Lending Club Case Study

upGrad

Table of Contents

Problem Statement

How it works:

1. This company is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures. Borrowers can easily access lower interest rate loans through a fast online interface

3. With help of this Data Set we need to identify risky loan applications and In other words, the company wants to understand the **driving factors (or driver variables)** behind loan default, i.e., the variables which are strong indicators of default

Solution Approach at High-level

- Understanding the Data
- Data Curation and Manipulation
- Exploratory Data Analysis
- Data Visualization

Understanding the Data

- Lending Club Loan data set provided along with the Data dictionary

- This data set for the **consumer finance company** which specializes in lending various types of loans to urban customers and Two types of risks associate with bank's decision
  - If the applicant is **likely to repay the loan**, then not approving the loan results in a **loss of business** to the company
  - Approving the loan may lead to a **financial loss** for the company If the applicant is **not likely to repay the loan**

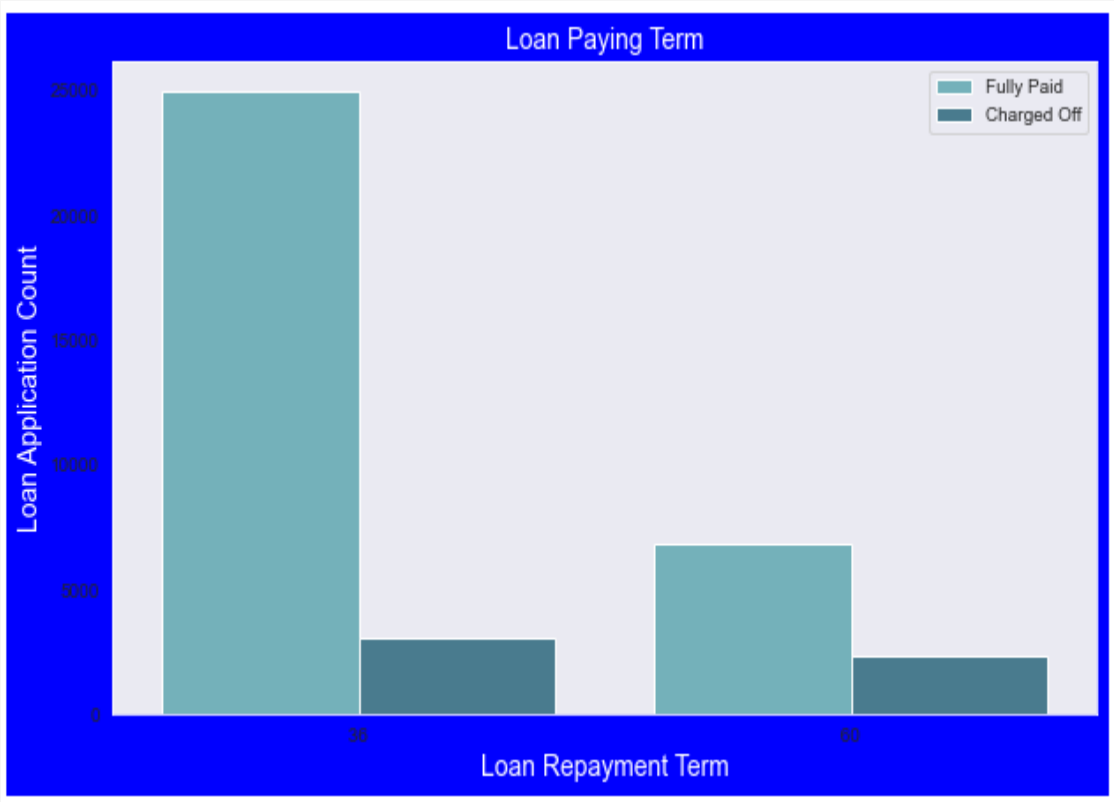
- Dataset has loan issued data from 2007 to 2011.
- Borrowers who default cause the largest amount of loss to the lenders. In this case, the customers labelled as '**charged-off**' are the '**defaulters**'

Data Curation and Manipulations

- Initial dataset had 111 Columns and 39117 Row in which 68 columns had Null values .
- Removed 53 Columns which had complete Null values
- There were 24 columns has been removed which had more district values or columns which are not need for this Analysis
- % symbol in fields, ‘months’ from term fields has been removed
- Rows which had the Null values in the emp_length has been removed since which doesn’t make sense for this Analysis
- Ongoing loan customer data has been deleted with help of the loan_status filed
- Month and Year filed being derived from the issue date filed
- Removed Outliers quantile .99 from the Annual Income filed in the Dataset

