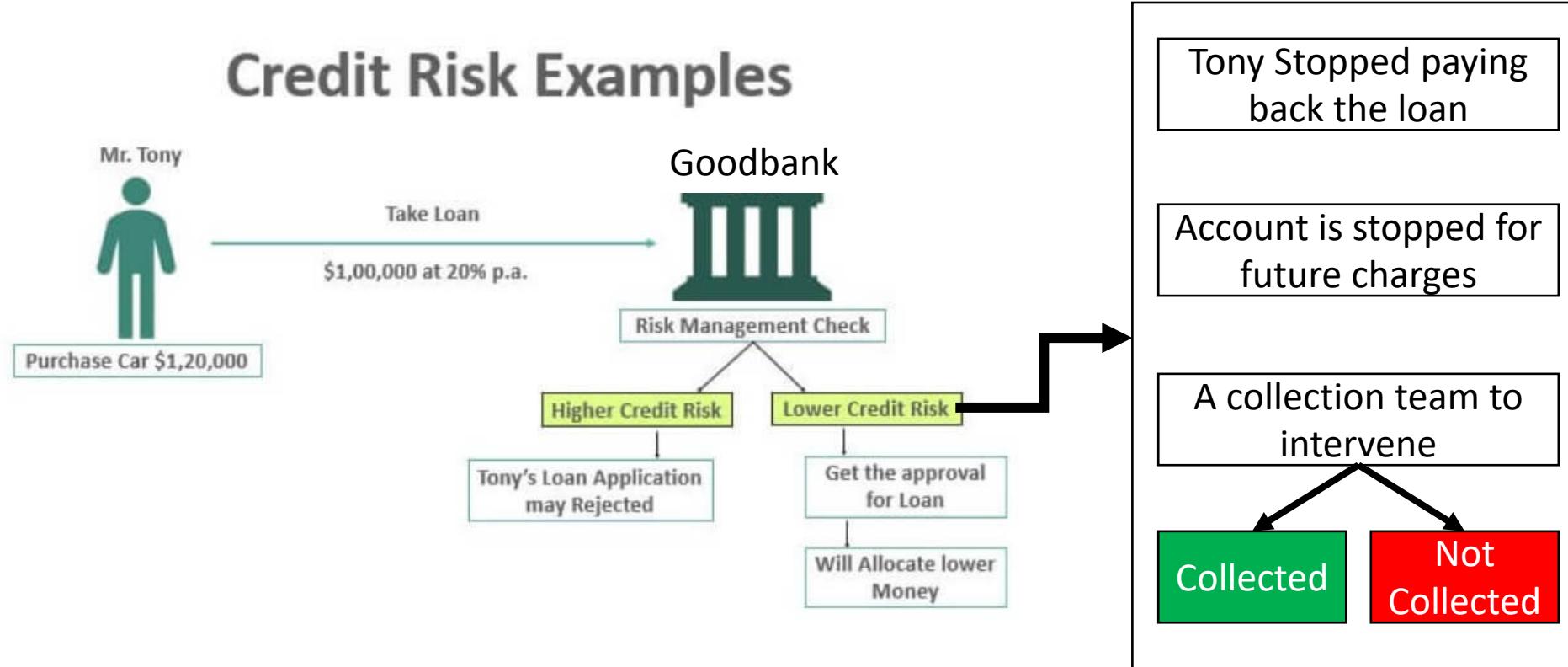


Credit Risk Analysis

Context

- Goodbank is a Canadian fintech that offers installment loans
- Automated decisions to account management.



