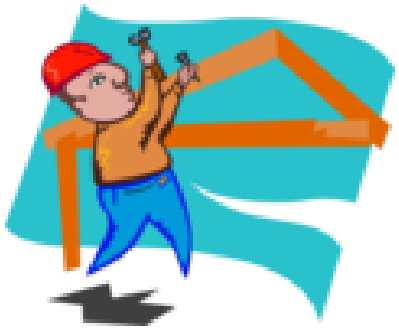
BUILDING ANALYTICAL APPLICATIONS



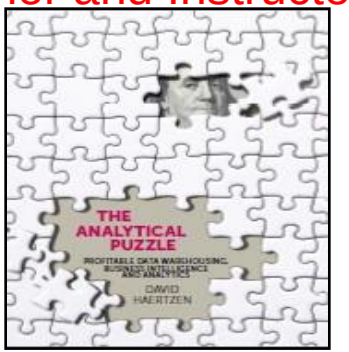
Presented by
David Haertzen
First Place Learning

About the Author

- Enterprise and data architect
- Provided services to organizations such as: Allianz Life, 3M, Mayo Clinic, IBM, Fluor Daniel, Procter & Gamble and Synchrono – from start up to multinational.
- Experienced author
- Frequent presenter in the areas of:
  - Data modeling
  - Data warehousing
  - Enterprise architecture
  - Analytics and Business Intelligence
  - SQL
- Instructor for First Place Learning and eLearningCurve
- University of Minnesota
  MBA, University of St Thomas
- **Visit: http://www.firstplacelearning.com/**
- **Visit: http://www.linkedin.com/davidhaertzen**
- **Visit: http://ecm.elearningcurve.com/David_Haertzen_s/89.htm**

David Haertzen

Author and Instructor



Topic Objectives

Upon finishing this presentation you will:

- Understand what Analytics is, its goals, and its components
- Know some of the most profitable applications of analytics
- Know how to organize an Analytics Project using the CRISP-DM methodology
- Know how to evaluate the effectiveness of an analytical model
- Understand what is involved with deploying an analytic application
- Be prepared to learn more about Analytics

Session Structure

Topic 1: What is an Analytical Application?

- What is Analytics?
- Predictive Analytics Impacts the Bottom Line
- Analytical Applications Examples

Topic 2: Analytical Methodology

- Analytics Methodology
- Team Roles
- Data Preparation
- Analytical Modeling
- Model Evaluation
- Deployment - Productionizing

Topic 3: Analytics Architecture

- Analytics Architecture Components
- Increasing Analytics Performance

Topic 4: Analytical Application Examples

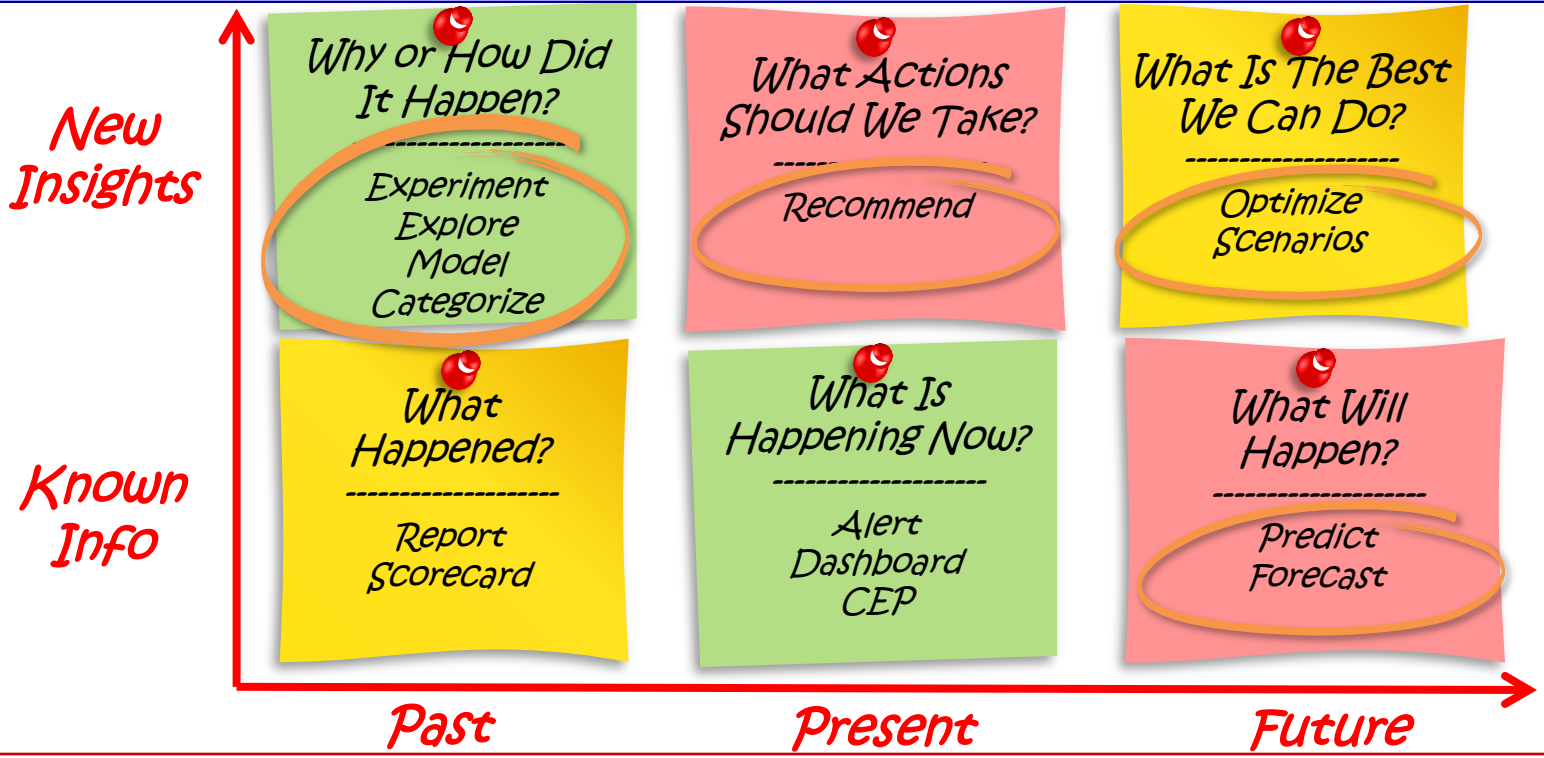
- Financial Services Applications
- Retail Analytical Applications
- Recommendation Engine

Topic I:
What is an Analytical Application?

