DAI-101 Asgn-1 Report

Palmer Archipelago Penguin Dataset EDA

Name: Kunal Virwal
Enrollment no.: 23116049
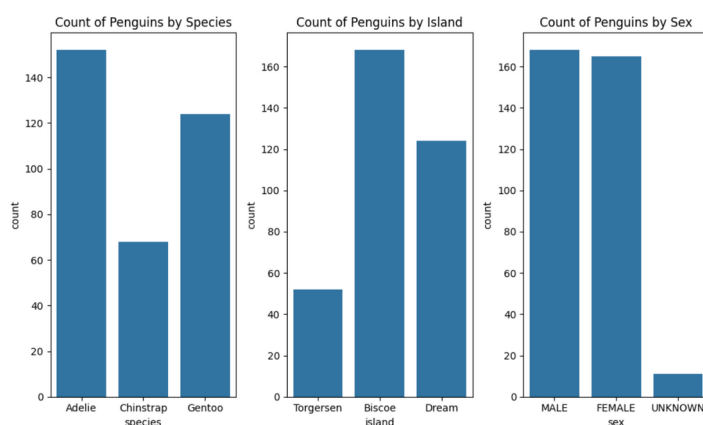Branch: Electronic and Communication

Introduction

This report provides a comprehensive exploratory data analysis of the Palmer Archipelago Penguin Dataset . Data were collected and made available by Dr. Kristen Gorman and the Palmer Station, Antarctica LTER, a member of the Long Term Ecological Research Network. The data examines morphological characteristics across three penguin species namely Adelie, Chinstrap and Gentoo. The data investigates four key measurements-culmen length, culmen depth, flipper length, and body mass—and explores how these traits vary across species and between sexes.
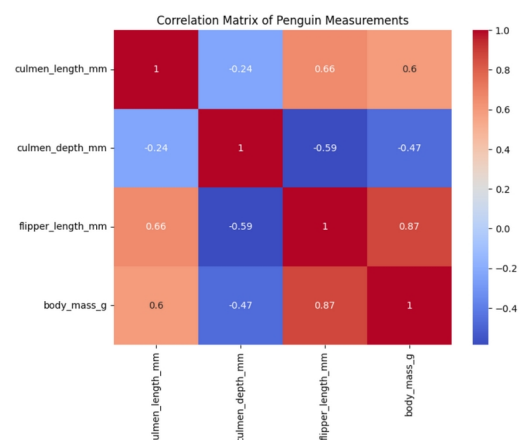
Data Cleaning

The data had to be cleaned before extracting any meaningful insights from it. This included the following changes to the data

- The numerical null values were replaced by column medians. This can also be done using mean but medians is a better choice as their values are unaffected by existance of outliers.
- Any categorical null values have to be replaced by a valid value or with UNKOWN as it were done for gender in this case
- Any data points which lied outside the 1.5*Interquatile range of the 25% and 75% quartiles are considered as outliers and are rounded to the the minum or maximum allowed values of the column.
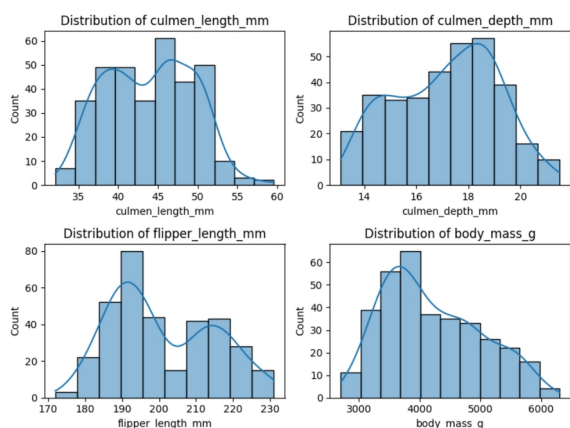

Count of Penguins Per category

Pairwise Heatmap of Numerical Arttributes
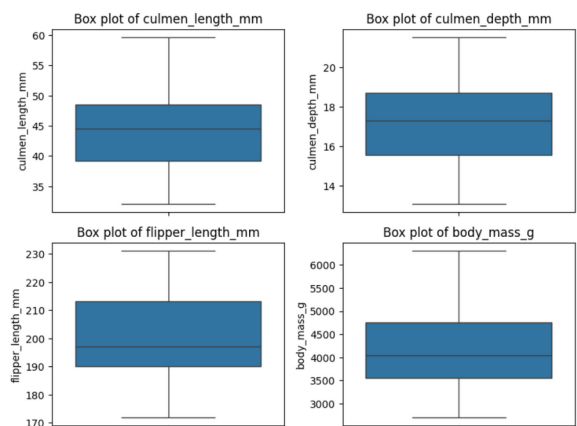
Exploratory Data Analysis

1. **Univariate analysis:**
    Performing univariate analysis include analyzing trends in the values of individual variables which are culmen length, culmen depth, flipper length, body mass, species, sex and island they live on. Trends like mean, median, mode and variance.

    - **Culmen Length:** The measurements vary from 32 mm – 60 mm. Adelie penguins show the shortest culmen lengths (centered around 38 mm - 40 mm , while Chinstrap and Gentoo penguins exhibit longer bills around 45 mm – 50 mm. Males consistently display longer culmen lengths than females across all species.

