

Lending club Case Study

Prepared by:
Akhilesh Dwivedi
Abrar Lohia

Agenda

- ▶ Objective
- ▶ Analyse loan data
- ▶ Cleaning of data
- ▶ Identify Key features
- ▶ Univariate analysis
- ▶ Bivariate analysis
- ▶ Data Visualisation
- ▶ Conclusion

Objective

- ▶ Identify risky loan Applicants
- ▶ Minimizing the loss

Data ingestion and analysis

- Loaded loan data and analysed the columns, null values and data types:

```
In [79]: # Load the data
df = pd.read_csv('loan.csv')
```

```
In [80]: df.shape
```

```
Out[80]: (39717, 111)
```

```
In [81]: df.head(10)
```

```
Out[81]:
```

	id	member_id	loan_amnt	funded_amnt	funded_amnt_inv	term	int_rate	installment	grade	sub_grade	...	num_tl_90g_dpd_24m	num_tl_op_past_1
0	1077501	1296599	5000	5000	4975.0	36 months	10.65%	162.87	B	B2	...	NaN	1
1	1077430	1314167	2500	2500	2500.0	60 months	15.27%	59.83	C	C4	...	NaN	1
2	1077175	1313524	2400	2400	2400.0	36 months	15.96%	84.33	C	C5	...	NaN	1

```
In [83]: ## find the dtypes
df.dtypes
```

```
Out[83]:
```

id	int64
member_id	int64
loan_amnt	int64
funded_amnt	int64
funded_amnt_inv	float64
...	...
tax_liens	float64
tot_hi_cred_lim	float64
total_bal_ex_mort	float64
...	...

Data cleaning

► Removal of missing values

```
In [5]: #removing null columns from data
df.dropna(axis = 1, how = 'all', inplace=True)
df.head()
```

► data type conversion

```
In [6]: #columns which are irrelevant for us in analysis
#print(df.columns)
df.drop(['id', 'member_id', 'url', 'title', 'emp_title', 'zip_code', 'last_credit_pull_d', 'addr_state', 'desc', 'out_prncp_inv', 'total_pymnt_'], axis=1, inplace=True)
df.head()
```

```
Out[6]:
```

	loan_amnt	funded_amnt_inv	term	int_rate	installment	grade	sub_grade	emp_length	home_ownership	annual_inc	...	inq_last_6mths	open_acc
0	5000	4975.0	36 months	10.65%	162.87	B	B2	10+ years	RENT	24000.0	...	1	3

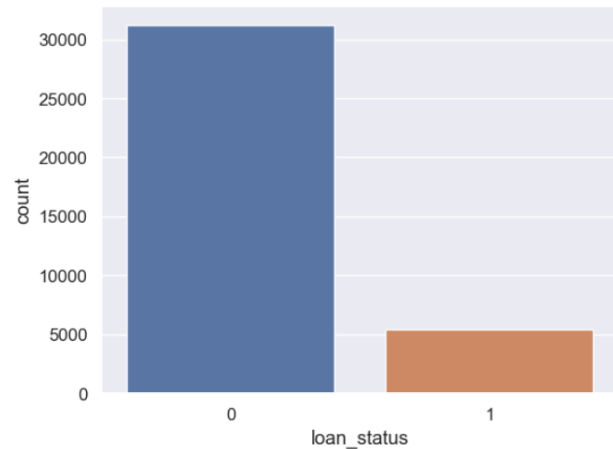
Key features

► Important variables are:

1. loan_amt
2. Term
3. Interest_rate
4. Grade
5. Subgrade
6. Annual Income
7. Purpose of loan
8. DTI
9. Emp_lenght
10. Loan date
11. Home Ownership
12. Verification Status

