What the hell is 'Data Science' anyway?
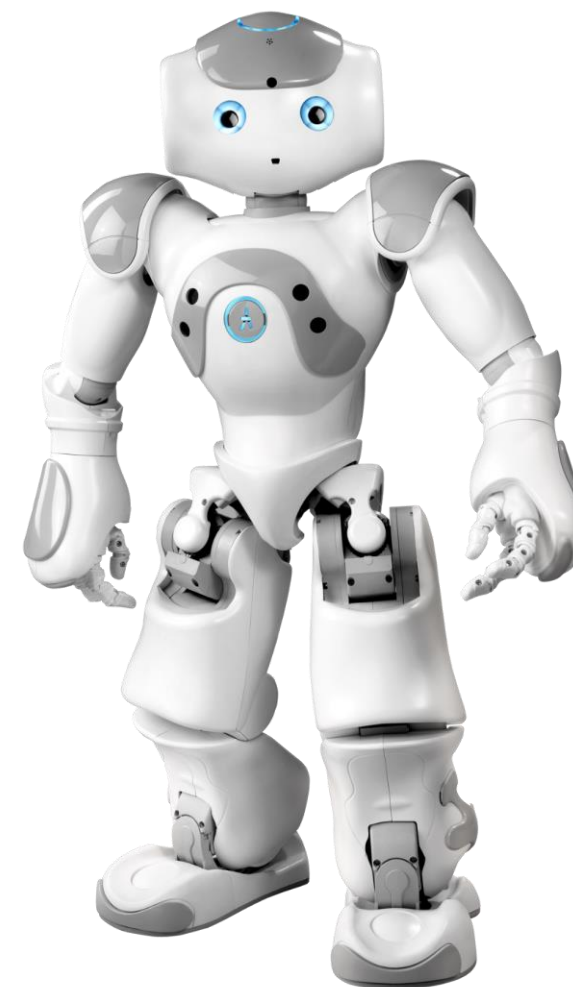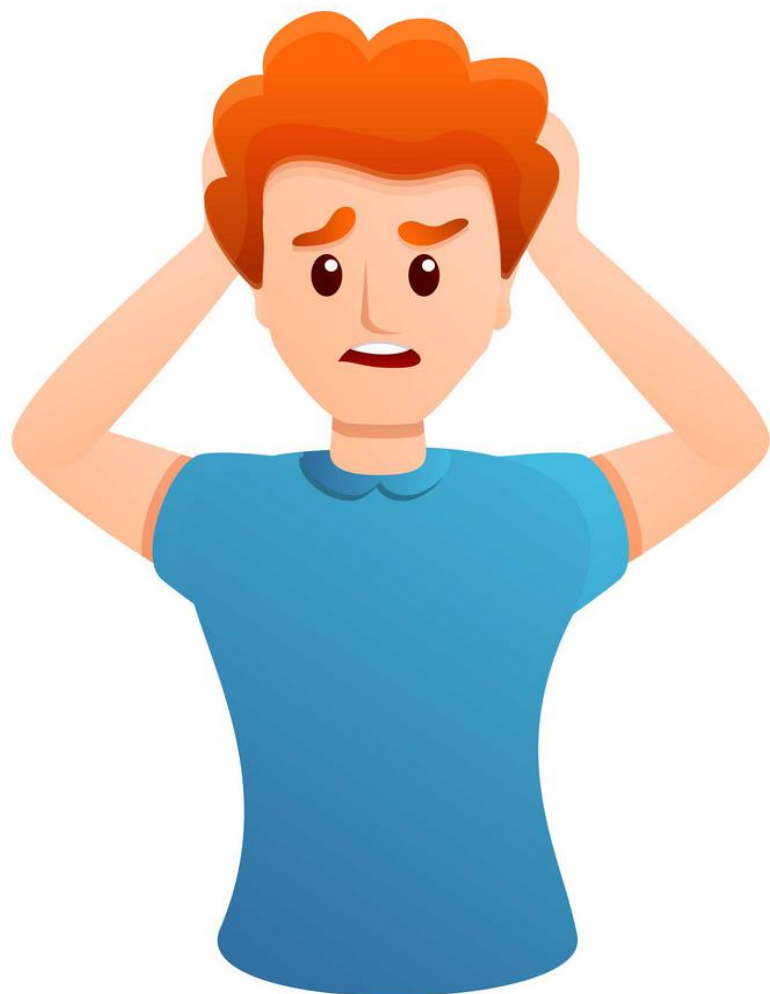
What is Data?

Data is defined as a collection of organized or unorganized facts, concepts, or instructions in a formalized manner suitable for communication, interpretation, processing by humans, or some automatic means such as computers, ATMs.

The main examples of data are phone numbers, weights, prices, costs, number of items sold, product names, addresses, registration marks, etc.
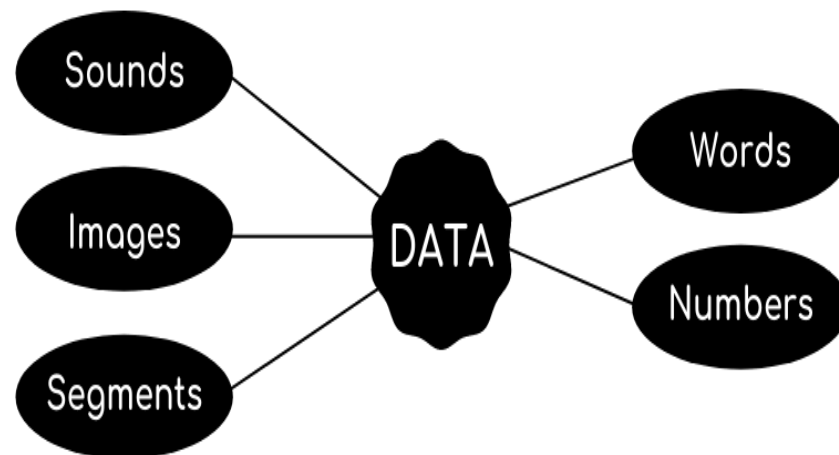
What is Information?

Information is defined as the **processed data** which helps us in making **decisions**. It is an intelligent form of data. such as bills, profits, reports, receipts, comparison of sales figures, merit list, printed documents, etc.

