

The main aim of this kernel is to answer some very common questions that arise while applying for loans using data.

The Home Mortgage Disclosure Act (HMDA) in America requires many financial institutions to maintain, report, and publicly disclose loan-level information about mortgages. These public data are important because they help show whether lenders are serving the housing needs of their communities; they give public officials information that helps them make decisions and policies; and they shed light on lending patterns that could be discriminatory.

Let's see a video to get a better view.

Code

Out[1]:



for a loan at her bank.

The dataset we have here is for all the mortgages in Washington during the year 2016. I'll analyze this dataset to find patterns in the lending process. It's indeed an interesting dataset, which has so far been left unexplored.

Code

```
['washington-state-home-mortgage-hdma2016', '2016-state-county-fips-codes']
```

```
[3]: df=pd.read_csv('../input/washington-state-home-mortgage-hdma2016/Washington_State_HDMA-2016.csv')# load the dataset
```

```
/opt/conda/lib/python3.6/site-packages/IPython/core/interactiveshell.py:3049: DtypeWarning:
Columns (23,24,25,28,29,30,38,39,40) have mixed types. Specify dtype option on import or set
low_memory=False.
```

A peak into the data.

```
In [4]: df.head()
```

Out[4]:

	tract_to_msamd_income	rate_spread	population	minority_population	number_of_owner_occupied_units	number_of_1_to_4_f
0	121.690002	NaN	8381.0	23.790001	2175.0	2660.0
1	83.370003	NaN	4915.0	23.990000	1268.0	1777.0
2	91.129997	NaN	5075.0	11.820000	1136.0	1838.0
3	146.169998	NaN	5032.0	8.590000	1525.0	1820.0
4	162.470001	NaN	5183.0	10.500000	1705.0	2104.0

```
In [5]: df['action_taken_name'].value_counts()
```

Out[5]:

```
Loan originated                263712
Application denied by financial institution    64177
Application withdrawn by applicant    60358
Loan purchased by the institution    48356
File closed for incompleteness    18176
Application approved but not accepted    11735
Preapproval request denied by financial institution    35
Preapproval request approved but not accepted    17
Name: action_taken_name, dtype: int64
```

Loan originated is a widely used term in finance , it means the loan application has been approved. We will remove the rows where the applicant has withdrawn the request for loan. Loan purchased means that the lender bought the loan on a secondary market.

What is a secondary market ?

It happens quite often that after lending the loan , the lender sells the loan and servicing rights to an investor in the secondary market.You might be thinking what is the need of a secondary market.

When a person takes out a home loan, the loan is underwritten, funded and serviced by a bank. Because the bank has used their own funds to make the loan, they will eventually run out of money to loan, so they will sell the loan to the secondary market to replenish their money available to make more home loans.

Since our analysis only concerns the primary market where borrowers and lenders are involved , I'll remove the rows where the action taken is "Loan purchased by financial institution".

```
In [6]: df=df[df['action_taken_name']!="Application withdrawn by applicant"]
df=df[df['action_taken_name']!="Loan purchased by the institution"]
```

Each county has a unique FIPS or county code that is useful in plotting the county map. I'll upload an additional dataset to add this feature.

```
In [7]: fips=pd.read_excel("../input/2016-state-county-fips-codes/all-geocodes-v2016.xlsx",converters=
{'County Code (FIPS)': lambda x: str(x)})
fips=fips[fips["State Code (FIPS)"]==53]# state code for washington
fips['county_code']=fips['State Code (FIPS)'].astype(str).str.cat(fips['County Code (FIPS)'],as
type(str))
fips=fips.drop(labels=['Summary Level','State Code (FIPS)','County Code (FIPS)','County Subdivi
sion Code (FIPS)','Place Code (FIPS)','Consolidated City Code (FIPS)'],axis=1)
fips.columns=['county_name','county_code']
```

```
In [8]: df=pd.merge(df,fips,how="left",on="county_name",sort=False)#merging it into original dataset
```

[Code](#)

Which county has the highest number of loan applications?

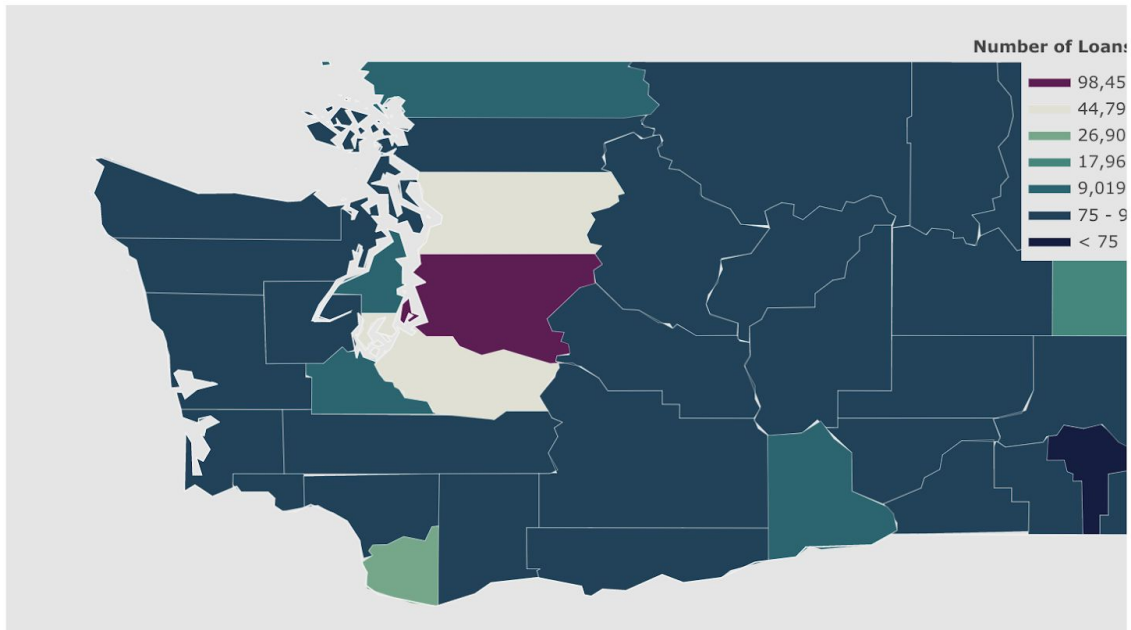
[Code](#)

```
/opt/conda/lib/python3.6/site-packages/pandas/core/frame.py:6211: FutureWarning:
```

```
Sorting because non-concatenation axis is not aligned. A future version
of pandas will change to not sort by default.
```

```
To accept the future behavior, pass 'sort=False'.
```

```
To retain the current behavior and silence the warning, pass 'sort=True'.
```



