RISK DEFAULT MANAGEMENT IN ONLINE PEER TO PEER LENDING

การจัดการความเสี่ยงในการผิดนัดชำระหนี้ส่วนบุคคลแบบออนไลน์

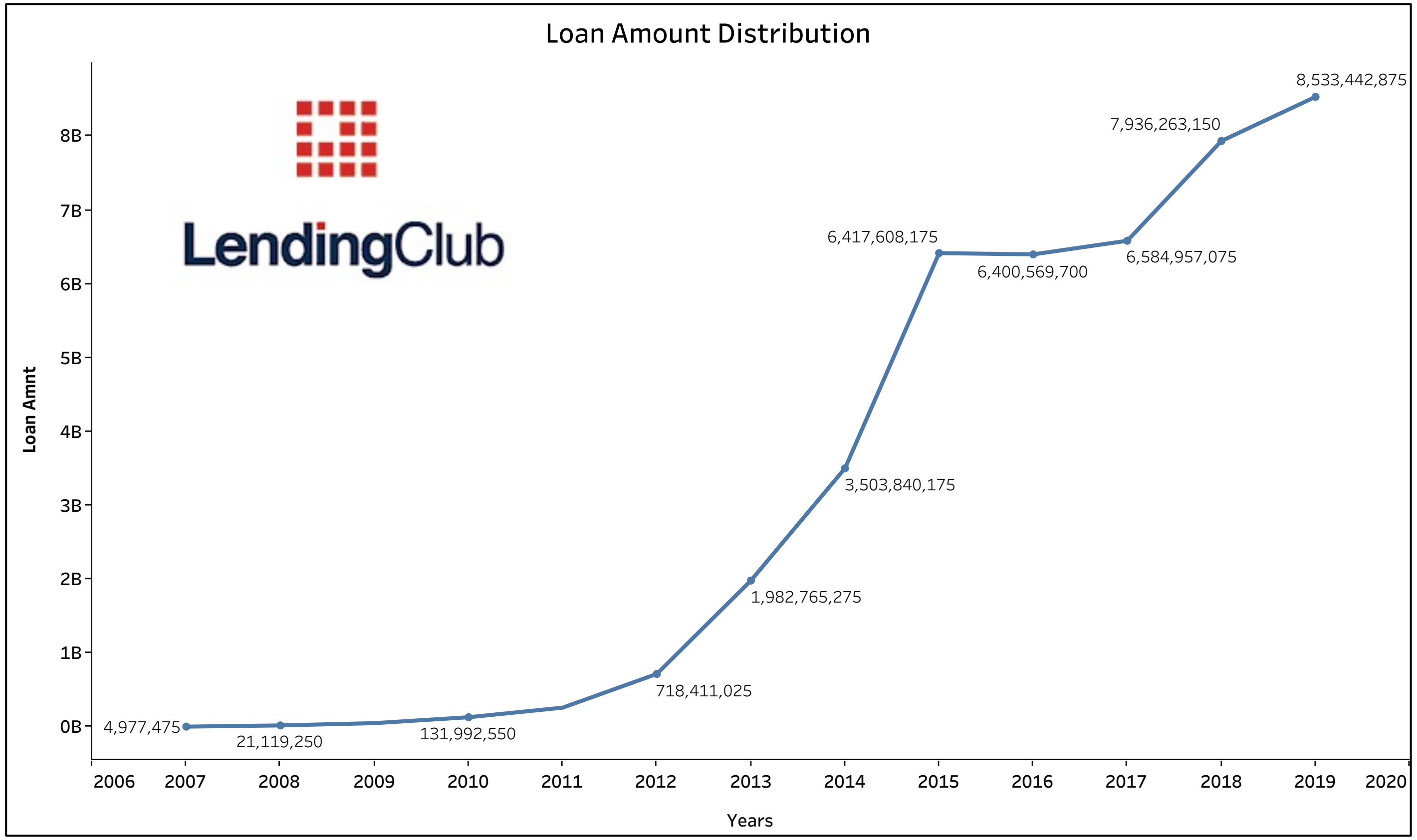
อาจารย์ที่ปรึกษา ผศ.ดร.เอกรัฐ รัฐกาญจน์

อติวิชญ์ ชนินทร์โชดึก รหัสนักศึกษา 6220412019 (DS)

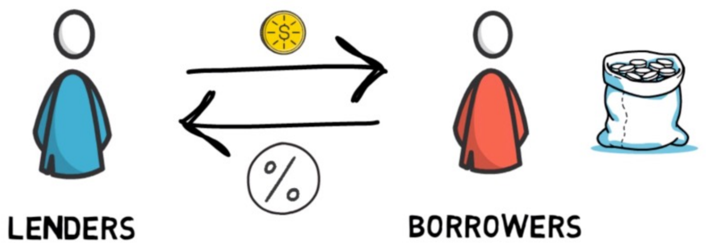
OUTLINE

- Introduction
  - Issue
  - Motivation
- Background
- Literature Review
- Problem Statement
- Methodology
- Experiment Result
- Conclusion
- Future Work
- Reference

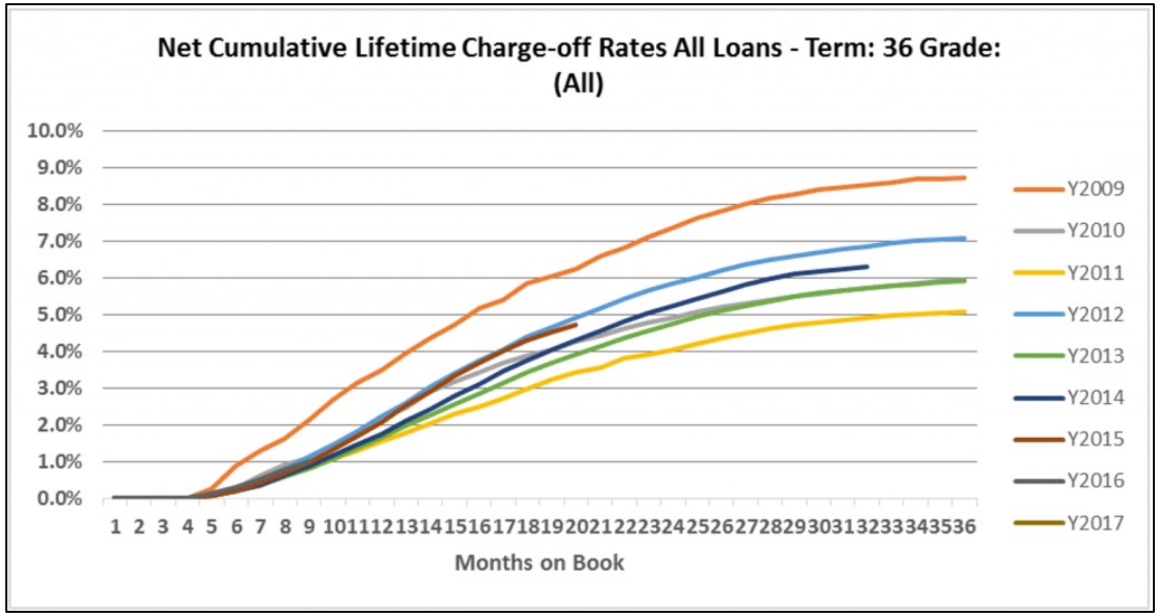
1. Introduction

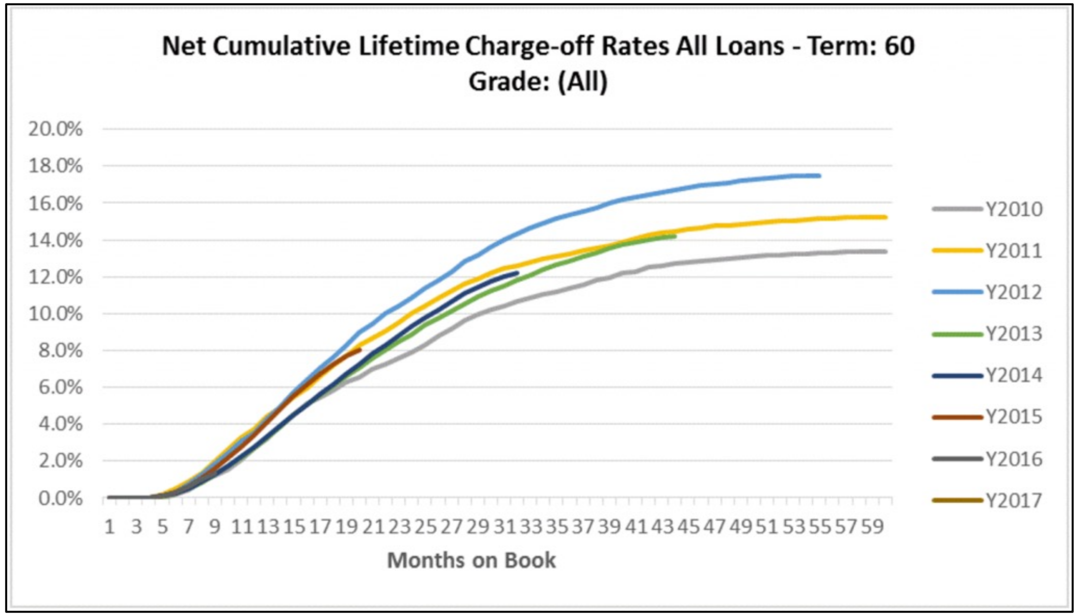


Currently, The personal loan in business online received Interest and continue to grow up because the borrower does not need to have assets or other persons to use as collateral for risks and the borrower is convenient to lend The money But the personal loan Online is a high risk that Borrower fails to pay on time make the bad debts in the future.



1. Introduction





ที่มา: https://www.lendacademy.com/lending-club-publishes-vintage-performance-data/

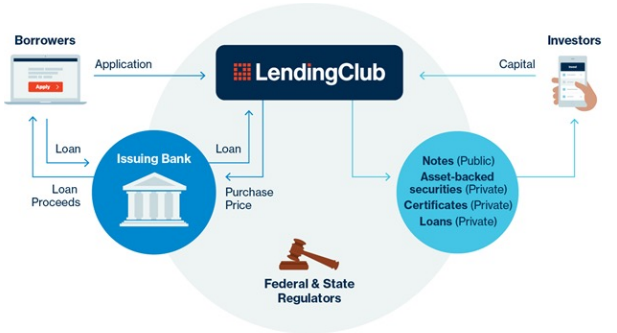
Issue

- The company cannot predict which borrowers are likely to default

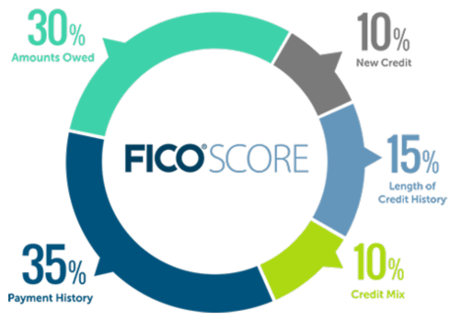
Motivation

- Using machine learning techniques to determine the pattern and predict of which borrowers are likely to default

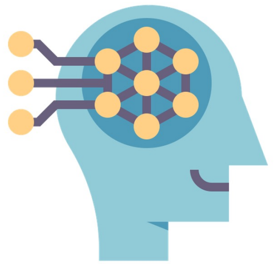
2. Background



P2P Lending



Credit Scoring



Traditional ML

- Logistic Regression
- Decision Tree
- Random Forest
- XGBoost
- LightGBM
- CatBoost
- ANN

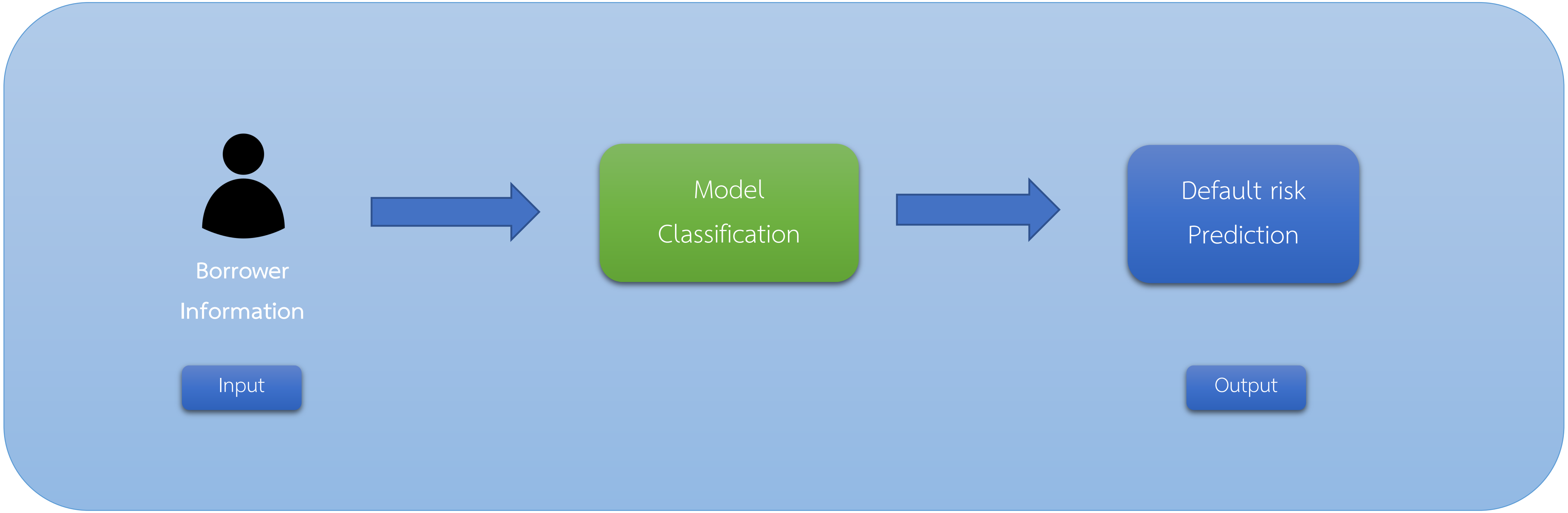
Sampling Method

- Under Sampling
- Over Sampling

3. Literature Review

Author	Topic	Dataset	#Data	#Attribute	Method
Zhiqiang Li et al. (2021)	Application of XGBoost in P2P Default Prediction	Lending club	2,260,699	40	1. Xgboost
Anahita Namvar et al. (2018)	Credit risk prediction in an imbalanced social lending environment	Lending club	66,376	43	1. Logistic Regression 2. Linear Discriminate Analysis, 3. Random Forest (Best Result)
Kim and Cho (2017)	Dempster-Shafer Fusion of Semi-supervised Learning Methods for Predicting Defaults in Social Lending	Lending club	332,844	17	1. Decision Tree
Fu (2017)	Combination of Random Forests and Neural Networks in Social Lending	Lending club	1,320,000	13	1. Combination of random forest and neural network
Zhang et al. (2017)	Determinants of loan funded successful in online P2P Lending	Paipai	193,614	21	1. Logistic regression
Serrano-Cinca and Gutiérrez-Nieto (2016)	Determinants of Default in P2P Lending	Lending club	40,907	26	1. Linear regression (Best Result) 2. decision tree

