**SCM 651 Study Guide**

* Week 1
  + Background
    - What drives analytics?
    - Why is analytics difficult?
    - What are business examples where analytics is important?
      * Tools
      * Formulas
      * Sorting
      * Filters
      * Pivot tables and charts
      * PowerviewExc**el**
* Week 2
  + What is NPV? IRR?
    - NPV
      * Calculates today’s value of a cash flow stream
      * Investments are entered as negative numbers
    - IRR
      * IRR > interest rate ➔ positive NPV
      * IRR < interest rate ➔ negative NPV
      * The rate of return of the outstanding investment for each period it remains invested
  + Eg: What is the difference between correlation (how, which direction) and linear regression(how much)?
    - Correlation
      * Strongest (highest positive or negative correlation): INTC & MSFT = 0.39
      * Weakest (closest to zero): CAT & MSFT = 0.08
    - Correlation versus regression
      * If one variable changes, does the other variable go up or down? (correlation)
      * If one variable changes, how much does the other change? (regression)
  + Linear regression
  + Exponential regression
    - Compounded growth
  + Power regression
    - Learning curve or volume efficiencies
  + Time series
* Week 3
  + Google Analytics
    - General measures
      * Visits, new visits, % new visits
      * Bounce rate, average visit duration
    - Locations
      * Countries, states, cities
      * Visits, new visits, % new visits
    - Behavior ◦ Frequency and recency
      * Engagement (length of time on site)
    - Technology
      * Operating system & browser
      * Network provider
      * Mobile device
    - Visitor flow
      * Entry and exit point
    - Acquisition
      * Channels (organic search, referral, direct, social, email)
      * Source/medium (google/organic, direct/none, syr.edu/referral, bing/organic, va.gov/referral)
    - Advertising
      * Advertising (key words, adwords)
      * Dayparting (time of day)
    - Social
      * Facebook
      * LinkedIn
      * Twitter
      * Google+
  + Sensitivity Analysis
    - One-way sensitivity analysis
      * Only one variable changes, allowing you to see the effect on the outcome
    - Two-way sensitivity analysis
      * Two variables changes simultaneously, allowing you to see the joint effect
  + Dashboards in Excel
    - Combine table and graphical representation
  + Conditional Formatting
    - Colors, bars, icons to facilitate interpretation
    - When are some better than others?
  + Eg: what are some difference between 1 way and 2 way sensitivity analysis?'
  + Eg: What are some common techniques in GA?
  + Eg: How would you evaluate the performance of a campaign in GA
* Week 4
  + **Access**
    - Tables
    - Relationships
    - Queries
      * Grouping
        + Collapses together rows of data according to the field grouped
        + It does not form calculations
      * Criteria
        + Identifies a subset of data
      * Calculations
        + Can be used to calculate min, max, sums, average, standard deviations, variance, counts, etc.
  + Eg: What does grouping do? Criteria? Calculations?
    - Grouping does aggregation of data.
    - Criteria creates a subset of data (filter)
    - Calculations creates a new column (numeric, text, logical)
* Week 5
  + PowerPivot
    - Importing
    - Relationships
    - Tables
    - Charts
    - Slicers
      * Creates a menu structure to slice the data by a specified characteristic
    - Timelines
      * Similar to a slicer, but using a date timeline rather than a categorical classification
    - Power Pivot charts
      * Same capability as Pivot Charts
      * Can be created using database tables, Excel spreadsheets and other files
  + EG: What is one main difference between power pivot and regular pivot?
    - Joins and connecting to external data
* Week 6
  + Goal Seek
    - Searches for one goal such as break even point (profit = 0)
    - Allows one variable to be changed in search of the goal
    - Does not allow constraints on the search
  + Solver (constrained and unconstrained)
    - Unconstrained
      * Does not constrain any variables in the search
      * Can search for maximum, minimum, or specific values
      * Can use linear programming (straight line functions) or non-linear programming (curved functions)
    - Constrained optimization
      * Can set variables to less than or greater than some constraint
      * Can set variables to be integer or binary
  + EG: What are some differences?
    - Goal seek 1 variable. No constraints in goal seek. MAX or MIN or specific value in Solver, Goal seek is always 1 specific point. Integers and binary constraints in Solver. Yes/No decision
  + EG: Similarities?
* Week 7
  + R: 3d Viz
    - Statistical summaries
      * Replicates capability seen in Excel: minimum, average, median maximum
