

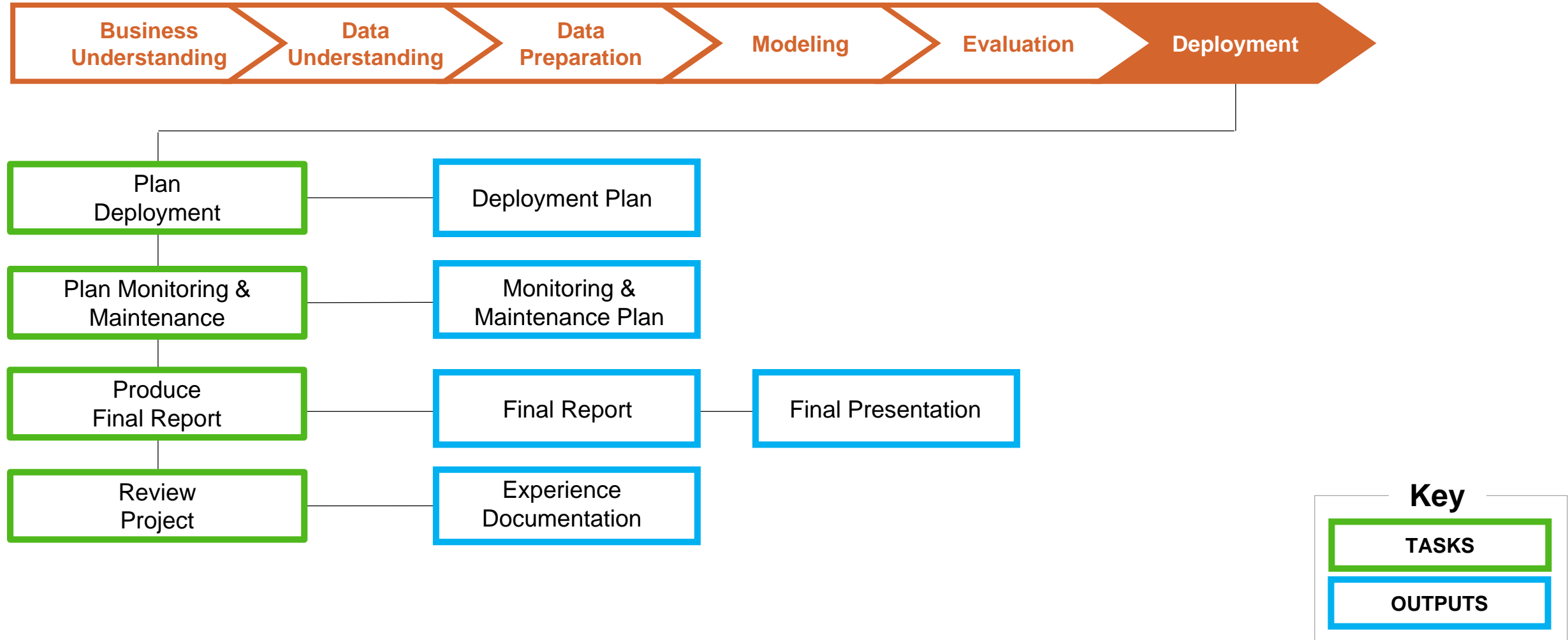
# Week 6 Unit 1: Deployment Phase – Overview





# Deployment Phase – Overview

## CRISP-DM – Phase 6: Deployment



# Deployment Phase – Overview

## Phase 6.1: Plan Deployment

- **Task**
  - Plan deployment.
- **Output – Deployment Plan**
  - Summarize the deployment strategy, including the necessary steps and how to perform them.



# Deployment Phase – Overview

## Phase 6.2: Plan Monitoring & Maintenance

- **Task**
  - Monitoring and maintenance of models and analysis are critical issues.
  - A careful preparation of a maintenance strategy is essential.
- **Output – Monitoring and Maintenance Plan**
  - Summarize the monitoring and maintenance strategy.



# Deployment Phase – Overview

## Phase 6.3: Produce Final Report

---

- **Task**
  - At the end of the project, the project leader and his team write up a final report.
- **Output – Final Report**
  - This is the final written report of the data mining engagement.
- **Output – Final Presentation**
  - There will also often be a meeting at the conclusion of the project where the results are verbally presented to the customer.



# Deployment Phase – Overview

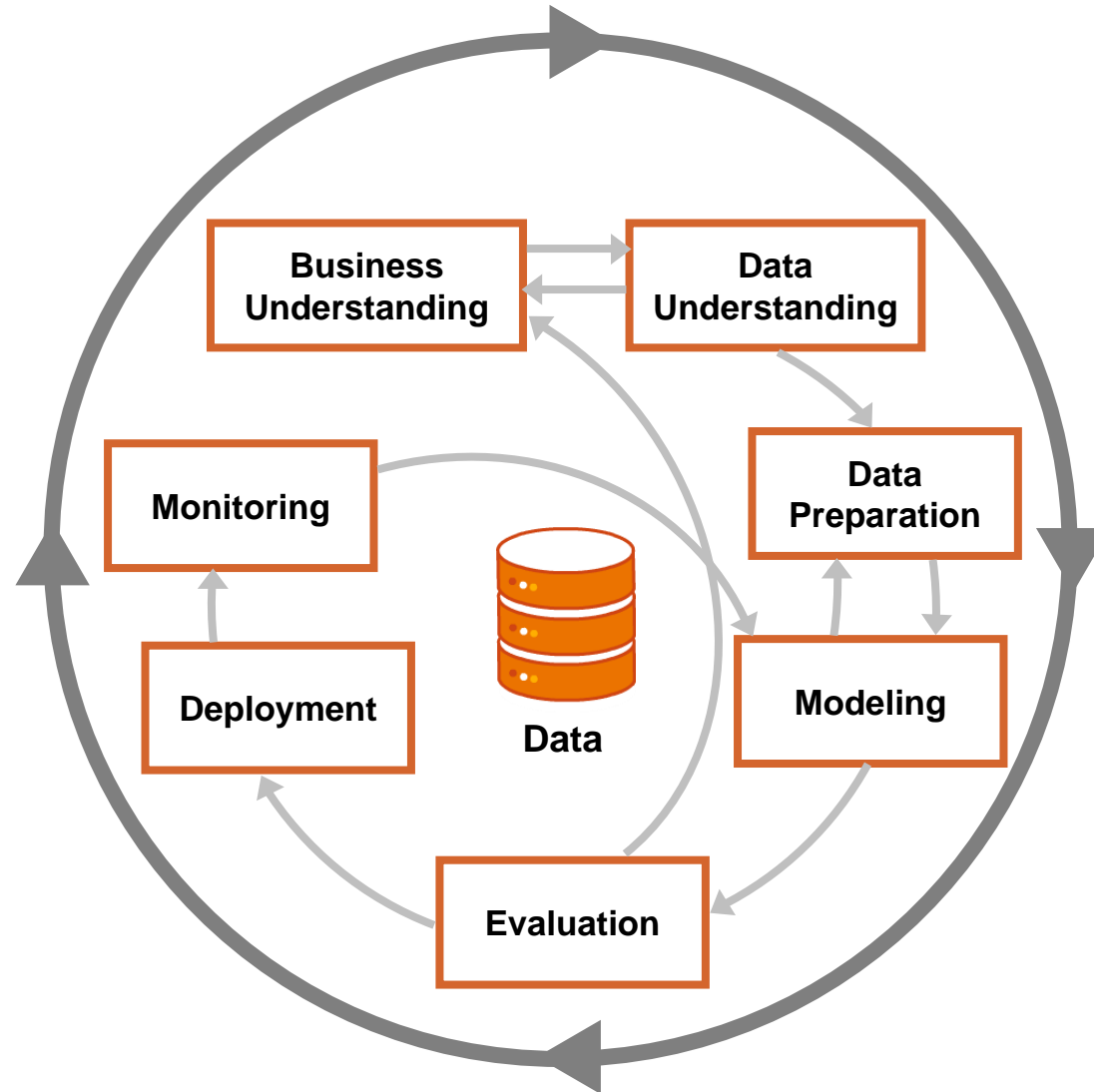
## Phase 6.4: Review Project

- **Task**
  - Assess what went right and what went wrong, what was done well and what needs to be improved.
- **Output – Experience Documentation**
  - Summarize important experiences made during the project.



# Deployment Phase – Overview

## CRISP-DM – Update





# Thank you

Contact information:

[open@sap.com](mailto:open@sap.com)

**openSAP**  
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 6 Unit 2: Deployment Options



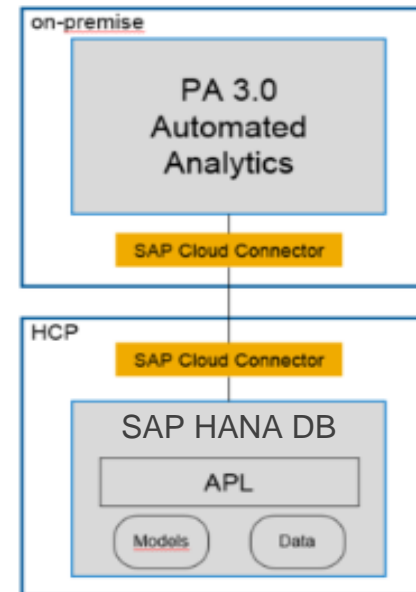


# Deployment Options

## Introduction

- There are a number of options to deploy the analysis:
  - Model scoring
  - Model integrated with BI Reporting
  - Model integrated with application
  - In the cloud

### PA with SAP HANA DB via HCP



- Requirements
  - Predictive Analytics 3.0.0 on Premise Suite License
  - SAP HANA Cloud Platform with HCP predictive services deployed
  - APL 3.0.0 Deployed on HCP
  - Database Connection is made through SAP Cloud Connector
  - User needs APL rights on Database
  - Configure PA to delegate to APL

# Deployment Options

## PMML (Predictive Model Markup Language)

---

The Predictive Model Markup Language (PMML) is an XML language for statistical and data mining models, which makes it easy to move models between different applications and platforms.

```
<?xml version="1.0"?>
<PMML version="4.1"
  xmlns="http://www.dmg.org/PMML-4_1"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">

  <Header copyright="Example.com"/>
  <DataDictionary> ... </DataDictionary>

  ... a model ...

</PMML>
```

General Structure of a PMML Document

# Deployment Options

## Real-time scoring

In some business applications, real-time scoring is necessary.





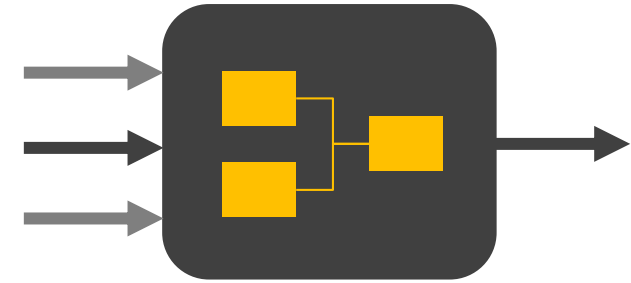
# Deployment Options

## What is event stream processing?

---

### Real-time event stream processing

1. **Capture** data arriving continuously from devices and applications
2. **Act** on new information as soon as it arrives: alerts, notifications, and immediate response to changing conditions
3. **Stream** information to live operational dashboards



