Mortgage Interest Rates

Exploratory Data Analysis (EDA)

Background

Project Proposal

Think of all the people out there paying more interest on their mortgages than they need to.

Their stories are waiting to be told.

Objectives

- Analyze mortgage interest rate and demographics data
- Gain insights into how different customer classes, loan types, and demographics affect mortgage interest rates
- Enhance financial literacy, promote informed decision-making, and contribute to a more equitable mortgage market

Mortgage Data

Exploration

- Federal Housing Finance Agency
  https://www.fhfa.gov/
- Zillow API
  https://www.zillowgroup.com/developers/

- Redfin
  https://www.redfin.com/news/data-center/
- Freddie Mac API
  https://developer.freddiemac.com/public/#/

Exploration - HMDA

https://ffiec.cfpb.gov/

The Home Mortgage Disclosure Act (HMDA) requires many financial institutions to maintain, report, and publicly disclose loan-level information about mortgages.

HMDA was originally enacted by Congress in 1975.

Project Scope: Minnesota from 2018 to 2022

HMDA Data Collection

https://ffiec.cfpb.gov/documentation/category/developer-apis

1279023 rows × 99 columns

```python
def download_hmda_data(fd, state, year):
    url = f'https://ffiec.cfpb.gov/v2/data-browser-api/view/csv?states={state}&years={year}'

    try:
        with requests.get(url, stream=True) as response:
            response.raise_for_status()
            with open(output_file, 'wb') as fd:
                for chunk in response.iter_content(chunk_size=1024):
                    if chunk:
                        fd.write(chunk)
        return True

    except requests.exceptions.RequestException as e:
        print(f"Error downloading data: {e}")
        return False
```

HMDA Data Cleaning

```python
# remove null values and 'Exempt' interest rate from dataframe

df = unclean_df[unclean_df['interest_rate'].notnull()]
df = df.query('interest_rate != "Exempt"')
df = df.query('loan_to_value_ratio != "Exempt"')
```

```python
# remove null loan terms from dataframe

df = df[df['loan_term'].notnull()]
```

```python
# data type conversions

df['interest_rate'] = pd.to_numeric(df['interest_rate'], errors='raise')
df['loan_to_value_ratio'] = pd.to_numeric(df['loan_to_value_ratio'], errors='raise')
```

```python
# rename values

df['derived_race'] = df['derived_race'].replace({
    'Black or African American': 'Black',
    'American Indian or Alaska Native': 'Native',
    'Native Hawaiian or Other Pacific Islander': 'Pacific Islander'
})
```

Analysis: Understanding the data

```python
for c in df.columns:
    print(c)
```

```
activity_year
lei
derived_msa-md
state_code
county_code
census_tract
conforming_loan_limit
derived_loan_product_type
derived_dwelling_category
derived_ethnicity
derived_race
derived_sex
action_taken
purchaser_type
preapproval
loan_type
loan_purpose
lien_status
```

```python
for c in df.columns:
    print(f'Examining column: {c}')
    print(df[c].value_counts())
    print()
```

```
Examining column: derived_race
derived_race
White                     882643
Race Not Available        288832
Asian                      47390
Black                      30486
Joint                      23764
Native                      3857
Pacific Islander            1106
2 or more minority races     879
Free Form Text Only           66
Name: count, dtype: int64

Examining column: derived_sex
derived_sex
Joint              498530
Male               323859
Sex Not Available  236510
```