- What is Analytics?
- Decisions Impact the Bottom Line
- Example Analytical Applications

What is Analytics?

Analytical Applications are software components that apply algorithms to data to enable improved understanding and decision making.



Inspired by Tom Davenport

Predictive Analytics Impacts Bottom Line

Predictive Analytics can have a major impact on profit: increasing sales, decreasing costs and mitigating risks. This mailing campaign is a good example.

Without Predictive Analytics	
Customer Sale Percent	2.00%
Cost Per Mailing Piece	$1
Mail Count	100,000
Mailing Cost	$100,000
Unit Sale Price	$100
Gross Margin Per Sale	$80
Est Sales Units	2,000
COGS	$40,000
Gross Margin Total	$160,000
Profit after Mailing Cost	$60,000

Improving Customer Response Percentage by 7% improves profit by 417%

With Predictive Analytics	
Customer Sale Percent	9.00%
Cost Per Mailing Piece	$1
Mail Count	50,000
Mailing Cost	$50,000
Unit Sale Price	$100
Gross Margin Per Sale	$80
Est Sales Units	4,500
COGS	$90,000
Gross Margin Total	$360,000
Profit after Mailing Cost	$310,000

TARGET PROGRESSIVE CapitalOne Allianz BEST BUY Google WELLS FARGO Walmart amazon.com usbank

