

Lending Case Study

Introduction

At the forefront of consumer finance, our company specializes in providing diverse loan options tailored to urban customers. When evaluating loan applications, we face the crucial task of determining loan approval based on the applicant's profile. This decision carries two distinct types of risks that impact our operations and financial health.

1. If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company.
2. If the applicant is not likely to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company.

Objective

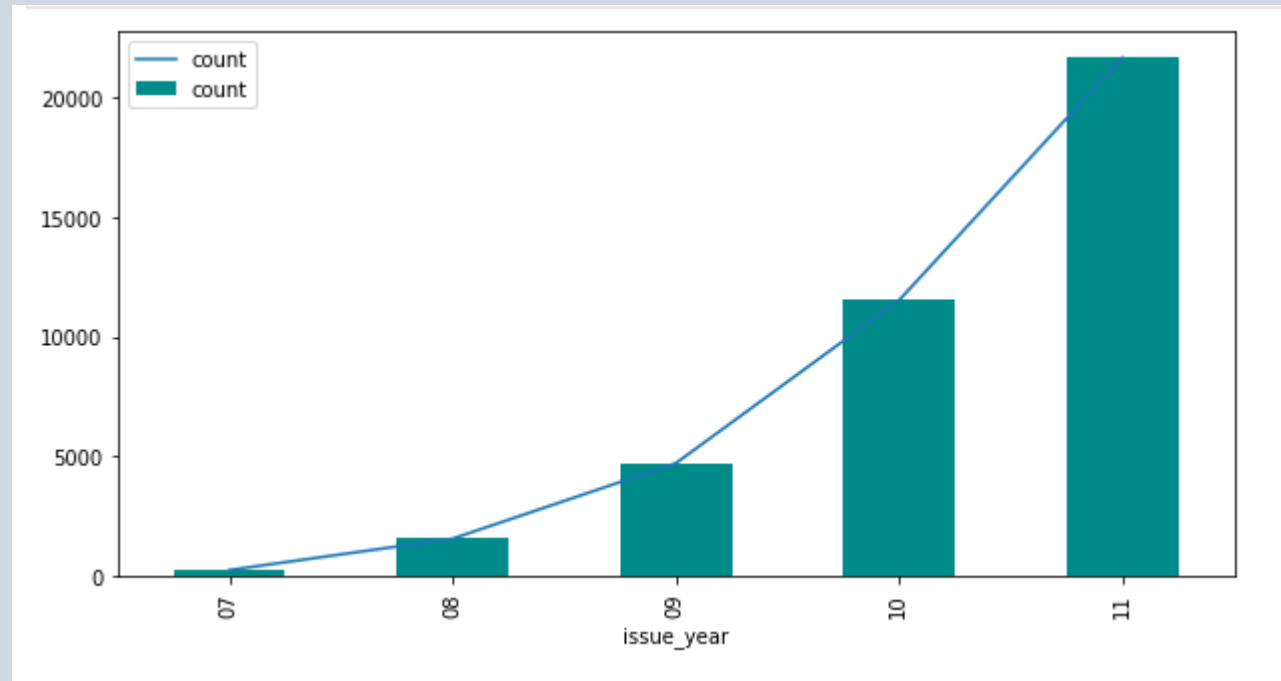
Our goal is to identify risky loan applicants by discerning patterns and correlations within the provided dataset. This endeavor aims to equip our organization with insights that will effectively mitigate business losses associated with loan defaults.

EDA

After analyzing the data we performed following

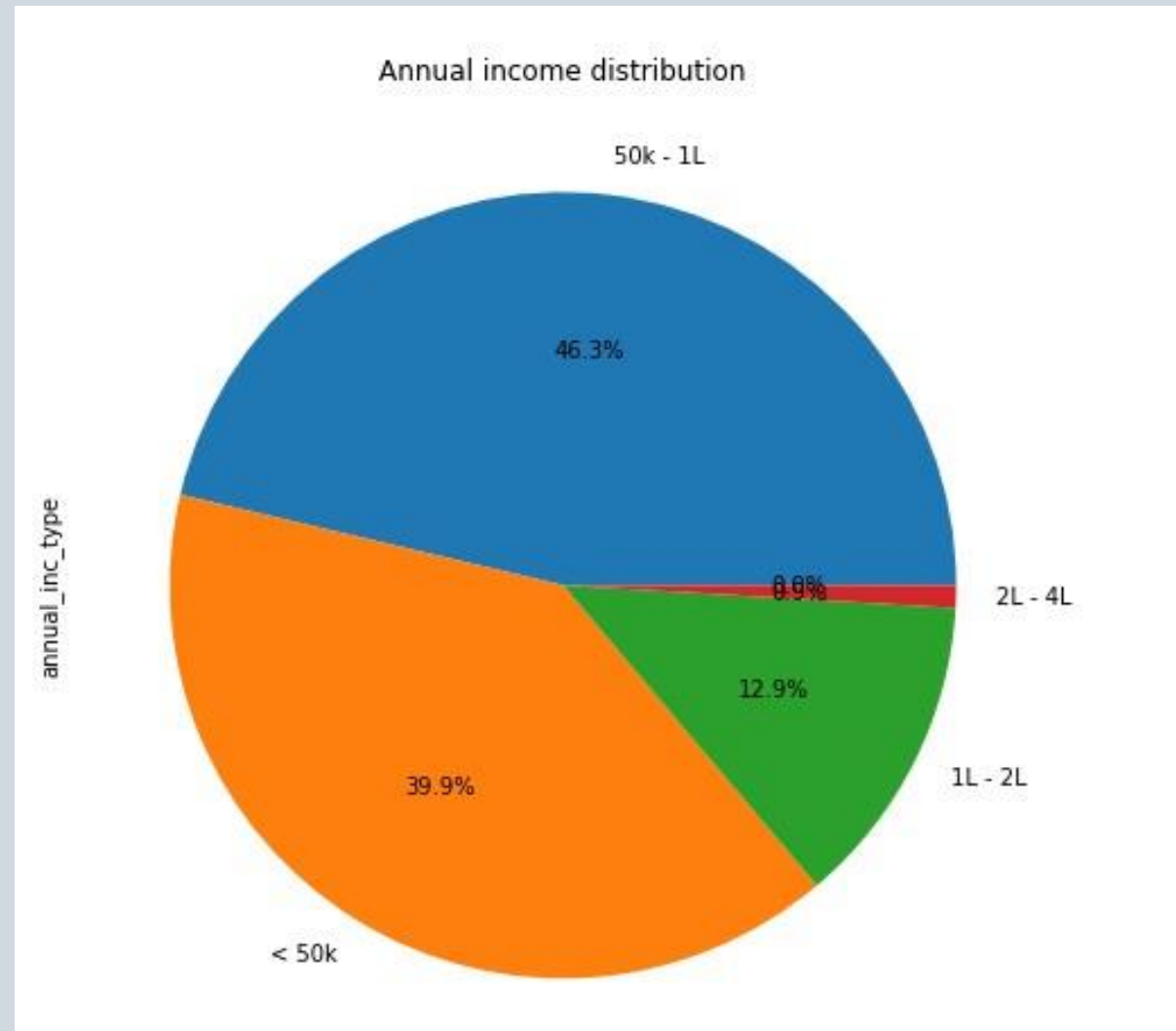
1. Removal of Columns with 100% Null Values: Eliminated columns devoid of useful data to streamline the dataset.
2. Data Cleaning: Employed various techniques to ensure data accuracy and consistency.
3. Missing Value Treatment: Implemented strategies such as imputation or removal to address missing values.
4. Data Imputation: Filled missing values using appropriate methods to maintain data integrity.
5. Outlier Detection and Treatment: Identified and treated outliers to prevent them from skewing the analysis.
6. Derivation of Various Columns: Created new columns from existing data to extract additional insights.
7. Creation of Correlation Matrix: Generated a correlation matrix to analyze relationships between different attributes in the dataset.

