

# **HOLTZBRINCK PUBLISHING GROUP**

## Interne Revision und Datenanalysen

kivitando Jahrestreffen

Jörg Meyer | Bonn | 11. April 2018

**SPRINGER NATURE**

 **macmillan**  
Publishers

 **holtzbrinck**  
Digital, Information & Services

 **holtzbrinck**  
Publishing Group

# Holtzbrinck Publishing Group

## Interne Revision und Datenanalysen

### Gliederung

- Die Verlagsgruppe im Überblick
- Interne Revision – Vorgehen und Ziele
- Datenanalysen – Praktische Erfahrungen mit Journal Entry Testing
- Datenanalysen – Technische Durchführung und Perl-Einsatz

# Holtzbrinck Publishing Group

## Die Verlagsgruppe im Überblick

1

# Holtzbrinck Publishing Group

## Werte



### Qualität

Fleiß, Talent und Geschick haben eine große Tradition begründet. Wir sind zuallererst unseren Autoren zu großem Dank verpflichtet. Viele von ihnen haben prestigereiche Auszeichnungen erhalten.



### Wachstum

Wachstum ist für uns kein Selbstzweck, der die Natur übergeht, sondern die Folge eines inspirierten und umsichtigen Unternehmergeistes.



### Verantwortung

Verantwortung übernehmen wir nicht nur für unsere Mitarbeiter und die uns anvertrauten oder von uns veröffentlichten Werke, sondern auch aktiv als Teil der Gesellschaft, in der wir leben.

<http://www.holtzbrinck.com>

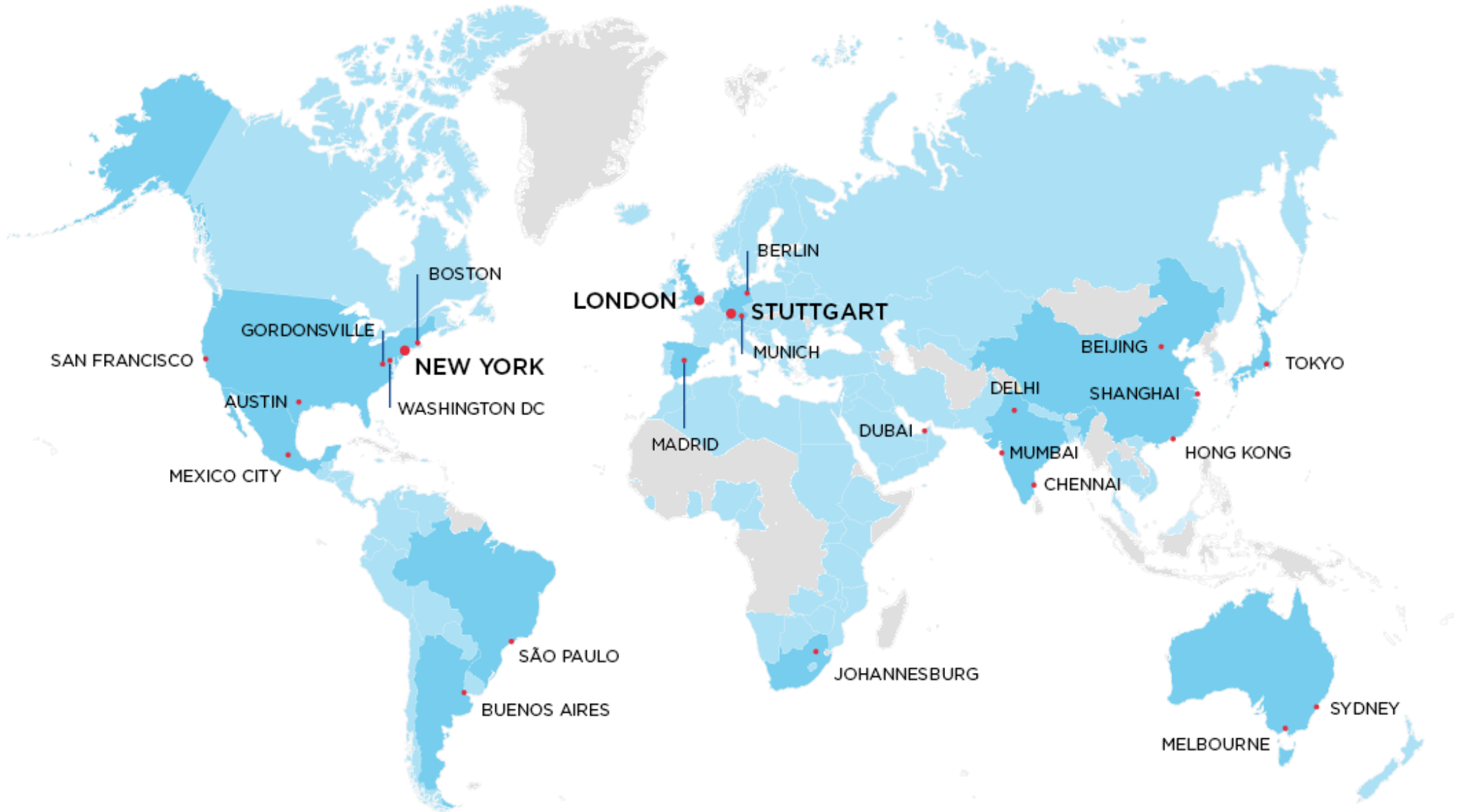
# Holtzbrinck Publishing Group

## Geschäftsbereiche



# Holtzbrinck Publishing Group

Weltweite Präsenz



# Interne Revision

## Vorgehen und Ziele

2

# Interne Revision

## Der „Three lines of defence“-Ansatz

**Die Interne Revision vieler Unternehmen orientiert sich am Ansatz „Three lines of defence“, der u.a. von der European Confederation of Institutes of Internal Auditing (ECIIA) empfohlen wird:**



Betriebliche Prozesse	Management	Interne Revision
<ul style="list-style-type: none"><li>– Definition von Prozessen und Verantwortlichkeiten</li><li>– Kenntnis von Aufgaben und Pflichten</li><li>– Wirksamkeit der Kontrollen</li><li>– Dokumentation</li></ul>	<ul style="list-style-type: none"><li>– Führung und Verantwortung</li><li>– Definition der Arbeitsprozesse</li><li>– Definition der Zuständigkeiten</li><li>– Beurteilung von Risiken</li><li>– Kontrolle der Einhaltung</li></ul>	<ul style="list-style-type: none"><li>– Geplante Prüfungen sowie Untersuchungen bei Bedarf</li><li>– Unterstützung bei Verbesserung der Organisation</li><li>– Überprüfung der Wirksamkeit von Kontrollen</li><li>– Unterstützung bei Erkennung und Bewertung von Risiken</li></ul>



# Interne Revision

## Vier Kontrollziele

### 1. Definition der Abläufe / Schaffung von Prozessqualität

*„Wie soll der Ablauf gestaltet sein? Wer soll was tun bzw. überprüfen?  
In welchen Abständen? Wie dokumentiert?“*

### 2. Dokumentation und Information

- Dokumentation von Abläufen und Pflichten (erstmalig und nach Änderungen)
- Geschäftsvorfälle und Bewegung von Gütern und Werten (kontinuierlich)
- Einhaltung rechtlicher inkl. steuerlicher Pflichten (gemäß Vorschrift)

### 3. Funktionstrennung („Vier-Augen-Prinzip“) / Delegation / Zuordnung

*Beispiele: Abteilungen, Abzeichnen, Wertgrenzen, EDV-Zugänge und Berechtigungen.*

### 4. Übersicht und Qualitätskontrolle

- Vergleich von „Soll“ mit „Ist“.
- Eingreifen, wenn erforderlich.

# Interne Revision

## Prüfungsmethodik

### Inhalte

- Auswählen ggfs. abgrenzen

### Zeitplanung

- Ablaufplan
- Anpassung während der Prüfung

### Arbeitspapiere

- Gemäß Ablaufplan gegliedert
- Qualitätssicherung der Prüfung
- Dokumentation

### Bericht

- Mehrstufiger Berichtsprozeß

Internal Audit -- Meeting Schedule Unit XYZ, October 2017					
	Mon, Oct 24.	Tue, Oct 25.	Wed, Oct 26.	Thu, Oct 27.	Fri, Oct 28.
9 - 10			Office tour	Finance and Accounting	Human Resources
10 - 11					
11 - 12		<b>Introduction Meeting and Audit Scope</b>	Editorial Department	Controlling and Reporting	Royalty & Permissions
between 12 - 14	Lunch Break			Tailgate Lunch	
14 - 15		Market and Competition	Information Technology	Digital Projects	Risk Management and Business Continuity
15 - 16		Internal Organization of Business	Operations Department	Marketing	Document review / audit documents
16 - 17					
17 - 18		Document review	Document review	Sample testing	HM Olympics -- Lawn Games and Fun in the afternoon
	Mon, Oct 31.	Tue, Nov 1.	Wed, Nov 2.	Thu, Nov 3.	Fri, Nov 4.
9 - 10					
10 - 11	Credit Collection	Royalties and Permissions	Sample testing	Preparation of audit documentation	
11 - 12	Travel and Expenses	Risk Management and Business Continuity	(TBD)		
between 12 - 14	Lunch Break				
14 - 15	OPTIONAL: Risk Management Methodology training	Function testing	Function testing	Debriefing Meeting	
15 - 16					
16 - 17		Document review	Remaining questions	Transfer to airport	
17 - 18					

**1 Ausgangssituation**

**1.1 Gesprächspartner**

1: Name, Funktion

2: Name, Funktion

Datum, Uhrzeit

Sonstige Angaben

---

**1.2 Gegenstand der Prüfung**

Welche Gesellschaften

Geschäftsbereiche / Marken

Was gehört nicht zum Prüfungsgegenstand?

---

**1.3 Marktwirtschaftliche Situation**

Wesentliche Geschäftsaktivitäten benennen

SWOT-Analyse

- Stärken (S)
- Schwächen (W)
- Chancen (O)
- Bedrohungen und Risiken (T)

Mittel- / Langfristige Planungen und Ziele der Gesellschaft?

---

**1.4 Wesentliche Kunden / Lieferanten / Autoren / Geschäftspartner?**

Welches sind die wesentlichen Kunden und Vertriebskanäle?

Hat die Gesellschaft öffentliche Kunden (Government Business)?

Wie wurde das herausgefunden?

Gelten für Wettbewerber AB&C-Regelungen?

---

**1.6 Volkswirtschaftliche Situation**

Allgemeine Einschätzung und Entwicklung abgeben lassen

Gesetzliche Rahmenbedingungen und erwartete Entwicklung?

Welche Grundsätze zum Rechnungswesen gelten im Land?

