Lending_Club_Case_Study_Project_Ajit_Siddhant
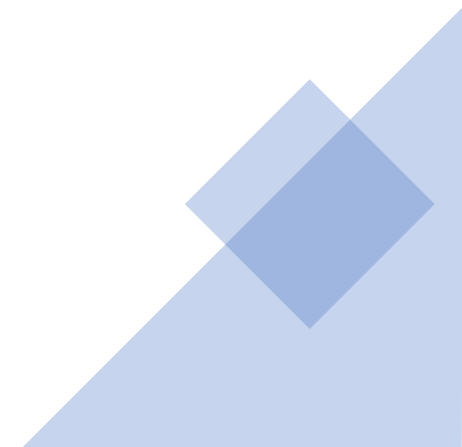
Project done by :-
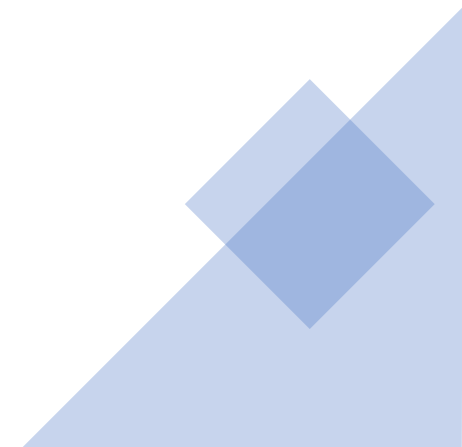
Ajit John

Siddhant Srivastava

Lending Club :EDA case study

- There are four major parts that are needed to be done for this case study:
  - 1. Data understanding
  - 2. Data cleaning (cleaning missing values, removing redundant columns etc.)
  - 3. Univariate analysis
  - 4. Bivariate analysis
  - 5.Observations and Results

Data Cleaning

1. Firstly, we are going to check the percentage of missing values

2. Remove all those with very high missing percentage

3. For columns with less missing percentage: perform Imputations

4. Then, we will just identify the correct metric to impute the column.

5. Finally, we will be dropping rows where the missing percentage is quite high

Data Analysis

i. The objective is to identify predictors of default so that at the time of loan application, we can use those variables for approval/rejection of the loan.

  i. There are broadly three types of variables –
  ii. those which are related to the applicant (demographic variables such as age, occupation, employment details etc.),
  iii. Loan characteristics (amount of loan, interest rate, purpose of loan etc.) and
  iv. Customer behavior variables (those which are generated after the loan is approved such as delinquent 2 years, revolving balance, next payment date etc.).

ii. Now, the customer behavior variables are not available at the time of loan application, and thus they cannot be used as predictors for credit approval.

iii. The ones marked 'current' are neither fully paid not defaulted, so get rid of the current loans. Also, tag the other two values as 0 or 1 to make your analysis simple and clean.

Data Analysis

- **Data Analysis: Univariate Analysis**

- For univariate analysis, you may check the default rate across various categorical features.

- For continuous features, you may perform binning and then you may perform univariate analysis.

