



Business Analytics

IBM
Software
Solutions
Group

Real World Application and Usage of IBM Advanced Analytics Technology

Anthony J. Young – Pre-Sales Architect for IBM Advanced Analytics

February 21, 2014



Welcome



Anthony J. Young

- Lives in Austin, TX
- Focused on Finance, Performance Management, & Analytics for 14+ Years
- Statistics & Mathematics Degrees from 
- 3 Years with  prior at **Deloitte**.
- First job was a Mathematical Statistician with the  on Capitol Hill
- Relocated to Austin in 2011 after 12 years in Washington, DC

Agenda

- Introduction
- IBM Approach to Predictive Analytics
- IBM SPSS Modeler Overview
- End to End demonstration of IBM Analytics Technology
- A Practitioner's Thoughts about a Data Miner
- Closing Discussion

What I want you to leave with today

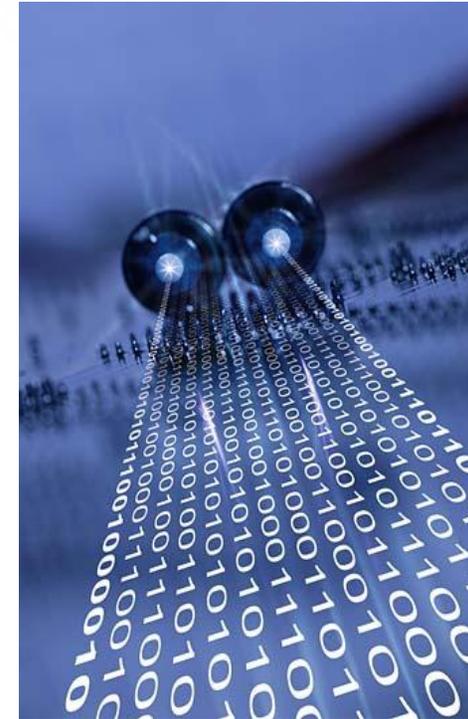
- Insight into the IBM Analytics Platform
- Understanding of our approach to Predictive Analytics
- Ideas on how to incorporate Data Mining into your curriculum
- Just how easy it is to get started with IBM Predictive Analytics
- More questions than you came in with; *e.g. how can we get involved asap?*

An Opening Question...

What if we could use the data that we are collecting to see farther, deeper, and with greater detail to support what we want to accomplish?

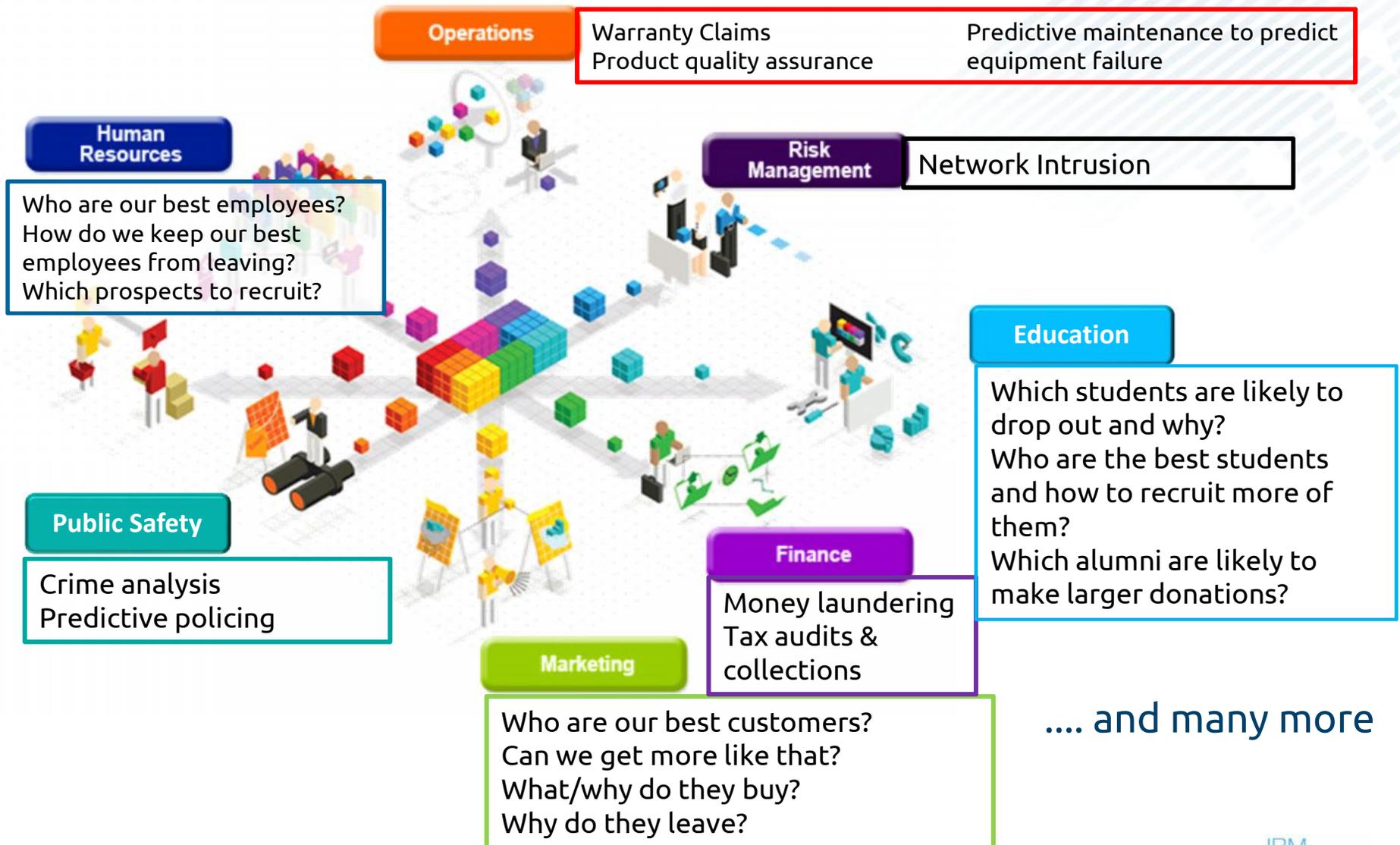
Another Opening Question...

