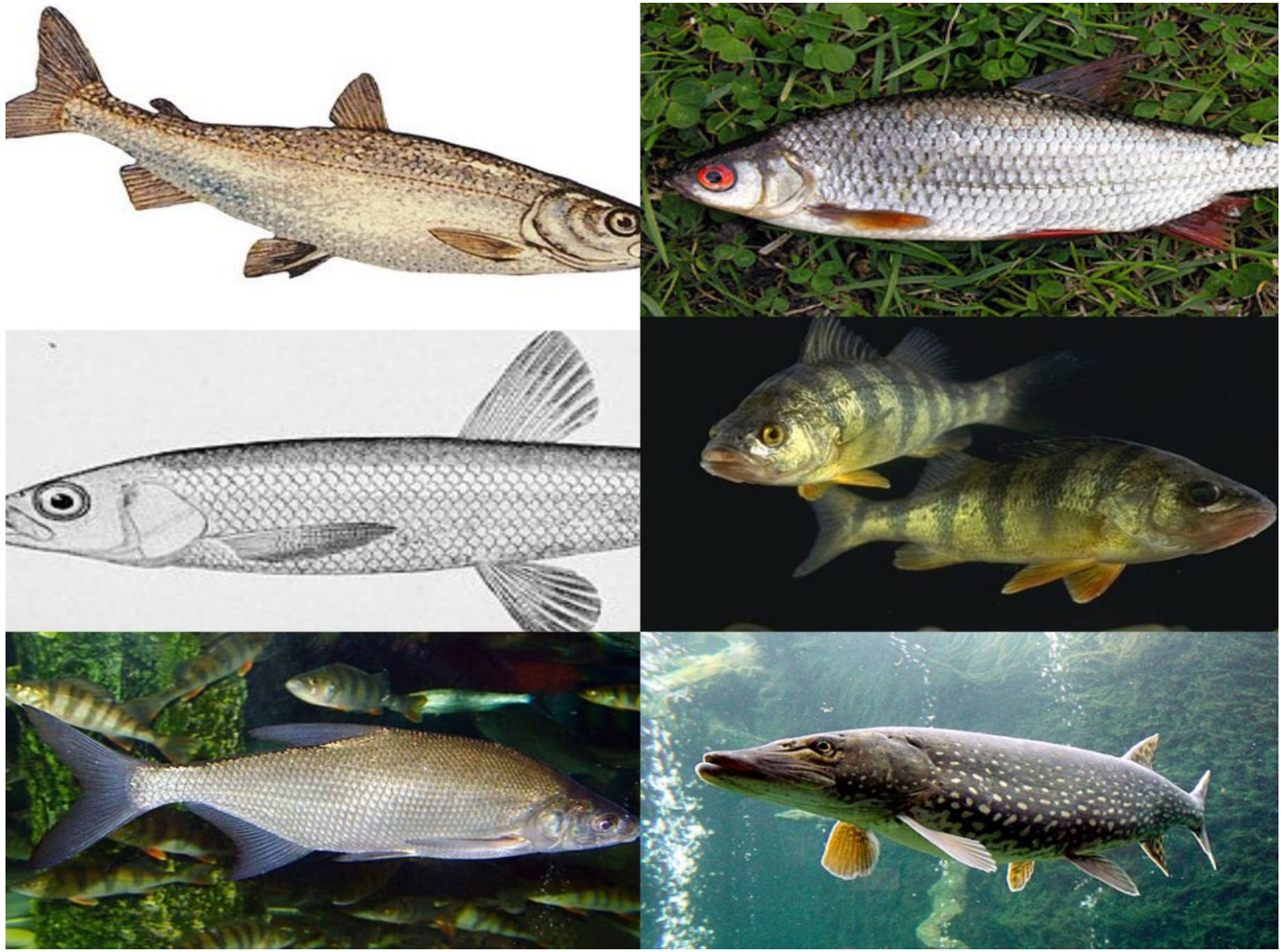
Analysis Of Fish Species



By Nigel Karikari

Goal Of The Project

This project will analyse the fish market data set on kaggle. The main focus will be on the correlation between weight and height of the fish species.

