

Autorin: Dr. Heike Püttmann  
 Tabellen und Grafiken: Dr. Heike Püttmann

## Aufbereitung hierarchisch-lückenhafter Daten für Datenbank-gestützte Auswertungen Am Beispiel der Pendlerstatistik der Bundesagentur für Arbeit

*Hierarchien ermöglichen das Aufbohren von Daten*

Hierarchisch gegliederte Daten bieten die Möglichkeit des „Drills“, d.h. das Aufbohren von Daten, von einer hohen, allgemeinen auf eine niedrigere, detaillierte Aggregationsebene bzw. auf Detaildaten. Die Hierarchie kann auf unterschiedlichste Aspekte bezogen sein: räumlich (Bundesland → Regierungsbezirk → ...), zeitlich (Jahr → Quartal → ...) oder sachlich (Amt → Abteilung → ...).

*Datenlücken in Hierarchien müssen geschlossen werden*

Voraussetzung für die Drill-Fähigkeit sind lückenlose, Redundanz-freie Daten auf der untersten, detailliertesten Hierarchieebene. Die Daten selber werden nur auf der kleinsten Ebene für die Auswertung vorgehalten. Aus der Summe der Einzeldaten dieser niedrigsten Ebene müssen immer die Summen der nächsthöheren Aggregationsebenen errechnet werden können. Bei Datenlücken der Ausgangsdaten sind diese entsprechend anzureichern. In einer Datenbank ist es unerheblich, ob in der auszuwertenden Tabelle die gewünschte Hierarchieebene aus der Drill-Spalte durch ein SQL-Sub-SELECT ausgewählt wird oder die Hierarchieebenen in verschiedenen Drill-Spalten enthalten sind – die Berechnung entspricht immer der Summenbildung über einem Gruppierungsmerkmal (=Hierarchiemerkmal):

### Listing 1: SQL-Darstellung Summenbildung über Gruppierungsmerkmal (Oracle-SQL)

```
select substr(drillspalte,1,3) HIERARCHIE_A,
       sum(kennzahl) SUMME
from tabelle_einfach
group by substr(drillspalte,1,3);

select drillspalte_B HIERARCHIE_B,
       sum(kennzahl) SUMME
from tabelle_mehrfach
group by drillspalte_B;
```

In einer Tabellenkalkulation muss ein Sub-SELECT auf jeden Fall in einer eigenen Spalte zuvor vorgerechnet sein, um dann z.B. die Drill-Funktionalität von Pivot-Tabellen nutzen zu können (siehe Tabelle 1).

**Tabelle 1: Hierarchische Auswertung der Pendlerstatistik 2006 für München mit der Pivot-Funktion der Tabellenkalkulation CALC**

Berichtsmonat

Summe - INSGESAMT		PENDLERSCHLUESSEL		
BUNDESLAND	REGIERUNGSBEZIRK	Auspendler	Einpendler	Gesamt Ergebnis
9	91	89300	215015	304315
	92	1289	12673	13962
	93	717	4649	5366
	94	297	2589	2886
	95	1722	5659	7381
	96	499	2618	3117
	97	3022	21614	24636
<b>Gesamt Ergebnis</b>		<b>96846</b>	<b>264817</b>	<b>361663</b>



- Es sind für die oberen drei Hierarchieebenen des amtlichen Gemeindegchlüssels Zwischensummen aufgeführt (Bundesgebiet, Bundesland, Regierungsbezirk).
- Kreisfreie Städte können doppelt vorkommen, einmal mit achtstelligem und einmal mit fünfstelligem Gemeindegchlüssel (Bsp. „Ingolstadt“). In letzterer Fassung wird dem Gemeindegnamen „Stadt“ nachgestellt.
- Die fünfstelligen Schlüssel für kreisfreie Städte und Landkreise können zwei mal aufgeführt werden (Bsp. Schlüssel „09190“). In der einen Variante wird der Schlüssel für „Übrige Gemeinden (Kreis)“ verwendet, in der anderen Variante für den Kreis.
- Ferner gibt es noch ergänzende Datenangaben wie „Insgesamt“, „Bundesgebiet“, „davon aus den alten Bundesländern“, „davon aus den neuen Bundesländern“, „Übrige Bundesländer“, „darunter Ein-/Auspendler“ (=Gesamtsumme aller Pendler, bei denen sich Wohn- und Arbeitsort unterscheiden). Diese Zeilen haben anstelle des amtlichen Gemeindegchlüssels einen künstlichen Schlüssel 'ZZ'.
- Die Datenzeile mit amtlichem Gemeindegchlüssel Wohnort = '09162000' und amtlichem Gemeindegchlüssel Arbeitsort = '09162000' ist für die Münchener Pendlerstatistik nicht relevant, da diese die in München wohnenden und arbeitenden Personen darstellt.
- Tests haben schließlich gezeigt, dass die Summen einer kleineren Hierarchieebene nicht die angegebenen Kennzahlen der nächsthöheren ergeben. Dies resultiert aus den aus Datenschutzgründen mit „X“ maskierten Kennzahlen.

