Introduction:

In this project we are going to check if there is any racial disparity on mortgage loans in the US by checking interest rate, loan amount to mortgage value ratio and other loan data for 2019. In this project we try to determine if there is any difference in the pattern in the features for different races.

Data Wrangling:

In the Data Wrangling part , we have checked missing values, converted categorical columns to categorical, drop columns with no information.Saved the dataframe to process the EDA part.

EDA :

In EDA , we check the distribution of the features for different races and check if the patterns are similar or different. Here are the findings from our exploratory data analysis.

1. Most loans have a maximum of 2 borrowers so focus on the race of the first two borrowers.
2. There are high numbers of White borrowers compared to other races.
3. Most Banks are from Chicago, Topeka, and Des Moines and all the cities have the majority of white borrowers to whom the loan was approved.
4. All the states are having a similar distribution of race with a majority of the white race except states with state codes 15 and 66.

5. A major disparity is found in areas with a high percentage of the minority population. In this area the majority of borrowers are white even if the majority population in the area is a minority.

6. The difference between local area median income and Total monthly income used to underwrite the loan follow the same pattern for all the races. Therefore there is no disparity based on income.

7. All the races have got most of the loans worth 80% of the property value except American Indians who got most of the loans worth 90% of the property value which may indicate some disparity.

8. The major purpose of borrowing the loan for all the races is the purchase and the most popular loan among all races is the Conventional loan.

9. Majority of white and native Hawaiian loans have co-borrowers whereas all the other races have majority loans with one borrower.

10. Most of the borrowers are male and borrower's ages lie between 20 and 80 years and the number of loans approved tends to decrease after age 35.

11. Interest rate seems to be following a similar pattern for all the races.

12. NoteAmount, HousingExpenseRatioPercent, TotalDebtExpenseRatioPercent have a similar pattern for all the races.

13. We see a different pattern in the credit score for different races. White and Asian have a high number of borrowers with credit score group 5 whereas, for Black and Native Hawaiian, a large number of borrowers falls in groups 4 and 5. American Indians have a large number of borrowers fall in groups 4 and 5 with maximum numbers of borrowers in group 5.

Data Modelling :

In the Data Modelling part of the project we try to predict the race of the person and in doing so we try to find out the features which are important in determining the race. If the model is significant the

accuracy of the model will be high which indicates the correlation among features and race. We extract the important features from all the trained models and rank them according to its importance.

We found there are few features which are common among all the models.

- CensusTractMinorityRatioPercent
- LoanAcquisitionActualUPBAmt
- NoteAmount
- HousingExpenseRatioPercent
- LocalAreaMedianIncomeAmount
- MortgageType_1
- Borrower1EthnicityType_1

Conclusions:

From this project we conclude that there are few features based on which there is some disparity.