Daten im Advertising: Endlich Schluss mit GiGo!

Amit Ghosh & Steffen Wagner

INWT Statistics GmbH

d3con University

Hamburg, 13. März 2017

INWT Statistics

INWT Statistics GmbH

Als Spezialist für Datenanalyse und Predictive Analytics bieten wir schwerpunktmäßig Lösungen in den folgenden Bereichen an:

Online Marketing

- Customer Journey Analysis
- Conversion Optimization
- Fraud Detection
- TV Impact

CRM

- Customer Lifetime Value
- Kundensegmentierung
- Churn Management
- 360° Brand

BI/Reporting

- Datenmanagement
- Datenkonsolidierung
- Dashboards

Ausgewählte Kunden:

CHARITÉ UNIVERSITÄTSMEDIZIN BERLIN

BIOMET

CosmosDirekt. Die Versicherung.

OTTO

SIXT rent a car

tipico

Webtrekk

zalando

Agenda

1. Stellenwert von Daten
2. Eine unbequeme Wahrheit
3. Beispiele: Daten im Advertising
4. 6 Tipps für bessere Datenqualität
5. Fazit

INWT Statistics

„Daten sind das Öl des 21. Jahrhunderts.“

Stefan Gross-Selbeck

Quelle: Von Unbekannt - US Coast Guard - 100421-G-XXXXL- Deepwater Horizon fire (Direct link), Gemeinfrei

INWT Statistics

Daten sind Teil des Alltags

Stellenwert von Daten

- Tech-Unternehmen wie Google, Facebook oder Amazon sind wegen Daten erfolgreich
- Geschäftsmodelle unzähliger Startups basieren auf Daten
- „Datengetriebenes“ Arbeiten ist Selbstverständlichkeit
- Marketing: SEO-Analysen, Conversion Optimization (A/B-Testing), RTB, Churn Management, . . .

INWT Statistics

Daten können „sexy“ sein!

„Die Daten haben wir. – Alles kein Problem.“

Unbekannt

Not quite as sexy (aber leider wahr)...

- **G**arbage **i**n/**G**arbage **o**ut
- das Ergebnis einer Analyse ist nie besser als die Daten auf denen sie basiert
- Effektivität: Bessere Daten haben einen größeren Hebel als ausgefeiltere Methoden
- nicht mehr Daten sind die Lösung, sondern **geeignete Daten**!

„There are a lot of small data problems that occur in big data.
They don't disappear because you've got lots of the stuff.
They get worse."

David Spiegelhalter

INWT Statistics

Das arme Schwein

Faustregel in der „klassischen“ Statistik: ca. **10-20%** der Daten sind fehlerhaft

Ein trauriges Beispiel:

- medizinische Studie zur Zulassung eines Herzschrittmachers
- Untersuchung erfolgt im Tierversuch mit Schweinen
- an einem OP-Tag liefert ein Messinstrument fehlerhafte Daten
- die Daten zu zwei OPs können nicht genutzt werden
- finanzieller Schaden: > 50.000 €

Von Kalumet aus der deutschsprachigen Wikipedia, CC BY-SA 3.0, Link

INWT Statistics

Agenda

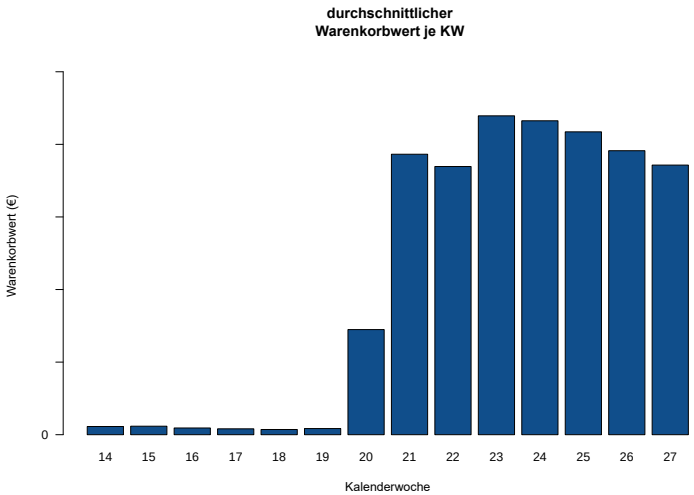
1. Stellenwert von Daten
2. Eine unbequeme Wahrheit
3. Beispiele: Daten im Advertising
4. 6 Tipps für bessere Datenqualität
5. Fazit

