Fairness in Mortgage Lending

Mike Labadie

The Importance of Fair Lending

Benefits of Home Ownership

- Proximity to opportunities
- Wealth generation

Historic problems: Redlining of the 1930s

Civil rights movement of the 1960s lead to legislation to address problems in lending

Home Mortgage Disclosure Act (HMDA)

My analysis was limited to the District of Columbia

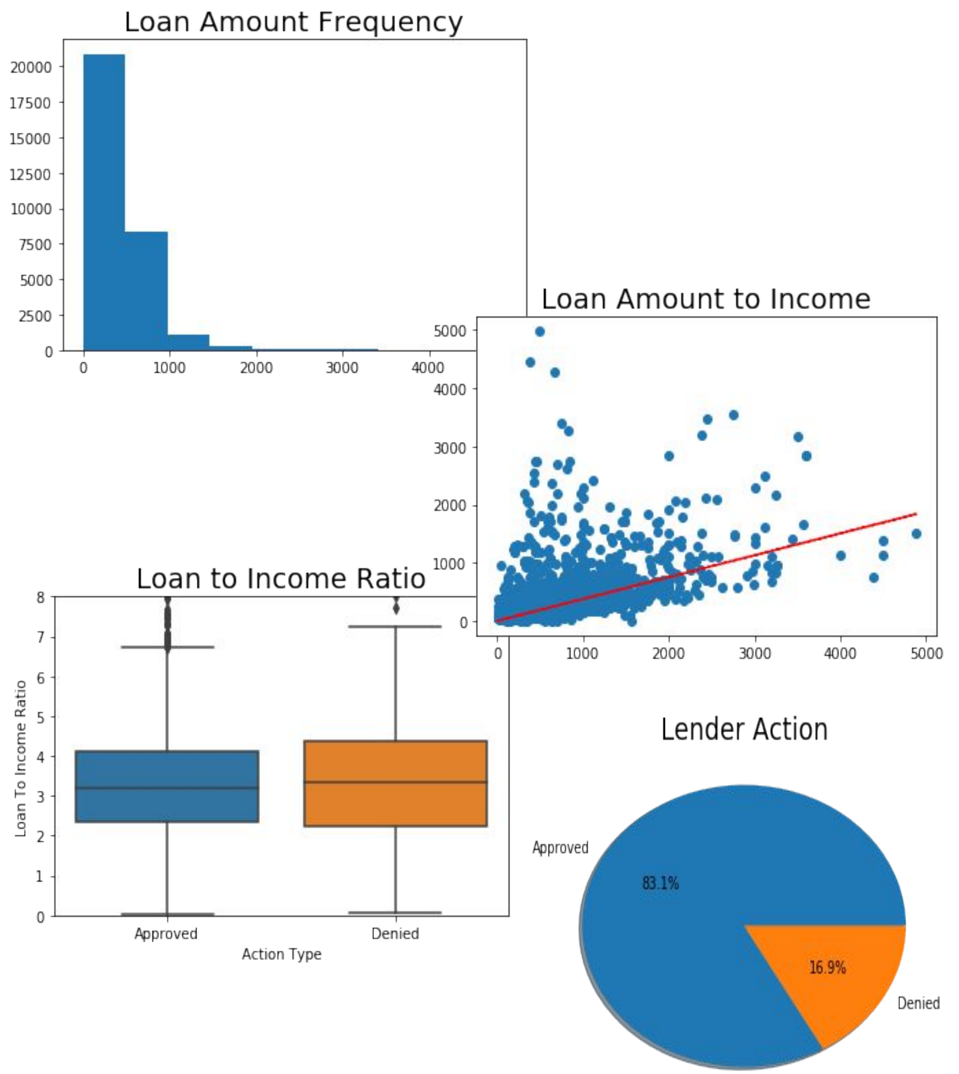
- Enacted in 1975 to increase transparency into lending practices
- Requires lenders (of a certain size) to report loan mortgage applications
- 2017 Dataset
  - Over 14 million records
  - Nearly 6,000 lenders
  - Nearly 50 fields describing borrower, application, location, etc

