

Lending Club Case Study

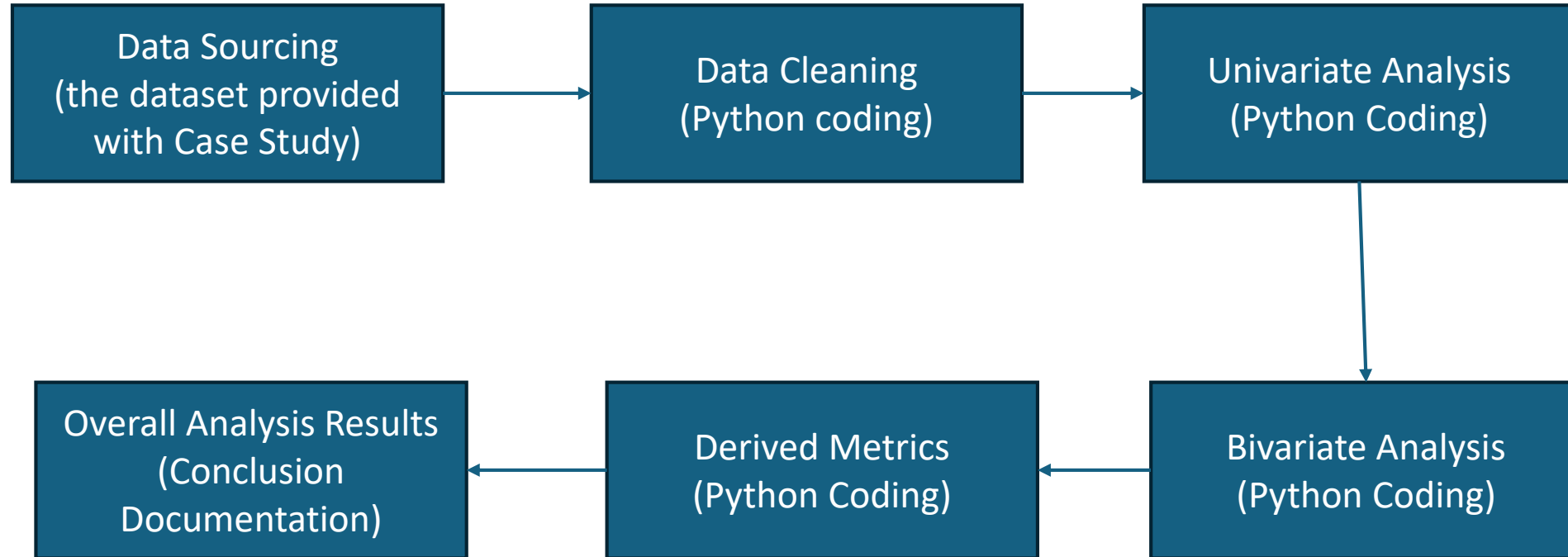
By - Gaurav Chaudhary & Sarita Behera

Problem Statement

- A **consumer finance company** which specialises in lending various types of loans to urban customers. When the company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile. Two **types of risks** are associated with the bank's decision:
 - ✓ If the applicant is **likely to repay the loan**, then not approving the loan results in a **loss of business** to the company
 - ✓ If the applicant is **not likely to repay the loan**, i.e. he/she is likely to default, then approving the loan may lead to a **financial loss** for the company

Problem statement is to identify patterns that can provide guidance to the company to predict applicant's loan repayment behavior and accordingly approve loan application.

Overall Approach for Analysis



Data Sourcing & Cleaning

Data Sourcing : Data provided for the case study in the module has been used

Raw Data Details	
Total Data Entries in Raw Data(“Loan Data Set.csv”)	39,717
Total Columns present in dataset	111

Data Cleaning Steps:

- 1) Missing Value Check
- 2) Dropping columns with more than or equal to 60% null values
- 3) Checking for unique counts
- 4) Identifying sample dataset by applying filter on “Loan Status” = ‘Charged off’ for the analysis

