# The most comprehensive Data Science learning plan for 2017

## Overview

- A learning plan for Data Science is necessary to become a successful data scientist
- For beginners and transitioners, R, Python, basic of statistics, basic and advanced machine learning algorithms form the plan.
- For intermediate students, advanced machine learning algorithms, big data, deep learning and reinforcement learning are required to be understood
- Practicing with datasets and an online github profile are helpful in showcasing your skills

## Note  – Here is ["The Ultimate Learning Path to Becoming a Data Scientist in 2019"](#)

I joined Analytics Vidhya as an intern last summer. I had no clue what was in store for me. I had been following the blog for some time and liked the community, but did not know what to expect as an intern.

The initial few days were good – all the interns were smart, motivated and fun to be around. We played cricket in office, did internal hackathons over weekends and learnt a lot of data science. But, if there was one defining moment for me in the internship – it was when I realized the impact Analytics Vidhya was having in data science community.

I saw thousands of people following Analytics Vidhya religiously. I saw people looking up for guidance in our meetups and hackathons. I saw people transitioning their careers because of the resources we provide them. That is when this good internship transformed into a mind blowing experience.

That is the day I decided that this is my calling. It just felt that this is what I would want to do daily.

## Why create this learning path?

Among various resources on Analytics Vidhya, learning paths are special. The amount of effort and thinking they need is tremendous. The number of drafts they undergo is mind-boggling. But, the kind of impact they create for our audience is HUGE. That is why I decided that I will create a learning plan for 2017 for all our followers.

We created a similar [plan for 2016](#) and we saw [transitions happening by people following](#) this learning plan. This time we have created a much granular and a more detailed learning plan. The sole aim behind creating this comprehensive plan is to create a much bigger impact for our followers this year.

# Who should use this learning path?

This learning path would be extremely useful for any one who wants to learn machine learning, deep learning or data science in this year. If you plan to wait for a year, we will publish something similar in 2018 as well 🙂

But, for the people looking for action this year, this framework and plan of action should be extremely useful. Whether you are a complete fresher or a transitioner or you are looking to up-skill yourself, this plan should give you the necessary direction.

We published a similar [plan in 2016](#) and we saw followers making transition by simply following the plan. This year's plan is more nuanced than last year's one – so if you plan to pick up / improve data science skills – this plan will guide you through the journey.

# How can you use this learning path?

In creating this plan, we have removed the confusion from the process of learning. The biggest challenge which people face while learning is not dearth of learning material – but too much of it. You are not sure where to start learning, what to practice, how much time to spend on a concept, where to get the useful resources etc. For most of the beginners, this becomes overwhelming and they simply drop out before even learning a single skill.

This plan takes this confusion out. This path contains both theoretical resources as well practical examples. We have also provided you with resources / tests to apply your learning and benchmark yourself. As part of this plan, you will apply the concepts you learn on real-world problems and gain hands-on experience.

# Table of Contents

# 1. A few definitions before we start

The first thing you need to do is identify which kind of learner are you. Have a look at the definitions / descriptions below and identify which category you belong to.

- **Who is a beginner data scientist?**
  - A beginner has no prior experience in data science or machine learning

- Does not know any analytical tool or languages like R, SAS or Python
- No prior knowledge of subjects like mathematics & statistics.
- A person who has prior exposure to some of the sections in this article like probability, linear algebra can feel free to skip the initial sections of the learning path to pace up their learning.

- **Who is a transitioner data scientist?**
  - A transitioner has no prior experience in any of the analytics tools like R/Python
  - Does not know Machine Learning concepts etc and
  - Has work experience more than 3 years in industry other than Analytics.
  - A person who has prior exposure to some of the sections in this article like probability, linear algebra can feel free to skip the appropriate sections of the learning path and pace up their learning.

- **Who is an Intermediate data scientist?**
  - People who already know Data Science, are comfortable with building predictive Machine Learning models
  - They participate in Data Science competitions and hackathons on a regular basis.
  - Prior knowledge of Basic and Advanced Machine Learning algorithms is necessary.
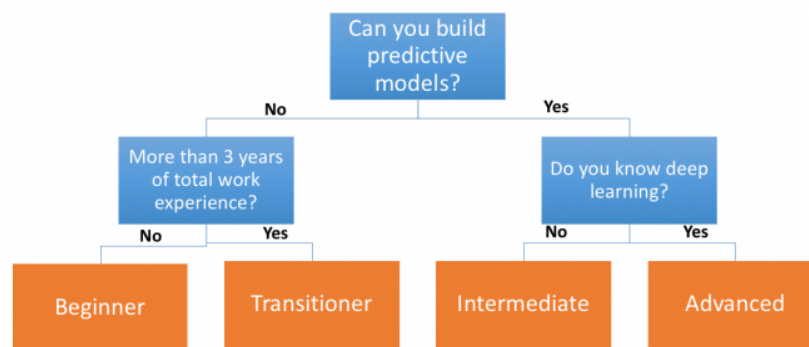