King County has the highest number of loan applications while Garfield County has the lowest. Now, we will add a new feature representing approved for originated loan and not approved for everything else.

King County has the highest number of loan applications while Garfield County has the lowest. Now, we will add a new feature representing approved for originated loan and not approved for everything else.

```
In [11]: df['loan_status'] = ["approved" if x=="Loan originated" else "not approved" for x in df['action_taken_name']]
```

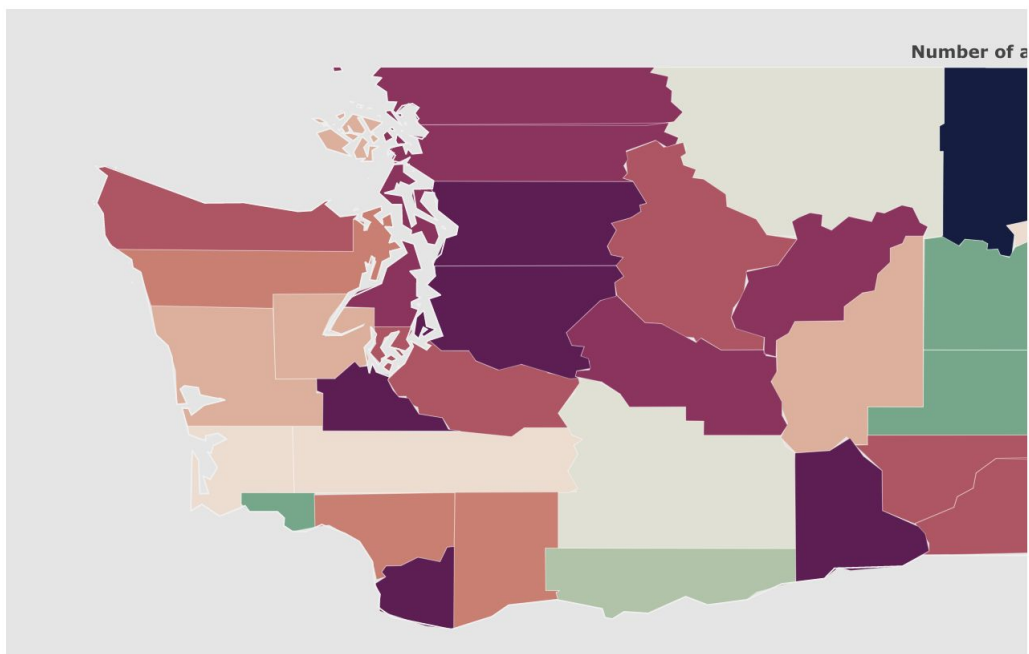
Which county has the highest approval rate?

```
In [12]: df_approved = df[df['loan_status'] == 'approved']
df_notapproved = df[df['loan_status'] == 'not approved']
```

Code

In [14]:

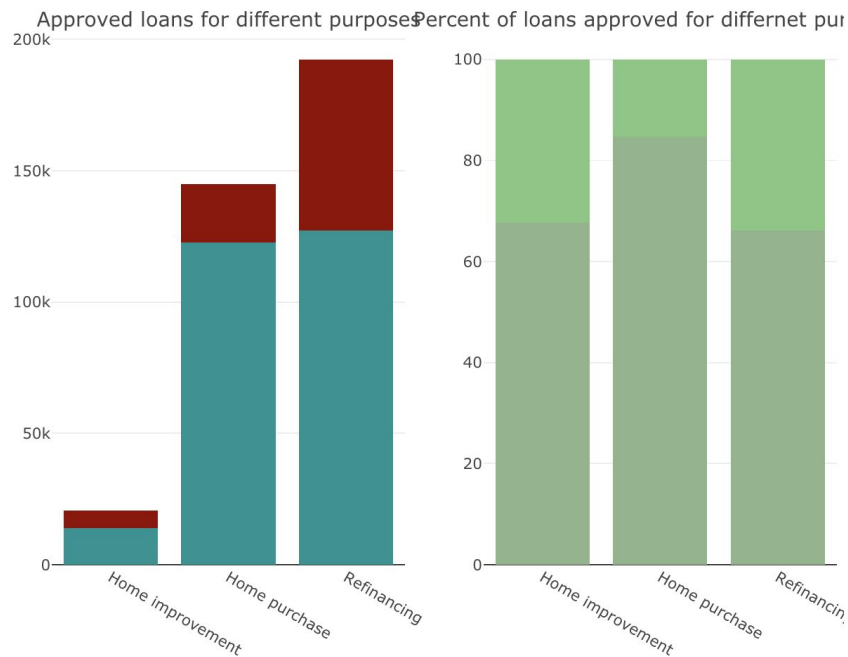
```
fips=county2['county_code'].tolist()
values=county2['approval_rate'].tolist()
endpts = list(np.mgrid[min(values):max(values):13j])
colorscale = ["#141d43", "#15425a", "#0a6671", "#26897d", "#67a989", "#acc5a6", "#e0e1d2",
               "#f0dbce", "#e4ae98", "#d47c6f", "#bb4f61", "#952b5f", "#651656", "#330d35"]
fig = ff.create_choropleth(
    fips=fips, values=values, scope=['Washington'], show_state_data=True,
    colorscale=colorscale, binning_endpoints=endpts, round_legend_values=True,
    plot_bgcolor='rgb(229,229,229)',
    paper_bgcolor='rgb(229,229,229)',
    legend_title='Number of approved Loans by county',
    county_outline={'color': 'rgb(255,255,255)', 'width': 0.2},
    exponent_format=True,
)
iplot(fig, filename='approved_loans_washington')
```



