Fish Market EDA and Machine Learning Models

Hilotin, Llamis, Candel, & Javier

2022

DS3A

Introduction

Fish has always been a valuable resource for humanity, primarily as a food source.
Approximately **35,000 species** worldwide, the most species diversity of any vertebrate group of various sizes and shapes. The price of a fish in a fish market is usually determined by the type of fish (or species), its weight, height, and length.

Fish markets range in size from small fish stalls, medium and large depending on the demand. In the fish market, it is important to determine the price of the fish correctly, and 2 important factors that impact on the price are fish species and also the weight

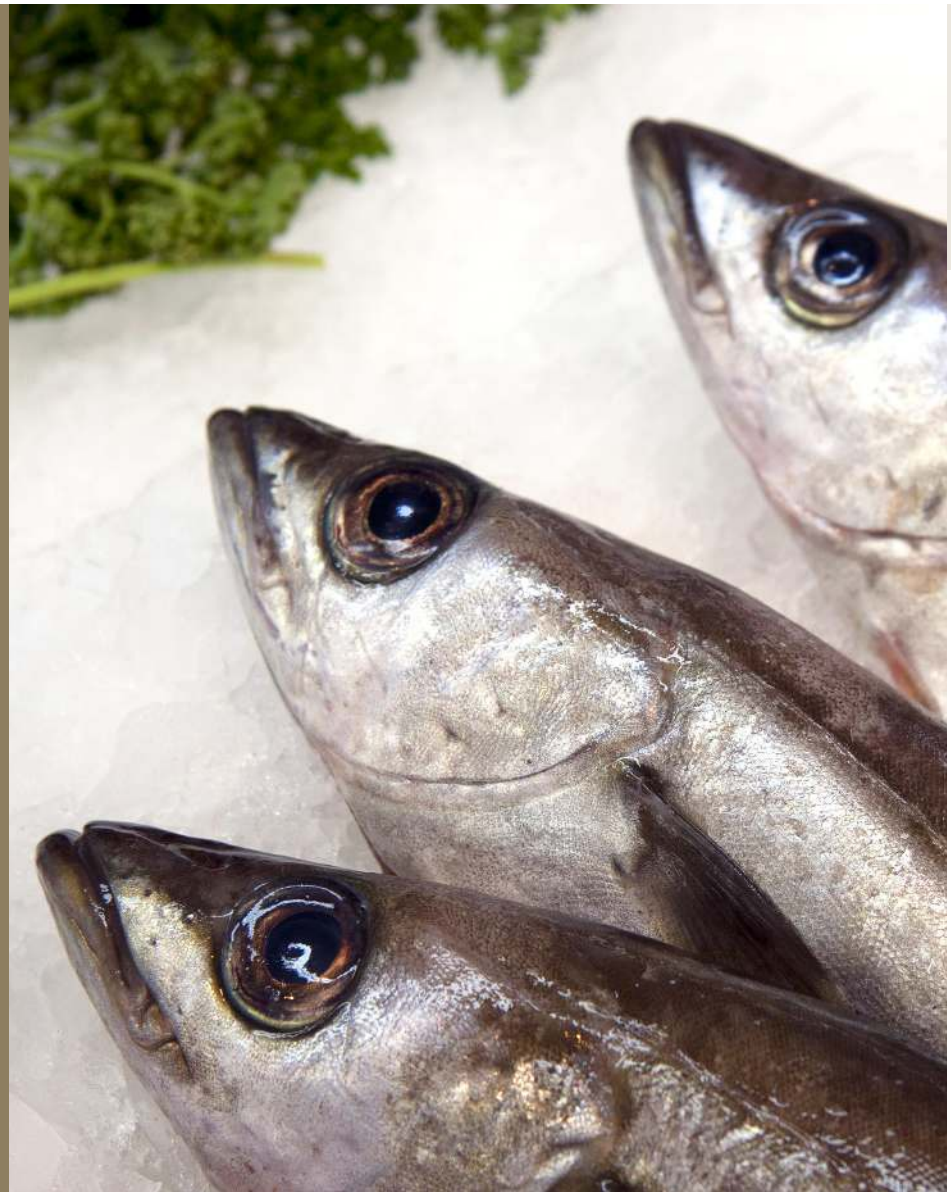
The Fish Market dataset was from Kaggle

This dataset is a record of 7 common different fish species in fish market sales. Mainly the:

- Bream
- Roach
- Whitefish
- Perch
- Pike
- Smelt
- Parkki

The dataset consist of 7 features.

- Species
- Weight (g)
- Length1 (cm)
- Length2 (cm)
- Length3 (cm)
- Height (cm)
- Width (cm)



Problem Statement & Objective

- Why the consumer prefers to buy this type of fish?
- What is the species of the fish?
- Without using weighing scale, what would be the weight of the fish?
- That is what we are going to find out using Classification and Regression.

The main objective in this project is to determine fish price according to its:

1. Fish Species/Type and its Weight.
2. ML model of the fish's weight based on the length, height and width of the fish and also to classify fish's species based on the weight, length, width and height.