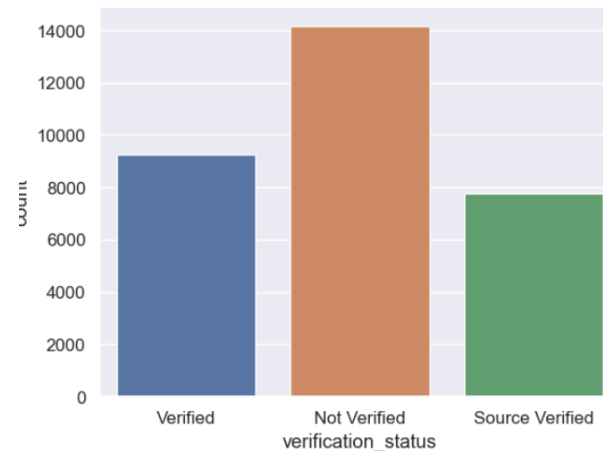
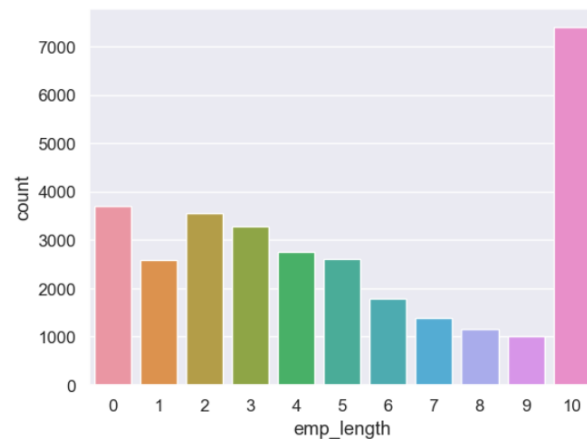
Univariate Analysis

- ▶ More than 5K loans are risky.
- ▶ 0: fully paid 1: Charged off
- ▶ Maximum number of people who fully paid the loan are having rented home.



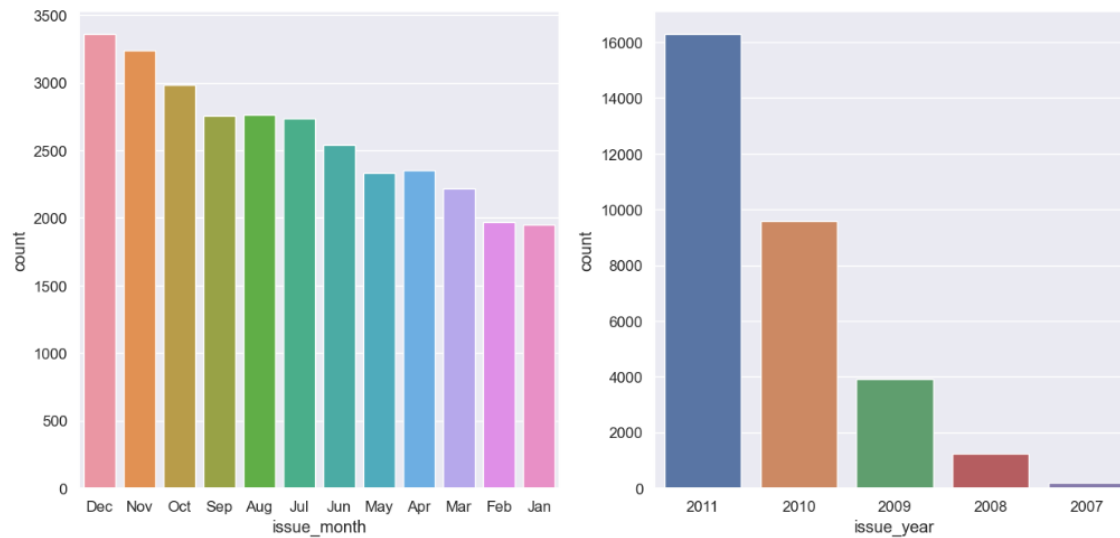
Continued...

- ▶ Maximum number of applicants who paid the loan were having service length 10 year or more.
- ▶ Maximum no of applicants who paid loan fully were not verified.



