LENDING CLUB CASE STUDY

PRESENTED BY:
SEETARAM PANDEY
ABHISHEK SHANMUKHA

PROBLEM STATEMENT

Context: Consumer company focusing on lending loans to urban customers.

Decision Point: Approving or Rejecting loans based on applicant’s profile.

Key challenges:

- Rejecting creditworthy candidates = Business Loss
- Approving risky applicants = Potential Financial Loss

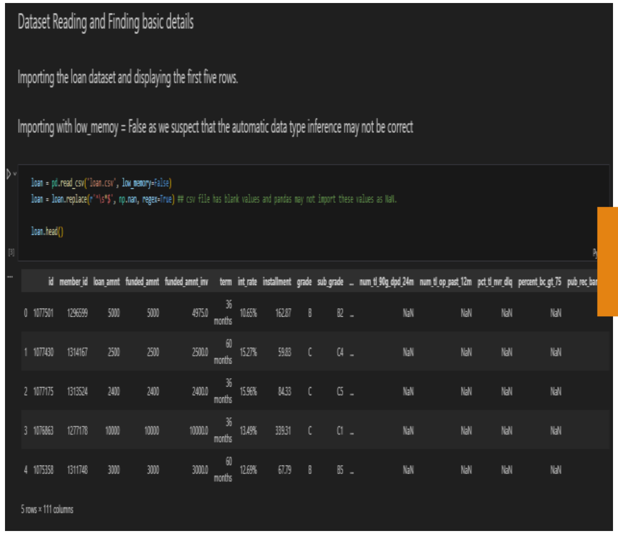
OBJECTIVE

- Identify the driving factors behind loan defaulters.
- Maximize profitability by approving loans for likely-to-repay customers.

Exploratory Data Analysis – Data Cleaning

- Importing the data and displaying the top 5 rows to understand the dataset.
- Check for null values if any – here most of the columns have 100% null values which can be dropped.
- Dropping the columns with records > 50 % NULL Values and confirming the shape after the drop.

Check before dropping



Dataset Reading and Finding basic details

Importing the loan dataset and displaying the first five rows.

Importing with low_memoy = False as we suspect that the automatic data type inference may not be correct

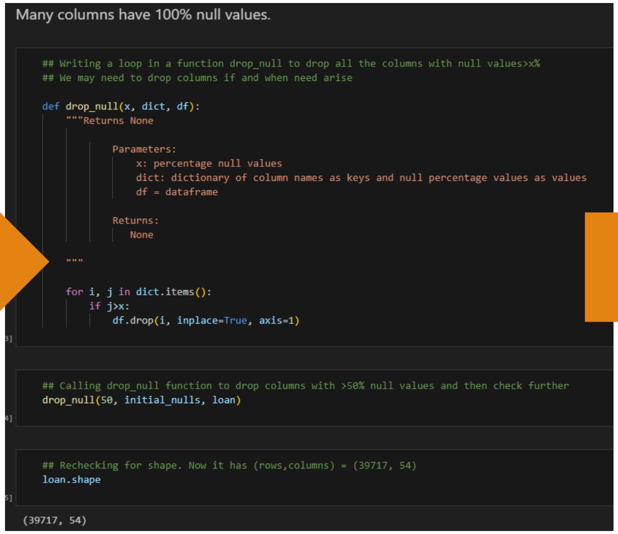
```
loan = pd.read_csv('loan.csv', low_memory=False)
loan = loan.replace(r'^\s*$', np.nan, regex=True) ## csv file has blank values and pandas may not import these values as NaN.

loan.head()
```

	id	member_id	loan_amnt	funded_amnt	funded_amnt_inv	term	int_rate	installment	grade	sub_grade	...	num_tl_90g_dpd_24m	num_tl_op_past_12m	pct_tl_nvr_dlq	percent_bc_gt_75	pub_rec_ban
0	1077501	1296599	5000	5000	4975.0	36 months	10.65%	162.87	B	B2	...	NaN	NaN	NaN	NaN	
1	1077430	1314167	2500	2500	2500.0	60 months	15.27%	59.83	C	C4	...	NaN	NaN	NaN	NaN	
2	1077175	1313524	2400	2400	2400.0	36 months	15.96%	84.33	C	C5	...	NaN	NaN	NaN	NaN	
3	1076863	1277178	10000	10000	10000.0	36 months	13.49%	339.31	C	C1	...	NaN	NaN	NaN	NaN	
4	1075358	1311748	3000	3000	3000.0	60 months	12.69%	67.79	B	B5	...	NaN	NaN	NaN	NaN	

