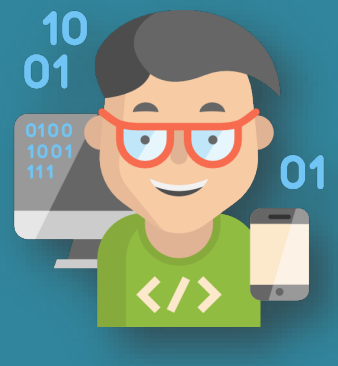



greatlearning


# Uber Data Analysis with Python

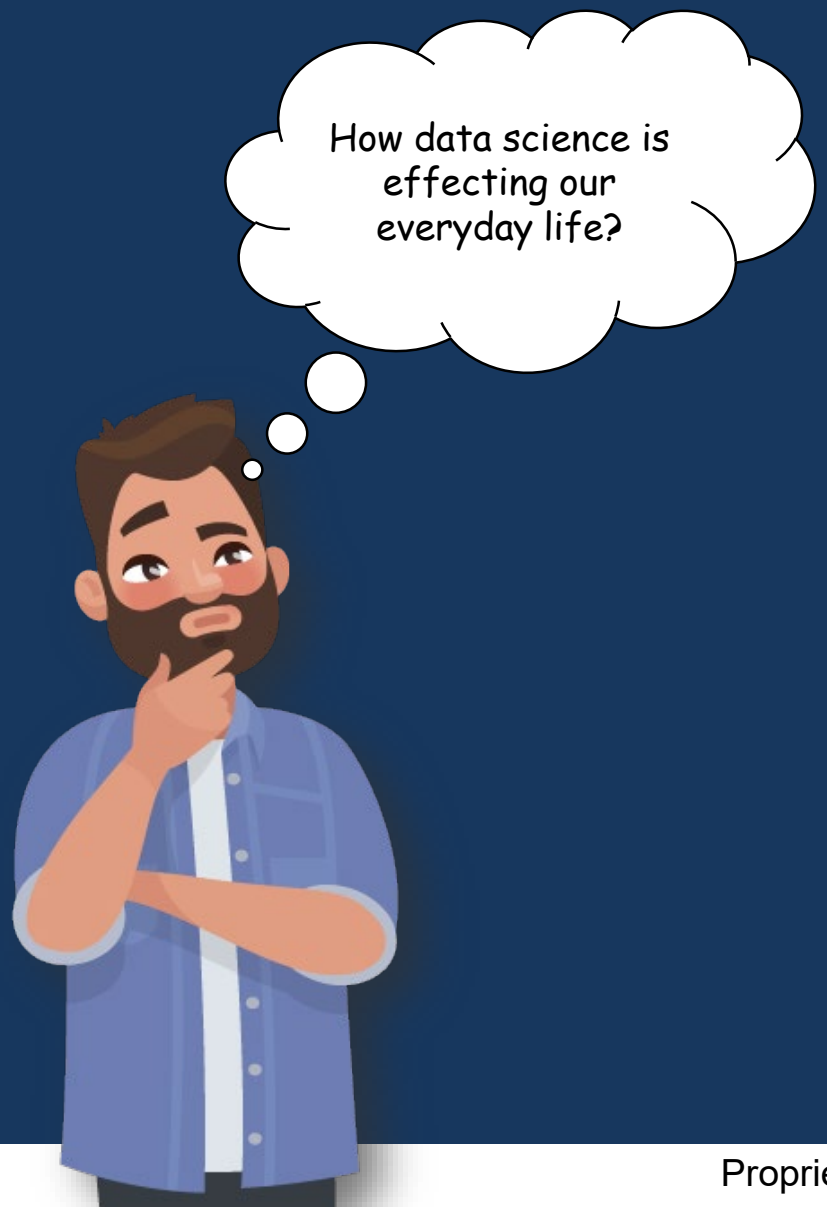
**Sampriti Chatterjee (Great Learning)**

# Agenda

greatlearning
Learning for Life

1. Why do we need data science?
2. What is Data science?
3. Life cycle of Data science
4. Why Python is so popular?
5. Install python
6. Statistical visualization on Python user
7. Basic Syntax for Python
8. List,Set.Tuple
9. Data manipulation using Numpy and Pandas
10. What is machine Learning?
11. Supervised Learning: Linear Regression
12. Uber data analysis to predict price

# Why do we need Data Science?




- In the past, we used to have data in a structured format but now as the volume of the data is increasing, so the number of structured data becomes very less, so to handle the massive amount of data we need data science techniques
- Those data can be used to get the proper business insights and the hidden trends from them.
- These insights helps the organization to predict the Future
- Using data science decision making can be faster and effective
- Helps to reduce the production cost
- Build model based on the data to give the ability to the machine to predicts on its own

