Lending Club Case Study

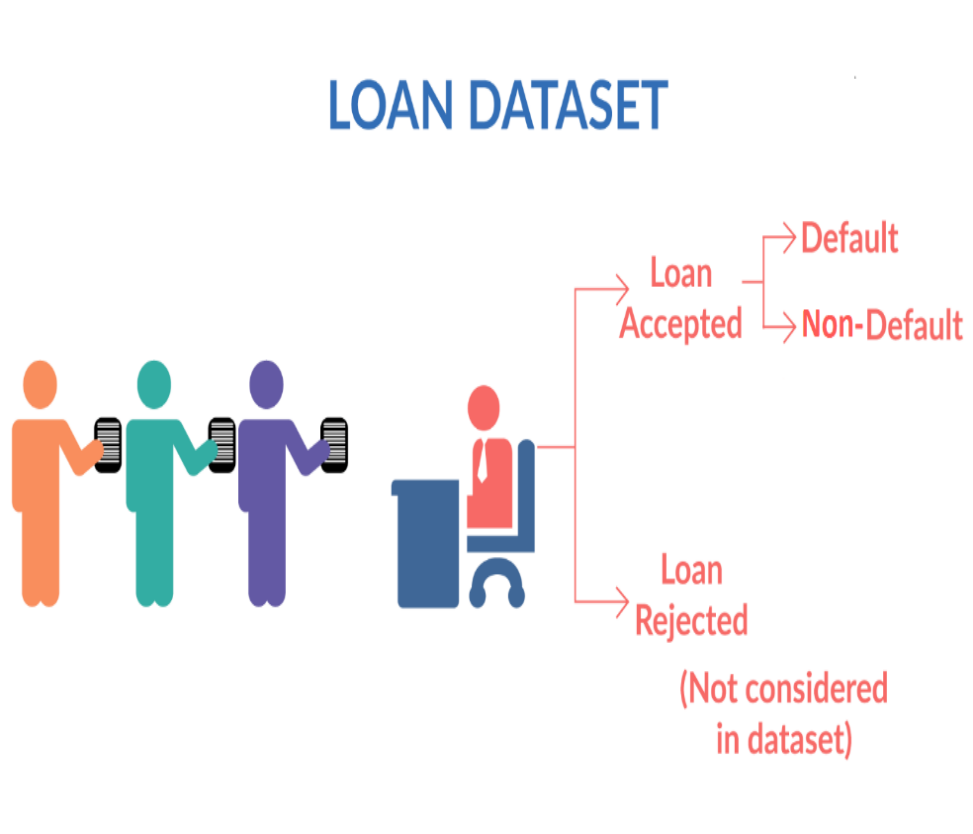
Recommending optimal Consumers for loan approval

Lending Club – what it is? How it works?

- Lending Club is a third-party system, which acts like a bridge between Consumer/ loan applicant and the Borrowers/ Lenders.
- When the company receives a loan application, the company has to take decision for loan approval based on the applicant's profile.
- Two types of risks are associated with the bank's decision:
  - If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company.
  - If the applicant is not likely to repay the loan, that is he/ she is likely to default, then approving the loan may lead to a financial loss for the company.

Lending Club – Problem Statement

- The data given for analysis contains information about past loan applicants and whether they 'defaulted' or not. The aim is to identify patterns which indicate if a person is likely to default, which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.
- If one is able to identify these risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss. Identification of such applicant's using EDA is the aim of this case study.
- In other words, the company wants to understand the driving factors(or driver variables) behind loan default, i.e., the variables which are strong indicators of default. The company can utilize this knowledge for its portfolio and risk assessment.



