Technical Report Proposal

I am using ” Fish Market ” Data Set from Kaggle

Student name - Sahil Rajesh Gidwani
Student number - L00171407
student Email - L00171407@atu.ie

I. Problem Description

I have selected the Fish Market Dataset from Kaggle, which has data for at least 7 different species of fish. This dataset of fish has different characteristics, such as vertical length , diagonal length , cross length , weight, and height, which are nessesary for me to make the analytic pipeline as well as models. With this dataset, a predictive model can be performed using machine-friendly data to estimate the weight of fish.

II. Goal

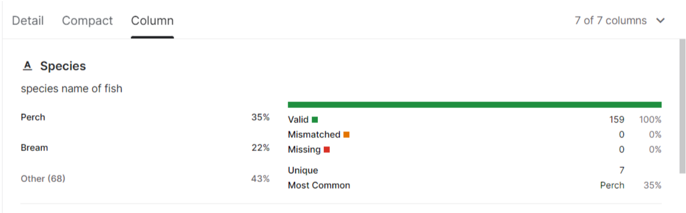
MY goal is to find biggest weighting fish , largest fish , widthest fish . All this data can be extracted from the fish market dataset on Kaggle . This dataset has all the data about the fishes and their measurements, along with their species names.

This dataset has 7 different dataset with information about weight , height and length in form of diagonal , vertical, cross .

With this dataset I can create a predictive model to estimate weight of the fish

Therefore my goal will be to make this model work for the predictive analytic.

III. Description of data



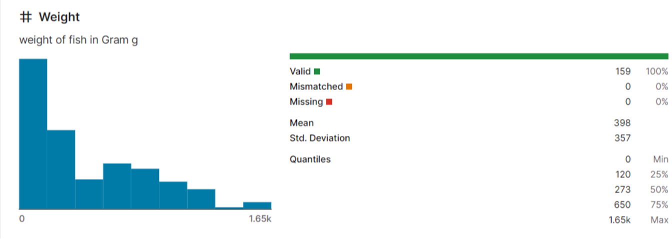
Species

species name of fish

Perch	35%	Valid 🟩	159	100%
		Mismatched 🟧	0	0%
Bream	22%	Missing 🟥	0	0%
		Unique	7	
Other (68)	43%	Most Common	Perch	35%

Fig. 1. Example of Figure 1.the image above shows how accurate the data is .

The dataset contains only precise and accurate data about fish species and their sub-data.



Weight

weight of fish in Gram g

Valid 🟩	159	100%
Mismatched 🟧	0	0%
Missing 🟥	0	0%
Mean	398	
Std. Deviation	357	
Quantiles	0	Min
	120	25%
	273	50%
	650	75%
	1.65k	Max

Fig. 2. Example of Figure 2. Above image shows visualized data of the column ’weight’

The dataset has sufficient data to obtain the results for the weight and length of the fish , which are the main focus of this project.

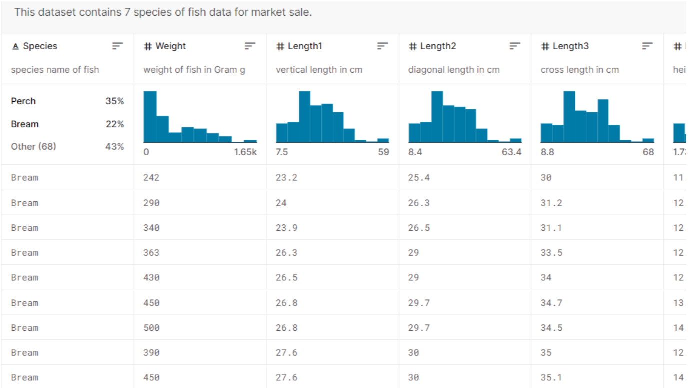
The data is consistent and reliable for the use.

The data is unique and of high quality , perfect for my use .

I have taken this dataset from Kaggle.

The title of the dataset is ”Fish Market.” it contains data of fish.

The link of the dataset is https://www.kaggle.com/datasets/aungpyaeap/fish-market?resource=download



This dataset contains 7 species of fish data for market sale.

Species	# Weight	# Length1	# Length2	# Length3	#
species name of fish	weight of fish in Gram g	vertical length in cm	diagonal length in cm	cross length in cm	hei
Perch 35% Bream 22% Other (68) 43%					
Bream	242	23.2	25.4	30	11
Bream	290	24	26.3	31.2	12
Bream	340	23.9	26.5	31.1	12
Bream	363	26.3	29	33.5	12
Bream	430	26.5	29	34	12
Bream	450	26.8	29.7	34.7	13
Bream	500	26.8	29.7	34.5	14
Bream	390	27.6	30	35	12
Bream	450	27.6	30	35.1	14

Fig. 3. Example of Figure 3. Above table shows the column and data I am going to use in this project

IV. Approach

I will extract the data from kaggle through API .
Then I will manipulate the data from data set.

Then cleanse and convert it to SQL format for further computing.
Then I will create the model for decision making.
I will create two predictive models. first to predict large weighting fish and second to predict largest fish in length
with above steps I will be able to create analytic pipeline.
I will be using Apache Spark in this pipeline. Spark utilizes optimized query execution and in-memory caching for rapid queries across any size of data. It is simply a general and fast engine for much larger-scale processing of data.
I am going to use DataBrick , which has cluster optimization for big data. Databricks Runtime for Machine Learning (Databricks Runtime ML) automates the creation of a cluster optimized for machine learning. Databricks Runtime ML clusters include the most popular machine learning libraries

V. Bibliography

As a result, using the Kaggle data set ”Fish Market,” I will complete the desired predictive model and answer the question. I will create the full analytic pipeline required for the model . The model’s results will also be visualized in a graph.

References

Kaggle - https://www.kaggle.com/datasets/aungpyaeap/fish-market