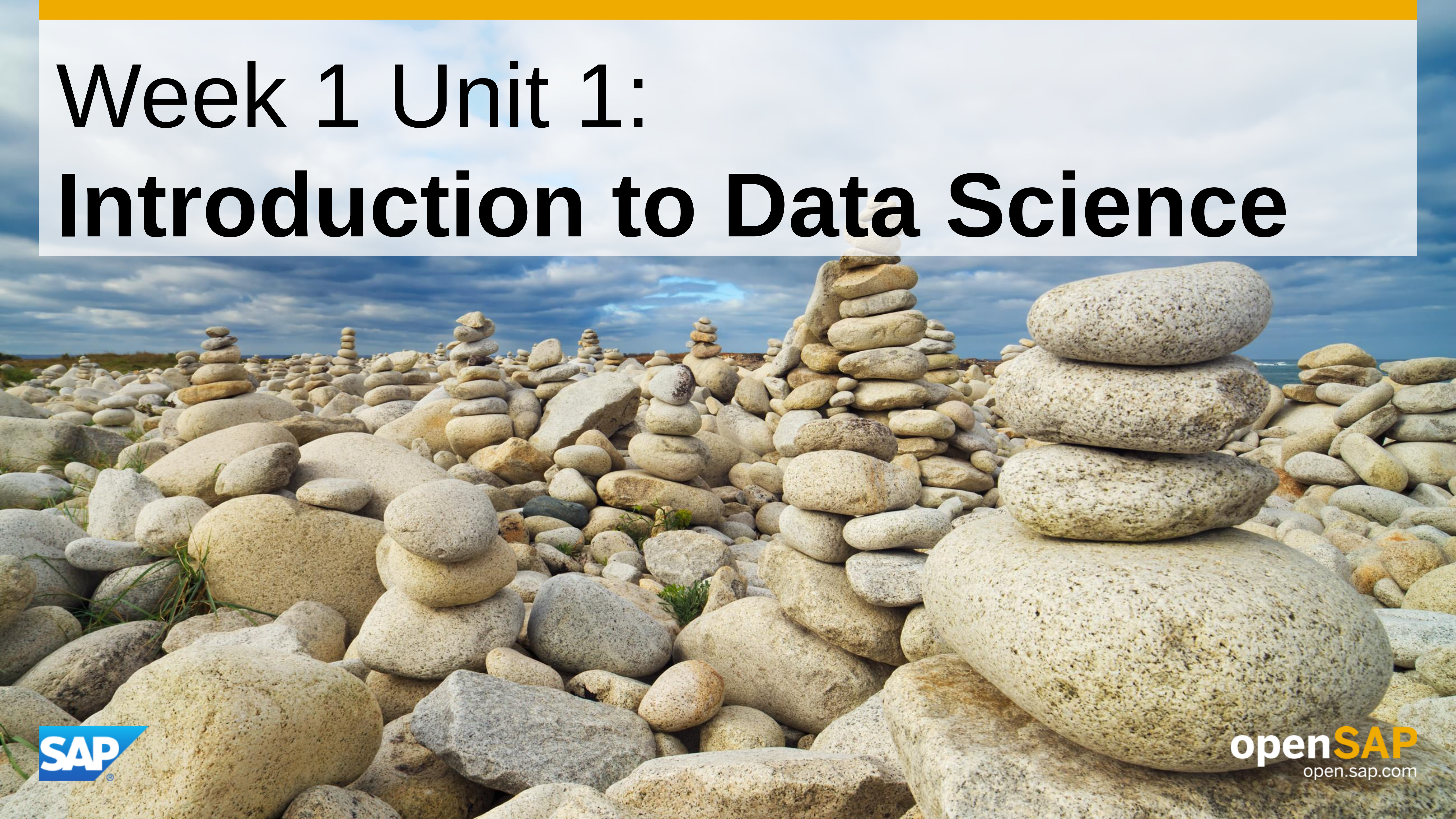


Week 1 Unit 1:
Introduction to Data Science

openSAP
open.sap.com

Introduction to Data Science

The next 6 weeks

What to expect in the next 6 weeks?

Introduction to Data Science

Curriculum flow (weeks 1-3)

1 Business & Data Understanding

- Introduction to Data Science
- Introduction to Project Methodologies
- Business Understanding Phase – Overview
- Defining Project Success Criteria
- Data Understanding Phase – Overview
- Initial Data Analysis & Exploratory Data Analysis

Weekly Assignment

2 Data Preparation

- Data Preparation Phase – Overview
- Predictive Modeling Methodology – Overview
- Data Manipulation
- Selecting Data – Variable and Feature Selection
- Data Encoding

Weekly Assignment

3 Modeling (1)

- Modeling Phase – Overview
- Detecting Anomalies
- Association Analysis
- Cluster Analysis
- Classification Analysis with Regression

Weekly Assignment

Introduction to Data Science

Curriculum flow (weeks 4-6)

4 Modeling (2)

- Classification Analysis with Decision Trees
- Classification Analysis with KNN, NN, and SVM
- Time Series Analysis
- Ensemble Methods
- Simulation & Optimization
- Automated Modeling

Weekly Assignment

5 Evaluation

- Evaluation Phase – Overview
- Model Performance Metrics
- Model Testing
- Improving Model Performance

Weekly Assignment

6 Deployment & Maintenance

- Deployment Phase – Overview
- Deployment Options
- Monitoring & Maintenance
- Automating Deployment & Maintenance
- Myths & Challenges
- Data Science Applications and References

Weekly Assignment

Final Exam

Introduction to Data Science

Cumulative points lead to record of achievement

Watch the deadlines! ⚠

Participate in Weekly Assignment (Weeks1-6)

6 assignments
6 x 30 = 180 points

Final Exam (Week 7)

180 points

Record of Achievement

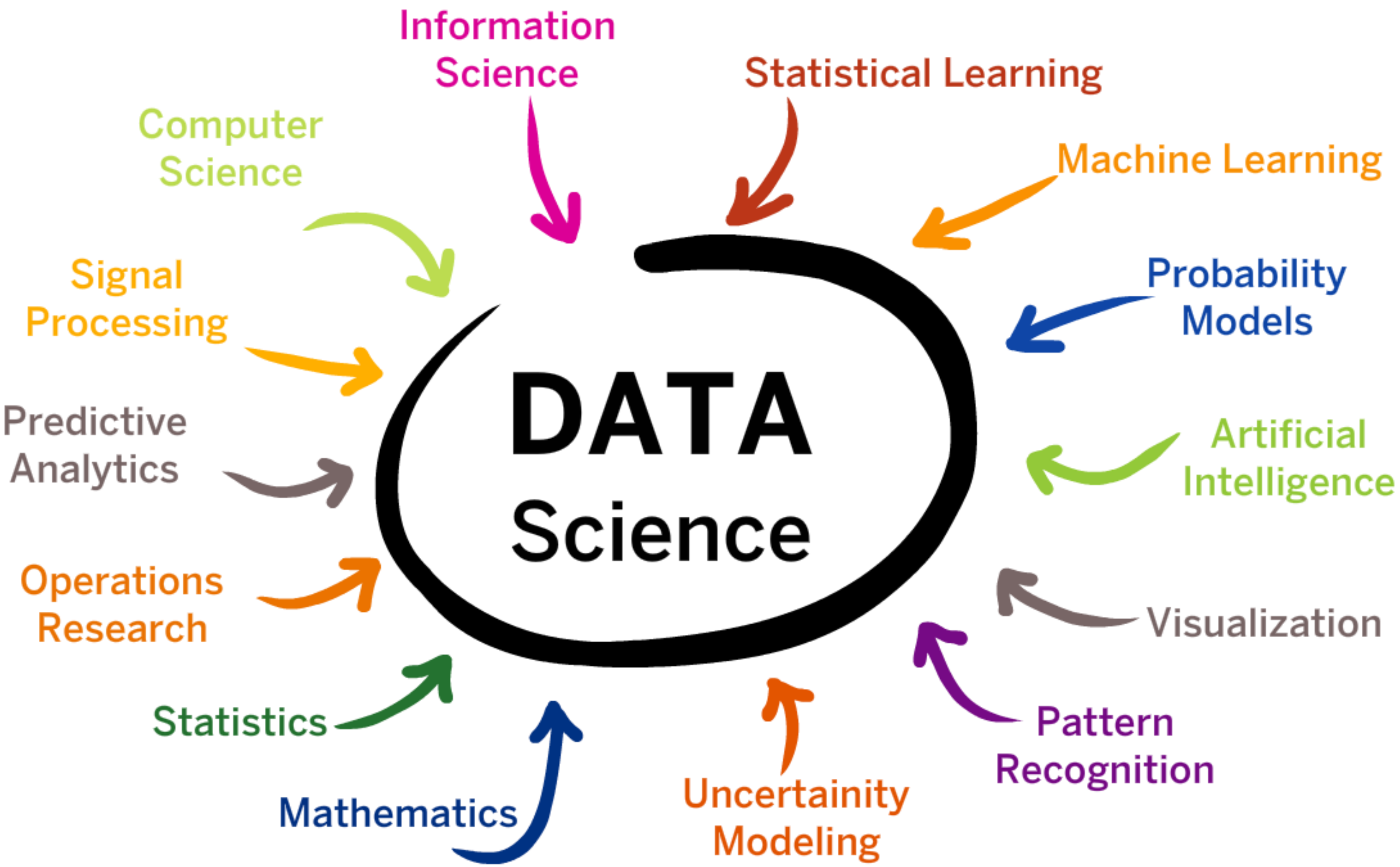
When results above
180 points

Introduction to Data Science

What is data science?

Data science is an interdisciplinary field about processes and systems that enable the extraction of knowledge or insights from data.

Data science employs techniques and theories drawn from a wide range of disciplines.



DATA Science

Information Science

Statistical Learning

Machine Learning

Probability Models

Artificial Intelligence

Visualization

Pattern Recognition

Uncertainity Modeling

Mathematics

Statistics

Operations Research

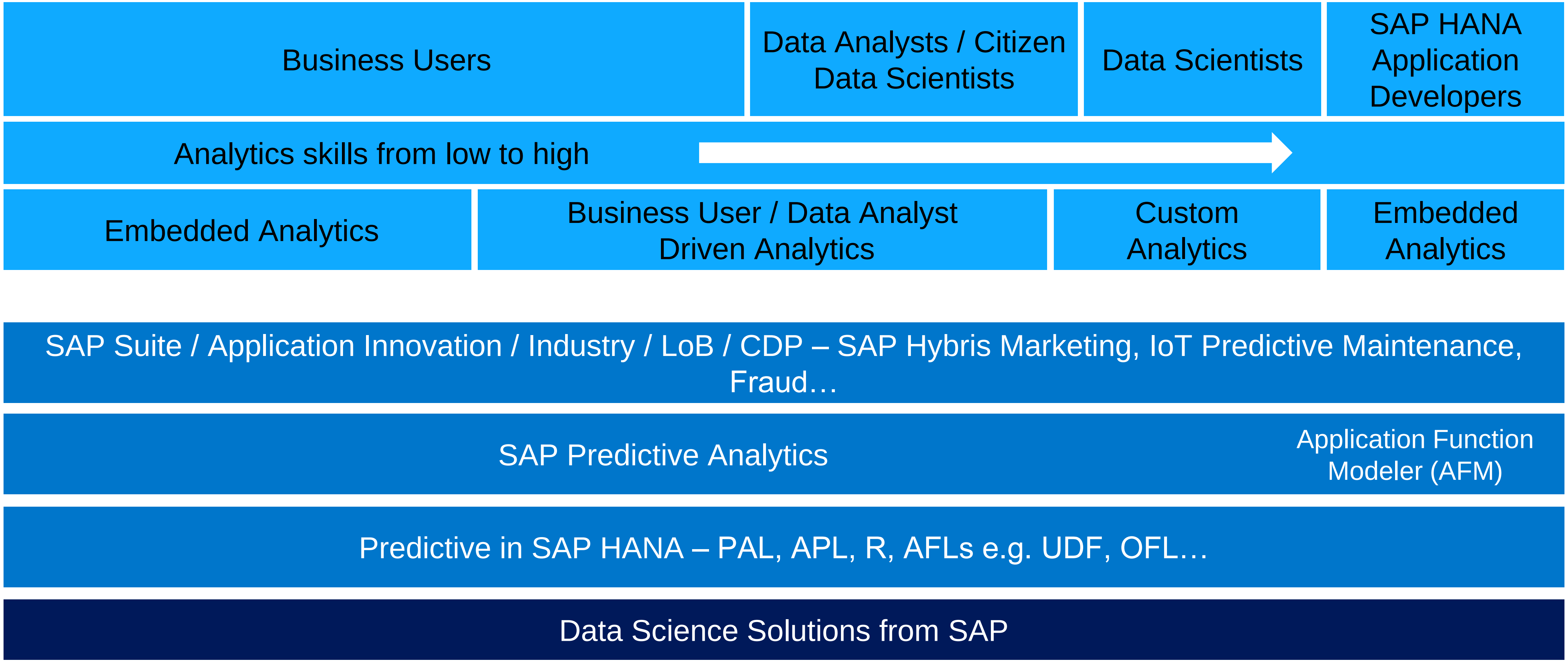
Predictive Analytics

Signal Processing

Computer Science

Introduction to Data Science

Data science personas



Business Users

Data Analysts / Citizen Data Scientists

Data Scientists

SAP HANA Application Developers

Analytics skills from low to high

Embedded Analytics

Business User / Data Analyst Driven Analytics

Custom Analytics

Embedded Analytics

SAP Suite / Application Innovation / Industry / LoB / CDP – SAP Hybris Marketing, IoT Predictive Maintenance, Fraud…

SAP Predictive Analytics

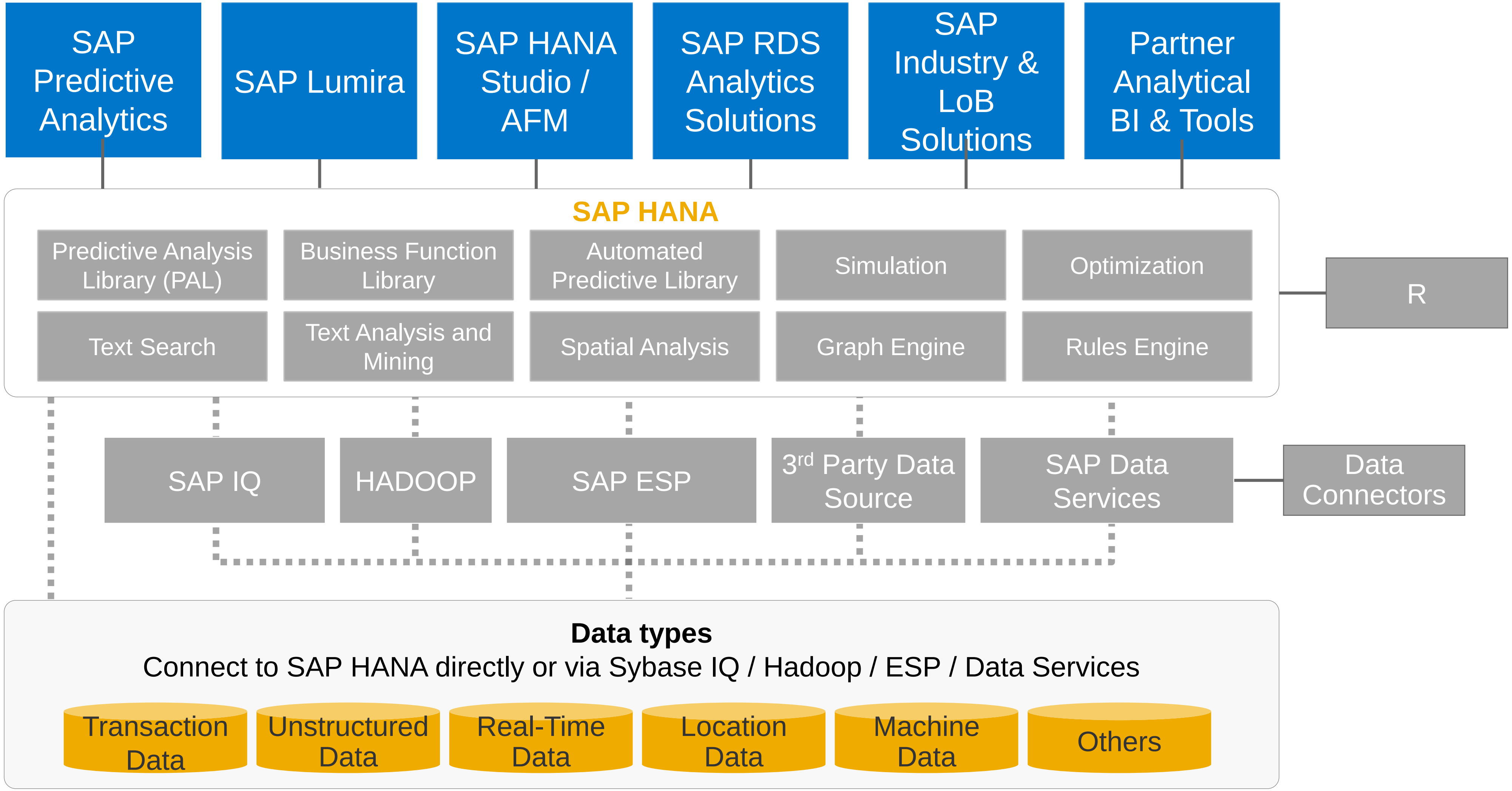
Application Function Modeler (AFM)

Predictive in SAP HANA – PAL, APL, R, AFLs e.g. UDF, OFL…

Data Science Solutions from SAP

Introduction to Data Science

Data science solutions from SAP



Data types

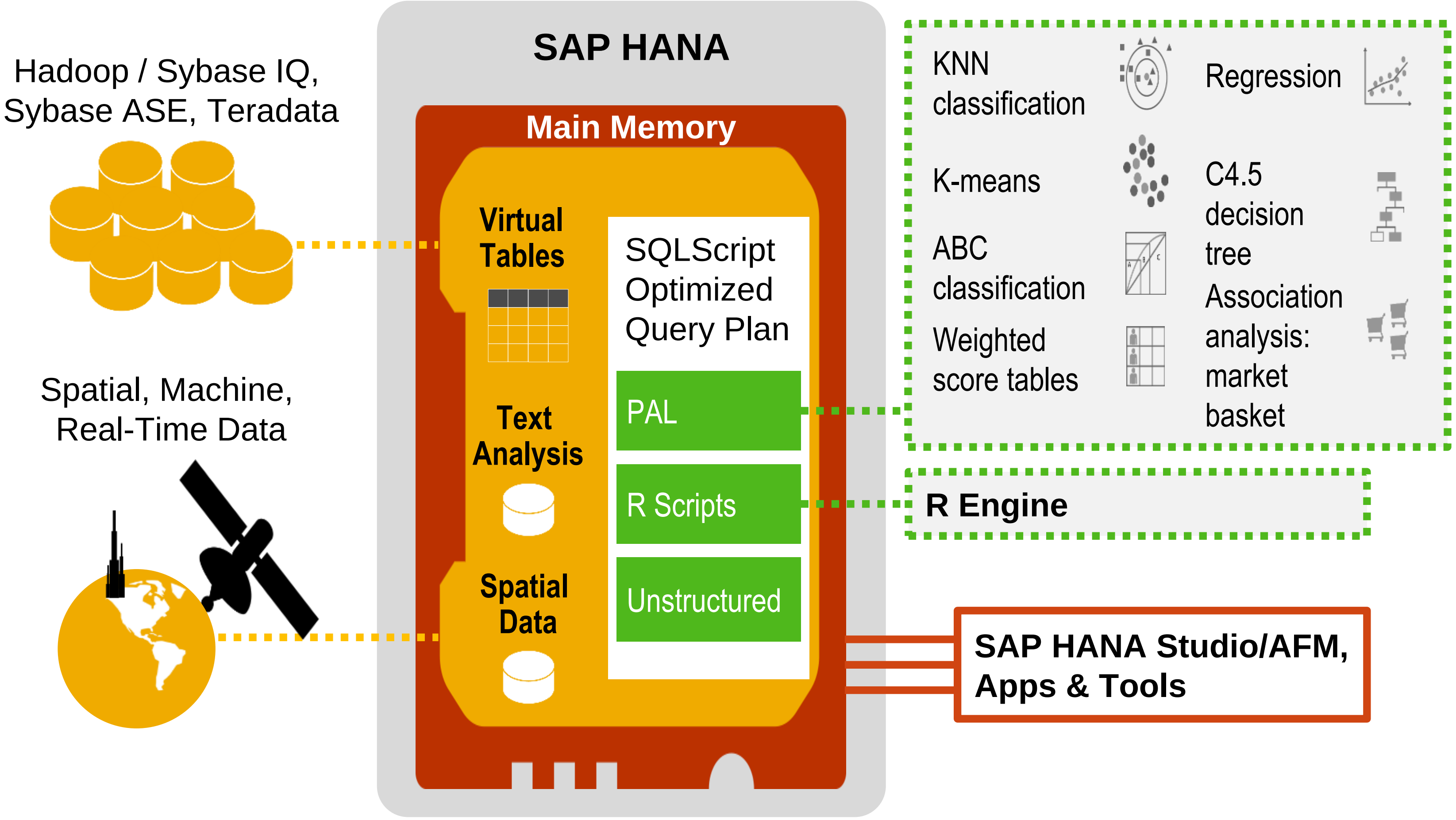
Connect to SAP HANA directly or via Sybase IQ / Hadoop / ESP / Data Services

Introduction to Data Science

SAP HANA Predictive Analysis Library (PAL)

Build High-Performance Predictive Apps

- The SAP HANA Predictive Analysis Library (PAL) is a built-in C++ library for performing in-memory data mining and statistical calculations.
- PAL is designed to provide high performance on large datasets for real-time analytics.



