Lending Club Case Study

Mandeep Singh Sachdeva

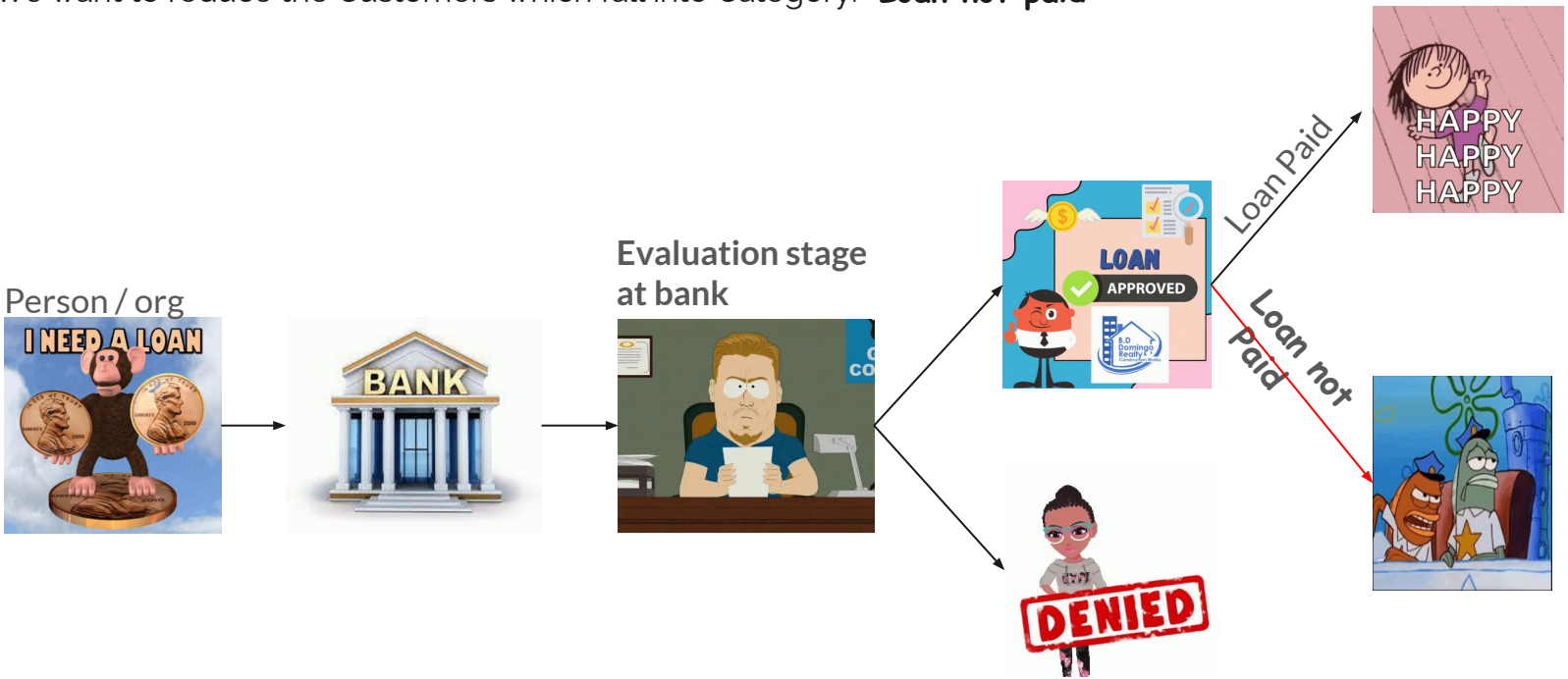
Objective

To implement EDA technique on a Loan Dataset which

- Helps the business to understand the patterns of the data.
- Helps to understand the various reasons why the loans are defaulted
- Make right choice of giving loan for the next time by analysing the previous data
- Understand the various risk associated with the loan business of the bank

Problem Statement

- We want to reduce the Customers which fall into Category: **"Loan not paid"**



How would we do it?

