LendingClub

Data Cleaning & Analysis

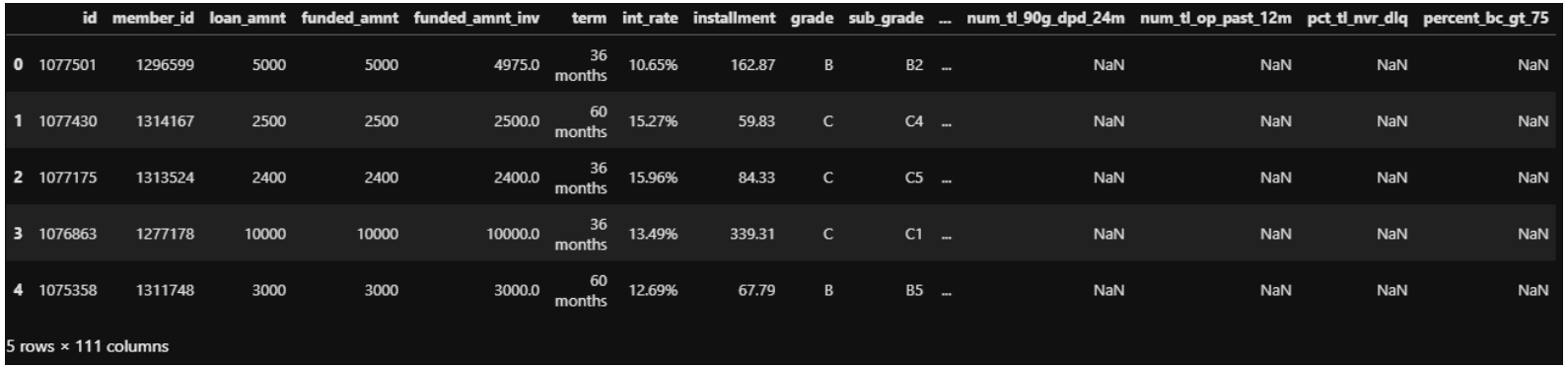
Objective

We need to find patterns in the data to make recommendations to the investors to avoid funding to any future loan applicants who have higher chances of default.

What we have?

We have 2 files:

1. Dataset: It contains records of all the customers whose loan application have been accepted



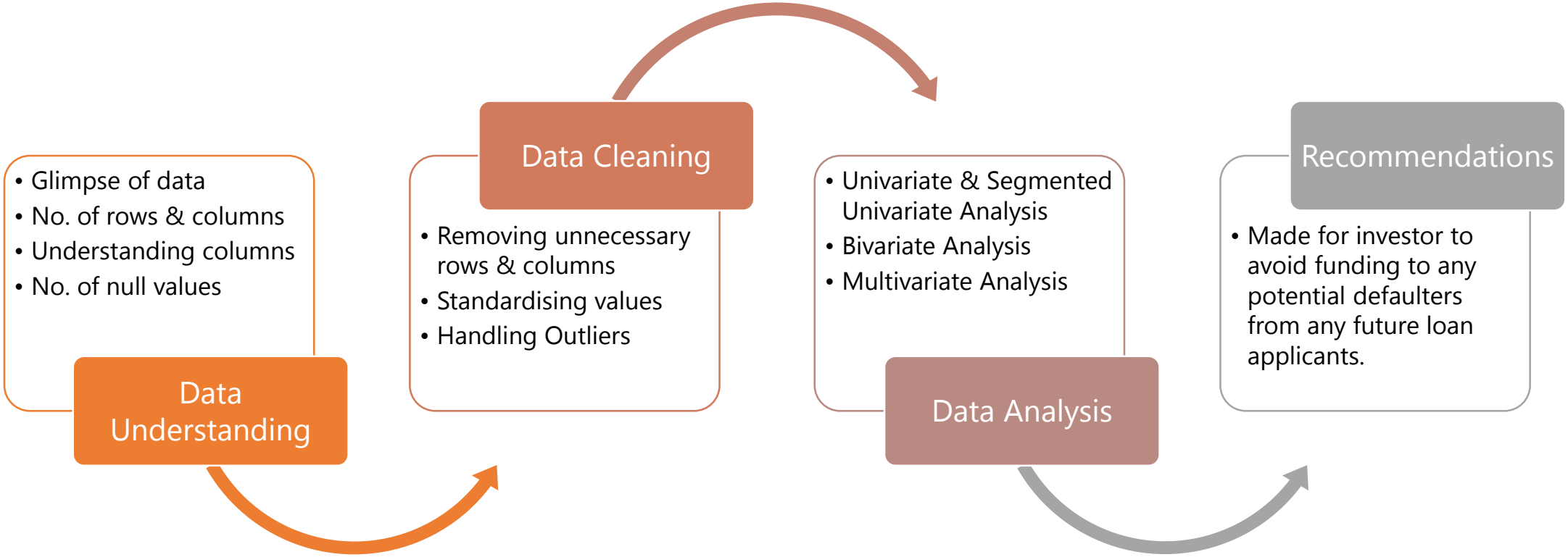
	id	member_id	loan_amnt	funded_amnt	funded_amnt_inv	term	int_rate	installment	grade	sub_grade	...	num_tl_90g_dpd_24m	num_tl_op_past_12m	pct_tl_nvr_dlq	percent_bc_gt_75
0	1077501	1296599	5000	5000	4975.0	36 months	10.65%	162.87	B	B2	...	NaN	NaN	NaN	NaN
1	1077430	1314167	2500	2500	2500.0	60 months	15.27%	59.83	C	C4	...	NaN	NaN	NaN	NaN
2	1077175	1313524	2400	2400	2400.0	36 months	15.96%	84.33	C	C5	...	NaN	NaN	NaN	NaN
3	1076863	1277178	10000	10000	10000.0	36 months	13.49%	339.31	C	C1	...	NaN	NaN	NaN	NaN
4	1075358	1311748	3000	3000	3000.0	60 months	12.69%	67.79	B	B5	...	NaN	NaN	NaN	NaN

