

# Lending Club case Study

By – Ribha Singh



## Introduction

This case study is part of course “upGrad and IITB Machine Learning & AI”

This is an EDA based case study to get the output to help business of a company called Lending Club

## Business Understanding

- ❑ Lending Club is a "consumer finance company" that specializes in lending various types of loans to urban customers.
- ❑ The business must decide whether to approve or reject a loan application after receiving it.
- ❑ The purpose of lending money to borrowers is to benefit from interest.
- ❑ In order to determine the variables causing defaults, the company gave the analyst access to a history of borrowers and their prior records (in loan.csv).
- ❑ This allowed the company to identify loan applications that were hazardous and make appropriate selections.
- ❑ The focus column is loan status, where we have details about default or not.

## Business Objectives

- ❑ In the given data set borrowers who default cause the largest amount of loss to the lenders. In this case, the customers labelled as 'charged-off' are the 'defaulters'
- ❑ Identification of such applicants who are likely to be defaulters using EDA is the aim of this case study.
- ❑ The Lending Club company wants to understand the driving factors (or driver variables) behind loan default.

## Data Understanding

Approach taken to understand data are below.

- ❑ Understand the all information including metadata, like no of row, columns, Data types with the help of provided data dictionary and Python/ NumPy/Pandas syntax.

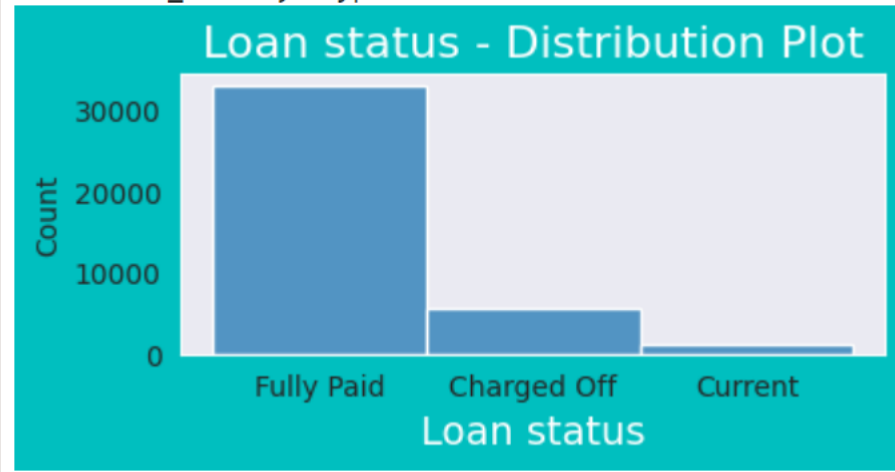
## Data Cleaning and Manipulation

For data cleaning, the below approach has been taken.

- ❑ Identify whether the columns have a null value or not.
- ❑ Based on the result, remove the column with >30% null, as the columns with less data are the descriptive data, which won't help in analysis.
- ❑ Check the columns having the same value and remove those that don't have any impact on the analysis.
- ❑ Find one column to act as a key and remove a column with a unique value, like title.
- ❑ Standardized the data for the column having symbols like %
- ❑ The final dataset is ready to start the analysis.

