BIG DATA & BUSINESS ANALYTICS
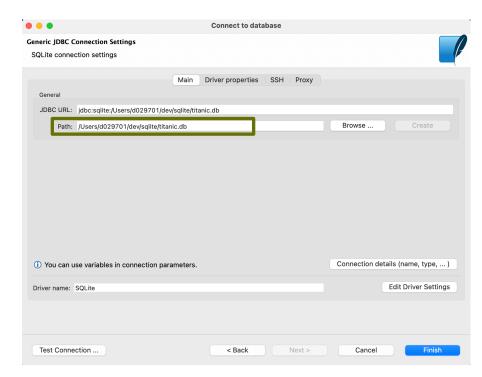
DATA ENGINEERING

SQL EXERCISE 2

CREATE NEW DATABASE FOR TITANIC

Create a new database connection:

- Database -> New Database Connection

- Choose SQLite

- As databae file, enter titanic.db
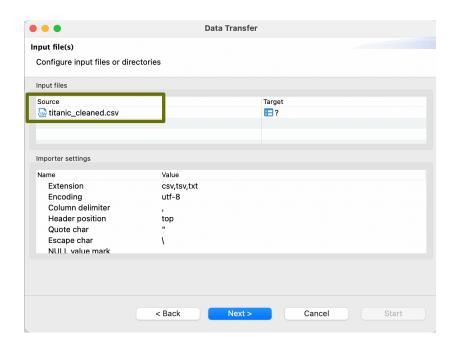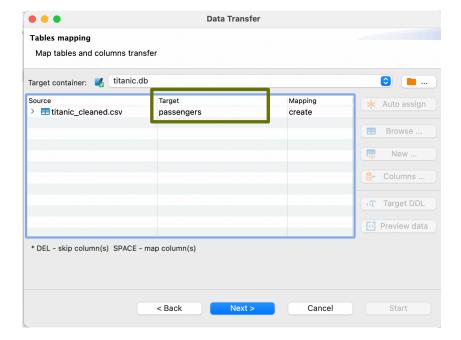 (file does not exist yet)

IMPORT CSV DATA

On the "titanic.db" connection, choose "Import Data"

- Select the source file "titanic_cleaned.csv"
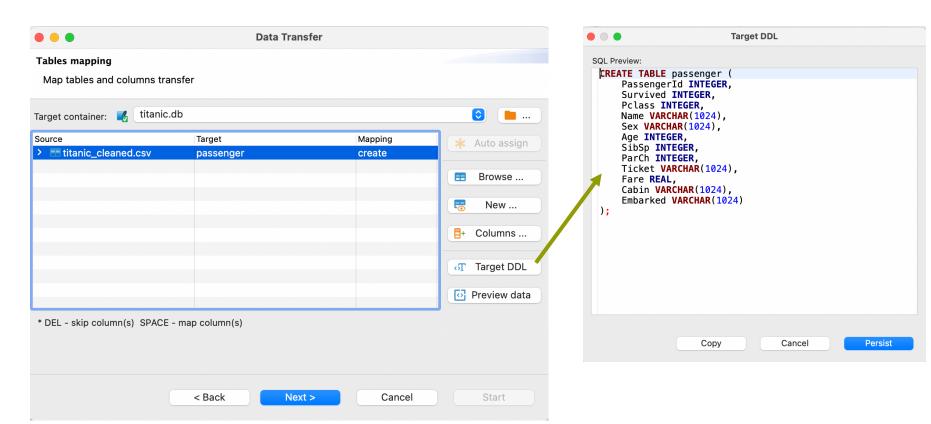- Enter "passengers" as name of target table

NO PRIMARY KEY

Apparently DBeaver does not support the creation of primary key constraints during CSV import.

Adding a primary key constraint afterwards is not supported by Sqlite. Something like

  ALTER TABLE passengers **ADD CONSTRAINT** pk_id **PRIMARY KEY** PassengerId

is not possible, see https://www.sqlite.org/omitted.html

EXERCISES
ON TITANIC DATA

EXERCISES

In the Titanic passenger list,

1. Count the number of passengers and the average age in each class

2. Count the number of passengers in each class that have survived / not survived

3. Determine the embarkment port (S, C, Q) where most first-class passengers boarded

4. For passenger Mr. James Moran, set the missing age value to 28
   Hint: UPDATE ... SET ... WHERE

5. Determine whether more male or female passengers travelled with at least one sibling or spouse

6. Select all last names (defined as prefix of Name before the comma)
   Hint: https://sqlite.org/lang_corefunc.html

7. For each last name, count the number of passengers →

	lastname	count
1	Andersson	9
2	Sage	7
3	Skoog	6
4	Rice	6

8. The average age is 24.064 years. What is the median age?
   Hint: SELECT ... LIMIT ... OFFSET