```
https://colab.research.google.com/drive/1k1rqQXAgj6jCwf_jnaM3RVTQyzrBvrAS?usp=sharing
# link for accesing entire note book in the attached pdf some pages are lost  so please view through link
```

```
import pandas as pd
```

```
data=pd.read_csv(r'/content/logistic_regression.csv')
```

```
data.columns
data['loan_status']
```

```
0           Fully Paid
1           Fully Paid
2           Fully Paid
3           Fully Paid
4          Charged Off
              ...
169673     Fully Paid
169674    Charged Off
169675     Fully Paid
169676    Charged Off
169677            NaN
Name: loan_status, Length: 169678, dtype: object
```

```
data
```

	loan_amnt	term	int_rate	installment	grade	sub_grade	emp_title	emp_length	home_ownership	annual_inc	...	open_acc	pub_rec	revol_bal	revol_util	total_acc	initial_list_status	application_type	mort_acc	pub_rec_bankruptcies	address
0	10000.0	36 months	11.44	329.48	B	B4	Marketing	10+ years	RENT	117000.0	...	16.0	0.0	36369.0	41.8	25.0	w	INDIVIDUAL	0.0	0.0	0174 Michelle Gateway\r\nMendozaberg, OK 22690
1	8000.0	36 months	11.99	265.68	B	B5	Credit analyst	4 years	MORTGAGE	65000.0	...	17.0	0.0	20131.0	53.3	27.0	f	INDIVIDUAL	3.0	0.0	1076 Carney Fort Apt. 347\r\nLoganmouth, SD 05113
2	15600.0	36 months	10.49	506.97	B	B3	Statistician	< 1 year	RENT	43057.0	...	13.0	0.0	11987.0	92.2	26.0	f	INDIVIDUAL	0.0	0.0	87025 Mark Dale Apt. 269\r\nNew Sabrina, WV 05113
3	7200.0	36 months	6.49	220.65	A	A2	Client Advocate	6 years	RENT	54000.0	...	6.0	0.0	5472.0	21.5	13.0	f	INDIVIDUAL	0.0	0.0	823 Reid Ford\r\nDelacruzside, MA 00813
4	24375.0	60 months	17.27	609.33	C	C5	Destiny Management Inc.	9 years	MORTGAGE	55000.0	...	13.0	0.0	24584.0	69.8	43.0	f	INDIVIDUAL	1.0	0.0	679 Luna Roads\r\nGreggshire, VA 11650
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
169673	24000.0	36 months	7.89	750.86	A	A5	Senior VP	1 year	MORTGAGE	170000.0	...	11.0	0.0	42261.0	67.2	31.0	w	INDIVIDUAL	4.0	0.0	834 Maldonado Forks\r\nShawnside, ND 30723
169674	12000.0	36 months	12.49	401.39	B	B5	Loan Agent	5 years	RENT	48000.0	...	10.0	0.0	7477.0	44.2	14.0	w	INDIVIDUAL	0.0	0.0	6150 Skinner Inlet Suite 792\r\nEmilyborough, ...
169675	11575.0	36 months	6.99	357.35	A	A3	Memorial Sloan Kettering Cancer Center	< 1 year	RENT	40000.0	...	13.0	0.0	9688.0	19.8	23.0	f	INDIVIDUAL	NaN	0.0	0763 Cindy Manors Apt. 125\r\nPort April, AR 2...
169676	12000.0	60 months	18.55	308.33	E	E2	Assistant AD	10+ years	RENT	97600.0	...	13.0	0.0	18302.0	63.1	32.0	f	INDIVIDUAL	3.0	0.0	0744 Robertson Grove Apt. 748\r\nEast Emilyhav...
169677	2000.0	36 months	21.49	75.86	E	E3	Food giant i	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

169678 rows × 27 columns

```
# data preprocessing
data.isnull().sum()
```

```
loan_amnt                   0
term                        0
int_rate                    0
installment                 0
grade                       0
sub_grade                   0
emp_title                9835
emp_length               7909
home_ownership              1
annual_inc                  1
verification_status         1
issue_d                     1
loan_status                 1
purpose                     1
title                     752
dti                         1
earliest_cr_line            1
open_acc                    1
pub_rec                     1
revol_bal                   1
revol_util                113
total_acc                   1
initial_list_status         1
application_type            1
mort_acc                16192
pub_rec_bankruptcies      233
address                     1
dtype: int64
```

Start coding or generate with AI.

```
# columns with   null values will be replaced with 0
data
```

	loan_amnt	term	int_rate	installment	grade	sub_grade	emp_title	emp_length	home_ownership	annual_inc	...	open_acc	pub_rec	revol_bal	revol_util	total_acc	initial_list_status	application_type	mort_acc	pub_rec_bankruptcies	address
0	10000.0	36 months	11.44	329.48	B	B4	Marketing	10+ years	RENT	117000.0	...	16.0	0.0	36369.0	41.8	25.0	w	INDIVIDUAL	0.0	0.0	0174 Michelle Gateway\r\nMendozaberg, OK 22690
1	8000.0	36 months	11.99	265.68	B	B5	Credit analyst	4 years	MORTGAGE	65000.0	...	17.0	0.0	20131.0	53.3	27.0	f	INDIVIDUAL	3.0	0.0	1076 Carney Fort Apt. 347\r\nLoganmouth, SD 05113
2	15600.0	36 months	10.49	506.97	B	B3	Statistician	< 1 year	RENT	43057.0	...	13.0	0.0	11987.0	92.2	26.0	f	INDIVIDUAL	0.0	0.0	87025 Mark Dale Apt. 269\r\nNew Sabrina, WV 05113
3	7200.0	36 months	6.49	220.65	A	A2	Client Advocate	6 years	RENT	54000.0	...	6.0	0.0	5472.0	21.5	13.0	f	INDIVIDUAL	0.0	0.0	823 Reid Ford\r\nDelacruzside, MA 00813
4	24375.0	60 months	17.27	609.33	C	C5	Destiny Management Inc.	9 years	MORTGAGE	55000.0	...	13.0	0.0	24584.0	69.8	43.0	f	INDIVIDUAL	1.0	0.0	679 Luna Roads\r\nGreggshire, VA 11650
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
169673	24000.0	36 months	7.89	750.86	A	A5	Senior VP	1 year	MORTGAGE	170000.0	...	11.0	0.0	42261.0	67.2	31.0	w	INDIVIDUAL	4.0	0.0	834 Maldonado Forks\r\nShawnside, ND 30723
169674	12000.0	36 months	12.49	401.39	B	B5	Loan Agent	5 years	RENT	48000.0	...	10.0	0.0	7477.0	44.2	14.0	w	INDIVIDUAL	0.0	0.0	6150 Skinner Inlet Suite 792\r\nEmilyborough, ...
169675	11575.0	36 months	6.99	357.35	A	A3	Memorial Sloan Kettering Cancer Center	< 1 year	RENT	40000.0	...	13.0	0.0	9688.0	19.8	23.0	f	INDIVIDUAL	NaN	0.0	0763 Cindy Manors Apt. 125\r\nPort April, AR 2...
169676	12000.0	60 months	18.55	308.33	E	E2	Assistant AD	10+ years	RENT	97600.0	...	13.0	0.0	18302.0	63.1	32.0	f	INDIVIDUAL	3.0	0.0	0744 Robertson Grove Apt. 748\r\nEast Emilyhav...
169677	2000.0	36 months	21.49	75.86	E	E3	Food giant i	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

