Exploratory Data Analysis

Lending Club Case Study

By
Arun James P

General Information

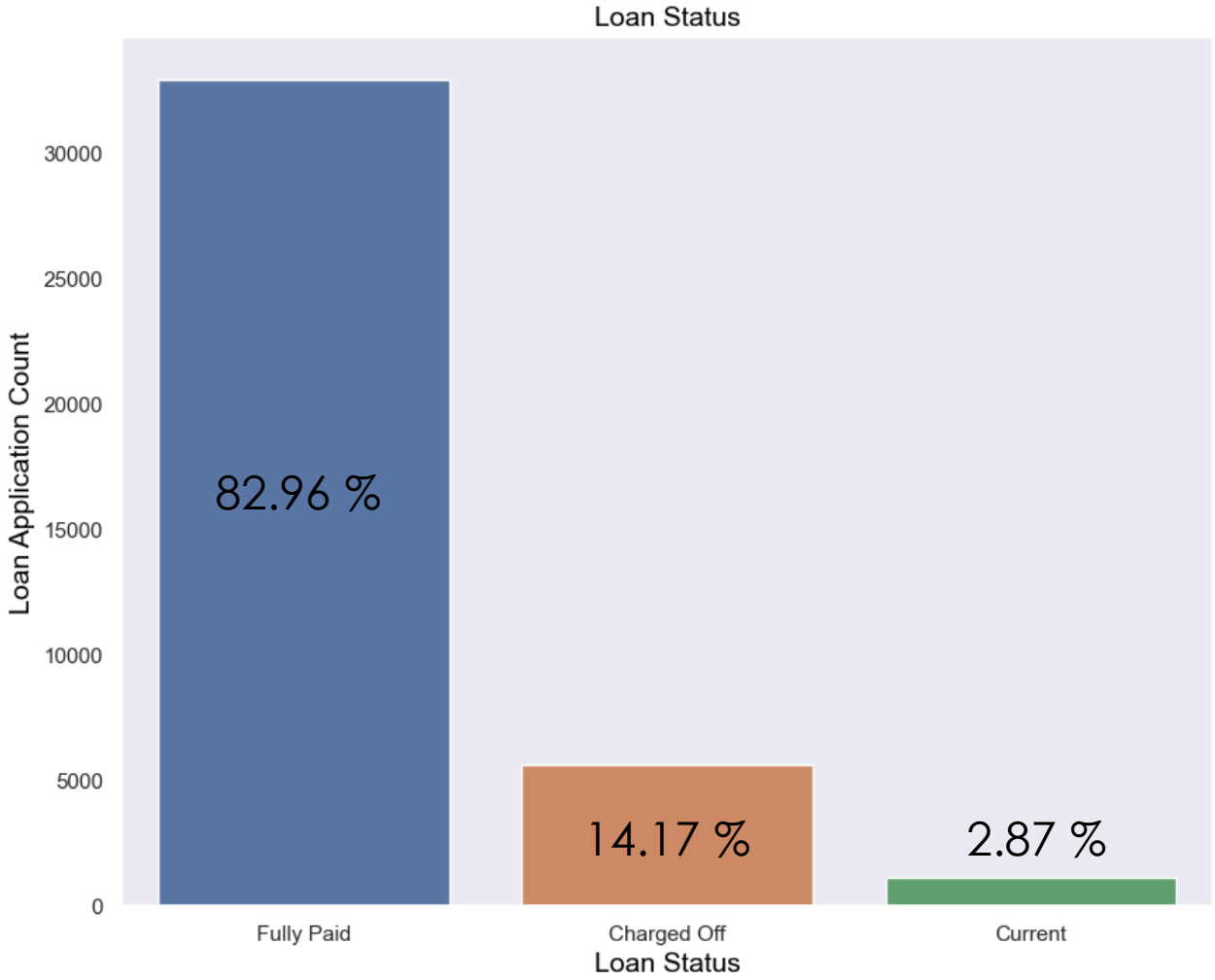
Project Overview: This project focuses on risk analytics in the largest online loan marketplace, aiming to minimize credit losses by identifying factors indicating loan default. The company seeks to understand the variables strongly associated with defaulting borrowers, enabling better portfolio management and risk assessment.

