GRAMENER CASE STUDY

SUBMISSION

Group Name:

1. Ankit Singhai
2. Mohak Bumb
3. Saurabh Jain
4. Vamsi Timmaraju

Abstract

Summary

"Consumer Finance Company" specialize in lending loans to urban customers. They are interested in understanding the **driving factors (or driver variables)** behind loan default, which can help in identifying such loans applications and improving there portfolio and risk assessment .

Design

The analysis is designed aligned to exploring data set, identifying variables aligned to domain. Some important variables which we consider from domain are as below

- *Charged Off*: Loan accounts which are defaulted, will be a primary variable in relation to which other variable are analysed.
- *Fully Paid*: Loan accounts which are fully paid, will serve as an important variable for ratio and comparative analysis.
- *Open accounts*: Number of open credit lines which can help us to understand a credit scoring for an application
- *Interest rate*: Interest rate at which loan is granted

Assumption

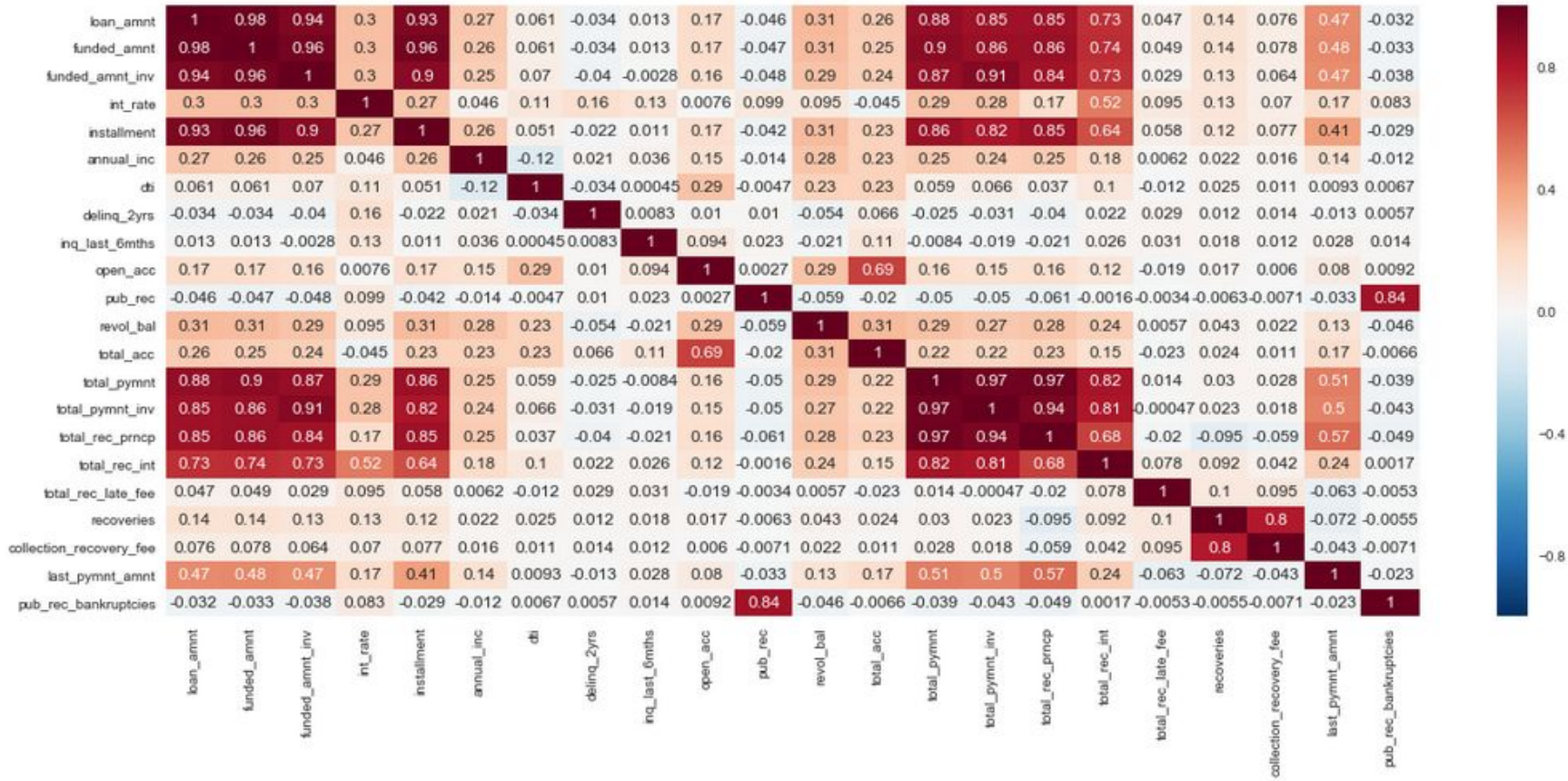
As we can't determine the status of Current accounts whether they will be defaulted or not we are not going to consider them for analysis. The ratio analysis will be based on Charged-Off and Fully Paid accounts.

Exploring Data Sets

We try explore the complete dataset *holistically* by establishing a correlation among them as to understand how various variables are related.

Assumption : To establish a correlation for some of the categorical variables we use label encoding to represent categorical bins to number.

Note: In next consecutive exploratory analysis we are going to detail out specific variables



Exploring Data Sets

We try explore the complete dataset *holistically* by establishing a correlation among them as to understand how various variables are related.

Assumption : To establish a correlation for some of the categorical variables we use label encoding to represent categorical bins to number.

