

Abalone Analysis

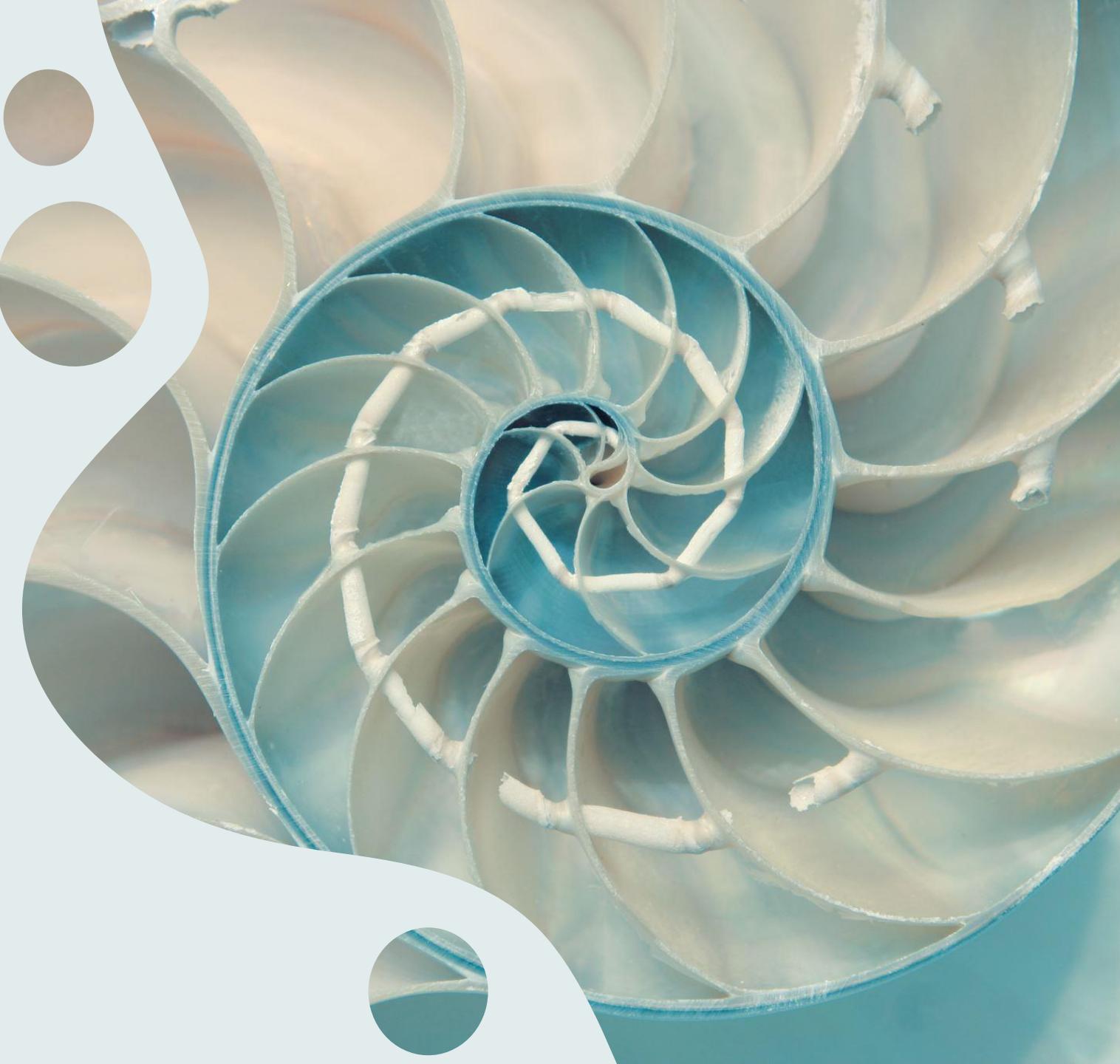
DDS 6306

Presented by:

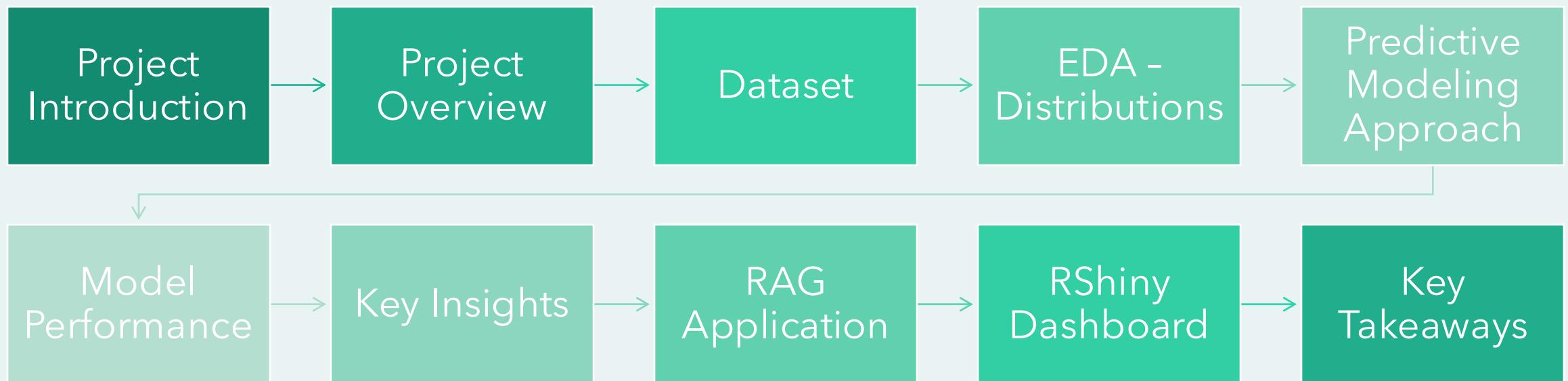
Nashita Khandaker

&

Chloe Prowse



Agenda





Project Introduction

Problem Context

- Marine growth and environmental impacts
- Manual, time-consuming, and error-prone
- Data science and AI = scalable, interpretable alternative

Goal

Develop an end-to-end solution that predicts abalone age and enables natural-language exploration of abalone data using AI

Project Overview

What We Built

- **Predictive Model:** Linear regression to estimate abalone age, optimized for low MAE
- **Data Insights:** Exploratory analysis to identify key features influencing age
- **RAG Application:** Natural-language Q&A system for abalone data
- **Deployment:** Models, dashboards, and apps via GitHub and Hugging Face Spaces

Why It Matters

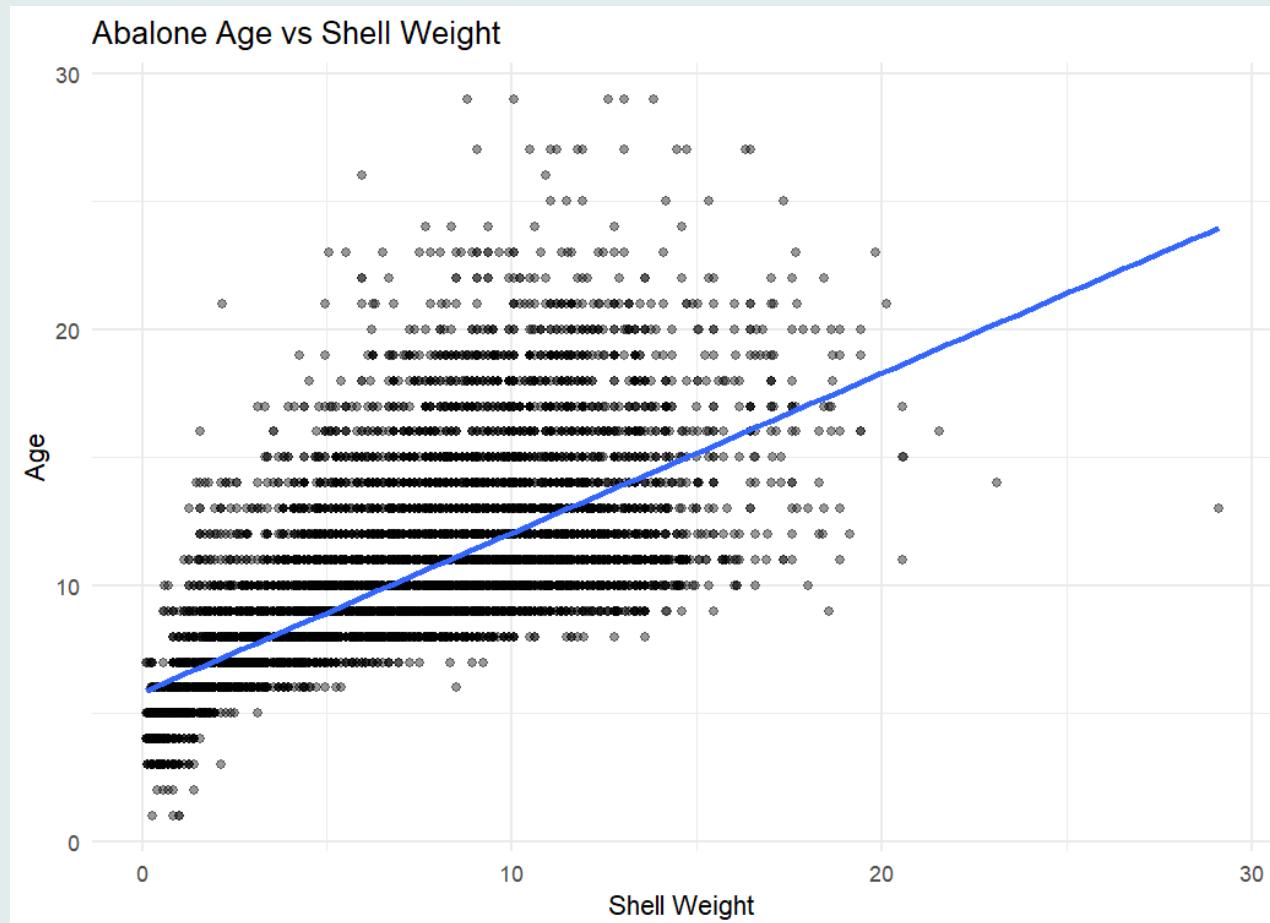
Combines statistical modeling with generative AI, transforming raw biological data into both accurate predictions and explainable, interactive insights



