Lending Club case study

Module

Created by:

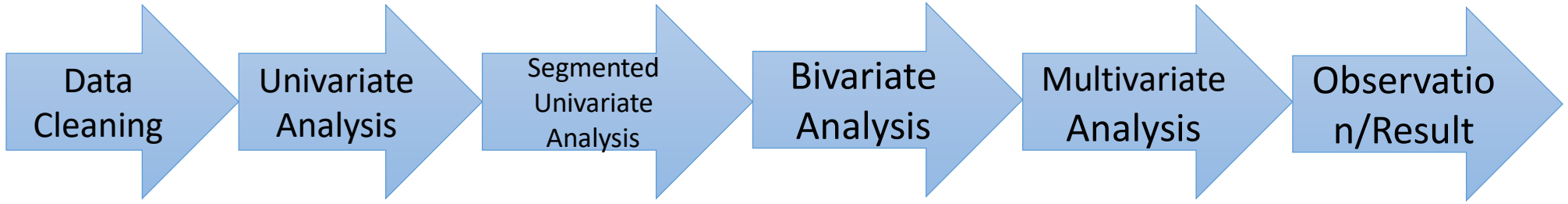
Bhavana Kshirsagar-ML60

Business Problem

- In Financial Organization Loan approval is important process as it can cause business loss if the applicant can repay loan and loans gets rejected and credit loss if applicany cannot pay loan and marked the loan as charged off
- To Define factors or pattern that will incline to applicant as "Defaulter" will help the risk area of the loan approval process
- Pattern about taking action against the risk of loan approval like denying loan, reducing ammountof loan etc.

Objective and Approach

Performed Initial analysis of CSV file provided also check the data dictionary for data column understating and marking the required and unrequired column ad per objective to find defaulted application.



Data Cleaning And Manipulation :

1. Removing null values row and column
2. Removing single value column
3. Removing personal information column
4. Removing the data column required after loan sanction
5. Checking the null value percentage and updating values
6. Giving mindful values for calculative analysis
7. Removing the outlier
8. Data manipulation for making necessary column more readable

Segmented Univariate Analysis:

1. Segmenting the continuous numeric data
2. Segmentation to make categorial type of data
3. Derived column for ordered categorial segments

Derived Bivariate Analysis:

1. Created bins or buckt of data
2. Conitnious data can be segmented and analaysied
3. Bin made of ranges to check factors incline to

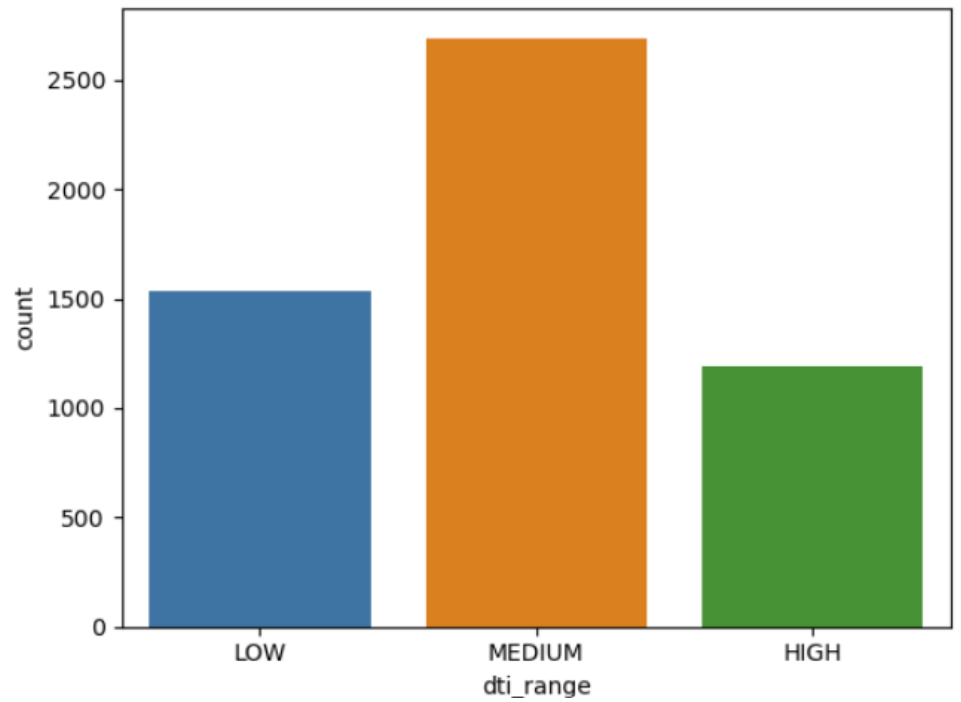
Segmentation

Continuous numeric column as per below logic

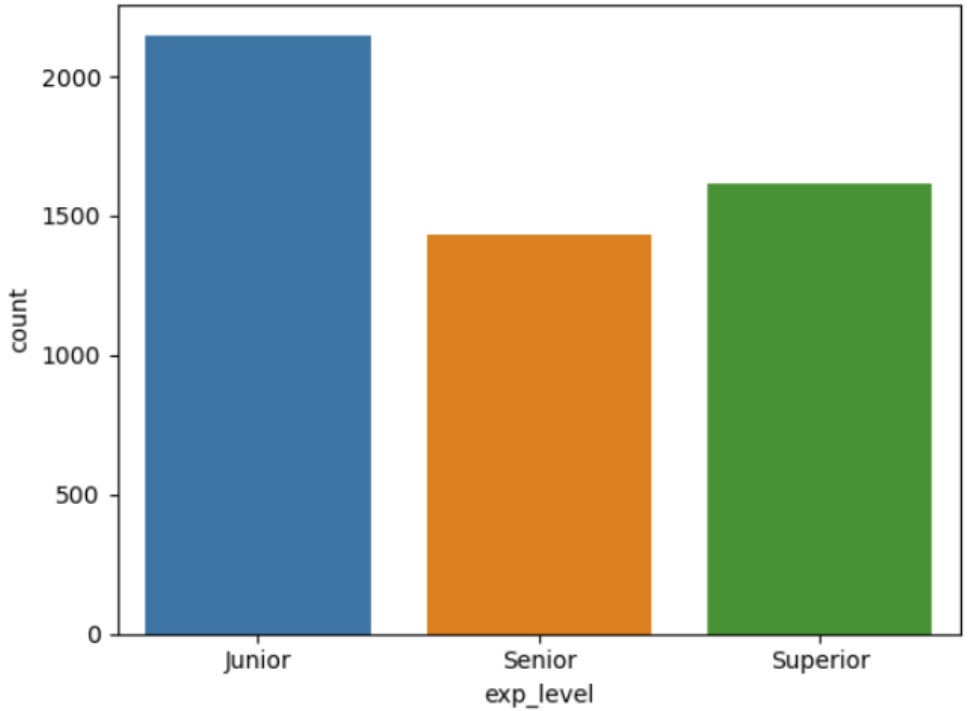
- dti with range(<=10 is Low , dti >10.00 and dti <= 20.00 is Medium,>20 is High).
- Exp_level with range(<=3 is Junior , exp >3 and exp <= 7 is Senior ,>7 is Superior).
- Issue_year derive from IssueDate.

```
sns.countplot(x = 'dti_range', data = loan_df[loan_df.loan_status == 'Charged Off'])
```

```
<Axes: xlabel='dti_range', ylabel='count'>
```




```
<Axes: xlabel='exp_level', ylabel='count'>
```

