INVESTMENT CASE STUDY SUBMISSION

Group Name: ML-AI C6 Batch

1. **Aniket Sahoo**
2. **Khushboo Singh**
3. **Puneet Hasija**
4. **Vishwas P**

Business Requirement

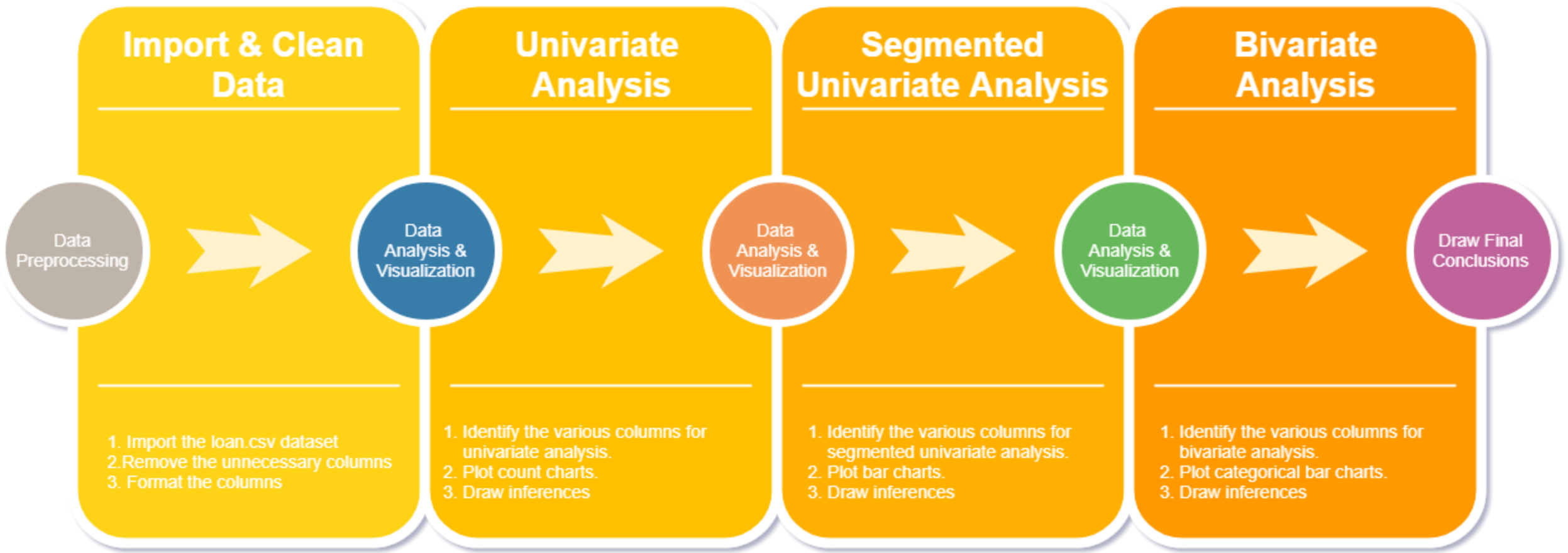
Lending Club is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures. Borrowers can easily access lower interest rate loans through a fast online interface. Lending loans to 'risky' applicants is the largest source of financial loss.

If one is able to identify these risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss. Identification of such applicants using EDA is the aim of this case study

Goals

- The aim is to identify patterns which indicate if a person is likely to default, which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.
- To understand how consumer attributes and loan attributes influence the tendency of default.
- To understand the driving factors (or driver variables) behind loan default.

