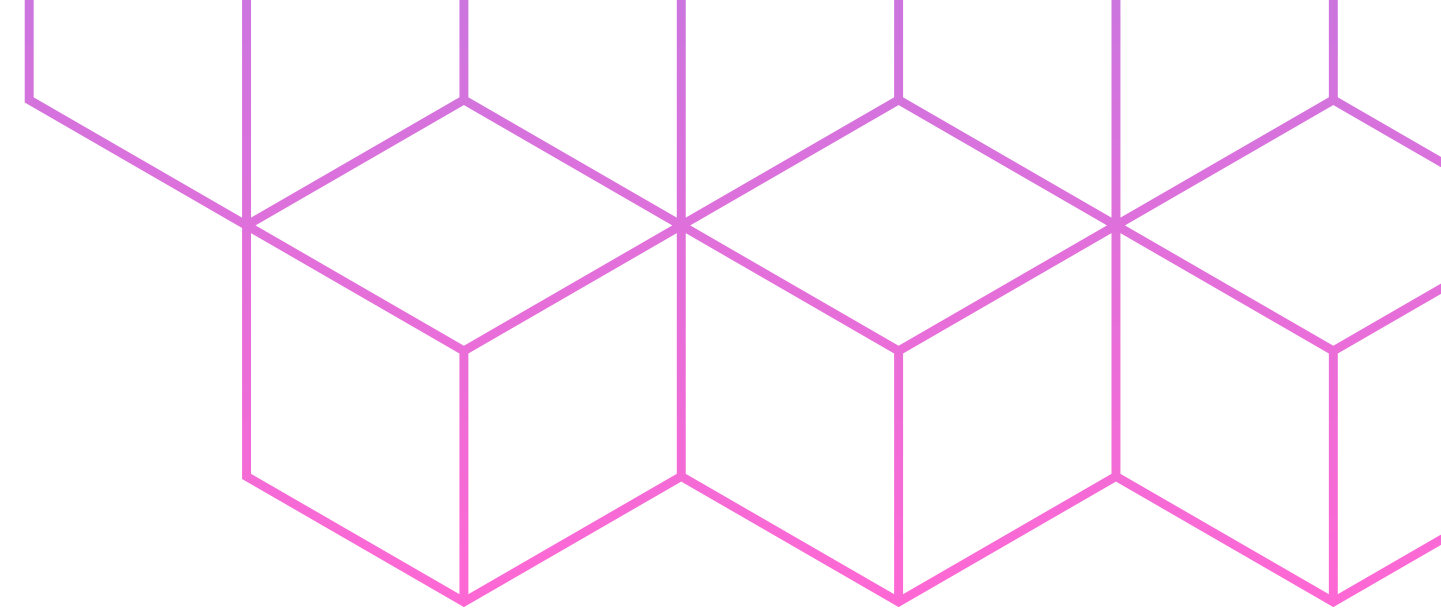


MS - AI and ML- EDA



Lending Case Study

Presented by: Abhishek Singh and Fouziya Memon

Problem Statement

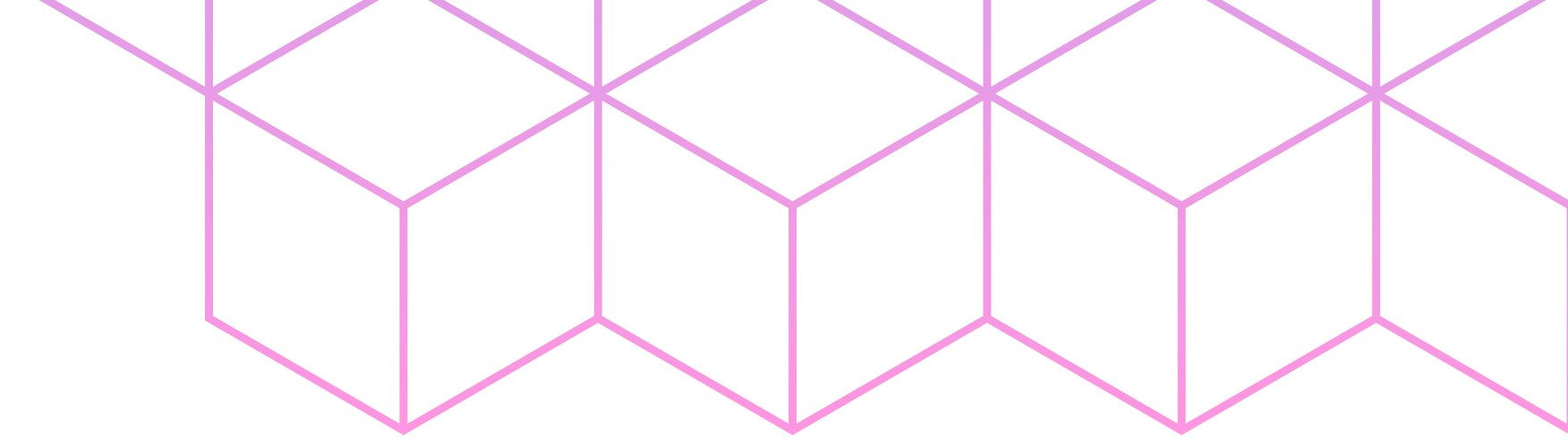
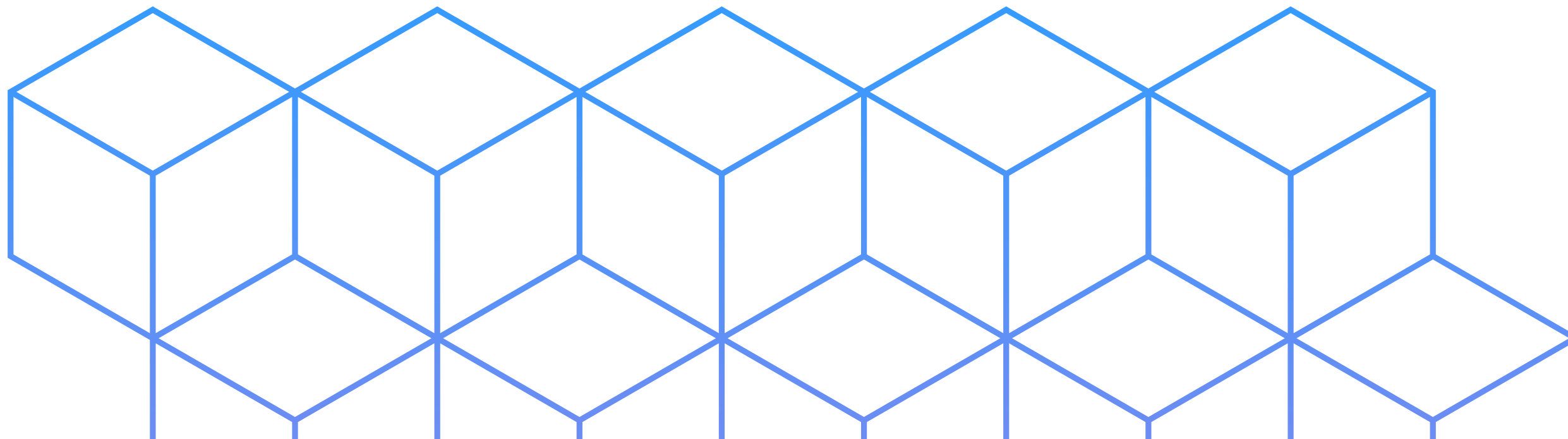
Analysis Approach

Univariate Analysis results

Bivariate Analysis results

Recommendations

Agenda



Problem Statement

01

Avoid credit loss by preventing company to lend to risky customers who could default on loans

02

Identify risky loan applicants using EDA methodology on loan dataset provided

03

Identify the driver variables that leads to loan defaults, this will help company to build its risk assesment portfolio

Analysis Approach

Data cleaning	Fixing rows and columns	Univariate Analysis	Bivariate Analysis
<ul style="list-style-type: none">• Check for duplicate rows and columns and remove them.• Identified that there are no duplicate rows but null columns.• Remove the clear null columns from the dataset. We have use percentage to make it more easier to understand.	<ul style="list-style-type: none">• Publish a heatmap of columns that have null values shown as yellow in visualization.• Handling the missing values until there are no columns left with null values and we get clean blue heat map.• Removing all columns related to customer demographics and post loan approval columns as they provide less value.	<ul style="list-style-type: none">• Removing data with loan status as Current as we can analyze the trend from Charged off customers.• We are finally left with 24 columns and all the data with Charged off loan. We conduct univariate analysis using below column which can point on trend for default.• Univariate Analysis of numerical columns• Univarait analysis of Ordered and Unordered Categorical variables	<ul style="list-style-type: none">• Derived variables using two or more columns like Month and year and plotted visual graph to analyze the trend for the same.<ul style="list-style-type: none">◦ Printing correlation matrix between all numerical columns to analyze the trend of defaults.

