

Day 1

- **Overview of Data Science**
 - Definitions
 - CRISP-DM
- **Data We Will Use**
 - Tools for Analysis
 - Titanic
 - KDD Cup 1998
- **Business Understanding**
 - Formulating the Problem
 - Setting Up Data
- **Data Understanding / Data Visualization**
 - Data Audit
 - Data Visualization Graph Types
- **Data Preparation**
 - Data Cleaning
 - Feature Creation / Selection / Sampling

Day 2

- **Modeling and Evaluation: Classification**
 - Supervised Learning
 - Classification vs. Estimation
 - Decision Trees
 - Evaluation
 - Logistic Regression
 - Neural Networks, kNN, Naïve Bayes, SVM
- **Modeling and Evaluation: Regression**
 - What is Regression
 - Linear Regression, Prep, Eval
- **Data Storytelling**
 - Data Visualization
 - Converting Models to Narratives
 - YMCA Case Study
- **Model Ensembles**
 - What are Ensembles; heterogeneous

Day 3

- **Model Ensembles Part 2**
 - Bagging

- Boosting
- RF
- GBM
- **Clustering and Evaluation**
 - What is Clustering
 - Kmeans. Kohonen SOMs
 - Data Prep
 - Interpreting Clustering
- **Deployment**
 - Deployment Strategies
- **Building Models together on the fly (live data demos); Case Studies**
 - Potential Case Studies:
 - Fraud Detection
 - Customer Lifetime Value
 - Titanic Data Start to Finish
 - KDD Cup Data Start to Finish