Lending Club Case Study

Background

- Lending Club is a confluence of investors looking for structured alternative investments and creditworthy borrowers looking to fulfill their monetary needs. There is a large part of the Indian population not covered by traditional credit-providing institutions. Simultaneously, there is an investor class looking to diversify their portfolio with alternative investment options that yield high returns.

- Lending Clubfill this gap and bring these demographics together, creating an ecosystem for people to meet their financial goals. We, as market leaders, strive to fulfill the demands that have been left unattended by others and are the fastest growing P2P lending platform in the country. We are led by passionate problem solvers and backed by investors around the world to realise this dream and become the most trustworthy platform for both our investors and borrowers.

Objective

In this case study, you will use EDA to understand how consumer attributes and loan attributes influence the tendency of default. When a person applies for a loan, there are two types of decisions that could be taken by the company:

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
- Loan rejected: The company had rejected the loan (because the candidate does not meet their requirements etc.). Since the loan was rejected, there is no transactional history of those applicants with the company and so this data is not available with the company (and thus in this dataset)

Objective

The analysis is divided into four main parts:

1. Data understanding
2. Data cleaning
3. Data Analysis
4. Conclusion

Data Understanding

Types of variables

- Customer (applicant) demographic
- Loan related information & characteristics
- Customer behavior (if the loan is granted)

E.g.

Customer demographics	Loan Information	Customer Behavior
Employment Length	Loan Amount	Delinquency Year
Employment Title	Funded Amount	Revolving Balance
Annual Income	Interest Rate	Recoveries
Description	Loan Status	Application Type

Data Cleaning

Please refer to python file for cleaning and hypothesis

Data Understanding

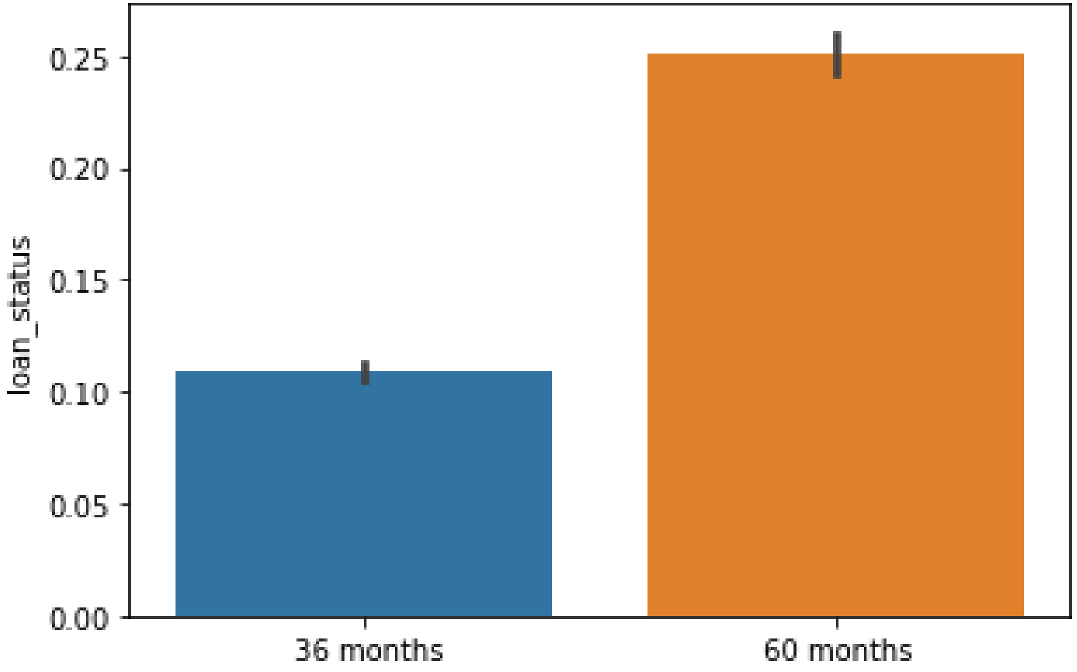
1. Overall Default Rate is 14%

First, let's look at the overall default rate.

```
# default rate
round(np.mean(df['loan_status']), 2)
```

4]

