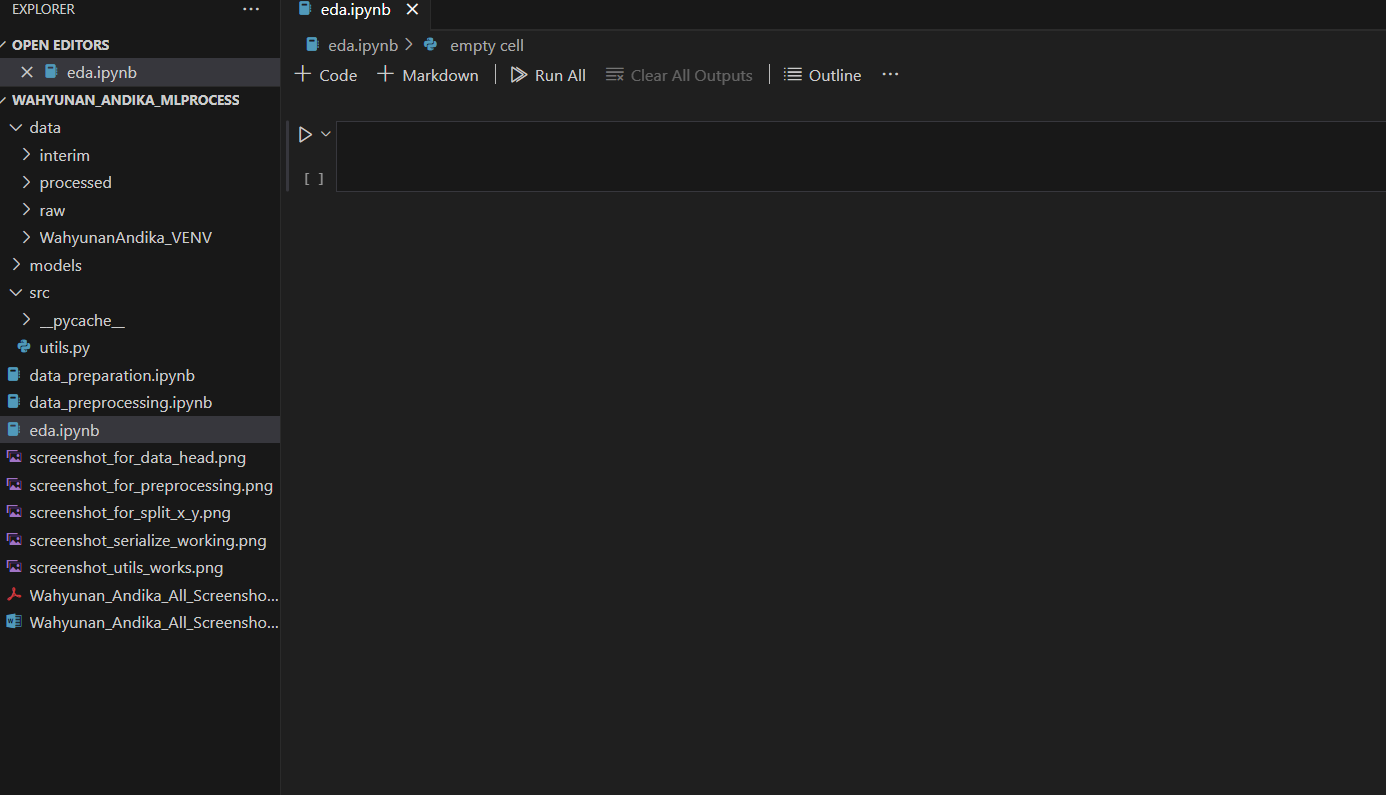
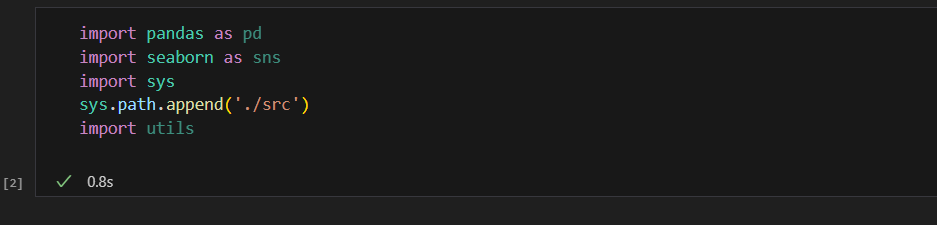
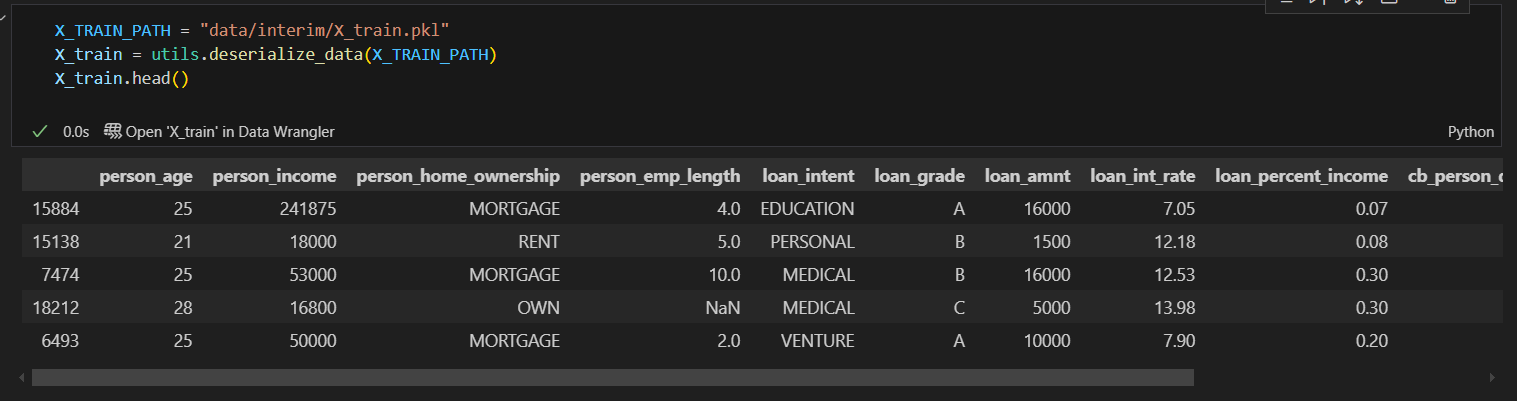
Buatlah satu file bernama eda.ipynb di root folder project Anda



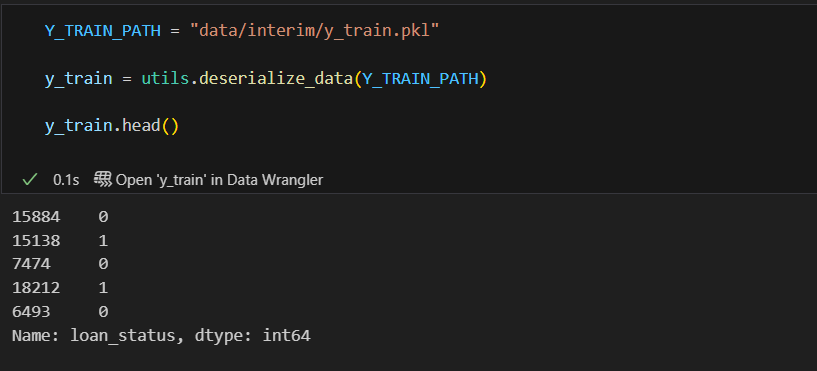
Import library yang dibutuhkan Buka eda.ipynb dan import library utils dari folder src, pandas yang memiliki alias pd dan seaborn yang memiliki alias sns