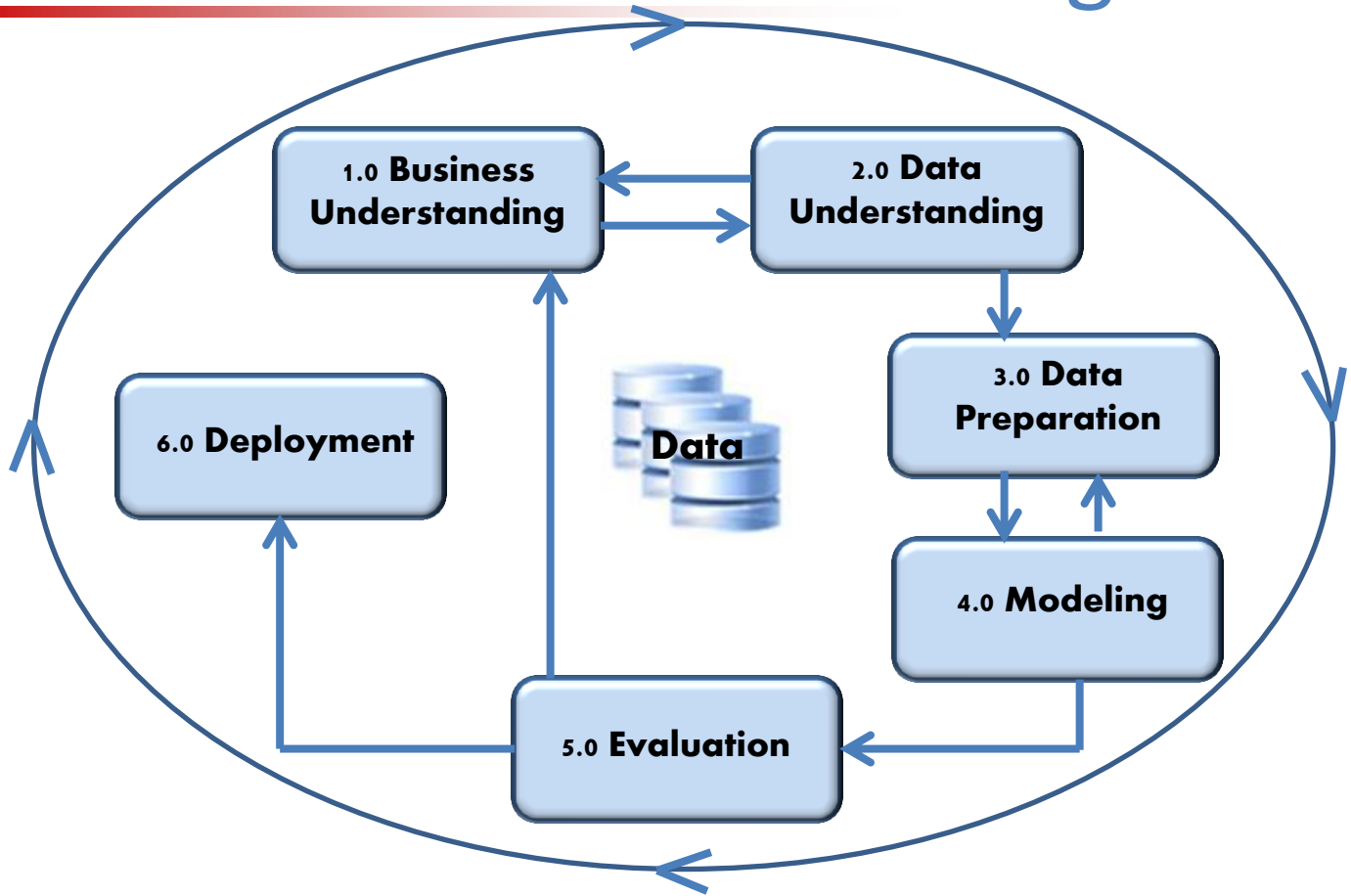
Analytical Application Examples

Model Type	Description
Acquisition Model	A model that predicts the probability that a prospect will buy the company's products or services.
Cross-sell Model	A model that predicts the probability that an existing customer will buy additional products or services of a different type than currently bought. Goods are at the same level.
Up-sell Model	A model that predicts the probability that an existing customer will buy an upgraded product or service.
Attrition Model	A model that predicts the probability that an existing customer will stop purchasing the company's products or services. This also known as a churn model.
Value Model	A model that predicts a numeric value such as customer lifetime value (CLV) or value resulting from the sale of a specific product to a customer.
Tone-Of-Voice Model	A model that identifies the most effective message for each targeted customer.
Risk Model	A model that predicts potential negative activities by customers such as: fraud, loan defaults, or excess service costs.
Customer Segmentation Model	A model that assigns customers to groups with similar characteristics.
Recommendation Engine	A model that provides advice on a near real-time basis – such as advice about offers that should be made to a customer or additional products to show to a customer.
Look-alike Modeling	A model where the target-marketed group (e.g. for a marketing campaign, product offering etc.) is an expanded list of parties whose profiles look like the selected party.

Topic II:
Analytics Methodology

- CRISP-DM Data Mining Methodology
- Team Roles
- Data Preparation
- Analytical Modeling
- Model Evaluation
- Deployment - Productionizing

CRISP-DM Data Mining Methodology



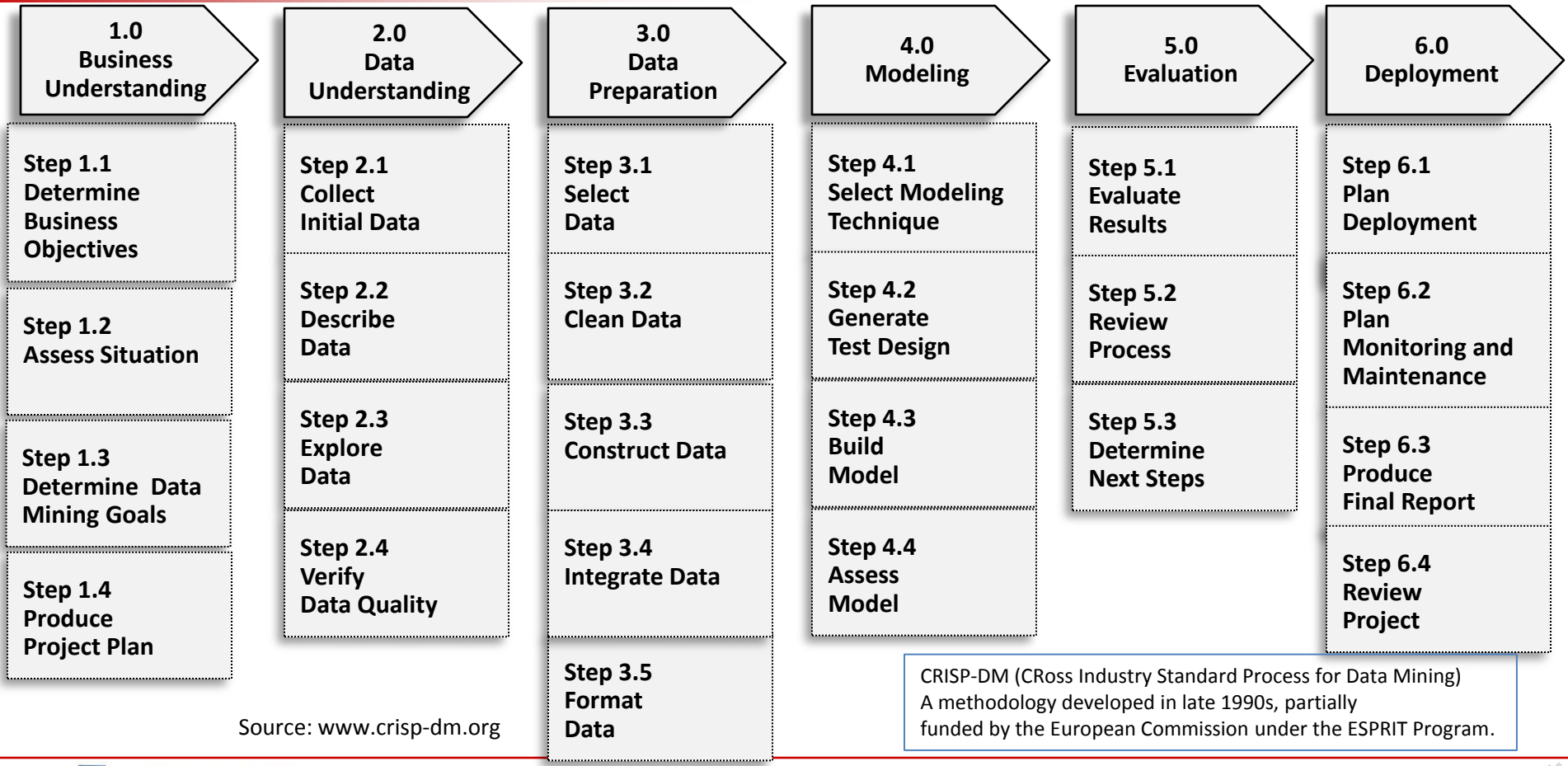
CRISP-DM (CRoss Industry Standard Process for Data Mining)

A methodology developed in late 1990s, partially funded by the European Commission under the ESPRIT Program.

