Bird Migration Analysis

Akshitha Paidi
2211CS010011
S2-2

INTRODUCTION

Bird migration is a critical ecological phenomenon that influences biodiversity, ecosystem balance, and conservation planning. Understanding migration patterns, distances, durations, and the environmental conditions affecting them helps researchers, policymakers, and conservationists develop strategies to protect species and habitats. This study analyzes a dataset of bird migration records using PySpark-based data processing and visualization techniques. The aim is to identify species-specific migration trends, seasonal variations, and factors influencing successful migrations.

1. DATASET OVERVIEW

1.1 SOURCE

The dataset analyzed is bird_migration_with_origin_destination.csv, containing 10,000 records of tracked bird migrations. Analysis was conducted using PySpark for data handling and visualization.

1.2 COLUMNS / FEATURES

The dataset includes the following feature groups:

- Identification and Taxonomy: Bird_ID, Species
- Geographical Attributes: Region, Origin, Destination, Start_Latitude, Start_Longitude, End_Latitude, End_Longitude
- Migration Metrics: Flight_Distance_km, Flight_Duration_hours, Average_Speed_kmph, Max_Altitude_m, Min_Altitude_m, Rest_Stops, Flock_Size
- Environmental Factors: Temperature_C, Wind_Speed_kmph, Humidity_%, Pressure_hPa, Visibility_km, Weather_Condition, Habitat
- Tracking Data: Tag_Type, Tag_Battery_Level_%, Signal_Strength_dB, Tag_Weight_g, Tracking_Quality
- Outcomes and Metadata: Migration_Start_Month, Migration_End_Month, Migration_Success, Migration_Interrupted, Interrupted_Reason, Observation_Counts, Observation_Quality, Tagged_By, Recovery_Location_Known, Recovery_Time_days

1.3 DATA QUALITY

- Total records: 10,000
- Missing values: Interrupted_Reason has 1,981 nulls. Other fields show strong completeness.
- Duplicate records: Not reported in the analysis.
- Observation and tracking quality fields indicate reliability levels of the data.

1.4 KEY STATISTICS

- Total Records: 10,000 bird migration observations

- Species Distribution: Diverse representation, with Stork, Eagle, Hawk, Swallow, Warbler, Goose, and Crane among the most frequently tracked species

- Flight Distance: Ranges from 527.7 km to 4,428.32 km, with an overall average of 2,504.04 km

- Flight Duration: Varies between 12.6 hours and 91 hours, averaging 49.99 hours

- Average Speed: Between 30.43 km/h and 68.95 km/h, with a mean of 49.95 km/h

- Migration Start Months: Observed in January–May and September–November, indicating strong seasonal migration patterns

- Dataset : Comprehensive coverage of flight, environmental, and tracking attributes that enable detailed multi-dimensional analysis of migration behavior

2. OPERATIONS PERFORMED

2.1 DATA CLEANING AND EXPLORATION

- Loaded the dataset into Spark using spark.read.csv with header and schema inference.

- Verified schema and inspected column names and datatypes.

- Checked the total row count (10,000 records).

- Inspected sample records using .show() for initial understanding.

- Performed null value analysis for each column: identified that Interrupted_Reason contained 1,981 nulls, while other columns were complete.

- Removed duplicate records to ensure clean data.

2.2 DESCRIPTIVE ANALYSIS

- Computed statistical summaries (describe()) for key numerical columns such as Flight_Distance_km, Flight_Duration_hours, and Average_Speed_kmph.

- Generated histogram of Flight_Distance_km to analyze distance distribution.

- Created pie chart showing the top five bird species by record count.

- Produced line plot of average flight distance by migration start month.

- Developed scatter plot of Flight_Distance_km versus Flight_Duration_hours to observe correlation.

- Created bar chart for the top ten species by frequency of records.

2.3 RELATIONSHIPS ANALYSIS

- Species × Average Distance – Identified species with the longest average migration distances.

- Species × Maximum Distance – Highlighted the species with the greatest recorded migration (e.g., Stork at 4,428 km).

- Migration Start Month × Average Distance – Analyzed monthly variation in migration distances.

- Flight Distance × Duration – Observed a proportional relationship between migration distance and flight duration.

- Species × Frequency – Determined the most commonly observed species in the dataset.

3. KEY INSIGHTS/ FINDINGS

3.1 SPECIES-LEVEL MIGRATION PATTERNS

- Stork, Eagle, Hawk, Swallow, Warbler, Goose, and Crane emerged as dominant long-distance migratory species.

- Storks recorded the longest migration distance of 4,428 km, highlighting their endurance and adaptability.

- Species-specific patterns indicate ecological preferences and migratory strategies.

3.2 DISTANCE–DURATION–SPEED CONSISTENCY

- Strong correlation observed between migration distance and duration; longer journeys proportionally require more time.
- Average migration speed across species remained stable at approximately 50 km/h, reflecting reliable flight behavior.
- Consistent speed and duration patterns suggest efficient energy utilization during migration.

3.3 SEASONAL MIGRATION TRENDS

- Migration activity is concentrated in spring (Jan–May) and autumn (Sep–Nov).
- Seasonal peaks align with breeding, nesting, and resource availability, reflecting established global migration cycles.
- Some species exhibit highly predictable seasonal movement, enabling targeted monitoring.

3.4 FLIGHT DISTANCE DISTRIBUTION

- Majority of flights cluster around an average distance of 2,500 km, representing a balanced migration range.
- Extreme long-distance migrations (>4,000 km) are rare but biologically significant.
- Distribution analysis aids in understanding energy demands and flight endurance across species.

3.5 SPECIES FREQUENCY & REPRESENTATION

- A few species dominate the majority of recorded observations, providing robust data for detailed analysis.
- Dataset diversity enables comparative insights across species with different migratory strategies.
- Observational patterns support conservation prioritization and species-specific studies.

3.6 OVERALL INSIGHTS

- Bird migration is multifactorial, influenced by species, season, distance, and regional factors.
- Data visualizations help identify peak migration periods and high-activity species quickly.
- Insights can guide conservation planning, habitat monitoring, and protection of migratory routes.

4. RECOMMENDATIONS AND FUTURE ENHANCEMENTS

4.1 Geospatial Mapping of Migration Routes

- Visualize bird migration paths on maps.
- Identify important flyways and stopover hotspots.
- Support habitat conservation planning.

4.2 Integration of Environmental Data

- Combine dataset with external sources (temperature, precipitation, vegetation indices).
- Understand environmental drivers of migration success or interruption.
- Improve prediction models of bird movement patterns.

4.3 Predictive Modeling of Migration Success

- Use historical migration data to build machine learning models.
- Forecast migration timing, duration, and likely routes.
- Anticipate risks such as interruptions and unfavorable conditions.

5. CONCLUSION

The bird migration dataset provides valuable insights into migration distances, durations, and species-level variations. Storks, Eagles, and Hawks emerge as the most significant long-distance migrants, with average migration distances around 2,500 km. Seasonal trends confirm migration peaks in spring and autumn. The data shows good overall quality. Future enhancements through geospatial mapping, integration of environmental data, and predictive modeling will strengthen the dataset's role in advancing conservation strategies and ecological understanding.