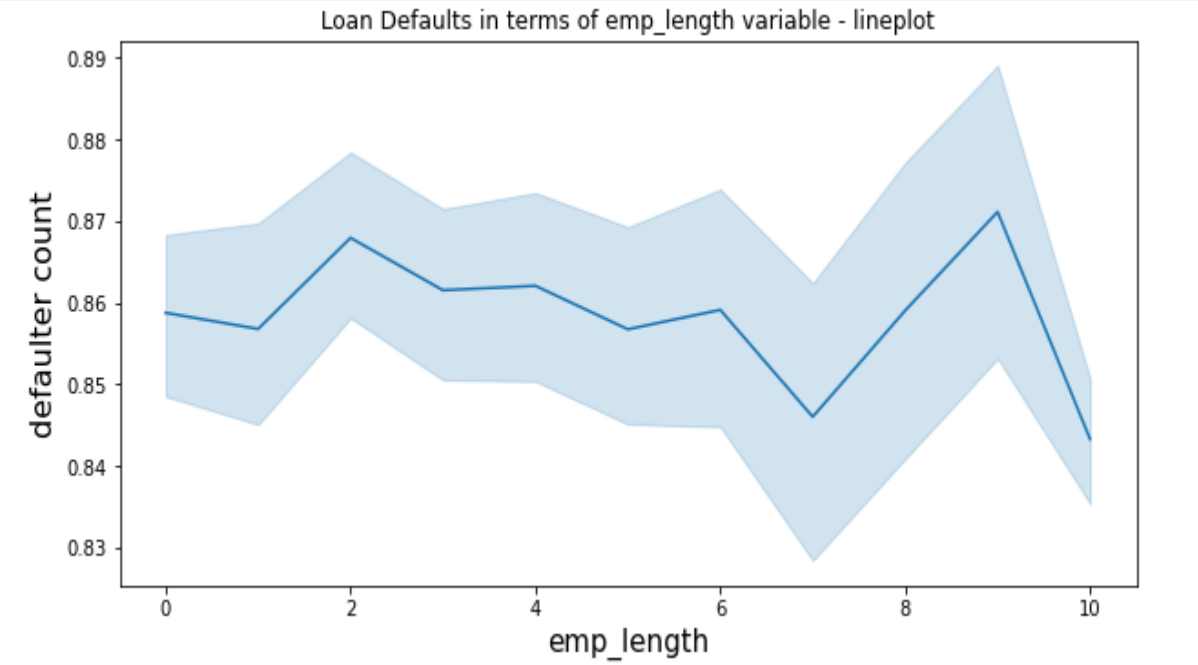
Exploratory Data Analysis

1.Univariate Analysis



Observation

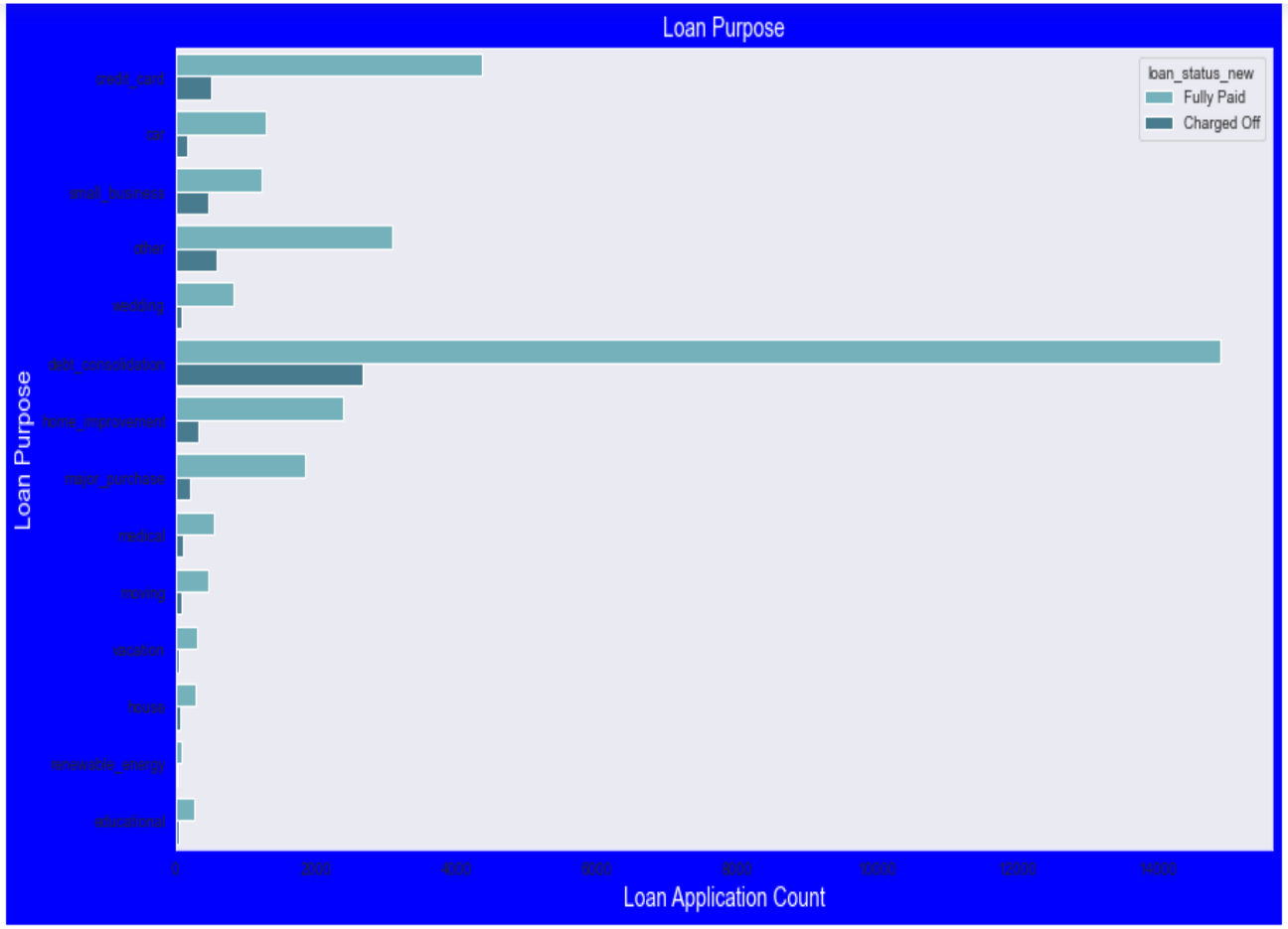
- Charged off is high for those who taken loan for 60 months compared to 36 months
- 36 months loan application is high compared to 60 months
- Defaulter ratio is high for more experience applicants