Lending Club – Data Understanding

- Loan dataset contains 111 columns and 39717 rows:

```
loan.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 39717 entries, 0 to 39716
Columns: 111 entries, id to total_il_high_credit_limit
dtypes: float64(74), int64(13), object(24)
memory usage: 33.6+ MB
```

```
loan = pd.read_csv('loan.csv')
loan.head()
```

	id	member_id	loan_amnt	funded_amnt	funded_amnt_inv	term	int_rate	installment	grade	sub_grade	...	num_tl_90g_dpd_24m	num_tl_op_past_
0	1077501	1296599	5000	5000	4975.0	36 months	10.65%	162.87	B	B2	...	NaN	
1	1077430	1314167	2500	2500	2500.0	60 months	15.27%	59.83	C	C4	...	NaN	
2	1077175	1313524	2400	2400	2400.0	36 months	15.96%	84.33	C	C5	...	NaN	
3	1076863	1277178	10000	10000	10000.0	36 months	13.49%	339.31	C	C1	...	NaN	
4	1075358	1311748	3000	3000	3000.0	60 months	12.69%	67.79	B	B5	...	NaN	

5 rows × 111 columns

Lending Club – Data Cleaning

- After Data cleaning process we got Loan dataset with 37 columns and 32,521 rows:

```
loans_df_final.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 38521 entries, 0 to 38520
```

```
loans_df_final
```

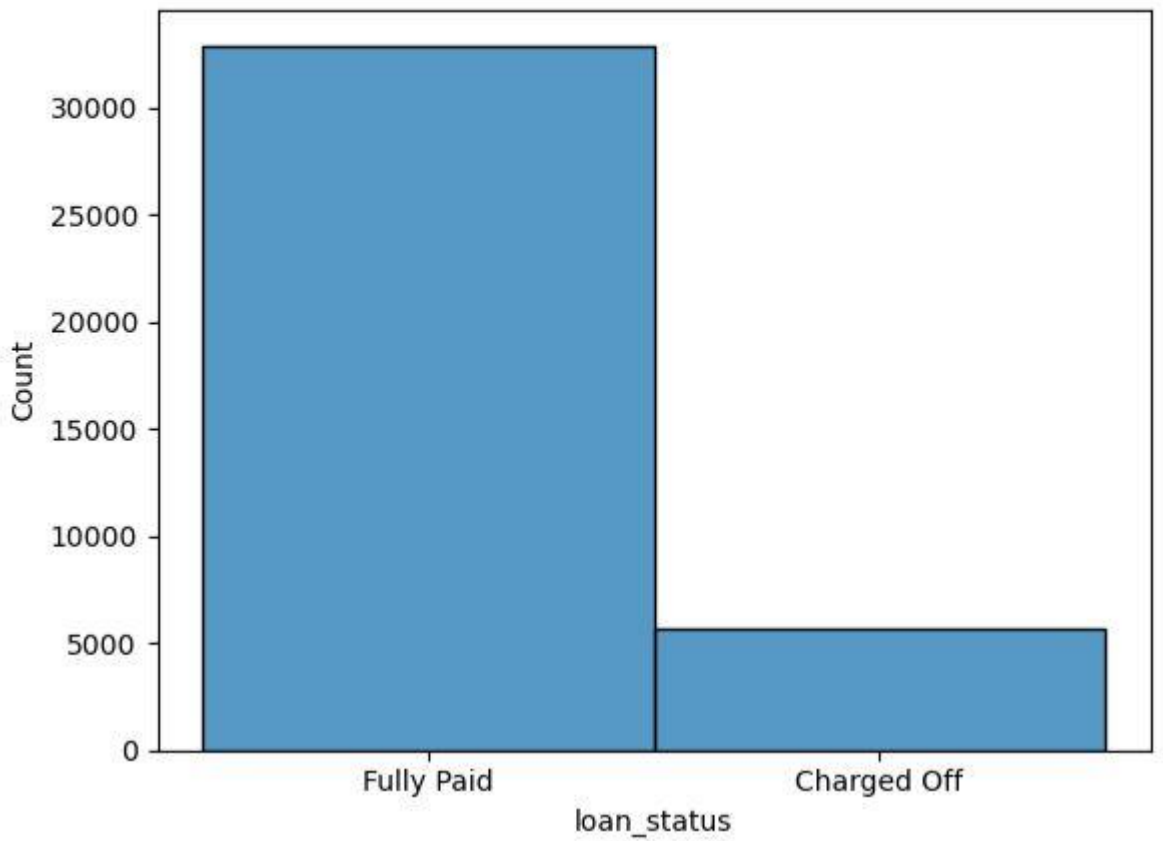
	index	member_id	loan_amnt	funded_amnt	funded_amnt_inv	term	int_rate	installment	grade	verification_status	...	last_pymnt_d	last_pymnt_amnt
0	0	1296599	5000	5000	4975.0	36 months	10.65%	162.87	B	Verified	...	Jan-15	171.62
1	1	1314167	2500	2500	2500.0	60 months	15.27%	59.83	C	Source Verified	...	Apr-13	119.66
2	2	1313524	2400	2400	2400.0	36 months	15.96%	84.33	C	Not Verified	...	Jun-14	649.91
3	3	1277178	10000	10000	10000.0	36 months	13.49%	339.31	C	Source Verified	...	Jan-15	357.48
4	5	1311441	5000	5000	5000.0	36 months	7.90%	156.46	A	Source Verified	...	Jan-15	161.03
...	...	...	...	...	...	...	...	...	...	...	...	...	...
38516	39661	121527	3000	3000	2225.0	36 months	7.75%	93.67	A	Not Verified	...	Apr-08	2563.55
38517	39662	121423	3000	3000	2975.0	36 months	7.75%	93.67	A	Not Verified	...	Apr-08	2563.59
38518	39663	120227	4000	4000	1475.0	36 months	10.91%	130.79	C	Not Verified	...	Aug-10	134.78
38519	39664	119635	2000	2000	475.0	36 months	8.70%	63.32	B	Not Verified	...	Aug-10	64.27
38520	39665	118760	4000	4000	3575.0	36 months	7.43%	124.31	A	Not Verified	...	Aug-10	119.24

