DADT MIDTERM REPORT

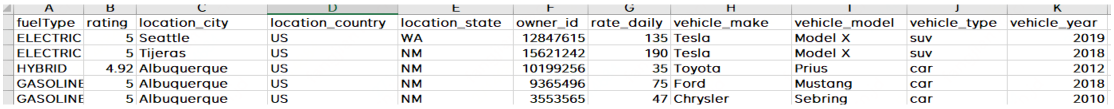
Part1: About the dataset:

Data Resource: https://www.kaggle.com/datasets/kushleshkumar/cornell-car-rental-dataset

- **Quality:** This is an open dataset on car rental information in major US cities in July 2020. It was scraped from the web, but the author didn't offer its original source. This dataset may not be authoritative and accurate enough, but it is fully capable of being used as a source for machine learning, data analysis, and database testing.
- **Detail:** This dataset has information of More than 5500 Rental Cars, including Make, Model, Year, Fare, Location and Fuel Type.
- **Documentation:** This dataset is standard tabular data. The author has marked the detailed description of the data for each column of data, so that users can accurately grasp the meaning of each data. All the data are marked prominently on the webpage, and there are corresponding text and graphic descriptions. When users encounter problems, it is very easy to find the corresponding reference content.
- **Interrelation:** The data contains the detailed geographic location of the vehicle, including latitude and longitude information. This information can be connected with databases such as Google Maps to obtain accurate positioning. The data also includes city, state, and country information of the city where the vehicle is located, which can be connected with other geographic location (city, state, etc.) databases to obtain similar information such as local population, average income level, and weather etc. More access to data information will allow us to provide more comprehensive and useful data information for further use by database users.
- **Use:** I would like to analyze the preferences of the US car rental market from this database.
  - What is main stream cars by fuel type? Gasoline or electric?
  - Which car is the most popular one? Made by which car maker?
  - Which city/state owns most renting cars?
  - Which car model has the highest renting fee rate?
- **Discoverability:** Car rental is a business. Large car rental companies such as HERZ and DAVIS have more detailed car rental data, but these data involve commercial confidentiality, and unless the above companies voluntarily provide some data sets, it is not easy to obtain.

Part2:Data Modeling

A glimpse of the data:



fuelType	rating	location_city	location_country	location_state	owner_id	rate_daily	vehicle_make	vehicle_model	vehicle_type	vehicle_year
ELECTRIC	5	Seattle	US	WA	12847615	135	Tesla	Model X	suv	2019
ELECTRIC	5	Tijeras	US	NM	15621242	190	Tesla	Model X	suv	2018
HYBRID	4.92	Albuquerque	US	NM	10199256	35	Toyota	Prius	car	2012
GASOLINE	5	Albuquerque	US	NM	9365496	75	Ford	Mustang	car	2018
GASOLINE	5	Albuquerque	US	NM	3553565	47	Chrysler	Sebring	car	2010

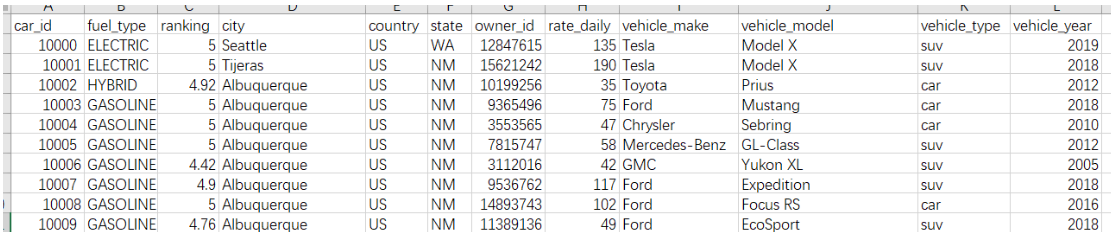
1、 This data is almost in 1NF because:

a) There are no mixing data types within the same column.

b) There are no repeating rows.

c) Row order does not convey any information.

But we still need to decide a Primary Key. As each row is about the rental summary of a specific car, we should add a car_id as the primary key for it, and it is finally in 1NF.



car_id	fuel_type	ranking	city	country	state	owner_id	rate_daily	vehicle_make	vehicle_model	vehicle_type	vehicle_year
10000	ELECTRIC	5	Seattle	US	WA	12847615	135	Tesla	Model X	suv	2019
10001	ELECTRIC	5	Tijeras	US	NM	15621242	190	Tesla	Model X	suv	2018
10002	HYBRID	4.92	Albuquerque	US	NM	10199256	35	Toyota	Prius	car	2012
10003	GASOLINE	5	Albuquerque	US	NM	9365496	75	Ford	Mustang	car	2018
10004	GASOLINE	5	Albuquerque	US	NM	3553565	47	Chrysler	Sebring	car	2010
10005	GASOLINE	5	Albuquerque	US	NM	7815747	58	Mercedes-Benz	GL-Class	suv	2012
10006	GASOLINE	4.42	Albuquerque	US	NM	3112016	42	GMC	Yukon XL	suv	2005
10007	GASOLINE	4.9	Albuquerque	US	NM	9536762	117	Ford	Expedition	suv	2018
10008	GASOLINE	5	Albuquerque	US	NM	14893743	102	Ford	Focus RS	car	2016
10009	GASOLINE	4.76	Albuquerque	US	NM	11389136	49	Ford	EcoSport	suv	2018

2、 To make it 2NF compliant, we need to make sure the non-key attribute in the table depend on the entire primary key.