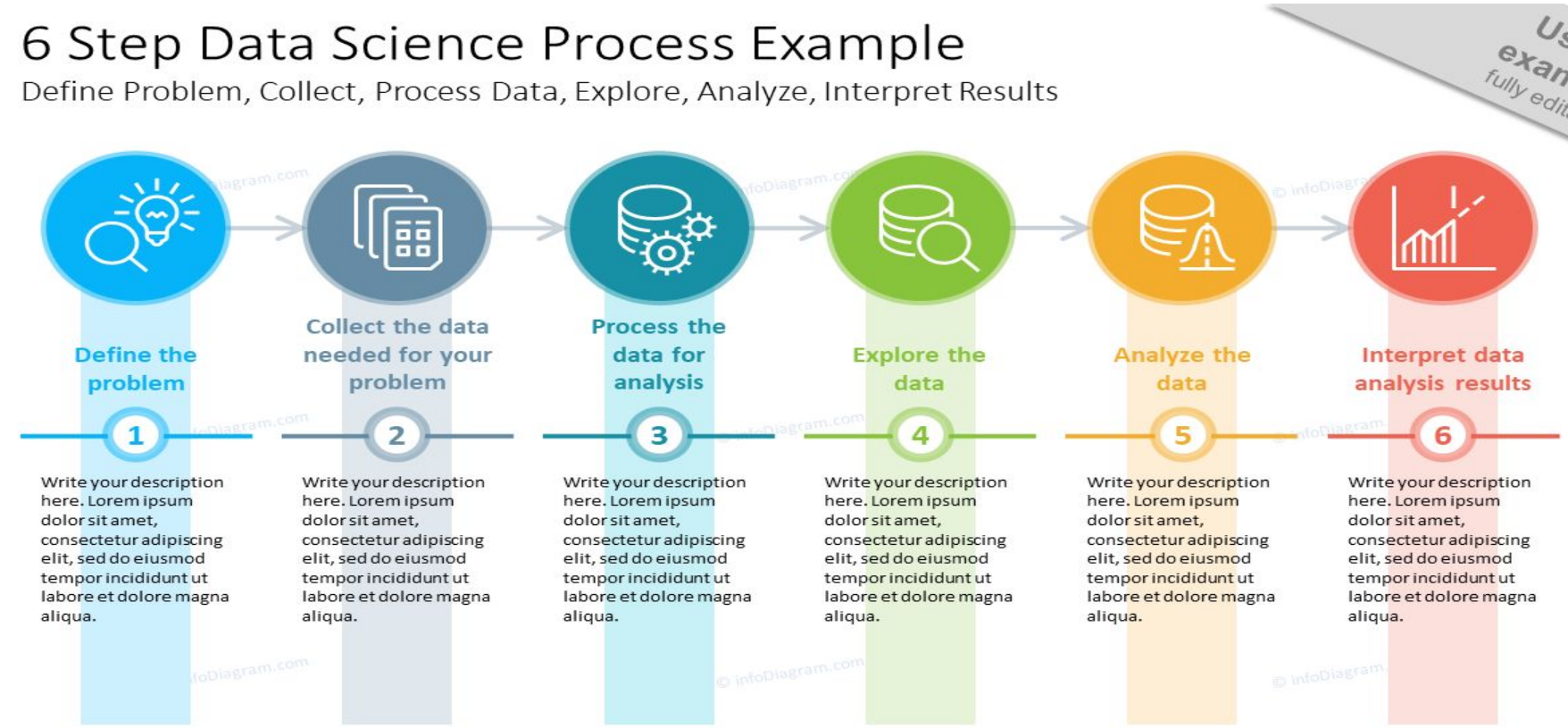
By analyzing the data of the past customers and number of variables associated to it like:

- Grade
- DTI ratio
- Owns a house
- State of residency
- Interest Rate
- Loan Amount

and many more..