Problem Statement

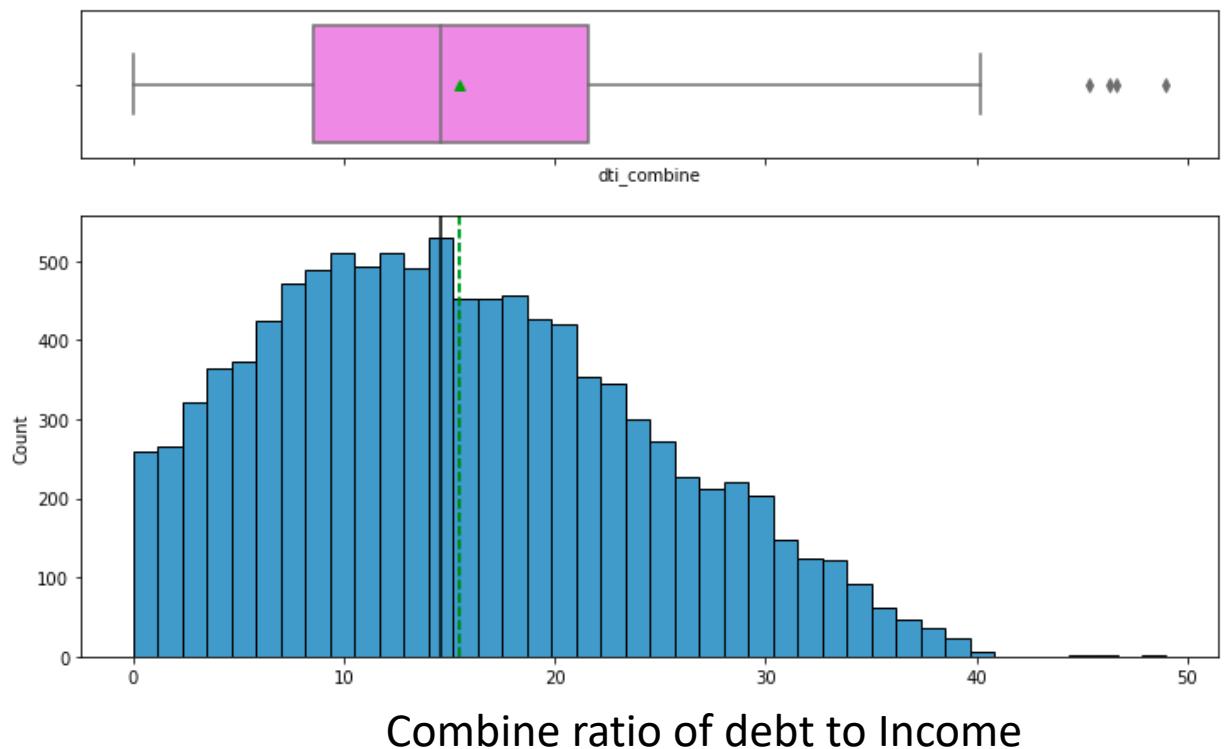
- Collections are expensive and can be more expensive than the recovery itself.
- Identify at the time of application that the collection is possible or not.
- How to improve collection efforts on charged off accounts using data?

Raw Data

- Rows and the columns of the Data: **150 columns and 2.6 million rows**
 - data provided by the customer at the time of application
 - data generated in the automated system after the approval of loan application.
- Data includes financial aspects of the customer
- Large number of missing values
- Selected Features: **25**
 - 6 categorical features
 - data generated in the automated system after the approval of loan application.

