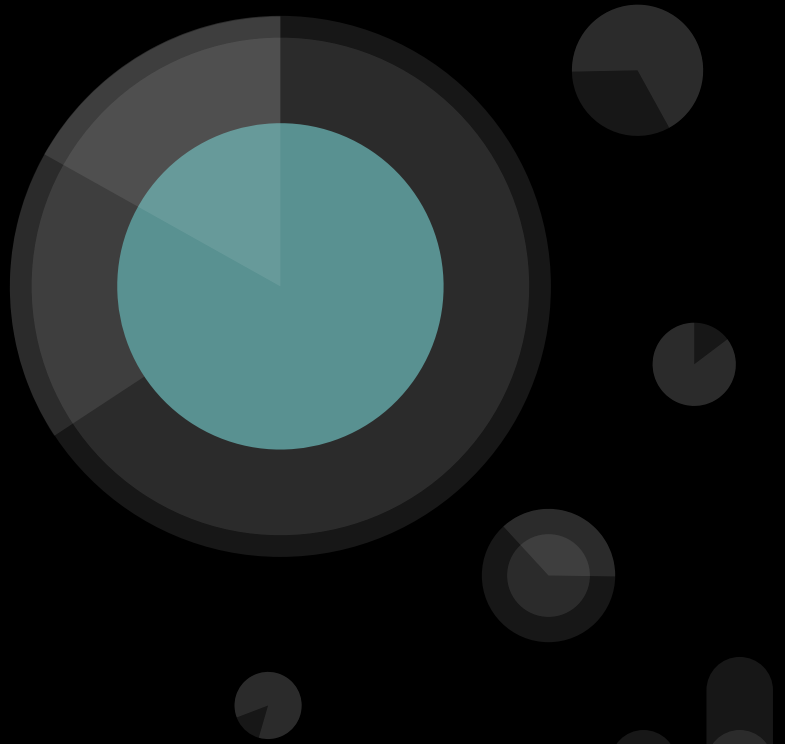
Data Science and the Hype around it

An introduction to NumPy, Pandas and Matplotlib



A Session by ACM Student Chapter IITH

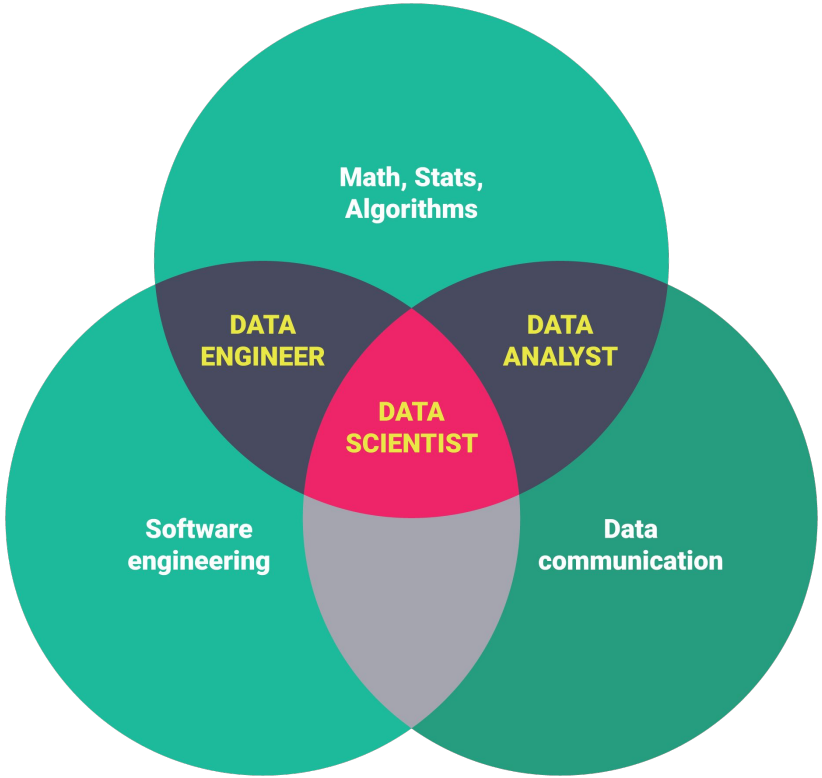
Presented by: Tanmay Goyal

DATA

1. There are approximately 400,000 bytes of data for every grain of sand on earth.

BIG
DATA
USE
ONE
SINCE
SOFTWARE
WORKING
USING
TYPES
GARTNER
MASSIVELY
REQUIRING
UBIQUITOUS
RADIO-FREQUENCY
COMPLEXITY
SAN
SOLID
WIRELESS
BUSINESS
MOVING
ANALYTICS
PARALLEL
TOLERABLE
PERFORMANCE
RELATED
ALSO
BIOLOGICAL
HUNDREDS
CREATED
SIZE
NEEDED
QUALITIES
BURIED
ORGANIZATIONS
DESKTOP
DISK
RELATIONAL
ZETTABYTES
GENOMICS
COMPUTING
TOOLS
SET
SHARED
TIME
PETABYTES
PROCESSING
LOGS
CURRENTLY
TENS
INTERNET
WITHIN
COMBAT
SIGNIFICANT
PROCESS
INCLUDE
SYSTEMS
CAPACITY
PRACTITIONERS
TECHNOLOGIES
DEFINITION
BIOGEOCHEMICAL
NETWORKS
FC
FORMS
USED
DISTRIBUTED
CAPTURE
RECONSIDER
SEARCH
INFORMATION
OPPORTUNITIES
MAY
MANAGEMENT
MANAGE
NOW
PRESENTATIONS
CONNECTOMICS
RECORDS
WORLD'S
SOCIAL
LARGE
DEFINING
STORE
GROW
COST
CONTINUES
TERABYTES
MPP
RESEARCH
STORAGE
EVERY
CASE
CITATION
COMPLEX
DATABASES
LARGER
INDEXING
DIFFICULTY
TARGET
ABILITY
SETS
EXAMPLES
ELAPSED
CURRENT
THOUGHT
APPLIED
SENSOR
ARCHIVES
AMOUNT
DESCRIBING

So is it just Data Science?



Requirements

- Technical Skills varying in the following:
  - Excel
  - SQL
  - Visualisation Tools such as Tableau, PowerBI
  - Some notion of coding -> R, Python, etc
  - Cloud Computing tools such as AWS/Azure
  - Apache Spark
  - Machine Learning

- What about mathematical or statistical skills?
- Remember, lot of overlap between the roles. Skills required as per demand!

Is that it though?

- Inquisitiveness
- Communication Skills
- Presentation Skills
- Report Writing and Compilation and Summarization Skills
- A good eye!