	A	B	C	D	E
	HMDA Data Dictionary				
	File Edit View Insert Format Data Tools Extensions Help				
A1	Column Name				
1	Column Name	Keep or Drop	Summary	Actions	Sample Value Counts
2	activity_year	Keep			
3	lei	Keep	Lending Instance	Translate	
4	derived_msa-md	Unknown			
5	state_code	Keep	State		
6	county_code	Keep	County		
7	census_tract	Unknown			
8	conforming_loan_limit		C, NC, or U		
9	derived_loan_product_type	Keep	Loan Type		Conventional:First Lien 86546 FHA:First Lien 5521 VA:First Lien 4681 Conventional:Subordinate Lien 2102 FSA/RHS:First Lien 1149
10	derived_dwelling_category	Keep	Building Type		Single Family (1-4 Units):Site-Built 98532 Single Family (1-4 Units):Manufactured 1168 Multifamily:Site-Built 290 Multifamily:Manufactured 9
11	derived_ethnicity	Keep			
12	derived_race	Keep			
13	derived_sex	Keep			
14	action_taken	Unknown	Unknown	Translate	1 57984 6 20342 4 10440 3 6790 5 3569 2 681 7 108 8 85
15	purchaser_type	Unknown	Unknown	Translate	0 42546 3 26519 1 20176 2 4789 71 2133 9 1803 6 1664 5 369
16	preapproval	Unknown	Unknown	Translate	2 99593 1 406
17	loan_type	Keep	Unknown	Translate	1 88648 2 5521 3 4681 4 1149 21 25450

Analysis: Statistics Summaries

```python
# Statistics Summary Table - Interest Rates by Race
race_group = df.groupby('derived_race')

summary_table = pd.DataFrame({
    "Mean Interest Rate": race_group['interest_rate'].mean(),
    "Median Interest Rate": race_group['interest_rate'].median(),
    "Interest Rate Variance": race_group['interest_rate'].var(),
    "Interest Rate Std. Dev.": race_group['interest_rate'].std(),
    "Interest Rate Std. Err.": race_group['interest_rate'].sem()
})

summary_table
```

derived_race	Mean Interest Rate	Median Interest Rate	Interest Rate Variance	Interest Rate Std. Dev.	Interest Rate Std. Err.
2 or more minority races	3.947778	3.625	2.205659	1.485146	0.050093
Asian	3.687475	3.375	1.426056	1.194176	0.005486
Black	3.820617	3.500	10.695898	3.270458	0.018731
Free Form Text Only	4.121212	3.750	2.153253	1.467397	0.180624
Joint	3.803048	3.500	1.382663	1.175867	0.007628
Native	3.943490	3.625	2.024991	1.423022	0.022913
Pacific Islander	3.903947	3.625	1.771620	1.331022	0.040023
Race Not Available	4.013744	3.625	425.640704	20.631062	0.038388
White	3.788169	3.500	63.506752	7.969112	0.008482