---

**1.7 Situation Gebäude und Infrastruktur**

Grobe Sichtung mit Bezug auf den Prüfungsstandard Safety & Security

Gibt es Nebengebäude oder weitere Betriebsstätten?

- Wer ist Eigentümer der Gebäude?
- Gibt es einen schriftlichen Mietvertrag? Welche Dauer?
- Bestehen Aufbau-, Umbau- oder Ausbauprojekte?

# Holtzbrinck Publishing Group

## Datenanalysen – Praktische Erfahrungen

3

# Datenanalysen – Praktische Erfahrungen

## Auswertbare Datenquellen und Geschäftsprozesse

### Bestandsführende Systeme

- Finanzbuchhaltung Hauptbuch → **Journal Entry Testing (JET)**
  - Nebenbücher Kreditoren und Debitoren
  - Materialwirtschaft und Lagerhaltung
  - Personal?
- Sachanwendungen
- Infrastruktursysteme
  - Zeiterfassung / Zugangskontrollsysteme?

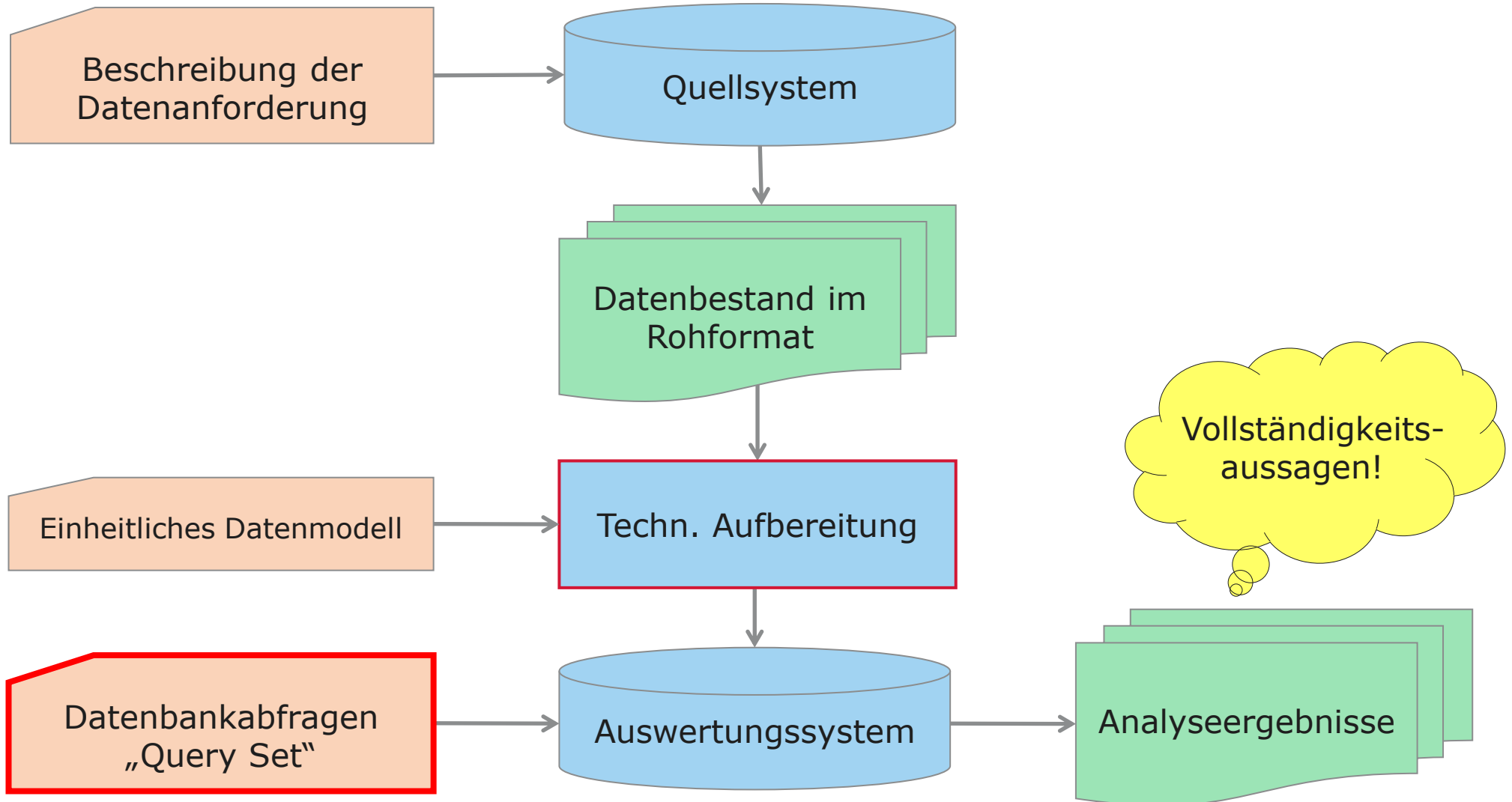
**DATENSCHUTZ BEACHTEN!**

### Geschäftsprozesse (Auswahl)

- Bestellwesen (inkl. Ausschreibungen) und Zahlungswesen
- Verarbeitung von Eingangsrechnungen → Fibu-Performance im Allgemeinen
- Versicherungen: Policierung und Schadensabwicklung
- Warenströme in produzierenden Unternehmen
- Compliance-Aktivitäten
- Systemwechsel, Jahresübergänge und Unternehmenskäufe

# Datenanalysen – Praktische Erfahrungen

Ablauf einer Datenanalyse mit Datenextraktion



# Datenanalysen – Praktische Erfahrungen

## Journal Entry Testing – Auswahl

### **JET-00\_Gesamtübersicht**

JET-11\_Gebuchte\_Belege\_je\_Periode

JET-12\_Erfaßte\_Zeilen\_je\_Periode

JET-13\_Erfaßte\_Belege\_je\_Tag\_im\_Monat

JET-14\_Erfaßte\_Belege\_je\_Stunde\_im\_Tag

JET-15\_Erfaßte\_Belege\_an\_WE\_und\_Feiertagen

### **JET-16\_Buchungen\_nach\_Periodenende**

JET-16a\_Liste\_später\_Buchungen

JET-17\_Erfassung\_vor\_Belegdatum

### **JET-18\_Altersschichtung\_späte\_Bearbeitung**

### **JET-19\_Belegnummersequenzen**

JET-20\_Schichtung\_Beträge

JET-21\_Größte\_Beträge

JET-23\_Schichtung\_glatte\_Beträge

JET-24\_Benford

JET-26\_Wechselkurse\_Fremdwährungen

### **JET-27\_Übersicht\_Wechselkurse**

### **JET-29\_Unausgeglichene\_Belege**

JET-31\_Längste\_Belege

JET-33\_Häufigste\_Texte

JET-34\_Belege\_ohne\_Text\_manuell\_maschinell

JET-37\_Belegtexte\_mit\_sonstigen\_Stichworten

### **JET-40\_Kontenübersicht**

### **JET-41\_Konten\_mit\_Ausreißern**

### **JET-42\_Selten\_genutzte\_Konten**

JET-43\_Besondere\_Konten

JET-44\_Konten\_nicht\_im\_Kontenplan

JET-49\_Schichtung\_negative\_Beträge

JET-50\_Periodenübersicht\_je\_Kostenstelle

### **JET-51\_Große\_und\_lange\_Belege\_je\_Periode**

### **JET-52\_Salden\_durchlaufender\_Posten**

JET-62\_Große\_und\_lange\_Belege\_je\_Kategorie

JET-63\_Schichtung\_Beträge\_je\_Kategorie

JET-65\_Schichtung\_Kategoriehäufigkeit

### **JET-66\_Kategoriekombinationen**

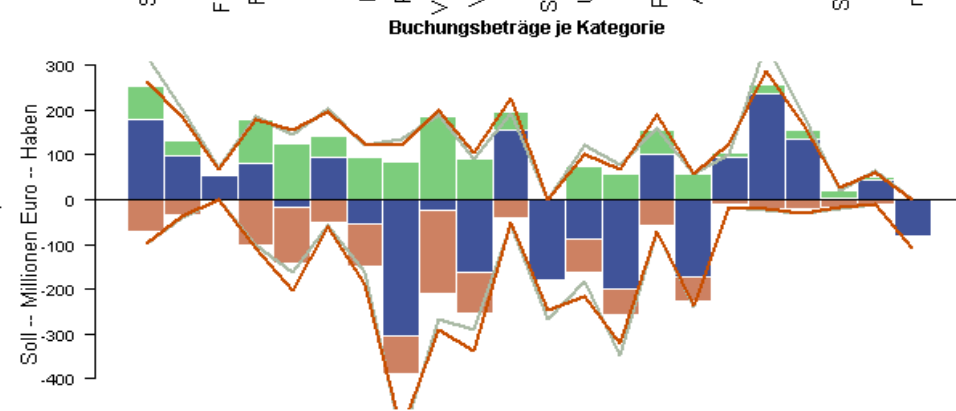
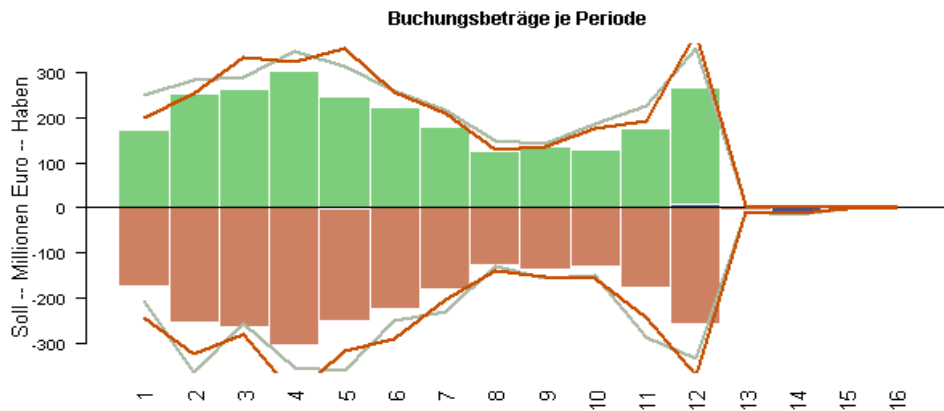
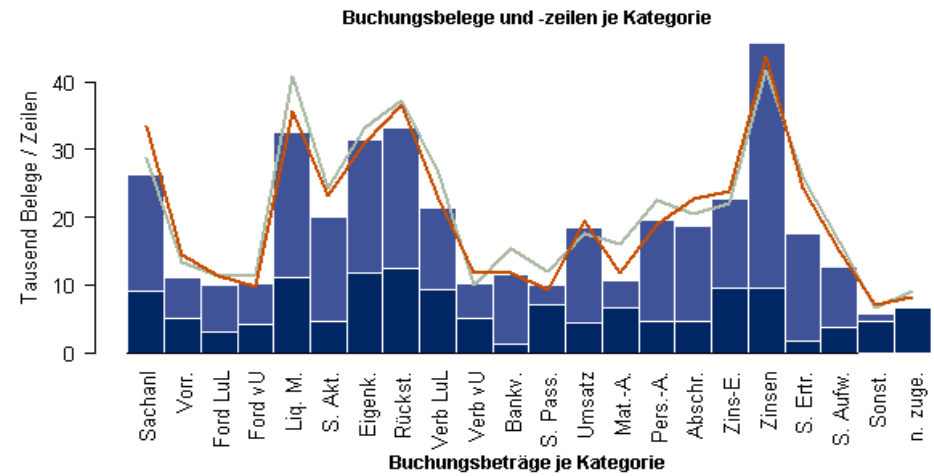
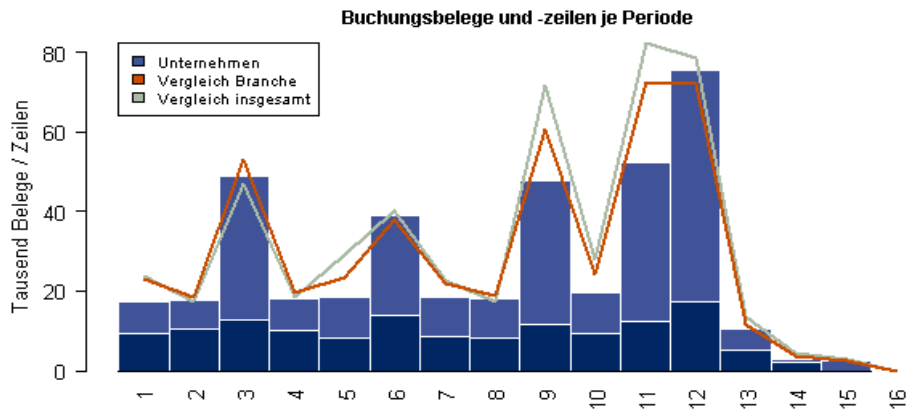
### **JET-68\_Kategorien\_Kreuztabelle**

