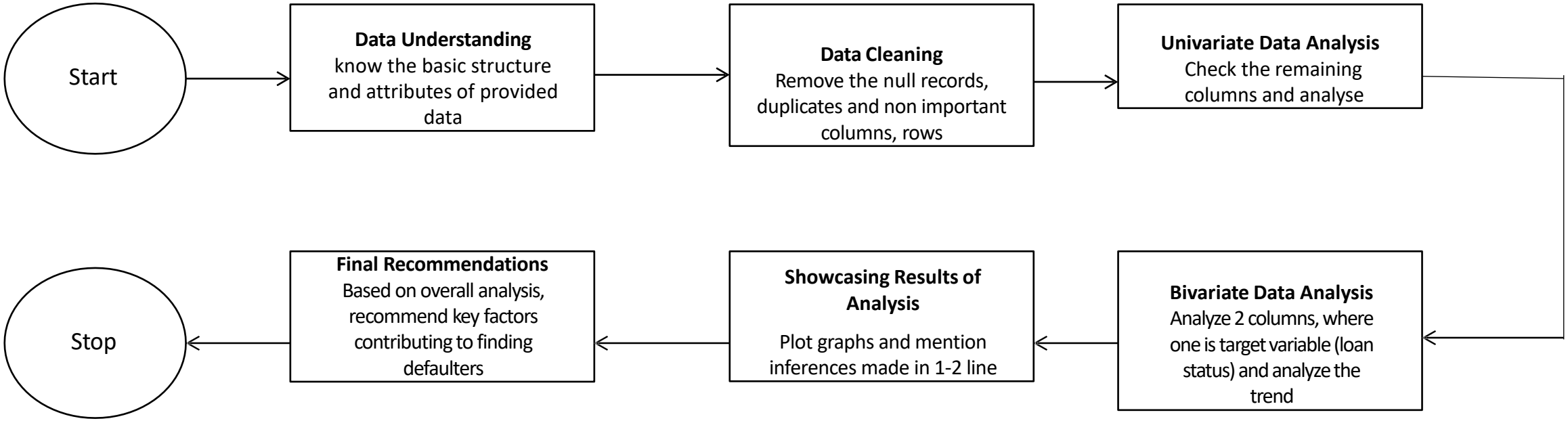
Lending Club Case study

Rajeev Ranjan & Rajesh Sapkota

Objectives: Lending Club Case study

- To understand the **driving factors (or driver variables)** behind loan default, i.e. the variables which are strong indicators of default. The company can utilize this knowledge for its portfolio and risk assessment.

Problem solving methodology



Data Understanding

- After loading, Checked head & tail of data to understand the type of data
- Check overall shape & Size of data – 39717 rows, 111 columns
- Used describe function to find mean median mode, max, etc.
- Checked for null value columns. There are many. Will remove them later.
- Check data type. dtypes: float64(74), int64(13), object(24). Need to correct some data type (terms, int rate etc.)
- Analyzed the target column, i.e. 'loan_status'
  - Fully paid
  - Current
  - Charged-off
- "Current" status is for the customers still paying and hence not confirmed as fully paid or defaulters. Will need to remove them from the analysis later.
- Studied Data dictionary file to understand some of the column headings

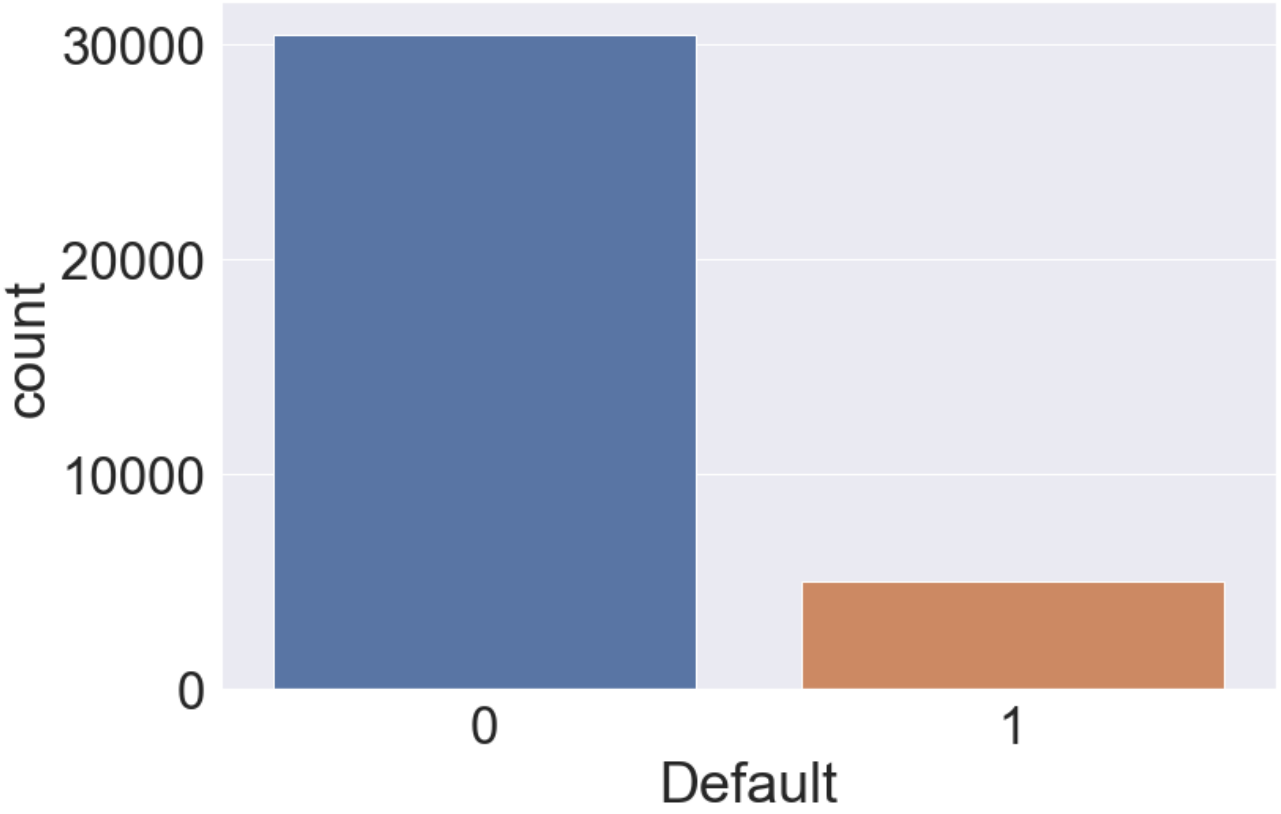
Data Cleaning

- At the end there were many columns with all entries as NaN. All such columns were removed using dropna function. Columns reduced from 111 to 57.
- Checked again columns with null values and removed some of the columns with high number of null values and not useful for our target variable analysis (e.g. desc….as we have purpose column for same thing). Now remaining 53 columns.
- Removing columns with customer behaviour attributes, which were not available before approval of loan and were populated after approval. These columns will not help in finding driving factors for identifying defaulters.
- Removed columns having same value in all entries. They will not help in analysis.
- Removed Rows containing “Current” in loan_status column (1140 entries – 2.87% of total entries 39717), as they are not confirmed as defaulters or non defaulters. Not needed for analysis.
- Removed some more unnecessary columns (id, member id, url, funded amt, funded amt inv) as they are not required for analysis. Only Loan amt kept for analysis, as the same was requested by borrower and mostly it is same or very close to the funded value

