DBI Übung DWH Bevölkerungsveränderung OÖ

9. Jänner 2022 | Version 0.4 | 5BHIF Andreas Aigner und Janik Neißl

Inhaltsverzeichnis

1 Einführung	3
1.1 Aufgabenstellung:	3
1.1 Adigabenstellung.	3
2 Datenquellen	3
3 Metadaten der Datensätze	4
3.1 Entwicklung Gesamt 1869 - 2021	4
3.2 Staatsangehörigkeit laut Volkszählung 1971 - 2011	4
3.3 Staatsangehörigkeit 2011 - 2021	5
3.4 Geschlecht und Alter nach Volkszählung 1971 - 2011	5
3.5 Geschlecht und Alter 2011 - 2021	5
4 Analyse der Datenstruktur	6
4.1 Staatsangehörigkeit	6
4.2 Geschlecht und Alter	6
5 ERD-Diagramm der DWH Struktur	7
6 Aufbereiten der Daten zum Laden in die Datenbank	8
7 Laden in die Datenbank	8
7.1 Laden der Daten in Staging Tabellen	8
7.2 Erstellen der DWH Struktur	10
7.3 Laden der Daten	11

1 Einführung

Dieser Bericht befasst sich mit der Erstellung eines Data Warehouse zum Thema "Veränderung der oö. Bevölkerung". Es müssen Datensätze (von <u>data.gv.at</u> oder <u>statistik.at</u>) zum Thema analysiert und ausgewertet werden. Hierfür wird eine DWH-Struktur erstellt und die Daten werden für die Auswertung in die Datenbank geladen.

Ressourcen sowie weitere Anhänge von diesem Bericht finden Sie online unter https://github.com/SchoolRepos/5B_DBI-DWH_Bevoelkerungsveraenderung.

1.1 Aufgabenstellung:

- 1. Finden der Daten (https://www.data.gv.at oder www.statistik.at
- 2. Kurzbericht über die Metadaten (Zeitraum der Daten, Ersteller der Daten, Aktualisierung, welche Informationen gibt es, ...)
- 3. Analyse der Datenstrukturen
- 4. Erstellen der DWH-Struktur
- 5. Aufbereiten der Daten zum Laden in die Datenbank
- 6. Laden in die Datenbank (inkl. eventueller Transformationen)
- 7. Auswerten laut Angaben

2 Datenquellen

Aufgabe 1: Finden der Daten

Datenquelle 1 (Entwicklung Gesamt 1869 - 2021):

https://www.data.gv.at/katalog/dataset/8efff49e-00e8-47c9-85b9-c63e42851494

Datensatz 1 (Entwicklung Gesamt 1869 - 2021):

http://data.ooe.gv.at/files/cms/Mediendateien/OGD/ogd_abtStat/OOE_Bevoelkerung_Zeitreih@e.csv

Datenquelle 2 (Staatsangehörigkeit laut Volkszählung 1971 - 2011):

https://www.data.gv.at/katalog/dataset/aa4e297a-9f9d-40d5-814c-6583c5e71d1f

Datensatz 2 (Staatsangehörigkeit laut Volkszählung 1971 - 2011):

http://data.ooe.gv.at/files/cms/Mediendateien/OGD/ogd_abtStat/OOE_Bev_laut_Volkszaehlung_Staatsangehoerigkeit.csv

Datenquelle 3 (Staatsangehörigkeit 2011 - 2021):

https://www.data.gv.at/katalog/dataset/e8d2e799-e495-4929-bbfa-53672778f025

Datensatz 3 (Staatsangehörigkeit 2011 - 2021):

http://data.ooe.gv.at/files/cms/Mediendateien/OGD/ogd_abtStat/OOE_Bev_Staatsangehoerigkeit.csv

Datenquelle 4 (Geschlecht und Alter nach Volkszählung 1971 - 2011):

https://www.data.gv.at/katalog/dataset/0c4505de-0128-44b4-9d86-8d396c388272

Datensatz 4 (Geschlecht und Alter nach Volkszählung 1971 - 2011):

http://data.ooe.gv.at/files/cms/Mediendateien/OGD/ogd_abtStat/OOE_Bev_laut_Volkszaehlung_Geschl_Alt5J.csv

