

# Home Mortgage Decisions in NY

Deepsikha Sarma | GIPE



# Setting the Scene



Basics and Starting off with the Analytic Journey

## Data Source



<https://www.consumerfinance.gov/data-research/hmda/explore>

## Contents



This dataset covers all mortgage decisions made in 2015 for the state of New York.

## Business Question

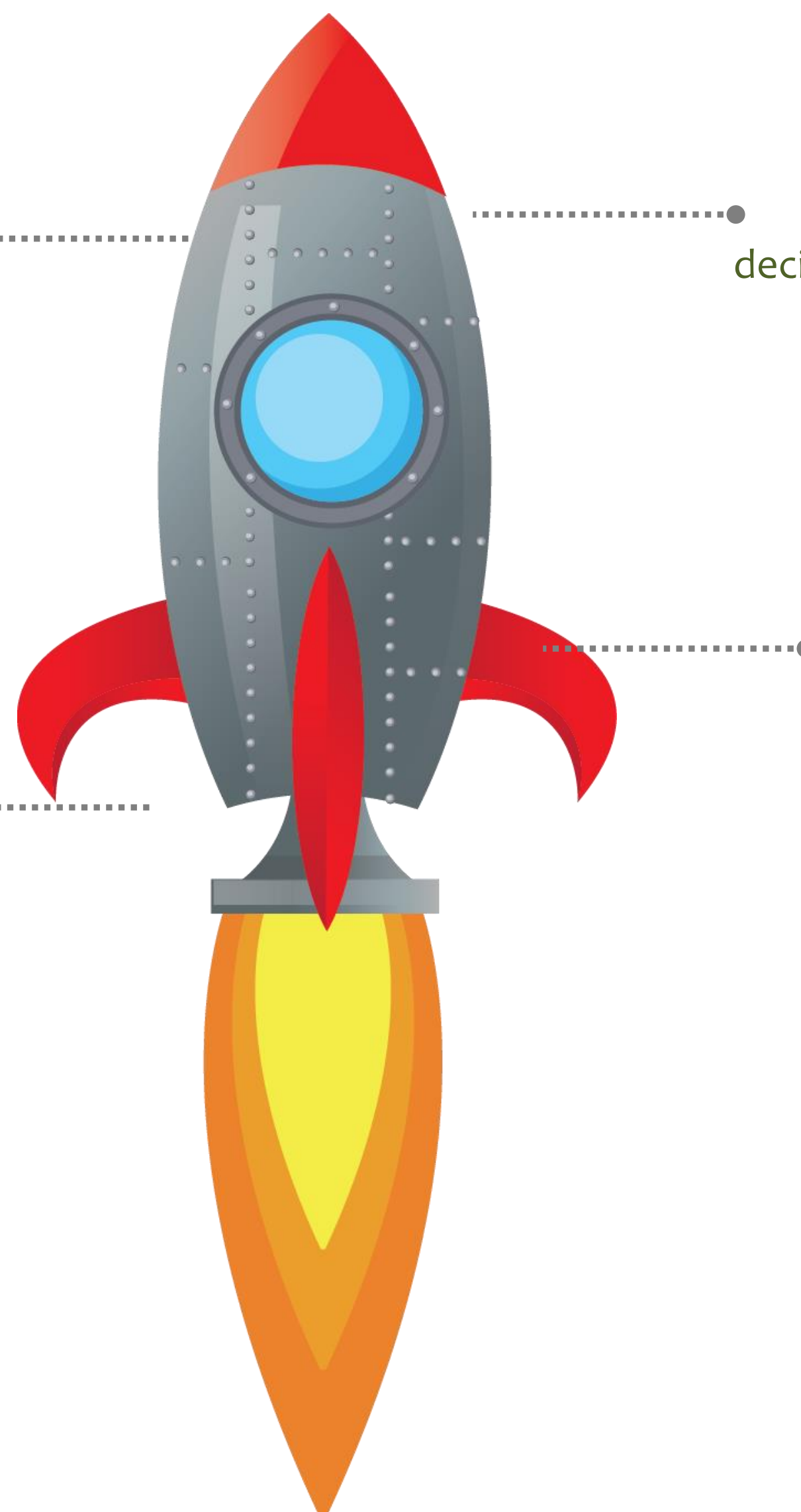
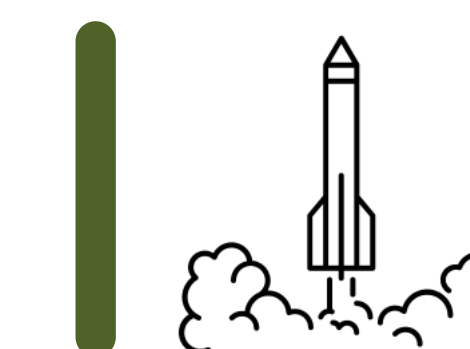


Where are mortgages most likely to be approved?  
Can you predict mortgage decisions based on the criteria provided here?

