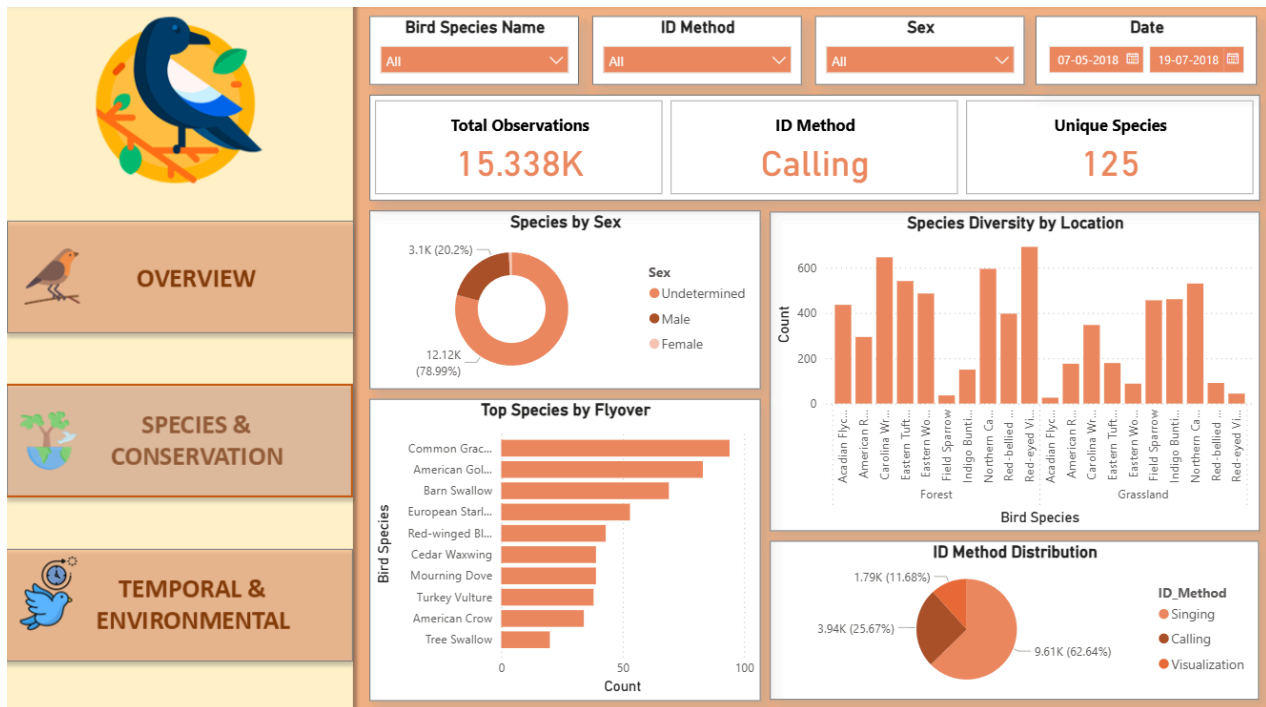
● Observations By Date



Insights:

The line graph displays a time series of observation counts, showing significant fluctuations over a period from early 2018 to late 2018. There are several sharp peaks, with the highest observation counts occurring in **early March** and **early July** of 2018, each surpassing **600**.

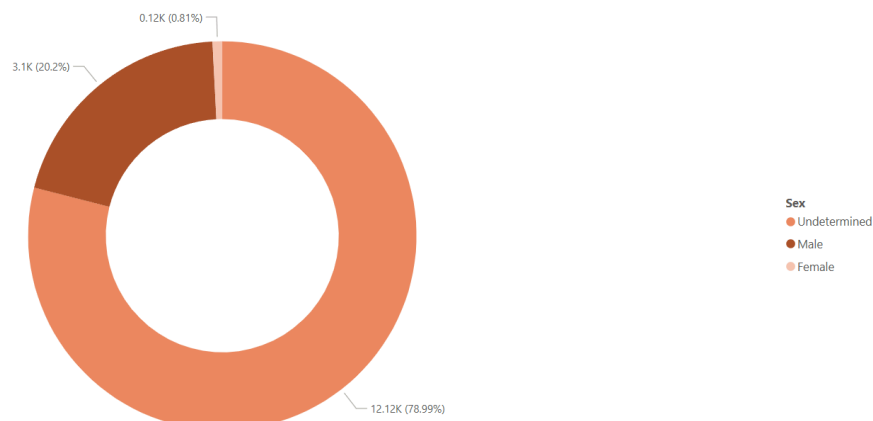
Page 2: Species and Conservation



- **KPI Cards:** Total Observations, ID Method, Unique Species
- **Slicers:** Bird Species Name, ID Method, Sex, Date