38521 rows × 38 columns

Lending Club – Data Analysis - Univariate Analysis(Histograms)

```
sns.histplot(loans_df_final["loan_status"])
```

```
<Axes: xlabel='loan_status', ylabel='Count'>
```

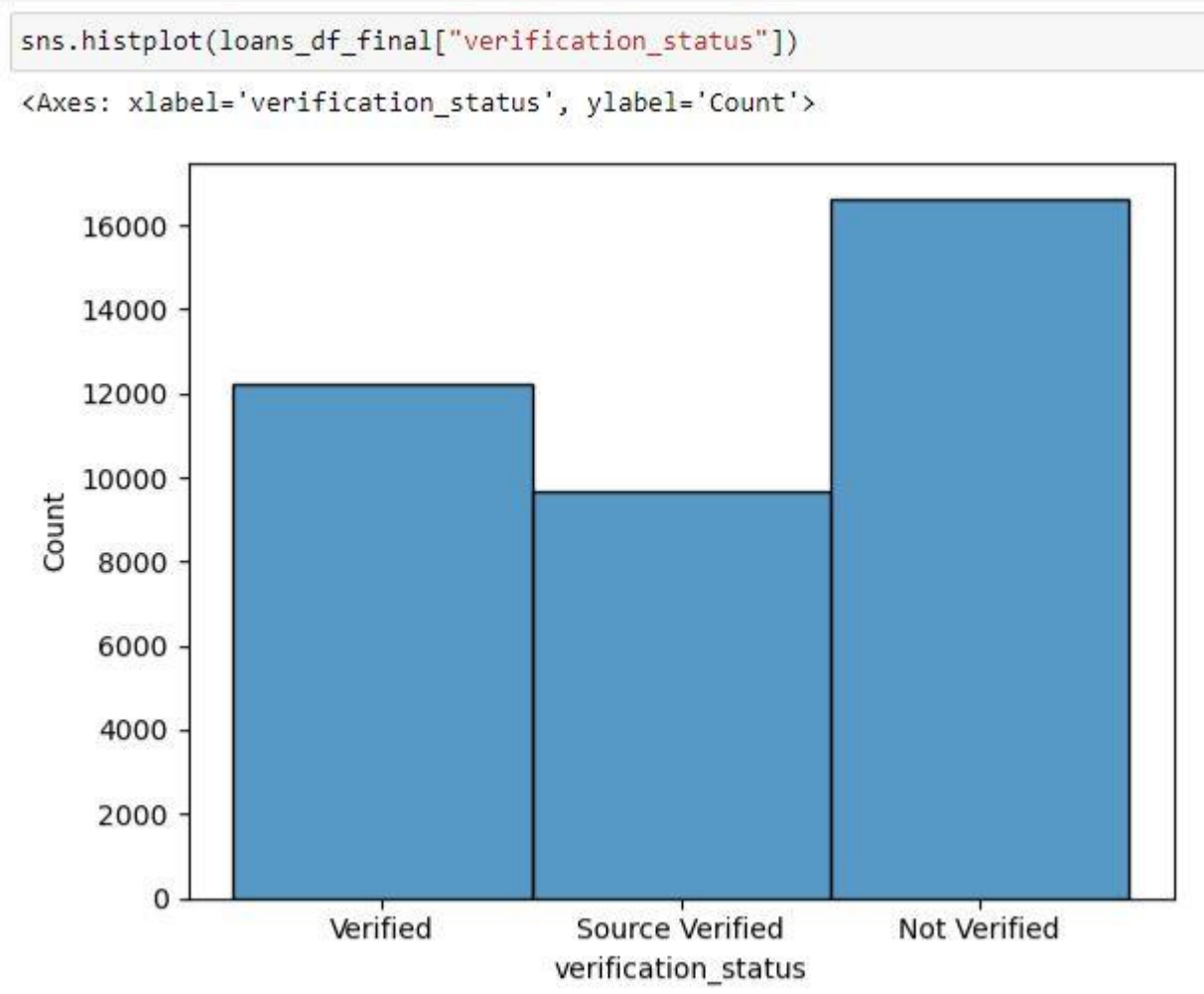


- There are 32,950 Fully paid and 5627 Charged off consumers, we do not consider Consumers of Loan Status as 'Current'.