4. Problem Statement



Objective

- To create a predictive model to classify each borrower with a tendency to default
- To find the factor that influence the risk of borrowers defaulting

Dataset

LendingClub
kaggle

Dataset Lending Club (2007-2020Q3)

Total: **2,925,493 borrower** records and **141 Feature,** Borrower: 123, Co-Borrower: 15, Investor: 3

Borrower Information

- acc_now_delinq
- annual_inc
- dti
- emp_length
- fico_range_high
- fico_range_low
- grade
- initial_list_status
- loan_amnt
- mort_acc
- open_acc
- sub_grade
- Etc.

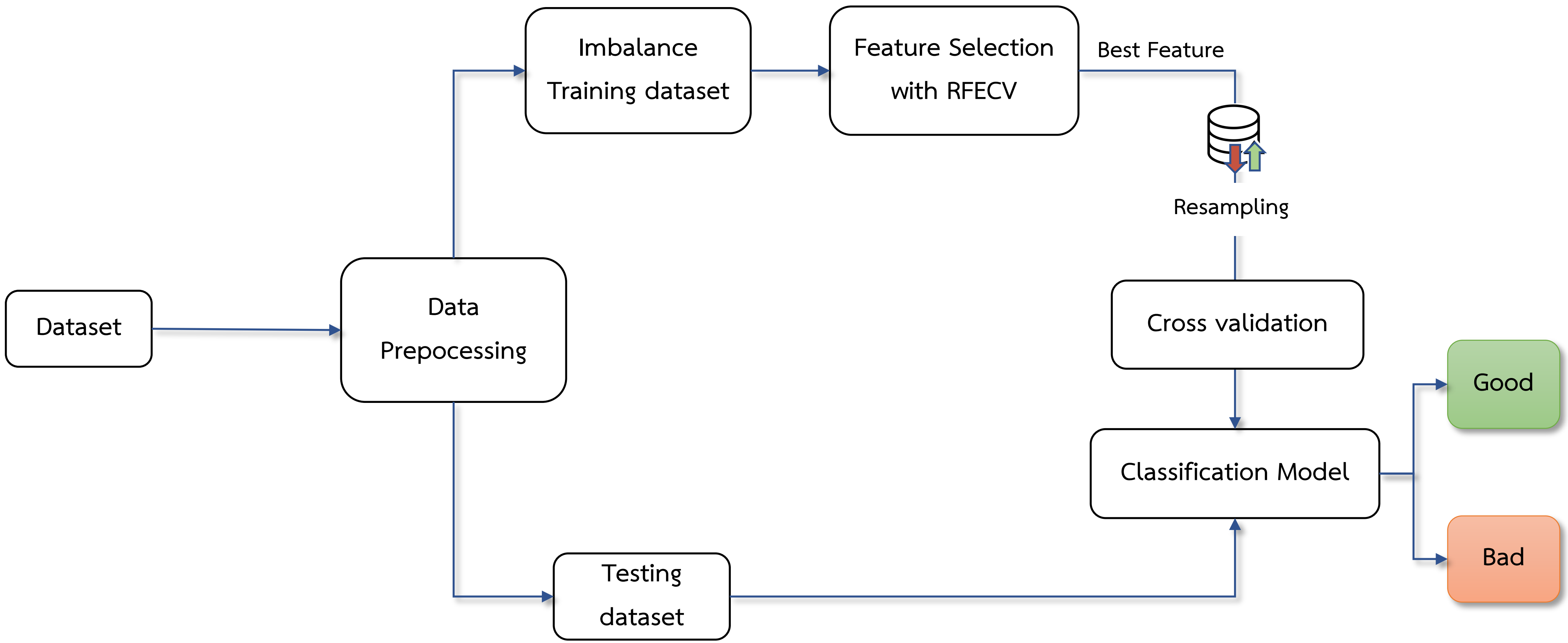
Co-borrower Information

- annual_inc_joint
- dti_joint
- revol_bal_joint
- sec_app_chargeoff_within_12_mths
- sec_app_collections_12_mths_ex_med
- sec_app_earliest_cr_line
- sec_app_fico_range_high
- sec_app_fico_range_low
- sec_app_inq_last_6mths
- sec_app_mort_acc
- sec_app_mths_since_last_major_derog
- sec_app_num_rev_accts
- Etc.

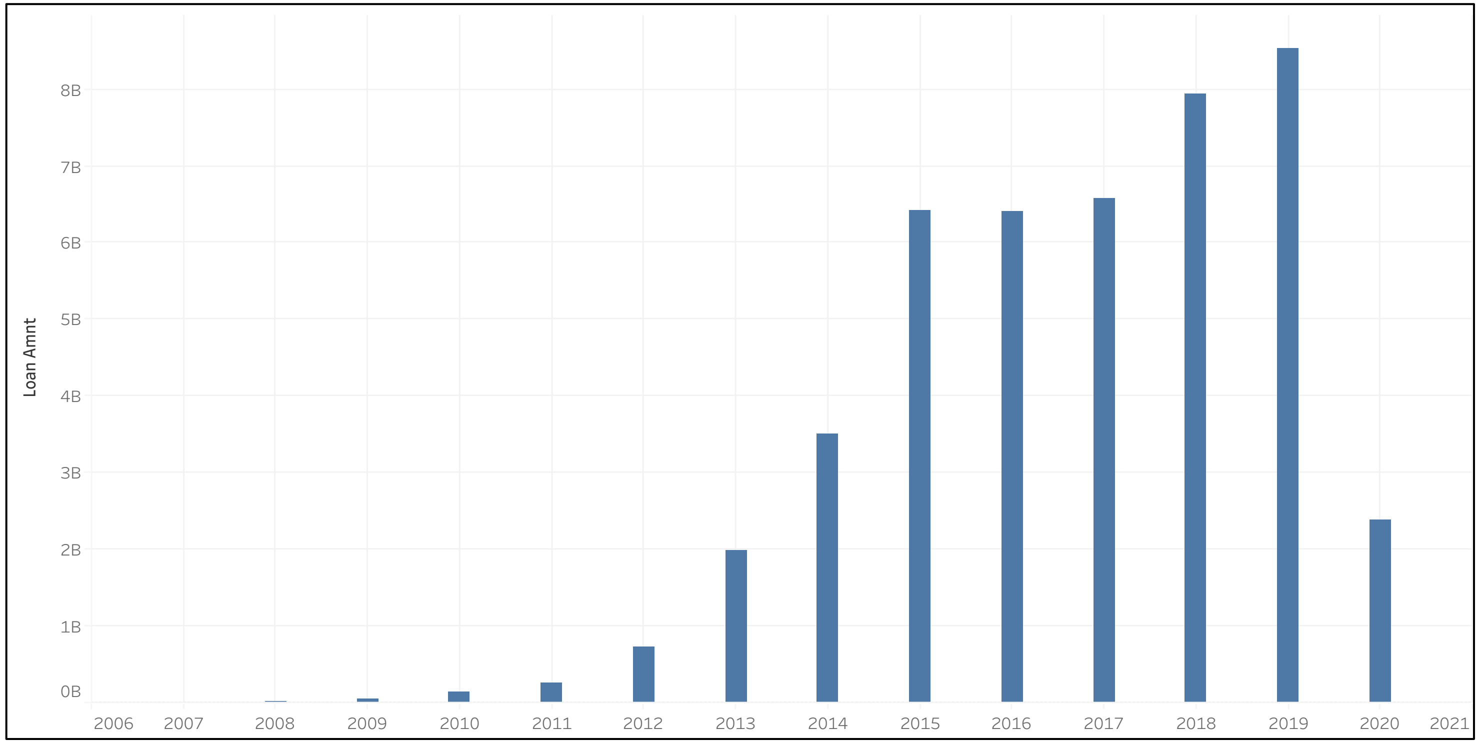
Investor Information

- funded_amnt_inv
- total_pymnt_inv
- out_prncp_inv

5. Methodology



5. Methodology (2/10)



5 years

- Raw Data
  - 141 Columns
  - 2,925,493 Row
- Filtering 5 years recently
  - 2,038,052 Row

5. Methodology (3/10)

Borrower

Variable	Description
annual_inc	Balance to credit limit on all trades
application_type	Indicates whether the loan is an individual application or a joint application with two co-borrowers
avg_cur_bal	Average current balance of all accounts
dti	A ratio calculated using the borrower's total monthly debt payments on the total debt obligations, excluding mortgage and the requested LC loan, divided by the borrower's self-reported monthly income.
Emp_length	Employment length in years. Possible values are between 0 and 10 where 0 means less than one year and 10 means ten or more years.
Emp_title	The job title supplied by the Borrower when applying for the loan.
Fico_range_low	The lower boundary ranges the borrower's FICO at loan origination belongs to.
Grade	LC assigned loan grade
home_ownership	The home ownership status provided by the borrower during registration or obtained from the credit report. Our values are RENT, OWN, MORTGAGE, OTHER
installment	The monthly payment owed by the borrower if the loan originates.
Int_rate	Interest Rate on the loan.
loan_amnt	The listed amount of the loan applied for by the borrower. If at some point in time, the credit department reduces the loan amount, then it will be reflected in this value.
Loan_status	Current status of the loan.
mort_acc	Number of mortgage accounts.
Pub_rec_bankruptcies	Number of public record bankruptcies
purpose	A category provided by the borrower for the loan request.
Revol_bal	Total credit revolving balance
revol_util	Revolving line utilization rate, or the amount of credit the borrower is using relative to all available revolving credit.
Sub_grade	LC assigned loan subgrade
term	The number of payments on the loan. Values are in months and can be either 36 or 60.
Verification_status	Indicates if income was verified by LC, not verified, or if the income source was verified
issue_d	The month which the loan was funded
last_pymnt_d	Last month payment was received
open_acc	Number of open trades at time of application for the secondary applicant
total_acc	The total number of credit lines currently in the borrower's credit file

Co - Borrower

Variable	Description
annual_inc_joint	The combined self-reported annual income provided by the co-borrowers during registration
dti_joint	A ratio calculated using the co-borrowers' total monthly payments on the total debt obligations, excluding mortgages and the requested LC loan, divided by the co-borrowers' combined self-reported monthly income
revol_bal_joint	Sum of revolving credit balance of the co-borrowers, net of duplicate balances
sec_app_fico_range_low	FICO range (high) for the secondary applicant
sec_app_mort_acc	Number of mortgage accounts at time of application for the secondary applicant
sec_app_open_acc	Number of open trades at time of application for the secondary applicant

5. Methodology (4/10)

- Data Overview

```
Sum of null values in each feature:
-----------------------------------
annual_inc                        0
application_type                  0
avg_cur_bal                     100
dti                            3106
emp_length                   160385
emp_title                    212613
fico_range_low                    0
grade                             0
home_ownership                    0
installment                       0
int_rate                          0
loan_amnt                         0
loan_status                       0
mort_acc                          0
pub_rec_bankruptcies              0
purpose                           0
revol_bal                         0
revol_util                     2158
sub_grade                         0
term                              0
verification_status               0
issue_d                           0
last_pymnt_d                   4251
open_acc                          0
total_acc                         0
annual_inc_joint            1828050
dti_joint                   1828052
revol_bal_joint             1840229
sec_app_fico_range_low      1840228
sec_app_mort_acc            1840228
sec_app_open_acc            1840228
dtype: int64
```

```
annual_inc                 float64
application_type            object
avg_cur_bal                float64
dti                        float64
emp_length                  object
emp_title                   object
fico_range_low             float64
grade                       object
home_ownership              object
installment                float64
int_rate                   float64
loan_amnt                  float64
loan_status                 object
mort_acc                   float64
pub_rec_bankruptcies       float64
purpose                     object
revol_bal                  float64
revol_util                 float64
sub_grade                   object
term                        object
verification_status         object
issue_d                     object
last_pymnt_d                object
open_acc                   float64
total_acc                  float64
annual_inc_joint           float64
dti_joint                  float64
revol_bal_joint            float64
sec_app_fico_range_low     float64
sec_app_mort_acc           float64
sec_app_open_acc           float64
dtype: object
```