Dataset Overview

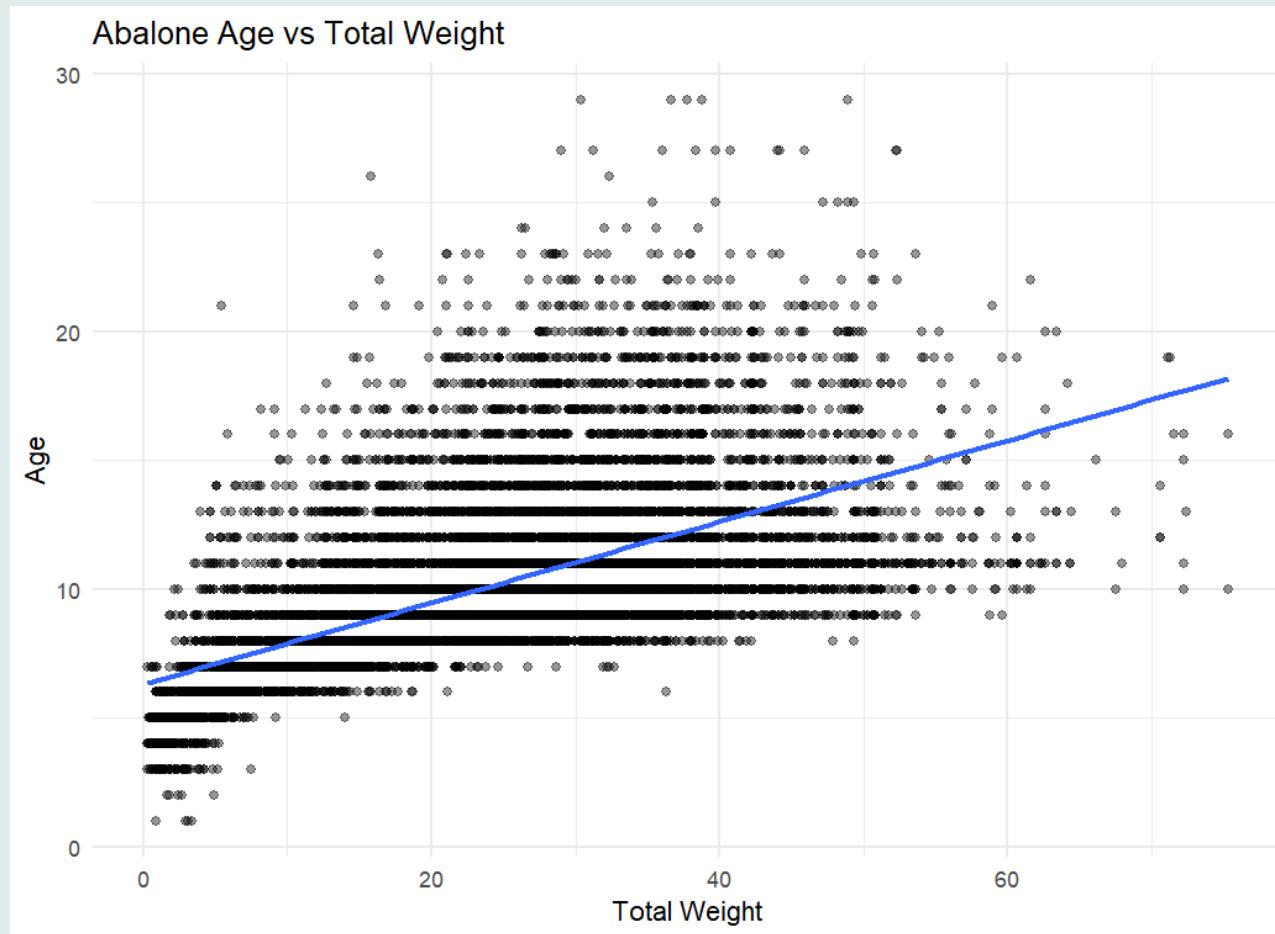
Feature	Description
sex	male, female, and indeterminant
length	longest shell measurement
diameter	perpendicular to length
height	with meat in shell
weight	whole abalone
shucked weight	weight of meat
visceral weight	gut weight (after bleeding)
shell weight	after being dried
age	number of rings on shell

