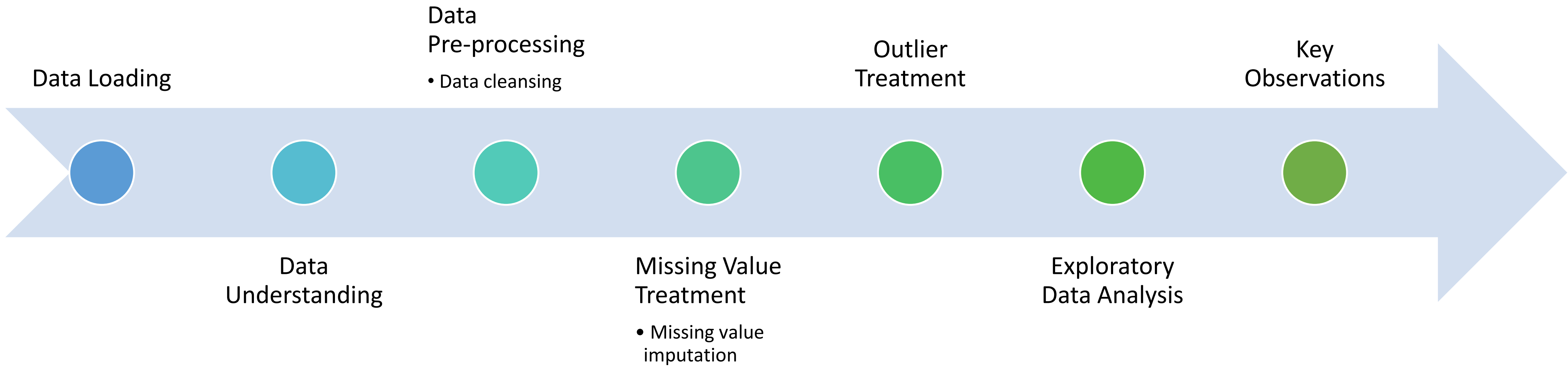
Lending Club Case Study

This EDA(Exploratory Data Analysis) based case study attempts to identify ‘driver’ features to help decide if a new loan application should be approved or rejected for the overall growth of the business.

by Feroz Shaikh

Approach

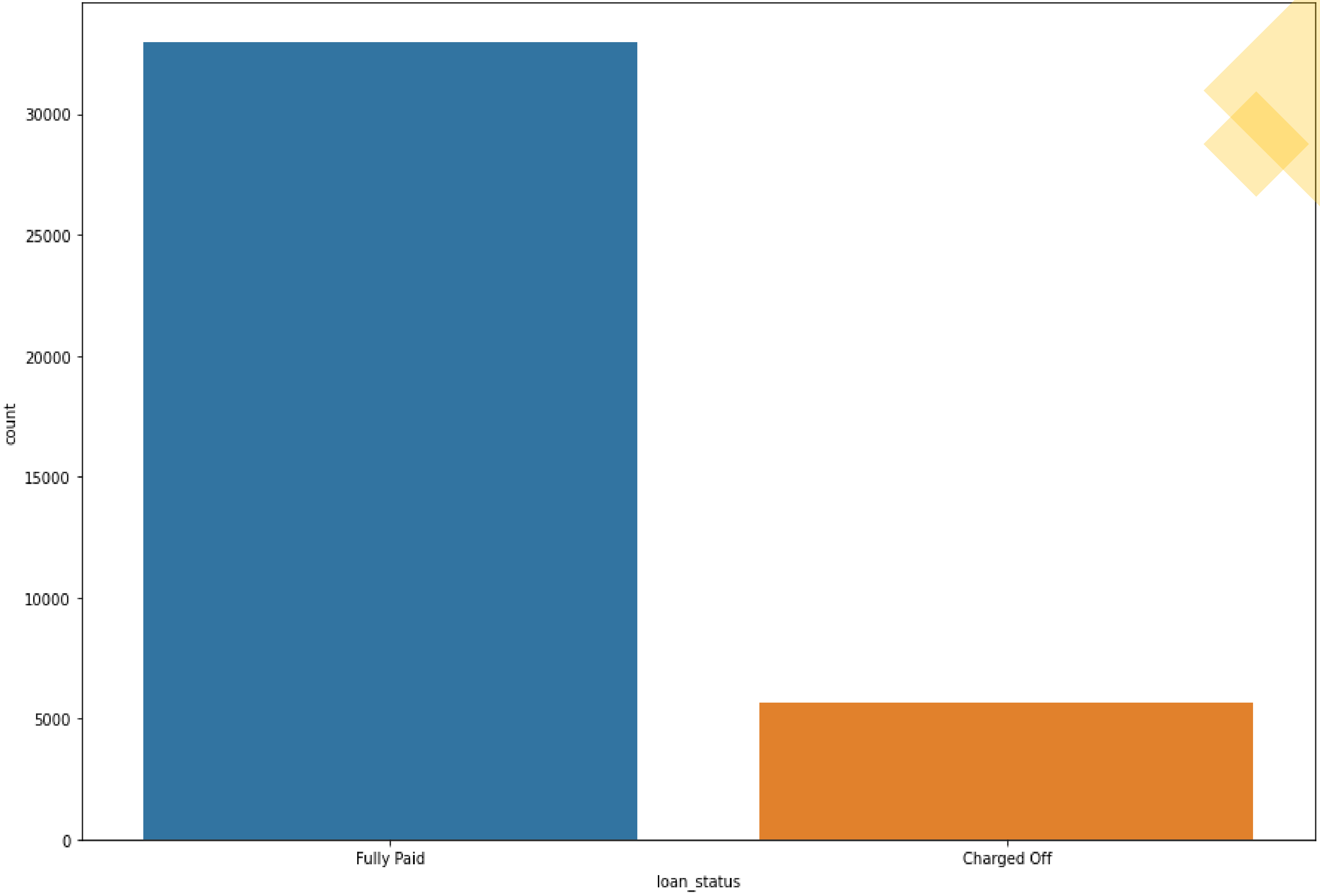


Data Loading
Data Understanding
Data Pre-processing
• Data cleansing
Missing Value Treatment
• Missing value imputation
Outlier Treatment
Exploratory Data Analysis
Key Observations

EDA- 1

Univariate Analysis - Count of Fully Paid vs Charged Off loans.