King county has the highest approval rate of 76.2% , so if you apply for a home loan in this county your chances are pretty better. The Ferry County in Washington has the lowest approval rate of 52.1%.

Does your purpose play a role?

Loan applications are applied for purchasing loans, home improvements or for refinancing an existing mortgage.

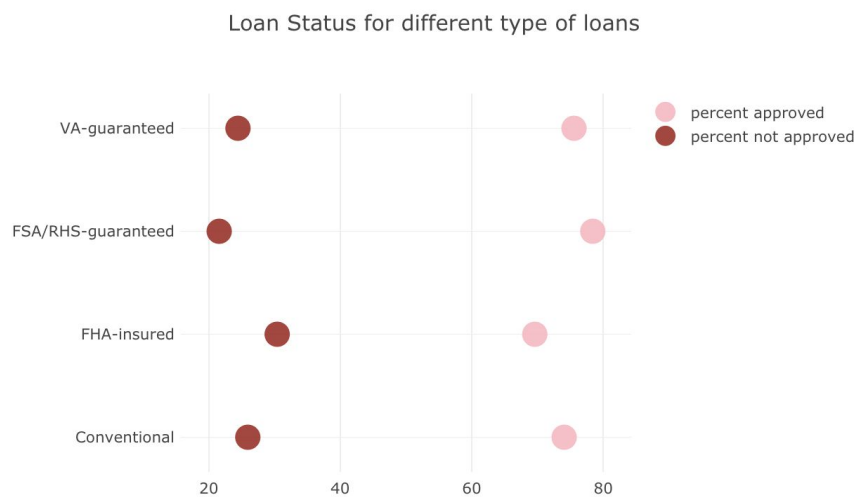


Which type of loans have a better chance of being approved?

Many loans are insured or guaranteed by government programs offered by Federal Housing Administration (FHA), the Department of Veterans Affairs (VA), or the Department of Agriculture's Rural Housing Service (RHS) or Farm Service Agency (FSA). All other loans are classified as conventional.

Code

Code

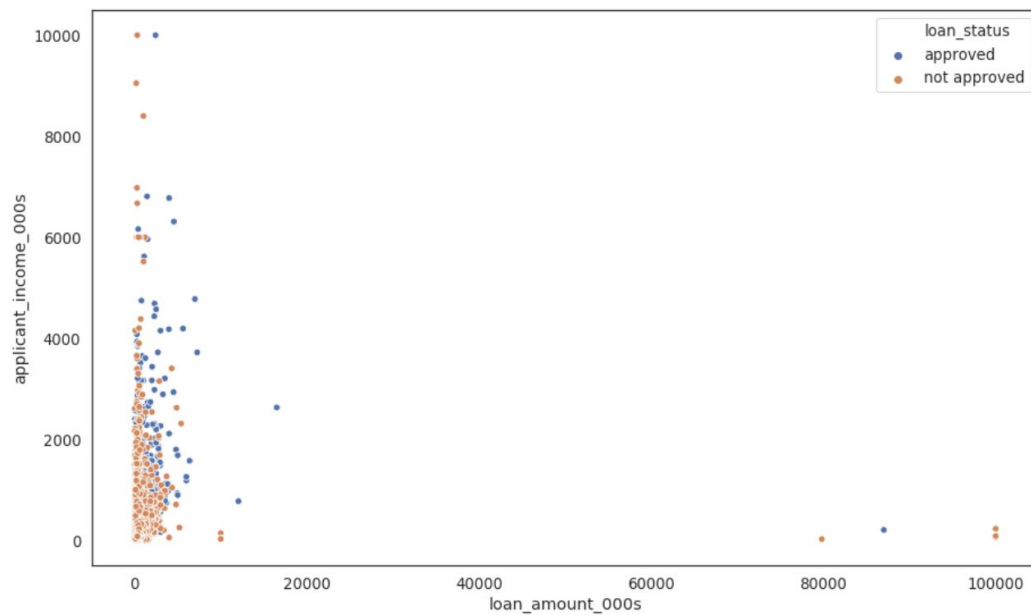


I use percentage as a measure to avoid any scaling issues. FSA/RHS loans have a good record of getting approved. And even Conventional Loans not issued by government programmes have good chances.

Let's have a look at the distribution of applicant's income and loan amount.

