Lending Club Casestudy

Sadasivuni Nikhita (Group Facilitator)

Objectives

Project Brief

We analyze the historical loan data of a large online loan marketplace specializing in lending various kind of loans to urban customers.

Problem statement

We analyze the historical loan data of a large online loan marketplace specializing in lending various kind of loans to urban customers.

Data Cleaning and correction

Identify and correct quality issues , Interpret the correct meaning of variables (Columns)

Data Analysis

Univariate & Segmented univariate analysis, BiVariate Analysis

Step 1

Imported Required Libraries

- Pandas : To work with dataset
- Numpy : To work with math library
- Matplotlib: to plot some variables
- Seaborn: Graph library that use matplotlib

Step 2

Loaded Datasets

‘Loan.csv’ was loaded into a google colabs to perform analysis

Step 3

Data Cleaning

Data Cleaning

- Drop all the up blank columns
- Identify the columns with more than 50% Null values , remove them
- Find unique values and filter these columns
- Delete all the columns with just 1 (Unique) value as they would not add any value in analysis.
- Identify duplicate or very close Columns.

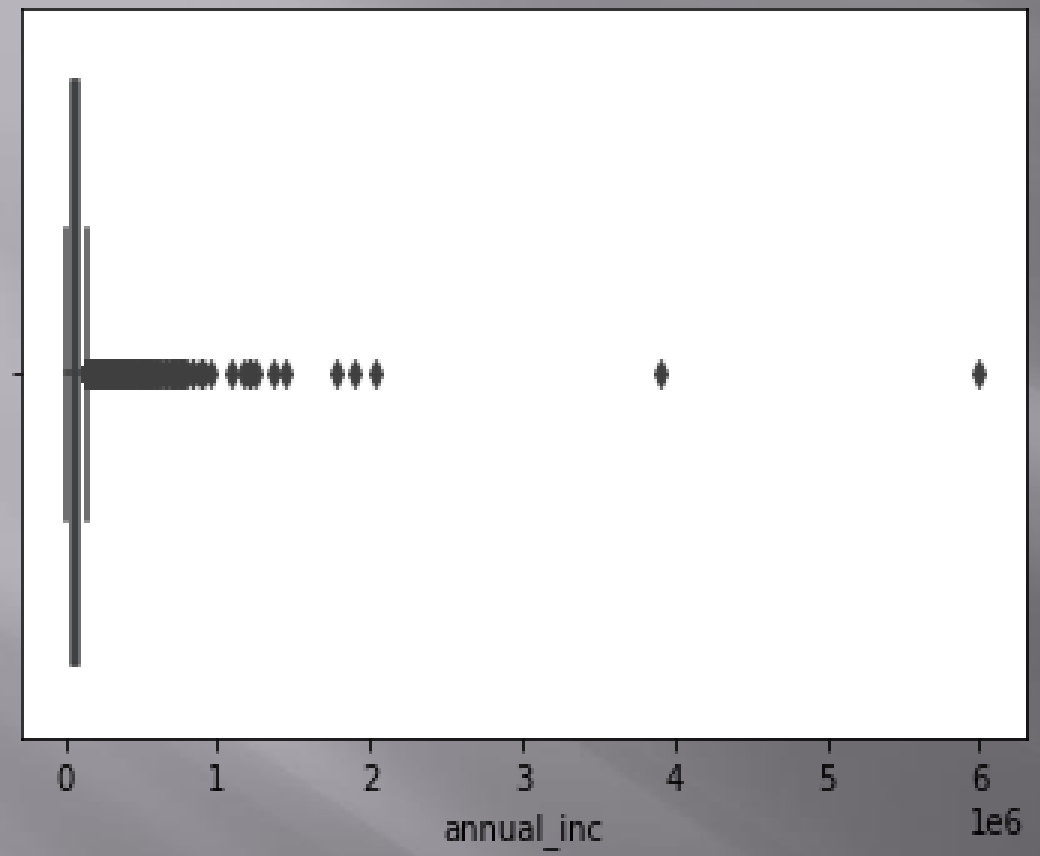
Data Correction

- Removing strings/characters from emp_length
- changing 'NONE' entry in 'home_ownership' with 'OTHER'
- Replacing Blank employer name value with Unknown