**Tabelle 2: Auszug aus Originaldatei Einpendler (Datenstand: 30.06.2011)**

 <b>Bundesagentur für Arbeit</b> Statistik		Wohnort	Insgesamt	Männer	Frauen	Deutsche	Ausländer + Sonstige	Auszubildende
			1	2	3	4	5	6
09162000	München, Landeshauptstadt	...						
		09161000 Ingolstadt	2 103	1 101	1 002	1 915	188	115
		09161 Ingolstadt, Stadt	2 103	1 101	1 002	1 915	188	115
		...						
		11000000 Berlin, Stadt	3 891	2 191	1 700	3 580	311	64
		11000 Berlin, Stadt	3 891	2 191	1 700	3 580	311	64
		...						
		09190 Übrige Gemeinden (Kreis)	33	21	12	X	X	3
		09190 Weilheim-Schongau	4 507	2 645	1 862	4 279	228	199
		...						
		091 Oberbayern	609 511	303 473	306 038	502 770	106 741	22 122
		09 Bayern	661 184	335 065	326 119	551 466	109 718	24 215
		...						
		IND Indien, einschl. Sikkim und G	10	X	X	7	3	X
		IND00000 Indien, einschl. Sikkim und G	10	X	X	7	3	X
		IND Indien	10	X	X	7	3	X
		IN Indien	10	X	X	7	3	X
		...						
		ZZ Übrige (Bundes)Länder	129	100	29	82	47	X
		ZZ BUNDESGBIET	707 706	364 641	343 065	594 487	113 219	25 181
	ZZ INSGESAMT	709 580	365 931	343 649	595 551	114 029	25 196	
	ZZ Darunter: EINPENDLER	325 204	183 042	142 162	298 778	26 426	12 144	
	ZZ davon: aus den ABL	309 826	172 963	136 863	284 788	25 038	11 739	
	ZZ davon: aus den NBL	13 504	8 789	4 715	12 926	578	390	

Ziel der Datenaufbereitung ist es, lückelose, Redundanz-freie Daten auf kleinster Raumhierarchieebene, also auf dem achtstelligen amtlichen Gemeindegchlüssel (=Gemeindeebene), zusammenzustellen.

Die Analyseergebnisse zeigen, dass dazu die Daten auf verschiedene Art und Weisen überarbeitet werden müssen:

1. „X“ wird durch „0“ ersetzt, damit Berechnungen erfolgen können
2. Auffüllen mit Datenzeilen mit künstlichen Schlüsseln und Bezeichnungen, die die Summendifferenzen einer niedrigeren räumlichen Hierarchieebene zur nächsthöheren füllen
3. Entfernen von Duplikaten
4. Entfernen von Zwischen- und Gesamtsummenzeilen
5. Entfernen der Datenzeile des Berichtsortes (=München).

### Datenanreicherung

Um die Datenlücken in der Hierarchie des amtlichen Gemeindegchlüssels zu schließen, werden durch Mengenoperationen künstliche Datensätze erzeugt. Hierzu werden durch Differenzrechnung für jede Kennzahl je Hierarchieebene die fehlenden Zahlen errechnet. Die neu entstandenen Datensätze werden mit künstlichen Schlüsseln und mit künstlicher Ortsbezeichnung bzw. Nationenbezeichnung (ab 2007) versehen.

