Shobhit Sinha
Hello, everyone. My name is Shobhit Sinha , and I am here with my group, Group 15, to present our project "Home Credit Default Risk".
We hope you enjoy our presentation. In this presentation we will cover the following. The aim was to accurately predict the likelihood of a loan applicant defaulting on their loan. This would help make informed decision, which can lead to less default on the loan. Next Akash will take you through the next sections.

Akash
EDA and modelling

Step1 was to understand the data and the requirement and align with the project goal. Step2 was to understand the demographic analysis . Step3 included handling the missing values and step 4 and step5 summary statistics, univariate analysis and bivariate analysis and encoding of the categorical data respectively

Some the the correlation charts and other graphs in our finding are presented here.

Dhairya Shah
Now, we will talk about modelling pipeline. This is the typical flow diagram of the project. Since, we did EDA and visual so we will talk about that. First we have analyzed the columns and then we have categorized it. Then we have dropped missing values to better analyze the data. Last two parts are Correlation analysis and pair based visualization. Finally, we plotted all the graphs. Now, Laya will take over

Laya Harwin
Coming to Comparing results, we implemented three algorithm and we got the highest accuracy for Decision Tree Algorithm. The highest ROC score is for Random Forest. We got the highest private and public score for Decision Tree in Kaggle submission.

Further we are planning to improvise the feature engineering, perform hyperparameter tuning for our models alongside using K-Fold cross validation and GridSearchCV, we might also use some advanced gradient boosting models so that we could get as close to the best accuracy as we can.

Thank you