Lending Club Case Study

Andrea and Shagun

Background

Lending club: An online loan marketplace facilitating personal loans, business loans, and financing of medical procedures.

Two types of decisions that could be taken by the company during processing a new loan application:

1. Loan accepted: If the company approves the loan, there are 3 possible scenarios described below

- **Fully paid**: Applicant has fully paid the loan (the principal and the interest rate)
- **Current**: Applicant is in the process of paying the instalments, i.e. the tenure of the loan is not yet completed. These candidates are not labelled as 'defaulted'.
- **Charged-off**: Applicant has not paid the instalments in due time for a long period of time, i.e. he/she has **defaulted** on the loan

2. Loan Rejected*:* If the company does not approve the loan. Based on the right decision taken i.e. whether to accept or reject the loan application, company makes profit.

Business Objectives

To understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default. The company can utilize this knowledge for its portfolio and risk assessment.

Dictionary of Driving factors

1. **loan_amt :** loan amount requested by the borrower

2. **funded_amnt :** The total amount committed to that loan at that point in time by the Lending Club.

3. **funded_amnt_inv :** The total amount committed by investors for that loan at that point in time.

4. **term :** term of the loan

5. **int_rate :** interest rate of the loan

6. **grade :** grade of loan from A-G in increasing level of risk

7. **sub-grade:** sub-grade of grades from 1-5 in increasing level of risk

8. **emp_length :** Employment length in years

9. **home_ownership :** The home ownership status provided by the borrower during registration

10. **annual_inc:** The self-reported annual income provided by the borrower during registration.

11. **issue_d :** The month which the loan was funded

12. **loan_status:** Current status of the loan. The possible values are Fully Paid & Charged off.

13. **purpose :** A category provided by the borrower for the loan request.

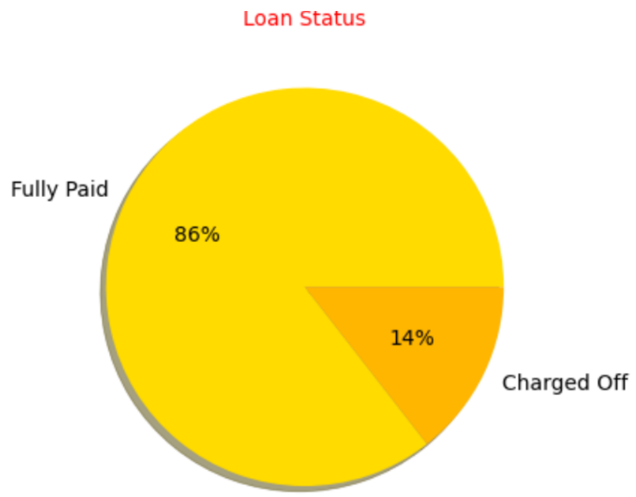
14. **addr_state :** The state provided by the borrower in the loan application

15. **dti:** A ratio calculated using the borrower's total monthly debt payments on the total debt obligations, excluding mortgage and the requested Lending Club loan, divided by the borrower's self-reported monthly income.

Analysis

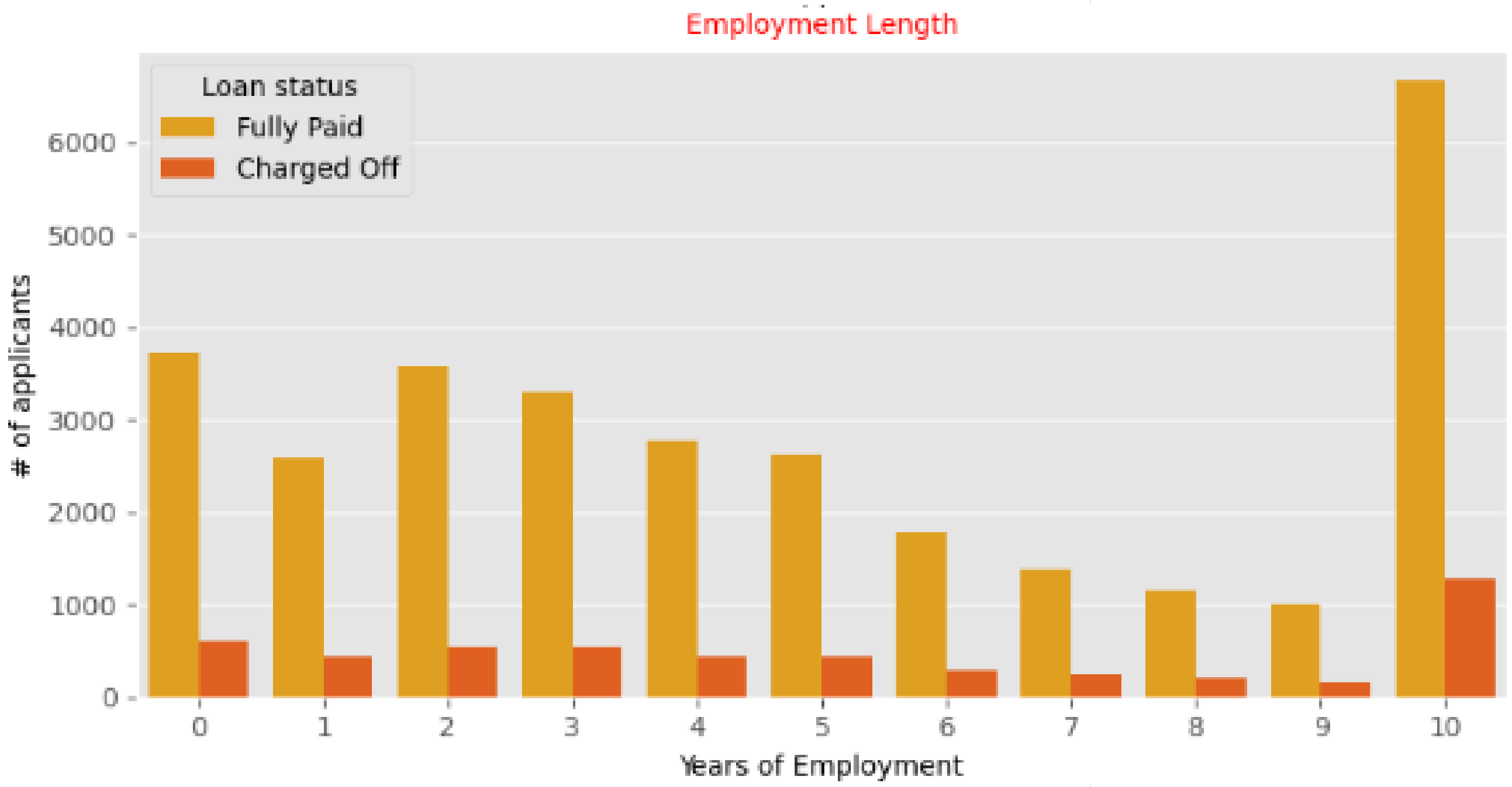
1. Remove columns and rows with high missing values.
2. Remove rows with loan status: current.
3. Remove customer behavioural variables.
4. Remove columns with descriptive text values.
5. Removed invalid rows e.g. rows with funded amount invested =0 and loan status='Charged Off'.
6. Add derived metrics on date and funded amount.
7. Format string fields to int/float values and create buckets to perform segmented analysis
8. Assumption made on Annual Income column based on business derived metric income ranges, removed outliers.

Univariate Analysis



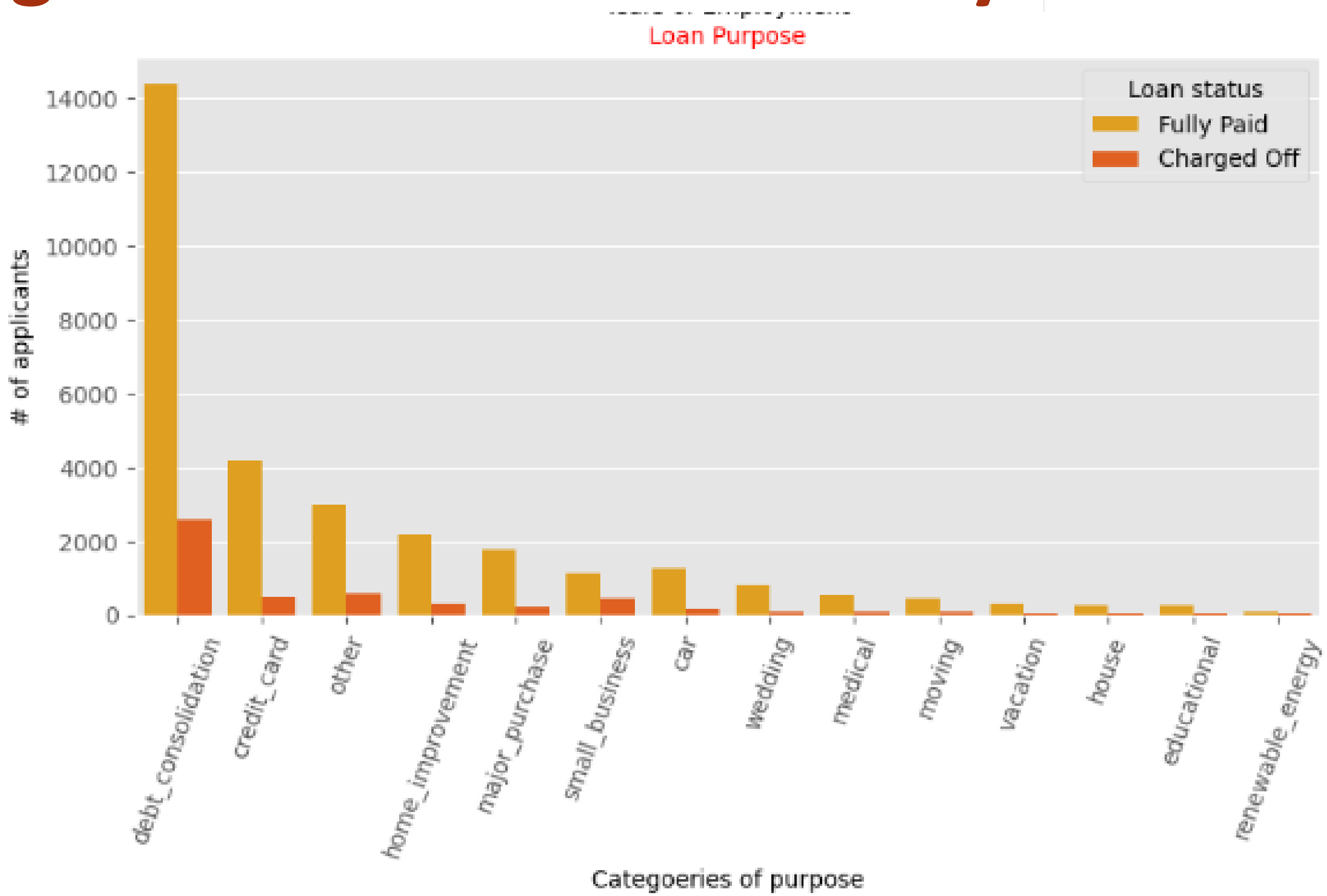
Observation: 85% of the applicants paid the loan amount fully and 15% were defaulters.