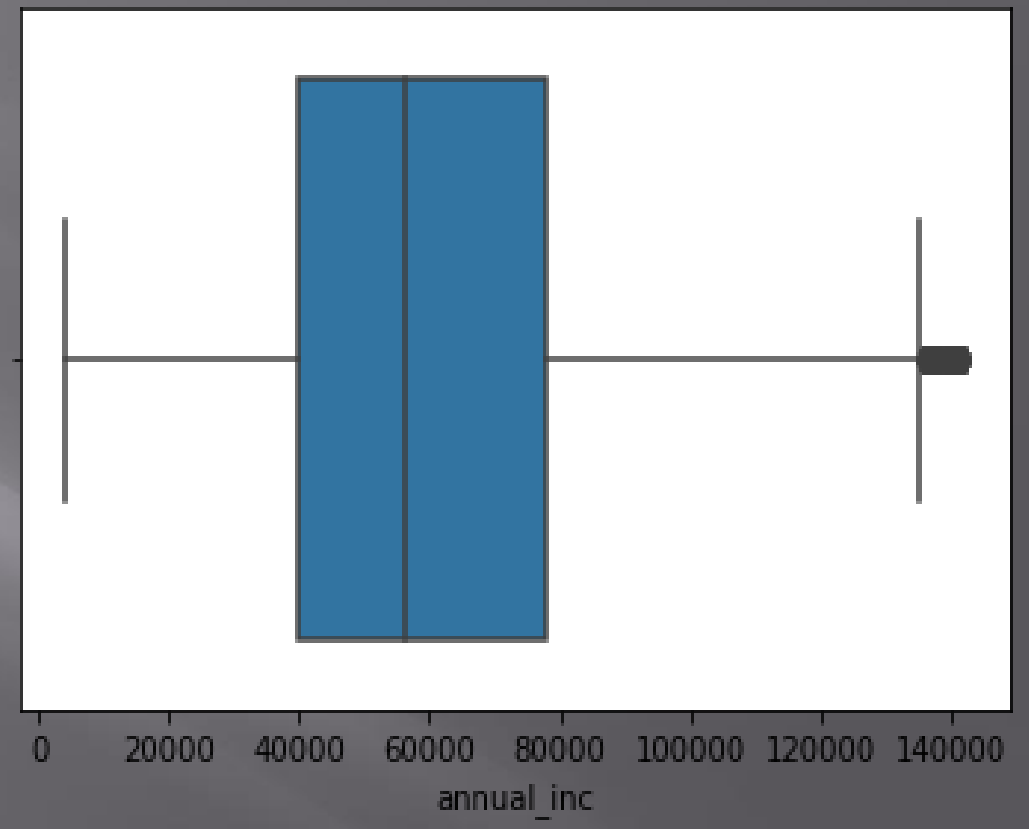
- Replacing NAN value with 0 bankrupcties
- Removing non number charactors from revol_util columns and change to int type