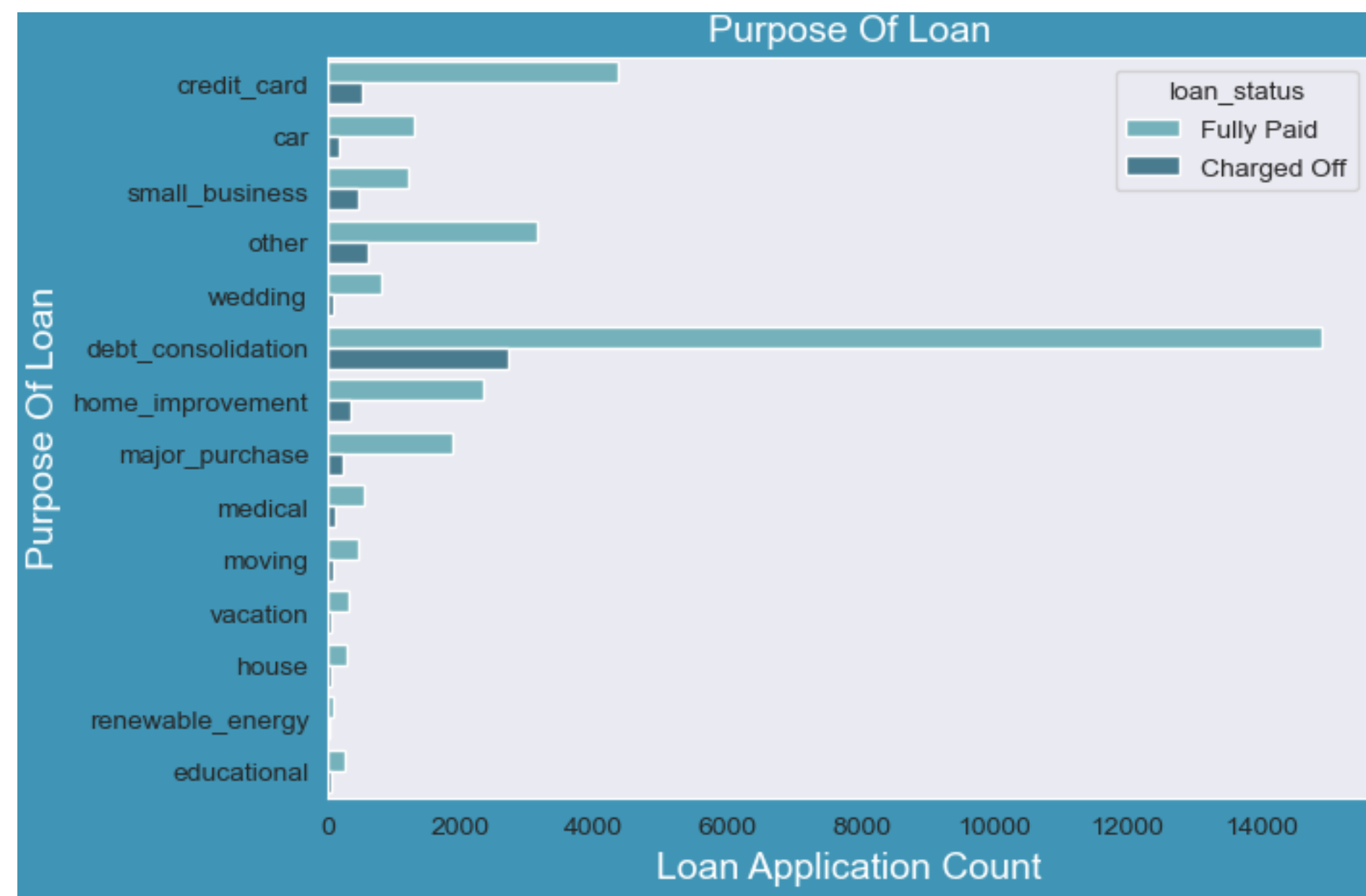
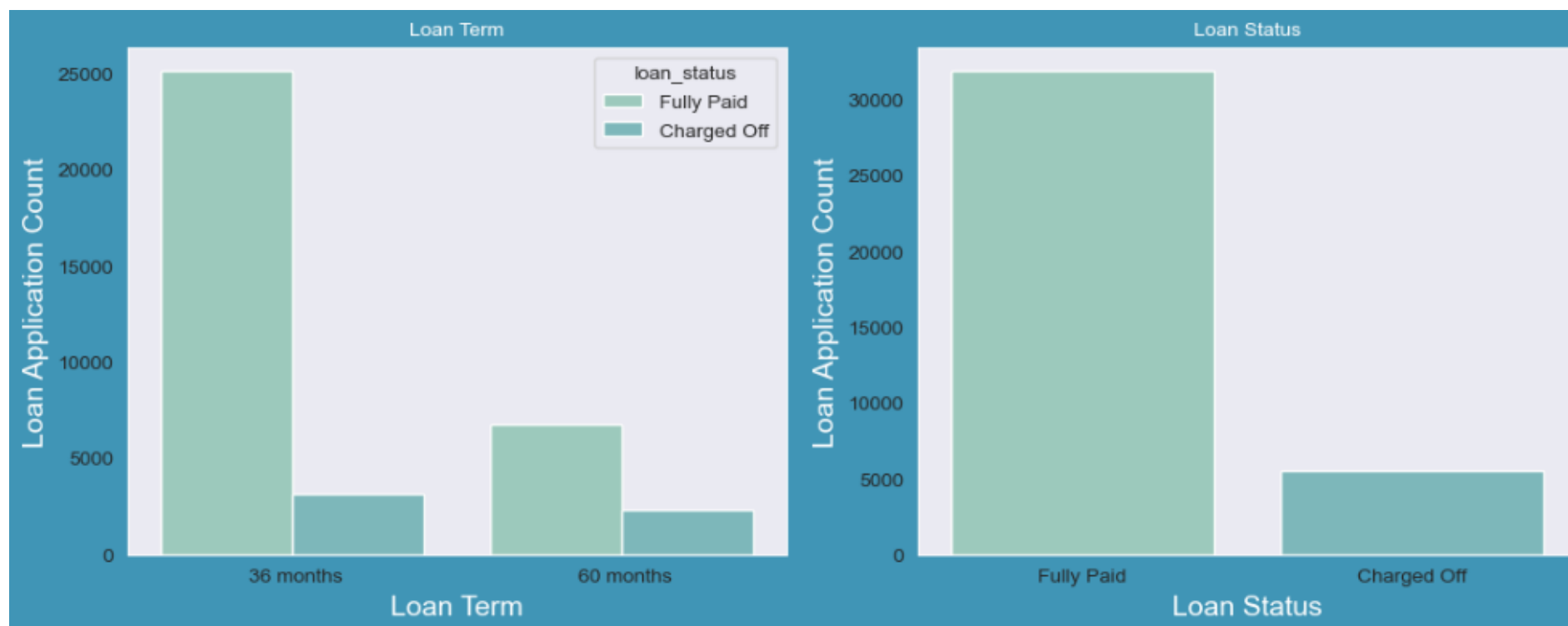
Univariate Analysis

Observation summary of from univariate analysis.

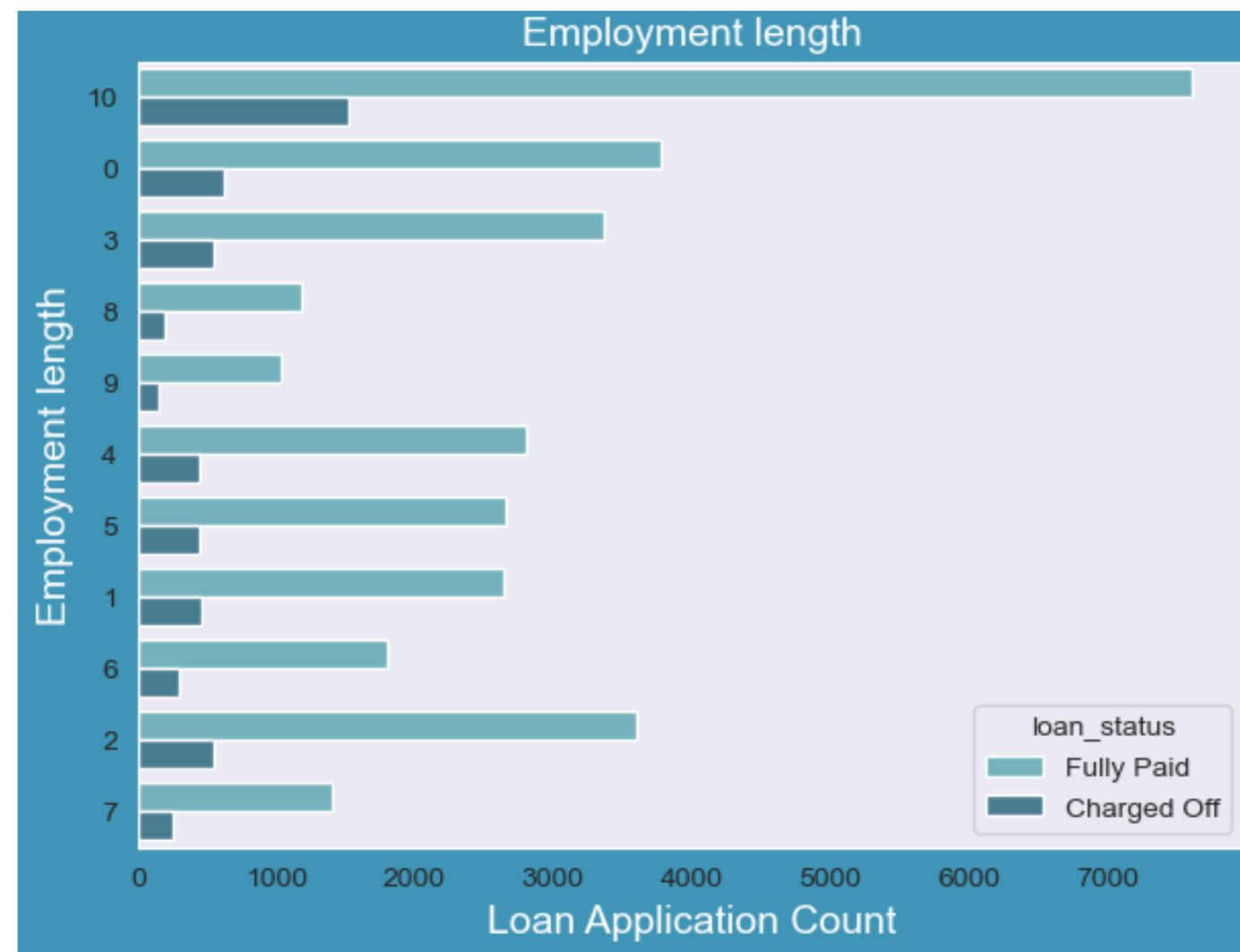
There is a more probability of loan being Charged Off when :