In [19]:

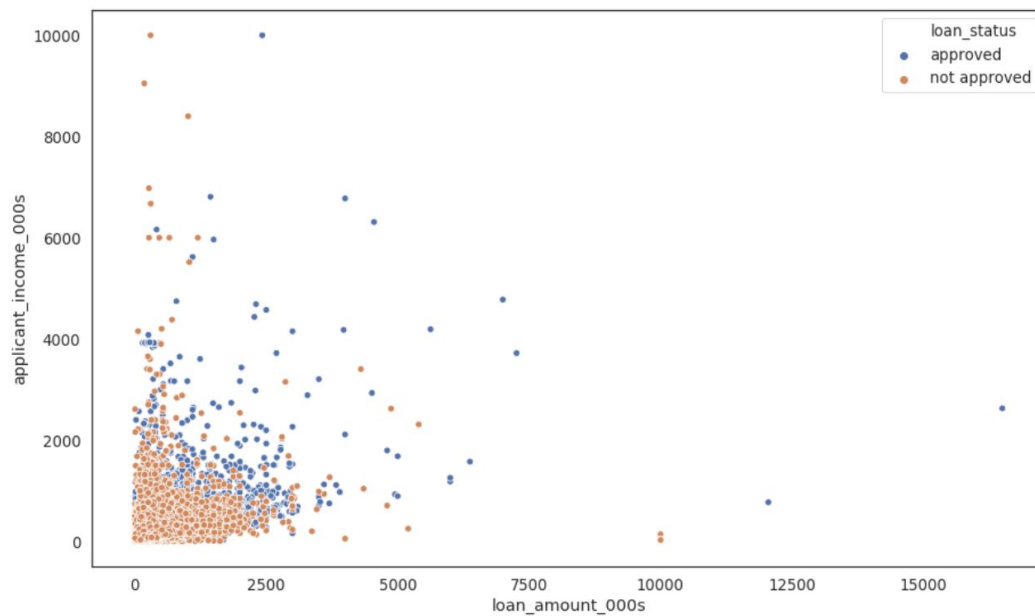
```
sns.set(style="white", palette="deep", font_scale=1.2,
        rc={"figure.figsize":(15,9)})
ax = sns.scatterplot(x="loan_amount_000s", y="applicant_income_000s", hue="loan_status", data=df)
```



I'm quite amazed by the blue dot in the far-right. A person with a low income gets a loan for a very large amount. However the orange dots in the top-left corner can be justified as applications get rejected for lack of documents or unverifiable credentials.

In [20]:

```
df_temp=df[df['loan_amount_000s']<20000]
ax = sns.scatterplot(x="loan_amount_000s", y="applicant_income_000s", hue="loan_status", data=df_temp)
```



Do homes occupied by owners have a higher chance at getting loans?

owner_occupancy_name represents the owner-occupancy status of the property. Second homes, vacation homes, and rental properties are classified as "not owner-occupied as a principal dwelling".

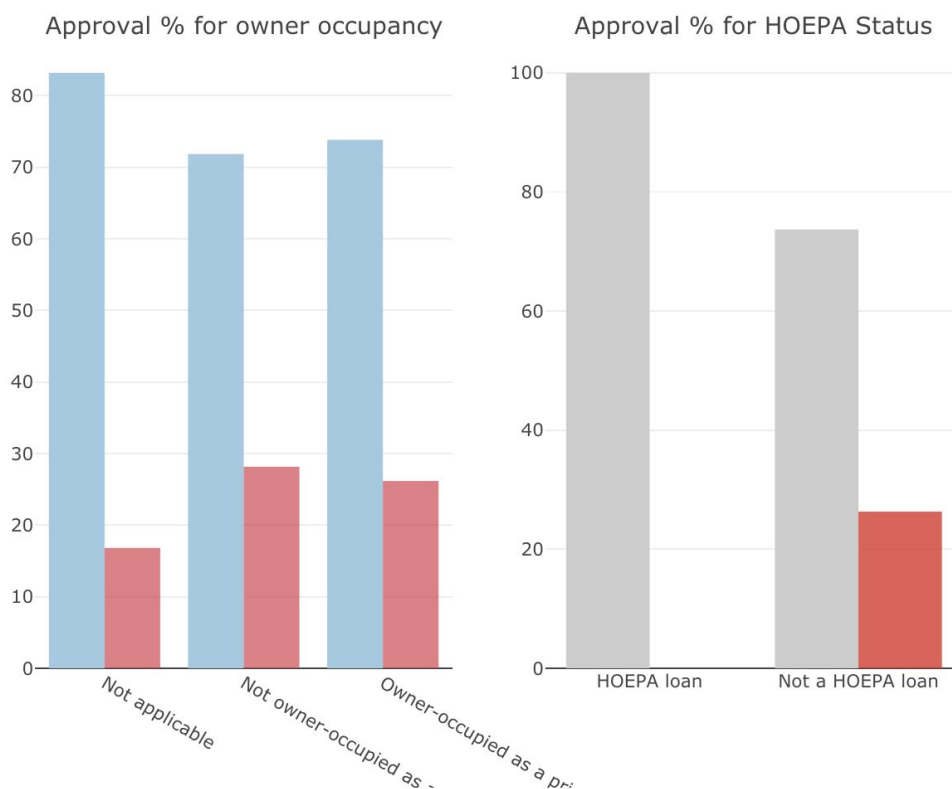
Code

Code

Code

This is the format of your plot grid:

```
[ (1,1) x1,y1 ] [ (1,2) x2,y2 ]
```

The HOEPA status tells us whether a loan is subject to the Home Ownership and Equity Protection Act or not. From the above chart it seems like 100% of the HOEPA loans got approved.

Homes occupied by owners as principle dwellings have a slightly better chance of getting approved when compared to non -owner occupied homes. But the difference is very less, we can't make a statement with much confidence.