5 rows × 111 columns

Check after dropping



Many columns have 100% null values.

```
## Writing a loop in a function drop_null to drop all the columns with null values>x%
## We may need to drop columns if and when need arise

def drop_null(x, dict, df):
    """Returns None

            Parameters:
                x: percentage null values
                dict: dictionary of column names as keys and null percentage values as values
                df = dataframe

            Returns:
                None

    """

    for i, j in dict.items():
        if j>x:
            df.drop(i, inplace=True, axis=1)
```

```
## Calling drop_null function to drop columns with >50% null values and then check further
drop_null(50, initial_nulls, loan)
```

```
## Rechecking for shape. Now it has (rows,columns) = (39717, 54)
loan.shape
```

(39717, 54)

Insights Gained

Many columns have 100 % Null values out of which the records with > 50% Null are dropped.

The **'Desc'** column has 30% Null values, not dropping the same due to potential importance of available descriptions

The remaining columns have Null values ranging from 0-7 %. Almost 50 columns have been dropped.

Int_rates % has been removed and converted to numeric value and there are no duplicates

Exploratory Data Analysis – Data Standardization

The int_rate column is converted to the numeric data type and also % is removed for easy computations.

Null values of the Revol_util column is filled using mode.

Revol_util converted to numeric data type and % sign is removed for easy computations.

Data Standardization

```python
## Filling null values of 'revol_util' with mode as it is currently object type

mode_revol_util = loan['revol_util'].mode()
loan['revol_util'] = loan['revol_util'].fillna(mode_revol_util)
```

✓ 0.0s

```python
## Will convert 'revol_util' (Revolving balance utility rate) to numeric type and remove '%' sign.
##Currently it is 'object' type which is wrong.


loan['revol_util'] = loan['revol_util'].apply(lambda x: str(x)) ## Since it already has some float values. We need to perform below action
loan['revol_util'] = loan['revol_util'].apply(lambda x: float(x.strip('%'))) ## converting each value to float and stripping the % symbol
```

✓ 0.0s

```python
loan['revol_util'].dtype
```

✓ 0.0s

```
dtype('float64')
```

Univariate Analysis

Loan Amount:

The maximum loan amount is 35,000.

The count of people who applied for a 35k loan is almost 800 based on the distribution plot below.

35,000 can be treated as an outlier, however, it is not required for us to drop those records as there could be the possibility of potential defaulters at this level.

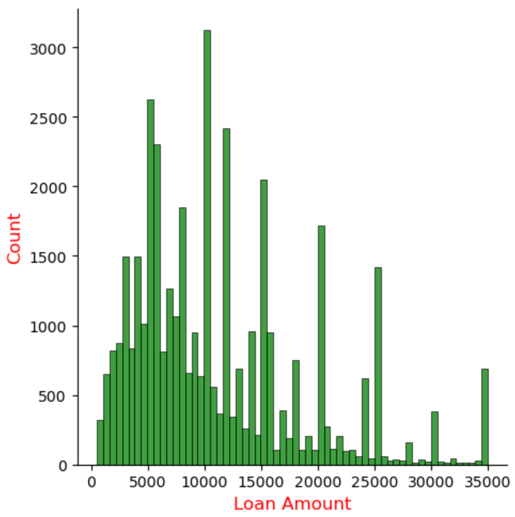
Interest Rates:

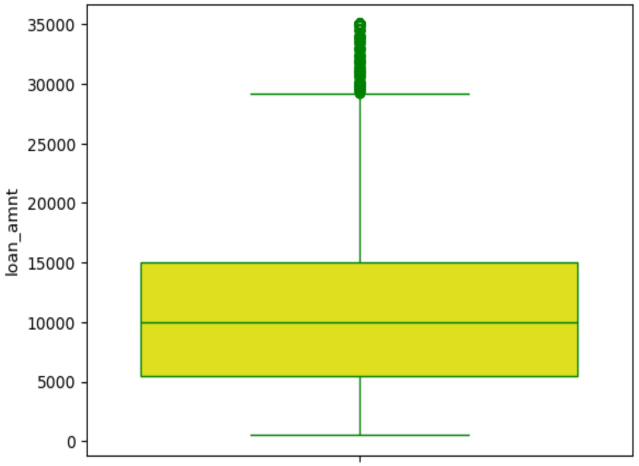
The maximum interest rate is 24.59%.

However, people have been granted loans at 10.99% at the highest.

