Avian Journeys: Uncovering Global Migration Patterns Through Data Visualization

A Data Visualization Project

Drashti Shah, Dhruvkumar Kamleshbhai Patel, Sanjay Kanabhai Bharvad, Keerthika Loganathan

DATA 230 - Data Visualization, San Jose State University

Instructor: Dr. Venkata Duvvuri

Intro : Why We Follow the Flock Migration as a Vital Sign for Our Planet

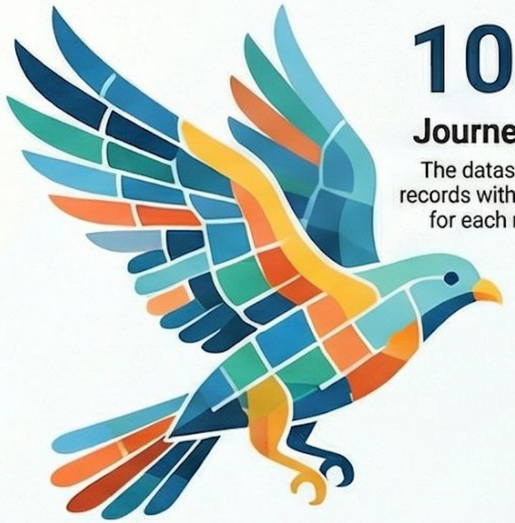
Bird migration is a critical indicator of global ecosystem health. Understanding migration routes is essential for:

- **Conservation Planning:** Identifying and protecting crucial habitats and stopover points to support species protection initiatives.
- **Climate Change Research:** Monitoring biodiversity and predicting how climate-related behavioral shifts alter age-old patterns.
- **Ecological Insight:** Revealing the intricate connections within our global ecosystem through animal movement data.

A Data-Driven Journey: Uncovering Bird Migration Patterns

Using Python for data exploration and Power BI for dashboarding to analyse global movement patterns.

THE MIGRATION DATASET AT A GLANCE



10,000

Journeys Analysed

The dataset contains 10,000 records with 42 distinct variables for each migratory journey.

High-Quality Data

The dataset is largely complete, with only one field ("Interrupted, Reason") having missing values (19.8%).

What's Being Tracked?

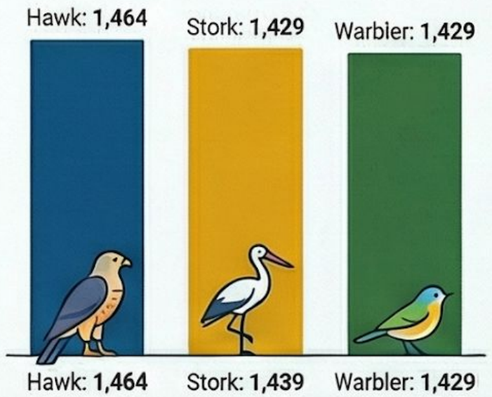
Species

Location

Environmental Conditions

Migration Success

Most Observed Species



CORE MIGRATION INSIGHTS

Survival Instincts Drive Migration

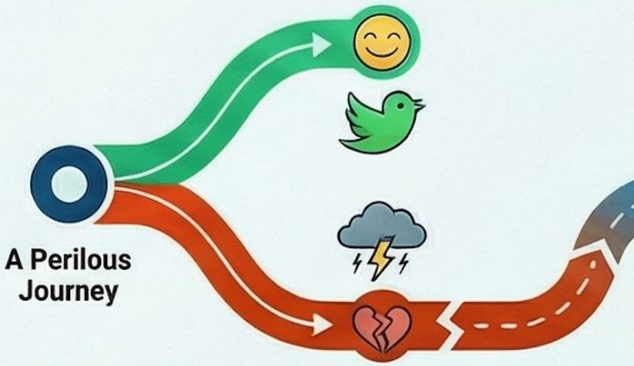
Feeding

Primary reasons are evenly split.

Avoiding Predators

Breeding

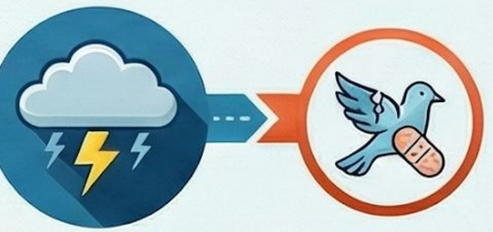
5,094
Successful Journeys (51%)



4,906
Failed Journeys (49%)

Top Threats

Storms and Injury are the Top Threats



Storms:
Leading Cause of Interruption

Injury:
Followed Closely

The Digital Footprints: Deconstructing the Bird Migration Dataset

Source
Bird Migration Dataset by Sahir Maharaj (2024)

Scale
10,000 records

Richness
42 distinct features

Key Data Categories

Spatiotemporal

'Start_Latitude', 'Start_Longitude', 'End_Latitude', 'End_Longitude', 'Flight_Distance_km'

Environmental

'Weather_Condition', 'Temperature_C', 'Humidity_%'

Behavioral

'Species', 'Migration_Reason', 'Flock_Size', 'Migration_Success'

```
df = pd.read_csv('./bird_migration_data.csv')
print("Shape:", df.shape)
display(df.head())
```

```
Shape: (10000, 41)
```

