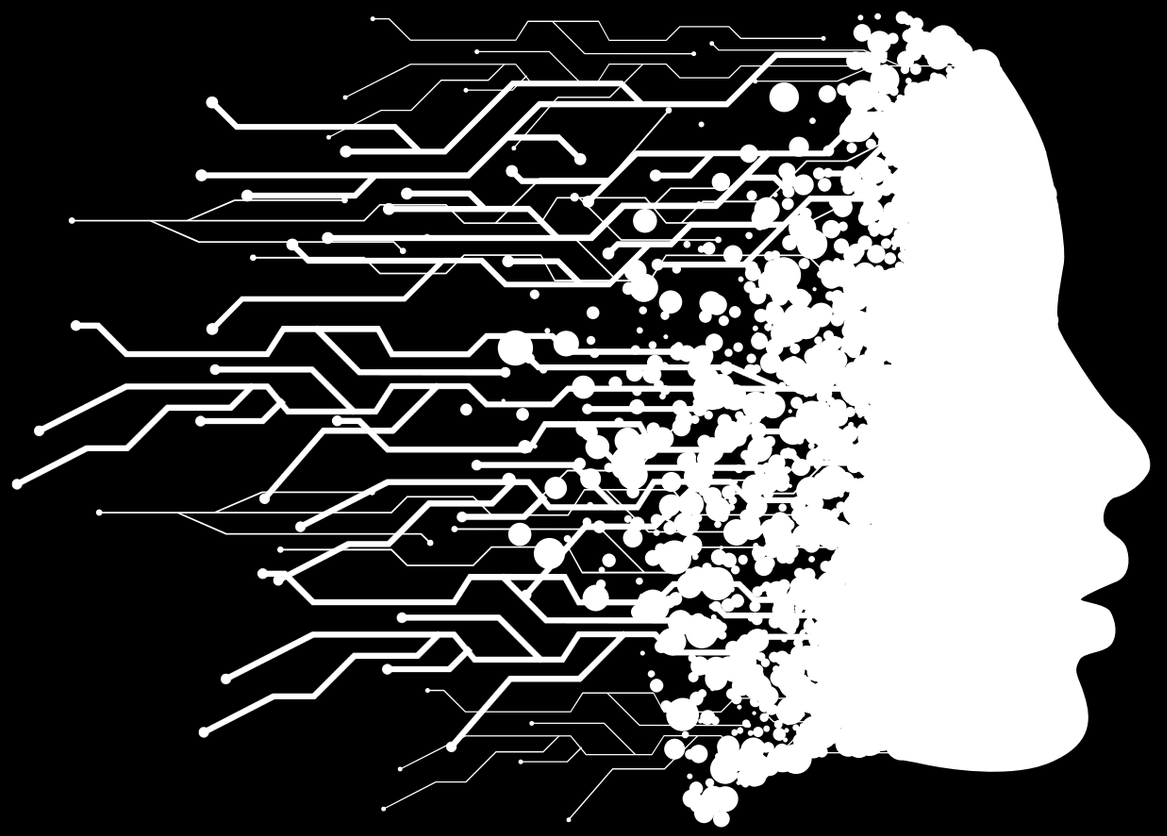
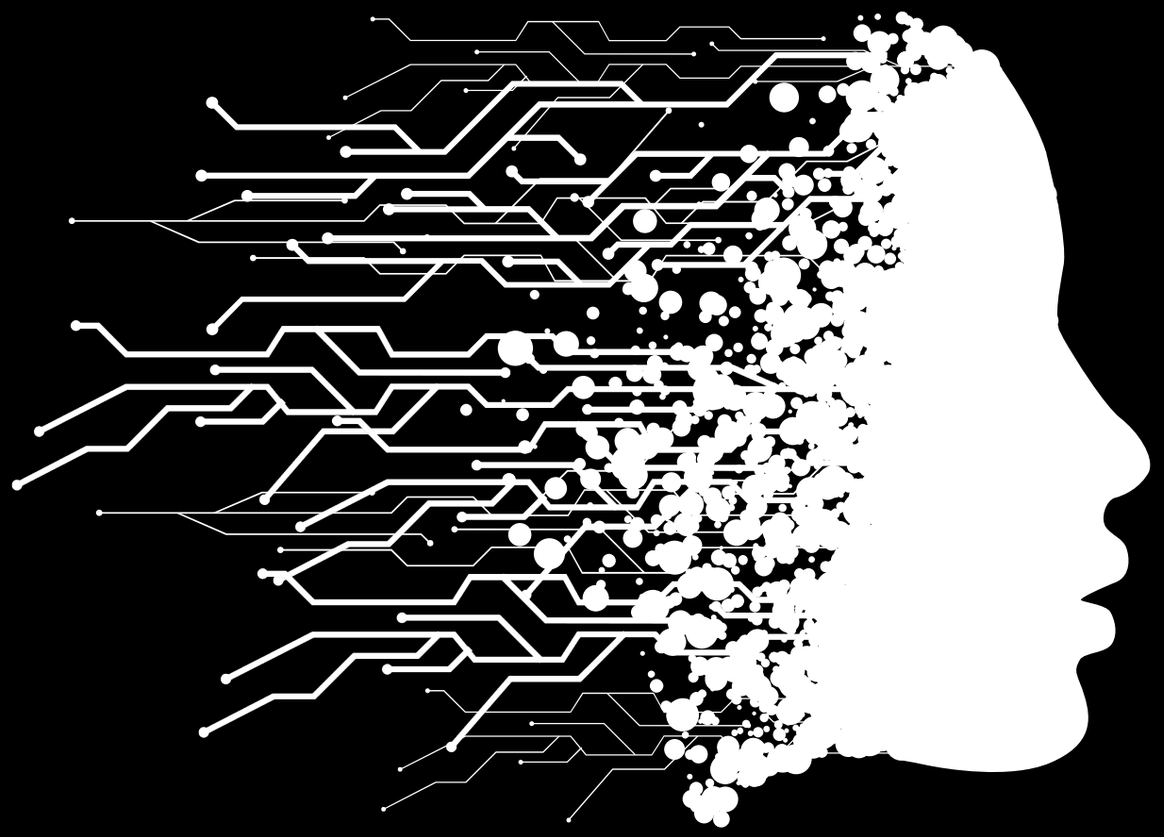


