LENDING CLUB CASE STUDY

Name:
ANIRUDH BABSHET

Lending Club Case Study

A **consumer finance company** which specializes in lending various types of loans to urban customers.
It 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.

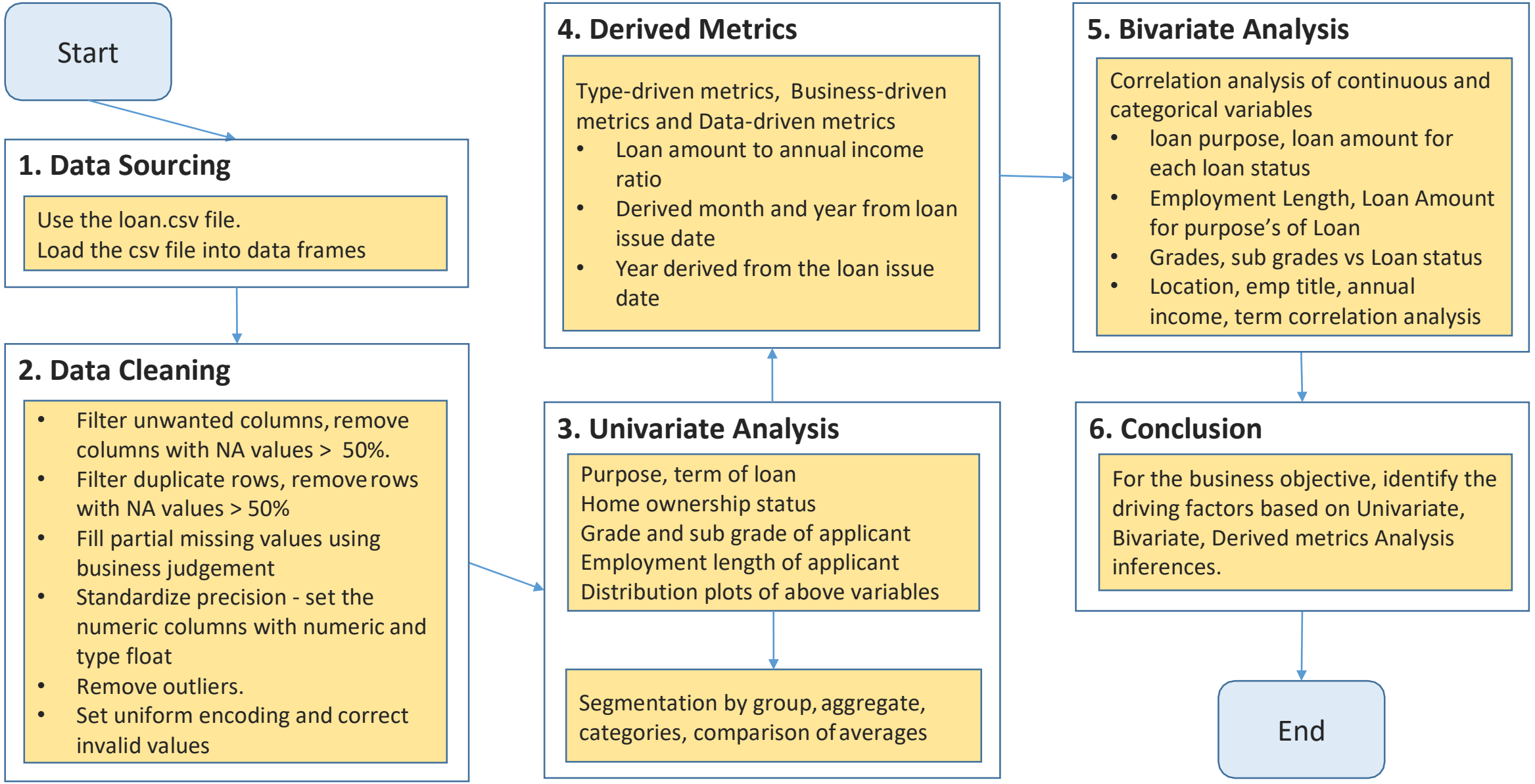
Business objective:

When the company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile. Two **types of risks** are associated with the bank's decision:

1. If the applicant is **likely to repay the loan**, then not approving the loan results in a **loss of business** to the company
2. If the applicant is **not likely to repay the loan,** i.e. he/she is likely to default, then approving the loan may lead to a **financial loss** for the company

The company wants to identify the risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss. The company wants to understand the **driving factors (or driver variables)** behind loan default, i.e. the variables which are strong indicators of default. The company can utilize this knowledge for its portfolio and risk assessment.