Methodology of Analysis



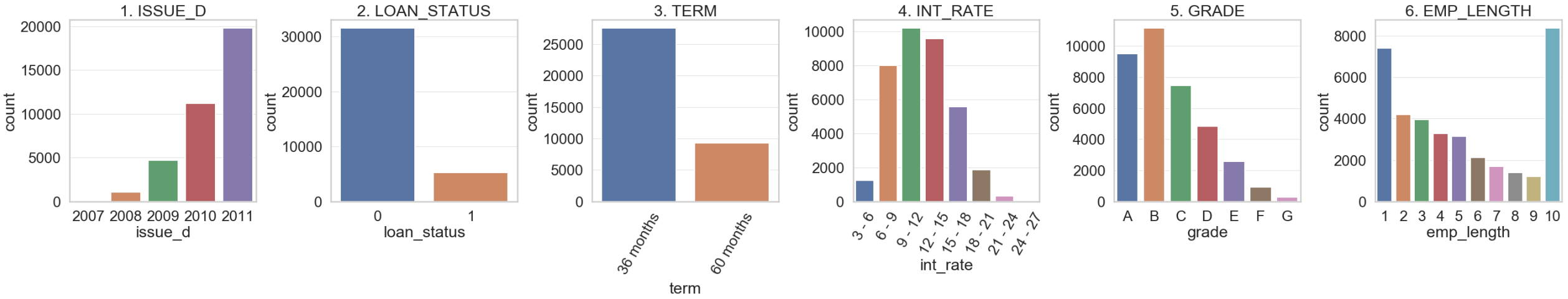
Data Analysis

As per the objective of the problem one needs to analyse the consumer attributes and loan attributes that influence the tendency of default. So, one should consider the details available prior to the loan approval i.e the consumer attributes and loan attributes only, for the analysis. Thus, the columns providing post-loan details i.e loan repayment attributes were not used.

Based on the descriptions provided for each column in the Data Dictionary and some research on the domain along with analysis of the data, the below columns were only used for analysis.

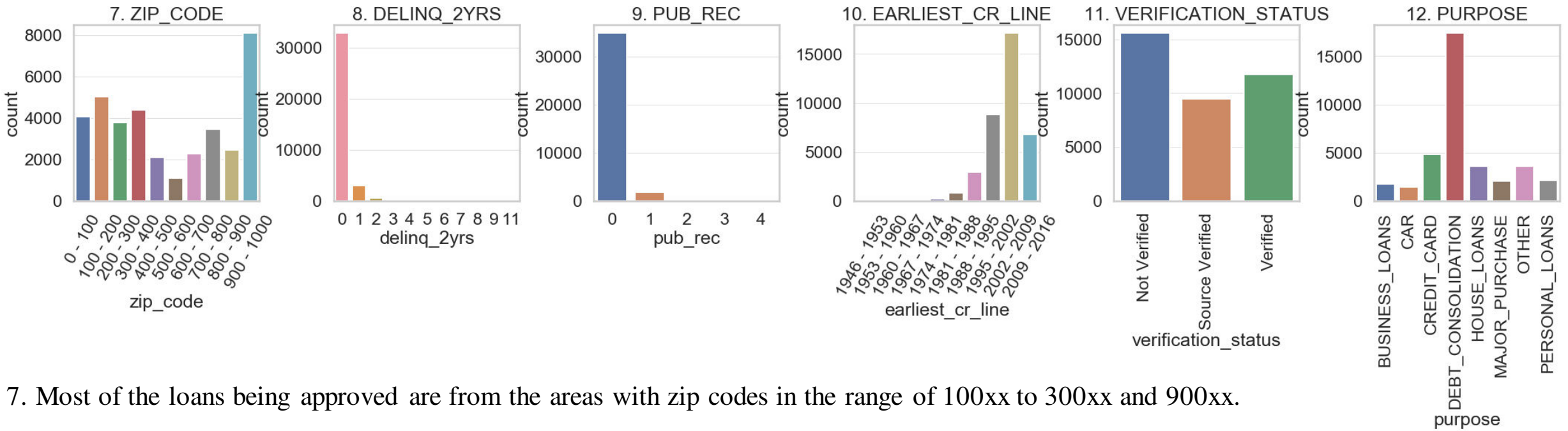
Loan Amount	Home Ownership	Debt to Income Ratio
Loan Term	Annual Income	Delinquency in past 2 years
Interest Rate	Income Verification Status	Earliest Credit Line
Instalment Amount	Loan Issue Date	Inquiry in last 6 months
Grade	Loan Status	Number of Open Accounts
Sub-Grade	Loan Purpose	Public Derogatory Records
Employment Length	Zip Code	Revolving Utilization Percentage

Univariate Analysis



1. The number of loans being funded has been sky rocketing starting with just 7 in 2007 to 4699 in 2009 to 19794 loans in the year 2011.
2. The overall default rate is 14% (5000 loans) out of the total of 36800 loans approved
3. The number of loans approved for short term (36 months) is almost thrice that of the long term (60 months)
4. The rate of interest for most of the loans approved lies between 6% to 18%
5. A major chunk of the loans approved belong to Grades A, B, C & D
6. People with employment length of 1 year or less and 10 years or more have the highest number of approvals for the loans.