    - **Culmen Depth:** The culmen depth distribution ranges from approximately 13 mm – 22 mm. The distribution looks bimodal which primarily reflects the difference between Gentoo penguins with shallower culmens around 14 mm -16 mm versus Adelie and Chinstrap penguins with deeper culmens around 17 mm – 20 mm. In all species males have deeper bills than females, though this difference is less pronounced in Gentoo penguins.

    - **Flipper Length:** Flipper length exhibits the most distinctive separation between species, ranging from approximately 170 mm – 230 mm. Gentoo penguins have substantially longer flippers around 210 mm -230 mm compared to Adelie around 185 mm – 195 mm and Chinstrap 190 mm – 205 mm. Males consistently have longer flippers than females across all species, with the greatest difference observed in Gentoo penguins.

    - **Body Mass:** Body mass shows the widest variation, ranging from approximately 2700g - 6300g with a right-skewed distribution. Gentoo penguins are substantially heavier around 4500g – 6000g than both Adelie and Chinstrap penguins which are around 3000g – 4000g.
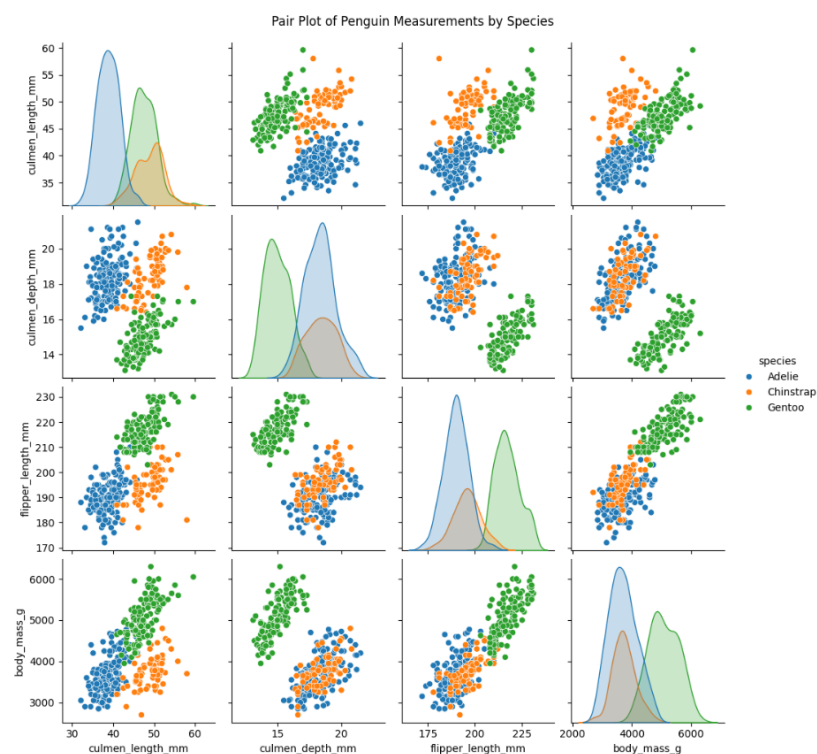

Bucketwise Bar Graphs for Each Numerical Attributes

Boxplots for Numerical Attributes

2. **Multivariate analysis:**
   Performing multivariate analysis includes analysing mutual trennds using correlation marices and scatter plots. This analysis helps us recognise relationships between factors, detect direct or inverse proportionalities and deduce their reasons. Some major pairs of attributes which reveal usefull insights are:

   - **Culmen Length vs Culmen Depth:** Gentoo penguins have the longest and shallowest culmens whereas Adelie penguins have the smallest but deepest culmens. Chinstrap penguins fall right in between with long and deep culmens.

   - **Island vs Species:** The bar plots of Island vs species give striking results that Gentoo Penguin population with a count of around 120 are completety located at the Biscoe Island. The Chinstrap population around 65 are also completely located at the Dream Island. The Adilie population of around 130 unlike the other two is uniformly distributed among the 3 islands.

   - **Adelie, Chinstrap vs Gentoo:** By plotting the pairpot for the data and analyzing the graphs we can easily deduce that the Adelie and Chinstrap penguin species are very mich identical with similar culmen depths, body weights and flipper lengths whereas the Gentoo species are considerably different with high flipper lengths and body weight andlower culmen depths.

Pair Plot of Penguin Measurements by Species

Pairplot for Numerical Attributes

Conclusion

This exploratory data analysis of the Palmer Archipelago penguin dataset reveals distinct morphological patterns across Adelie, Chinstrap, and Gentoo penguin species. The analysis demonstrates clear species differentiation through multiple physical measurements, with each species exhibiting unique combinations of culmen dimensions, flipper length, and body mass that likely reflect their evolutionary adaptations to specific ecological niches.
The multivariate analysis uncovered increasing levels of morphological integration from Adelie to Chinstrap to Gentoo penguins, suggesting different evolutionary pressures and adaptation strategies. Consistent sexual dimorphism was observed across all species, with males exhibiting larger measurements than females in all four morphological traits, though the magnitude varies by species and trait.

More data and insights on Univariate, Bivariate and Multivariate Analysis of the Palmer archipelago penguine dataset are available in main.ipynb provided along with this report.