In [ ]:

```
# 1. Business Problem
# 2. Data Acqisition
#    1. web servers
#    2. logs
#    3. data bases
#    4. APIs
#    5. Online Repositeries
# 3. Data Preperation
#    1. Data Cleaning
#        1. Inconsistent Data types
#        2. Misspelled Attributes
#        3. Missing and Duplicate Values
#    2. Data Transformation
#        1. talend
#        2. informatica
# 4. Exploratory Data Analysis
#    1. Defined Refines
#    2. The selection feature
#    3. Variables that will be used in the modle deployment
# 5. Data Modeling
#    KNN, decision tree, Navie bayes (Identify the model that best fit the business
#    Trains the models ont the training data set and test ---> select the best perfo
# 6. Visualization and Communication
#    (tools: tableu, powerBI, QlikView)
# 7. Deploy and Maintains
```

In [2]:

```
# There are Various roles offered to a daata scientist like
# 1. Data Analyst
# 2. Machine Learning Engineer
# 3. Deep Learning Engineer
# 4. Data Engineer
# 5. Data Scientist
```

In [4]:

```
# Top 5 Python libraries for Data Science
#1. Tensorflow
#  This library for high performance numerical computation and used across many scie
#  Basically it is a framework
#  Data in tensorflow are represented as tensors, which are multidimensional arrays

# Features of Tensorflow
#  1. Better computational graph visualizations
#  2. In neural machine translation, reduces error by 50-60%
#  3. parallel computing to execute complex models
#  4. Seamless library management

# Applications of Tensorflow
#  1. Speech / Image recognization
#  2. Text based applications
#  3. Time Series
#  4. Video detection
```

In [5]:

```
#2. Numpy
# it stands for Numerical Python
# it is used for general purpose array processing package
# Numpy is the funadamental package for numerical computation with python, it contai

# Features of Numpy
# 1. provides fast precompiled functions for numerical routines
```

```
# 2. Array oriented computing for better efficiency
# 3. Supports object-oriented approach
# 4. compact & faster computations with vectorization

# Appliations of Numpy
# 1. Extensively used in data analysis
# 2. Creating powerful N-Dimensional arrays
# 3. Forms the base of other libraries like scipy, scikit-learn
# 4. Replacement of MATLAb when used with scipy,matplotlib
```

In [6]:

```
#3. Scipy
#  it stands for scintific python
#  it is used for scintific & technical computation
#  it provides many user-friendly and efficient routines for scintific computation,

# Features of Scipy
# 1. A collection of mathematical algorithms and scintific functions built on the nu
# 2. High level commands and classes for manipulating and visualizing data
# 3. multi dimensional image processing with scipy, NDimage
# 4. Include functions for computing integrals numerically, solving diffenrential eq

# Applications of Scipy
# 1. Multi-Dimensional image operations
# 2. Solving differential equations & fourier transform
# 3. optimization algorithm
# 4. Linear algebra
```

In [7]:

```
#4. Pandas
# it stands for Python Data Analysis library
# it is used data analysis and cleaning
# it provides fast, flexible and expressive data structures designed to work with st

# Features of Pandas
# 1. Eloquent syntax and rich functionality
# 2. Apply()enables you to run a function across a series of data
# 3. High level abstraction
# 4. contains high level data structures & manipulation tools

# Applications of Pandas
# 1. General data wrangling
# 2. ETL jobs & data storage
# 3. used in a wide variety of academic and commercial domains, including statistics
# 4. Time-series specific functionality
```

In [ ]:

```
#5. Matplotlib
# it stands for plotting library for python
# Used for data visualization
# it provides and object oriented API for embedding plots into applications.

# Features of Matplotlib
# 1. As usable as matlabwith an advantage of being free and open source
# 2. Supports dozens of backends & output types
# 3. pandas itself can be used as wrappers around matplotlib's API
# 4. smaller memory consumtion & better runtime behaviour

# Applications of Matplotlib
# 1. corelation analysis of variables
# 2. Visualize 95% confidence intervals of the models
# 3. Outlier detection
# 4. Visualizing distributions to gain instant insights.
```