

ASSINGMENT-4

Problem Statement: Abalone Age Prediction

Predicting the age of abalone from physical measurements. The age of abalone is determined by cutting the shell through the cone, staining it, and counting the number of rings through a microscope - a boring and time-consuming task.

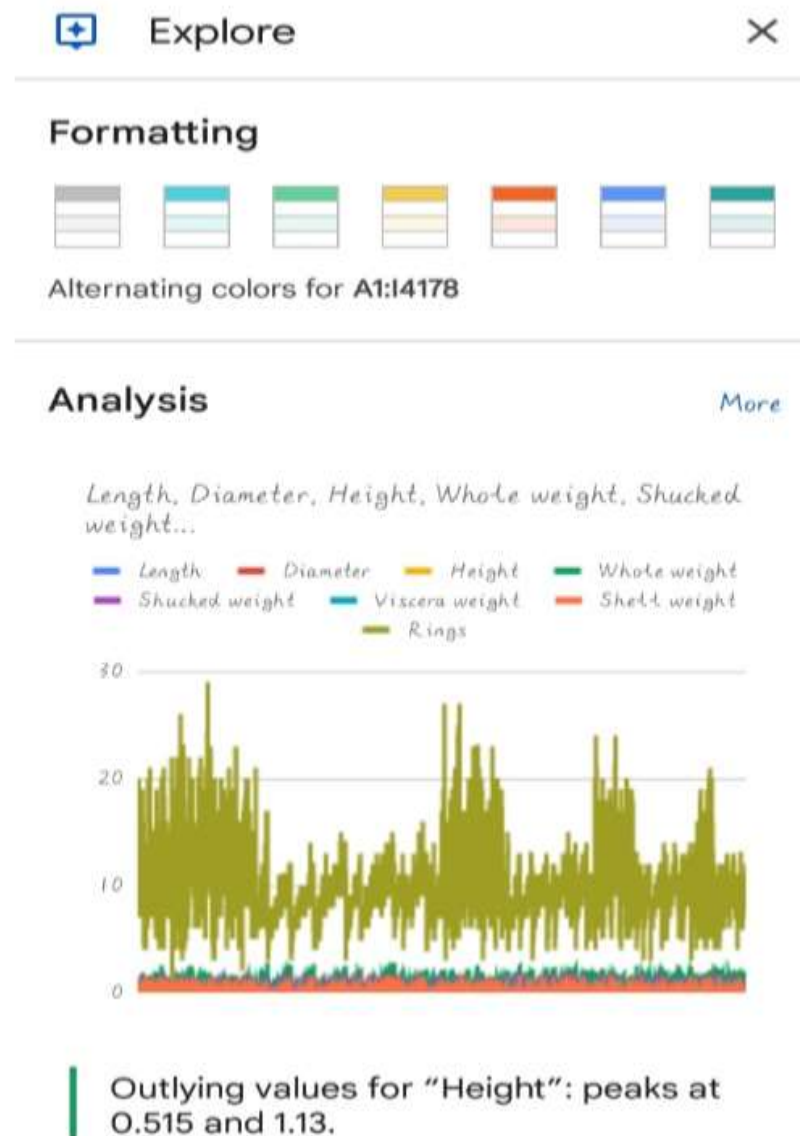
Attribute Information:

Given is the attribute name, attribute type, measurement unit, and a brief description. The number of rings is the value to predict: either as a continuous value or as a classification problem.