After Plotting a graph of loan_status vs issue_year



Inference:

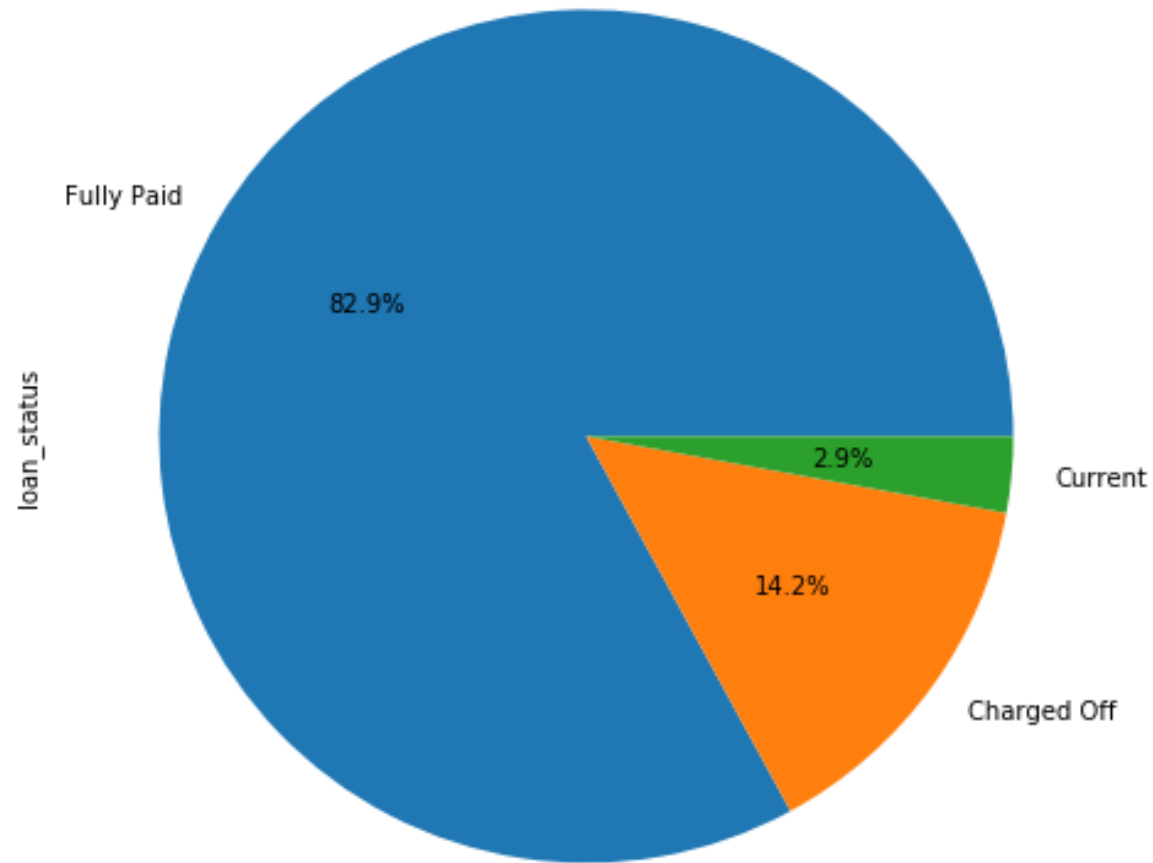
There is a steady increase in the issuance of loan



Inference:

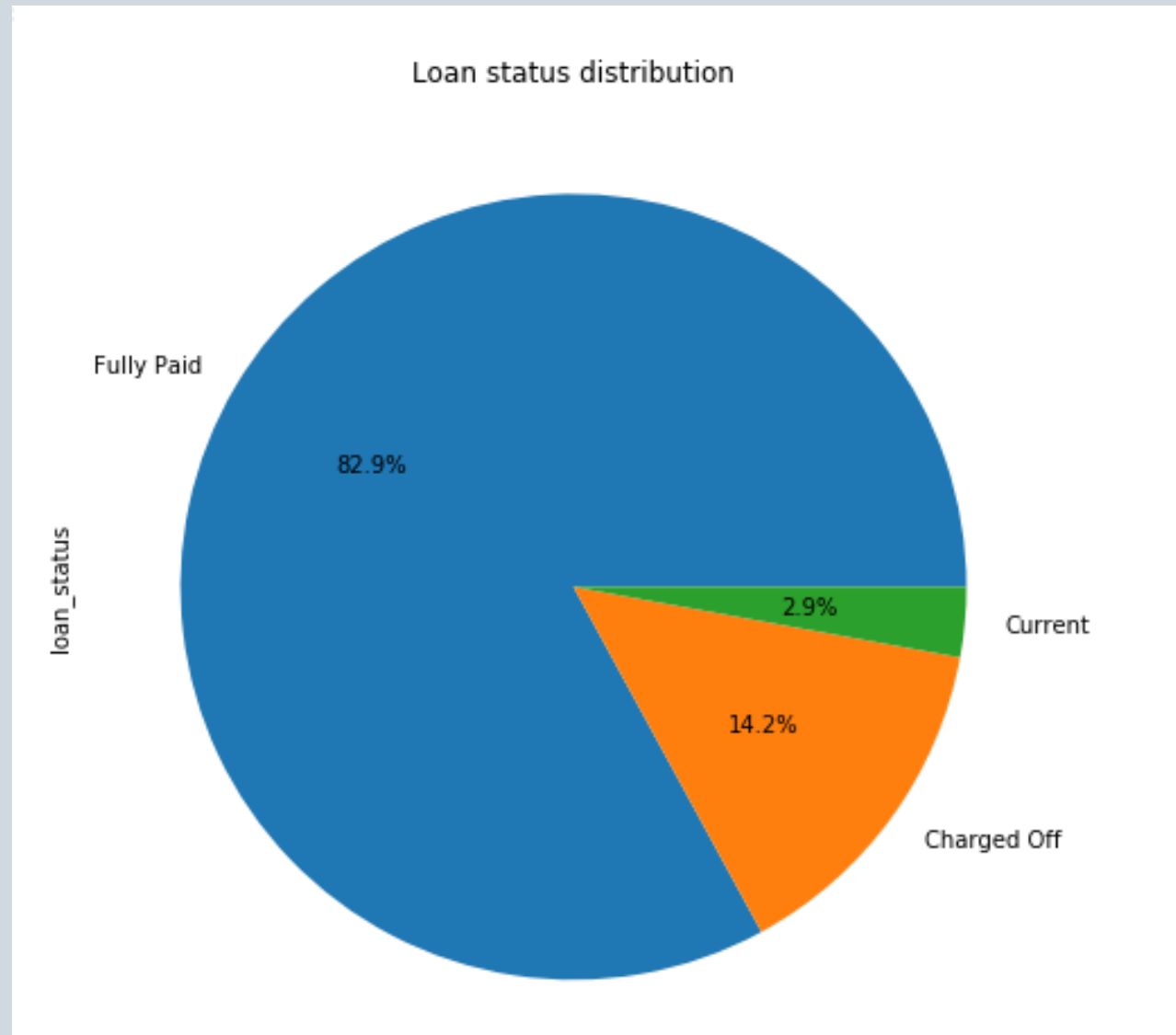
People with Annual income below 2 lakhs have taken more loans