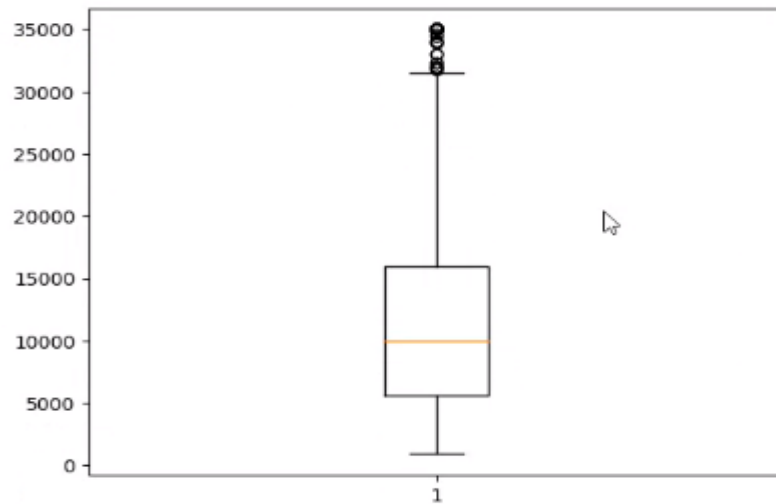
Cleaned Data Details	
Total Data Entries in cleaned data	5627
Total Columns present in dataset	15

Variable Identification for Analysis

Variable Type	Variable Name(Column)
Categorical variables	'addr_state','emp_length','issue_d','purpose','verification_status','grade','sub_grade','loan_status'
Numeric variables	'term','annual_inc','funded_amnt','int_rate','inq_last_6mths','delinq_2yrs'
Additional variables	'total_rec_late_fee','recoveries'

Univariate Analysis

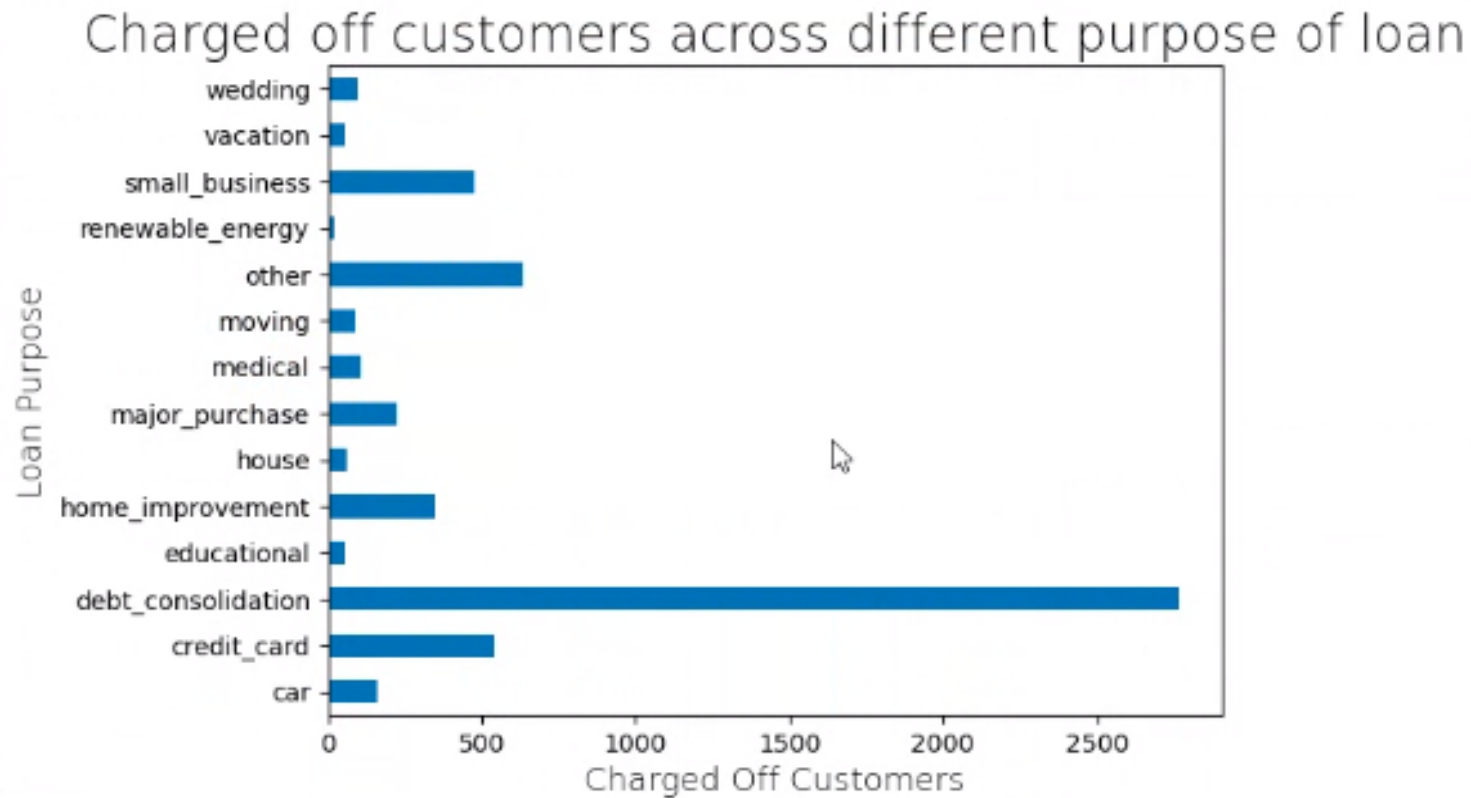
- Using the numeric columns identified, below analysis results found:
 - ✓ More loan was given on “Term” with 36 months in comparison to 60 months
 - ✓ More loan was given for a “Funded Amount” in the range of 5000 – 15000



