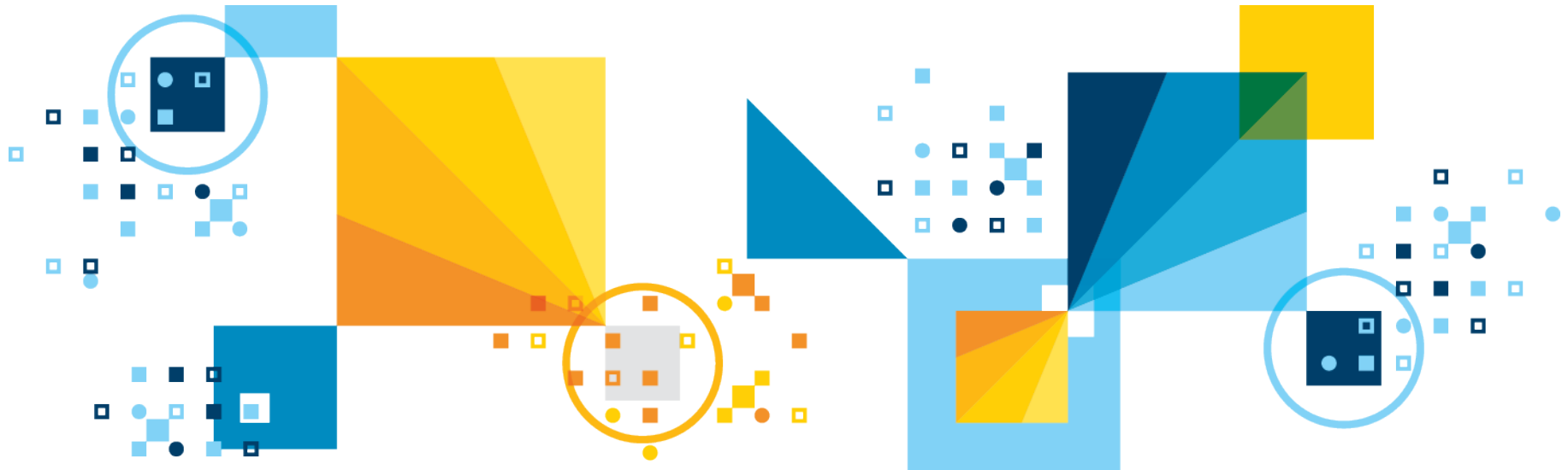# IBM SPSS Predictive Analytics Workshop
with IBM SPSS Modeler

Add Date and Location and Presenter Name

Agenda

- Welcome and Introductions
- Overview of Predictive Analytics
- IBM SPSS Modeler Overview
- Exercise 1: Predictive in 20 Minutes
- IBM SPSS Modeler Basics
- Predictive Analytics Methodology and Applications
- **Break**
- Exercise 2: Find Patterns and Groups
- Exercise 3: Understand the Past, Predict the Future
- Exercise 4: Deploy Insights
- Integrating Predictive Analytics with Planning and BI
- Wrap up and Next Steps

2

Introductions

1. Name and Organization you represent
2. Have you done any predictive analytics projects before today?
3. Do you have any experience with any IBM SPSS Products?
   - SPSS Statistics
   - SPSS Modeler (formerly Clementine)
   - SPSS Text Analytics for Surveys (TAFS)
4. What are your objectives for today?

Today's presenter: Lorem Ipsem

XXXX@us.ibm.com

XXX-XXX-XXX

linkedin.com/in/xxxxxx

Project Experience

- *XXX*
- *YYY*
- *ZZZ*

Purpose of the workshop

- Introduction to predictive analytics

- Stimulate thinking about how predictive analytics would benefit your organization

- Demonstrate ease of use of powerful technology

- Get experience in "doing" predictive analytics

- Explore multiple predictive analytics techniques

- Understand the role of predictive analytics in decision optimization, planning and business intelligence

Advanced Analytics is For Everyone

Advanced analytics is the analysis of all kinds of data using sophisticated quantitative methods (i.e., statistics, descriptive and predictive data mining, simulation and optimization) to produce insights that traditional approaches to business intelligence (BI) — such as query and reporting — are unlikely to discover.