Note: In next consecutive exploratory analysis we are going to detail out specific variables

Fig: Loan amount relationship with charged off

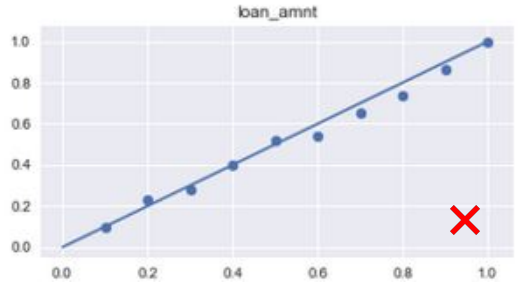


Fig: Int. rate relationship with charged off

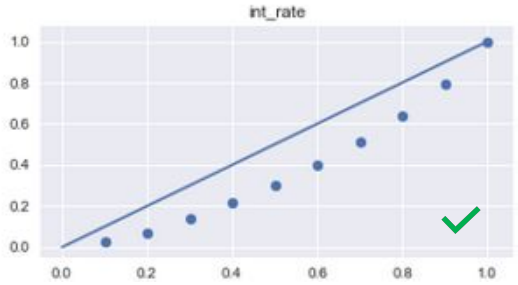


Fig: annual income relationship with charged off

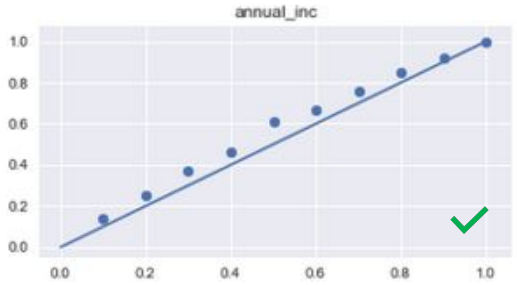


Fig DTI relationship with charged off

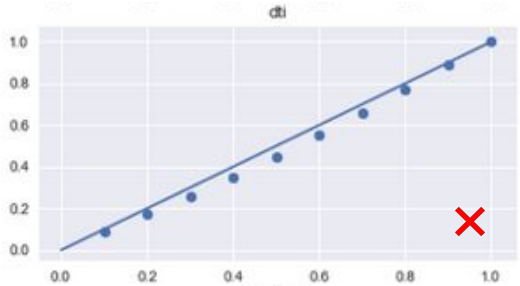


Fig: Int. Open Accounts relationship with charged off

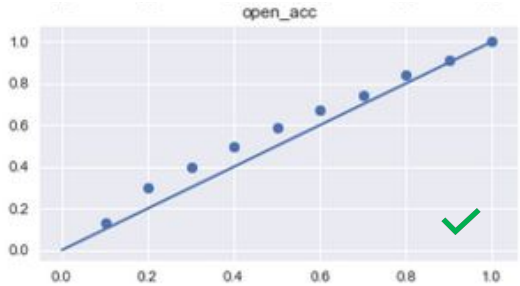


Fig: revolving balance relationship with charged off

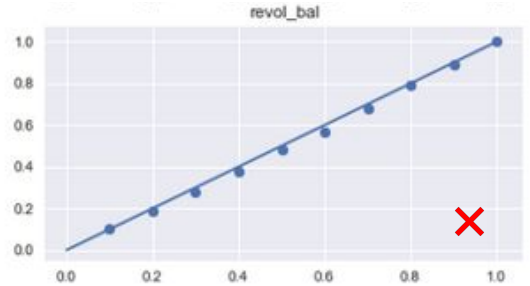
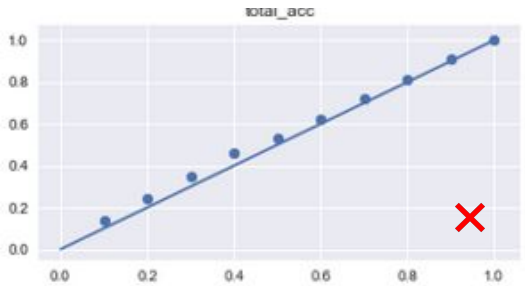