EDA

District of Columbia

- ~30k loan applications
- Applications mostly below $1mm
- Slightly higher loan to income ratio for denials; higher variability
- Heavy imbalance towards approvals

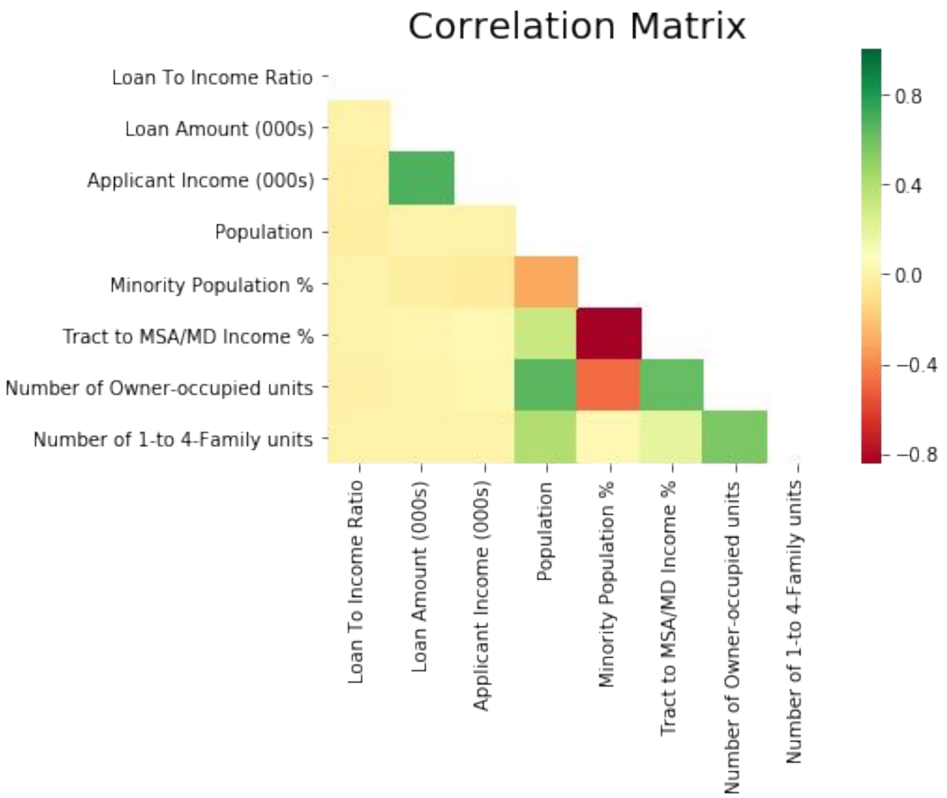
matplolib, seaborn



Correlations

Strong positive correlation between income and loan amount

Strong negative correlation between census tract minority population and income



seaborn, colormaps

Geographic Visualization

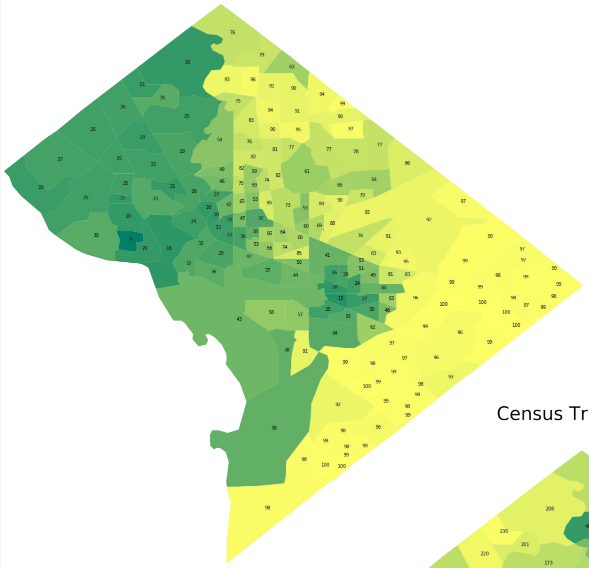
Visualization of census tract level data in D.C.

