LendingClubCase Study

MLC33-Sept 2021 Batch

-Sanjay Chhetri

-Mamatha Shetty

Problem Statement

1. In this Case study, we have to analyse the loan lending dataset and analyse the below factors:
   - To identify the risky customers based on different factors.
   - Identify the the variables which are strong indicators for the customers to be defaulters.
   - Understand the pattern of Customers getting charged off.

   Conclude based on analysis the and provide recommendations.

Data Understanding

1. Load the Csv File to Python DataFrame.
2. Print the head, tail to understand the data.
   - Head returns top 5row and 111 columns
   - Tail returns bottom 5 rows and 111 columns

   Understand the columns included in the data and its description
4. Print shape, description and info to get a better understanding of datatypes, number of rows number of columns etc.
   - Shape() provides results as (39717, 111). i.e. 39717 rows and 111 columns
   - Describe gives details detailed information for numeric columns
5. Analyze the data which are not important and are required to be cleaned up as part of Data Cleaning activity.

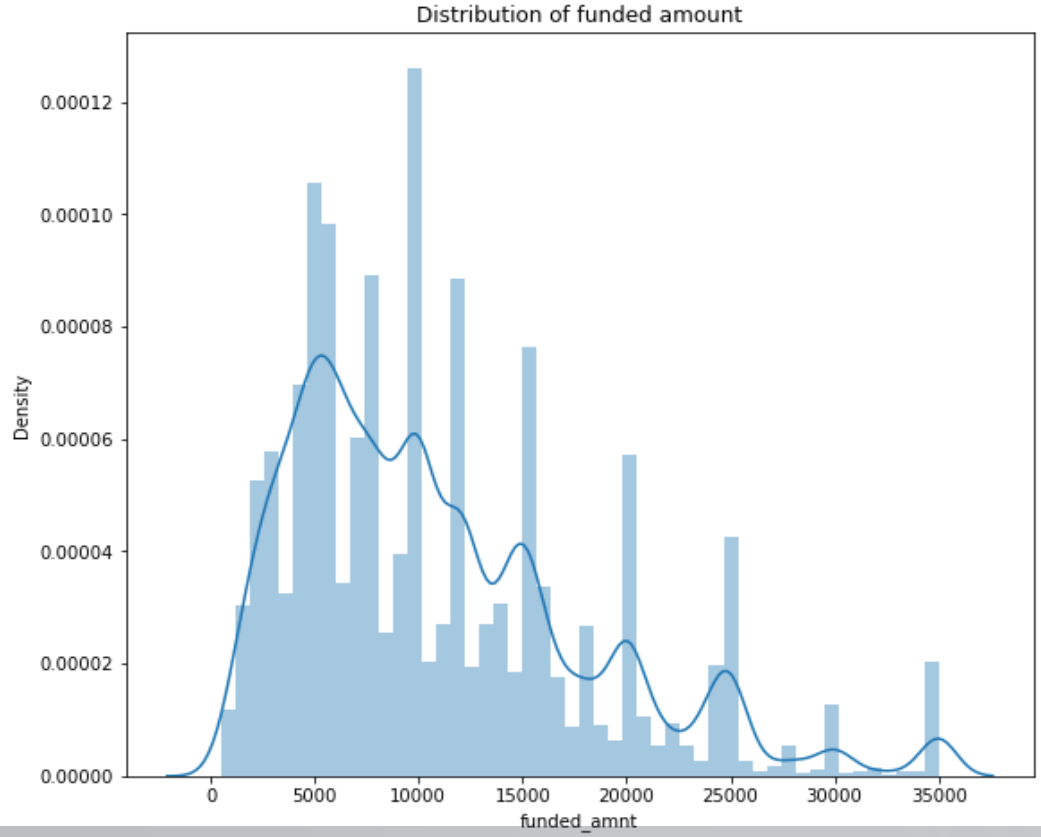
Data Cleaning and Manipulations

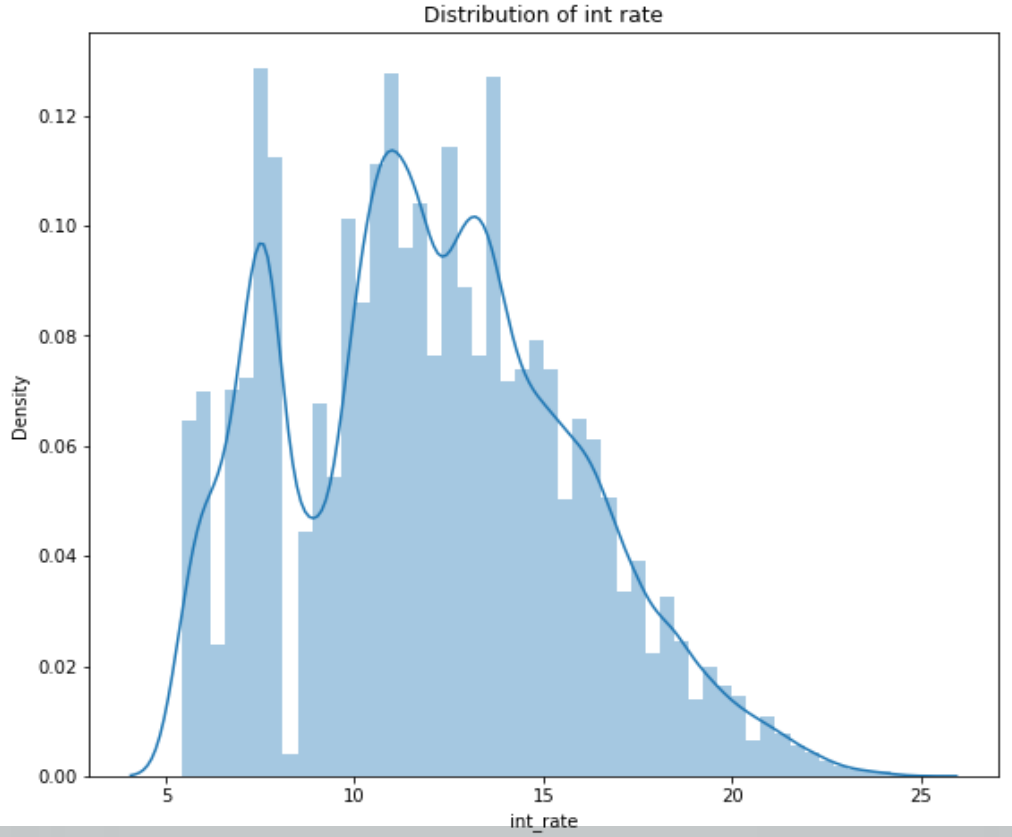
1. Based on analysis, all columns after index 49 seems to add no values to the case. Hence drop them.
2. Also drop **pymnt_plan** and **url** column as it has only duplicate data
3. Treat the data and manipulate as required.
    - **int_rate** : Convert string to float