For example, the students' marks and their roll numbers form the data, and their report card/sheet is the information that helps us decide which student stood at which rank in the class.

Data Science & Machine Learning
with Python

What is Database?

A database is an organized collection of related data, or information, which is stored and accessed electronically within a computer system.

For example, SQL, MongoDB, Oracle Database, etc. are all examples of different databases. These modern databases are managed by DBMS. Structured Query Language, or SQL as it is more widely known, is used to operate on the data in a database.

Data Science & Machine Learning with Python

WHY DATA SCIENCE ??

Have a look!

1024 KB = 1MB
1024 MB = 1GB

1000 GB = 1TB
1000 TB = 1PB

1000 PB = 1EB
1000 EB = 1 Zettabyte

Note: Zettabyte is approximately equal to a thousand Exabytes, a billion Terabytes, or a trillion Gigabytes.
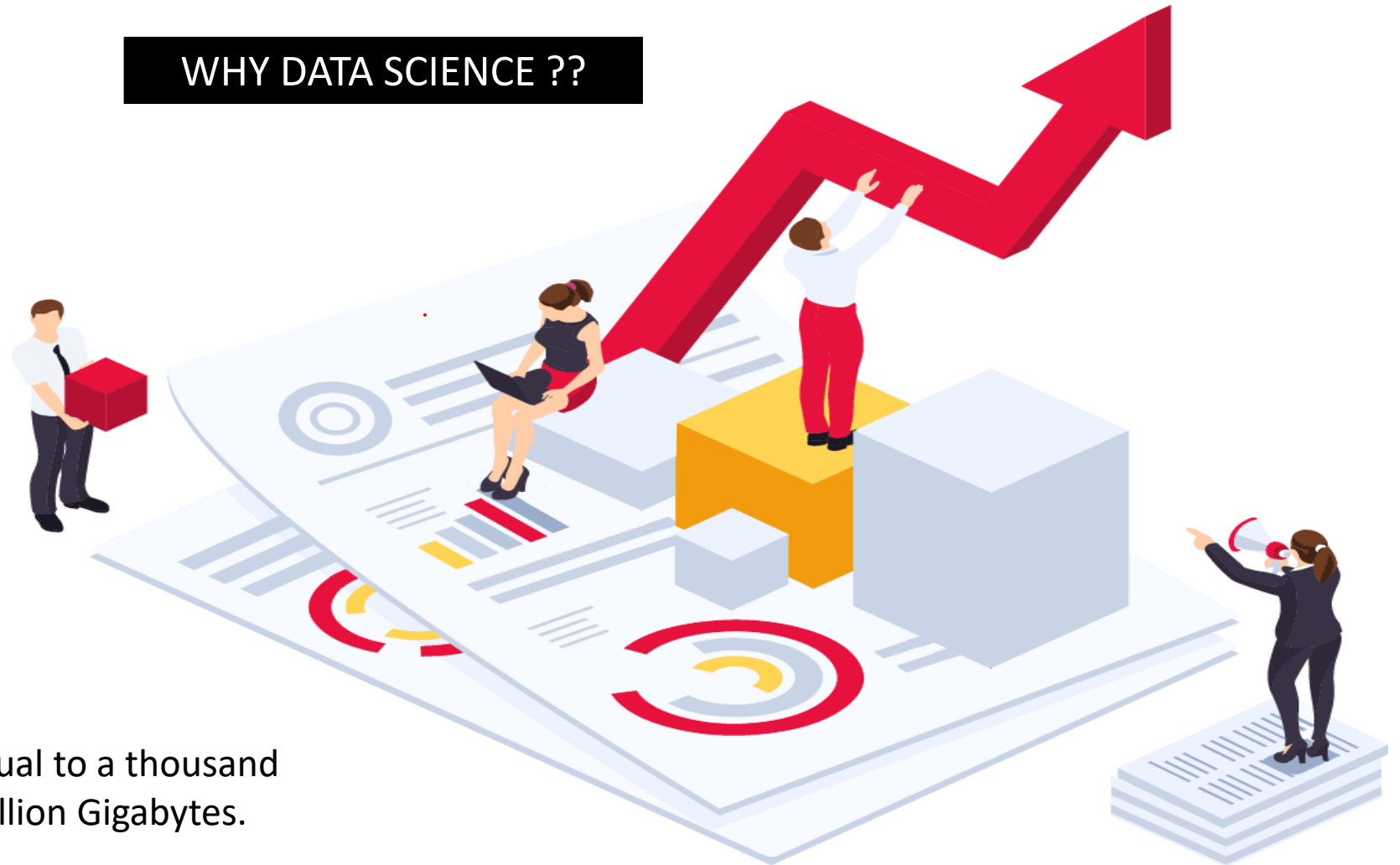
Data Science & Machine Learning
with Python

What is Data Science?