It will be done at the **Evaluation Stage** where we can prevent the bank to give the loan to a customer who is most likely to be a defaulter

Process of Analysis



Reference

Data Pre Processing

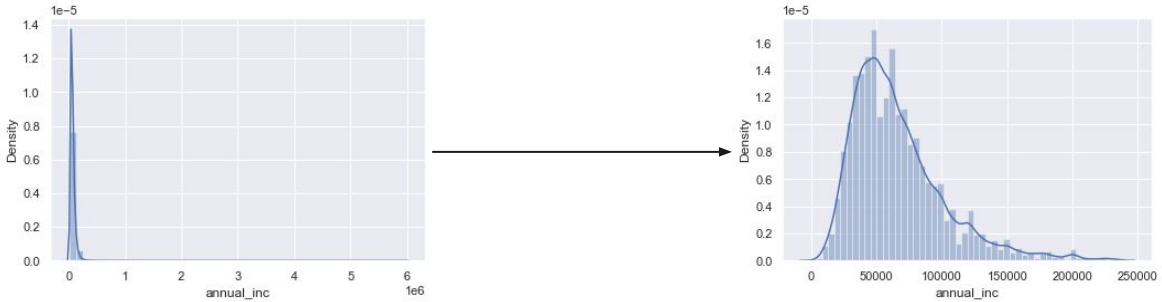
1. **Converted some columns from categorical to numerical**

   Eg: int_rate has values: 12.3 %, 11.2 %, ..

   Converted above to: 12.3, 11.2, .. (made it numerical)

2. **Removing Outliers:**

   Annual Income column had some values which was an outlier (basically too far from the 99 percentile of data). Hence it was removed.