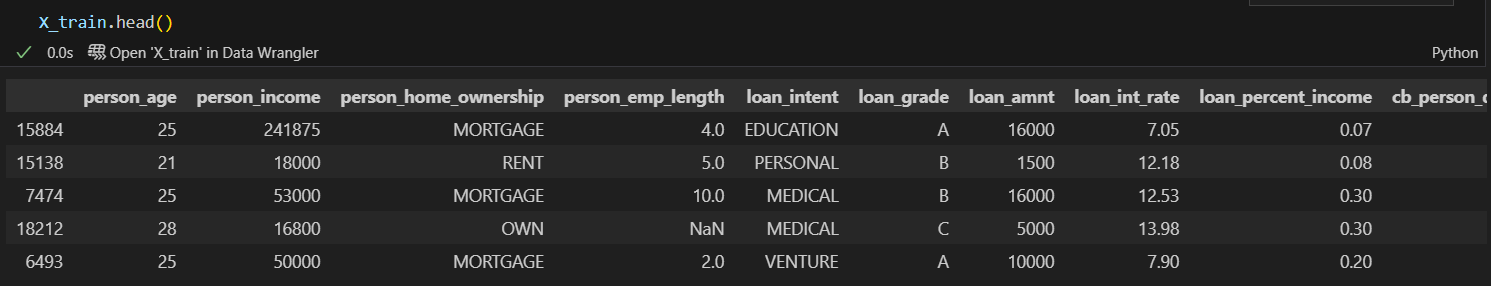
Muat X\_train data



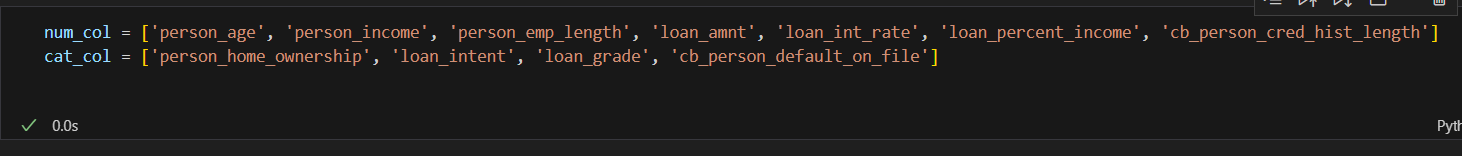
Muat Y\_train data



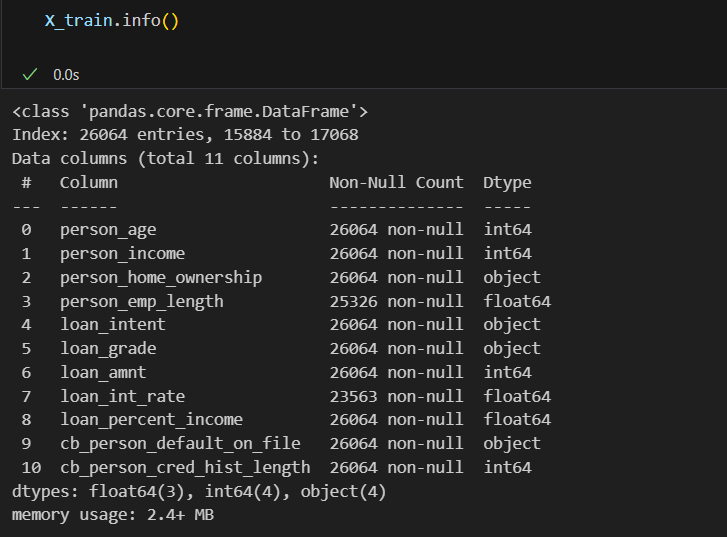
X\_train.head()



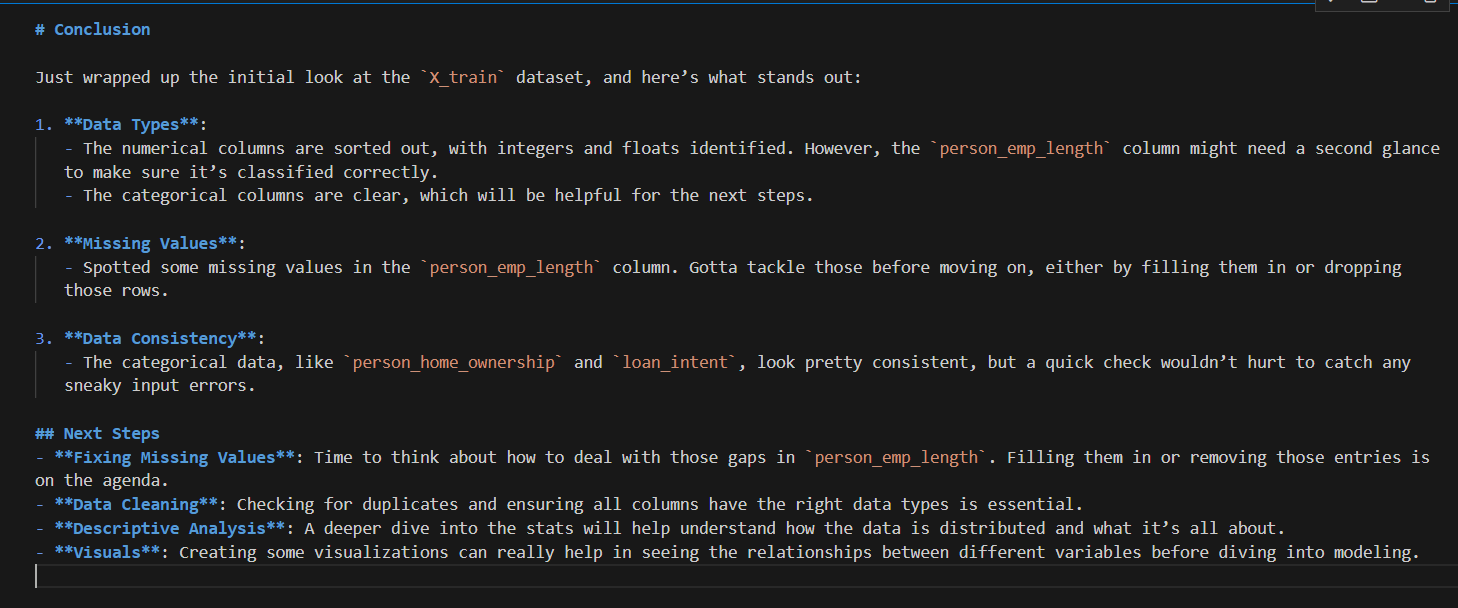
Split numerical and categorical



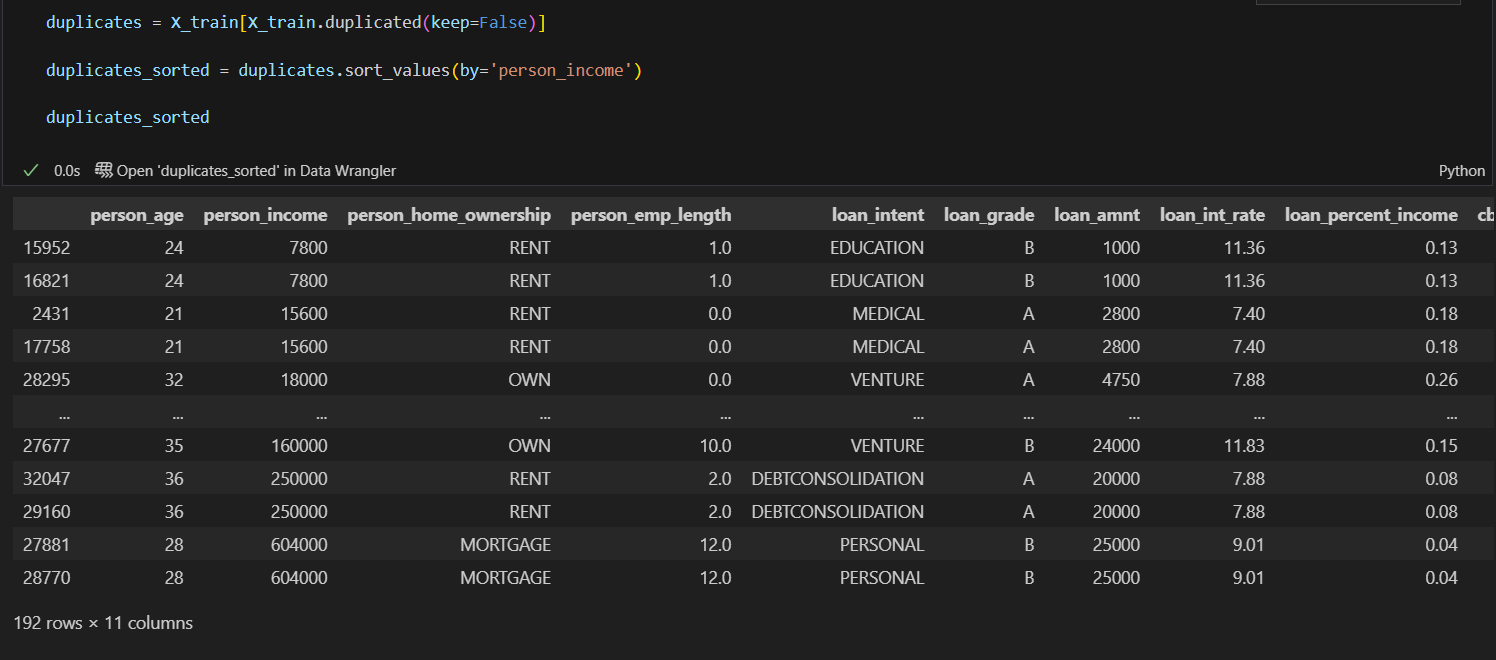
Info



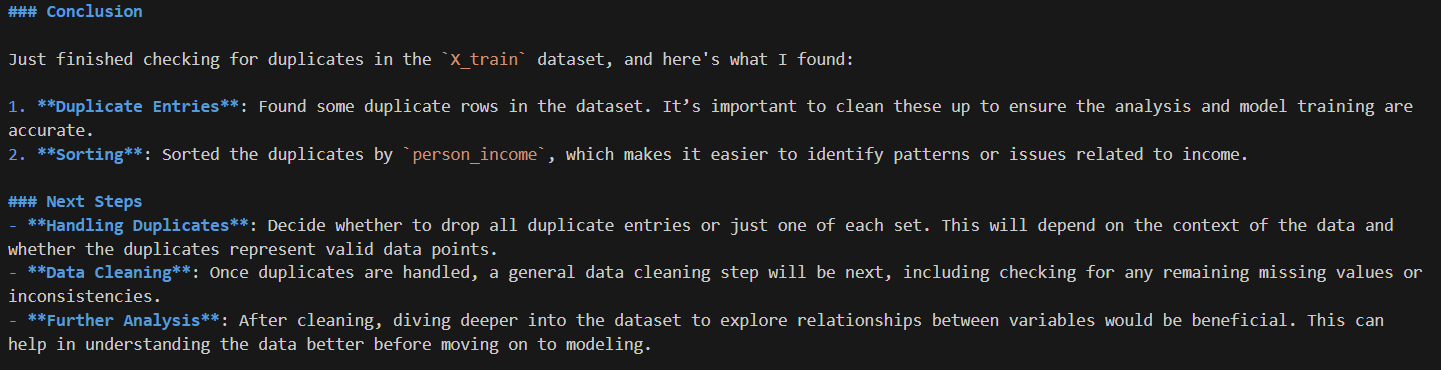
Conclusion



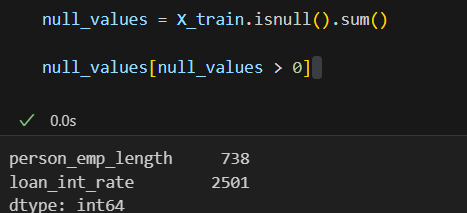
Urutkan data berdasarkan data pada kolom person income



