Geoscientist to Data Scientist 12 week Syllabus

12-Sep	12-Sep	18-Sep	25-Sep	2-Oct	9-Oct	16-Oct	23-Oct	30-Oct	6-Nov	13-Nov	20-Nov	27-Nov	4-Dec
0 (PreReq)	**0 (PreReq)**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Pre-Requisite	**Pre-Requisite**	**Introduction to Data Projects**	**Data Analysis**	**Data Analysis**	**Data Analysis**	**Data & Communication Sandbox**	**Data Fusion and Machine Learning**	**Data Fusion and Machine Learning**	**Special Data Types -Time-series Data**	**Data Fusion Sandbox**	**Special Data Types - Natural Language Processing and Text Mining**	**Special Data Types - Spatial Data**	**Capstone Project Development & Presentation**
Introduction to The Course Tools The Why	**Introduction to The Course Tools The History of Where Data Science Comes From**	**Zero to data science in a day**	**Getting to know the tools - data munging and exploratory data analysis**	**Simple predictions - regression and statistical model building**	**Multivariate analysis, building models**	**Effective data storytelling – communicating results to non-technical audiences**	**Commonly used statistical and machine learning techniques I**	**Commonly used statistical and machine learning techniques II**	**The 4th Dimension and Predictions**	**Consolidate approaches covered**	**Finding needles in wordstacks**	**Spatial analytics and predictions**	**Pitching Capstone Projects**
0.1 The Why Teamwork aspects This course vs MOOCs How your work will add value into the whole org	0.5 Intro to Python and Jupyter Notebooks	1.1 Intro to Data Science Workflows Workflow all filled in to give end to end overview	2.1 Tabular data munging with pandas	3.1 Models and regression	4.1 Multivariate analysis, compositional analysis, useful data transforms	5.1 Making decisions with data, data ethics	6.1 Morning 6.1 – Machine Learning Framework A look at the toolbox -, recap what gets carried forward from weeks 1-5	7.1 Distinguish ML, AI and Deep Learning Recurrent Neural Networks for time series predictions (50 mins)	8.1 Genetic/Evolutionary algorithms	9.1 Sandbox I - explore with own data or provided datasets	10.1 Handling text, aims and goals, meaning, ontology - why might we care about this?	11.1 Point-based spatial predictions	12.1 Thinking about my Problem
0.2 Introduction to tools Intro to notebooks and binder Version Control - Git	0.6 Starting with Data	1.2 Individual Data Sets and Statistics Missing data How to integrate datasets	2.2 Data QA/QC - formats, tidy data etc using mornings examples -	3.2 Regression and basic predictions	4.2 Dimensionality reduction, cluster analysis, feature extraction	5.2 Advanced data visualization and communication	6.2 Teaching machines to classify - training models and assessing performance	7.2 Deep learning: models and aspects of training I Deep MLP, backpropagation, LSTM, regularisation	8.2. Statistical process control	9.2 Sandbox II Report back - what did you try, did it work, peer feedback	10.2 Basic text processing – lemmatization Feature extraction	11.2 Spatial statistics, sampling theory, Gaussian Processes	12.2 Persuasive Pitches and Visualisation
0.3 Intro to Git	0.7 Data workflows and automations	1.3 Fusing data Communicating results to come in week 5 after learning some analysis	2.3 Distributions and visualisation of 1D and 2D data	3.3 Model selection how-to	4.3 Handling missing data, detection limits etc using regression for imputation	5.3 Developing communication design as group activity (Storyboard)	6.3 Improving your inputs - feature extraction and selection	7.3 Deep learning: models and aspects of training II CNN applied to an industry problem	8.3. Linear and Mixed integer programming	9.3 External guest speaker - Neural network industry examples (Woodside, Google)	10.3 Unsupervised learning - latent topic models with scikit-learn and gensim	11.3 Reflection - how can I apply what I have learned (work towards Capstone)	12.3 Pitch Presentations for Follow on Projects
0.4 Introduction to Git Hub Issue tracking and other benefits of Git Hub Using Gitter as a comms tool for asynchronous project	0.8 Making plots The History of Where Data Science Comes From	1.4 Mapping interfaces in your business, considering trade-offs when planning data projects	2.4 Exploratory data analysis reports for individual datasets.	3.4 Evaluating your models, building hierarchical models	4.4 Exploring multidimensional approaches on a given dataset	5.4 Communication design continued, peer assessment of group work	6.4 Working it out on your own - unsupervised learning & clustering	7.4 Deep learning: models and aspects of training III Meaningful compression - auto-encoders	8.4 Roadmap for Experimental design and Analysis method	9.4 Reflection - how can I apply what I have learned (work towards Capstone)	10.4 Supervised learning example - Entity recognition with neural nets (spacy + prodigy)	11..4 Reflection - how can I apply what I have learned (work towards Capstone)	12.4 Program Debrief