Loan status distribution



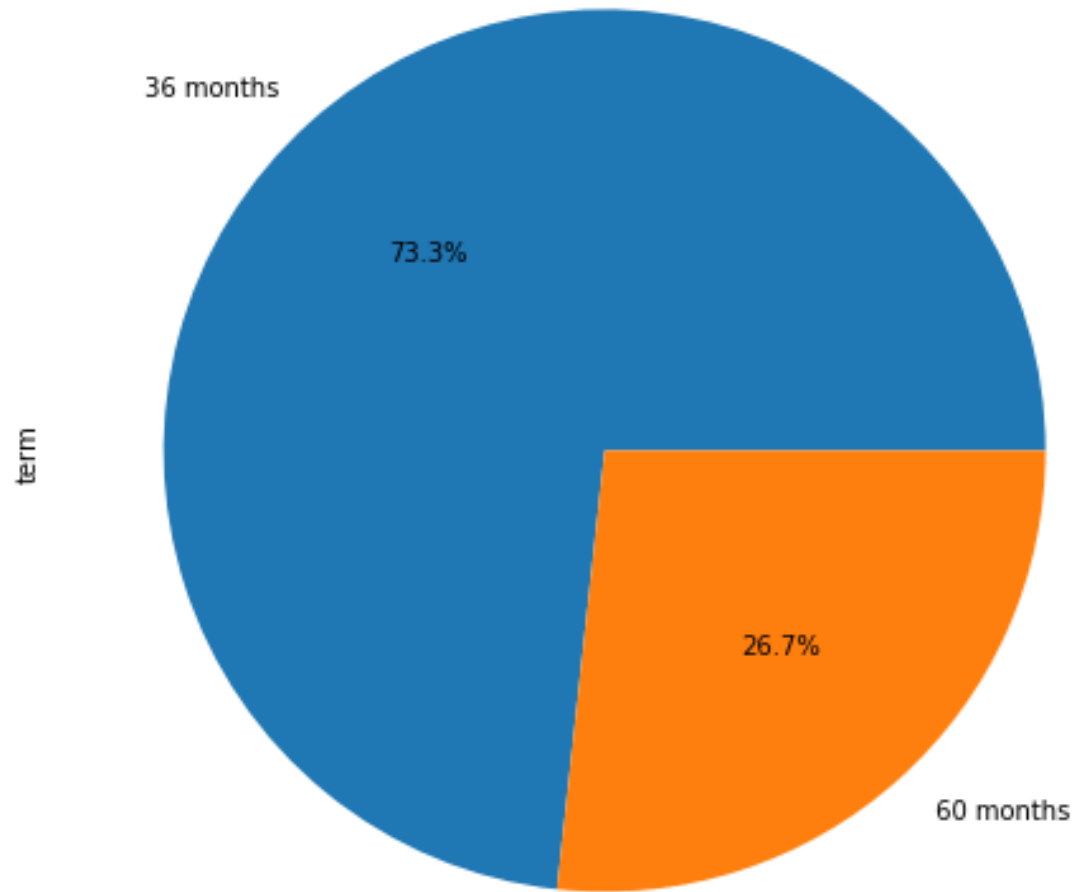
Inference:

14.2% persons are labelled as 'Charged Off'



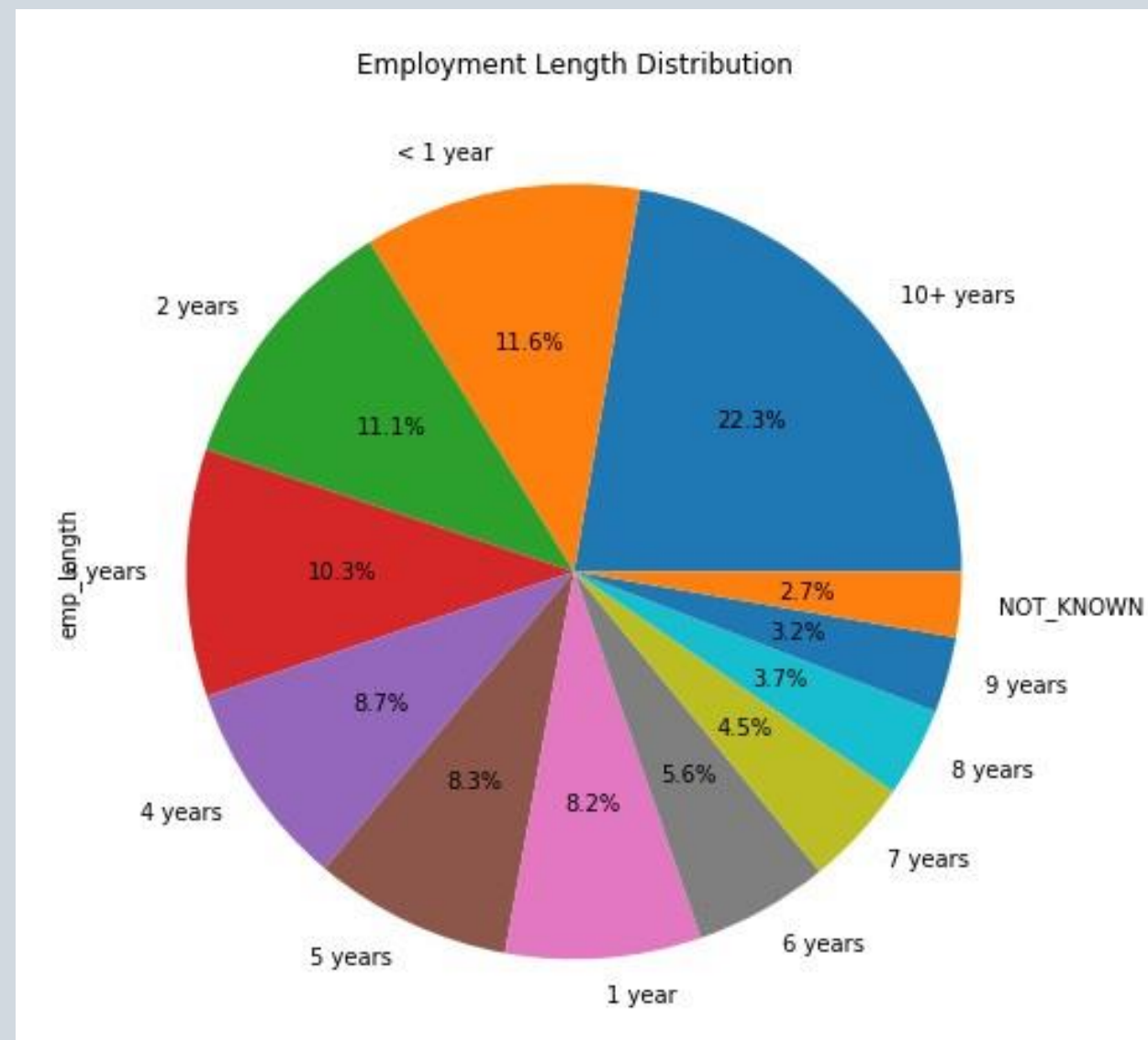
Inference:
14.2% persons are labelled as 'Charged Off'