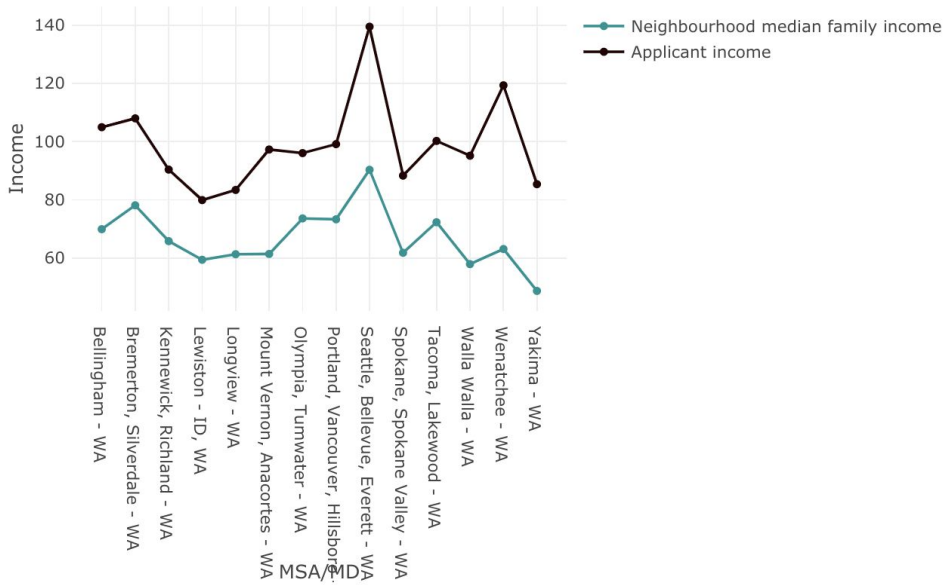
Does the neighbourhood family income affect your chances of getting a loan?

Instead of disclosing the address, lenders disclose the census tract , which is part of the community where the property is located. Each census tract is located in a Metropolitan Statistical Area/Metropolitan Division (MSA/MD). The `hud_median_family_income` is the median family income in dollars for the MSA/MD in which the tract is located.

Now , you must be expecting that for a loan to be approved the applicant's income must be quite similar or above the neighbourhood median family income . Right?

Well that turns out to be true!

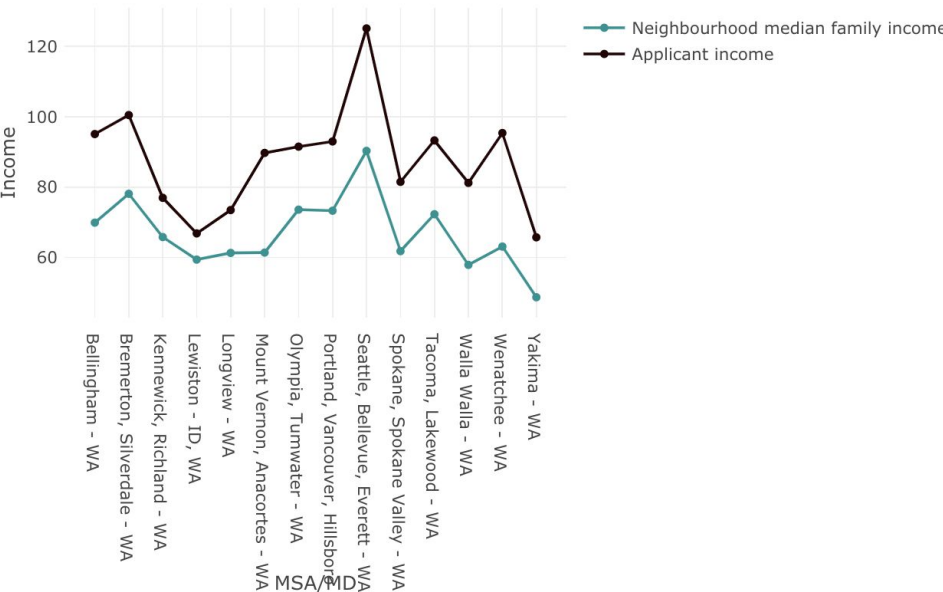
Difference in neighborhood median family income and applicant income for approved loans



What's very unsettling is that for the loans not approved the applicant's income is still greater than the neighbourhood median family income. Maybe , I did something wrong on this , **please leave a comment , if you have any idea on what went wrong.**

Code

Difference for the loans not approved

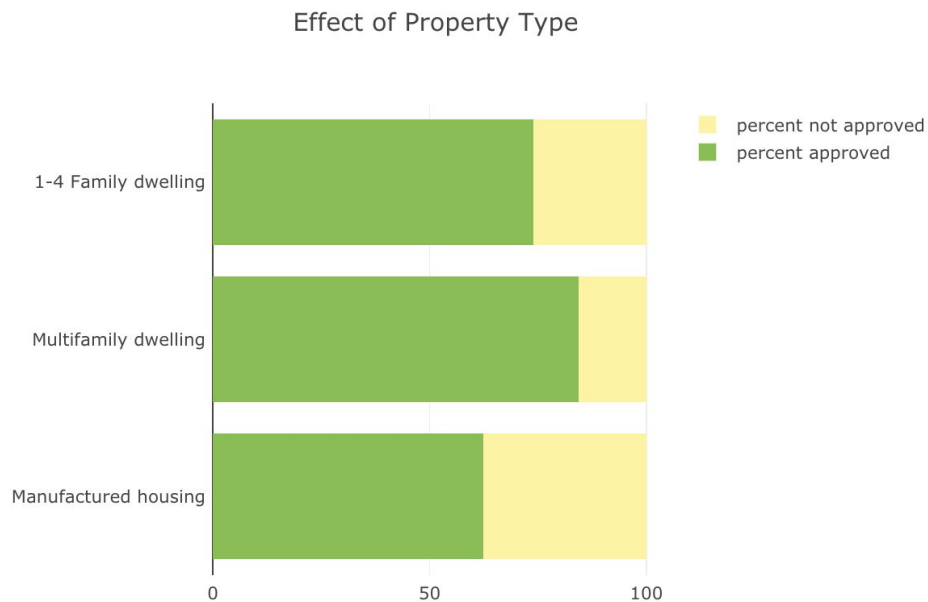


Which property type should you apply for?

There are basically 3 property types, 1-4 family dwelling, multifamily dwelling and manufactured housing. Manufactured homes are housing that is essentially ready for occupancy upon leaving the factory and being transported to a building site.