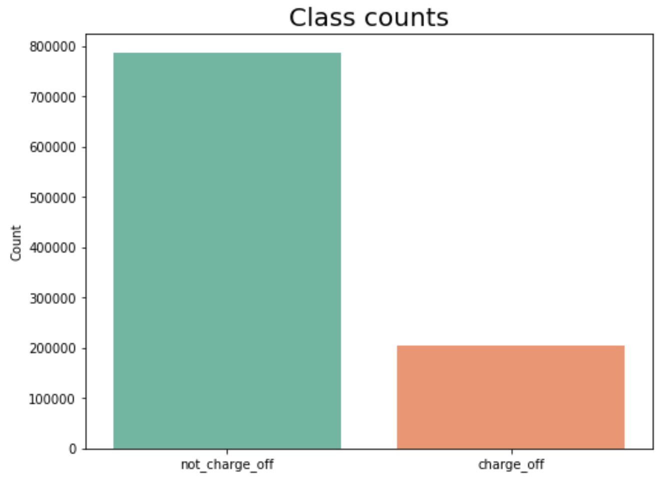
2,038,052 rows 31 columns

5. Methodology (5/10)

- Select target values

Loan status

Current	1013507
Fully Paid	788161
Charged Off	205768
Late (31-120 days)	15761
In Grace Period	9738
Late (16-30 days)	2643
Issued	2062
Default	412



Positive examples = 788,161

Negative examples = 205,768

Proportion of pos to neg examples = 383.03%

993,929 rows 31 columns

- Fully paid: Loan has been fully repaid
- Default: Loan has not been current for 121 days or more
- Charged off: Loan for which there is no longer a reasonable expectation of further payments. Generally, Charge Off occurs no later than 30 days after the Default status is reached
- We have changed Fully paid as 0 and Charged Off as 1 where 1 indicates the borrower as a defaulter

ที่มา: https://help.lendingclub.com/hc/en-us/articles/216109367-What-do-the-different-Note-statuses-mean-

5. Methodology - Feature Engineering

Calculate New Debt to incom Ratio

$$\text{New DTI} = \frac{\text{New monthly repayment amount}}{\text{monthly income}}$$

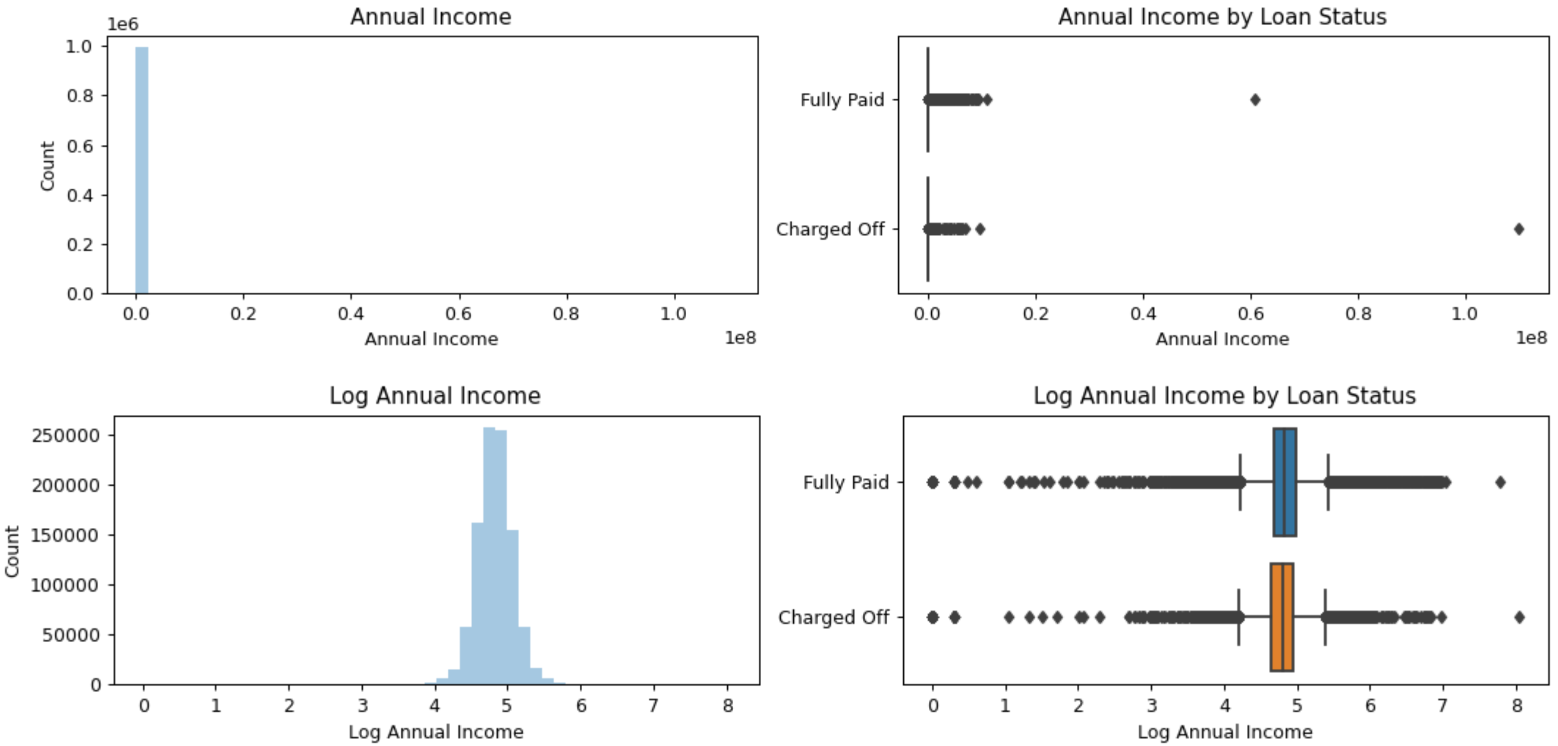
$$\text{New monthly repayment amount} = \left(\frac{installment}{Annual\ income} * Annual\ income\right) + \text{installment}$$

open_acc	total_acc	annual_inc_joint	dti_joint	revol_bal_joint	sec_app_fico_range_low	sec_app_mort_acc	sec_app_open_acc	new_dti
9.0	16.0	NaN	NaN	NaN	NaN	NaN	NaN	27.847394
7.0	10.0	NaN	NaN	NaN	NaN	NaN	NaN	18.859829
8.0	34.0	155000.0	18.73	28294.0	600.0	1.0	10.0	20.388134
10.0	23.0	NaN	NaN	NaN	NaN	NaN	NaN	22.687991
24.0	50.0	NaN	NaN	NaN	NaN	NaN	NaN	38.103979
25.0	68.0	NaN	NaN	NaN	NaN	NaN	NaN	14.418629
14.0	31.0	NaN	NaN	NaN	NaN	NaN	NaN	9.572040
7.0	16.0	50000.0	36.70	18909.0	655.0	1.0	7.0	42.239170

5. Methodology - Exploratory Data Analysis

1. Annual Income

```
count    9.939290e+05
mean     7.954268e+04
std      1.494380e+05
min      0.000000e+00
25%      4.700000e+04
50%      6.600000e+04
75%      9.500000e+04
max      1.100000e+08
```



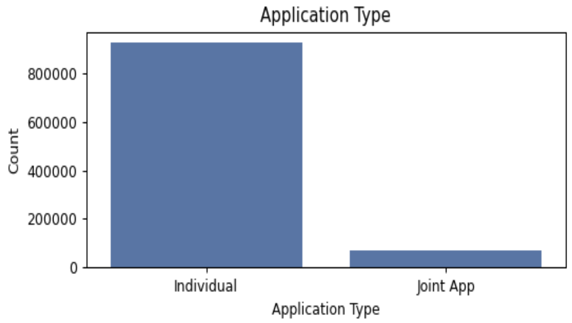

```
count    993929.000000
mean          4.819808
std           0.294266
min           0.000000
25%           4.672107
50%           4.819551
75%           4.977728
max           8.041393
```

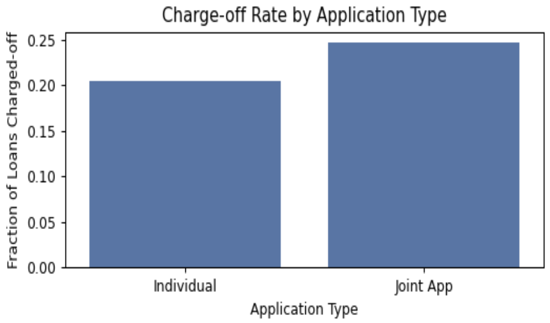
loan_status	count	mean	std	min	25%	50%	75%	max
Charged Off	205768.0	4.791363	0.290040	0.0	4.653222	4.795717	4.949395	8.041393
Fully Paid	788161.0	4.827234	0.294908	0.0	4.681250	4.832515	4.982276	7.785330

5. Methodology - Exploratory Data Analysis

2. Application type

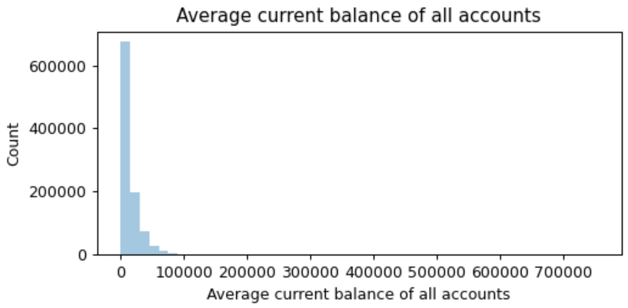
Individual	924879
Joint App	69050

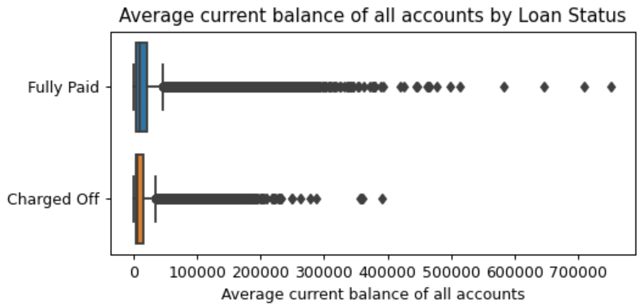




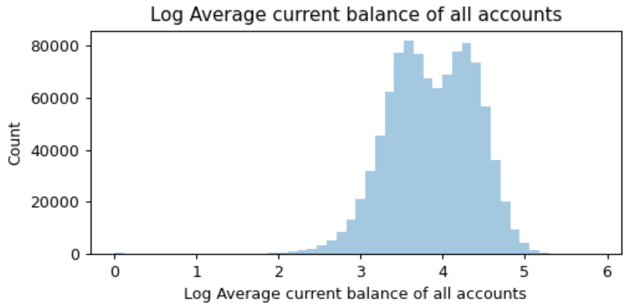
3. Average current balance of all accounts

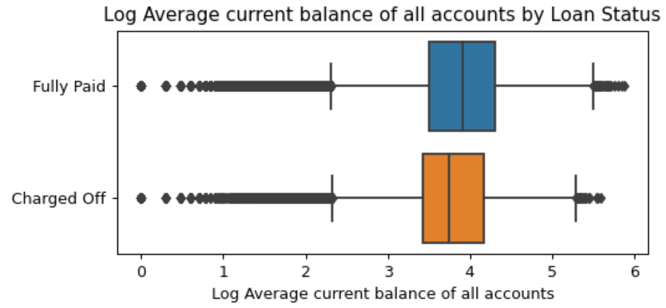
count	993896.000000
mean	13818.346037
std	16933.638951
min	0.000000
25%	3072.000000
50%	7440.000000
75%	19128.000000
max	752994.000000





count	993896.000000
mean	3.862554
std	0.535163
min	0.000000
25%	3.487563
50%	3.871631
75%	4.281692
max	5.876792