Ordered Categorical Variables

Exploratory Data Analysis

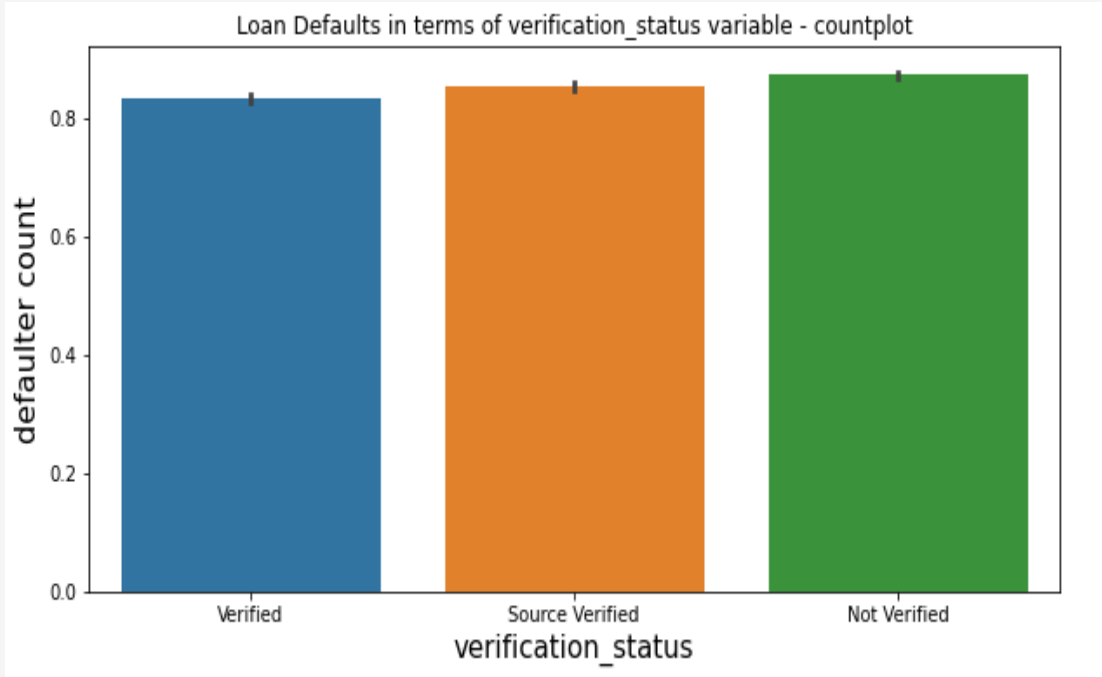
1.Univariate Analysis



Unordered Categorical Variables

Observation

- Plot shows more loans were charged off out of total loan issued for the not verified applicants.
- The plot showing that huge no of the loans was taken for debt consolidation purpose and credit card bill

