Loan approval on predicted real estate price using ML

Team 5
Yash Gupta - 191IT158
Sarthak Jain - 191IT145
Rishit - 191IT141

Introduction

Whenever a person is searching a property to buy, his major concerns are- What should be the appropriate price for that property ? Will I get a loan to buy this property?

By this project we have tried to solve both these problems using Machine Learning. The user will need to provide the details of the property such as location, size, bhk, etc. and the model will predict an appropriate price for that property.

Secondally, the user may require loan for the purchase, so he may also test if he can get a loan for the property. User will need to enter his financial details, etc to get the probability of getting the loan approved on predicted price.

Methodology

Price Prediction

- Web Scraping
- Data Cleaning.
- Training the dataset on algorithms such as XGBoost, lasso, decision tree for regression.
- Generation of pickle file.

Loan Approval

- Data Cleaning.
- Training the dataset using various classifiers to get the probability of getting loan.
- Generation of Pickle file.

Providing a UI for easy use of models, i.e, prediction of price, loan approval and buying and selling properties.

Housing Dataset

Dataset : Data of 10 cities scraped from makaan.com

www.kaggle.com/dataset/91529c1aa475816f7a1f5f1661e350917964f7624cb02058469673c4ba63f5c6

features : (name,price,squarefeet,bhk,type of property (villa,appartment etc))

Exact dimensions : 40000 rows and 6 columns

There were few tuples which had bhk as null so we filled it with the median of the data

Price Prediction

There are three columns location,city and type of property which are categorical in nature so label encoding is performed on them and than using the grid search cv best algorithm among (extreme gradient boosting,decision tree,ada boost,catboost and light gbm is picked along with best parameters)

Best picked algorithm is trained on scrapped dataset and pickle file is generated that will be used to make prediction of price in our application.

Loan Approval dataset

Dataset : Housing Loan dataset

https://www.kaggle.com/mayanksingh09/loanapprovaldataset

Important features : Gender, Loan_ID, Married, Dependants, Education, NAME_HOUSING_TYPE, BHK,' ApplicantIncome, TOTALAREA_MODE, AMT_GOODS_PRICE

Exact dimensions : 307511 rows, 12 columns

Loan Approval

Data Preprocessing

The dataset contained null values in the total area and Housing type column of the dataset which were filled with the mean and mode respectively.

The dependants columns had arbitrary vague values which were refined for label encoding to follow.

One hot encoding was performed on categorical features for the ML models to train on.

Loan Approval

Training and Testing

The cleaned dataset was trained on multiple classifiers - Logistic Regression, Naive Bayes Gaussian, Naive Bayes Multinomial and Decision trees.

Their cross validation scores were recorded with 5 cross fold splits.

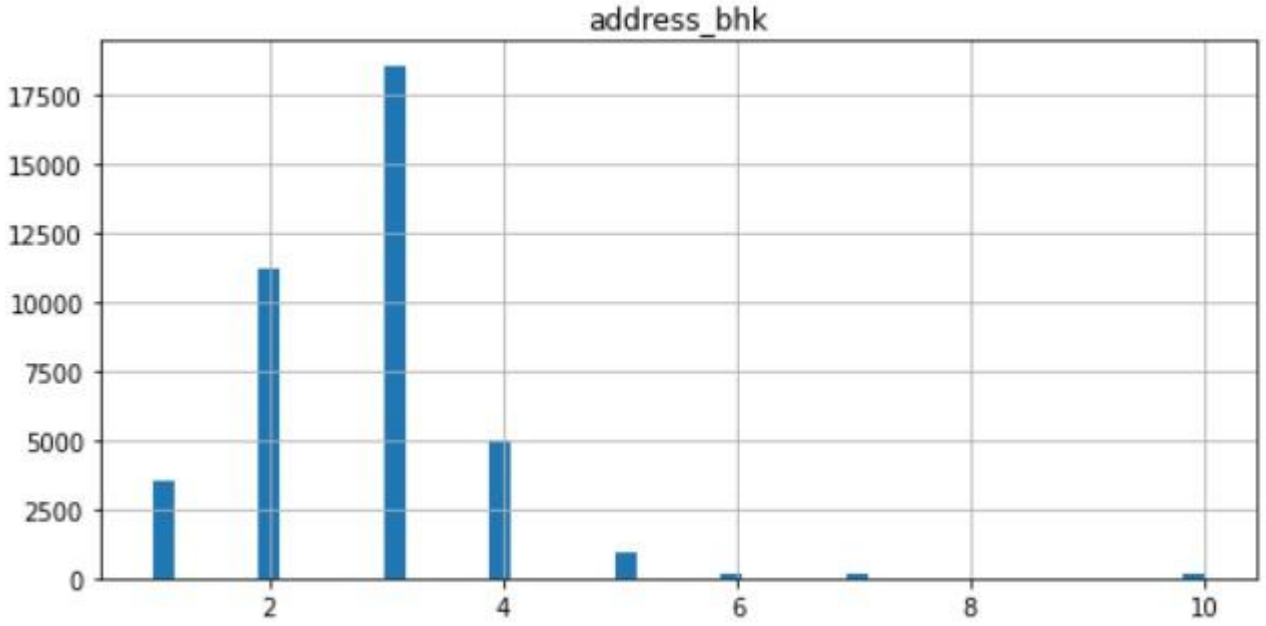
The best model was picked using GridSearchCV and the model was dumped into a pickle file for the server to make predictions.