Segmented Univariate Analysis



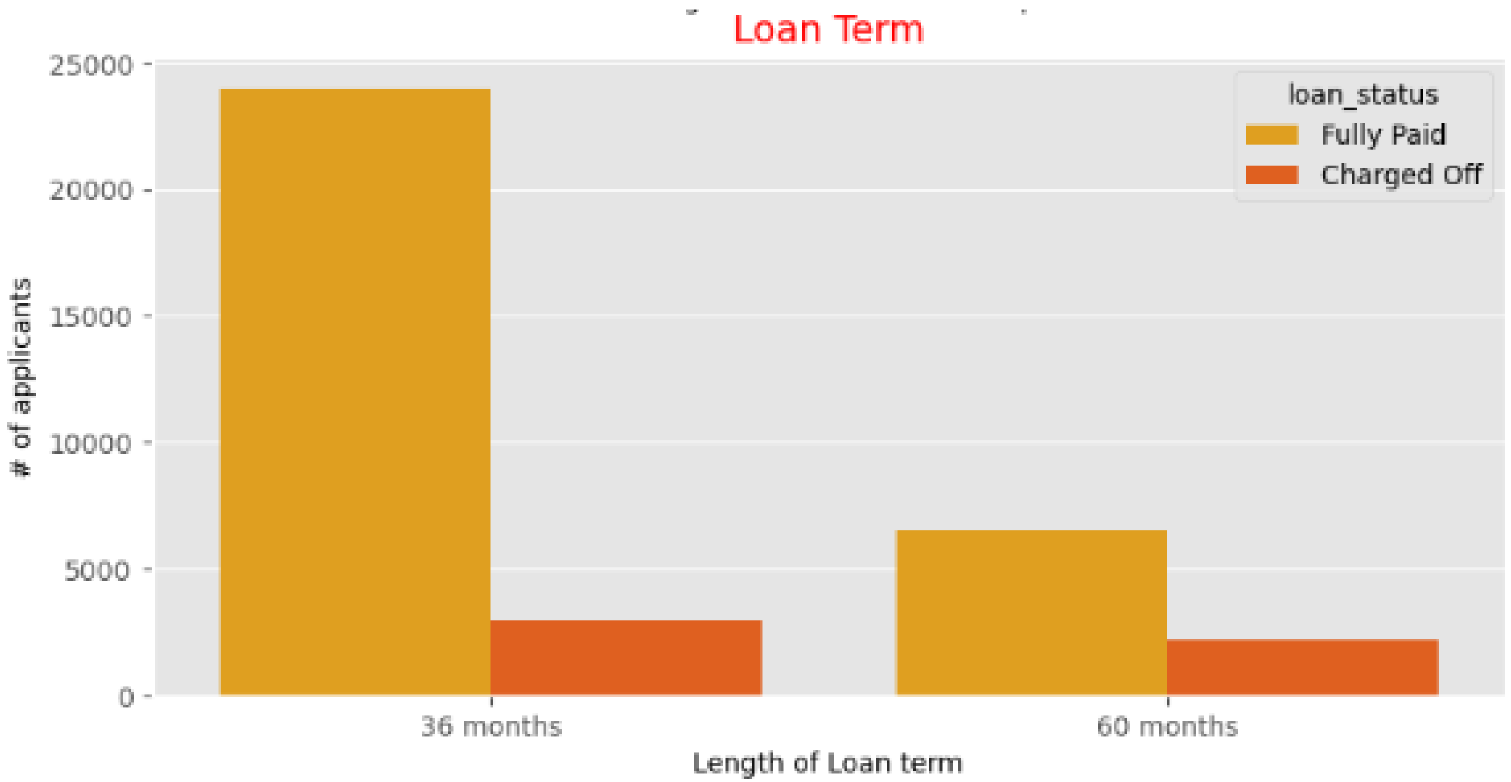
Observation: Maximum charged off loans are for those having experience more than 10 years. Also, it is seen that applicants with more than 10 years of experience also have a high number of fully paid loans.

Segmented Univariate Analysis



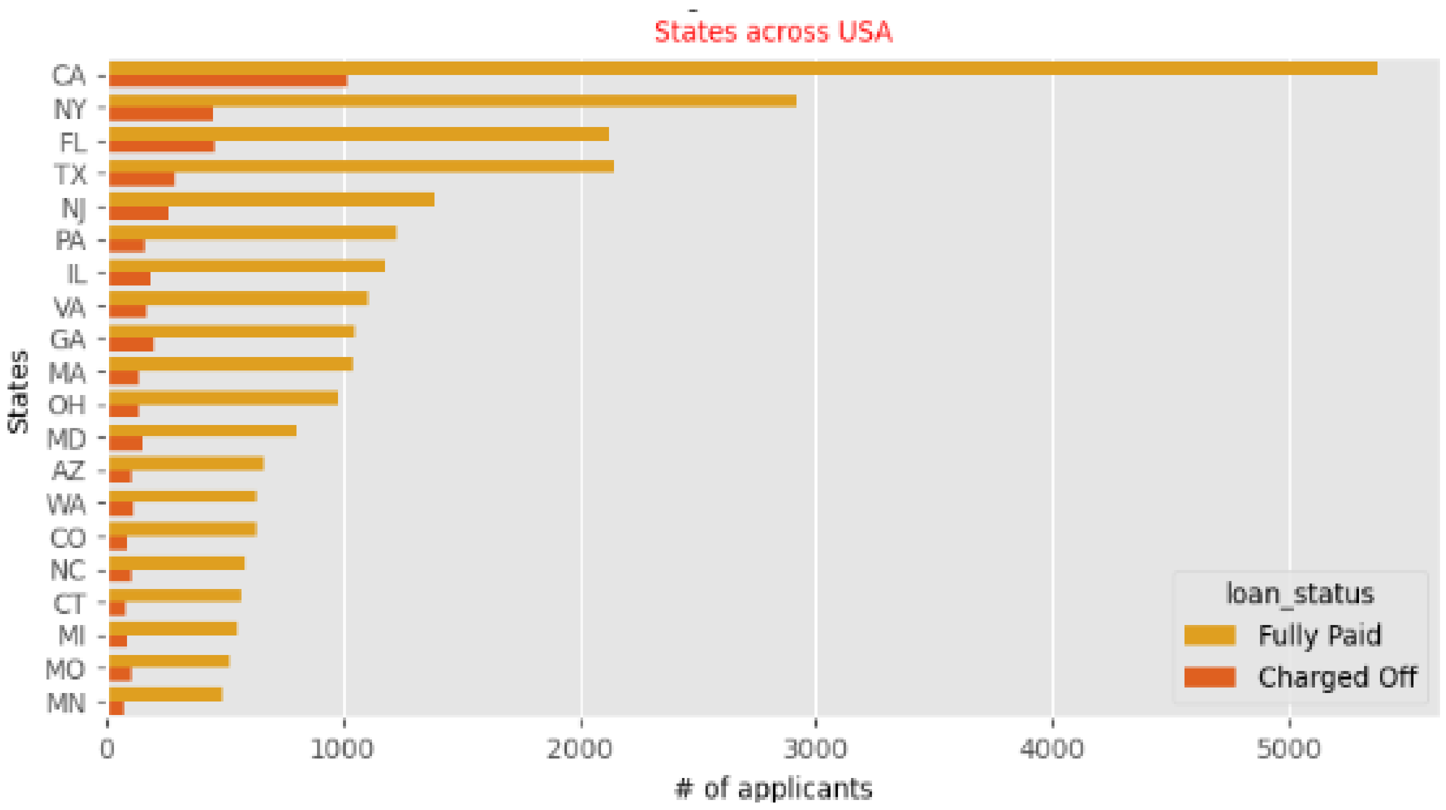
Observation: Majority of loans were taken for the purpose of debt consolidation, then the number is high for credit card loans and loans taken to do home improvements & major purchases.

Segmented Univariate Analysis



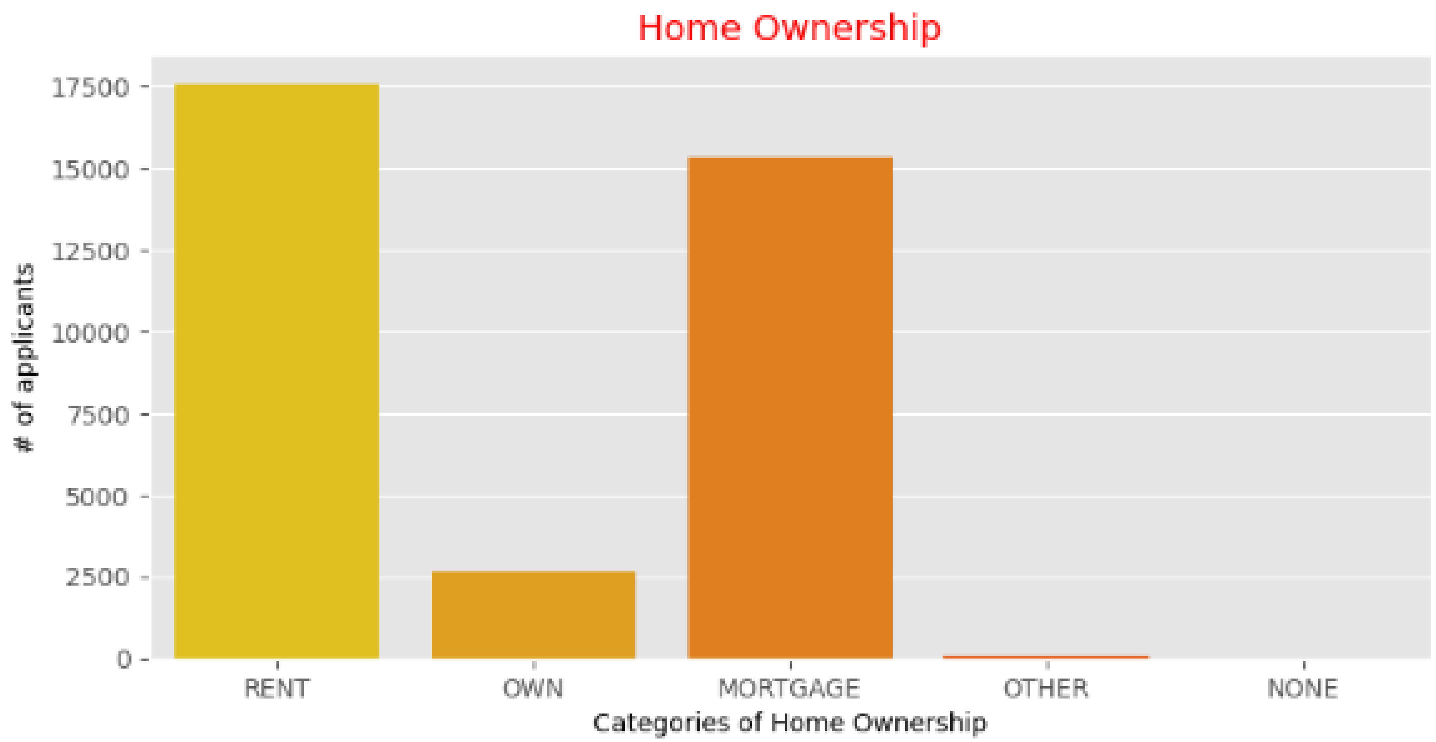
Observation: Majority of the applicants opted for 36 months duration of the loan.