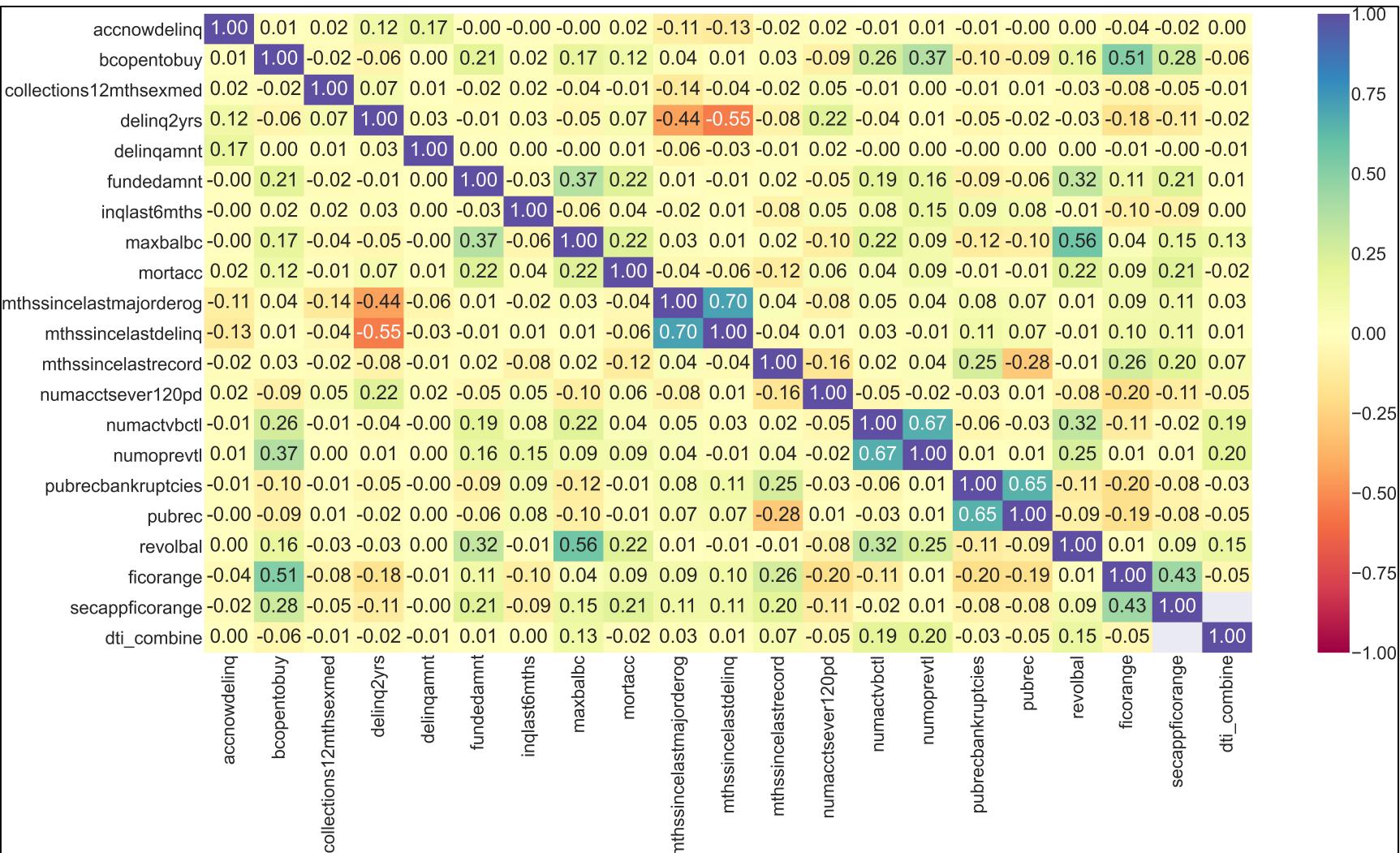
Selected Data

- 25 features are selected: 6 Categorical and 19 Numeric
 - Box plots
 - Histograms
 - Correlation Matrix
 - Intuition (
 - Null Values Count
 - Features Extraction)



Selected Data (Numerical)

- 25 features are selected: 6 Categorical and 19 Numeric

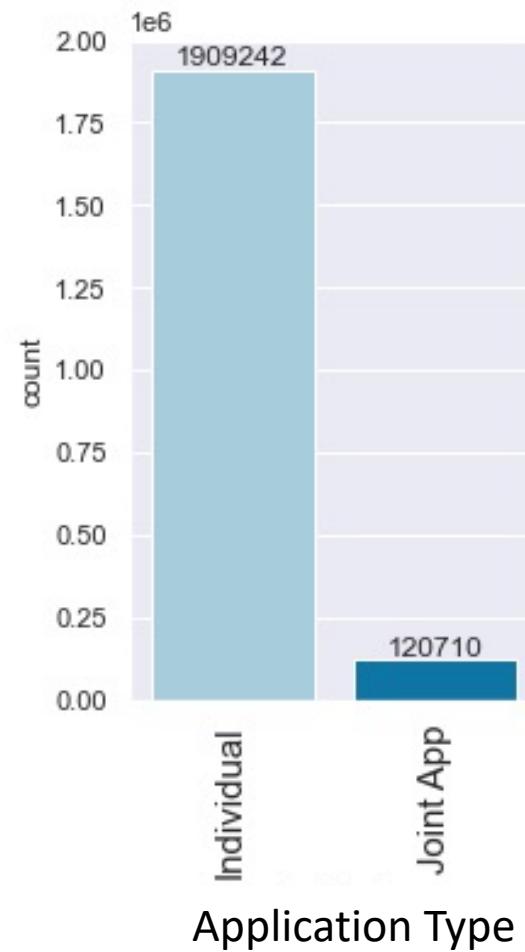


Selected Data (Categorical)

	applicationtype	emplength	homeownership	purpose	term	grade
count	2029952	1892044	2029952	2029952	2029952	2029952
unique	2	11	4	14	2	7
top	Individual	10+ years	MORTGAGE	debt_consolidation	36 months	C
freq	1909242	677437	995479	1146650	1434328	591304

