**Project 1 : House Loan Data**

**Screen Shots**

A screenshot of a social media post

Description automatically generated

A screenshot of a cell phone

Description automatically generated

A close up of a logo

Description automatically generated

A picture containing orange, sitting, light, table

Description automatically generated

**A screenshot of a cell phone

Description automatically generated**

A screenshot of a cell phone

Description automatically generated

A picture containing table, sitting, orange, computer

Description automatically generated

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

**A picture containing computer

Description automatically generated**

**A screenshot of a cell phone

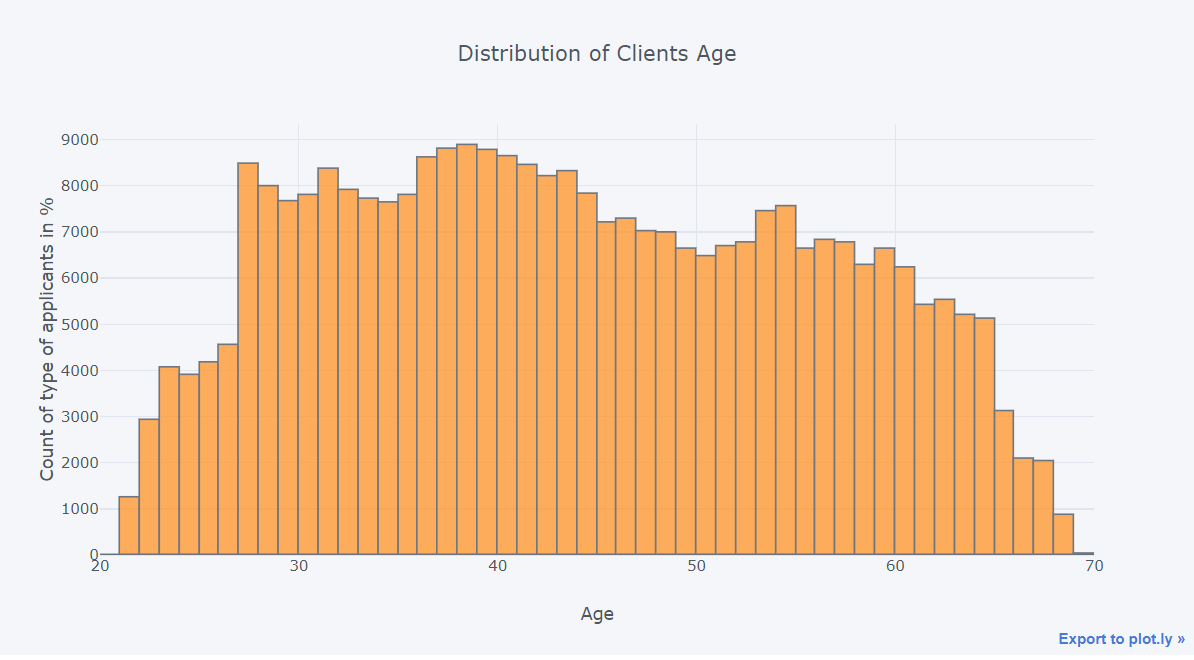
Description automatically generated**

A picture containing computer

Description automatically generated

A screenshot of a cell phone

Description automatically generated



A picture containing black, white

Description automatically generated

A screenshot of a cell phone

Description automatically generated

A screenshot of a computer screen

Description automatically generated

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

A screenshot of a social media post

Description automatically generated

A screenshot of a social media post

Description automatically generated

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

# **Machine Learning Models:**

**Logistic regression with selected features:**

Logistic Regression finds a hyperplane which best seperates the given positive and negative data points.

A close up of a map

Description automatically generated

Cross validation results and plot for Logistic Regression model.

A screenshot of a cell phone

Description automatically generated

Logistic Regression model results

A close up of a map

Description automatically generated

ROC curve for Logistic Regression model with AUC=0.754

**Random Forest with selected features:**

A screenshot of a cell phone

Description automatically generated

Cross validation results for Random Forest model.

A screenshot of a cell phone

Description automatically generated

Random Forest model results.

A close up of a map

Description automatically generated

ROC curve for Random Forest model with AUC=0.75

**LightGBM with selected features:**

A screenshot of a cell phone

Description automatically generated

LightGBM model Results

A close up of a map

Description automatically generated

**ROC curve for LightGBM model with AUC=0.787**

## Overview of Results:

|  |  |  |  |
| --- | --- | --- | --- |
| **Model** | **Train AUC** | **Validation AUC** | **Test AUC** |
| Logistic Regression with Selected features | 0.756 | 0.747 | 0.753 |
| Random Forest with Selected features | 0.841 | 0.751 | 0.751 |
| **LightGBM with Selected features** | **0.861** | **0.781** | **0.787** |

**LightGBM** gives the best performance and it is also faster to train when compared to Xgboost.