

# Mortgage Loan Predictor

By Alex Martin



# Poll

How many people would  
like to buy a house/or  
already own a house?







# The Situation

Owning a home is highly aspirational and a key life milestone:

- Financial stability.
- Potential property value appreciation.
- Stability.
- Free from the whims of landlords.
- Fosters a stronger sense of community.
- Offers financial security for future generations.
- A base to build from.
- And many more...



# The Problem

*“What do I have to do to get a mortgage?”*

- Mortgage approval process is a ‘black-box’
- Denied applications show on your credit report for a year.





# A Solution?

- Develop a mortgage prediction app that allows users to determine potential approval before submitting official applications.
- This will help prevent users from making unnecessary applications that could harm their chances of obtaining a mortgage in the future.
- Provide users with reasons for the decision so they know what they need to improve on in the future.



# Methodology

- Data Collection
- Data Wrangling
- EDA
- Predictive Analysis
- Predictor App



# Mortgage Disclosure Act

HOME

FILING

DATA BROWSER ▾

DATA PUBLICATION ▾

TO

Transparency & Accountability:

## U.S. Mortgage Market + Lenders

Enacted by Congress in 1975, HMDA provides the most comprehensive source of publicly available information on the U.S. Mortgage Market.

See How It Works

Explore the Data

\$ Rate Spread  
Calculator



HMDA Filing  
Platform



Frequently Asked  
Questions

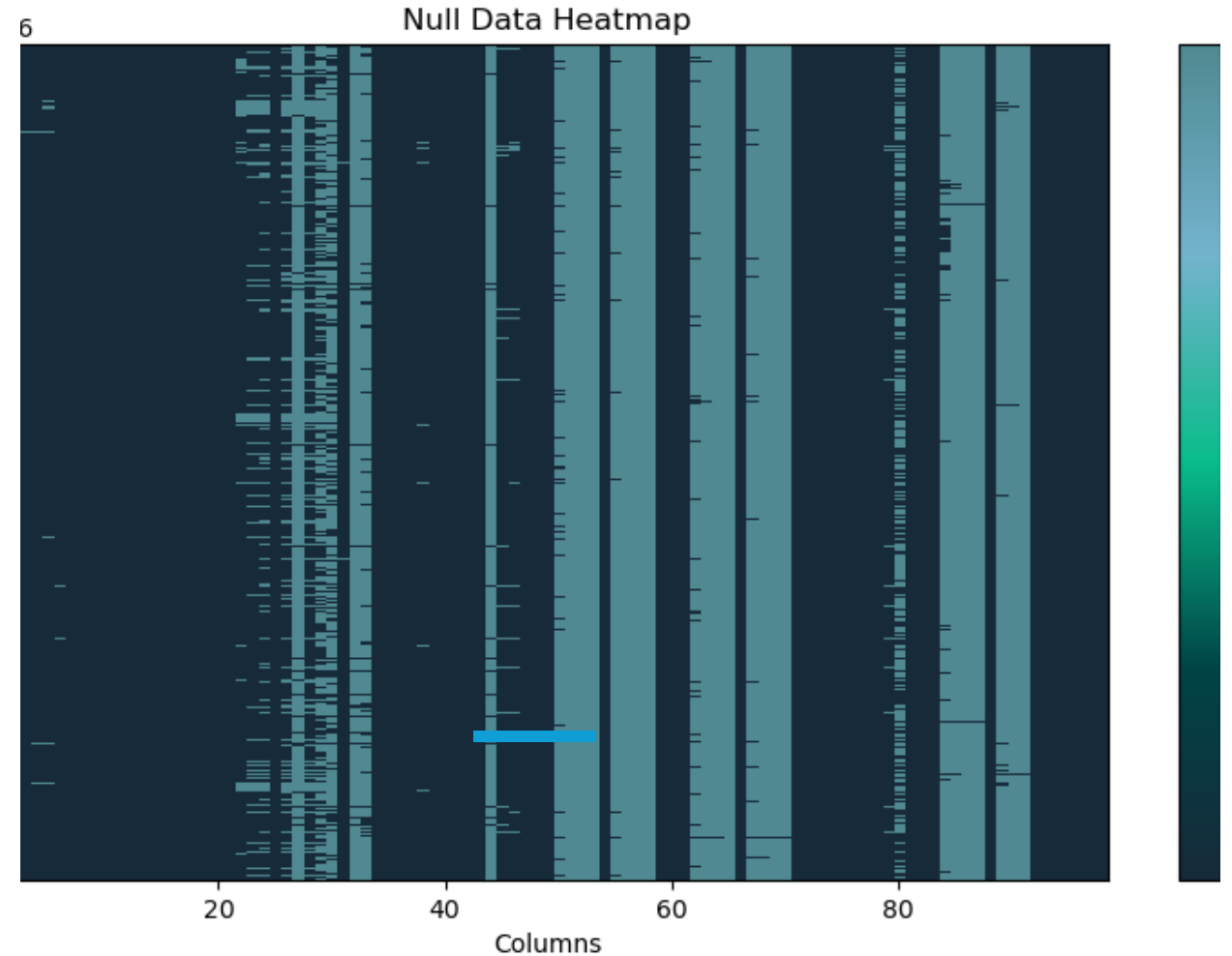
## Data Collection

- <https://ffiec.cfpb.gov/data-browser/data/2022?category=states>

# Data Wrangling

Dimension	Count
Row	5,081,179
Columns	99

Exemption	Count
'Exempt'	942,042
1111	1,756,100
8888	8,547



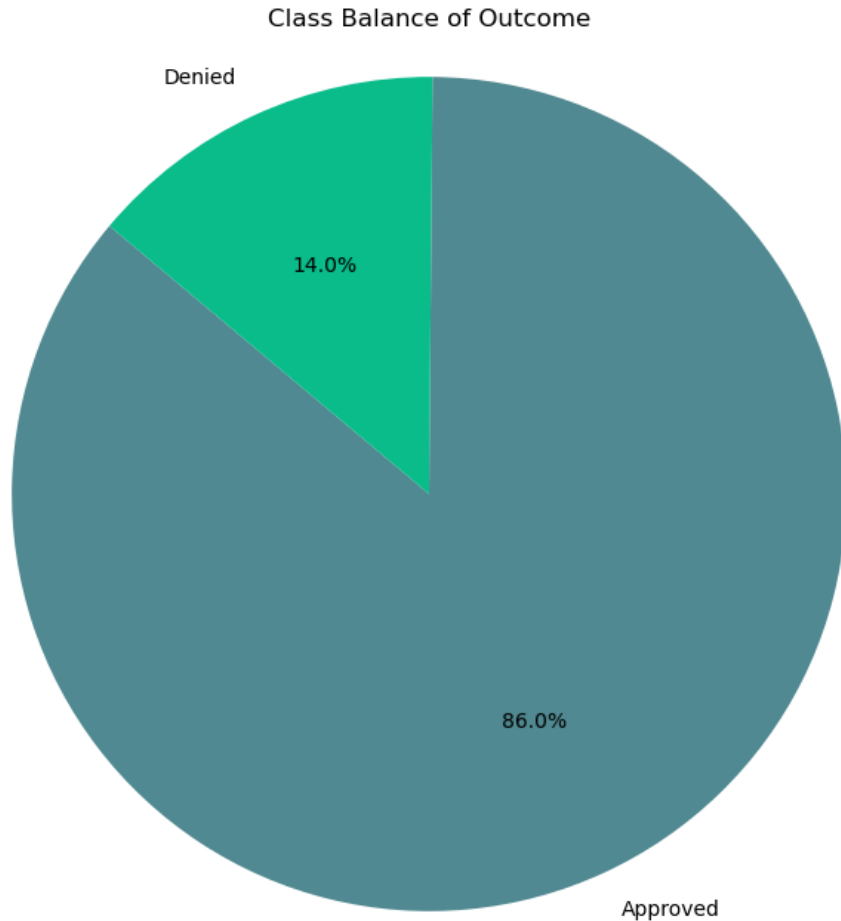


---

# Exploratory Data Analysis



# Approved v. Denied



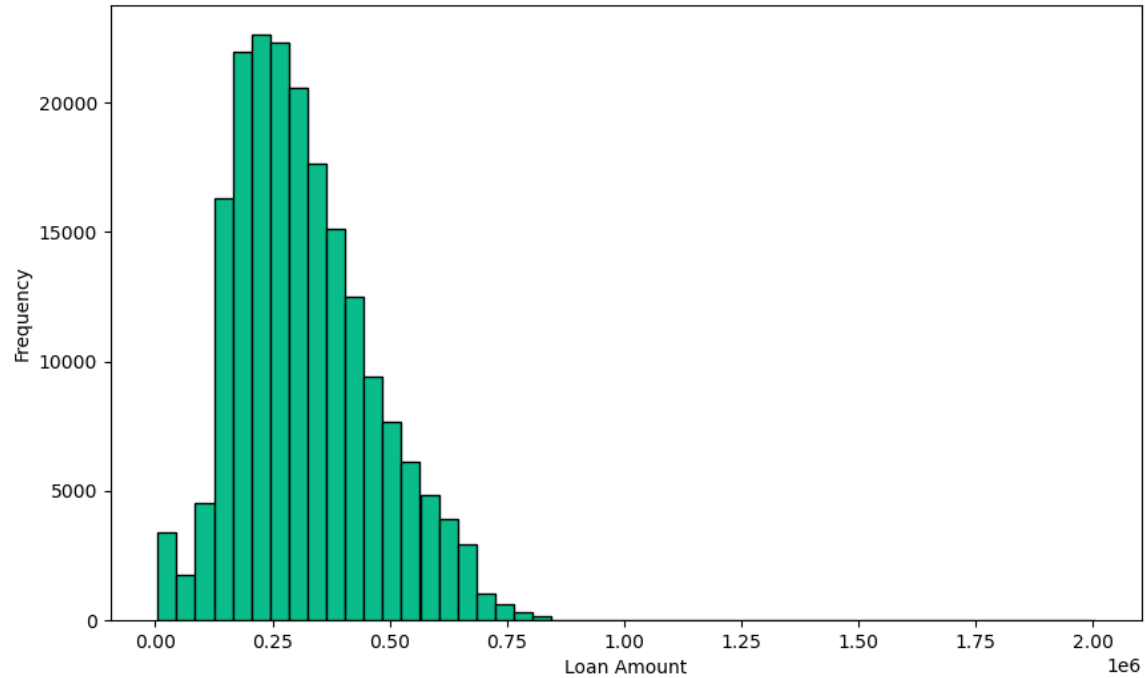
- 86% of Mortgage loans were approved
- 14% of Mortgage loans were denied
- Large class imbalance in the dataset

