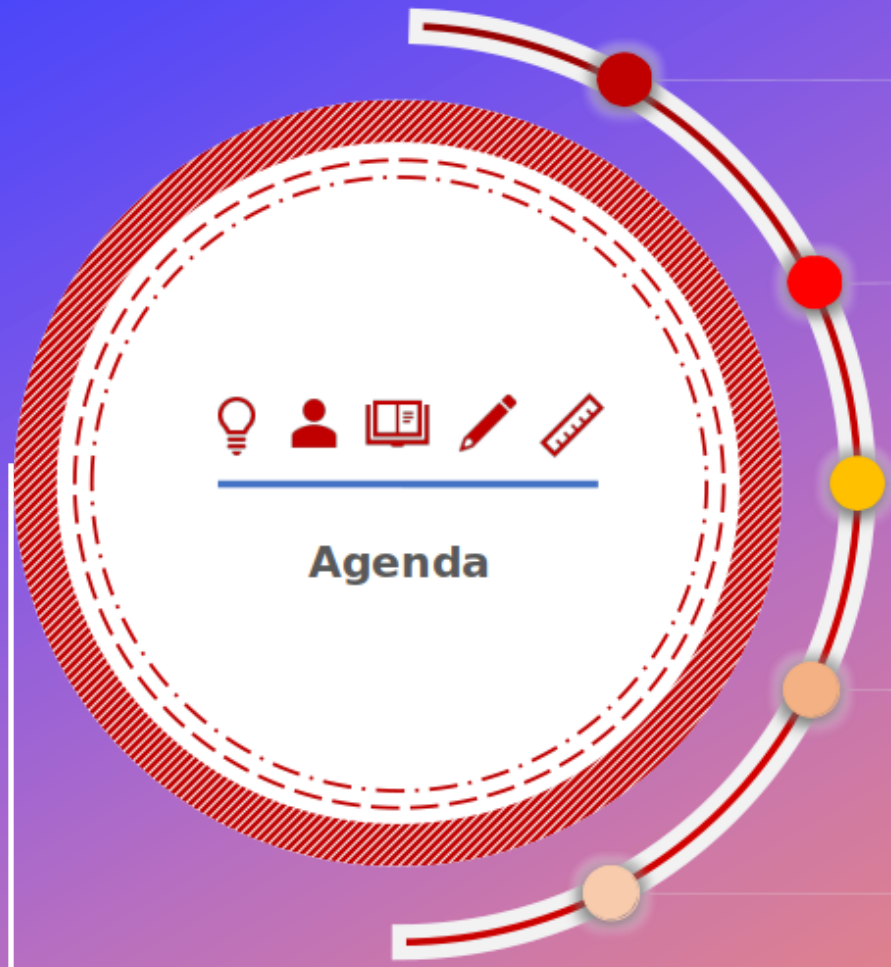
LENDING CLUB CASE STUDY



Agenda

1. Understanding the business problem
2. Explore EDA on the DATA
3. Data Cleaning, Univariate/ Bivariate/ Multivariate Analysis
4. Data Visualization and Find trends/ interesting patterns
5. Conclusion and Key Takeaways

BUSINESS PROBLEM

The aim of the assignment is to apply EDA (Exploratory Data Analysis) concepts to help a finance company make informed decisions on lending loans. The objectives are:

- Determine whether to approve a loan. If approved, decide on the appropriate loan amount and interest rate, ensuring the company minimizes credit loss while retaining the customer.
- Identify cases where loan repayment is highly unlikely and decide not to approve the loan in such instances.

DATASET OVERVIEW AND DATA CLEANING

1

Check the important columns which should not be zero/null. For example "ID" and "Member ID" and should not have repeatation.

2

Check for the columns which have null values in percentage and fill them with mode/median or place holder

3

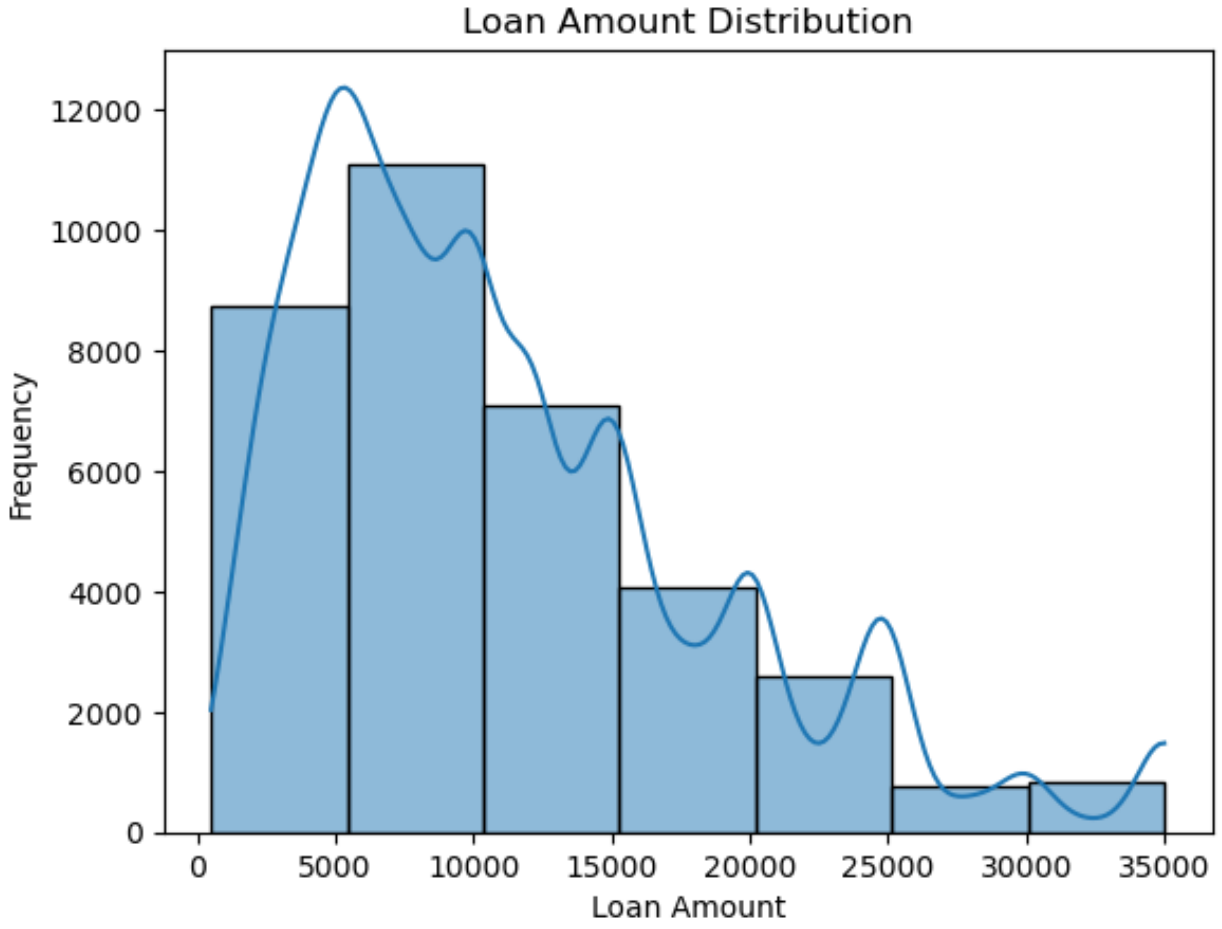
Check for the columns which have macimum nulls and no unique values. Ideally we can drop those columns from rows.

