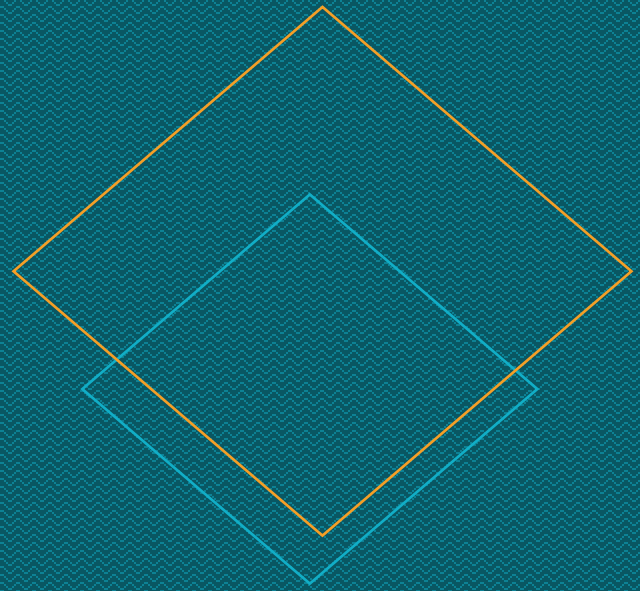
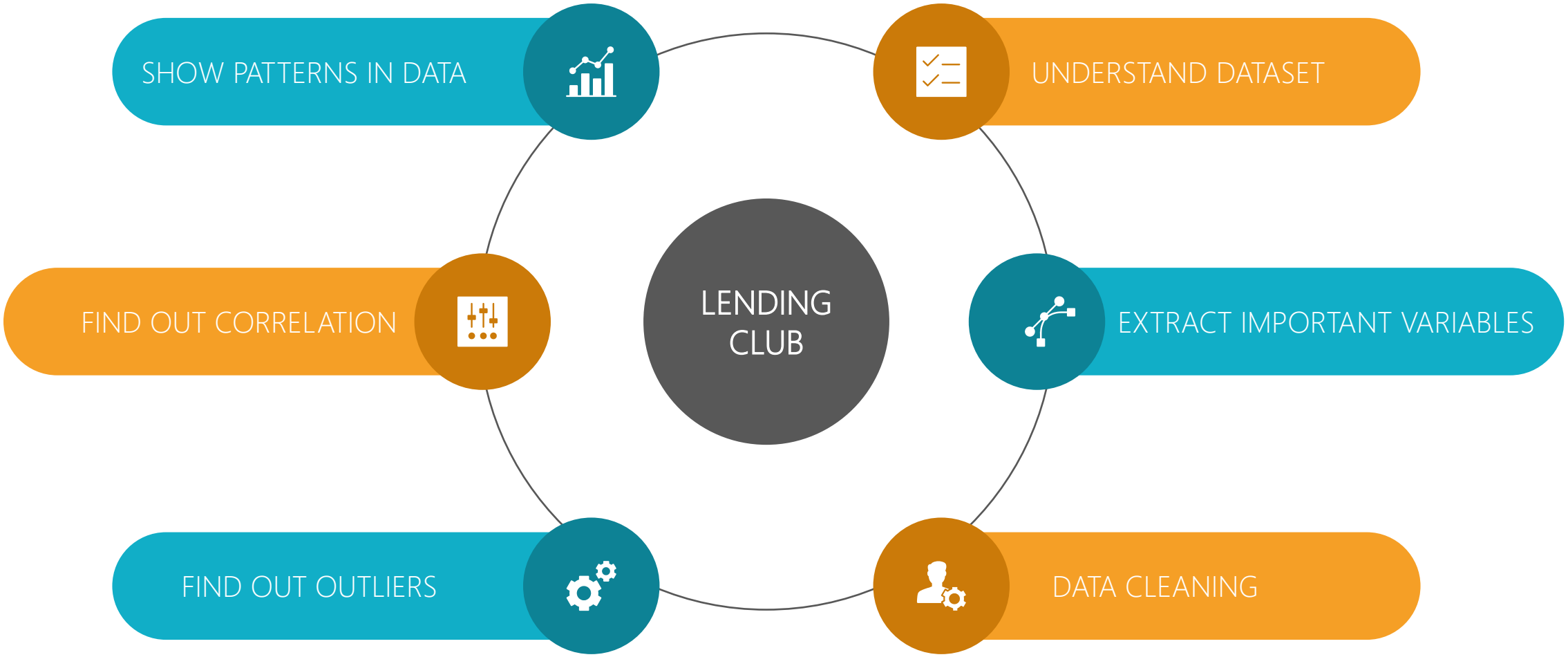


LENDING CLUB GROUP CASE STUDY

Manish Virmani
Spoorthi N

EDA Approach



Analysis

Recommendation

Note down the spread in default percentage for each variable analysis and at the end, decide based on the variance of this percentage that who are the strong indicators.

