



EDA

Problem Statement



- Identify the risky loan applicants
- Mitigate the risk
- Identify Driving factors (or driver variables) behind loan default

Approach to the find solution



- Perform EDA is the first action to perform any analysis
- In this analysis we collect the data which is public or private based on the requirement .
- Check for the quality of the data if quality is okay then no need to perform any cleansing activities.
- If quality is not upto the mark then perform data cleansing activities involves
 - ❖ 1. Fix rows and columns
 - ❖ 2. Fix missing values
 - ❖ 3. Standardise values
 - ❖ 4. Fix invalid values
 - ❖ 5. Filter data

Approach to the find solution

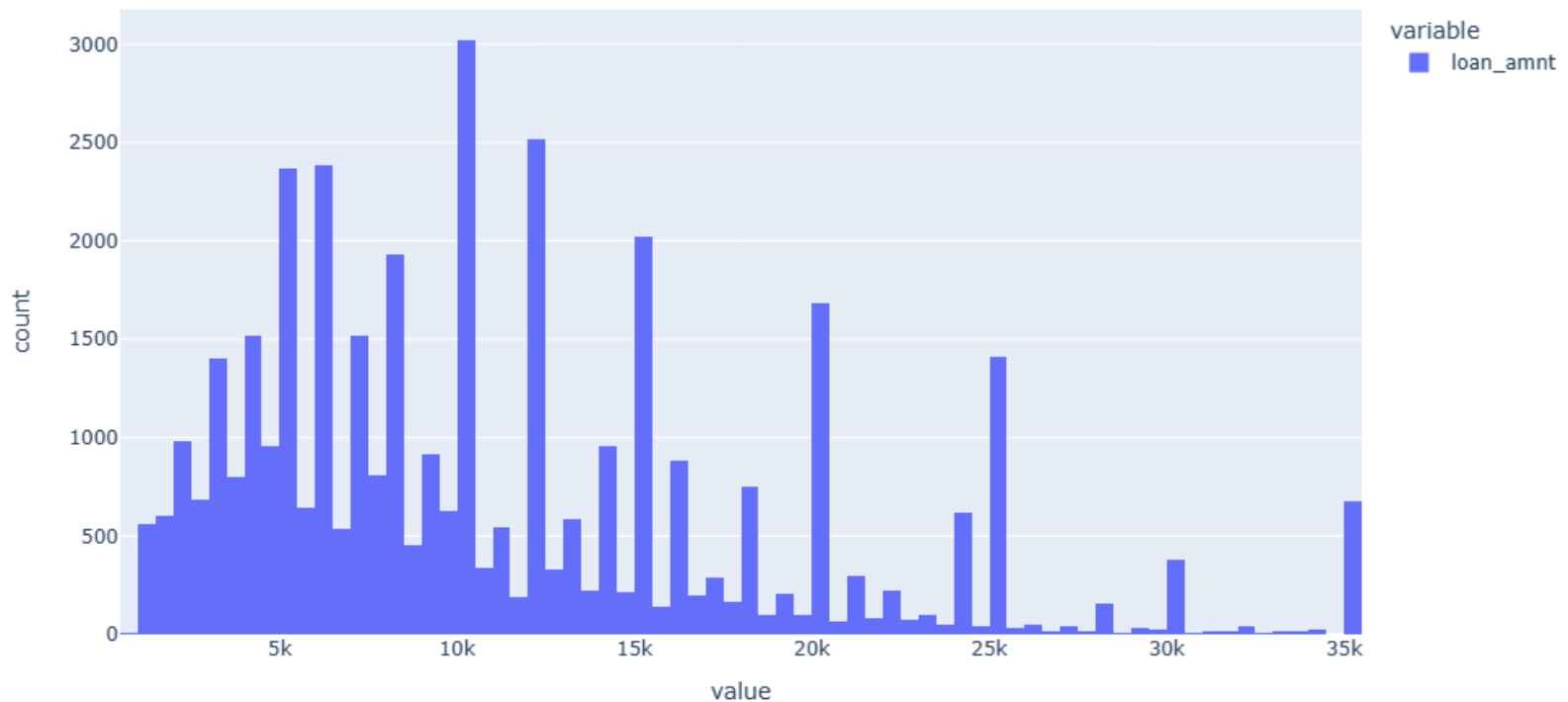


- Perform Analysis on basis of variable classified as Univariate Analysis ,Bivariate analysis
- Types of Variables in univariant Analysis
- Ordered and unordered categorical variables

Check we can Derived Matrix from existence variables:

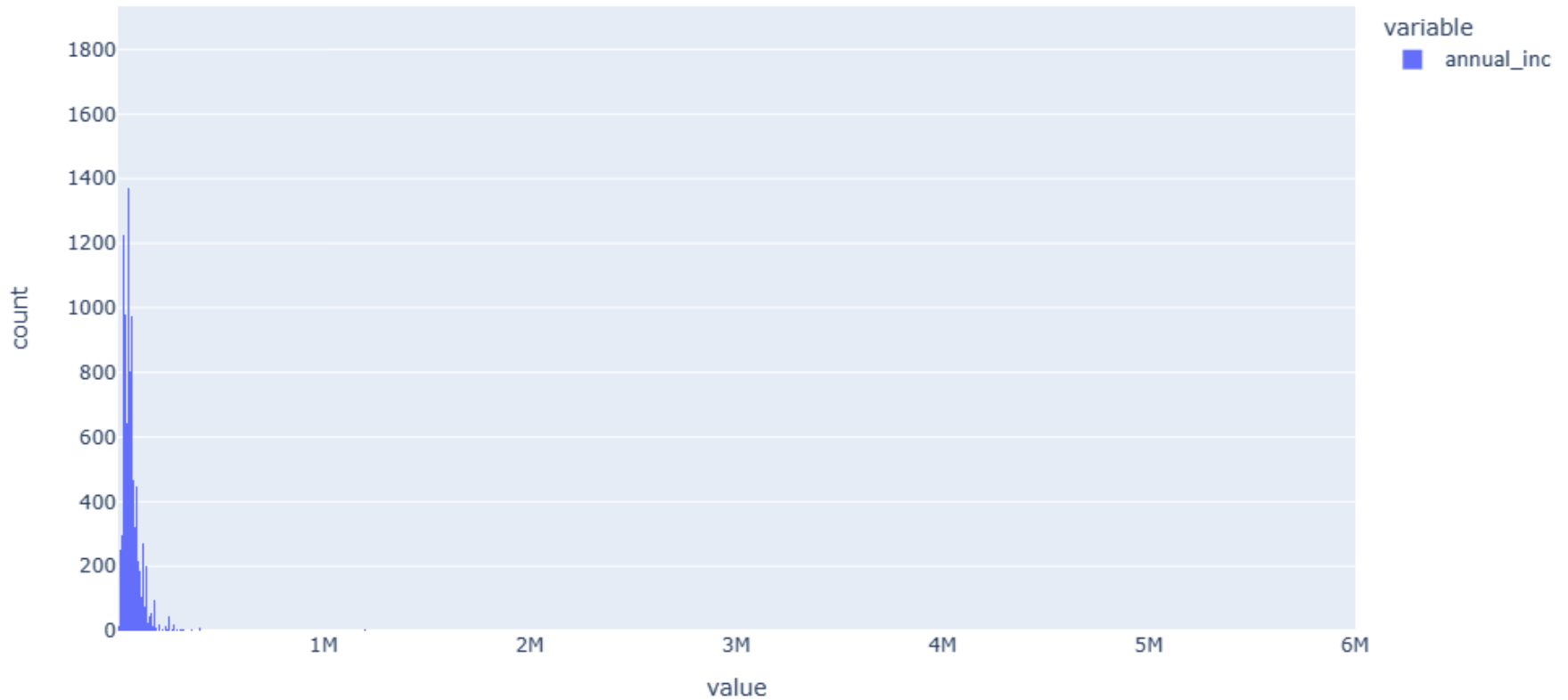
- Three different types of derived metrics::
- **Type-driven metrics**
- **Business-driven metrics**
- **Data-driven metrics**