- **Data Analysis: Bivariate Analysis**

- Here you may choose two or more features to understand the Default variable
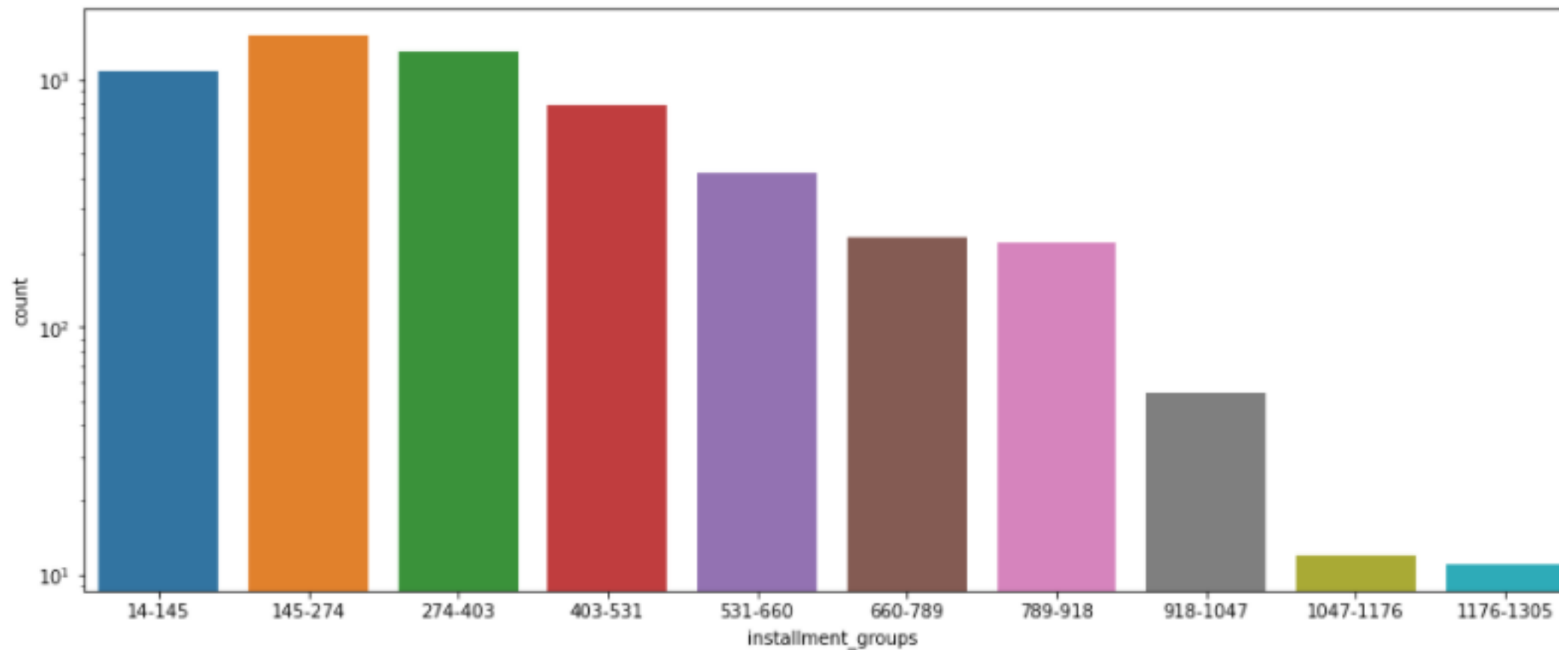
Out[65]: `<AxesSubplot:xlabel='issue_year', ylabel='count'>`