Segmented Univariate Analysis



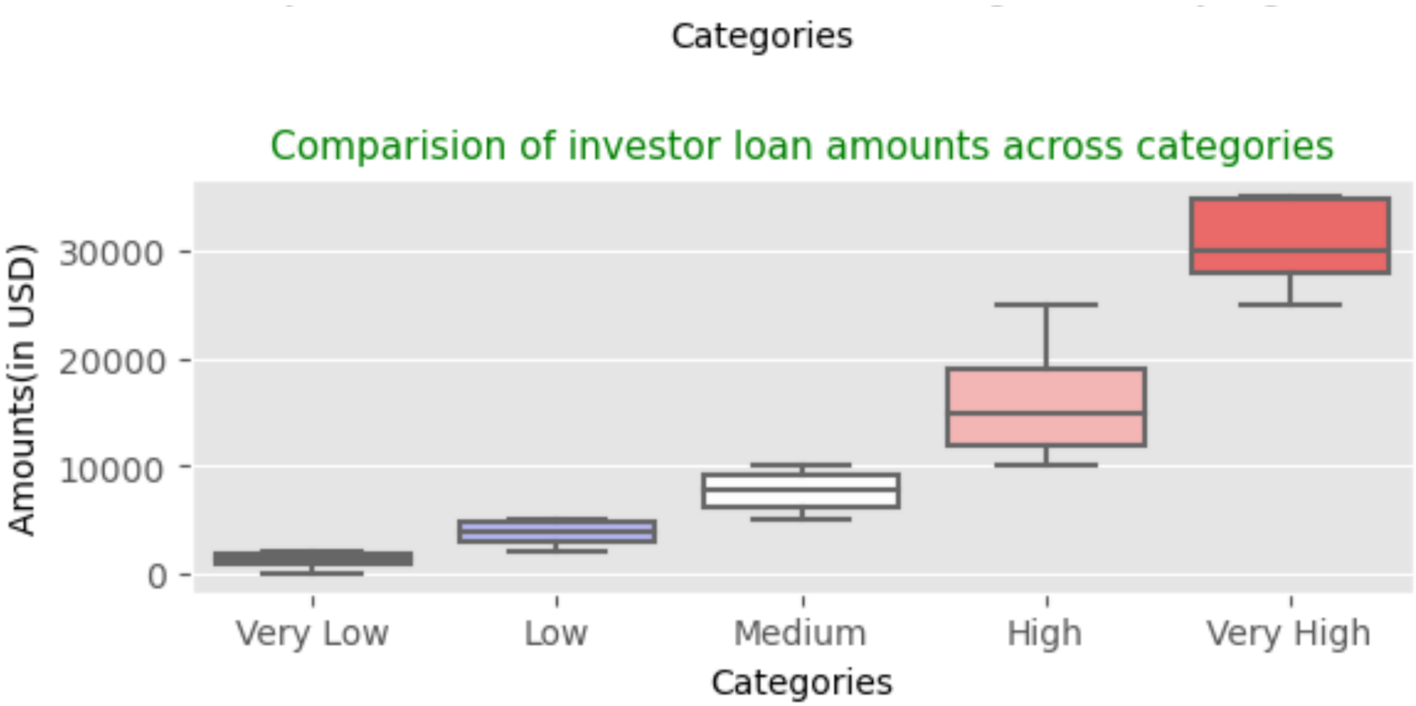
Observation: CA state has the highest number of loan applicants and also with maximum number of charged offs followed by NY and FL.

Univariate Analysis



Observation: A large number of loan applicants were on Rent or were paying Mortgage on their houses implying that many applications in the future would come from such profiles. Candidates who own their houses do not generally apply for loans that frequently. For some candidates, the status of their residence could not be determined but since the data is negligible, they can be ignored.

Bivariant Analysis

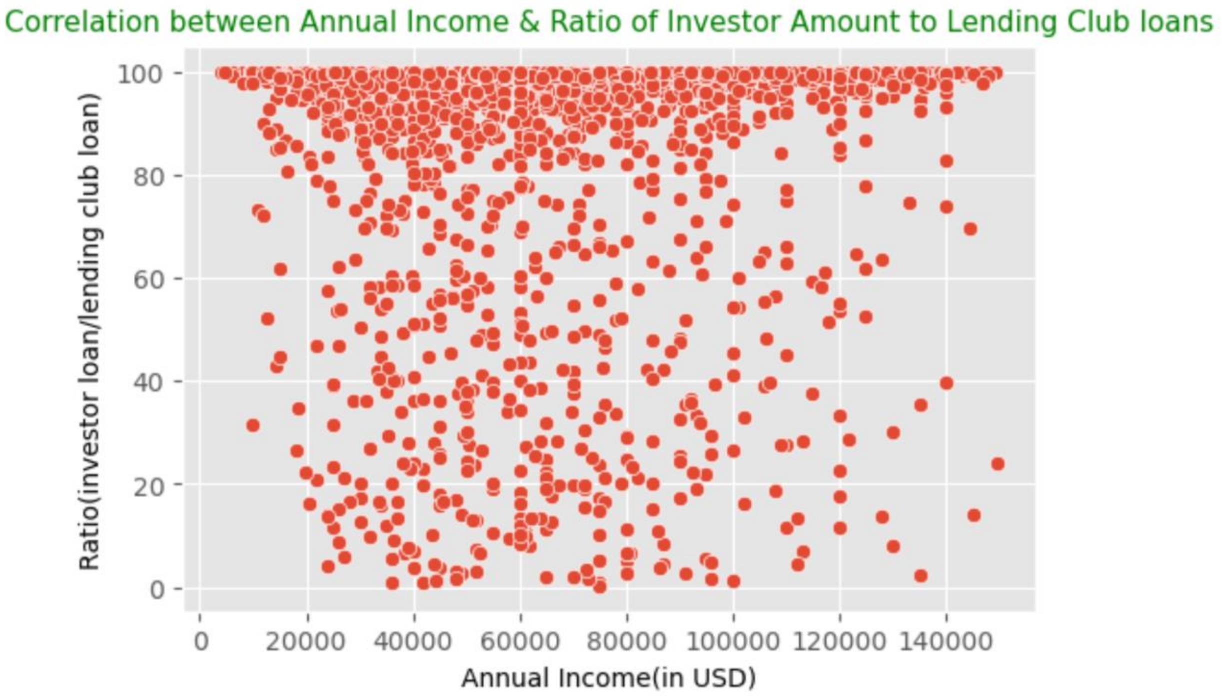


Observation:

1.In general, borrowers tend to request for higher loan amounts(USD 10,000-25,000), however the investor tends to sanction lower loans which is seen by looking at the sizes of box plots across all categories and seen prominently for categories High & Very High. However the median of these 2 categories are very close.

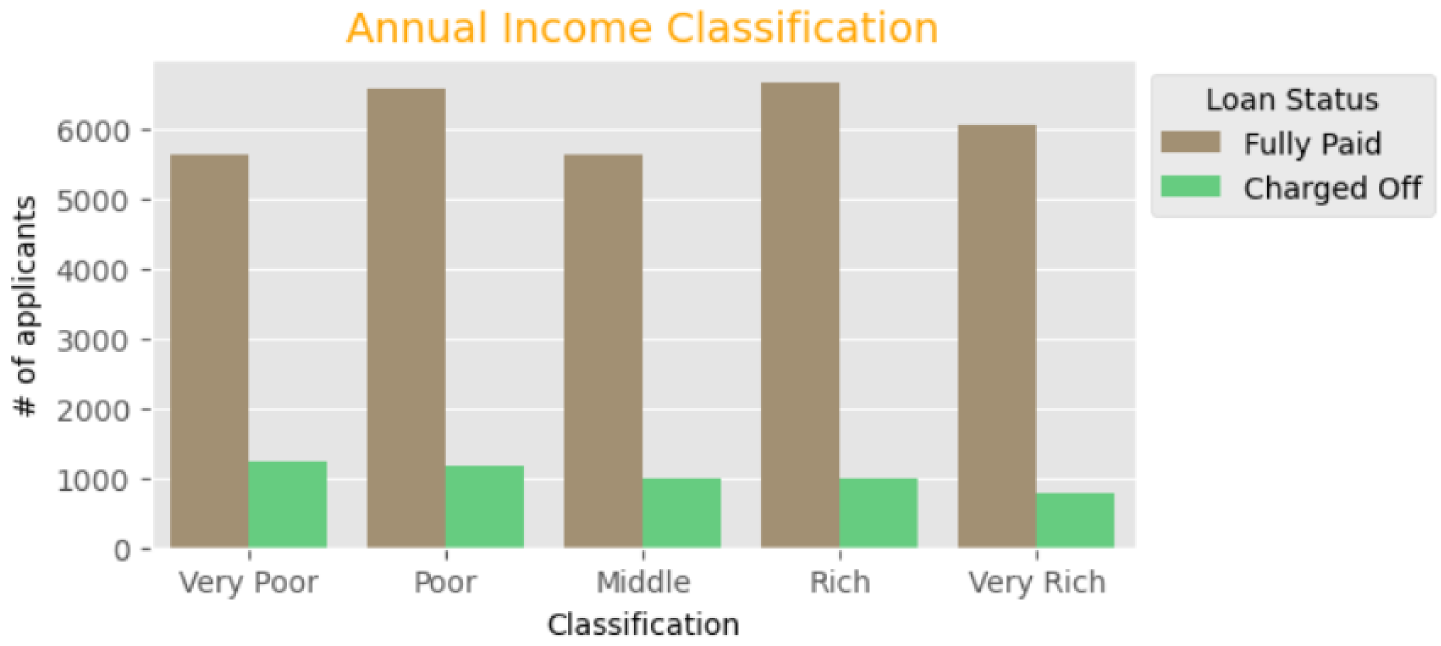
2.Almost of the requested loans from the defaulted borrower are matched up with the loans funded by the investor.