Inferences univariient



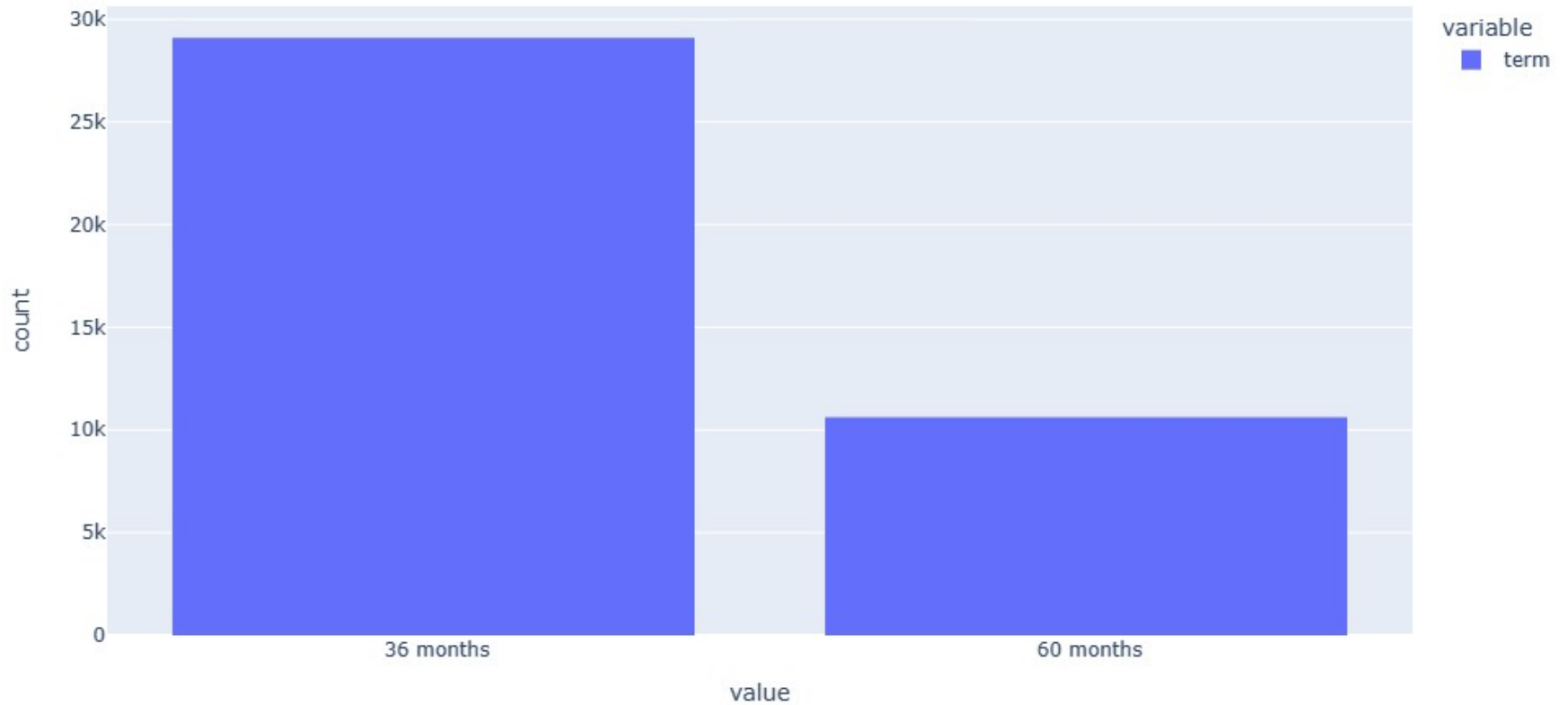
The frequency of loan amount 10k is more across the data set with frequency 3017 people requirement of 10k loan is greater in number

Inferences univariert



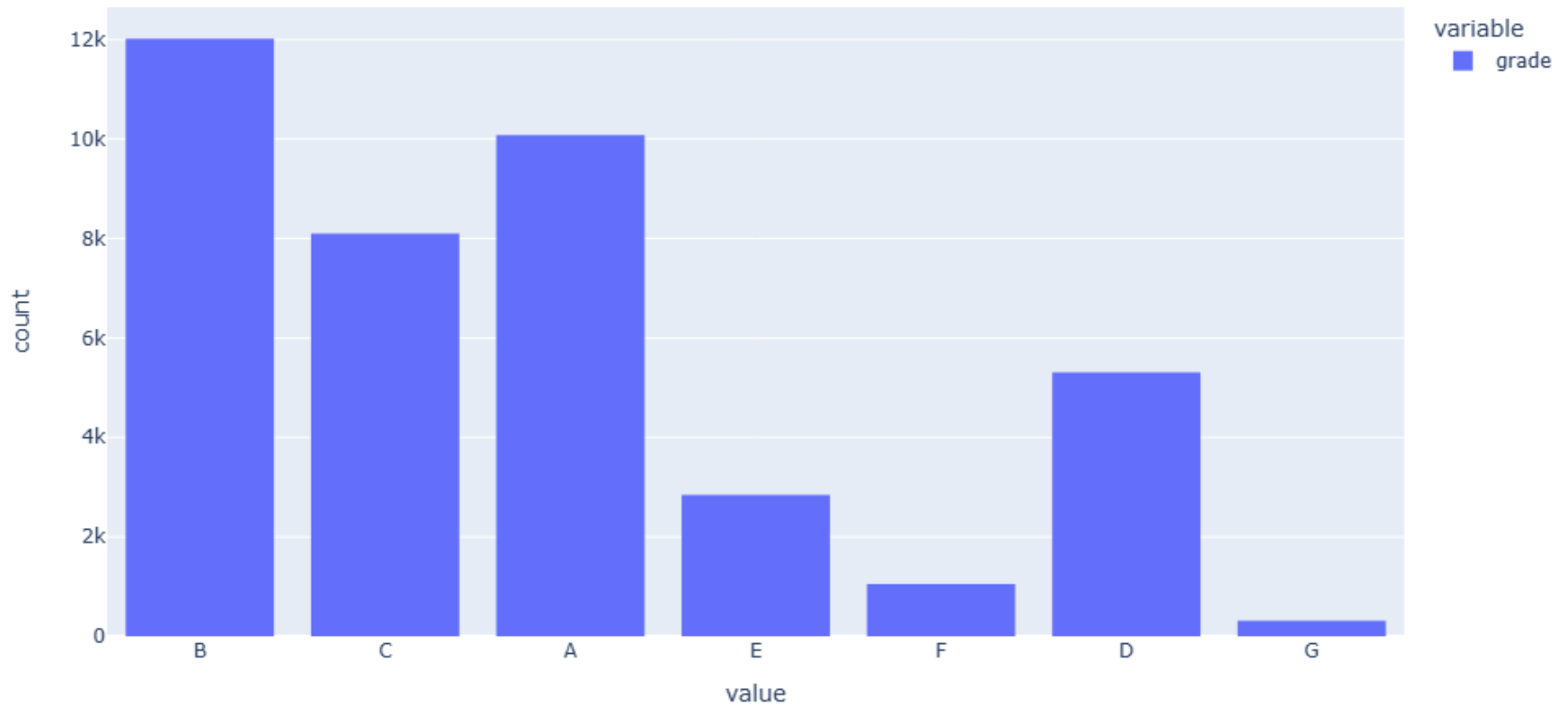
The majority of the people opted for loan there annual income is less that 0.5M

