```
/Users/rakesh/anaconda3/lib/python3.7/site-packages/IPython/core/intera
ctiveshell.py:3020: DtypeWarning: Columns (34,36,38,44,46,48) have mixe
d types.Specify dtype option on import or set low_memory=False.
  interactivity=interactivity, compiler=compiler, result=result)
```

Rakesh Senthilvelan and Shin Ehara

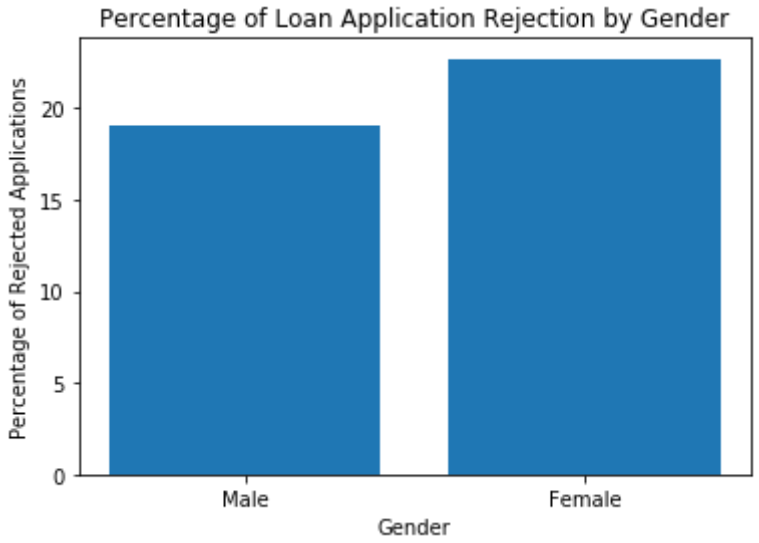
Introduction

Through our observation of the Housing Mortgage Disclosure Act in relation to fairness based on racial group in 2017, we noticed significant discrepancies when it came to the access to mortgages between Black applicants and White applicants across all income groups. In order to better understand the discrepancies and inequities in mortgage lending, we wanted to look into the areas of inframarginality and threshold testing, as well as intersectionality analysis, when it comes to algorithmic decision making in this area. Through this, we believe that a more fair and equitable classification model can be developed to determine the most likely scenario of someone getting their loan application originated or not.

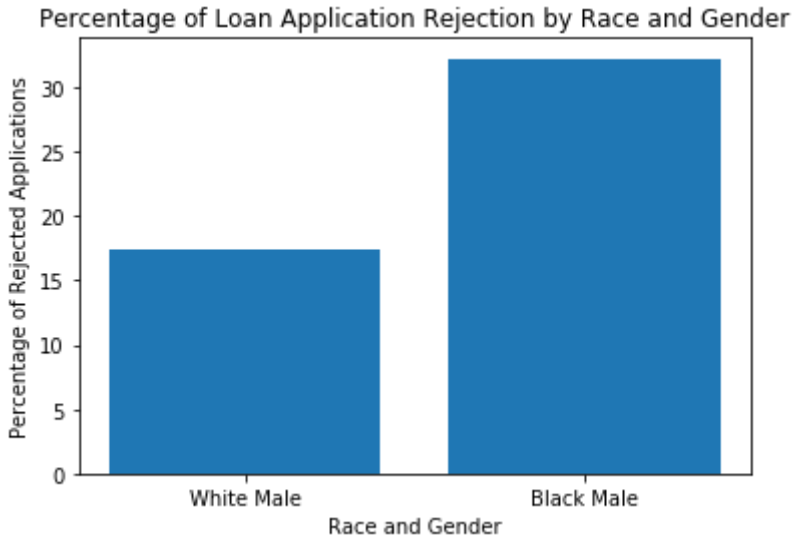
Data Preparation

For the case of our classification system, we wanted to compare between whether or not a loan was approved or denied. To do this, we had to prepare our data as there are numerous data points that have end results such as "loan was withdrawn", "file closed for incompleteness", or "loan was purchased by financial institution" among others. To handle this, we merged the values for "loan originated" and "application approved but not accepted" into one value, "application denied by financial institutio" into another, and then dropped the remaining columns. We will use this subset of the data as the source for our classification model.