Bivariant Analysis



Observation: Across all annual income data for charged off applicants, the investor tend to sanction the loans at a higher ratio of 80% & above & strongly approves 100% for many applicants. However, there is no other pattern that is visible for lower ratios.

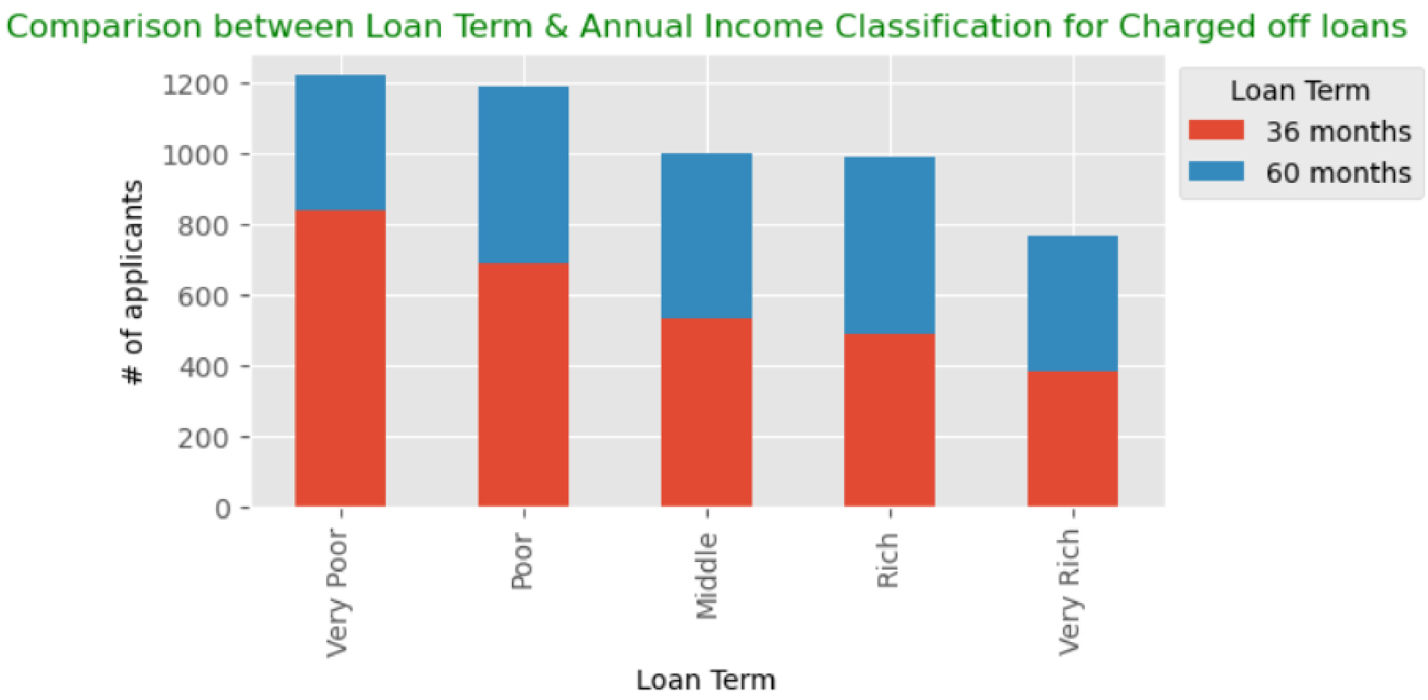
Bivariant Analysis



Observation:
High number of defaulted(charged off) loans belong to those form the Very Poor or Poor income families. It is surprising to note that those with very high income also have measureable defaulted loans which implies that most likely they have unhealthy financial habits.

Generally, it is noted that those who belong to the Rich or Poor classifications tend to fully pay their loans even though there are applicants in both categories with measurable defaulted loans.

Bivariant Analysis



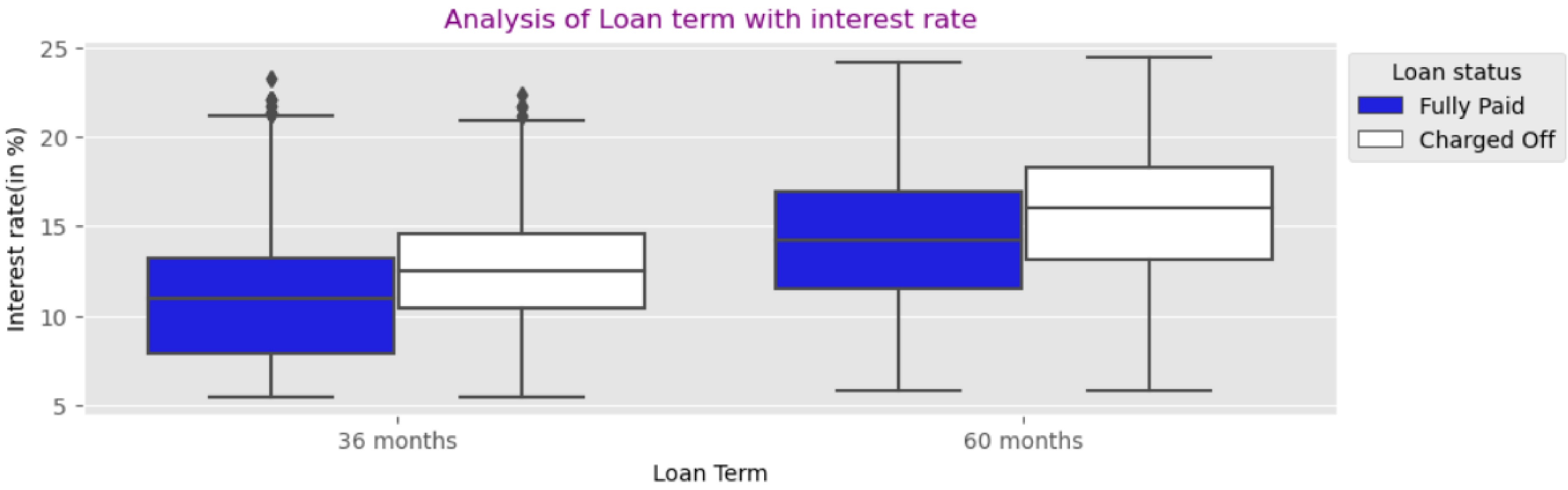
Observation:

1.Defaulted applicants have a loan term of 36 months on a much higher side for Very Poor & Poor income classifications.

2.Middle applicants have almost same ratio between the different loan terms.

3.Rich & Very Rich applicants have slightly larger defaulters with a 60 months loan term.

Multivariant Analysis



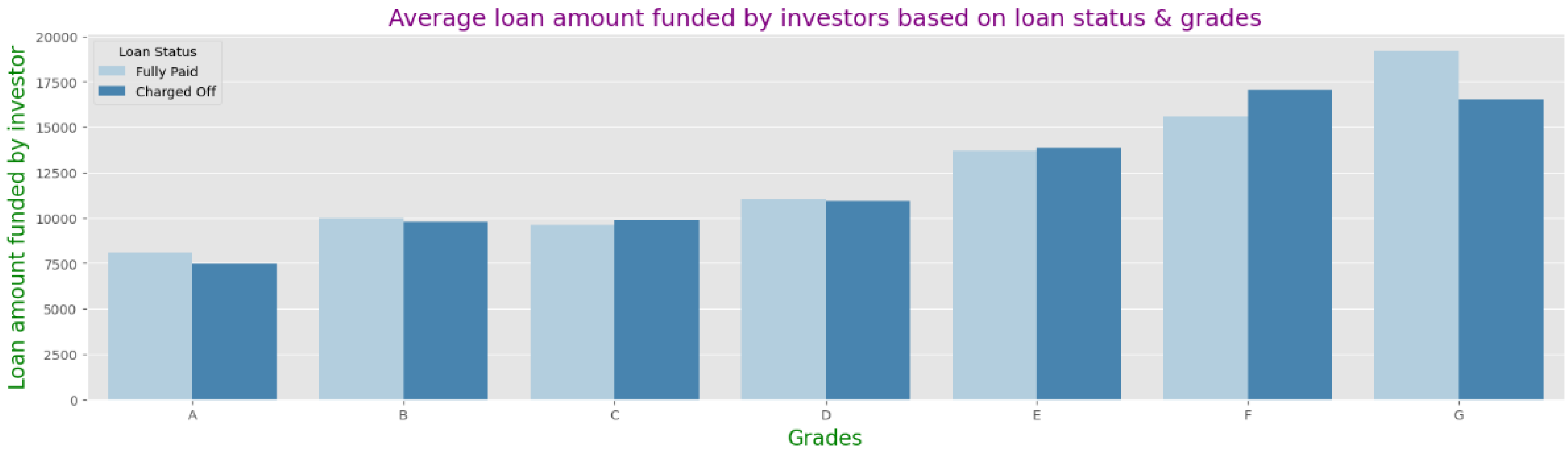
Observation:

1.There is a significant difference in the median interest rates between defaulters and non-defaulted applicants in both types of loan terms.

2.Applicants with higher interest rate in either loan terms tend to default which is around 13% for loan term of 36 months & 16% for loan term of 60 months.

3.Loan terms of 60 months have on general much higher interest rates compared to loan terms of 36 months which may imply that those who opt for loan term of 60 months may most likely tend to default.