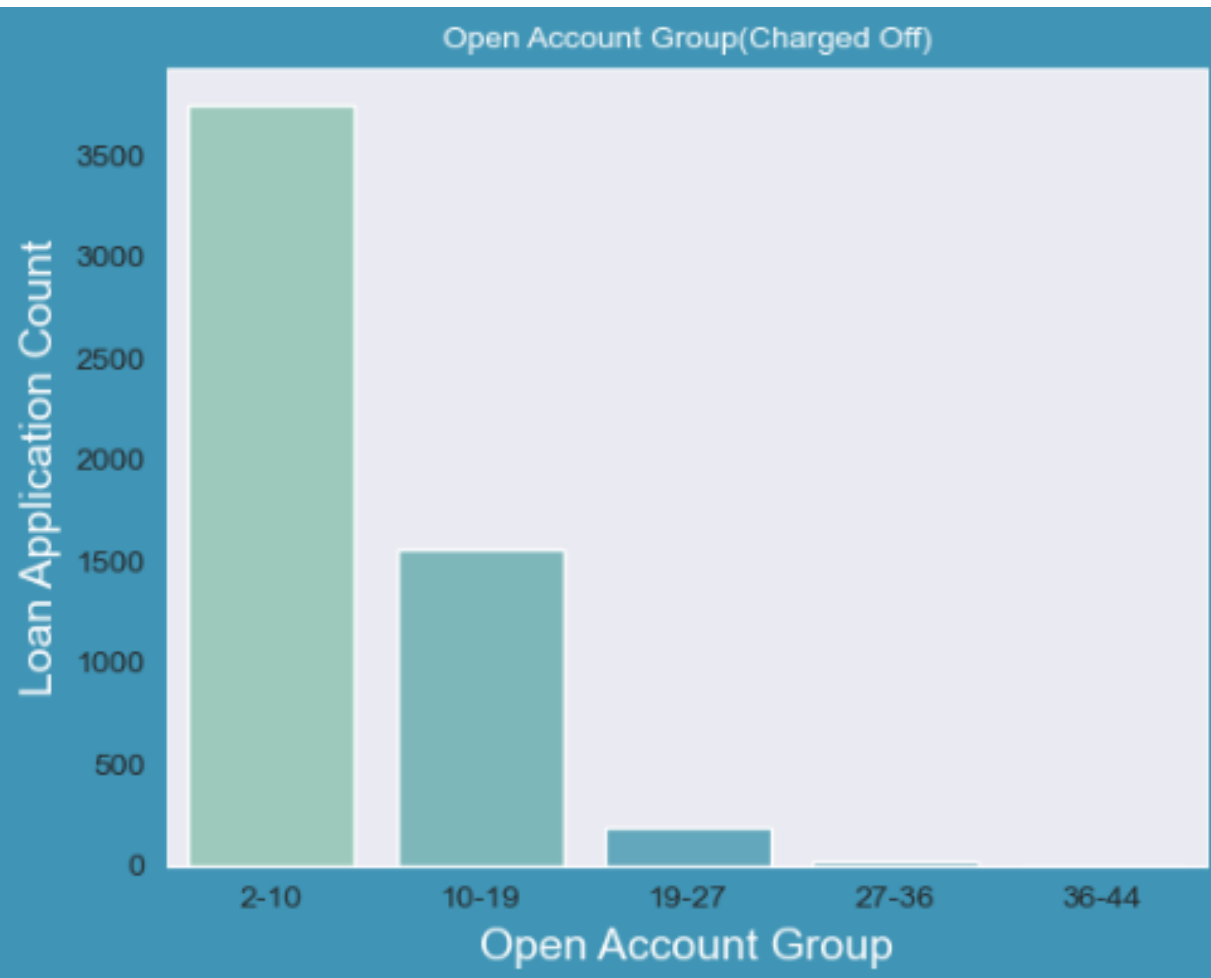
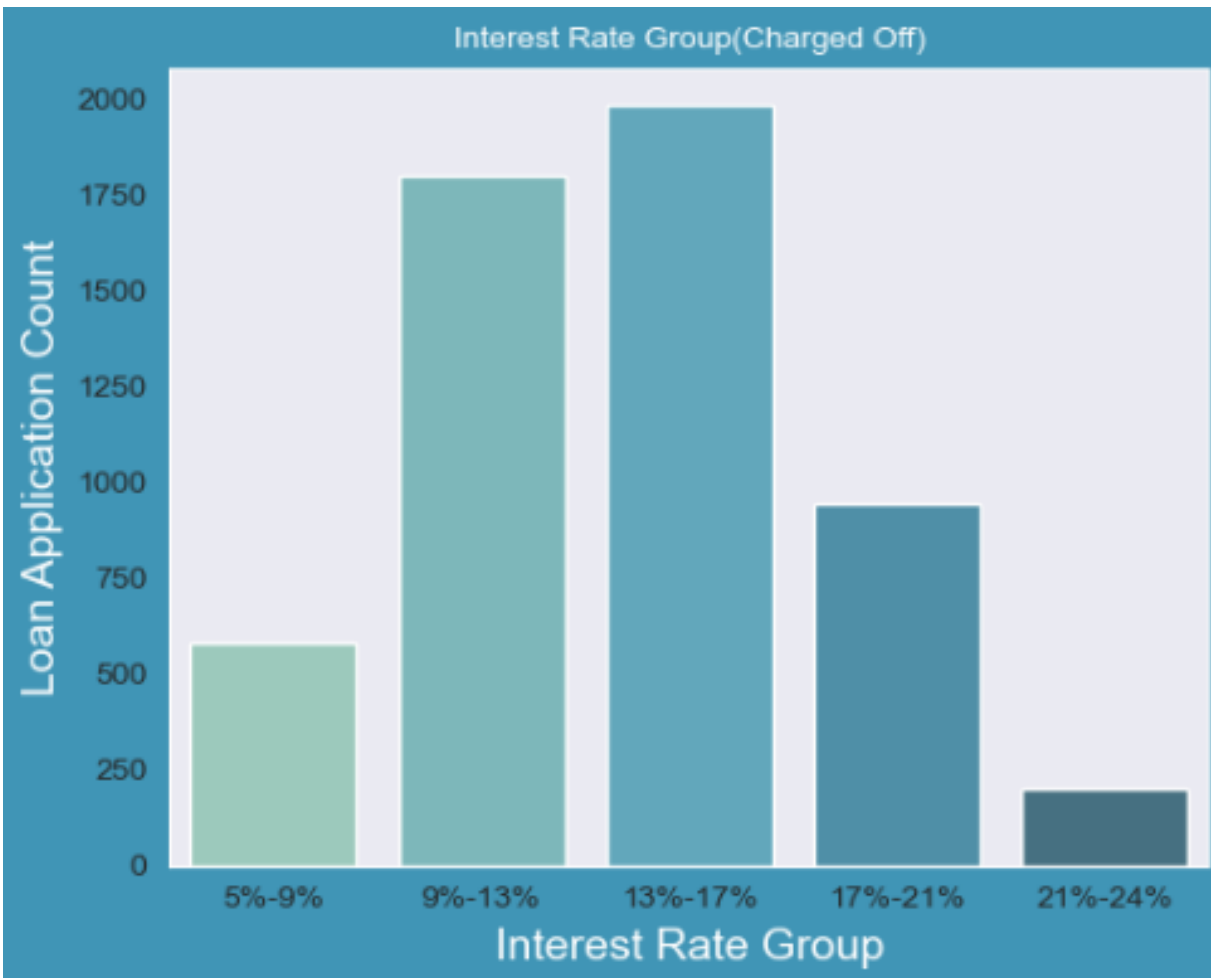
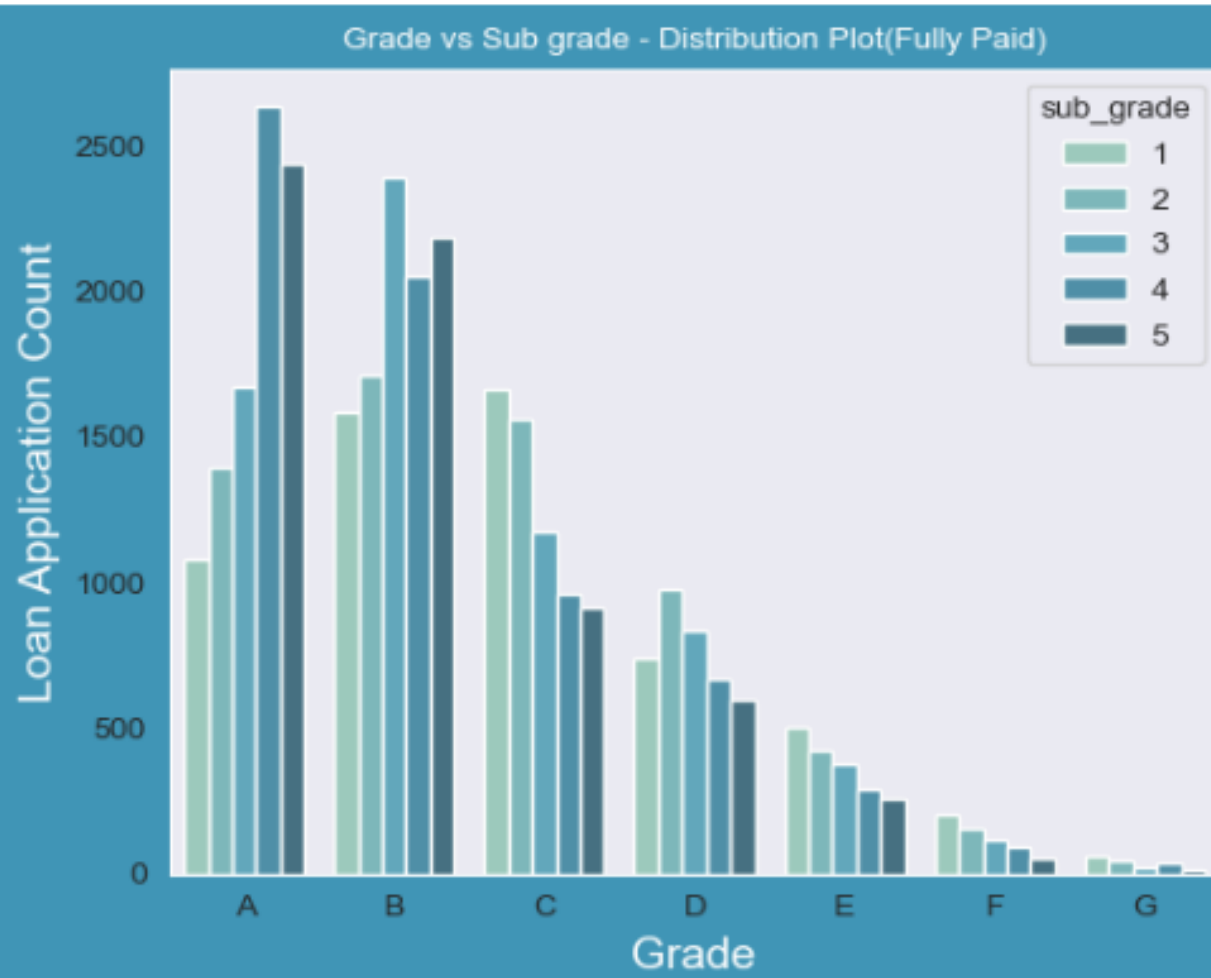
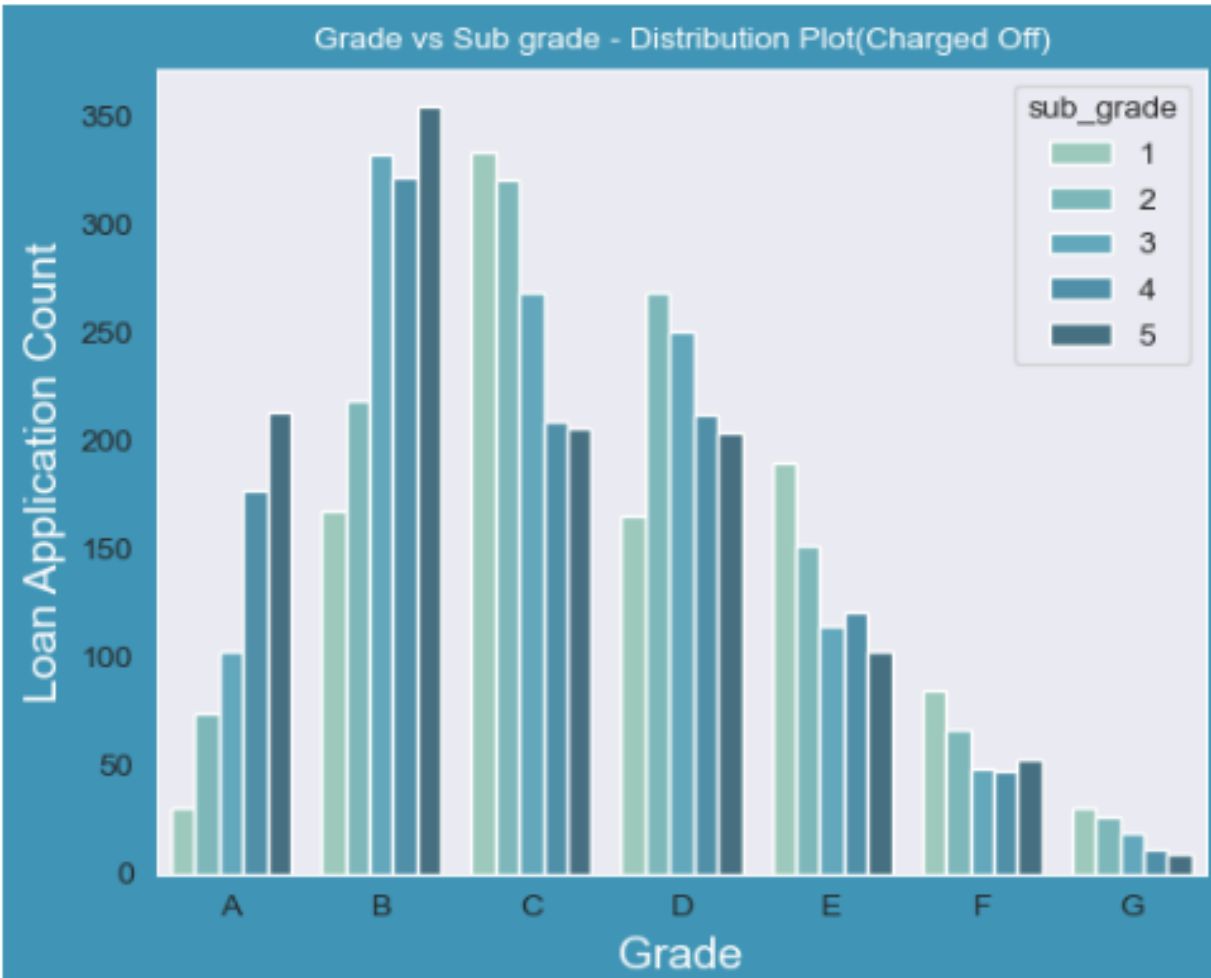
- Consumer taking loan for term of 36 months
- Consumer whose purpose of the loan is for 'Debt consolidation'
- Consumer whose home ownership is 'RENT' or 'MORTGAGES'
- Consumer who is in employment for 10 years and GRADE B
- Consumer who have taken a loan interest rate of 13-17%.
- Consumer who has 2-10 or 2-20 open accounts.
- Consumer who has utilization rates between 60-100
- Consumer who has DTI ratio between 12-18.
- Consumer whose number of derogatory public records is 0(It can be due to not having proper data for all consumer)
- Consumer whose funded Loan amount is between 5000-10000
- Consumer whose monthly installments are between 145-274

Supporting graphs are added in upcoming pages.