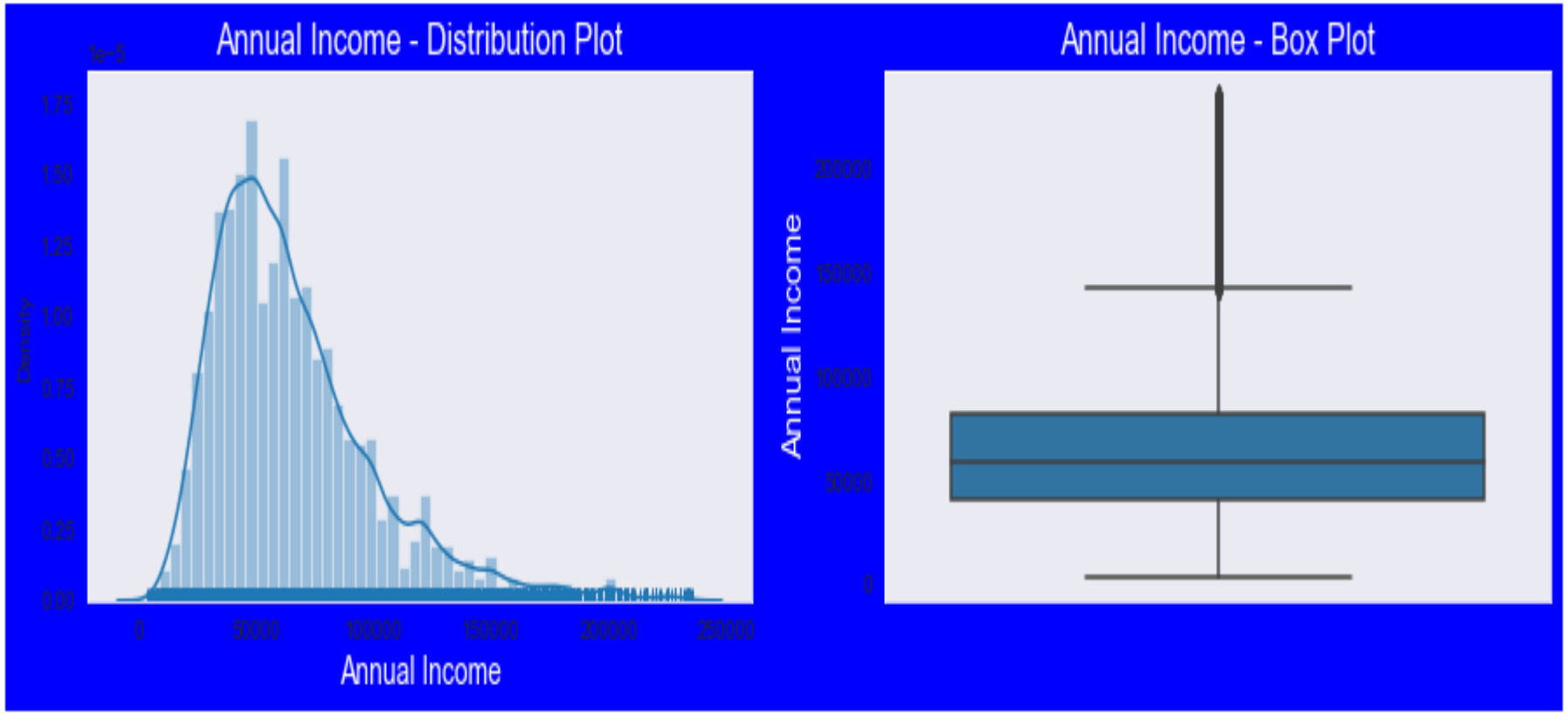


Exploratory Data Analysis

1.Univariate Analysis

Observation

- Below plots show that most of the borrower's Annual incomes are in range of 40000- 80000.