What is predictive analytics?



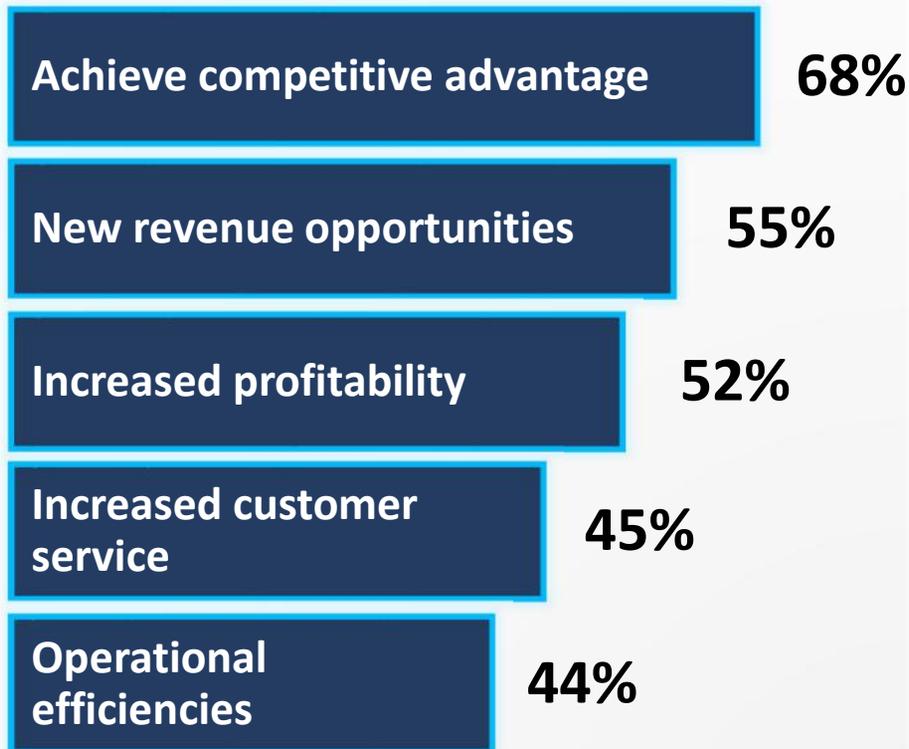
Predictive Analytics helps connect **data** to effective **action** by drawing reliable conclusions about current conditions and **future events**.
- Gareth Herschel, Research Director, Gartner Group

Focus Areas for Predictive Analytics



Top Five Benefits of Predictive Analytics

How has your organization benefited from predictive analytics:



Source: Ventana Research
Predictive Analytics Benchmark Research

Related Research Points:

- Management (76%) has no doubts that predictive analytics has given its business the competitive edge.
- Small (81%) businesses are especially convinced that they have achieved competitive advantage by using predictive analytics.