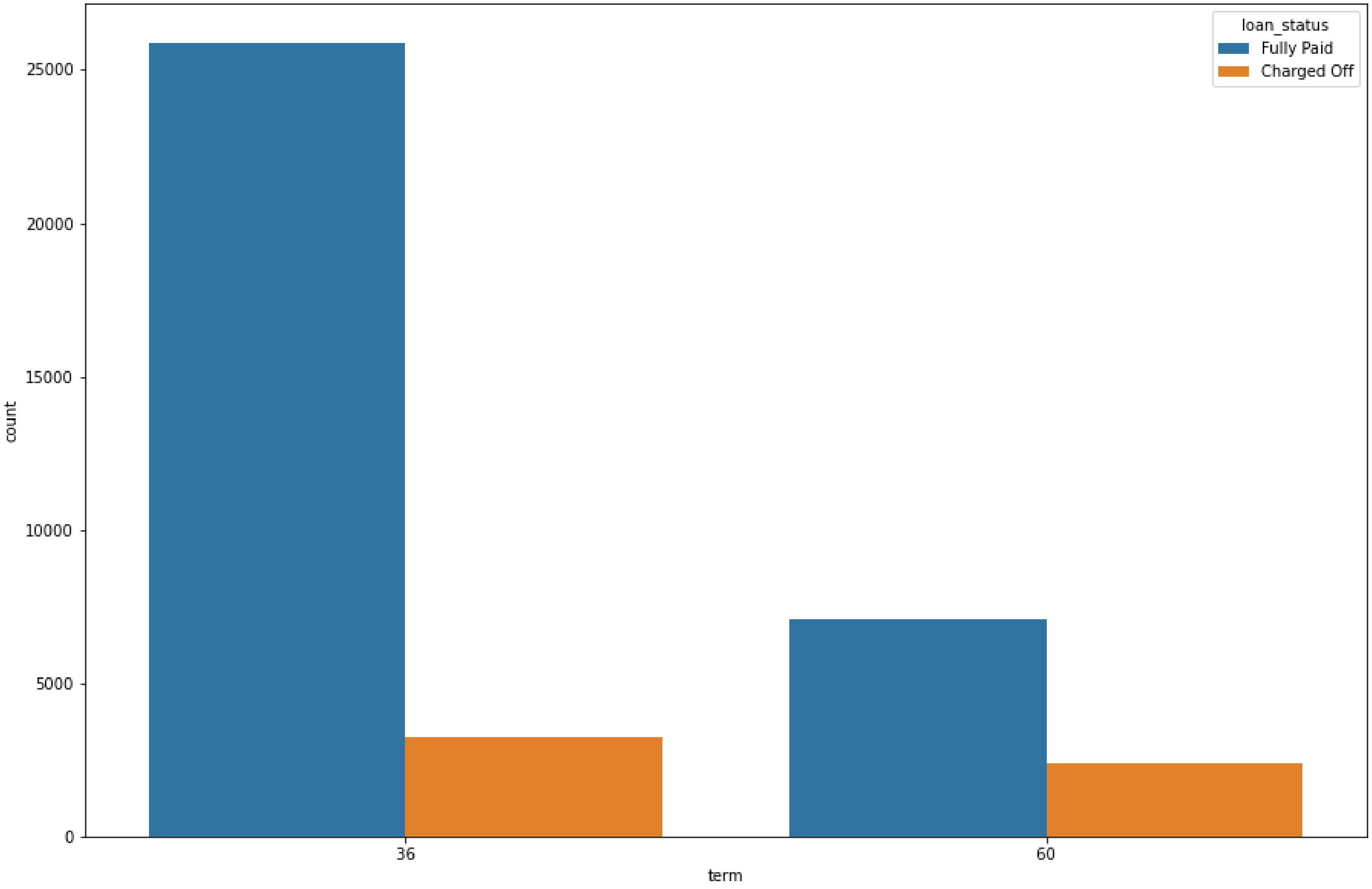
- *The number of defaulters is comparatively less than number of non-defaulters – a healthy sign for business.*



EDA- 2

Univariate Analysis – Loan Term.