Background: Operating in the financial domain, the company provides online loans, facing challenges of credit losses from defaulted loans. The project leverages exploratory data analysis (EDA) to discern patterns and variables predicting loan default.

Business Problem: The core business problem is identifying risky loan applicants to reduce credit losses. The project targets understanding variables strongly linked to loan default, specifically those labeled as 'charged-off,' improving lending decisions.

Dataset Used: While specifics about the dataset are not provided, it likely includes loan and borrower information, facilitating the identification of patterns and correlations indicative of potential loan default.

Univariate Analysis – Loan Status



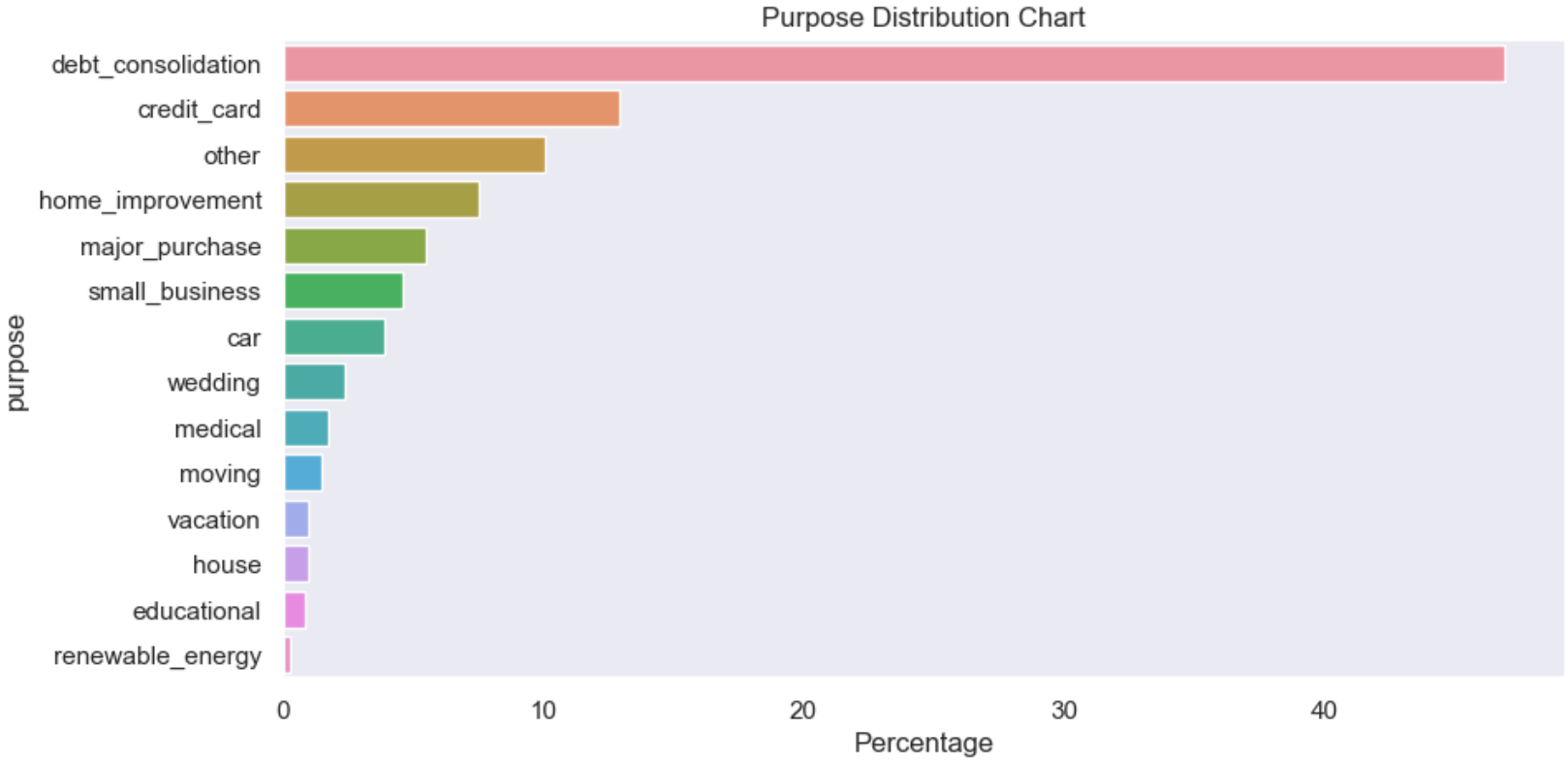
Observations :

