Case Study: Lending Club

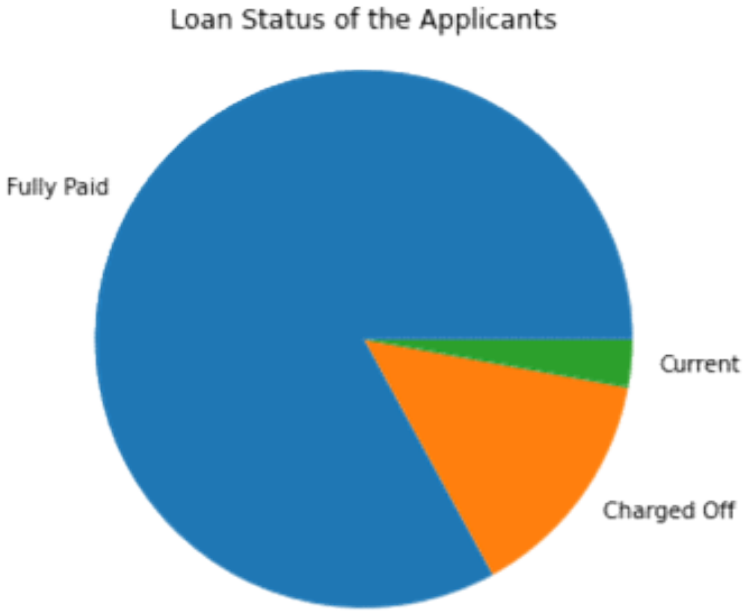
(Insights for loan dataset)

Author : Shiva Kumar Bandaru

Email : shivamkr.b271999@gmail.com

Dataset Overview:

- The below Pie Chart Shows the distribution of the applicants based on the loan status.
- The applicants are categorized as
  - Fully Paid loan applicants (who payed back their debt in time)
  - Current applicants (who is in paying process)
  - Charged Off (who did not payed back their debt in time)

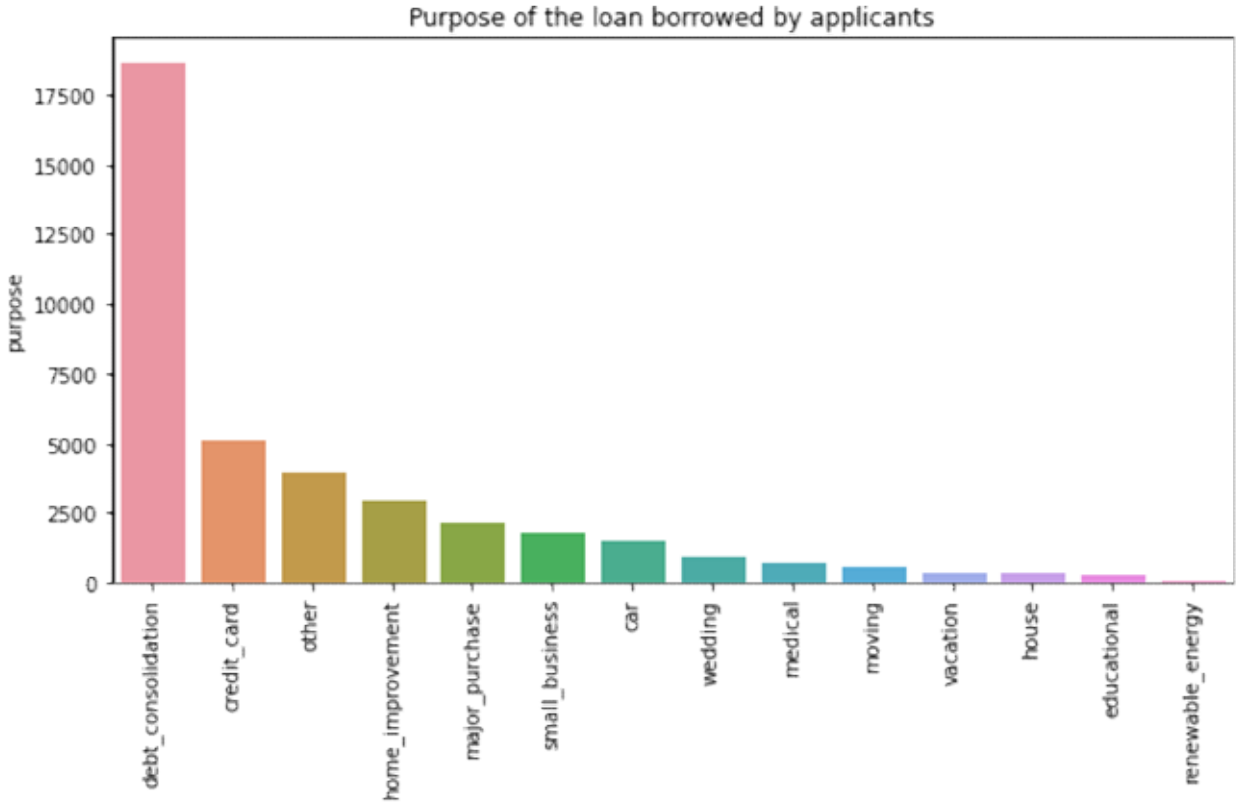


Distribution of loan_status of the applicants.

Status	Percent
Fully_Paid	82.9%
Charged_Off	4.1%
Current	2.9%

- The below Bar grap shows the distribution of purpose of the loan why did the applicants opt-in for a loan.

