Byte Academy

Data Science Syllabus

Data Science Tracks

"Data Science" can mean so many things. It encompasses Wrangling, Analysis, Prediction, Big Data, and much more. Because of the wide range of specialities and job roles, our program consists of three tracks its students can declare during phase 2 of the program.

Data Scientist

This path is ideal for those experienced with statistics or with a strong understanding of math. By the end of this course, students should be able to approach data from a data driven perspective. Not only will we focus on how to use different computational tools, but also on *when* to use them.

Data Engineer

Data Engineers are similar to the typical software engineering role; however, they specialize in data-driven projects. In this track, we will focus on the tools used for maintaining and building data systems.

Data Analyst

This specialty is often diverse in its description, but in this track of our course, we will focus on preprocessing data and analysis of such data. This specialty is particularly great if students are looking to have advanced knowledge of R.

Data Science Specialties

Our data science curriculum is designed with our students in mind. For those interested in a specific domain of tech, we offer three specialities which students can also declare during phase 2 of the program

MedTech

Code to save lives. Recent years have brought significant changes in the healthcare industry especially as health and medical technology has developed. Various areas of medical technology include: Biomedical databases. Bioinformatics, including statistics, mathematics and other data sciences. Medical devices. Genomics.

FinTech

Code for interesting problems. FinTech is a rapidly growing industry that's requiring more engineers to fit the industry's needs. In response to this transition of talent, we train individuals how to program for financial domain. Students graduate with the ability the analyze financial data.

Areas of FinTech Include: Peer to Peer Lending, Cryptocurrency, Mobile Payments, Analytics, Developer APIs, Cybersecurity, Cloud Security.

Social Sciences

Code to make society better. This specialty explores important questions about how data science will transform society as a whole. The computational social sciences generates interdisciplinary projects that utilize data science tools to tackle policy problems.

Computational social sciences raises questions about the politics and ethics of data science research, particularly when it focuses on sociopolitical problems with applications in government and the private sector.

Online Pre-work

We want student to be as prepared as possible for future careers. To get the most out of the experience, we require all students to complete pre-work designed by our instructors. This curriculum consists solely of introductory material and should be expected to take 10-35 hours, depending on prior experience.

Topics

Python - designed to introduce students to fundamental computer science concepts and successfully develop programming skills during the bootcamp

Statistics - brief but thorough overview of statistical concepts, designed to prepare students for statistical modeling and prediction

Linear Algebra - brief overview of vectors and matrices, designed to prepare students for machine learning

Github - github is a version control system students will utilize throughout the entirety of the bootcamp

Command Line - being comfortable with the terminal and all its commands is incredibly important for those interested in entering the tech field.
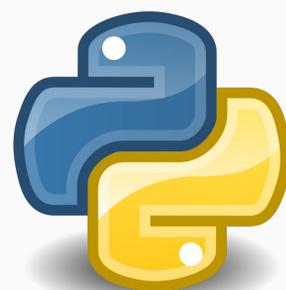
Week 1

Python Fundamentals

Week 1 begins with acquiring an in-depth knowledge of the Python programming language. By the end of the week, students will be expected to program intermediate level scripts in Python.

Tools Utilized:

Python

Day 1
  Github Workflow & Integration
  Data Types and Operators
  Control Flow

Day 2
  Functions
  Lambda Functions
  Decorator Functions

Day 3
  Object Oriented Programming
  Classes & Inheritance

Day 4
  Modules
  Input/Output

Day 5
  Exercises & Examination

Week 2

Statistics & Linear Algebra

Week 2 of our curriculum is dedicated to creating a deep understanding of mathematical concepts we'll later see in topics like machine learning and statistical analysis. Contrary to the traditional mathematics course, students will learn statistics and linear algebra through a computational lens.

Tools Utilized:

numpy

SciPy

Day 1
> Descriptive Statistics
> Distributions & Histograms

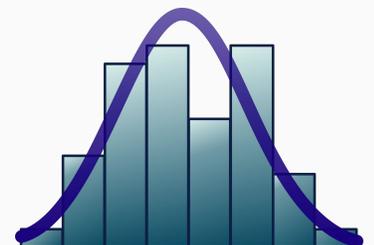**Day 2**
> Cumulative Distribution Functions
> Skewness

Day 3
> Conditional Probability
> Bayes Theorem
> Estimation

Day 4
> Hypothesis Testing
> Correlation

Day 5
> Vectors & Matrices
> Matrix Operations

Week 3

Data Wrangling

Week 3 begins the true start of industry focused tools, such as scrapy, pandas, beautifulsoup, and more. We'll learn the different tools in which we can acquire, prepare, clean, and manipulate data to fit a problem's needs.

Tools Utilized:

json

BeautifulSoup

Scrapy

Requests

Pandas

Day 1
CSV, json, and zipfiles
APIs

Day 2
HTML & CSS basics
Web Scraping

Day 3
DataFrames
Data Merging
Data Normalization

Day 4
Missing Values
Outlier Detection

Day 5
Exercises & Examination

CSV

Week 4

Data Visualization &
Exploratory Analysis

On week 4 we'll begin
curriculum focused on
various data visualization
techniques and how they
can help us engage and
learn from our data.