# 2. Setting target and timelines for yourself

We have created these guides with the following target in mind:

- **Beginner Data Scientist**
  - Learn basic mathematics and statistics required for data science
  - Develop a basic understanding of machine learning algorithms and solving real life problems from them
  - Skills required to land you first data science internship / job.
  - Time spent ~ 3 hours / day

- **Transitioner Data Scientist**
  - Learn basic mathematics and statistics required for data science
  - Develop a basic understanding of machine learning algorithms
  - Work on projects and create a portfolio of projects
  - Skills required to land your first data science internship / job.
  - Time spent ~ 5 hours / day

- **Intermediate Data Scientist**
  - Understand Deep Learning techniques and algorithms to the extent of applying them on real world problems.
  - Learn to create awesome Interactive Visualizations and improve your story telling capabilities.
  - Understanding of recent development (Reinforcement Learning) in the field of Data Science and incorporate them into the existing Machine Learning frameworks.
  - Web Frameworks and cloud computing to create independent Data / machine learning products.
  - Time spent ~ 3 hours / day

# 3. Ultimate Beginner's path for 2017

**Structure for your 2017 journey:**

- [Step 1: Getting started and testing the waters](#)
- [Step 2: Mathematics & Statistics](#)
- [Step 3: Introducing the tool – R / Python](#)
- [Step 4: Basic & Advanced machine learning tools](#)
- [Step 5: Building your profile](#)
- [Step 6: Applying for Jobs / Internships](#)



Beginner's Plan to become a data scientist - 2017
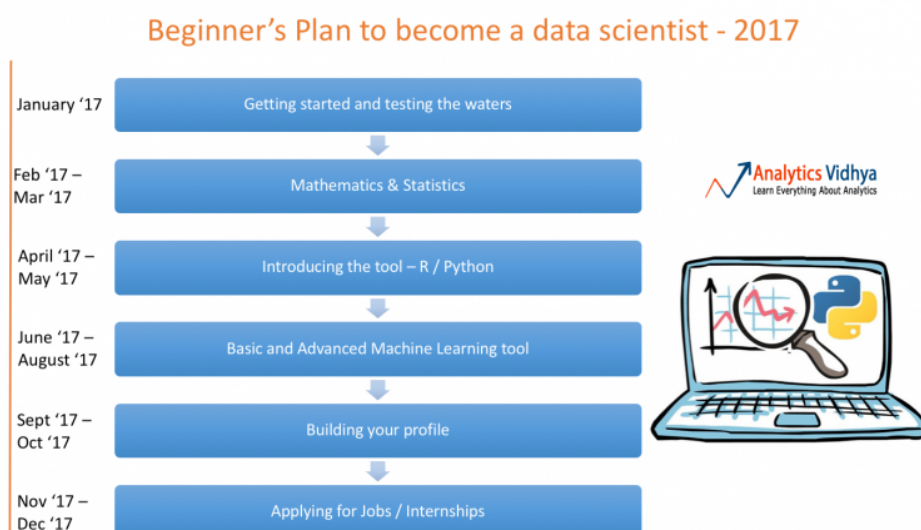
# 3.1: Getting Started and testing the waters

**Time suggested: 4 weeks (January 2017)**

At this stage, it is important to understand why you want to become a data scientist? What are your strengths and weaknesses? Do you know what it takes to be a Data Scientist? You must answers these questions before jumping on the boat of Data Science journey.

Watch [this excellent video](#) where **Tetiana Ivanova** describes how she became a Data Scientist without going through a Masters or doctorate program in data science and with help of Meetups.

Here are some additional resources you can use to answer these questions:

1. [What is Data Science?](#) – This article by Data Jobs will give you a broad perspective of how data science is being used in Netflix and Amazon. Also, it will highlight the skill set required for Data Science.
2. [Should I become a Data Scientist?](#) This article points out some questions for you to decide whether you are fit for a Data Scientist role. I suggest you must go through this article before proceeding further.
3. Next, you should attend local meetups in your area. Go out and find out what people are talking about Data Science / Machine Learning. Meetups not only help you learn the tools and techniques, they provide you with a network of people in similar industry which helps you in finding the right jobs and internships later on.