Data Cleaning

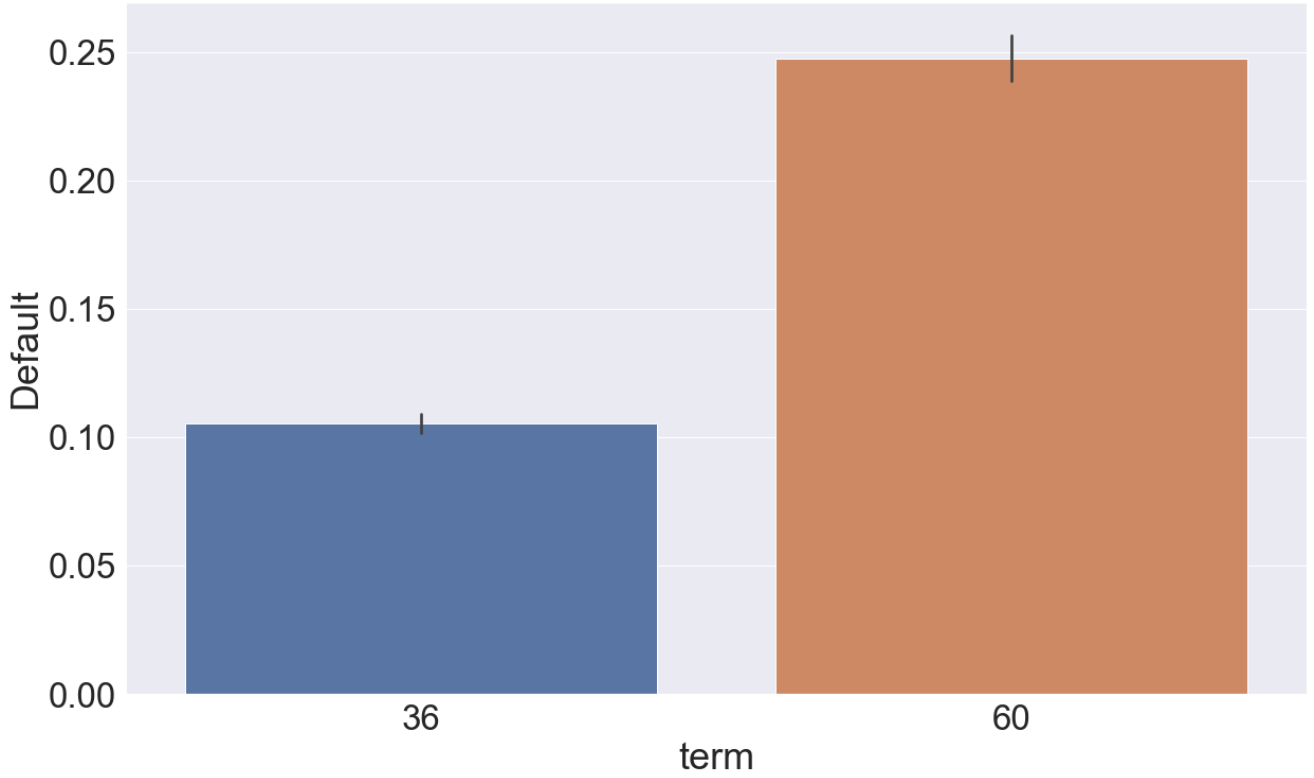
- Removed unnecessary characters of string from int rate and term columns and changed their data type for making easy analysis.
- Removed rows with null values in some columns (emp title, emp length etc.), as they are not many and will not hamper our analysis.
- Did not use imputation in this dataset, due to above reason.
- Categorical variable may be imputed with mode
- Numeric columns may be imputed with
  - ✓median (when outliers are there) or
  - ✓mean (when no outliers)
- Removed unnecessary characters from emp length and issue d
- Created a column for month from issue date column for analysis
- Replaced Fully Paid to 0 and Charged Off as 1 for the loan status column
- Renamed loan_status column to ‘Default’ for easy understanding in plots

Data Analysis



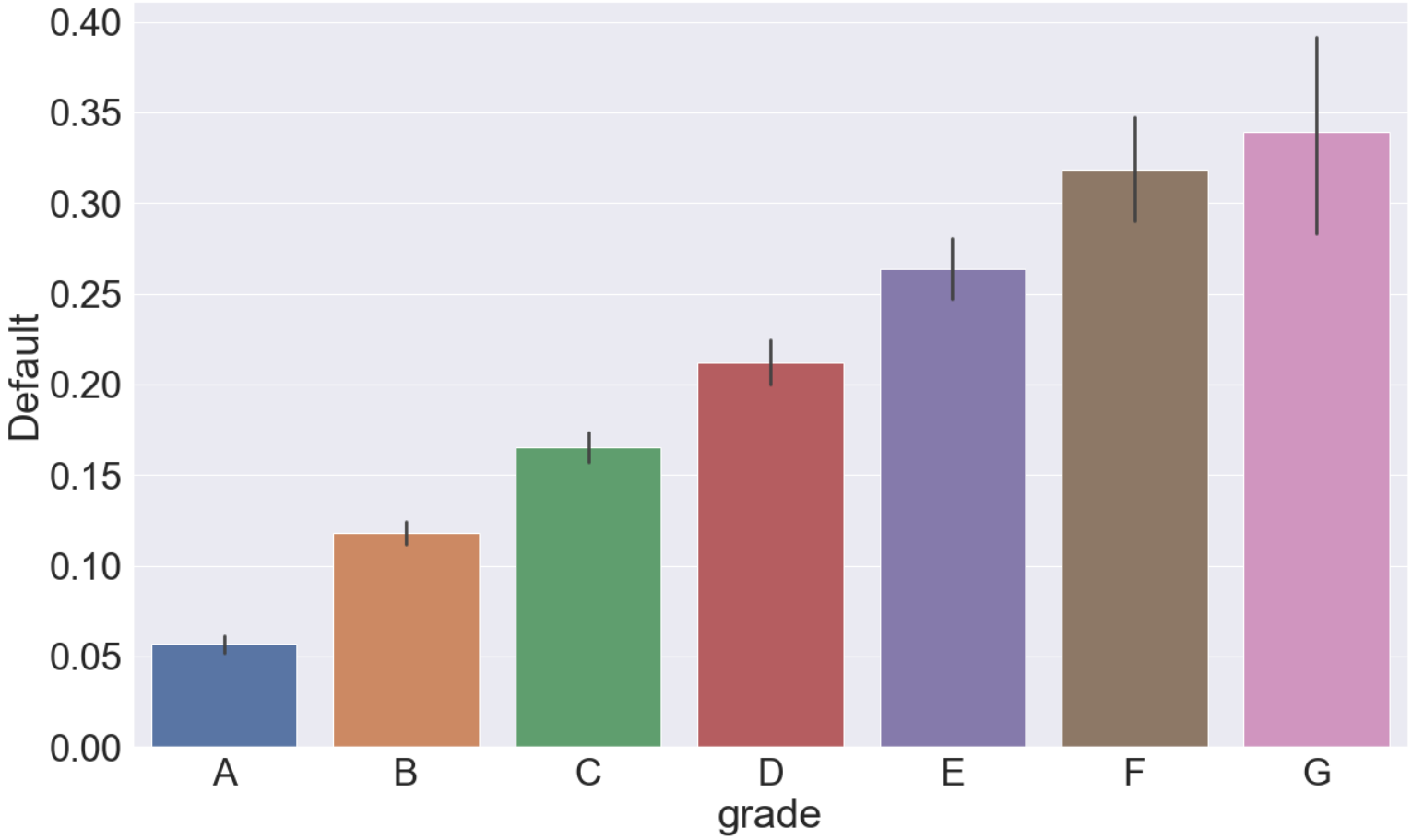
- The total number of Defaulters (1) are much lower (~5000) compared to total number of Non-Defaulters (or Fully Paid – ‘0’)

Univariate Analysis – Loan Term



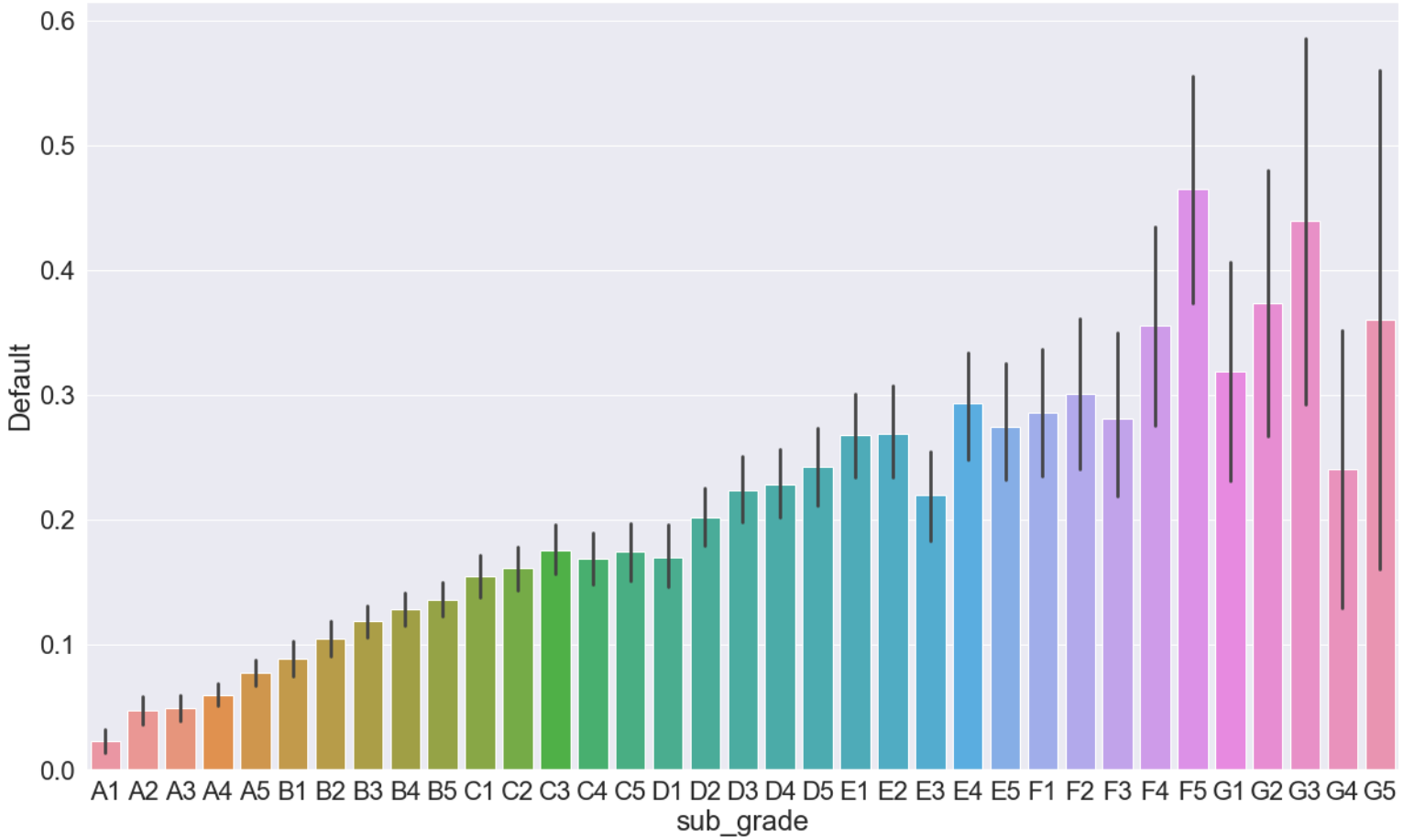
- There is an increase in defaulters with increase in Loan term duration (36 to 60 months)