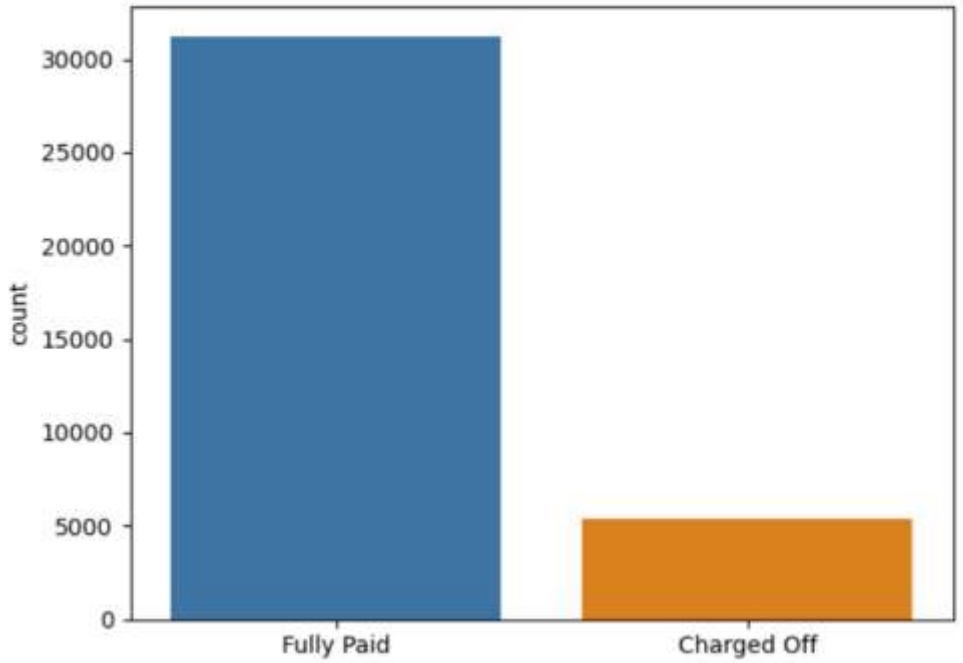


- Applicants with Less experience and tend to default the loan
- Dit-range of medium means segment value 11-20 are tend to default the application

Univariate Analysis:

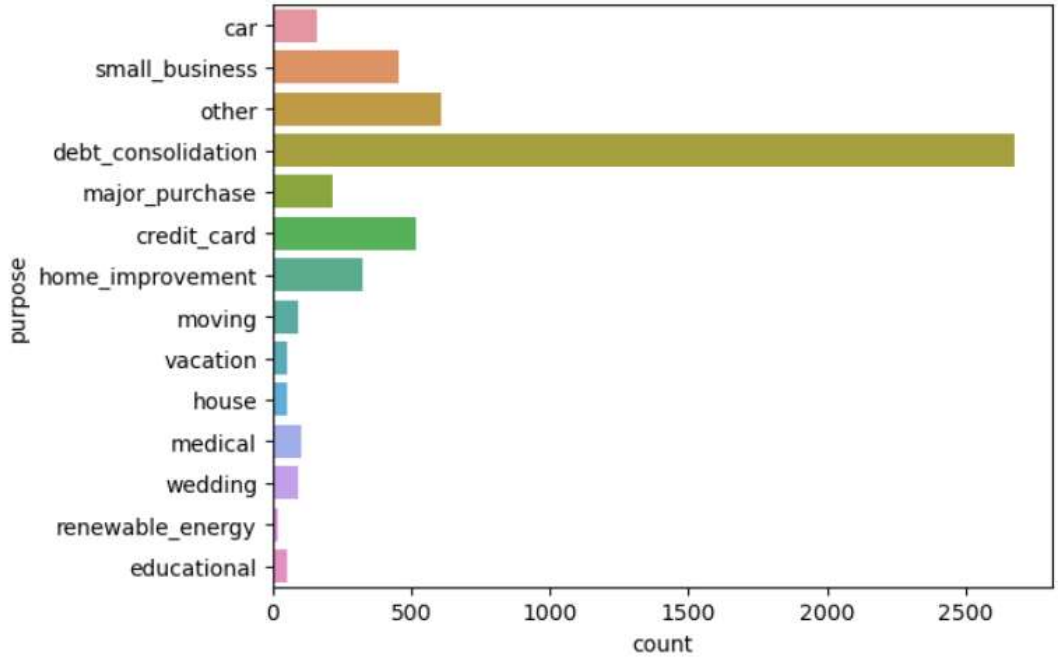
```
sns.countplot(x = 'loan_status', data = loan_df)
```

```
<Axes: xlabel='loan_status', ylabel='count'>
```




```
sns.countplot(y = 'purpose', data = loan_df[loan_df.loan_status == 'Charged Off'])
```

```
<Axes: xlabel='count', ylabel='purpose'>
```



