DAY 1 (6th June)

Two types of data

Qualitative data (videos, images)

Quantitative data(numbers)

From these two types we can further categorize data as

Structured

Unstructured

Semi structured

Python library

Pandas delas almost everything and data frames

Numpi deals with numbers

Jason and xml format is to be understood

NUMPI

Training data and testing data

Converts data into matrix

For data visualization we use following library

Matplotlib

seeborn

Exploratory data analysis

NUMPI

Pandas

Lambda expression

For python environment

Anaconda

Jupyton Nework

Data structure

Array

Dictionary

Linked list

DATA SCIENCE LIFECYCLE

OSEMN

Obtain

Data Collection and preparation

Scrub

Data cleaning

Explore

Select mutate filer join

Model

Train and test

Interpret

DATA SCIENCE DATA ANALYSIS DOMAIN

Descriptive Analysis

What happened

Diagnostic Analysis

Why did it happen

Predictive Analysis

Prescriptive Analysis

Data science majorly revolves around predictive and prescriptive analysis

R is preferred for data science as it provides more libraries for quantitate analysis

Stock market is time series Analysis

Raw data are not altered we duplicate the data. This is considered for accuracy

Data modeling has 3 algorithms

1. NB
2. KNN
3. SVM(Support Vector Machine)

DATA SOURCE

Kaggle.com (open an account)

Open data (for country related data)

This contains data related to country made public for the general citizens)

Research topics

Structured data VS unstructured data VS Semi structured data

SQL vs No SQL VS New SQL

Python operators and data type

Python libraries NUMPY and PANDAS

Import pandas as pd

Import numpy as np

If you are initial python uses the apply package installation ie pip install numpy, pandas

check data science Road Map