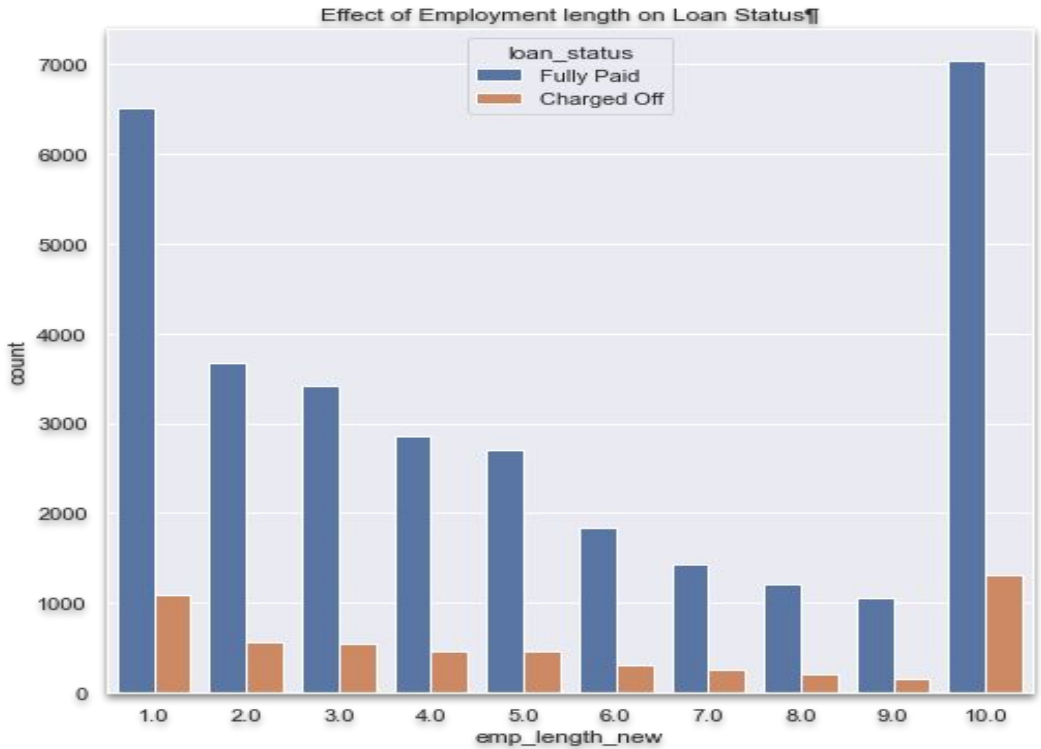


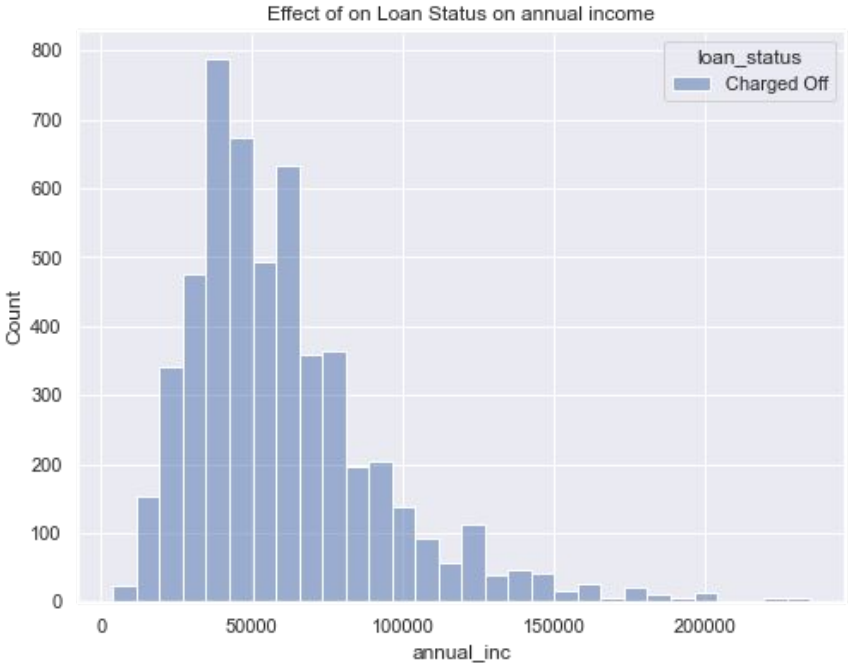
Interpretation Of Analysis

Employment vs Loan Status

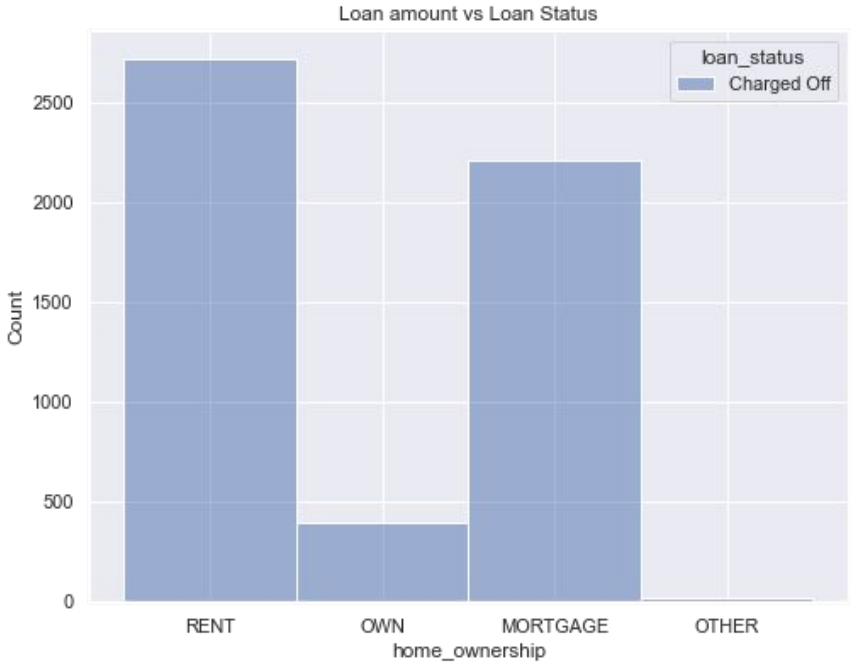
Here we can see the emp_length is a factor for the **Charged Off** loans, if not that much.

