In [1]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
from datetime import datetime
```

In [2]:

```
data = pd.read_csv(r"C:\Users\omkar navale\OneDrive\Desktop\lending_club_loans 1.csv")
```

```
D:\anaconda3\lib\site-packages\IPython\core\interactiveshell.py:3444: DtypeWarning: Columns (0,49) have mixed
types.Specify dtype option on import or set low_memory=False.
  exec(code_obj, self.user_global_ns, self.user_ns)
```

In [3]:
```
1 data
```

Out[3]:

	id	member_id	loan_amnt	funded_amnt	funded_amnt_inv	term	Unnamed: 6	installment	grade	sub_grade	...	num_tl_
0	1077501	1296599.0	5000.0	5000.0	4975.0	36 months	10.65%	162.87	B	B2	...	
1	1077430	1314167.0	2500.0	2500.0	2500.0	60 months	15.27%	59.83	C	C4	...	
2	1077175	1313524.0	2400.0	2400.0	2400.0	36 months	15.96%	84.33	C	C5	...	
3	1076863	1277178.0	10000.0	10000.0	10000.0	36 months	13.49%	339.31	C	C1	...	
4	1075358	1311748.0	3000.0	3000.0	3000.0	60 months	12.69%	67.79	B	B5	...	
...	...	...	...	...	...	...	...	...	...	...	...	...
42537	70686	70681.0	5000.0	5000.0	0.0	36 months	7.75%	156.11	A	A3	...	
42538	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	
42539	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	
42540	Total amount funded in policy code 1: 460296150	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	
42541	Total amount funded in policy code 2: 0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	

42542 rows × 115 columns

In [4]:
```
1 data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 42542 entries, 0 to 42541
Columns: 115 entries, id to total_il_high_credit_limit
dtypes: float64(90), object(25)
memory usage: 37.3+ MB
```

In [5]:
```
1 data.head(5)
```

Out[5]:

	id	member_id	loan_amnt	funded_amnt	funded_amnt_inv	term	Unnamed: 6	installment	grade	sub_grade	...	num_tl_90g_dp
0	1077501	1296599.0	5000.0	5000.0	4975.0	36 months	10.65%	162.87	B	B2	...	
1	1077430	1314167.0	2500.0	2500.0	2500.0	60 months	15.27%	59.83	C	C4	...	
2	1077175	1313524.0	2400.0	2400.0	2400.0	36 months	15.96%	84.33	C	C5	...	
3	1076863	1277178.0	10000.0	10000.0	10000.0	36 months	13.49%	339.31	C	C1	...	
4	1075358	1311748.0	3000.0	3000.0	3000.0	60 months	12.69%	67.79	B	B5	...	

5 rows × 115 columns

In [6]:
```
1 data.shape
```

Out[6]: (42542, 115)

In [7]:
```
1 data.columns
```

Out[7]:
```
Index(['id', 'member_id', 'loan_amnt', 'funded_amnt', 'funded_amnt_inv',
       'term', 'Unnamed: 6', 'installment', 'grade', 'sub_grade',
       ...
       'num_tl_90g_dpd_24m', 'num_tl_op_past_12m', 'pct_tl_nvr_dlq',
       'percent_bc_gt_75', 'pub_rec_bankruptcies', 'tax_liens',
       'tot_hi_cred_lim', 'total_bal_ex_mort', 'total_bc_limit',
       'total_il_high_credit_limit'],
      dtype='object', length=115)
```

In [8]:
```
1 data.isnull().sum()
```

Out[8]:
```
id                            4
member_id                     7
loan_amnt                     7
funded_amnt                   7
funded_amnt_inv               7
                            ...
tax_liens                   112
tot_hi_cred_lim           42542
total_bal_ex_mort         42542
total_bc_limit            42542
total_il_high_credit_limit    42542
Length: 115, dtype: int64
```

In [9]:
```
1 data['loan_status'].unique()
```

Out[9]:
```
array(['Fully Paid', 'Charged Off', 'Current', 'In Grace Period',
       'Late (31-120 days)', 'Late (16-30 days)', 'Default', nan,
       'Does not meet the credit policy. Status:Fully Paid',
       'Does not meet the credit policy. Status:Charged Off'],
      dtype=object)
```

In [10]:
```
1 data=data[(data['loan_status'] == 'Fully Paid')| (data['loan_status'] == 'Charged Off')]
```

In [11]:
```
1 data['loan_status'].unique()
```

Out[11]:
```
array(['Fully Paid', 'Charged Off'], dtype=object)
```

In [12]:
```
1 data.shape
```

Out[12]:
```
(39239, 115)
```

In [13]:
```
1 data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 39239 entries, 0 to 39785
Columns: 115 entries, id to total_il_high_credit_limit
dtypes: float64(90), object(25)
memory usage: 34.7+ MB
```

In [14]:
```
1 data.dropna(axis=1, how='any', thresh=None, subset=None, inplace=False)
```

Out[14]:

	id	member_id	loan_amnt	funded_amnt	funded_amnt_inv	term	Unnamed: 6	installment	grade	sub_grade	...	total_rec_
0	1077501	1296599.0	5000.0	5000.0	4975.0	36 months	10.65%	162.87	B	B2	...	
1	1077430	1314167.0	2500.0	2500.0	2500.0	60 months	15.27%	59.83	C	C4	...	
2	1077175	1313524.0	2400.0	2400.0	2400.0	36 months	15.96%	84.33	C	C5	...	
3	1076863	1277178.0	10000.0	10000.0	10000.0	36 months	13.49%	339.31	C	C1	...	
5	1075269	1311441.0	5000.0	5000.0	5000.0	36 months	7.90%	156.46	A	A4	...	
...	...	...	...	...	...	...	...	...	...	...	...	
39781	92187	92174.0	2500.0	2500.0	1075.0	36 months	8.07%	78.42	A	A4	...	
39782	90665	90607.0	8500.0	8500.0	875.0	36 months	10.28%	275.38	C	C1	...	
39783	90395	90390.0	5000.0	5000.0	1325.0	36 months	8.07%	156.84	A	A4	...	
39784	90376	89243.0	5000.0	5000.0	650.0	36 months	7.43%	155.38	A	A2	...	
39785	87023	86999.0	7500.0	7500.0	800.0	36 months	13.75%	255.43	E	E2	...	

39239 rows × 47 columns

In [15]:
```
1 data.isnull().sum()
```

Out[15]:
```
id                                0
member_id                         0
loan_amnt                         0
funded_amnt                       0
funded_amnt_inv                   0
                                 ...
tax_liens                        39
tot_hi_cred_lim               39239
total_bal_ex_mort             39239
total_bc_limit                39239
total_il_high_credit_limit    39239
Length: 115, dtype: int64
```

In [16]:
```
1 null_val = []
2 for val in data:
3     if data[val].isna().sum()>35000:
4         null_val.append(val)
```

In [17]:
```
1 data[null_val]
```

Out[17]:

	mths_since_last_record	next_pymnt_d	mths_since_last_major_derog	annual_inc_joint	dti_joint	verification_status_joint	tot_col
0	NaN	NaN	NaN	NaN	NaN	NaN	
1	NaN	NaN	NaN	NaN	NaN	NaN	
2	NaN	NaN	NaN	NaN	NaN	NaN	
3	NaN	NaN	NaN	NaN	NaN	NaN	
5	NaN	NaN	NaN	NaN	NaN	NaN	
...	...	...	...	...	...	...	
39781	0.0	NaN	NaN	NaN	NaN	NaN	
39782	0.0	NaN	NaN	NaN	NaN	NaN	
39783	0.0	NaN	NaN	NaN	NaN	NaN	
39784	0.0	NaN	NaN	NaN	NaN	NaN	
39785	0.0	NaN	NaN	NaN	NaN	NaN	

39239 rows × 56 columns

In [18]:
```
1 data.drop(columns=null_val,axis=1,inplace=True)
```

```
D:\anaconda3\lib\site-packages\pandas\core\frame.py:4906: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy (https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)
  return super().drop(
```

In [19]:
```
1 data
```

Out[19]:

	id	member_id	loan_amnt	funded_amnt	funded_amnt_inv	term	Unnamed: 6	installment	grade	sub_grade	...	last_fico_r
0	1077501	1296599.0	5000.0	5000.0	4975.0	36 months	10.65%	162.87	B	B2	...	
1	1077430	1314167.0	2500.0	2500.0	2500.0	60 months	15.27%	59.83	C	C4	...	
2	1077175	1313524.0	2400.0	2400.0	2400.0	36 months	15.96%	84.33	C	C5	...	
3	1076863	1277178.0	10000.0	10000.0	10000.0	36 months	13.49%	339.31	C	C1	...	
5	1075269	1311441.0	5000.0	5000.0	5000.0	36 months	7.90%	156.46	A	A4	...	
...	...	...	...	...	...	...	...	...	...	...	...	
39781	92187	92174.0	2500.0	2500.0	1075.0	36 months	8.07%	78.42	A	A4	...	
39782	90665	90607.0	8500.0	8500.0	875.0	36 months	10.28%	275.38	C	C1	...	
39783	90395	90390.0	5000.0	5000.0	1325.0	36 months	8.07%	156.84	A	A4	...	
39784	90376	89243.0	5000.0	5000.0	650.0	36 months	7.43%	155.38	A	A2	...	
39785	87023	86999.0	7500.0	7500.0	800.0	36 months	13.75%	255.43	E	E2	...	

39239 rows × 59 columns

In [20]:
```
1 data.isnull().sum()
```

Out[20]:

id	0
member_id	0
loan_amnt	0
funded_amnt	0
funded_amnt_inv	0
term	0
Unnamed: 6	0
installment	0
grade	0
sub_grade	0
emp_title	2427
emp_length	1057
home_ownership	0
annual_inc	0
verification_status	0
issue_d	0
loan_status	0
pymnt_plan	0
url	0
desc	12766
purpose	0
title	11
zip_code	0
addr_state	0
dti	0
delinq_2yrs	0
earliest_cr_line	0
fico_range_low	0
fico_range_high	0
inq_last_6mths	0
mths_since_last_delinq	25352
open_acc	0
pub_rec	0
revol_bal	0
revol_util	50
total_acc	0
initial_list_status	0
out_prncp	0
out_prncp_inv	0
total_pymnt	0

```
total_pymnt_inv                0
total_rec_prncp                0
total_rec_int                  0
total_rec_late_fee             0
recoveries                     0
collection_recovery_fee        0
last_pymnt_d                  71
last_pymnt_amnt                0
last_credit_pull_d             2
last_fico_range_high           0
last_fico_range_low            0
collections_12_mths_ex_med    56
policy_code                    0
application_type               0
acc_now_delinq                 0
chargeoff_within_12_mths      56
delinq_amnt                    0
pub_rec_bankruptcies         697
tax_liens                     39
dtype: int64
```

In [21]:

```
1 data.drop(columns=['desc','mths_since_last_delinq'],inplace=True)
```

```
D:\anaconda3\lib\site-packages\pandas\core\frame.py:4906: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy (https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)
  return super().drop(
```

In [22]:
```
1 data.isnull().sum()
```

Out[22]:

id	0
member_id	0
loan_amnt	0
funded_amnt	0
funded_amnt_inv	0
term	0
Unnamed: 6	0
installment	0
grade	0
sub_grade	0
emp_title	2427
emp_length	1057
home_ownership	0
annual_inc	0
verification_status	0
issue_d	0
loan_status	0
pymnt_plan	0
url	0
purpose	0
title	11
zip_code	0
addr_state	0
dti	0
delinq_2yrs	0
earliest_cr_line	0
fico_range_low	0
fico_range_high	0
inq_last_6mths	0
open_acc	0
pub_rec	0
revol_bal	0
revol_util	50
total_acc	0
initial_list_status	0
out_prncp	0
out_prncp_inv	0
total_pymnt	0
total_pymnt_inv	0
total_rec_prncp	0

```
total_rec_int                  0
total_rec_late_fee             0
recoveries                     0
collection_recovery_fee        0
last_pymnt_d                  71
last_pymnt_amnt                0
last_credit_pull_d             2
last_fico_range_high           0
last_fico_range_low            0
collections_12_mths_ex_med    56
policy_code                    0
application_type               0
acc_now_delinq                 0
chargeoff_within_12_mths      56
delinq_amnt                    0
pub_rec_bankruptcies         697
tax_liens                     39
dtype: int64
```

In [23]:
```
1 data['emp_title'].fillna('Other',inplace=True)
```

```
D:\anaconda3\lib\site-packages\pandas\core\generic.py:6392: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy (https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)
  return self._update_inplace(result)
```

In [24]:
```
1 data.isnull().sum()
```

Out[24]:

id	0
member_id	0
loan_amnt	0
funded_amnt	0
funded_amnt_inv	0
term	0
Unnamed: 6	0
installment	0
grade	0
sub_grade	0
emp_title	0
emp_length	1057
home_ownership	0
annual_inc	0
verification_status	0
issue_d	0
loan_status	0
pymnt_plan	0
url	0
purpose	0
title	11
zip_code	0
addr_state	0
dti	0
delinq_2yrs	0
earliest_cr_line	0
fico_range_low	0
fico_range_high	0
inq_last_6mths	0
open_acc	0
pub_rec	0
revol_bal	0
revol_util	50
total_acc	0
initial_list_status	0
out_prncp	0
out_prncp_inv	0
total_pymnt	0
total_pymnt_inv	0
total_rec_prncp	0

```
total_rec_int                 0
total_rec_late_fee            0
recoveries                    0
collection_recovery_fee       0
last_pymnt_d                 71
last_pymnt_amnt               0
last_credit_pull_d            2
last_fico_range_high          0
last_fico_range_low           0
collections_12_mths_ex_med   56
policy_code                   0
application_type              0
acc_now_delinq                0
chargeoff_within_12_mths     56
delinq_amnt                   0
pub_rec_bankruptcies        697
tax_liens                    39
dtype: int64
```

In [25]:

```
1 data['title'].fillna('Other',inplace=True)
```

In [26]:
```
1 data.isnull().sum()
```

Out[26]:

id	0
member_id	0
loan_amnt	0
funded_amnt	0
funded_amnt_inv	0
term	0
Unnamed: 6	0
installment	0
grade	0
sub_grade	0
emp_title	0
emp_length	1057
home_ownership	0
annual_inc	0
verification_status	0
issue_d	0
loan_status	0
pymnt_plan	0
url	0
purpose	0
title	0
zip_code	0
addr_state	0
dti	0
delinq_2yrs	0
earliest_cr_line	0
fico_range_low	0
fico_range_high	0
inq_last_6mths	0
open_acc	0
pub_rec	0
revol_bal	0
revol_util	50
total_acc	0
initial_list_status	0
out_prncp	0
out_prncp_inv	0
total_pymnt	0
total_pymnt_inv	0
total_rec_prncp	0

```
total_rec_int                 0
total_rec_late_fee            0
recoveries                    0
collection_recovery_fee       0
last_pymnt_d                 71
last_pymnt_amnt               0
last_credit_pull_d            2
last_fico_range_high          0
last_fico_range_low           0
collections_12_mths_ex_med   56
policy_code                   0
application_type              0
acc_now_delinq                0
chargeoff_within_12_mths     56
delinq_amnt                   0
pub_rec_bankruptcies        697
tax_liens                    39
dtype: int64
```

In [27]:

```
1 data['emp_length'].fillna('< 1 year',inplace=True)
```

In [28]:
```
1 data.isnull().sum()
```

Out[28]:

id	0
member_id	0
loan_amnt	0
funded_amnt	0
funded_amnt_inv	0
term	0
Unnamed: 6	0
installment	0
grade	0
sub_grade	0
emp_title	0
emp_length	0
home_ownership	0
annual_inc	0
verification_status	0
issue_d	0
loan_status	0
pymnt_plan	0
url	0
purpose	0
title	0
zip_code	0
addr_state	0
dti	0
delinq_2yrs	0
earliest_cr_line	0
fico_range_low	0
fico_range_high	0
inq_last_6mths	0
open_acc	0
pub_rec	0
revol_bal	0
revol_util	50
total_acc	0
initial_list_status	0
out_prncp	0
out_prncp_inv	0
total_pymnt	0
total_pymnt_inv	0
total_rec_prncp	0

```
total_rec_int                 0
total_rec_late_fee            0
recoveries                    0
collection_recovery_fee       0
last_pymnt_d                 71
last_pymnt_amnt               0
last_credit_pull_d            2
last_fico_range_high          0
last_fico_range_low           0
collections_12_mths_ex_med   56
policy_code                   0
application_type              0
acc_now_delinq                0
chargeoff_within_12_mths     56
delinq_amnt                   0
pub_rec_bankruptcies        697
tax_liens                    39
dtype: int64
```

In [29]:

```
1 data['inq_last_6mths'].fillna(0,inplace=True)
```

In [30]:
```
1 data.isnull().sum()
```

Out[30]:

id	0
member_id	0
loan_amnt	0
funded_amnt	0
funded_amnt_inv	0
term	0
Unnamed: 6	0
installment	0
grade	0
sub_grade	0
emp_title	0
emp_length	0
home_ownership	0
annual_inc	0
verification_status	0
issue_d	0
loan_status	0
pymnt_plan	0
url	0
purpose	0
title	0
zip_code	0
addr_state	0
dti	0
delinq_2yrs	0
earliest_cr_line	0
fico_range_low	0
fico_range_high	0
inq_last_6mths	0
open_acc	0
pub_rec	0
revol_bal	0
revol_util	50
total_acc	0
initial_list_status	0
out_prncp	0
out_prncp_inv	0
total_pymnt	0
total_pymnt_inv	0
total_rec_prncp	0

```
total_rec_int                 0
total_rec_late_fee            0
recoveries                    0
collection_recovery_fee       0
last_pymnt_d                 71
last_pymnt_amnt               0
last_credit_pull_d            2
last_fico_range_high          0
last_fico_range_low           0
collections_12_mths_ex_med   56
policy_code                   0
application_type              0
acc_now_delinq                0
chargeoff_within_12_mths     56
delinq_amnt                   0
pub_rec_bankruptcies        697
tax_liens                    39
dtype: int64
```

In [31]:

```
1 data['revol_util'].fillna('0%', inplace=True)
```

In [32]:
```
1 data.isnull().sum()
```

Out[32]:

id	0
member_id	0
loan_amnt	0
funded_amnt	0
funded_amnt_inv	0
term	0
Unnamed: 6	0
installment	0
grade	0
sub_grade	0
emp_title	0
emp_length	0
home_ownership	0
annual_inc	0
verification_status	0
issue_d	0
loan_status	0
pymnt_plan	0
url	0
purpose	0
title	0
zip_code	0
addr_state	0
dti	0
delinq_2yrs	0
earliest_cr_line	0
fico_range_low	0
fico_range_high	0
inq_last_6mths	0
open_acc	0
pub_rec	0
revol_bal	0
revol_util	0
total_acc	0
initial_list_status	0
out_prncp	0
out_prncp_inv	0
total_pymnt	0
total_pymnt_inv	0
total_rec_prncp	0

```
total_rec_int                 0
total_rec_late_fee            0
recoveries                    0
collection_recovery_fee       0
last_pymnt_d                 71
last_pymnt_amnt               0
last_credit_pull_d            2
last_fico_range_high          0
last_fico_range_low           0
collections_12_mths_ex_med   56
policy_code                   0
application_type              0
acc_now_delinq                0
chargeoff_within_12_mths     56
delinq_amnt                   0
pub_rec_bankruptcies        697
tax_liens                    39
dtype: int64
```

In [33]:

```
1 data['pub_rec_bankruptcies'].fillna('0',inplace=True)
```

In [34]:
```
1 data.isnull().sum()
```

Out[34]:

```
id                          0
member_id                   0
loan_amnt                   0
funded_amnt                 0
funded_amnt_inv             0
term                        0
Unnamed: 6                  0
installment                 0
grade                       0
sub_grade                   0
emp_title                   0
emp_length                  0
home_ownership              0
annual_inc                  0
verification_status         0
issue_d                     0
loan_status                 0
pymnt_plan                  0
url                         0
purpose                     0
title                       0
zip_code                    0
addr_state                  0
dti                         0
delinq_2yrs                 0
earliest_cr_line            0
fico_range_low              0
fico_range_high             0
inq_last_6mths              0
open_acc                    0
pub_rec                     0
revol_bal                   0
revol_util                  0
total_acc                   0
initial_list_status         0
out_prncp                   0
out_prncp_inv               0
total_pymnt                 0
total_pymnt_inv             0
total_rec_prncp             0
```

```
total_rec_int                 0
total_rec_late_fee            0
recoveries                    0
collection_recovery_fee       0
last_pymnt_d                 71
last_pymnt_amnt               0
last_credit_pull_d            2
last_fico_range_high          0
last_fico_range_low           0
collections_12_mths_ex_med    56
policy_code                   0
application_type              0
acc_now_delinq                0
chargeoff_within_12_mths      56
delinq_amnt                   0
pub_rec_bankruptcies          0
tax_liens                     39
dtype: int64
```

In [35]:

```
1 data['collections_12_mths_ex_med'].fillna('0',inplace=True)
```

In [36]:
```
1 data.isnull().sum()
```

Out[36]:

```
id                          0
member_id                   0
loan_amnt                   0
funded_amnt                 0
funded_amnt_inv             0
term                        0
Unnamed: 6                  0
installment                 0
grade                       0
sub_grade                   0
emp_title                   0
emp_length                  0
home_ownership              0
annual_inc                  0
verification_status         0
issue_d                     0
loan_status                 0
pymnt_plan                  0
url                         0
purpose                     0
title                       0
zip_code                    0
addr_state                  0
dti                         0
delinq_2yrs                 0
earliest_cr_line            0
fico_range_low              0
fico_range_high             0
inq_last_6mths              0
open_acc                    0
pub_rec                     0
revol_bal                   0
revol_util                  0
total_acc                   0
initial_list_status         0
out_prncp                   0
out_prncp_inv               0
total_pymnt                 0
total_pymnt_inv             0
total_rec_prncp             0
```