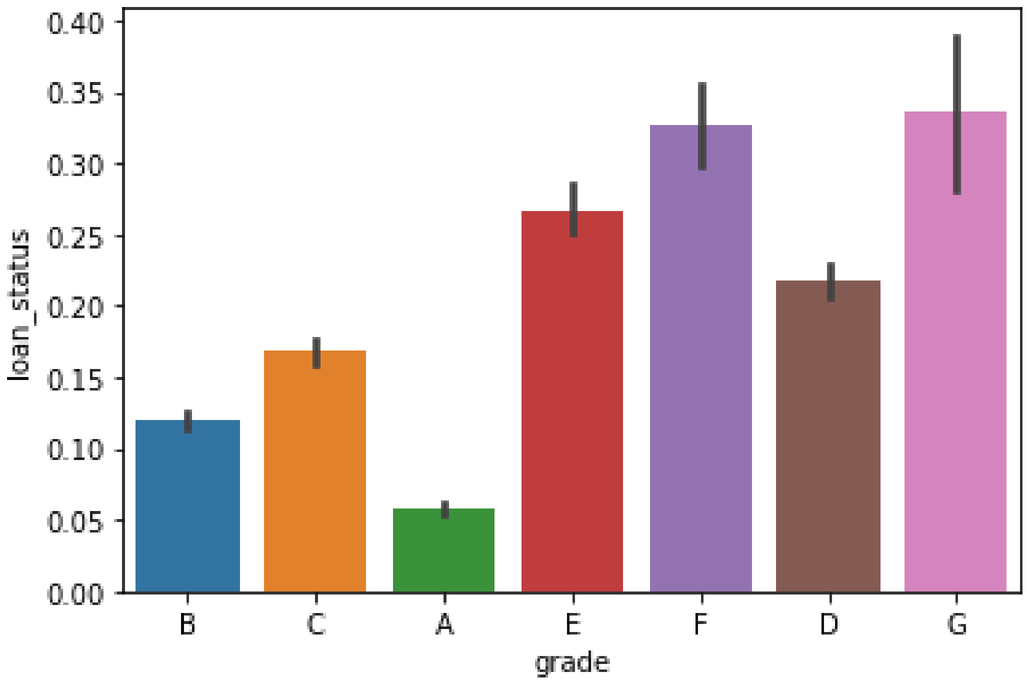
```
0.14
```



Data Analysis

1. Verification Status vs Loan Status

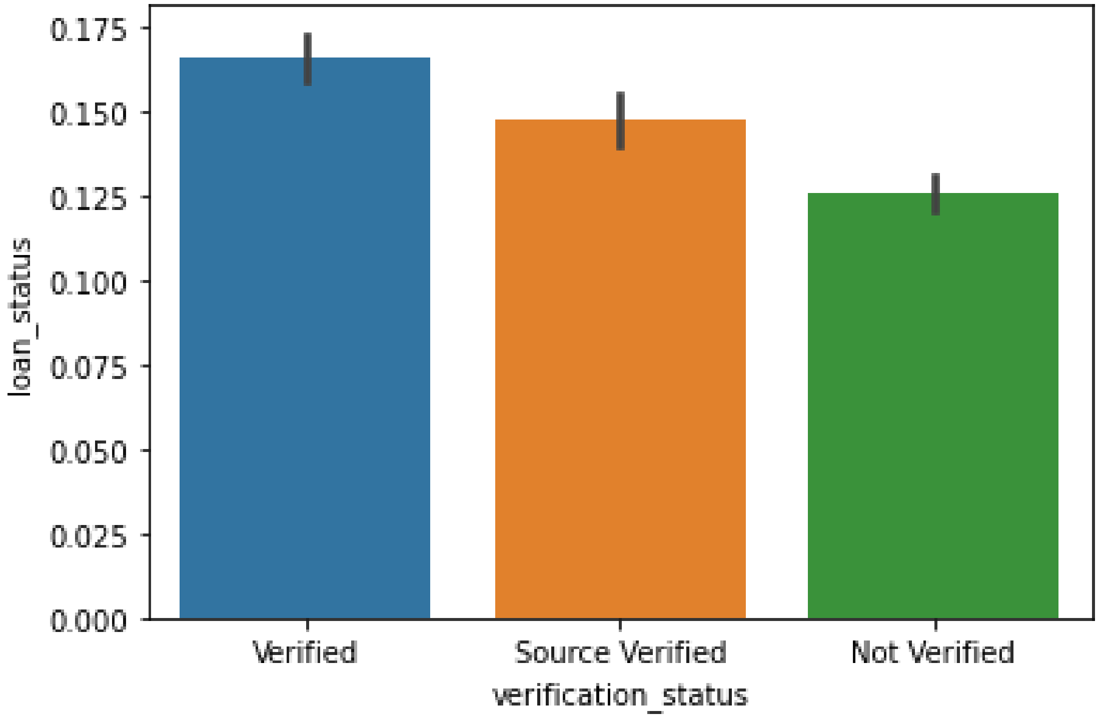
Clearly, as the grade of loan goes from A to G, the default rate increases. This is expected because the grade is decided by Lending Club based on the riskiness of the loan.



Data Analysis

2. Verification Status vs Loan Status

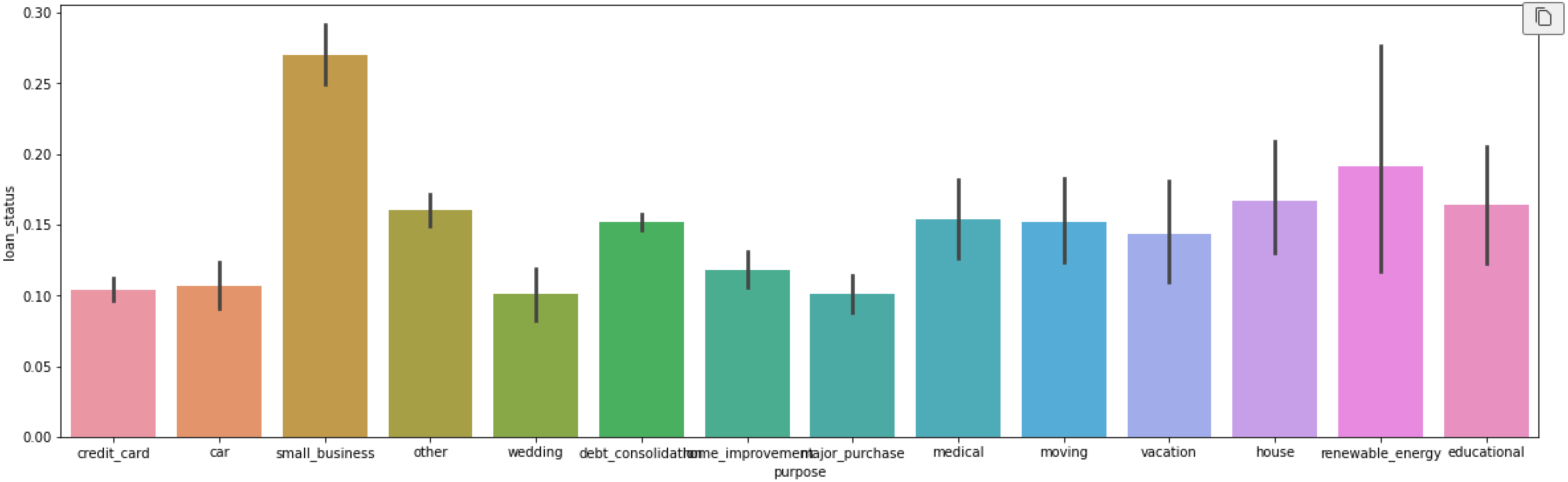
Contradiction to the expected values the verified loans default more than not verified



