Joerg Niessing
Affiliate Professor of Marketing

[Big]-Data Analytics for Businesses

Understand the world. Expand your world.

Today's plan

1. Course introduction
2. Example applications
3. Data analytics processes
4. Introduction to key tools

Our three goals

1. "Big picture": Develop your intuition about identifying data analytics opportunities and their implementation challenges
2. "Medium picture": learn how to approach data analytics projects
3. "Dirty hands": Learn how to perform, read, and use key data analytics methods.

Our three goals

1. "Big picture": Develop your intuition about identifying data analytics opportunities and their implementation challenges
2. "Medium picture": learn how to approach data analytics projects
3. "Dirty hands": Learn how to perform, read, and use key data analytics methods.

What is big data? What is new?

… It is not only about technology but business value (WIN-WIN)

Where have you used data analytics? …or would like to use in the future?

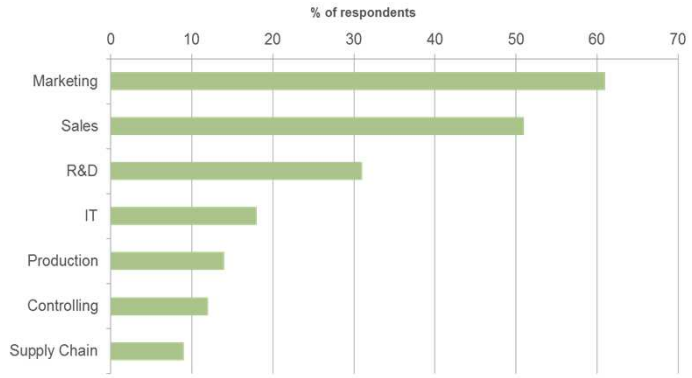
Many uses

- Marketing (i.e. acquisition, cross-sell, churn, etc)
- Personalization (i.e. targeted advertisement)
- Market positioning
- Fraud detection
- Monitoring of processes
- Credit scoring and risk management
- Supply chain optimization
- Financial modeling
- Knowledge management and text mining
- ……

Where

Marketing and **sales** are the key areas for big data…

Where do organizations plan to use big data in 1-3 years?

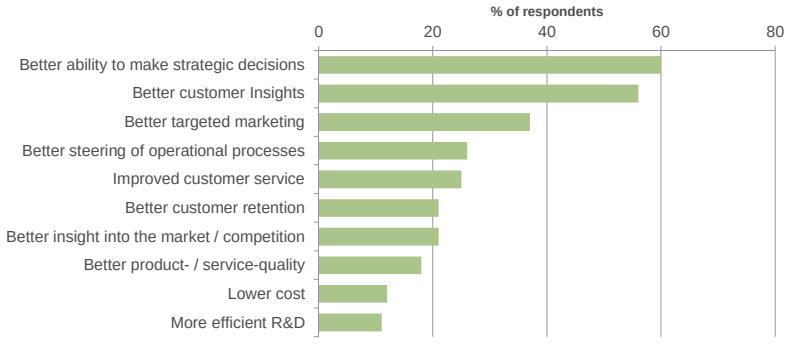


9

Why

Making **better decisions** and getting to **better customer insights** are the main benefits of big data technologies and analysis…

The main benefits of Big Data technologies



10

Why

The business value of big data: companies already using big data to make decisions show a **competitive edge**…

