

LENDING CLUB CASE STUDY

--PRADEEP HEGDE

DATA METADATA CHECK

Shape of
Dataframe

Dataset
Information

Dataset
Datatypes

```
Number of rows and Columns: (39717, 111)
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 39717 entries, 0 to 39716
Columns: 111 entries, id to total_il_high_credit_limit
dtypes: object(111)
memory usage: 33.6+ MB
Data set INFO : None
Data Types: id                                     object
            member_id                               object
            loan_amnt                               object
            funded_amnt                            object
            funded_amnt_inv                      object
            ...
            tax_liens                                object
            tot_hi_cred_lim                         object
            total_bal_ex_mort                        object
            total_bc_limit                           object
            total_il_high_credit_limit               object
Length: 111, dtype: object
```

DATA CLEANING



UNIVARIATE ANALYSIS - UNORDERED CATEGORICAL VARIABLES FROM CHARGED OFF LIST- HOME OWNERSHIP

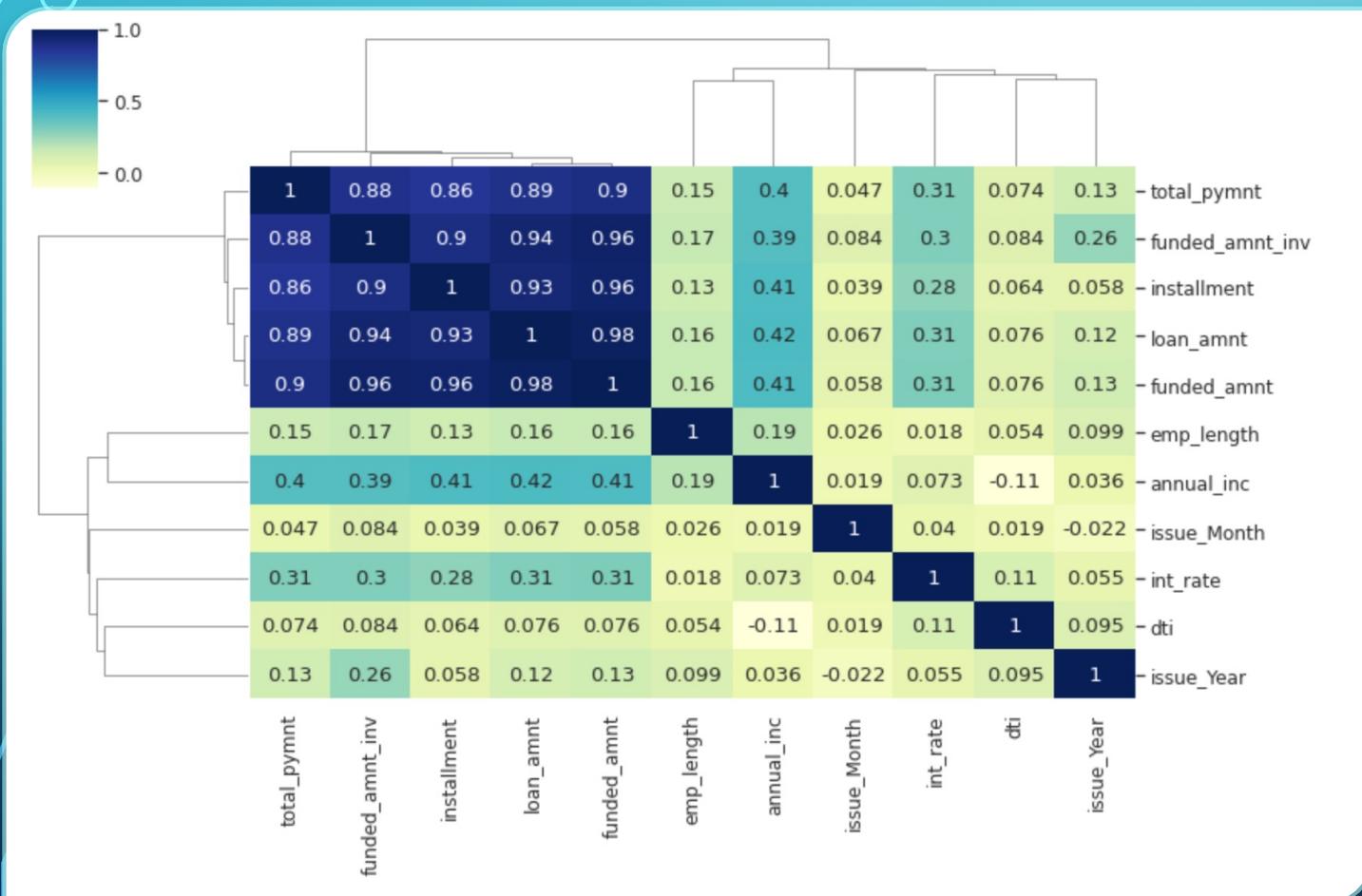
Count of loan_status	Column Labels	MORTGAGE	OTHER	OWN	RENT	Grand Total
Row Labels						
car		70	1	15	74	160
credit_card		229	2	37	269	537
debt_consolidation		1137	7	194	1414	2752
educational		12		5	38	55
home_improvement		258	1	41	38	338
house		19		3	36	58
major_purchase		75		26	119	220
medical		42		10	54	106
moving		8	1	6	76	91
other		185	2	60	382	629
renewable_energy		9		1	9	19
small_business		212	4	35	221	472
vacation		13		4	36	53
wedding		33		3	58	94
Grand Total		2302	18	440	2824	5584

Chart shows that most of them living in rented home or mortgaged their home.