Realizing Power and Insight in your Data

Descriptive Analytics

Prescriptive Analytics

Predictive Analytics
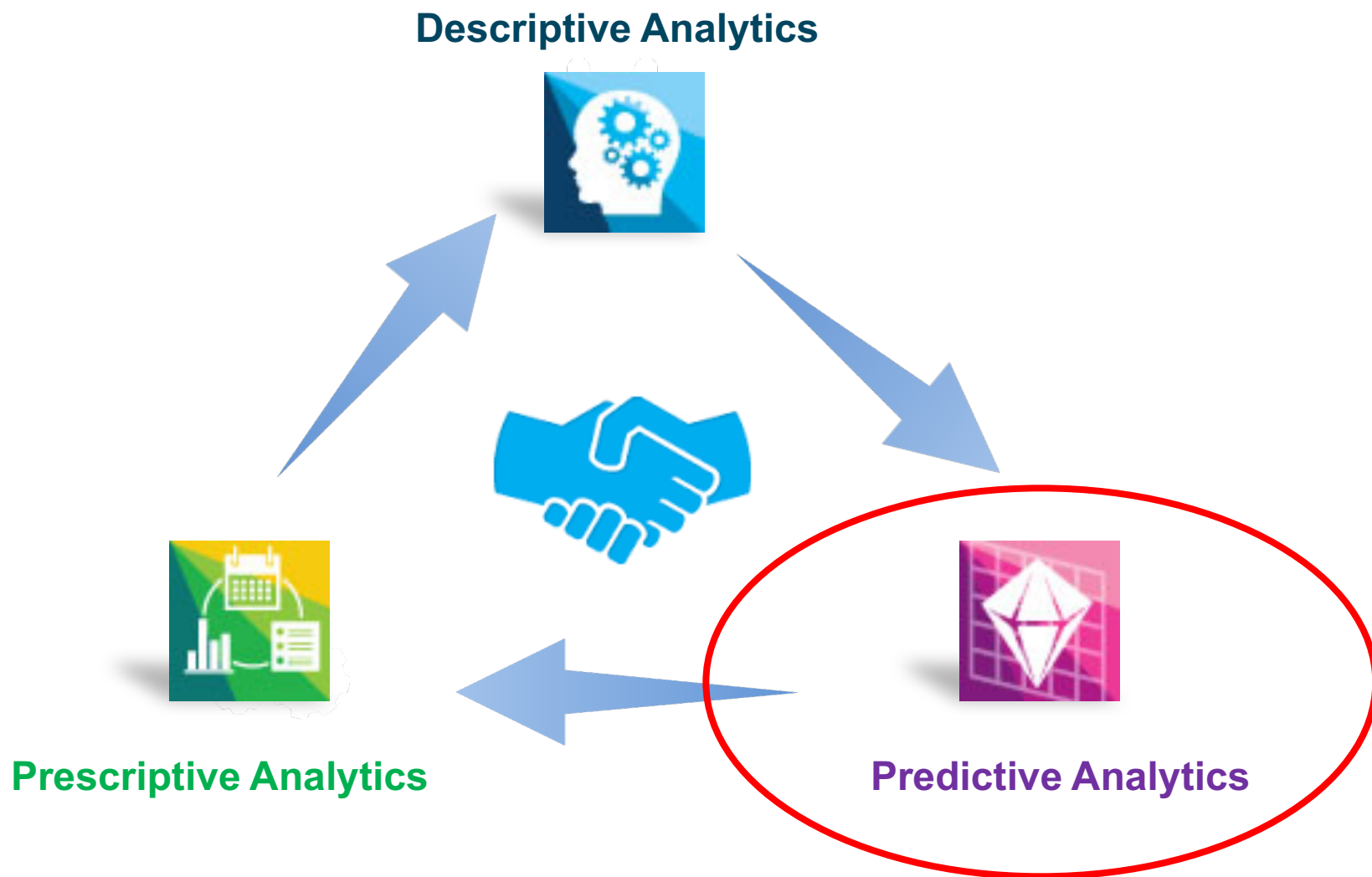
Analytics Driven Organizations Reap Rewards

Front Runners outperform on

Business outcomes (69%)

Competitive advantage (53%)

Revenues (60%)

by using data and analytics

Source: IBM Institute for Business Value (IBV)

Areas for Predictive Analytics

Operations

Warranty Claims
Product quality assurance

Predictive maintenance to
predict equipment failure

Human Resources

Who are our best
employees?
How do we keep our best
employees from leaving?
Which prospects to recruit?

Risk Management

Network Intrusion

Education

Which students are likely to
drop out and why?
Who are the best students and
how to recruit more of them?
Which alumni are likely to
make larger donations?

Public Safety

Crime analysis
Predictive policing

Finance

Money laundering
Tax audits &
collections

Marketing

Who are our best customers?
Can we get more like that?
What/why do they buy?
Why do they leave?

.... and many more

9

Predictive Analytics Use Cases

Consumer Insights
- Increasing revenue by understanding purchase drivers
- Reducing costs with strategically targeted marketing
- Forecasting new product sales

Human capital management
- Acquiring, growing and retaining employees
- Helping ensure optimal staff levels
- Increasing performance, efficiency and engagement

Higher Education
- Identifying students most likely to drop out
- Reducing student loan defaults
- Improving admit rate with targeted admissions

Supply chain management
- Increasing visibility into virtually all areas of the supply chain
- Decreasing downtime and unpredictability
- Improving customer satisfaction

Process optimization
- Improving accurate responses at the point of impact
- Decreasing costs through operational efficiency
- Transforming threat and fraud identification processes

What are some use cases from your organization?

Predictive Analytics in Action

FleetRisk Advisors

Helps trucking operators prevent accidents by building stronger & faster risk prediction models

- **20% reduction** in minor accidents
- **30% increase** in driver retention rates

U.S. Insurer

Analyzes and links claims and medical data, to prevent fraud and fast-track legitimate claims

- **USD 22 Millon ROI** anticipated, and
- **100% payback** anticipated from fraud reduction
- **Over 85% accuracy** for predicting independent medical failures

ASTRON

Uses streaming analytics to deliver insights from the world's largest radio telescope

- **99% faster identification** of data
- **Analyzes >1 exabyte of data** daily
- **Integrates data from >3,000** dishes & antennas to form the largest & fastest radio telescope in the world

Oak **Lawn Marketing**

Understand customer buying patterns for targeting infomercials

- **159% boost** in the average monthly rate of customers who return to shop compared to the previous year
- **400% increase** in expected total revenue over a three year period

