

## Project 2

# Bird Species Observation Analysis

<b>Workflow:</b>
------------------

## 1. Define Goals

The objective of this project is to analyze bird sightings and seasonal trends using structured data. Key tasks include:

- **Extracting Data:** Retrieve bird monitoring data from various sources.
- **Storing Data:** Save the data in a **MySQL database** for structured analysis.
- **Visualizing Data:** Develop an **interactive Streamlit app** to explore seasonal trends and patterns in bird sightings.

## 2. Define Scope

### Project Tasks

- Fetching data from **bird monitoring sources** (CSV, API, or manual input).
- Storing structured data in a **MySQL database**.
- Building a **Streamlit dashboard** for data exploration and visualization.
- Applying **data cleaning techniques** to remove duplicates and missing values.
- Generating **seasonal trend insights** using **Matplotlib**.

### 3. Define Requirements

#### Functional Requirements

Feature	Description
Data Extraction	Read data from CSV files / APIs.
Database Storage	Store bird sightings, weather, and observer data in MySQL.
Data Cleaning	Remove duplicates, handle missing values, and standardize timestamps.
Seasonal Analysis	Identify trends based on months and locations.
Visualization	Generate interactive graphs using Matplotlib and Streamlit.

#### Technical Requirements

- **Backend:** Python (Pandas, MySQL Connector, Matplotlib)
- **Frontend:** Streamlit
- **Database:** MySQL
- **Deployment:** Local

## 4. Implementation Plan

### Step 1: Data Extraction

- Read bird monitoring data from Excel / CSV.
- Convert date columns to **datetime** format.

### Step 2: Data Storage

- Create **MySQL tables** for storing data.
- Insert cleaned and structured data.

### Step 3: Data Cleaning

- Drop **empty values and duplicates**.
- Standardize **date formats**.

### Step 4: Data Visualization

- **Seasonal Trends:** Bird sightings per month.
- **Species Analysis:** Frequency of bird species observed.
- **Location Insights:** Distribution of birds in **forest vs. grassland**.

### Step 5: Build the Streamlit App


- Allow users to **filter bird species and locations**.
- Display **interactive charts** for seasonal trends.
- Implement **search and filtering features**.

### Step 6: Deployment

- **Local Deployment:** Run using `streamlit run bird_monitoring_dashboard.py`

## Schema Design

Schema design is crucial for structuring data and ensuring efficiency. It includes:

 **Data Modeling** – Designing tables, relationships, and entities.

- In this project, I have carefully **designed the tables** to reflect the relationships between competitions, events, and players.
- I also ensured that each table had the appropriate **primary keys** and **foreign keys** to maintain data integrity.

## Challenges Faced in the Project

### 1. Learning New Technologies

- This being my first project, I spent significant time learning **Streamlit**, **MySQL**
- Retrieving data from Excel was bit tough as it is new.
- Creating graphs was also a challenge.

### 2. Frontend: Implementing Multiple Filters

- Building an interactive **Streamlit** interface with **multiple filters** (such as SelectBox) posed challenges.
- Each filter required me to write separate queries and ensure the data updated dynamically with each user interaction.

### 3. Database Queries

- Writing efficient **SQL queries** for dynamic data retrieval, particularly when applying multiple filters, was quite complex.
- Ensuring that the queries were **optimized** for performance, especially with large datasets, and managing relationships between tables (e.g., competitions, complexes, and competitors) proved to be a challenging task.

## Insights and Lessons Learned from the Project

1. **Comfort with Streamlit** – Since this was my second project, using Streamlit was much easier.
2. **Better Data Cleaning** – Handled missing values, removed duplicates, and standardized timestamps efficiently.
3. **Optimized SQL Queries** – Improved performance by writing better queries and handling table relationships effectively.
4. **Advanced Data Visualization** – Used Matplotlib to analyze seasonal bird trends clearly.
5. **Faster Development** – Prior experience made the process smoother and more efficient.