Lending Club Case Study

Name: Harshitha S, Phanidhar Gadiyaram

Problem statement

- Lending loans to customers is one of the main purposes of a bank. Lending to qualified customers is crucial for profitability. If customers pay their loans on time, the bank makes a profit.
- Conversely, lending to customers who cannot repay or not lending to eligible customers results in losses.
- It's important to identify customers as eligible or non-eligible for loans based on specific criteria.
- In this case study, we analyze loan histories to derive parameters that determine customer credibility in terms of loan repayment.
- We also seek to understand generic loan patterns: who takes loans, when, where, and why.

Solutioning…

- The dataset provided contains a large number of rows and columns, necessitating careful selection of appropriate data for analysis.
- **Data Cleaning** involves:
  - Removing Rows/Columns: Eliminating irrelevant or incomplete data.
  - Standardizing Formats: Ensuring consistent data types and formats.
  - Handling Outliers: Addressing data points that significantly deviate from the norm.
- **Univariate Analysis**:
  - Parameter Visualization: Independently examining parameters to identify customer patterns.
- **Segmented Analysis**:
  - Impact on Loan Default: Assessing how specific parameters directly influence loan default rates.
- **Bivariate/Multivariate Analysis**:
  - Parameter Relationships: Exploring interactions between different parameters and their collective impact on loan status.

Data cleaning and variables

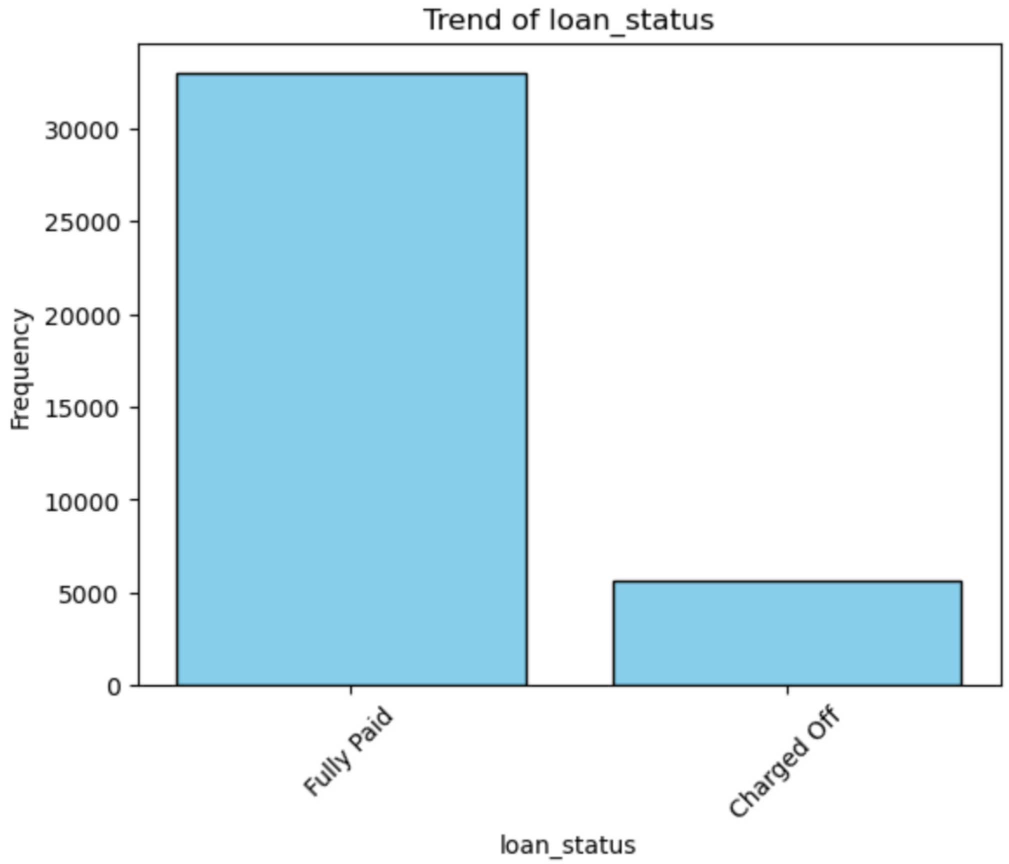
Data cleaning steps

Understanding data, data types, missing values.
Dropping all columns with max null/Na value, insignificant column, having same values for all rows
Removing all rows with loan status as “current”
Correcting the data types removing %, converting Date and time variables and converting string – int/float
Imputing missing categorical variable with Mode value
Imputing missing numerical variable with Median value
Removing Outliners

Final Variable

loan_amnt	funded_amnt	funded_amnt_inv
term	int_rate	installment
grade	sub_grade	emp_length
home_ownership	annual_inc	verification_status
issue_d	loan_status	purpose
addr_state	dti	delinq_2yrs
inq_last_6mths	open_acc	revol_bal
revol_util	total_acc	pub_rec_bankruptcies

