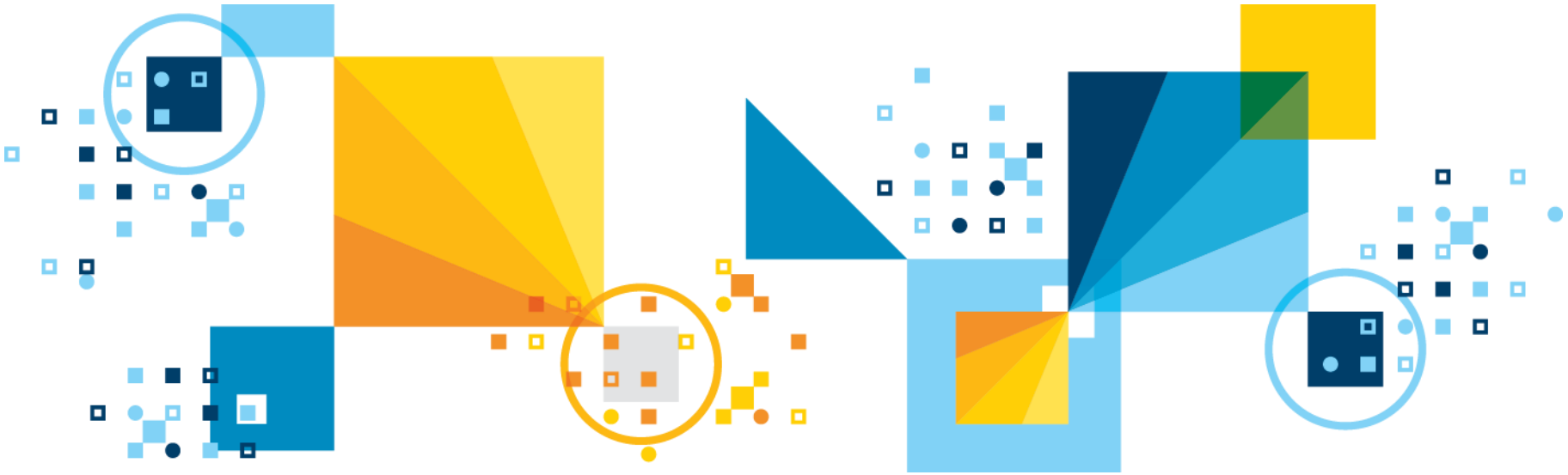
Predictive Modeling Fundamentals I
Lesson 1: Introduction



Setting the Stage….

Why this is important to know…

1. Fundamental introduction to Data Mining and its application to business problems
2. Ability to utilize software tools for advanced analytics

After this session, you will be able to…

1. Understand what Data Mining is (and isn't)
2. Understand the CRISP-DM Methodology
3. Access and explore the SPSS Modeler workbench

