Bank Loan

MACHINE LEARNING CLASSIFICATION

Amal Althaqafi Soaad Aljafr Muzoon Alshahrani

Introdution

- loan is one of the most important schemes of banks..

- We mean by Bank Loan is when a bank offers to lend money to consumers for a certain time period. As a condition of the bank loan, the borrower will need to pay a certain amount of interest..

- Bank loans can be short term or long term, depending on the purpose of the loan.

Bank Loan Status Dataset

Kaggle

100,000 loan records
18 features

Target Column(Term):

- Short Term
- Long Term

Tools:

- Python
- Numpy
- Pandas
- Sklearn
- Seaborn
- Matplotlib
- imblearn
- xgboost

EDA

Exploratory data analysis

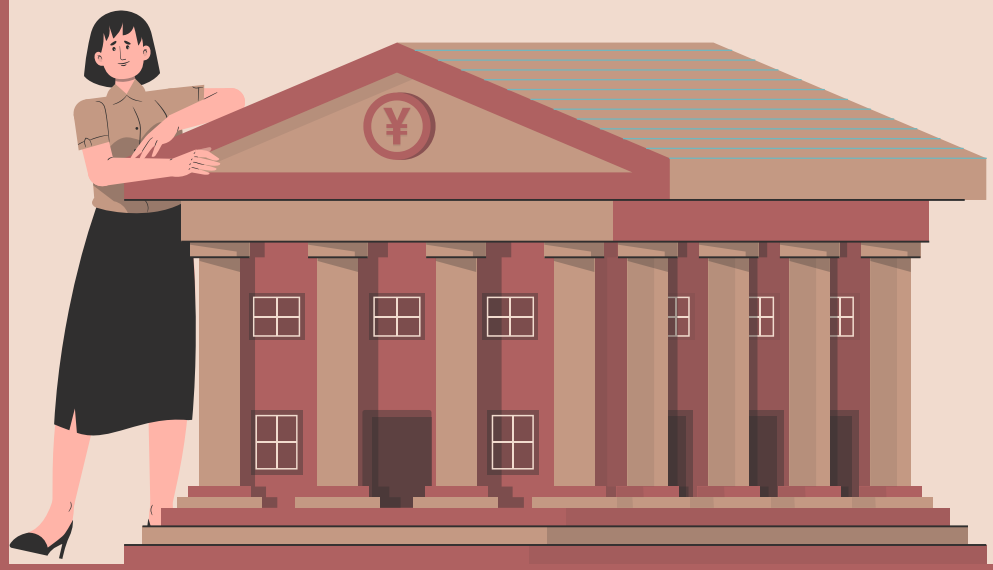
Data Cleaning

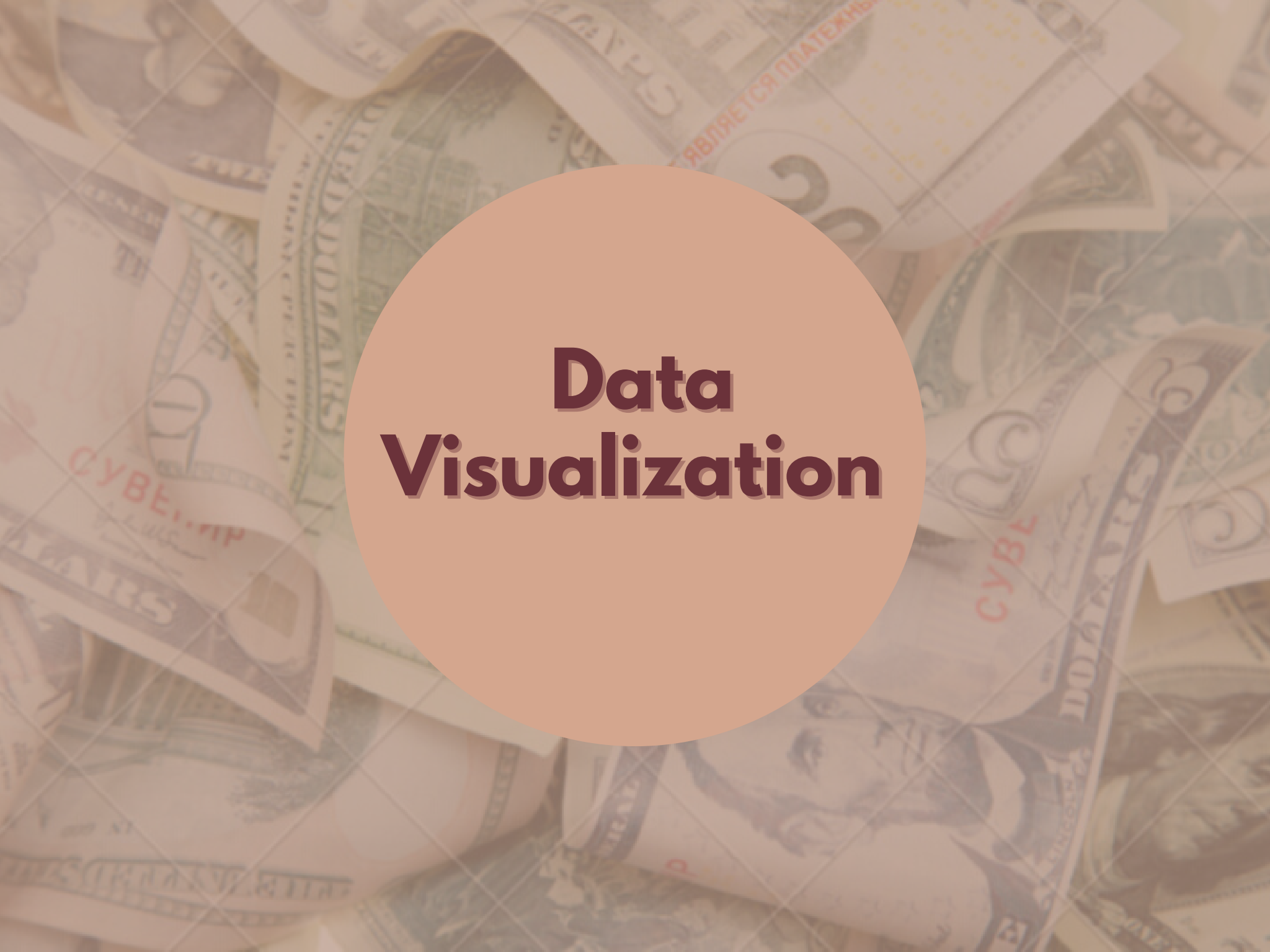
- Check for NaN and deal with them
- Drop Duplicate
- Drop unwanted columns
- Droping outliers