Analysis: Hypothesis Testing

ANOVA Tests

```python
# Extract individual groups
group0 = df[df["derived_race"].str.fullmatch('White')]["interest_rate"]
group1 = df[df["derived_race"].str.fullmatch('Race Not Available')]["interest_rate"]
group2 = df[df["derived_race"].str.fullmatch('Asian')]["interest_rate"]
group3 = df[df["derived_race"].str.fullmatch('Joint')]["interest_rate"]
group4 = df[df["derived_race"].str.fullmatch('Black')]["interest_rate"]
group5 = df[df["derived_race"].str.fullmatch('Native')]["interest_rate"]
group6 = df[df["derived_race"].str.fullmatch('2 or more minority races')]["interest_rate"]
group7 = df[df["derived_race"].str.fullmatch('Pacific Islander')]["interest_rate"]
group8 = df[df["derived_race"].str.fullmatch('Free Form Text Only')]["interest_rate"]

# Perform the ANOVA test
stats.f_oneway(group0, group1, group2, group3, group4, group5, group6, group7, group8)
```
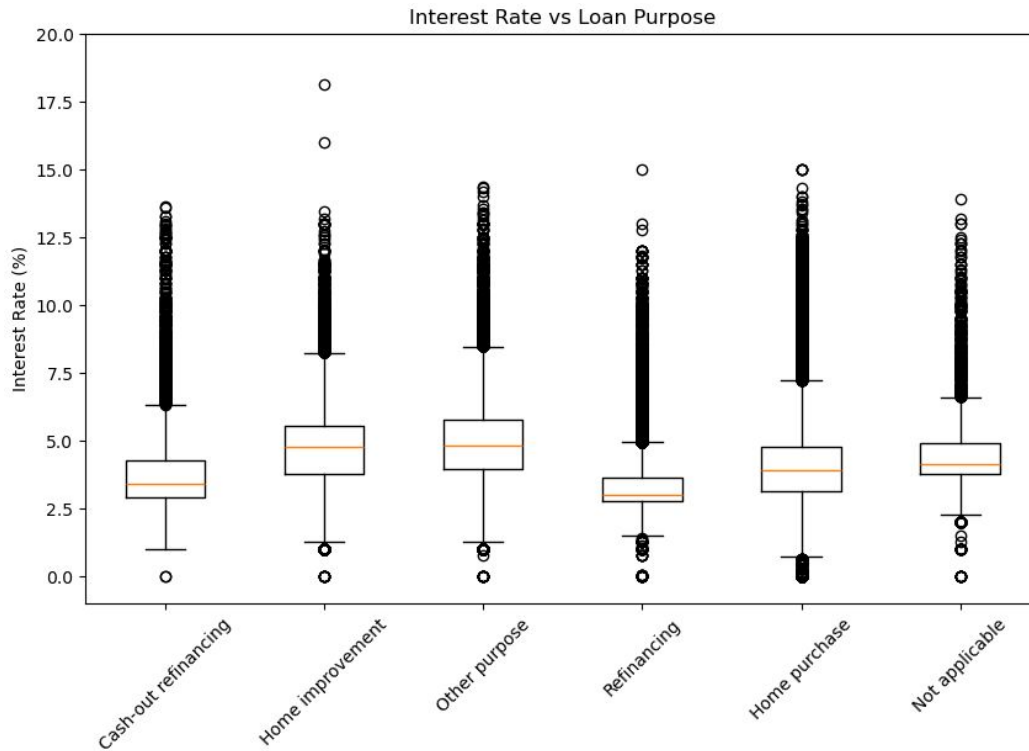
F_onewayResult(statistic=10.948260161132048, pvalue=1.4375699707682176e-15)

Key Observations

Interest Rate vs Loan Purpose

Impact of Loan Purpose

Home Refinancing interest rates are lower than initial Home Purchase interest rates ... which in turn are lower than Home Improvement interest rates.

Home Refinancing interest rate "box" is tighter than the Home Purchase and Home Improvement boxes ... suggesting more controlled or less risky.

Regarding outliers ... they tend heavily to be negative in nature , to the disadvantage of the borrower.

Impact of Income Groups
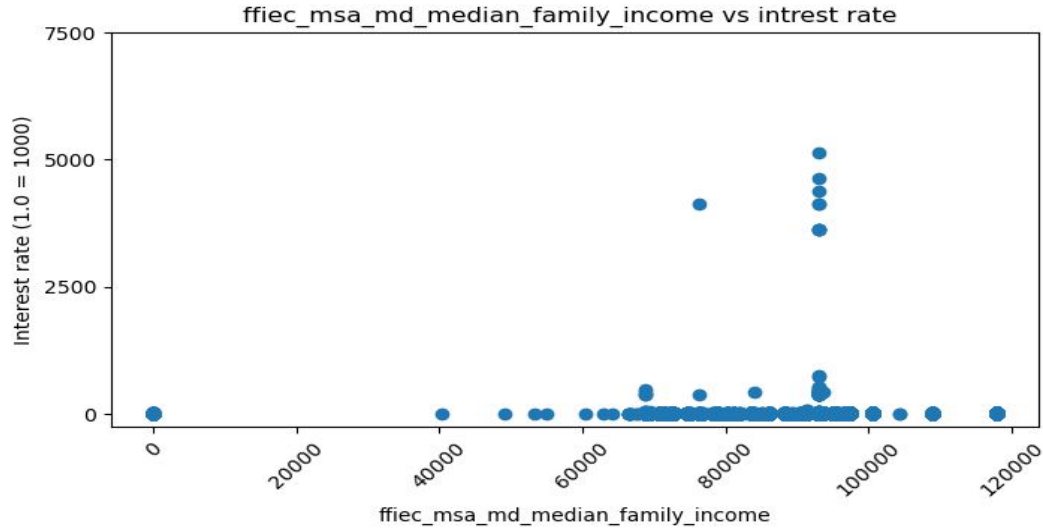
ffiec_msa_md_median_family_income vs intrest rate