- 82.96 % loans were fully paid.
- 14.17 % loans were charged off.
- 2.89 % loans were current.

Loan Status

Fully Paid - 82.96 %
Charged Off -14.17 %
Current - 2.87 %

Univariate Analysis – Purpose of Loan

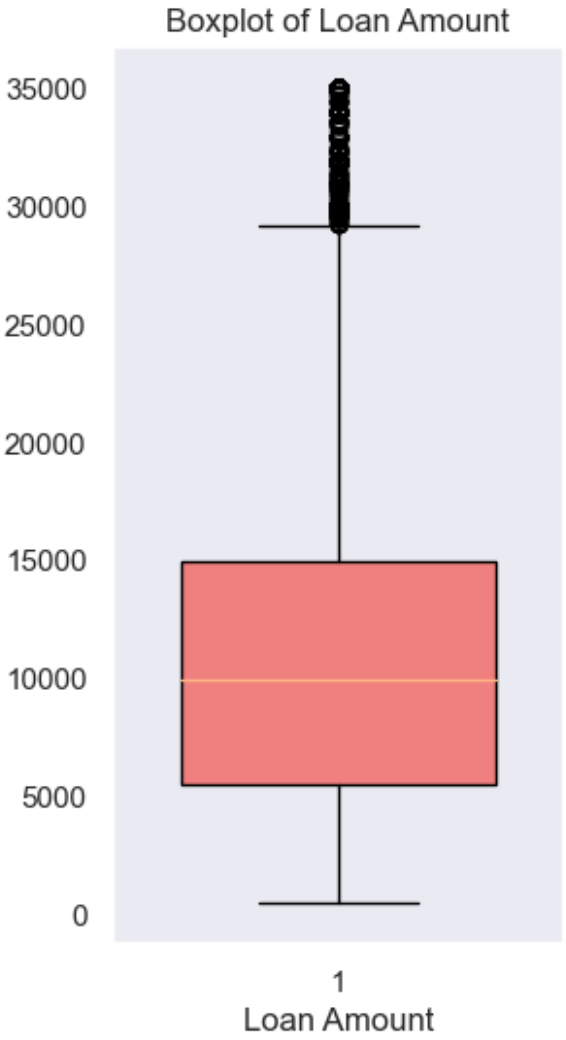


Purpose	
Debt consolidation	46.93
Credit card	12.92
other	10.05
Home improvement	7.49
Major purchase	5.51
Small business	4.60
car	3.90
wedding	2.38
medical	1.74
moving	1.47
vacation	0.96
house	0.96
educational	0.82
Renewable energy	0.26

Observations :

- Most of the loans taken for debt consolidation(47%)
- Second one is Credit card bill payment (13%)