▪ Computer Science
▪ Mathematics
▪ Statistics
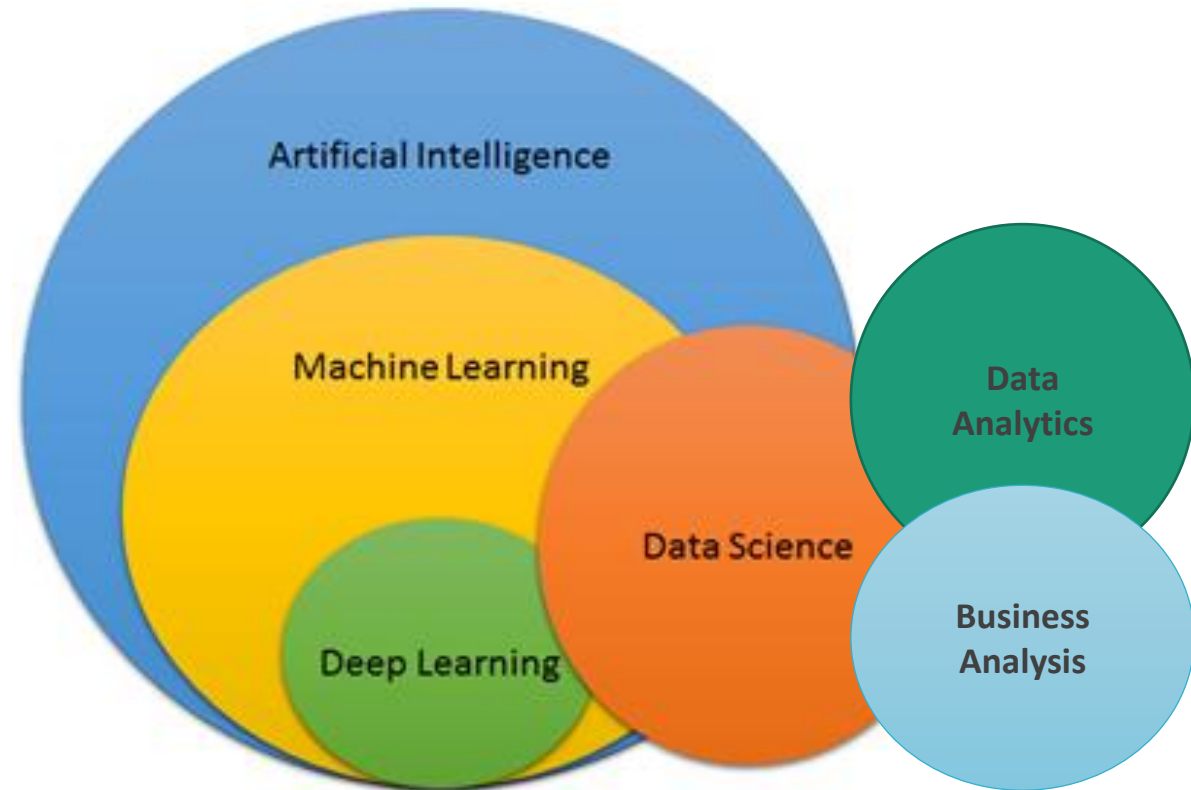
Fig: Venn Diagram

Data Science & Machine Learning with Python

Machine Learning vs Deep Learning
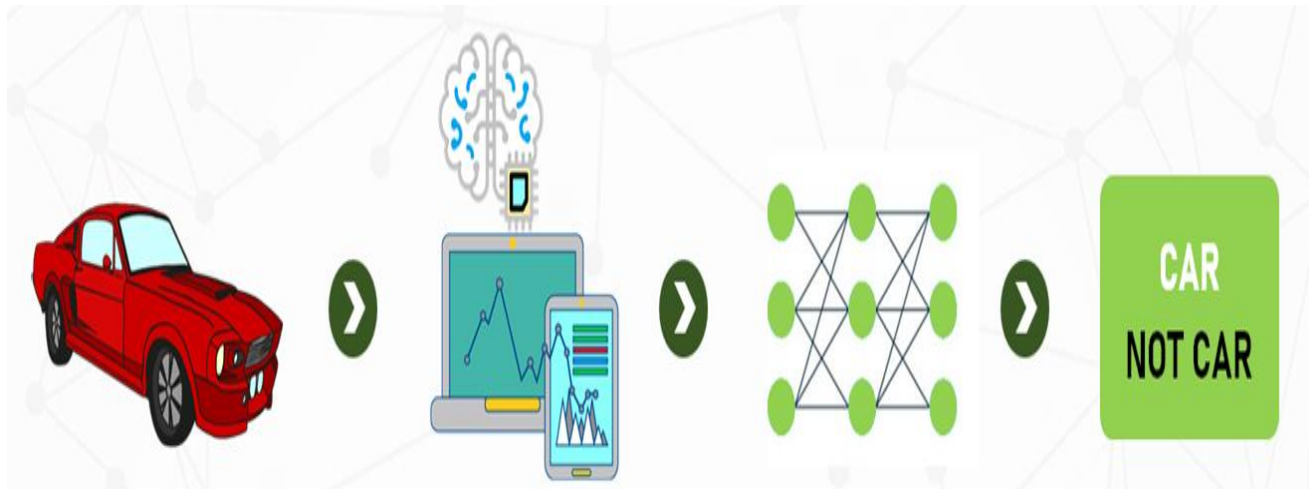
Input

Extraction

Classification

Output

Machine Learning

Input Layer

Hidden Layer 1

Hidden Layer 2

Output Layer

Deep Learning

Why Data Science So Popular?

- DDDM
- Market Basket
- Machine Learning
- Deep Learning
- Artificial Intelligence
- Data Analysis
- Big Data
- Industry 4.0

Data Science & Machine Learning
with Python

How Much Does a Data Scientist Make?

Data Scientists made a median salary of $94,280 in 2019. The best-paid 25 percent made $125,140 that year, while the lowest-paid 25 percent made $70,090.

$94,280

$70,090 $125,140

25%   Median   75%

Resource: www.glassdoor.com
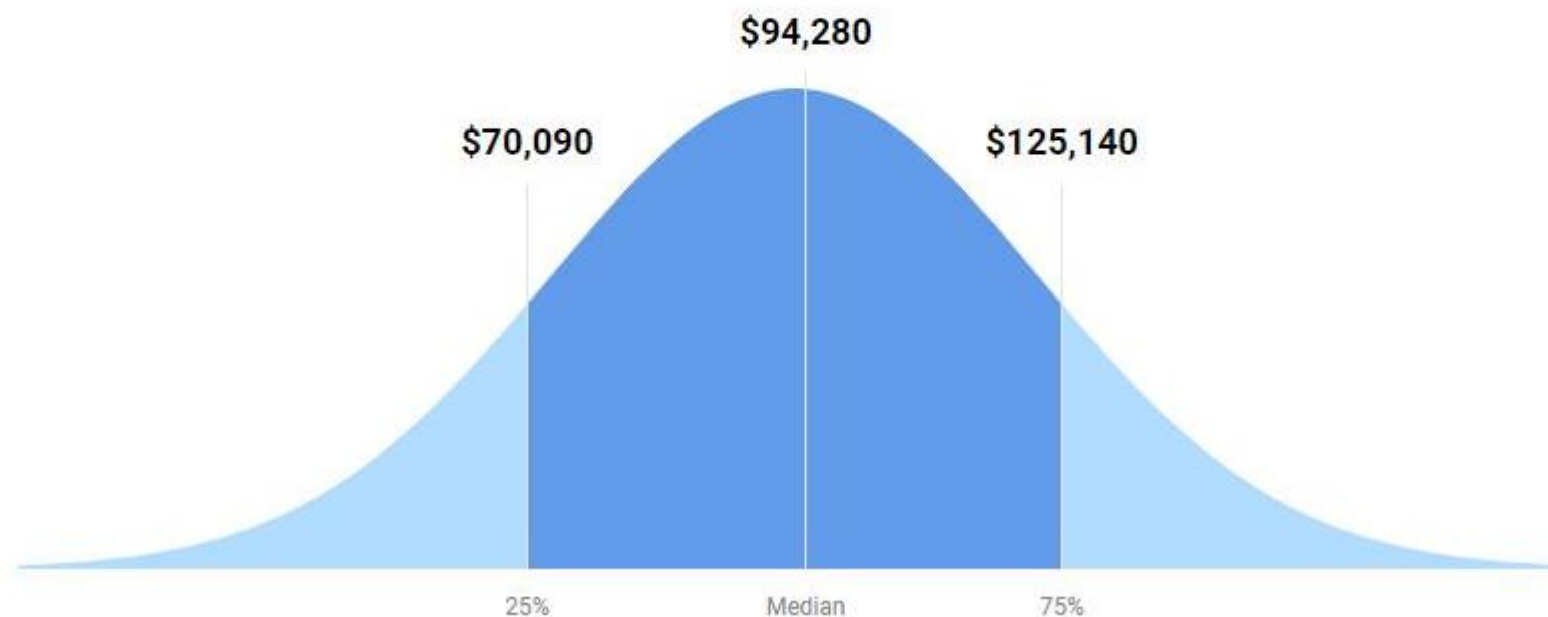
Learn: Become a Data Scientist

