Standard Deviation

Customer ID	Name	Surname	Gender	Age	Age Grou	Height	egion	Job Classification	Tenure Months	Balance	Spend On Groceries
200000262	Zoe	Clarkson	Female	59	5	62	cotland	Other	24	23550.89	70.77
200001214	Carolyn	McDonald	Female	58	5	61.2	cotland	Other	24	69027.62	67.1
400000497	Anna	Chapman	Female	26	2	65.1	orthern Ireland	White Collar	46	5789.63	46.23
400001939	Richard	Dowd	Male	21	2	70.9	orthern Ireland	White Collar	23	10248.59	36.48
300002298	Phil	Arnold	Male	37	3	70.4	Vales	Blue Collar	15	80824.89	36.11

$$\{ 61.2, 62, 65.1, 70.4, 70.9 \}$$

$$\text{Mean} = \frac{61.2 + 62 + 65.1 + 70.4 + 70.9}{5} = 65.92$$

$$\{\ 61.2,\ 62,\ 65.1,\ 70.4,\ 70.9\ \}$$

$$\mu \quad \text{Mean} = \frac{61.2 + 62 + 65.1 + 70.4 + 70.9}{5} = 65.92$$

$$\text{Variance} = \frac{(61.2 - 65.92)^2 + (62 - 65.92)^2 + (65.1 - 65.92)^2 + (70.4 - 65.92)^2 + (70.9 - 65.92)^2}{5}$$

$$\text{Variance} = \sigma^2 = \frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N} = 16.64$$

$$\text{Std. Dev.} = \sigma = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}} = 4.08$$

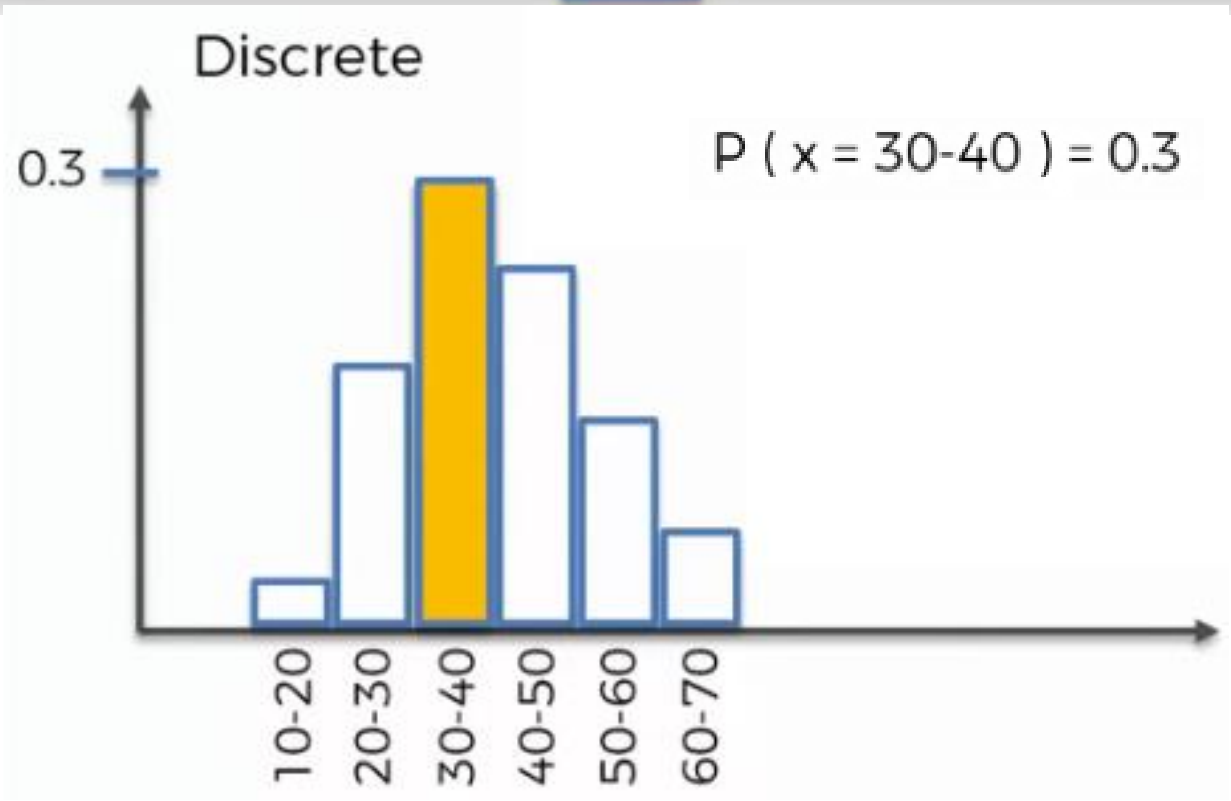
What is Distribution?