Top Left: Lighter region shows higher minority population

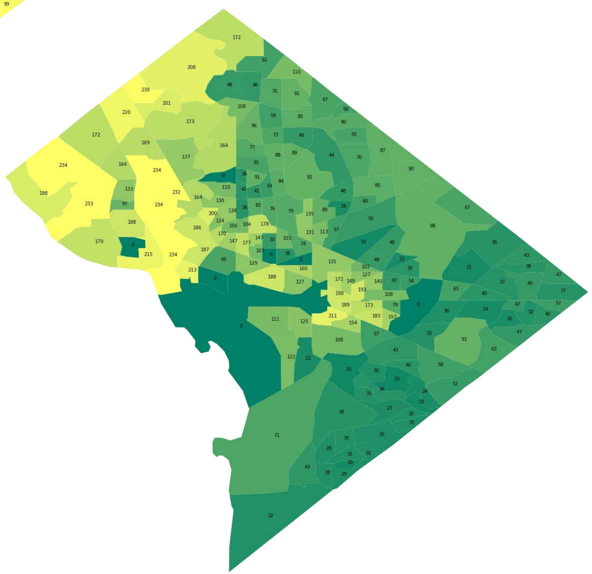
Bottom Right: Darker region show lower income level

pyshp, shapely, descartes, matplotlib

Minority Population % by Census Tract



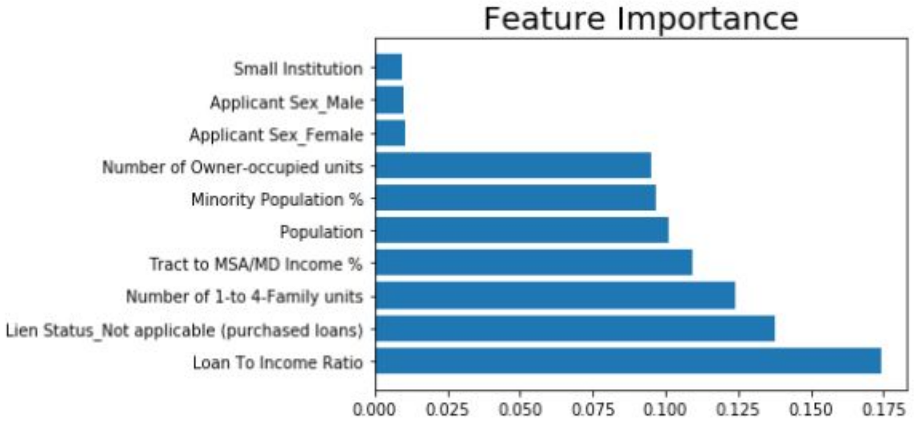
Census Tract to MSA Income Ratio

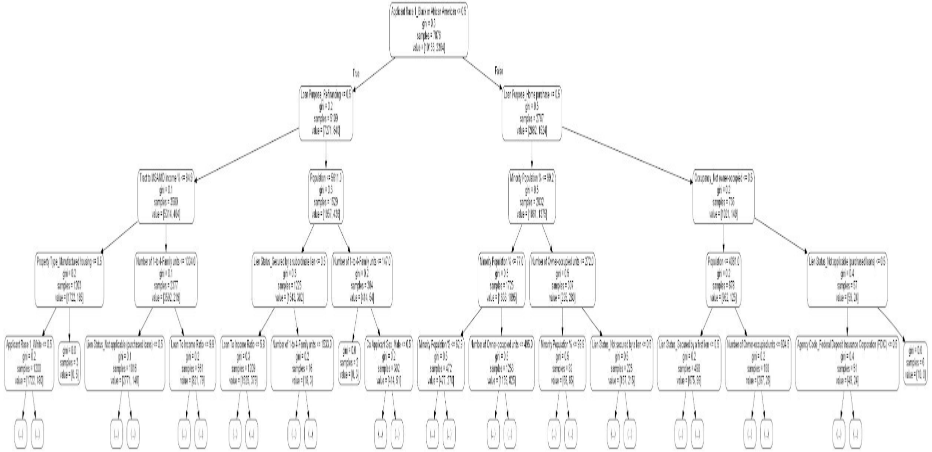


Random Forest

Loan to Income Ratio was the important feature

Bottom Right: example tree generated during the random forest process





sklearn, export_graphviz, pydot

Logistic Regression

Lien Status and Loan to Income Ratio among the largest weights

sklearn

	0
Lien Status_Secured by a first lien	2.111928
Lien Status_Not secured by a lien	1.980851
Loan To Income Ratio	1.718018
Lien Status_Secured by a subordinate lien	1.448096
Loan Purpose_Refinancing	1.308184
Agency Code_Consumer Financial Protection Bureau (CFPB)	1.238898
Applicant Race 1_Black or African American	1.221188
Loan Purpose_Home improvement	1.188488
Occupancy_Not owner-occupied	1.076439
HOEPA Status_Not a HOEPA loan	1.070377
Loan Type_FHA-insured (Federal Housing Administration)	1.057503
