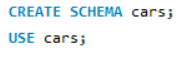
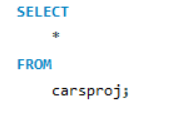
**CAR DEALER DATASET**

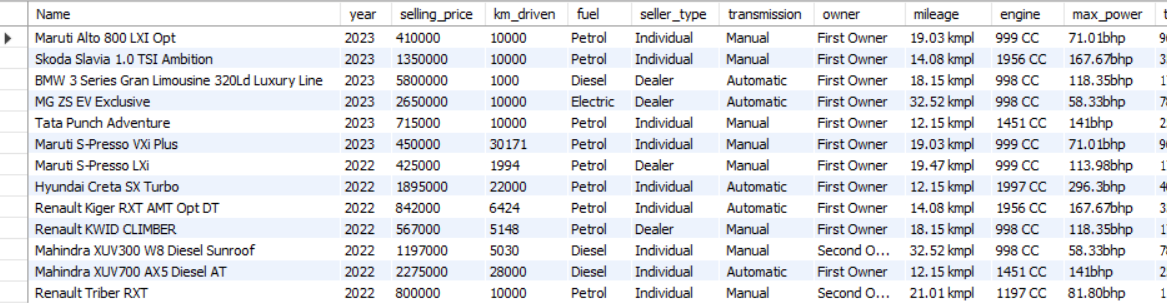
**DATA CLEANING USING MYSQL**

1. **Schema named ‘CARS’ created**

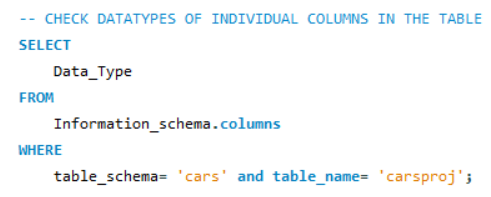
****

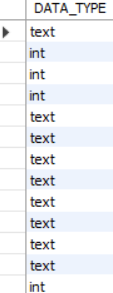
1. **Data Discovery**

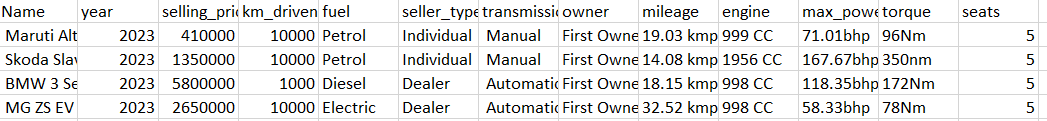
****

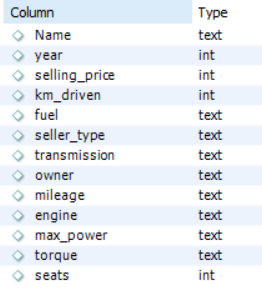
****

1. **Check datatypes of individual columns of the cars table**

****

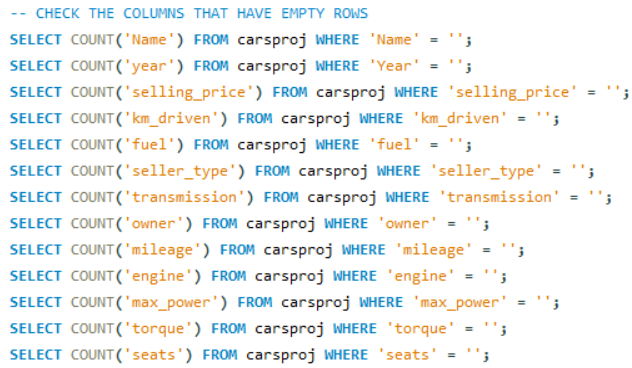
****

****

****

No datatype mis-match

1. **Check for missing rows in each column**

****

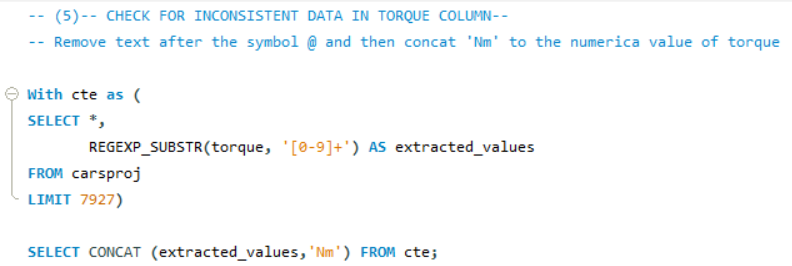
****

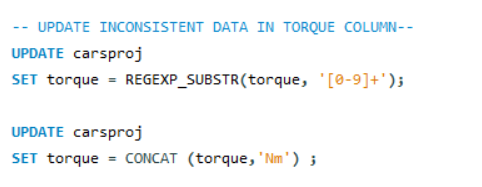
No missing rows

1. **Check for data inconsistencies**

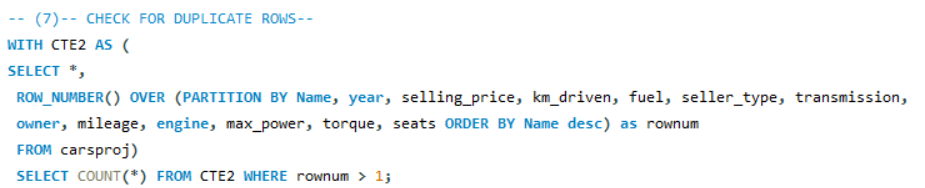
The column ‘torque’ contains data with inconsistency in the units. Therefore, only the numeric value was extracted and separately added in a new column. The unit ‘Nm’ was concatenated to the torque value.