Lending Club – Data Analysis – Univariate Analysis(Histograms)

```
sns.histplot(loans_df_final["verification_status"])
```

```
<Axes: xlabel='verification_status', ylabel='Count'>
```

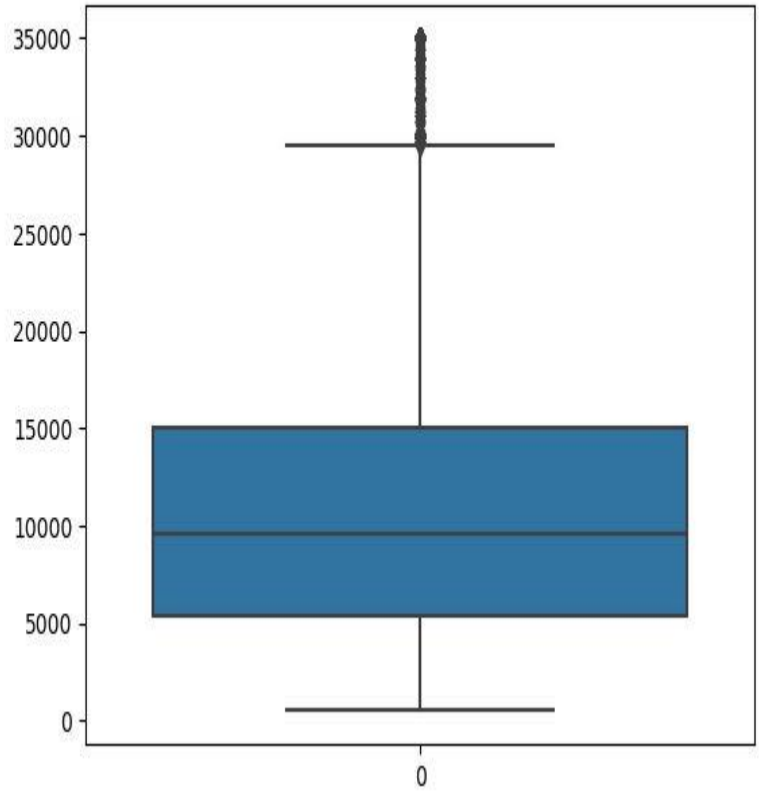


- Verification Status has to be in 'Verified' or in 'Source Verified'. Here 12,809 Consumers are Verified, and 9,987 Consumers are Source Verified.