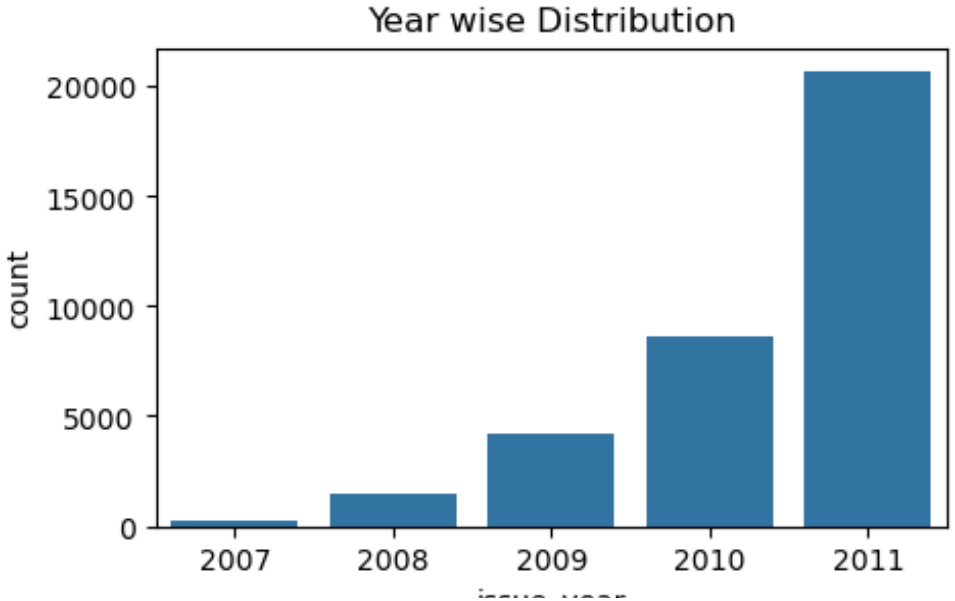
- **Interesting Observation**

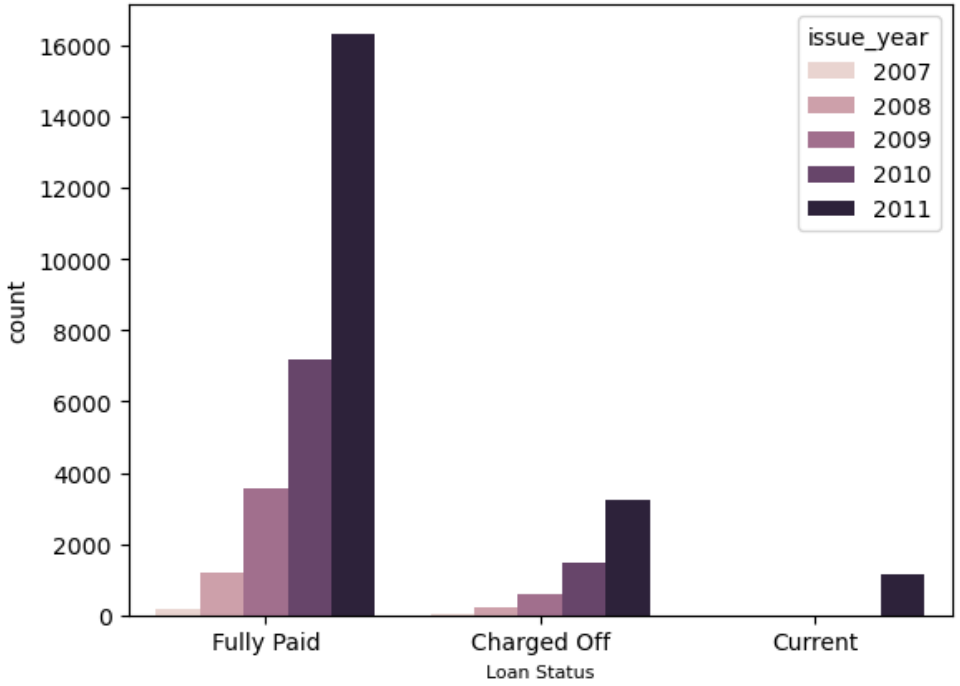
During dataset overview and cleaning process, observed that the total payments made exceeded the expected repayment amount (Installments * Term).