Novelty

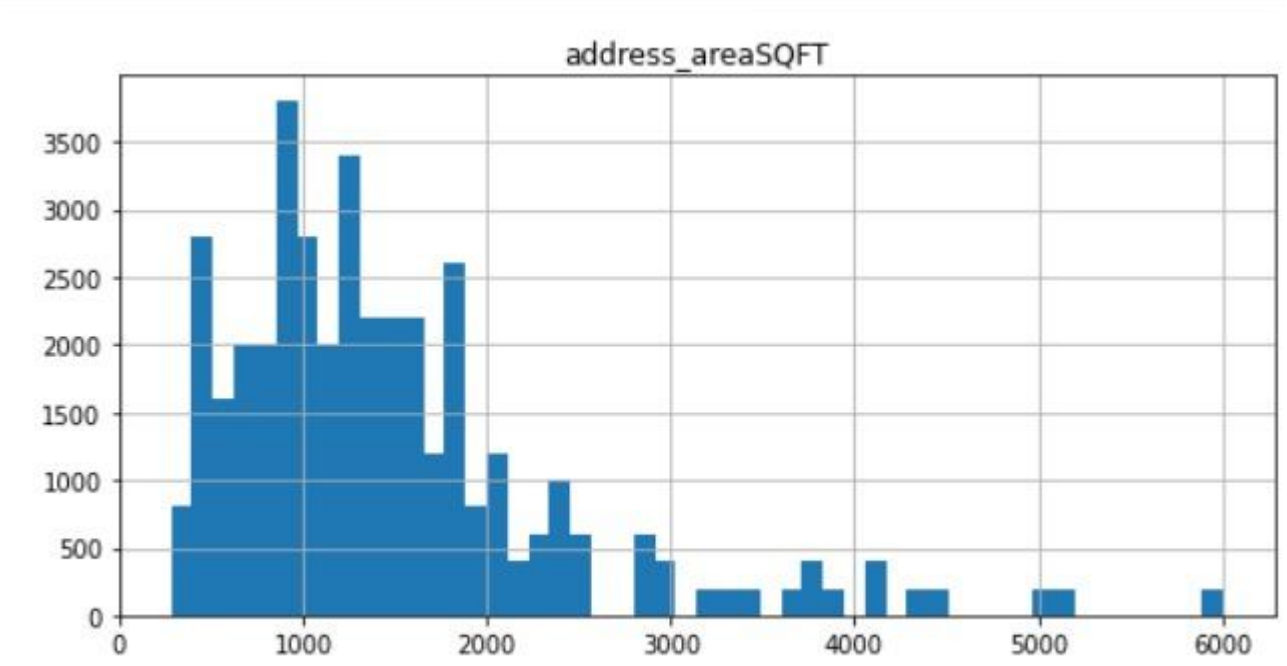
- New housing Dataset scraped from makaan.com
- Interconnection between housing predicted price and loan approval datasets

