# Lending Club Case Study

**Case Study Group**

**Gurpreet Singh** (Group Facilitator)  [LinkedIn Profile]

**Karan Prinja** (Group Member )  [LinkedIn Profile]

# Problem Statement

**The Problem** : Lending Club is a consumer finance company that specializes in lending various loans to urban customers. On receiving a loan application , the company needs to take decision whether a loan should be sanctioned or not.

This involves two types of risk :

a) A consumer who is capable to repay the loan but is denied during application process leads to **business loss**.

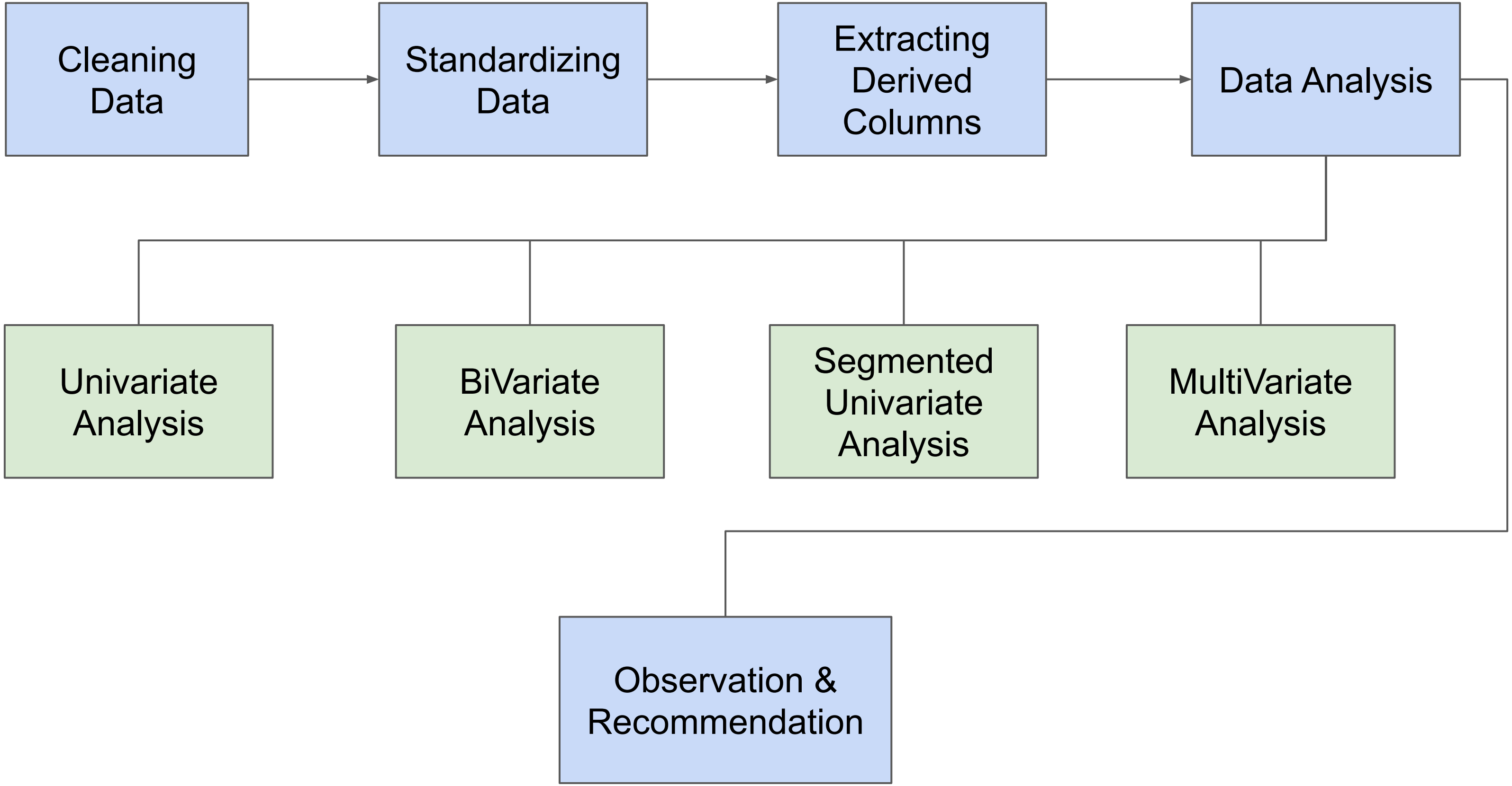
b)

c) A consumer who is not capable to repay the loan but is approved during application process leads to **financial loss**.