Outcome	Count
Approved	4,362,232
Denied	713,641



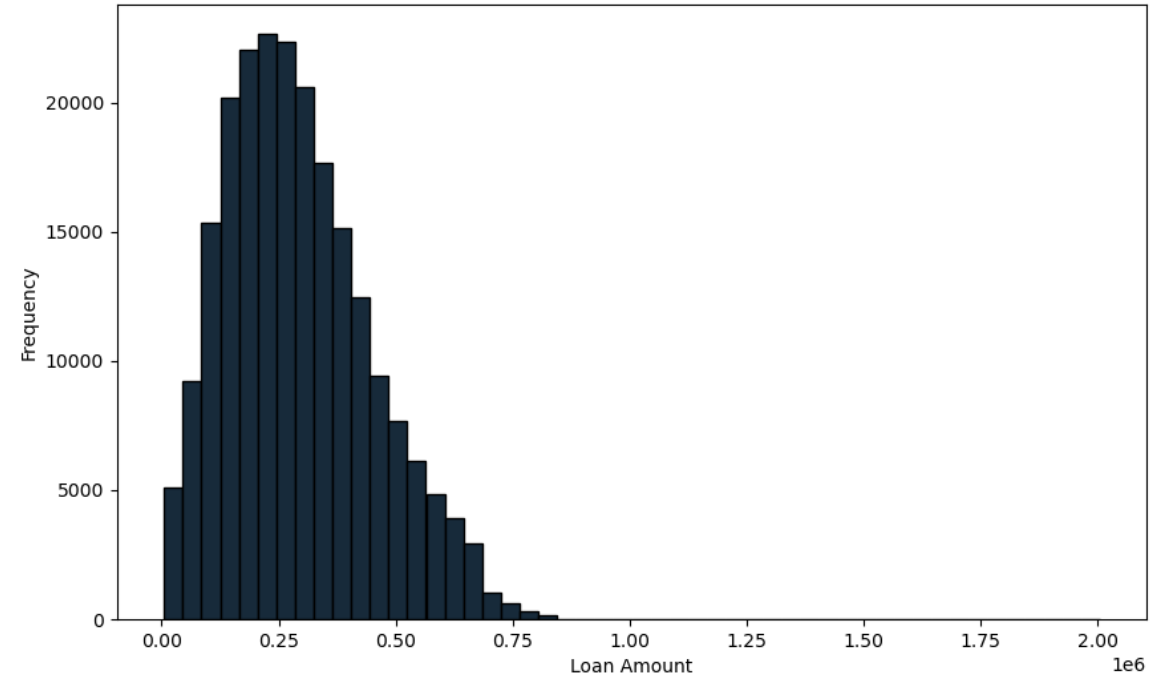
# Data Distribution

Distribution of Loan Amount



Median loan amount: \$285,000  
Mean loan amount: \$382,000  