- ✓ More loan in on lower “Interest Rate” provided. Interest rates have been categorized from “Low” to “Very High” for analysis.

Bivariate Analysis - Details

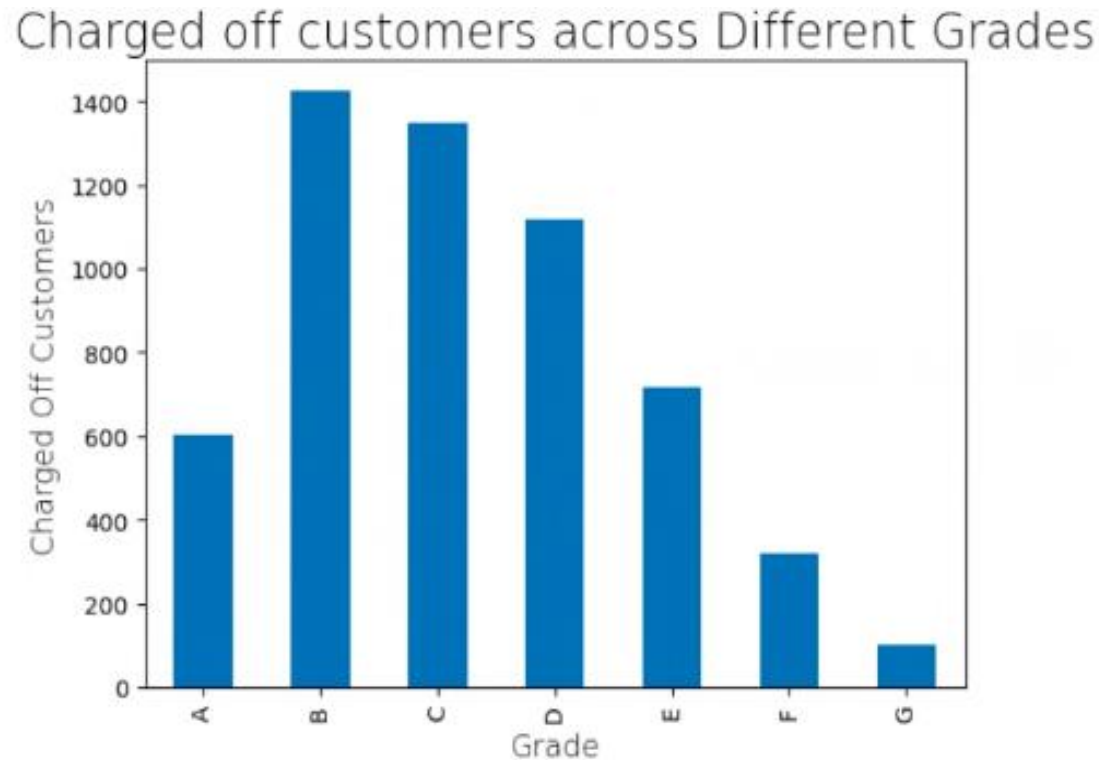
Bivariate analysis has been conducted on the Target Variable of “Loan Status = ‘Charged off’” for ‘Purpose’ to identify the pattern for probable defaulting