Data analysis flow



Data Cleaning

- Fix rows and columns, Filter data
  - There are 111 columns and not every column has different amount of NA values checking for only those columns whose NA values are more than 50%.
  - Remove columns and rows whose NA values more 50%
  - Remove columns whose values same throughout or have just one value for all columns
  - Removing extra columns which are not required for analysis
  - Deleting duplicated rows

- Fix missing values, Fix invalid values
  - *Employment length in years. Possible values are between 0 and 10 where 0 means less than one year and 10 means ten or more years. So we are assuming that any NA will be considered as 'self-employed' which is 0 years of employment. Replace all nan in emp_length to 0*

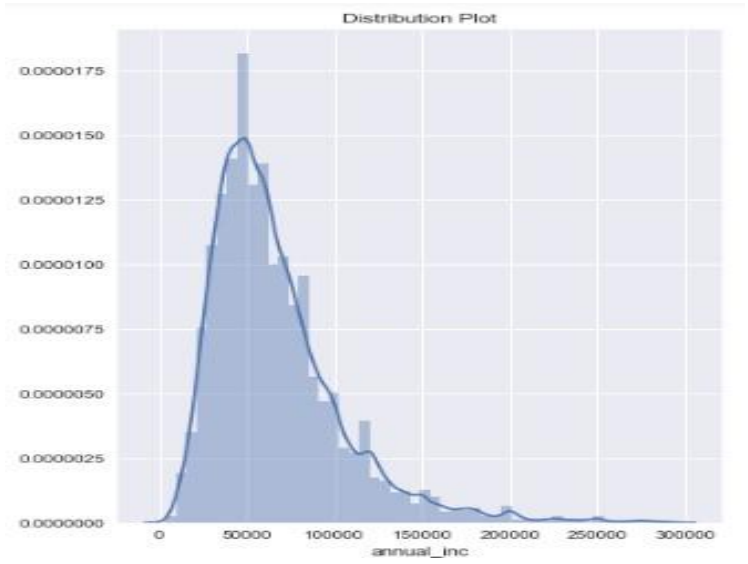
- Standardize values
  - Type case all number columns to numeric

Univariate Analysis

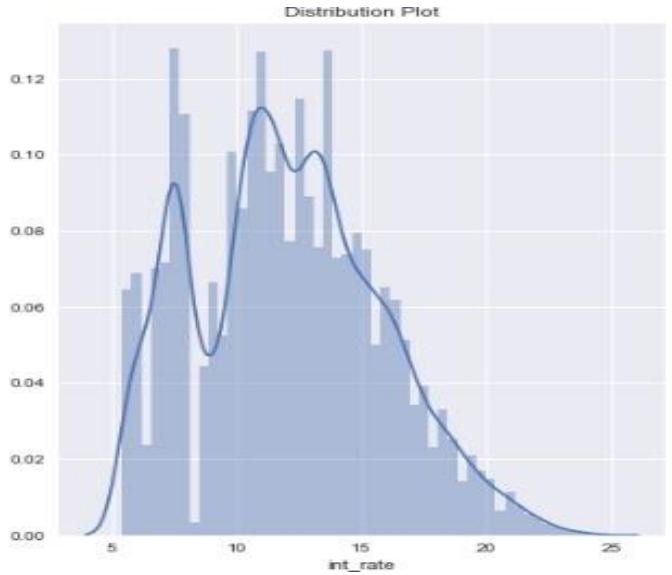
- Ordered and unordered categorical
  - Purpose of loan
  - Home Ownership
  - Loan Term
  - Grade, Sub grade

- Segmented Univariate
  - Create groups for range of loan amounts.
  - Create groups for range of annual income
  - Create buckets for range of interest rates for the loan applicant(s)