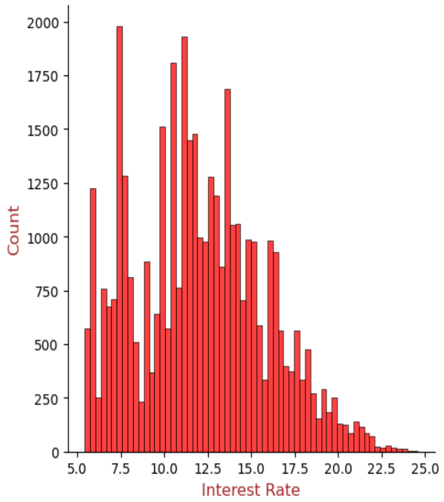
There are a few cases where the maximum interest rate is provided at 24. 59 % which is an outlier, however, it cannot be removed as this could be a potential defaulter.

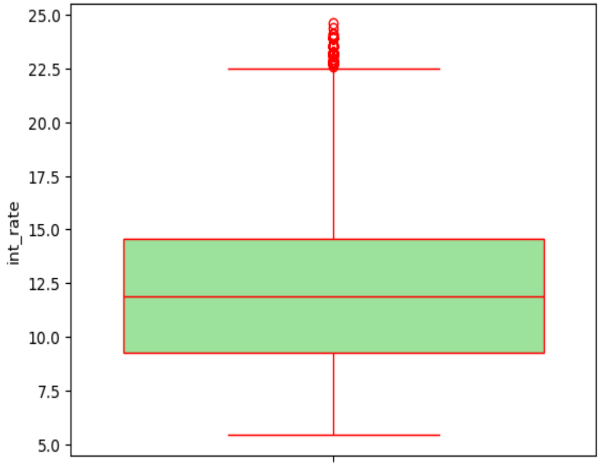
Loan Amount Distribution





Interest Rate Distribution

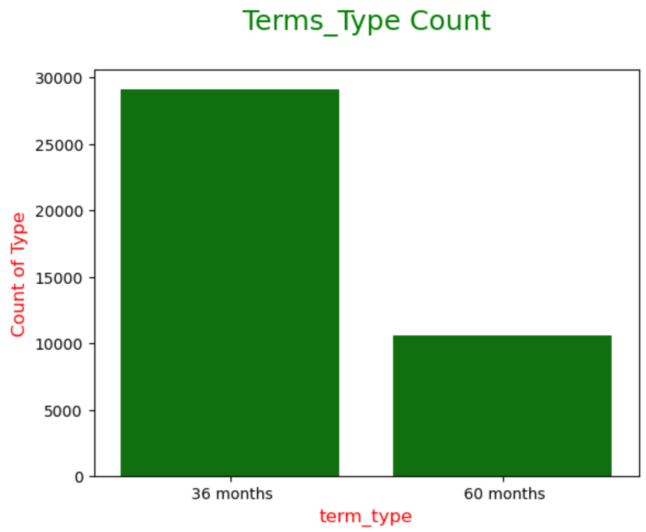




Univariate Analysis

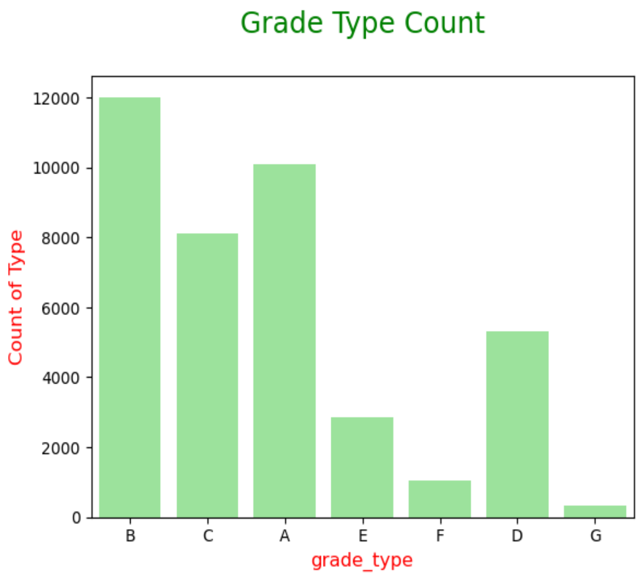
Loan Term:

- There are two types of loan terms provided/opted which are 36 months and 60 months.
- The graph below shows that almost 29,000 applicants opted for 36 months of tenure.



Grade Type:

- There are multiple grades applicants are divided into before the loan is sanctioned.
- The graph below shows that grades A, B, and C have a high chance of getting the loan sanctioned.



