Lending Club Case Study

By: Nishadh Shrestha

About this case study

- Lending Club is a financial services company that offer peer-to-peer lending services. When the company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile.
- In this case study we analyse the data that contains historical information about loan applicants from lending club and look closely at whether they defaulted or not.
- The aim is to study features in the data set identify patterns which indicate if a person is likely to default. These finding may then be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.

In this case study, we will Exploratory Data Analysis to explore and acquire insights out of the provided loan data.

Exploratory Data Analysis Process

There are four major parts of this study:

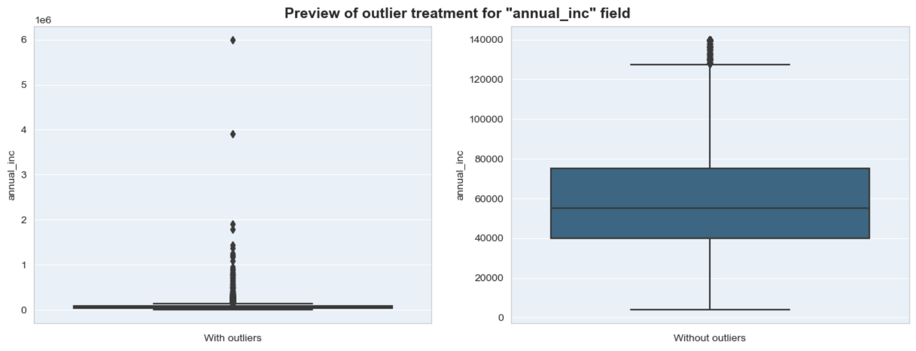
- Data Understanding
- Data Cleaning
- Data Analysis
- Recommendations

Data Understanding

- There are **39717** rows and **111** columns in this dataset
- This seems like high-dimensional dataset and might have one or more features that aren't important for this analysis
- First impression is that this dataset might by somewhat noisy (like **null values**) and will need cleaning up

Data Cleaning

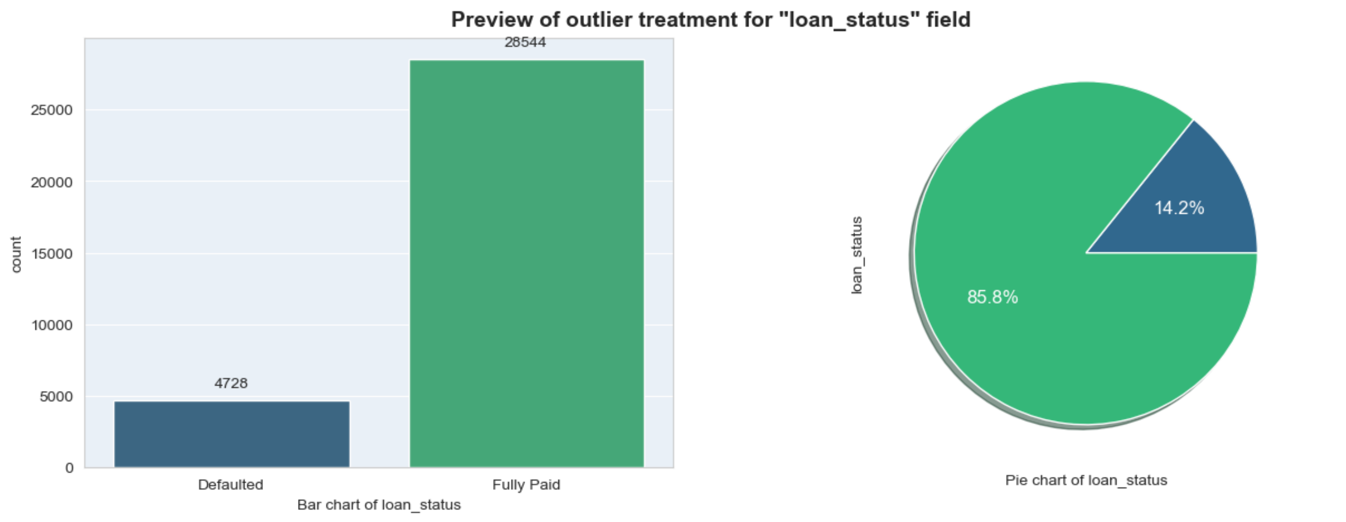
- These 55 columns seems to contain only null values and they were dropped
- Some other columns like loan behaviour data, unique id and fixed value columns were also removed
- Data types of each columns were corrected
- Outliers for numerical columns were removed by using the IQR method