      * Also can calculate by brand or other category
    - Histograms, Boxplots, and other charts
      * Boxplots show minimum, 25%-ile, median, 75%-ile, maximum
      * Histograms show frequency of data within intervals
      * Scatterplots add trend line, average, standard deviation
  + ANOVA (Analysis of Variance)
    - Compares the means of two populations
  + Regression
    - Linear regression in Excel is similar to multivariate regression in R
  + Dummy Variables
    - Dummy variables are used to measure the differences in intercepts between two groups, for example, different brands
  + Moderating effects
    - • Moderating effects (interaction effects) measures the difference in slope between two groups
  + EG: What does ANOVA stand for and what does it do?
    - Study of the means, compares the means between two different calculations
  + EG: What is a business example of ANOVA?
    - Difference of Marketing Promotions in a retail setting. Are the difference in means statistically significant?
  + EG: What do dummy variables do to a regression line? Moderating effects?
    - Moderating effects change slope to show acceleration or deceleration
    - Dummy variables change the intercept to show comparison.
      * EG: price differentiation in different product markets.
  + EG: When evaluating dummy variables why is there a variable left out?
    - One is a base level, and we add to the base level each layer.
  + EG: Dummy variables, categorical or other?
    - CATEGORICAL
* Week 8
  + Regression Assumptions
    - Regression Assumption #1: linearity
      * Violation: non-linear data
      * Solution: logarithm, square, inverse, other
    - Regression Assumption #2: X variable are not correlated
      * Violation: multicollinearity
      * Solution: drop or combine variables
    - Regression Assumption #3a: errors are random with constant variance
      * Violation: heteroscedasticity, or wedge shape to error terms in scatterplot
      * Solution: logarithm, square, inverse, or Huber regression
    - Regression assumption #3b: error terms are correlated
      * Violation: serial correlation
      * Solution: rho differencing
    - Regression Assumption #3c: outliers
      * Violation: outlier influences slope of line
      * Solution: drop outlier data points
    - EG: What are the 5 different assumptions and how they look? (screenshot with an outlier, what is the violation?
    - EG: What do you do with the outlier? (solutions)
  + Solutions
  + Benford's Law
    - Financially reported numbers tend to start with smaller digits
  + Decision Trees
    - Use entropy reduction to reduce the amount of error in the data to make a decision
    - Identify the most important variables in making a decision
    - Create a series of rules to make a decision
  + EG: Decision Tree example, weather data. What does it do? Screenshot how would we interpret the numbers?
* Week 9
  + Logit
    - Predict probabilities
  + Probit
    - Logistic distribution
    - More sensitive in detecting differences at extreme values of your variables
  + Perceptrons
    - Normal distribution
    - More sensitive in detecting differences at values near the mean of your variables
  + Neural Networks
    - Subject to Local optima (always different starts, may not be most optimum point)
    - Uses the logistic function to build relationships
    - Also has at least three levels, the X input variables, one or more hidden layers of variables, and the Y output variables
    - To predict the neural network outcome:
      * First, predict the hidden variables from the inputs, just like a logit prediction
      * Second, predict the Y output variables from the hidden variables, again like a logit prediction
  + EG: Difference between Logit and Probit?
    - Logit is logistic dist and more sensitive at extremes values of x variables. Probit is normal dist and more sensitive at values of variables near their means?
  + EG: When would you use linear regression, and when logit or probit?
    - Matters what your Y variable is. If it is continuous we use linear. If it is probability we use logit/probit.
* Week 10
  + Tableau
    - Importing data
    - Creating relationships
    - Tables and charts
    - Dashboards
    - Can connect to Excel, Access, Text files, etc.
    - Joins:
      * Inner joins only create exact matches
      * For left joins, all records in the left table are used to match with those on the right.
      * For right joins, all records in the right table are used to match with those on the left.
    - Differences in data can be highlighted by numbers, picture sizes, colors, etc.
    - Geographic data can be displayed by city, state and country
    - All mathematical calculations can be performed in Tableau
    - Filters can be applied to tables and graphs
    - Dashboards can include tables and graphs simultaneously