Univariate Analysis



7. Most of the loans being approved are from the areas with zip codes in the range of 100xx to 300xx and 900xx.

8. Almost all of the loans being approved do not have any past incidences of delinquency in the last 2 years

9. Almost all of the loans approved do not have any derogatory public records to their name.

10. Most of the loans approved had their credit lines starting around the 2000's.

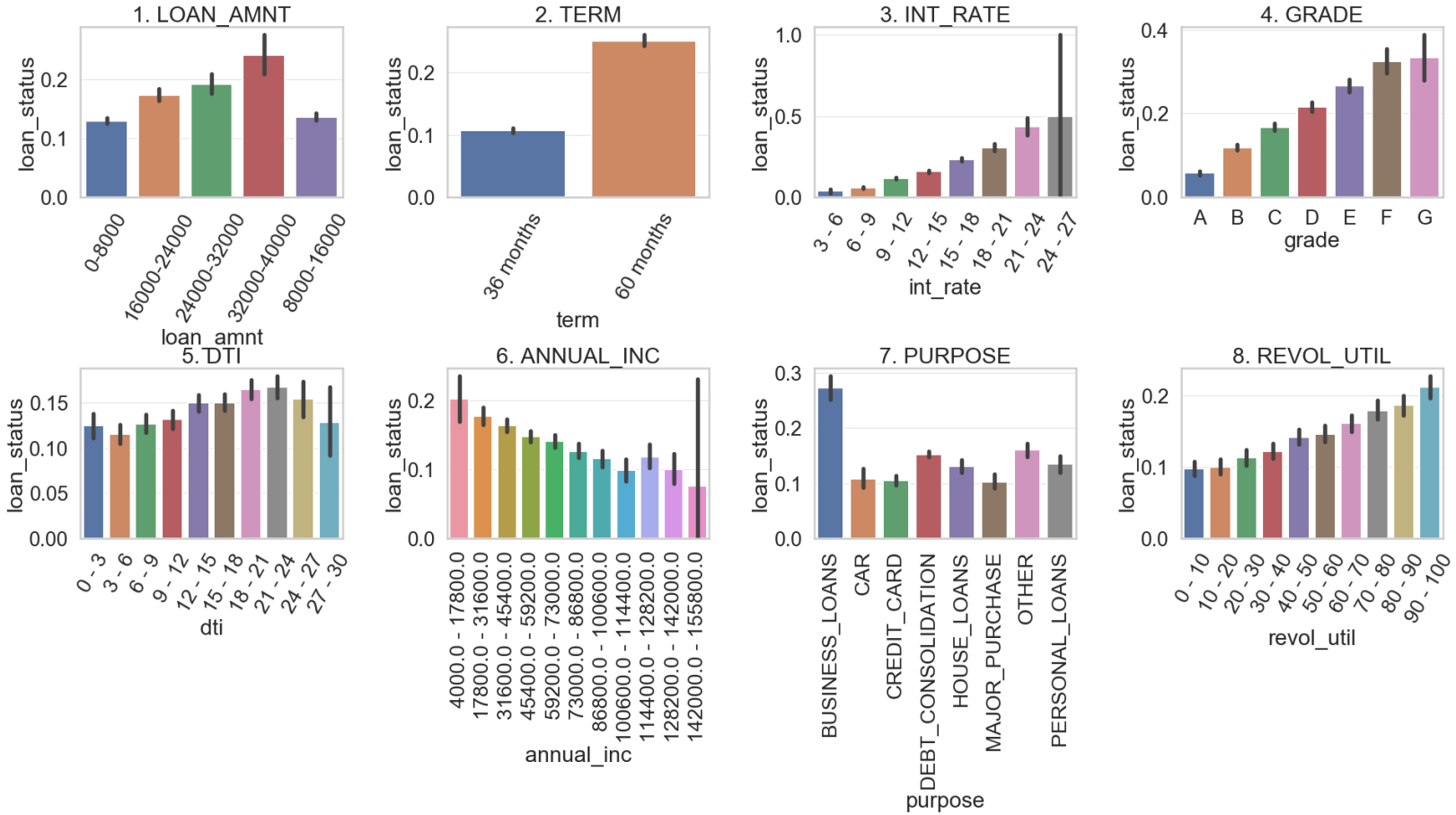
11. Almost 40% of the loans approved do not have their income sources verified