**Zusätzlich Auswertung Kreditoren und Debitoren**  
**Zusätzlich Auswertung CPD-Zahlungen**

# Interne Revision

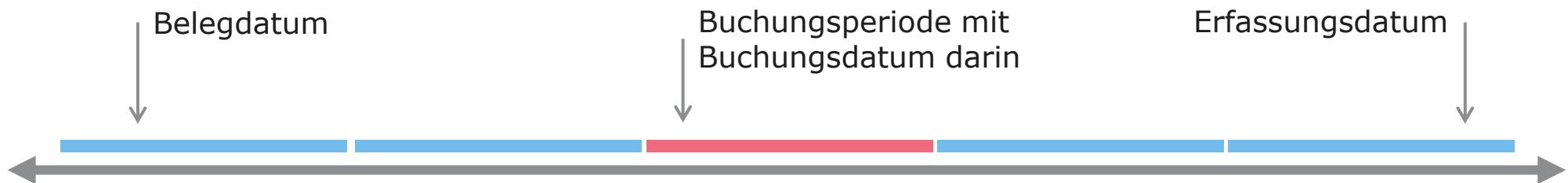
## Explorative Analyse von Unternehmensdaten

Bild 2: Branche Veranstalter -- Buchungskreis 1 -- Geschäftsjahr 2010



# Datenanalysen – Praktische Erfahrungen

## Analyse der Zeitachse



### Typische Auffälligkeiten

Buchungsperiode << Erfassungsdatum

Buchungsperiode wurde nicht geschlossen oder später wieder geöffnet. → JET-16

Belegdatum > Erfassungsdatum

Theoretisch unmöglich. Was ist das für ein Beleg? → JET-17

Belegdatum << Buchungsperiode

Eingangsrechnungen erst sehr spät verarbeitet. Rechnungseingangskontrolle evtl. nicht effizient. → JET-18

Buchungsdatum  $\notin$  Buchungsperiode

Inkonsistenz, ebenso bei unsinnigen Datumswerten.



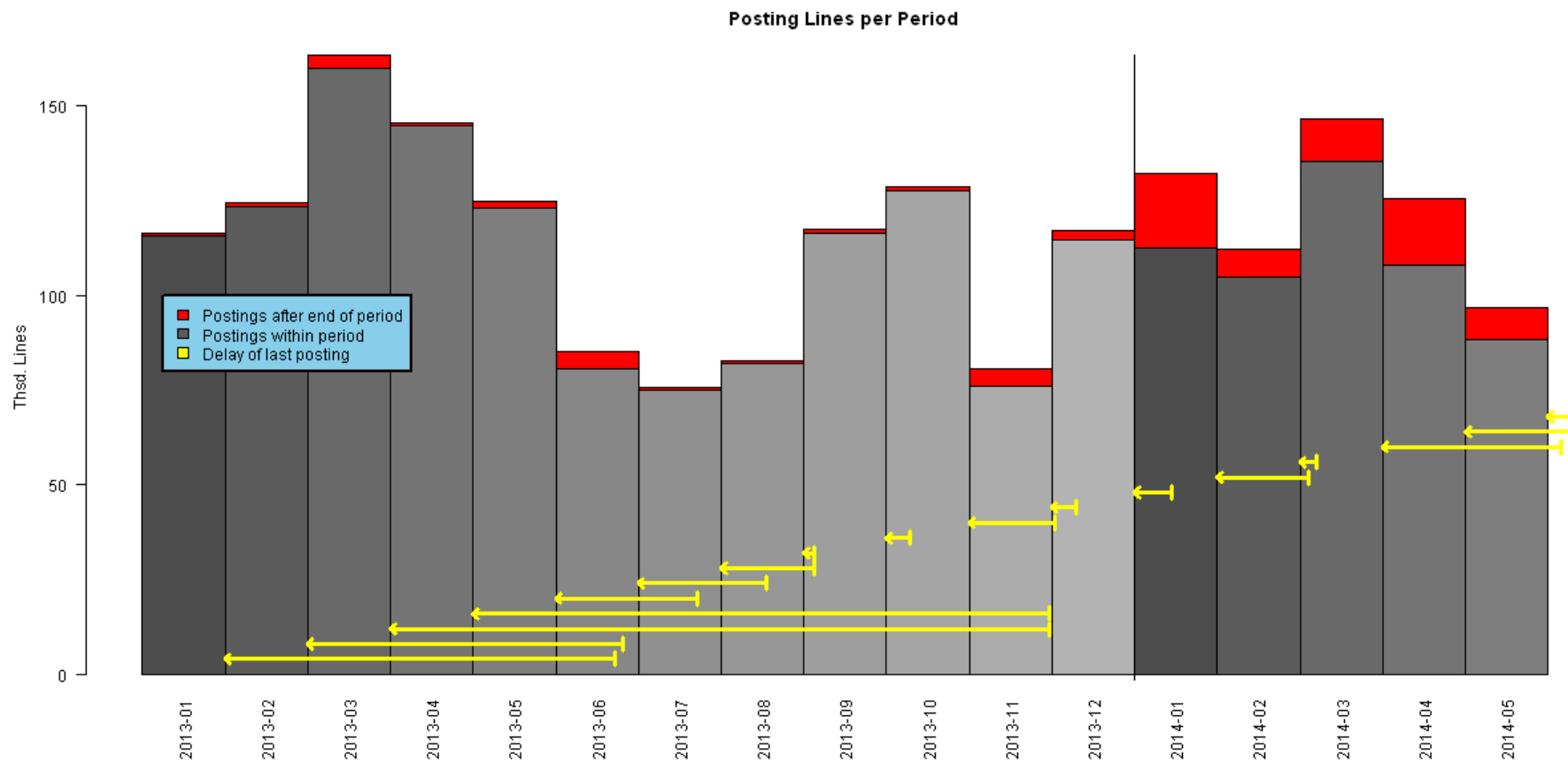
# Datenanalysen – Praktische Erfahrungen

## Praxisbeispiel JET-16 „Buchungen nach Periodenende“

Periode	Buchungen pro Periode	Davon im Folgemonat oder später erfasst	Davon später als nach zwei Wochen erfasst	Letzte Buchung für die Periode	Anzahl Tage nach Periodenende
2016-01	116.313	690	213	26.06.2016	146
2016-02	124.515	1.085	21	26.06.2016	118
2016-03	163.393	3.380	1.731	03.12.2016	247
2016-04	145.437	558	25	18.05.2016	18
2016-05	124.949	1.768	0	03.06.2016	3
2016-06	85.032	4.495	16	17.08.2016	48
2016-07	75.701	605	15	16.08.2016	16
2016-08	82.718	803	0	04.09.2016	4
2016-09	117.512	1.185	0	09.10.2016	9
2016-10	128.562	1.000	0	12.11.2016	12
2016-11	80.631	4.459	0	09.12.2016	9
2016-12	116.953	2.455	0	14.01.2017	14

# Datenanalysen – Praktische Erfahrungen

## Datenvisualisierung JET-16 „Buchungen nach Periodenende“



# Datenanalysen – Praktische Erfahrungen

## Praxisbeispiel JET-52 „Salden durchlaufender Posten“

Periode	Konto	Bezeichnung	Summe Soll	Summe Haben	Verkehrs-Saldo
2016-08	384711	Durchlaufende Posten	15.324,74	0,00	15.324,74
2016-09	384711	Durchlaufende Posten	7.291,66	0,00	7.291,66
2016-10	384711	Durchlaufende Posten	8.118,25	826,59	7.291,66
2016-11	384711	Durchlaufende Posten	7.291,66	0,00	7.291,66
2016-12	384711	Durchlaufende Posten	7.291,66	43.750,00	-36.458,34
2017-01	384711	Durchlaufende Posten	16.335,66	7.600,00	8.735,66
2017-02	384711	Durchlaufende Posten	16.335,66	7.600,00	8.735,66
2017-03	384711	Durchlaufende Posten	16.669,02	14.266,66	2.402,36
2017-04	384711	Durchlaufende Posten	26.176,33	7.600,00	18.576,33
2017-05	384711	Durchlaufende Posten	82.154,57	88.845,95	-6.691,38

