Marketing Effectiveness and Data Quality

Dr. Mathias Klier

University of Innsbruck
School of Management

Information Systems II

Universitätsstraße 15

A-6020 Innsbruck

Mathias.Klier@uibk.ac.at

Agenda

- Motivation
- A Probability-based Data Quality Metric for Currency
- Application of the Metric at a Mobile Services Provider
- Summary and Discussion

Project at a Mobile Services Provider

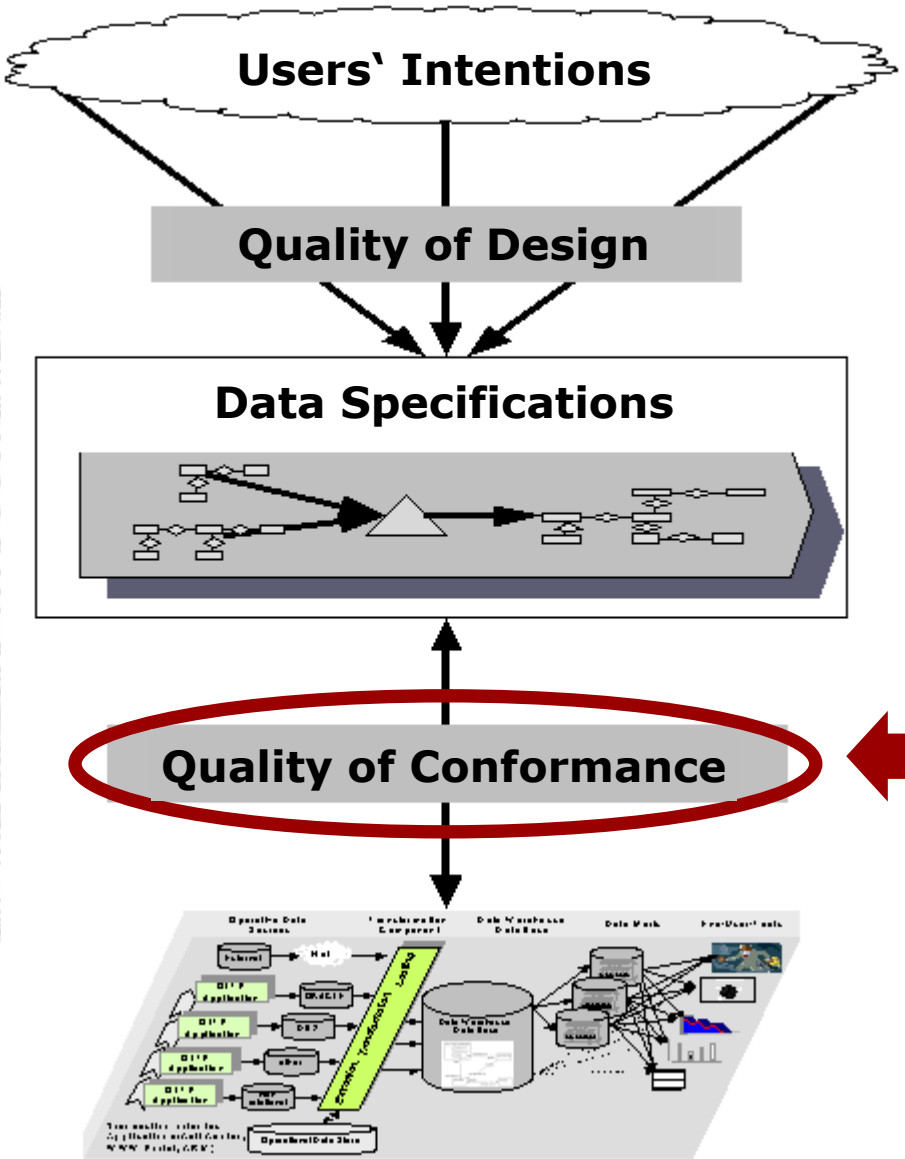
Focus of the Project

- Analyzing the cost benefit ratio of data quality measures
- Starting point: campaign management process

Relevant Questions

- How can data quality be assessed by means of metrics?
- How can the data quality level be improved by taking measures?
- Which measures shall be taken from an economic point of view?

