**Phase 3 Presentation Text**

**Kunal Singh**

Team and Project Description:  
This is group 10. I am Kunal and I my teammates are Pragat, Sahil and Mark.

HCDR is machine learning-based classification model to make accurate lending decisions for individuals by predicting loan default risk.

**Sahil Dhingra**

Visual EDA and Feature Engineering:  
For feature engineering we followed 3 approaches. First one is to plot distributions for transformations and choose one which is more normal and has less skewness score. Another is adding new columns based upon the data analysis.

In the end, we have put collinearity reducer utility which reduces features by measuring collinearity between the input variables and target. The variables with the lowest target variable correlation are dropped from the input X.

**Pragat Wagle**

Results and Discussion  
The pipeline consisted of a categorical and numerical pipeline which were both combined using logistic regression. The best hyper-parameters were C 0.01, penalty l1, solver liblinear . Results of Test AUC improved from .74 from phase 1 to .7734 in phase 3 after specific feature engineering. Xgboost came in second to Logistic Regression with a Test AUC of 0.7574. Our Kaggle submission improved from .762 to .769 for our public score.

**Mark Green**Hi this is Mark. Phase 3 was mostly spent on running pipelines, doing some more feature engineering, and trying different hyperparameters on our model. The accuracy was about the same on the top performing model as the baseline, but the AUC-ROC improved by about 0.034. In the previous phase we completed the data aggregation and merging, and some exploratory data analysis In this phase we tuned hyperparameters and engineered features and got a really good model going. And in the next phase we are going to improve upon our model by implementing a Multilayer Perceptron.