Term Distribution



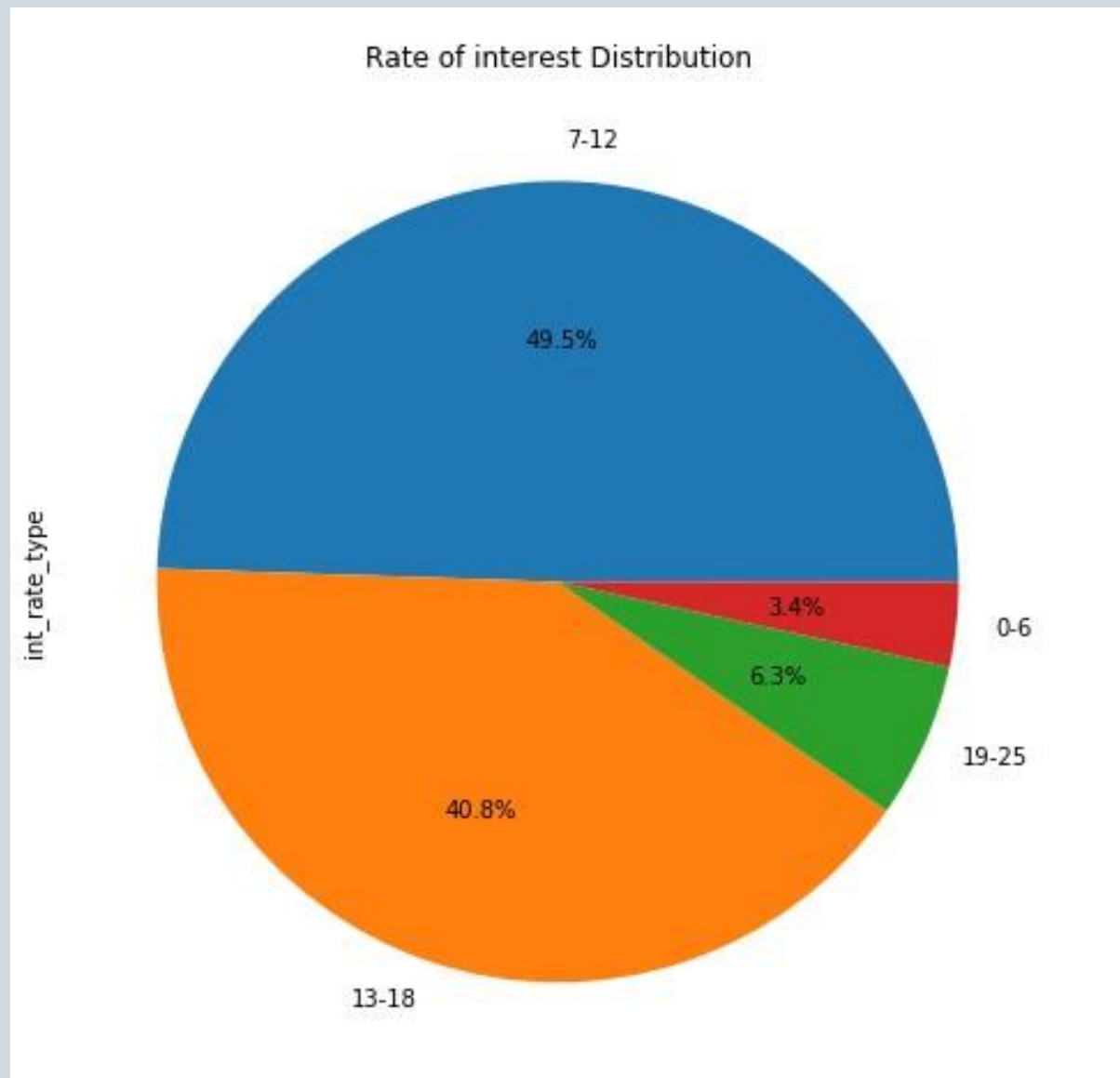
Inference:

14.2% persons are labelled as 'Charged Off'



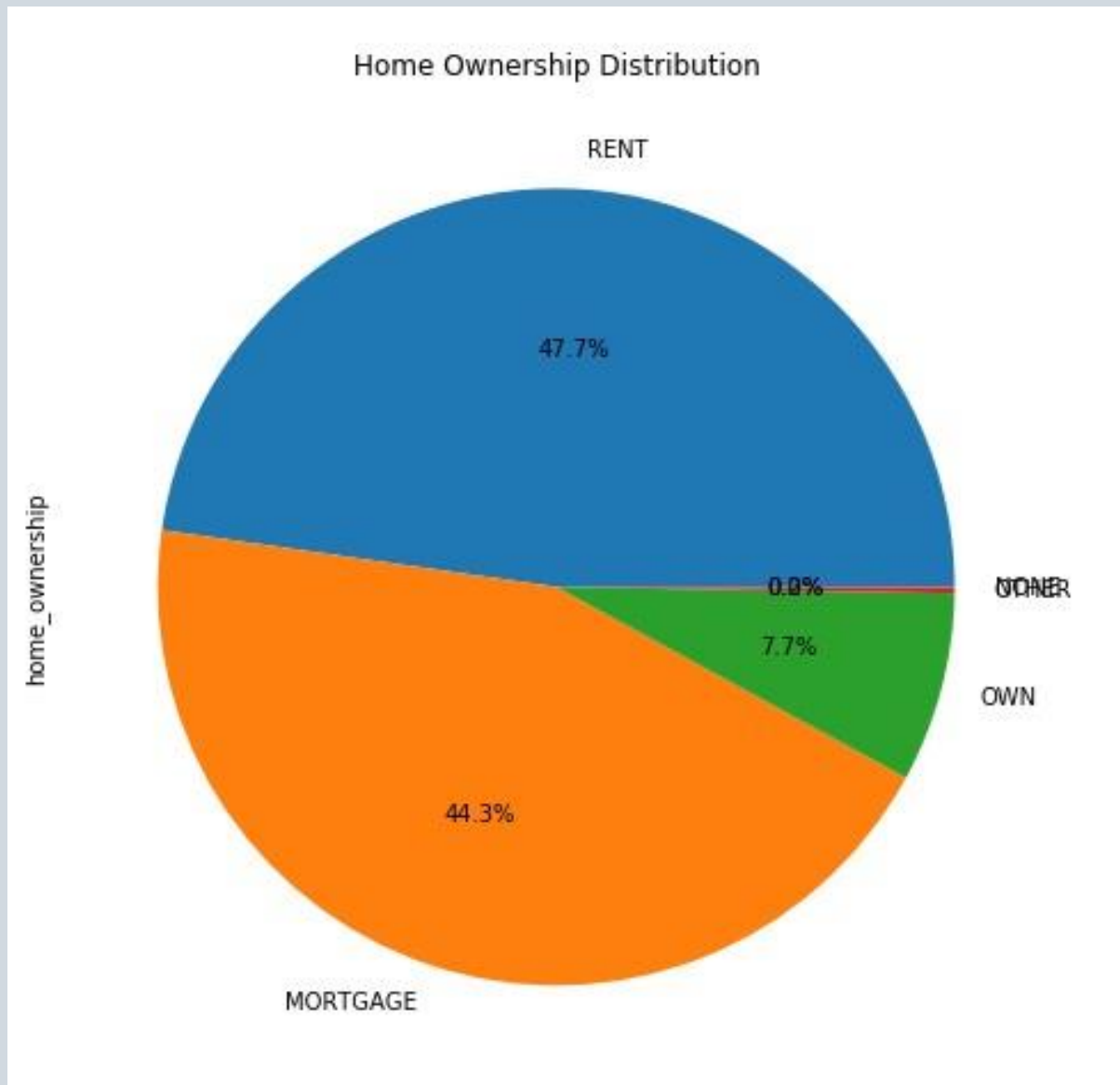
Inference:

People with Employment Length > 10 years have taken more loan

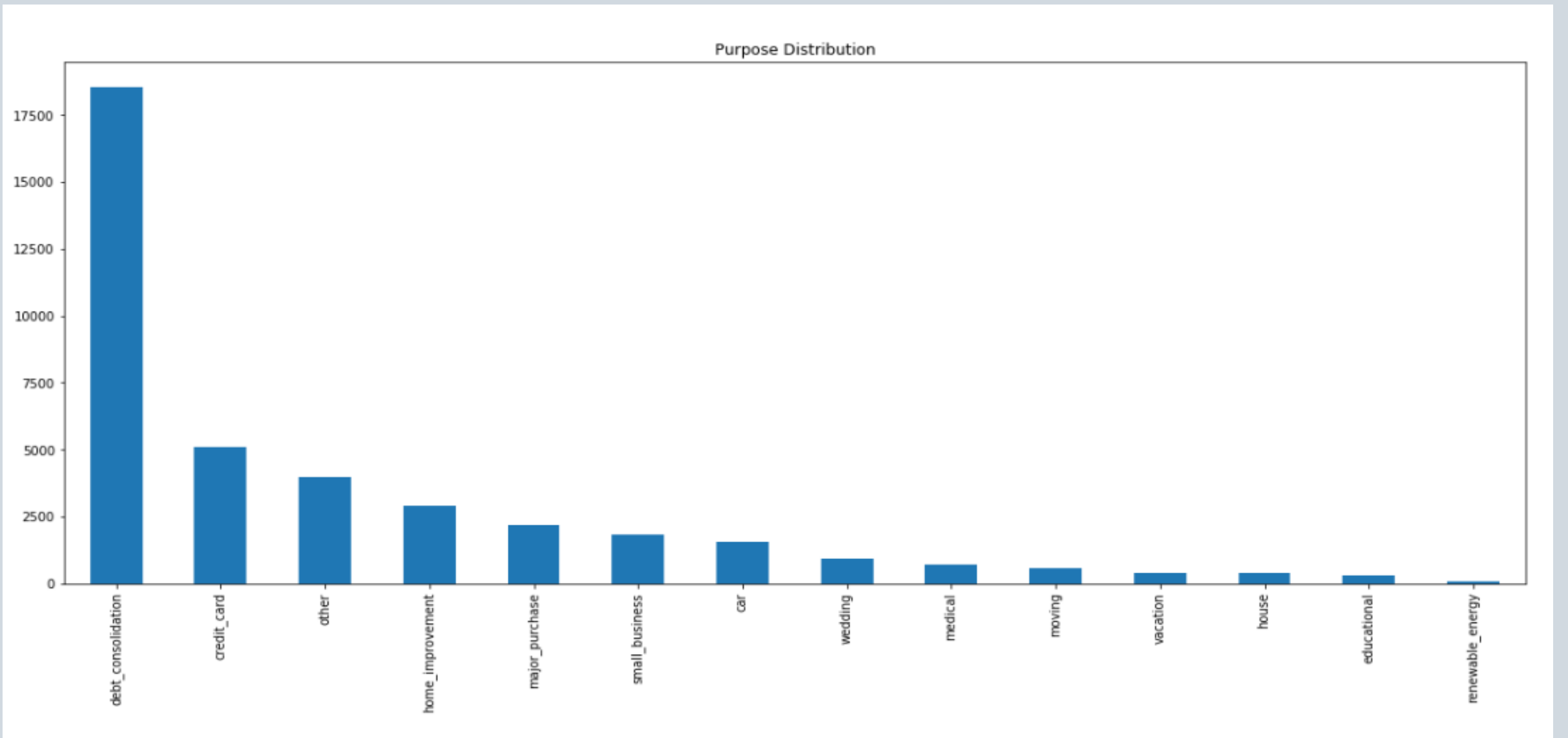


Inference:

People have 7%-12% of rate of interest have opted more loans



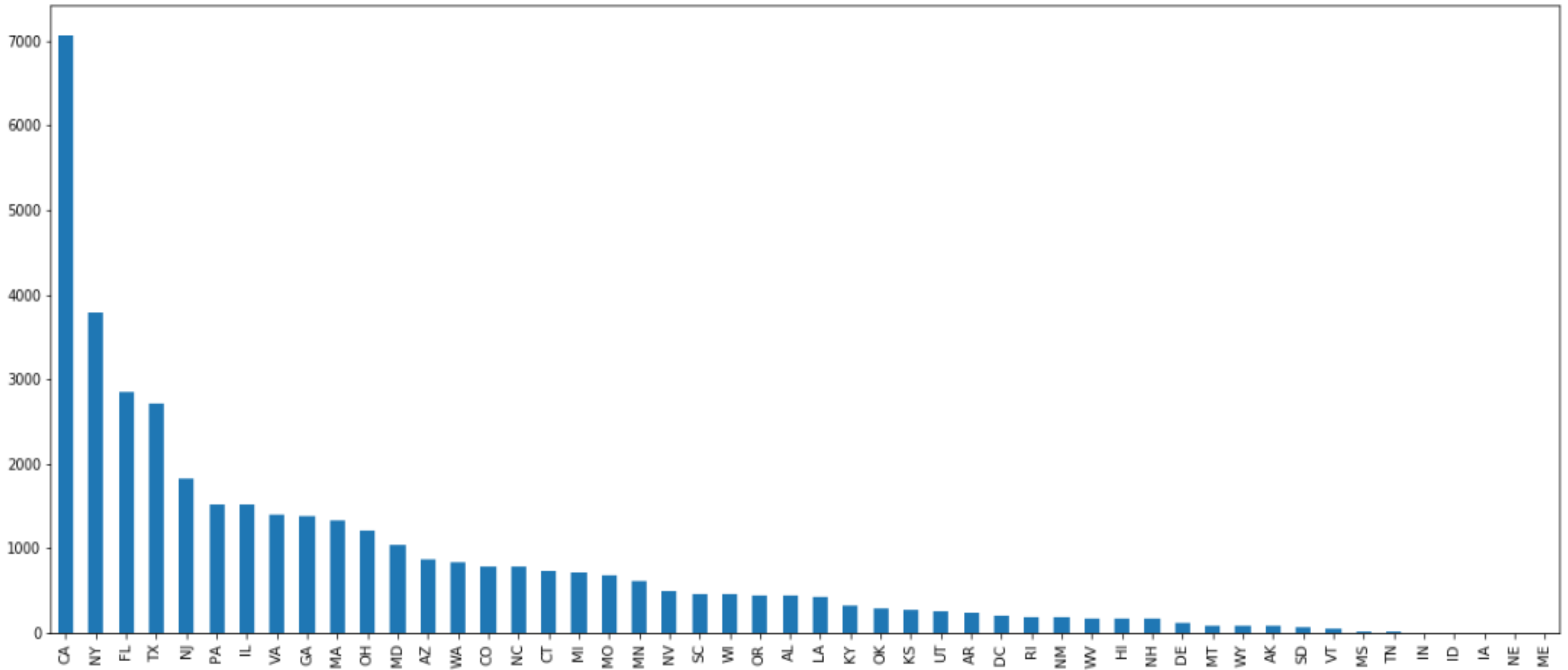
Inference:
People who has RENT or MORTGAGE takes loan



Inference:

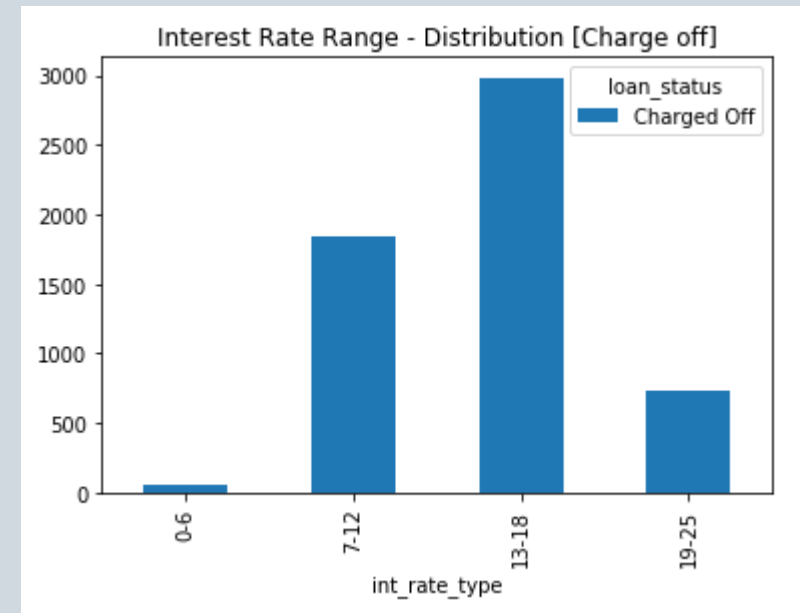
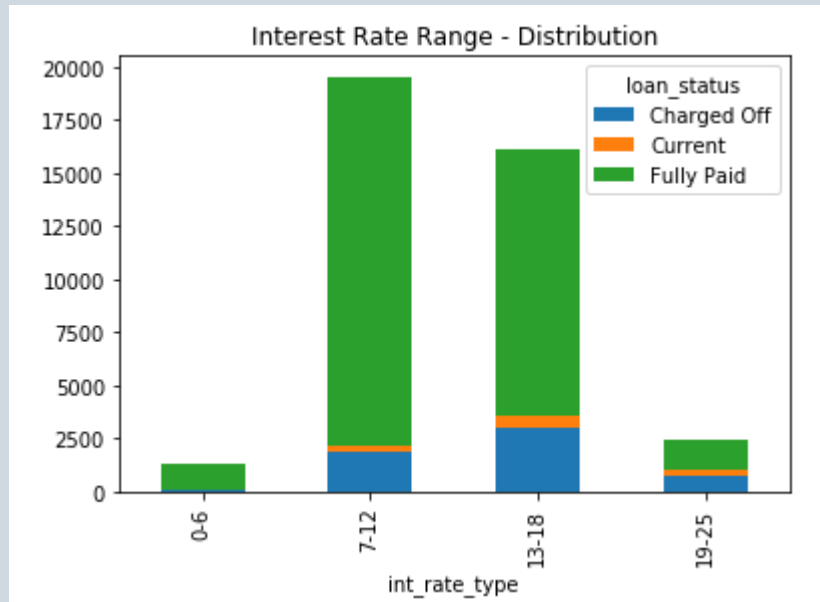
Pepole get loans more for the category 'debt_consolidation'