Different Perspectives on Data Quality



Required Data Set

Specified Data Set

Data Quality Dimensions
Focus here: Currency

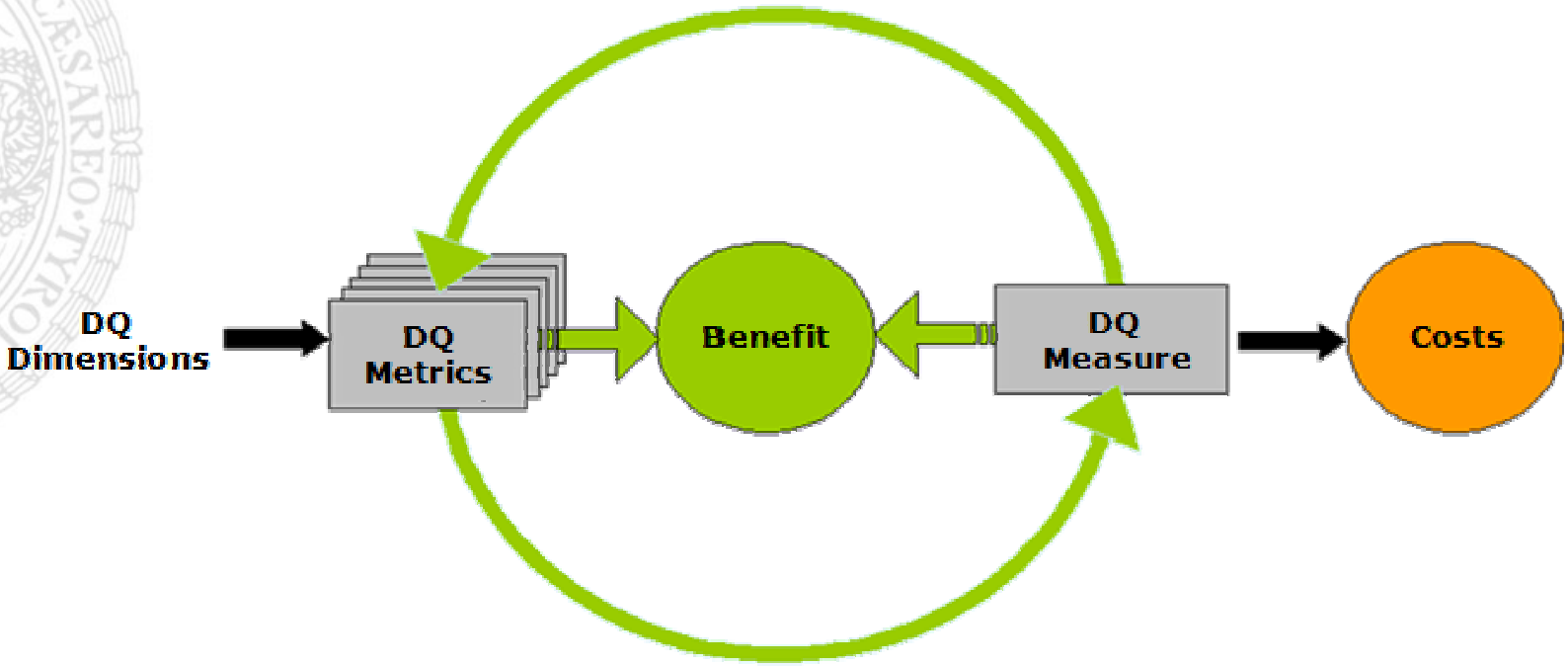
Existing Data Set

Motivation: Data Quality Assessment

Data quality metrics are indispensible for an economic management of data quality.

→"What doesn't get measured doesn't get managed!"

Illustration: Data Quality (DQ) Loop



Requirements on Data Quality Metrics

R1 Normalization

R2 Interval Scale

R3 Interpretability

R4 Aggregation

R5 Adaptivity

R6 Feasibility

Focus: Data Quality Dimension Currency (I)

What is Currency?

- Currency expresses whether data are still up-to-date.
- Attribute values in a database still correspond to the current values of their real-world counterparts at the instant when data quality is assessed (i.e. they have not become outdated).
- Quantifying currency does not necessarily require a real-world test (in contrast to the data quality dimension correctness).

Focus: Data Quality Dimension Currency (II)

Practical Questions (Examples)

- To what extent are the customer contact data (e.g. addresses) in the information system still up-to-date?
- Does the professional status of a customer (e.g. "student"), which was stored once, still correspond to the real-world?

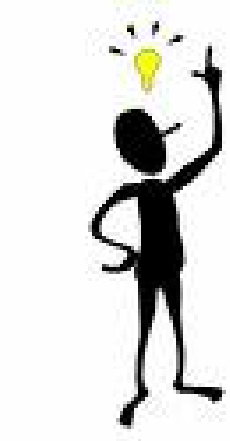
Problem

- In science as well as in practice there is a lack of appropriate data quality metrics for currency.
- To the best of our knowledge, there are no metrics which fulfill the requirements R1 to R6.

Design of a Metric for Currency (I)

Basic Idea

Design of a probability-based data quality metric to enable a widely automated data quality assessment.