Code

Code



Multifamily-dwelling is a safe choice when going for house loans.

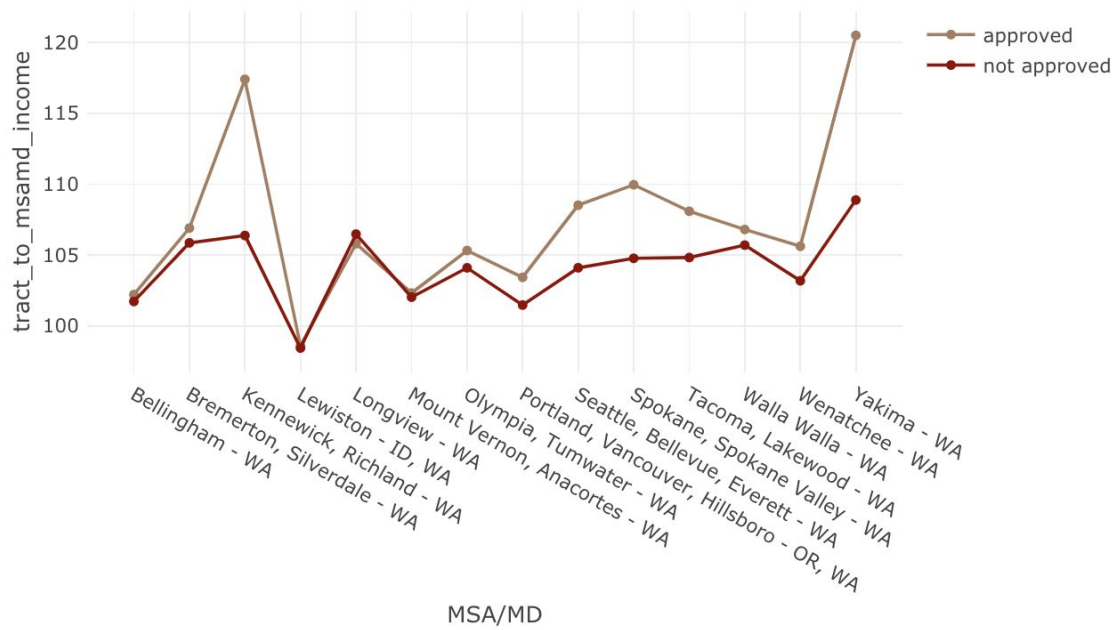
Getting a higher residential mortgage on a multifamily dwelling is easier based on the rental income generated, which can cover or reduce the mortgage.

It's tougher to get a loan for manufactured housing.

Many manufactured home loan programs have strict guidelines about the property condition and age. That's because manufactured housing tends to depreciate, while traditional home values tend to increase over time.

What is the effect of tract_to_msamd_income?

tract_to_msamd_income is the percentage of the median family income for the census tract for which the applicant has applied compared to the median family income for the MSA/MD, rounded to two decimal places.



This ratio is greater for approved loans than for non-approved loans for all MSAMD. And this difference matters a lot if you are applying in Kennewick or Yakima.

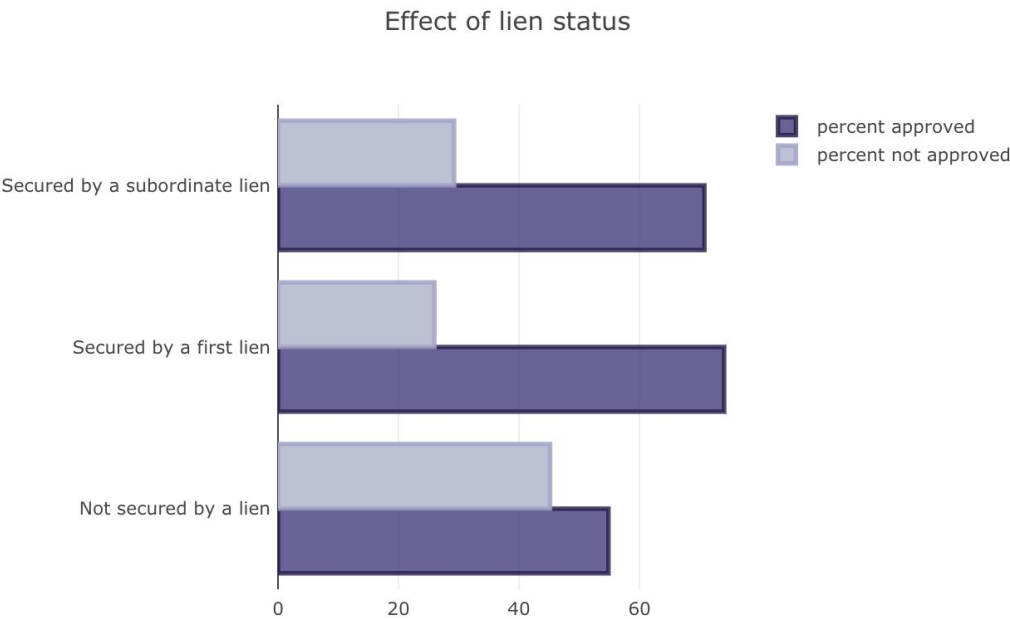
How does the lien status affect the loan?

A lien serves to guarantee an underlying obligation of the repayment of a loan. If the underlying obligation is not satisfied, the creditor(lender) may be able to seize the asset that is the subject of the lien.Once executed, a lien becomes the legal right of a creditor to sell the collateral property of a debtor who fails to meet the obligations of a loan or other contract.

Most mortgages are secured by a lien against the property. In the event of a forced liquidation, first lien holders will generally get paid before subordinate lien holders.