Data Analysis

3. Purpose vs Loan Status

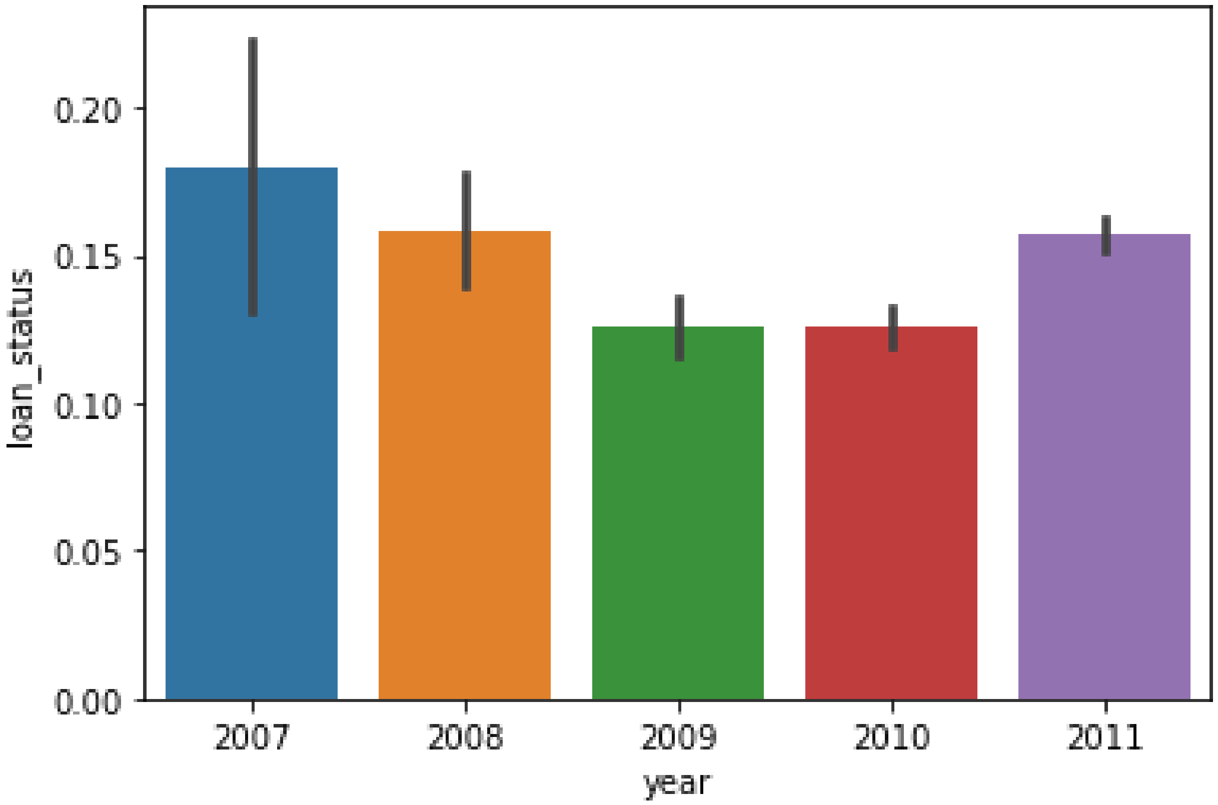
Small business loans default the most, then renewable energy and education