12. Almost 50% of the loans being approved is being used for Debt Consolidations and Credit Card payment.

Segmented Univariate Analysis

Driving factors of Default Rate

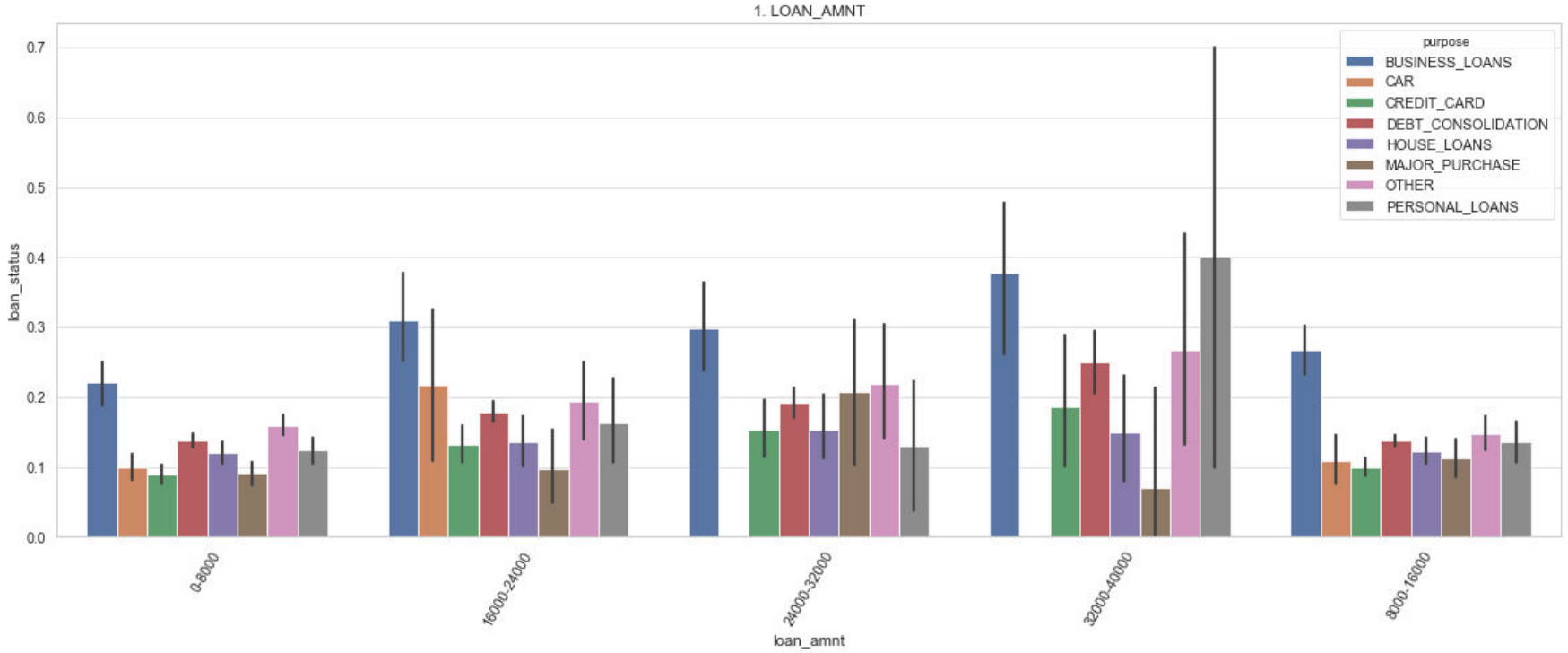
1. Loan Amount (Directly Proportional)
2. Loan Term (Directly proportional)
3. Interest Rate (Directly Proportional)
4. Grade (Directly Proportional)
5. Debt to Income Ratio
6. Annual Income (Inversely Proportional)
7. Loan Purpose
8. Revolving Line of Utilization Rate (Directly Proportional)



Bivariate Analysis

The loans on which borrowers tend to default the most are given below.