IBM SPSS Modeler: At a glance
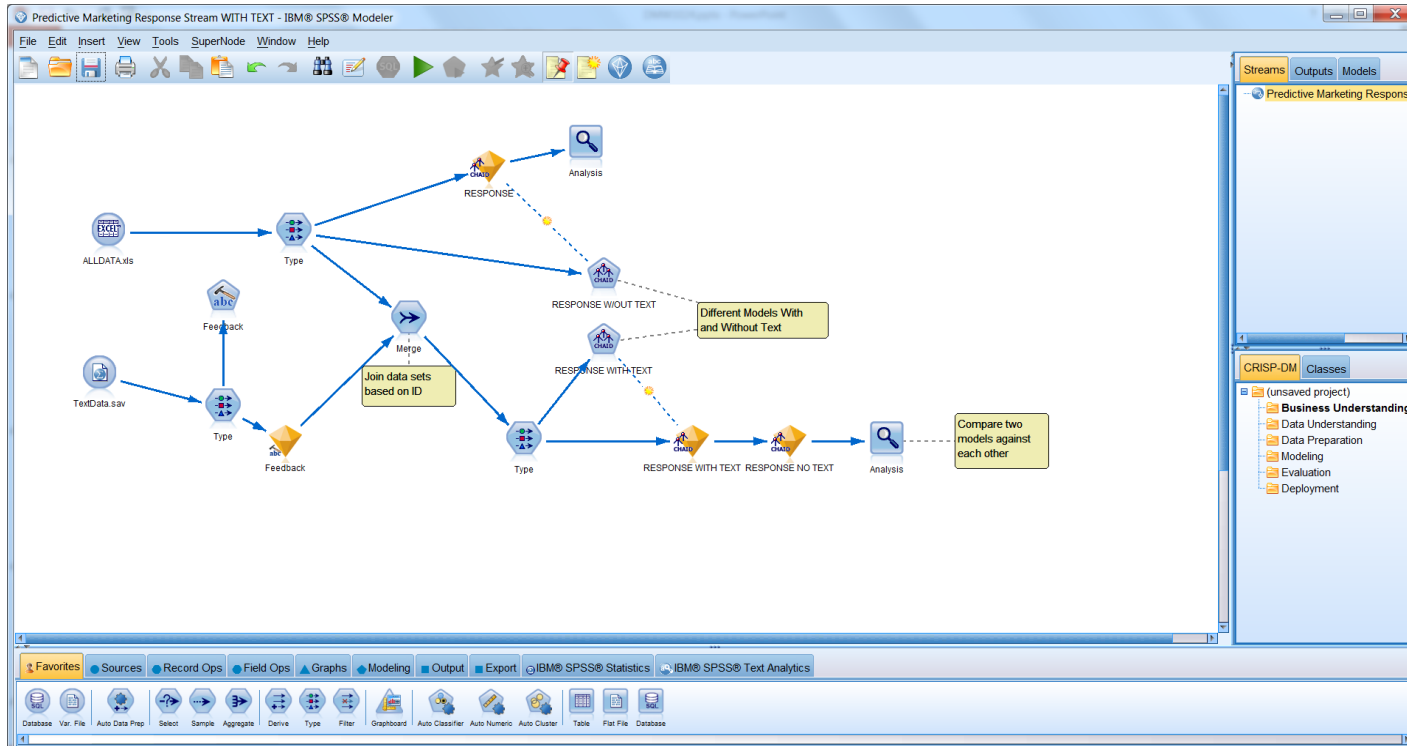
Flexible adoption and configuration

**On premises
Cloud
Hybrid
Bluemix**

Scale from **personal usage, point solution(s) to enterprise deployment**

Comprehensive predictive analytics platform

Identify patterns to generate predictive intelligence and improve outcomes

Access a range of advanced analytics; **decision management, text, entity, social network analysis and optimization**

12

Exercise: Predictive in 20 Minutes

Goal:

- **Identify who has responded to a marketing campaign**

Approach:

- Use a data extract from a CRM

- Prepare data for modeling

- Define which fields to use

- Choose the modeling technique

- Automatically generate a model to identify who has responded

- Review results

Why?

- To save marketing cost and increase marketing response, identify those likely to respond and focus marketing efforts on those prospects.

SPSS MODELER OVERVIEW CONTINUED

Methodology

CRISP-DM

- **CR**oss-**I**ndustry **S**tandard **P**rocess for **D**ata **M**ining

- Describes Components of Complete Data Mining Project Cycle

- Shows Iterative Nature of Data Mining

- Vendor and Industry Neutral
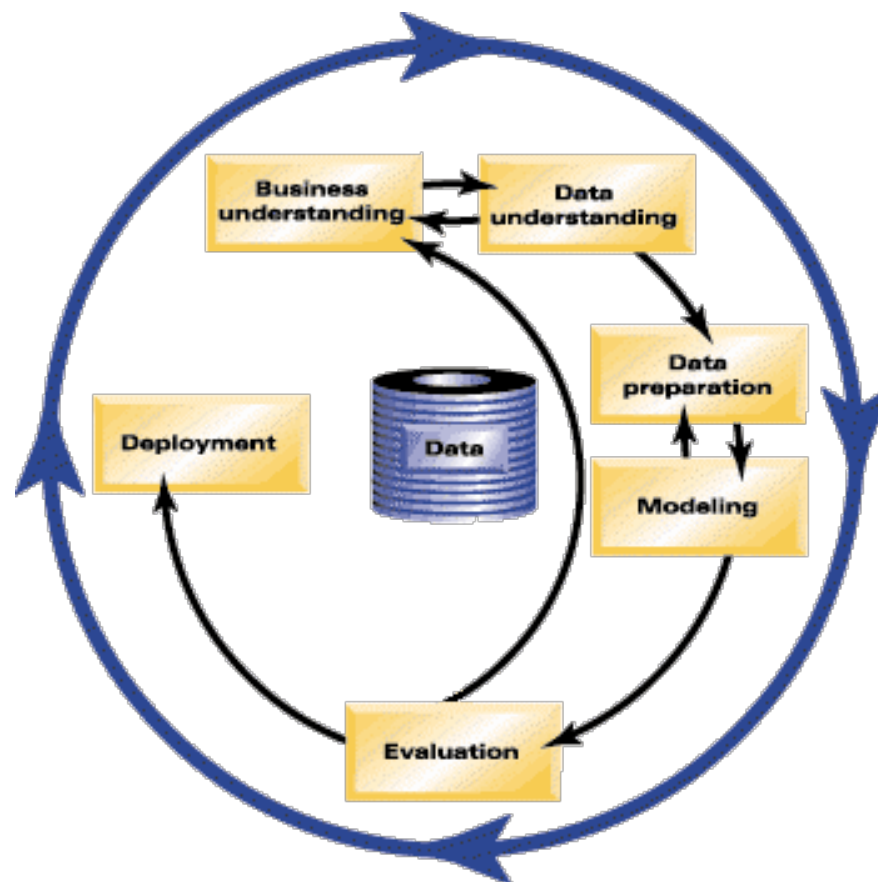
Data mining is a key discipline for applying predictive analytics