In probability theory and statistics, a probability distribution is a mathematical function that provides the probabilities of occurrence of different possible outcomes in an experiment.

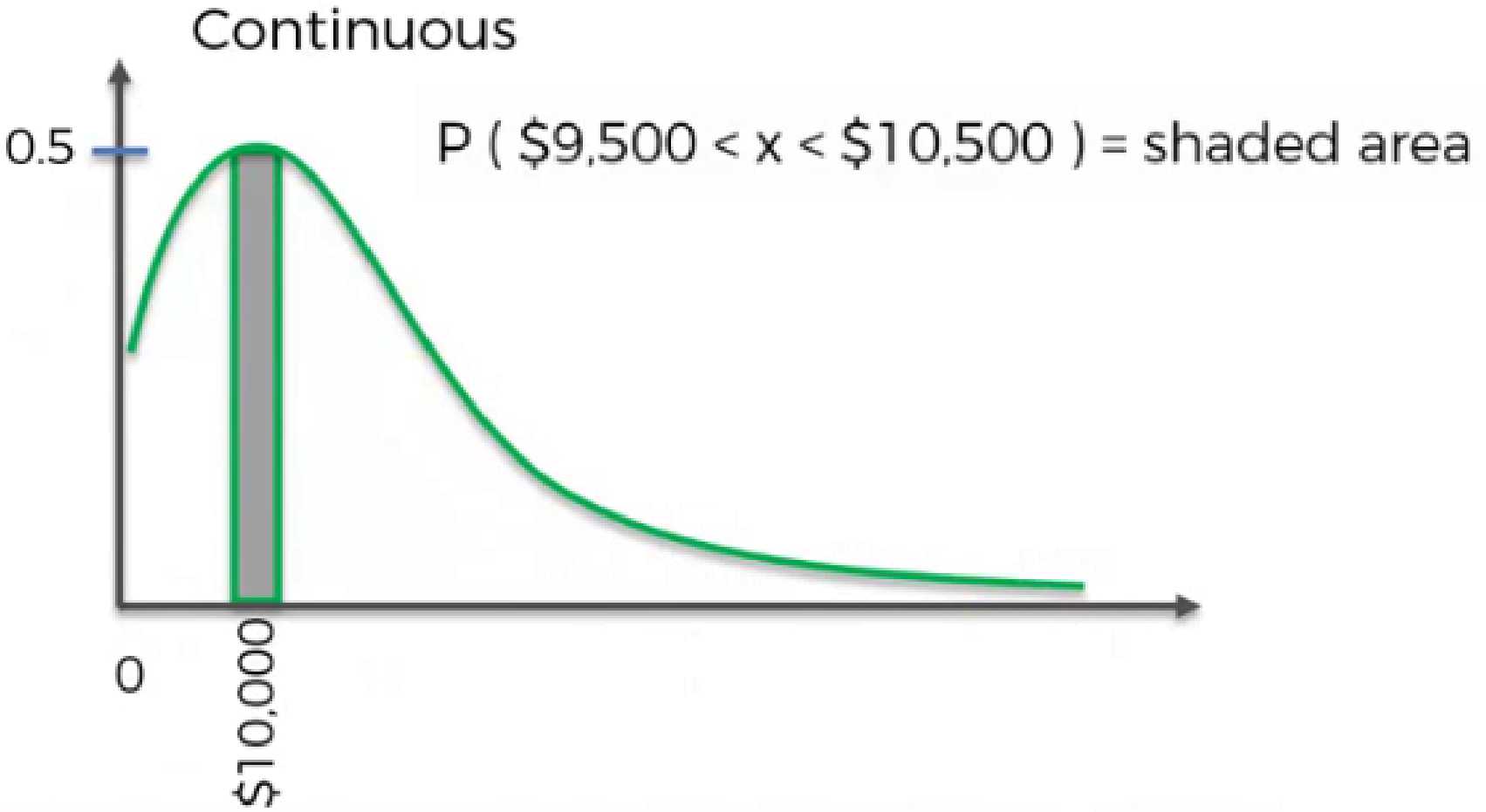
What is Distribution?