In order to train our model in a fair way, we wanted to look into which intersectional group of race and gender saw the highest level of loan acceptance in this dataset. To do this, we did preliminary data analysis comparisons between gender, then we look into the mixture of that gender and race to determine which group sees the highest acceptance rates. Our hypothesis is that this group will be represented by white males. Also, due to the nature of the dataset, the analysis of gender data only include the binary of male and female.



Through the above graph, we can see that males face lower rates of loan rejection and higher rates of loan acceptance than females do. To further analyze, we will look into the intersection of race and gender to see which group has the lowest rejection rate.



Based on the above graph, we can see that White males have a lower rate of rejection than Black males. From this analysis, we will develop our training set off of this demographic.

Building the Model

Now, we will build our model that takes fairness into account. First, we will build the model using training data derived from White males. We believe that due to this intersectional group holding the lowest rejection rate in the dataset in the scope of our research, this will present the most fair subset of the data to pull from as we do not have training data that we can call inherently fair to build the model upon. In terms of the variables used, we will remove sensitive variables including race, ethnicity, sex, minority population percentage in an area, and geographic location to eliminate any biases. We will train the model using the variables of property type, loan type, loan purpose, owner occupancy, our engineered feature for loan to income ratio, and number of owner occupied units. We will develop this model using a Random Forest Classifier.

The reason why we picked these traits is based on correlation as well as the nature of the problem at hand. We created a correlation matrix for all the variables in this dataset and found significant correlations between property type, loan purposes, and purcharser type in relation to action taken with correlation coefficients of 0.1322, 0.1544, and 0.3888 respectively. In addition, we trained on applicant income, loan amount, and loan to income ratio as well due to the financial nature of this situation as such factors would naturally be taken into consideration in any loan-based situation.

```
RandomForestClassifier(bootstrap=True, ccp_alpha=0.0, class_weight=Non
e,
                       criterion='gini', max_depth=10, max_features='au
to',
                       max_leaf_nodes=None, max_samples=None,
                       min_impurity_decrease=0.0, min_impurity_split=No
ne,
                       min_samples_leaf=1, min_samples_split=2,
                       min_weight_fraction_leaf=0.0, n_estimators=100,
                       n_jobs=None, oob_score=False, random_state=None,
                       verbose=0, warm_start=False)
```

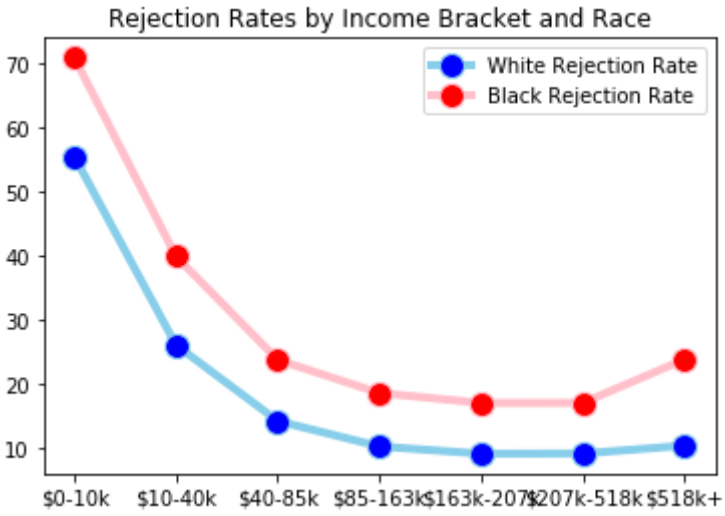
Next, we will run our model on our full dataset and look into how its predictions look. Below, you will see a preview of the model's predictions.

```
['3' '1' '3' ... '3' '1' '1']
```