Understanding the relationship between age, income, and interest rates is crucial for financial planning. While age seems to be a more consistent predictor, the unexpected spikes warrant further exploration. As we move forward, let's delve deeper into these patterns to make informed decisions.

Notably, there were outlier spikes in interest rates at approximately $70,000 and again at $90,000.

Further investigation into these spikes could provide valuable insights.

Income Impact:

- Surprisingly, income did not significantly affect interest rates during the period from 2018 to 2022.
- This suggests that other factors may play a more dominant role in determining interest rates.

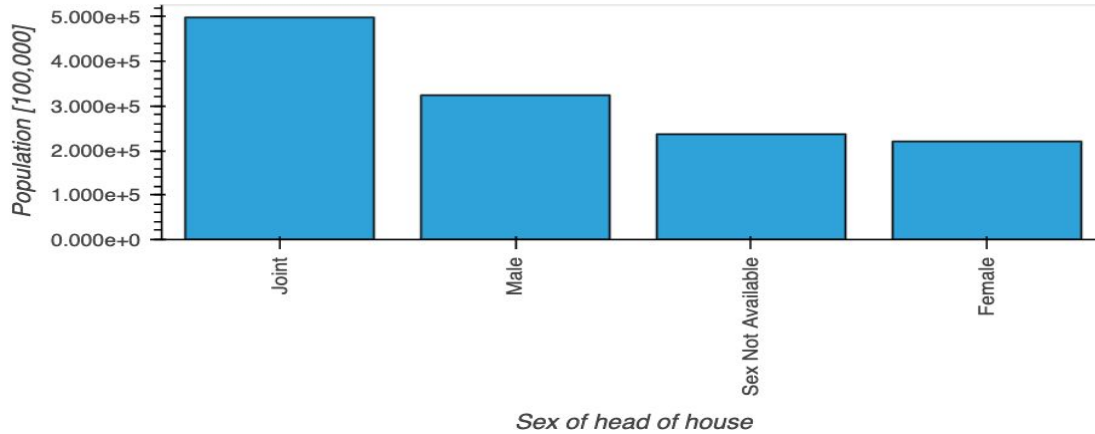# Impact of Age
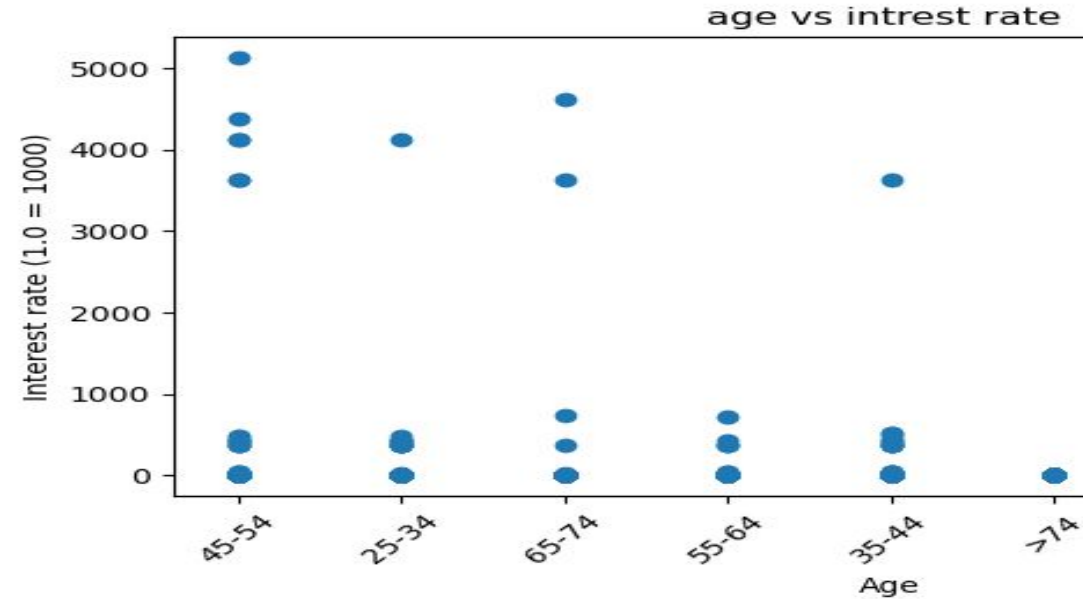
age vs intrest rate