# Datenanalysen – Praktische Erfahrungen

## Journal Entry Testing – Auswahl

JET-00\_Gesamtübersicht

JET-11\_Gebuchte\_Belege\_je\_Periode

JET-12\_Erfaßte\_Zeilen\_je\_Periode

JET-13\_Erfaßte\_Belege\_je\_Tag\_im\_Monat

JET-14\_Erfaßte\_Belege\_je\_Stunde\_im\_Tag

JET-15\_Erfaßte\_Belege\_an\_WE\_und\_Feiertagen

JET-16\_Buchungen\_nach\_Periodenende

JET-16a\_Liste\_später\_Buchungen

JET-17\_Erfassung\_vor\_Belegdatum

JET-18\_Altersschichtung\_späte\_Bearbeitung

JET-19\_Belegnummersequenzen

JET-20\_Schichtung\_Beträge

JET-21\_Größte\_Beträge

JET-23\_Schichtung\_glatte\_Beträge

JET-24\_Benford

JET-26\_Wechselkurse\_Fremdwährungen

JET-27\_Übersicht\_Wechselkurse

JET-29\_Unausgeglichene\_Belege

JET-31\_Längste\_Belege

JET-33\_Häufigste\_Texte

JET-34\_Belege\_ohne\_Text\_manuell\_maschinell

JET-37\_Belegtexte\_mit\_sonstigen\_Stichworten

JET-40\_Kontenübersicht

JET-41\_Konten\_mit\_Ausreißern

JET-42\_Selten\_genutzte\_Konten

JET-43\_Besondere\_Konten

JET-44\_Konten\_nicht\_im\_Kontenplan

JET-49\_Schichtung\_negative\_Beträge

JET-50\_Periodenübersicht\_je\_Kostenstelle

JET-51\_Große\_und\_lange\_Belege\_je\_Periode

JET-52\_Salden\_durchlaufender\_Posten

**JET-62\_Große\_und\_lange\_Belege\_je\_Kategorie**

**JET-63\_Schichtung\_Beträge\_je\_Kategorie**

JET-65\_Schichtung\_Kategoriehäufigkeit

JET-66\_Kategoriekombinationen

JET-68\_Kategorien\_Kreuztabelle

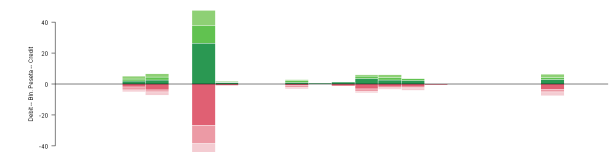
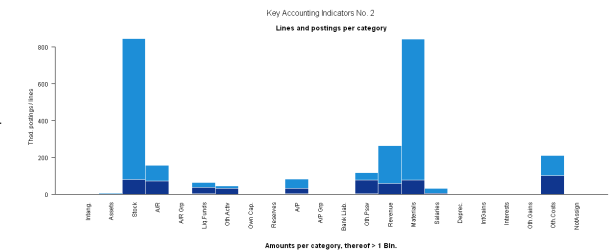
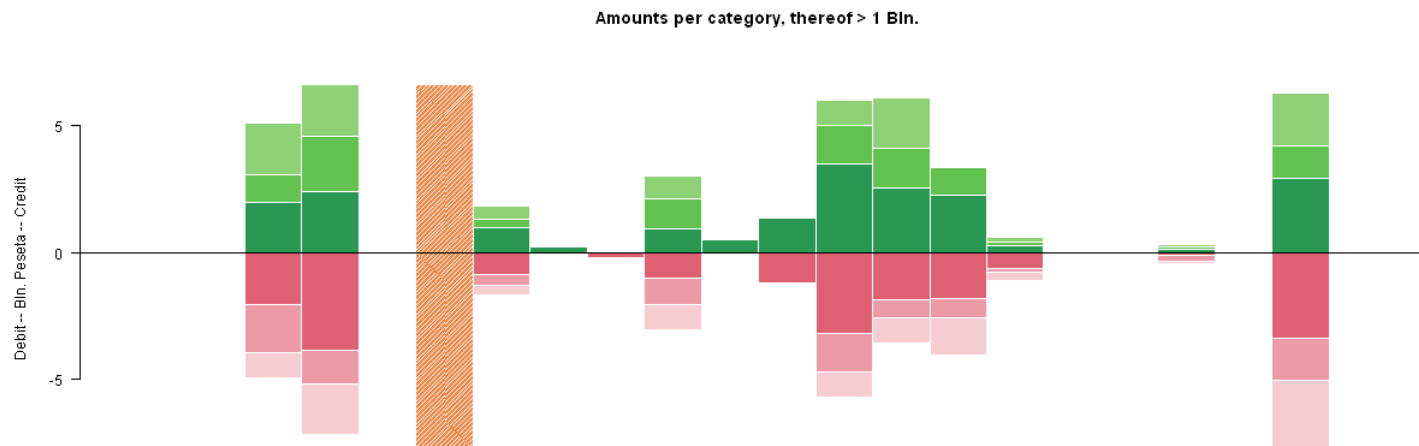
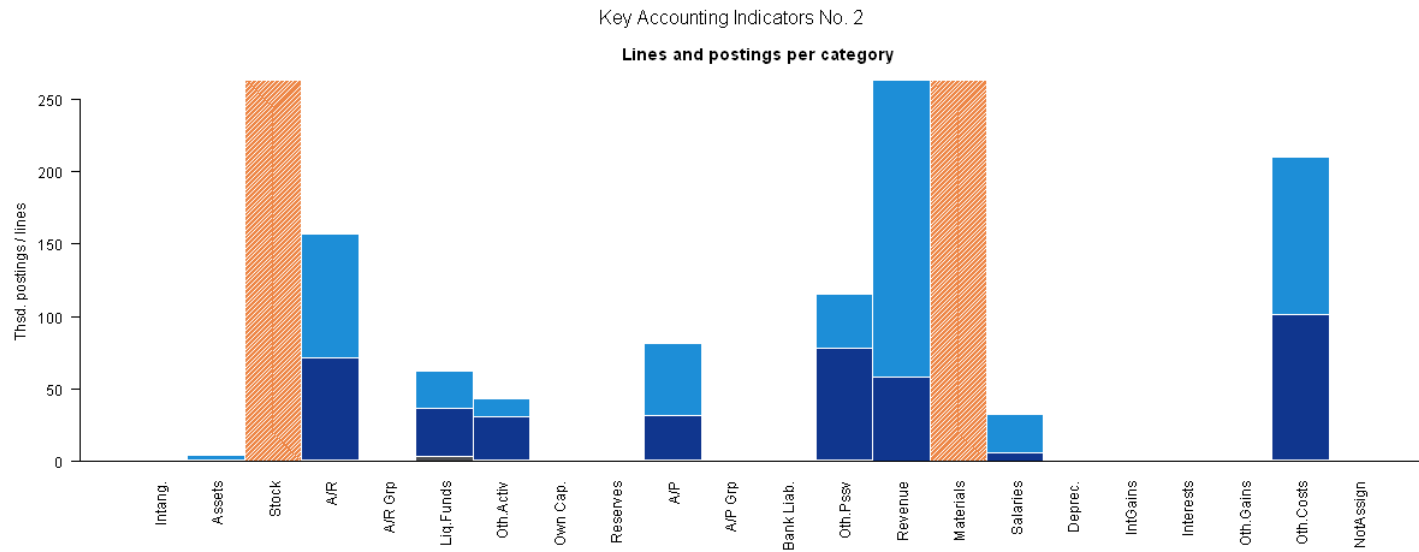
# Datenanalysen – Praktische Erfahrungen

## Kontenkategorisierung nach IAS 1.54

Aktiva		Passiva		G & V		Sonstiges	
<b>AIW</b>	Immaterielle Vermögenswerte	<b>PEK</b>	Eigenkapital	<b>GUM</b>	Umsatz	<b>SDP</b>	Durchlaufende Posten
<b>ASA</b>	Sachanlagen	<b>PBV</b>	Bankverbindlichkeiten	<b>GSE</b>	Sonstige Erträge	<b>SNZ</b>	Nicht zugeordnet
<b>AFI</b>	Finanzanlagen	<b>PRS</b>	Rückstellungen	<b>GBV</b>	Bestandsveränderungen		
<b>AVO</b>	Vorräte	<b>PLS</b>	Latente Steuern	<b>GMA</b>	Materialaufwand		
<b>AFL</b>	Forderungen aus Lieferungen und Leistungen	<b>PVL</b>	Verbindlichkeiten aus Lief. und Leistungen	<b>GPA</b>	Personalaufwand		
<b>AFV</b>	Forderungen an verbundene Unternehmen	<b>PVV</b>	Verbindlichkeiten gegenüber verbund. Unternehmen	<b>GAB</b>	Abschreibungen		
<b>ALM</b>	Liquide Mittel	<b>PSV</b>	Steuerverbindlichkeiten	<b>GSA</b>	Sonstiger Aufwand		
<b>ARP</b>	Aktive Rechnungsabgrenzung	<b>PRP</b>	Passive Rechnungsabgr.	<b>GZE</b>	Zinserträge		
<b>ASO</b>	Sonstige Aktiva	<b>PSO</b>	Sonstige Passiva	<b>GZI</b>	Zinsen		
				<b>GST</b>	Steuern		