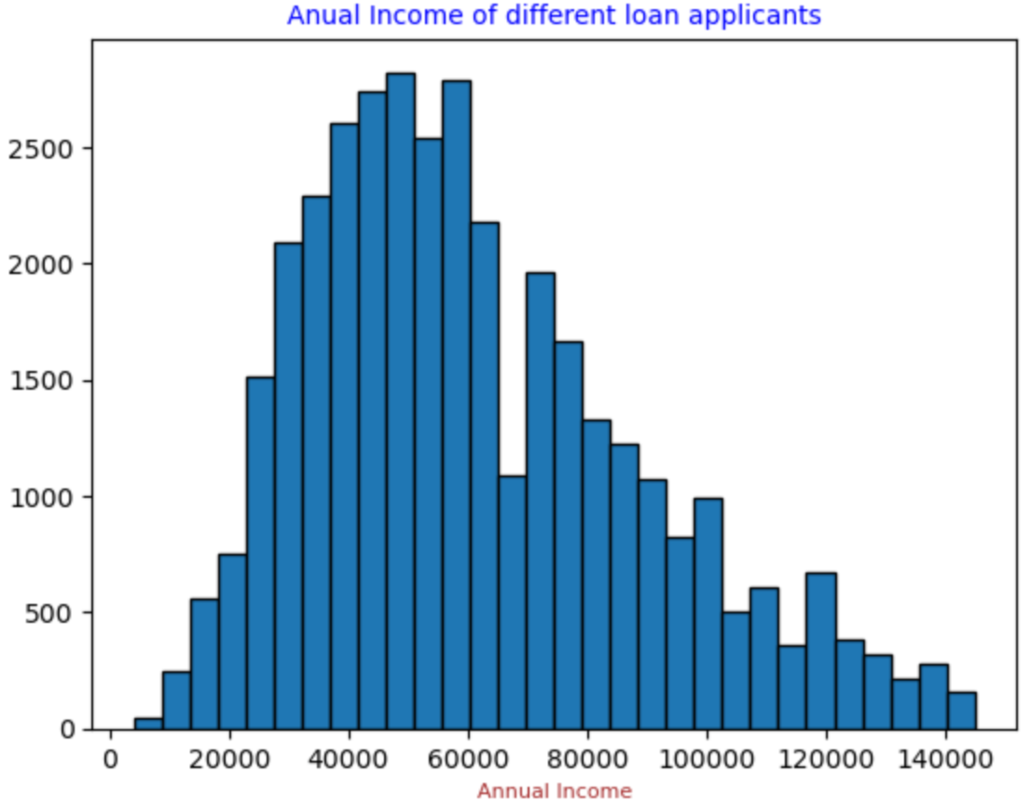
Understanding Default vs Fully paid

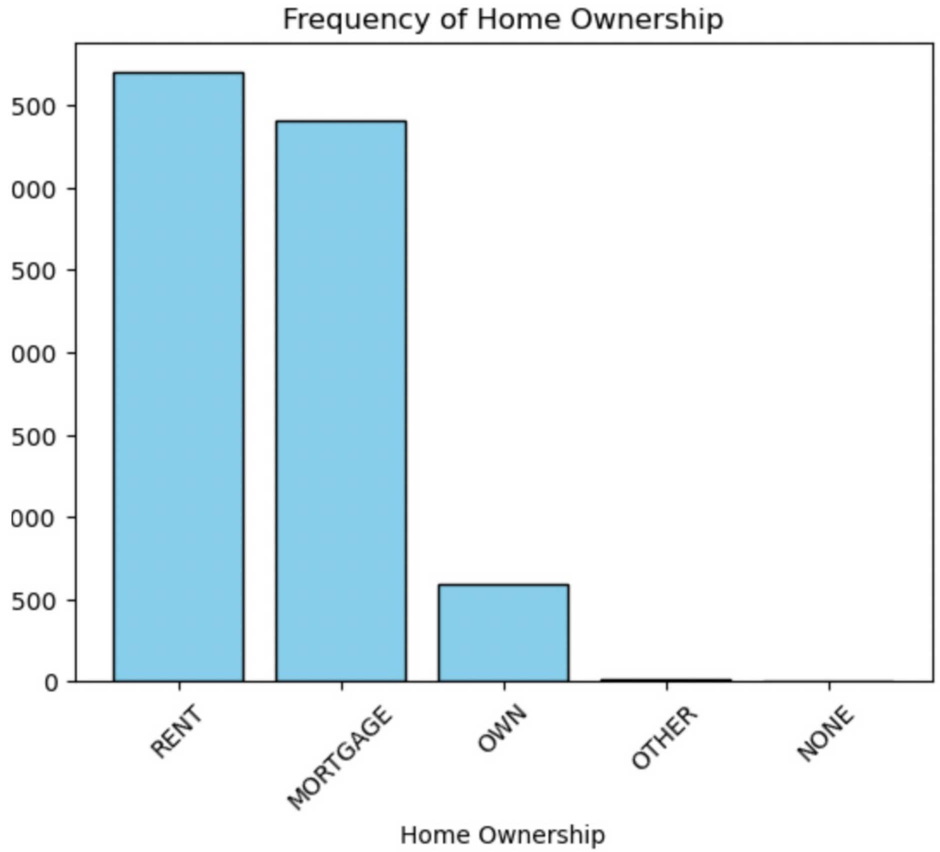


Observation: Loan Repayment Outperforms Charge-offs

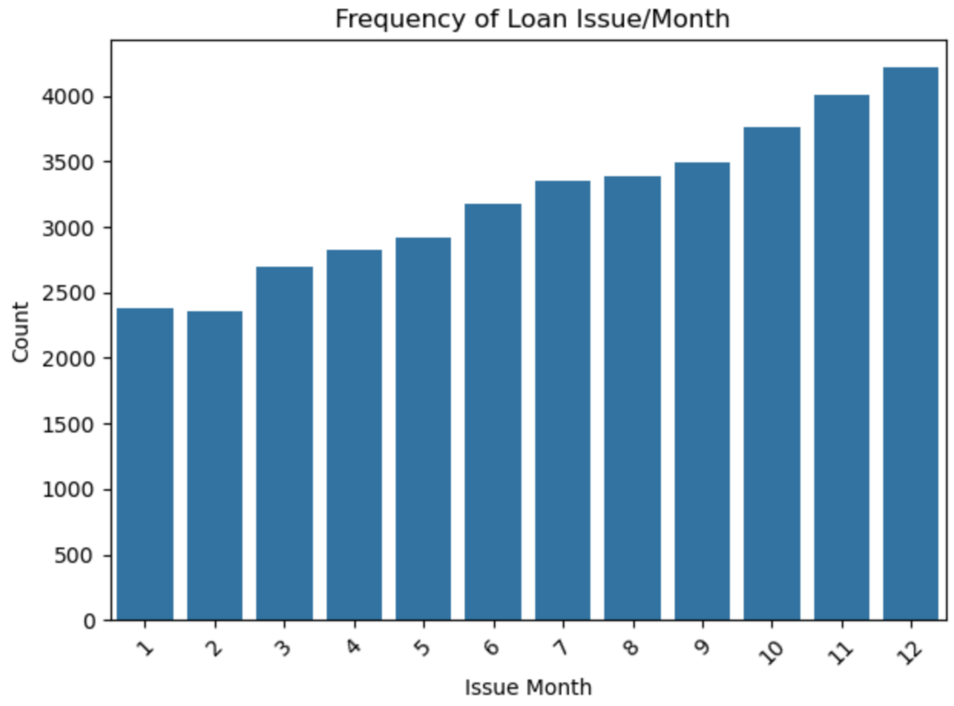
•The data clearly shows a higher number of customers are paying off their loans compared to those who have defaulted.