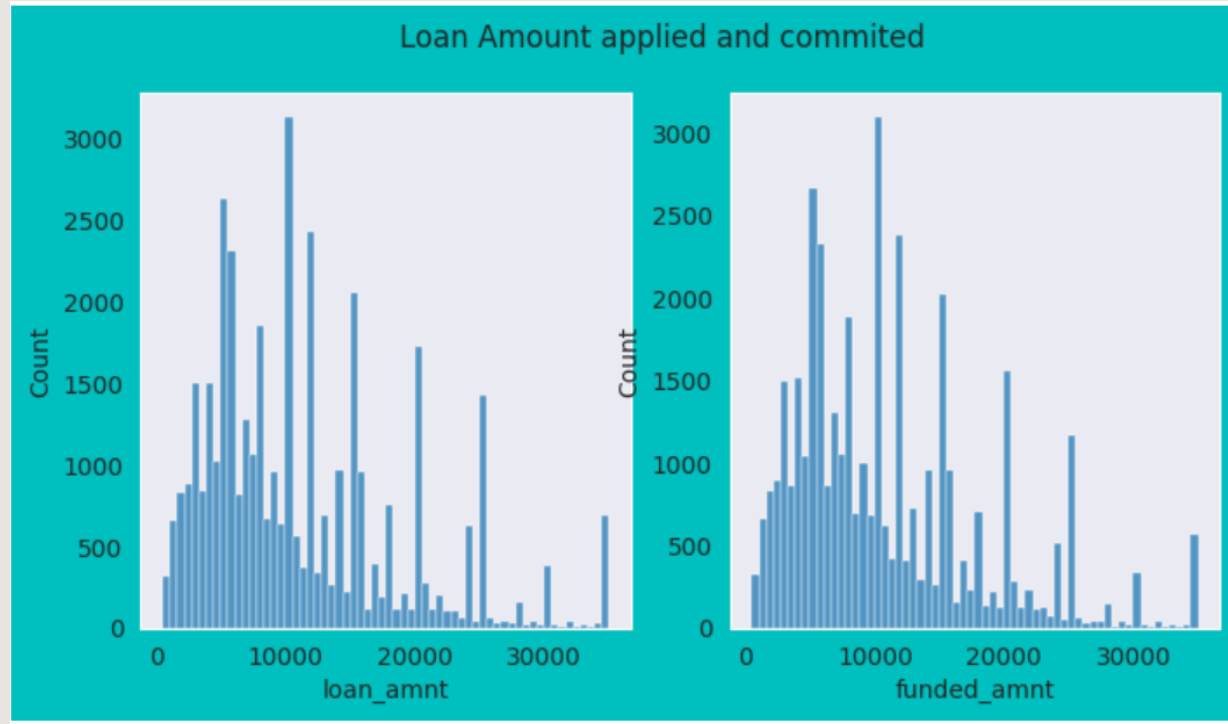
## Data Analysis – Univariate Analysis

```
Fully Paid    32950  
Charged Off   5627  
Current       1140  
Name: loan_status, dtype: int64
```



- 20-25% Borrowers are defaulters.
- So, depth analysis is required to find the defaulter factors.

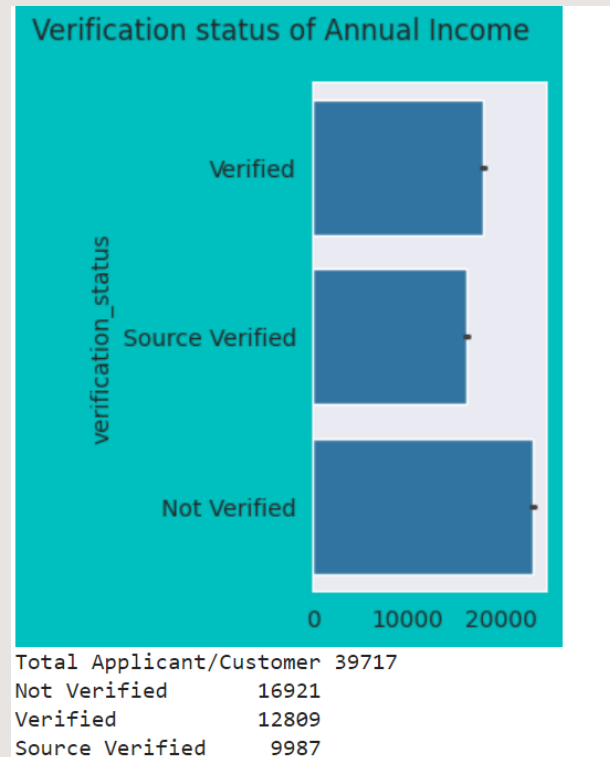
## Data Analysis – Univariate Analysis



- From data it looks like mostly approved borrowers got the requested amount.
- Hence, verification status needs to be checked whether amount has been approved without verification or not.