Description Of The Data

The data consists of a table with seven columns which are variables of the fish species namely: Species, Weight, Length1, Length2, Length3, Height, and Width.

Introduction

Using R, let us attach the file and take a look at some summary statistics.

	Species	Weight	Length1	Length2	Length3	Height	Width
1	Bream	242	23.2	25.4	30.0	11.5200	4.0200
2	Bream	290	24.0	26.3	31.2	12.4800	4.3056
3	Bream	340	23.9	26.5	31.1	12.3778	4.6961
4	Bream	363	26.3	29.0	33.5	12.7300	4.4555
5	Bream	430	26.5	29.0	34.0	12.4440	5.1340
6	Bream	450	26.8	29.7	34.7	13.6024	4.9274

	Species	Weight	Length1	Length2	Length3	Height	Width
154	Smelt	9.8	11.4	12.0	13.2	2.2044	1.1484
155	Smelt	12.2	11.5	12.2	13.4	2.0904	1.3936
156	Smelt	13.4	11.7	12.4	13.5	2.4300	1.2690
157	Smelt	12.2	12.1	13.0	13.8	2.2770	1.2558
158	Smelt	19.7	13.2	14.3	15.2	2.8728	2.0672
159	Smelt	19.9	13.8	15.0	16.2	2.9322	1.8792

