THINKFUL

Data Science Syllabus

Fundamentals

This 4 week course is required to get you up to speed for the main course. At the end of it, you will submit a capstone and present it to move on to the main course.

Unit 1 – Introductory Programming in Python

Concepts covered: Data Types, Application Logic, Loops and lists, Dictionaries, Functions, Objects, Classes, Inheritance, Modules

Project(s) you'll build:
- Extensive drills to master programming fundamentals

Unit 2 – Introduction to Data Science Toolkit

Concepts covered: NumPy, Pandas, Data Visualization, matplotlib, Basic Plot and Scatter, Subplots, Statistical Plots

Project(s) you'll build:
- Using your chosen data source, you will generate at least four different data visualizations using the learned concepts.

Unit 3 – Statistics for Data Science

Concepts covered: Population vs sample, Central Tendency, Measures of Variance, Randomness, Sampling and Selection Bias, Independence and Dependence, Bayes' Rule, Normal Distribution, Central Limit Theorem

Project(s) you'll build:
- Drills to master statistics fundamentals.
- Solve the Monty Hall problem

Unit 4 – Career Planning and Capstone Report

Concepts covered: Career planning, Capstone

Project(s) you'll build:

- Career Plan - Explore the variety of data science work being done, understand the skills companies are looking for, find your future professional community, and create a preliminary vision for your career.
- Prep Course Capstone - You will complete an Analytic Report and Research Proposal on a data set of your choosing.

Data Science — Main program

Unit 1 – Data and Analysis
Concepts covered: Matplotlib, SQL, SQLite, Data Cleaning, Data Visualization, Seaborn, Experimental design, A/B Testing

Project(s) you'll build:
- SQL Challenge - Solve questions about AirBnB data using SQL queries with a database that you'll set up locally.
- Data Cleaning & Validation - Practice data cleaning & validation using a dataset of your choice from the web
- Your First Research Proposal - Using a dataset of your own choice, create your first Research Proposal (also known as an Experimentation RFC).

Unit 2 – Supervised Learning
Concepts covered: PCA, Feature engineering, Naive Bayes, Regression models, Classification models, Least Squares Regression, Multivariable Regression, Class Imbalance

Project(s) you'll build:
- Prepare a Dataset for Modeling - Using a dataset of your choice, you will explore variables using univariate and bivariate methods
- Build your Own Naive Bayes Classifier - Perform a sentiment analysis on feedback left on a website to determine if it is positive or negative.
- Classifier Validation - Test the performance of your classifier from the previous project and learn how to improve it.
- Your First Multivariate Linear Regression Model - Build a regression model using FBI UCR Crime data in order to predict property crimes.
- Validating a Linear Regression - Validate your property crime model and based on the results create a revised model. Test both old and new models on a new holdout or set of folds.

Unit 3 – Deeper into Supervised Learning
Concepts covered: Similarity Models, KNN, Decision Trees, Random Forest, ID3 Algorithm, Ensemble Modeling, Advanced Regression, Support Vector Machines, Boosting Models

- Model Comparison - Using your own chosen data set build a KNN and an OLS regression and compare them.
- Random Forests & Decision Trees - Compare the relative accuracy of random forests and decision trees using a data set of your choosing.
- Support Vector Machines Challenge - Translate a weak SVR into a more accurate SVC
- Boosted Models - Give your model a boost in the Boosted Model Challenge.

Unit 4 - Unsupervised Learning

Concepts covered: Unsupervised learning, Basic Clustering, K - Means, Clustering Evaluation, NLP (Natural Language Processing), Neural Networks, Deep Learning

Project(s) you'll build:
- Supervised vs Unsupervised Drill - Determine whether a problem is best solved using supervised or unsupervised techniques.
- .Applying K Means - Use your knowledge of basic clustering to determine variance with changes in K.

Unit 5 – Other Topics in Data Science

Concepts covered: Algorithms, Data Scraping, Big Data, Survey Design, Privacy and Data Science

Project(s) you'll build:
- Data Scraping - Learn the value of Data Scraping and practice on a source of your choosing.
- Survey Design - Create a survey on the topic of your choosing and gather data from users.
- Algorithms - Build your own algorithm for some of the models we've gone over so far!

Unit 6 – Specializations

This is an opportunity to work with a mentor to cover topics more aligned to your chosen specialization. This can include but is not limited to Natural Language Processing, TensorFlow, Econometrics, Social Genomics/Quantified Self, Time Series and Forecasting, Anomaly Detection, Big Data using Spark or Visualization.

Unit 7 – Final Capstone Project

To cap off the course, you'll build:
- A custom project showcasing your data science skills as well as specialization knowledge.