In this capstone we were given a Kaggle database with data containing information about used cars. This data included things like:

* 'region'
* 'price'
* 'year'
* 'manufacturer'
* 'model'
* 'condition’
* 'cylinders'
* 'fuel'
* 'odometer'
* 'title\_status'
* 'transmission'
* ‘VIN',
* 'drive'
* 'size'
* 'type’
* 'paint\_color'
* 'state'

My goal was to use different prediction models to determine which of the above categories had the most impact on the price of the vehicle. To begin I used a linear regression model to predict the price of the vehicles. With this model I saw that the top 10 features that impacted the price of a used car were:

* Price per Mile
* Odometer (mileage on the car)
* Transmission is manual
* Manufactures are:
  + Volvo
  + Mercedes Benz
  + Datsun
  + Pontiac
  + Toyota
  + Mercury
  + Harley Davidson

I then used a more advanced model called random forest. With this model we saw the top 10 features that were most important were:

* Odometer
* Price Per Mile
* Fuel type is gas
* Year (car Age)
* Manufactures are:
  + Totyota
  + Mercedes benz
  + Jeep
  + Volvo
  + Nissan

Based off these results I can conclude that the features impacting the price of the used car the most is odometer readings and price per mile.