Univariate Analysis – Grades



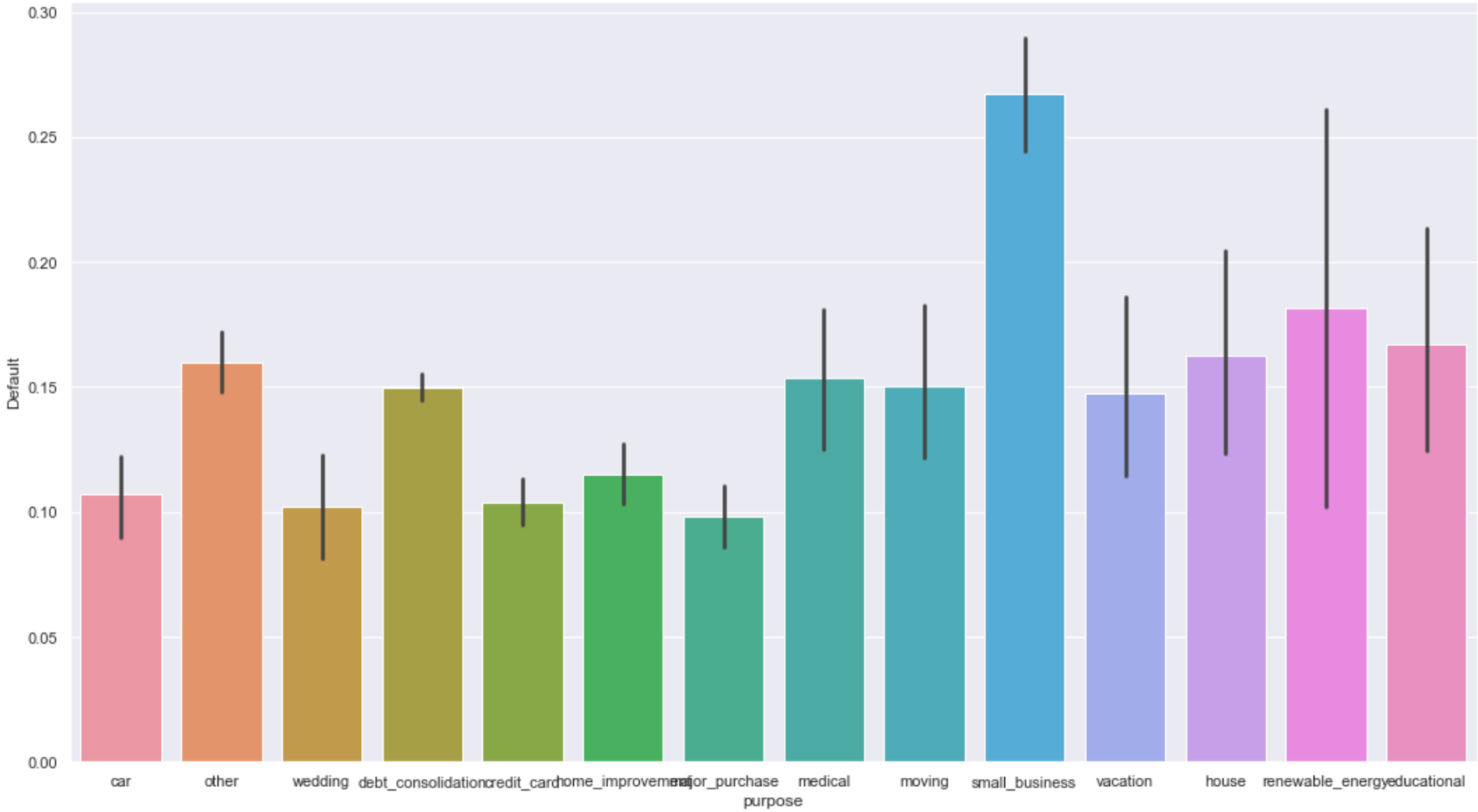
- There is an increase in defaulters as we move Grade A to Grade G (from good to bad grades)

Univariate Analysis – Sub Grades



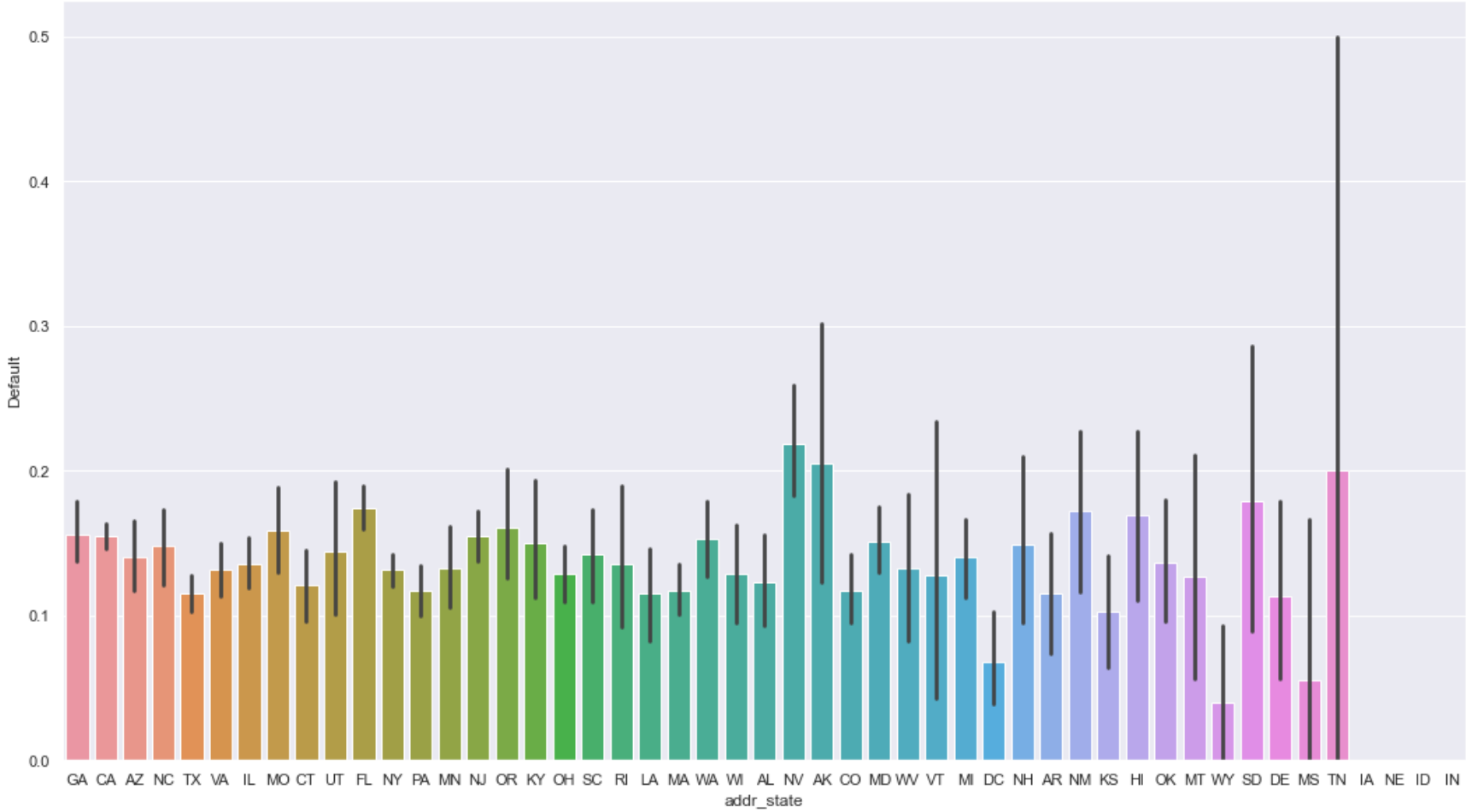
- In General, there is an increase in defaulters while moving from good to bad subgrades, within a Grade.
- However F5 has more default that any of the G grade subgrades