AI Quest

Programming Language
- PYTHON
- R

Data Analysis Skills
- SQL
- Data Maintenance
- Feature Engineering
- Exploratory Data Analysis
- Pandas, Seaborn, Matplotlib, GGplot

Data Analysis Tools
- Google Data Studio
- Power Bi
- Tableau
- Stata
- Excel
- SPSS

Big Data Tools
- Hadoop
- Hive
- Spark

Environment: IDE
- Jupyter Notebook
- Colab
- R Studio
- Jupyter Lab

Mathematics
- Statistics & Probability
- Calculus
- Linear Algebra
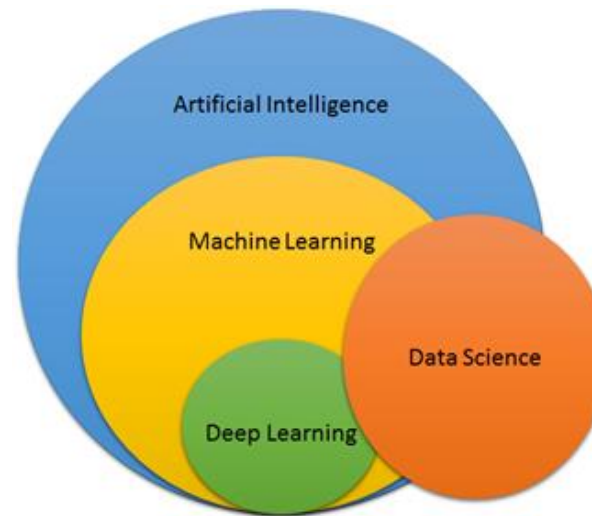
Cloud
- Amazon AWS
- MS Azure

Artificial Intelligence: ML,DL
- Supervised Learning
- Unsupervised Learning
- Neural Network (DL,NLP,CV,OD)

Artificial Intelligence
Machine Learning
Data Science
Deep Learning

Data Science & Machine Learning
with Python

What is Machine Learning?

ML = Machine + Learning
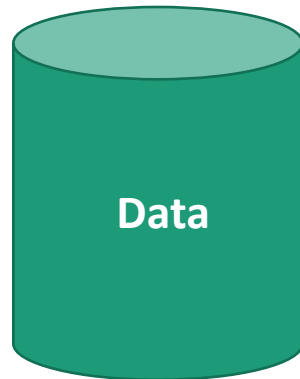
Data

Training →

Testing →

Create Artificial Brain Using ML

Able to Make Prediction
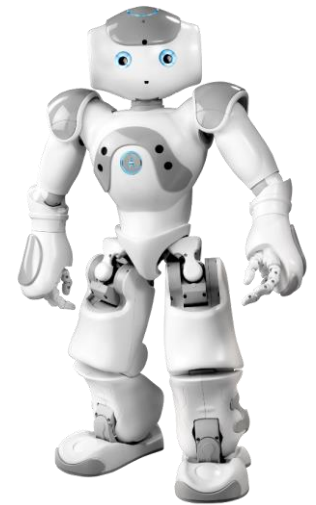
Artificial Intelligence