1. Unit mismatch

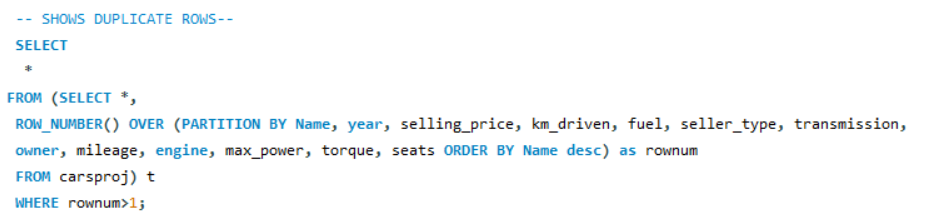


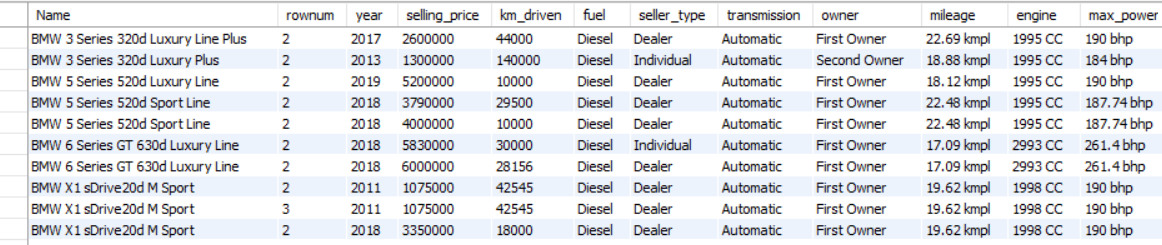


1. **Check for duplicate rows**

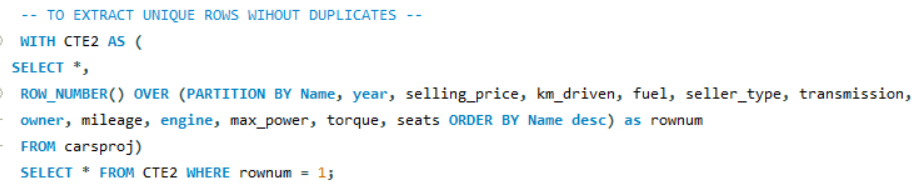


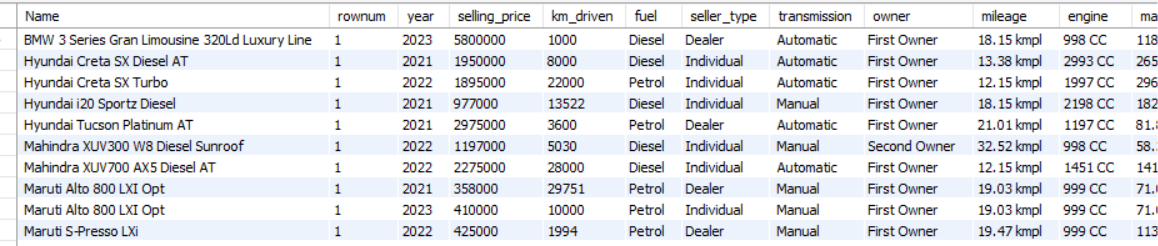






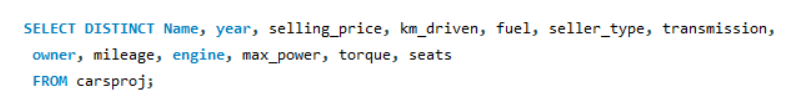






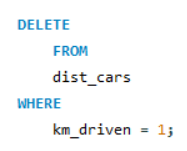


To display table with unique 6738 rows



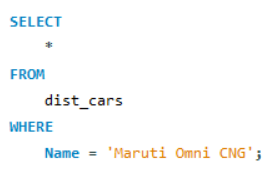
1. **Check for data inconsistencies in the unique rows**

* **km\_driven** data irrelavant for a single record. Hence it need to be removed

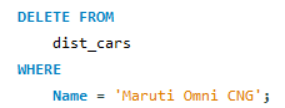




A row containing max\_power as only unit ‘bhp’ and torque as ‘NULL’ is unnecessary to be included. Hence to be removed.



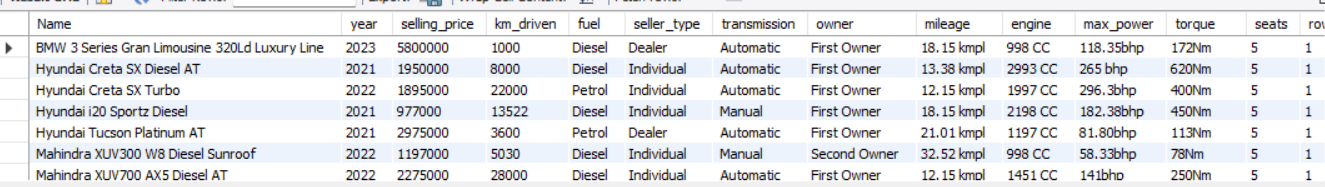












Delete column named rownum