169678 rows × 27 columns

```
data=data.fillna(0)
```

```
data.isnull().sum()
```

```
loan_amnt               0
term                    0
int_rate                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
purpose                 0
title                   0
dti                     0
earliest_cr_line        0
open_acc                0
pub_rec                 0
revol_bal               0
revol_util              0
total_acc               0
initial_list_status     0
application_type        0
mort_acc                0
pub_rec_bankruptcies    0
address                 0
dtype: int64
```

```
data.duplicated().any()
```

```
False
```

```
data
```

	loan_amnt	term	int_rate	installment	grade	sub_grade	emp_title	emp_length	home_ownership	annual_inc	...	open_acc	pub_rec	revol_bal	revol_util	total_acc	initial_list_status	application_type	mort_acc	pub_rec_bankruptcies	address
0	10000.0	36 months	11.44	329.48	B	B4	Marketing	10+ years	RENT	117000.0	...	16.0	0.0	36369.0	41.8	25.0	w	INDIVIDUAL	0.0	0.0	0174 Michelle Gateway\r\nMendozaberg, OK 22690
1	8000.0	36 months	11.99	265.68	B	B5	Credit analyst	4 years	MORTGAGE	65000.0	...	17.0	0.0	20131.0	53.3	27.0	f	INDIVIDUAL	3.0	0.0	1076 Carney Fort Apt. 347\r\nLoganmouth, SD 05113
2	15600.0	36 months	10.49	506.97	B	B3	Statistician	< 1 year	RENT	43057.0	...	13.0	0.0	11987.0	92.2	26.0	f	INDIVIDUAL	0.0	0.0	87025 Mark Dale Apt. 269\r\nNew Sabrina, WV 05113
3	7200.0	36 months	6.49	220.65	A	A2	Client Advocate	6 years	RENT	54000.0	...	6.0	0.0	5472.0	21.5	13.0	f	INDIVIDUAL	0.0	0.0	823 Reid Ford\r\nDelacruzside, MA 00813
4	24375.0	60 months	17.27	609.33	C	C5	Destiny Management Inc.	9 years	MORTGAGE	55000.0	...	13.0	0.0	24584.0	69.8	43.0	f	INDIVIDUAL	1.0	0.0	679 Luna Roads\r\nGreggshire, VA 11650
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
169673	24000.0	36 months	7.89	750.86	A	A5	Senior VP	1 year	MORTGAGE	170000.0	...	11.0	0.0	42261.0	67.2	31.0	w	INDIVIDUAL	4.0	0.0	834 Maldonado Forks\r\nShawnside, ND 30723
169674	12000.0	36 months	12.49	401.39	B	B5	Loan Agent	5 years	RENT	48000.0	...	10.0	0.0	7477.0	44.2	14.0	w	INDIVIDUAL	0.0	0.0	6150 Skinner Inlet Suite 792\r\nEmilyborough, ...
169675	11575.0	36 months	6.99	357.35	A	A3	Memorial Sloan Kettering Cancer Center	< 1 year	RENT	40000.0	...	13.0	0.0	9688.0	19.8	23.0	f	INDIVIDUAL	0.0	0.0	0763 Cindy Manors Apt. 125\r\nPort April, AR 2...
169676	12000.0	60 months	18.55	308.33	E	E2	Assistant AD	10+ years	RENT	97600.0	...	13.0	0.0	18302.0	63.1	32.0	f	INDIVIDUAL	3.0	0.0	0744 Robertson Grove Apt. 748\r\nEast Emilyhav...
169677	2000.0	36 months	21.49	75.86	E	E3	Food giant i	0	0	0.0	...	0.0	0.0	0.0	0.0	0.0	0	0	0.0	0.0	0

169678 rows × 27 columns

```
data
```

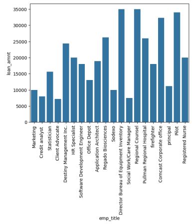
	loan_amnt	term	int_rate	installment	grade	sub_grade	emp_title	emp_length	home_ownership	annual_inc	...	open_acc	pub_rec	revol_bal	revol_util	total_acc	initial_list_status	application_type	mort_acc	pub_rec_bankruptcies	address
0	10000.0	36 months	11.44	329.48	B	B4	Marketing	10+ years	RENT	117000.0	...	16.0	0.0	36369.0	41.8	25.0	w	INDIVIDUAL	0.0	0.0	0174 Michelle Gateway\r\nMendozaberg, OK 22690
1	8000.0	36 months	11.99	265.68	B	B5	Credit analyst	4 years	MORTGAGE	65000.0	...	17.0	0.0	20131.0	53.3	27.0	f	INDIVIDUAL	3.0	0.0	1076 Carney Fort Apt. 347\r\nLoganmouth, SD 05113
2	15600.0	36 months	10.49	506.97	B	B3	Statistician	< 1 year	RENT	43057.0	...	13.0	0.0	11987.0	92.2	26.0	f	INDIVIDUAL	0.0	0.0	87025 Mark Dale Apt. 269\r\nNew Sabrina, WV 05113
3	7200.0	36 months	6.49	220.65	A	A2	Client Advocate	6 years	RENT	54000.0	...	6.0	0.0	5472.0	21.5	13.0	f	INDIVIDUAL	0.0	0.0	823 Reid Ford\r\nDelacruzside, MA 00813
4	24375.0	60 months	17.27	609.33	C	C5	Destiny Management Inc.	9 years	MORTGAGE	55000.0	...	13.0	0.0	24584.0	69.8	43.0	f	INDIVIDUAL	1.0	0.0	679 Luna Roads\r\nGreggshire, VA 11650
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
169673	24000.0	36 months	7.89	750.86	A	A5	Senior VP	1 year	MORTGAGE	170000.0	...	11.0	0.0	42261.0	67.2	31.0	w	INDIVIDUAL	4.0	0.0	834 Maldonado Forks\r\nShawnside, ND 30723
169674	12000.0	36 months	12.49	401.39	B	B5	Loan Agent	5 years	RENT	48000.0	...	10.0	0.0	7477.0	44.2	14.0	w	INDIVIDUAL	0.0	0.0	6150 Skinner Inlet Suite 792\r\nEmilyborough, ...
169675	11575.0	36 months	6.99	357.35	A	A3	Memorial Sloan Kettering Cancer Center	< 1 year	RENT	40000.0	...	13.0	0.0	9688.0	19.8	23.0	f	INDIVIDUAL	0.0	0.0	0763 Cindy Manors Apt. 125\r\nPort April, AR 2...
169676	12000.0	60 months	18.55	308.33	E	E2	Assistant AD	10+ years	RENT	97600.0	...	13.0	0.0	18302.0	63.1	32.0	f	INDIVIDUAL	3.0	0.0	0744 Robertson Grove Apt. 748\r\nEast Emilyhav...
169677	2000.0	36 months	21.49	75.86	E	E3	Food giant i	0	0	0.0	...	0.0	0.0	0.0	0.0	0.0	0	0	0.0	0.0	0

169678 rows × 27 columns

```
# data visualization
import seaborn as sns
p=data.head(20)
```

```
import matplotlib.pyplot as plt
sns.barplot(x='emp_title',y='loan_amnt',data=p)
plt.xticks(rotation=90)
```

```
([0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19],
 [Text(0, 0, 'Marketing'),
  Text(1, 0, 'Credit analyst '),
  Text(2, 0, 'Statistician'),
  Text(3, 0, 'Client Advocate'),
  Text(4, 0, 'Destiny Management Inc.'),
  Text(5, 0, 'HR Specialist'),
  Text(6, 0, 'Software Development Engineer'),
  Text(7, 0, 'Office Depot'),
  Text(8, 0, 'Application Architect'),
  Text(9, 0, 'Regado Biosciences'),
  Text(10, 0, 'Sodexo'),
  Text(11, 0, 'Director Bureau of Equipment Inventory'),
  Text(12, 0, 'Social Work/Care Manager'),
  Text(13, 0, 'Regional Counsel'),
  Text(14, 0, 'Pullman Regional Hospital'),
  Text(15, 0, 'firefighter'),
  Text(16, 0, 'Comcast Corporate office'),
  Text(17, 0, 'principal'),
  Text(18, 0, 'Pilot'),
  Text(19, 0, 'Registered Nurse')])
```



New section

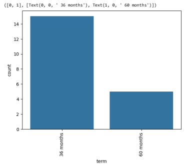
```
data.head(5)
```

	loan_amnt	term	int_rate	installment	grade	sub_grade	emp_title	emp_length	home_ownership	annual_inc	...	open_acc	pub_rec	revol_bal	revol_util	total_acc	initial_list_status	application_type	mort_acc	pub_rec_bankruptcies	address
0	10000.0	36 months	11.44	329.48	B	B4	Marketing	10+ years	RENT	117000.0	...	16.0	0.0	36369.0	41.8	25.0	w	INDIVIDUAL	0.0	0.0	0174 Michelle Gateway\r\nMendozaberg, OK 22690
1	8000.0	36 months	11.99	265.68	B	B5	Credit analyst	4 years	MORTGAGE	65000.0	...	17.0	0.0	20131.0	53.3	27.0	f	INDIVIDUAL	3.0	0.0	1076 Carney Fort Apt. 347\r\nLoganmouth, SD 05113
2	15600.0	36 months	10.49	506.97	B	B3	Statistician	< 1 year	RENT	43057.0	...	13.0	0.0	11987.0	92.2	26.0	f	INDIVIDUAL	0.0	0.0	87025 Mark Dale Apt. 269\r\nNew Sabrina, WV 05113
3	7200.0	36 months	6.49	220.65	A	A2	Client Advocate	6 years	RENT	54000.0	...	6.0	0.0	5472.0	21.5	13.0	f	INDIVIDUAL	0.0	0.0	823 Reid Ford\r\nDelacruzside, MA 00813
4	24375.0	60 months	17.27	609.33	C	C5	Destiny Management Inc.	9 years	MORTGAGE	55000.0	...	13.0	0.0	24584.0	69.8	43.0	f	INDIVIDUAL	1.0	0.0	679 Luna Roads\r\nGreggshire, VA 11650