Data Analysis

4. Year vs Loan Status

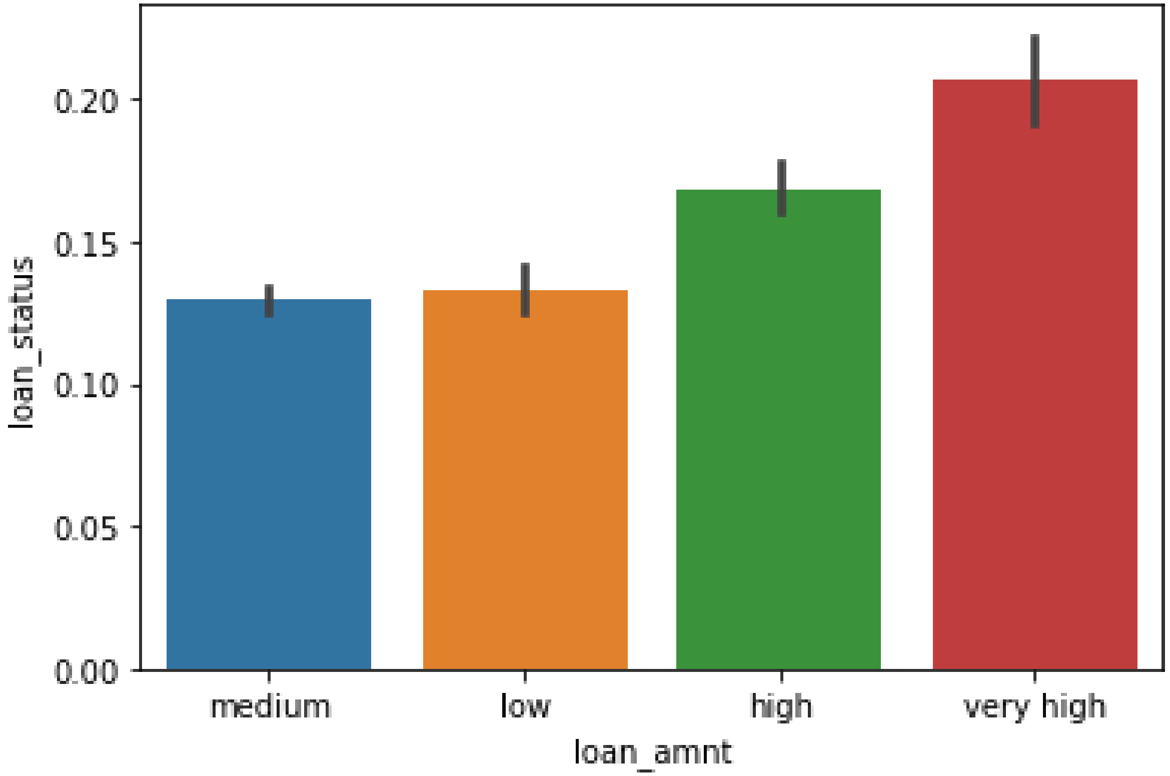
The rate had suddenly increased in 2011



Data Analysis

5. rates across loan amount type

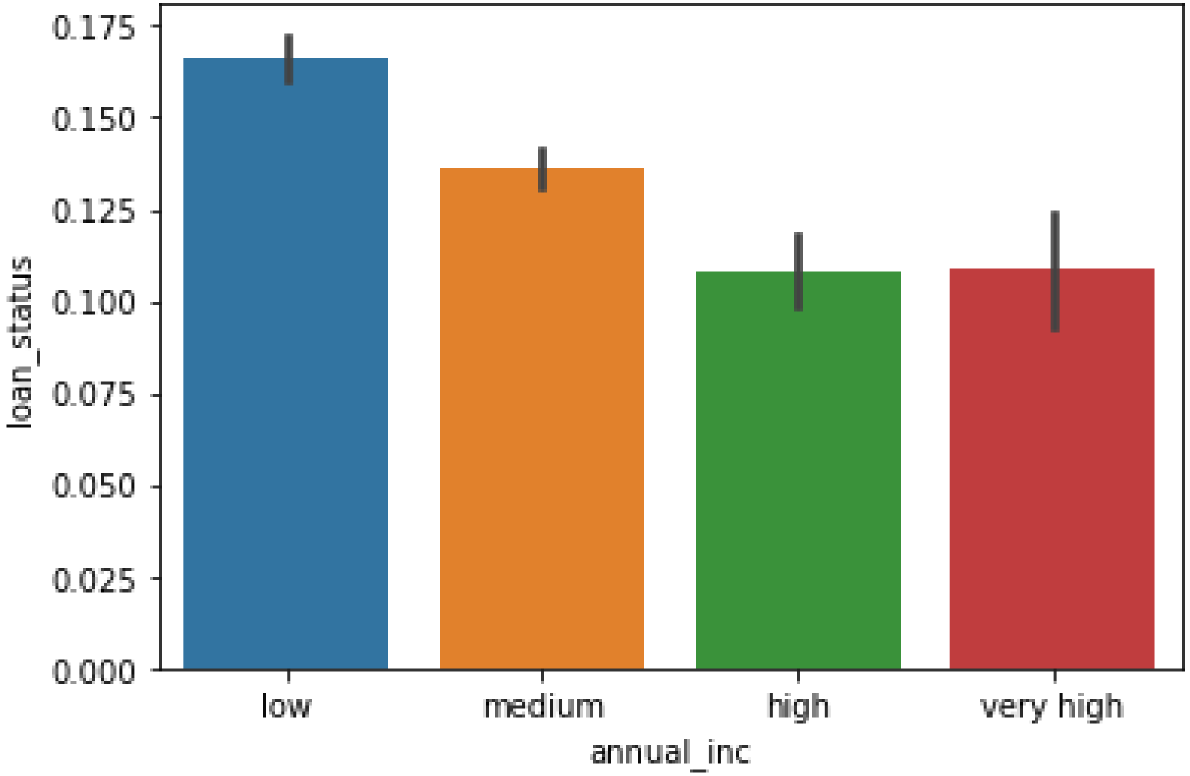
Binning the loan amount to ease analysis