Name /Data Type / Measurement Unit / Description

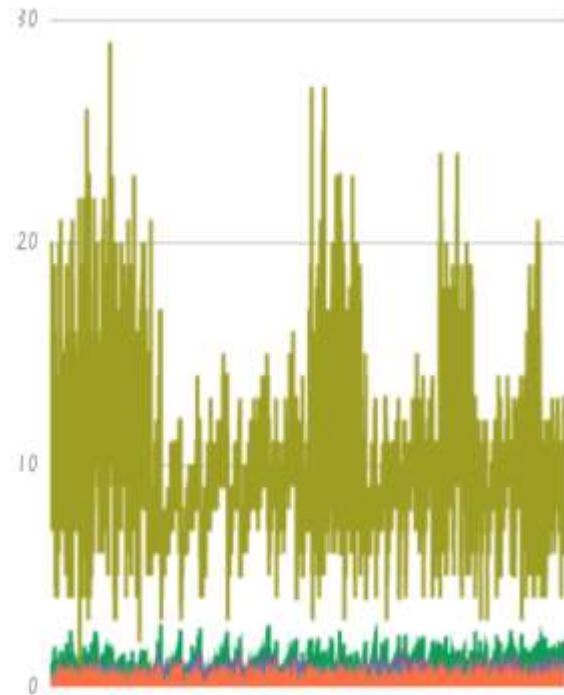
	Sex	Length	Diameter	Height	Whole weight	Shucked weight	Viscera weight	Shell weight	Rings
2	M	0.455	0.365	0.095	0.514	0.2245	0.101	0.15	15
3	M	0.35	0.265	0.09	0.2255	0.0995	0.0485	0.07	7
4	F	0.53	0.42	0.135	0.677	0.2565	0.1415	0.21	9
5	M	0.44	0.365	0.125	0.516	0.2155	0.114	0.155	10
6	I	0.33	0.255	0.08	0.205	0.0895	0.0395	0.055	7
7	I	0.425	0.3	0.095	0.3515	0.141	0.0775	0.12	8
8	F	0.53	0.415	0.15	0.7775	0.237	0.1415	0.33	20
9	F	0.545	0.425	0.125	0.768	0.294	0.1495	0.26	16
10	M	0.475	0.37	0.125	0.5095	0.2165	0.1125	0.165	9
11	F	0.55	0.44	0.15	0.8945	0.3145	0.151	0.32	19
12	F	0.525	0.38	0.14	0.6065	0.194	0.1475	0.21	14
13	M	0.43	0.35	0.11	0.406	0.1675	0.081	0.135	10

Download the dataset:
Load the dataset into the tool.



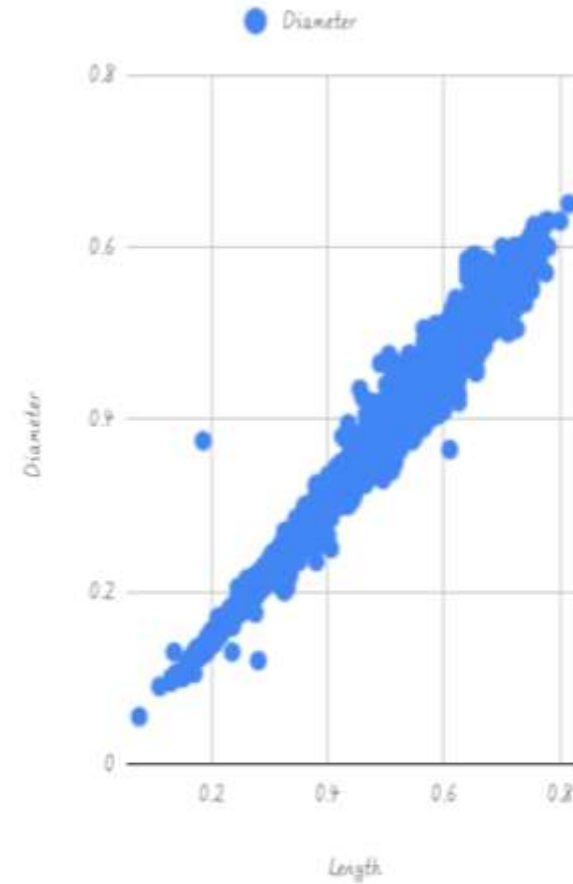
Length, Diameter, Height, Whole weight, Shucked weight...

Length Diameter Height Whole weight
Shucked weight Viscera weight Shell weight
Rings



Outlying values for "Height": peaks at 0.515 and 1.13.