```
total_rec_int                 0
total_rec_late_fee            0
recoveries                    0
collection_recovery_fee       0
last_pymnt_d                 71
last_pymnt_amnt               0
last_credit_pull_d            2
last_fico_range_high          0
last_fico_range_low           0
collections_12_mths_ex_med    0
policy_code                   0
application_type              0
acc_now_delinq                0
chargeoff_within_12_mths     56
delinq_amnt                   0
pub_rec_bankruptcies          0
tax_liens                    39
dtype: int64
```

In [37]:
```
1 data['last_credit_pull_d'].value_counts()
```

Out[37]:
```
Sep-16    14555
Mar-16      827
Aug-16      716
Apr-16      664
Feb-13      663
           ...
May-08        1
Jun-08        1
Jul-08        1
May-07        1
Jul-07        1
Name: last_credit_pull_d, Length: 110, dtype: int64
```

In [38]:
```
1 data['last_credit_pull_d'].fillna('Jul-08',inplace=True)
```

In [39]:
```
1 data.isnull().sum()
```

Out[39]:

id	0
member_id	0
loan_amnt	0
funded_amnt	0
funded_amnt_inv	0
term	0
Unnamed: 6	0
installment	0
grade	0
sub_grade	0
emp_title	0
emp_length	0
home_ownership	0
annual_inc	0
verification_status	0
issue_d	0
loan_status	0
pymnt_plan	0
url	0
purpose	0
title	0
zip_code	0
addr_state	0
dti	0
delinq_2yrs	0
earliest_cr_line	0
fico_range_low	0
fico_range_high	0
inq_last_6mths	0
open_acc	0
pub_rec	0
revol_bal	0
revol_util	0
total_acc	0
initial_list_status	0
out_prncp	0
out_prncp_inv	0
total_pymnt	0
total_pymnt_inv	0
total_rec_prncp	0

```
total_rec_int                 0
total_rec_late_fee            0
recoveries                    0
collection_recovery_fee       0
last_pymnt_d                 71
last_pymnt_amnt               0
last_credit_pull_d            0
last_fico_range_high          0
last_fico_range_low           0
collections_12_mths_ex_med    0
policy_code                   0
application_type              0
acc_now_delinq                0
chargeoff_within_12_mths     56
delinq_amnt                   0
pub_rec_bankruptcies          0
tax_liens                    39
dtype: int64
```

In [40]:

```
1 data['last_pymnt_d'].fillna('Dec-07',inplace=True)
```

In [41]:
```
1 data.isnull().sum()
```

Out[41]:

```
id                          0
member_id                   0
loan_amnt                   0
funded_amnt                 0
funded_amnt_inv             0
term                        0
Unnamed: 6                  0
installment                 0
grade                       0
sub_grade                   0
emp_title                   0
emp_length                  0
home_ownership              0
annual_inc                  0
verification_status         0
issue_d                     0
loan_status                 0
pymnt_plan                  0
url                         0
purpose                     0
title                       0
zip_code                    0
addr_state                  0
dti                         0
delinq_2yrs                 0
earliest_cr_line            0
fico_range_low              0
fico_range_high             0
inq_last_6mths              0
open_acc                    0
pub_rec                     0
revol_bal                   0
revol_util                  0
total_acc                   0
initial_list_status         0
out_prncp                   0
out_prncp_inv               0
total_pymnt                 0
total_pymnt_inv             0
total_rec_prncp             0
```

```
total_rec_int                 0
total_rec_late_fee            0
recoveries                    0
collection_recovery_fee       0
last_pymnt_d                  0
last_pymnt_amnt               0
last_credit_pull_d            0
last_fico_range_high          0
last_fico_range_low           0
collections_12_mths_ex_med    0
policy_code                   0
application_type              0
acc_now_delinq                0
chargeoff_within_12_mths     56
delinq_amnt                   0
pub_rec_bankruptcies          0
tax_liens                    39
dtype: int64
```

In [42]:
```
1 data['chargeoff_within_12_mths'].fillna(0,inplace=True)
```

In [43]:
```
1 data['tax_liens'].fillna(0,inplace=True)
```

In [44]:
```
1 data.isnull().sum()
```

Out[44]:

id	0
member_id	0
loan_amnt	0
funded_amnt	0
funded_amnt_inv	0
term	0
Unnamed: 6	0
installment	0
grade	0
sub_grade	0
emp_title	0
emp_length	0
home_ownership	0
annual_inc	0
verification_status	0
issue_d	0
loan_status	0
pymnt_plan	0
url	0
purpose	0
title	0
zip_code	0
addr_state	0
dti	0
delinq_2yrs	0
earliest_cr_line	0
fico_range_low	0
fico_range_high	0
inq_last_6mths	0
open_acc	0
pub_rec	0
revol_bal	0
revol_util	0
total_acc	0
initial_list_status	0
out_prncp	0
out_prncp_inv	0
total_pymnt	0
total_pymnt_inv	0
total_rec_prncp	0

```
total_rec_int                0
total_rec_late_fee           0
recoveries                   0
collection_recovery_fee      0
last_pymnt_d                 0
last_pymnt_amnt              0
last_credit_pull_d           0
last_fico_range_high         0
last_fico_range_low          0
collections_12_mths_ex_med   0
policy_code                  0
application_type             0
acc_now_delinq               0
chargeoff_within_12_mths     0
delinq_amnt                  0
pub_rec_bankruptcies         0
tax_liens                    0
dtype: int64
```

In [45]:
```
data.drop(["tax_liens","delinq_amnt","chargeoff_within_12_mths","acc_now_delinq","collections_12_mths_ex_me
```

In [46]:
```
data.shape
```

Out[46]: (39239, 52)

In [47]:
```
1 data.info().sum()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 39239 entries, 0 to 39785
Data columns (total 52 columns):
 #   Column                   Non-Null Count  Dtype
---  ------                   --------------  -----
 0   id                       39239 non-null  object
 1   member_id                39239 non-null  float64
 2   loan_amnt                39239 non-null  float64
 3   funded_amnt              39239 non-null  float64
 4   funded_amnt_inv          39239 non-null  float64
 5   term                     39239 non-null  object
 6   Unnamed: 6               39239 non-null  object
 7   installment              39239 non-null  float64
 8   grade                    39239 non-null  object
 9   sub_grade                39239 non-null  object
 10  emp_title                39239 non-null  object
 11  emp_length               39239 non-null  object
 12  home_ownership           39239 non-null  object
 13  annual_inc               39239 non-null  float64
 14  verification_status      39239 non-null  object
 15  issue_d                  39239 non-null  object
 16  loan_status              39239 non-null  object
 17  pymnt_plan               39239 non-null  object
 18  url                      39239 non-null  object
 19  purpose                  39239 non-null  object
 20  title                    39239 non-null  object
 21  zip_code                 39239 non-null  object
 22  addr_state               39239 non-null  object
 23  dti                      39239 non-null  float64
 24  delinq_2yrs              39239 non-null  float64
 25  earliest_cr_line         39239 non-null  object
 26  fico_range_low           39239 non-null  float64
 27  fico_range_high          39239 non-null  float64
 28  inq_last_6mths           39239 non-null  float64
 29  open_acc                 39239 non-null  float64
 30  pub_rec                  39239 non-null  float64
 31  revol_bal                39239 non-null  float64
 32  revol_util               39239 non-null  object
 33  total_acc                39239 non-null  float64
 34  initial_list_status      39239 non-null  object
```

```
 35  out_prncp                39239 non-null  float64
 36  out_prncp_inv            39239 non-null  float64
 37  total_pymnt              39239 non-null  float64
 38  total_pymnt_inv          39239 non-null  float64
 39  total_rec_prncp          39239 non-null  float64
 40  total_rec_int            39239 non-null  float64
 41  total_rec_late_fee       39239 non-null  float64
 42  recoveries               39239 non-null  float64
 43  collection_recovery_fee  39239 non-null  float64
 44  last_pymnt_d             39239 non-null  object
 45  last_pymnt_amnt          39239 non-null  float64
 46  last_credit_pull_d       39239 non-null  object
 47  last_fico_range_high     39239 non-null  float64
 48  last_fico_range_low      39239 non-null  float64
 49  policy_code              39239 non-null  float64
 50  application_type         39239 non-null  object
 51  pub_rec_bankruptcies     39239 non-null  object
dtypes: float64(28), object(24)
memory usage: 15.9+ MB


---------------------------------------------------------------------------
AttributeError                            Traceback (most recent call last)
C:\Users\OMKARN~1\AppData\Local\Temp/ipykernel_30320/2945938051.py in <module>
----> 1 data.info().sum()

AttributeError: 'NoneType' object has no attribute 'sum'
```

In [48]:
```
1 data['member_id'] = data['member_id'].astype('int')
```

```
C:\Users\OMKARN~1\AppData\Local\Temp/ipykernel_30320/4088084349.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy (https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)
  data['member_id'] = data['member_id'].astype('int')
```

In [49]:
```
1 data.info().sum()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 39239 entries, 0 to 39785
Data columns (total 52 columns):
 #   Column                   Non-Null Count  Dtype
---  ------                   --------------  -----
 0   id                       39239 non-null  object
 1   member_id                39239 non-null  int32
 2   loan_amnt                39239 non-null  float64
 3   funded_amnt              39239 non-null  float64
 4   funded_amnt_inv          39239 non-null  float64
 5   term                     39239 non-null  object
 6   Unnamed: 6               39239 non-null  object
 7   installment              39239 non-null  float64
 8   grade                    39239 non-null  object
 9   sub_grade                39239 non-null  object
 10  emp_title                39239 non-null  object
 11  emp_length               39239 non-null  object
 12  home_ownership           39239 non-null  object
 13  annual_inc               39239 non-null  float64
 14  verification_status      39239 non-null  object
 15  issue_d                  39239 non-null  object
 16  loan_status              39239 non-null  object
 17  pymnt_plan               39239 non-null  object
 18  url                      39239 non-null  object
 19  purpose                  39239 non-null  object
 20  title                    39239 non-null  object
 21  zip_code                 39239 non-null  object
 22  addr_state               39239 non-null  object
 23  dti                      39239 non-null  float64
 24  delinq_2yrs              39239 non-null  float64
 25  earliest_cr_line         39239 non-null  object
 26  fico_range_low           39239 non-null  float64
 27  fico_range_high          39239 non-null  float64
 28  inq_last_6mths           39239 non-null  float64
 29  open_acc                 39239 non-null  float64
 30  pub_rec                  39239 non-null  float64
 31  revol_bal                39239 non-null  float64
 32  revol_util               39239 non-null  object
 33  total_acc                39239 non-null  float64
 34  initial_list_status      39239 non-null  object
```