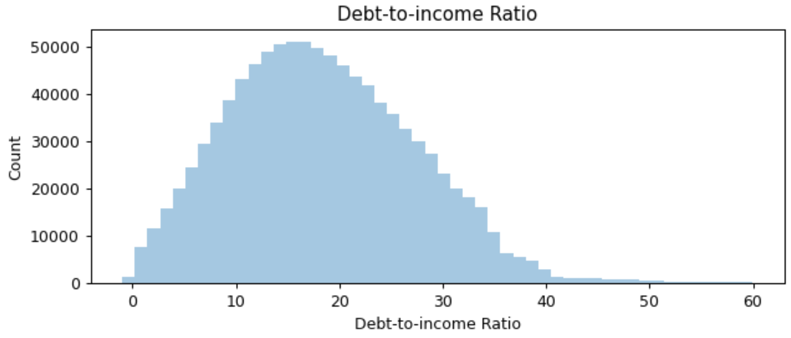


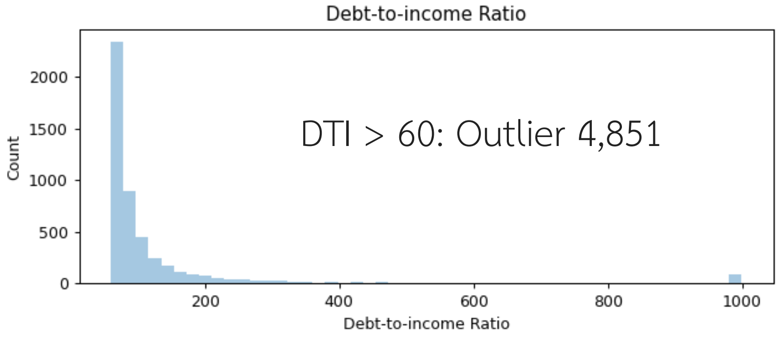


5. Methodology - Exploratory Data Analysis

4. Debt to income ratio

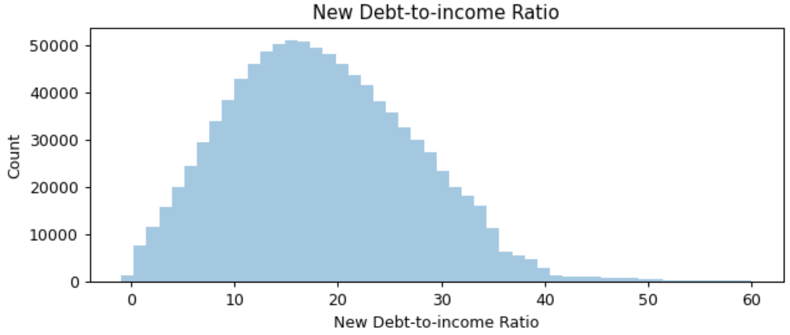
count	992823.000000
mean	18.989967
std	16.095942
min	-1.000000
25%	11.780000
50%	17.820000
75%	24.670000
max	999.000000

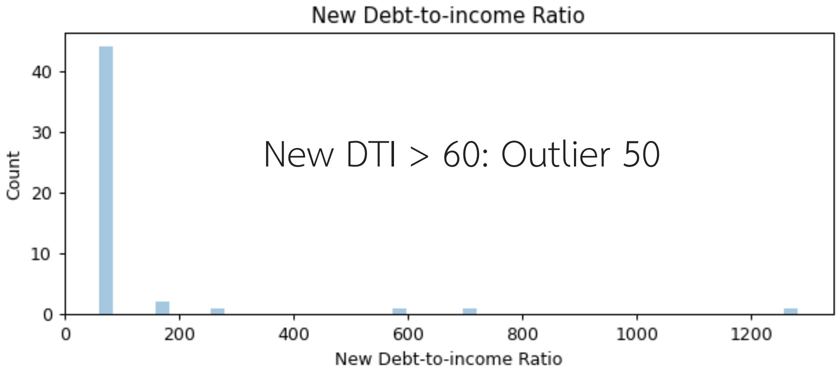




5. New Debt to income ratio

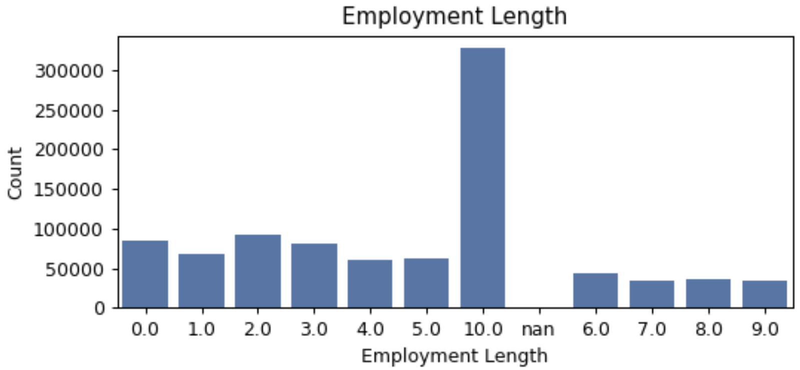
count	987972.000000
mean	18.529589
std	9.219309
min	-0.938436
25%	11.815078
50%	17.835926
75%	24.637410
max	1281.900000

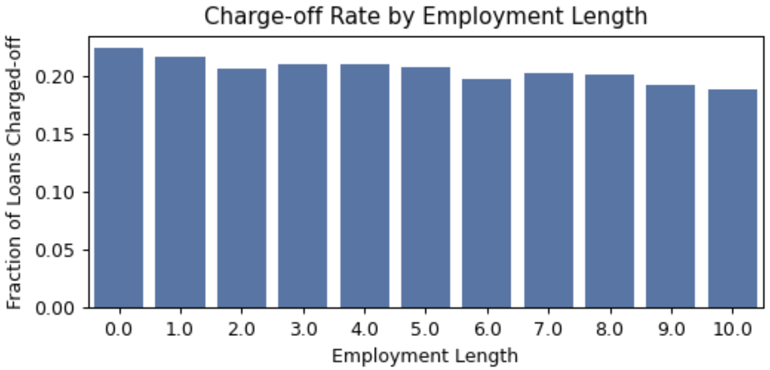




6. Employee Length

0 years	84953
1 year	67229
10+ years	327277
2 years	91444
3 years	80830
4 years	60860
5 years	61355
6 years	43257
7 years	34404
8 years	35765
9 years	33341
NaN	73214

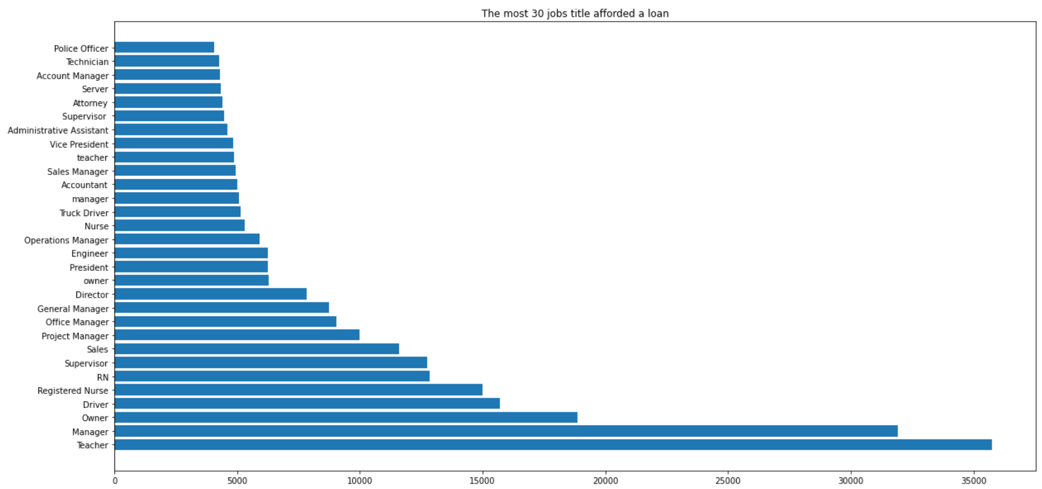




5. Methodology - Exploratory Data Analysis

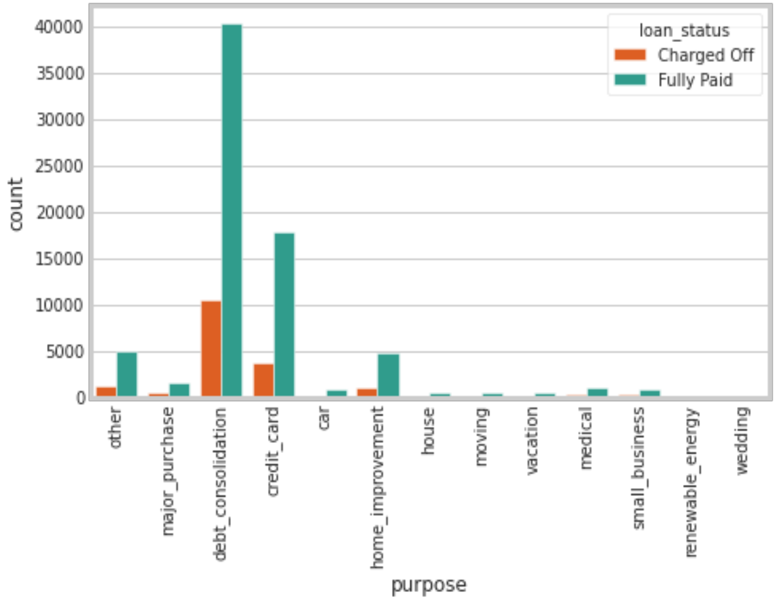
7. Occupation

count	912962
unique	218590
top	Teacher
freq	17725



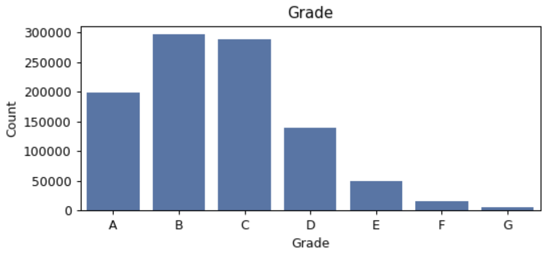
8. Purpose

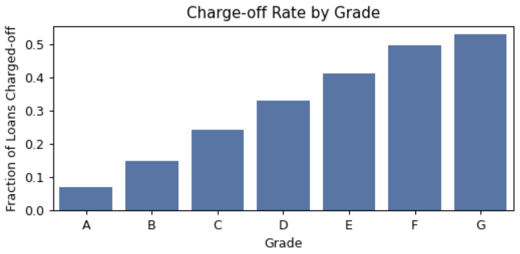
debt_consolidation	551274
credit_card	212002
home_improvement	73299
other	71621
major_purchase	24871
medical	14210
car	11529
small_business	10415
vacation	8645
moving	7879
house	7484
renewable_energy	692
wedding	7
educational	1

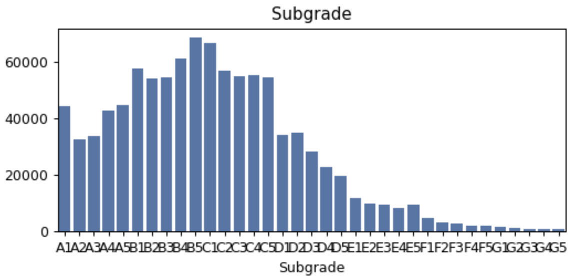


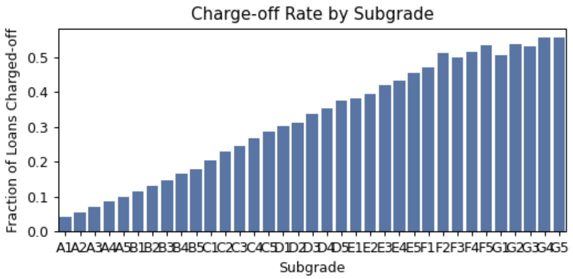
5. Methodology - Exploratory Data Analysis

9. Grade and Subgrade





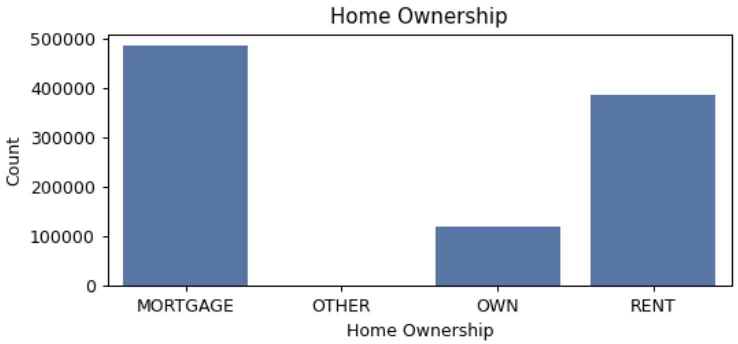


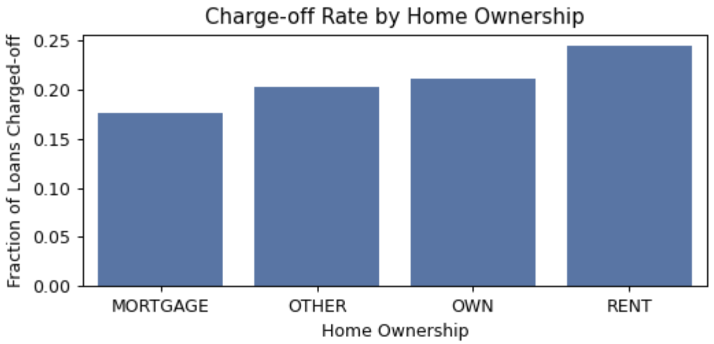


10. Home ownership

MORTGAGE	485792
RENT	386631
OWN	120331
ANY	1171
NONE	4

MORTGAGE	485792
RENT	386631
OWN	120331
OTHER	1175

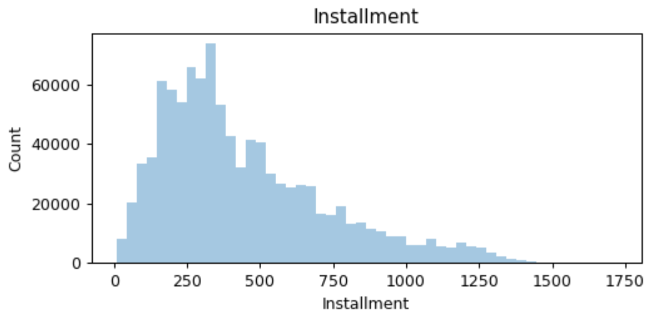


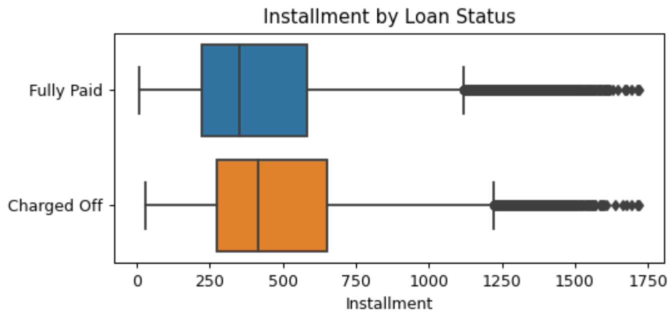


5. Methodology - Exploratory Data Analysis

11. Installment

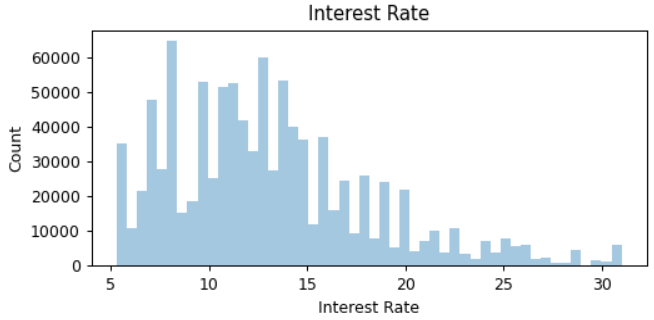
count	993929.000000
mean	443.638352
std	284.366689
min	7.610000
25%	234.230000
50%	366.040000
75%	597.780000
max	1719.830000

