

What's Data? Why is it important?

Data can come in many forms:-

- Text
- Video
- Spread Sheets
- Databases
- Images
- Audio



And There are many other forms of Data.

Utilizing the Data is so important in many fields:-

- Insurance
- Banking
- Medicine
- Education
- Agriculture
- Automotive
- Manufacturing

**You can utilize data to make better decisions
and accomplish your goals.**

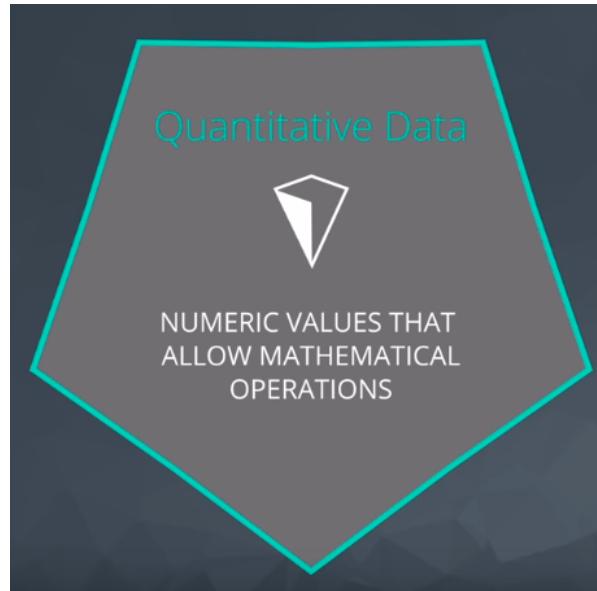


Data Types (Quantitative vs. Categorical)

Quantitative	Categorical
NUMBER OF DOGS	BREED OF DOGS
0	Lab
1	Pug
2	Poodle

Quantitative Data:-

- ### Numeric values that allow mathematical operations.



Example:

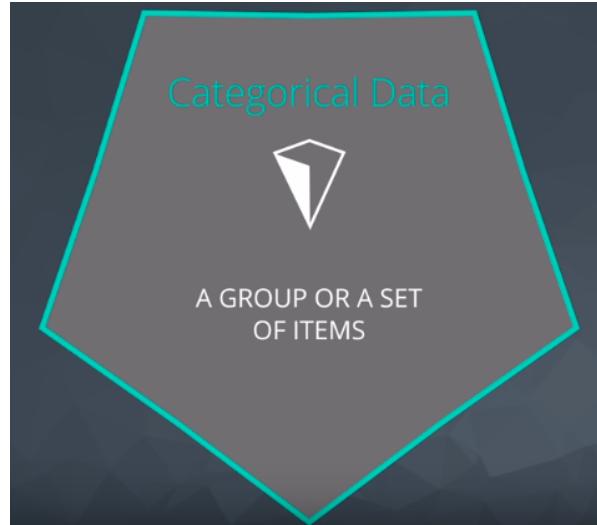
- #### If I saw 5 dogs on manday and 6 dogs on tuesday that mean, I saw 11 dogs this week.

MATH CAN BE DONE
WITH QUANTITATIVE DATA

$$\begin{array}{r} 5 \text{ DOGS} \\ + \\ 6 \text{ DOGS} \\ \hline 11 \text{ DOGS} \end{array}$$

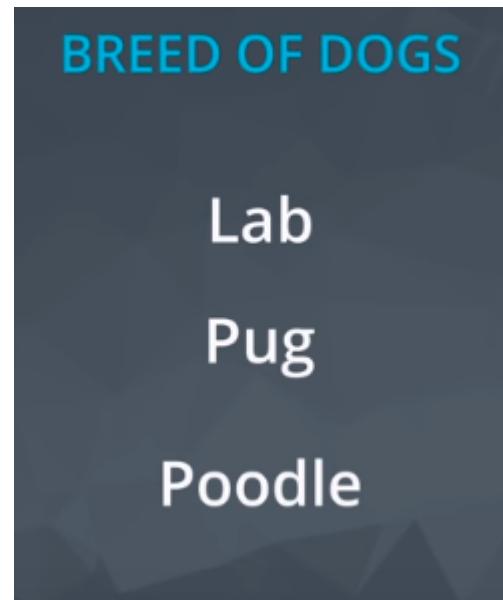
Categorical Data:-

- ### A Group or set of items.

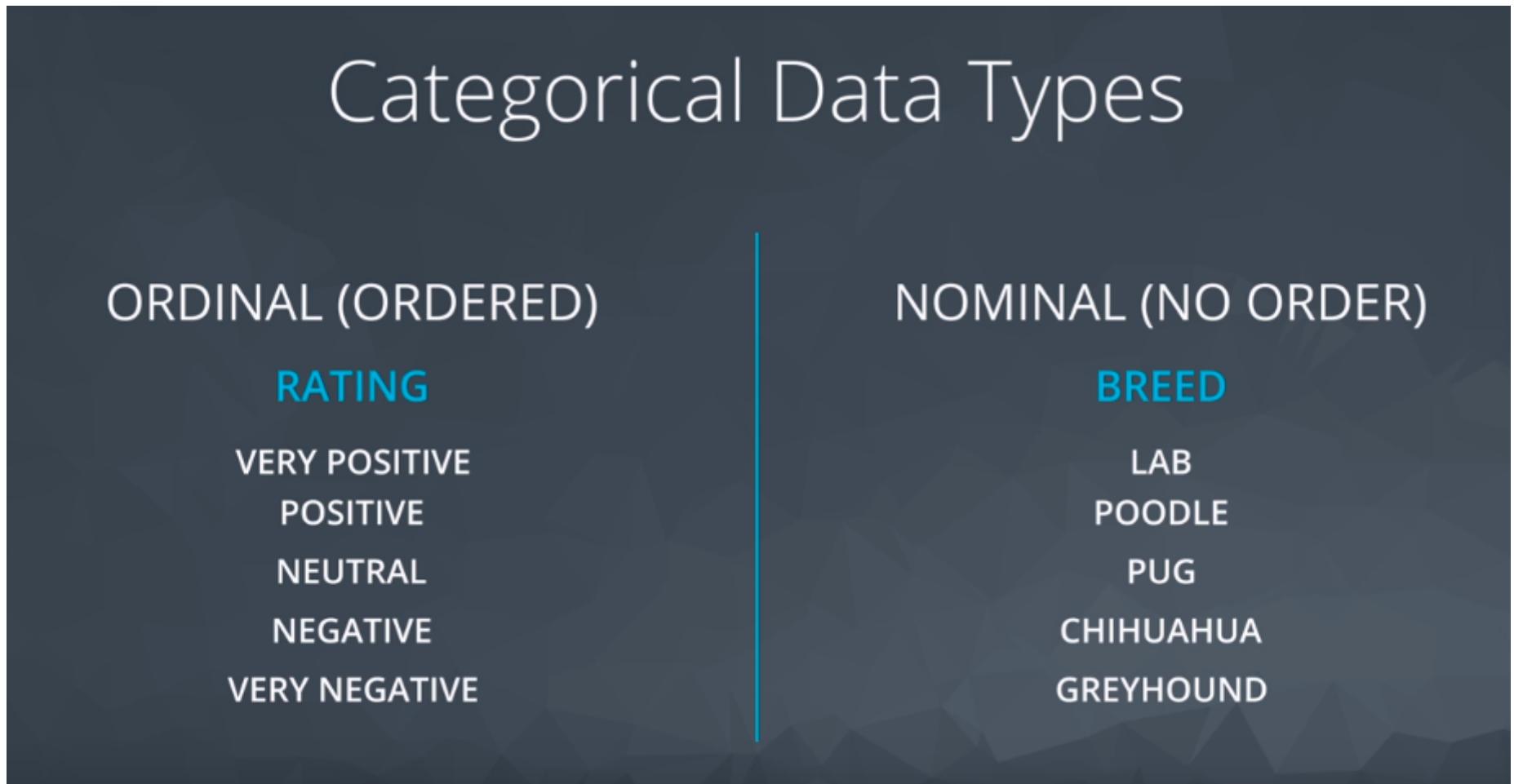


Example:

- ##### Breed of Dogs.