Articles

**1.1 Business Analytics Insight: Hype or Here to Stay**

**10 Important Insights**

1. Analytics has a longer history than many think
   1. Decision Support Systems (DSS) invented in late 60s early 70s.
   2. 90s data warehousing (DW) and business intelligence (BI)
2. 'Analytics' means different things to different people
   1. Optimization analytics
      1. Mathematical programming
   2. Predictive Analytics
      1. Decision trees, algorithms, neural networks
   3. Descriptive Analytics
      1. Data visualization, dashboards, drillable OLAP reports, SQL
3. Analytics is becoming a competitive requirement
4. Analytics is overhyped but is here to stay
5. Bid data is changing the scope of technologies for analytics
   1. So much data coming in, it is difficult to capture, store and analyze.
6. BI platforms are changing
7. Analytics are used in new places
8. Analytics requires a diverse set of skills
9. There is a shortage of people with analytics skills
10. Advanced analytics are packaged better

**1.2 GE and the Culture of Analytics**

GE is a conglomerate, but with oil and gas is betting heavy on analytics.

They use two large categories

* Big data: trying to identify from data points if there is a technical issue with a customer asset
* How do you take what available data you have and drive it toward commercial objectives?

Use data for visualization, lean six, sophisticated modelling to determine accuracy of decision and relevance of the market.

Do not increase sophistication, but to instead make them simpler and easier to use

Evaluate different projects ROI for prioritization

Fail fast. Ability to determine potential for success and integrate into ops

Democratize analytics. Fusion of sales and analytics.

**1.3 Location Analytics: Bringing Geography Back**

GIS (geographic information systems) is making a surge and including social, geographic, physical and emotional indicators to better predict trends.

Exploiting the idea of connections based on the location of things

* Why they got there?
* How they got there?
* What does this mean?

Used for city planning, store location

Credit card can track, and determine where and when purchased were made

Social media is unstructured data but a gold mine

Correlation of "clicks to bricks"

Location - where something is. Geography - the context it is in.

In IT, people think big on the T and little on the I

Households, individuals, relation to neighbors (like or unlike)

It is scary and can be misused. Ask not what can you do with the information about a person but instead what can you do for that person?

**Article 2.1: Sustaining an Analytics Advantage**

* **What are some examples of creating competitive advantage with analytics (companies and their techniques)?** 
  + Wal-Mart and supply-chain management algorithm, while specifically concealing the details of its analytics. In-house analytics team that stays together.
  + ABB Electric implemented analytics fast to defeat competitors before they can react (1970s). Developed sophisticated customer-choice analytics.
  + Proctor & Gamble used analytics to reengineer its global supply chains. While other companies focused analytics on "low-hanging fruit," P&G focused on big systems change.
  + ICBC partnered with IBM to develop branch network optimization system and has enabled them to be the largest bank in the world.
  + American Airlines sold its cost saving algorithms but safeguarded its own data. Value of the algorithm came from vast library of data.
  + Amazon, FedEx, and IBM applied analytics to so many small projects that they have truly transformed into a "data-driven" company.

**Article 2.2: Creating Business Values with Analytics**

* **What are the differences between competencies in information management and analytics expertise?** 
  + Level of reliance. Aspirational, Experienced, and Transformed.
  + The higher the level of reliance, the higher the degree of competency in both areas (information management and analytic experience). Mutually reinforcing.
  + Analytic expertise built from talent, tools and tech where information management is built from strong data governance, data management practices and the capability to deliver the right information to the right people at the right time.
* **What are the advantages of focusing first on information management versus analytics expertise?** 
  + Data creation and sharing. Data integration for better results. Specialized and collaborative organizations seek this.

