**TASK 1**

1. Выбрал дамб базы данных из кагла:

**NYC Yellow Taxi Trip Records -> yellow\_tripdata\_2015-01**

Содержит записи всех поездок Yellow Taxi в Нью-Йорке за январь 2015 года.

**Формат:** CSV.

**Размер файла:** ~1,94 ГБ.

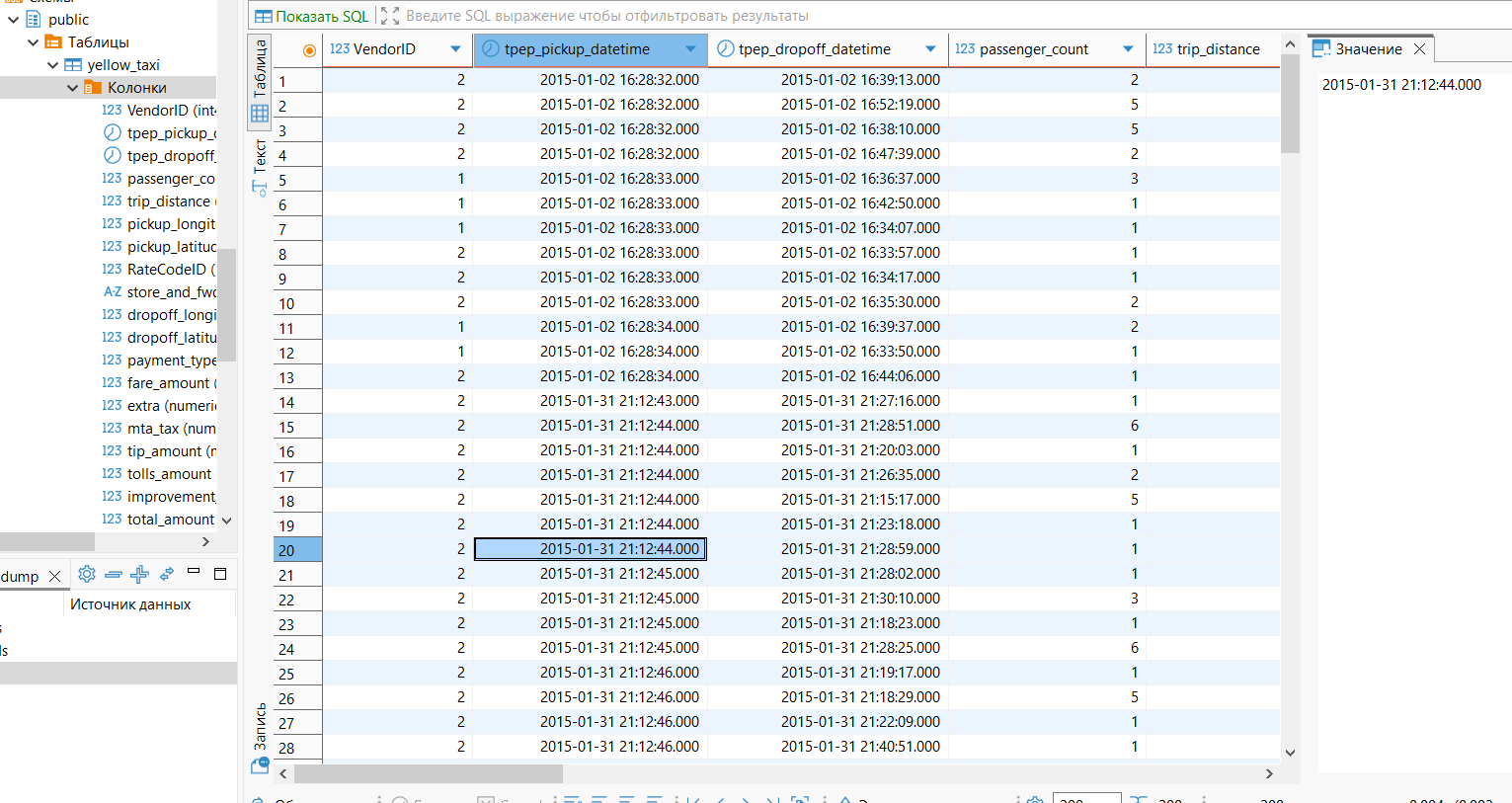
1. Установил **PostgreSQL** , сделал соединение, и создал бд в **DBeaver**-е:

**Название БД:** taxi\_source.  
**Таблица:** public.yellow\_taxi.

Вставил csv файл получил:

**Количество строк:** 12 748 986.

**Количество столбцов:** 19.



1. Создал вторую базы данных (**хранилище**):

**Название БД:** taxi\_dwh.  
**Схема:** dwh.  
**Таблица:** fact\_trips.

Сделал перенос данных с прошлой бд.

Перенос выполнялся с использованием расширения **dblink.**

Сократил количество столбцов и так же добавил -**trip\_duration\_minutes**(tpep\_dropoff\_datetime" -"tpep\_pickup\_datetime).

**Количество строк:** 12 748 986 .

**Количество столбцов:** 5.