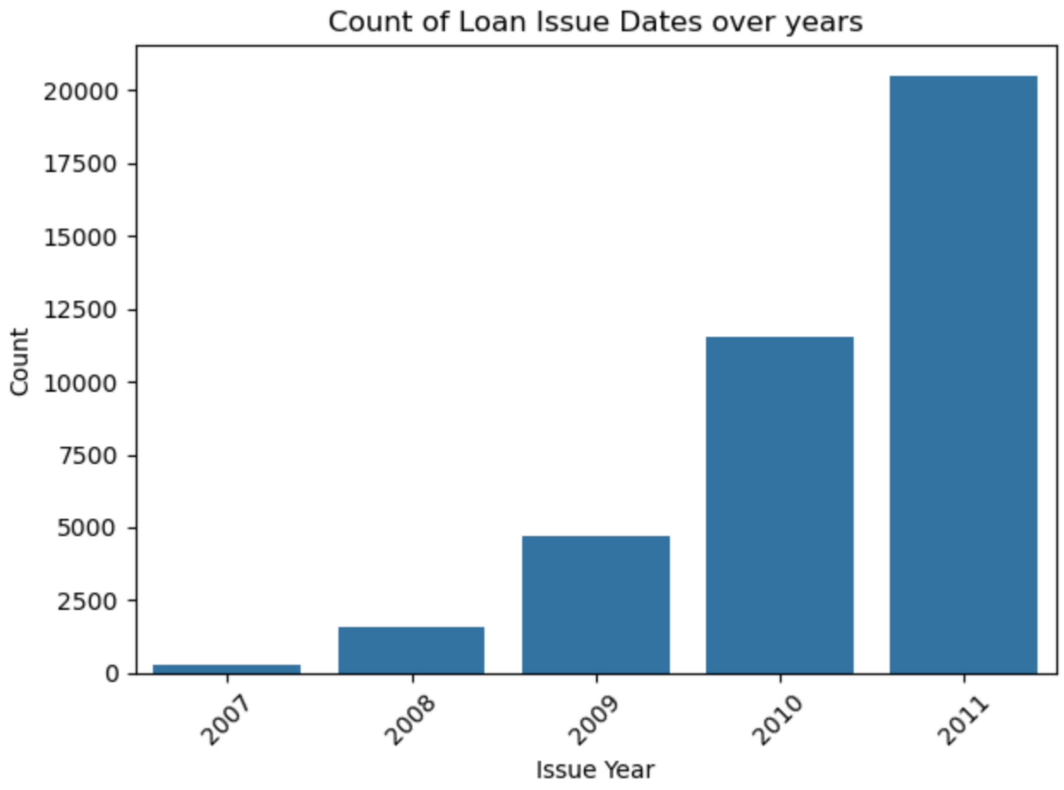
•This positive trend is a favourable indicator for our analysis.





Analyzing the annual Income and home ownership graph we can understand people with middle income level are taking more loan. Compared to people having high annual income with own house.





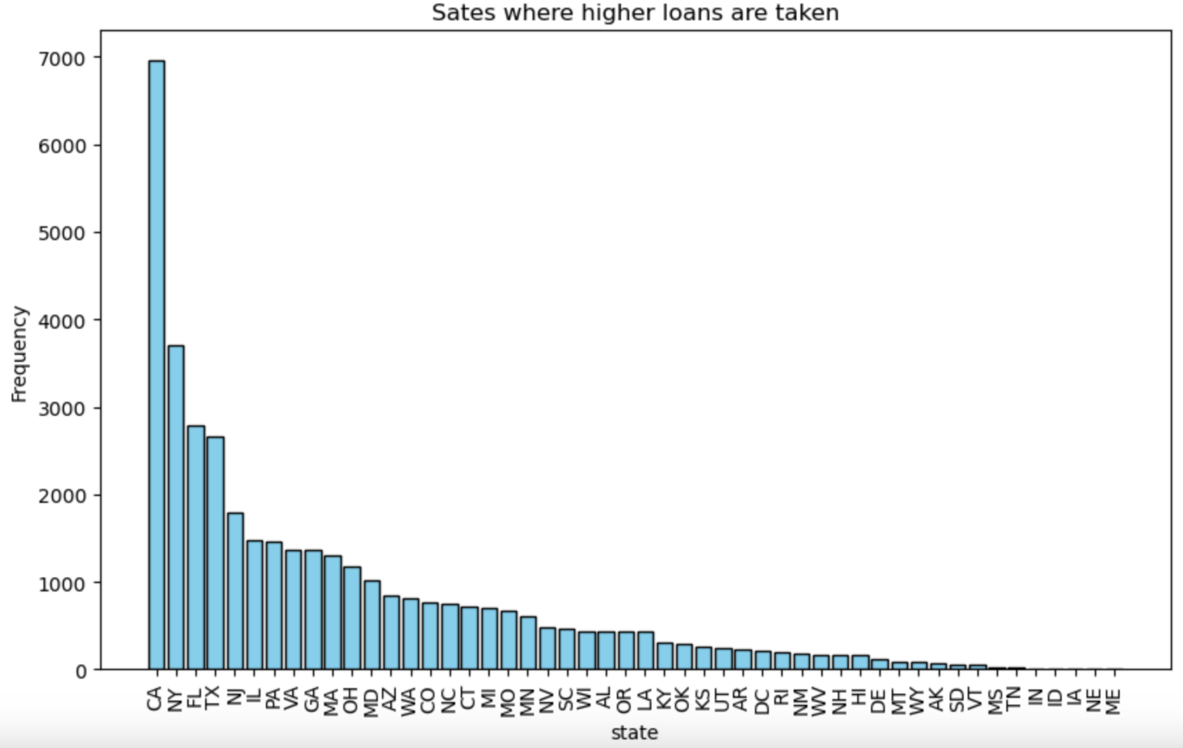
Trend Analysis: Increasing Loan Applications Over Time