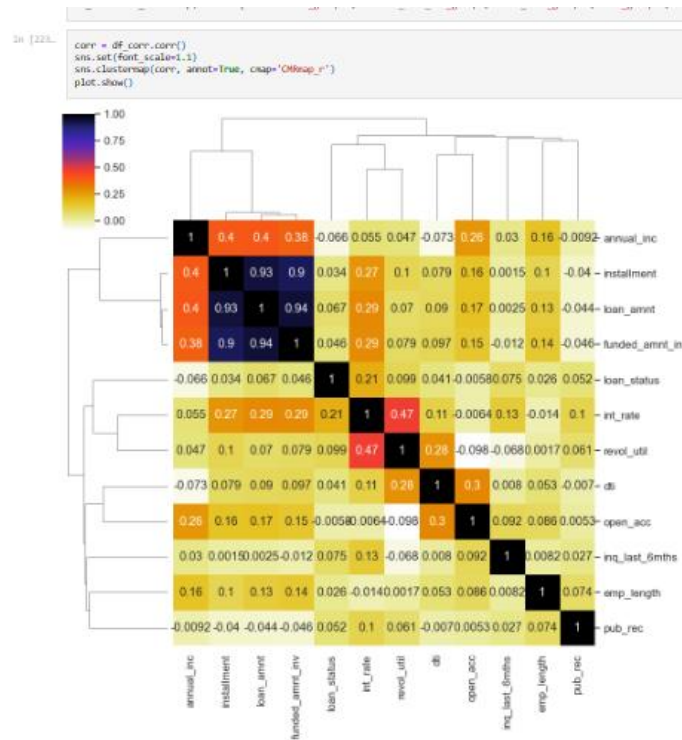
Continued....

- ▶ Maximum loan we given in year 2011
- ▶ Dec is the month when maximum loans are disbursed