Lending Club – Data Analysis – Univariate Analysis(Boxplots)

```
sns.boxplot(loans_df_final["loan_amnt"])
```

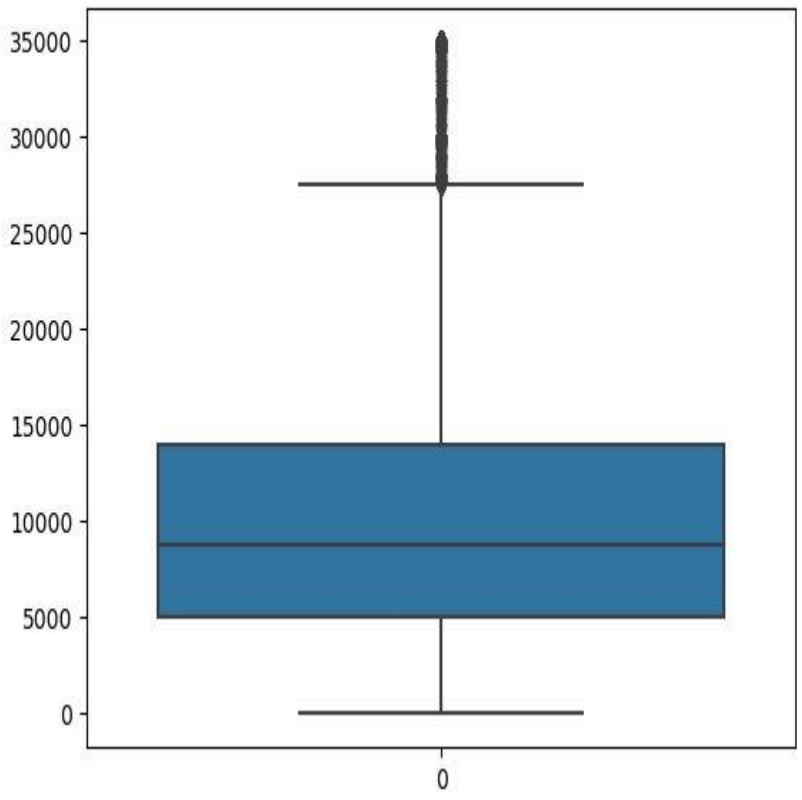
```
<Axes: >
```



Consumer requested Loan Amount

```
sns.boxplot(loans_df_final["funded_amnt_inv"])
```