5 rows × 27 columns

```
sns.countplot(x='term',data=p)
plt.xticks(rotation=90)
# in this visualizing how many people opted 36 months plan and 60 months plan
# here more number of people are opted for 36 months plan
```

```
([0, 1], [Text(0, 0, ' 36 months'), Text(1, 0, ' 60 months')])
```



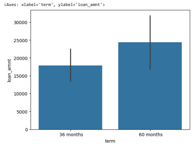

```
data.head(3)
```

	loan_amnt	term	int_rate	installment	grade	sub_grade	emp_title	emp_length	home_ownership	annual_inc	...	open_acc	pub_rec	revol_bal	revol_util	total_acc	initial_list_status	application_type	mort_acc	pub_rec_bankruptcies	address
0	10000.0	36 months	11.44	329.48	B	B4	Marketing	10+ years	RENT	117000.0	...	16.0	0.0	36369.0	41.8	25.0	w	INDIVIDUAL	0.0	0.0	0174 Michelle Gateway\r\nMendozaberg, OK 22690
1	8000.0	36 months	11.99	265.68	B	B5	Credit analyst	4 years	MORTGAGE	65000.0	...	17.0	0.0	20131.0	53.3	27.0	f	INDIVIDUAL	3.0	0.0	1076 Carney Fort Apt. 347\r\nLoganmouth, SD 05113
2	15600.0	36 months	10.49	506.97	B	B3	Statistician	< 1 year	RENT	43057.0	...	13.0	0.0	11987.0	92.2	26.0	f	INDIVIDUAL	0.0	0.0	87025 Mark Dale Apt. 269\r\nNew Sabrina, WV 05113

3 rows × 27 columns

```
sns.barplot(x='term',y='loan_amnt',data=p)
# for 36 months the highest loan_amnt 18000
# for 60 months the highest loan_amnt is 24000
# more months for more loan amount in order to clear thats why term will be more for that
```

```
<Axes: xlabel='term', ylabel='loan_amnt'>
```



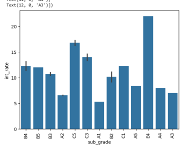

```
data
```

	loan_amnt	term	int_rate	installment	grade	sub_grade	emp_title	emp_length	home_ownership	annual_inc	...	open_acc	pub_rec	revol_bal	revol_util	total_acc	initial_list_status	application_type	mort_acc	pub_rec_bankruptcies	address
0	10000.0	36 months	11.44	329.48	B	B4	Marketing	10+ years	RENT	117000.0	...	16.0	0.0	36369.0	41.8	25.0	w	INDIVIDUAL	0.0	0.0	0174 Michelle Gateway\r\nMendozaberg, OK 22690
1	8000.0	36 months	11.99	265.68	B	B5	Credit analyst	4 years	MORTGAGE	65000.0	...	17.0	0.0	20131.0	53.3	27.0	f	INDIVIDUAL	3.0	0.0	1076 Carney Fort Apt. 347\r\nLoganmouth, SD 05113
2	15600.0	36 months	10.49	506.97	B	B3	Statistician	< 1 year	RENT	43057.0	...	13.0	0.0	11987.0	92.2	26.0	f	INDIVIDUAL	0.0	0.0	87025 Mark Dale Apt. 269\r\nNew Sabrina, WV 05113
3	7200.0	36 months	6.49	220.65	A	A2	Client Advocate	6 years	RENT	54000.0	...	6.0	0.0	5472.0	21.5	13.0	f	INDIVIDUAL	0.0	0.0	823 Reid Ford\r\nDelacruzside, MA 00813
4	24375.0	60 months	17.27	609.33	C	C5	Destiny Management Inc.	9 years	MORTGAGE	55000.0	...	13.0	0.0	24584.0	69.8	43.0	f	INDIVIDUAL	1.0	0.0	679 Luna Roads\r\nGreggshire, VA 11650
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
169673	24000.0	36 months	7.89	750.86	A	A5	Senior VP	1 year	MORTGAGE	170000.0	...	11.0	0.0	42261.0	67.2	31.0	w	INDIVIDUAL	4.0	0.0	834 Maldonado Forks\r\nShawnside, ND 30723
169674	12000.0	36 months	12.49	401.39	B	B5	Loan Agent	5 years	RENT	48000.0	...	10.0	0.0	7477.0	44.2	14.0	w	INDIVIDUAL	0.0	0.0	6150 Skinner Inlet Suite 792\r\nEmilyborough, ...
169675	11575.0	36 months	6.99	357.35	A	A3	Memorial Sloan Kettering Cancer Center	< 1 year	RENT	40000.0	...	13.0	0.0	9688.0	19.8	23.0	f	INDIVIDUAL	0.0	0.0	0763 Cindy Manors Apt. 125\r\nPort April, AR 2...
169676	12000.0	60 months	18.55	308.33	E	E2	Assistant AD	10+ years	RENT	97600.0	...	13.0	0.0	18302.0	63.1	32.0	f	INDIVIDUAL	3.0	0.0	0744 Robertson Grove Apt. 748\r\nEast Emilyhav...
169677	2000.0	36 months	21.49	75.86	E	E3	Food giant i	0	0	0.0	...	0.0	0.0	0.0	0.0	0.0	0	0	0.0	0.0	0

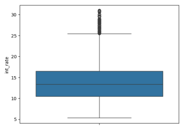
169678 rows × 27 columns

```
sns.barplot(x='sub_grade',y='int_rate',data=p)
plt.xticks(rotation=90)
# here subgrade implies based on credit score,history,
#employment status of a person,annual income the borrowers are classified in to subgrades
# A1,A2 are safest or risk free borrowers  e1,e2 are not safest  they are risky borrowers
# for banks also providing more intrest rate for risky borrowers
# less intrest rate for non risky borrowers
```

```
([0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12],
 [Text(0, 0, 'B4'),
  Text(1, 0, 'B5'),
  Text(2, 0, 'B3'),
  Text(3, 0, 'A2'),
  Text(4, 0, 'C5'),
  Text(5, 0, 'C3'),
  Text(6, 0, 'A1'),
  Text(7, 0, 'B2'),
  Text(8, 0, 'C1'),
  Text(9, 0, 'A5'),
  Text(10, 0, 'E4'),
  Text(11, 0, 'A4'),
  Text(12, 0, 'A3')])
```




```
# checking for outliers
data.head(5)
iqr=data['int_rate'].quantile(0.75)-data['int_rate'].quantile(0.25)
iqr
sns.boxplot(data['int_rate'])
iqr
upper_limit=data['int_rate'].quantile(0.75)+1.5*(iqr)
lower_limit=data['int_rate'].quantile(0.25)-1.5*(iqr)
```




```
outliers=data[data['int_rate']>upper_limit]
```

```
outliers
```