Visuals

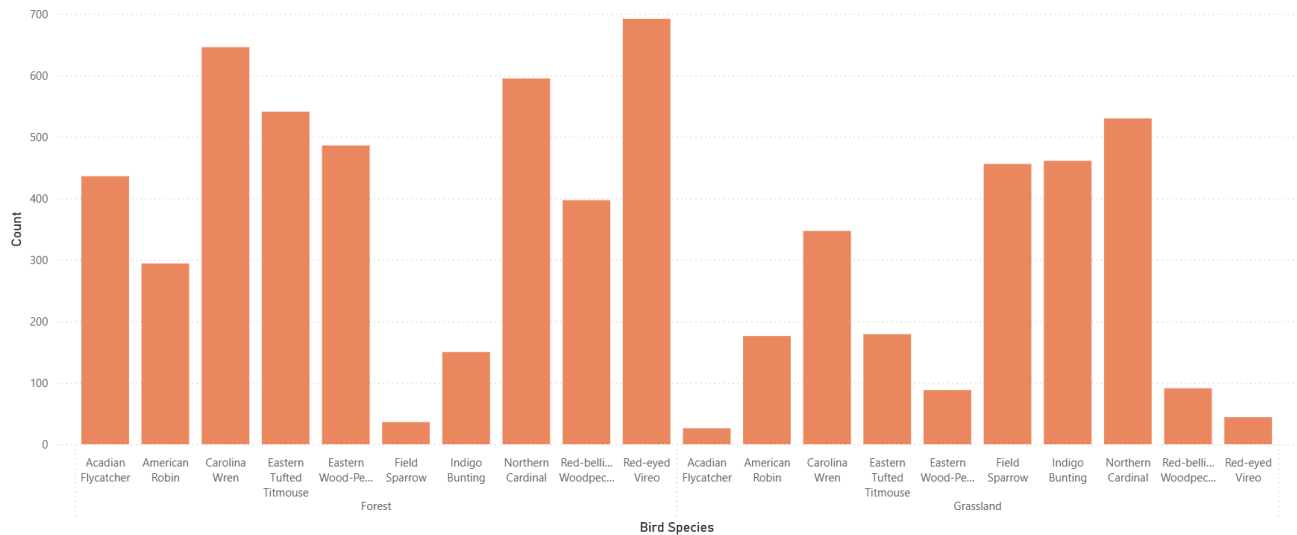
- **Species By Sex**



Insights:

The doughnut chart displays a breakdown of the population by sex. The majority of the population is **Male**, making up **78.99% (12.12K)**. **Female** and **Undetermined** categories account for a much smaller portion, at **20.2% (3.1K)** and **0.81% (0.12K)**, respectively.