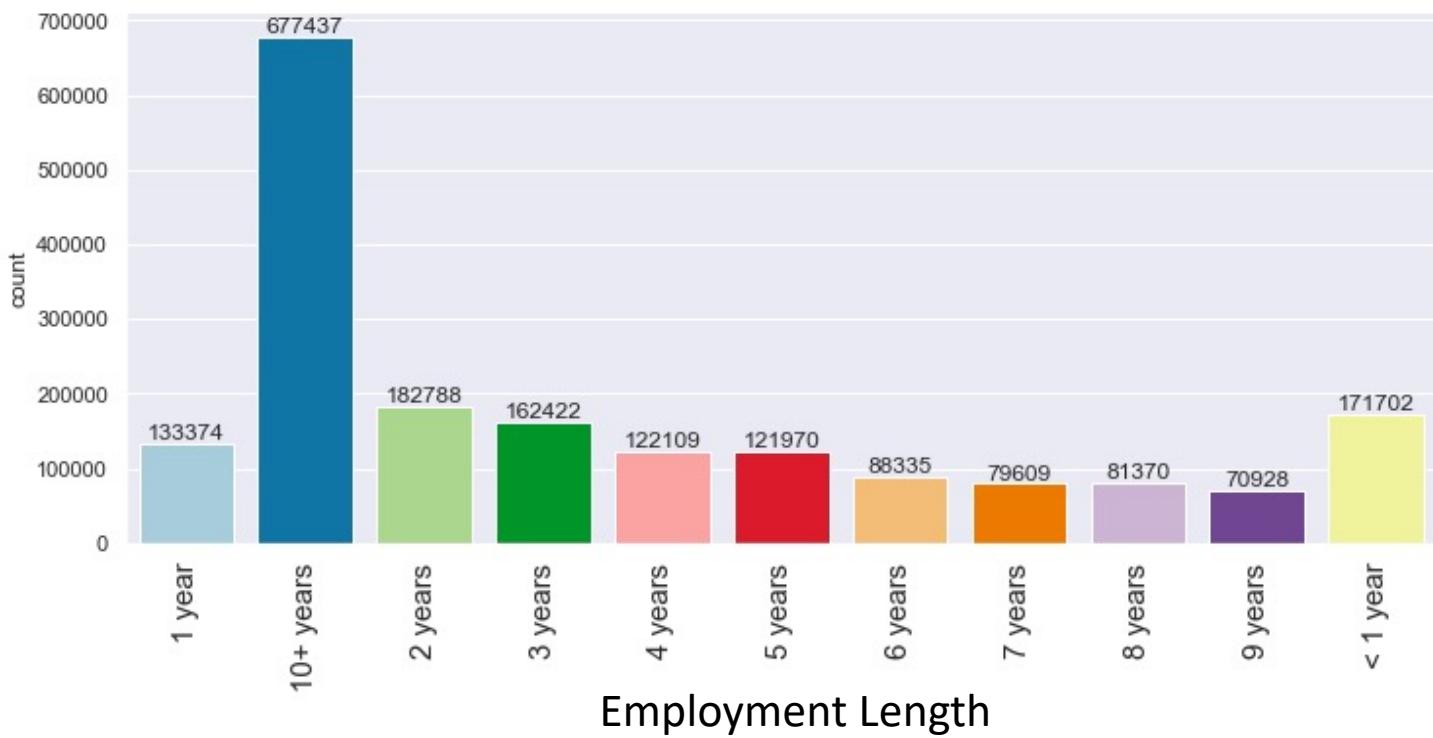
Selected Data (Categorical)

applicationtype	
count	2029952
unique	2
top	Individual
freq	1909242



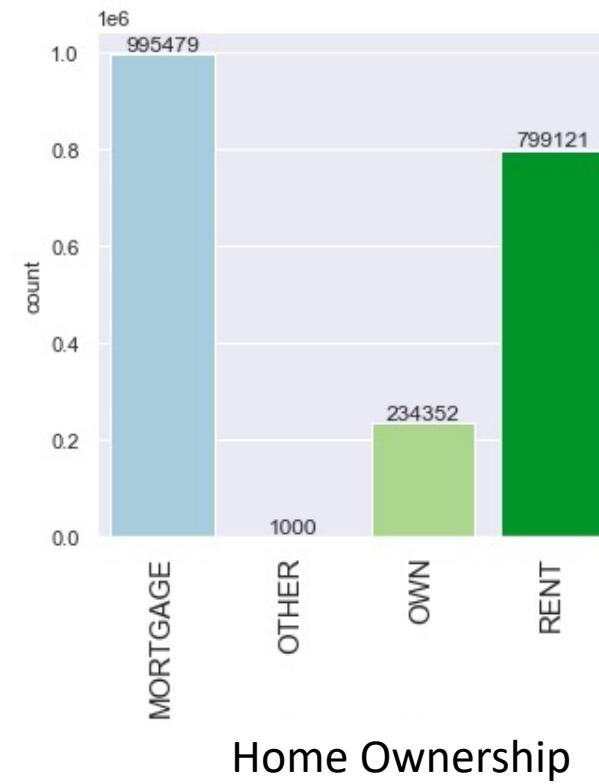
Selected Data (Categorical)

	applicationtype	emplength
count	2029952	1892044
unique	2	11
top	Individual	10+ years
freq	1909242	677437