Methodology

Data Collection

Data Exploration

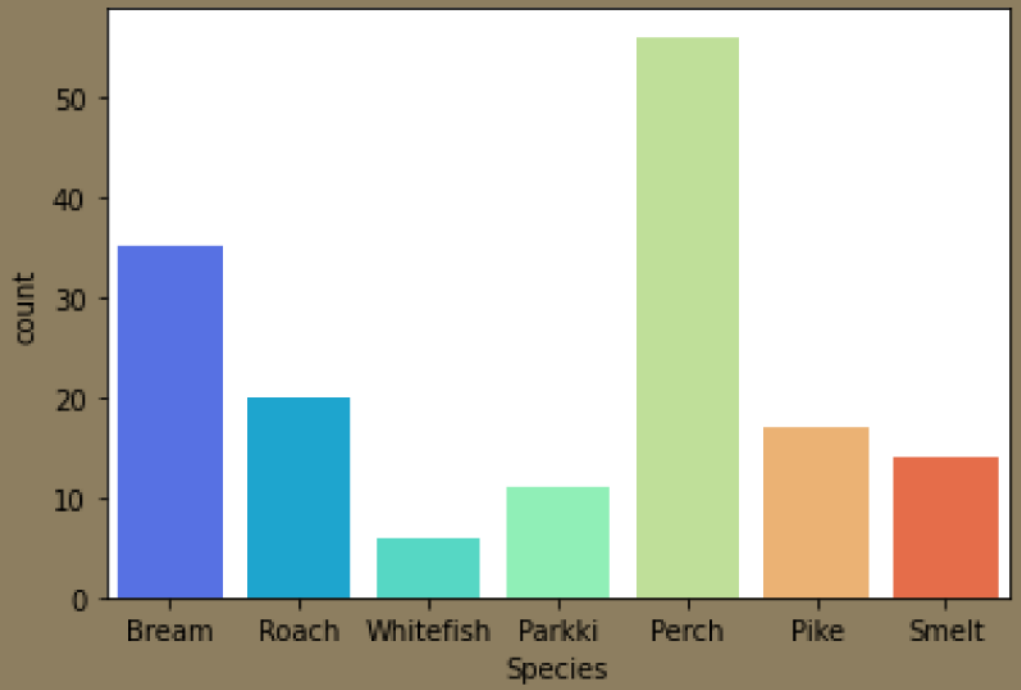
Exploratory Data Analysis

Data Modeling

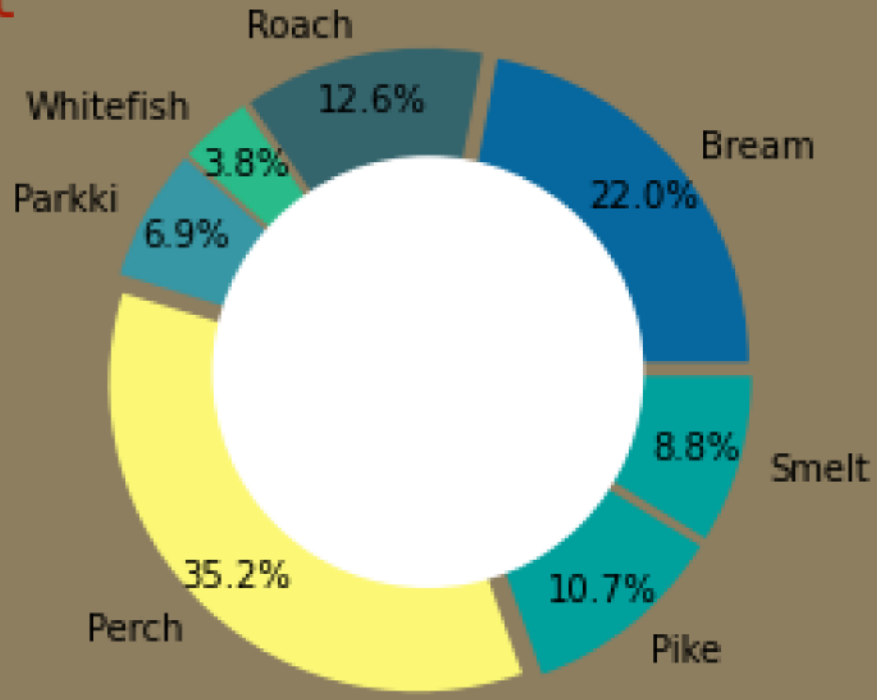
```
data.iloc[[40]] = data.iloc[[40]].replace({0.0 : 120})
b=data.iloc[[40]]
b.style.background_gradient(cmap='Pastel2')
```

	Species	Weight	Vertical_Length	Diagonal_Length	Cross_Length	Height	Width
40	Roach	120.000000	19.000000	20.500000	22.800000	6.475200	3.351600