Applicant Race 1_American Indian or Alaska Native	1.054636
Co Applicant Sex_No co-applicant	1.048064
Co Applicant Race 1_No co-applicant	1.048064
Loan Type_VA-guaranteed (Veterans Administration)	1.040858
Co Applicant Race 1_American Indian or Alaska Native	1.023040
Applicant Sex_Male	1.021492
Minority Population %	1.020019
Property Type_Manufactured housing	1.015992
Population	1.013053
Co Applicant Race 1_Asian	1.007369
Co Applicant Sex_Male	1.000883
FFIEC Median Family Income	1.000000
Co Applicant Race 1_Black or African American	0.995697
Small Institution	0.994034
Applicant Race 1_Native Hawaiian or Other Pacific Islander	0.992735
Property Type_One to four-family (other than manufactured housing)	0.984260
Applicant Sex_Female	0.978960
Number of 1-to 4-Family units	0.977451
Co Applicant Race 1_Native Hawaiian or Other Pacific Islander	0.968469
Applicant Race 1_Asian	0.964263
Agency Code_Federal Deposit Insurance Corporation (FDIC)	0.961202
Co Applicant Race 1_White	0.950527
Agency Code_National Credit Union Administration (NCUA)	0.949695
Co Applicant Sex_Female	0.947243