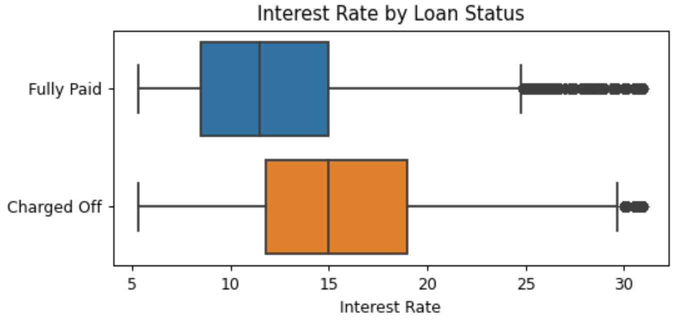




12. Interest rate

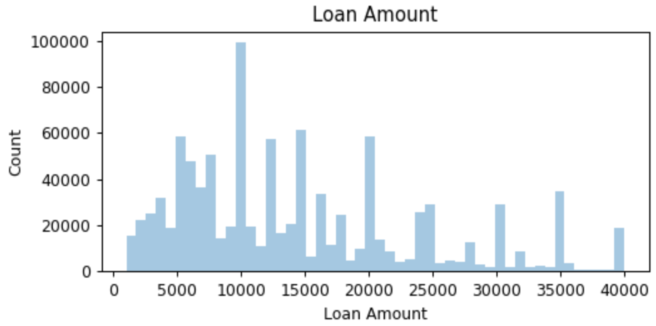
count	993929.000000
mean	13.125613
std	5.176469
min	5.310000
25%	9.440000
50%	12.400000
75%	15.990000
max	30.990000

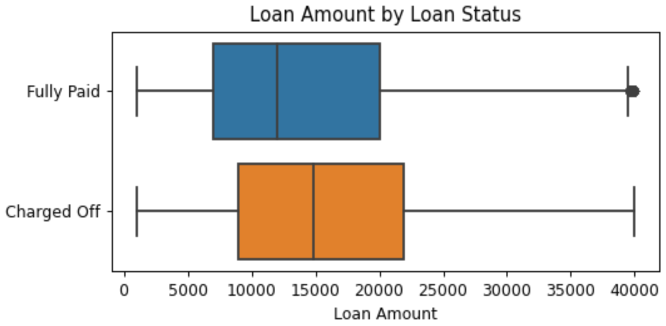




13. Loan Amount

count	993929.000000
mean	14526.980675
std	9423.153348
min	1000.000000
25%	7200.000000
50%	12000.000000
75%	20000.000000
max	40000.000000



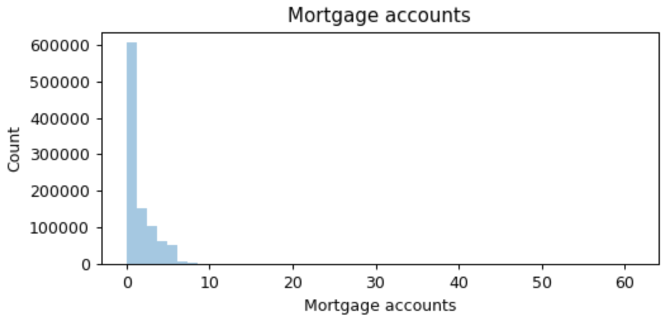


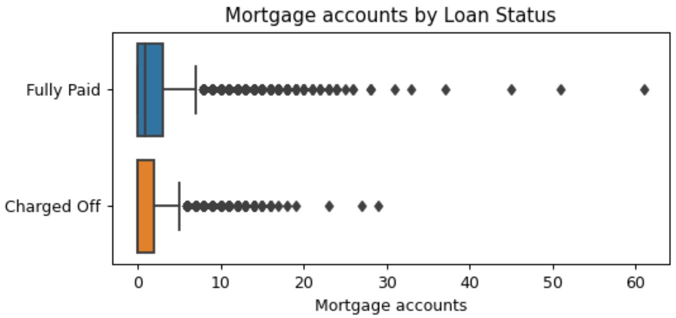
5. Methodology - Exploratory Data Analysis

14. Mort account

count	993929.000000
mean	1.481162
std	1.808551
min	0.000000
25%	0.000000
50%	1.000000
75%	2.000000
max	61.000000

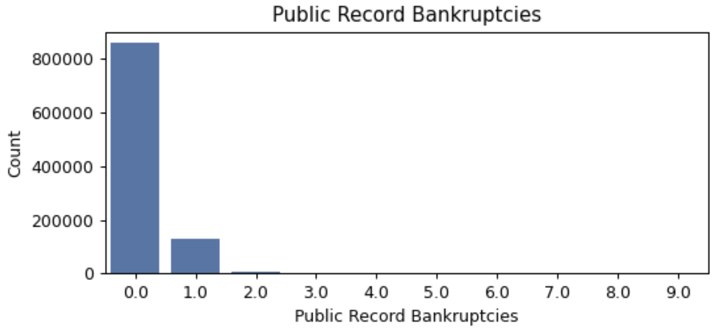
0.0	422225
1.0	182579
2.0	151458
3.0	103903
4.0	64105
5.0	34856
6.0	17831
7.0	8608
8.0	4092
9.0	2008

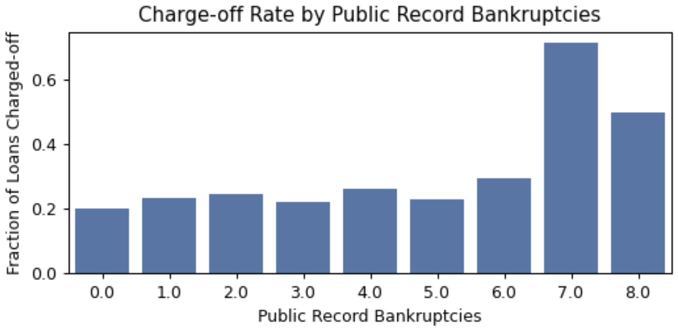




15. Public record bankruptcies

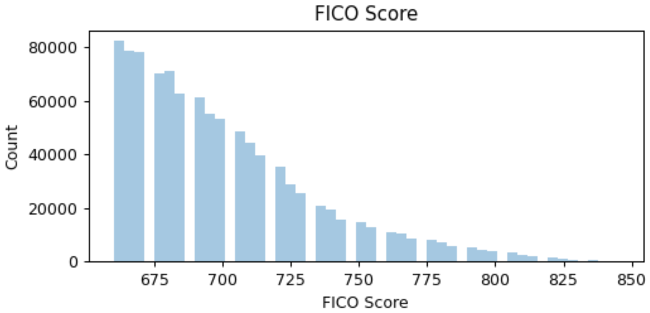
0.0	859572
1.0	127695
2.0	5211
3.0	1025
4.0	284
5.0	96
6.0	34
7.0	7
8.0	4
9.0	1

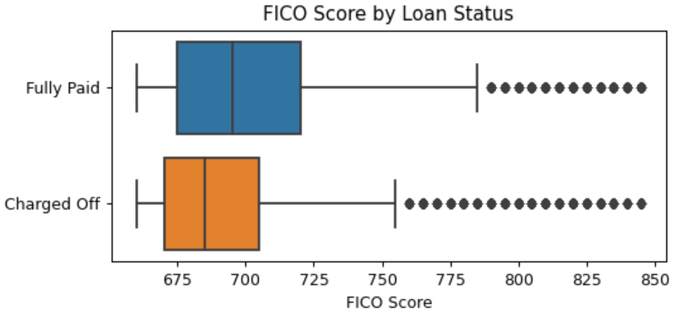




16. FICO Score

count	993929.000000
mean	700.036471
std	34.316292
min	660.000000
25%	675.000000
50%	690.000000
75%	720.000000
max	845.000000

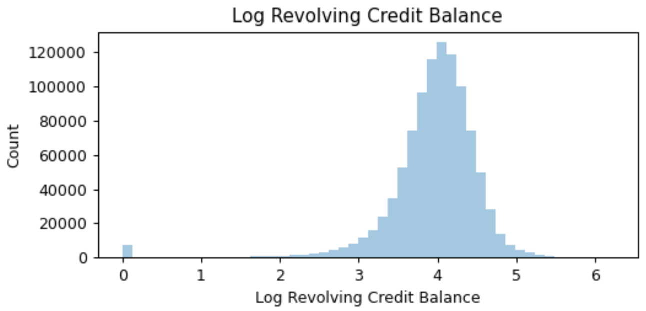


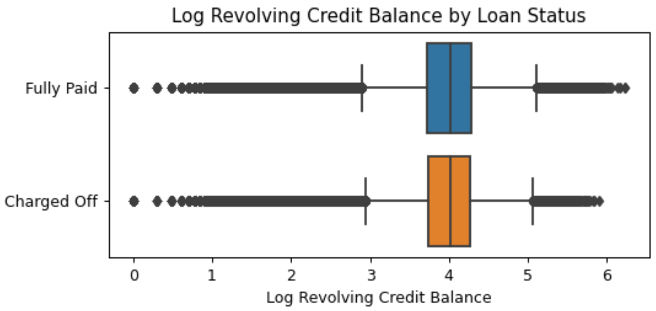


5. Methodology - Exploratory Data Analysis

17. Total credit revolving balance

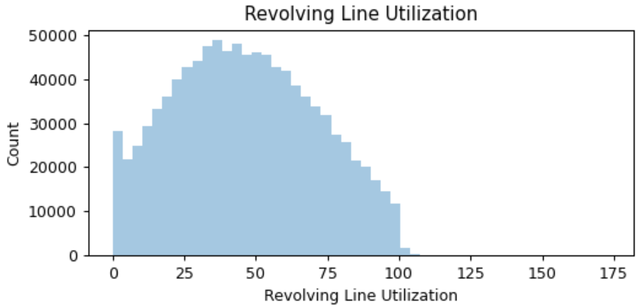
```
count    9.939290e+05
mean     1.587528e+04
std      2.285231e+04
min      0.000000e+00
25%      5.374000e+03
50%      1.043900e+04
75%      1.894100e+04
max      1.698749e+06
```

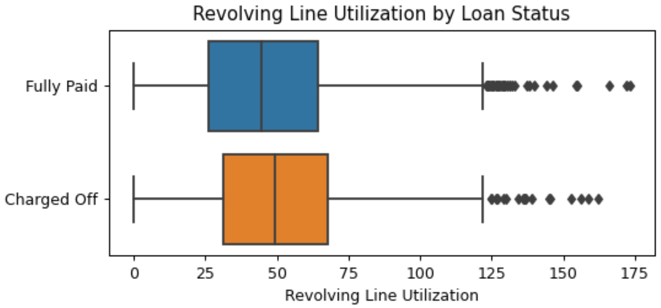




18. Revolving line utilization rate

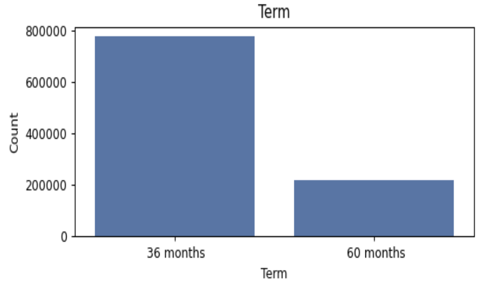
```
count    992986.000000
mean         46.540928
std          24.945235
min           0.000000
25%          27.100000
50%          45.500000
75%          65.200000
max         173.200000
```

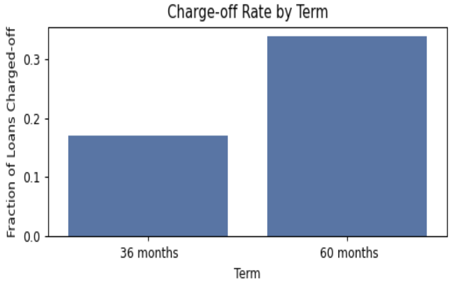




19. Term

```
term        loan_status
 36 months  Fully Paid     642877
            Charged Off    131691
 60 months  Fully Paid     145284
            Charged Off     74077
```