**Highly scalable – processing hundreds of thousands or even millions of events per second**

# Deployment Options

## Smart data streaming capabilities

---

### Stream Capture

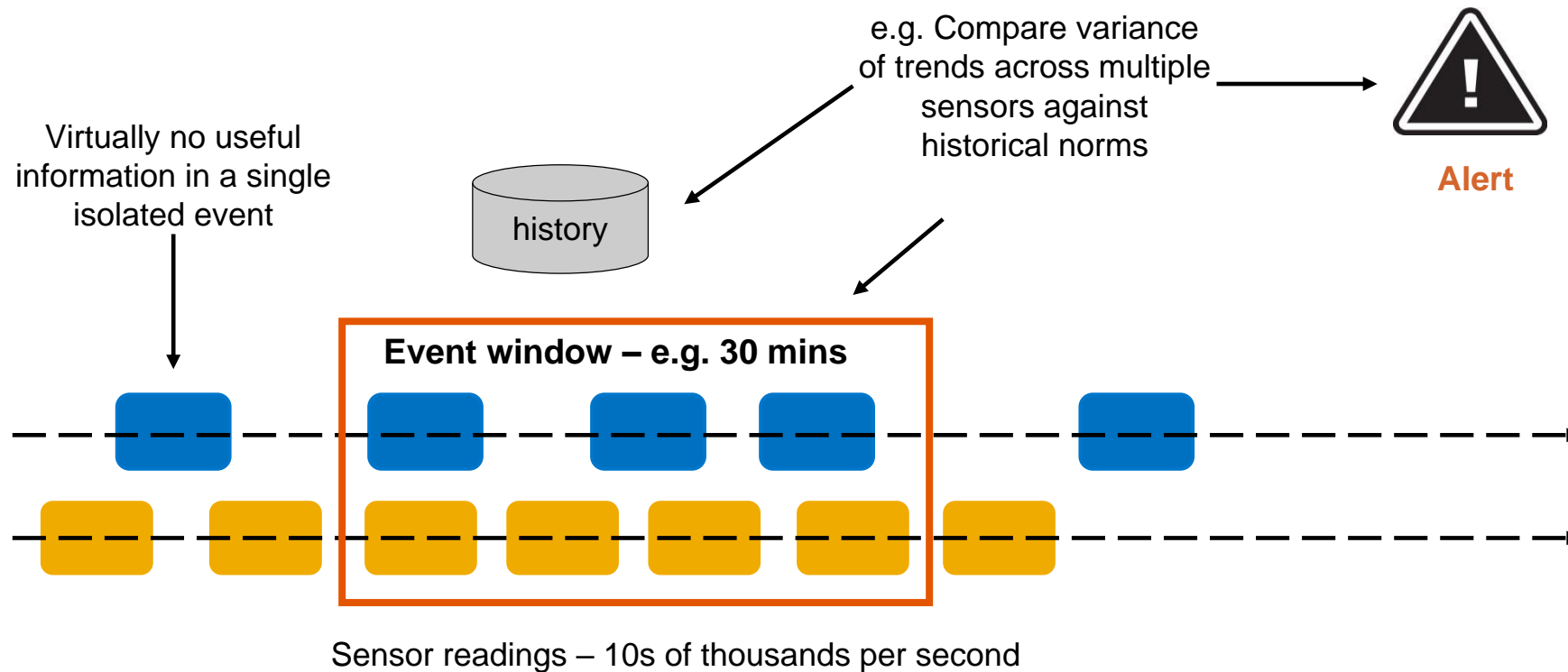
- **Capture data arriving as individual events – at potentially high speeds**
  - Hundreds of thousands or millions of events per second
  - Micro-batching and parallel processing to optimize load speeds
- **Capture events that are published from streaming sources**
  - e.g. message bus
- **Filter, transform, or enrich the data on the way in**
  - Capture only the data you want, in the form you need it
- **Prioritize data**
  - Capture high value data in SAP HANA and direct other data into Hadoop

### Immediate Response

- **Monitor incoming event streams**
  - Watch for **trends or patterns**
  - Monitor **correlations**
  - Detect **missing events**
  - **Continuously update and monitor** aggregate statistics
- **Generate alerts, notifications**
- **Initiate immediate response**

# Deployment Options

Complex event processing extracts insight from events





# Thank you

Contact information:

[open@sap.com](mailto:open@sap.com)

**openSAP**  
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 6 Unit 3: Monitoring & Maintenance



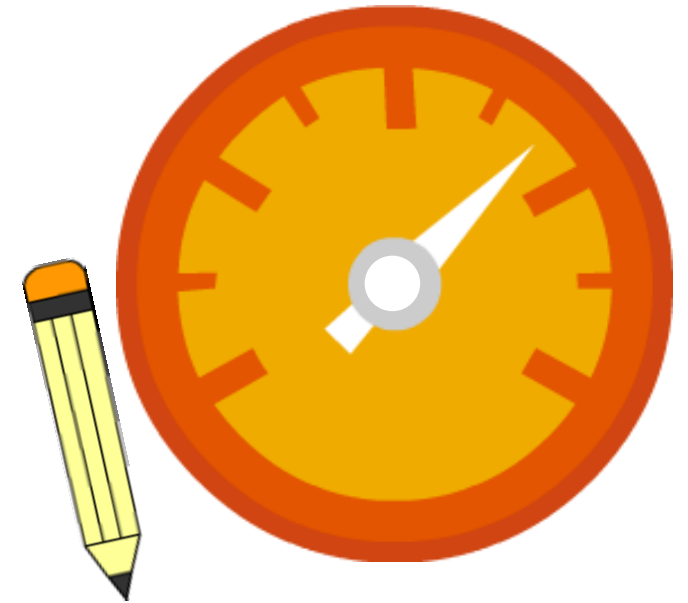


# Monitoring & Maintenance

## Introduction

---

- The predictive models that have been deployed into the business environment must be regularly monitored for acceptable performance.
- Remember – model performance will reduce over time.

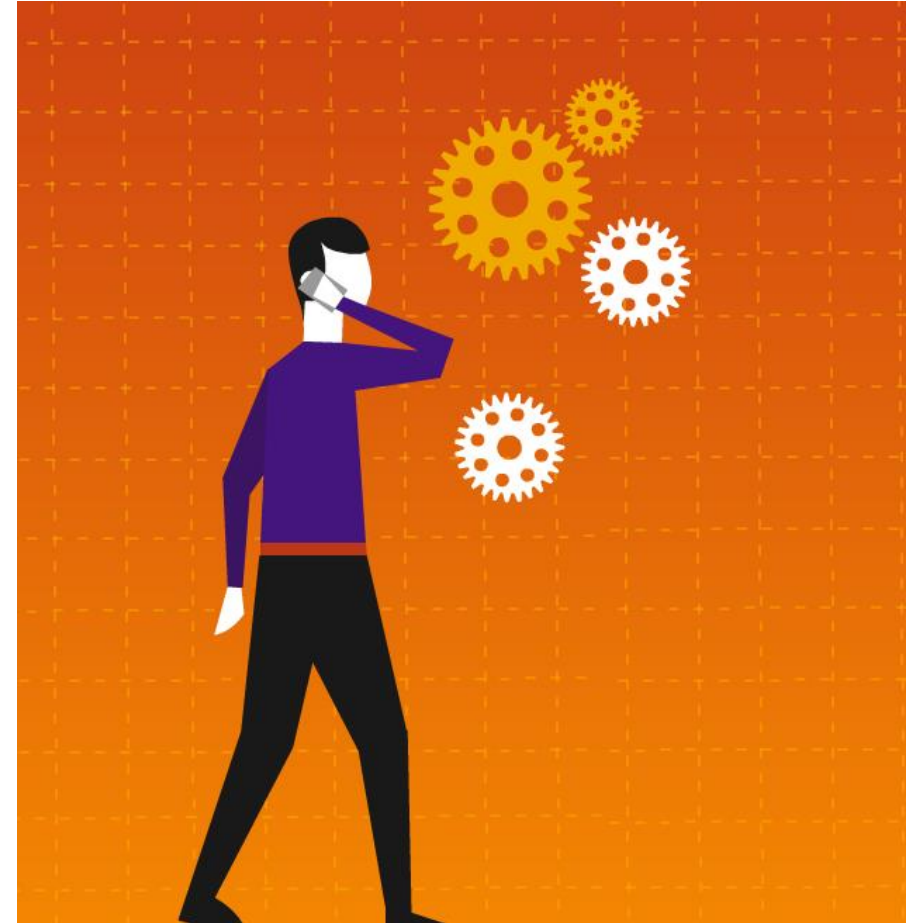


# Monitoring & Maintenance

## Approach

---

- Each model should be thoroughly evaluated and measured for accuracy, predictive power, stability over time, and other appropriate metrics that are defined by each model's objectives.



# Monitoring & Maintenance

## Update frequency

- Factors that influence when you should update your models:
  - Business environment change
  - Frequency of model usage
  - The age of a model
  - Data readiness



# Monitoring & Maintenance Process

---

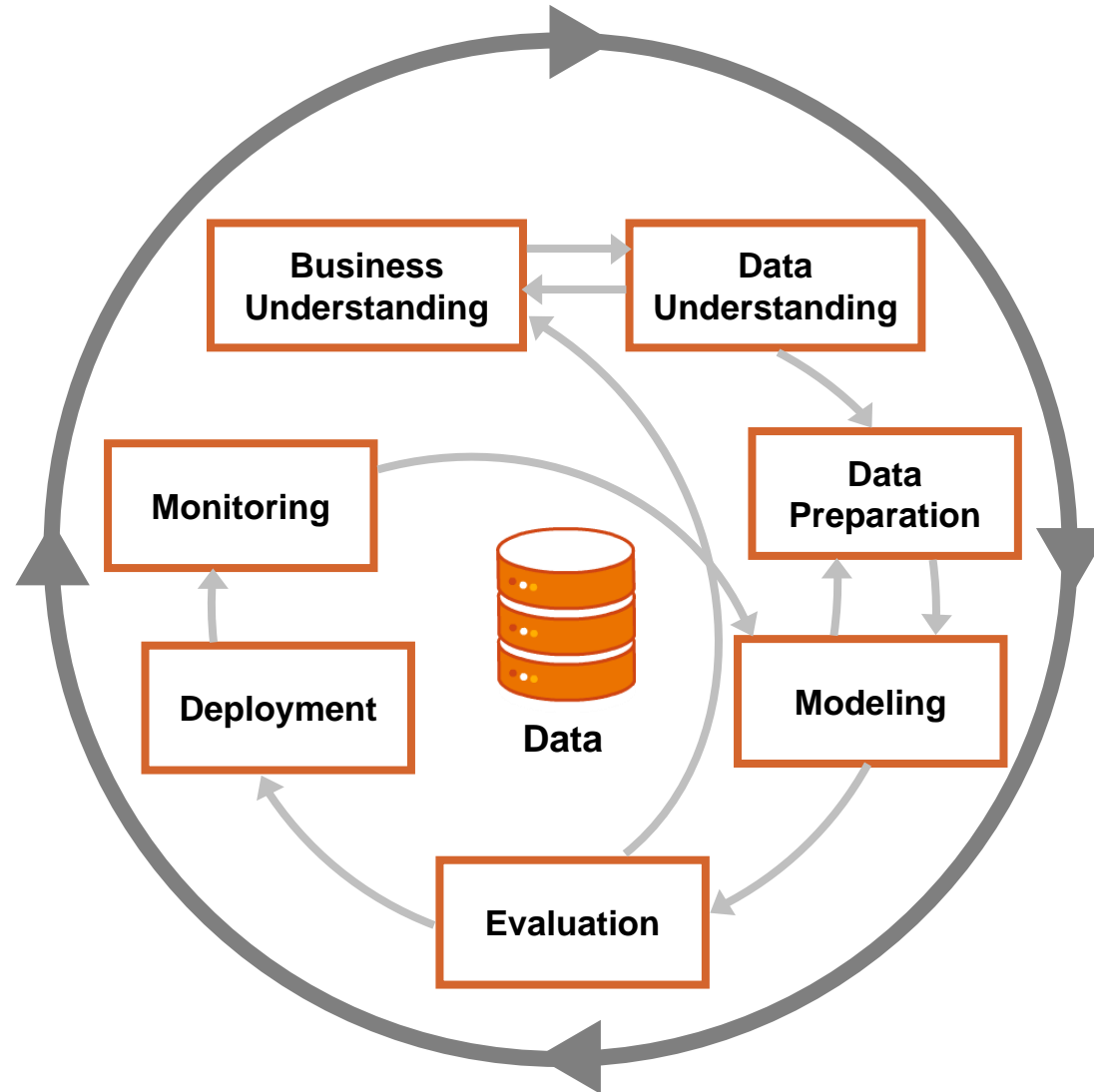
- Those models that are not running at their best may need to be refreshed (or recalibrated) with current data or rebuilt completely.





# Monitoring & Maintenance

## CRISP-DM – Update





# Thank you

Contact information:

[open@sap.com](mailto:open@sap.com)

**openSAP**  
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 6 Unit 4: Automating Deployment & Maintenance



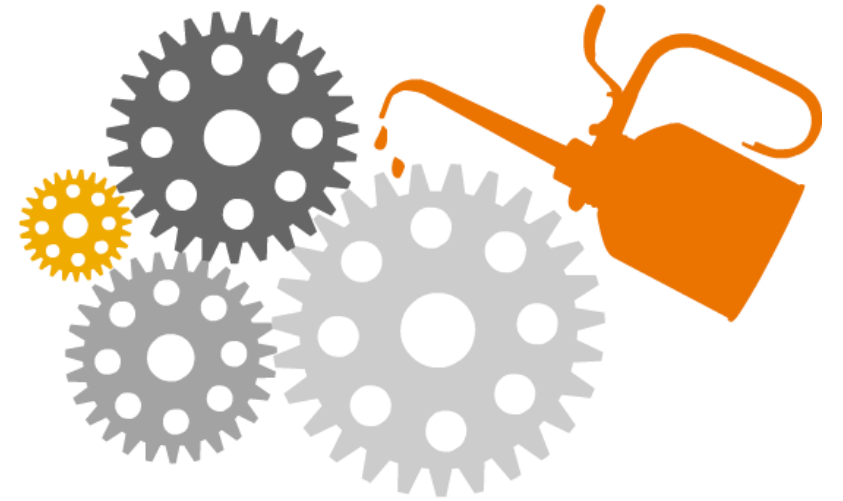


# Automating Deployment & Maintenance

## Introduction

---

- Once an organization has developed 100s or possibly 1000s of predictive models, it becomes very difficult to apply them on new data and maintain them over time.
- Therefore, it is important that the processes of deploying and maintaining models can be automated, so that they can be achieved as efficiently as possible and free-up scarce data scientist resources.