Data Analysis

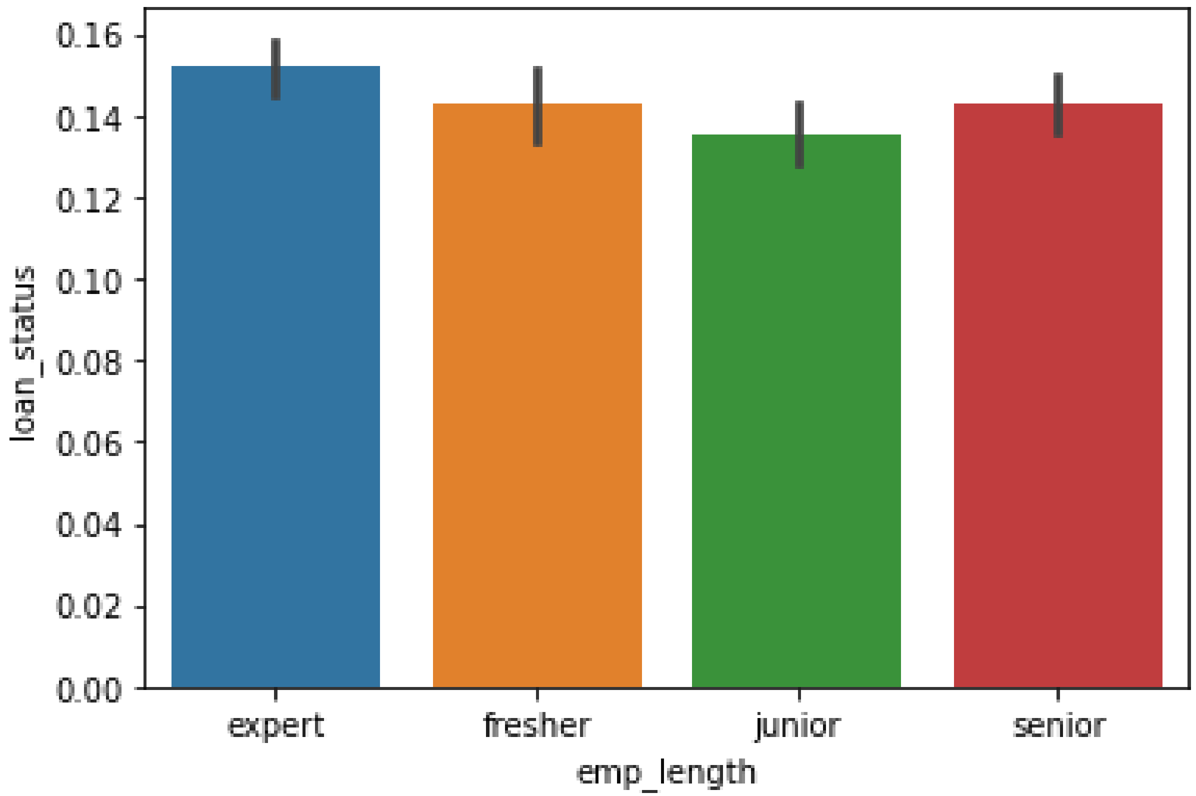
6. Annual income and default rate

Lower the annual income, higher the default rate



Data Analysis

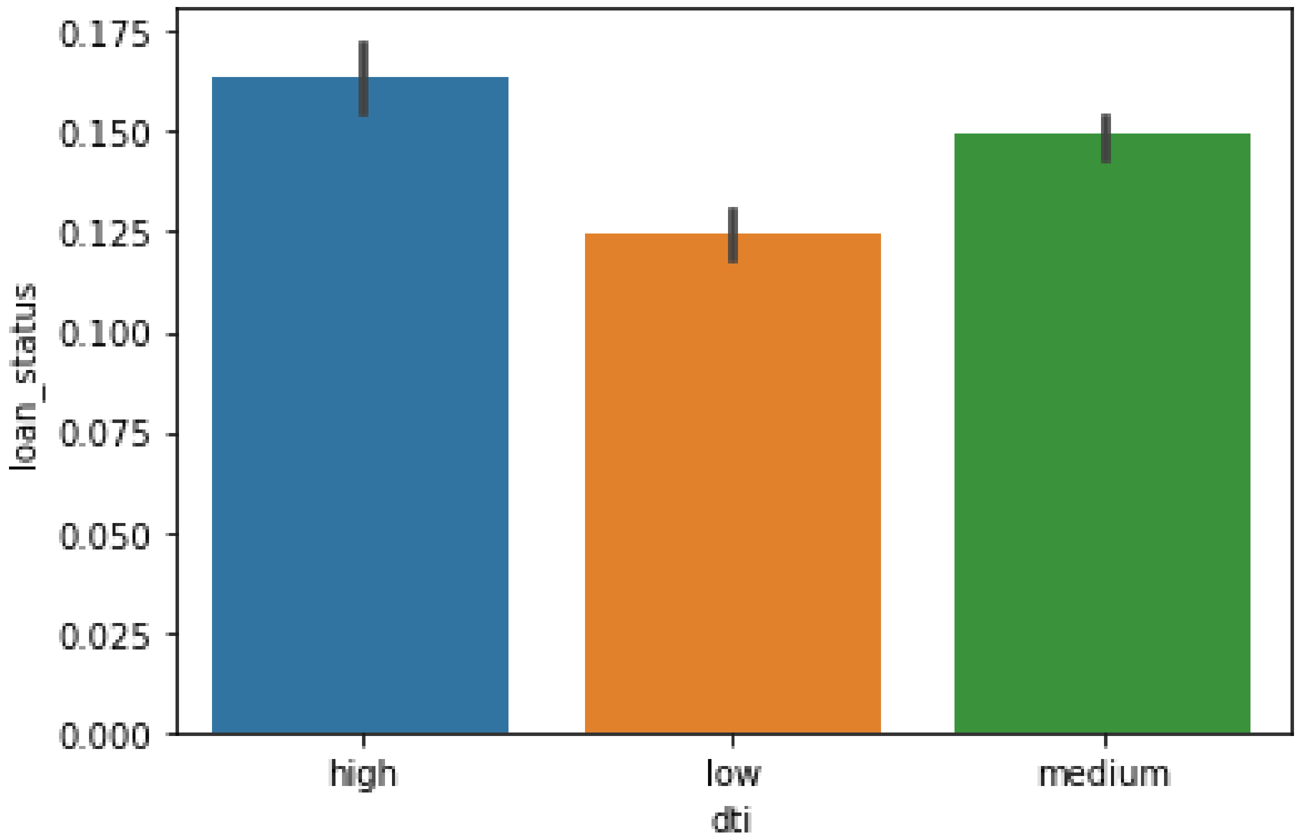