Data Types (Ordinal vs. Nominal)



Categorical Ordinal Data:

- *categorical values that are ranked.*



Categorical Nominal Data:

- *categorical values that don't have ranked order*

Categorical Nominal Data

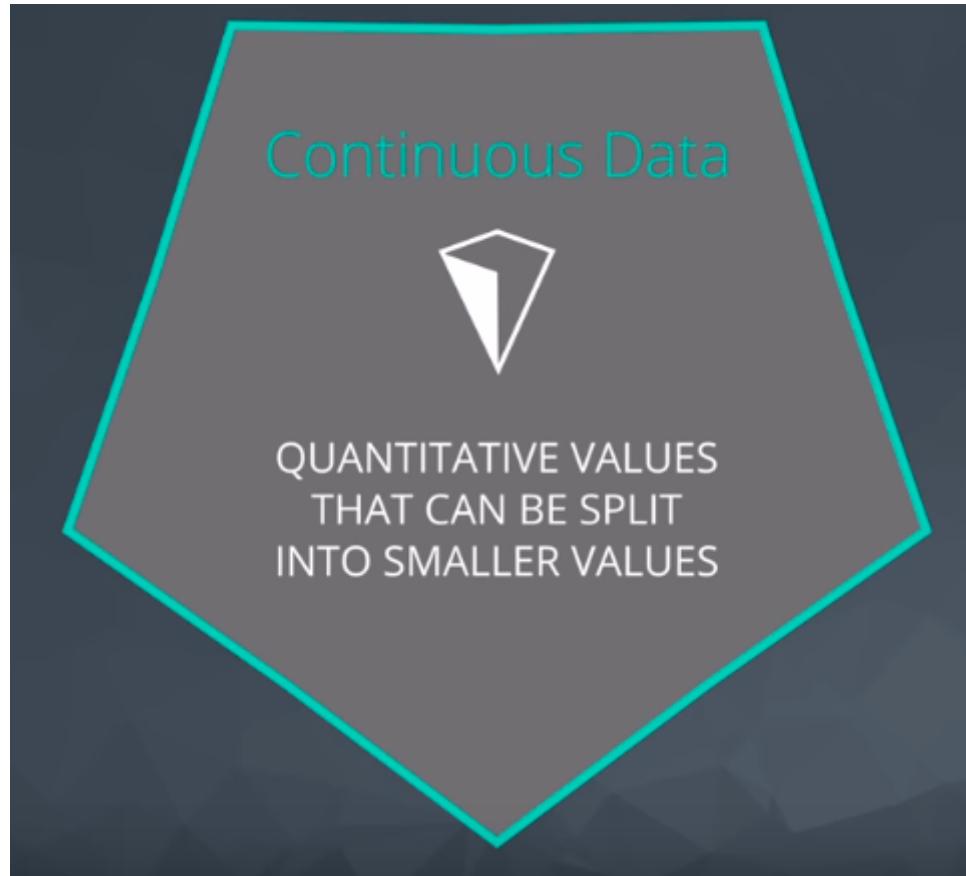


CATEGORICAL VALUES
THAT DO NOT HAVE
RANKED ORDER

Data Types (Continuous vs. Discrete)

Continuous Data:

- *Quantitive values that can be split into smaller values.*



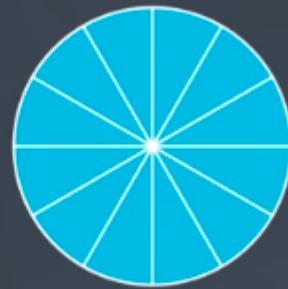
- *For Example:*

Continuous Data

YEAR



MONTHS



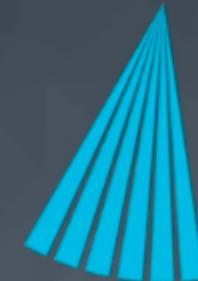
DAYS



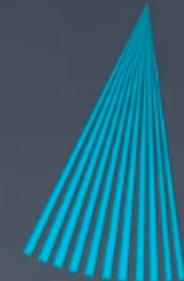
HOURS



MINUTES



SECONDS



Discrete Data:

- *Quantitive values that are countable*

Discrete Data



QUANTITATIVE VALUES
THAT ARE COUNTABLE

- **For Example:** *The number of dogs we interact with*



Data Types Summary



QUANTITATIVE

▽ DISCRETE | 0,1,2,...

▽ CONTINUOUS | 0.1,0.2,1.5,...



CATEGORICAL

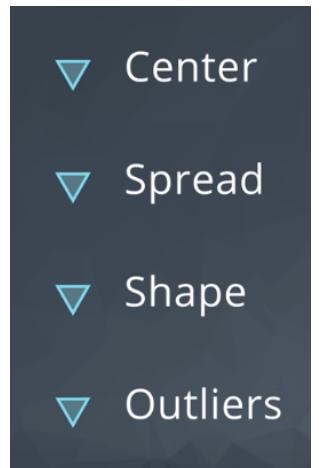
▽ ORDINAL | High, Med, Low

▽ NOMINAL | Lab, Pug, Poodle ...