# Automating Deployment & Maintenance

## Basic processes

---

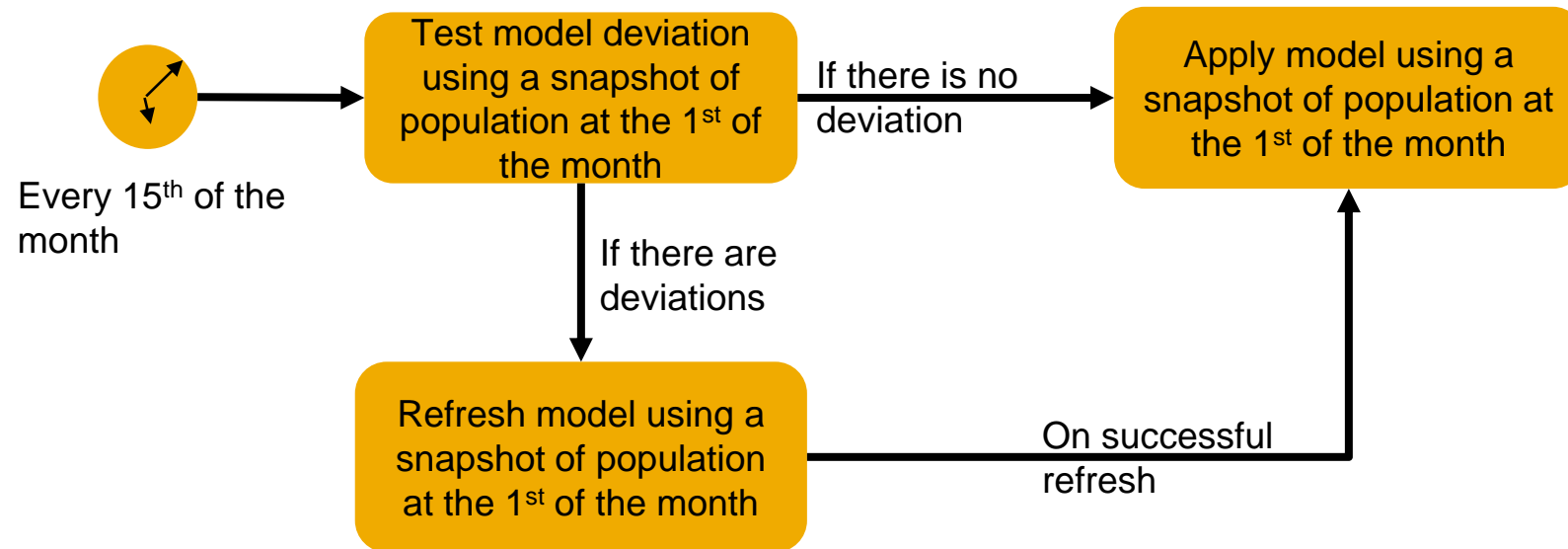
- There are a number of basic processes that require automation:
  - Model scheduling
  - Apply dataset build
  - Data deviation
  - Model deviation analysis
  - Time series anomaly
  - Automated model refresh or rebuild



# Automating Deployment & Maintenance

## Model automation scenario (workflow)

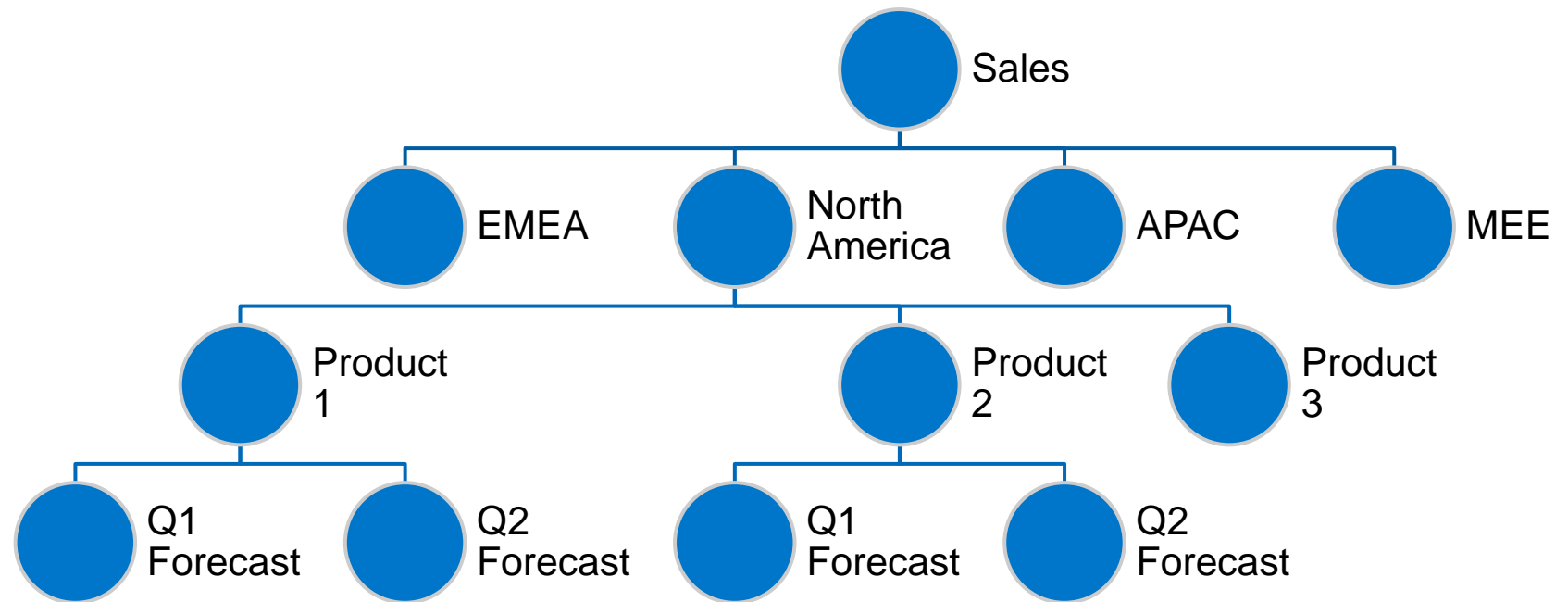
Here is an example workflow of tasks we might want to set up:



# Automating Deployment & Maintenance

## Segmented modeling in SAP Predictive Factory

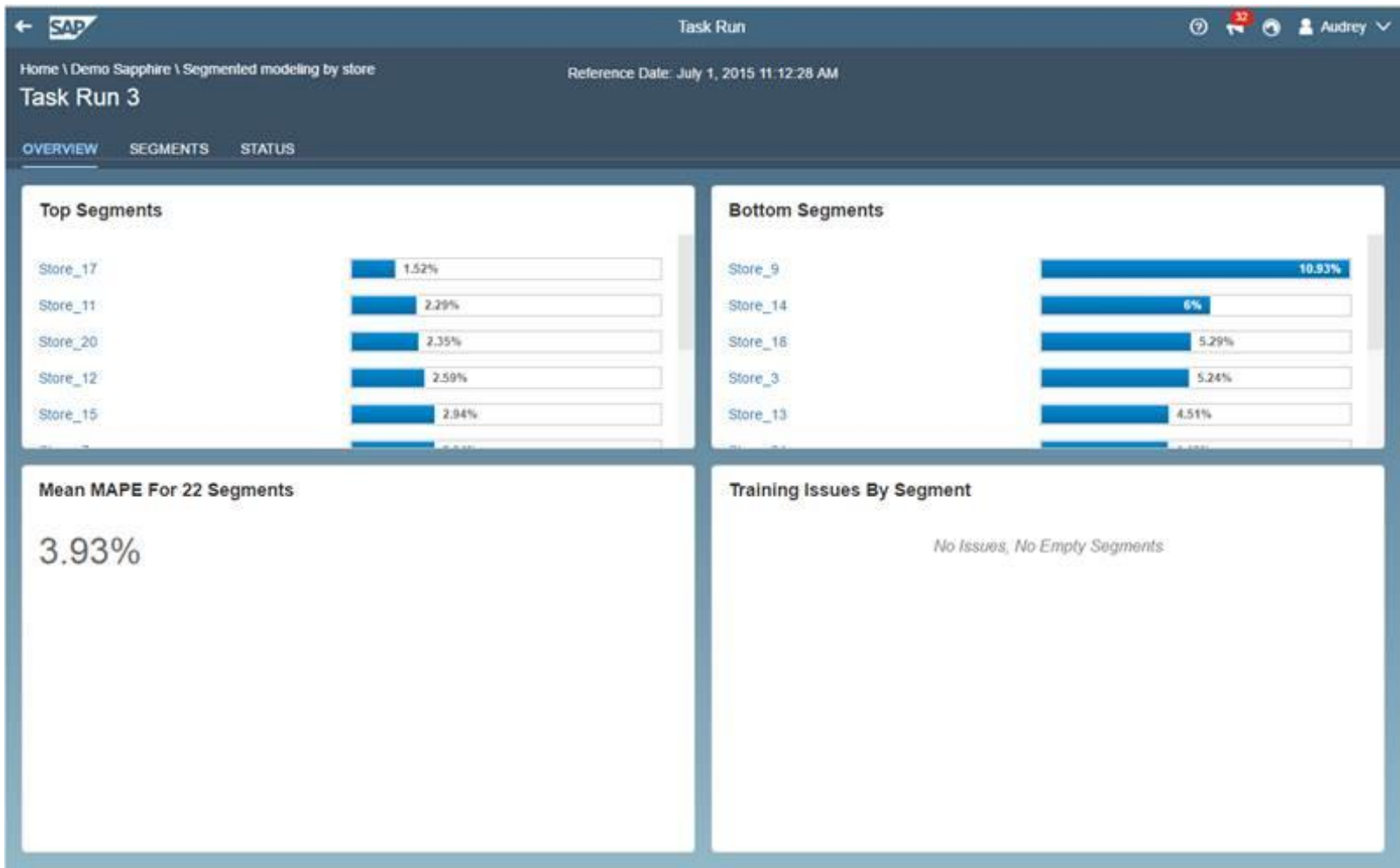
One of the features of the SAP Predictive Factory is that it enables automated segmented modeling for a time series model.



Build thousands of models in a single operation

# Automating Deployment & Maintenance

## Segmented modeling in SAP Predictive Factory





# Thank you

Contact information:

[open@sap.com](mailto:open@sap.com)

**openSAP**  
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 6 Unit 5: Myths & Challenges





# Myths & Challenges

## Myth 1 – Data science is all algorithms

$$W_p(\mu, \nu) := \left( \inf_{\gamma \in \Gamma(\mu, \nu)} \int_{M \times M} d(x, y)^p d\gamma(x, y) \right)^{1/p}$$

$$\begin{aligned} \text{Var}[a^T X] &= \frac{1}{n} \sum_{i=1}^n \left\{ a^T \left( X_i - \frac{1}{n} \sum_{j=1}^n X_j \right) \right\}^2 \\ &= a^T V_{XX} a. \end{aligned}$$

where

$$V_{XX} = \frac{1}{n} \sum_{i=1}^n \left( X_i - \frac{1}{n} \sum_{j=1}^n X_j \right) \left( X_i - \frac{1}{n} \sum_{j=1}^n X_j \right)^T$$

$$H(P, Q) := \sqrt{\frac{1}{2} \int \left( \sqrt{\frac{dP}{d\lambda}} - \sqrt{\frac{dQ}{d\lambda}} \right)^2 d\lambda}$$

$$f(x; \lambda, k) = \begin{cases} \frac{k}{\lambda} \left( \frac{x}{\lambda} \right)^{k-1} e^{-(x/\lambda)^k} & x \geq 0, \\ 0 & x < 0, \end{cases}$$