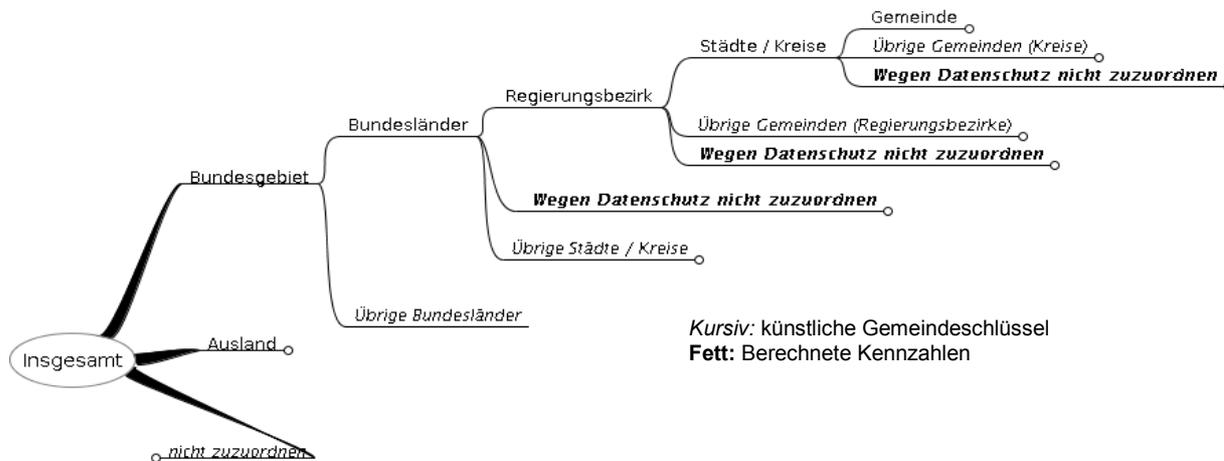
Es werden somit folgende Datensätze erstellt:

1. Gemeindegchlüsselformat: 99999AAA, Ortsbezeichnung: „Wegen Datenschutz nicht zuzuordnen“, wenn Kreis mit sechsstelligem AGS - Summe der dem Kreis zugehörigen Gemeinden mit achtstelligem AGS > 0
2. Gemeindegchlüsselformat: 999AAAAA, Ortsbezeichnung: „Wegen Datenschutz nicht zuzuordnen“, wenn Regierungsbezirk mit dreistelligem AGS - Summe der dem Regierungsbezirk zugehörigen Gemeinden mit achtstelligem AGS > 0
3. Gemeindegchlüsselformat: 99AAAAAA, Ortsbezeichnung: „Wegen Datenschutz nicht zuzuordnen“, wenn Bundesland mit zweistelligem AGS - Summe der dem Bundesland zugehörigen Gemeinden mit achtstelligem AGS > 0
4. Gemeindegchlüsselformat: AAAAAAAA, Ortsbezeichnung: „Wegen Datenschutz nicht zuzuordnen“, wenn Bund - Summe alle dem Bund zugehörigen Gemeinden mit achtstelligem AGS > 0
5. Künstlicher Kfz-Trennschlüssel (vgl. Kapitel Datenmodell) mit Nation: „nicht gefüllt“ und Ortsbezeichnung: „Wegen Datenschutz nicht zuzuordnen“, wenn (Insgesamt - Bund - übrige Bundesländer - keine Zuordnung möglich) - Ausland > 0

Die künstlichen amtlichen Gemeindegchlüssel für diese Datenzeilen sind allesamt achtstellig und damit auf Gemeindeebene. Im Ergebnis liegt damit eine komplette Hierarchie vor, wie sie Abbildung 1, Seite 20 dargestellt.

*Datensätze mit künstlichen Schlüsseln schließen Datenlücken*

Abbildung 1: hierarchische Gliederung des Raumbezuges der Pendlerdaten



Mehrstufige Verarbeitungsschritte führen zum Ziel

**Datenmodell**

Da eine Trennung nach Ein- und Auspendlern in zwei Datenmodelle nicht sinnvoll ist, werden diese in einem zusammengefasst. Um entweder die Ein- oder Auspendler auswerten zu können, muss lediglich nach dem entsprechenden Pendlerschlüssel ('E'=Einpendler bzw. 'A'=Auspendler) gefiltert werden.