Measures of Center (Mean)

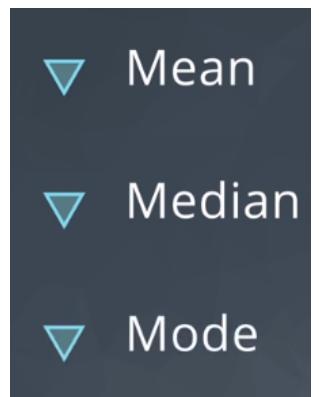
We need 4 main aspects when analyzing both discrete and quantitative data:-

- #### Center
- #### Spread
- #### Shape
- #### Outliers



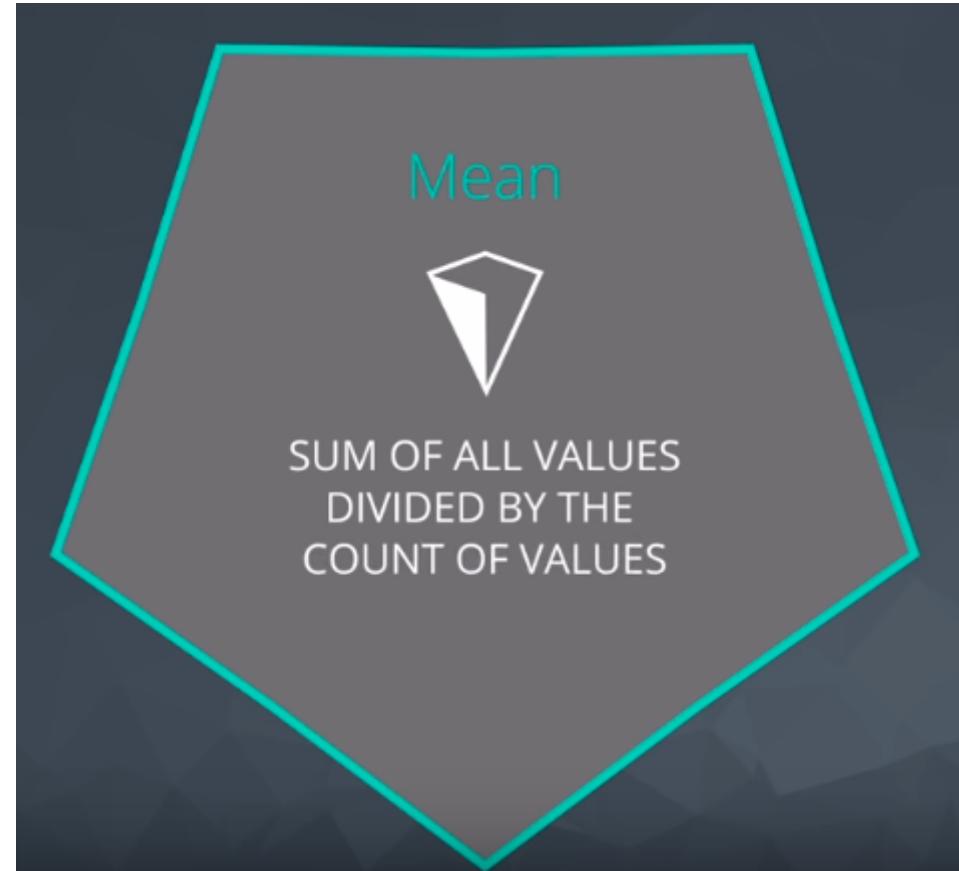
Measures of Center:-

- #### Mean
- #### Median
- #### Mode



Mean:

- Sum of all values divided by the count of values.*



For Example: Data Of the dogs were seen in the coffe shop through the week:-

MON	TUE	WED	THU	FRI	SAT	SUN
5	3	8	3	15	45	9

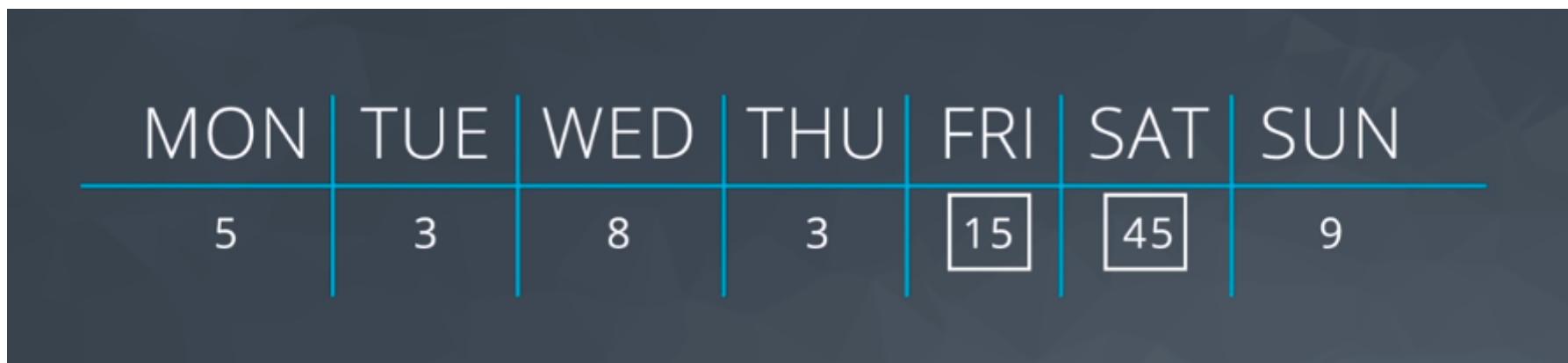
Mean

$$5 + 3 + 8 + 3 + 15 + 48 + 9$$

7

$$= 12.57 \text{ dogs}$$

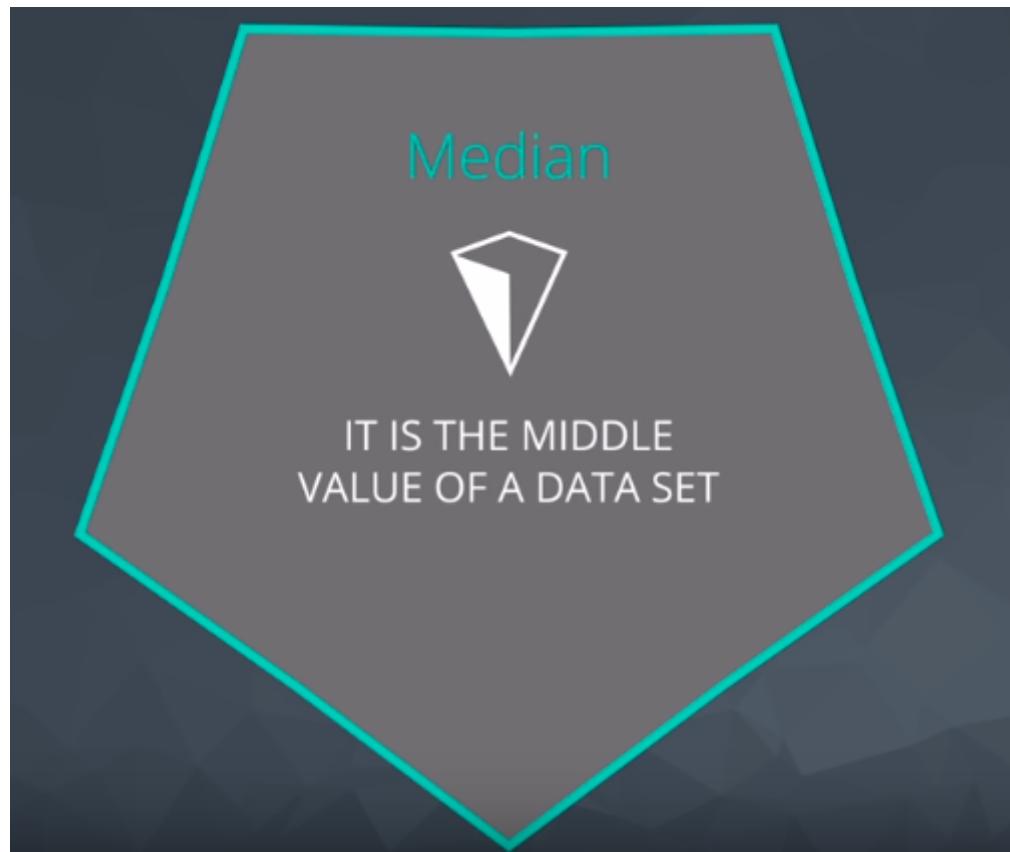
- But it doesn't seem that the mean isn't in the middle of the data and it also is splitting our dogs into decimal values which will seem strange when we reporting it back.