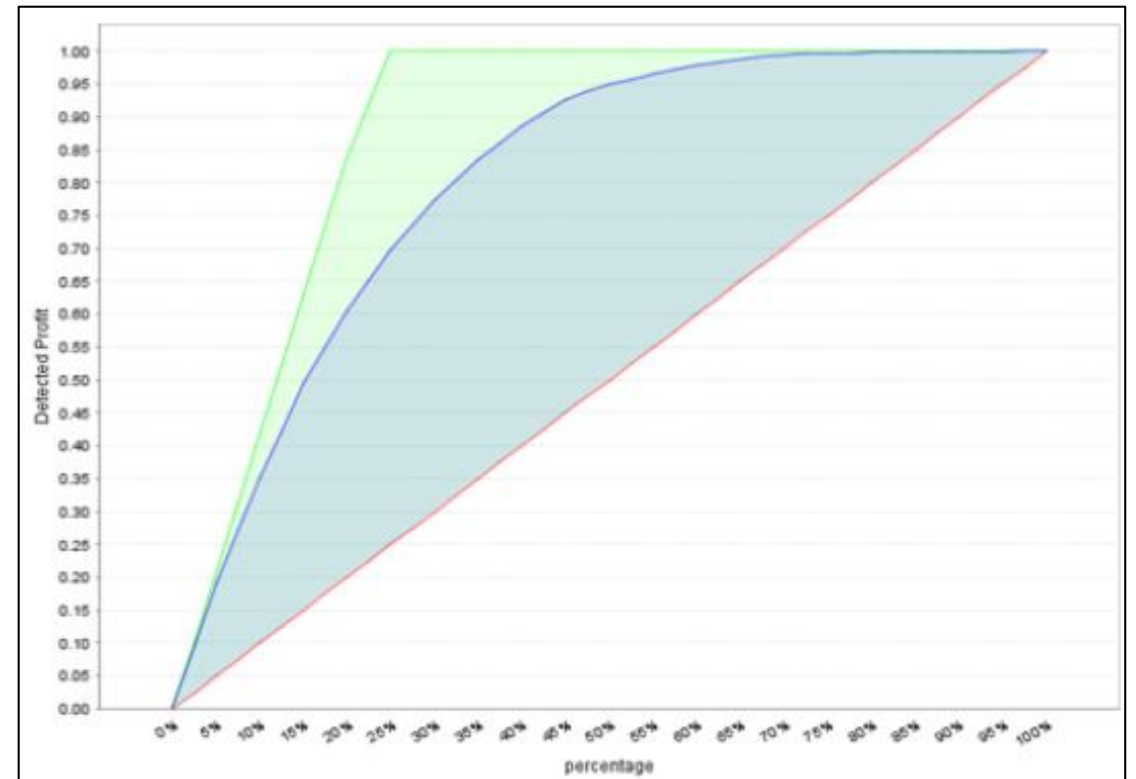
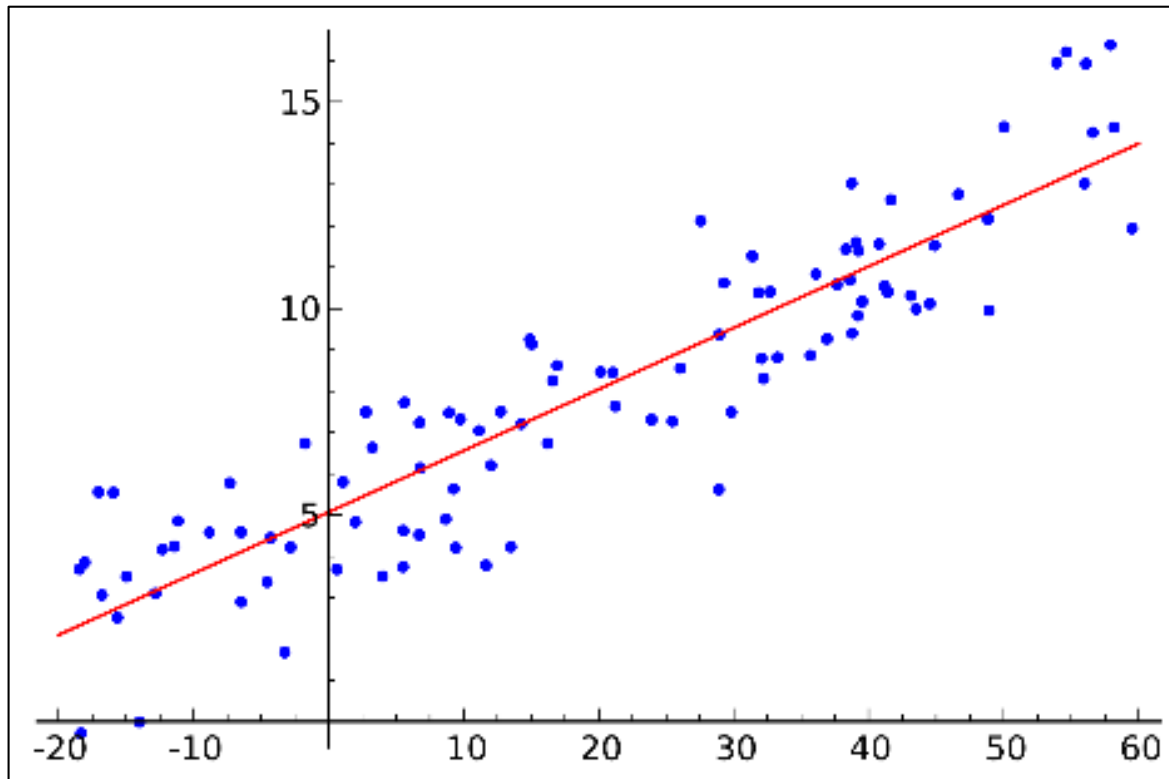
$$W(P, Q) = \min \left\{ \sum_{x \in M, y \in M} \text{dist}(x, y) \cdot T(x, y) \mid T: M \times M \rightarrow [0, 1] \text{ with } \sum_{y \in M} T(x, y) = P(x), \sum_{x \in M} T(x, y) = Q(y) \right\}$$

- You will often hear that “all you need for data science are good algorithms”!
- “Advances in data science = advances in algorithms”

# Myths & Challenges

## Myth 2 – Data science is all about accuracy

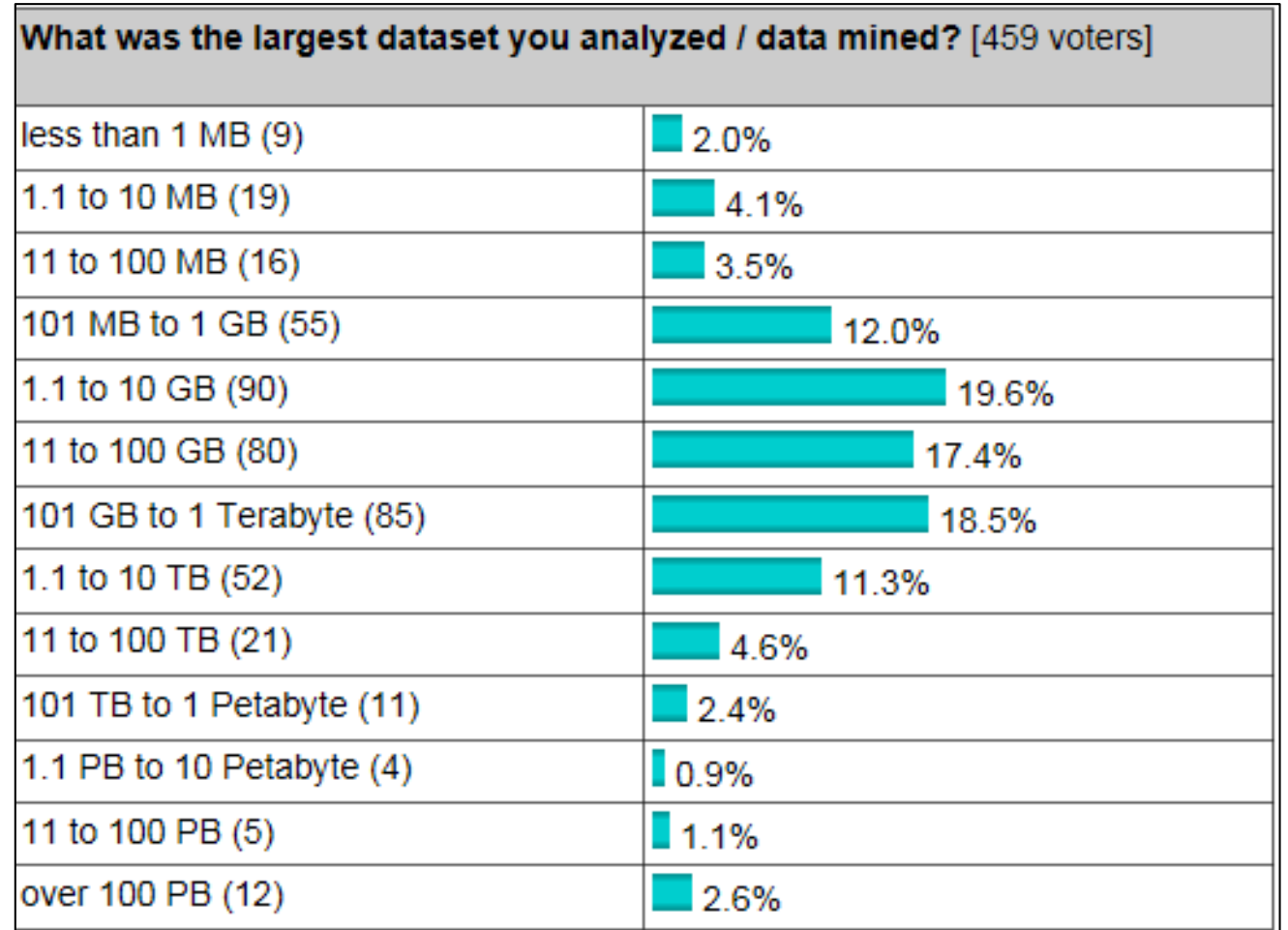
- You will also hear that “the main criterion for a good data mining algorithm is predictive accuracy”!



# Myths & Challenges

## Myth 3 – Data science is all about vast quantities of data

Many of the datasets we use in data science projects are relatively small, as this survey shows.

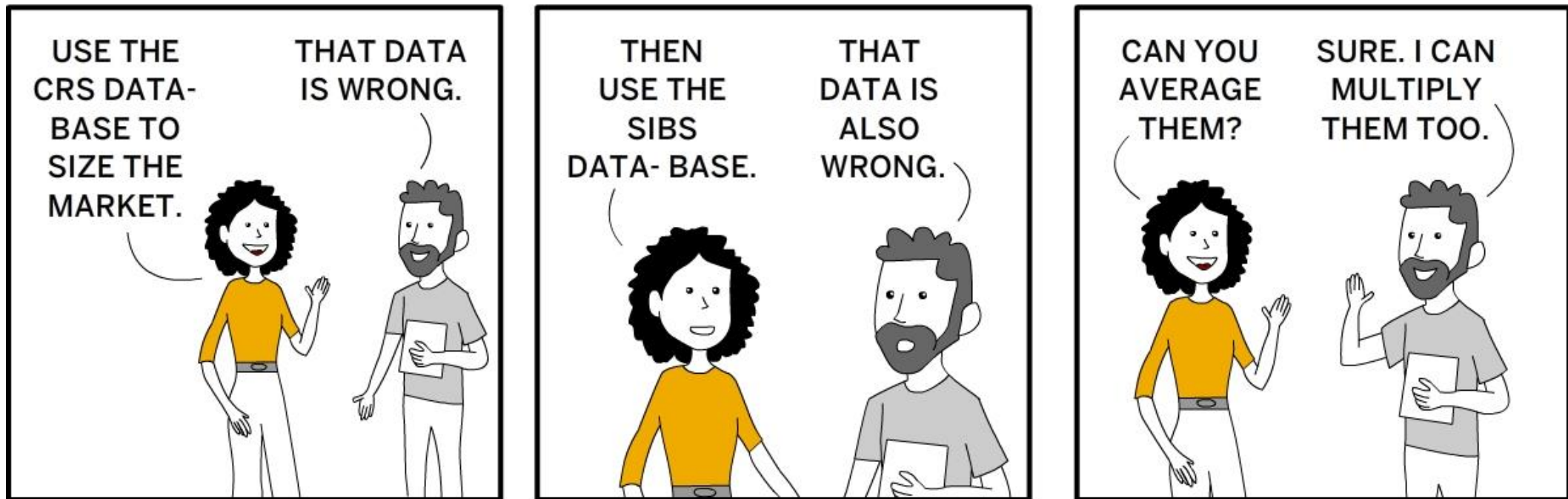


From a KDnuggets poll (2015)

# Myths & Challenges

## Challenge 1 – Throwing in data without thinking

- You must think about the data you are going to use before blindly using it to be a model.
- Rubbish in = Rubbish out





# Myths & Challenges

## Challenge 2 – Lack of business knowledge

- Business knowledge is crucial to guide the data mining process towards useful results, and to recognize and use them once found.



# Myths & Challenges

## Challenge 3 – Lack of data knowledge

- Data science projects involve lots of low-level questions about the data.

