Project Name: Lending club case study

Lending Club provides loan based of "creditworthiness," Once Lending Club determines that an applicant is "creditworthy," it issues the applicant an "A" through "G" grade and a 1 through 5 subgrade based off of an applicant's credit history. 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. This project using data analysis tries to find the potential attributes for loan default.

Table of Contents

General Information

Lending Club enables borrowers to create loan listings on its website by supplying details about themselves and the loans that they would like to request. On the basis of the borrower's credit score, credit history, desired loan amount and the borrower's debt-to-income ratio, Lending Club determined whether the borrower was creditworthy and assigned to its approved loans a credit grade that determined the payable interest rate and fees. The standard loan period was three years; a five-year period was available at a higher interest rate and additional fees. Personal loan up to $ 40,000 and Business Loans ranging from $ 5,000 to $ 500,000 are provided

Background

The project is being submitted as partial fulfilment of Advanced Certificate in Machine Learning and Deeping Learning course conducted jointly by IIIT, Bangalore and Upgrad.

Business Problem

Lending loans to 'risky' applicants, who are likely to default , is the largest source of financial loss (called credit loss). The credit loss is the amount of money lost by the lender when the borrower refuses to pay or runs away with the money owed. In other words, borrowers who default cause the largest amount of loss to the lenders. In this case, the customers labelled as 'charged-off' are the 'defaulters'.

How consumer attributes and loan attributes influence the tendency of default.

Identify risky loan applicants who are likely to default. Identification of such applicants using EDA is the aim of this case study.

Dataset Used

Loan.csv

The data given contains the information about past loan applicants and whether they 'defaulted' or not. It contains the complete loan data for all loans issued through the time period 2007 t0 2011.

Approach

The dataset was loaded in python. Data was cleaned to address the data quality issues.

Loan status rows having current status were removed since it is not required for analysis.

Member id , id , url column not required for analysis were dropped.

Columns having more than 30% null values were dropped.

Numeric columns having % , + were cleaned to remove these characters and were converted to numeric.

Date was split to year and month for analysis.

Univariate, Segmented Univariate and bivariate analysis was done to analyse the data set.

Conclusions

Charge off percent was calculated for relevant categories.
Charge off Percent = Charge off / Total
where total = charge off + fully paid.

Interest Rate VS Charged off Percent

- Interest rate less than 10% has very less chances of charged off.
- Interest rate more than 15% has the highest chances of charged off.
- Charged off percent is increasing with higher interest rates.
- Highest number of loans are provided in interest rate range of 10% to 15%

loan_status	int_rate_buckets	Charged Off	Fully Paid	Total	Chargedoff_Percent
3	15 +	1794	5432	7226	24.83
2	10-15	2707	15558	18265	14.82
1	7.5-10	574	6372	6946	8.26
0	0-7.5	256	5114	5370	4.77

Grade VS Charged off Percent

- 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"

loan_status	grade	Charged Off	Fully Paid	Total	Chargedoff_Percent
6	G	101	198	299	33.78
5	F	319	657	976	32.68
4	E	715	1948	2663	26.85
3	D	1118	3967	5085	21.99
2	C	1347	6487	7834	17.19
1	B	1425	10250	11675	12.21
0	A	602	9443	10045	5.99

Annual Income VS Charged off Percent

- Higher Income range Above 80000 has less chances of charged off.
- Low Income range between 0 and 20000 has high chances of charged off.
- As income increases charged off percent decreases.

loan_status	annual_inc_buckets	Charged Off	Fully Paid	Total	Chargedoff_Percent
0	0-20000	237	943	1180	20.08
1	20000-40000	1514	7004	8518	17.77
2	40000-60000	1729	9534	11263	15.35
3	60000-80000	1024	6597	7621	13.44
4	80000 +	1122	8859	9981	11.24

Purpose VS Charged off Percent

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

loan_status	purpose	Charged Off	Fully Paid	Total	Chargedoff_Percent
11	small_business	475	1279	1754	27.08
10	renewable_energy	19	83	102	18.63
3	educational	56	269	325	17.23
9	other	633	3232	3865	16.38
5	house	59	308	367	16.08
8	moving	92	484	576	15.97
7	medical	106	575	681	15.57
2	debt_consolidation	2767	15288	18055	15.33
12	vacation	53	322	375	14.13
4	home_improvement	347	2528	2875	12.07
1	credit_card	542	4485	5027	10.78
0	car	160	1339	1499	10.67
13	wedding	96	830	926	10.37
6	major_purchase	222	1928	2150	10.33

