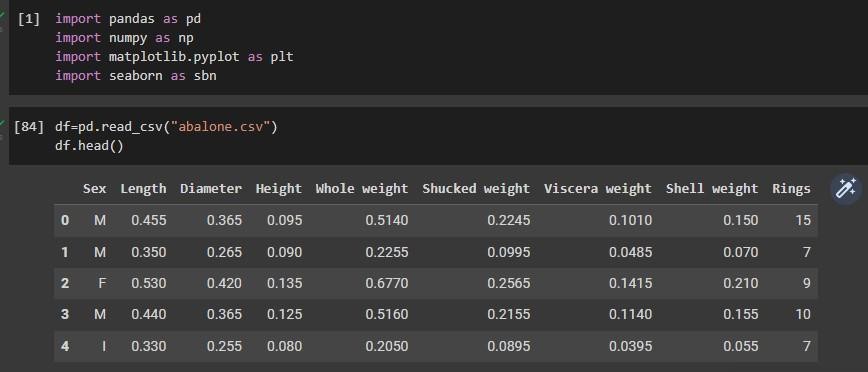
ASSIGNMENT - 4

|  |  |
| --- | --- |
| Assignment Date | 15 October 2022 |
| Student Name | Sadhana M |
| Student Roll Number | 720719104140 |
| Maximum Marks | 2 Marks |

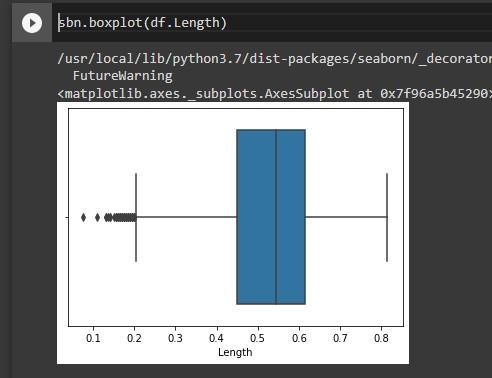
**Dataset**: [https://drive.google.com/file/d/1sIv7x7CE0zAPAt0Uv6pbO2ST2LVp5u/view](https://drive.google.com/file/d/1sIv-7x7CE0zAPAt0Uv-6pbO2ST2LVp5u/view)

**Loading the dataset:**

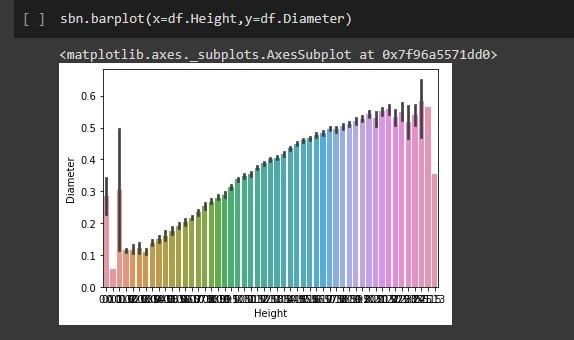


**Perform Below Visualizations.**

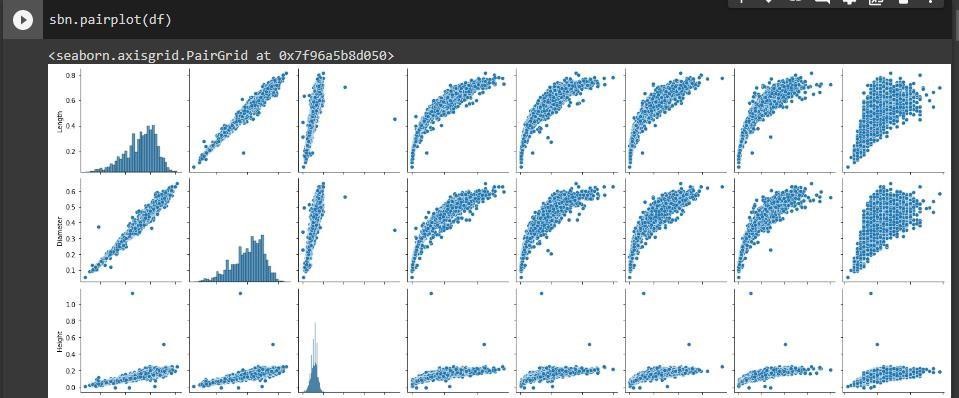
**∙ Univariate Analysis**



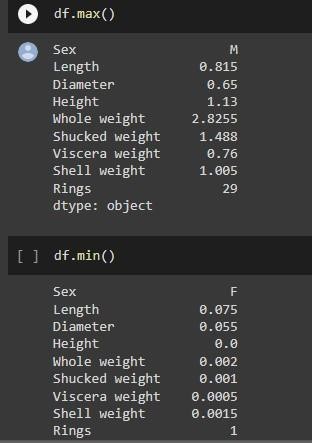
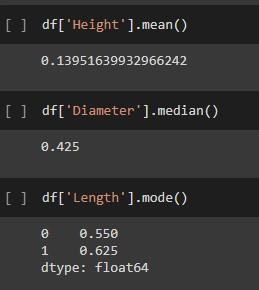
**Bi-Variate Analysis**