UNIVARIENT ANALYSIS CATAGORICAL



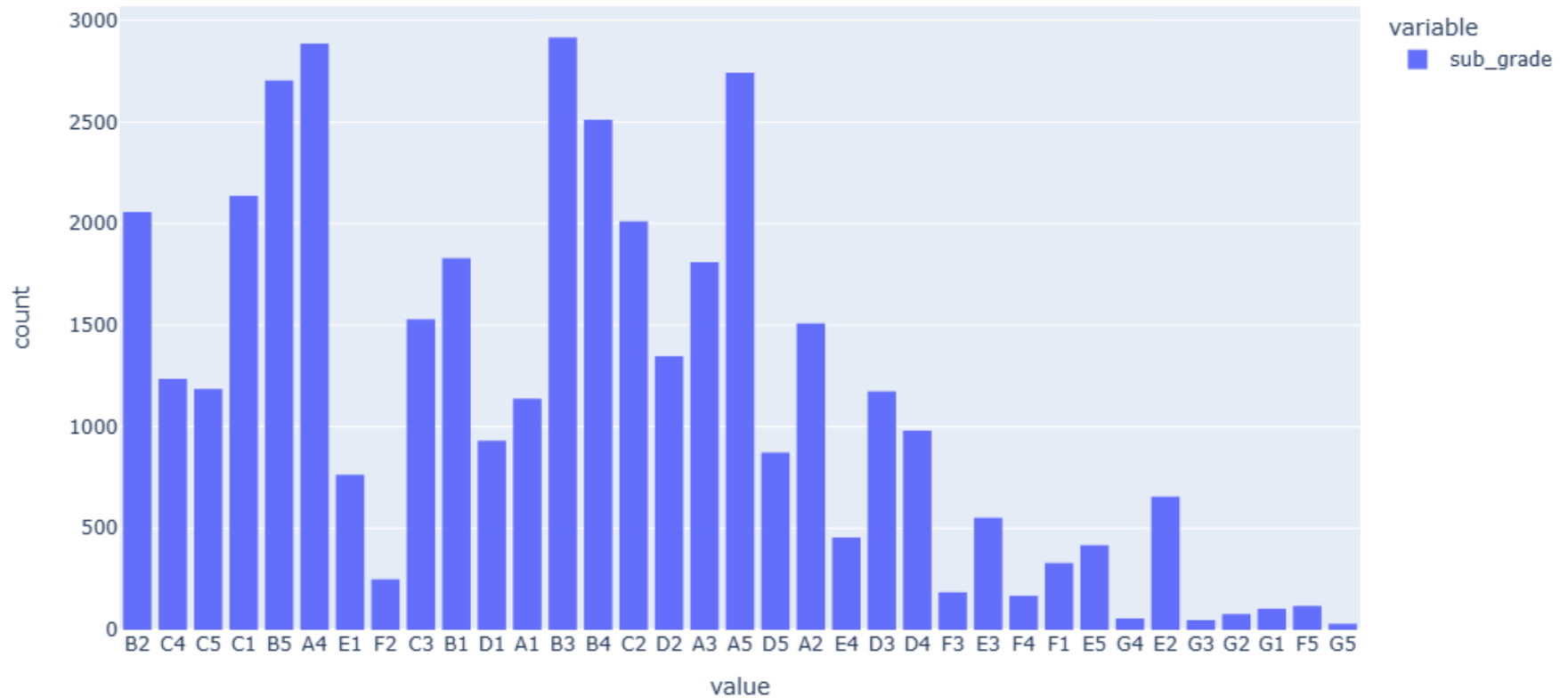
Loan opted for 36 term is more compared 60 months

UNIVARIANT ANALYSIS CATEGORICAL



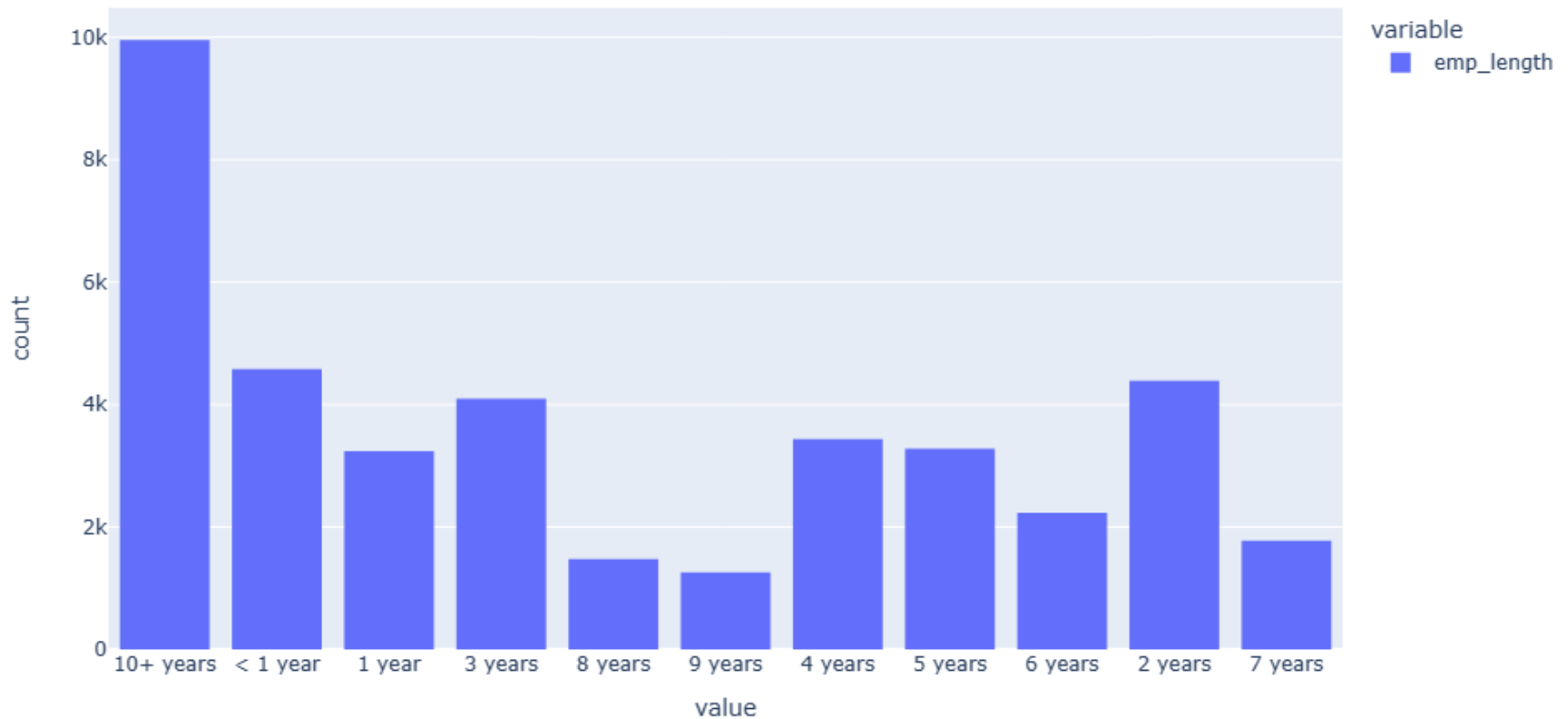
Grade B loans are opted more than any other grade and the next Grade is A followed by C, D, E, F and G

UNIVARIANT ANALYSIS CATEGORICAL



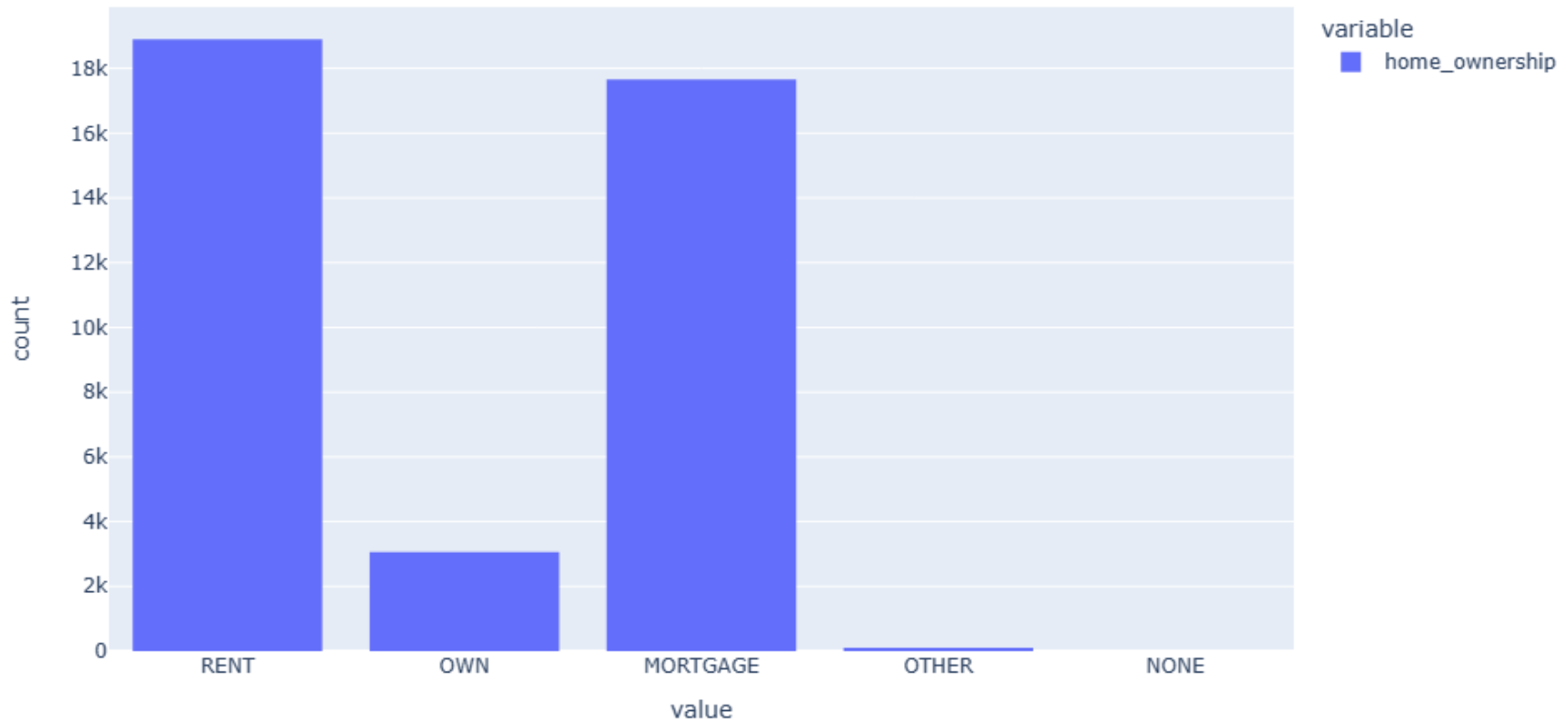
Sub Grades B3 opted more than any other grades followed by A4 and A5

