



Lending Club Case Study

SREENIVASA REDDY

THE CASE STUDY

We are collaborating with Lending Club, a finance company specializing in providing various types of loans to urban customers. When the company receives a loan application, it must decide whether to approve the loan based on the applicant's profile. Two types of risks are associated with the company's decision:

- If the applicant is likely to repay the loan, not approving the loan results in a loss of business for the company.
- If the applicant is likely to default on the loan, approving the loan may lead to a financial loss for the company.

Business Objective

The primary objective is to analyze past loan applicant data to identify patterns that predict the likelihood of default. By understanding these patterns through Exploratory Data Analysis (EDA), the company aims to make informed decisions such as denying loans, reducing loan amounts, or lending to risky applicants at higher interest rates. This analysis will help Lending Club minimize financial losses and optimize their loan approval process.

EDA & Feature Engineering

PRE-PROCESSING

| [Missing Values](#) | [Outliers](#) | [Normalization](#) | [Health Checks](#)

DATA UNDERSTANDING

| [Univariate, Bivariate and Multivariate Analysis](#) | [Exploration and Data Visualization](#)

POST-PROCESSING

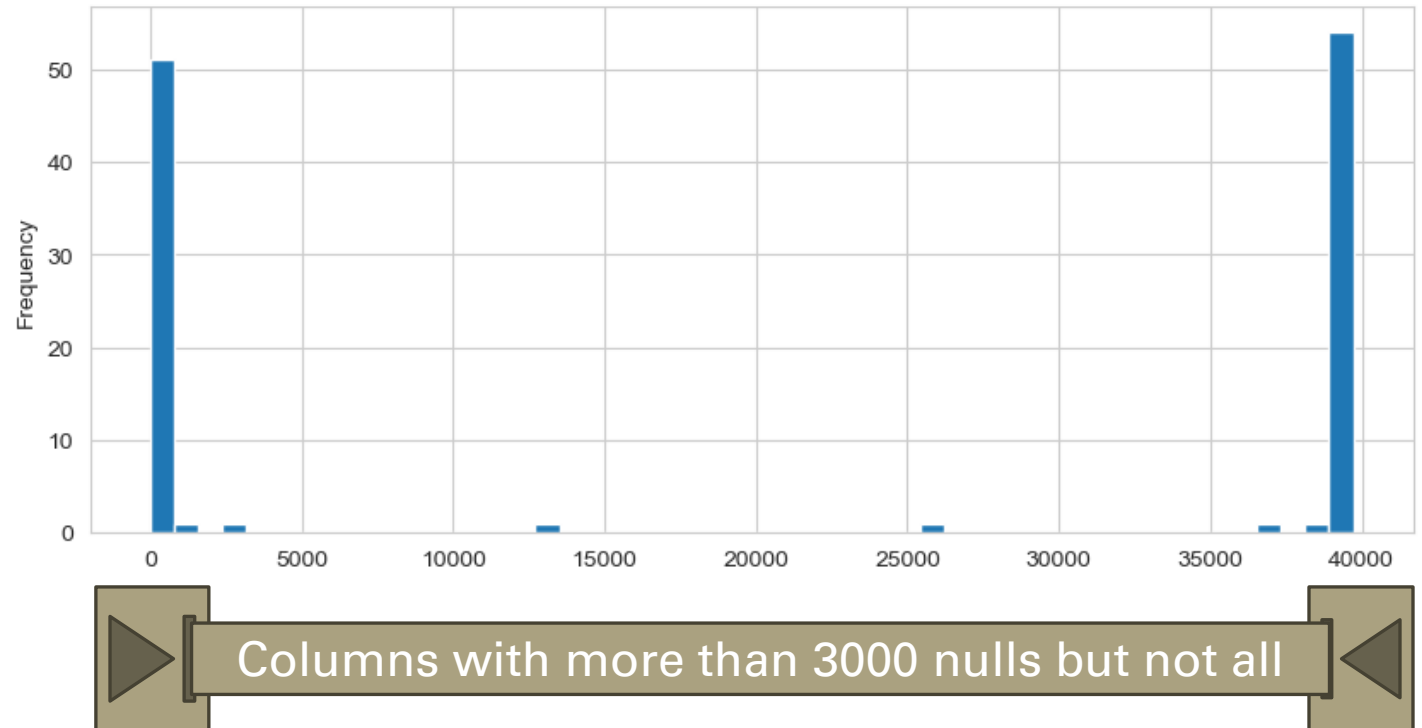
| [Collinearity](#) | [Outliers](#) | [Dimensionality Reduction](#)

DATA PREPROCESSING

Dataset has

- 39717 rows and
- 111 columns.

Distribution of Columns with Null and NaN values



DATA PREPROCESSING

Removed

- 24 columns Which are irrelevant to the context.