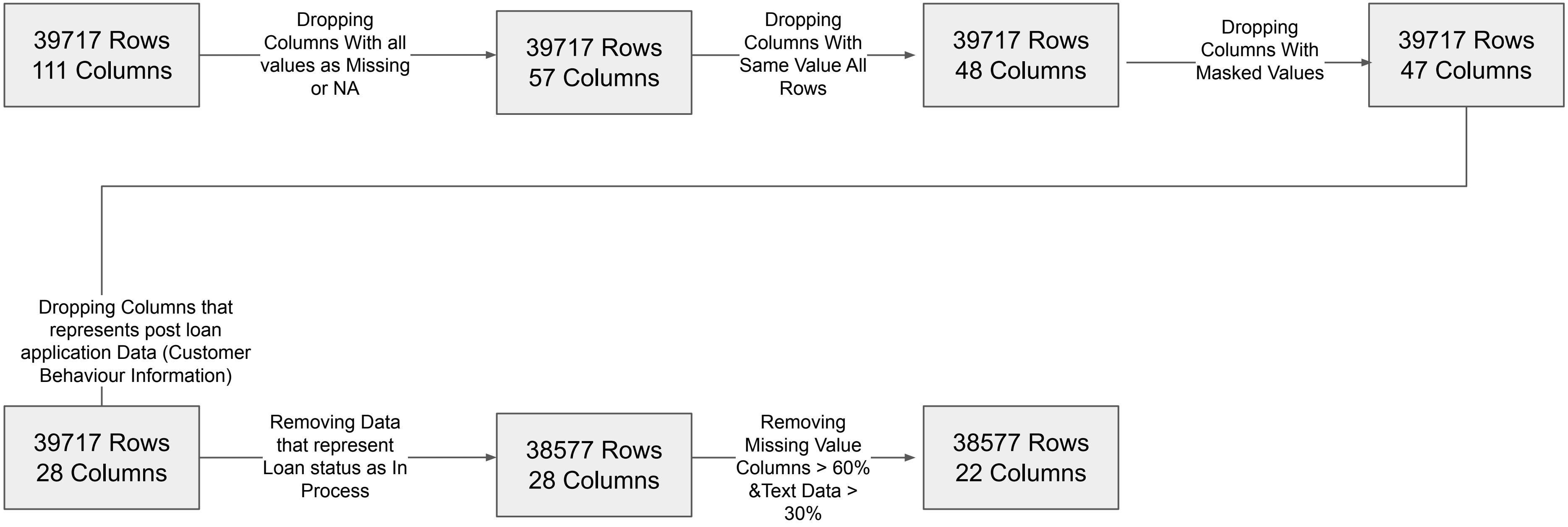
**The Goal** : To Identify pattern based on the historical data provided, that suggest whether an applicant will be able to repay the loan or not. Further , the observation should also help Lending club to develop effective strategies to reduce the risk of default and make informed lending decisions.

**The Approach** : Using EDA Techniques to grasp understanding of the data and find out how customer attributes and loan attributes impact the likelihood of loan default

# Exploratory Data Analysis Flow



# The Data Cleaning Flow



upGrad

# Imputing Missing Values

| Columns | Missing Values | Approach to Impute | Imputed With |
|---|---|---|---|
| Employee Job Title | 6.19% | It has 28k+ records with unique Values and are user specific title, instead of removing this column, we prefer to keep it for future records and to ensure no data changes is done on this column , we impute it with a specific value only | Not Specified |
| Employee Period of Service | 2.68% | Since it has few records mentioned and has a very less unique values so we replaced this with mode value | 10+ years |
| Public record bankruptcies | 1.81% | This represents event of bankruptcies and there could be borrower that do not have this event on their record , so we assumed as no such event has taken place | 0 |
| Title | 0.03% | Title also has 19k+ unique records so we apply the same approach that we applied for Employee Job TItle | Not Specified |

