Diploma in Data Analytics

Basic Computer Skills:

Basic Computer Organization: Components – CPU, Primary (Random Access Memory, Read Only Memory) and secondary storage unit (Hard disk), Input and Output devices, Removable storage media

Introduction to OS and its functions, Windows operating system – booting, logging in/out of windows, restart/shut down a system, taskbar, desktop, folders – creating a new folder, moving around folder, renaming a folder, creating a new file, deleting/renaming a file, moving/copying a file to another folder, shortcuts, display properties, zip/unzip a file/folder, control panel and using it, searching for an app/file/settings in a computer system, administrative tasks – adding a new user/deleting an old user, change password, system settings, installing and removing software, anti-virus and its installation; File management through Windows Explorer; maximizing/minimizing/closing a window and moving around windows, command shell/prompt in windows

Basics of networking: Internet, IP addresses, networking cables and devices, world wide web, domain name server, what is proxy, what is firewall, browser and its functions, chrome browser and its setting, URL, using browser to open a webpage; working with email

Labs should introduce students not only to OS but office suite (word processor, spreadsheet, presentations)

Understanding the data:

Binary number system; Conversion from decimal to binary and vice versa, 2's complement for subtraction; octal and hexadecimal numbers;

Data types – integer, real number; fixed and floating point notation; maximum and minimum values of a numeric data in computer; precision and accuracy; understanding overflow and underflow; numerical errors;

Characters and strings; Attributes of the data, Continuous variables, Categorical variables, binning for transforming a continuous variable to categorical variable;

Vectors – addition, subtraction, scalar and vector product, unit vector; tuples; Matrices – addition, subtraction, transposition, multiplication, identity matrix, inverse; Rank of matrix, determinant, Eigen value Searching and sorting the data

Introduction to data Visualization: Histogram, box-plot, bar charts, line graphs, scatter plots.

Labs related to this course should emphasize on using spreadsheets for data sorting, data statistics, visualization

Introduction to Data Analytics:

Descriptive Statistics – graphical approaches, measures of central tendency, measures of dispersion

Data Statistics: First order statistics – mean; Rank ordered statistics - Maximum, minimum, quartile, median; Second order statistics – variance and standard deviation, introduction to higher order statistics – moments, skewness, kurtosis; correlation and covariance. Introduction to machine learning Associative rule mining, Big data Case studies – stock market, social analytics

Introduction to R Programming:

Basic data types in R – logical, integer, numeric, complex, character, raw ; Variables and operators; Reading and writing data; expressions

Decision making – if, if else, switch; Loop control statements – repeat, for, while, Vectors, matrices, arrays, strings, factors, data frames and lists

Functions in R, built-in functions, user defined functions, functions with arguments, default arguments, lazy evaluations

Creating bars and charts in R – pie charts, bar charts, scatter plots, line graphs, box plots, histogram

Statistics with R – mean, median, mode, variance calculations on data, linear/logistics/multiple regressions, correlatrion and covariance Generating probabilistic distributions with R

Probability/Statistics:

Counting; Random variables, distributions, quantiles, mean variance; Central limit theorem

Conditional probability, Bayes' theorem, base rate fallacy; Joint distributions, covariance, correlation, independence

Data Statistics: First order statistics – mean; Rank ordered statistics - Maximum, minimum, quartile, median; Second order statistics – variance and standard deviation, introduction to higher order statistics – moments, skewness, kurtosis; correlation and covariance. Bayesian inference with known priors, probability intervals, Bayesian inference with unknown priors Linear regression

Re-sampling methods: bootstrapping

1. The Visual Display of Quantitative Information (2nd Edition). E. Tufte. Graphics Press.
2. Envisioning Information. E. Tufte. Graphics Press.
3. Exploratory data analysis
4. Perception and interaction
5. Data-drive documents (D3) and their uses
6. Using space effectively
7. Animation, color and graph layout