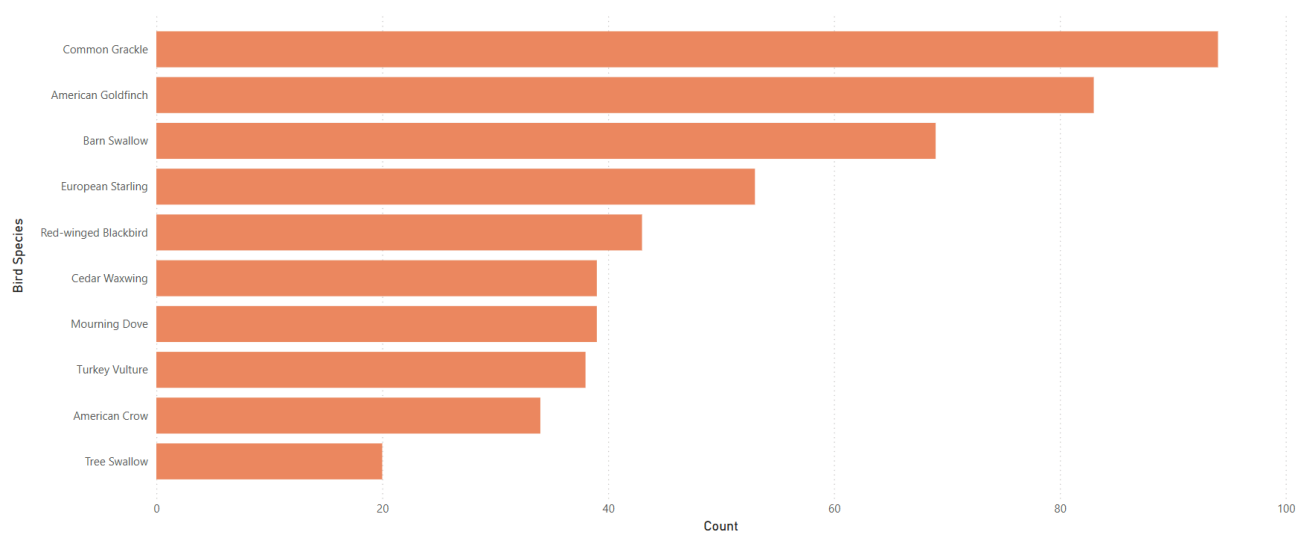
Species Diversity By Location



Insights:

The grouped bar chart compares the count of different bird species in Forest and Grassland habitats. The **Red-eyed Vireo** has the highest count in **Forests**, while the **Northern Cardinal** has the highest count in **Grasslands**. Most species show a preference for one habitat over the other, with significant counts in one group and very low counts in the other.

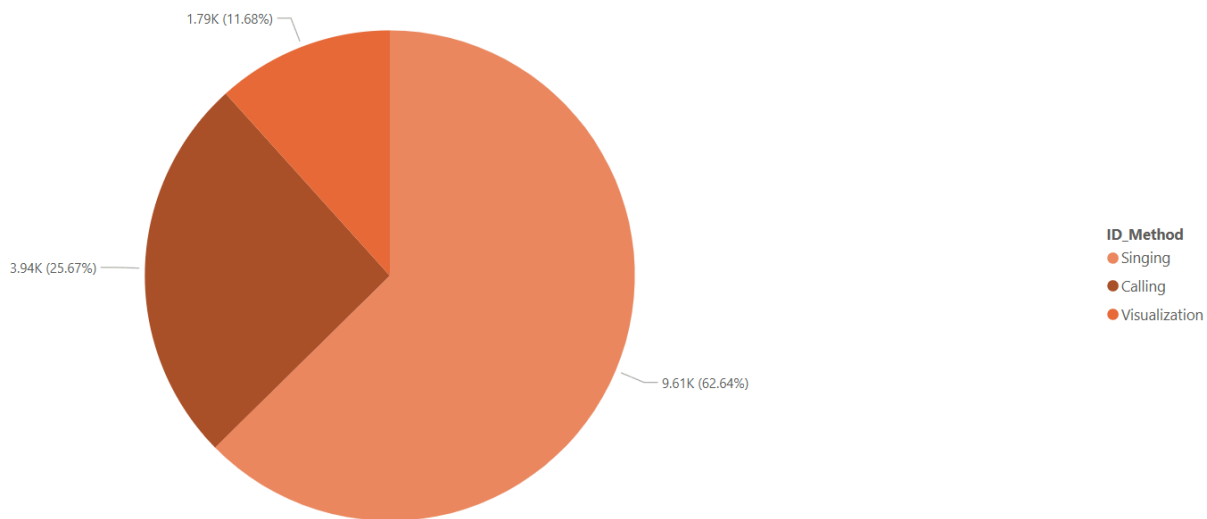
Top Species By Flyover



Insights:

The horizontal bar chart shows the count of different bird species. The **Common Grackle** has the highest count, approaching 100, while the **Tree Swallow** has the lowest count, at just over 20. The chart clearly ranks the species by their observed numbers, with the top three being the **Common Grackle**, **American Goldfinch**, and **Barn Swallow**.