Contributors:

- NCR Systems Engineering
- SPSS Inc.
- DaimlerChrysler
- OHRA Verzekering en Bankk Groep

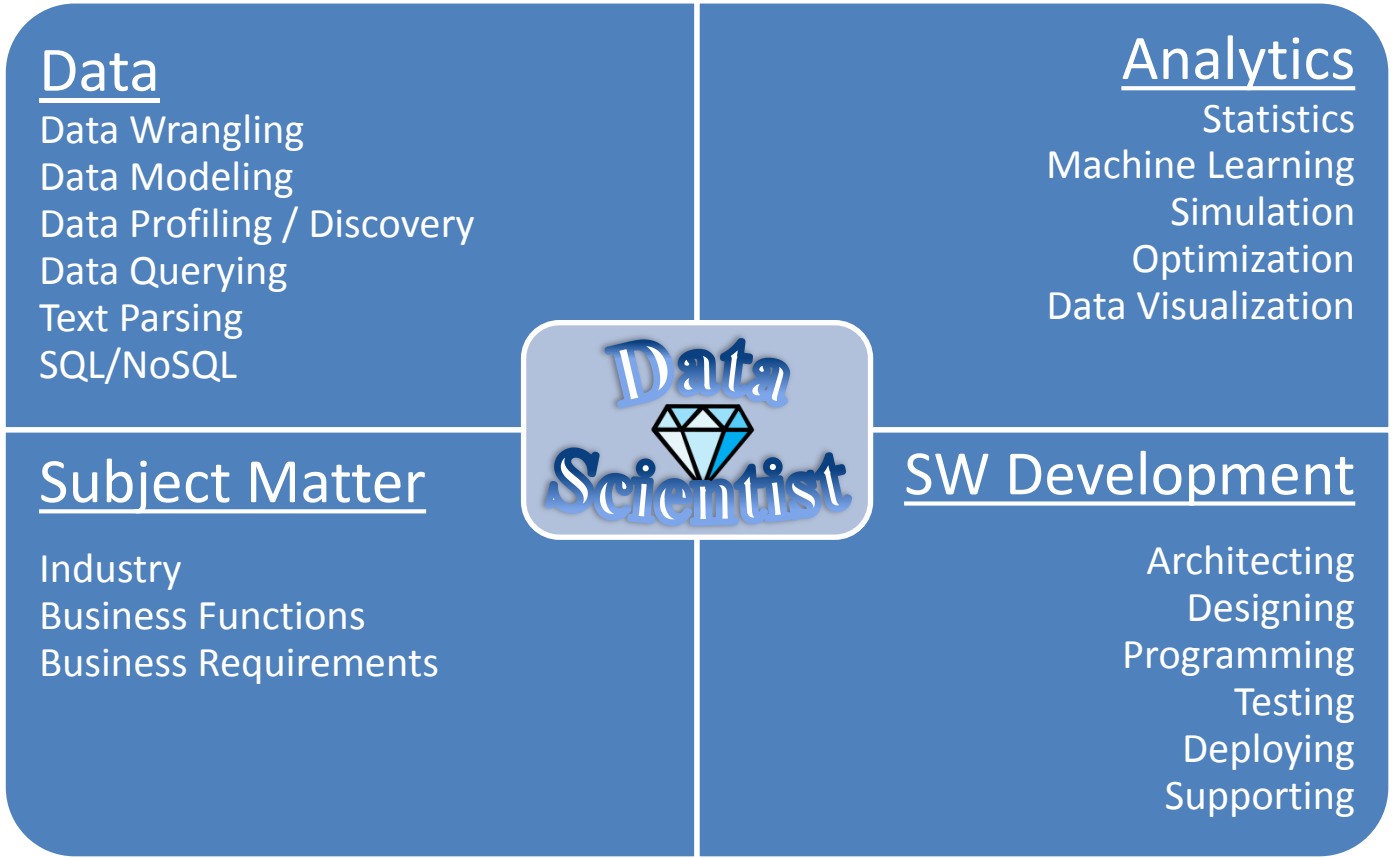
CRISP-DM Flow



Source: www.crisp-dm.org

CRISP-DM (CRoss Industry Standard Process for Data Mining)
A methodology developed in late 1990s, partially
funded by the European Commission under the ESPRIT Program.

Data Scientist – Quadruple Threat



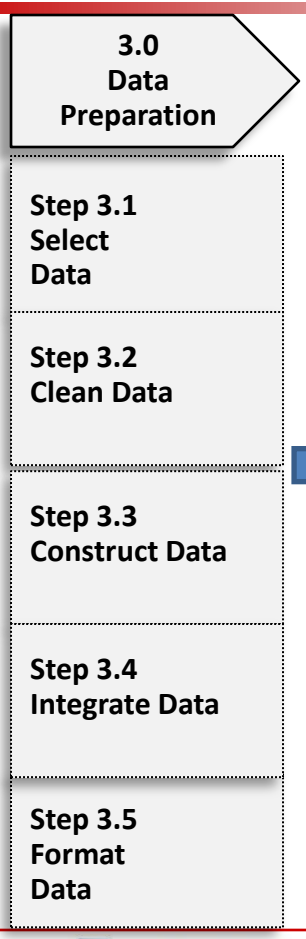
Analytical / Data Roles

Qualification	ETL Developer	Database Admin	Data Modeler	Data Architect	Data Developer	BI Developer	BI Analyst	Data Analyst	Data Scientist
Education Level	Tech	Tech	Bachelor	Bachelor	Tech or Bachelor	Tech or Bachelor	Bachelor	Bachelor	Master to PhD
Business Understanding	Low	Low	High	High	Medium	Medium	High	High	High
SQL and Data Languages	High + SQL Procedures	High + SQL Procedures	Medium	Medium	High + SQL Procedures	High	High	High	High
Data Modeling	Medium	Medium	High	High	Medium	Medium	Medium	Medium	High
Database Admin Tools	Medium	Highest	Low	Low	Low	Low	Low	Low	High
Data Exploration and Preparation	High	Medium	Medium	Medium	Low	Medium	Medium	High	Highest
BI Reporting	N/A	N/A	Low	Medium	Medium	High	High	High	High
Visual Analytics	N/A	N/A	Low	Medium	Medium	High	High	High	Highest
Statistics and Prediction	N/A	N/A	Low	Medium	Low	Low	Low	Medium (Uses GUI)	High
Programming	N/A	N/A	N/A	Low	Java, C#	Java, C#	N/A	N/A	R, Python, Java + more