Modeling Techniques in IBM SPSS Modeler

Technique	Usage	Algorithms
Classification (or prediction)	• Used to predict group membership (e.g., will this employee leave?) or a number (e.g., how many widgets will I sell?)	• Auto Classifiers, Decision Trees, Logistic, SVM, Time Series, etc.
Segmentation	• Used to classify data points into groups that are internally homogeneous and externally heterogeneous. • Identify cases that are unusual	• Auto Clustering, K-means, etc. • Anomaly detection
Association	• Used to find events that occur together or in a sequence (e.g., market basket)	• APRIORI, Carma, Sequence

Extend Capabilities through R
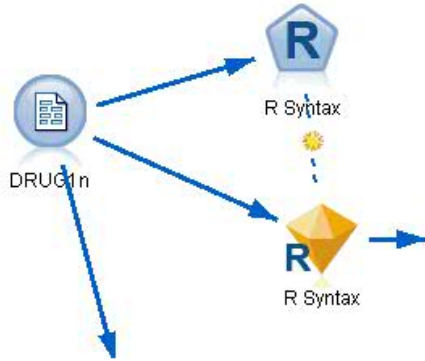
R Integration

R Build/Score, Process and Output node support

Scale R execution by leveraging database vendor provided R engines

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

SPSS Modeler Editions

Edition	Deployment	Techniques	Capabilities	Enhancements
Modeler **Personal**	**Desktop Only**	**Classification Segmentation Association**		
Modeler **Professional**	Desktop/Server **Hybrid**	**Classification Segmentation Association**		
Modeler **Premium**	Desktop/Server **Hybrid**	**Classification Segmentation Association**	Text Analytics Entity Analytics Social Network Analysis	
Modeler **Gold**	**Desktop/ Server/ Cloud/ Hybrid**	**Classification Segmentation Association**	Text Analytics Entity Analytics Social Network Analysis	**ADM C&DS**

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

The importance of text

Because people communicate with **words**, not numbers, it has become critical to be able to **mine text** for its **meaning** and to sort, analyse, and understand it in the same way that data has been tamed. In fact, the two basic types of information complement each other, with data supplying the "what" and text supplying the "**why**".

Source IDC: "Text Analytics: Software's Missing Piece?"

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:
-Documents
-Survey responses
-Call center Notes
-Social Media
-Web Pages

Entity Analytics Overview
## Identify Matching Entities

- **Analysts Struggle to Match Entities, Especially From Diverse Sources**
  - Natural variability (Bob vs. Robert)
  - Errors (transposed month and day)
  - Fabrications (fake identities)
- **Enhances Model Accuracy – Model Against a Complete, Accurate Entity**
- **Multiple Applications**
  - Business: is this the same order the customer submitted?
  - Fraud: is this the same person who already defaulted on a loan?
  - Government: is this the same vehicle that was carrying illegal content last time?
  - Policing: Is this the same person who called us before?

22

Social Network Analysis

- Used to indentify depth and reach of social media content.

- Focuses around identifying groups, leaders and probabilities that others will churn based on influence