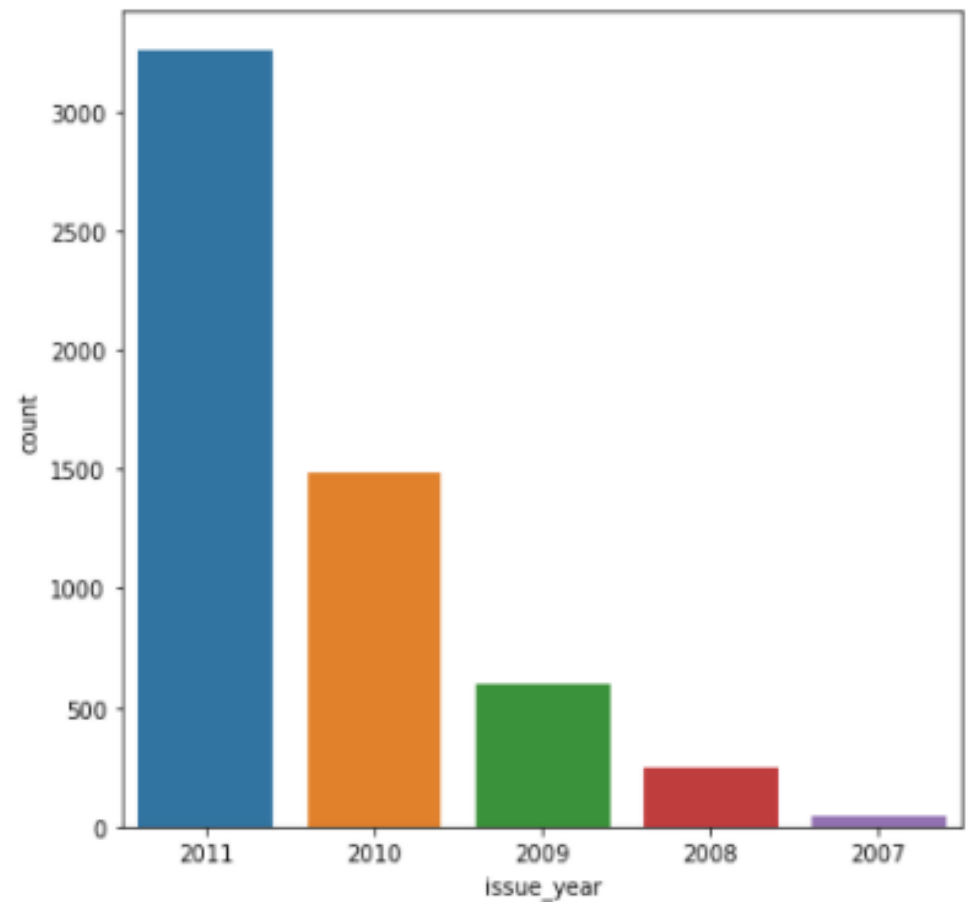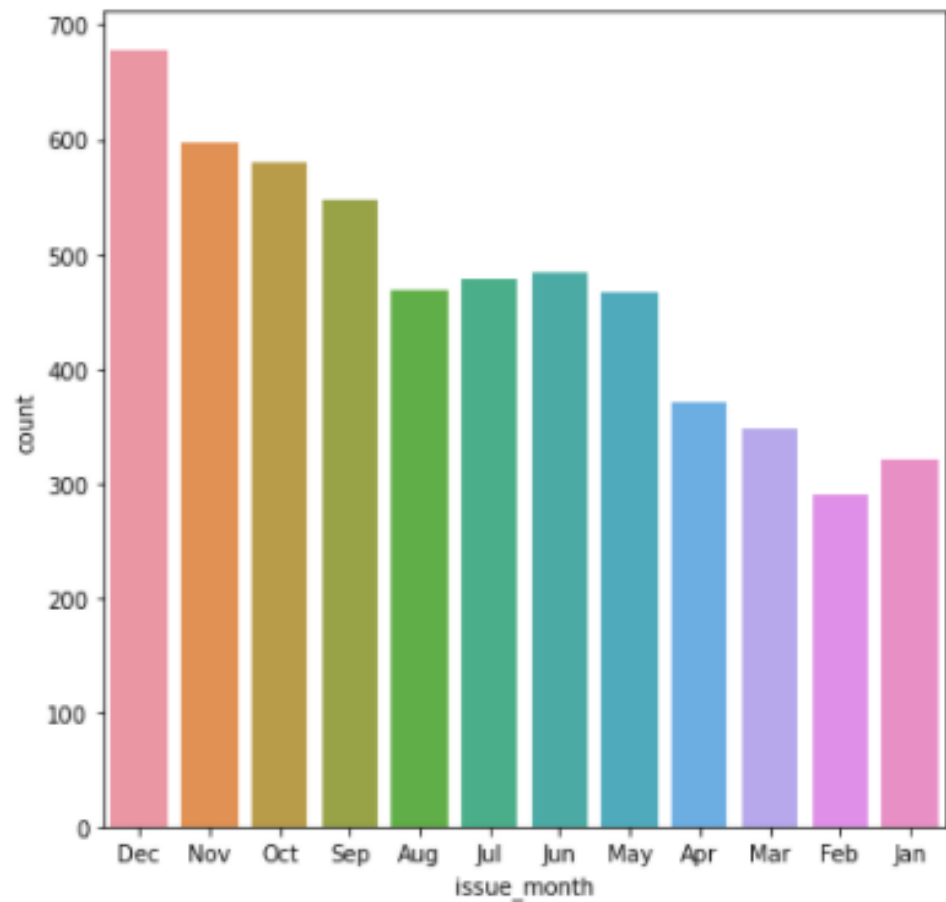
Maximum number of defaults occured when the loan was sanctioned/issued in Dec. Loan issued in the year 2011 were also as compared to other years
