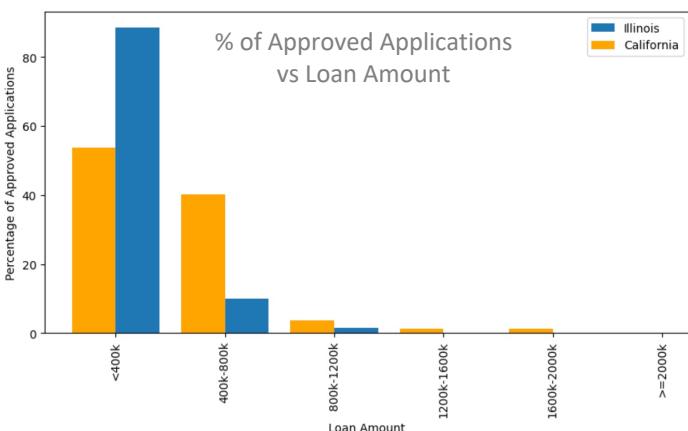
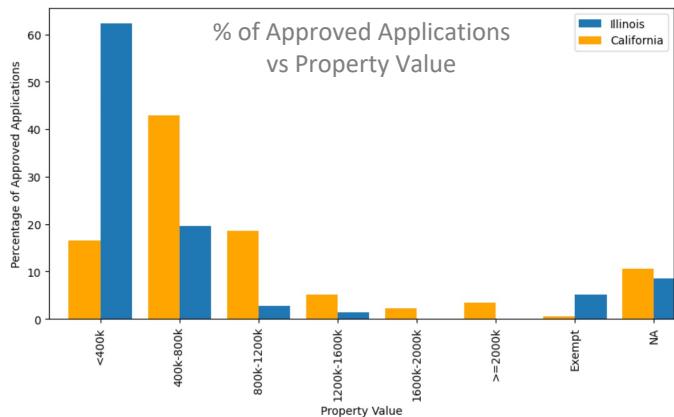


Illinois vs California

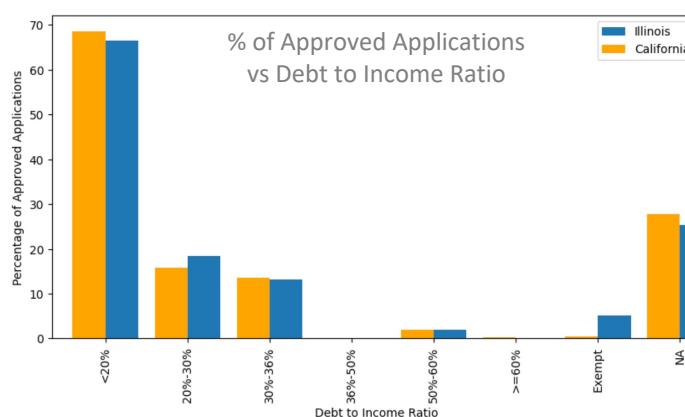
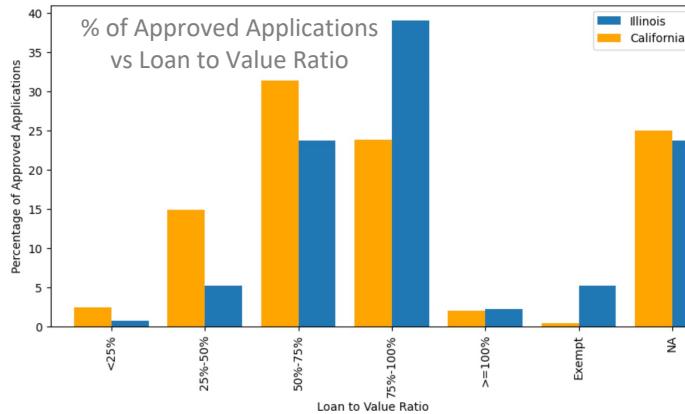
Loan Size & Collateral

- Property values:** substantially higher in California than in Illinois, reflecting structural differences in housing markets.
- Loan amounts:** higher in California than Illinois, consistent with higher underlying property values.