Age vs Shell Weight



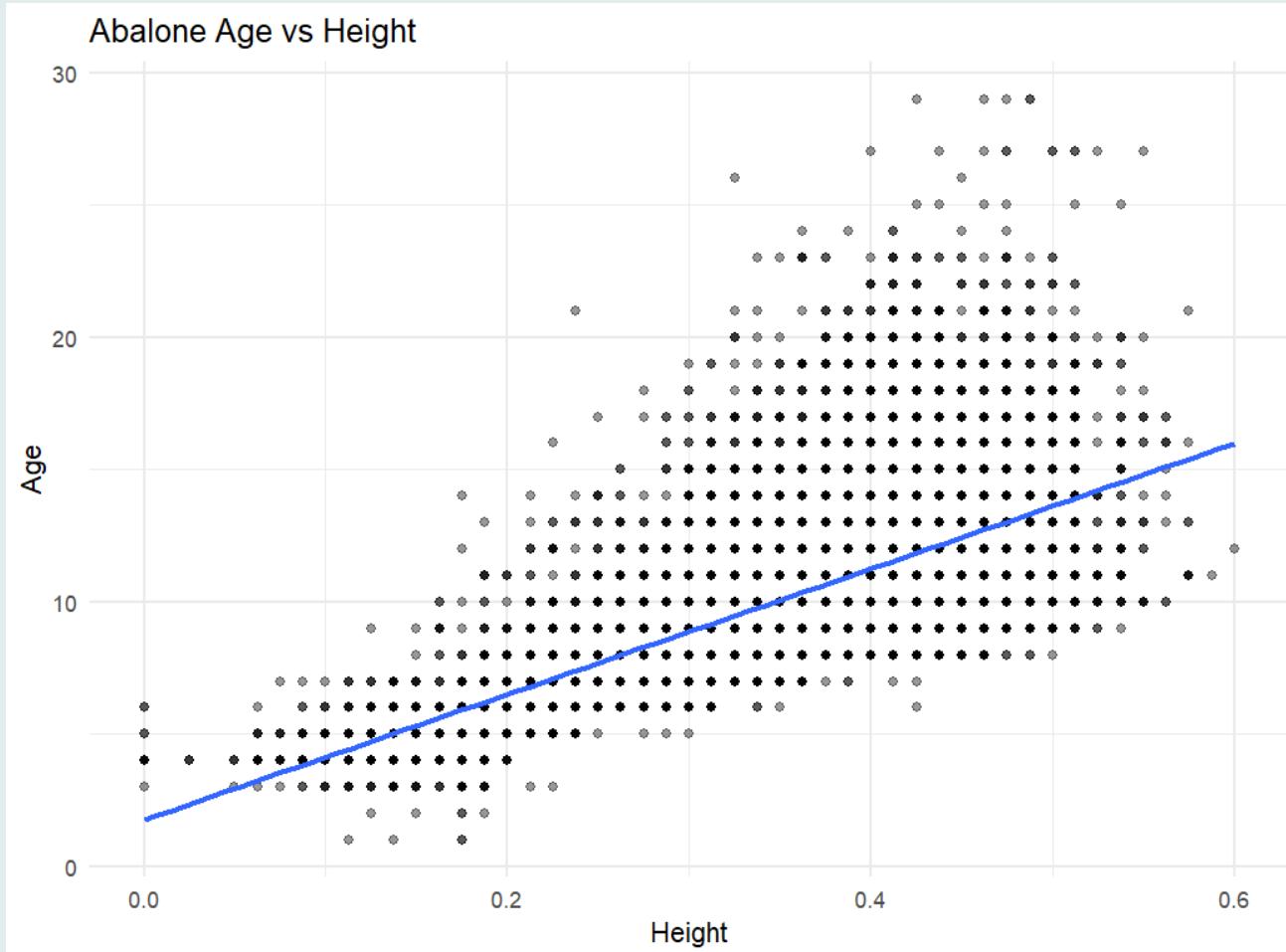
- Clear positive relationship between shell weight and age
- Older abalones tend to have heavier shells due to accumulated shell growth
- Shell weight shows a consistent upward trend across ages
- Variability suggests additional biological factors also influence age
- Supports shell weight as a strong predictor in the regression model