Step 3

Univariate Analysis

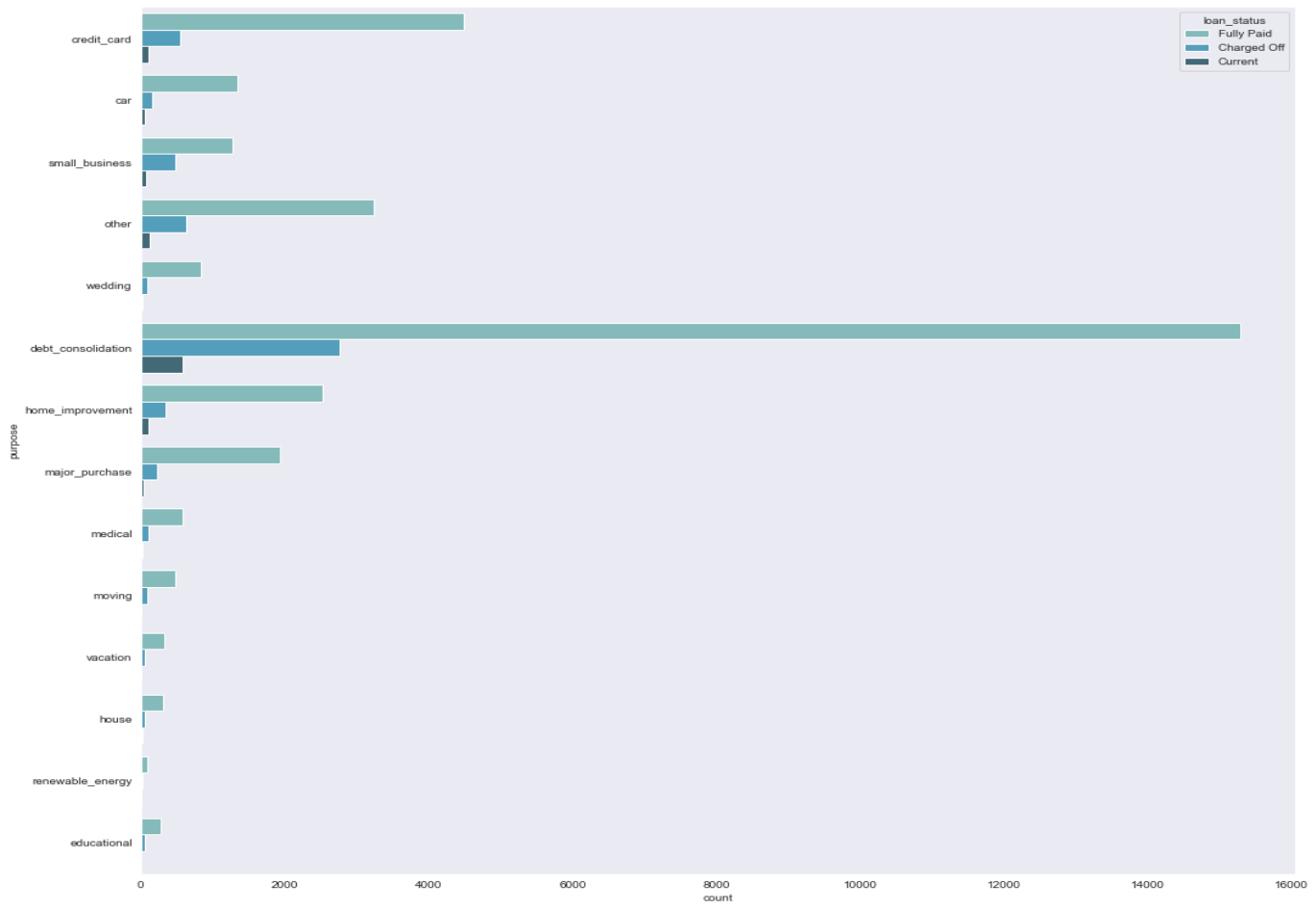


Analysis of annual_inc using boxplot identified outliers



Removed outliers considering threshold as 0.95

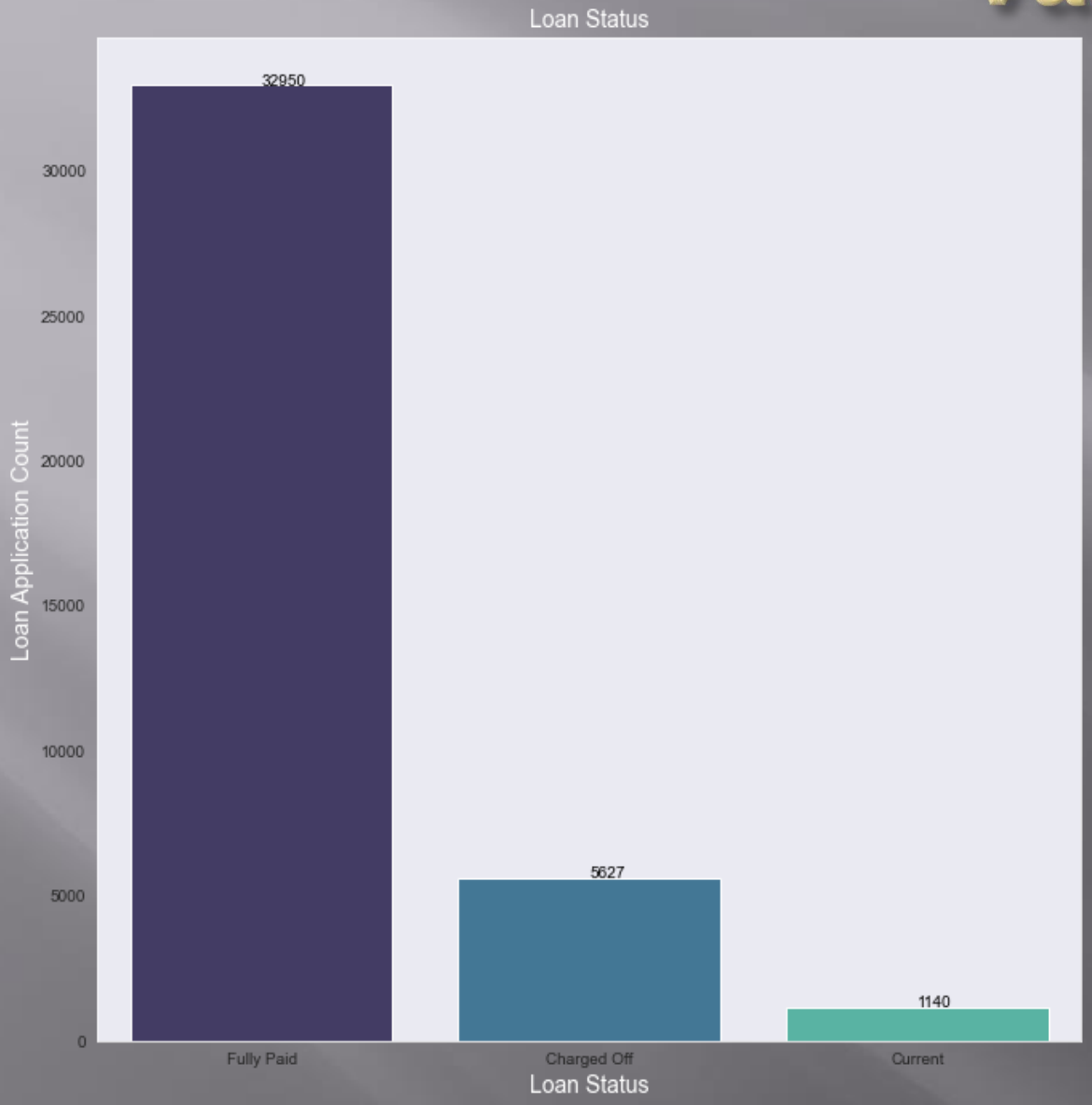
Unordered Categorical Variables



Purpose Of Loan with Loan Status

We can see from the countplot that debt consolidation loans are fully paid and charged off higher compared to other loan purposes and education loan has least charged off and renewable energy is lowest paid fully loans.