Go ahead and think through these aspects of choosing a career in data science. This decision is going to decide the next 11 months of your life.

# 3.2: Basics of Mathematics and Statistics

**Time suggested: 8 weeks (February 2017 – March 2017)**

Topics to be covered:

- Descriptive Statistics – 1 week
- Probability – 2 weeks
- Inferential Statistics – 2 weeks
- Linear Algebra – 1 week
- Structured Thinking – 2 weeks

## Descriptive Statistics – 1 week

- **Course (mandatory) –** [Descriptive Statistics from Udacity](#) is a basic and must do course to get started.
- **Books (optional)** – Supplement your online course with [online stats book](#). A good book for any one looking for learning basic statistics.

# Probability – 2 weeks

- **Course (mandatory)** – [Introduction to probability – The science of uncertainty](#) is an excellent course on edX to learn concepts of probability like conditional probability and probability distributions.

- **Books (optional)** – The textbook [Introduction to probability – Berkley's stats 134 standard textbook](#) will supplement the course above and can be used as a good reference material.

# Inferential Statistics – 2 weeks

- **Course (mandatory)** – [Intro to Inferential Statistics from Udacity](#) – Once you have gone through the descriptive statistics course, this course will take you through statistical modeling techniques and advanced statistics.

- **Books (optional)** – [Online Stats Book](#) – This online book can be used for a quick reference for inference tasks.

# Linear Algebra – 1 week

- **Course (mandatory)**
  - [Linear Algebra – Khan Academy](#) : This concise and an excellent course on Khan Academy will equip you with the skills necessary for Data Science and Machine Learning.
- **Books (optional)**
  - [Linear Algebra/ Levandosky](#) – This is an often cited book to Stanford graduates for Linear Algebra.
  - [The Manga guide to Linear Algebra](#) – This is a fun filled Linear Algebra book which keeps Machine Learning in context. You will never forget these Algebra lessons for sure.

# Structured Thinking – 2 weeks

- **Articles (mandatory):** These articles will guide you to structure your thinking process to approach problems in a better way so as to improve your efficiency.
  - [How to train your mind for analytical thinking?](#)
  - [Tools for improving structured thinking](#)
  - [The art of structured thinking and analyzing](#)

- **Competitions (mandatory):** No amount of theory can beat practice. This is a [strategic thinking problem](#) which will test you on your thinking process. Also, keep an eye on business case studies as

they help in structuring your thoughts tremendously.

# 3.3: Introducing the tool – R / Python

**Time suggested: 8 weeks (April 2017 – May 2017)**

**Topics to be covered:**

- Tools (R/Python) – 4 weeks
- Exploration and Visualization (R/Python) – 4 weeks
- Feature Selection/ Engineering

## Tools

## 1. R

- **Course** – [Interactive Intro to R Programming Language by DataCamp](#) – An excellent course by DataCamp to give you hands-on experience in R. The course includes interactive examples You will never feel bored while learning R.

- **Books** – [R for Data Science](#) – This is your one stop solution for referencing basic materials on R.

- **Blogs/Articles**
    - [This](#) article will serve a great point for collating the entire process of model building starting from installation of RStudio/R.
    - [R-bloggers](#) – This is one of the most recommended blog for R- users. Every R practitioner should keep this blog bookmarked. It has some of the most effective and practical R tutorials. Bookmark it now.

## 2. Python

- **Course (mandatory)** – [Intro to Python for Data Science](#) – An interactive course developed by DataCamp to facilitate Data Science learning using Python.

- **Books (mandatory)** – [Python for Data Analysis](#) – This book covers various aspects of Data Science including loading data to manipulating, processing, cleaning and visualizing data. Must keep reference guide for Pandas users.

- **Blogs/Articles (optional)**
    - [A Complete Tutorial to Learn Data Science with Python from Scratch](#): This article will serve as a quick guide to learning Data Science using Python.

# Exploration and Visualization

## 1. R

- **Course**
  - [Exploratory Data Analysis](#) – This is an awesome course by Johns Hopkins University on Coursera. You will need no other course to perform visualization and exploratory work in R.

- **Blogs/Articles**
  - [Comprehensive guide to Data Exploration in R](#) – This will be a one-stop article that I will suggest you to go through carefully and follow every step. This is because the steps mentioned in the article are the same steps you will be using while solving any data problem or a hackathon problem.
  - [Cheat sheet – Data Exploration in R](#) – This cheat sheet contains all the steps in data exploration with codes. I suggest you to take out a print and paste it on your wall for quick reference.