As we have 1 single primary key, the other attributes are indeed dependent on it, so it is in 2NF

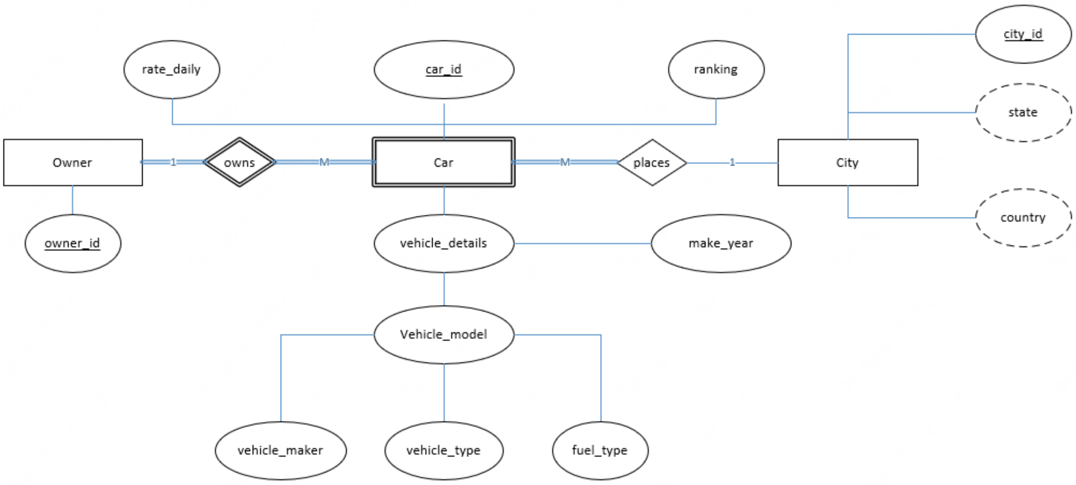
3、 Let's move on to see if it is 3NF compliant. We need to make sure every non-key attribute in the table depend on the key, the whole key, and nothing but the key.

And we can see several attributes do not meet this standard in this table.

a) We can deduce vehicle_model from the car_id, then from the vehicle_model, we further get information about the vehicle_maker, vehicle_type. So these attributes are indirectly dependent on the car_id. It violates the rule that it must depend on the key, nothing but the key. We can fix it by create a new vehicle_model table to avoid it.

b) Each car must locate in some place(here we have City, State,Country). But we can deduce State from City and thus Country from State(not directly from the car_id). So again it violates the rule. We can fix it by create a City table, State table and a Country(optional as there is only one country involved here) table to avoid the violation.

And we can create E/R model based on the above principles:

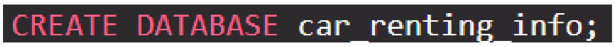


It is now 3NF compliant and as there's no multivalued dependency and each table can not be describe as being the logical result of joining other tables, so there's no need to further check 4NF and 5NF.

Part3: Create the database

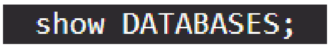
It is time to create the required data table based on the ER model.

1.Create a new database named car_renting_info

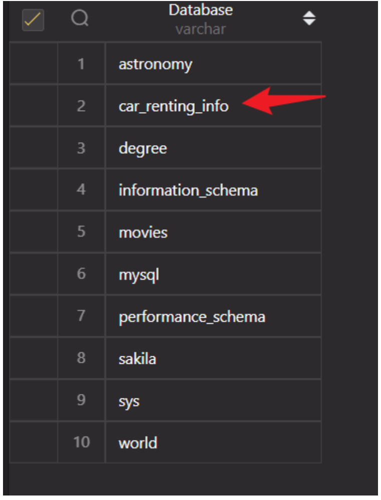


```
CREATE DATABASE car_renting_info;
```

Show databases;

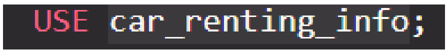


```
show DATABASES;
```



	Database varchar
1	astronomy
2	car_renting_info
3	degree
4	information_schema
5	movies
6	mysql
7	performance_schema
8	sakila
9	sys
10	world

And use it

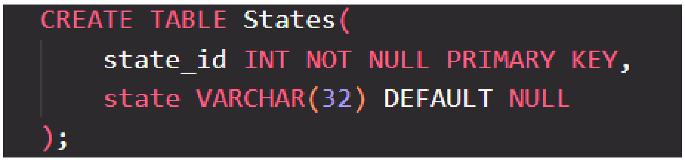


```
USE car_renting_info;
```

2. Create 4 tables, named Cars, Vehicle_details,Cities and States

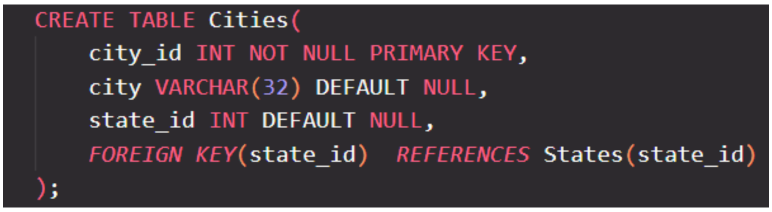
As some tables contain foreign keys constraints, so we need to create tables without constraints first.

2.01: Create table States, which includes state_id as PK and state name



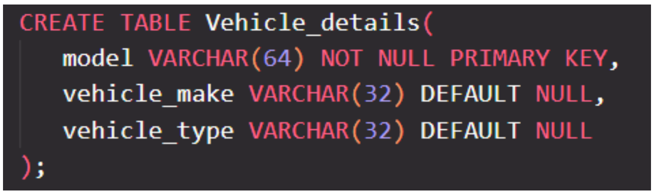
```
CREATE TABLE States(
    state_id INT NOT NULL PRIMARY KEY,
    state VARCHAR(32) DEFAULT NULL
);
```