Measures of Center (Median)

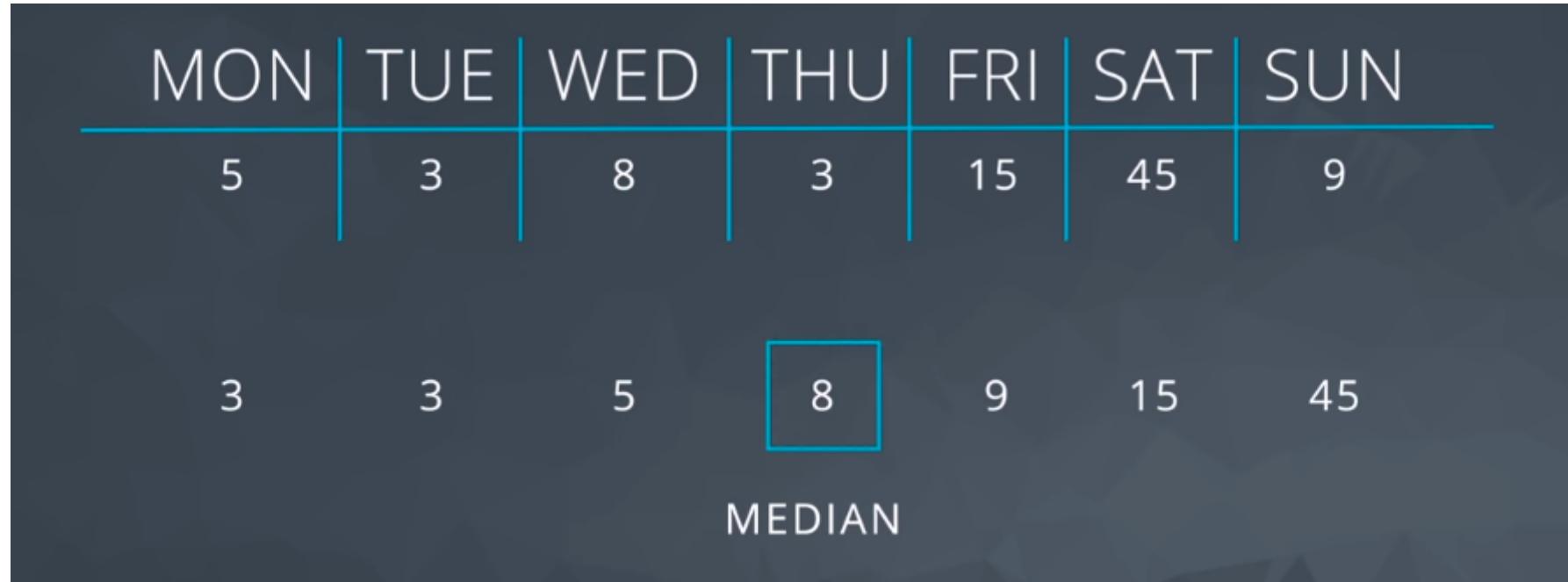
Median:-

- *It's the middle value of a dataset*



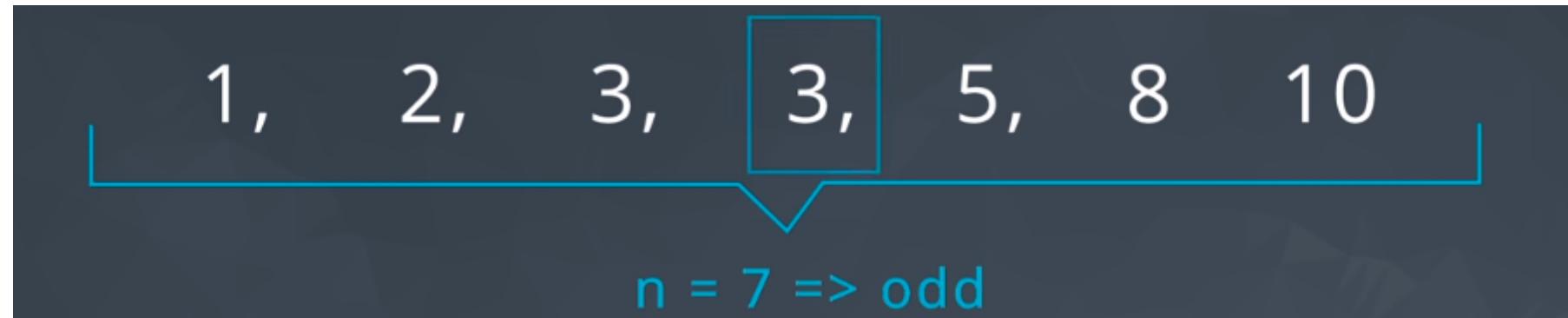
For Example:-

- *The median of the data we have which seem more reasonable more than the mean.*



Examples of Median Calculation:-

- *when the count of numbers is an odd number.*



- *when the count of the number is an even number.*

5, 8, 3, 2, 1, 3, 10, 105

$n = 8 \Rightarrow \text{Even}$

- put the values in order.

1, 2, 3, 3, 5, 8, 10, 105

- get the 2 numbers in the middle of the data.