7. Employment Length and default rate



Data Analysis

8. default rates across debt to income ratio

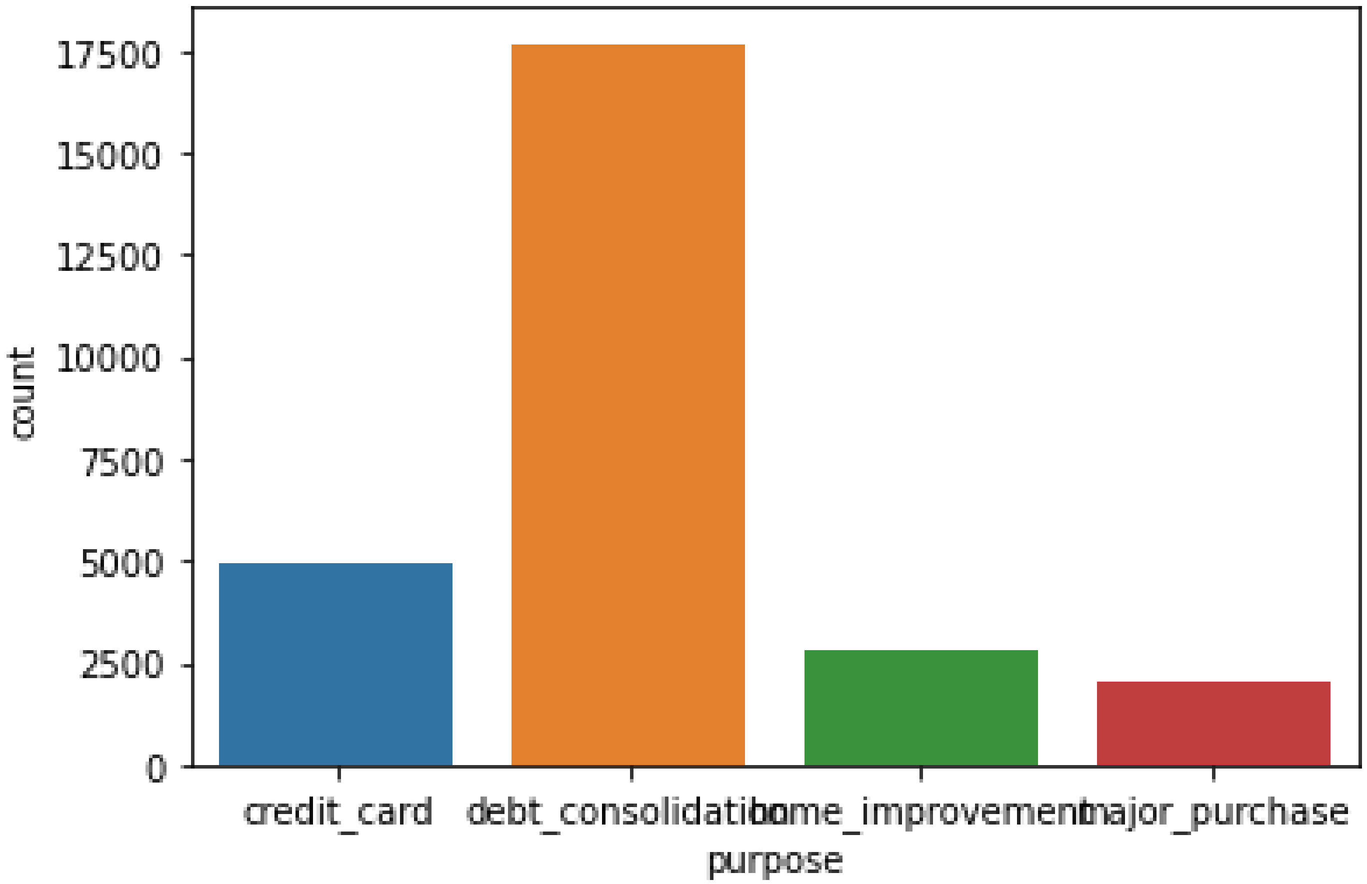
High dti translates into higher default rates, as expected