More Critical Roles

Qualification	Executive Sponsor	Project Manager	Enterprise Architect	Infrastructure Architect	Security Architect	Business Architect	Business Analyst	QA Leader	Infrastructure Support
Education Level	Master	PMI Cert	Master	Bachelor	Bachelor + Cert	Master	Bachelor	Bachelor	Tech
Project Management	Medium	High	Medium	Low	Low	High	Medium	Medium	Low
Business Accumen	High	High	High	Low	Medium	High	High	Medium	Low
Enterprise Architecture	Low	Medium	High	High	High	High	Medium	Medium	Low
Hardware / Software Stack	Low	Medium	Medium	High	Medium	Medium	Low	Medium	High
QA	Low	Medium	Medium	Low	Medium	Medium	Medium	High	Medium
Customer Support	Medium	Medium	Medium	Medium	Medium	High	High	Medium	High
Infrastructure Management	Low	Medium	Medium	High	Medium	Low	Low	Low	High
Communication / Presentation	High	High	High	High	High	High	High	Medium	Medium

3.0 Data Preparation



3.1 Select Data

- Consider business goals and objectives
- Consider technical constraints
- Consider sample size and boosting

3.2 Clean Data – raise data quality level

- Handle incorrect and missing values

3.3 Construct Data

- Calculate derived data
- Generate missing records such as zero orders

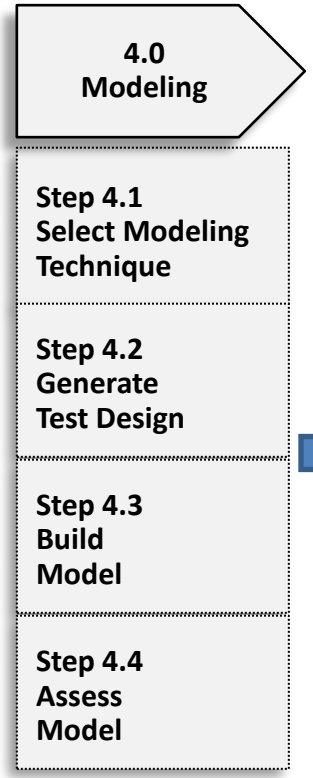
3.4 Integrate Data

- Combine data from multiple sources or rows
- Create a flat data structure

3.5 Format data – modify input for data mining models and tools

- Re-sequenced attributes
- Re-sequenced records (change sort order)
- Modified datatypes such as floating point to integer
- Binned data to reduce noise

4.0 Modeling



4.1 Select Modeling Techniques

- Determine which modeling techniques to use
- Document assumptions about the model such as data distribution

4.2 Generate Test Design

- Determine how testing will be performed
- Decide how data will be allocated to training, test and validation datasets

4.3 Build Model

- Generate the model using a modelling tool or manually create
- Document the parameters generated by the tool: formulas, scoring, decision conditions

4.4 Assess Model

- Determine and document accuracy and generality of models
- Document the computing performance of the models
- Order the models by accuracy and performance

Select Modeling Technique

Criteria

- Business objectives
- Inputs are ratio numbers
- Inputs are discrete values
- Inputs are multiple events or actions
- Objective is a ratio number
- Objective is a category value such as yes or no
- Multiple scenarios may occur
- Mix optimization
- Number of inputs

Modeling Techniques

- Decision Tree
- Regression
- Neural Net
- Affinity Analysis / Clustering
- Montecarlo Scenario
- Graph Optimization
- Multiple Model Voting (Ensemble)

Dimension Reduction

Dimension Reduction is the process of simplifying input factors to predictive analytics algorithms to reduce the number and/or complexity. The process may reduce 100s of factors to a handful.

Methods:

- Drop Missing Values
- Drop Low Variance
- High Correlation
- Backward Feature Elimination
- Factor Analysis
- Principal Component Analysis (PCA)

Benefits:

- Calculations are faster.
- Storage space needed is reduced.
- Models are easier to explain.
- Model is easier to productionize.

Examples:

Wine Price depends on:

- Winter Rainfall
- Average Growing Season Temperator
- Harvest Rainfall

UPS Route Safety depends on Left Turns

Dimension Reduction
http://www.analyticsvidhya.com/blog/2015/07/dimension-reduction-methods/?utm_source=feedburner&utm_medium=email&utm_campaign=Feed%3A+AnalyticsVidhya+%28Analytics+Vidhya%29

Assessing Numeric Predictions

Assessing Numeric Predictions includes a determination of the accuracy of the prediction compared to actual.

- Descriptive statistics – standard deviation, variance, etc.
- Quantify cases where prediction is inside and outside business tolerance.

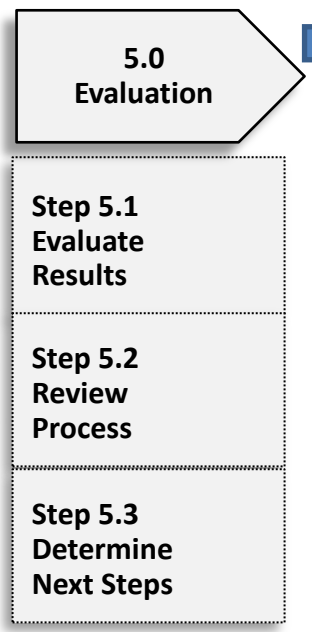
Assessing Classification Models

A Confusion Matrix is a method for evaluating Classification Models that quantifies the number and proportion of correct and incorrect classifications through use of a table.

- True Positive (TP)
- True Negative (TN)
- False Positive (FP)
- False Negative (FN)
- **Accuracy** = (TN+TP)/n (60 + 105) / 190 = 87%
- **Error Rate** = (FN+FP)/n (10 + 15) / 190 = 13%

n=190	Predicted: No	Predicted: Yes	
Actual: No	TN=60	FP=15	75
Actual: Yes	FN=10	TP=105	115
	70	120	

5.0 Evaluation



5.1 Evaluate Results

- Compare model results to business objectives
- Identify any business objective shortcomings
- Approve model(s) that meet business criteria