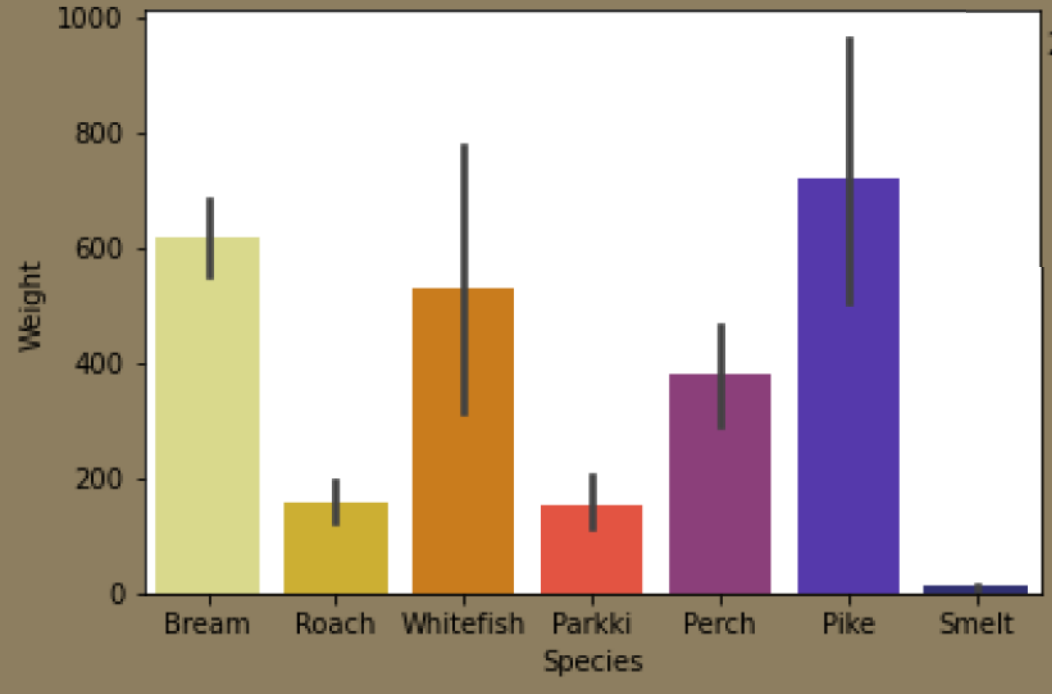
Popular Choice of Fish in the Market

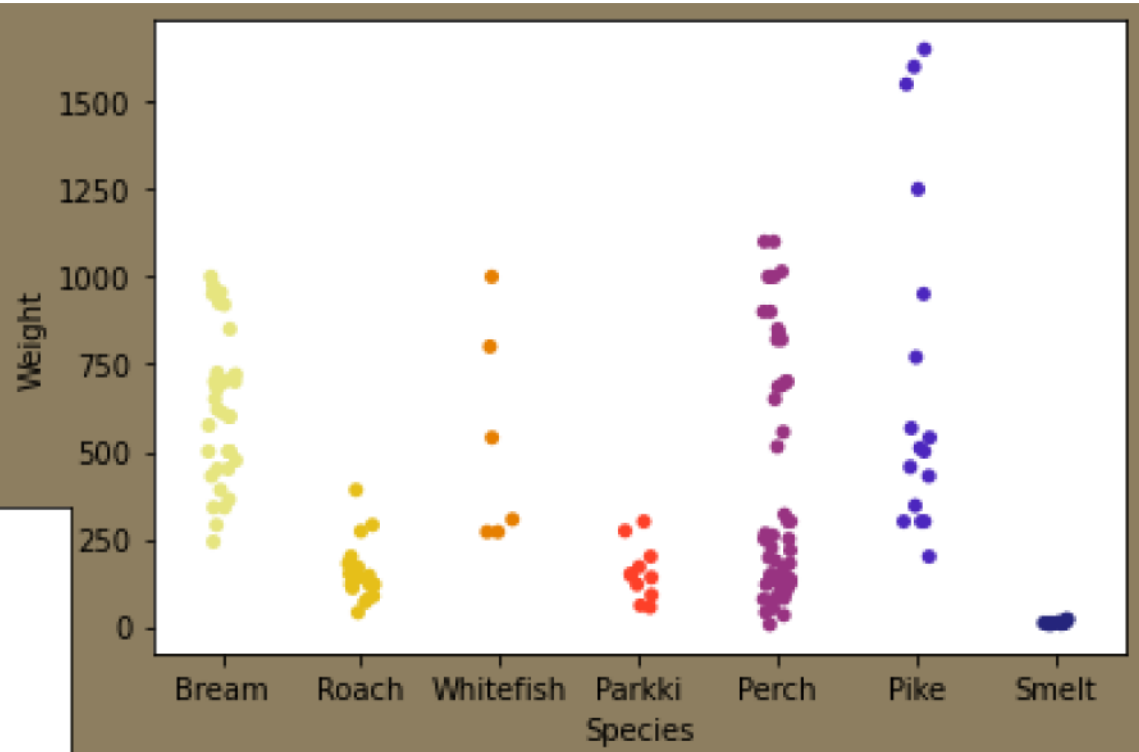


Species Proportion

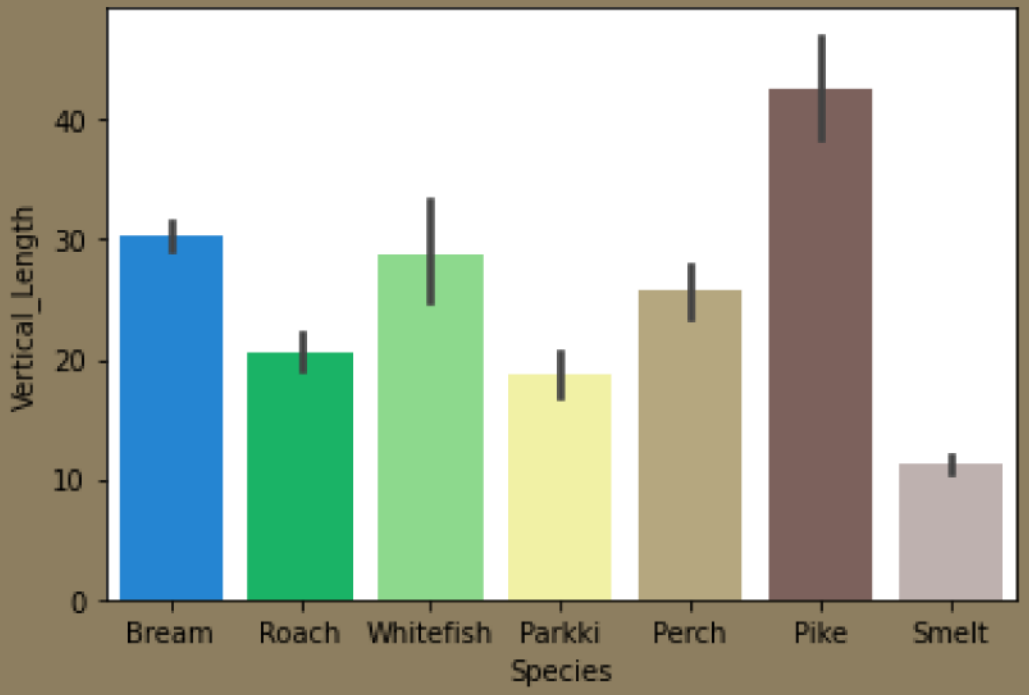


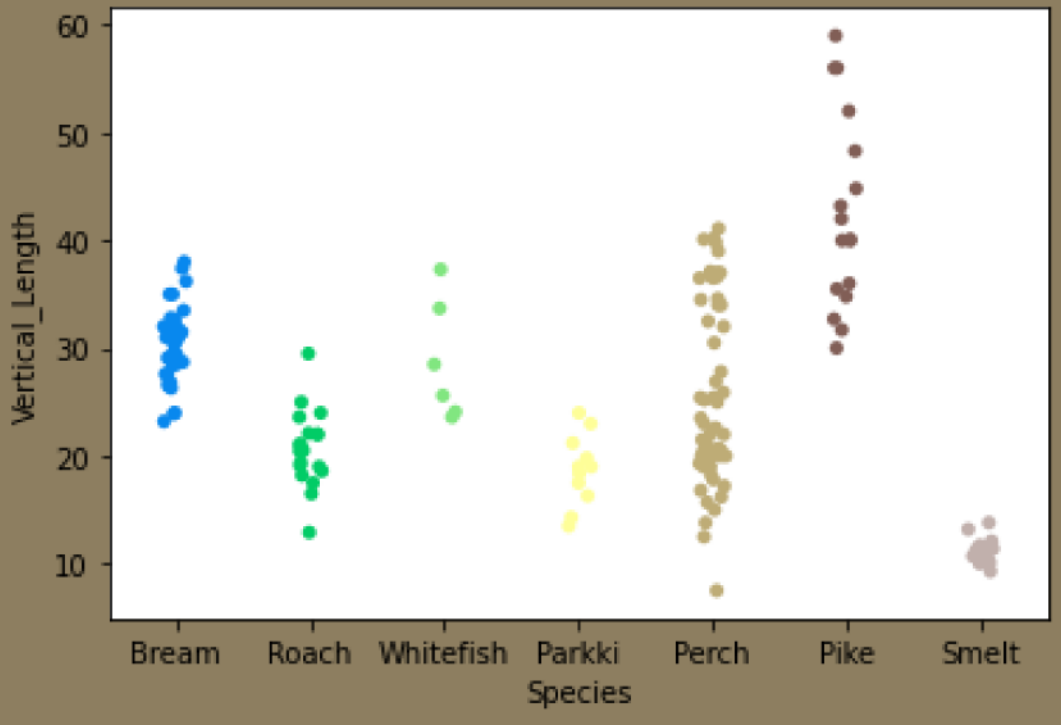
Weight Analysis