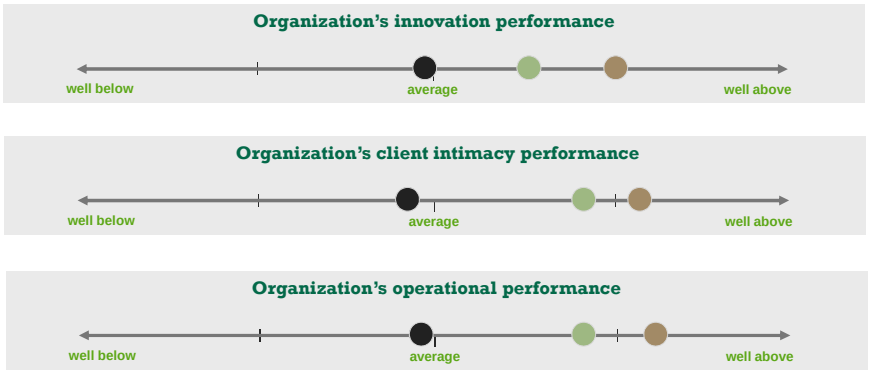
Which of the following statements best describes your organization's stage in using big data to help make business decisions?

- = currently not using big data to inform business decisions
- = already using big data to inform business decisions

11

Why

…and companies that are also using big data more **efficiently** are outperforming their peers even more



How efficient is your organization in internally **sharing and reusing** data analytics knowledge and research?

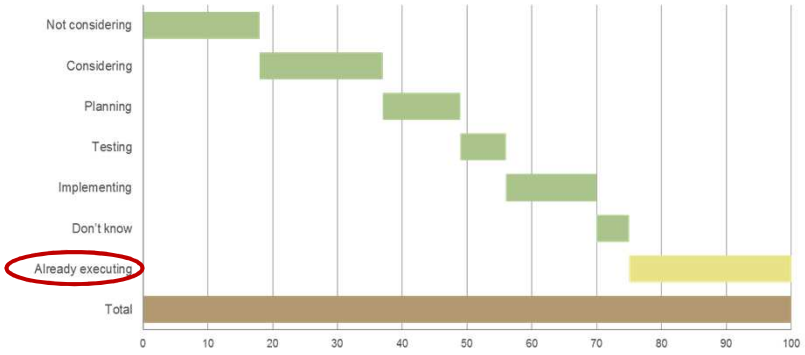
- = (somewhat) inefficient in sharing and reusing data analytics
- = average
- = (somewhat) efficient in sharing and reusing data analytics

12

How

… but 75% of the companies participated in the study are still in early stages…

Companies already using big data

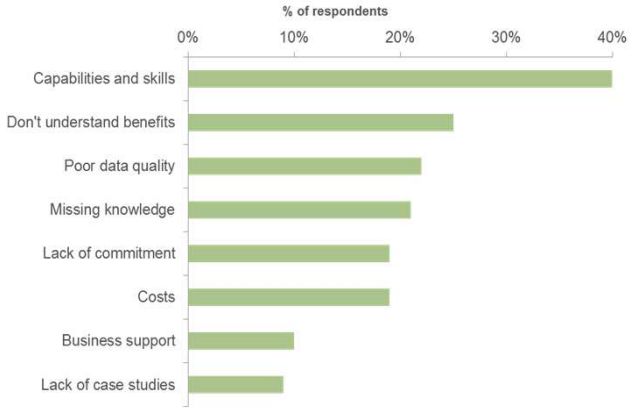


13

How

Missing **capabilities and skills** are the key reason why organizations do <u>not</u> use big data…

Key reasons why organizations are <u>not</u> considering or further exploring the use of big data

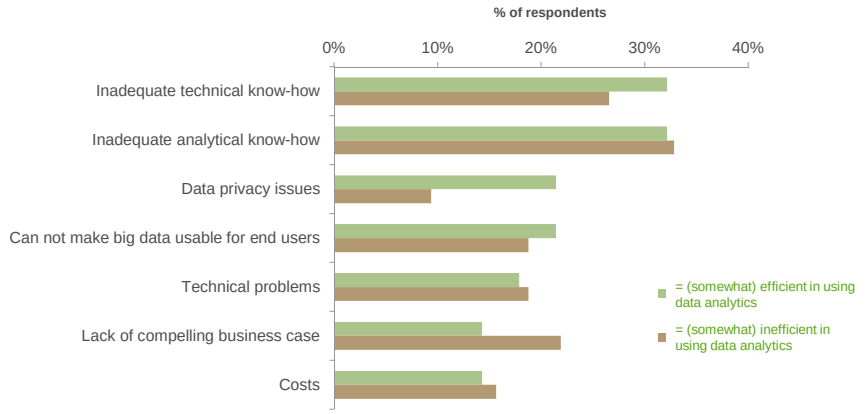


14

How

…but analyzing data effectively is also challenging….