# Datenanalysen – Praktische Erfahrungen

## Praxisbeispiel JET-62 „Große und lange Belege je Kategorie“



# Datenanalysen – Praktische Erfahrungen

## Journal Entry Testing – Auswahl

JET-00\_Gesamtübersicht

JET-11\_Gebuchte\_Belege\_je\_Periode

JET-12\_Erfaßte\_Zeilen\_je\_Periode

JET-13\_Erfaßte\_Belege\_je\_Tag\_im\_Monat

JET-14\_Erfaßte\_Belege\_je\_Stunde\_im\_Tag

JET-15\_Erfaßte\_Belege\_an\_WE\_und\_Feiertagen

JET-16\_Buchungen\_nach\_Periodenende

JET-16a\_Liste\_später\_Buchungen

JET-17\_Erfassung\_vor\_Belegdatum

JET-18\_Altersschichtung\_späte\_Bearbeitung

JET-19\_Belegnummersequenzen

JET-20\_Schichtung\_Beträge

JET-21\_Größte\_Beträge

JET-23\_Schichtung\_glatte\_Beträge

JET-24\_Benford

JET-26\_Wechselkurse\_Fremdwährungen

JET-27\_Übersicht\_Wechselkurse

JET-29\_Unausgeglichene\_Belege

JET-31\_Längste\_Belege

JET-33\_Häufigste\_Texte

JET-34\_Belege\_ohne\_Text\_manuell\_maschinell

JET-37\_Belegtexte\_mit\_sonstigen\_Stichworten

JET-40\_Kontenübersicht

JET-41\_Konten\_mit\_Ausreißern

JET-42\_Selten\_genutzte\_Konten

JET-43\_Besondere\_Konten

JET-44\_Konten\_nicht\_im\_Kontenplan

JET-49\_Schichtung\_negative\_Beträge

JET-50\_Periodenübersicht\_je\_Kostenstelle

JET-51\_Große\_und\_lange\_Belege\_je\_Periode

JET-52\_Salden\_durchlaufender\_Posten

JET-62\_Große\_und\_lange\_Belege\_je\_Kategorie

JET-63\_Schichtung\_Beträge\_je\_Kategorie

JET-65\_Schichtung\_Kategoriehäufigkeit

**JET-66\_Kategoriekombinationen**

**JET-68\_Kategorien\_Kreuztabelle**

# Datenanalysen – Praktische Erfahrungen

## Kontenkategorisierung – Übersicht

Category	AIW	ASA	AVD	AFL	AFV	ALM	ASO	PEK	PBV	PRS	PVL	PVV	PSV	PSD	GUM	GSE	GMA	GPA	GAB	GSA	GZE	GZI	
AIW	Rot								Rot	Rot													
ASA		Rot							Rot	Rot													
AVD			Rot						Rot	Rot													
AFL		Rot	Rot	Rot	Rot	Rot			Rot	Rot	Rot	Rot	Rot	Rot	Rot	Rot	Rot	Rot	Rot	Rot	Rot	Rot	Rot
AFV		Rot	Rot	Rot	Rot	Rot			Rot	Rot	Rot	Rot	Rot	Rot	Rot	Rot	Rot	Rot	Rot	Rot	Rot	Rot	Rot
ALM		Rot							Rot	Rot													
ASO									Rot	Rot													
PEK									Rot	Rot													
PBV									Rot	Rot													
PRS									Rot	Rot													
PVL									Rot	Rot													
PVV									Rot	Rot													
PSV									Rot	Rot													
PSD									Rot	Rot													
GUM									Rot	Rot													
GSE									Rot	Rot													
GMA									Rot	Rot													
GPA									Rot	Rot													
GAB									Rot	Rot													
GSA									Rot	Rot													
GZE									Rot	Rot													
GZI									Rot	Rot													

### Legende

- Grün      Übliche Kombination
- Gelb      Seltene Kombination
- Orange    Ungewöhnliche Kombination
- Rot        Kritische Kombination
- Grau      Umbuchung innerhalb Kategorie



# Datenanalysen – Praktische Erfahrungen

## Kontenkategorisierung – Beispiel Auszug

Kategorie	GUM	GSE	GBV	GMA	GPA	GAB	GSA	GZE	GZI
AIW						9: 6.000			
ASA				1: 0		19: 26.000	1: 0		
AFI									
AVO									
AFL	242: 10.418.000	40: 19.000		1: 2.000			3: 0		
AFV	13: 141.000	78: 152.000		1: 0			1: 0		1: 0
ALM	1: 0	2: 9.000		1: 0	1: 1.000		5: 1.000		
ARP		1: 0		10: 6.000			15: 9.000		
ASO		1: 0			16: 18.000				
PEK									
PBV									
PRS		11: 153.000		40: 366.000	44: 4.894.000		71: 377.000		
PLS									
PVL	1: 0	1: 2.000		264: 34.000		2: 0	70: 50.000		
PVV							2: 1.000		
PSV		14: 0		59: 390.000	1: 0	2: 0	205: 17.000		
PRP		1: 2.000					2: 13.000		
PSO		2: 0			20: 32.000				
GUM	9: 74.000	1: 0		95: 680.000	1: 0	1: 0	7: 82.000		
GSE	2: 0	22: 68.000		19: 30.000	18: 62.000	19: 62.000	22: 62.000		
GBV									
GMA	174: 1.447.000	58: 30.000		100: 504.000	177: 93.000	118: 71.000	325: 489.000		
GPA	5: 0	62: 63.000		52: 94.000	582: 2.317.000	170: 76.000	745: 689.000		
GAB	5: 0	61: 62.000		48: 71.000	342: 76.000	181: 79.000	535: 79.000		
GSA	32: 82.000	76: 74.000		93: 520.000	378: 689.000	180: 79.000	1067: 799.000		
GZE									
GZI									

### Legende

**Gelb** Seltene Kombination  
**Rot** Kritische Kombination

**Grün**  
**Orange**  
**Grau**

Übliche Kombination  
 Ungewöhnliche Kombination  
 Umbuchung innerhalb Kategorie

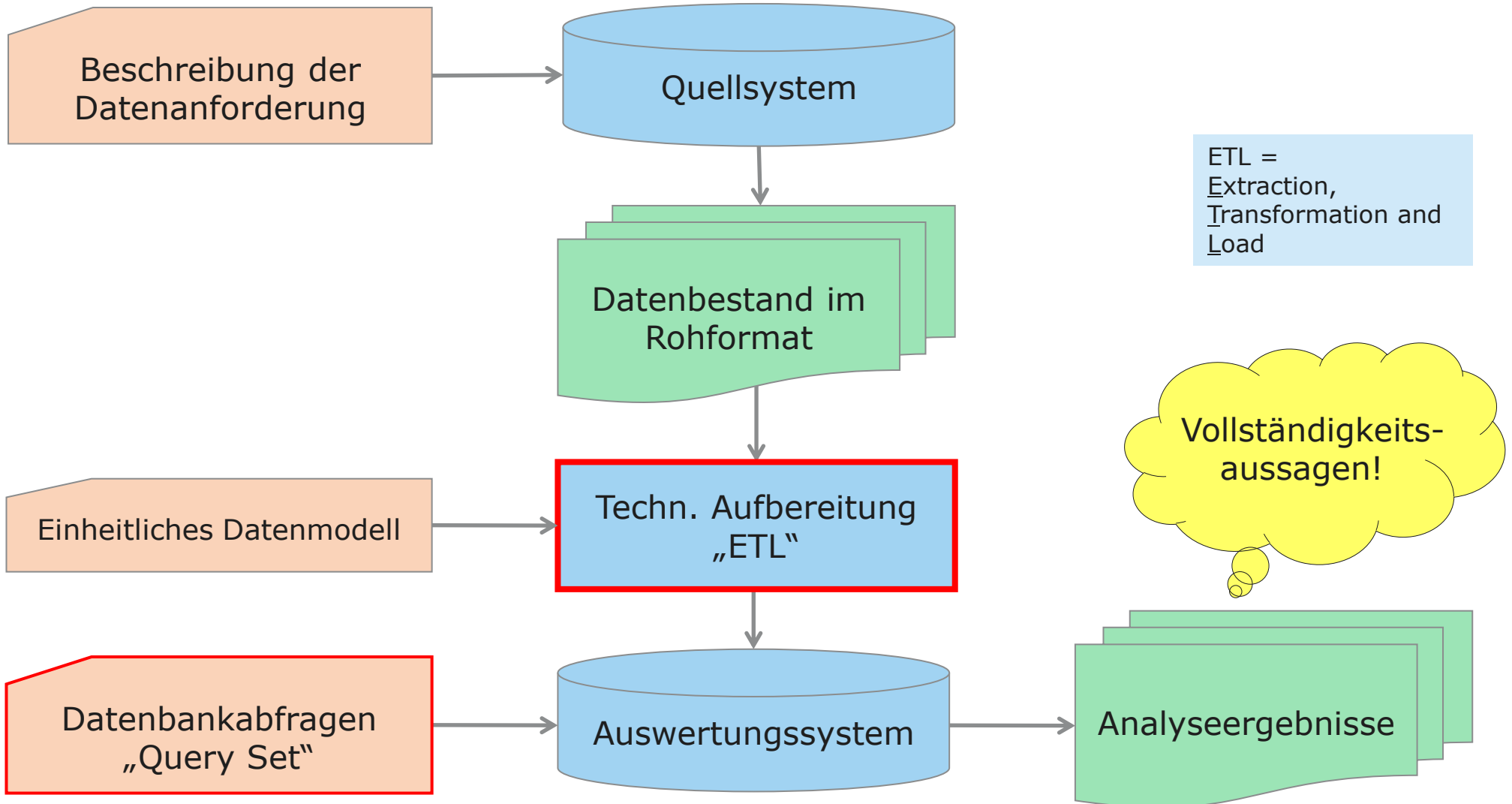
**Holtzbrinck Publishing Group**

Datenanalysen – Technische Durchführung

4

# Datenanalyse – Technische Durchführung

Ablauf einer Datenanalyse mit Datenextraktion



# Datenanalyse – Technische Durchführung

## Rohdaten

Cuenta	Fecha Doc.	T.D.	Documento	Tercero	Nombre del Tercero	Descripción del Movimie
1						
11						
1105						
110510						
110510.05	SALDO ANTERIOR				"\$4,000,000.00"	
	28/11/2013	AC	165	900444029	MACMILLAN PUBLISHERS SAS	RECLASIFICAMOS LOS
	\$0.00		"\$3,000,000.00"			
1110						
111005						
111005.05	SALDO ANTERIOR				"\$148,189,659.19"	
	01/01/2013	FP	496	900314828	PROYECTOS INVERSIONES SAN JOSE PLAZA	Fra. H
	03/01/2013	RC	187	830037946	PANAMERICANA LIBRERIA Y PAPELERIA SA	"CANCELACI
	08/01/2013	CE	922	800153993	COMUNICACION CELULAR SA	CANCELACION FACTUR
	09/01/2013	CE	980	444	MACMILLAN PUBLISHERS S.A. DE CV MEXICO	R.F.C
	14/01/2013	CE	924	900097575	CAJAS SANTAFE LTDA	CANCELACIÓN SALDO FAC 9
	17/01/2013	RC	191	900169783	S&O BOOKSTORE SAS	CANCELACIÓN FACTURA FC 0
	18/01/2013	RC	193	800137399	GRUPO KTDRA S.A.S	"CANCELACION FACTURAS
	18/01/2013	CE	932	900101529	LYN COMPUTADORES S.A.S	CANCELACIÓN FACTUR
	18/01/2013	CE	927	860512330	SERVIENTREGA S.A.	"CANCELACIÓN FACTURAS
	18/01/2013	CE	928	860066749	AGENCIA DE ADUANAS ADUANAMIENTOS LTDA NIVEL 1	
	18/01/2013	CE	926	80796561	CASTAÑEDA PRECIADO JOSE	
...						
17						
1705						
170520						
170520.05	SALDO ANTERIOR				"\$4,250,369.00"	
	31/01/2013	DA	38	860009578	SEGUROS DEL ESTADO S.A	Amortizaciones
	31/01/2013	DA	38	860039988	LIBERTY SEGUROS	Amortizaciones
	31/01/2013	DA	38	890903790	SURAMERICANA	Amortizaciones Enero 2013
	15/02/2013	FP	520	860039988	LIBERTY SEGUROS	Poliza 9476 - Seguro
...						