Univariate Analysis – Loan Amount

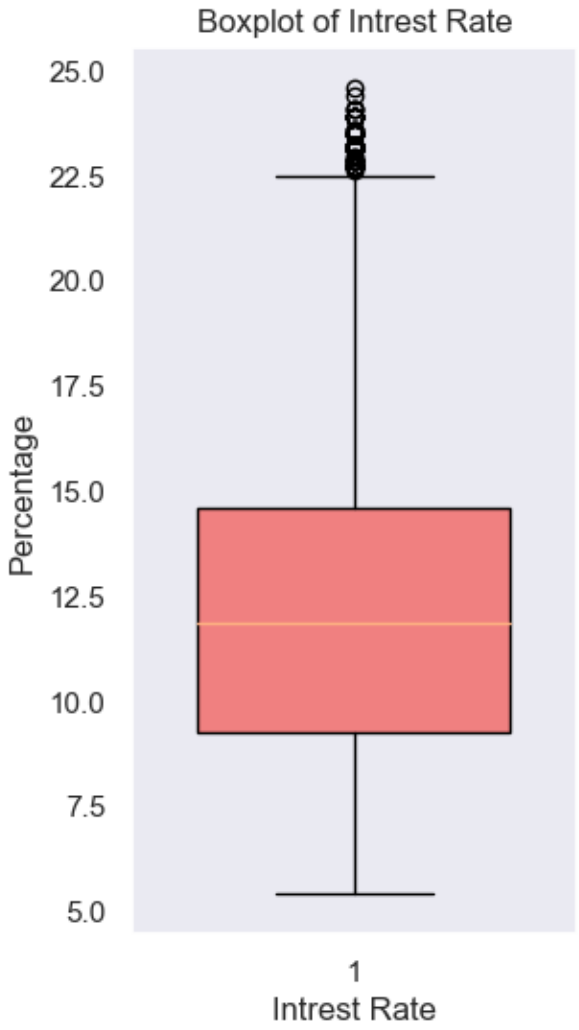


count	39,717.00
mean	11,219.44
std	7,456.67
min	500.00
25%	5,500.00
50%	10,000.00
75%	15,000.00
max	35,000.00

Observations :

- Most of the Loan amounts are in range of 5000 - 15000.

Univariate Analysis – Interest Rate

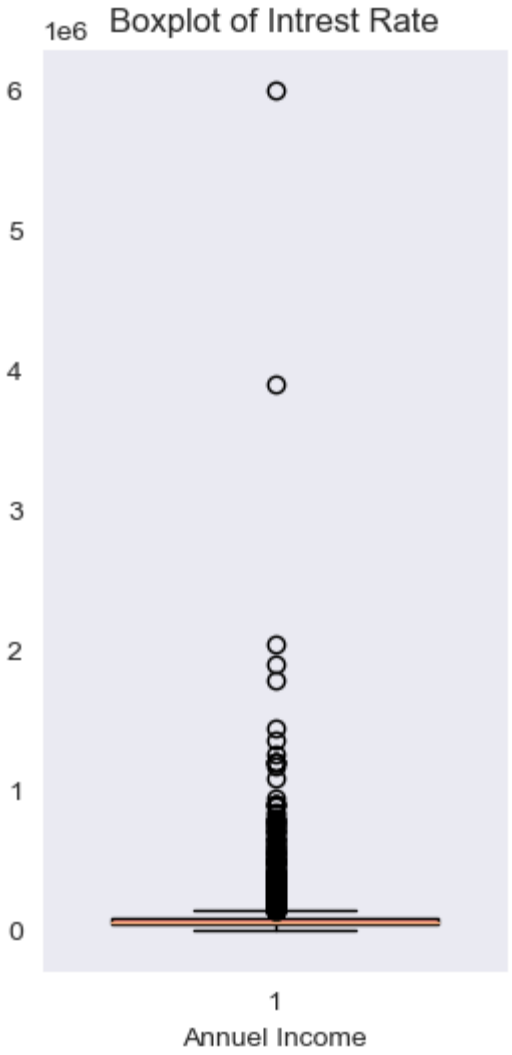


count	39,717.00
mean	12.02
std	3.72
min	5.42
25%	9.25
50%	11.86
75%	14.59
max	24.59

Observations :

- Most of the interest rates are in range of 9 % - 14.5 %.