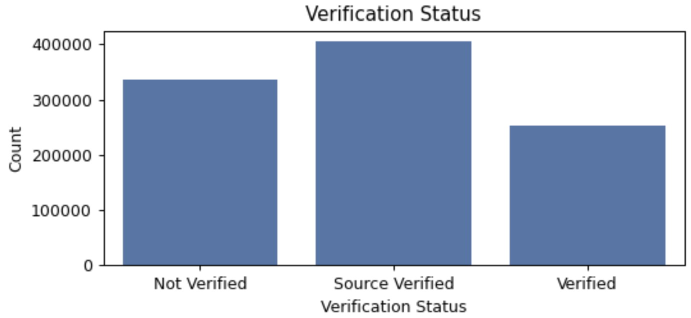


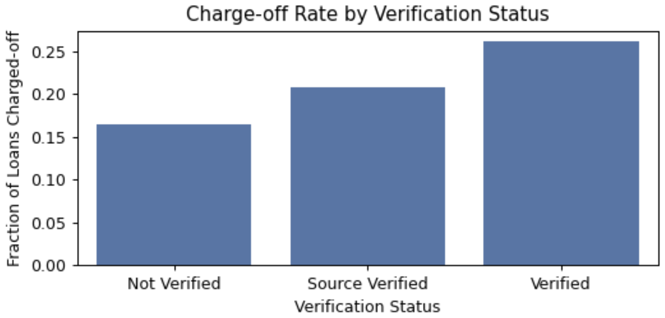


5. Methodology - Exploratory Data Analysis

20. Verification status

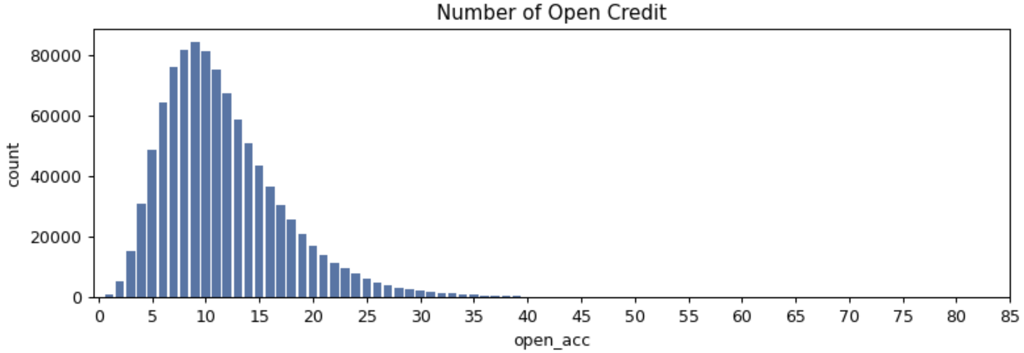
```
Source Verified     405190
Not Verified        336515
Verified            252224
```





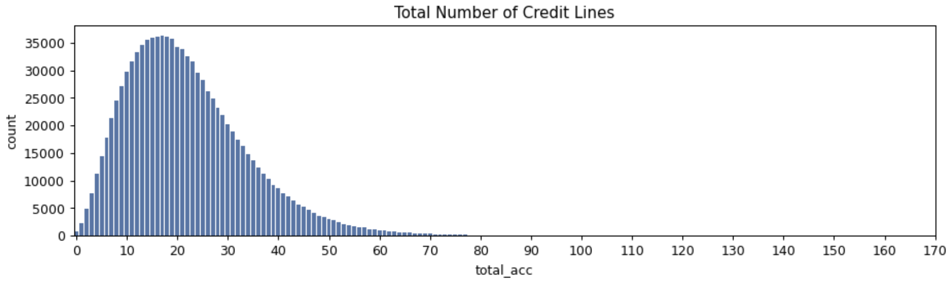
21. Number of Open Credit

	count	mean	std	min	25%	50%	75%	max
loan_status								
Charged Off	205768.0	11.801981	5.894573	0.0	8.0	11.0	15.0	78.0
Fully Paid	788161.0	11.634768	5.778010	0.0	8.0	11.0	15.0	88.0

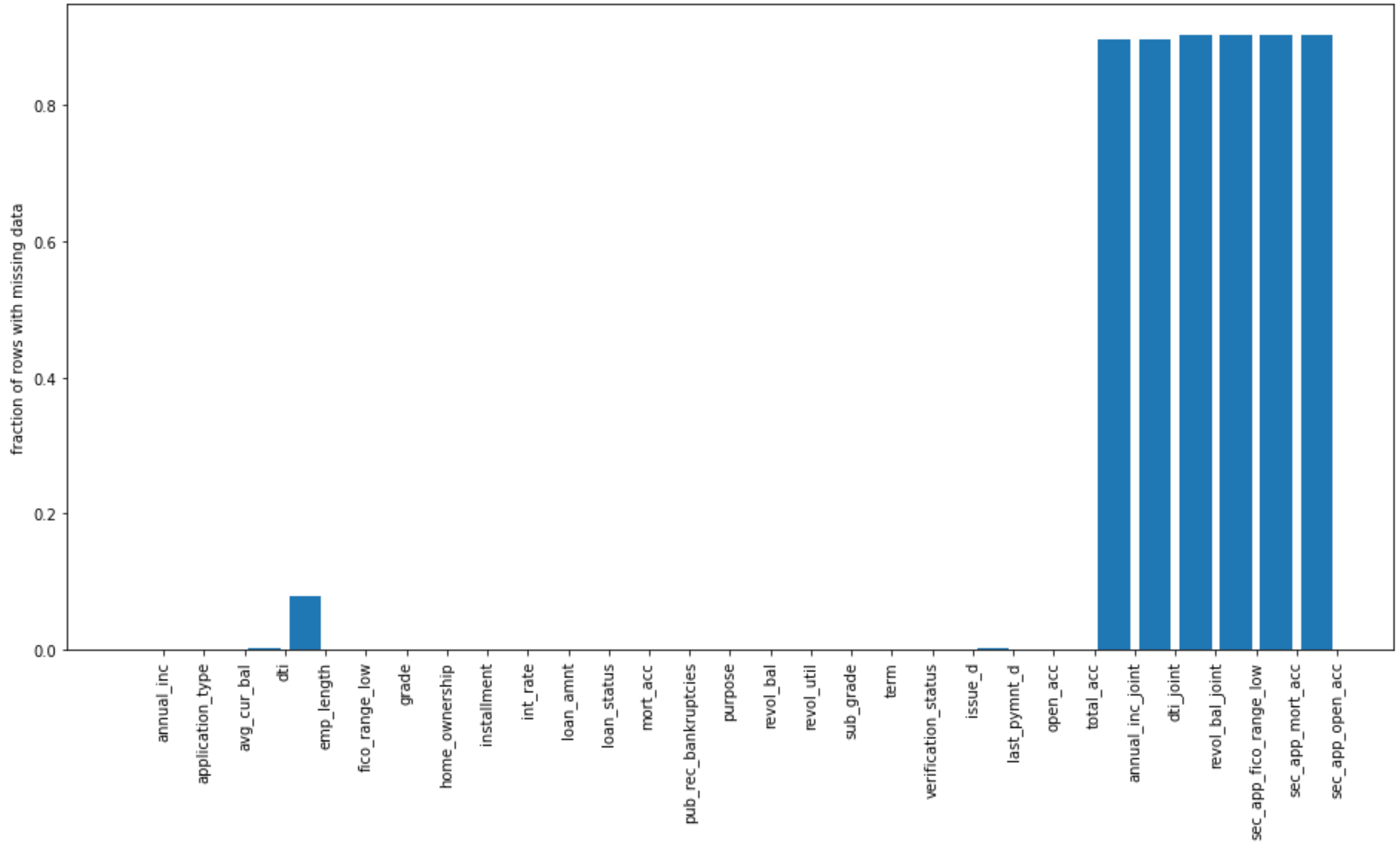


22. Total number of Credit Lines

	count	mean	std	min	25%	50%	75%	max
loan_status								
Charged Off	205768.0	23.366631	12.189346	2.0	14.0	21.0	30.0	176.0
Fully Paid	788161.0	24.174228	12.190440	2.0	15.0	22.0	31.0	165.0



5. Methodology – Cleansing Data



- Remove columns missing values > 50%
- Remove rows with null values
- Data remaining 914173, 26

5. Methodology - Feature Transformation

- **Convert Categorical data using Label Encoder**

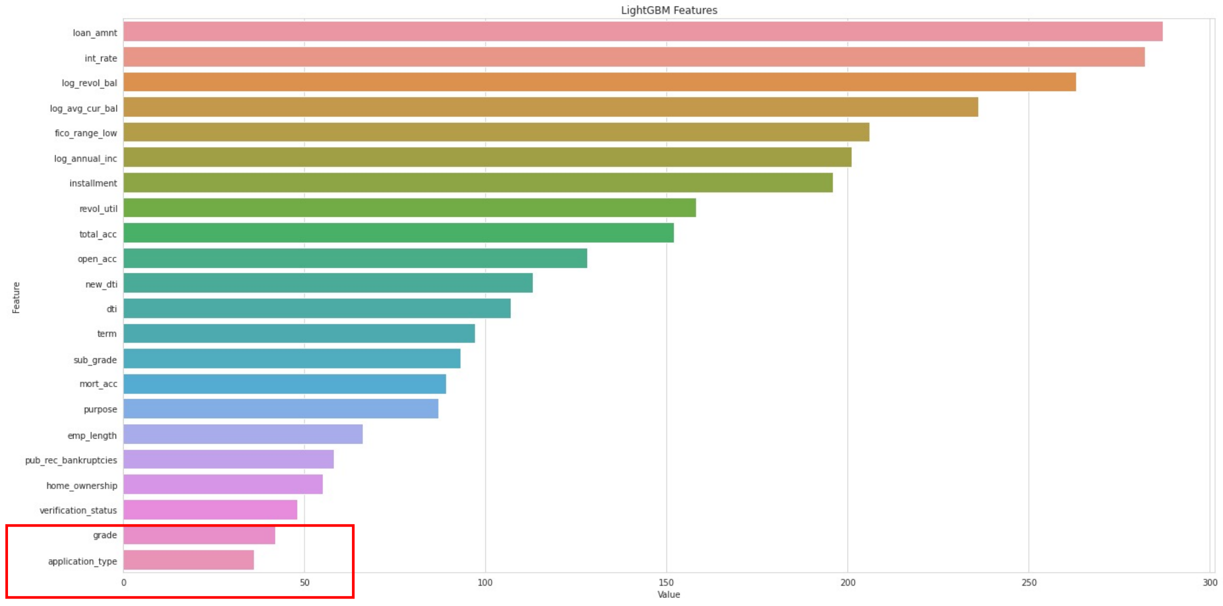
  application_type, grade, home_ownership, purpose, sub_grade, term, verification_status, issue_d, last_pymnt_d
- **Rescale numerical data using standardization**

  avg_cur_bal, dti, fico_range_low, installment, int_rate, loan_amnt, revol_util, new_dti, log_annual_inc, log_avg_cur_bal, log_revol_bal

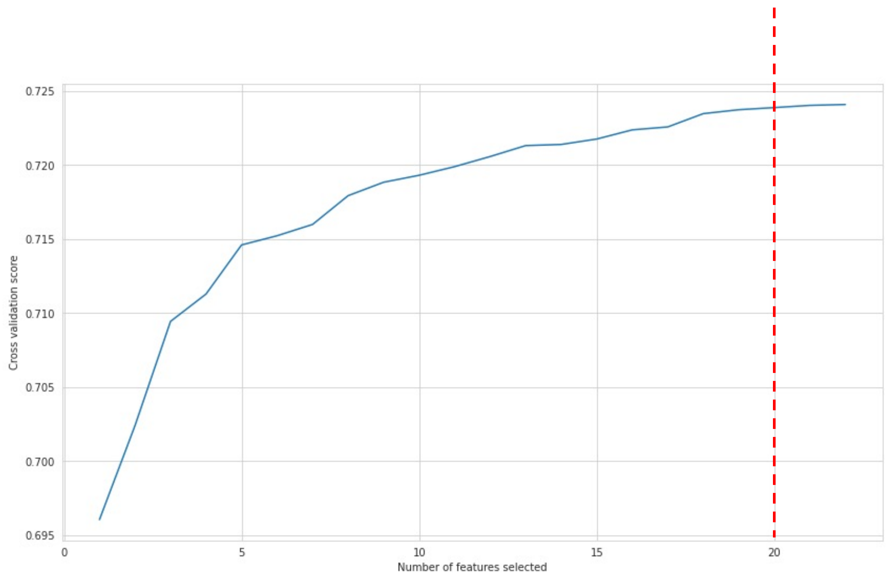
dti	emp_length	fico_range_low	grade	home_ownership	installment	int_rate	loan_amnt	loan_status	mort_acc
1.046168	10.0	0.448951	0.0	2.0	-0.256863	-0.991638	-0.288343	1.0	0.0
0.048078	3.0	-0.139589	1.0	0.0	-0.452421	-0.707229	-0.499786	1.0	2.0
0.223162	0.0	0.007546	2.0	0.0	-0.588013	0.565841	-0.711230	1.0	6.0
0.476309	5.0	-1.169533	2.0	3.0	-0.050025	0.095695	-0.203766	1.0	0.0
2.198153	4.0	-0.580994	2.0	0.0	0.211840	0.095695	0.028822	1.0	4.0

5. Methodology - Feature Selection

- GridSearchCV tuning parameters of LightGBM model to find feature importance
- Using RFECV (Recursive feature elimination) to select feature