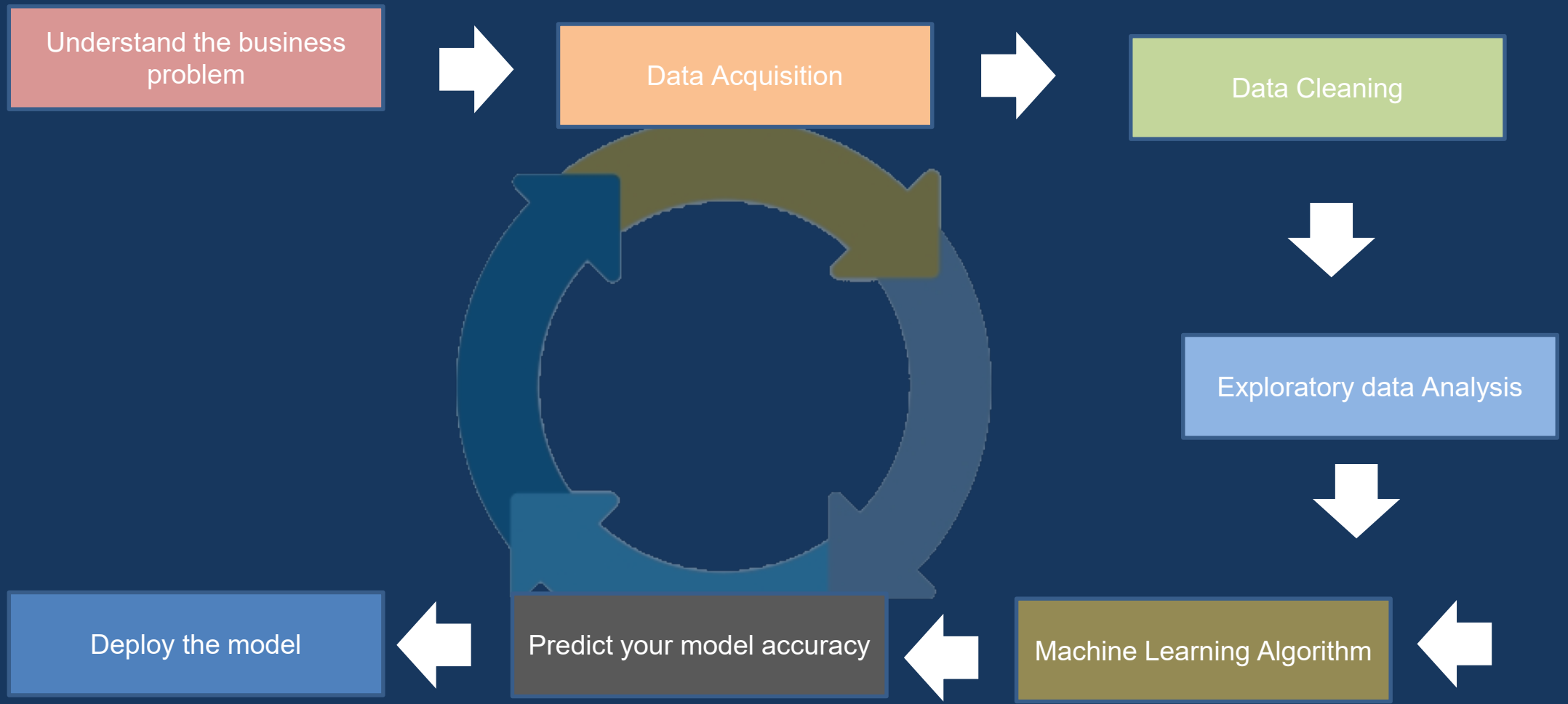
# What is Data Science?

greatlearning
Learning for Life

Data science is a process to get some meaningful information from the massive amount of data. In simple terms, read and study the data to get proper intuitive insights. Data Science is a mixture of various tools, algorithms, and machine learning and deep learning concepts to discover hidden patterns from the raw and unstructured data

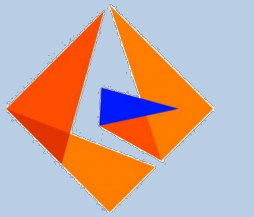
# Life cycle of Data Science?

greatlearning
Learning for Life



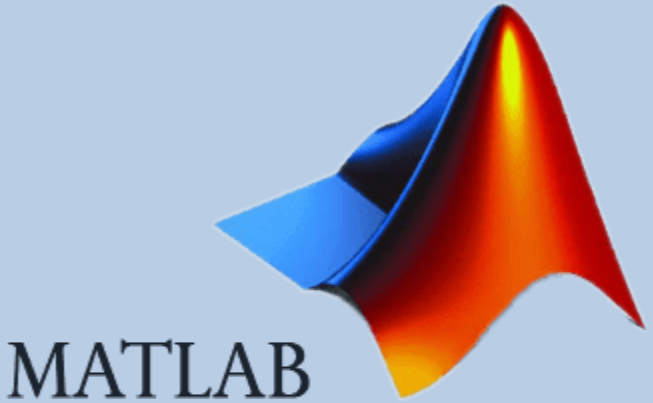
# Most Popular Programming Languages For Data Science