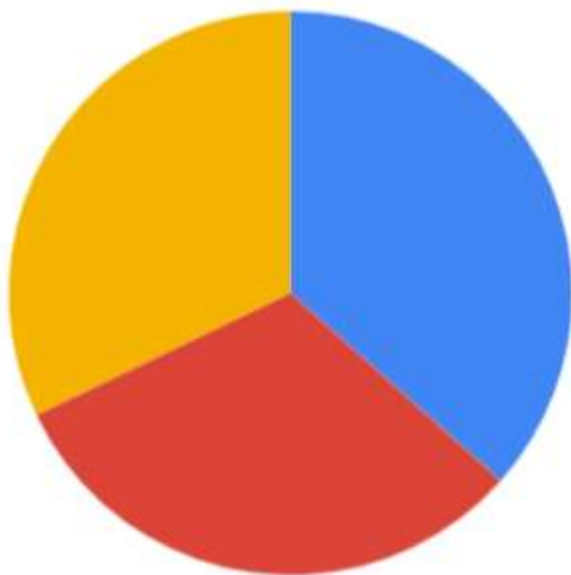
Diameter vs. Length



For every increase of 0.1 in "Length", "Diameter" increases by about 0.0813.

Count of Sex

● M ● F ● I



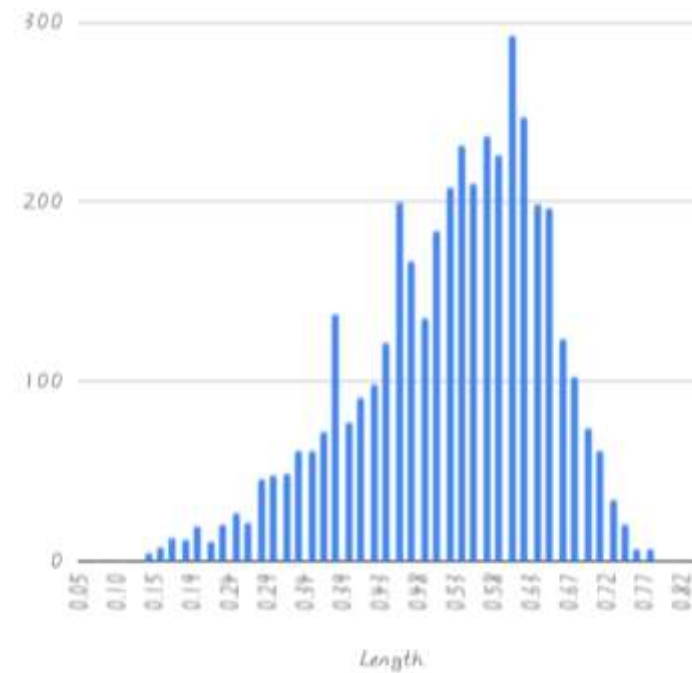
<

3/9

>

Histogram of Length

■ Length (count)



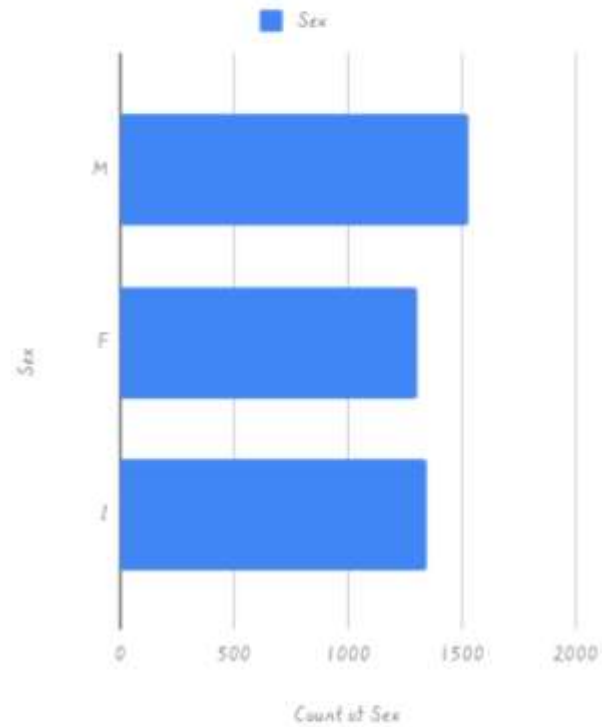
Ranges from 0.075 to 0.815, but most values are around 0.524, plus or minus 0.101.