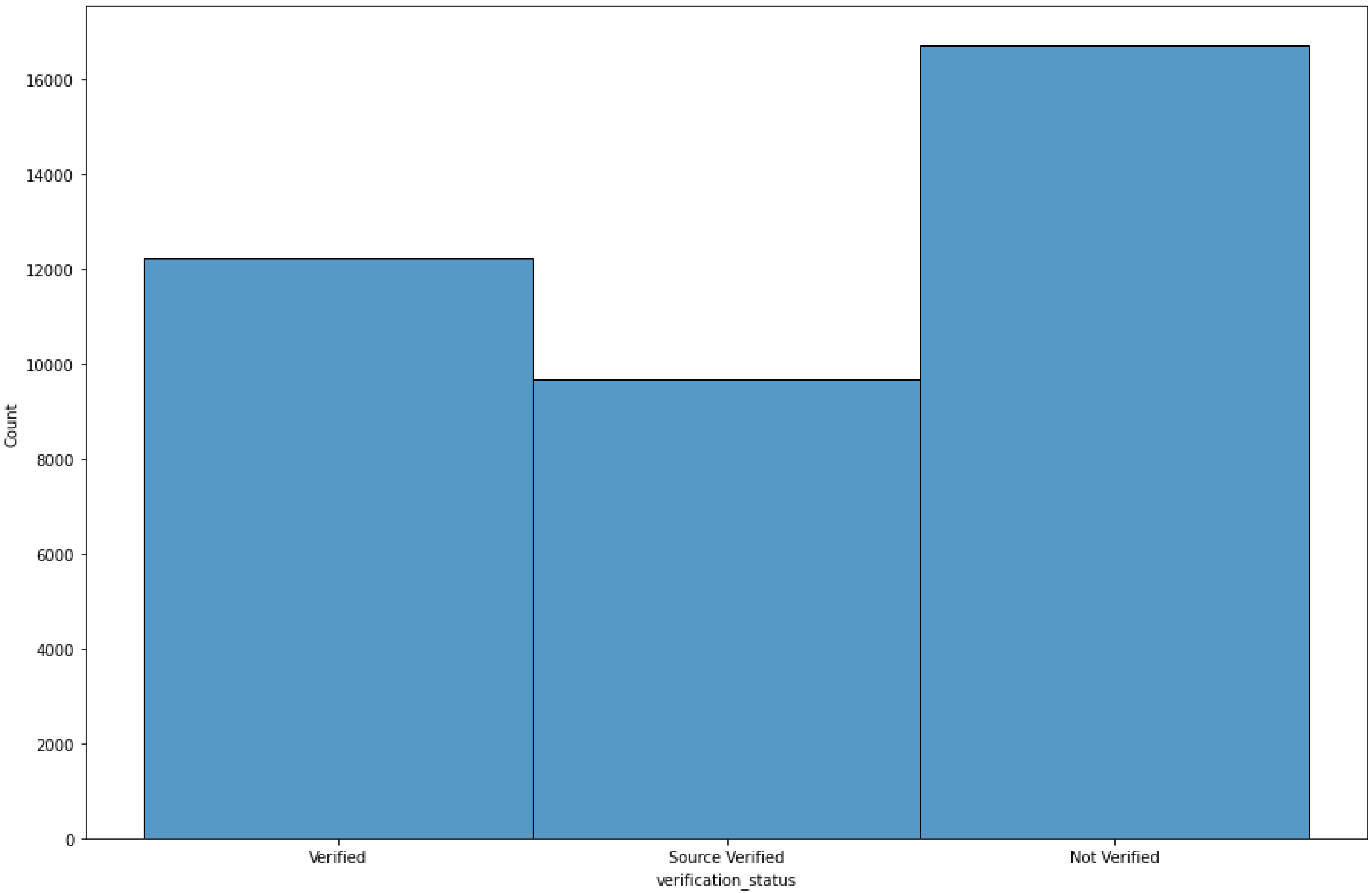
- *Number of defaulters are more for shorter term loans(36 months), compared to longer term loans.*



EDA- 3

Univariate Analysis – Loan Verification Status.

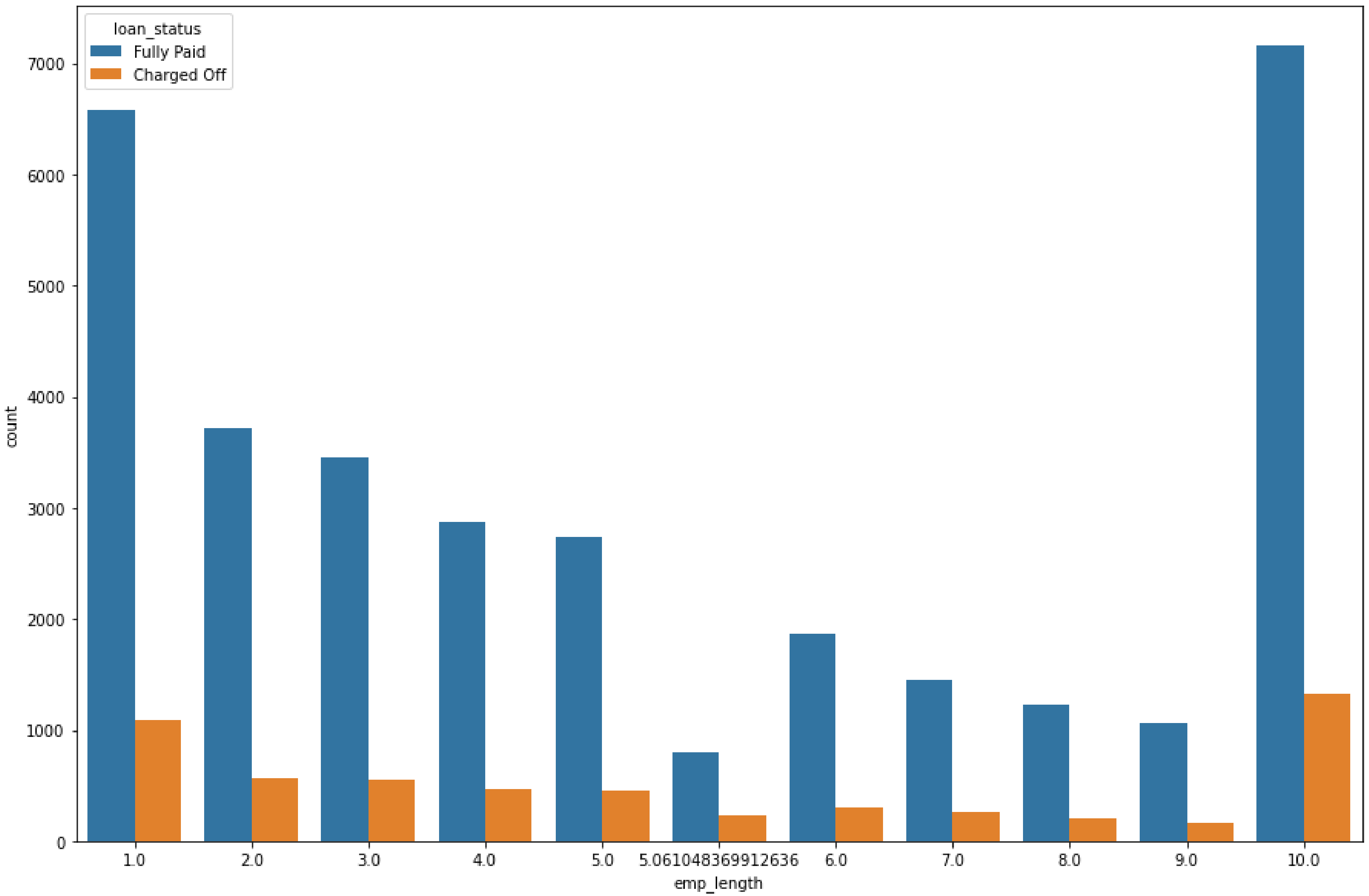
- *The number of NOT VERIFIED loans provided are more than verified loans. This could trickle down to defaulters in future.*



EDA- 4

Univariate Analysis – Employment length.