After careful consideration with univariate analysis and loan status, removed to ensure a more accurate analysis.

LOAN AMOUNT ANALYSIS

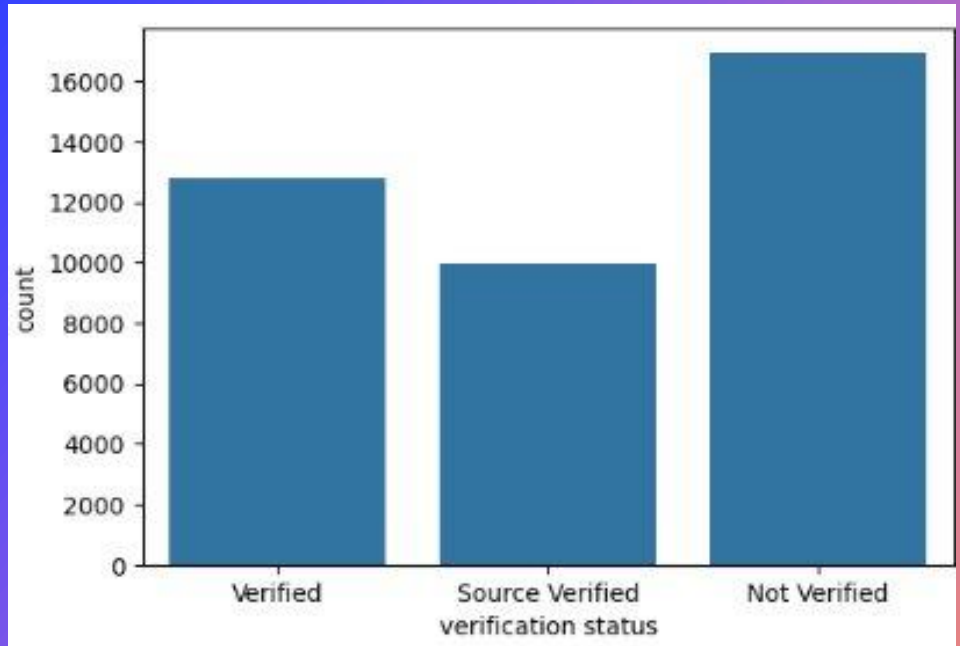


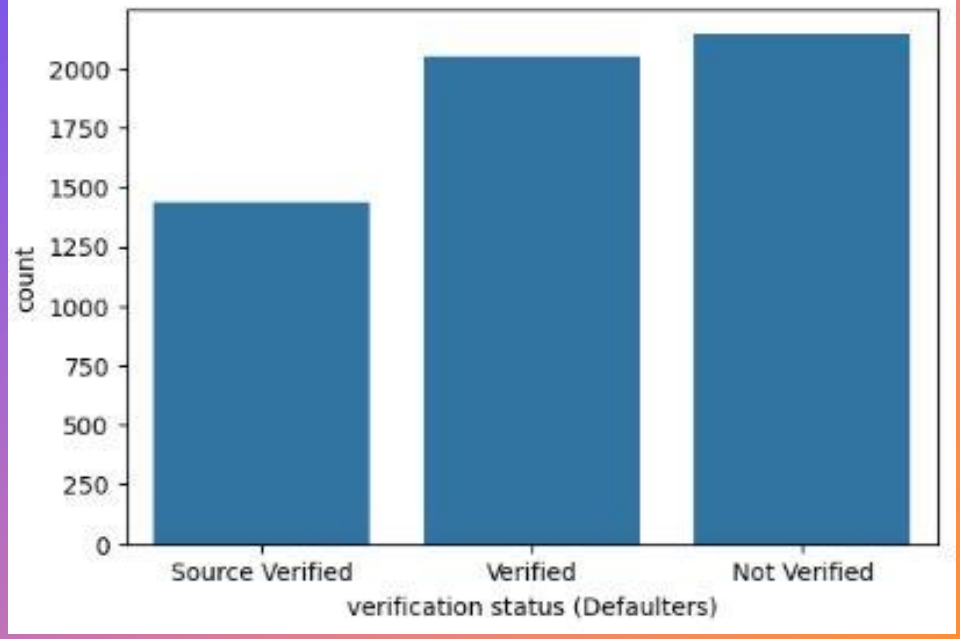




UNIVARIATE ON VERIFICATION STATUS

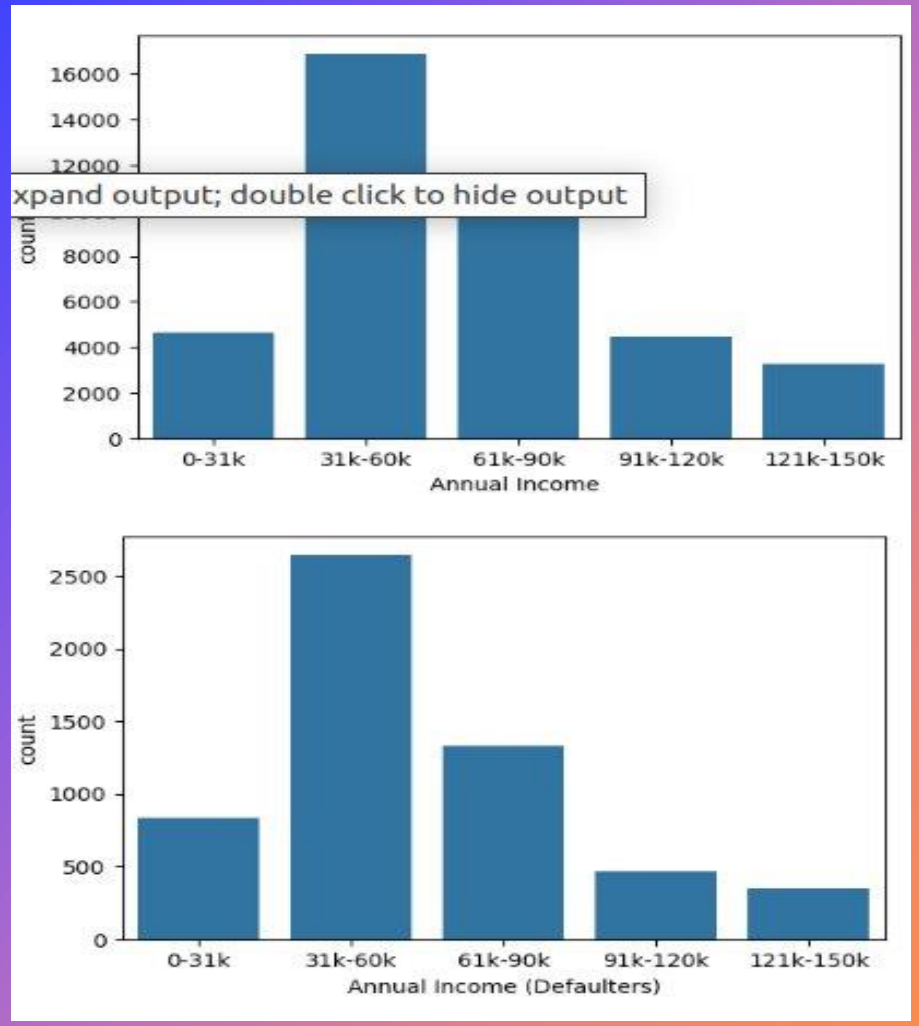
Interestingly, "Verified" profiles show a higher number of defaults, which is a significant concern. Management may need to investigate this further and implement stricter verification processes. This could be a key factor contributing to the issue.





UNIVARIATE ON ANNUAL INCOME

From the univariate analysis of annual income, the majority of loans were issued to individuals with incomes ranging between 31K to 60K. However, defaults appear proportional across this range, and no clear pattern is observed with respect to annual income.