Columns Removed from List

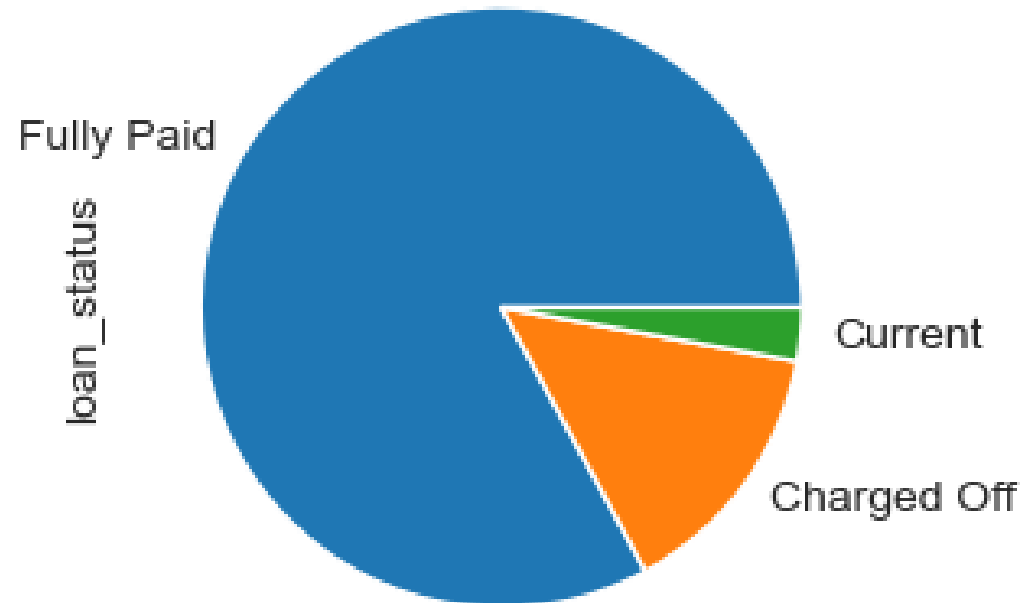
id	title	url	'chargeoff_within_12_mths	'zip_code	revol_bal
recoveries	out_prncp_inv	member_id	emp_title	total_rec_late_fee	
out_prncp	'last_credit_pull_d	total_pymnt	total_rec_int	tax_liens	
last_pymnt_amnt	'last_pymnt_d	total_rec_prncp	earliest_cr_line		
	'delinq_2yrs	collections_12_mths_ex_med			

DATA PREPROCESSING

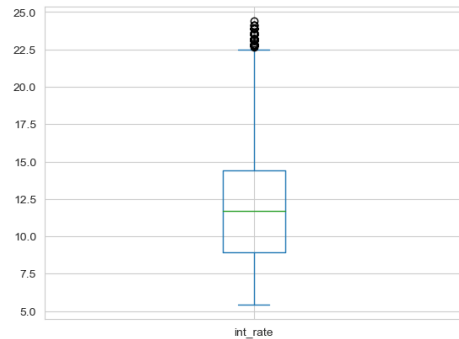
Rows

- 39717 to 38577 after removing current.

Remove rows w/ `loan_status` as `current`



Health Check of features



Outliers in *int_rate*

```
date and month issue in 'issue_d'

m, y = loan_df['issue_d'].str.split('-').str
loan_df['issue_y'] = pd.to_numeric(y)
loan_df['issue_m'] = m
loan_df.drop(columns=['issue_d'], inplace=True)
```

Format problem in *issue_d* and *int_rate*

```
% text issue in 'int_rate'

loan_df['int_rate'] = loan_df['int_rate'].apply(lambda x:
pd.to_numeric(x.split("%")[0]))
loan_df['int_rate'].plot.box()
```

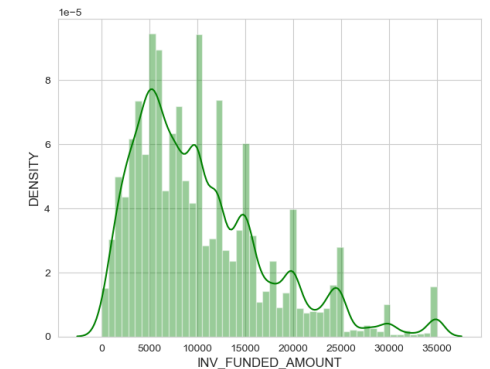
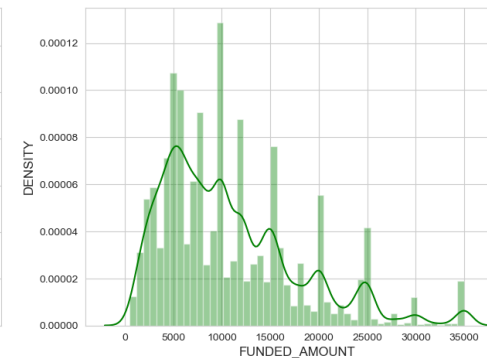
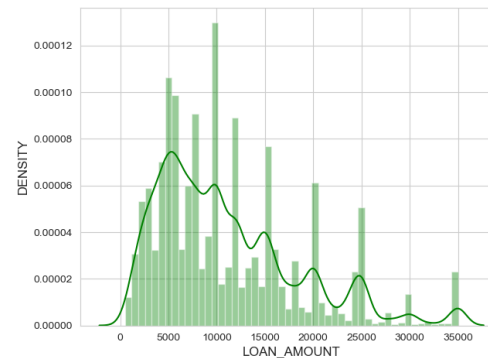