## 2. Python

- **Course (optional)**
  - [Intro to Data Analysis](#) – This is an excellent course by Udacity on Data Exploration using Numpy and Pandas.

- **Blogs/Articles (mandatory)**
  - [Comprehensive guide to Data Exploration using Python NumPy, Matplotlib and Pandas](#) – This is a sufficient and comprehensive article which uses the most popular Python libraries for exploration and visualization purposes.
  - [9 popular ways to perform Data Visualization in Python](#) – This article presents the most commonly used graphs and plots used in Data Exploration along with Python codes. This is a must bookmarked article for people working in Data Science using Python.

- **Books (optional)** – [Python for Data Analysis](#) – A one stop solution for your Data Exploration and Visualization in Python.

# Feature Selection/ Engineering

- **Blog** – [A Comprehensive Guide to Data Exploration](#): This article will explain underlying techniques of feature engineering and different methods for feature creation

- **Books (optional)** – [Mastering Feature Engineering](#): This book is master piece to learn feature engineering. Not only will you learn how to implement feature engineering in a systematic way. You will also learn different methods involved in feature engineering.

# 3.4: Basic & Advanced machine learning tools

**Time suggested: 12 weeks (June 2017 – August 2017)**

Topics to be covered (June 2017 – July 2017):

- Basic Machine Learning Algorithms.
  - Linear Regression
  - Logistic Regression
  - Decision Trees
  - KNN (K- Nearest Neighbours)
  - K-Means
  - Naïve Bayes
  - Dimensionality Reduction
- Advanced algorithms (August 2017)
  - Random Forests
  - Dimensionality Reduction Techniques
  - Support Vector Machines
  - Gradient Boosting Machines
  - XGBOOST

## Linear Regression

- **Course**
  - [Machine Learning by Andrew Ng](#) – There is no better resource to learn Linear Regression than this course. It will give you a thorough understanding of linear regression and there is a reason why Andrew Ng is considered the rockstar of Machine Learning.

- **Blogs/Articles**
  - [This](#) lesson out of PennState Stat 501 course outlines the main features of Linear Regression ranging from a simple definition of a Linear Regression to determining the goodness of fit of a regression line.
  - [This](#) is an excellent article with practical examples to explain Linear Regression with code.

- **Books**
  - [The Elements of Statistical Learning](#) – This book is sometimes considered the holy grail of Machine Learning and Data Science. It explains Machine Learning concepts mathematically from a Statistics perspective.
  - [Machine Learning with R](#) – This is a book I personally use to have a brief understanding of Machine Learning algorithms along with their implementation code.

- **Practice**
  - [Black Friday](#) – Like I already said – No amount of theory can beat practice. Here is a regression problem that you can try your hands on for a deeper understanding.

# Logistic Regression

- **Course (mandatory)**
  - [Machine Learning by Andrew Ng](#)– The week 3 of this course will give you a deeper understanding of the one of the most widely used classification algorithm.
  - [Machine Learning: Classification](#) – Week 1 and 2 of this practical oriented Specialization course using Python will satiate your knowledge thirst about Logistic Regression.

- **Blogs/Articles (optional)**
  - [Logistic Regression by Machine Learning Mastery](#) – This is an excellent non-code based approach to Logistic regression to deepen your knowledge. I suggest you to have a look at it.

- **Books (optional)**
  - [Introduction to Statistical Learning](#) – This is an excellent book with a quality content on Logistic Regression's underlying assumptions, statistical nature and mathematical linkage.

- **Practice (mandatory)**
  - [Loan Prediction](#) – This is an excellent competition to practice and test your new Logistic Regression skills to predict whether loan status for a person was approved or not.

# Decision Trees

- **Course (mandatory)**
  - [Machine Learning: Classification](#) – Week 3 and 4 in this course is about the working of decision trees, preventing overfitting and handling missing values

- **Blogs/Articles (mandatory)**
  - [Technical Overview of decision trees](#) – This is a quick overview of decision trees and a must read for anyone new to decision trees.
  - [Complete tutorial on tree based modeling](#) – This is a python based tutorial on decision trees. For the sake of decision trees, read only sections 1-6 in this article.

- **Books (mandatory)**
  - [Introduction to Statistical Learning](#) – Section 8.1 and 8.3 explain the basics of decision trees through theory and practical examples.
  - [Machine Learning with R](#) – Chapter 5 of this book provides you the best explanation of Machine Learning Algorithms available in the market. Here, the decision trees are explained in an extremely non-intimidating and easier style.