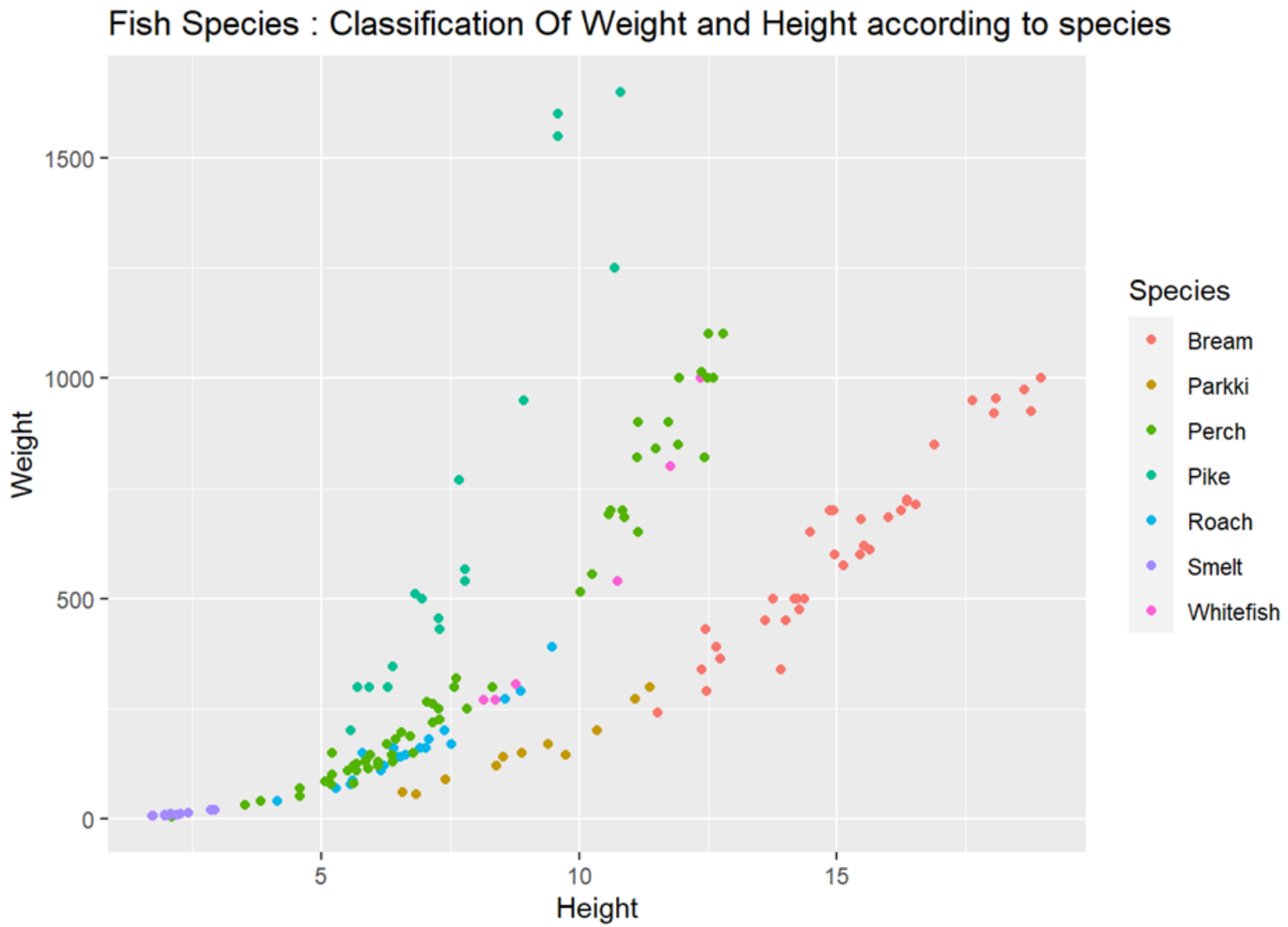
Descriptive Analysis

Here, we perform descriptive statistics such as the mean, for all the variables in the data set.

```
  Species            Weight        Length1        Length2
 Length:159        Min.   :   5.9   Min.   : 7.50   Min.   : 8.40
 Class :character   1st Qu.: 122.5   1st Qu.:19.05   1st Qu.:21.00
 Mode  :character   Median : 273.0   Median :25.20   Median :27.30
                    Mean   : 399.2   Mean   :26.25   Mean   :28.42
                    3rd Qu.: 650.0   3rd Qu.:32.70   3rd Qu.:35.50
                    Max.   :1650.0   Max.   :59.00   Max.   :63.40
    Length3         Height          Width
 Min.   : 8.80   Min.   : 1.728   Min.   :1.048
 1st Qu.:23.15   1st Qu.: 5.945   1st Qu.:3.386
 Median :29.40   Median : 7.786   Median :4.248
 Mean   :31.23   Mean   : 8.971   Mean   :4.417
 3rd Qu.:39.65   3rd Qu.:12.366   3rd Qu.:5.585
 Max.   :68.00   Max.   :18.957   Max.   :8.142
```