	loan_amnt	term	int_rate	installment	grade	sub_grade	emp_title	emp_length	home_ownership	annual_inc	...	open_acc	pub_rec	revol_bal	revol_util	total_acc	initial_list_status	application_type	mort_acc	pub_rec_bankruptcies	address
96	12625.0	60 months	25.78	376.36	F	F5	Nursing Assistant	7 years	MORTGAGE	43000.0	...	18.0	0.0	19993.0	55.7	31.0	w	INDIVIDUAL	1.0	0.0	522 Melissa Village\r\nEast Janice, MT 00813
97	13400.0	60 months	25.83	399.86	G	G2	DESIGNER	10+ years	MORTGAGE	56000.0	...	23.0	0.0	28845.0	84.5	44.0	w	INDIVIDUAL	6.0	0.0	0326 Jacob Via\r\nJacksonstad, MI 93700
133	13075.0	60 months	27.31	401.68	G	G2	Clerk-specialist level	10+ years	MORTGAGE	54777.0	...	4.0	0.0	3845.0	59.2	27.0	f	INDIVIDUAL	2.0	0.0	1532 Gill Greens Apt. 236\r\nSouth Cassandrafu...
168	11800.0	60 months	28.99	374.49	G	G5	0	0	RENT	44011.0	...	9.0	1.0	6313.0	67.2	13.0	f	INDIVIDUAL	0.0	1.0	3019 Smith Manors Suite 402\r\nLake Anthonybur...
204	34350.0	60 months	28.99	1090.13	G	G5	Quality improvement data analyst	3 years	RENT	84000.0	...	11.0	1.0	8400.0	85.7	23.0	w	INDIVIDUAL	0.0	1.0	5050 Gomez Mill Suite 557\r\nEast Adamton, SD ...
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
168943	20350.0	60 months	25.83	607.25	G	G2	inventory superviser	10+ years	RENT	50000.0	...	2.0	0.0	722.0	90.2	6.0	w	INDIVIDUAL	0.0	0.0	18522 Newton Alley Apt. 113\r\nLopezside, VA 1...
169350	35000.0	60 months	27.31	1075.24	G	G2	Director of Global Social Marketing	2 years	RENT	166000.0	...	16.0	0.0	21203.0	61.5	41.0	f	INDIVIDUAL	0.0	0.0	95755 Jenkins Well\r\nEast Tyler, MI 11650
169442	14975.0	60 months	25.78	446.42	F	F5	dockmanager	10+ years	MORTGAGE	52000.0	...	8.0	0.0	2855.0	18.1	24.0	w	INDIVIDUAL	1.0	0.0	3588 Evan Drives Apt. 136\r\nDanahaven, NC 30723
169449	18000.0	60 months	26.77	547.17	G	G1	Senior Graphic Design Coordinator	10+ years	OWN	50000.0	...	14.0	0.0	16540.0	64.9	32.0	w	INDIVIDUAL	0.0	0.0	48201 Edward Mountains Apt. 492\r\nLake Gregor...
169637	6575.0	60 months	25.80	196.09	G	G1	School Nurse	8 years	RENT	81600.0	...	10.0	0.0	5933.0	79.1	22.0	w	INDIVIDUAL	0.0	0.0	92945 Gentry Isle\r\nNorth Kayla, AR 30723

1573 rows × 27 columns

```
data=data[data['int_rate']<upper_limit]
# here the values greater than upper limit are considered as an outliers less than that is normal data
# 742 rows are dropped because of outliers
```

Start coding or generate with AI.

```
data
```

	loan_amnt	term	int_rate	installment	grade	sub_grade	emp_title	emp_length	home_ownership	annual_inc	...	open_acc	pub_rec	revol_bal	revol_util	total_acc	initial_list_status	application_type	mort_acc	pub_rec_bankruptcies	address
0	10000.0	36 months	11.44	329.48	B	B4	Marketing	10+ years	RENT	117000.0	...	16.0	0.0	36369.0	41.8	25.0	w	INDIVIDUAL	0.0	0.0	0174 Michelle Gateway\r\nMendozaberg, OK 22690
1	8000.0	36 months	11.99	265.68	B	B5	Credit analyst	4 years	MORTGAGE	65000.0	...	17.0	0.0	20131.0	53.3	27.0	f	INDIVIDUAL	3.0	0.0	1076 Carney Fort Apt. 347\r\nLoganmouth, SD 05113
2	15600.0	36 months	10.49	506.97	B	B3	Statistician	< 1 year	RENT	43057.0	...	13.0	0.0	11987.0	92.2	26.0	f	INDIVIDUAL	0.0	0.0	87025 Mark Dale Apt. 269\r\nNew Sabrina, WV 05113
3	7200.0	36 months	6.49	220.65	A	A2	Client Advocate	6 years	RENT	54000.0	...	6.0	0.0	5472.0	21.5	13.0	f	INDIVIDUAL	0.0	0.0	823 Reid Ford\r\nDelacruzside, MA 00813
4	24375.0	60 months	17.27	609.33	C	C5	Destiny Management Inc.	9 years	MORTGAGE	55000.0	...	13.0	0.0	24584.0	69.8	43.0	f	INDIVIDUAL	1.0	0.0	679 Luna Roads\r\nGreggshire, VA 11650
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
169673	24000.0	36 months	7.89	750.86	A	A5	Senior VP	1 year	MORTGAGE	170000.0	...	11.0	0.0	42261.0	67.2	31.0	w	INDIVIDUAL	4.0	0.0	834 Maldonado Forks\r\nShawnside, ND 30723
169674	12000.0	36 months	12.49	401.39	B	B5	Loan Agent	5 years	RENT	48000.0	...	10.0	0.0	7477.0	44.2	14.0	w	INDIVIDUAL	0.0	0.0	6150 Skinner Inlet Suite 792\r\nEmilyborough, ...
169675	11575.0	36 months	6.99	357.35	A	A3	Memorial Sloan Kettering Cancer Center	< 1 year	RENT	40000.0	...	13.0	0.0	9688.0	19.8	23.0	f	INDIVIDUAL	0.0	0.0	0763 Cindy Manors Apt. 125\r\nPort April, AR 2...
169676	12000.0	60 months	18.55	308.33	E	E2	Assistant AD	10+ years	RENT	97600.0	...	13.0	0.0	18302.0	63.1	32.0	f	INDIVIDUAL	3.0	0.0	0744 Robertson Grove Apt. 748\r\nEast Emilyhav...
169677	2000.0	36 months	21.49	75.86	E	E3	Food giant i	0	0	0.0	...	0.0	0.0	0.0	0.0	0.0	0	0	0.0	0.0	0

168105 rows × 27 columns

```
data
```

	loan_amnt	term	int_rate	installment	grade	sub_grade	emp_title	emp_length	home_ownership	annual_inc	...	open_acc	pub_rec	revol_bal	revol_util	total_acc	initial_list_status	application_type	mort_acc	pub_rec_bankruptcies	address
0	10000.0	36 months	11.44	329.48	B	B4	Marketing	10+ years	RENT	117000.0	...	16.0	0.0	36369.0	41.8	25.0	w	INDIVIDUAL	0.0	0.0	0174 Michelle Gateway\r\nMendozaberg, OK 22690
1	8000.0	36 months	11.99	265.68	B	B5	Credit analyst	4 years	MORTGAGE	65000.0	...	17.0	0.0	20131.0	53.3	27.0	f	INDIVIDUAL	3.0	0.0	1076 Carney Fort Apt. 347\r\nLoganmouth, SD 05113
2	15600.0	36 months	10.49	506.97	B	B3	Statistician	< 1 year	RENT	43057.0	...	13.0	0.0	11987.0	92.2	26.0	f	INDIVIDUAL	0.0	0.0	87025 Mark Dale Apt. 269\r\nNew Sabrina, WV 05113
3	7200.0	36 months	6.49	220.65	A	A2	Client Advocate	6 years	RENT	54000.0	...	6.0	0.0	5472.0	21.5	13.0	f	INDIVIDUAL	0.0	0.0	823 Reid Ford\r\nDelacruzside, MA 00813
4	24375.0	60 months	17.27	609.33	C	C5	Destiny Management Inc.	9 years	MORTGAGE	55000.0	...	13.0	0.0	24584.0	69.8	43.0	f	INDIVIDUAL	1.0	0.0	679 Luna Roads\r\nGreggshire, VA 11650
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
169673	24000.0	36 months	7.89	750.86	A	A5	Senior VP	1 year	MORTGAGE	170000.0	...	11.0	0.0	42261.0	67.2	31.0	w	INDIVIDUAL	4.0	0.0	834 Maldonado Forks\r\nShawnside, ND 30723
169674	12000.0	36 months	12.49	401.39	B	B5	Loan Agent	5 years	RENT	48000.0	...	10.0	0.0	7477.0	44.2	14.0	w	INDIVIDUAL	0.0	0.0	6150 Skinner Inlet Suite 792\r\nEmilyborough, ...
169675	11575.0	36 months	6.99	357.35	A	A3	Memorial Sloan Kettering Cancer Center	< 1 year	RENT	40000.0	...	13.0	0.0	9688.0	19.8	23.0	f	INDIVIDUAL	0.0	0.0	0763 Cindy Manors Apt. 125\r\nPort April, AR 2...
169676	12000.0	60 months	18.55	308.33	E	E2	Assistant AD	10+ years	RENT	97600.0	...	13.0	0.0	18302.0	63.1	32.0	f	INDIVIDUAL	3.0	0.0	0744 Robertson Grove Apt. 748\r\nEast Emilyhav...
169677	2000.0	36 months	21.49	75.86	E	E3	Food giant i	0	0	0.0	...	0.0	0.0	0.0	0.0	0.0	0	0	0.0	0.0	0