In the Public Sector¹:

- To best position governments and citizens for desired results, four focus areas now require integrated execution and measurement: information, engagement model, digital platform and **analytics competency**.
- Leaders acknowledge both a perpetual information explosion and a persistent “data paradox” – the management dilemma associated with having **too much data and too little insight**.

¹IBM Institute for Business Value, “Opening up government: How to unleash the power of information for new economic growth”

Our Approach to the Predictive Analytics Process

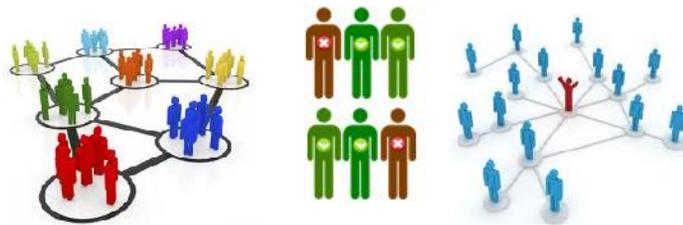
Capture Detect & Capture

Transactions
Demographics
Interactions
Opinions



Predict Analyze & Predict

Predictive Modeling
Data Mining
Text Analytics
Social Network Analysis
Statistical Analysis



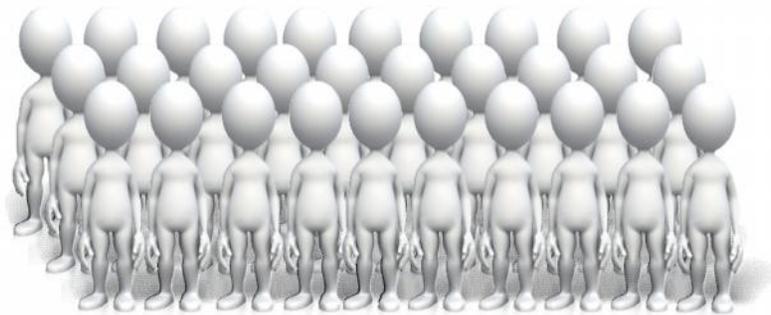
Act Engage & Act

Prediction
Rules
Optimization
Process



Automation and Deployment

The Mathematics of Predictive Analytics



Without Predictive Analytics

- 30 Customers
- We expect some churn, but who knows which ones will leave
- Spend avg of \$250 per customer on retention offers at a cost of \$7,500 (what we normally spend on retention programs)
- 5 actually churn, and we spend \$1,500 to acquire new customers
- We are now out **\$9,000** to reactively manage the customer base



Leveraging Predictive Analytics

- Same 30 Customers, but clustered into High, Medium, & Low Risk
- Predict which High & Medium Risk customers are likely to churn, excluding Low Risk customers
- Spend \$250 per customer on very focused retention offers for the 6 High & Medium Risk customers predicted to churn only at a cost of \$1,500
- 3 actually churn from this group, in addition to 2 Low Risk customers
- We spend \$1,250 to acquire 5 new customers
- We are now out **\$2,750** to proactively manage the customer base
- We saved almost 70% to maintain the same customer base

How do we do it?