Data Analysis - Univariate

- Univariate Analysis

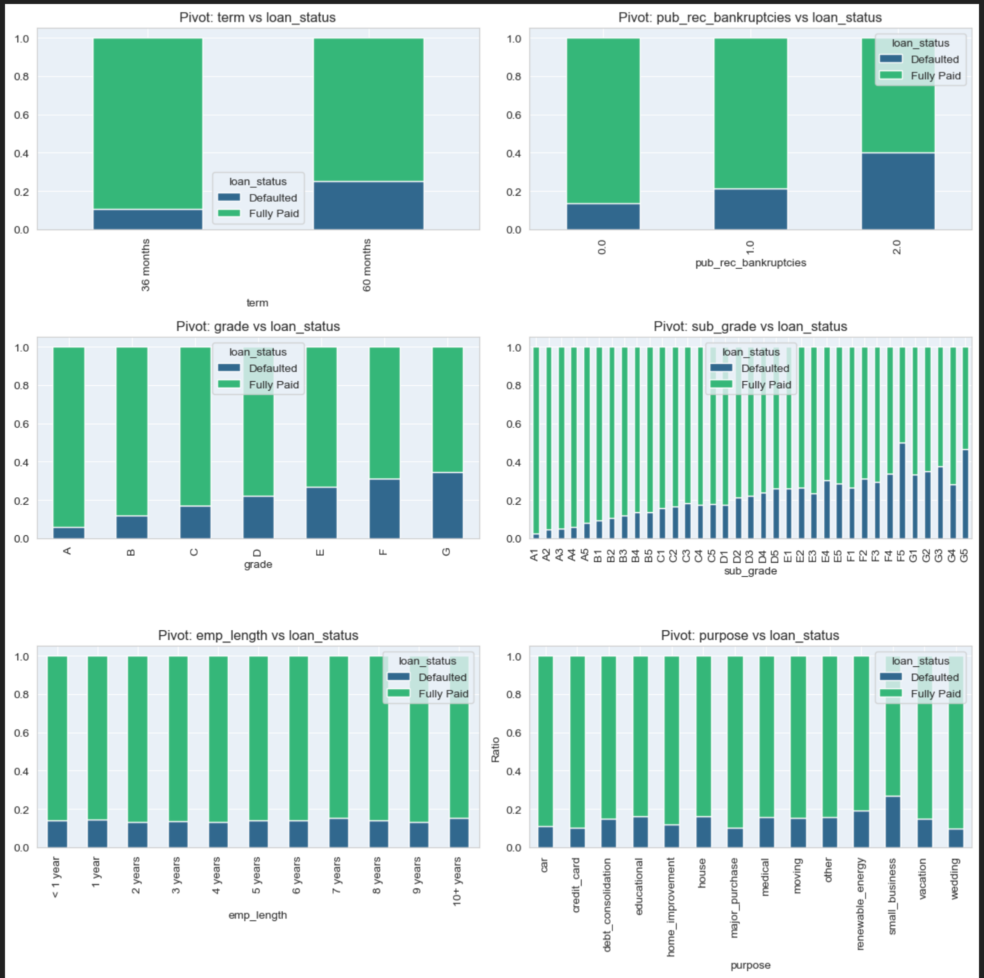
Univariate Analysis was performed on several single fields to gain insights about individual variables in the data