Categorical Variables

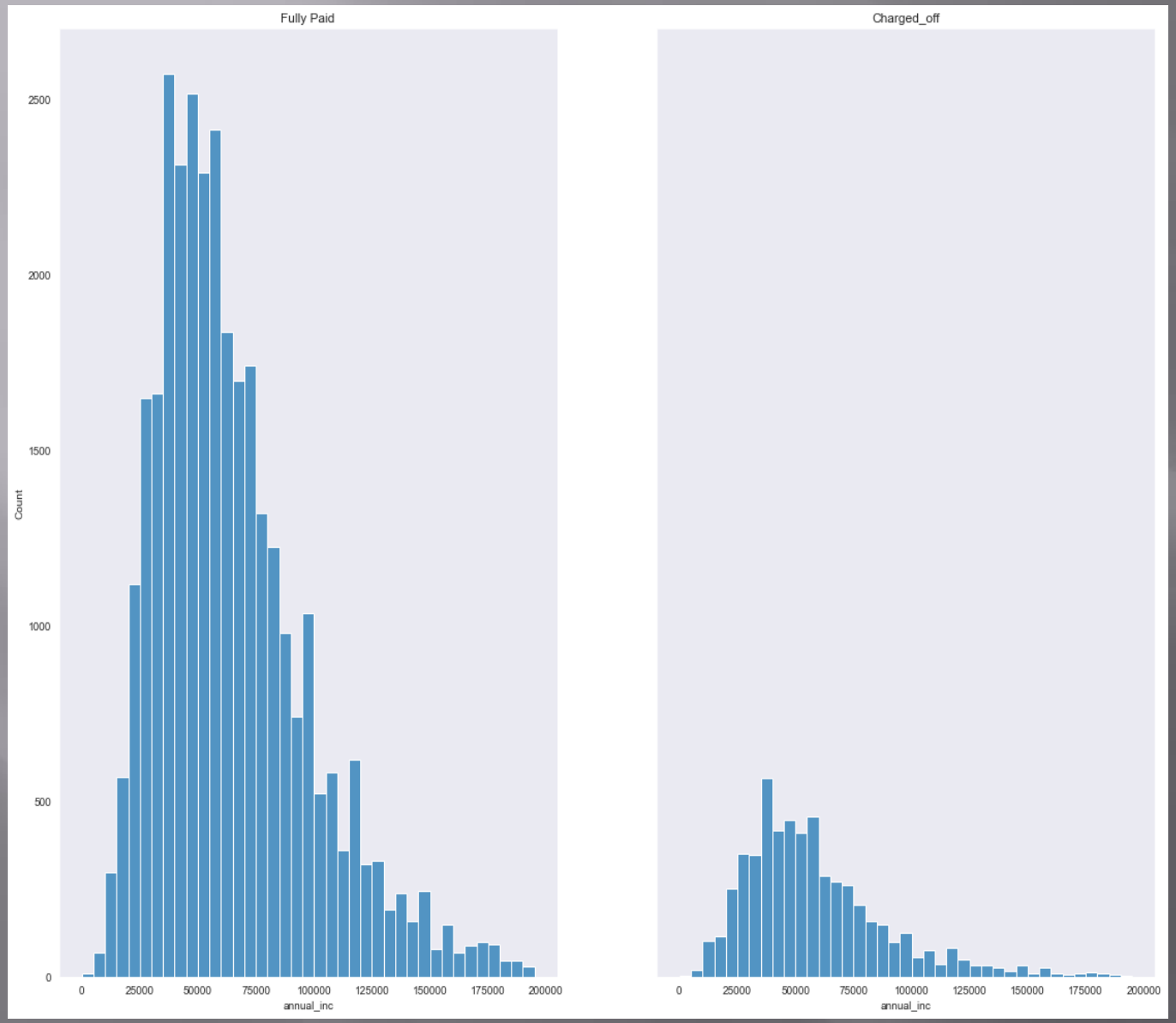


Loan Status with countplot

From the given Data, there are 32950 customers have fully Paid the loans on the other side 5627 customers have been charged off.