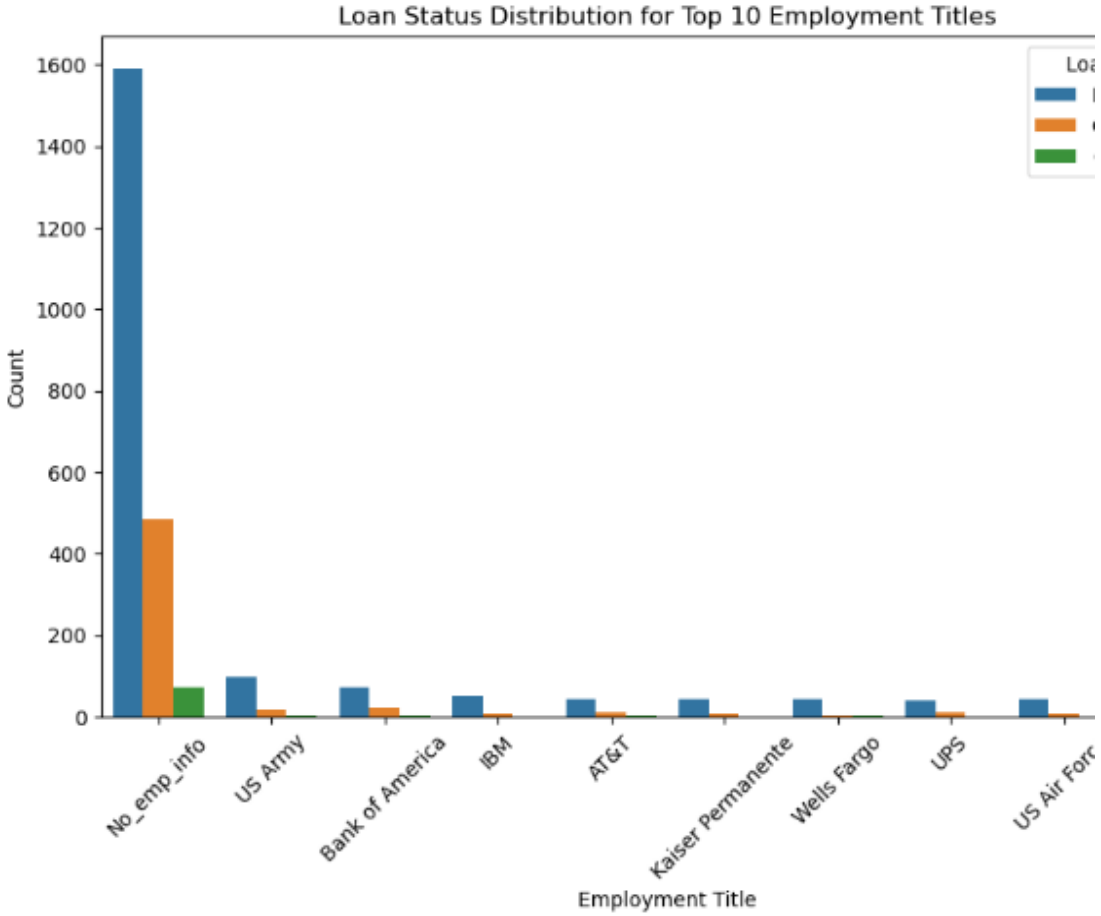
EMPLOYMENT TITTLE ANALYSIS

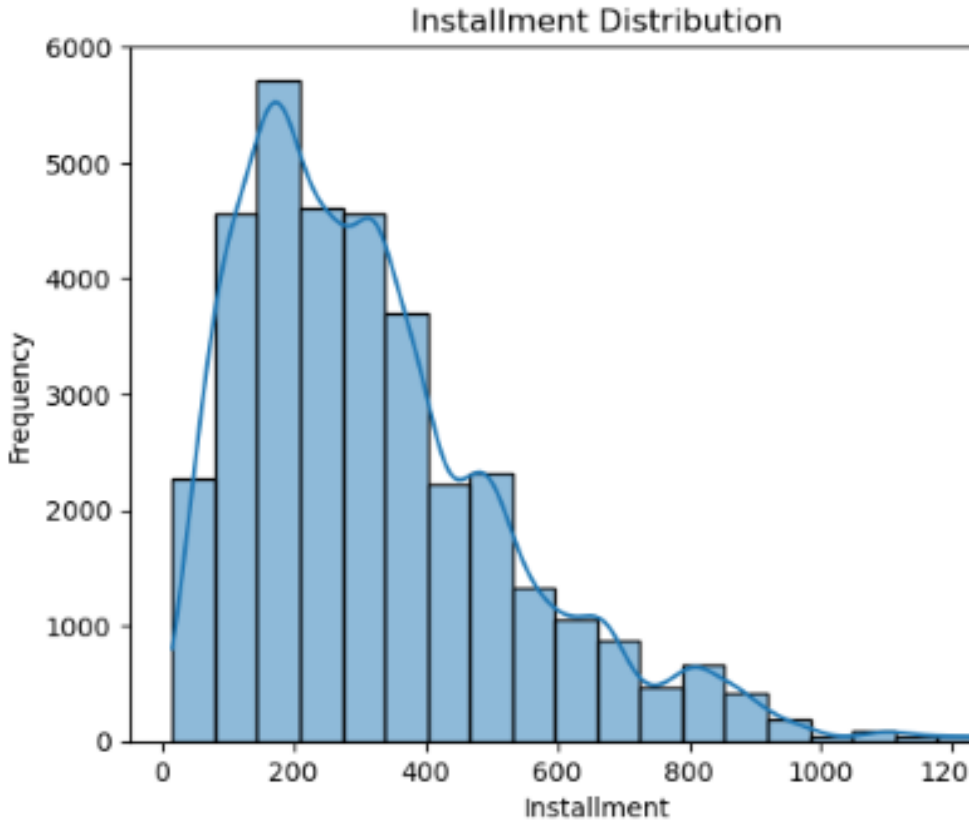
During the data cleaning process, it was observed that many rows were missing employment titles.

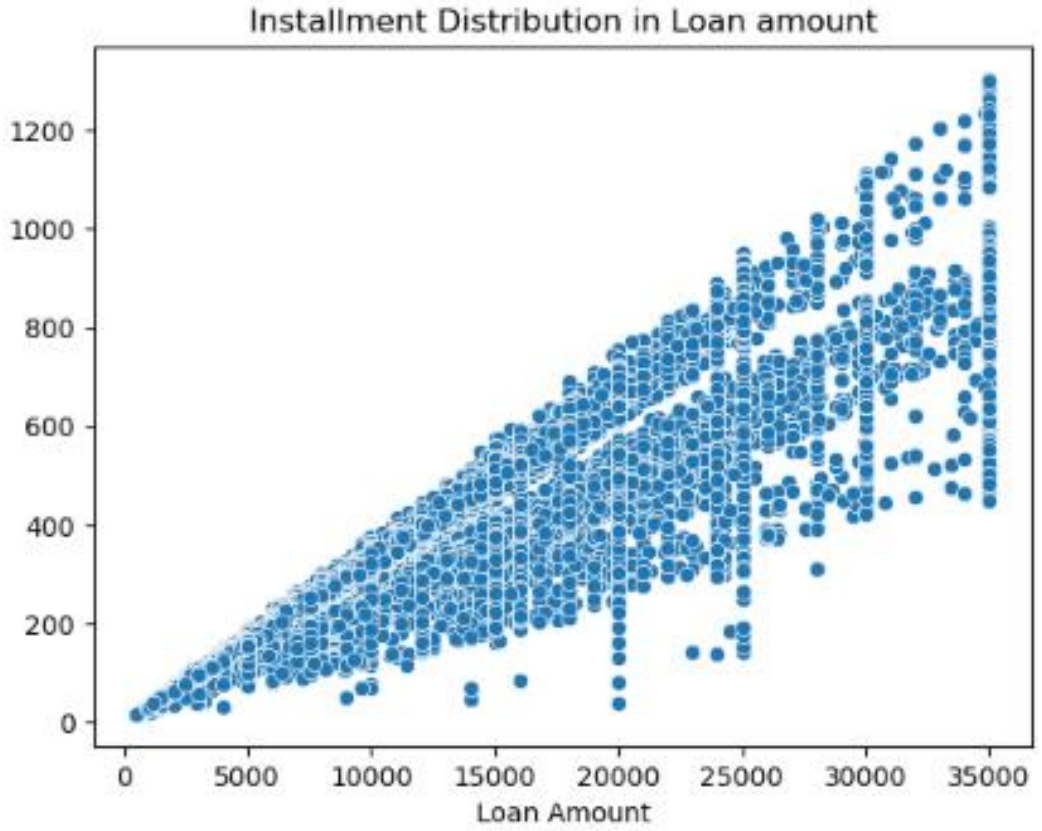
For the analysis, a placeholder "No_emp_info" was used.

A notable concern is that a significant number of loans were granted without employment details.

This highlights a key area for stricter enforcement in the loan approval process.