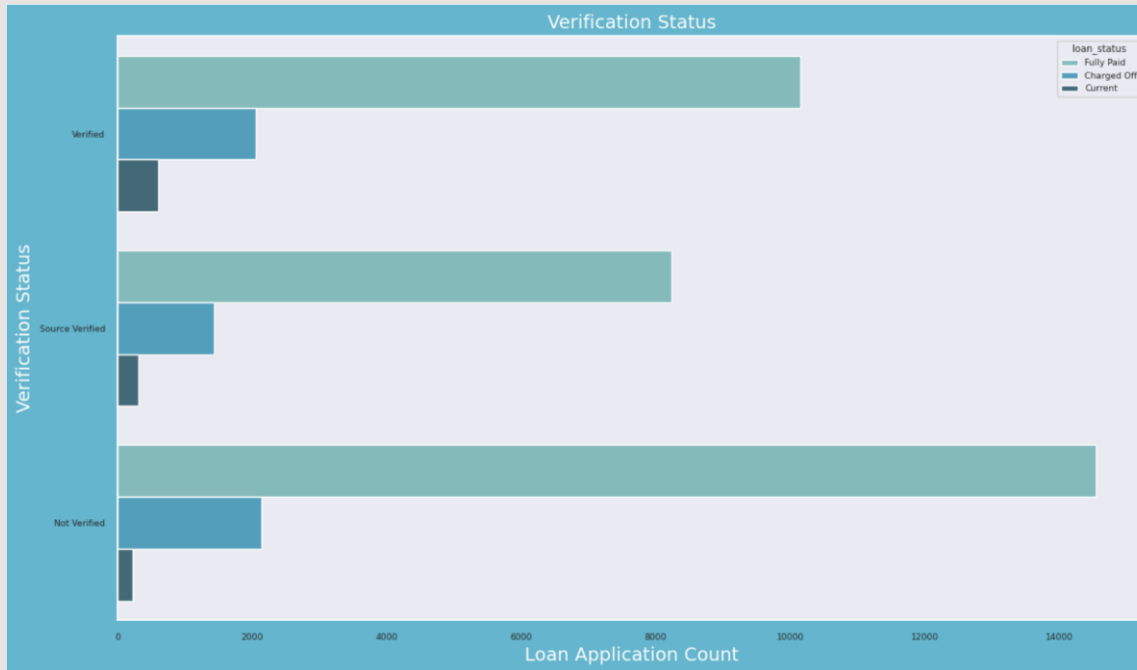


## Data Analysis – Univariate Analysis



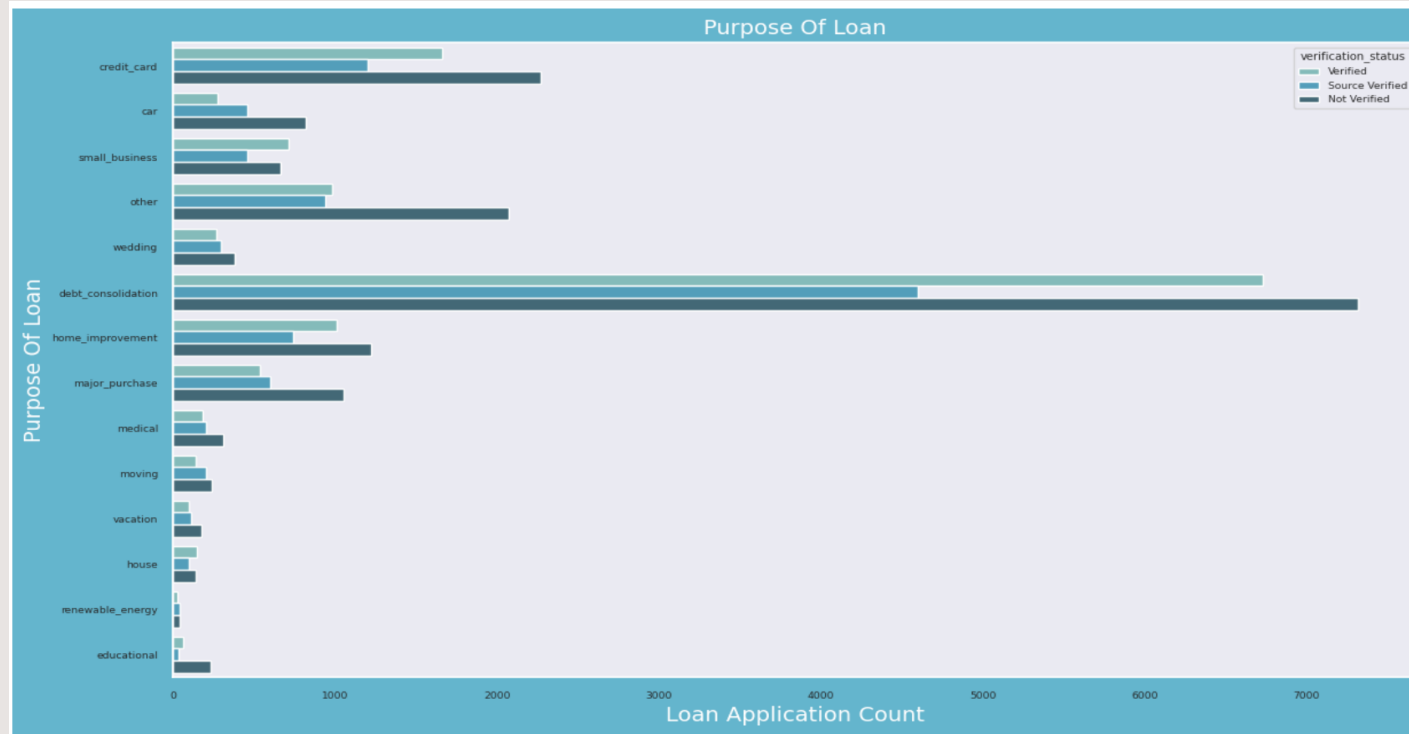
- From visual and count its clear that approx. 42% Borrower's annual income is not verified.
- So next step needs to check the other column with respect to verification status as the