<

4/9

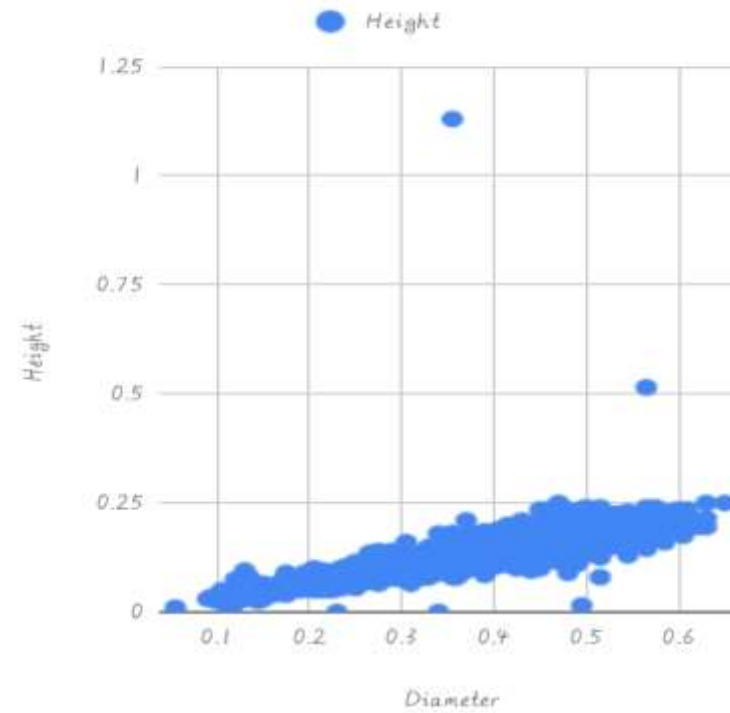
>

Count of Sex



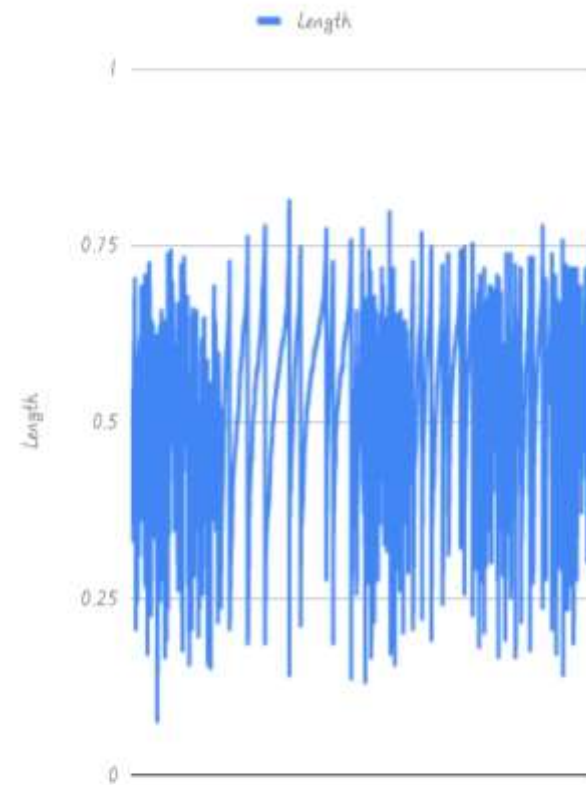
"M" has the highest value for "Sex" (1528).