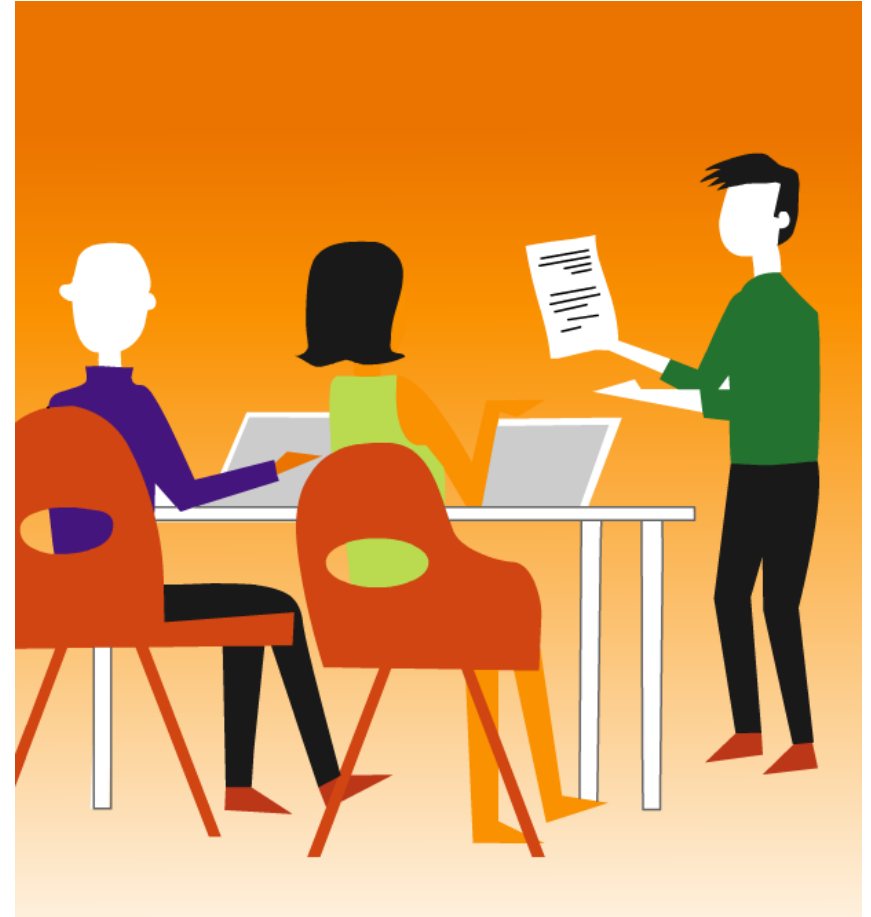


# Myths & Challenges

## Challenge 4 – Erroneous assumptions

- Even with a lot of data knowledge, we can be misled.

$$2 + 2 = 5$$



# Myths & Challenges

## Challenge 5 – Cause and effect

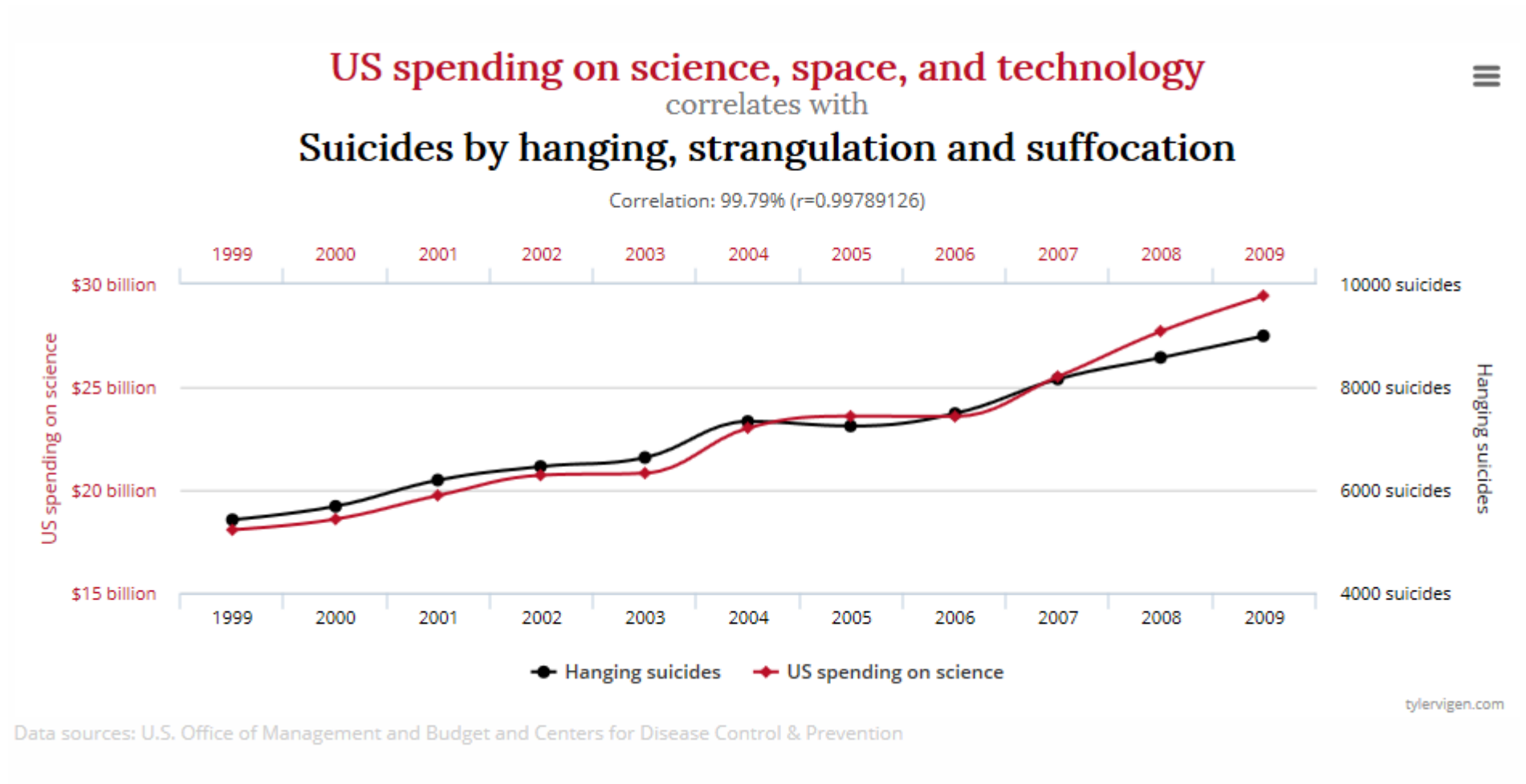
- *Cause and effect* is a relationship between events or things, where one is the result of the other or others.