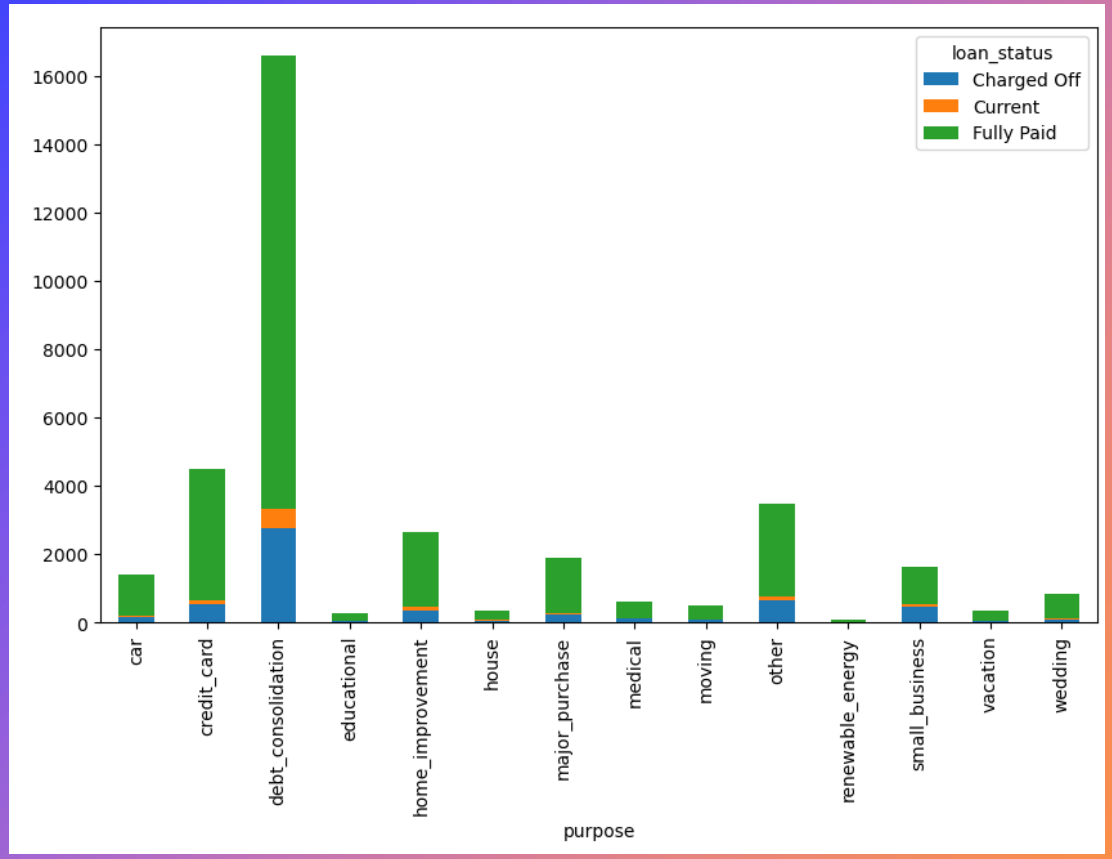




ANALYSIS ON INSTALLMENT AMOUNT

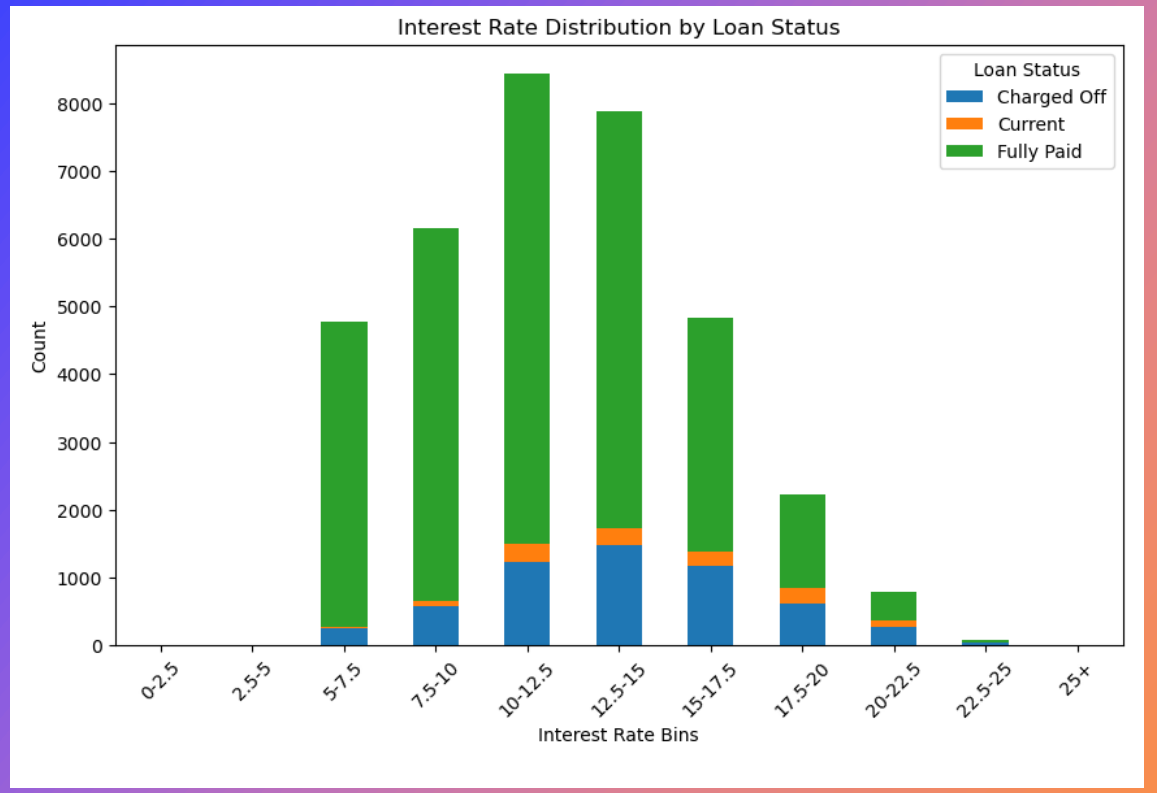
PURPOSE OF LOAN

loan_status	Charged Off	Current	Fully Paid
purpose			
car	159.0	50.0	1208.0
credit_card	540.0	102.0	3852.0
debt_consolidation	2758.0	585.0	13241.0
educational	55.0	NaN	221.0
home_improvement	346.0	101.0	2213.0
house	59.0	14.0	280.0
major_purchase	221.0	37.0	1650.0
medical	104.0	12.0	493.0
moving	92.0	7.0	404.0