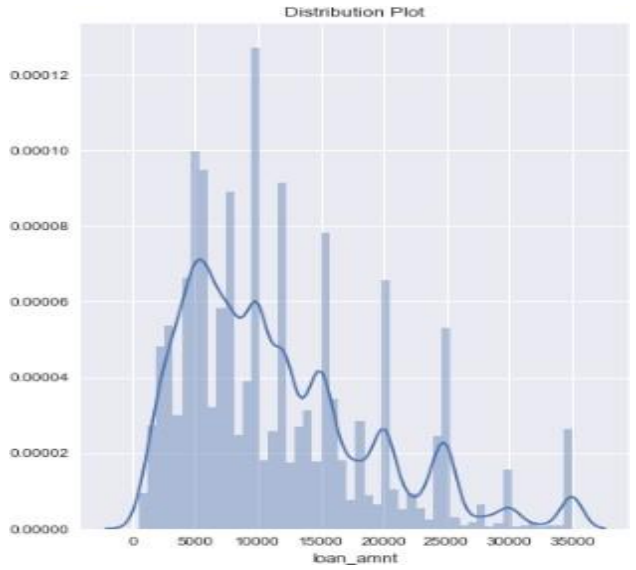
- Distribution Plots



Annual income



Interest Rates



Loan amount

Univariate Analysis – Plots and Inference

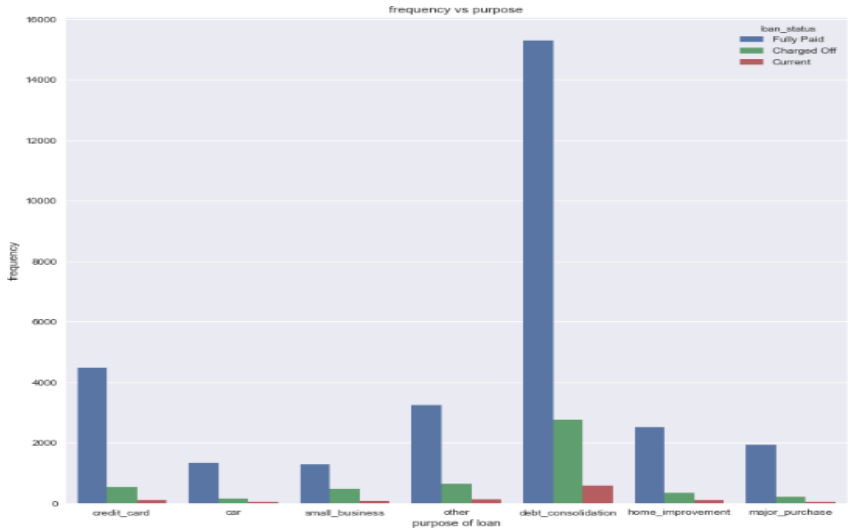


Fig 1

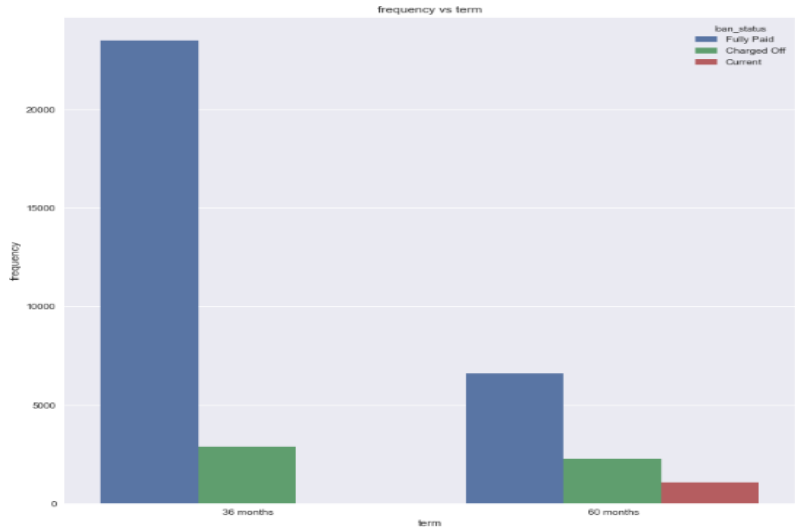


Fig 2

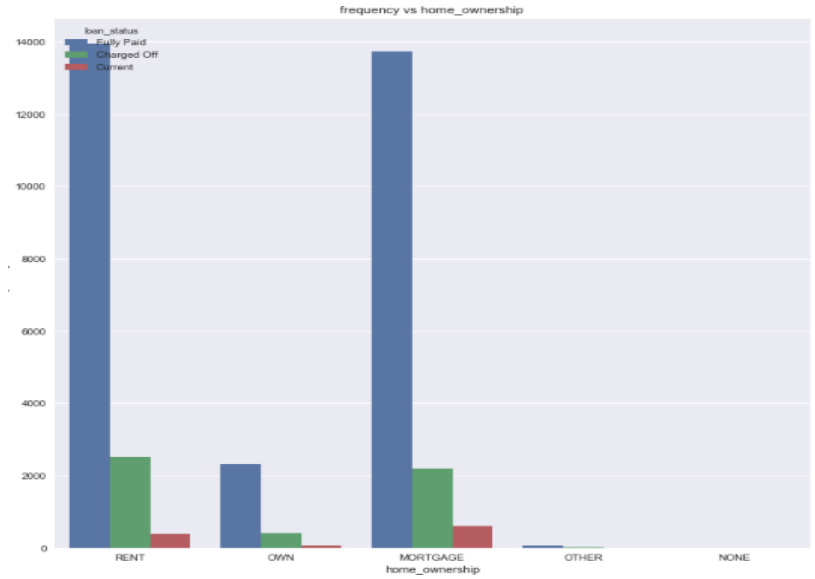


Fig 3

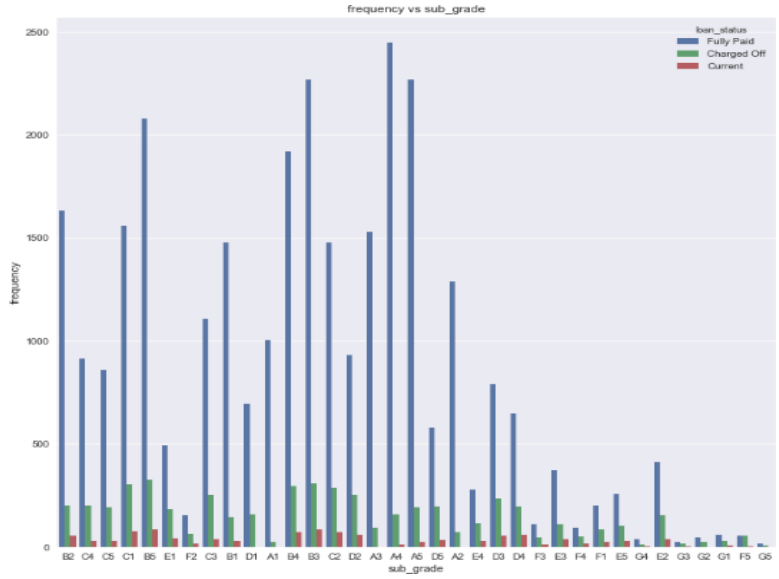


Fig 4

Inferences:

Fig 1 - Purpose of loan

- Debt_consolidation, Credit_card, others and small_business show higher charged off values.

Fig 2 - Loan Term

- The term 60 months has higher probability of charged off.

Fig 3 - Grade, Sub grade

- The grades from A to H has steep increase in charged off values

Fig 4 - Home Ownership

- The Rent and Mortgage factors show high values of charged off.

Bivariate Analysis

- **Bivariate analysis on categorical variables**
  - Purpose of loan vs loan amount for each loan status
  - Bivariate corelation matrix with numeric values
  - Frequency of loan status for purpose and employee length
  - Address_State vs Probability Charge Off
  - Purpose of Loan vs Probability Charge Off
  - Annual Income Range vs Probability Charge Off
  - Interest rate Range vs Probability Charge Off
- **Bivariate analysis on continuous variables**
  - Grade/Subgrade vs Probability Charge Off
  - Employment Length vs Probability Charge Off
  - Term converted to integer and probability vs charged off