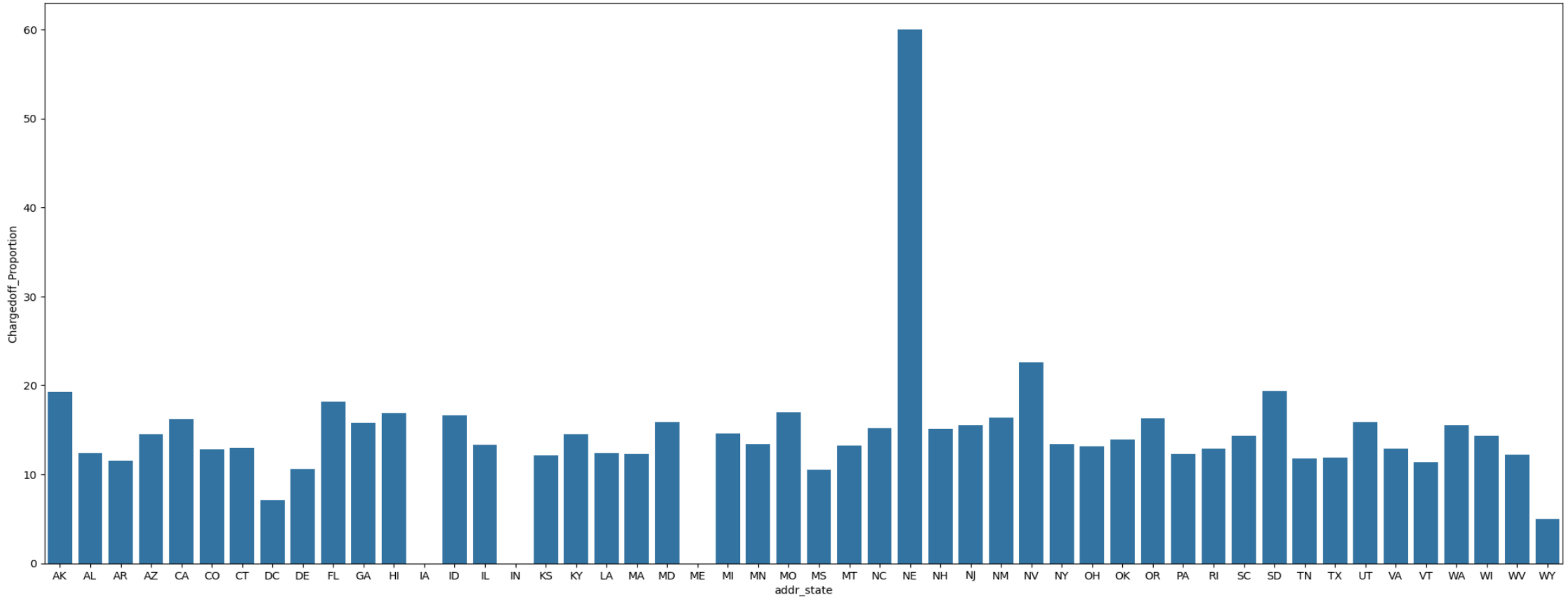
•There is a noticeable upward trend in loan applications over the years, peaking towards the end of each year.

•This trend coincides with various economic factors and the holiday season, when consumer spending typically rises

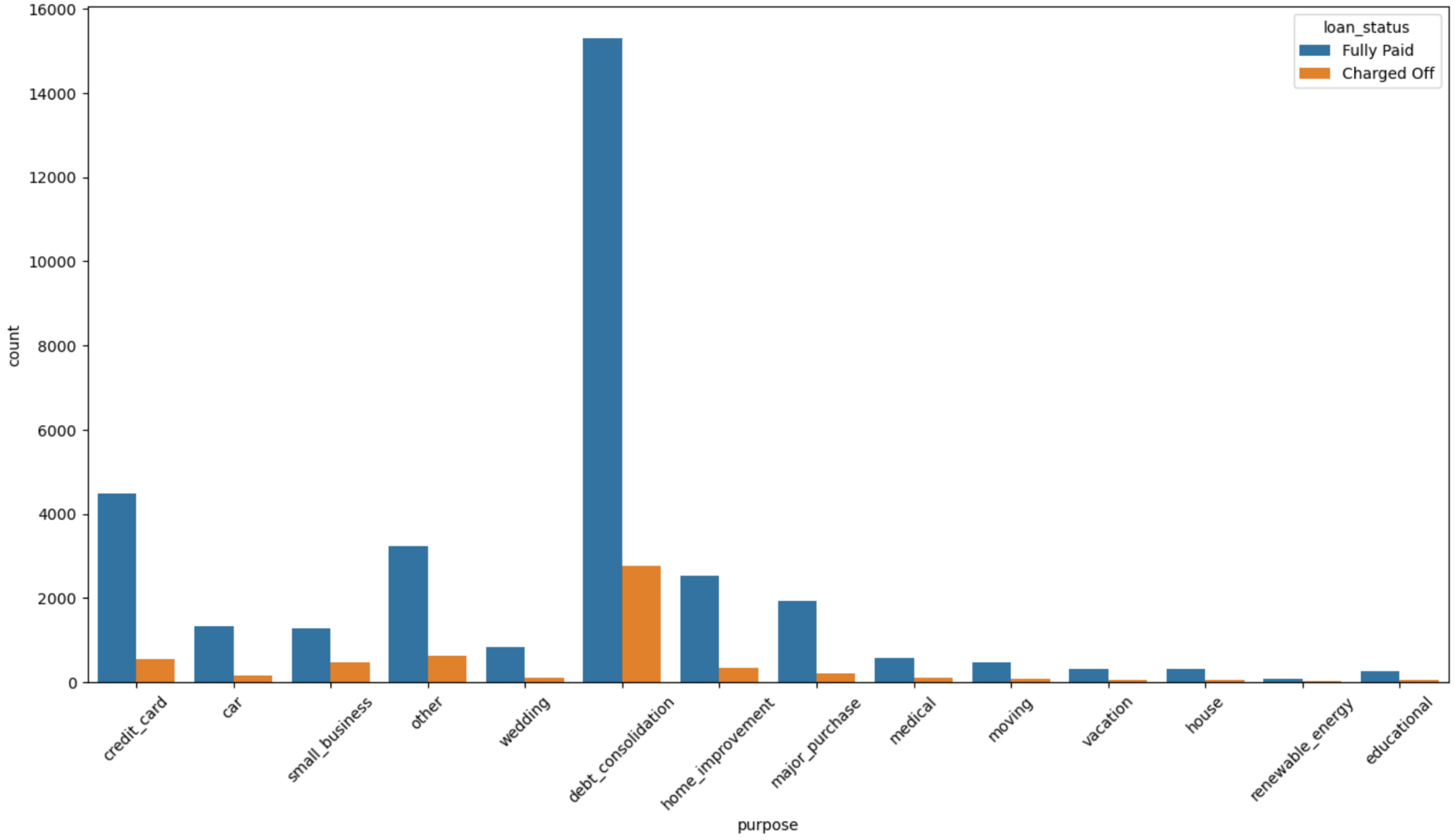
Statewise Loan Distribution and Default Analysis