# Data Standardization

| Columns | Action taken |
| --- | --- |
| Employee Period of Service | We replaced less than 1 year as 0 & greater than +10 as 10 and Converting rest of the value into number format |
| Interest Rate | We removed % sign and converted text to number format |
| Term | We removed months and converted text to number format |

# Derived Columns

| Columns | Purpose |
| --- | --- |
| **Issue Month and Issue Year** | To analyze data by month and year, we converted loan issue date to 2 different columns by the name of month and year. |
| **Annual Income Range Bins** | We created bracket of income ranges to convert continuous values to categorical column for doing analysis |
| **Interest Range Bins** | For this column also , we created bins to better understand this column in combination of other columns ( bivariate analysis) |
| **Loan Amount Range** | Loan Amount Range is also splitted in to bins for understanding and categorize what type of loan amount range belongs to what type of behaviour over them. |
| **DTI Range** | This is ratio of debt to income and the same is converted to bins for doing bivariate and multivariate analysis. |

upGrad

# Univariate Analysis - Observations

58% of the loan count disbursed fall in the category of Grade B and Grade A

75% of the Loans are given in smaller amount ranging below $14k

18% of the loan disbursements are for California

*77% of the loans have term of 36 months*

Most of the installments are between 150-300 months

Most of the loans are given to mid range income group. The highest income range taking maximum loan belongs to 40K-60K $

## Univariate Analysis

*24% of the Loans are given to people with 10 or more years of experience*

*Most of the installments are between 150-300*

47% of loan have purpose for loan as Debt Consolidation

50% of the loan are given to applicants who are living on the rent

Interest Rates between 10.5 and 12 are the highest

45% of the Loans are not verified

# Bivariate Analysis [ With Charged Off ] Observations

Loans issued in 2007 got the most charged off which got improved in the next 3 years gradually but a rise was again visible in 2011 , if we look into the economical situation of the country , we will find that The performance of the US economy from 2007 to 2011 was characterized by a severe recession that began in late 2007 and lasted until mid-2009. So somehow , **the country economy plays a role in the charged off status**. When the economy is not performing well , the chances of charged off increases.

Loan Defaulted for Other category is high but we can't find what does other denote , by business understanding "other" category belongs to either co owner property or named property keeping in consideration the rest of categories listed. **No concrete conclusion** can be made here as to which type of home ownership has higher proportion of defaulters

Loan Defaulted Percent is **higher for State NE , but it also needs to be mentioned here that no. of loan application processed for this state are too low** , the state that has higher number of loans processed is **CA which has default ratio of 16.12 %** and **DC stands with just 5%** default rate with total cases processed as 187

## Bivariate Analysis

DTI is the ratio to calculate debt vs monthly income of the borrower and it is clearly visible that **as this ratio increases , the risk of getting charged off also increases.**

This was an unexpected conclusion came out which shows that verified borrowers got higher charged off but ideally it should not be the case since a verification process ensures that borrowers repayment capability is fine. The data **suggest to check the verification process Standard Operating Procedures ( SOP ) and make necessary changes** required to ensure that a verfiied status ensures low risk of getting charged off

Loans Application with **higher term have higher chances of getting defaulted**. It could be because a longer period of time does have higher chances of loan repayment capability as compare to shorter periods

26% of the Charged off Cases have purpose as **"small business"** making this as the **highest risk candidate**

Charged off rate increase as the **employee term** of service increases

Higher the **income range** , less the chances of getting charged off

As the **grade increases , the risk of getting default also increases**. Loans that are marked as Grade A has the lowest risk of getting defaulted. Grade G are the riskiest loan

As the **interest rate increases** , the chances of getting charged off also increase.

**Higher the loan amount , more the chances of getting charged off** but this is applicable only up to a certain loan range. As the loan amount reaches the upper max level, it is evident that chances of getting charged off also reduces.

# Bivariate Analysis [ With Different Features] Observations



Grade A has charged off rate for just 6% and charged off rate has increased as the grade is decreasing but loans are given mostly to other Grades at a higher rate of interest

Interest rate increases as we have loans for lower grades and the charge off rate for such loans is also very high.

DTI Ratio is not given weightage while approving loan as we can see that it is same for almost all the grades excluding grade A. Additionally it could be also interpret that DTI ratio is similar to all the grades which is not possible. There should be some measure to check while a customer is mentioning annual income to ensure it is filled correctly.