# Fundamentals of Data Science

www.BigDataTrunk.com

# Welcome to Fundamentals of Data Science



PRE-WORK

- Sign-in for Attendance
- Put Name on Name tag
- Download Anaconda if using your laptop (python v3.x)
  **https://www.anaconda.com/distribution/**
- Download material from
  **https://tinyurl.com/yy3fv3f5**

# Agenda

**01** Data Science Overview

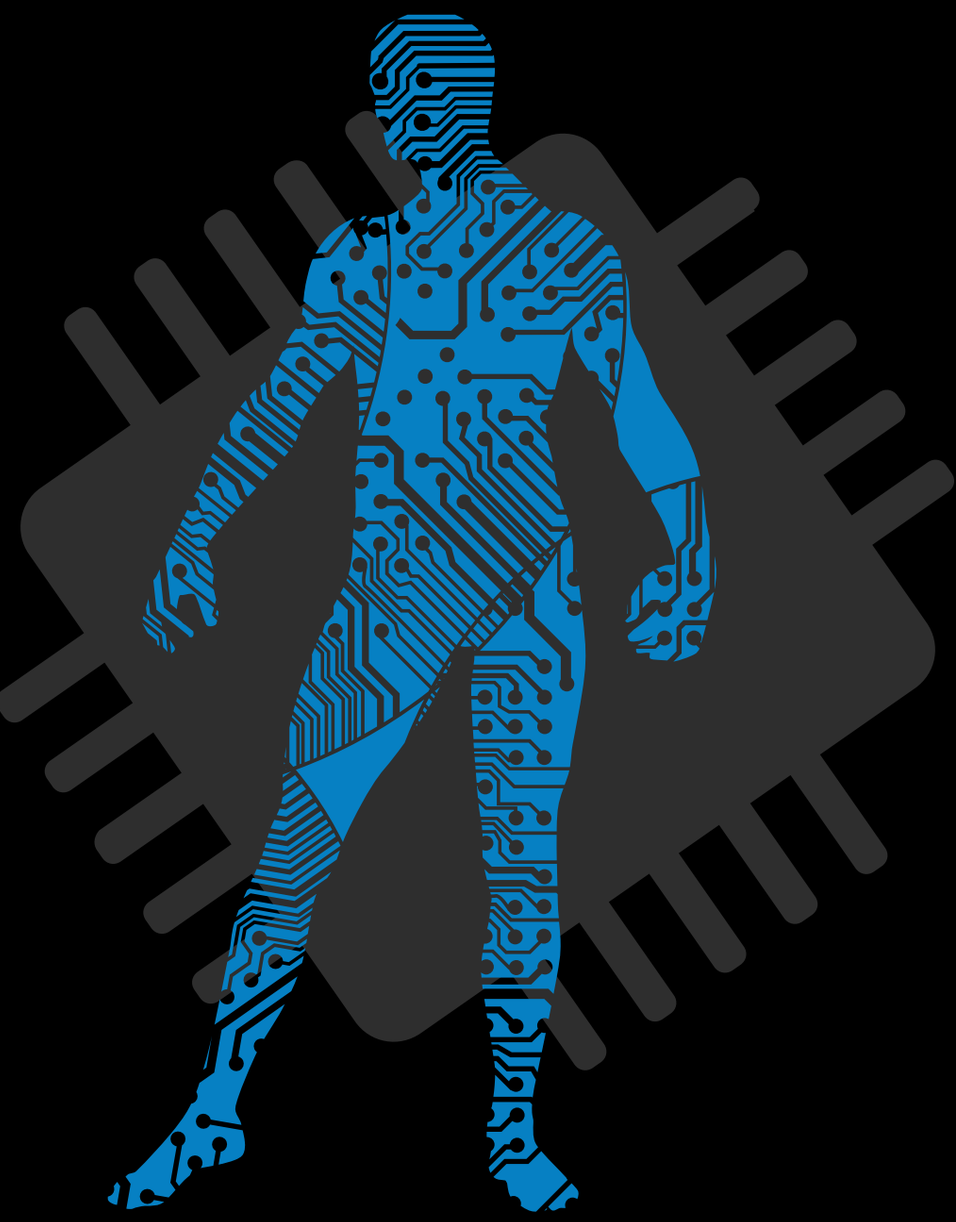
**02** Data Analytics vs. DS

**03** Descriptive Statistics

**04** The Data Science Process

**05** Python Libraries for DS

**06** Next Steps

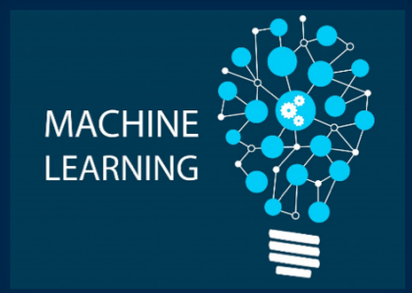


# Logistics

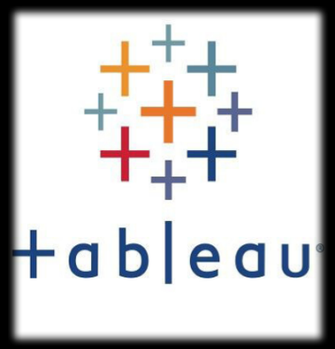
- Timing – 9 to 4 PM PST
- Lunch – Approx. noon to 1 PM PST
- Periodic Breaks – At logical points

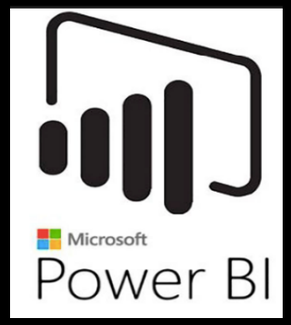
# Stanford &

# Big Data Trunk











Data Science

*www.bigdatatrunk.com*

# Big Data Trunk

## Highlights

- Headquartered in Bay Area, California
- Offshore Development Center in India
- Consulting - Practice Areas
  - Big Data and Data Science
  - Business Intelligences & Analytics
  - Data Warehouses
  - Cloud (Azure/AWS)
- Corporate and Individual trainings
- Products
- E-Verified Company

## Technology Partners

amazon web services  Microsoft  Datameer

www.bigdatatrunk.com

# Training Roadmap

## AI,ML & Data Science

- Fundamentals of Python (2 Days)
- Python for Data Science (1 Day)
- Fundamentals of Data Science (1 Day)
- Data Science : Deep Dive (3 Days)
- Fundamentals of R Programming (2 Days)
- Data Science using R Programming (2 Days)
- Fundamentals of Machine Learning (1 Day)
- Machine Learning : Deep Dive (2 Days)
- Code Free Fundamentals of AI,ML & DS (1 Day)
- Fundamentals of Deep Learning (1 Day)
- TensorFlow : Deep Dive (2 Days)
- Keras : Deep Dive (2 Days)

## Big Data

- Fundamentals of Big Data (1 Day)
- Big Data : Deep Dive (3 Days)