Discrete **Continuous**

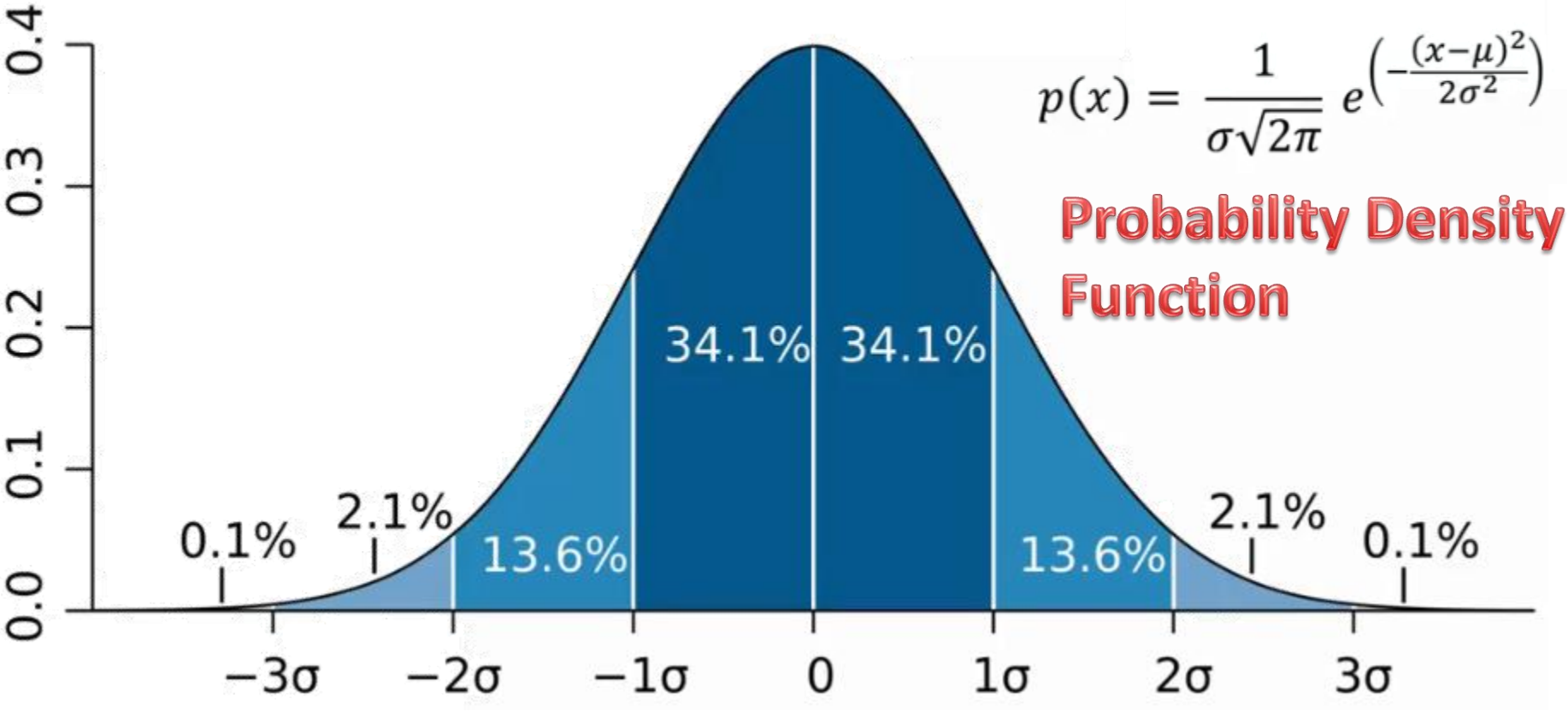
Customer ID	Name	Surname	Gender	Age	Age Group	Height	Region	Job Classification	Tenure Months	Balance	Spend On Groceries
200000262	Zoe	Clarkson	Female	59	50	62	Scotland	Other	24	23550.89	70.77
200001214	Carolyn	McDonald	Female	58	50	61.2	Scotland	Other	24	69027.62	67.1
400000497	Anna	Chapman	Female	26	20	65.1	Northern Ireland	White Collar	46	5789.63	46.23
400001939	Richard	Dowd	Male	21	20	70.9	Northern Ireland	White Collar	23	10248.59	36.48
300002298	Phil	Arnold	Male	37	30	70.4	Wales	Blue Collar	15	80824.89	36.11