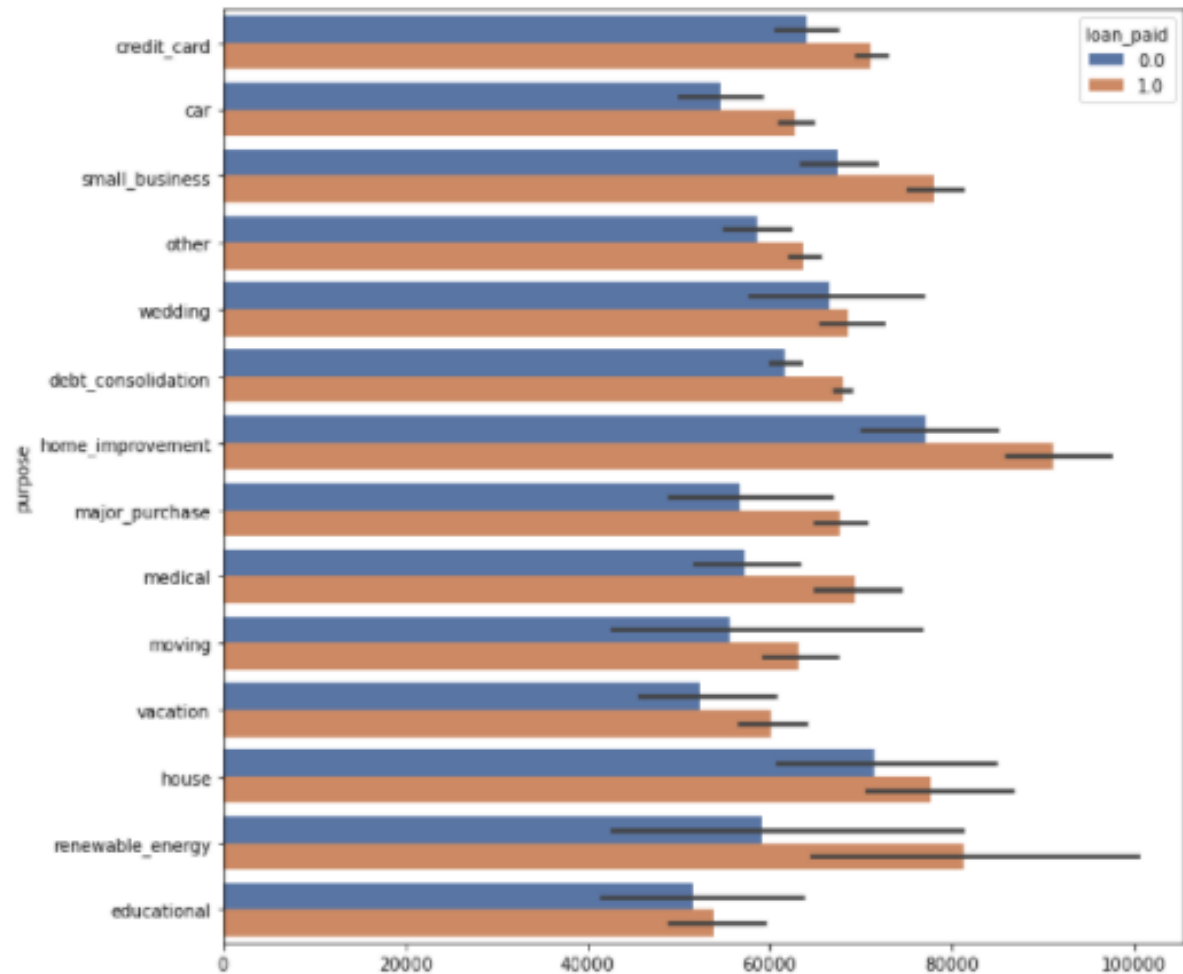Analyzing installment,dti, loan_amnt

```
In [69]: fig,ax = plt.subplots(figsize = (15,6))
         ax.set_yscale('log')
         sns.countplot(x='installment_groups', data=loan_data[loan_data['loan_paid']==0])
```