Data Analysis

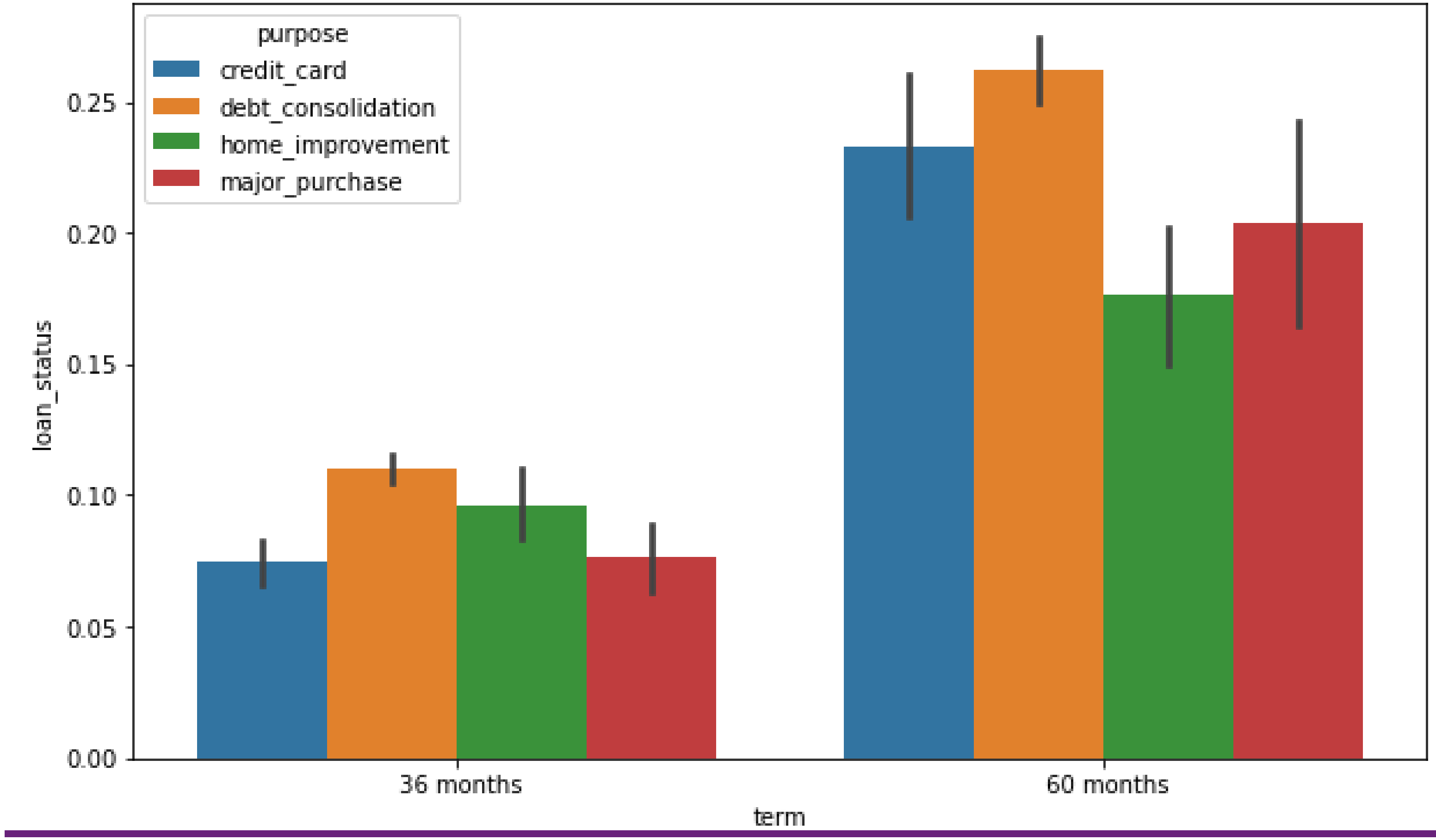
Analyze the top 4 types of loans based on purpose:

- consolidation
- credit card
- home improvement
- major purchase.



Data Analysis

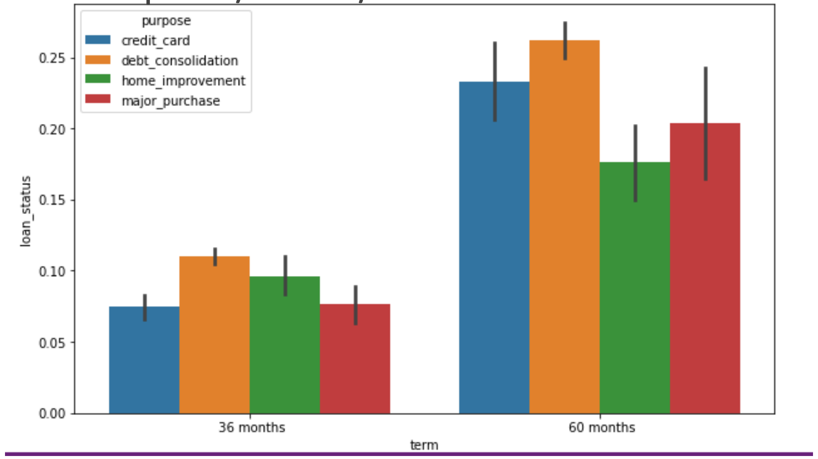
default rates across two types of categorical variables purpose of loan (constant) and another categorical variable (which changes)