Bivariate Analysis

Considering the major categories, work with other variables to get better insight

Univariate Analysis

Perform segmented univariate analysis and check the defaulting nature.
Analysing the spread of each numeric variable

Data Organizing & Cleaning

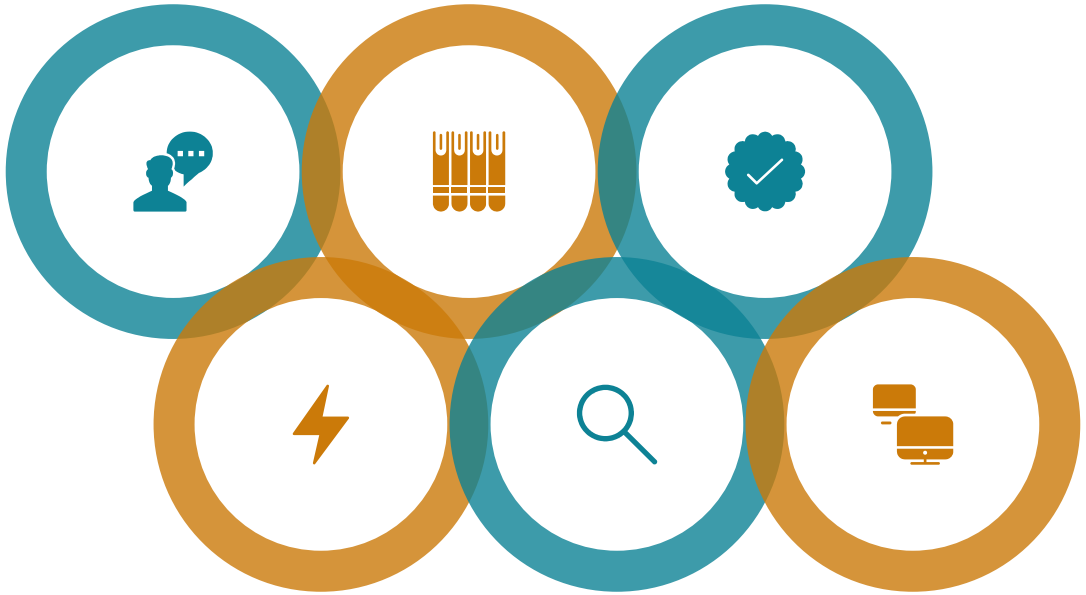
Drop all the rows with missing values

Drop all the columns with single values and high missing percentages

Data Cleaning

We found 54 such columns where all the values were null and removed those columns.

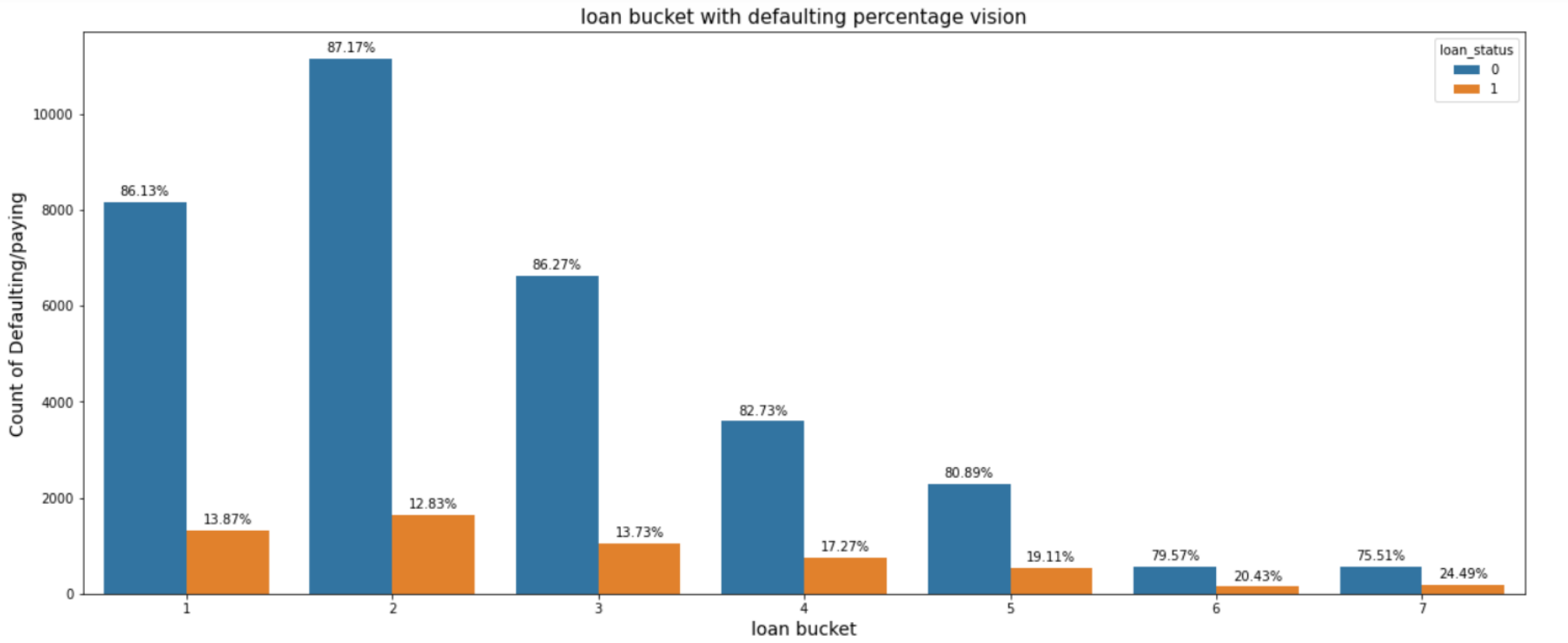
Removed columns with single value through out rows.