Public Derogatory Record VS Charged off Percent

- Charged off percent is high where 1 or 2 derogatory public records exist.
- For 3 or 4 derogatory public records the data is insufficient.

loan_status	pub_rec	Charged Off	Fully Paid	Total	Chargedoff_Percent
1	1	457.00	1556.00	2013.00	22.70
2	2	10.00	38.00	48.00	20.83
0	0	5160.00	31347.00	36507.00	14.13
3	3	0.00	7.00	7.00	0.00
4	4	0.00	2.00	2.00	0.00

Public Recorded Bankruptcies VS Charged off Percent

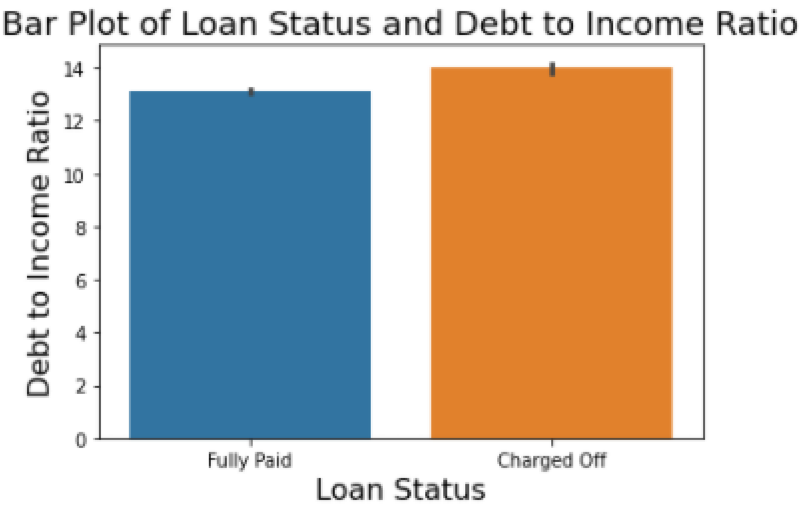
- Charged off percent is high where 1 or 2 public record bankruptcies exist.
- For few cases the public record bankruptcies is not known.

loan_status	pub_rec_bankruptcies	Charged Off	Fully Paid	Total	Chargedoff_Percent
2	2.00	2	3	5	40.00
1	1.00	366	1271	1637	22.36
3	NA	118	579	697	16.93
0	0.00	5141	31097	36238	14.19

Status of loan with respect to categories:

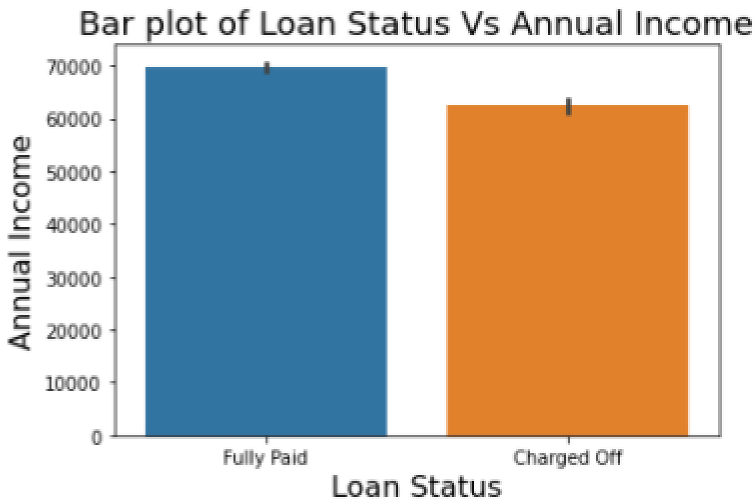
Debt to Income Ratio VS Loan Status

- Fully paid loans have a lower mean debt to income ratio of 13.15%.
- Charged off loan have a higher mean debt to income ratio 14%.