Speaking to you today…

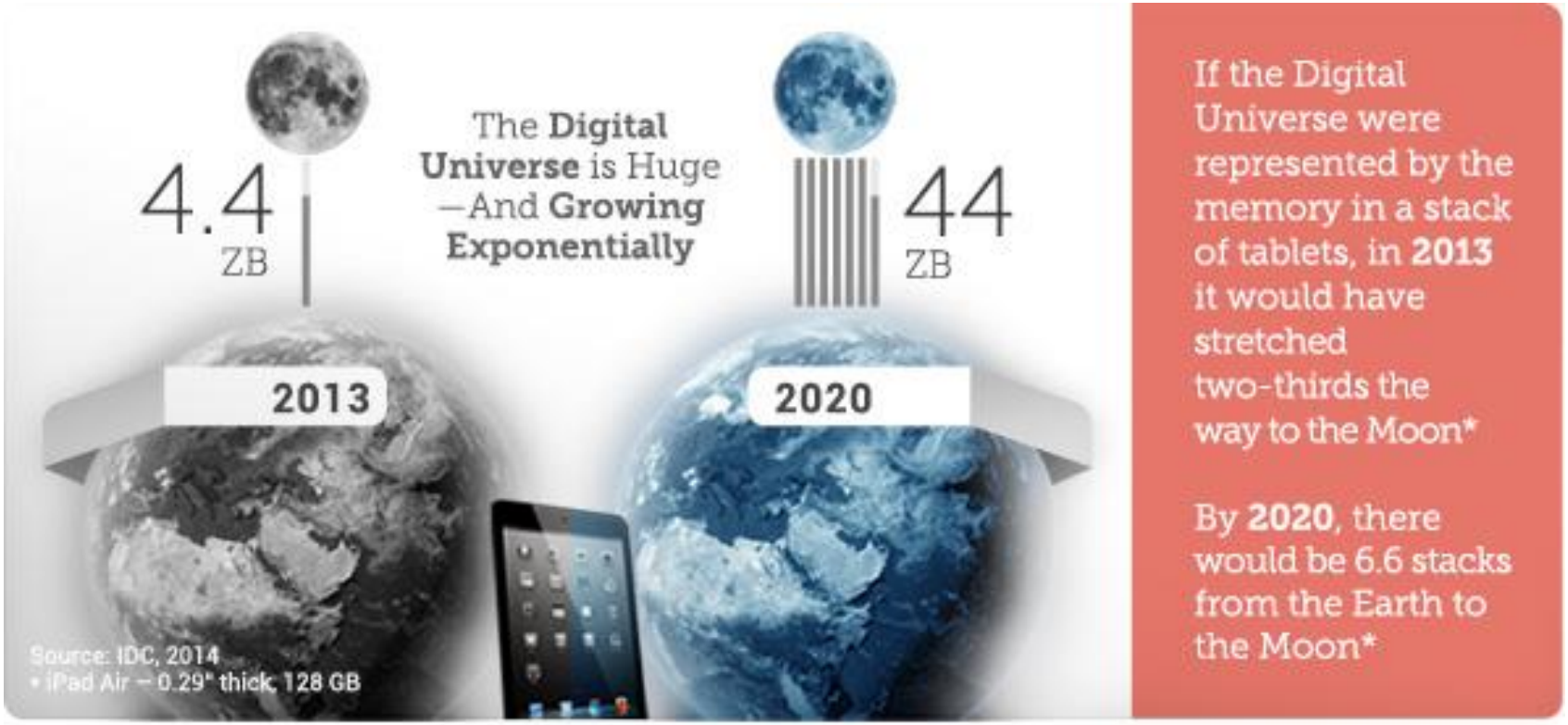
Armand Ruiz
Product Manager

Mikhail Lakirovich
Product Marketing Manager

Agenda

- Course Overview
- Introduction to Data Mining
- CRISP-DM Methodology
- Introduction to SPSS Modeler – predictive data mining workbench
- Lab 1: Installation of SPSS Modeler trial and first steps
- Lab Solution Review