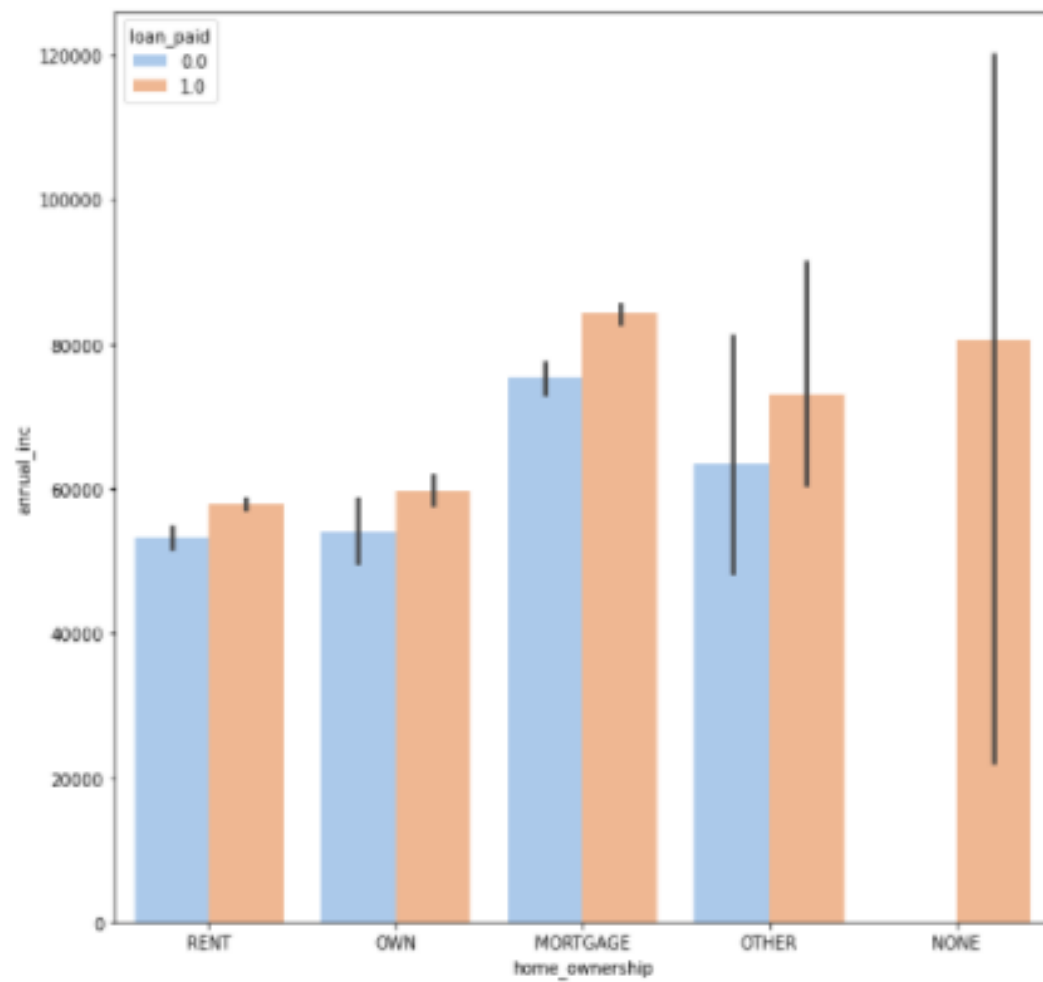Out[69]: <AxesSubplot:xlabel='installment_groups', ylabel='count'>