UNIVARIANT ANALYSIS CATEGORICAL



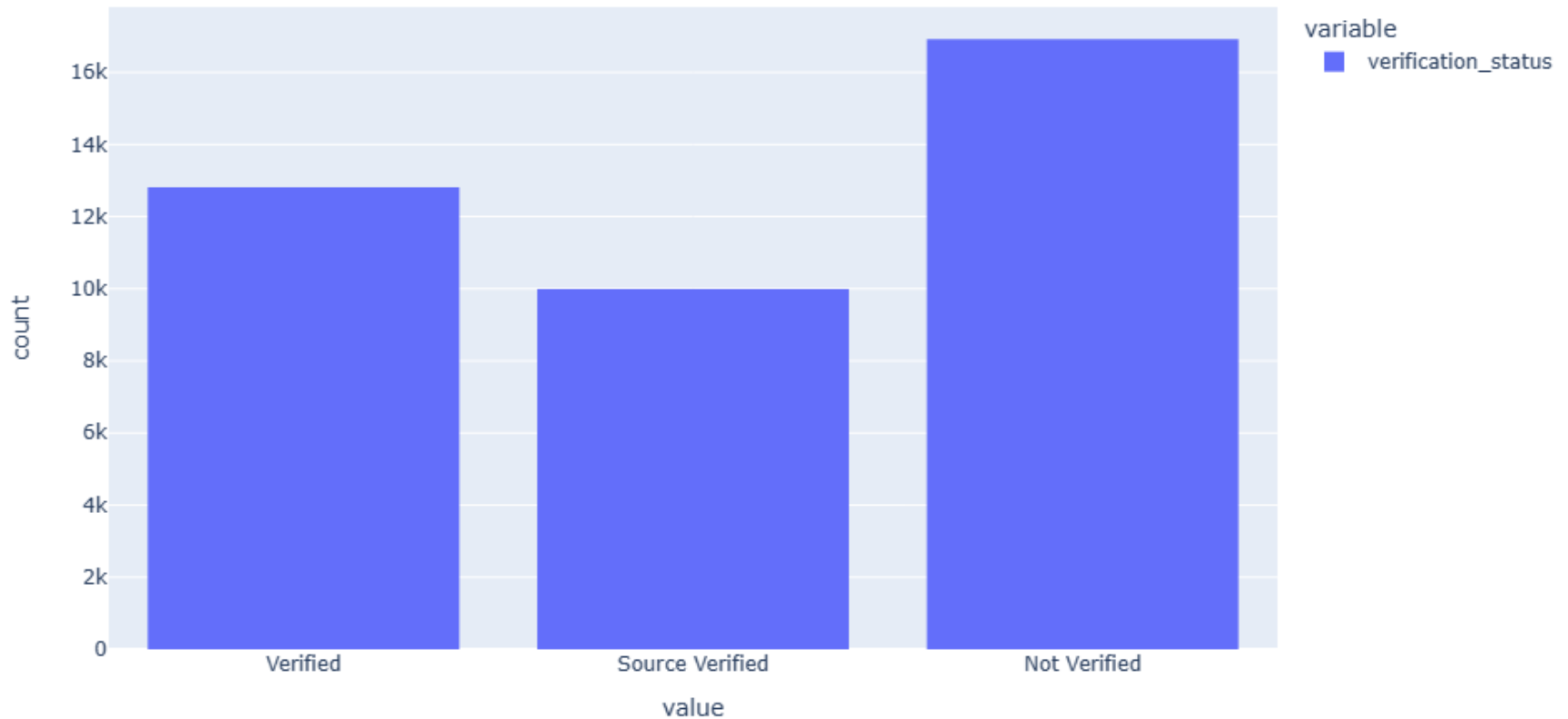
Employees with 10 years has opted loan more compared to others followed by <1 year and 2 years

UNIVARIANT ANALYSIS CATEGORICAL



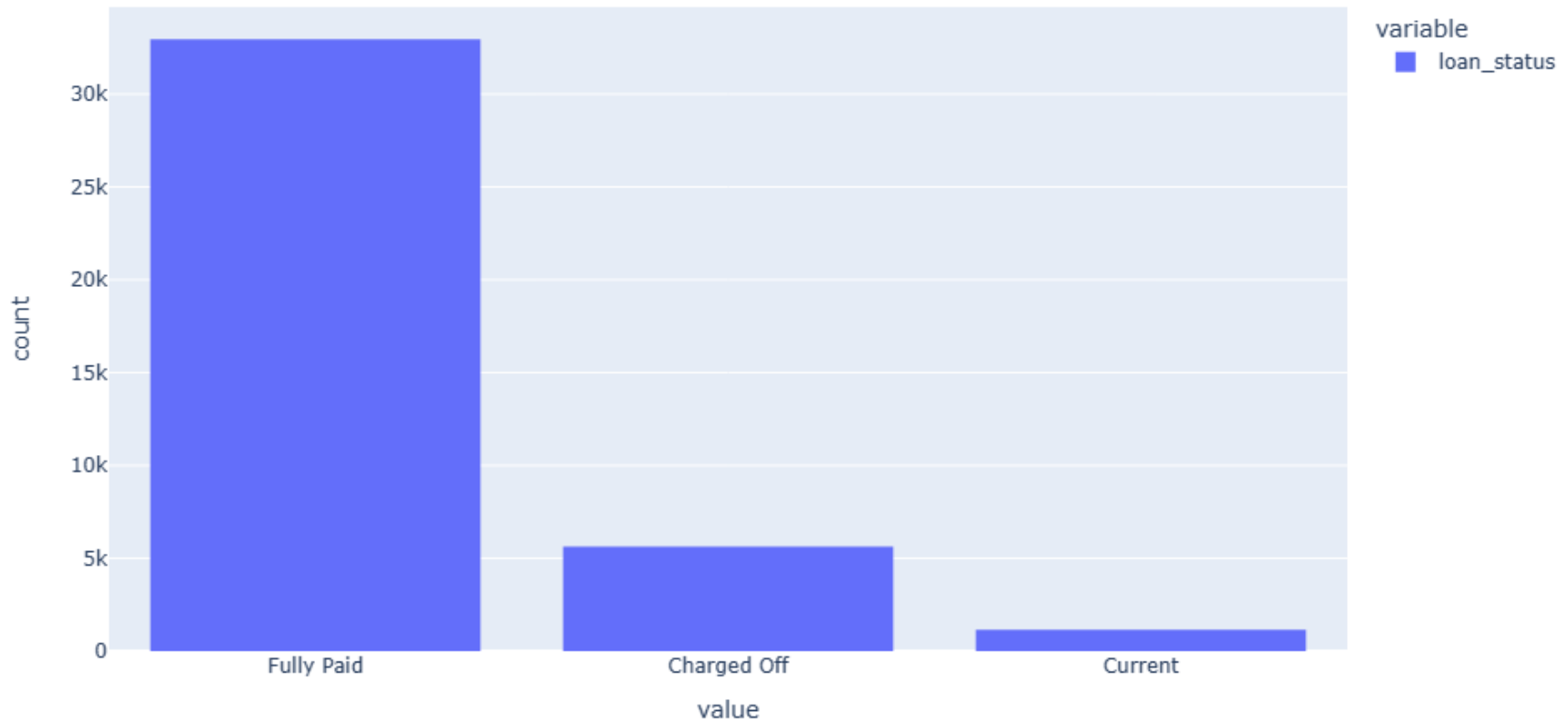
People with home ownership as Rent and mortgage has taken loan when compared to own and others