People having more than 10 years of experience tend not to pay the loans fully

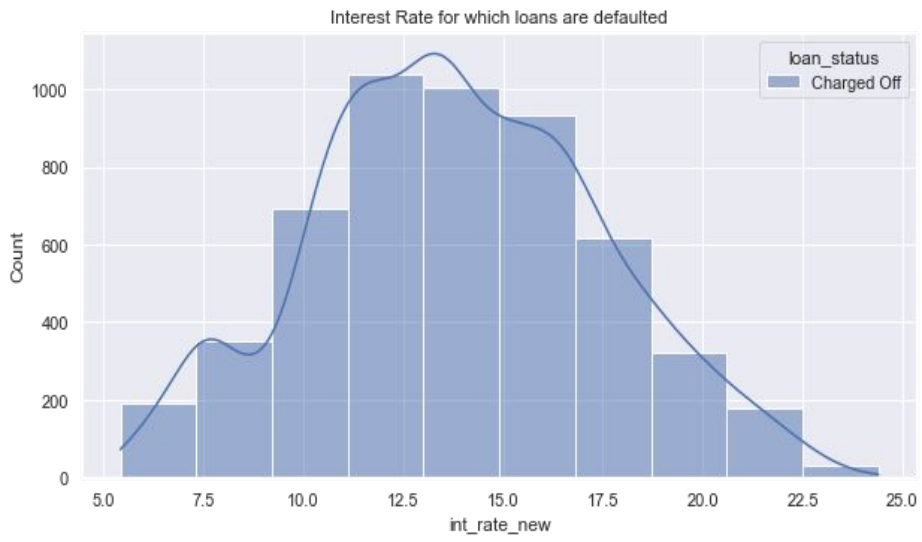




Mostly the **annual income** of person's who are **defaulters** ranges from **30k to 60k**



If the payer does not own his/her own house, then there is a **high** chance that the loan will be Charged off