Univariate Analysis – Purpose of loan



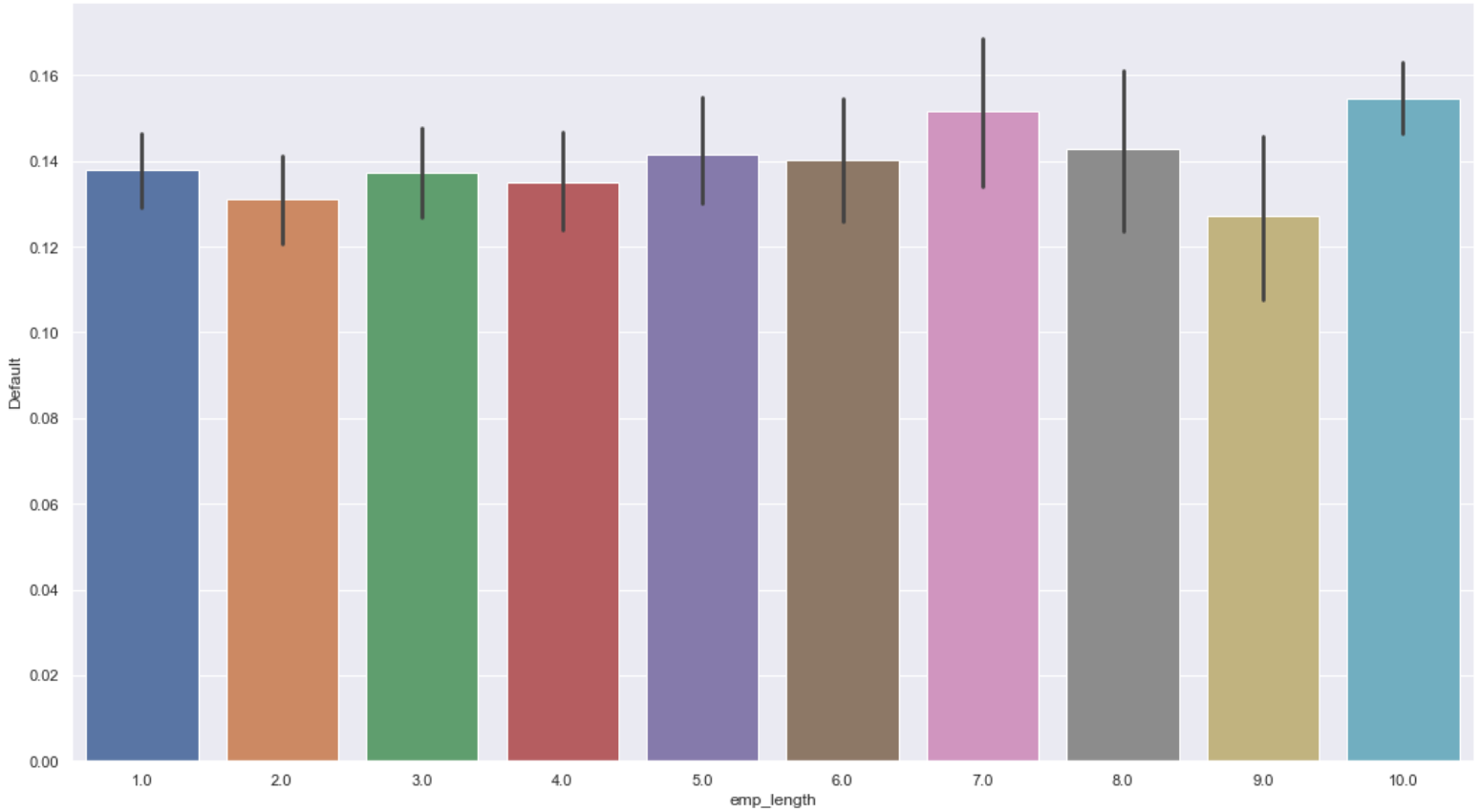
- Small Business as the purpose of loan has the highest proportion of defaulters,

Univariate Analysis – State address



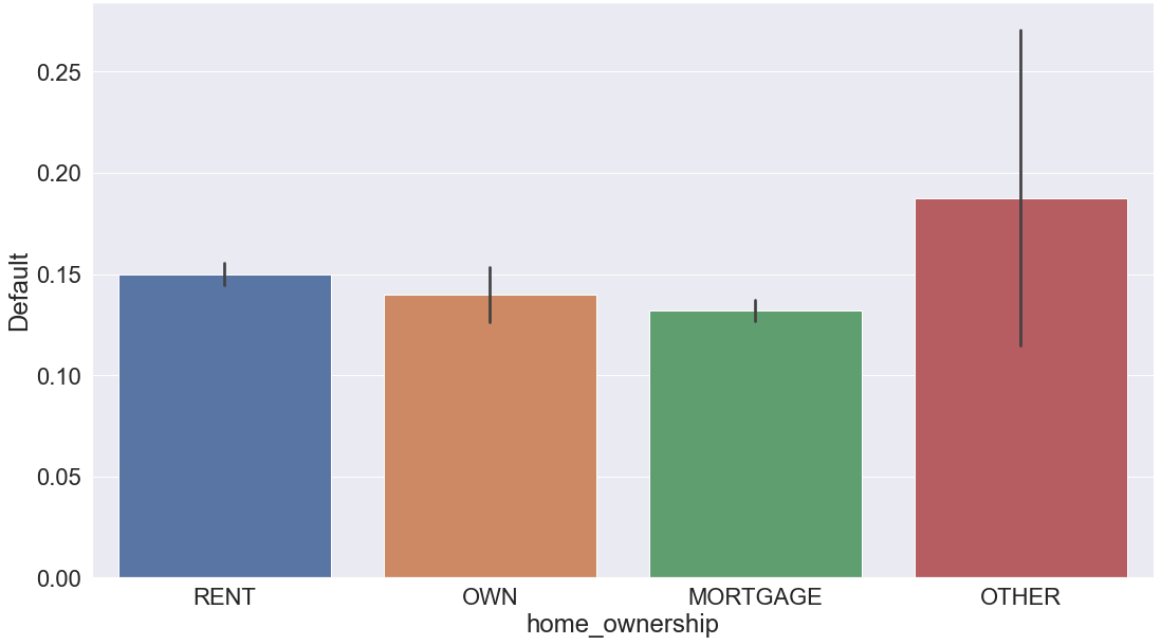
- NV,AK & TN states have highest defaults

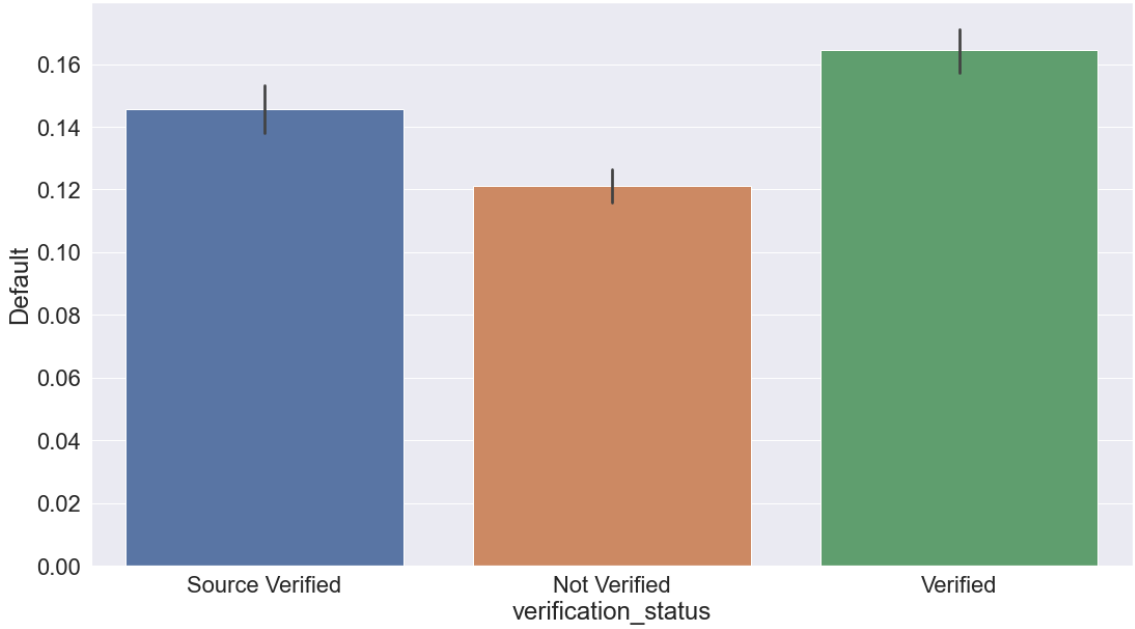
Univariate Analysis – Employment length (years)