Understanding the relationship between age, income, and interest rates is crucial for financial planning. While age seems to be a more consistent predictor, the unexpected spikes warrant further exploration. As we move forward, let's delve deeper into these patterns to make informed decisions.

Age Groups and Consistency:

- The age group 55-64 exhibited the most consistent interest rates across the board.
- In contrast, the 45-54 age group had the greatest number of outliers in interest rates.
- Understandably, the >74 age group experienced the lowest interest rates overall.

Bar Graph for derived_sex

derived_sex vs interest_rate

Impact of Gender

Gender Groups and Consistency:

- Joint ownership is the most common form of mortgage ownership, with 498,530 cases.
- Interestingly, joint owners tend to have higher interest rates compared to other ownership types.
- Female mortgage owners are the least represented group, with 220,124 individuals.

Bar Graph for derived_race


derived_race vs interest_rate
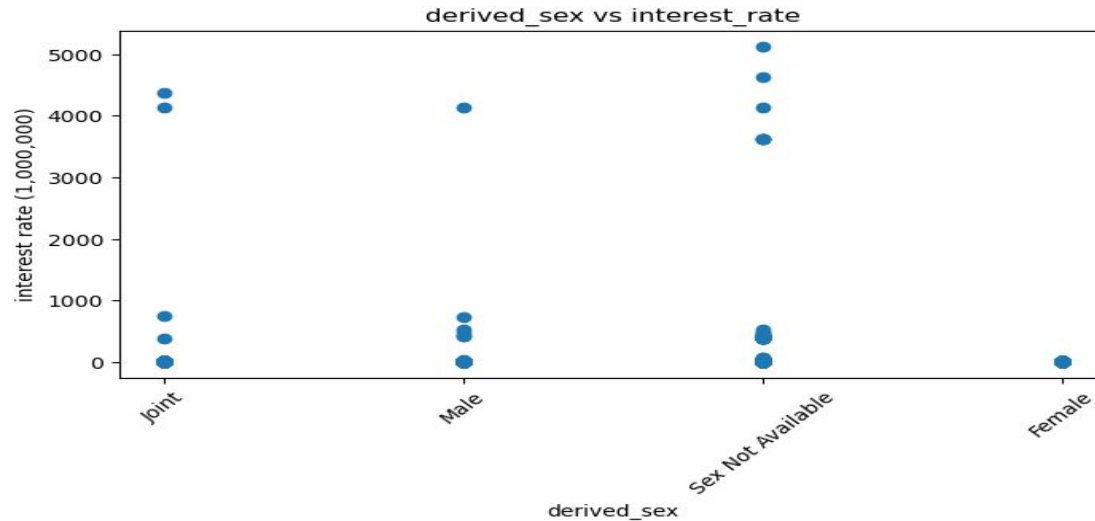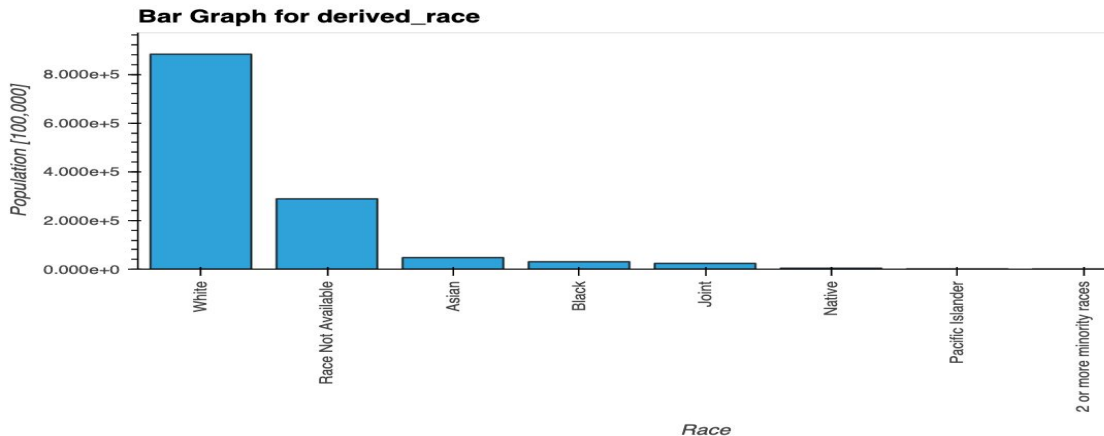
Impact of Ethnicity