# Myths & Challenges

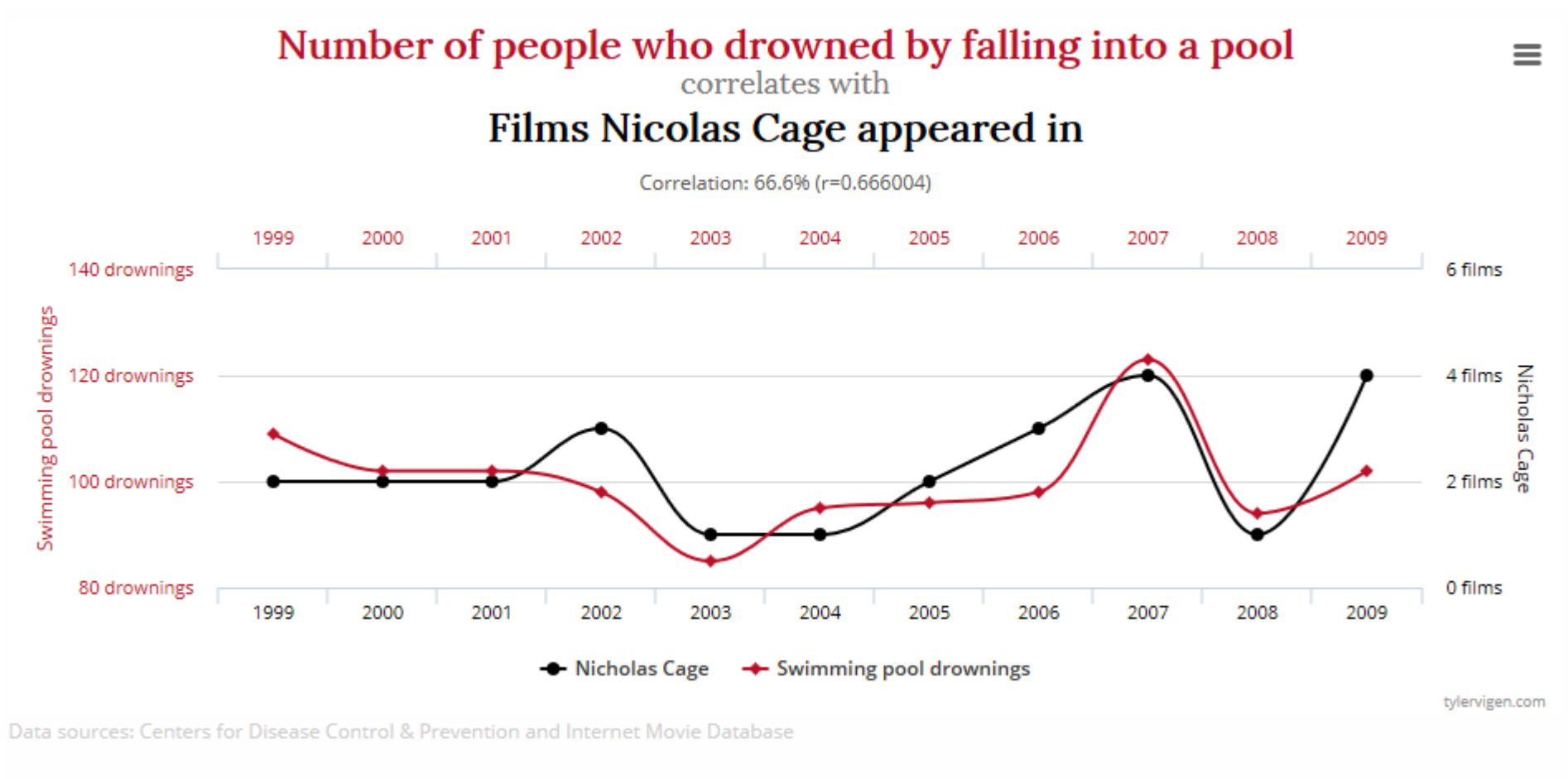
## Challenge 5 – Cause and effect



<http://tylervigen.com/spurious-correlations>

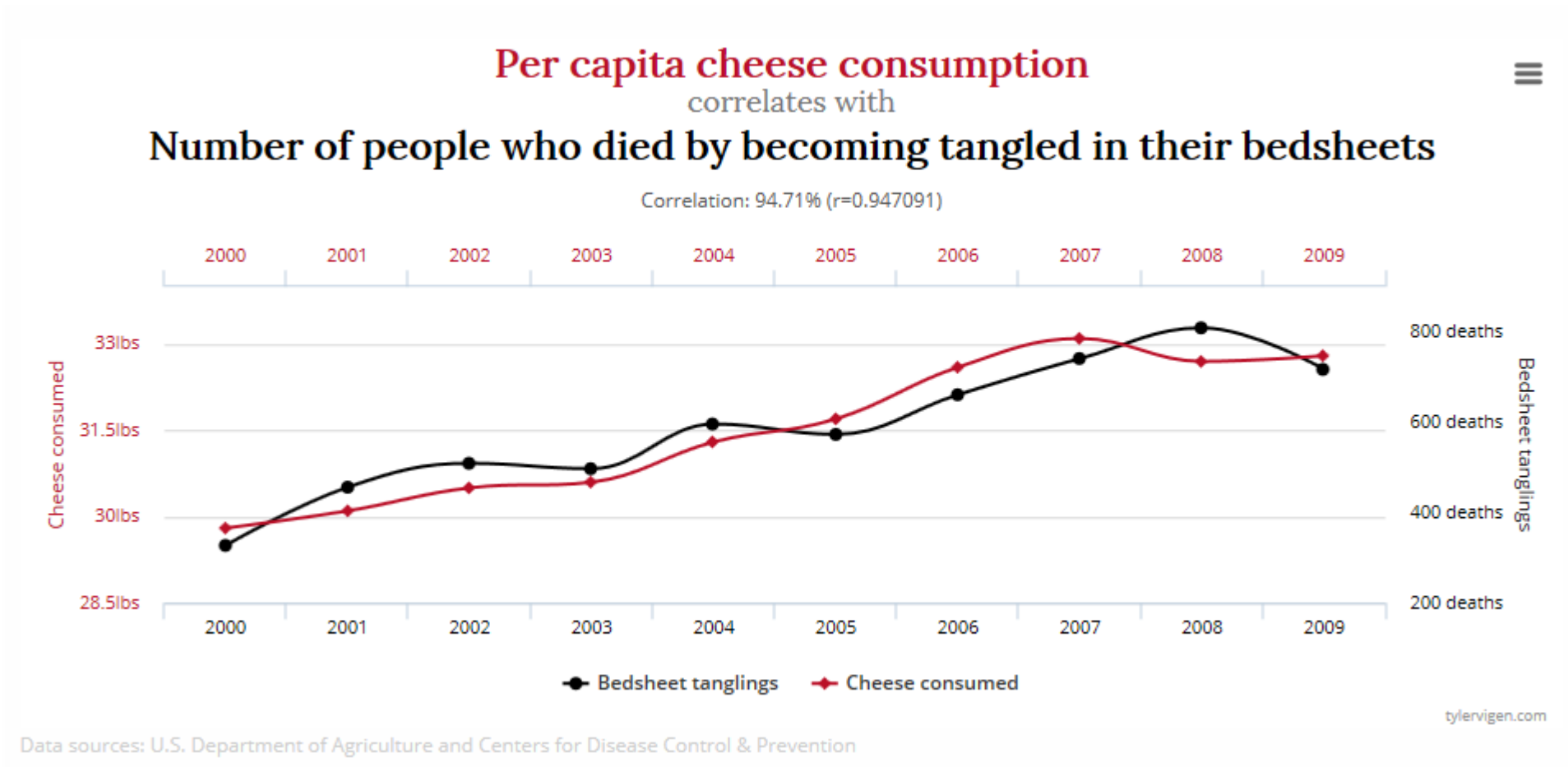
# Myths & Challenges

## Challenge 5 – Cause and effect



# Myths & Challenges

## Challenge 5 – Cause and effect



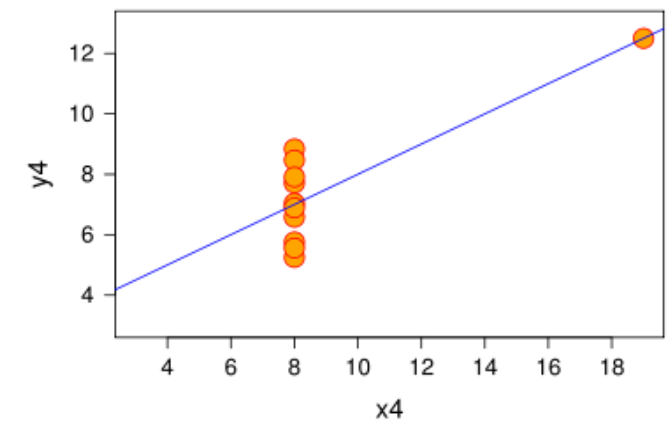
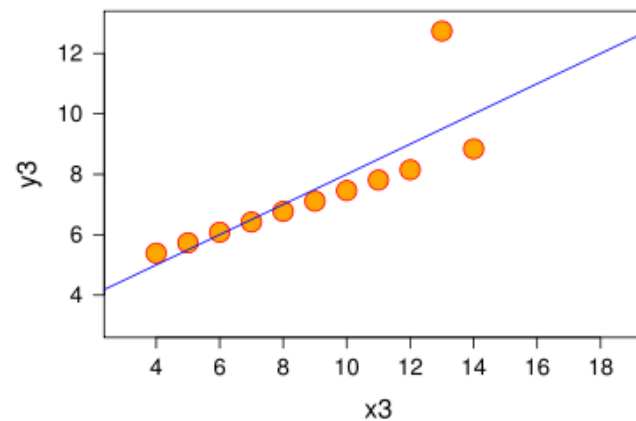
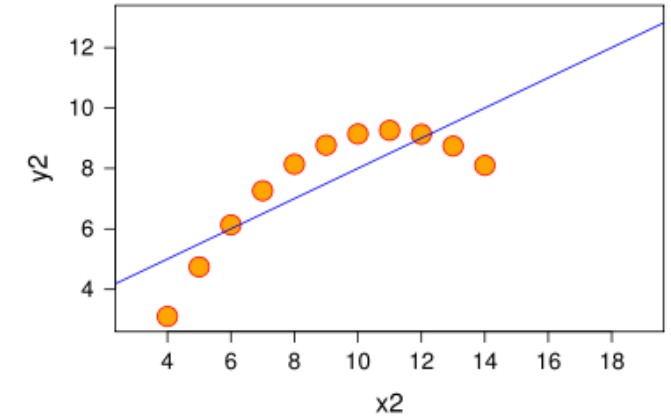
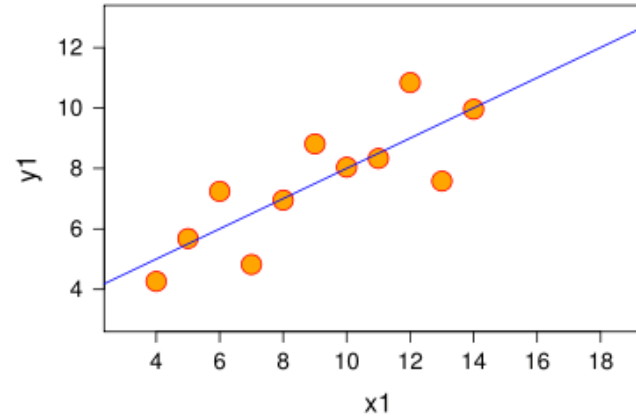
<http://tylervigen.com/spurious-correlations>

# Myths & Challenges

## Challenge 6 – Do not forget the importance of data visualization

I		II		III		IV	
x	y	x	y	x	y	x	y
10	8.04	10	9.14	10	7.46	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.1	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
4	4.26	4	3.1	4	5.39	19	12.5
12	10.84	12	9.13	12	8.15	8	5.56
7	4.82	7	7.26	7	6.42	8	7.91
5	5.68	5	4.74	5	5.73	8	6.89

Property	Value
Mean of each x variable	9
Variance of each x variable	10
Mean of each y variable	7.5
Variance of each y variable	3.75
Correlation between each x and y variable	0.816
Linear regression line	$y = 3 + 0.5x$





# Myths & Challenges

Remember...

---



- “Essentially, all models are wrong, but some are useful.”
- “Remember that all models are wrong: The practical question is how wrong do they have to be to not be useful.”

George E. P. Box,

1919 - 2013

Professor Emeritus of Statistics at the University of Wisconsin



# Thank you

Contact information:

[open@sap.com](mailto:open@sap.com)

**openSAP**  
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 6 Unit 6: Data Science Applications and References





# Data Science Applications and References

## Applications of data science in SAP

---

<b>SAP Hybris Marketing</b>	Campaign optimisation, product recommendations, segmentation...
<b>SAP Predictive Maintenance and Service</b>	Anomaly detection, rule mining, survival analysis, failure prediction...
<b>SAP Fraud Management</b>	Anomaly detection, rule discovery
<b>Unified Demand Forecast</b>	Modeling and forecasting sales
<b>SAP Integrated Business Planning</b>	Demand sensing and multistage inventory optimization
<b>SAP Digital Boardroom</b>	Revenue forecasting
<b>SAP Transportation Resource Planning</b>	Resource optimisation allocation
<b>SAP Assortment Planning for Retail</b>	Optimization of product selection
<b>SAP S/4HANA Finance</b>	Cash management, liquidity forecast
<b>SAP Real-Time Offer Management</b>	Self-learning real-time recommendation engine
<b>Big Data Analytics</b>	Forecasting marketing opportunities, revenue, opportunity scoring...

# Data Science Applications and References

## Applications of data science by SAP customers



**VELUX: From Annual Plan to Rolling Forecasts with SAP® Business Planning and Consolidation powered by SAP HANA®**



**Thélem assurances: Founding a Customer-Oriented Culture with SAP® Predictive Analytics**



**Covenant Transportation Group: Reducing Driver Turnover and Building a Safer Workforce with SAP® Predictive Analytics**



**Validating Sales Forecasts to Improve Production Planning with SAP® Predictive Analytics**

“With SAP Predictive Analytics, we can better analyze sales demand forecasts and market fluctuations. This helps us meet our customer needs quickly without overstocking our inventory. The result is happier customers and lower costs – which is great news for our bottom line.”

Tijani Ben Said, Manager Production Planning and Control, EMS-GRILTECH, EMS-CHEMIE AG



**mBank: Delivering a Personalized Banking Experience for 4.5 Million Customers with SAP® Predictive Analytics**

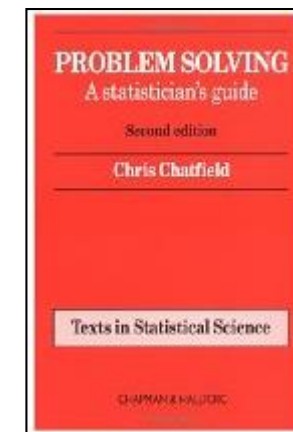
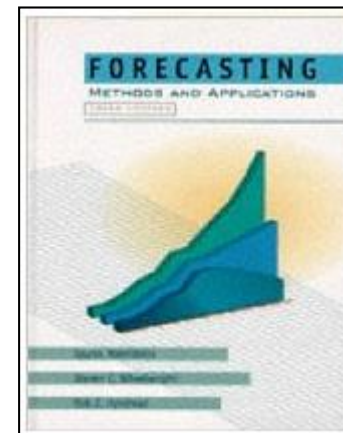
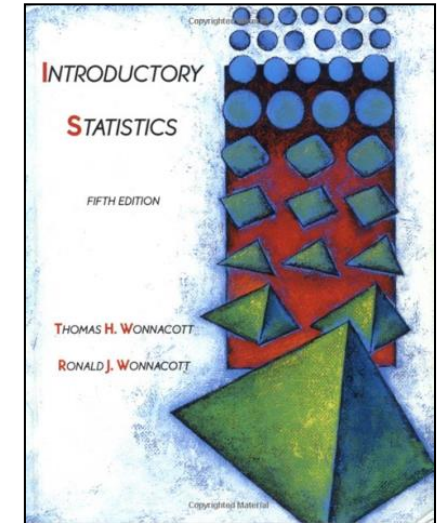
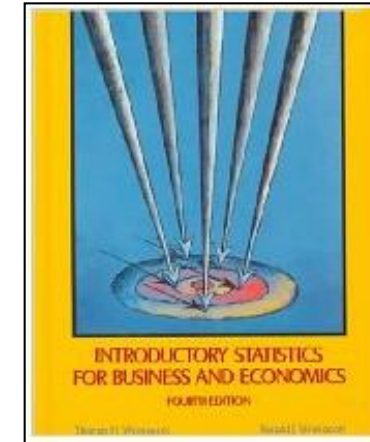
“SAP Predictive Analytics has allowed mBank to discover individual customer preferences and identify the next best activity for our marketing efforts. Now we are able to initiate more direct conversations, resulting in a better understanding of our clients on a personal level.”

Bartosz Witorzenc, Strategic Initiatives Manager, Retail Banking Department, mBank S.A.