- **Practice (mandatory)**
  - [Loan Prediction](#) – This is an excellent competition to practice and test your new Logistic Regression skills to predict whether loan status for a person was approved or not.

# KNN (K- Nearest Neighbors)

- **Course (mandatory)**
  - [Machine Learning – Clustering and Retrieval](): Week 2 of this course progresses to k-nearest neighbors from 1-nearest neighbor and also describes the best ways to approximate the nearest neighbors. It explains all the concepts of KNN using python.

- **Blogs/Articles (mandatory)**
  - [Introduction to k-nearest neighbors: simplified]() – This basic article describes when to use KNN, the ways in which k can be chosen and the way in which KNN algorithm works.
  - [Learning KNN algorithm using R]() – This article is a comprehensive guide to learning KNN with hands-on codes for future references.

# K-Means

- **Course**
  - [Machine Learning Course – Unsupervised Learning with K-means algorithm](): Week 8 of this discusses how to use course how K-means algorithm is used for handling unstructured data.

- **Blog**
  - [An Introduction to Clustering and different methods of clustering](): In this article, you will learn what is k-means clustering and the intricacies involved in that. It will give you a step by step approach how K-means algorithm works.

# Naive Bayes

- **Course**
  - [Intro to Machine Learning](): Take this course to see Naive Bayes in action. In this course, Sebastian Thrun has explained Naive Bayes in Simple English.

- **Blog / Article**
  - [6 Easy Steps to Learn Naive Bayes Algorithm (with code in Python)]() : This article will take you through Naive Bayes algorithm in detail. In this guide, you will learn how Naive Bayes algorithm works, applications and many more. It will also give you hands-on knowledge of building a model using Naive Bayes.
  - [Naive Bayes for Machine Learning]() : This is one of the most comprehensive articles I have come across. Go through this article to have a complete understanding of why naive bayes algorithm is important for machine learning.

# Dimensionality Reduction

- **Course**
  - [Machine Learning – Dimensionality Reduction:](#) Week 8 of this course will walk you through dimensionality reduction and how Principal Components Analysis can be used for data compression of complex data.

- **Blog / Article**
  - [Beginners Guide To Learn Dimension Reduction Techniques:](#) In this article, you will learn why dimension reduction is important in machine learning and the various techniques of dimension reduction.

# Random Forests

- **Videos (mandatory)**
  - [How Random Forest algorithm works?](#) – Watch this video to have a visual perspective of how the Random Forest algorithm works.

- **Books (optional)**
  - [Introduction to Statistical Learning](#) – Section 8 explains the basics of Random Forests including bagging and boosting through theory and practical examples.
  - [Applied predictive modeling](#) – Chapter 8

- **Blogs/Articles (mandatory)**
  - [A tutorial on tree based modeling from scratch](#) – This is an excellent article on trees based modeling using python. I suggest you to bookmark it right now.
  - [Random Forests](#) – This blog explains the entire working, nuts and bolts of Random Forest.

# Gradient Boosting Machines

- **Blogs/Articles (mandatory)**
  - [Guide on Boosting methods](#)
  - [Parameter tuning GBM](#)
  - [Machine Learning Mastery- GBM](#)

- **Presentation (mandatory):** [Here](#) is an excellent presentation on GBM. It contains the prominent features of GBM and the advantages and disadvantages of using it to solve real-world problems. It is must see article for somebody trying to understand GBM.

# XGBOOST

- **Blogs /Articles (mandatory)**
  - [Official Introduction XGBOOST](#) – Read the documentation of hackathons winning algorithm. It is an improvement over GBM and is right now the most widely used algorithm for winning competitions.
  - [Using XGBOOST in R](#) – An excellent article on deploying XGBOOST in R using a practical problem at hand.
  - [XGBOOST for applied Machine Learning](#) – An article by Machine Learning Mastery to evaluate the performance of XGBOOST over other algorithms.

# Support Vector Machines

- **Course (mandatory)**
  - [Machine Learning by Andrew Ng](#) – Week 7 of this course is an interesting place to start your SVM journey.

- **Books (mandatory)**
  - [Introduction to Statistical Learning](#) – Chapter 9 of the book contains a detail discussion about SVMs and the ways to deploy them.

- **Blogs/Articles (optional)**
  - [Understanding support vector machines](#) – This is an excellent article to understand an algorithm practically using examples.
  - [SVM by Machine Learning Mastery](#) – This article discusses the different types of kernels employed in SVM and their uses.

