1:Assignments should be submitted electronically, as MS Word or PDF document containing text  
and embedded graphics. Label each plot, chart or table, add detailed title and reference it in your text (i.e. “Figure 1 shows … “).  
2. Submit code if requested.  
3. The deadline is firm. Late submissions are not allowed

1(30%)Use the Vehicle dataset CAR DETAILS FROM CAR DEKHO.csv to build a linear regression model  
that predicts the “selling\_price” for each car. Split the data to 70% train and 30% test and report  
your results.  
a. Which variables did you use? Try to select a subset of the variables and check whether you  
get better results.  
b. Which evaluation measures did you use? Why? Describe your results and their meaning.  
c. Predict the selling price for the new cars in stock presented in Table 1.  
2(40%) Build both logistic regression model and a decision tree model to predict whether the car  
brand be sold above their brand’s mode price or not.  
a. Describe your experiment, the results and their meaning.  
b. Which model perfumed better on your test set?  
c. Use your model to predict whether the new cars in stock presented in Table 1 be sold above  
their brand’s mode price.  
d. Repeat your experiment with 5-folds cross validation and report your results. Are they  
different from your 70%-30% split?  
\* Remember: This is binary classification, not regression like above. You need to calculate the mode  
price for each brand and then create the target variable based on the selling\_price > mode\_price.  
3. (40%) Build a decision tree model to predict whether the car be sold at their brand’s 25th percentile,  
50th percentile, 75th percentile or top percentile.  
a. Where all variables used by the tree?  
b. Show your tree.  
c. Describe your results or your model.  
d. Use your model to predict at which percentile the new cars in stock presented in Table 1 be  
sold.  
\* You may use df.selling\_price.quantile([.25, .5, .75] to calculate the percentile values and then pd.cut() to get the corresponding categorical value.  
**Table 1: New Cars in Stock**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| brand | year | km\_driven | fuel | seller\_type | transmission | owner |
| Honda | 2017 | 40000 | Petrol | Dealer | Automatic | Second Owner |
| Hyundai Creta | 2016 | 6000 | Petrol | Dealer | Manual | Third Owner |
| Mercedes-Benz | 2015 | 9500 | Diesel | Dealer | Manual | First Owner |