IBM SPSS Modeler at a Glance

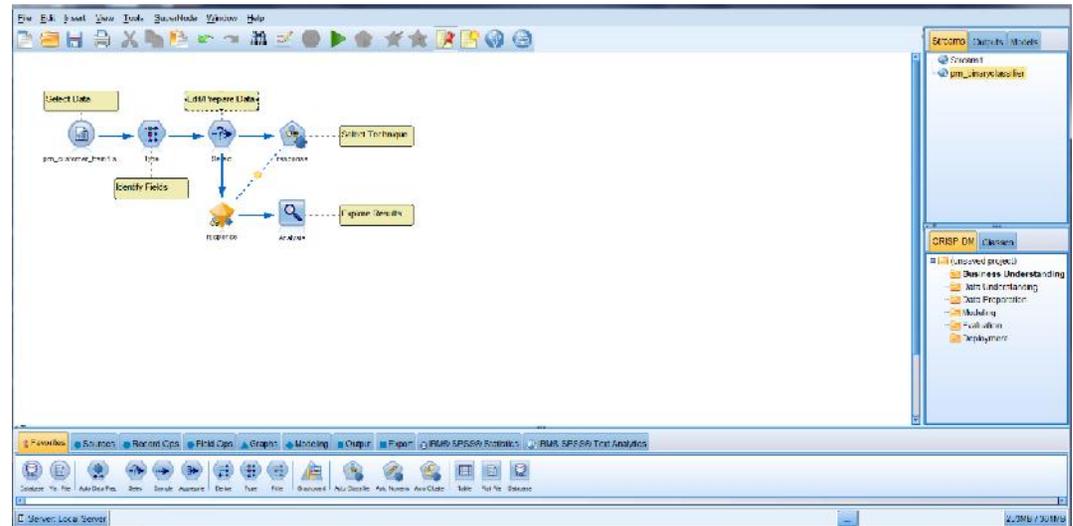
- Comprehensive predictive analytics platform
- All types of users at multiple analytical abilities
- Integrated deployment, optimization and decision management capabilities
- A visual interface with built-in guidance
- Structured and unstructured data
- Deployed on a desktop or integrated within operational systems
- Bring predictive intelligence to a single user or an entire enterprise

Model Building for Users of Multiple Analytical Abilities

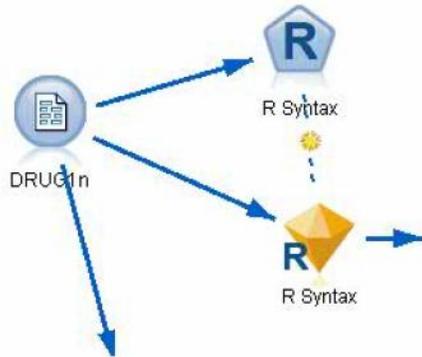


One-click modeling for the business user

Visual analytical workbench to expand and extend analysis



Extend Capabilities through R



R Integration

R Build/Score, Process and Output node support
 Scale R execution by leveraging database vendor provided R engines

The image shows two screenshots from the Custom Dialog Builder. The left screenshot shows the 'Tools' pane with various control types like Source List, Target List, Check Box, etc. The main area shows a 'Variables' list with Variable1, Variable2, and Variable3, and a 'Predictors List' with a single 'Variable'. The 'Automatic Learning' checkbox is unchecked. The right screenshot shows the 'Syntax Template' dialog box with the following content:

Enter your syntax template below. The syntax template is used to generate syntax for your dialog. Enter replaceable values using the corresponding control identifier enclosed within '%' symbols.

For example, one line of a syntax template may contain:
`/VARIABLES %var_list%`

Tip: You can select from a list of all available control identifiers to insert by positioning the text caret where you would like to insert an identifier and then pressing the <CTRL>-<SPACEBAR> keys together. The list contains the control identifiers followed by the elements of command syntax available for the current context.

```

1 # model building
2 attach(modelerData)
3 model<-lm(%stgt_var% ~ %%pred_var%%)
4 model
5
6
7
8 # scoring
9 attach(modelerData)
10 output<-predict(model)
11 output
12
    
```

At the bottom, the 'Target Variable List Properties' table is visible:

Property	Value
Identifier	pred_var
Title	Predictors List
ToolTip	
Target list type	Multiple item list
Mnemonic Key	
Required for execution	True

Custom Dialog Builder for R

Provides the ability to create new Modeler Algorithm nodes and dialogs that run R processes
 Makes R usable for non-programmers

Uncovering Patterns in Unstructured Data

- Text Analytics
 - Natural Language
 - Sentiment Analysis
- Entity Analytics
 - Disambiguate identity
 - People, places, things
- Social Network Analysis
 - Uncover relationships
 - Find leaders and followers