Step 3

Segmented Univariate Analysis

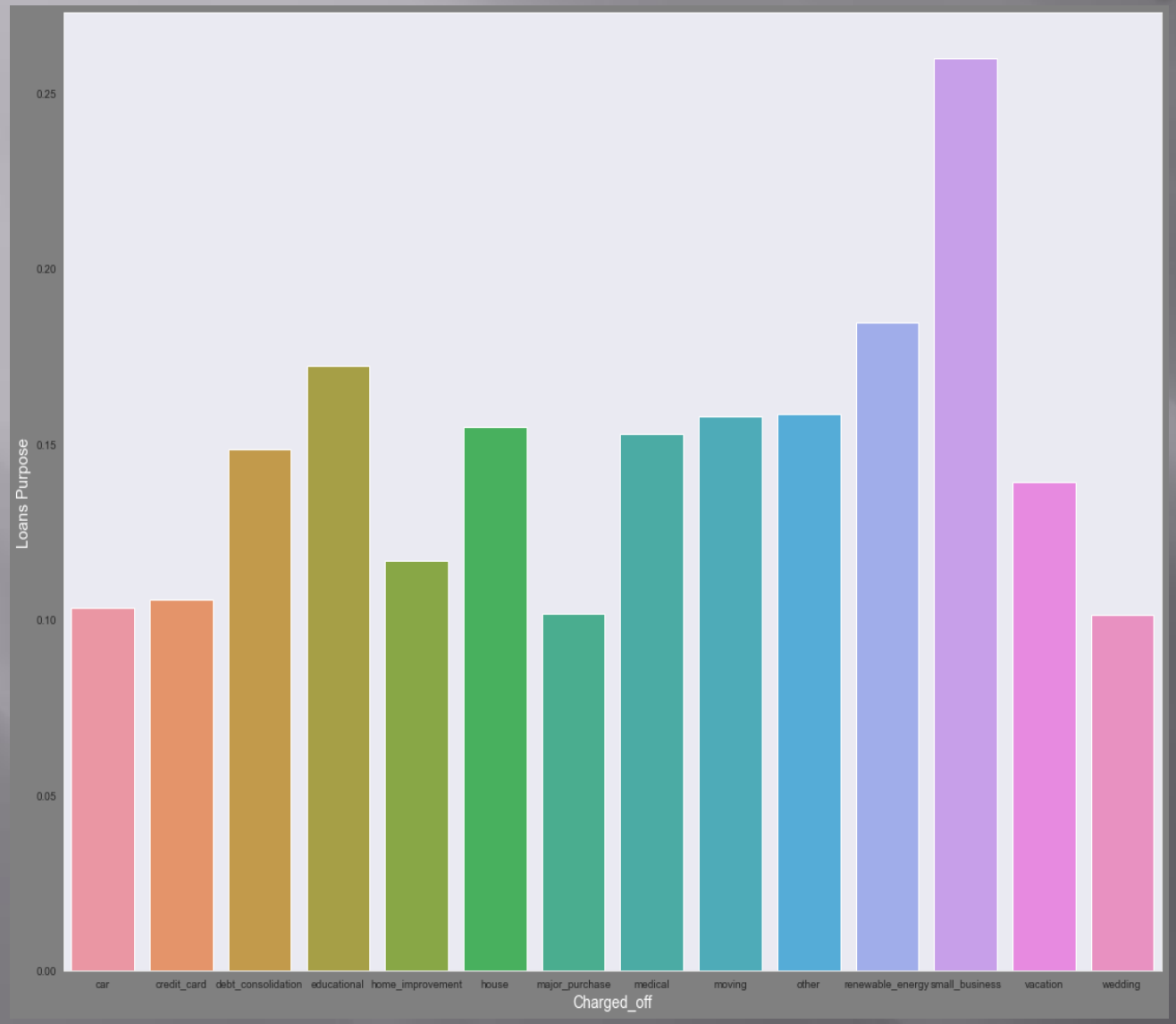


By the histplots we analyse that customers with annual income range between 25000 to 125000 fully paid the loan¶

