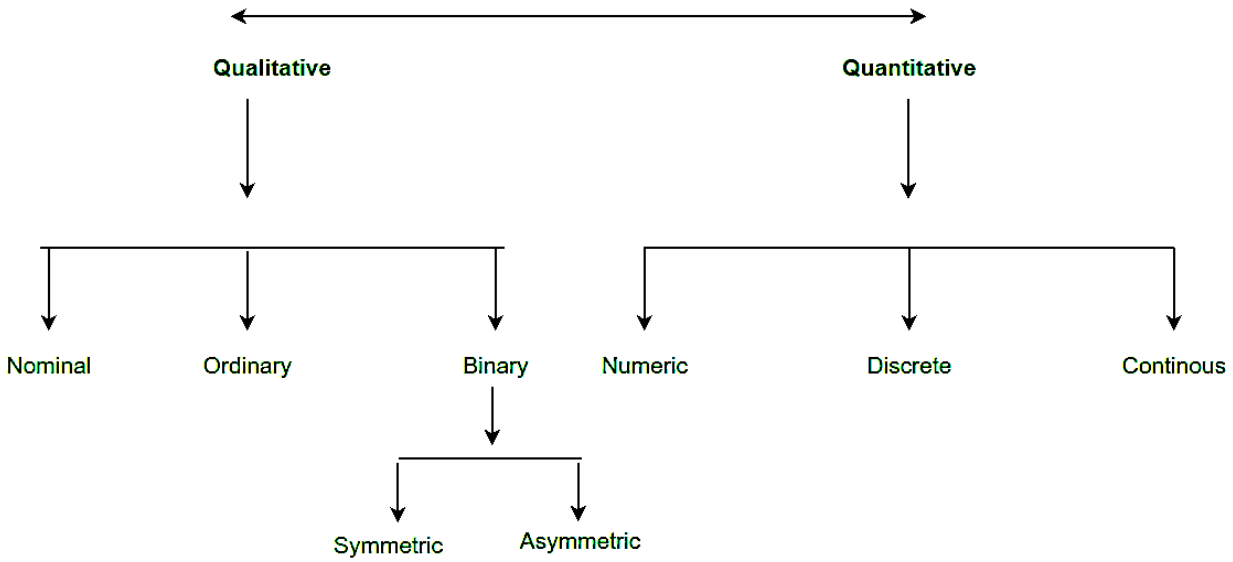
Data Types in Data Analytics

Attribute:

- It can be seen as a data field that represents the characteristics or features of a data object.
- For a customer, object attributes can be customer Id, address, etc.
- It can be seen as a data field that represents the characteristics or features of a data object. For a customer, object attributes can be customer Id, address, etc.

Type of attributes :

- Qualitative (Nominal (N), Ordinal (O), Binary(B)).
- Quantitative (Numeric, Discrete, Continuous)



Qualitative Attributes:

- **Nominal Attributes – related to names:** The values of a Nominal attribute are names of things, some kind of symbols. Values of Nominal attributes represents some category or state and that's why nominal attribute also referred as **categorical attributes** and there is no order (rank, position) among values of the nominal attribute.

Attribute	Values
Colours	Black, Brown, White
Categorical Data	Lecturer, Professor, Assistant Professor

- **Binary Attributes:** Binary data has only 2 values/states. For Example yes or no, affected or unaffected, true or false.
- **Symmetric:** Both values are equally important (Gender).
- **Asymmetric:** Both values are not equally important (Result).

Attribute	Values
Gender	Male , Female

Attribute	Values
Cancer detected	Yes, No
result	Pass , Fail

- Ordinal Attributes : The Ordinal Attributes contains values that have a meaningful sequence or ranking(order) between them, but the magnitude between values is not actually known, the order of values that shows what is important but don’t indicate how important it is.

Attribute	Value
Grade	A,B,C,D,E,F
Basic pay scale	16,17,18

Quantitative Attributes:

- **1. Numeric:** A numeric attribute is quantitative because, it is a measurable quantity, represented in integer or real values. Numerical attributes are of 2 types, **interval**, and **ratio**.
  - An **interval-scaled** attribute has values, whose differences are interpretable, but the numerical attributes do not have the correct reference point, or we can call zero points. Data can be added and subtracted at an interval scale but can not be multiplied or divided. Consider an example of temperature in degrees Centigrade. If a day's temperature of one day is twice of the other day we cannot say that one day is twice as hot as another day.
  - A **ratio-scaled** attribute is a numeric attribute with a fix zero-point. If a measurement is ratio-scaled, we can say of a value as being a multiple (or ratio) of another value. The values are ordered, and we can also compute the difference between values, and the mean, median, mode, Quantile-range, and Five number summary can be given.

