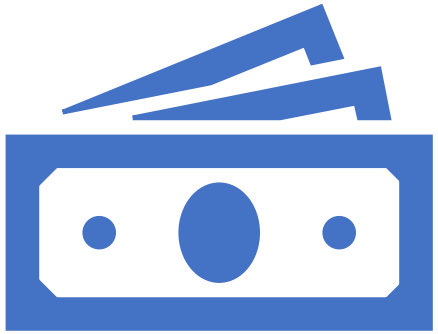


Lending Club Case Study

Problem Statement : Based on the data provided, identify risky loan applicants to reduce credit loss to the lending company.

Solution Approach

- Identify null values in the dataset
- Replace / Drop null values from the dataset
- Remove unwanted columns which are not relevant for the analysis
- Find Outliers

Identify Null Values in the dataset

```
] loanData.isna().sum()
```

```
id                              0
member_id                       0
loan_amnt                       0
funded_amnt                     0
funded_amnt_inv                 0
                               ...
tax_liens                      39
tot_hi_cred_lim             39717
total_bal_ex_mort           39717
total_bc_limit              39717
total_il_high_credit_limit  39717
Length: 111, dtype: int64
```

Identifying Null Values

Drop Null Values from the dataset

```
loanData.dropna(how='all', inplace=True)
loanData.dropna(axis = 1, how='all', inplace=True)
```

Drop Null Values

Drop Current Status records

Dropping records which have "loan_status" as current.

```
loanData = loanData.drop(loanData[loanData['loan_status'] == 'Current'].index)

len(loanData)
```

```
38577
```

Remove unwanted columns which are not relevant for analysis

```
loanData.describe()
```

	id	member_id	loan_amnt	funded_amnt	funded_amnt_inv	installment	annual_inc	dti	delinq_2yrs
count	3.857700e+04	3.857700e+04	38577.000000	38577.000000	38577.000000	38577.000000	3.857700e+04	38577.000000	38577.000000
mean	6.763787e+05	8.422843e+05	11047.025430	10784.058506	10222.481123	322.466318	6.877797e+04	13.272727	0.146668
std	2.092639e+05	2.644519e+05	7348.441646	7090.306027	7022.720644	208.639215	6.421868e+04	6.673044	0.492271
min	5.473400e+04	7.069900e+04	500.000000	500.000000	0.000000	15.690000	4.000000e+03	0.000000	0.000000
25%	5.120330e+05	6.611310e+05	5300.000000	5200.000000	5000.000000	165.740000	4.000000e+04	8.130000	0.000000
50%	6.564230e+05	8.392920e+05	9600.000000	9550.000000	8733.440000	277.860000	5.886800e+04	13.370000	0.000000
75%	8.291460e+05	1.037336e+06	15000.000000	15000.000000	14000.000000	425.550000	8.200000e+04	18.560000	0.000000
max	1.077501e+06	1.314167e+06	35000.000000	35000.000000	35000.000000	1305.190000	6.000000e+06	29.990000	11.000000

```
loanData.drop(['url', 'desc', 'pymnt_plan', 'zip_code','earliest_cr_line', 'mths_since_last_delinq', 'mths_since_last_record', 'ou
```

Remove unwanted columns not relevant for analysis

Remove unwanted columns which are not relevant for analysis

```
loanData.describe()
```