# Data Science Applications and References

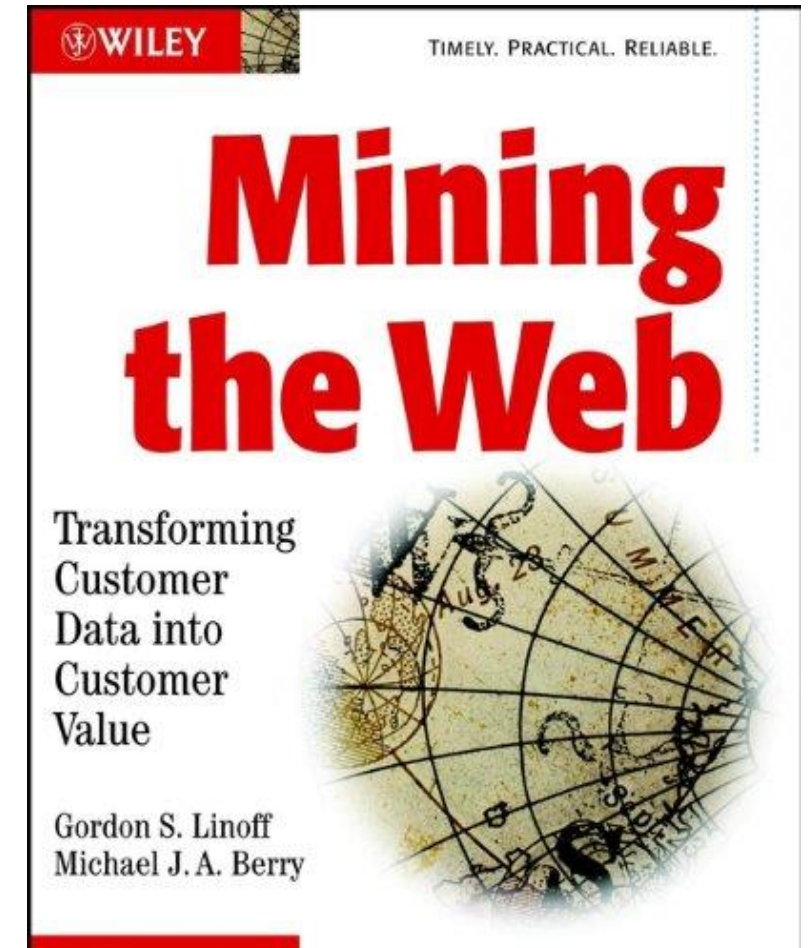
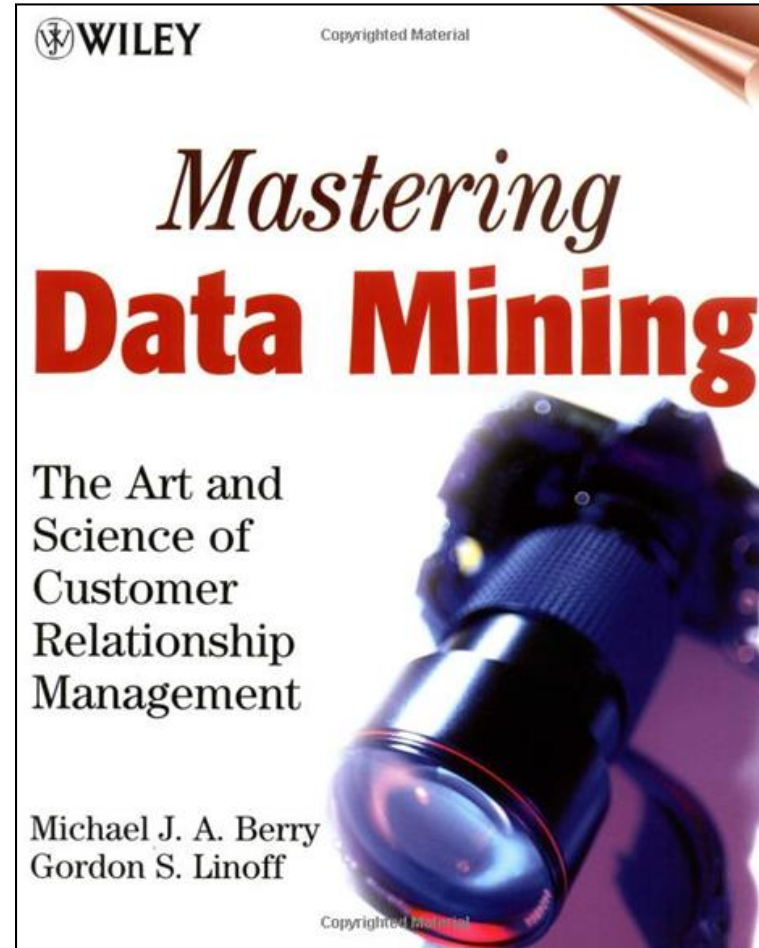
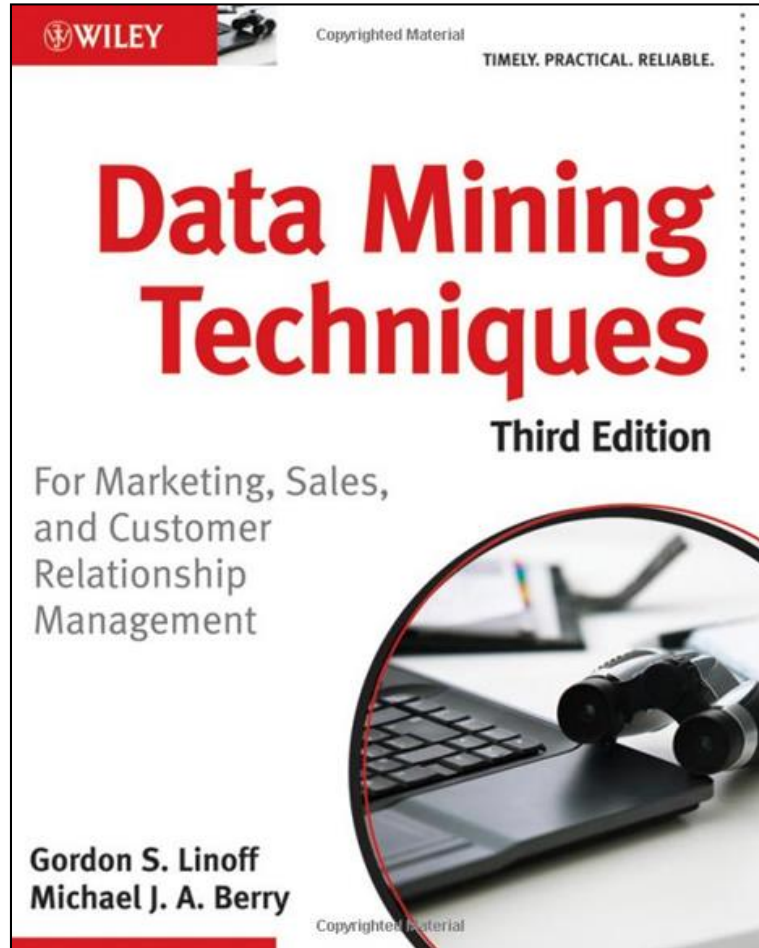
## References – Statistics

- Introductory Statistics For Business and Economics  
T.H. Wonnacott & R.J. Wonnacott
- Forecasting – Methods and Applications  
Makridakis, Wheelwright, & McGee
- Problem Solving: A Statistician's Guide  
Christopher Chatfield
- Introductory Statistics  
T.H. Wonnacott and R.J. Wonnacott



# Data Science Applications and References

## References – Data mining

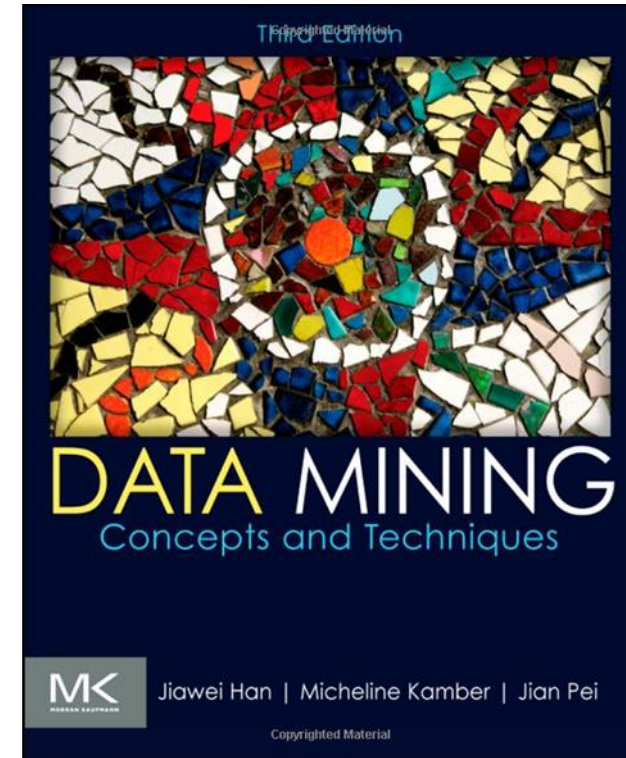
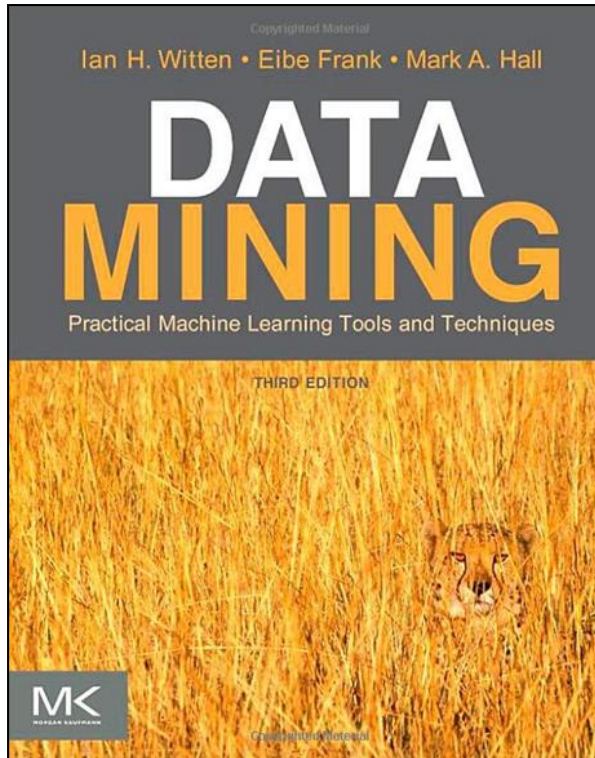
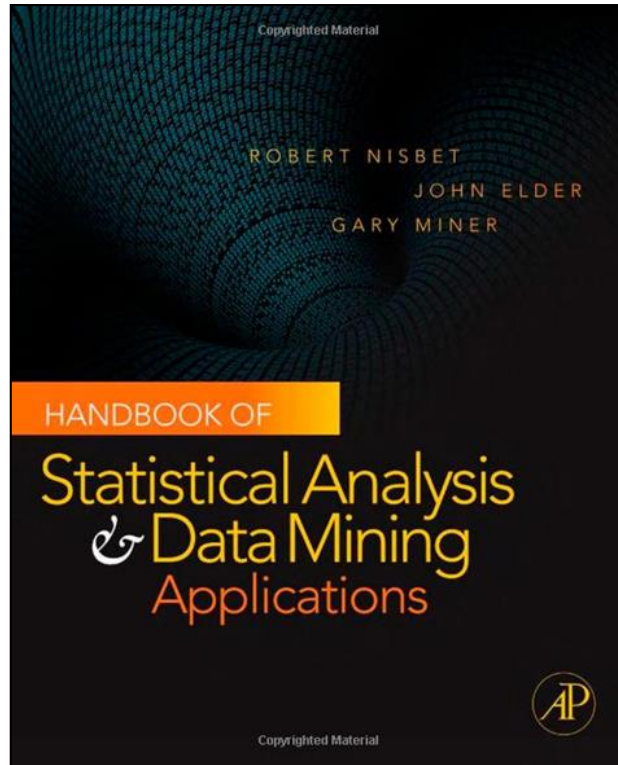