- 46% of applicants are applied loan for debt_consolidation



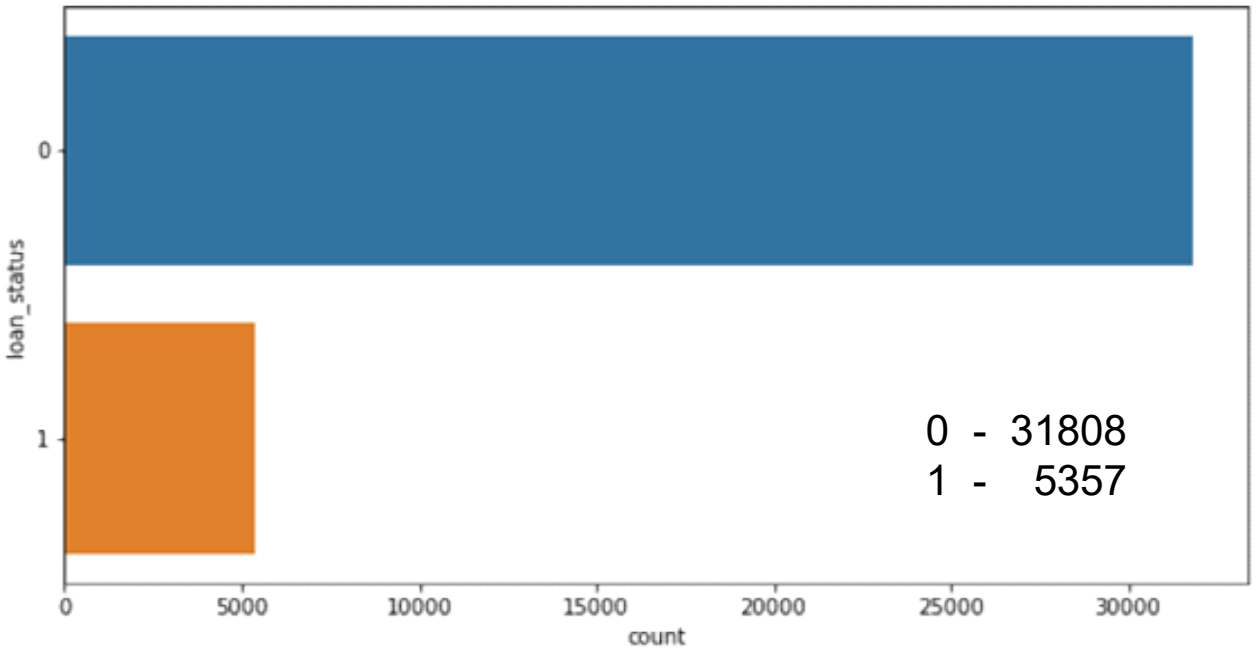
Purpose for the Loan Borrowed.

Purpose	Percentage
debt_consolidation	#46.9%
credit_card	#12.9%
other	#10.0%
home_improvement	#7.4%
major_purchase	#5.5%
small_business	#4.6%
car	#3.9%
wedding	#2.3%
medical	#1.7%
moving	#1.4%
vacation	#0.9%
house	#0.9%
educational	#0.8%
renewable_energy	#0.2%

After Data Cleaning

We filtered the "current" applicants.
As we are accessing the Risk detection based on the previous applicants data

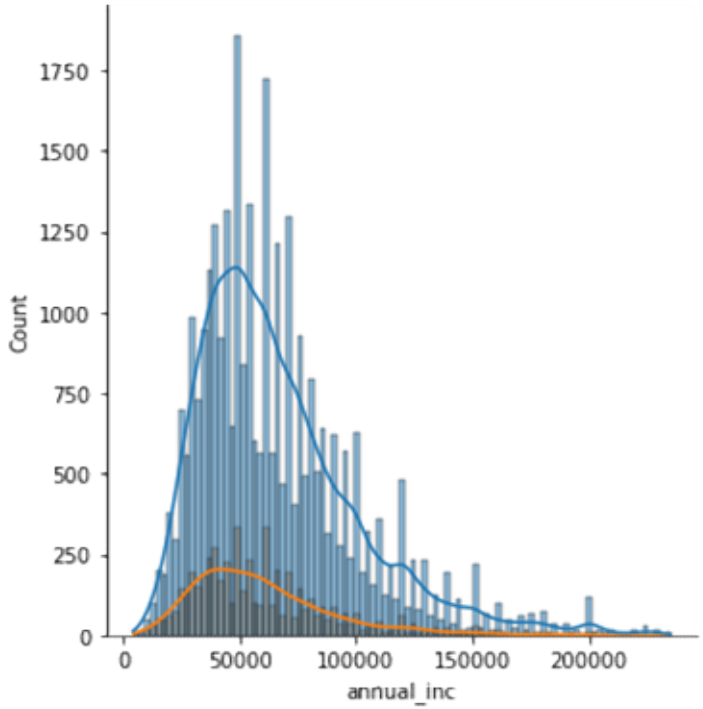
Graph below shows the count of Defaulters and Non-Defaulters.