## Data Features

0.4 Million Records

Variables 77



# The Basics | Concepts

3

## Understanding the Business Context

### What is HMDA ?

Each year thousands of banks and other financial institutions report data about mortgages to the public, thanks to the **Home Mortgage Disclosure Act**, or “**HMDA**” for short. These public data are important because:

- Help show whether lenders are serving the housing needs of their communities;
- Give public officials information that helps them make decisions and policies; and
- Shed light on lending patterns that could be discriminatory

## Loan Origination Journey

Meet Alex. He wants to buy a home but doesn't have the money to pay for it in cash, so he applies for a loan at his bank. He tells the bank about his finances, the house he wants to buy, and other information the bank needs to make a decision about whether or not to lend to him, and the terms of the loan.

The bank reviews Alex's application, decides that he meets their criteria, and he gets approved. Once all the papers are signed, Alex closes the loan... or in mortgage-speak, **the loan is “originated.”**

Therefore the last stage of the loan is **Loan Origination**.

# Section 1 : The Variables

Telling the Data Story

## Understanding the problem - Data Perspective

The data provided can be grouped into the following subjects

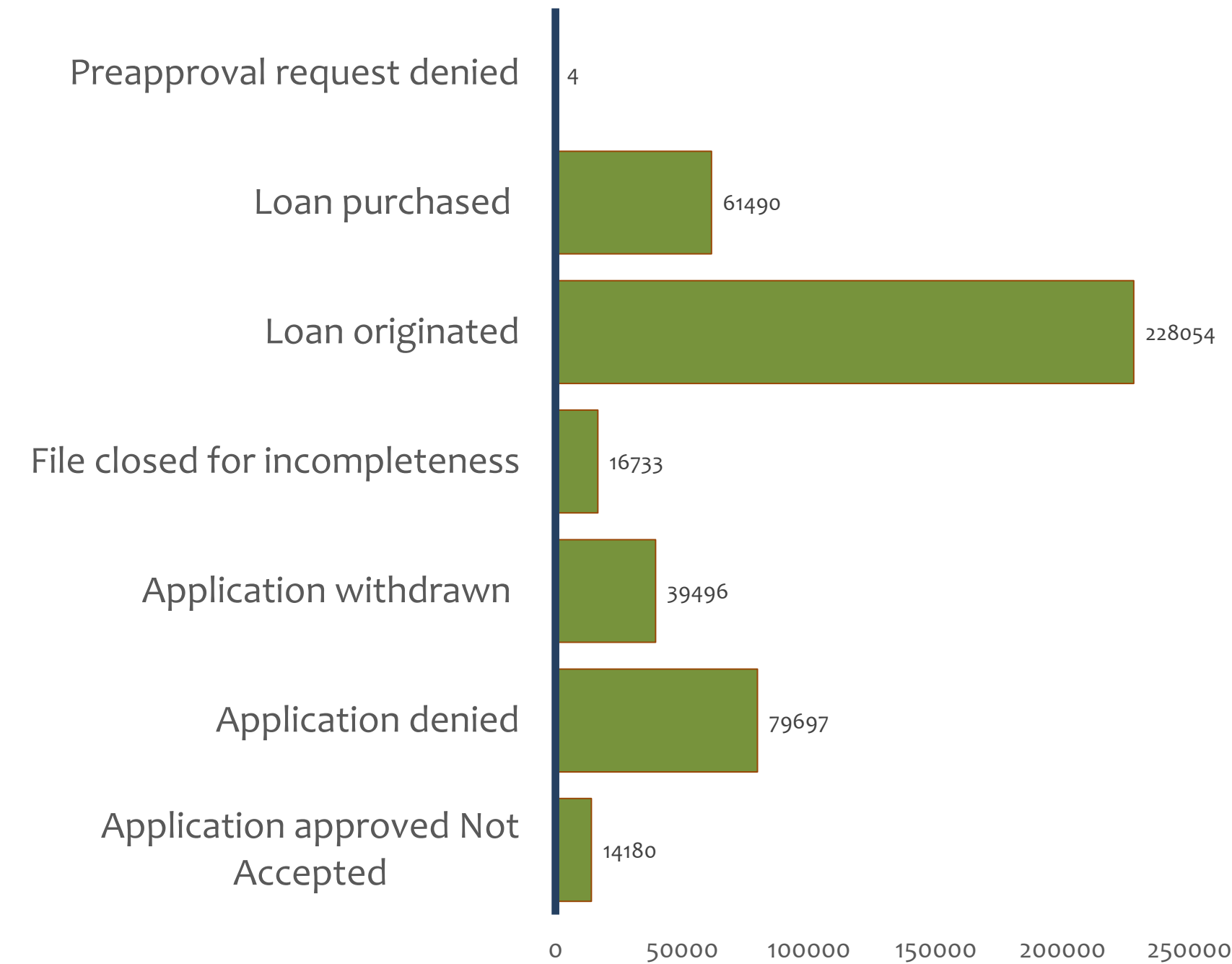
- **Location** describes the State, metro area and census tract of the property
- **Property Type** describes the **Property Type and Occupancy of the property**. Property type values include One-to-four family dwelling, Manufactured housing and Multifamily dwelling. This also answers the question “Will the owner use the property as their primary residence ?” . The values include Owner occupied as principal dwelling , Not owner occupied as principal dwelling and Not Applicable.
- **Loan** describes the action taken on the Loan, purpose of the Loan , Type of the loan ,Loan’s lien status.
- **Lender** describes the lender associated with the loan and the Federal agency associated with the loan.
- **Applicant** describes the demographic information for the applicants and the co-applicants. This has the applicant sex , co- applicant sex , applicant race and ethnicity, co- applicant race and ethnicity.