INWT Statistics

Agenda

INWT Statistics

Conversion Tracking



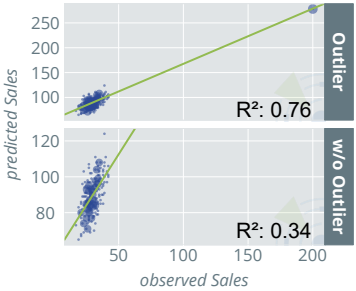
- durchschnittlicher Wert von Conversions
- Daten wurden für ein Attributionsprojekt vom Tracking-Dienst-leister des Kunden bereitgestellt
- Exportzeitraum: 3 Monate
- Umstellung des Trackings genau in der Mitte des exportieren Zeitraums

INWT Statistics

Ausreißer/extreme Werte

CRM-Daten

- Ein einzelner Kunde im B2C-Geschäft hat innerhalb von 12 Monaten ca. 65.000 Transaktionen
- Der durchschnittliche Kunde hat weniger als 2 Transaktionen pro Jahr
- Abhängig vom gewählten Verfahren
  ⇒ Einfluss auf das Ergebnis



Tracking-Daten

- > 1.000 Visits in einer Customer Journey (⌀ = 1.9)
- Session mit einer Dauer von 6.9 Tagen (⌀ = 3.5 Minuten)
- verursacht z.B. durch Bots/Crawler/Monitoring-Tools

INWT Statistics

Fehler im Datenexport

Fehlerhafte Exporteinstellungen:

```
;visit;pi;start;end;duration;sale;saletarget;saletime;orderValue;cl
.teo;utm_id=uLxxKpKr;3016969584897950854;1;12.09.2016 07:30:00;;0;;;
.inet;https://www.sparwelt.de/suche?searchform=H&ouml;kspiel;3016946
)1,41;true
```

- Strings nicht in Anführungszeichen
- HTML-kodierte Zeichen enthalten das Feldtrennzeichen (Semikolon)

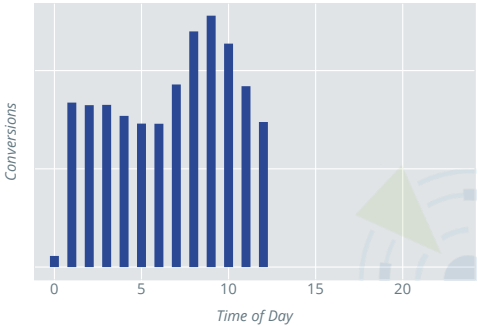
Flüchtigkeitsfehler bei manuellem Export:

- Bereitstellung eines Datenbankdumps (14 GB)
- darin eine Mapping-Tabelle mit Partnern, die genau 1.000 Zeilen enthält

INWT Statistics

Datumswerte

Unterschiedliche Datumsformate:



- Sales Timestamp aus alternativer Datenquelle (<1% der Daten)
- Sales Timestamp verwendet englisches Datum mit AM/PM
- Speicherung in Textfeld mit fester Länge (AM/PM wird abgeschnitten)

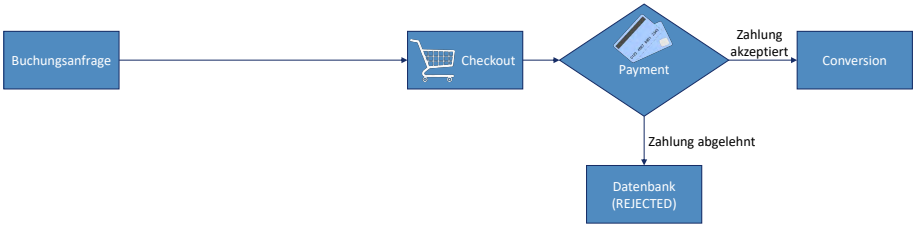
Zeitzonen:

- Projekt zur Erweiterung der Attribution um TV-Impact
- Analyse von Sendeplänen und Tracking-Daten
- es kann kein Einfluss von TV nachgewiesen werden
- Ursache: Zeiten in Sendeplänen von der Agentur in GMT+2 vs. Zeiten im Tracking in UTC

INWT Statistics

Daten sind **„selbsterklärend“**

- Datenbank-Export enthält das Feld `status`
- Werte: `SUCCESS`, `REJECTED`, . . .
- Annahme: `REJECTED` – Buchung kann nicht abgeschlossen werden
- erst aufgrund von Unstimmigkeiten stellt sich heraus:

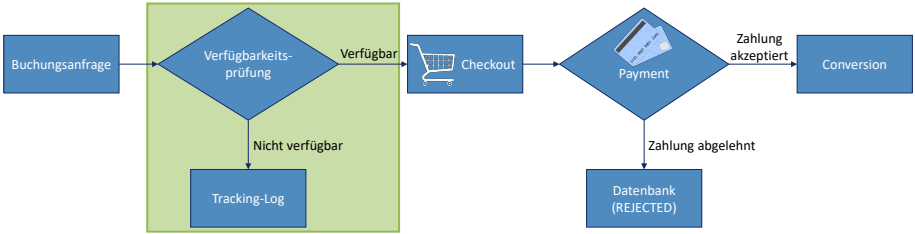


- die interessierende Prüfung findet vor dem Einstieg in den Checkout-Prozess statt und wird nicht in der Datenbank geloggt, sondern nur in den Tracking-Daten des Webservers