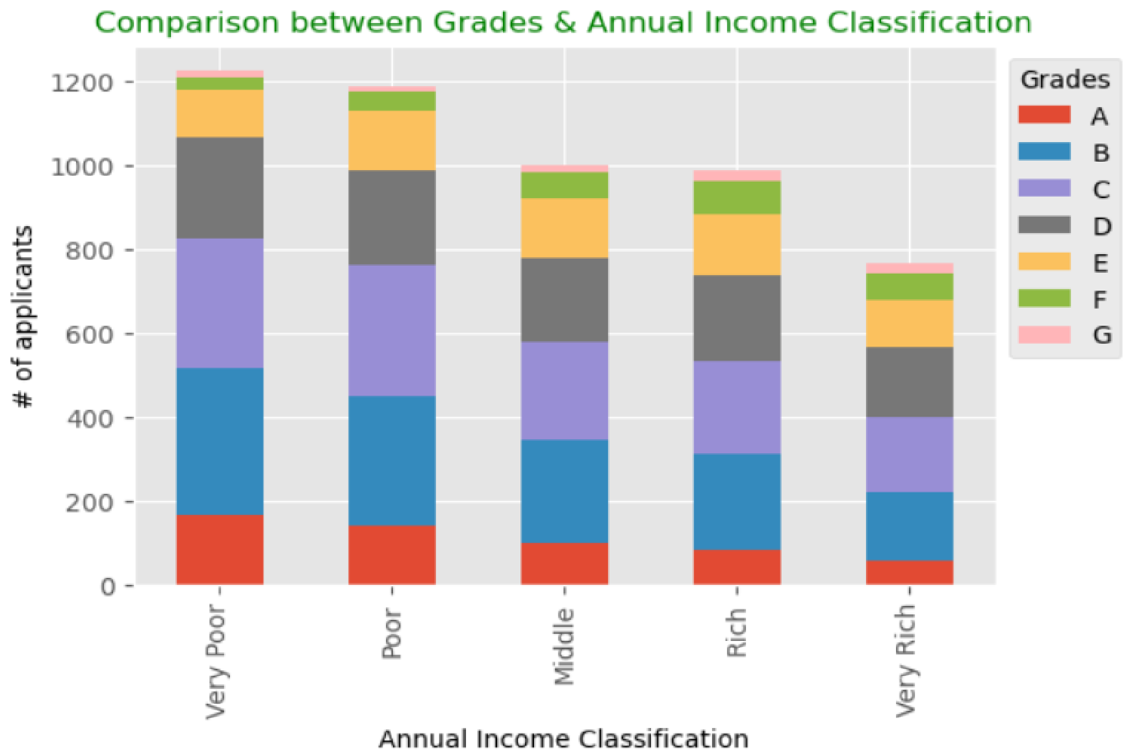
Multivariant Analysis



Observation:
Generally applicants whose loans have grades B,C & D have highest defaulted loans with median sanctioned loan amounts by the investor fall in the medium bin(USD 5,000-10,000) which would form a huge chunk of credit loss for the company.

However, loans with riskier grades E,F,G have a high median sanctioned loan from the investor, yet the number of defaults are lesser in comparison to those in all of the lower less riskier grades. Since the sanctioned loan amounts are high, the company maybe advised to sanction lower loan amounts for riskier grades.

Bivariant Analysis



Observation:

1.All defaulted applicants have loans across all grades with the total being least for Very Rich applicants and highest for Very Poor applicants.

2.Very Rich defaulted applicants tend to receive loans with grades C and lowest with grade A.

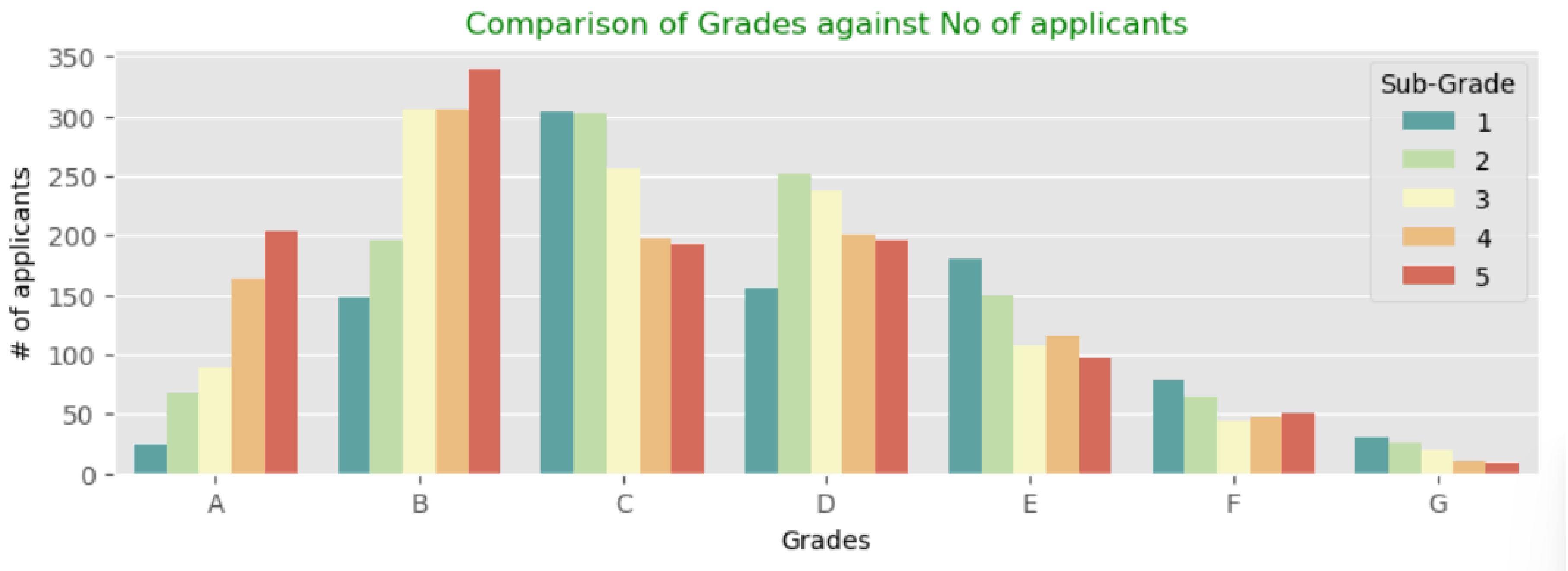
3.Rich defaulted applicants have the highest number of loans with grade F in comparison with other income classifications.

4.Middle income defaulted applicants receive loans with grades almost similar in proportion with the remaining incomes.

5.Poor income defaulted applicants tend to receive loans with grades mostly in B & C.

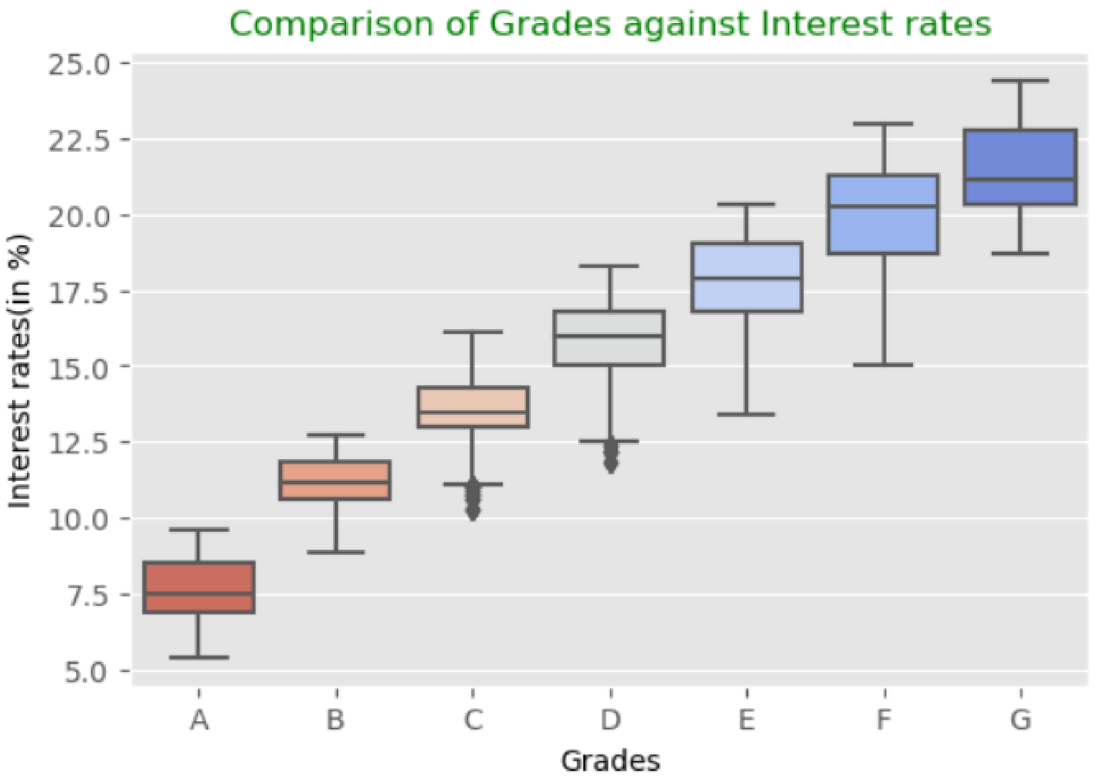
6.Very poor income defaulted applicants have similar characteristics to Poor income defaulted applicants in proportion except for grade F which it has the least.

Bivariant Analysis



Observation:
On comparing grades & sub-grades for defaulted applicants:
1.Sub-grade B5 has the highest number of defaulters followed by B3, B4, C1 & C2.
2.The next highest peaks are seen for C3, D2 & D3.
3.The lowest peaks are seen within all sub-grades of G followed by F and then A1.

