

DBI Übung DWH

Bevölkerungsveränderung OÖ

10. Jänner 2022 | Version 1.0 | 5BHIF

Andreas Aigner und Janik Neißl

Inhaltsverzeichnis

1 Einführung	3
1.1 Aufgabenstellung	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
4.3 Das Jahr 2011 Problem	7
5 ERD-Diagramm der DWH Struktur	7
6 Aufbereiten der Daten zum Laden in die Datenbank	9
7 Laden in die Datenbank	9
7.1 Laden der Daten in Staging Tabellen	9
7.2 Erstellen der DWH Struktur	11
7.3 Laden der Daten	12
8 Auswertungen laut Angaben	16
8.1 Übersicht Bevölkerungswachstum	16
8.2 Altersgruppen und Geschlecht	18
8.3 Geschlechterverteilung über die Jahre in Oberösterreich	19
8.4 Bevölkerung mit Staatsbürgerschaft	19
8.5 Staatsbürgerschaft in den Städten 2011 / 2021	20
8.6 Geschlechterverteilung in den Städten	21
8.7 Altersverteilung Linz	21
8.8 Sonstige Ausreißer und Unregelmäßigkeiten	22
8.9 Pivot Statements Anhang	22
9 Resümee	28
10 Bildverzeichnis	Fehler! Textmarke nicht definiert.

1 Einführung

Dieser Bericht befasst sich mit der Erstellung eines Data Warehouse zum Thema "Veränderung der öö. Bevölkerung". Es müssen Datensätze (von data.gv.at oder statistik.at) 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_Bevoelkerungsveraeenderung.

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/8eff49e-00e8-47c9-85b9-c63e42851494>

Datensatz 1 (Entwicklung Gesamt 1869 - 2021):

http://data.ooe.gv.at/files/cms/Mediendateien/OGD/ogd_abtStat/OOE_Bevoelkerung_Zeitreihe.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 lt. 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

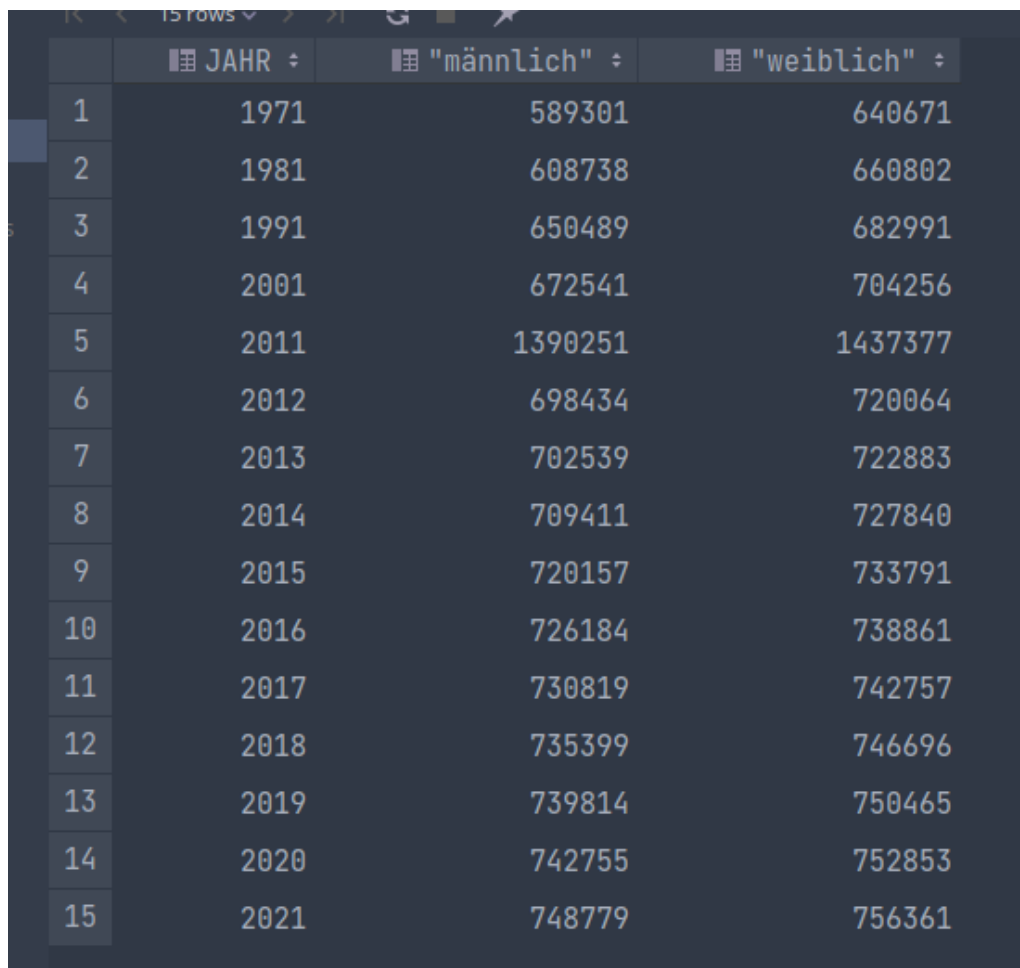
Im Datensatz 3.4 Geschlecht und Alter nach Volkszählung 1971 - 2011 werden alle Über-90-Jährigen in der Kategorie "AGE_90_PLUS" zusammengefasst.

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)

4.3 Das Jahr 2011 Problem

Da Volkszählungen (genauer Überblick über den Bevölkerungsstand) nur alle 10 Jahre stattfinden, haben wir 2 Arten von Datensätzen für Staatsangehörigkeit sowie Geschlecht und Alter. Einmal die Daten der Volkszählungen bis inkl. 2011 (die Volkszählung 2021 ist noch nicht abgeschlossen) und die "neueren", detaillierteren Daten, welche seit inkl. 2011 jährlich dokumentiert wurden. In beiden Datensätzen wurden die Daten für 2011 erfasst, was während der Auswertung in Punkt 8 dazu führte, dass in diesem Jahr die Zahlen stark von den restlichen Werten abweichen. Dies wurde korrigiert, indem alle Werte für das Jahr 2011 aus der Staging-Tabelle des älteren Datensatzes gelöscht wurden (aufgrund deren niedrigeren Detailgrades).



	JAHR	"männlich"	"weiblich"
1	1971	589301	640671
2	1981	608738	660802
3	1991	650489	682991
4	2001	672541	704256
5	2011	1390251	1437377
6	2012	698434	720064
7	2013	702539	722883
8	2014	709411	727840
9	2015	720157	733791
10	2016	726184	738861
11	2017	730819	742757
12	2018	735399	746696
13	2019	739814	750465
14	2020	742755	752853
15	2021	748779	756361

Abb. 1 - Ungewöhnlich starker Ausschlag der Zahlen im Jahr 2011 aufgrund der doppelten Daten