      emp_length : Fill the missing values.i.e. fill 'n/a' values with 0

Data Analysis : Univariate

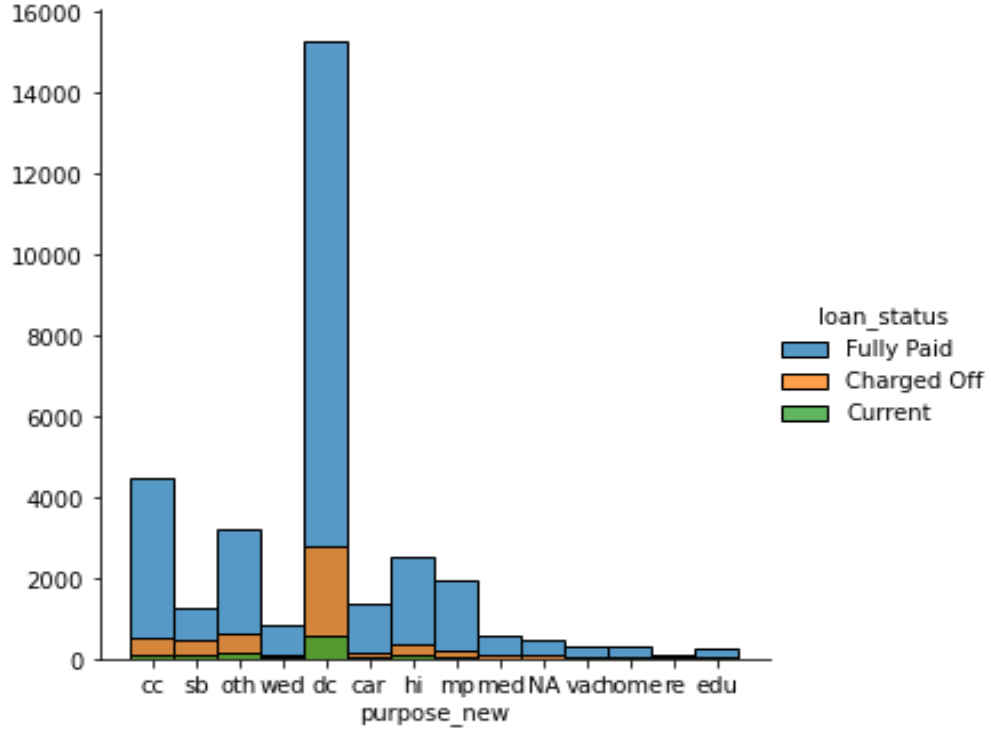
1. Funded Amount : Plot the distribution plot of **funded_amnt** and **int_rate**
   - *most of the funded amount range between 5000 to 12000 and no.of customers are reducing as the amount increases*
   - *Most of the loan int range is spread between 8 to 14%*





Data Analysis : Univariate

2. Purpose : Percentage Charged off on **Purpose**
   - *As per the analysis on purpose of the loan taken, we can clearly see that small business has the highest percentage of Charged off followed by renewable energy.*
   - *Charged off percentage for **small business** is **25.9%**.*
   - *Charged off percentage for **renewable energy** is **18.4%**.*

