



---

## Introduction to Data Science

### Duration- 2 Months

#### **Introduction**

What is Data Science? - Big Data and Data Science hype – and getting past the hype - Why now? – Datafication - Current landscape of perspectives - Skill sets needed.

#### **Introduction to Statistics**

Statistical Inference - Populations and samples - Statistical modeling, probability distributions, fitting a model - Intro to R.

#### **Data Analysis**

Exploratory Data Analysis and Data Science Process - Basic tools (plots, graphs and summary statistics) of EDA - Philosophy of EDA - The Data Science Process - Case Study: RealDirect (online real estate firm).

#### **Machine Learning**

Three Basic Machine Learning Algorithms - Linear Regression - k-Nearest Neighbors (k-NN) - k-means.

#### **Application of Machine Learning**

One More Machine Learning Algorithm and Usage in Applications - Motivating application: Filtering Spam - Why Linear Regression and k-NN are poor choices for Filtering Spam - Naive Bayes and why it works for Filtering Spam - Data Wrangling: APIs and other tools for scrapping the Web.

#### **Introduction to Feature**

Feature Generation and Feature Selection (Extracting Meaning From Data) - Motivating application: user (customer) retention - Feature Generation (brainstorming, role of domain expertise, and place for imagination) - Feature Selection algorithms – Filters; Wrappers; Decision Trees; Random Forests.

#### **Recommendation Systems**

Building a User-Facing Data Product - Algorithmic ingredients of a Recommendation Engine - Dimensionality Reduction - Singular Value Decomposition - Principal Component Analysis - Exercise: build your own recommendation system.

#### **Social-Network Graphs**

Mining Social-Network Graphs - Social networks as graphs - Clustering of graphs - Direct discovery of communities in graphs - Partitioning of graphs - Neighborhood properties in graphs.

#### **Data Visualization**

Data Visualization - Basic principles, ideas and tools for data visualization 3 - Examples of inspiring (industry) projects - Exercise: create your own visualization of a complex dataset.

#### **Data Science and Ethical Issues**