Race Groups and Consistency:

- White mortgage owners constitute the largest group, with a total of 882,643 individuals.
  - Other racial groups follow, but none come close to the White population in terms of mortgage ownership.
- Among all racial groups, Whites exhibit the highest peaks in interest rates.
  - Further investigation into the factors driving these differences would be valuable.
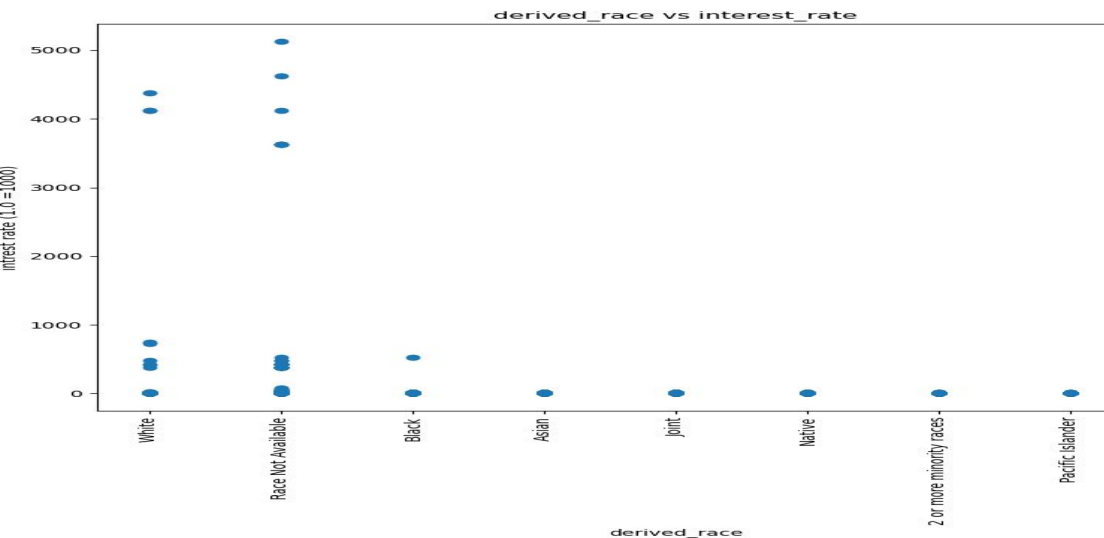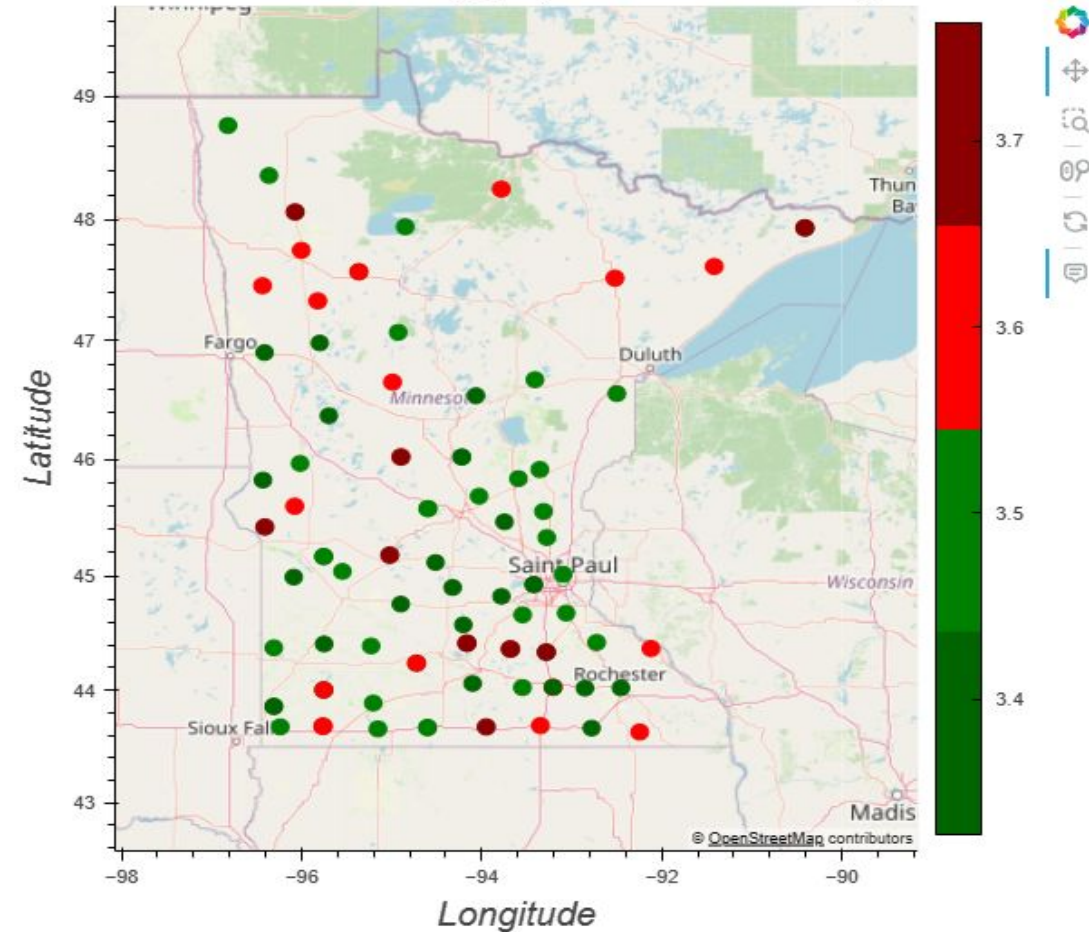
Impact of Property Location

Observations:

- Twin Cities Metro Area has favorable interest rates.
- Outstate Areas are more likely to have less favorable interest rates.

