Exercise: Categorical Feature Encoding

Using the data in ***car-data-category-encoding.xlsx*** do the following:

1. Read the data into a dataframe
2. Create Python functions (one for each of the encoding methods) to accept as inputs a dataframe and a target column, and return a new dataframe with all columns encoded, except the target column, using the following methods: (Can you create just one function to handle all the below methods, instead of creating separate one for each?)

   (Which Python module implements all the below methods? Find out)

   a. Integer or Ordinal encoding
   b. One-hot encoding
   c. Binary encoding
   d. Target encoding
   e. Frequency encoding
3. Using each of the above encoded dataframes carry out the following steps:
   a. Split the available data into train / test sets
   b. Create a classification model (e.g. Random Forest Classification) to predict the "**class**" using all the other columns.
   c. Derive the test and train metrics for each class: Precision, recall, F1-score, overall accuracy, and save these metrics in a dataframe.
   d. Create and print the Confusion Matrix.
4. Compare the metrics created in the above steps and identify which encoding method gives the best results. Try and understand why.
5. Carry out Step '3' using multiple classification methods (Logistic Regression, Support Vector Classification, RF classification, etc.), and identify which combination, of encoding and classification method, gives the best results!

oooOOOooo