UNIVARIANT ANALYSIS CATEGORICAL



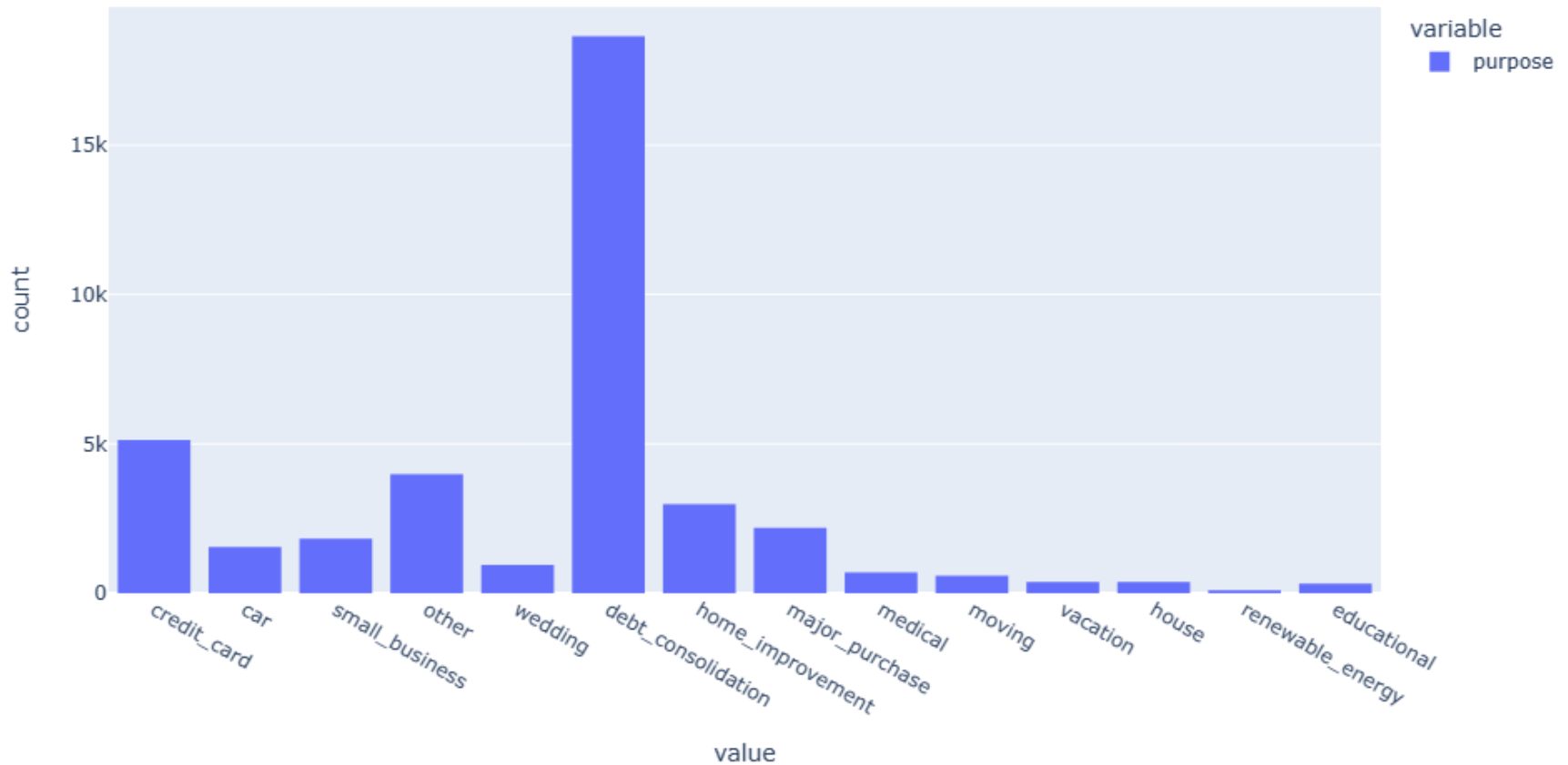
people with not verified income source has taken loan more in number when compared with others

UNIVARIANT ANALYSIS CATEGORICAL



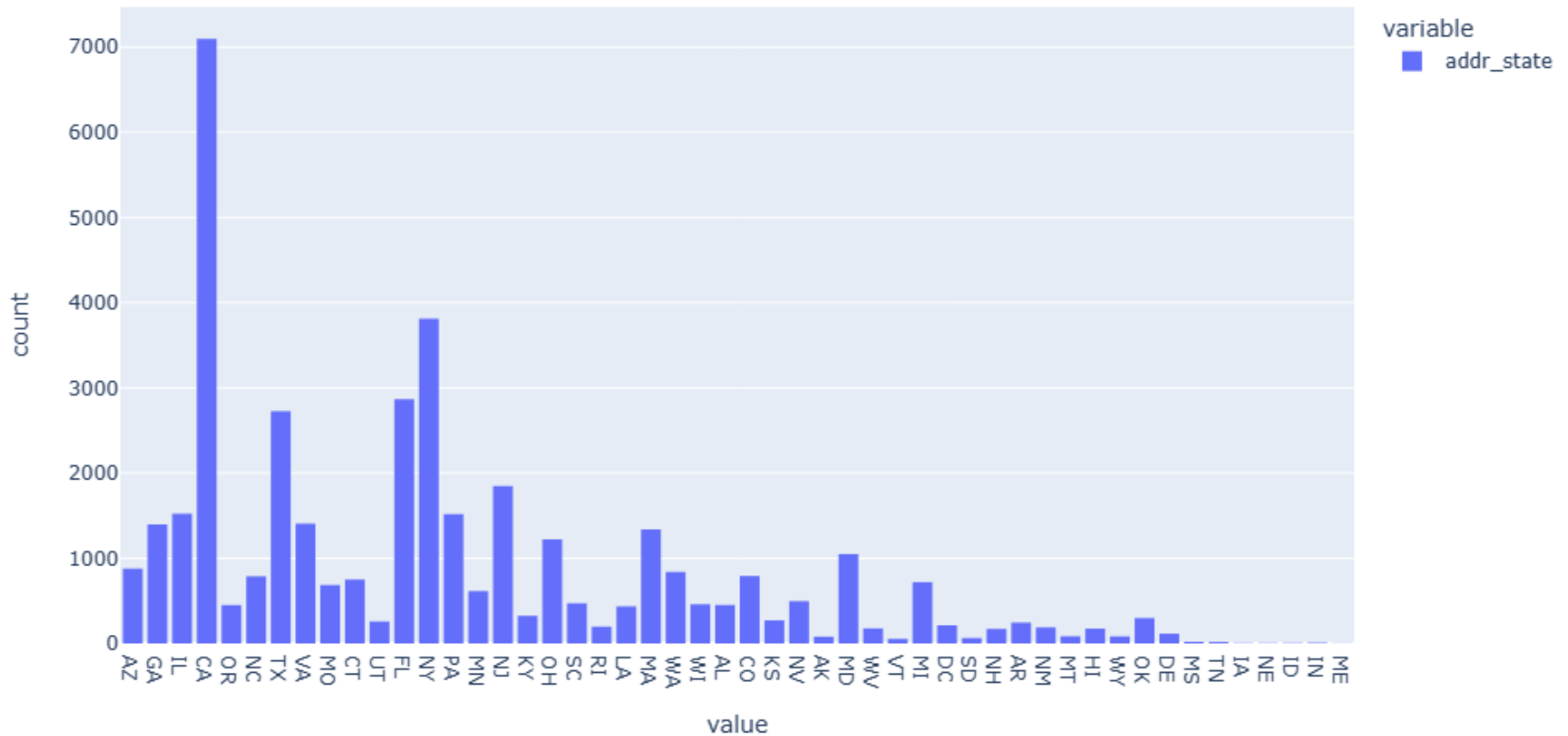
People with loan status charged off are 5k where fully paid are 32k and current are 1k