2.02:Create table Cities, which has city_id as PK and state_id as foreign key



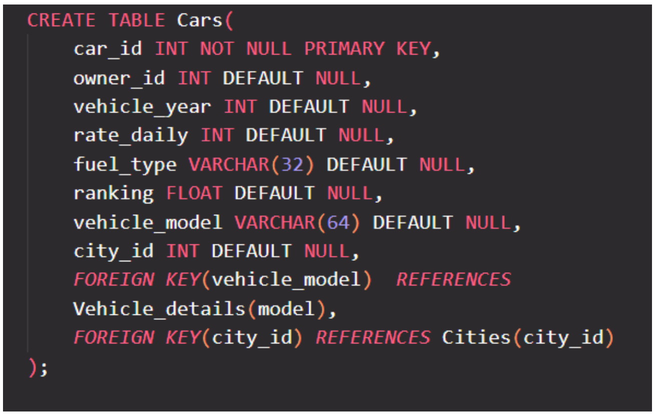
```
CREATE TABLE Cities(
    city_id INT NOT NULL PRIMARY KEY,
    city VARCHAR(32) DEFAULT NULL,
    state_id INT DEFAULT NULL,
    FOREIGN KEY(state_id)  REFERENCES States(state_id)
);
```

2.03:Create table Vehicle_details:



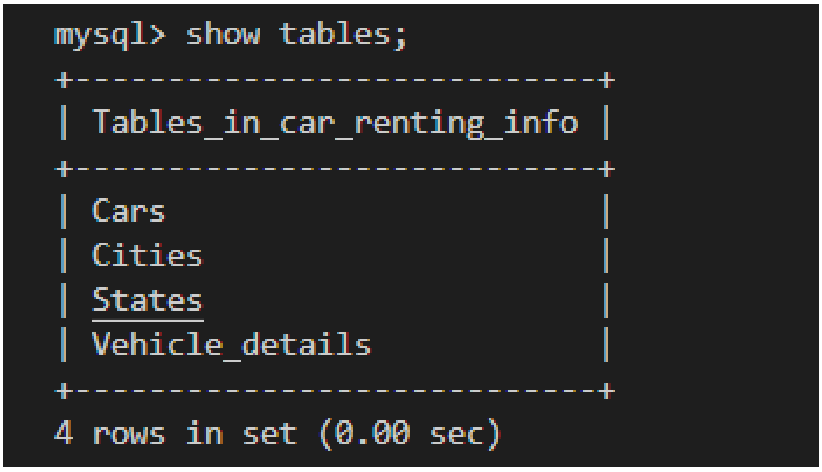
```
CREATE TABLE Vehicle_details(
    model VARCHAR(64) NOT NULL PRIMARY KEY,
    vehicle_make VARCHAR(32) DEFAULT NULL,
    vehicle_type VARCHAR(32) DEFAULT NULL
);
```

2.04:Create table Cars:



```
CREATE TABLE Cars(
    car_id INT NOT NULL PRIMARY KEY,
    owner_id INT DEFAULT NULL,
    vehicle_year INT DEFAULT NULL,
    rate_daily INT DEFAULT NULL,
    fuel_type VARCHAR(32) DEFAULT NULL,
    ranking FLOAT DEFAULT NULL,
    vehicle_model VARCHAR(64) DEFAULT NULL,
    city_id INT DEFAULT NULL,
    FOREIGN KEY(vehicle_model)  REFERENCES
    Vehicle_details(model),
    FOREIGN KEY(city_id) REFERENCES Cities(city_id)
);
```

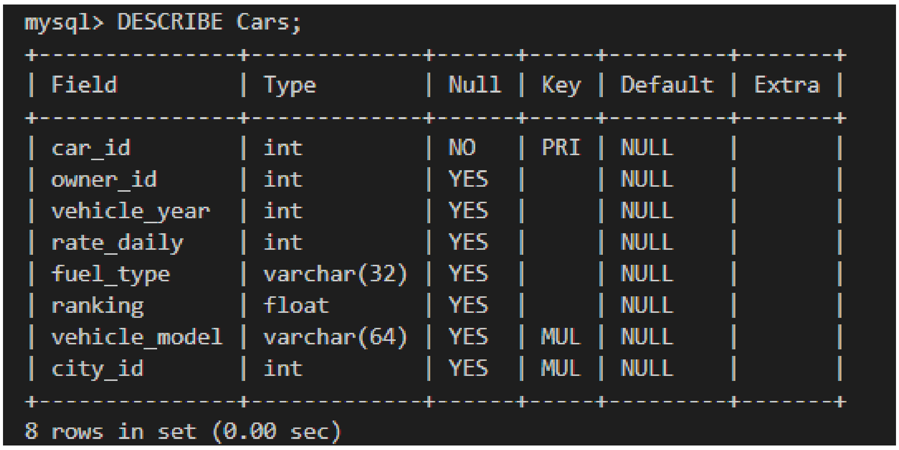
The final result:



```
mysql> show tables;
+---------------------------+
| Tables_in_car_renting_info |
+---------------------------+
| Cars                      |
| Cities                    |
| States                    |
| Vehicle_details           |
+---------------------------+
4 rows in set (0.00 sec)
```

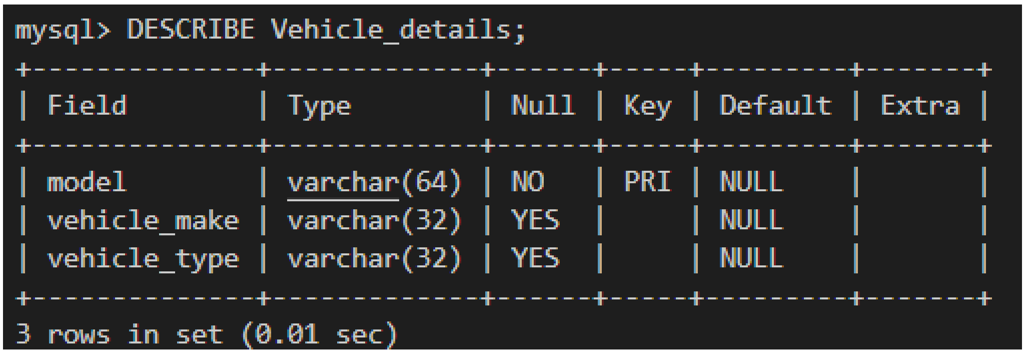
And table description:

Cars:



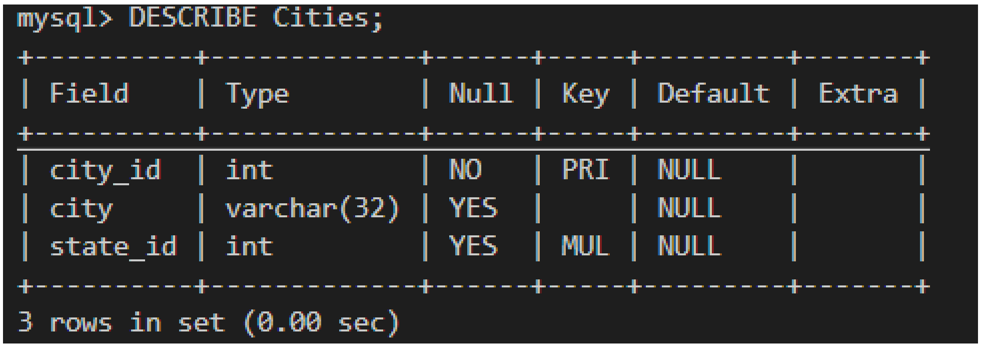
```
mysql> DESCRIBE Cars;
+---------------+-------------+------+-----+---------+-------+
| Field         | Type        | Null | Key | Default | Extra |
+---------------+-------------+------+-----+---------+-------+
| car_id        | int         | NO   | PRI | NULL    |       |
| owner_id      | int         | YES  |     | NULL    |       |
| vehicle_year  | int         | YES  |     | NULL    |       |
| rate_daily    | int         | YES  |     | NULL    |       |
| fuel_type     | varchar(32) | YES  |     | NULL    |       |
| ranking       | float       | YES  |     | NULL    |       |
| vehicle_model | varchar(64) | YES  | MUL | NULL    |       |
| city_id       | int         | YES  | MUL | NULL    |       |
+---------------+-------------+------+-----+---------+-------+
8 rows in set (0.00 sec)
```

Vehicle_details:



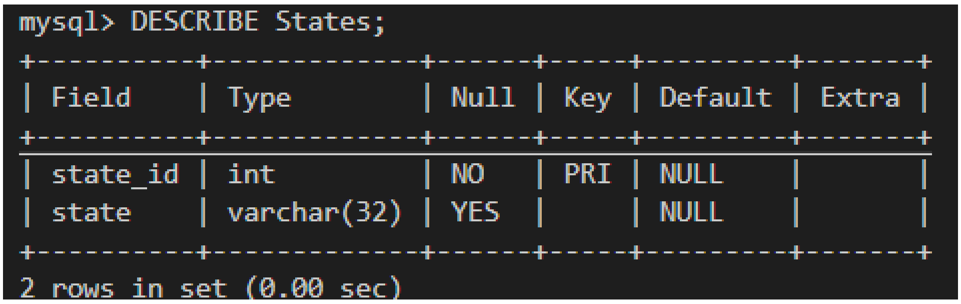
```
mysql> DESCRIBE Vehicle_details;
+--------------+-------------+------+-----+---------+-------+
| Field        | Type        | Null | Key | Default | Extra |
+--------------+-------------+------+-----+---------+-------+
| model        | varchar(64) | NO   | PRI | NULL    |       |
| vehicle_make | varchar(32) | YES  |     | NULL    |       |
| vehicle_type | varchar(32) | YES  |     | NULL    |       |
+--------------+-------------+------+-----+---------+-------+
3 rows in set (0.01 sec)
```

Cites:



```
mysql> DESCRIBE Cities;
+----------+-------------+------+-----+---------+-------+
| Field    | Type        | Null | Key | Default | Extra |
+----------+-------------+------+-----+---------+-------+
| city_id  | int         | NO   | PRI | NULL    |       |
| city     | varchar(32) | YES  |     | NULL    |       |
| state_id | int         | YES  | MUL | NULL    |       |
+----------+-------------+------+-----+---------+-------+
3 rows in set (0.00 sec)
```

States:

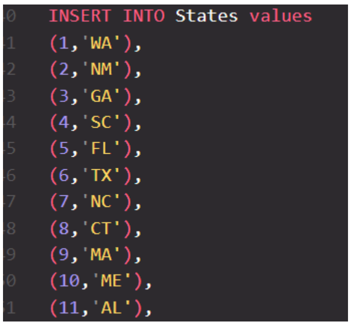


```
mysql> DESCRIBE States;
+----------+-------------+------+-----+---------+-------+
| Field    | Type        | Null | Key | Default | Extra |
+----------+-------------+------+-----+---------+-------+
| state_id | int         | NO   | PRI | NULL    |       |
| state    | varchar(32) | YES  |     | NULL    |       |
+----------+-------------+------+-----+---------+-------+
2 rows in set (0.00 sec)
```

3.Populate data into the right table.

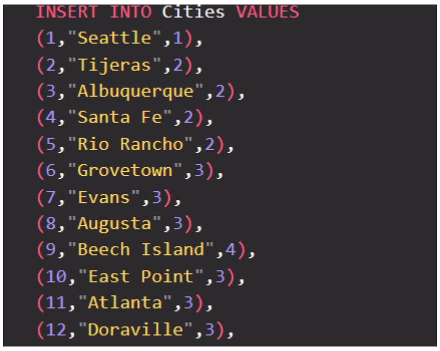
There are several ways of populating data into tables, such as using **INSERT** statement or load the .csv file into the table. As we don't have permission to upload files to the cloud project file, here I use INSERT and I give part of the code screenshots of how I did it.