greatlearning
Learning for Life

Informatica™

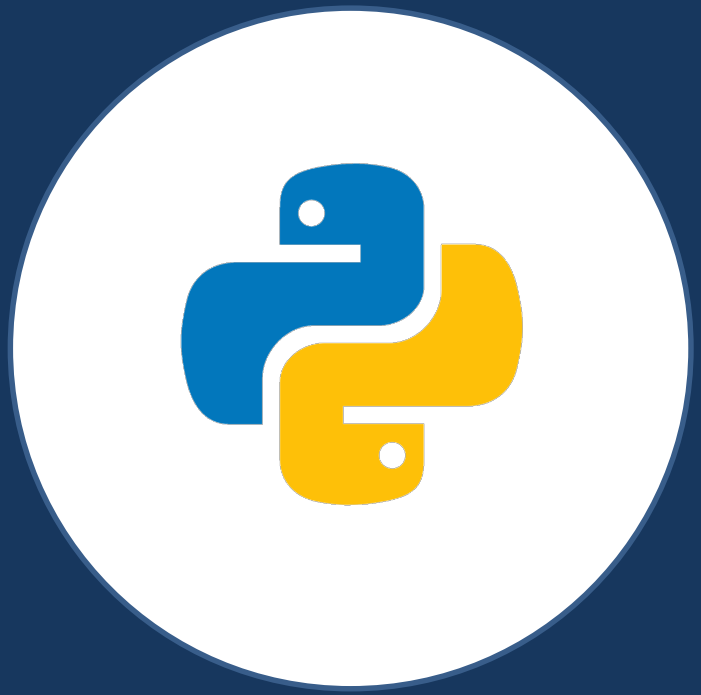


greatlearning
Learning for Life

## Python is a popular high level, object oriented and interpreted language

High level



Interpreted

Object oriented

# History of Python


Inventor of Python

## Important Facts

- Python is invented by Guido van Rossum in 1989
- Rossum used to love watching comedy movies from late seventies
- He needed a short, unique, and slightly mysterious name for his language
- In that time he was watching Monty Python's Flying Circus and from that series he decided to keep his language name python.
- This how Python invented