Age vs Weight



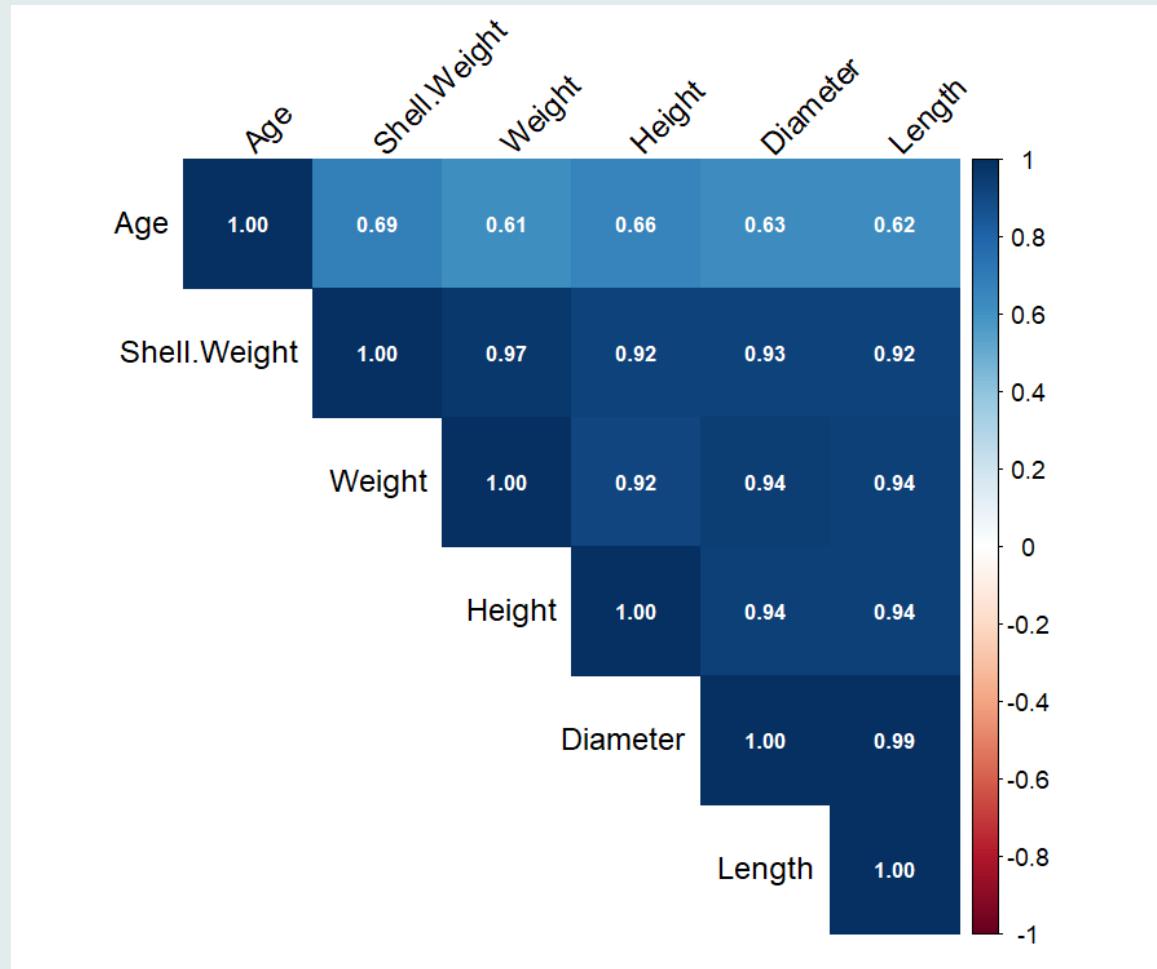
- Positive relationship between total weight and age
- Older abalones generally have higher total weight
- Greater spread in weight at higher ages indicates natural biological variability
- Total weight captures accumulated growth but is not a perfect indicator of age
- Supports using weight-based variables in the predictive model