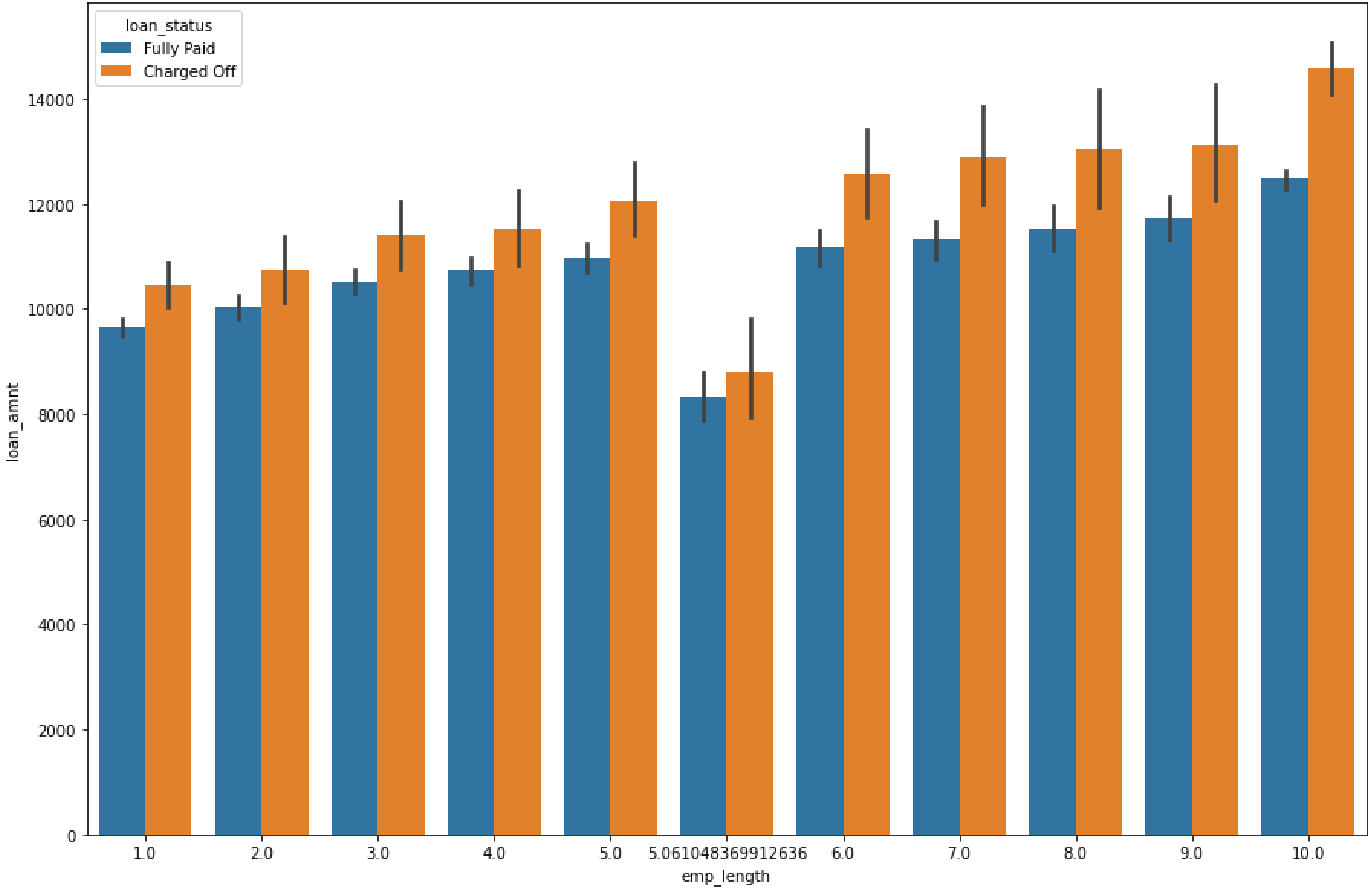
- *Applicants with employment length of less than or equal to 1 year & employees with 10 years of employment are the highest potential defaulters.*



EDA- 5

Bivariate Analysis – Loan Amount v/s Employment Length.