5 rows × 111 columns

2. Data dictionary: It contains a detail description about each column in the dataset

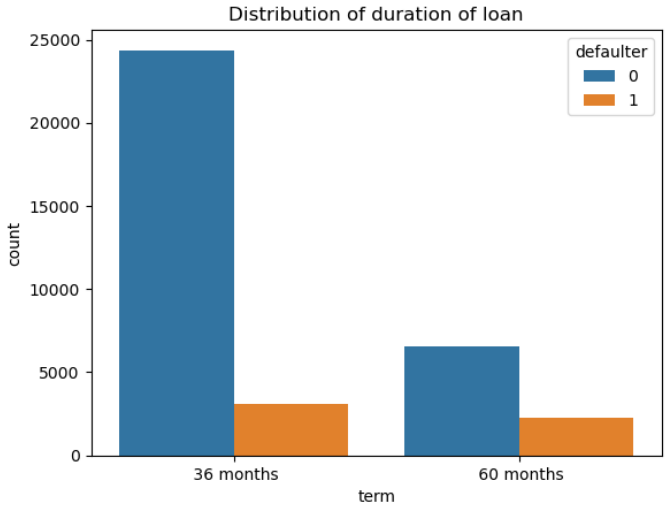
LoanStatNew	Description
acc_now_delinq	The number of accounts on which the borrower is now delinquent.
acc_open_past_24mths	Number of trades opened in past 24 months.
addr_state	The state provided by the borrower in the loan application
all_util	Balance to credit limit on all trades
annual_inc	The self-reported annual income provided by the borrower during registration.
annual_inc_joint	The combined self-reported annual income provided by the co-borrowers during registration
application_type	Indicates whether the loan is an individual application or a joint application with two co-borrowers