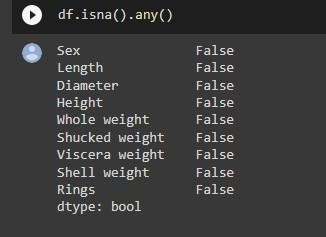
**Multi-Variate Analysis**



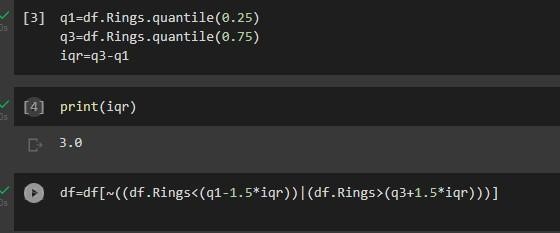
# Perform descriptive analytics on the dataset



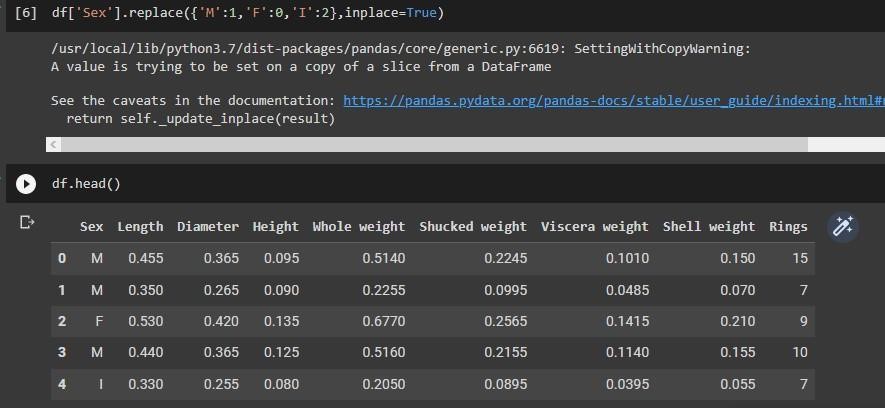
**Check for Missing values and deal with them.**



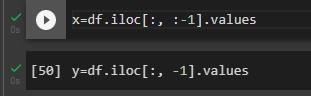
# Find the outliers and replace them outliers



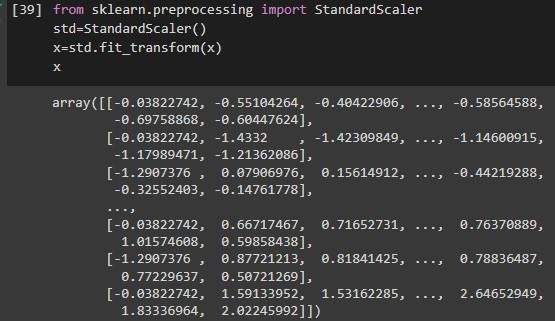
**Check for Categorical columns and perform encoding.**



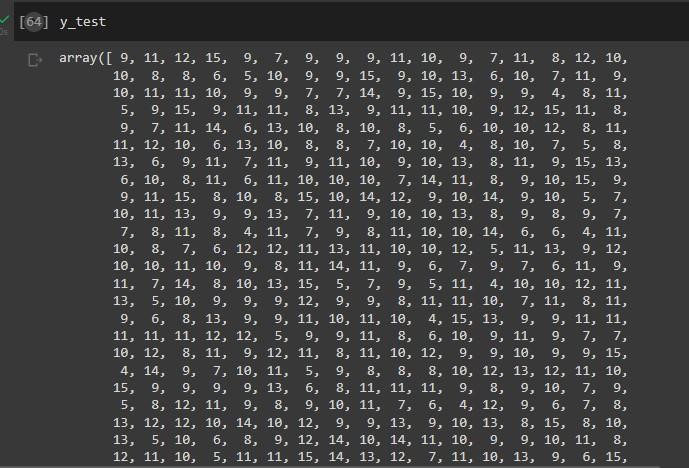
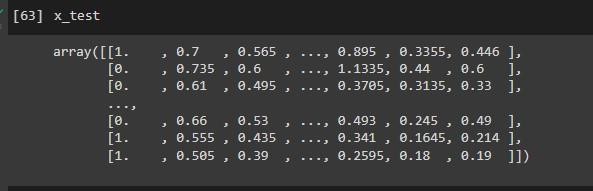
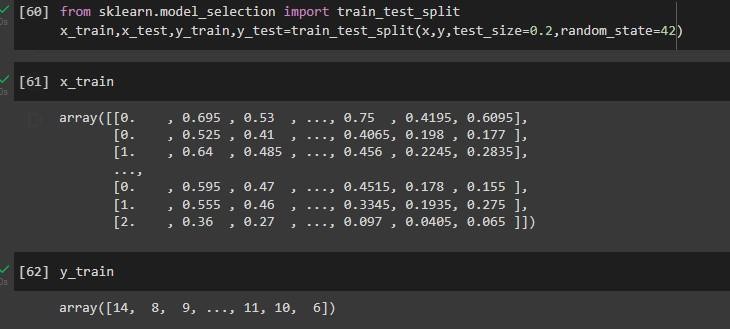
**Split the data into dependent and independent variables.**