Bivariant Analysis



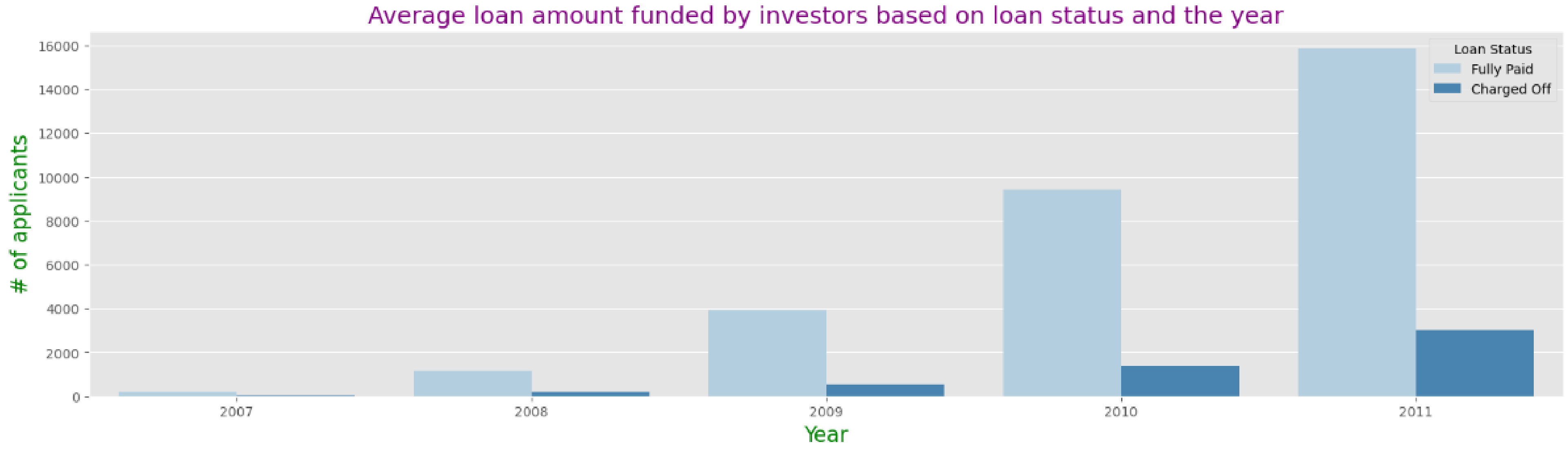
Observation:

On comparing grades & interest rates for defaulted applicants:

1.There is an increasing trend of interest grade with the increase in the grade risk with the lowest being for grade A and highest for grade G. There is a also a wider range of interest rates provided for grades E,F and G.

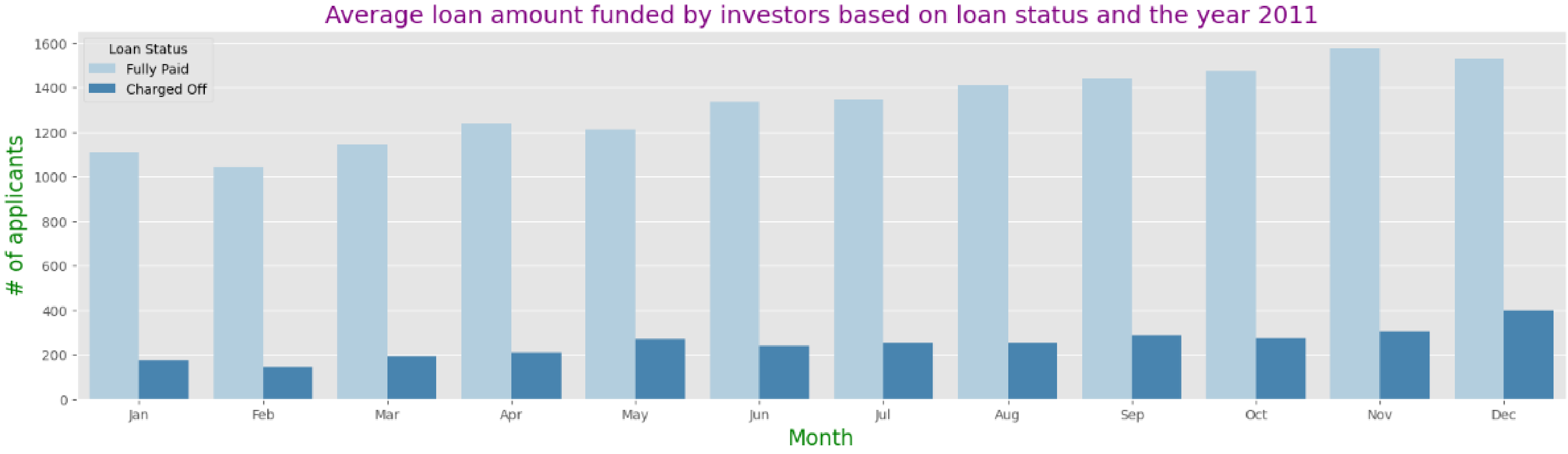
2.The grades C & D have outliers out of which sub-grade C1,D1 & D2 are the largest outliers implying that these sub-grades have lower interest rate and yet the applicant defaulted.

Bivariant Analysis



Observation:
Maximum loans that were defaulted was in the year 2011 and the least was in the year 2007. The exact same behavior exists also for the those loans that were Fully paid.

Bivariant Analysis

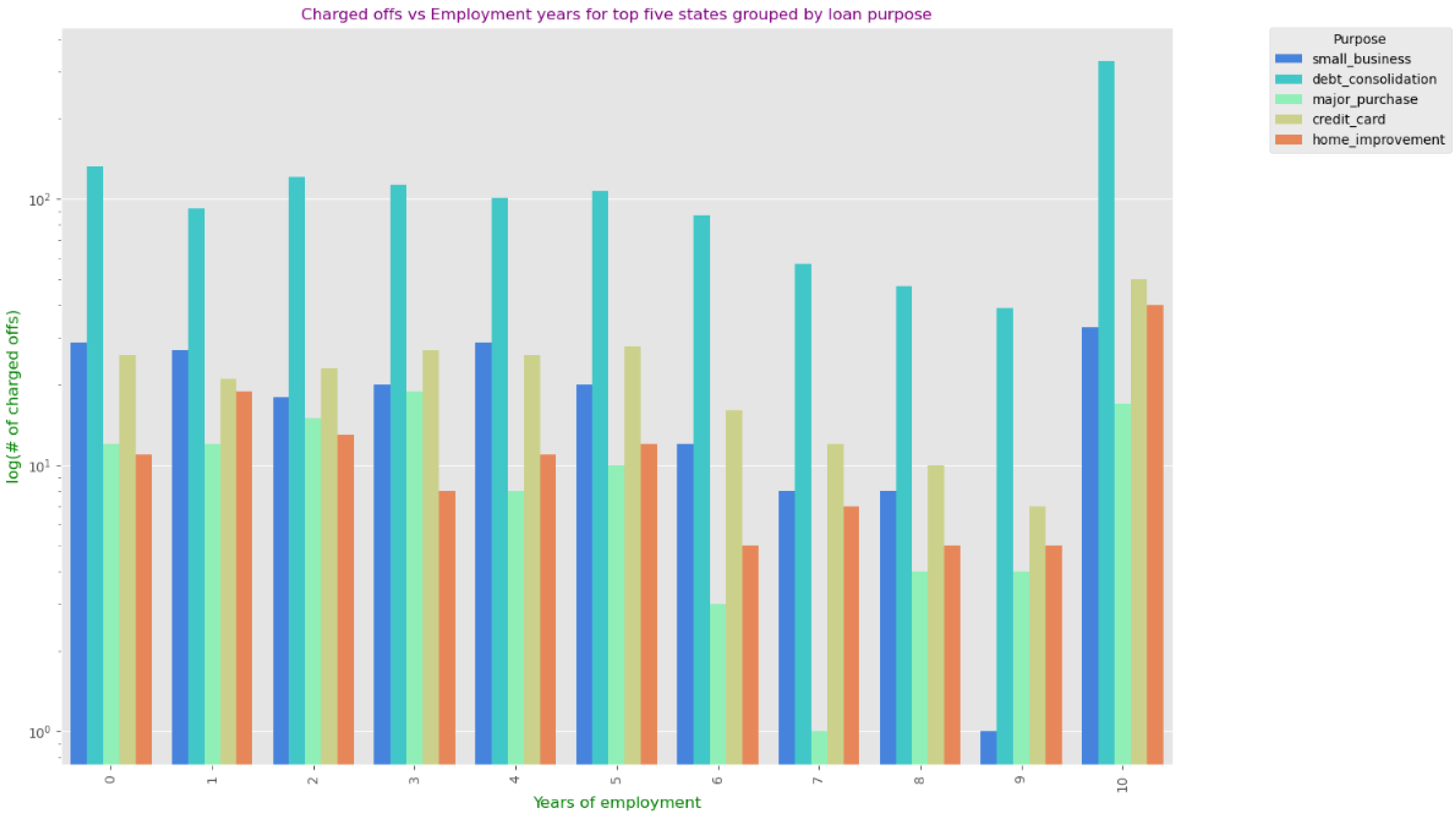


Observation:
Maximum defaults occurred in Dec 2011 and during later half of the year which are Sept, Oct & Nov and May & July during mid of the year. There is a consistency in loans that were fully paid off in the year 2011 which has maximum defaults.
In the year 2011, maximum loans were taken for the purpose of **debt_consolidation**.

As per www.investopedia.com, Debt consolidation is the act of using a new loan to pay off older debts and liabilities. By combining multiple high-interest debts into one, you may be able to save money, pay lower EMIs, and manage one's finances better.

According to https://www.federalreservehistory.org/essays/great-recession-and-its-aftermath, The Great Recession had bottomed out in the middle of 2009, however the economic weakness lasted for few more years which impacted the credit conditions of US households and resulted in historically elevated unemployment rates. This could explain why there were many defaults in later half of the year 2011 and therefore can be attributed as a rare event.
Investopedia
Investopedia is the world's leading source of financial content on the web, ranging from market news to retirement strategies, investing education to insights from advisors.

