



## 12-WEEK DATA SCIENCE BOOTCAMP CURRICULUM 2015

### WEEK 1 - BASIC DATA SCIENCE TOOLKIT

#### Become proficient with Linux, Git, Bash, and SQL

##### Linux system

- Introduce Linux environment
- Learn Linux commands
- IO redirection and Pipe
- Introduce server-side Linux usage

##### Git

- Introduce modern source code management
- Learn common git operations
- Setup github and personal portfolio page

##### Other server related topics

- Text editors and IDEs
- ssh: how to communicate with a remote server
- Linux environment variables
- Linux pipeline

##### SQL

- Introduction to relational database
- Introduction to Structured Query Language
- MySQL setup and configuration
- SQL major commands

### WEEK 2 - DATA SCIENCE WITH PYTHON

#### Python Data Analytics – Part I

##### Python Programming Language 1

- Overview of syntax
- Built-in functions
- Data structures
- Standard libraries
- Object oriented programming

##### Python Programming Language 2

- List comprehension
- Data copy
- Introduction to algorithm concepts

##### Common packages

- numpy
- scipy
- pandas

##### Python visualization

- Matplotlib
- seaborn

### WEEK 3 - ALGORITHMS IN COMPUTER SCIENCE, 1ST PROJECT REVIEW

#### Python Data Analytics – Part II

##### Python web scraping

- Regular expressions
- Introduce to web scraping: Ajax, XPath, BeautifulSoup
- Accessing APIs
- Building web applications (Django, Flask, optional)

### Introduce high performance computing and algorithms

- Introduce to algorithms and data structure
- Big-O notation
- Break-out sessions for code reviews
- Daily algorithm challenges

#### 1st Project delivery and review

- Project planning and presentation
- Peer review
- Guidance from instructors

### WEEK 4 - DATA SCIENCE WITH R

#### R Data Analytics – Part I

##### Programming foundation 1

- Syntax
- Data object: Vectors, Matrices, Data Frames, and

##### Lists

- Common functions
- Rstudio environment and package management
- Local data input/output
- Introduction to R data visualization

##### Programming foundation 2

- Data sorting and merging
- String manipulation
- Dates and times
- Connecting to an external database

##### Data manipulation with dplyr

- Tables in R
- Join
- Subset
- Advance manipulations with dplyr

### WEEK 5 - DATA SCIENCE WITH R STATISTICS FOUNDATION FOR MACHINE LEARNING

#### R Data Analytics – Part II

##### Data visualization with ggplot

- Histogram
- Point graphics
- Columnar graphics
- Line charts
- Pie charts
- Box Plots
- Scatter plots
- Visualizing multivariate data
- Matrix-based visualizations
- Maps

#### Statistics Foundation for Machine Learning

##### Review of probability and statistics

- Descriptive statistics
- Hypothesis testing & experiment design
- Introducing non-parametric statistics

##### Regression model

- Linear Regression
- Logistic Regression
- Introduction to generalized linear regression model
- Variable selection
- Bias-variance balance

#### 1st Project delivery and review

- Project planning and presentation
- Peer review
- Guidance from instructor

## R Machine Learning – Part I

Introduction: R in machine learning

R in machine learning 1

- Linear regression
- Logistic regression
- Generalized linear models
- Performance measurements
- Dimension reduction

## WEEK 6 - DATA SCIENCE WITH R STATISTICS FOUNDATION FOR MACHINE LEARNING

### Statistics Foundation for Machine Learning

Classification and support vector machine

- Introduction to classification problem
- Cross validation
- SVM
- Real world example

## R Machine Learning – Part II

K-Nearest Neighbors model

Naïve Bayes model

Laplace estimator

Support Vector Machine (SVM)

## WEEK 7 - DATA SCIENCE WITH R STATISTICS FOUNDATION FOR MACHINE LEARNING

### Statistics Foundation for Machine Learning

Tree based model

- Classification and regression trees (CART)
- Random forest

Unsupervised learning

- Principal Component Analysis (PCA)
- Clustering

## R Machine Learning – Part III

Tree models

GBM

Principle Component Analysis (PCA)

Neural Networks (optional)

Associations Rule and other models

Clustering models

Time series models (optional)

## R Data Analytics – Part III

R packages for project

- Shiny
- knitr

## WEEK 8 - REVIEW ALL TOPICS, 2ND AND 3RD PROJECT DELIVERY AND REVIEW

Project review by peers

Reach out for team work

Review course content

Interview preparation starts

## WEEK 9 - DATA SCIENCE WITH PYTHON

### Machine Learning with Python scikit-learn

Regression and Classification

- Naive Bayes Classifiers
- k-Nearest Neighbors
- Logistic Regression
- Linear Discriminant Analysis
- Lab: Supervised Learning

Resampling and Model selection

- Cross-validation
- Bootstrap
- Feature selection

- Lab: Model selection and regularization

## Support Vector Machines and Decision Trees

- Support Vector Machines
- Decision Trees
- Forests
- Lab: Decision Trees and SVMs

## Unsupervised Learning

- Principal Component Analysis
- Clustering with K-Means
- State Estimation
- Lab: PCA and clustering

## WEEK 10 - PARALLEL COMPUTING, HADOOP, SPARK

### Parallel Computing

Introduction to parallel computing

Parallel computing in R

### Hadoop

Hadoop Ecosystem

- Introduction to the origin and functions of Hadoop
- The principle operations of Hadoop Distributed File System (HDFS)
- The principle system and working mechanisms of Map-Reduce
  - Hadoop data flow
- Map-Reduce programming
- Advanced Hadoop applications (i.e. Hive)
- Data Mining with Mahout

### Spark

Advantages of Spark

Introduction to PySpark

Spark Ecosystem

Lab: text mining

## WEEK 11 - PROJECT REVIEW

Final week of project

*"Show yourself, and potential employers, you have real world data science skills."*

From the beginning of Bootcamp, you will work on hands-on projects. Now your Capstone Project lets you create a useable/viewable data product that showcases your interests and your talents.

Interview Preparation

D3.js (optional)

## WEEK 12 - PROJECT COMPLETION AND INTERVIEW PREP

*"We don't just help you understand it, we help you show it."*

We work in the data industry every day. So we know the skills you need to learn to be the Data Scientist they want to hire and promote. The process starts with our Interview and Coding Prep Sheets, which prepare you with the real-world interview and coding questions. With strong alumni network, we will help you connect with companies hiring data scientists.