

# GRAMENER CASE STUDY

## SUBMISSION

**Group Name:**

- 1. Divij Jawarani**
- 2. Himanshu Singh**
- 3. Roushan Anand**
- 4. Dinesh Challa**

Lenders club would like to identify the driving factors behind loan defaults and we will be analyzing the borrower data from 2007 to 2011 to identify the risk factors.

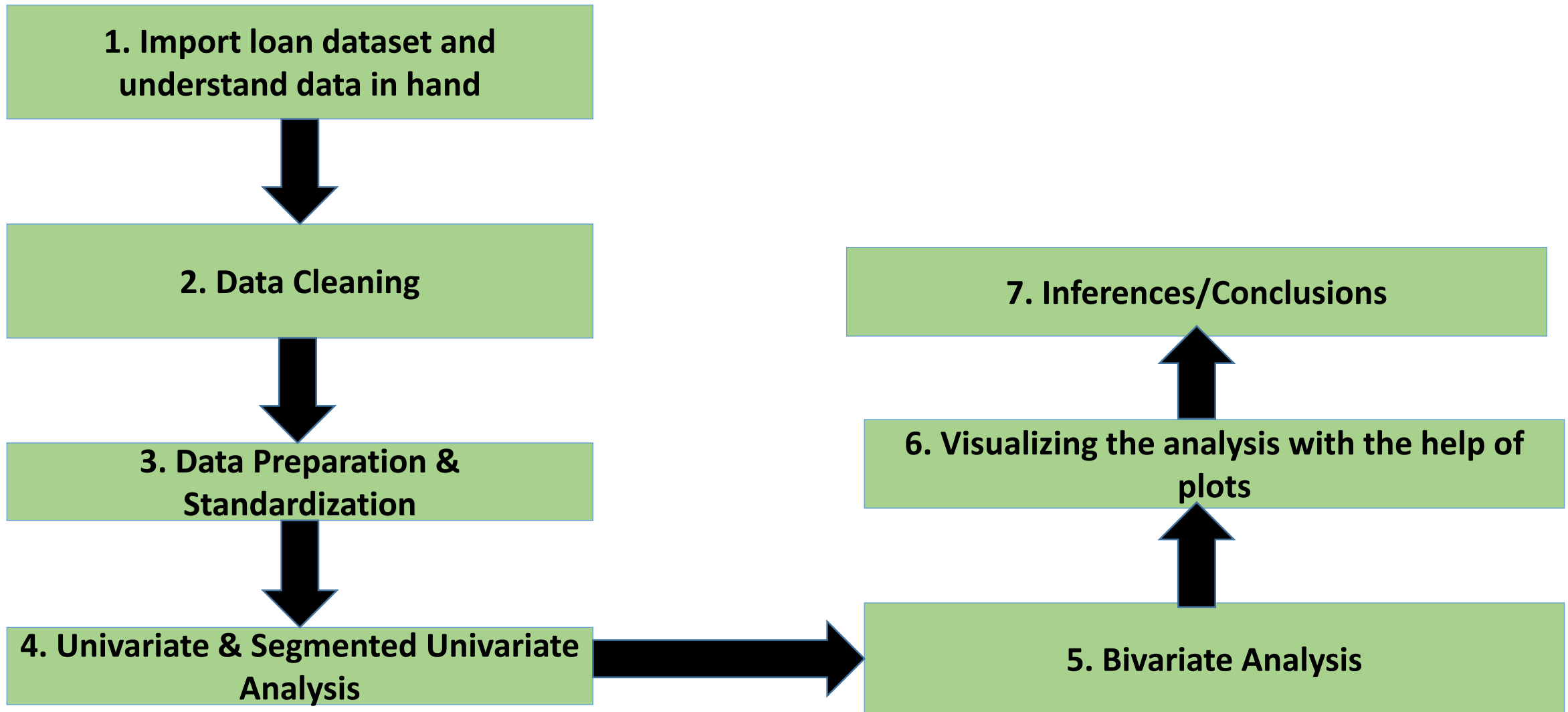
Two types of risks are associated with bank's decision:

- If the applicant is **likely to repay the loan**, then not approving the loan results in a **loss of business** to the company. Identify the combination of risk variables that the bank needs to cautions about.
- 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.

**Business objective:** The objective is to identify the best sectors, countries, and a suitable investment type for making investments.

**Goals of data analysis:** The goals are divided into three main sub-goals

- Analyse the risk variables to identify the possible defaulters.
- Identify the combination of risk variables that the bank needs to cautions about.



