Data Analyticsz

- Pavan

Agenda

- Introduction
- Why Data Analytics
- What is data Analysis
- Course Overview
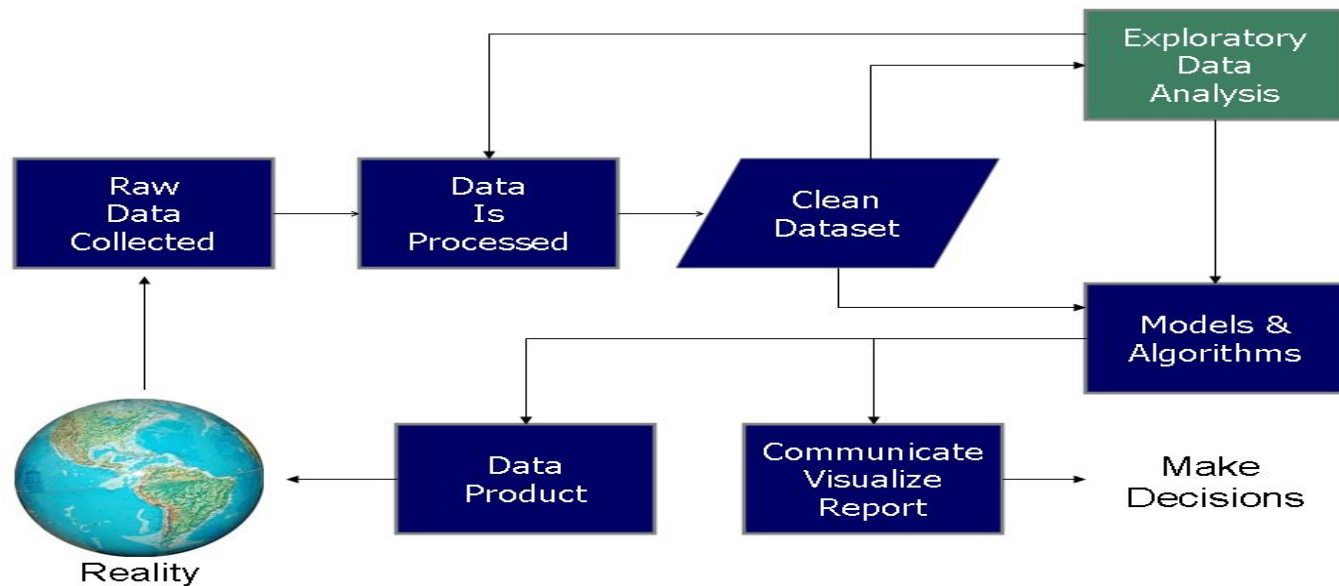- Q & A

Introduction

Data Analytics is the science of analyzing data to convert information to useful knowledge.

We are drowning in information and starving for knowledge - Rutherford D.Roger
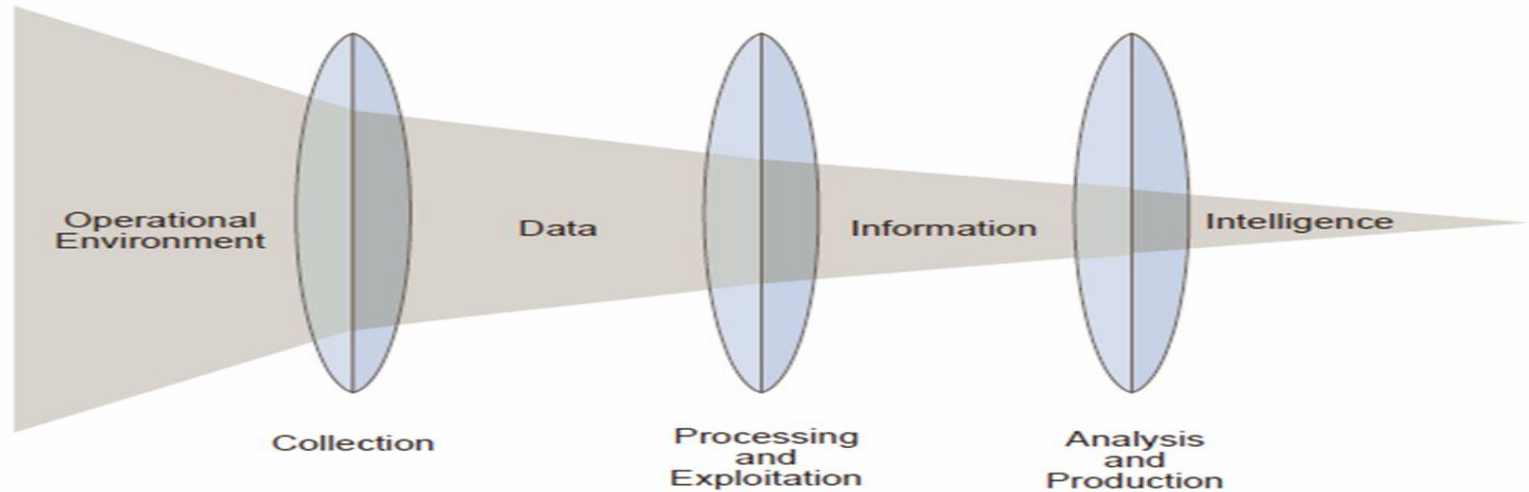
Data Analytics

Data Science Process

Why data analytics

- Companies that use data analytics drive better than companies who do not
    - Ex :Google vs Yahoo , Amazon vs Barnes and Noble
- Data Analytics drives results
    - Increased revenue
    - Decreases cost
    - Increases productivity
- Automate process

Relationship of Data, Information and Intelligence

Operational Environment → Collection → Data → Processing and Exploitation → Information → Analysis and Production → Intelligence

Examples

- Predict security threat
- Fraud detection
    - Geo-location , authorization, transaction etc
- Market pricing
    - House price, product cost
- Predictive support
    - Identify potential failure in system and detect before they recur
- Customer segmentation
    - Target customers based on segment they belong
- Recommendation system
    - Recommend useful products to customers (ex: Amazon)

Overview

- Data collection
- Data Preprocessing
- Cleaning
- Exploratory analysis
- Models (Data = Model + Error)
  - Estimation, prediction, Regression, Classification, clustering, Ensemble methods, decision trees etc.
- Supervised learning
- Clustering analysis
- reinforcement learning
- Data visualization
- Tools
  - Hands on with R/Python application

Tools

- [R](#) – a programming language and software environment for statistical computing and graphics.
- Python libraries : numpy, pands , scipy , scikitlearn