5 ERD-Diagramm der DWH-Struktur

Aufgabe 4: Erstellen der DWH-Struktur

DBI-ERD Aufgabe 3/4

Andreas Aigner - Janik Neißl | November 22, 2022

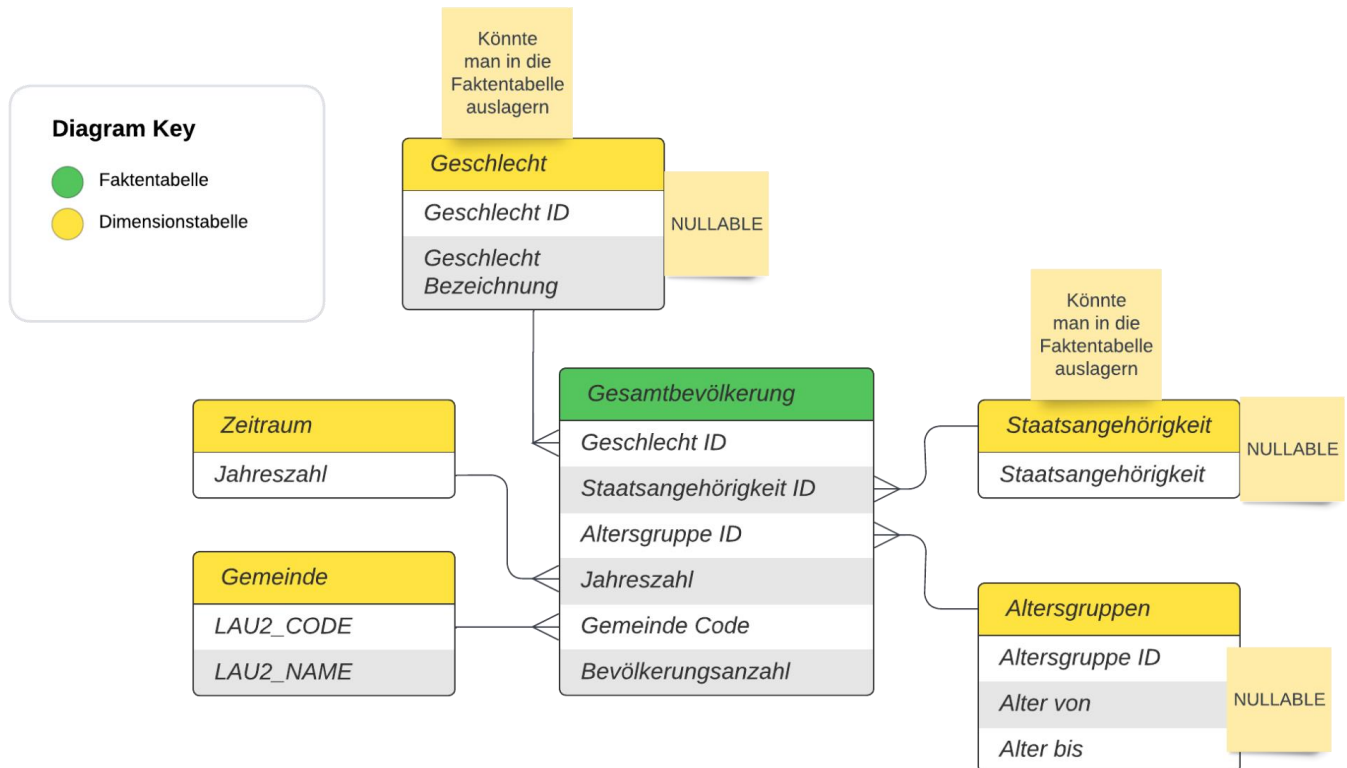


Abb. 2 - ERD-Diagramm der DWH-Struktur

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:

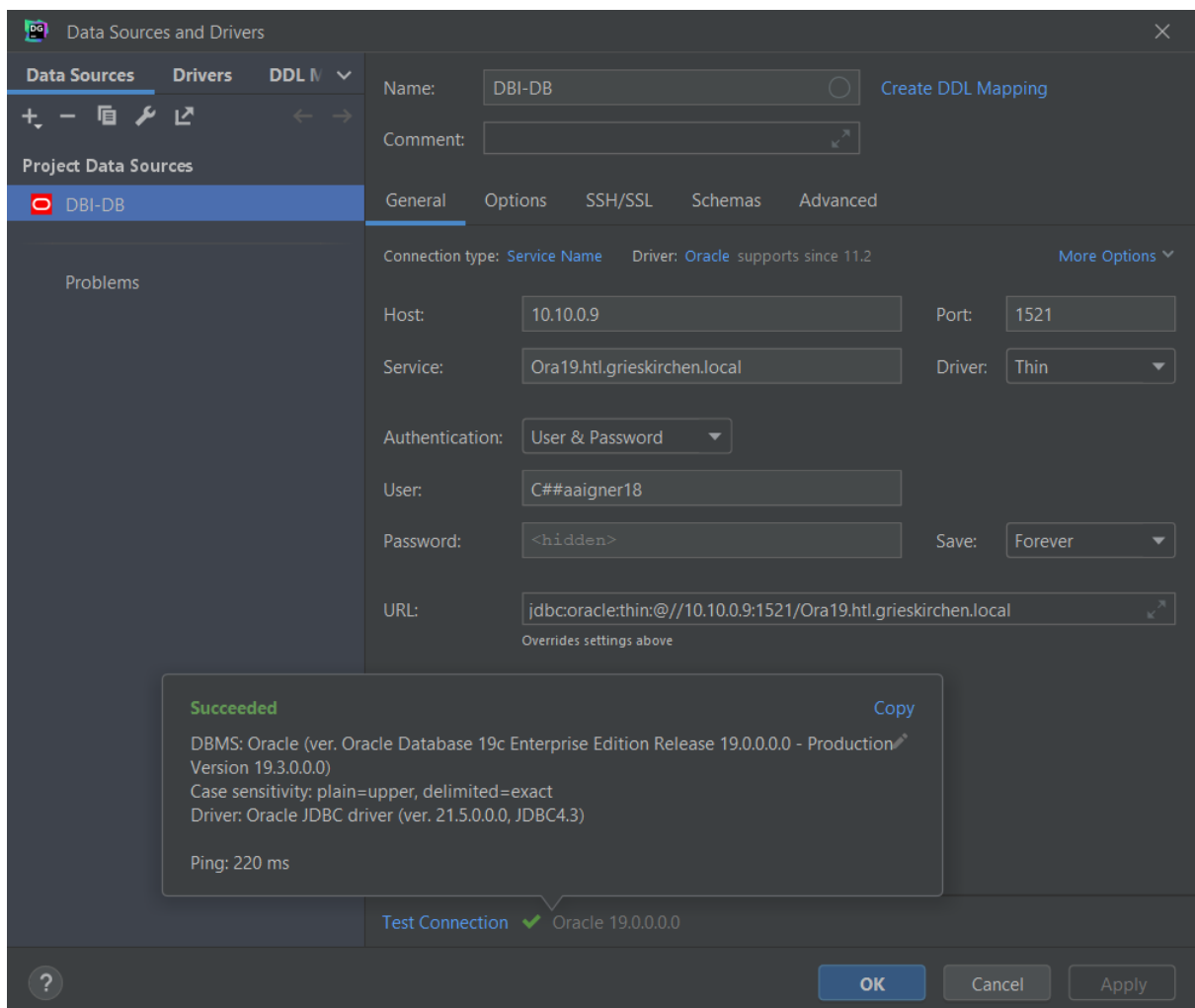


Abb. 3 - Konfigurieren der Zugangsdaten zum Verbindungsaufbau mit dem Datenbank-Server

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:

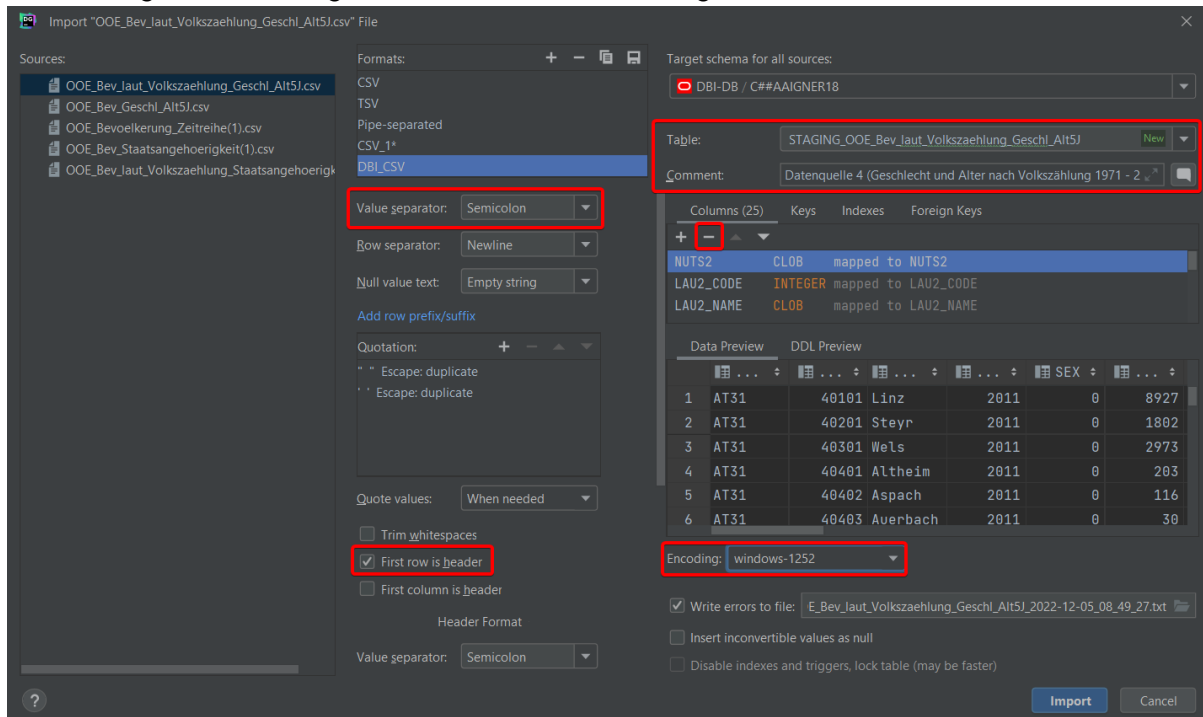


Abb. 4 - Import der CSV-Files mithilfe des Wizards von Datagrip

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 Gemeindennamen 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.