# Why should you learn Python?

greatlearning
Learning for Life

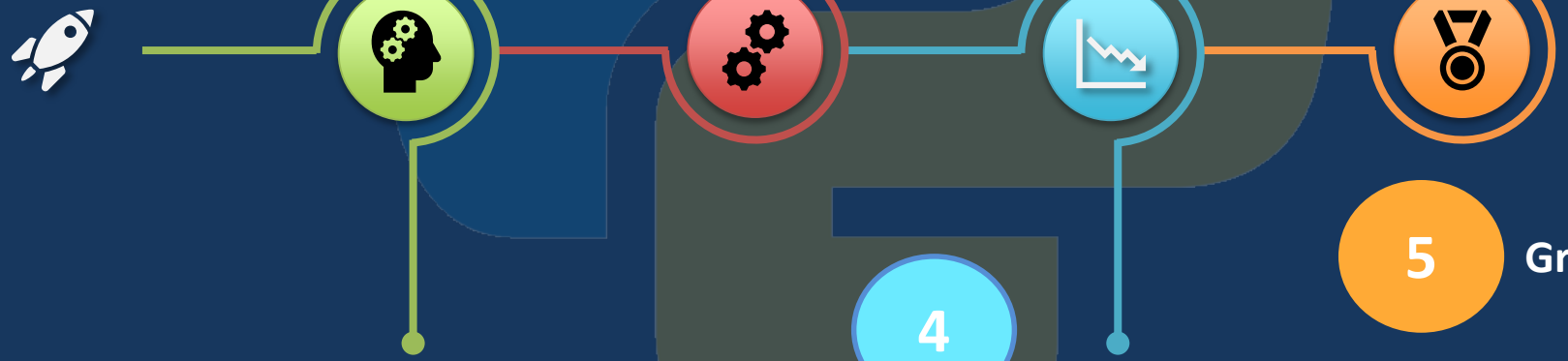
1 Web development using Python

2 Length of the program is short

3 Python is simple and beginner friendly language

4 Mathematical computation can be done easily

5 Graphical user interface



# Why Python is so popular?

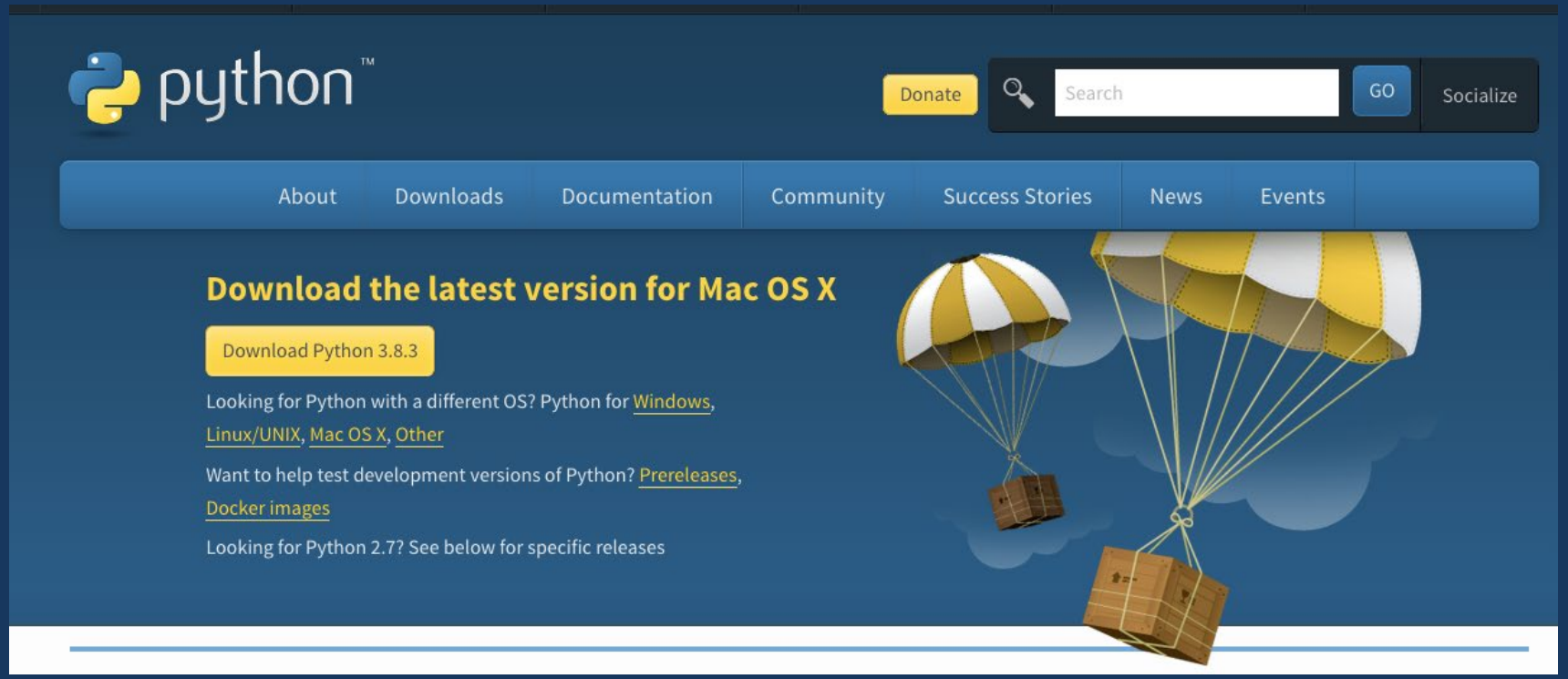
greatlearning
Learning for Life

1. Largest community for Learners and Collaborators
2. Open source
3. Easy to learn and usable flexibility
4. Huge numbers of Python libraries and Frame work
5. Supports Big Data, Machine Learning and Cloud computing
6. Supports Automation

# Installing Python

greatlearning
Learning for Life

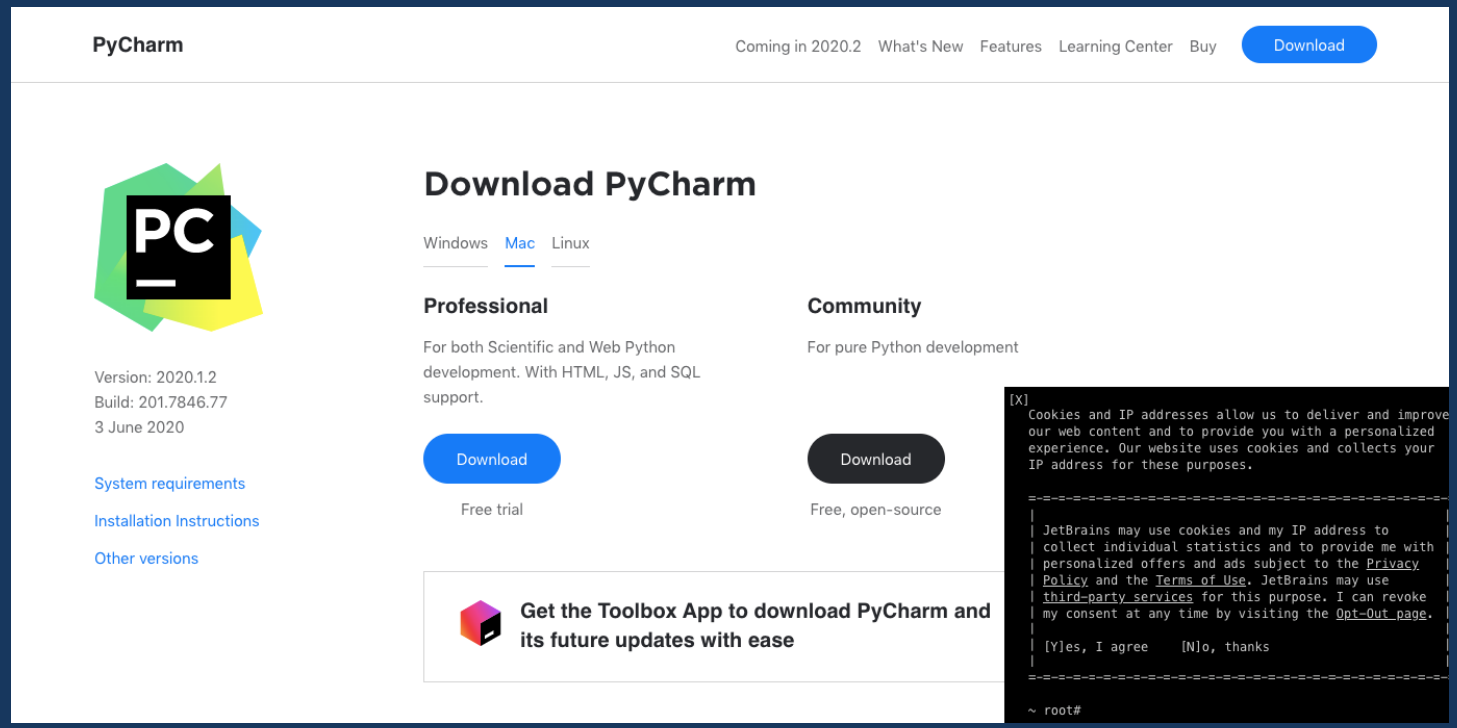
This is the site to install Python -> https://www.python.org/downloads/



