**Lesson Learned Report**

**Learned potential business value from this analysis**

The main purpose of this analysis is to identify what factors affect clients whether they pay their bills on time, which group of customers should we target at based on their gender, marital status, education level etc. Once we find out the group of customers who are likely to default their payment, then we can take appropriate actions to reduce loss and generate more revenue for Credit One.

**Main Lessons Learned**

* Data Cleaning - Missed the duplicate header row first, codes only worked after this duplicate row was deleted, need to be more diligent in the future
* Learned how to convert object to numeric
* Learned how to use One Hot Encoding ( convert all objects to numeric at once ) and Label Encoding ( convert separately for each variable )
* Learned how to obtain data from SQL and how to extract a CSV file from Jupyter notebook
* Learned how to create a HTML EDA report in Jupyter Notebook
* Further practice on plotting different visualizations with more variables than Course 1 to find out their relationships and meanings
* Problem of importing Pandas\_Profiling to Jupyter notebook, after receiving guidance from mentor, the solution is the newest version of PP needs to be installed in terminal

**Recommendations Regarding My Findings**

The Correlation relationship between different default payment and other variables such as Limit\_bal, Sex, Education etc. are very weak and even little to none relationship, which indicates that these factors probably wouldn’t affect whether customers pay their bills on time that much, which also means that when we apply predictive model to predict the result, the outcome may not be as accurate as it should be. Credit one might need to consider other factors such as clients’ salary, credit history etc. to take into account for better profit for the company.