Introduction to Data Science

SAP HANA Predictive Analysis Library (PAL) algorithms

- SAP HANA Predictive Analysis Library (PAL) contains a wide range of algorithms that can be deployed for in-HANA and standalone data science applications.
- A wide range of algorithms are available for the following types of analysis:

SAP HANA Predictive Analysis Library

Association Analysis

Classification Analysis

Regression

Cluster Analysis

Time Series Analysis

Probability Distribution

Outlier Detection

Link Prediction

Data Preparation

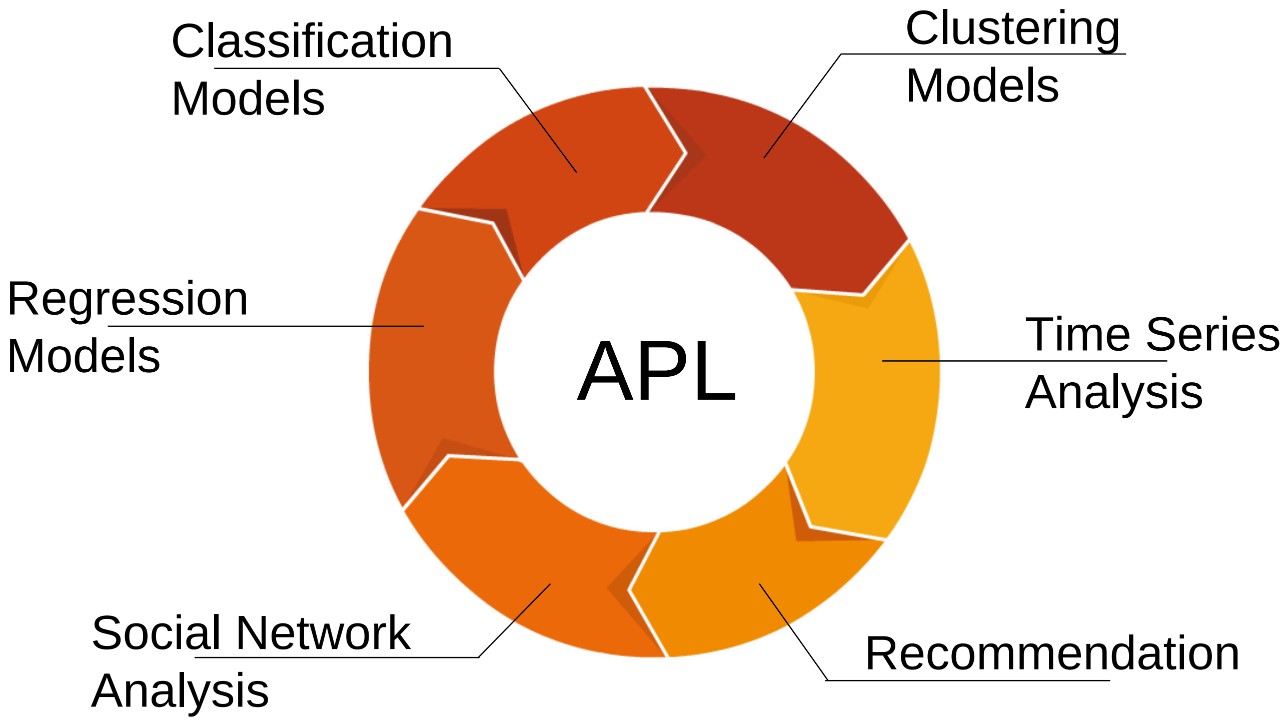
Statistic Functions (Univariate)

Statistic Functions (Multivariate)

Introduction to Data Science

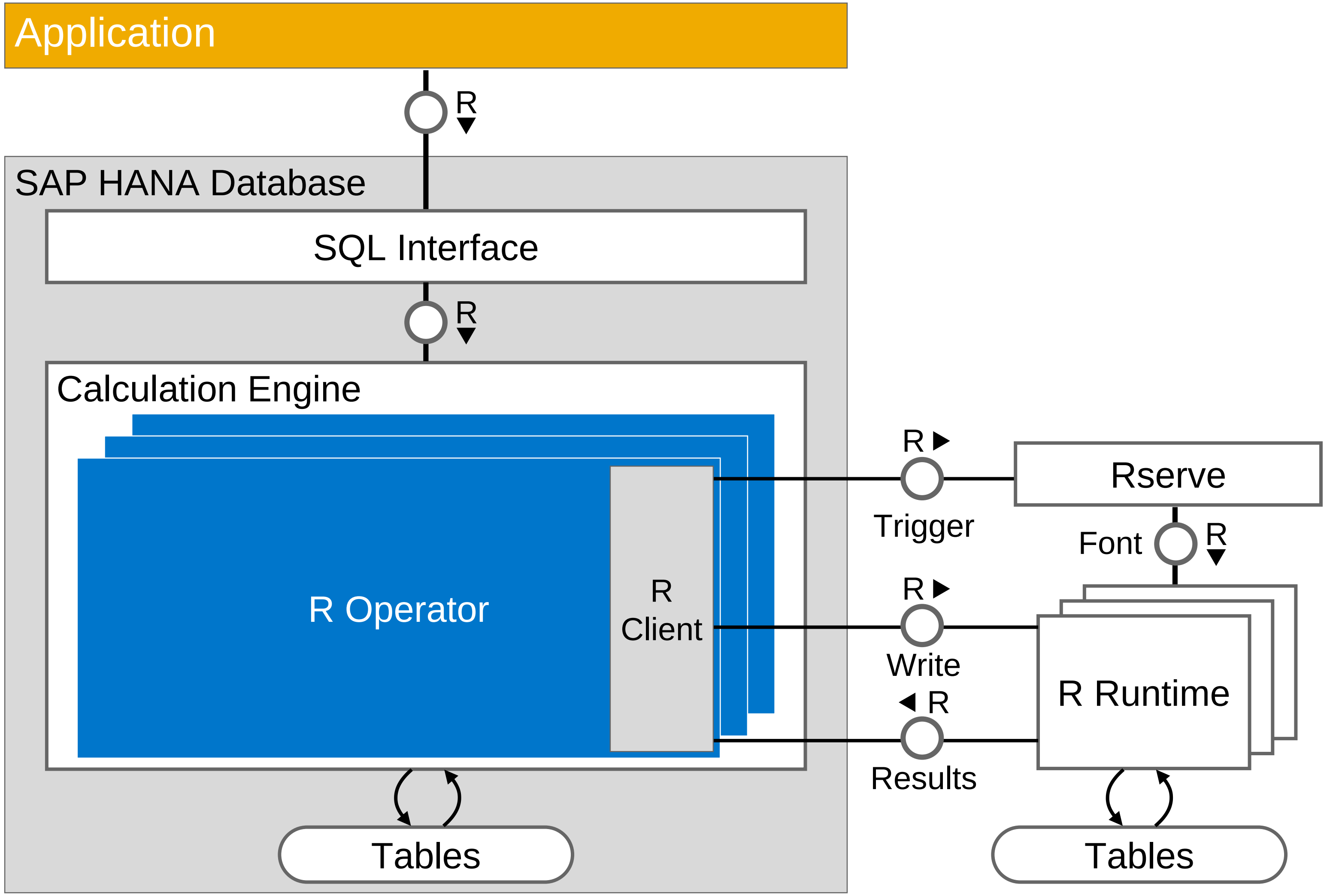
SAP HANA Automated Predictive Library (APL) algorithms

- SAP HANA APL is an application function library (AFL) that lets you use the data mining capabilities of the SAP Predictive Analytics automated analytics engine on your customer datasets stored in SAP HANA.
- You can create a wide range of models to answer your business questions.



Introduction to Data Science

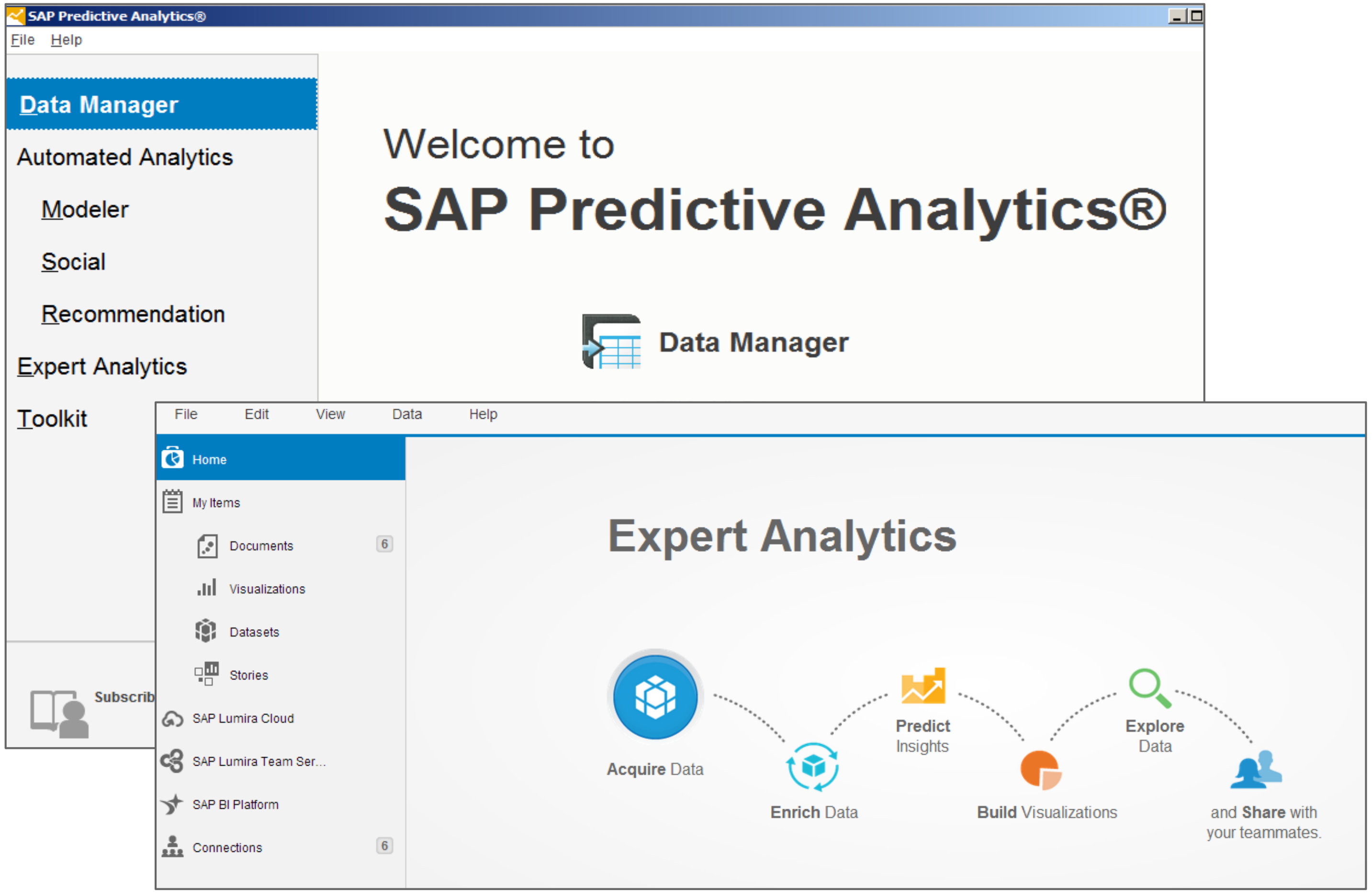
R integration for SAP HANA and standalone



Introduction to Data Science

SAP Predictive Analytics

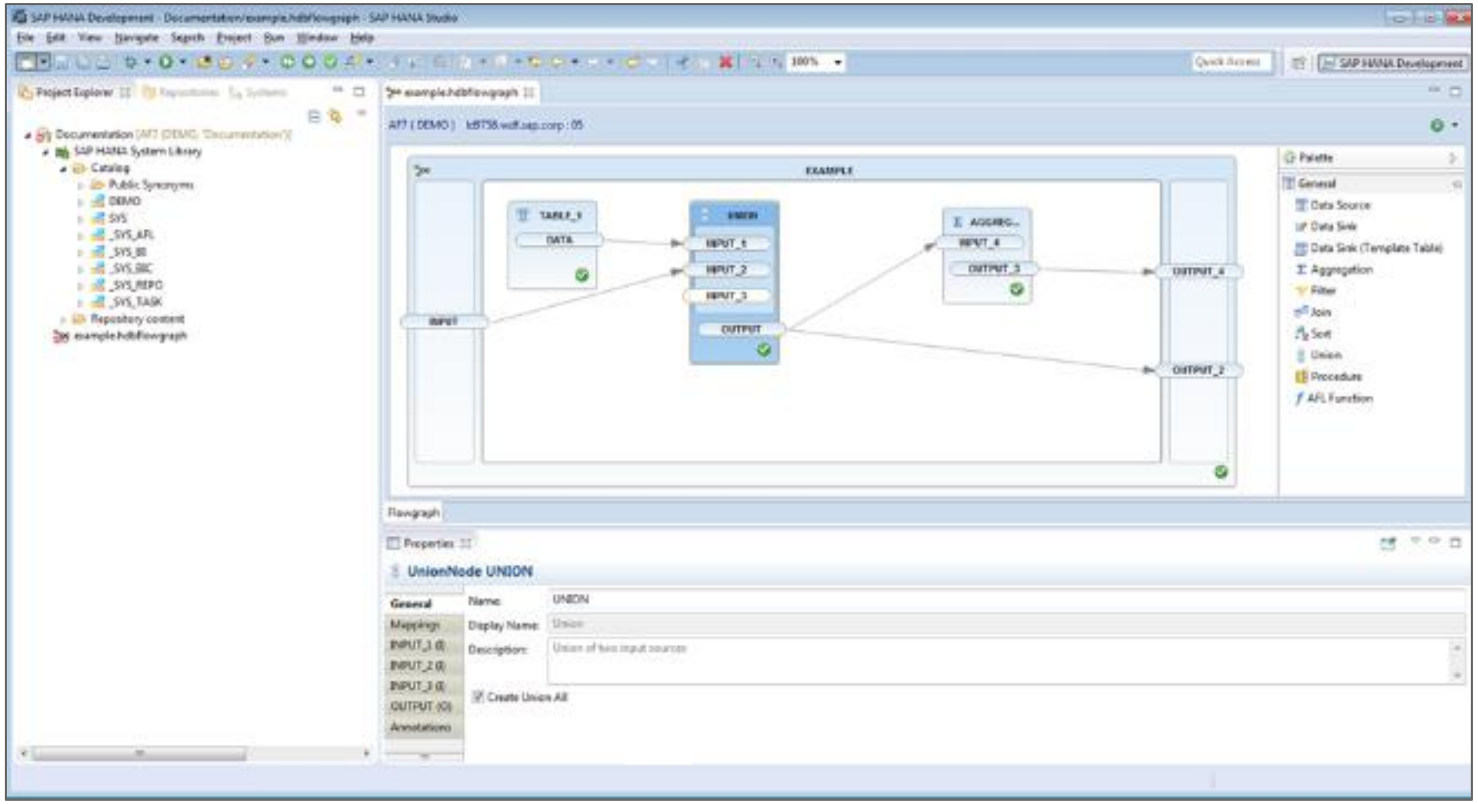
- SAP Predictive Analytics is built for both data scientists and business / data analysts, making predictive analytics accessible to a broad spectrum of users.
- Automated and expert modes
- Used to automate data preparation, predictive modeling, and deployment tasks
- Rich pre-built modelling functionality
- PAL, APL, and R language support
- Advanced visualization
- Native integration with SAP HANA



Introduction to Data Science

Application function modeler (AFM)

- Graphical tool to build advanced applications in SAP HANA
- Web-based flow-graph editor
  - Support for AFL, R, SDI, & SDQ
  - Used to create procedures or task runtime operations
  - Interoperability with SAP HANA studio AFM
- SAP HANA studio-based AFM
  - PAL function support including time series, clustering, classification, and statistics
  - General usability enhancements for an easier, simpler, and more functional experience



A77 (DEMO) ld8758.wdf.sap.corp : 05

UnionNode UNION

Name: UNION

Display Name: Union

Description: Union of two input sources

☑ Create Union All

Thank you

Contact information:

open@sap.com

openSAP
open.sap.com

© 2016 SAP SE or an SAP affiliate company. All rights reserved.

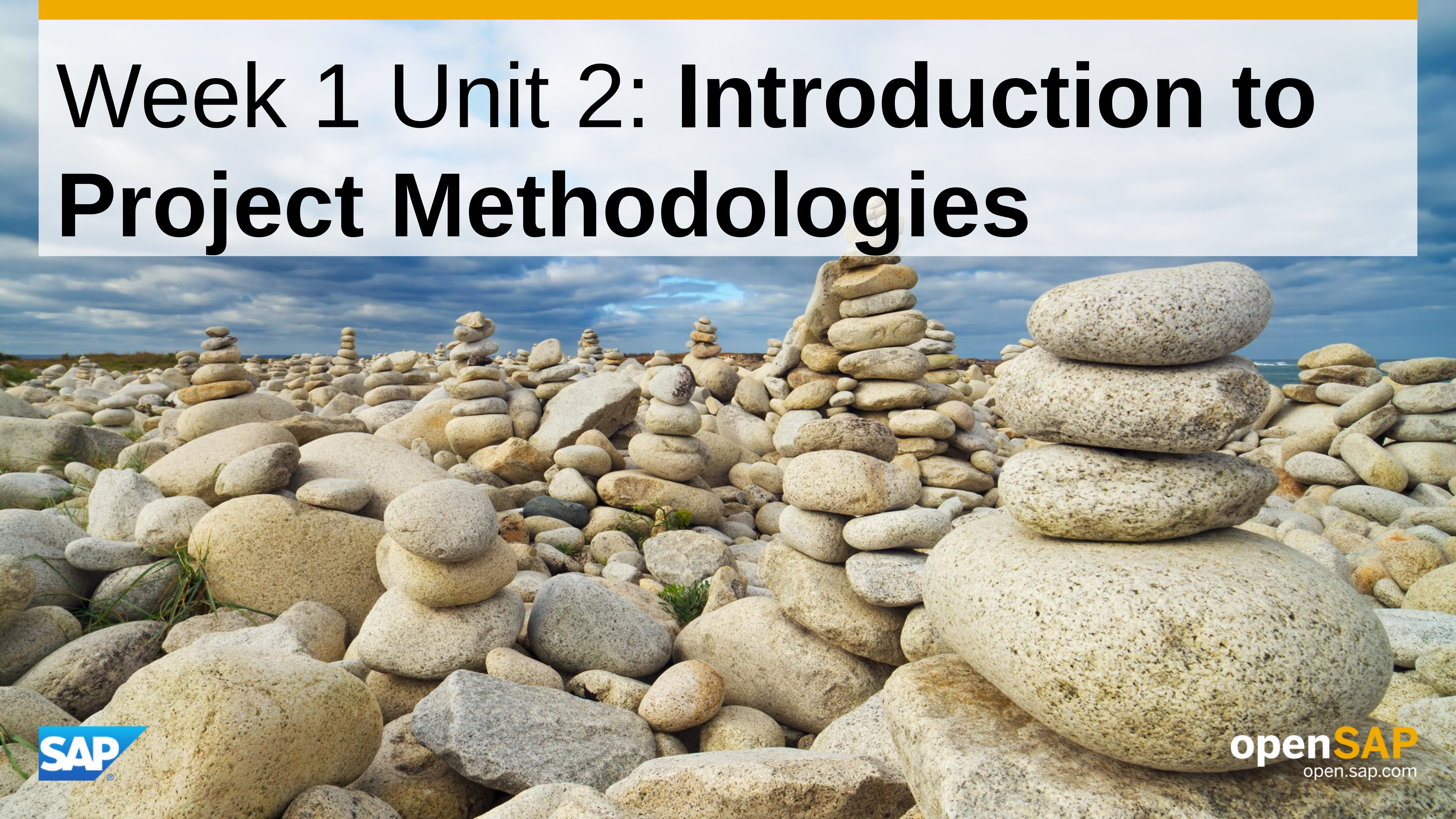
Some software products marketed by SAP SE and its distributors contain proprietary software components of other software vendors.