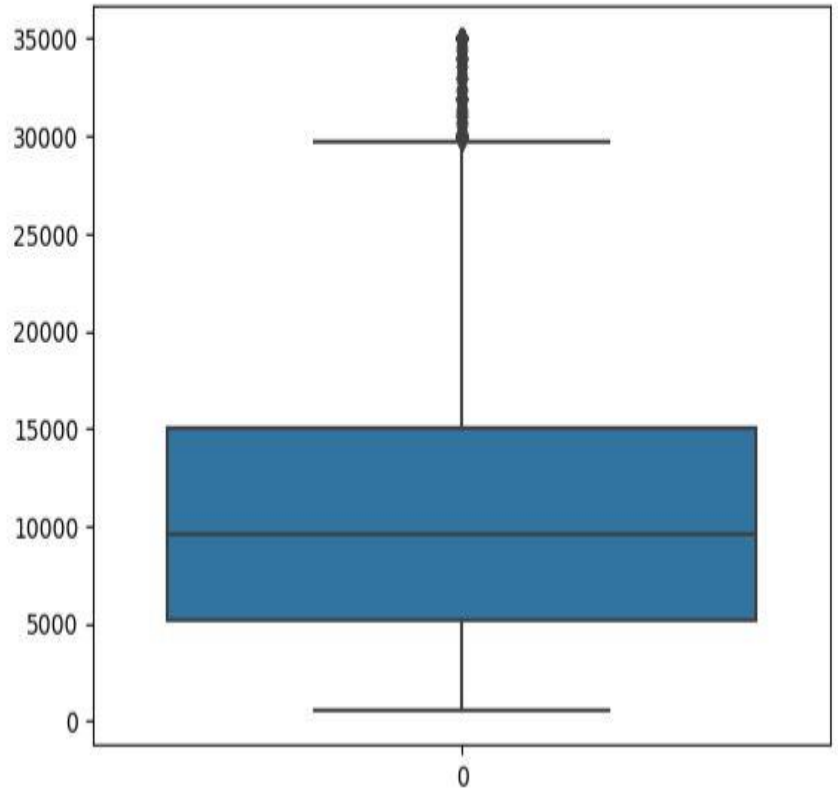
```
<Axes: >
```



Investors funded Amount

```
sns.boxplot(loans_df_final["funded_amnt"])
```

```
<Axes: >
```



Lending Club funded Amount

Lending Club – Data Analysis – Bivariate Analysis

Row Labels	Count of member_id
Charged Off	5627
Current	1140
Fully Paid	32950
Grand Total	**39717**

- There are 32,950 Fully paid and 5627 Charged off consumers, we do not consider Consumers of Loan Status as 'Current'.

Row Labels	Count of member_id
Not Verified	16921
Source Verified	9987
Verified	12809
Grand Total	**39717**

- Verification Status has to be in 'Verified' or in 'Source Verified'. Here 12,809 Consumers are Verified, and 9,987 Consumers are Source Verified.

Row Labels	Count of member_id
MORTGAGE	17659
NONE	3
OTHER	98
OWN	3058
RENT	18899
Grand Total	**39717**

- Home Ownership should ideally be 'OWN' as this factor improves the Consumers credibility for loan approval. Here most of the Consumers are in 'RENTED' or in 'MORTGAGE' status. Those 3,058 Consumers who own the house can be given highest priority for the Loan approval.

Lending Club – Data Analysis – Bivariate Analysis

Row Labels	Count of member_id
< 1 year	4583
1 year	3240
10+ years	8879
2 years	4388
3 years	4095
4 years	3436
5 years	3282
6 years	2229
7 years	1773
8 years	1479
9 years	1258
n/a	1075
Grand Total	39717

- Employment Tenure or Employment length will play impactful role in Loan repayment process, it is suggested to choose Consumers whose Employment Tenure is >3 years.

Row Labels	Count of member_id
2007	251
Qtr2	1
Qtr3	81
Qtr4	169
2008	1562
Qtr1	581
Qtr2	292
Qtr3	186
Qtr4	503
2009	4716
Qtr1	775
Qtr2	965
Qtr3	1231
Qtr4	1745
2010	11532
Qtr1	1953
Qtr2	2776
Qtr3	3283
Qtr4	3520
2011	21656
Qtr1	4120
Qtr2	5078
Qtr3	5861
Qtr4	6597
Grand Total	39717

- Loan Funded Month and Year will explain about how the business has grown or fallen over the years in quarterly wise.
- As we have available data of Loan funding status from year 2007 till 2011, we can see how the loan processing/ funding targets were met based on each quarter for all these years.
- We can see a quite significant growth in business year over year here: in 2007 total loans funded are 251, in 2008 it is 7X times more when compared.
- 2011 > 2010 > 2009 > 2008 > 2007.