# A Look Into the Data EDA

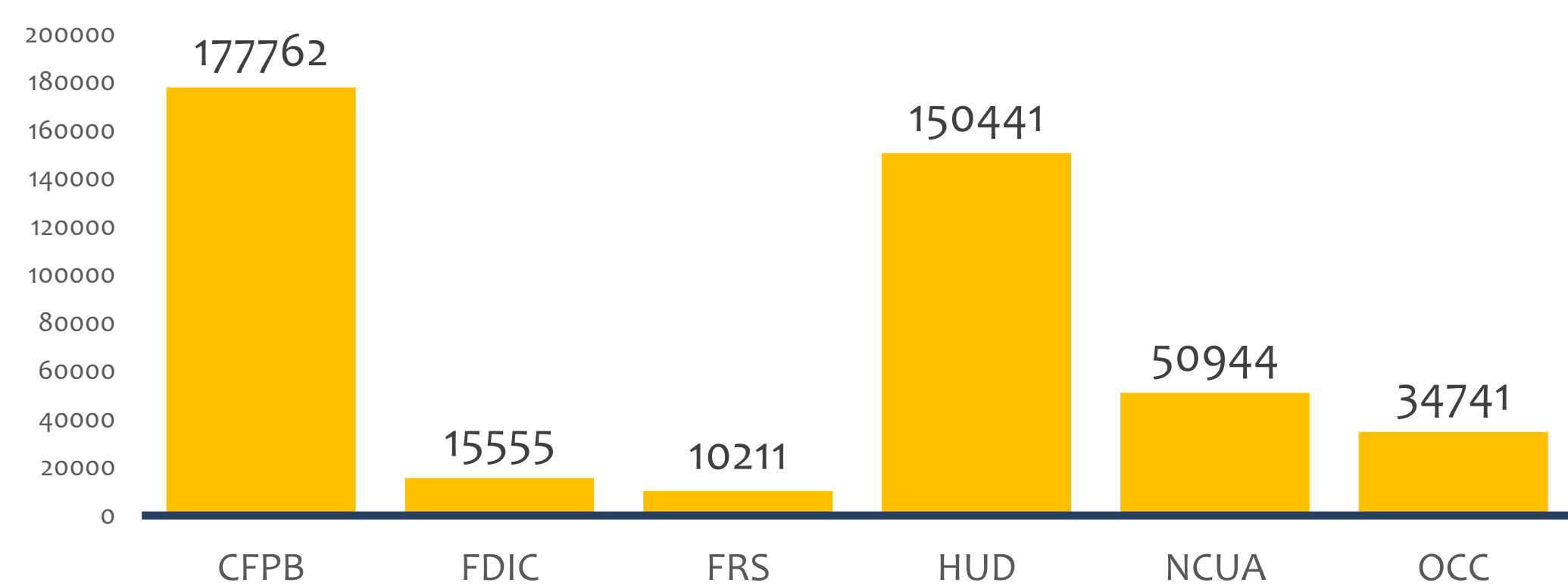


Telling the Data Story

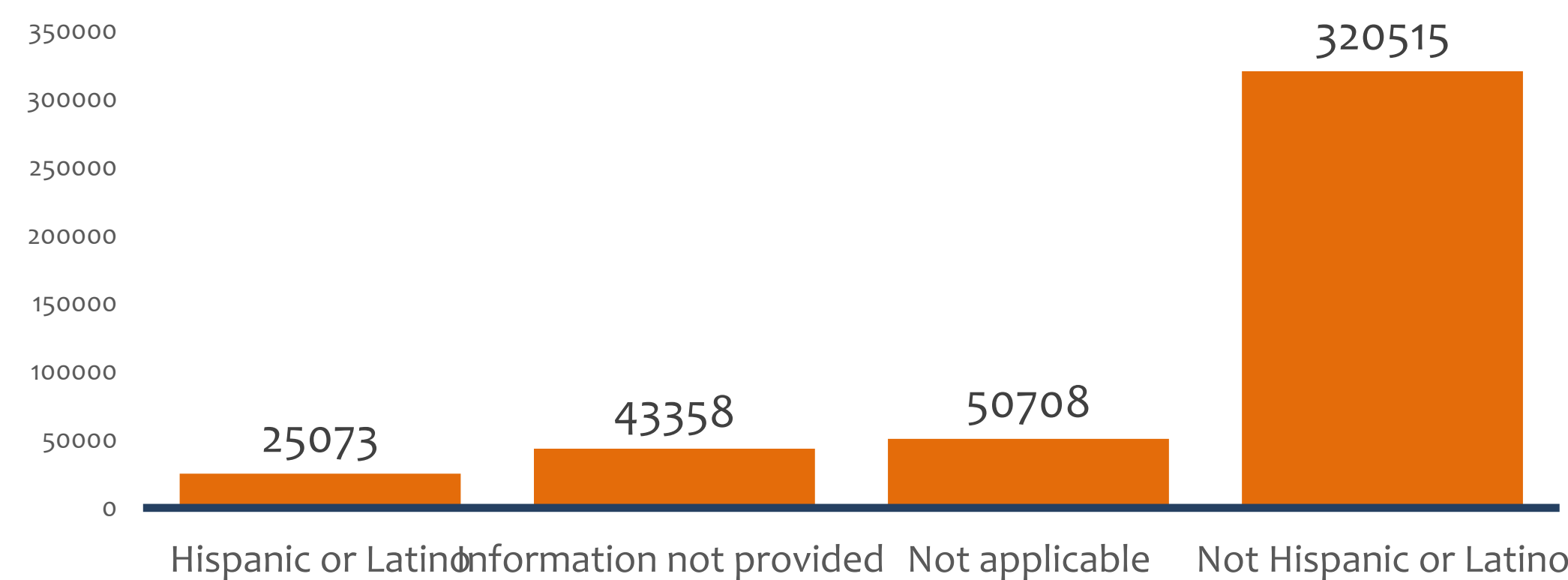
Action Taken on Loans | Counts



Cases By Agencies| Counts



Ethnicity| Counts



Notes:-

Agency Abb	Agency Name
CFPB	Consumer Financial Protection Bureau
HUD	Department of Housing and Urban Development
NCUA	National Credit Union Administration
FDIC	Federal Deposit Insurance Corporation
OCC	Office of the Comptroller of the Currency
FRS	Federal Reserve System