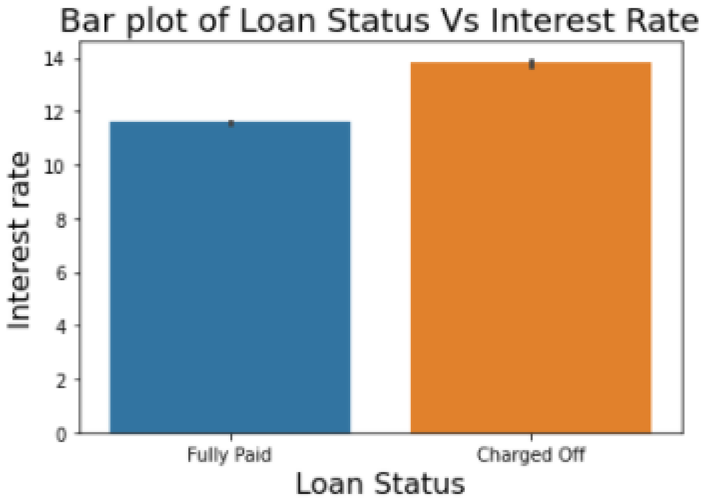
Annual Income VS Loan Status

- Fully paid loans have a higher mean annual income of around $69862.50.
- Charged off loan have a lower mean annual income of around $62427.30.



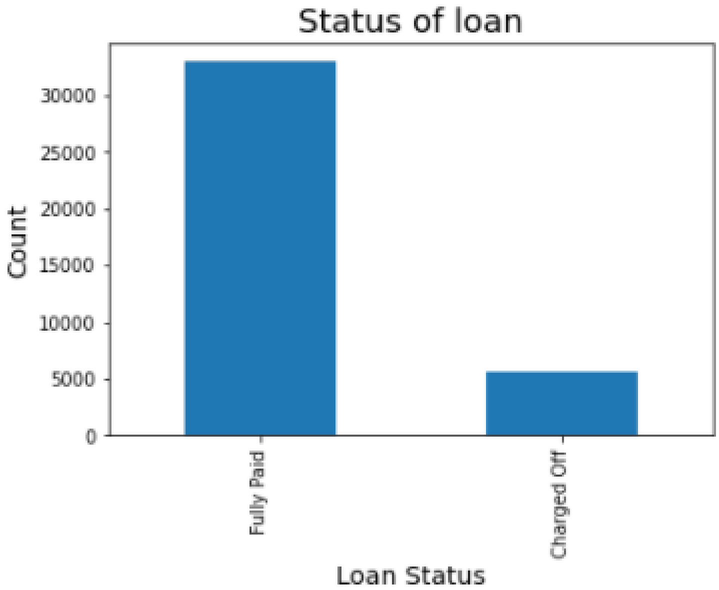
Interest VS Loan Status

- Fully paid loans have a lower mean interest rate of 11.61%.
- Charged off loan have a higher mean interest rate of 13.82%.