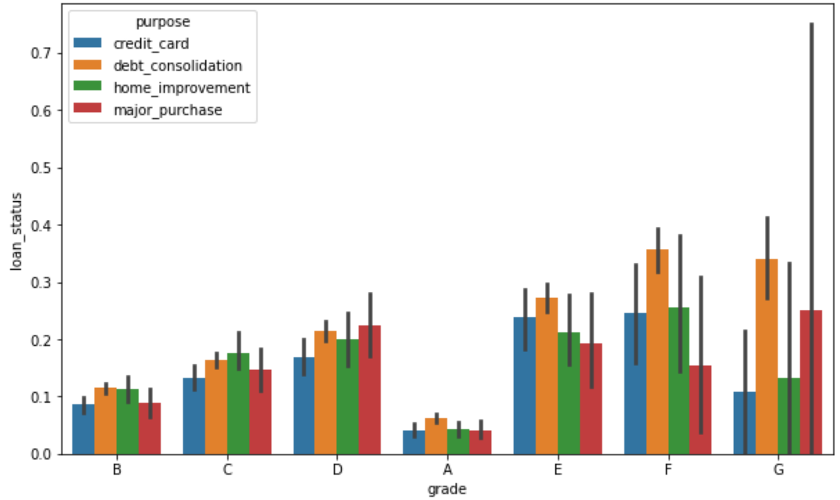
Data Analysis

Default rates across two types of categorical variables

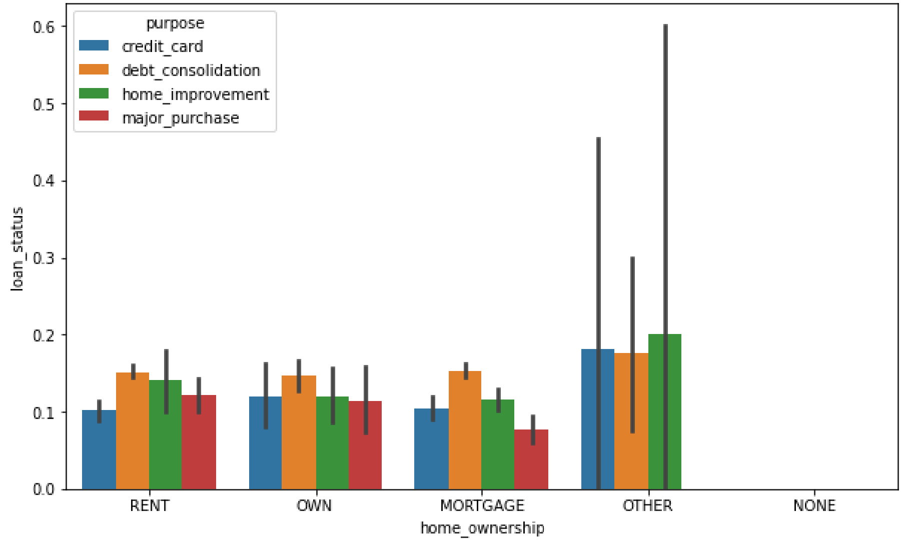
Purpose / Term / status



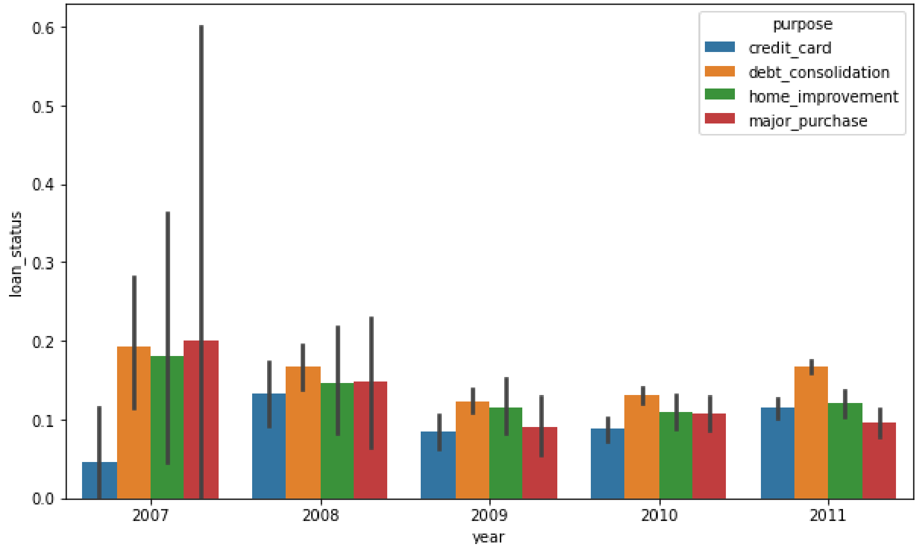
Grade / purpose / status



Home ownership / Status / purpose



Year / purpose / status



Conclusion

- The grade of loan goes from A to G, the default rate increases
- Contradiction to the expected values the verified loans default more than not verified
- Small business loans default the most, then renewable energy and education
- Lower the annual income, higher the default rate
- High dti translates into higher default rates, as expected