Preprocessing steps: Done using python code

1. Loading the dataset Vehicle Dataset 2024.
2. Find the null values for each field
3. Handle null/missing values
    a. Replace numerical values with median of the field
    b. Replace categorical values with value “Unknown”.
    Reason: To avoid bias in the data and errors in querying the data.
4. Convert the data into right datatypes.
5. Export the processed/ cleaned dataset to a new *preprocessed_vehicles_dataset.csv* file for querying in MongoDB.

Queries for the given questions using MongoDB.

1. Calculate the number of vehicles which are Jeeps.
   Query : db.processedVehicle.countDocuments({make: "Jeep"})
   Result : 194

2. Write a query to find no of vehicles which are SUV’s
   Query: db.processedVehicle.countDocuments({body: "SUV"})
   Result: 704

3. Find the number of vehicles which have less than four doors in the entire dataset.
   Query: db.processedVehicle.countDocuments({doors: {$lt: 4}})
   Result : 47

4. Write a query to find the number of vehicles which have mileage less than 20,000.
   Query: db.processedVehicle.countDocuments({mileage: {$lt: 20000}})
   Result : 1002

5. Retrieve the details of make and mileage of the first vehicle.
   Query : db.processedVehicle.find({}, {make: 1, mileage: 1}).limit(1)
   Result: [{_id: ObjectId('66842f6eb8a37d3222ec4dd0'),make: 'Jeep',mileage: 10}]

6. Find the number of vehicles bought in the year 2023.
   Query: db.processedVehicle.countDocuments({year: 2023})
   Result : 90

7. Find the maximum no.of doors that a car can have.
   Query: db.processedVehicle.aggregate([
   { $group: { _id: null, maxDoors: { $max: "$doors" } } },
   { $project: { _id: 0, maxDoors: 1 } }
   ])
   Result: { maxDoors: 5 }

8. Calculate the average price of SUV cars.
   Query: db.processedVehicle.aggregate([
   {$match: {body: "SUV"}},
   {$group: {_id: null, avgPrice: {$avg: "$price"}}},
   { $project: { _id: 0, avgPrice: 1 } }
   }])
   Result: { avgPrice: 47844.146306818184 }

9. Identify the no of vehicles with price greater than 40,000 less than 70,000.
   Query: db.processedVehicle.countDocuments({price: {$gt: 40000, $lt: 70000}})
   Result : 558

10. Calculate the no of vehicles with fuel type only hybrid (not PHEV Hybrid)
    Query: db.processedVehicle.countDocuments({fuel: "Hybrid"})
    Result: 137

11. Find the no of fuel types.
    Query: db.processedVehicle.distinct("fuel").length
    Result: 7

12. Retrieve the details (make,model,type,year,price,mileage) of jeeps which have white as exterior_colour, Global black as interior color.
    Query: db.processedVehicle.countDocuments(
    {make: "Jeep",exterior_color: “White”,interior_color: "Global Black"},
     {make: 1, model: 1, type: 1, year: 1, price: 1, mileage: 1})
    Result: 1