Datenquelle 5 (Geschlecht und Alter 2011 - 2021):

https://www.data.gv.at/katalog/dataset/e108dcc3-1304-4076-8619-f2185c37ef81

Datensatz 5 (Geschlecht und Alter 2011 - 2021):

http://data.ooe.gv.at/files/cms/Mediendateien/OGD/ogd_abtStat/OOE_Bev_Geschl_Alt5J.csv

3 Metadaten der Datensätze

Aufgabe 2: Kurzbericht über die Metadaten

Sämtliche Datensätze behandeln 438 Gemeinden It. Gebietsstand 1.1.2022.

3.1 Entwicklung Gesamt 1869 - 2021

Zeitraum: 1869 - 2021

Ersteller: Land Oberösterreich, Abteilung Trends und Innovation, Statistik Oberösterreich

Aktualisierung: jährlich, zuletzt aktualisiert am 10.08.2022 23:19:03

Welche Informationen gibt es:

- 1. Bundesland Oberösterreich (AT31 NUTS2)
- 2. Gemeindekennzahl
- 3. Gemeindename
- 4. Referenzjahr
- 5. Bevölkerung insgesamt

Datenherkunft: Volkszählungen bzw. ZMR-Populationsregister; Statistik Austria; Daten bis inkl. 2001 aus STATCUBE (Abfragedatum Juli 2015)

3.2 Staatsangehörigkeit laut Volkszählung 1971 - 2011

Zeitraum: 1971 - 2011

Ersteller: Land Oberösterreich, Abteilung Trends und Innovation, Statistik Oberösterreich

Aktualisierung: kontinuierlich, zuletzt aktualisiert am 15.08.2022 00:59:00

Welche Informationen gibt es:

- 1. Bundesland Oberösterreich (AT31 NUTS2)
- 2. Gemeindekennzahl
- 3. Gemeindename
- 4. Referenzjahr
- 5. Österreichische Staatsangehörigkeit
- 6. Ausländische Staatsangehörigkeit (inkl. staatenlos, ungeklärt, unbekannt)
- 7. Wohnbevölkerung insgesamt

Datenherkunft: Volks- bzw. Registerzählungen (1971-2001, 2011); Statistik Austria; Daten bis inkl. 2001 aus STATCUBE (Abfragedatum Juli 2015)

3.3 Staatsangehörigkeit 2011 - 2021

Zeitraum: 2011 - 2021

Ersteller: Land Oberösterreich, Abteilung Trends und Innovation, Statistik Oberösterreich

Aktualisierung: jährlich, 26.09.2022 06:20:27

Welche Informationen gibt es:

- 1. Bundesland Oberösterreich (AT31 NUTS2)
- 2. Gemeindekennzahl
- 3. Gemeindename
- 4. Referenzjahr
- 5. Österreichische Staatsangehörigkeit
- 6. Staatsangehörigkeit EU-Mitgliedsstaat (laut aktuellem Stand (EU 27), ohne Österreich und Vereinigtes Königreich)
- 7. Staatsangehörigkeit ehemaliges Jugoslawien (ohne Kroatien und Slowenien)
- 8. Staatsangehörigkeit Türkei
- 9. übrige Staatsangehörigkeit (inkl. Vereinigtes Königreich, staatenlos, ungeklärt, unbekannt)
- 10. Wohnbevölkerung insgesamt

Datenherkunft: ZMR-Populationsregister; Statistik Austria

3.4 Geschlecht und Alter nach Volkszählung 1971 - 2011

Zeitraum: 1971 - 2021

Ersteller: Land Oberösterreich, Abteilung Trends und Innovation, Statistik Oberösterreich Aktualisierung: kontinuierlich, letzte Aktualisierung am 15.08.2022 00:58:16 Welche Informationen gibt es:

- 1. Bundesland Oberösterreich (AT31 NUTS2)
- 2. Gemeindekennzahl
- 3. Gemeindename
- 4. Geschlecht (0 = total, 1 = männlich, 2 = weiblich)
- 5. Referenzjahr
- 6. Bevölkerungszahl nach Altersgruppe