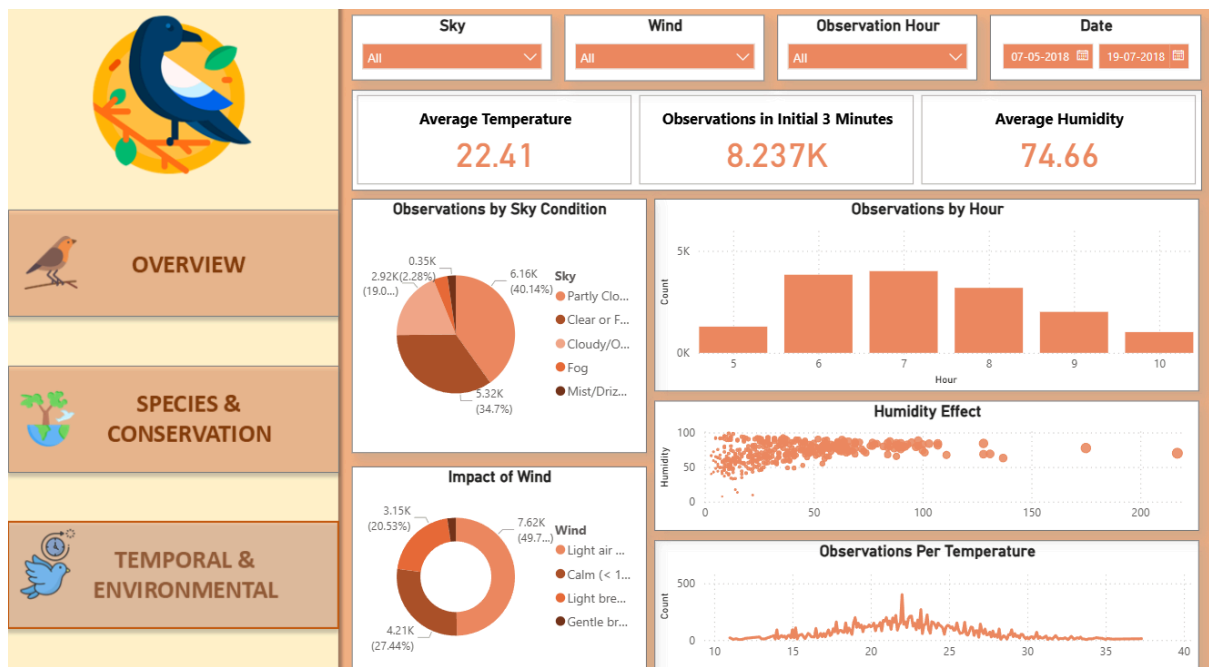
● ID Method Distribution



Insights:

The pie chart illustrates that **Singing** is the most common method of identification, comprising **62.64%** of the data. **Calling** and **Visualization** are used far less frequently, accounting for **25.67%** and **11.68%**, respectively.