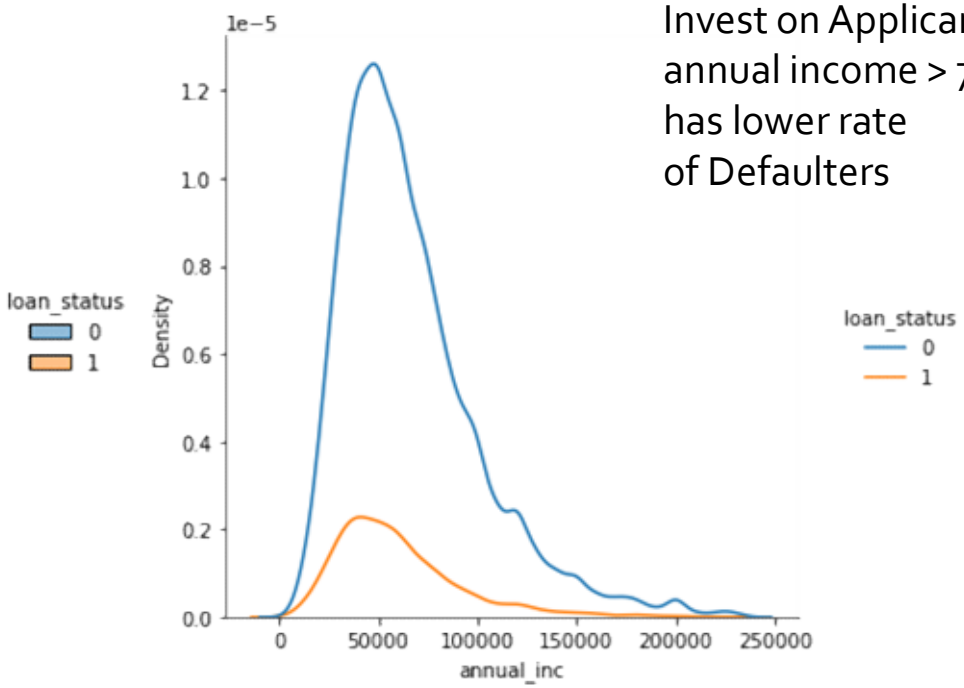


Data for Loan Status is categorised as

Non- Deafulters 0 (Fully Paid)
Defaulters 1 (Charged Off)

Distribution Plot of applicants with annual income w.r.t Defaulters and Non-defaulters





Recommendation :

Invest on Applicants with annual income > 70000 has lower rate of Defaulters

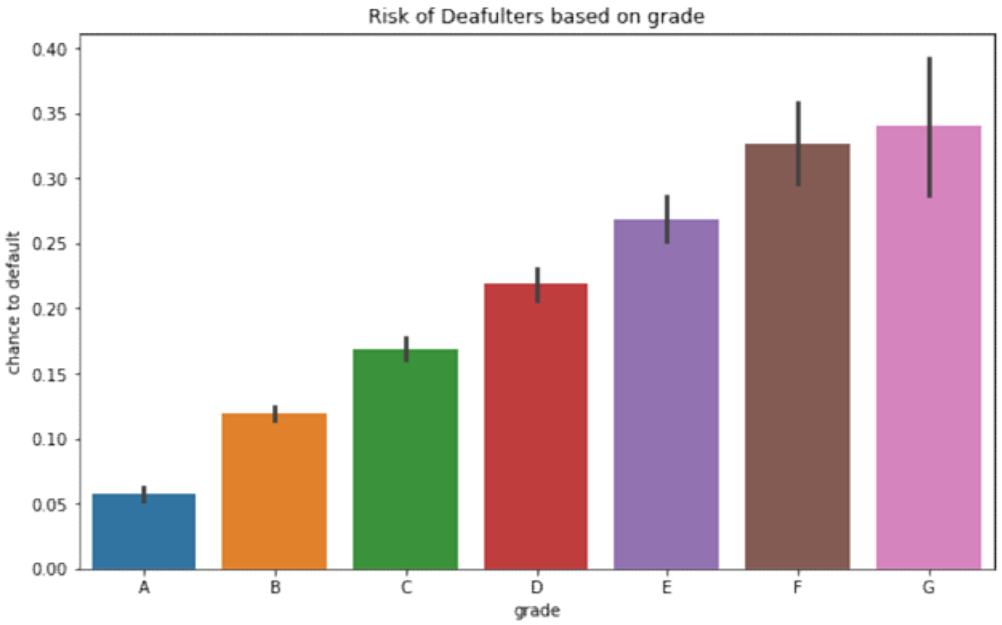
The Bar-Plots below shows the relation between

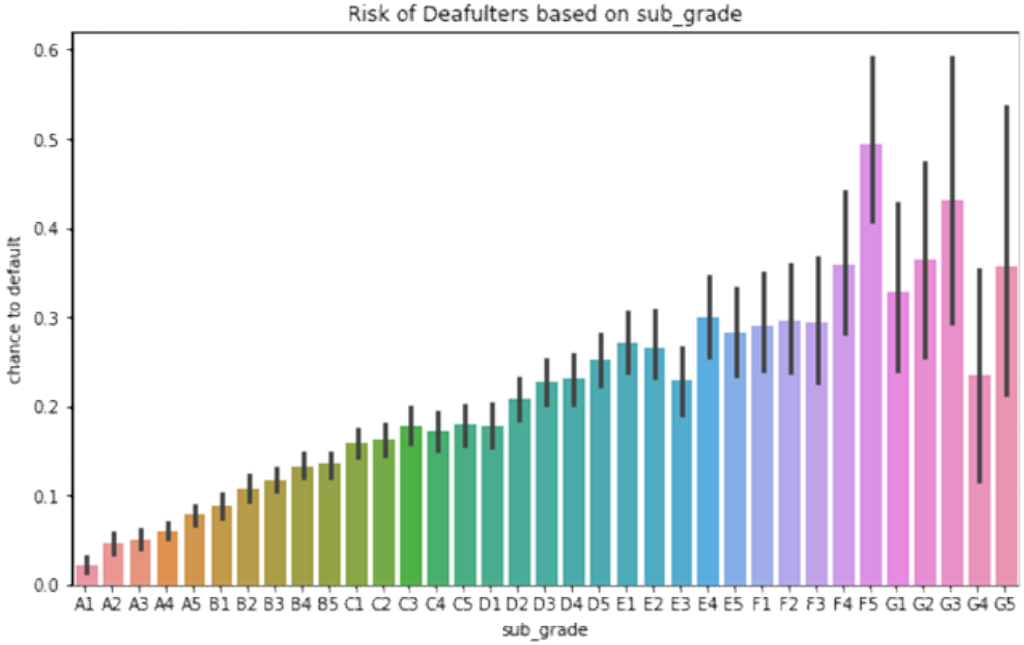
1. Grade vs. Chance of the Applicant to become default.
2. Sub-Grade vs. Chance of the Applicant to become default.