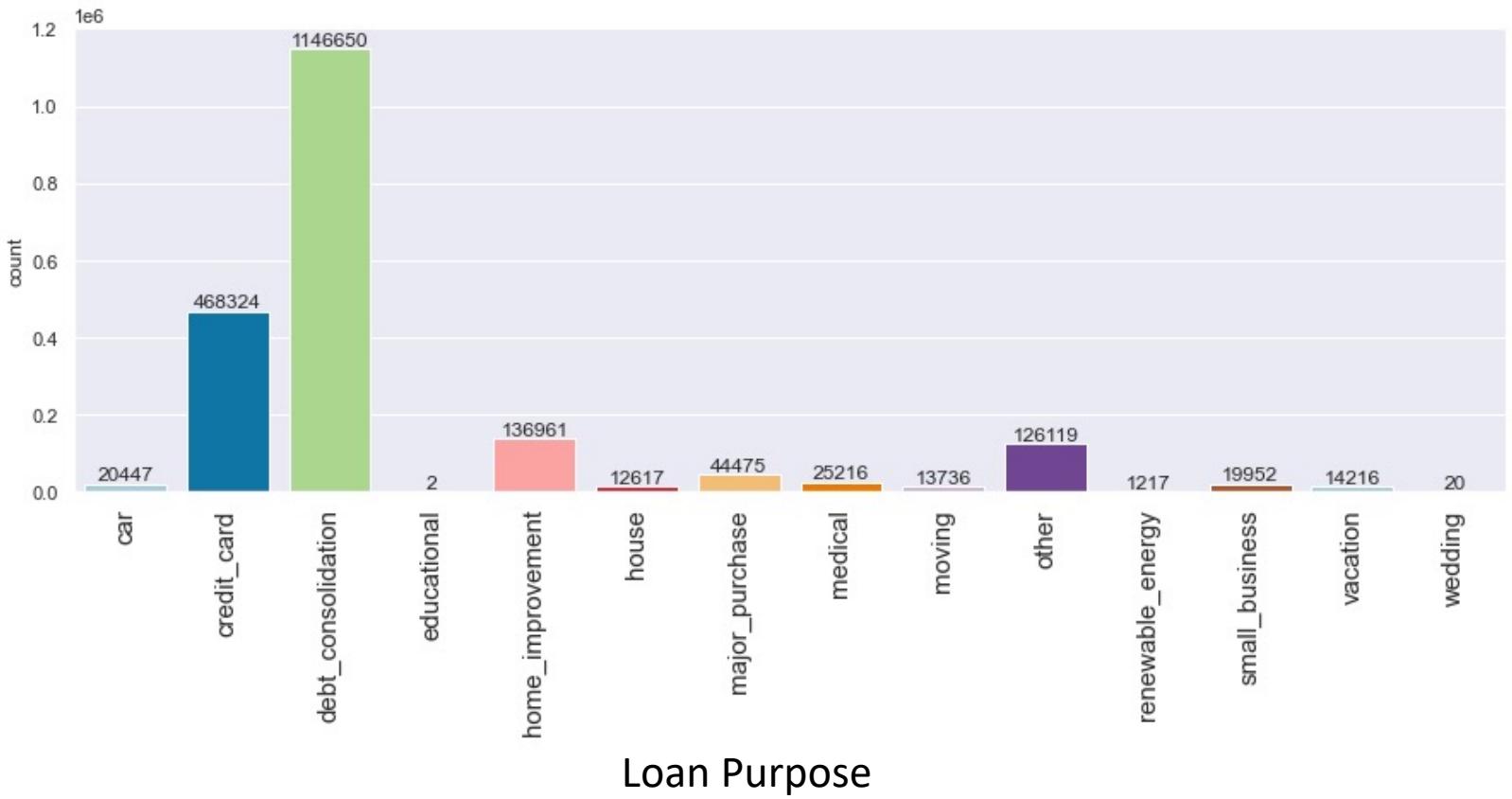
Selected Data (Categorical)

	applicationtype	emplength	homeownership
count	2029952	1892044	2029952
unique	2	11	4
top	Individual	10+ years	MORTGAGE
freq	1909242	677437	995479



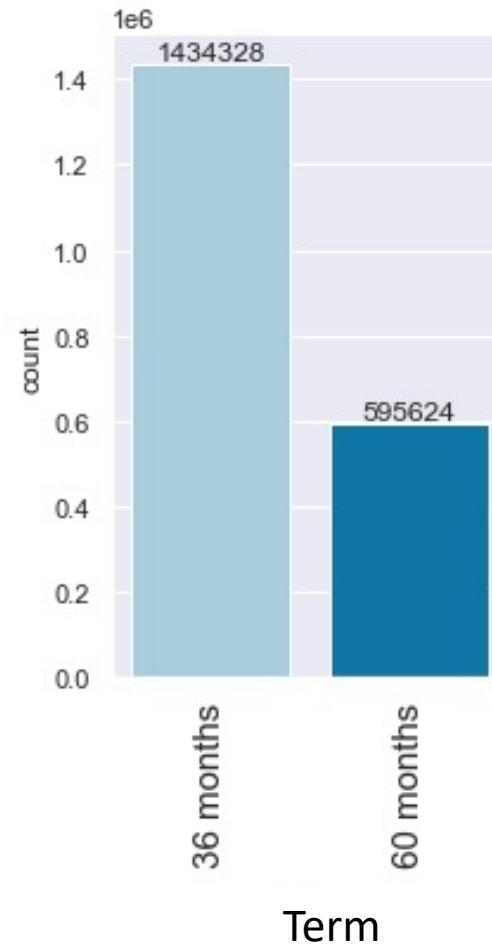
Selected Data (Categorical)