Fig Total account relationship with charged off



Details: Decile Distribution

- X Axis: column decile
- Y Axis: Cumulative % Default

Legends

✓ variable to consider

✗ variable not to consider

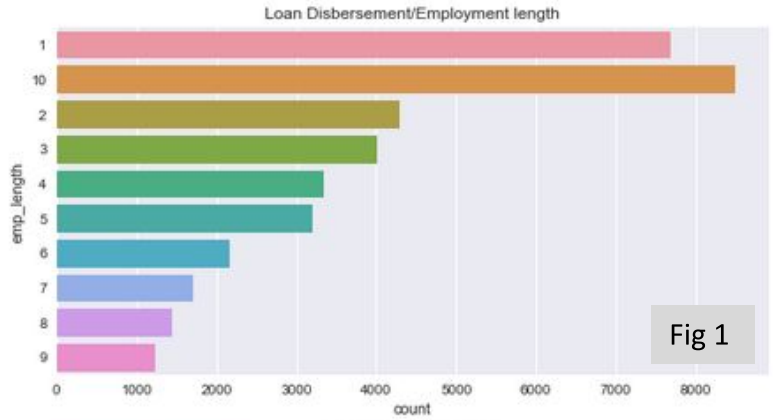
Exploring Data Sets

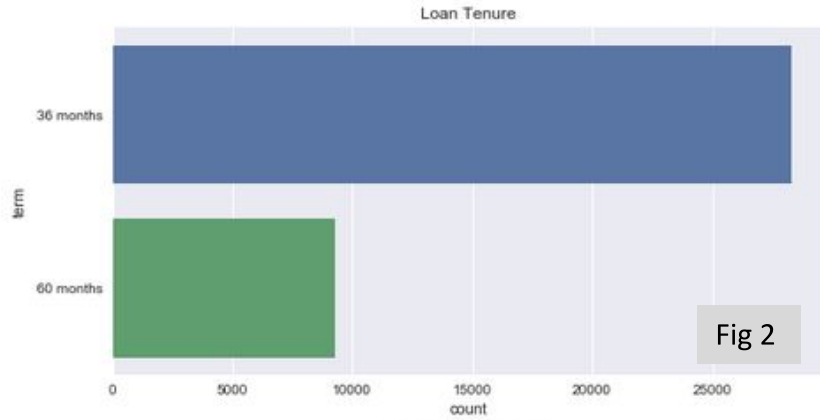
Univariate Exploration

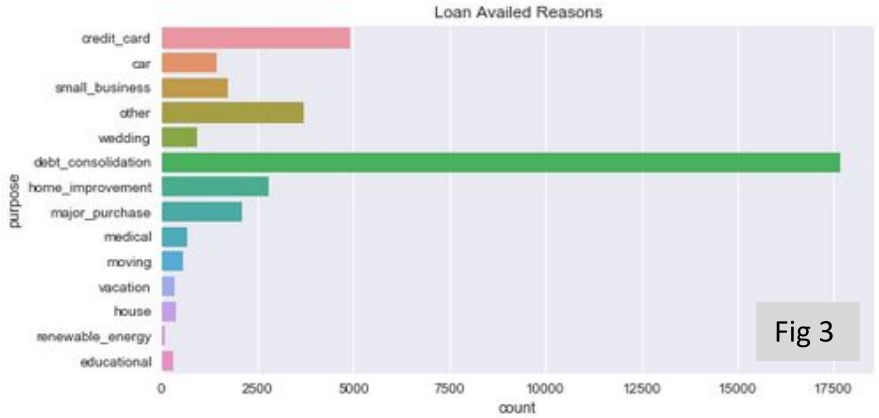
We explore data set available to understand how the loan accounts are distributed around various variables

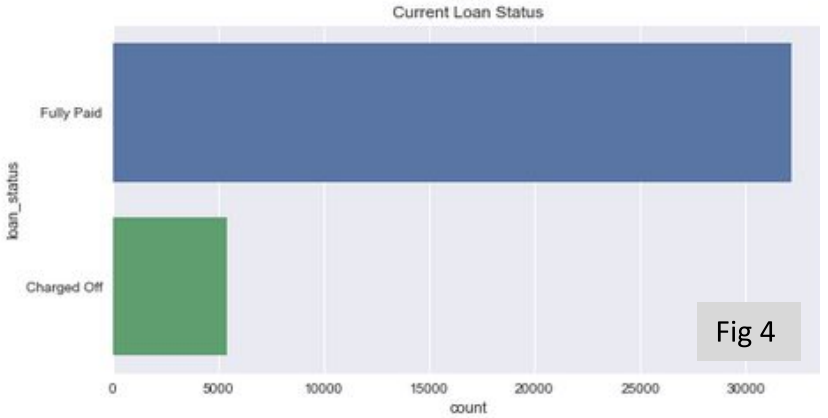
Identification:

1. Fig 1: Top 3 employment length got maximum loans are 10,2 and 1
2. Fig 2: All loans tenure are either 36 month (short term) and 60 months (long term)
3. Fig 3: Loan distribution across various purpose for which load been disbursed
4. Fig 4: Loan application proportion of Fully Paid over Charged Off









Exploring Data Sets

Univariate Exploration

We explore data set available to understand how the loan accounts are distributed around various variables

Identification:

1. Fig 1: Home ownership distribution of loan applicants
2. Fig 2: Verification status of loan applications

Fig 1

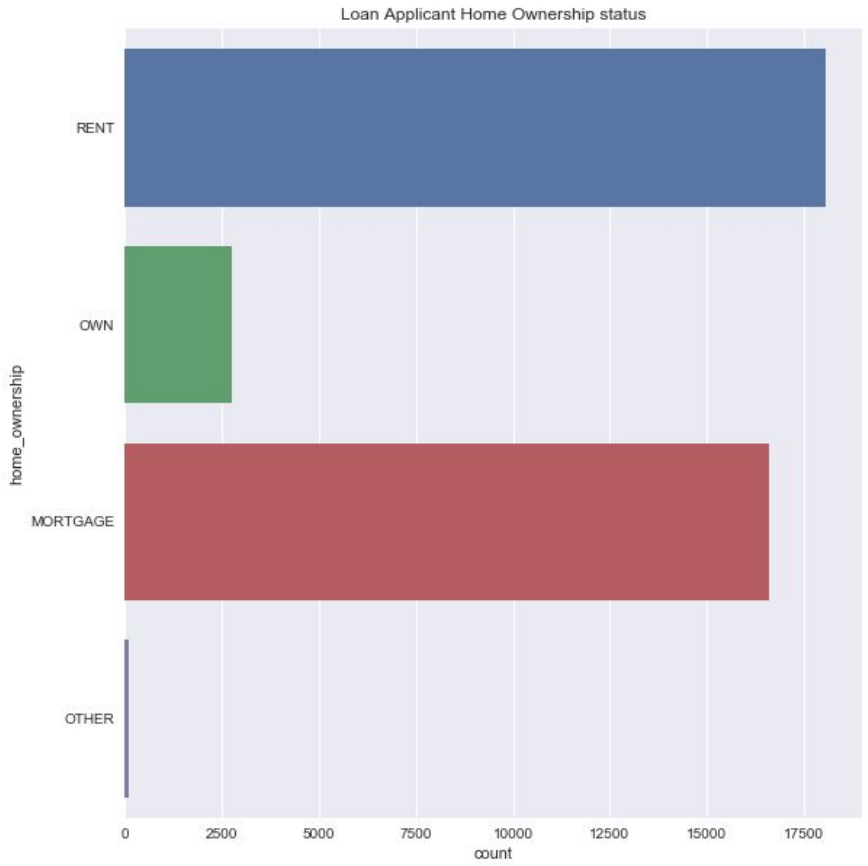
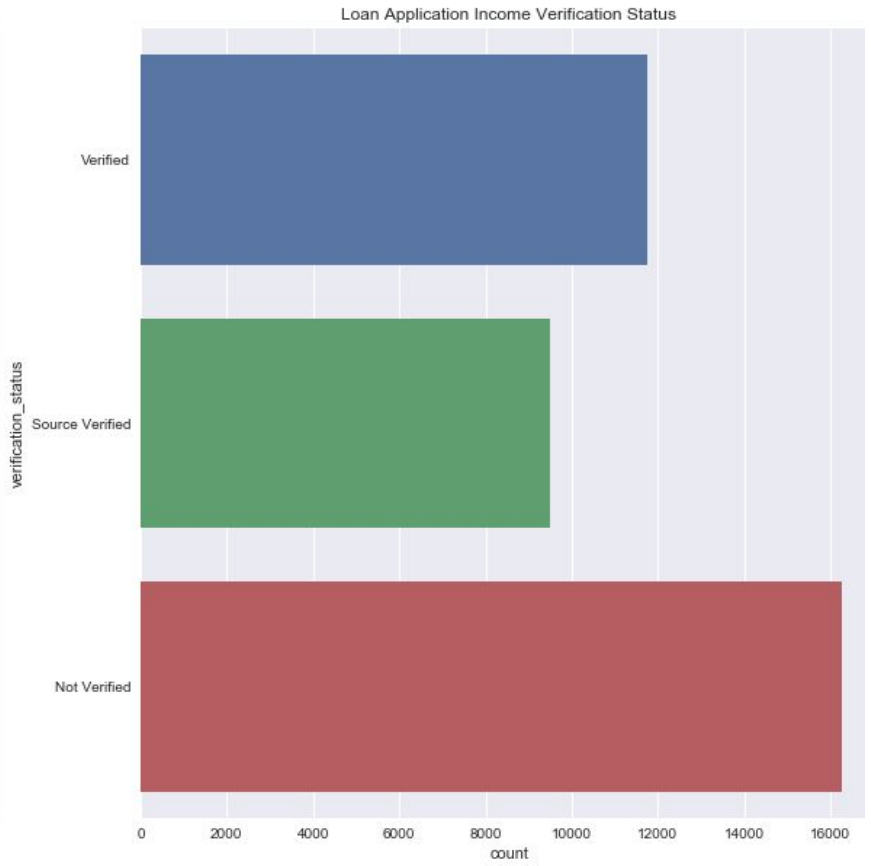


Fig 2



Exploring Data Sets

Univariate Exploration

We explore data set available to understand how the loan accounts are distributed around various variables

Identification:

1. Fig 1: Loan accounts distribution over Grade
2. Fig 2: Delinquency incidents reported for loan accounts
3. Fig 3: Inquires distribution
4. Fig 4: Public recorded bankruptcies