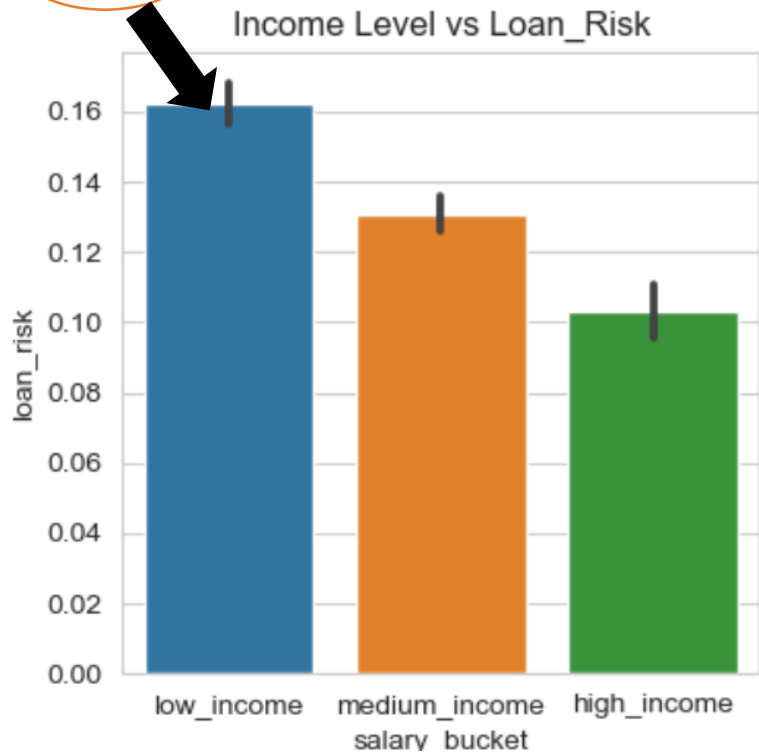
As a part of the univariate & segmented univariate analysis the following variables have been analyzed:

- Annual Income
- Funded Amount
- Interest Rate
- Employee Experience
- Grades
- Home Ownership
- Credit Utilization
- Debt to Income Ration (DTI)
- Purpose
- Verification Status
- Open Credit Lines

After analyzing the variables we have identified 7 important variables that would help us identify the loan defaulters

- Annual Income
- Funded Amount
- Interest Rate
- Credit Utilization
- Debt to Income Ration (DTI)
- Verification Status
- Open Credit Lines

16.23%

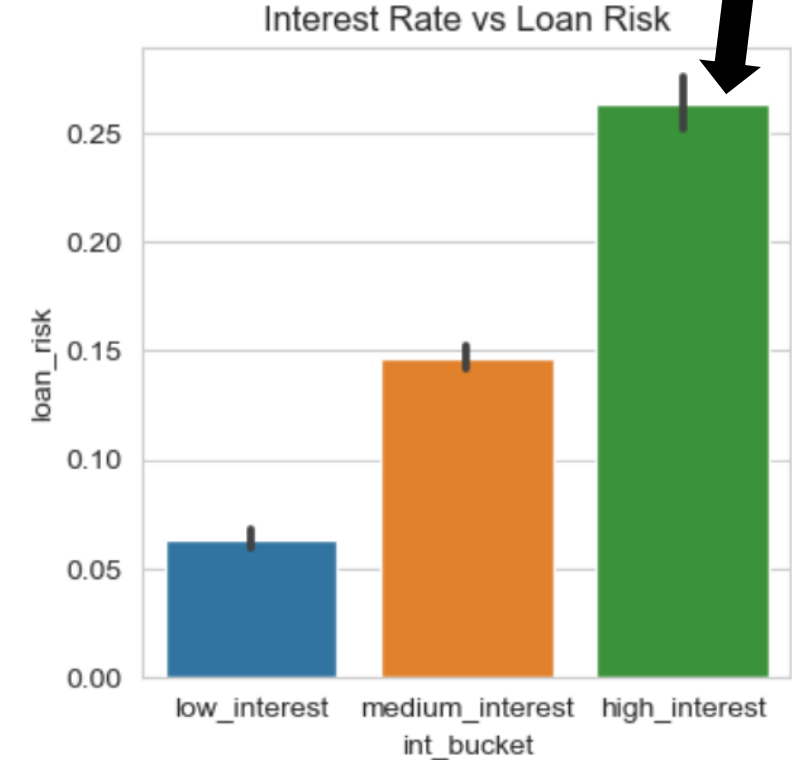


- Low income group has the highest default rate among the lot.