Tools Utilized:

Matplotlib

Seaborn

Ggplot

Geojsonio

Shapely

Descartes

Day 1
Line and Scatter Plots
Histograms
Visualization Customization

Day 2
Seaborn & ggplot
Exploratory Computing
Box Whisker Plots
Heatmaps

Day 3
Factorplots & FacetGrids
Advanced Graphs

Day 4
Geospatial Visualizations
Map Visualizations

Day 5
Exercises & Examination

plotly

Week 5

Regression Analysis

Week 5 begins the official start of the statistical analysis and prediction portion of this course. We'll spend week 5 engaging with the basics of machine learning and work our way towards learning and implementing several regression models.

Tools Utilized:

scikit-learn

R

Day 1
Intro to Machine Learning
Types of Learning & Data

Day 2
Maximum Likelihood
Linear Regression
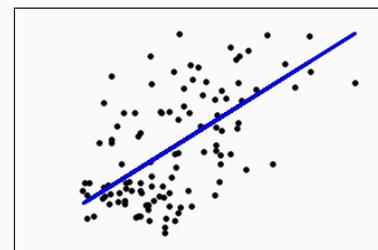Multiple Linear Regression

Day 3
Non-Linear Regression
Logistic Regression

Day 4
Time Series Analysis
Stepwise Regression

Day 5
Ridge & Lasso Regression
Exercises & Examination

Week 6-7

Machine Learning

Working off of week 5's curriculums, we'll enter the sphere of machine learning algorithms, including topics in both supervised and unsupervised learning. These two weeks will be heavily project and exercise based, but with the mathematical implications of this work heavily in mind.

Tools Utilized:

scikit-learn

caret

Day 1
Review of Bayes
Naive Bayes & Joint Models

Day 2
Classification
Support Vector Machines

Day 3
Mini Project Day with Kaggle

Day 4
K-Means Clustering
K Nearest Neighbors

Day 5
ROC Curves
Cross Validation

Day 6
Decision Trees
Random Forests

Day 7
Optimization
Regularization

Day 8
Project Day

Day 9
Principle Component Analysis
Dimension Reduction

Day 10
Boosting & Bagging
Examination

kaggle

Week 8-9

Natural Language Processing & Deep Learning

Once again expanding on the knowledge gained from weeks 5-7, we will enter the realm of machine learning involving textual analyses and artificial neural networks. Because the two can be complementary, we will also engage with topics like word2vec and more.

Tools Utilized:

re & nltk

gensim

Tensorflow

Theano

Keras

Day 1
 Regular Expressions
 Components of Speech
 Text Normalization
 Word Tagging

Day 2
 Sentiment Analysis
 Information Extraction
 Named Entity Extraction

Day 3
 Topic Modeling
 Summarization

Day 4
 Neural Networks
 BackProp & Gradient Descent

Day 5
 Mini Project

Day 6
 Feedforward Neural Networks
 Recurrent Neural Networks

Day 7
 Autoencoders & Embeddings
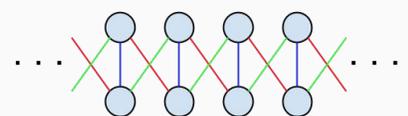 Vector Space Models
 Word2Vec

Day 8
 Mini Project

Day 9
 Convolution Neural Networks
 LSTM Networks

Day 10
 GPUs & Hardware
 Examination

Week 10

Databases

Week 10 dives into database systems headfirst. During this week, becoming fluent with SQL, NoSQL, and MySQL is a must that will carry over into the rest of this course.

Tools Utilized:

sqllite

mongoDB

PostgreSQL

Day 1
    Intro to Databases
    SQL Basics
    Database Modeling

Day 2
    Advanced SQL
    Database Design

Day 3
    NoSQL
    MongoDB

Day 4
    MySQL
    PostgreSQL

Day 5
    Exercises & Examination

Week 11

Big Data

With the consistent growing of data every day, engineers are forced to become equipped to handle, prepare, and process this data in a computationally efficient manner. This week reviews the different big data architecture tools available in the data science industry today.

Tools Utilized:

AWS

hadoop

docker

Kafka

Spark

Storm

Day 1
MapReduce
Hadoop

Day 2
Spark
Hadoop Ecosystem

Day 3
Kafka
Storm

Day 4
Amazon Web Services
Cloud Computing

Day 5
Project Day

Week 12–14

Final Project

We emphasize projects and pair-programming so that you graduate with a portfolio to showcase to potential employers, as well as to prove to them that you can develop in a team environment similar to the real world.

Students are required to use at least three of the technologies we used throughout the course. Projects for this program last about three weeks that will ultimately result in a demo for others to see.

On Demo Day, students present their final projects to hiring partners.

•   •   •   •   •   •   •   •   •   •   •   •   •   •

As part of the final project, students will be guided by instructors at Byte Academy, as well as working professionals in the field. We believe this heavy mentorship will provide students with the proper support for success, while simultaneously ramping the skillset developed throughout the fourteen weeks of our course.