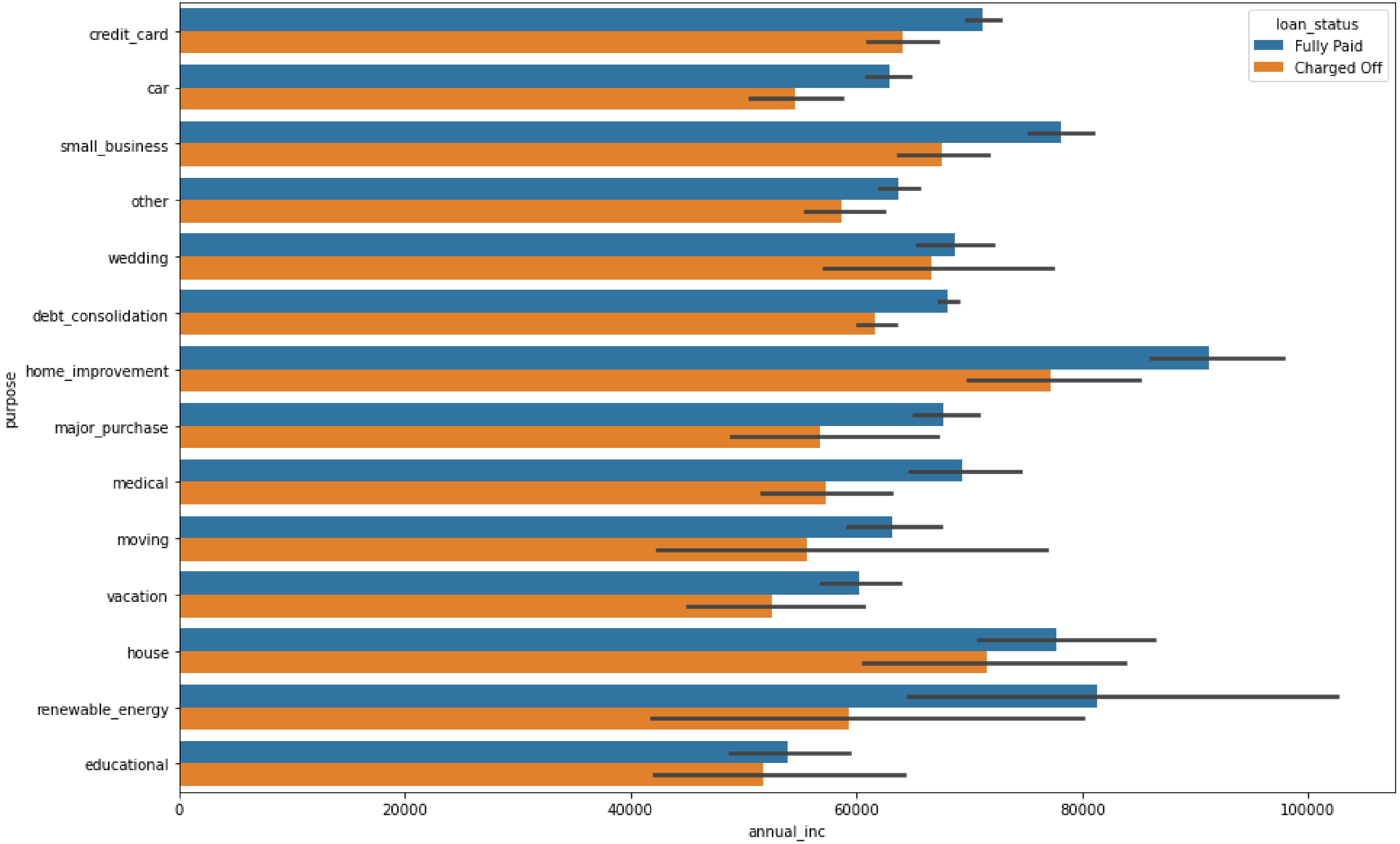
- *It cannot be concluded that applicants with longer employment length end up fully repaying the loan.*



EDA- 6

Bivariate Analysis – Loan Amount v/s Employment Length.

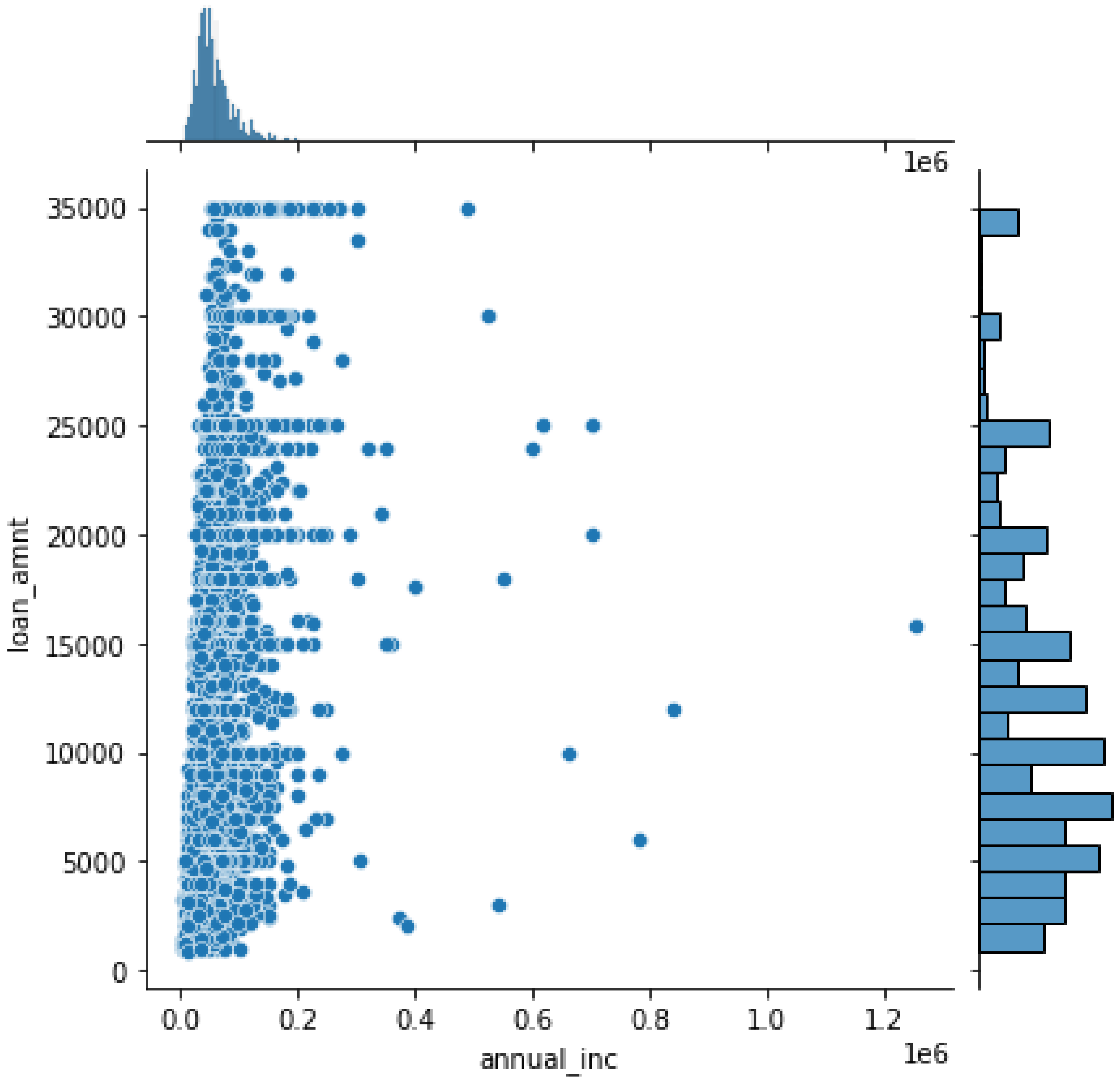
- *Higher annual incomes applicants should be considered for 'home improvement' loan approval.*
- *Even lower annual incomes applicants can be considered for 'education' loan approval.*



EDA- 7

Bivariate Analysis – Annual Income vs loan amount.