- **2. Discrete :** Discrete data have finite values it can be numerical and can also be in categorical form. These attributes has finite or countably infinite set of values.

Attribute	Value
Profession	Teacher, Business man, Peon
ZIP Code	301701, 110040

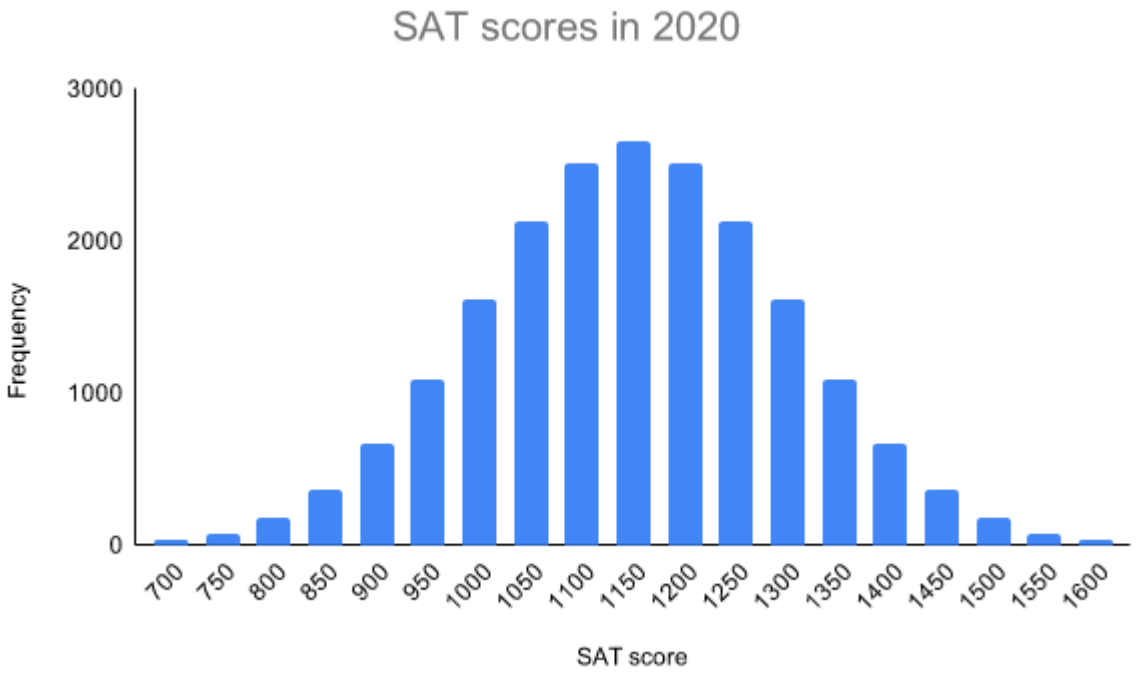
- 3. **Continuous**: Continuous data have an infinite no of states. Continuous data is of float type. There can be many values between 2 and 3.

Attribute	Value
Height	5.4, 6.2 ...etc
weight	50.33 .....etc

Normal Distribution of data

- In a normal distribution, data is symmetrically distributed with no skew. When plotted on a graph, the data follows a bell shape, with most values clustering around a central region and tapering off as they go further away from the center.

- Normal distributions are also called Gaussian distributions or bell curves because of their shape.


SAT scores in 2020
Frequency
3000
2000
1000
0
700
750
800
850
900
950
1000
1050
1100
1150
1200
1250
1300
1350
1400
1450
1500
1550
1600
SAT score