Market Context

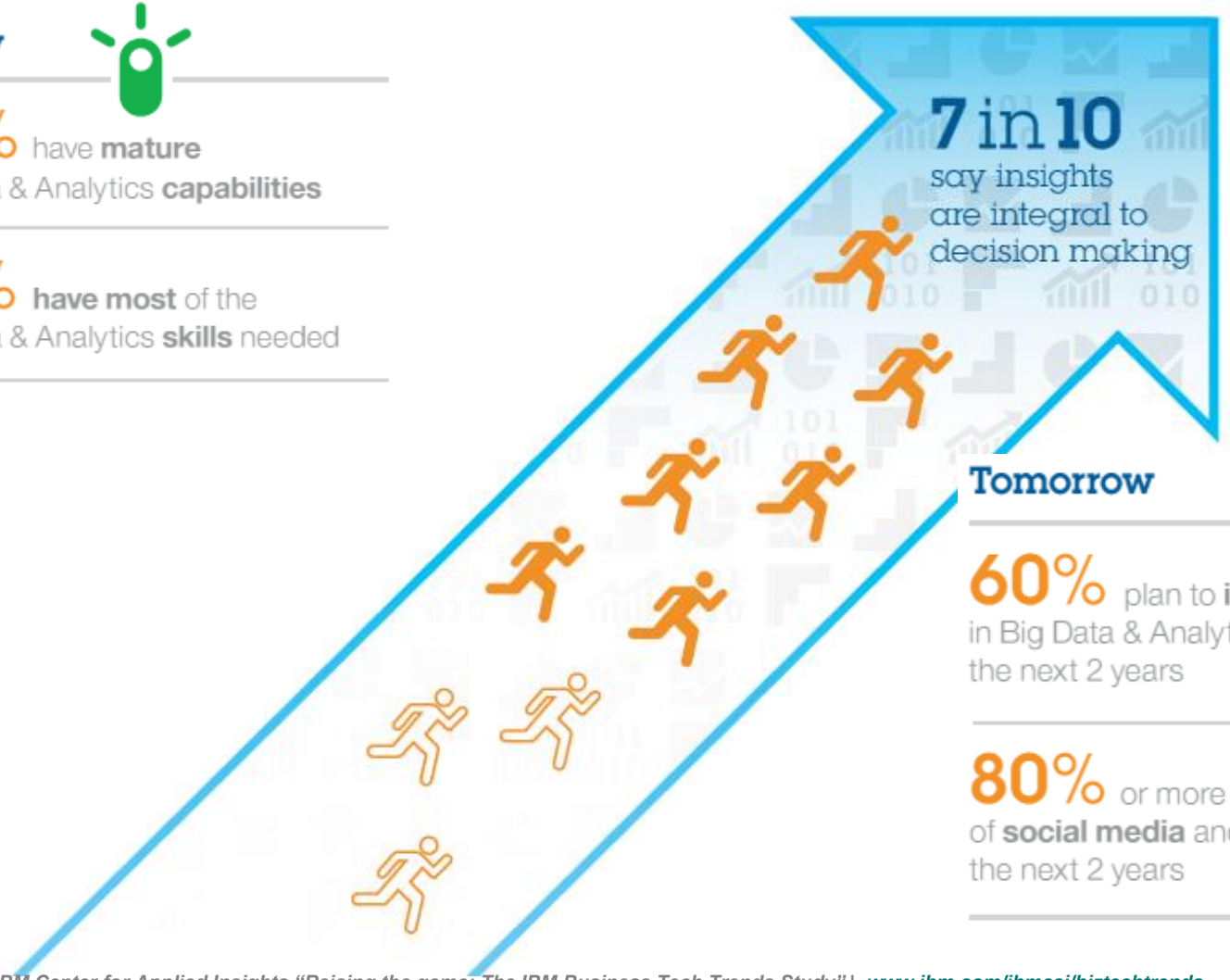


Analytics Drive Decision-Making

Today

89% have **mature** Big Data & Analytics **capabilities**

74% **have most** of the Big Data & Analytics **skills** needed



Tomorrow

60% plan to **increase investment** in Big Data & Analytics 10% or more over the next 2 years

80% or more plan to **increase** use of **social media** and **mobile analytics** over the next 2 years

Source: IBM Center for Applied Insights "Raising the game: The IBM Business Tech Trends Study" | www.ibm.com/ibmcai/biztechtrends

Introduction into Data Mining

Why Data Mining?

- Information age: terabytes and petabytes of data available. How do we consume this data, translate it into information and make it usable?

What is Data Mining?

- Process of discovering insights, patterns and relationships from large amounts of data.

What knowledge can be extracted?

- Descriptive: What has happened and why did it happen?
- Predictive: What is likely to happen next?

What can we learn?

Association Rules: Rules that indicate relationships. For example, people who buy diapers also buy beer

Classification: Finding a model that describes the data and classifies it to a set of categories. For example, people who drink and drive are more likely to have higher insurance rates

Segmentation: Grouping objects by similarity. For example, prospective customers are broken up into clusters of suburban families with children, single college students, urban empty nesters.

Why is Data Mining Important

Create new business models *(CEO)*

Transform management and financial processes *(CFO)*

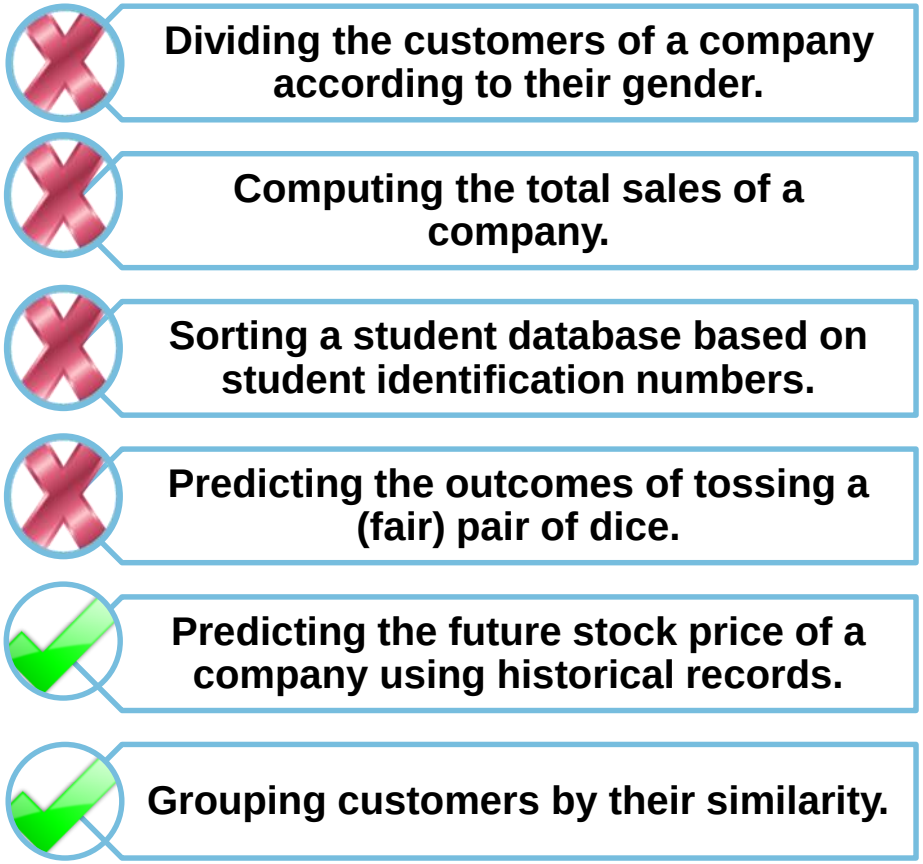
Attract, grow, and retain customers *(CMO)*