States table:



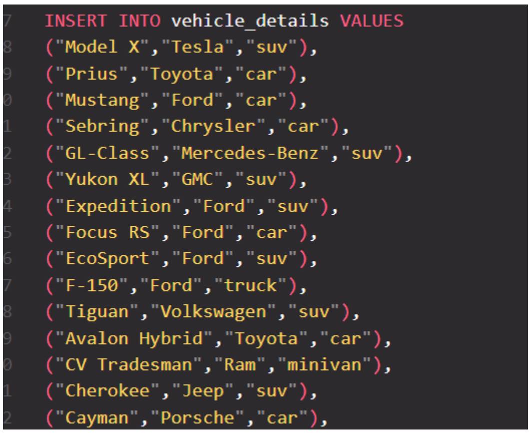
```
INSERT INTO States values
(1,'WA'),
(2,'NM'),
(3,'GA'),
(4,'SC'),
(5,'FL'),
(6,'TX'),
(7,'NC'),
(8,'CT'),
(9,'MA'),
(10,'ME'),
(11,'AL'),
```

Cities table:



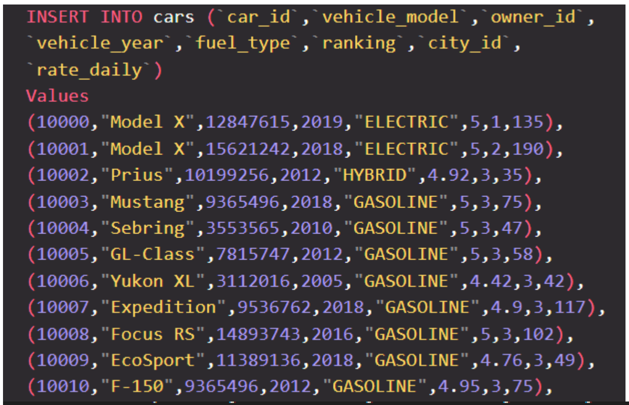
```
INSERT INTO Cities VALUES
(1,"Seattle",1),
(2,"Tijeras",2),
(3,"Albuquerque",2),
(4,"Santa Fe",2),
(5,"Rio Rancho",2),
(6,"Grovetown",3),
(7,"Evans",3),
(8,"Augusta",3),
(9,"Beech Island",4),
(10,"East Point",3),
(11,"Atlanta",3),
(12,"Doraville",3),
```

Vehicle_details table:

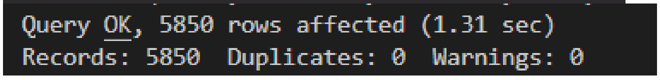


```
INSERT INTO vehicle_details VALUES
("Model X","Tesla","suv"),
("Prius","Toyota","car"),
("Mustang","Ford","car"),
("Sebring","Chrysler","car"),
("GL-Class","Mercedes-Benz","suv"),
("Yukon XL","GMC","suv"),
("Expedition","Ford","suv"),
("Focus RS","Ford","car"),
("EcoSport","Ford","suv"),
("F-150","Ford","truck"),
("Tiguan","Volkswagen","suv"),
("Avalon Hybrid","Toyota","car"),
("CV Tradesman","Ram","minivan"),
("Cherokee","Jeep","suv"),
("Cayman","Porsche","car"),
```

Cars table:



```
INSERT INTO cars (`car_id`,`vehicle_model`,`owner_id`,
`vehicle_year`,`fuel_type`,`ranking`,`city_id`,
`rate_daily`)
Values
(10000,"Model X",12847615,2019,"ELECTRIC",5,1,135),
(10001,"Model X",15621242,2018,"ELECTRIC",5,2,190),
(10002,"Prius",10199256,2012,"HYBRID",4.92,3,35),
(10003,"Mustang",9365496,2018,"GASOLINE",5,3,75),
(10004,"Sebring",3553565,2010,"GASOLINE",5,3,47),
(10005,"GL-Class",7815747,2012,"GASOLINE",5,3,58),
(10006,"Yukon XL",3112016,2005,"GASOLINE",4.42,3,42),
(10007,"Expedition",9536762,2018,"GASOLINE",4.9,3,117),
(10008,"Focus RS",14893743,2016,"GASOLINE",5,3,102),
(10009,"EcoSport",11389136,2018,"GASOLINE",4.76,3,49),
(10010,"F-150",9365496,2012,"GASOLINE",4.95,3,75),
```



```
Query OK, 5850 rows affected (1.31 sec)
Records: 5850  Duplicates: 0  Warnings: 0
```

Now let's answer the questions regarding car renting market with the sql scripts.

1. **What is main stream cars by fuel type? Gasoline or electric?**

```
"SELECT COUNT(*) as car_counts ,fuel_type from Cars GROUP BY fuel_type ORDER BY car_counts DESC",
```

2. **What car model is the most popular? Made by which car maker?**

```
"SELECT AVG(Cars.ranking) as car_ranking, Vehicle_details.vehicle_make,Cars.vehicle_model,COUNT(*) as count FROM Cars INNER JOIN Vehicle_details ON Cars.vehicle_model = Vehicle_details.model GROUP BY Cars.vehicle_model HAVING count > 50 ORDER BY car_ranking DESC;",
```

3. **Which city/state owns most renting cars?**

```
"SELECT COUNT(*) as car_counts,state from Cars INNER JOIN Cities ON Cars.city_id = Cities.city_id    INNER JOIN States  ON Cities.state_id = States.state_id    GROUP BY States.state    ORDER BY car_counts DESC     LIMIT 10",
```

