

### Data Science

- Business Analytics** - the skills, technologies, applications, and practices for continuous, iterative exploration and investigation of past business performance to gain insight and drive business planning
- Data Mining** - the computational process of discovering patterns in large data sets ("big data")
- Big Data** - the term for a collection of data sets so large and complex that it becomes difficult to process using on-hand database management tools or traditional data processing applications
- Business Intelligence** - a set of theories, methodologies, architectures, and technologies that transform raw data into meaningful and useful information for business purposes
- Descriptive Analytics** - to gain insight from historical data with reporting, scorecards, clustering, etc.
- Predictive Analytics** - a variety of techniques from statistics, modeling, machine learning, and data mining that analyze current and historical facts to make predictions about future, or otherwise unknown, events

### Data Science (cont)

- Prescriptive Analytics** - automatically synthesizes big data, multiple disciplines of mathematical sciences and computational sciences, and business rules, to make predictions and then suggests decision options to take advantage of the predictions

### CRISP-DM

Cross Industry Standard Process for Data Mining (CRISP-DM):

1. Business understanding
2. Data understanding
3. Data preparation
4. Modeling
5. Evaluation
6. Deployment

### Cluster Analysis

Defini- Partitional vs Hierar- ret  
tions chical

- Type of analysis: –Unsupervised (looking for natural relationships - not trying to predict a target variable)
- Type of data: –Quantitative (interval/ratio) and/or qualitative (ordinal/nominal) data may be used in cluster analysis
- Types of business questions answered: –Do cases (e.g., customers, employees, etc.) tend to cluster into natural groups that we can use to take some action?