Applicants with Grades A,B,C, D has lower chance to become defaulters in both grades and sub-grades

Recommendation :

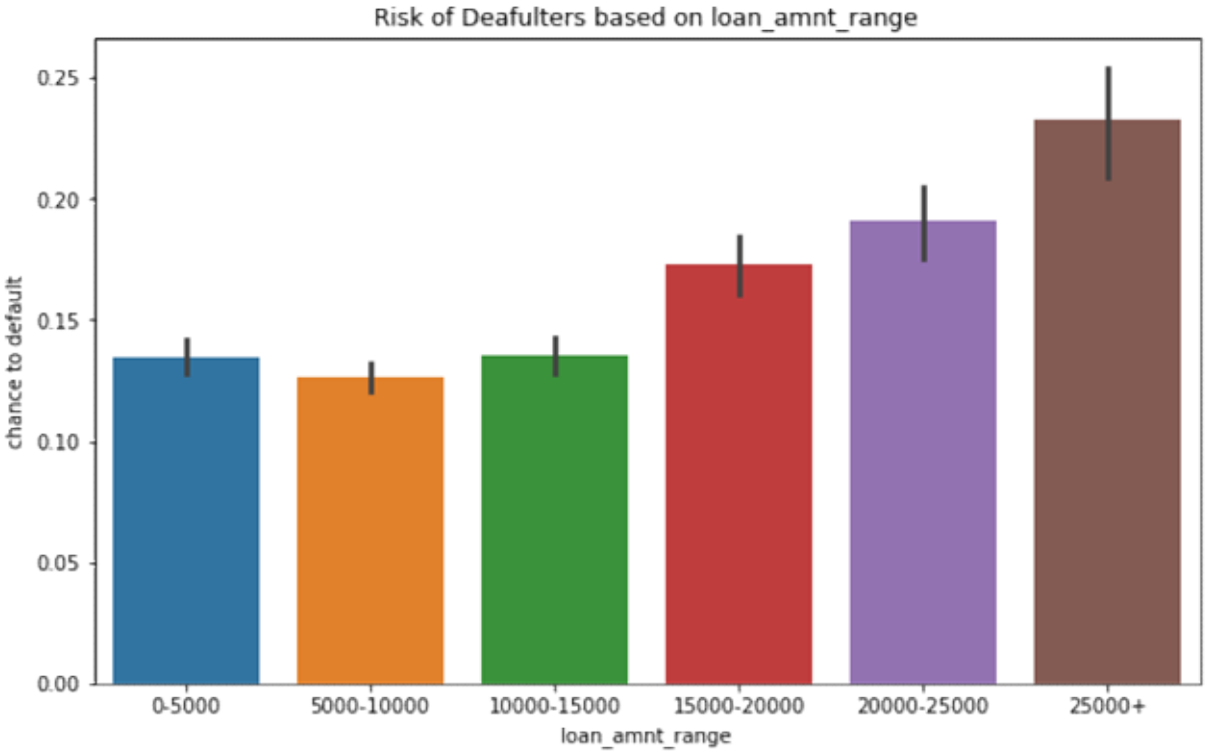
Invest on Applicants with Grades A, B, C, D doing better and pay back debt in time





Features – loan_amnt, int_rate, dti, annual_inc are categorised into ranges so that will get better insight on applicants