The main challenges when using big data

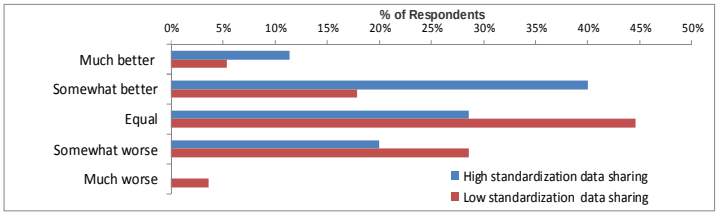


15

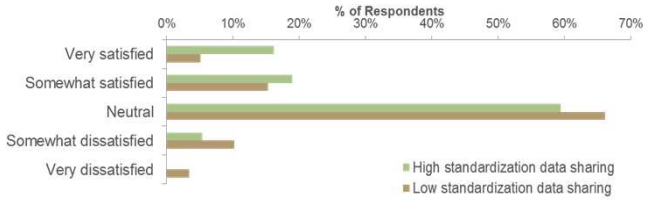
How

To bring value, data analytics need organizational enablers

Quality of data analytics compared to biggest competition



Satisfaction with ROI of big data



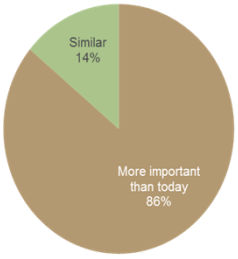
16

Next

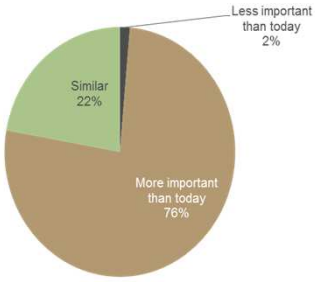
Companies already using big data are saying that it will become even more important for their business in the future

Relative importance of data analysis and data reporting in the next 1-3 years for your business

Companies already USING big data to inform business decisions



Companies currently NOT USING big data to inform business decisions

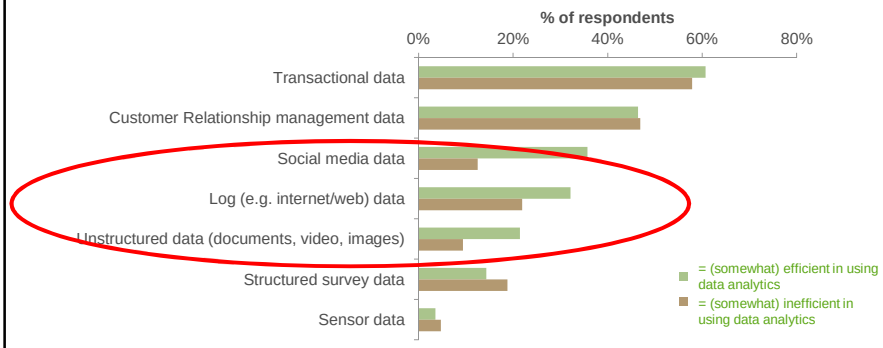


17

Next

There is a big potential in understanding unstructured data…

The main types of data analyzed



BUT

- 30% analysed data from just ONE source
- Over 50% analysed data from TWO source's
- Less than 20% analysed data from MORE THAN TWO source's

