This Folder is the HMDA research starting January 2024

#LAR Data Description

#https://ffiec.cfpb.gov/documentation/publications/loan-level-datasets/lar-data-fields

#TS Data Description

#https://ffiec.cfpb.gov/documentation/publications/loan-level-datasets/ts-data-fields

#2019-2021 HMDA Data

#https://ffiec.cfpb.gov/data-publication/one-year-national-loan-level-dataset/2021

#2022 HMDA Data

#https://ffiec.cfpb.gov/data-publication/snapshot-national-loan-level-dataset/2022

#Community bank list

#https://www.fdic.gov/resources/community-banking/cbi-data.html

#Credit union list

#https://catalog.data.gov/dataset/ncua-active-federally-insured-credit-unions-list-59c68

#CDFI list

#https://www.cdfifund.gov/sites/cdfi/files/2021-04/CDFI%20Cert%20List%2004-14-2021%20Final.xlsx

#Traditional Bank list and fintech list from Bushak et. al.

https://sites.google.com/view/fintech-and-shadow-banks

Trad Bank and Fintech List (from [Bushak et al](https://sciencedirect.com/science/article/abs/pii/S0304405X1830237X))

https://sites.google.com/view/fintech-and-shadow-banks

#Climate Data

2020 census tract level data: for HMDA data 2022

<https://resilience.climate.gov/datasets/FEMA::national-risk-index-census-tracts/about>

Climate data glossery: <https://hazards.fema.gov/nri/data-resources>

2010 Census tract level data for HMDA data 2019, 2020, 2021

<https://hazards.fema.gov/nri/data-archive>

In 2022 we use RISK\_SCORE and in 2019-2021 we use RISK\_NPCTL ([I think these are the same](https://hazards.fema.gov/nri/understanding-scores-ratings)).

We use data from [HMDA](https://www.consumerfinance.gov/data-research/hmda/) to examine how home mortgage data from MDIs compared to other bank types

1. Data houses all the public data
2. Data intermediate: holds the cleaned data for the summary statistics
3. Data Output: holds all the summary statistics and excel sheets that create the graphs and tables
4. Graphs: houses all graphs created
5. Tables: houses all tables as HTMLs PDF and JPGs
6. Outmoded: Previous versions of the study
7. Final Product: Houses the final draft to be published online

In the main folder is the code

1. 0. HMDA TS Fuzzy Matching.py – imports and fuzzy matches HMDA TS (name/bank identifying information) data held in 0.Data and exports the matched data to 1. Intermediate
2. 1. HMDA LAR matched banks to loan.py -- takes the matched data from 1. Data Intermediate then matches the TSLEI (bank identifier) to the LAR (loan level data) LEI, we do this by bank type, MDI, CDFI, Community Bank, Trad Bank, Fintech
3. 2. Non MDI Separate.py – does the same as 1. HMDA Lar but for all non-MDI banks
4. 3. Summary Statistics.py – get summary statistics for loans of MDI, CDFI, and community banks (have to separate out non-mdi, trad banks, and fintechs as they’re too large for this file
5. 4. NonMDI Summary Statistics.py – does the same as 3. Summary Statistics but for non-MDI
6. 5. Trad Banks.py - does the same as 3. Summary Statistics but for trad banks
7. 6. Fintechs.pydoes the same as 3. Summary Statistics but for fintechs

In [2. Data Output](https://nationalbankers-my.sharepoint.com/:f:/g/personal/cromer_nationalbankers_org/Eq5F7iGL1gZDh6gjWs9Sro8BpwqPIFNNg0OcPEVj-PrSjA?e=EiRNLf) the summary statistics are held.

1. Overall – Overall summary statistics, total loans, loan dollars and median loan by year
2. By bank –Total loans, loan dollars and median loan by bank by year
3. By race – Total loans, loan dollars and median loan by race by year
4. By bank race – Total loans, loan dollars and median loan by MDI race ownership by year
5. By bank race by race– Overall summary statistics, total loans, loan dollars and median loan by loan by MDI race ownership by race by year

Convo 1/16/24

Is there an association between HMDA and asset size? Who loans?   
Wealth from housing in climate-vulnerable communities is precarious -- we can pull in from literature (But also do a little bit of analysis of this analysis -- 1st street foundation has some publicly available data, also could pay for Moody's)   
How does DEI/ESG backlash impact this story -- we can look at how Insurance companies are refusing to insure markets and how climate is impacting property taxes and emergency services cost (a nonpartisan fiscal concern)/health of community -- not AS relevant for initial publication

Convo 1/30/24

And then after that, I'd say the priority is getting the list of lending locations and matching to the climate data from First Street since that it will be the heart of the climate analysis. Once we have all those assembled with directionality of findings, I can start drafting the broad strokes of the report itself, while we backfill with any additional HMDA (refinancing rates, etc.) based on time, etc.