Plot of the data

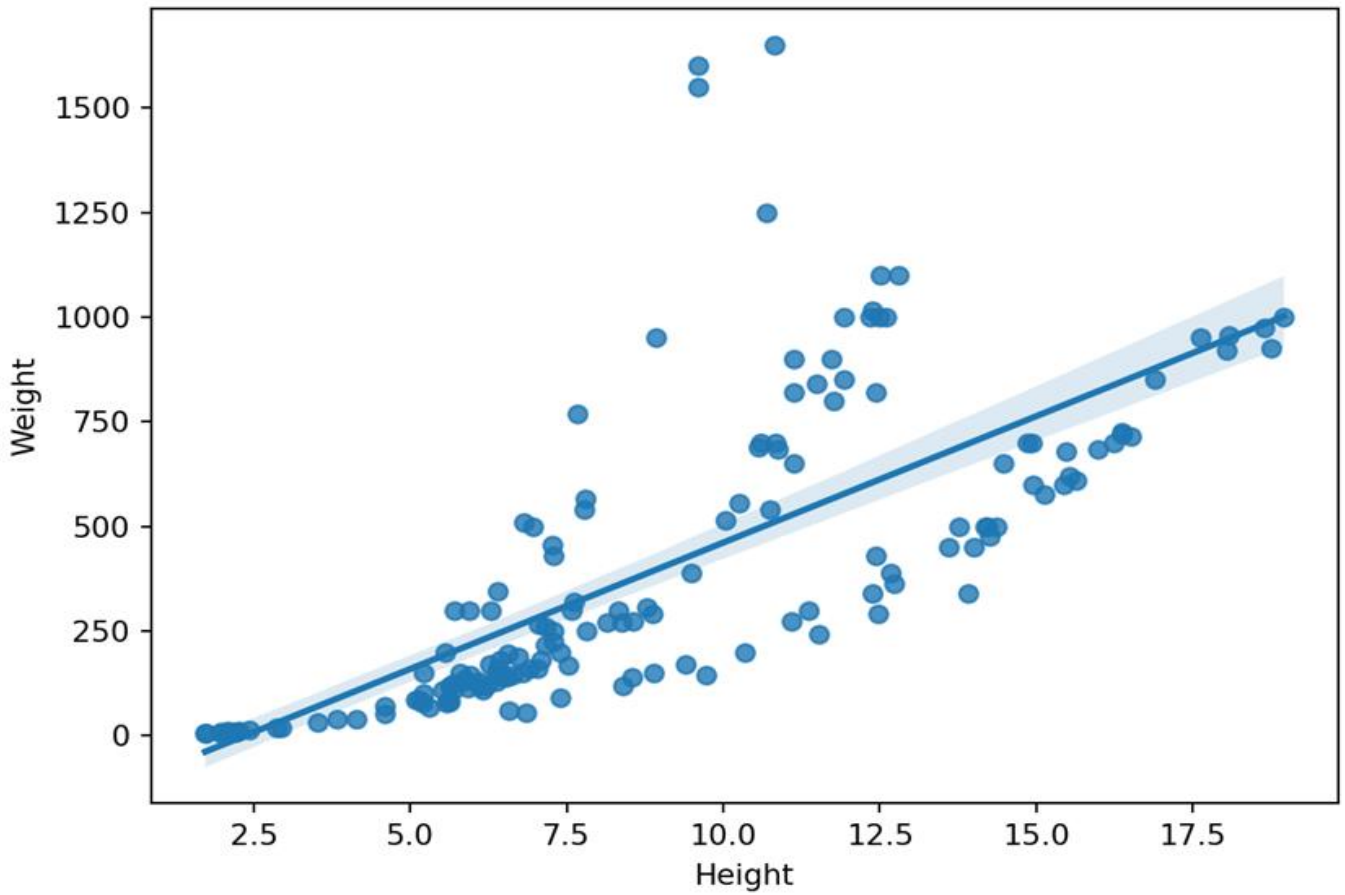
In this plot, the relationship between the variables, weight and height was studied for each specie.



The plot shows a positive correlation between the weight and height of the fish species.
The following analysis was done using python
A linear model was tested on the data set.
The intercept is -142.42559228455798 whereas the coefficient is [60.37597706].
The coefficient of determination is 0.5249770555319087.
The resulting equation would be : Weight = -142.43 + 60.38 * Height. This means per unit height, weight increases by 60.38.
The coefficient of determination is close to 1, which means the model is appropriate for the data.