# Datenanalyse – Technische Durchführung

## Rohdaten

Cuenta	Fecha Doc.	T.D.	Documento	Tercero	Nombre del Tercero	Descripción del Movimie
1						
11						
1105						
110510						
110510.05	SALDO ANTERIOR				"\$4,000,000.00"	
	28/11/2013	AC	165	900444029	MACMILLAN PUBLISHERS SAS	RECLASIFICAMOS LOS
	\$0.00		"\$3,000,000.00"			
1110						
111005						
111005.05	SALDO ANTERIOR				"\$148,189,659.19"	
	01/01/2013	FP	496	900314828	PROYECTOS INVERSIONES SAN JOSE PLAZA	Fra. H
	03/01/2013	RC	187	830037946	PANAMERICANA LIBRERIA Y PAPELERIA SA	"CANCELACI
	08/01/2013	CE	922	800153993	COMUNICACION CELULAR SA	CANCELACION FACTUR
	09/01/2013	CE	980	444	MACMILLAN PUBLISHERS S.A. DE CV MEXICO	R.F.C
	14/01/2013	CE	924	900097575	CAJAS SANTAFE LTDA	CANCELACIÓN SALDO FAC 9
	17/01/2013	RC	191	900169783	S&O BOOKSTORE SAS	CANCELACIÓN FACTURA FC 0
	18/01/2013	RC	193	800137399	GRUPO KTDRA S.A.S	"CANCELACION FACTURAS
	18/01/2013	CE	932	900101529	LYN COMPUTADORES S.A.S	CANCELACIÓN FACTUR
	18/01/2013	CE	927	860512330	SERVIENTREGA S.A.	"CANCELACIÓN FACTURAS
	18/01/2013	CE	928	860066749	AGENCIA DE ADUANAS ADUANAMIENTOS LTDA NIVEL 1	
	18/01/2013	CE	926	80796561	CASTAÑEDA PRECIADO JOSE	
...						
17						
1705						
170520						
170520.05	SALDO ANTERIOR				"\$4,250,369.00"	
	31/01/2013	DA	38	860009578	SEGUROS DEL ESTADO S.A	Amortizaciones
	31/01/2013	DA	38	860039988	LIBERTY SEGUROS	Amortizaciones
	31/01/2013	DA	38	890903790	SURAMERICANA	Amortizaciones Enero 2013
	15/02/2013	FP	520	860039988	LIBERTY SEGUROS	Poliza 9476 - Seguro
...						

# Datenanalyse – Technische Durchführung

## Rohdaten

Cuenta	Fecha Doc.	T.D.	Documento	Tercero	Nombre del Tercero	Descripción del Movimie
1						
11						
1105						
110510						
110510.05	SALDO ANTERIOR				"\$4,000,000.00"	
	28/11/2013	AC	165	900444029	MACMILLAN PUBLISHERS SAS	RECLASIFICAMOS LOS
	\$0.00		"\$3,000,000.00"			
1110						
111005						
111005.05	SALDO ANTERIOR				"\$148,189,659.19"	
	01/01/2013	FP	496	900314828	PROYECTOS INVERSIONES SAN JOSE PLAZA	Fra. H
	03/01/2013	RC	187	830037946	PANAMERICANA LIBRERIA Y PAPELERIA SA	"CANCELACI
	08/01/2013	CE	922	800153993	COMUNICACION CELULAR SA	CANCELACION FACTUR
	09/01/2013	CE	980	444	MACMILLAN PUBLISHERS S.A. DE CV MEXICO	R.F.C
	14/01/2013	CE	924	900097575	CAJAS SANTAFE LTDA	CANCELACIÓN SALDO FAC 9
	17/01/2013	RC	191	900169783	S&O BOOKSTORE SAS	CANCELACIÓN FACTURA FC 0
	18/01/2013	RC	193	800137399	GRUPO KTDRA S.A.S	"CANCELACION FACTURAS
	18/01/2013	CE	932	900101529	LYN COMPUTADORES S.A.S	CANCELACIÓN FACTUR
	18/01/2013	CE	927	860512330	SERVIENTREGA S.A.	"CANCELACIÓN FACTURAS
	18/01/2013	CE	928	860066749	AGENCIA DE ADUANAS ADUANAMIENTOS LTDA NIVEL 1	
	18/01/2013	CE	926	80796561	CASTAÑEDA PRECIADO JOSE	
...						
17						
1705						
170520						
170520.05	SALDO ANTERIOR				"\$4,250,369.00"	
	31/01/2013	DA	38	860009578	SEGUROS DEL ESTADO S.A	Amortizaciones
	31/01/2013	DA	38	860039988	LIBERTY SEGUROS	Amortizaciones
	31/01/2013	DA	38	890903790	SURAMERICANA	Amortizaciones Enero 2013
	15/02/2013	FP	520	860039988	LIBERTY SEGUROS	Poliza 9476 - Seguro
...						

# Datenanalyse – Technische Durchführung

## Rohdaten

Cuenta	Fecha Doc.	T.D.	Documento	Tercero	Nombre del Tercero	Descripción del Movimie
1						
11						
1105						
110510						
110510.05	<b>SALDO ANTERIOR</b>				<b>"\$4,000,000.00"</b>	
	28/11/2013	AC	165	900444029	MACMILLAN PUBLISHERS SAS	RECLASIFICAMOS LOS
	<b>\$0.00</b>				<b>"\$3,000,000.00"</b>	
1110						
111005						
111005.05	<b>SALDO ANTERIOR</b>				<b>"\$148,189,659.19"</b>	
	01/01/2013	FP	496	900314828	PROYECTOS INVERSIONES SAN JOSE PLAZA	Fra. H
	03/01/2013	RC	187	830037946	PANAMERICANA LIBRERIA Y PAPELERIA SA	"CANCELACI
	08/01/2013	CE	922	800153993	COMUNICACION CELULAR SA	CANCELACION FACTUR
	09/01/2013	CE	980	444	MACMILLAN PUBLISHERS S.A. DE CV MEXICO	R.F.C
	14/01/2013	CE	924	900097575	CAJAS SANTAFE LTDA	CANCELACIÓN SALDO FAC 9
	17/01/2013	RC	191	900169783	S&O BOOKSTORE SAS	CANCELACIÓN FACTURA FC 0
	18/01/2013	RC	193	800137399	GRUPO KTDRA S.A.S	"CANCELACION FACTURAS
	18/01/2013	CE	932	900101529	LYN COMPUTADORES S.A.S	CANCELACIÓN FACTUR
	18/01/2013	CE	927	860512330	SERVIENTREGA S.A.	"CANCELACIÓN FACTURAS
	18/01/2013	CE	928	860066749	AGENCIA DE ADUANAS ADUANAMIENTOS LTDA NIVEL 1	
	18/01/2013	CE	926	80796561	CASTAÑEDA PRECIADO JOSE	
...						
17						
1705						
170520						
170520.05	<b>SALDO ANTERIOR</b>				<b>"\$4,250,369.00"</b>	
	31/01/2013	DA	38	860009578	SEGUROS DEL ESTADO S.A	Amortizaciones
	31/01/2013	DA	38	860039988	LIBERTY SEGUROS	Amortizaciones
	31/01/2013	DA	38	890903790	SURAMERICANA	Amortizaciones Enero 2013
	15/02/2013	FP	520	860039988	LIBERTY SEGUROS	Poliza 9476 - Seguro
...						

# Datenanalyse – Technische Durchführung

## Umformung mittels Perl-Skript

Cuenta	Fecha Doc.	T.D.	Documento	Tercero	Nombre del Tercero	Descripción del Movimie
1						
11						
1105						
110510						
110510.05						
1110						
111005						
111005.05						
...						
17						
1705						
170520						
170520.05						
...						

```
#!/bin/perl -w

print "Kontenjournal umwandeln\n";
# ...
open DATEI, '<' . $datei || die "Datei nicht gefunden.\n";
open JRN, '>imp_journal.txt';

$_ = <DATEI>; # Spaltenköpfe
print JRN $_;

while (<DATEI>) {
    chomp;
    $lines++;

    if (m/^\d*\t/) {
#         ...
        print JRN $_, "\n";
    }
    elsif (m/^\t\d/) {
#         ...
        print JRN $konto, $_, scalar @felder >= 13 ? "\n" : '';
    }
    elsif (m/^[0-9.]+\t/) {
        ($konto, $rest) = split /\t/;
    }
    elsif (m/^\t\t\t\t\t/) {
        print 'Zeile ', $lines, " ist eine Saldozeile.\n" unless $QUIET;
    }
    elsif (m/^\$/) {
        print 'Zeile ', $lines, " ist eine Leerzeile.\n" unless $QUIET;
    }
    else {
        print 'Zeile ', $lines, ' ist anders: >>', substr($_, 0, 10), "<<\n";
    }
}

close DATEI;
close JRN;

...

print 'Fertig nach ', $lines, " Zeilen.\n";
```



# Datenanalyse – Technische Durchführung

## Ergebnis der Umformung

Konto	Datum	Belegart	BelegNr	GeschPart	GeschPName	Beschreibung
110510.05	28.11.2013	AC	165	900444029	MACMILLAN PUBLISHERS SAS	RECLASIFI
111005.05	01.01.2013	FP	496	900314828	PROYECTOS INVERSIONES SAN JOSE PLAZ	
111005.05	03.01.2013	RC	187	830037946	PANAMERICANA LIBRERIA Y PAPELERIA SA	
111005.05	08.01.2013	CE	922	800153993	COMUNICACION CELULAR SA	CANCELAC
111005.05	09.01.2013	CE	980	444	MACMILLAN PUBLISHERS S.A. DE CV M	F.C M
111005.05	14.01.2013	CE	924	900097575	CAJAS SANTAFE LTDA	
111005.05	14.01.2013	CE	923	860515627	INVERSIONES LESA DE COLO	CAN
111005.05	17.01.2013	CE	925	900285666	VIGIA PLUS SERVICES SA	CANCELACIÓN F
111005.05	17.01.2013	RC	191	900169783	S&O BOOKSTORE SAS	CANCELACIÓN F
111005.05	18.01.2013	RC	193	800137399	GRUPO KTDRA S.A.S	CANCELACION FACT
111005.05	18.01.2013	CE	932	900101529	LYN COMPUTADORES S.A.S	CANCELAC
111005.05	18.01.2013	CE	927	860512330	SERVIENTREGA S.A.	CANCELACIÓN FACT
111005.05	18.01.2013	CE	928	860066749	AGENCIA DE ADUANAS ADUANAMIENTOS L	
111005.05	18.01.2013	CE	926	80796561	CASTAÑEDA PRECIADO JOSE DAVID	LIQUIDAC
111005.05	21.01.2013	RC	198	830037946	PANAMERICANA LIBRERIA Y PAPELERIA SA	
111005.05	21.01.2013	CE	933	5653901	PLATA DELGADO JOSE JOAQUIN ANTICIPO MONT	
111005.05	25.01.2013	CE	978	067708645	MACMILLAN PUBLISHERS BRASIL	Pago