•Significantly more loans are *fully paid* when compared to *defaulted* loans

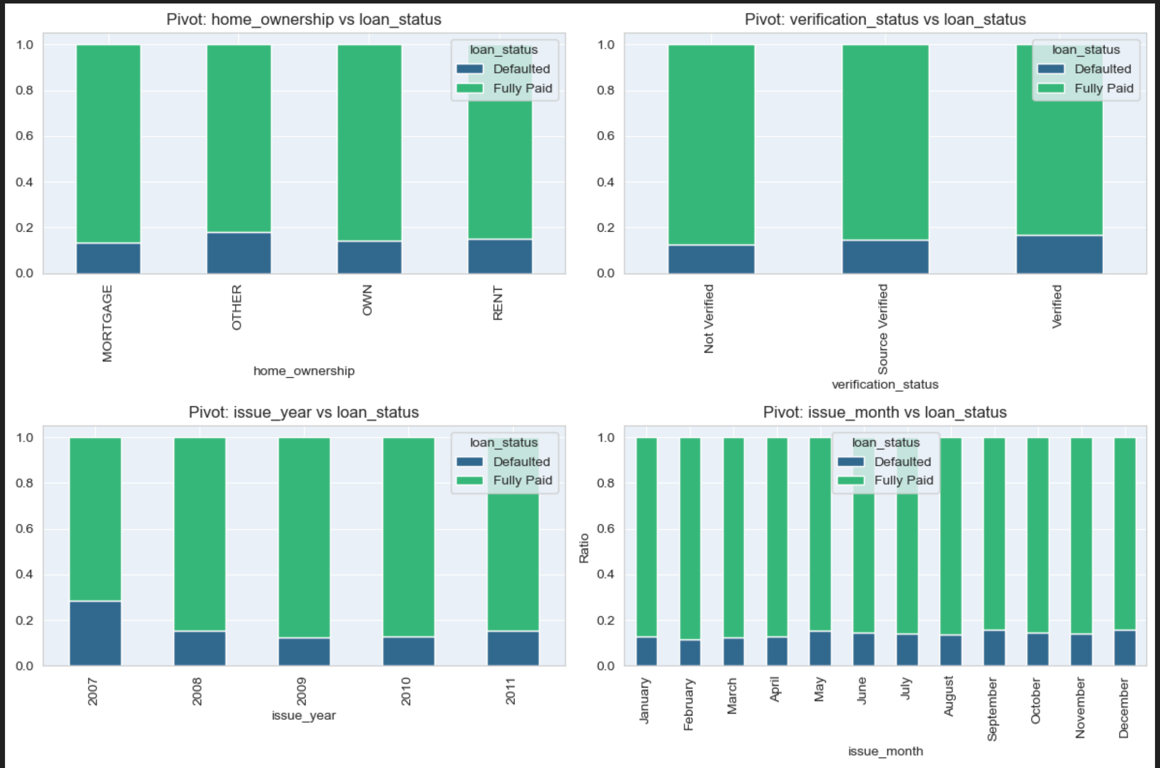
Observations from Univariate Analysis

- Defaults rate is higher for 60 month term loan
- Defaults rate is higher for borrowers with public bankruptcy records
- Defaults rate is higher for lower grade loans
- Default rate is slightly higher for small business loans