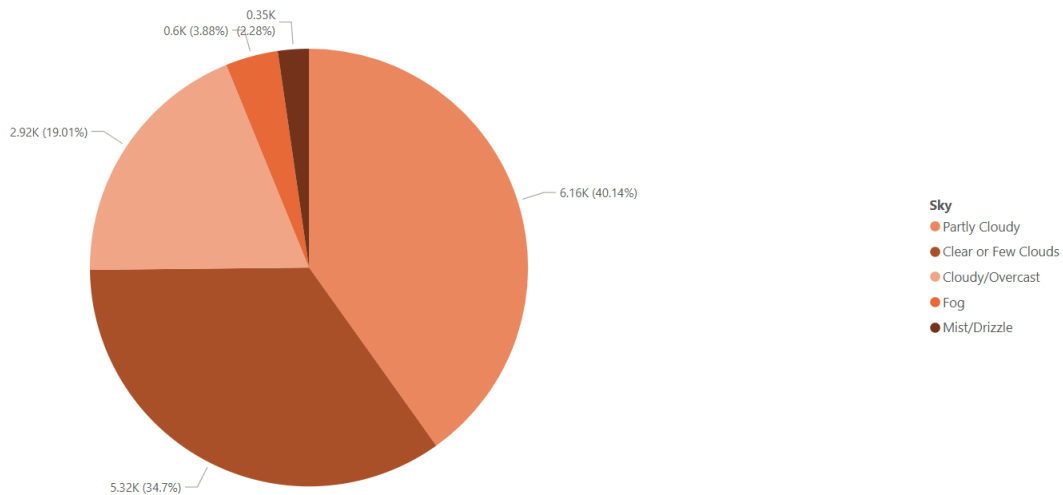
🐦 Page 3: Temporal And Environmental



- **KPI Cards:** Average Temperature, Observations in initial 3 Minutes, Average Humidity
- **Slicers:** Sky, Wind, Observations Hour, Date

Visuals:

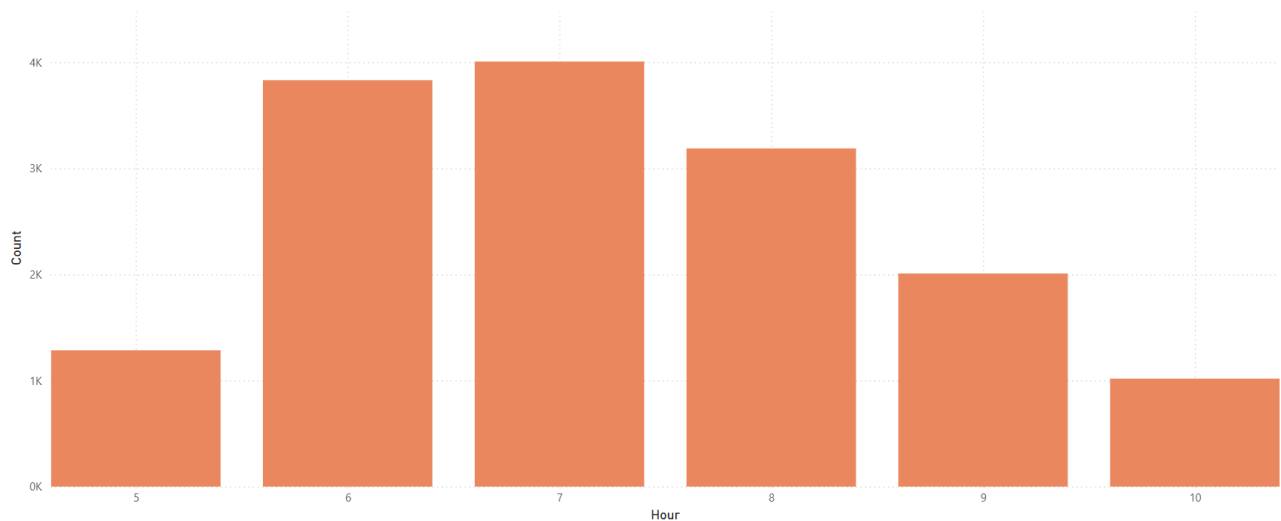
- **Observations By Sky Condition**



Insights:

The pie chart illustrates that **Partly Cloudy** conditions are the most prevalent, accounting for over **40%** of the total observations. **Clear or Few Clouds** and **Cloudy/Overcast** conditions make up the next largest portions at **34.7%** and **19.01%**, respectively, while **Fog** and **Mist/Drizzle** are rare.