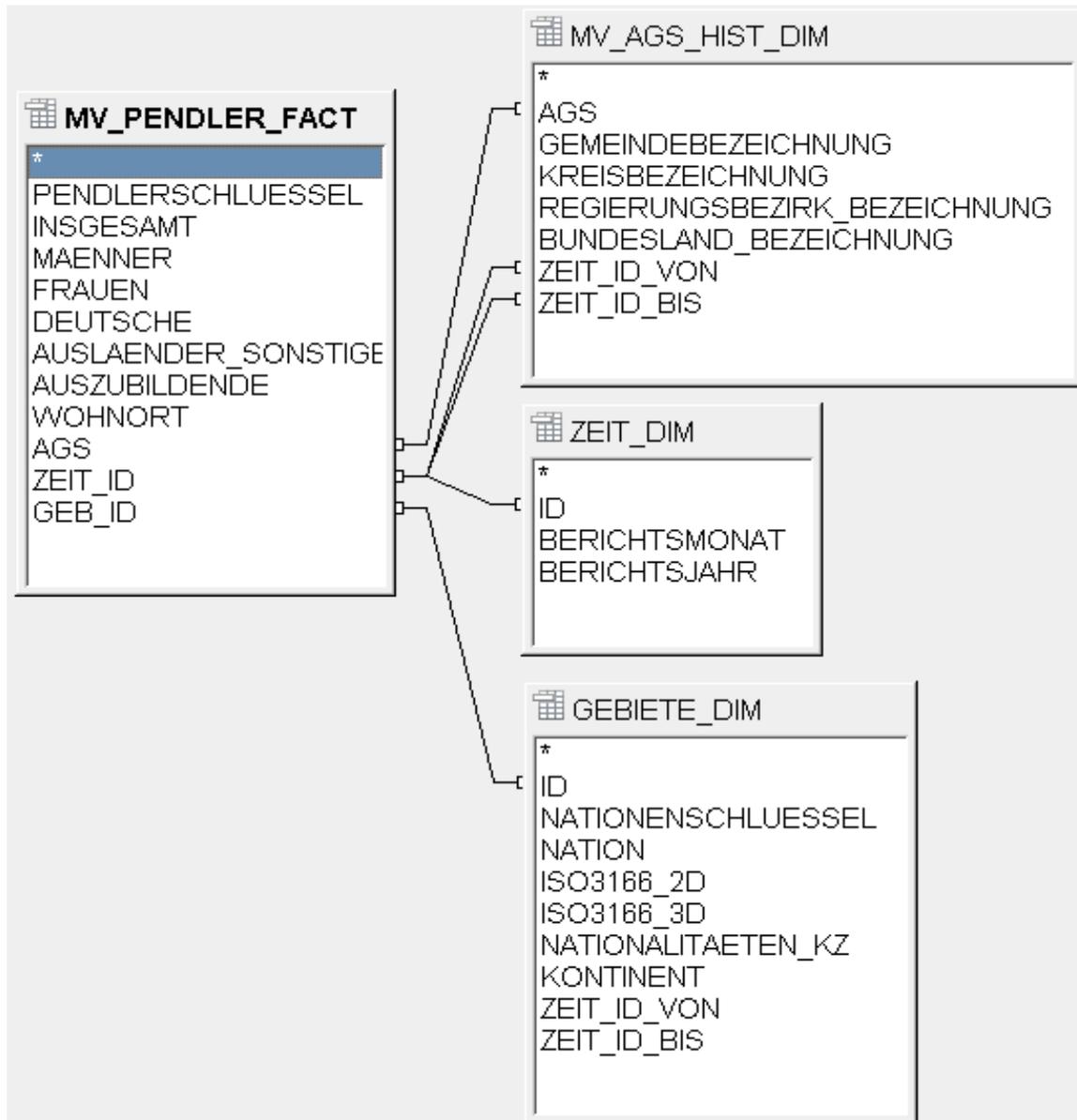
Für die verschiedenen Schritte der Datenverarbeitung werden unterschiedliche Tabellenstrukturen angelegt. Für den ersten Schritt, den Import der unveränderten Rohdaten in die Datenbank, ist eine Stagingtabelle (PENDLER\_STG) angelegt worden, welche die sechs Spalten der Kennzahlen sowie die zwei Spalten mit dem Raumbezug enthält.