**Article 2.3: Raising the Bar with Analytics**

* **What new opportunities did StyleSeek and Entravision encounter when they used analytics? What opportunity allowed MillerCoors to create efficiencies with analytics?** 
  + Organizations were hungry to be a part of their platform, and eager to learn about the insights on their data. (StyleSeek)
  + Already boasting a large user base, the ability to share and market data to external sources helped create insights that better targeted customers wants, and became valuable enough to shift the focus towards this division, (Entravision)
  + Analytics transformed the business by helping make better decisions to invest resources more effectively. (MillerCoors)

**Article 3.1: Web Analytics: Enhancing Customer Relationship Management**

**1. Describe the four main categories of metrics and relate to the Google analytics lessons**

* **Website usability**
  + Evaluates items such as page views, time on site, and click paths to determine how user-friendly or user-relevant a web site is.
* **Traffic sources**
  + Metrics identify traffic origination points, such as referral web sites or even offline advertising campaigns.
* **Visitor profiles**
  + Data from visitor profiles can provide information such as geographical origination of traffic, the time of day users most frequently visit, or what keywords are used in reaching the site
* **Conversion statistics** 
  + Measure which visitors are new, returning, or abandoning the site, as well as which are actually completing sales. Web analytic programs can arrange relevant data into a convenient dashboard for regular metric monitoring,

**2. Describe the common techniques for Web analytics**

* **Clustering/classification**
  + Clustering, or classification, is a means of developing profiles of items with similar characteristics. This ability enhances the discovery of relationships that are otherwise not obvious. For example, classification of Web access logs allows a business to discover the average age of customers who order a certain product. This information can be valuable when developing advertising strategies.
* **Association rules**
  + Association rules are the ones that govern data warehouses "Bid Data" of transaction where each transaction consists of a set of items. This technique is used to predict the correlation of items where the presence of one pattern of items in a transaction implies (with a certain degree of confidence) the presence of other items. For example, association rules can provide a prediction of the percentage of clients assessing a particular site who will place online orders for a certain product.
* **Path analysis**
  + Path analysis is a technique that involves the generation of some form of graph that represents relation[s] defined on Web pages. This can be the physical layout of a web site in which the web pages are nodes and the hypertext links between these pages are directed edges. Most graphs are involved in determining frequent traversal patterns or large reference sequences from physical layout, such as the most frequently visited paths in a website. Another example is what paths users travel before they go to a particular URL.
* **Sequential patterns** 
  + This technique is applied to Web access server transaction logs. The purpose is to discover sequential patterns that indicate user visit patterns over a certain period. For example, 25% of clients who visited a site had performed a search within the past week on a specific keyword or a product.

**3. What are some business applications of web analytics?**

* Evaluate promotional campaigns.
* Provide users dynamic information based on their interests.
* Target electronic ads based on user patterns.
* Design more effective marketing strategies.
* Predict user behavior.
* Provide better customer care and customer intimacy.
* Maintain and manage customer base.

**Article 3.2: How eBay Uses Data and Analytics to Get Closer to Its (Massive) Customer Base**

**1. What is an A/B test and what is its purpose?**

* A/B testing is comparing multiple versions of a webpage or an app to determine which version performs better.

**2. Describe the three biggest challenges of web data**

* **Data at a large scale**
  + Building or implementing hardware and software systems that can handle large scale data
  + Managing scale as you grow in data collection
* **Collecting the right data**
  + Tagging software systems to collect the right kind of data
  + Cleaning and cutting data the right way
* **New kinds of data** 
  + Being able to deal with new kinds of data as technology evolves (smartphone devices, image and video data, etc.)
  + Need to be prepared to process new forms of data really well

**3. How can Power Sellers use data better?**

* Naturally occurring experimentation.
* Power sellers are very smart about their business and know what to do. They are navigating the best routes and you can learn directly from them

**4. Why are web analytics better than surveys?**

* "Squeaky wheel" problem with surveys.
* User behavior through web analytics tells user behavior in a systematic way. "Friction points"

**Article 4.1: Minding the Analytics Gap**

1. **What is a barrier to using analytics?** 
   * **translating analytics into business actions**
2. **What can be done on the production and consumption side of analytics to overcome this barrier?** 
   * **From the production side, data analysts can learn more about the business. Organizations can also systemically improve infrastructure and processes; improved data quality, for example, can make it easier to turn data into competitive advantage**
   * **From the consumption side, managers can also take steps to become savvier at understanding analytical results. Beyond training, other known steps include identifying trustworthy analytics professionals within the organization, requiring straightforward explanations and asking detailed questions.**
3. **Describe the three levels of analytics maturity** 
   * **Analytically Challenged**
     + **generally rely more on management experience than data analysis and tend to lack data management and analytical skills**
   * **Analytical Practitioners**
     + **tend to use analytics for operational purposes, have “just good enough data” and are working to become more data driven**
   * **Analytical Innovators**
     + **more strategic in their application of analytics, place a high value on data, and have higher levels of data management and analytical skills**