Nachher

Cuenta	Fecha Doc.	T.D.	Documento	Tercero	Nombre del Tercer
11					
1105					
110510					
110510.05	SALDO ANTERIOR				"\$4,000,000.00"
	28/11/2013	AC	165	900444029	MACMILLAN PUBLISHERS SAS
					\$0.00
					"\$2,000,000.00"
1110					
111005					
111005.05	SALDO ANTERIOR				"\$148,189,659.19"
	01/01/2013	FP	496	900314828	PROYECTOS INVERSI
	03/01/2013	RC	187	830037946	PANAMERICANA LIBE
	08/01/2013	CE	922	800153993	COMUNICACION
	09/01/2013	CE	980	444	MACMILLAN PUBLISHERS S.A. DE CV M
	14/01/2013	CE	924	900097575	CAJAS
	17/01/2013	RC	191	900169783	S&O BOOKSTORE SAS
	18/01/2013	RC	193	800137399	GRUPO KTDRA S.A.S
	18/01/2013	CE	932	900101529	LYN COMPUTADORES
	18/01/2013	CE	927	860512330	SERVIENTREGA S.A.
	18/01/2013	CE	928	860066749	AGENCIA DE ADUANA
	18/01/2013	CE	926	80796561	CASTAÑEDA PRECIAD

Vorher

# Datenanalyse – Technische Durchführung

## Andere Rohdaten

```
"Date:"      Thursday, September 26, 2013      "Macmillan"  "Page:"      "of"      1
"Time:"      02:22PM      "GL Transactions - Posted"  "Report:"    "01680P.rpt"
"User:"      „RQUINON"      "Periods: 01-12 Through 06-13 As of: 9/26/2013"      "Company:"    "HB PUB"
```

```
"Ledger ID:" "ACTUAL"
```

"Per"	"Jrnl"	"Original"	"Tran"	"Batch"	"Per"	"Reference"	"Tran"	"Transaction,"
"Post"	"Type"	"Company"	"Account"	"Subaccount"	"Type"	"Number"	"Entr"	"Number,"
"01-12"	" "	"GJ"	"HB PUB"	"1009000000"	"21-905-000"	"GL"	"081873"	"01-12,"
"01-13"	" "	"GJ"	"HB PUB"	"1009000000"	"21-905-000"	"GL"	"087024"	"01-13"
"Account Sub Total"		2,351.89	1,304.59					

"01-12"	" "	"GJ"	"HB PUB"	"1009009012"	"21-905-000"	"GL"	"081845"	"01-12"
"01-12"	" "	"GJ"	"HB PUB"	"1009009012"	"21-905-000"	"GL"	"081849"	"01-12"
"01-12"	" "	"GJ"	"HB PUB"	"1009009012"	"21-905-000"	"GL"	"081873"	"01-12"
"02-12"	" "	"GJ"	"HB PUB"	"1009009012"	"21-905-000"	"GL"	"082202"	"02-12"
"03-12"	" "	"GJ"	"HB PUB"	"1009009012"	"21-905-000"	"GL"	"082614"	"03-12"
"03-12"	" "	"GJ"	"HB PUB"	"1009009012"	"21-905-000"	"GL"	"082693"	"03-12"
"04-12"	" "	"GJ"	"HB PUB"	"1009009012"	"21-905-000"	"GL"	"082693"	"03-12"
"05-12"	" "	"GJ"	"HB PUB"	"1009009012"	"21-905-000"	"GL"	"082693"	"03-12"
"06-12"	" "	"GJ"	"HB PUB"	"1009009012"	"21-905-000"	"GL"	"082693"	"03-12"
"Account Sub Total"		13,37						

```
➤ grep -P '^"[01][0-9]-1[23]"' export.txt > daten.csv
```

"01-12"	" "	"GJ"	"HB PUB"	"1039000000"	"21-905-000"	"GL"	"081873"	"01-12"
"01-12"	" "	"GJ"	"HB PUB"	"1039000000"	"21-905-000"	"GL"	"081873"	"01-12"
"01-13"	" "	"GJ"	"HB PUB"	"1039000000"	"21-905-000"	"GL"	"087024"	"01-13"
"01-13"	" "	"GJ"	"HB PUB"	"1039000000"	"21-905-000"	"GL"	"087024"	"01-13"
"Account Sub Total"		16,488,997.70	16,488,997.70					

"01-12"	" "	"GJ"	"HB PUB"	"1039009001"	"21-905-000"	"GL"	"081841"	"01-12"
"01-12"	" "	"GJ"	"HB PUB"	"1039009001"	"21-905-000"	"GL"	"081873"	"01-12"
"02-12"	" "	"GJ"	"HB PUB"	"1039009001"	"21-905-000"	"GL"	"082179"	"02-12,"

# Datenanalyse – Technische Durchführung

## Andere Rohdaten

```
"01-12"      " "      "GJ"      "HB PUB"      "1009000000" "21-905-000" "GL"      "081873"      "01-12"
"01-13"      " "      "GJ"      "HB PUB"      "1009000000" "21-905-000" "GL"      "087024"      "01-13"
"01-12"      " "      "GJ"      "HB PUB"      "1009009012" "21-905-000" "GL"      "081845"      "01-12"
"01-12"      " "      "GJ"      "HB PUB"      "1009009012" "21-905-000" "GL"      "081849"      "01-12"
"01-12"      " "      "GJ"      "HB PUB"      "1009009012" "21-905-000" "GL"      "081873"      "01-12"
"02-12"      " "      "GJ"      "HB PUB"      "1009009012" "21-905-000" "GL"      "082202"      "02-12"
"03-12"      " "      "GJ"      "HB PUB"      "1009009012" "21-905-000" "GL"      "082614"      "03-12"
"03-12"      " "      "GJ"      "HB PUB"      "1009009012" "21-905-000" "GL"      "082693"      "03-12"
"04-12"      " "      "GJ"      "HB PUB"      "1009009012" "21-905-000" "GL"      "083059"      "04-12"
"05-12"      " "      "GJ"      "HB PUB"      "1009009012" "21-905-000" "GL"      "083704"      "05-12"
"06-12"      " "      "GJ"      "HB PUB"      "1009009012" "21-905-000" "GL"      "083955"      "06-12"
"01-12"      " "      "GJ"      "HB PUB"      "1039000000" "21-905-000" "GL"      "081873"      "01-12"
"01-12"      " "      "GJ"      "HB PUB"      "1039000000" "21-905-000" "GL"      "081873"      "01-12"
"01-13"      " "      "GJ"      "HB PUB"      "1039000000" "21-905-000" "GL"      "087024"      "01-13"
"01-13"      " "      "GJ"      "HB PUB"      "1039000000" "21-905-000" "GL"      "087024"      "01-13"
"01-12"      " "      "GJ"      "HB PUB"      "1039009001" "21-905-000" "GL"      "081841"      "01-12"
"01-12"      " "      "GJ"      "HB PUB"      "1039009001" "21-905-000" "GL"      "081873"      "01-12"
"02-12"      " "      "GJ"      "HB PUB"      "1039009001" "21-905-000" "GL"      "081873"      "01-12"
"10-12"      " "      "GJ"      "HB PUB"      "1039009001" "21-905-000" "GL"      "081873"      "01-12"
"11-12"      " "      "GJ"      "HB PUB"      "1039009001" "21-905-000" "GL"      "081873"      "01-12"
"12-12"      " "      "GJ"      "HB PUB"      "1039009001" "21-905-000" "GL"      "081873"      "01-12"
"01-13"      " "      "GJ"      "HB PUB"      "1039009001" "21-905-000" "GL"      "081873"      "01-12"
"01-13"      " "      "GJ"      "HB PUB"      "1039009001" "21-905-000" "GL"      "081873"      "01-12"
"02-13"      " "      "GJ"      "HB PUB"      "1039009001" "21-905-000" "GL"      "081873"      "01-12"
"03-13"      " "      "GJ"      "HB PUB"      "1039009001" "21-905-000" "GL"      "081873"      "01-12"
"04-13"      " "      "GJ"      "HB PUB"      "1039009001" "21-905-000" "GL"      "081873"      "01-12"
"05-13"      " "      "GJ"      "HB PUB"      "1039009001" "21-905-000" "GL"      "081873"      "01-12"
"06-13"      " "      "GJ"      "HB PUB"      "1039009001" "21-905-000" "GL"      "081873"      "01-12"
"01-12"      " "      "GJ"      "HB PUB"      "1045000000" "21-905-000" "GL"      "081873"      "01-12"
"01-12"      " "      "GJ"      "HB PUB"      "1045000000" "21-905-000" "GL"      "081873"      "01-12"
"01-12"      " "      "GJ"      "HB PUB"      "1045000000" "21-905-000" "GL"      "081873"      "01-12"
"01-13"      " "      "GJ"      "HB PUB"      "1045000000" "21-905-000" "GL"      "087024"      "01-13"
. . .
```

```
➤ grep -P '^"[01][0-9]-1[23]"' export.txt > daten.csv
```

```
➤ wc -l export.txt daten.csv
```

```
34581 export.txt
```

```
33717 daten.csv
```

```
68298 total
```

