Problem Statement:

Carry out an exploratory data analysis on the provided Used Car Dataset. Identify how to optimally build a model to predict used car price. Tune the model and explain your process and the results. Try using various techniques to clean your data, perform feature engineering and tune your model to get optimal results.

Expected Output:

Please provide a program/script that can take a tab delimited file with these given features and output a file with the index and the predicted price.

Also attach a file with your output for the Test values (see attached as “Test.txt”) in an “Index<tab-delimited>Price” format.

Training data :

[https://drive.google.com/file/](https://drive.google.com/file/d/147OEDk7G3f2iJV1UTstu90O99h3kY4ml/view?usp=sharing)

Test Data :

[https://drive.google.com/file/](https://drive.google.com/file/d/1JPCZMm5JdoozQHRxmLvC4vXKJsI7cmle/view?usp=sharing)

We would need four files, one is the script (in Python/R) that can be used on a input file with the given features and returns the predicted price for each data point, second is the script you used to train your model (if separate), third would be the output file for the test file and most important would be a document detailing your journey with this problem with a detailed explanation about

1. your understanding of this data,
2. the features/ feature transformations you used,
3. your choice of the technique,
4. the metrics you use to evaluate the model
5. the results/ performance of your model/model system and
6. any visualizations and scripts you have used to arrive at the result besides the two scripts expected above.

Kindly acknowledge receipt of the problem and please let us know if you have any questions/concerns.