Annual Income' verification is a suspected factor.

## Data Analysis – Univariate Analysis



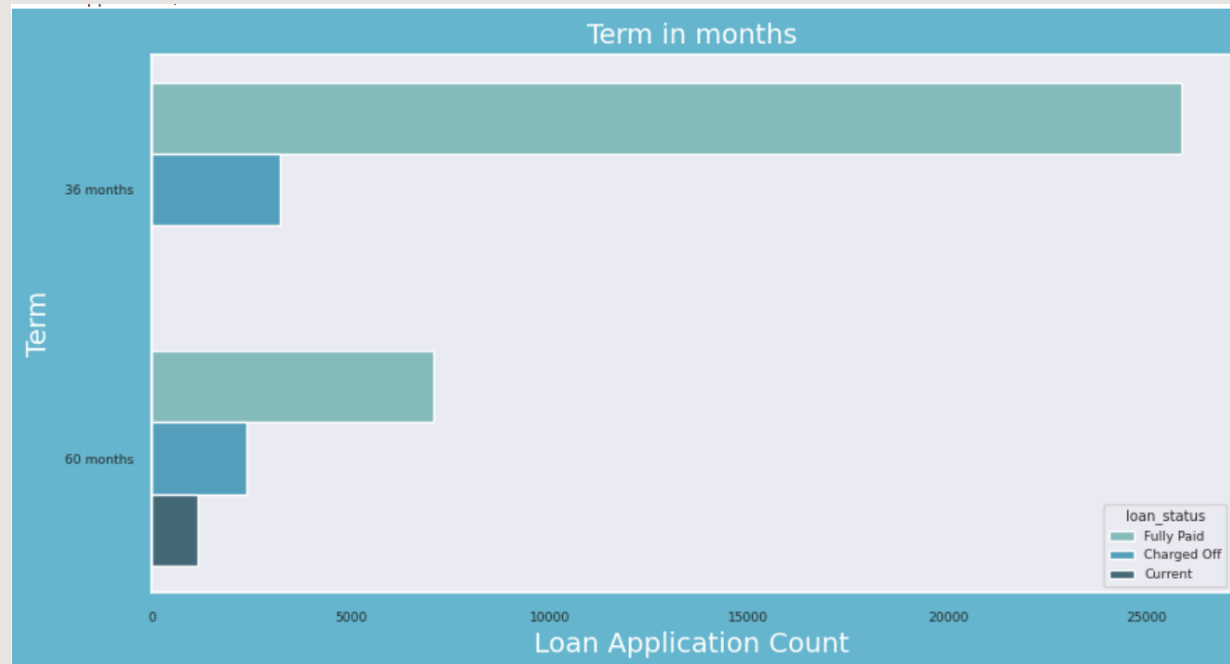
- Charged Off is highest under not verified category but not much difference.
- As verified also have Charged off, Hence need to Annual Income in bivariate analysis also.