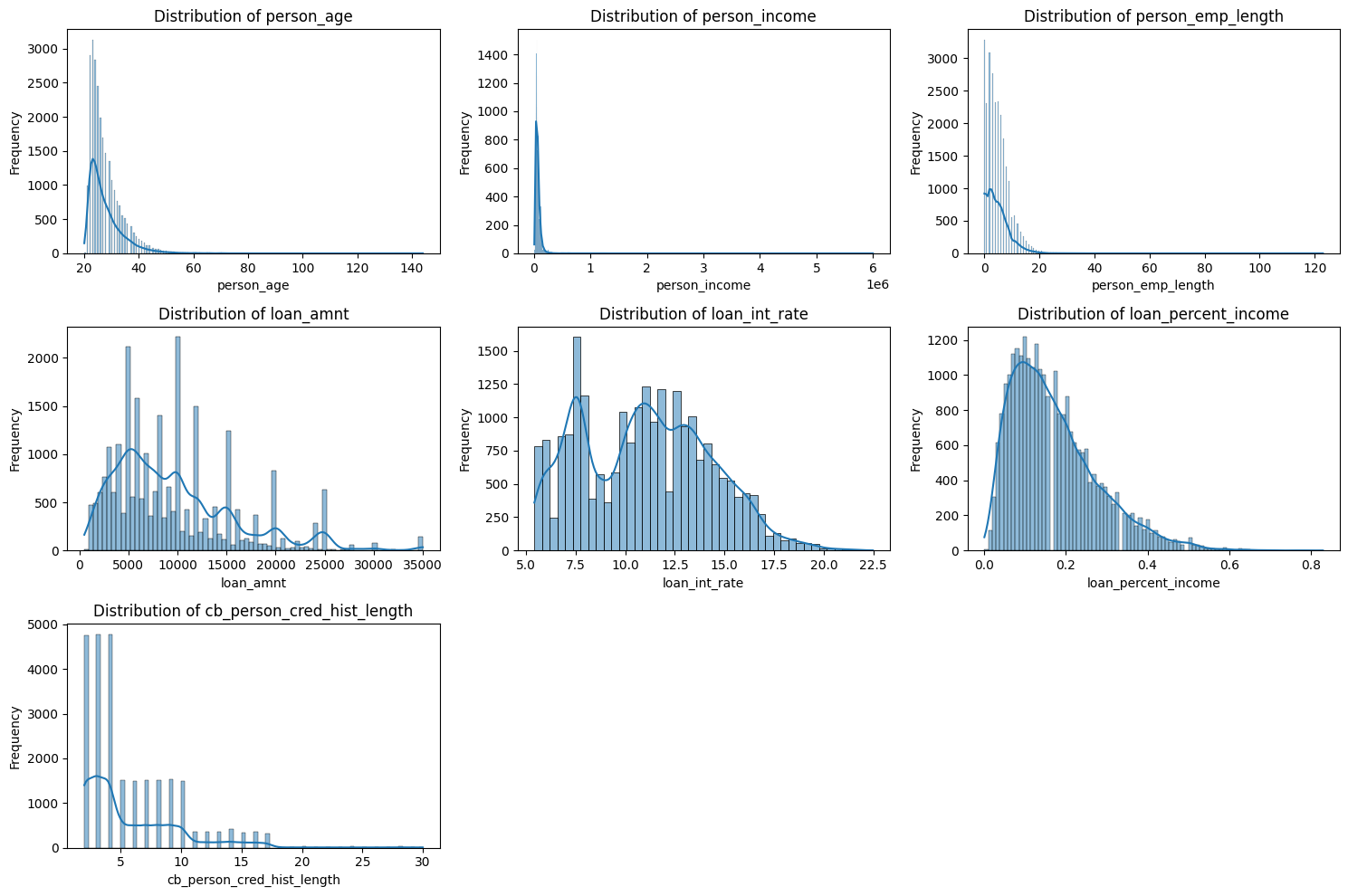
Buatlah kesimpulan dari yang telah dilakukan serta tindakan kedepannya pada sel jupyter notebook selanjutnya



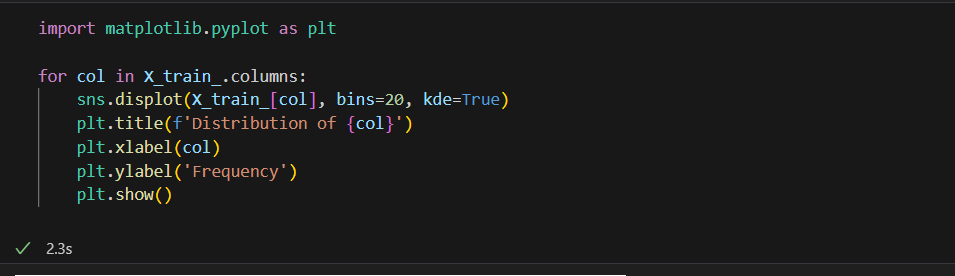
Null value



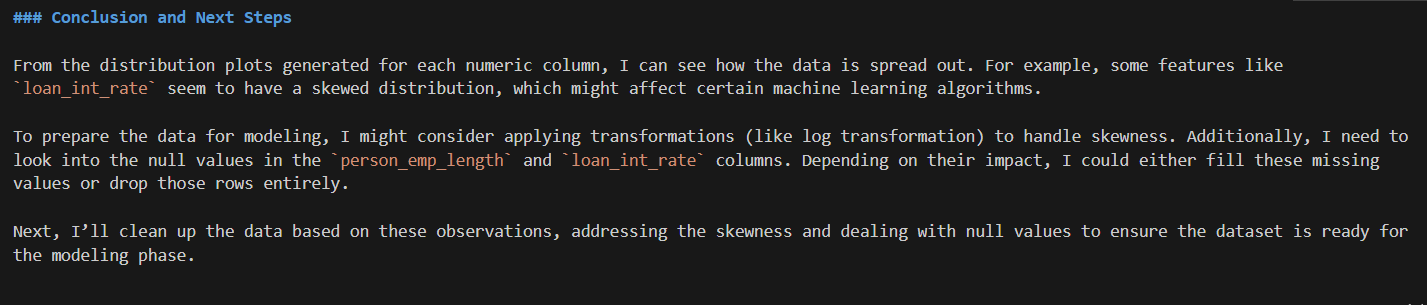
Seleksi hanya kolom numerik pada variabel X\_train dan simpan pada variabel X\_train\_ menggunakan daftar nama kolom yang telah dibuat sebelumnya (variabel num\_col)



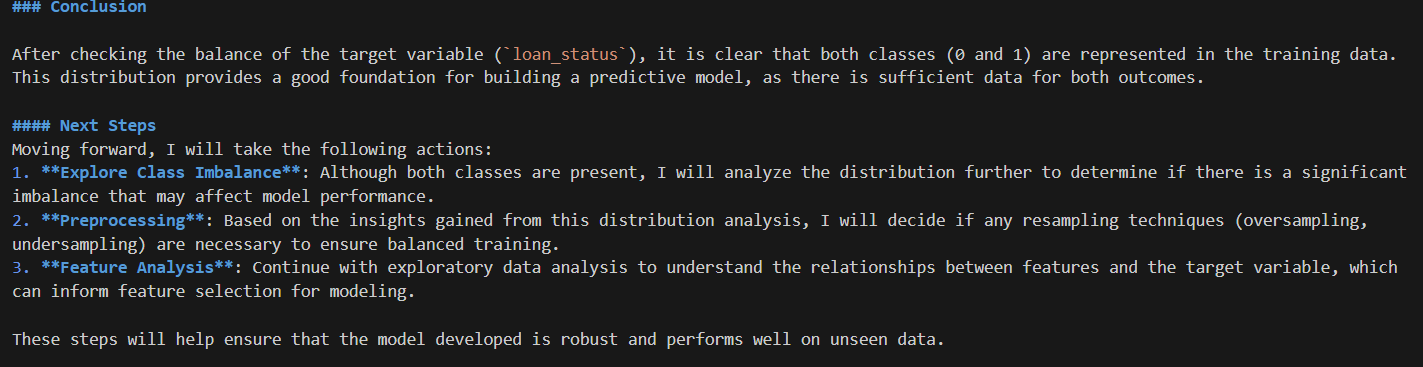
Pada tiap iterasi perulangan, panggil fungsi displot() dari library seaborn