- The first graph displays states with the highest loan uptake
- Notably California (CA) leading among all states.
- This trend likely correlates with the state's strong economy and high GDP.
- Interestingly, despite CA having the highest loan uptake, it is not the state with the highest default rates.
- The Northeast (NE) region shows the highest default rates, indicating a need to assess state-specific economic factors before approving loans.

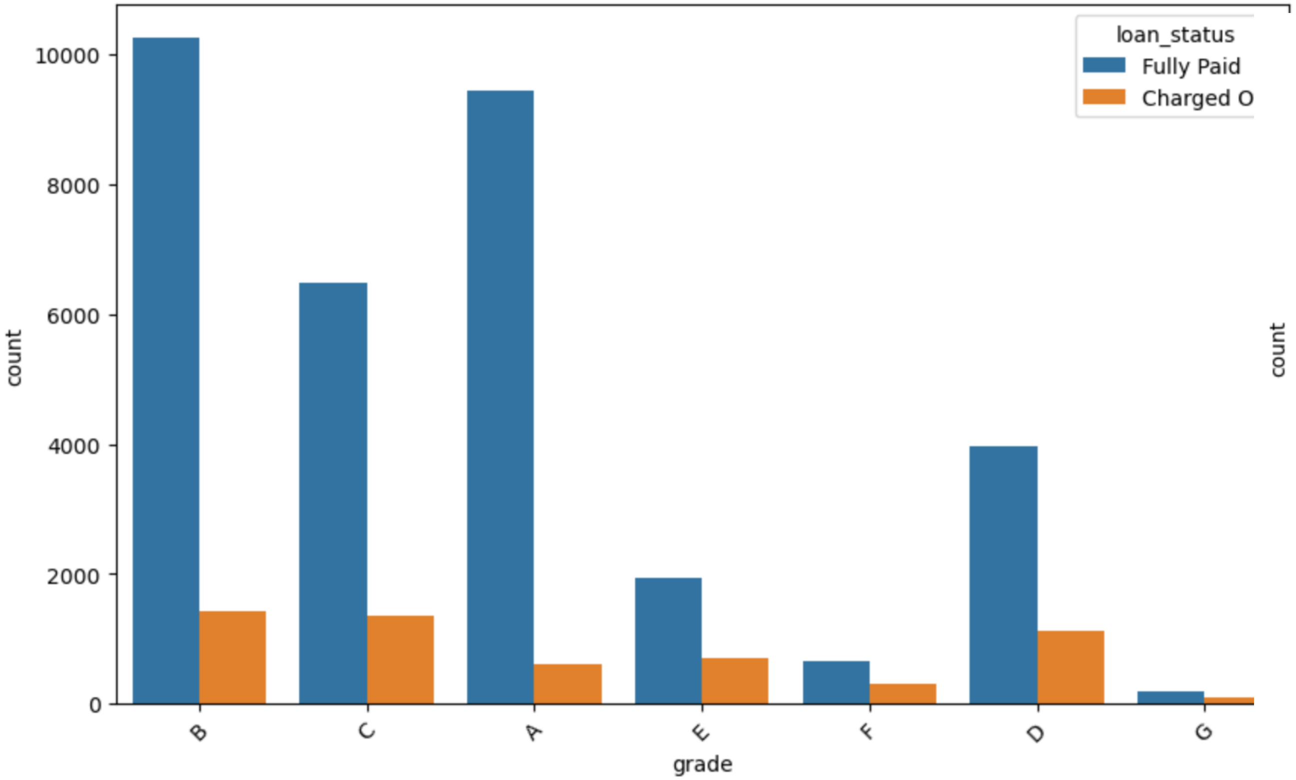


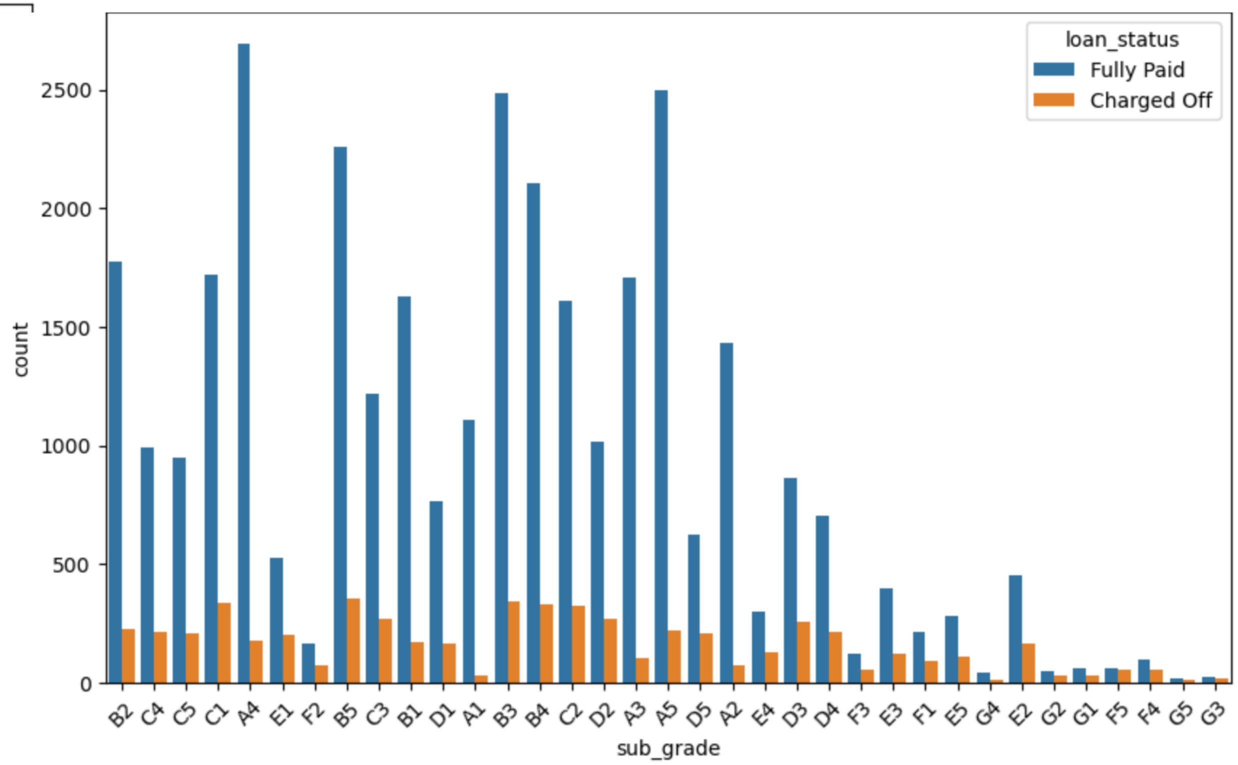


Loan Reason vs Loan status(Paid completely)



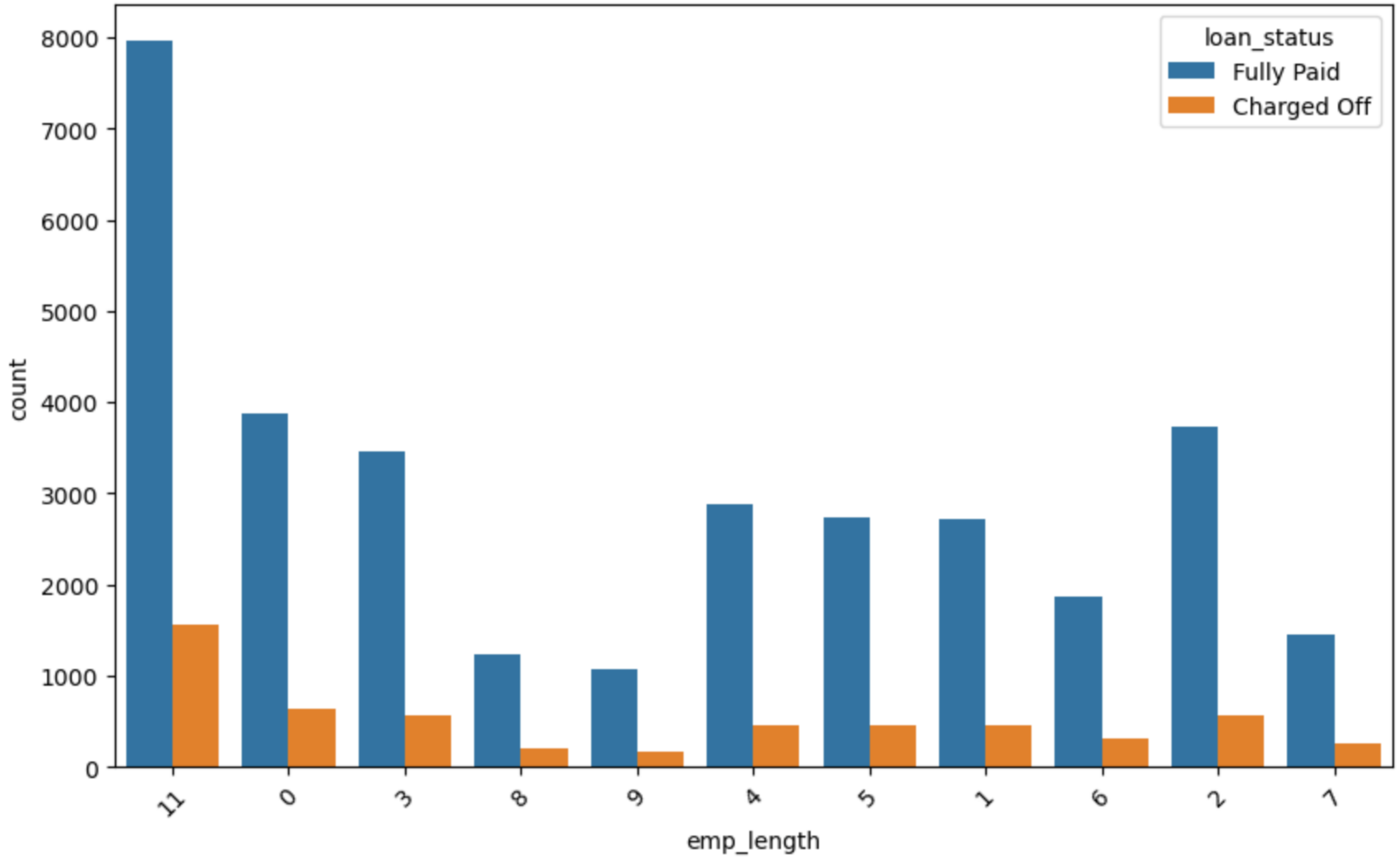
- Many customers are taking loans major for debt consolidation.
- However, the repayment rate for these loans is not significantly high
- This suggests a significant trend: customers using loans for debt consolidation often struggle with repayment, indicating a recurring cycle of debt.