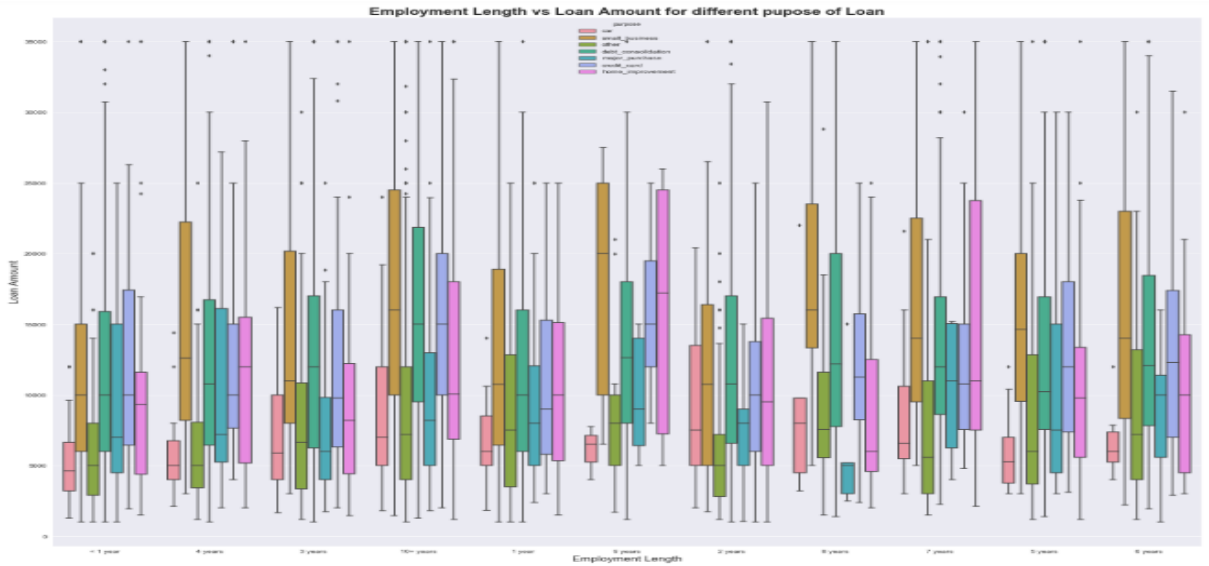


Fig 2

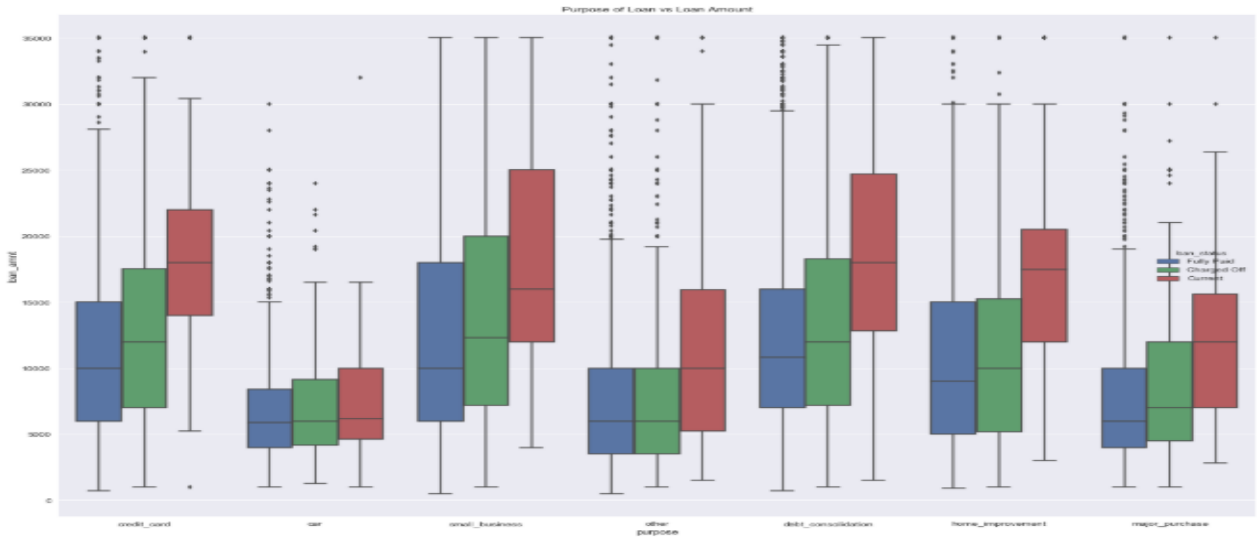


Fig 1

Inferences:

Fig 1 - Purpose of loan vs loan amount for each loan status

- Debt_consolidation, Credit_card, others and small_business show higher charged off values.