**Article 4.2: Innovating with Analytics**

1. **Describe the three characteristics of analytics innovators** 
   * **Analytical Innovators Tend to Use More Data**
     + **strong correlation between how much a given company uses analytics to create competitive advantage and advance innovation and how much of their data that company uses**
   * **Analytical Innovators Manage Information More Effectively**
     + **strong correlation between driving competitive advantage and innovation with analytics and a company’s effectiveness at managing the information transformation cycle, that is: capturing data, analyzing information, aggregating and integrating data, using insights to guide future strategy and disseminating information and insights**
   * **Speed Is Very Important to Most Analytical Innovators**
     + **very important to have the ability to process and analyze data more quickly**

**Article 5.1: Innovating with Airborne Analytics**

1. **What are the three areas of focus for their strategy for the next 3-4 years? Give examples of what they are doing.**
   1. Operational Efficiency (or improved performance)
      1. Using analytics on engine performance data for the fleet, increasing reliability and reducing fuel burn (40% of costs)
      2. Optimize crew and shift deployment (match to plane types, destinations, special breaks)
   2. Customer Intimacy
      1. Frequent flyer preferences and assess flying preferences
      2. Click stream data on the site to understand stage in cycle customers are at
      3. Complaint data for trends (meals, seats, service, etc.)
   3. Innovation
      1. Using the data on customer preferences to tailor services (streamline cache for newspapers to be available)
      2. Innovation center in IT (developing a tag for baggage tracking)
2. **What are the challenges with hiring data scientists?** 
   1. They are expensive, not needed all of the time. Some complex predictions take months and therefore is sometimes too far out for expense to be worth it
3. **What are the risks trying to use technology in the organization? Give an example of each**
   1. Interoperability
   2. Data security
   3. Proper taxonomies
   4. Data ownership
   5. Good judgement on data produced

**Article 5.2: A New, Analytics-Based Era of Banking Dawns at State Street**

1. **What were the key questions that they asked which analytics could help?** 
   1. How do I better manage, measure, visualize and report on risk?
   2. How do I deal with all of this regulatory change?
   3. How do I continue to find new sources of return?
   4. What new types of investments, asset classes and geographies do I have to expand into to achieve better yields?
   5. How do I keep my costs down, and how do I make sure I’m getting a fair shake in my trading activity?
2. **What was the challenge of launching analytics at State Street?** 
   1. Very large company can cause decisions to take time and require a lot of discussion
   2. Inspiring everyone to become part of the new program, as to remaining opposed or culturally assimilated to the different component org they came from
   3. Nimble. A chance to evolve the culture that allows it to be more adaptive in the future
3. **How do they continue to push analytics forward?** 
   1. Lead by example, executive level support, good communication from the start, desire to stay ahead and continue to grow

**Article 6.1: Modern Analytics and the Future of Quality Performance Excellence**

1. **Define analytics (page 6)** 
   1. The use of data, information technology, statistical analysis, quantitative methods, and mathematical or computer-based models to help managers gain improved insight about their business operations and make better, fact-based decisions.
2. **How are companies using analytics in (page 7)?** 
   1. **Banks**
      1. Predict and prevent credit fraud
   2. **Manufacturing**
      1. Production, planning, purchasing, inventory management
   3. **Retail**
      1. Recommend products to customers and optimize marketing promotions
   4. **Pharmaceuticals**
      1. To get life-saving drugs to market more quickly
   5. **Sports** 
      1. To determine both game strategy and optimal ticket prices
3. **Modern analytics integrates which three fields (page 8)?** 
   1. Business Intelligence/Information Systems
   2. Statistics
   3. Quantitative Methods/Operations Research
4. **What are some examples of data sources (page 9)?** 
   1. Numerical, textual, audio, video
5. **What are examples of data visualization (page 11)?** 
   1. Space and time, multivariate, text, graph, network.