Manage risk *(CRO)*

Optimize operations; counter fraud and threats *(COO)*

Maximize insight, ensure trust, improve IT economics *(CIO, CDO)*

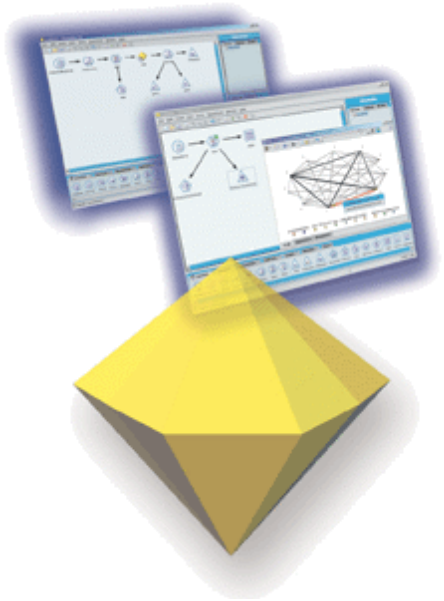
What Data Mining is and is not



8

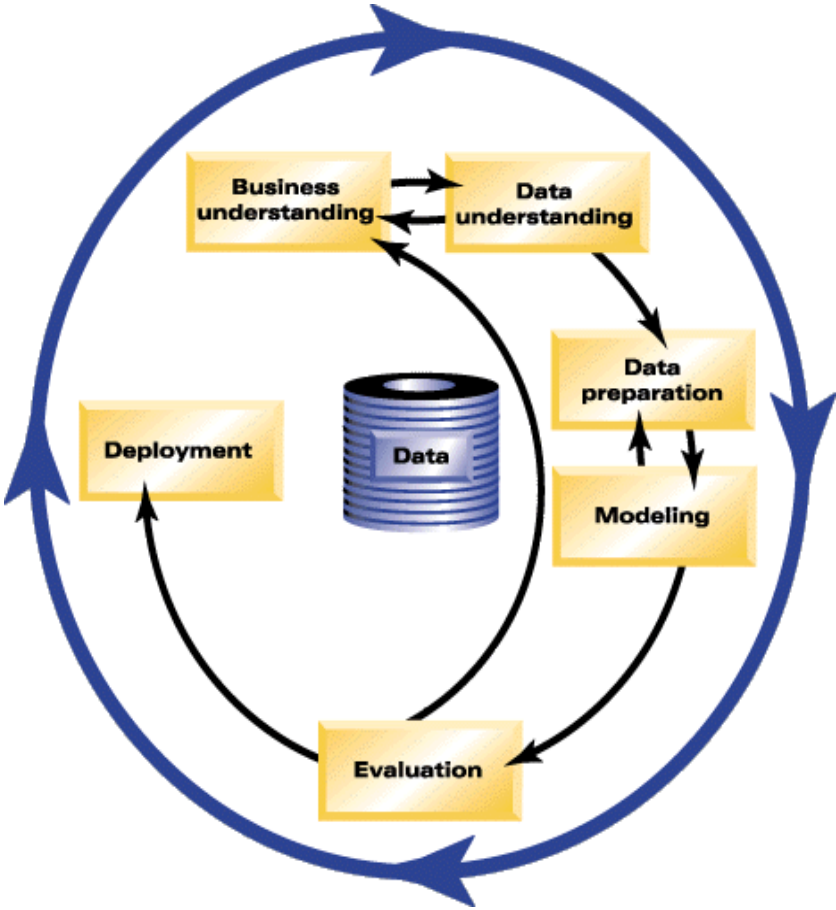
Predictive Models and Data Mining Algorithms