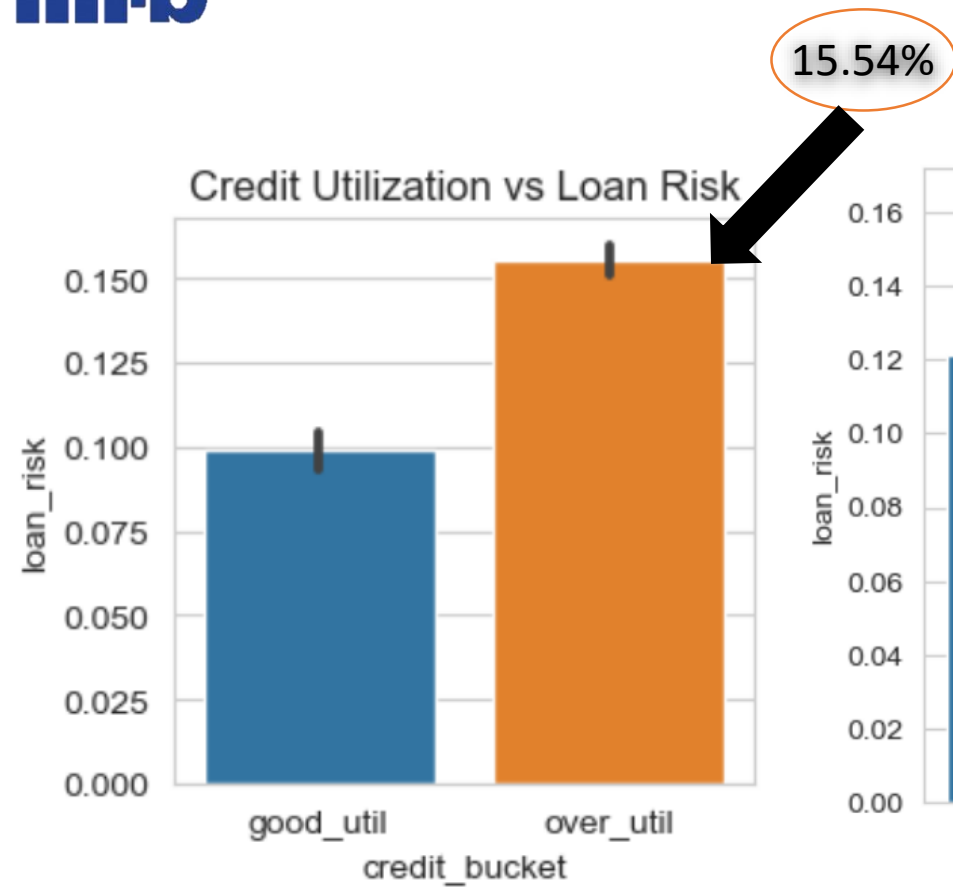
26.38%



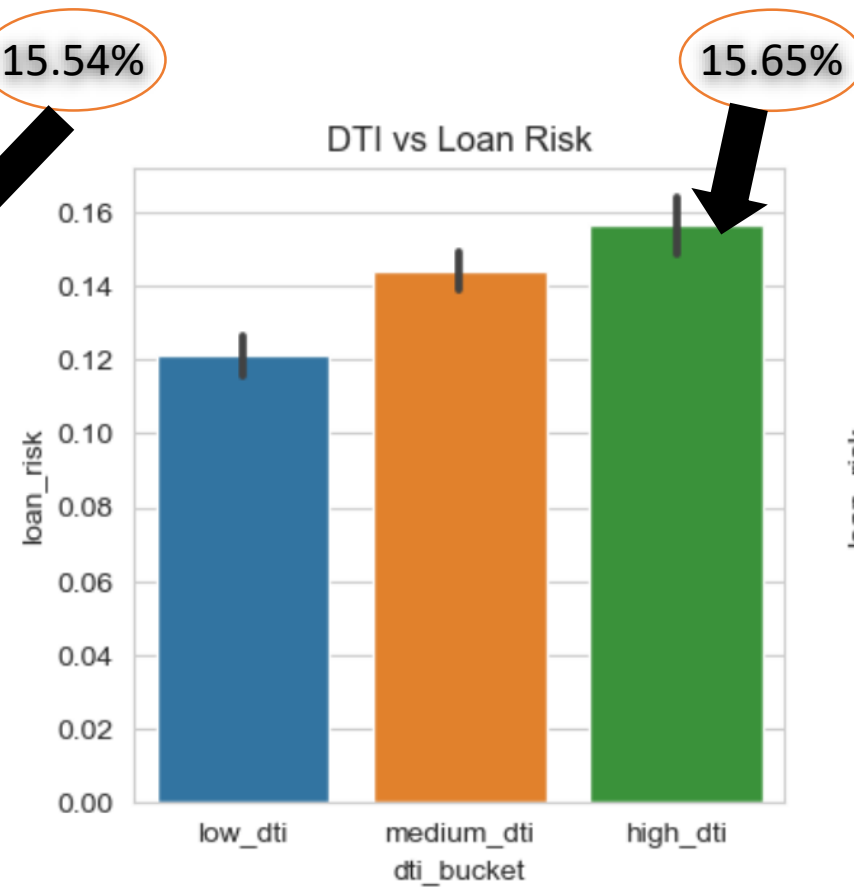
- Big loans have high default rate among the lot.