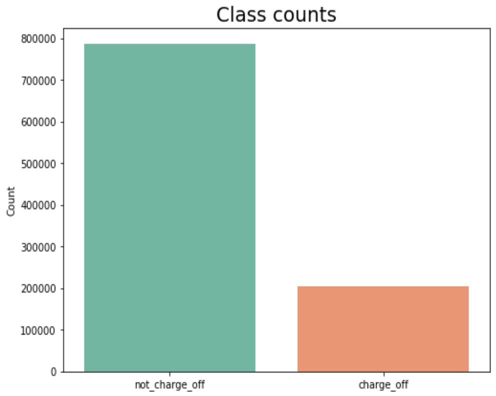
Record: 914,173 Feature: 22

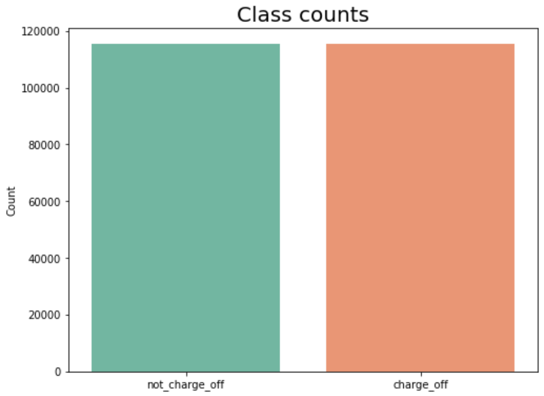


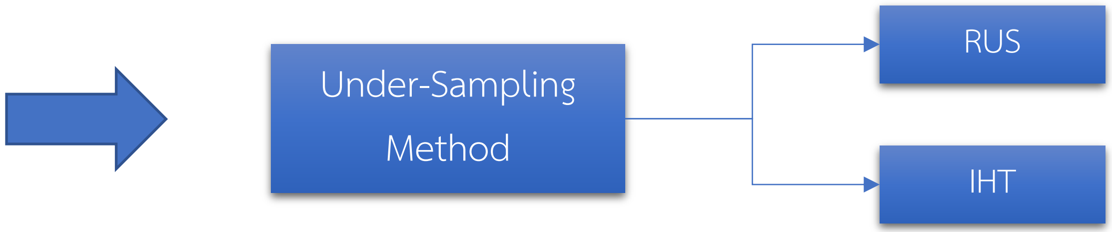
Optimal number of features: 20

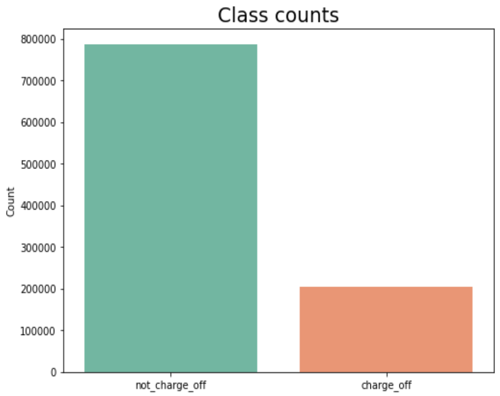
5. Methodology - Imbalance Problem

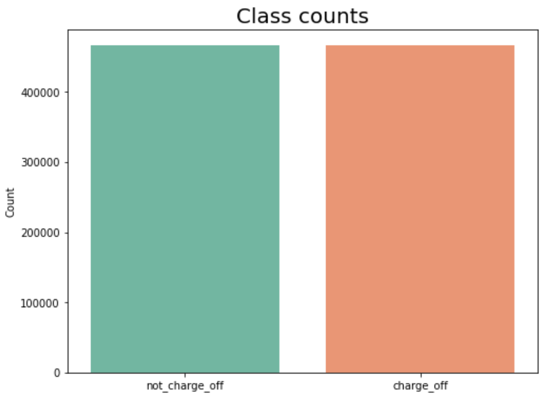
- Resampling Techical

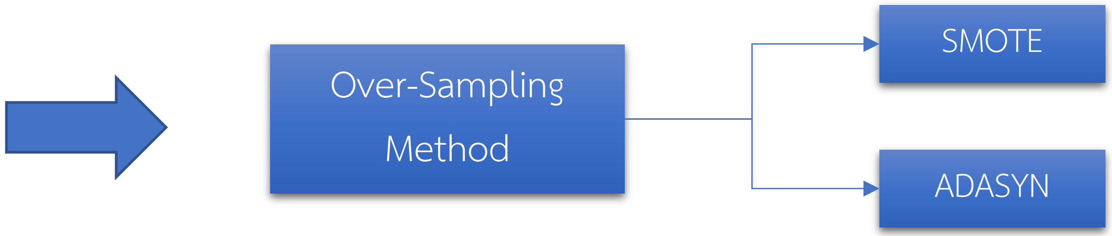












6. Experiment Result

	Classifier	Accuracy	Precision	Recall	F1-score	AUC
Non sampling	Logistic Regression	80.39	52.03	7.69	13.39	52.97
	Decision Tree	80.38	52.40	5.83	10.49	52.26
	Random Forest	80.44	56.60	3.47	6.54	51.41
	Xgboots	80.57	56.06	6.93	12.34	52.80
	Catboost	80.58	55.77	7.37	13.01	52.97
	LightGBM	80.60	55.50	8.31	14.45	53.33
	ANN	80.47	52.43	10.61	17.65	54.12
RUS	Logistic Regression	65.83	32.1	64.27	42.81	65.13
	Decision Tree	61.81	30.16	69.86	42.13	64.83
	Random Forest	63.81	31.3	68.49	42.97	65.57
	Xgboots	64.85	31.97	67.98	43.49	66.02
	Catboost	64.98	32.11	68.19	43.66	66.18
	LightGBM	65.06	32.11	67.84	43.59	66.1
	ANN	67.83	33.39	61.96	43.39	65.63
IHT	Logistic Regression	40.48	23.97	92.1	38.04	59.9
	Decision Tree	43.78	24.64	89.08	38.6	60.82
	Random Forest	44.48	24.86	88.95	38.86	61.21
	Xgboots	42.18	24.36	90.91	38.42	60.52
	Catboost	41.63	24.24	91.4	38.32	60.35
	LightGBM	40.95	24.09	91.85	38.16	60.1
	ANN	38.73	23.57	93.13	37.62	59.2
SMOTE	Logistic Regression	65.17	31.6	64.92	42.51	65.07
	Decision Tree	72.99	34.72	41.02	37.61	60.97
	Random Forest	69.47	33.56	55	41.69	64.03
	Xgboots	79.65	46.54	17.24	25.16	56.17
	Catboost	80.33	52.11	10.45	17.41	54.04
	LightGBM	80.35	52.35	10.82	17.93	54.19
	ANN	74.74	35.79	33.92	34.83	59.4
ADASYN	Logistic Regression	63.89	30.98	66.78	42.32	64.98
	Decision Tree	73.6	34.16	35.61	34.87	59.31
	Random Forest	71.12	34.25	49.54	40.5	63
	Xgboots	79.53	45.71	16.91	24.68	55.97
	Catboost	80.37	52.81	9.78	16.51	53.81
	LightGBM	80.4	53.19	10.08	16.95	53.94
	ANN	74.01	35.58	38.22	36.85	60.55

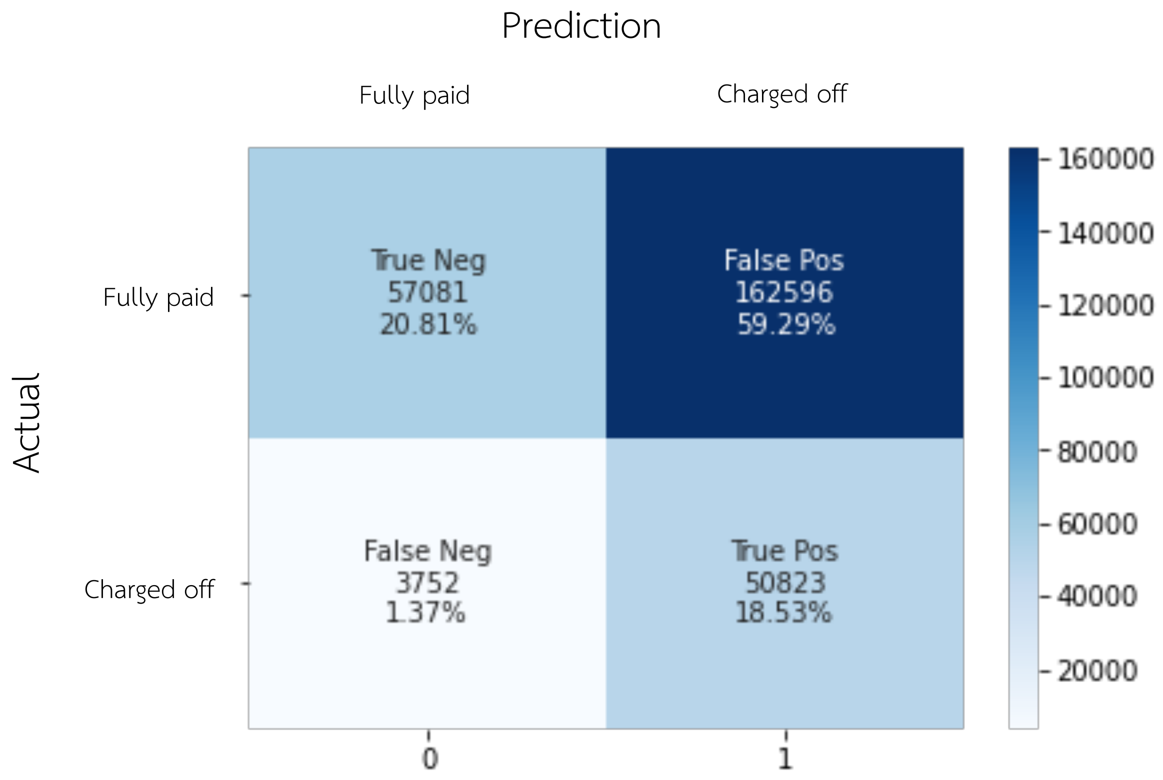
Evaluation Approach

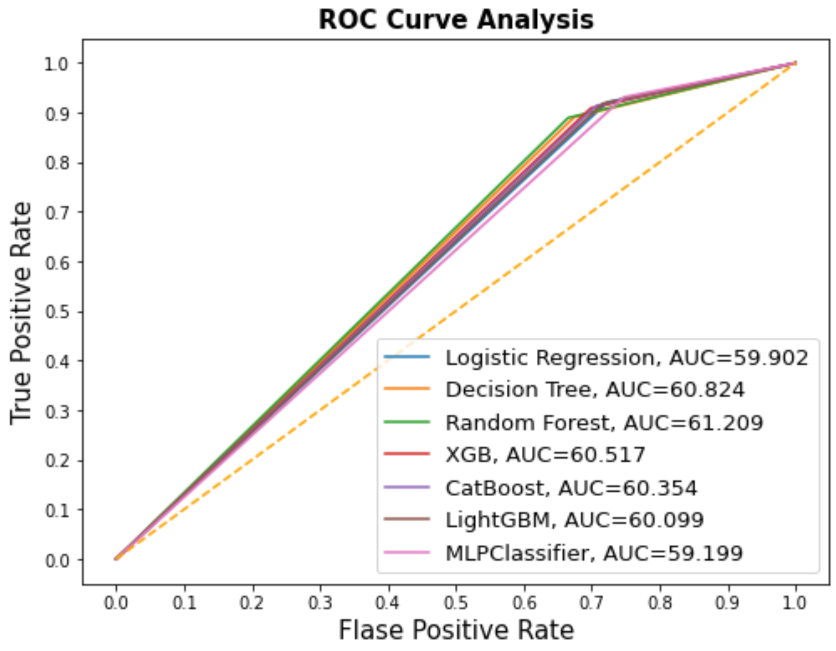
- For loan default prediction, the False Negatives Rate is metric to evaluate the model. Lower the number of false negatives, better the model is

- A false negative is when the model predicting "a borrower will not default a loan even though he will" Our model cannot afford having higher False Negative as it leads to negative impact on the investors and the credibility of the credibility of the company. So, we evaluated our models using the number of False negative and roc curve