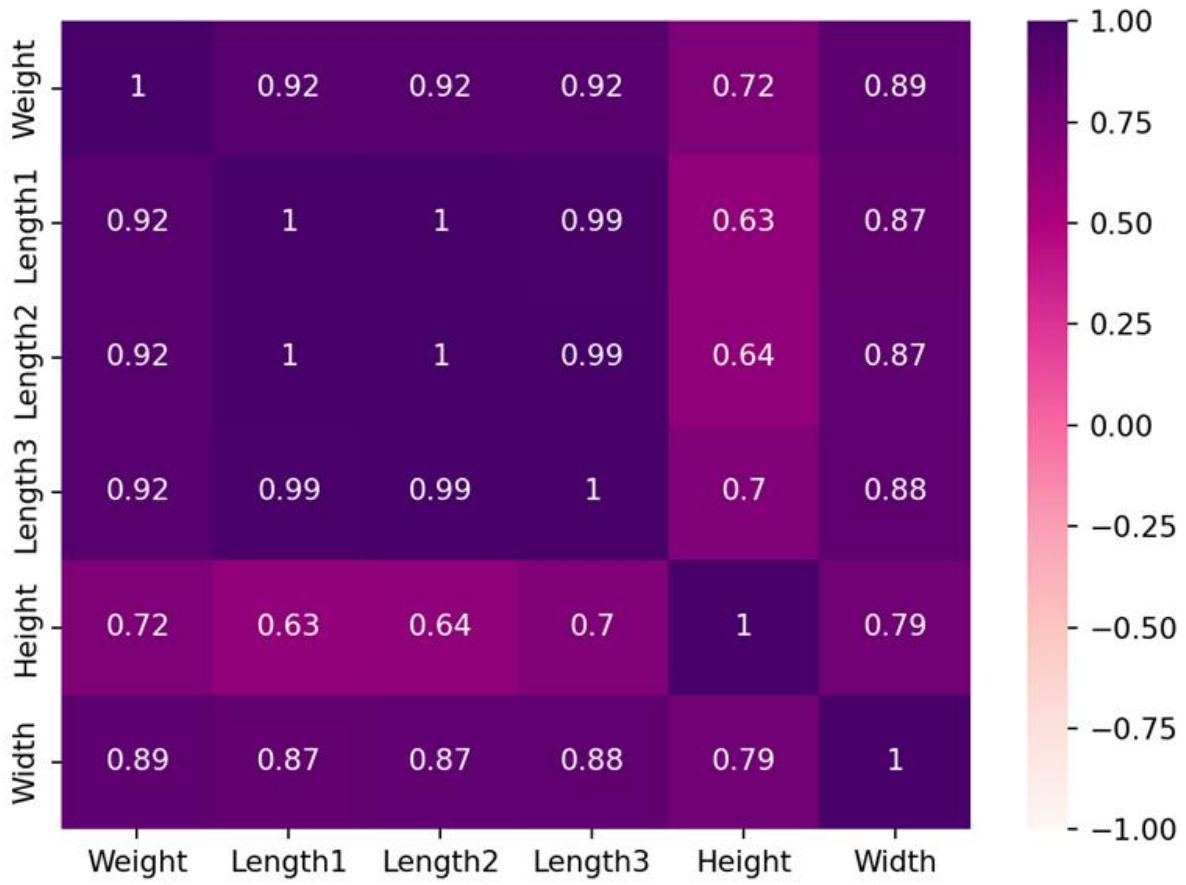
Evaluating the model

This could be done using a regression plot:



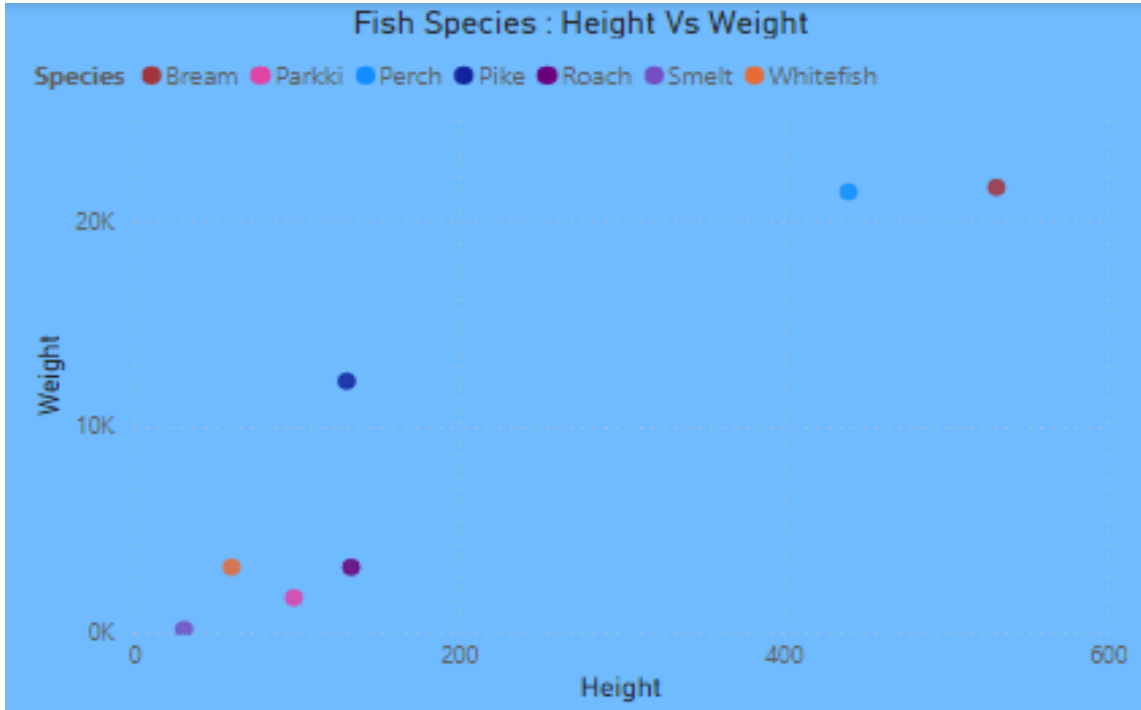
Judging by the graph, the linear model seems appropriate for the data. This was initially confirmed by the coefficient of determination on the previous page.