Results

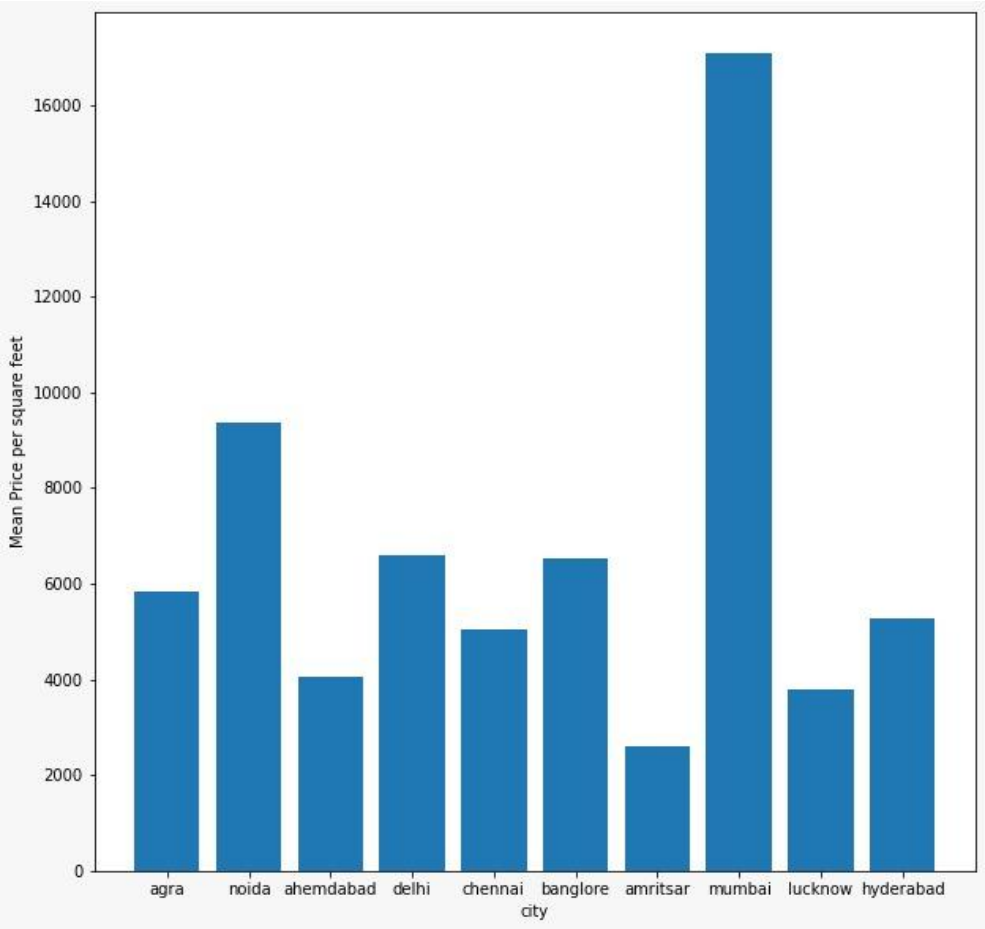
Analysis on scrapped dataset



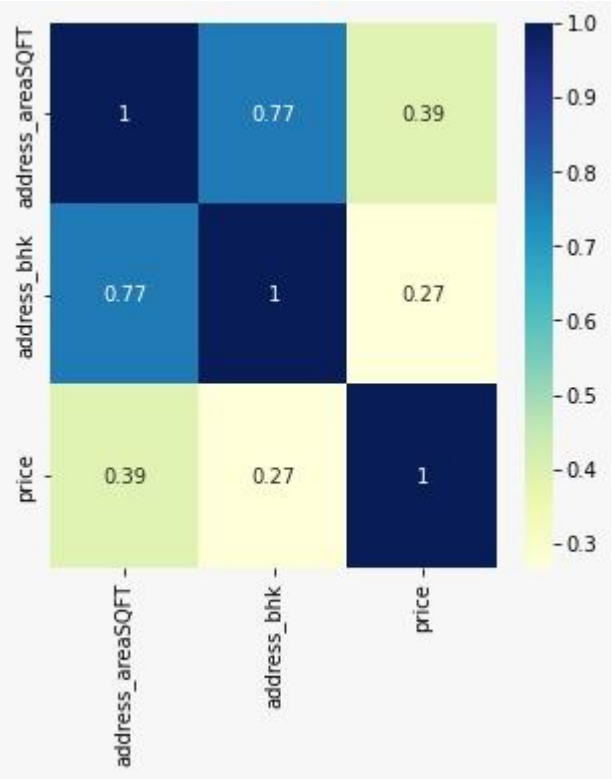
No of rows (y-axis) corresponding to bhk(x axis)



No of rows (y-axis) corresponding to area in square feet(x-axis)