## Cloud Computing

- Fundamentals of Cloud Computing (1 Day)
- Google Cloud Platform Deep Dive (3 Days)
- Amazon Web Services Deep Dive (3 Days)
- Microsoft Azure Deep Dive (3 Days)

## Visualization & Reporting

- Tableau Desktop : Beginners (1 Day)
- Tableau Desktop : Intermediate (2 Days)
- Visualization using Matplotlib (1 Day)
- Power BI : Beginners (1 Day)
- Power BI : Intermediate (2 Days)

*No Coding to deep coding*

*www.bigdatatrunk.com*

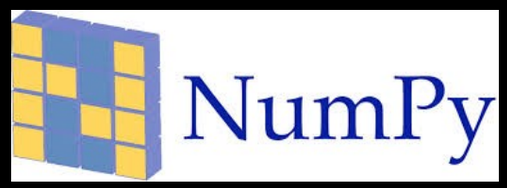
# Introductions

# Hands-on Lab
# Jupyter Notebook

Intro_Jupyter_Notebook.ipynb

# Python Libraries for Data Analysis

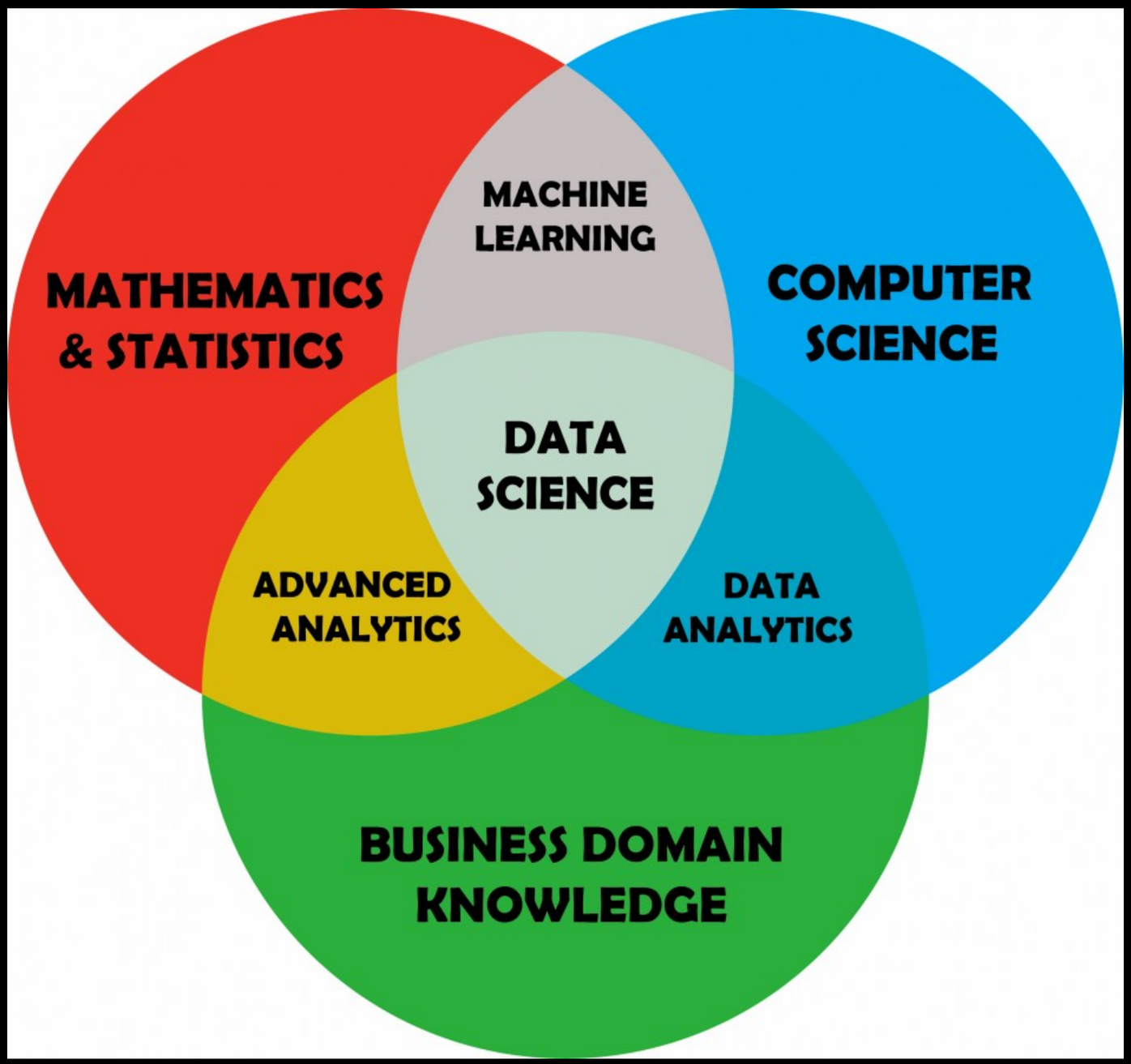






# Data Science

# Data Science



# Data Analytics vs. Data Science

# Python

# Hands-on Lab
# Python Exercise

Python_Exercise.ipynb

# Descriptive Statistics

# The Data Science Process (DIAPERS)

1. Define Problem Statement
2. Ingest Data
3. Analyze Data
4. Prepare Data for ML
5. Evaluate Models
6. Refine Model
7. Ship It

# Define Problem Statement

## Ask good questions:

- answerable
- actionable
- specific
- narrow

# Ingest Data

## Google Dataset Search

toolbox.google.com/datasetsearch

## Kaggle

kaggle.com/datasets

## ProPublica Data Store

propublica.org/datastore

## World Bank Open Data

data.worldbank.org

# Some challenges with data

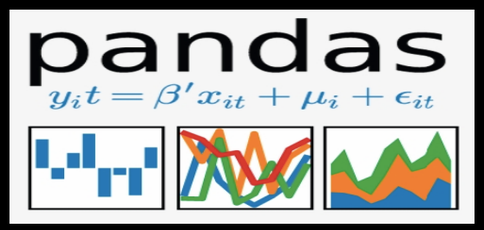
- Insufficient quantity of data
- Non-representative data
- Poor quality of data
- Irrelevant features

# Analyze Data

# Analyze Data

# Prepare Data

# Prepare Data



# Pandas

# Hands-on Lab
# Pandas Exercise

Pandas_Exercise.ipynb

# Evaluate Models

# Survey Time

http://ttseval.stanford.edu

Course Code – ITS-1905

# Machine Learning Model

# Machine Learning Model


.fit : learn the model

.predict : make predictions from learned model

.score : display performance metric

# Scikit-learn

# Hands-on Lab
# Scikit-learn Exercise

Scikit-learn_Exercise.ipynb