Analysis Outcome – It is noted that loan applicants for “Debt Consolidation” purpose have more Charged off loan status. This concludes that there is a more probability of defaulting a loan in the Debt Consolidation category

Bivariate Analysis - Details

Bivariate analysis has been conducted on the Target Variable of “Loan Status = ‘Charged off’” for ‘Grade’ to identify the pattern for probable defaulting



Analysis Outcome – It is noted that loan applicants for “Grade B” have more Charged off loan status. This concludes that there is a more probability of defaulting a loan in the Grade B category

Bivariate Analysis - Details

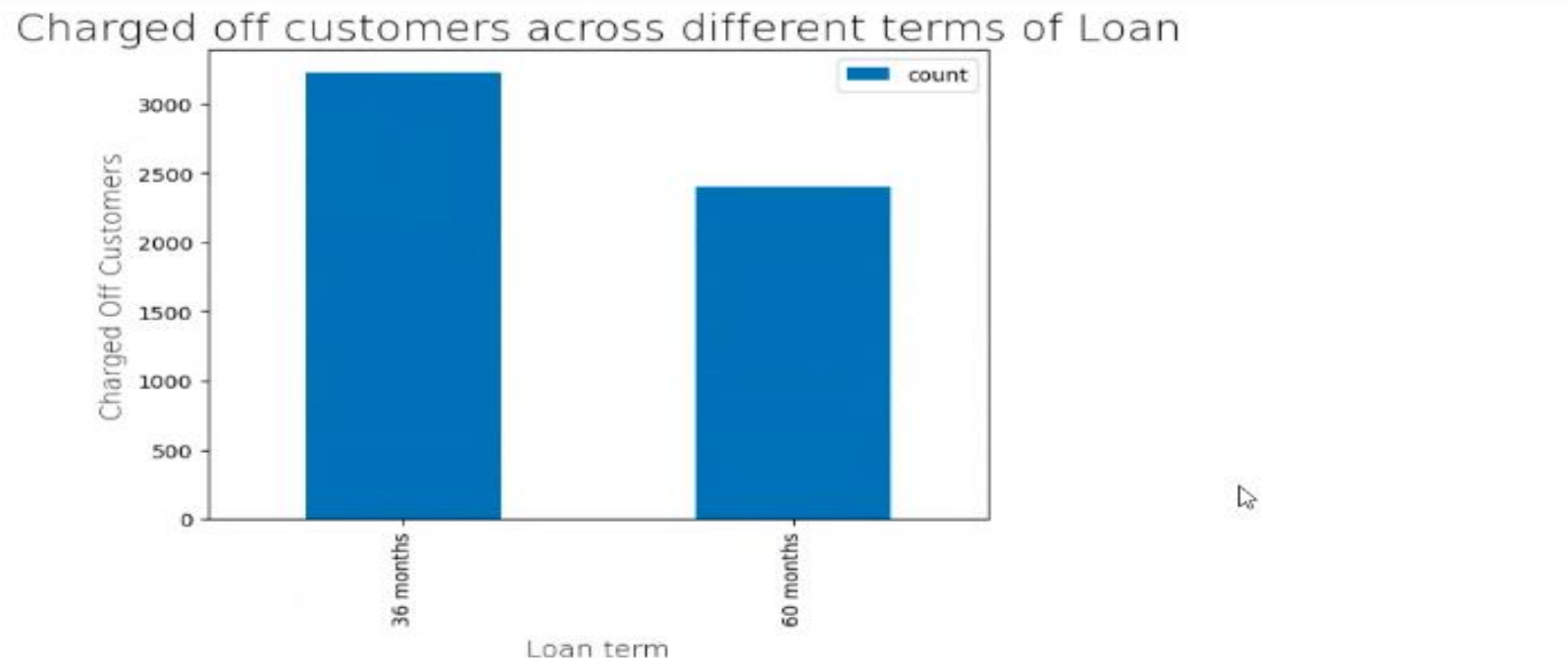
Bivariate analysis has been conducted on the Target Variable of “Loan Status = ‘Charged off’” for ‘Verification_Status’ to identify the pattern for probable defaulting



Analysis Outcome – It is noted that loan applicants for “Not Verified” Annual Salary have more Charged off loan status. This concludes that there is more probability of defaulting a loan in the Not verified category. The probable reasons could be the annual salary has been wrongly indicated by the loan applicant or could have changed(reduced) during the loan period due various reasons which might have led to the defaulting of loan repayment.