Exploratory Data Analysis

2.Bivariate Analysis

Following columns derived for the Bivariate Analysis

- Bin created for the loan_amnt, annual_inc and int_rate

loan_status_new	annual_inc_bin	Charged Off	Fully Paid	Total	Chargedoff_percentage
0	0-20000	187	809	996	18.775100
1	20000-40000	1415	6697	8112	17.443294
2	40000-60000	1671	9320	10991	15.203348
3	60000-80000	1001	6502	7503	13.341330
4	80000 +	1068	8449	9517	11.222024

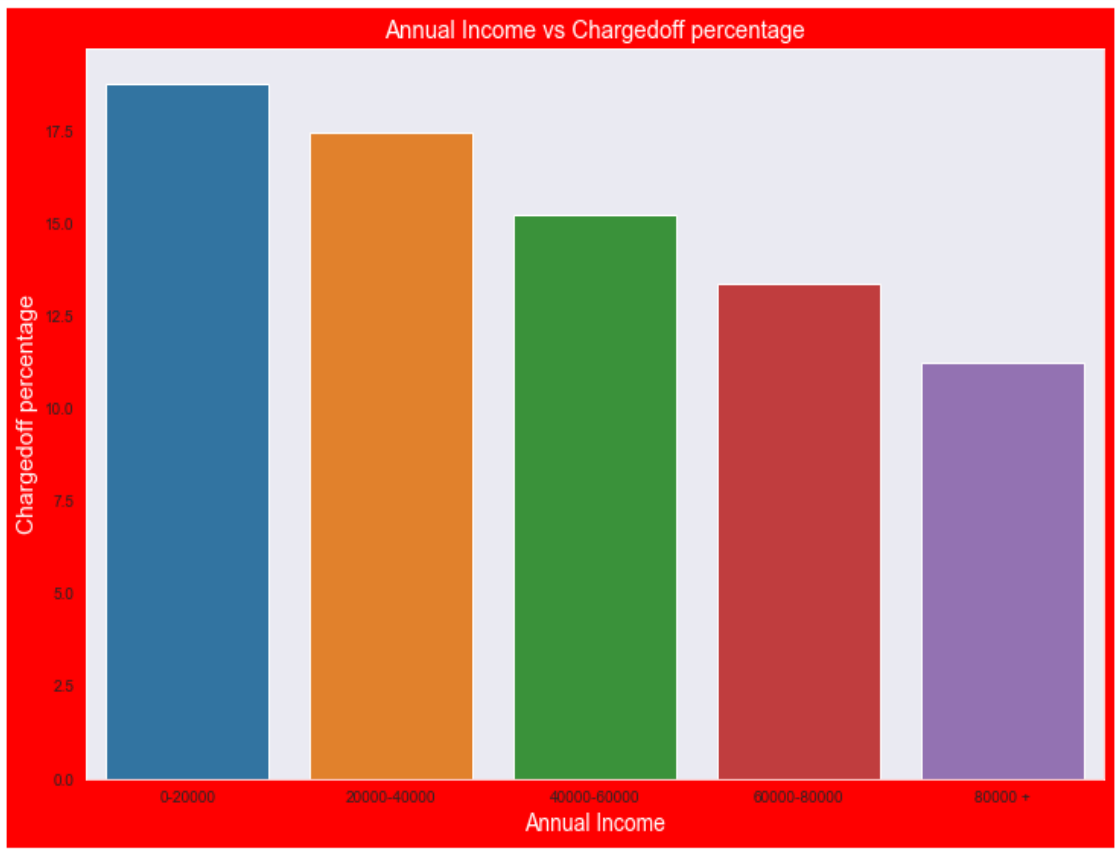
loan_status_new	int_rate_bin	Charged Off	Fully Paid	Total	Chargedoff_Percentage
3	16 +	1208	3242	4450	27.146067
2	12.5-16	1918	9185	11103	17.274610
1	10-13	1175	7915	9090	12.926293
0	0-10	758	10988	11746	6.453261