4. Which car model has the highest renting fee rate?

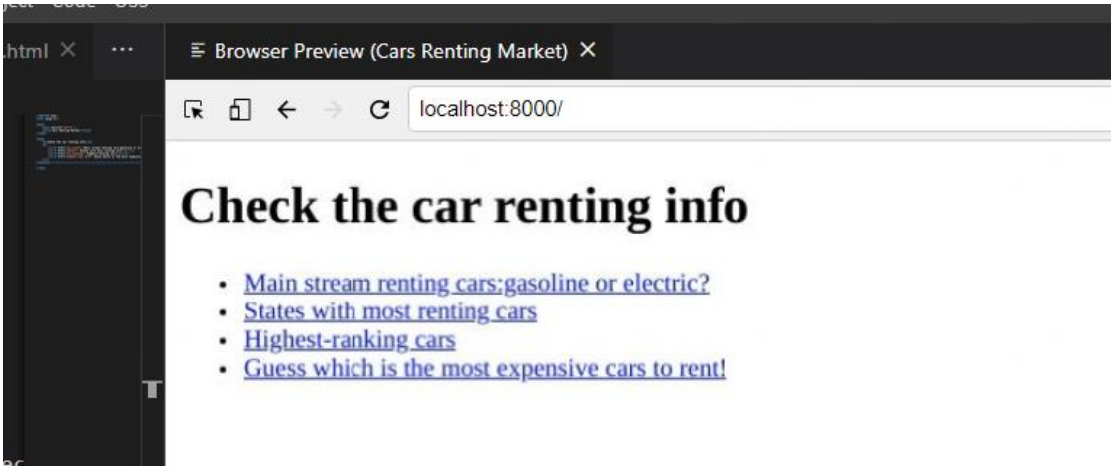
```
"SELECT AVG(Cars.rate_daily) as rate_daily_avg,Vehicle_details.vehicle_make,Vehicle_details.model FROM Cars INNER JOIN Vehicle_details ON Cars.vehicle_model = Vehicle_details.model GROUP BY Vehicle_details.model ORDER BY rate_daily_avg DESC LIMIT 20;"
```

I actually would like to do further analysis on the users or owners of the cars, however, as the data is limited(we do not have any user or the owner information), I had to give it up.

Part 4. Create a simple web application

I created a main page(default page) for this project. User can click the info they want and it will route to the right page.

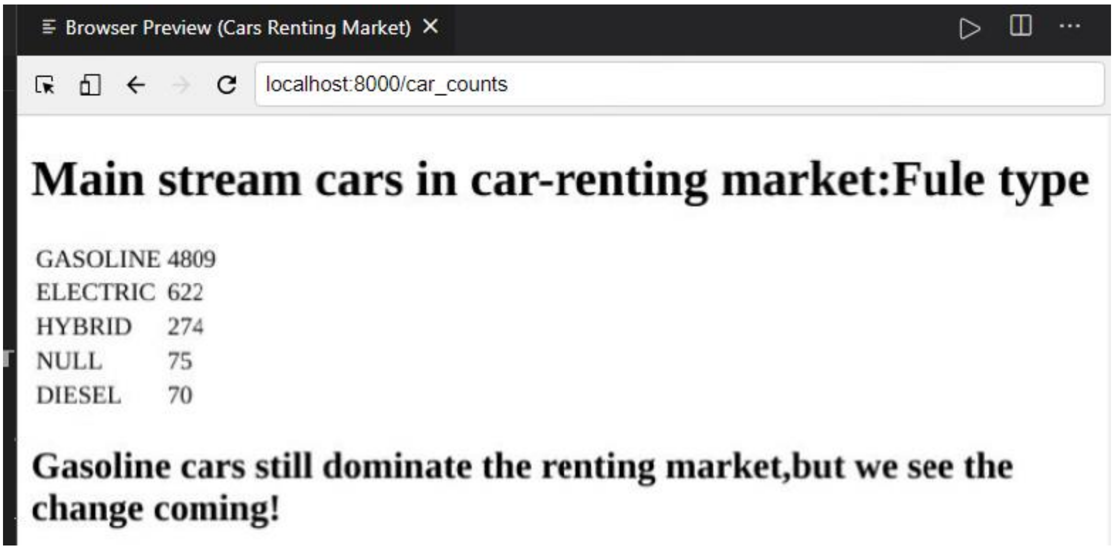
Default page:



Check the car renting info

- Main stream renting cars:gasoline or electric?
- States with most renting cars
- Highest-ranking cars
- Guess which is the most expensive cars to rent!

Main stream Fuel type:



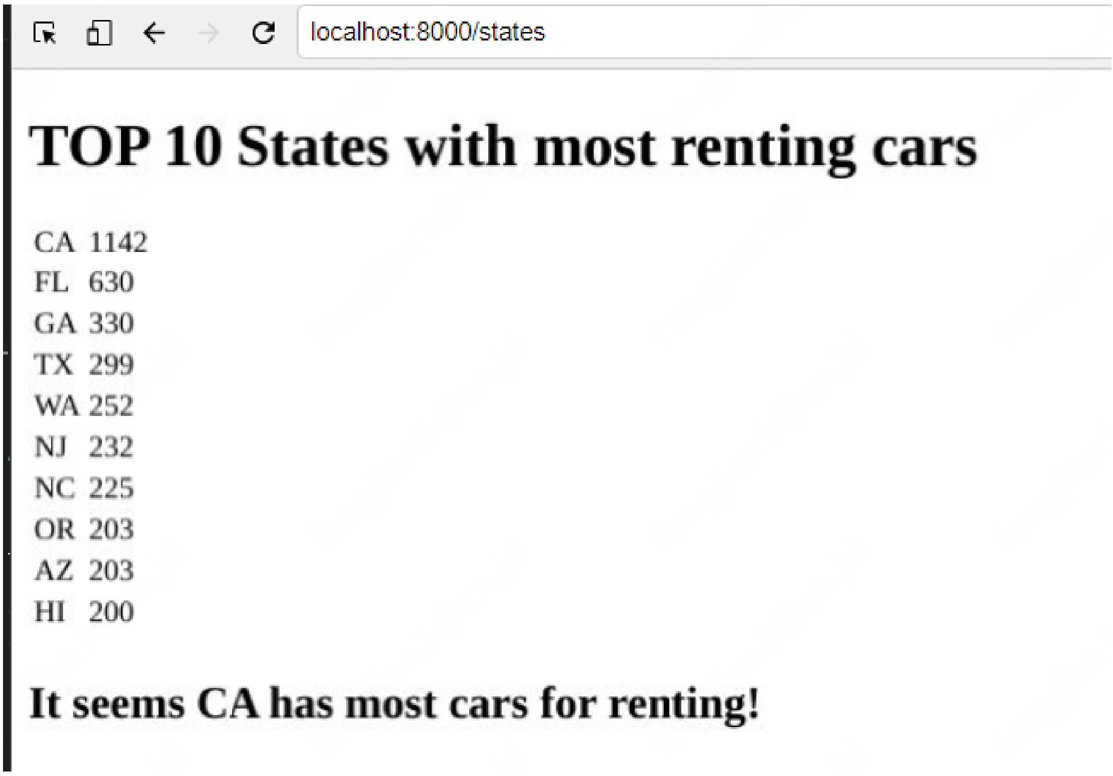
Browser Preview (Cars Renting Market)
localhost:8000/car_counts