purpose	
count	2029952
unique	14
top	debt_consolidation
freq	1146650



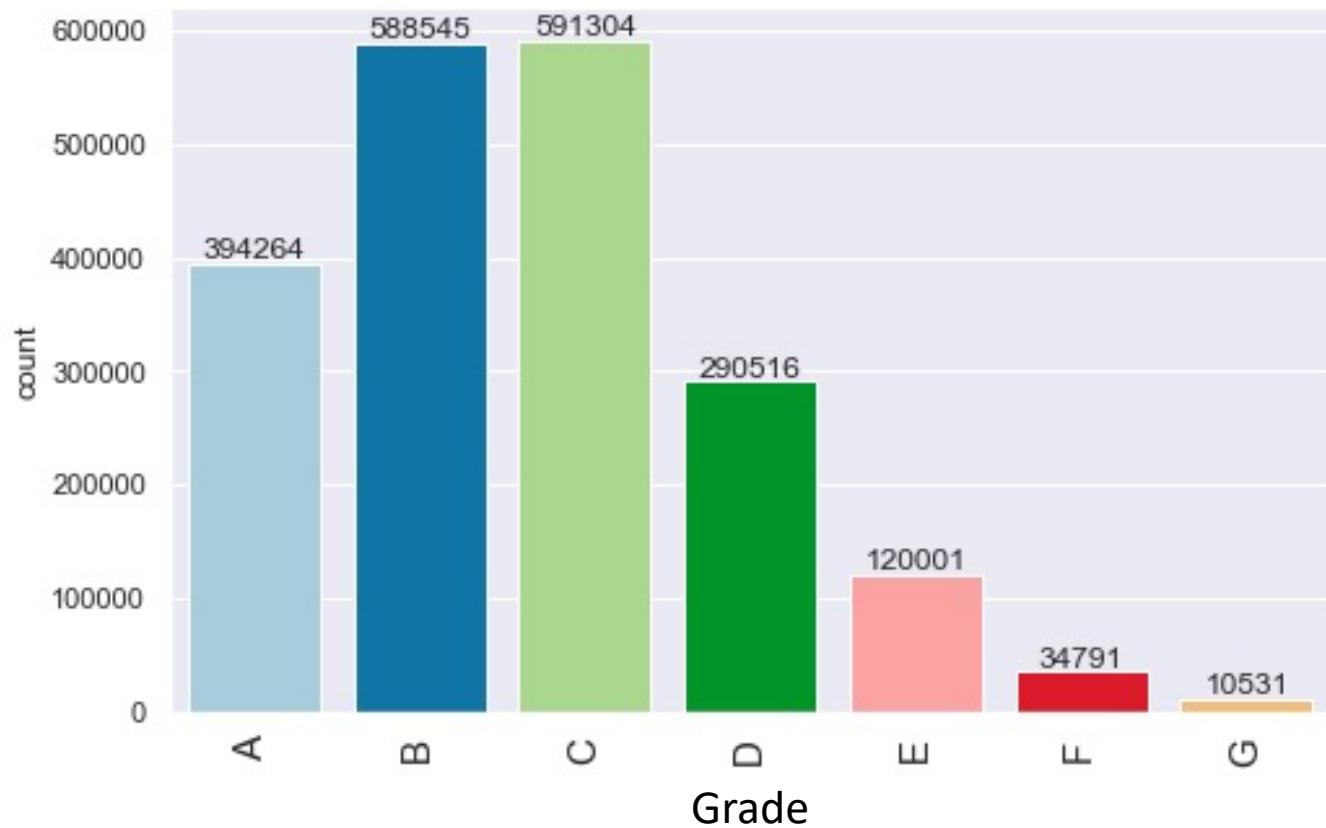
Selected Data (Categorical)

	term
count	2029952
unique	2
top	36 months
freq	1434328



Selected Data (Categorical)