Mean price per square feet city wise analysis

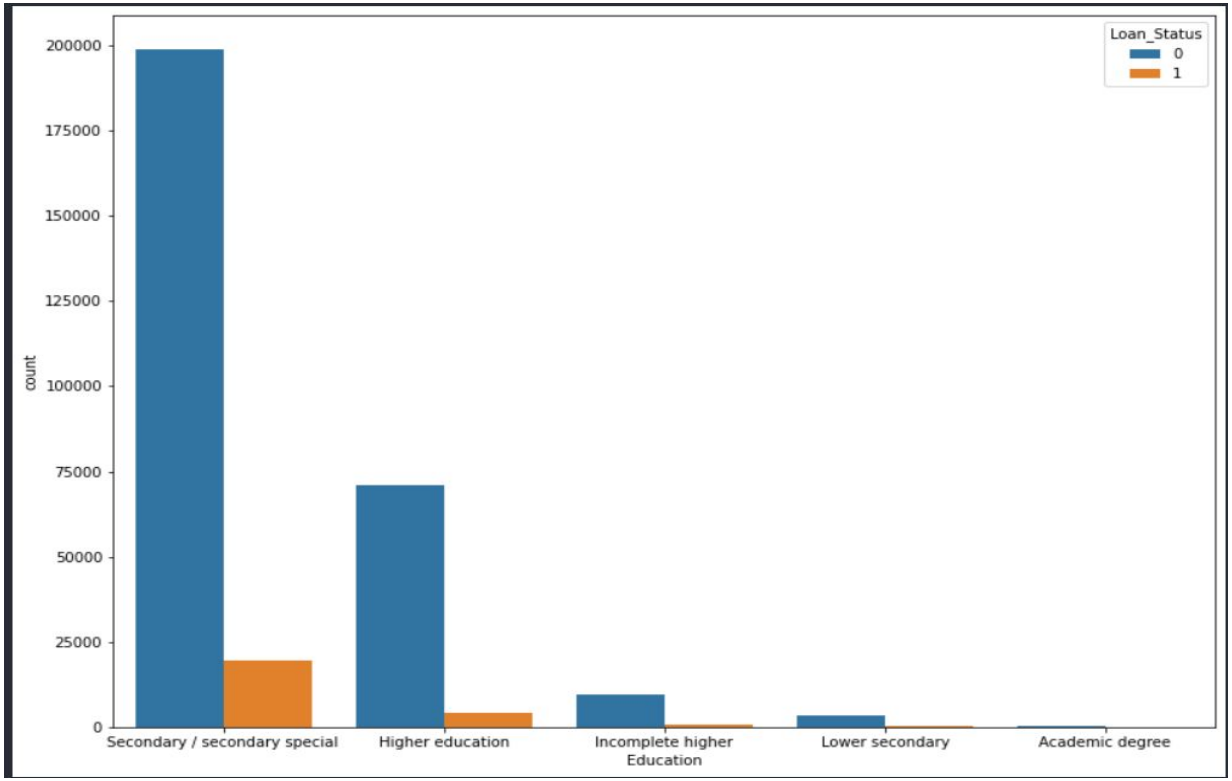


Corelation matrix on columns with numerical value of scrapped dataset

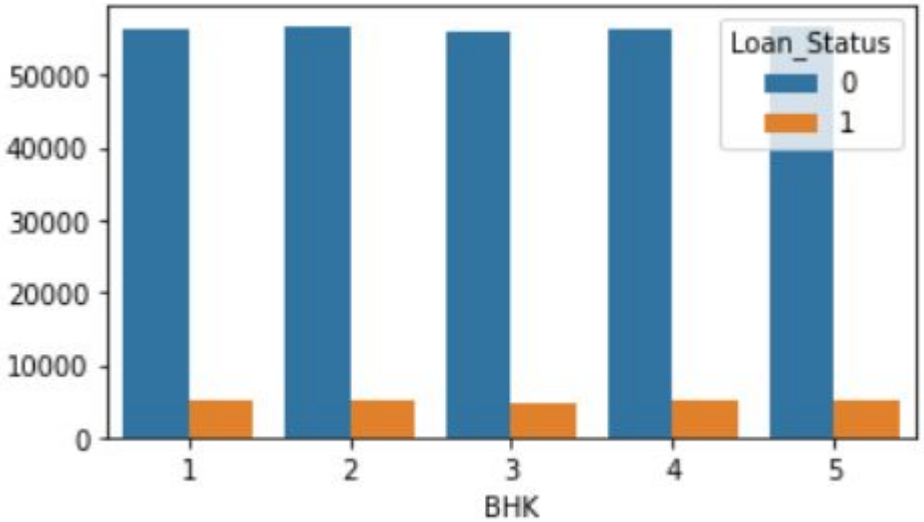
Score of various algorithms on scrapped dataset (for price prediction)

	model	best_score	best_params
0	decision_tree	0.975019	{'criterion': 'mse', 'max_depth': 9, 'splitter...
1	ada	0.656284	{'learning_rate': 1, 'n_estimators': 200}
2	xgboost	0.913266	{'max_depth': 15, 'n_estimators': 15}
3	lightGBM	0.978750	{'learning_rate': 0.09, 'max_depth': 4}
4	catboost	0.875169	{'depth': 3}