Lets check correlation, with our main focus on height and weight :



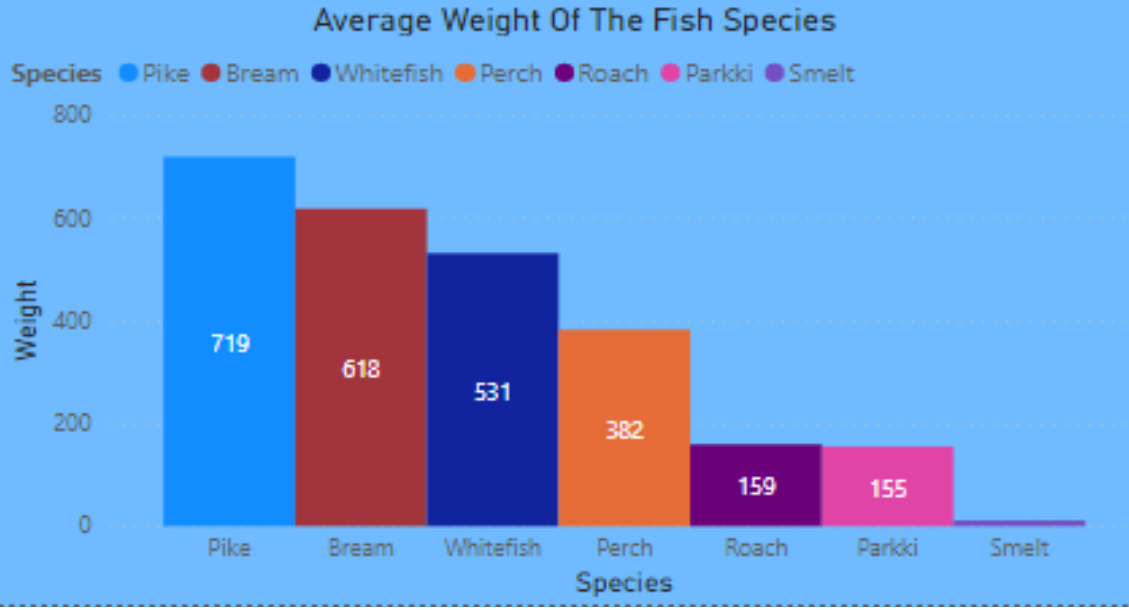
A strong positive correlation (0.72) exists between the height and the weight of the fish species, which implies that as weight increases, height also increases in a similar fashion and vice-versa.

The next page shows the visualization of the fish market data in Microsoft Power BI.
The next page after the visualization page, is the summary page for the insights observed in the data.



As a group, the bream species have the largest weight whereas, the smelt species are the smallest.
The smelts are also the least in comparison to the other species, whilst the perch species have the largest population.

Averagely, The pikes have the greatest weight, whiles the smelts weigh the least.



Summary Of Key Insights

- Averagely, pikes have the greatest weight whereas smelts weigh the least.
- There is a strong positive correlation between the weight and height of the fish species.
- As a group, breams have the greatest weight whiles the smelts are least.

Thanks for reading,
Visit kacelykodes.github.io to view similar projects.