Das Zwischenziel des Datenaufbereitungsprozesses ist ein (unvollständiges) Datenmodell mit einer Faktentabelle (PENDLER\_FACT), welche die noch redundanten, jedoch bereinigten Daten für alle Hierarchieebenen (= mit zwei-, drei-, fünf- und achstelligen Gemeindegchlüssel) zusammenstellt. Diese Form ist erforderlich, um die Berechnung der fehlenden Differenzsummen – wie im vorangegangenen Abschnitt Datenanreicherung beschrieben – bestimmen und hinzufügen zu können.

Da der amtliche Gemeindegchlüssel und die Kfz-Nationalitätenkennzeichen unabhängig voneinander sind, werden für den Raumbezug entsprechend zwei Dimensionsspalten benötigt: eine, die die innerdeutschen Gemeindegchlüssel enthält und eine weitere, welche die ausländischen Kfz-Nationalitätenkennzeichen abbildet. Erstere enthält die echten bzw. künstlichen Gemeindegchlüssel, letztere einen Fremdschlüssel auf die Dimensionstabelle GEBIETE\_DIM, welche neben dem Kfz-Nationalitätenkennzeichen noch weitere Information auf Staatenebene enthält. Ein Datensatz kann dabei immer nur einen von beiden Schlüsseln haben (innerdeutsch, aus dem Ausland).

Im nachfolgenden letzten Arbeitsschritt werden die Daten dann in ein Drillfähiges Datenformat gebracht (siehe Abbildung 2, Seite 21). Dabei werden aus der vorgenannten Faktentabelle PENDLER\_FACT des zuvor beschriebenen Datenmodells alle Datensätze mit achtstelligem Gemeindegchlüssel und ohne Kfz-Nationalitätenkennzeichen (ohne München) bzw. mit Kfz-Nationalitätenkennzeichen und ohne Gemeindegchlüssel extrahiert und in eine weitere Faktentabelle (MV\_PENDLER\_FACT) geschrieben.

Abbildung 2: Datenmodell Pendlerdaten



Auch gibt es in der Faktentabelle wieder eine Fremdschlüsselspalte zur Dimensionstabelle GEBIETE\_DIM. Die Spalte mit den Gemeindegeschlüsseln ist in diesem Modell zugleich Fremdschlüssel zur Dimensionstabelle MV\_AGS\_HIST\_DIM, welche alle Gemeindegeschlüssel in besonderer Form zusammenstellt: So sind neben den offiziellen amtlichen Gemeindegeschlüsseln (Quelle: <sup>2)</sup>) auch die künstlich generierten Gemeindegeschlüssel u.a. mit folgendem Muster (9 = Zahl zwischen 0-9, Buchstabe wie angegeben), enthalten :

*Muster der Gemeindegeschlüssel*

- AAAAAAAAA - Wegen Datenschutz nicht zuzuordnen
- 99AAAAAAAA - Wegen Datenschutz nicht zuzuordnen
- 999AAAAA - Wegen Datenschutz nicht zuzuordnen
- 99999AAA - Wegen Datenschutz nicht zuzuordnen
- NNNNNNNN - Übrige Bundesländer
- 99NNNNNN - Übrige Regierungsbezirke
- 999NNNNN - Übrige Kreise
- 99999NNN - Übrige Gemeinden
- XXXXXXXX - Keine Zuordnung möglich
- 99999999 - echte amtliche Gemeindegeschlüssel

Die Gemeindegeschlüssel in dieser Tabelle sind seit 1987 historisiert.

*Fast automatisierter Verarbeitungsablauf*

**Datenverarbeitung**

Für den Betrieb von ZIMAS (Zentrales Informationsmanagement- und Analysesystem), dem auf einer Oracle Datenbank 11g Enterprise Edition basierendem Data Warehouse des Statistischen Amtes der Landeshauptstadt München, wird ein möglichst automatisierter und standardisierter Betrieb angestrebt. So müssen zunächst die auf München eingegrenzten Daten der Excel-Datenlieferung noch manuell in eine CSV-Datei umgewandelt werden. Erst mit dem Import erfolgt die nachfolgende Auf- und Weiterverarbeitung (ETL = extraction, transformation, loading) vollautomatisch per Shell- und SQL-Skripte.