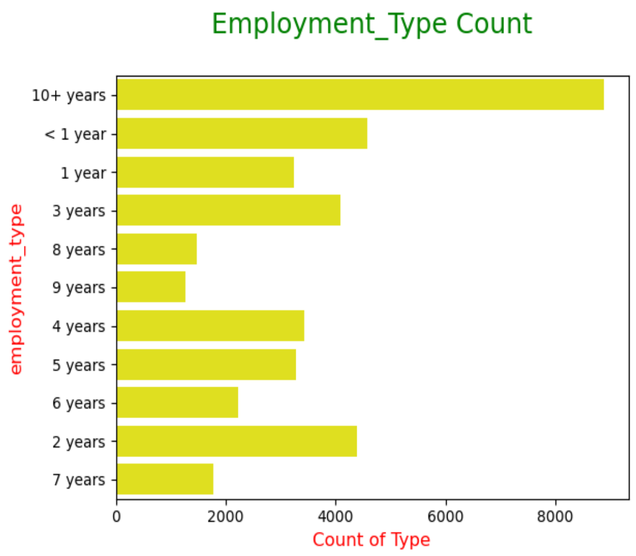
Univariate Analysis

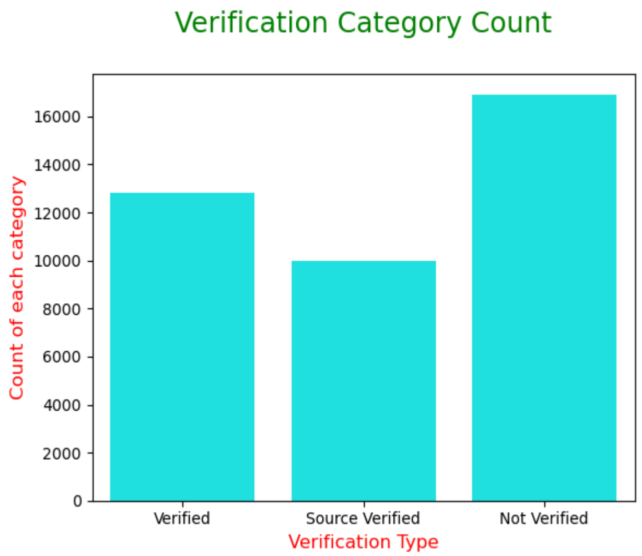
Employment Length:

- Employment Length gives the idea about the number of years the applicant has experience working in an industry.
- The graph gives us the information that most of the applicants who have 10+ years of experience are considered for loan approval.

Verification Status:

- Before the loan Is approved, the company verifies the information provided by the applicant.
- The below insight shows us that most loans are approved but are not verified.
- This can be the information about the potential defaulters.

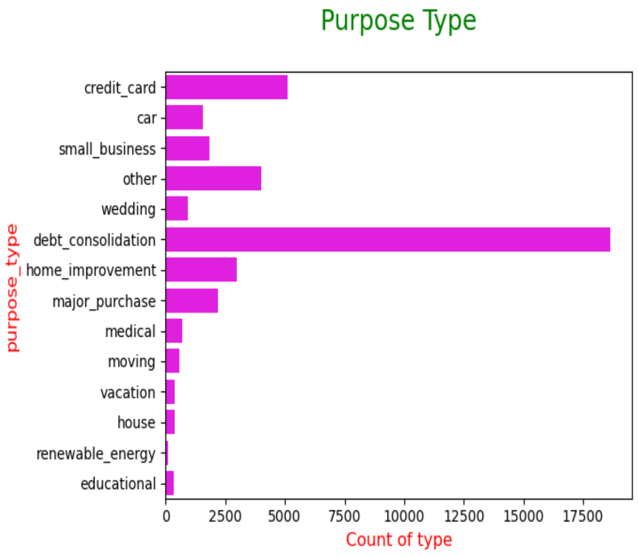




Univariate Analysis

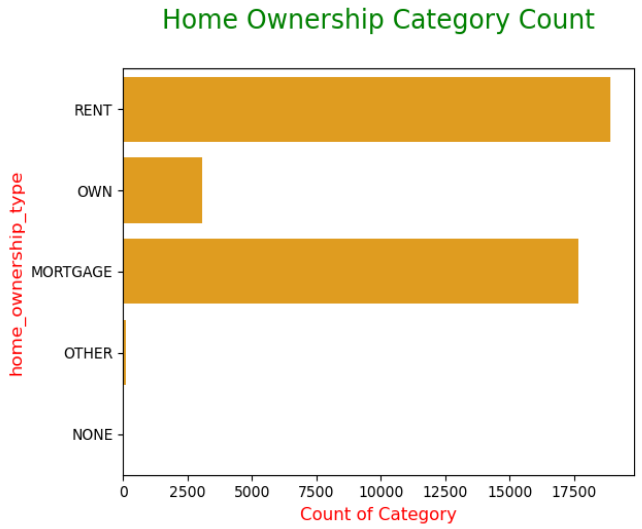
Purpose:

- Usually the company also understands the reason for the loan application.
- As per the below visual, we understand that major loan applications are for debt consolidation.