Data Science & Machine Learning
with Python

Let's have an Example -

Training Data

Apples   Cupcakes

Machine
Learning Model

Class : Cupcake

Unseen and
Unlabeled data
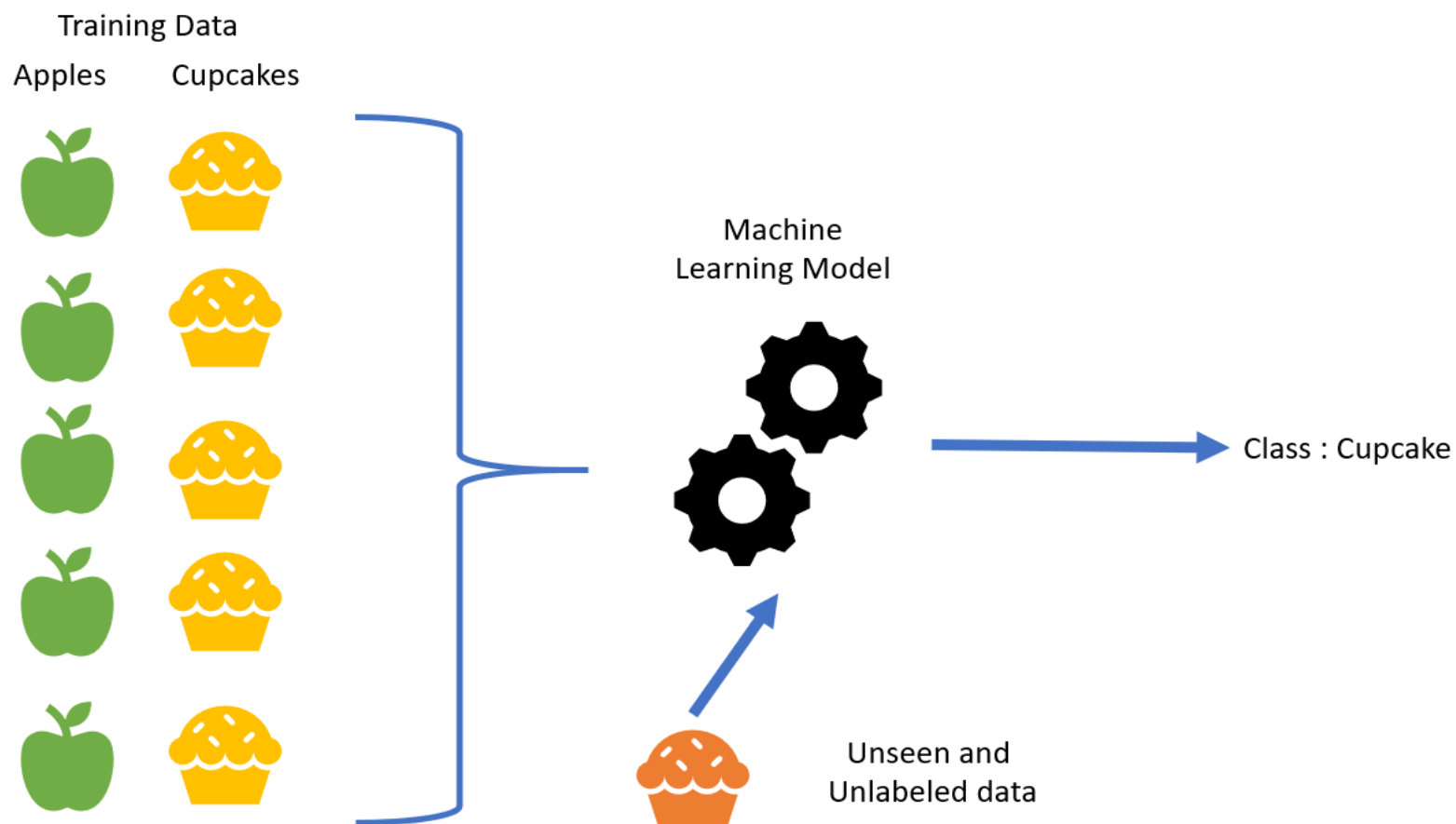
Data Science & Machine Learning with Python

Real Life Applications!

- Spam Emails or Comments
- Stock Market Analysis
- Healthcare Industry
- Drug Discovery
- Medical Diagnosis
- Agriculture
- Transport
- Weather Forecasting
- Market Basket Analysis
- Cyber Risk Analysis
- Speech Recognition
- Financial Services
- And So On.....

Data Science & Machine Learning for ALL with Python

> Types of Learning –

- Supervised Learning
- Unsupervised Learning
- Reinforcement learning

Data Science & Machine Learning for ALL with Python
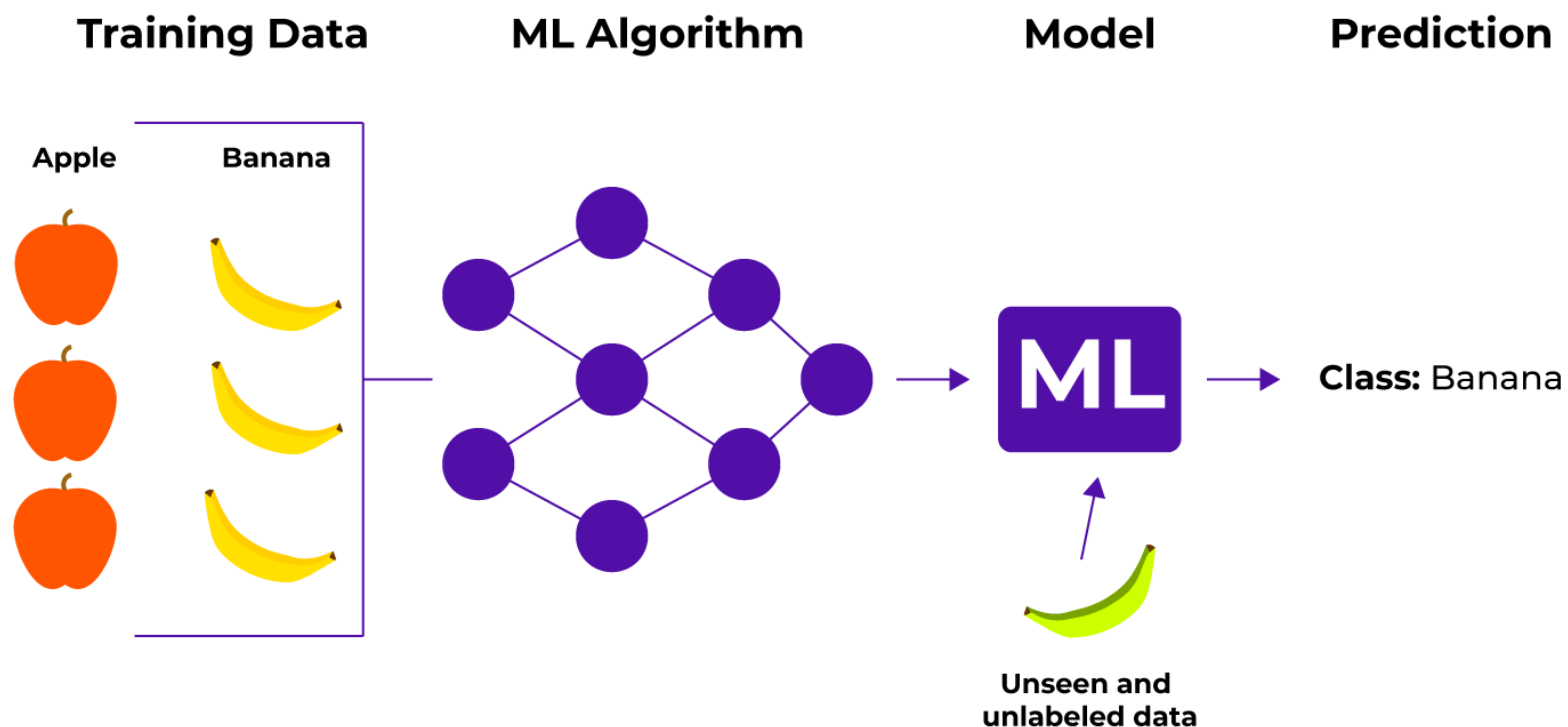
Supervised Learning:

Training Data   **ML Algorithm**   **Model**   **Prediction**

Apple   Banana

Class: Banana

Unseen and unlabeled data

Mind it

- Labeled
- At least two classes

Data Science & Machine Learning for ALL with Python

Supervised Learning:

We Have -

Regression

Classification

Mind it

- Labeled
- At least two classes

Data Science & Machine Learning for ALL with Python

Regression

Let's have an another Example–

Output Column

Features:

- X1
- X2
- X3
- X4
- X5
- X6

Mind it

- Labeled
- Continuous Data

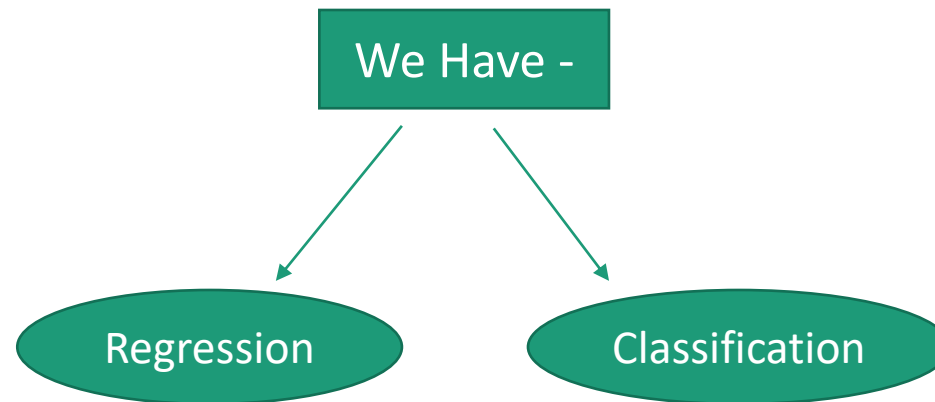| No	X1 transaction	X2 house age	X3 distance	X4 number	X5 latitude	X6 longitude	Area
1	2012.917	32	84.87882	10	24.98298	121.54024	37.9
2	2012.917	19.5	306.5947	9	24.98034	121.53951	42.2
3	2013.583	13.3	561.9845	5	24.98746	121.54391	47.3
4	2013.5	13.3	561.9845	5	24.98746	121.54391	54.8
5	2012.833	5	390.5684	5	24.97937	121.54245	43.1
6	2012.667	7.1	2175.03	3	24.96305	121.51254	32.1
7	2012.667	34.5	623.4731	7	24.97933	121.53642	40.3
8	2013.417	20.3	287.6025	6	24.98042	121.54228	46.7
9	2013.5	31.7	5512.038	1	24.95095	121.48458	18.8
10	2013.417	17.9	1783.18	3	24.96731	121.51486	22.1
11	2013.083	34.8	405.2134	1	24.97349	121.53372	41.4
12	2013.333	6.3	90.45606	9	24.97433	121.5431	58.1
13	2012.917	13	492.2313	5	24.96515	121.53737	39.3
14	2012.667	20.4	2469.645	4	24.96108	121.51046	23.8
15	2013.5	13.2	1164.838	4	24.99156	121.53406	34.3
16	2013.583	35.7	579.2083	2	24.9824	121.54619	50.5
17	2013.25	0	292.9978	6	24.97744	121.54458	70.1
18	2012.75	17.7	350.8515	1	24.97544	121.53119	37.4
19	2013.417	16.9	368.1363	8	24.9675	121.54451	42.3
20	2012.667	1.5	23.38284	7	24.96772	121.54102	47.7

Data Science & Machine Learning
for ALL with Python

Classification

Let's have another Example-

Label / Class

Features:

- Age
- Income
- Gender
- M_status

	A	B	C	D	E
1	age	income	gender	m_status	buys
2	<25	high	male	single	no
3	<25	high	male	married	no
4	25-35	high	male	single	yes
5	>35	medium	male	single	yes
6	>35	low	female	single	yes
7	>35	low	female	single	no
8	25-35	low	female	married	yes
9	<25	medium	male	married	no
10	<25	low	female	single	yes
11	>35	medium	female	married	yes
12	<25	medium	female	single	yes
13	25-35	medium	male	married	yes
14	25-35	high	female	single	yes
15	>35	medium	male	married	no
16	<25	high	male	single	no
17					