# 3.5: Building your profile

**Time suggested: 8 weeks (September 2017 – October 2017)**

Topics to be covered:

1. GitHub Profile Building
2. Practice via competitions
3. Discussion Portals

# GitHub Profile Building (mandatory)

It is very important for a Data Scientist to have a GitHub profile to host all the codes of the project he/she has undertaken. Potential employers not only see what you have done, how you have coded and how

frequently / how long you have been practicing data science.

Also, codes on GitHub open up avenues for open source projects which can highly boost your learning. If you don't know how to use Git, you can learn from Git and GitHub on Udacity. This is one of the best and easy to learn course to manage the repositories through terminal.

## Practice via competitions (mandatory)

Time and again, I have stressed on the fact that practice beats theory. Moreover coding in hackathons brings you closer to developing data products in real life for solving real world problems. Below are most popular platforms to participate in Data Science/ Machine Learning Competitions.

1. Analytics Vidhya Datahack
2. Kaggle competitions
3. Crowd Analytix human layer

## Discussion Forums (optional)

Discussions are a great way to learn in a peer-to-peer setup from finding an answer to a question you stuck to providing answers to someone else's questions. Below are some of the discussion rich platforms which you should keep a tab on to clear your doubts.

1. Analytics Vidhya Discussion Portal
2. Kaggle Discussion
3. StackExchange

## 3.6: Apply for Jobs & Internships

**Time suggested: 8 weeks (November 2017 – December 2017)**

**Topics to be covered: Jobs / Internships**

If you are here after diligently following the above steps, then you can be sure that you are ready for a Job / Internship position at any Data Science / Analytics or Machine Learning firms. But it becomes quite difficult to identify the right jobs. So, for the purpose of saving the trouble, I have created a list of portals which lists down Data Science/ Machine Learning jobs and Internships.

1. Analytics Vidhya Job Portal
2. Datajobs
3. Kaggle Job portal
4. Internshala

In order to prepare for these interviews, you should go through this Damn Good Hiring Guide

# 4. Transitioner's path for 2017

Let me start by giving you the bad news – it is not going to be easy to transition in data science. Also, the more your work experience, the more difficult your transition would typically be. You would need a strong resolve – there will be times when you might question, whether this is the right domain for you.

The good news is that once you get your first break in the industry, there is no looking back. Also, because of the salary differential from other industry, you may not need to compromise on your earnings during transition.

To achieve your goal all you have to do is follow this learning path diligently. We have covered all the skills, techniques you need to gain to take your first steps in data science.

## The Ultimate Path for transitioners

Simply put, if you are looking for a transition under a year, you will need to learn everything we laid out for the beginner above. Additionally, you will need to carve out additional time to showcase your skills. You will need to overcome the doubts of your potential employers through your projects and work.

I am sure you are beginning to understand why transition is not an easy thing.

**Structure for your 2017 journey:**

The structure of the path is similar, but you will need to accelerate your learning in the first half of the plan. Start by going through this article and go through a few success stories to understand what a transition would entail. Once you are set for the journey, follow the plan by sticking to these timelines.

- Step 1: Getting started and testing the waters (1 week in January '17)
- Step 2: Mathematics & Statistics (Jan '17 – March '17)
- Step 3: Introducing the tool – R / Python (March '17 – April '17)
- Step 4: Basic & Advanced machine learning tools (May '17 – July '17)
- Step 5: Building your profile (Aug '17 – Oct '17)
- Step 6: Applying for Jobs  (Nov '17 – Dec '17)

## 5. Intermediate's path for 2017

If you can build predictive models, but don't necessary know deep learning and some recent development in the domain, this learning path can help you out. Depending on your skills and learning plan for the year, you can pick and choose the areas you want to learn.

**Structure of intermediate path for 2017:**

- Step 1: Assessing your technical & Structured thinking skills

## 5.1: Assess your technical & structured thinking skills – Jan 2017

The first step in creating your learning plan is to benchmark yourself on various skills – both technical and structured thinking. You can go through the [skill tests on Analytics Vidhya](#) to judge whether you need to review the old material. If you do well, go ahead with acquiring new skills. Else, go back to practice for some more time.

If you feel the need to go through the old material once again, refer to beginner's path which contains various useful resources.

**Skill tests:**

1. [Statistics 1](#) & [Statistics 2](#)
2. [R for Data Science](#)
3. [Python for Data Science](#)
4. [Machine Learning](#)
5. [Regression](#)
6. [Tree-based algorithms](#)
7. [SQL](#)

## Structured Thinking

- **Articles (mandatory) –** These articles will guide you to structure your thinking to solve business problems efficiently.
  - [How to train your mind for analytical thinking?](#)
  - [Tools for improving structured thinking](#)
  - [The art of structured thinking and analyzing](#)

- **Competitions (mandatory):** Check out [strategic thinking problem](#) to test your structured thinking. Also, keep an eye on business case studies as they help in structuring your thought process.