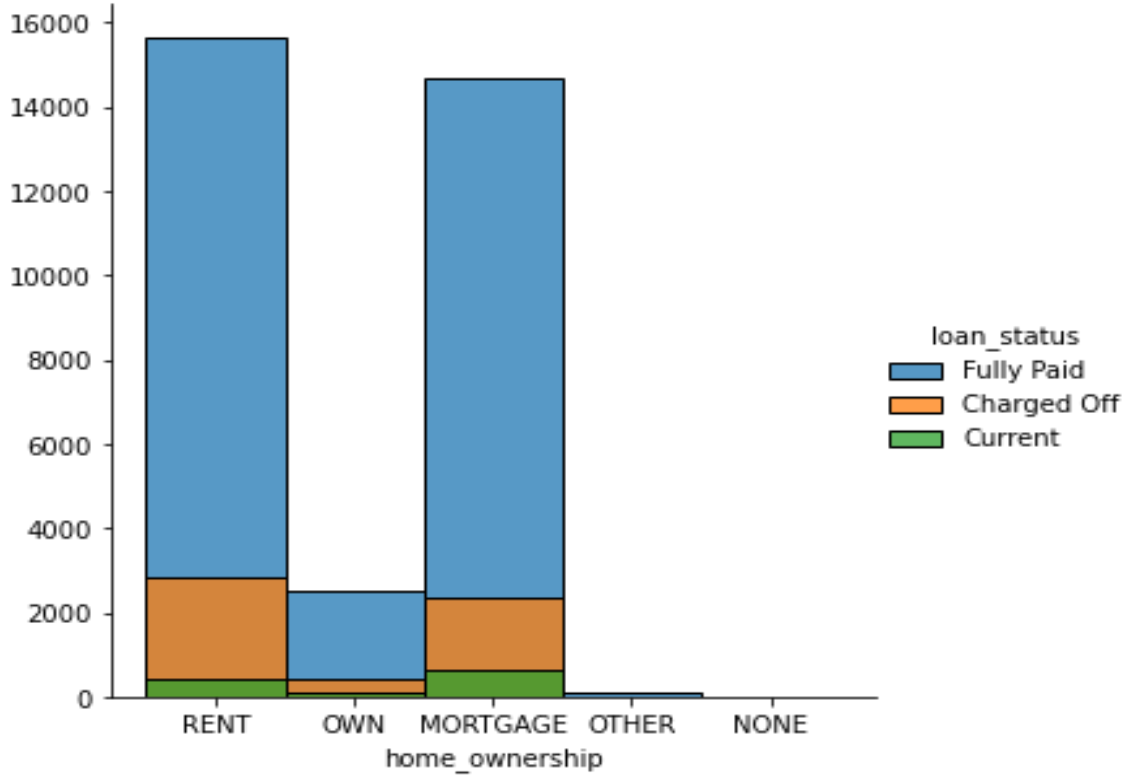


loan_status	Cu	T	per_chargedoff	per_fullyPaid	per_Current
purpose					
small_business	74.0	1828.0	25.984683	69.967177	4.048140
renewable_energy	1.0	103.0	18.446602	80.582524	0.970874
educational	0.0	325.0	17.230769	82.769231	0.000000
other	128.0	3993.0	15.852742	80.941648	3.205610

Data Analysis : Univariate

3. Home Ownership : Percentage Charged off on **Home Ownership**

- *As per the results table and histogram below, loans provided with home ownership as **'Other'** has highest percentage of charge off followed by home ownership as "**Rent**".*
- *Charged off percentage for loans provided with home ownership as **'Other' is 18%.***
- *Charged off percentage for loans provided with home ownership as **Rent - 15%.***



	Cu	T	per_chargedoff	per_fullyPaid	per_Current
loan_status					
home_ownership					
OTHER	0.0	98.0	18.367347	81.632653	0.000000
RENT	419.0	18899.0	15.021959	82.760993	2.217049
OWN	83.0	3058.0	14.486593	82.799215	2.714192
MORTGAGE	638.0	17659.0	13.177417	83.209695	3.612889
NONE	0.0	3.0	0.000000	100.000000	0.000000