```
 35  out_prncp                39239 non-null  float64
 36  out_prncp_inv            39239 non-null  float64
 37  total_pymnt              39239 non-null  float64
 38  total_pymnt_inv          39239 non-null  float64
 39  total_rec_prncp          39239 non-null  float64
 40  total_rec_int            39239 non-null  float64
 41  total_rec_late_fee       39239 non-null  float64
 42  recoveries               39239 non-null  float64
 43  collection_recovery_fee  39239 non-null  float64
 44  last_pymnt_d             39239 non-null  object
 45  last_pymnt_amnt          39239 non-null  float64
 46  last_credit_pull_d       39239 non-null  object
 47  last_fico_range_high     39239 non-null  float64
 48  last_fico_range_low      39239 non-null  float64
 49  policy_code              39239 non-null  float64
 50  application_type         39239 non-null  object
 51  pub_rec_bankruptcies     39239 non-null  object
dtypes: float64(27), int32(1), object(24)
memory usage: 15.7+ MB
```

```
---------------------------------------------------------------------------
AttributeError                            Traceback (most recent call last)
C:\Users\OMKARN~1\AppData\Local\Temp/ipykernel_30320/2945938051.py in <module>
----> 1 data.info().sum()

AttributeError: 'NoneType' object has no attribute 'sum'
```

In [58]:
```
1 data.shape
```

Out[58]:
```
(39239, 52)
```

Data Cleaning And New Data

In [51]:
```
1 new_data = data.copy()
```

In [52]:
```
1 new_data.shape
```

Out[52]:
```
(39239, 52)
```

In [53]:
```
1 new_data.describe()
```

Out[53]:

	member_id	loan_amnt	funded_amnt	funded_amnt_inv	installment	annual_inc	dti	delinq_2yrs	fico_range_l
count	3.923900e+04	39239.000000	39239.000000	39239.000000	39239.000000	3.923900e+04	39239.000000	39239.000000	39239.000(
mean	8.456767e+05	11134.730115	10863.217080	10307.088303	323.273499	6.888432e+04	13.293984	0.146563	715.0007
std	2.637546e+05	7398.238030	7126.007647	7060.672210	208.463559	6.400031e+04	6.676607	0.491534	35.8681
min	7.069900e+04	500.000000	500.000000	0.000000	15.690000	4.000000e+03	0.000000	0.000000	625.000(
25%	6.642775e+05	5400.000000	5300.000000	5000.000000	166.305000	4.001400e+04	8.160000	0.000000	685.000(
50%	8.459240e+05	10000.000000	9600.000000	8875.000000	279.010000	5.900000e+04	13.390000	0.000000	710.000(
75%	1.040192e+06	15000.000000	15000.000000	14200.000000	427.280000	8.200000e+04	18.570000	0.000000	740.000(
max	1.314167e+06	35000.000000	35000.000000	35000.000000	1305.190000	6.000000e+06	29.990000	11.000000	825.000(

8 rows × 28 columns

In [ ]:
```
1
```

Data Visualization

In [54]:
```
1 new_data
```

Out[54]:

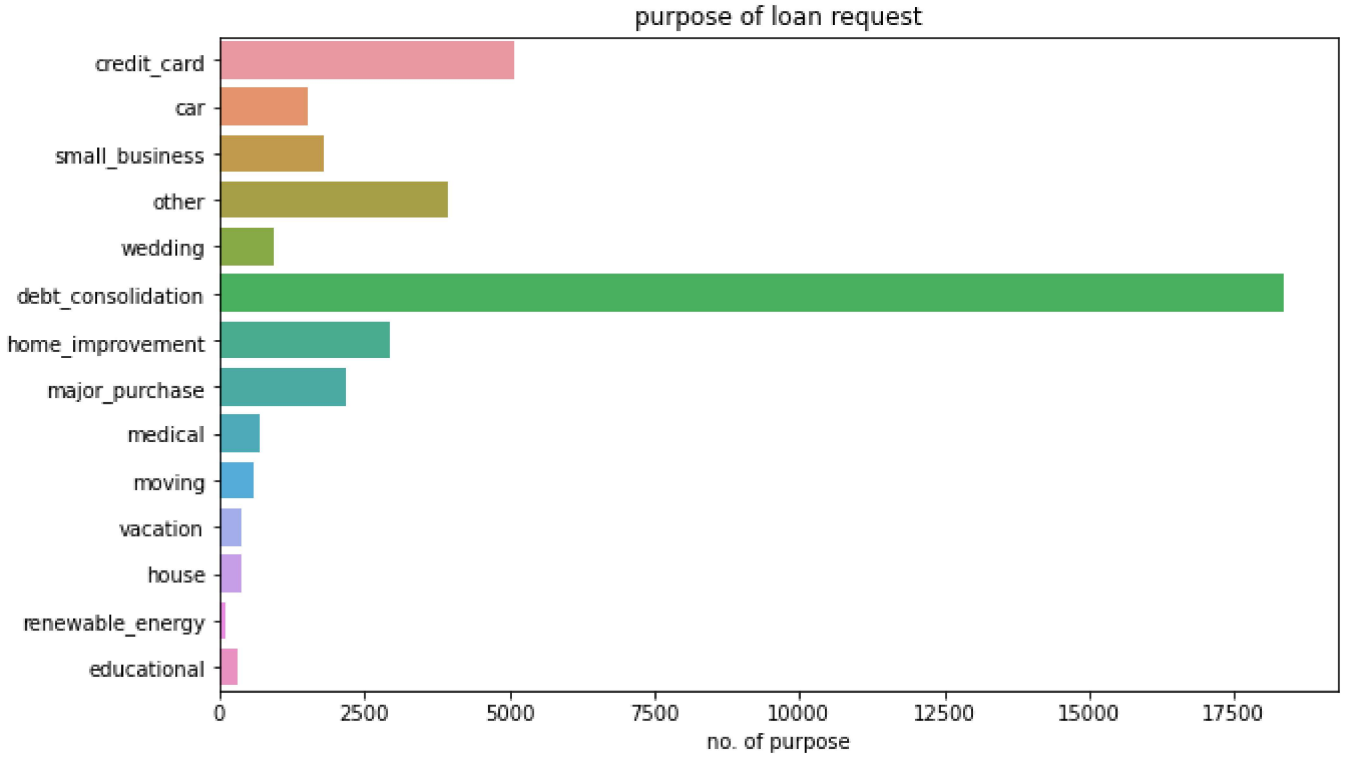
	id	member_id	loan_amnt	funded_amnt	funded_amnt_inv	term	Unnamed: 6	installment	grade	sub_grade	...	recoveries
0	1077501	1296599	5000.0	5000.0	4975.0	36 months	10.65%	162.87	B	B2	...	0.0(
1	1077430	1314167	2500.0	2500.0	2500.0	60 months	15.27%	59.83	C	C4	...	117.08
2	1077175	1313524	2400.0	2400.0	2400.0	36 months	15.96%	84.33	C	C5	...	0.0(
3	1076863	1277178	10000.0	10000.0	10000.0	36 months	13.49%	339.31	C	C1	...	0.0(
5	1075269	1311441	5000.0	5000.0	5000.0	36 months	7.90%	156.46	A	A4	...	0.0(
...	...	...	...	...	...	...	...	...	...	...	...	..
39781	92187	92174	2500.0	2500.0	1075.0	36 months	8.07%	78.42	A	A4	...	0.0(
39782	90665	90607	8500.0	8500.0	875.0	36 months	10.28%	275.38	C	C1	...	0.0(
39783	90395	90390	5000.0	5000.0	1325.0	36 months	8.07%	156.84	A	A4	...	0.0(
39784	90376	89243	5000.0	5000.0	650.0	36 months	7.43%	155.38	A	A2	...	0.0(
39785	87023	86999	7500.0	7500.0	800.0	36 months	13.75%	255.43	E	E2	...	0.0(

39239 rows × 52 columns

Countplot

In [55]:

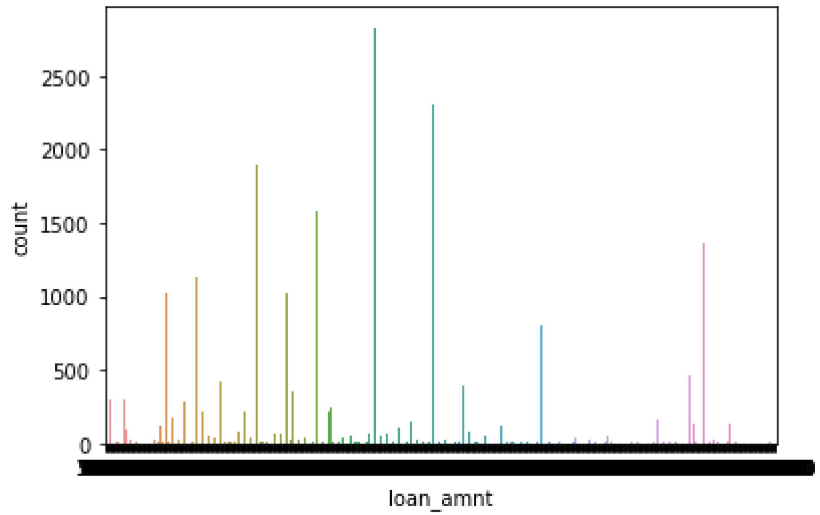
```
plt.figure(figsize=(10,6))
sns.countplot(y='purpose',data=new_data)
plt.title('purpose of loan request')
plt.xlabel('no. of purpose')
plt.ylabel('')
plt.show()
```



In [56]:
```
1 sns.countplot(new_data['loan_amnt'])
```

```
D:\anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variable as a keyw
ord arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments
without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
```

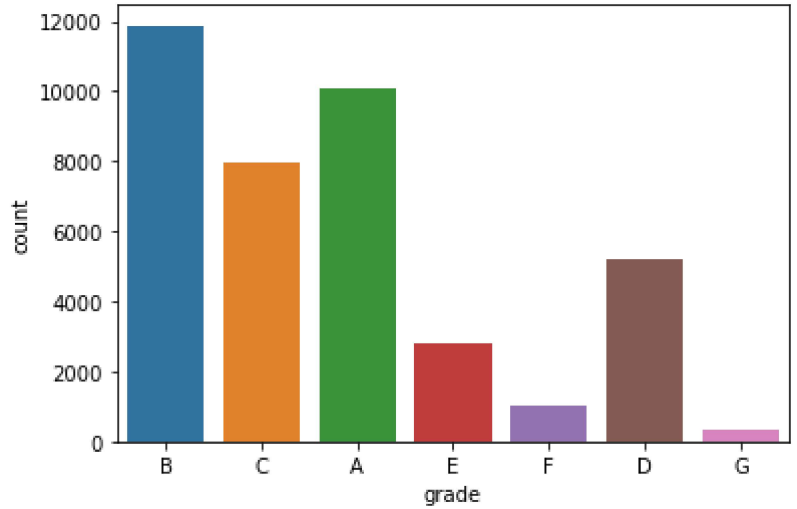
Out[56]: <AxesSubplot:xlabel='loan_amnt', ylabel='count'>



In [57]:
```
1 sns.countplot(new_data['grade'])
```