Fig 1

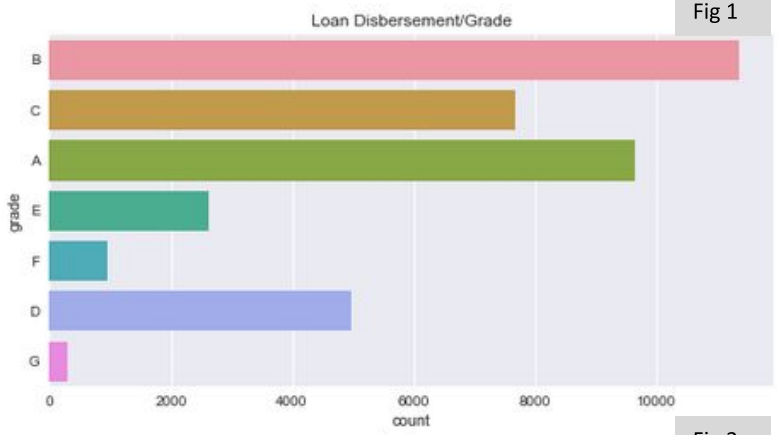


Fig 2

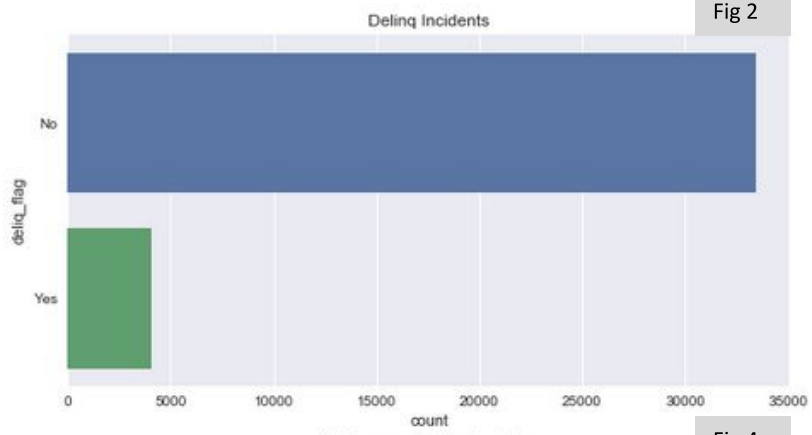


Fig 3

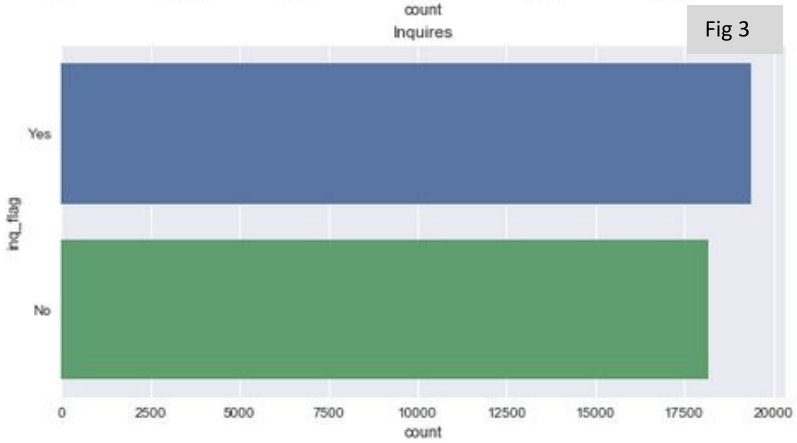
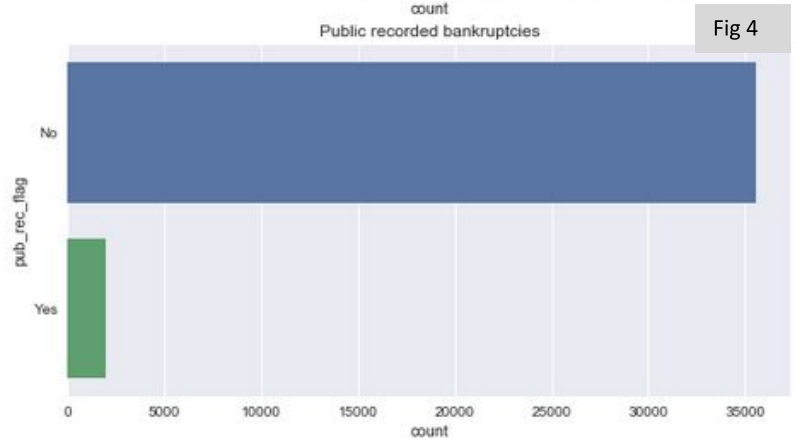