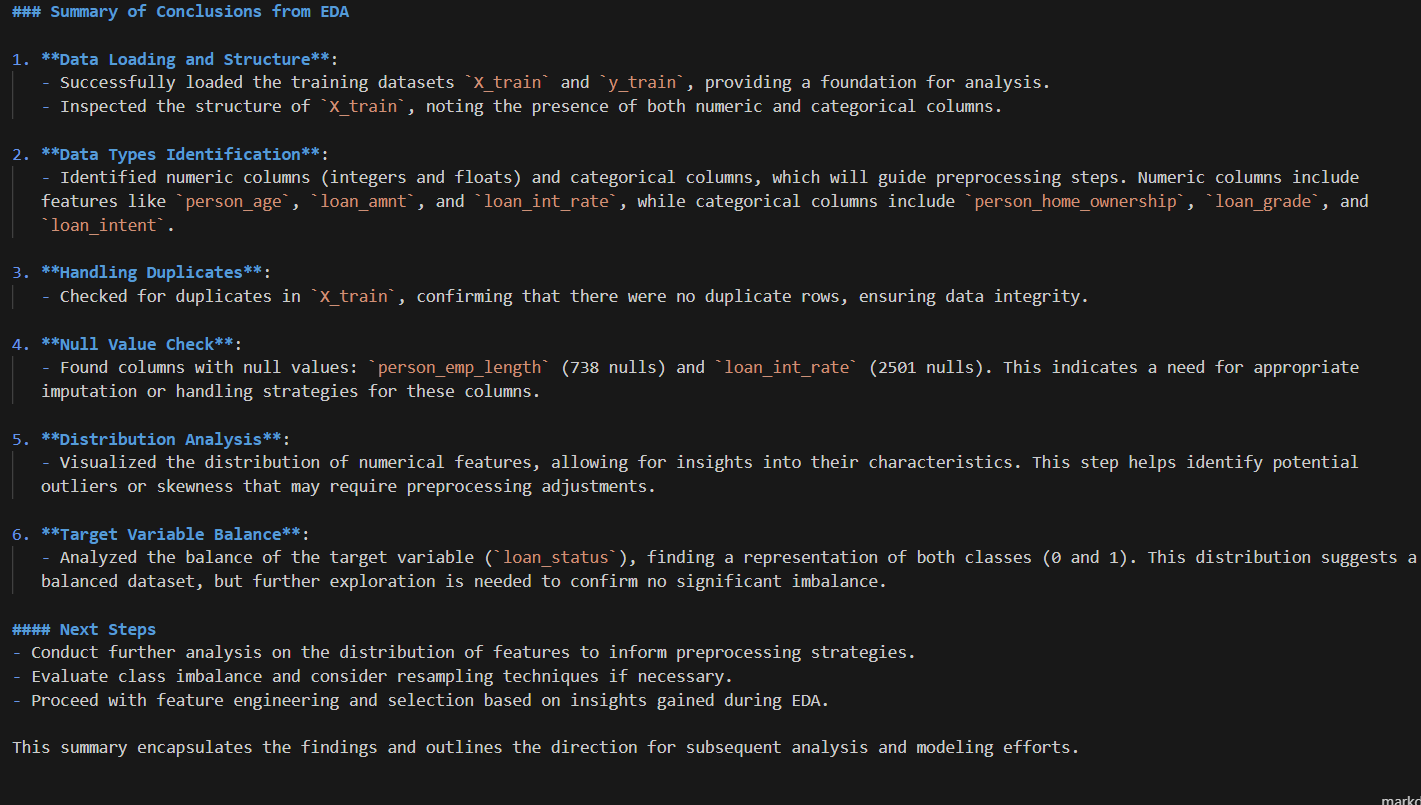
Conclusion



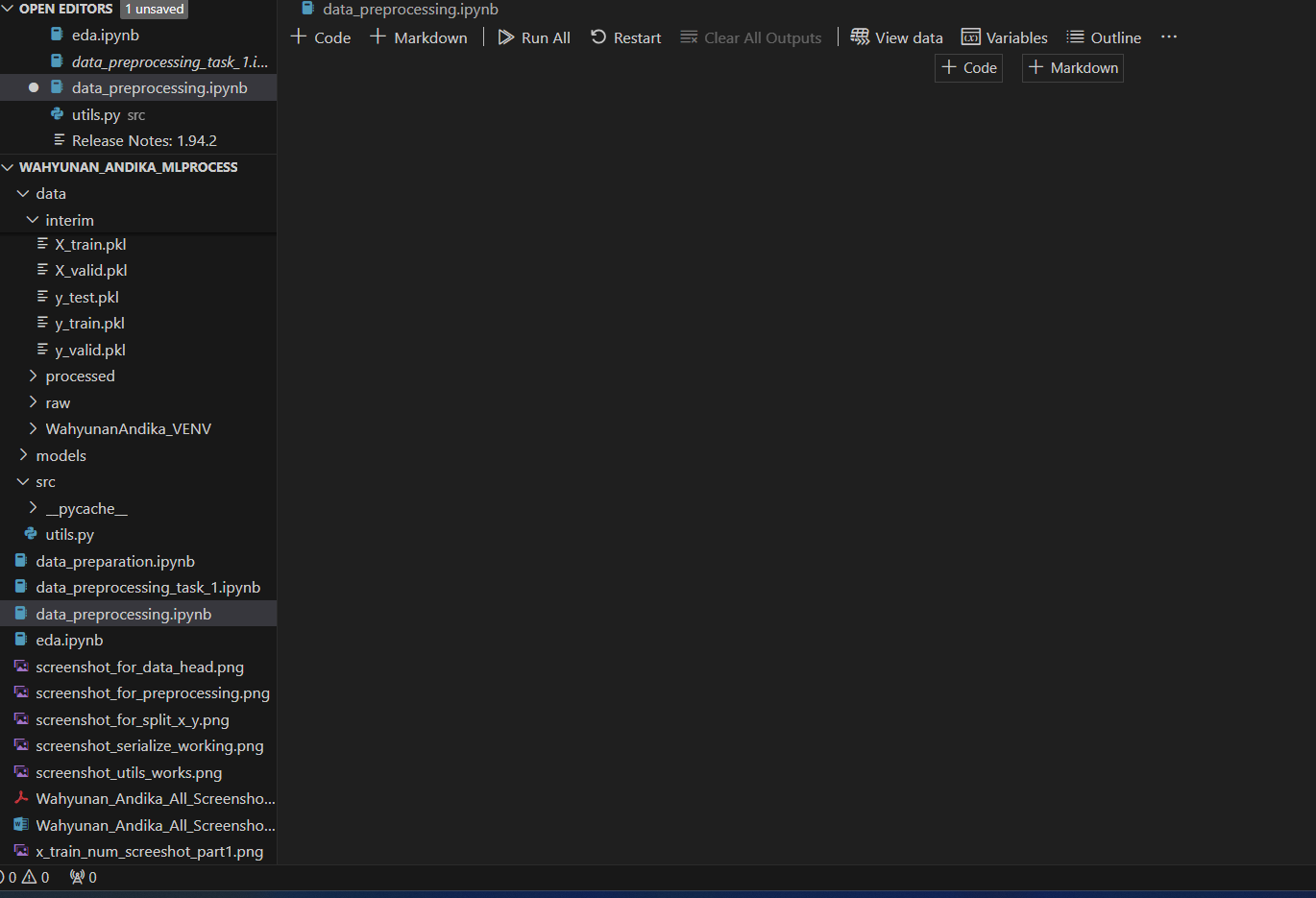
Conlusion y\_target



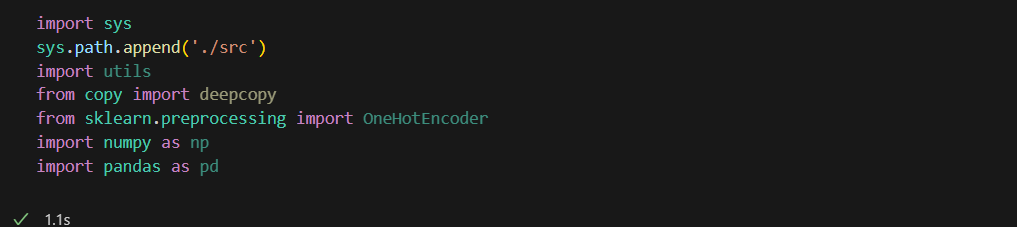
Summary eda



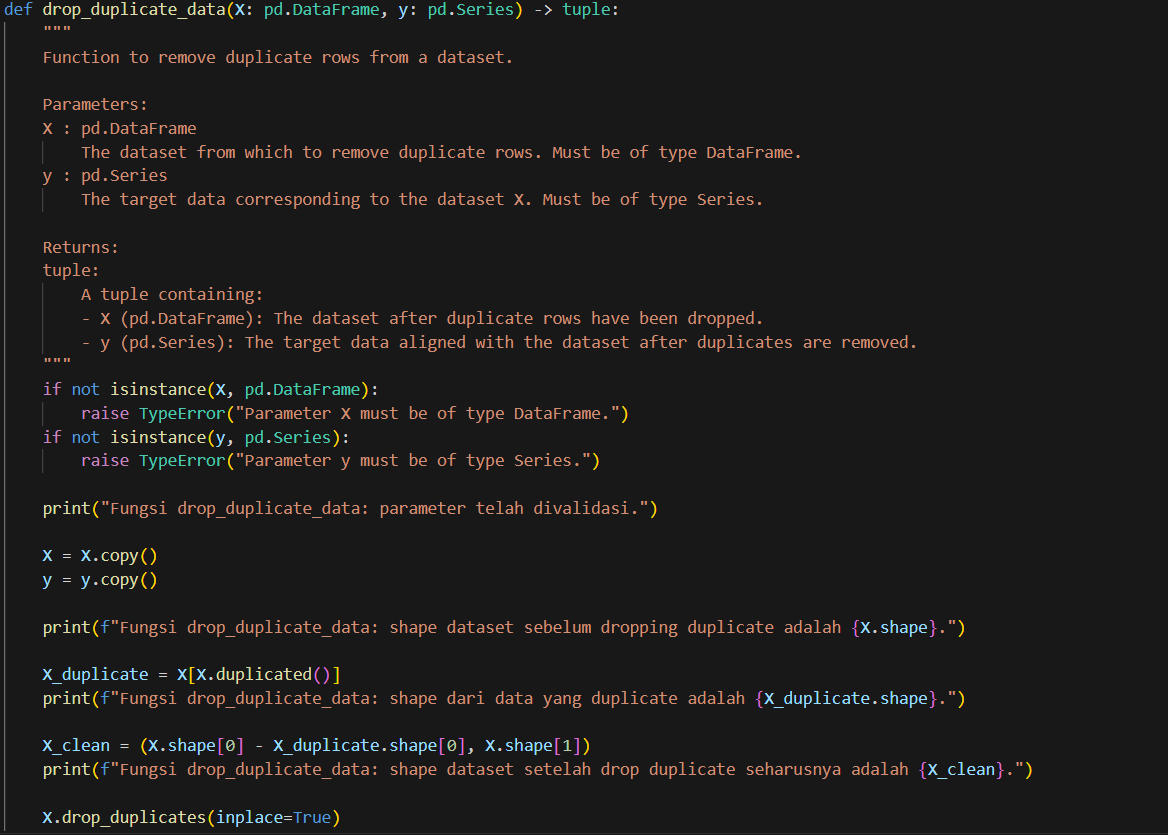
Create data\_preprocessing

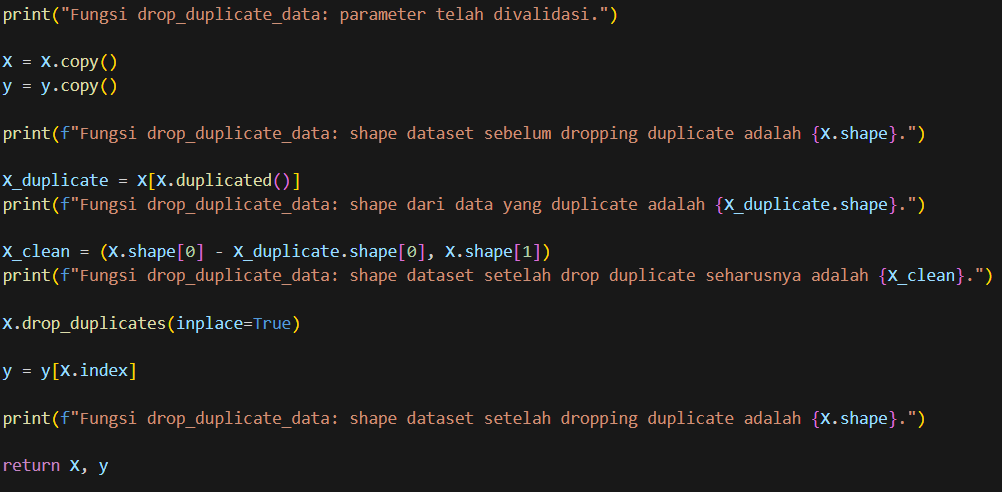