	Bird_ID	Species	Region	Habitat	Weather_Condition	Migration_Reason	Start_Latitude	Start_Longitude	End_Latitude	End_Longitude	...	Tracking_Quality	Migration_Interrupted	Interrupte
0	B1000	Warbler	South America	Grassland	Stormy	Feeding	11.906566	-169.378251	30.377647	-21.366879	...	Excellent	Yes	
1	B1001	Stork	North America	Grassland	Stormy	Breeding	62.301546	-111.475069	39.921092	47.963436	...	Good	Yes	
2	B1002	Hawk	South America	Mountain	Stormy	Avoid Predators	87.861164	-78.727327	66.990980	19.448466	...	Fair	No	
3	B1003	Warbler	South America	Urban	Stormy	Climate Change	35.770590	153.104341	-49.003145	-157.868744	...	Good	Yes	
4	B1004	Crane	Europe	Urban	Windy	Avoid Predators	-21.611614	106.674824	11.681051	-115.022863	...	Good	No	

5 rows × 41 columns

The Questions Guiding Our Quest

Our exploration was driven by three core questions aimed at transforming raw data into ecological understanding:

1. **What are the fundamental characteristics of these migration journeys?** (e.g., What species and regions are represented? How far do they typically fly?)

2. **Can we identify key factors that influence migration outcomes?** (e.g., Does weather impact interruptions? Do success rates vary by species or region?)

3. **How can we best visualize these complex spatiotemporal patterns for interactive exploration by stakeholders?**

Business Context : Why these questions matter

These questions reflect real challenges in wildlife conservation and environmental planning:

1. **Conservation teams need to understand which habitats or regions need protection and predict interruption risks.**

2. **Research labs want to track how environmental factors shape migration success.**

3. **Technology teams want to evaluate which tracking systems provide the best data.**

Our Blueprint: A Dual-Tool Approach from Analysis to Interaction

We executed a two-phase strategy to move from raw data to a dynamic, shareable story.

Clean Python

Phase 1: Deep Exploration with Python

- **Tools:** Pandas, Seaborn, Matplotlib.
- **Role:** Data cleaning, validation, and comprehensive Exploratory Data Analysis (EDA) to compute descriptive statistics and identify initial patterns.
- **Output:** Statistical summaries and static visualizations (histograms, boxplots, heatmaps).

Microsoft Power BI

Phase 2: Interactive Storytelling with Power BI

- **Tool:** Microsoft Power BI.
- **Role:** Importing the cleaned dataset to build an interactive dashboard with map visualizations.
- **Output:** A dynamic map with slicers for species, region, and season, enabling stakeholders to explore migration routes interactively.

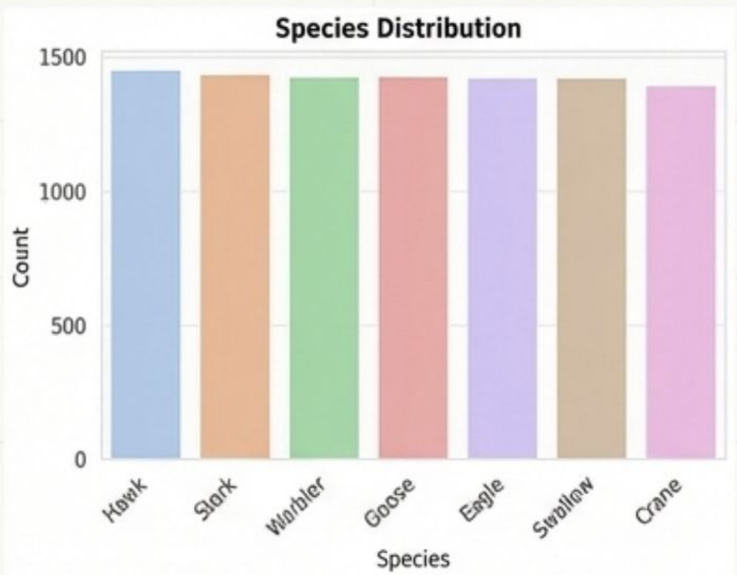
Exploratory Analysis Revealed Surprisingly Uniform Data Distributions

Demo

A deep dive into the dataset's characteristics surfaced several unexpected patterns across categorical, numerical, and geographical features.