- Loans with high interest have high default rates among the lot.



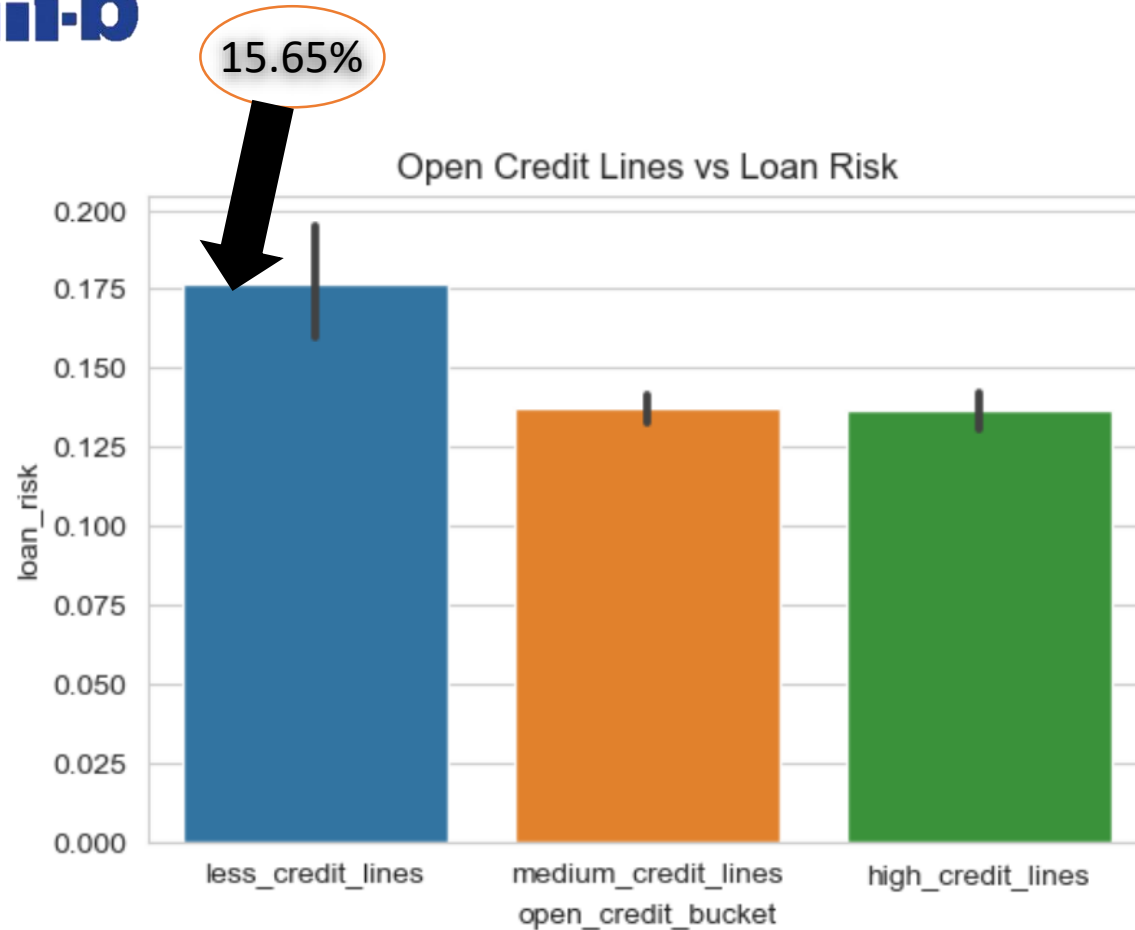
- Borrowers with high credit utilization have more default rate.



- Borrowers with high debt to income ration have high default rates among the lot.



- Interestingly, verified & source verified loans are defaulting the most.