Lei	Count	Interest rates
549300CN50N3250U7V79	27	12.036111
254900JI5JVEBN2QKM97	1	10.200000
549300YX5RPSSMR88K92	6	10.166667
254900O2VYWY7LKZ6T60	90	10.076944
254900SJDUBD5DVVWZ69	1	10.000000
549300F570XAWYN8Z237	2	2.312500
549300FK3AFCFVAPH234	1	2.250000
549300Y0F8X17ADZK505	1	2.250000
549300MPGZVO0YIGL418	1	2.125000
549300JGMQJ4R419LR70	12	1.967225
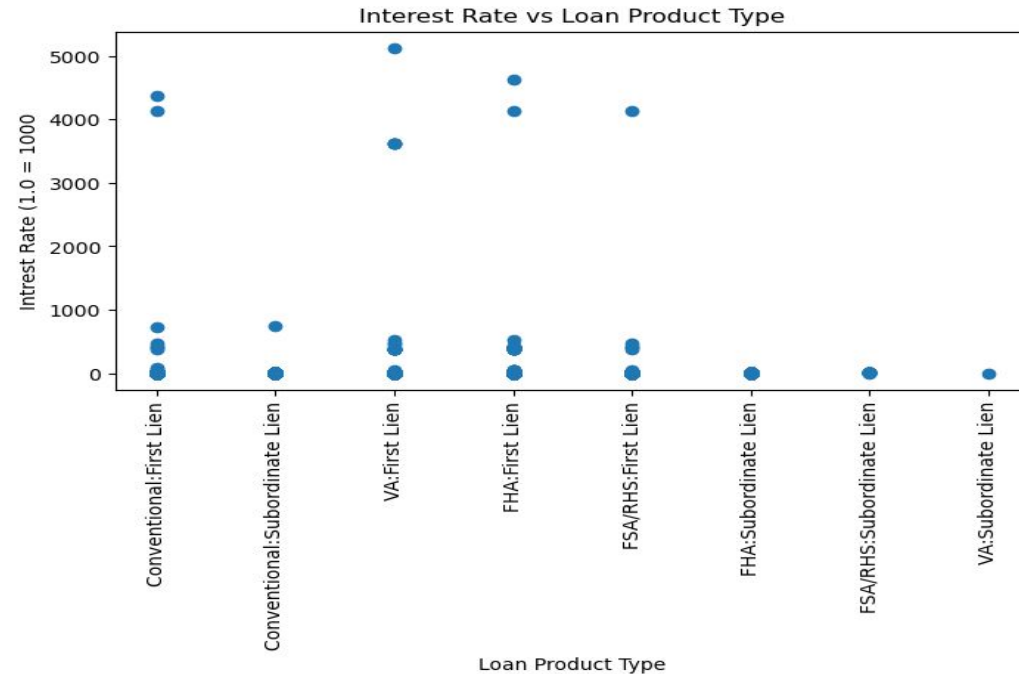
Impact of Lending Practices by Lender

Interest Rate vs Loan Product Type

Observations:

- Lei institute 549300JGMQJ4R419LR70 had the lowest average interest rate given out at   1.967225
- FHA:Subordinate lien, FSA/RHS: Subordinate Lien and VA: Subordinate Lien have the lowest interest based on loan type
- Conventional: First Lien at the highest average.

Summary

Observations and Implications

The key observations of the exploration include:

1. Interest rates are higher for first-time homebuyers than for refinancers
2. Older borrowers tend to have lower interest rates than younger borrowers
3. Demographics impact interest rates
4. Geographic disparities - higher rates in rural areas than in urban areas
5. Loan type and lending institution can impact interest rates

Lessons Learned: Class Concepts in the Real World

- Applied class concepts to real-world data, resulting in a successful project
- Project provided valuable hands-on experience and learnings:
  - Data collection
  - Importance of documentation
  - Scope creep

Thank you!

Project data and analysis available on GitHub:
https://github.com/PonchoBeBallin/Group_project_Data_Class