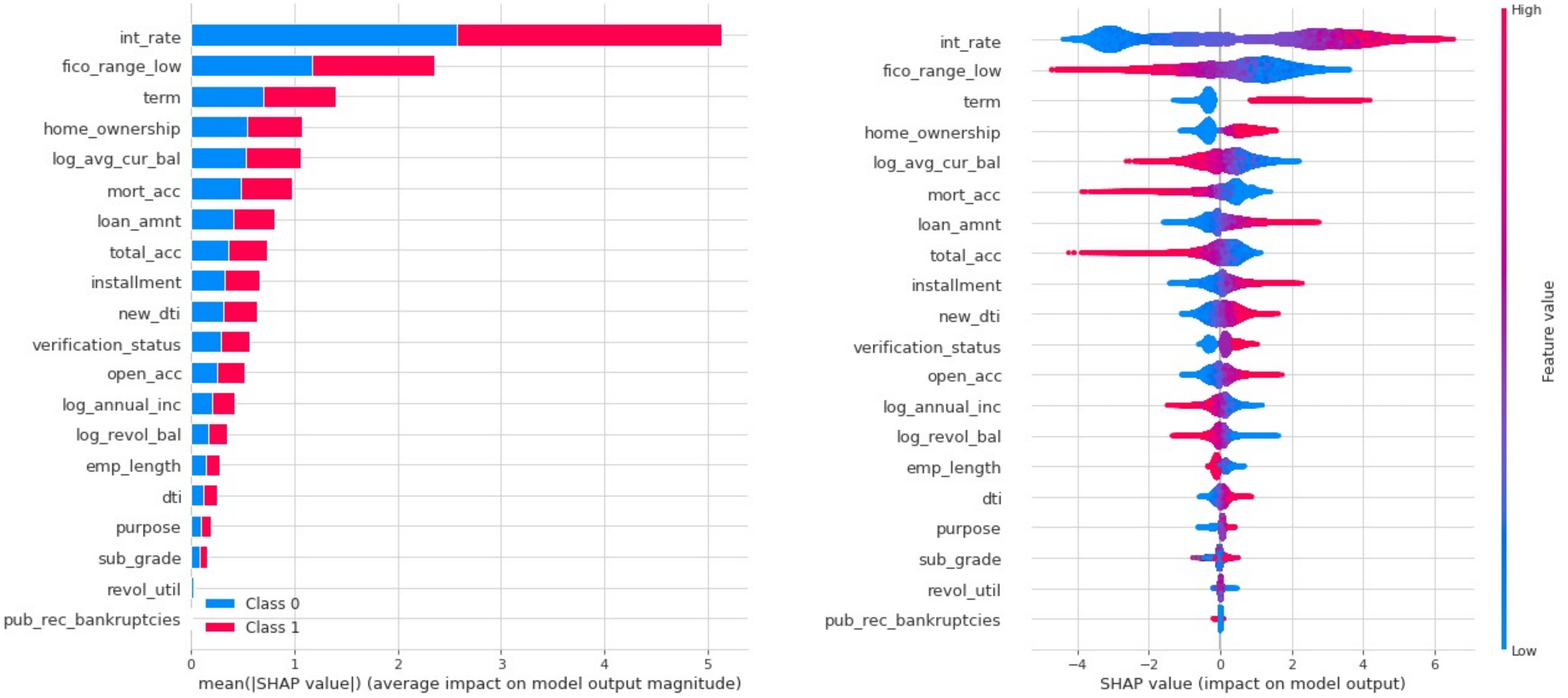
$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

6. Experiment Result

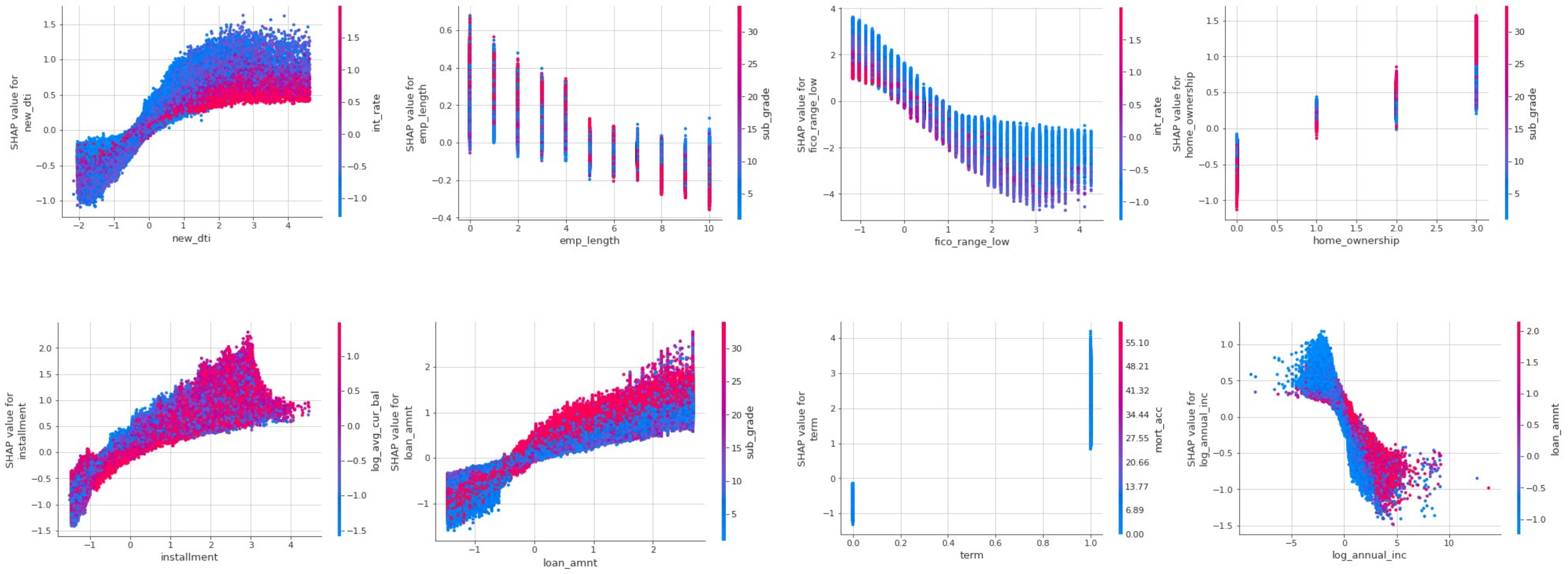




6. Experiment Result

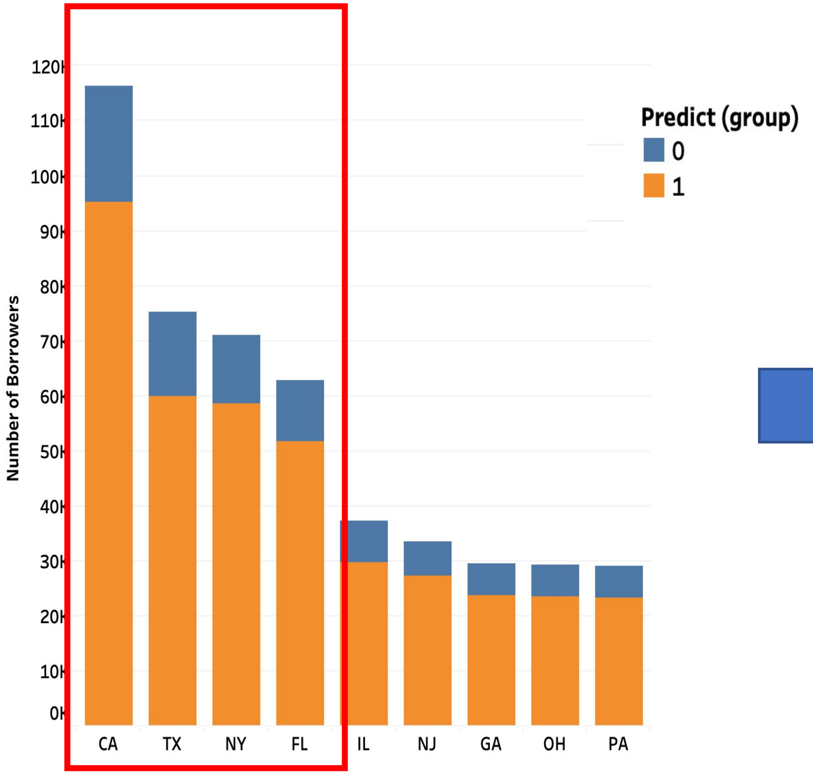


6. Experiment Result - Partial dependence



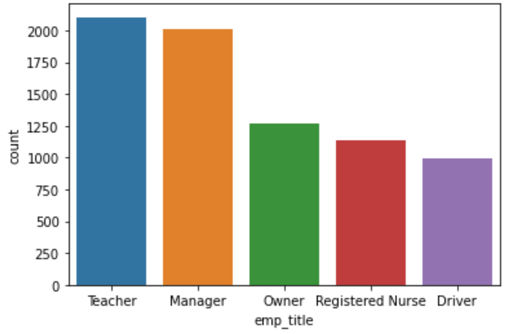
7. Discussion

Status by location

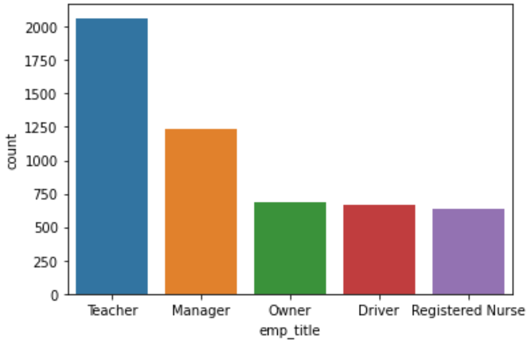


Top 5 of career

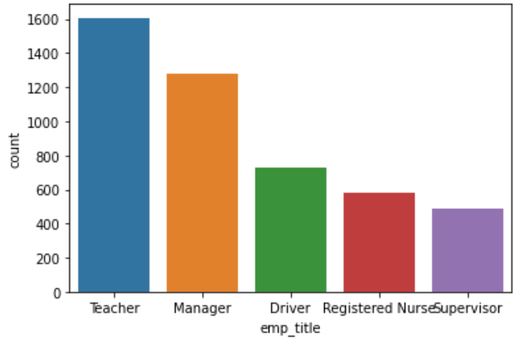
1. California



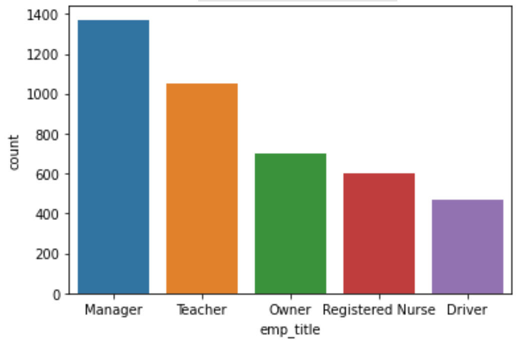
2. Texas



3. New York



4. Florida



8. Conclusion

- The information of the lending club with the most effective method for finding potential defaulters
- Resampling techniques can increase the model's efficiency well. The best performance of model is combination between the IHT Under-Sampling Technique and Neural Network Model, Recall efficiency is 93.13% and AUC is 59.2%
- The Lending club can use this method to predict the tendency of the borrower's default and used it as a strategy for planning bad debt management in the future
- The technique can be applied to other loans

9. Future Work & Suggestion

- This research is an analysis of personal loans only. This can be used to analyze similar loans such as car loans, home loans, agricultural loans, and others.
- Using Concept Diff techniques with imbalance data to find patterns of defaulters
- Using Clustering techniques can be used to group borrowers.
- Using Social Networks to find the relationship between investors and borrowers. to be used to find interesting patterns such as investor behavior, borrower behavior

10. Reference

- นพมาศ อัครจันทโชติ และ นพมาศ อัครจันทโชติ (2019). **การเปรียบเทียบวิธีการแก้ปัญหาข้อมูลไม่สมดุลสำหรับการจำแนกกลุ่มรายได้ของผู้ประกอบการร้านยาประเภท ข.ย.1.** การประชุมเสนอผลงานวิจัยระดับชาติ มหาวิทยาลัยสุโขทัยธรรมาธิราช ครั้งที่ 9
- พรชนก เทพขาม (2019). **เศรษฐกิจแบ่งปัน: บทเรียนการกู้ยืมผ่านช่องทางอิเล็กทรอนิกส์ (Peer-to-peer Lending)**. ธนาคารแห่งประเทศไทย
- อโนทัย พุทธารี จิตรภณ หรูเจริญพรพานิช เพ็ญสิริ บำรุงเชาว์เกษม และชินวัฒน์ เทพหัสดิน ณ อยุธยา. (2018). CREDIT **SCORING MODEL: เครื่องมือในการประเมินคุณภาพสินเชื่อ**. ธนาคารแห่งประเทศไทย
- เปมิกา ฐิติเพิ่มพงศ์ (2020, 17 กุมภาพันธ์). **มาทำความรู้จักกับ Peer-to-Peer Lending กันเถอะ | ตอนที่ 3** สืบค้นจาก https://www.nestifly.com/post/what-is-p2p-ep3
- Aleum Kim and Sung-Bae Cho. (2017). **Dempster-Shafer Fusion of Semi-supervised Learning Methods for Predicting Defaults in Social Lending**. Seoul: Yonsei University.
- Anahita Namvar, Mohammad Siami, Fethi Rabhi and Mohsen Naderpour. (2018). **Credit risk prediction in an imbalanced social lending environment**. Sydney: University of New South Wales.
- Haibo He, Yang Bai, Edwardo A. Garcia, and Shutao Li. (2008) **ADASYN: Adaptive Synthetic Sampling Approach for Imbalanced Learning**. 2008 International Joint Conference on Neural Networks
- Naufal Azmi Verdikha, Teguh Bharata Adji and Adhistya Erna Permanasari. (2018). **Study of Undersampling Method: Instance Hardness Threshold with Various Estimators for Hate Speech Classification**. IJITEE, Vol. 2, No. 2, June 2018
- Yijie Fu (2017). **Combination of Random Forests and Neural Networks in Social Lending**. Shanghai: Shanghai Foreign Language School.
- Yuejin Zhanga, Haifeng Lia, Mo Hai a, Jiaxuan Lia and Aihua Lib (2017). **Determinants of loan funded successful in online P2P Lending.** Beijing: Central University of Finance and Economics
- Zhiqiang Li, Shouyan Li, Zhilong Li, Yixiang Hu and Hanlin Gao. (2021). **Application of XGBoost in P2P Default Prediction**. College of Business Jiangxi University of Science and Technology Nanchang