Applicant numbers are high from these categories so charged off is high too.

Number shows that most of the loans were taken for the purpose of debt consolidation & paying credit card bill.

BIVARIATE ANALYSIS - CORRELATION MATRIX



- Observation is loan_amount, investor amount, funded amount are highly correlated
- Employee length increase , annual income increases - Positive correlation
- Annual income to DTI is negative correlated - when annual income is low , dti is high and vice versa.

BIVARIATE ANALYSIS ON ANNUAL INCOME AGAINST CHARGEDOFF PERCENTAGE

Observations:

- Income range 80000+ has less chances of charged off.
- Income range 0-20000 has high chances of charged off.
- Notice that with increase in annual income charged off proportion got decreased.

loan_status	annual_inc_breakup	Charged Off	Current	Fully Paid	Total	Chargedoff_Proportion
0	0-20000	237	9	943	1189	0.199327
1	20000-40000	1514	170	7004	8688	0.174263
2	40000-60000	1729	345	9534	11608	0.148949
3	60000-80000	1024	240	6597	7861	0.130263
4	80000 +	1080	362	8531	9973	0.108292

BIVARIATE ANALYSIS ON GRADE AGAINST CHARGEDOFF_PROPORTION

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"

loan_status	grade	Charged Off	Current	Fully Paid	Total	Chargedoff_Proportion
6	G	99	17	190	306	0.323529
5	F	312	71	642	1025	0.304390
4	E	704	176	1906	2786	0.252692
3	D	1112	216	3916	5244	0.212052
2	C	1342	262	6411	8015	0.167436
1	B	1414	344	10167	11925	0.118574
0	A	601	40	9377	10018	0.059992

BIVARIATE ANALYSIS ON EMPLOYMENT LENGTH AGAINST CHARGEDOFF_PROPORTION

Observations:

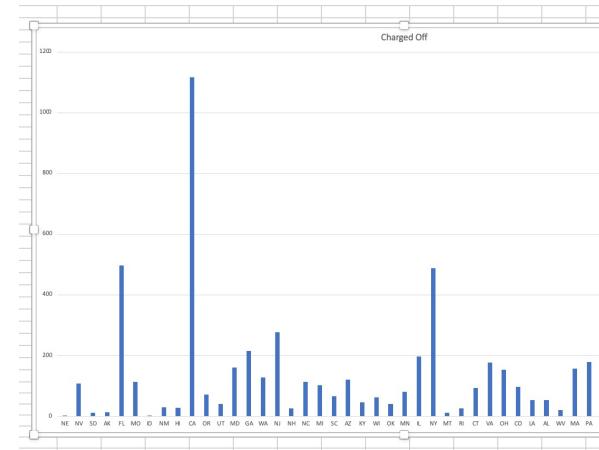
- Those who are not working or have less than 1 year of work experience have high chances of getting charged off.
- It makes sense as with less or no experience they don't have source of income to repay loan.
- Rest of the applicants have more or less same chances of getting charged off.

A	B	C	D	E	F
Employee Length	Charged Off	Current	Fully Paid	Grand Total	chargedoff-Percentage
0	2003200	575050	6665875	9244125	22%
1	11349075	2107050	62639800	76095925	15%
2	5978100	1586825	36475625	44040550	14%
3	6225050	1283625	35649700	43158375	14%
4	5212350	1493325	30542225	37247900	14%
5	5485350	1521675	29562625	36569650	15%
6	3842025	887975	20534550	25264550	15%
7	3356375	1026225	16140125	20522725	16%
8	2649325	708225	13864300	17221850	15%
9	2040500	517175	12354250	14911925	14%
10	19105725	7387700	87011250	113504675	17%

BIVARIATE ANALYSIS ON ADDRESS STATE AGAINST CHARGEDOFF_PROPORTION

Observations:

- States NE has very high chances of charged off but number of applications are too low to make any decisions.
- NY, CA and FL states shows good number of charged offs in good number of applications.



State	Charged Off	Applications	Charged Off %
FL	497	85	2252 2834 18%
MO	114	16	551 681 17%
ID	1	0	5 6 17%
NM	30	5	149 184 16%
HI	28	8	138 174 16%
CA	1116	149	5767 7032 16%
OR	71	16	364 451 16%
UT	40	6	211 257 16%
MD	161	26	849 1036 16%
GA	215	38	1135 1388 15%
WA	127	22	684 833 15%
NJ	277	59	1490 1826 15%
NH	25	5	139 169 15%
NC	114	38	629 781 15%
MI	102	16	596 714 14%
SC	66	13	389 468 14%
AZ	121	30	718 869 14%
KY	45	13	266 324 14%
WI	63	20	376 459 14%
OK	40	12	246 298 13%
MN	81	10	522 613 13%
IL	196	47	1262 1505 13%
NY	488	111	3156 3755 13%
MT	11	2	72 85 13%

THANK YOU