Mind it

- Labeled
- At least two classes

Data Science & Machine Learning
for ALL with Python

Unsupervised Learning

INPUT DATA

OUTPUT DATA

ML Model
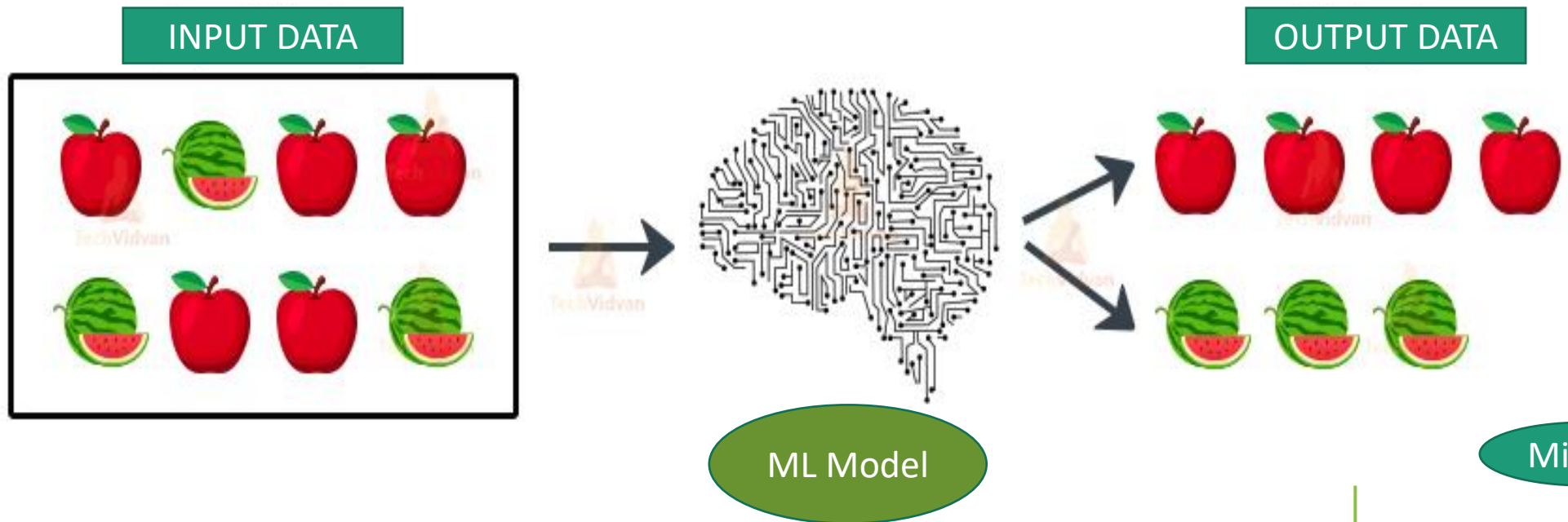
Mind it

- Un-Labeled
- At least two classes

Data Science & Machine Learning
for ALL with Python

Unsupervised Learning

Features:

- Gender
- Age
- Income
- Score

Customer	Gender	Age	Annual Income (k$)	Spending Score (1-100)
1	Male	19	15	39
2	Male	21	15	81
3	Female	20	16	6
4	Female	23	16	77
5	Female	31	17	40
6	Female	22	17	76
7	Female	35	18	6
8	Female	23	18	94
9	Male	64	19	3
10	Female	30	19	72
11	Male	67	19	14
12	Female	35	19	99
13	Female	58	20	15
14	Female	24	20	77
15	Male	37	20	13
16	Male	22	20	79
17	Female	35	21	35
18	Male	20	21	66
19	Male	52	23	29
20	Female	35	23	98

Mind it

- Un-Labeled
- At least two classes

Data Science & Machine Learning
for ALL with Python

AI Quest

12 Most Popular Algorithms in ML

Regression Algorithm:
- Linear Regression
- XGBoost (Regressor & Classifier)
- AdaBoost (Regressor & Classifier)
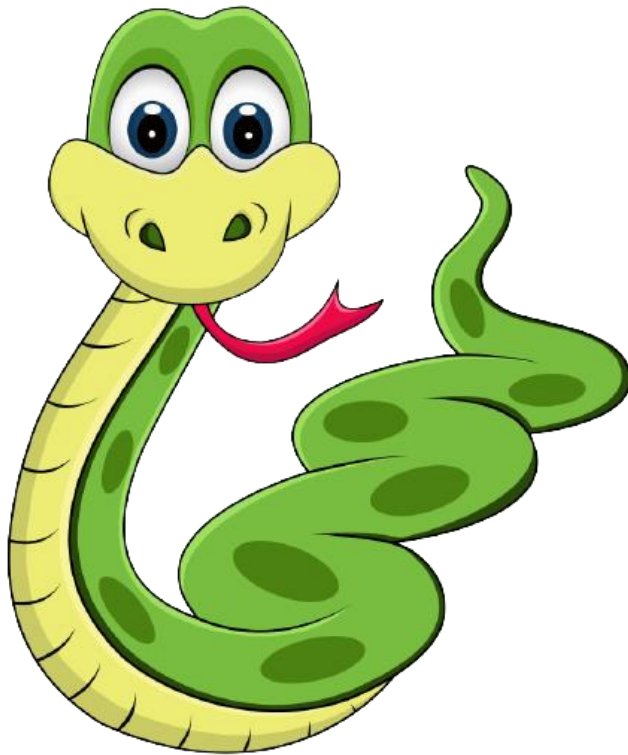
Classification Algorithms: (Also Have Regressor)
- Logistic Regression
- Decision Tree
- Naïve Bayes
- Random Forest
- SVM
- KNN

Unsupervised Algorithm:
- Clustering (K-means, O-Cluster)
- Association Rule Learning
- PCA

Data Science & Machine Learning
for ALL with Python

Why Python for Data Science?

- Easy Syntax, Flexible, Support OOP & Faster
- Python has Machine Learning Libraries
- Python has Data Analysis Library
- Python has Data Frame Library
- Python has Calculator Library
- Python is Significant for Deep Learning
- Keras, Tensorflow, Pytorch
- Web (Django & Flask)
- Open Resources
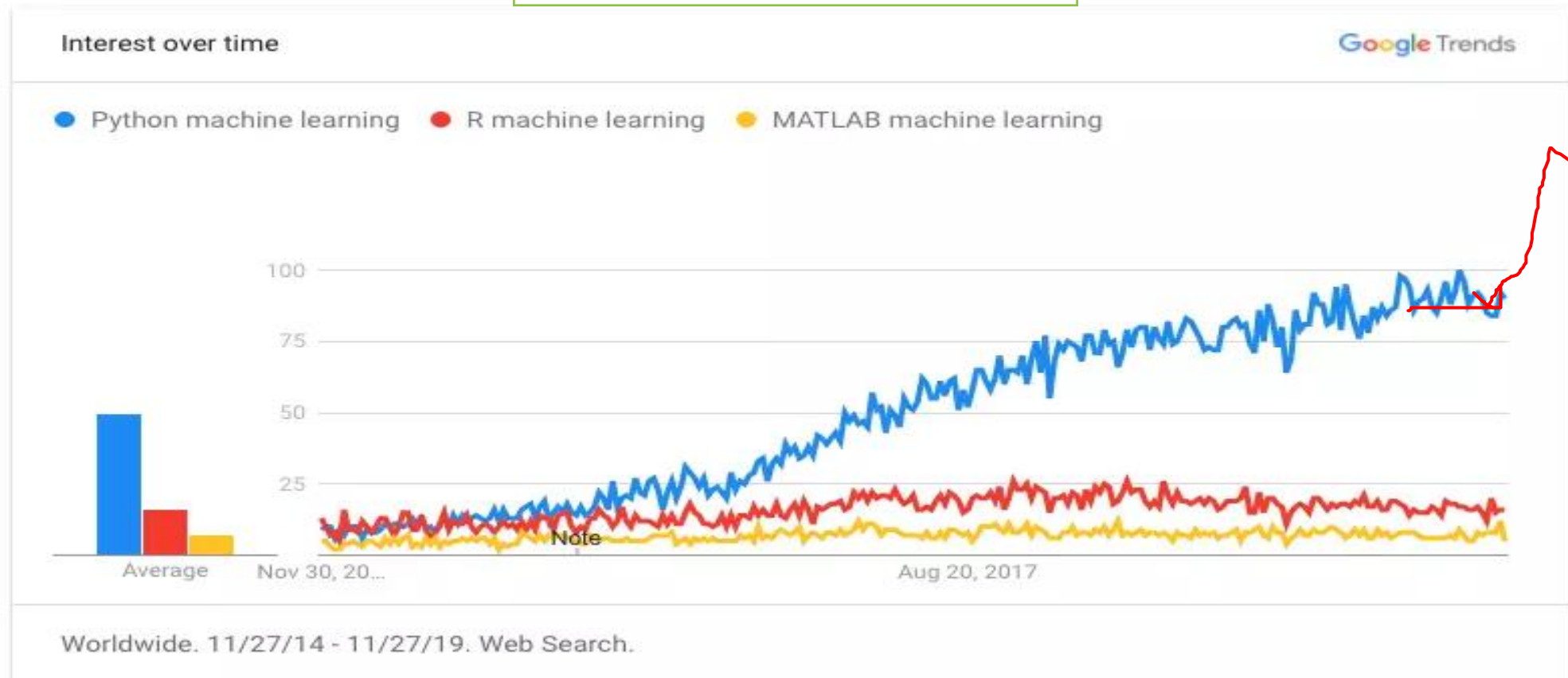
Data Science & Machine Learning
for ALL with Python

