

Modul ddi: Datenbankdesign und Implementierung

Mini-Challenge zu LE2: Relationale Datenbanken (SQL)

Thema: Datenmigration

Stand des Dokuments: 18.09.2023, Andrea Kennel

Für LE2 ist Teil des Leistungsausweises eine benotete Abgabe. Dazu kann dieser Mini-Challenge abgegeben werden. Diese Mini-Challenge darf als Gruppe von maximal 4 Personen abgegeben werden.

Zum gleichen Thema gibt es ein Mini-Challenge im Modul daw: Data Wrangling. Die Zusatzfragen am Schluss der Aufgabenstellung können von Teams beantwortet werden, die beide Mini-Challenges gelöst haben.

Ausgangslage

Die Geschichte des schweizerischen Parlaments soll recherchiert werden. Dazu solle eine Datenbank für die Analysen erstellt werden. Die Daten werden vom Internet geholt:

<https://www.parlament.ch/de/ratsmitglieder>

Die Daten können vom obigen Link geholt werden oder stehen im Excel Ratsmitglieder_1848_DE.xlsx zur Verfügung.

Die Daten sind nicht in 3 Normalform und es interessieren nicht alle Attribute. Daher sollen die Daten in einer Datenbank besser strukturiert werden. Ziel ist es, dass mindestens folgende Fragestellungen beantwortet werden können.

Fragestellungen

Wer war von wann bis wann in welchem Rat?

Erstelle eine Liste mit folgender Struktur:

Name, Vorname, Kanton, Erster Eintritt in Nationalrat, Letzter Austritt aus Nationalrat, Erster Eintritt in Ständerat, Letzter Austritt aus Ständerat, Erster Eintritt in Bundesrat, Letzter Austritt aus Bundesrat

Wer war am längsten in einem Rat?

Erstelle eine Liste mit folgender Struktur:

Name, Vorname, Kanton, Rat, Länge der Ratsmitgliedschaft ohne Unterbruch (in Jahren), Datum des Eintritts in den Rat, Datum des Austritts aus dem Rat

Wer kam direkt in den Bundesrat?

Erstelle eine Liste mit folgender Struktur:

Name, Vorname, Kanton, Erster Eintritt in Bundesrat, Letzter Austritt aus Bundesrat
Dies nur für Personen, die weder im Nationalrat noch im Ständerat waren.

Wie hat sich die Parteienstärken über die Jahrhunderte verändert?

Wähle als Stichtag immer den 1.1 jedes Jahres.

Erstelle eine Liste mit folgender Struktur:

Rat, Jahr, Partei, Parteistärke (Anzahl Personen dieser Partei). Sortieren Sie diese Liste nach Rat, Jahr und Parteistärke.

Erstelle eine geeignete Grafik, die die Parteienstärke über die Zeit darstellt.

Wie war die «Zauberformel» (Parteienverteilung im Bundesrat) und wie gut stimmte diese mit den Parteistärken in Nationalrat und Ständerat überein?

Wähle als Stichtag immer den 1.1 jedes Jahres.

Erstelle eine Liste mit folgender Struktur:

Jahr, Partei, prozentuale Parteistärke in der Bundesversammlung (Nationalrat und Ständerat zusammen), prozentuale Parteistärke im Bundesrat. Sortieren Sie diese Liste nach Jahr und Parteistärke in der Bundesversammlung.

Erstelle eine geeignete Grafik, die das Verhältnis Parteienstärke in der Bundesversammlung und im Bundesrat über die Zeit darstellt.

Wie hat sich die Frauenquote in den drei Räten historisch entwickelt?

Erstelle eine Liste mit folgender Struktur:

Jahr, Rat, Anzahl Mitglieder, Anzahl Frauen, Anzahl Männer, Frauenquote

Visualisiere die Frauenquote über alle Jahre im Datensatz. Stelle zusätzlich auch die jeweilige Ratsgrösse graphisch dar, um zu prüfen, ob die Zahlen einigermaßen verlässlich sind.

Aufgabe 1

Die Daten aus dem Excel sollen so wie diese vorhanden sind, in eine Tabelle in die Datenbank geladen werden, um dann weiter verarbeitet zu werden.

Erstelle eine entsprechende Tabelle und lade die Daten aus dem Excel hinein.

Hinweis: Am einfachsten ist es, wenn alle Daten als String in die Datenbank geladen werden.

Hinweis: Die meisten Tools erlauben einen Import ab Excel.

Aufgabe 2

Definiere ein Datenmodell, mit welchem die Fragestellungen einfach und sinnvoll beantwortet werden können. Achte darauf, dass das Datenmodell keine Redundanzen aufweist. Bei Daten, die sich über die Zeit ändern, wie Parteinaamen, soll der zuletzt gültige Wert gespeichert werden. Lasse dieses Datenmodell von der FE überprüfen.

Aufgabe 3

Fülle das Zieldatenmodell, so dass die Daten ohne Redundanz in der Datenbank gespeichert sind. Falls nötig sollen Daten bereinigt werden.

Tipp: Es kann sich lohnen, Zwischenresultate in Hilfstabellen zwischen zu speichern.

Dazu kann man folgende Syntax verwenden:

```
CREATE TABLE temp1 AS
```

```
SELECT ....
```

```
FROM ...
```

Aufgabe 4

Beantworte die Fragestellungen und halte die Resultate in einem Dokument fest.

Aufgabe 5 (fakultativ)

Dieselbe Fragestellung wird in einer Mini-Challenge im Modul daw angeboten. Das Ziel dahinter ist ein Vergleich von verschiedenen Lösungsansätzen. Es soll ersichtlich werden, für welche Probleme welche Tools sich besser eignen.

Insbesondere können folgende Fragen beantwortet werden:

- Welche der konkreten Fragen konnten einfacher mit Python, welche mit SQL gelöst werden und wieso?
- Mit welcher Technologie war die Datenanalyse einfacher und wieso?
- Mit welcher Technologie war das Strukturieren der Daten einfacher und wieso?
- Sind die Resultate mit beiden Technologien identisch und wieso?

Wer die Antworten abgibt kann in einer modulübergreifenden Sprechstunde mit den FE die Antworten reflektieren.