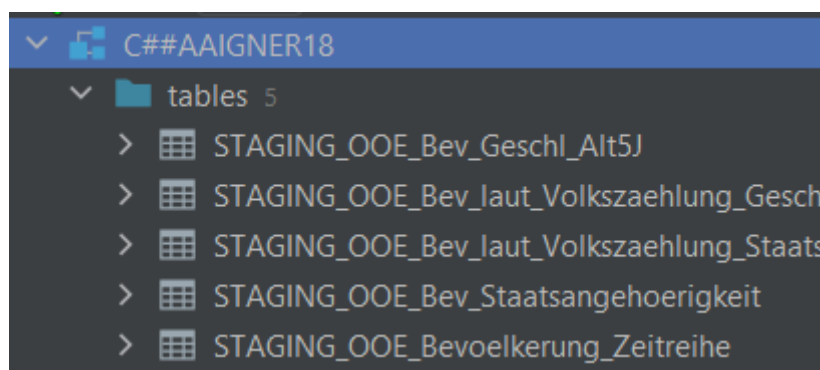


Abb. 5 - Staging-Tabellen mit den importierten CSV-Daten

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:

```
CREATE TABLE DIM_GESCHLECHT
(
    geschlecht_id    CHAR(1)          NOT NULL PRIMARY KEY,
    geschlecht_name  VARCHAR2(50) NOT NULL
);

CREATE TABLE DIM_ALTERSGRUPPE
(
    altersgruppe_id VARCHAR2(15) NOT NULL PRIMARY KEY,
    von              number       NOT NULL,
    bis              number       NOT NULL,
    UNIQUE (von, bis)
);

CREATE TABLE DIM_ZEITRAUM
(
    jahr NUMBER NOT NULL PRIMARY KEY
);

CREATE TABLE DIM_GEMEINDEN
(
    LAU2_CODE CHAR(5)          NOT NULL PRIMARY KEY,
    LAU2_NAME VARCHAR2(150) NOT NULL
);

CREATE TABLE DIM_STAATSANGEHOERIGKEIT
(
    staatsangehoerigkeit_id CHAR(3),
    staatsangehoerigkeit_name VARCHAR(100) NOT NULL
);
```

Faktentabelle:

```
CREATE TABLE FAKT_GESAMTBEVOELKERUNG
(
    geschlecht_id          CHAR(1)          NULL,
    altersgruppe_id        VARCHAR2(15)     NULL,
    jahr                   NUMBER            NOT NULL,
    LAU2_CODE               CHAR(5)          NOT NULL,
    staatsangehoerigkeit_id CHAR(3),

    bevoelkerungsanzahl     NUMBER            NOT NULL,

    FOREIGN KEY (geschlecht_id) REFERENCES DIM_GESCHLECHT
(geschlecht_id),
    FOREIGN KEY (altersgruppe_id) REFERENCES DIM_ALTERSGRUPPE
(altersgruppe_id),
    FOREIGN KEY (jahr) REFERENCES DIM_ZEITRAUM (jahr),
    FOREIGN KEY (LAU2_CODE) REFERENCES DIM_GEMEINDEN (LAU2_CODE),
    FOREIGN KEY (staatsangehoerigkeit_id) REFERENCES
DIM_STAATSANGEHOERIGKEIT (staatsangehoerigkeit_id)
);
```

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";
```

```

INSERT INTO DIM_ALTERSGRUPPE (altersgruppe_id, von, bis)
SELECT 'AGE_0_TO_4', 0, 4 FROM DUAL UNION ALL
SELECT 'AGE_5_TO_9', 5, 9 FROM DUAL UNION ALL
SELECT 'AGE_10_TO_14', 10, 14 FROM DUAL UNION ALL
SELECT 'AGE_15_TO_19', 15, 19 FROM DUAL UNION ALL
SELECT 'AGE_20_TO_24', 20, 24 FROM DUAL UNION ALL
SELECT 'AGE_25_TO_29', 25, 29 FROM DUAL UNION ALL
SELECT 'AGE_30_TO_34', 30, 34 FROM DUAL UNION ALL
SELECT 'AGE_35_TO_39', 35, 39 FROM DUAL UNION ALL
SELECT 'AGE_40_TO_44', 40, 44 FROM DUAL UNION ALL
SELECT 'AGE_45_TO_49', 45, 49 FROM DUAL UNION ALL
SELECT 'AGE_50_TO_54', 50, 54 FROM DUAL UNION ALL
SELECT 'AGE_55_TO_59', 55, 59 FROM DUAL UNION ALL
SELECT 'AGE_60_TO_64', 60, 64 FROM DUAL UNION ALL
SELECT 'AGE_65_TO_69', 65, 69 FROM DUAL UNION ALL
SELECT 'AGE_70_TO_74', 70, 74 FROM DUAL UNION ALL
SELECT 'AGE_75_TO_79', 75, 79 FROM DUAL UNION ALL
SELECT 'AGE_80_TO_84', 80, 84 FROM DUAL UNION ALL
SELECT 'AGE_85_TO_89', 85, 89 FROM DUAL UNION ALL
SELECT 'AGE_90_TO_94', 90, 94 FROM DUAL UNION ALL
SELECT 'AGE_95_TO_99', 95, 99 FROM DUAL UNION ALL
SELECT 'AGE_90_PLUS', 90, 1000 FROM DUAL UNION ALL
SELECT 'AGE_100_PLUS', 100, 1000 FROM DUAL;

INSERT INTO DIM_ZEITRAUM
SELECT DISTINCT YEAR FROM "STAGING_OOE_Bevoelkerung_Zeitreihe";

INSERT INTO DIM_GEMEINDEN (LAU2_CODE, LAU2_NAME)
SELECT DISTINCT LAU2_CODE, TO_CHAR(LAU2_NAME)
FROM "STAGING_OOE_Bevoelkerung_Zeitreihe";

INSERT INTO DIM_STAATSANGEHOERIGKEIT (STAATSANGEHOERIGKEIT_ID,
STAATSANGEHOERIGKEIT_NAME)
SELECT 'AT', 'NATION_AUSTRIA' FROM DUAL UNION ALL
SELECT 'FE', 'NATION_FOREIGN' FROM DUAL UNION ALL
SELECT 'TOT', 'NATION_TOTAL' FROM DUAL UNION ALL
SELECT 'EU', 'NATION_EU' FROM DUAL UNION ALL
SELECT 'FYS', 'NATION_FORMER_YUGOSLAVIA' FROM DUAL UNION ALL
SELECT 'TR', 'NATION_TURKEY' FROM DUAL UNION ALL
SELECT 'OTH', 'NATION_OTHERS' FROM DUAL;

```

Faktentabelle:

```
INSERT INTO FAKT_GESAMTBEVOELKERUNG (geschlecht_id, altersgruppe_id,
jahr, LAU2_CODE, staatsangehoerigkeit_id, bevoelkerungsanzahl)
SELECT sex, 'AGE_0_TO_4', year, LAU2_CODE, NULL, AGE_0_TO_4 FROM
"STAGING_OOE_Bev_laut_Volkszaehlung_Geschl_Alt5J" UNION ALL
SELECT sex, 'AGE_5_TO_9', year, LAU2_CODE, NULL, AGE_5_TO_9 FROM
"STAGING_OOE_Bev_laut_Volkszaehlung_Geschl_Alt5J" UNION ALL
SELECT sex, 'AGE_10_TO_14', year, LAU2_CODE, NULL, AGE_10_TO_14 FROM
"STAGING_OOE_Bev_laut_Volkszaehlung_Geschl_Alt5J" UNION ALL
SELECT sex, 'AGE_15_TO_19', year, LAU2_CODE, NULL, AGE_15_TO_19 FROM
"STAGING_OOE_Bev_laut_Volkszaehlung_Geschl_Alt5J" UNION ALL
SELECT sex, 'AGE_20_TO_24', year, LAU2_CODE, NULL, AGE_20_TO_24 FROM
"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
SELECT sex, 'AGE_30_TO_34', year, LAU2_CODE, NULL, AGE_30_TO_34 FROM
"STAGING_OOE_Bev_laut_Volkszaehlung_Geschl_Alt5J" UNION ALL
SELECT sex, 'AGE_35_TO_39', year, LAU2_CODE, NULL, AGE_35_TO_39 FROM
"STAGING_OOE_Bev_laut_Volkszaehlung_Geschl_Alt5J" UNION ALL
SELECT sex, 'AGE_40_TO_44', year, LAU2_CODE, NULL, AGE_40_TO_44 FROM
"STAGING_OOE_Bev_laut_Volkszaehlung_Geschl_Alt5J" UNION ALL
SELECT sex, 'AGE_45_TO_49', year, LAU2_CODE, NULL, AGE_45_TO_49 FROM
"STAGING_OOE_Bev_laut_Volkszaehlung_Geschl_Alt5J" UNION ALL
SELECT sex, 'AGE_50_TO_54', year, LAU2_CODE, NULL, AGE_50_TO_54 FROM
"STAGING_OOE_Bev_laut_Volkszaehlung_Geschl_Alt5J" UNION ALL
SELECT sex, 'AGE_55_TO_59', year, LAU2_CODE, NULL, AGE_55_TO_59 FROM
"STAGING_OOE_Bev_laut_Volkszaehlung_Geschl_Alt5J" UNION ALL
SELECT sex, 'AGE_60_TO_64', year, LAU2_CODE, NULL, AGE_60_TO_64 FROM
"STAGING_OOE_Bev_laut_Volkszaehlung_Geschl_Alt5J" UNION ALL
SELECT sex, 'AGE_65_TO_69', year, LAU2_CODE, NULL, AGE_65_TO_69 FROM
"STAGING_OOE_Bev_laut_Volkszaehlung_Geschl_Alt5J" UNION ALL
SELECT sex, 'AGE_70_TO_74', year, LAU2_CODE, NULL, AGE_70_TO_74 FROM
"STAGING_OOE_Bev_laut_Volkszaehlung_Geschl_Alt5J" UNION ALL
SELECT sex, 'AGE_75_TO_79', year, LAU2_CODE, NULL, AGE_75_TO_79 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
SELECT sex, 'AGE_85_TO_89', year, LAU2_CODE, NULL, AGE_85_TO_89 FROM
"STAGING_OOE_Bev_laut_Volkszaehlung_Geschl_Alt5J" UNION ALL
SELECT sex, 'AGE_90_PLUS', year, LAU2_CODE, NULL, AGE_90_PLUS FROM
"STAGING_OOE_Bev_laut_Volkszaehlung_Geschl_Alt5J" UNION ALL