UNIVARIANT ANALYSIS CATEGORICAL



People with debt consolidation purpose have opted loan more

UNIVARIANT ANALYSIS CATEGORICAL



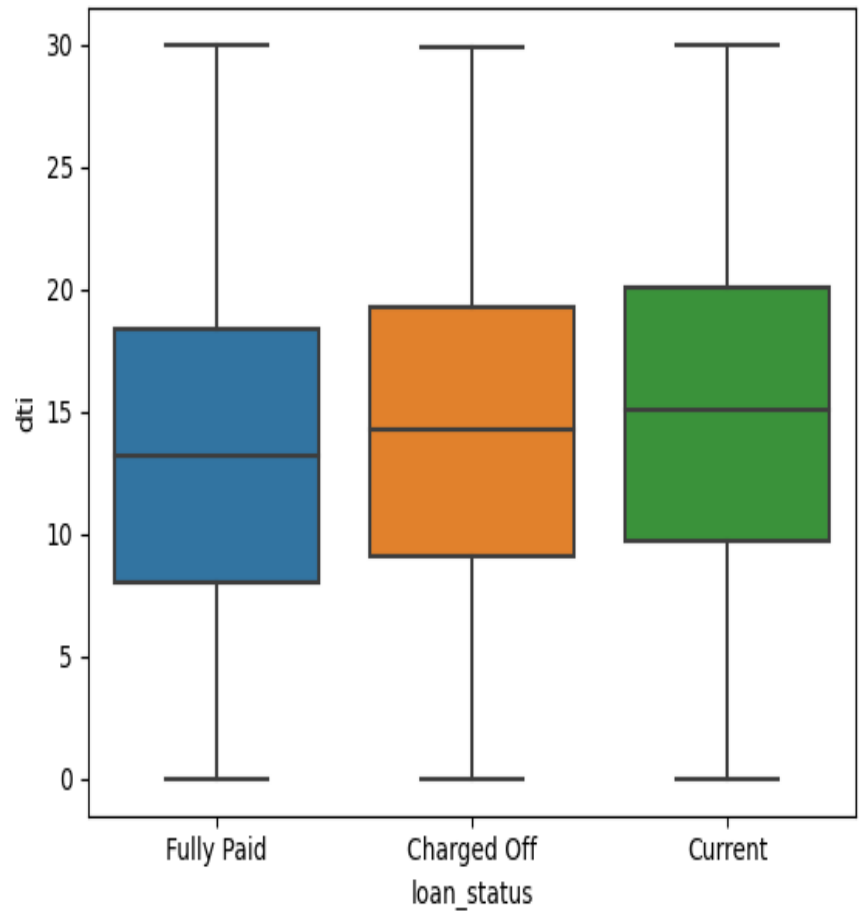
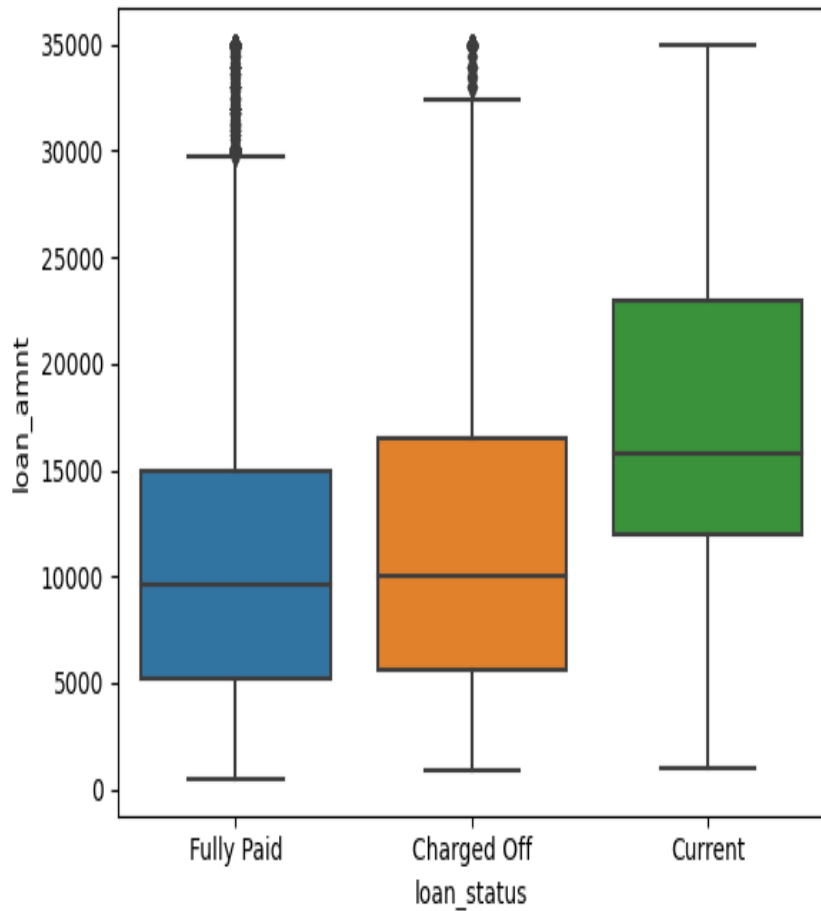
People who belong to CA state has opted/taken loan more in number when compared to other states follows NY and FL.

BIVARIANT ANALYSIS CATEGORICAL

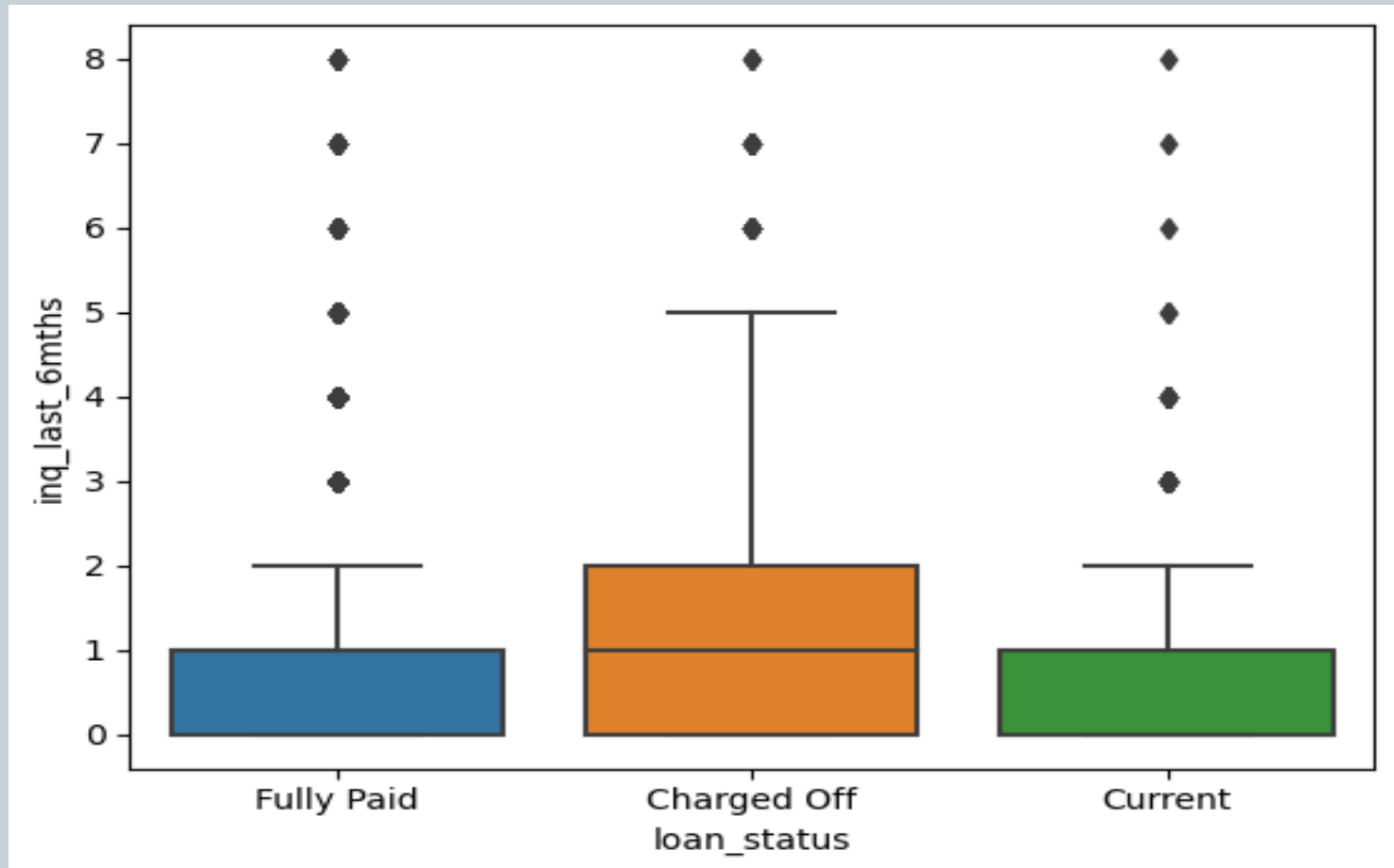


- Loan amount vs loan status --the median is slight higher for charged off but no major difference with full paid
- 2) DTI average/median is slight more for the charged off status when compared to fully paid and current
- 3) The number of enquires for past 6 MONTHS more is more for charged off status.
- Visualisation next page