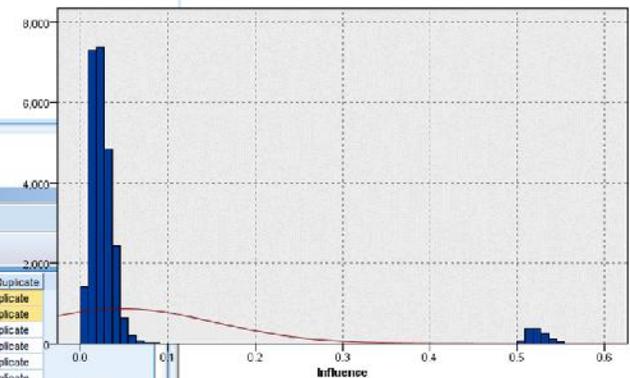
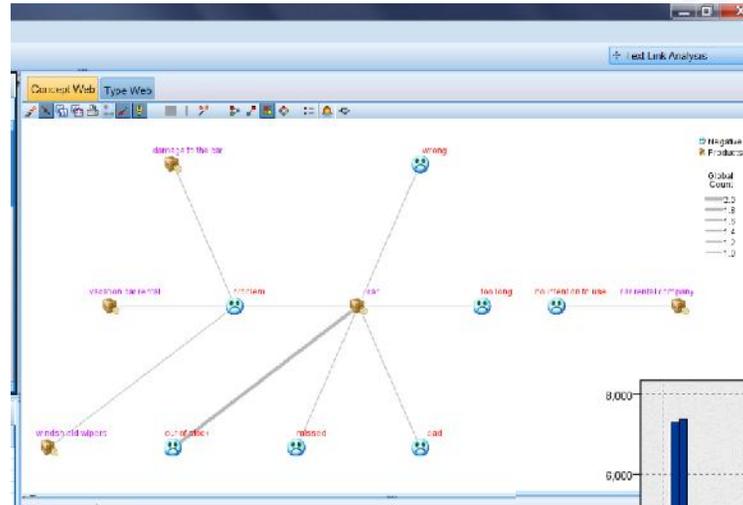
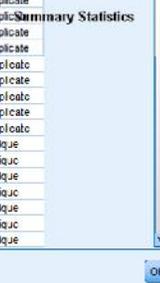