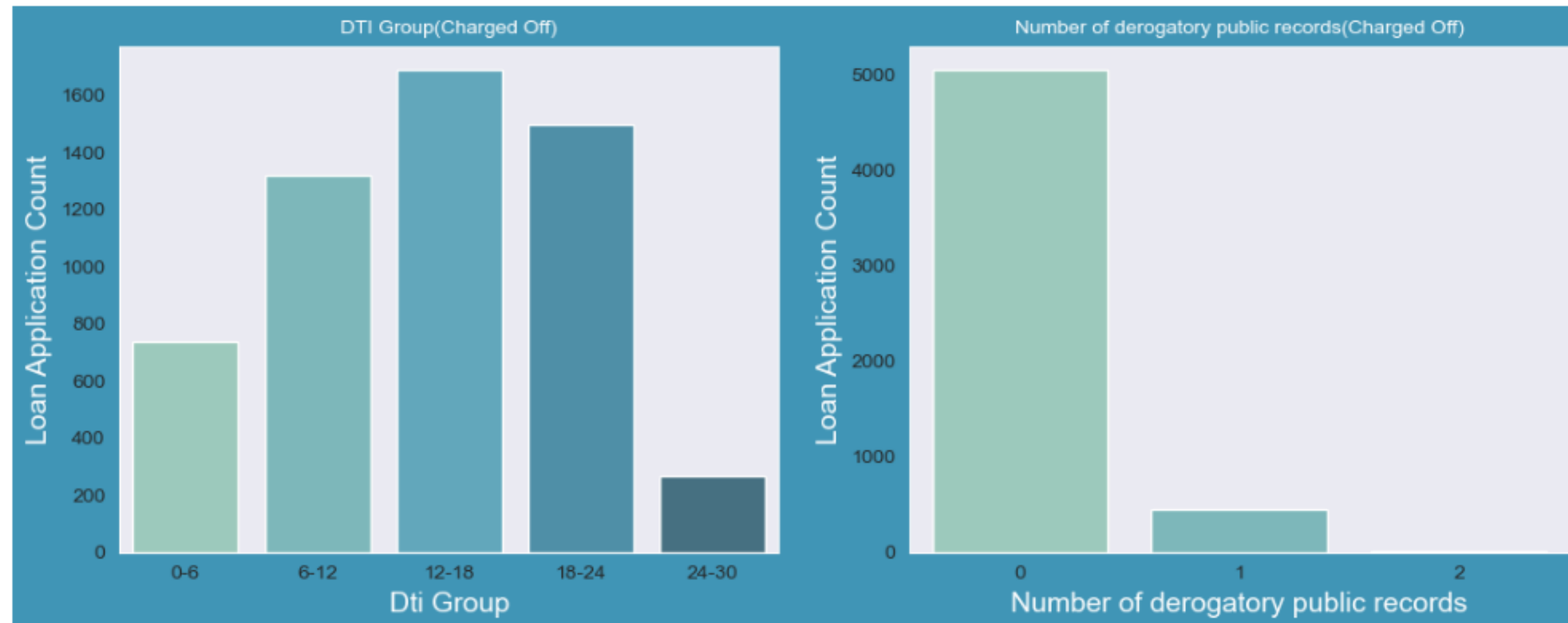
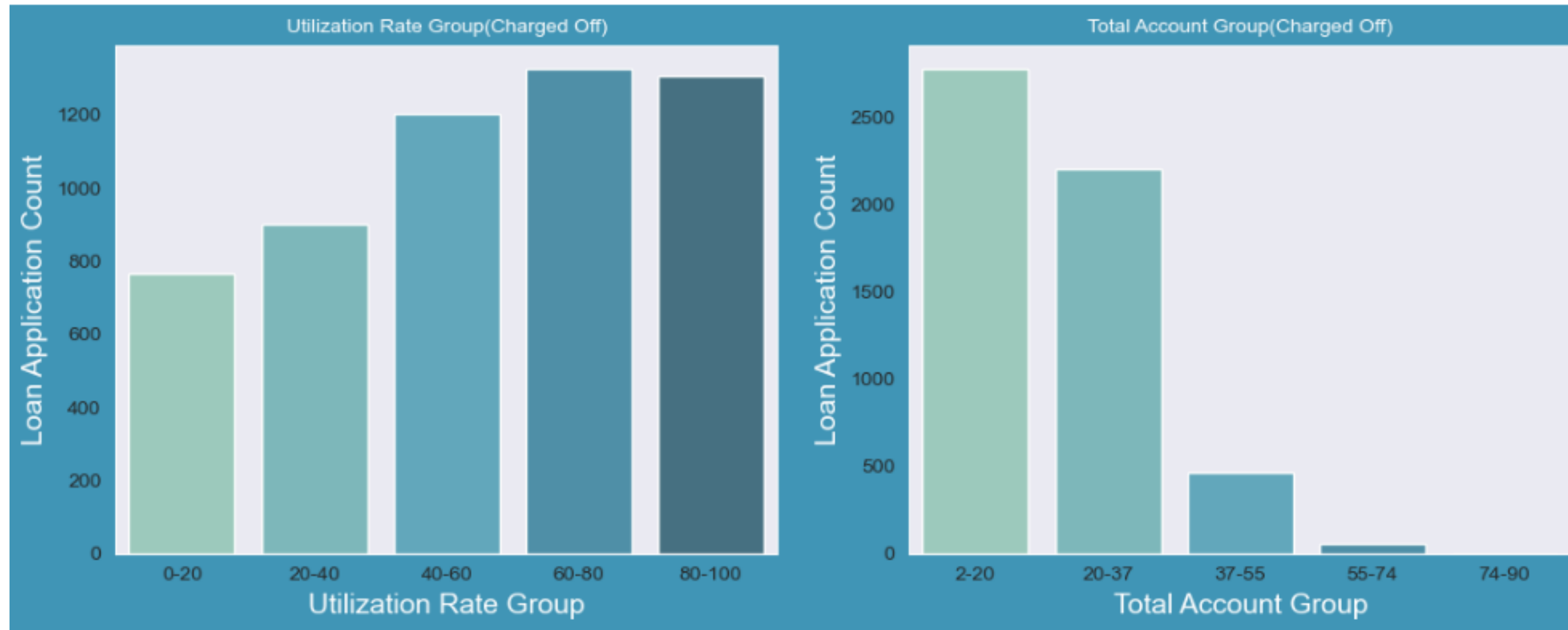


UA Cont.