Mode loan amount: \$205,000

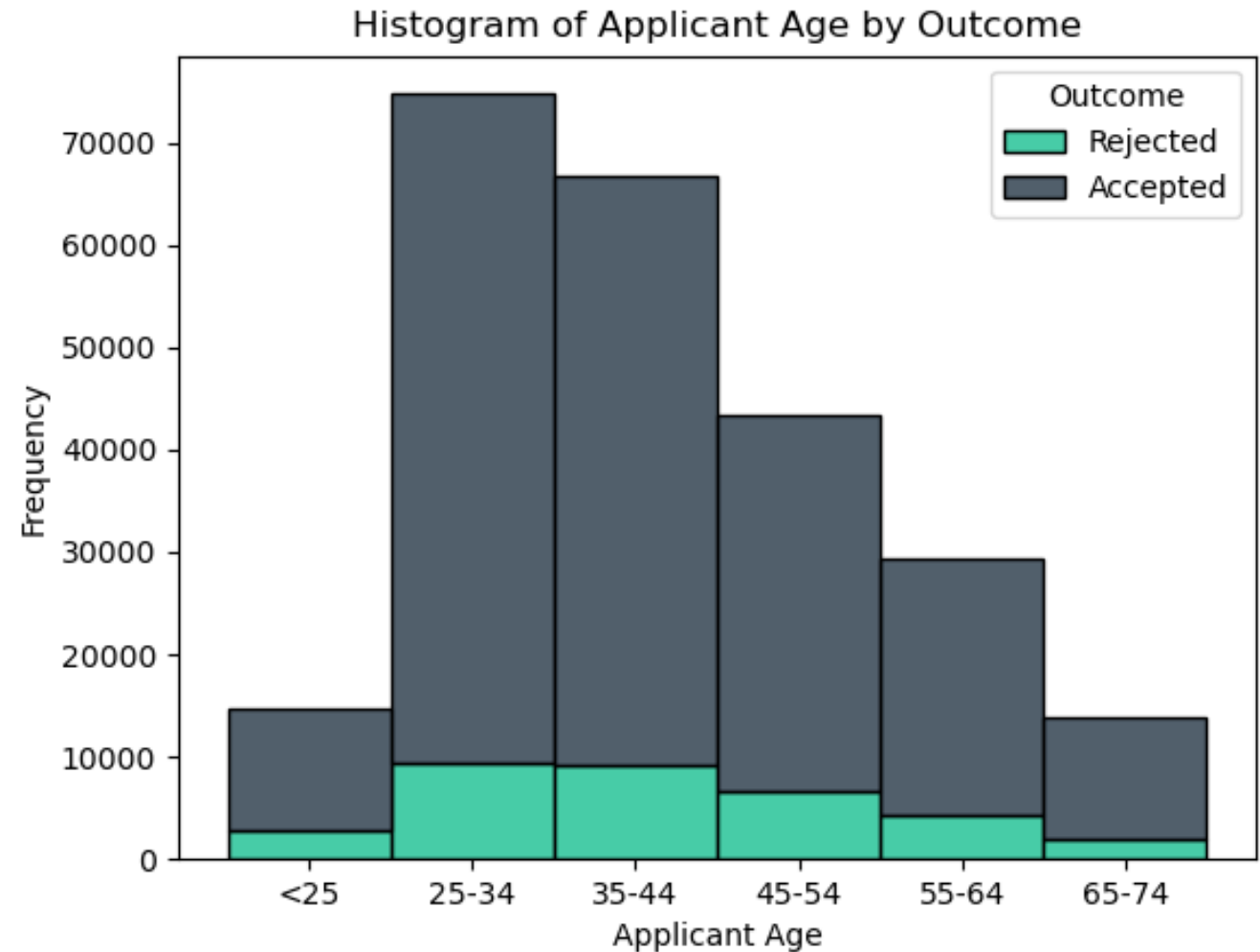
Distribution of Property Value



Median loan amount: \$355,000  
Mean loan amount: \$522,000  
Mode property value: \$255,000