DATA UNDERSTANDING

Numerical Summary:

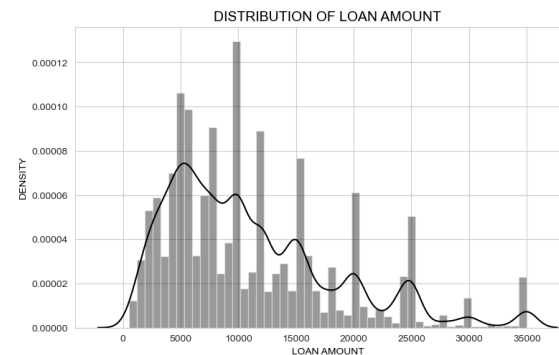
- Mean: 68777.97368120901
- Median: 58868.0
- 90% Data: 115000.0
- 99% Data: 234143.99999999878
- 100% Data: 6000000.0

UNIVARIATE ANALYSIS

- **loan_amnt** The listed amount of the loan applied for by the borrower. If at some point in time, the credit department reduces the loan amount, then it will be reflected in this value.
- **funded_amnt** The total amount committed to that loan at that point in time.
- **funded_amnt_inv** The total amount committed by investors for that loan at that point in time.



Density distribution for all the three features are similar

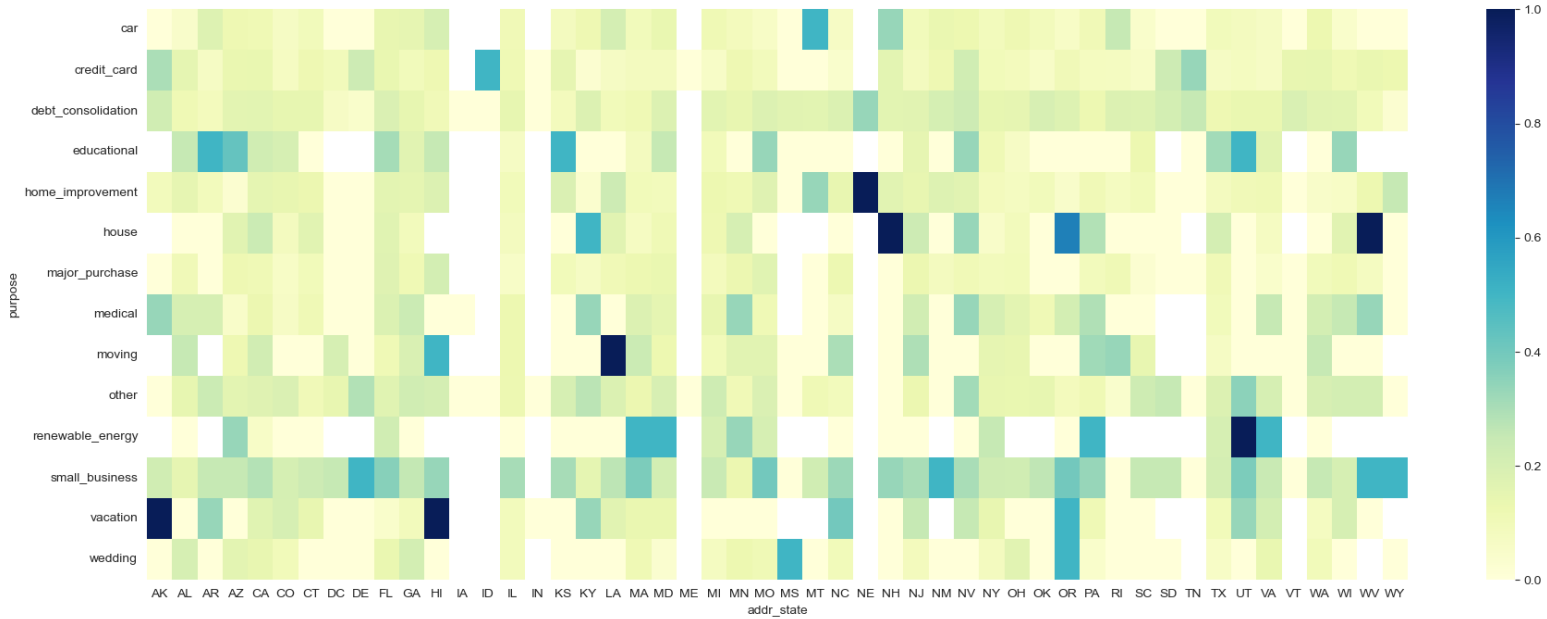


DATA UNDERSTANDING

Bivariate Categorical:

- **vacation loans in AK, HI, OR**
- **education loans in AR, KS, UT**
- **small business loans in DE, NM, WV, WY**
- **wedding loans in MS, OR**

BIVARIATE ANALYSIS

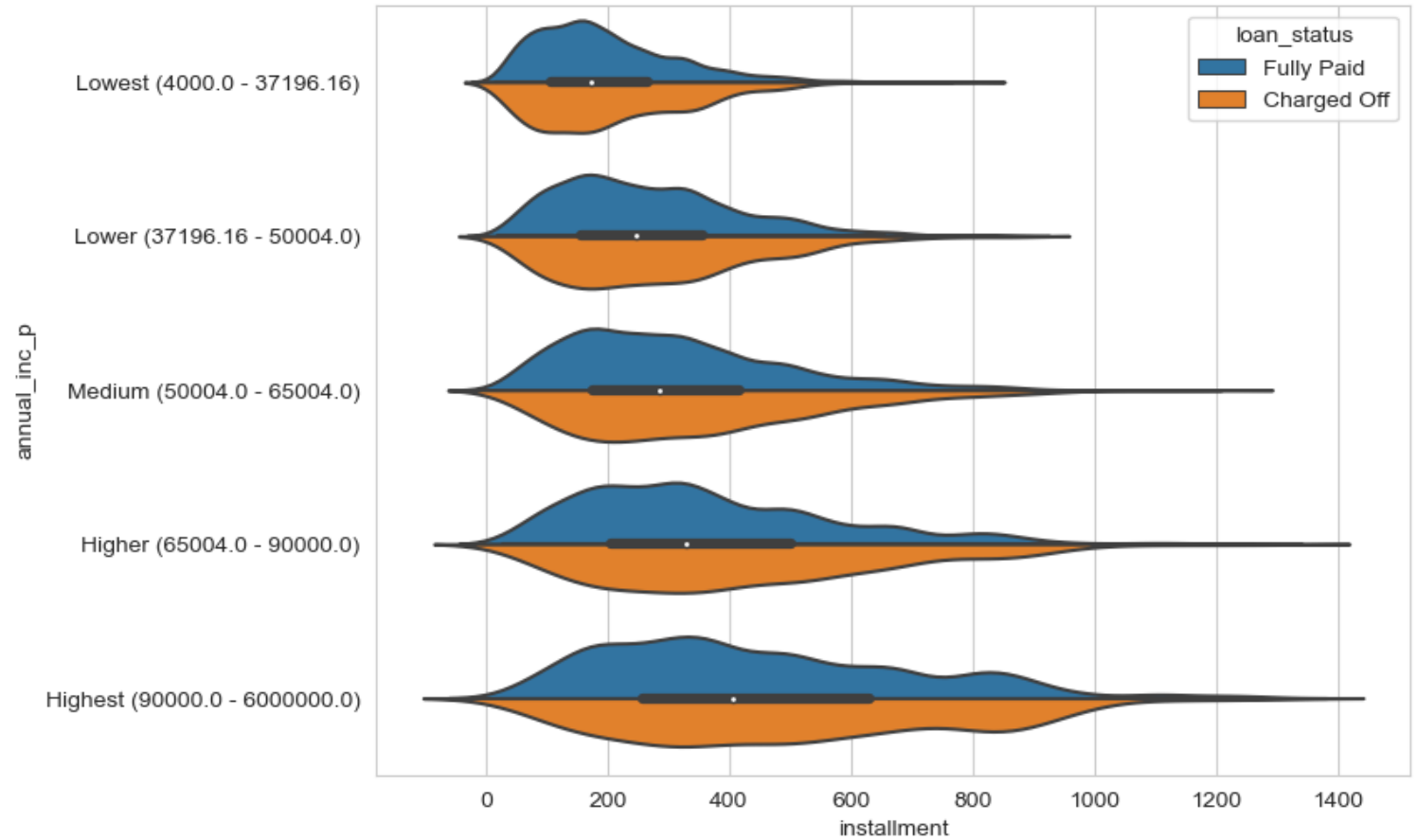


DATA UNDERSTANDING

Bivariate Analysis:

variables and their impact on the loan-status

BIVARIATE ANALYSIS



Bivariate Analysis:

- # MULTIVARIATE ANALYSIS

