INSTRUCTIONS FOR RUNNING PYTHON CODE

Here I have divided python file into 2 parts as the code is taking long time to execute in a single ipynb file. Here are the guidelines regarding the two parts:

Part1:

It consists of all the pre-processing techniques that are to be done on the data set such as outlier analysis, missing value analysis, feature extraction and the distributions plots of various characteristics on the train, test datasets also the datatypes of test and train attribution are understood and conclusion made via seaborn library (sns) count plot clearly states that the given data set is imbalanced distribution of various characteristics along both column and row are plotted.

Part2:

It consists of modelling various ml algorithms and their cross-validation performance measured in terms of accuracy and AUC scores. Moreover, smote analysis is done for all the models and evaluation of all the classification reports is done. To quantify the need for better accuracy lightGBm is introduced and results are submitted with lightGBm+stratified kfold+smote,

Here I combined both the files and made a file **combined.ipynb** and attaching it with all the other files.

The output files are attached as well and named as submission1 and submission