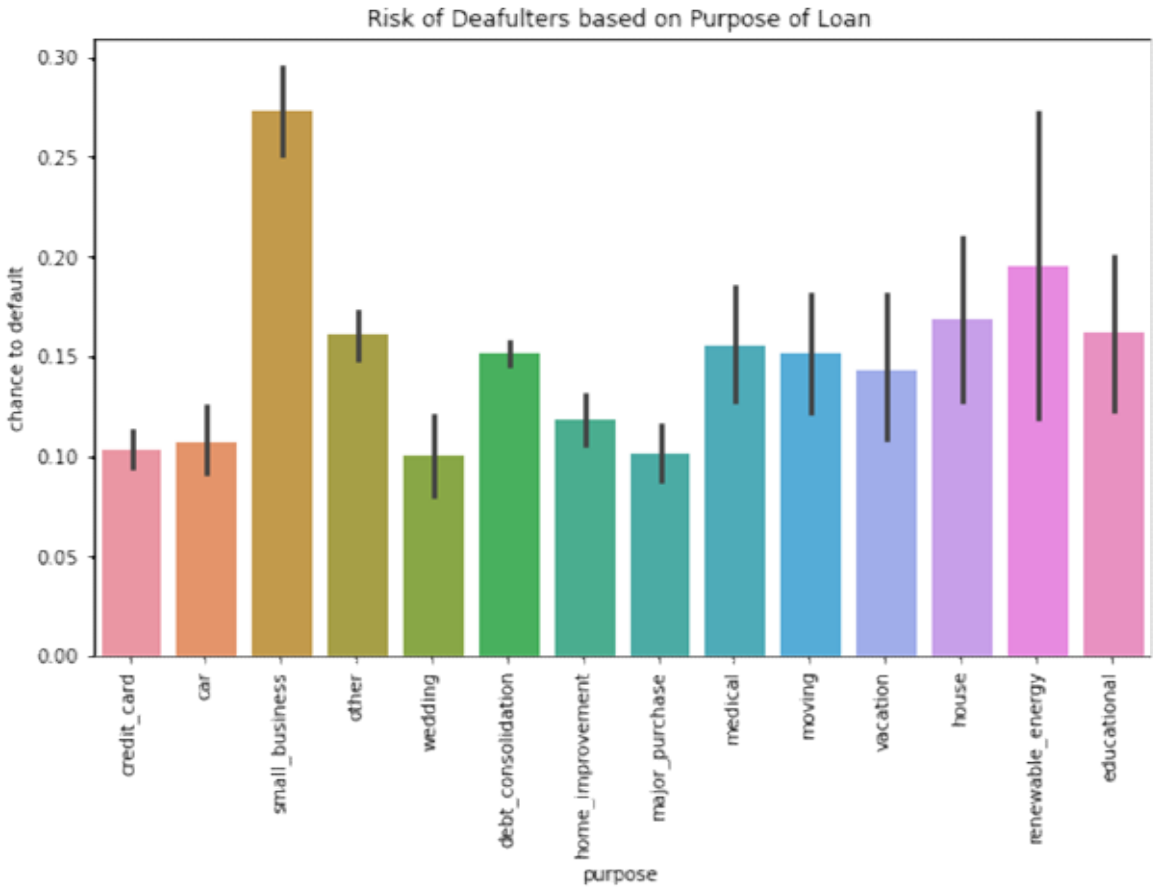
The Below Plot shows that applicants how took the loan in range from 0 to 25000 payed back their loan money with low Defaulter ratio



Recommendation :

Invest on loan amount less than 25000

The below plot shows the purpose of loan applicants to the chance to default by an applicant for particular loan purpose



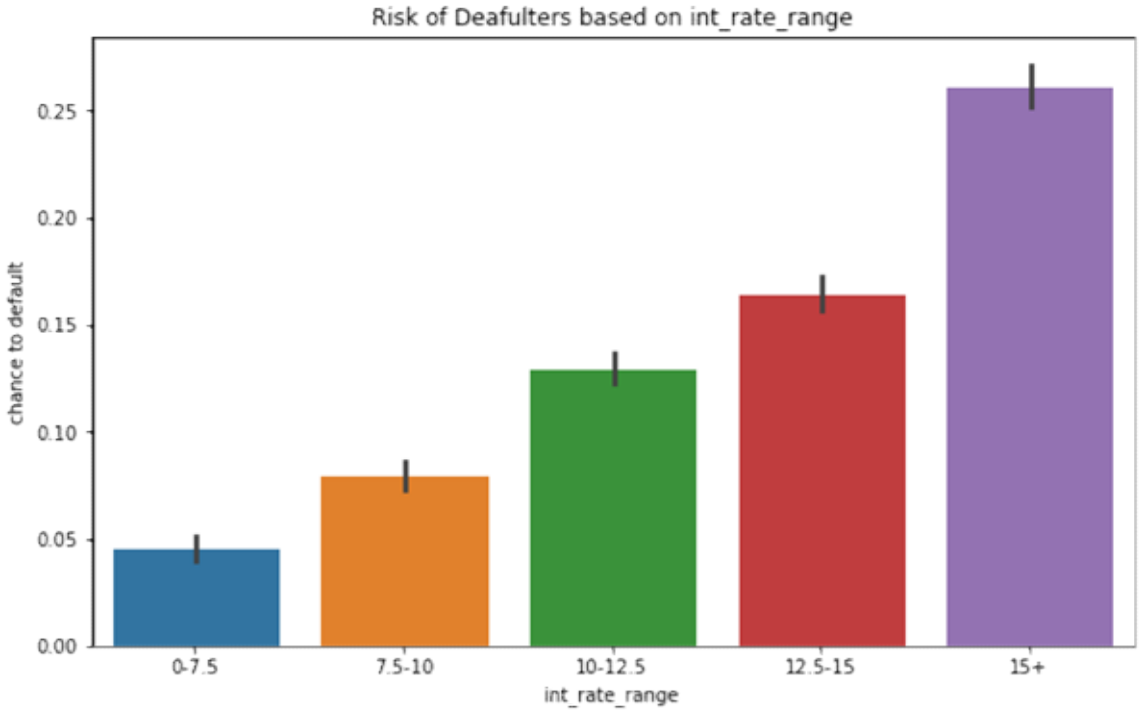
We can infer that loans on Small business and Renewable_energy are not doing well

Recommendation :

Investment on Small business and Renewable_energy has higher Risk.
Rest of them are doing good.

The below plot shows to the relation between the int_rate and chance of applicant to become default.