Age vs Height



- Positive relationship between height and age
- Taller abalones tend to be older, reflecting vertical shell growth over time
- Height increases steadily with age but shows noticeable variability
- Height alone does not fully explain age differences across abalones
- Supports height as a contributing predictor but may be better paired with other features

Correlation Heat Map



- Age is strongly correlated with shell weight, total weight, and height
- Shell weight has the strongest relationship with age
- Length and diameter are highly correlated with each other.
- Strong correlations indicate overlap among physical measurements
- This overlap helps explain why some variables were not individually significant

Predictive Modeling Approach

Cleaned and prepared the abalone data by removing values that did not make biological sense



Started with a baseline linear regression model using the physical features of an abalone



Checked model assumptions and significance to understand which variables mattered



Chose the final model based on overall predictive accuracy



Added interaction terms to better capture the realistic growth relationships

Model Performance

Used a 70/30 train-test split

Evaluated performance using Mean Absolute Error (MAE)

The interaction model achieved a test MAE of 1.38, while the baseline achieved a test MAE of 1.40

The interaction model was selected for the competition set predictions



Key Insights

- Shell.Weight and Weight were the most important predictors of age
- Length and Diameter alone added little explanatory power
- Length × Diameter and Weight × Shell.Weight interactions improved the model
- Biological interactions led to better predictive performance

Abalones Q&A

Ask anything about Abalones!

Chatbot

What do young abalones feed on?

Young abalones, called larvae, feed on plankton.

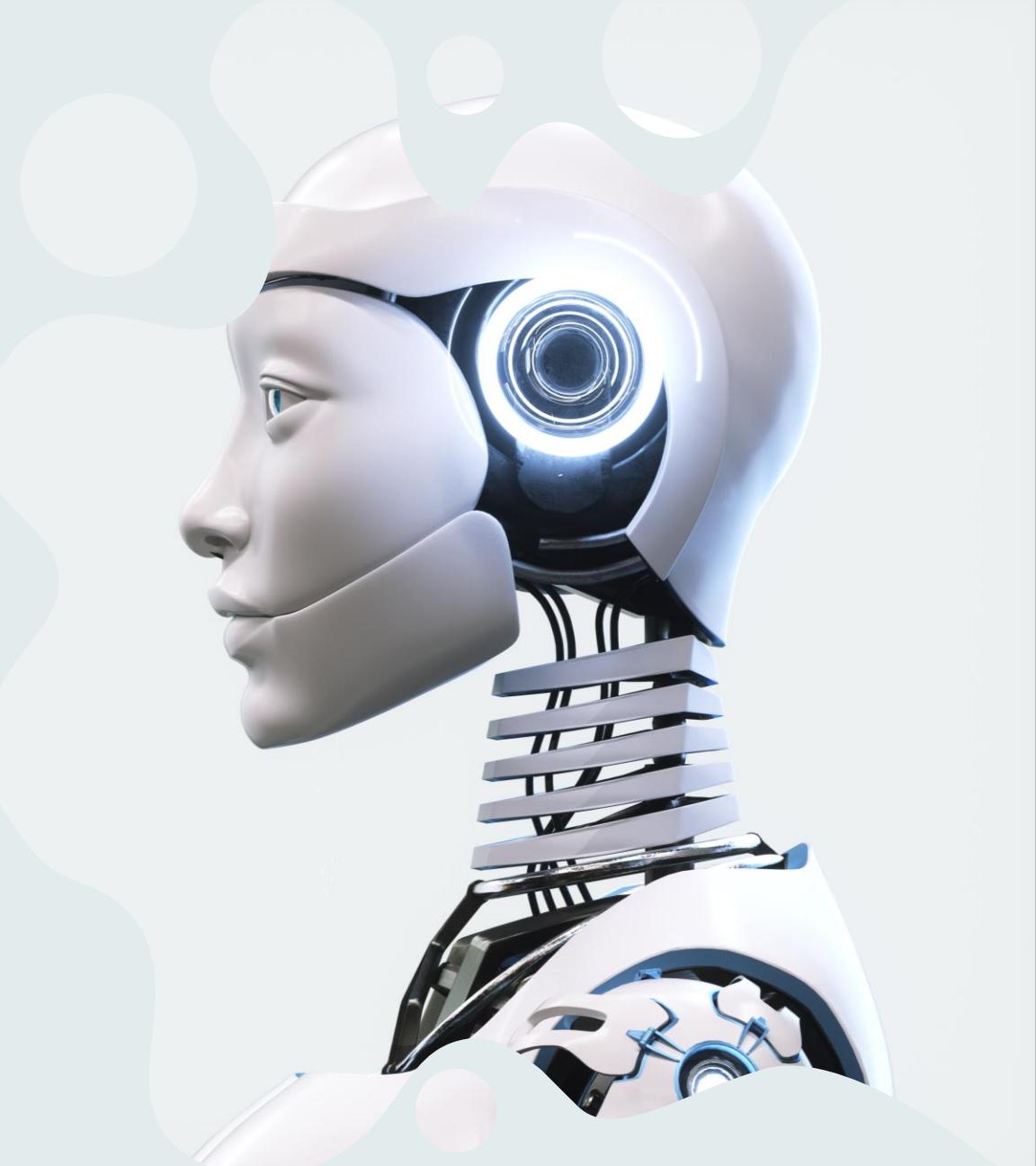
What about adult abalones, what do they like to eat?