Import libs

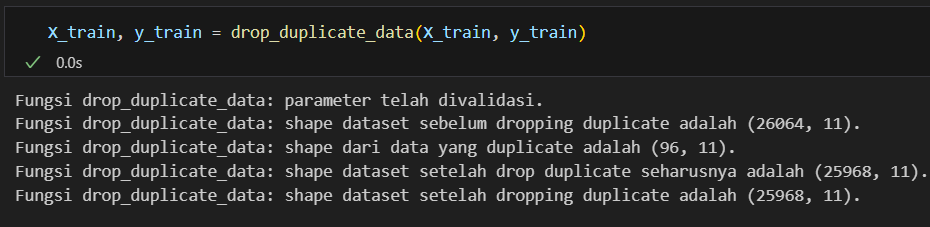


Drop\_duplicate\_data

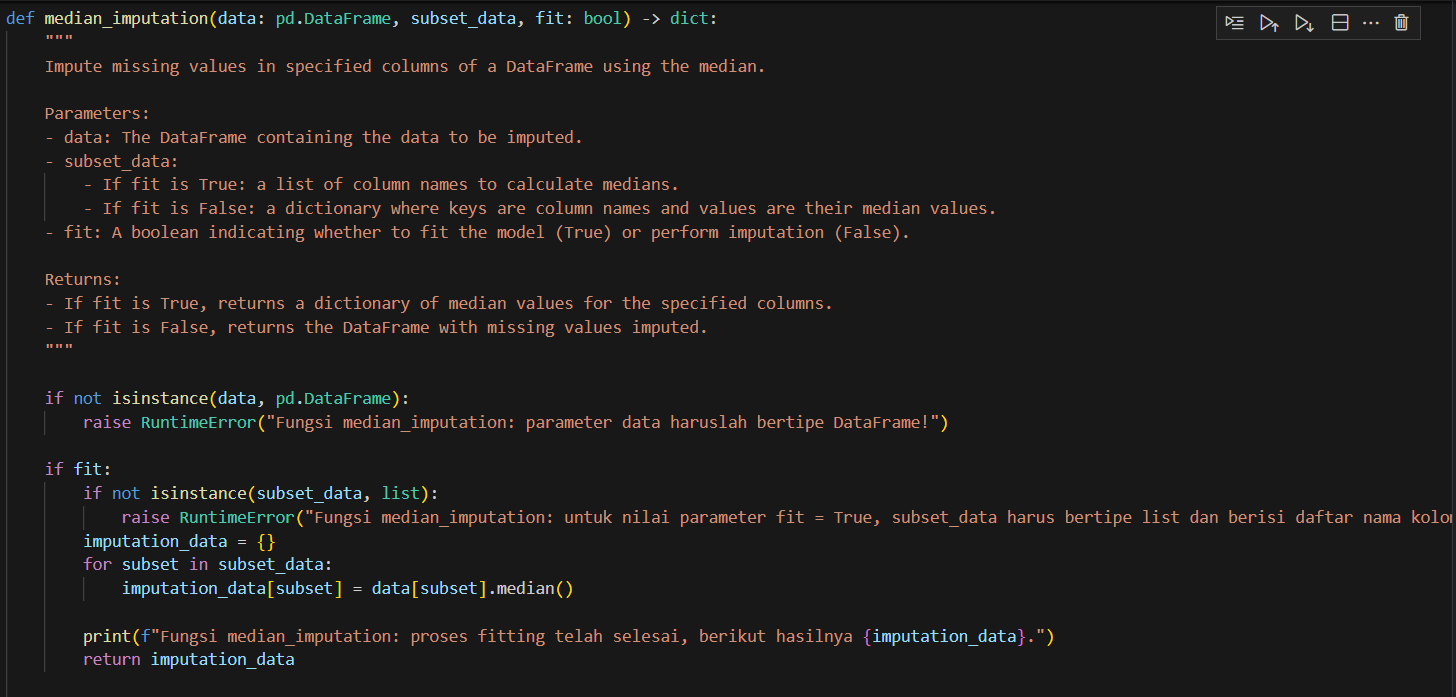


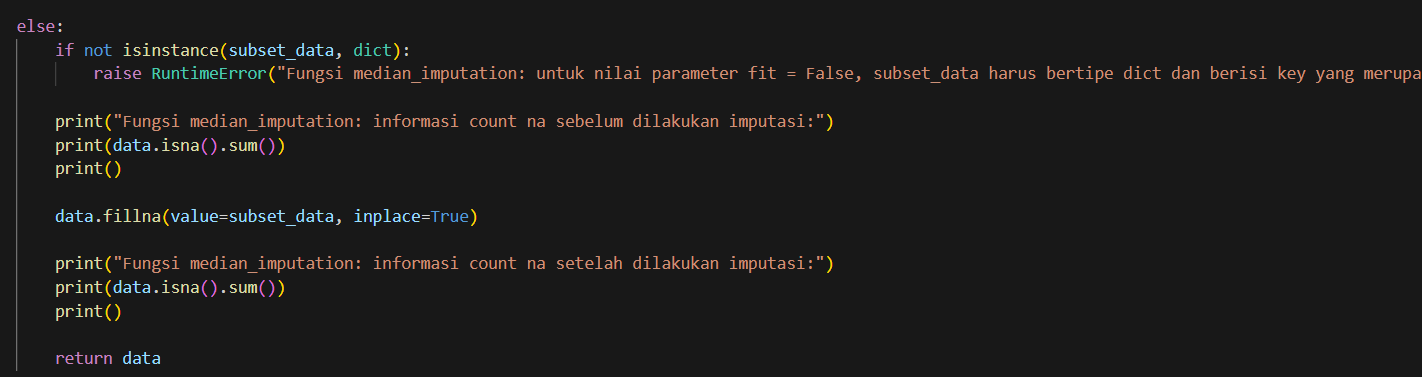


Executing drop\_duplicates

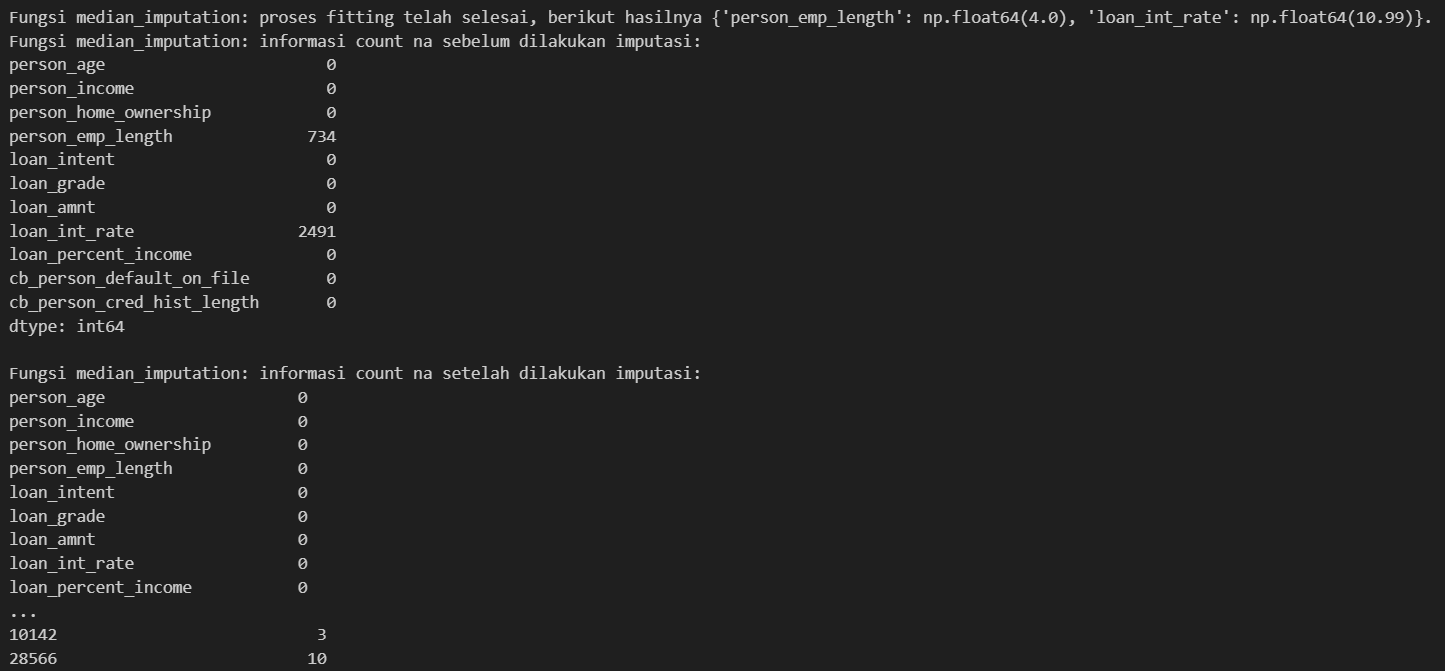


Function median