```
D:\anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variable as a keyw
ord arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments
without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
```

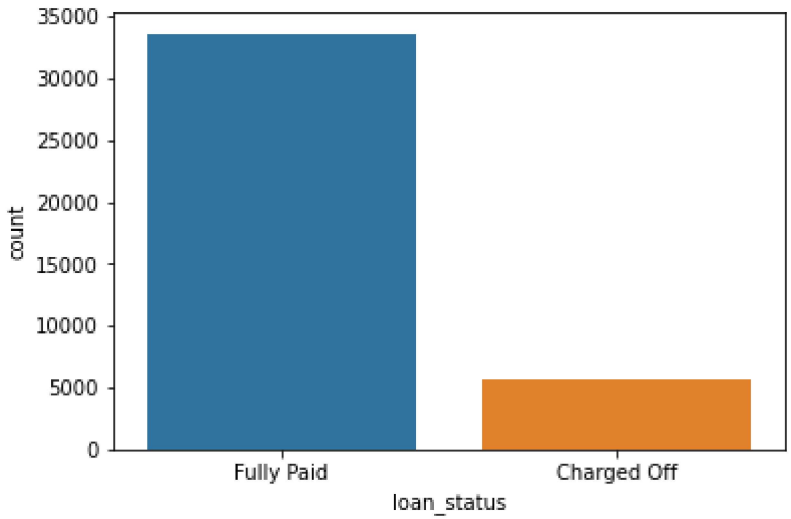
Out[57]: <AxesSubplot:xlabel='grade', ylabel='count'>



In [59]:
```
1 sns.countplot(new_data['loan_status'])
```

```
D:\anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variable as a keyw
ord arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments
without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
```

Out[59]: <AxesSubplot:xlabel='loan_status', ylabel='count'>

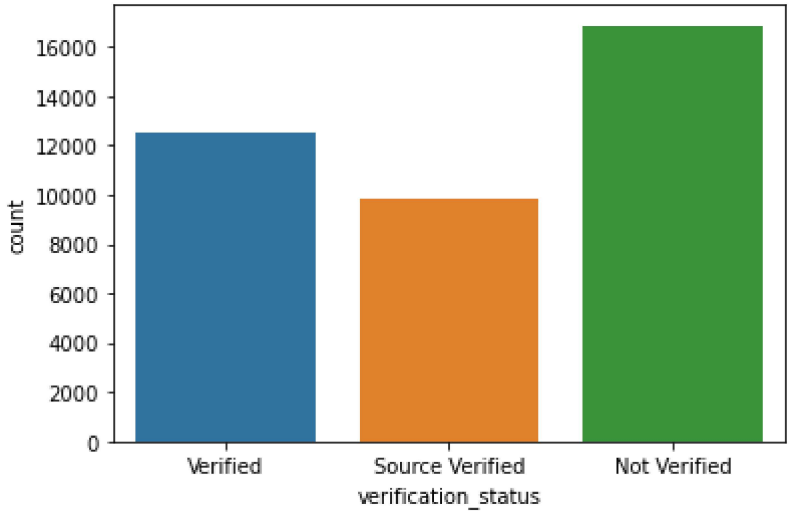


In [ ]:
```
1
```

In [60]:
```
1 sns.countplot(new_data['verification_status'])
```