Bivariate Analysis - Details

Bivariate analysis has been conducted on the Target Variable of “Loan Status = ‘Charged off’” for ‘Term’ to identify the pattern for probable defaulting



Analysis Outcome – It is noted that loan applicants for “36 months” loan period have more Charged off loan status. This concludes that there is a more probability of defaulting a loan in the 36 month term category. The probable reasons could be that the Monthly Installment of repayment amount(EMI) would have been higher than the repayment capacity of the loan applicant. Therefore, checking the monthly salary and assessing the repayment capacity month-wise would be recommended. There could be multiple other parallel loans that an applicant has which could be leading to defaulting. Therefore, checking for other existing loans would be recommended.

Bivariate Analysis - Summary

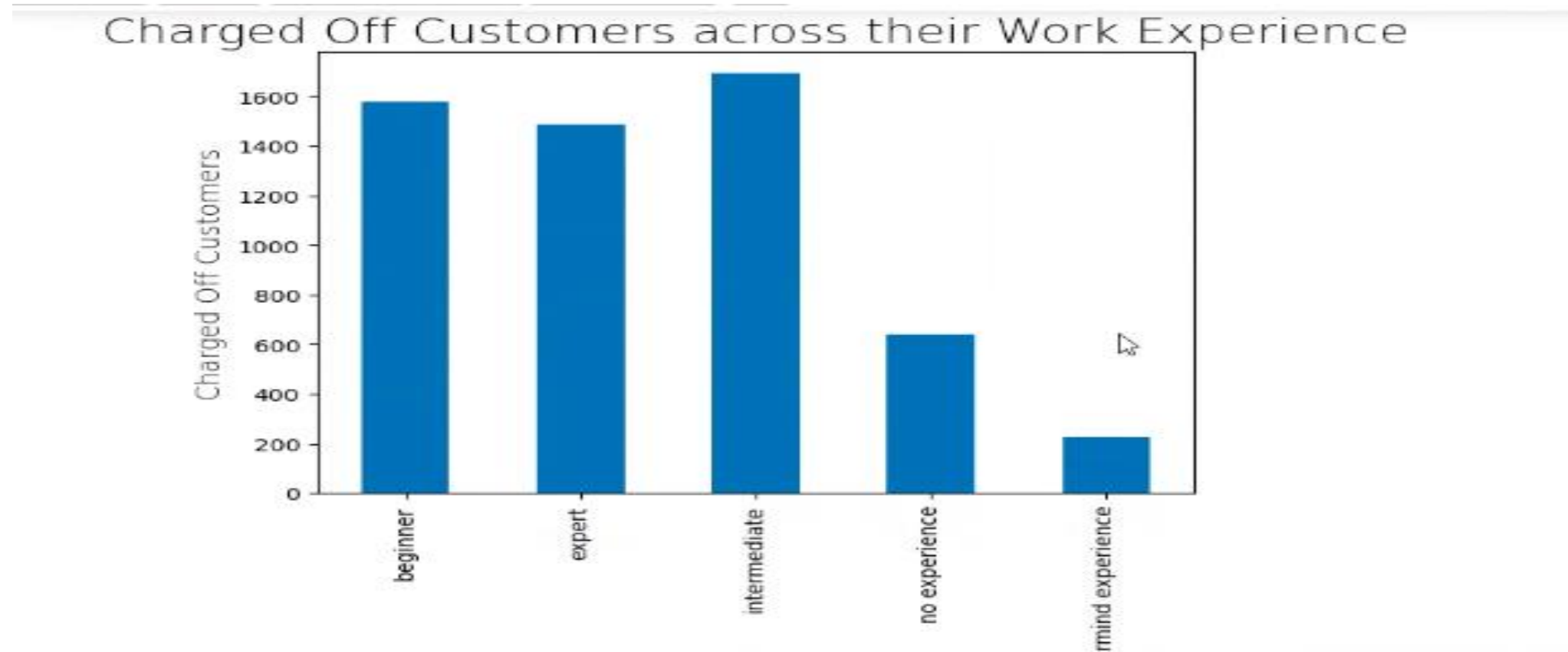
- Bivariate analysis has been conducted on the Target Variable of “Loan Status” = ‘Charged off’ for the below columns to identify the pattern for probable defaulting
 - ✓ Purpose
 - ✓ Grade
 - ✓ Verification Status
 - ✓ Term
- Based on the univariate analysis conducted we notice that, the highest loans that were charged off were
 - ✓ Related to loans provided for the purpose of “Debt Consolidation”
 - ✓ Related to “B” and “C” Grade
 - ✓ Pertained to a non verified source as the “Verification Status” is “Not Verified”
 - ✓ Pertain to 36 months “Term”
- **Analysis Conclusion:**
 - ✓ It is noted that loan applicants for “Debt Consolidation” purpose have more Charged off loan status. This concludes that there is a more probability of defaulting a loan in the Debt Consolidation category
 - ✓ It is noted that loan applicants for “Grade B” have more Charged off loan status. This concludes that there is a more probability of defaulting a loan in the Grade B category
 - ✓ It is noted that loan applicants for “Not Verified” Annual Salary have more Charged off loan status. This concludes that there is more probability of defaulting a loan in the Not verified category. The probable reasons could be the annual salary has been wrongly indicated by the loan applicant or could have changed(reduced) during the loan period due various reasons which might have led to the defaulting of loan repayment.
 - ✓ It is noted that loan applicants for “36 months” loan period have more Charged off loan status. This concludes that there is a more probability of defaulting a loan in the 36 month term category. The probable reasons could be that the Monthly Installment of repayment amount(EMI) would have been higher than the repayment capacity of the loan applicant. Therefore, checking the monthly salary and assessing the repayment capacity month-wise would be recommended. There could be multiple other parallel loans that an applicant has which could be leading to defaulting. Therefore, checking for other existing loans would be recommended.