SELECT sex, 'AGE_0_TO_4', year, LAU2_CODE, NULL, AGE_0_TO_4 FROM
"STAGING_OOE_Bev_Geschl_Alt5J" UNION ALL
SELECT sex, 'AGE_5_TO_9', year, LAU2_CODE, NULL, AGE_5_TO_9 FROM
"STAGING_OOE_Bev_Geschl_Alt5J" UNION ALL
SELECT sex, 'AGE_10_TO_14', year, LAU2_CODE, NULL, AGE_10_TO_14 FROM
"STAGING_OOE_Bev_Geschl_Alt5J" UNION ALL
```

```

SELECT sex, 'AGE_15_TO_19', year, LAU2_CODE, NULL, AGE_15_TO_19 FROM
"STAGING_OOE_Bev_Geschl_Alt5J" UNION ALL
SELECT sex, 'AGE_20_TO_24', year, LAU2_CODE, NULL, AGE_20_TO_24 FROM
"STAGING_OOE_Bev_Geschl_Alt5J" UNION ALL
SELECT sex, 'AGE_25_TO_29', year, LAU2_CODE, NULL, AGE_25_TO_29 FROM
"STAGING_OOE_Bev_Geschl_Alt5J" UNION ALL
SELECT sex, 'AGE_30_TO_34', year, LAU2_CODE, NULL, AGE_30_TO_34 FROM
"STAGING_OOE_Bev_Geschl_Alt5J" UNION ALL
SELECT sex, 'AGE_35_TO_39', year, LAU2_CODE, NULL, AGE_35_TO_39 FROM
"STAGING_OOE_Bev_Geschl_Alt5J" UNION ALL
SELECT sex, 'AGE_40_TO_44', year, LAU2_CODE, NULL, AGE_40_TO_44 FROM
"STAGING_OOE_Bev_Geschl_Alt5J" UNION ALL
SELECT sex, 'AGE_45_TO_49', year, LAU2_CODE, NULL, AGE_45_TO_49 FROM
"STAGING_OOE_Bev_Geschl_Alt5J" UNION ALL
SELECT sex, 'AGE_50_TO_54', year, LAU2_CODE, NULL, AGE_50_TO_54 FROM
"STAGING_OOE_Bev_Geschl_Alt5J" UNION ALL
SELECT sex, 'AGE_55_TO_59', year, LAU2_CODE, NULL, AGE_55_TO_59 FROM
"STAGING_OOE_Bev_Geschl_Alt5J" UNION ALL
SELECT sex, 'AGE_60_TO_64', year, LAU2_CODE, NULL, AGE_60_TO_64 FROM
"STAGING_OOE_Bev_Geschl_Alt5J" UNION ALL
SELECT sex, 'AGE_65_TO_69', year, LAU2_CODE, NULL, AGE_65_TO_69 FROM
"STAGING_OOE_Bev_Geschl_Alt5J" UNION ALL
SELECT sex, 'AGE_70_TO_74', year, LAU2_CODE, NULL, AGE_70_TO_74 FROM
"STAGING_OOE_Bev_Geschl_Alt5J" UNION ALL
SELECT sex, 'AGE_75_TO_79', year, LAU2_CODE, NULL, AGE_75_TO_79 FROM
"STAGING_OOE_Bev_Geschl_Alt5J" UNION ALL
SELECT sex, 'AGE_80_TO_84', year, LAU2_CODE, NULL, AGE_80_TO_84 FROM
"STAGING_OOE_Bev_Geschl_Alt5J" UNION ALL
SELECT sex, 'AGE_85_TO_89', year, LAU2_CODE, NULL, AGE_85_TO_89 FROM
"STAGING_OOE_Bev_Geschl_Alt5J" UNION ALL
SELECT sex, 'AGE_90_TO_94', year, LAU2_CODE, NULL, AGE_90_TO_94 FROM
"STAGING_OOE_Bev_Geschl_Alt5J" UNION ALL
SELECT sex, 'AGE_95_TO_99', year, LAU2_CODE, NULL, AGE_95_TO_99 FROM
"STAGING_OOE_Bev_Geschl_Alt5J" UNION ALL
SELECT sex, 'AGE_100_PLUS', year, LAU2_CODE, NULL, AGE_100_PLUS FROM
"STAGING_OOE_Bev_Geschl_Alt5J" UNION ALL

