# Fish Market Mini Project

By Ethan, Ishani, Nicholas, Renee, and Sameer

# Model and Loss Function

- We first tested both the Ridge model and the Lasso model and determined that the ridge model consistently gave a lower MAE

- The sklearn Ridge function was used to create the multilinear regression model
  - It was fitted using the training x and y data

- Model uses the least-squares loss function to minimize error

- Loss Function: Mean Absolute Error (MAE) was used to determine the accuracy of the model
  - It better accounts for outliers in the dataset

$$\text{MAE} = \frac{1}{n} \sum_{j=1}^{n} |y_j - \hat{y}_j|$$

# Feature Transformation and Regularization

Feature Transformation: The feature data was normalized using preprocessing.scale.
Regularizer: Ridge Regression, fit to both xtrain and ytrain. This was a method of regularization to mitigate the effects of overfitting.

Hyper Parameters: The best value for alpha was 0.04. This value was found by applying ridge regression via a for loop through alpha values starting at 0 and incrementing by .01 to reach a maximum of 1. These results were then used to determine the alpha value that produced the lowest MAE.

```python
min_train_mae = np.amin(maetrain_values)
min_test_mae = np.amin(maetest_values)

best_train_alpha = alpha_values[np.where(maetrain_values == min_train_mae)]
best_test_alpha = alpha_values[np.where(maetest_values == min_test_mae)]
```

```python
for i in range(0,len(alpha_values)):
    yhat_train, yhat_test = RidgeRegression(alpha_values[i], x_train, x_test, y_train, y_test)
```
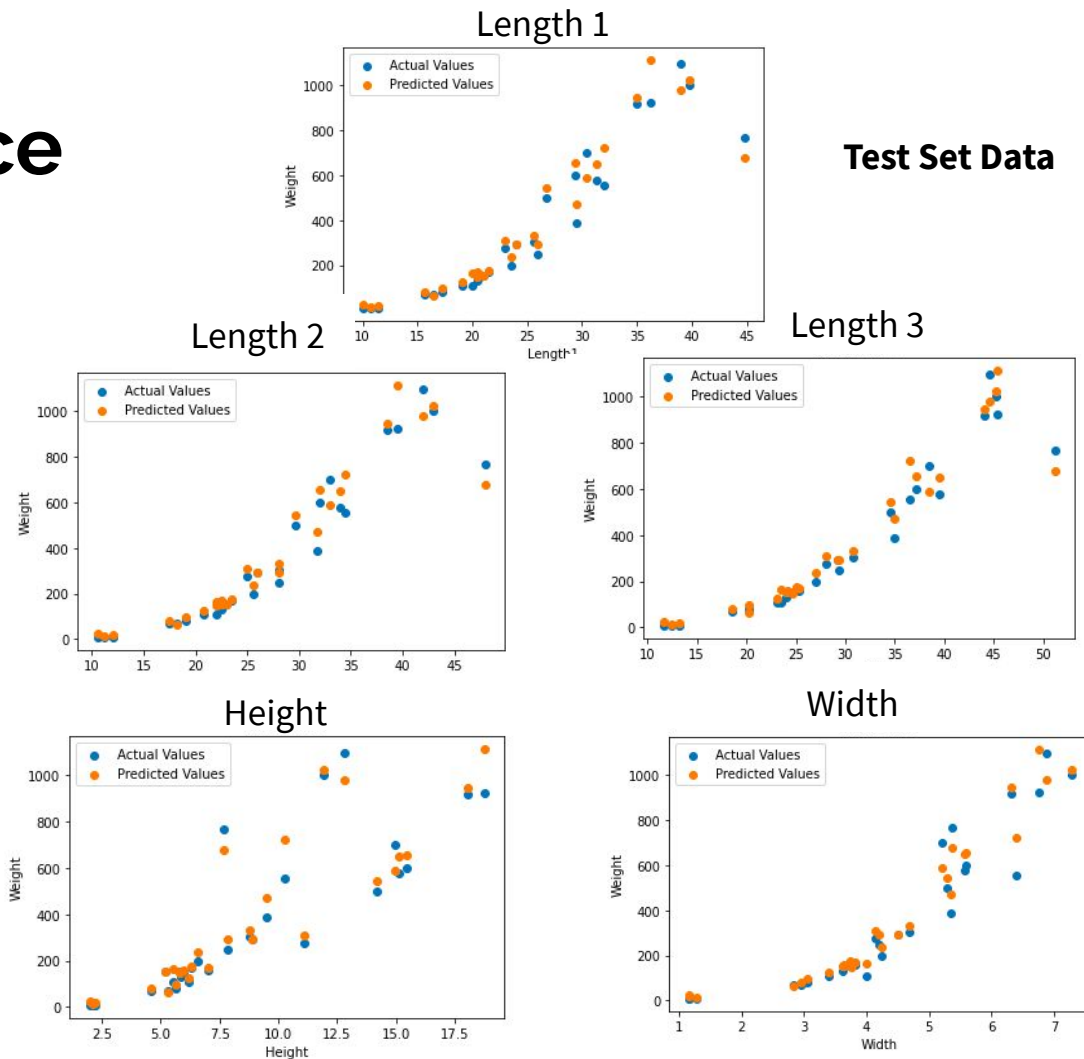
# Model Performance

**Test Set Data**

**Training Set:**

- MAE: 14.3
- Best Alpha: 0.04

**Test Set:**

- MAE: 44.3
- Best Alpha:



Length 1

Length 2

Length 3

Height

Width

# Challenges

- Throughout our project, we encountered some difficulties regarding our MSE and MAE values.
- Though we went through our code multiple times, our MSE and MAE values were abnormally high. We tried adjusting the test size, adding a train size, reordering our data while splitting, but the differences were minimal.
- We reached out for help from others, but we soon realized this was perhaps the intended range for the MSE and MAE was larger.
- We also initially found the best alpha was 0.0, which was later fixed.
- We also had several error messages and took a while to edit the code to make sure we defined all variables and classes, but we worked together to find any mistakes.

# Conclusion

Overall, our model performed well on the test set. It accurately predicted fish weight based on length, width, and height, even when there was not a strong correlation between the features and label.

Even though we faced some challenges in terms of virtual collaboration and high MSE and MAE values, in the end, our model performed well on the data.