- Enhances existing churn predictions of Modeler

- Expressed as two new nodes in the Sources Palette

  - Group Analysis – what are the groups in my data and who are the leaders

  - Diffusion Analysis – uses existing churn information to determine who else that churner is likely to influence to leave
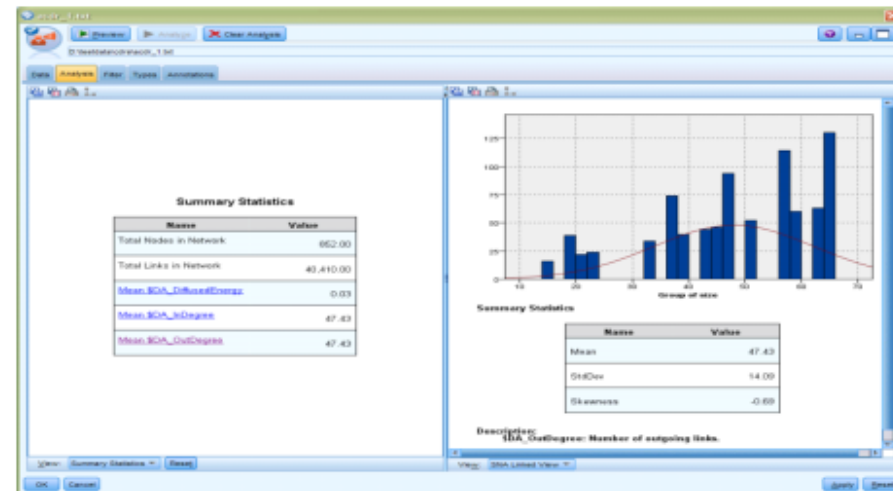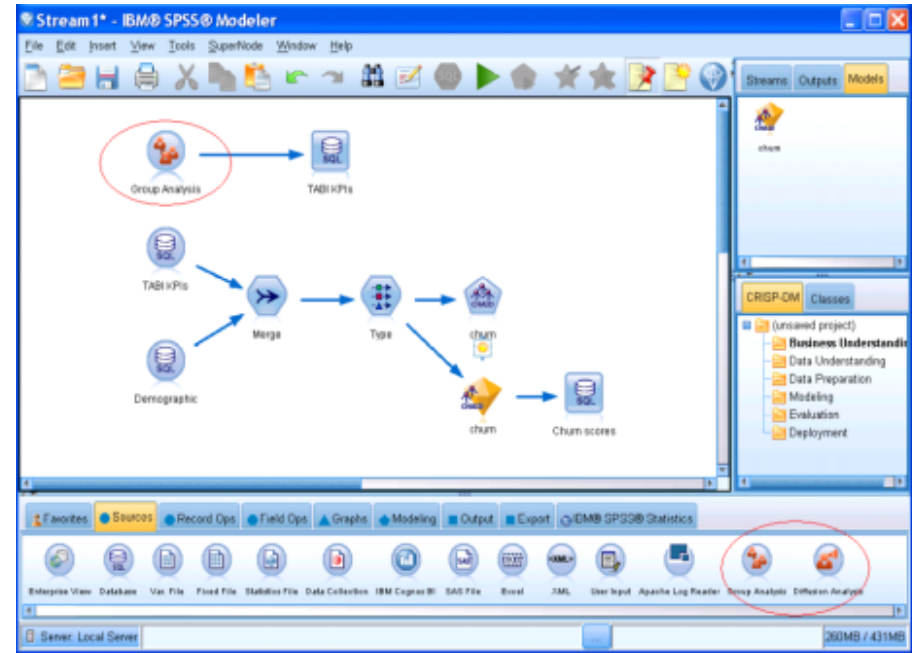
Break - please return in 10 minutes

HANDS ON EXERCISES

Find Patterns and Groups	Understand the Past, Predict the Future	Deploy Insights on new projects
Classify customers into groups based on underlying characteristics	**Model response to marketing offers using historical data**	**Leverage models to create insights on new customers**

Find Patterns and Groups

Understand the Past, Predict the Future

Deploy Insights on new projects

Classify customers into groups based on underlying characteristics

Model response to marketing offers using historical data

Leverage models to create insights on new customers

Exercise 2: Find Patterns and Groups

Goal:

- Create segments of customers

Approach:

- Merge disparate data sources
- Define which fields to use
- Automatically generate a model to group customers
- Apply business terms to new groups
- Export newly created groups to database

Why?

- Better customer understanding (demographics, socio-economic etc)
- Tailored messages for each group/segment
- Personal and more relevant for consumers

Find Patterns and Groups
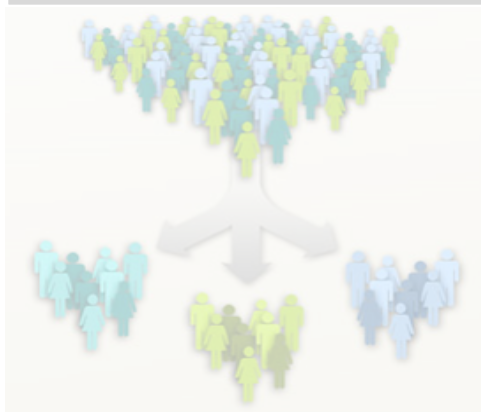
Understand the Past, Predict the Future

Deploy Insights on new projects

Classify customers into groups based on underlying characteristics

Model response to marketing offers using historical data

Leverage models to create insights on new customers

Exercise 3: Understand the Past, Predict the Future

Goal:

- Identify who is likely to respond to a marketing offer

Approach:

- Use a data extract from a CRM
- Extract concepts from the open ended comments in a customer survey
- Define which fields to use
- Choose the modeling technique
- Automatically generate a model to identify who is likely to respond
- Review results

Why?

- Target those likely to respond to offers to increase revenue, cut costs
- Using unstructured data improves modeling accuracy and provides more insight

Find Patterns and Groups
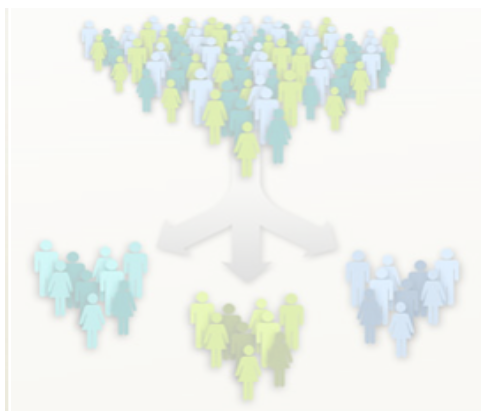
Understand the Past, Predict the Future

Deploy Insights on new projects

Classify customers into groups based on underlying characteristics

Model response to marketing offers using historical data

Leverage models to create insights on new customers

Exercise 4: Deploy Insights on New Projects

Goal:

- ✗ Leverage predictive insights to score new consumers

Approach:

- ✗ Use new customer records who have never received an offer
- ✗ Leverage the text extraction and classification models used in exercise 3
- ✗ Automatically generate scores of who is likely to respond
- ✗ Review results
- ✗ Deploy results for use by marketing team

Why?

- ✗ Create models on large populations for robust, representative results, and easily score smaller record batches when necessary, on-demand.
- ✗ Apply consistent and seamless methodology over many model scoring iterations, ensuring quality control.