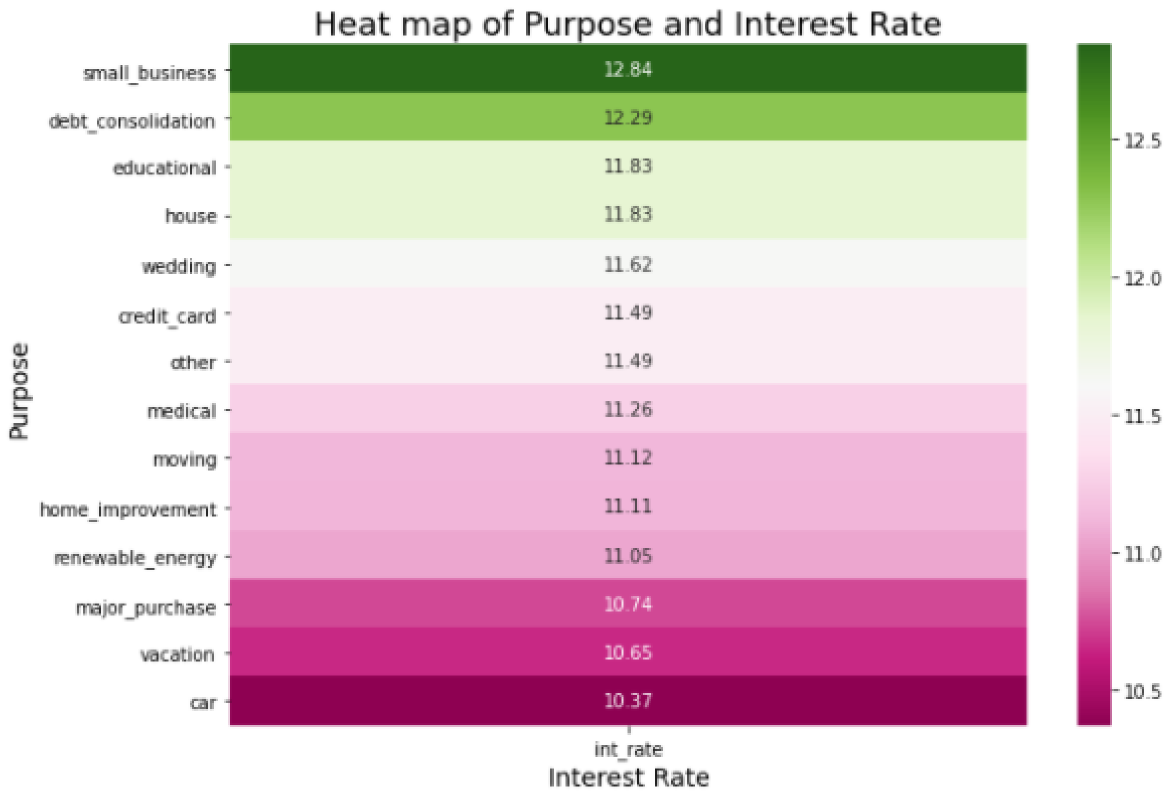
Loan Status

- 85.41% loan is fully paid and 14.59% loan is charged off



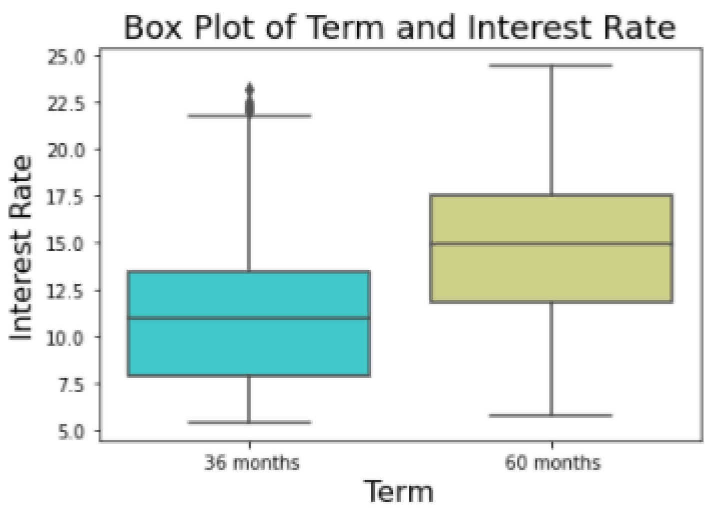
Interest rate VS Purpose

- Highest interest rate is charged for small business followed by debt consolidation, educational, house, wedding etc.
- Interest rate purpose wise varies from 12.84% to 10.37%



Interest rate VS Purpose

- 36 months loans have lower interest rates compared to 60 months interest rates.
- Median interest rate of 36 month loan = 10.99
- Median interest rate of 60 month loan = 14.91



Recommendations

The following should be closely monitored:

- Loan having Interest rate greater than 15% has high chances of charged off.
- Loan Grade "F" and "G" have very high chances of charged off. Low Income range between 0 and 20000 has high chances of charged off.
- Small Business applicants have high chances of getting charged off.
- Charged off percent is high where 1 or 2 derogatory public records exist.
- Charged off percent is high where 1 or 2 public record bankruptcies exist.

Technologies Used

Python	3.8.8
Pandas	1.2.4
Numpy	1.20.1
Matplotlib	3.3.4
Seaborn	0.11.1

Acknowledgements

This project was inspired by IIT, Bangalore and Upgrad, Bangalore.

References

https://en.wikipedia.org/wiki/LendingClub
https://www.lendingclub.com/
Python Documentation
Upgrad Study Material
https://corporatefinanceinstitute.com/resources/knowledge/finance/fico-score/
https://corporatefinanceinstitute.com/resources/knowledge/credit/purpose-of-credit-risk-analysis/
https://corporatefinanceinstitute.com/resources/knowledge/finance/debt-to-income-ratio/
https://stackoverflow.com/
https://pbpython.com/pandas-qcut-cut.html

Contact
Created by [@rohitrs0908]