other	631.0	128.0	2706.0
renewable_energy	19.0	1.0	74.0
small_business	474.0	74.0	1074.0
vacation	53.0	6.0	286.0
wedding	96.0	21.0	730.0

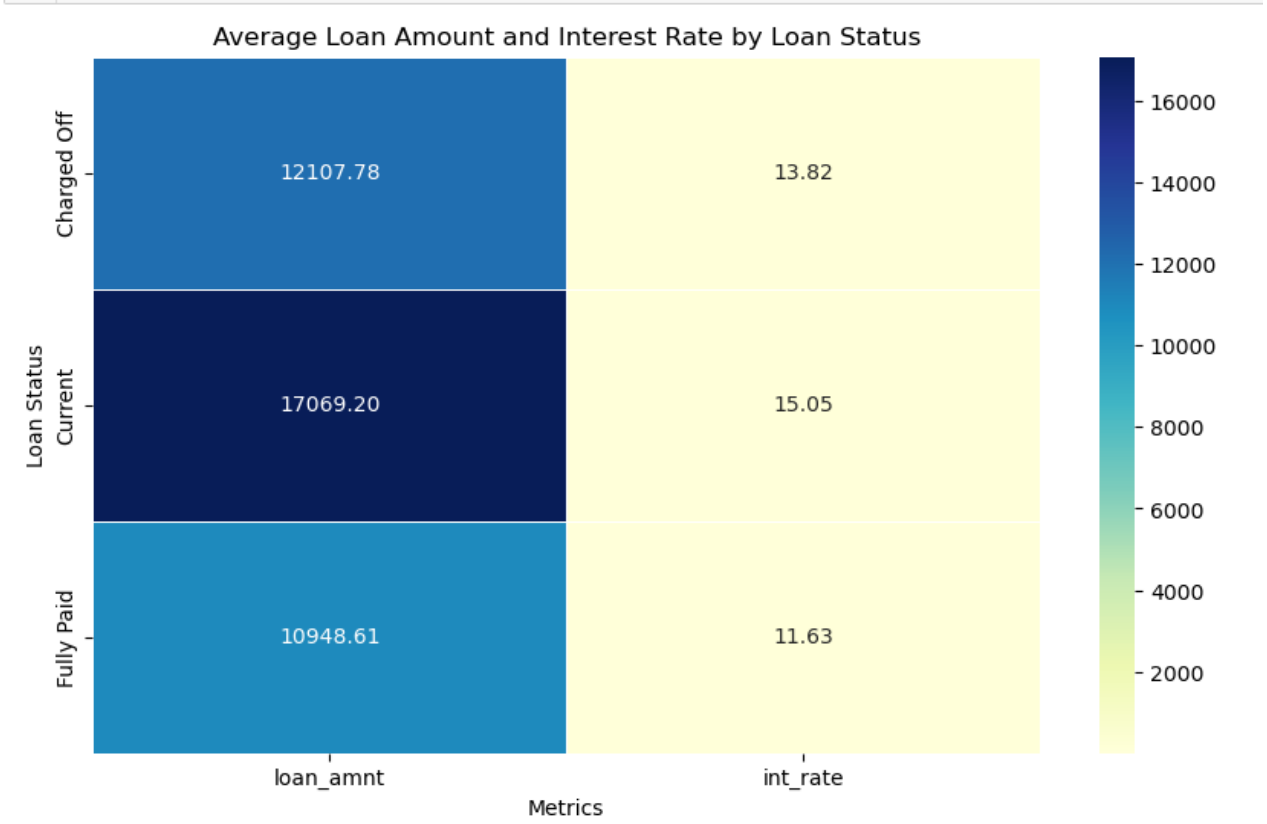


PURPOSE OF LOAN

loan_status in_rate_bin	Charged Off	Current	Fully Paid
0-2.5	0	0	0
2.5-5	0	0	0
5-7.5	255	5	4515
7.5-10	572	70	5512
10-12.5	1226	270	6949
12.5-15	1467	259	6156
15-17.5	1178	208	3453
17.5-20	613	224	1392
20-22.5	266	89	421
22.5-25	30	13	34
25+	0	0	0