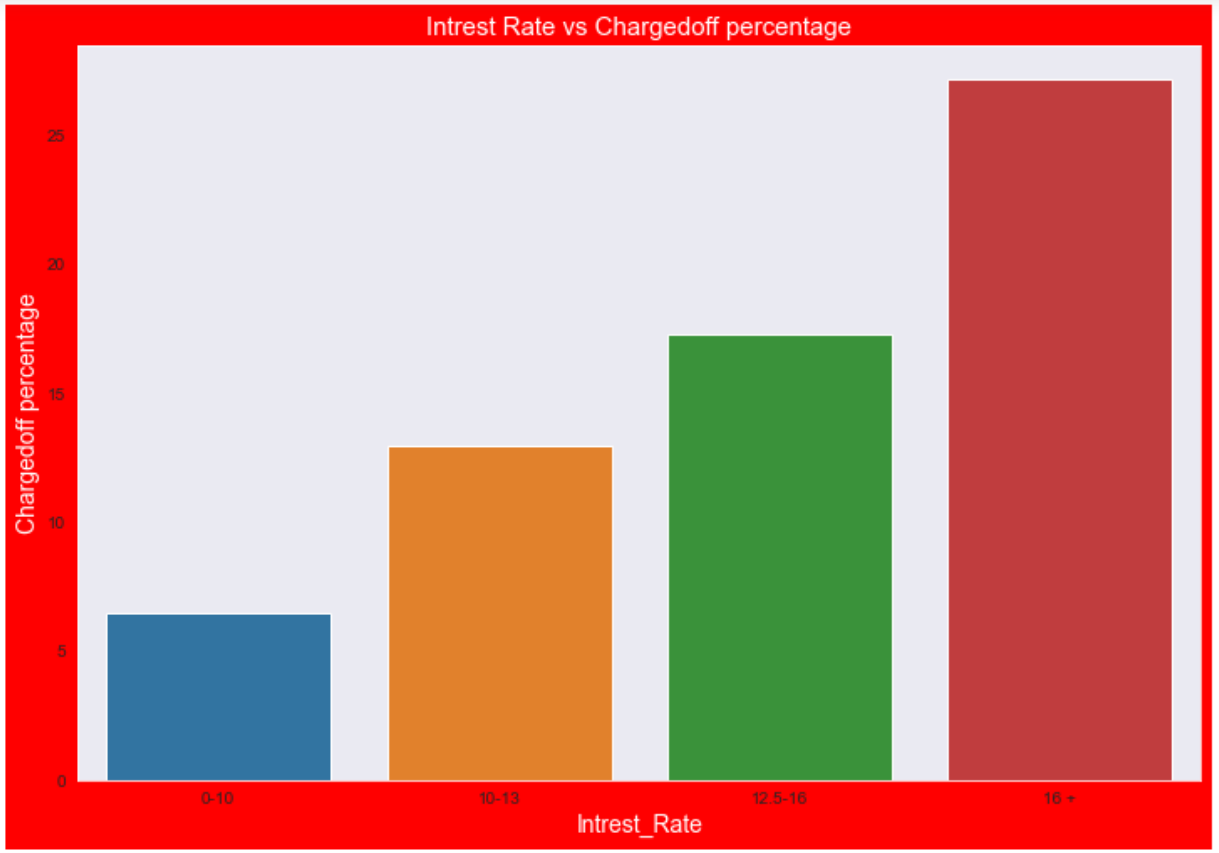
Exploratory Data Analysis

2.Bivariate Analysis

Observation

- More Interest rate leading to high charged off percentage
- who has income more than 80000+ has less charge off and 0-20000 has the more defaulters



