Geoscientist to Data Scientist 12 week – Weekly Aims and Learning Outcomes by Topic

12-Sep	12-Sep	18-Sep	25-Sep	2-Oct	9-Oct	16-Oct
0 (PreReq)	0 (PreReq)	1	2	3	4	5
Pre-Requisite	**Pre-Requisite**	**Introduction to Data Projects**	**Data Analysis**	**Data Analysis**	**Data Analysis**	**Data & Communication Sandbox**
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**
Aim 1: To introduce the structure and tools of the 12W program Aim 2: To clarify participant expectations and desired outcome of the program	Aim 1: To get over the initial 'inertia hump' of beginning your data science journey Aim 2: To be aware of what data science is and where the field has come from	Aim 1: Overview the entire sequence of a 'typical' data science workflow	Aim 1: Create and interpret statistics and visualisations after completing appropriate QA/QC Aim 2: Be able to implement an EDA and apply pattern recognition principles while avoiding pitfalls	Aim 1: Perform statistical model building Aim 2: Conduct regression Aim 3: Generate simple predictions - regression	Aim 1. Understanding multivariate EDA Aim 2. Analysing central tendency, dispersion and correlations Aim 3. Multivariate techniques PCA, MDS, SOFM and clustering	Aim 1: Employ a variety of effective data communication strategies Aim 2: Appraise data driven narratives Aim 3: To refine and appropriately frame questions
1. Be able to define the WHY, the WHAT and the HOW of the 12 week program 2. Be able to explain the need for version control 3. Be able to explain how to track files on own computer using Git 4. Be able to demonstrate how to collaborate (e.g. using remote repositories)	1. Be able to open and use the tools 2. Recognise and understand basic Python syntax 3. Perform analysis by loading and using packages 4. Appraise data cleaning and analysis requirements of a data set 5. Conduct automation of workflows (e.g. writing functions and conditional statements) 6. To appreciate what data science is	1. To understand the stages of a data science project and to be able to define a mental model of it 2. To practice using git, Python and Jupyter notebooks 3. Demonstrate how to connect to and explore data sets using the sci-kit learn and binder interfaces 4. Analyse the opportunity and potential value of data science in your organizations	1. Understand the pitfalls of different data types 2. Appreciate the importance of choosing i.e. 'clean data' and be aware of some QA/QC approaches for enforcing this 3. Perform basic data visualisations given tabular data 4. Construct reasoning to explain links between data and statistical distributions including pattern recognition 5. Critique basic summary statistics after implementing an EDA	1. Understand regression as the basis for prediction 2. Understand how outliers and noisy data affect results 3. Understand the impact of missing data and recall practical solutions to work with incomplete data sets 4. Understand how to choose between basic statistical models and evaluate their effectiveness (e.g. linear vs polynomial) 5. Develop an understanding of hierarchical models as a means of modelling connections between datasets or processes	1. Understand the differences between univariate and multivariate data and apply multivariate techniques and EDA 2. Understand the basis of dimensionality reduction 3. Execute measures of correlations 4. Critique whether a correlation is spurious 5. Identify if the process used in the activities is reproducible	1. To invite participants to consider avenues for communicating project results beyond the usual dashboard/graphs approach. 2. Participants are aware of interpretation biases in their audiences and how design principles can be used to mitigate these 3. Participants can critique data presentations (both good and bad) in a variety of media 4. Develop effective data-driven communication in each context 5. Examine 'data ethics' – handling (personal) data appropriately

Geoscientist to Data Scientist 12 week – Weekly Aims and Learning Outcomes by Topic

23-Oct	30-Oct	6-Nov	13-Nov	20-Nov	27-Nov	4-Dec
6	7	8	9	10	11	12
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**
Pros and cons of commonly used statistical and machine learning techniques I	**Pros and cons of 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**
Aim 1: supervised and unsupervised classification and the crucial steps in end-to-end machine learning strategy from data. Aim 2: Explore traditional supervised classification techniques and the evaluation metrics for error analysis.	Aim 1: Introduce recurrent neural networks for time series prediction Aim 2: Introduce the main concepts of deep learning with industry examples. Aim 3: Present the practical strategies to train neural networks	Aim 1: Extract text to run ML pipelines over	Aim 1: Practice and questions session	Aim 1: Extract text to run ML pipelines over Aim 2: Present the practical strategies for using text	Aim 1: Extract geospatial features to run ML pipelines over Aim 2: Present the practical strategies for using geospatial data	Aim 1: Deliver a compelling pitch describing a data project to create value Aim 2: Program debrief and evaluation Aim 3: Introduce Community of Practice & connect Enabling leaders to Professionals
1. Understand the important steps to develop a machine learning investigation from data 2. Understand how to evaluate, validate and prevent problems such as overfitting and underfitting. 3. Understand which techniques are best suited for a classification problem and how to design the learning model 4. Know where to find packages and toolboxes applicable to mining/energy data 5. Understand the advantages and disadvantages for the techniques to select the correct type of model given available data	1. Know the powerful framework given by neural networks and deep learning techniques 2. Know what types of data are required by such algorithms 3. Know how to train a neural network (and its different flavours) 4. Know how to develop a project for time series prediction 5. Know how to develop a project for classification with deep neural networks	1. Genetic (Evolutionary) algorithms 2. Time based considerations such as lags, time stamps, periodicy, resampling 3. Statistical process control 4. Linear and mixed integer programming 5. Experimental design	1: Practice and questions session	1. Handling text 2. Text preprocessing 3. Text with supervised and unsupervised learning	1. Vector and raster geospatial formats with feature extraction 2. Point based spatial predictions 3. Spatial statistics and sampling theory	1. Be able to demonstrate the proposed value of an action by explaining your project metrics and expected outcomes 2. Formulate persuasive visualisations to give your 'story' structure