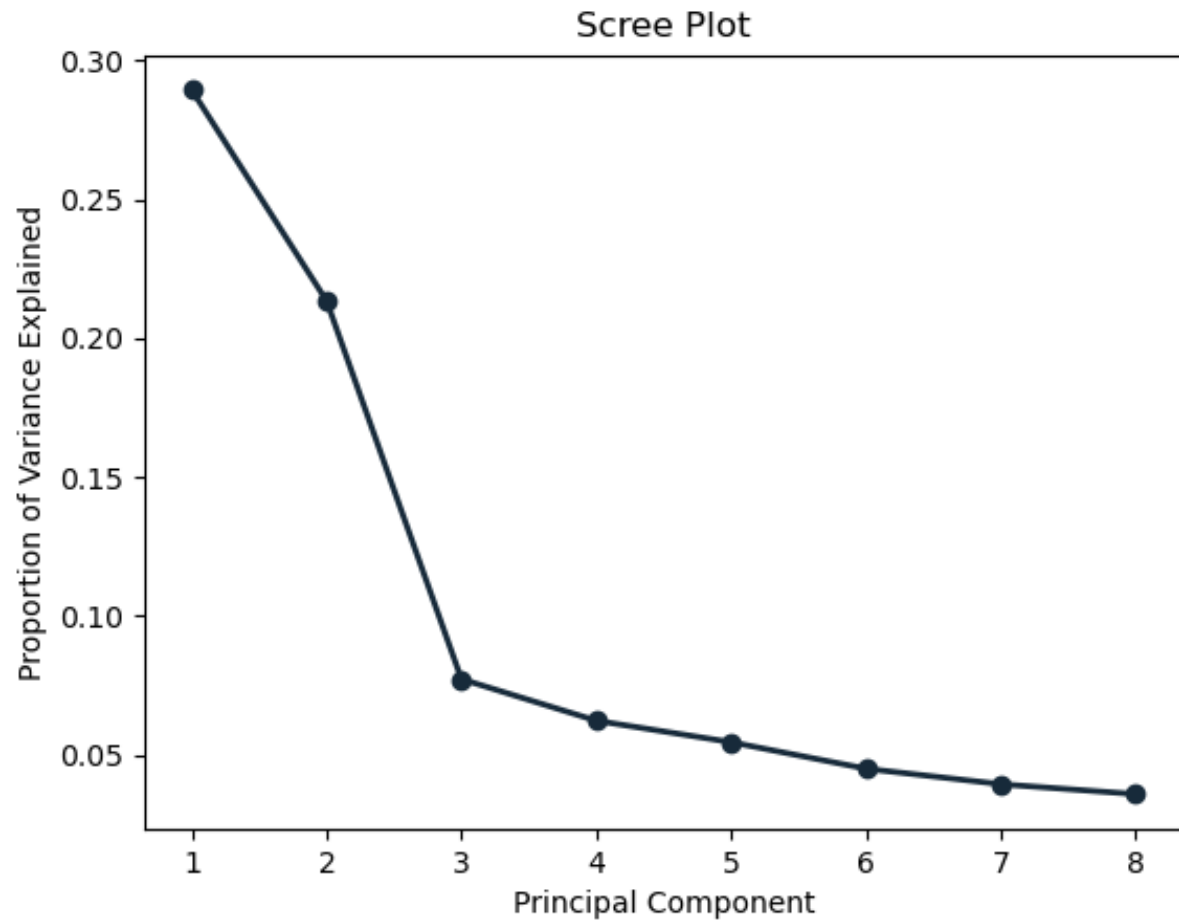
# Mortgage Application by Age

- The most common age to apply for a mortgage is between 25-34.
- Rejection and Acceptance follow the same pattern
- There is no age discrimination for mortgage applicants.