- Currency is interpreted as the probability that an attribute value still corresponds to the current value of its real-world counterpart (e.g. probability that a customer's address is still up-to-date).
- Attribute values "grow older" and are characterized by a limited, but unknown shelf life.
- For the most attributes, the probability that an attribute value is still up-to-date decreases over time.

Design of a Metric for Currency (II)

Definition (Level of Attribute Values)

•Let A be an attribute, w an attribute value, and $age(w,A) \in R^+$ the age of the attribute value.

•In addition, let $decline(A) \in R^+$ be the decline rate of attribute values of attribute A.

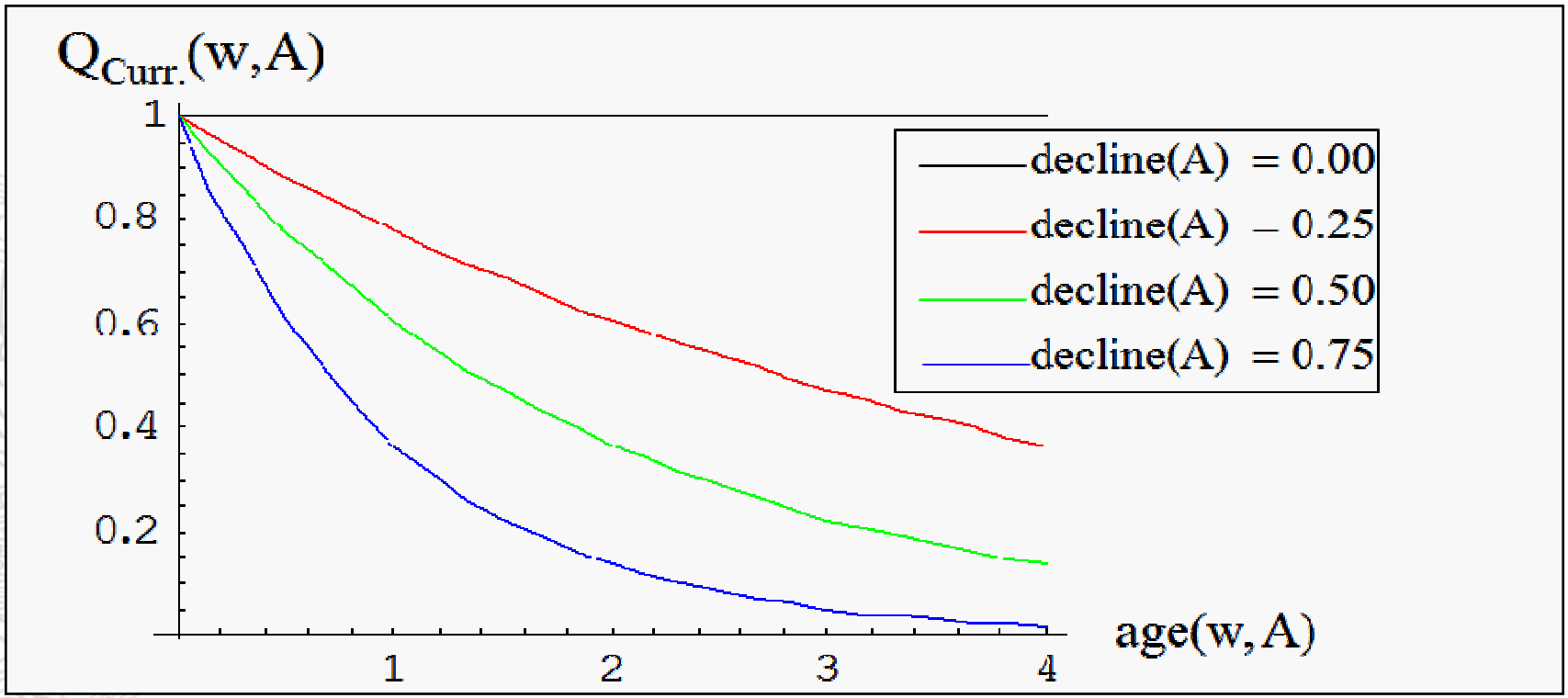
•Based on an exponential distribution (example) for the shelf life of the attribute values, the metric is defined as follows:

$$Q_{Curr.}(w, A) := e^{-decline(A) \cdot age(w,A)}$$

•Note: The value of the metric $Q_{Curr.}(w,A)$ represents the probability that the attribute value w is still up-to-date.

Design of a Metric for Currency (III)

Illustration (Level of Attribute Values)



Note: To quantify currency for data sets (e.g. tuples, relations, views), the values of the metric on the level of attribute values can be aggregated easily.