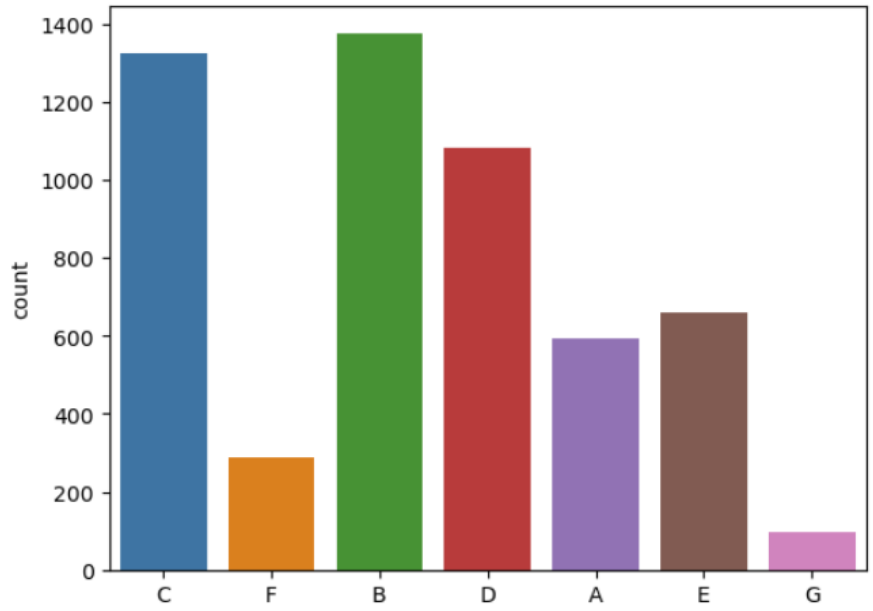
- Debt_consolidation is the purpose to take loan to repay other loan are more tended to fall under the defaulter case

Univariate Analysis:

3 Grade

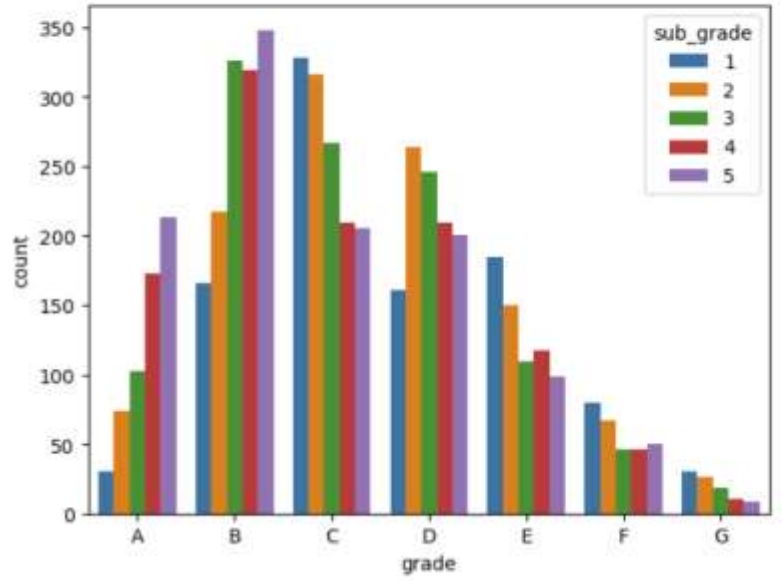
```
sns.countplot(x = 'grade', data = loan_df[loan_df.loan_status == 'Charged Of
```

```
<Axes: xlabel='grade', ylabel='count'>
```




```
sns.countplot(x = 'grade', order = ['A', 'B', 'C', 'D', 'E', 'F', 'G'] , hue = 'sub_grade',
              data = loan_df[loan_df.loan_status == 'Charged Off'])
```

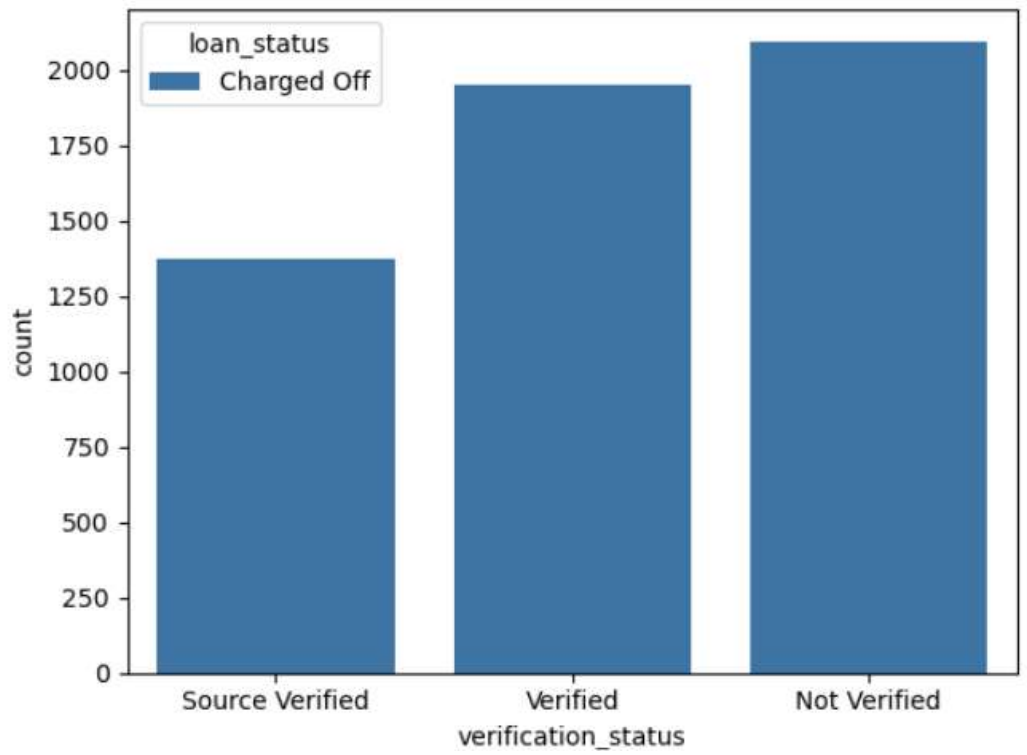
```
<Axes: xlabel='grade', ylabel='count'>
```



- To get more precise idea of grade and sub graded the grade lies under B,C,D are makes defaulters in to that B5,B3 and C! are having respectively higher changes to default