# Principle Component Analysis



Principal Component	Proportion of Variance Explained	Cumulative Proportion of Variance Explained
PC1	0.289	0.289
PC2	0.214	0.503
PC3	0.077	0.58
PC4	0.062	0.643
PC5	0.055	0.697
PC6	0.045	0.742
PC7	0.039	0.782
PC8	0.036	0.818



Modelling

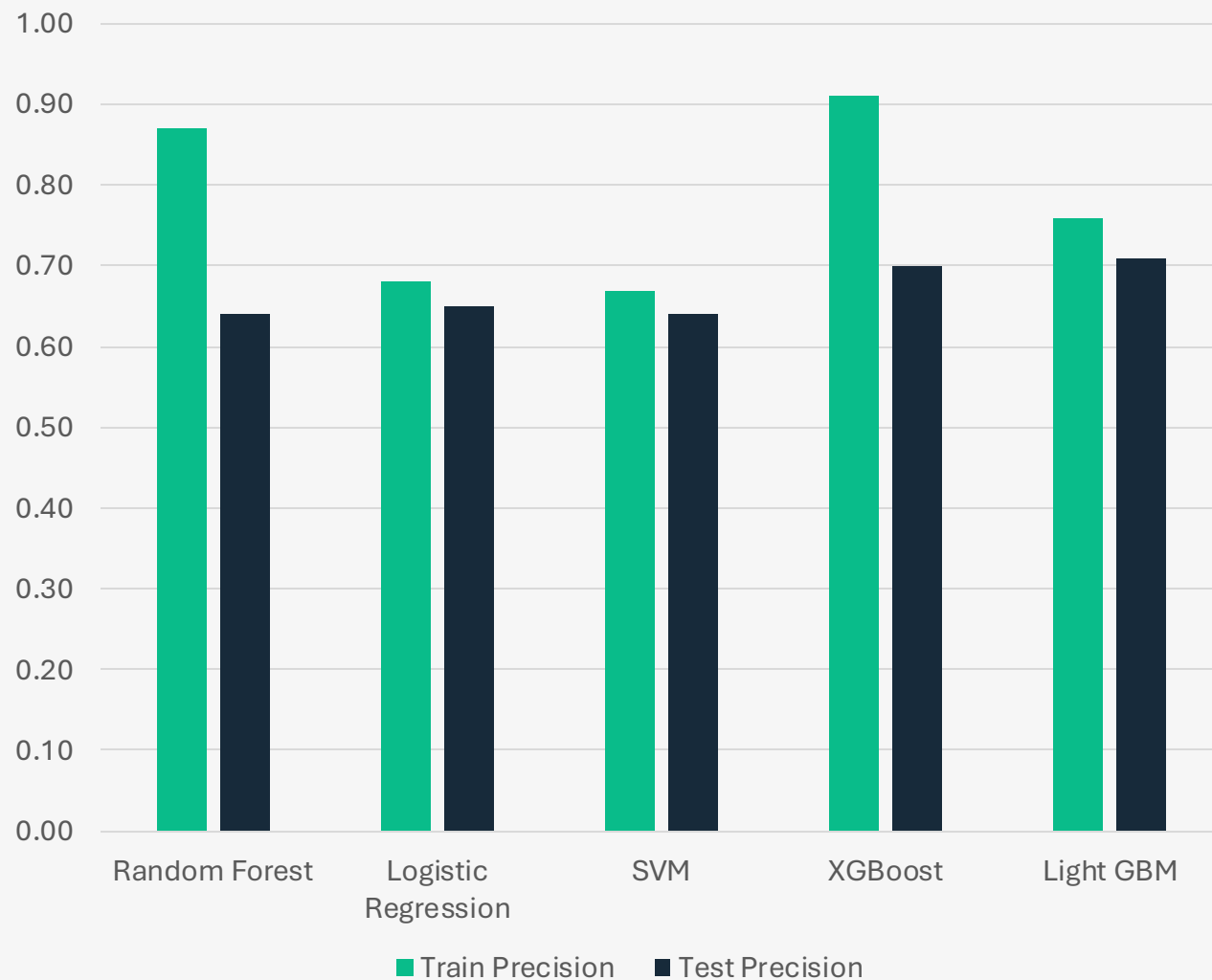


# Measuring model performance

- Precision, Recall, F1 score and ROC-AUC to measure the performance.
- **Precision** measures the accuracy of positive predictions.
  - This is important in my model as a false positive would mean giving someone a prediction of approval for a mortgage and then getting denied.
  - This would be the worst outcome for my model as customers could damage their credit cores with failed applications.
  - Furthermore, it would undermine customer trust in the model/function.