Address Distribution

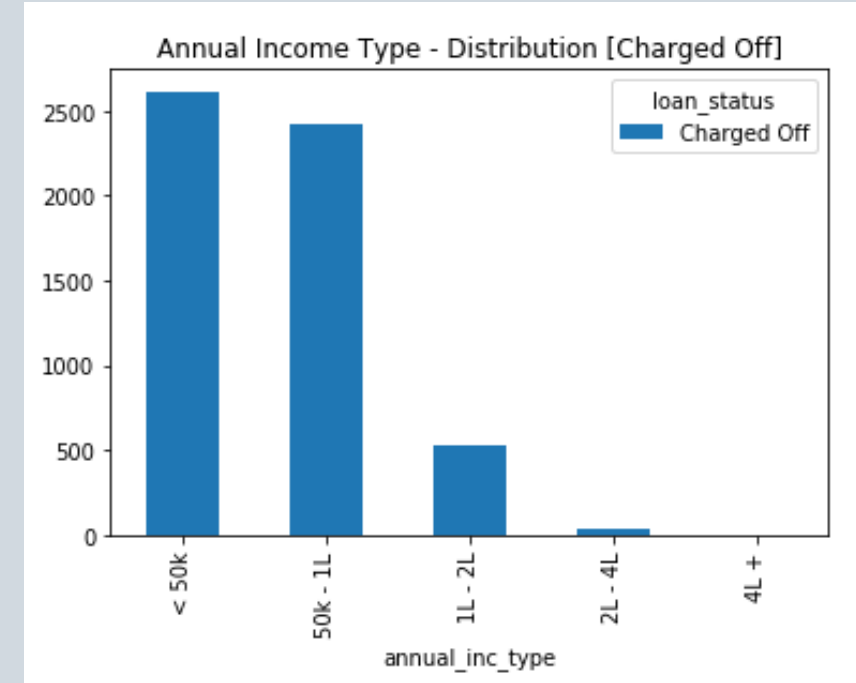
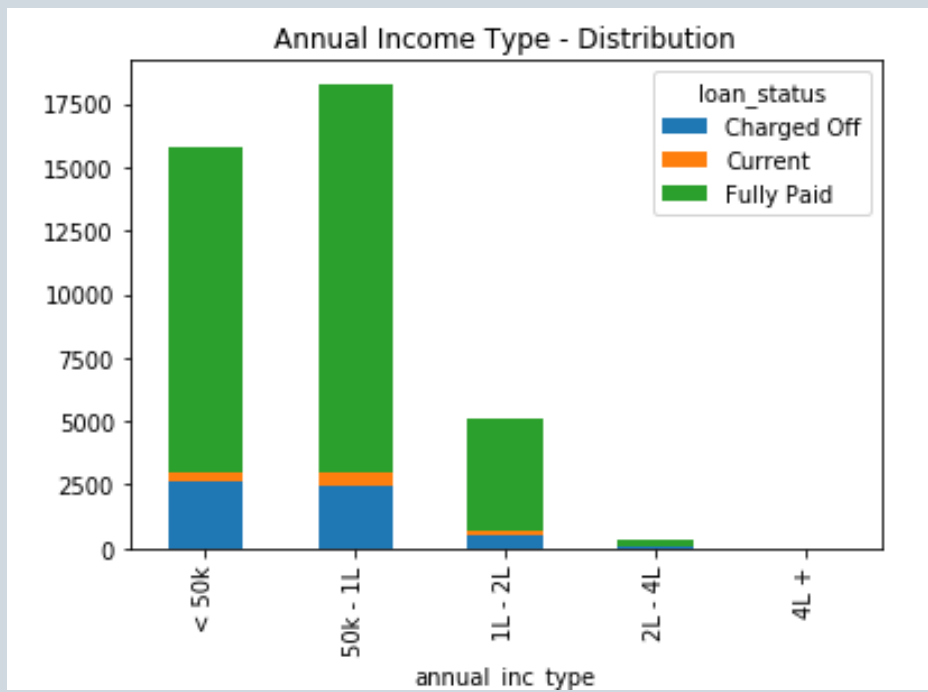


Inference:

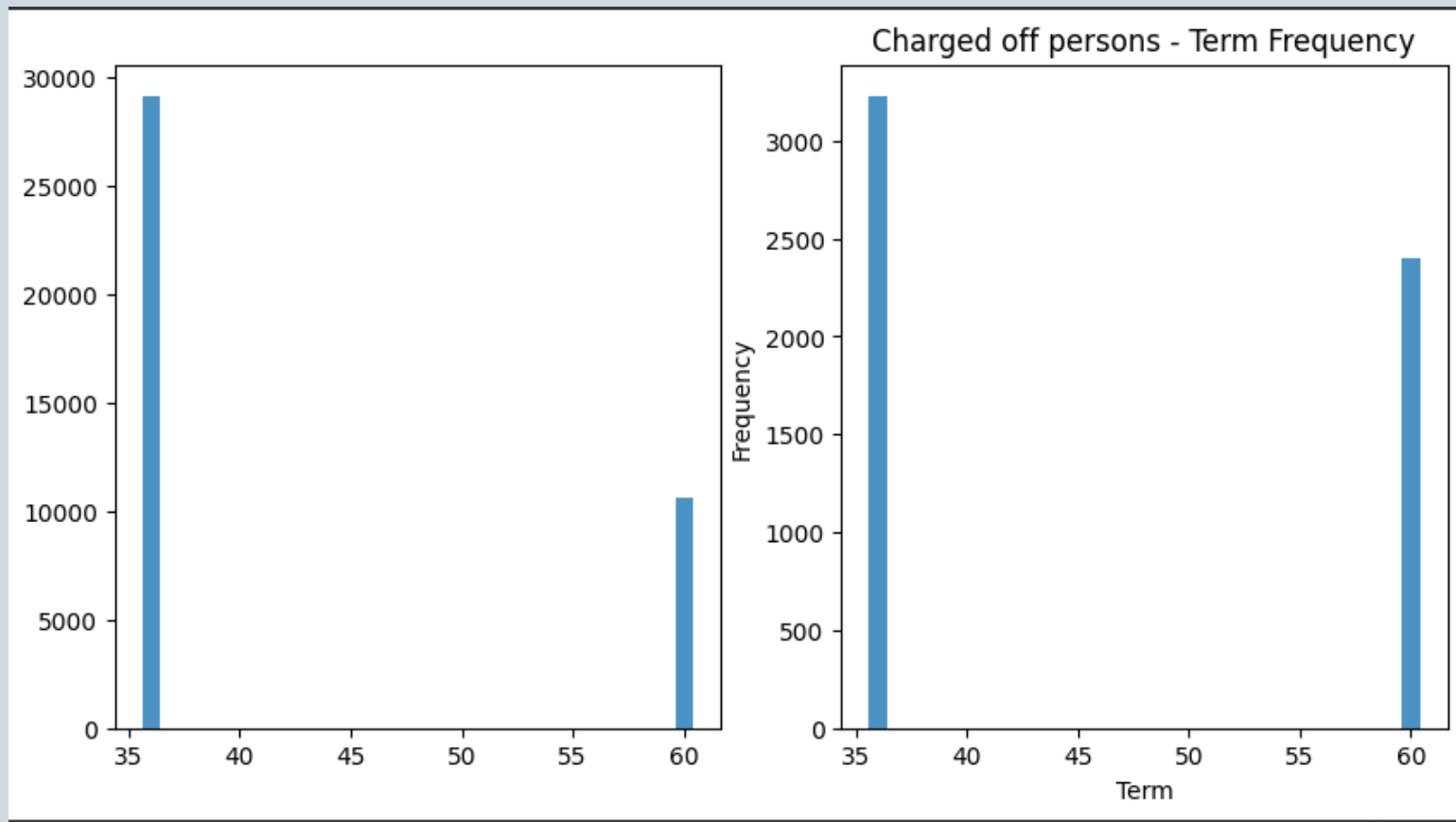
People in 'CA' have taken more loans



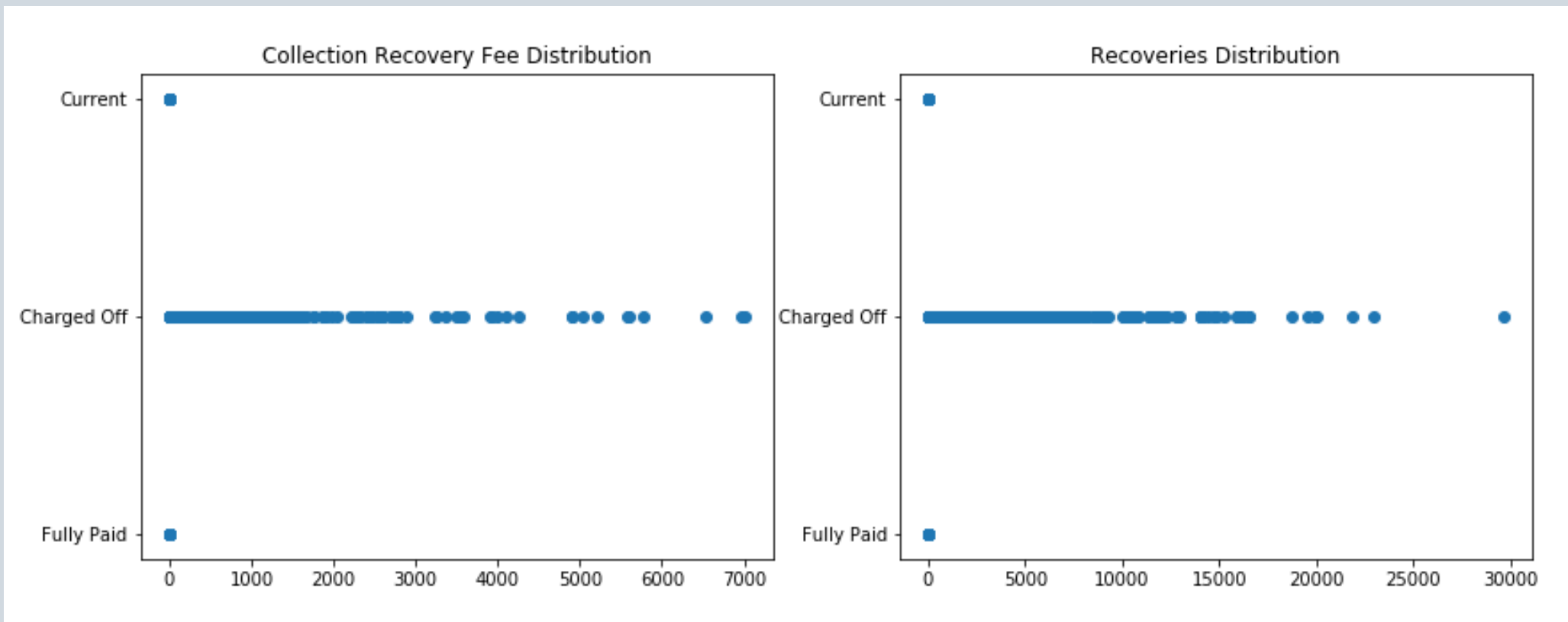
Inference on Rate of Interest (Univariate Analysis)



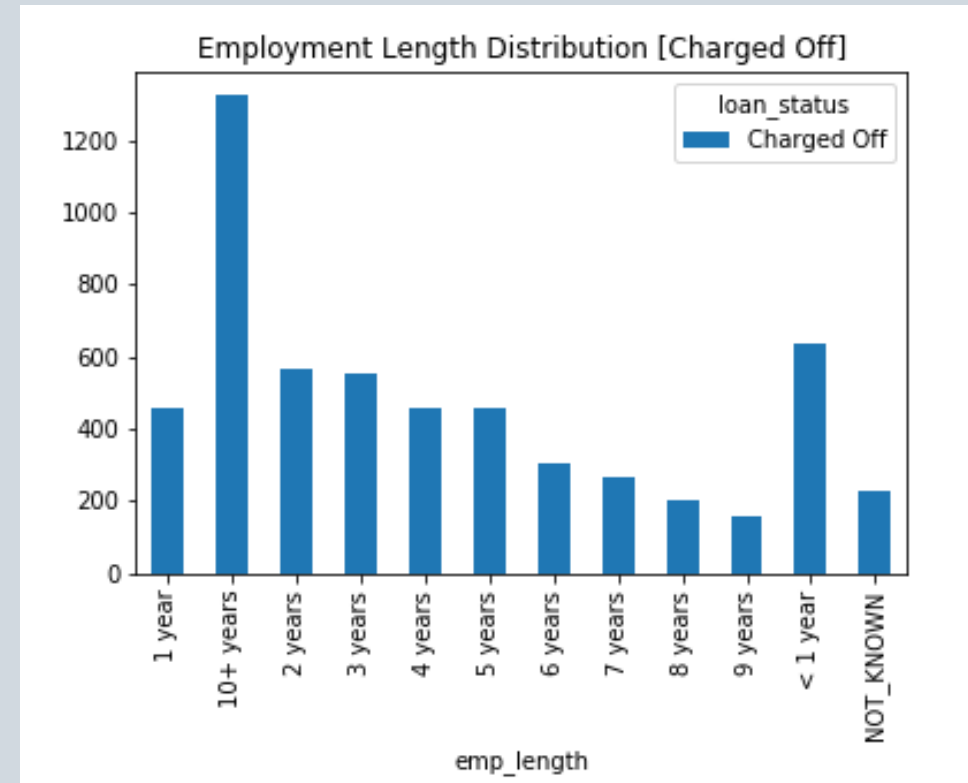
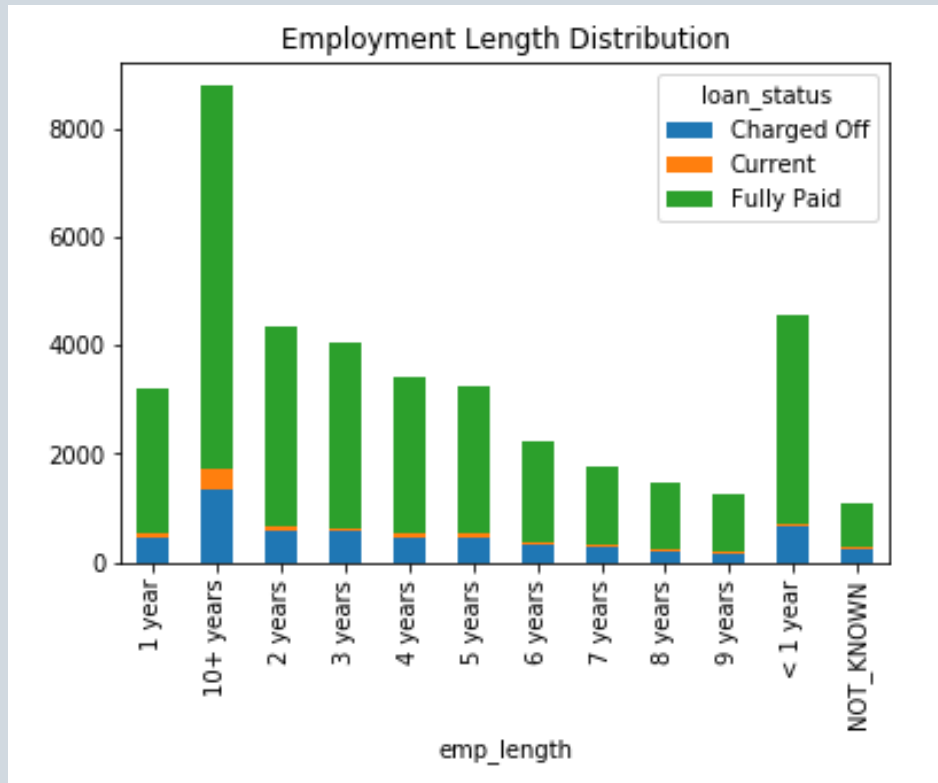
Inference on Annual Income (Univariate Analysis)