Datenherkunft: Volks- bzw. Registerzählungen (1971-2001, 2011); Statistik Austria; Daten bis inkl. 2001 aus STATCUBE (Abfragedatum Juli 2015)

3.5 Geschlecht und Alter 2011 - 2021

Zeitraum: 2011 - 2021

Ersteller: Land Oberösterreich, Abteilung Trends und Innovation, Statistik Oberösterreich Aktualisierung: jährlich, letzte Aktualisierung am 05.09.2022 00:58:09

Welche Informationen gibt es:

- 1. Bundesland Oberösterreich (AT31 NUTS2)
- 2. Gemeindekennzahl
- 3. Gemeindename
- 4. Geschlecht (0 = total, 1 = männlich, 2 = weiblich)
- 5. Referenzjahr
- 6. Bevölkerungszahl nach Altersgruppe

Datenherkunft: ZMR-Populationsregister; Statistik Austria

4 Analyse der Datenstruktur

Aufgabe 3: Analyse der Datenstrukturen

4.1 Staatsangehörigkeit

Die Daten aus 3.2 Staatsangehörigkeit laut Volkszählung 1971 - 2011 unterteilen die Staatsangehörigkeit in 3 Kategorien:

- NATION_AUSTRIA (Österreichische Staatsangehörigkeit)
- NATION_FOREIGN (Ausländische Staatsangehörigkeit inkl. staatenlos, ungeklärt, unbekannt)
- NATION_TOTAL (Wohnbevölkerung insgesamt)

Bei den Daten aus 3.3 Staatsangehörigkeit 2011 - 2021 wird die Staatsangehörigkeit spezifischer unterteilt. Daraus ergeben sich 6 Kategorien:

- NATION AUSTRIA (Österreichische Staatsangehörigkeit)
- NATION_EU (Staatsangehörigkeit EU-Mitgliedsstaat (laut aktuellem Stand (EU 27), ohne Österreich und Vereinigtes Königreich)
- NATION_FORMER_YUGOSLAVIA (Staatsangehörigkeit ehemaliges Jugoslawien ohne Kroatien und Slowenien)
- NATION TURKEY (Staatsangehörigkeit Türkei)
- NATION_OTHERS (übrige Staatsangehörigkeit inkl. Vereinigtes Königreich, staatenlos, ungeklärt, unbekannt)
- NATION_TOTAL (Wohnbevölkerung insgesamt)

Die Daten aus 3.3 könnten wir mit den 3 Kategorien aus 3.2 zusammenführen, indem wir jene Kategorien, die nicht in 3.2 vorkommen, in NATION_FOREIGN kombinieren, ohne diese weiter nach Herkunft zu unterteilen.

4.2 Geschlecht und Alter

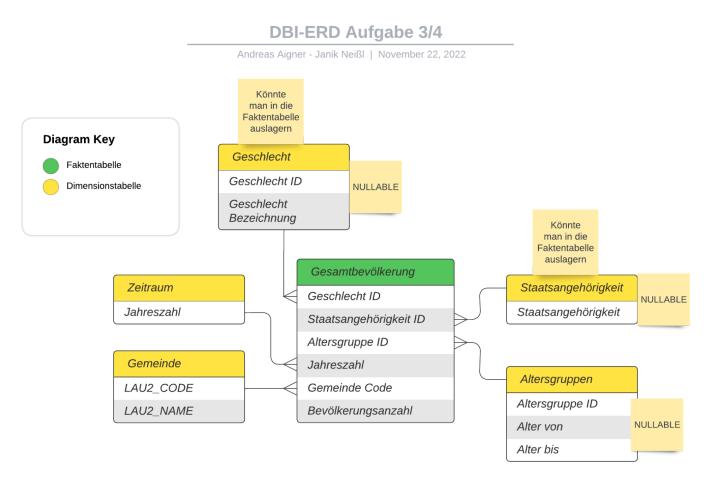
Im Datensatz 3.4 Geschlecht und Alter nach Volkszählung 1971 - 2011 werden alle Über-90-Jährigen in der Kategorie "AGE_90_PLUS" zusammengefasst.