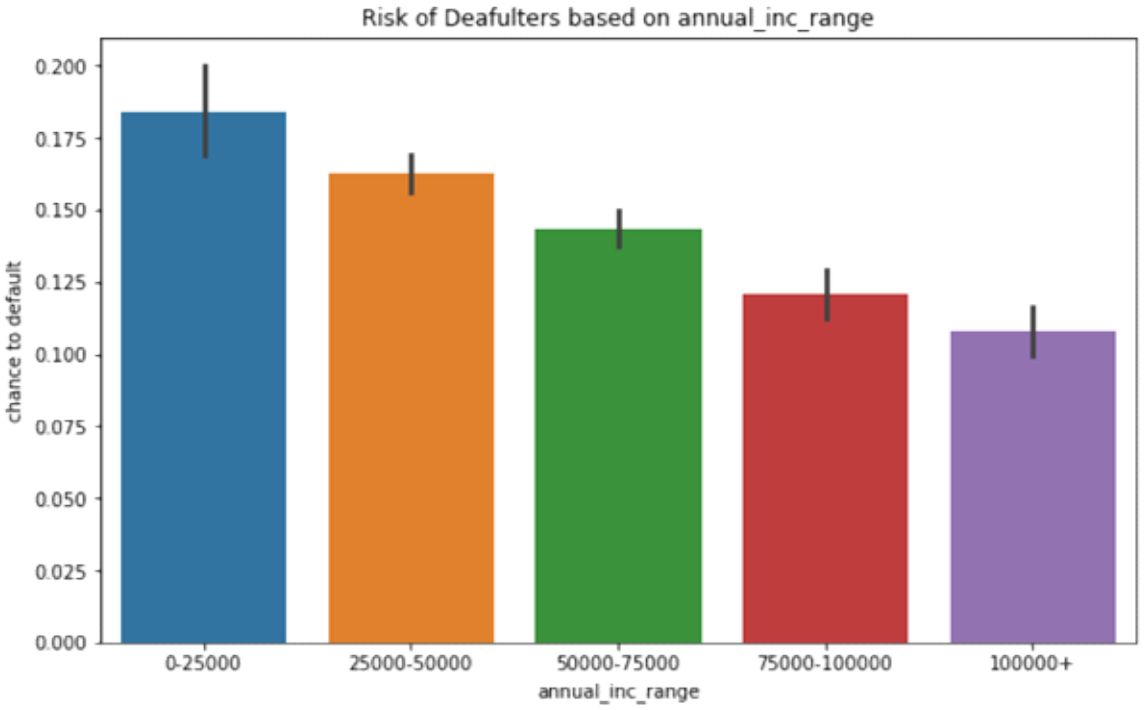
We can infer that:
Interset rate below 15% has the lower deafulters ratio.



Recommendation :

Invest on loans with interest rate
Less that 15% has higher applicants
Paid back their loans on time

The below plot shows the relation between the annual income of an applicant and Chance of an applicant to become defaulter.



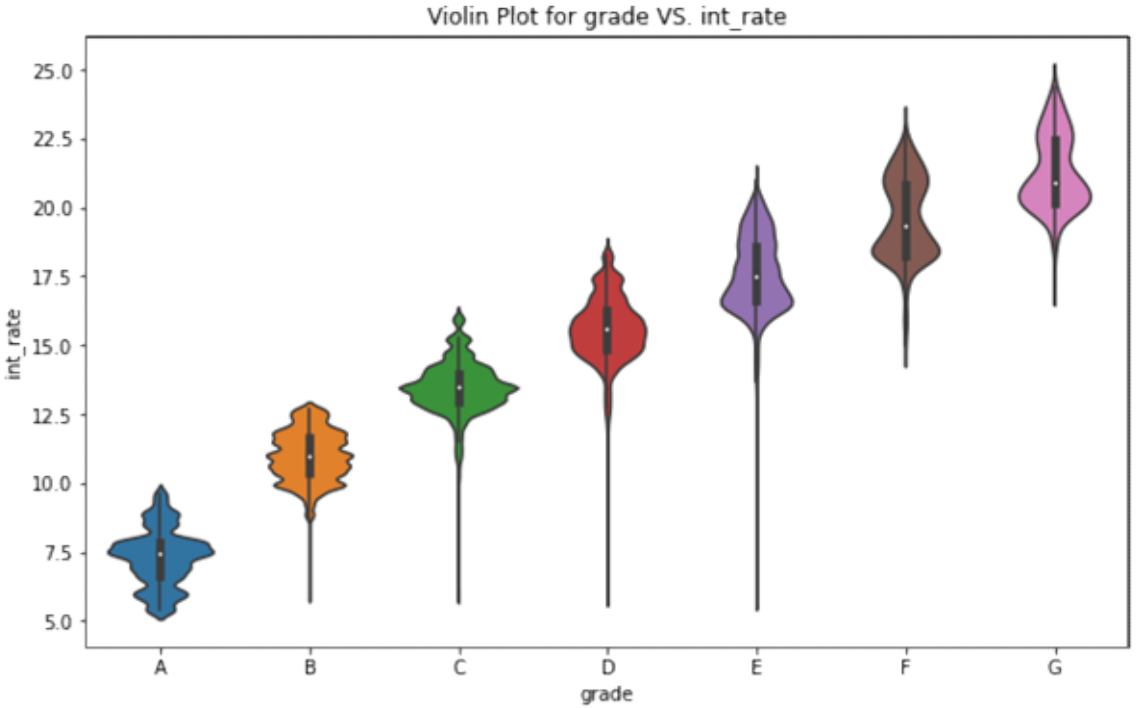
Applicants with annual income between 0-50000 has higher chance of deafulters

Recommendation :

Invest on applicants with annual Income above than 50000.

