Data Analytics in Organisations and Business

Dr. Isabelle Flückiger

E-mail: isabelle.flueckiger@math.ethz.ch

Data Analytics in Organisations and Business

Some organisational information:

Tutorship:
Gian Thanei: thanei@stat.math.ethz.ch
Sylvain Robert: robert@stat.math.ethz.ch

Beginning of lecture: Wednesday, 16th of September.
Beginning of exercises: 23rd of September (bi-weekly)
Room: HG D7.1

Lecture notes: https://stat.ethz.ch/education/semesters/as2015/analytics

Data Analytics in Organisations and Business

Content

1 Introduction

- 1.1 What is (Data) Analytics?
- 1.2 What is this Lecture about? - ...and what it is not . .

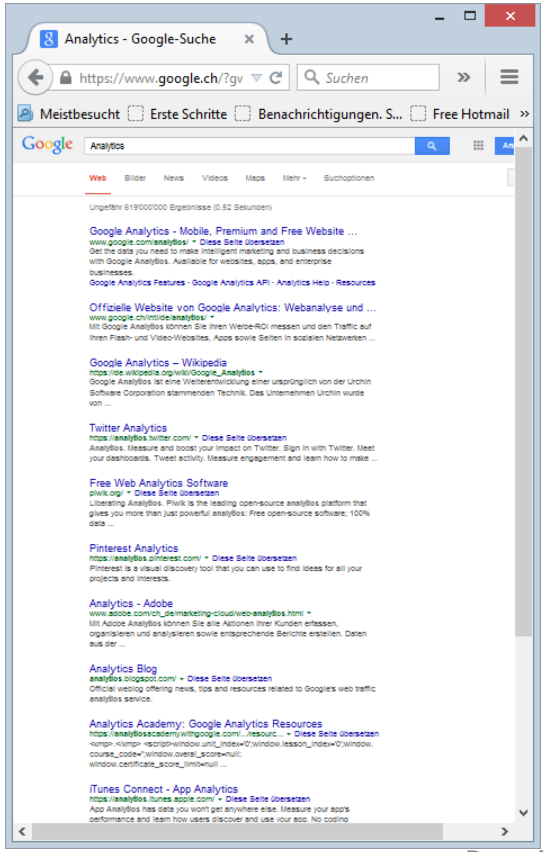
2 Framing the Business Problem

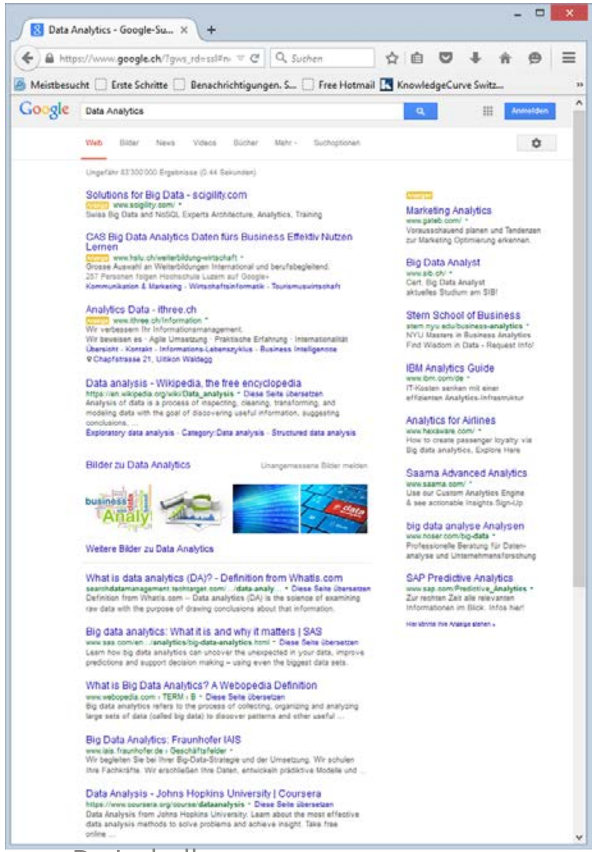
- 2.1 Content of this Chapter
- 2.2 Obtain or work out the description of the business problem and what should be the usability
- 2.3 Identifying all stakeholders i.e. all direct or indirect stakeholders
- 2.4 Analyse whether the business problem is amenable to an analytics solution
- 2.5 Refinement of the problem statement and if necessary depict known or possible constraints
- 2.6 Determine the business benefits
- 2.7 Obtain stakeholder agreement on the business problem statement