# Data Science Applications and References

## References – Data mining

---

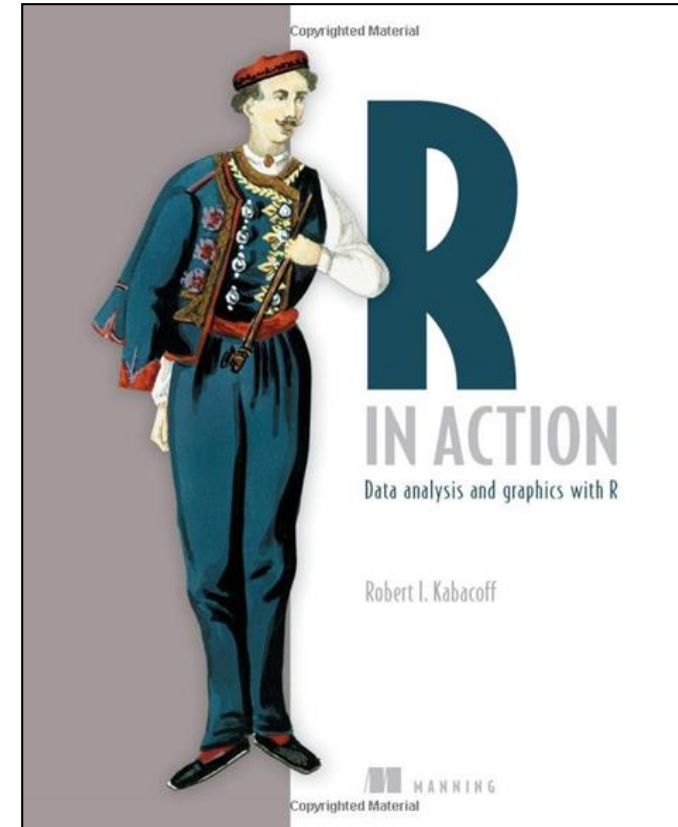
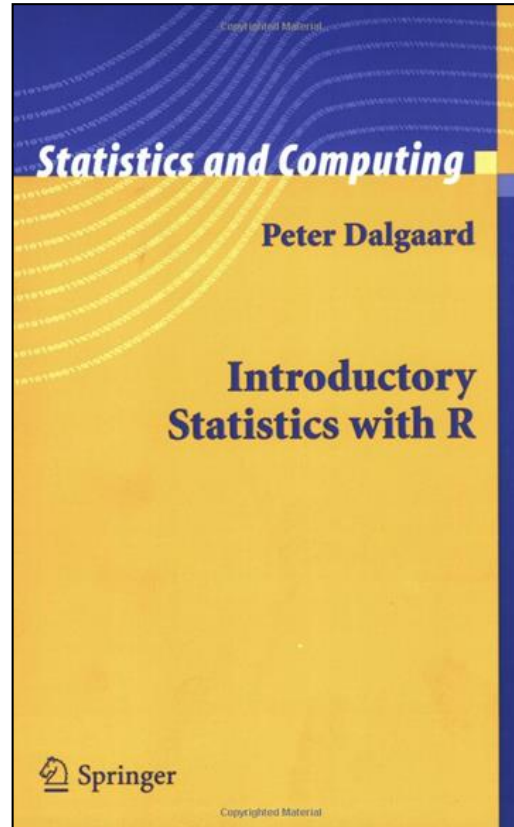
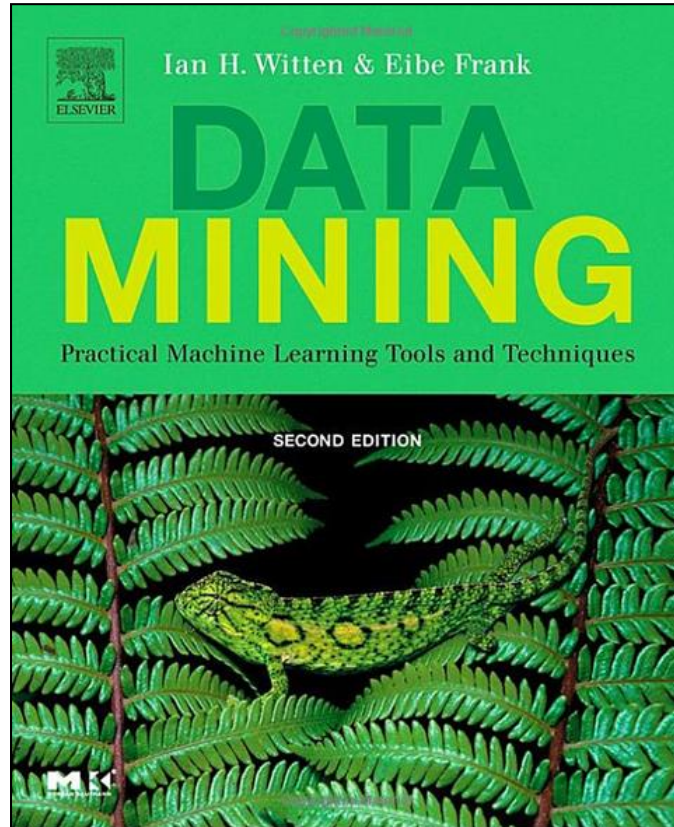




# Data Science Applications and References

## References – Data mining and R

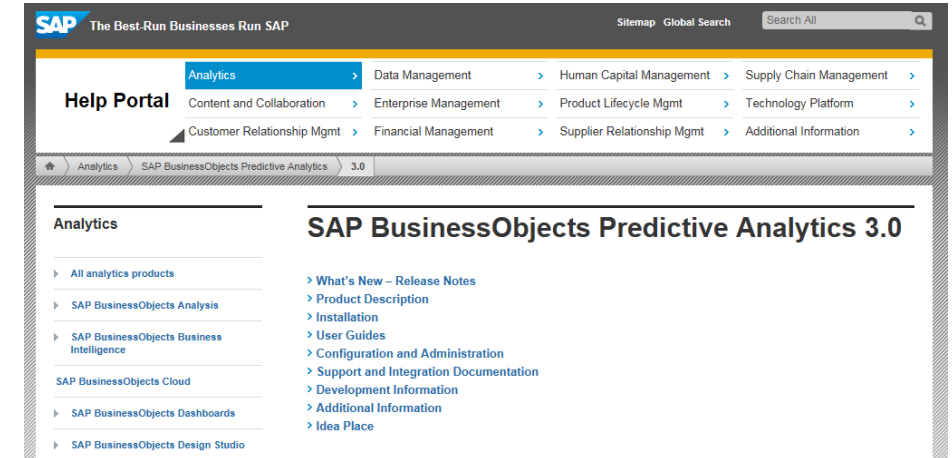
---



# Data Science Applications and References

## References – SAP predictive analytics products

- SAP BusinessObjects Predictive Analytics 3.0
  - <http://help.sap.com/pa>
- SAP HANA Predictive Analysis Library (PAL)
  - [http://help.sap.com/hana/SAP\\_HANA\\_Predictive\\_Analysis\\_Library\\_PAL\\_en.pdf](http://help.sap.com/hana/SAP_HANA_Predictive_Analysis_Library_PAL_en.pdf)
- Roadmaps on SAP Service Marketplace
  - <http://service.sap.com/saproadmaps>
- SAP predictive analytics information and 30 day free trial
  - <https://go.sap.com/solution/platform-technology/analytics/predictive-analytics.html>



SAP HANA Platform SPS 12  
Document Version: 1.1 – 2016-09-13

PUBLIC

**SAP HANA Predictive Analysis Library (PAL)**

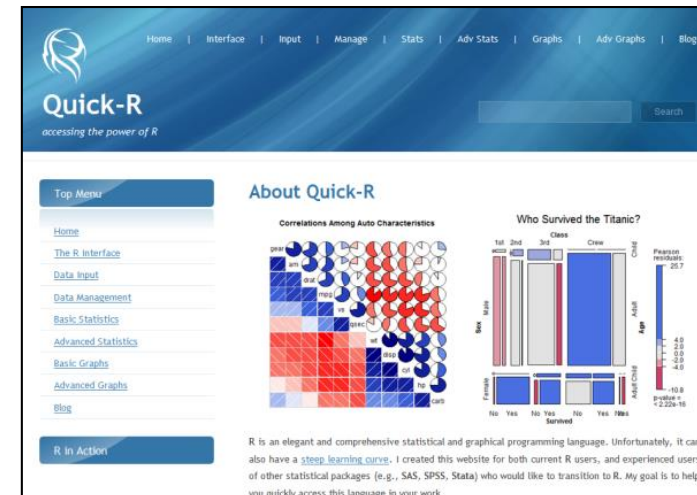
# Data Science Applications and References

## References – R

- SAP HANA R Integration Guide - [http://help.sap.com/hana/SAP\\_HANA\\_R\\_Integration\\_Guide\\_en.pdf](http://help.sap.com/hana/SAP_HANA_R_Integration_Guide_en.pdf)



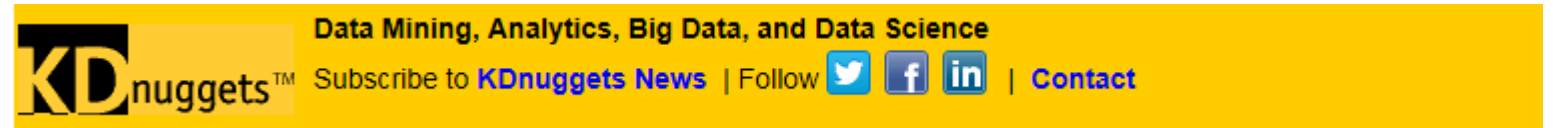
- The R manual - <http://cran.r-project.org/doc/manuals/R-intro.html>
- For beginners:
  - <http://www.statmethods.net/>
  - <http://www.ats.ucla.edu/stat/R/>
  - <http://www.r-bloggers.com/>
  - <http://www.inside-r.org/r-resources-web>



# Data Science Applications and References

## References – Web resources


- KDnuggets
  - [www.kdnuggets.com/](http://www.kdnuggets.com/)
- Data Science Central
  - [www.datasciencecentral.com/](http://www.datasciencecentral.com/)
- analyticbridge
  - [www.analyticbridge.com/](http://www.analyticbridge.com/)
- To learn more about PMML
  - <http://dmg.org/>



THE DATA MINING GROUP (DMG) IS AN INDEPENDENT,  
VENDOR LED CONSORTIUM THAT DEVELOPS DATA MINING  
STANDARDS




# Data Science Applications and References

## KDnuggets

**Data Mining, Analytics, Big Data, and Data Science**

search KDnuggets

Search

Subscribe to **KDnuggets News** | Follow    | [Contact](#)

[SOFTWARE](#) | [NEWS](#) | [Top stories](#) | [Opinions](#) | [Tutorials](#) | [JOBS](#) | [Academic](#) | [Companies](#) | [Courses](#) | [Datasets](#) | [EDUCATION](#) | [Certificates](#) | [Meetings](#) | [Webinars](#)

### Latest News, Stories

- [Process Mining: Where Data Science and Process Science...](#)
- [KDD 2016: Watch Talks by Top Data Science Researchers](#)
- [Deep Learning Reading Group: Skip-Thought Vectors](#)
- [Why Fluid Intelligence Makes You a Better Data Scientist](#)
- [Top tweets, Nov 9-15: #Trump, limits of #prediction; ...](#)

More [News & Stories](#) | [Top Stories](#)


### Search results for top books

Found 99 documents, 5904 searched:

- **Top 10 Essential [Books](#) for the Data Enthusiast**  
...not to mention the myriad tools for automating the various aspects of our professional lives which seem to pop up on a daily basis. There are a lot of topics to keep abreast of. Fortunately (unfortunately?) there is no shortage of [books](#) available on all of these subjects. There are a lot of lists av...  
<http://www.kdnuggets.com/2016/04/top-10-essential-books-data-enthusiast.html>
- **Top 10 Amazon [Books](#) in Data Mining, 2016 Edition**  
The recent explosion of interest in data science, data mining, and related disciplines has been mirrored by an explosion in book titles on these same topics. One of the best ways to decide which [books](#) could be useful for your career is to look at which [books](#) others are reading. This post details th...  
<http://www.kdnuggets.com/2016/11/top-10-amazon-books-data-mining.html>
- **KDnuggets™ News 16:n41, Nov 16: Top 10 Amazon [Books](#) in Data Mining; Intuitive Explanation of Convolutional Neural Nets**  
[Features](#) | [Software](#) | [Tutorials](#) | [Opinions](#) | [News](#) | [Webcasts](#) | [Meetings](#) | [Jobs](#) | [Tweets](#) | ... and ..... don't forget to check .... Image of the week [Features Top 10 Amazon \[Books\]\(#\) in Data...](#)  
<http://www.kdnuggets.com/2016/n41.html>

## Visual Statistics

Your data.  
Powerful predictive modeling.




Join Us and Discover What Open Data Science Means

AUSTIN, TX | 2017

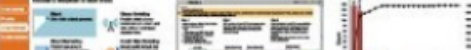
REGISTER NOW

**AnacondaCON 2017 conference for Open Data Science leaders, Austin, Feb 7-9**



**rapidminer**  
DAILY DEMO

Watch a Quick Demo Followed by Live Q&A with a RapidMiner Data Scientist





# Data Science Applications and References

## Top 10 books recommended by KDnuggets

1. **The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd Edition**  
Trevor Hastie, Robert Tibshirani, Jerome Friedman  
4.1 out of 5 stars ([78 reviews](#))
2. **Data Science from Scratch: First Principles with Python, 1st Edition**  
Joel Grus  
4.2 out of 5 stars ([65 reviews](#))
3. **Data Science for Business: What You Need to Know About Data Mining and Data-Analytic Thinking, 1st Edition**  
Foster Provost, Tom Fawcett  
4.6 out of 5 stars ([152 reviews](#))
4. **Data Analytics Made Accessible**  
Anil Maheshwari  
4.8 out of 5 stars ([175 reviews](#))
5. **Python Machine Learning**  
Sebastian Raschka  
4.3 out of 5 stars ([78 reviews](#))



# Data Science Applications and References

## Top 10 books recommended by KDnuggets

---

6. Data Smart: Using Data Science to Transform Information into Insight, 1st Edition

John W. Foreman

4.7 out of 5 stars ([105 reviews](#))

7. Hadoop: The Definitive Guide, 4th Edition

Tom White

4.6 out of 5 stars ([41 reviews](#))

8. R in Action: Data Analysis and Graphics with R, 2nd Edition

Robert Kabacoff

4.8 out of 5 stars ([33 reviews](#))

9. Doing Data Science: Straight Talk from the Frontline, 1st Edition

Cathy O'Neil, Rachel Schutt

4.0 out of 5 stars ([50 reviews](#))

10. Big Data: Principles and best practices of scalable realtime data systems, 1st Edition

Nathan Marz, James Warren

4.4 out of 5 stars ([33 reviews](#))



# Thank you

Contact information:

[open@sap.com](mailto:open@sap.com)

**openSAP**  
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.