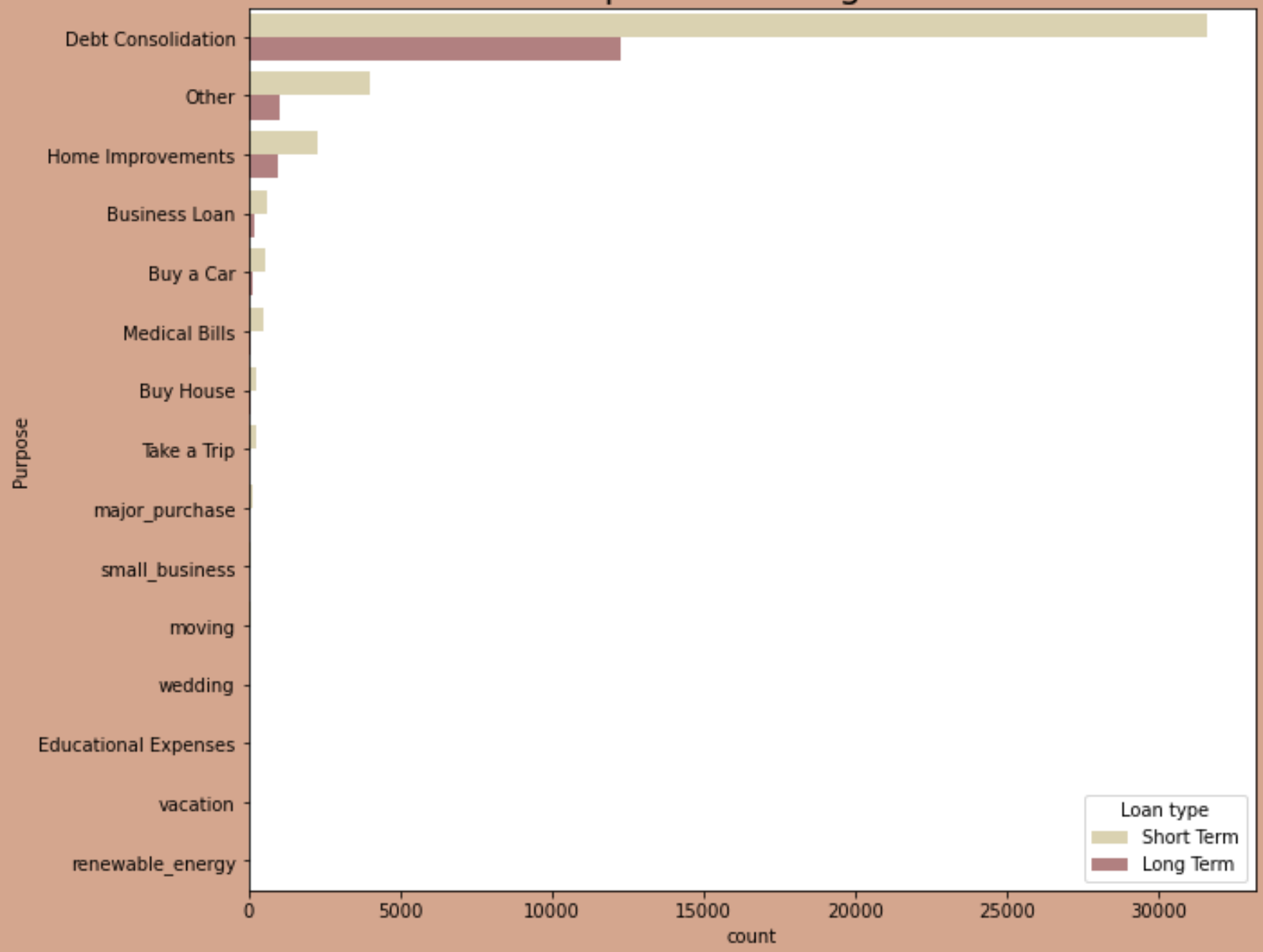
Feature Engneering

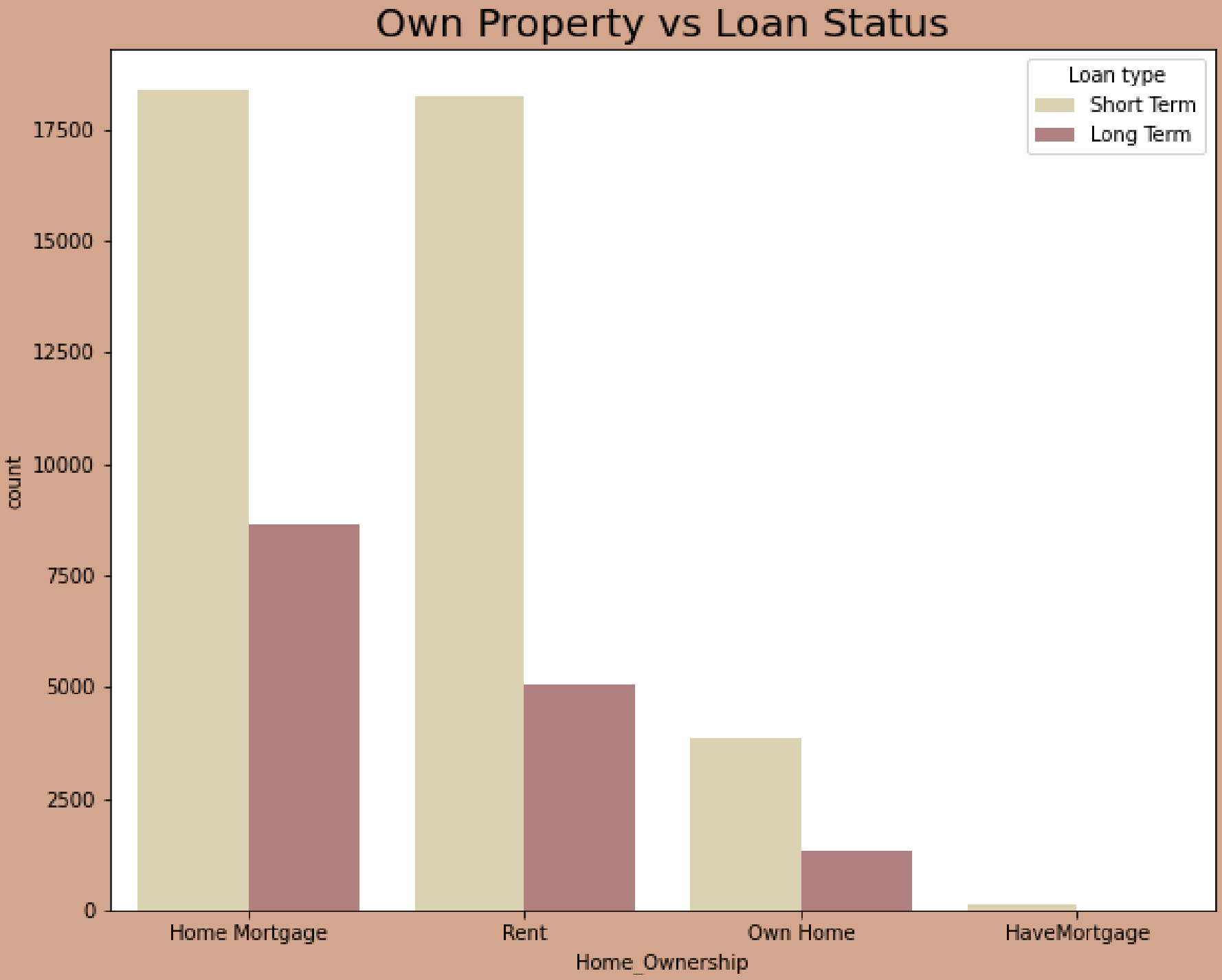
One hot Encoding

Correlation

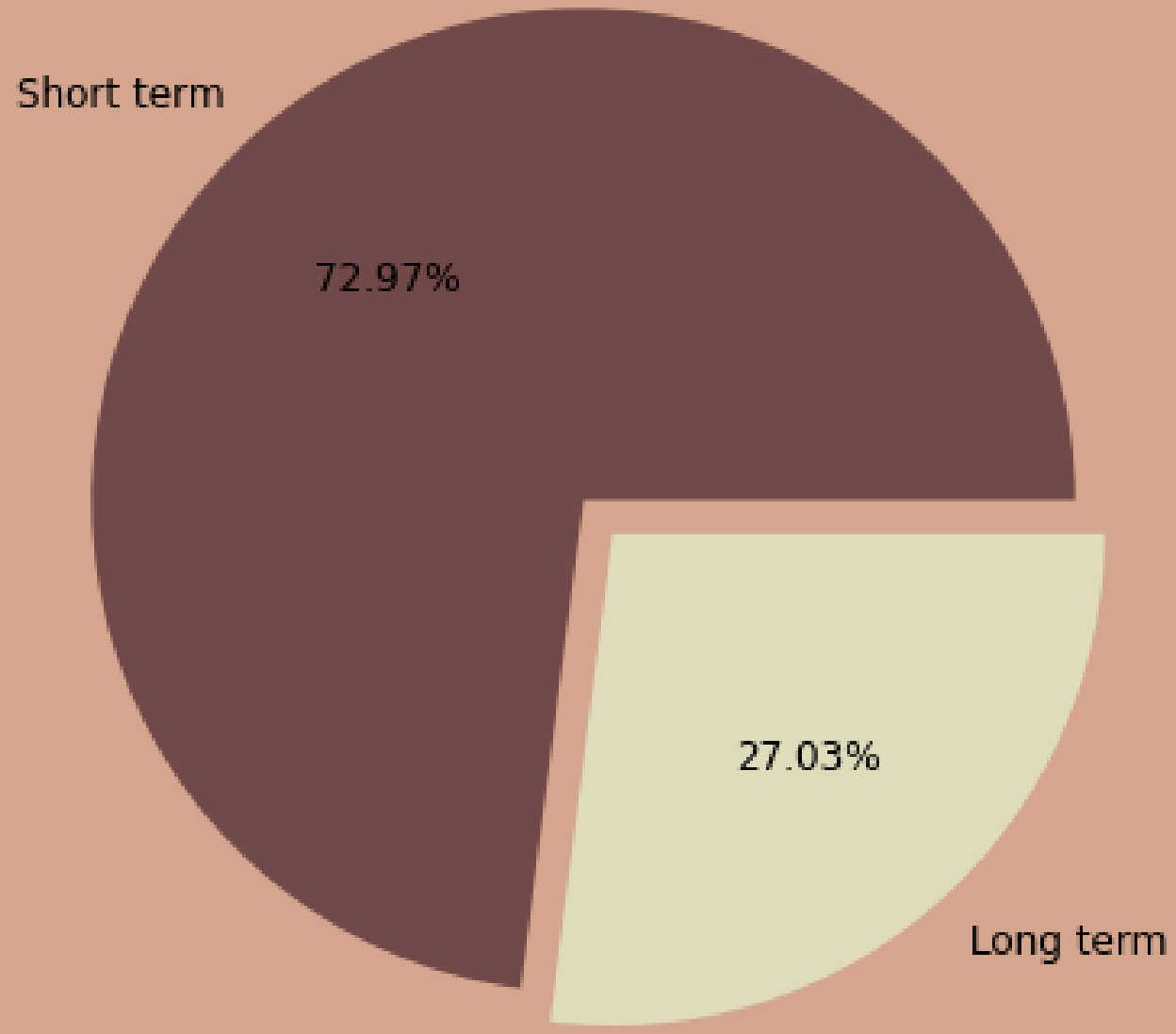



Data
Visualization


Purpose of taking Loan
Debt Consolidation
Other
Home Improvements
Business Loan
Buy a Car
Medical Bills
Buy House
Take a Trip
major_purchase
small_business
moving
wedding
Educational Expenses
vacation
renewable_energy
Purpose
0
5000
10000
15000
20000
25000
30000
count
Loan type
Short Term
Long Term


Own Property vs Loan Status
Loan type
Short Term
Long Term
17500
15000
12500
10000
7500
5000
2500
0
count
Home Mortgage
Rent
Own Home
HaveMortgage
Home_Ownership

Time Period of Taking Loan



Classification Models

Logistic Regression

Mean Smote Logistic Regression Valdition F1:

0.6901

KNN Model

Best estimator:
KNeighborsClassifier(n_neighbors=9)
Best f1_score for cross validation:

0.8520

Decision Tree Classifier

Best param after grid search {'criterion': 'entropy', 'max_depth': 8}
Best score after grid search

0.8703

Random Forest Classifier

Best param after grid search {'bootstrap': True, 'criterion': 'gini', 'n_estimators': 500}
Best score after grid search

0.8649

Extra Tree

Mean f1_score for validation:

0.8589

Stacking

Mean f1_score stacking for cross validation :

0.8531

Bagging

Mean f1_score Bagging for cross validation :

0.8529

Voting Classifer (HARD)