Design of a Metric for Currency (IV)

Example (Level of Attribute Values)

1. Determine the relevant attributes A_i
2. Determine the parameter $decline(A_i)$ (e.g. based on empirical data)
3. Compute the parameter $age(w_i,A_i)$ based on metadata (instant of assessment and data acquisition)
4. Compute the value of the metric $Q_{Curr.}(w_i,A_i)$:

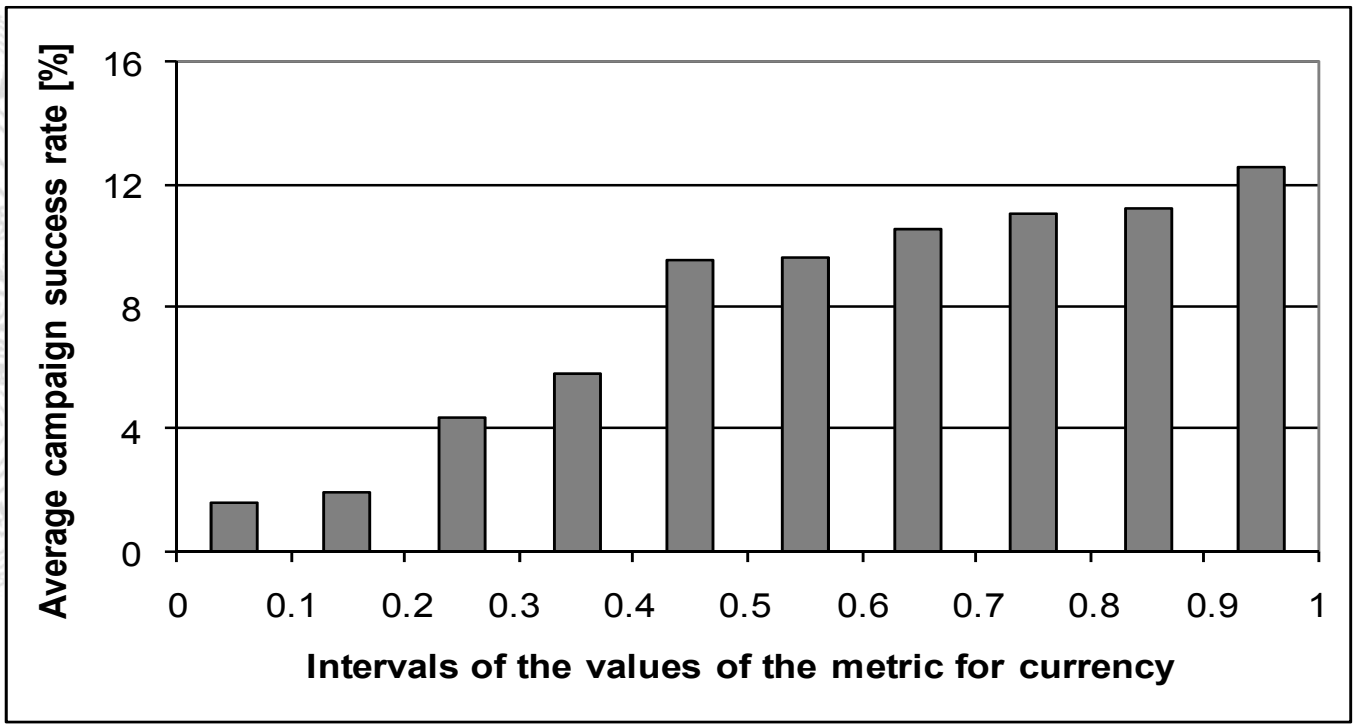
A_i		*surname*	*first_name*	*address*	*current_tariff*
$decline(A_i)$ [1/year]		0.02	0.00	0.10	0.40
Customer A	w_i	"Meier"	"Hans"	"Alstertor 1, 20095 Hamburg"	"Mobile 1000"
	$age(w_i,A_i)$ [year]	0.5	0.5	0.5	0.5
	$Q_{Curr.}(w_i,A_i)$	**0.99**	**1.00**	**0.95**	**0.82**

Application of the Metric for Currency (I.I)

Campaign Management: Prepaid2Postpaid campaign *)

1. Analysis of a former, very similar campaign

Ex post success rates of the *former* campaign depending on the values of the metric for currency