Height vs. Diameter



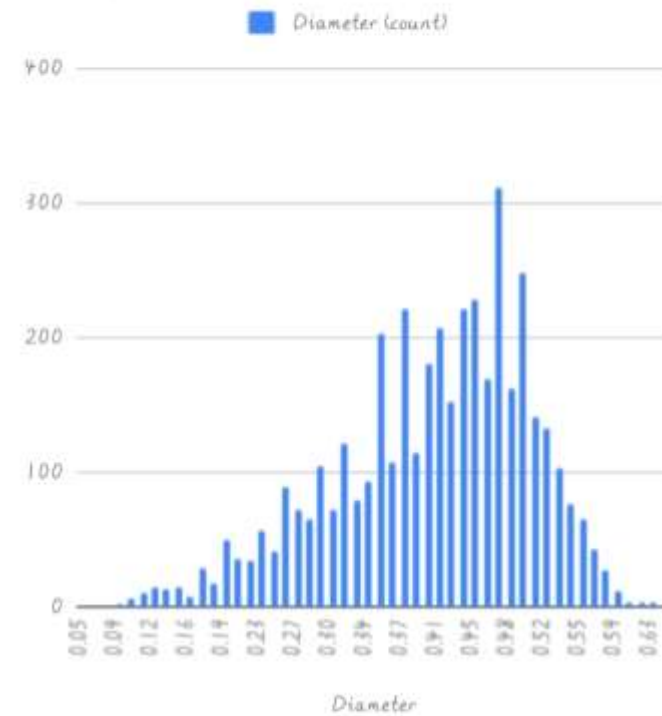
For every increase of 0.1 in "Diameter", "Height" increases by about 0.0371.

Length



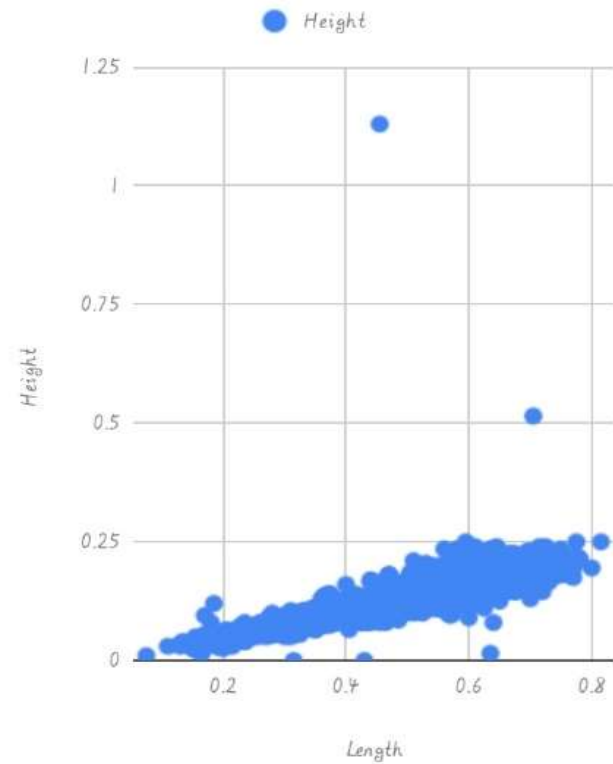
Outlying values for "Length": dips to 0.075, 0.11, 0.13, and 0.14.

Histogram of Diameter



Ranges from 0.055 to 0.65, but most values are around 0.408, plus or minus 0.0829.