Dropped few columns as they will not add any value to the data like: 'title','zip_code' etc.

Dropped rows for "current" loan_status as they would not add much value

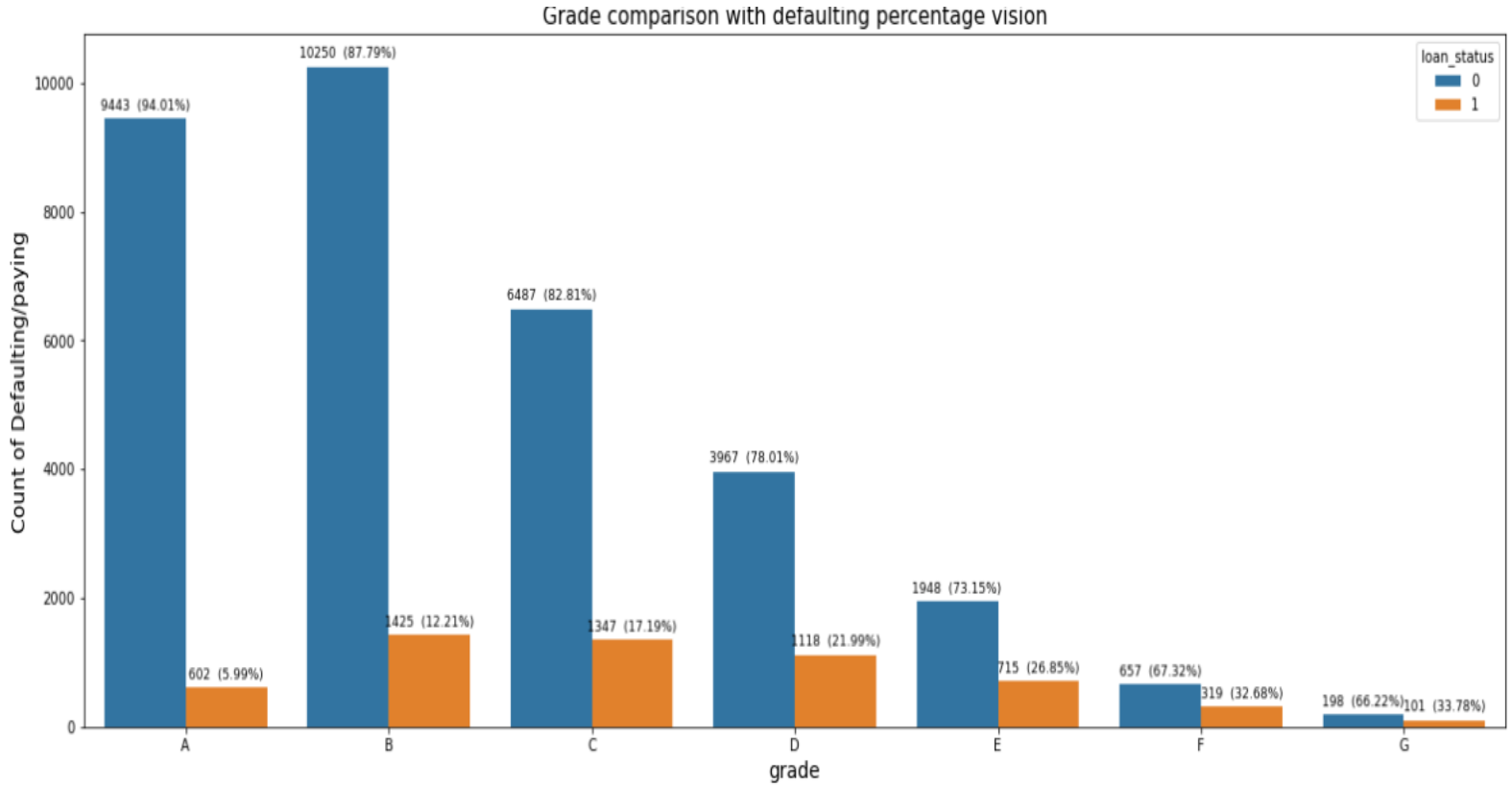
Univariate Analysis



With this we can see some pattern that above 15,000 (bucket 4) defualting rate is increasing.