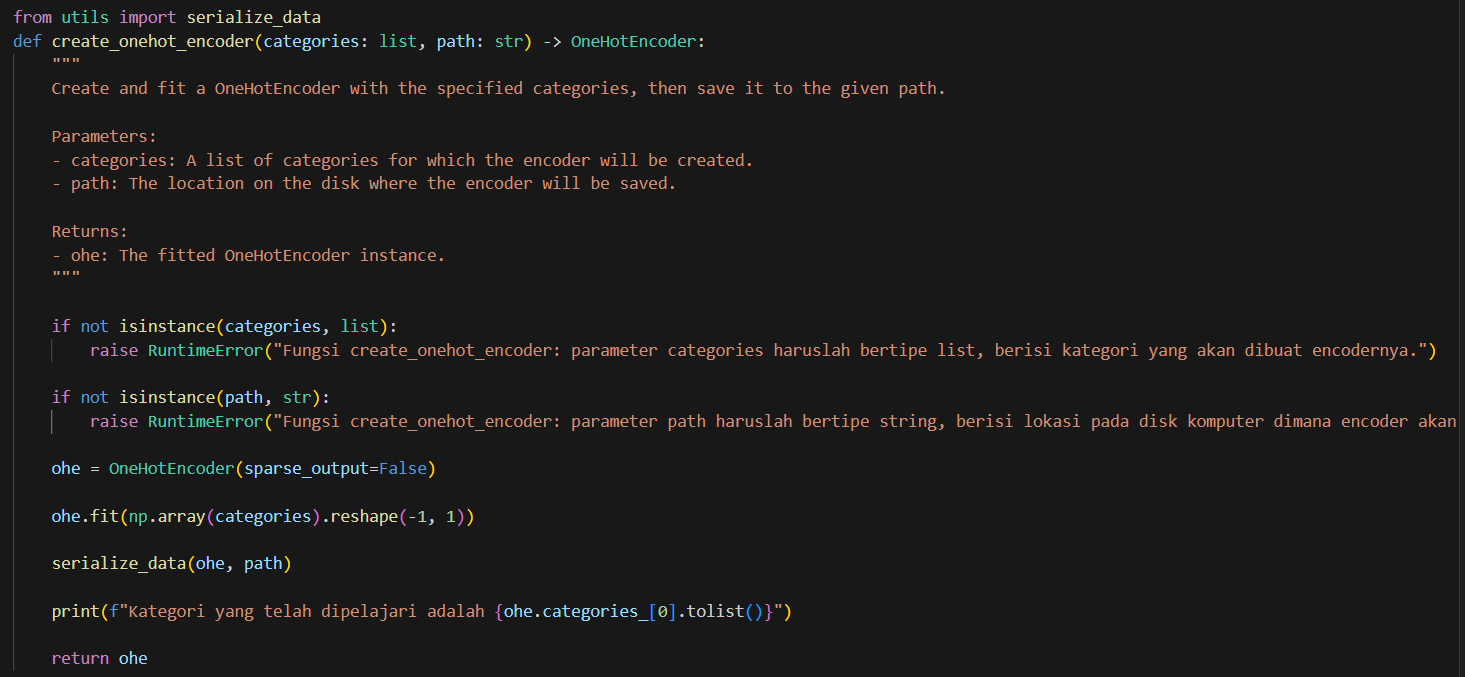




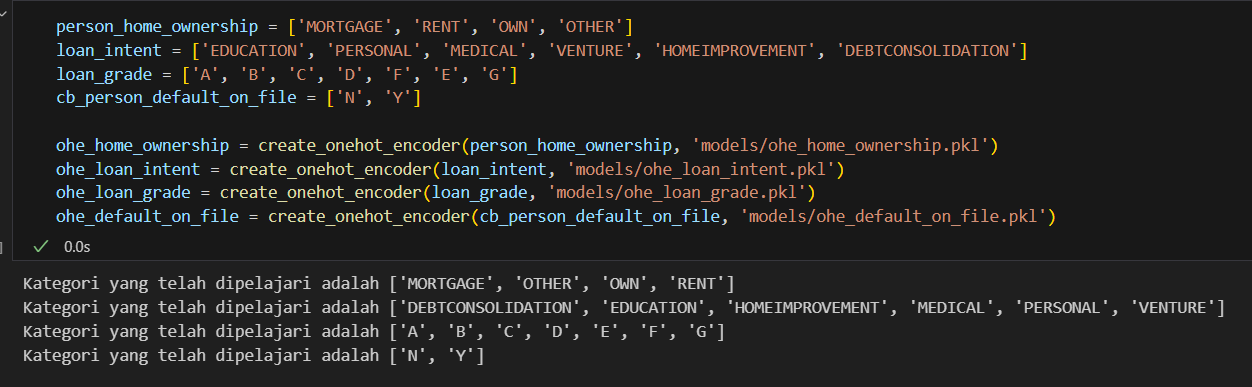
Execute median



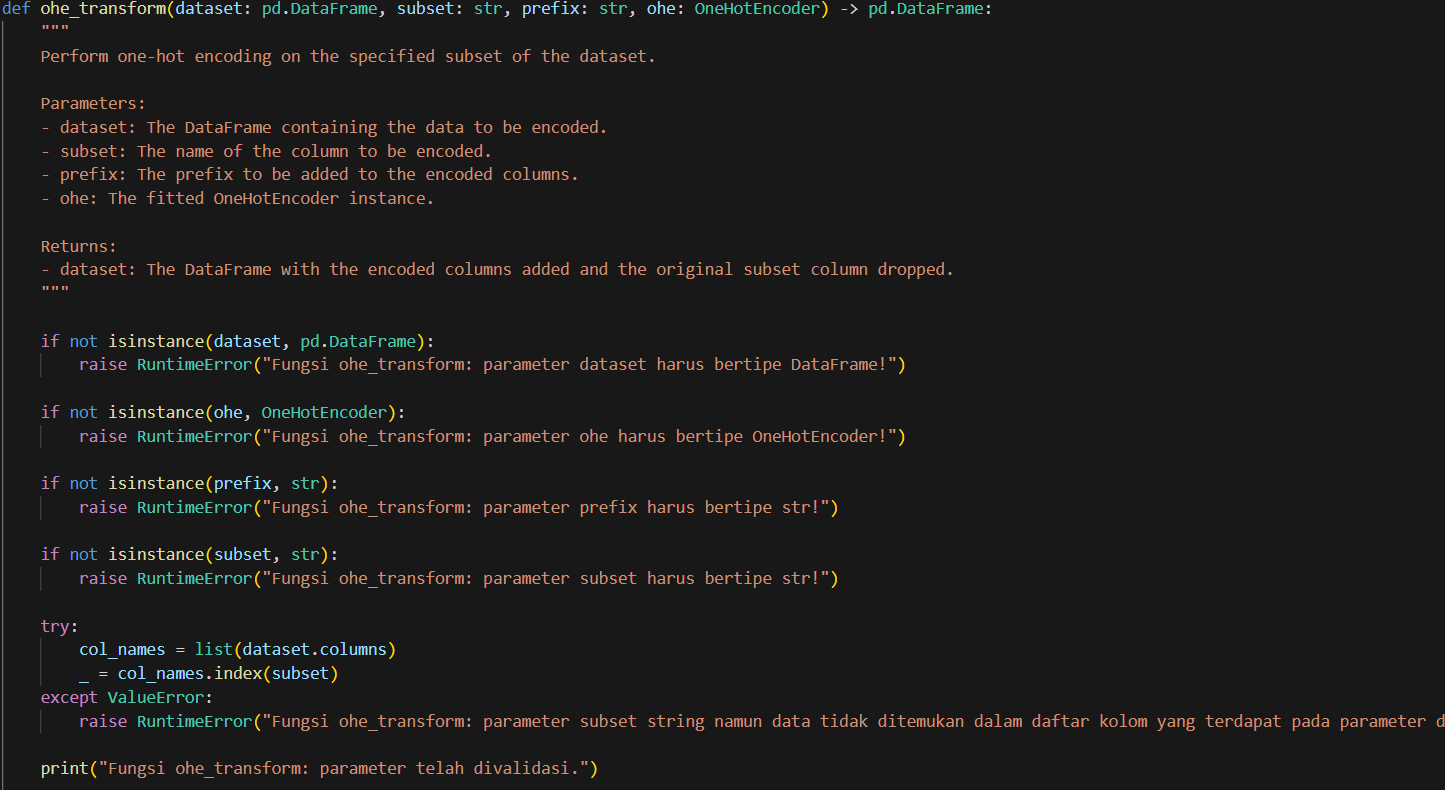
Def OHE

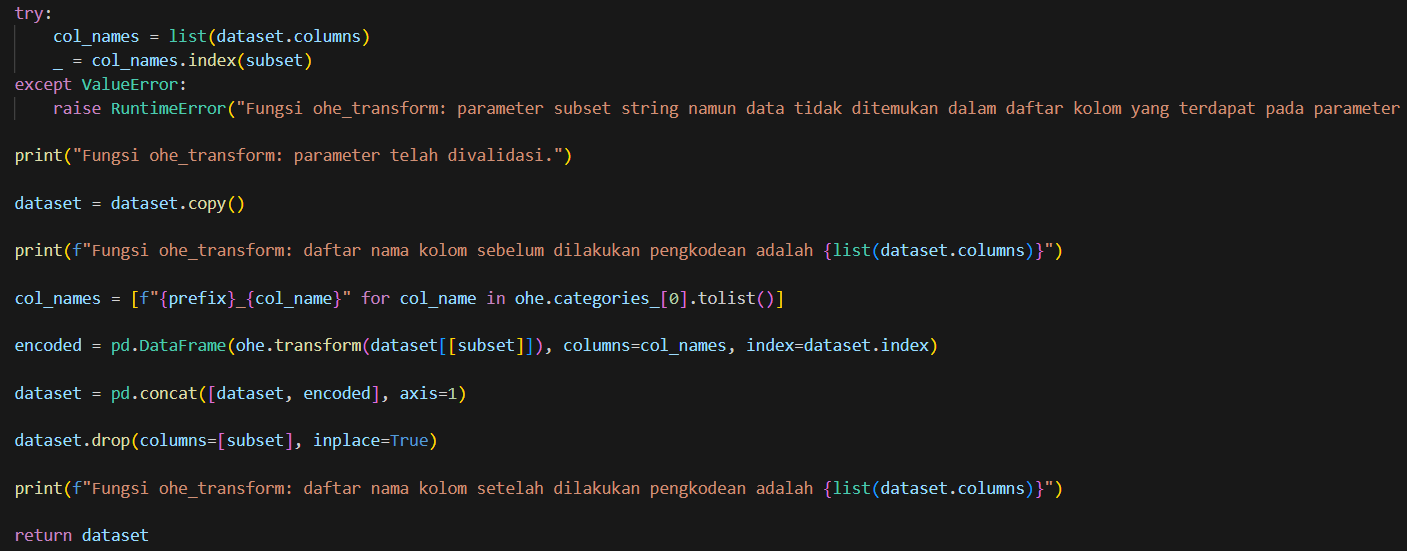


Execute ohe

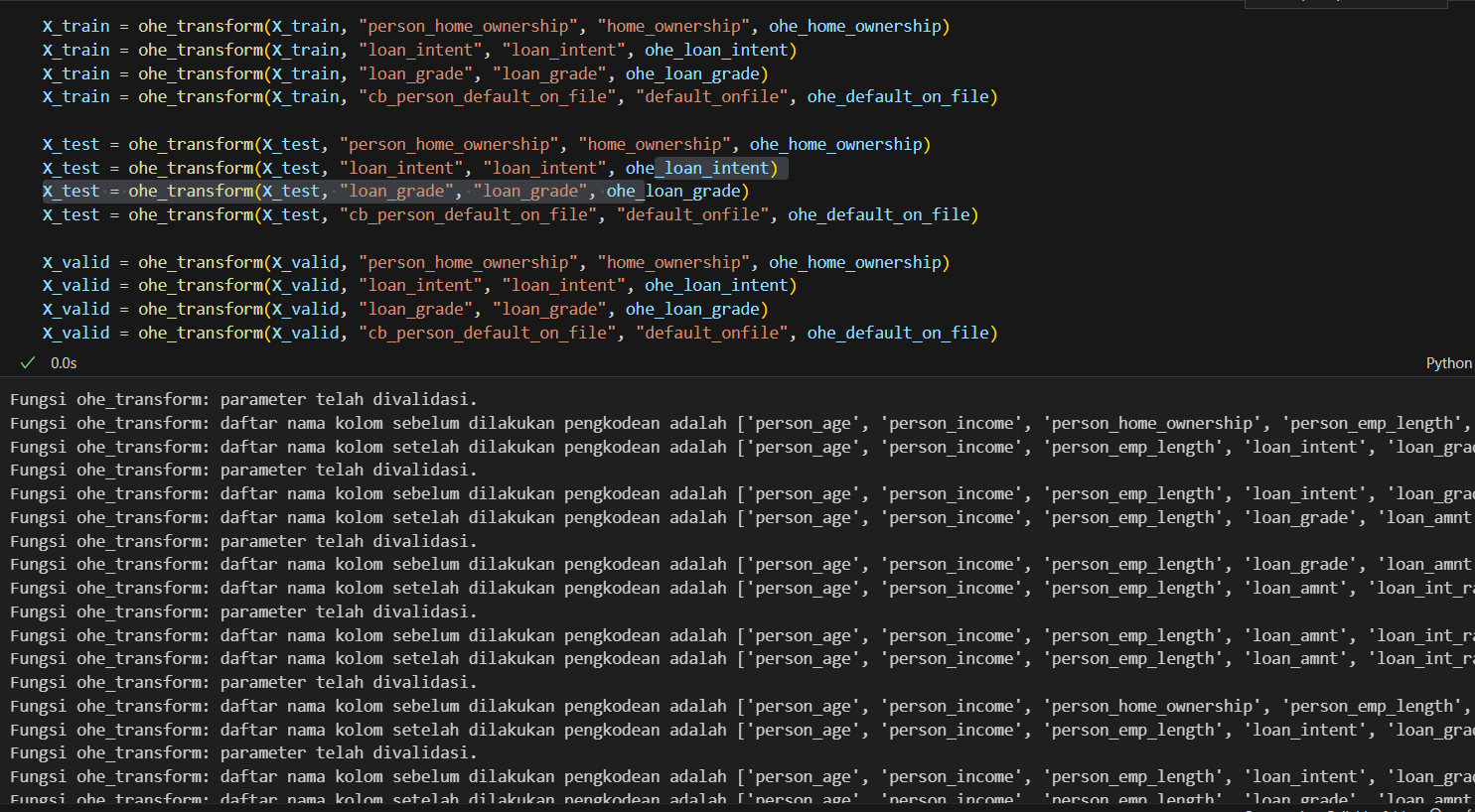


Def ohe\_transform