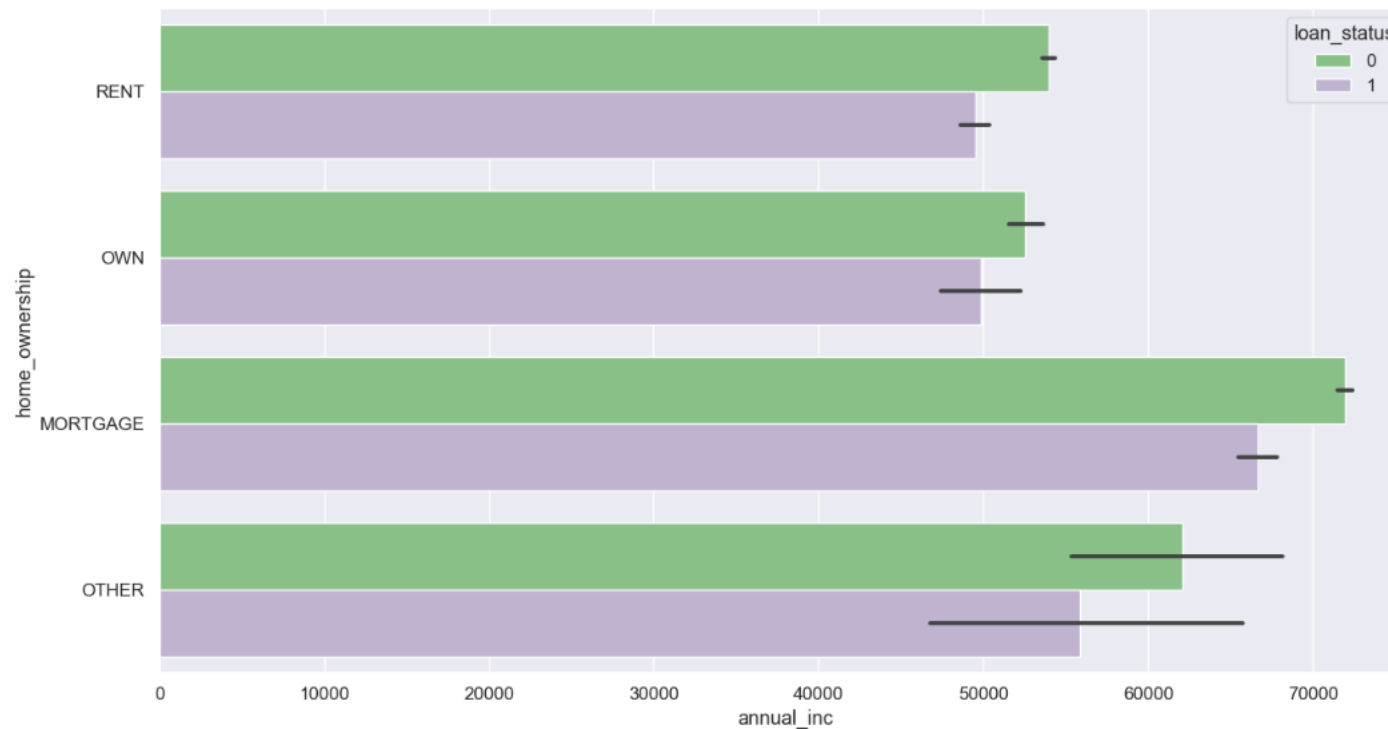
Correlation

- ▶ Annual Income to dti Ratio are negatively correlated
- ▶ Loan Amount,Investor Amount and Funding Amount are strongly correlated
- ▶ Positive correlation between Annual Income and emp_length



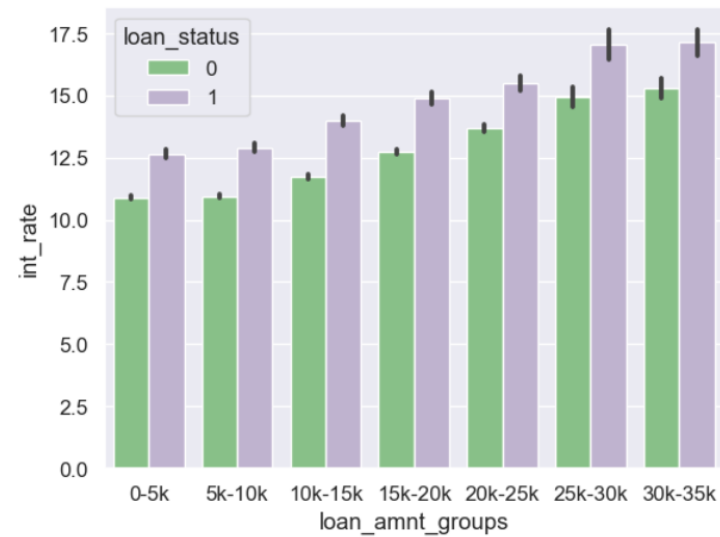
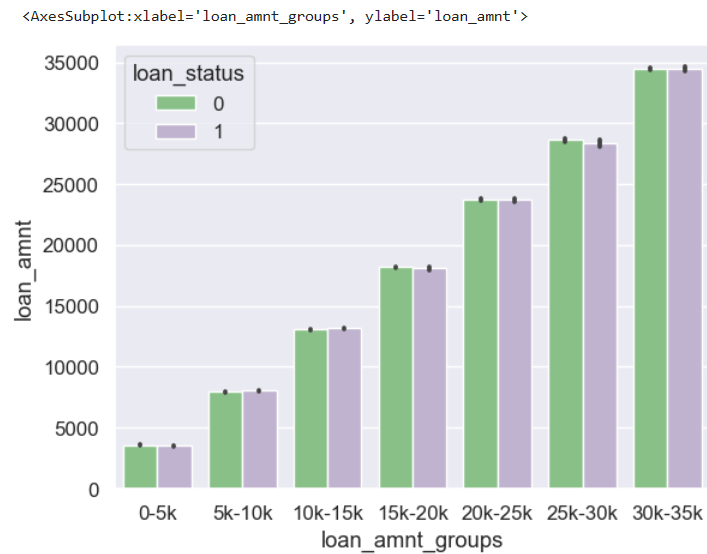
Bivariate Analysis

- ▶ Maximum number of applicants who didn't paid load are having home ownership as Mortgage.
- ▶ People with rented home defaulted the loan least.