# Datenanalyse – Technische Durchführung

## Andere Rohdaten

```
"01-12"      " "      "GJ"      "HB PUB"      "1009000000"  "21-905-000"  "GL"      "081873"      "01-12"
"01-13"      " "      "GJ"      "HB PUB"      "1009000000"  "21-905-000"  "GL"      "087024"      "01-13"
"01-12"      " "      "GJ"      "HB PUB"      "1009009012"  "21-905-000"  "GL"      "081845"      "01-12"
"01-12"      " "      "GJ"      "HB PUB"      "1009009012"  "21-905-000"  "GL"      "081849"      "01-12"
"01-12"      " "      "GJ"      "HB PUB"      "1009009012"  "21-905-000"  "GL"      "081873"      "01-12"
"02-12"      " "      "GJ"      "HB PUB"      "1009009012"  "21-905-000"  "GL"      "082202"      "02-12"
"03-12"      " "      "GJ"      "HB PUB"      "1009009012"  "21-905-000"  "GL"      "082614"      "03-12"
"03-12"      " "      "GJ"      "HB PUB"      "1009009012"  "21-905-000"  "GL"      "082693"      "03-12"
"04-12"      " "      "GJ"      "HB PUB"      "1009009012"  "21-905-000"  "GL"      "083059"      "04-12"
"05-12"      " "      "GJ"      "HB PUB"      "1009009012"  "21-905-000"  "GL"      "083704"      "05-12"
"06-12"      " "      "GJ"      "HB PUB"      "1009009012"  "21-905-000"  "GL"      "083955"      "06-12"
"01-12"      " "      "GJ"      "HB PUB"      "1039000000"  "21-905-000"  "GL"      "081873"      "01-12"
"01-12"      " "      "GJ"      "HB PUB"      "1039000000"  "21-905-000"  "GL"      "081873"      "01-12"
"01-13"      " "      "GJ"      "HB PUB"      "1039000000"  "21-905-000"  "GL"      "087024"      "01-13"
"01-13"      " "      "GJ"      "HB PUB"      "1039000000"  "21-905-000"  "GL"      "087024"      "01-13"
"01-12"      " "      "GJ"      "HB PUB"      "1039009001"  "21-905-000"  "GL"      "081841"      "01-12"
"01-12"      " "      "GJ"      "HB PUB"      "1039009001"  "21-905-000"  "GL"      "081873"      "01-12"
"02-12"      " "      "GJ"      "HB PUB"      "1045000000"  "21-905-000"  "GL"      "081873"      "01-12"
"10-12"      " "      "GJ"      "HB PUB"      "1045000000"  "21-905-000"  "GL"      "087024"      "01-13"
"11-12"      " "      "GJ"      "HB PUB"      "1045000000"  "21-905-000"  "GL"      "087024"      "01-13"
"12-12"      " "      "GJ"      "HB PUB"      "1045000000"  "21-905-000"  "GL"      "081841"      "01-12"
"01-13"      " "      "GJ"      "HB PUB"      "1045000000"  "21-905-000"  "GL"      "081873"      "01-12"
"01-13"      " "      "GJ"      "HB PUB"      "1045000000"  "21-905-000"  "GL"      "081873"      "01-12"
"02-13"      " "      "GJ"      "HB PUB"      "1045000000"  "21-905-000"  "GL"      "081873"      "01-12"
"03-13"      " "      "GJ"      "HB PUB"      "1045000000"  "21-905-000"  "GL"      "081873"      "01-12"
"04-13"      " "      "GJ"      "HB PUB"      "1045000000"  "21-905-000"  "GL"      "081873"      "01-12"
"05-13"      " "      "GJ"      "HB PUB"      "1045000000"  "21-905-000"  "GL"      "081873"      "01-12"
"06-13"      " "      "GJ"      "HB PUB"      "1045000000"  "21-905-000"  "GL"      "081873"      "01-12"
"01-12"      " "      "GJ"      "HB PUB"      "1045000000"  "21-905-000"  "GL"      "081873"      "01-12"
"01-12"      " "      "GJ"      "HB PUB"      "1045000000"  "21-905-000"  "GL"      "081873"      "01-12"
"01-12"      " "      "GJ"      "HB PUB"      "1045000000"  "21-905-000"  "GL"      "081873"      "01-12"
"01-12"      " "      "GJ"      "HB PUB"      "1045000000"  "21-905-000"  "GL"      "081873"      "01-12"
"01-13"      " "      "GJ"      "HB PUB"      "1045000000"  "21-905-000"  "GL"      "087024"      "01-13"
. . .
```

```
➤ grep -P '^"[01][0-9]-1[23]"' export.txt > daten.csv
```

```
➤ wc -l export.txt daten.csv
```

```
34581 export.txt
```

```
33717 daten.csv
```

```
68298 total
```

```
➤ perl analyzeCSV.pl -s '\t' datei.csv
```

# Datenanalyse – Technische Durchführung

## Andere Rohdaten

```
"01-12"      " "      "GJ"      "HBPUB"      "1009000000" "21-905-000" "GL"      "081873"      "01-12"
"01-13"      " "      "GJ"      "HBPUB"      "1009000000" "21-905-000" "GL"      "087024"      "01-13"
"01-12"      " "      "GJ"
"01-12"      " "      "GJ"
"01-12"      " "      "GJ"
"02-12"      " "      "GJ"
"03-12"      " "      "GJ"
"03-12"      " "      "GJ"
"04-12"      " "      "GJ"
"05-12"      " "      "GJ"
"06-12"      " "      "GJ"
"01-12"      " "      "GJ"
"01-12"      " "      "GJ"
"01-13"      " "      "GJ"
"01-13"      " "      "GJ"
"01-12"      " "      "GJ"
"01-12"      " "      "GJ"
"02-12"      " "      "GJ"
"10-12"      " "      "GJ"
"11-12"      " "      "GJ"
"12-12"      " "      "GJ"
"01-13"      " "      "GJ"
"01-13"      " "      "GJ"
"02-13"      " "      "GJ"
"03-13"      " "      "GJ"
"04-13"      " "      "GJ"
"05-13"      " "      "GJ"
"06-13"      " "      "GJ"
"01-12"      " "      "GJ"
"01-12"      " "      "GJ"
"01-12"      " "      "GJ"
"01-12"      " "      "GJ"
"01-13"      " "      "GJ"
. . .
```

```
➤ perl analyzeCSV.pl -s '\t' datei.csv
Struktur einer Datei ermitteln
-- Lese Datei datei.csv
-- Keine Titelzeile
-- Datei enthaelt Feldbegrenzer ab Zeile 1
-- Zeile 27 hat nur 4 Feld(er), anstatt 9
-- Zeile 144 hat nur 4 Feld(er), anstatt 9
-- Zeile 234 hat nur 4 Feld(er), anstatt 9
```

Lfd	Feldname	Typ	Brt	Len	Hinweis
1	Feld_1	T	17	5	
2	Feld_2	R	10	10	Ab Zl. 27
3	Feld_3	T	10	2	
4	Feld_4	T	7	5	
5	Feld_5	L	12	10	
6	Feld_6	?	12	10	
7	Feld_7	T	4	2	
8	Feld_8	I	8	6	
9	Feld_9	T	7	5	

```
Fertig nach 33717 Datenzeilen in 33717 Dateizeilen.
```

# Datenanalyse – Technische Durchführung

## Vorbereitung und Werkzeuge

### Vorbereitung und Validierung des Datenmaterials

- Auswahl und Anforderung der Daten,
- Technische Datenaufbereitung,
- Plausibilitätskontrollen zur Echtheit / Vollständigkeit / Richtigkeit.

### Werkzeugkasten des Analysten

- Text-Viewer bzw. Editoren: SciTe, VIM, Total Commander Viewer [F3],
- Skripte: bash mit GNU-Utills, Perl,
- (Fertige Analyseprogramme: ACL, IDEA, und viele mehr),
- Data Warehouse, Datenbank mit Analyseroutinen,
- Visualisierungsprogramme, zB. R,
- **Kritische Grundeinstellung.**

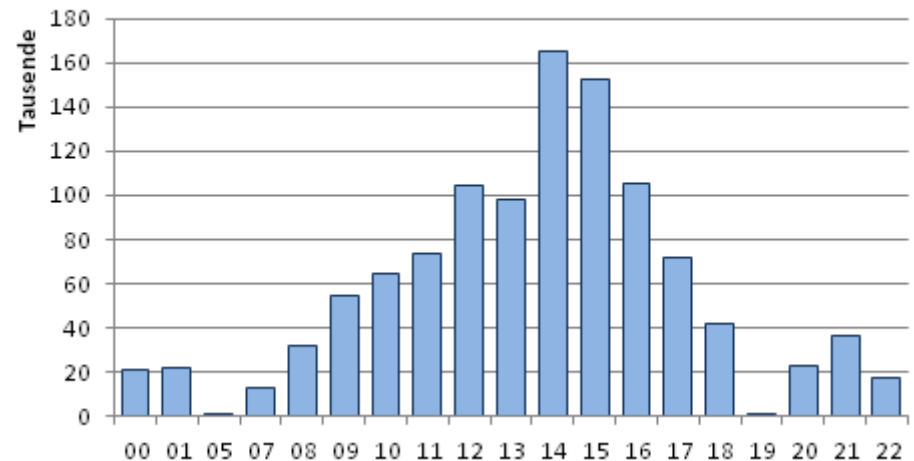
# Datenanalyse – Technische Durchführung

## Analysetechniken

### Datenbankanalysen

- Query Sets zur Übersicht und Qualitätssicherung,
- Explorative Datenanalyse bildet Grundverständnis,
- Vertiefung durch individuelle Abfragen,
- Modellierung von Handlungsmustern,
- Wie entsteht eine qualifizierte Negativ-Aussage?
- Visuelle Analyse und Ergebnisdarstellung.

Buchungszeilen je Stunde des Tages



# Datenanalyse – Technische Durchführung

## Grundsatzfragen

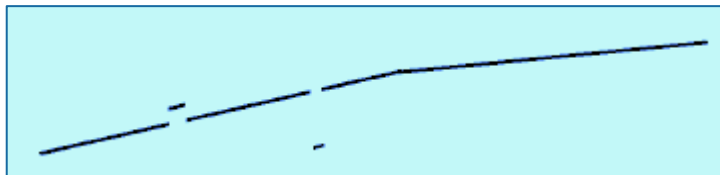
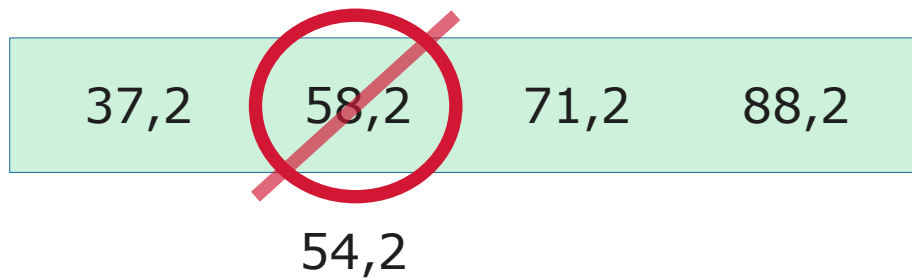
### **Entscheidung „Machen oder kaufen?“**

- Was kann man intern selbst?
- Was kauft man hinzu?
  - Komponenten
  - Dienstleistung
- Welchen Einblick hat man in fertige Komponenten?
- Welche Erweiterungsmöglichkeiten bestehen?
  - Wiederholbarkeit?



# Datenanalyse – Technische Durchführung

## Visualisierung



# Holtzbrinck Publishing Group

## Kontakt Daten

5

# Holtzbrinck Publishing Group

## Kontaktdaten



### **Jörg Meyer**

Leiter Interne Revision  
Holtzbrinck Publishing Group  
Stuttgart

[joerg.meyer@holtzbrinck.com](mailto:joerg.meyer@holtzbrinck.com)  
Telefon 0711 / 21 50-2 29