1, 2, 3, 3, 5, 8, 10, 105

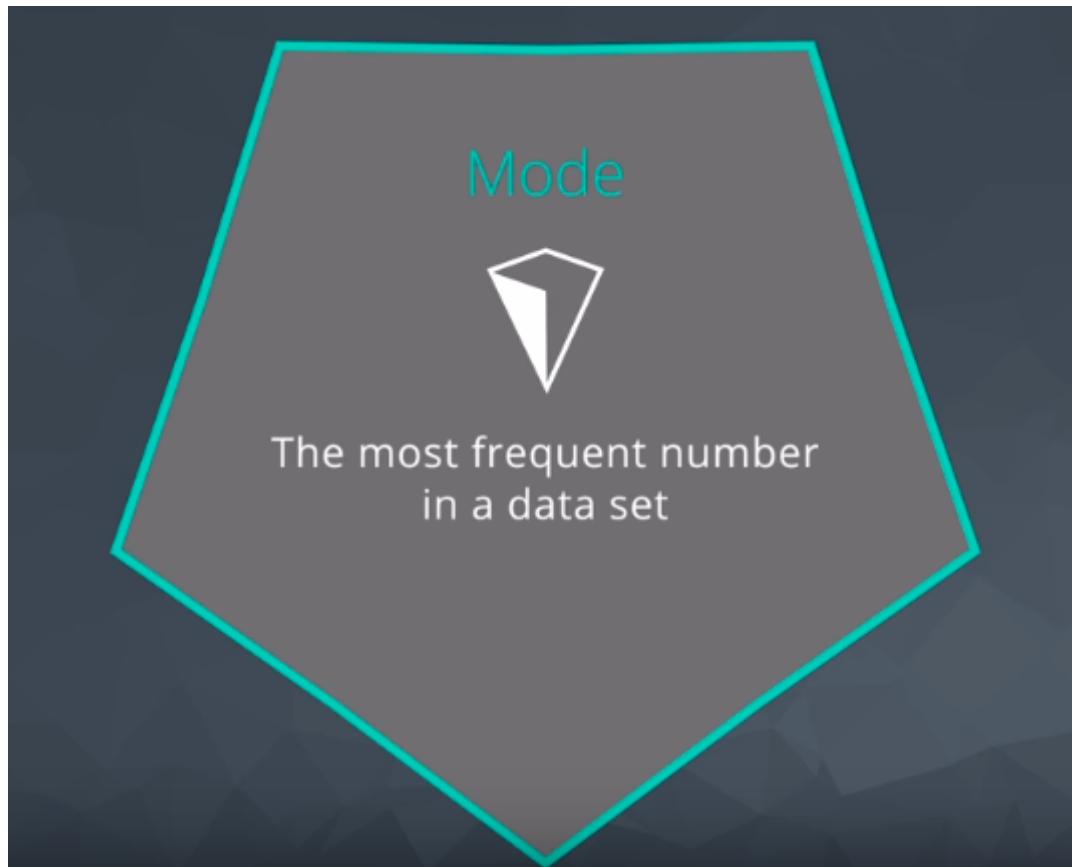
3, 5,

$$\frac{3 + 5}{2} = 4$$

Measures of Center (Mode)

Mode:-

- *the most frequent number in the data set.*



For Example:-

1,

2,

3,

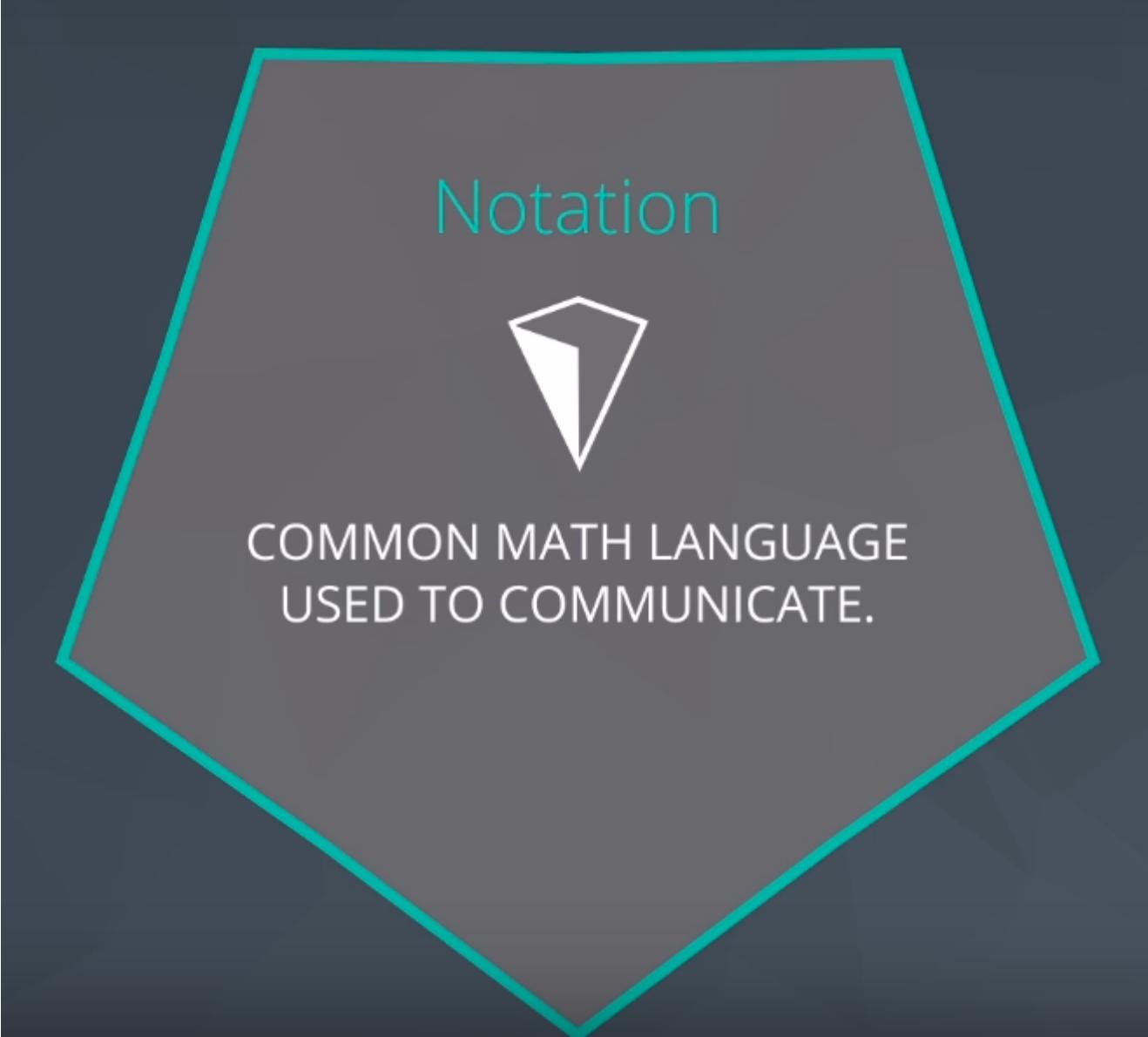
3,

5,

8

10

What is Notation?