National product specifications may vary.

These materials are provided by SAP SE or an SAP affiliate company for informational purposes only, without representation or warranty of any kind, and SAP SE or its affiliated companies shall not be liable for errors or omissions with respect to the materials. The only warranties for SAP SE or SAP affiliate company products and services are those that are set forth in the express warranty statements accompanying such products and services, if any. Nothing herein should be construed as constituting an additional warranty.

In particular, SAP SE or its affiliated companies have no obligation to pursue any course of business outlined in this document or any related presentation, or to develop or release any functionality mentioned therein. This document, or any related presentation, and SAP SE's or its affiliated companies' strategy and possible future developments, products, and/or platform directions and functionality are all subject to change and may be changed by SAP SE or its affiliated companies at any time for any reason without notice. The information in this document is not a commitment, promise, or legal obligation to deliver any material, code, or functionality. All forward-looking statements are subject to various risks and uncertainties that could cause actual results to differ materially from expectations. Readers are cautioned not to place undue reliance on these forward-looking statements, which speak only as of their dates, and they should not be relied upon in making purchasing decisions.

Week 1 Unit 2: **Introduction to Project Methodologies**

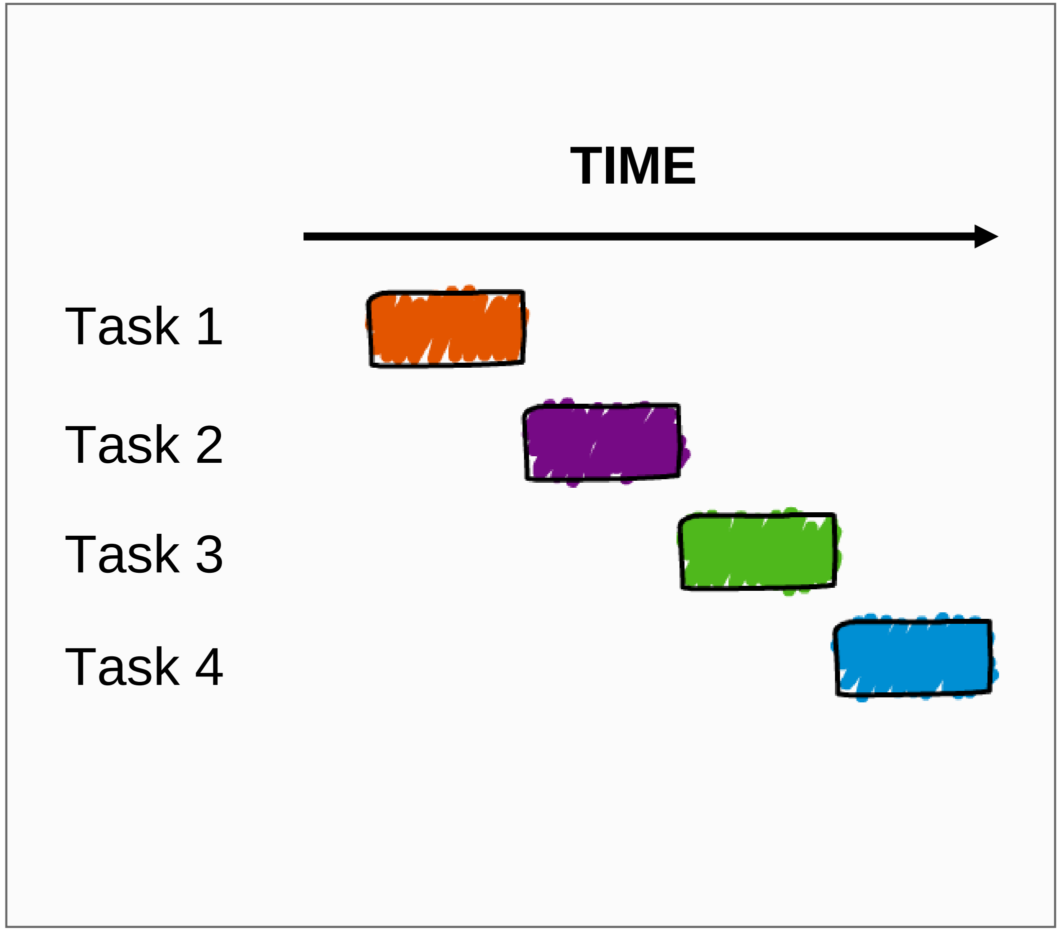


open.sap.com

Introduction to Project Methodologies

Why should there be a project methodology?

- The data science process must be reliable and repeatable by people with little data science background.
- A project methodology:
  - Provides a framework for recording experience
  - Allows projects to be replicated
  - Provides an aid to project planning and management
  - Is a “comfort factor” for new adopters
  - Reduces dependency on “stars”



TIME

Task 1

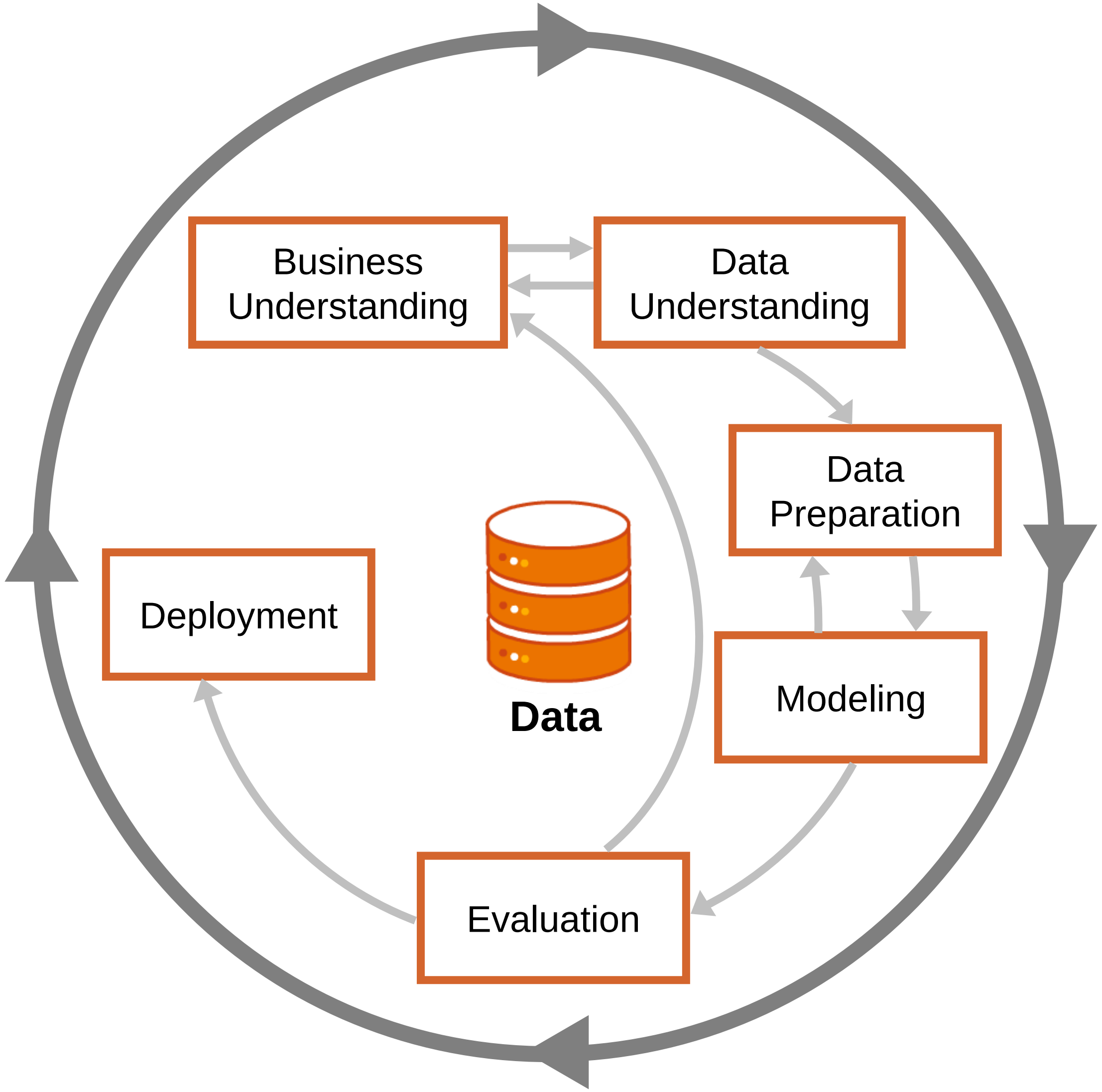
Task 2

Task 3

Task 4

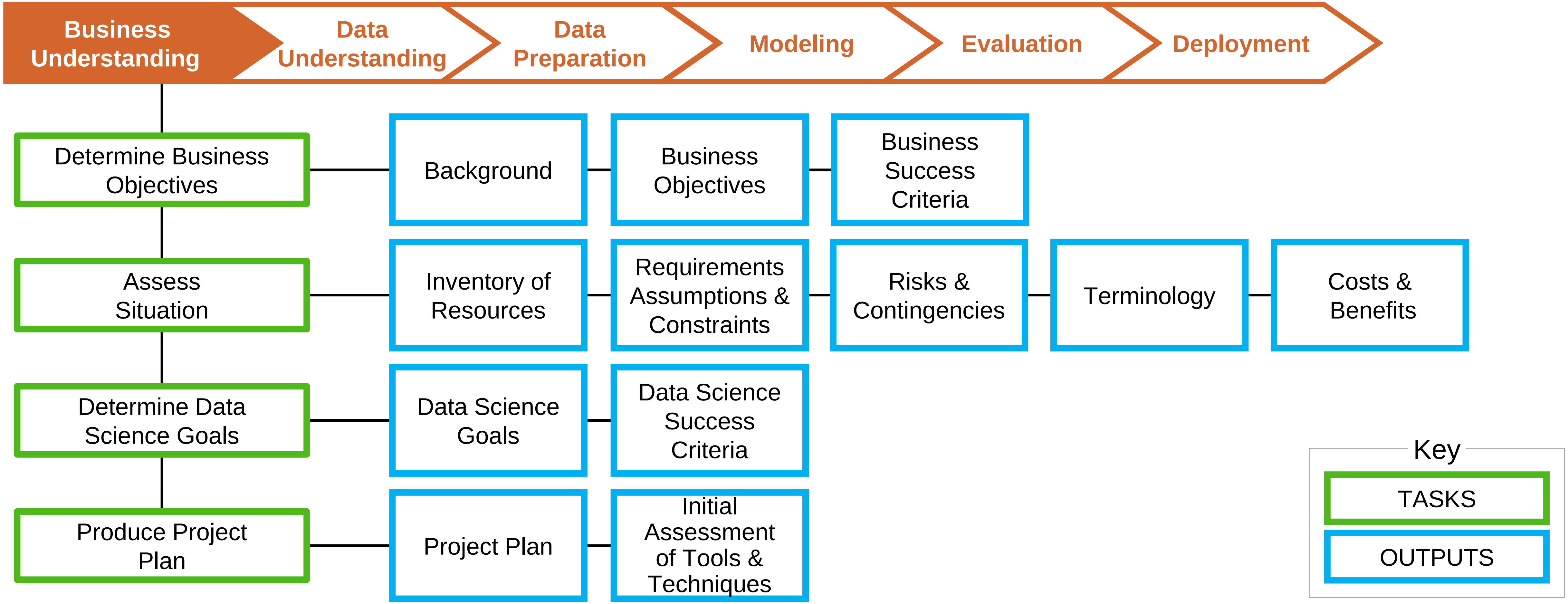
Introduction to Project Methodologies

Cross-industry standard process for data mining (CRISP-DM)