168105 rows × 27 columns

```
# finding relationship between different features  how much correlation is there among different independent features
data=data[data['loan_status']!=0]
data=data[data['emp_title']!=0]
```

```
import re

def xtract_numeric(value):

  match = re.search(r'\d+', value)

  if match:

    return int(match.group())

  else:

    return None
```

Generate a slider using jupyter widgets Close

Generate is available for a limited time for unsubscribed users. Upgrade to Colab Pro

```
data['term']=data['term'].apply(xtract_numeric)
```

Start coding or generate with AI.

```
# label encoding is required because features contains text data that data we cant use it in the mathematical model
# thats why we are going to convert text data into numerical values
data=data.head(2000)
```

```
y=data['loan_status']
x=data.drop(columns=['loan_status'],axis=1)
```

```
x.columns
```

```
Index(['loan_amnt', 'term', 'int_rate', 'installment', 'grade', 'sub_grade',
       'emp_title', 'emp_length', 'home_ownership', 'annual_inc',
       'verification_status', 'issue_d', 'purpose', 'title', 'dti',
       'earliest_cr_line', 'open_acc', 'pub_rec', 'revol_bal', 'revol_util',
       'total_acc', 'initial_list_status', 'application_type', 'mort_acc',
       'pub_rec_bankruptcies', 'address'],
      dtype='object')
```

```
y=y.to_frame()
```

```
y=y['loan_status'].map({'Fully Paid':0,'Charged Off':1})
```

Start coding or generate with AI.

```
from sklearn.model_selection import train_test_split
```

```
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.3,random_state=78)
```

Start coding or generate with AI.

```
y_train
```

```
768     0
1723    1
117     0
1880    0
2038    1
       ..
114     0
43      0
2070    0
507     0
762     0
Name: loan_status, Length: 1400, dtype: int64
```

```
from sklearn.preprocessing import LabelEncoder
```

```
lr=LabelEncoder()
x_train['emp_title']
```

```
768      Pro are home health and private duty
1723                Empire Ozark Food Brokers
117                  UGHS AUTIMIS BILLING INC
1880                   Regional Sales Manager
2038                                  Teacher
                        ...
114                                 Principal
43                                  Vegas.com
2070                                    Sales
507                           wells fargo bank
762     Performance Consultant and Instructor
Name: emp_title, Length: 1400, dtype: object
```

```
x_train['emp_title'],x_train['home_ownership'],=lr.fit_transform(x_train['emp_title']),lr.fit_transform(x_train['emp_title'])
```

```
x_train['initial_list_status']=lr.fit_transform(x_train['initial_list_status'])
x_train
y_train
```

```
768     0
1723    1
117     0
1880    0
2038    1
       ..
114     0
43      0
2070    0
507     0
762     0
Name: loan_status, Length: 1400, dtype: int64
```

Start coding or generate with AI.

```
x_train['application_type']=lr.fit_transform(x_train['application_type'])
x_train
```

	loan_amnt	term	int_rate	installment	grade	sub_grade	emp_title	emp_length	home_ownership	annual_inc	...	open_acc	pub_rec	revol_bal	revol_util	total_acc	initial_list_status	application_type	mort_acc	pub_rec_bankruptcies	address
768	18000.0	36	13.11	607.45	B	B4	677	10+ years	677	55000.0	...	14.0	0.0	10586.0	92.1	34.0	0	0	2.0	0.0	4226 Robinson Drive\r\nPort Hailey, MS 29597
1723	20000.0	36	11.55	660.00	B	B3	311	10+ years	311	105000.0	...	15.0	1.0	27801.0	86.1	33.0	1	0	4.0	1.0	54780 Steven Crest Suite 481\r\nEast Tonyastad...
117	18525.0	36	15.88	650.19	C	C4	955	1 year	955	42000.0	...	7.0	0.0	6665.0	41.9	10.0	0	0	1.0	0.0	64445 Jessica Mountains Apt. 496\r\nLake Kimbe...
1880	19000.0	60	12.99	432.22	C	C2	735	6 years	735	75000.0	...	5.0	0.0	7788.0	50.2	15.0	1	0	2.0	0.0	01223 Ray Shoal Suite 209\r\nNorth John, WI 70466
2038	5600.0	36	13.33	189.58	C	C3	909	10+ years	909	50000.0	...	9.0	0.0	4477.0	60.5	21.0	0	0	1.0	0.0	1738 Maria Ridges\r\nPort Meganshire, ME 93700
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
114	8000.0	36	12.99	269.52	B	B4	674	10+ years	674	110000.0	...	18.0	0.0	52572.0	84.9	31.0	0	0	3.0	0.0	932 Reed Track Suite 467\r\nNorth Anna, CA 70466
43	8400.0	36	12.12	279.49	B	B3	994	10+ years	994	120000.0	...	11.0	0.0	28739.0	71.3	34.0	0	0	6.0	0.0	976 Smith Bridge\r\nWest Michael, NE 30723
2070	5000.0	36	15.61	174.83	C	C4	775	8 years	775	45000.0	...	8.0	0.0	17372.0	89.5	18.0	1	0	0.0	0.0	71992 Turner Orchard\r\nLake Joanne, OK 00813
507	18000.0	36	10.99	589.22	B	B2	1204	10+ years	1204	95000.0	...	10.0	0.0	18146.0	70.1	23.0	1	0	0.0	0.0	4430 Chavez Station Apt. 638\r\nJordanchester,...
762	24375.0	60	18.55	626.29	D	D2	650	10+ years	650	88500.0	...	18.0	0.0	4713.0	20.6	52.0	0	0	0.0	0.0	8677 Dawn Overpass\r\nJorgeton, ME 00813

1400 rows × 26 columns

```
x_train['grade']=lr.fit_transform(x_train['grade'])
```

```
x_train['sub_grade']=lr.fit_transform(x_train['sub_grade'])
```

```
x_train['sub_grade']=lr.fit_transform(x_train['sub_grade'])
```

```
x_train['address']=lr.fit_transform(x_train['address'])
```

```
x_train
```

	loan_amnt	term	int_rate	installment	grade	sub_grade	emp_title	emp_length	home_ownership	annual_inc	...	open_acc	pub_rec	revol_bal	revol_util	total_acc	initial_list_status	application_type	mort_acc	pub_rec_bankruptcies	address
768	18000.0	36	13.11	607.45	1	8	677	10+ years	677	55000.0	...	14.0	0.0	10598.0	92.1	34.0	0	0	2.0	0.0	538
1723	20000.0	36	11.55	660.00	1	7	311	10+ years	311	105000.0	...	15.0	1.0	27801.0	86.1	33.0	1	0	4.0	1.0	682
117	19325.0	36	15.88	682.19	2	13	955	1 year	955	63000.0	...	7.0	0.0	6465.0	41.9	10.0	0	0	1.0	0.0	802
1880	19000.0	60	12.99	432.22	2	11	735	6 years	735	79000.0	...	5.0	0.0	7788.0	92.2	15.0	1	0	2.0	0.0	14
2038	5600.0	36	13.33	189.58	2	12	909	10+ years	909	50000.0	...	9.0	0.0	6477.0	60.5	21.0	0	0	1.0	0.0	231
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
114	8000.0	36	12.99	269.32	1	8	674	10+ years	674	110000.0	...	18.0	0.0	52072.0	84.9	31.0	0	0	3.0	0.0	1171
43	8400.0	36	12.12	279.49	1	7	994	10+ years	994	120000.0	...	11.0	0.0	29739.0	71.3	34.0	0	0	6.0	0.0	1223
2070	5000.0	36	15.61	174.93	2	13	775	8 years	775	45000.0	...	8.0	0.0	17372.0	89.5	18.0	1	0	0.0	0.0	899
507	19000.0	36	10.99	589.22	1	6	1264	10+ years	1264	95000.0	...	15.0	0.0	19164.0	70.1	23.0	1	0	0.0	0.0	566
762	24375.0	60	18.55	624.39	3	16	650	10+ years	650	88500.0	...	18.0	0.0	4713.0	20.6	52.0	0	0	0.0	0.0	1085

1600 rows × 26 columns

```
x_test['grade']=lr.fit_transform(x_test['grade'])
x_test['sub_grade']=lr.fit_transform(x_test['sub_grade'])
```

```
#This is the code you are required to use
```

Start coding or generate with AI.