```
D:\anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variable as a keyw
ord arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments
without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
```

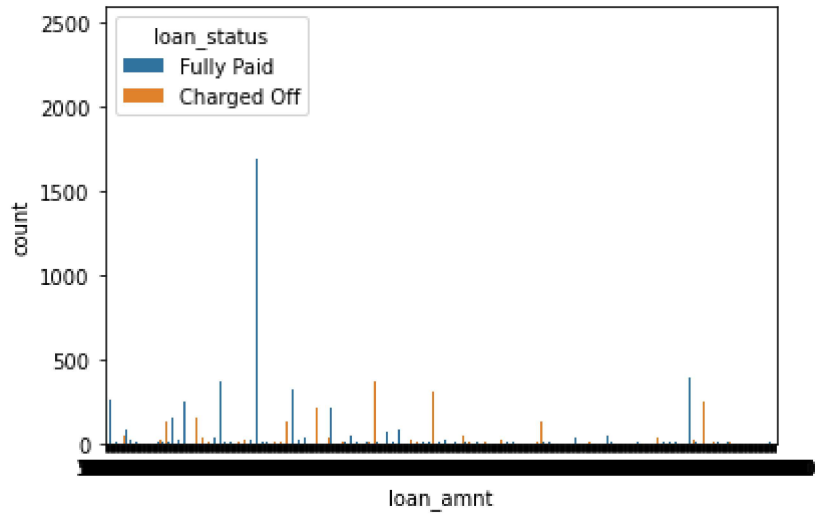
Out[60]: <AxesSubplot:xlabel='verification_status', ylabel='count'>



In [ ]:
```
1
```

In [61]:
```
1 sns.countplot(x='loan_amnt',data=new_data,hue='loan_status')
```

Out[61]: <AxesSubplot:xlabel='loan_amnt', ylabel='count'>

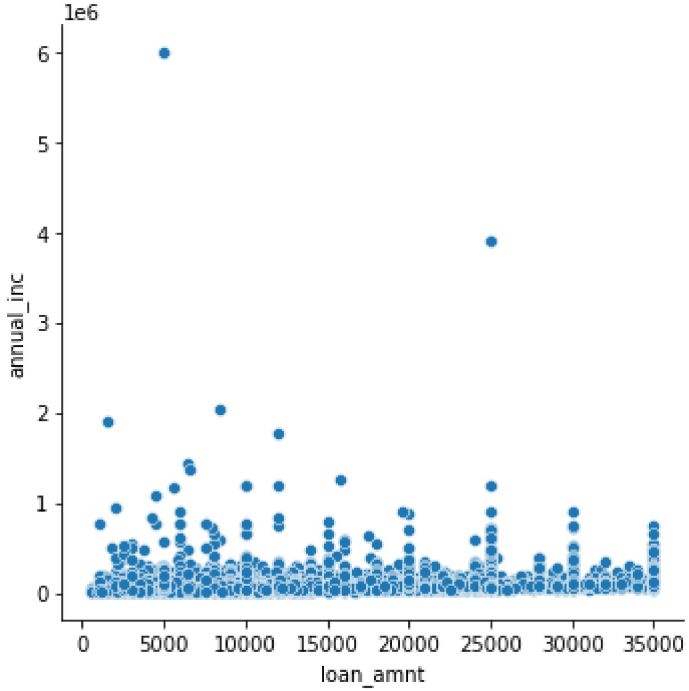


In [ ]:
```
1
```

Replot

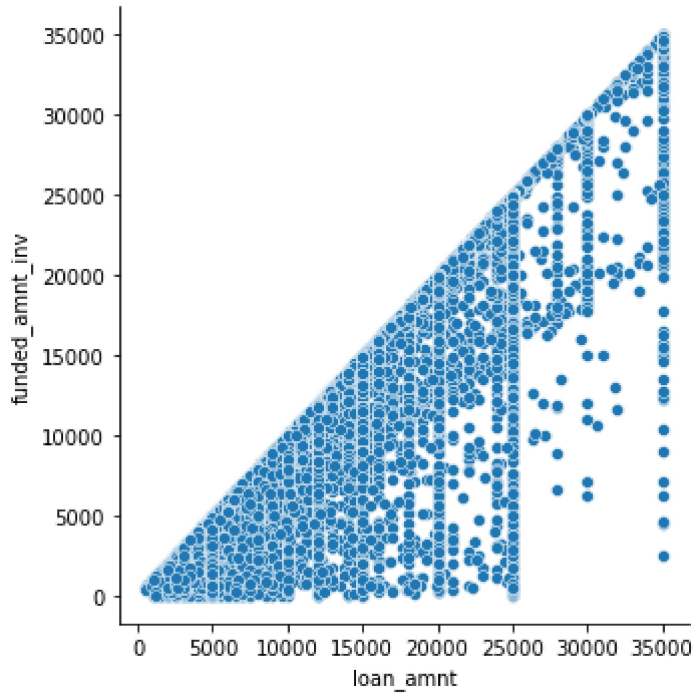
In [62]:
```
1 sns.relplot(x='loan_amnt',y='annual_inc',data=new_data)
```

Out[62]: <seaborn.axisgrid.FacetGrid at 0x1ff29773100>



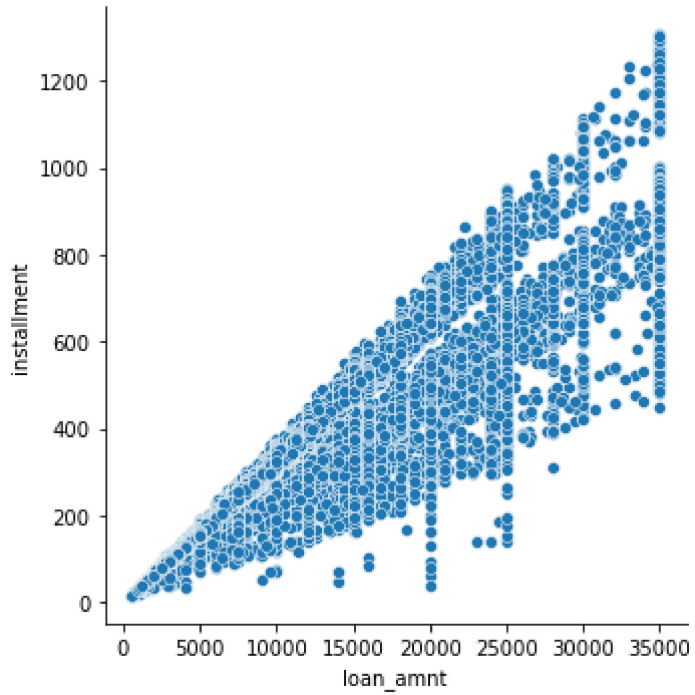
In [63]:
```
1 sns.relplot(x='loan_amnt',y='funded_amnt_inv',data=new_data)
```

Out[63]: <seaborn.axisgrid.FacetGrid at 0x1ff28ab5a30>



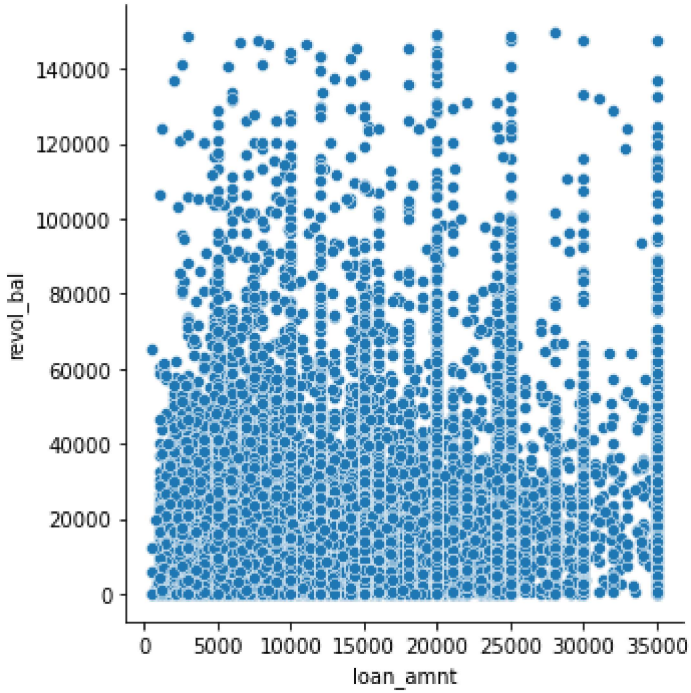
In [64]:
```
1 sns.relplot(x='loan_amnt',y='installment',data=new_data)
```

Out[64]: <seaborn.axisgrid.FacetGrid at 0x1ff36106e50>



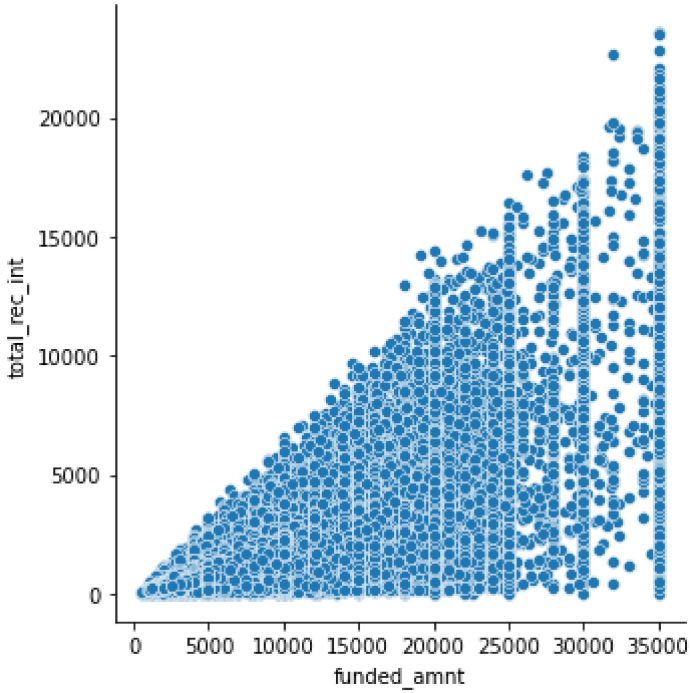
In [65]:
```
1 sns.relplot(x='loan_amnt',y='revol_bal',data=new_data)
```

Out[65]: <seaborn.axisgrid.FacetGrid at 0x1ff2983b190>



In [66]:
```
1 sns.relplot(x='funded_amnt',y='total_rec_int',data=new_data)
```

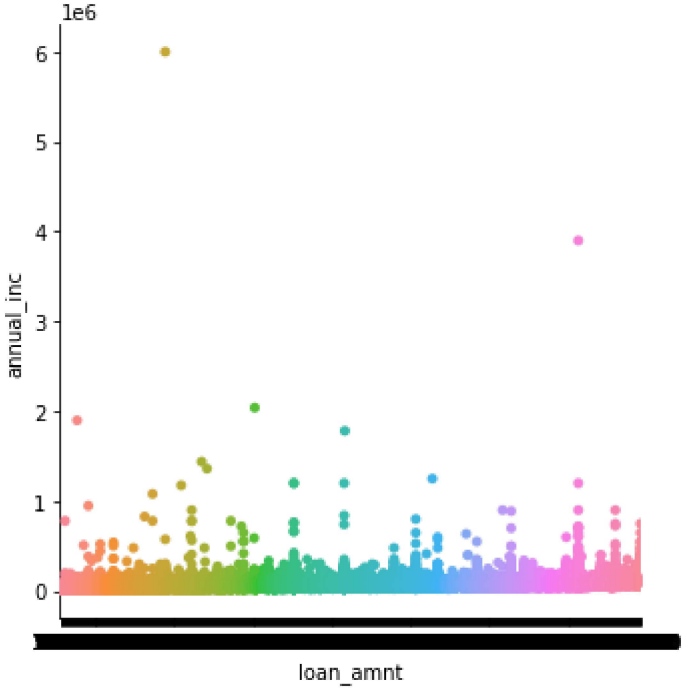
Out[66]: <seaborn.axisgrid.FacetGrid at 0x1ff35fea970>



Catplot

In [67]:
```
1 sns.catplot(x='loan_amnt',y='annual_inc',data=new_data,jitter=False)
```

Out[67]: <seaborn.axisgrid.FacetGrid at 0x1ff360f74c0>

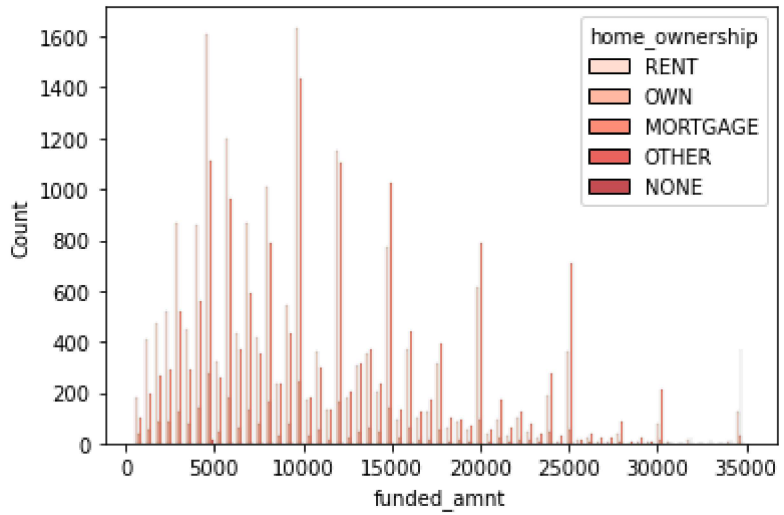


In [ ]:
```
1
```

Histplot

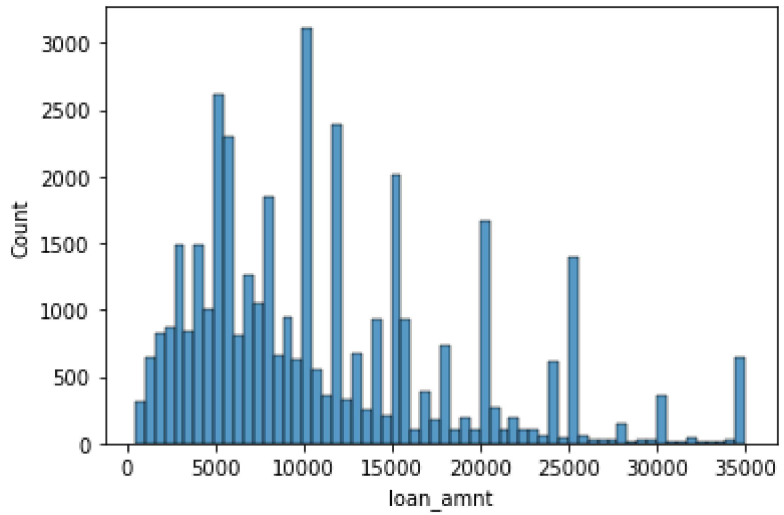
In [68]:
```
1 sns.histplot(data=new_data,x='funded_amnt',hue='home_ownership',multiple='dodge',shrink=.9,palette='Reds')
```

Out[68]: <AxesSubplot:xlabel='funded_amnt', ylabel='Count'>



In [69]:
```
1 sns.histplot(new_data['loan_amnt'])
```

Out[69]: <AxesSubplot:xlabel='loan_amnt', ylabel='Count'>

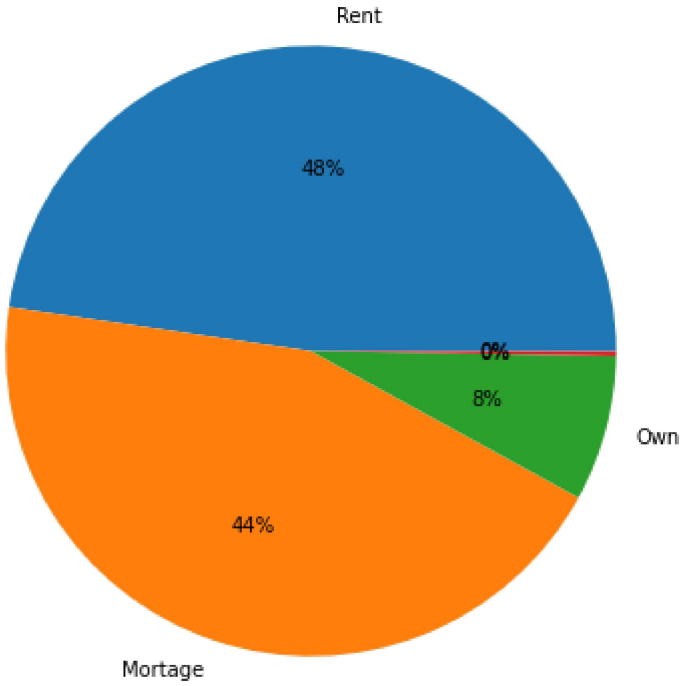


Pie Chart

In [ ]:
```
1
```

In [70]:

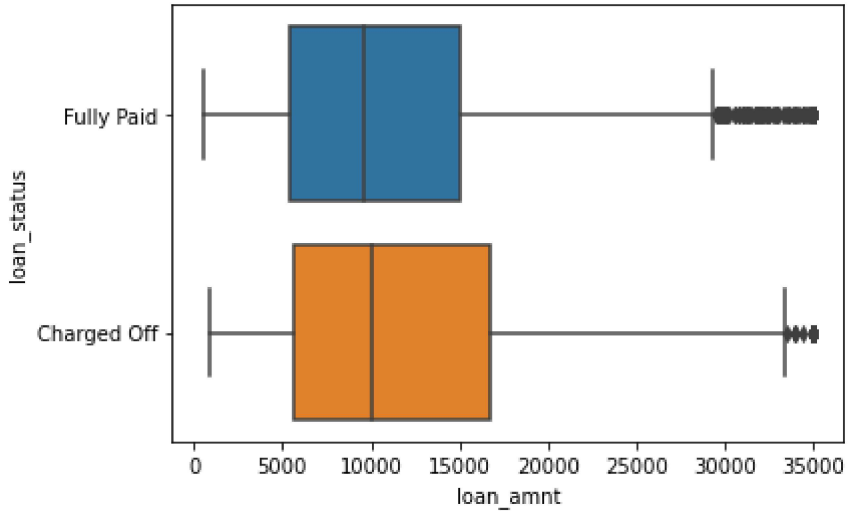
```
plt.figure(figsize=(12,7))
plt.pie(new_data['home_ownership'].value_counts(),
        autopct='%.0f%%',
        labels=['Rent','Mortage','Own','',''])
plt.show()
```



Box Plot

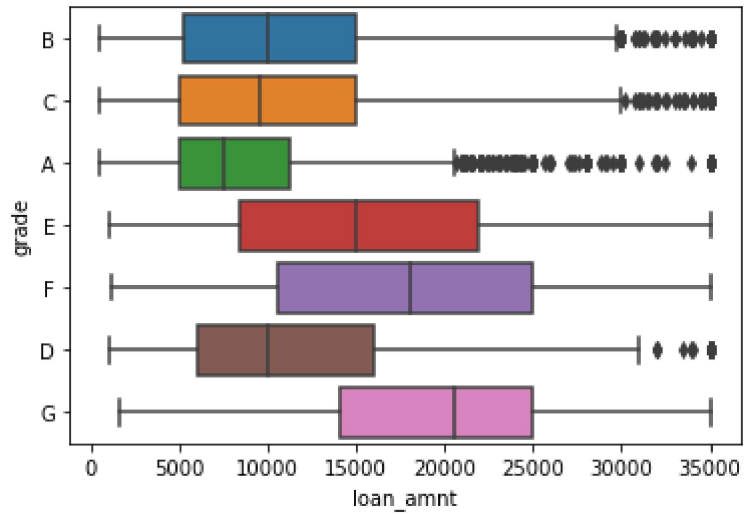
In [71]:
```
1 sns.boxplot(x='loan_amnt',y='loan_status',data=new_data)
```

Out[71]: <AxesSubplot:xlabel='loan_amnt', ylabel='loan_status'>



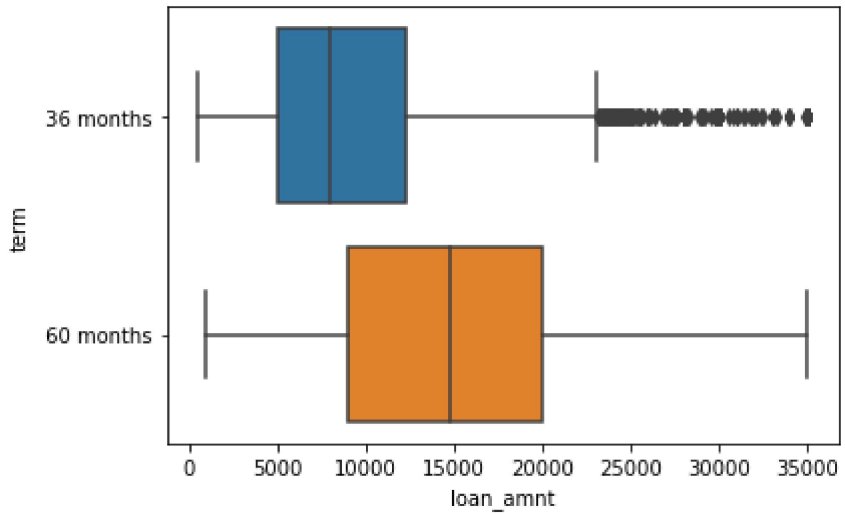
In [72]:
```
1 sns.boxplot(x='loan_amnt',y='grade',data=new_data)
```

Out[72]: <AxesSubplot:xlabel='loan_amnt', ylabel='grade'>



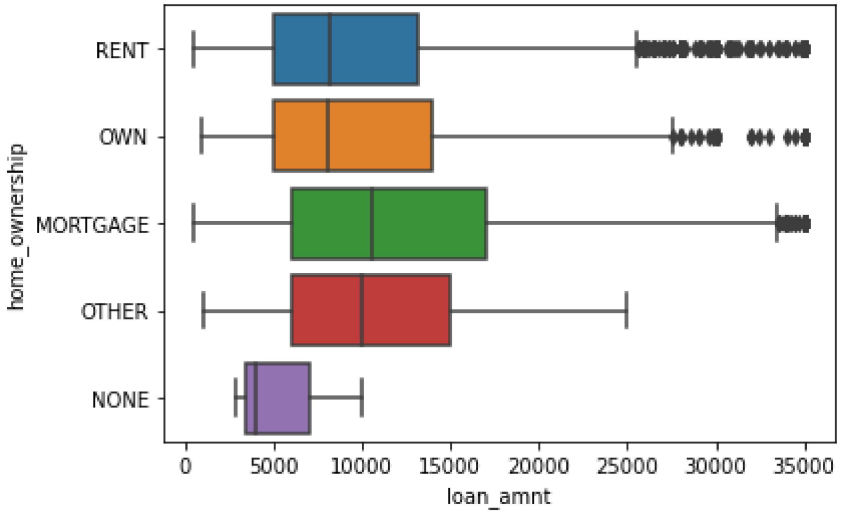
In [73]:
```
1 sns.boxplot(x='loan_amnt',y='term',data=new_data)
```

Out[73]: <AxesSubplot:xlabel='loan_amnt', ylabel='term'>



In [74]:
```
1 sns.boxplot(x='loan_amnt',y='home_ownership',data=new_data)
```

Out[74]: <AxesSubplot:xlabel='loan_amnt', ylabel='home_ownership'>



Machine Learning Algoritham

In [75]:
```
1 new_data['loan_status'].value_counts()
```

Out[75]:
```
Fully Paid     33586
Charged Off     5653
Name: loan_status, dtype: int64
```

In [76]:
```
1 new_data.sample()
```

Out[76]:

	id	member_id	loan_amnt	funded_amnt	funded_amnt_inv	term	Unnamed: 6	installment	grade	sub_grade	...	recoveries
26570	568246	731025	6000.0	6000.0	6000.0	36 months	14.35%	206.09	C	C4	...	0.0

1 rows × 52 columns

In [77]:
```
1 new_data['pymnt_plan'].value_counts()
```

Out[77]:
```
n    39238
y        1
Name: pymnt_plan, dtype: int64
```

In [78]:
```
1 new_data['application_type'].value_counts()
```

Out[78]:
```
INDIVIDUAL    39239
Name: application_type, dtype: int64
```

In [79]:
```
1 new_data['addr_state'].value_counts()
```

Out[79]:

CA	7029
NY	3759
FL	2834
TX	2698
NJ	1830
IL	1514
PA	1497
VA	1389
GA	1381
MA	1323
OH	1202
MD	1041
AZ	867
WA	831
CO	778
NC	773
CT	738
MI	718
MO	678
MN	611
NV	489
SC	469
WI	448
OR	442
AL	442
LA	432
KY	319
OK	295
KS	265
UT	255
AR	242
DC	213
RI	197
NM	187
WV	174
HI	171
NH	169
DE	114
MT	84
WY	83