# Popular IDE for Python: Pycharm

greatlearning
Learning for Life

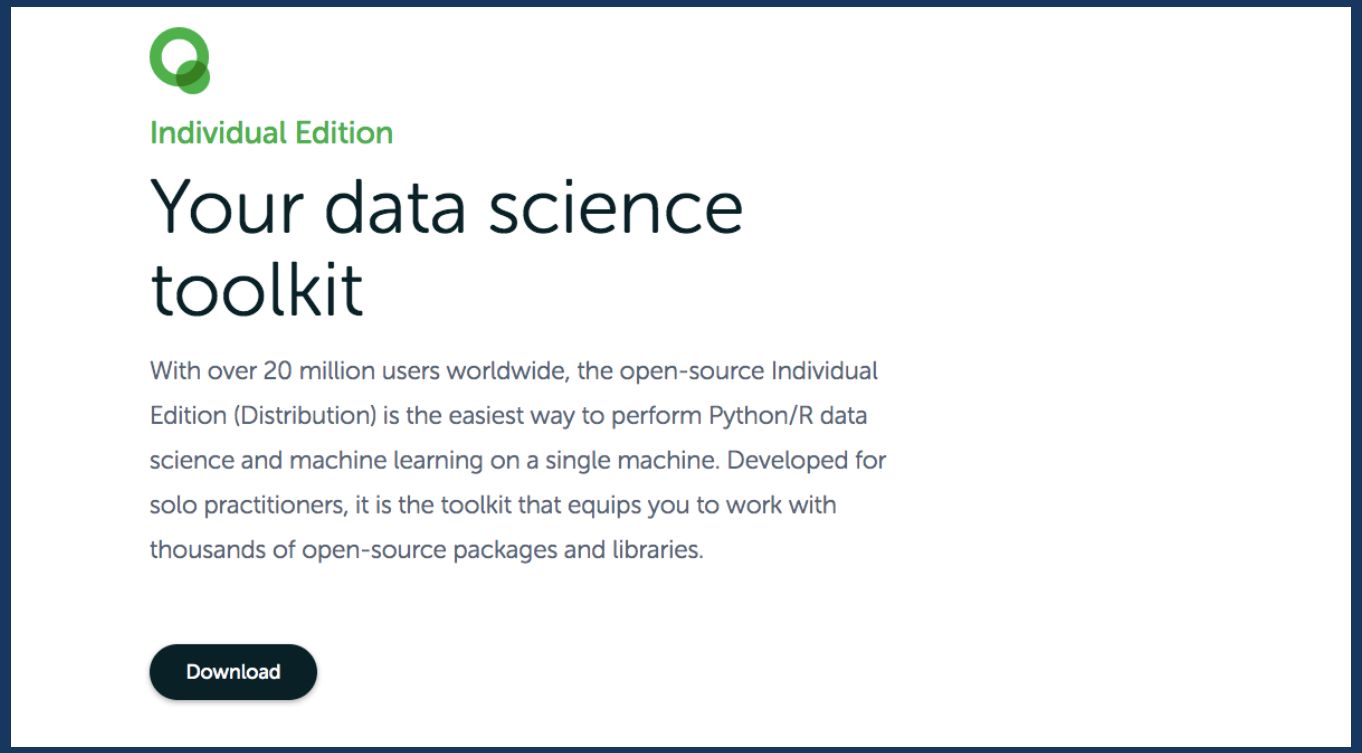
Site to install Python ->
https://www.jetbrains.com/pycharm/download/#section=mac



# Popular IDE for Python: Anaconda

greatlearning
Learning for Life

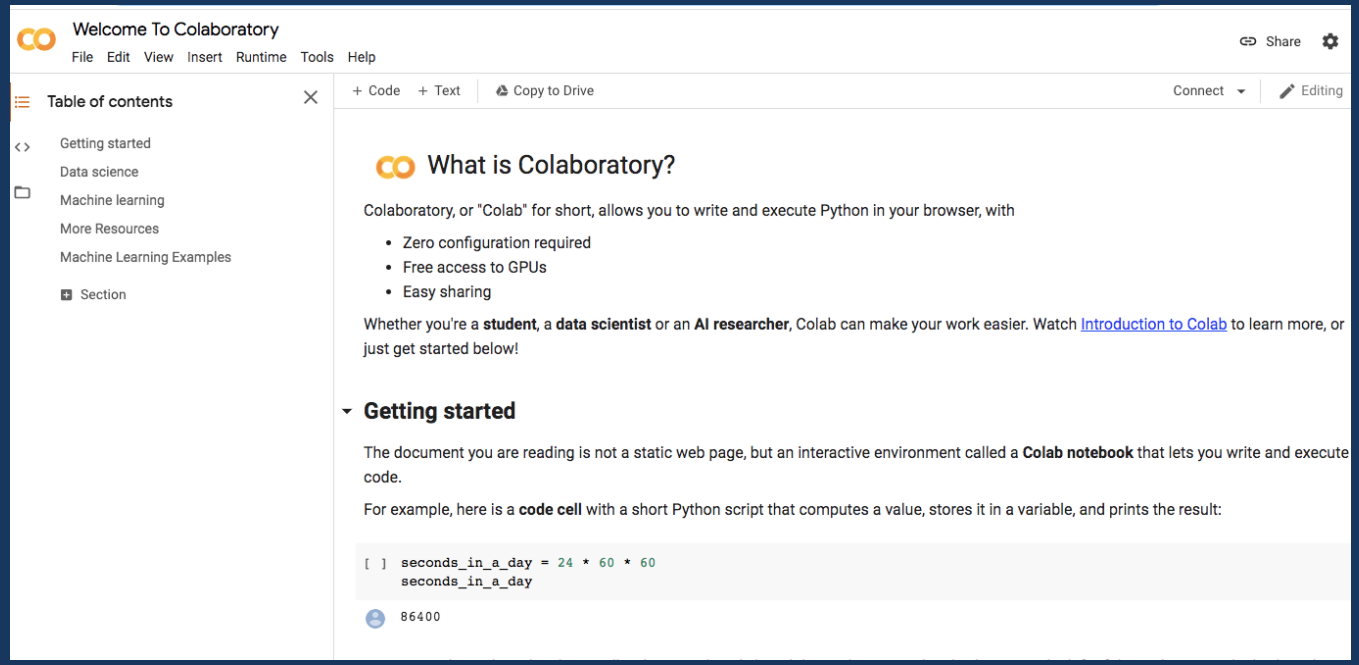
Anaconda installation site->
https://www.anaconda.com/products/individual