- Data mining algorithms create predictive models by analyzing the data automatically to find patterns
- Predictive models contain the patterns that have been found, and use them to make predictions
- Some potential predictions
  - Customer likely to leave
  - Typical high-value customer
  - Credit risk score
  - Legitimacy of transaction
  - Cancer diagnosis

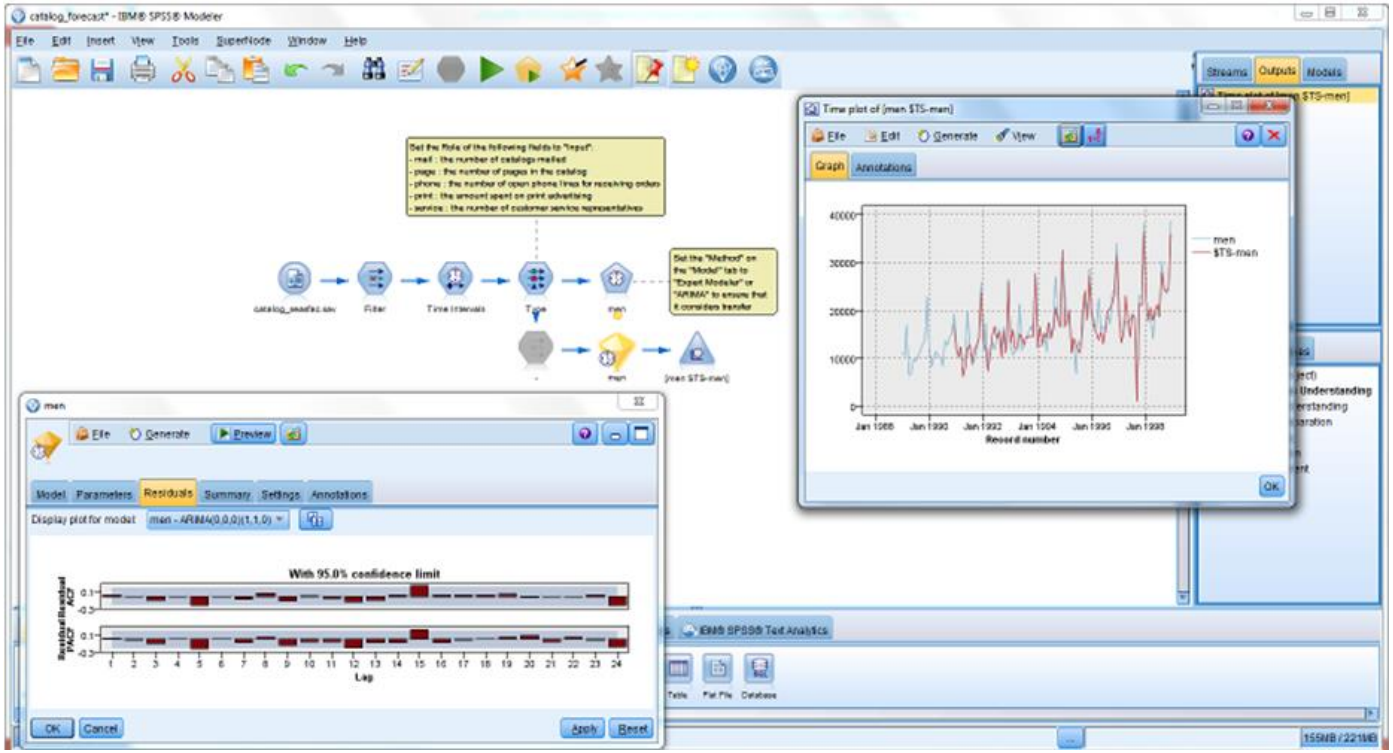


Data Mining Process: CRISP-DM

- Cross-Industry Process for Data Mining
- Iterative process: continuously learns and improves through accumulation of insight and predictive power
- Business Understanding: what should be accomplished from a business perspective
- Data Understanding: Acquiring the data needed to accomplish the objective
- Data Preparation: Selecting and cleaning the data. May transform/aggregate for analysis
- Modeling: Selecting technique, building and training the model, assessing accuracy
- Evaluation: Does the model meet business objectives
- Deployment: Strategy for deploying the model



IBM SPSS Modeler



- Puts predictive power into the hands of a business analyst
- Provides the sophistication needed by an expert
- Includes a range of advanced data manipulation and analytical algorithms
- Flexible Deployment Options