**Article 6.2: A Process of Continuous Innovation: Centralizing Analytics at Caesars**

1. **Why does Caesars use analytics (pages 1 & 2)?** 
   1. Ultimately, they believe and are devoted to analytics in order to build a deeper understanding of customers as well as operations.
2. **What are four lessons learned from their experience (page 3)?**
   1. Scaling a large analytics team requires appropriate leaders to run them, and the leaders need to be as deliberate as possible
   2. Infrastructure takes time to create (willingness to relearn)
   3. Communication is necessary, and requires a tremendous amount of time
   4. Have visible and meaningful wins throughout the process

**7.1: Big Data in Health Care: Using Analytics to Identify and Manage High-Risk and High-Cost Patients**

* **What are the six opportunities to reduce costs through analytics? How can cost be reduced in each?** 
  + High-cost patients
    - 5% of patients account for 50% of US health care spending
      * Identify and manage these patients more effectively
        + Doesn’t always lower costs…
      * Better analytic systems to effectively identify potentially high-cost patients
      * Many cases of outcomes in predictive models come from low-risk groups
        + Need for more accurate modeling
  + Readmissions
    - Frequency and high-cost of hospital readmissions
      * Use an algorithm to better predict readmission
      * Use millions of patients records
  + Triage
    - Estimating risk of complications
    - Reduce the risk of adding unnecessary medication
  + Decompensation(when a patient’s condition worsens)
    - Determine patients risk for decompensation
  + Adverse events
    - Predict when patients are at risk of adverse events of several types
    - Adverse events are expensive and cause substantial morbidity and mortality, yet many are preventable
  + Treatment optimization for diseases affecting multiple organ systems (such as autoimmune diseases, including lupus)
    - Chronic conditions that expand into multiple systems are costly
    - Accurately predicting the trajectory of a patients disease could allow the caregiver to better target complicated and expensive therapies

**7.2: A Review of Analytics and Clinical Informatics in Health Care**

* What are some methods for improvement in health care using analytics? (page 2)
  + Utilizing risk assessment analytics to process HER data to identify the patients at greatest risk for utilizing more resources than their peers with the goal of improving patient outcomes and managing costs
  + Potential value of aggregating data enhanced with real-time analytics to provide point-of-care information to oncologists that was tailored to individual patients
  + Application of predictive analytics for better targeting of disease management and innovative patient care approaches while also warning of the unintended consequences that may arise.
  + Enhance less sophisticated rules-based systems already in use
  + Financial cost savings based on improved patient care and outcomes, as well as identification of simple billing anomalies.
  + Better manage resource allocation
* What are some challenges for analytics in health care? (page 4)
  + Managing massive data files
  + Poor data application could produce faulty conclusions
  + Demand for medical analytics professionals

**8.1: What Businesses Can Learn from Sports Analytics?**

* **Describe the five key lessons of analytics in sports (give an example of each)** 
  + Align leadership at multiple levels
    - Analytics needs to be involved with the corporate team/coaches as well.
  + Focus on the human dimension
    - There is all the normal stats, but what about how a group performs with or without a specific member
  + Exploit video and locational data
    - Video and GPS data. Habits and behaviors of customers
  + Work with a broader ecosystem
    - May be hard to hire analysts, so work with data technologies
  + Support "analytical amateurs"
    - Allow data and analytics to be in the hands of your employees, sot hey can measure themselves

**8.2: Team GB: Using Analytics (and Intuition) to Improve Performance**

* **What is the value of predicting team performance? (page 2)** 
  + You can compare your performance to expected performance, as well as to the competition.
* **What is the biggest challenge? (page 2)** 
  + Difficulty in collecting the data
* **What are some of the barriers? (page 3)** 
  + Combination of experience of the coaches and their terms of reference when planning training and the lack of information on how robust new data are when it comes to decision making
* **Where is the power of the data? (page 5)** 
  + Good longitudinal information rather than snapshots