INTEGRATING PREDICTIVE ANALYTICS WITH PLANNING AND BI

IBM SPSS Predictive Analytics

Predictive customer analytics	Predictive operational analytics	Predictive threat & Fraud analytics
■ Acquire ■ Grow ■ Retain	■ Manage ■ Maintain ■ Maximize	■ Monitor ■ Detect ■ Prevent

IBM® SPSS® Predictive Analytics

Statistics

Modeler

Analytic Server

Analytical Decision Management

Integration with IBM Cognos Analytics

1) Leveraging BI, identify problem or situation needing attention

Common Business Model

Cognos BI
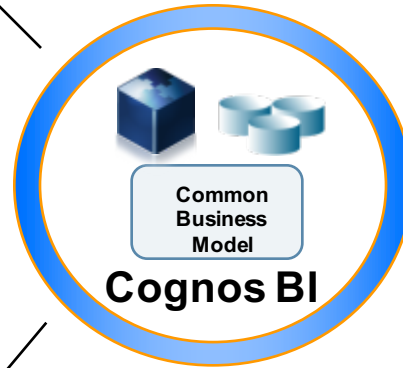
2) SPSS predictive analytics feed results back into the BI layer
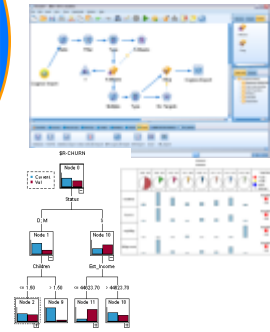
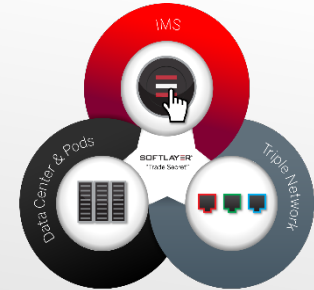**3) Results widely distributed via BI for consumption by business Users**

Flexible deployment

Flexible

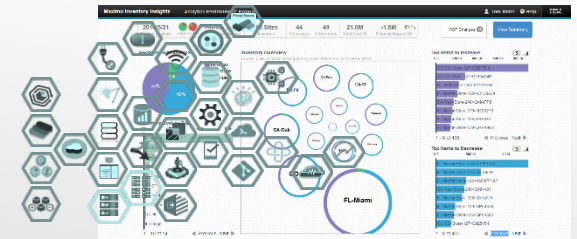Flexible Deployment Options, Including Cloud, Supported By An Infrastructure-agnostic Platform

Integrated

Integrated With IBM's High-performance Systems Built for Big Data Analytics

Embedded

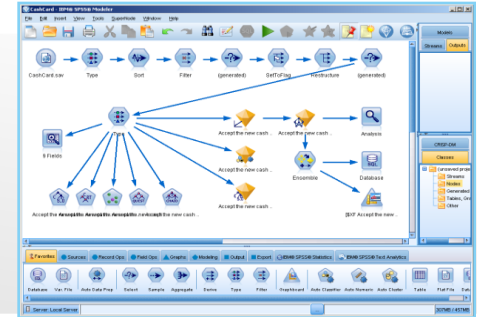Embedded Into Operational, Mobile and/or Cloud-based Applications

WRAP UP

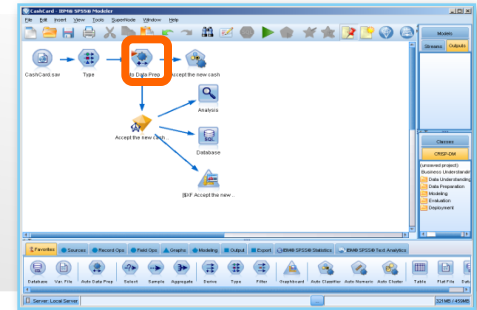# Predictive Analytics for All

Expert User

Deep Predictive Capabilities With In-database, R and Python Integration, and Integrated Deployment
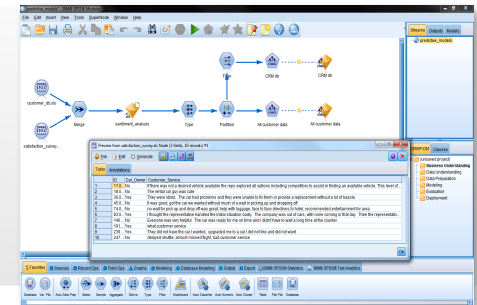

Intermediate User

Visual Workbench for Building Models of Any Complexity With Ability to Automate and Combine Tasks

Novice

Simple, Easy-to-Use, Visual, Guided Analytical Discovery, Intelligent Automation and Visual Storytelling

Workshop 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

Call To Action

Demo the software download a trial version of SPSS modeler and create your own use-cases with your own data

IBM SPSS Resource Center stay connected to this site to learn more about Modeler capabilities, latest use-case stories, and upcoming training opportunities