grade
count 2029952
unique 7
top C
freq 591304

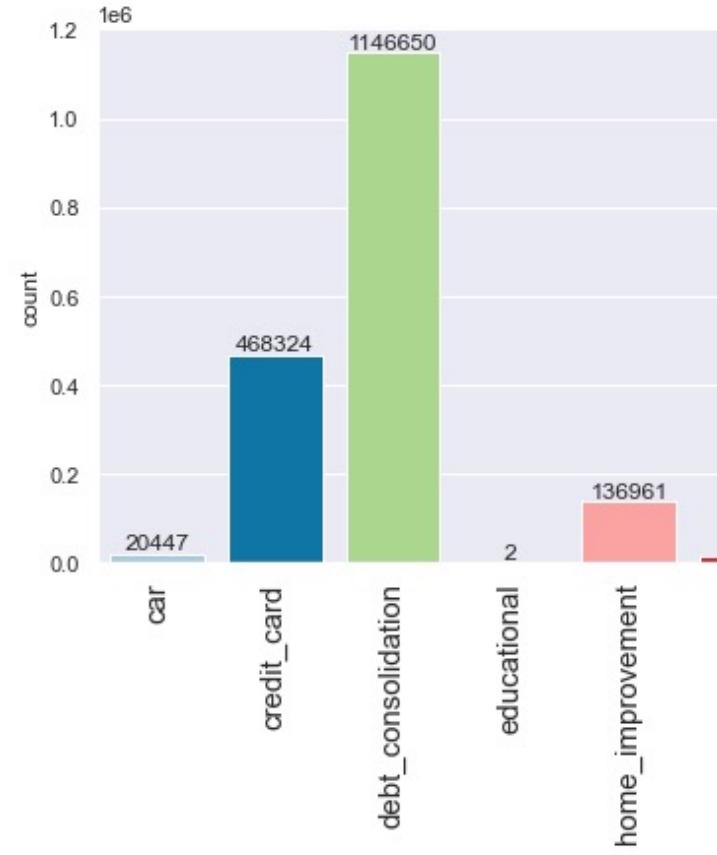


Findings

- Risk credit for each category require a separate predictive model.
 - For example, customer with loan purpose car could have different dynamics as compared to the one with debt consolidation.
 - It is possible that there are further division in the data, like debt consolidation could have different reasons. So we need to apply clustering to find out which features are more appropriate to apply the predictive model.
- Feature elimination before developing the model is intuitive. And using 25 features data is computationally expensive.

Feature Selection

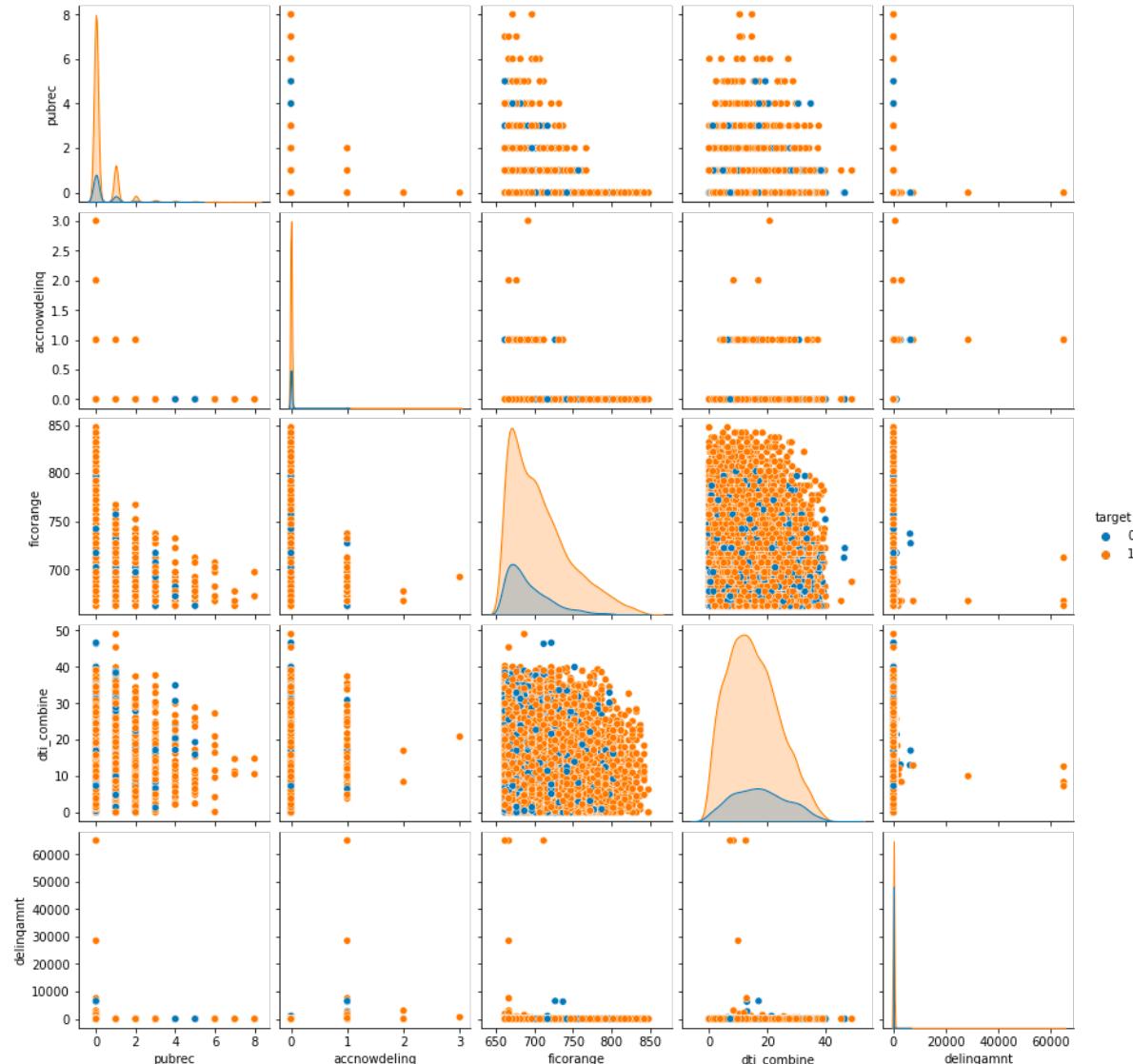
- Focus on one category of a feature “Purpose”: CAR
- There are 20447 rows data.
- Within this category, 5 variable are chosen
 - Public record
 - Number of accounts delinqt
 - Ficorange
 - Combined debt to income ratio
 - Delinq Amount
- Set the target as “Loan Status”. This is given in the data