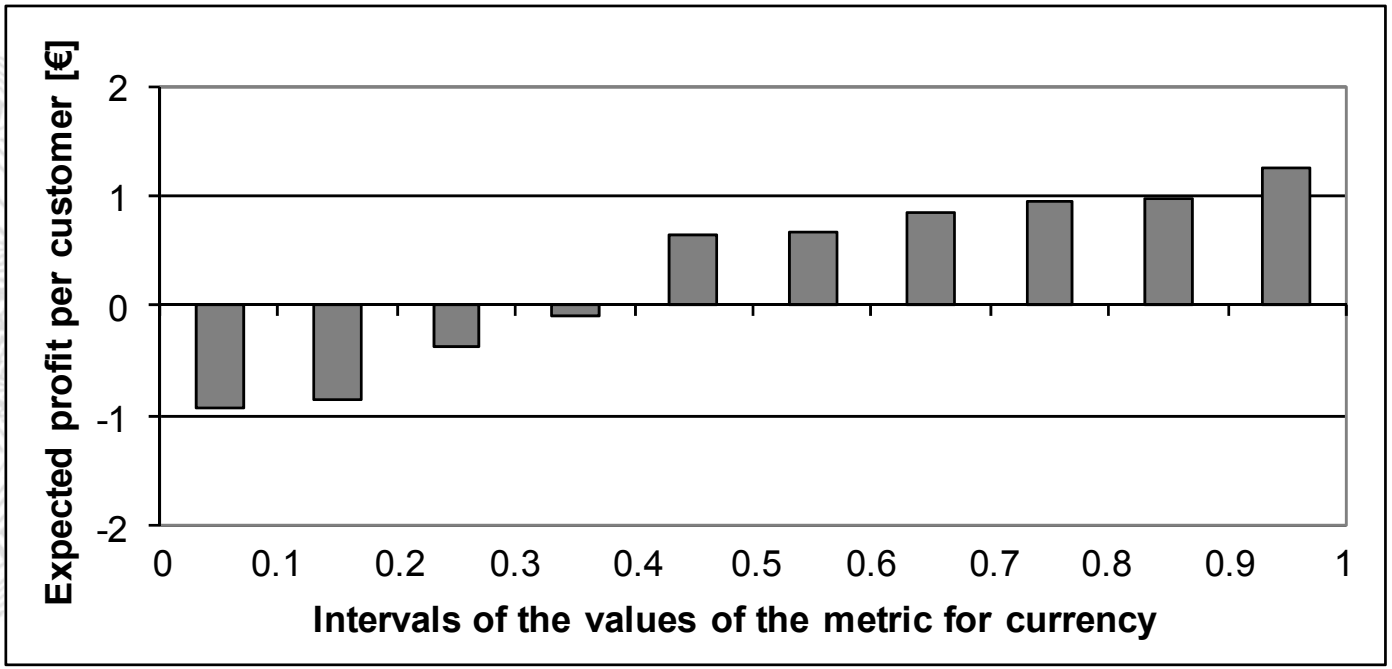


*) The following examples were simplified and made anonymous!

Application of the Metric for Currency (I.II)

2. Estimation of the profits of the *new/current* campaign

Expected (net) profits based on the success rates of the former campaign, mailing costs of 1.25 Euro per mailing and profits of 20 Euro per customer accepting the offer



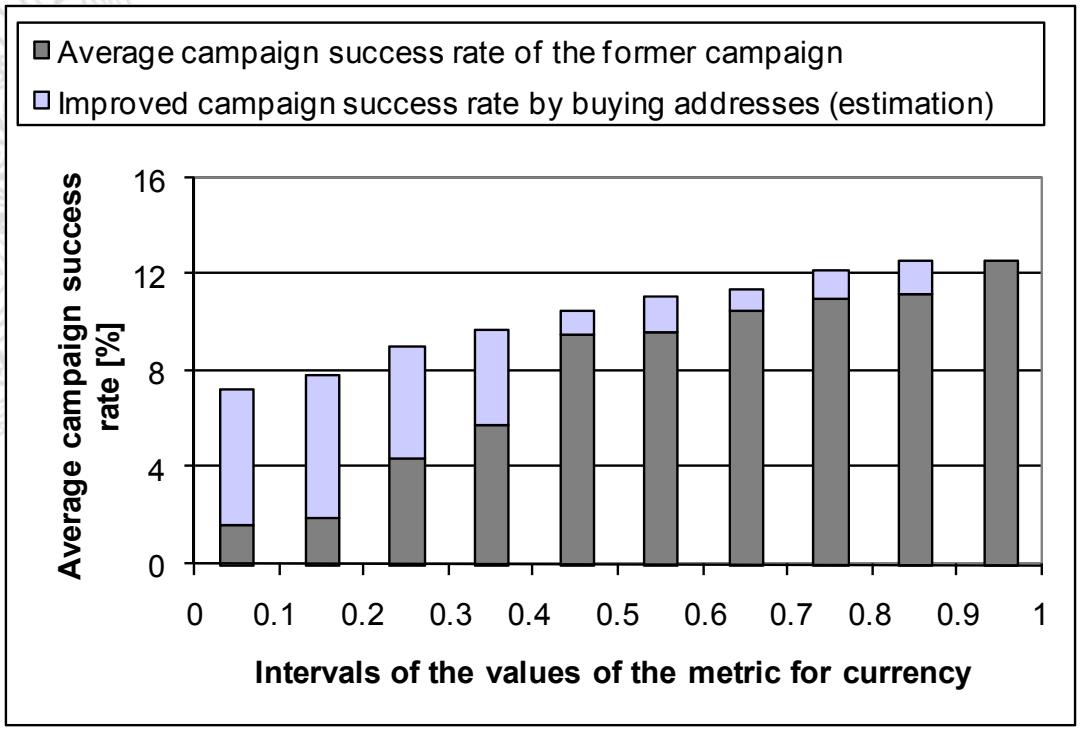
- Interval [0.0;0.4]: campaign not profitable
- Interval]0.4;1.0]: campaign profitable