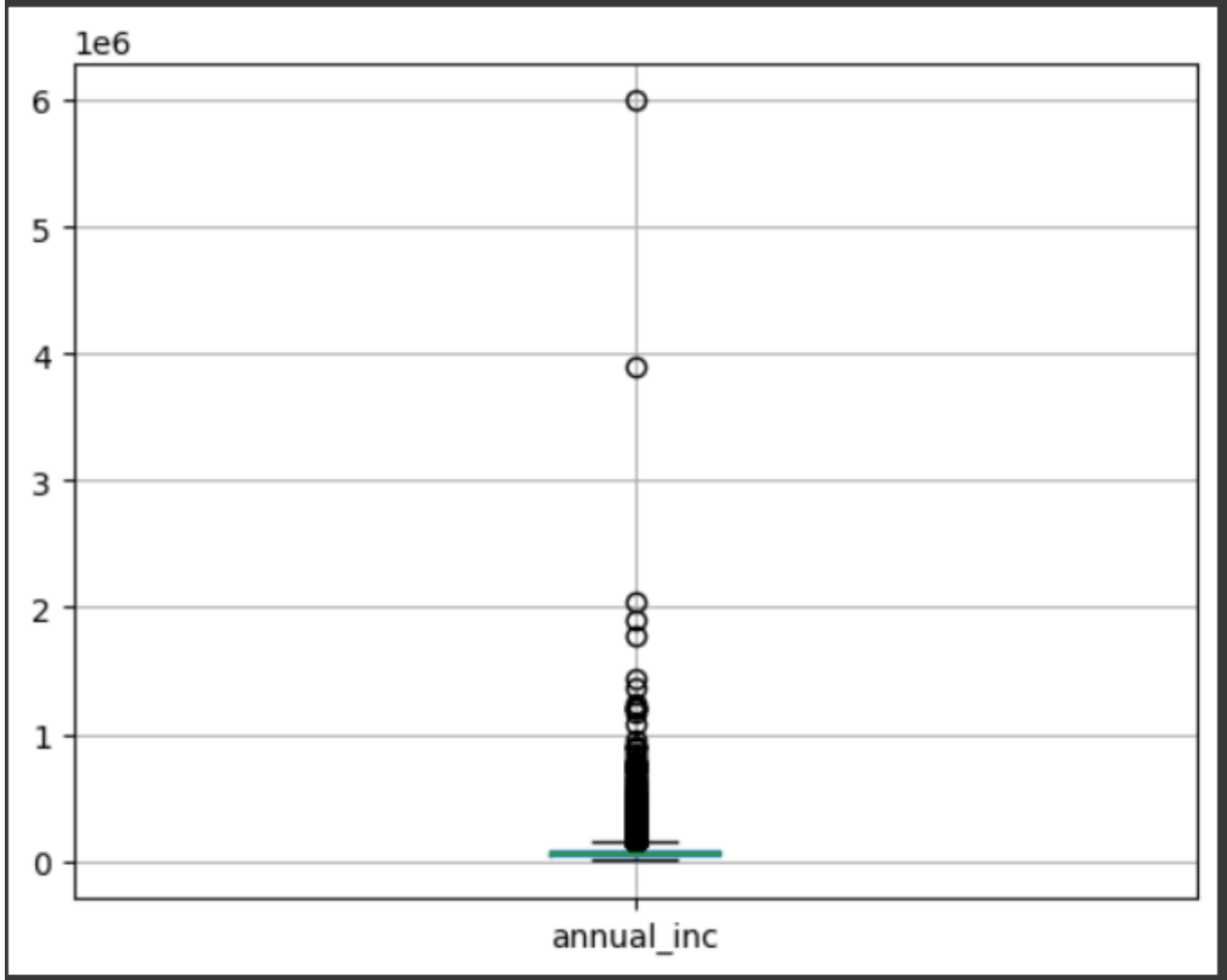
	id	member_id	loan_amnt	funded_amnt	funded_amnt_inv	installment	annual_inc	dti	delinq_2yrs
count	3.857700e+04	3.857700e+04	38577.000000	38577.000000	38577.000000	38577.000000	3.857700e+04	38577.000000	38577.000000
mean	6.763787e+05	8.422843e+05	11047.025430	10784.058506	10222.481123	322.466318	6.877797e+04	13.272727	0.146668
std	2.092639e+05	2.644519e+05	7348.441646	7090.306027	7022.720644	208.639215	6.421868e+04	6.673044	0.492271
min	5.473400e+04	7.069900e+04	500.000000	500.000000	0.000000	15.690000	4.000000e+03	0.000000	0.000000
25%	5.120330e+05	6.611310e+05	5300.000000	5200.000000	5000.000000	165.740000	4.000000e+04	8.130000	0.000000
50%	6.564230e+05	8.392920e+05	9600.000000	9550.000000	8733.440000	277.860000	5.886800e+04	13.370000	0.000000
75%	8.291460e+05	1.037336e+06	15000.000000	15000.000000	14000.000000	425.550000	8.200000e+04	18.560000	0.000000
max	1.077501e+06	1.314167e+06	35000.000000	35000.000000	35000.000000	1305.190000	6.000000e+06	29.990000	11.000000

```
loanData.drop(['url', 'desc', 'pymnt_plan', 'zip_code','earliest_cr_line', 'mths_since_last_delinq', 'mths_since_last_record', 'ou
```