Home Ownership:

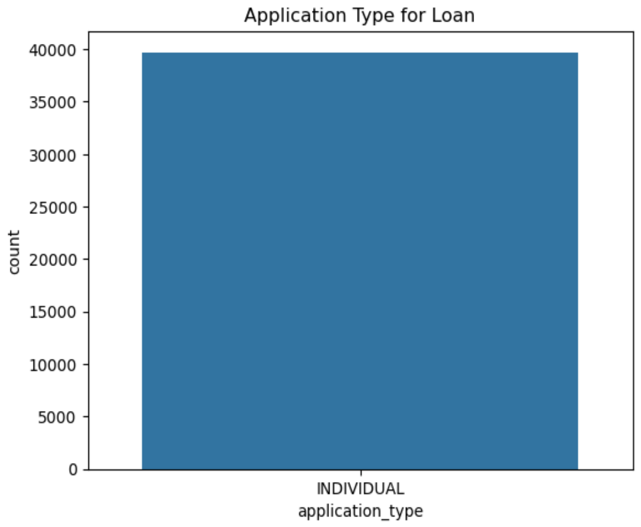
- Before the loan Is approved, the company verifies if the applicant owns a house which can become the surety if the applicant is charged off.
- The below insight shows us that most loans are approved but the applicants are on rent or mortgage.
- This can be the information about the potential defaulters and high risk for the company.



Univariate Analysis

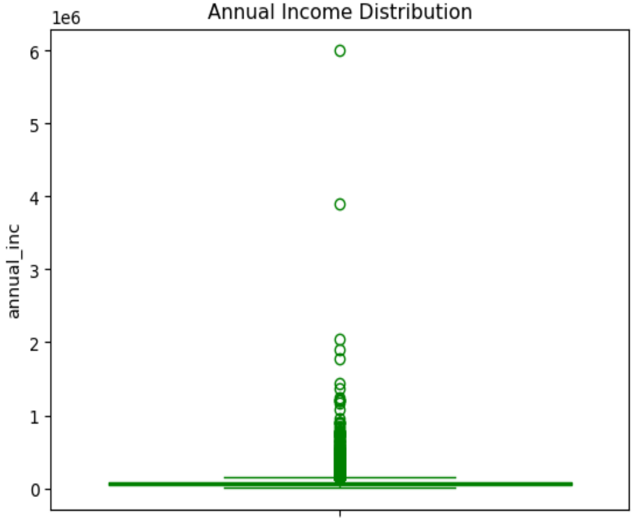
Application Type:

- Application type can be individual or can have a co-applicant.
- The given dataset has records where the loan applicants were all Individuals.



Annual Income:

- Major loans are provided for the applicants with 60k Annual Income
- However, we do see some outliers which we are not dropping as sense likely a wrong information as some data is not verified.



Bivariate Analysis

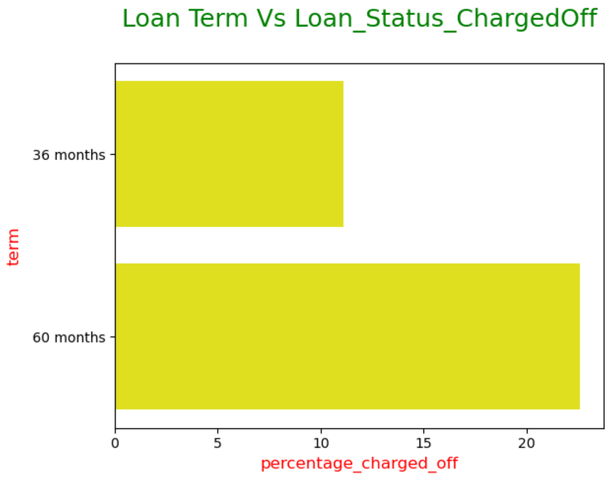
Loan Term vs Loan Status:

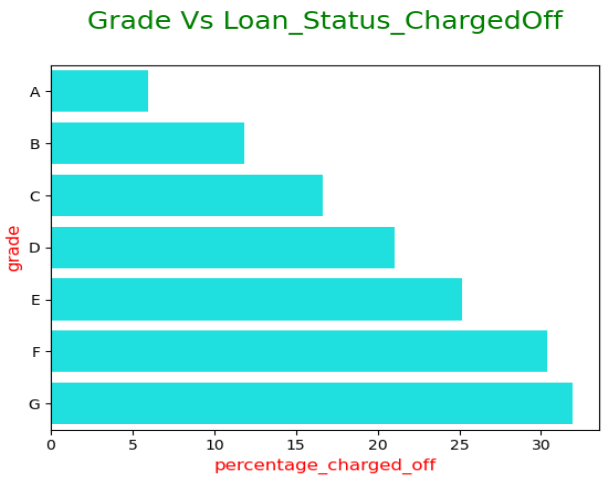
- We observe that the tenure of 60 months has much of a charged-off percentage.
- Additionally, there are current pending cases in the 60-month tenure.
- Avoiding a 60-month tenure would ideally be the best way to avoid the financial loss for the company.

Grade vs Loan Status:

- As mentioned earlier, there are multiple grade levels based on which the loan is approved.
- From the below table, it is understood that the grades E,F and G have percentage charged off higher than 25%.
- Suggested to avoid the loan approval for the above mentioned grades.

loan_status	term	Charged Off	Current	Fully Paid	percentage_charged_off
0	36 months	3227.0	0.0	25869.0	11.090872
1	60 months	2400.0	1140.0	7081.0	22.596742





loan_status	grade	Charged Off	Current	Fully Paid	percentage_charged_off
0	A	602	40	9443	5.969261
1	B	1425	345	10250	11.855241
2	C	1347	264	6487	16.633737
3	D	1118	222	3967	21.066516
4	E	715	179	1948	25.158339
5	F	319	73	657	30.409914
6	G	101	17	198	31.962025

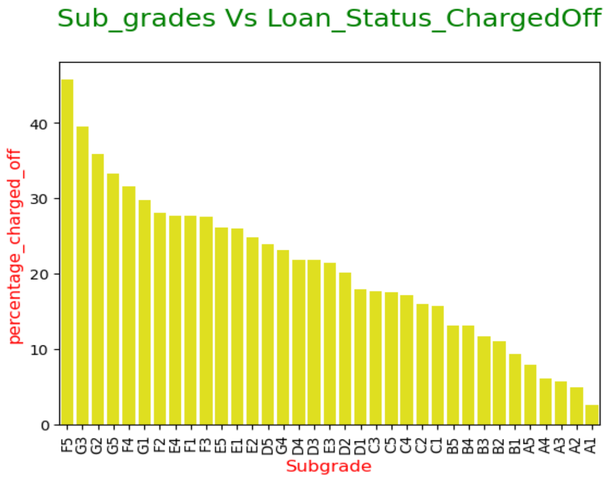
Bivariate Analysis

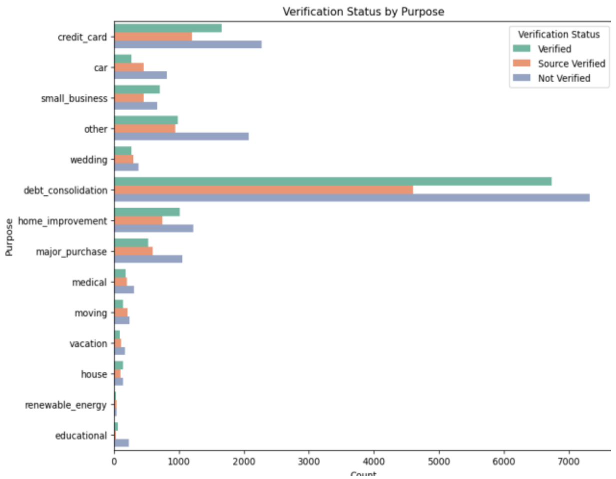
Sub-Grade vs Loan Status:

- The grades are divided into sub-grades in the dataset.
- This analysis also proves that grades E, F, and G exhibit more charged-off cases.

Verification Status vs Purpose:

- A lot of approvals are for debt consolidation for those who were not verified.
- This might lead to financial loss for the company.