Lending Club – Detecting Aberrant & Missing values

- After Data wrangling and Data cleaning process, final Loan Dataset with required Attributes is as below:

```
loans_df_final.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 38521 entries, 0 to 38520
Data columns (total 38 columns):
 #   Column                      Non-Null Count  Dtype
---  ------                      --------------  -----
 0   index                       38521 non-null  int64
 1   member_id                   38521 non-null  int64
 2   loan_amnt                   38521 non-null  int64
 3   funded_amnt                 38521 non-null  int64
 4   funded_amnt_inv             38521 non-null  float64
 5   term                        38521 non-null  object
 6   int_rate                    38521 non-null  object
 7   installment                 38521 non-null  float64
 8   grade                       38521 non-null  object
 9   verification_status         38521 non-null  object
 10  issue_d                     38521 non-null  object
 11  loan_status                 38521 non-null  object
 12  pymnt_plan                  38521 non-null  object
 13  purpose                     38521 non-null  object
 14  title                       38510 non-null  object
 15  mths_since_last_delinq      13622 non-null  float64
 16  revol_bal                   38521 non-null  int64
 17  revol_util                  38471 non-null  object
 18  initial_list_status         38521 non-null  object
 19  out_prncp                   38521 non-null  float64
 20  out_prncp_inv               38521 non-null  float64
 21  total_pymnt                 38521 non-null  float64
 22  total_pymnt_inv             38521 non-null  float64
 23  total_rec_prncp             38521 non-null  float64
 24  total_rec_int               38521 non-null  float64
 25  total_rec_late_fee          38521 non-null  float64
 26  recoveries                  38521 non-null  float64
 27  collection_recovery_fee     38521 non-null  float64
 28  last_pymnt_d                38450 non-null  object
 29  last_pymnt_amnt             38521 non-null  float64
 30  last_credit_pull_d          38519 non-null  object
 31  collections_12_mths_ex_med  38521 non-null  float64
 32  policy_code                 38521 non-null  int64
 33  application_type            38521 non-null  object
 34  acc_now_delinq              38521 non-null  int64
 35  chargeoff_within_12_mths    38521 non-null  float64
 36  delinq_amnt                 38521 non-null  int64
 37  pub_rec_bankruptcies        37878 non-null  float64
dtypes: float64(16), int64(8), object(14)
memory usage: 11.2+ MB
```

Lending Club – Recommendations

- This is the final Dataset to consider with almost 32,521 consumers for loan process:

```
loans_df_final
```

	index	member_id	loan_amnt	funded_amnt	funded_amnt_inv	term	int_rate	installment	grade	verification_status	...	last_pymnt_d	last_pymnt_amnt
0	0	1296599	5000	5000	4975.0	36 months	10.65%	162.87	B	Verified	...	Jan-15	171.62
1	1	1314167	2500	2500	2500.0	60 months	15.27%	59.83	C	Source Verified	...	Apr-13	119.66
2	2	1313524	2400	2400	2400.0	36 months	15.96%	84.33	C	Not Verified	...	Jun-14	649.91
3	3	1277178	10000	10000	10000.0	36 months	13.49%	339.31	C	Source Verified	...	Jan-15	357.48
4	5	1311441	5000	5000	5000.0	36 months	7.90%	156.46	A	Source Verified	...	Jan-15	161.03
...	...	...	...	...	...	...	...	...	...	...	...	...	...
38516	39661	121527	3000	3000	2225.0	36 months	7.75%	93.67	A	Not Verified	...	Apr-08	2563.55
38517	39662	121423	3000	3000	2975.0	36 months	7.75%	93.67	A	Not Verified	...	Apr-08	2563.59
38518	39663	120227	4000	4000	1475.0	36 months	10.91%	130.79	C	Not Verified	...	Aug-10	134.78
38519	39664	119635	2000	2000	475.0	36 months	8.70%	63.32	B	Not Verified	...	Aug-10	64.27
38520	39665	118760	4000	4000	3575.0	36 months	7.43%	124.31	A	Not Verified	...	Aug-10	119.24

38521 rows × 38 columns

Lending Club – Recommendations

Few points to consider before providing loan to consumers:

- ✓ Consumer's 'Loan_Status' has to be in 'Fully Paid' or 'Charged Off' status but not 'Current' we can not determine a good Consumer if is still paying back the Loan amount until current loan tenure completes
- ✓ Consumer's 'chargeoff_within_12_mths' attribute has to be '0', indicates he/ she did not close any defaulted loan in last 12 months. Consumers who closed defaulted loans in last 12 months are not eligible for fresh loans.
- ✓ Consumer's 'open_acc' attribute should be in less number, if Consumer already have handful of loans to re-pay at the end of the month from his/ her most of the Salary, he/ she can not afford to pay the new loan.
- ✓ Re-pay capacity should be considered after calculating how much salary is being deducted in a month under loan installments, ideally after all deductions Consumer's in hand money should be more than 50% of total Salary.

Lending Club Case Study

Thank You