Transcript of video:

Hello everyone this is group 11 and we are gonna present phase 1 of our hdcr project.

This is how our project phase 2 is going to look like in outline.

The goal of our project is to create a model that can forecast if a potential borrower of a new mortgage will be a "default risk" or not.

In this phase 2, we have examined the HCDR data and conducted the exploratory data analysis and built baseline pipeline

In the Visual EDA you can see the different features and how they effect the default risk.

We have done the correlation analysis and you can see the heat map top 4 features showing corrrelation again target variable

For modeing pipeline we did feature scaling and divided our dataset into train test amd validation and trained our baseline model for different classifiers.

We can now see the performance metrics of baseline modes and comparison of these models for selecting best performing model.

After submitting to Kaggle, we received an AUC private score of 71.54% and a public score of 71.49%.

Based on the results that we just presented we have concluded that our best baseline model is logistic reg and XGBoost classifier.

Talking about the four P’s of this phase. These are the past present and future work done or to be done by the team and the problems we faced during this phase.