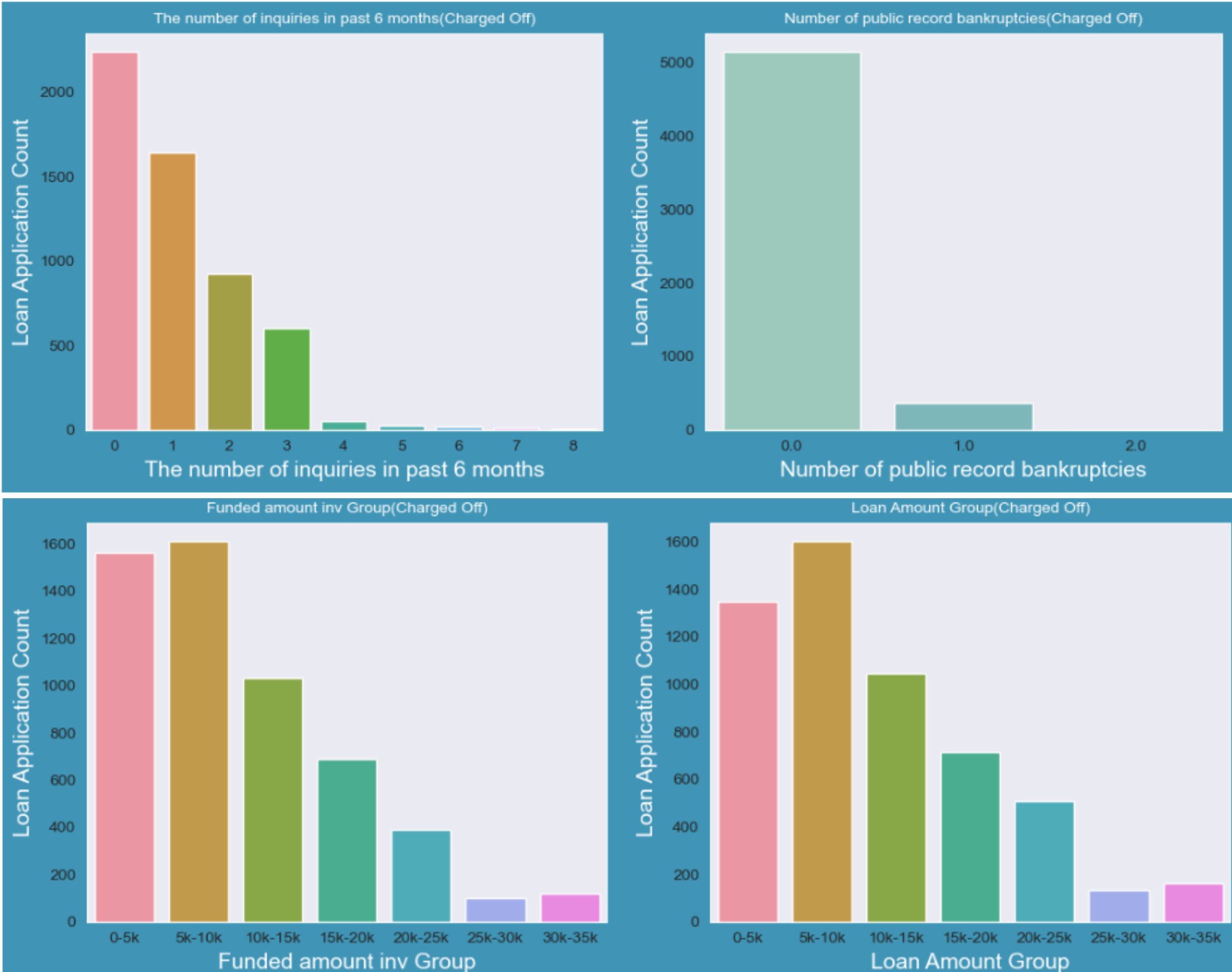




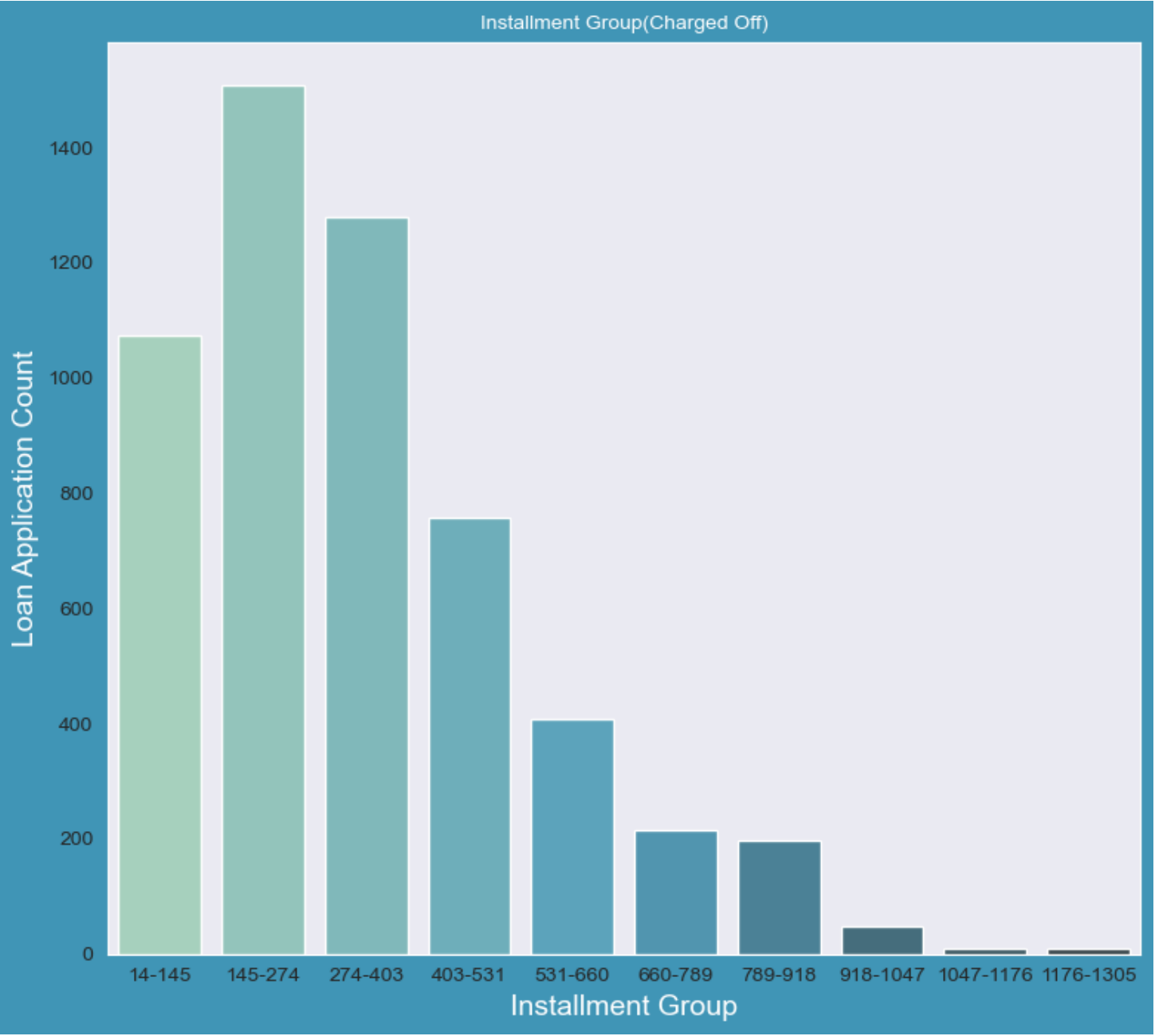
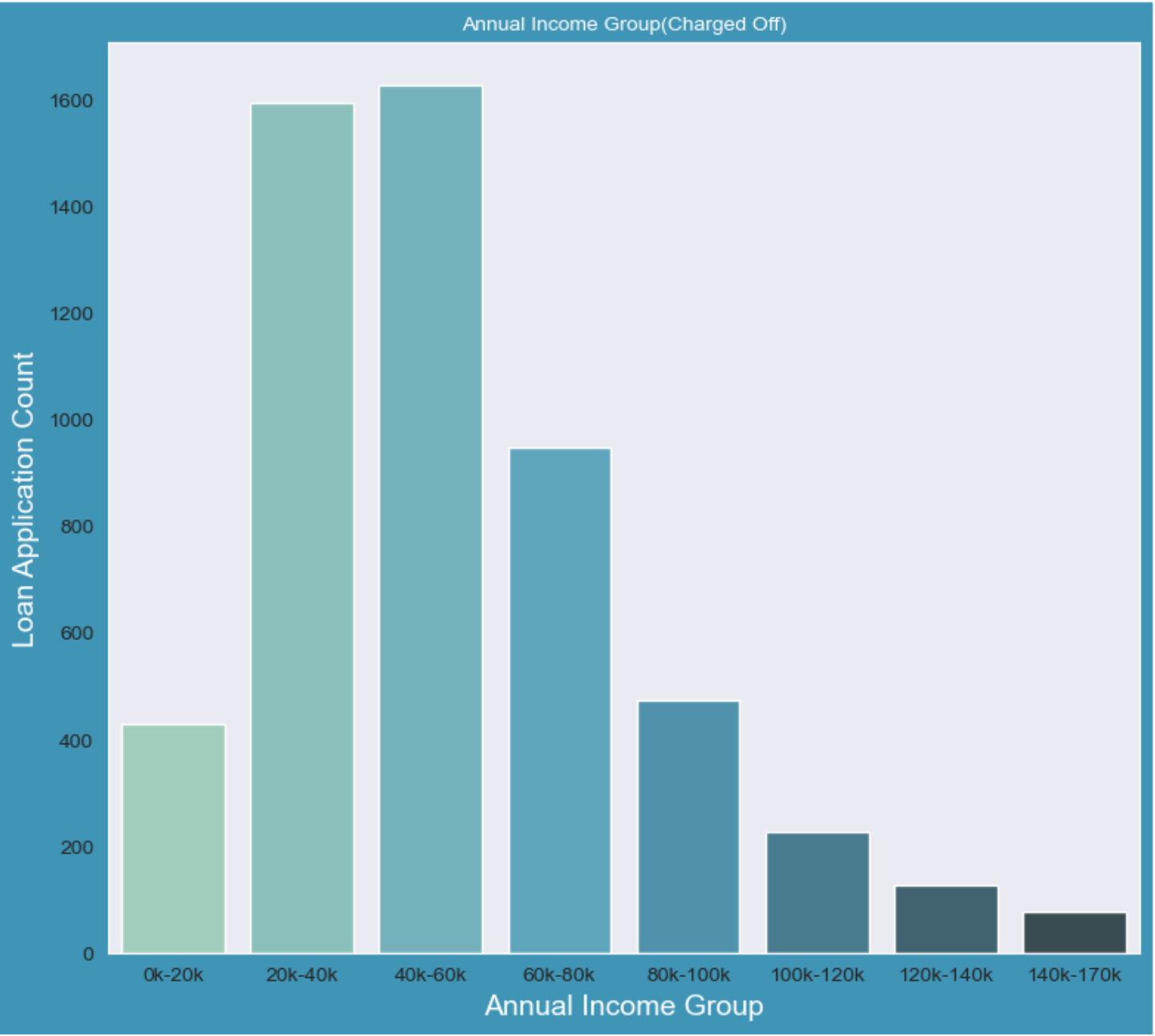
UA Cont.



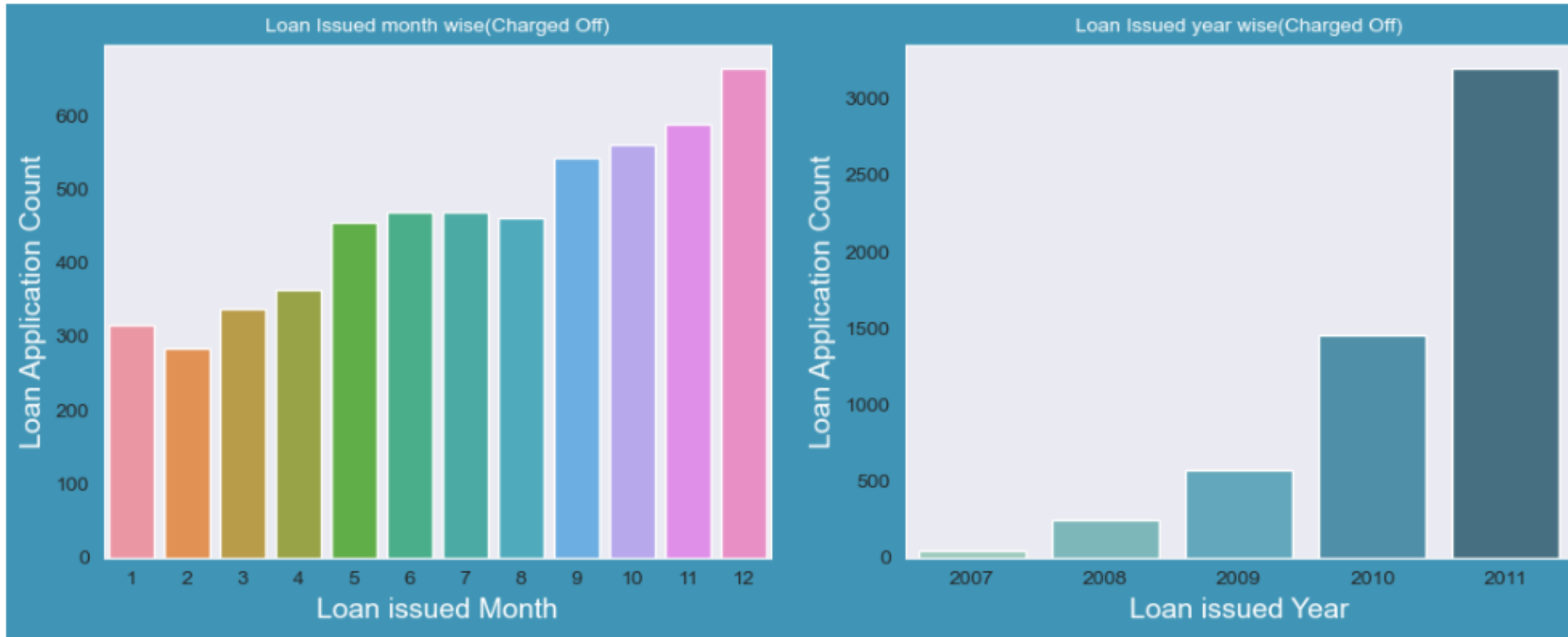
UA Cont.