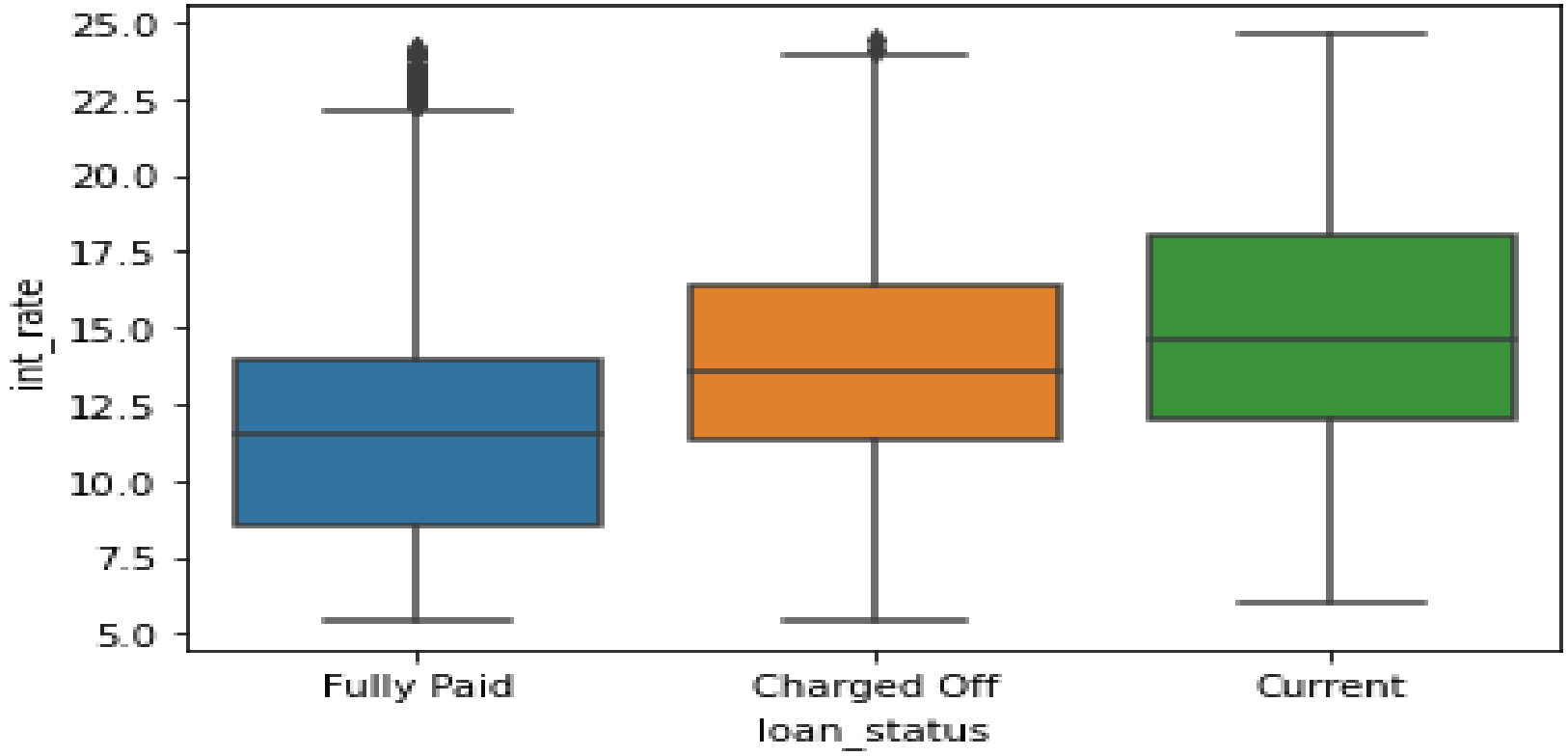
Data Analysis : Univariate

4. Home Ownership : Percentage Charged off on **Employment Length**
   - *As per the results table below, loans provided with employment length with **10+ years** has highest percentage of charge off followed by employment length with **7years** .*
   - *Charged off percentage for loans provided with employment length with **10+ years is 14.9%.***
   - *Charged off percentage for loans provided with employment length with **7years is 14.8%.***

loan_status	per_chargedoff	per_fullyPaid	per_Current
emp_length			
10+ years	14.990427	80.605924	4.403649
7 years	14.833615	81.669487	3.496898
1 year	14.074074	83.734568	2.191358
5 years	13.954906	83.363803	2.681292
< 1 year	13.942832	84.420685	1.636483
6 years	13.772992	83.490354	2.736653
8 years	13.725490	83.299527	2.974983