Which means that when who have taken higher loan amounts are likely to be contributing towards defaulting rate

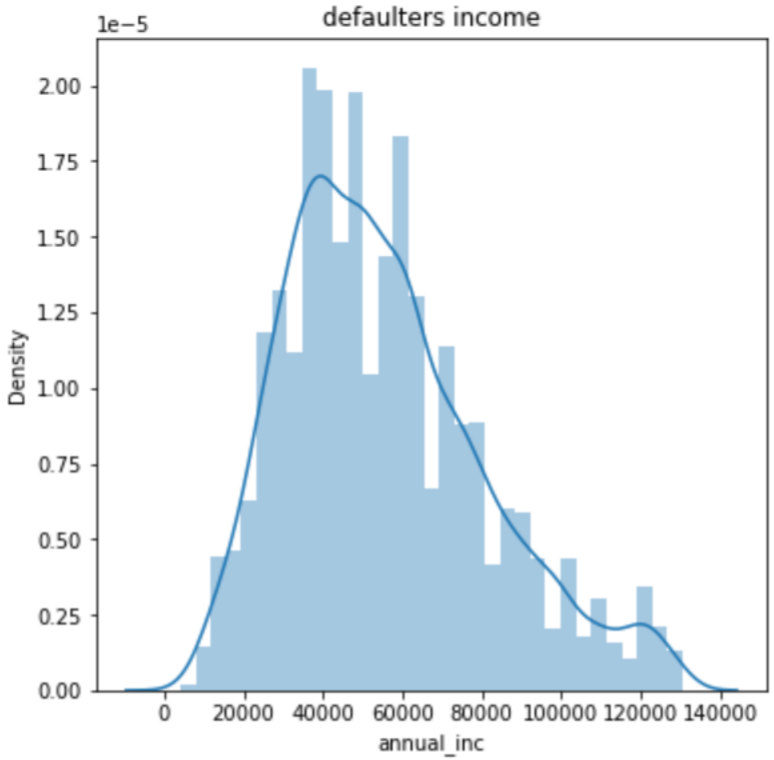
Univariate Analysis

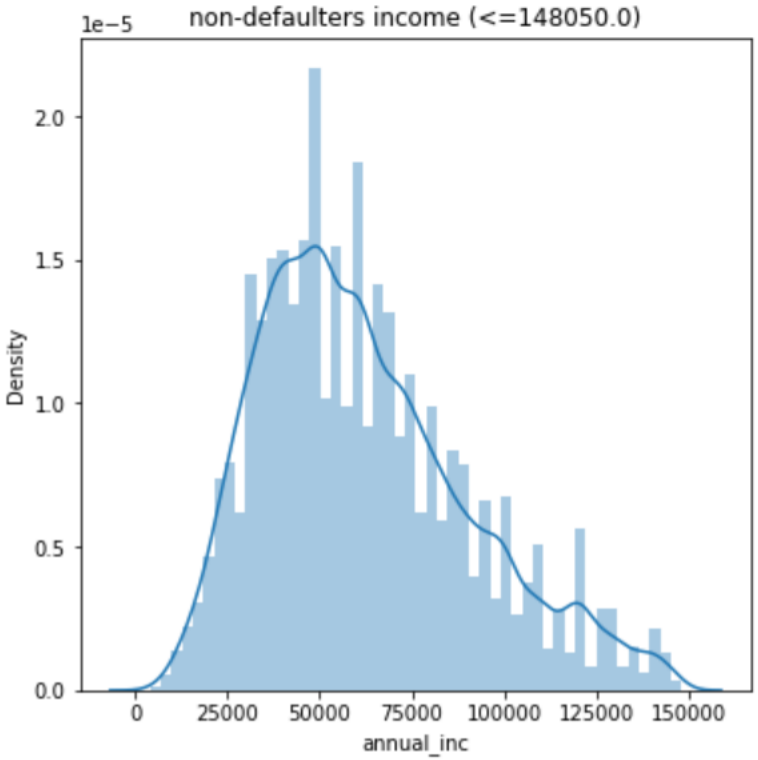


With this we can analysis that with Grade defaulting pattern is increasing.