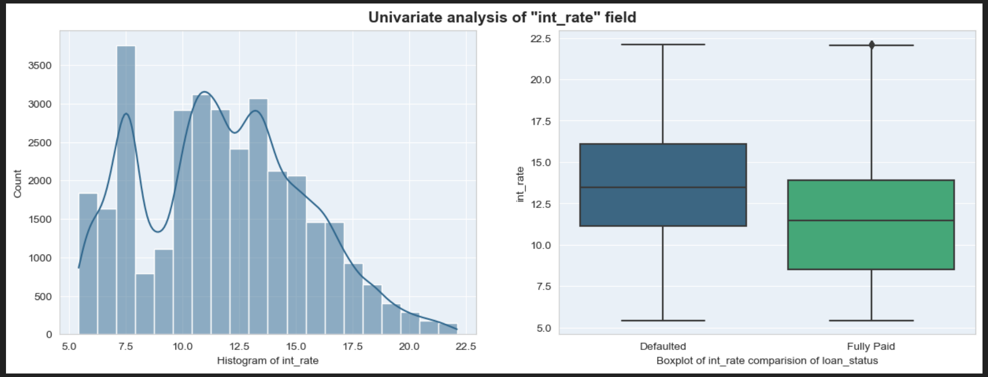
Univariate Analysis – Part 1

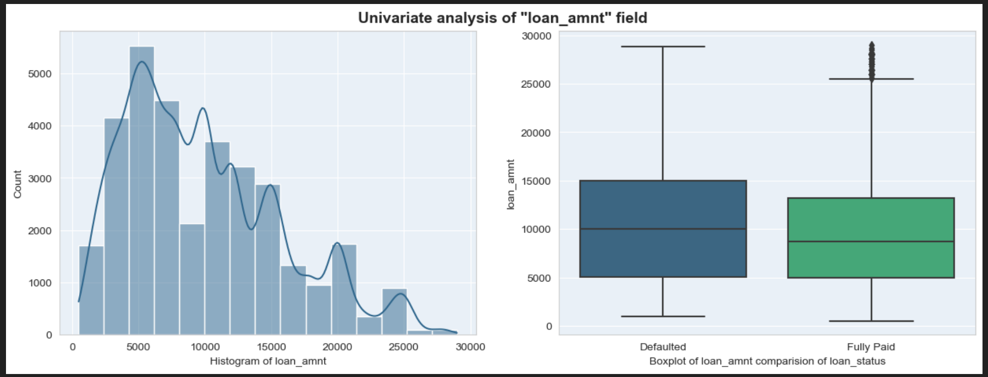
- Default rate is higher in year 2007



Univariate Analysis – Part 2

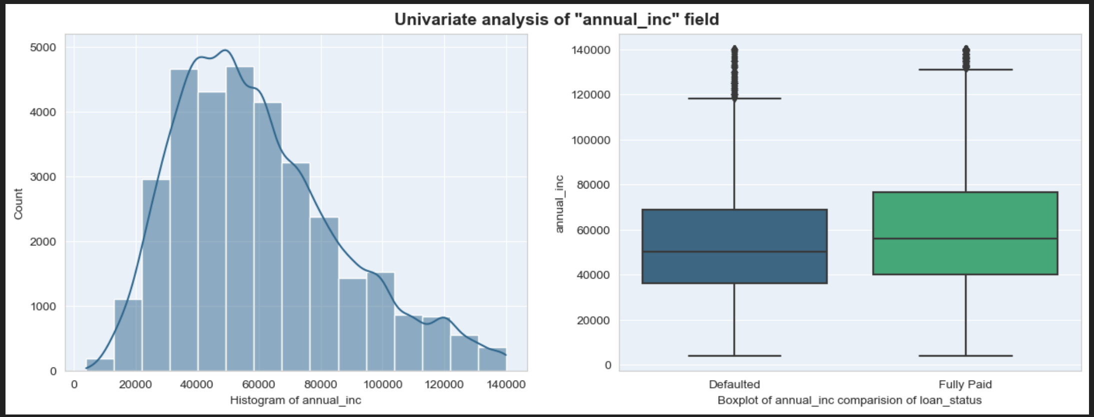
- Loan defaults rate is higher with higher interest rates
- Loan defaults rate is higher with higher loan amount