Fig 4



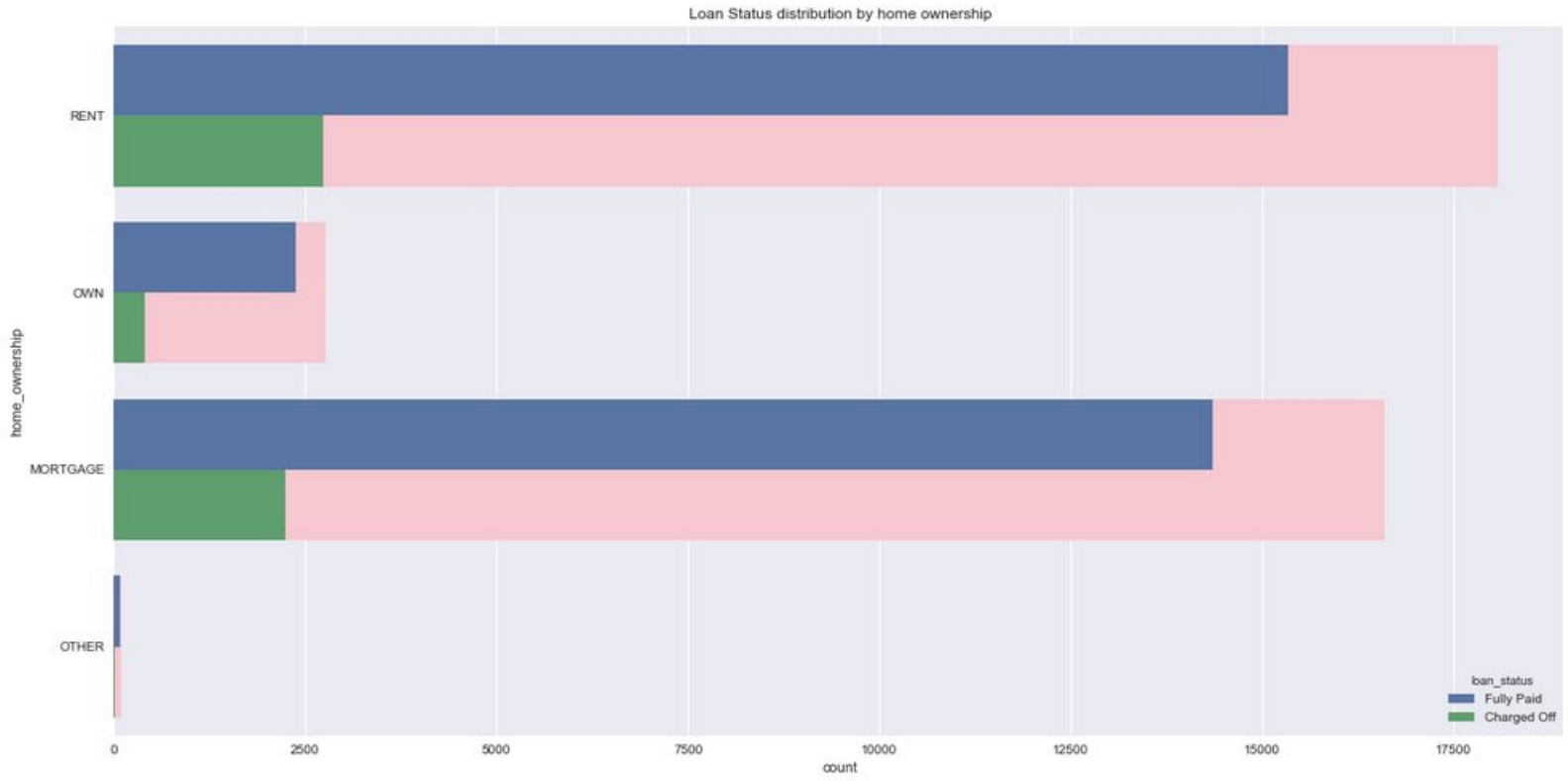
Exploring Data Sets

Bivariate Exploration

We explore data set bivariable to understand how the loan accounts and home ownership are related

Identification:

1. It indicates that maximum ***charged-off*** occurred for loan accounts where home ownership status are either ***rent*** *or* ***mortgage***



Exploring Data Sets

Bivariate Exploration

We explore data set bivariable to understand how the loan accounts and loan purpose are distributed

Identification:

1. Fig 1: It indicates that maximum ***charged-off*** occurred for loan accounts where purpose of loan is ***debt consolidation, small business, other and credit cards***
2. Fig 2: It shows that loan charged-off % is very evenly distributed across employment length.

Fig 1

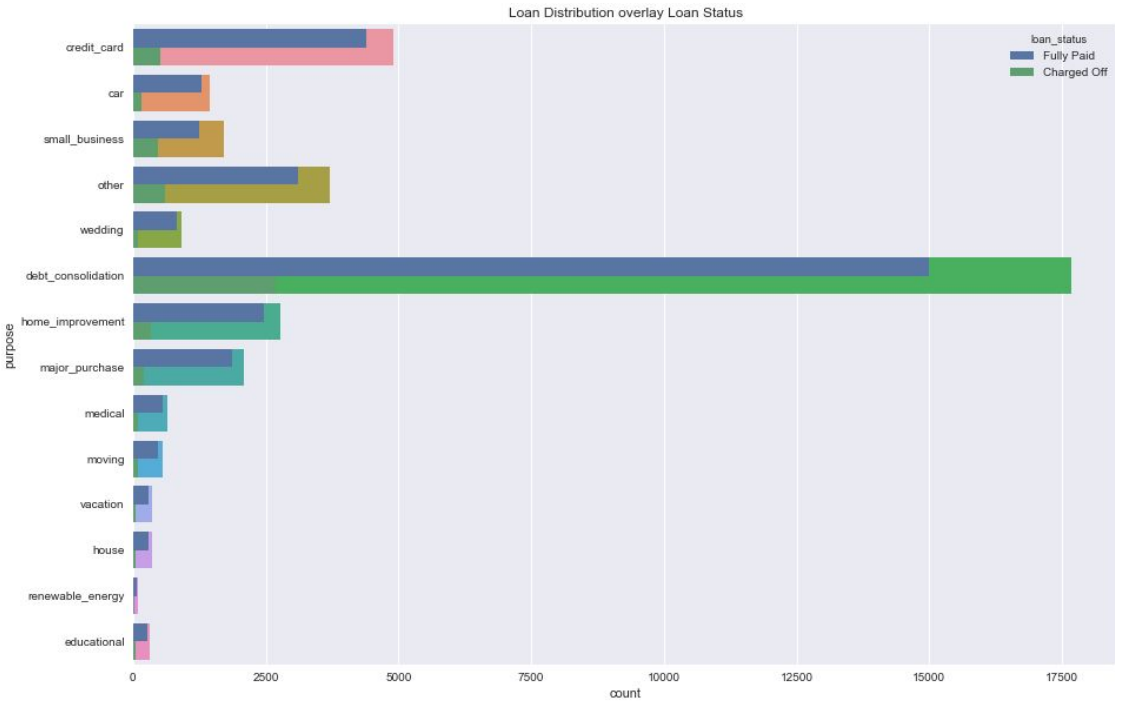
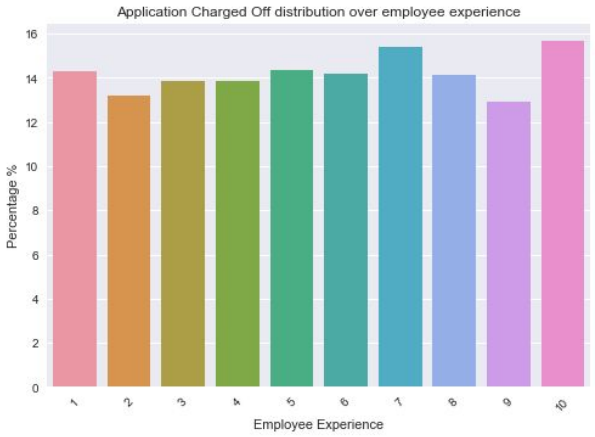


Fig 2



Pre-analysis Summary

After exploring various variables of data set in relationship with charged off or Fully Paid accounts we concluded and identified a list of variables listed in below table.

All variables identified below are attributed by if they hold direct/indirect relationship with charged off, do they have a conjunct variable which forms the basis of bivariate analysis and if they need to be considered at there value or need to be ranged.