## Data Analysis – Univariate Analysis



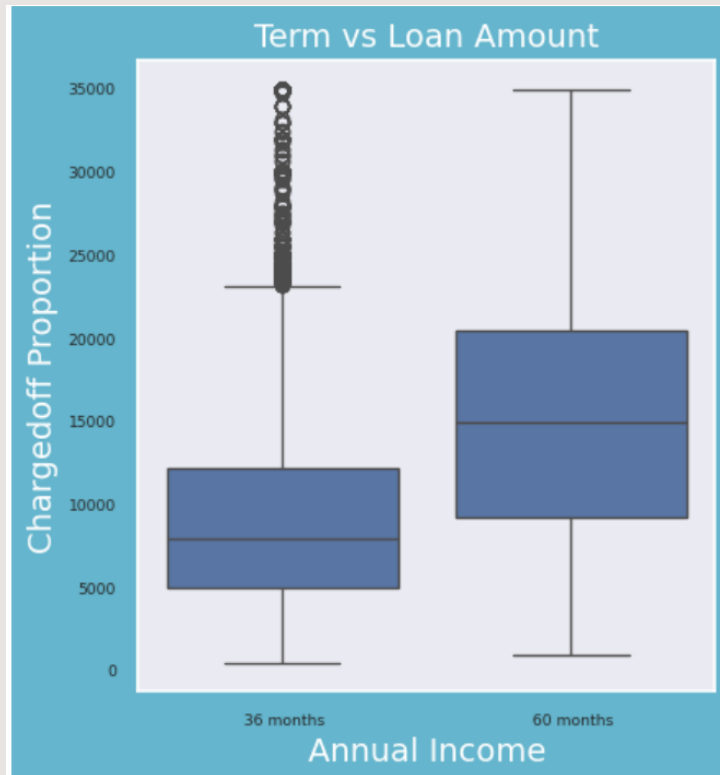
- The main purpose of borrower is to clear the debt as bar is showing high value for "debt\_consolidation" where source not verified is with high value.