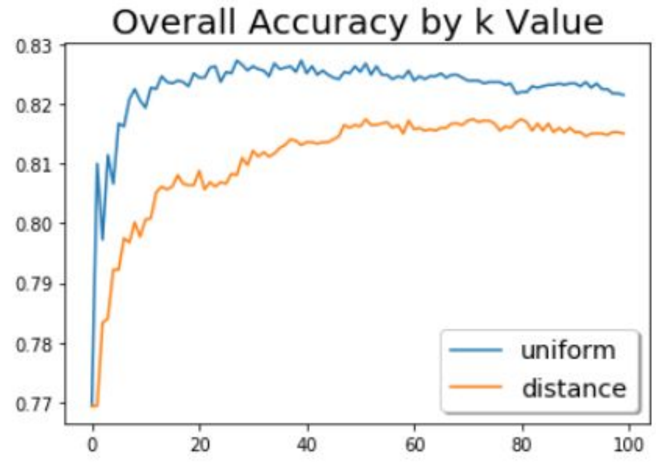
k-Nearest Neighbors

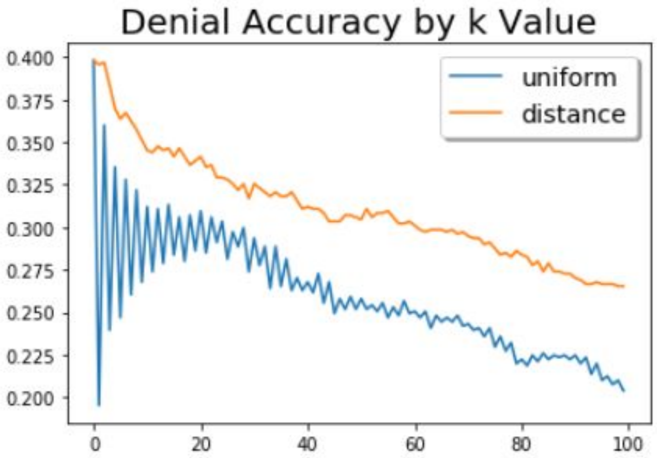
Simple algorithm to classify unseen samples based on similarity to some number (k) of seen samples

Results dependent upon parameters:

- k
- Weighting of Neighbors

sk-learn





Classification Model Results

k-Nearest Neighbors

Predicted / Actual	Approved	Denied
Approved	3005	354
Denied	513	301

Overall Accuracy: 79%
Denials Accuracy: 37%

Random Forest

Predicted / Actual	Approved	Denied
Approved	2615	744
Denied	277	537

Overall Accuracy: 76%
Denials Accuracy: 66%

Logistic Regression

Predicted / Actual	Approved	Denied
Approved	2435	924
Denied	187	627

Overall Accuracy: 73%
Denials Accuracy: 77%

Challenge: How to deal with class imbalance?

- Answer: adjust class weights

Linear Regression

Goal: Identify particular institutions with skeptical lending patterns to minority census tracts

Steps:

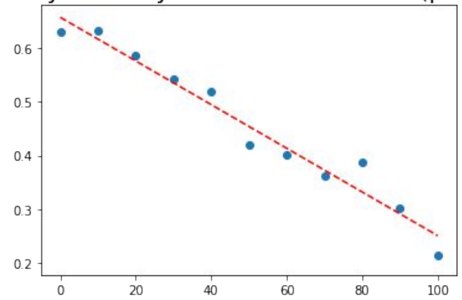
- Group by institution and minority population % (bucketed @ 10%)
- Calculate approval rates for each bucket
- Linear regression on Minority % and Approval Rate

Left: Low p-value indicate a statistically significant relationship

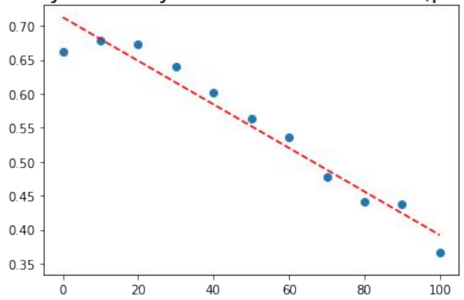
scipy

Three Questionable Lenders

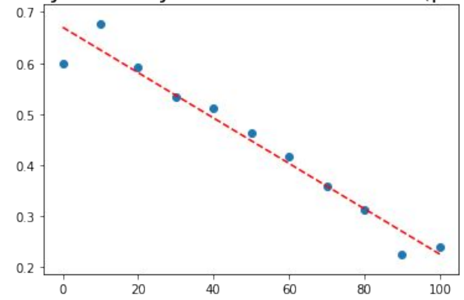
Approval Rate by Minority % for 0000723112 (p-value:1.30E-07)



Approval Rate by Minority % for 0000233031 (p-value:2.76E-07)



Approval Rate by Minority % for 0000712504 (p-value:3.01E-07)



Lending Practices Findings

Good:

- Loan to Income Ratio appears to be key in lending decision as seen in classification modeling

Bad:

- Clear demographic divides with respect to race, income, and housing location
- Identified questionable lenders as it relates to minority neighborhoods

Next Steps:

- Need more data (FICO scores, Loan to Value, Debt to Income)
- Combine with other data sources: census data