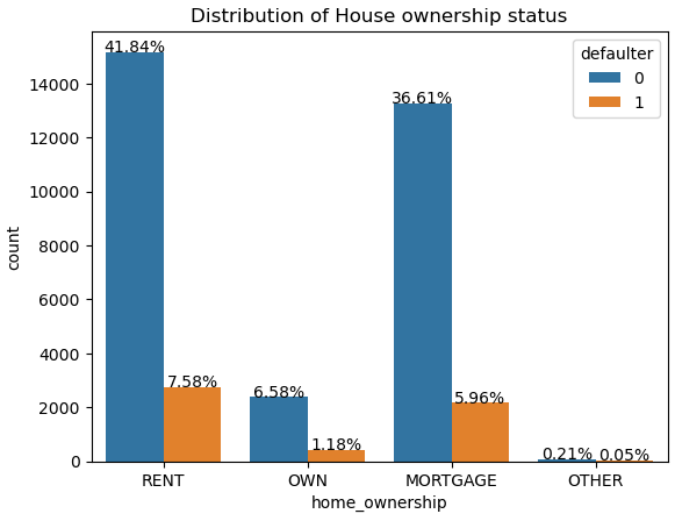
Plan of Action

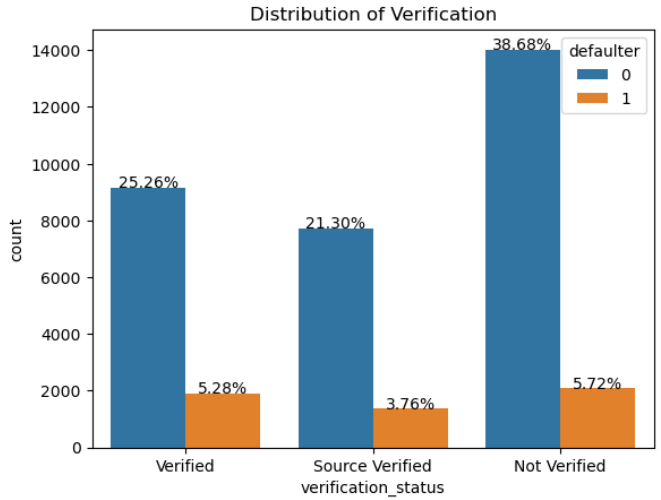


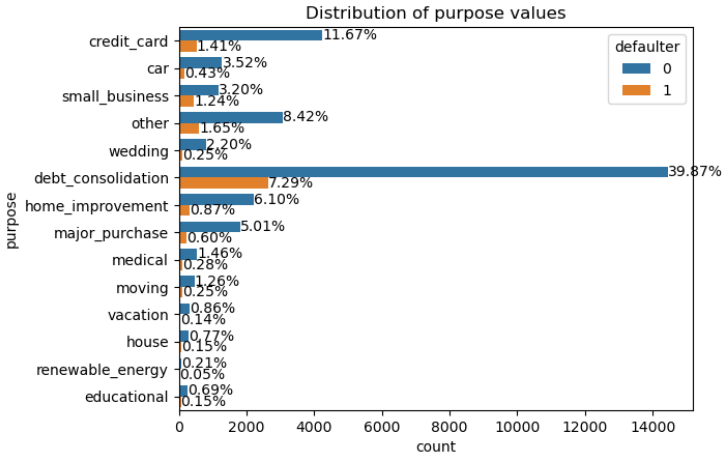
Univariate Analysis

- **Unordered Categorical Variable :** term, home_ownership, verification_status & purpose









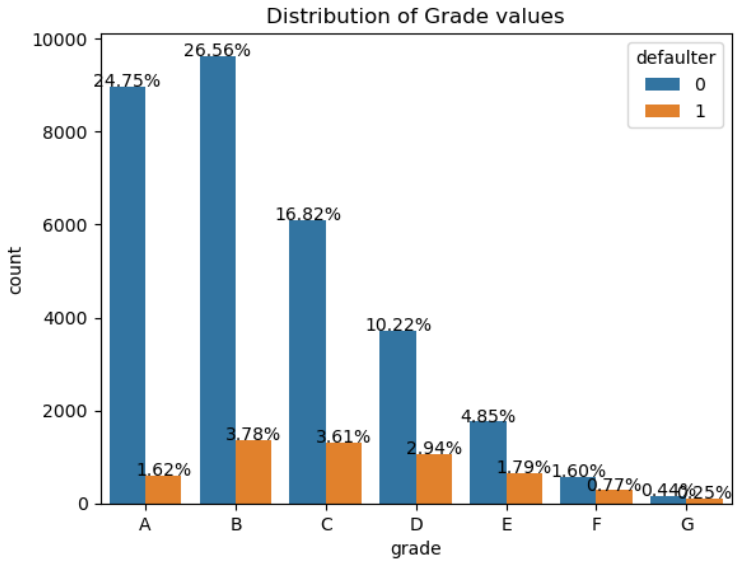
Recommendations:

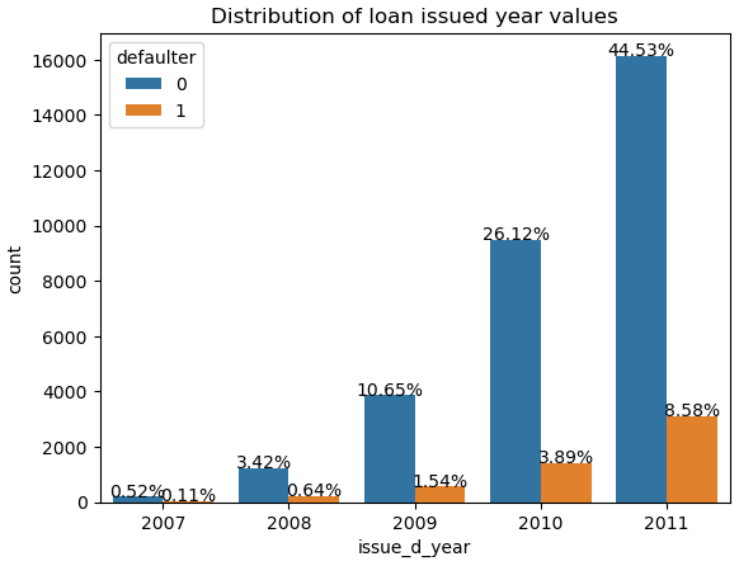
Based on the this analysis, we can say that the chances of a person defaulting on loan is high when

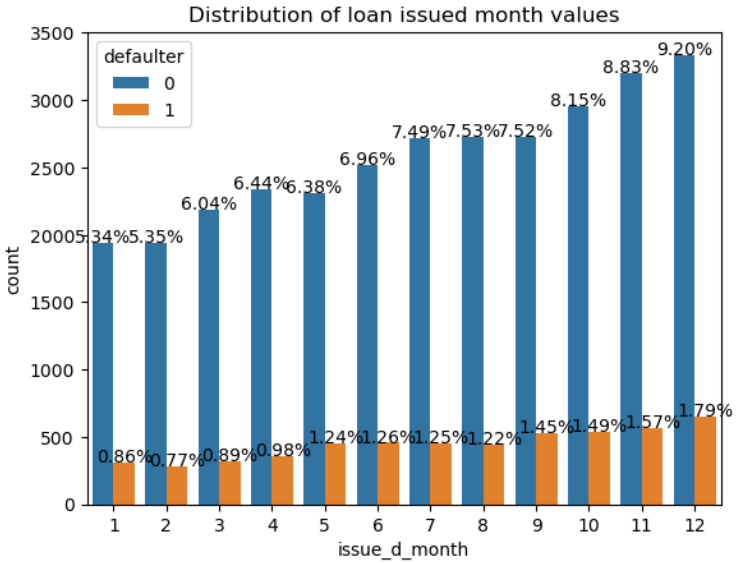
1. term = '60 months'
2. home_ownership = 'OTHER'
3. verification_status = 'Verified'
4. purpose = 'Small_business'

Univariate Analysis

- **Ordered Categorical Variable :** grade, issue_d_year, issue_d_month







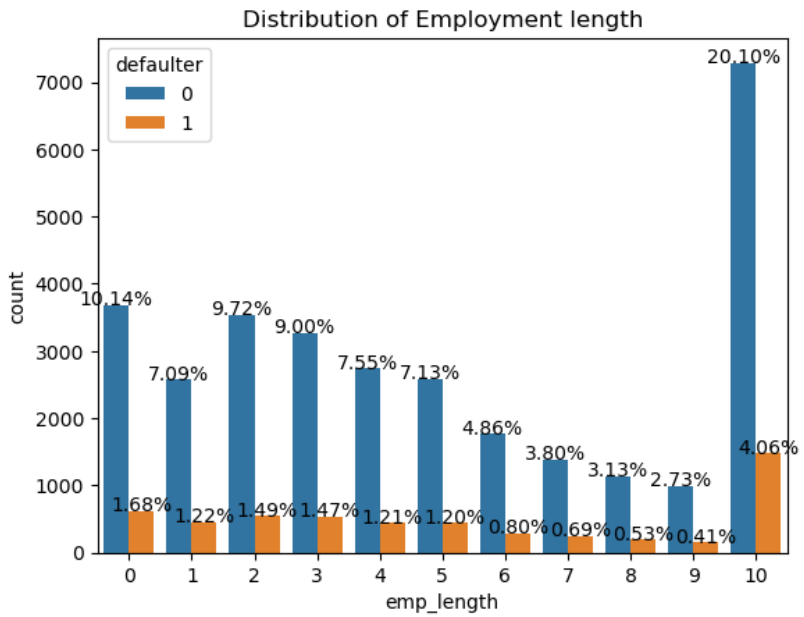
Recommendations:

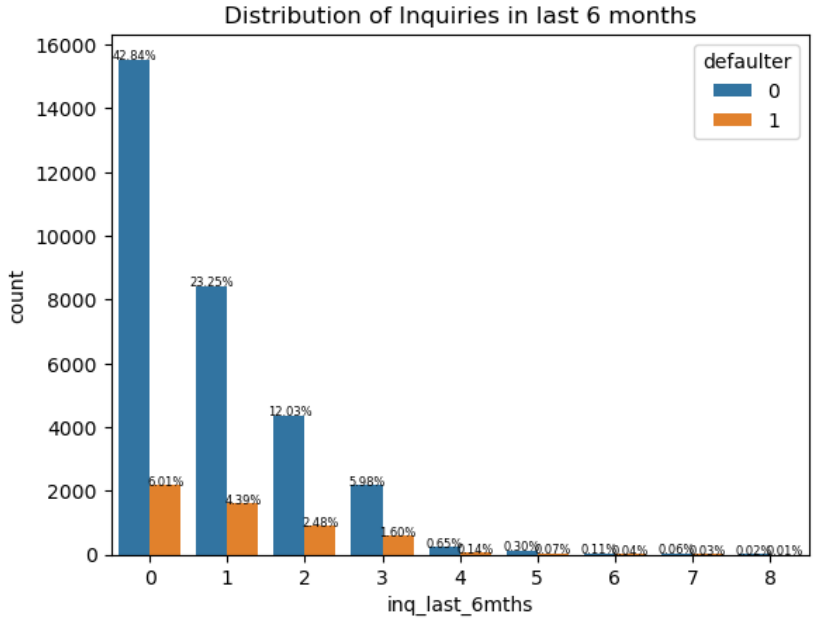
Based on this analysis, we can say that the chances of a person defaulting on loan is high when