Derived Metric- Identification & Approach

- Target column = Emp_length (purpose is to identify the cohort of users who are more likely to default loan repayment)
- Performed String to Integer conversion for Emp_Length
- Replaced null values with value “Undetermined Experience”
- Created a derived column named as “Emp_Exp” showing the categorization of employee_length

Emp_length value range	Emp_Exp Category
$x < 0$	No Experience
$0 < x < 3$	Beginner
$3 < x < 8$	Intermediate
$x > 8$	Expert
Null Values	Un-determined experience

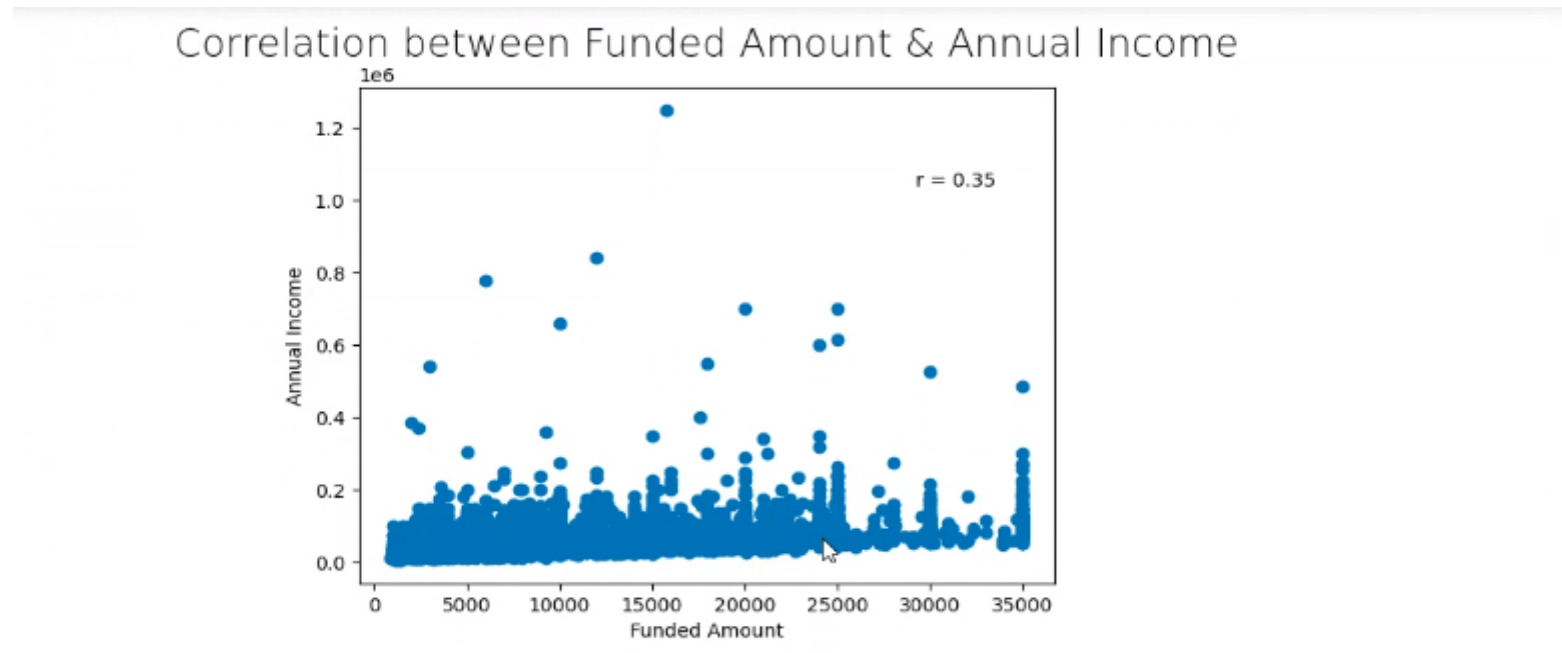
Derived Metric- Analysis



Conclusion: Based on analysis it can be noted that people with “Intermediate” experience level are more likely to default loan repayment.

Correlation Analysis on Numeric variables

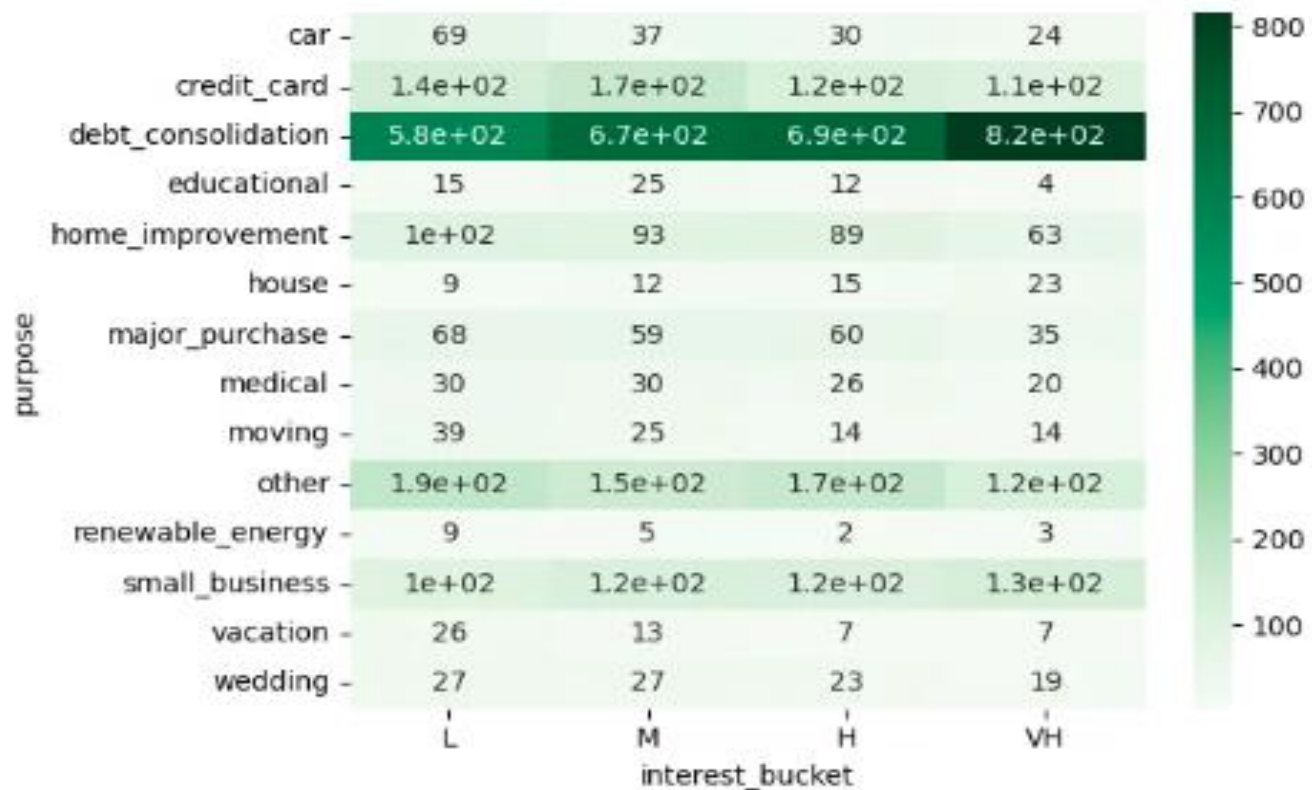
Performed correlation analysis between Funded Amount & Annual Income. The correlation is 0.35 for these variables.