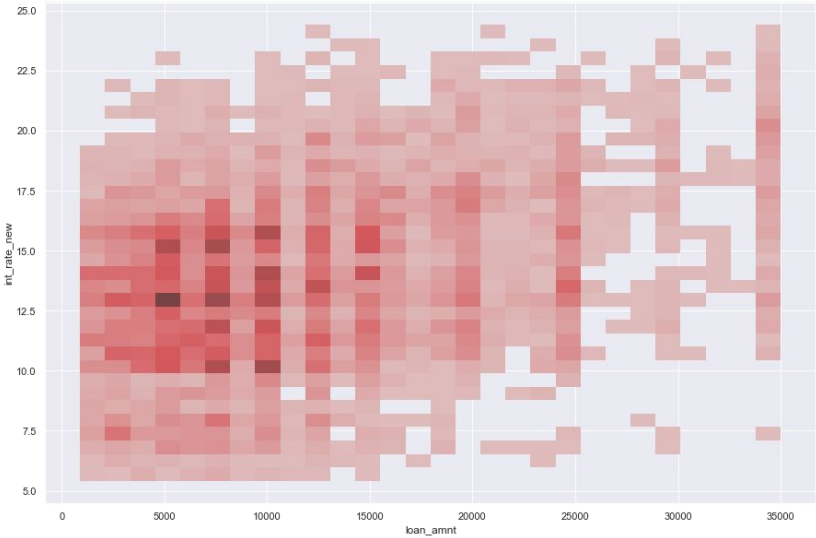


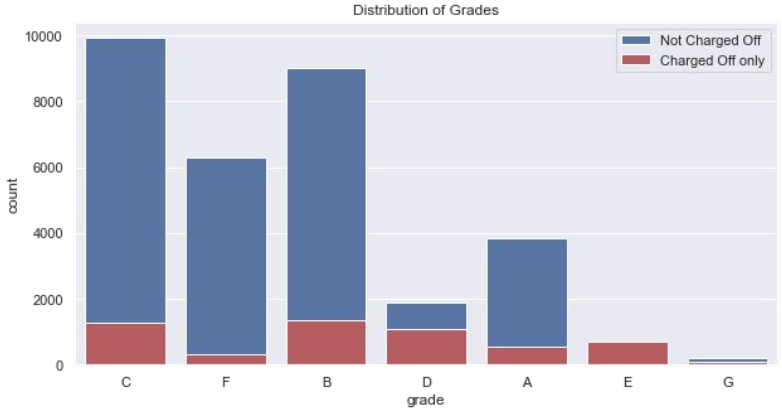
Interestingly, interest rate has a major effect on loan_status. Rates ranging from **11.0-16.0** have higher chance of getting defaulters

The interesting part comes in where

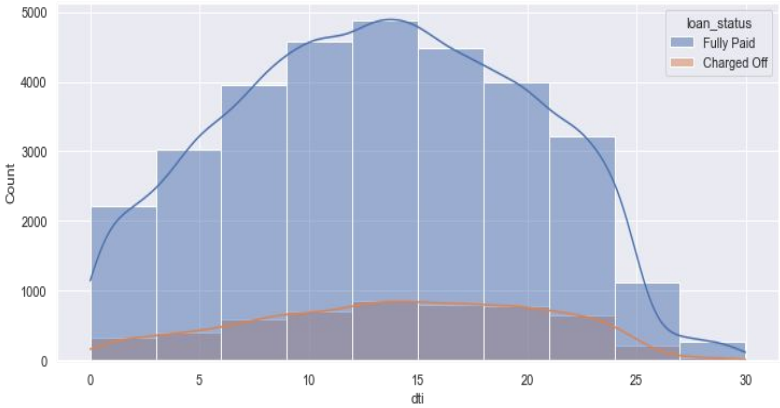
- The loan amount which is **>5K and <14k**
- The interest rate **>10 and <16**

Mostly the defaulters lie in these regions

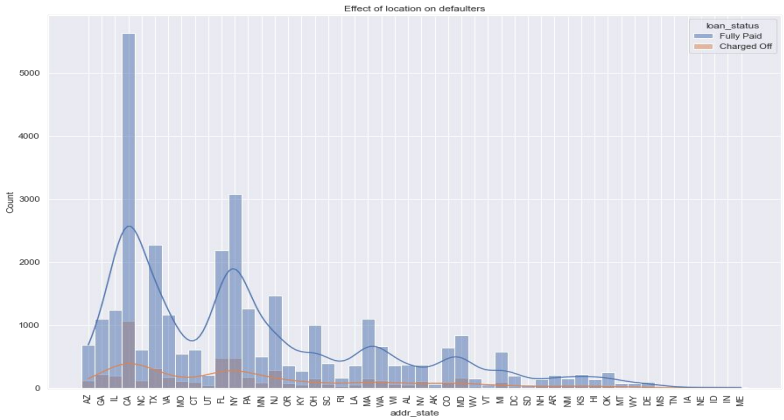




- Grade **E** is the most likely to be the loan defaulters.
- Grade **D** is also somewhat likely to be the defaulters.