1. grade= ‘G’
2. issue_d_year = 2007
3. issue_d_month= 5 or 12

Univariate Analysis

- **Quantitative Variable :** emp_length, inq_last_6mths



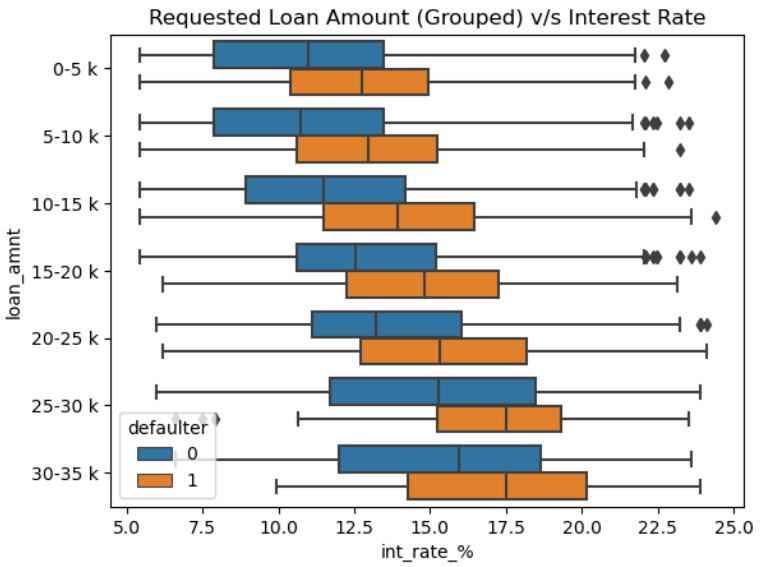


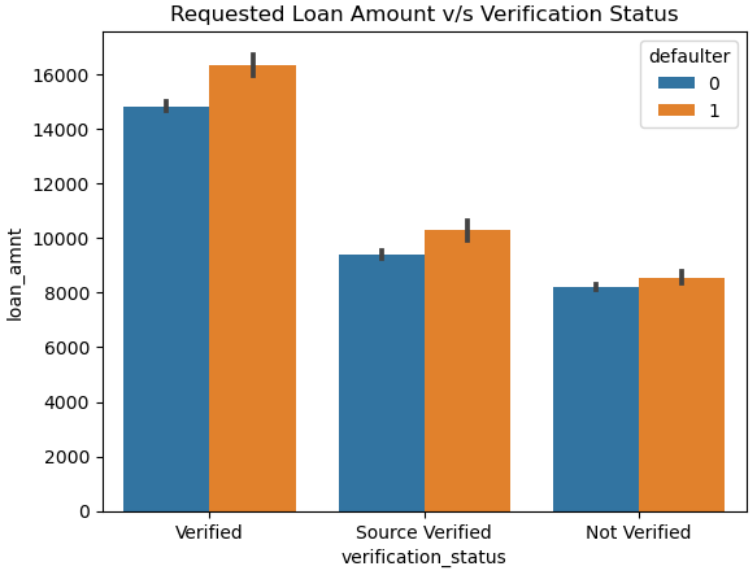
Recommendations:

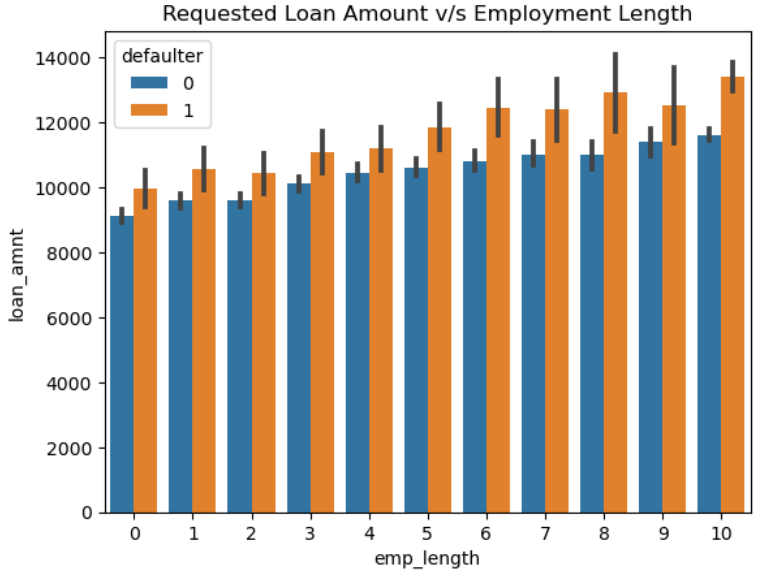
Based on this analysis, we can say that the chances of a person defaulting on loan is high when