Univariate Analysis – Annual Income

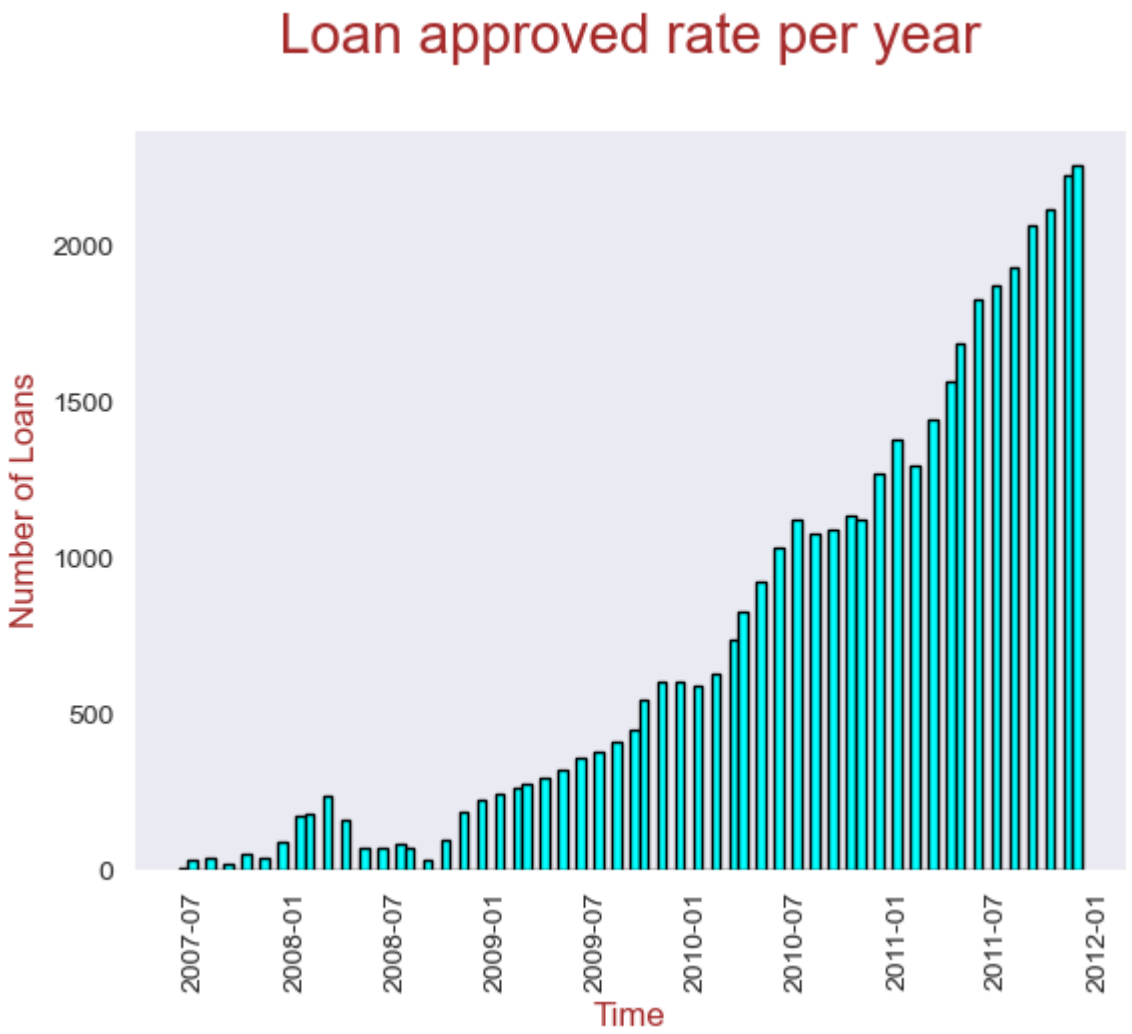


count	39,717.00
mean	68,968.93
std	63,793.77
min	4,000.00
25%	40,404.00
50%	59,000.00
75%	82,300.00
max	6,000,000.00