The Violin Plot between (grade Vs. int_rate)

Applicants with Grade E, F, G has higher interest rate on the loans



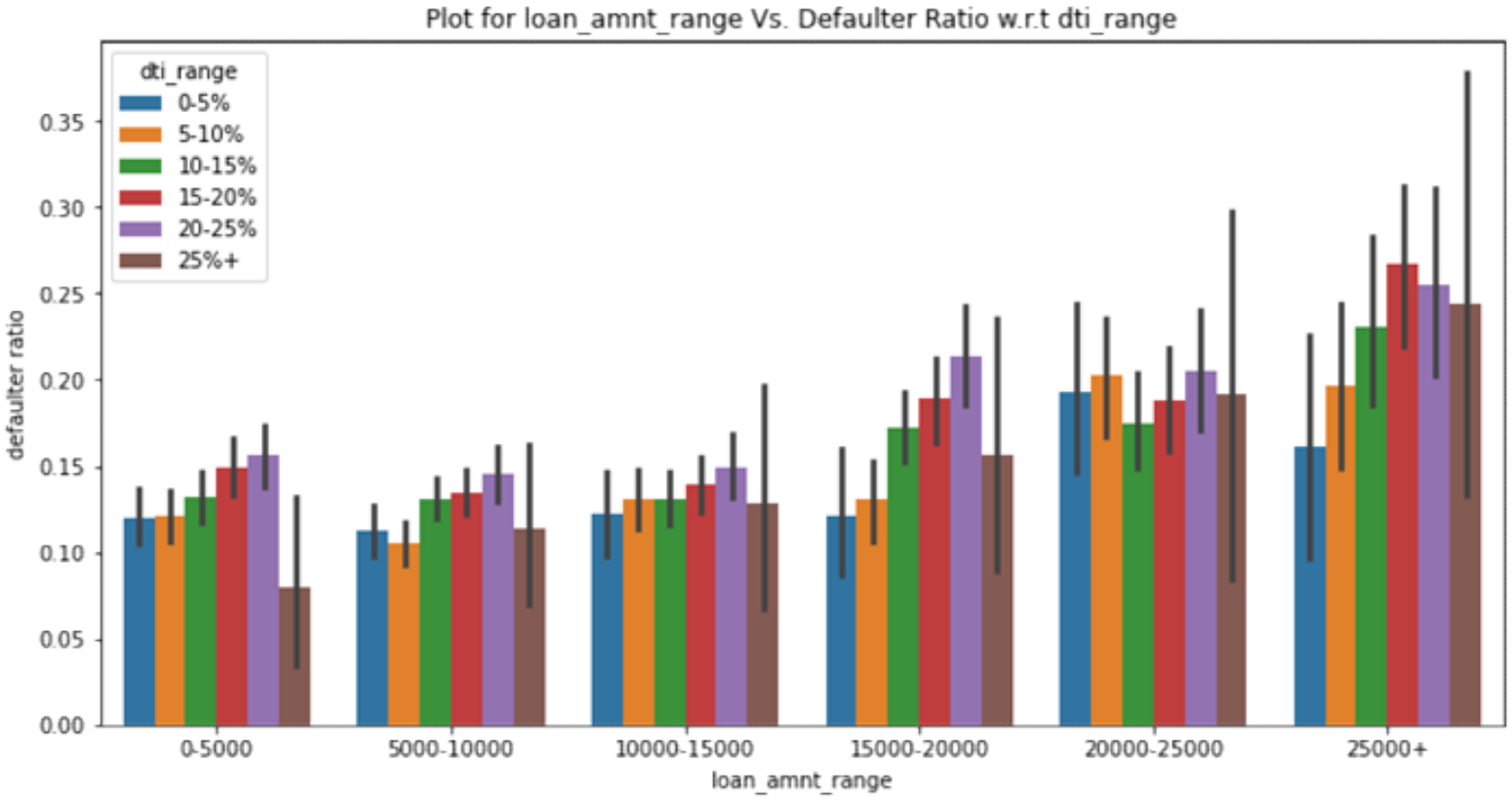
Applicants with grade A, B, C, D has lower interest rates compared to E, F, G

Recommendation :

Invest on applicants who has good Grade A, B, C are more likely to repay their loans.