Variables	Relationship with charge off	Conjunct variable	Range variable
int_rate	Direct	-	False
open_acc	Indirect	-	True
emp_length	Indirect	annual_inc	False
annual_inc	Direct	emp_length	True
purpose	Direct	-	False
term	Direct	-	False

Charged Off Analysis

As we have observed in data exploration there is not much **strong one to one** relationship between variable and *'Charged-Off %'*. We are going to perform analysis based on bivariable

Bivariate Analysis

We explore data set bivariable how various variables are in groups are contributing to default rated i.e. *'Charged-Off %'*

Identification:

We able to identify some of the variables which we will be looking for our analysis due to there relationship to the *'Charged –Off'* accounts.

	Category_1	Category_2	Level_1	Level_2	Def%	Def%_LC1	Def%_LC2	Pop%	Pop%_LC1	Pop%_LC2	base_Percent
307	term	home_ownership	60 months	RENT	27.9	25.1	15.1	9.9	24.7	48.2	14.4
343	term	inq_flag	60 months	Yes	27.3	25.1	16.6	13.6	24.7	51.6	14.4
4	grade	deliq_flag	E	No	27.2	26.7	14.2	5.8	7.0	89.1	14.4
381	term	purpose	60 months	debt_consolidation	26.2	25.1	15.2	12.9	24.7	47.1	14.4
139	inq_flag	grade	Yes	D	23.7	16.6	21.8	7.7	51.6	13.3	14.4
3	grade	deliq_flag	D	No	22.5	21.8	14.2	10.9	13.3	89.1	14.4
43	inq_flag	deliq_flag	Yes	Yes	18.4	16.6	16.2	5.8	51.6	10.9	14.4
115	inq_flag	grade	Yes	C	18.3	16.6	16.8	13.0	51.6	20.4	14.4
337	purpose	inq_flag	other	Yes	18.2	16.1	16.6	5.1	9.9	51.6	14.4
301	purpose	home_ownership	other	RENT	18.1	16.1	15.1	5.4	9.9	48.2	14.4

Index

- Def%: default rate aggregated
- Def%_LC1: default rate of Level 1 & Category 1
- Def%_LC2: default rate of Level 2 & Category 2
- Pop%: overall population %
- Pop%_LC1: population % of Level 1 & Category 1
- Pop%_LC2: population % of Level 2 & Category 2

Charged Off Analysis

Bivariate Analysis

We grouped open accounts with range index of 2 and explored loan account distribution of charged off and fully paid accounts

Observation:

Fig 1: The distribution of charged off accounts and fully paid accounts on open-accounts range is a uniform normal distribution

Identification:

Fig 2: The account having open accounts in range of 2-10 will be contributing to highest default rate as there credit lines are low.

Fig 1

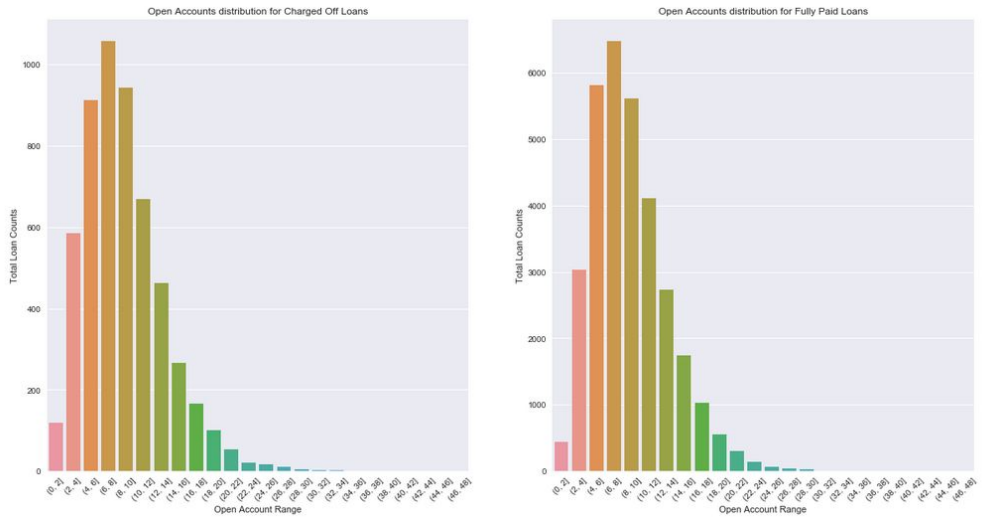
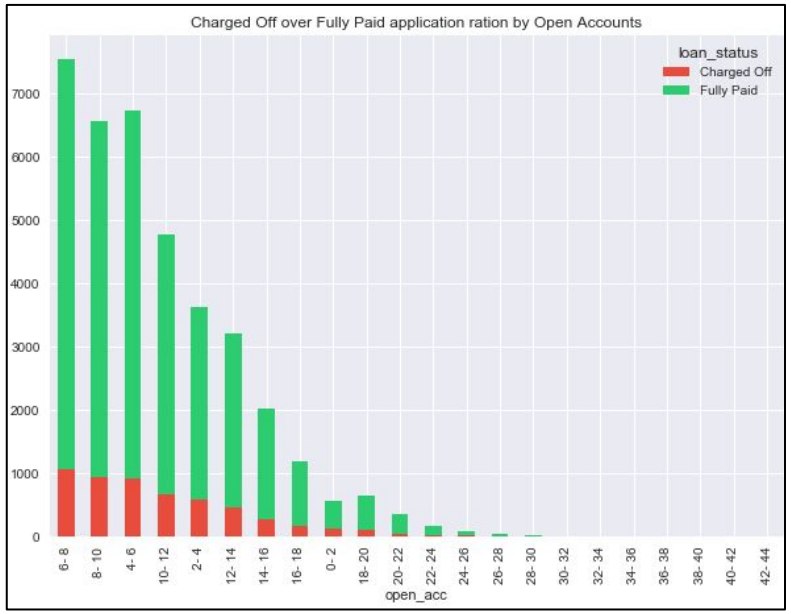