Leverage and Borrowers' Risk

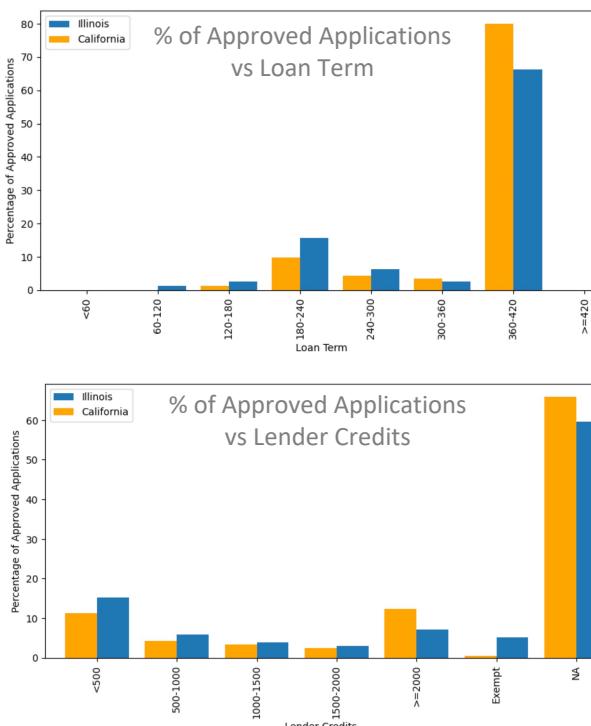
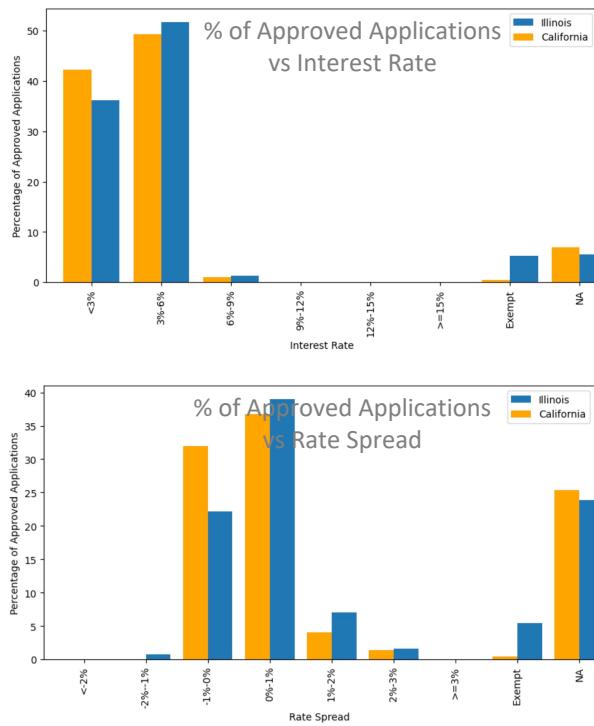
- Loan to value (LTV) ratios:** Approved loans in California exhibit higher LTV ratios than those in Illinois, indicating higher borrower leverage at origination.
- Debt to income ratios:** broadly similar between Illinois and California, suggesting comparable income-based affordability despite differences in housing prices.



Illinois vs California

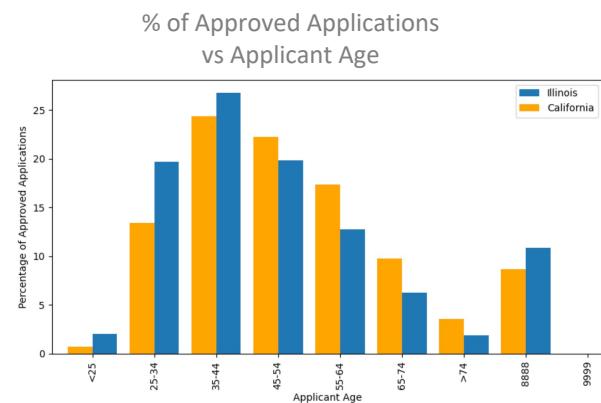
Product Pricing & Cost

- Interest rate & rate spread:** slightly higher for loans originated in Illinois than those in California, suggesting marginally higher borrowing costs and perceived risk for Illinois borrowers.
- Loan term:** longer for loans originated in California than those in Illinois, suggesting extended repayment periods.
- Lender credits:** more often exempt in Illinois and slightly higher for California borrowers.



Borrowers' Demographic

- Applicant age:** younger for loans originated in Illinois than those in California, suggesting Illinois borrowers were generally at earlier life stages.



Mortgage Application Prediction

Illinois 2020 Dataset

- **Goal:** Predict mortgage application decision
- **Data engineering:**
 - Removed withdrawn applications
 - Separated “Exempt” as distinct features
 - Transformed categorical variables using one-hot encoding
- **Algorithm:** XGBoost
 - Chosen for robustness to multicollinearity.
 - Tuned “scale_pos_weight” to address class imbalance.
- **Evaluation:**
 - Optimal decision threshold selected based on F1-score.
 - Performance:

Metrics	Train	Test
Accuracy	0.9957	0.9946
ROC-AUC	0.9914	0.9894
Precision	0.9981	0.9927
Recall	0.9969	0.9962
F1-score	0.9975	0.9970

- Top 10 features using absolute mean SHAP value:

Feature	SHAP value	Description
interest_rate	2.52	Higher interest rate increases probability of approval.
hoepa_status_3	1.31	Missing HOEPA high-cost mortgage status increases probability of approval.
rate_spread	0.76	Higher rate spread increases probability of approval.
hoepa_status_2	0.43	Loan not being a high-cost mortgage increases probability of approval.
loan_to_value_ratio	-0.34	Higher LTV ratio increases probability of rejection.
property_value	-0.19	Higher property value increases probability of rejection.
occupancy_type_3	-0.1	Investment property increases probability of rejection.
co-applicant_credit_score_type_10	-0.09	No co-applicant increases probability of rejection.
open-end_line_of_credit_2	-0.09	Not an open-end line of credit increases probability of rejection.
tract_to_msa_income_percentage	-0.09	Higher tract-to-MSA income ratio increases probability of rejection.