Chapter 1 Introduction

1.1 What is (Data) Analytics?

Google search: «Analytics» & «Data Analytics» (619'000'000 results in 0.52 seconds):





1.1 What is (Data) Analytics?

Wikipedia says about Analytics: “*Analytics is the discovery and communication of meaningful patterns in data. Especially valuable in areas rich with recorded information, analytics relies on the simultaneous application of statistics, computer programming and operations research to quantify performance. Analytics often favours data visualization to communicate insight.*” [Wikipedia, 2015]

1.1 What is (Data) Analytics?

Many similar terms used, and often the users of such expressions have similar but not the same understanding of their meanings:

- Predictive Analytics
- Data Mining
- Advanced Analytics
- Business Analytics
- Web or Online Analytics
- Big Data Analytics
- Data Analysis

1.1 What is (Data) Analytics?

But why Data Analytics becomes increasingly more and more important?

In 2015 there will be data generated of

8500 exabytes

and 1 exabyte is 1 billion gigabytes, i.e. 10^{18} bytes.

And in 2040 there will be data generated of

40'000 exabytes.

1.1 What is (Data) Analytics?

Three facts about data and analytics.

1. The amount of data cannot anymore be grasped with the human brain nor proceeded for extracting the relevant information.
2. And today's business world is changing from making decisions based on knowledge and intuition to so-called fact-based decisions.
3. The reason is that on one side the world becomes more global and more complex networked but on the other side there is an ongoing decentralisation of information and data storage.

1.1 What is (Data) Analytics?

A definition of (Data) Analytics:

Definition 1. Analytics is the extensive use of data, statistical and quantitative analysis, explanatory and predictive models, and fact-based management to drive decisions and add value [Davenport and Harris, 2007].

1.1 What is (Data) Analytics?

Classes of analytics:

- Descriptive
- Predictive
- Prescriptive

Classification	Questions	Examples from Business
Prescriptive	What is the best outcome? What if?	Optimisations Scenario testing Randomised tests
Predictive	What could happen? What is happening next? Why is this happening?	Statistical modelling Forecasting
Descriptive	What happened? How many, how often? What action is needed?	Standard and ad hoc reports Queries Alerts

1.1 What is (Data) Analytics?

Descriptive Analytics

- Describes what happened in the past
- Contains of gathering and organising data, plotting the data and giving characteristics
- Used for classifying groups of data e.g. customer groups
- Historically used for reporting and to show how and organisation were performing
- Says nothing about the cause why something happened nor what could happen in the future

1.1 What is (Data) Analytics?

Predictive Analytics

- Uses models and data from the past to forecast the future
- Associations among the variables are identified and then the dependent variable is forecasted e.g. how many customer would buy
- In predictive analytics we do not necessarily assume causal effects
- In fact, causal effects are not always necessary to predict accurately certain behaviour

Example: A grocery store found out that women who stopped buying cookies will be lost as a customer within the next three month

1.1 What is (Data) Analytics?

Prescriptive Analytics

- Gives actions to perform
- It includes experimental design and optimisation.
  - Experimental Design makes causal inference by conducting experiments for answering questions of WHY something happened.
  - Optimisation wants to achieve the optimised level of a particular variable in its relationship to another variables

Example: Determination of the price of a product which leads to the highest profitability, highest margin or highest market share

1.1 What is (Data) Analytics?

Analytics can typically be further classified in

- Qualitative and
- Quantitative

Data Analytics.