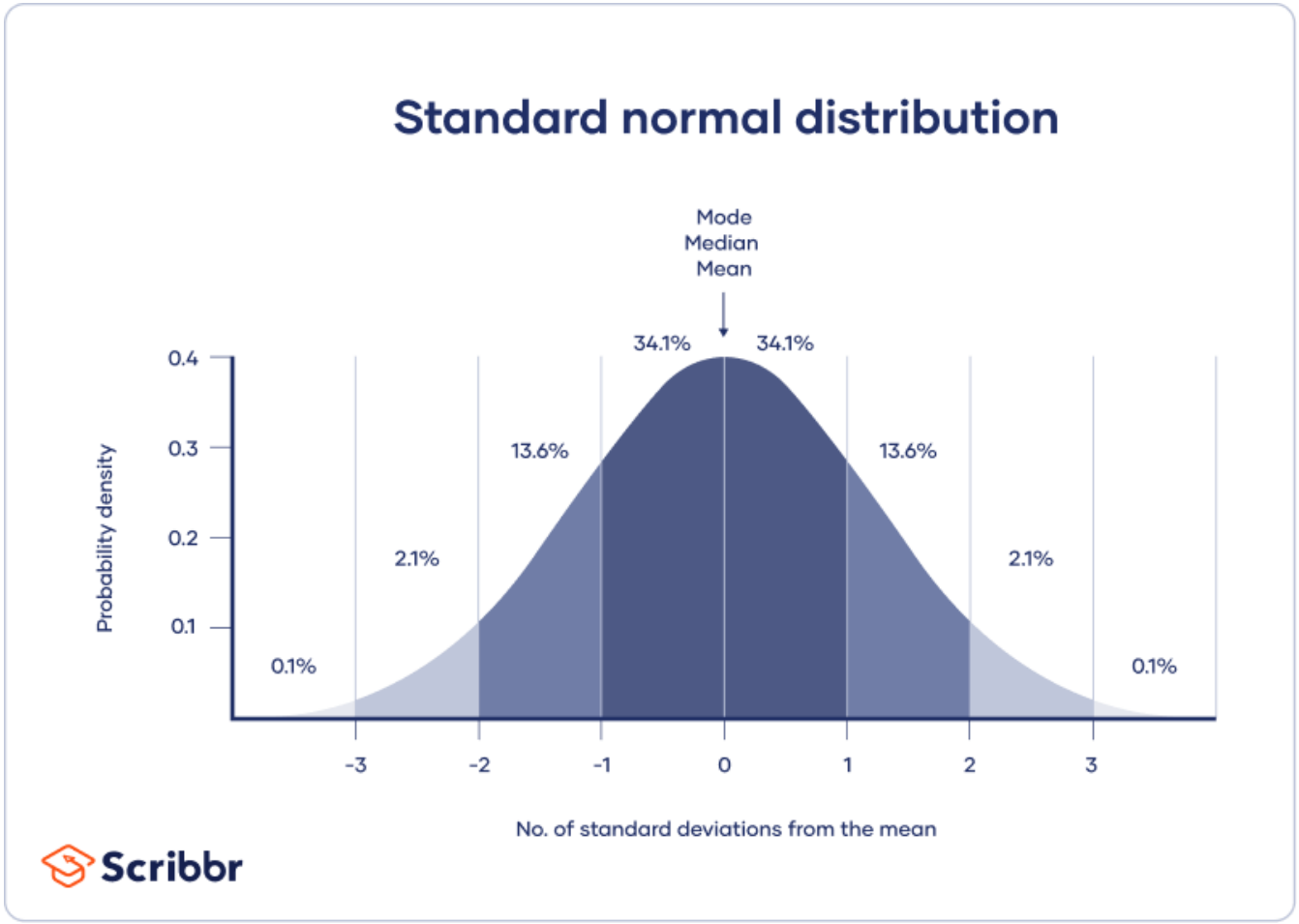
- All kinds of variables in natural and social sciences are normally or approximately normally distributed. Height, birth weight, reading ability, job satisfaction, or SAT scores are just a few examples of such variables.

- Because normally distributed variables are so common, many statistical tests are designed for normally distributed populations.

- Understanding the properties of normal distributions means you can use inferential statistics to compare different groups and make estimates about populations using samples.