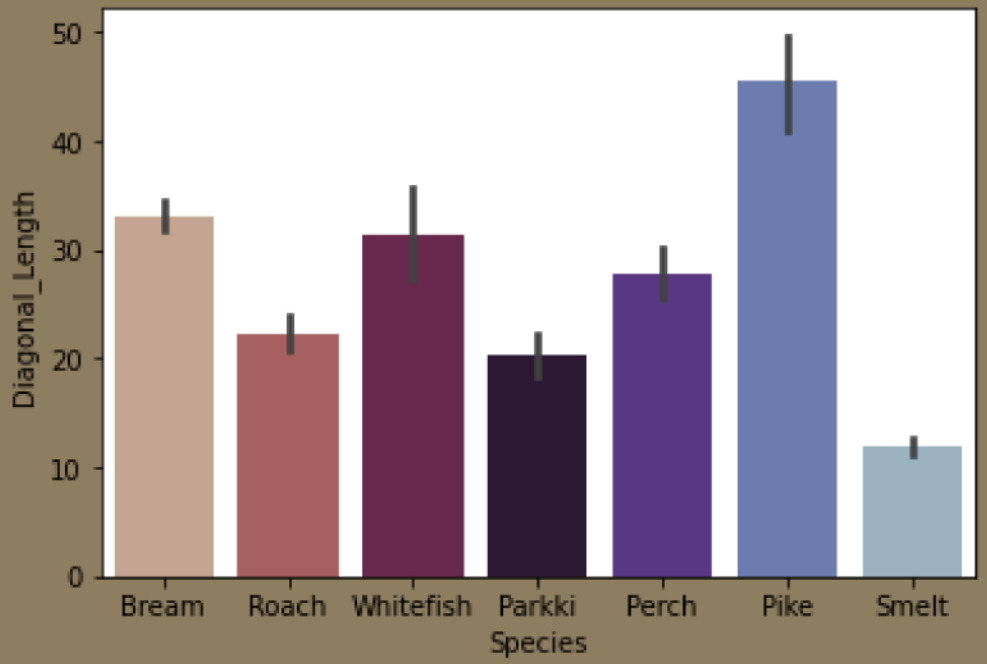


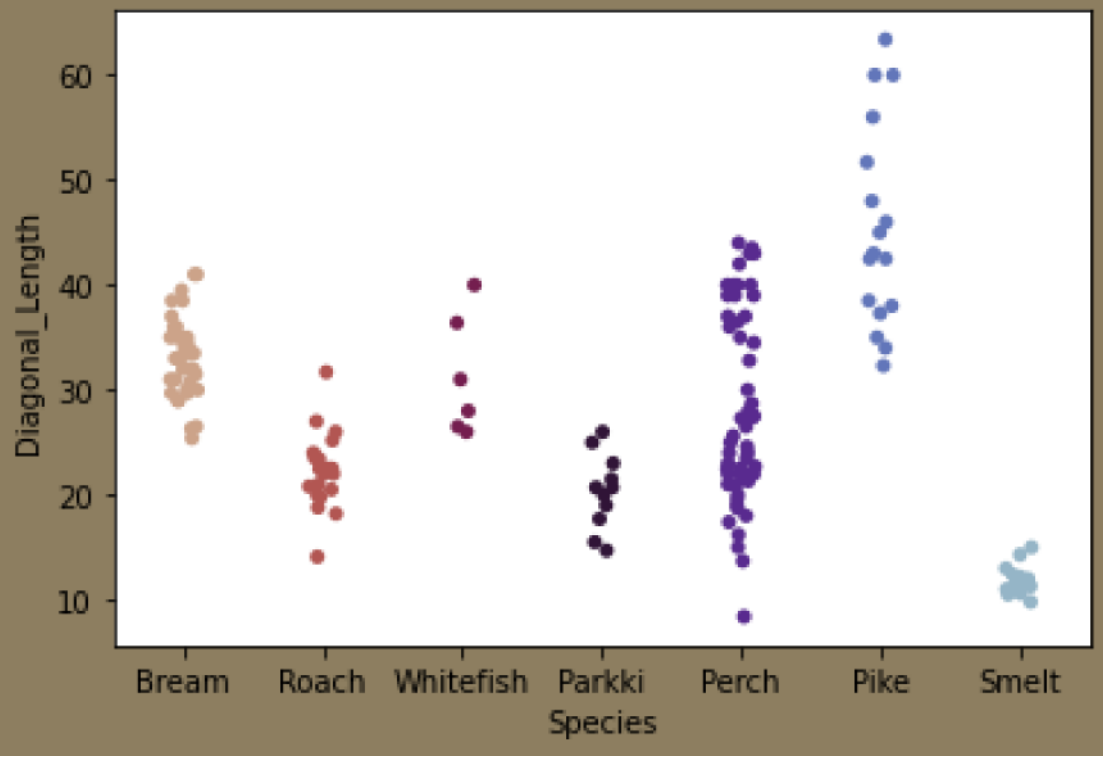
Vertical Length Analysis



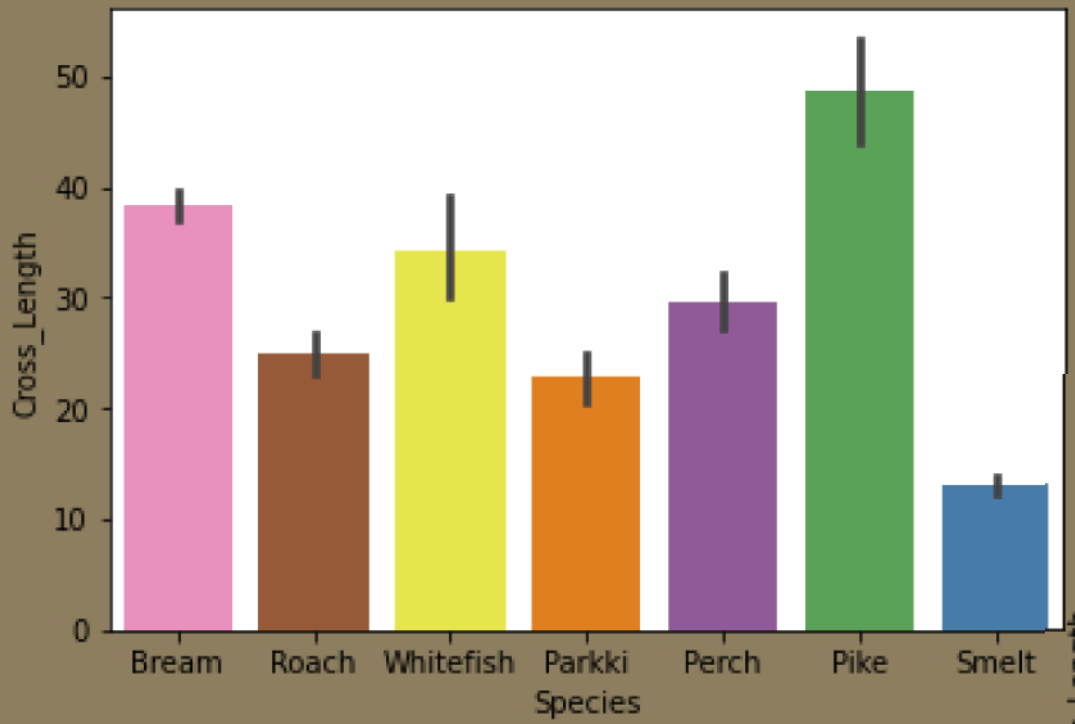


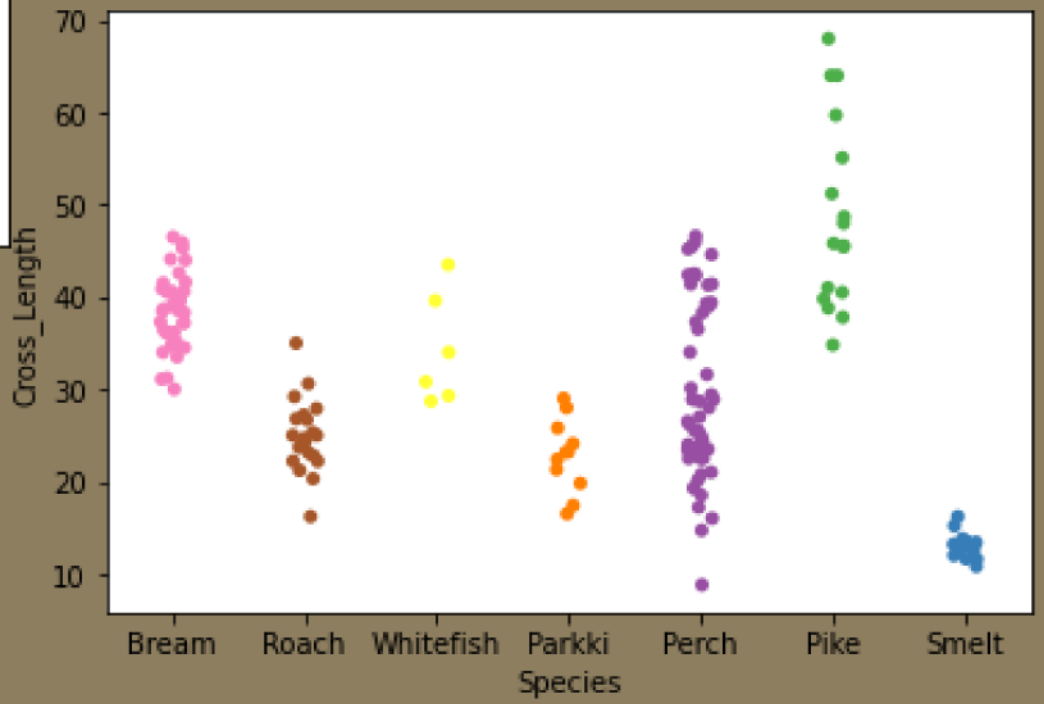
Diagonal Length Analysis



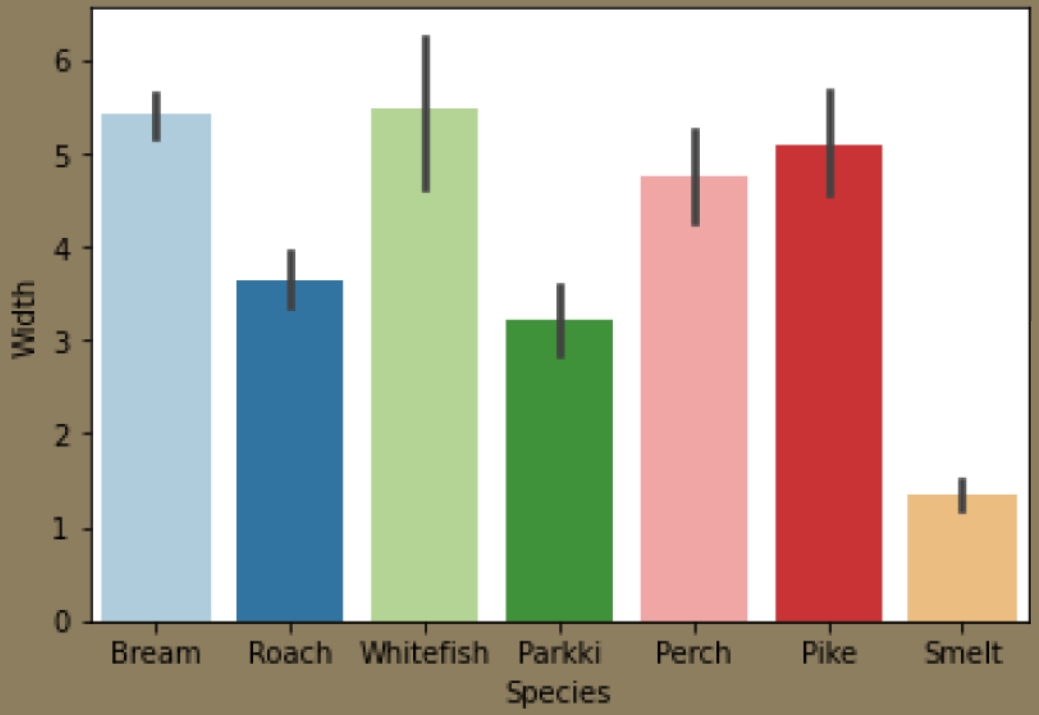


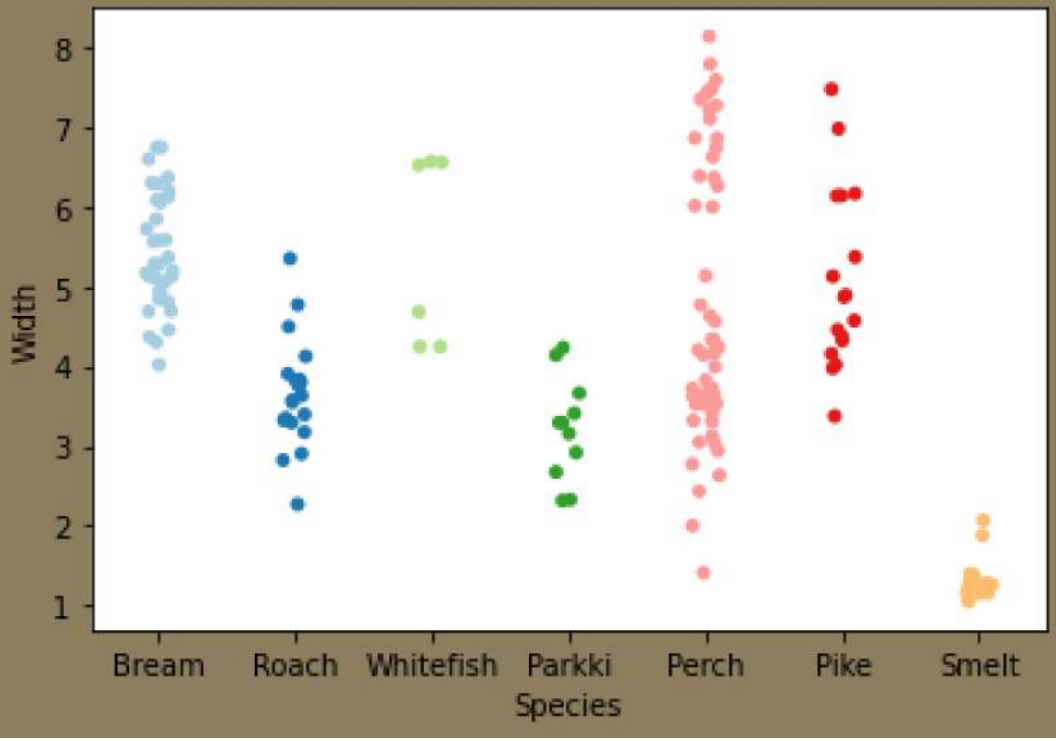