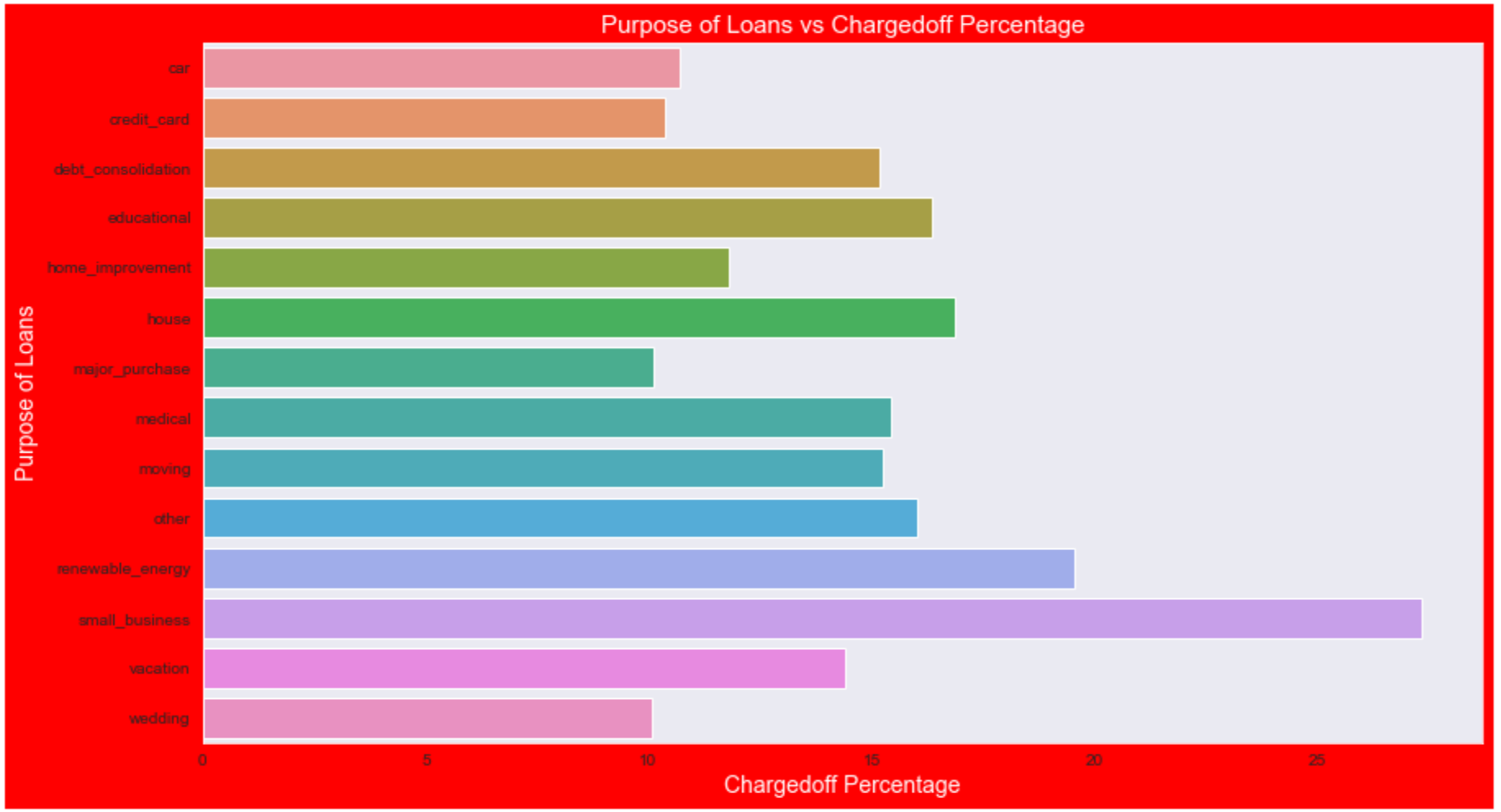


Exploratory Data Analysis

2.Bivariate Analysis

Observation

- small Business applicants has high chances of getting more defaulters.