SELECT NULL, NULL, year, LAU2_CODE, 'AT', NATION_AUSTRIA FROM
"STAGING_OOE_Bev_laut_Volkszaehlung_Staatsangehoerigkeit" UNION ALL
SELECT NULL, NULL, year, LAU2_CODE, 'FE', NATION_FOREIGN FROM
"STAGING_OOE_Bev_laut_Volkszaehlung_Staatsangehoerigkeit" UNION ALL
SELECT NULL, NULL, year, LAU2_CODE, 'TOT', NATION_TOTAL FROM
"STAGING_OOE_Bev_laut_Volkszaehlung_Staatsangehoerigkeit" UNION ALL
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";
```

8 Auswertungen laut Angaben

Aufgabe 7:

Die folgenden Auswertungen wurden mithilfe der obigen DWH Struktur und einem PIVOT Statement erstellt. Die grafische Darstellung wurde mit Excel realisiert. Zur Auswertung herangezogen wurde die Verbindung aller 2er Kombinationen der 5 Dimensionen des DWH.

Die Verbindungen zwischen Staatsangehörigkeit, Geschlecht und Altersgruppe sind nicht möglich, da sich diese Datensätze nicht überschneiden, wodurch sie nicht angeführt werden können.

Auffälligkeiten und Unregelmäßigkeiten werden bei den einzelnen Grafiken beschrieben.

Etwasige Pivot Statements befinden sich für die einzelnen Unterpunkte im Anhang.

8.1 Übersicht Bevölkerungswachstum

Es zeigt sich ein stetiges Bevölkerungswachstum seit Beginn der Aufzeichnungen. Ein großer Anstieg der Geburten in den 1950er - 1960er Jahren ist hervorzuheben (Baby-Boom - <https://de.wikipedia.org/wiki/Baby-Boomer>).

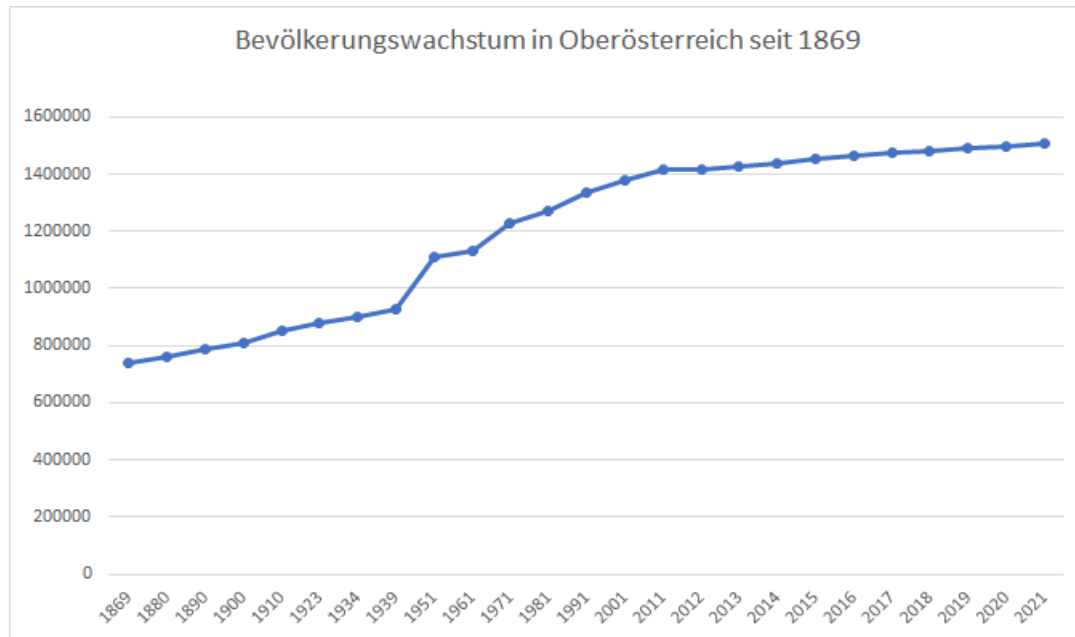


Abb. 6 - Bevölkerungswachstum in Oberösterreich seit 1869

In den Städten Linz, Steyr, Wels und Grieskirchen gab es nur sehr kleine Bevölkerungssprünge über die Jahre hinweg.

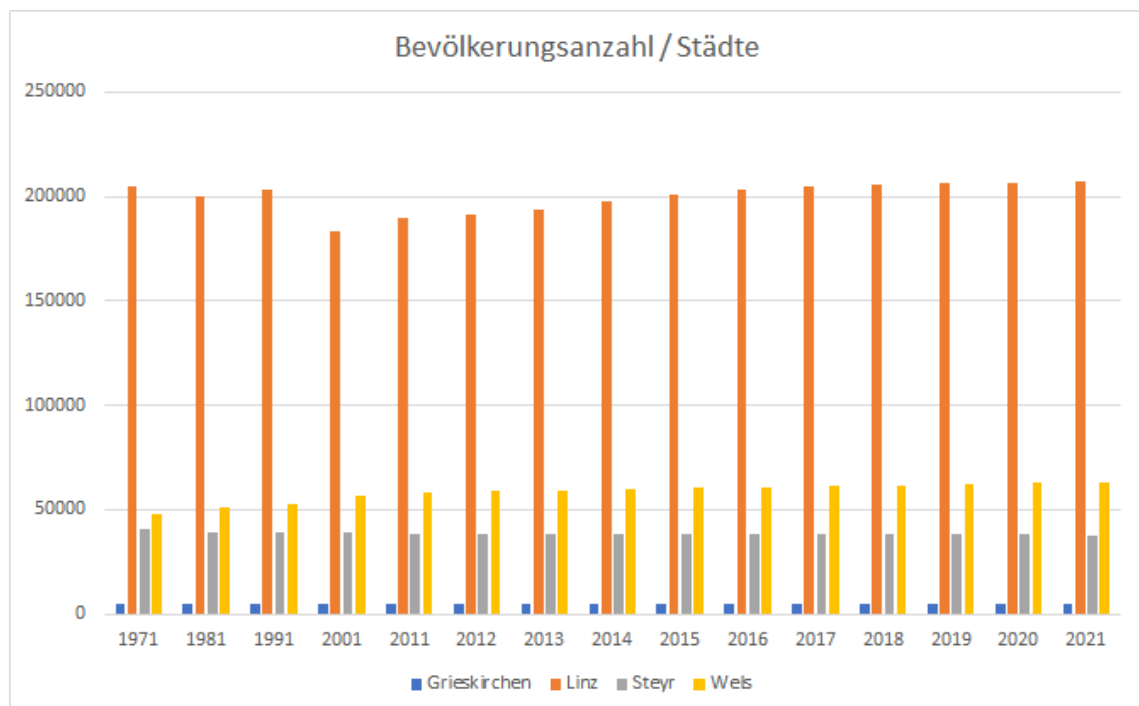


Abb. 7 - Bevölkerungsanzahl in den Städten

8.2 Altersgruppen und Geschlecht

Die folgenden 2 Bevölkerungspyramiden zeigen die Altersverteilung aus den Jahren 2001 und 2021. Man kann erkennen, dass das Geschlechterverhältnis bis auf einige Ausnahmen annähernd ausgeglichen ist. Die Altersstrukturen haben die Urnenform. Man erkennt in beiden Grafiken die Generation des Baby-Booms anhand der Ausschreitungen in den mittleren Bereichen.

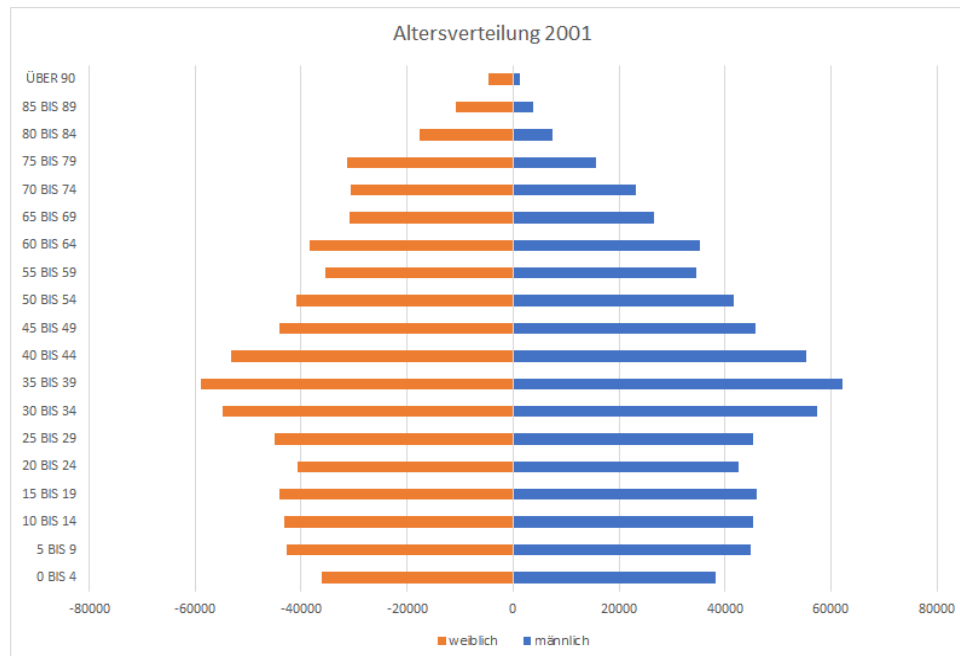


Abb. 8 - Altersverteilung OÖ 2001

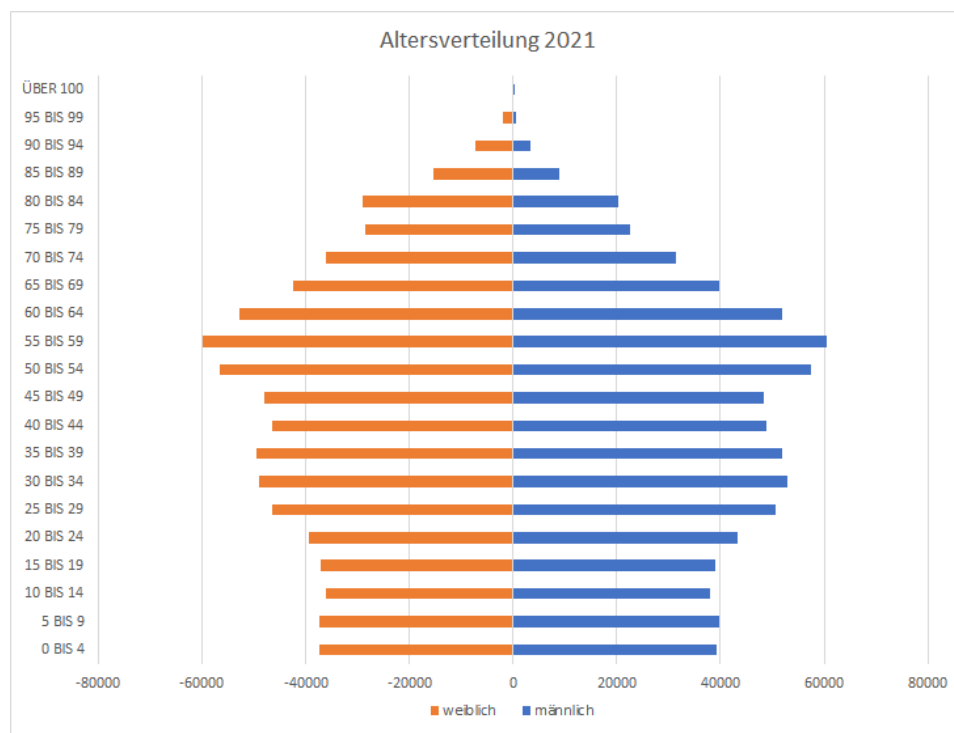


Abb. 9 - Altersverteilung OÖ 2021

8.3 Geschlechterverteilung über die Jahre in Oberösterreich

Die Geschlechterverteilung in Oberösterreich ist sehr ausgeglichen (annähernd 50 / 50 verteilt).