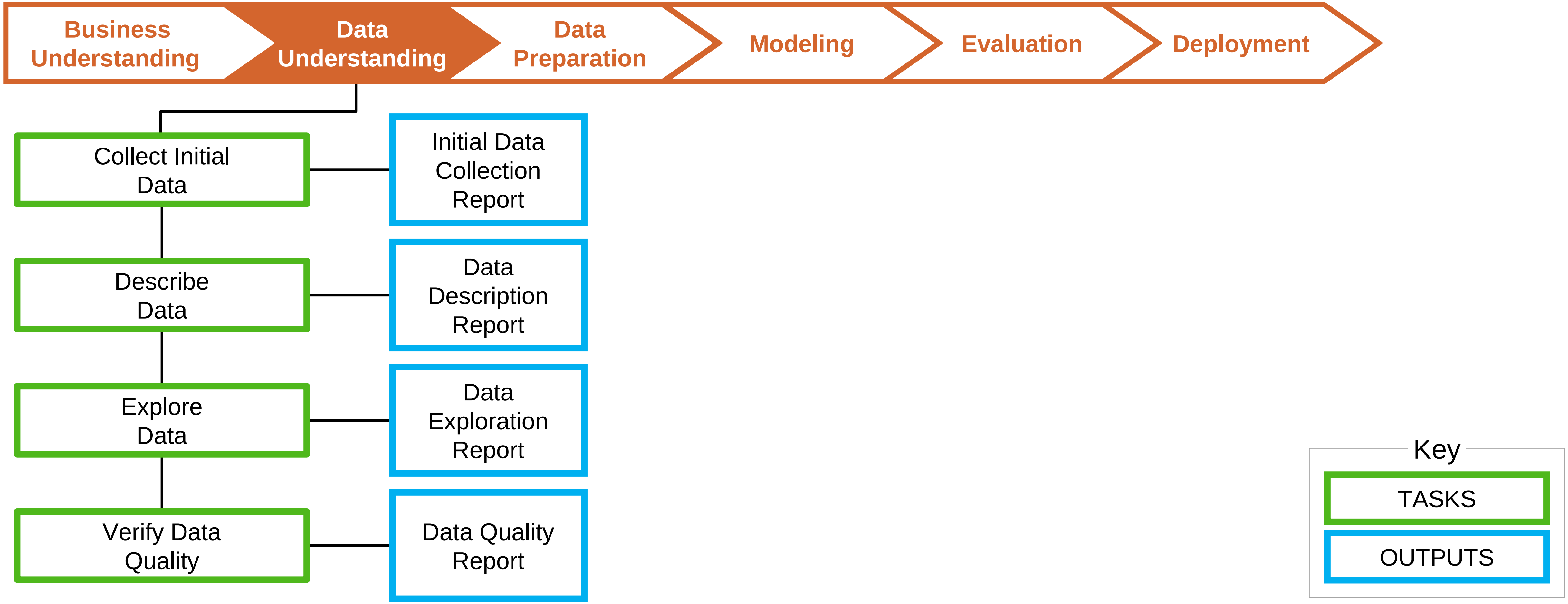
Introduction to Project Methodologies

CRISP-DM – Phase 1: Business Understanding



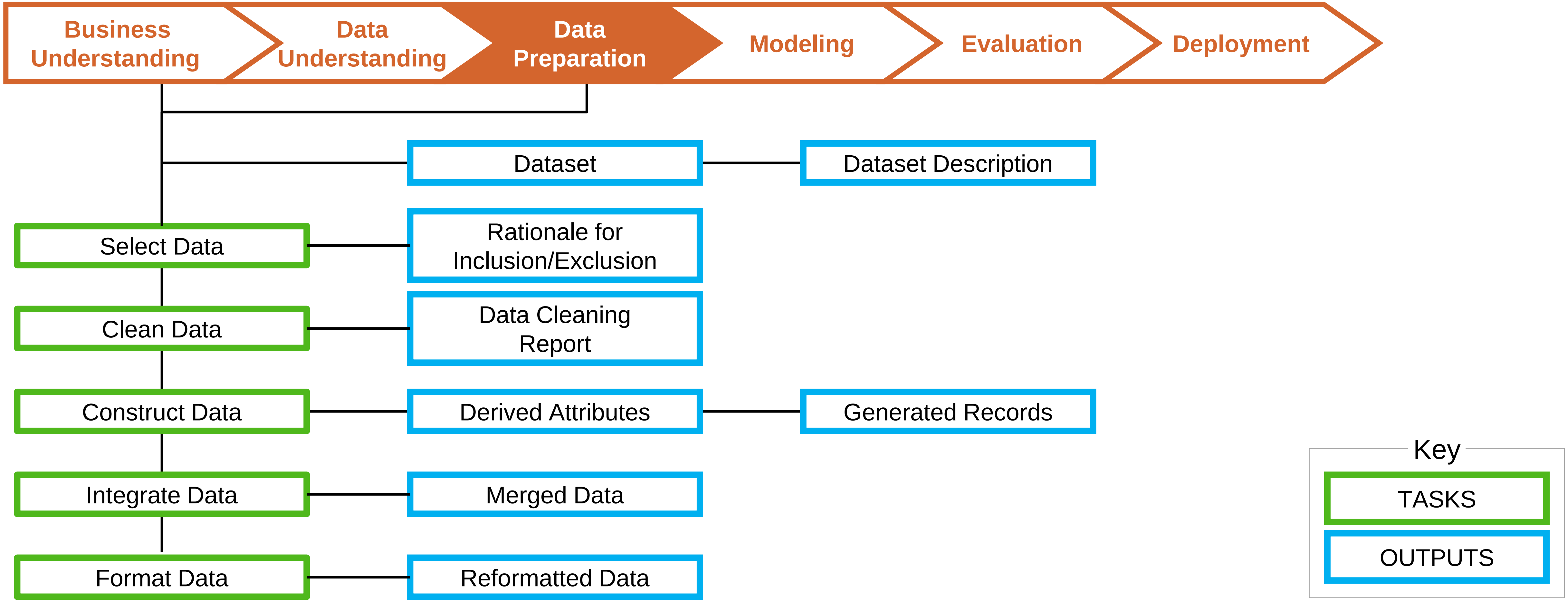
Introduction to Project Methodologies

CRISP-DM – Phase 2: Data Understanding



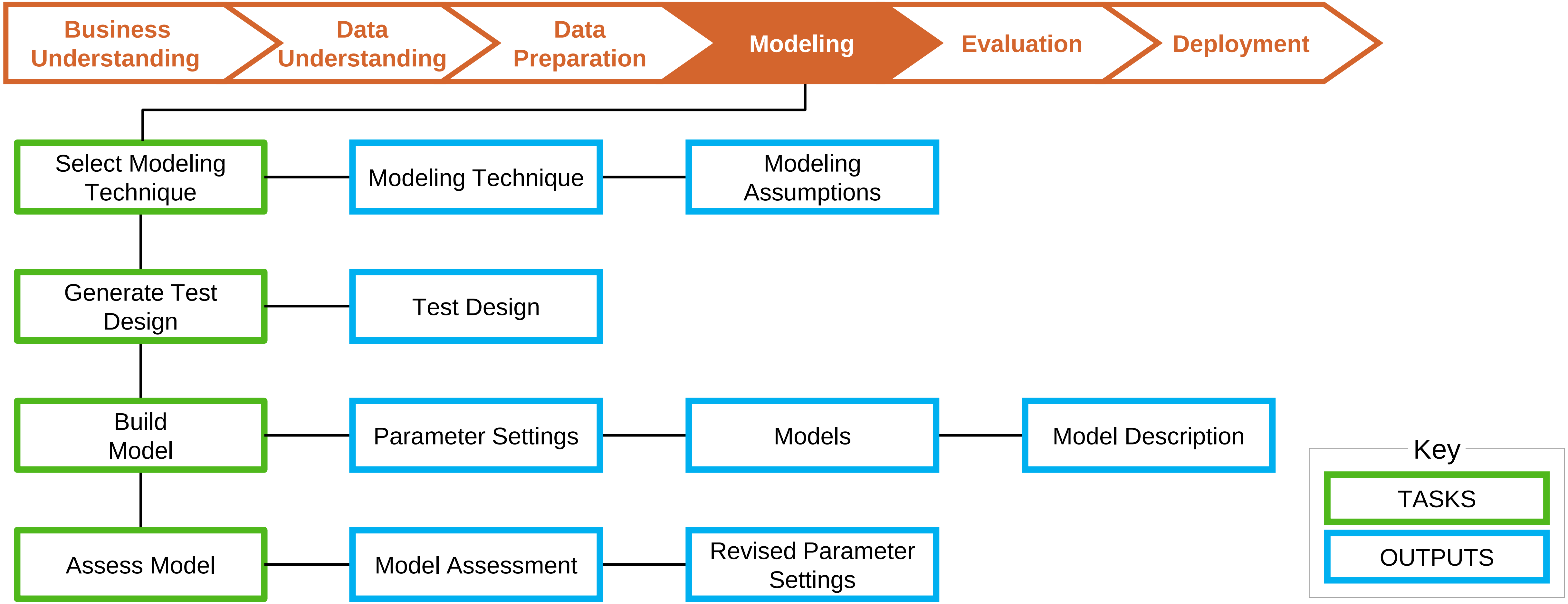
Introduction to Project Methodologies

CRISP-DM – Phase 3: Data Preparation



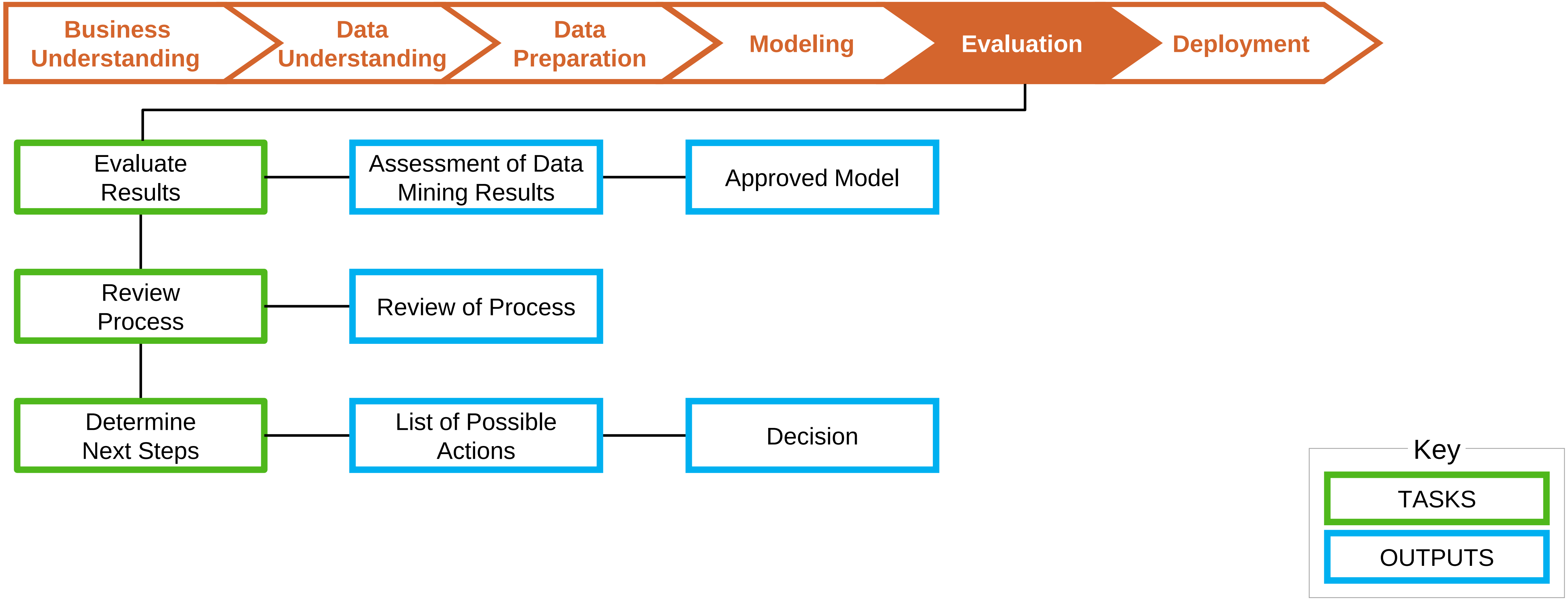
Introduction to Project Methodologies

CRISP-DM – Phase 4: Modeling



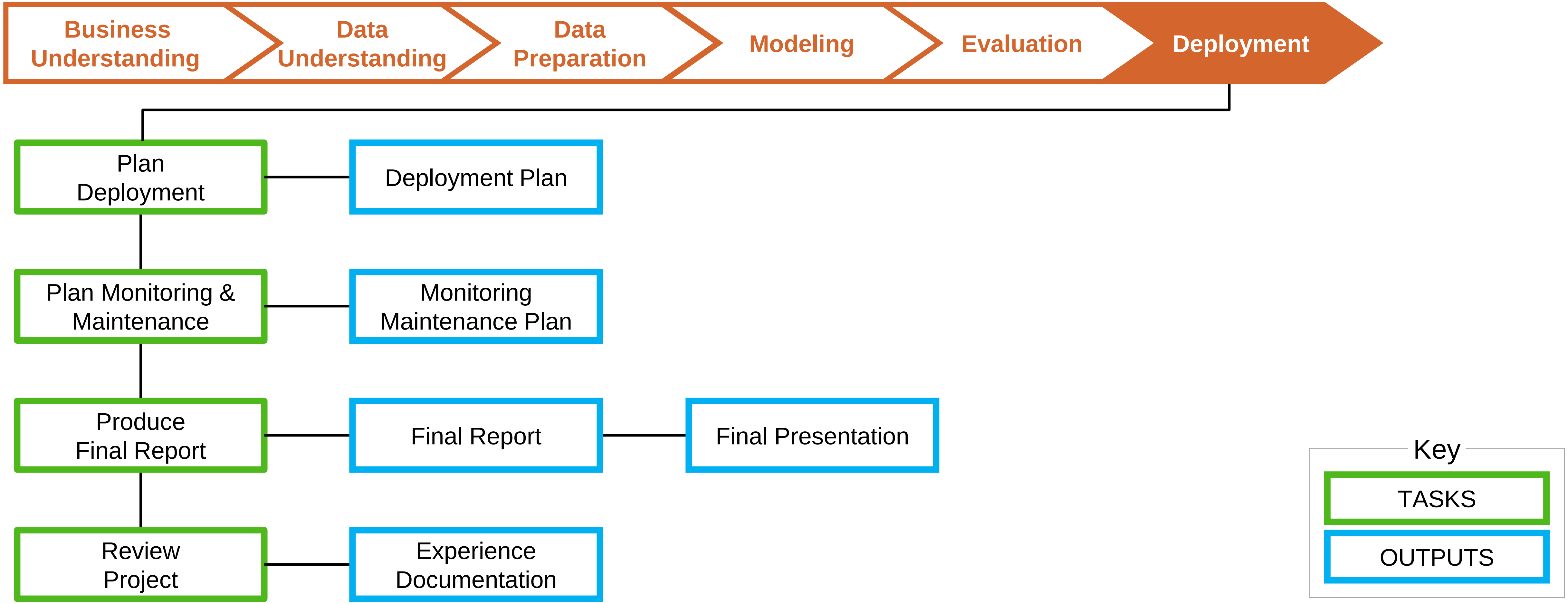
Introduction to Project Methodologies

CRISP-DM – Phase 5: Evaluation



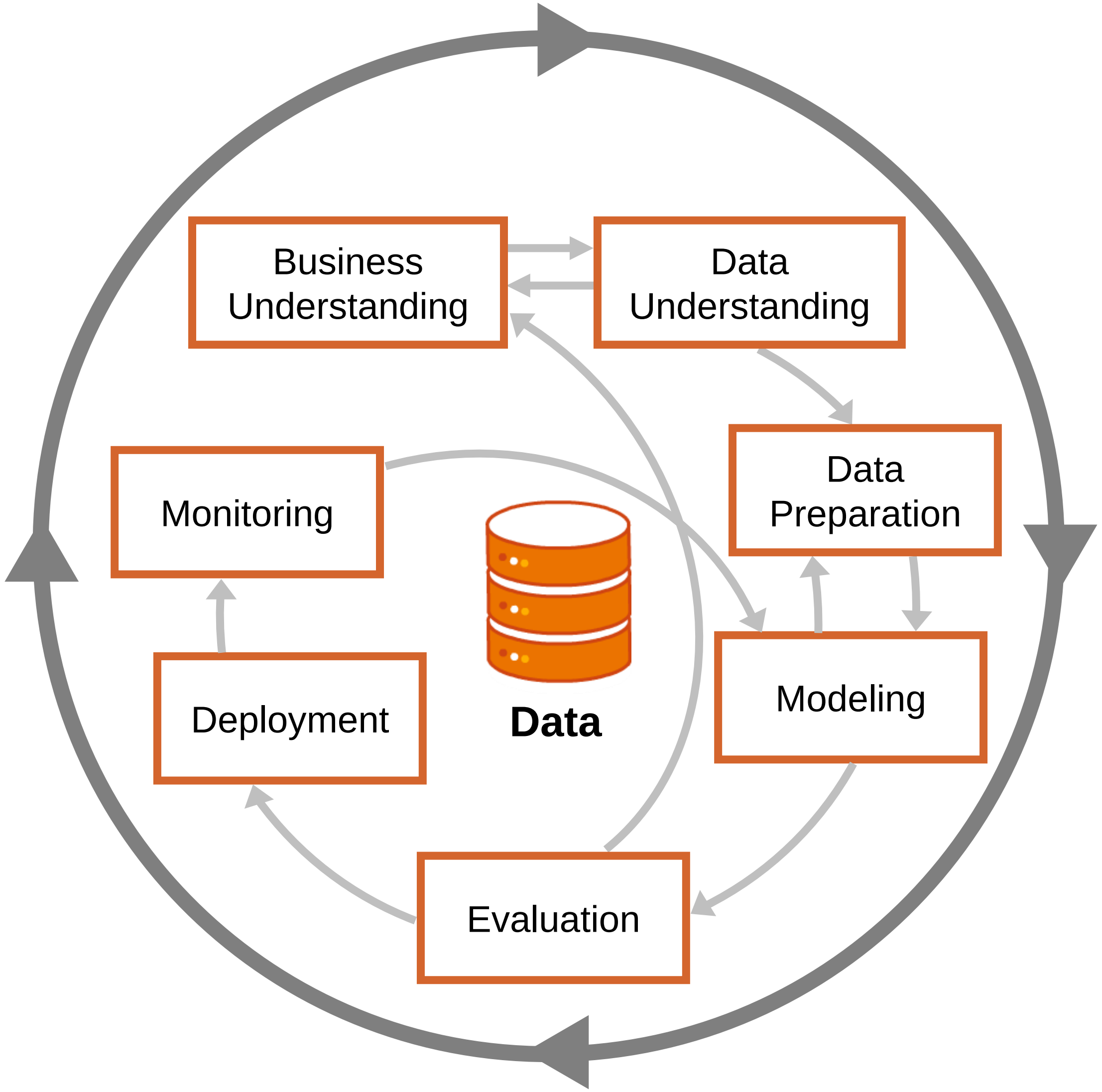
Introduction to Project Methodologies

CRISP-DM – Phase 6: Deployment



Introduction to Project Methodologies

CRISP-DM – Update



Thank you

Contact information:

open@sap.com

© 2016 SAP SE or an SAP affiliate company. All rights reserved.

Some software products marketed by SAP SE and its distributors contain proprietary software components of other software vendors.

National product specifications may vary.

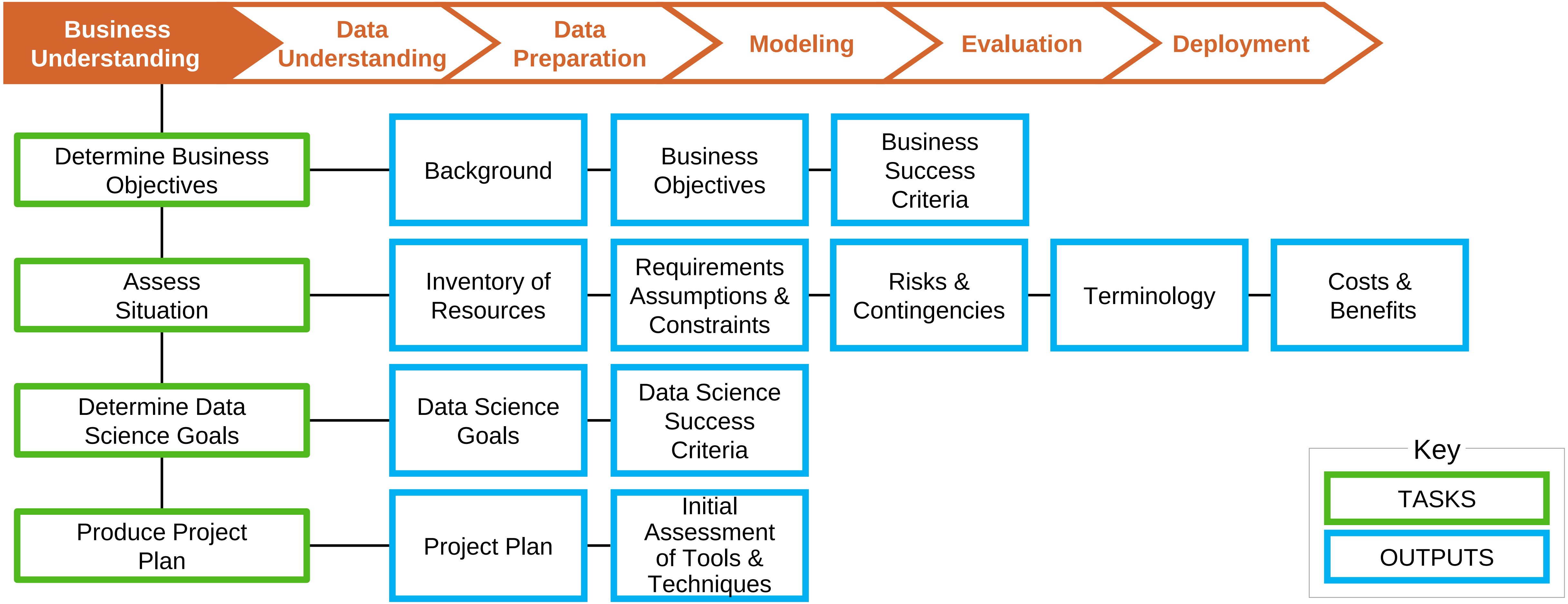
These materials are provided by SAP SE or an SAP affiliate company for informational purposes only, without representation or warranty of any kind, and SAP SE or its affiliated companies shall not be liable for errors or omissions with respect to the materials. The only warranties for SAP SE or SAP affiliate company products and services are those that are set forth in the express warranty statements accompanying such products and services, if any. Nothing herein should be construed as constituting an additional warranty.

In particular, SAP SE or its affiliated companies have no obligation to pursue any course of business outlined in this document or any related presentation, or to develop or release any functionality mentioned therein. This document, or any related presentation, and SAP SE's or its affiliated companies' strategy and possible future developments, products, and/or platform directions and functionality are all subject to change and may be changed by SAP SE or its affiliated companies at any time for any reason without notice. The information in this document is not a commitment, promise, or legal obligation to deliver any material, code, or functionality. All forward-looking statements are subject to various risks and uncertainties that could cause actual results to differ materially from expectations. Readers are cautioned not to place undue reliance on these forward-looking statements, which speak only as of their dates, and they should not be relied upon in making purchasing decisions.

Week 1 Unit 3: **Business Understanding Phase – Overview**

Business Understanding Phase – Overview

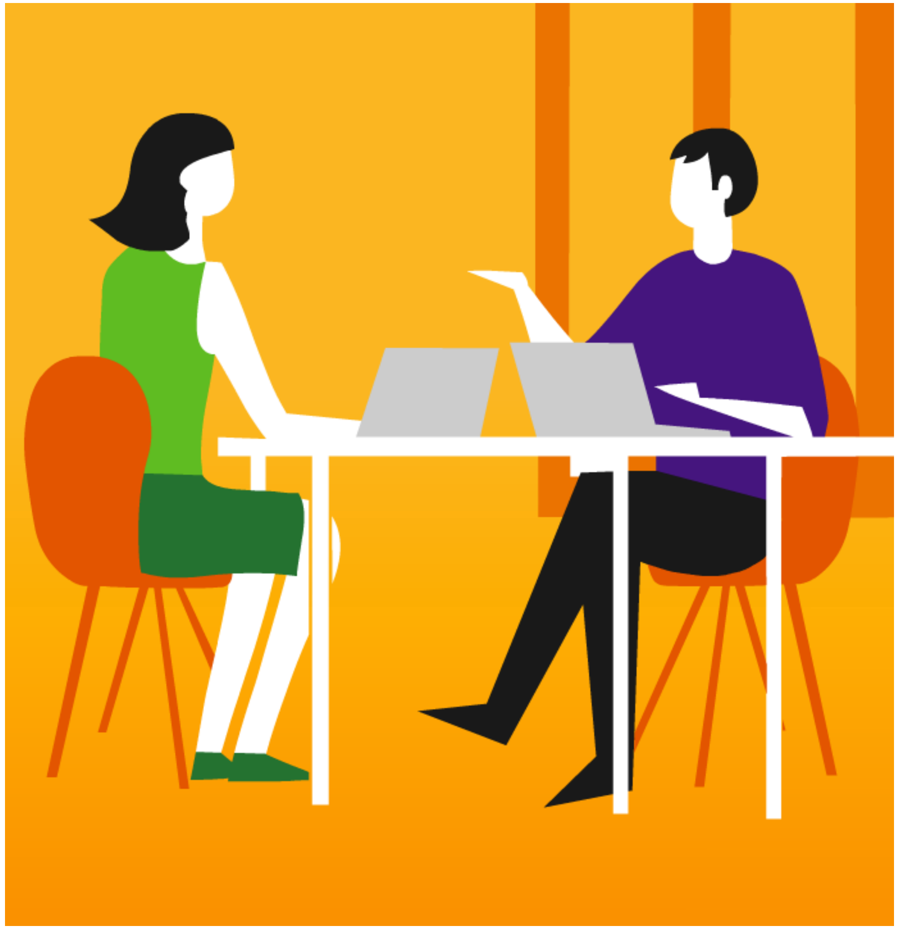
CRISP-DM – Phase 1: Business Understanding



Business Understanding Phase – Overview

Phase 1.1: Determine Business Objectives

- **Task**
  - The first objective of the data analyst is to thoroughly understand, from a business perspective, what the client really wants to accomplish.
- **Outputs**
  - Background
  - Business Objectives
  - Business Success Criteria



Business Understanding Phase – Overview

Phase 1.2: Assess Situation

- **Task**
  - In the previous task, your objective is to quickly get to the crux of the situation. Here, you want to flesh out the details.
- **Outputs**
  - Inventory of Resources
  - Requirements, Assumptions, & Constraints
  - Risks & Contingencies
  - Terminology
  - Costs & Benefits

Business Understanding Phase – Overview

Phase 1.3: Determine Data Science Goals

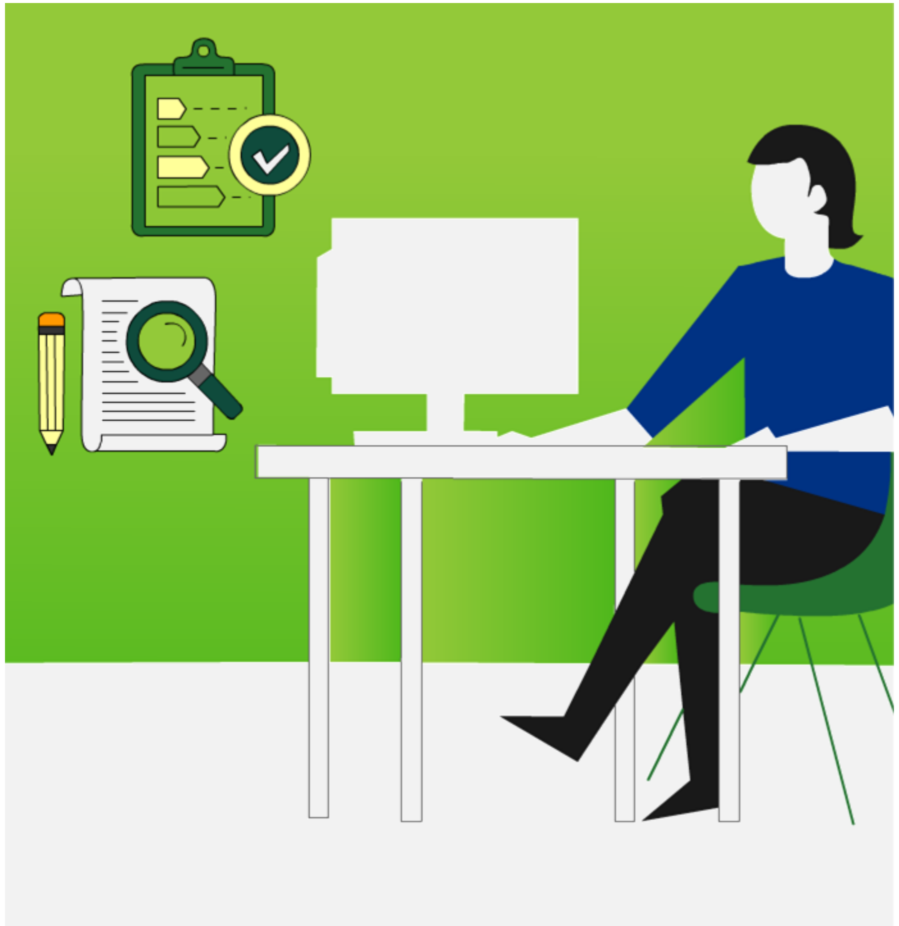
- **Task**
  - A ***business goal*** states objectives in business terminology.
  - A ***data science goal*** states project objectives in technical terms.