1. emp_length= 10
2. inq_last_6mths is 7 or 8

Bivariate Analysis





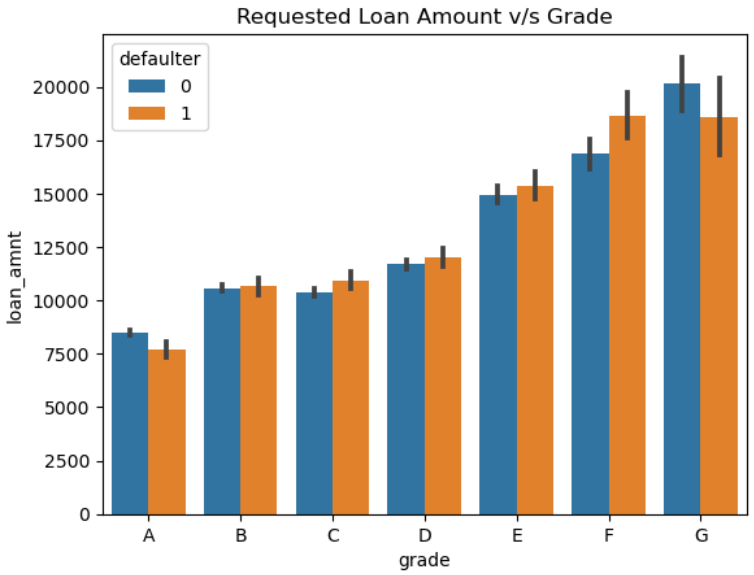


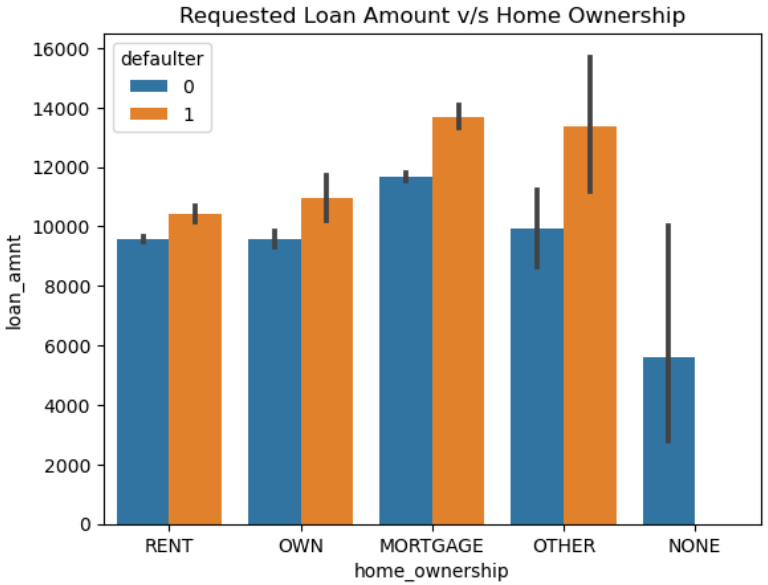
Recommendations:

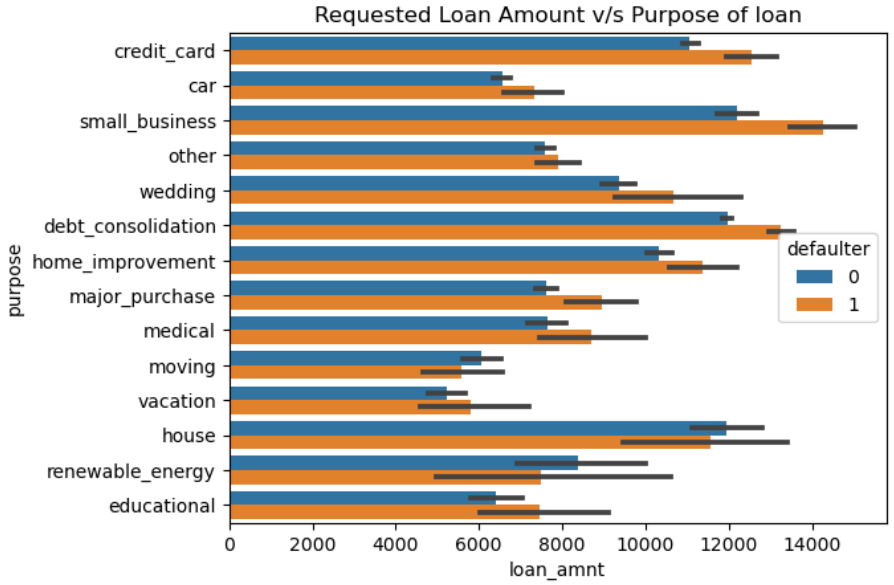
Based on this analysis, we can say that the chances of a person defaulting on loan is high when

1. int_rate_% is between 14.5 to 20 & loan_amnt is between 30 to 35 k
2. verification_status='Verified' & loan_amnt > 16,000
3. emp_length = 10 & loan_amnt is b/w 12,000 to 14,000