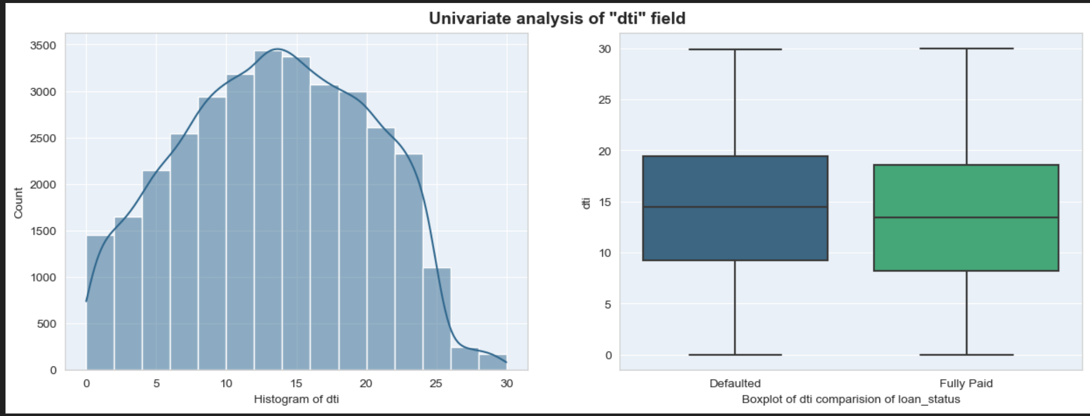




Univariate Analysis – Part 3

- Loan defaults rate is lower with higher annual income
- Loan defaults rate is higher with higher debt to income ratio



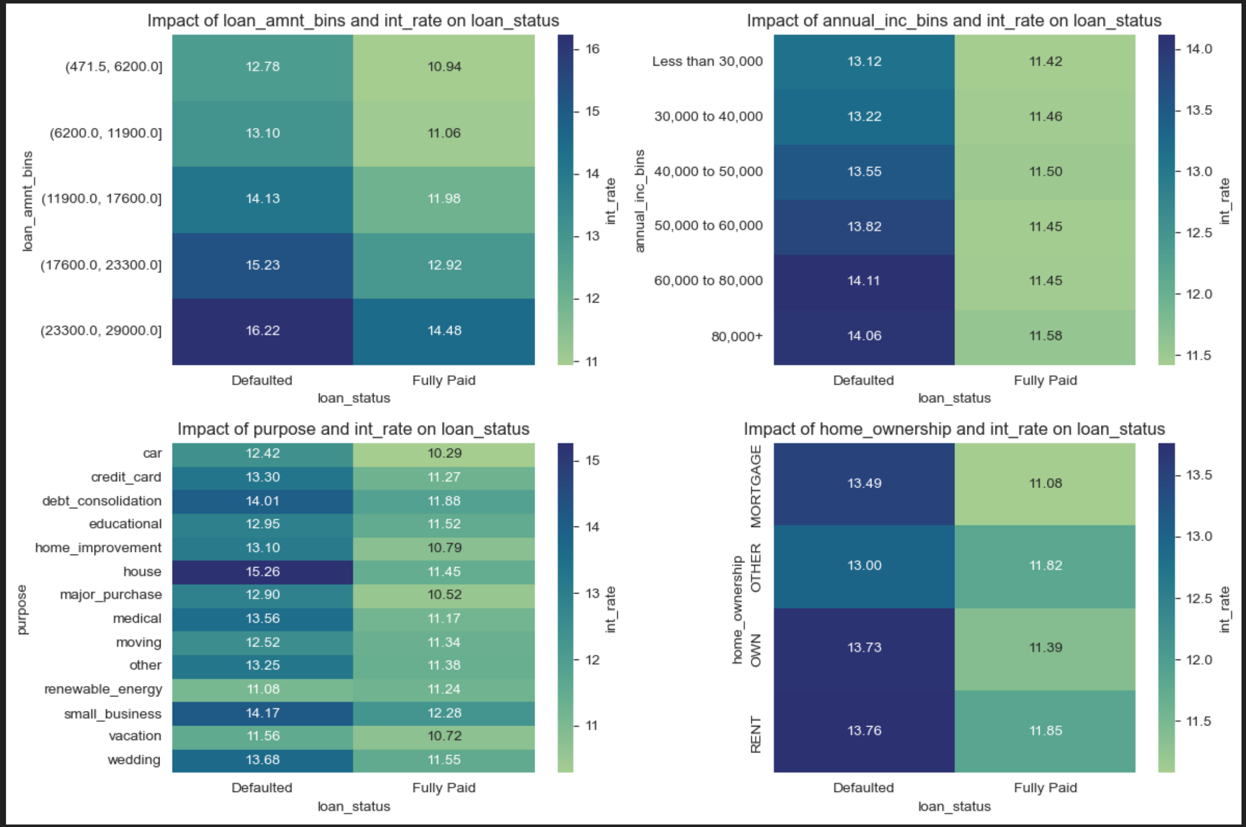


Bivariate Analysis – Part 1

- Loan amount, Funded amount, Funded amount invested, and Installment are highly correlated.