- **Outputs**
  - Describe data science goals.
  - Define data science success criteria.

Business Understanding Phase – Overview

Phase 1.4: Produce Project Plan

- **Task**
  - Describe the intended plan for achieving the data mining goals and thereby achieving the business goals.
- **Output**
  - Project plan with project stages, duration, resources, etc.
  - Initial assessment of tools & techniques.



Thank you

Contact information:

open@sap.com

© 2016 SAP SE or an SAP affiliate company. All rights reserved.

Some software products marketed by SAP SE and its distributors contain proprietary software components of other software vendors.

National product specifications may vary.

These materials are provided by SAP SE or an SAP affiliate company for informational purposes only, without representation or warranty of any kind, and SAP SE or its affiliated companies shall not be liable for errors or omissions with respect to the materials. The only warranties for SAP SE or SAP affiliate company products and services are those that are set forth in the express warranty statements accompanying such products and services, if any. Nothing herein should be construed as constituting an additional warranty.

In particular, SAP SE or its affiliated companies have no obligation to pursue any course of business outlined in this document or any related presentation, or to develop or release any functionality mentioned therein. This document, or any related presentation, and SAP SE’s or its affiliated companies’ strategy and possible future developments, products, and/or platform directions and functionality are all subject to change and may be changed by SAP SE or its affiliated companies at any time for any reason without notice. The information in this document is not a commitment, promise, or legal obligation to deliver any material, code, or functionality. All forward-looking statements are subject to various risks and uncertainties that could cause actual results to differ materially from expectations. Readers are cautioned not to place undue reliance on these forward-looking statements, which speak only as of their dates, and they should not be relied upon in making purchasing decisions.

Week 1 Unit 4: **Defining Project Success Criteria**

Defining Project Success Criteria

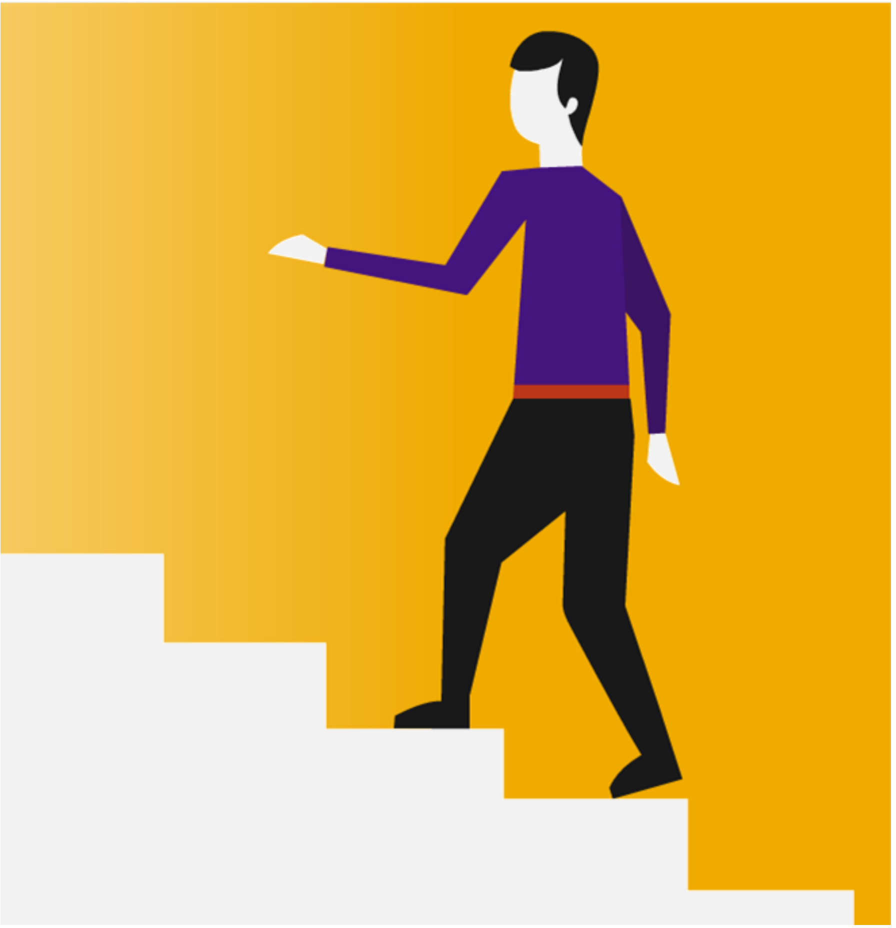
Business and data science project success criteria: reminder

Business success criteria

- Describe the criteria for a successful or useful outcome to the project from the business point of view.

Data science success criteria

- Define the criteria for a successful outcome to the project in technical terms.



Defining Project Success Criteria

Recent industry surveys

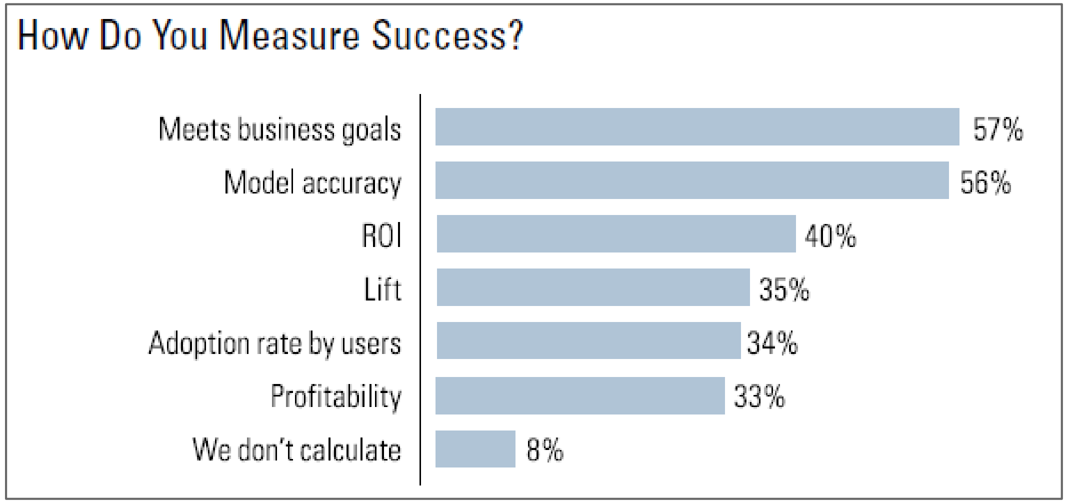


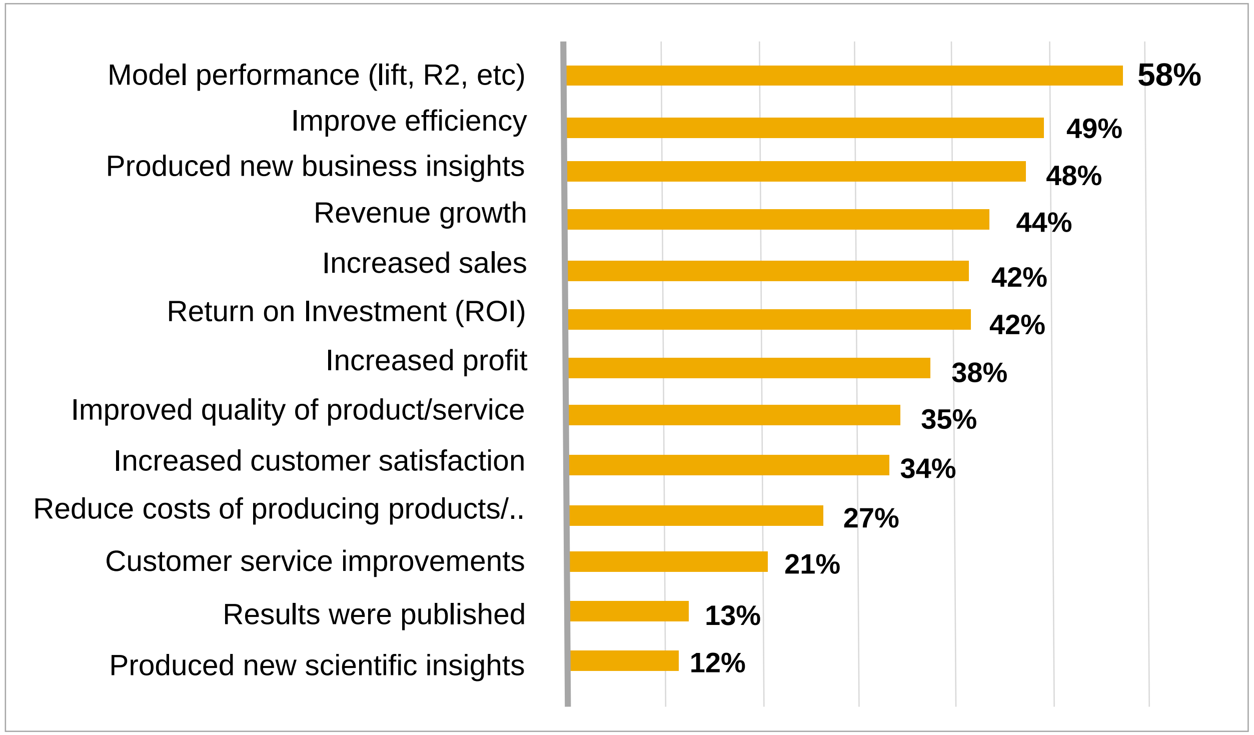
Figure 5. Based on 110 users who have implemented predictive analytics initiatives that offer "very high" or "high" value. Respondents could select multiple choices.

PREDICTIVE ANALYTICS

Extending the Value of Your Data Warehousing Investment

By Wayne W. Eckerson

In their Third Annual Data Miner Survey, Rexer Analytics, an analytics and renowned CRM consulting firm based in Winchester, Massachusetts asked the BI community "How do you evaluate project success in data mining?" Out of 14 different criteria, a massive 58% ranked "Model performance" (lift, R2, etc) as the primary factor.



Defining Project Success Criteria

Model success criteria: descriptive or predictive models

Descriptive Models

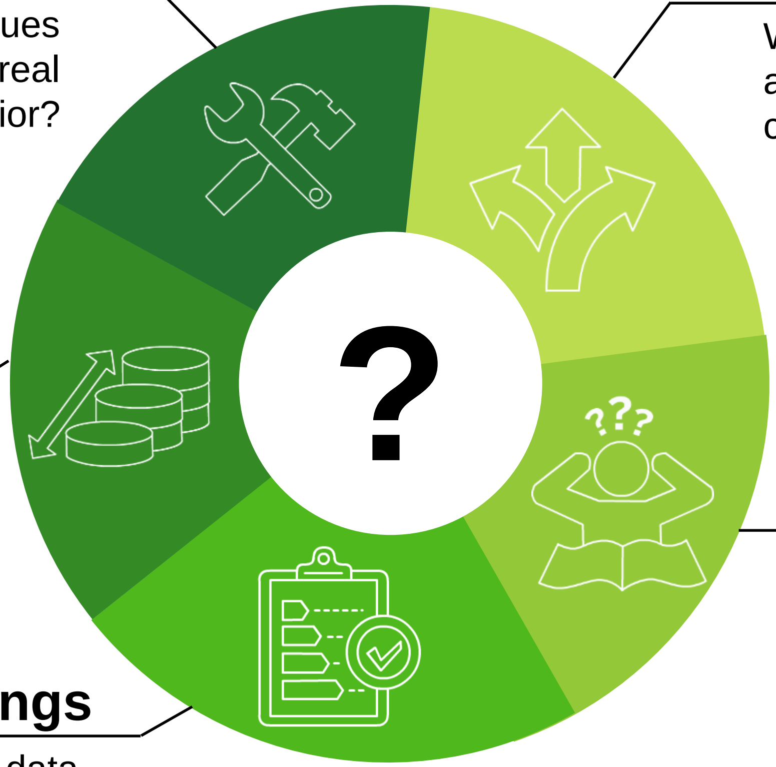
- Descriptive analysis describes or summarizes raw data and makes it more interpretable. It describes the past – i.e. any point of time that an event occurred, whether it was one minute ago or one year ago.
- Descriptive analytics are useful because they allow us to learn from past behaviors and understand how these might influence future outcomes.
- Common examples of descriptive analytics are reports that provide historical insights regarding a company's production, financials, operations, sales, finance, inventory and customers.
- Descriptive analytical models include cluster models, association rules, and network analysis.

Predictive Models

- Predictive analysis predicts what might happen in the future – providing estimates about the likelihood of a future outcome.
- One common application is the use of predictive analytics to produce a credit score. These scores are used by financial services to determine the probability of customers making future credit payments on time.
- Typical business uses include: understanding how sales might close at the end of the year, predicting what items customers will purchase together, or forecasting inventory levels based upon a myriad of variables.
- Predictive analytical models include classification models, regression models, and neural network models.

Defining Project Success Criteria

Model success criteria: choosing the algorithm



Anomalies

What anomalies or unusual values might exist? Are they errors or real changes in behavior?

Trends

What are the trends, both historical and emerging, and how might they continue?

Associations

What are the correlations in the data? What are the cross-sell opportunities?

Relationships

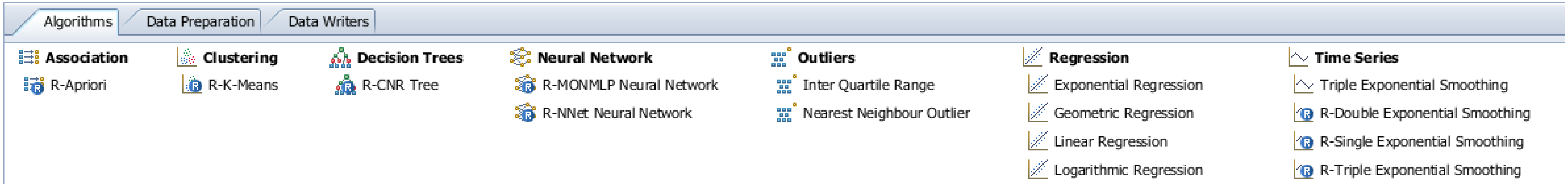
What are the main influencers, for example customer churn, employee turnover etc.?

Groupings

Are there any clear groupings of the data, for example customer segments for specific marketing campaigns?

Defining Project Success Criteria

There is a wide range of algorithms to choose from…



Algorithms | Data Preparation | Data Writers

Association
- R-Apriori

Clustering
- R-K-Means

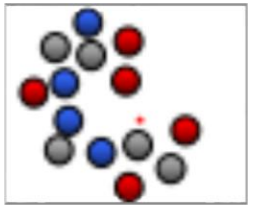
Decision Trees
- R-CNR Tree

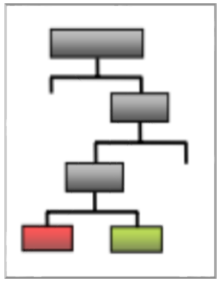
Neural Network
- R-MONMLP Neural Network
- R-NNet Neural Network

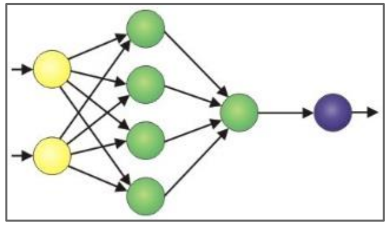
Outliers
- Inter Quartile Range
- Nearest Neighbour Outlier

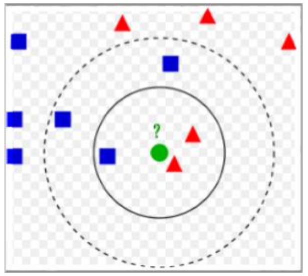
Regression
- Exponential Regression
- Geometric Regression
- Linear Regression
- Logarithmic Regression

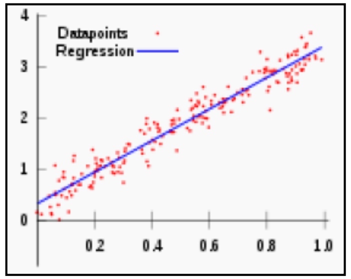
Time Series
- Triple Exponential Smoothing
- R-Double Exponential Smoothing
- R-Single Exponential Smoothing
- R-Triple Exponential Smoothing

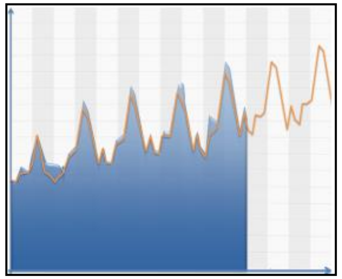












Association

Clustering

Classification – bivariate target variable

Detect anomalies or outliers (data cleansing or decision support)

Regression – continuous target variable

Forecasting with time series data

Descriptive Models

Predictive Models

Defining Project Success Criteria

Which business question do you need to answer?

Classification

Who will (buy | fraud | churn …) next (week | month | year…)?

Regression

What will the (revenue | # churners) be next (week | month…)?

Segmentation or Clustering

What are the groups of customers with similar (behavior | profile …)?

Forecasting (Time Series Analysis)

What will the (revenue | # churners…) be over next year on a monthly basis?