## 5.2: Few more ML algorithms – Feb 2017

There are a few specific machine learning algorithms, which come in handy while solving specific problems. For example, try solving online click prediction on large data sets with out applying online learning algorithms and you would know what I am talking about. Here are a few advanced ML algorithms you should learn this month:

**Online Machine Learning**

- Course: [Online Methods In Machine Learning by MIT](#)
- Books:
    - [Understanding Machine Learning: From Theory to Algorithms](#)
    - [Online Learning and Online Convex Optimization](#)
- Blogs : [Langford's hunch.net](#)

**Vowpal Wabbit**

- Tutorial : [By the creator of Vowpal Wabbit – John Langford](#)

**FTRL- Algorithms**

- Course: [Online Learning by University of Washington](#)
- Paper : [Google's ad click prediction](#)

**Exercise:** Practice on one of the old Kaggle competitions or open click through rate data sets as provided by Criteo.

## 5.3: Pick up a data visualization tool (March 2017)

Ideally you should pick up D3.js for sure and either one of QlikView and Tableau. While D3.js provides the most flexibility, QlikView and Tableau are both handy for creating dashboards or less complex story creation and narration.

**Topics to be covered:**

- Interactive Visualization using d3.js (3 weeks)
- Creating Visualizations in QlikView (1 week)
- Creating Visualizations in Tableau (1 week)

### Interactive Visualization using d3.js

The reason d3.js is not so much popular among Data Scientist is because it requires an entire different skill test like HTML, CSS, Javascript which is not typical of a Data Scientist.

But knowing D3.js can take your story telling capabilities to a different level. You can create non-static Interactive graphs embedded right in a browser for a much richer experience. Below are the list of

resources to master d3.js

- **Course** [Data Visualization and d3.js](#) : This is an excellent course provided by Zipfian experts on Udacity and a part of Facebook's Data Analyst Nanodegree program.

- **Books** [Interactive Data Visualization for the Web](#) – An excellent book by Scott Murray and your one stop reference material. It has a web version which is free to use.

- **Code-Oriented Resource** [Dashing d3.js](#) – This is a code oriented tutorial which will help you create your Interactive Visualizations. This is also the same tutorial I am currently undergoing to learn d3.js

- **Blogs/Articles** [Complete path from being a noobie to an expert at d3.js](#) – This was the original article which got me started into learning d3.js. It contains a list of resources as well as codes for some basic graph elements which you can always refer back to.

## Creating Visualizations using QlikView

- [Learning Path from a starter to a QlikView expert](#) – This is an exhaustive article which hosts the required materials and resources required for mastering QlikView.

## Creating Visualizations in Tableau

- **Course (mandatory)** – [Data Visualization and Communication in Tableau – Coursera](#) – This is an excellent course provided by Duke University to help people to learn to create stories using Tableau.

- **Blogs/Articles (mandatory)** – [Your guide to become a Tableau expert](#) – This is a comprehensive learning path to become an expert at Tableau. The article is very well structured and detailed. Keep it bookmarked to reference often.

- **Books** – [Communicating Data with Tableau](#) – An excellent book to keep by your side for quick referencing.

# 5.4: Big Data tools and techniques (April 2017)

## Big Data

- **Course (mandatory)** – [Introduction to Big Data by University of California, San Diego](#)

- **Book (optional)** – [Big Data: Using Smart Big Data, Analytics and Metrics to make better decisions and improve Performance](#)

## Other useful tools:

- [H2O](#)
- [SparkR](#) & [PySpark](#)
- Apache Spark
  - **Course –** [Big Data analysis with Apache Spark](#) by edx
  - **Book –** [Learning Spark – Lightening fast Big Data Analysis](#)

# 5.5: Deep Learning Basics & Advanced (May 2017 – August 2017)

## Deep Learning Basics (May 2017 – June 2017)

- **Course (mandatory)**
  - [Machine Learning by Andrew Ng](#) – There is no better introductory material to Deep Learning and Neural Networks than Week 4 and Week 5 material of this course.
  - [Deep learning by Google | Udacity](#) – This is an excellent basic course on transition from Machine Learning to Deep Learning, deep neural networks, Convolutional Neural Networks and Deep Learning for texts.