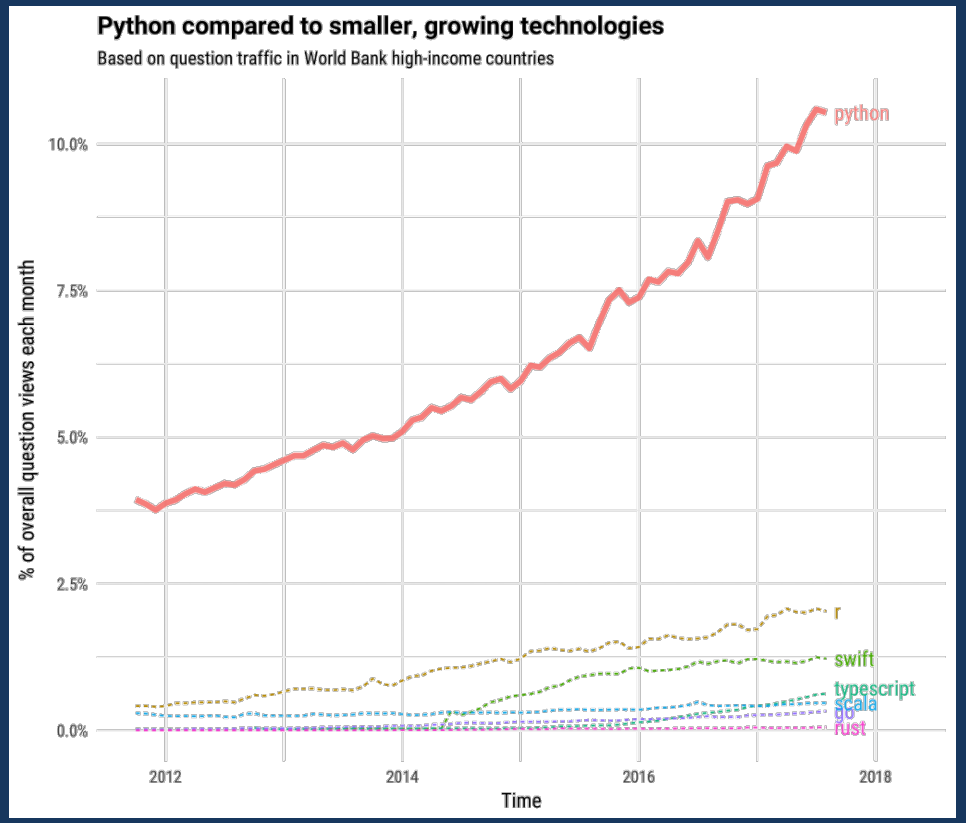
# Popular IDE for Python: Google colab

greatlearning
Learning for Life

Google collaboratory link->
https://colab.research.google.com/notebooks/intro.ipynb



# Statistical measurement on Python user




In recent time it is prominent that Python is one of the most popular language because of it's simplicity

greatlearning


# Uber Data Analysis

# Exploratory Data Analysis

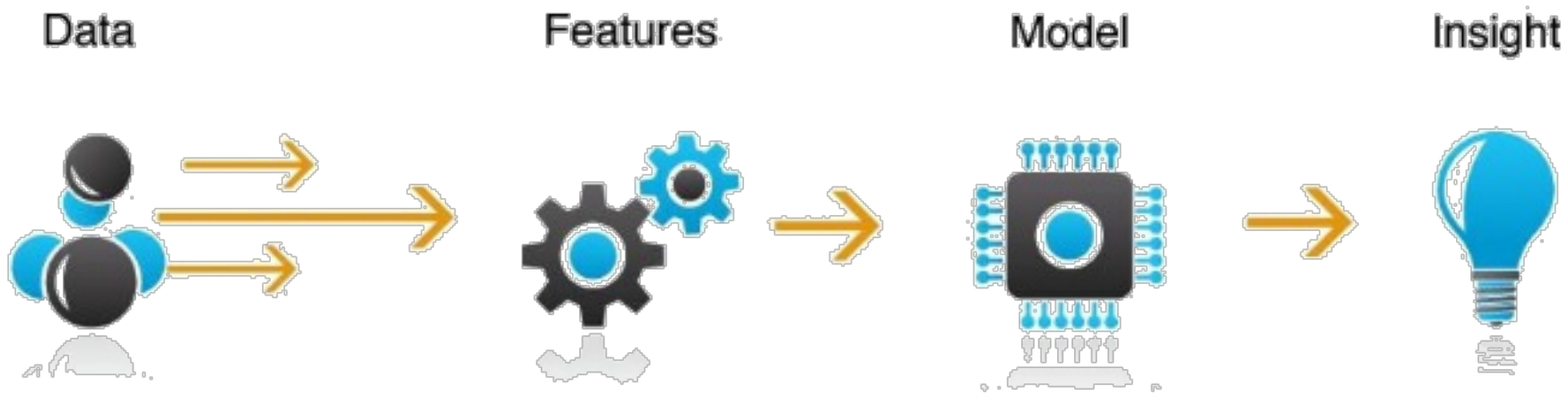
greatlearning

**Exploratory data analysis is basically a technique to understand your data using statistics concept**

# Feature Engineering

greatlearning

**Feature engineering process is basically used to create new feature from the existing data which helps to understand the data more deeply**