Fig 2 - Frequency of loan status for given purpose and employee length

- Loan applicants years of experience would matter the loan of amount sanctioned as well a trade off with loan status

Bivariate Analysis – Plots and Inference

Fig 1

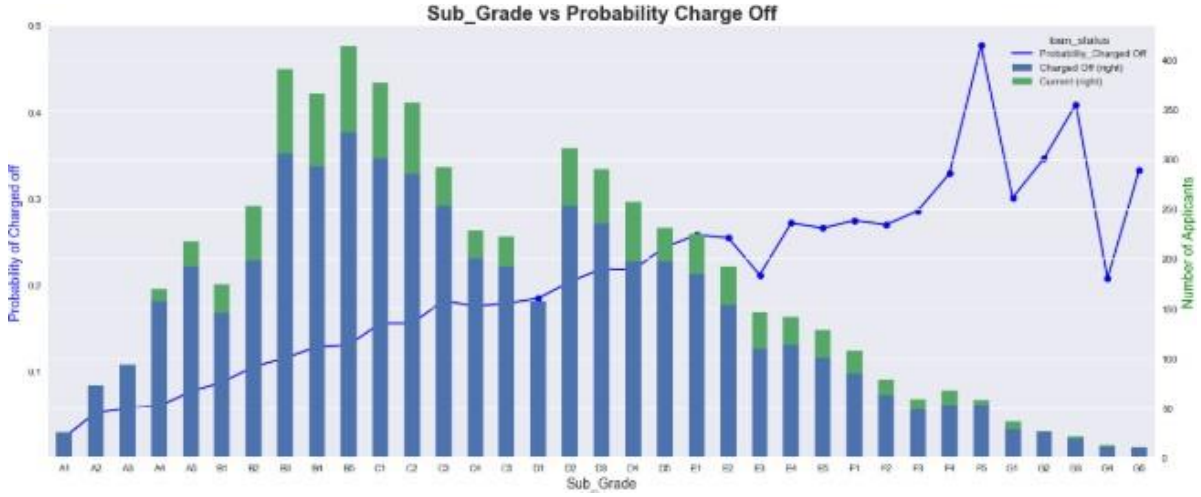


Fig 2

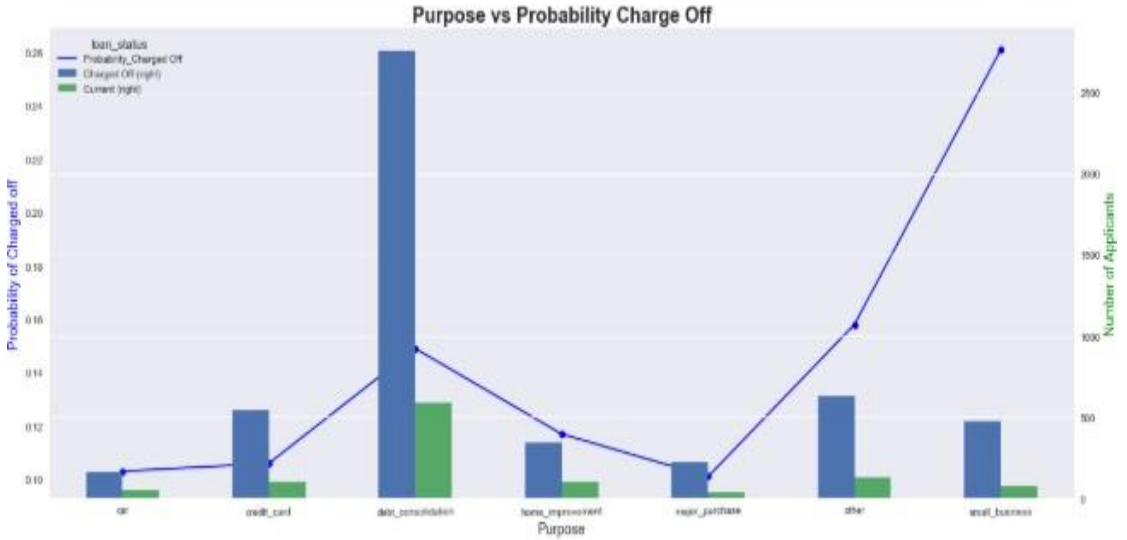


Fig 3

Inferences:

Fig 1 - Address_State vs Probability Charge Off

- There are multiple States/Provinces with high probability of charge off, among which highest being 'NV' above 20%

Fig 2 - Grade/Subgrade vs Probability Charge Off

- From grades A to G from left to right, probability of the applications be charged off are higher.

Fig 3 - Purpose of Loan vs Probability Charge Off

- Applicants who has taken the Loan with purpose of 'small business' has the highest probability of charge off of 26%. So loan lending company needs to take extra care about these 'small business' applicants.

Derived metrics

- Loan amount to annual income ratio
- Derived month and year from loan issue date
- Loan term split to numeric
- year derived from the loan issue date
- **Calculation probability of charged off as p=number of charged off/total loan applicants**

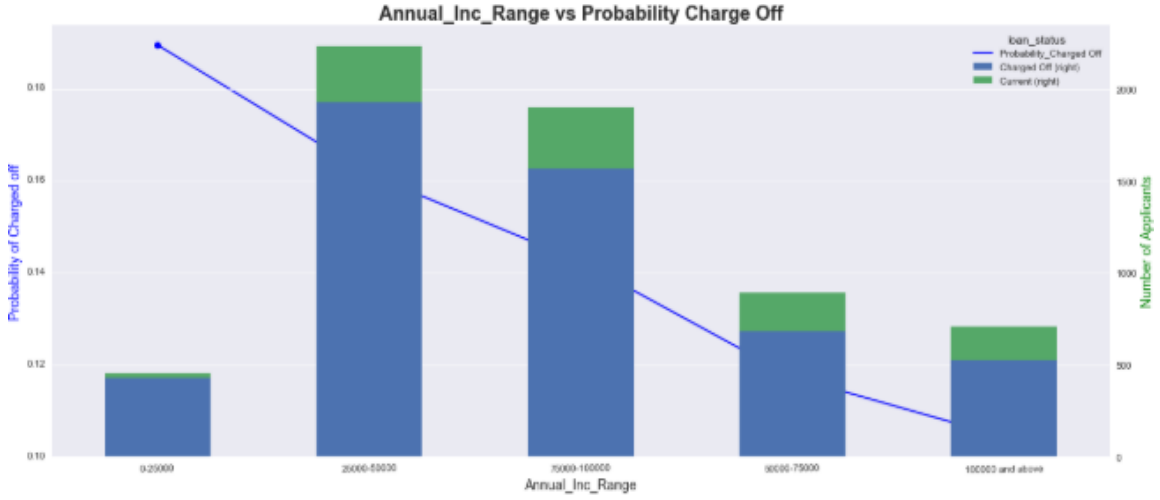


Fig 1

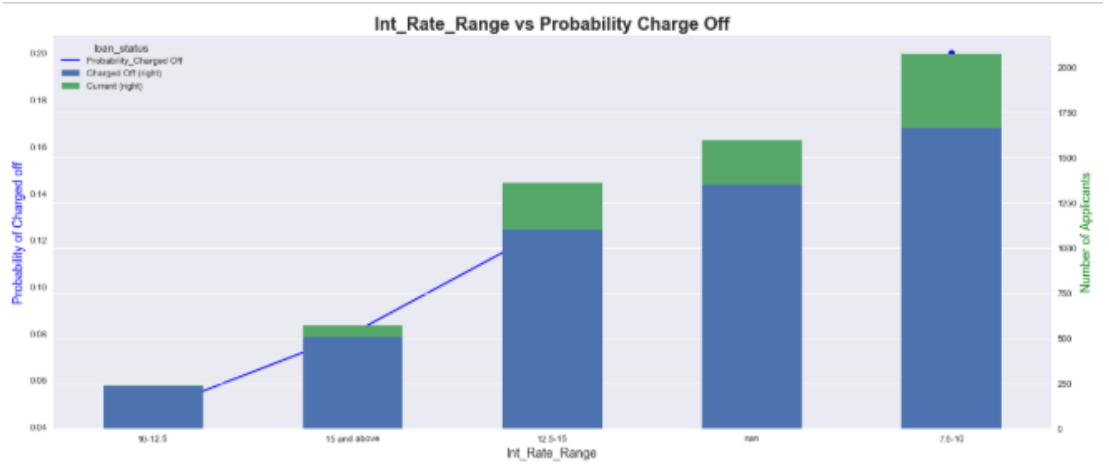


Fig 2

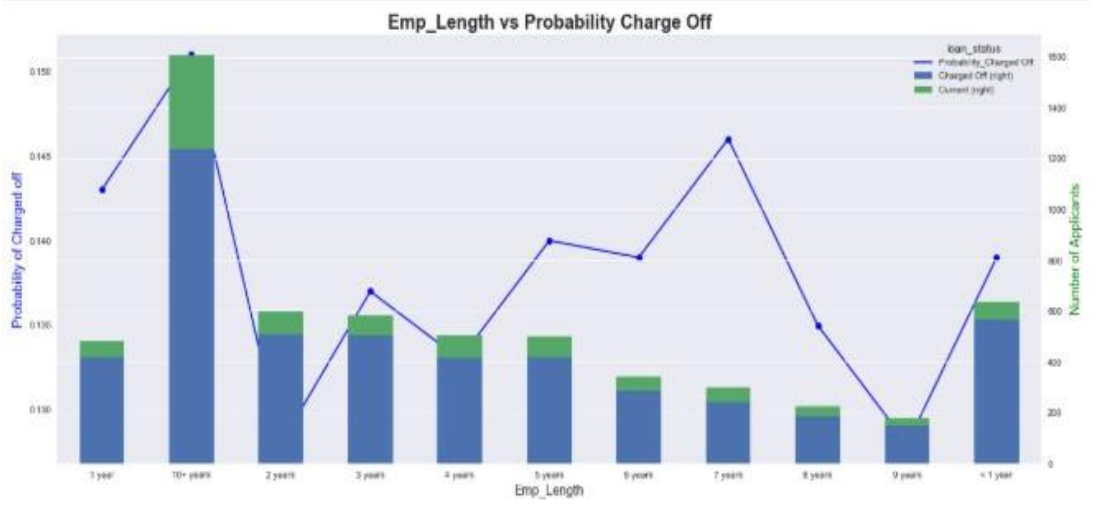


Fig 3

Inferences:

Fig 1 - Loan amount to annual income ratio

- Annual income in the bracket of 0 - 25000 have a probability of charged off nearing 20%

Fig 2 – Interest rate Range vs Probability Charge Off

- Interest rate range of 10 to 12.5% with a probability of charged off being 22.5%. Second peak at 25% with an interest rate above 15%.

Fig 3 - Employment Length vs Probability Charge Off

- Applicants less than 1 year and 0 year considered as self employed are more probable to be charged off