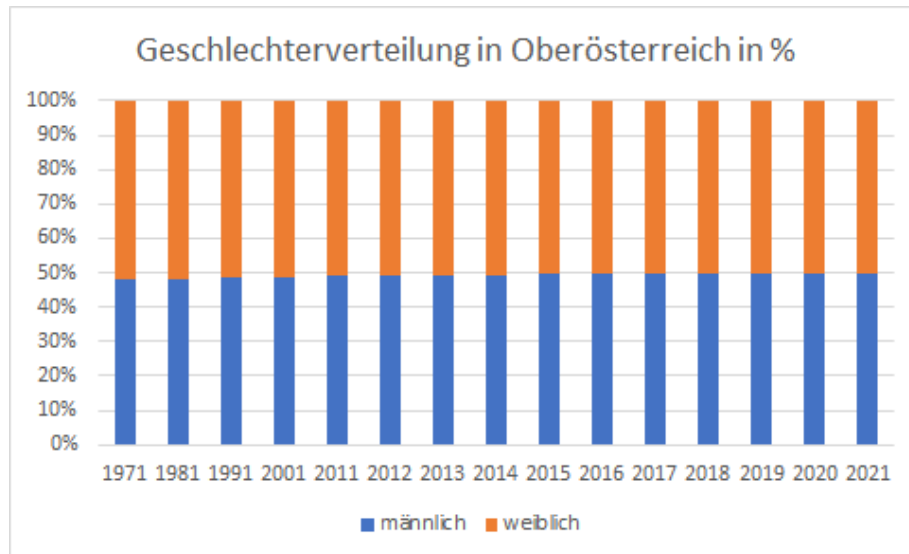


Abb. 10 - Geschlechterverteilung in OÖ in %

8.4 Bevölkerung mit Staatsbürgerschaft

(Datensätze vor 2001 haben keine weitere Untergliederung in Staaten, daher wurden sie auf "Sonstige" zusammengefasst)

Auch ein stetiger Zuwachs an Menschen ohne österreichische Staatsbürgerschaft ist ersichtlich. Genaue Ereignisse wie Flüchtlingswellen konnten in unseren Daten nicht hervorgehoben werden. Es handelt sich eher um einen stetigen Anstieg als um abrupte Ereignisse.

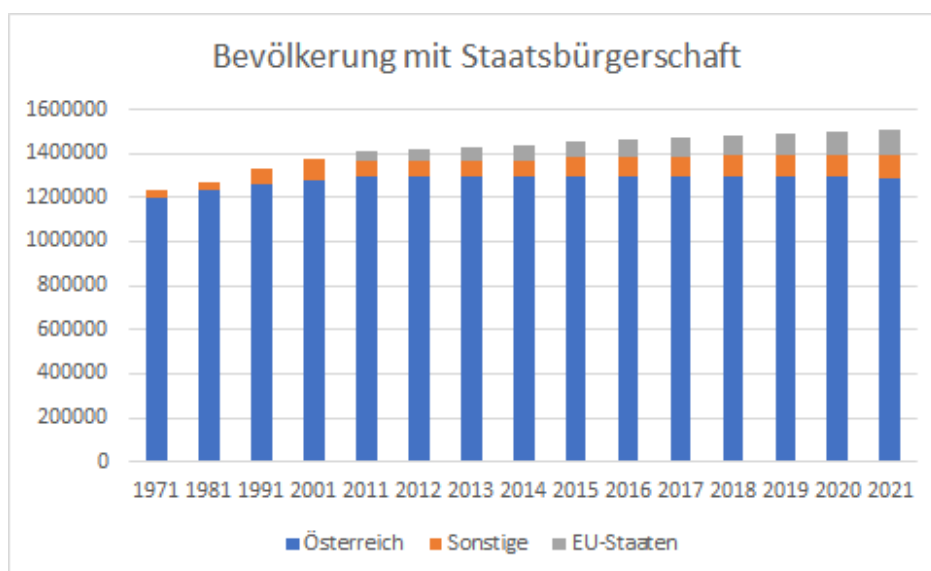


Abb. 11 - Bevölkerung mit Staatsbürgerschaft

8.5 Staatsbürgerschaft in den Städten 2011 / 2021

Die meisten Einwanderer stammen aus EU-Staaten, was keine Überraschung darstellt, wenn man berücksichtigt, dass diese Kategorie die meisten Staaten umfasst. Die Reihung der Städte ist stets dieselbe: Linz ist mit Abstand am attraktivsten, Wels belegt den zweiten Platz und Steyr verzeichnet am wenigsten Zuwachs. Die Zahl der österreichischen Einwohner nahm zwischen 2011 und 2021 in allen drei Städten ab, allein in Linz um etwa 10000.

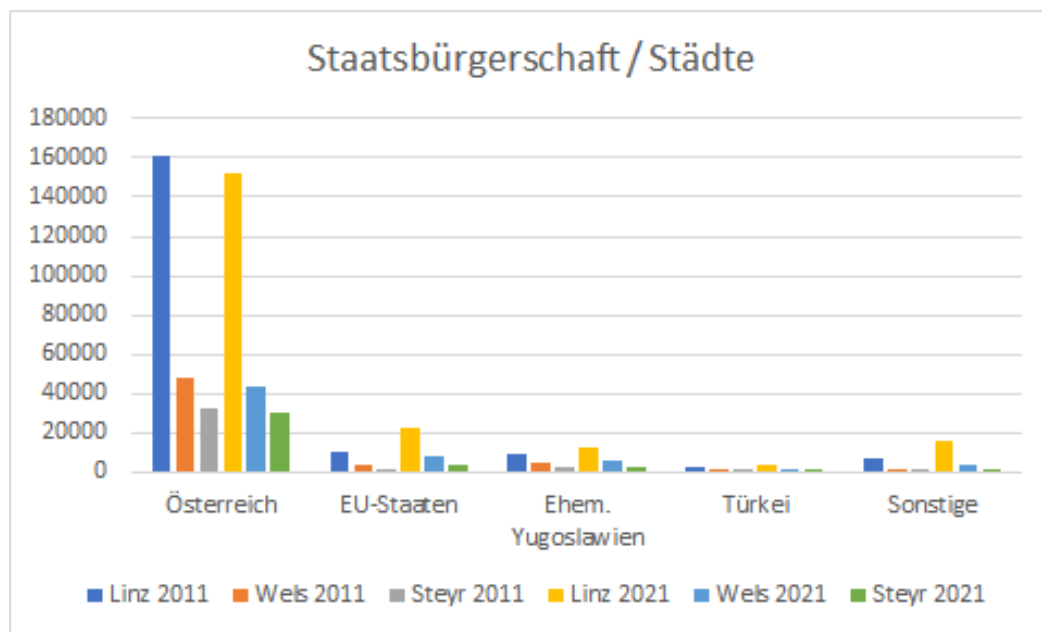


Abb. 12 - Staatsbürgerschaft Verteilung in den Städten 2011 / 2021

8.6 Geschlechterverteilung in den Städten

Diese Werte bestätigen die Ergebnisse aus 8.3, auch in dieser Darstellung ist das Verhältnis zwischen Männern und Frauen etwa 50/50.

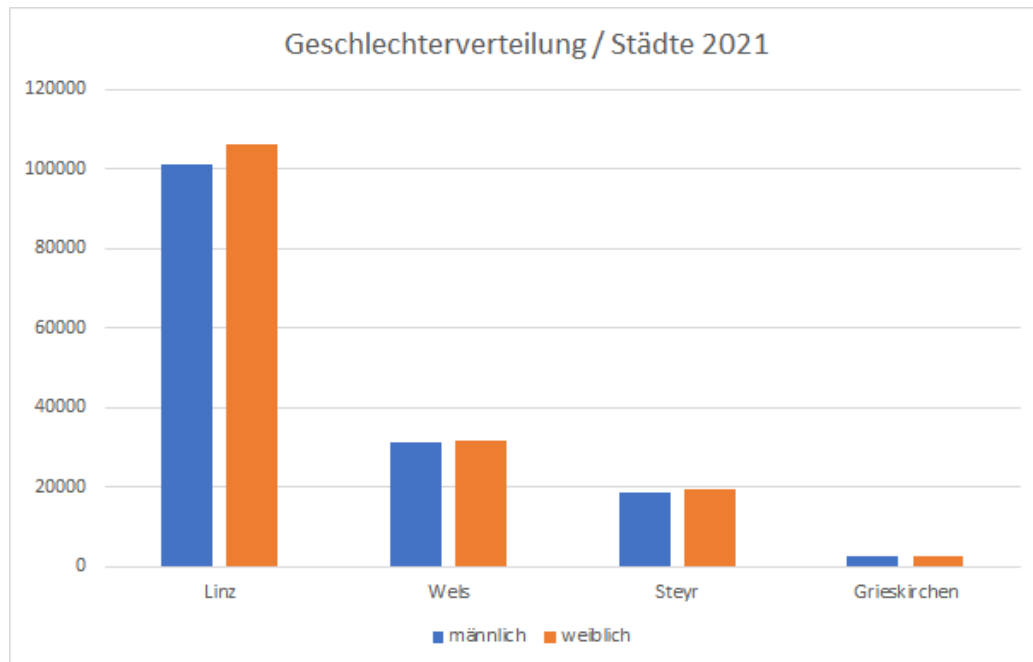


Abb. 13 - Geschlechterverteilung Städte 2021

Die "Gender Gap" in Linz schließt sich langsam, der Anteil der männlichen Einwohner wuchs in den letzten 20 Jahren etwas stärker als der der Weiblichen.

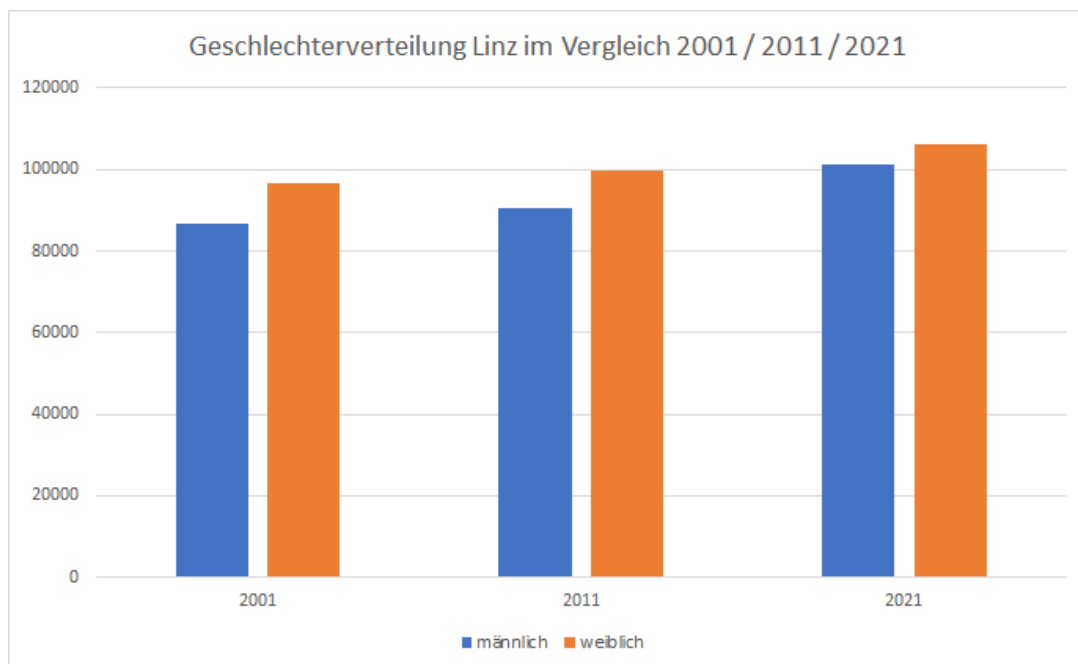


Abb. 14 - Geschlechterverteilung Linz 2001 / 2011 / 2021

8.7 Altersverteilung Linz

Im Bereich über 65 Jahren ist die Anzahl der Personen über die Jahrzehnte hinweg ziemlich ähnlich. Im Gegensatz dazu steht die Zahl der 25- bis 34-jährigen, dieser Altersgruppe gehörten bei der letzten Erhebung erheblich mehr Menschen an als bei den vorherigen Zählungen.

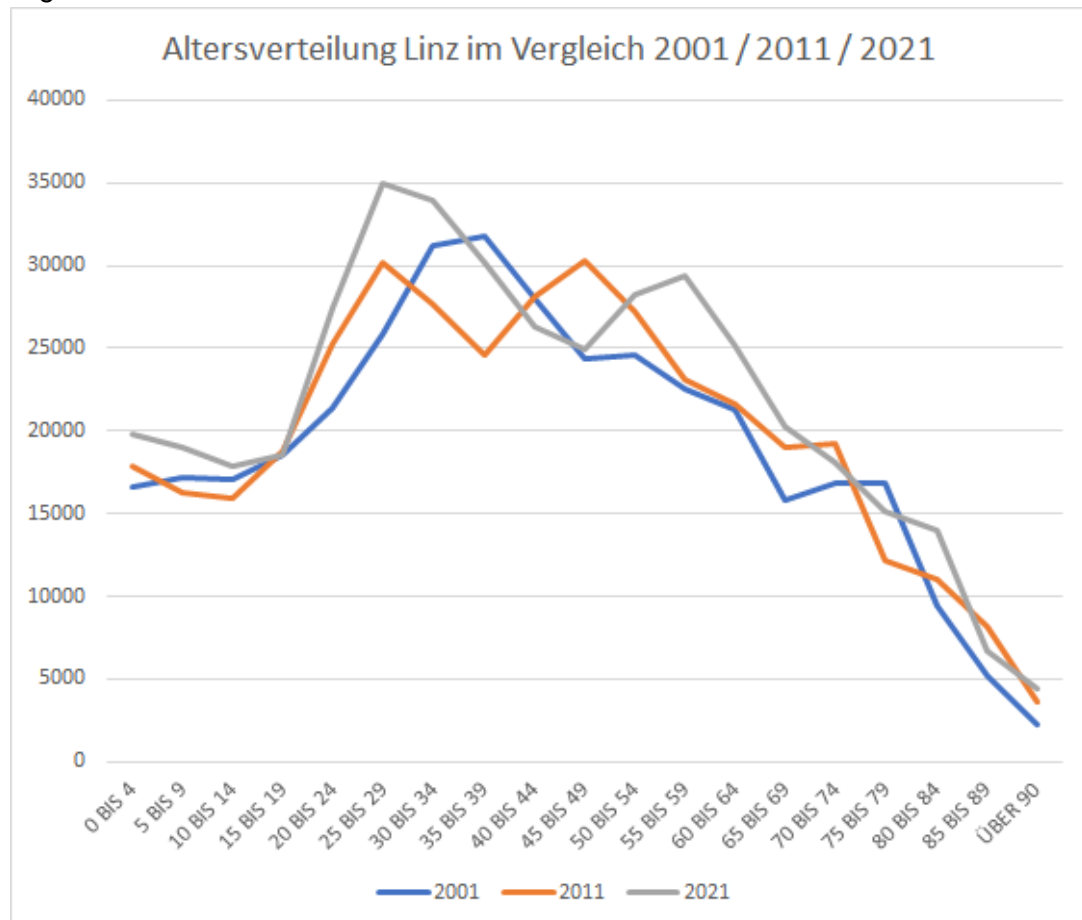


Abb. 15 - Altersverteilung in Linz 2001 / 2011 / 2021

8.8 Sonstige Ausreißer und Unregelmäßigkeiten

Siehe Punkt 4.3 über Unregelmäßigkeiten bei den Daten im Jahr 2011.

8.9 Pivot Statements Anhang

In manchen Statements haben wir die Jahreszahlen ergänzt, um mehrere Datensätze von verschiedenen Jahren zu erhalten.

Übersicht Bevölkerungswachstum