Code

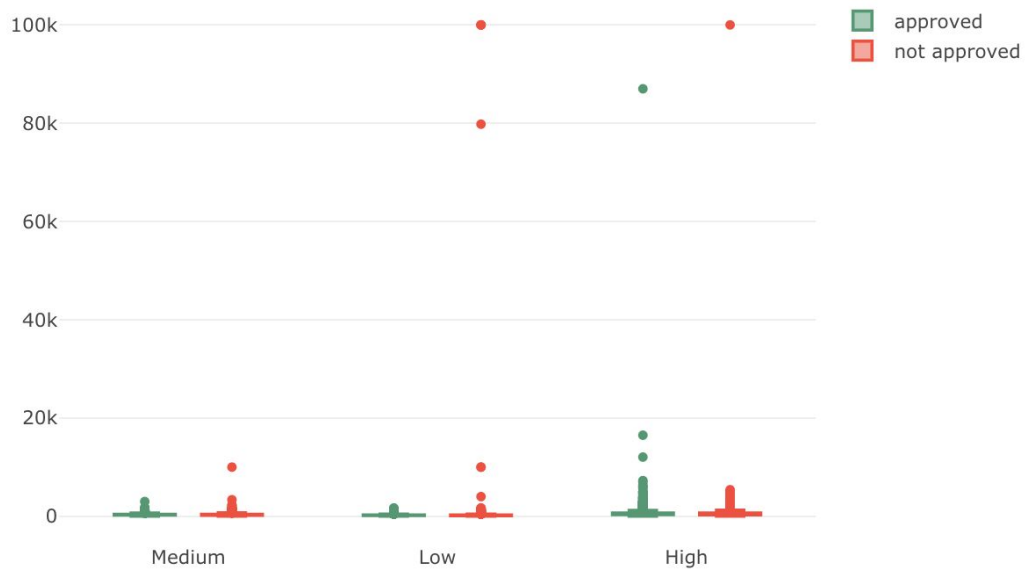
Code



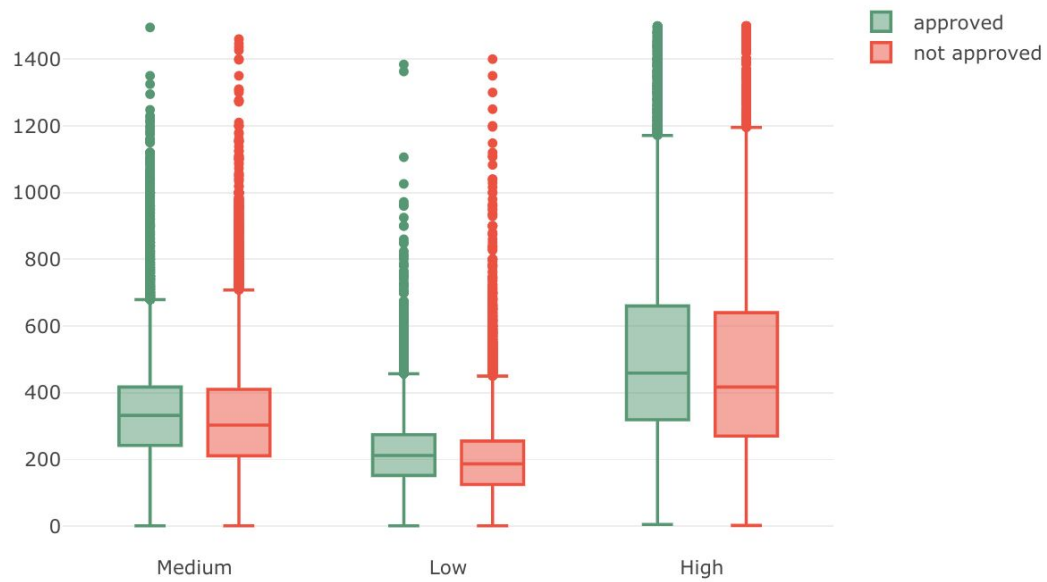
Loans secured by first lien get approved 70% of the time. Loans not secured by any lien has the highest percentage of not getting approved.

What is the loan amount pattern for people in different income categories?

To study this first I'll categorize people as falling in low (for applicant income less than 100k), middle(between 100k and 200k) and high(more than 200k) income range.