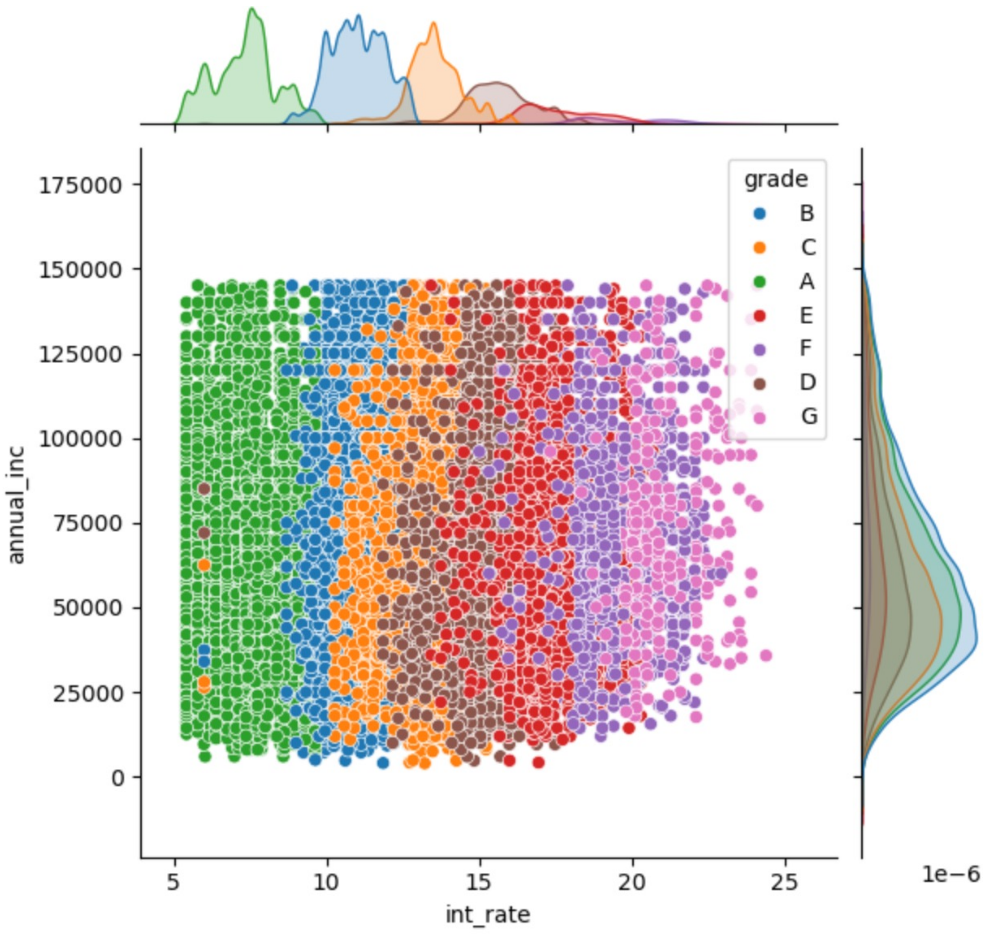
Observing Grade wise loan approval

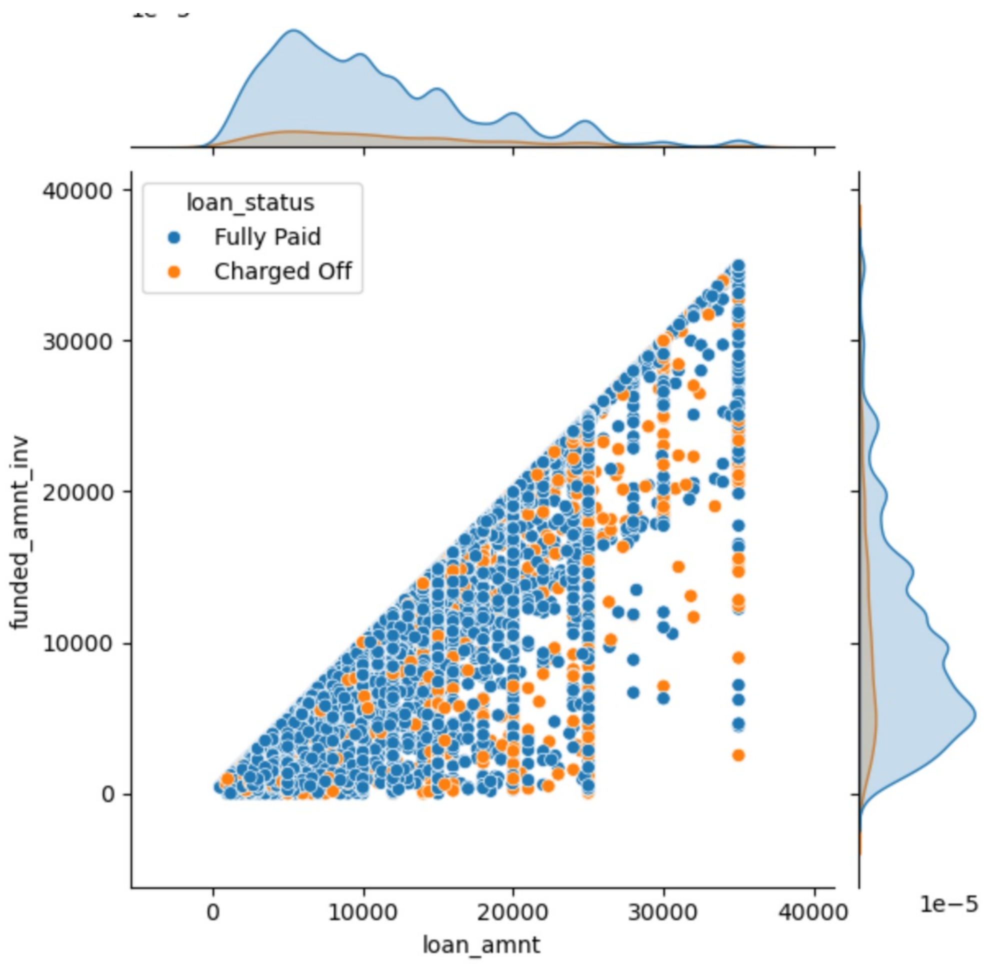
- Applicants in Grade A, B, and C have received the majority of loan approvals.
- Grade C and D borrowers show a higher likelihood of repayment compared to other grades.
- Historical data highlights that subgrades A4, B5, B3, and A5 have notably high numbers of loans approved.