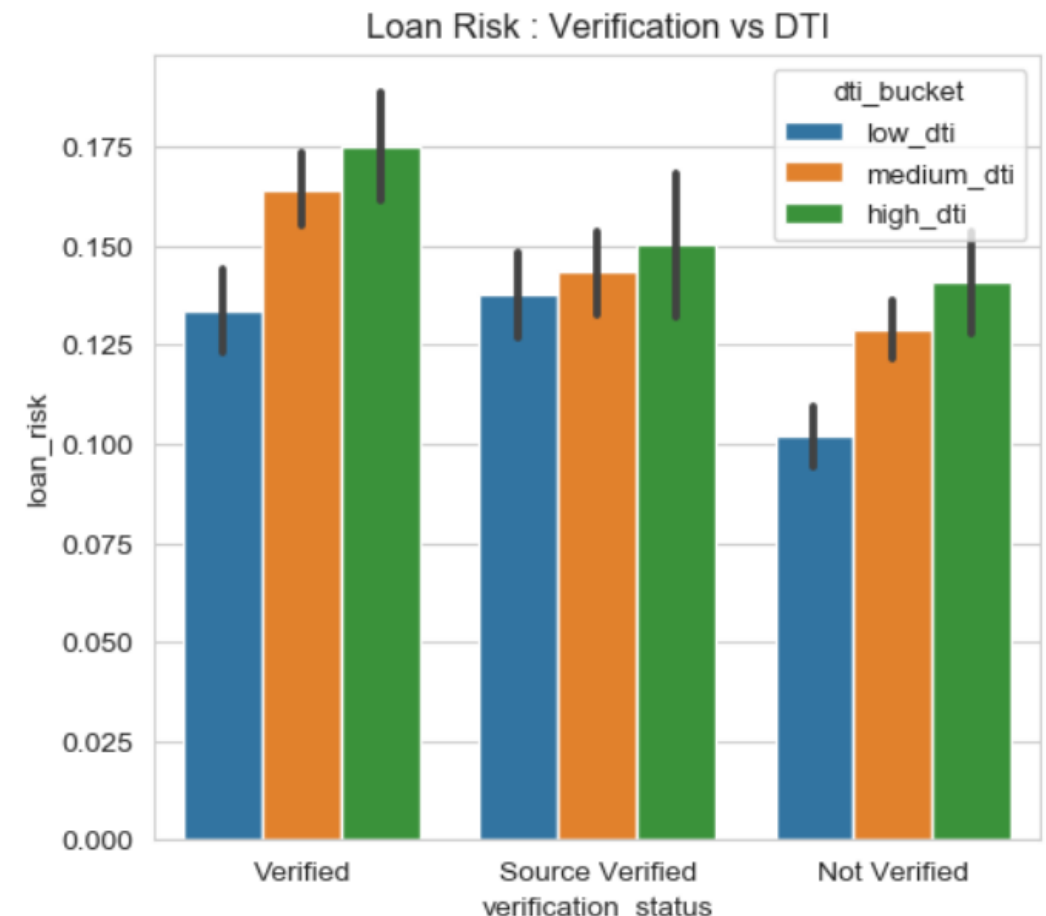
- Borrowers who have less credit lines have high default rate among the lot.

We have now identified 7 driving variables that would help us reduce the risk of loan default. We will now check for different combinations of these variables to check for insights as a part of the bivariate analysis.

The 7 driving variables that were identified in the previous step are now analyzed against one another to observe any useful insights. The following combinations have been analyzed:

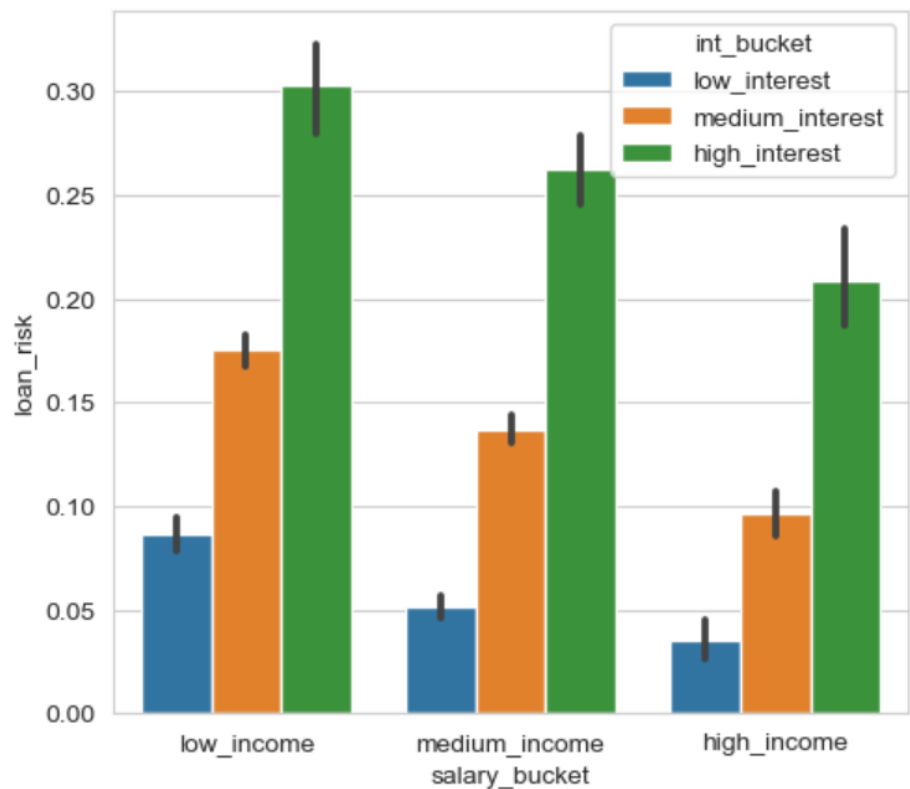
- Verification vs DTI
- Income vs Interest Rate
- Income vs Loan Size
- Credit Utilization vs Verification
- Open Credit Lines vs Credit Utilization

Average of loan_risk	Column Labels			
Row Labels	high_dti	medium_dti	low_dti	Grand Total
Verified	17.5%	16.4%	13.4%	15.8%
Source Verified	15.0%	14.3%	13.8%	14.2%
Not Verified	14.1%	12.9%	10.2%	12.2%
Grand Total	15.7%	14.4%	12.1%	13.9%