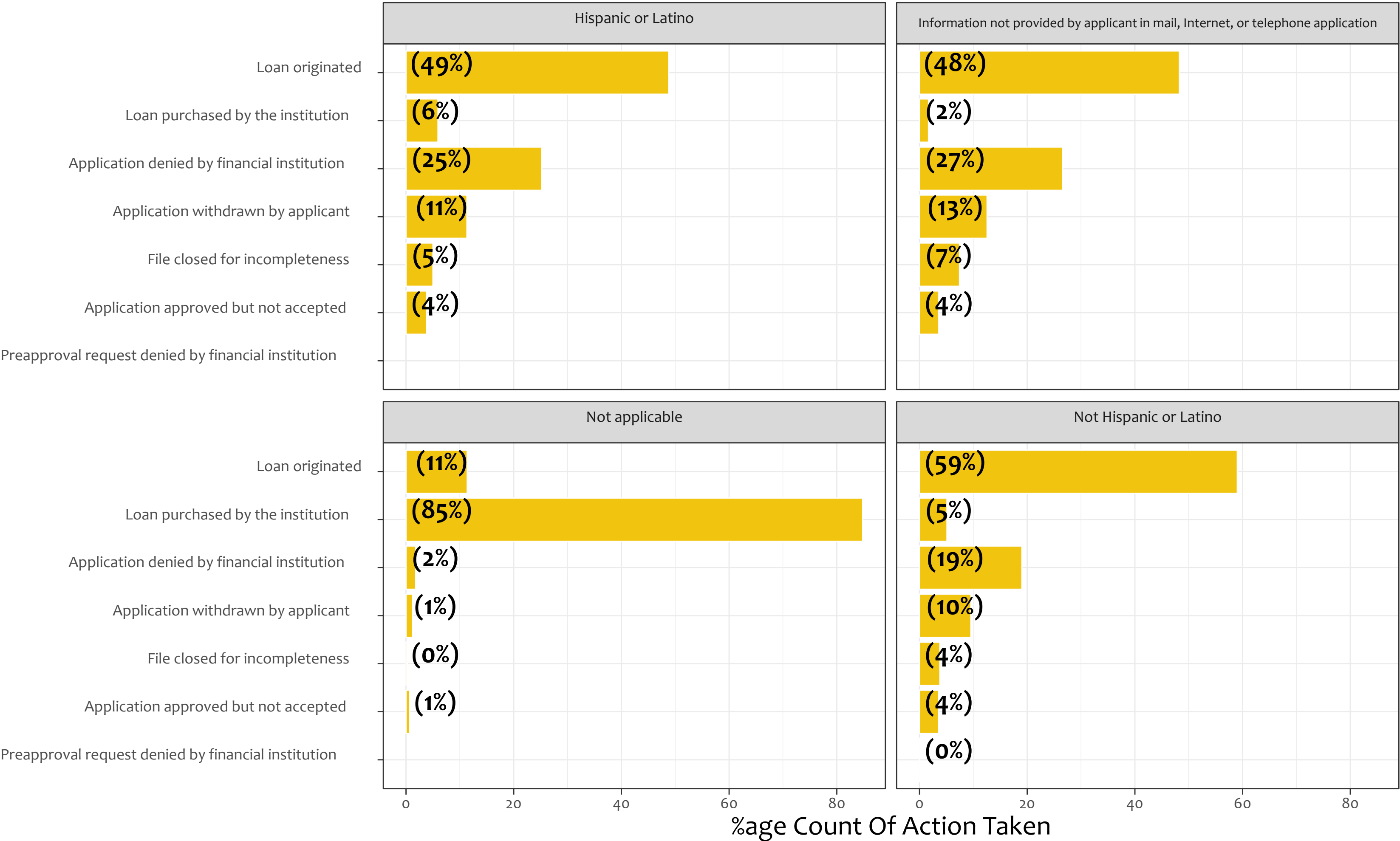
# EDA Continued



## Telling the Data Story

### Actions in Loans

Action Taken Name



### Insights

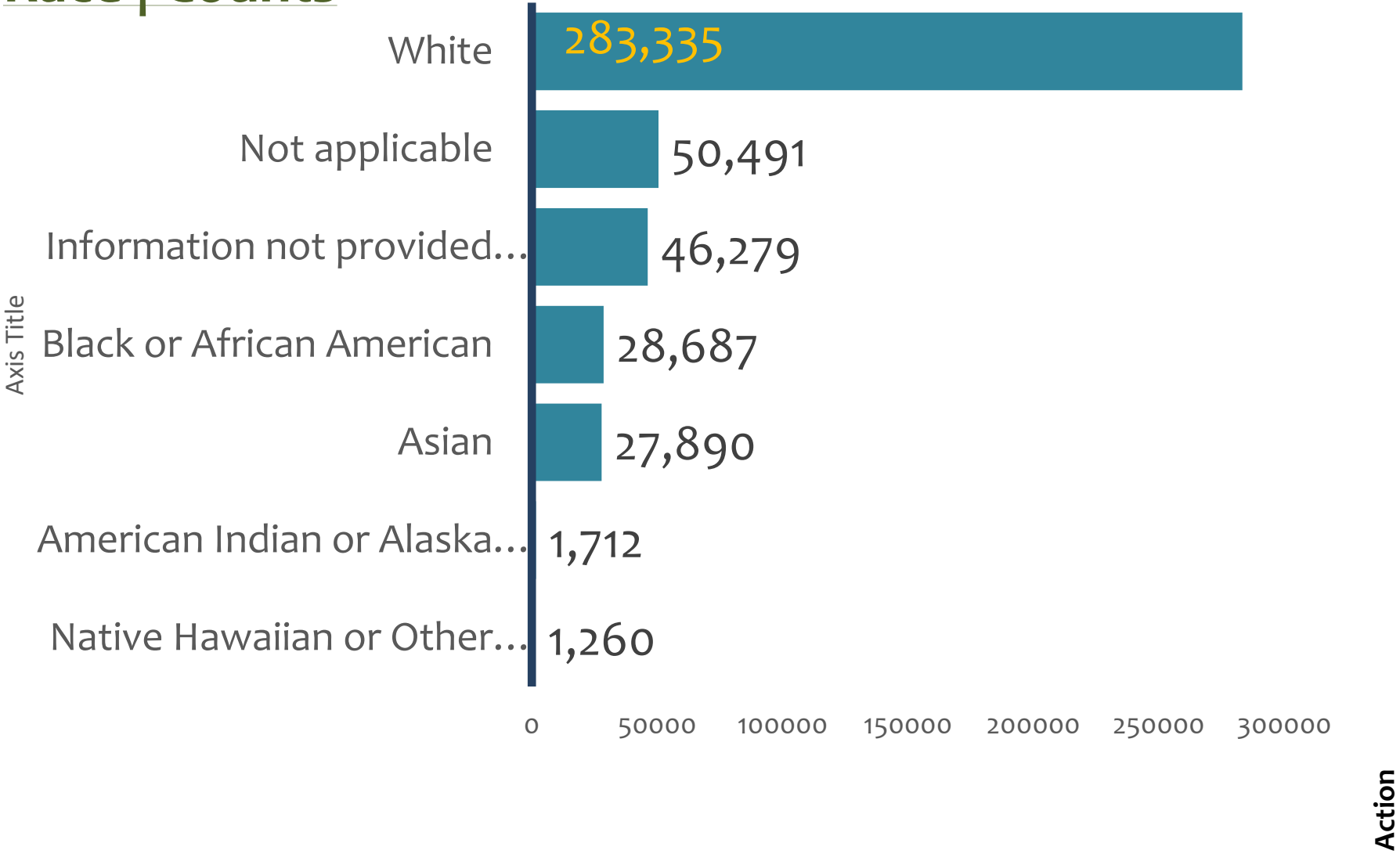
- The **Not Hispanic or Latino** ethnic community has a **greater percentage (59%)** of the **Loans originated** compared to the **Hispanic or Latino** community which has around **(49%)**.
- The **Not Hispanic or Latino** ethnic community has a **Lesser percentage (19%)** of the **Application denied by the Financial institution** compared to the **Hispanic or Latino** community which has around **(25%)**.