Main stream cars in car-renting market:Fule type

```
GASOLINE 4809
ELECTRIC  622
HYBRID    274
NULL      75
DIESEL    70
```

Gasoline cars still dominate the renting market,but we see the change coming!

States with most cars for renting

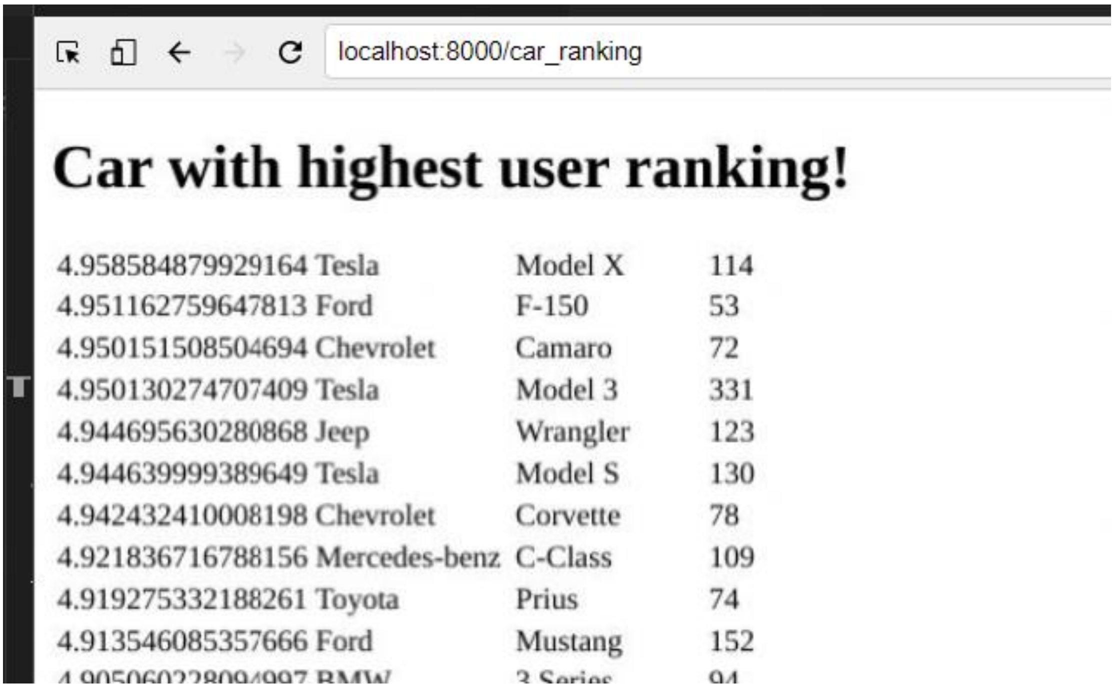


TOP 10 States with most renting cars

```
CA  1142
FL  630
GA  330
TX  299
WA  252
NJ  232
NC  225
OR  203
AZ  203
HI  200
```

It seems CA has most cars for renting!