Convo: 2/1/24

Look at location data of loans (match to census tract data: poverty, % minority) \*DONE

Look at get a list of community banks and hopefully match that, top 5-10 fintechs and top 5-10 banks \* DONE

2/16/24

Next steps: keep working on presenting climate data

Work on model for how we’re going to show MDI rejection vs non-MDI rejection maybe rejection by MDI type by race?

Present differences in rejection reasons \* DONE

2/22/24

I’d like to use logistic regression and interact MDI and Race categories.

There are a lot of missing data, I’d like to use multiple imputation to fix this, but I don’t think it’s realistic given our time constraints. I’ll just drop data that is incomplete (I looked and it does not look like its disproportionately MDIs that are missing data – may be disproportionately Fintechs)

2/27/24

Need to go back to Climate and figure out how to present that

3/4/22

Results

==============================================================================

Dep. Variable: Denial No. Observations: 4649412

Model: Logit Df Residuals: 4649340

Method: MLE Df Model: 71

Date: Mon, 04 Mar 2024 Pseudo R-squ.: -inf

Time: 12:41:25 Log-Likelihood: -inf

converged: False LL-Null: -1.7865e+06

Covariance Type: nonrobust LLR p-value: 1.000

=====================================================================================================

coef std err z P>|z| [0.025 0.975]

-----------------------------------------------------------------------------------------------------

const -2.8742 0.010 -293.515 0.000 -2.893 -2.855

Hispanic 0.4160 0.005 90.704 0.000 0.407 0.425

Black 0.7125 0.005 147.052 0.000 0.703 0.722

AAPI 0.3165 0.007 47.980 0.000 0.304 0.329

AIAN 0.7794 0.014 54.032 0.000 0.751 0.808

NoRace 0.3223 0.005 69.322 0.000 0.313 0.331

MDI -1.0396 0.050 -20.921 0.000 -1.137 -0.942

BlackMDI 0.1950 0.099 1.961 0.050 0.000 0.390

HispanicMDI 1.0513 0.060 17.472 0.000 0.933 1.169

AAPIMDI 0.6604 0.087 7.619 0.000 0.490 0.830

AIANMDI 1.1772 0.102 11.490 0.000 0.976 1.378

NoRaceMDI 0.2786 0.112 2.492 0.013 0.059 0.498

income 3.995e-06 9.69e-07 4.124 0.000 2.1e-06 5.89e-06

income2 -4.69e-13 1.59e-13 -2.958 0.003 -7.8e-13 -1.58e-13

property\_value -5.035e-07 6.51e-09 -77.354 0.000 -5.16e-07 -4.91e-07

debt\_to\_income\_ratio 0.0651 0.000 388.923 0.000 0.065 0.065

tract\_to\_msa\_income\_percentage -0.0091 4.27e-05 -213.309 0.000 -0.009 -0.009

tract\_minority\_population\_percent -0.0096 7.27e-05 -132.427 0.000 -0.010 -0.009