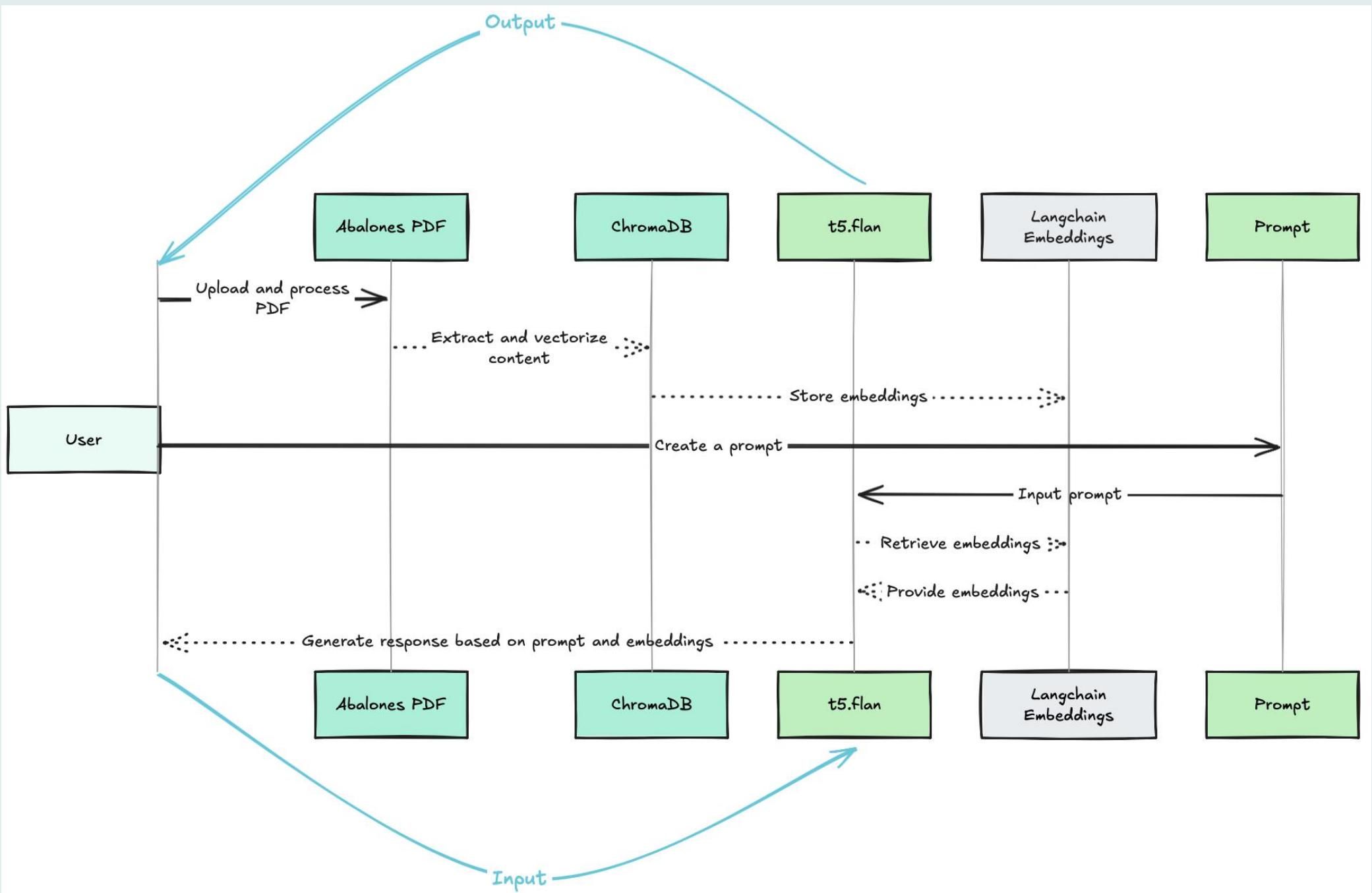
The adults eat plants and feed on algae. They like to eat red algae.

