Machine Learning Summer Group 2020 : Final Project
Used Cars Analysis

Problem Statement:

You've been hired as a data scientist at a used car company to help them analyze their sales and improve their business. Your main aim is to predict the price of their cars. You are given a CSV file containing several details for each vehicle sold.

Your first task is to analyze the data. Here you are supposed to ask relevant questions and write down the code to find those answers (mostly in pandas). Example questions include 'Which companies sell the most number of cars ?', 'Do petrol cars have lesser mileage ?', etc. (answer these two questions too). Write down the question in markdown script and the code after that.

Your next task is to visualize data. Try to find some patterns in data and make some more meaningful inferences.
Note: Please put your conclusions and inferences in markdown script.

Another critical preprocessing step is feature engineering. Remember to convert categorical data into one hot encoding, fill in any missing data, dropping unwanted features etc. You will then find the relationship between the price and other features. Try finding out what might be the most crucial factor that affects car price and state why.
Here is a reference link for Feature Engineering:
https://towardsdatascience.com/feature-engineering-for-machine-learning-3a5e293a5114

You will then try to predict the car price, first with linear regression, then using a neural network. Remember to solve any underfitting/overfitting issues to make it generalize well.

Your company also wants to find out different types of customers to improve their advertising. You are supposed to cluster the data into **three different clusters** and state the characteristics of each cluster.

General Instructions:
- Complete the entire project in a single Jupyter/Colab Notebook.
- Use Markdown and comments whenever necessary to make your project presentable.
- You are allowed to use Numpy, Pandas, Matplotlib, Seaborn, Scikit Learn and Keras.