

Data Science Readiness Training
A Mosaic Data Science Case Study

Background

The client, one of the largest oil and gas manufacturers and commodity marketers in the world, felt they had the right people in place to develop an internal data science capability. The company had been collecting mass amounts of data for a very long time, but lacked the skill-sets to develop business cases and produce results with the data.

As data collection technologies develop, companies have access to information they previously never had. Companies need to mine this data and act upon the predictive insights they develop, or else they are just spending money on housing information in databases, spending time and money while gaining no insight into their business.

The company formed a group of 13 business solution advisors, whose main focus was to receive requirements from whichever line of business they were assigned to (upstream, downstream, production, marketing, etc.) and provide a data-centric solution to the question or issue. The company had invested hundreds of millions of dollars in data collection technology and believed their infrastructure was complete.

Objectives

The oil and gas client realized they needed their team to be able to start taking advantage of their data assets. They wanted their team to be able to identify areas around the business where they could use data science and predictive analytics to improve their decision making.

Mosaic was asked to deliver training which would bring everyone on the analytic solution team to the same knowledge level. Once the team had a base level of knowledge, they could begin to apply predictive analytics and optimization techniques to the business challenges.

The goal was to insure that solution advisors and solution engineers were well versed in mathematical and data lingo to develop actionable analytics solutions which would bring high impact results.

Approach

Having been in the data science business for over a decade, Mosaic knows a thing or two about data science. They have seen what works and what doesn't work. Mosaic developed, in conjunction with the customer, a four-day training which focused on business case development, data science theory, managing a data science project, and model development and implementation.

To make the training as relevant as possible, Mosaic asked the team for several project ideas. Being able to work with tangible examples help ground the training to the day-to-day jobs of the team.

On the first day Mosaic presented a session focused on the application of data science techniques. The session taught solution advisors how to determine which methods a well-defined business problem requires. The seven classes of data science methods used were:

- Prediction
- Classification
- Recommendation
- Text analysis
- Risk analysis
- Decision analysis
- Optimization

Day two surveyed the modeling process – that is, the process of developing a scientific model in a business context. Topics included:

- Building the business case
- Working with sponsors and stakeholders
- Problem definition
- Concept development
- Data assessment, enrichment, storage, and preparation
- Variable selection
- Model selection
- Model evaluation

Once the analytic solution team felt they could confidently select a model, Mosaic moved on to Day three. Day three consisted of implementing a scientific business model as a decision support application. Topics included:

- Working with sponsors and stakeholders
- What agile, scrum and lean engineering require in the context of developing a decision-support system (DSS)
- Data architecture for data science
- Choosing proprietary, open source, and homegrown tools for data storage, data preparation, data analysis, and testing
- Correct and efficient implementation of analytical algorithms
- Testing for DSS development
- Deploying a DSS
- Maintaining a DSS

On Day four, Mosaic focused on data science management and the unique challenges managers face as they shift from a business intelligence (BI) focus to a collaborative effort with the business lines required by advanced analytics. Topics included:

- Data strategy and the strategic use of data
- Preparing a business case for a data-science project
- Team structure
- When to train, hire, and retain data scientists
- Relationship with IT and BI

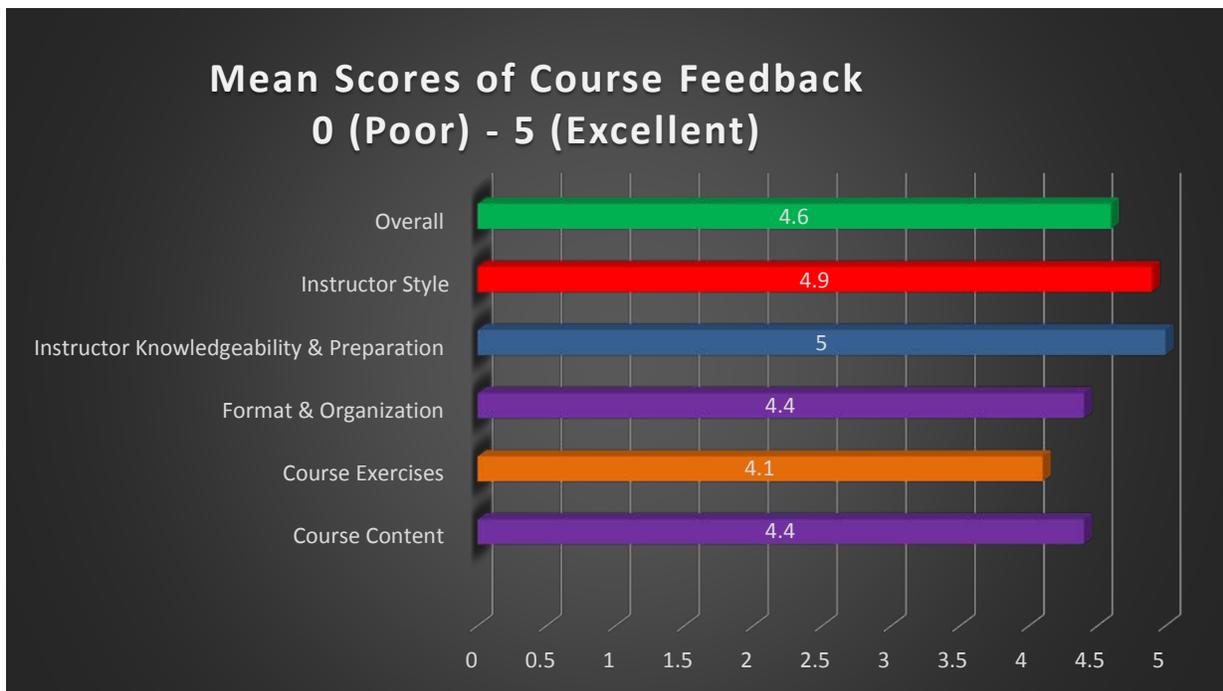
- Technology stack
- Managing scientific-model development
- Managing scientific-model implementation.

Results

The team is now able to understand what is possible with data science and where it can be most effective. This gives them an incredible competitive advantage in today’s landscape, as well as saving them millions on their bottom line.

The analytic solution advisors now have a common terminology and approach to developing use cases and working through the data science process. They are now seen as experts and a bridge between IT and the lines of business.

Mosaic handed out a survey at the end of the training to receive feedback from the twelve participants.



Course Content – “Loved the discussion & theory the most” “Topics were relevant in business world as well as provided a good glimpse of advanced / scientific analysis”

Format & Organization – “Good format, well balanced” “Well organized and well presented, flows well”

Instructor Knowledgeability & Preparation – “Impressive educational and practical experience” “Ridiculously smart”

Instructor Style – “Good presenters” “Incredibly well prepared instructors”

Overall Course Rating – “Good class, very dense in material” “Would highly suggest to others”

Technologies Involved

Mosaic trained the participants on R and the RStudio integrated development environment (IDE). Technical exercises leveraged some of the powerful statistical and machine learning packages available from the CRAN repository.

<https://www.r-project.org/>

<https://www.rstudio.com/>

For More Information

[Contact Mosaic](#) today to learn how we can help your organization achieve equally dramatic returns on your data science investment!