



Database Systems - Assignment Project

Prof. Dr. A. Voisard, M.-U. Karagülle

A.10

Deadline: **Monday, July 1, 2024, 2 pm**

Task 1: Project idea

(10 %)

- 1 EN:** Have a look at the data sets *hochschulen.csv*¹ and *bevoelkerung.csv*².
DE: Schaue dir die Datensätze *hochschulen.csv*¹ und *bevoelkerung.csv*² an.
- 2 EN:** Conceptualize a data visualization that integrates and effectively presents information from the given datasets. Prepare a brief written description of your project goal.
DE: Konzipiere eine Datenvisualisierung, die Informationen aus den gegebenen Datensätzen integriert und effektiv präsentiert. Bereite eine kurze schriftliche Beschreibung deines Projektziels auf.

Task 2: Data schema and database set up

(20 %)

- 1 EN:** Identify the relevant data fields for analysis and visualization.
DE: Identifiziere die relevanten Datenfelder für die Analyse und Visualisierung.
- 2 EN:** Draw an Entity-Relationship model (ERM) for the datasets.
DE: Zeichne ein Entity-Relationship-Modell (ERM) für die Datensätze.
- 3 EN:** Transform the ERM into a relational model.
DE: Transformieren Sie das ERM in ein relationales Modell.
- 4 EN:** Set up a PostgreSQL database based on the relational model.
DE: Richte eine PostgreSQL-Datenbank basierend auf Ihrem relationalen Modell ein.

¹Attached; Source: <https://daten.berlin.de/datensaetze/universitaeten-berlin-und-deutschland>

²Attached; Source: <https://www.statistikportal.de/de/bevoelkerung#datenbanken>



Task 3: Data pre-processing and import (10 %)

- 1 **EN:** Pre-process (data cleansing) the data sets.
DE: Verarbeite die Datensätze vor (Datenbereinigung).
- 2 **EN:** Import the data sets into the created database.
DE: Importiere die Datensätze in die erstellte Datenbank.

Task 4: Queries and Analysis (25 %)

- 1 **EN:** Formulate specific SQL queries to extract relevant information from the data. Examples: number of students per city, gender distribution of students in federal states, comparison of student numbers and population in various regions.
DE: Formuliere spezifische SQL-Abfragen, um relevante Informationen aus den Daten zu extrahieren. Beispiele: Anzahl der Studierenden pro Stadt, Geschlechterverteilung der Studierenden in den Bundesländern, Vergleich der Studierendenzahlen und Bevölkerungszahlen in verschiedenen Regionen.

Task 5: Visualization of Results (20 %)

- 1 **EN:** Create visualizations of the query results. Prefer a map visualizations. Use appropriate visualization tools such as *Plotly*, *Geopandas*, *Matplotlib*, and *Seaborn* (or other).
DE: Erstelle Visualisierungen der Abfrageergebnisse. Bevorzuge Kartenvisualisierungen. Nutze geeignete Visualisierungstools wie *Plotly*, *Geopandas*, *Matplotlib* und *Seaborn* (oder andere).

Task 6: Presentation (10 %)

- 1 **EN:** Prepare presentation slides with appropriate formatting and complete information.
DE: Bereite Präsentationsfolien mit geeigneter Formatierung und vollständigen Informationen vor.
- 2 **EN:** Present the approach with a sufficient description of all work steps (5 min).
DE: Präsentiere den gewählten Ansatz mit einer ausreichenden Beschreibung aller Arbeitsschritte (5 min).
- 3 **EN:** Show a live demo of the solution (1 min)
DE: Zeige eine Live-Demo der Lösung (1 min)

Task 7: Documentation

(15 %)

- 1 EN:** Document all key steps, challenges, and decisions in a PDF in an appropriate format.
DE: Dokumentiere in einer PDF in geeigneter Form alle wesentlichen Schritte, Herausforderungen und Entscheidungen.
- 2 EN:** Upload the documentation, the source code ZIP, and the presentation file to KVV/Whiteboard by the specified deadline.
DE: Lade die Dokumentation, die ZIP des Quellcodes und die Präsentationsdatei fristgerecht auf KVV/Whiteboard hoch.

Task 8: Additional task for 7 CP students

(20 %)

Hint: Only students of the according study regulation can get the 'extra' CP.

Both subtasks have to be implemented.

Hinweis: Nur Studierende der entsprechenden Studienordnung können die 'zusätzlichen' LP erhalten.

Beide Teilaufgaben müssen implementiert werden.

- 1 EN:** Integrate information from at least one publicly available additional demographic data set (e.g., average age, average income, unemployment rate) into the visualization concept. Provide a brief description of how this data enhances your analysis.
DE: Integriere Informationen aus mindestens einem zusätzlichem und öffentlich zugänglichen demografischen Datensatz (z.B., Durchschnittsalter, Durchschnittseinkommen, Arbeitslosenquote) in das Visualisierungskonzept. Gebe eine kurze Beschreibung, wie diese Daten deine Analyse verbessern.
- 2 EN:** Implement advanced sorting and filter functions, which are applied on the result set of your queries. For example, implement functions that allow sorting by different criteria and filtering based on specific conditions.
DE: Implementiere erweiterte Sortier- und Filterfunktionen, die auf das Ergebnis deiner Abfragen angewendet werden. Implementiere zum Beispiel Funktionen, die das Sortieren nach verschiedenen Kriterien und das Filtern basierend auf spezifischen Bedingungen ermöglichen.