

Comprehensive Data Analysis and Statistical Exploration

Question: Data Loading and Preliminary Analysis Load the "penguins.csv" dataset into a DataFrame. Display the first few rows of the dataset and summarize its structure, including the number of rows/columns and data types. Data Cleaning and Preprocessing Check for and handle missing values in the dataset. Exploratory Data Analysis.

1. Data Loading and Preliminary Analysis:

- Load the "penguins.csv" dataset into a DataFrame.
- Display the first few rows of the dataset and summarize its structure, including the number of rows/columns and data types.

2. Data Cleaning and Preprocessing: Check for and handle missing values in the dataset & Exploratory Data Analysis

- 3. Create visualizations:
 - Histograms or box plots for the distribution of numerical features.
- Bar charts for categorical attributes.
- Pair plots or scatter plots to explore relationships between features.

4. Advanced Statistical Analysis:

- Compute the Multivariate Mean Vector: Calculate the mean for each attribute in the dataset.
- Compute the Sample Covariance Matrix (Inner Product): Use the inner product between the columns of the centered data matrix (Numerical Attributes only).
- Compute the Sample Covariance Matrix (Outer Product): Use the outer product between the centered data points (Numerical Attributes only).
- Compute the Correlation Between Two Attributes (bill_length_mm, bill_depth_mm)
- Compute the correlation between the two attributes (bill_length_mm, bill_depth_mm) by calculating the cosine of the angle between the centered attribute vectors. Include a scatter plot of these two attributes.
- Compute the Covariance matrix of the three categorical attributes (species, island, and sex).
- Probability Density Function: Assuming the first numerical attribute is normally distributed, plot its probability density function.
- Empirical CDFs: Plot the Empirical Cumulative Distribution Functions (CDFs) of attributes bill_length_mm, bill_depth_mm, and body_mass_g..
- Variance Analysis: Identify the numerical attribute with the largest variance and the one with the smallest variance. Print these values.
- Covariance Analysis: Determine which pair of attributes has the largest and smallest covariance. Print these values.