Datensatz 3.5 Geschlecht und Alter 2011 - 2021 beinhaltet zwei zusätzliche Unterteilungen für diese Altersgruppe:

- AGE 90 TO 94 (Personen zwischen 90 und 94 Jahre)
- AGE 95 TO 99 (Personen zwischen 95 und 99 Jahre)
- AGE 100 PLUS (Personen über 100 Jahre)

5 ERD-Diagramm der DWH Struktur

Aufgabe 4: Erstellen der DWH-Struktur



Grafik selbst erstellt mit https://www.lucidchart.com

6 Aufbereiten der Daten zum Laden in die Datenbank

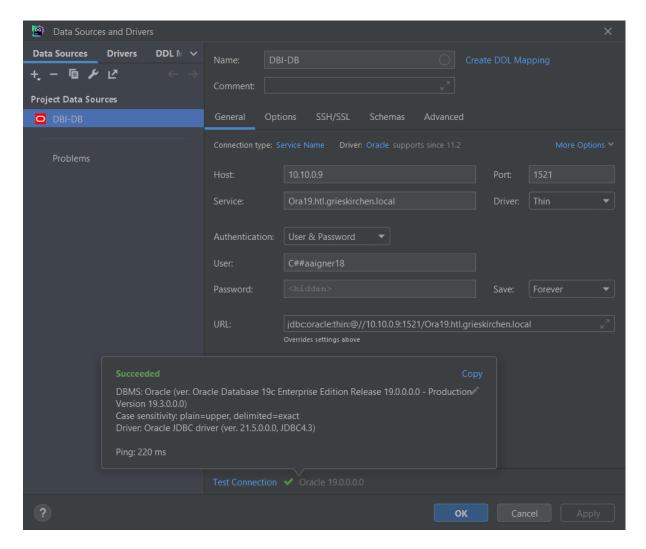
Aufgabe 5: Aufbereiten der Daten zum Laden in die Datenbank Bearbeiten der CSV-Dateien (Löschen von Kommentarzeilen, Beschreibungen, ...)

7 Laden in die Datenbank

Aufgabe 6: Laden in die Datenbank (inkl. eventueller Transformationen) Laden der Daten in Staging Tabellen und erstellen der DWH Struktur.

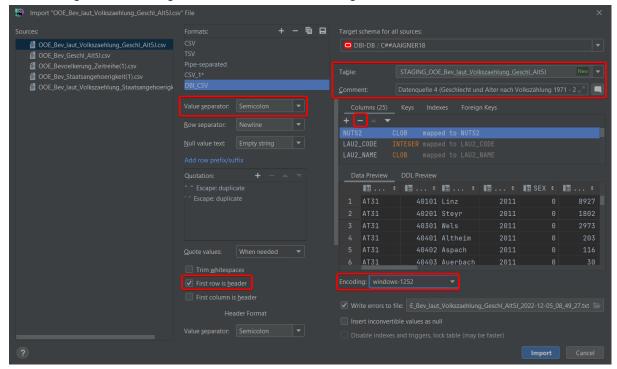
7.1 Laden der Daten in Staging Tabellen

Zum Laden der Daten verwenden wir das Jetbrains Tool Datagrip. Die Verbindung wird wie beim SQL Developer hergestellt mit folgenden Daten als Data Source:



Nun können wir die 5 CSV Datensätze mithilfe des File-Imports in der Datenbank als Staging Tabellen erstellen.

Wir benötigen eine Konfiguration der Tabellen wie folgt:



In einem Kommentar erwähnen wir noch einmal die für die Daten verwendete Quelle. Da die Spalte NAT_2 in unseren Datensätzen immer gleich bleibt können wir diese direkt über das Spaltenmenü rechts entfernen.

Weitere wichtige Einstellungen sind der "Value separator" auf "Semicolon", um die Unterteilung der Spalten im CSV durch Strichpunkte festzulegen und die Checkbox "First row is header", um die Spaltennamen der ersten Zeile im CSV zu übernehmen. Bei unseren Datensätzen müssen wir weiters die Kodierung "windows-1252" um Umlaute der Gemeindenamen richtig darstellen zu können.

Wenn wir das bei allen CSV Dateien konfiguriert haben, können wir auf "Import" drücken. Nun haben wir alle Staging Tabellen die wir benötigen.