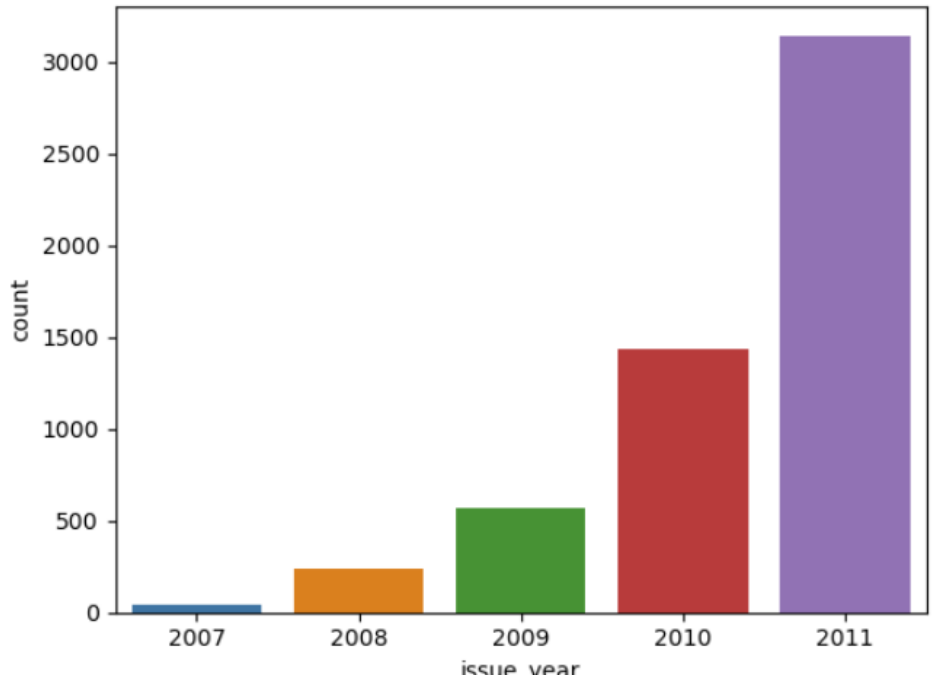
Univariate Analysis:

```
<Axes: xlabel='verification_status', ylabel='count'>
```




```
sns.countplot(x = 'issue_year', data = loan_df[loan_df.loan_status == 'Charged Off'])
```

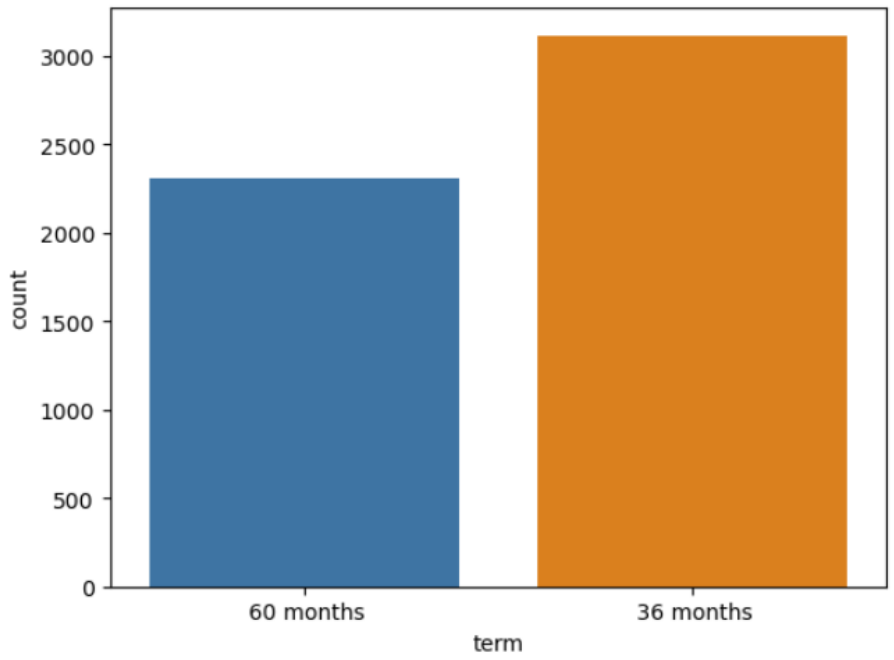
```
<Axes: xlabel='issue_year', ylabel='count'>
```



- Application Which is not verifies and granded loan in 2011 are more likely to default

Univariate Analysis:

<Axes: xlabel='term', ylabel='count'>



<Axes: xlabel='home_ownership', ylabel='count'>

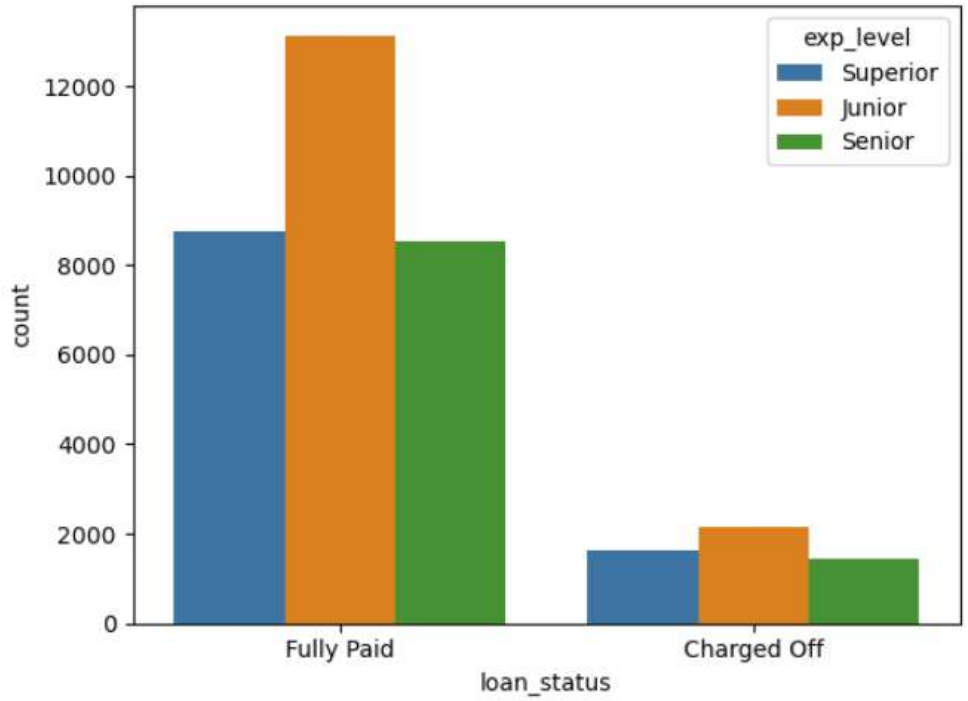
- 36 month loan is having more charge of loans than 60
- Also Rent and mortgare category of home ownership is incline to default

Univariate Anlaysis:

Experiance level and Loan status

```
sns.countplot(x = 'loan_status', hue = 'exp_level',
              data = loan_df[loan_df.loan_status != 'Current'])
```