## Data Analysis – Univariate Analysis



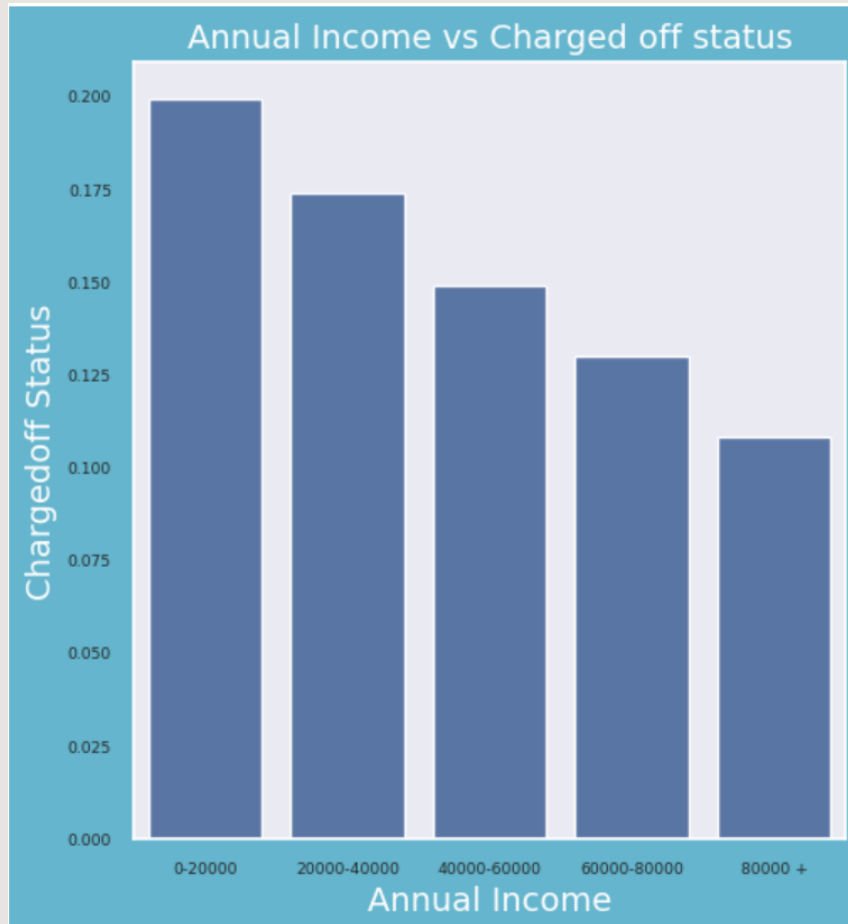
- Short duration having high chances of charge off.

## Data Analysis – Segmented Univariate Analysis



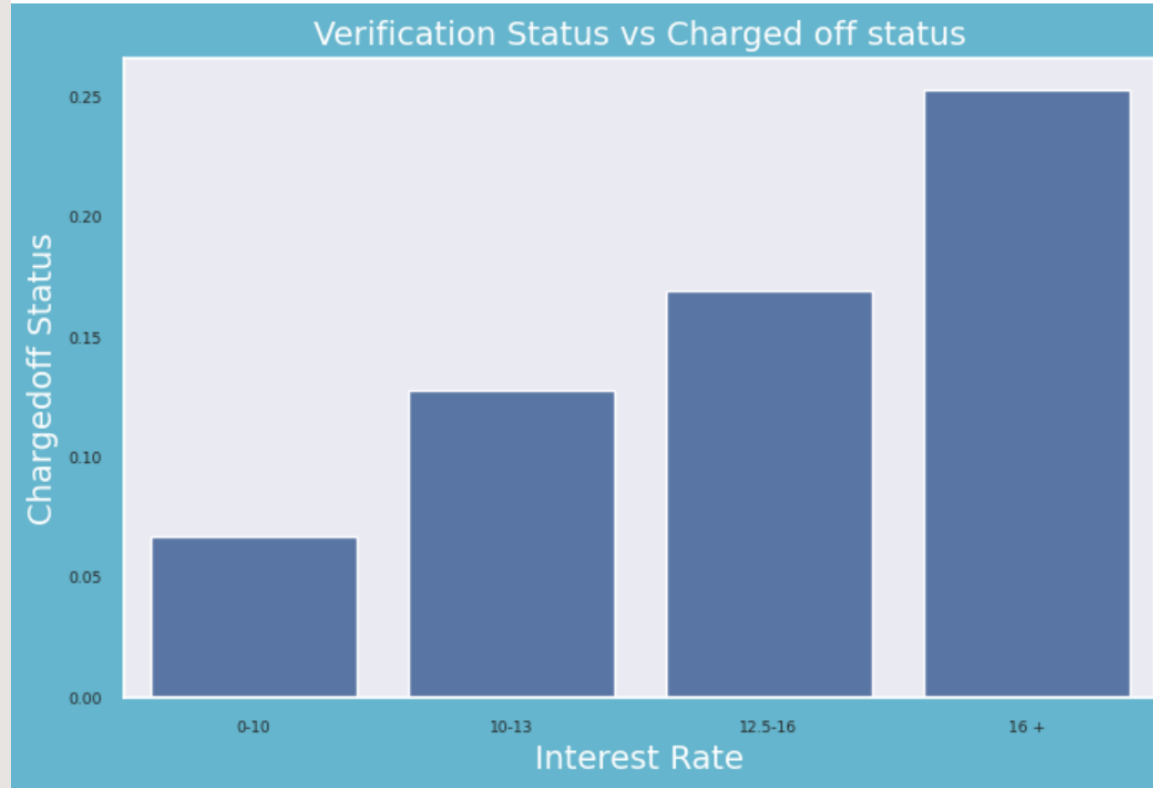
- High loan amount for short term is risk to be defaulters.