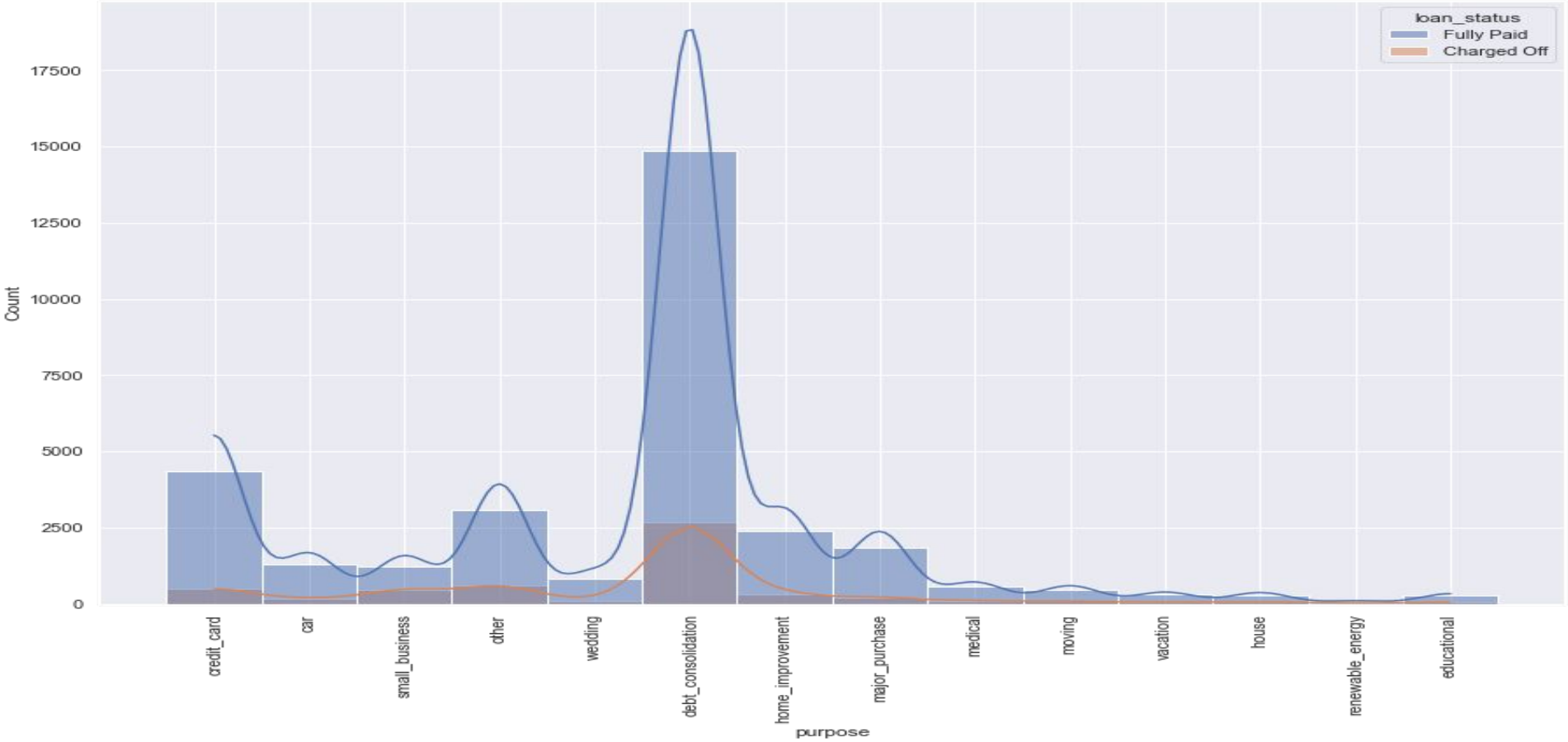


It can be seen that the loans having dti value **10-15** have higher number of defaulters. Also, higher the dti higher the chance of defaulting.



Loans from states like **CA**, **FL** and **NY** are more likely to be defaulters compared to other states

Low risk loans includes: car, wedding, medical, house. These loans are likely to be fully paid



Final Insights

- **Home Ownership** is one of the major thing to look out when giving out the loans. The persons who don't own their own house are the most defaulters.
- Customers having **Grade E** should be avoided giving loans, as there are almost 100% chance that the loan will be defaulted.
- The verifier should be **cautious** for people who have **Grade D**
- A person is most likely to be a defaulter when, the **Loan Amount** is ranging from: **5K - 14K** and the **Interest Rate** is in the range of **10 - 16.**
- The likelihood of a person to be a defaulter increases when the **DTI Ratio** starts from **10** and decreases when the **DTI Ratio** ends at **25**.
- The persons who have the following **purposes** to get the loan are **likely to return** the loan amount:
    - Car
    - Wedding
    - Medical
    - House