```
SELECT YEAR AS "Jahr", SUM(POP_TOTAL) AS "Gesamtbevölkerung OÖ"  
FROM "STAGING_OOE_Bevoelkerung_Zeitreihe"  
GROUP BY YEAR  
ORDER BY YEAR;
```

Altersgruppen und Geschlecht

```
SELECT *
FROM (
    SELECT GESCHLECHT_NAME, ALTERSGRUPPE_ID,
    SUM(BEVOELKERUNGSANZAHL) AS anzahl
    FROM FAKT_GESAMTBEVOELKERUNG f
    LEFT JOIN DIM_GESCHLECHT g
    ON f.GESCHLECHT_ID = g.GESCHLECHT_ID
    WHERE JAHR IN (2021)
    GROUP BY GESCHLECHT_NAME, ALTERSGRUPPE_ID
)
PIVOT (
    SUM(anzahl)
    FOR ALTERSGRUPPE_ID
    IN (
        'AGE_0_TO_4' AS "0 BIS 4",
        'AGE_5_TO_9' AS "5 BIS 9",
        'AGE_10_TO_14' AS "10 BIS 14",
        'AGE_15_TO_19' AS "15 BIS 19",
        'AGE_20_TO_24' AS "20 BIS 24",
        'AGE_25_TO_29' AS "25 BIS 29",
        'AGE_30_TO_34' AS "30 BIS 34",
        'AGE_35_TO_39' AS "35 BIS 39",
        'AGE_40_TO_44' AS "40 BIS 44",
        'AGE_45_TO_49' AS "45 BIS 49",
        'AGE_50_TO_54' AS "50 BIS 54",
        'AGE_55_TO_59' AS "55 BIS 59",
        'AGE_60_TO_64' AS "60 BIS 64",
        'AGE_65_TO_69' AS "65 BIS 69",
        'AGE_70_TO_74' AS "70 BIS 74",
        'AGE_75_TO_79' AS "75 BIS 79",
        'AGE_80_TO_84' AS "80 BIS 84",
        'AGE_85_TO_89' AS "85 BIS 89",
        'AGE_90_TO_94' AS "90 BIS 94",
        'AGE_90_PLUS' AS "[Legacy] ÜBER 90",
        'AGE_95_TO_99' AS "95 BIS 99",
        'AGE_100_PLUS' AS "ÜBER 100"
    )
)
ORDER BY GESCHLECHT_NAME;
```

Jahre und Gemeinden

```
SELECT *
FROM (
    SELECT JAHR, LAU2_CODE, SUM(BEVOELKERUNGSANZAHL) AS anzahl
    FROM FAKT_GESAMTBEVOELKERUNG
```

```

WHERE STAATSANGEHOERIGKEIT_ID IS NULL AND GESCHLECHT_ID = 0
GROUP BY JAHR, LAU2_CODE
)
PIVOT (
    SUM(anzahl)
    FOR LAU2_CODE
    IN (
        40808 AS "Grieskirchen",
        40101 AS "Linz",
        40201 AS "Steyr",
        40301 AS "Wels"
    )
)
ORDER BY jahr;

```

Jahre und Staatsangehörigkeiten