We can see that A type loan has very less defaulting rate of 5.99% and Grade G has high defaulting rate.

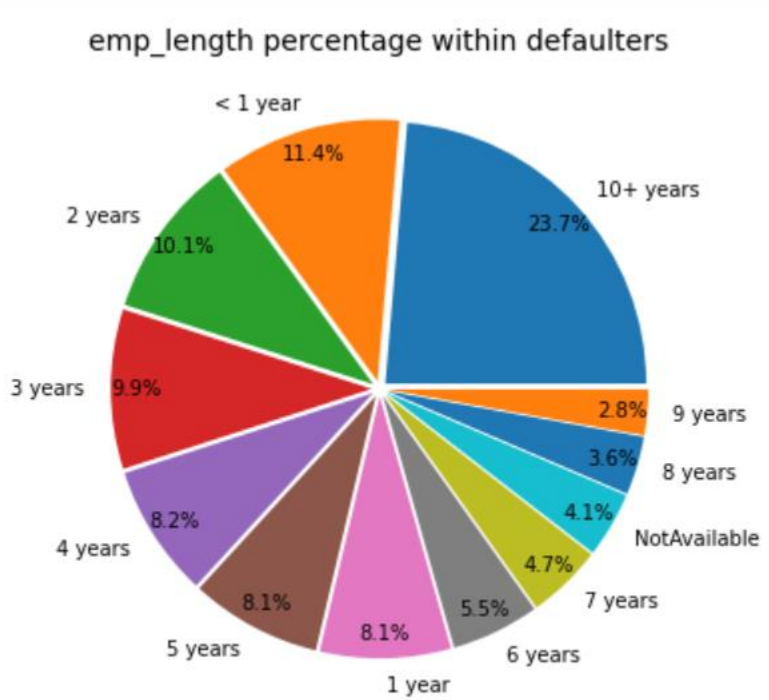
Univariate Analysis





From histogram we can see there is normal distribution in income of defaulters and non-defaulters