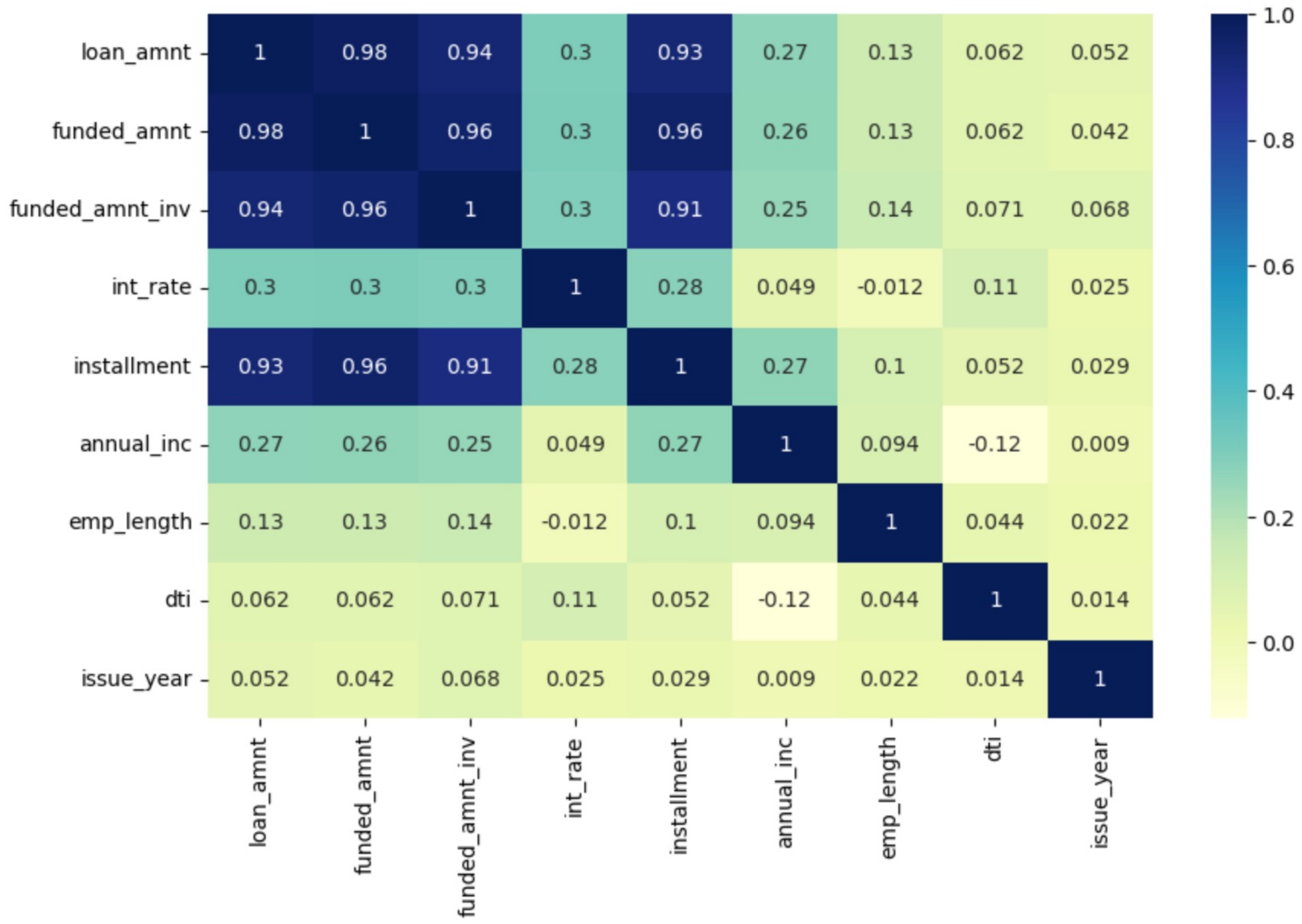
We can infer from the graf that:

- Applications from borrowers with higher experience levels show a trend of profitable repayment.
- However, it's noteworthy that the charged-off percentage among higher experience borrowers is moderately higher compared to borrowers with lower experience.





- From the graph, it is evident that higher loan amounts, higher interest rates, and longer terms are associated with higher charge-off rates.
- These loans typically fall under grade G in our analysis.



- Loan Amount, Funded Amount, Funded Amount Investment, and Installment show a positive correlation.

- Annual Income and Debt-to-Income Ratio (DTI), as well as Interest Rate and Employment Length, exhibit negative correlations.

- Specifically:
  - •Lower Annual Income correlates with higher DTI.
  - •Lower Interest Rates correlate with longer Employment Length, and vice versa.

Conclusion

- Annual Income with mid range are the customers who are generally taking loan
- Increase in loan amount increase interest rate
- Higher Employee experience is like a relatively safe option
- Higher interest rates are also leading to higher charge-off
- Some of the variables are negatively correlated so those combination has to be well chosen before lending