Observations :

- Most of the borrower's Annual incomes are in range of 40000 - 82000

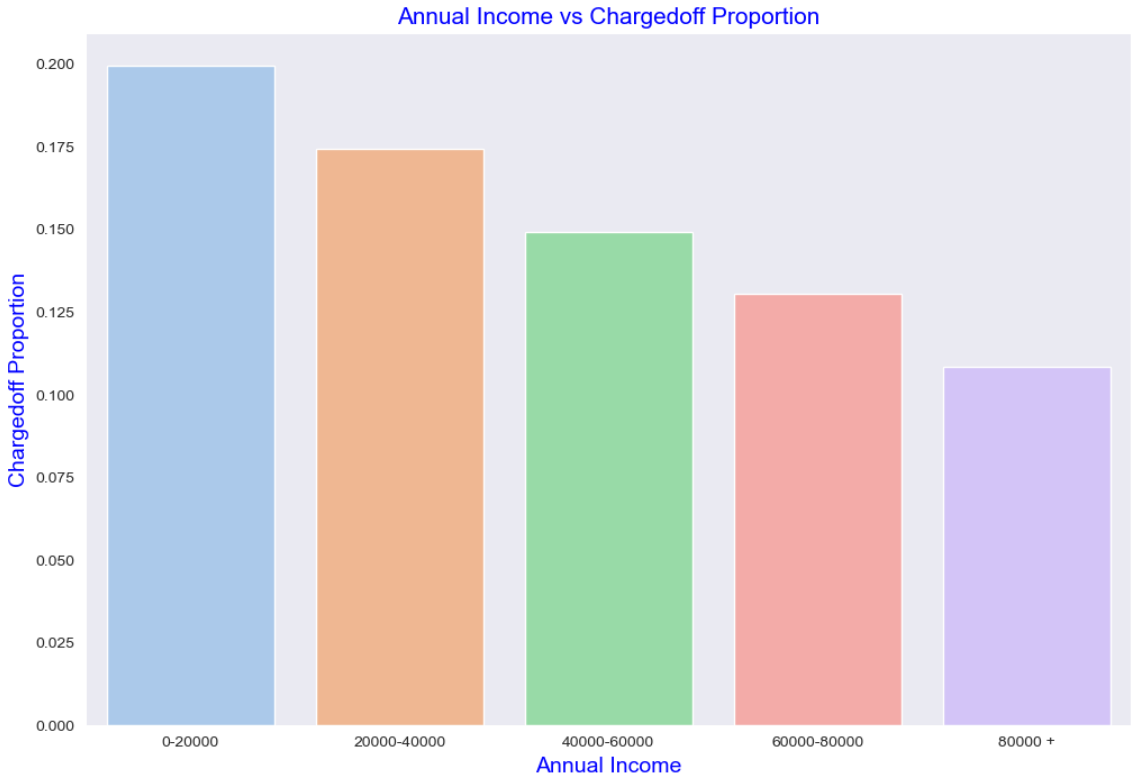
Univariate Analysis – Loan approved rate per year



Observations :