1.1 What is (Data) Analytics?

Qualitative Analysis

- Has the purpose gaining an understanding of the underlying (qualitative) reasons or motivations for a behavior
- Has the goal to gain insight in causal effects from a behavioural perspective
- Includes collection of unstructured data of some “small” and “non-representative” samples which are “analysed” non-statistically
- Typically used as part of exploratory research in the earliest stage of an analytics process

1.1 What is (Data) Analytics?

Quantitative Analysis

- Is the systematic empirical investigation of a phenomena by statistical, mathematical or computational methods
- Contrary to a qualitative analysis, data are collected in structured manner out of a large representative sample and analysed

1.1 What is (Data) Analytics?

Different types of analytical methods [Davenport and Harris, 2007]

- *Statistics*: The science of collection, organisation, analysis, interpretation and presentation of data.
- *Forecasting*: The estimation of some variable of interest at some specified future point in time as a function of past data.
- *Data Mining*: The automatic and semiautomatic extraction of previously unknown, interesting patterns in large quantities of data through the use of computational and statistical techniques.
- *Text Mining*: The process of deriving patterns and trends from the text in manner similar to data mining.
- *Optimisation*: The use of mathematical techniques to find optimal solutions with regard to some criteria while satisfying constraints.
- *Experimental Design*: The use of test and control groups, with random assignment of subject or cases to each group, to elicit the cause and effect relationship in a particular outcome.

1.1 What is (Data) Analytics?

Small Data vs Big Data

Examples:

- Twitter produces 15 terabytes data per day ((1 terabyte = 1 TB = 10^{12} bytes = 1000 gigabytes)
- Google can proceed 1 petabyte of data each hour (1 petabyte = 1 PB = 10^{15} bytes = 1000 terabytes)

1.2 What is this Lecture about? - ...and what it is not

If you are looking for a lecture where you can learn “fancy” techniques for process data, then you are in the wrong place!

1.2 What is this Lecture about? - ...and what it is not

What you will learn is this lecture

- How to conduct data analytics projects in the real world with many people who are not experts in that area but who are the buyer of such services
- That in data analytics projects the rule of thumb is that 20% is data analysis and the other 80% is something else whereas
    - this consists of project management,
    - finding out the underlying problem which has to be analysed,
    - stakeholder management,
    - explaining to non-experts what and why
    - search for, discover and cleansing data
- Often the "real" problem which has to be solved is originally not known but has to be defined
- How to deal with data which are somewhere and somehow available

1.2 What is this Lecture about? - ...and what it is not

…and a lot of Case Studys

…and become familiar with all the business terms and phrases used

1.2 What is this Lecture about? - ...and what it is not

Thus, the lecture will contain:

1. How to frame the business problem
2. How to transfer it to a problem which can be solved with analytics methods
3. Data identification and prioritisation, data collection and data harmonisation
4. Identification of problem solving approaches and appropriate tools (not only R even though this is important)
5. How to set up and validate models
6. The deployment of a model
7. Model lifecycle
8. Some words about soft skills needed by statistical and mathematical professionals

Chapter 2
Framing the Business Problem

2.1 Content of this Chapter

How to frame the business problem:

1. Obtain or work out the description of the business problem and what should be the usability
2. Identifying all stakeholders i.e. all direct or indirect stakeholders
3. Analyse whether the business problem is amenable to an analytics solution
4. Refinement of the problem statement and if necessary depict known or possible constraints
5. Determine the business benefits
6. Obtain stakeholder agreement on the business problem statement

2.2 Obtain or work out the description of the business problem and what should be the usability

Definition: A description of a business problem is a *business problem statement*.

This business problem statement contains a description about the business opportunity or threat, or an issue.

Example: “We are experiencing production problems and cannot deliver in time”

But this information is still insufficient to identify the full detailed problem.

