# CHILLING STATISTICS:
## A DEEP DIVE INTO PENGUIN DATA

*By Rucha Deo*

## INTRODUCTION

The Palmer penguins dataset is about three penguin species in Antarctica's Palmer Archipelago. Data were collected and made available by Dr. Kristen Gorman and the Palmer Station, Antarctica LTER.
The Objectives of project are:

1.Showcase the distribution of physical traits across species and islands,

2.Explore the relationships between various morphometric traits and how they influence one another,

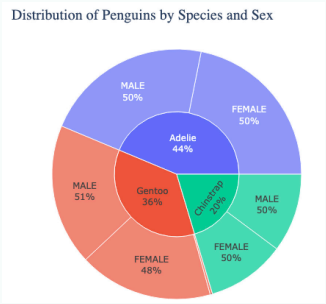
3.Predict body mass based on linear measurements.

## METHODS

Tools used for this project are Python libraries like Pandas, Matplotlib, Seaborn, Plotly, Scikitlearn, and Numpy. I standardized and cleaned the data, followed by statistical analyses to understand species distribution and physical traits. Predictive modeling helped explore relationships between traits, validated by cross-validation. Visualizations such as violin plots and scatter plots were used to present the findings.

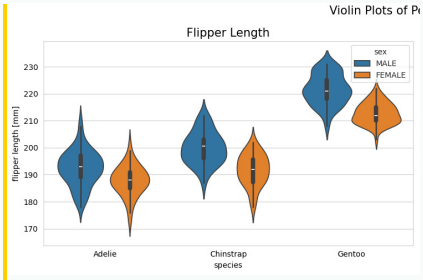
## DATA

The Palmer penguins dataset is a collection of data about three penguin species in Antarctica's Palmer Archipelago. The dataset has 7 columns including Sex, Flipper Length, Culmen Leng, Culmen Depth, body mass, island and species.




## ANALYSIS

Adelie penguins are the most common, accounting for approximately 43.7% of the sample, followed by Gentoo and Chinstrap penguins. Torgersen Island supports the highest number of penguins, with nearly half of the total population observed there. The sex distribution within the dataset is nearly balanced, with males slightly outnumbering females. The Violin plots of all the features show that the average fipper length(shown here), body mass and culmen length is higher in males thn in females. The scatter plot showed a visible positive correlation between body mass and Flipper length.




## RESULTS

Results of linear regression model Flipper length as a feature and Body mass as target
**********

Training Data R-squared: 76.04%
Test Data R-squared: 77.04%
Mean Absolute Error: 300.12 g
Root Mean Squared Error: 374.51 g
Coefficients: [50.46746483]
Intercept: -5931.563273314287

## CONCLUSION

While there is some variation in the spread of body masses compared to the lenghth of the flipper on a penguin, there is a relatively strong correlation. Using this measurement to estimate the mass of a penguin would be within a reasonable margin of error on most estimates.
The correlation differences across species might reflect different ecological strategies, growth patterns, or evolutionary histories.These insights can help biologists and ecologists understand how different physical traits are interrelated within each species and can inform studies on the ecological roles and adaptive strategies of these penguins in their respective environments.