Link Analysis

Analyze interactions to identify (communities | influencers…)

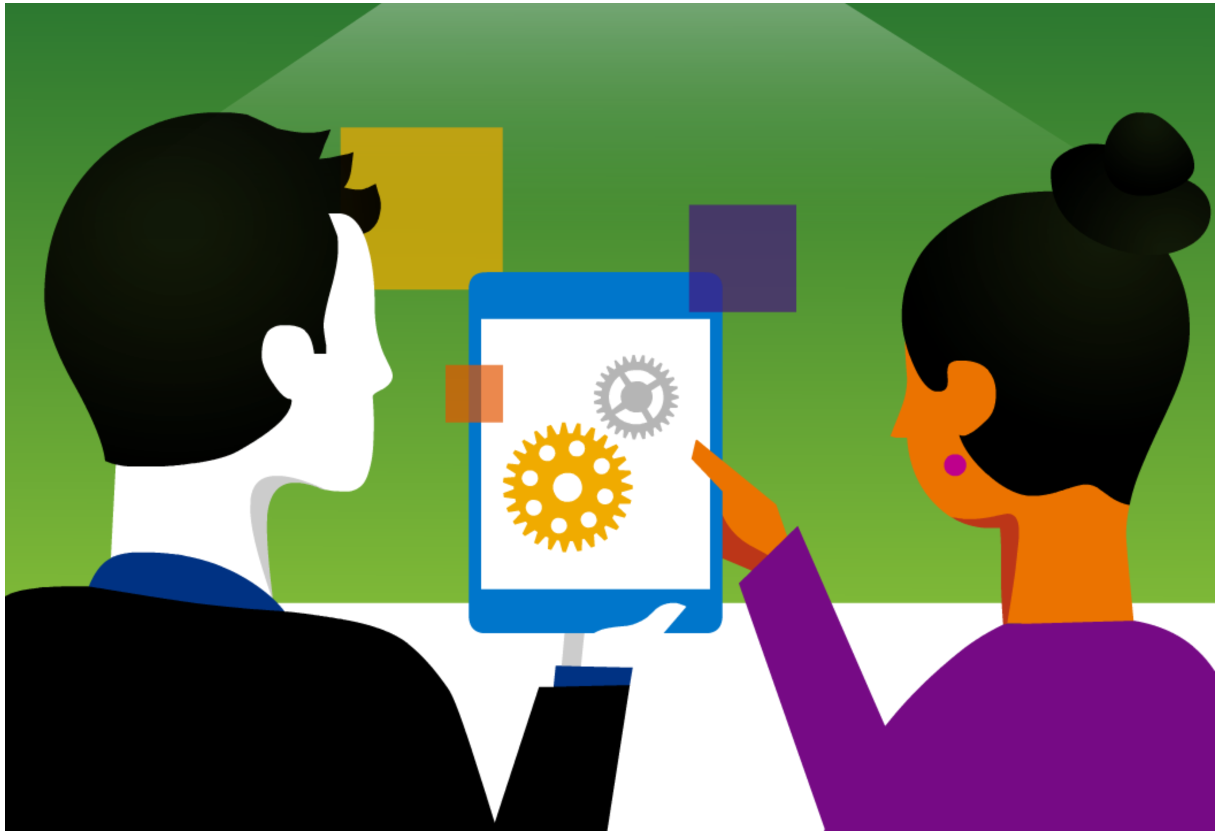
Association or Recommendation Engines

Provides recommendations on web sites or to retailers – basket analysis

Defining Project Success Criteria

Model accuracy and robustness

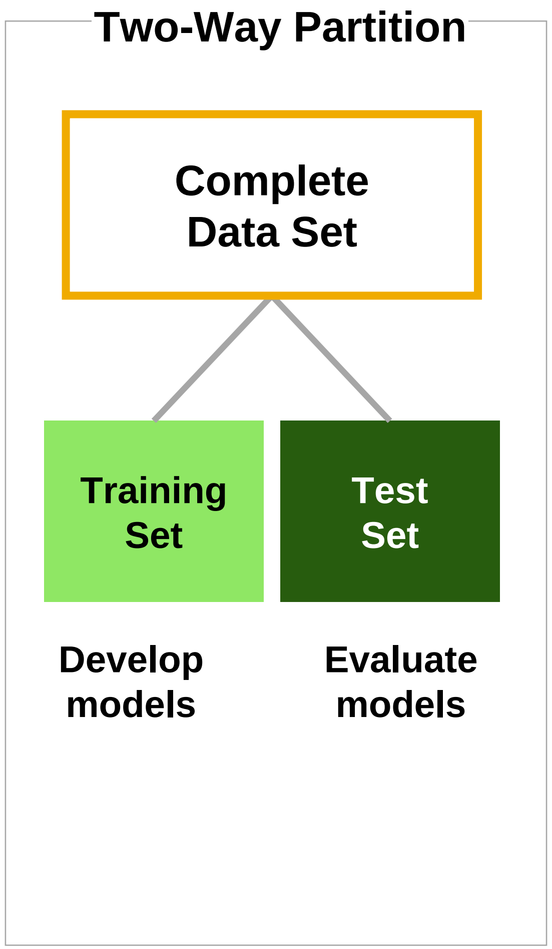
- The accuracy and robustness of the model are two major factors to determine the quality of the prediction, which reflects how successful the model is.
- **Accuracy** is often the starting point for analyzing the quality of a predictive model, as well as an obvious criterion for prediction. Accuracy measures the ratio of correct predictions to the total number of cases evaluated.
- The **robustness** of a predictive model refers to how well a model works on alternative data. This might be hold-out data or new data that the model is to be applied onto.

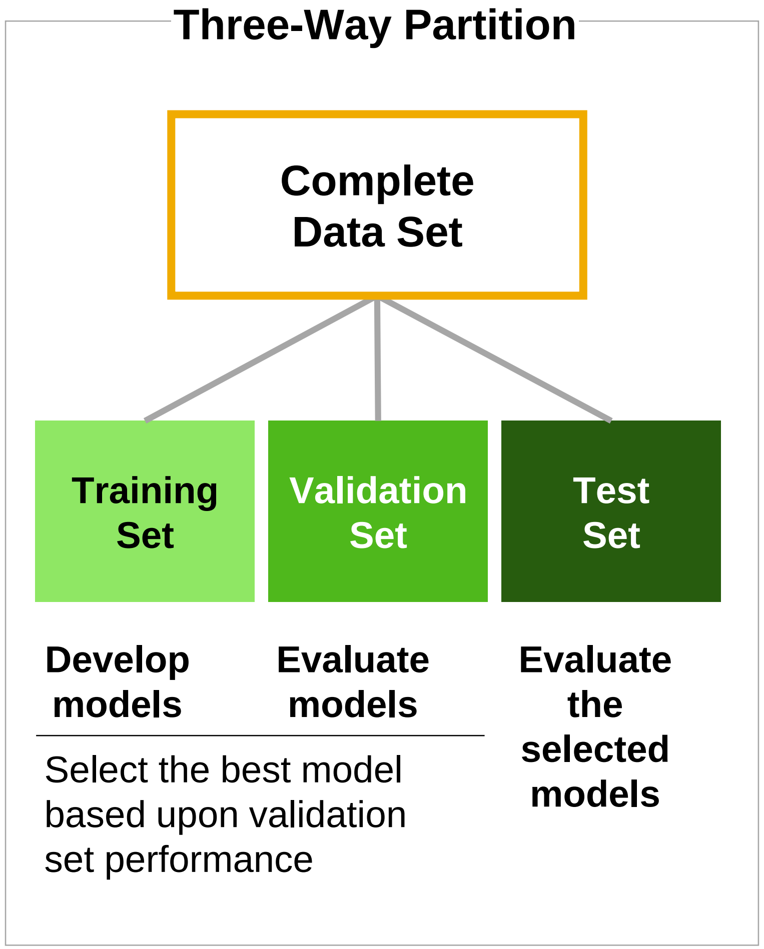


Defining Project Success Criteria

Training and testing: data cutting strategies

- Central to developing predictive models and assessing if they are successful is a **train-and-test regime**.
- Data is partitioned into **training** and **test** subsets. There are a variety of **cutting strategies** (e.g. random/sequential/periodic).
- We build our model on the training subset (called the **estimation** subset) and evaluate its performance on the test subset (a **hold-out** sample called the **validation** subset).
- Simple two and three-way data partitioning is shown in the diagram.





Defining Project Success Criteria

Example success criteria for predictive models

Business Criteria:

- Models meet business goals – the model meets the business objectives specified in CRISP-DM phase 1.1 as defined by the customer
- The model's contributing variables and the variable categories make "business sense"

Model Performance Criteria:

- Depends on the algorithm. For a classification model for example:
  - – Model accuracy compared to any previous, similar models. There are a variety of accuracy measures that will be discussed in this course
  - – Model robustness – Models have acceptable robustness

Software Usability Criteria:

- Speed and ease of model development – so the customer can build new models and update existing models quickly
- Speed and ease of model deployment – so the customer can create Apply datasets easily and deploy models quickly with the required outputs (probabilities, deciles, etc.) – speed to market
- Ease of model maintenance – so the customer can easily define when models require refreshing/rebuilding and undertake this quickly and easily
- Integration capability with other systems

Defining Project Success Criteria

Example success criteria for descriptive models

Business Criteria:

- Models meet business goals – the model meets the business objectives specified in CRISP-DM phase 1.1
- Contributing variables and categories make business sense

Model Performance Criteria:

- Depends on the algorithm. For a cluster model for example:
  - Determining the clustering tendency of a set of data (distinguishing whether non-random structure actually exists in the data)
  - Comparing the results to given class labels (comparing model results to existing cluster groups)
  - Evaluating how well the results of the analysis fit the data without reference to external information
  - Comparing the results of different cluster models to determine which is better
  - Determining the best number of clusters

Software Usability Criteria:

- Speed and ease of model development – so the customer can build new models and update existing models quickly
- Speed and ease of model deployment – so the customer can create Apply datasets easily and deploy models quickly with the required outputs (probabilities, deciles, etc.) – speed to market
- Ease of model maintenance – so the customer can easily define when models require refreshing/rebuilding and undertake this quickly and easily
- Integration capability with other systems

Thank you

Contact information:

open@sap.com

© 2016 SAP SE or an SAP affiliate company. All rights reserved.

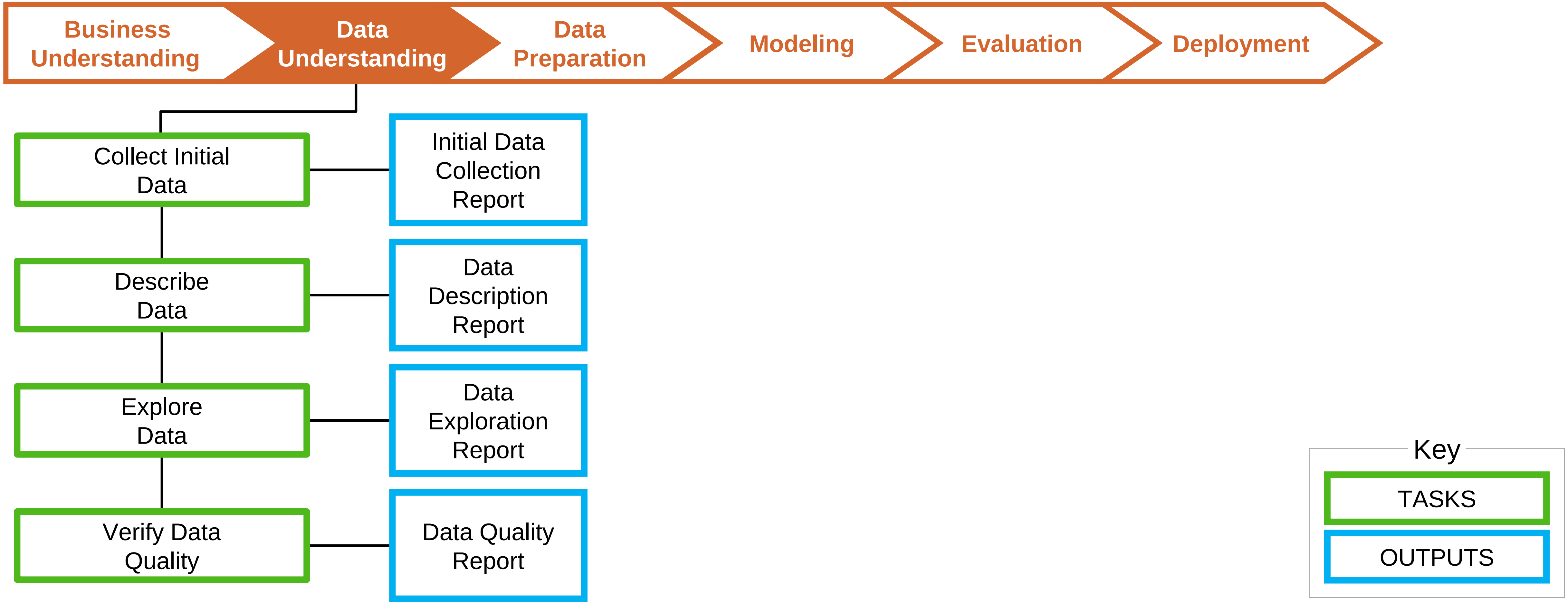
Week 1 Unit 5: **Data Understanding Phase – Overview**

openSAP

open.sap.com

Data Understanding Phase – Overview

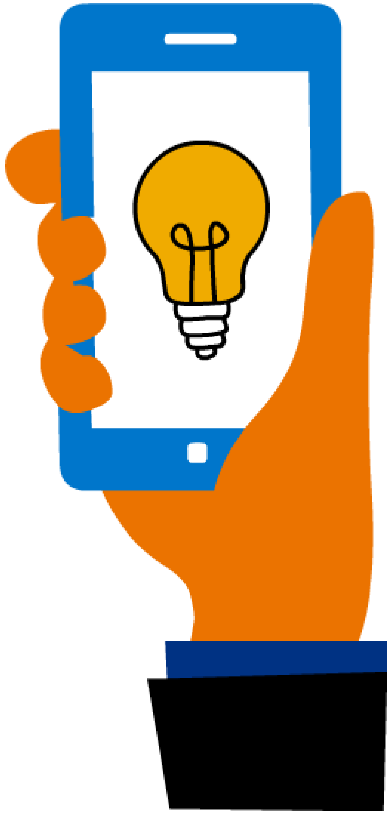
CRISP-DM – Phase 2: Data Understanding



Data Understanding Phase – Overview

Phase 2.1: Collect Initial Data

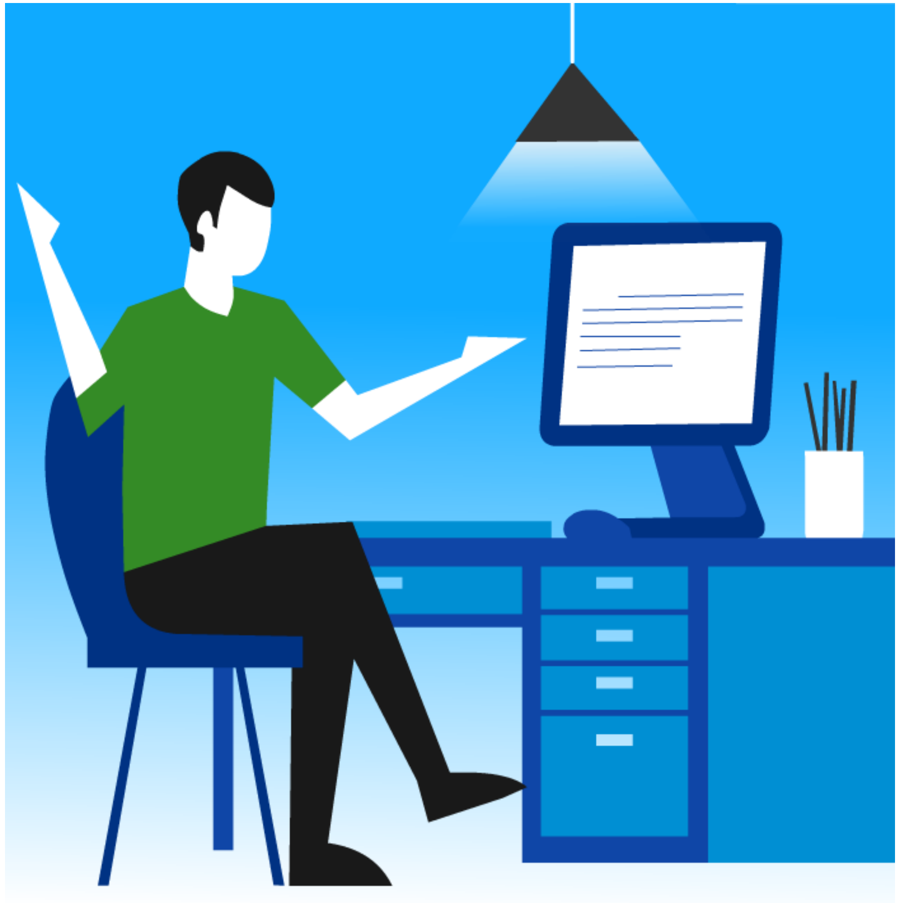
- **Task**
  - Acquire the data (or access to the data) listed in the project resources.
  - This initial collection includes data loading into the data exploration tool and data integration if multiple data sources are acquired.
- **Output – Initial Data Collection Report**
  - List the following:
    - The dataset (or datasets) acquired
    - The dataset locations
    - The methods used to acquire the datasets
    - Any problems encountered
  - Record problems encountered and any solutions.



Data Understanding Phase – Overview

Phase 2.2: Describe Data

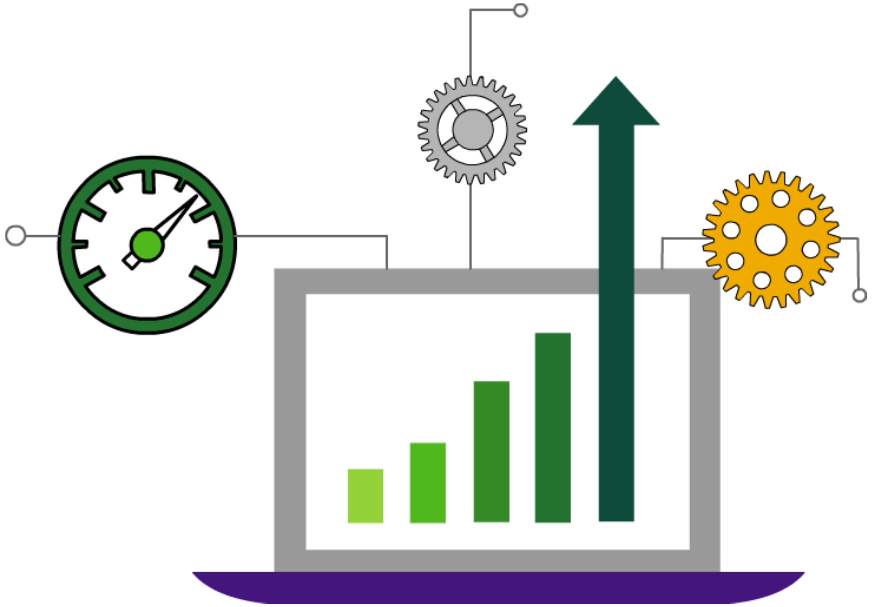
- **Task**
  - – Examine the “gross” or “surface” properties of the acquired data and report on the results.
- **Output – Data Description Report**
  - – Describe the data that has been acquired, including:
    - The format of the data.
    - The quantity of data, e.g. the number of records and fields in each table.
    - The identities of the fields.
    - Any other surface features of the data that have been discovered.



Data Understanding Phase – Overview

Phase 2.3: Explore Data

- **Task**
  - This task tackles the data mining questions, which can be addressed using querying, visualization, and reporting.
- **Output – Data Exploration Report**
  - Describe results of this task including:
    - First findings or initial hypothesis and their impact on the remainder of the project.
    - If appropriate, include graphs and plots.



Data Understanding Phase – Overview

Phase 2.4: Verify Data Quality

- **Task**
  - Examine the quality of the data, addressing questions such as:
    - Is the data complete?
    - Is it correct or does it contain errors?
    - Are there missing values in the data?
- **Output – Data Quality Report**
  - List the results of the data quality verification.
  - If quality problems exist, list possible solutions.

Thank you

Contact information:

open@sap.com

© 2016 SAP SE or an SAP affiliate company. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or for any purpose without the express permission of SAP SE or an SAP affiliate company.

SAP and other SAP products and services mentioned herein as well as their respective logos are trademarks or registered trademarks of SAP SE (or an SAP affiliate company) in Germany and other countries. Please see http://global12.sap.com/corporate-en/legal/copyright/index.epx for additional trademark information and notices.

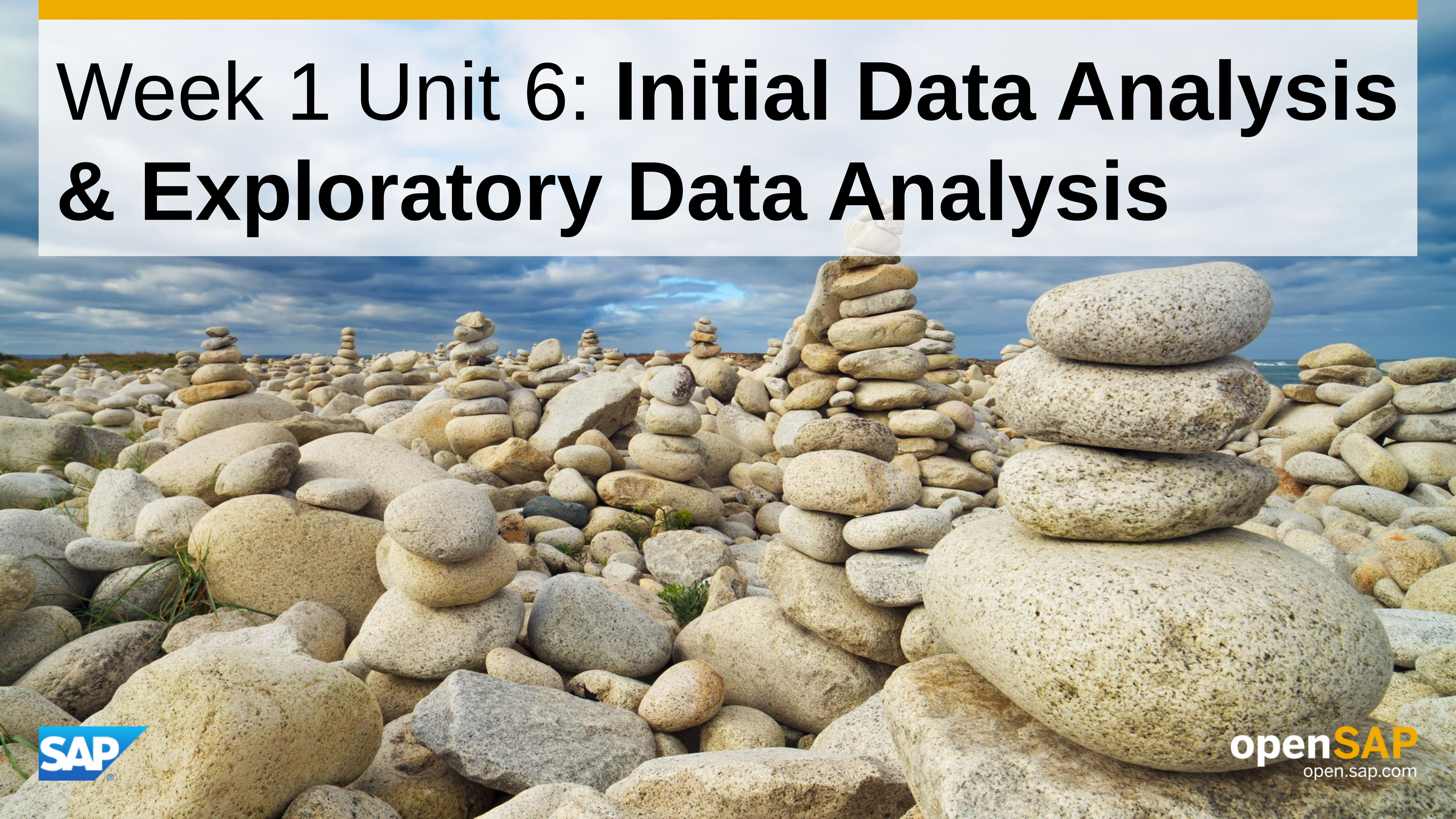
Some software products marketed by SAP SE and its distributors contain proprietary software components of other software vendors.