Bivariant Analysis

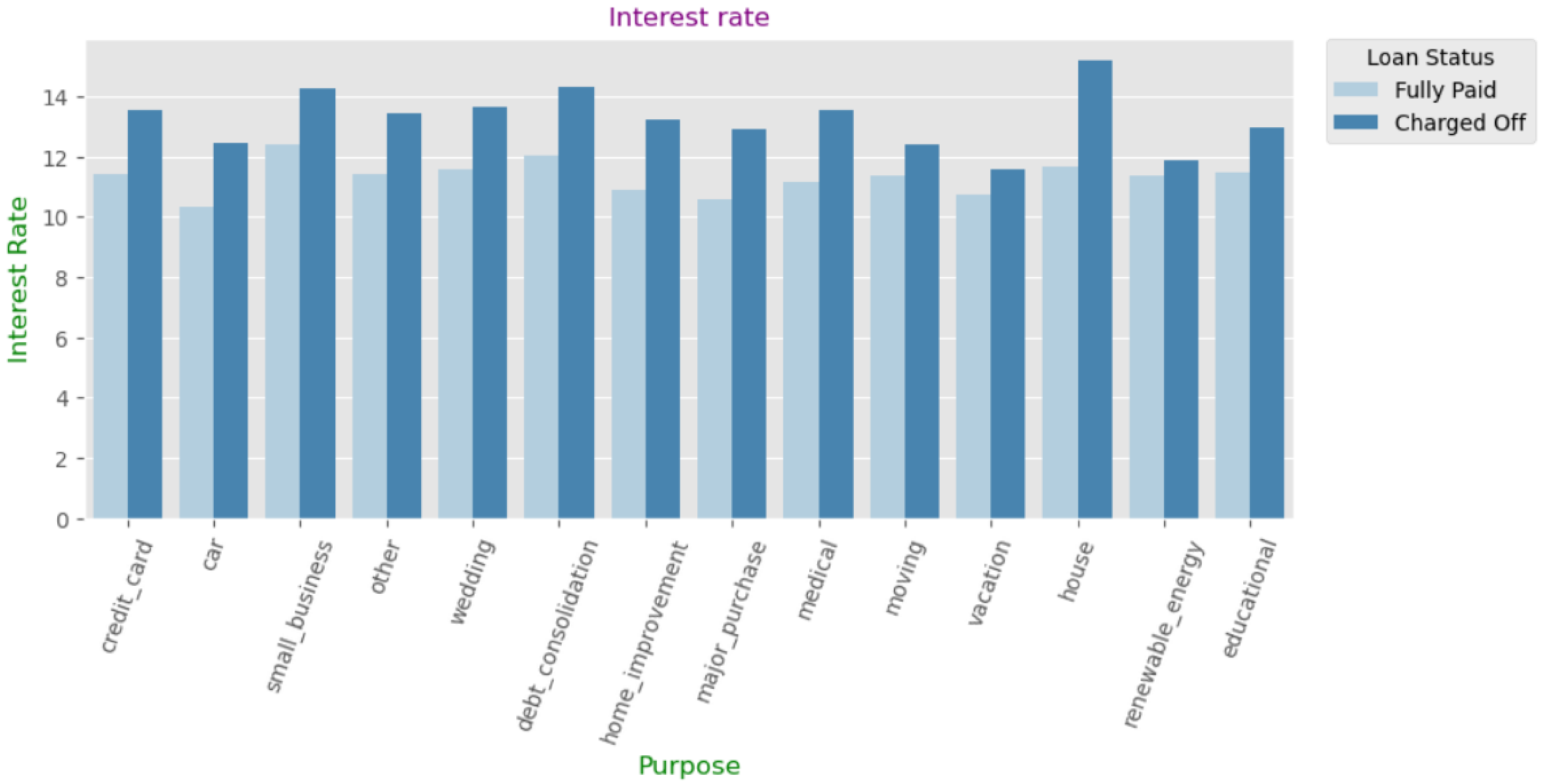


Observation:

1.Charged offs are maximum for applicants with employment years greater than or equal to 10.

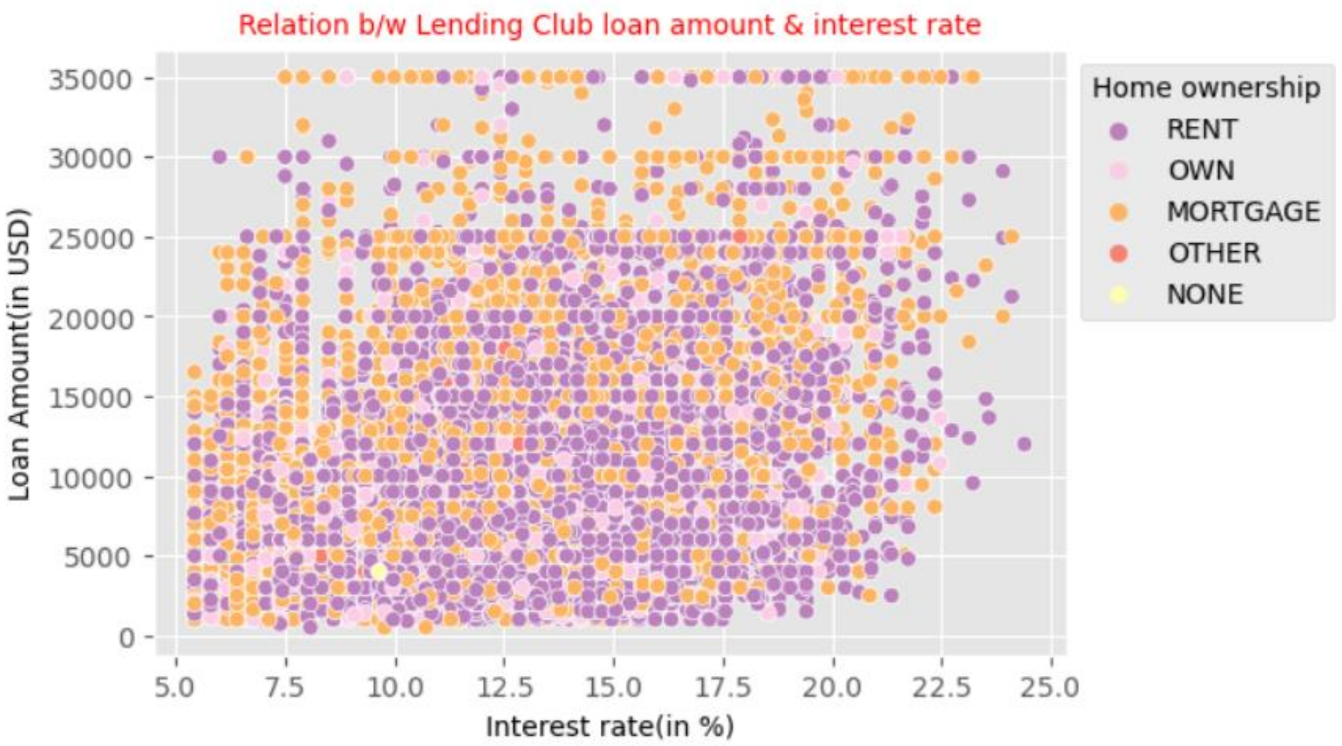
2.Debt consolidation is the major contributer across different employment lengths

Multivariant Analysis



Observation: Interest rate is maximum for house loan followed by small businesses and debt consolidation

Multivariant Analysis



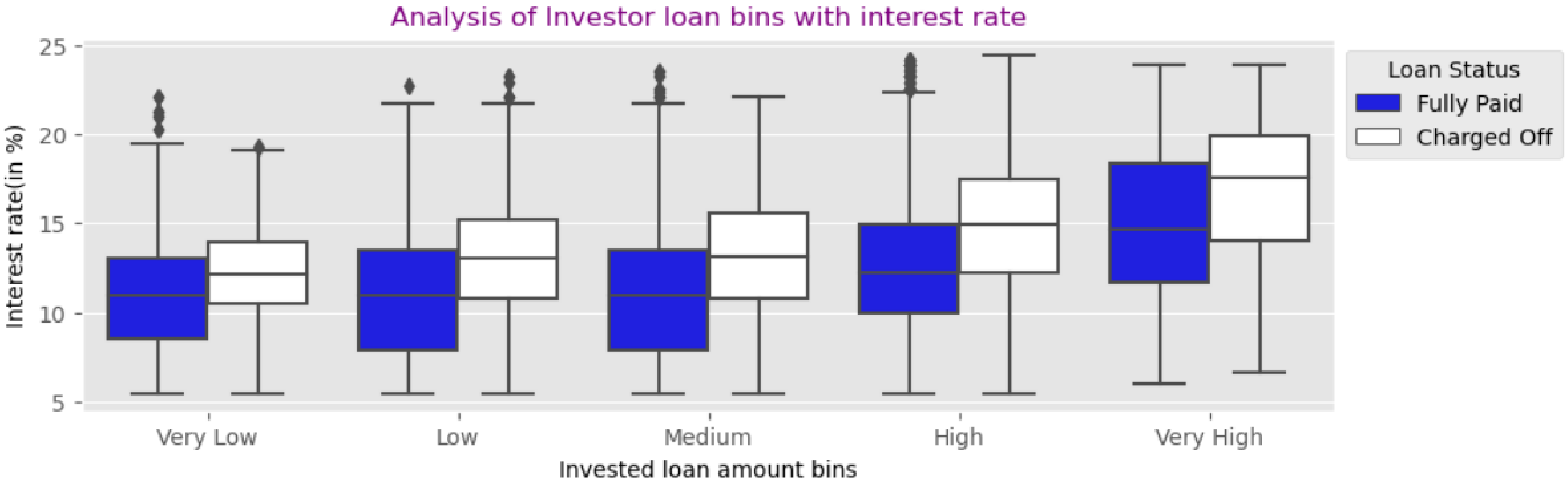
Observation:

1.Investor sanctioned amounts around USD 35,000 mostly belong to defaulted applicants with mortgage which implies that any applicants on mortgage and with high investor funded amount tend to default on these loans.

2.Majority of the clustering of data points towards to the lower end of sanctioned loan amount for those applicants living on rent. This implies that even though the loans sanctioned are for the most part low, applicants still default especially if they live on rent.

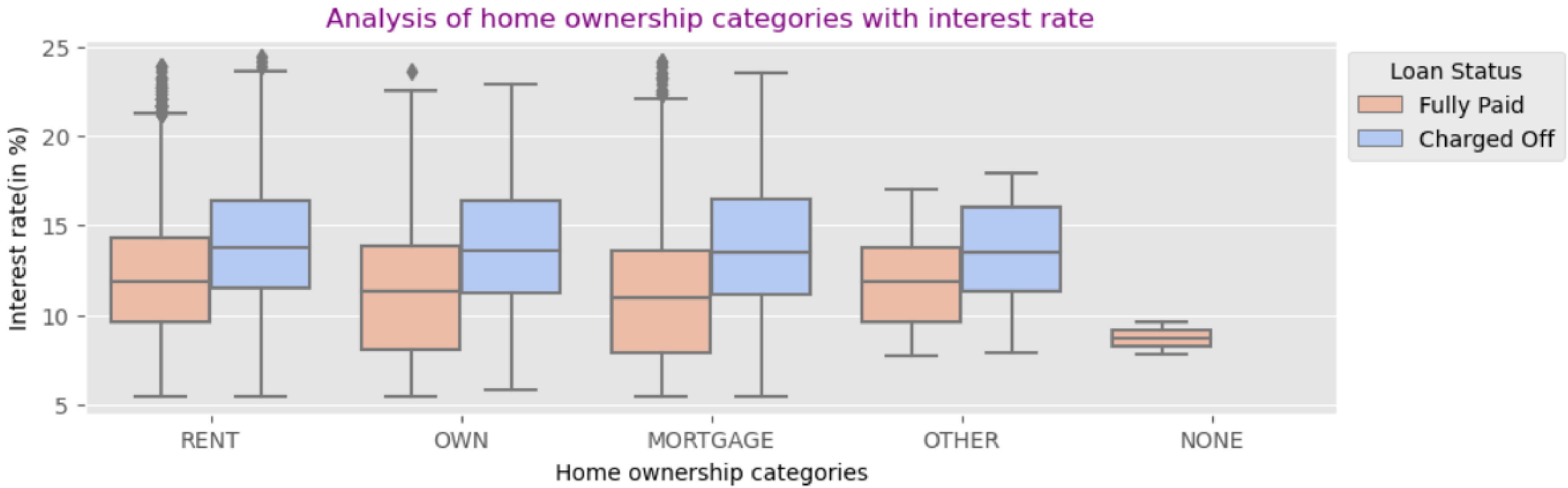
3.A cluster of defaulted applicants owning their own residence tend to default with higher interest rates but with low to middle sanctioned loan amounts.