**9.1: An introduction to data mining and other techniques for advanced analytics**

* **What are the key differences between statistical analysis and data mining? (page 140)** 
  + The data miner is no longer restricted to working with ***small samples***, different in approach from traditional statistical methods — ***statistical techniques may give misleading results if applied to a vast sample size****,* which carries risks of overﬁtting the model or producing unhelpful results in which every variable appears to be statistically signiﬁcant.
  + In DM, the dataset is liable to contain a huge number of candidate predictor attributes (variables), for example, volumes and values of transactions by product, channel, brand or period — far too many to be individually assessed and transformed manually.
    - ***DM solutions ideally provide automated tools for selecting relevant attributes and recoding them in the form of variables for use in analysis****.*
  + A further key difference is that ***statistical analysis will aim to identify a model that is statistically signiﬁcant — that is, outperforms a random prediction — based on a set of signiﬁcant predictor variables.*** However, this provides no guarantee that the model will perform sufﬁciently well to be of business value. ***DM goes further than that, by including diagnostic results to indicate likely business beneﬁts from the model.*** The assessment is produced by using two methods in combination:
  + Lastly, having built and evaluated a DM model on a sample dataset, the model will be deployed by applying the scoring algorithm to all ‘ X ’ million records in the customer database. Therefore, facilities for large-scale model deployment are essential — the form that this takes will vary from package to package, as we will see below.
  + Both DM and statistical analysis require that the data are organized as a simple rectangular table, where the rows (or records) represent individuals (eg customers) and the columns contain structured variables (eg demographics, usage or purchasing behavior). Often, much effort is required in order to assemble this analytic dataset, as we discuss in the following section.

* **Describe tools for advanced analytics (page 149-151)**
  + **Data visualization**
    - Neural Networks
    - Genetic Algorithms
    - Scatter plots
    - Heat Maps
    - Maps
  + **Text mining**
    - Linguistic analysis
    - Tools from Attensity, clarabridge, IBM (SPSS), KXEN, SAS
  + **Social network analysis**
    - Uses Nodes and Links
    - Churn Predictions
    - Tools from Idiro, KXEN
  + **Contact optimization** 
    - Extension of CRM system
    - Sits "above" DM
    - Tools from Experian, SAS, IBM (SPSS), TCP Marketing Solutions, and Unica
* **How do you mitigate the risks of data mining? (page 152)** 
  + Good Data Quality
  + Strong Business Focus
  + Sound User Training

**10.1: Business Analytics: Transforming the Role of Management Accountants**

* **What are some external and internal data sources for accountants? (page 3)** 
  + Internal
    - Files
      * Spreadsheet Files
      * CSV
      * Microsoft Access
    - Databases
      * SQL Queries
      * ERP data
      * Data warehouses
  + External Sources
    - Internet Sources
      * Google Analytics
      * SEC XBRL database
      * Salesforce
      * Zendesk
  + Other Data\*\* (not really these)
    - Sensors, videos, tweets, emails
* **What are four challenges for accountants using analytics? (page 4)** 
  + Awareness
    - Understanding of how accessible and valuable BA is to their companies
  + Interoperability
    - Some data is structured, some unstructured. There is a movement to make them all structured and linkable, but it is not the standard
  + Security
    - Data is an asset and needs to be protected
  + Analysis Quality
    - Focus on data integrity and well defined objectives
    - Garbage in garbage out philosophy.
* **What are five areas for leveraging analytics in accounting? (page 4)** 
  + franchise sales analysis
    - BA to analyze sales metrics to better use POS promotions
  + accounts receivable and credit analysis
    - Tracking DSO.
    - AR collection history
  + accounts payable analysis and payment monitoring
    - Monitoring payments against payment-decision rules that flag fraud
  + mergers and acquisitions (M&A) due diligence
    - Accurate bus. Valuations for M&A
  + forensic accounting
    - Fraud Detection Analytics (FDA) strategies to identify fraud, bribery, and corruption in companies.

**10.2: Elevating Data, Analytics to the C-Suite**

* **What are the steps to elevate a department using analytics? (page 5)** 
  + Have your data first (not always perfect data)
  + Then, have your reporting
  + Next, have your analytics,
  + Last your quantitative and predictive modeling.
* **How should you address non-perfect data? (page 5)** 
  + Has to be good, clean but never perfect. Think of it as an 80% solution. A decision is being made regardless, and the data can help guide that decision even if imperfect.
  + Analysts have to articulate the limitations of data.
* **Should analytics teams be centralized or decentralized? (page 6)** 
  + Centralizing has all the benefits of better training, skill growth, and career paths, but has the down side of losing connectivity to the business. Decentralization has the exact opposite issues.
  + He recommends to be decentralized in a large org, centralized in a small one?