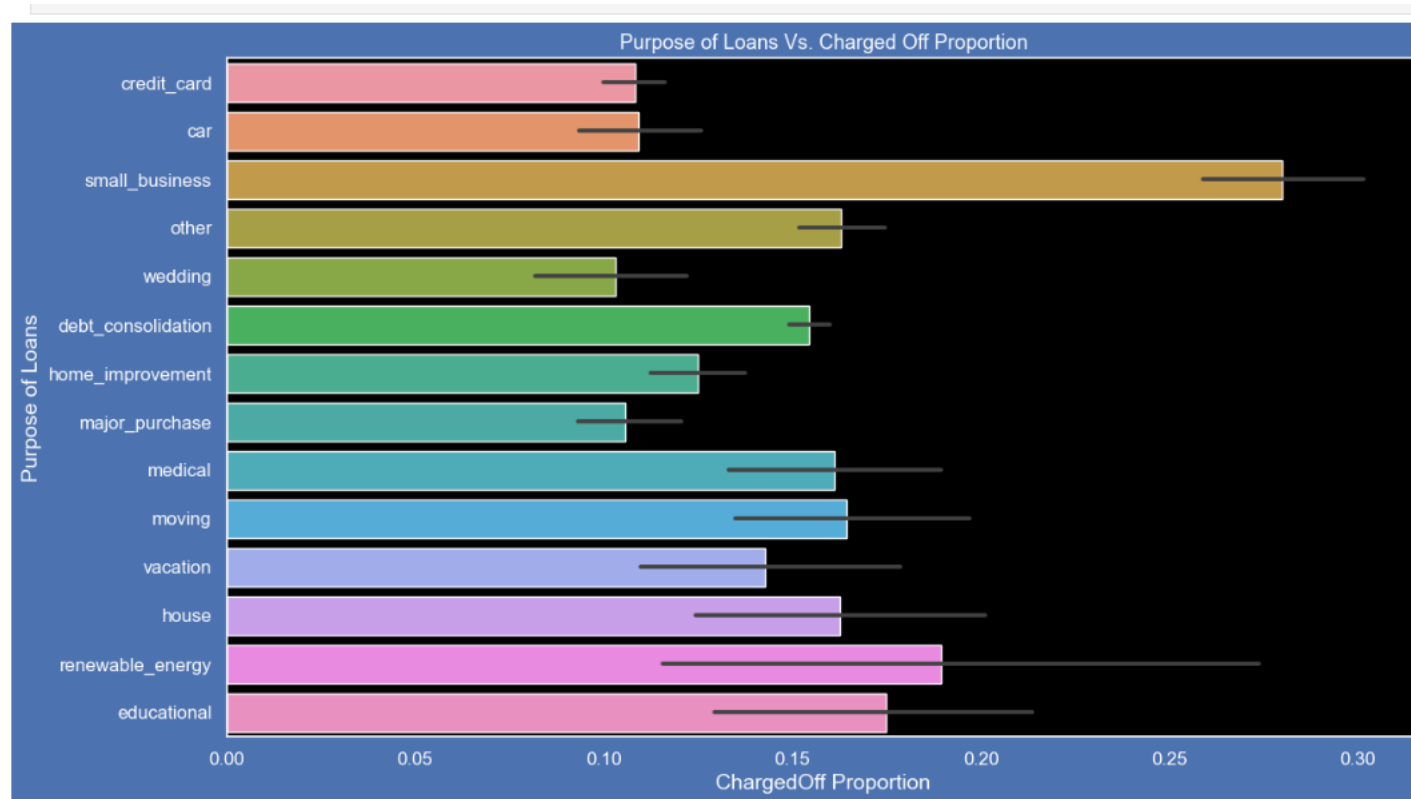
Continued...