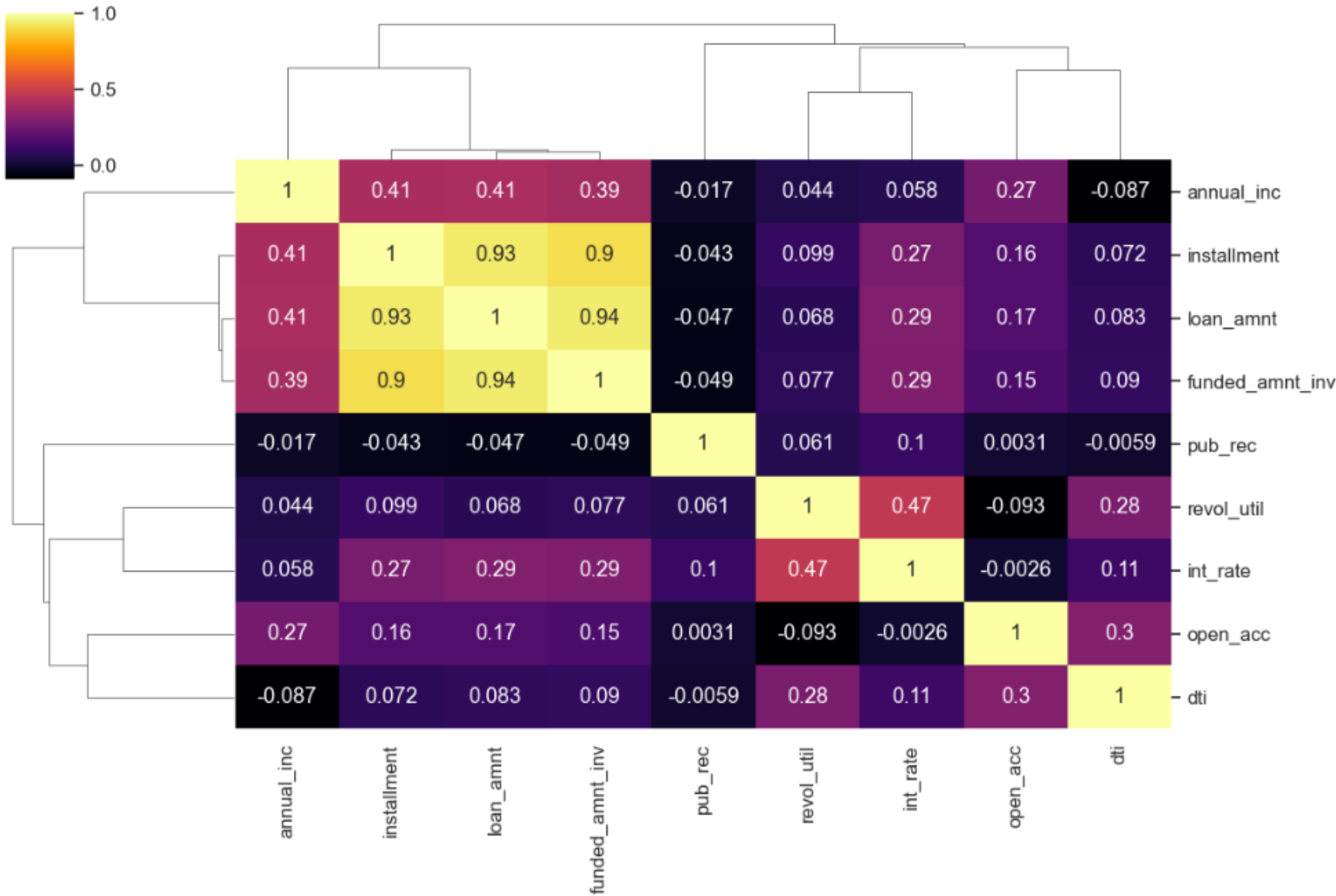
UA Cont.



UA Cont.



UA Cont.



Bivaraiate Analysis

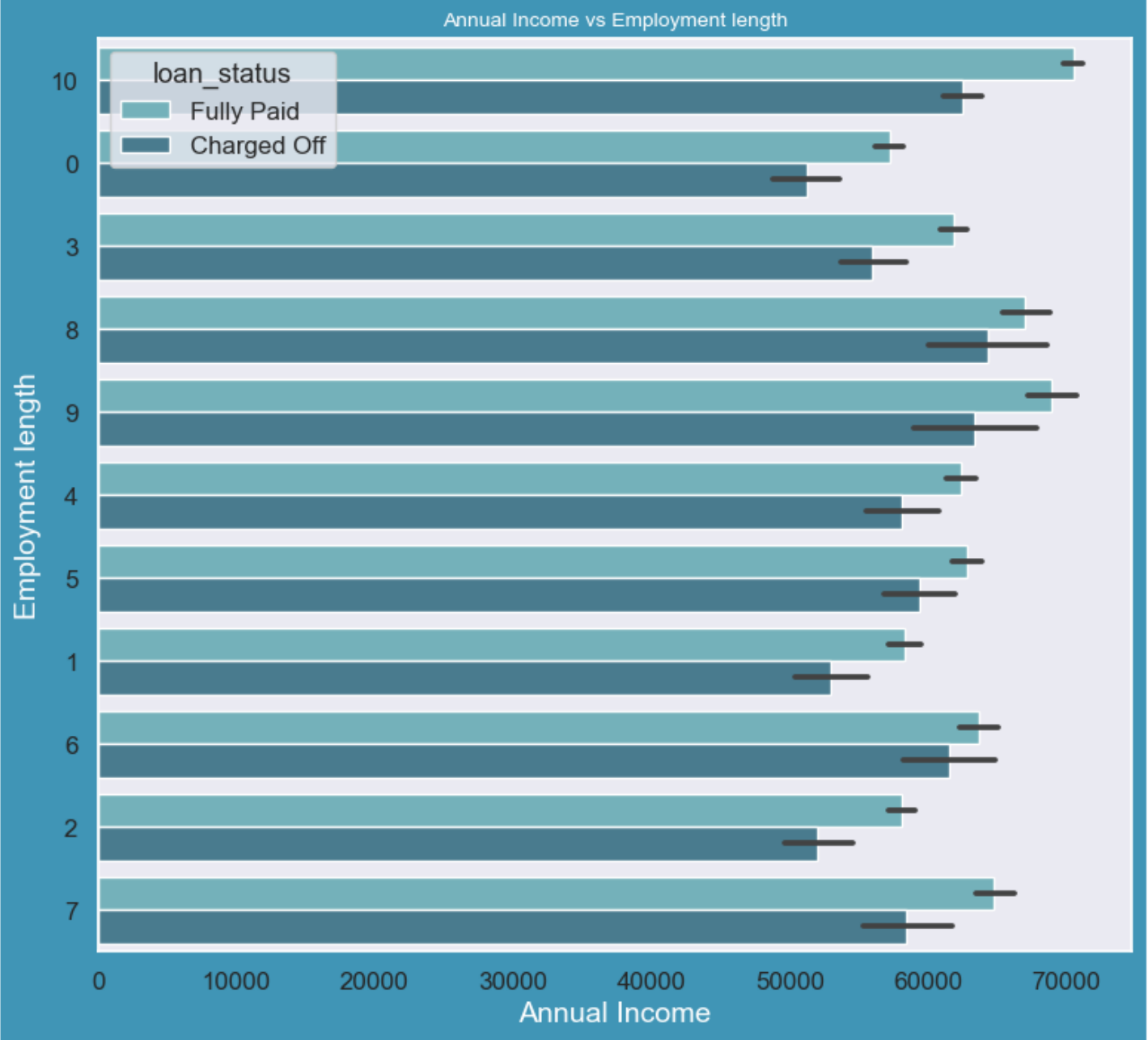
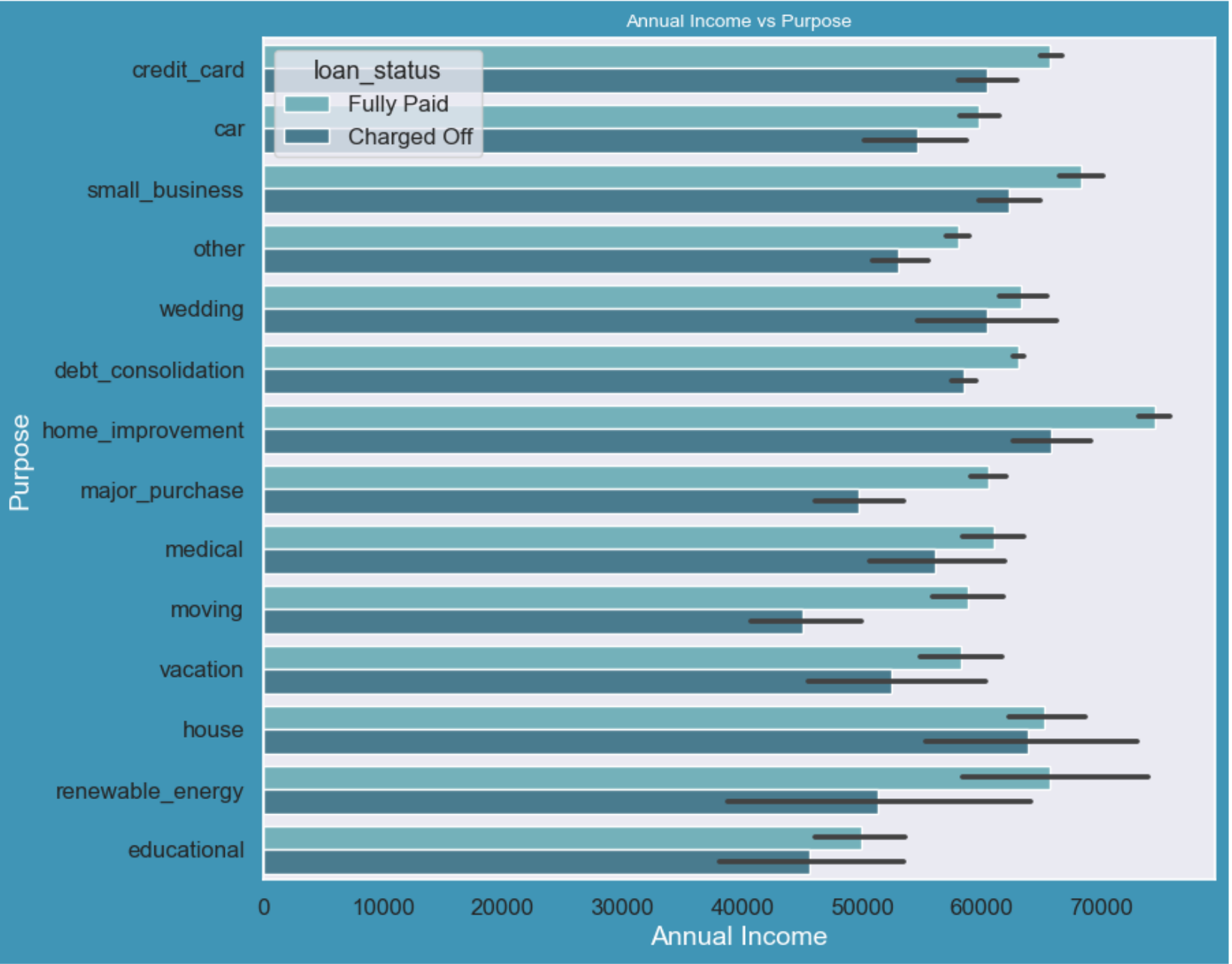
Observation summary of from bivariate analysis.