Analysis Outcome: With the above analysis we can conclude that if the Annual Income is more then Funded Amount can be more. This can suggest positive repayment behavior of the loan applicant.

Heat Map Analysis

- Interest Rate has been converted to float from percentage and then categorized into various categories(L – Low, M – Medium, H –High, VH-Very High)
- Stacked the Loan Purpose into these interest categories to determine the category to determine where most default has happened.



Analysis Conclusion – Based on the above we notice that Loans provided for “Debt Consolidation” purpose with a higher interest rate has more defaulting of loan repayment.

Overall Analysis Conclusion

- Based on the **Univariate Analysis** it is observed that maximum defaulting of repayment of loan happens in:
 - ✓ Loans approved for the 36 months term
 - ✓ Loan amount approved is the range of 5000 – 15000
- **Bivariate Analysis Conclusion:**
 - ✓ It is noted that loan applicants for “Debt Consolidation” purpose have more Charged off loan status. This concludes that there is a more probability of defaulting a loan in the Debt Consolidation category
 - ✓ It is noted that loan applicants for “Grade B” have more Charged off loan status. This concludes that there is a more probability of defaulting a loan in the Grade B category
 - ✓ It is noted that loan applicants for “Not Verified” Annual Salary have more Charged off loan status. This concludes that there is more probability of defaulting a loan in the Not verified category. The probable reasons could be the annual salary has been wrongly indicated by the loan applicant or could have changed(reduced) during the loan period due various reasons which might have led to the defaulting of loan repayment.
 - ✓ It is noted that loan applicants for “36 months” loan period have more Charged off loan status. This concludes that there is a more probability of defaulting a loan in the 36 month term category. The probable reasons could be that the Monthly Installment of repayment amount(EMI) would have been higher than the repayment capacity of the loan applicant. Therefore, checking the monthly salary and assessing the repayment capacity month-wise would be recommended. There could be multiple other parallel loans that an applicant has which could be leading to defaulting. Therefore, checking for other existing loans would be recommended.
- Based on **Derived Metric** analysis it can be noted that people with “Intermediate” experience level are more likely to default loan repayment.

Conclusion

➤ Final Recommendation:

- ✓ Company can provide loan based on Annual Salary Verification Status, Purpose, Term and Grade. Verification of Annual Salary, Monthly repaying capacity of loan applicants are crucial in determining the repayment status.
- ✓ Loan applicants having higher annual income are more likely to repay loan for a higher funded amount in comparison to a lower annual income applicant.
- ✓ If “Debt Consolidation” is the purpose of the loan and the interest rate is Very High (>75% of overall distribution of interest rate) then the chances of defaulting a loan repayment is “High”.