## Data Analysis – Bivariate Analysis (Annual Income Range vs Charged off status)



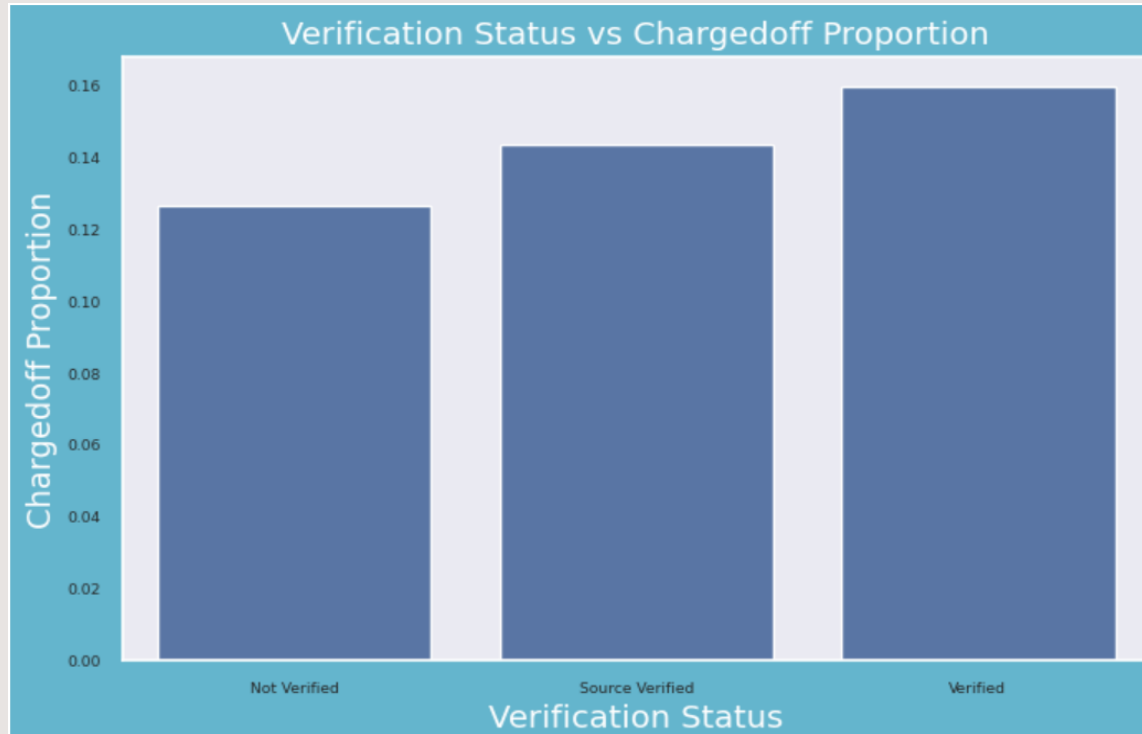
- Low annual income range under 20000 are high chance of being defaulter.
- High annual income 80000+ having chance of not being defaulter.

## Data Analysis – Bivariate Analysis (Interest Rate Range vs Charged off status)



- Interest rate having highest range 16+% is have high chance of being defaulters.
- Interest Rate Below 10% having less chances of being defaulter.

## Data Analysis – Bivariate Analysis (Verification Status vs Charged off status)



- There is not much difference in charged off but this factor is having relation with Annual income.



## Conclusion

- Borrowers with unverified annual income and short-term loans are likely to be defaulters.
- When the purpose of the loan is debt consolidation, small businesses are likely to be defaulters.
- The company should reduce the interest rate for the long term, like 60 months.
- Low annual income range under 20000 has a high chance of being a defaulter.

**Thank you!**