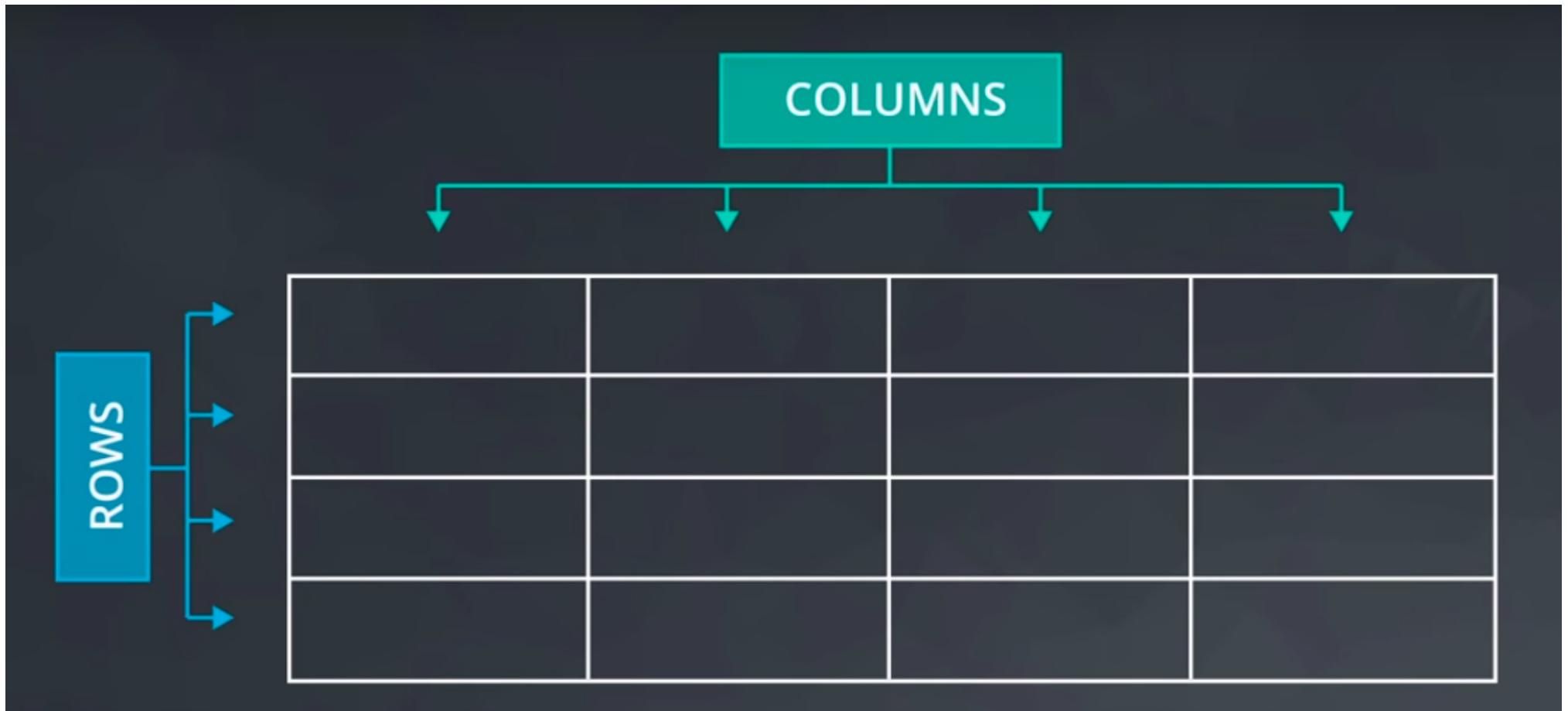
Notation



COMMON MATH LANGUAGE
USED TO COMMUNICATE.

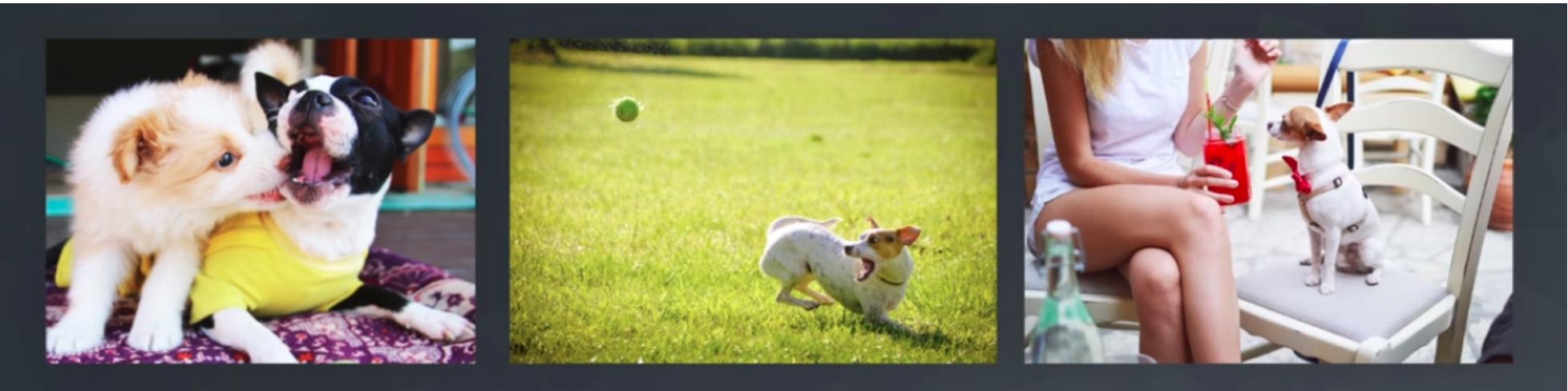
Random Variables

The best way to introduce Notation we will use spreadsheets



Example:-

- ### Consider a websites for dogs'needs



- ### So We need to answer some questions to make a better decisions about our website*

QUESTIONS

- ▽ How many people visit the site?
- ▽ How much time do visitors spend on the site?
- ▽ Are there differences in traffic depending on the day of the week?
- ▽ How many visitors purchase an item through the blog?

- ### In order to answer these question we need to keep track of (*Date of the visit, Day of the week of the visit, Amount of time spend on the site, Whether or not an individual buys an item*)

DATE	DOW	TIME	BUY

- *### Each column of them assossiated with random variable*

Random Variable:



PLACEHOLDER FOR THE POSSIBLE
VALUES OF SOME PROCESS
NOTATION = X

- ### With this spreadsheet we can answer our questions

How much time do visitors spend on the site?

DATE	DOW	TIME	BUY
Jun 15	Thur	5	No
Jun 15	Thur	10	Yes
Jun 16	Fri	7	Yes
Jun 16	Fri	9	Yes
Jun 16	Fri	12	No

Are there differences in traffic depending on the day of the week?

DATE	DOW	TIME	BUY
Jun 15	Thur	5	No
Jun 15	Thur	10	Yes
Jun 16	Fri	7	Yes
Jun 16	Fri	9	Yes
Jun 16	Fri	12	No

How many purchases occur through our blog?

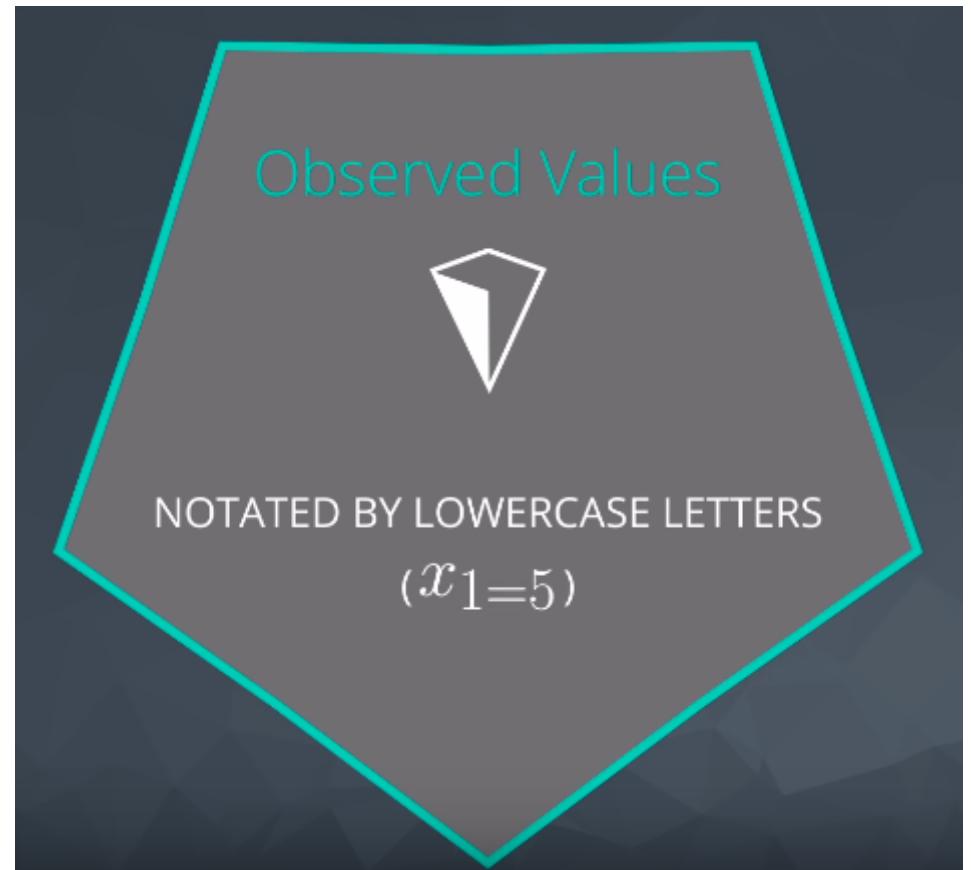
DATE	DOW	TIME	BUY
Jun 15	Thur	5	No
Jun 15	Thur	10	Yes
Jun 16	Fri	7	Yes
Jun 16	Fri	9	Yes
Jun 16	Fri	12	No