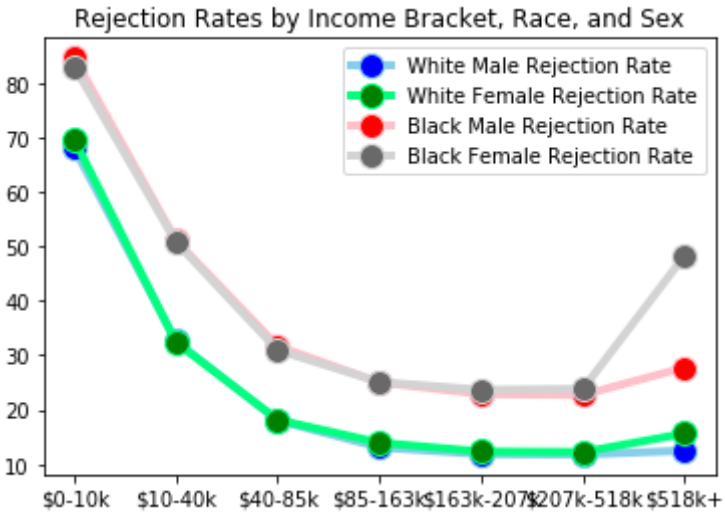
Our model, using Random Forest Classifier and the selected variables, approved 87.28% of the applicants compared to the 80.71% that were approved in our original sample. We believe that developing this training model using the fairness metrics brought up earlier were beneficial in creating a more equal lending system. Now, we will look into the fairness of this algorithm through the lenses of Individual Fairness and Intersectionality.

Intersectionality Analysis

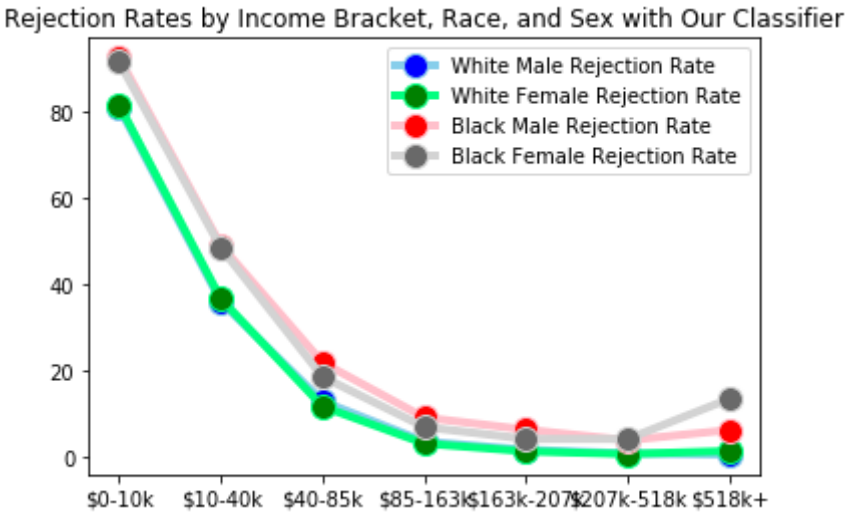
One of the main characteristics we noticed in our data was that Black applicants were rejected a signficantly higher rate than White applicants, as seen earlier in the section titled "Data Preparation" as well as our first paper. When looking into the intersectionality between race and income in our first paper, we noticed that Black applicants in the dataset were consistently rejected at a higher rate than White applicants. That is represented by the following graph, brought from our first paper.



For this intersectionality analysis, we will once again look into the areas of race and income, but we will also look into the area of sex as well as we were able to find a difference in rejection rates based on sex. For the case of our analysis, we will first look at the rejection rates on these different identities in the original data set and then compare those statistics on the predictions of our model.



Based on the above graph, we can see that the main factor that affects rejection rates is race but there is a minor disparity between the genders. There is a significant jump in rejection rates for Black women in the highest income bracket which can be attributed to a small sample size of that demographic. Now, we will do the same analysis but this time, we will swap out the "action taken" column from the original dataset with the predictions from our Random Forest Classifier, which we believe will be more fair.



From the above graph, we can see that using our unsensitive variables, a more equitable system was created compared to the original dataset. One flaw we saw was that there was a high rejection rate at low income ranges and a very low rejection rate at high income brackets. In addition, there is still a higher rejection rate for Black women in the highest income bracket than other groups which is a trait we noticed in the original dataset as well. We believe that there is additional work that would need to be done to create an even more fair system, but that our model shows an improvement over the original dataset.