```
AK      79
SD      62
VT      53
MS      19
TN      17
IN       9
ID       6
IA       5
NE       5
ME       3
Name: addr_state, dtype: int64
```

In [80]:
```
1 new_data['title'].value_counts()
```

Out[80]:
```
Debt Consolidation         2144
Debt Consolidation Loan    1671
Personal Loan               650
Consolidation               502
debt consolidation          495
                           ...
Excavator 4 Hire              1
Skidsteer                     1
black bike                    1
new Home                      1
JAL Loan                      1
Name: title, Length: 19502, dtype: int64
```

In [81]:
```
1 droped_cols = ['pymnt_plan','addr_state','application_type','title']
```

In [82]:
```
1 new_data.drop(columns=droped_cols, axis=0, inplace=True)
```

In [83]:
```
1 new_data.sample()
```

Out[83]:

	id	member_id	loan_amnt	funded_amnt	funded_amnt_inv	term	Unnamed: 6	installment	grade	sub_grade	...	total_rec_la
26200	573580	737776	4000.0	4000.0	4000.0	36 months	7.88%	125.13	A	A5	...	

1 rows × 48 columns

In [84]:
```
1 drop_cols = ['id',
2              'member_id',]
```

In [85]:
```
1 new_data.drop(columns=drop_cols, axis=1, inplace=True)
```

In [86]:
```
1 new_data.sample()
```

Out[86]:

	loan_amnt	funded_amnt	funded_amnt_inv	term	Unnamed: 6	installment	grade	sub_grade	emp_title	emp_length	...	total_re
37733	16000.0	16000.0	7041.138096	36 months	16.32%	565.02	E	E3	phoenixville hospital	4 years	...	

1 rows × 46 columns

In [87]:
```
1 drop_cols = ['funded_amnt',
2              'funded_amnt_inv',
3              'sub_grade',
4              'emp_title',
5              'issue_d',
6              'zip_code',
7             ]
```

In [88]:
```
1 new_data.drop(columns=drop_cols, axis=1, inplace=True)
```

In [89]:
```
drop_cols = ['out_prncp',
             'out_prncp_inv',
             'total_pymnt',
             'total_pymnt_inv',
             'total_rec_prncp',
             'total_rec_int',
             'total_rec_late_fee',
             'recoveries',
             'collection_recovery_fee',
             'last_pymnt_d',
             'last_pymnt_amnt',
            ]
```

In [90]:
```
new_data.drop(columns=drop_cols, axis=1, inplace=True)
```

In [91]:
```
new_data.sample()
```

Out[91]:

	loan_amnt	term	Unnamed: 6	installment	grade	emp_length	home_ownership	annual_inc	verification_status	loan_status	...	pu
4554	14000.0	36 months	10.65%	456.03	B	2 years	MORTGAGE	75000.0	Verified	Fully Paid	...	

1 rows × 29 columns

In [92]:
```
droped_cols = ['initial_list_status',
               'last_credit_pull_d',
               'policy_code',
               'pub_rec_bankruptcies',
              ]
```

In [93]:
```
1 new_data[droped_cols].describe()
```

Out[93]:

	policy_code
count	39239.0
mean	1.0
std	0.0
min	1.0
25%	1.0
50%	1.0
75%	1.0
max	1.0

In [94]:
```
1 new_data.drop(columns=droped_cols, axis=0, inplace=True)
```

In [95]:
```
1 new_data.sample()
```

Out[95]:

	loan_amnt	term	Unnamed: 6	installment	grade	emp_length	home_ownership	annual_inc	verification_status	loan_status	...	f
15808	12000.0	36 months	10.99%	392.81	B	10+ years	MORTGAGE	100000.0	Verified	Fully Paid	...	

1 rows × 25 columns

In [96]:
```
1 new_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 39239 entries, 0 to 39785
Data columns (total 25 columns):
 #   Column                Non-Null Count  Dtype  
---  ------                --------------  -----  
 0   loan_amnt             39239 non-null  float64
 1   term                  39239 non-null  object 
 2   Unnamed: 6            39239 non-null  object 
 3   installment           39239 non-null  float64
 4   grade                 39239 non-null  object 
 5   emp_length            39239 non-null  object 
 6   home_ownership        39239 non-null  object 
 7   annual_inc            39239 non-null  float64
 8   verification_status   39239 non-null  object 
 9   loan_status           39239 non-null  object 
 10  url                   39239 non-null  object 
 11  purpose               39239 non-null  object 
 12  dti                   39239 non-null  float64
 13  delinq_2yrs           39239 non-null  float64
 14  earliest_cr_line      39239 non-null  object 
 15  fico_range_low        39239 non-null  float64
 16  fico_range_high       39239 non-null  float64
 17  inq_last_6mths        39239 non-null  float64
 18  open_acc              39239 non-null  float64
 19  pub_rec               39239 non-null  float64
 20  revol_bal             39239 non-null  float64
 21  revol_util            39239 non-null  object 
 22  total_acc             39239 non-null  float64
 23  last_fico_range_high  39239 non-null  float64
 24  last_fico_range_low   39239 non-null  float64
dtypes: float64(14), object(11)
memory usage: 8.8+ MB
```

In [97]:
```
1 new_data['loan_status']
```

Out[97]:
```
0         Fully Paid
1        Charged Off
2         Fully Paid
3         Fully Paid
5         Fully Paid
            ...
39781     Fully Paid
39782     Fully Paid
39783     Fully Paid
39784     Fully Paid
39785     Fully Paid
Name: loan_status, Length: 39239, dtype: object
```

In [98]:
```
1 new_data['loan_status'].value_counts()
```

Out[98]:
```
Fully Paid     33586
Charged Off     5653
Name: loan_status, dtype: int64
```

In [99]:
```
1 new_data = new_data[(new_data['loan_status']=='Fully Paid') | (new_data['loan_status']=='Charged Off')]
```

In [100]:
```
1 new_data.reset_index(drop=True, inplace=True)
```

In [101]:
```
1 new_data.shape
```

Out[101]:
```
(39239, 25)
```

In [102]:
```
1 X_df = new_data.drop(columns=['loan_status'], axis=1)
2 y_df = new_data['loan_status']
```

In [103]:
```
1 X_df.shape, y_df.shape
```

Out[103]:
```
((39239, 24), (39239,))
```

In [104]:
```
cat_cols = []
for data in X_df:
    if X_df[data].dtype == 'object':
        cat_cols.append(data)
print(cat_cols)
```