- ### Mathematically we consider random variable or column using a capital letter, so consider (X related to "Time" column and Y related to "Buy" column)

X	Y
TIME	BUY
5	No
10	Yes
7	Yes
9	Yes
12	No

Capital vs. Lower

Random variables & Its subset:-



For Example:-

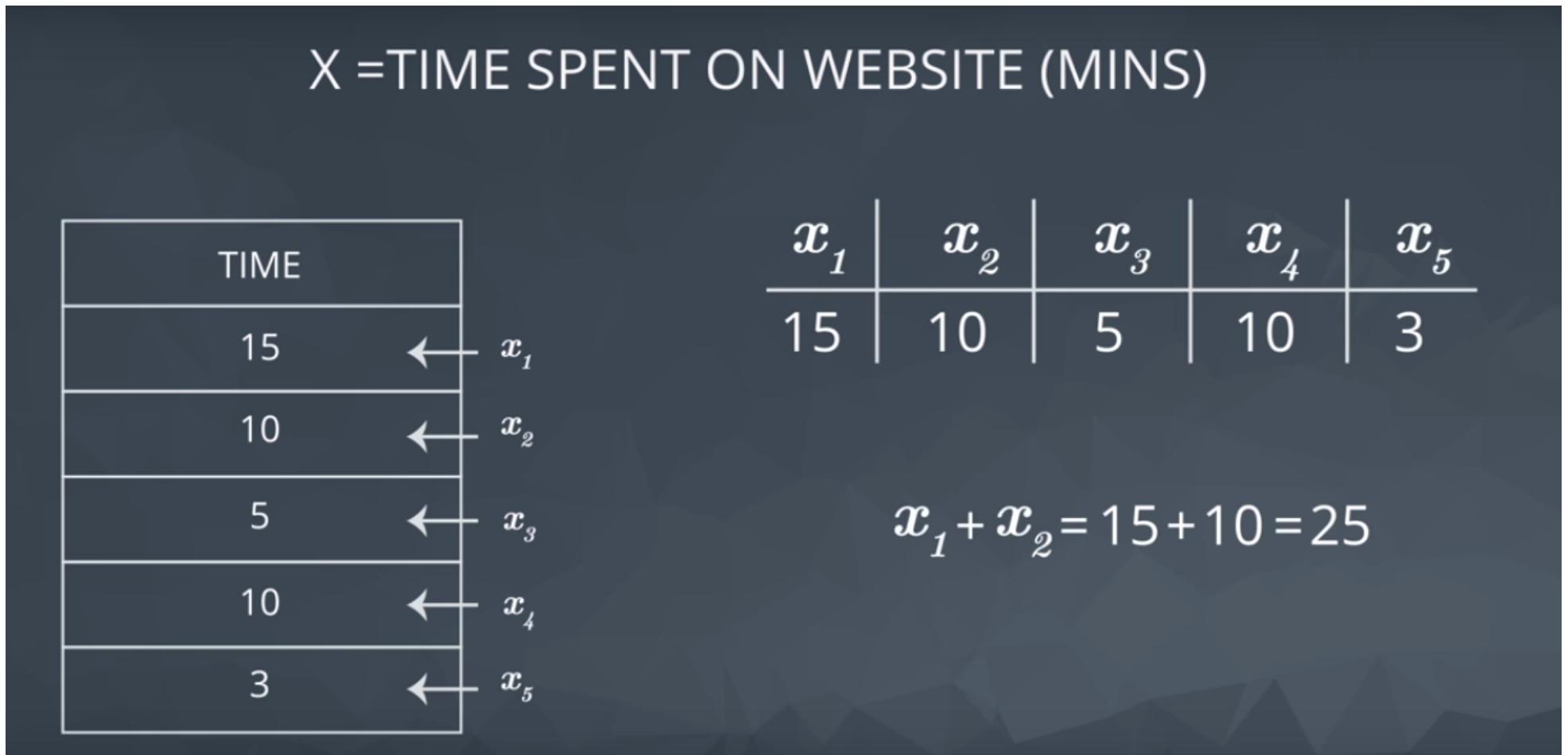
X

DATE	DOW	TIME	BUY
		5 ←	x_1
		10 ←	x_2
		...	
		...	
		...	
		n ←	x_n

How to calculate the mean with Notation

For Example:-

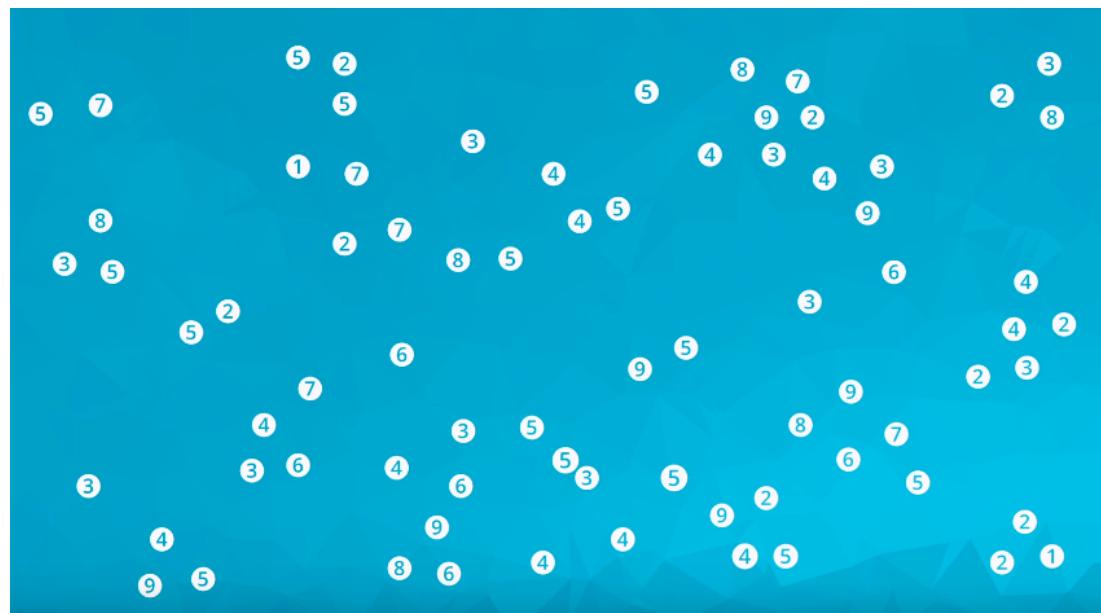
- *### If we want to sum the first 5 elements of "Time" column*