Remove unwanted columns not relevant for analysis

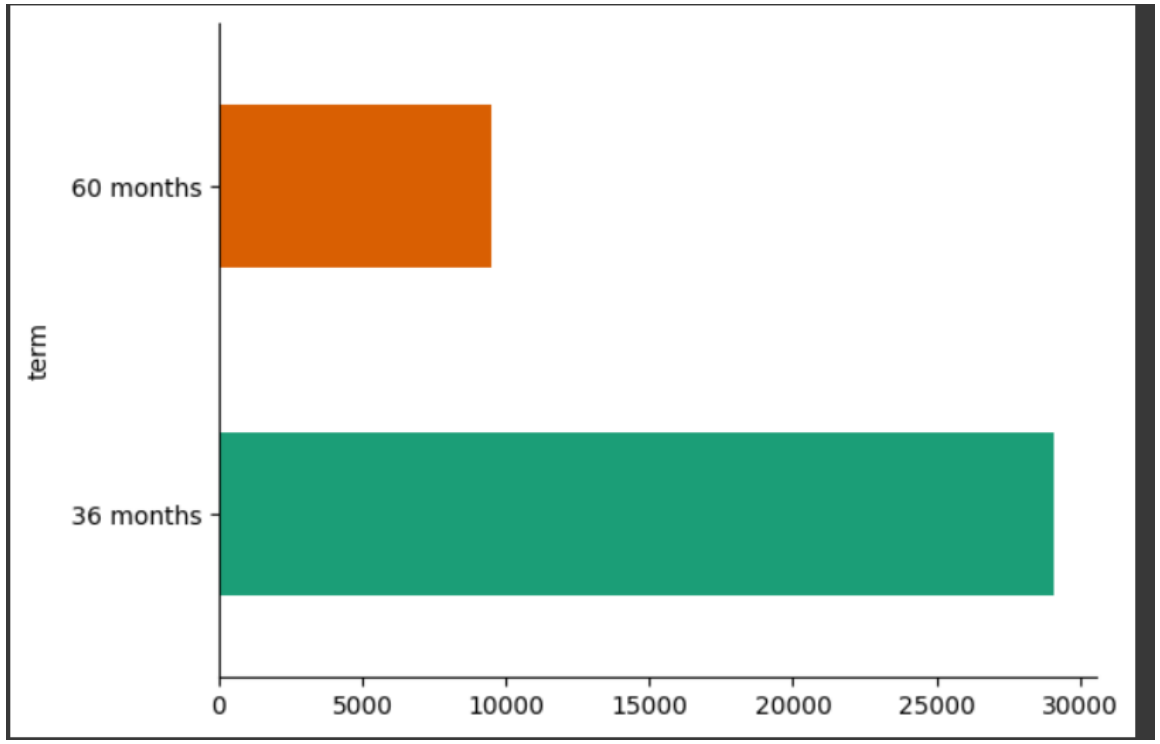
```
loanData.boxplot(column=['annual_inc'])
```

```
<Axes: >
```