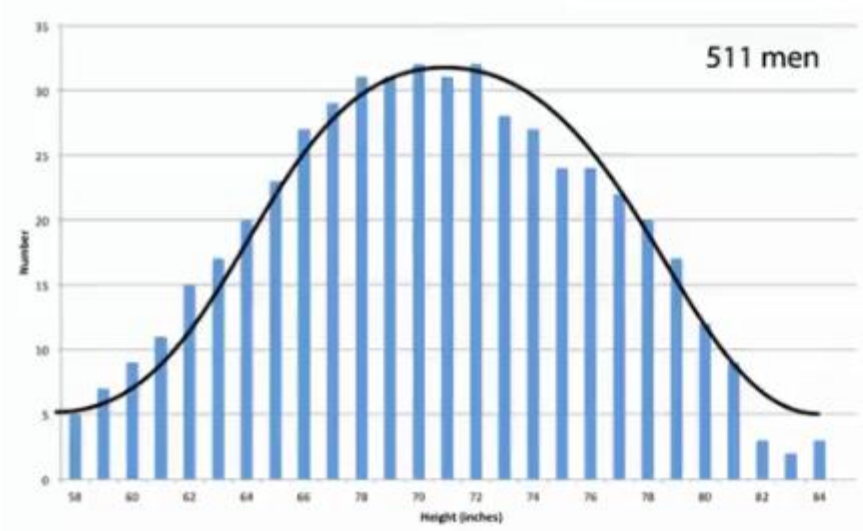
....

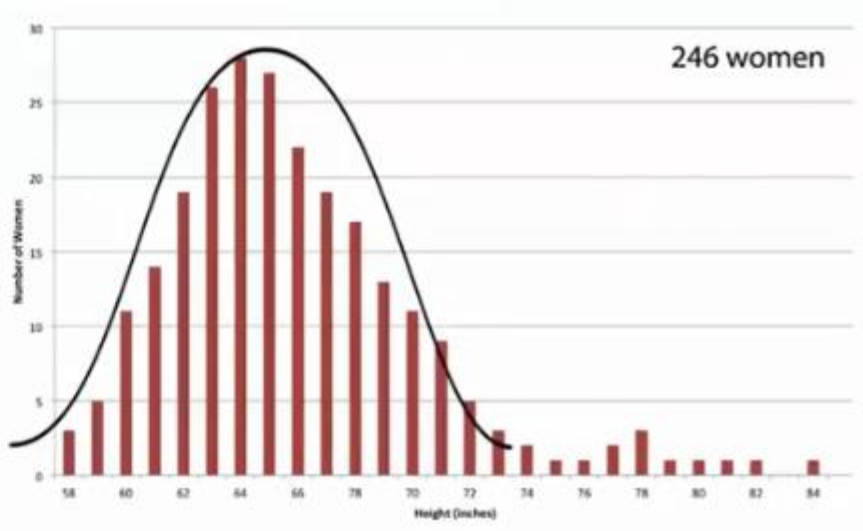

Continuous
0.5
P ($9,500 < x < $10,500) = shaded area
0
$10,000

Normal Distribution



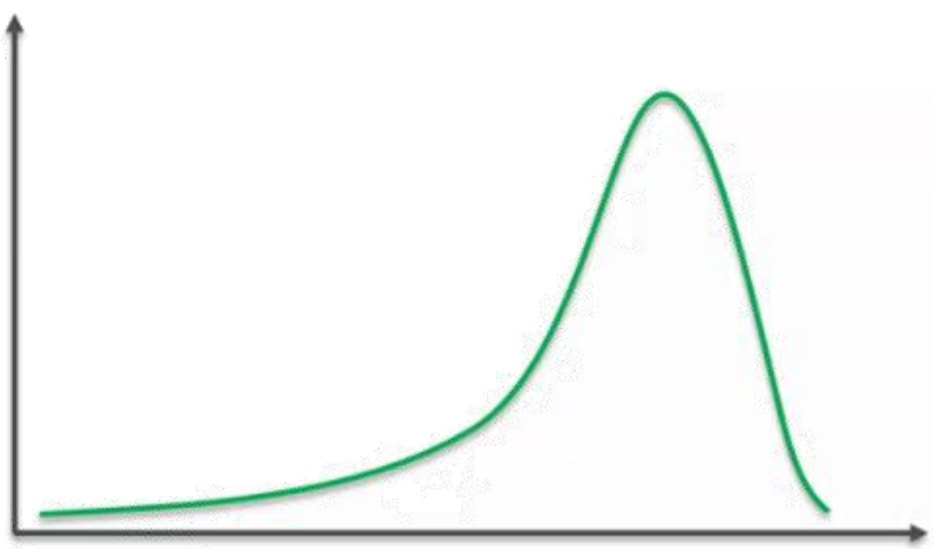
Height distribution of 20-yr old men and women in India