Target variable

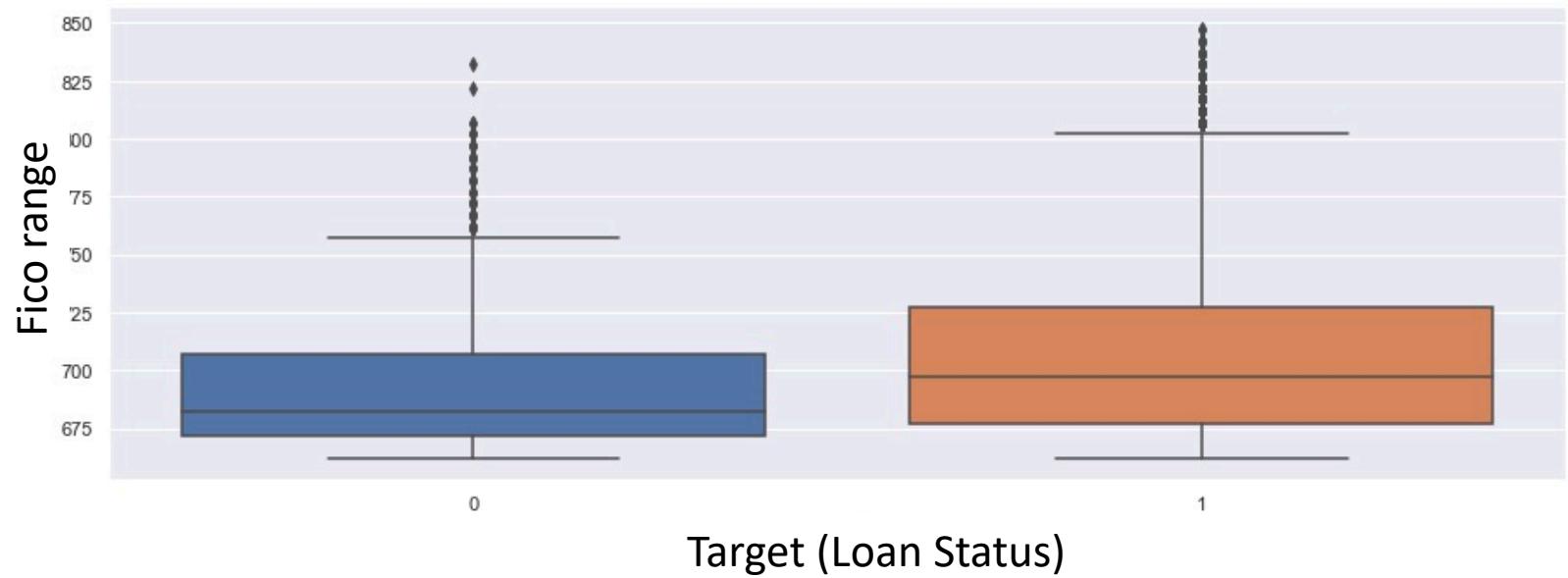
- Public record
- Number of accounts delinqt
- Ficorange
- Combined debt to income ratio
- Delinq Amount

Orange: Fully Paid
Blue: Charged Off



Fico range

Orange: Fully Paid
Blue: Charged Off



Model Preparation

- Model Predictions could be:
 - Predicting a customer will pay the loan and customer do not - Loss of resources
 - Predicting a customer will not pay the loan and the customer pays - Loss of opportunity
- Need to reduce False Negatives, Recall to be maximized, the greater the Recall lesser the chances of False Negatives.

Next steps

- Applying Machine Learning Models
- Cross validation
- Model Tuning and Regularization

Recommendations

- Important features
 - purpose of the loan
 - past delinquencies
 - ownership of property
 - an individual or a business
 - Amount of the loan required and the term required
 - Fico score of primary and secondary applicant
- Need separate model for each purpose of loan application
- Clustering is required in case of the purposes like debt consolidation and payment of the credit card.