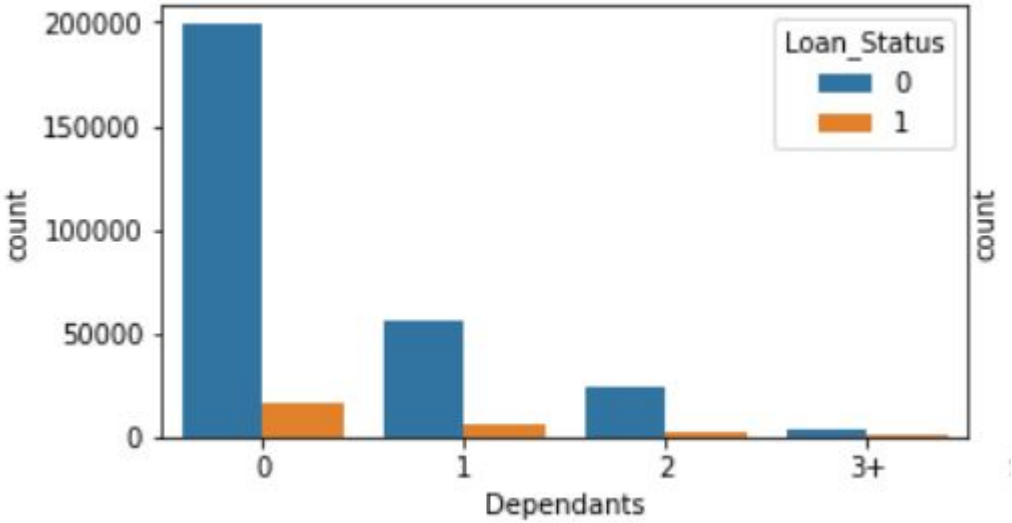
Results - Education vs Loan Status



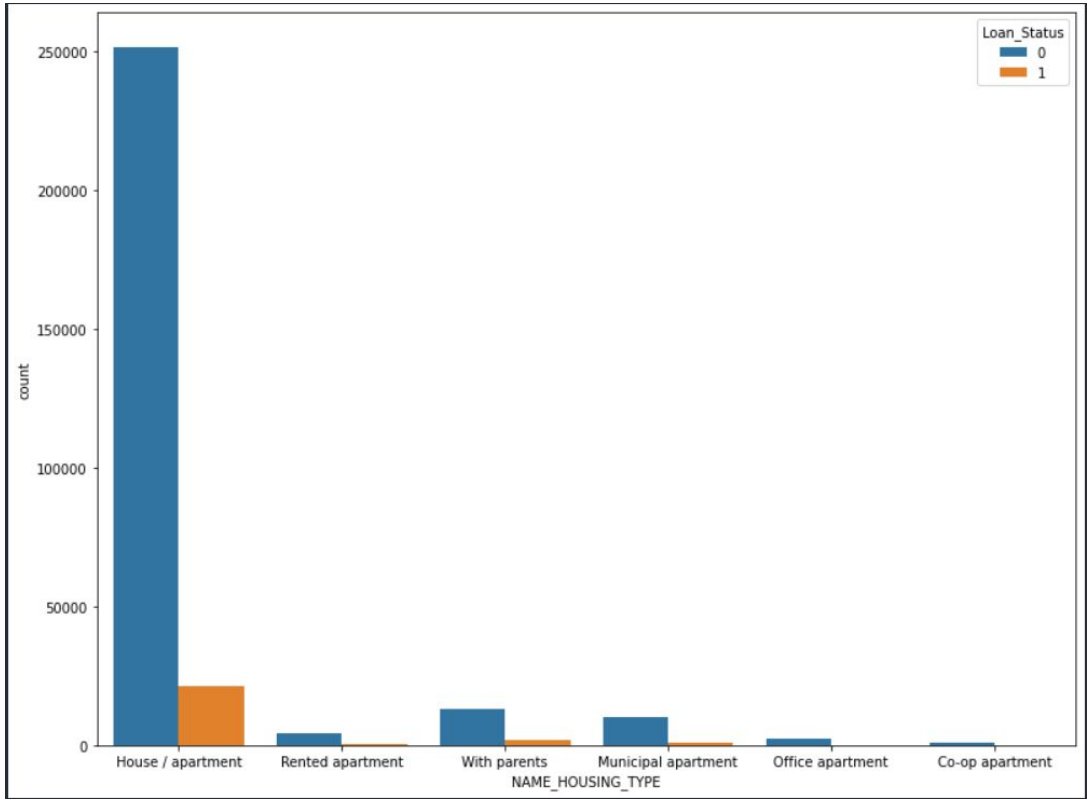
BHK vs Loan Status



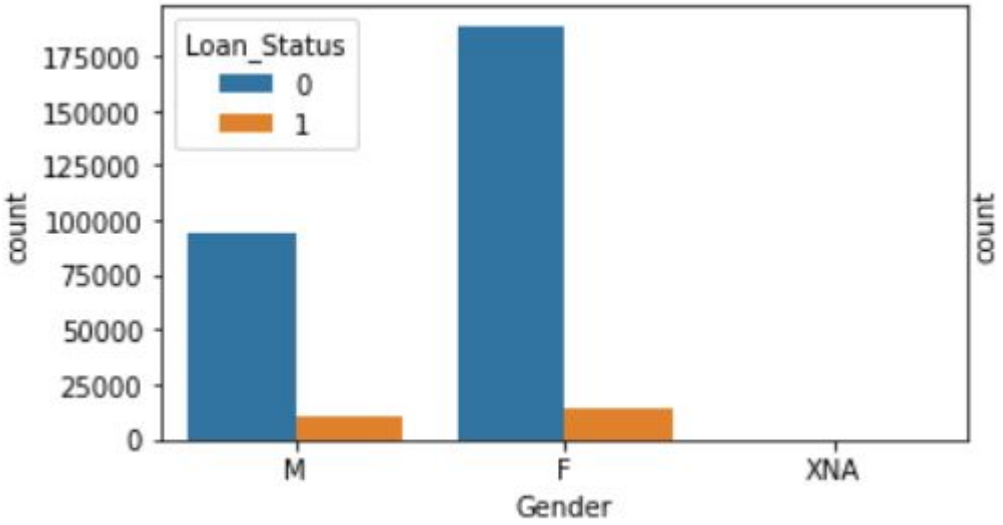
Dependants vs Loan Status