Cross Length Analysis



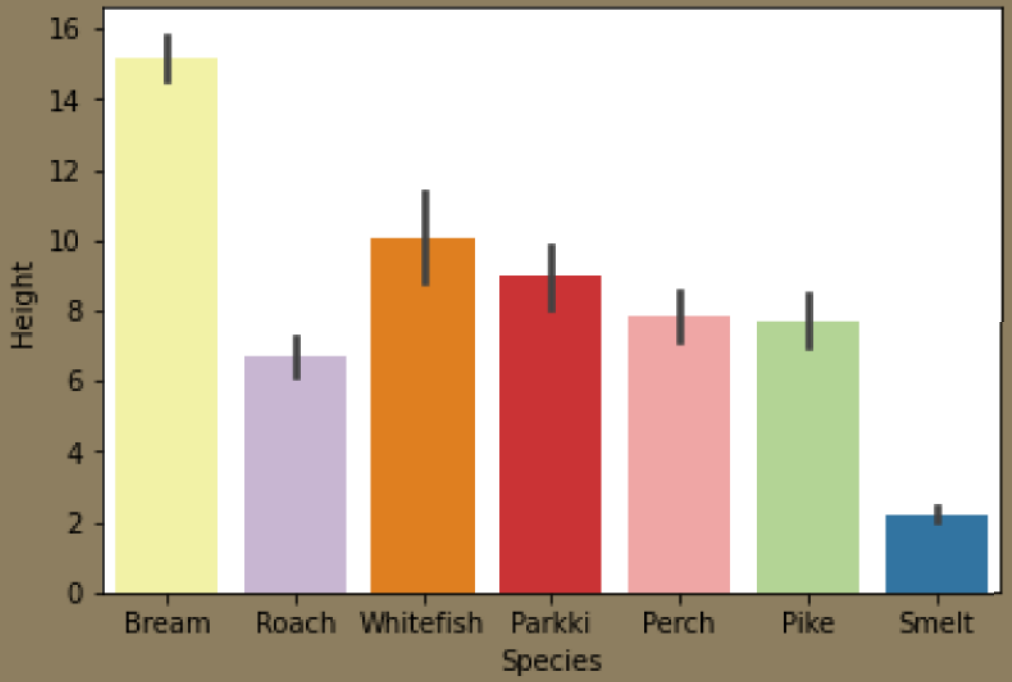


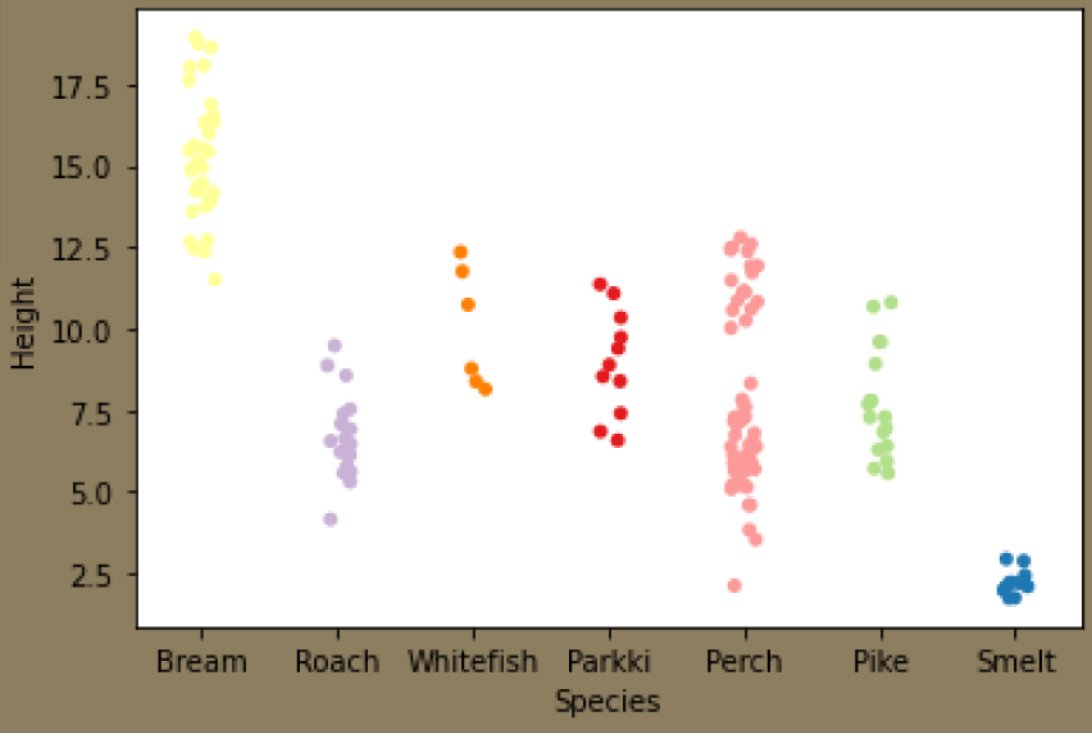
Width Analysis





Height Analysis





Machine Learning Method	Test Accuracy	Feautures
L2-Ridge Regression	82.64%	1 Features
L2-Ridge Regression	80.35%	42 features
L1-Lasso Regression	82.41%	1 Features
L1-Lasso Regression	97.50%	41 features

Machine Learning Method	Estimator	Test Accuracy	Best Parameter
Linear Regression	None	82.64%	None
Decision Tree Regressor	None	94.46%	max_depth=10, random_state=0
Decision Tree Regressor	GridSearchCV	96.25%	max_depth= 6, max_features= 5, min_samples_leaf= 1
K-Nearest Neighbor Regression	GridSearchCV	96.46%	'n_neighbors': 1
Gradient Boosting Regressor	None	78.07%	learning_rate= 0.01, max_depth= 2
Gradient Boosting Regressor	GridSearchCV	97.18%	'learning_rate': 0.04, 'max_depth': 4, 'n_estimators': 500, 'subsample': 0.5
Random Forest Regressor	None	94.92%	n_estimators=4, random_state=42
Random Forest Regressor	GridSearchCV	97.19%	'bootstrap': True, 'max_depth': 10, 'n_estimators': 1000