Analysis Report on Fish Species Characteristics and Growth Patterns

Shashank shivakoti|ID: 23103084|Github

A.Introduction:

This report presents a data analysis on a fish dataset obtained from Kaggle. The goal is to examine various fish species and understand their physical characteristics, including length, height, and width. These insights contribute to understanding fish biology and ecology, as well as informing fisheries management.

B. Dataset Summary:

The dataset, known as the Fish Market Dataset, contains 159 entries and 7 features:
Species: Categorical, indicating fish type (e.g., Bream, Roach, Whitefish, Parkki, Perch, Pike, Smelt).
Weight: Measured in grams, representing overall biomass.
Length1, Length2: Length measurements in centimeters, critical for analyzing growth patterns.
Height: Physical height in centimeters.
Width: Width in centimeters, which indicates body shape and health.

C. Data Exploration:

Structure: Initial examination using `df.head()` and `df.info()` confirmed 159 rows and 7 columns.
Species Distribution: Analyzed using `df.Species.unique()` and `df.value_counts('Species')`.

D. Data Quality Checks:

Missing Values: Verified no missing data with `df.isnull().sum()`.
Duplicates: Confirmed absence of duplicates using `df.drop_duplicates()`.

E. Descriptive Statistics:

Weight: Mean of 398.33 grams with high variability (standard deviation 357.98).
Lengths: Length1 and Length2 have means of 26.25 cm and 28.42 cm, showing consistent distributions.
Height & Width: Means of 8.97 cm and 4.42 cm, indicating moderate variability.

This analysis provides a foundation for understanding fish species characteristics, supporting further research on growth trends and ecological patterns.

Count Plot:I extracted a count plot to display the number of entries for each fish species in the dataset. Bream is the most common, followed by Perch, while Roach, Whitefish, and Parkki have moderate counts. Pike and Smelt have fewer entries, which may indicate sampling biases or habitat availability.
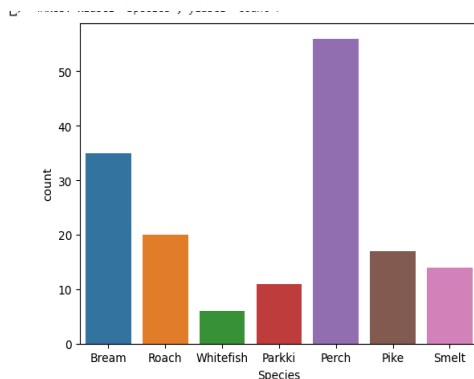

Box Plot:I extracted a box plot to visualize the distribution of heights across different species. For example, by looking at the Bream species, the minimum height is around 7 cm, the maximum height reaches up to 15 cm, and the average height typically falls between 10-12 cm. The box plot visually presents the minimum,

maximum, and average values, making it easier to identify the spread and central tendency of each species' height. This visualization aids in comparing species and understanding their physical variation
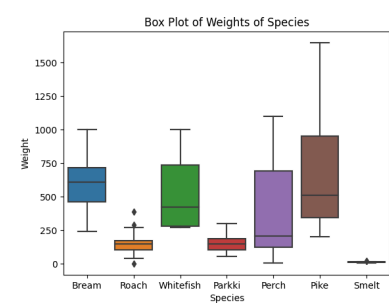
Correlation Analysis:A correlation heatmap using `df.corr()` reveals strong relationships within the dataset. Length1, Length2, and Length3 exhibit near-perfect positive correlations (0.99), suggesting they represent similar aspects of fish length. Additionally, Length3 and Weight show a strong positive correlation (0.89), indicating that longer fish generally have more weight. Moderate correlations are observed between Height and Weight (0.72) and Height and Width (0.63), implying that taller fish tend to be heavier and slightly broader, offering valuable insights into fish growth patterns.

.
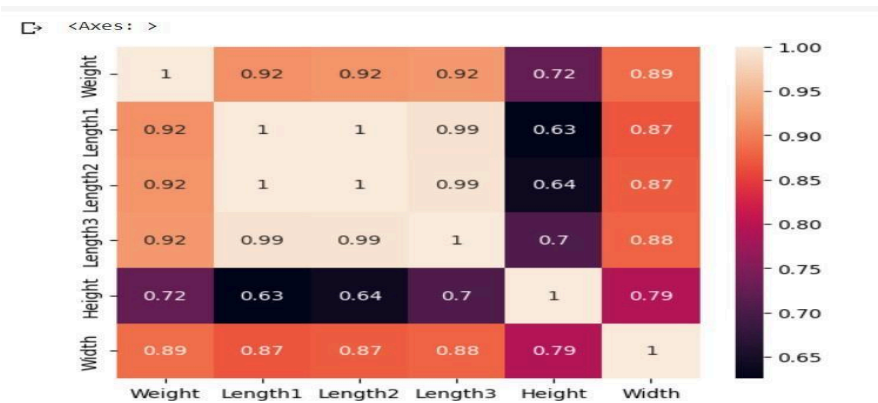
Conclusion:

Based on the analysis, I found strong correlations between fish length and weight, with larger fish being heavier. The species distribution showed Bream and Perch as the most frequent. The box plot revealed key insights into height variations, offering a clearer understanding of species characteristics and growth patterns.