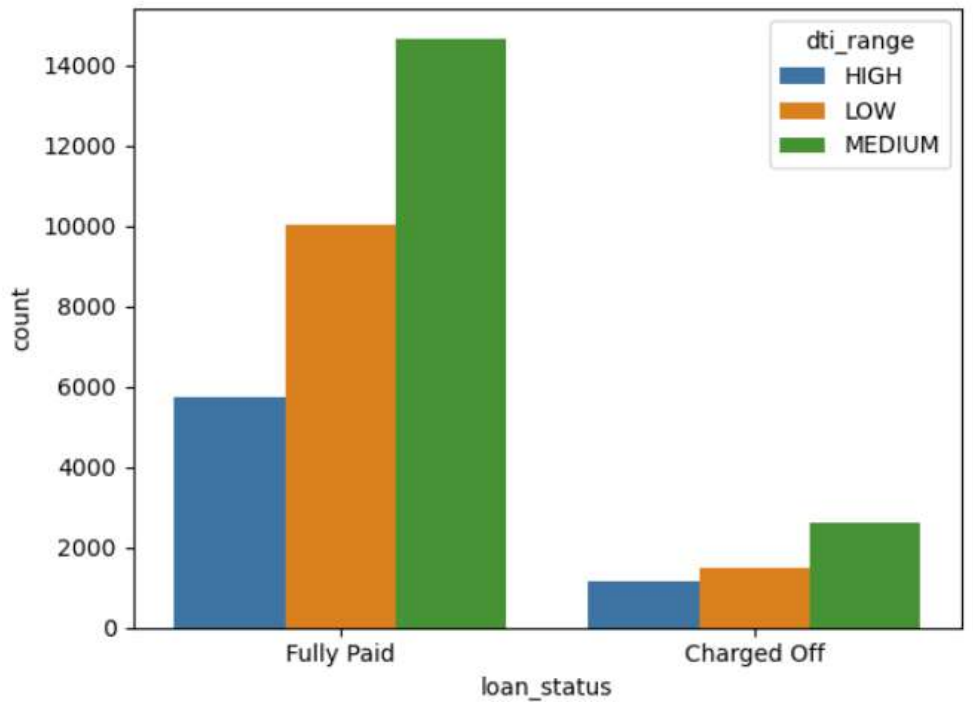
```
<Axes: xlabel='loan_status', ylabel='count'>
```



2 DTI range and Loan status

```
sns.countplot(x = 'loan_status', hue = 'dti_range',
              data = loan_df[loan_df.loan_status != 'Current'])
```

```
<Axes: xlabel='loan_status', ylabel='count'>
```

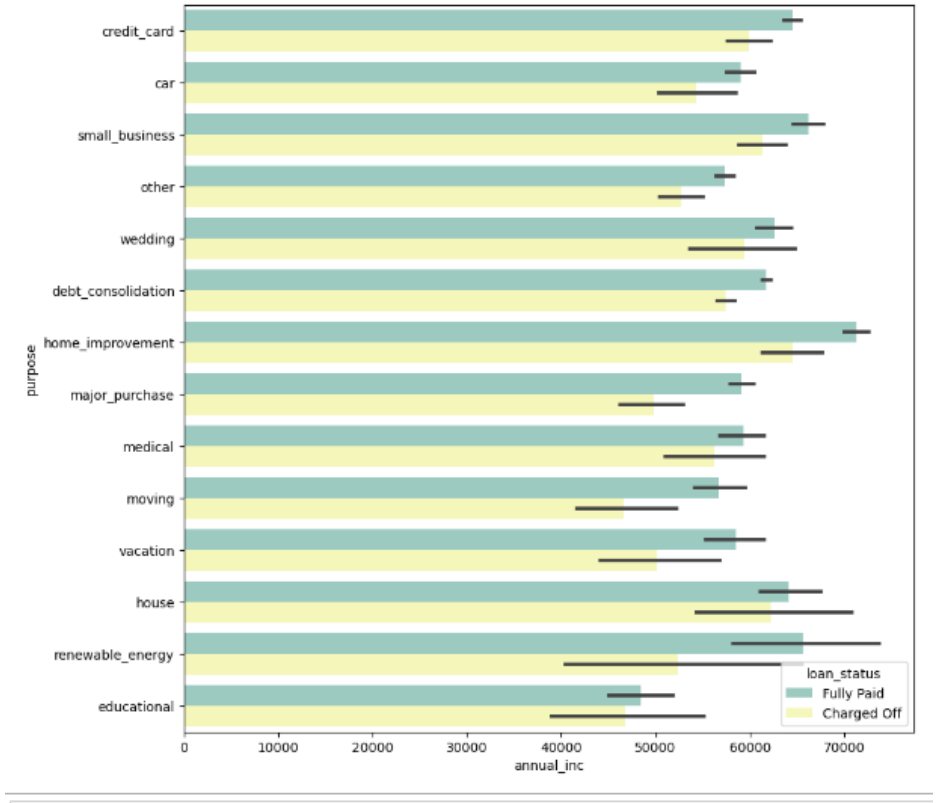


As group of Charged off and fully paid of loan status can be defauted when experience of applicant is low and dti rage is Medium

Bivariate Anlaysis:

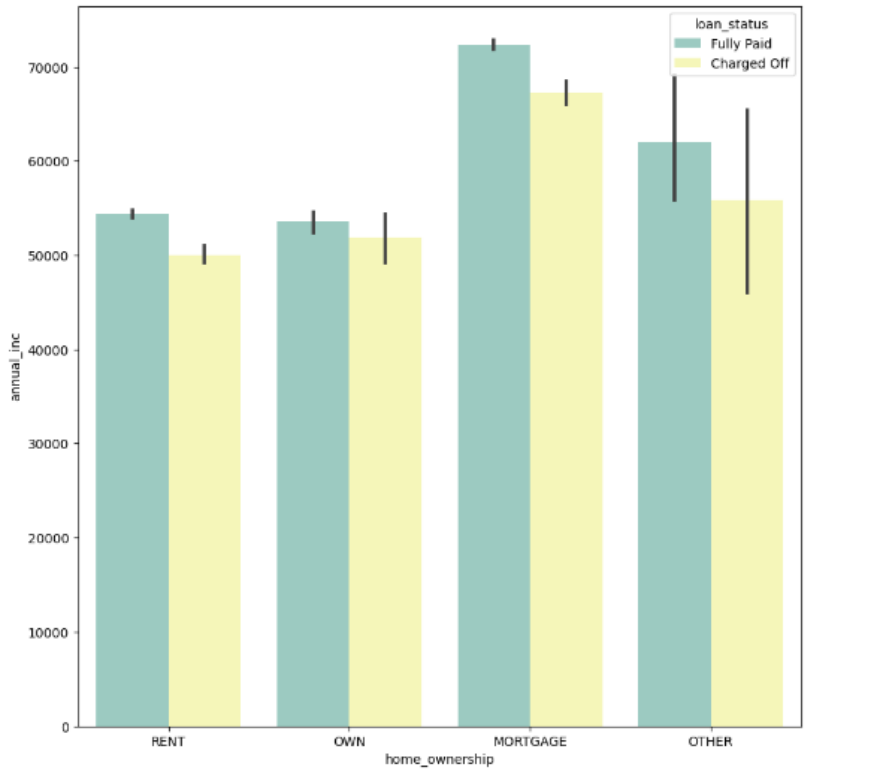
3 Annual income and purpose

```
plt.figure(figsize=(10,10))
sns.barplot(data =loan_df[loan_df.loan_status!='Current'],x='annual_inc', y='purpose', hue ='loan_status',palette="Set3")
plt.show()
```