Start coding or generate with AI.

Generate print hello world using cuda

Generate is available for a limited time for unsubscribed users. Upgrade to Colab Pro

```
x_test['initial_list_status']=lr.fit_transform(x_test['initial_list_status'])
x_test['application_type']=lr.fit_transform(x_test['application_type'])
```

```
x_test['home_ownership']=lr.fit_transform(x_test['home_ownership'])
x_test['address']=lr.fit_transform(x_test['address'])
x_test
```

	loan_amnt	term	int_rate	installment	grade	sub_grade	emp_title	emp_length	home_ownership	annual_inc	...	open_acc	pub_rec	revol_bal	revol_util	total_acc	initial_list_status	application_type	mort_acc	pub_rec_bankruptcies	address
1108	10000.0	36	14.33	343.38	2	10	Hubbardton Forge	< 1 year	2	54000.0	...	12.0	0.0	27926.0	52.1	39.0	0	0	1.0	0.0	208
824	11500.0	36	12.12	382.63	1	7	Alto ISD	2 years	0	35000.0	...	12.0	0.0	14199.0	35.4	18.0	1	0	1.0	0.0	123
757	4525.0	36	12.12	150.56	1	7	NAF	5 years	0	67000.0	...	14.0	0.0	12936.0	52.6	19.0	0	0	1.0	0.0	549
93	5000.0	36	10.99	163.67	1	6	Electrician	7 years	2	52000.0	...	11.0	0.0	9655.0	27.3	19.0	0	0	0.0	0.0	511
476	2500.0	36	8.67	79.12	1	5	Front Desk Reception	10+ years	2	36000.0	...	11.0	0.0	4483.0	56.0	39.0	0	0	3.0	0.0	331
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1431	18000.0	36	11.47	595.03	1	8	Associated Director	1 year	0	115000.0	...	8.0	0.0	13916.0	70.5	18.0	0	0	3.0	0.0	434
895	12000.0	36	6.62	368.45	0	1	Store Manage	6 years	2	95000.0	...	15.0	0.0	17430.0	48.7	32.0	1	0	3.0	0.0	538
898	9000.0	36	6.49	275.76	0	1	cell technician	10	1	106568.0	...	4.0	0.0	4642.0	11.6	14.0	1	0	2.0	0.0	409
1893	25000.0	60	17.86	632.96	3	19	Spa Admin	10+ years	1	103000.0	...	12.0	0.0	27455.0	74.2	21.0	1	0	2.0	0.0	586
1235	17000.0	36	10.64	553.67	1	6	southern california gas company	10+ years	0	85280.0	...	8.0	0.0	25642.0	89.0	32.0	1	0	5.0	0.0	551

600 rows × 26 columns

```
def extract(numeric):
  return(''.join(filter(str.isdigit,str(numeric))))
x_test['emp_length']=x_test['emp_length'].apply(extract)
```

```
x_test
```

	loan_amnt	term	int_rate	installment	grade	sub_grade	emp_title	emp_length	home_ownership	annual_inc	...	open_acc	pub_rec	revol_bal	revol_util	total_acc	initial_list_status	application_type	mort_acc	pub_rec_bankruptcies	address
1108	10000.0	36	14.33	343.38	2	10	Hubbardton Forge	1	3	54000.0	...	12.0	0.0	27926.0	52.1	39.0	0	0	1.0	0.0	208
824	11500.0	36	12.12	382.63	1	7	Alto ISD	2	0	35000.0	...	12.0	0.0	14199.0	35.4	18.0	1	0	1.0	0.0	123
757	4525.0	36	12.12	150.56	1	7	NAF	5	0	67000.0	...	14.0	0.0	12936.0	52.6	19.0	0	0	1.0	0.0	549
93	5000.0	36	10.99	163.67	1	6	Electrician	7	2	52000.0	...	11.0	0.0	9655.0	27.3	19.0	0	0	0.0	0.0	511
476	2500.0	36	8.67	79.12	1	5	Front Desk Reception	10	2	36000.0	...	11.0	0.0	4483.0	56.0	39.0	0	0	3.0	0.0	331
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1431	18000.0	36	11.47	595.03	1	8	Associated Director	1	0	115000.0	...	8.0	0.0	13916.0	70.5	18.0	0	0	3.0	0.0	434
895	12000.0	36	6.62	368.45	0	1	Store Manage	6	2	95000.0	...	15.0	0.0	17430.0	48.7	32.0	1	0	3.0	0.0	538
898	9000.0	36	6.49	275.76	0	1	cell technician	10	1	106568.0	...	4.0	0.0	4642.0	11.6	14.0	1	0	2.0	0.0	409
1893	25000.0	60	17.86	632.96	3	19	Spa Admin	10	1	103000.0	...	12.0	0.0	27455.0	74.2	21.0	1	0	2.0	0.0	586
1235	17000.0	36	10.64	553.67	1	6	southern california gas company	10	0	85280.0	...	8.0	0.0	25642.0	89.0	32.0	1	0	5.0	0.0	551

600 rows × 26 columns

```
x_train=x_train.dropna()
x_train
```

	loan_amnt	term	int_rate	installment	grade	sub_grade	emp_title	emp_length	home_ownership	annual_inc	...	open_acc	pub_rec	revol_bal	revol_util	total_acc	initial_list_status	application_type	mort_acc	pub_rec_bankruptcies	address
768	18000.0	36	13.11	607.45	1	8	677	10+ years	677	55000.0	...	14.0	0.0	10598.0	92.1	34.0	0	0	2.0	0.0	538
1723	20000.0	36	11.55	660.00	1	7	311	10+ years	311	105000.0	...	15.0	1.0	27801.0	86.1	33.0	1	0	4.0	1.0	682
117	19325.0	36	15.88	682.19	2	13	955	1 year	955	63000.0	...	7.0	0.0	6465.0	41.9	10.0	0	0	1.0	0.0	802
1880	19000.0	60	12.99	432.22	2	11	735	6 years	735	79000.0	...	5.0	0.0	7788.0	92.2	15.0	1	0	2.0	0.0	14
2038	5600.0	36	13.33	189.58	2	12	909	10+ years	909	50000.0	...	9.0	0.0	6477.0	60.5	21.0	0	0	1.0	0.0	231
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
114	8000.0	36	12.99	269.32	1	8	674	10+ years	674	110000.0	...	18.0	0.0	52072.0	84.9	31.0	0	0	3.0	0.0	1171
43	8400.0	36	12.12	279.49	1	7	994	10+ years	994	120000.0	...	11.0	0.0	29739.0	71.3	34.0	0	0	6.0	0.0	1223
2070	5000.0	36	15.61	174.93	2	13	775	8 years	775	45000.0	...	8.0	0.0	17372.0	89.5	18.0	1	0	0.0	0.0	899
507	19000.0	36	10.99	589.22	1	6	1264	10+ years	1264	95000.0	...	15.0	0.0	19164.0	70.1	23.0	1	0	0.0	0.0	566
762	24375.0	60	18.55	624.39	3	16	650	10+ years	650	88500.0	...	18.0	0.0	4713.0	20.6	52.0	0	0	0.0	0.0	1085

1600 rows × 26 columns

Start coding or generate with AI.