- No concrete trend with increasing employment experience of borrower

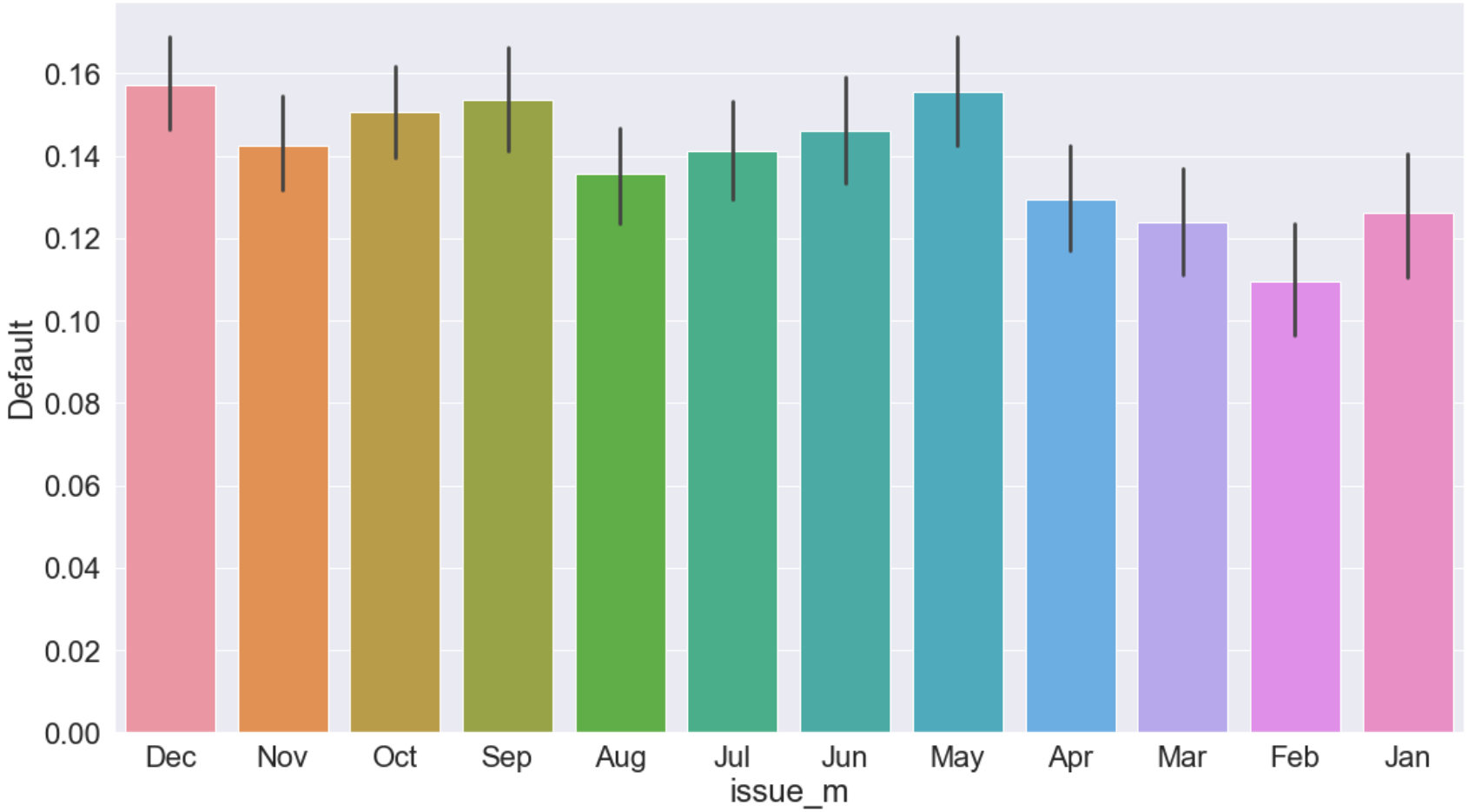
Univariate Analysis – Home ownership and verification status





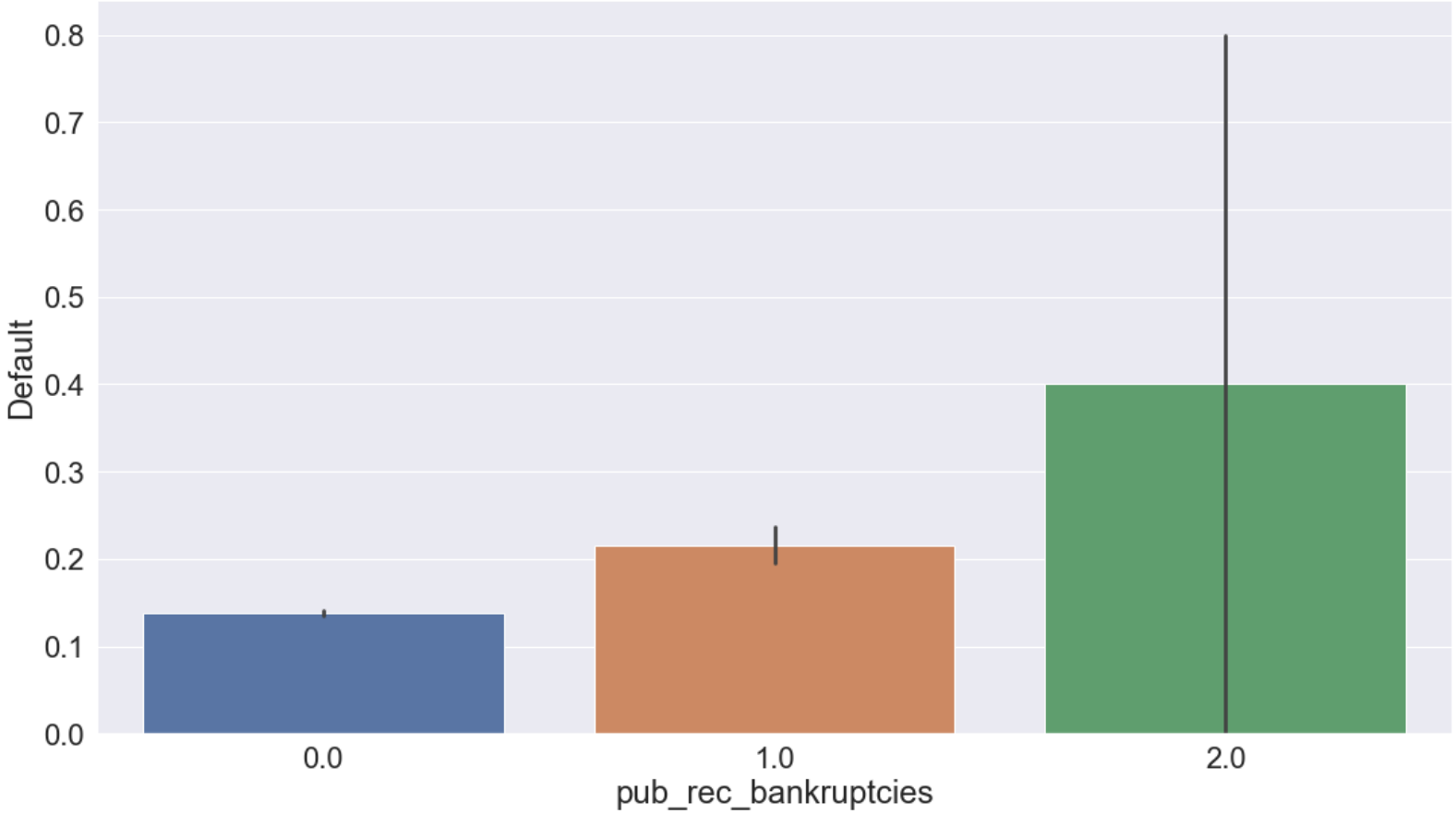
- No firm trend in home ownership. Others have highest default
- No logical trend in Verification status. Not verified borrowers has less default.