# EDA Continued



## Telling the Data Story

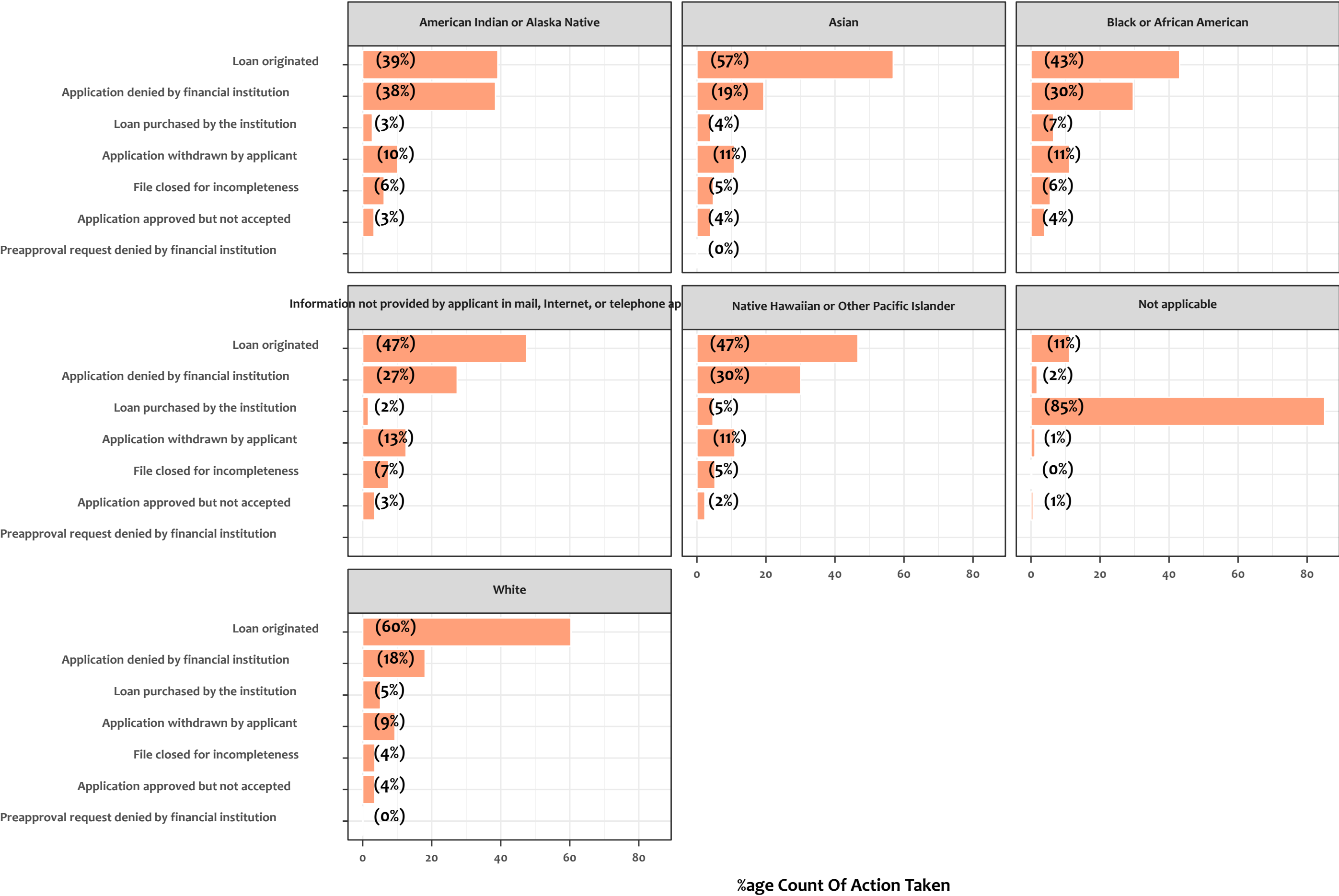
### Race | Counts



### Insights

- The data was split across race and analyzed
- Whites dominated loan applications however, a significant chunk didn't mention race on the application with also cases where race was missing in the dataset
  - The Loan Origination percentage for the White and Asian community is higher than the Black or Asian community.