Application of the Metric for Currency (I.III)

3. Analysis of a possible data quality measure (I)

Buying external address data: Costs in the example: 0.40 Euro / address

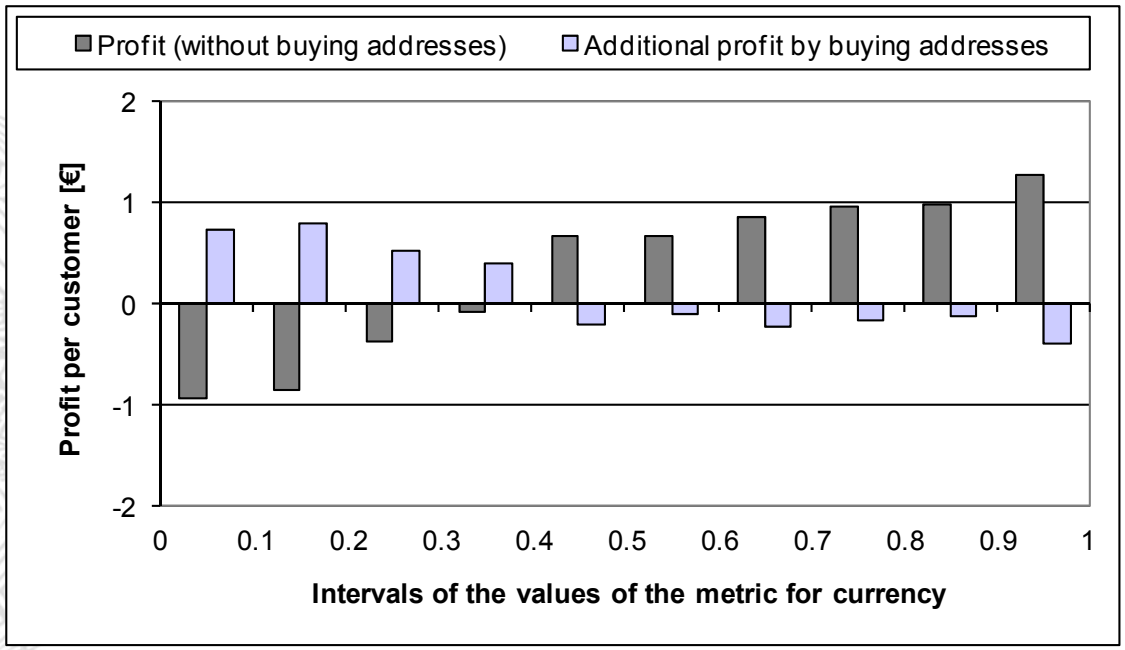
- Calculation of the values of the metric in case of buying address data for the customers of each interval
- Estimation of the improved campaign success rates due to increased data quality in case of buying address data



Application of the Metric for Currency (I.IV)

3. Analysis of a possible data quality measure (II)

Additional profit by buying addresses: profits resulting from improved success rates minus costs for buying addresses



- Interval]0.4;1.0]: buying addresses is not profitable
- Interval]0.0;0.4]: buying addresses is profitable
- However, only within the interval]0.2;0.4] holds:
  Profit + Additional profit by buying addresses >0

Application of the Metric for Currency (I.V)

4. Ex post analysis

- The mobile services provider bought addresses for about 27,500 customers within the interval]0.2;0.4].