Univariate Analysis – Loan issue month



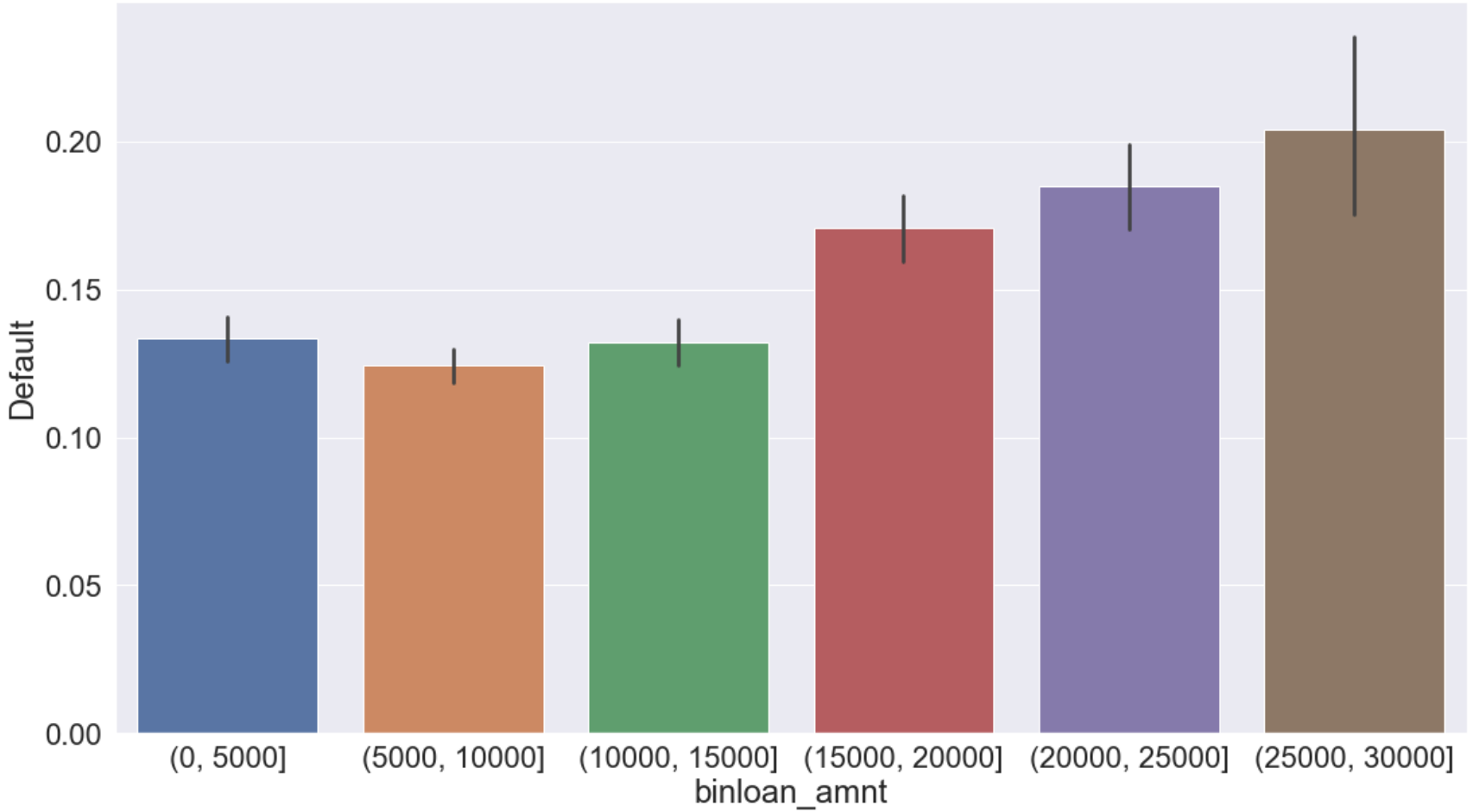
- No revealing trend as such.
- However, December month has highest loan issued defaulters (Festival or new year spendings may be the reason) .

Univariate Analysis – Public Record bankrupcies



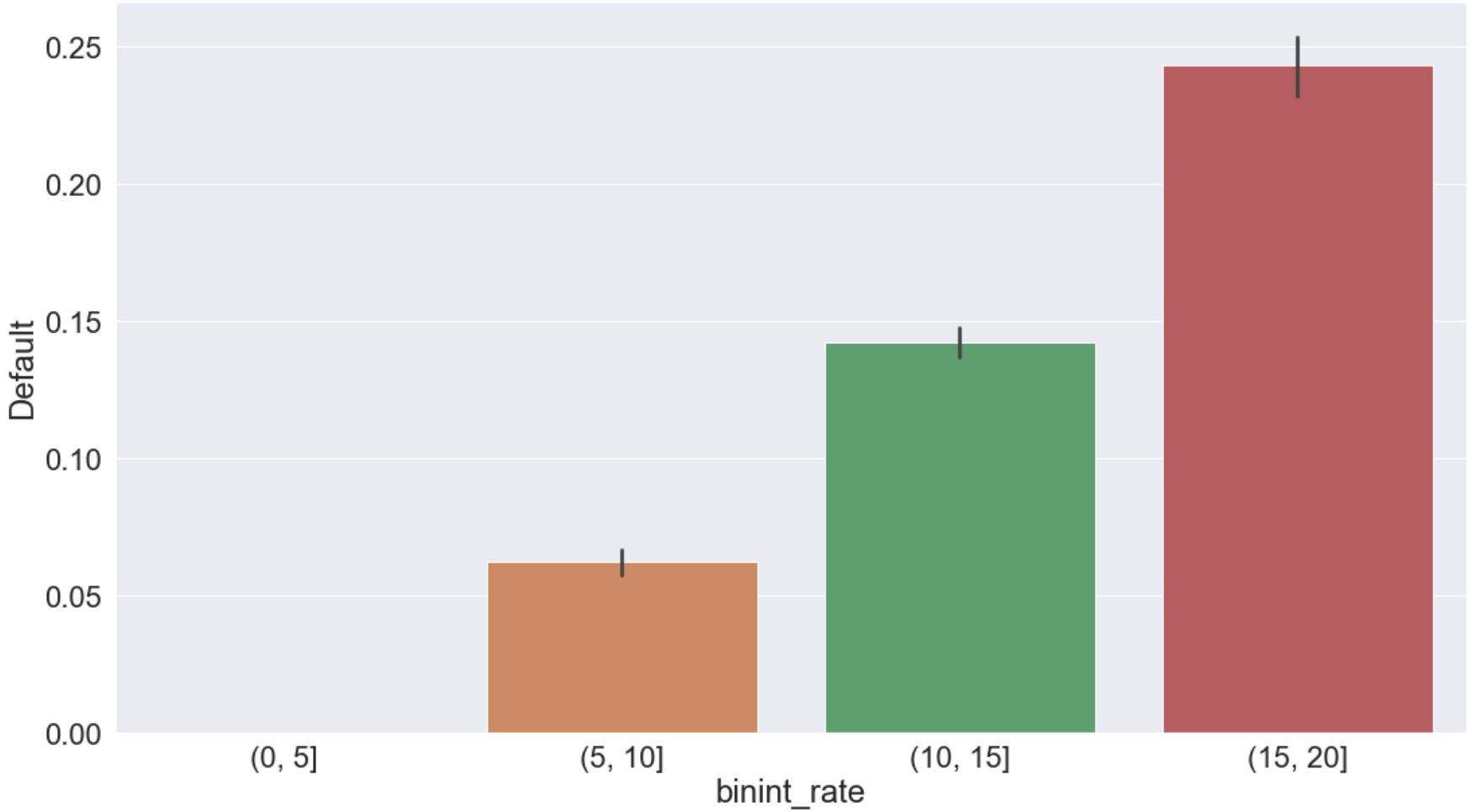
- Clear trend that borrowers having more number of public bankruptcies records are more likely to be defaulters

Univariate Analysis – Loan amount



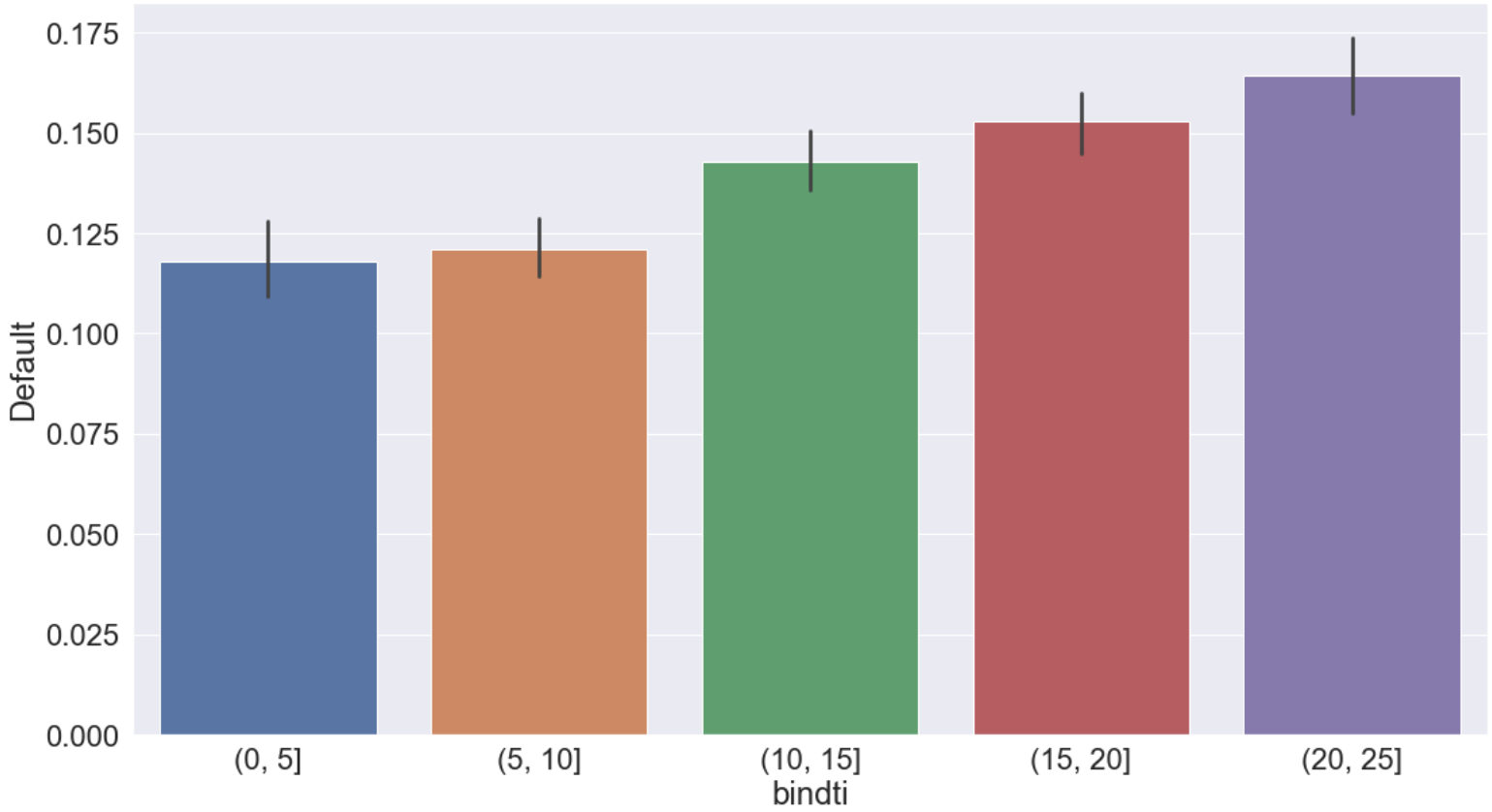
- gradual trend after 10000, that higher loan amount has greater risk of defaulting

