Vroom - Analytics

The company purchases cars from private clients and sells them either to merchants (C2B) or to other private clients (C2C).

Material:

You will find a dataset including 3 different tables:

- **car_sales** :
  - ***This table includes the sales information of the car.***
  - *Merchant_id* : ID of the merchant who purchased the car.
  - *Country* : country of the merchant who purchased the car.
  - *Car_id* : ID of the car.
  - *Selling_date* : date at which the merchant signed the sell contract.
  - *Payment_date* : date at which the merchant sent the payment for the car.
  - *Sell_price* : self explanatory, in EUR.
  - *Transport_to_merchant* (Boolean): if 1 then the merchant ordered a transport of the car from one of Vroom compounds to his place, if 0 then he is picking up the car himself from the compound.
  - *Transport_date* : date at which the merchant receives the car (pick up or transport).
- **Car_details :**
  - ***This table gives you a list of car characteristics.***
  - Some are booleans : has_tuning, has_airbags, has_alarm_system, etc.
  - Some are references : ac_type, fuel_type, gear_type, etc.
- **Car_claims :**
  - ***This table is referring to the aftersale claims merchants can have on the car.***
  - *Full_refunds* : in case the claim is leading to the full refund of a car, then the sell price of the car is fully refunded to the merchant. The car comes back to Vroom.
  - *Partial_refunds*: the car stays with the merchant but he receives a voucher to cover the claim.
  - *Voucher_amount* : in EUR, the amount of the partial refund.
  - *Claim_status* : tells you if the claim was processed or refused. It is worth noting that for partial refunds, even if the claim was rejected, a voucher amount may have been agreed along the process but ultimately not paid out to the merchant.

Over the past year, the claims department has seen its expenses going through the roof.
To mitigate the problem, the Director of Operations wants to understand how he can reduce his budget by investigating cars that are causing refunds.

1) Based on the data you find attached, please perform an analysis that will help us understand what types of cars cause the most claims and thus require a refund more frequently.
2) Based on your findings, please provide some practical business recommendations in a structured way to your stakeholder.

3) Predict the dataset based on the feature applied to the scenario or the use case e.g. types of cars might cause the most claims and thus require a refund more frequently.