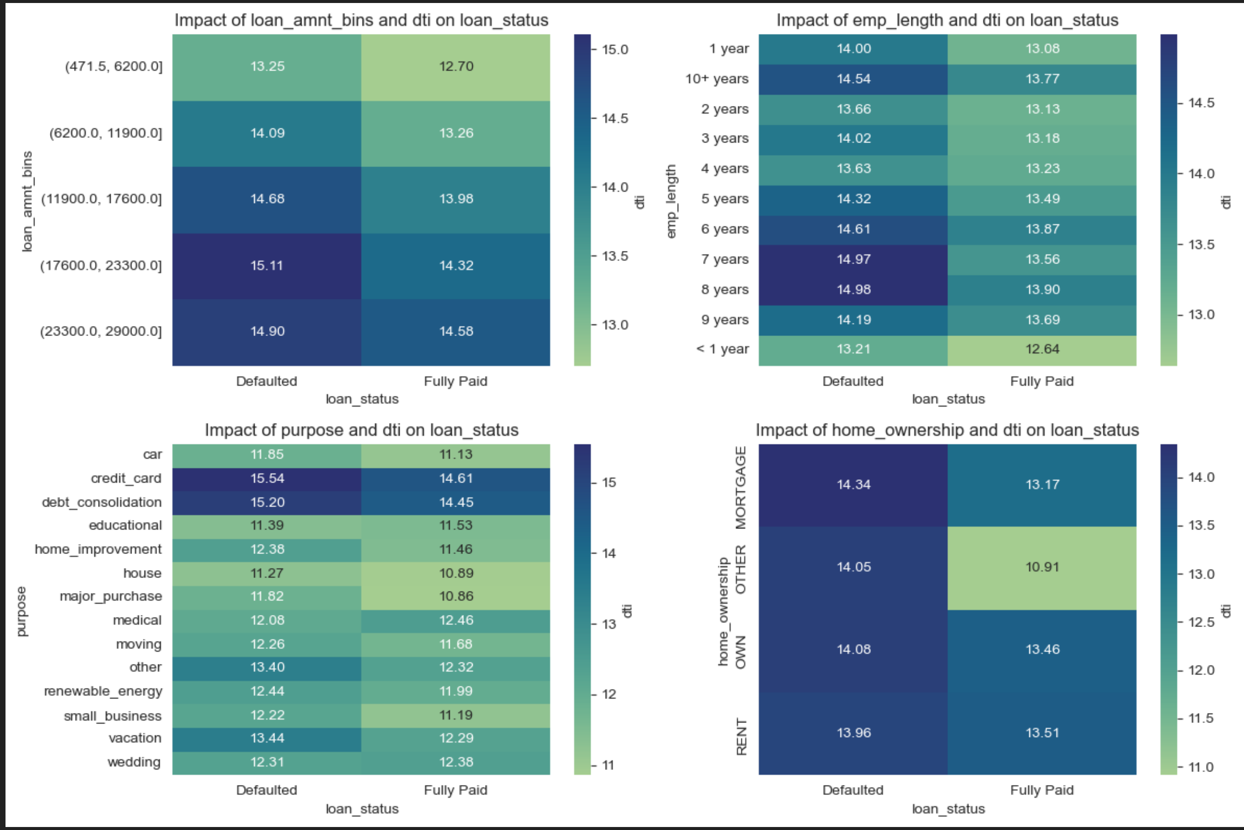
Bivariate Analysis – Part 2

- Higher interest rate indicates higher loan defaults across most range of other variables