National product specifications may vary.

These materials are provided by SAP SE or an SAP affiliate company for informational purposes only, without representation or warranty of any kind, and SAP SE or its affiliated companies shall not be liable for errors or omissions with respect to the materials. The only warranties for SAP SE or SAP affiliate company products and services are those that are set forth in the express warranty statements accompanying such products and services, if any. Nothing herein should be construed as constituting an additional warranty.

In particular, SAP SE or its affiliated companies have no obligation to pursue any course of business outlined in this document or any related presentation, or to develop or release any functionality mentioned therein. This document, or any related presentation, and SAP SE's or its affiliated companies' strategy and possible future developments, products, and/or platform directions and functionality are all subject to change and may be changed by SAP SE or its affiliated companies at any time for any reason without notice. The information in this document is not a commitment, promise, or legal obligation to deliver any material, code, or functionality. All forward-looking statements are subject to various risks and uncertainties that could cause actual results to differ materially from expectations. Readers are cautioned not to place undue reliance on these forward-looking statements, which speak only as of their dates, and they should not be relied upon in making purchasing decisions.

Week 1 Unit 6: **Initial Data Analysis & Exploratory Data Analysis**



open.sap.com

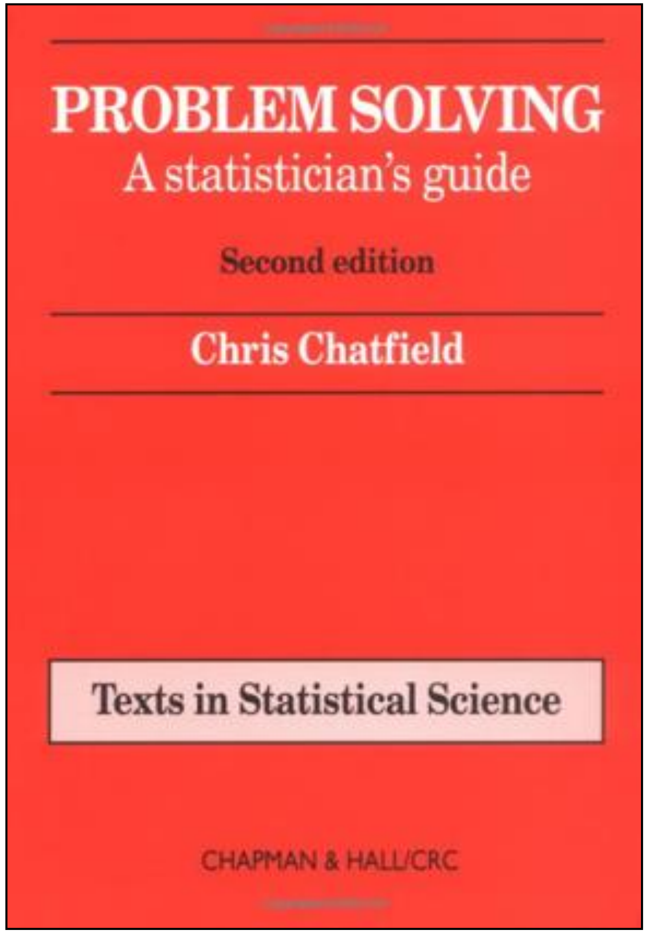
Initial Data Analysis & Exploratory Data Analysis

Initial data analysis

- *“Initial data analysis (IDA) is an essential part of nearly every analysis”*

  Problem Solving, A Statisticians Guide
  Christopher Chatfield

- Chatfield defines the various steps in IDA. It includes analysis of:
  - The structure of the data
  - The quality of the data
    - errors, outliers, and missing observations
  - Descriptive statistics
  - Graphs
- The data are modified according to the analysis:
  - Adjust extreme observations, estimate missing observations, transform variables, bin data, form new variables.



PROBLEM SOLVING
A statistician’s guide

Second edition

Chris Chatfield

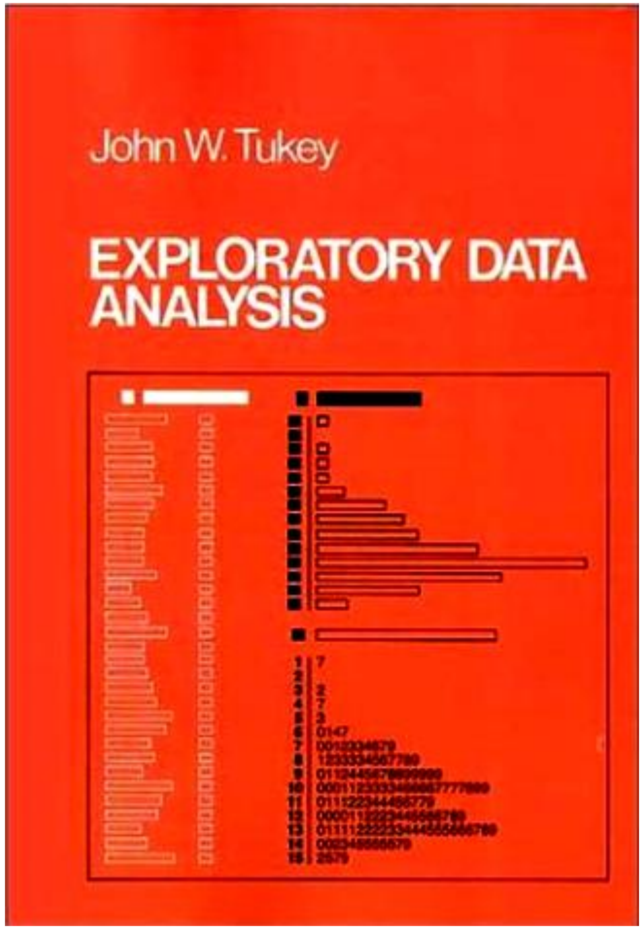
Texts in Statistical Science

CHAPMAN & HALL/CRC

Initial Data Analysis & Exploratory Data Analysis

Exploratory data analysis

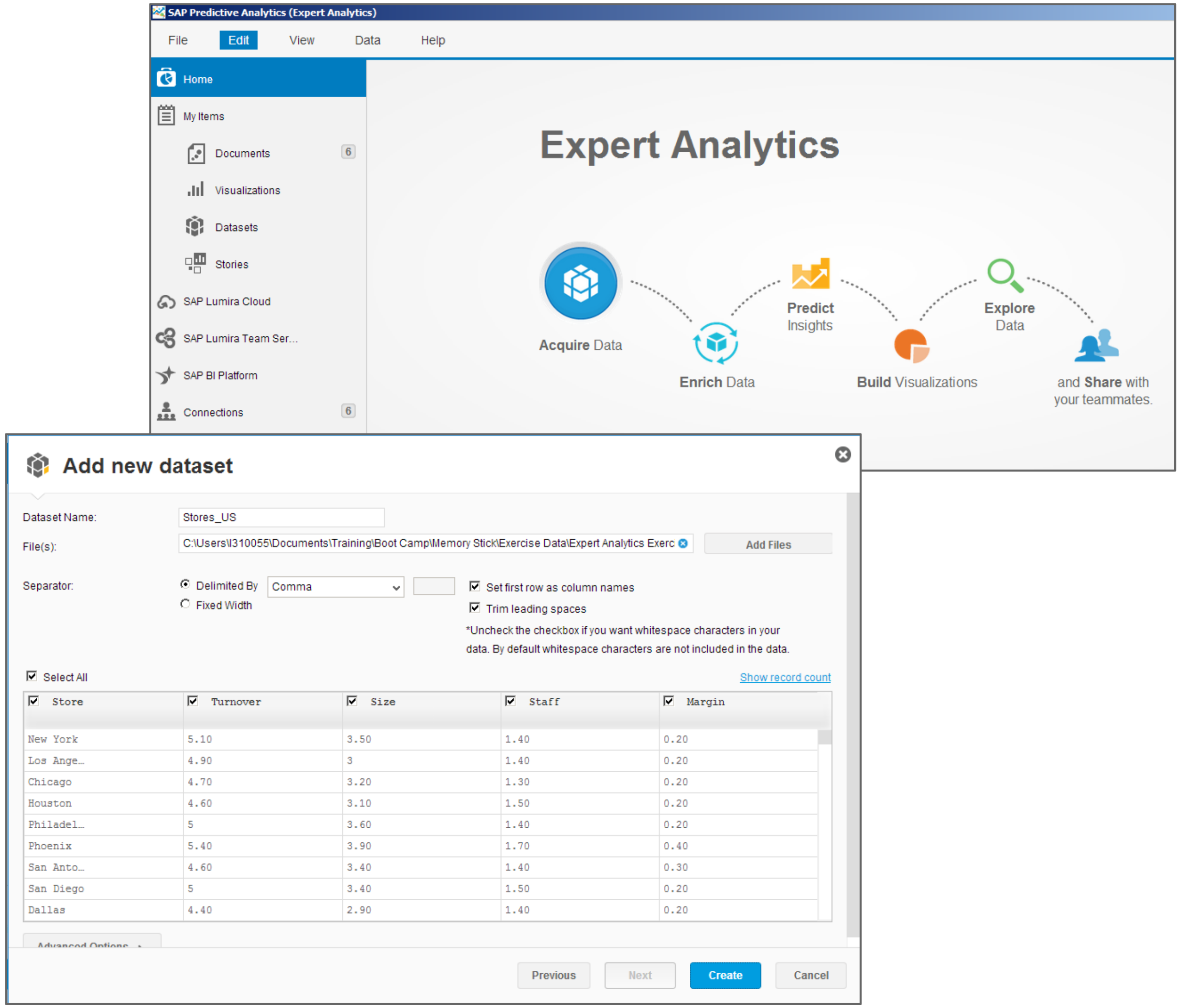
- Exploratory data analysis (EDA) is an approach to analyzing data for the purpose of formulating hypotheses that are worth testing, and complements the tools of conventional statistics for testing hypotheses.
- It was so named by John Tukey.
  Wikipedia
- *“It is important to understand what you CAN DO before you learn to measure how WELL you seem to have done it.”*
- *“To learn about data analysis, it is right that each of us try many things that do not work – that we tackle more problems than we make expert analyses of. We often learn less from an expertly done analysis than from one where, by not trying something, we missed an opportunity to learn more.”*
  John Tukey, Exploratory Data Analysis



Initial Data Analysis & Exploratory Data Analysis

Example – US Stores data demonstration

- Demonstration using the
  SAP Predictive Analytics expert system.
- We will use US Stores retail data to walk you through this topic.
- The dataset contains the following variables:
  - Store location
  - Turnover
  - Margin
  - Staff
  - Store size



Dataset Name: Stores_US

File(s): C:\Users\I310055\Documents\Training\Boot Camp\Memory Stick\Exercise Data\Expert Analytics Exerc

Separator: ● Delimited By Comma ○ Fixed Width

☑ Set first row as column names

☑ Trim leading spaces

*Uncheck the checkbox if you want whitespace characters in your data. By default whitespace characters are not included in the data.

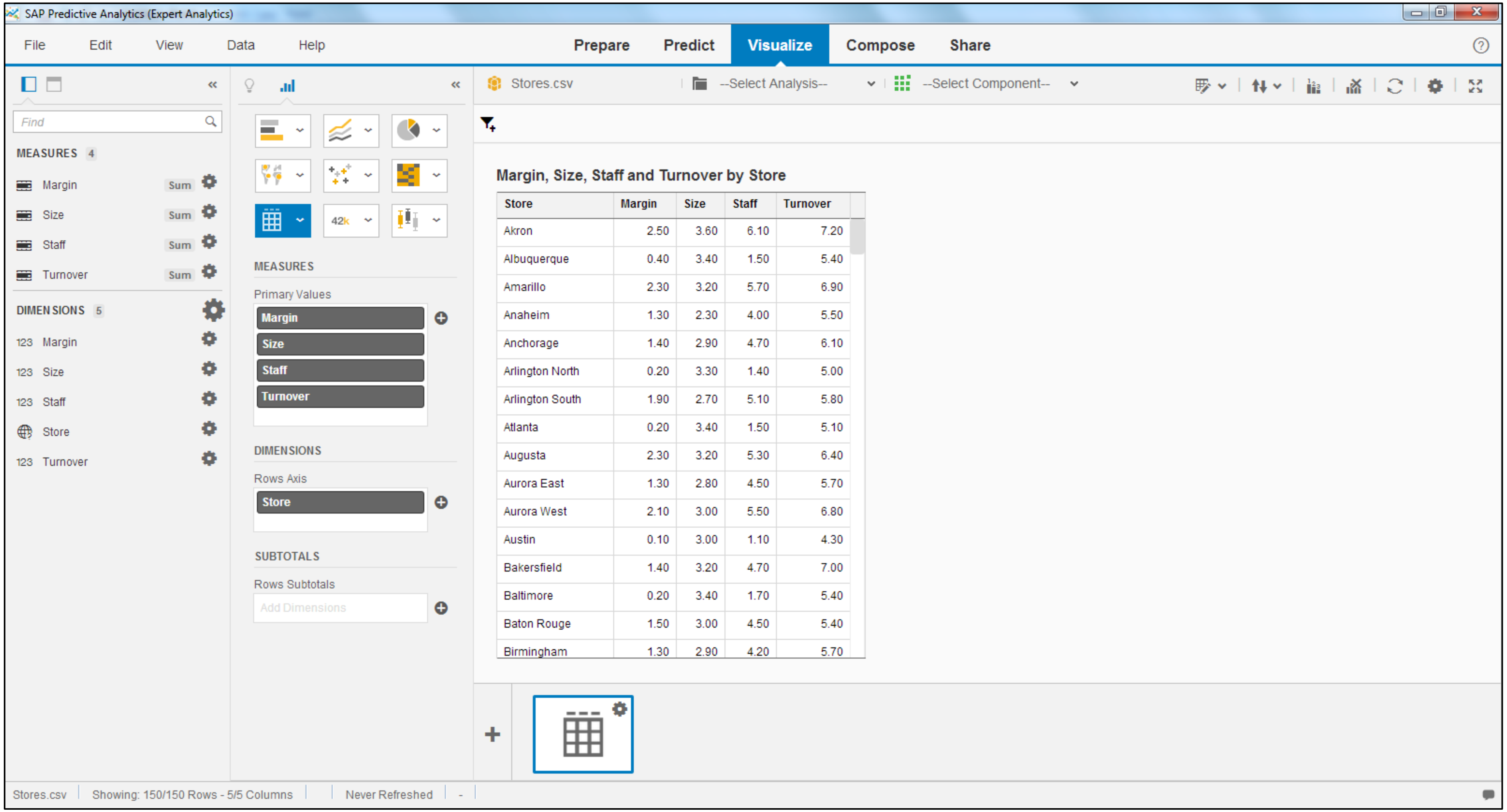
☑ Select All

Show record count

☑ Store	☑ Turnover	☑ Size	☑ Staff	☑ Margin
New York	5.10	3.50	1.40	0.20
Los Ange…	4.90	3	1.40	0.20
Chicago	4.70	3.20	1.30	0.20
Houston	4.60	3.10	1.50	0.20
Philadel…	5	3.60	1.40	0.20
Phoenix	5.40	3.90	1.70	0.40
San Anto…	4.60	3.40	1.40	0.30
San Diego	5	3.40	1.50	0.20
Dallas	4.40	2.90	1.40	0.20

Initial Data Analysis & Exploratory Data Analysis

Example – Data visualization

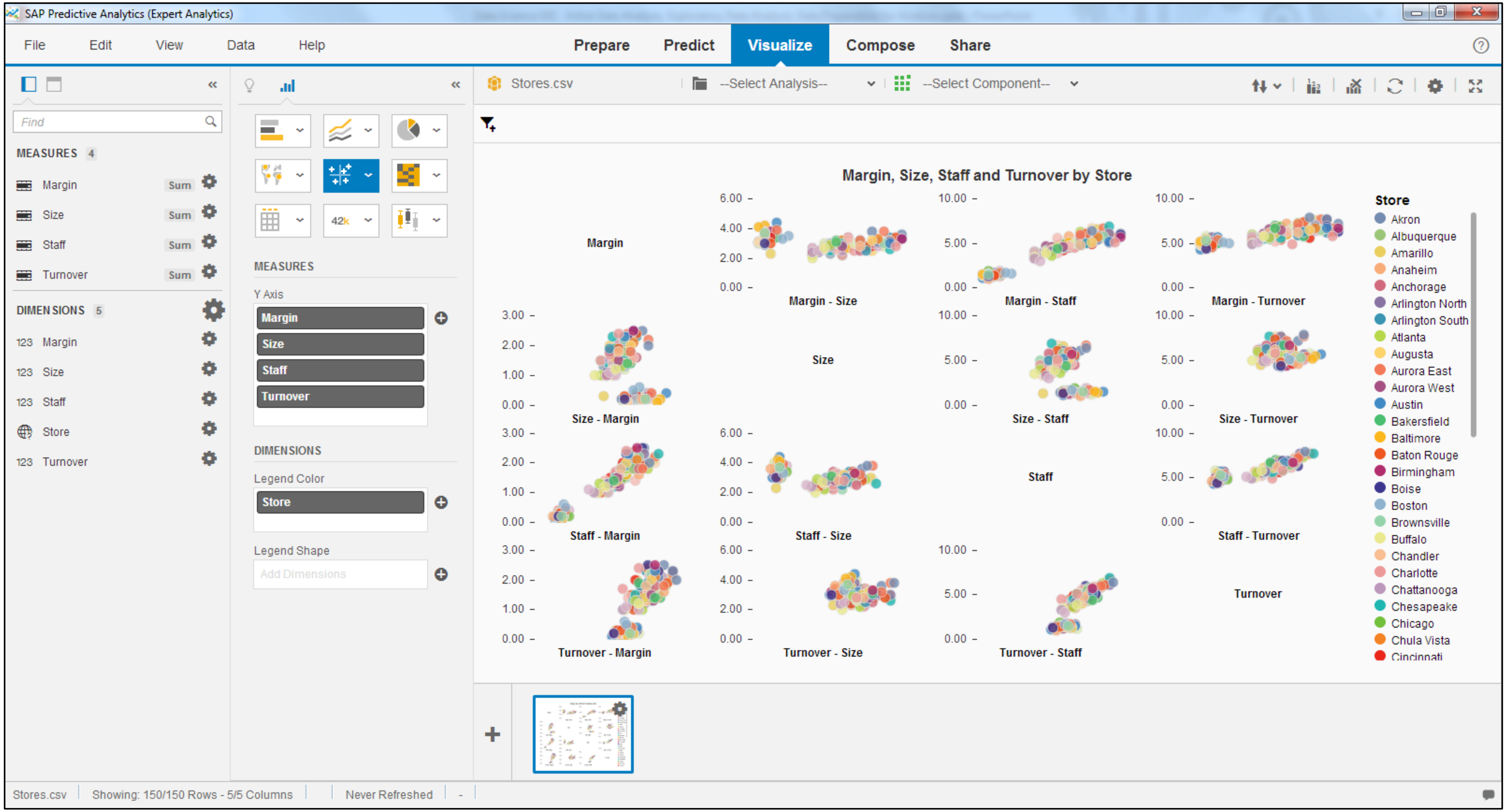


Margin, Size, Staff and Turnover by Store

Store	Margin	Size	Staff	Turnover
Akron	2.50	3.60	6.10	7.20
Albuquerque	0.40	3.40	1.50	5.40
Amarillo	2.30	3.20	5.70	6.90
Anaheim	1.30	2.30	4.00	5.50
Anchorage	1.40	2.90	4.70	6.10
Arlington North	0.20	3.30	1.40	5.00
Arlington South	1.90	2.70	5.10	5.80
Atlanta	0.20	3.40	1.50	5.10
Augusta	2.30	3.20	5.30	6.40
Aurora East	1.30	2.80	4.50	5.70
Aurora West	2.10	3.00	5.50	6.80
Austin	0.10	3.00	1.10	4.30
Bakersfield	1.40	3.20	4.70	7.00
Baltimore	0.20	3.40	1.70	5.40
Baton Rouge	1.50	3.00	4.50	5.40
Birmingham	1.30	2.90	4.20	5.70

