Introduction to Data Analytics

Data Analytics - General Meaning

In layman terminology, Data Analysis is simply understanding the given data to get proper insights (useful information).

- **Data** - Facts and Figures
- **Information**
  - Data which is processed is called information.
  - Information is thus understood better.

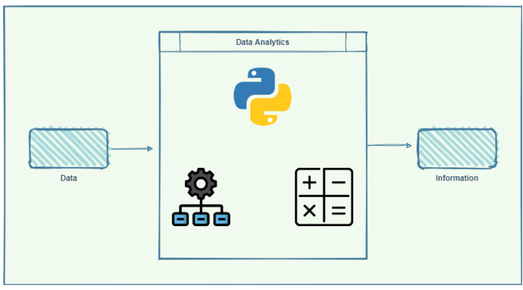

Image by author

How do we get the data?

First identify in which domain you want to collect the data.

For example →

- Study to analyse peoples' habits on YouTube platform
- Study to analyse the changes occurred in peoples' life due to Demonotization
- Study to analyse the students' overall development due to online education

We can collect data by the following methods

```
- Data Collection → Collection of data through
                    1. Person to Person Survey
                    2. Online Questionnaire (Google Forms)
                    3. Online Tracking
                    4. Many more
```

How do we understand the data?

Through

```
- Inspection       → Careful Observation
- Cleansing        → Data Cleaning
- Transforming     → Change the structure of the data
- Modelling        → Mathematical and Statistical study
```

For implementing the above methods we need to have data.

Let's understand in a broader perspective

Imagine you are a class teacher of a certain class and you want to make sure your students do participate in the `Sports Meet` coming in the near future. To ensure that you will need the data of the students like height and weight.

Data Collection

Data variables that are most commonly required are -

- Student Roll number/Name
- Student Height
- Student Weight

Since you are the class teacher, data is obtained with less effort.

Roll no	Height (ft)	Weight (kg)
1	5	40
2	4.8	45
3	5.3	48
4	6	52
5	5.5	53
...	...	...
60	5.8	50

Case Study

If you, as a teacher give this collected data to the principal, does the principal be able to understand how many are going to take part in the Sports Meet and also which game they are more suitable for?

Solution

Since the principal is the one who decides the future of your employment, you will take measures to get the proper information from the data.

Like -

First Stage

- Average Height → `5.6 ft`
- Average Weight → `70 kgs`
- Min and Max values

Second Stage

- Correlation - Relationship between height and weight → `89%`
  - Studying or understanding how one parameter/variable affect another variable
  - If the height of person is more, then what is the weight (more or less)

Third Stage

- Interpretation
  - If the height is more, the person is more suitable for long jump or high jump
  - If the weight is more, the person is more suitable for weight lifting
  - If both height and weight is more, the person is suitable for both long jump and weight lifting

Conclusion

Based on the above results the institue can be able to decide what to do further.

- Either proceed to take part in the Sports Meet or not.
  - If yes, then which game there needs more focus.

How was this possible?

It was all possible because of the availability of certain facts (data) which got processed into information and with certain statistical measure we were able to decide what to do (whether to take part or not).

$$D_m = f(A, B, C, D)$$

where

- A → Adequate data
- B → Relevant data
- C → Reliable data
- D → Timely data

and

- f → function of data processing
- D_m → Decision making

In Reality

As a Data Analyst your role would be something like the above example

- Collecting the right data
- Converting the right data into information
- Applying statistical methods for decision making
  - Interpreting the results
  - predicting the future results
  - Concluding

Reference - https://bit.ly/3ufPxn7

Activity → Homework

- Take any one of these domains below or if you want take your own domain, (You are good to go) -
  - Study to analyse peoples' habits on YouTube platform
  - Study to analyse the changes occurred in peoples' life due to Demonotization
  - Study to analyse the students' overall development due to online education
- You are required to come up with a list of questions (**Maximum questions 5**) from which you can get the data that is needed.
- After forming the questions try to identify the variables.

For example → Topic - `Study to analyse peoples' habits on YouTube platform`

Questions

```
1. How much time do you spend on youtube daily?
    a) 1 hour       b) 2 hours
    c) 3 hours      d) NIL
2. What kind of videos do you watch?
    a) Education    b) Fun
    c) Movies       d) Random videos
3.
4.
5.
```

Variables

```
1. spending_time
2. video_type
3.
4.
5.
```