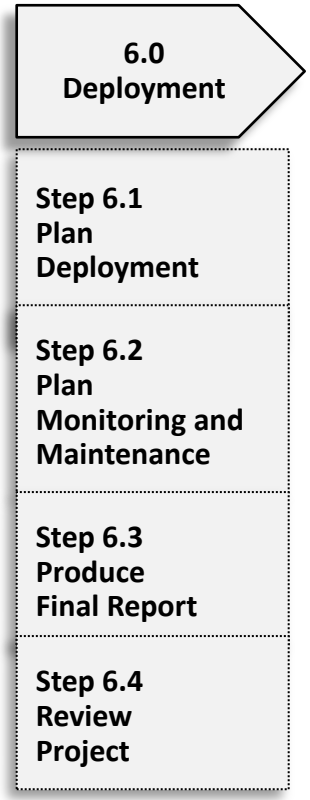
5.2 Review Process

- Conduct "Lessons Learned" review of the modeling process

5.3 Determine Next Steps

- Determine if iterations are needed
- Determine if ready for deployment

6.0 Deployment



6.1 Plan Deployment

- Create deployment plan document
- Consider model findings to create implementation approach

6.2 Plan Monitoring and Maintenance

- Determine how the deployed application will be controlled
- Determine how the deployed application will be maintained

6.3 Produce Final Report

- Produce a report that summarizes project activities and findings

6.4 Review Project

- Determine lessons learned from the overall project
- Add to best practices knowledgebase

Deploying / Productionizing

Deploying is the process of moving prototype computer applications to industrial strength applications that run in a managed environment.

Modeling and Evaluation	Deployment
Data broken into development and test groups	Data is all production data
Data may be boosted – heavily weighted for exceptions	Data is not boosted and will contain fewer exception
Documentation supports research and experimentation	Documentation supports ongoing operations and decision making
Run from test harness – analytical tool / IDE	Batch or plugged into transactional systems
Developed using modeling tools: Excel, Python, SAS, R, Visual IDE	Translated to IT language: Java, C#, COBOL, SQL
Run from workstation or desktop	Run from production server in a managed, secure environment that includes error logging and monitoring
Batch processes are run manually as needed	Batch processes are automated and scheduled
Application is available on an intermittent basis	Application is highly available and supported by disaster recovery

Topic III:
Analytics Architecture

- Analytics Architecture Components
- Increasing Analytics Performance

Analytics Architecture Components

Data Sources


Data Brokers


Social Media

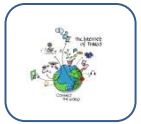

Operational Systems

Text Files

Images


Logs


Government


Internet of Things


Spreadsheets

More ➤

Logical Data Lake


Operational Systems


EDW


Data Marts

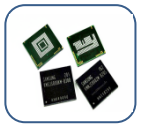

Hadoop


NoSQL


Sand Boxes

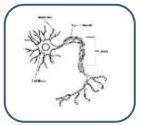

tbd

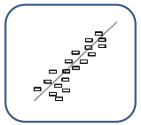

Cloud


In Memory

More ➤

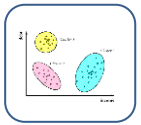
Analytical Models


Neural Network

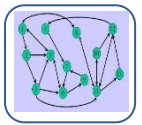

Regression

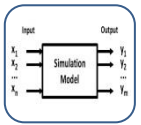

Decision Tree


Machine Learning


Cluster


Montecarlo


Optimization



Simulation


Forecast

More ➤

Uses


Report


Dashboard


Scorecard


Recommend Engine

Anti-fraud


Campaign Management


Algorithmic Trading


Customer Intelligence


Location Analysis

More ➤

Increasing Analytics Performance

Database Technology:
- ✓ SQL Traditional Database
- ✓ Data Warehouse Appliance
- ✓ Columnar Database
- ✓ In Memory Database
- ✓ OLAP / Cube Database
- ✓ NoSQL Database

Scale It Out:
- ✓ Grid - Data Synapse
- ✓ In memory Grid - Apache Ignite, others
- ✓ DIY Grid
- ✓ GPU - CUDA, OpenCL, BOINC
- ✓ Supercomputer / Minisuper Computer
- ✓ Hadoop - bring calculations to the data
- ✓ Cloud

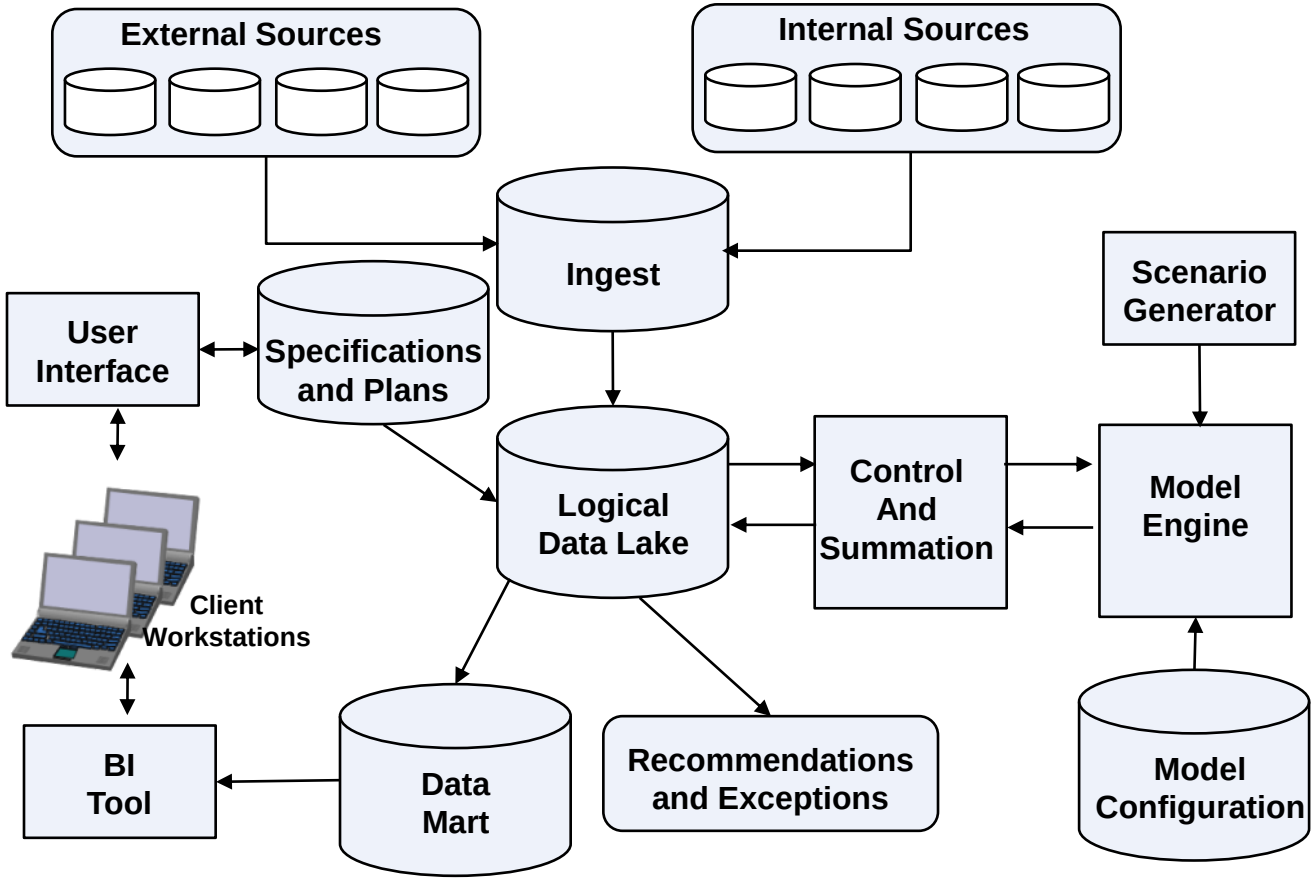
Scale It Up:
- ✓ *Memory*
- ✓ *Flash / SSD*
- ✓ *CPUs and Cores*
- ✓ *Dedicated Fast Disks*