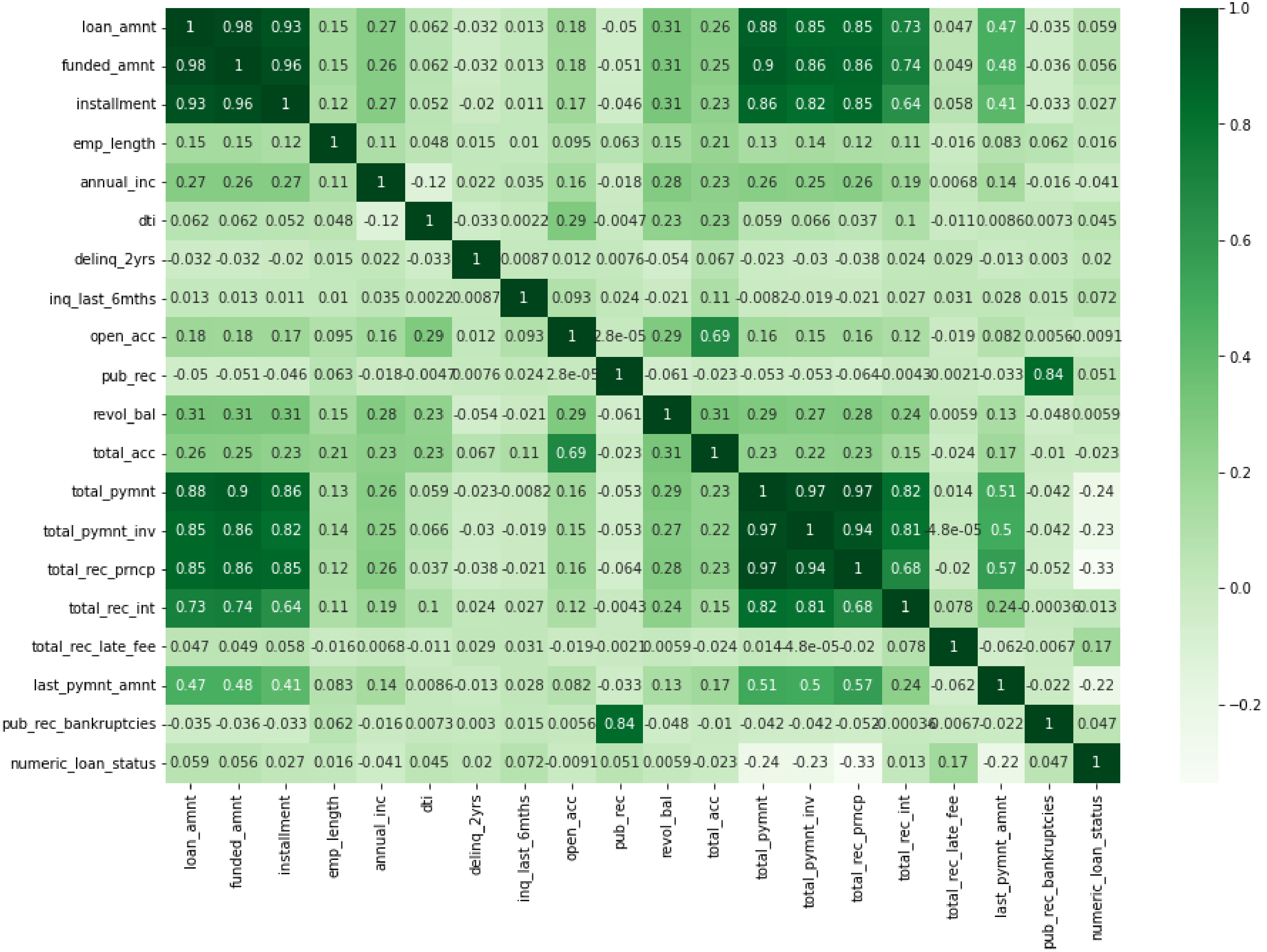
- *Lower annual income applicants applying for lower loan amounts tend to default more.*



EDA- 8

Multivariate Analysis – Heat Map.