- Business Loans (28%)
- Other Loans (16%)
- Debt Consolidation (15%)
- House Loans (13%)
- Personal Loans (13%)
- Car Loans (11%)
- Credit Cards (11%)
- Major Puchase (10%)



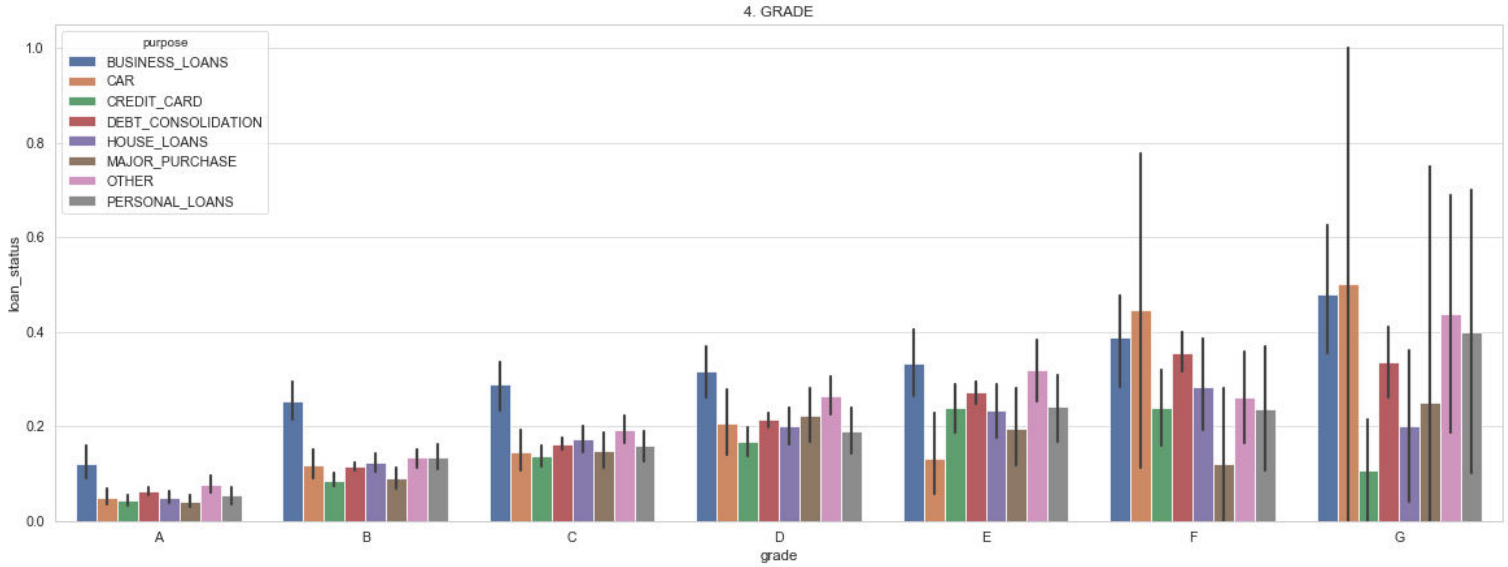
1. Borrowers with a loan amounts of 32,000 to 40,000 as Personal loans tend to default by almost 40% as compared to around 15% for those in other loan amount brackets, suggesting people with high amount of Personal loans tend to default a lot.
2. The default rate for Car loans is pretty high almost 22% for borrowers with loan amounts of 16,000 to 24,000 as compared to almost 10% in other loan amount brackets, suggesting that people with higher loan amounts as Car loans default twice as much as other amounts.
3. Nobody in the loan amounts of 24,000 to 40,000 has defaulted for Car loans, suggesting that the people avail for car loans mostly below 24,000.