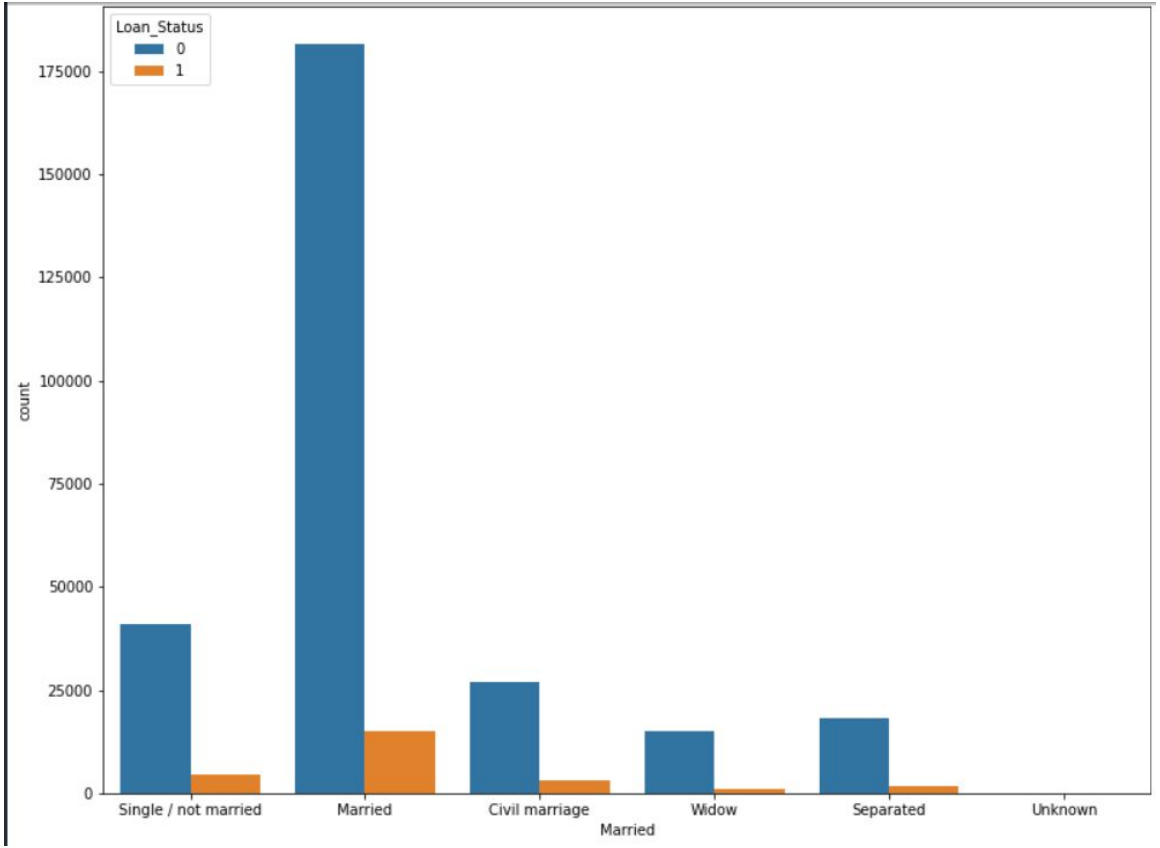
Housing Type vs Loan Status



Gender vs Loan Status



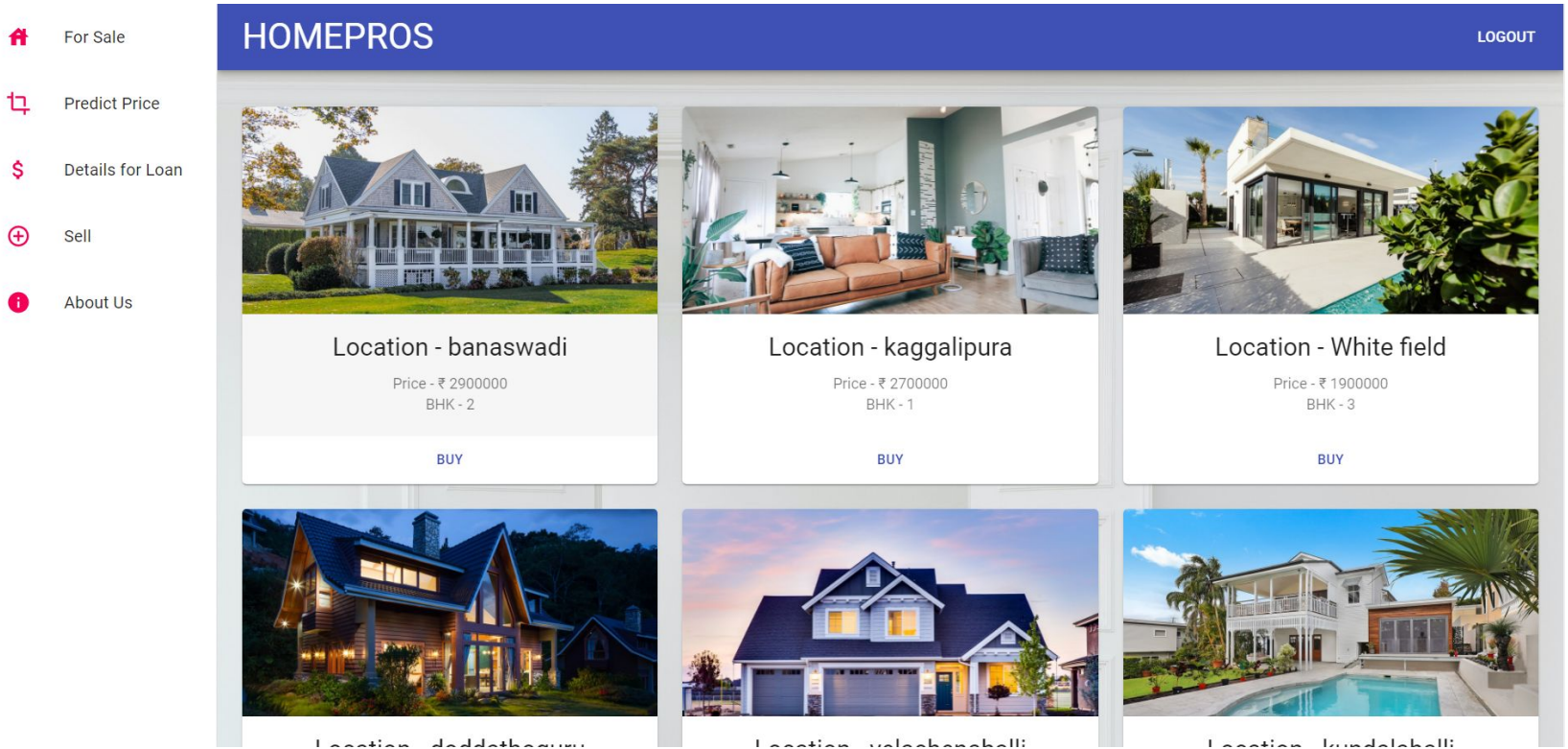
Married vs Loan Status



GridSearchCV Loan Approval

	model	best_score	best_params
0	logistic_regression	0.919271	{'C': 1}
1	naive_bayes_gaussian	0.913847	{'var_smoothing': 1e-12}
2	naive_bayes_multinomial	0.481845	{}
3	decision_tree	0.841606	{'criterion': 'entropy', 'splitter': 'best'}