Inference on TERM (Univariate Analysis)

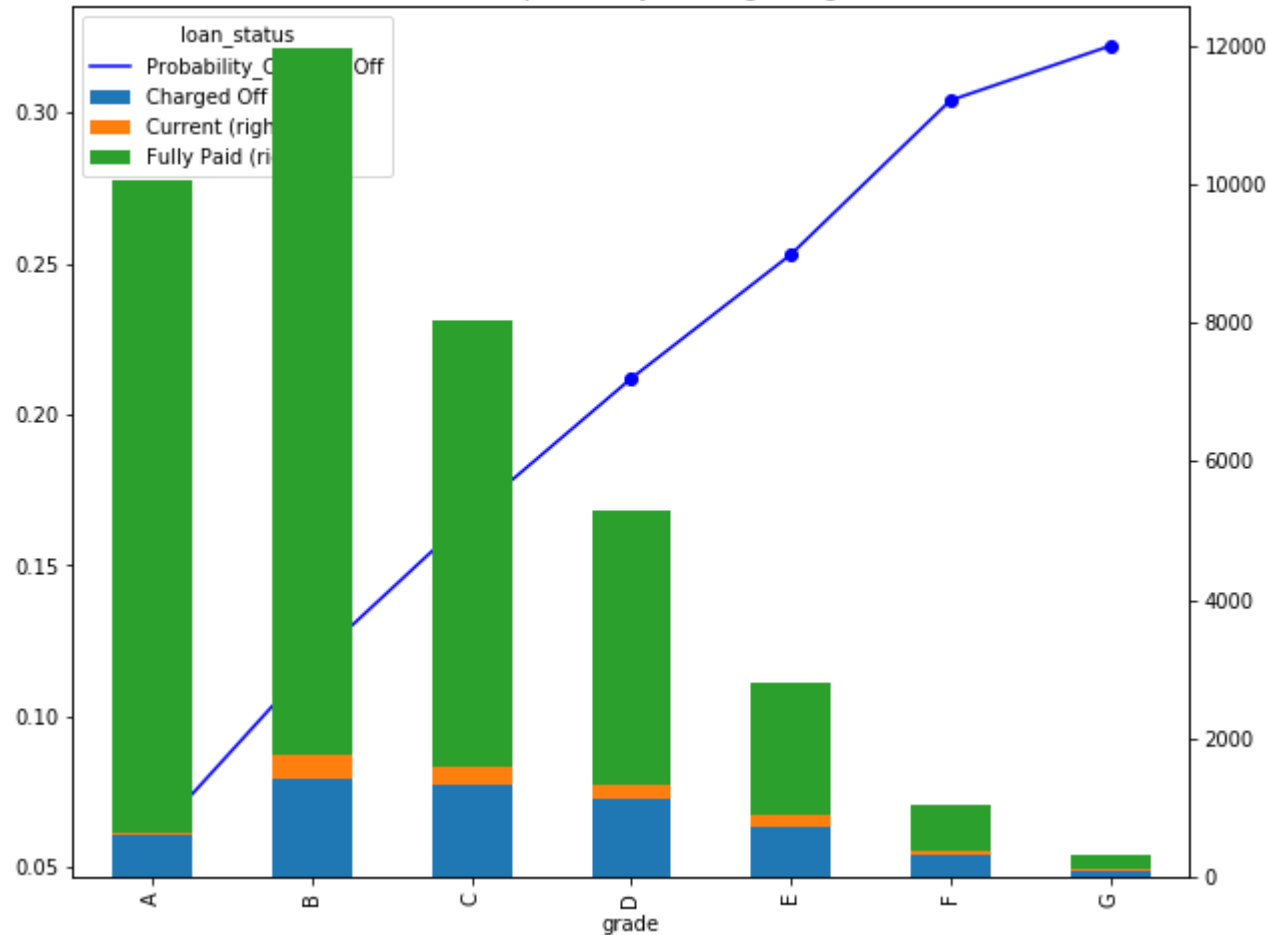


Inference on Collection Recovery Fee & Recoveries (Univariate Analysis)

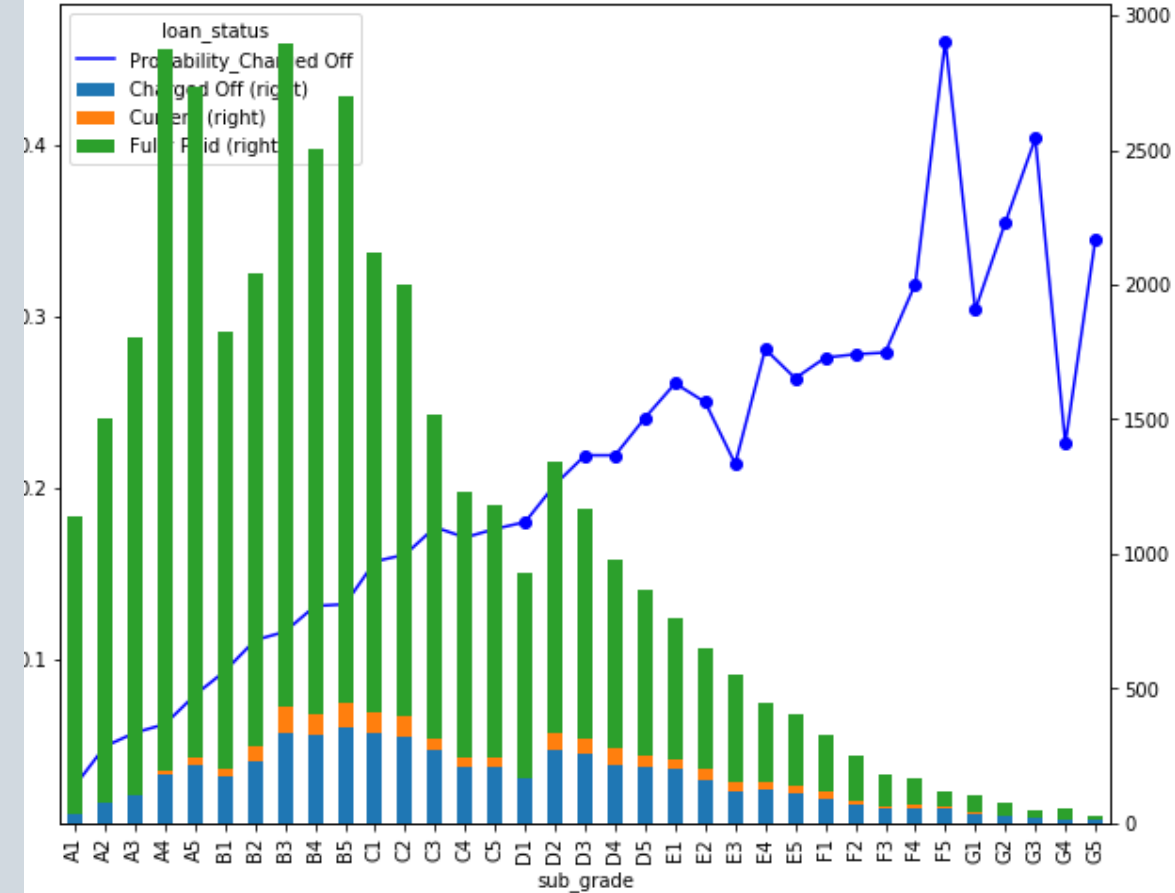


Inference on emp_length (Univariate Analysis)

Grade and its probability of being Charged Off



Sub Grade and its probability of being Charged Off



Inference - Grade/SubGrade (Bivariate Analysis):

Grade increases, the probability of charge off is also increased.