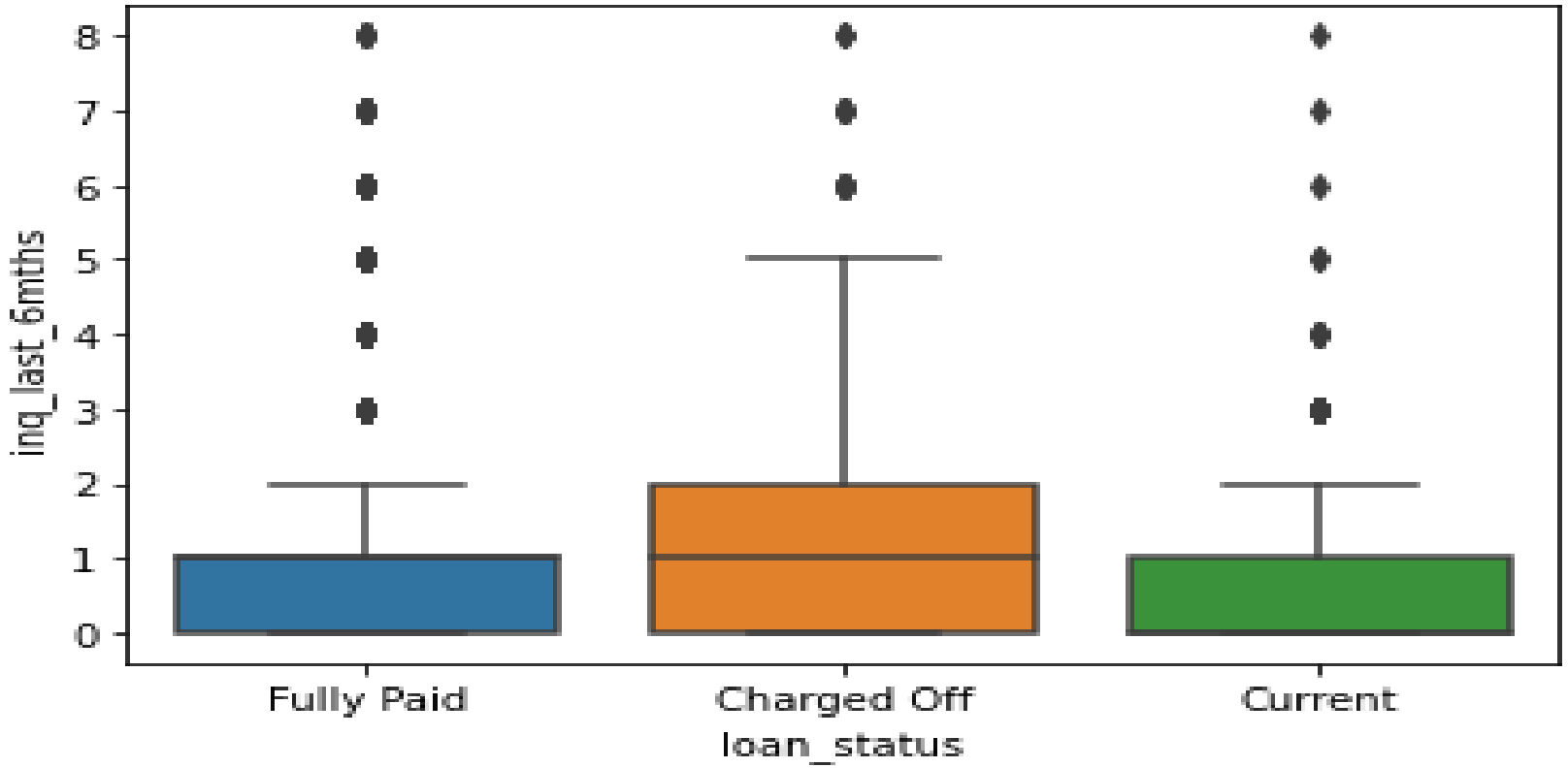
Data Analysis : Bivariate

1. Box plot : Bivariate Analysis on **interest rate** and **loan status**
   - *Based on the box plot below, customers with interest rate **higher than 12** are more likely to be charged off.*
   - *As we can see, there are no customers who have fully paid the loan with interest rate higher than 12, though we have few loans with current status.*