Univariate Analysis – Interest rate



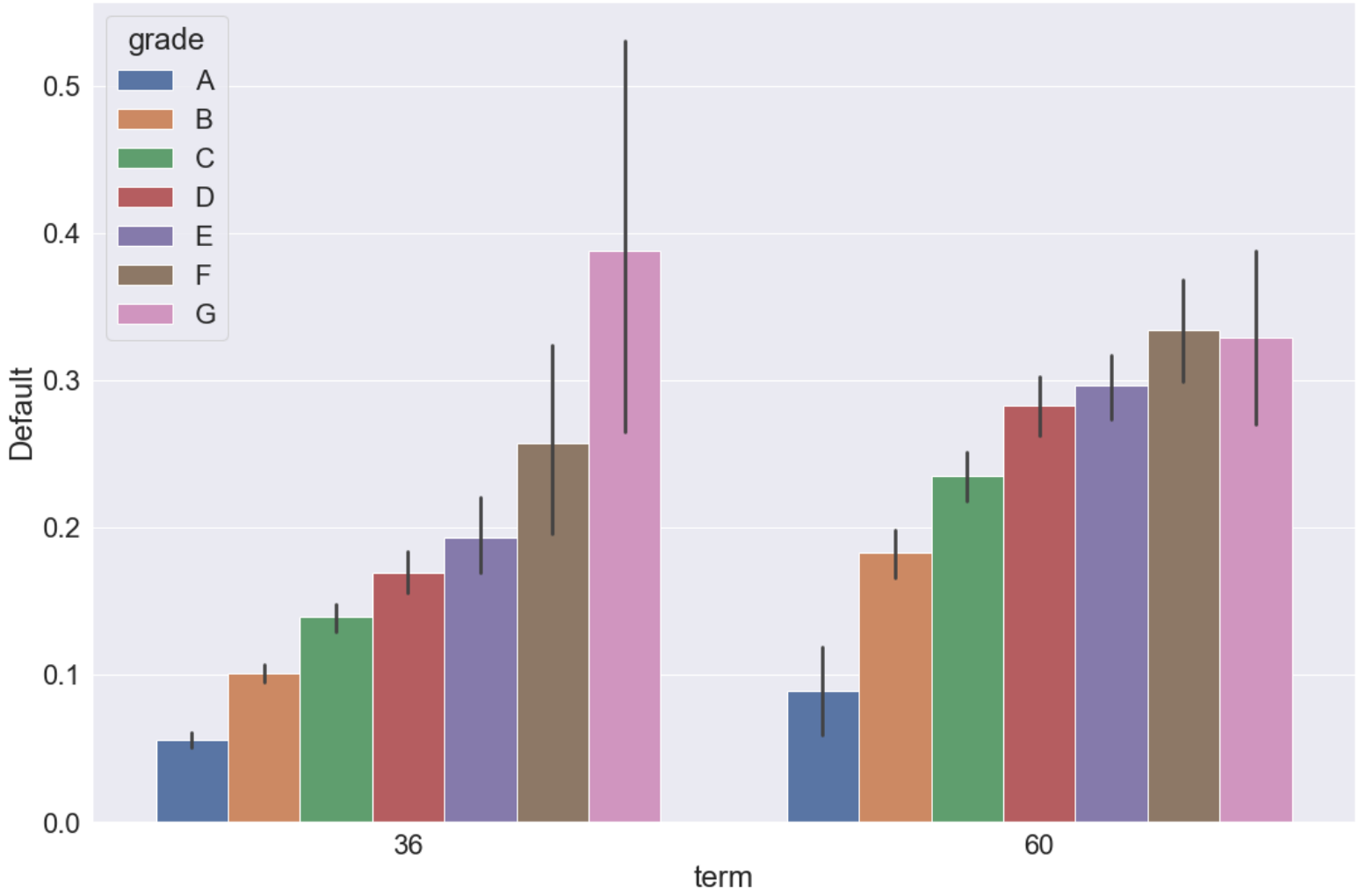
- Strong trend seen that higher interest rates are more prone to defaulting from the borrower

Univariate Analysis – Debt to Income ratio



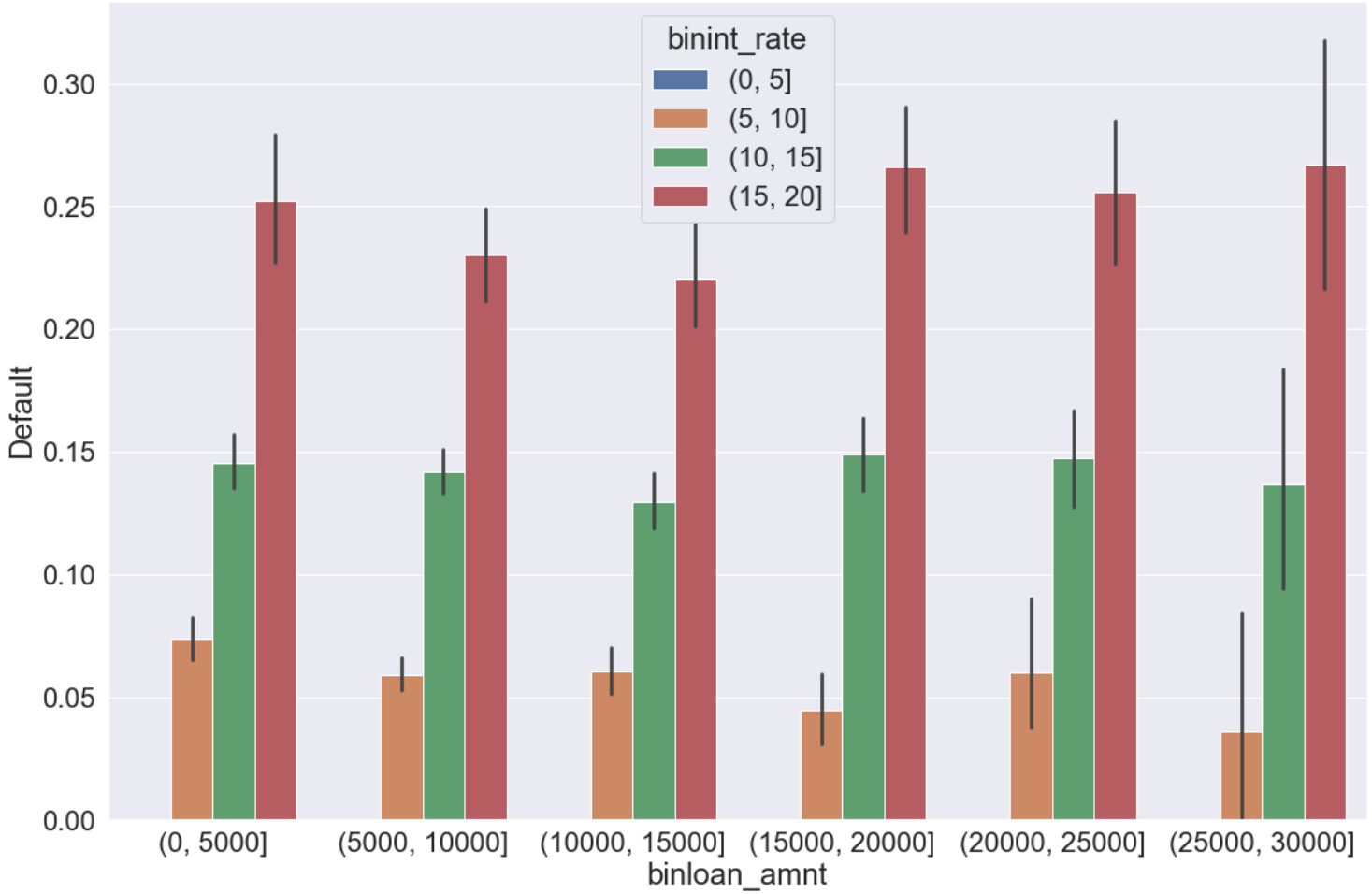
- Clear trend that chances of defaulting increases with increase in debt to income ratio