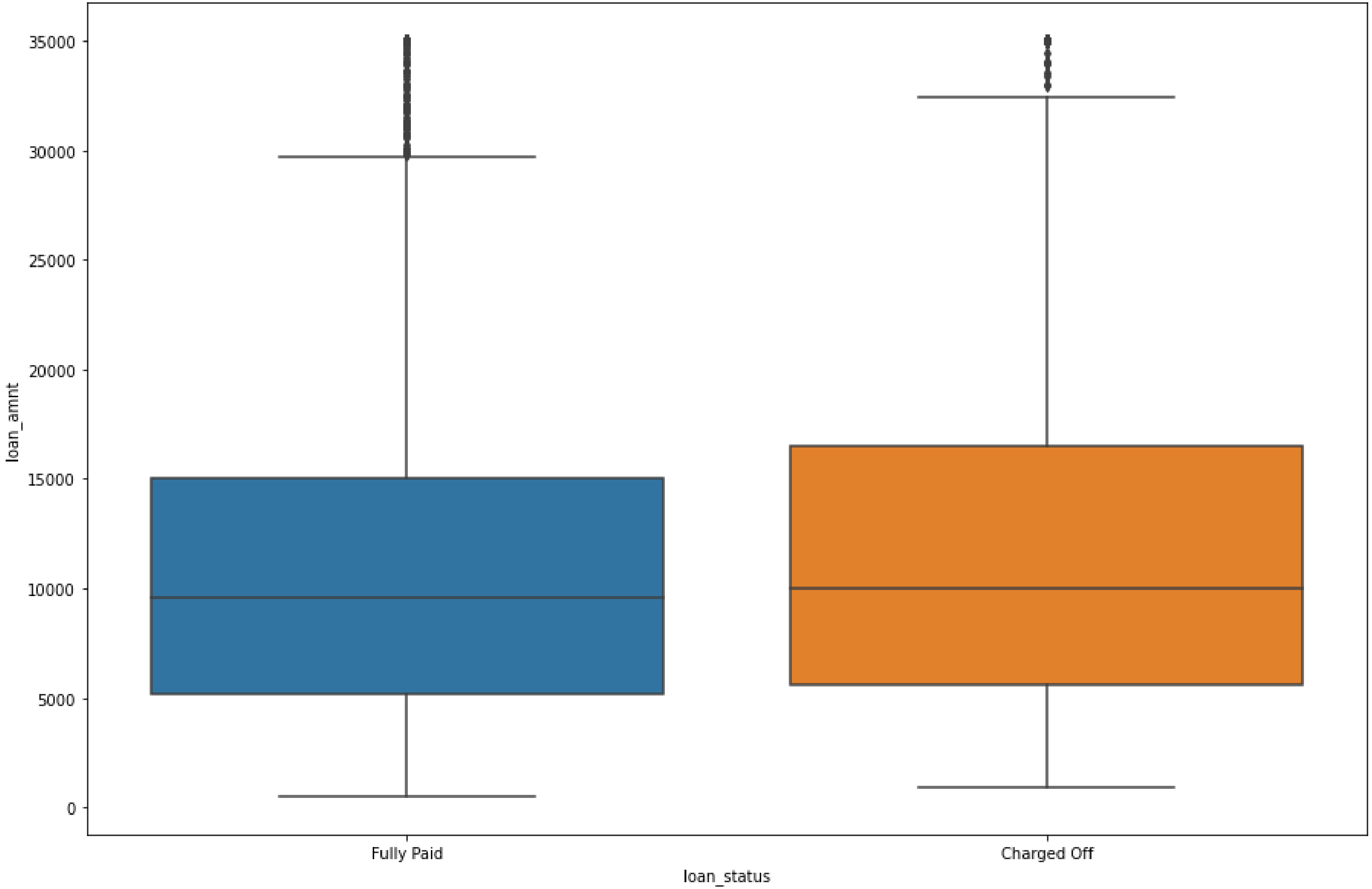
- *Loan applicants with higher annual income are slightly better candidates for loan approval.*
- *Employees working for longer years cannot be considered as strong candidates for higher loan amounts.*



EDA- 9

Segmented Univariate Analysis – Loan Amount.

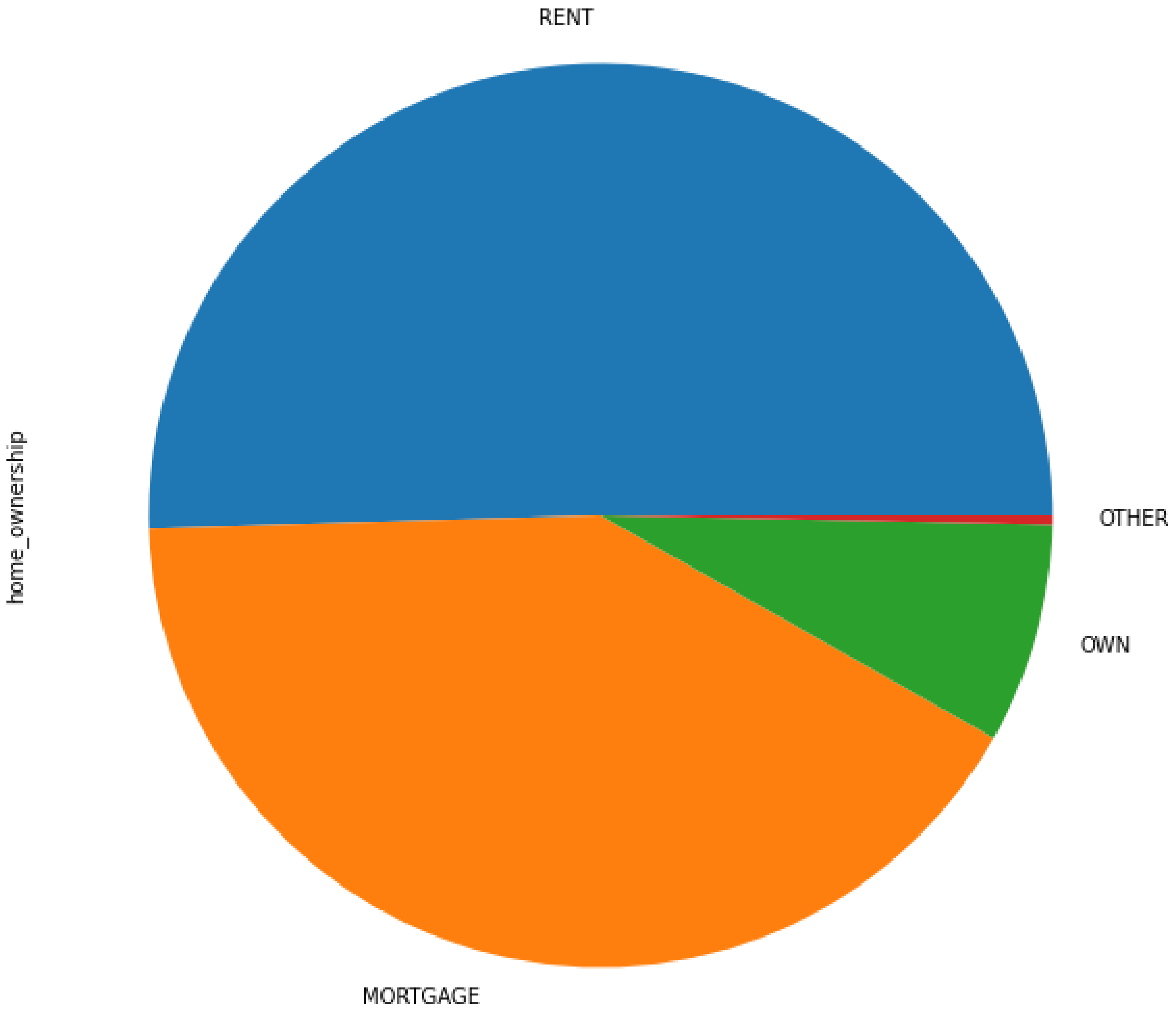
- *Higher loan amount applications end up defaulting more.*