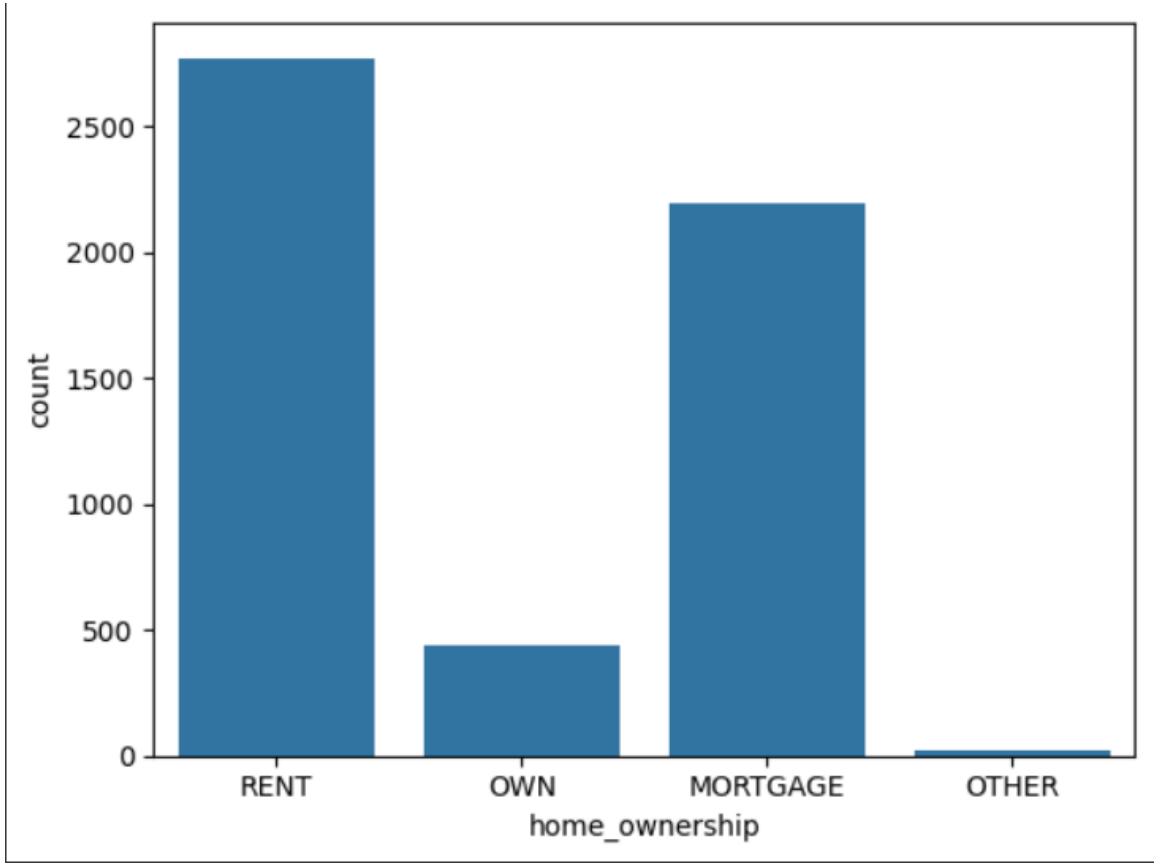


Finding outliers

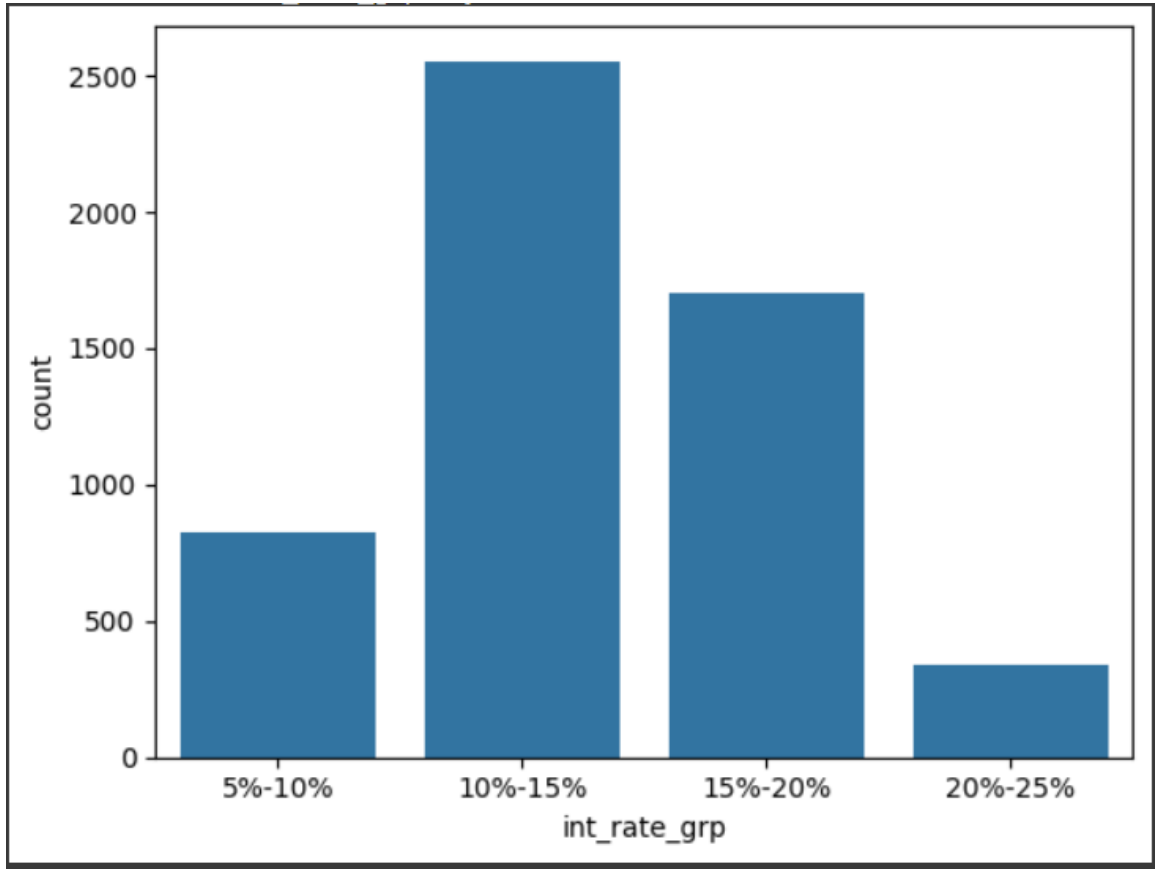