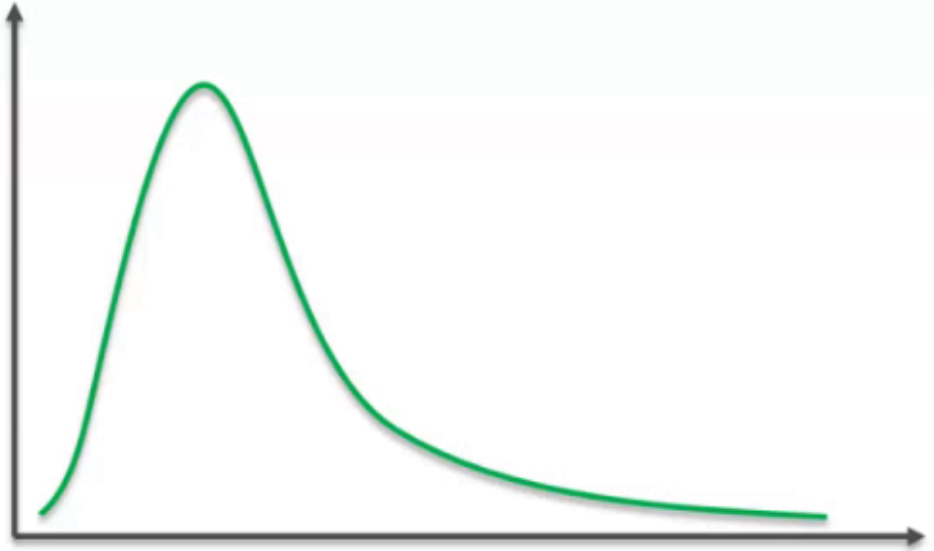


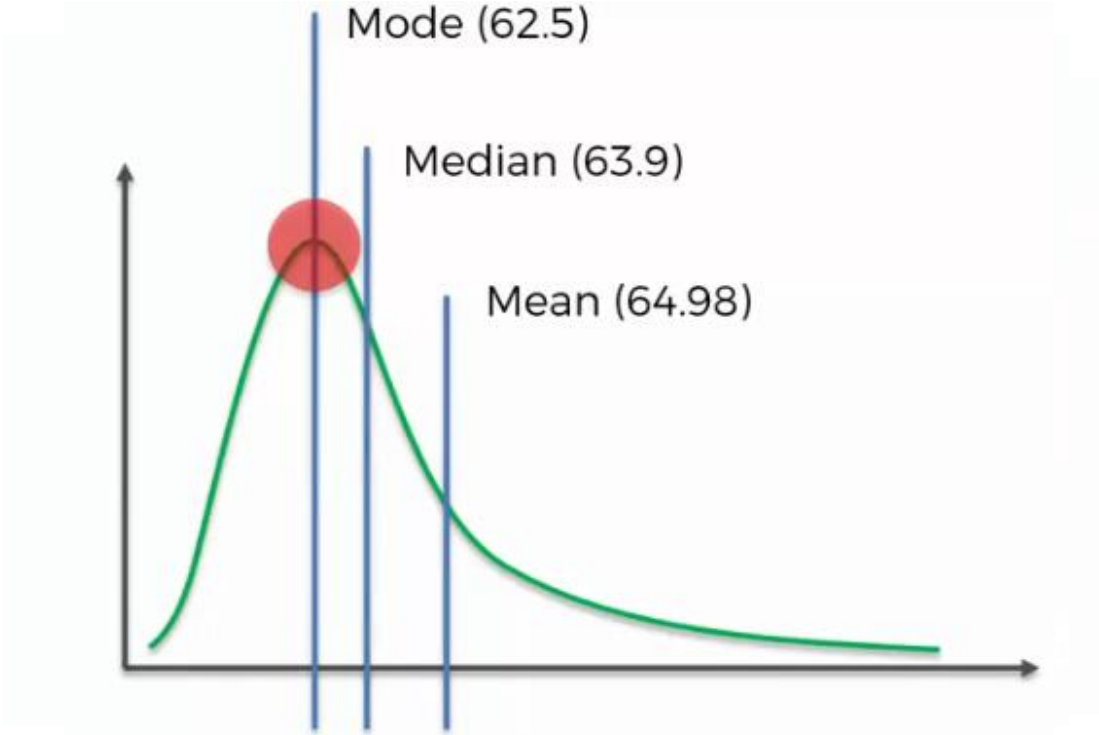
Skweness

Left (Negative) Skew



Right (Positive) Skew



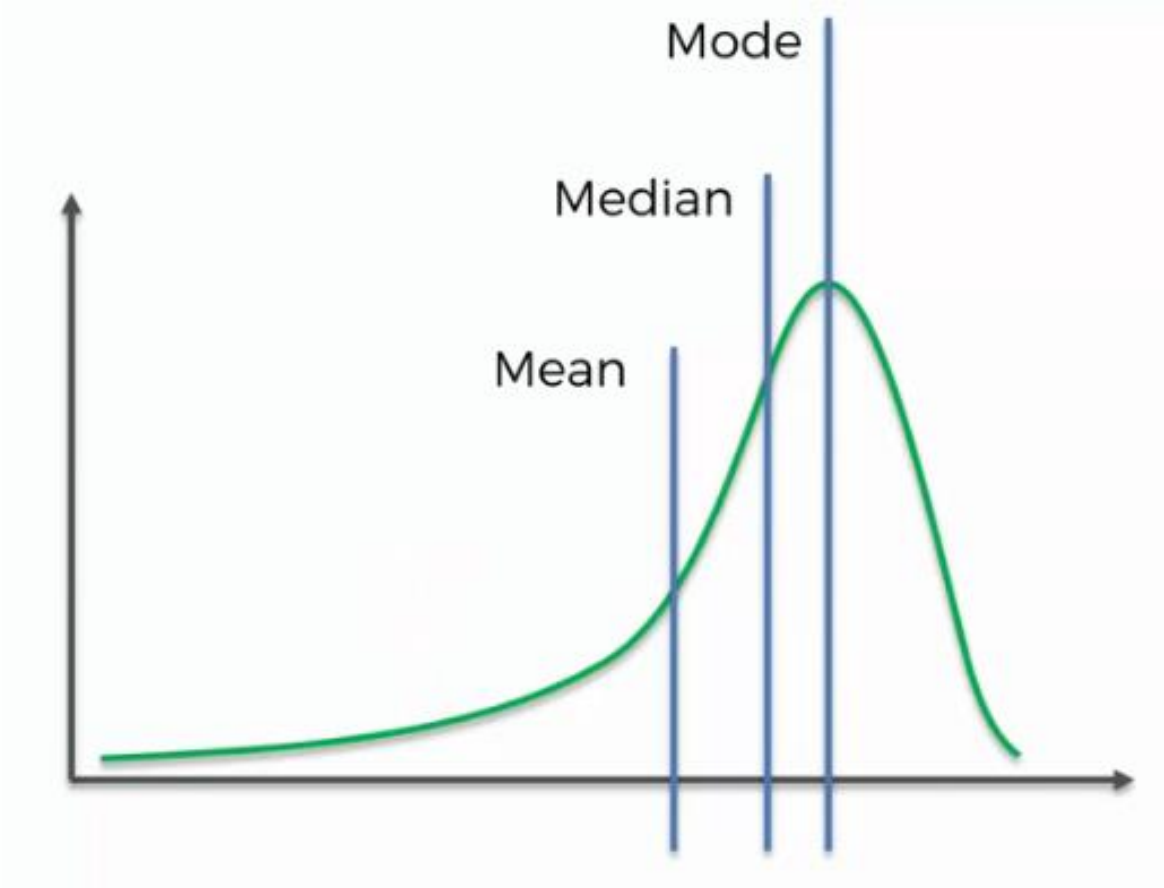


{ 58.8 59.9 61.2 61.3 61.5 62 62.5 62.5 62.5 63 63 63.7 63.9 64 64 65.1 65.2 66.7 67.8 68.3 69.9 70.7 71.8 72.2 73 }

{ 58.8 59.9 61.2 61.3 61.5 62 62.5 62.5 62.5 63 63 63.7

63.9

64 64 65.1 65.2 66.7 67.8 68.3 69.9 70.7 71.8 72.2 73 }


Mode
Median
Mean

Seaborn Package

- Seaborn is a Python visualization library based on matplotlib.
- It provides a high-level interface for drawing attractive statistical graphics.

import seaborn

seaborn.distplot()

Creating a univariate distribution in seaborn with distplot()

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings("ignore")
plt.figure(figsize=(3,4))
sns.distplot(DF.InternetAccess )
plt.show()
```

Probability density : Probability per unit on the x-axis