Improve Design and Implementation:
- ✓ *Buy Pre-analyzed and Aggregated Data*
- ✓ *Dimension Reduction*
- ✓ *Faster Algorithms*
- ✓ *Data Filter*
- ✓ *Data Vault*
- ✓ *Indexing*
- ✓ *Query Optimization*
- ✓ *Change Data Capture*

Topic IV:
Analytical Application Examples

- Financial Services Applications
  - Back Office Analysis
  - Customer Risk Analysis
- Retail Analytics Applications
  - Market Basket Analysis
  - Customer Profitability Analysis
- Next Best Action / Recommendation Engine

Financial Services Back Office Analysis



Back Office Applications:

- Reserve Calculations
- Solvency Analysis
- Investment Hedging
- Portfolio Value
- Asset / Liability Modeling
- Value of New Business

Data Sources

Internal Data Sources are sources of data found inside the organization – i.e. in the organizations databases, files, emails, etc.

- Customer data
- Customer accounts
- Product data
- Organization and employee data
- Transactions and events
- Financial assets and liabilities
- Inventory
- Employees
- Emails
- Audio Recordings

External Data Sources are sources of data found outside the organization – i.e. in the records of other entities such as: government, data brokers and social media companies.

- Census data
- Customer surveys
- Customer lifestyle profiles
- Financial instrument prices
- Economic indicators
- Internet of Things (IoT)
- Social Media (Facebook, LinkedIn, Twitter)

Specifications and Plans Database contains data that drives the process. It is better than uncontrolled spreadsheets or hardcoding values.

- Plans and budgets
- Calculation rules
- Interest and other rates
- Organizational hierarchies
- Products
- Sensitivities
- Control rules

Model Engine

The Model Engine is the part of the analytical application that performs calculations. This is where number crunching happens and input is transformed into useful output. For analysis based on Monte Carlo simulation methods, multiple scenarios are processed based on a combination of scenarios and system data. Use of repeated randomized cycles is called a stochastic approach. It differs from the deterministic approach which returns a single known answer.

The model engine calculations can be based on a number of models and approaches, including:

- Statistics
- Data mining
- Monte Carlo simulation
- Custom models

Production expands use of the model:

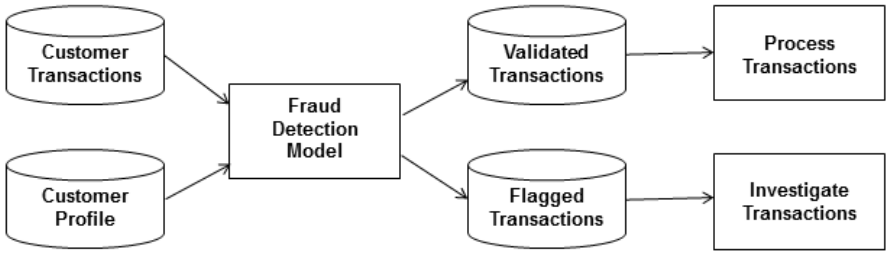
- Run for millions of accounts or policies
- Run for thousands of scenarios

Customer Risk Analysis

Customer Risk Analysis requires tracking customer data and using analytics to identify and anticipate issues.

Loss of Profit:

- Switch to a Competitor
- Cancel an Order
- Drop Out from School
- Return a Product
- Quit a Job
- Have an Automobile Accident
- Make Expensive Requests



Criminals:

- Commit Fraud
- Commit a Violent Crime
- Commit Bribery
- Embezzle Funds
- File Improper Insurance Claims
- Steal Private Information
- Improperly Invoice for Goods or Services
- Launder Money
- Make a Terrorist Attack
- Skip Bail
- Stranger Owned Life Insurance (STOLI)
- Stranger Owned Annuities (STOA)

Risk – Red Flag Models

Red Flag Models identify conditions that may indicate fraud or other undesirable activity. Many are known to Fraud Investigators.