Outcome of the analysis



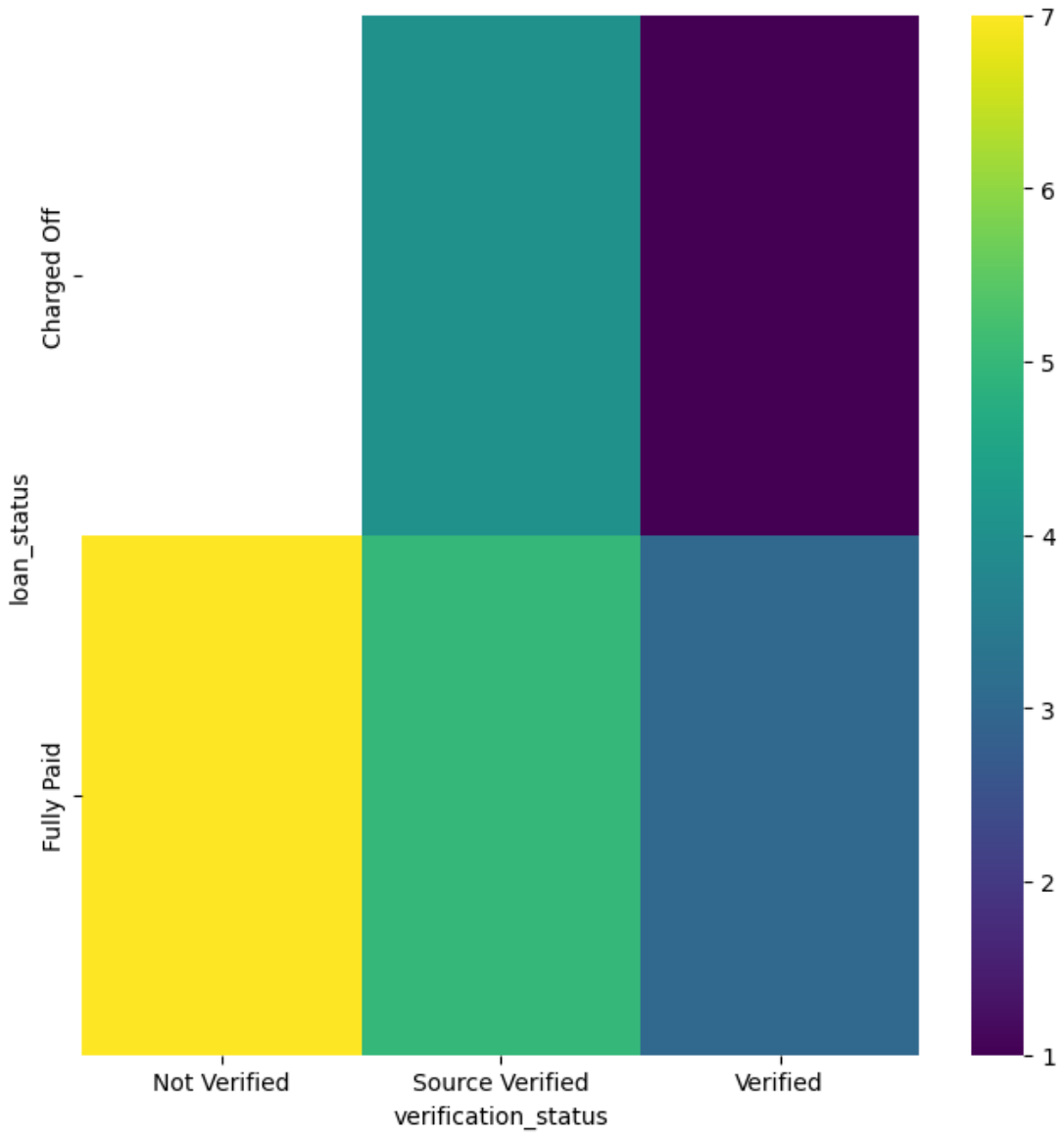
- People with tenure of 36 months have defaulted more