BIVARIANT ANALYSIS CATEGORICAL



BIVARIANT ANALYSIS CATEGORICAL



Important inferences

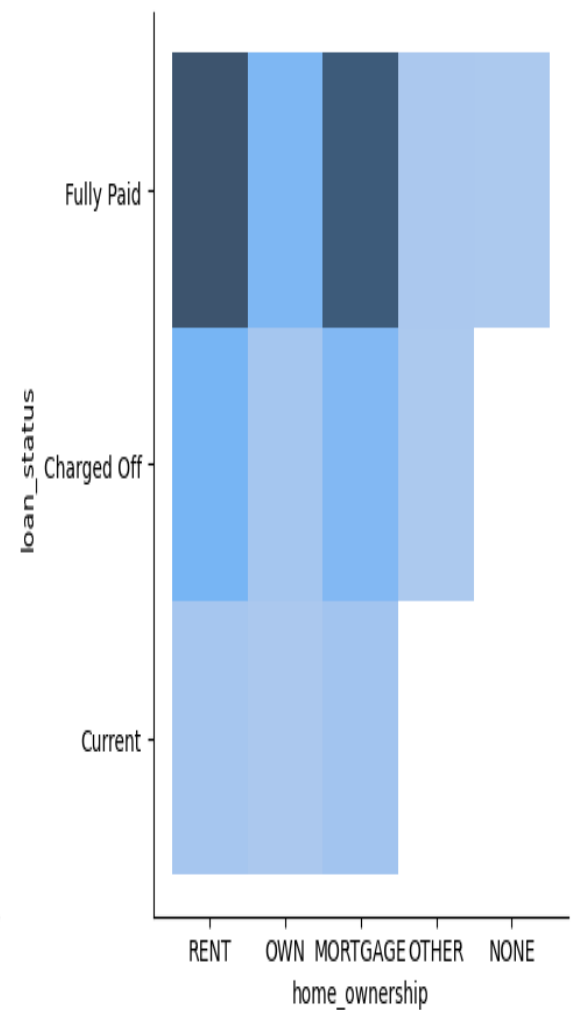
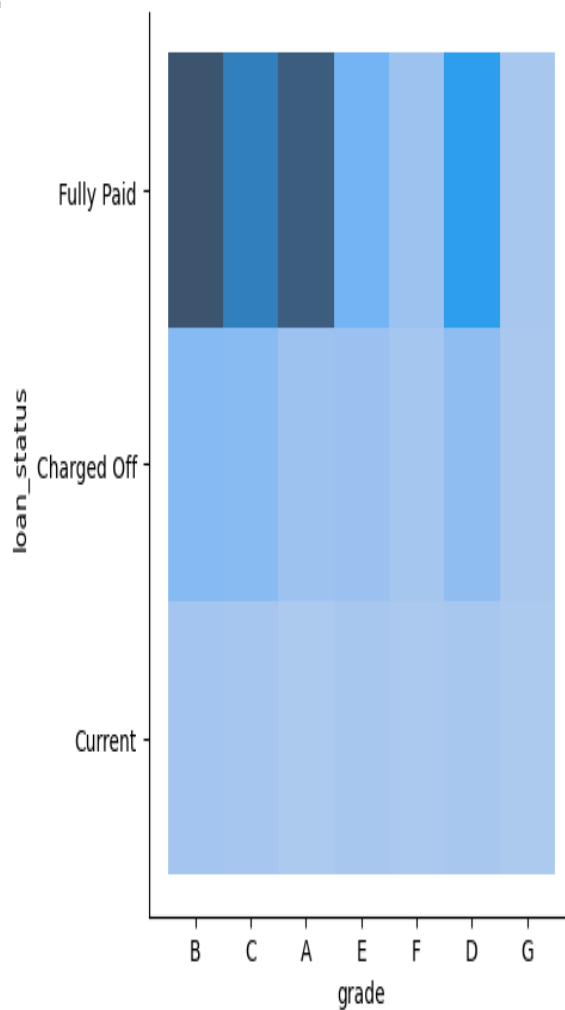
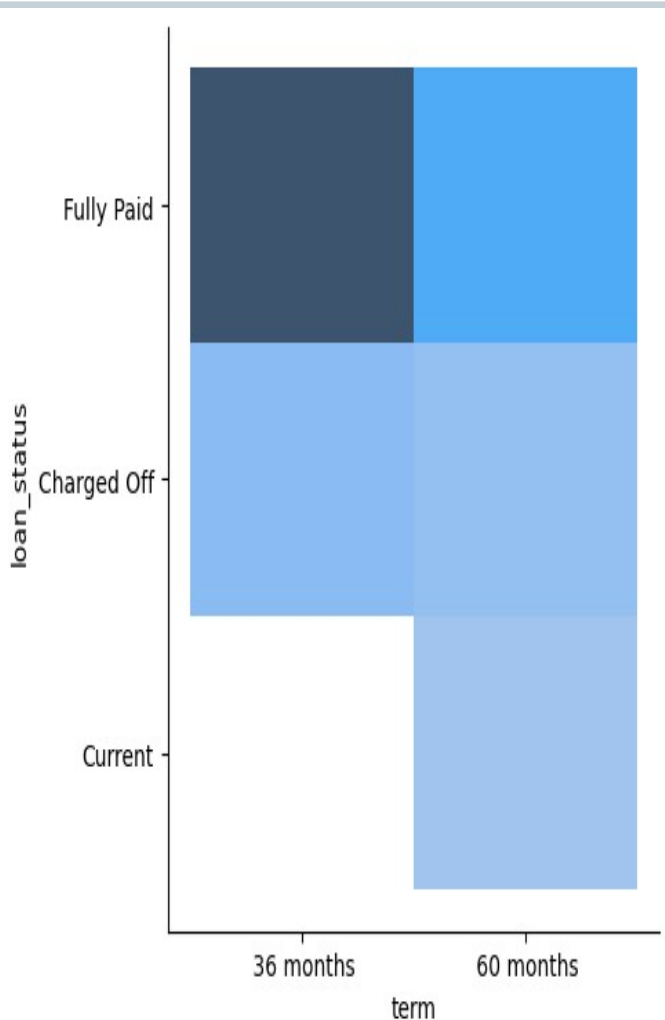