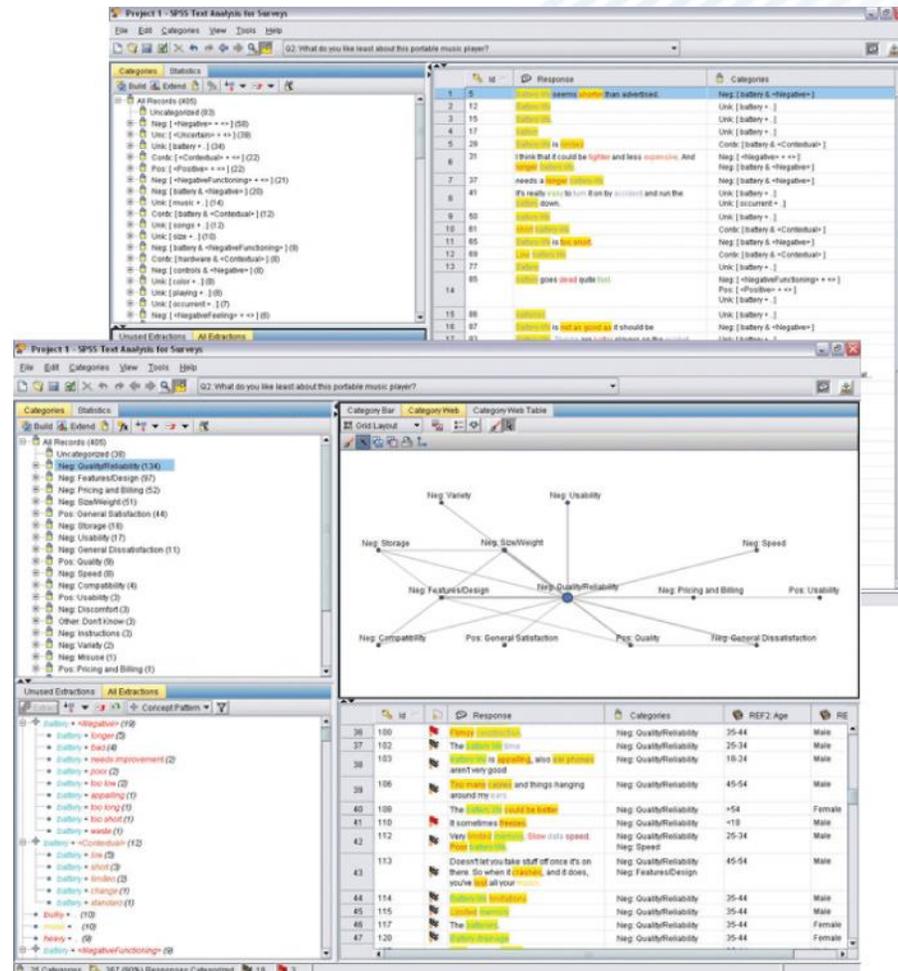
Table (13 fields, 650 records) #1

SEA-ID	SEA-SRC	key	SEA-SC	SEA-ROLE	SEA-DOB.DATE	SEA-GENDER	CENDER	SEA-NAVE	CLR_NAVE	SEA-NAME	CVEN_NAME	SEA-PASSPORT_ID_NUM	SEA-PHONE	PHONE_NUM	SEA-SSN_ID_NUM	IsDuplicate
11	273	TEST	REF0000273	0.000	SnUS	1928-04-03	M	VARTINEZ	KEVIN	317451		SnUS	320-19-4		Duplicate	
12	273	TEST	REF0000273	10.000	SF	1028-04-03	M	VARTINEZ	EUGENE	317451		SnUS	320-10-4		Duplicate	
13	324	TEST	REF0000324	0.000	SnUS	1995-02-3	M	BUTLER	JOSHUA	692453		SnUS	510-86-6		Duplicate	
14	324	TEST	REF0000324	10.000	SF	1095-02-3	M	BUTLER	TODD	692453		SnUS	510-86-6		Duplicate	
15	342	TEST	REF0000342	8.000	SF	1967-12-02	F	HUGHES	DEBRA	989511		SnUS	156-24-7		Duplicate	
16	342	TEST	REF0000342	0.000	SnUS	1073-12-3	F	HUGHES	DEBRA	989511		SnUS	156-24-7		Duplicate	
17	377	TEST	REF0000377	0.000	SnUS	1950-09-28	M	WHITE	DANIEL	844897		SnUS	389-32-9		Duplicate	
18	377	TEST	REF0000377	8.000	SF	1040-12-01	M	WHITE	DAN	844897		SnUS	389-32-9		Duplicate	
19	388	TEST	REF0000388	0.000	SnUS	1937-03-26	M	HILL	RUSSELL	124791		SnUS	551-95-0		Duplicate	
20	388	TEST	REF0000388	10.000	SF	1037-03-26	M	HILL	RUSSELL	124791		SnUS	551-95-0		Duplicate	
21	437	TEST	REF0000437	9.000	SF	1937-09-6	F	JENKINS	DORIS	482623		SnUS	688-19-5		Duplicate	
22	437	TEST	REF0000437	0.000	SnUS	1037-09-6	F	JENKINS	DORIS	482623		SnUS	688-10-6		Duplicate	
23	501	TEST	REF0000501	0.000	SnUS	1978-04-03	F	GRIFFIN	RUTH	571206		SnUS	516-52-9		Duplicate	
24	501	TEST	REF0000501	10.000	SF	1078-04-03	F	GRIFFIN	RUTH	571206		SnUS	516-52-9		Duplicate	
25	1	TEST	REF0000001	0.000	SnUS	1938-1-6	M	BROCKS	JOE	147882		SnUS	338-14-3		Unique	
26	2	TEST	REF0000002	0.000	SnUS	1080-04-05	F	HALL	AWNE	554947		SnUS	413-31-6		Unique	
27	3	TEST	REF0000003	0.000	SnUS	1931-06-25	M	BROWN	AARON	858779		SnUS	997-89-0		Unique	
28	4	TEST	REF0000004	0.000	SnUS	1090-0-02	M	WASHINGTON	WAYNE	642815		SnUS	465-80-0		Unique	
29	5	TEST	REF0000005	0.000	SnUS	1983-0-21	F	ADAMS	KIMBERLY	762208		SnUS	440-90-5		Unique	
30	6	TEST	REF0000006	0.000	SnUS	1049-05-25	M	EMIT	CLARENCE	530356		SnUS	823-80-1		Unique	
31	7	TEST	REF0000007	0.000	SnUS	1974-08-2	M	BROWN	TERRY	213824		SnUS	274-75-1		Unique	



Text Analytics

- Uses natural language processing heuristic rules and statistical techniques to reveal conceptual meaning in text
- Extracts concepts from text and categorizes them
- Makes unstructured qualitative data more quantifiable, enabling the discovery of key insights from sources such as survey responses, documents, emails, call center notes, web pages, blogs, forums and more