Multivariant Analysis



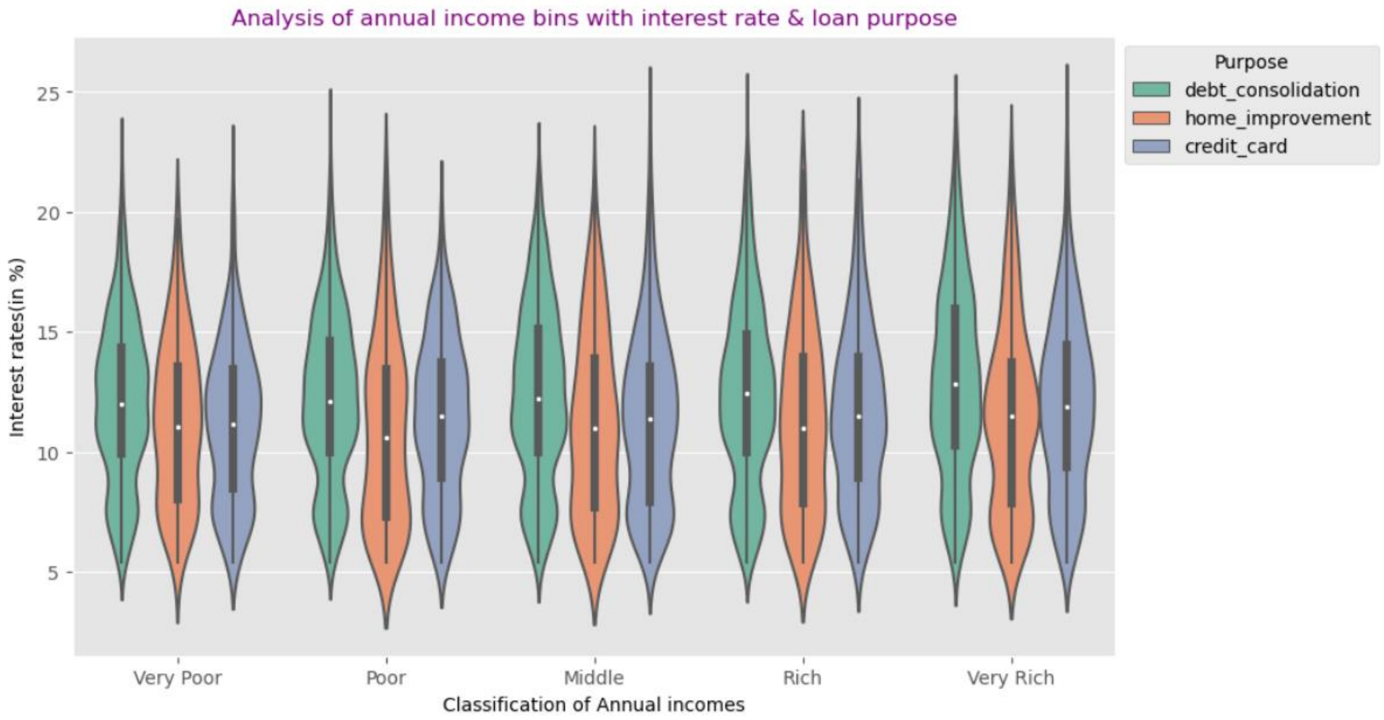
Observation:
1.Defaults happen for high interest rates for all bins.
2.There is an increasing trend of interest rate against the increasing loan amount bins.
3.Outliers are present only for Fully paid loans where the interest rates are very high ie above 20% except for Very High income applicants. Only Very Low & Low income defaulted applicants have outliers.
4.Median interest rate for fully paid applicants falls in the range between 10%-15%.
5.Median interest rate for defaulted applicants falls in same range, except for Very high which is closer to 18%.

Multivariant Analysis



Observation:
1.There is no data for defaulted applicants whose home ownership is unknown. Since the fully paid applicants are very less, this can be ignored.
2.The range of interest rate is within 8%-16% for all categories of home ownership, however defaulted applicants have a median interest rate of around 14% which is on a higher side in comparison with fully paid applicants who have it around 11%-12%.
3.Outliers exist for those on rent & paying mortgage with very high interest rates though a majority of them are for those who have paid their loans fully.
4.Only outlier for defaulted applicants are for those living on rent with a very high interest rate above 24%.
5.Boxplot for charged off are similar for all categories, however for fully paid, applicants on mortgage or owning their own homes are the largest which implies that this applicants get a wide range of interest rates compared to those living on rent or other types of accommodation(perhaps business offices/factories)

Multivariant Analysis



Observation:

1.Median interest rate for debt consolidation is the highest(around 15%) for Rich & Very Rich defaulted applicants.

2.Median interest rate for debt consolidation is lowest(around 13%) for Very poor defaulted applicants.

3.Very rich defaulted applicants tend to default more with credit card loans with an interest rate between 14.5%-15%.

4.For the rich defaulted applicants, it is mostly for credit card followed by home improvement loans. This is similar for those in middle & poor income classifications.

5.For the very poor defaulted applicants, they tend to default with lower interest rate(around 13%) for credit cards and more so for home improvement with higher interest rates (around 14%).

Observations

Profile of **defaulted applicants** typically contain following indicators:

1.Home ownership:
1.1 Applicants lives on **rent** or are paying **mortgage**.
1.2 Applicants tend to be accept **interest rates** around **14%**.
1.3 Applicants living on **rent** and with **interest rate** above **25%** most likely always defaults.

2. Grade & sub-grade:
2.1 **Very rich** applicant will most likey default with **grade G** at an **interest rate** of **20%**.
2.2 **Rich** applicants tend to apply & default with **grades F & G** for large number of loan applications.
2.3 **High-Very High** sanctioned loans from investors have the highest interest rates.
2.4 **Median** amount sanctioned loans are assigned **grades B,C & D** and are higher in applications and which have the highest defaults resulting in the maximum credit loss for the company.
2.5 **Very poor** & *Poor** income groups have the highest number of **median** sanctioned loan amounts with grades B(sub-grades B3,B4,B5(highest)), C(sub-grades C1,C2,C3) & D(D2,D3). They also tend to default **lesser on riskier grades** because of **lower interest rates**.
2.6 Loan **sub-grade A1** has the least number of defaulted applicants which is also the least for Very high, Poor and Very poor income groups.

3. Term
3.1 **Very Poor** & **Poor** income groups have loan term of **36 months**.
3.2 **Rich** & **Very Rich** income groups have loan term of **60 months**.
3.3 Defaults happen at an interest rate of **13%** for loan term of **36 months** while it is **16%** for **60 months**.
3.4 Loan term of **60 months** have higher interest rates.

4. Purpose

4.1 **Debt consolidation** has the highest defaults with interest rate at 15% in the **year 2011** & **month December** which may be due to after-effects of The Great Recession. This is an event that may repeat during any global event that affects a USA economic cycle.

4.2 Debt consolidation defaults is high for **Very Rich** & **Rich** applicants & lowest for **Very Poor** applicants.

4.3 **Very rich** applicants tend to default more with **credit card loans** with an interest rate between **14.5%-15%**.

4.4 **Very Poor** applicants tend to default more with **home improvement** with higher interest rates of around **14%** and least with **credit card** with interest rates of around **13%**.

5. Interest rate

5.1 The higher the interest rate, the higher likelihood that applicant defaults.

5.2 The higher the loan sanctioned by the investor, higher is the interest rate.

5.3 The median of interest rate is in the **range of 10%-15%** for all income groups.

5.4 **Very Low** & **Low** sanctioned loan amounts have very high interest rates.

5.5 Higher the DTI, higher is the interest rate.

5.6 Very few applicants had applied for loan with very high rate of interest and the proportion of charged off's is maximum.

5.7 For very low and low rate of interest, proportion of charged off's is lesser.

6. Employment Length

6.1 Between 2-9 years, with increase in employment length, applicants tend to default less.

6.2 With employment length more than or equal to 10 years, there is more likelihood that the applicant defaults.

6.3 Debt consolidation is the major contributer for defaults across all employment lengths.

6.4 Annual income increases with increase in employment length.

7. States

7.1 CA has maximum number of defaults among all US states followed by NY, FL and TX.

Recommendations

1.Investor may sanction a much lower loan amount against the requested loan amount for those applicants belonging to the medium income group with sub-grade B5.

2.Whenever there is an economic down-turn in USA, the investor can try to be more cautious towards those applicants who have applied for loan requests against debt consolidation and either decide to reject the loans or reduce the sanctioned loan amounts with an increase in the loan terms while maintaining the same interest rates for the very poor & poor income groups.

3.For those applicants living on rent, the investor may decide to reject those applications which would have an interest rate of 25% and above.

4.Applicants can be offered lower interest rates for those living on rent or paying mortgage for those belonging to lower end of the income groups for those who apply for home improvement since they tend to repay their loans in general.

5.The higher end of income groups tends to go for credit card loans of riskier grade G with high interest rates which the investor may decide to reduce the sanctioned loan amounts and increase the term of the loan or reject in case the requested loan amounts are significantly high.

6.Care to be taken when lower income applicants apply for loans that qualify under grades B, C & D due to high defaults. The investor may decide to either reduce the interest rate or sanction lower loan amounts.

7.In general, if a loan applicant is against debt consolidation or credit card esp from the higher income groups, care may be taken to reject them or to reduce interest rates or sanction lower loan amounts.

8.Investor may reject the requested loans with high DTI since most likely they have high interest rates which are likley to default

9.Investor may take precaution while lending loans to applicants coming from California and take measures for those from New York, Florida & Texas especially for loans against debt consolidation.

10.Take caution when applicants have more than 10 years of employment when the loans are for debt consolidation.

Q & A