Bivariate Analysis





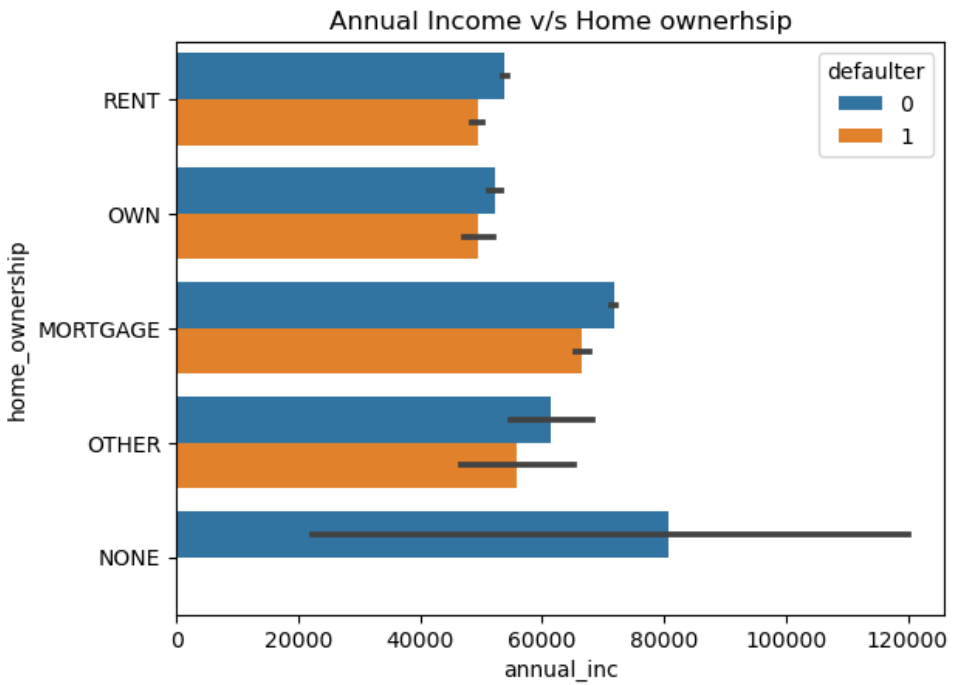


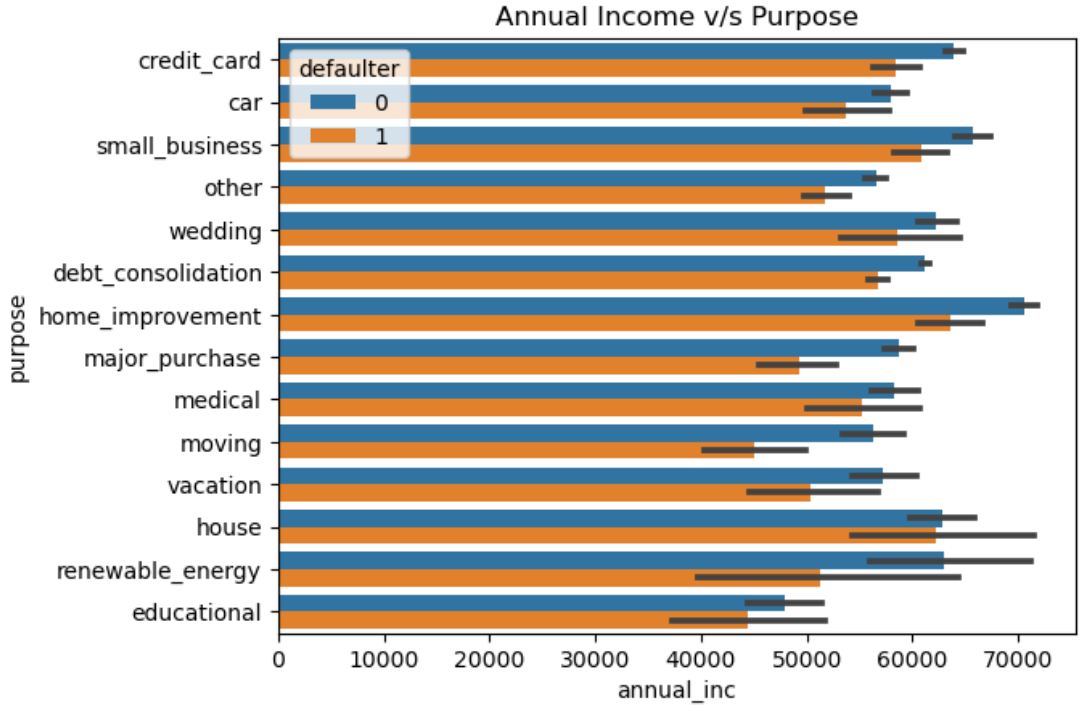
Recommendations:

Based on this analysis, we can say that the chances of a person defaulting on loan is high when

1. int_grade is either 'F' or 'G' & loan_amnt is b/w 17,500 to 20,000
2. home_ownership is either 'MORTGAGE' or 'OTHER' & loan_amnt is b/w 12,000 to 14,000
3. purpose is any one of these - credit_card, small_business, debt_consolidation - and loan_amnt is b/w 12,000 to 14,000

Bivariate Analysis





Recommendations:

Based on this analysis, we can say that the chances of a person defaulting on loan is high when

1. home_ownership is 'MORTGAGE' and annual_inc is b/w Rs 60,000 to 80,000.
2. purpose is either 'house' or 'home_improvement' and annual_inc is b/w Rs 60,000 to 70,000

GitHub Repository Link

https://github.com/AnirudhJayant06/LendingClub-Data_Cleaning_and_Analysis