Final Project: Statistical Analysis and Shiny App Development

Project Overview

This project challenges students to apply Advanced statistical analysis techniques to the palmerpenguins data set, an inbuilt data set within R. find the correct package and load the data. must make a separate GitHub repository and I want to see all the codes used for cleanup as well.

Tasks and Deliverables

Task 1: Data preprocessing

- Load and inspect the palmerpenguins dataset.

- Handle missing values

  - Impute missing numerical values using the median but you have to do this logically you cannot just replace all the missing values with the median. Maybe take the median by species and by variable. you have to explain your thought process.

  - impute missing categorical Values using a proper method one possibility is to perform K means clustering and then find the categorization or simply replace by mode provide you justification and how you replace the missing values if any

- Generate summary statistics and initial visualizations such as box plots and Scatter Plots. there should be an extensive exploratory data analysis. go variable by variable then see if you can visually discover any patterns or relationships between variables. comment on your findings.

- This visualization needs to be implemented in the shiny app. Meaning the app will have a function that prompts the user to press a button which will do the analysis and the visualizations and summaries will be shown.
- Add a column modifier to the shiny app. That means I need a place to create a new column by doing modifications to the existing column. example: assume you have weight and height, I might want to create a new variable called bmi. Or there could be a categorical variable based on which you could create a new column. for any example: if attendance == high and age > 90 then new variable grade = 5
- Deliverable: submit an R script for data preprocessing with comments

Task 2: Model Building Statistical Analysis

- Model building: so basically you should create a space to build a predictive model. similar to the linear regression app that we made. but should be more sophisticated. space to select response and explanatory variables. Then check the model assumptions. if the assumptions are not satisfied do an appropriate transformation. further, make sure there's a mechanism to build multiple models and compare then using Adj R^2, AIC and BIC. Read about these comparison Critera
- This needs to be implemented in the shiny app

Task 3: Dimensionality reduction and clustering

- Create a shiny app with the following features
  - Filters for species Island and sex
  - Visualizations for exploratory analysis (e.g., Scatter plots, box plots)

- ○ Interactive PCA and clustering visualizations

- ○ Dynamic summary statistics and model outputs

- ● publish the app on shinyapps.io

- ● Deliverable: provide a link to the shiny app

Task 5: Github Repository

- ● Create a GitHub repository for your project

- ● Include

  - ○ all R scripts and data sets

  - ○ a detailed README file with project description and instructions

  - ○ link to your shiny app

- ● example GitHub repository:

  - ○ https://github.com/zzactag/US-Car-Accidents-Shiny-App?tab=readme-ov-file

- ● Example Shiny App

  - ○ https://zmactag.shinyapps.io/US_Accidents_Sample_Test/

- ● Deliverable: Submit the Github Repository Link