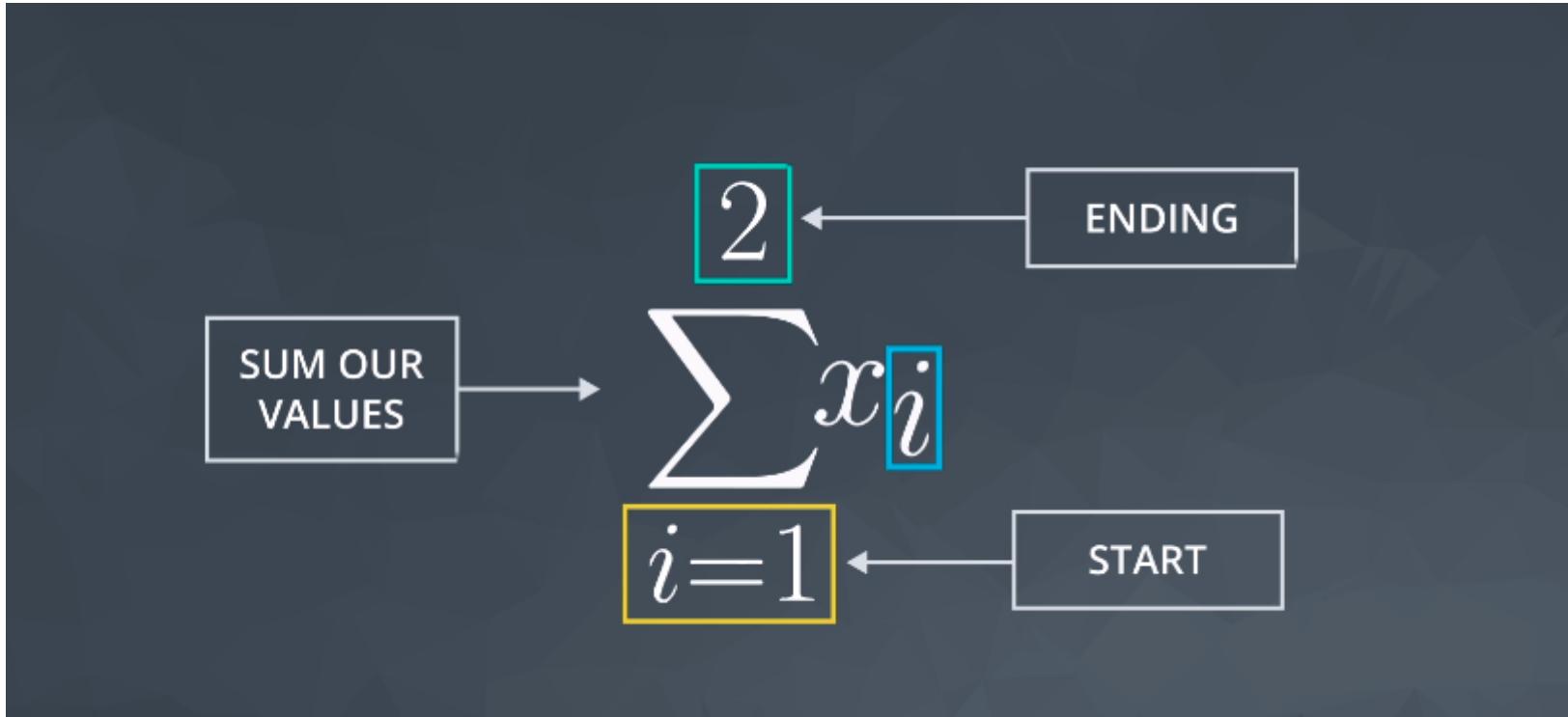
- *### What if we want to sum 100 elements of the "Time" column*

$$x_1 + x_2 + x_3 + \dots + x_{100}$$

we will use the concept of Aggregation to make simple to represent multiple numbers into fewer numbers



Using Upper case Sigma \sum to represent those numbers :-



For Example:-

$$\begin{aligned}\sum_{i=1}^2 x_i &= x_1 + x_2 \\ &= 15 + 10 = 25\end{aligned}$$

Other Aggregations

The \sum sign is used for aggregating using summation, but we might choose to aggregate in other ways. Summing is one of the most common ways to need to aggregate. However, we might need to aggregate in alternative ways. If we wanted to multiply all of our values together we would use a product sign \prod , capital Greek letter pi. The way we aggregate continuous values is with something known as integration (a common technique in calculus), which uses the following symbol \int which is just a long s. We will not be using integrals or products for quizzes in this class, but you may see them in the future!

$$Mean : \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

Summary

Notation Recap

Notation is an essential tool for communicating mathematical ideas. We have introduced the fundamentals of notation in this lesson that will allow you to read, write, and communicate with others using your new skills!

Notation and Random Variables

As a quick recap, **capital letters** signify **random variables**. When we look at **individual instances** of a particular random variable, we identify these as **lowercase letters** with subscripts attached to each specific observation.

For example, we might have **X** be the amount of time an individual spends on our website. Our first visitor arrives and spends 10 minutes on our website, and we would say **x₁** is 10 minutes.

We might imagine the random variables as columns in our dataset, while a particular value would be notated with the lower case letters.

Notation	English	Example
X	A random variable	Time spent on website
x ₁	First observed value of the random variable X	15 mins
$\sum_{i=1}^n x_i$	Sum values beginning at the first observation and ending at the last	5 + 2 + ... + 3
$\frac{1}{n} \sum_{i=1}^n x_i$	Sum values beginning at the first observation and ending at the last and divide by the number of observations (the mean)	(5 + 2 + 3)/3
\bar{x}	Exactly the same as the above - the mean of our data.	(5 + 2 + 3)/3

Notation for the Mean

We took our notation even farther by introducing the notation for summation \sum . Using this we were able to calculate the mean as:

$$\frac{1}{n} \sum_{i=1}^n x_i$$

In the next section, you will see this notation used to assist in your understanding of calculating various measures of spread. Notation can take time to fully grasp. Understanding notation not only helps in conveying mathematical ideas, but also in writing computer programs - if you decide you want to learn that too! Soon you will analyze data using spreadsheets. When that happens, many of these operations will be hidden by the functions you will be using. But until we get to spreadsheets, it is important to understand how mathematical ideas are commonly communicated. **This isn't easy, but you can do it!**