AK -1.1832 0.041 -28.761 0.000 -1.264 -1.103

AL 0.0151 0.012 1.215 0.225 -0.009 0.039

AR -0.0385 0.015 -2.510 0.012 -0.069 -0.008

AZ -0.5925 0.013 -46.006 0.000 -0.618 -0.567

CA -0.4081 0.011 -38.462 0.000 -0.429 -0.387

CO -0.8176 0.014 -59.572 0.000 -0.844 -0.791

CT -0.8839 0.019 -46.726 0.000 -0.921 -0.847

DC -0.7631 0.048 -15.941 0.000 -0.857 -0.669

DE -0.5657 0.025 -22.186 0.000 -0.616 -0.516

FL -0.3019 0.010 -31.548 0.000 -0.321 -0.283

GA -0.4031 0.011 -36.214 0.000 -0.425 -0.381

GU -1.6823 0.263 -6.399 0.000 -2.198 -1.167

HI -0.5904 0.037 -16.080 0.000 -0.662 -0.518

IA -0.9865 0.020 -49.257 0.000 -1.026 -0.947

ID -1.0507 0.024 -44.544 0.000 -1.097 -1.005

IL -0.6982 0.012 -58.173 0.000 -0.722 -0.675

IN -0.8220 0.014 -60.702 0.000 -0.849 -0.795

KS -0.8183 0.021 -38.624 0.000 -0.860 -0.777

KY -0.1141 0.014 -8.431 0.000 -0.141 -0.088

LA 0.1552 0.013 11.984 0.000 0.130 0.181

MA -1.0097 0.017 -60.996 0.000 -1.042 -0.977

MD -0.8693 0.015 -58.356 0.000 -0.898 -0.840

ME -0.8576 0.027 -31.357 0.000 -0.911 -0.804

MI -0.6017 0.012 -51.341 0.000 -0.625 -0.579

MN -1.2413 0.017 -73.503 0.000 -1.274 -1.208

MO -0.8672 0.015 -59.245 0.000 -0.896 -0.839

MS 0.3292 0.015 22.102 0.000 0.300 0.358

MT -0.6030 0.029 -20.561 0.000 -0.660 -0.545

NC -0.4672 0.011 -42.153 0.000 -0.489 -0.446

ND -1.1047 0.041 -26.723 0.000 -1.186 -1.024

NE -1.1979 0.028 -42.608 0.000 -1.253 -1.143

NH -0.9736 0.028 -34.560 0.000 -1.029 -0.918

NJ -0.6570 0.014 -48.050 0.000 -0.684 -0.630

NM 0.0577 0.019 3.106 0.002 0.021 0.094

NV -0.5113 0.017 -29.403 0.000 -0.545 -0.477

NY -0.5312 0.012 -45.083 0.000 -0.554 -0.508

OH -0.9229 0.012 -75.328 0.000 -0.947 -0.899

OK -0.2780 0.015 -17.977 0.000 -0.308 -0.248

OR -0.8910 0.017 -53.456 0.000 -0.924 -0.858

PA -0.8572 0.012 -69.530 0.000 -0.881 -0.833

PR 0.5882 0.025 23.828 0.000 0.540 0.637

RI -1.0911 0.035 -31.352 0.000 -1.159 -1.023

SC -0.1371 0.012 -11.589 0.000 -0.160 -0.114

SD -1.1478 0.039 -29.520 0.000 -1.224 -1.072

TN -0.4252 0.012 -34.770 0.000 -0.449 -0.401

TX -0.1388 0.010 -14.338 0.000 -0.158 -0.120

UT -0.8924 0.018 -50.766 0.000 -0.927 -0.858

VA -0.8115 0.013 -62.926 0.000 -0.837 -0.786

VI -9.6535 40.725 -0.237 0.813 -89.474 70.167

VT -0.8155 0.043 -18.781 0.000 -0.901 -0.730

WA -0.9293 0.014 -67.680 0.000 -0.956 -0.902

WI -1.0690 0.017 -64.530 0.000 -1.101 -1.037

WV 0.0116 0.020 0.592 0.554 -0.027 0.050

WY -0.7741 0.040 -19.125 0.000 -0.853 -0.695

=====================================================================================================

C:\Users\csromer\anaconda3\Lib\site-packages\statsmodels\discrete\discrete\_model.py:1819: RuntimeWarning: overflow encountered in exp

return 1/(1+np.exp(-X))

Results

==============================================================================

Dep. Variable: Denial No. Observations: 4528469

Model: Logit Df Residuals: 4528451

Method: MLE Df Model: 17

Date: Mon, 04 Mar 2024 Pseudo R-squ.: -inf

Time: 12:59:08 Log-Likelihood: -inf

converged: True LL-Null: -1.6637e+06

Covariance Type: nonrobust LLR p-value: 1.000

=====================================================================================================

coef std err z P>|z| [0.025 0.975]

-----------------------------------------------------------------------------------------------------

const -3.8308 0.009 -412.446 0.000 -3.849 -3.813

Hispanic 0.4430 0.005 96.256 0.000 0.434 0.452

Black 0.7368 0.005 151.225 0.000 0.727 0.746

AAPI 0.2926 0.007 43.683 0.000 0.279 0.306

AIAN 0.8035 0.015 55.184 0.000 0.775 0.832

NoRace 0.3387 0.005 70.737 0.000 0.329 0.348

MDI -0.9093 0.050 -18.331 0.000 -1.006 -0.812

BlackMDI 0.2624 0.099 2.645 0.008 0.068 0.457

HispanicMDI 1.3461 0.059 22.670 0.000 1.230 1.463

AAPIMDI 0.5472 0.086 6.327 0.000 0.378 0.717

AIANMDI 1.1908 0.103 11.597 0.000 0.990 1.392

NoRaceMDI 0.1740 0.112 1.559 0.119 -0.045 0.393

income 5.376e-06 9.42e-07 5.708 0.000 3.53e-06 7.22e-06

income2 -6.42e-13 1.87e-13 -3.438 0.001 -1.01e-12 -2.76e-13

property\_value -5.899e-07 6.37e-09 -92.533 0.000 -6.02e-07 -5.77e-07

debt\_to\_income\_ratio 0.0720 0.000 404.317 0.000 0.072 0.072

tract\_to\_msa\_income\_percentage -0.0083 4.28e-05 -193.336 0.000 -0.008 -0.008

tract\_minority\_population\_percent -0.0068 6.53e-05 -104.882 0.000 -0.007 -0.007

=====================================================================================================