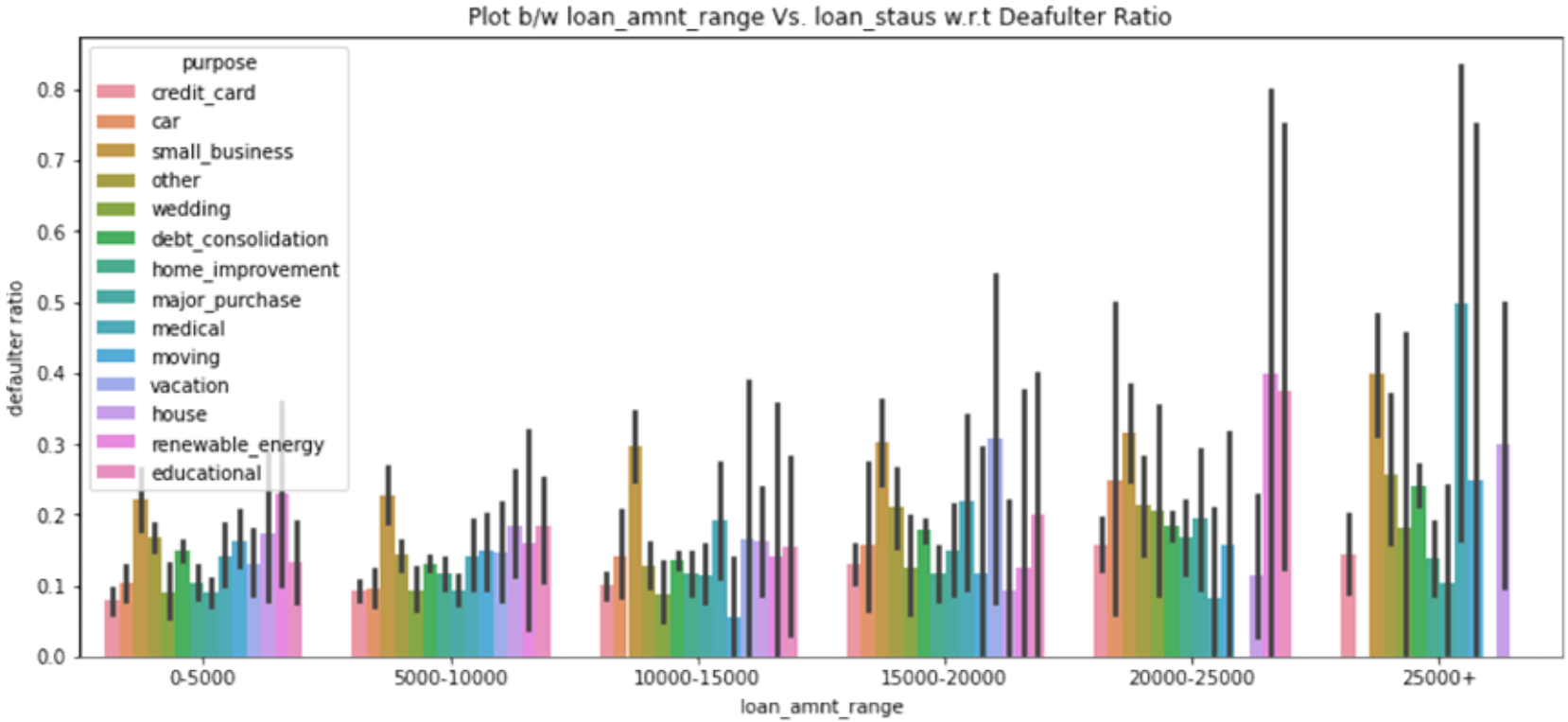
The Below Plot shows relation between loan_amnt and Deafulter ratio of applicants wrt dti of applicant.

We can infer: that applicants how took the loan amount less than 25000 payed back their loan money with low defaulter ratio



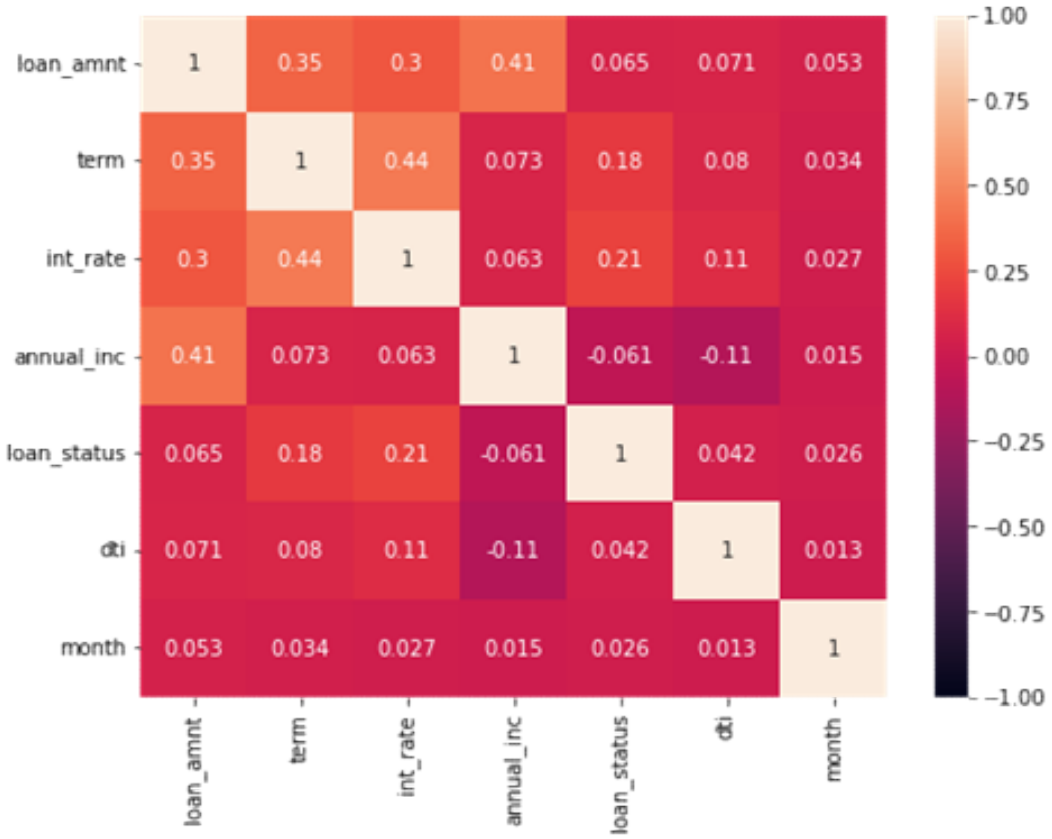
The below plot shows the loan-amnt vs defaulter ratio based on the purpose of applicant's took the loan

We can infer that : loan_amnt rage from 20000 to 25000 and 25000+ has highest variation.



The below heat map shows the correlation between all the features.

We can infer that: Relation between features neither does show the strong positive correlation nor the strong negative correlation between the features
loan_amnt, term, int_rate, annual_inc, loan_status, dti, month



Features that show positive correalation

1. term and int_rate
2. Annual_inc and loan_amnt
3. Term and loan_amnt