Connect with IBMers stay connected with IBM SPSS representatives. We would love to hear your success and help with any roadblocks to enablement
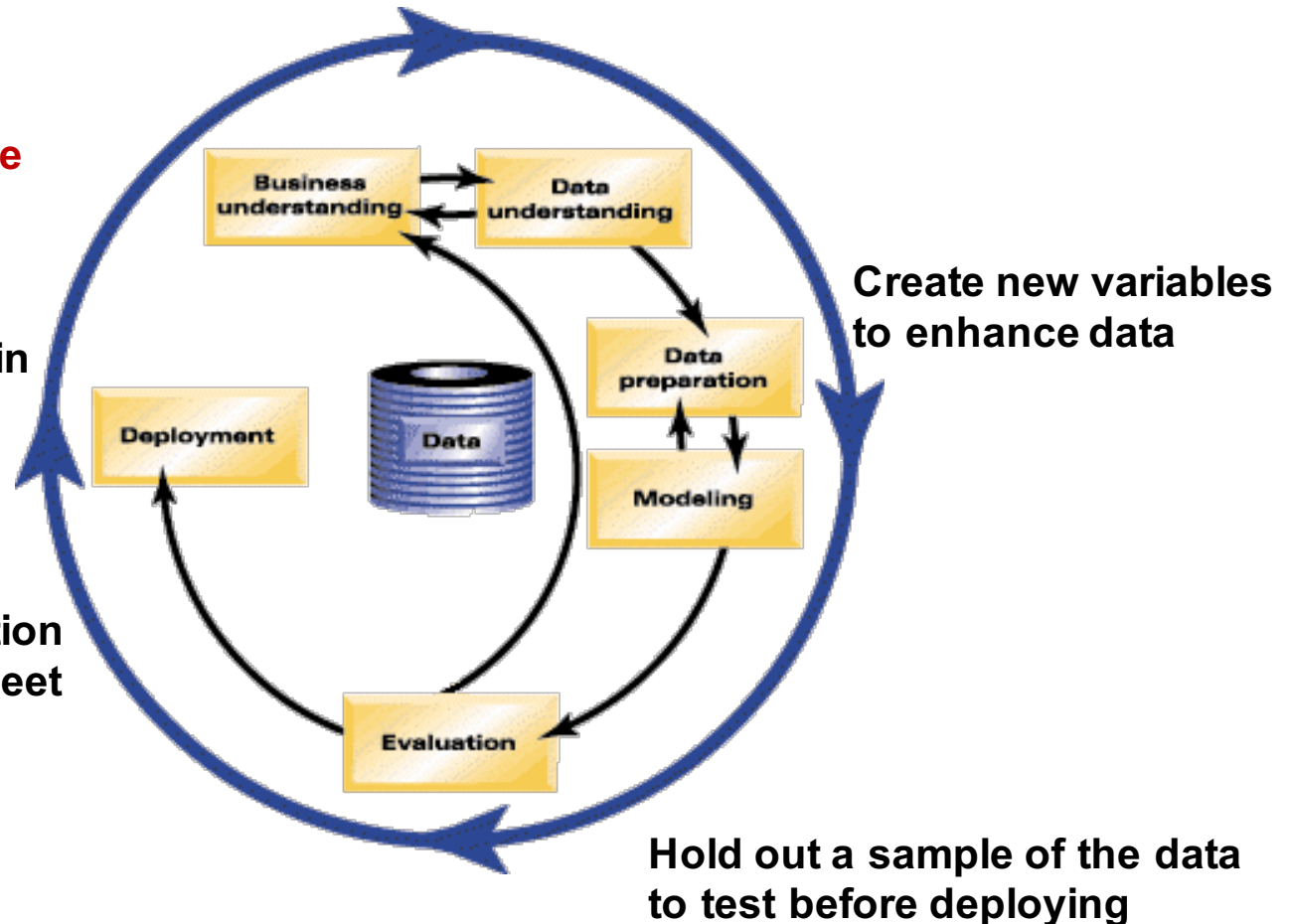
TELL ME HOW YOU CAN APPLY THIS TO **YOUR** BUSINESS

Getting Started

- **What do you want to achieve?**
- **Short project with the biggest impact**
- **ROI and Success Criteria are key to identifying early**

- **Measure results**
- **Quantify your ROI**
- **97% of customers are able to pay for their investment within 11 months!**
- **Use return to invest in the next project**
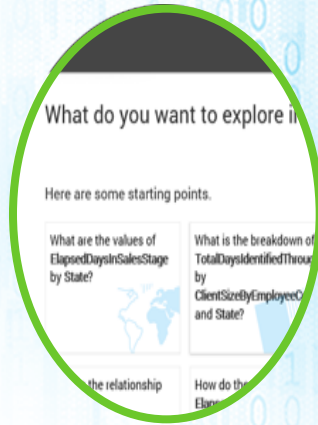
- **Embed in an application**
- **Export to a spreadsheet**

Create new variables to enhance data

Business understanding

Data understanding

Data preparation

Data

Deployment

Modeling

Evaluation

Hold out a sample of the data to test before deploying
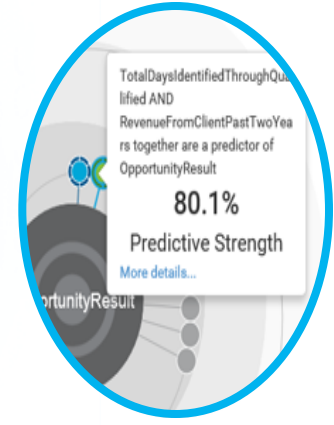
Partner projects

IBM Watson Analytics

Single Analytics Experience

Fully Automated Intelligence

Natural Language Dialogue

Guided Analytic Discovery

Single Interface … Explore > Predict > Assemble

Appendix

Suggested Books

IBM SPSS Modeler Cookbook

- From Amazon.com

  - IBM SPSS Modeler Cookbook
  - by Keith McCormick, Dean Abbott, Meta S. Brown, Tom Khabaza, Scott R. Mutchler
 - Paperback - 382 pages (October 2013) (also on Kindle)
  - ISBN : 1849685460