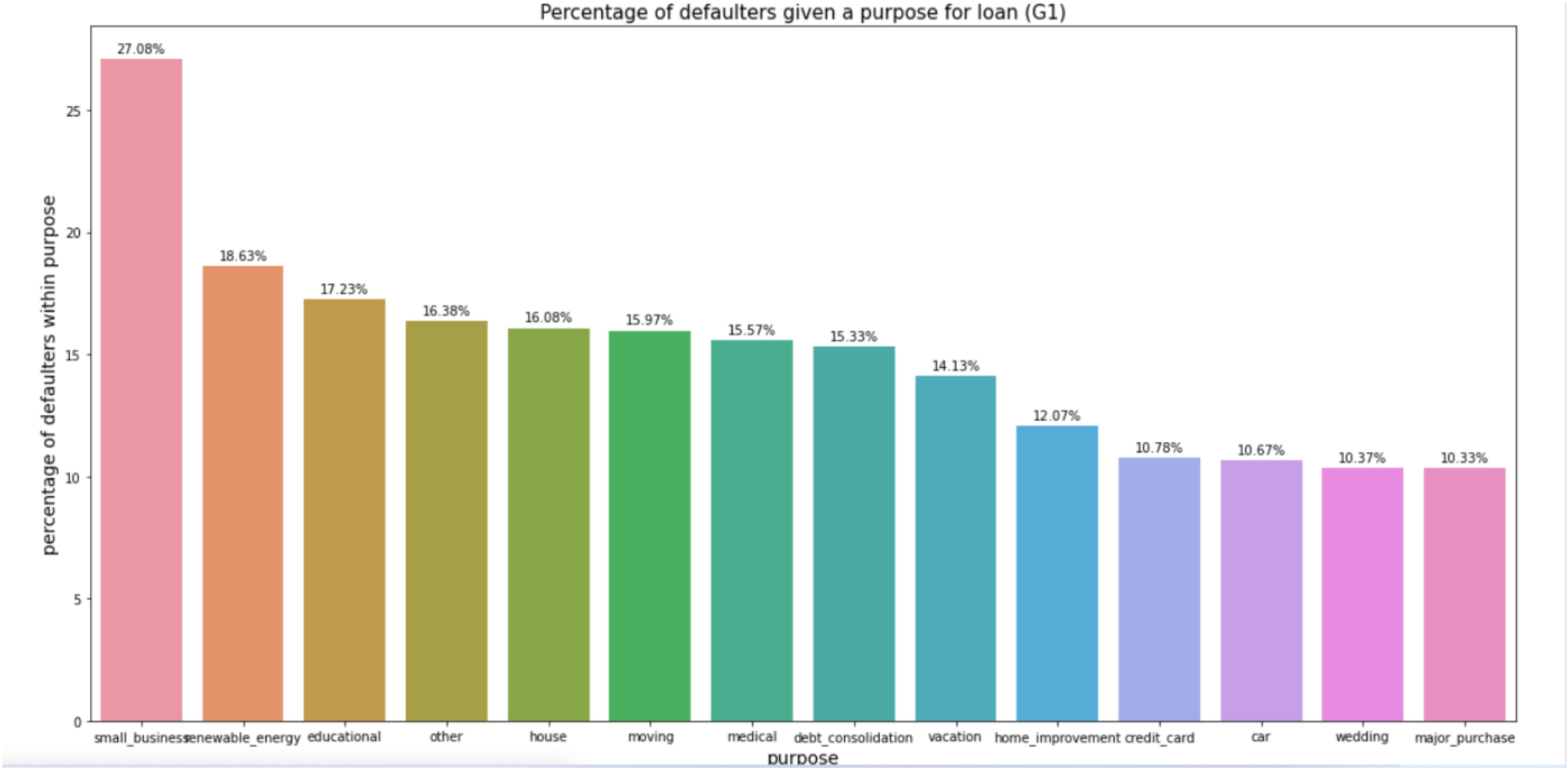
Univariate Analysis



"

From this pie chart we can see there are maximum defaulters over 10 years of experience

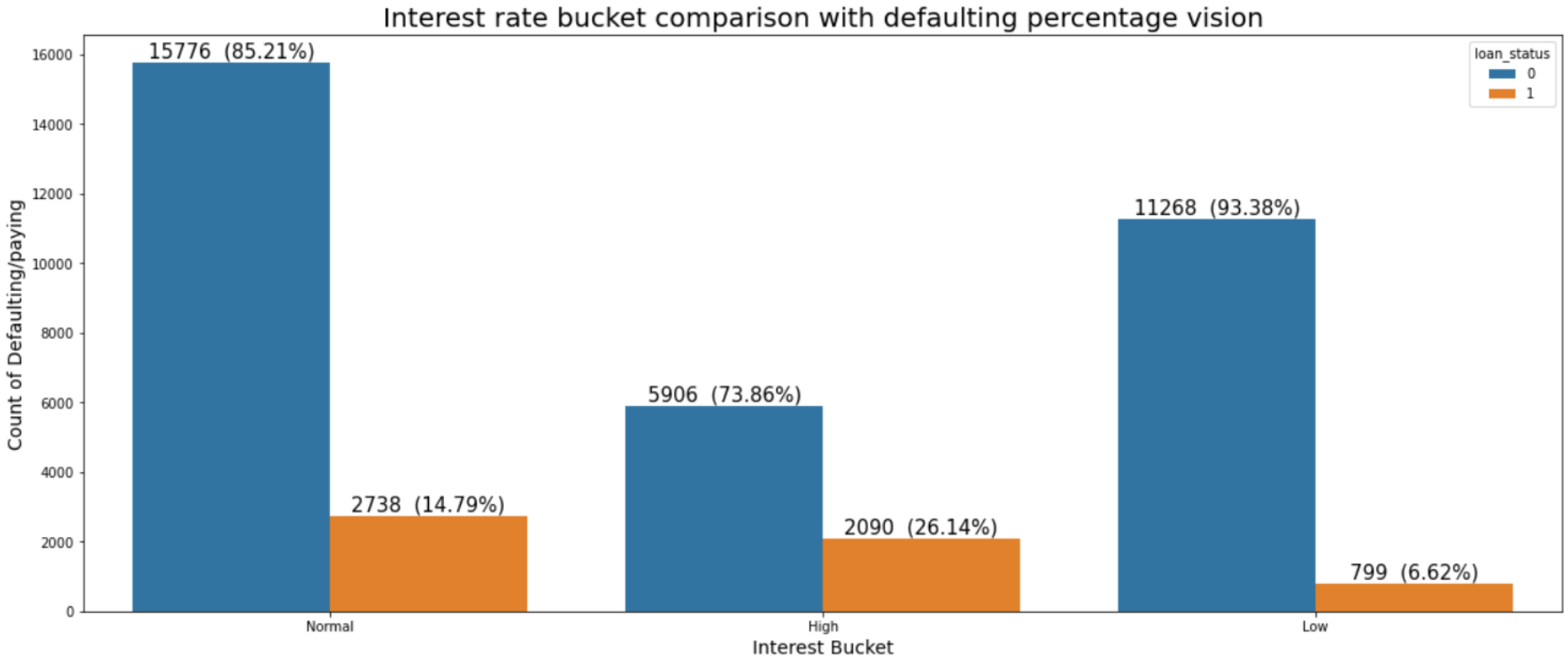
Univariate Analysis



"