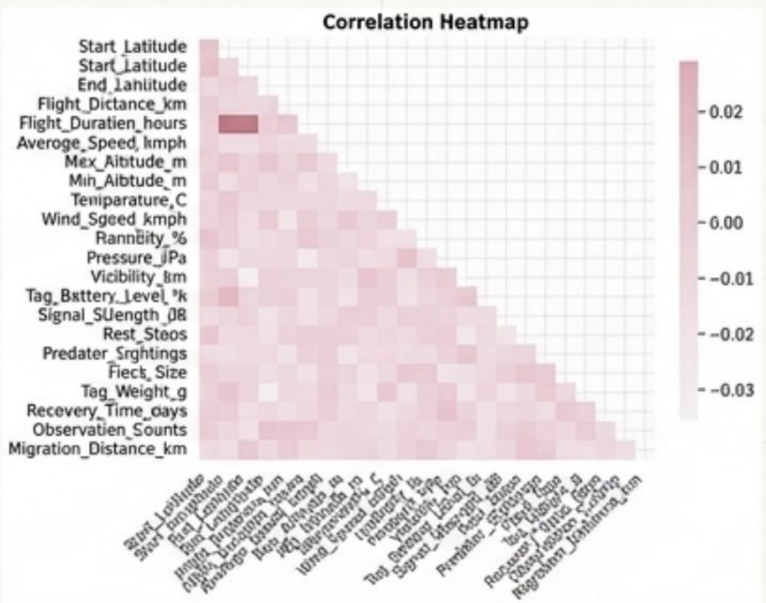
Uniform Categorical Distributions

The distribution of bird species is almost perfectly even. This flat distribution is also observed across regions, habitats, and weather conditions.



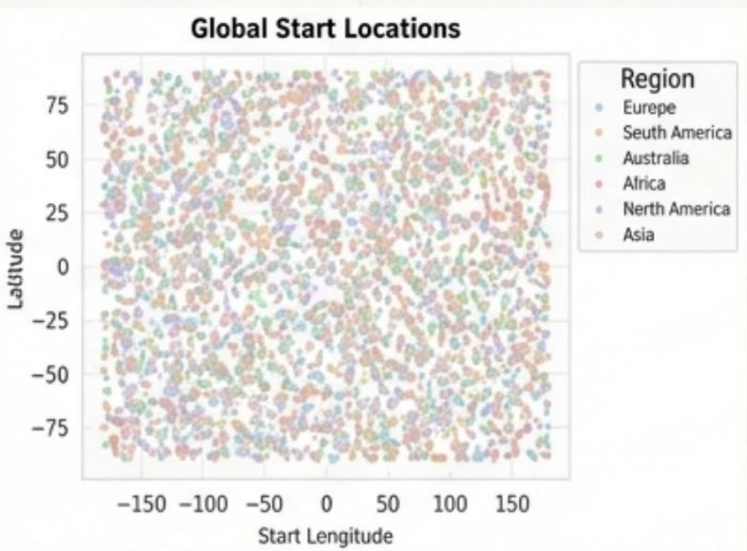
No Significant Correlations

The correlation heatmap of all 22 numerical variables shows an almost complete absence of strong relationships. Flight distance and duration show a moderate positive correlation as expected, but other links are negligible.



Random Geographic Spread

When plotting the start and end coordinates of the 10,000 migration records, the points are distributed uniformly across the globe in a near-perfect rectangular pattern, rather than clustering along known real-world migratory flyways.



POWER BI DEMO

Demo Link

What the Data Taught Us: Key Findings

Our analysis uncovered four primary insights about the patterns within this dataset.

A Perfectly Balanced World

The synthetically generated dataset represents all the species data distributed uniformly.

Success is the Norm, Universally

Migration success rates are consistently high (~51% Successful vs. 49% Failed) and do not significantly vary by species or region. The journey is perilous, but this data shows location is not a deciding factor.

Weather's Surprising Indifference

Counter-intuitively, migrations were interrupted at nearly the same rate across all weather conditions (Stormy, Windy, Clear, etc.). This challenges the simple assumption that bad weather is the primary cause of interruption.

Uniform Geographic Spread

Start and end locations are spread evenly across the globe rather than being concentrated in known migratory flyways, reinforcing the balanced nature of the dataset.

From 10,000 Rows to a Coherent Story

Objective Achieved

We successfully transformed a complex, 42-column dataset into a clear set of insights and an interactive visualization tool.

Process Validated

Our dual-methodology, leveraging Python for rigorous EDA and Power BI for accessible storytelling, proved highly effective for moving from exploration to communication.

Value Delivered

The project provides a static, evidence-based analysis of the dataset's unique characteristics and a dynamic dashboard for stakeholders to conduct their own explorations, answering our initial research questions.

Charting the Path Forward

While this project is complete, our analysis opens doors for further exploration:

Predictive Modeling

Develop a machine learning model to predict 'Migration_Success' based on environmental and behavioral factors, as originally scoped in our project plan.

Time-Series Analysis

Conduct a deeper dive into the 'Migration_Start_Month' and 'Migration_End_Month' columns to uncover seasonal patterns not explored in this initial EDA.

Dashboard Enhancement

Integrate external datasets, such as real-world protected ecological zones or historical weather data, to provide richer contextual insights within the Power BI dashboard.

References:

- Mahara, S. (2025). *Bird Migration Dataset – Data Visualization & EDA*. Kaggle. https://www.kaggle.com/datasets/sahirmaharajj/bird-migration-dataset-data-visualization-eda
- Microsoft. *Power BI Documentation*. Microsoft Learn. https://learn.microsoft.com/en-us/power-bi/
- Waskom, M. L. (2021). *Seaborn: Statistical Data Visualization*. https://seaborn.pydata.org/
- Freepik. Free Images & Graphic Resources. https://www.freepik.com/