INWT Statistics

Daten sind **„selbsterklärend“**

- Datenbank-Export enthält das Feld `status`
- Werte: `SUCCESS`, `REJECTED`, . . .
- Annahme: `REJECTED` – Buchung kann nicht abgeschlossen werden
- erst aufgrund von Unstimmigkeiten stellt sich heraus:



- die interessierende Prüfung findet vor dem Einstieg in den Checkout-Prozess statt und wird nicht in der Datenbank geloggt, sondern nur in den Tracking-Daten des Webservers

INWT Statistics

Agenda

1. Stellenwert von Daten
2. Eine unbequeme Wahrheit
3. Beispiele: Daten im Advertising
4. 6 Tipps für bessere Datenqualität
5. Fazit

INWT Statistics

Agenda

INWT Statistics

Don’t Save All The Data You Can Get. . .

but all that you **need!**

INWT Statistics

Tipp 2: Data Warehouse mit konsolidierten Daten

Wenn man nicht auf der „grünen Wiese“ starten kann...

- neues Data Warehous (DW) aufsetzen
- Daten gezielt im ETL-Prozess bei der Aufnahme konsolidieren
- für alle Tabellen im DW auf Dokumentation achten

$\Rightarrow$ Analysen auf konsolidierten Daten

INWT Statistics

Tipp 3: Dokumentation

Dokumentation ist lästig, aber wichtig...

- Daten/Prozesse sind selten so eindeutig wie gedacht
- gravierende Fehlinterpretationen vermeiden
- Katastrophen bei Personalfluktuation verhindern
- Dokumentation betrifft auch:
  - Umstellung Tracking, technische Ausfälle & besondere Ereignisse, Re-Launch Webseite, Marketing-Maßnahmen, ...

⇒ Richtlinie zu Dokumentation

INWT Statistics

Tipp 4: Mit Daten regelmäßig arbeiten

Eigentlich trivial, aber…

- werden Daten erstmalig “genutzt“ finden sich immer Fehler
- nachträgliche Korrektur der Daten evtl. nicht möglich
- häufig: Anpassung im datengenerierenden Prozess (Analyse verzögert sich)

⇒ BI-Anbindung & Dashboards, Black-Boxes vermeiden

INWT Statistics

Tipp 5: Daten bereinigen

Alte Daten sind oft wertlos...

- es sollten nur Daten gehalten werden, die benötigt werden
- Beispiel: Tracking-Daten der alten Webseite werden einige Monate nach Re-Launch nicht mehr benötigt
- Datenschutz / Prävention im Falle eines Leaks
- Ressourcen einsparen

$\Rightarrow$ interaktives Arbeiten ermöglichen

INWT Statistics

Tipp 6: Datenvalidierung als Teil des Projekts

Immer auf valide Daten achten...

- jeder Analyse eine (angemessene) Datenvalidierung voranstellen
- auch Dienstleister darauf verpflichten
- Bewertung anhand von KPIs/Gesamtbild statt Einzelfällen
- sinnvolles Maß finden (100% Datenqualität sind nicht erreichbar)
- robuste Verfahren bevorzugen

⇒ falsche Gewissheit ist schlimmer als Ungewissheit!

INWT Statistics

The Big Picture

- Daten sind „nur“ ein Rohstoff
- Mehrwert lässt sich nur im **Prozess** aus Daten generieren
- dieser ist nur so stark wie das schwächste Glied in der Kette

⇒ nicht nur um die Themen kümmern, die „sexy“ sind!

INWT Statistics

Fazit: Cut the Buzzwords

- Trend: Nach dem Hype um Big Data wird auf **Ergebnisse** geschaut
- Kunden hinterfragen den **Nutzen** von Analysen und Tools
- eine Analyse auf ungeeigneten und falschen Daten ist **im besten Fall nutzlos**
- auch für Daten gilt: **Qualität** vor Quantität
- 100% Datenqualität sind unrealistisch (aber auch nicht das Ziel $\rightarrow$ **robuste Methoden**)
- **Software unterstützt** Datenvalidierung, ersetzt den Menschen aber noch nicht

INWT Statistics

Vielen Dank für die Aufmerksamkeit!

INWT Statistics GmbH
E-Mail: info@inwt-statistics.de
Internet: www.inwt-statistics.de
Tel.: +49 30 609857990