# Scale the independent variables



# Split the data into training and testing

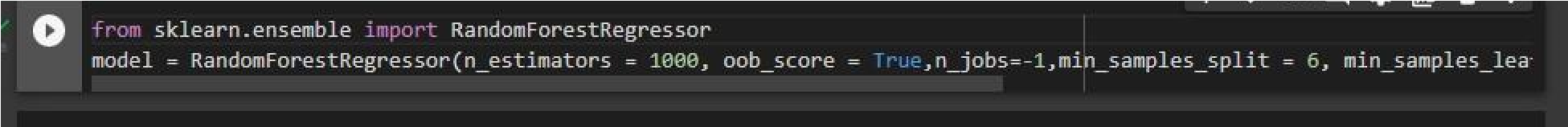


# Build the Model

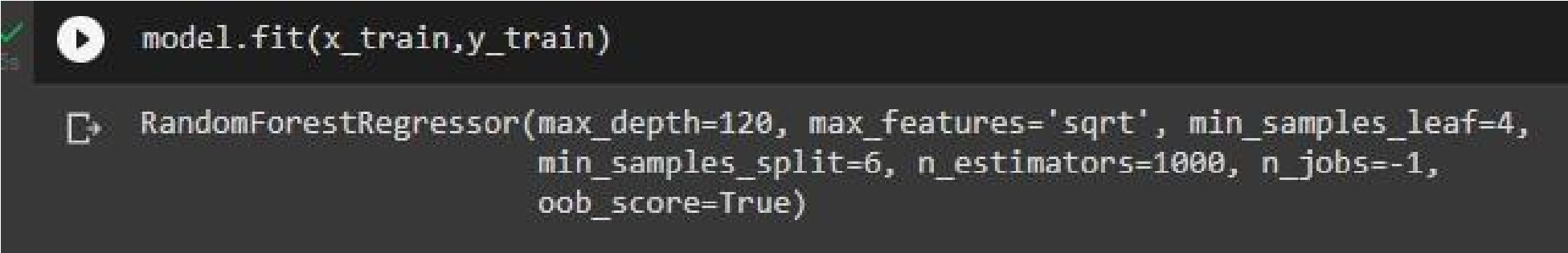
from sklearn.ensemble import RandomForestRegressor model =

RandomForestRegressor(n\_estimators = 1000, oob\_score =

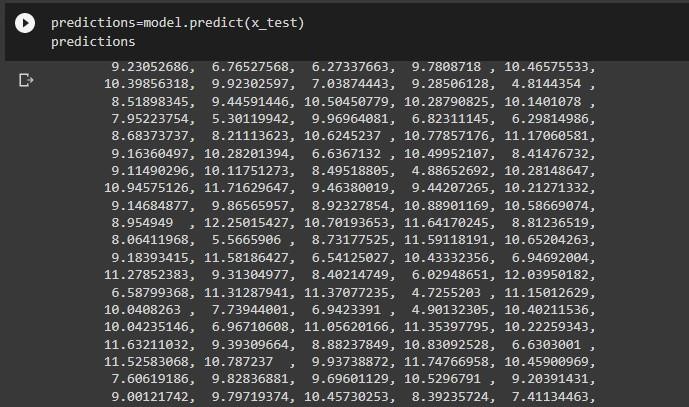
True,n\_jobs=1,min\_samples\_split = 6, min\_samples\_leaf= 4, max\_features = 'sqrt', max\_depth= 120, bootstrap=True)



# Train the Model



# Test the Model



**Measure the performance using Metrics.**