- Address is post office box
- Credit card charge outside of usual locations
- Digit distribution breaks Benford's Law
- Disbursements just under approval limit
- Dormant account is suddenly active
- Duplicate vendor invoice numbers
- Gaps in check numbers
- Line items do not match control totals
- Loans without repayments
- Numbers are outliers beyond standard deviation
- Parties with same tax id but different name
- Parties without postal address or telephone
- Party grouping often exchanges excess funds
- Party name is on watch list
- Payments to agents rather than policy holder
- Postal address changed frequently
- Postal address is on hot list
- Postal Address shared by unrelated parties
- Postal Addresses matches Employee Postal Address
- Tax id number is on death list
- Transactions are duplicated
- Unrelated parties share direct deposit account
- Unrelated parties share postal address
- Payee information matches employee information
- Vendor sequential vendor invoice

Market Basket Analysis (MBA)

Market Basket Analysis is an analytical method that identifies product and service combinations that customers tend to buy. It is typically based on records of customer purchases. Market Basket Analysis is also known as Product Affinity Analysis or Association Rule Learning.

Beer and Nappies

Walmart

SKU Rationalization Demands Market Basket Analysis (aka Customer Buying Patterns)

emcien

http://emcien.com/sku-rationalization-demands-market-basket-analysis-aka-customer-buying-patterns/

MBA Based Actions

Do this:

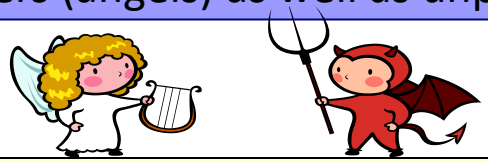
- Locate items in stores and websites
- Improve cross-sell and up-sell
- Offer attractive incentives
- Target offers to individuals or segments
- Attract traffic to stores and websites
- Obtain inventory to support promotions – a sale on one item can lead to increased sales on other items

Avoid this:

- Harming sales by dropping products
- Harming sales by increasing prices on related goods

Customer Profitability Analysis (CPA)

Customer Profitability Analysis is an analytic approach that determines the profitability of individual customers or segments of customers by identifying revenue and cost patterns associated with those customers. This includes identifying the most profitable customers (angels) as well as unprofitable customers (devils).



- Customer behavior and profitability identified by analysis of company databases
- 20% of customers are angels and result in bulk of profit
- 20% of customers are devils and reduce profits by 20%
- Profiles built of profitable and unprofitable customers
- Attracts most profitable customers by promotions, stocking desired products and providing best service
- Avoids unprofitable customers by dropping them from promotion lists, stopping loss-leader promotions and charging fees for restocking
- **Warning:** taken to extremes this could backfire and turnoff best customers!

Customer Profitability Factors

The Profitable Customer:

- Orders standard products
- Orders standard handling
- Orders via web or ecommerce
- Makes short service calls
- Almost never returns goods
- Orders large volume
- Pays on time
- Praises company on social media

The Unprofitable Customer:

- Orders exception products
- Orders via call centers
- Often makes lengthy service calls
- Frequently returns goods
- Orders special handling
- Orders small volume
- Requires low cost – price match
- Pays late – requires collection
- Complains on social media
- Only buys on sale

Next Best Action (NBA)

Next Best Action is an immediate action recommended by rules discovered through data mining or statistics that is intended to produce optimal results. This often includes providing service or making an recommendations to a customer.

amazon.com

- Amazon's revenue in 2009: $24.5B
- ~$5B came from product recommendations

NETFLIX

- Netflix offers prize for improved recommendation engine algorithm

- Telco analyzes customer profitability and behavior
- Builds a decision model based on customer profitability and responsiveness
- Customer calls Customer Service, requests lower rate or termination of account
- Telco service rep knows what to offer to customer – discounts or other accommodations

Recommendation Engine

- Customer Profitability Model
  - Customer Revenue Model
  - Customer Expense/Risk Model
- Customer Preferences Model
- Market Basket Model
- Product Profitability Model

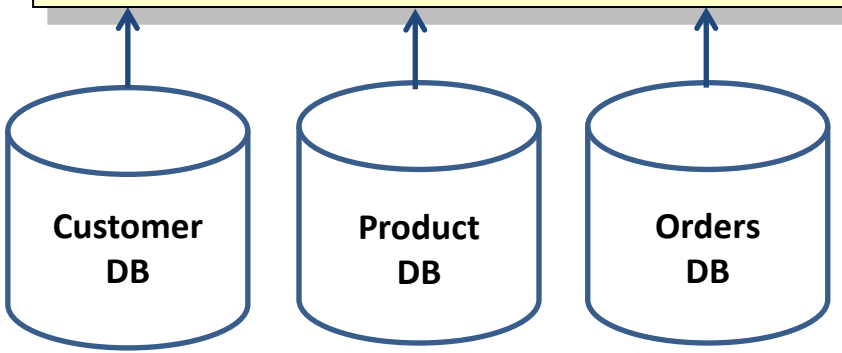
Recommendation Engine:

On Demand

Find intersection of products:

- Related to viewed product
- Preferred by Customer

Order by Product Profitability



Session Structure

Topic 1: What is an Analytical Application?

- What is Analytics?
- Predictive Analytics Impacts the Bottom Line
- Analytical Applications Examples

Topic 2: Analytical Methodology

- Analytics Methodology
- Team Roles
- Data Preparation
- Analytical Modeling
- Model Evaluation
- Deployment - Productionizing

Topic 3: Analytics Architecture

- Analytics Architecture Components
- Increasing Analytics Performance

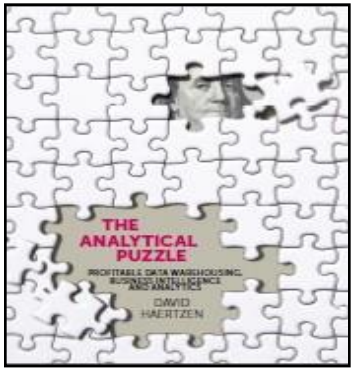
Topic 4: Analytical Application Examples

- Financial Services Applications
- Retail Analytical Applications
- Recommendation Engine

David Haertzen – Contact Information

David Haertzen

Author and Instructor



http://www.davidhaertzen.com/

http://www.linkedin.com/davidhaertzen

http://ecm.elearningcurve.com/David_Haertzen_s/89.htm

Twitter: #BigHeart7

david at davidhaertzen dotCom