Data Preprocessing

- Missing values were imputed using KNN imputation for numerical columns and mode imputation for categorical data.

Univariate Analysis

1. **Species Distribution**
    - The dataset contains three species: Adelie, Chinstrap, and Gentoo.
    - Adelie penguins are the most common, while Chinstrap penguins are the least represented.
2. **Island Distribution**
    - Penguins are distributed across three islands: Torgersen, Biscoe, and Dream.
    - Gentoo penguins are found exclusively on Biscoe Island, while Chinstrap penguins are only on Dream Island.
3. **Numerical Feature Distribution**
    - Bill length follows a **right-skewed distribution**, meaning most penguins have shorter bills, but some have exceptionally long ones.
    - Bill depth is **normally distributed**, indicating balanced variation.
    - Flipper length and body mass exhibit **bimodal distributions**, suggesting distinct size variations among species.
4. **Outliers**
    - Boxplots show that Gentoo penguins have significantly larger body masses compared to Adelie and Chinstrap.
    - Some extreme values in bill length and flipper length may indicate measurement errors or unique biological variations.
5. **Categorical Feature Distribution**
    - The sex ratio is almost balanced, with a slight male dominance.
    - Pie charts confirm that Gentoo penguins are the largest-bodied species, while Adelie penguins are the smallest.

Bivariate Analysis

1. **Pairplot Analysis**
    - Gentoo penguins tend to have longer flippers and heavier bodies compared to the other species.

    - Bill length and flipper length show a positive correlation, meaning penguins with longer bills tend to have longer flippers.
2. **Correlation Heatmap**

    - **Strong positive correlation (0.87) between flipper length and body mass** indicates that larger penguins generally have longer flippers.
    - **Weak correlation between bill depth and body mass** suggests that bill depth is not a strong predictor of a penguin's overall size.
    - **Negative correlation (-0.32) between bill depth and bill length** implies that penguins with longer bills tend to have shallower bills.

Multivariate Analysis

1. **Pairplot with Hue**
    - The distinction between species is evident in bill length vs. bill depth plots, where each species occupies a distinct cluster.
    - Chinstrap penguins tend to have the longest bills, while Adelie penguins have the shortest.
    - The separation in flipper length and body mass further highlights species differences.

Summary of Trends & Insights

- **Species and island distribution trends suggest distinct habitat preferences** (Gentoo on Biscoe, Chinstrap on Dream).
- **Penguins with longer flippers tend to be heavier**, which aligns with evolutionary adaptations for swimming.
- **Bill depth and length show an inverse relationship**, meaning species with long bills tend to have shallower ones.
- **Gentoo penguins are the largest and heaviest, while Adelie are the smallest**, reflecting dietary and environmental adaptations.