Generate print hello world using cuda

Generate is available for a limited time for unsubscribed users. Upgrade to Colab Pro

```
# we are going to create the model of logistic regression
from sklearn.preprocessing  import StandardScaler


ss=StandardScaler()
x_train['issue_d']=pd.to_numeric(x_train['issue_d'],errors='coerce')
#38  verification_status  1216 non-null   object
 #11  issue_d              1243 non-null   object
 #12  purpose              1216 non-null   object
 #13  title
#emp_length          1214 non-null   object
# we are going to convert strings to numeric  then only we can do scaling
```

Generate randomly select 5 items from a list

Generate is available for a limited time for unsubscribed users. Upgrade to Colab Pro

```
x_train['verification_status']=pd.to_numeric(x_train['verification_status'],errors='coerce')
```

```
x_train['emp_length']=pd.to_numeric(x_train['emp_length'],errors='coerce')
```

```
x_train['earliest_cr_line']=pd.to_numeric(x_train['earliest_cr_line'],errors='coerce')
```

```
x_train['purpose']=pd.to_numeric(x_train['purpose'],errors='coerce')
```

```
x_train['title']=pd.to_numeric(x_train['title'],errors='coerce')
```

```
x_train.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 1600 entries, 768 to 762
Data columns (total 26 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   loan_amnt             1600 non-null   float64
 1   term                  1600 non-null   int64
 2   int_rate              1600 non-null   float64
 3   installment           1600 non-null   float64
 4   grade                 1600 non-null   int64
 5   sub_grade             1600 non-null   int64
 6   emp_title             1600 non-null   int64
 7   emp_length            1600 non-null   float64
 8   home_ownership        1600 non-null   int64
 9   annual_inc            1600 non-null   float64
 10  verification_status   0 non-null      float64
 11  issue_d               0 non-null      float64
 12  purpose               0 non-null      float64
 13  title                 0 non-null      float64
 14  dti                   1600 non-null   float64
 15  earliest_cr_line      0 non-null      float64
 16  open_acc              1600 non-null   float64
 17  pub_rec               1600 non-null   float64
 18  revol_bal             1600 non-null   float64
 19  revol_util            1600 non-null   float64
 20  total_acc             1600 non-null   float64
 21  initial_list_status   1600 non-null   int64
 22  application_type      1600 non-null   int64
 23  mort_acc              1600 non-null   float64
 24  pub_rec_bankruptcies  1600 non-null   float64
 25  address               1600 non-null   int64
dtypes: float64(18), int64(8)
memory usage: 395.3 KB
```

```
x_test.info()
x_test['emp_length']=pd.to_numeric(x_test['emp_length'])
x_test['verification_status']=pd.to_numeric(x_test['verification_status'],errors='coerce')
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 600 entries, 1108 to 1335
Data columns (total 26 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   loan_amnt             600 non-null    float64
 1   term                  600 non-null    int64
 2   int_rate              600 non-null    float64
 3   installment           600 non-null    float64
 4   grade                 600 non-null    int64
 5   sub_grade             600 non-null    int64
 6   emp_title             600 non-null    object
 7   emp_length            600 non-null    object
 8   home_ownership        600 non-null    int64
 9   annual_inc            600 non-null    float64
 10  verification_status   600 non-null    object
 11  issue_d               600 non-null    object
 12  purpose               600 non-null    object
 13  title                 600 non-null    object
 14  dti                   600 non-null    float64
 15  earliest_cr_line      600 non-null    object
 16  open_acc              600 non-null    float64
 17  pub_rec               600 non-null    float64
 18  revol_bal             600 non-null    float64
 19  revol_util            600 non-null    float64
 20  total_acc             600 non-null    float64
 21  initial_list_status   600 non-null    int64
 22  application_type      600 non-null    int64
 23  mort_acc              600 non-null    float64
 24  pub_rec_bankruptcies  600 non-null    float64
 25  address               600 non-null    int64
dtypes: float64(12), int64(7), object(7)
memory usage: 126.6+ KB
```

```
x_test['earliest_cr_line']=pd.to_numeric(x_test['earliest_cr_line'],errors='coerce')
```

```
x_test['title']=pd.to_numeric(x_test['title'],errors='coerce')
```

```
x_test['issue_d']=pd.to_numeric(x_test['issue_d'],errors='coerce')
```

```
x_test['purpose']=pd.to_numeric(x_test['purpose'],errors='coerce')
```

```
x_test['emp_title']=pd.to_numeric(x_test['emp_title'],errors='coerce')
x_test.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 600 entries, 1108 to 1335
Data columns (total 26 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   loan_amnt             600 non-null    float64
 1   term                  600 non-null    int64
 2   int_rate              600 non-null    float64
 3   installment           600 non-null    float64
 4   grade                 600 non-null    int64
 5   sub_grade             600 non-null    int64
 6   emp_title             1 non-null      float64
 7   emp_length            600 non-null    int64
 8   home_ownership        600 non-null    int64
 9   annual_inc            600 non-null    float64
 10  verification_status   0 non-null      float64
 11  issue_d               0 non-null      float64
 12  purpose               0 non-null      float64
 13  title                 4 non-null      float64
 14  dti                   600 non-null    float64
 15  earliest_cr_line      0 non-null      float64
 16  open_acc              600 non-null    float64
 17  pub_rec               600 non-null    float64
 18  revol_bal             600 non-null    float64
 19  revol_util            600 non-null    float64
 20  total_acc             600 non-null    float64
 21  initial_list_status   600 non-null    int64
 22  application_type      600 non-null    int64
 23  mort_acc              600 non-null    float64
 24  pub_rec_bankruptcies  600 non-null    float64
 25  address               600 non-null    int64
dtypes: float64(18), int64(8)
memory usage: 126.6 KB
```

```
# all object data types are converted to integers and floats standard scaler will be applicable to numericals not strings
# we can do scaling
```

```
from sklearn.preprocessing import StandardScaler
```

```
st=StandardScaler()
x_train=pd.DataFrame(st.fit_transform(x_train),columns=x_train.columns)
```

```
/usr/local/lib/python3.10/dist-packages/sklearn/utils/extmath.py:1047: RuntimeWarning: invalid value encountered in divide
  updated_mean = (last_sum + new_sum) / updated_sample_count
/usr/local/lib/python3.10/dist-packages/sklearn/utils/extmath.py:1052: RuntimeWarning: invalid value encountered in divide
  T = new_sum / new_sample_count
/usr/local/lib/python3.10/dist-packages/sklearn/utils/extmath.py:1072: RuntimeWarning: invalid value encountered in divide
  new_unnormalized_variance -= correction**2 / new_sample_count
```

```
x_test=pd.DataFrame(st.fit_transform(x_test),columns=x_test.columns)
```

```
/usr/local/lib/python3.10/dist-packages/sklearn/utils/extmath.py:1047: RuntimeWarning: invalid value encountered in divide
  updated_mean = (last_sum + new_sum) / updated_sample_count
/usr/local/lib/python3.10/dist-packages/sklearn/utils/extmath.py:1052: RuntimeWarning: invalid value encountered in divide
  T = new_sum / new_sample_count
/usr/local/lib/python3.10/dist-packages/sklearn/utils/extmath.py:1072: RuntimeWarning: invalid value encountered in divide
  new_unnormalized_variance -= correction**2 / new_sample_count
```

```
x_train
# scaled values are placed in data frames  for x_train,y_train
```

	loan_amnt	term	int_rate	installment	grade	sub_grade	emp_title	emp_length	home_ownership	annual_inc	...	open_acc	pub_rec	revol_bal	revol_util	total_acc	initial_list_status	application_type	mort_acc	pub_rec_bankruptcies	address
0	0.410239	-0.550943	-0.087829	0.654267	-0.602081	-0.450243	0.190841	NaN	0.190841	-0.405366	...	0.538040	-0.351328	-0.262975	1.630198	0.752771	-0.860984	-0.026736	0.194244	-0.334096	-0.399609
1	0.647377	-0.550943	-0.452989	0.863940	-0.602081	-0.609863	-0.865803	NaN	-0.865803	0.610094	...	0.736526	1.815452	0.597452	1.378335	0.667643	1.161462	-0.026736	1.160288	2.535438	-0.018558
2	0.472488	-0.550943	0.560563	0.824798	0.183486	0.347858	0.993429	NaN	0.993429	-0.669385	...	-0.851365	-0.351328	-0.458951	-0.477059	-1.290290	-0.860984	-0.026736	-0.288778	-0.334096	0.253622
3	0.528808	1.815070	-0.115918	-0.044897	0.183486	0.028618	0.358288	NaN	0.358288	0.000818	...	-1.248338	-0.351328	-0.402822	-0.128648	-0.864653	1.161462	-0.026736	0.194244	-0.334096	-1.696173
4	-1.060019	-0.550943	-0.036332	-1.013026	0.183486	0.188238	0.860627	NaN	0.860627	-0.506912	...	-0.454392	-0.351328	-0.568310	0.303717	-0.353887	-0.860984	-0.026736	-0.288778	-0.334096	-1.159237
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1395	-0.775453	-0.550943	-0.115918	-0.694067	-0.602081	-0.450243	0.182180	NaN	0.182180	0.711640	...	1.331985	-0.351328	1.835538	1.327962	0.497388	-0.860984	-0.026736	0.677266	-0.334096	1.166660
1396	-0.728025	-0.550943	-0.319565	-0.654287	-0.602081	-0.609863	1.106022	NaN	1.106022	0.914732	...	-0.057419	-0.351328	0.644335	0.757072	0.752771	-0.860984	-0.026736	2.126332	-0.334096	1.295327
1397	-1.131160	-0.550943	0.497363	-1.071878	0.183486	0.347858	0.473768	NaN	0.473768	-0.608458	...	-0.652878	-0.351328	0.076198	1.521057	-0.609270	1.161462	-0.026736	-0.771800	-0.334096	0.493635
1398	0.410239	-0.550943	-0.584071	0.581529	-0.602081	-0.769484	1.712294	NaN	1.712294	0.407002	...	-0.255906	-0.351328	0.114884	0.706699	-0.183632	1.161462	-0.026736	-0.771800	-0.334096	-0.330327
1399	1.166117	1.815070	1.185548	0.729438	0.969052	0.826719	0.112892	NaN	0.112892	0.274993	...	1.331985	-0.351328	-0.556514	-1.371175	2.285066	-0.860984	-0.026736	-0.771800	-0.334096	0.953865