- Number of loan application is increasing every year

Bivariate Analysis – Annual Income Vs Charged off

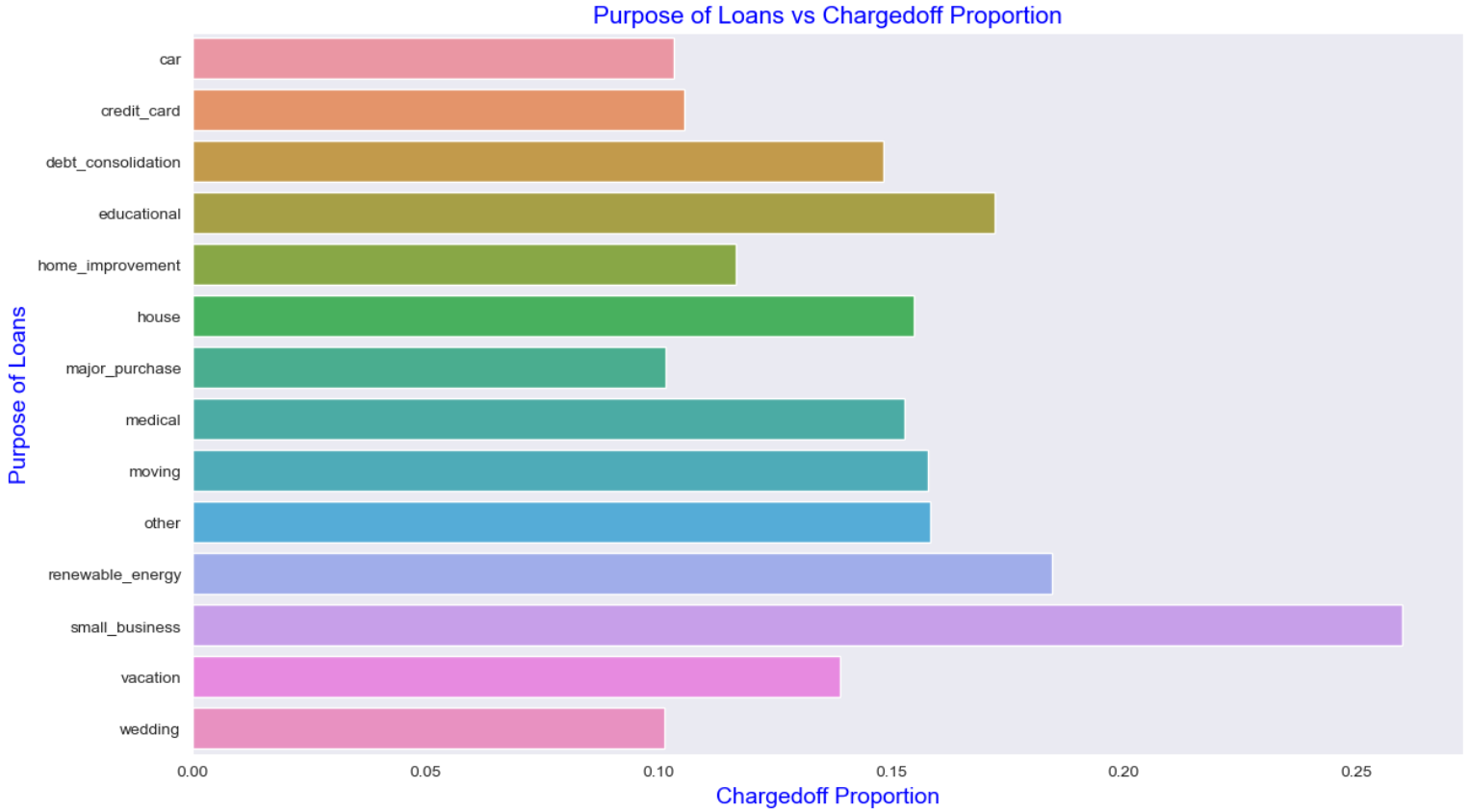


annual_inc_cats	Charged Off	Current	Fully Paid	Total	Chargedoff_Proportion
0-20000	237	9	943	1189	0.20
20000-40000	1514	170	7004	8688	0.17
40000-60000	1729	345	9534	11608	0.15
60000-80000	1024	240	6597	7861	0.13
80000 +	1122	376	8859	10357	0.11

Observations :

- Income range 80000+ has less chances of charged off.
- Income range 0 - 20000 has high chances of charged off.
- Annual in come is inversely proportional charged off