- Written by those who teach and have been working with IBM SPSS Modeler since its beginnings. Full of practical examples that span the full gamut of capabilities.
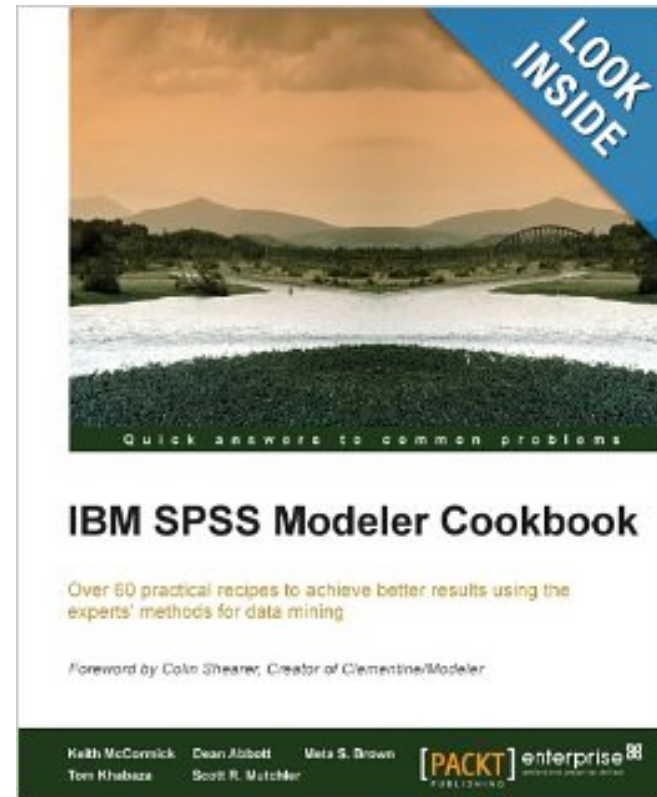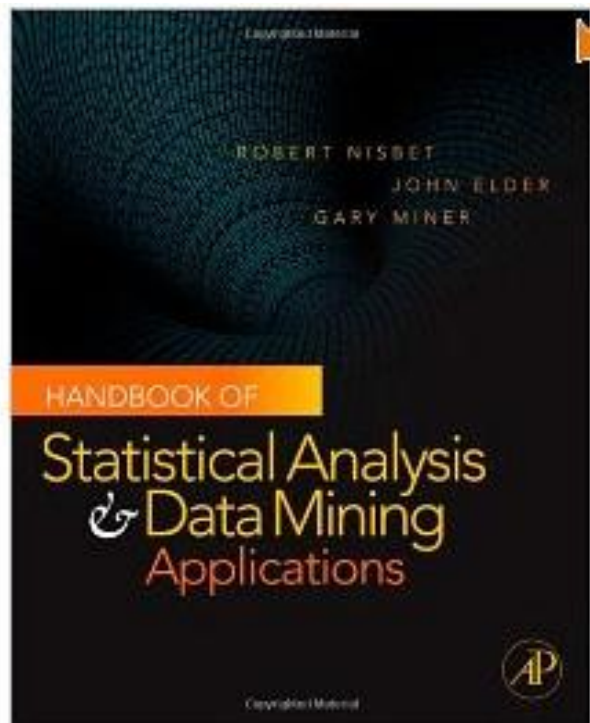
Data Mining Overview

- From Amazon.com

  - ☒**Paperback:** 512 pages
  - ☒**Publisher:** Wiley; 1 edition (December 28, 1999)
  - ☒**Language:** English
  - ☒**ISBN-10:** 0471331236
  - ☒**ISBN-13:** 978-0471331230 ;

- Good introductory text on data mining for marketing from two top communicators in the field
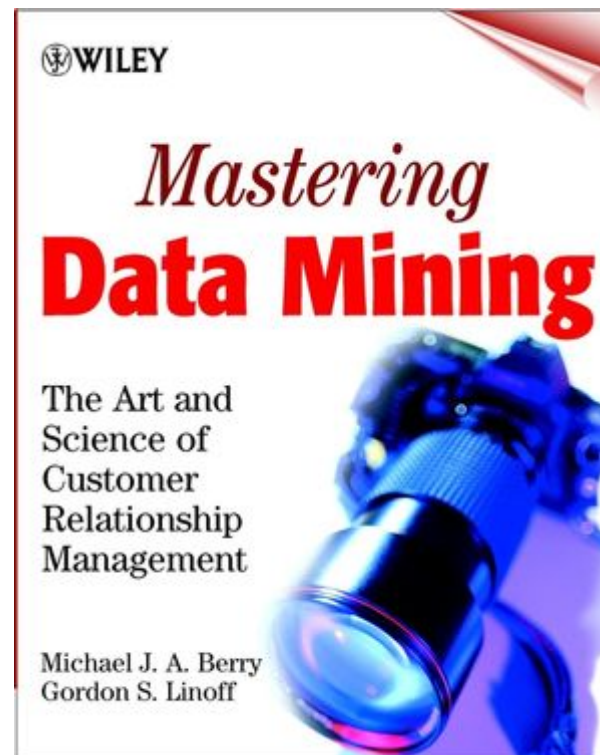
Handbook of Statistical Analysis and Data Mining Applications

- Handbook of Statistical Analysis and Data Mining Applications
- Robert Nisbet, John Elder IV, and Gary Miner
- Academic Press (2009)
- **ISBN-10:** 0123747651

- An excellent guide to many aspects of data mining including Text mining.

Data Mining Algorithms

- From Amazon.com

  ✗ Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations

  ✗ by Eibe Frank, Ian H. Witten

  ✗ Paperback - 416 pages (October 13, 1999)

  ✗ Morgan Kaufmann Publishers;

  ✗ ISBN: 1558605525;

- Best book I've found in between highly technical and introductory books. Good coverage of topics, especially trees and rules, but no neural networks.