Initial Data Analysis & Exploratory Data Analysis

Example – Scatter plot matrix



MEASURES

Y Axis: Margin, Size, Staff, Turnover

DIMENSIONS

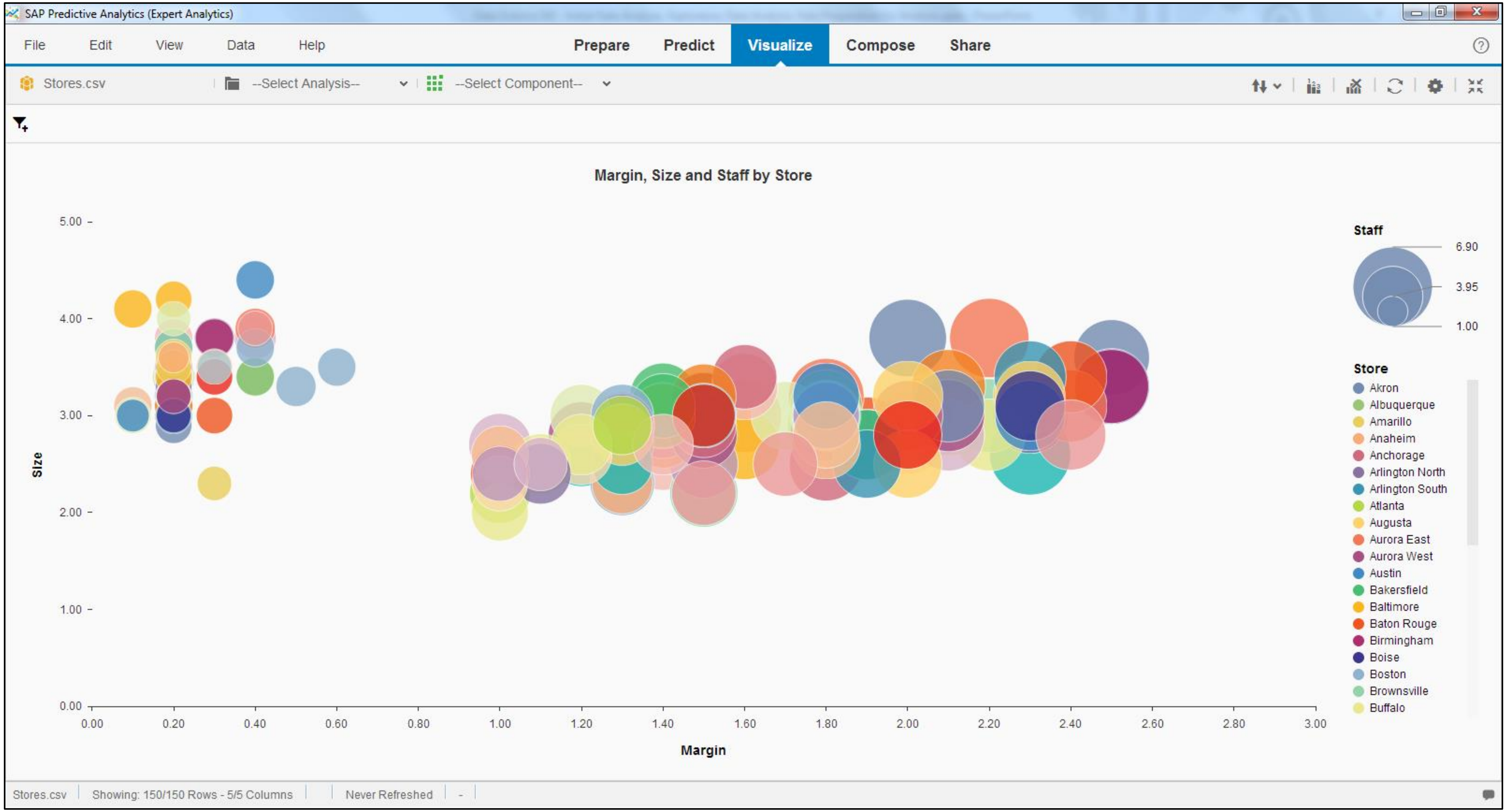
Legend Color: Store

Legend Shape: Add Dimensions

Stores.csv | Showing: 150/150 Rows - 5/5 Columns | Never Refreshed | -

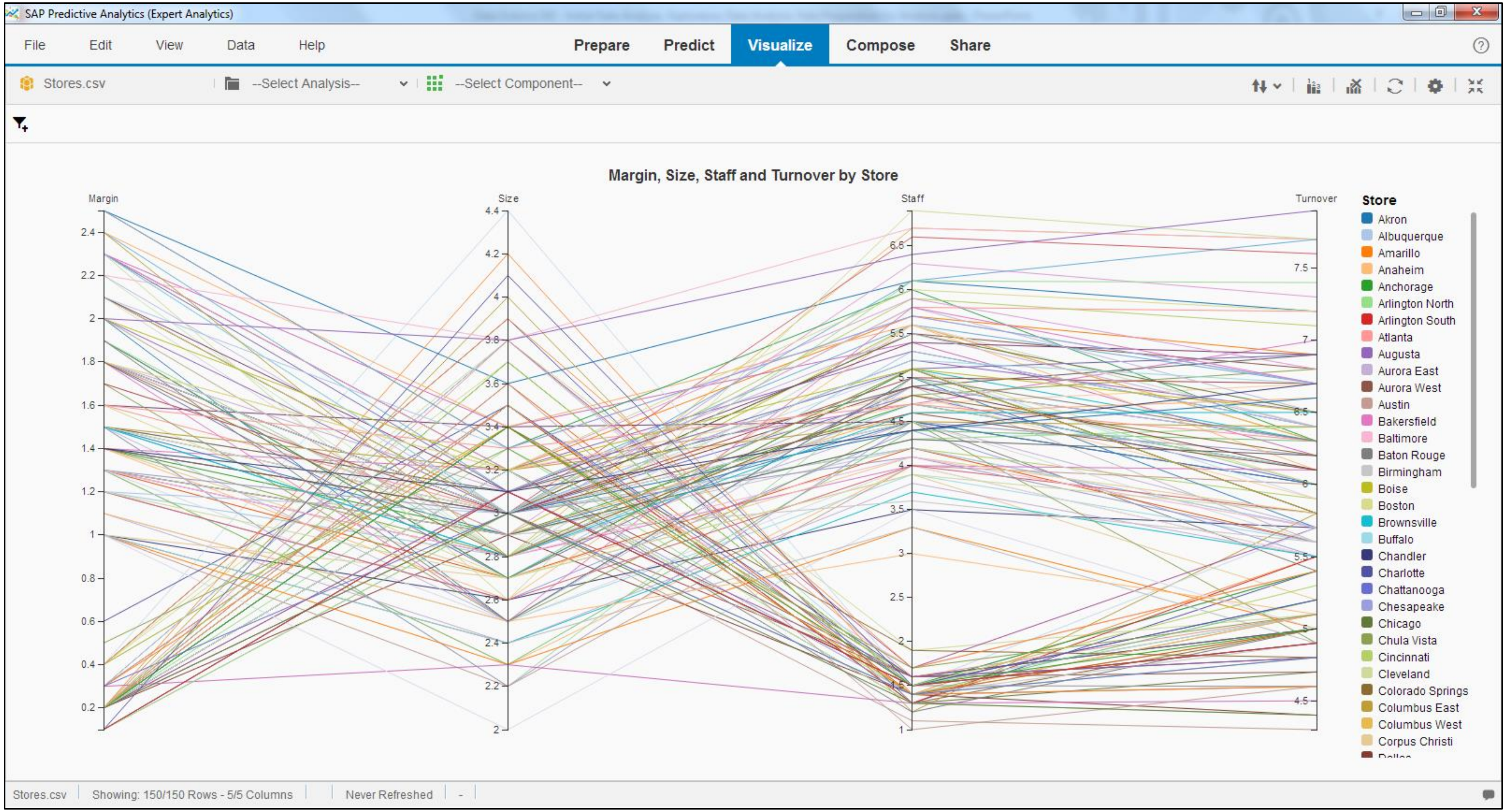
Initial Data Analysis & Exploratory Data Analysis

Example – Bubble plot



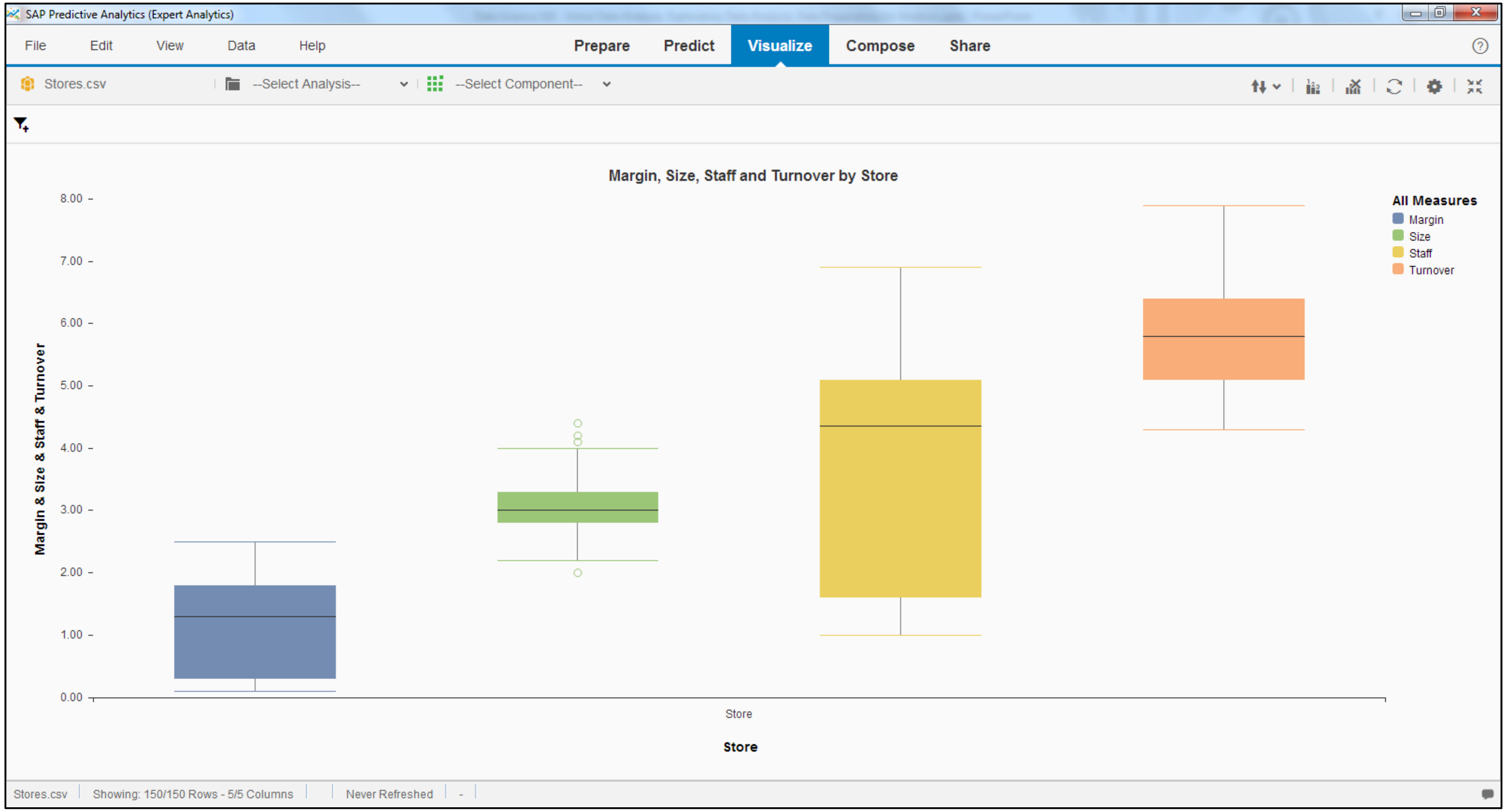
Initial Data Analysis & Exploratory Data Analysis

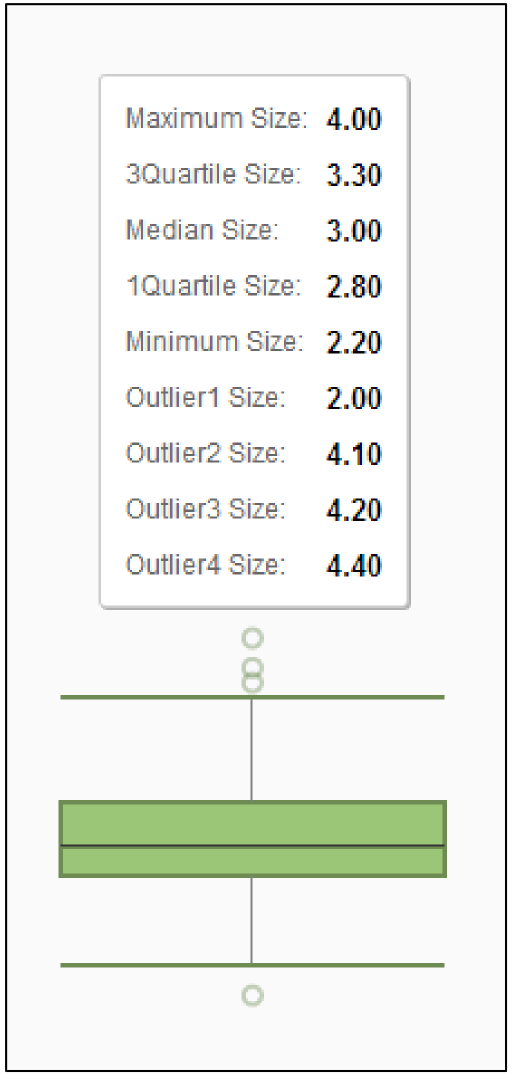
Example – Parallel co-ordinate plot



Initial Data Analysis & Exploratory Data Analysis

Example – Box plot

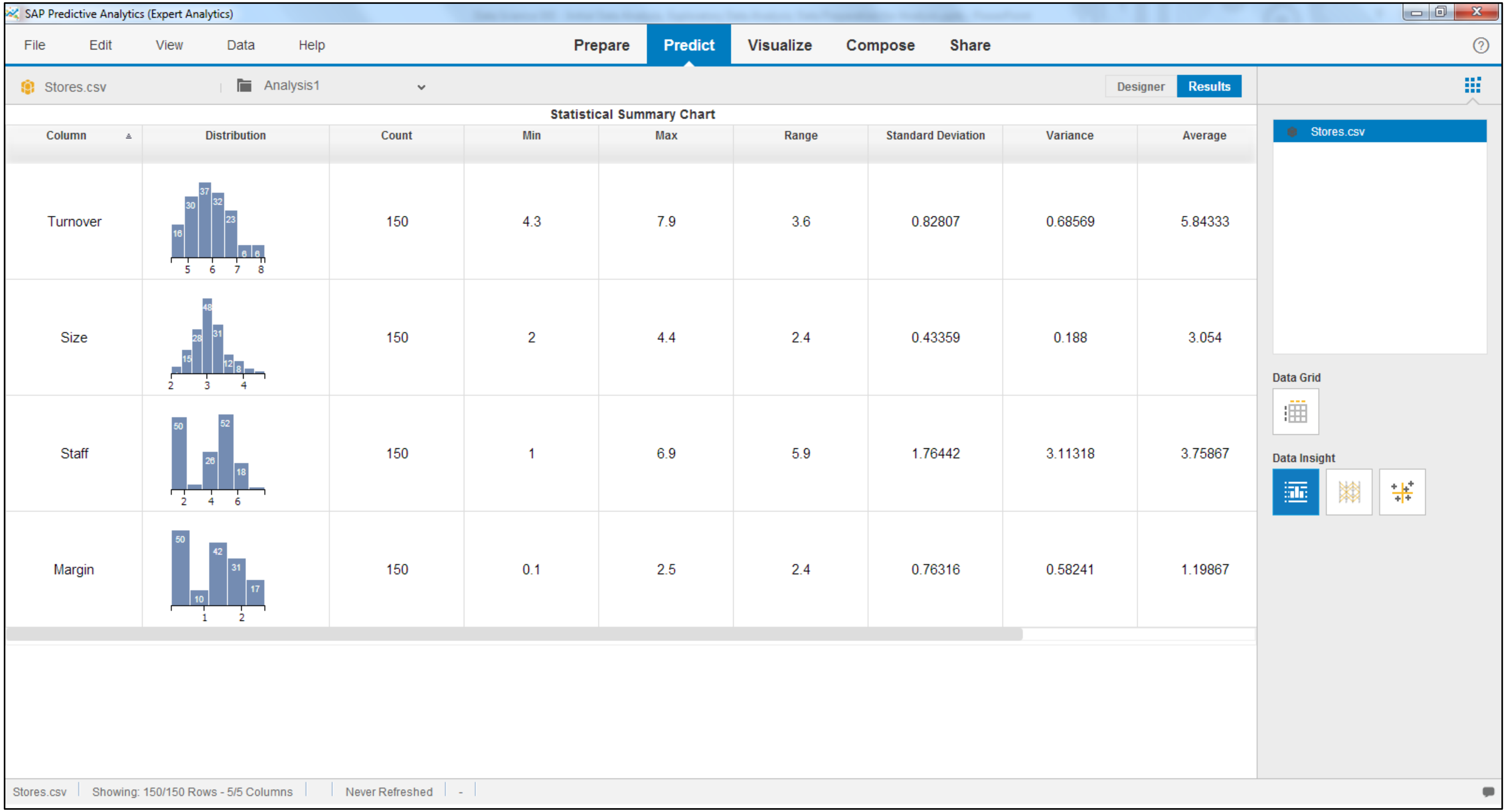




Maximum Size:	4.00
3Quartile Size:	3.30
Median Size:	3.00
1Quartile Size:	2.80
Minimum Size:	2.20
Outlier1 Size:	2.00
Outlier2 Size:	4.10
Outlier3 Size:	4.20
Outlier4 Size:	4.40

Initial Data Analysis & Exploratory Data Analysis

Example – Statistical summary chart



Statistical Summary Chart

Column	Distribution	Count	Min	Max	Range	Standard Deviation	Variance	Average
Turnover		150	4.3	7.9	3.6	0.82807	0.68569	5.84333
Size		150	2	4.4	2.4	0.43359	0.188	3.054
Staff		150	1	6.9	5.9	1.76442	3.11318	3.75867
Margin		150	0.1	2.5	2.4	0.76316	0.58241	1.19867

Stores.csv | Showing: 150/150 Rows - 5/5 Columns | Never Refreshed | -

Thank you

Contact information:

open@sap.com

© 2016 SAP SE or an SAP affiliate company. All rights reserved.

Some software products marketed by SAP SE and its distributors contain proprietary software components of other software vendors.

National product specifications may vary.

These materials are provided by SAP SE or an SAP affiliate company for informational purposes only, without representation or warranty of any kind, and SAP SE or its affiliated companies shall not be liable for errors or omissions with respect to the materials. The only warranties for SAP SE or SAP affiliate company products and services are those that are set forth in the express warranty statements accompanying such products and services, if any. Nothing herein should be construed as constituting an additional warranty.

In particular, SAP SE or its affiliated companies have no obligation to pursue any course of business outlined in this document or any related presentation, or to develop or release any functionality mentioned therein. This document, or any related presentation, and SAP SE's or its affiliated companies' strategy and possible future developments, products, and/or platform directions and functionality are all subject to change and may be changed by SAP SE or its affiliated companies at any time for any reason without notice. The information in this document is not a commitment, promise, or legal obligation to deliver any material, code, or functionality. All forward-looking statements are subject to various risks and uncertainties that could cause actual results to differ materially from expectations. Readers are cautioned not to place undue reliance on these forward-looking statements, which speak only as of their dates, and they should not be relied upon in making purchasing decisions.