### [Advanced] 2-day workshop Syllabus

**Teacher:** Michael Pyrcz, Ph.D., P.Eng. ([mpyrcz@daytum.org](mailto:mpyrcz@daytum.org)) / John T. Foster, Ph.D., P.E. ([jtfoster@daytum.org](mailto:jtfoster@daytum.org))

**Where and when do we meet?**   
July 11-12th, 2019 8:00 AM- 5:00 PM at [700 Louisiana St, Suite 2400, Houston, TX 77002**]**  
  
**What you’ll need:**   
Daytum’s 3-Day Workshop Certificate of Completion **OR**   
Oil and gas industry experience and extensive knowledgy of the Python Data Science Ecosystem (e.g. NumPy, SciPy, Matplotlib, Bokeh, Intake, Panel, Scikit-Learn)

**Workshop Objective: [To learn the theory and practical workflows for the application of data analytics, geostatistics and machine learning with energy industry data support development decision making.**

**Workshop Overview:**  
  
Our 2-day workshop focuses on the advanced application of data analytics, geostatistics and machine learning to **energy industry data**.

This [intermediate course is a critical step in laying the foundation necessary for thinking statistically and identifying the key signals from the noise that is data.

Specifically, our intermediate 2-day workshop will teach you:

* To effectively prepare data for deep dives with advanced analytic techniques and ensure that any drawn conclusions are trustworthy and reliable
* To glean insights and make predictions from your data, using such techniques such as outlier detection, data debiasing and imputation, feature engineering, anomaly detection, supervised and unsupervised learning, spatiotemporal modeling, uncertainty modeling.
* To understand the assumptions and limits of data precision, scale and coverage, spatial interpolation, multivariate models, analytics and uncertainty models, given that predictions are only as strong as your process
* What Big Data really means and how to apply robust machine learning tools to real Energy data problems (think Dask, Torch, TensorFlow etc.).

This course is designed to be highly interactive and tailored to the professional seeking to build from the ground up to deploying products to immediately add value. You will be coding python classes and functions, dashboards, models, and effective visualizations, using all free open-source tools.

By the end of our workshop, you will have a firm understanding of the role of statistics within data science and the power of advanced analytics and machine learning techniques. As a deliverable, you will have [x].

Note: Our program is rigorous, fast-paced, and focused on practical technical skills needed to solve data problems, automate workflows, visualize data, and leverage predictive analytics. It won’t be easy, but it will be fun. We hope to see you there!

**Course Format:**

Strategy  
Our workshops incorporate a just-in-time (JIT) teaching strategy. What this means is that we will show you final deliverable (think dashboards, models, visualizations, etc.) and teach you step-by-step how to build it, while exploring the Python data science ecosystem in a broader context.

Major course concepts will be covered in one-hour bundles, with 15 minutes of lecture, 15 minutes of live demonstration and 30 minutes of interactive coding, coaching, and troubleshooting.

**Course materials will consist of web-hosted lecture materials, instructor led, well-documented method and workflow demonstrations, and programming examples. Everything will be delivered via a cloud-based Jupyter Lab session, so the only thing you need to bring are your laptop and your fingers!**

**Class Size**We believe in tailored and focused instruction within our classrooms and have structured our courses with a 1:15 teacher to student ratio. Our goal is to give each student the attention and continuous feedback necessary to succeed.

**Class Material  
Daytum classes will include materials for content and reference. Students will be allowed to use class materials (e.g. video-lectures and course notes) for personal use following the completion of the course.**

**Other**

**Cancellation Policy   
Courses may be cancelled up to two weeks prior to the start of the class. Cancellation can be done by sending an email to** [contact@daytum.org](mailto:contact@daytum.org)**. If a class is cancelled after this time, the student is responsible for the full cost of the course.**

**Software Requirements   
A laptop with a working web browser (try Google Chrome!)**

**Feedback  
Following the course, Daytum may ask for feedback on instructor quality and course content. This information is kept highly confidential and is only used solely to improve the content offerings and the quality of instruction!**

**Detailed Course Outline:**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Time** | **Topic** | **Objective** |
| **Day 1** | 8:00 AM - 8:30 AM | Course Goal: [Deep dive into Data Analytics, Geostatistics and Machine Learning] |  |
| 8:30 AM - 8:45 AM | Prerequisites and Perspectives |  |
| 8:45 AM - 10:45 AM | Intro to Data Preparation – Debiasing, Statistics 101: Reporting Significance and Uncertainty |  |
| 10:45 AM - 12:00 PM | Feature Engineering |  |
| 12:00 PM - 1:00 PM | Lunch |  |
| 1:00 – 3:00 PM | Spatial Models and Spatial Estimation |  |
| 3:00 – 5:00 | Building Workflows: Putting it Together |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
| **Day 2** | 8:00 AM - 9:00 AM | Day 1 Review – Data Preparation, Feature Engineering, Spatial Models |  |
| 9:00 AM - 10:00 AM | Introduction to Machine Learning |  |
| 10:00 AM - 11:00 AM | Inference: Dimensionality Reduction |  |
| 11:00 AM - 12:00 PM | Inference: Clustering |  |
| 12:00 PM - 1:00 PM | **Lunch** |  |
| 1:00 PM - 2:00 PM | Prediction: Tree-based Regression |  |
| 2:00 PM - 3:00 PM | Prediction: Support Vector Machines and Neural Networks |  |
| 3:00 PM - 4:00 PM | Building Workflows: Putting it Together |  |
| 4:00 PM - 5:00 PM | Closing Remarks |  |