1400 rows × 26 columns

Start coding or generate with AI.

```
# we can import model and we can train models
# here iam facing null values
from sklearn.linear_model import LogisticRegression
x_train.isnull().sum()
x_train=x_train.fillna(0)
```

```
x_test=x_test.fillna(0)
y_train=y_train.fillna(0)
x_test=x_test.fillna(0)
y_test=y_test.fillna(0)
```

```
lo=LogisticRegression()
x_train
```

	loan_amnt	term	int_rate	installment	grade	sub_grade	emp_title	emp_length	home_ownership	annual_inc	...	open_acc	pub_rec	revol_bal	revol_util	total_acc	initial_list_status	application_type	mort_acc	pub_rec_bankruptcies	address
0	0.410239	-0.550943	-0.087829	0.654267	-0.602081	-0.450243	0.190841	0.0	0.190841	-0.405366	...	0.538040	-0.351328	-0.262975	1.630198	0.752771	-0.860984	-0.026736	0.194244	-0.334096	-0.399609
1	0.647377	-0.550943	-0.452989	0.863940	-0.602081	-0.609863	-0.865803	0.0	-0.865803	0.610094	...	0.736526	1.815452	0.597452	1.378335	0.667643	1.161462	-0.026736	1.160288	2.535438	-0.018558
2	0.472488	-0.550943	0.560563	0.824798	0.183486	0.347858	0.993429	0.0	0.993429	-0.669385	...	-0.851365	-0.351328	-0.458951	-0.477059	-1.290290	-0.860984	-0.026736	-0.288778	-0.334096	0.253622
3	0.528808	1.815070	-0.115918	-0.044897	0.183486	0.028618	0.358288	0.0	0.358288	0.000818	...	-1.248338	-0.351328	-0.402822	-0.128648	-0.864653	1.161462	-0.026736	0.194244	-0.334096	-1.696173
4	-1.060019	-0.550943	-0.036332	-1.013026	0.183486	0.188238	0.860627	0.0	0.860627	-0.506912	...	-0.454392	-0.351328	-0.568310	0.303717	-0.353887	-0.860984	-0.026736	-0.288778	-0.334096	-1.159237
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1395	-0.775453	-0.550943	-0.115918	-0.694067	-0.602081	-0.450243	0.182180	0.0	0.182180	0.711640	...	1.331985	-0.351328	1.835538	1.327962	0.497388	-0.860984	-0.026736	0.677266	-0.334096	1.166660
1396	-0.728025	-0.550943	-0.319565	-0.654287	-0.602081	-0.609863	1.106022	0.0	1.106022	0.914732	...	-0.057419	-0.351328	0.644335	0.757072	0.752771	-0.860984	-0.026736	2.126332	-0.334096	1.295327
1397	-1.131160	-0.550943	0.497363	-1.071878	0.183486	0.347858	0.473768	0.0	0.473768	-0.608458	...	-0.652878	-0.351328	0.076198	1.521057	-0.609270	1.161462	-0.026736	-0.771800	-0.334096	0.493635
1398	0.410239	-0.550943	-0.584071	0.581529	-0.602081	-0.769484	1.712294	0.0	1.712294	0.407002	...	-0.255906	-0.351328	0.114884	0.706699	-0.183632	1.161462	-0.026736	-0.771800	-0.334096	-0.330327
1399	1.166117	1.815070	1.185548	0.729438	0.969052	0.826719	0.112892	0.0	0.112892	0.274993	...	1.331985	-0.351328	-0.556514	-1.371175	2.285066	-0.860984	-0.026736	-0.771800	-0.334096	0.953865

1400 rows × 26 columns

```
log=lo.fit(x_train,y_train)
```

```
log.coef_
# these are weight values
```

```
array([[-0.09260207,  0.18365779,  0.07620565,  0.05045007,  0.2843696 ,
         0.21853674,  0.10791225,  0.        ,  0.10791225,  0.05871865,
         0.        ,  0.        ,  0.        ,  0.11498601,  0.24179935,
         0.        ,  0.16119405,  0.06687773, -0.0200392 ,  0.06267468,
        -0.40877191,  0.03757307, -0.10468335,  0.05925814,  0.01618686,
        -0.03399685]])
```

```
log.intercept_
```

```
array([-1.72145721])
```

```
y_predtest=log.predict(x_test)
# based on threshold values it will find out  firstly it will find the mathematical model based calculations and after that
# values will be transformed  to 0 or 1
y_predtest
```

```
array([0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1,
       0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0,
       1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
```

```
    0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0,
    0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
    0, 0, 0, 0, 0, 0])
```

```
from sklearn.metrics import accuracy_score,confusion_matrix,classification_report
y_test.value_counts() # for imbalanced dataset  we  should have to check precision and recall score for that
```

```
loan_status
0    485
1    115
Name: count, dtype: int64
```

```
# we are going to  find the model performance like how many are correct out of total predictions
accuracy_score(y_predtest,y_test)*100
# 81 percent records are correctly predicted  out of 100 records
```

```
81.0
```

```
confusion_matrix(y_predtest,y_test)
```

```
array([[475, 104],
       [ 10,  11]])
```

```
# here it is classified that 475 records are true negative values
# here it is classified that 10 records are false negative values
# 11 are true positive values
# 104 are  false positives
# type1 error is false positive
# type2 error is  false negative
```

```
from sklearn.metrics import  precision_score,recall_score
precision=precision_score(y_predtest,y_test)
precision
```

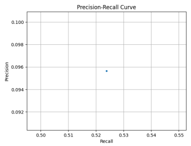
```
0.09565217391304348
```

```
recall=recall_score(y_predtest,y_test)
recall
```

```
0.5238095238095238
```

```
import matplotlib.pyplot as plt
```

```
plt.plot(recall, precision, marker='.')
plt.xlabel('Recall')
plt.ylabel('Precision')
plt.title('Precision-Recall Curve')
plt.grid(True)
plt.show()
```




```
classification_report(y_predtest,y_test)
```

```
'              precision    recall  f1-score   support\n\n           0       0.98      0.82      0.89       579\n           1       0.10      0.52      0.16        21\n\n    accuracy                           0.81       600\n   macro avg       0.54      0.67      0.53       600\nweighted avg       0.95      0.81      0.87       600\n'
```

```
# insights
# intrest_rate is also classified based on subgrade  for people who are low risk borrowers banks have provided low intrest rates
# people who are high risk borrowers are having high intrest rates
# compared to 36 months emi plan  the  maximum  loan amount  will be 15000
# for 60 months emi plan the maximum loan amount will  be 25000
# by using label encoder  we have converted  categorical values to numerical values of a particular features
# by using standard scaler we are scaling values  from 0 to 1 or -1 to +1
# because bigger values will play important role in finding target values mathematical function will give importance to higher values
# but it is not true the feature with related or play key role even though it is not a bigger value
# we are standardizing it  then weight values will plays a key role
```



```
# recommendations
# generalized model should be considered
# model should be simpler
# if model is complex there is a more chance for overfitting that takes place
# in train data set you should have to choose the model that is  accurate in 2nd or third position
# but it should not be  100 percent accurate if it is accurate then it is considered as overfitting model it will be bad
# it cannot perform well in test data set
# bias variance tradeoff in the sense in between model we should have to consider in performance wise
```