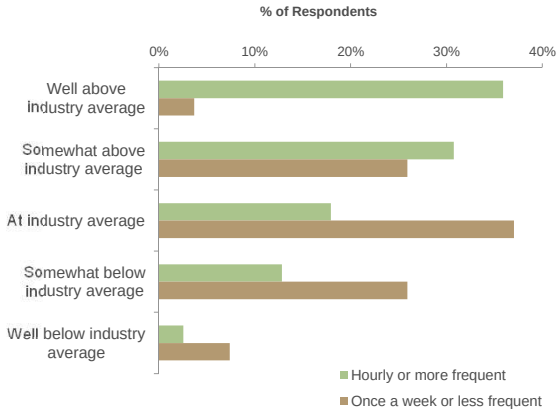
18

The business value of big data: it is not only about having large amounts of data but mainly about **analyzing data fast**….

Next

Frequency of data analysis…

…and innovation performance



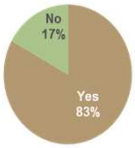
19

Next

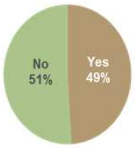
Companies that are already efficient in using big data will leverage new technologies more often in the future

Use of **cloud based** data analytics technologies the next 1-3 years

Companies already EFFICIENT in using big data

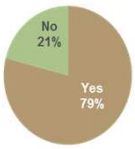


Company currently NOT EFFICIENT in using big data

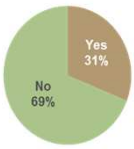


Use of **open source** data analytics technologies the next 1-3 years

Companies already EFFICIENT in using big data



Company currently NOT EFFICIENT in using big data



20

Summary of key pre-study findings

- Companies already using data analytics to make decisions show a **competitive edge** and could outperform their peers even more if **sharing and reusing** data analytics more efficiently
- **Marketing** and **sales** will still be the key areas of use
- **75%** of the companies are still in early stages
- Analyzing data effectively is **challenging** (e.g. lack of analytical of technical skills, lack of compelling business cases for investing in big data technologies, still data quality issues, etc.)
- Data analytics need organizational **enablers**. Two key enabler are **skills** and **standardization of data sharing**
  - The combination of high frequency of analysis and high data standardization is good for knowledge and innovation
- There is a big potential in understanding **unstructured** data
- **Cloud** and **open source** are expected to rise