PYTHON vs R vs MATLAB

Interest over time — Google Trends

- Python machine learning
- R machine learning
- MATLAB machine learning
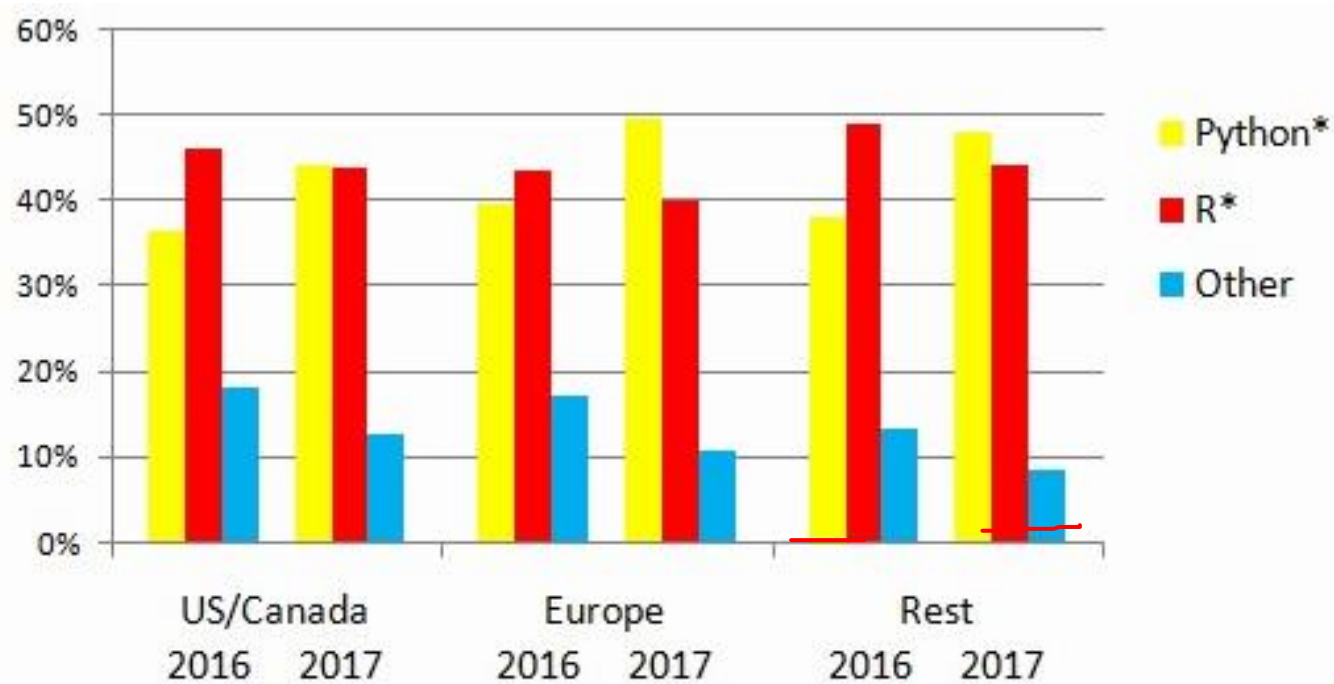
Worldwide. 11/27/14 - 11/27/19. Web Search.

Data Science & Machine Learning
for ALL with Python

PYTHON vs R for Data Analysis

Data Science & Machine Learning
for ALL with Python

Data Science Important Tools

- WEKA
- STATA
- SPSS
- Tableau
- PowberBi
- Hadoop
- Google Data Studio
- MS Excel
- Apache Spark

Future Jobs: Based on Data Science

- Data Scientist ($139,840/year)
- Machine Learning Engineer ($114,826/year)
- Data Architect ($108,278/year)
- Data Engineer ($102,864 /year)
- Business Intelligence (BI) Developer($81,514 /year)
- Statistician ($76,884/year)
- Database Administrator ($72,400/year)
- Data Analyst($62, 453/year)

Our all-paid courses on data science!

Our Paid Courses:

- ☑ **List of all Paid Courses:** www.aiquest.org
- ☑ **Deep Learning: https://www.aiquest.org/courses/deep-learning-nlp/**
- ☑ **SQL for DS: https://www.aiquest.org/courses/sql-for-data-science/**
- ☑ **Become a Data Analyst: https://www.aiquest.org/courses/become-a-data-analyst/**
- ☑ **Big Data Engineering: https://www.aiquest.org/courses/become-a-big-data-engineer/**
- ☑ **Machine Learning: https://www.aiquest.org/courses/data-science-machine-learning/**
- ☑ **To Learn About Paid Courses, Contact:** +8801704265972 (Sohan Khan)