4 Annual income and Home Ownership

```
plt.figure(figsize=(10,10))
sns.barplot(data =loan_df[loan_df.loan_status!='Current'],x='home_ownership', y='annual_inc', hue ='loan_status',palette="S
plt.show()
```

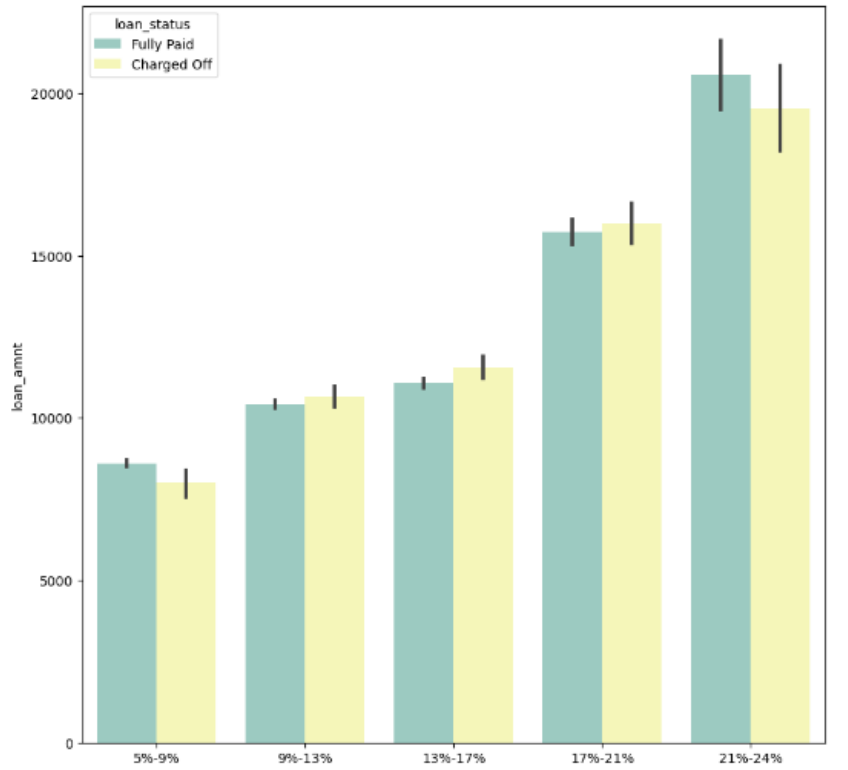


Home_immprovement are likely to take by higher income group then debt_consolidation ,also higher income group applicant has more Mortgage properties

Bivariate Anlaysis:

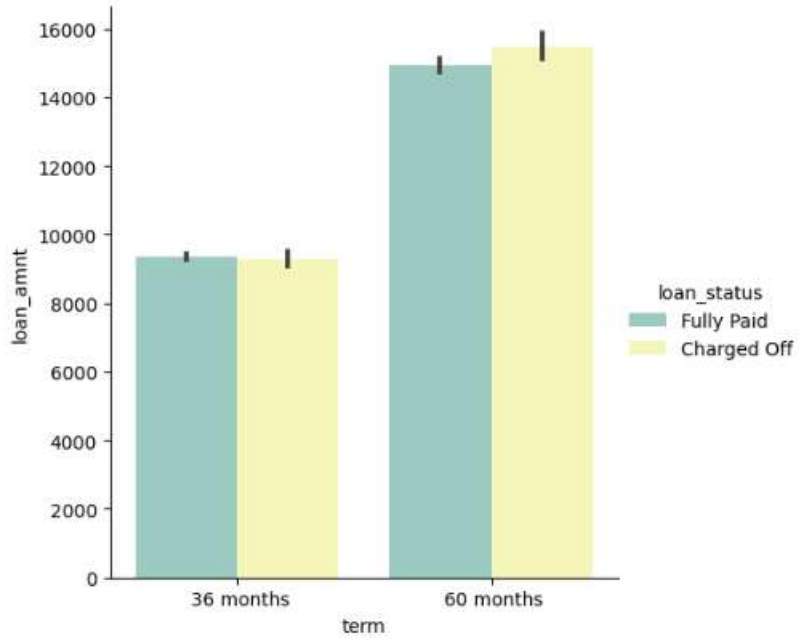
5 Interatest rate groups and Loan ammount

```
plt.figure(figsize=(10,10))
sns.barplot(data =loan_df[loan_df.loan_status!='Current'],x='int_rate_groups', y='loan_amnt', hue ='loan_status',p
plt.show()
```



6 Term and Loan ammount

```
sns.catplot(x = 'term', y = 'loan_amnt', data = loan_df,hue = 'loan_status', ki
<seaborn.axisgrid.FacetGrid at 0x23afae2fad0>
```