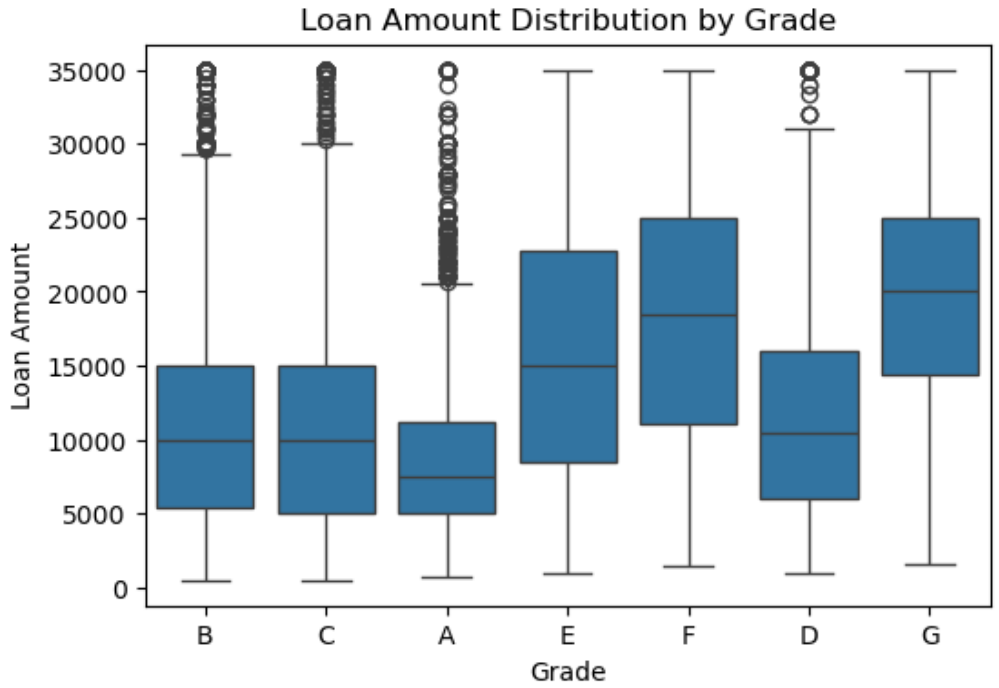
From the heat map distribution, it's clear that higher interest rates, combined with larger loan amounts, are associated with a higher likelihood of defaults. As a firm, this information should be carefully considered when issuing loans in the future.

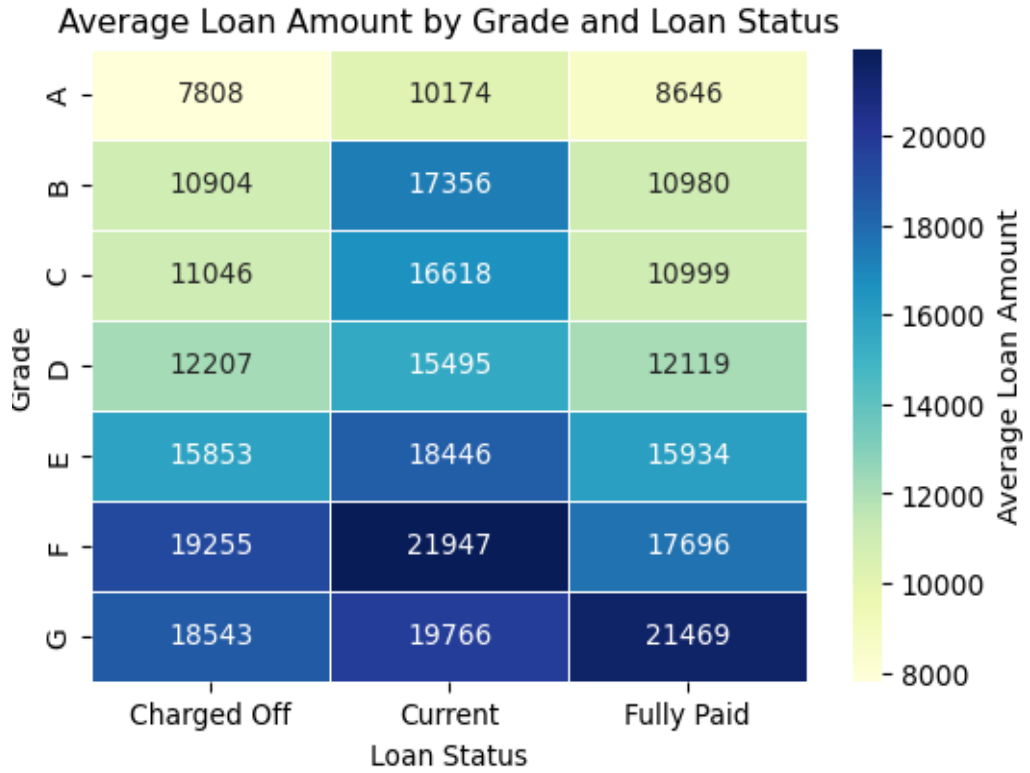


HEAT MAP – AVG LOAN AMOUNT/INTEREST RATE/LOAN STATUS

ANALYSIS ON GRADES

- From the analysis, Grades "E", "F" and "G" got the highest loan amount.
- Defaults are proposonal to the loan amount allotted to grades.
- Grade "A" have many outliers which might be again loophole in the process.





KEY TAKEAWAYS

Below are the important key takeaways.

Total payment made was more than expected amount(installments * term) - This can happen only on special cases where loan got compuonding interest rate. But no data is available. 11% is huge.

Verification process is compromised where maximum defaults are because of no strict process in the verification

Employment title not finding is also another case where confirms loop hole in the system

From all other univariate and bivariate analysis we could conclude that firm should consider moderate average rate with loan amount between 15k to 20k and avoid exceptional cases, with which helps in credit security.

Overall firm is not in great profit and should come with strict processes with in the firm.

```
# Understand the current buisness profti/loss

# total funded amount
total_funded = df1['funded_amnt'].sum()

# total amount which firm will get in return, add it as new column
df1['expected_return'] = df1['installment'] * df1['term']

# Calculate the total amount repaid
total_repaid = df1['total_pymnt'].sum()

print(f"Total Amount Funded: {total_funded}")
print(f"Total Amount Repaid: {total_repaid}")
print(f"Total Amount expected: {df1['expected_return'].sum()}")
```

```
Total Amount Funded: 434810325
Total Amount Repaid: 482704393.92338794
Total Amount expected: 555780868.56
```

THANK YOU