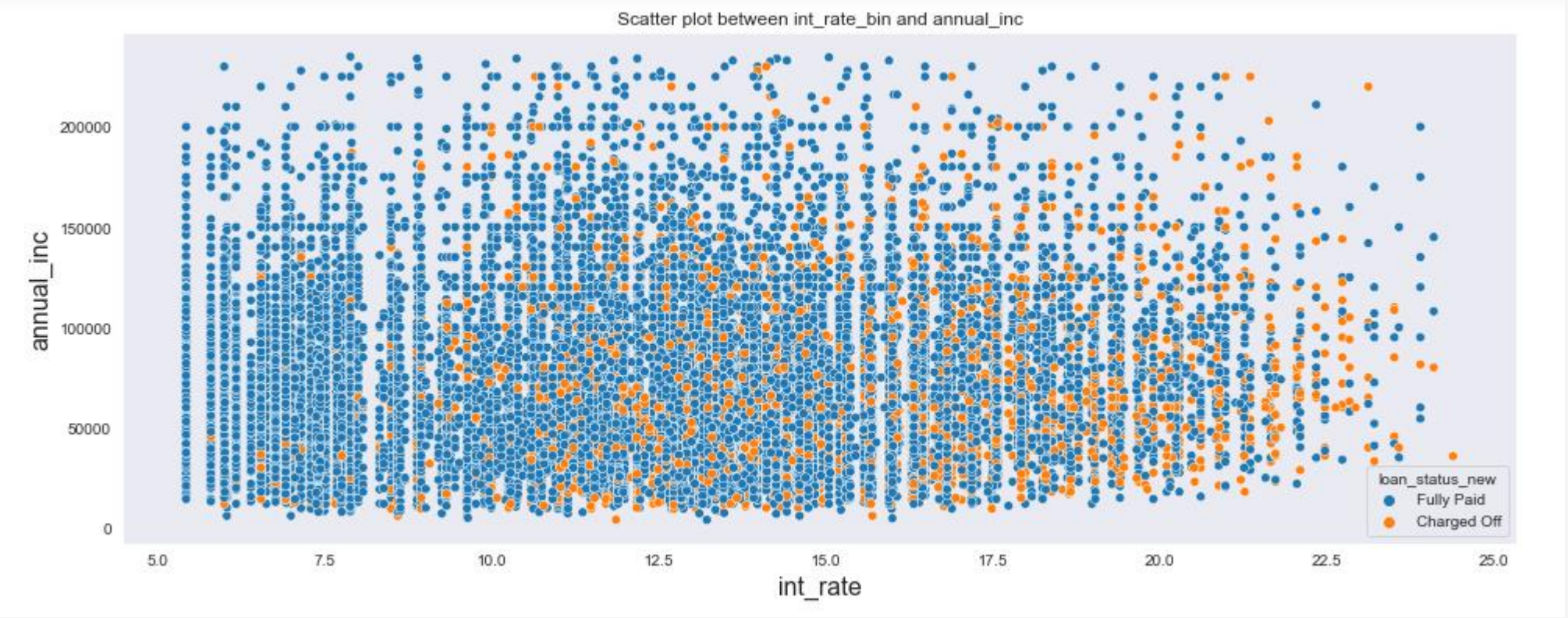


Exploratory Data Analysis

2.Bivariate Analysis

Observation

- defaulters are very high for above 17 percent interest rates

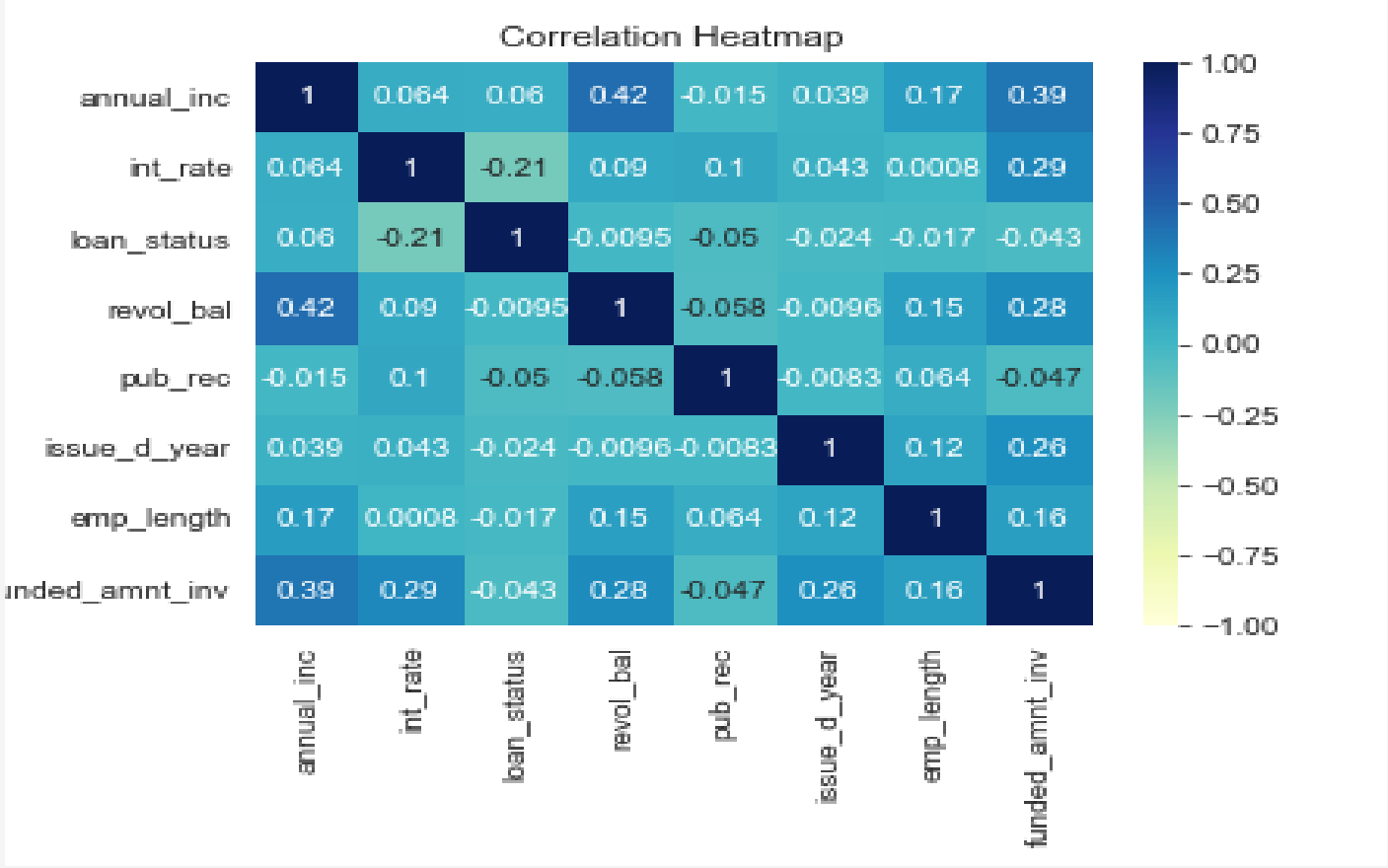


Exploratory Data Analysis

3.Multivariate Analysis

Observation

- A correlation map was created which provides a clear view on which columns are correlated.
- This also shows that income increases based on the work experience



Conclusion

- The Major **driving factors** to identify the defaulters are
    - **Employment Length**
    - **Interest Rate**
    - **Loan Paying term**
    - **Annual Income**

Thank You