1.Annual income vs loan purpose

```
In [71]: plt.figure(figsize=(10,10))
         sns.barplot(data =loan_data,x='annual_inc', y='purpose', hue ='loan_paid',palette="deep")
         plt.show()
```
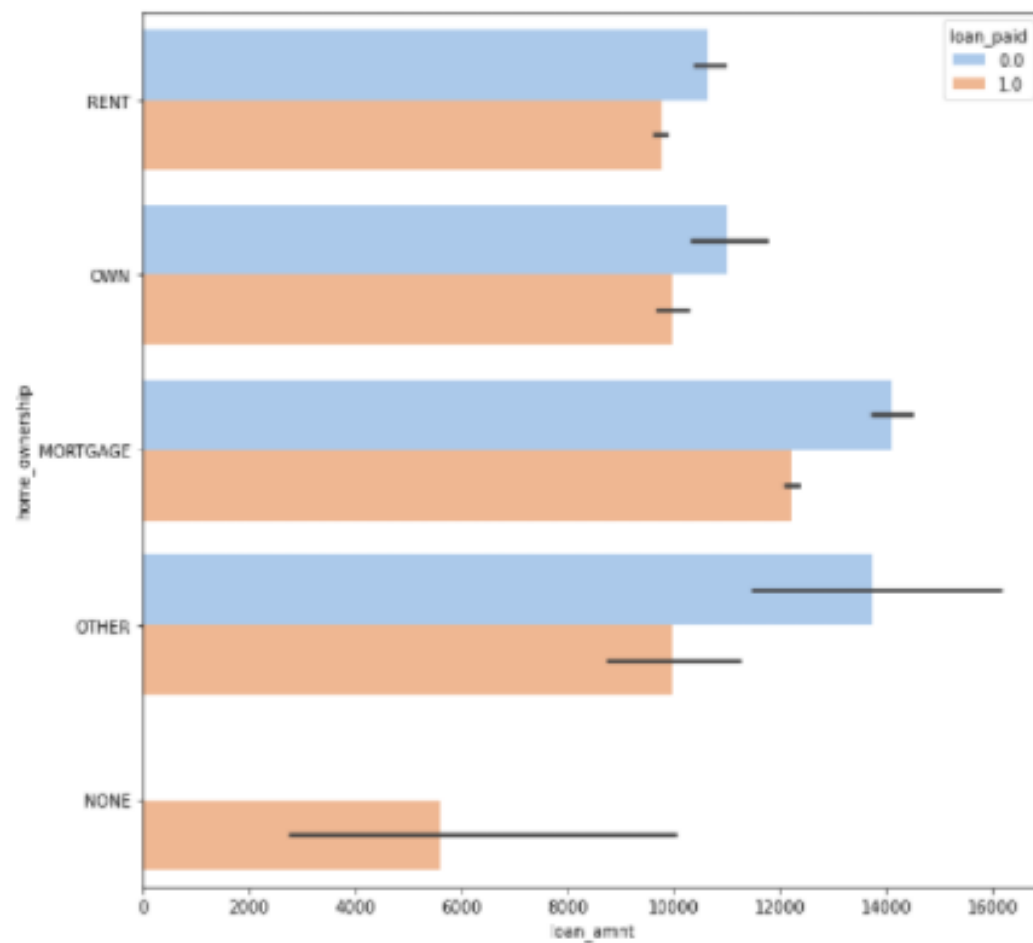
```
In [73]: plt.figure(figsize=(10,10))
         sns.barplot(data =loan_data,x='home_ownership', y='annual_inc', hue ='loan_paid',palette="pastel")
         plt.show()
```
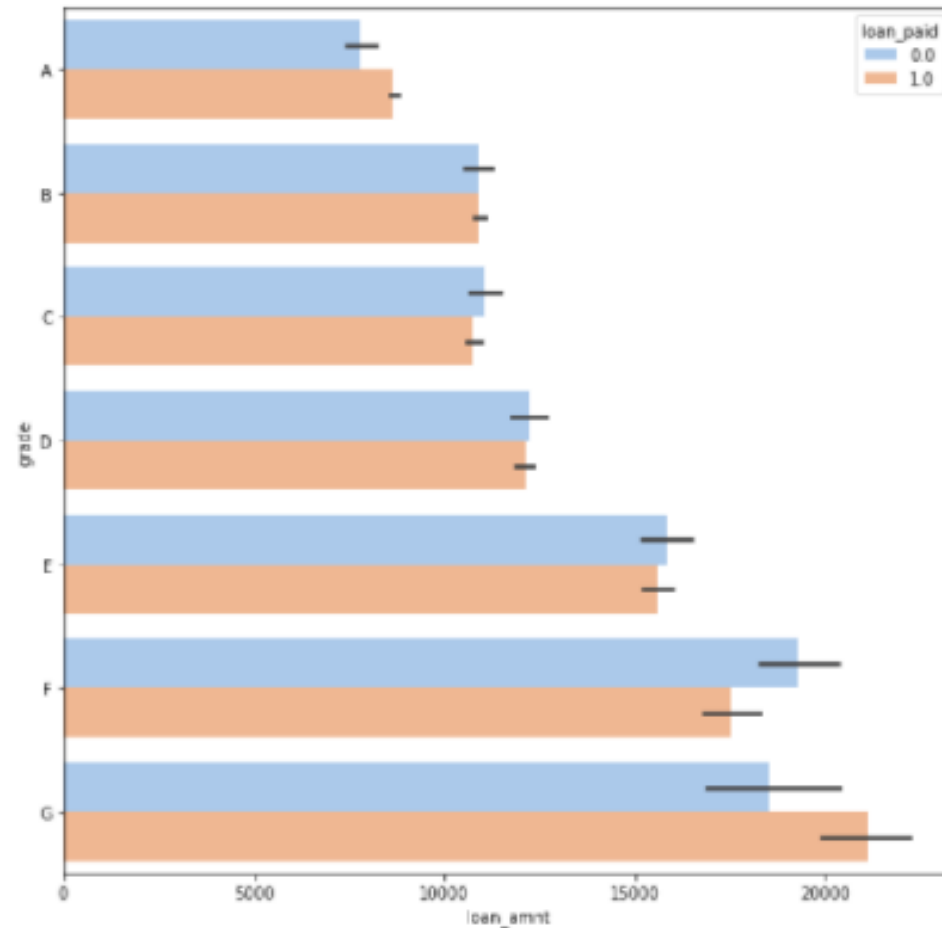
3.Loan vs House Ownership

```
In [80]: plt.figure(figsize=(10,10))
         sns.barplot(data =loan_data,x='loan_amnt', y='home_ownership', hue ='loan_paid',palette="pastel")
         plt.show()
```
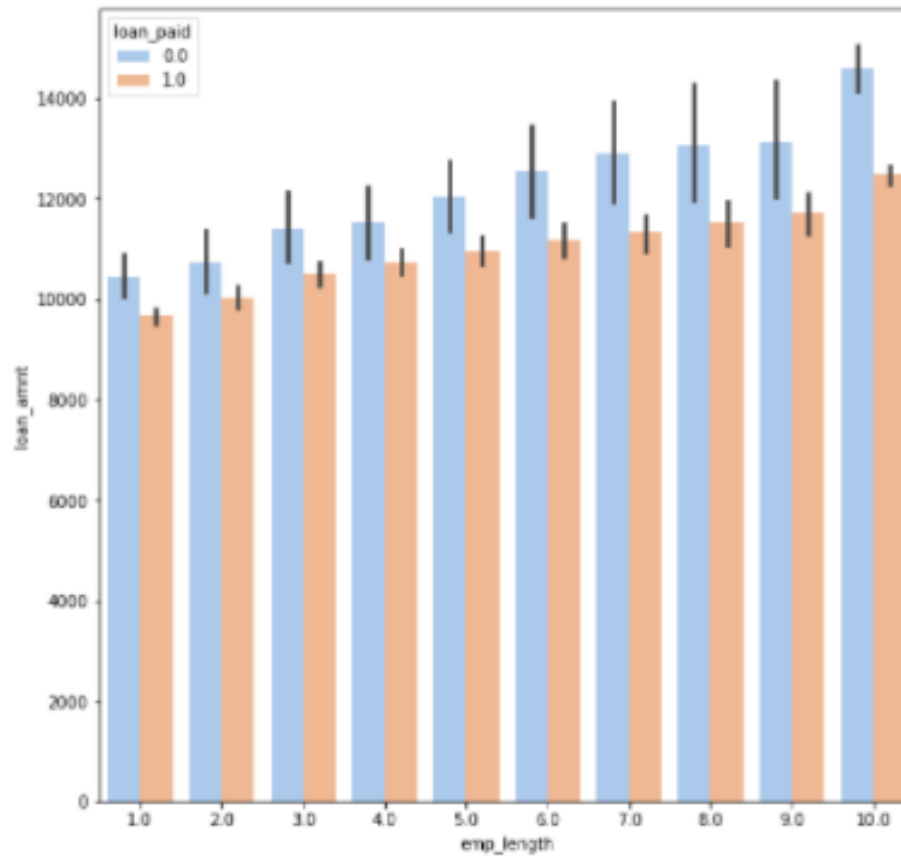
5.Loan amount vs Grade

```
In [83]: plt.figure(figsize=(10,10))
         sns.barplot(data =loan_data,x='loan_amnt', y='grade', hue ='loan_paid',palette="pastel", order=['A','B','C','D','E','F','G'])
         plt.show()
```

Employees with longer working history got the loan approved for a higher amount.

Looking at the verification status data, verified loan applications tend to have higher loan amount. Which might indicate that the firms are first verifying the loans with higher values.

Recommendations

- Applicants who applied and defaulted have no significant difference in loan amounts.

- Which means that applicants applying for long term has applied for more loan.

- Observations

- The above analysis with respect to the charged off loans. There is a more probability of defaulting when :

- 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 income of 70k-80k

- Applicants who have taken a loan in the range 30k - 35k and are charged interest rate of 15-17.5 %

- Applicants who have taken a loan for small business and the loan amount is greater than 14k

- Applicants whose home ownership is 'MORTGAGE and have loan of 14-16k

- When grade is F and loan amount is between 15k-20k

- When employment length is 10yrs and loan amount is 12k-14k

- When the loan is verified and loan amount is above 16k

- For grade G and interest rate above 20%