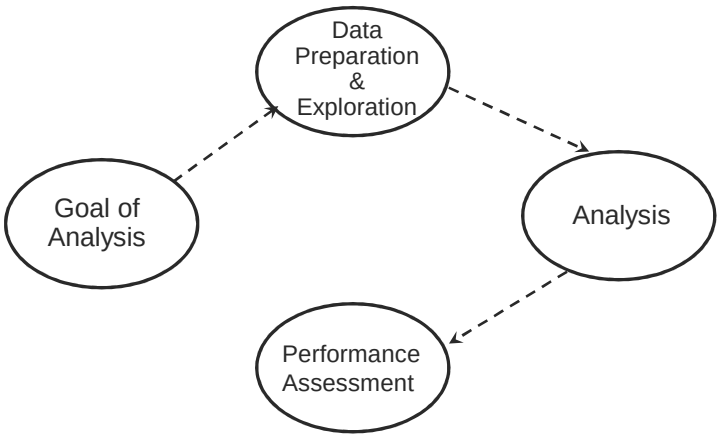
21

Our three goals

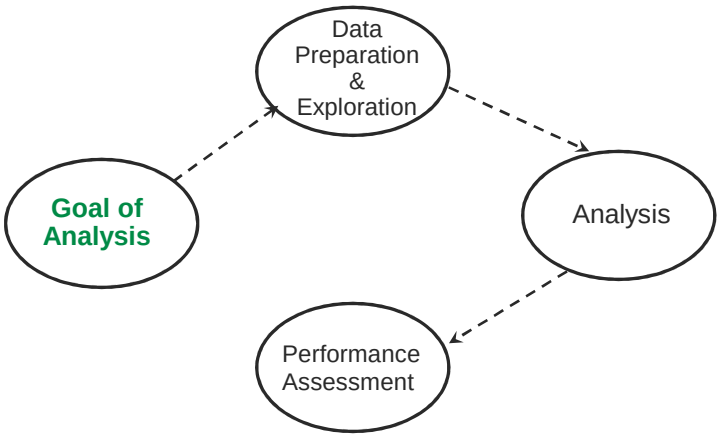
1. "Big Picture": Develop your intuition about identifying data analytics opportunities and their implementation challenges
2. "Medium Picture": learn how to approach data analytics projects
3. "Dirty Hands": Learn how to perform, read, and use key data analytics methods.

How could you use data for churn management?

The *Process*



The *Process*



Step 1: Goal of Analysis

A lot of good statistical analysis is directed at solving the wrong business problem.

- What Is the business objective?
- Example: Telco Churn
- Initial problem: Assign a churn score to all customers.
  - Recent customers with little call history? Elite customers?
  - Telephones? Individuals? Families?
  - Churn any time or within a time range?
- How will the results be used?
- Better objective: By September 24, provide a list of the 10,000 elite customers who are most likely to churn in October.

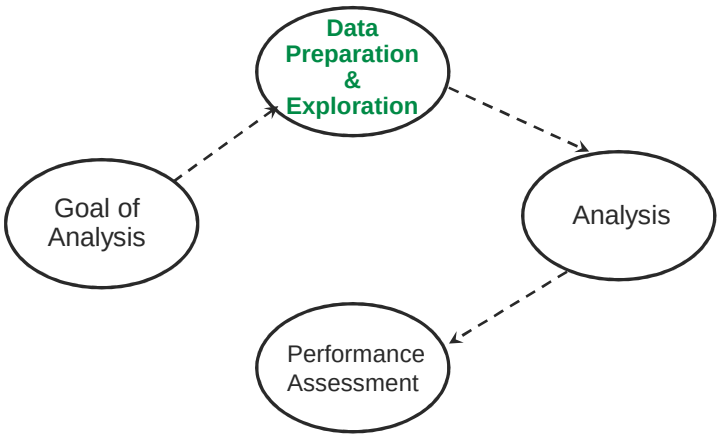
The new objective is actionable.

Step 1: Goal of Analysis

- Improve the response rate for a direct marketing campaign.
- Increase the average order size.
- Determine what drives customer acquisition.
- Forecast the size of the customer base in the future.
- Choose the right message for the right groups of customers.
- Target a marketing campaign to maximize incremental value.
- Recommend the next, best product for existing customers.
- Segment customers by 'WHAT(!)'

A lot of good statistical analysis is directed at solving the wrong business problem.

The *Process*



Step 2: Data Preparation and Exploration

- What is available?
- What is the right level of granularity?
- How much data is needed?
- What data quality issues do we have?
- How much history is required?
- Which variables should be used?
- What derived variables do we need?
- Is our data representative?

Representativeness of the Data Sample

The model set might not reflect the relevant population.

- Customers differ from prospects.
- Survey responders differ from non-responders.
- People who read e-mail differ from people who do not read e-mail.
- Customers who started three years ago might differ from customers who started three months ago.
- People with land lines differ from those without.

Availability of Relevant Data

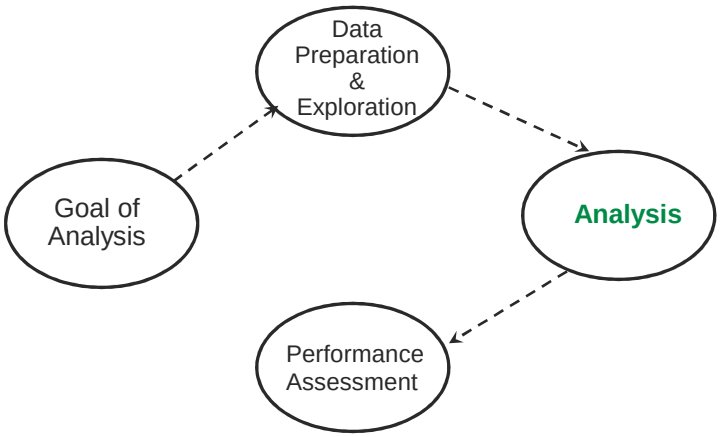
Poor coverage might be the number one reason for wireless subscribers canceling their subscriptions, but data about dropped calls is not in billing data.