Die Rohdaten der CSV-Datei werden zunächst in die Stagingtabelle (PENDLER\_STG) geladen. Von hier aus werden die Daten mit etlichen Überarbeitungsregeln, u.a. aus Tabelle 3, siehe unten, überarbeitet und schließlich in die Faktentabelle PENDLER\_FACT geschrieben. Die durch im Rahmen der Aufbereitung entstehenden Duplikate werden entfernt (Deduplizierung). Für die korrekte Berechnung der Hierarchie enthält die Faktentabelle auch noch den Datensatz für die Berichtsgemeinde (=München). Anschließend werden Daten in dieser Tabelle dann mit der im Kapitel Datenanreicherung beschriebenen Methode ergänzt.

Die Tabellenbeziehungen des Datenmodells werden durch Primärschlüssel – Fremdschlüsselbeziehungen realisiert. Hierzu werden zu den Kfz-Nationalitätenkennzeichen in der Dimensionstabelle die Primärschlüssel nachgeschlagen (lookup) und als Fremdschlüssel in die Faktentabelle eingebunden (GEB\_ID). Die Datenspalte mit den amtlichen und um künstliche erweiterte Gemeindegemeinschaften ist zugleich Fremdschlüsselspalte.

**Tabelle 3: Wichtigste Verarbeitungsregeln der Pedlerdaten**

Überarbeitungsziel	SQL *)	Erläuterung SQL
Wohnort / Herkunftsland vereinheitlichen auf erstes Wort	substr(WOHNORT ,1,case when regexp_instr(WOHNORT ,',')=0 then length(WOHNORT ) else regexp_instr(WOHNORT ,',')-1 end)	Nimmt, wenn Wohnort ein Komma enthält, nur den Text vor dem Komma Bsp. „Ingolstadt“, <b>Tabelle 2</b>
Künstlichen amtlichen Gemeindegemeinschaften aus fünfstelligen Gemeindegemeinschaften erstellen für Städte	GEMEINDESCHLUESSEL    '000'	Hängt drei "0"en an den Gemeindegemeinschaften, wenn Gemeindegemeinschaftenlänge = 5 Bsp. „Ingolstadt“, <b>Tabelle 2</b>
Künstlichen Gemeindegemeinschaften aus fünfstelligen Gemeindegemeinschaften erstellen für "Übrige..."	GEMEINDESCHLUESSEL    'NNN' bzw. GEMEINDESCHLUESSEL    'NNNNN' bzw. GEMEINDESCHLUESSEL    'NNNNNN'	Hängt "N" and den Gemeindegemeinschaften, wenn Wohnort wie "Übrige...", <b>Tabelle 2</b>
Kfz-Nationalitätenkennzeichen vereinheitlichen	regexp_replace(('\GEMEINDESCHLUESSEL', '^[:alpha:]]{1,3})(_{0,2}0{1,5}\$)', '\1')	Nimmt nur den ersten, linken Buchstabenteil Bsp. „Indien“, <b>Tabelle 2</b>

\*) Oracle-SQL

Im Anschluss werden alle Pendlerdaten auf kleinster räumlicher Hierarchieebene (mit achtstelligem Gemeindegemeinschaften und ohne Kfz-Nationalitätenkennzeichen (ohne München) bzw. mit Kfz-Nationalitätenkennzeichen und ohne Gemeindegemeinschaften) in dem Materialized View MV\_PENDERL\_FACT zusammengefasst, welches nach jedem Laden in die Faktentabelle PENDLER\_FACT automatisch aktualisiert wird.

**Tests**

*Tests, Tests, Tests*

Neben Tests im Rahmen des Data Profiling sind selbstverständlich auch bei der Fortschreibung der Daten immer Tests zur Erfolgskontrolle und Qualitätssicherung erforderlich (siehe auch <sup>3)</sup>). Hierzu werden zur Validierung in den verschiedenen Arbeitsschritten diverse Kennzahlen erhoben:

1. Anzahl Datenzeilen in den CSV-Quelldatendateien
2. Anzahl mittels dem Programm SQL Loader importierter Datenzeile
3. Anzahl mittels ETL verarbeiteter Datenzeilen
4. Anzahl künstlich erstellter Datenzeilen

5. Anzahl und Ausprägung aller gelieferten Gemeindeschlüssel, welche nicht in die Fakten übernommen worden sind (bedeutet Verarbeitungsfehler und erfordert Korrektur!)
6. Berechnung von Kennzahlen aus der Faktentabelle PENDLER\_FACT und Vergleich mit den Angaben in der XLS-Ursprungsdatei (SQL-Testskript-Ausgabe mit anschließender manueller Sichtkontrolle). Dabei sind mehr oder minder komplexe SQL-Befehle erforderlich, um auf die entsprechenden Untermengen zu filtern, wie in Tabelle 4 dargestellt.
7. Test Männer + Frauen = Insgesamt bzw. Deutsche + Ausländer = Insgesamt