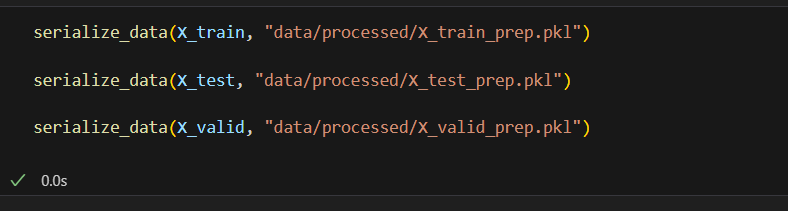




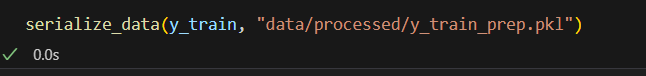
Execute ohe transform



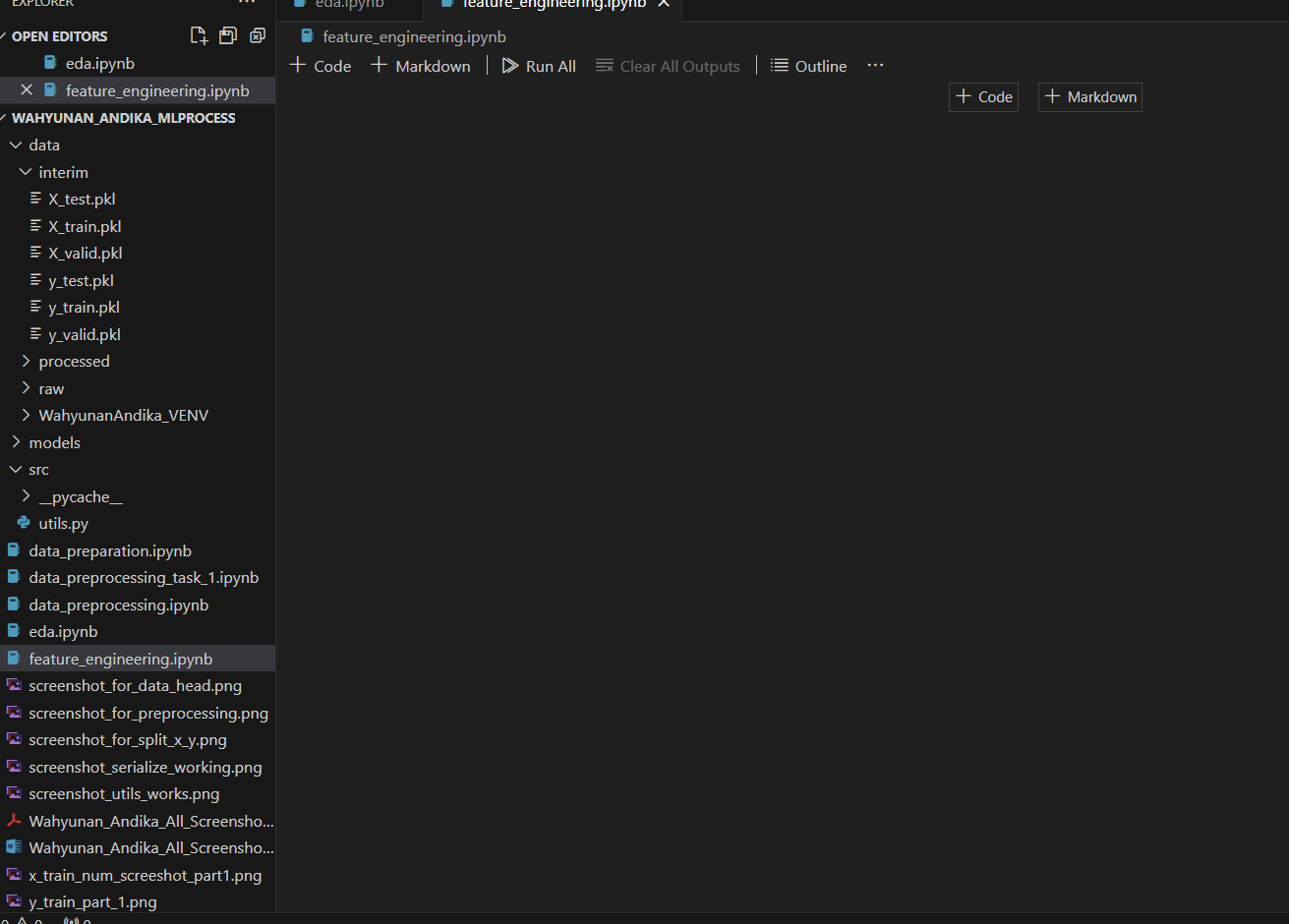
Serialize x data



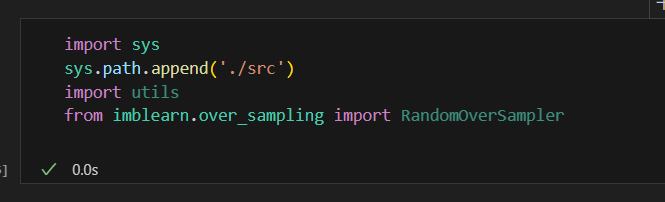
Serialize y data



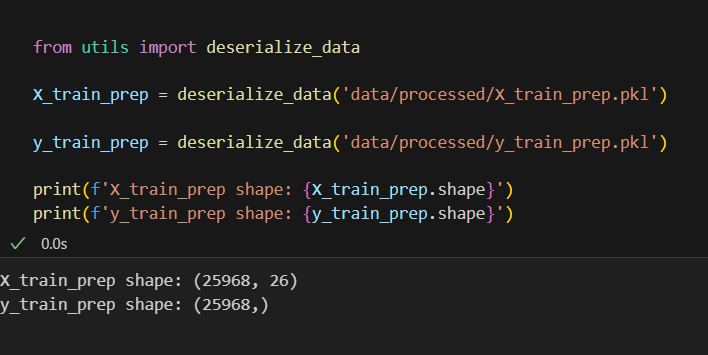