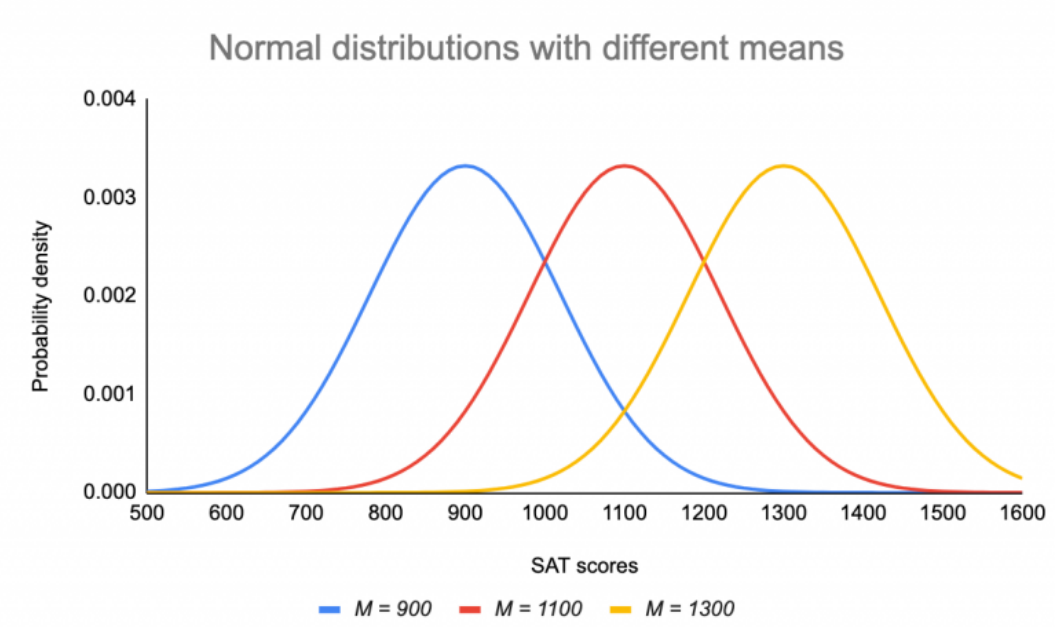
The properties of normal distributions

- Normal distributions have key characteristics that are easy to spot in graphs:
- The mean, median and mode are the same.
- The distribution is symmetric about the mean—half the values fall below the mean and half above the mean.
- The distribution can be described by two values: the mean and the standard deviation.

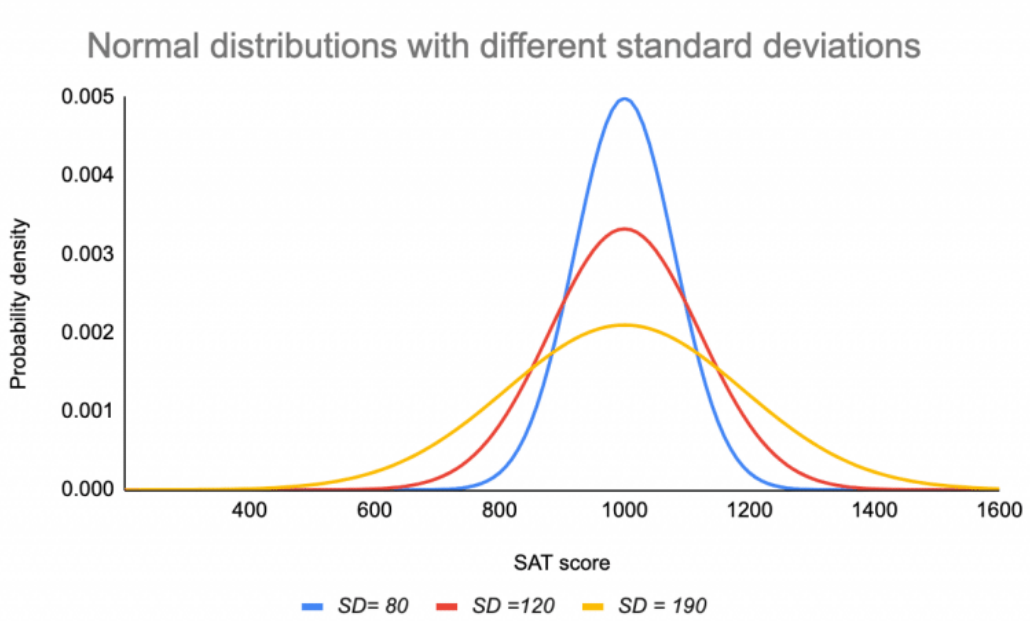

Standard normal distribution
Mode
Median
Mean
34.1%
34.1%
13.6%
13.6%
2.1%
2.1%
0.1%
0.1%
0.4
0.3
0.2
0.1
Probability density
-3
-2
-1
0
1
2
3
No. of standard deviations from the mean
Scribbr

- The mean is the location parameter while the standard deviation is the scale parameter.

- The mean determines where the peak of the curve is centered. Increasing the mean moves the curve right, while decreasing it moves the curve left.



- The standard deviation stretches or squeezes the curve. A small standard deviation results in a narrow curve, while a large standard deviation leads to a wide curve.



Question ?