Customers might already have potential cross-sell products from other companies, but that information is not available internally.

Key Idea: Ask the Data MANY Questions

- Why were some customers active for 31 days in February, but none were active for more than 28 days in January?
- How do some retail card holders spend more than $100,000 in a week in a grocery store?
- Why were so many customers born in 1911? Are they really that old?
- Why do Safari users never make second purchases?
- What does it mean when the contract **begin** date is after the contract **end** date?
- Why are there negative numbers in the sale price field?
- How can active customers have a non-null value in the cancellation reason code field?

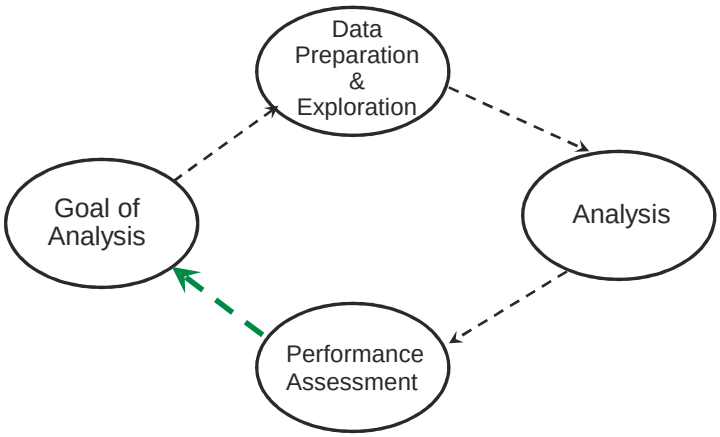
The *Process*



Step 2: Analysis

- What approach do I need to use?
- What is the right analytical method?
- Do I need complex tools?
- What is the benchmark approach?
- Plot the data and compare!
- Digest and interpret!
- What is relevant?
- Association vs. causation!
- Short vs. long-term!

The *Iterative Process Cycle*



This is an ITERATIVE process!

- Revisit business objectives.
- Define new objectives.
- Gather and evaluate new data.

Example:

- A model discovers that geography is a good predictor of churn.
  - What do the high-churn geographies have in common?
  - Is the pattern your model discovered stable over time?

Why this class? Our three goals

1. "Big Picture": Develop your intuition about identifying data analytics opportunities and their implementation challenges
2. "Medium Picture": learn how to approach data analytics projects
3. "Dirty Hands": Learn how to perform, read, and use key data analytics methods

Basic Types of Questions and Tools

1. Market Basket Analysis: which pairs of products are typically sold together? – "On Friday evenings, shoppers who buy diapers also buy beer".
2. Factor Analysis: Finding important dimensions ("factors") that summarize your data, and visualizing your data
3. Clustering: What are the main types of customers we have?
4. Regression Modeling: What are variables that drive a specific outcome?
5. Discriminant Analysis: How can we differentiate between the "high value" and "low value" customers?

Three key-takeaways

- Data is not big – it is diverse
- Organizational enablers (e.g. skills, standardization) are key for success
- Creative analytics require an iterative process

Course grading

- Group assignment (part A & B): 50%
- Group project: 20%
- Class participation: 30%

To do for next time

- Download and install SPSS (trial version – from course website)
- Download and explore the Coffee data
- Read the "Boating Case A"

INSEAD

The Business School for the World®