There is a more probability of loan being Charged Off when :

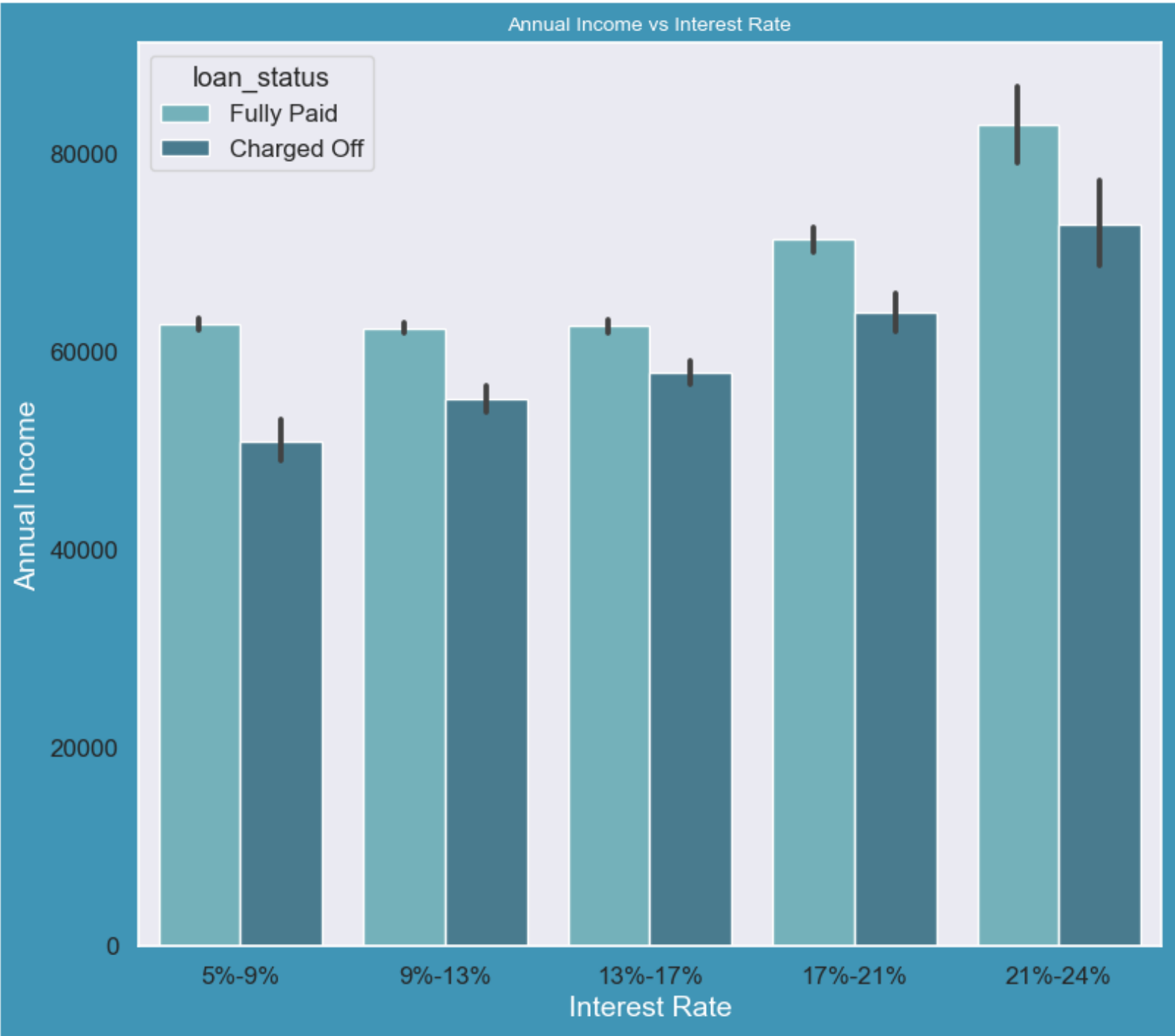
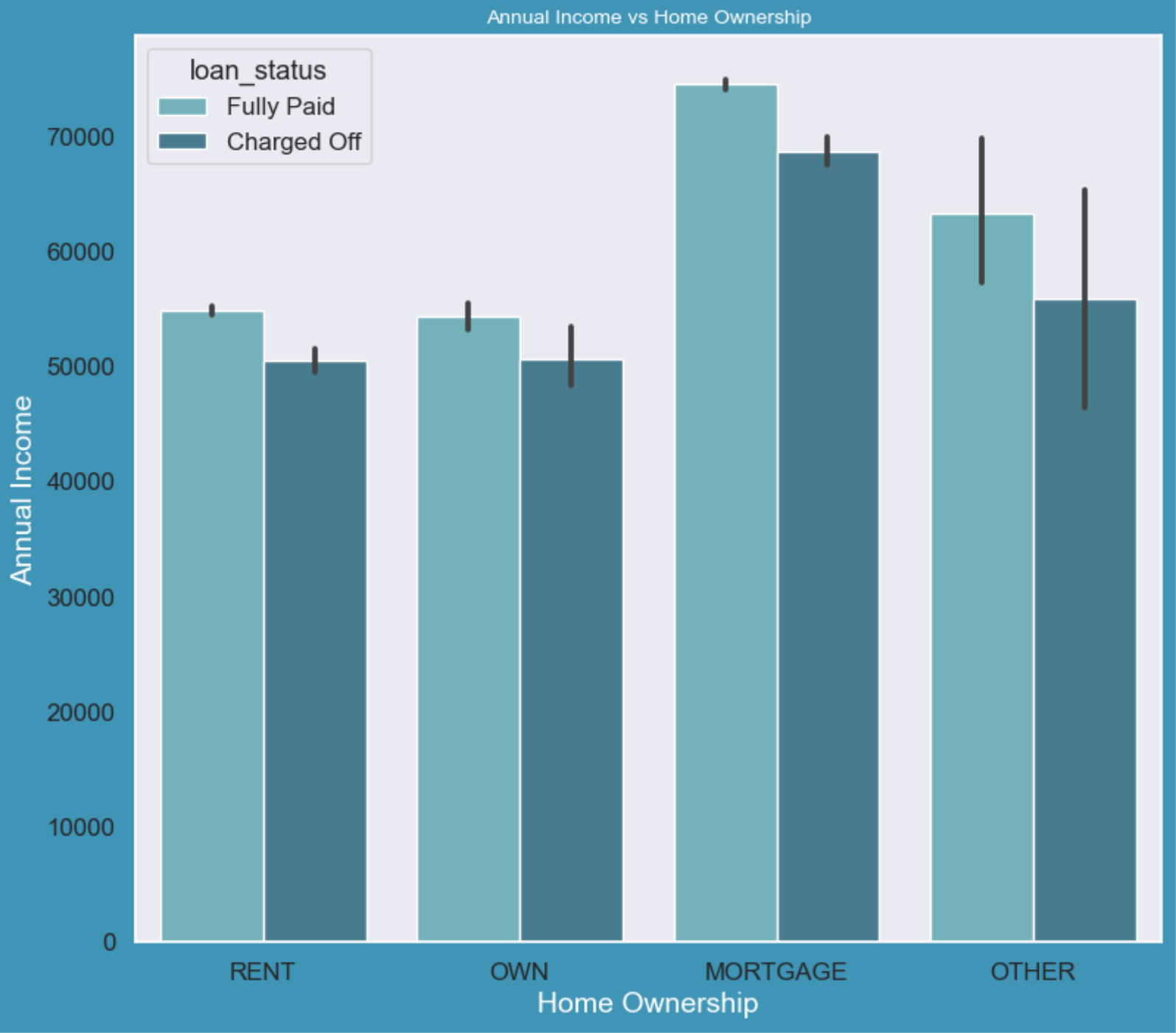
- Consumer taking loan for 'home improvement' and have annual income of 60k -70k
- Consumer employment length of 8 years and have annual income of 60k -70k
- Consumer whose home ownership is 'MORTGAGE and have annual income of 60-70k
- Consumer who receive interest at the rate of 21-24% and have an annual income of 60k-80k
- Consumer who have taken a loan for small business and the loan amount is in between 14k - 16k
- Consumer employment length of 10 years and the loan amount is in between 13k - 14k
- Consumer whose home ownership is 'MORTGAGE and have loan of 13k-15k
- Consumer who have taken a loan in the range 15k - 20K and are charged interest rate of 21%-24%

Supporting graphs are added in upcoming pages.

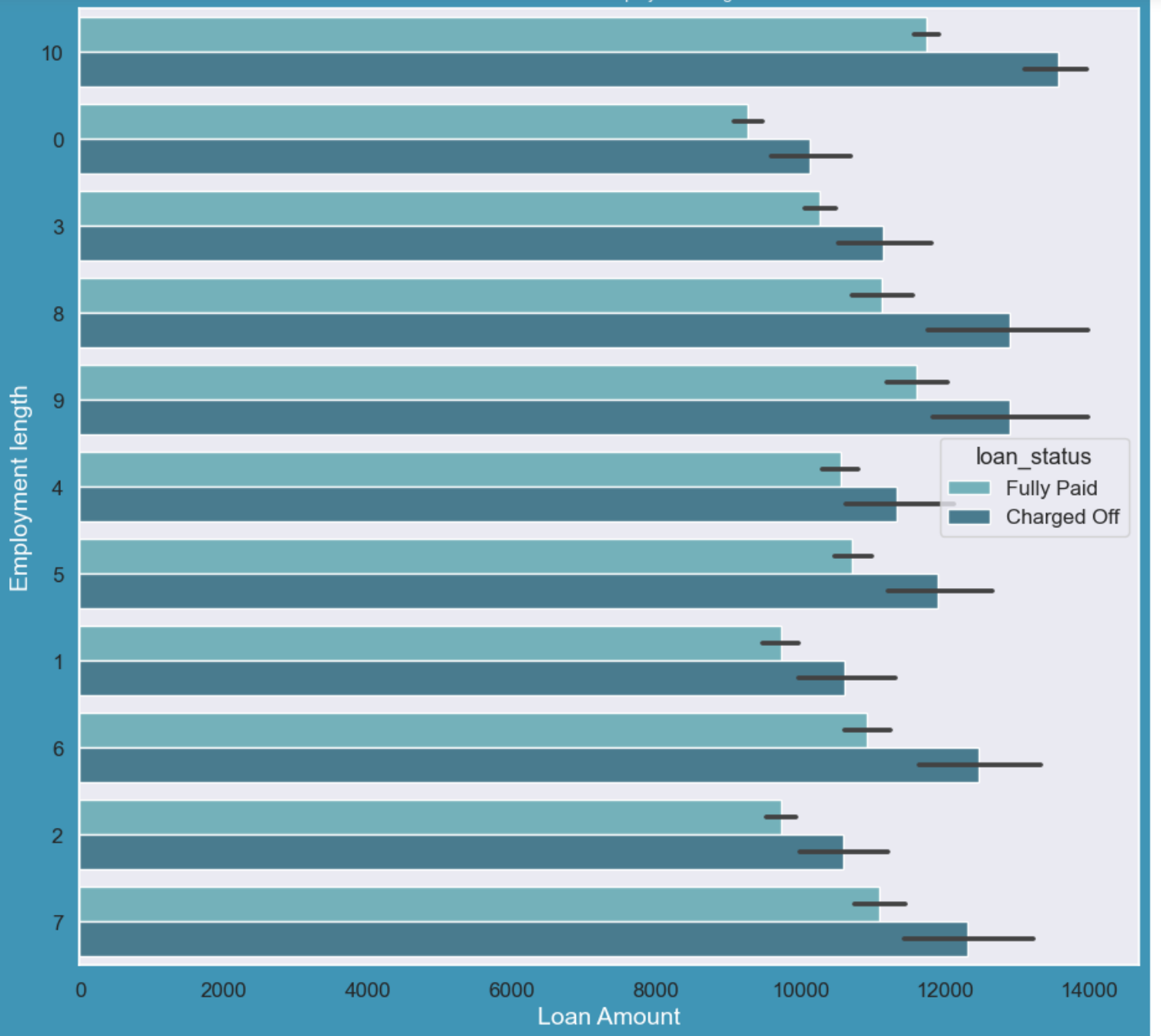
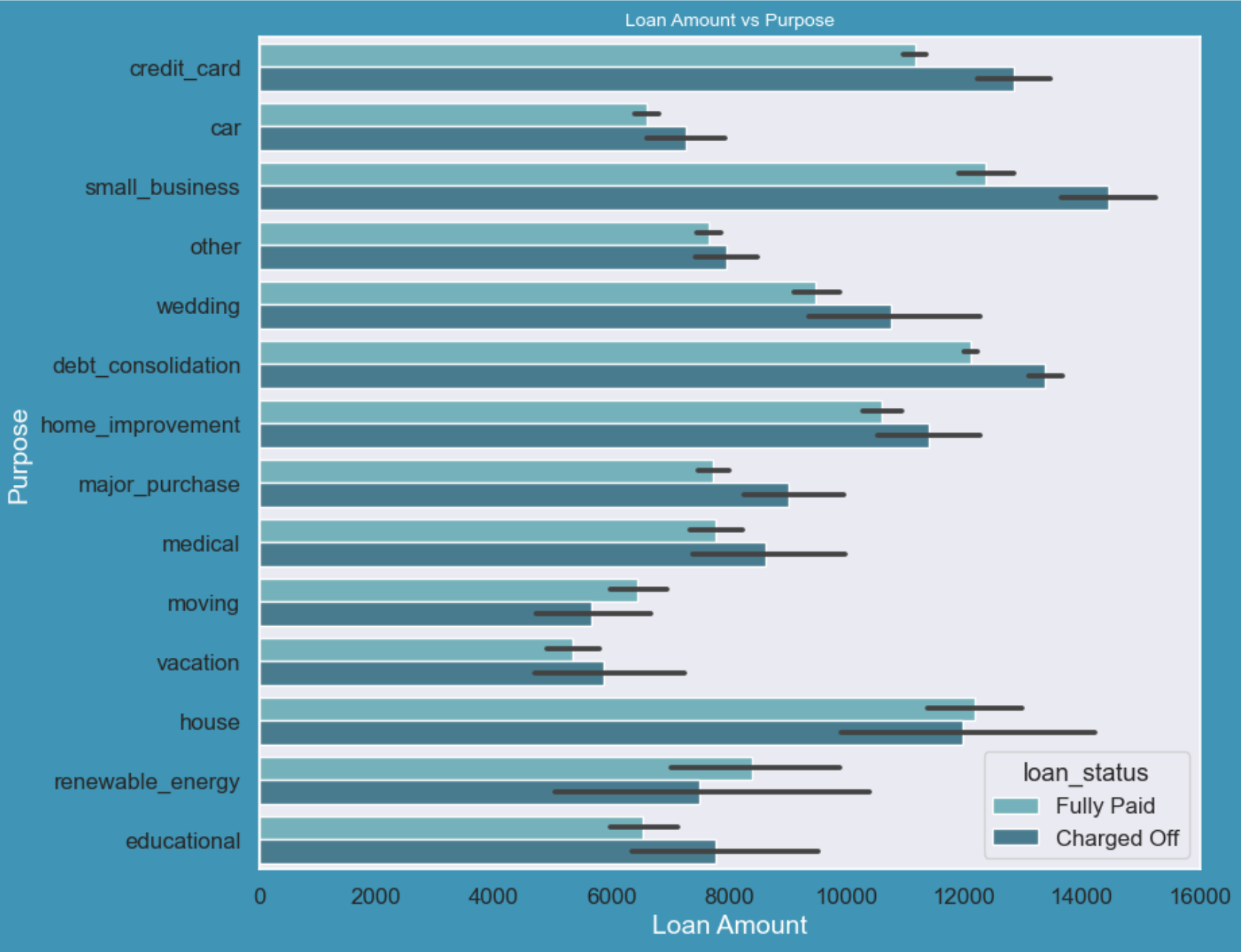
Bivaraiate Analysis