**Tabelle 4: Filter- und Rechenregeln zum Test der Fakten in PENDLER\_FACT**

Kennzahl Pendler	Vorkommen	Filterregel *)	Rechenregel
Übrige (Bundes)Länder	Einpendler	AGS = 'NNNNNNNN'	-
Bundesgebiet	Einpendler / Auspendler	AGS = 'ZZ'	<u>Auspendler</u> : Auspendler + München + Wegen Datenschutz nicht zuzuordnen <u>Einpendler</u> : Einpendler + München – Übrige (Bundes)Länder - Ausland
Insgesamt	Einpendler / Auspendler	AGS = 'ZZ' or AGS is null or AGS in ('NNNNNNNN', 'XXXXXXXX')	Bundesgebiet + Ausland + Keine Zuordnung möglich + Übrige (Bundes)Länder
Auspendler	Auspendler	length (AGS)=8	-
Einpendler	Einpendler	length (AGS)=8 or AGS is null oder (AGS = 'ZZ' or AGS is null or AGS in ('NNNNNNNN', 'XXXXXXXX')) or AGS != '09162000'	(Bundesgebiet + Ausland + Übrige (Bundes)Länder + Keine Zuordnung möglich) oder München
München	Einpendler / Auspendler	AGS = '09162000'	-
Ausland	Einpendler	AGS is null	Insgesamt – Bundesgebiet – Keine Zuordnung möglich – Übrige Bundesländer
Keine Zuordnung möglich	Einpendler	AGS = 'XXXXXXXX'	-
Wegen Datenschutz nicht zuzuordnen	Auspendler	AGS = 'AAAAAAA'	-

\*) Oracle-SQL

Die Summen aus der Drill-fähigen Tabelle MV\_PENDLER\_FACT lassen sich gemäß Listing 1 dann wie in Tabelle 5 dargestellt berechnen. Die Drill-Spalte ist AGS.

**Tabelle 5: Filterregeln für die Kennzahlberechnung aus MV\_PENDLER\_FACT**

Kennzahl Pendler	Vorkommen	Filterregel *)
Übrige (Bundes)Länder	Einpendler	AGS = 'NNNNNNNN' and pendlerschluessel = 'E'
Bundesgebiet	Einpendler / Auspendler	AGS = is not null
Insgesamt	Einpendler / Auspendler	-
Auspendler	Auspendler	pendlerschluessel = 'A'
Einpendler	Einpendler	pendlerschluessel = 'E'
Ausland	Einpendler	zgebi_id is not null and pendlerschluessel = 'E'
Keine Zuordnung möglich	Einpendler	AGS = 'XXXXXXXX' and pendlerschluessel = 'E'
Wegen Datenschutz nicht zuzuordnen	Auspendler	AGS = 'AAAAAAA' and pendlerschluessel = 'A'

\*) Oracle-SQL

**Fazit**

Der Aufbereitung der Pendlerdaten ging einmalig eine aufwendige Datenanalyse voran, die alle wichtigen Verarbeitungs- und Auswertungsregeln zum Ergebnis hatte. Die eigentliche Modellierung der beiden Datenmodelle nahm nur den geringsten Teil des Arbeitsaufwandes in Anspruch. Im Aufwand dazwischen lag die iterative Entwicklung des ETL und der Tests. Bis auf den ersten Arbeitsschritt, der Umwandlung der Rohdaten von Excel in das CSV-Format, können alle nachfolgenden Arbeiten automatisiert und damit zugleich standardisiert werden. Der Skript-gesteuerte Arbeitsablauf ermöglicht zugleich die Dokumentation der Verarbeitungs- und Testergebnisse in Protokolldateien. Die automatische Verarbeitung dauert weniger als zwei Minuten.