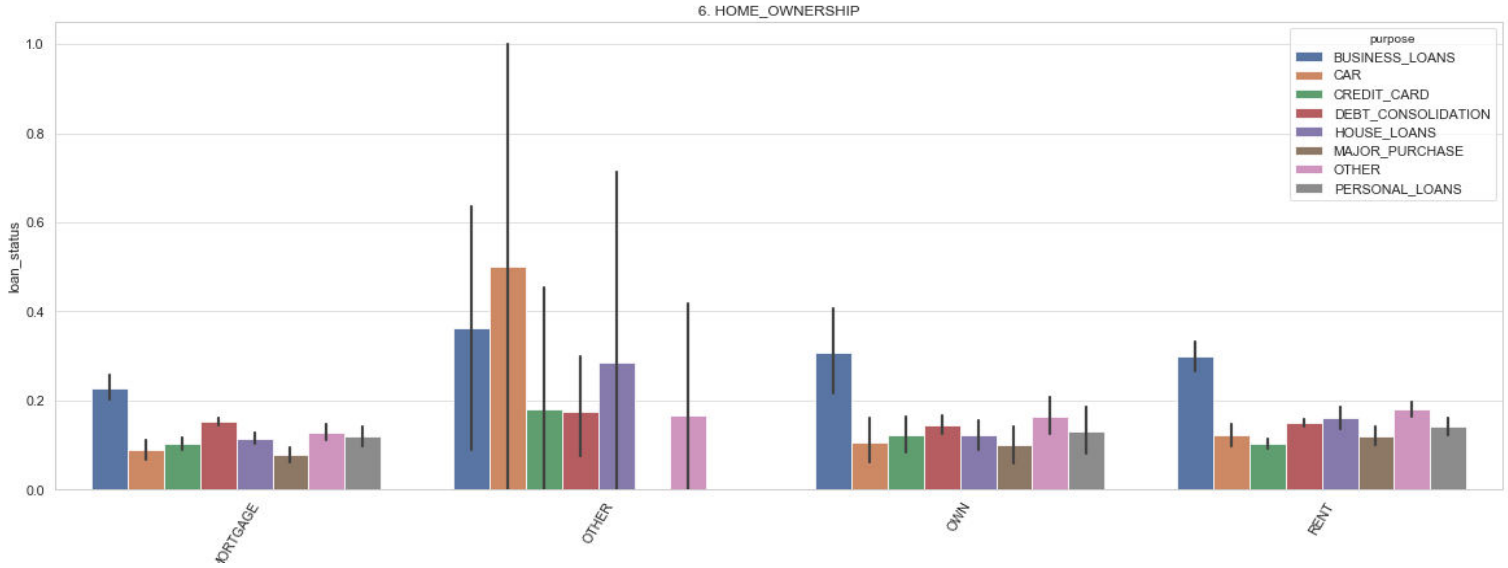
Bivariate Analysis

4. Borrowers belonging to grades F & G having a Car loan tend to default by almost 30% more than those belonging to grades A, B, C or D, suggesting Car loans for these grades should be scrutinized more.

5. The personal loan in the grade G is almost 20% more than that in other grades, suggesting Personal loans are a bad investment in this grade.

3. Borrowers having Car loans with no information regarding their home ownership default by almost 50% as compared to 10% for those with information, suggesting these loans should be avoided. Also there are no defaults in the Major Purchases for these borrowers, suggesting these borrowers may not have any permanent dwelling.





Conclusion

1. The default rate is pretty high almost 25% for long term loans as compared to 10% for short term loans. So, long term loans should be avoided.

2. The rate of default increases steadily with the increase in the rate of interest starting from almost 2% for interest rates of 3% to almost 50% for interest rates of 24% and above. So the rate of interest should be kept around 12%-15%.

3. The default rate keeps increasing significantly with the grades starting with 6% for grade A to 32% for grade G. Thus, for loans falling in higher grades the loan amounts be decreased.

4. The default rate is high for those borrowers whose house ownership has not been identified. So, these loans should be avoided as these are riskier.

5. The rate of default decreases steadily from almost 22% for applicants having annual income of 15,000 or less to 5% for applicants with annual incomes of 1,19,000 or more. Loan amounts should be decreased for low income borrowers.

6. There seems to be a loophole in the process of verification of source of income as the default rate for unverified income sources have a lower default rate almost 12.5% as compared to 16.5% for income source verified by LC.

7. Focus should be given to the business loans as these register an alarming rate of default of 28% as compared to 15% for other loan types.

8. Loans should be avoided or amounts decreased for borrowers with more number of inquiries for loan in the past 6 months, borrowers with delinquency in the past 2 years, borrowers with a derogatory public records or borrowers with very high number of open accounts.

Assumptions

1. All the analysis has been done after having removed the loan records with status currently active or in progress.

2. The Loan amounts have been grouped into various brackets of 10,000 each.

3. The purpose of loans have been grouped into various categories as given below
   - DEBT_CONSOLIDATION
   - CREDIT_CARD
   - OTHER
   - MAJOR_PURCHASE
   - BUSINESS_LOANS consisting of SMALL_BUSINESS, RENEWABLE_ENERGY
   - CAR
   - HOUSE_LOANS consisting of HOME_IMPROVEMENT, HOUSE, MOVING
   - PERSONAL_LOANS consisting of WEDDING, MEDICAL, VACATION, EDUCATIONAL

4. All the other numerical variables with unique elements more than 20 have been bracketed into 5 or 10 bins depending upon the type of analysis.