Ask a question...

Submit

Clear

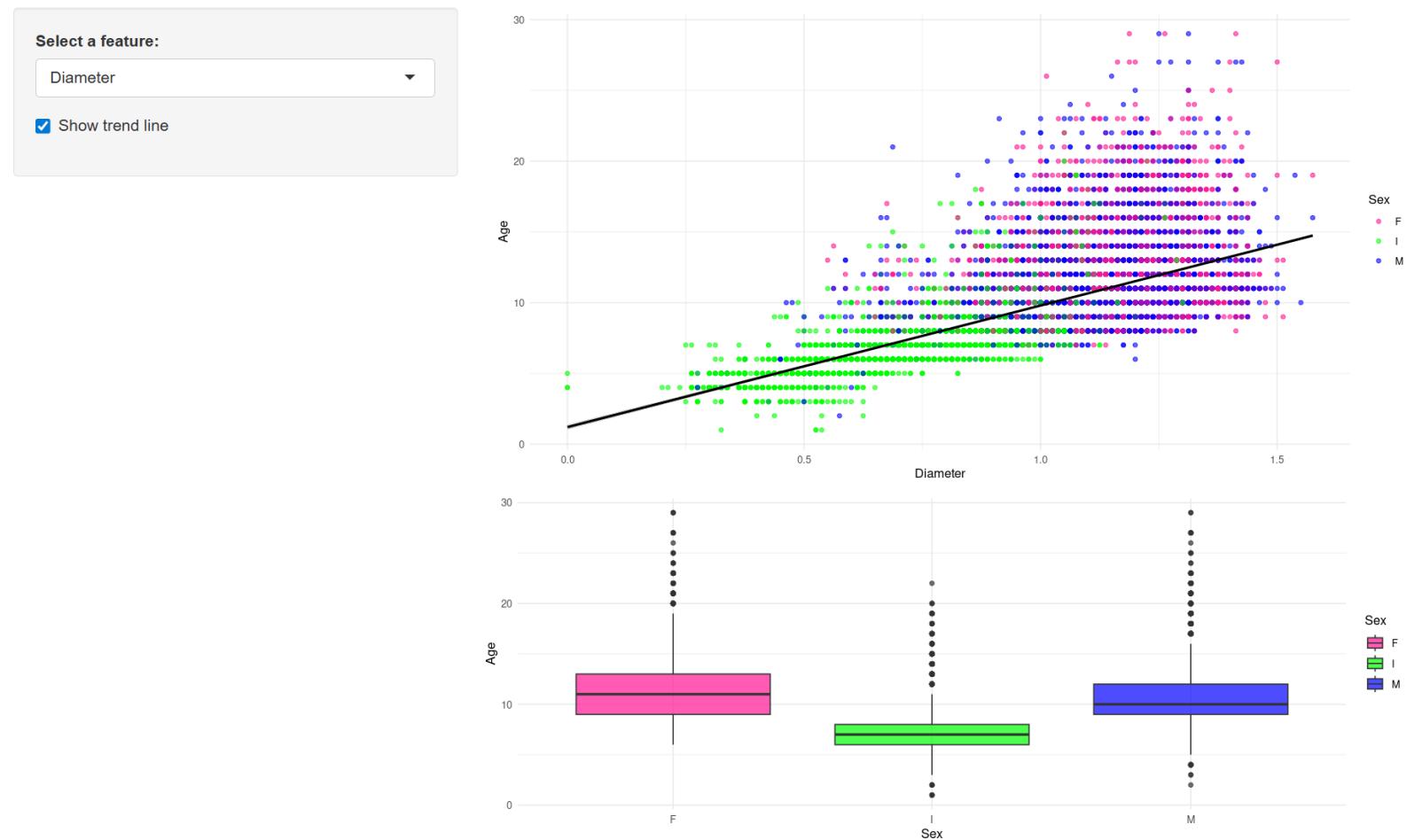




RShiny Dashboard

[Abalone: Exploring Relationships with Age](#)

Abalone: Exploring Relationships with Age



Key Takeaways

- Strongest predictors: shell weight and total weight
- Interaction terms > a simple linear approach
- Low error ($MAE \approx 1.38$), strong predictive performance
- Predictive modeling with RAG = accurate predictions and interactive data exploration





Thank You!

Emails:

nkhandaker@smu.edu

cprowse@smu.edu

Github:

[Github Repository](#)

Q&A Application:

[Q&A Application](#)

RShiny Dashboard:

[Abalone: Exploring Relationships with Age](#)