7.2 Erstellen der DWH Struktur

Um die Daten leichter aus den CSV Datensätzen zu extrahieren, haben wir die Spalten und Tabellen auf Deutsch benannt.

Dimensionstabellen:

Faktentabelle:

7.3 Laden der Daten

Die Dimensionstabellen befüllen wir mit folgenden Insert Statements:

```
INSERT INTO DIM_GESCHLECHT (geschlecht_id, geschlecht_name)

SELECT DISTINCT SEX,

CASE SEX

WHEN 0 THEN 'total'

WHEN 1 THEN 'männlich'

WHEN 2 THEN 'weiblich'

END

FROM "STAGING_OOE_Bev_Geschl_Alt5J";
```

```
SELECT 'AGE 0 TO 4', 0, 4 FROM DUAL UNION ALL
   SELECT 'AGE 45 TO 49', 45, 49 FROM DUAL UNION ALL
   SELECT 'AGE 55 TO 59', 55, 59 FROM DUAL UNION ALL
  SELECT 'AGE 90 TO 94', 90, 94 FROM DUAL UNION ALL
SELECT DISTINCT YEAR FROM "STAGING OOE Bevoelkerung Zeitreihe";
SELECT DISTINCT LAU2 CODE, TO CHAR(LAU2 NAME)
FROM "STAGING OOE Bevoelkerung Zeitreihe";
INSERT INTO DIM STAATSANGEHOERIGKEIT (STAATSANGEHOERIGKEIT ID,
STAATSANGEHOERIGKEIT NAME)
  SELECT 'FYS', 'NATION FORMER YUGOSLAVIA' FROM DUAL UNION ALL
```

Faktentabelle:

```
INSERT INTO FAKT GESAMTBEVOELKERUNG (geschlecht id,
altersgruppe id, jahr, LAU2_CODE, staatsangehoerigkeit_id,
bevoelkerungsanzahl)
"STAGING OOE Bev laut Volkszaehlung Geschl Alt5J" UNION ALL
SELECT sex, 'AGE 25 TO 29', year, LAU2 CODE, NULL, AGE 25 TO 29
FROM "STAGING OOE Bev laut Volkszaehlung Geschl Alt5J" UNION ALL
FROM "STAGING OOE Bev laut Volkszaehlung Geschl Alt5J" UNION ALL
FROM "STAGING OOE Bev laut Volkszaehlung Geschl Alt5J" UNION ALL
FROM "STAGING_OOE_Bev_laut_Volkszaehlung_Geschl_Alt5J" UNION ALL
SELECT sex, 'AGE 80 TO 84', year, LAU2 CODE, NULL, AGE 80 TO 84
FROM "STAGING OOE Bev laut Volkszaehlung Geschl Alt5J" UNION ALL
"STAGING OOE Bev laut Volkszaehlung Geschl Alt5J" UNION ALL
"STAGING OOE Bev Geschl Alt5J" UNION ALL
```

```
FROM "STAGING OOE Bev Geschl Alt5J" UNION ALL
"STAGING OOE Bev laut Volkszaehlung Staatsangehoerigkeit" UNION
"STAGING OOE Bev laut Volkszaehlung Staatsangehoerigkeit" UNION
"STAGING OOE Bev laut Volkszaehlung Staatsangehoerigkeit" UNION
```

```
SELECT NULL, NULL, year, LAU2_CODE, 'EU', NATION_EU FROM

"STAGING_OOE_Bev_Staatsangehoerigkeit" UNION ALL

SELECT NULL, NULL, year, LAU2_CODE, 'FYS',

NATION_FORMER_YUGOSLAVIA FROM

"STAGING_OOE_Bev_Staatsangehoerigkeit" UNION ALL

SELECT NULL, NULL, year, LAU2_CODE, 'TR', NATION_TURKEY FROM

"STAGING_OOE_Bev_Staatsangehoerigkeit" UNION ALL

SELECT NULL, NULL, year, LAU2_CODE, 'OTH', NATION_OTHERS FROM

"STAGING_OOE_Bev_Staatsangehoerigkeit" UNION ALL

SELECT NULL, NULL, year, LAU2_CODE, 'TOT', NATION_TOTAL FROM

"STAGING_OOE_Bev_Staatsangehoerigkeit";
```