Web App


For Sale
Predict Price
Details for Loan
Sell
About Us
HOMEPROS
LOGOUT
Location - banaswadi
Price - ₹ 2900000
BHK - 2
BUY
Location - kaggalipura
Price - ₹ 2700000
BHK - 1
BUY
Location - White field
Price - ₹ 1900000
BHK - 3
BUY

For Sale

Predict Price

Details for Loan

Sell

About Us

HOMEPROS

LOGOUT

Enter Property Details

City *
Banglore

Location *
Yeshwantpur

Property Type *
Apartment

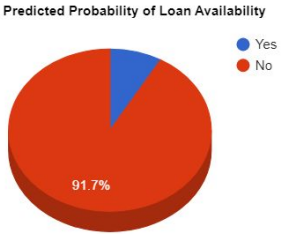
Size in Square Feet *
2500

BHK *
3

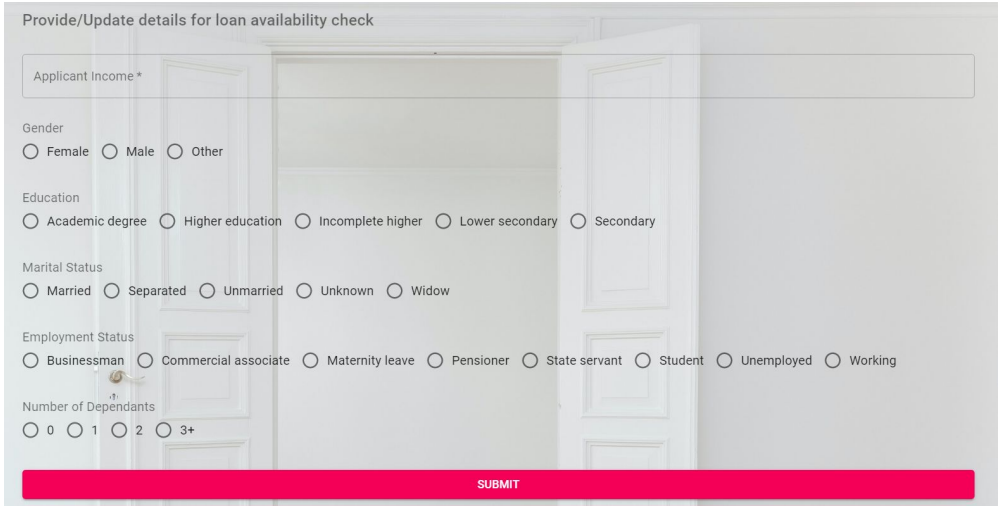
SUBMIT

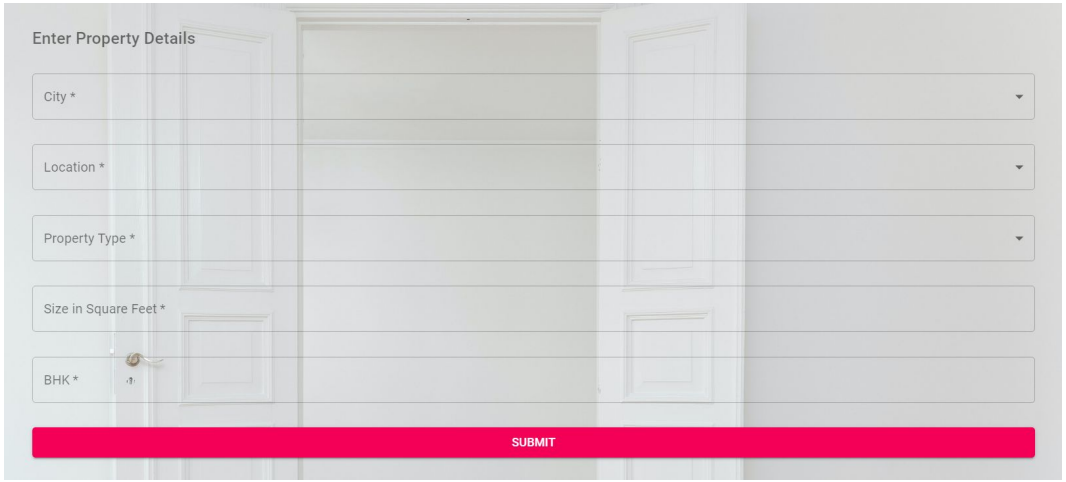
This could be your next house at just

₹ 122 lacks only



Loan Applicant Details Form





Property Details form