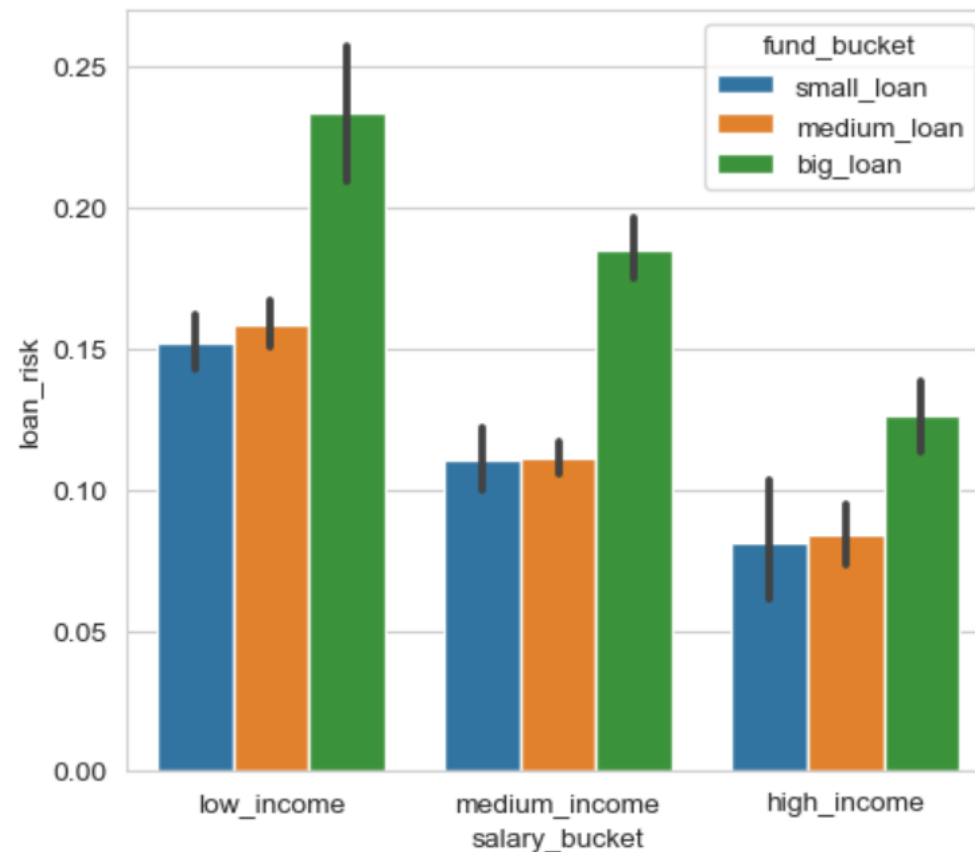


Loan Risk : Income vs Interest Rate



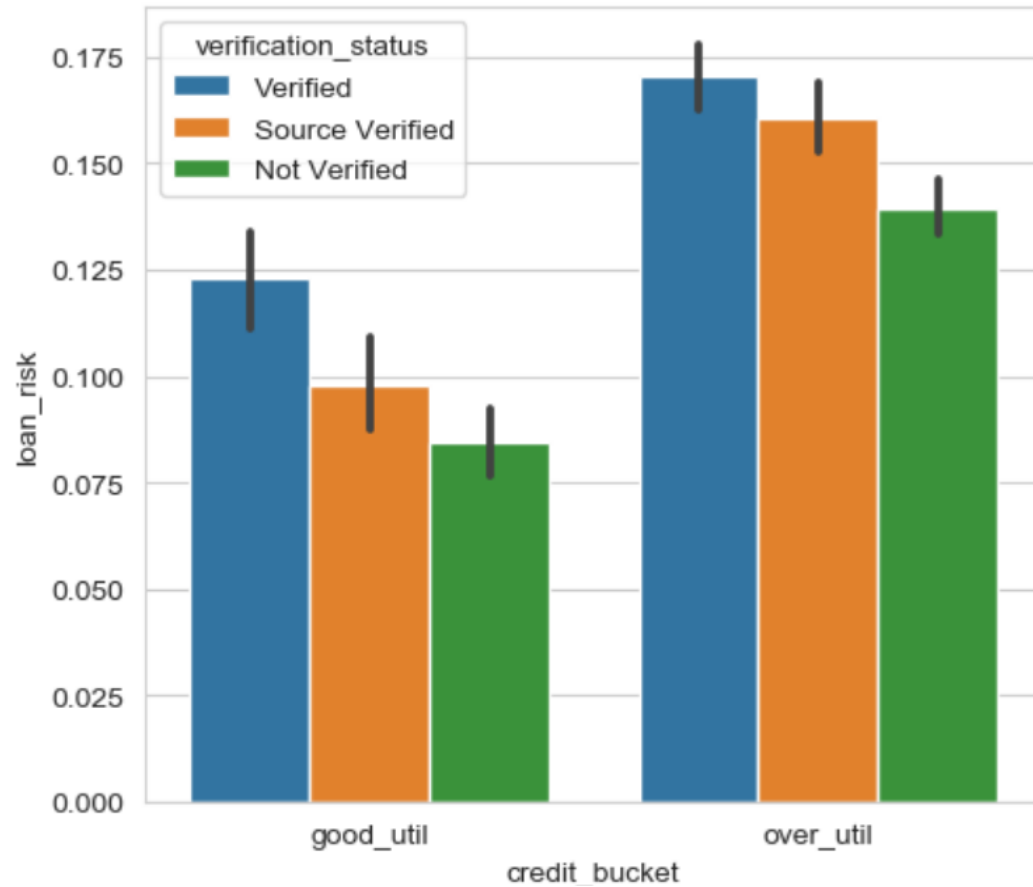
Average of loan_risk				
Column Labels				
Row Labels	high_interest	medium_interest	low_interest	Grand Total
low_income	30.3%	17.5%	8.7%	16.2%
medium_income	26.3%	13.7%	5.2%	13.1%
high_income	20.9%	9.6%	3.6%	10.3%
Grand Total	26.4%	14.7%	6.3%	13.9%