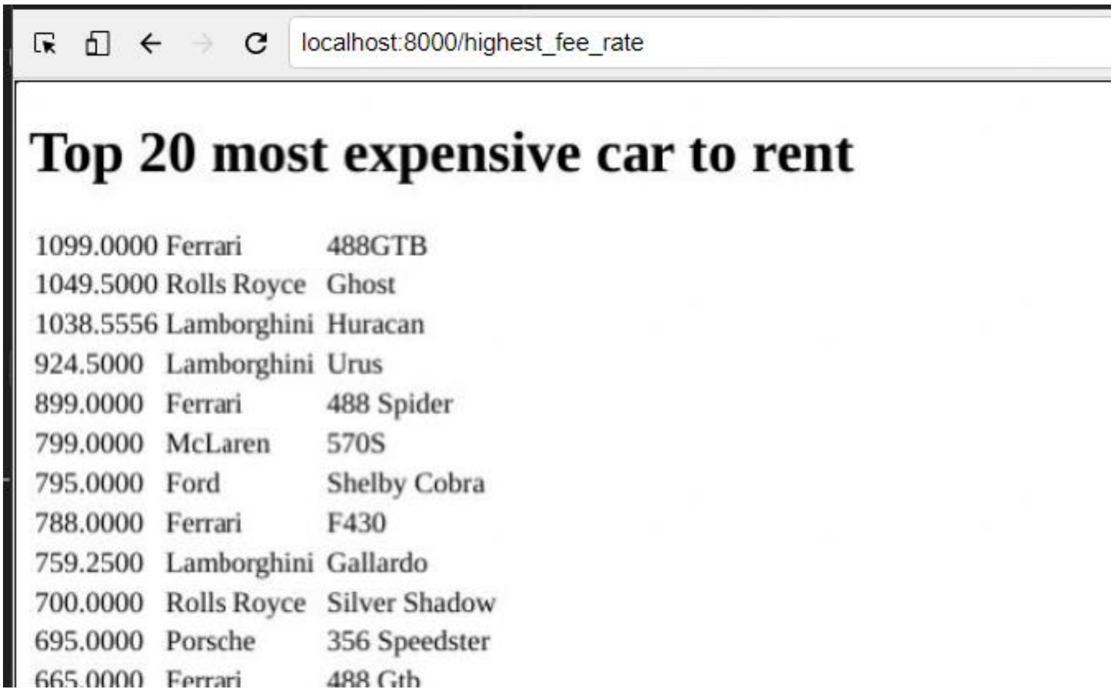
Most popular cars to rent



Car with highest user ranking!

```
4.958584879929164 Tesla          Model X     114
4.951162759647813 Ford           F-150       53
4.950151508504694 Chevrolet      Camaro      72
4.950130274707409 Tesla          Model 3     331
4.944695630280868 Jeep           Wrangler    123
4.944639999389649 Tesla          Model S     130
4.942432410008198 Chevrolet      Corvette    78
4.921836716788156 Mercedes-benz  C-Class     109
4.919275332188261 Toyota         Prius       74
4.913546085357666 Ford           Mustang     152
4.905060228094997 BMW            3 Series    94
```

Most expensive cars to rent

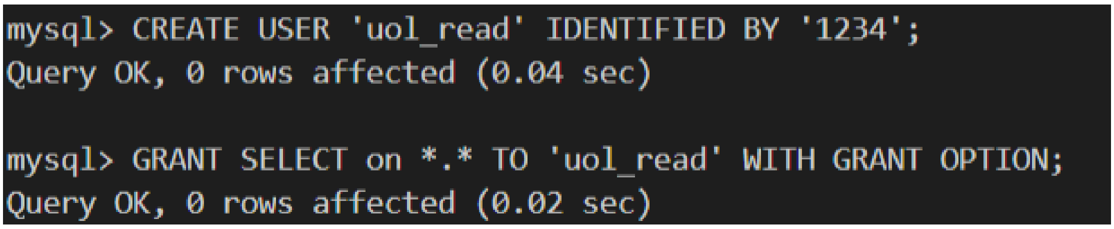


Top 20 most expensive car to rent

```
1099.0000 Ferrari      488GTB
1049.5000 Rolls Royce  Ghost
1038.5556 Lamborghini  Huracan
924.5000  Lamborghini  Urus
899.0000  Ferrari      488 Spider
799.0000  McLaren      570S
795.0000  Ford         Shelby Cobra
788.0000  Ferrari      F430
759.2500  Lamborghini  Gallardo
700.0000  Rolls Royce  Silver Shadow
695.0000  Porsche      356 Speedster
665.0000  Ferrari      488 Gtb
```

Grant only SELECT permission

The admin user(**root**) has all permissions and can perform operations such as create, add, delete, and update etc. I also created a new account(**uol_read**), and limit the account permissions to only **SELECT** with **GRANT OPTION**, but not making any changes to the data to assure security of the database and tables.

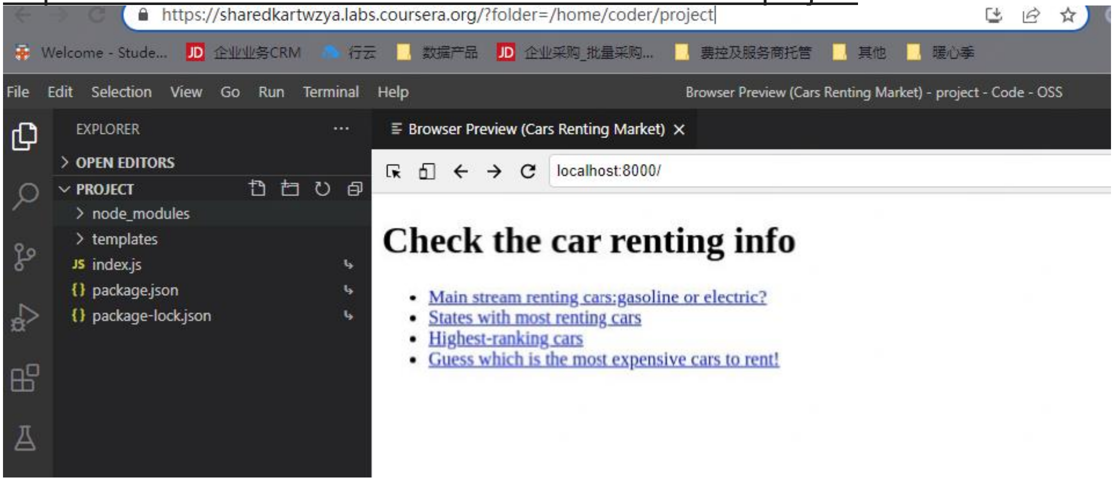


```
mysql> CREATE USER 'uol_read' IDENTIFIED BY '1234';
Query OK, 0 rows affected (0.04 sec)

mysql> GRANT SELECT on *.* TO 'uol_read' WITH GRANT OPTION;
Query OK, 0 rows affected (0.02 sec)
```

Test on UOL LAB with the shared link

https://hub.labs.coursera.org:443/connect/sharedkartwzya?forceRefresh=false&path=%2F%3Ffolder%3D%2Fhome%2Fcoder%2Fproject



https://sharedkartwzya.labs.coursera.org/?folder=/home/coder/project

localhost:8000/

Check the car renting info

- Main stream renting cars:gasoline or electric?
- States with most renting cars
- Highest-ranking cars
- Guess which is the most expensive cars to rent!