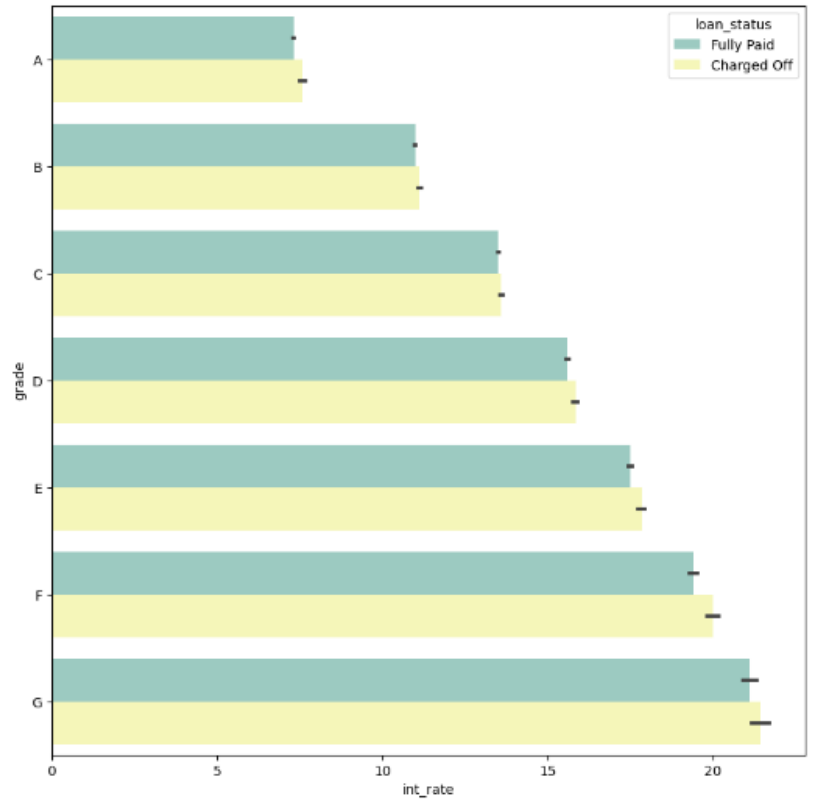


- Applicants with high loan amount tend to have more term and interest rate

Bivariate Anlaysis:

7 Interest rate and Grade

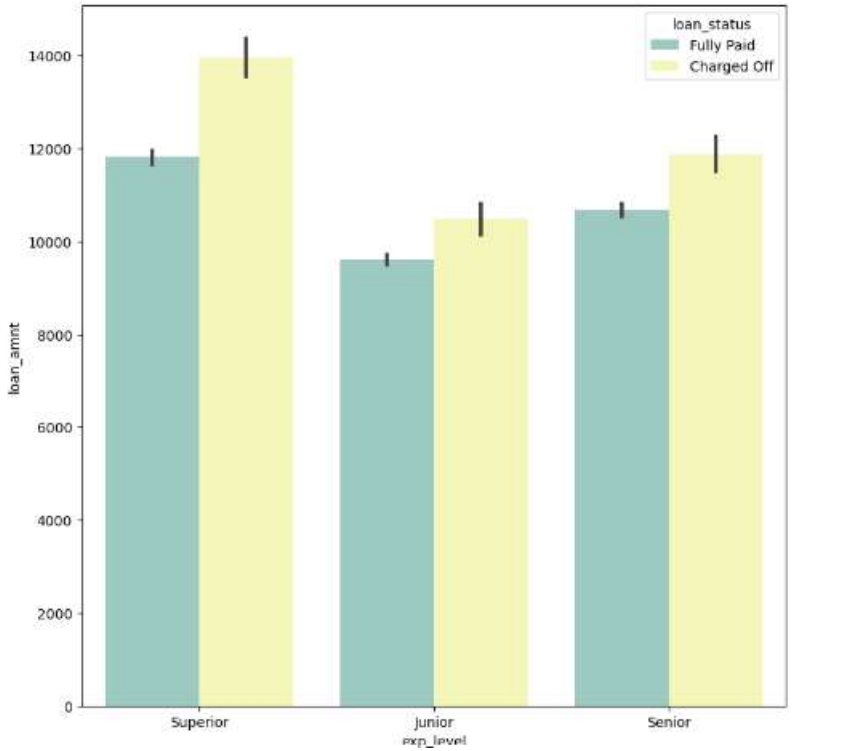
```
plt.figure(figsize=(10,10))
sns.barplot(data =loan_data,x='int_rate', y='grade', hue ='loan_status',palette="Set3", order=['A','B','C','
plt.show()
```



8 Loan ammount amd Experience level

```
plt.figure(figsize=(20,20))
plt.subplot(221)
sns.barplot(data =loan_df,y='loan_amnt', x='exp_level', hue ='loan_status',palette="Set3")
```

```
<Axes: xlabel='exp_level', ylabel='loan_amnt'>
```



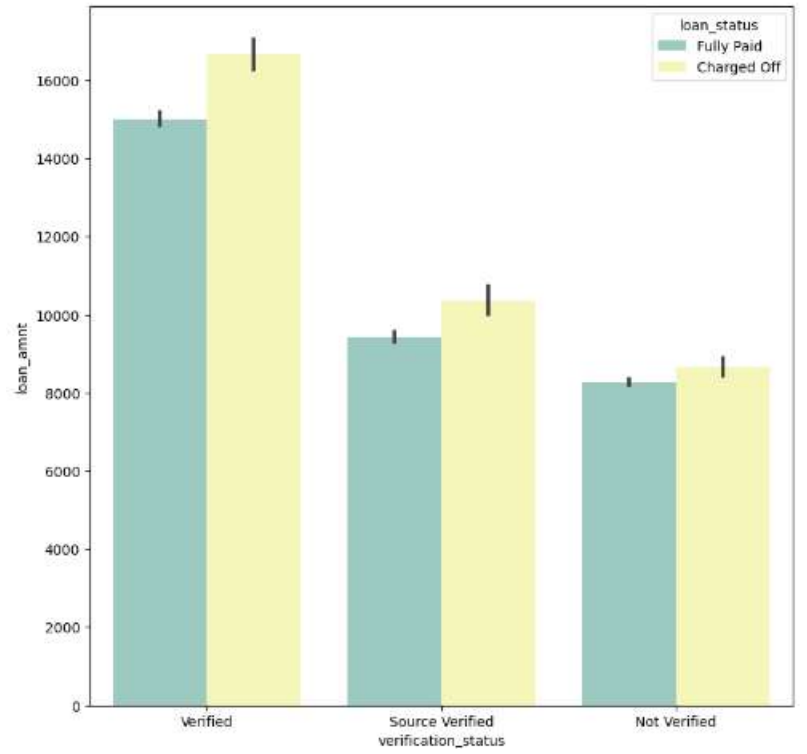
- interest rate for G grade seems to be more as the univariate analaysis the B grade are more tend to default
- Applicants which has more experience tend to take more loanas annual income will be high

Bivariate Anlaysis:

9 Loan ammount and Verification status

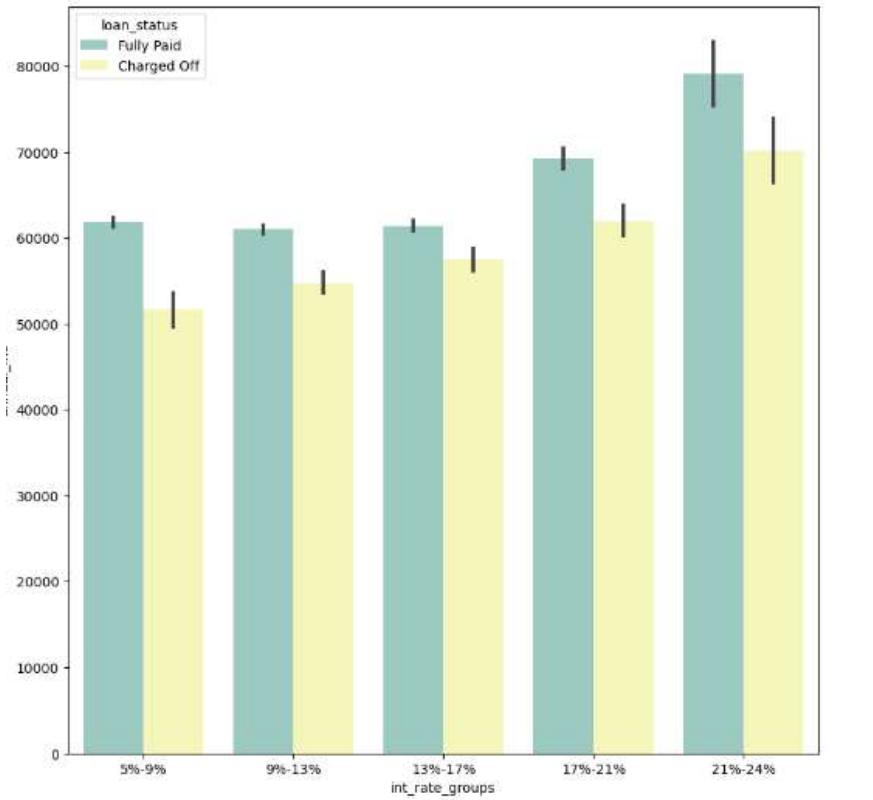
```
plt.figure(figsize=(20,20))
plt.subplot(222)
sns.barplot(data =loan_df,y='loan_amnt', x='verification_status', hue ='loan_status',palette="Set3")
```

```
<Axes: xlabel='verification_status', ylabel='loan_amnt'>
```



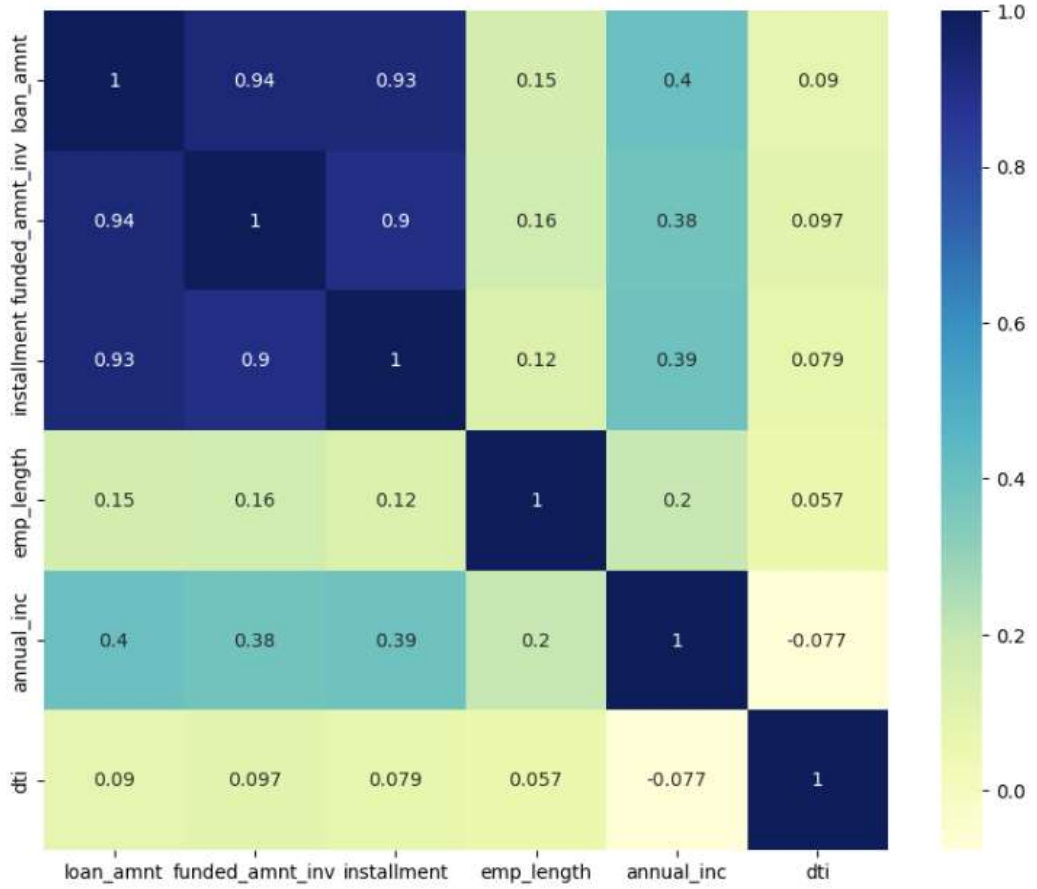
Annual income and Interest rate

```
.figure(figsize=(10,10))
.barplot(data =loan_df,x='int_rate_groups', y='annual_inc', hue ='loan_status',palette="Set3")
.show()
```



- Less loam amount is sanction to not verified applicants
- Annual income high is proporstinarte to interest rate

Multivariate Anlaysis:



As per the correlation the 0 is neutral-1 is negative and 1 is positive relation amount the entity as we can see in diagram 0.94 and 0.93 are respectively is positively corelated the loan amount - funded amount and loan amount - installments

As per observation Factors or variable to Defaulted application

- Applicants having Less work experience
- Applicants having DTI value 11-20 range
- Applicants having house_ownership as 'RENT'
- Applicants who use the loan to clear other debts
- Term of 36 months
- When the loan status is Not verified
- When the purpose is 'debt_consolidation'
- Grade is and a total grade of 'B5' level.
- Applicants taking loan for 'home improvement' and have income of 60k - 70k
- Applicants whose home ownership is 'MORTGAGE and have income of 60-70k
- Applicants who receive interest at the rate of 21-24% and have an loan amount of 70k-80k
- Applicants who have taken a loan in the range 15k - 20k and are charged interest rate of 21-24%
- Applicants whose experience level is ‘Junior ’= less than 3 years and have loan of 12-14k
- When the loan is verified and loan amount is above 15k
- For grade G and interest rate above 20%