Loan Risk : Income vs Loan Size



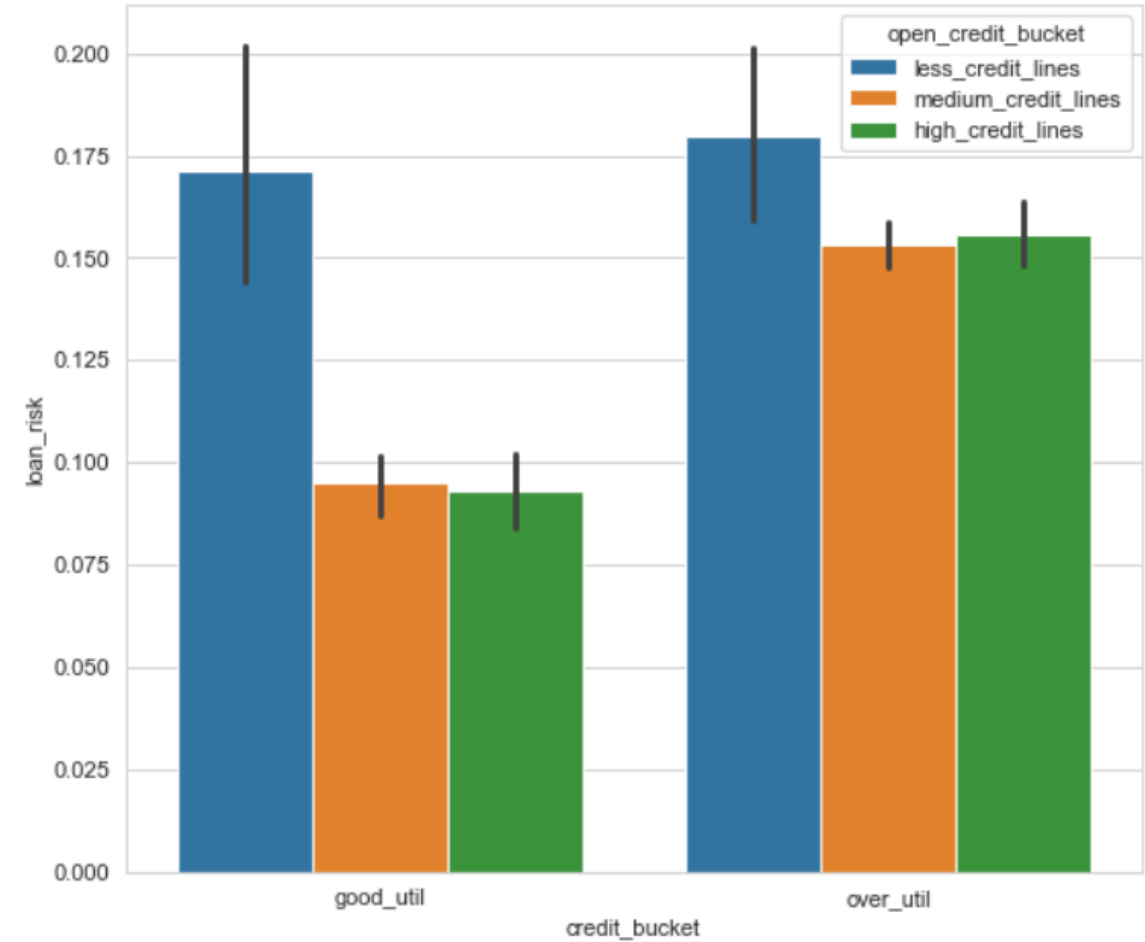
Average of loan_risk				
Column Labels				
Row Labels	big_loan	small_loan	medium_loan	Grand Total
low_income	23.4%	15.2%	15.9%	16.2%
medium_income	18.5%	11.0%	11.1%	13.1%
high_income	12.6%	8.1%	8.4%	10.3%
Grand Total	17.4%	13.3%	12.7%	13.9%

Loan Risk : Credit Utilization vs Verification



Average of loan_risk Column Labels				
Row Labels	Verified	Source Verified	Not Verified	Grand Total
over_util	17.0%	16.1%	14.0%	15.5%
good_util	12.3%	9.8%	8.4%	9.9%
Grand Total	15.8%	14.2%	12.2%	13.9%

Loan Risk : Open Credit lines vs Credit Utilization



Average of loan_risk Column Labels				
Row Labels	less_credit_lines	high_credit_line	medium_credit_line	Grand Total
over_util	18.0%	15.7%	15.2%	15.5%
good_util	17.1%	9.3%	9.5%	9.9%
Grand Total	17.7%	13.8%	13.6%	13.9%

We would like to make the following suggestions to Lenders Club based on the bivariate analysis:

- ❖ **Verification Status vs DTI:** It is recommended that LC regulate their loans to those borrowers having high and medium dti irrespective of whether a loan is verified or not to reduce the risk of default.
- ❖ **Income vs Interest Rate:** We would recommend LC that there is a high possibility of defaulting a high interest loan among any given salary bucket. So, LC should review their higher interest rates to reduce the number of defaulters.
- ❖ **Income vs Loan Size:** : We would recommend LC **NOT** to fund big loans for low income & middle income groups.
- ❖ **Credit Utilization vs Verification:** We would recommend LC **NOT** to give loans to borrowers with high credit utilization whether a loan is verified or not.
- ❖ **Open Credit Lines vs Credit Utilization:**
  - Though the borrower is having good credit utilization(<30%), we recommend **NOT** to give loans if the open credit line are less.
  - We also recommend **NOT** to give loans to those borrowers who have high credit utilization irrespective of the number of open credit line.