# What is Machine Learning?

greatlearning
Learning for Life

**Machine learning is a sub-set of artificial intelligence (AI) that allows the system to automatically learn and improve from experience without being explicitly programmed**

| | Time | V1 | V2 | V3 | V4 | V5 |
|---|---|---|---|---|---|---|
| **0** | 0.0 | -1.359807 | -0.072781 | 2.536347 | 1.378155 | -0.338321 |
| **1** | 0.0 | 1.191857 | 0.266151 | 0.166480 | 0.448154 | 0.060018 |
| **2** | 1.0 | -1.358354 | -1.340163 | 1.773209 | 0.379780 | -0.503198 |
| **3** | 1.0 | -0.966272 | -0.185226 | 1.792993 | -0.863291 | -0.010309 |
| **4** | 2.0 | -1.158233 | 0.877737 | 1.548718 | 0.403034 | -0.407193 |

**Training Data**

**Model Building**

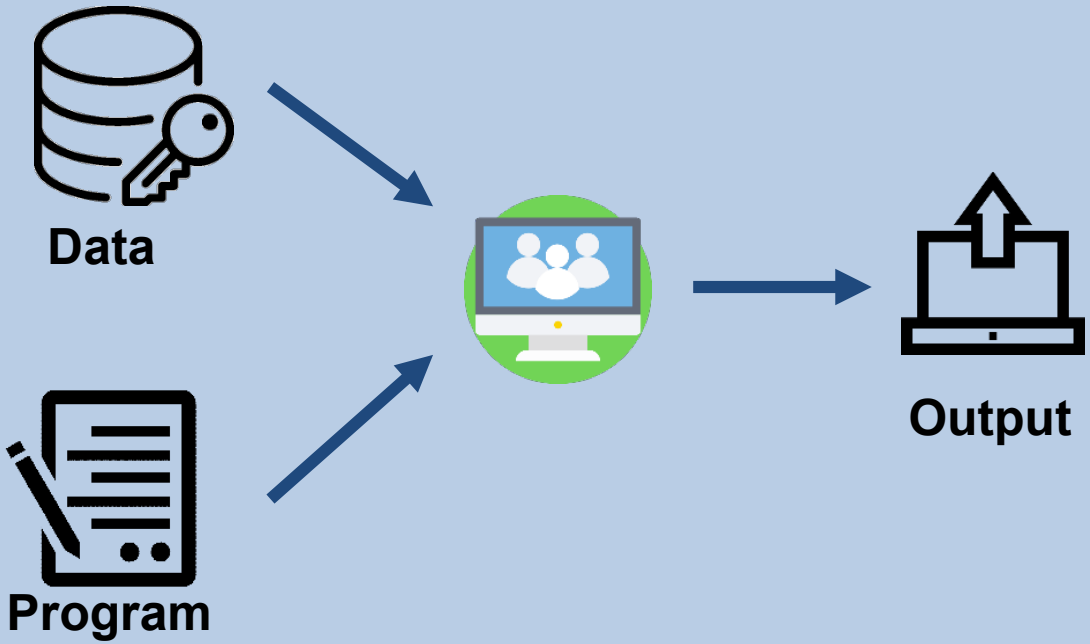
| | Time | V1 | V2 | V3 | V4 |
|---|---|---|---|---|---|
| **284802** | 172786.0 | -11.881118 | 10.071785 | -9.834783 | -2.066656 |
| **284803** | 172787.0 | -0.732789 | -0.055080 | 2.035030 | -0.738589 |
| **284804** | 172788.0 | 1.919565 | -0.301254 | -3.249640 | -0.557828 |
| **284805** | 172788.0 | -0.240440 | 0.530483 | 0.702510 | 0.689799 |
| **284806** | 172792.0 | -0.533413 | -0.189733 | 0.703337 | -0.506271 |