Bivariate Analysis – term and grade



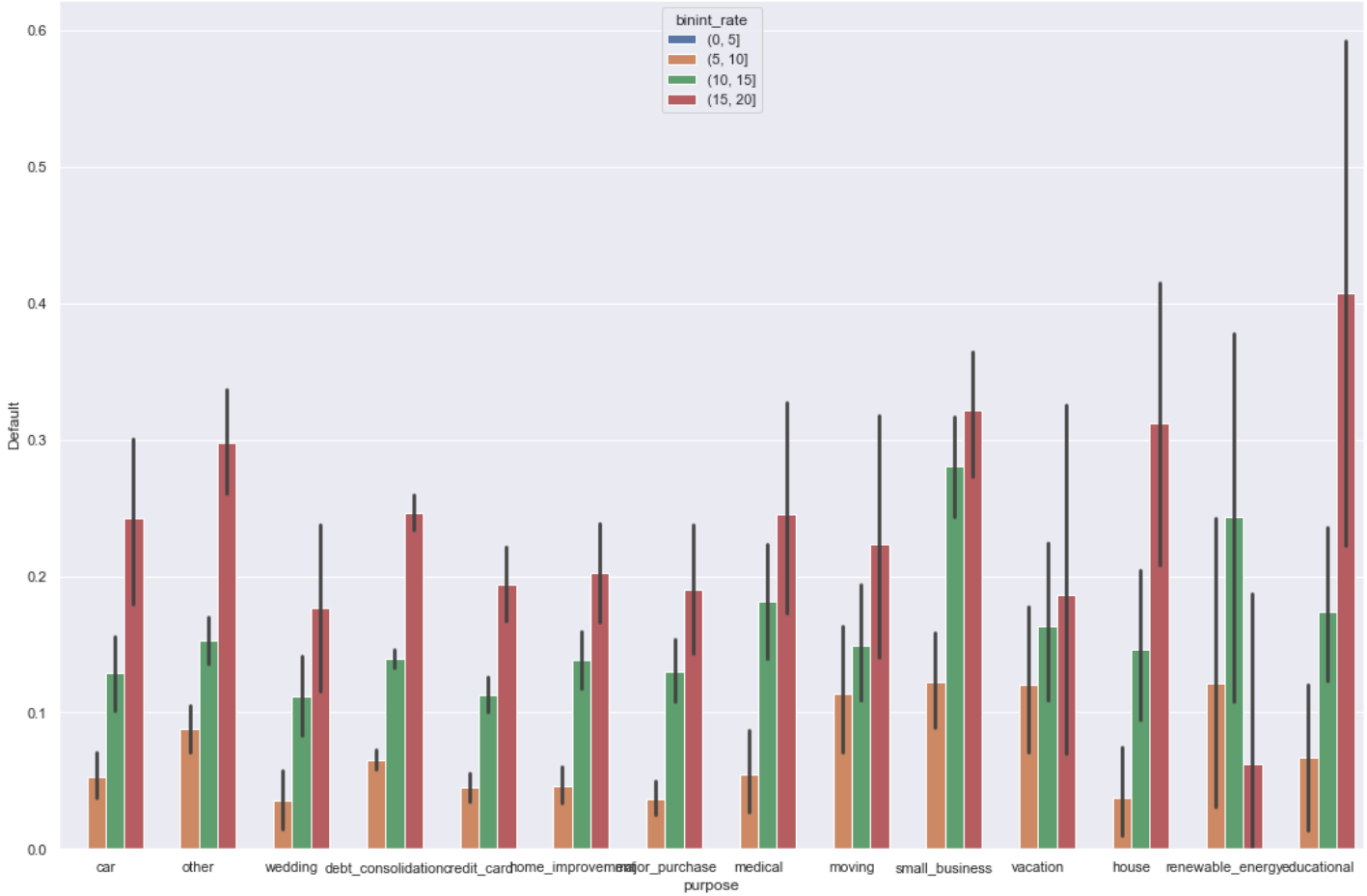
- It is a strong trend that for both loan terms, default rates increase when we move from good grades to bad grades. However, for higher term (60 months), the default rates of bad grades are consolidating at the end.

Bivariate Analysis – loan amnt and int. rate



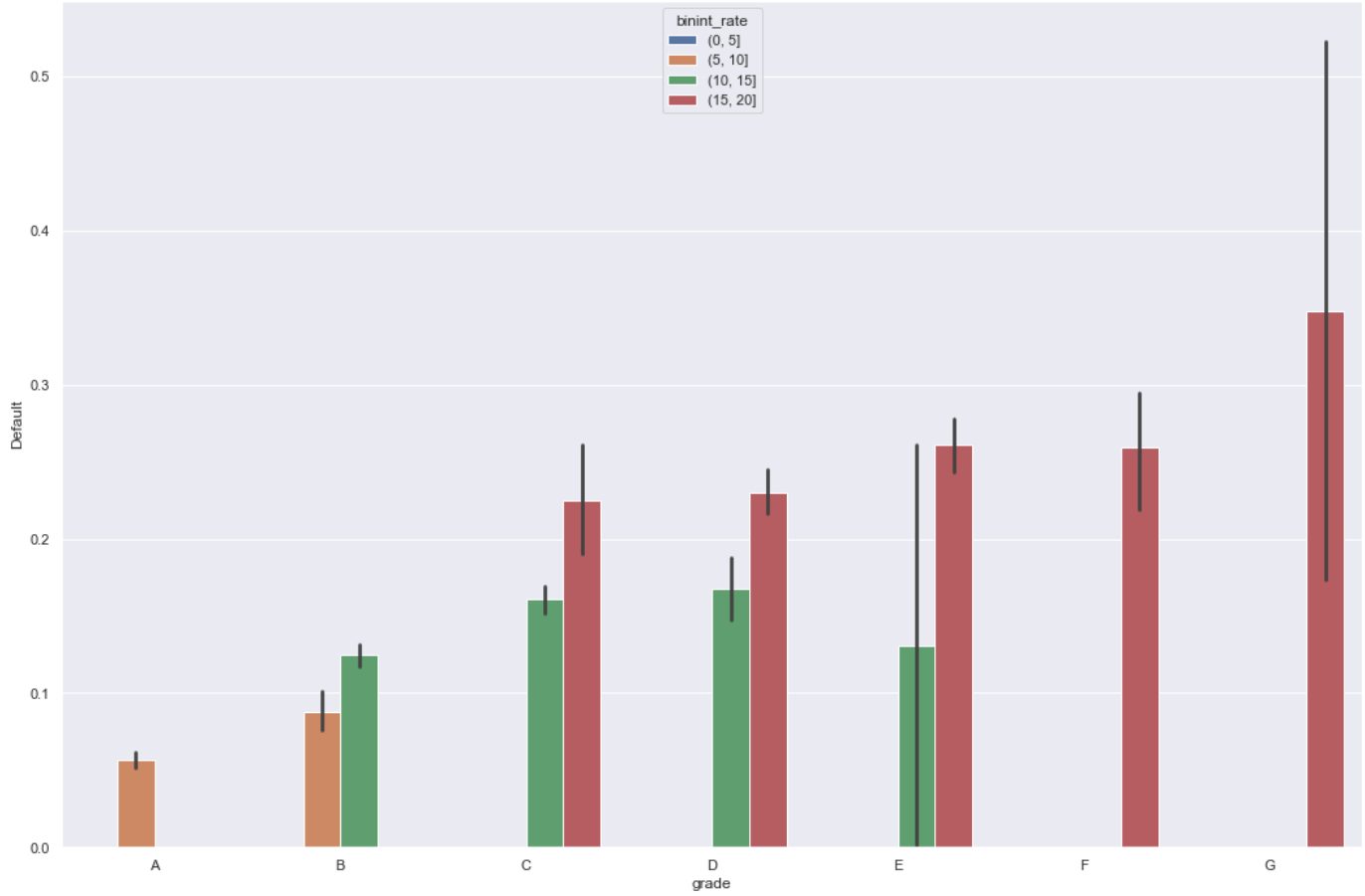
- Default rate increases with higher interest rates for all the loan amount buckets
- However, with increasing loan amount, the rate of increase in default rate is not significant in a fixed interest bucket.
- It means that high interest rate is a more stronger indicator for default, compared to high loan amount

Bivariate Analysis – Purpose & Interest rate



- Default rate increase with higher interest rate for all purpose of loans.
- Small business have higher default rates at even lower interest rates
- Education purpose has a sudden jump in default rate with increasing interest rate to >15%

Bivariate Analysis – Grades and Interest rate



- Default rate increases with higher interest rates from A to G grade
- However F & G grades are given loans on higher interest rates only

Final Recommendations

Key Drivers to identify defaulters

- Interest rates :- Avoid loans at high interest rates to low grade borrowers
- Purpose of loan :- Avoid high loan amount to borrowers with small business as purpose and avoid high interest (>15%) educational loans
- Public Records :- Avoid giving loans or high amount loans to people with more than 1 public records of bankruptcies
- Address :- Restrict loan disbursement to NV,AK & TN states
- Loan issue month :- More scrutiny on loans disbursed during December /New year is required