Project - Penguin Dataset

Dataset:

Click here for the dataset link

Data Description:

The dataset contains 7 columns describing penguins from the Palmer Archipelago in Antarctica. It includes information on species, physical measurements, location, and sex.

1. **species:** The penguin species, categorized as Chinstrap, Adélie, or Gentoo.
2. **culmen_length_mm:** Length of the penguin's culmen (beak) in millimeters (mm).
3. **culmen_depth_mm:** Depth of the penguin's culmen (beak) in millimeters (mm).
4. **flipper_length_mm:** Length of the penguin's flipper in millimeters (mm).
5. **body_mass_g:** Body mass of the penguin in grams (g).
6. **island:** Name of the island where the penguin was found, either Dream, Torgersen, or Biscoe.
7. **sex**: Sex of the penguin, either male or female (may include missing values).

Sprint 0: Perform EDA With Basic Plots

1. **Understand the Data:** Familiarize yourself with the dataset and its attributes.
2. **Define the Problem Statement:** Describe the primary question(s) this dataset can help you answer.
3. **Data Cleaning and Manipulation:** Handle missing values, correct data types, and resolve other data quality issues.
4. **Identify Data Types:** Classify attributes as discrete, continuous, or ordinal, and identify relationships among them.
5. **Formulate Analysis Questions:** Create specific questions to guide your analysis based on the problem statement.
6. **Visualization:**
    - Explore the distribution of numerical and categorical variables.
    - Identify apparent trends or patterns.
    - Investigate relationships between variables.
    - Identify potential correlations and interactions.
    - Visualize relationships and correlations among multiple variables.

Sprint 1: Preprocessing

1. **Data Preparation:**
    - Split the dataset into training and testing sets.
    - Perform feature engineering.
    - Conduct data preprocessing

Sprint 2: Model Development

1. **Model Training:** Train machine learning models
2. Evaluate model performance using metrics.

Sprint 3: Model Deployment

- **Deployment:** Deploy the model using your preferred method.

Submission:

- Submit your Python notebook (.ipynb) on the LMS, including a summary of your process and results.