**Testing Data**

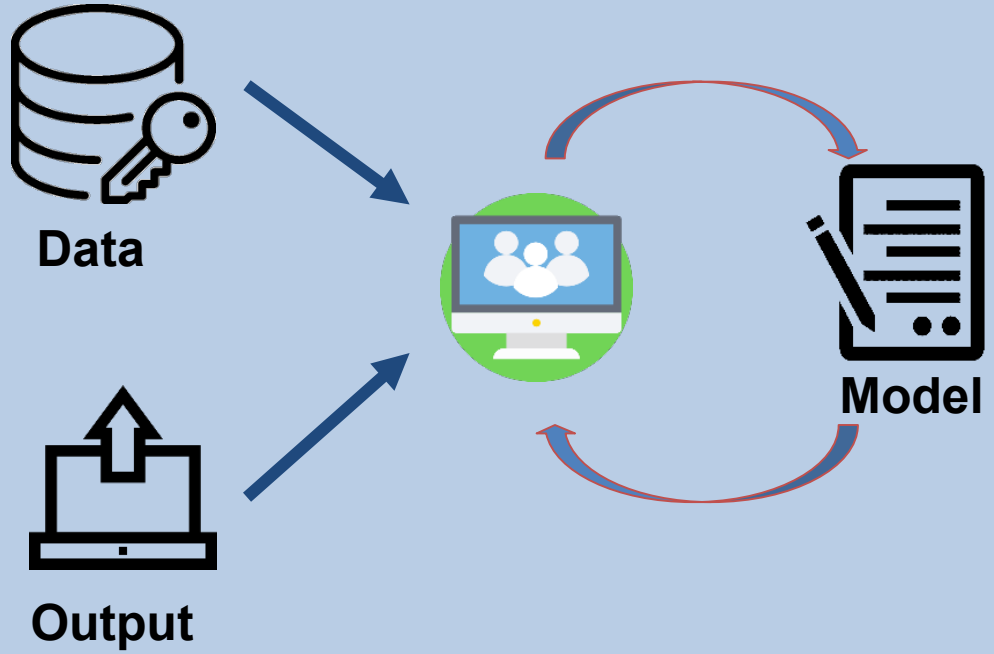
# Traditional Vs Machine Learning

greatlearning
Learning for Life

## Traditional Programming

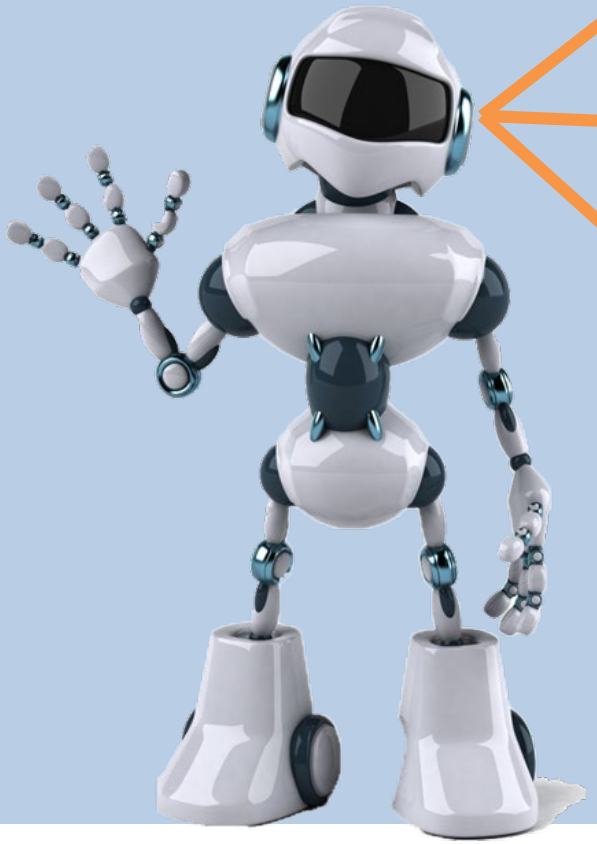



## Machine Learning

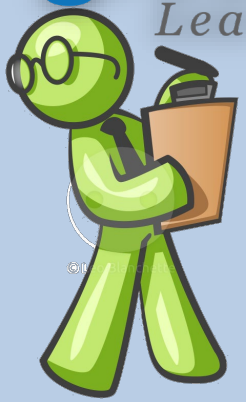


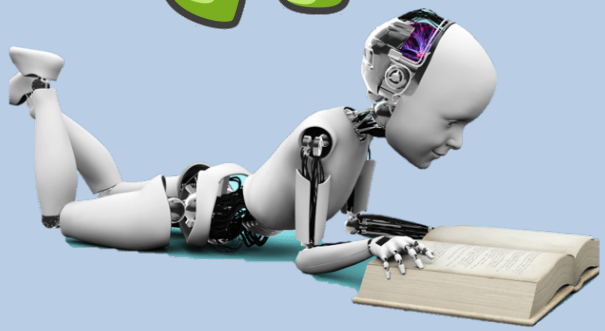
# Types Of Machine Learning

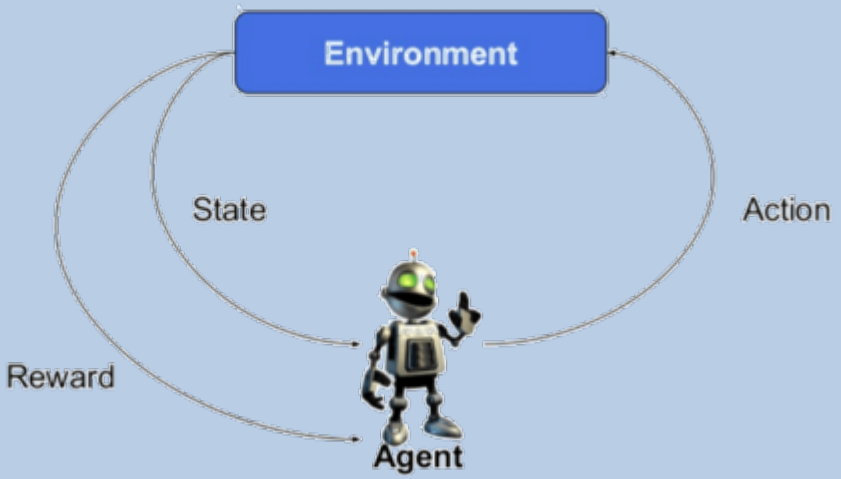
greatlearning
Learning for Life



- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning







# What is Supervised Learning?

Supervised learning works as a supervisor or teacher. Basically, In supervised learning, we teach or train the machine with labeled data (that means data is already tagged with some predefined class). Then we test our model with some unknown new set of data and predict the level for them

- Learning from the labelled data and applying the knowledge to predict the label of the new data(test data), is known as ***Supervised Learning***
- ***Types of Supervised Learning:***
  - **Linear Regression**
  - **Logistic regression**
  - **Decision Tree**
  - **Random Forest**
  - **Naïve Bayes Classifier**

# What is Linear Regression?

greatlearning
Learning for Life

Regression stands for to model a target value based on independent variables and Linear Regression is used to find the relationship between dependent(y) and independent variable(x)

- Linear regression is a supervised machine learning algorithm
- Always works with continuous value
- Formula: y = mx+c  m=slope of line and c= intercept
- Main target for linear regression to find the best value for X and Y

greatlearning
Learning for Life

# Thank You