- Applicants with 30K-35K loan amount are paying maximum interest.

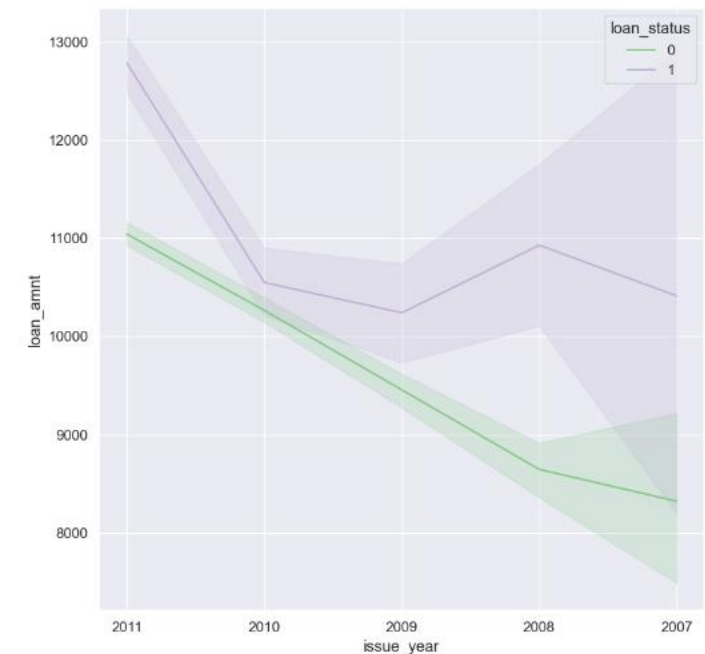
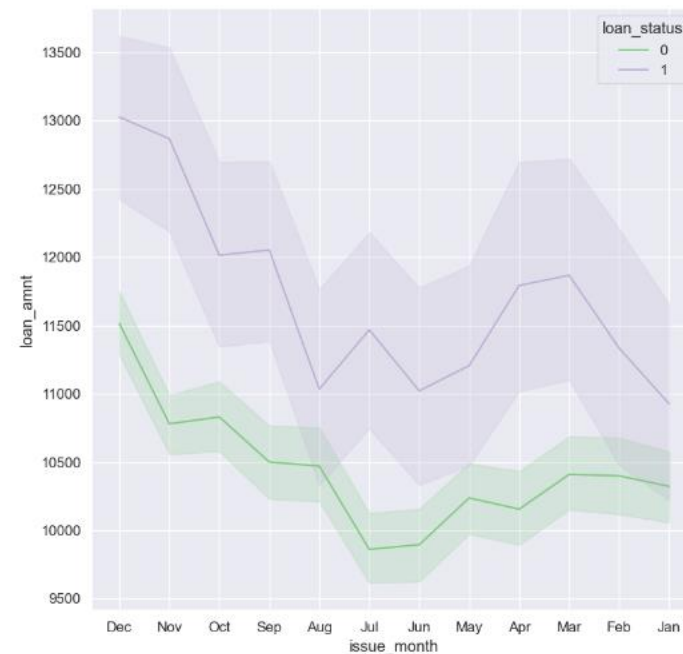


Continued...

- Loan for Small business are most risky.



Important visuals



Conclusion

- ▶ Loan with higher interest rate are having most chances to be charged off.
- ▶ Loan amount is directly proportional to annual income.
- ▶ Applicant taking loan have income between 60-70K
- ▶ Applicant having home_ownership status as mortgage have income between 60-70K
- ▶ Applicant having income ranging 30-35K have interest rate of 15-17.5%
- ▶ Applicant who have taken home ownership status as mortgage have loan amount between 12-14K
- ▶ Employees working for more than 10 years and whose source are verified tend to get loan amount above 16K
- ▶ Applicant who have taken loan for small business have amount greater than 14K