Prediction Scores of Production Models



## Modelling

These are the supervised learning classification models tested:

- Random Forest
- Logistic Regression
- Support Vector Machine
- eXtreme Gradient Boosting
- Light Gradient Boosting Machine

---

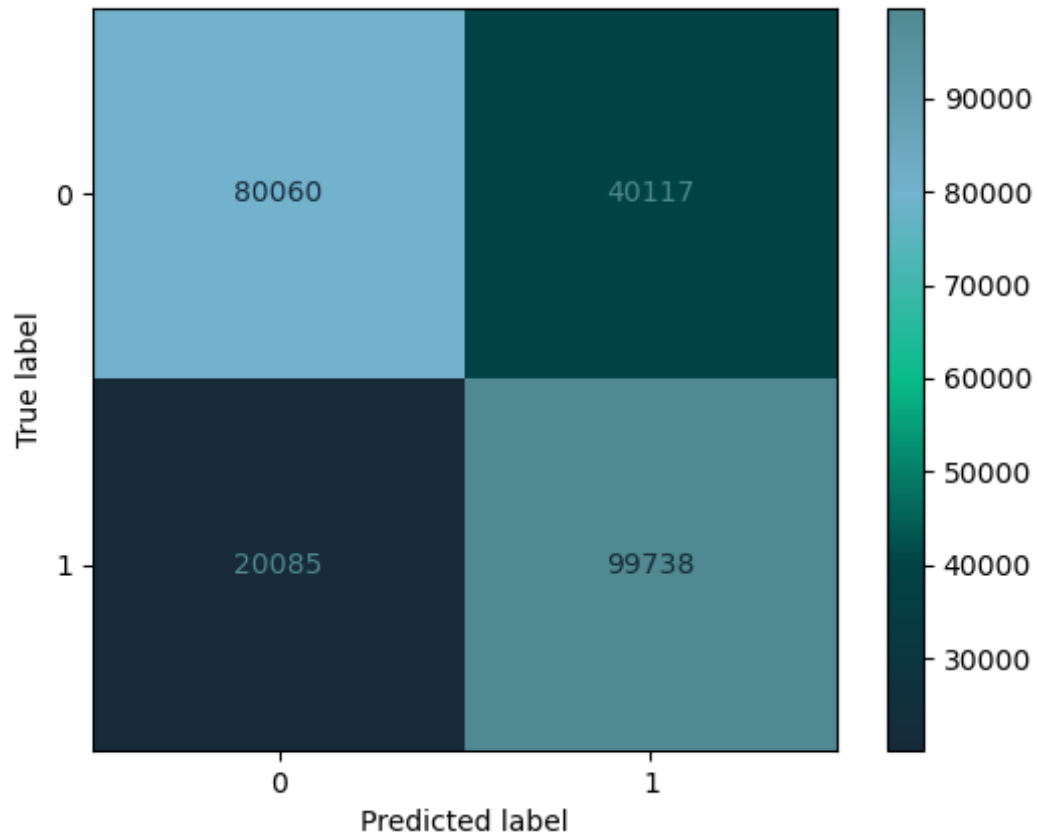
# Model Selection- LightGBM

- Gradient Boosting:
  - Builds models in a sequential manner where each new model corrects errors made by the previous ones.
- Leaf-wise Growth:
  - Unlike traditional boosting methods that grow trees level-wise, LightGBM grows trees leaf-wise. This means it splits the leaf with the maximum loss reduction, leading to a more complex and potentially more accurate model.
- Histogram-based Algorithm:
  - LightGBM uses a histogram-based algorithm to bucket continuous feature values into discrete bins, significantly speeding up the training process and reducing memory usage.