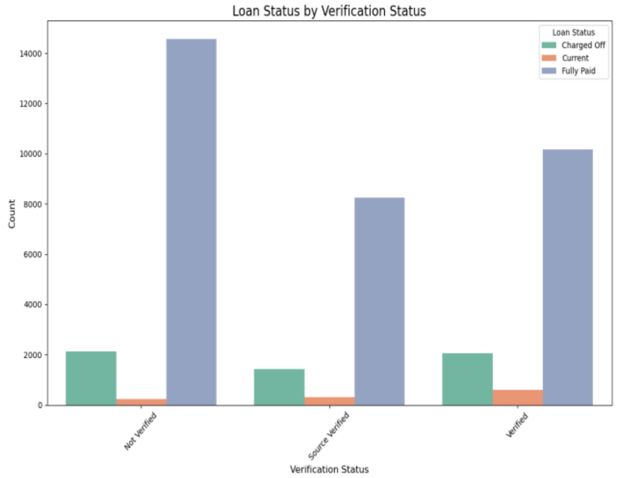




Bivariate Analysis

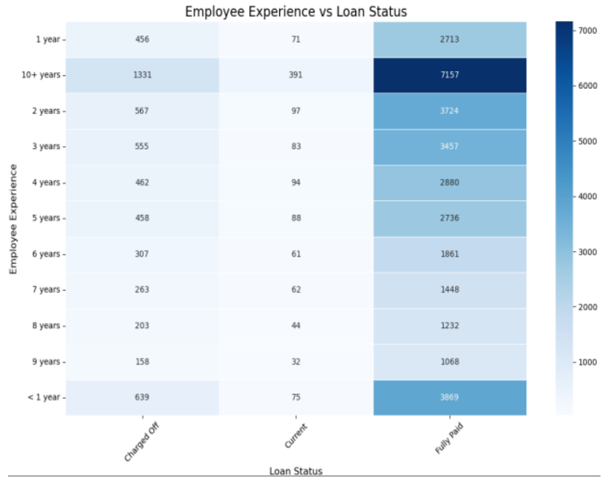
Verification Status vs Loan Status:

- We observe that the fully paid applicants are the ones among not verified.
- However, the charged-off cases remain almost the same in all three loan status types.



Employee Length vs Loan Status:

- The heatmap below shows that for both charged-off cases and fully paid cases, the employee length is 10+ years.
- Univariate Analysis showed that 10+ years are approved for loans, and thus they are also defaulters.



Derived Metrics Analysis

Public Record vs Loan Status:

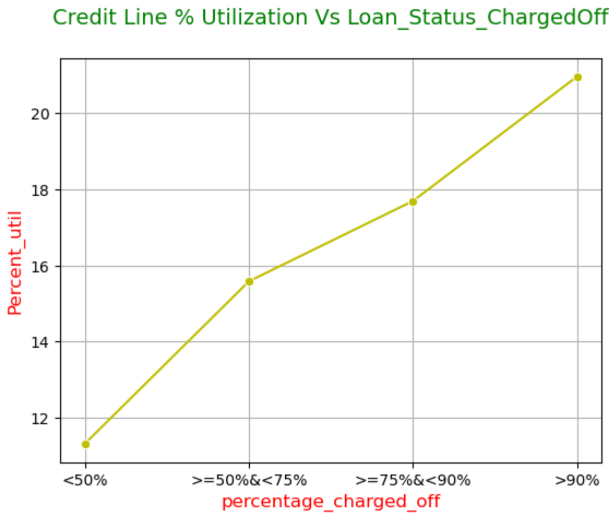
- Deriving a new column public_record from the pub_rec column.
- Creating pivot table with columns "public_record" and loan_status and aggregating on member count
- It is understood that people with 1 or more public records tend to have a higher percentage of charged-off cases.

loan_status	public_record	Charged Off	Current	Fully Paid	percentage_charged_off
1	Yes	467	46	1603	22.069943
0	No	5160	1094	31347	13.723039

Percentage Utilisation of Credit Line vs Loan Status:

- Deriving a new column perc_util from the revol_util column.
- It is understood that, as the credit line utilization increases, the charged-off cases also increase.