  Ex ante and ex post currency of the address data

	Interval]0.2;0.3]	Interval]0.3;0.4]
Total number	≈ 13,200	≈ 14,300
Valid addresses "ex post"	≈ 3,600 (~0,273)	≈ 4,820 (~0,337)

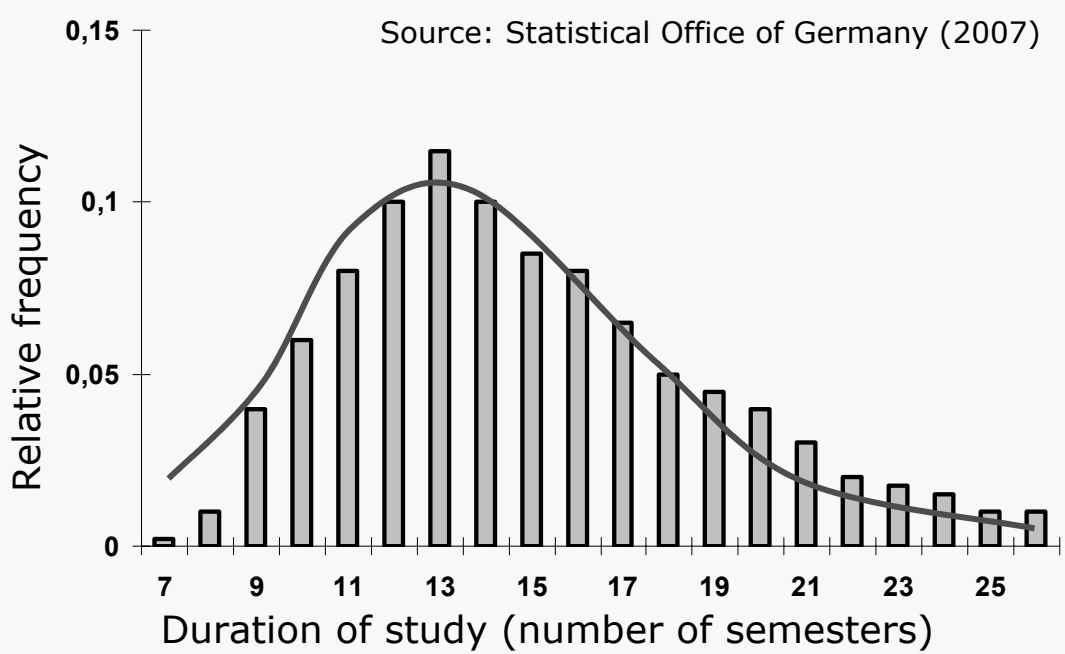
- The ex ante estimated campaign success rates were quite accurate.
- Using the data quality metric it was possible to take data quality measures in a more efficient way.
- The profitability of campaigns could be improved significantly.

But: The Metric does not Fit for all Attributes!

The exponential distribution - which served as a basis when defining the metric - is memoryless. This assumption can be justified for address data, for example.

→ But: This assumption does not hold for all attributes!

- However:
  The metric can be adapted easily by using other types of probability distribution functions!



Application of the Metric for Currency (II.I)

Campaign Management: Student campaign

Initial situation

The mobile services provider had problems to ensure the quality of customer data (e.g. addresses, transactions, products, etc.).

- Addressing customers adequately was quite difficult (e.g. wrong customer master data or contract data).
- Offering customer specific products and tariffs was quite limited.
- Campaign success rates of (selected) customer campaigns were quite low.

Focused problem

Offering a new premium tariff called *Student AAA* to customers with the professional status student.

Application of the Metric for Currency (II.II)

1. Former process for such campaigns

   Address the top 30% customers according to sales values whose professional status is "student" (according to the database).

   → Success rate: about 9%

2. Business case for the campaign (former process)

   a. The top 30% (46,800 customers) out of 156,000 customers with professional status "student" have average sales of 1,340 Euro.