Step 4

Bivariate Analysis

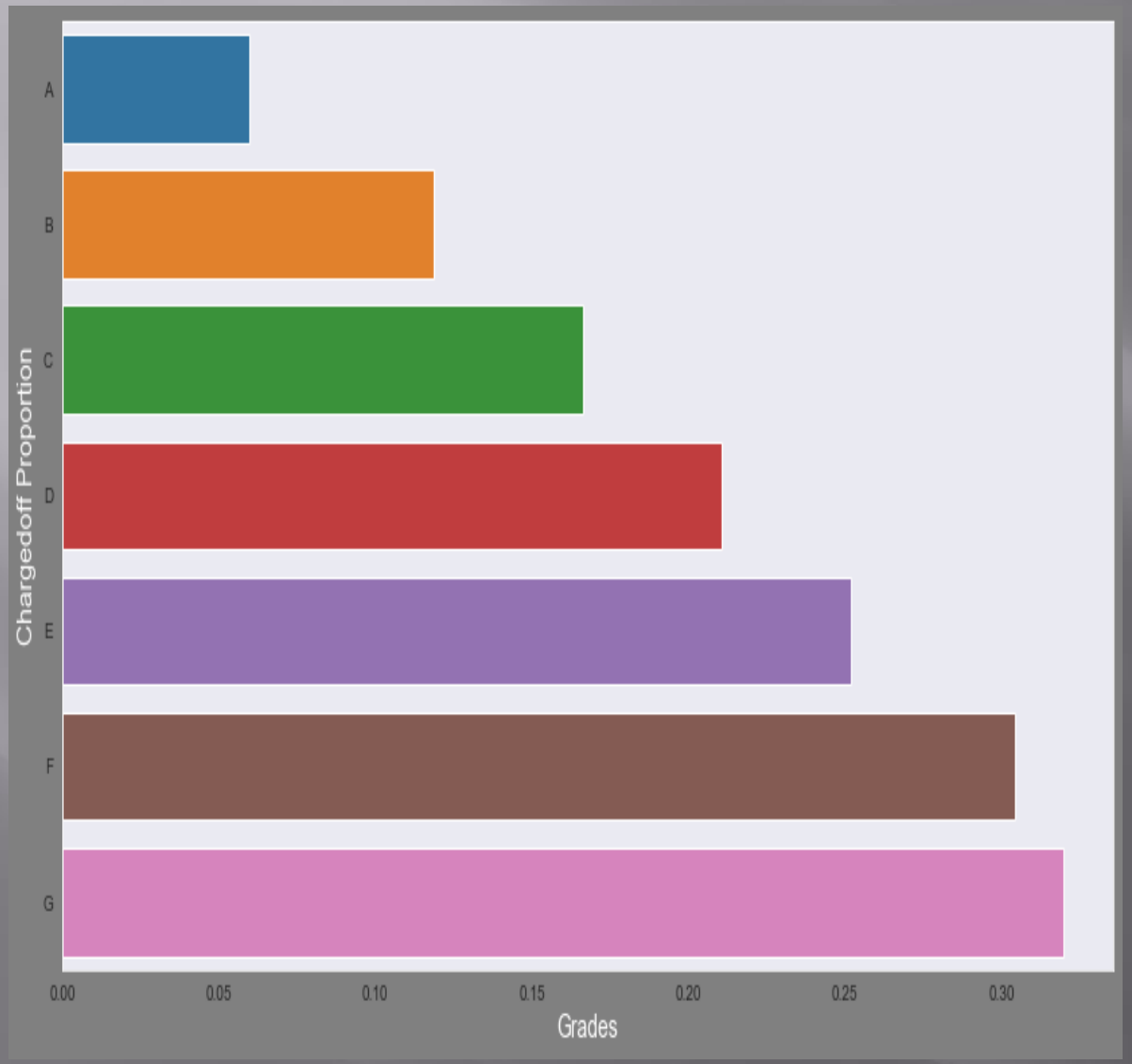
Bivariate Analysis



Analysis on purpose vs chargedoff_portion

- **Small Business applicants have high chances of getting charged off.**
- **renewable_energy where chanrged off proportion is better as compare to other categories.**

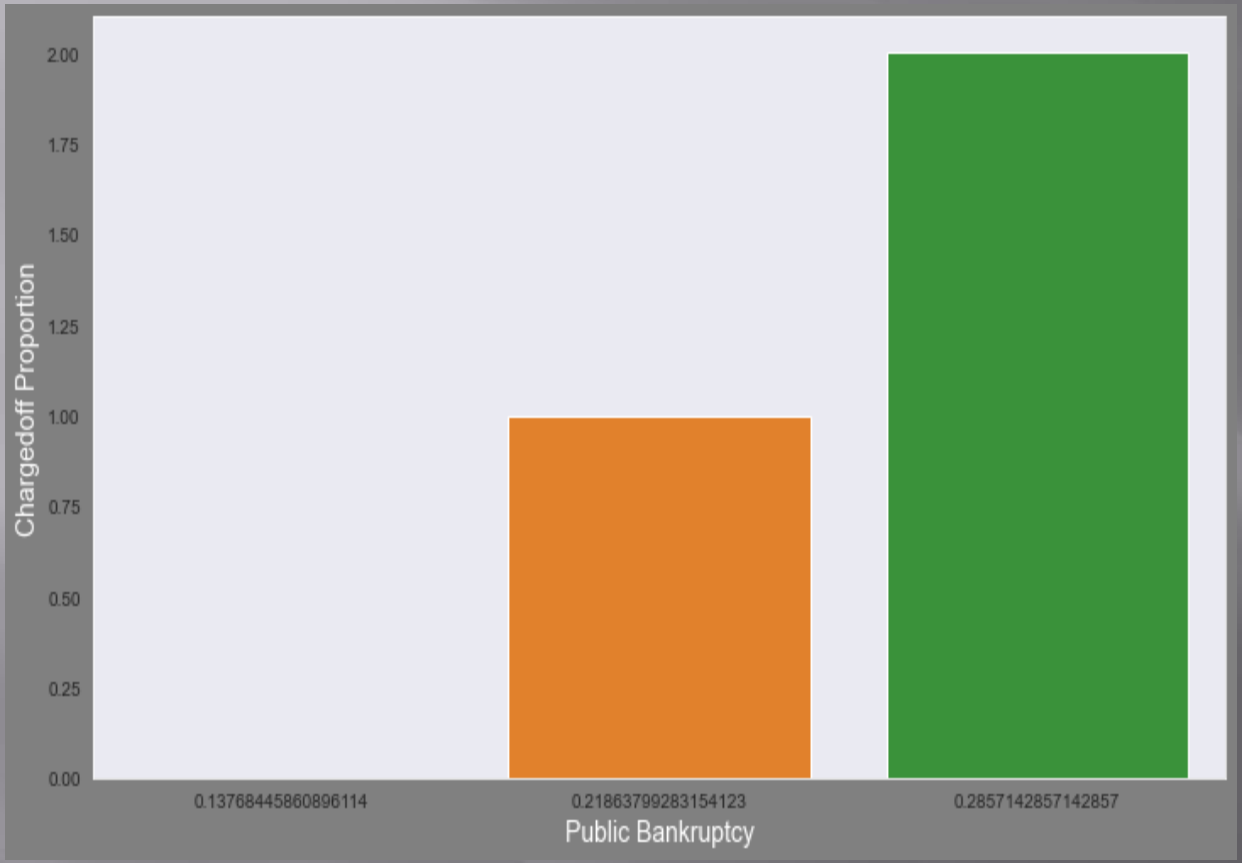
BiVariate Analysis



Analysis on grade vs Chargedoff_portion

- **Grade "A" has very less chances of charged off.**
- **Grade "F" and "G" have very high chances of charged off.**
- **Chances of charged off is increasing with grade moving from "A" towards "G"**