Height vs. Length

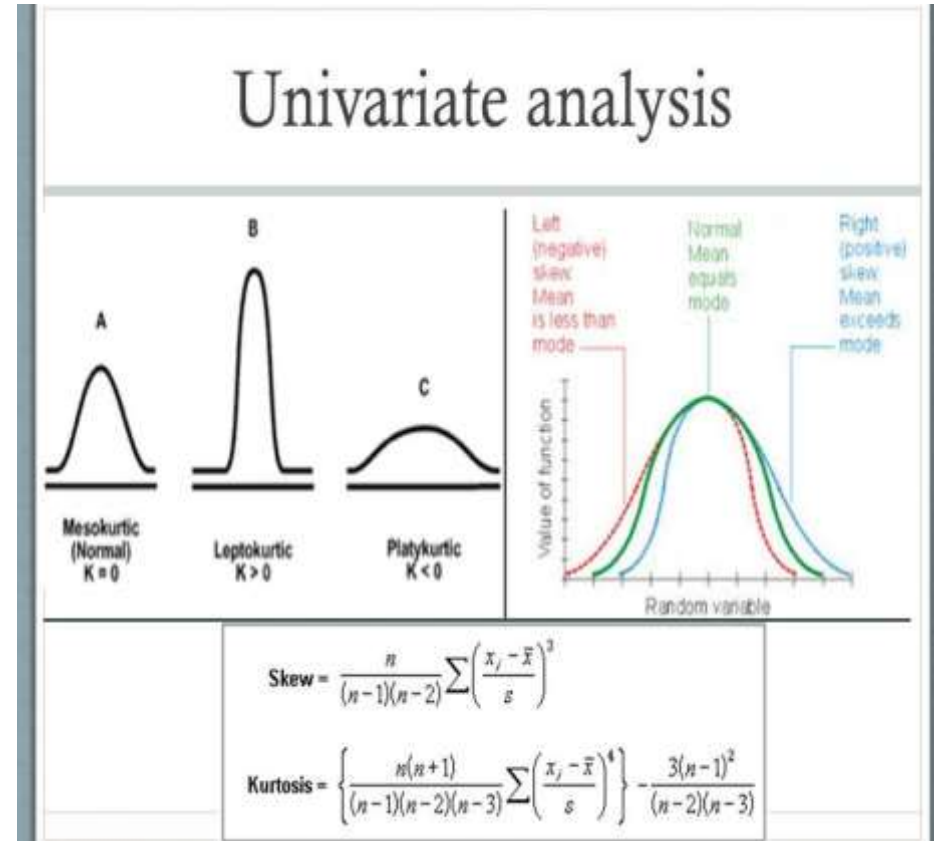


For every increase of 0.1 in "Length",
"Height" increases by about 0.0283.

- Other measurement, which are easier to obtain, are used to predict age.
- Further information, such as weather patterns and location (hence food availability) may be required to solve the problem

Perform Below Visualizations, Univariate Analysis,

Its commonly used in statistics to describe a type of data which consists of observations on only a single characteristic or attribute.

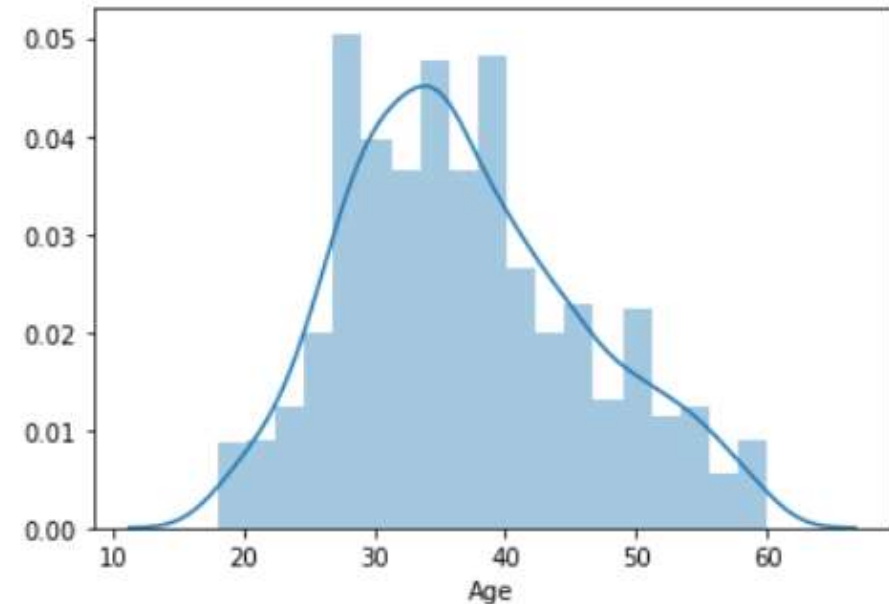


Bi-variate Analysis,

- Its one of the simplest forms of quantitative analysis.
- It involves the analysis of two variables, for the purpose of determining the empirical relationship between them.