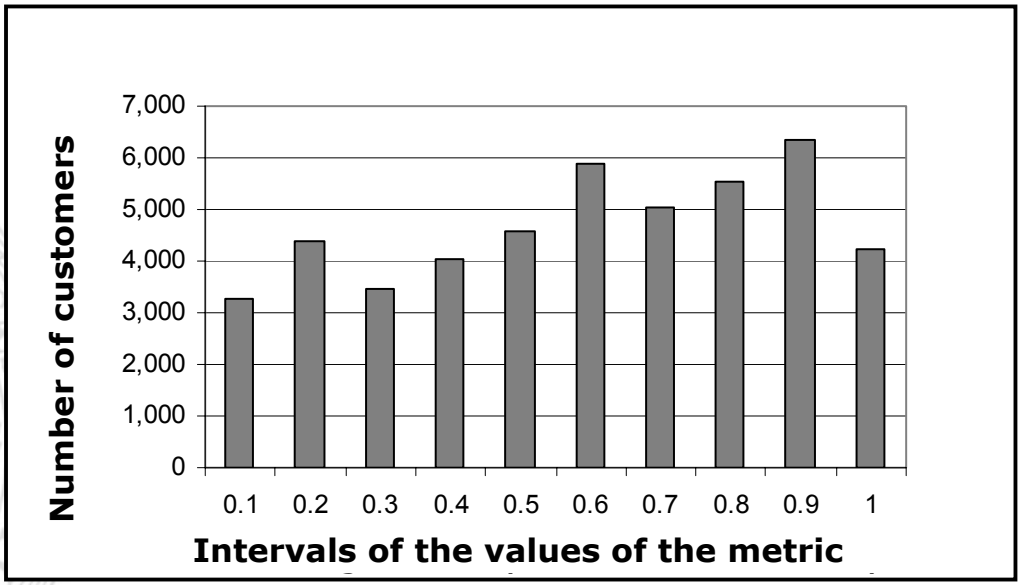
   b. Assuming the former success rate of 9%, about 4,200 customers will accept the offer. For these customers the return on sales will increase by 5% (→ additional profit).

   c. Estimated additional profit:
      4,200 * 1,340 Euro * 0.05 = 281,400 Euro

Application of the Metric for Currency (II.III)

3. New process using the metric for currency

   a. Calculate the values of the metric for all customers with status "student".



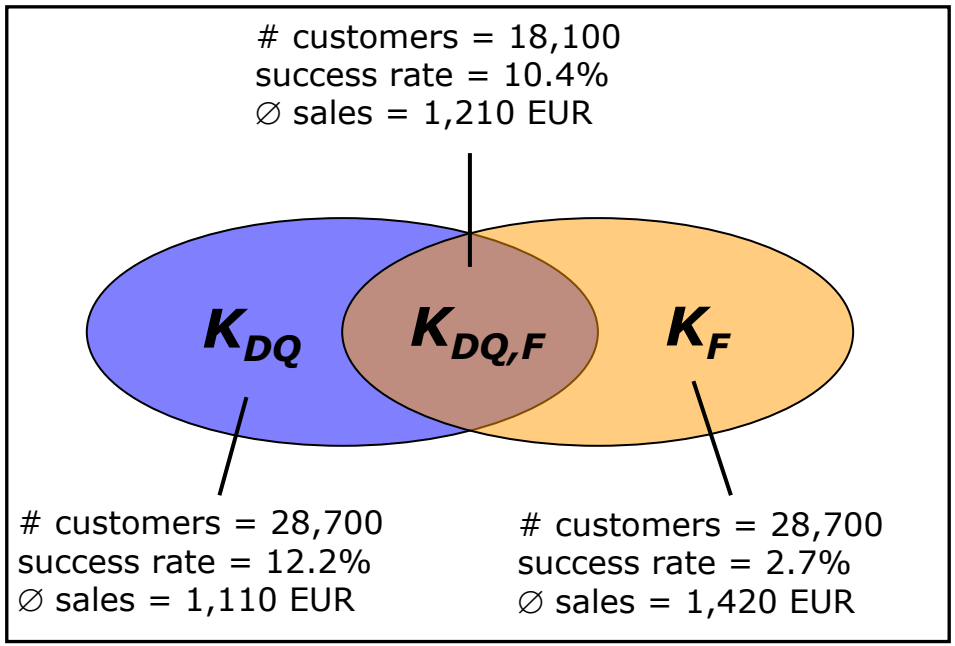
   b. Calculate the expected additional profit depending on the value of the metric for each customer (here: Value of the metric * Sales).

   c. Use this key figure as selection criterion to identify the top 30% customers.
   → From 46,800 customers who were selected this way, 28,700 were not selected according to the former process.

Application of the Metric for Currency (II.IV)

4. Results of the campaign

The mobile services provider decided to address all 75,500 customers who were selected (former (F) and new (DQ) process).



- Former process (F):
  - Success rate: 5.7%
  - ∅-sales: 1,340 Euro
  - Add. profit: 178,700 Euro

- New process (DQ):
  - Success rate: 11.5%
  - ∅- sales: 1,150 Euro
  - Add. profit: 309,200 Euro

Summary

- Development of a probability-based data quality metric for currency which fills a gap in both, science and practice.
- The metric enables an objective and automated data quality assessment and makes it possible to analyze the economic impact of quality measures adequately.
- The metric approved useful and appropriate in practice, led to a substantiated and comprehensible decision support, and resulted in considerable economic benefits.
- The metric is a powerful tool for anyone who is concerned with the economic management of data quality.

Thank you very much for your attention!
Any questions?