Now we can see that defaulting percentage is higher for small_businesses borrowers i.e. 27.08%

Segmentate Analysis

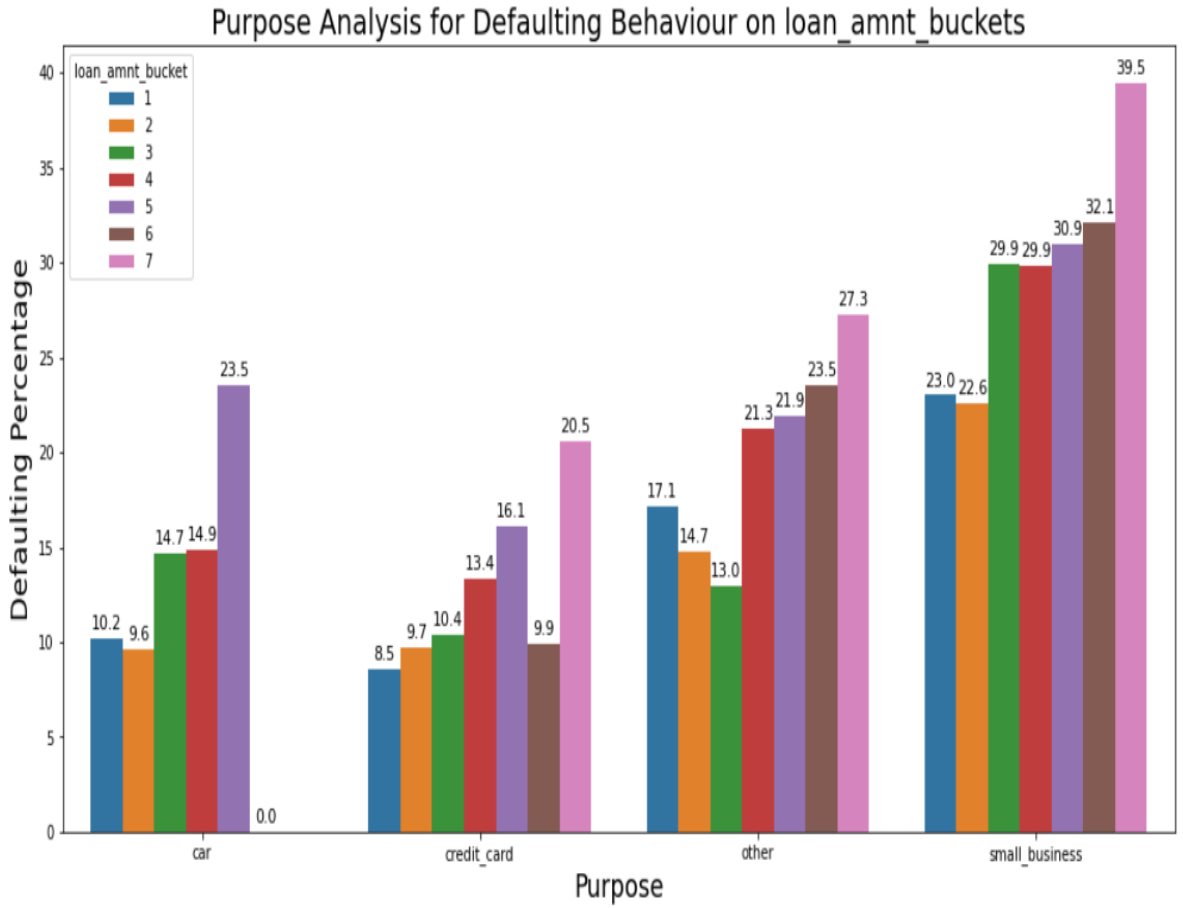


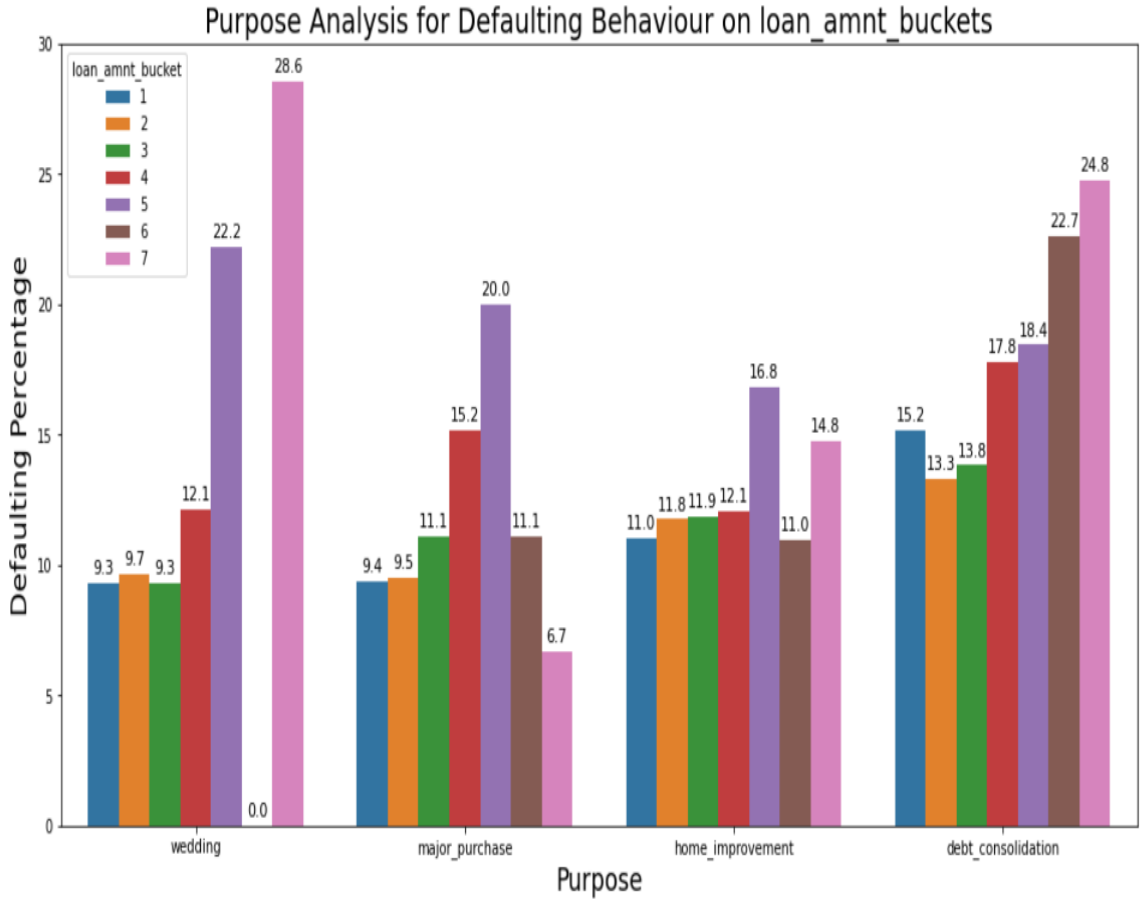
"

Now we can clearly see that defaulting percentage increases a lot from one bucket to the other bucket of interest rate

When grouped into three categories, we have observed that the higher the interest rate, more the defaulters.

Bivariate Analysis





Recommendations

POSITIVE

- Grade: We can see that A type loan has very less defaulting rate of 5.99%. We can recommend to go with Grade A or B as they have low defaulting rate
- Purpose: Giving loans in categories like marriage, major purchase, car and credit card have very less risk. Specially car loans have less default rates.
- Loan Amount: We can recommend to provide loan below 15000, as we see less defaulting rate
- Interest Rate: its better to give loans in the normal or lower range. - recommended to keep the interest rate below 15% -
- Term: It's recommended to give loan for 3 years instead of 5 years , as we see a default spike from 10% to 25% between these two.
- Income Bucket: We see less defaulting in between buckets 4-5-6, good to give loans in this range.

NEGATIVE

- Grade: As we go from A to G , the defaulting rate is increasing. It is not recommended to give loan to Grade F and G due to high defaulting rate
- Purpose: We saw that borrowers taking loans for small business send up defaulting around 25% of the times.
- Loan Amount: We can observe more defaulting higher loan amount loan bracket(7), so it is not recommended to give loans above 30000
- Interest Rate: When grouped into three categories, we have observed that the higher the interest rate, more the defaulters.
- Employment length: When looked at only defaulter, we see that 10+ years had 23% of them which is huge.
- Income Bucket Borrowers with income 5000-10000 are having the highest defaulting percentage, so better to avoid loans in this range.

Thank You