Inferences

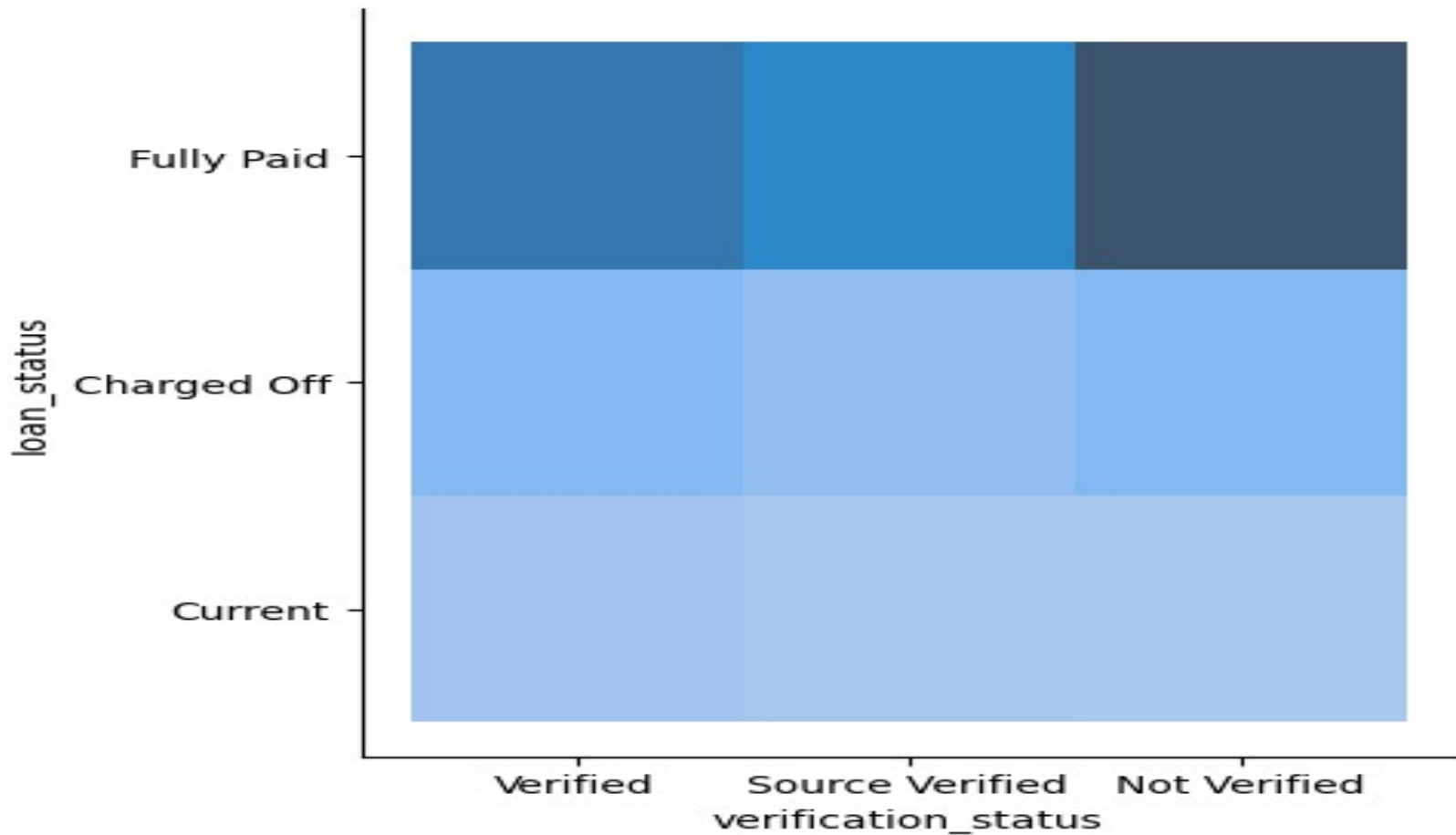
- 1) loan with 36 term have more number of charged off
- 2) loan with grade B,C and D have more charged off
- 3) people with home ownership as Rent and Mortgage have high charged off loan status
- 4) people with income source not verified have high charged off loan status
- 5) employee with 10+ years have high charged off status
- 6) People who opted loan for the purpose of debt_consolidation have high charged off loan status
- 7) People belong to CA state has high charged off status of loan
- 8) People belong to IN, ME and IA has no signs of charged off status of loan

Visualisation next page

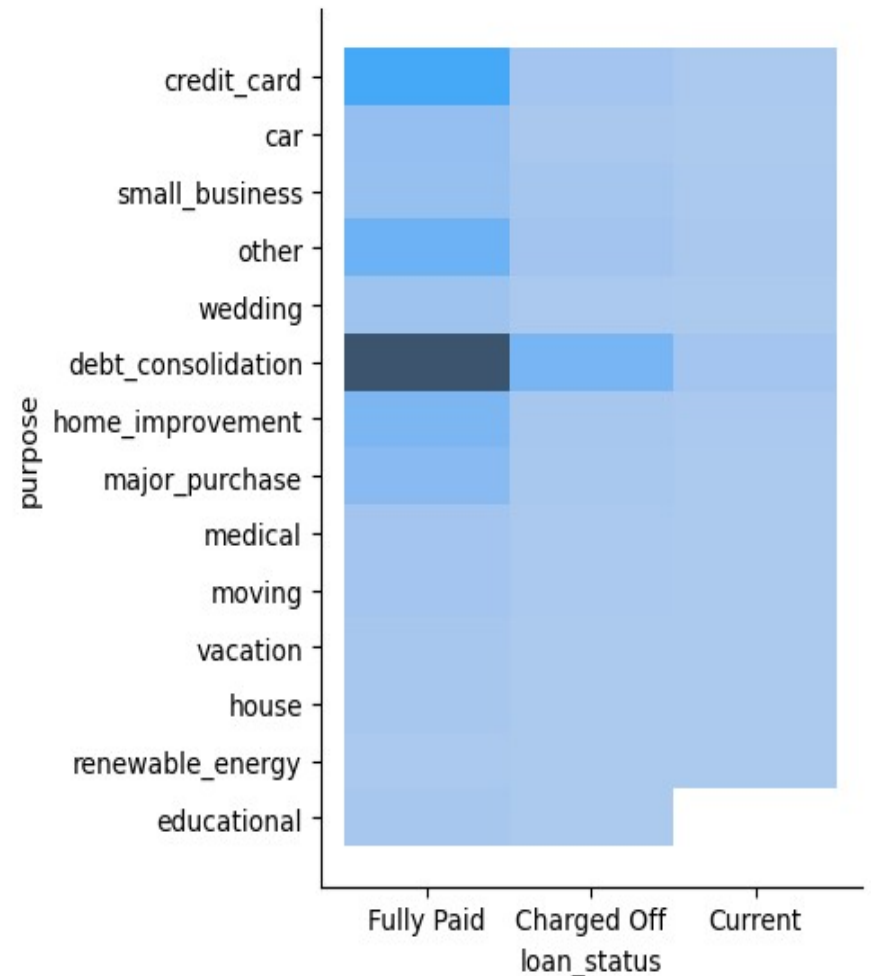
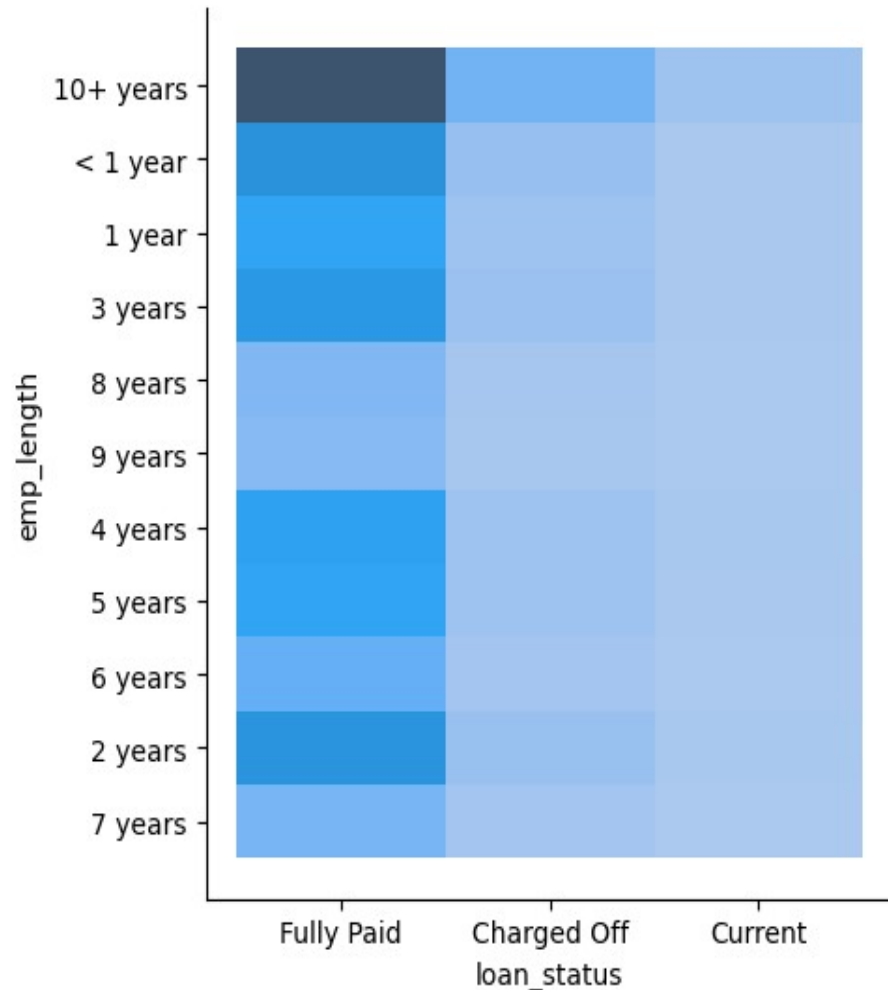
Important inferences



Important inferences



Important inferences



Important inferences

