Data -Mining

1. Select a domain-appropriate question that you want data mining to answer

* “what factors enable rapid sale of a vehiclevs.it taking a long time to sell?”
  + Classification
* “what factors determine the price of a used vehicle”
  + Clustering
* **“what make and/or model sell most in a given geography?”**
  + Classification
* What causes a dealer to sell the most amount of cars
  + Classification
* What cars are most popular?
  + Classification

2. Select a technique or techniques that will be appropriate to the question you

are investigating

* Clustering not as relevant for us, go for classification
  + Classification: 1.) putting things into classes 2.) use those classes to predict which classes new items will belong in
* Decision Tree (decision algorithm)
  + **Best method** due to a lack of numbers in our schema
* Bayesian Classifiers (probabilities)
* Support Vector Classifiers (lines/points/distance)

3. Implement said technique

Efficiently to build a data model

* Examples in python for decision trees

4. Determine the validity of your model

5. Report the results of your investigation

Note: use ½ of the entries to build the model and ½ half to validate