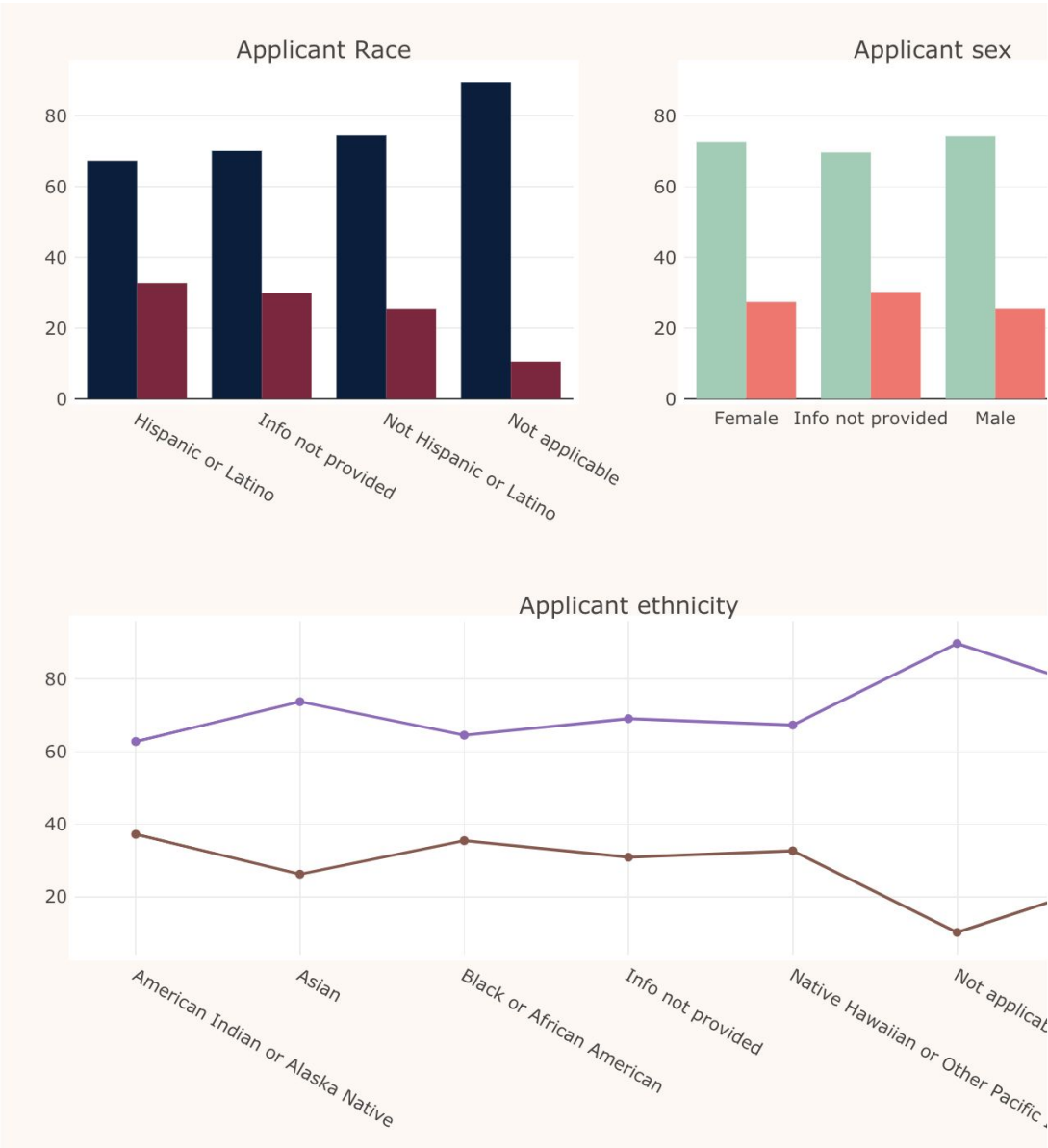
Okay, this is a very compressed view. Loans with very high loan amount are always rejected for the low and medium salary range applicant. While they may get accepted for applicants falling in the high income range. Let's get a zoomed in view by limiting the loan amount to less than 1500000.



In all the income ranges the approved loans have requested for loan amount greater than loans not approved. A usual trend, people generally request for loan amount proportionate to their incomes. In most of the cases , applications where the loan amount far exceeds the applicant's income the loans get rejected.

Is the lending process discriminatory?

Now, I analyze to know whether one gender or race is preferred over the other or not. Not applicable means the applicant is not a natural person , is an organisation or something.

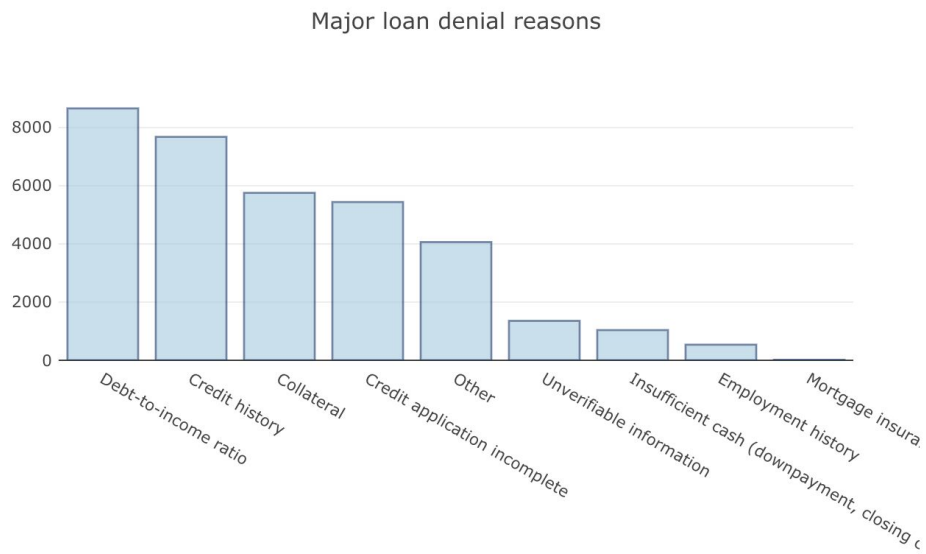


There isn't much difference between approval rates for females and males, however there seems to be a little discrimination based on ethnicity and race favoring Non-hispanic or Latino Whites and Asians.

What are the major reasons for loan denial?

Code

Code



Mostly applications are denied for high debt-to-income ratio, poor credit history, lack of suitable collateral. The most easily avoidable reason is incomplete credit application. Though a total of 5433 applications are denied because of it.

Summary :-

To increase the chances of getting your loan approved in Washington, you must keep the following points in mind.

- Get a home loan in King County (highest approval rate).
- Application for a home purchase loan has a better chance at getting approved.
- Go for a FSA/RHS guaranteed loan. They also have lenient requirements when compared to conventional loans, but this comes at the cost of higher mortgage interest.
- However, there isn't much difference but owner occupied loans have a better chance.
- Must go for HOEPA. Loans with HOEPA always get approved.
- In whichever MSAMD you are buying a home, your income should be greater than or equal to the median family income of that MSAMD.
- Apply for a multifamily or 1-4 family dwelling.
- Get your loan secured by a lien preferably by a first lien.
- Apply for loans proportional to your income. Most loans get rejected for high debt-to-income ratio.
- I don't think you can change any demographic features.