

Session 1

21 May 2021 17:09

Agenda

- 1) Basics
- 2) What is Data science
- 3) How python is effective
- 4) Hands on installation and open source materials for practice

Organize:

- 1) ETL - Extract Transform Load
- 2) Data Pre -Processing

Data - Data Extract	Predominantly done by 'Data Engineers'
Data Extract - Data Organize - EDA	Data Analytcs specialist
Organise to Analyse(Model Building and Implementation)	Data Scientist

Data Engineers - SQL an they should be able to extract the date from multiples sources

If we need 15 columns:

- 1) 1st 10 columns are available in the database
 - a. We will write queries to extract the data
- 2) 2 columns are available in the HTML
 - a. Need to do web scrap
- 3) 3 columns text (Eg colour of a car customer uses frequently)
 - a. We need to derive that using

Active customers

Descriptive analysis

90% of the people whio buy bread will buy jam

predictive analysis

What if a person tomorrow buys bread?

To practice python:

Possible install anaconda(Personal laptop)

Google colab