Around 43% of the total loan amount sanctioned is done for the purpose of debt_consolidation that has an average charge off % as 15% where as small business tops the list by 27% and wedding is the purpose with the lowest risk of 10.18%

## Bivariate Analysis

With lower loan amount also there are states that are getting defaulted which clearly identifies that state has less potential for repaying the loan as in the above case NE.Further, we can see that for states that have low loan amount and even low interest rate , the charged off rate is still high. Specific marker should be there to grade State also before sanctioning loan

As the income increases, the loan requirement gets decreased . This does not add any conclusion to the target variable and can be just an observation while finding relations

# Segmented Univariate Analysis ( Observations)

Interest rate increases as we have loans for lower grades and the charge off rate for such loans is also very high.

Grade A is the least riskiest grade with charge off rate as compared to other Grades. As the Grade decreases in order from (A..Z ) , the charged off ratio is also getting high

DTI Ratio is not given weightage while approving loan as we can see that it is same for almost all the grades excluding grade A. Additionally it could be also interpret that DTI ratio is similar to all the grades which is not possible. There should be some measure to check while a customer is mentioning annual income to ensure it is filled correctly.

## Segmented Univariate Analysis

With lower loan amount also there are states that are getting defaulted which clearly identifies that state has less potential for repaying the loan as in the above case NE.Further, we can see that for states that have low loan amount and even low interest rate , the charged off rate is still high. Specific marker should be there to grade State also before sanctioning loan

As the income increases, the loan requirement gets decreased . This does not add any conclusion to the target variable and can be just an observation while finding relations

upGrad

# Driver Features

iiit-b

## Grade :

**Observation :** Grade A has charged off rate for just 6% and charged off rate has increased as the grade is decreasing but loans are given mostly to other Grades at a higher rate of interest.

**Recommendation :** Grade plays an important role in predicting customer behaviour.

## Purpose :

**Observation** : Around 43% of the total loan amount sanctioned is done for the purpose of debt_consolidation that has an average charge off as 15% where as small business tops the list by 27% and wedding is the purpose with the lowest risk of 10.18%

**Recommendation** : Purpose of the loan if used tactically can reduce risk and ensure repayment of loan

## State Economic Condition:

**Observation** : Loans issued in 2007 got the most charged off which got improved in the next 3 years gradually but a rise was again visible in 2011 , if we look into the economical situation of the country , we will find that The performance of the US economy from 2007 to 2011 was characterized by a severe recession that began in late 2007 and lasted until mid-2009. So somehow , the country economy plays a role in the charged off status. When the economy is not performing well , the chances of charged off increases.

**Recommendation** : If a certain state economic condition is not performing well , we need to control the sanctions and ensures that rest of the indicator are favouring approvals.

# Driver Features

## Income Range :

**Observation :** Higher the income range , less the chances of getting charged off

**Recommendation :** Income Range is yet another important variable that ensures repayment.

## Tenure:

**Observation :** 25% Charge off rate is for loan that have higher duration. It also needs to be noticed that predicting stability of the borrower for a longer tenure is tough as compare smaller tenure.

**Recommendation :** Loan Tenure needs to be played around where other variables are not giving strong indication that the person is less riskier.

## Home Ownership :

**Observation :** Rented Ownership has a charge off rate for 15% but due to information provided as “other” , it's tough to identify it as strong driver.

**Recommendation :** It could be an effective driver but not a strong one.

## Lower Loan Amount:

**Observation :** Data Suggest that loan amount upto 14K have lower risk as compared to upper but once the loan amount crosses a certain limit ( in this case ) 28K, the charge off rate decreases.

**Recommendation :** It could be used strategically as and when required but cannot be a strong driver.

upGrad

# Recommendations

iiit-b

## Verification process :

**Observation :** Verified Loans have a charged off rate for around 17% as compare to 12% for not verified. This should not be the case since the verified applicants chances of getting default should be less.

**Recommendation:** Verification SOP needs to be checked and implemented to ensure low risk for Verified Applicants.

## DTI ratio :

**Observation:** DTI ratio is basically ratio of debt vs income. Grade , on the other hand identifies income as one of the important criteria. But on comparing these 2 fields, no major difference was seen **across** different grades.

**Recommendation**: Measures should be taken to ensure that information mentioned during application have some government authorized documentation to proves the income levels.

# Thank you