Mean f1_score Voting Classifer for cross validation :
0.8584

Boosting

- **AdaBoost**

Best score after grid search:
0.8198

- **Gradient Boosting**

Mean f1_score Gradient Boosting for cross validation :
0.8714

- **XGBoost**

Mean f1_score XGBoost for cross validation :
0.8731

Best Score

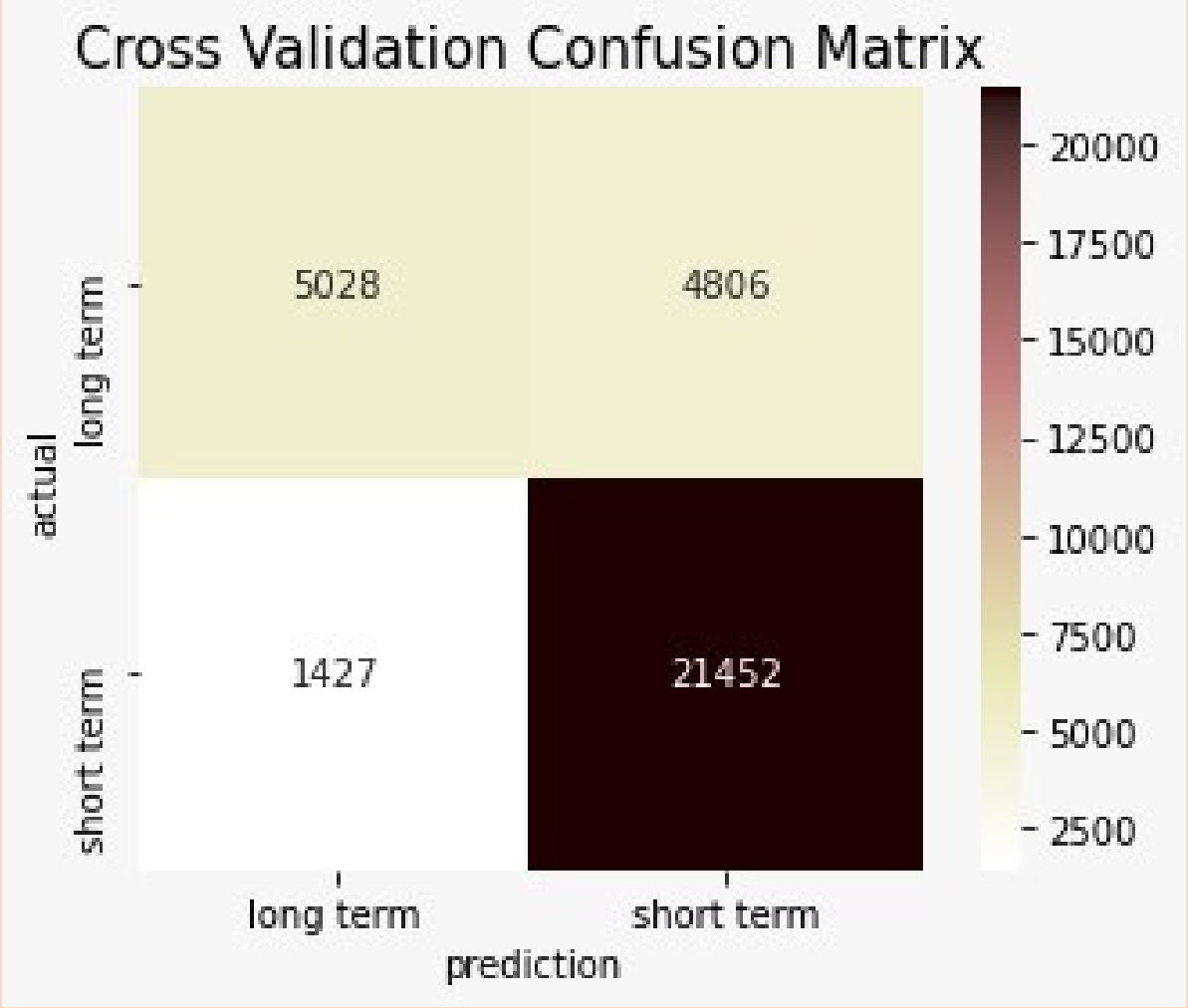
XGBoost

Mean f1_score XGBoost for cross validation :

0.8731

Test Score:

0.8691



Conclusion

The best model to predict whether a loan is a short term or a long term is **XGBoost classification model**, which gave the **highest cross validation F1**.
F1 for the ***cross validition*** set is **0.8731** and for the ***testing*** set **0.8691 .**

Thank You

FOR YOUR TIME