Brings repeatability to ongoing decision making

Optimize your operational business with Decision Management

Optimized Decisions

Business Rules + Optimization + Predictive Analytics

All Data

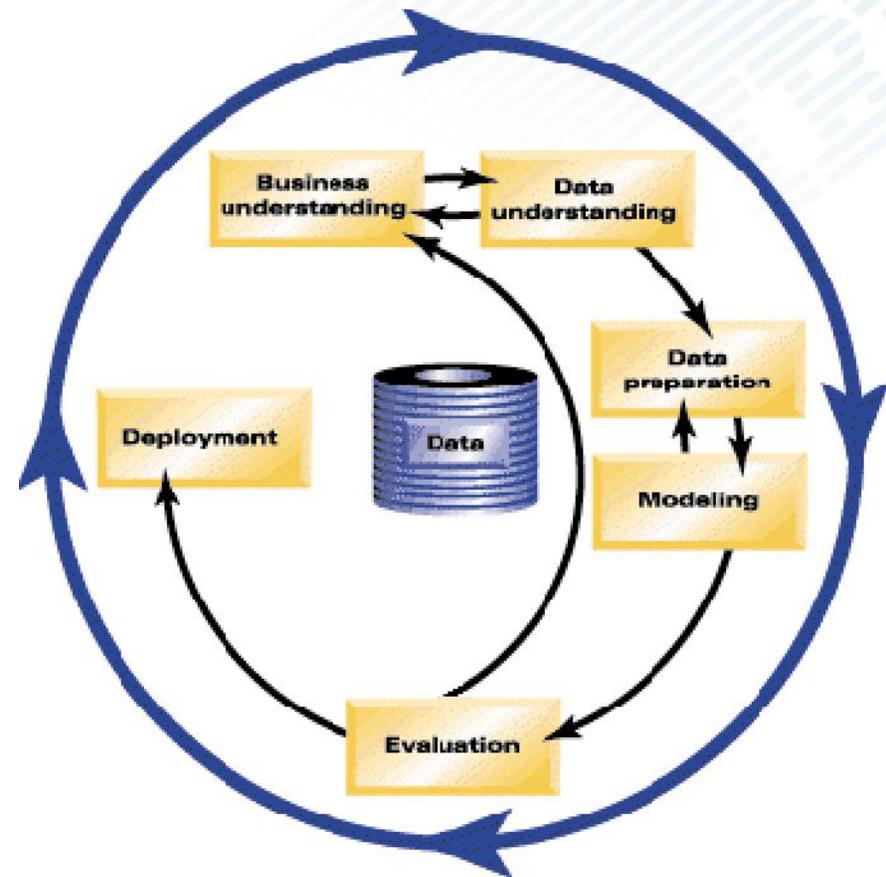
Optimize Actions Within Resource Constraints, Aligning Execution With Strategy

Empower Real-time and Adaptive Decisions Accommodating Changing Conditions

Provide Front-line Employees and Systems With Recommended Actions

Methodology

- **Cross-Industry Standard Process Model for Data Mining**
- **Describes Components of Complete Data Mining Project Cycle**
- **Shows Iterative Nature of Data Mining**
- **Vendor and Industry Neutral**



Data mining is a key discipline for applying predictive analytics

IBM SPSS Modeler Within the Business Analytics Ecosystem

Predictive Customer Analytics

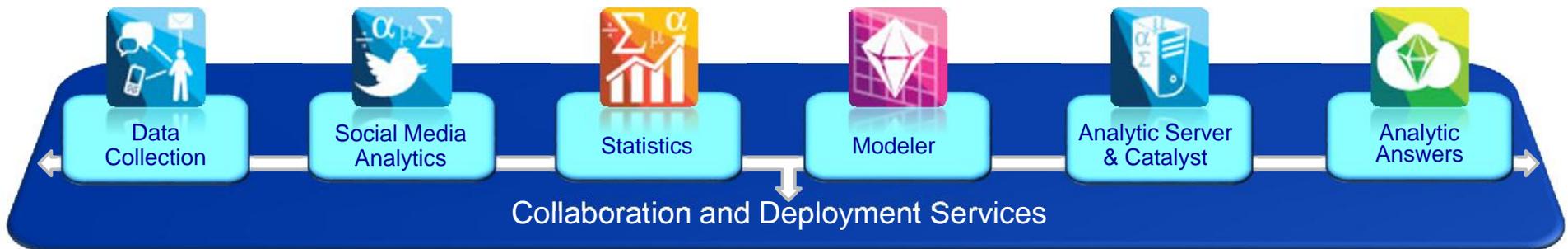
Acquire
Grow
Retain

Predictive Operational Analytics

Manage
Maintain
Maximize

Predictive Threat & Fraud Analytics

Monitor
Detect
Control



Cognos software

InfoSphere

ASSETS MANAGEMENT IBM maximo

IBM

N*NETEZZA an IBM Company

WebSphere software

IBM Research

Etc...