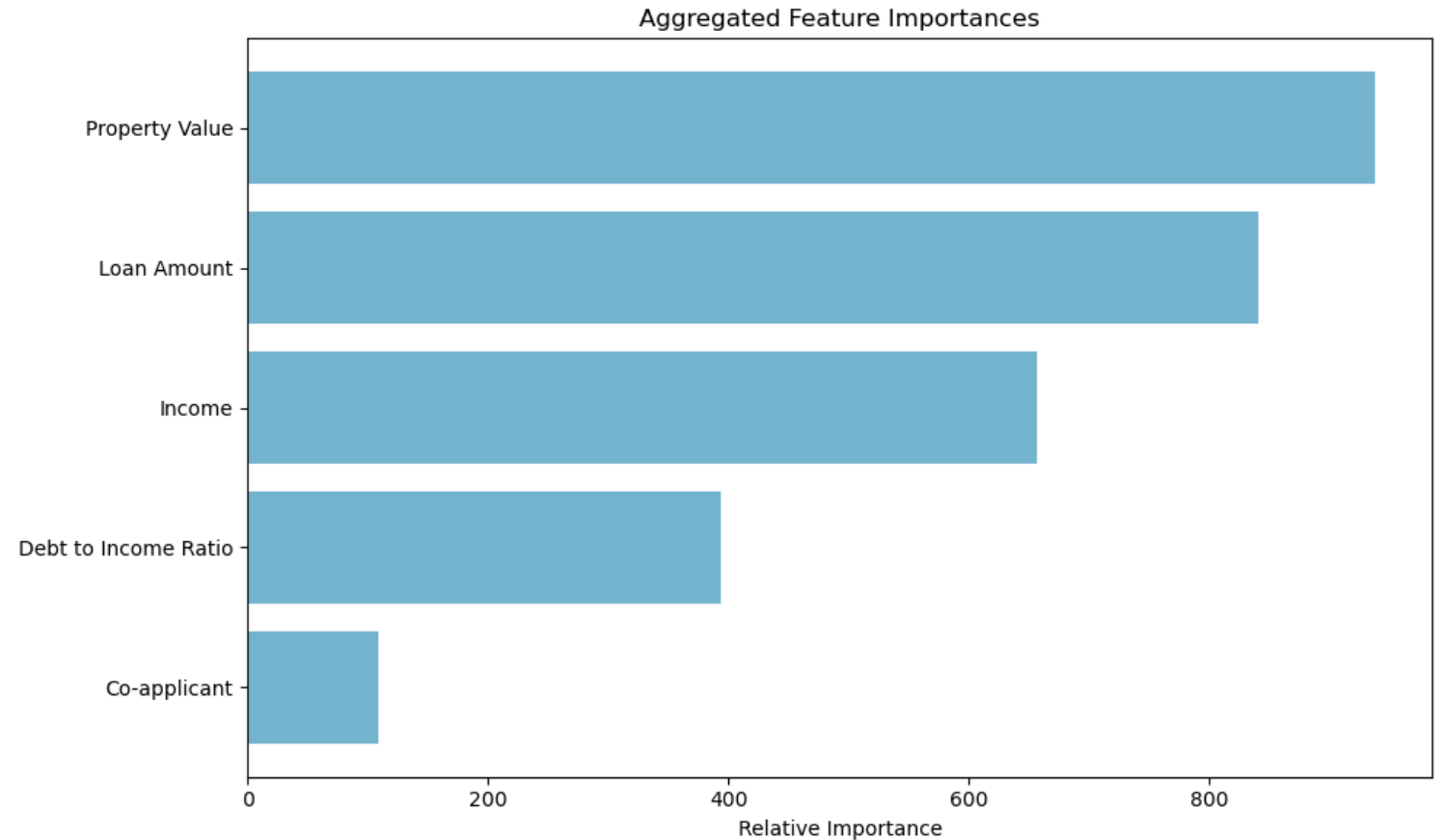
# Performance Evaluation



	Precision	Recall	f1-score
0	0.80	0.67	0.73
1	0.71	0.83	0.77
Accuracy			0.75
Macro - Avg	0.76	0.75	0.75
Weighted - Avg	0.76	0.75	0.75

# Model Transparency

- Model transparency is of utmost importance for this app
- Enables customers to understand the factors that made the decision.
- Property Value and Loan Amount have the largest effect on likelihood of mortgage approval.



The image shows four white, three-dimensional house models with red roofs, arranged in a diagonal line from the bottom-left towards the top-right. Each successive house is larger than the one before it, creating a sense of growth or progression. They are placed on a dark brown wooden surface with a visible grain. The text "Demo time..." is centered over the middle of the houses in a white, sans-serif font.

Demo time...



# Future Developments

- More time cleaning the data.
- Computational power for more model tuning.
- UK data.
- Information on macro lending decisions.



A close-up photograph of a person's hand holding a small, detailed model of a two-story house. The house has a brown roof, a chimney, a balcony with a metal railing, and several windows. The background is a blurred image of a person's torso wearing a light-colored shirt.

# Thank you for your time

Any questions?