Fig 2



Charged Off Analysis

Bivariate Analysis

As we observed in employment length alone was unable to provide an insight grouping same with annual income provide an interesting insight

Observation:

Fig 1: uniform normal distribution of charged off accounts over employment length

Identification:

Fig 2: Loan account with employee experience of *1,2 and 10 years if in annual income bracket of 4,000-1,24,000 will hold high probability of defaulting loans.*

Fig 1

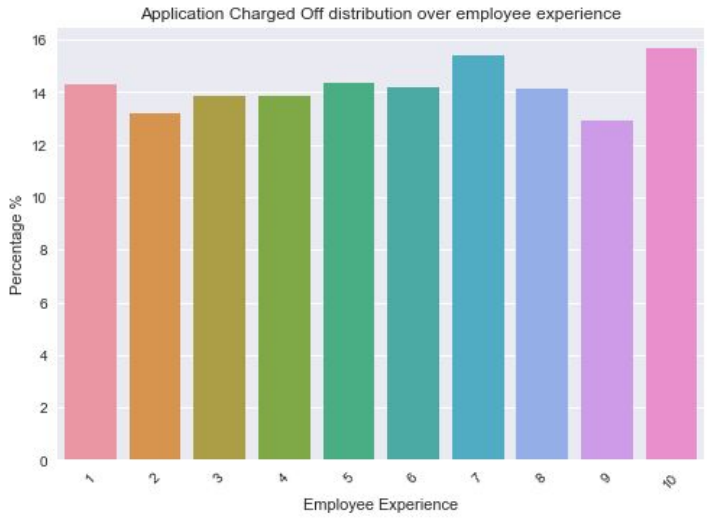
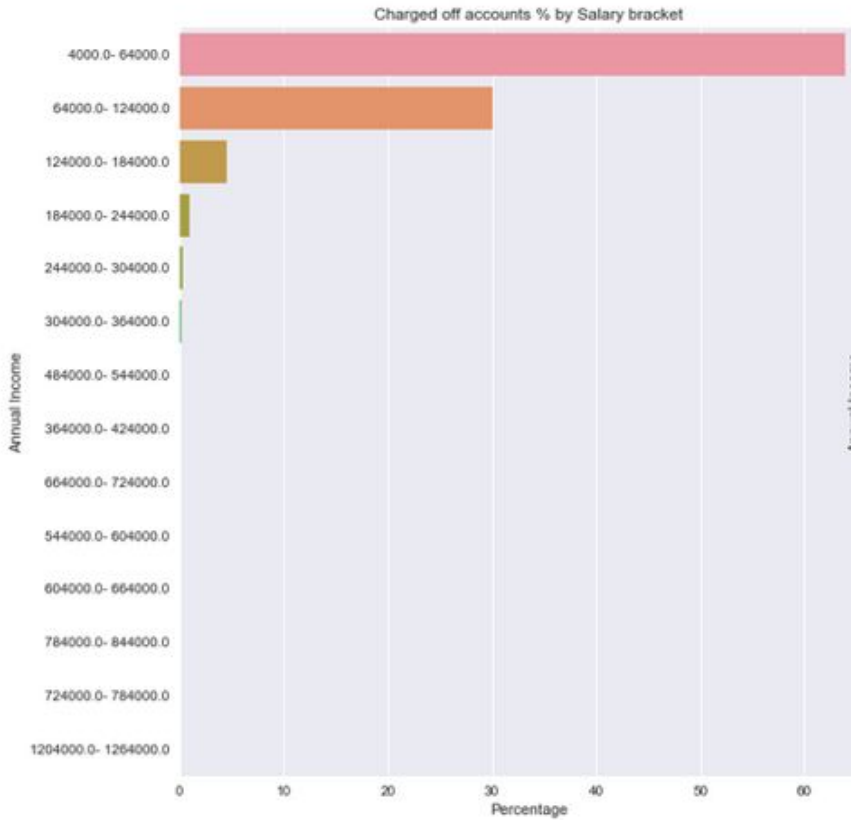


Fig 2



Summary

As per analysis made below is the summary, which can help in filtering loan application which can account to default.

S.No.	Variables	Conjunct variable	Analysis
1	emp_len		A loan give to an applicant bearing employment **experience of 1,2 & 10** year hold high chances of default. The possible reason can be either job instability for an experience employee, a less experience employee may be pursuing higher education
2	Int_rate	term	A high interest rate reflects a relationship with charged off and in conjunction if the loan term is of **long term i.e. 60 months** it is likely that a person may default the loan.
3	term	home_ownership, purpose,int_rate,emp_len	A long term loan i.e. 60 months if taken for **debt_consolidation or for small business** with home_ownership status of **Rent or mortgage** it's very likely to default.
4	open_acc		An open_acc between **range of 0-10** reflect a very instable/low credit line and hold very high probability of defaulting a loan account.
5	annual_inc	emp_len,term,int_rate,open_acc	An annual income range of 4000-124000/annum if in conjunction with employment experience of 1/2/10 offered a loan of long term with high interest rate & low credit lines holds very high probability to default