Feature engineering



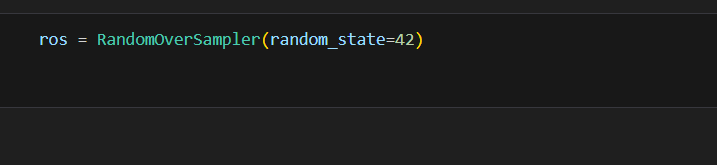
Import



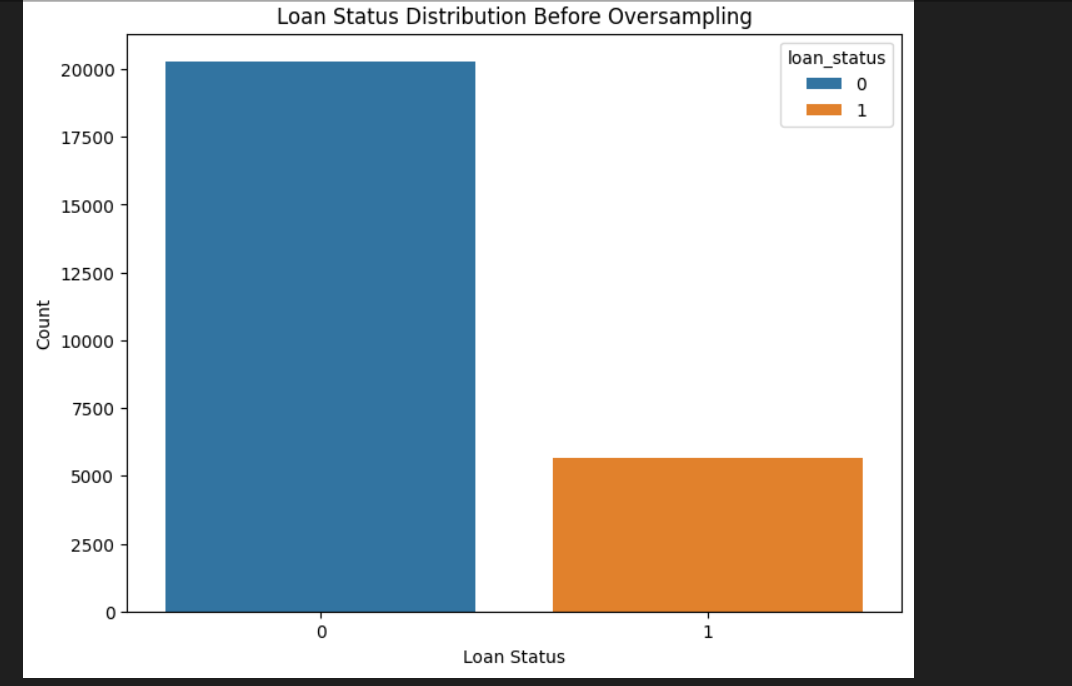
Execute deseriealize



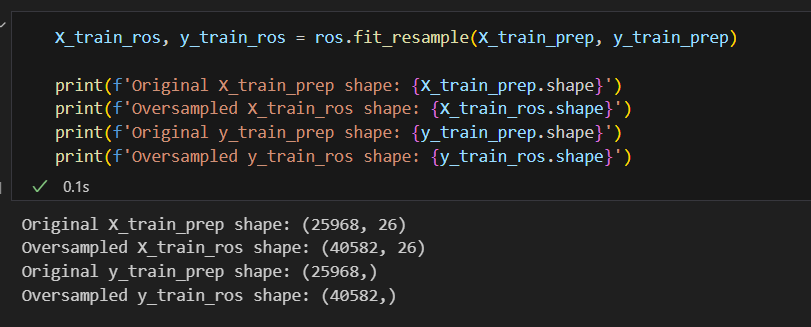
Ros



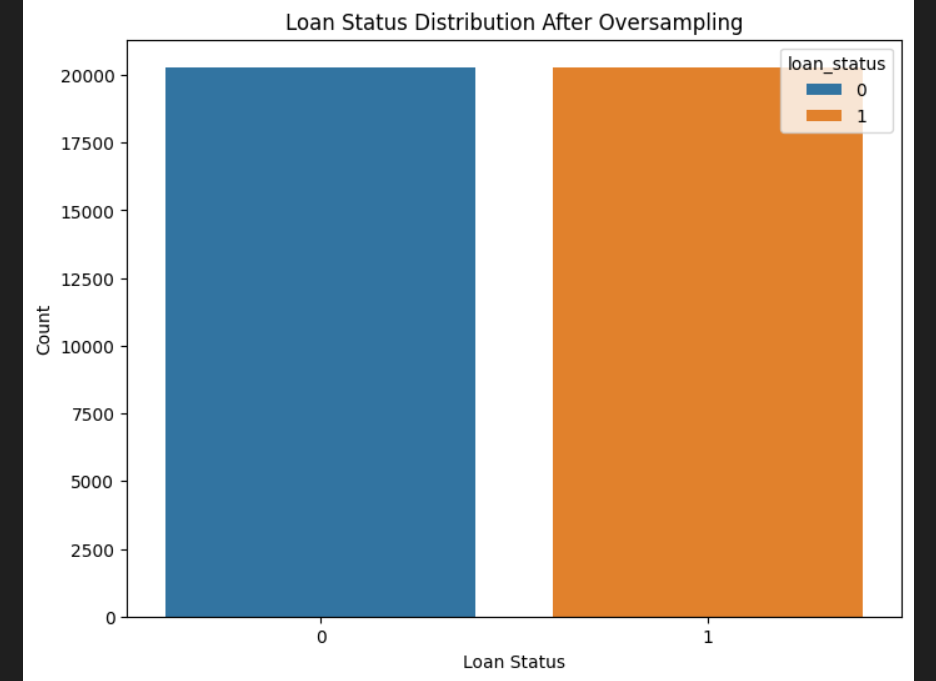
Distribution



X\_train\_ros



Result ros



Serialize