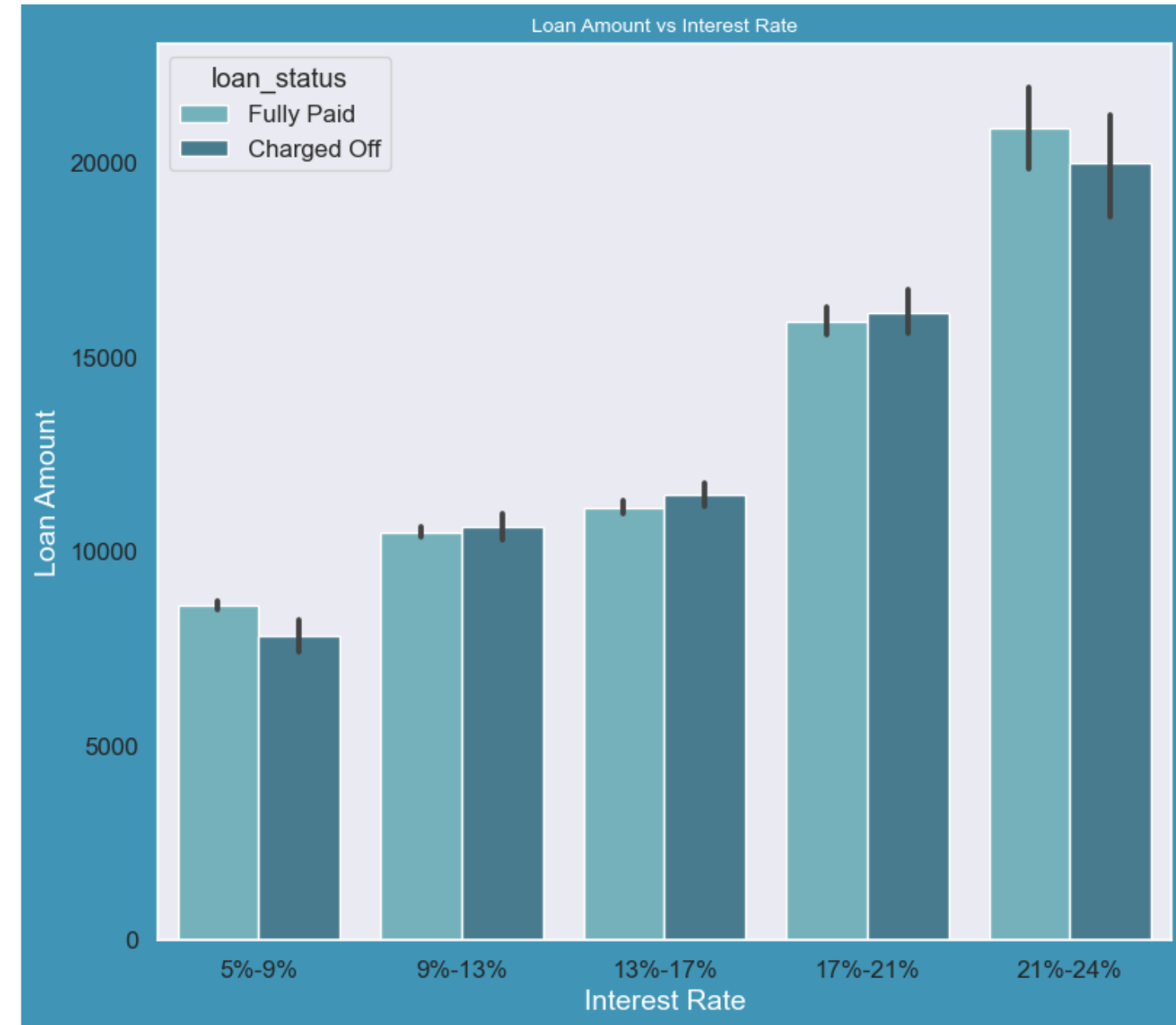
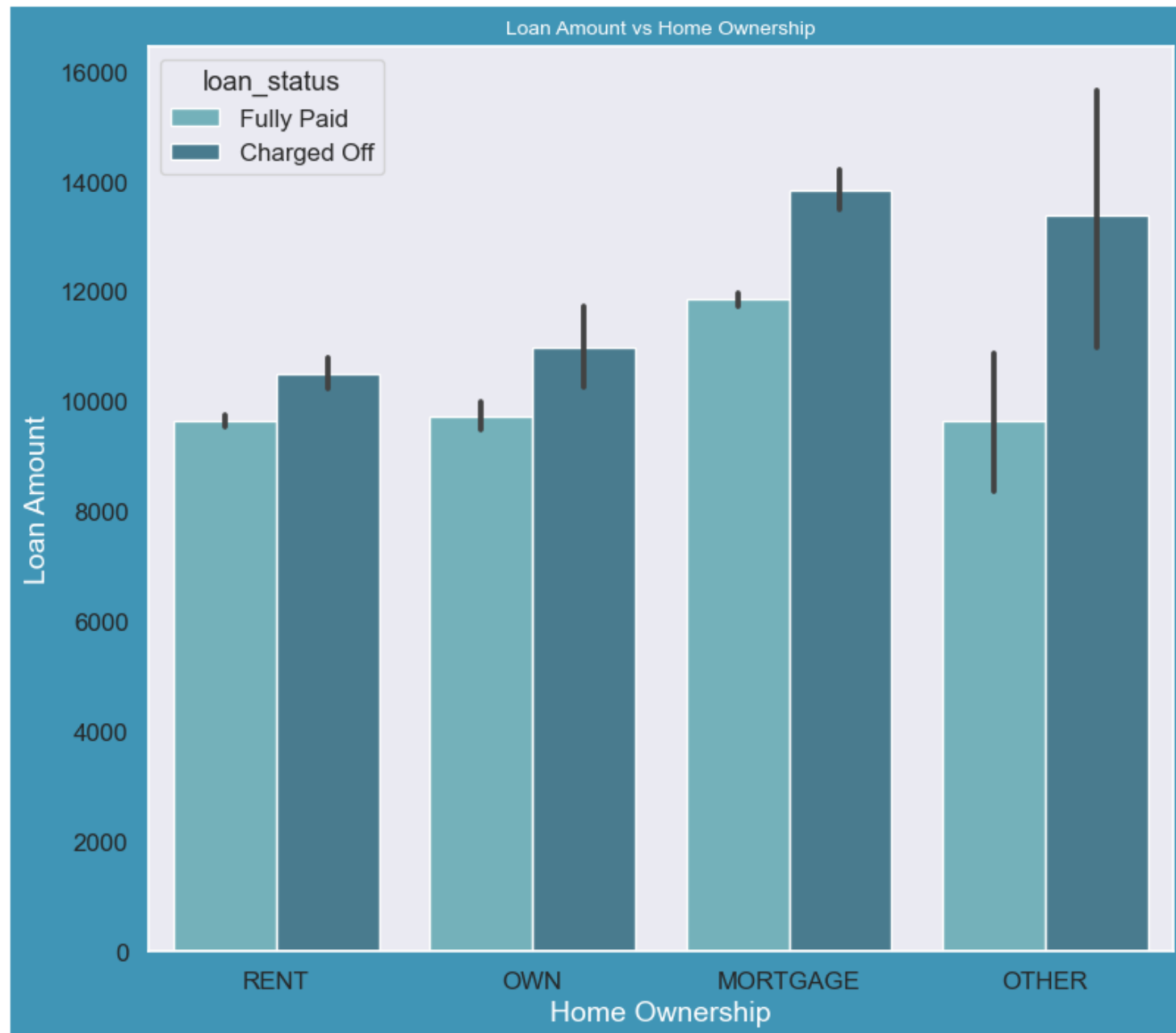
Bivaraiate Analysis



Bivaraiate Analysis



Bivaraiate Analysis



Recommendations for Risk Management

Parameters shared in this document from univariate & bivariate analysis can be potential driving factors for loan defaults and hence company should consider the field value highlighted as red before issuing a loan to customer. Customer who has maximum parameters matching this criteria should be critically checked by underwriters before issuing any loan with sufficient securities. Although this decision can vary from case to case basis.

Thank You