```
['term', 'Unnamed: 6', 'grade', 'emp_length', 'home_ownership', 'verification_status', 'url', 'purpose', 'earl
iest_cr_line', 'revol_util']
```

In [105]:
```
X_df.drop(columns=cat_cols, inplace=True)
```

In [106]:
```
new_data.columns
```

Out[106]:
```
Index(['loan_amnt', 'term', 'Unnamed: 6', 'installment', 'grade', 'emp_length',
       'home_ownership', 'annual_inc', 'verification_status', 'loan_status',
       'url', 'purpose', 'dti', 'delinq_2yrs', 'earliest_cr_line',
       'fico_range_low', 'fico_range_high', 'inq_last_6mths', 'open_acc',
       'pub_rec', 'revol_bal', 'revol_util', 'total_acc',
       'last_fico_range_high', 'last_fico_range_low'],
      dtype='object')
```

In [107]:
```
x = new_data[['loan_amnt', 'term', 'installment', 'grade', 'home_ownership',
       'annual_inc', 'verification_status', 'loan_status', 'purpose', 'dti',
       'delinq_2yrs', 'fico_range_low', 'fico_range_high', 'inq_last_6mths',
        'open_acc', 'pub_rec', 'revol_bal',
       'total_acc', 'last_fico_range_high', 'last_fico_range_low',
        ]]
y = new_data['loan_status']
```

In [108]:
```
1 x
```

Out[108]:

	loan_amnt	term	installment	grade	home_ownership	annual_inc	verification_status	loan_status	purpose	dti	del
0	5000.0	36 months	162.87	B	RENT	24000.0	Verified	Fully Paid	credit_card	27.65	
1	2500.0	60 months	59.83	C	RENT	30000.0	Source Verified	Charged Off	car	1.00	
2	2400.0	36 months	84.33	C	RENT	12252.0	Not Verified	Fully Paid	small_business	8.72	
3	10000.0	36 months	339.31	C	RENT	49200.0	Source Verified	Fully Paid	other	20.00	
4	5000.0	36 months	156.46	A	RENT	36000.0	Source Verified	Fully Paid	wedding	11.20	
...	...	...	...	...	...	...	...	...	...	...	
39234	2500.0	36 months	78.42	A	MORTGAGE	110000.0	Not Verified	Fully Paid	home_improvement	11.33	
39235	8500.0	36 months	275.38	C	RENT	18000.0	Not Verified	Fully Paid	credit_card	6.40	
39236	5000.0	36 months	156.84	A	MORTGAGE	100000.0	Not Verified	Fully Paid	debt_consolidation	2.30	
39237	5000.0	36 months	155.38	A	MORTGAGE	200000.0	Not Verified	Fully Paid	other	3.72	
39238	7500.0	36 months	255.43	E	OWN	22000.0	Not Verified	Fully Paid	debt_consolidation	14.29	

39239 rows × 20 columns

In [109]:
```
1 y
```

Out[109]:
```
0         Fully Paid
1        Charged Off
2         Fully Paid
3         Fully Paid
4         Fully Paid
            ...
39234     Fully Paid
39235     Fully Paid
39236     Fully Paid
39237     Fully Paid
39238     Fully Paid
Name: loan_status, Length: 39239, dtype: object
```

Traning And Testing Data

In [110]:
```
1 from sklearn.model_selection import train_test_split
```

In [111]:
```
1 x_train,x_test,y_train,y_test = train_test_split(X_df, y_df, test_size=0.25, random_state=42)
```

In [112]:
```
1 x_train
```

Out[112]:

	loan_amnt	installment	annual_inc	dti	delinq_2yrs	fico_range_low	fico_range_high	inq_last_6mths	open_acc	pub_rec	revol
36549	19750.0	676.02	65000.0	18.85	0.0	700.0	704.0	1.0	8.0	0.0	116
2584	8900.0	229.12	77000.0	11.89	0.0	675.0	679.0	0.0	19.0	0.0	129
16823	20000.0	435.95	51000.0	26.26	0.0	735.0	739.0	0.0	11.0	0.0	130
26248	2700.0	83.08	66000.0	0.27	0.0	790.0	794.0	0.0	9.0	0.0	11
2425	11200.0	400.82	36840.0	14.07	0.0	695.0	699.0	0.0	2.0	0.0	76
...	...	...	...	...	...	...	...	...	...	...	
6265	29300.0	658.00	65000.0	2.05	1.0	765.0	769.0	0.0	12.0	0.0	18
11284	15000.0	329.82	85000.0	20.51	0.0	755.0	759.0	0.0	22.0	0.0	673
38158	7500.0	253.58	129996.0	16.78	0.0	660.0	664.0	3.0	12.0	0.0	899
860	7200.0	270.86	88800.0	20.46	0.0	670.0	674.0	3.0	34.0	0.0	179
15795	16000.0	383.00	28800.0	17.04	0.0	705.0	709.0	0.0	5.0	0.0	192

29429 rows × 14 columns

In [113]:
```
1 x_test
```

Out[113]:

	loan_amnt	installment	annual_inc	dti	delinq_2yrs	fico_range_low	fico_range_high	inq_last_6mths	open_acc	pub_rec	revol
27519	19750.0	321.34	78000.0	24.60	0.0	715.0	719.0	0.0	7.0	0.0	163
29108	25000.0	879.80	182004.0	14.03	0.0	680.0	684.0	0.0	11.0	0.0	20
34867	13000.0	437.24	43000.0	14.93	0.0	715.0	719.0	2.0	5.0	0.0	
24512	3000.0	93.34	75000.0	15.07	0.0	725.0	729.0	1.0	12.0	0.0	135
30287	15000.0	519.70	75000.0	8.52	0.0	675.0	679.0	1.0	21.0	1.0	99
...	...	...	...	...	...	...	...	...	...	...	
2616	14825.0	422.37	45600.0	23.68	0.0	660.0	664.0	1.0	25.0	0.0	150
4525	12000.0	368.45	158367.0	2.31	0.0	740.0	744.0	0.0	8.0	0.0	66
14009	10000.0	311.02	60000.0	24.46	0.0	750.0	754.0	1.0	11.0	0.0	27
27592	15000.0	364.38	50000.0	8.90	0.0	685.0	689.0	2.0	7.0	0.0	115
31026	25000.0	875.27	47000.0	10.85	0.0	695.0	699.0	0.0	6.0	0.0	139

9810 rows × 14 columns

In [114]:
```
1 y_train
```

Out[114]:
```
36549    Charged Off
2584     Charged Off
16823     Fully Paid
26248     Fully Paid
2425      Fully Paid
             ...
6265      Fully Paid
11284     Fully Paid
38158    Charged Off
860      Charged Off
15795     Fully Paid
Name: loan_status, Length: 29429, dtype: object
```

In [115]:
```
1 y_test
```

Out[115]:
```
27519     Fully Paid
29108     Fully Paid
34867     Fully Paid
24512     Fully Paid
30287    Charged Off
             ...
2616     Charged Off
4525      Fully Paid
14009     Fully Paid
27592     Fully Paid
31026     Fully Paid
Name: loan_status, Length: 9810, dtype: object
```

In [116]:
```
1 from sklearn.linear_model import LogisticRegression
2 from sklearn.tree import DecisionTreeClassifier
3 from sklearn.ensemble import RandomForestClassifier
4 from sklearn.svm import SVC
5 from sklearn import metrics
```

Logistic Regression

In [119]:
```
log_model = LogisticRegression()

log_model.fit(x_train, y_train)

log_pred = log_model.predict(x_test)
```

```
D:\anaconda3\lib\site-packages\sklearn\linear_model\_logistic.py:763: ConvergenceWarning: lbfgs failed to conv
erge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://scikit-learn.org/stable/modules/prepro
cessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression (https://scikit-learn.org/st
able/modules/linear_model.html#logistic-regression)
  n_iter_i = _check_optimize_result(
```

In [120]:
```
print(metrics.classification_report(y_test, log_pred))
```

```
              precision    recall  f1-score   support

 Charged Off       0.62      0.32      0.42      1424
  Fully Paid       0.89      0.97      0.93      8386

    accuracy                           0.87      9810
   macro avg       0.76      0.64      0.68      9810
weighted avg       0.85      0.87      0.86      9810
```

Accuracy Score Testing Data

In [122]:
```
print('Accuracy score is',log_model.score(x_test,y_test))
```

```
Accuracy score is 0.8728848114169215
```

In [ ]:
```

```

In [ ]:
```
1
```

In [ ]:
```
1
```