Bivariate Analysis – Purpose of Loan Vs Charged off

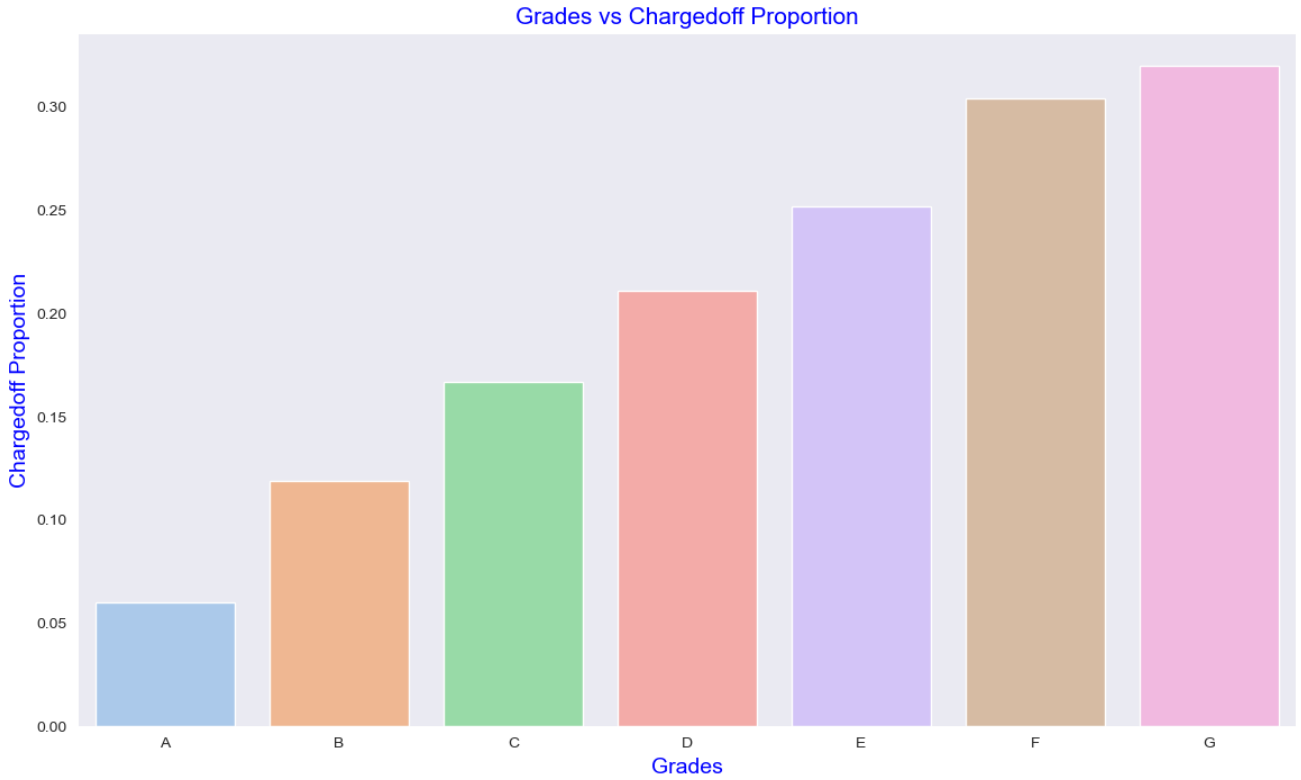


purpose	Chargedoff_Proportion
small_business	0.26
renewable_energy	0.18
educational	0.17
other	0.16
moving	0.16
house	0.15
medical	0.15
debt_consolidation	0.15
vacation	0.14
home_improvement	0.12
credit_card	0.11
car	0.10
major_purchase	0.10
wedding	0.10

Observations :

- Small Business applicants have high chances of getting charged off.

Bivariate Analysis – Grade Vs Charged off



grade	Charged Off	Current	Fully Paid	Total	Chargedoff_Proportion
G	101	17	198	316	0.32
F	319	73	657	1049	0.30
E	715	179	1948	2842	0.25
D	1118	222	3967	5307	0.21
C	1347	264	6487	8098	0.17
B	1425	345	10250	12020	0.12
A	602	40	9443	10085	0.06

Observations :

- 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"

Python Library Used for Analysis

Thank You