An End-To-End Demonstration

Thoughts about a Data Miner

What qualities do I look for in someone to be a good data miner?



A Practitioner's Thoughts about being a Data Miner

Creative

- Looks at a data set and thinks about what else can we incorporate?
- Constantly asks, "what if?"
- ...and tenacious

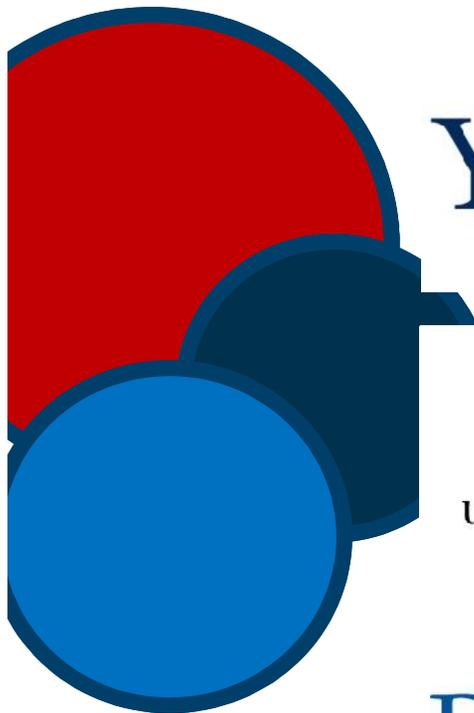
Proactive

- Can I build another model from the results of this model?
- Where can we insert this into workflow and make it repeatable?

Big Picture

- Why did we do what we just did?
- Who owns this problem?
- How much is a solution worth?
- What does an exec need vs. an analyst?
- Can answer the "so what" and "who cares" questions

Universities that have taken the first steps



Yale



MICHIGAN STATE UNIVERSITY



uOttawa



RUTGERS

NC STATE UNIVERSITY

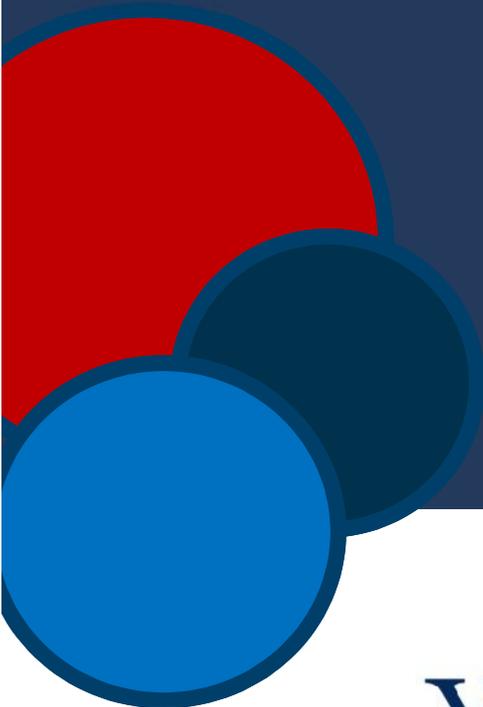


Duke UNIVERSITY



Experiential Learning: SMART Projects

- Students
- University
- Client Partner
- IBM



Yale



MICHIGAN STATE
UNIVERSITY



Takeaways

Easy to use, visual interface

- Short timeframe to be productive with actionable results
- Does not require knowledge of programming language
- No proprietary data formats
- Open architecture

Business results focused

- Leverages the investments already made in technology
- Cost effective solution that delivers powerful results across organization
- Full range of algorithms for your business problems
- Big Data enabled (Hadoop, SQL Pushback)

End-to-end solution

- Data preparation through real time interactions
- Use structured, unstructured and semi-structured data
- Integrated portfolio for business analytics
- Scales from a single desktop to an enterprise deployments



Thank YOU