- **Reading Material/Books**
  - [Deep learning Textbook](#) – Written by people like Ian Goodfellow, Yoshua Bengio and Aaron Courville, this book is bound to become the de-facto for people trying to learn Deep Learning.
  - [Stanford Deep Learning tutorial](#) – This is an all text and images resource provided by Stanford which starts from Linear Regression and goes to Convolutional Neural Networks with ease.

- **Practice** – [Identify the digits](#) –  An awesome contest to check the basics you have learned to identify handwritten digits.

## Deep Learning advanced (June 2017 – August 2017)

- **Course (mandatory)**
  - [Deep Learning by Oxford](#)

- [Deep learning summer school at Montreal 2016](#) – This is a treasure trove of knowledge with many experts researching in the field of Deep Learning delivering keynote lectures.

- **Specialization Material**
  - **Deep Learning for Computer Vision**
    - Primer: ["DL for Computer Vision"](#)
    - Project: ["Facial Keypoint Detection"](#) Tutorial
    - Required libraries: Nolearn
    - Associated Course: ["CS231n: Convolutional Neural Networks for Visual Recognition"](#)
  - **Deep Learning for Natural Language Processing**
    - Primer : ["Deep Learning, NLP, and Representations"](#)
    - Project : "Deep Learning for Chatbots": ["Part 1"](#), ["Part 2"](#)
    - Required library : Tensorflow
    - Associated Course : ["CS224d: Deep Learning for Natural Language Processing"](#)
  - **Deep Learning for Speech/Audio**

    - Primer : ["Deep Speech: Lessons from Deep Learning"](#) news article and corresponding video.
    - Project : ["Music Generation using Magenta (Tensorflow)"](#)
    - Required library : Magenta
    - Associated Course : ["Deep Learning (Spring 2016), CILVR Lab@NYU"](#)

## 5.6: Reinforcement Learning (September 2017 – October 2017)

Topics to be covered: Reinforcement Learning (Theory)

- **Course**
  - [Course on Reinforcement Learning by David Silver UCL [Theory]](#)
  - [Udacity – Machine learning Reinforcement Learning](#) [Theory +Example]
  - [Reinforcement Learning video by Andrew NG](#) [Theory]

- **Code** [Reinforcement Learning Introductory Codes](#)[Code]

  - **Books** [Reinforcement Learning by MIT press](#) – This will be good reference material for the reinforcement learning taught by the professors at MIT.

- **Competitions:**
  - [An awesome website for reinforcement Learning Competitions](#)
  - [Kaggle's first reinforcement style learning competition- Two sigma](#)

## 5.7: Web frameworks & Cloud Computing (November 2017 – December 2017)

**Web Frameworks**

Now that you know machine learning well, you might want to apply it to web products. What you need to learn is a working knowledge about web frameworks. Web frameworks allow you to quickly build and prototype web based products, with out getting into the complications of coding.

Given that you would already have working knowledge of Python, you can choose any of the Python based web frameworks. I would recommend Flask for its simplicity. Flask is a simple and light web framework, which should serve your needs well. If you are looking to build a complex web product, you might want to consider Django as well.

Resources for learning Flask:

- Flask [Official documentation](#)
- [Mega Tutorial on Flask](#)
- [Python Web applications using Flask](#)

**Exercises:**

Additionally, you should do a side project to merry your machine learning skills and web development skills. You can build a simple web application where users can upload pictures and find which make and model the car is. Or may be tells people about their age.

## Cloud computing

Now that you know how to build web applications, you should also get your hands dirty on cloud computing. A few popular platforms are Amazon Web Services (AWS), Google Cloud platform and Microsoft Azure.

Each of these platform provide extensive documentation for their offering. If you have to pick only one – AWS is the way to go because of its popularity, wide spread use and comprehensive offerings.

# End Notes

I hope you found this learning path helpful. I have made it as specific and comprehensive as possible. If you think I have missed out on any specific areas or resources, do let me know.

If you want to progress in your data science journey all you have to do is choose your category and follow the learning diligently.

If you have any questions, doubts or suggestions drop in your comment below and I will be happy to answer them.

If you want to make your own learning path share it with me how are you planning to follow your journey of becoming a data scientist.

# [Learn](#), [compete, hack](#) and [get hired](#)!

Article    Url    -    https://www.analyticsvidhya.com/blog/2017/01/the-most-comprehensive-data-science-learning-plan-for-2017/

## NSS

I am a perpetual, quick learner and keen to explore the realm of Data analytics and science. I am deeply excited about the times we live in and the rate at which data is being generated and being transformed as an asset. I am well versed with a few tools for dealing with data and also in the process of learning some other tools and knowledge required to exploit data.