2.2 Obtain or work out the description of the business problem and what should be the usability

To collect and structure this information in an understandable context: The five W's: who, what, where, when, and why.

- **Who** are the stakeholders who are sponsoring the project, who are using the results, who are making decisions based on the outcome and who are affected by the results?
- **What** problem has to be solved? What would be the perfect solution of that problem? What happen if the problem is not solved?
- **Where** does the problem occur? Where does the function requires to perform?
- **When** does the issue occur, or the function requires to be performed? When does the project need to be completed?
- **Why** does the problem occur, or function need to perform? Why this problem should be solved?

2.2 Obtain or work out the description of the business problem and what should be the usability

Example: A bank suffers the movement of customers.

Who are the stakeholder? Who is interested in this issue? Who is affected by solving this Problem?

- The management (CEO, CFO, and so on): sponsor or buyer of the analytics and decision makers
- Division managers: affected by project support and by client segment and product decisions, is interested in solving the issue
- Product managers: affected by project involvement and product decisions
- Client advisors: affected by project e.g. information gathering and by client segment and product decisions, is interested in solving the problem
- IT department: affected by project support e.g. data, affected by decisions which would affect the IT landscape
- Compliance officer: affected by the project e.g. sensitive information or if decisions are compliant with regulations, affected by implementation of solutions
- Risk management: affected by decisions and affected by setting up possible new risk management processes

2.2 Obtain or work out the description of the business problem and what should be the usability

Example cont'd: A bank suffers the movement of customers.

Why does the problem occur? Why this problem should be solved?

- Wrong products?
- Insufficient customer service?
- Too less diversified customer basis?
- Not an attractive brand?

If it will not be solved the bank is loosing more clients and thus, more revenues. The bank is losing profit.

2.2 Obtain or work out the description of the business problem and what should be the usability

Important:

The full understanding of the problem is the most important aspect and is guiding the whole analytics process.

2.2 Obtain or work out the description of the business problem and what should be the usability

How to guide the stakeholder to a problem definition and for characterising the problem?

1. What problem are we addressing?	
Why this question is important?	Having a clear understanding is key for success Users might not clearly understand what data anyalytics is There are unspoken and maybe unreasonable expectations
What is this question seeking?	Definition of success for the project List of project deliverables State and quantify elements of performance
Likely responses and meaning	Vague descriptions Widely divergent answers They probably do not know what they want
Follow-up Questions	What are the requirements? What are desired results What are measures of performance? What are the project deliverables How is success and failure determined?
Special considerations	Questions from people who are evaluating the project Expectations of the sponsor or buyer

2.2 Obtain or work out the description of the business problem and what should be the usability

How to guide the stakeholder to a problem definition and for characterising the problem? (cont’d)

3. How would you characterise the desired solution?	
Why this question is important?	Lack of understanding of the different solution possibilities Working together through this decision Avoidance to solve the wrong problem
What is this question seeking?	This answer drives the analytics process and methods used
Likely responses and meaning	Vague and confusing answers
Follow-up Questions	What are special features needed as outcome?
Special considerations	Most users have little knowledge about this aspect Look for indications that additional explanations are needed

2.2 Obtain or work out the description of the business problem and what should be the usability

How to guide the stakeholder to a problem definition and for characterising the problem? (cont'd)

6. What is good performance and what is bad performance?	
Why this question is important?	Just improving the performance does not mean success The cost - benefit ratio have to be considered
What is this question seeking?	To find the benefit of an organisation
Likely responses and meaning	Business answers vs technical answers
Follow-up Questions	Asking for quantitative performance measures Vague responses and room for interpretation leads to failure
Special considerations	Maybe the problem cannot be solved by data analytics

2.2 Obtain or work out the description of the business problem and what should be the usability

Review of previous analyses of the problem:

- All previous findings connected to this problem should be investigated
- Is helping to think about how the problem has been structured so far and how it should be newly structured

Important: Often your problem is not as unique as you think, and it is likely that many people have already done something similar