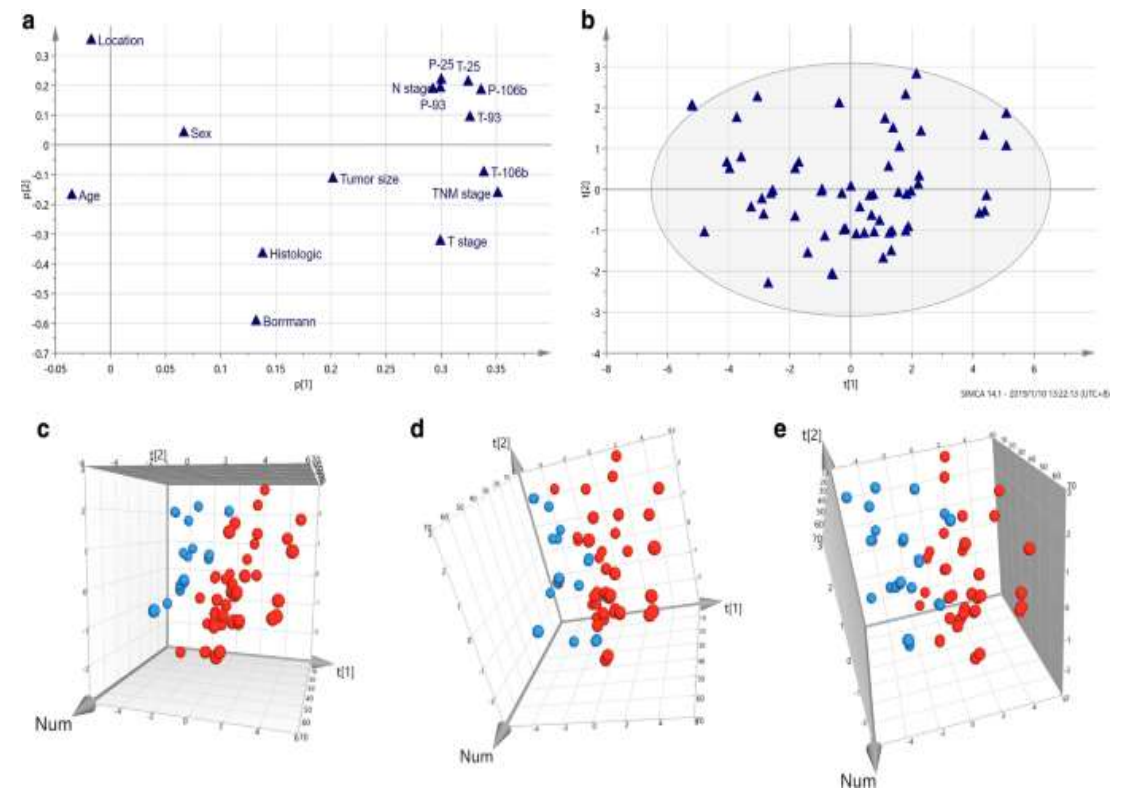
```
#Distribution of Age variable  
sns.distplot(data['Age'])
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x1f68a445a20>
```