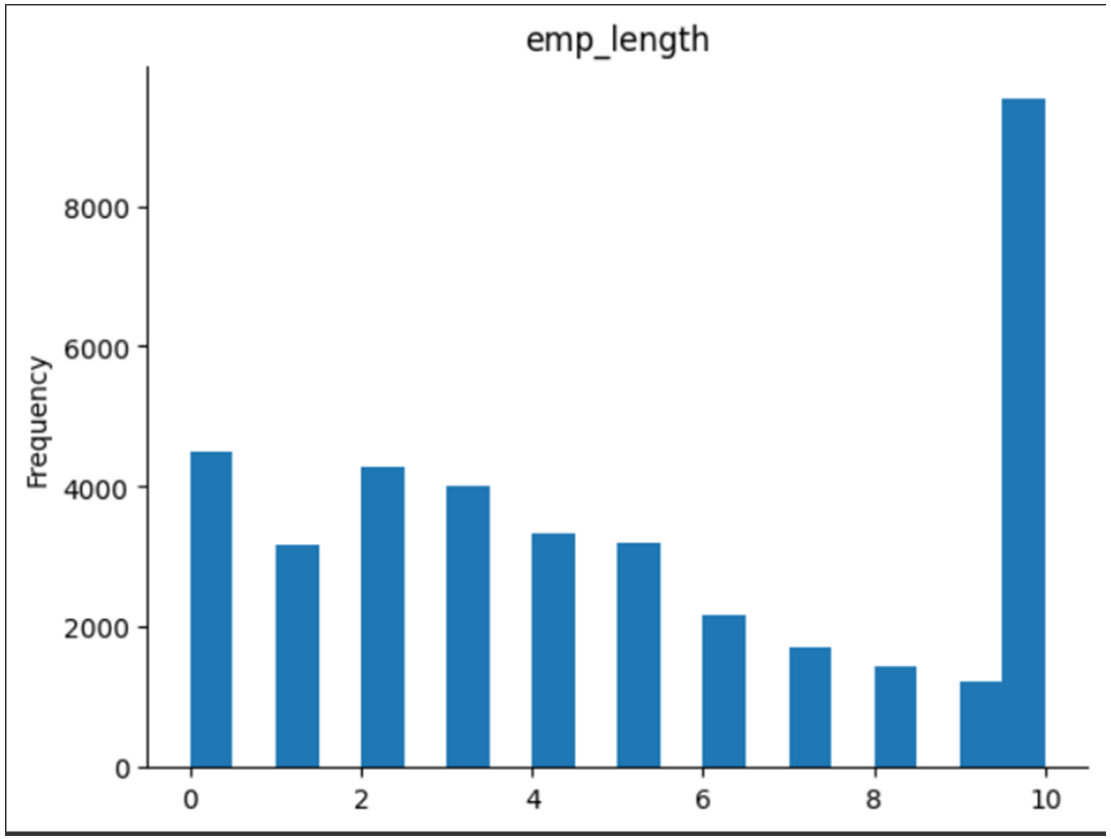
- People home ownership as “Rent” have defaulted



- People who have an interest rate between 10-20% have defaulted



- People who have been “verified” have defaulted



- People who have been employed for more than 10 yrs. have defaulted