# Summary



# Next steps

- Make a custom plan for yourself to continue this journey
- Improve some of your skills (Stats, Python, ML, Domain, etc.)
- Take some additional courses
- Try a Kaggle.com competition
- Work on a personal project to improve understanding

# Useful Data Science and Machine Learning books for beginners to intermediate:

| Book | Author |
|---|---|
| Data Smart: Using Data Science to Transform Information Into Insight | John W. Foreman |
| Naked Statistics: Stripping the Dread from the Data | Charles Wheelan |
| An Introduction to Statistical Learning | Robert Tibshirani and Trevor Hastie |
| Python Machine Learning | Sebastian Raschka and Vahid Mirjalili |
| The Hundred-Page Machine Learning Book | Andriy Burkov |

# Training Roadmap

## AI,ML & Data Science

- Fundamentals of Python (2 Days)
- Python for Data Science (1 Day)
- Fundamentals of Data Science (1 Day)
- Data Science : Deep Dive (3 Days)
- Fundamentals of R Programming (2 Days)
- Data Science using R Programming (2 Days)
- Fundamentals of Machine Learning (1 Day)
- Machine Learning : Deep Dive (2 Days)
- Code Free Fundamentals of AI,ML & DS (1 Day)
- Fundamentals of Deep Learning (1 Day)
- TensorFlow : Deep Dive (2 Days)
- Keras : Deep Dive (2 Days)

## Big Data

- Fundamentals of Big Data (1 Day)
- Big Data : Deep Dive (3 Days)

## Cloud Computing

- Fundamentals of Cloud Computing (1 Day)
- Google Cloud Platform Deep Dive (3 Days)
- Amazon Web Services Deep Dive (3 Days)
- Microsoft Azure Deep Dive (3 Days)

## Visualization & Reporting

- Tableau Desktop : Beginners (1 Day)
- Tableau Desktop : Intermediate (2 Days)
- Visualization using Matplotlib (1 Day)
- Power BI : Beginners (1 Day)
- Power BI : Intermediate (2 Days)

*No Coding to deep coding*

*www.bigdatatrunk.com*

# Stanford Upcoming Training

- Click here for latest schedule

# Thank you

www.BigDataTrunk.com

Reminder : Remove Name tags ☺