EDA- 10

Segmented Univariate Analysis – Home ownership.

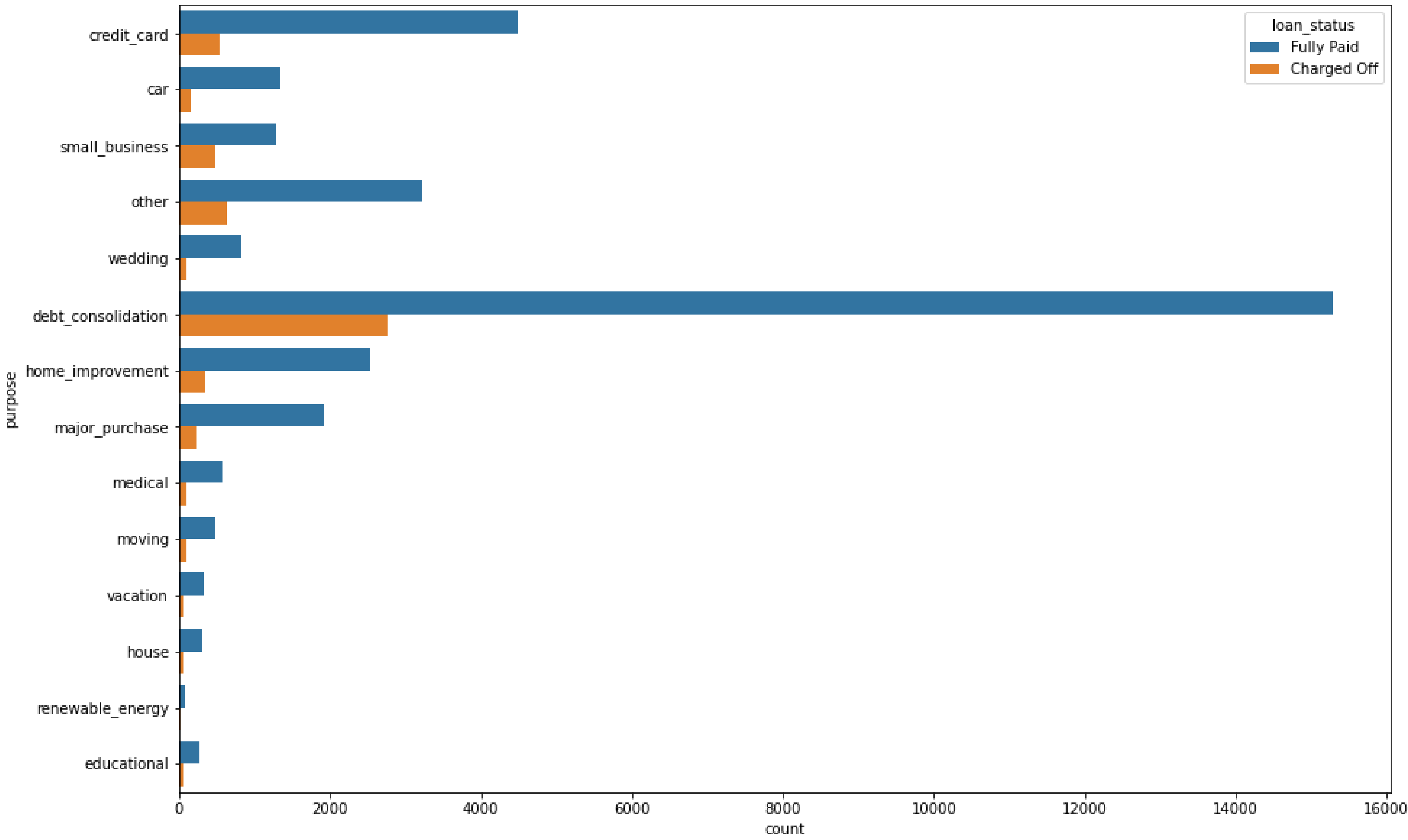
- *It is better to consider applicants with ‘OWN’ houses for approval.*



EDA- 10

Segmented Univariate Analysis – Loan Purpose.

- *It cannot be concluded that ‘debt consolidation’ loan purpose will end up defaulting more.*



Conclusion / Observations

- *Number of defaulters are more for shorter term loans(36 months), compared to longer term loans.*
- *Applicants with employment length of less than or equal to 1 year & employees with 10 years of employment are the highest potential defaulters.*
- *While longer employment length might sound intuitive, but it cannot be concluded that applicants with longer employment length end up fully repaying the loan.*
- *Higher annual incomes applicants should be considered for 'home improvement' loan approval.*
- *Even lower annual incomes applicants can be considered for 'education' loan approval.*
- *Lower annual income applicants applying for lower loan amounts tend to default more.*
- *Loan applicants with higher annual income are slightly better candidates for loan approval.*
- *Employees working for longer years cannot be considered as strong candidates for higher loan amounts.*
- *Higher loan amount applications end up defaulting more.*
- *It is better to consider applicants with 'OWN' houses for approval.*
- *It cannot be concluded that 'debt consolidation' loan purpose will end up defaulting more.*