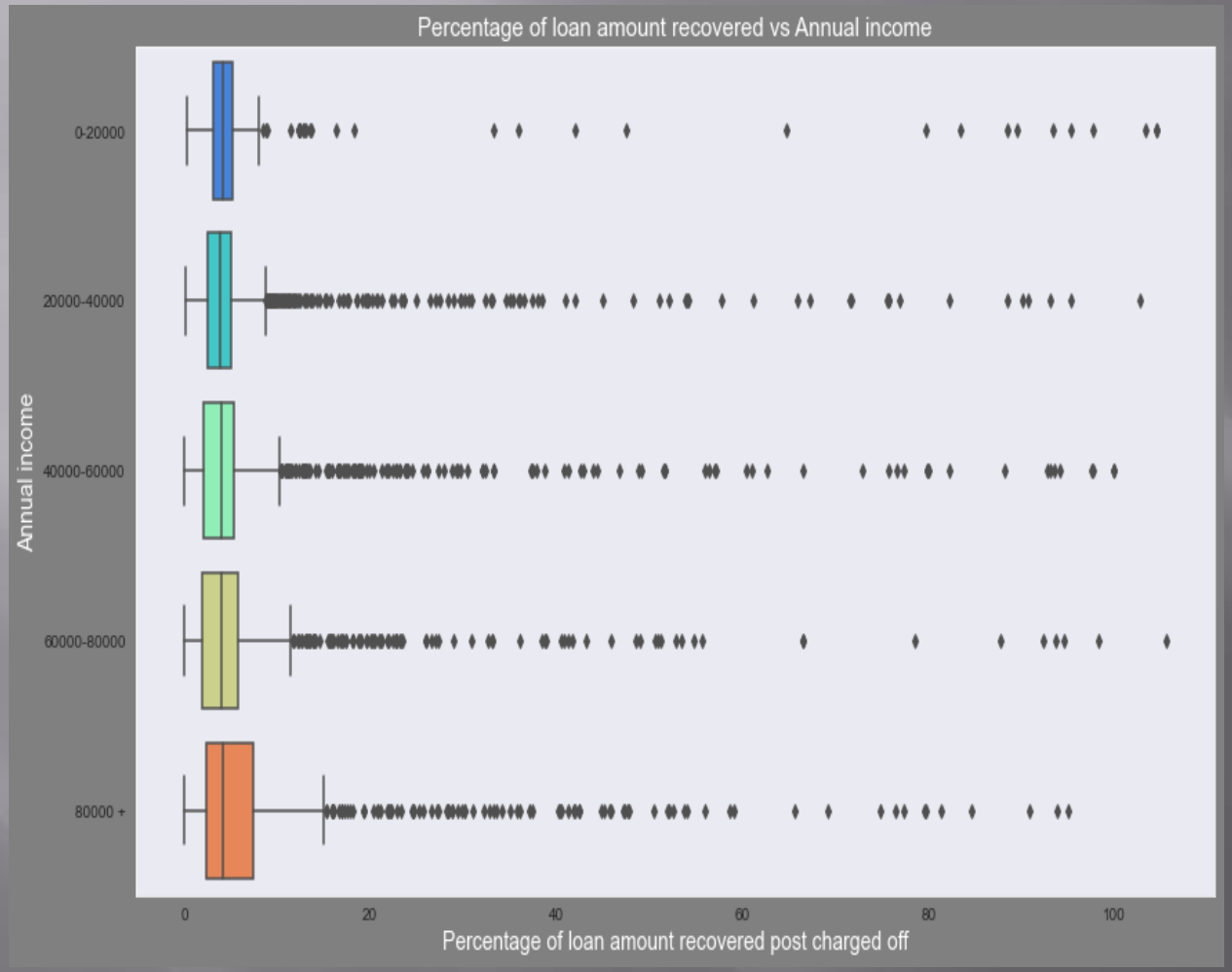
Bivariate Analysis



Analysis on pub_rec_bankruptcies vs Chargedoff_Proportion

- **Those who already have pub_rec_bankruptcies value 1, have charged off proportion higher than who have no pub_rec_bankruptcies.**
- **pub_rec_bankruptcies count 2 has even higher charged off proportion but those numbers are not significant**

Bivariate Analysis



Bivariate Analysis on Recoveries vs Annual income

- **Higher percentage of loan amount is recovered when annual income is high.**
- **Plot shows no significant variation but there is slight increase in recovery percentage with increase in annual income.**