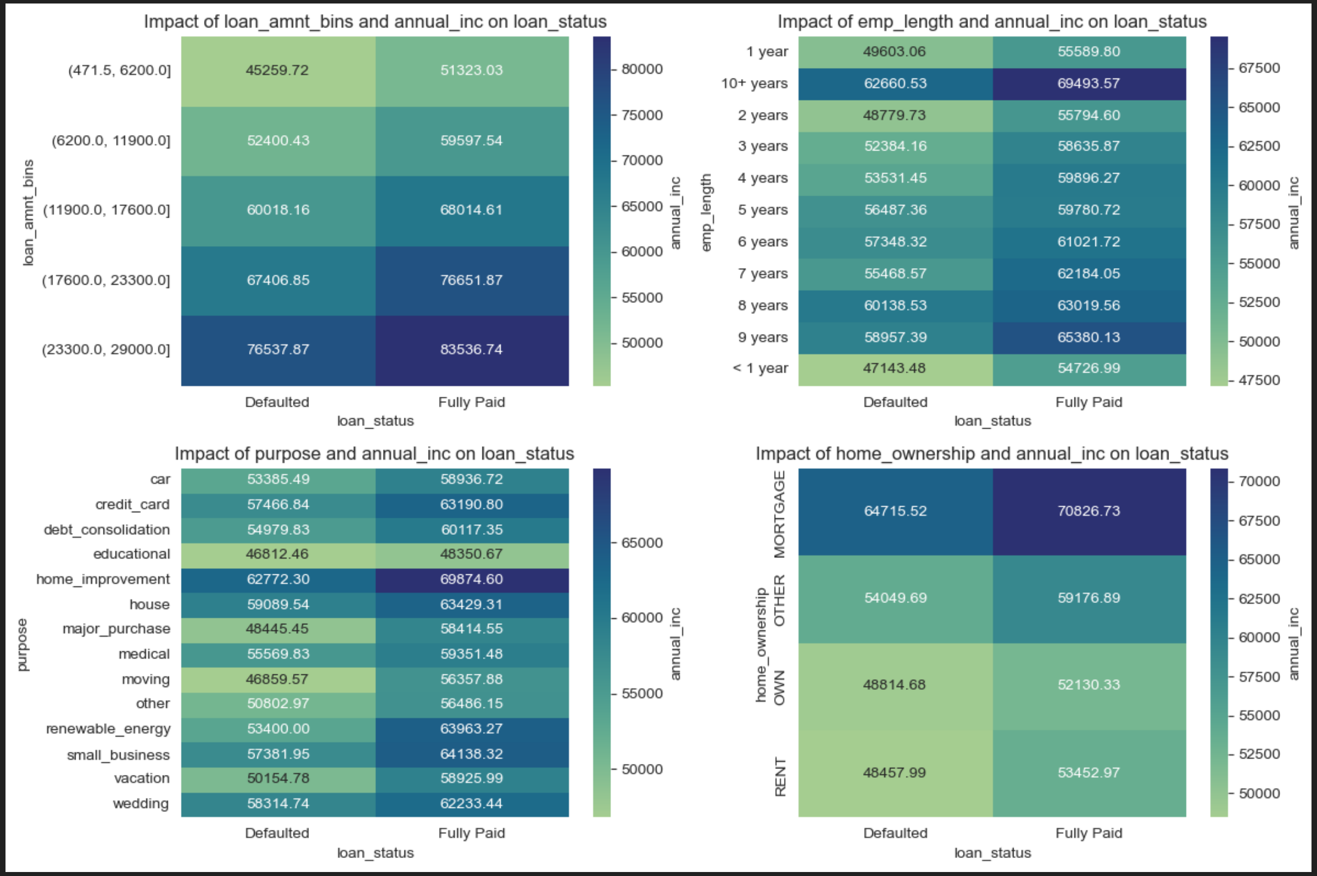
Bivariate Analysis – Part 3

- Higher Debt to Income ratio indicates higher loan default rates across most range of other variables



Bivariate Analysis – Part 4

- Higher Annual Income indicates lower loan default rates across most range of other variables



Recommendations

The key factors that drive higher loan default rates include the following.

- Borrower has low salary
- Loan amount is high
- Interest rate is high
- 60 months term instead of 30 months term
- Lower Grade loans
- Loan type of Small Business
- Borrower has been bankrupt before
- Higher Debt to Income ratio