Conclusions

Driving factors/variables:

1. Employment Parameters

- Employment length
  - Applicants years of experience would matter the loan of amount sanctioned as well a trade off with loan status
  - Applicants less than 1 year and 0 year considered as self employed are more probable to be charged off

- Annual Income
  - Annual income in the bracket of 0 - 25000 have a probability of charged off nearing 20%
  - Annual income in the bracket of 50000 - 750000 have a probability of charge off more than 25%

- Employment Title
  - A set of emp titles (At&T, Bank of America, UPS, US Army, WalMart) combined with median annual income of < 45000 show high probability of charged off.

2. Purpose of loan

- Applicants who has taken the Loan with purpose of 'small business' has the highest probability of charged off of 26%.
- Debt consolidation, Others and Small Business are the loan purposes with high risk of Charged off.

3. Loan Parameters:

- Term of the loan
  - 60 months of loan have more probability of charged off
- Interest rate
  - Interest rate range of 10 to 12.5% with a probability of charged off being 22.5%
  - Second peak at 25% with an interest rate above 15%

4. Grade, Sub-grade

- From grades A to G, higher the grade the probability of charged off increases.

5. Location

- 'NV' States has highest probability of charge off with 20%
- SD, FL, MO, NM, AK, CA are other states in order with more probability of Charged off