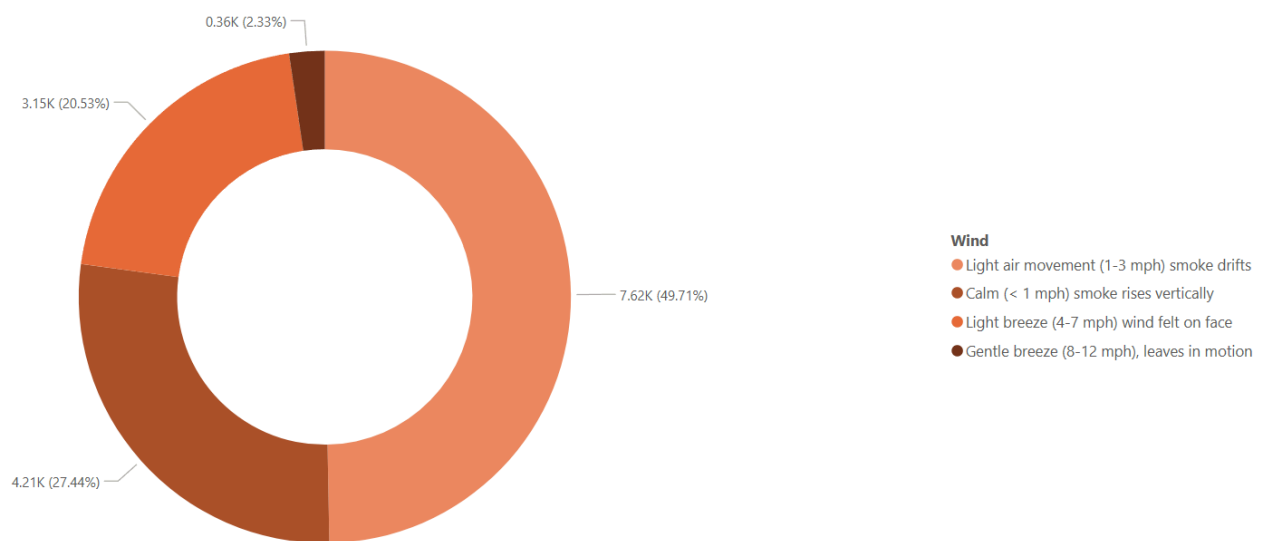
- **Observations By Hour**



Insights:

The bar chart displays observation counts by hour, indicating a peak in activity between **6 th Hour** and **7th Hour**, with a count of **4,000**. The number of observations steadily decreases after this peak, reaching its lowest point at **10th Hour** with a count of just over **1,000**.

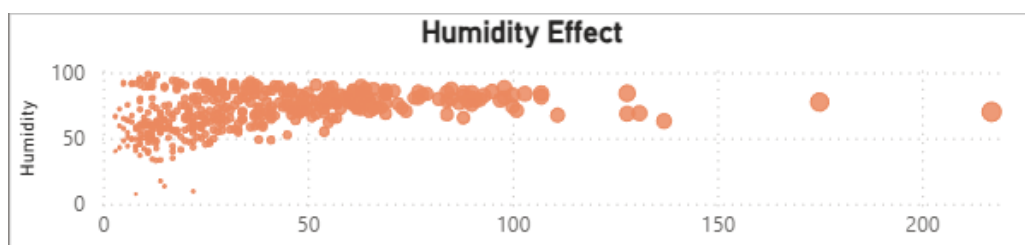
● Impact of Wind



Insights:

The doughnut chart illustrates that **light air movement (1-3 mph)** is the most common wind condition, accounting for nearly half of all observations at **49.71%**. **Calm (< 1 mph)** and **Light breeze (4-7 mph)** conditions are the next most frequent, at **27.44%** and **20.53%**, respectively, while a **Gentle breeze (8-12 mph)** is rare.