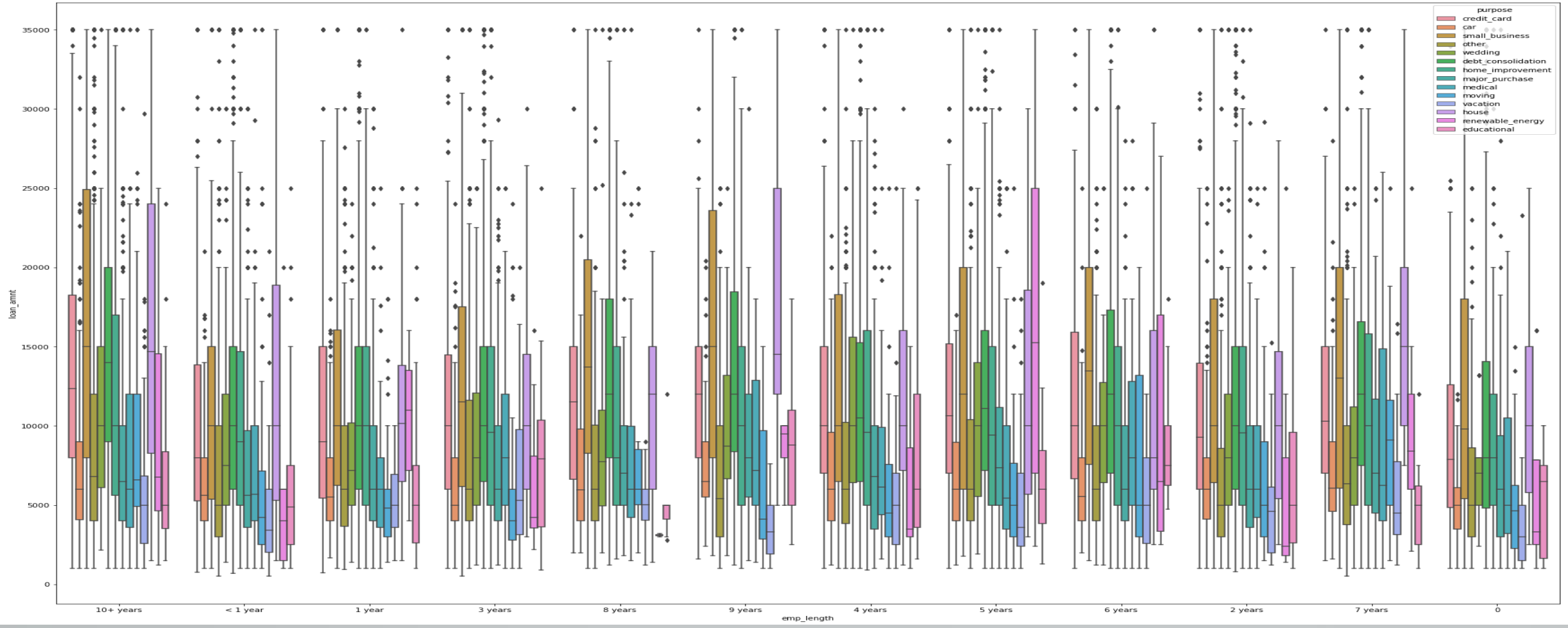
Data Analysis : Bivariate

2. Box plot : Bivariate Analysis on **inquiry last 6 months** and **loan status**
   - *Based on the box plot below, customers with enquiry **more than once** in last 6 months more likely to be charged off.*
   - *There are no current loan customers with enquiry more than 1 since last 6 months.*



Data Analysis : Multivariate

1. Multivariate analysis on **emp_length, loan_amnt, purpose**
   - *Based on the graph boxplot, we can see small business loan amount is higher followed by renewable energy and house.*



Modelling

1. Target Variable:
   - Loan_Status
2. Decision Making Variable:
   - Purpose
   - Interest Rate
   - Home Ownership

   Inquiry last 6 months

   Employment Length

Conclusion

Based on analysis, we can conclude that:

- Funded **amount ranging between 5000 to 12000** has the most no. of customers and the customer decreases as the funded amount increases
- Customer with **higher interest rates**(>12) are more likely to be charged off.
- Customer **inquiring more than once in last 6 months** are more likely to be charged off.

  Customer whose **purpose of loan is "small business"** is most likely to be charged off
- Customer with **home ownership with "Other" and "RENT"** are more likely to be charged off.

Thank You…