Individual Fairness

In order to analyze individual fairness, we wanted to look into the distances between our cases with the usage of the Jaccard and Euclidean distance formulas. This will help us determine whether or not, based on the individual traits that we have selected which include our non-sensitive traitts such as loan purpose, loan amount, income, and purchaser type, and loan type while also including sensitive atttributes such as race and sex. This will allow us to look into fairness from multiple angles, particularly the more sensitive ones as mentioned previously. In our case, we will take a sample of 500 entries from the original set and we will normalize the quantitative variables that we have selected. In addition, this will be an analysis into the fairness of our model, so we will replace the action taken column with the predictions that were made by our classifier. From here, we will judge the individual fairness performance of our model. We want to examine the individual fairness of our model, that is, whether the model produces the same outcome for individuals whose input variables are very close. For example, suppose one person with a 100 thousand dollar income and a 50 thousand dollar loan gets his loan accepted. We consider our model individually fair if another person with a 100 thousand dollar income and 50 thousand dollar loan also gets his loan taken regardless of race and gender.

Next, we define how to compare "distance/similarity" between different individuals. To compare the similarity of different instances, we employ different measures for categorical and numerical values. For categorical values ("loan_type", "purchaser_type", "loan_purpose"), we use Jaccard distance to measure the similarity between pairs of individuals in the sample. For numerical values ("loan_amount_000s", "applicant_income_000s"), we use Euclidean distance to calculate the similarity between instances. To calculate the Euclidean distance with better accuracy, we normalized each of these two columns first, then calculated the distance.

This table below consists of all possible pairs of instances from the sample with each pair's Jaccard and Euclidean distances. For clarification, the higher the Jaccard distance, the more similar a pair is. And the lower Euclidean distance, the more similar a pair is. The table below is a small sample of the distances, showing the head, or the first 5 entries, of the dataframe.

	instance 1	instance 2	jaccard	euc
0	2530809	4452886	0.500000	1.651296
1	8214276	4452886	0.250000	0.911146
2	8214276	2530809	0.333333	0.747706
3	1315020	4452886	0.250000	2.252186
4	1315020	2530809	0.333333	3.903101

Now, we want to define "close distance" for each of Jaccard and Euclidean distances. For Jaccard distance, we define it to be 1.00 only as it's the highest similarity score of 7 possible scores (1.0, 7.5, 0.67, 0.5, 0.33, 0.25, 0.0). For Euclidean distance, we define it to be the lowest 10% distance of all, which is 0.304677 as calculated below.

```
0.1    0.304677
0.2    0.568797
0.3    0.841212
0.4    1.126108
0.5    1.438945
0.6    1.792070
0.7    2.220929
0.8    2.763580
0.9    3.693413
Name: euc, dtype: float64
```

With these definitions of "similar instance," we slice rows of similar pairs and compare outcomes of two instances within each pair, as shown below.

	instance 1	instance 2	jaccard	euc
0	5641634	8583422	1.0	0.289155
1	10689407	9130746	1.0	0.210693
2	12706871	8583422	1.0	0.066672
3	12706871	5641634	1.0	0.235142
4	12557361	11923108	1.0	0.137827

For these selected pairs of instances, which we consider to be "similar," we want to know what action was taken in our model for their loan application.

	instance 1	instance 2	jaccard	euc	action_1	action_2	same outcome
0	5641634	8583422	1.0	0.289155	1.0	1.0	True
1	10689407	9130746	1.0	0.210693	1.0	1.0	True
2	12706871	8583422	1.0	0.066672	1.0	1.0	True
3	12706871	5641634	1.0	0.235142	1.0	1.0	True
4	12557361	11923108	1.0	0.137827	1.0	1.0	True

```
True    1267
Name: same outcome, dtype: int64
```

As shown above, all 1,267 pairs of similar individuals experienced the same outcome (loan action). Thus, it seems that our model performs very fair at the individual level; that is, the model produces the same result for individuals whose input variables are very close. Finally, we explore whether these same outcomes between similar individuals are attributed to their race or gender backgrounds. To do this, we compare the race and gender of each instance within a pair. First, we will look at race.