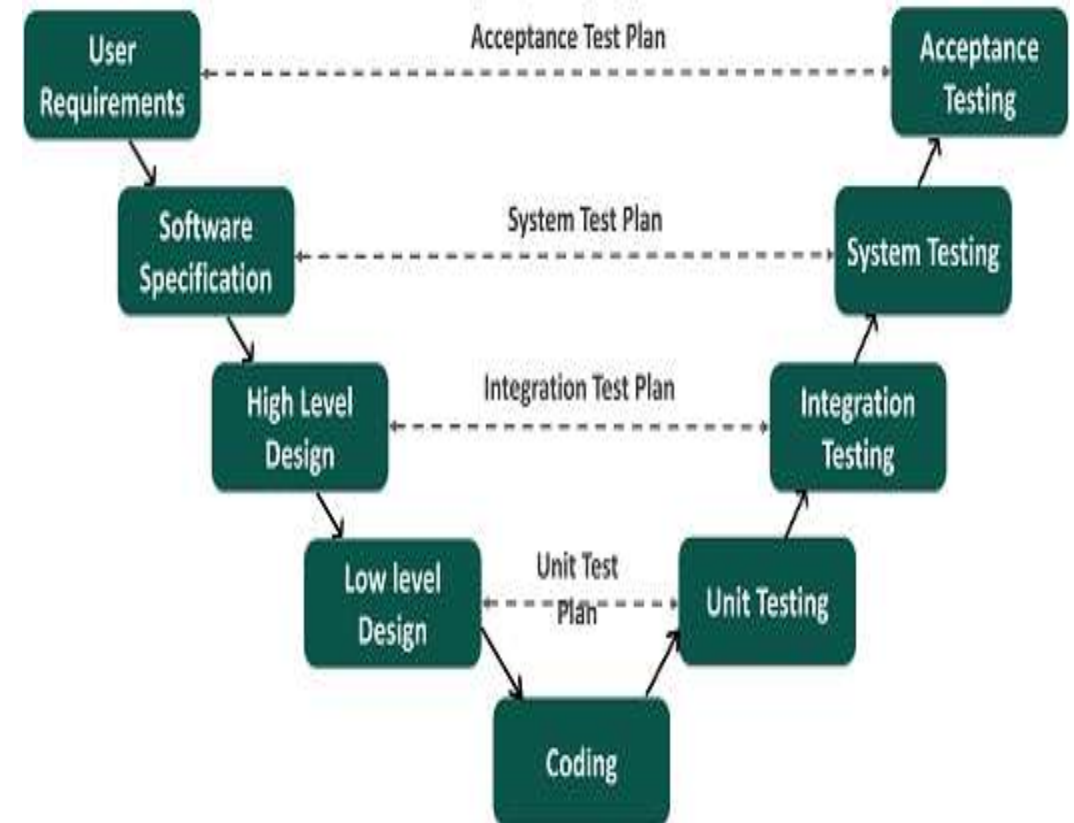
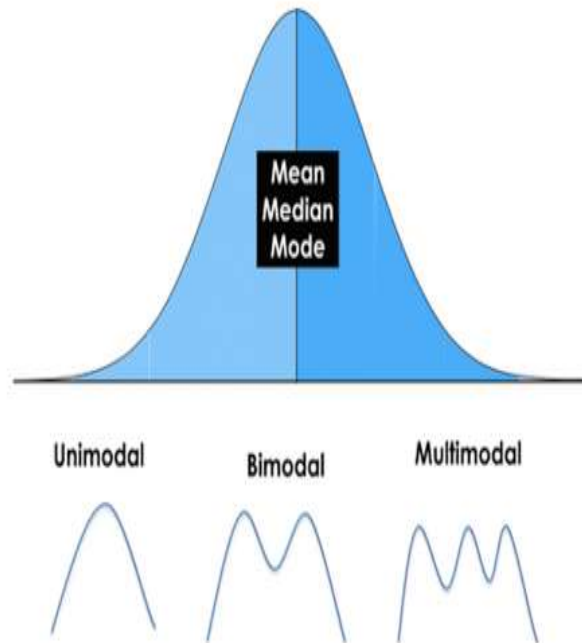
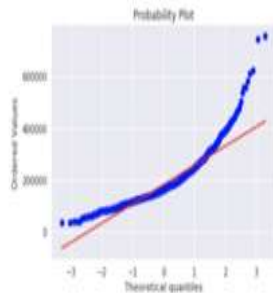
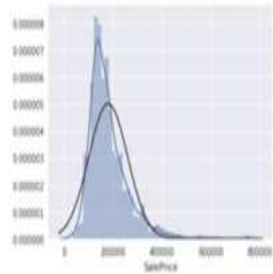


Multi variate Analysis,

Its defined as: The statistical study of data where multiple measurements are made on each experimental unit and where the relationships among multivariate measurements and their structure are important.



Perform descriptive statistics on the dataset,
Scale the independent variables,
split the data into training and testing,
Measure the performance using metrics,



- Performance metrics are data used to track processes within a business.
- Descriptive statistics are brief informational coefficients that summarize a given data set, which can be either a representation of the entire population or a sample of a population.
- A training model is a dataset that is used to train an ML algorithm.
- Model-based testing is an application of model-based design for designing and optionally also executing artifacts to perform software testing or system testing.