```

SELECT *
FROM (
    SELECT JAHR, STAATSANGEHOERIGKEIT_ID, SUM(BEVOELKERUNGSANZAHL)
    AS anzahl
    FROM FAKT_GESAMTBEVOELKERUNG
    GROUP BY JAHR, STAATSANGEHOERIGKEIT_ID
)
PIVOT (
    SUM(anzahl)
    FOR STAATSANGEHOERIGKEIT_ID
    IN (
        'AT' AS "Österreich",
        'FE' AS "[Legacy] Ausland",
        'TOT' AS "Gesamt",
        'EU' AS "EU-Staaten",
        'FYS' AS "Ehem. Jugoslawien",
        'TR' AS "Türkei",
        'OTH' AS "Sonstige"
    )
)
ORDER BY jahr;

```

Gemeinden und Staatsangehörigkeiten (Jahr anpassen)

```

SELECT *
FROM (
    SELECT LAU2_NAME, STAATSANGEHOERIGKEIT_ID,
    SUM(BEVOELKERUNGSANZAHL) AS anzahl
    FROM FAKT_GESAMTBEVOELKERUNG f

```



```

        LEFT JOIN DIM_GEMEINDEN g
        ON f.LAU2_CODE = g.LAU2_CODE
    WHERE JAHR IN (2021)
    GROUP BY LAU2_NAME, STAATSANGEHOERIGKEIT_ID
)
PIVOT (
    SUM(anzahl)
    FOR STAATSANGEHOERIGKEIT_ID
    IN (
        'AT' AS "Österreich",
        'FE' AS "[Legacy] Ausland",
        'TOT' AS "Gesamt",
        'EU' AS "EU-Staaten",
        'FYS' AS "Ehem. Jugoslawien",
        'TR' AS "Türkei",
        'OTH' AS "Sonstige"
    )
)
ORDER BY LAU2_NAME;

```

Gemeinden und Altersgruppen

```

SELECT *
FROM (
    SELECT LAU2_NAME, ALTERSGRUPPE_ID, SUM(BEVOELKERUNGSANZAHL) AS
anzahl
    FROM FAKT_GESAMTBEOELKERUNG f
    LEFT JOIN DIM_GEMEINDEN g
    ON f.LAU2_CODE = g.LAU2_CODE
    WHERE JAHR IN (2021)
    GROUP BY LAU2_NAME, ALTERSGRUPPE_ID
)
PIVOT (
    SUM(anzahl)
    FOR ALTERSGRUPPE_ID
    IN (
        'AGE_0_TO_4' AS "0 BIS 4",
        'AGE_5_TO_9' AS "5 BIS 9",
        'AGE_10_TO_14' AS "10 BIS 14",
        'AGE_15_TO_19' AS "15 BIS 19",
        'AGE_20_TO_24' AS "20 BIS 24",
        'AGE_25_TO_29' AS "25 BIS 29",
        'AGE_30_TO_34' AS "30 BIS 34",
        'AGE_35_TO_39' AS "35 BIS 39",
        'AGE_40_TO_44' AS "40 BIS 44",
        'AGE_45_TO_49' AS "45 BIS 49",

```

```

        'AGE_50_TO_54' AS "50 BIS 54",
        'AGE_55_TO_59' AS "55 BIS 59",
        'AGE_60_TO_64' AS "60 BIS 64",
        'AGE_65_TO_69' AS "65 BIS 69",
        'AGE_70_TO_74' AS "70 BIS 74",
        'AGE_75_TO_79' AS "75 BIS 79",
        'AGE_80_TO_84' AS "80 BIS 84",
        'AGE_85_TO_89' AS "85 BIS 89",
        'AGE_90_TO_94' AS "90 BIS 94",
        'AGE_90_PLUS' AS "[Legacy] ÜBER 90",
        'AGE_95_TO_99' AS "95 BIS 99",
        'AGE_100_PLUS' AS "ÜBER 100"
    )
)
ORDER BY LAU2_NAME;

```

Gemeinden und Geschlechter

```

SELECT *
FROM (
    SELECT LAU2_NAME, GESCHLECHT_ID, SUM(BEVOELKERUNGSANZAHL) AS
anzahl
    FROM FAKT_GESAMTBEOELKERUNG f
    LEFT JOIN DIM_GEMEINDEN g
    ON f.LAU2_CODE = g.LAU2_CODE
    WHERE JAHR IN (2001, 2011, 2021)
    GROUP BY LAU2_NAME, GESCHLECHT_ID
)
PIVOT (
    SUM(anzahl)
    FOR GESCHLECHT_ID
    IN (
        1 AS "männlich",
        2 AS "weiblich"
    )
)
ORDER BY LAU2_NAME;

```

Jahre und Altersgruppen

```

SELECT *
FROM (
    SELECT JAHR, ALTERSGRUPPE_ID, SUM(BEVOELKERUNGSANZAHL) AS
anzahl
    FROM FAKT_GESAMTBEOELKERUNG
    GROUP BY JAHR, ALTERSGRUPPE_ID
)

```

```

PIVOT (
    SUM(anzahl)
    FOR ALTERSGRUPPE_ID
    IN (
        'AGE_0_TO_4' AS "0 BIS 4",
        'AGE_5_TO_9' AS "5 BIS 9",
        'AGE_10_TO_14' AS "10 BIS 14",
        'AGE_15_TO_19' AS "15 BIS 19",
        'AGE_20_TO_24' AS "20 BIS 24",
        'AGE_25_TO_29' AS "25 BIS 29",
        'AGE_30_TO_34' AS "30 BIS 34",
        'AGE_35_TO_39' AS "35 BIS 39",
        'AGE_40_TO_44' AS "40 BIS 44",
        'AGE_45_TO_49' AS "45 BIS 49",
        'AGE_50_TO_54' AS "50 BIS 54",
        'AGE_55_TO_59' AS "55 BIS 59",
        'AGE_60_TO_64' AS "60 BIS 64",
        'AGE_65_TO_69' AS "65 BIS 69",
        'AGE_70_TO_74' AS "70 BIS 74",
        'AGE_75_TO_79' AS "75 BIS 79",
        'AGE_80_TO_84' AS "80 BIS 84",
        'AGE_85_TO_89' AS "85 BIS 89",
        'AGE_90_TO_94' AS "90 BIS 94",
        'AGE_90_PLUS' AS "[Legacy] ÜBER 90",
        'AGE_95_TO_99' AS "95 BIS 99",
        'AGE_100_PLUS' AS "ÜBER 100"
    )
)
ORDER BY jahr;

```

Jahre und Geschlechter

```

SELECT *
FROM (
    SELECT JAHR, GESCHLECHT_ID, SUM(BEVOELKERUNGSANZAHL) AS anzahl
    FROM FAKT_GESAMTBEVOELKERUNG
    GROUP BY JAHR, GESCHLECHT_ID
)
PIVOT (
    SUM(anzahl)
    FOR GESCHLECHT_ID
    IN (
        1 AS "männlich",
        2 AS "weiblich"
    )
)
ORDER BY jahr;

```

9 Resümee

Wie zu erwarten war, geht aus den Daten hervor, dass die Bevölkerung in Oberösterreich grundsätzlich wächst. Spannend zu beobachten war vor allem die Entwicklung der Altersgruppen, etwa in Form von Bevölkerungspyramiden (siehe 8.2). Auch interessant zu sehen ist, dass in Linz die größte Differenz zwischen männlichen und weiblichen Einwohnern besteht, doch diese im Laufe der letzten Jahrzehnte stetig schrumpfte (siehe 8.6). Diese Entwicklung könnte man in Zukunft weiterverfolgen, um zu sehen, ob die Männer die Frauen irgendwann überholen. Auch den Anteil der nicht-Österreich stämmigen Einwohner (siehe 8.4) zu beobachten war nicht uninteressant, die Zuwanderung verläuft stetig, Flüchtlingswellen oder dergleichen gingen aus den Daten nicht hervor.

10 Bildverzeichnis

Abb. 1 - Ungewöhnlich starker Ausschlag der Zahlen im Jahr 2011 aufgrund der doppelten Daten	7
Abb. 2 - ERD-Diagramm der DWH-Struktur	8
Abb. 3 - Konfigurieren der Zugangsdaten zum Verbindungsaufbau mit dem Datenbank-Server	9
Abb. 4 - Import der CSV-Files mithilfe des Wizards von Datagrip	10
Abb. 5 - Staging-Tabellen mit den importierten CSV-Daten	10
Abb. 6 - Bevölkerungswachstum in Oberösterreich seit 1869	17
Abb. 7 - Bevölkerungsanzahl in den Städten	17
Abb. 8 - Altersverteilung OÖ 2001	18
Abb. 9 - Altersverteilung OÖ 2021	18
Abb. 10 - Geschlechterverteilung in OÖ in %	19
Abb. 11 - Bevölkerung mit Staatsbürgerschaft	19
Abb. 12 - Staatsbürgerschaft Verteilung in den Städten 2011 / 2021	20
Abb. 13 - Geschlechterverteilung Städte 2021	21
Abb. 14 - Geschlechterverteilung Linz 2001 / 2011 / 2021	21
Abb. 15 - Altersverteilung in Linz 2001 / 2011 / 2021	22