	instance 1	instance 2	jaccard	euc	action_1	action_2	same outcome	race_1	race_2	same race
0	5641634	8583422	1.0	0.289155	1.0	1.0	True	5.0	5.0	True
1	10689407	9130746	1.0	0.210693	1.0	1.0	True	5.0	5.0	True
2	12706871	8583422	1.0	0.066672	1.0	1.0	True	5.0	5.0	True
3	12706871	5641634	1.0	0.235142	1.0	1.0	True	5.0	5.0	True
4	12557361	11923108	1.0	0.137827	1.0	1.0	True	5.0	3.0	False

```
True     1117
False     150
Name: same race, dtype: int64
```

As shown above, 10+% of pairs in the sample consisted of two individuals with different races yet all of them experienced the same loan action. None of the pairs with different racial backgrounds experienced different outcomes with our model. Next, we will look into sex.

	instance 1	instance 2	jaccard	euc	action_1	action_2	same outcome	race_1	race_2	same race
0	5641634	8583422	1.0	0.289155	1.0	1.0	True	5.0	5.0	True
1	10689407	9130746	1.0	0.210693	1.0	1.0	True	5.0	5.0	True
2	12706871	8583422	1.0	0.066672	1.0	1.0	True	5.0	5.0	True
3	12706871	5641634	1.0	0.235142	1.0	1.0	True	5.0	5.0	True
4	12557361	11923108	1.0	0.137827	1.0	1.0	True	5.0	3.0	False

```
True     688
False    579
Name: same sex, dtype: int64
```

As shown above, nearly half of the pairs in the sample consisted of two individuals with different genders, yet all of them experienced the same loan action. Furthermore, none of the pairs with different genders experienced different outcomes with our model. Thus, to summarize what we found in this individual fairness analysis, our model is very fair in producing the same loan action for individuals with similar input variables. This fairness is not significantly attributed to their race and gender.

Conclusion

Based on our prior analysis, it is evident that the original dataset of mortgage data from 2017 showed levels of unfairness on the basis of sensitive traits such as race and sex. Through our analysis in our first paper and this paper, we found that there was a difference in acceptance and denial rates for people based on race and sex, even when comparing within similar income brackets. Developing our model that took in less-sensitive traits allowed us to develop a more fair solution to determining whether a loan gets accepted or denied.

In our intersectionality analysis, we were able to see through our graphs that there was a significant difference in denial rates between Black and White applicants in the original dataset. Through our model's predictions, this difference shrank drastically, indicating a more fair system for evaluating mortgage loan acceptances when taking into account the attributes of race, sex, and income bracket together. From here, we looked into an audit of the individual fairness our model. What we found was that, when taking into account the race, sex, and overall traits of each individual in our sample set, we noticed no major differences in the Jaccard and Euclidean distances between each entry in relation to action taken. This indicates that there is similar treatment between the people based on their circumstances when our model's predictions are taken into acccount, showing that our model displays individual fairness. While we believe that there is still room for improvement when it comes to fairness in mortgage lending data, we see that our model takes a step in further achieving true fairness.

Works Cited

"Download Historic HMDA Data." Consumer Financial Protection Bureau, www.consumerfinance.gov/data-research/hmda/historic-data/.

Fleisher, Will. "What's Fair About Individual Fairness?" SSRN Electronic Journal, 2021, doi:10.2139/ssrn.3819799.

Foulds, James R., et al. "Bayesian Modeling of Intersectional Fairness: The Variance of Bias." Proceedings of the 2020 SIAM International Conference on Data Mining, 2020, pp. 424–432., doi:10.1137/1.9781611976236.48.

Fraenkel, Aaron. "Lecture 10: Limits of Observational Fairness: Intersectionality & Subgroup Validity." DSC 167. 28 May 2021, San Diego, CA.

Fraenkel, Aaron. "Lecture 11: Individual Notions of Fairness." DSC 167. 3 June 2021, San Diego, CA.

Senthilvelan, Rakesh, and Shin Ehara. "Analysis of Inequities in Mortgage Lending Data." 7 May 2021.