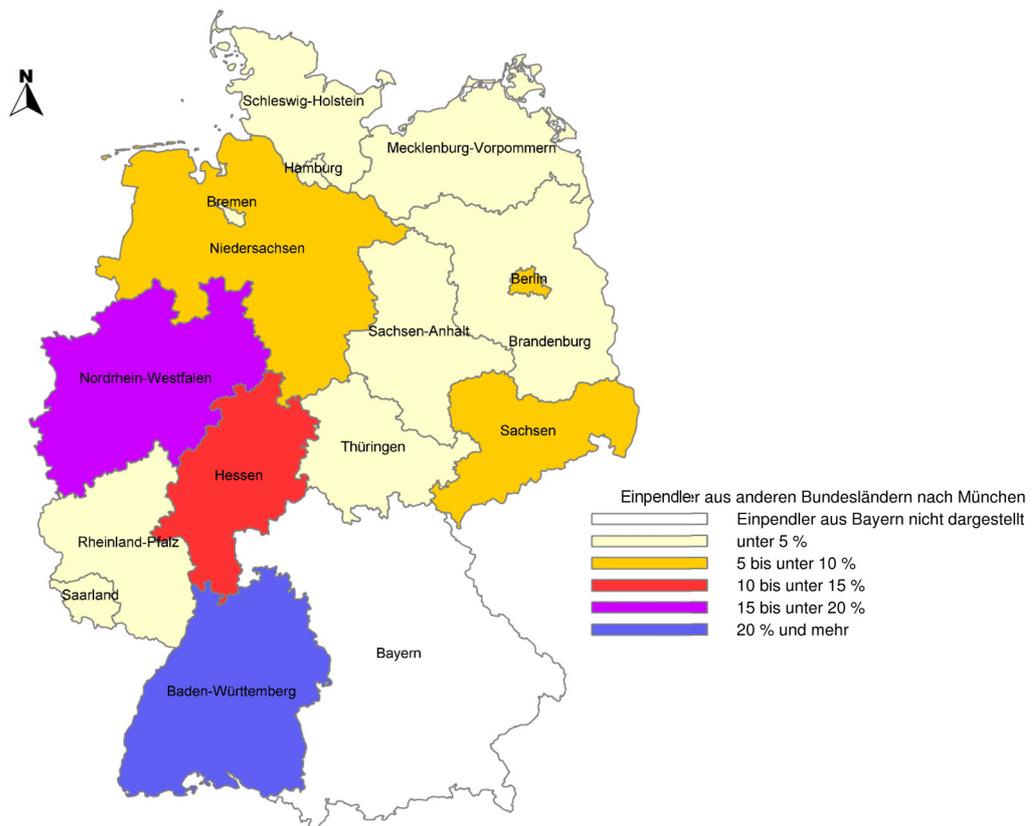
Die Datenlücken konnten geschlossen werden, so dass im Ergebnis eine räumlich hierarchische Aggregation bzw. Auswertung möglich ist. Hierfür bedarf es dann nicht unbedingt eines BI-Werkzeuges, es reicht auch ein Tabellenkalkulationsprogramm <sup>(4)</sup>, vgl. Tabelle 1).

Ferner entstehen hierdurch keine Verfälschungen mehr durch Fehlwerte, z.B. bei einer kartographischen Darstellung (vgl. Abbildung 3). Die leichte Verfälschung der Daten zumeist auf Detailebene durch Ersetzen der wegen Datenschutz als „X“ dargestellten Werte mit „0“ ist vernachlässigbar.

Schließlich können aus den Daten benutzerdefinierten Summen, welche auf die Gemeindeschlüssel-Hierarchie aufbauen, wie z.B. die Planungsregionen in Bayern, errechnet werden.

Neue Auswertungsmöglichkeiten

**Abbildung 3: Einpendler aus anderen Bundesländern nach München in Prozent zum Stand 30.6.2011**



**Quellen**

1) PÜTTMANN, H. (2011): Zusammenführung von Sach- und Geodaten in einer Oracle-Datenbank. In: Münchner Statistik, 4. Quartalshft, Jahrgang 2010, S. 12-22. - 2) <http://www.destatis.de>, Startseite > Zahlen & Fakten > Länder & Regionen > Regionales > Gemeindeverzeichnis - 3) APEL, D., BEHME, W., EBERLEIN, R. Und MERIGHI, C.: Datenqualität erfolgreich Steuern. Hanser 2009. - 4) PÜTTMANN, H. (2008): Einfaches Reporting mit OpenOffice Calc. In: DOAG News, Heft 2, S. 61-63.