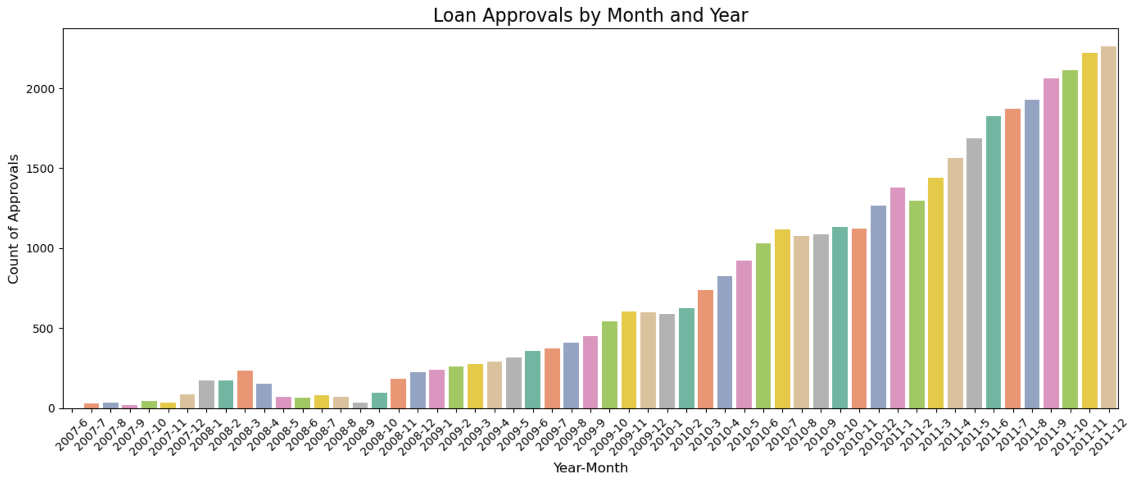
loan_status	perc_util	Charged Off	Current	Fully Paid	percentage_charged_off
0	<50%	2281	494	17364	11.326282
2	>=50%&<75%	1659	363	8622	15.586246
3	>=75%&<90%	1005	190	4485	17.693662
1	>90%	682	93	2479	20.958820



Derived Metrics Analysis

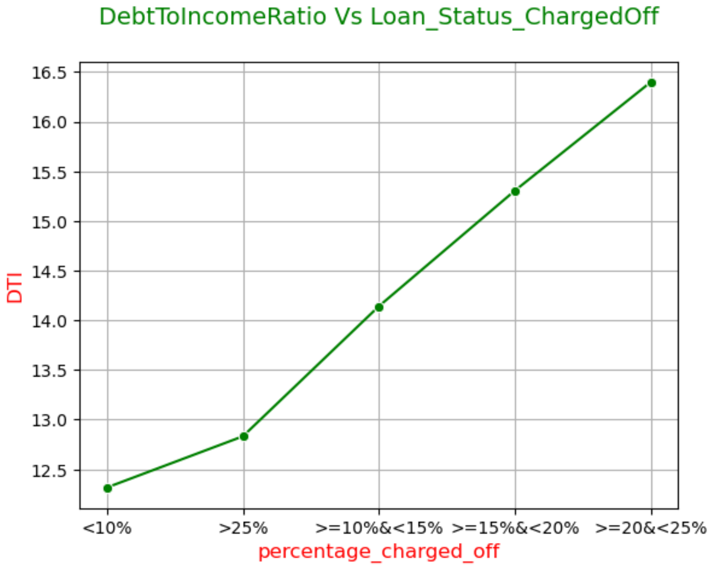
Count of Loan Approvals over the years:

- Deriving two new columns issue_month, issue_year from the issue_d column.
- The graph below gives us insight into the number of loan approvals has been increasing monthly over the years.
- The company is trying to increase the business without inferring the financial loss.



Debt to income ratio and loan status:

- Deriving a new column dti_buck from the dti column.
- The graph below gives us insight as the dti increases, the percentage charged off also increases.



Conclusion

Loan Term: The organization offers a 36-month loan repayment period for approximately 78% of its clients. However, customers with a 60-month loan repayment term are generally more prone to default. This was further confirmed when analyzed alongside the length of employment. Regardless of how long a customer has been employed, those with a 60-month tenure have a greater likelihood of defaulting. Therefore, the company should have a stringent policy to minimize lending for 60 60-month tenure.

Grades: Most of the members are part of grades A and B. However, those in grades G, F, and E have a higher probability of defaulting. Therefore, the company should refrain from extending loans to customers in grades G, F, and E.

Purpose of Loan: The company should be wary or exercise additional caution when providing loans for small business ventures or renewable energy projects, as these are more prone to default. This was further confirmed when analyzed in conjunction with the length of employment.

Credit History: The company needs to exercise extreme caution when lending to customers who have any credit inquiries within the past six months, those with any public record entries, or individuals with a history of bankruptcy.

Customer Financial Situation: The organization needs to thoroughly evaluate the Debt-to-Income (DTI) ratio before making decisions on loan approvals or setting interest rates. The interest rates should not be excessively high. It has been noted that as both the DTI and interest rates rise, the likelihood of default also escalates.

Customer Credit Limit: Customers who make greater use of credit lines are at a higher risk of default. Consequently, the company should focus on this significant variable as well.

Institution Research: The lending institution has failed to consider crucial factors before granting loans. It appears that there is almost no correlation between aspects such as bankruptcy records and loan amounts, debt-to-income ratios and loan amounts, delinquency records and loan amounts, and so forth.

Institution Loss: Typically, around 14% to 15% of members default, which is quite significant.

THANK YOU!!!