Actions in Loans by Race

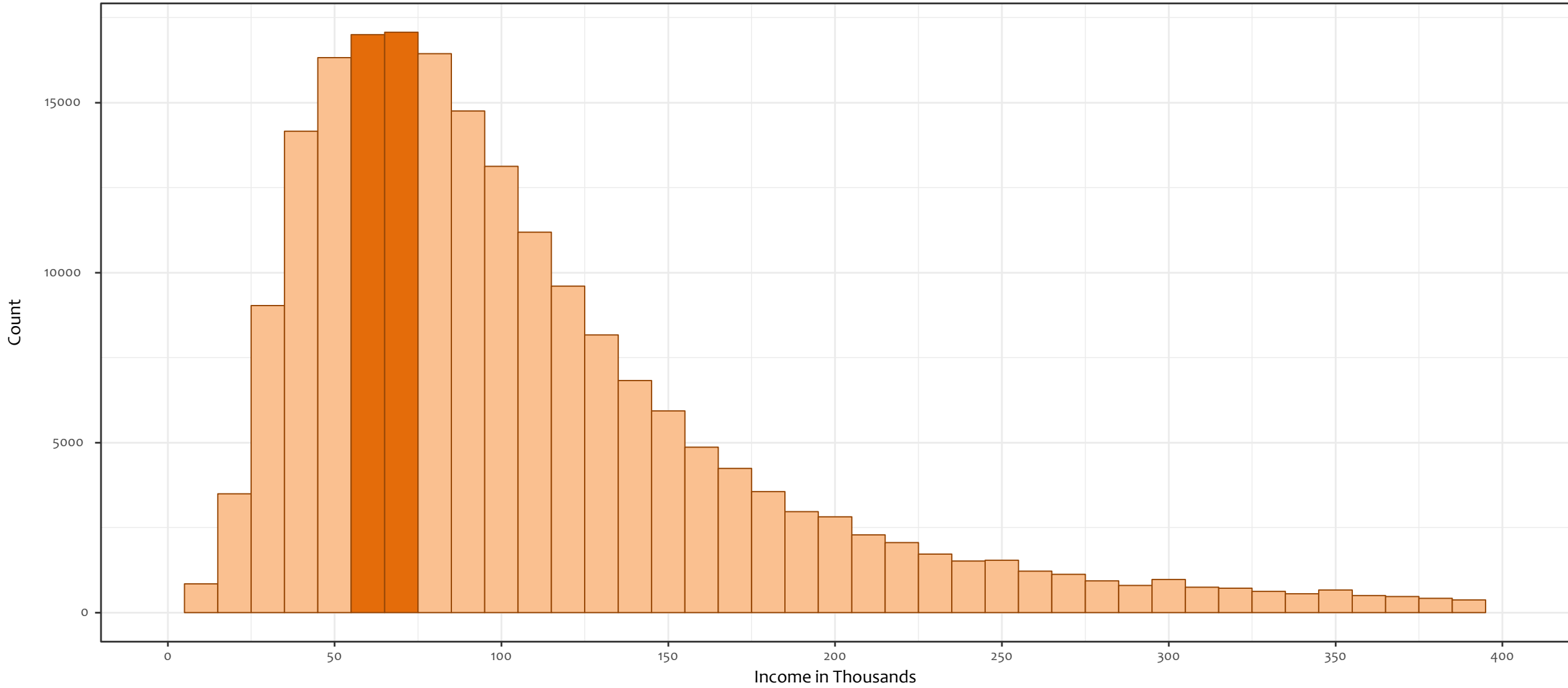


# EDA Continued

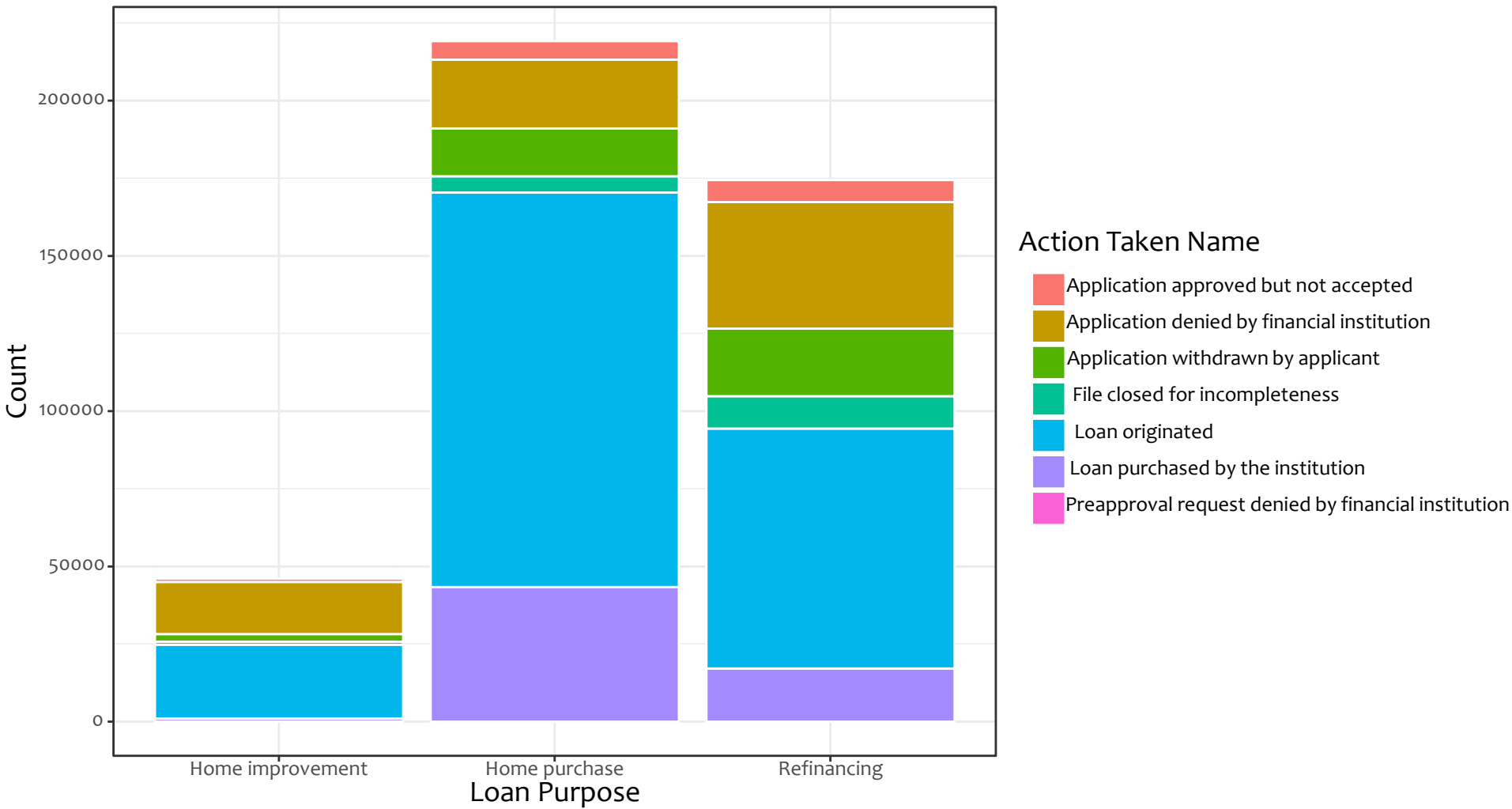


## Telling the Data Story

Loan Originated Applicant Income distribution



Loans Purpose Types Distribution with Action Types

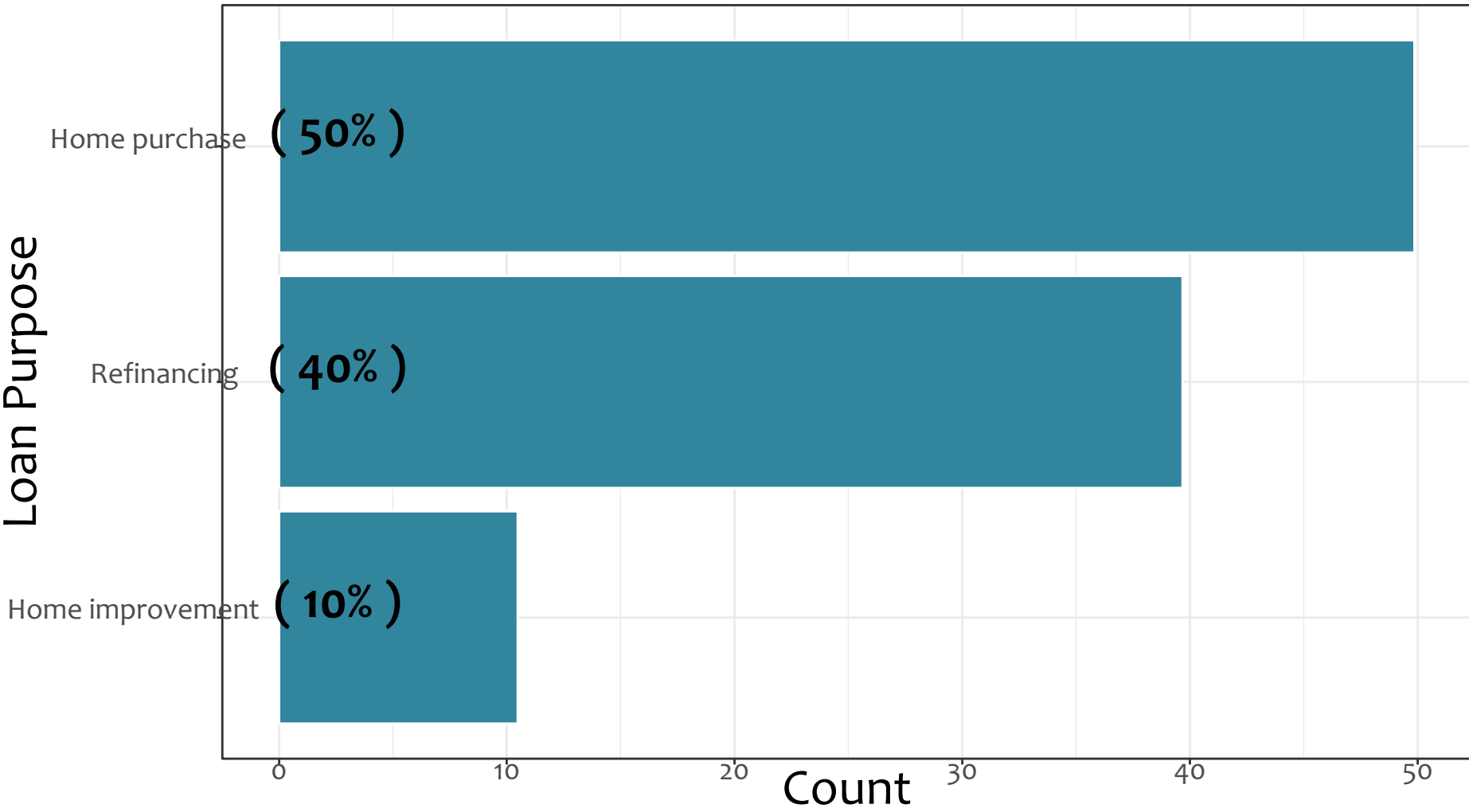


## Insights

### Moving On ...

- We observe that **MOST** of the loans which are originated have applicants with income around **Sixty Thousand to Seventy Five thousand dollars**.
- **Home Purchase** and **Refinancing** are the major Loan Purpose types.

Loans Purpose Types





## Section 2 : The Model

Building the Model and Analyzing the results

### **Modelling Purpose:-**

We predict whether the status of the Loan would be **Loan originated** or not. The following tree shows the conditions which would be used to determine whether the would be **Loan originated** or not.