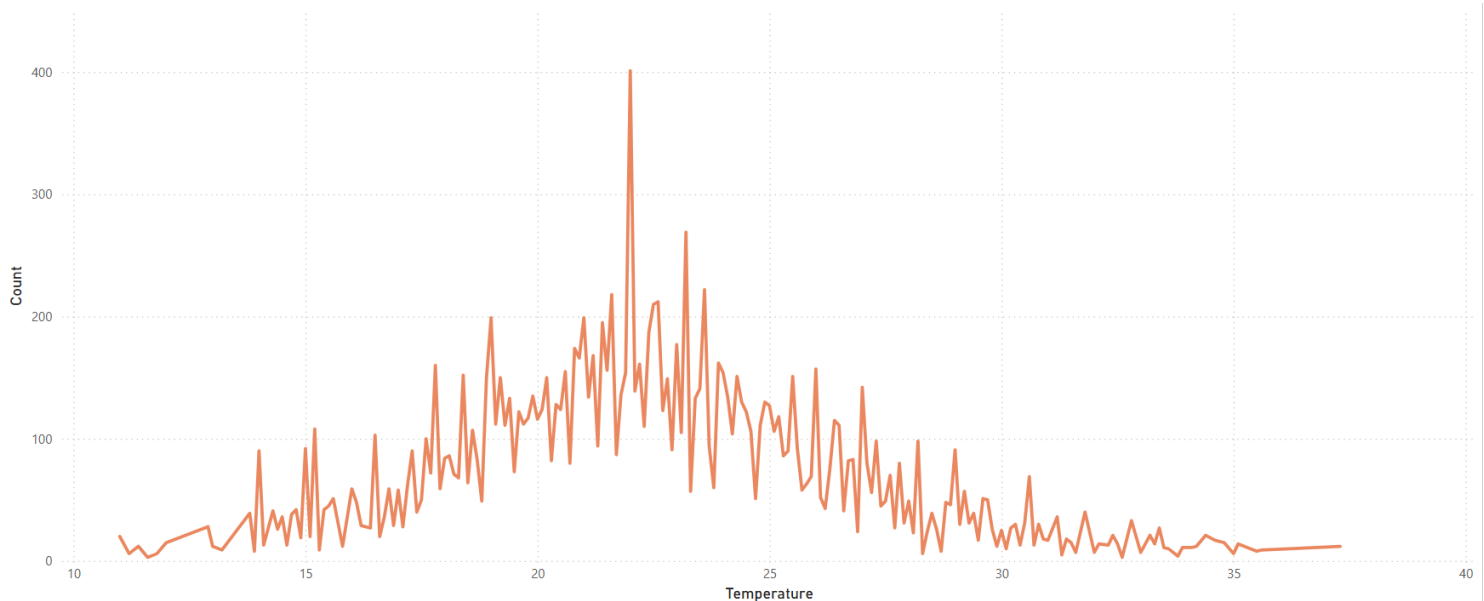
● Humidity Effect



Insights:

The scatter plot shows the relationship between two variables, with **humidity** on the y-axis. The data points are highly concentrated between 50% and 100% humidity, suggesting that most of the observations were recorded under conditions of high humidity. The larger size of some dots indicates that these data points represent a higher number of observations.

- **Observations Per Temperature**



Insights:

The line graph displays the frequency of observations across a range of temperatures, with the count peaking sharply at approximately **22°C** with a count of over **400**. The distribution is somewhat bell-shaped, showing that temperatures between **20°C** and **25°C** are the most common, while counts decrease significantly as temperatures move toward the extremes of the observed range (10°C to 40°C).

5. Recommendations

- **Improve Data Quality:** Reduce undetermined sex records (currently 79%) via better observer training.
- **Diversify ID Methods:** Promote visualization and calling along with singing to ensure accurate species identification.
- **Standardize Distance Records:** Minimize “Unknown” entries by ensuring consistent data collection.

- **Time Fieldwork Strategically:** Conduct observations between 6–8 hour when bird activity peaks.
 - **Habitat-Based Planning:** Allocate more resources to forests (55.6% of data) but monitor grasslands for underreported trends.
 - **Leverage Top Observers:** Recognize and support contributors like Elizabeth Oswald to maintain engagement.
 - **Environmental Targeting:** Schedule birdwatching in calm, slightly humid, partly cloudy weather (~22°C).
 - **Geographical Focus:** Prioritize biodiversity-rich zones like the U.S., Brazil, and India.
 - **Watchlist Monitoring:** Investigate even low-frequency species on the PIF Watchlist for early conservation signals.
 - **Observer-Habitat Alignment:** Assign observers based on their efficiency in specific habitat types.
-

6. Conclusion

This project presents an insightful analysis of bird species across forest and grassland habitats using observational data and advanced visualization. Key findings highlight species diversity, peak activity hours, environmental influences, and observer contributions. Forests host the majority of observations, with species like the Northern Cardinal and Carolina Wren dominating. Environmental factors such as humidity, temperature, and wind show significant influence on sightings. The Power BI dashboard effectively visualizes these trends. Strategic recommendations aim to enhance data quality, optimize observation timing, and inform conservation planning. This study contributes valuable guidance for biodiversity monitoring and eco-tourism strategies through data-driven decision-making.