



# Lending Club Case Study Analysis & Recommendation

Shiv Malhotra



## Problem Statement:

To understand the **driving factors (or at least 5 driver variables )** behind loan default, i.e. the variables which are strong indicators of default.



## Analysis & Approach

1. Perform EDA by importing the data & creating the relevant data frame for numerical & data analysis by importing libraries.
2. Understanding the variables by taking the reference from Data dictionary to understand the relevance based on problem statement.
3. Transforming the data by dropping all the drivers with large amount(>80%-85%) null values & which are not relevant based on definition & problem statement. Conversion of selected drivers into correct data types.
4. Validating the data by understanding the total unique values for each driver & summation of null values if any in each driver. Also, replacing the null values with the measures of central tendency based on the nature of driver i.e. categorical or continuous.
5. Interpretation of data or selected drivers with the help of visualization libraries to get meaningful insights from each driver & provide relevant recommendations on the basis of such insights.



# Recommendation & Insights

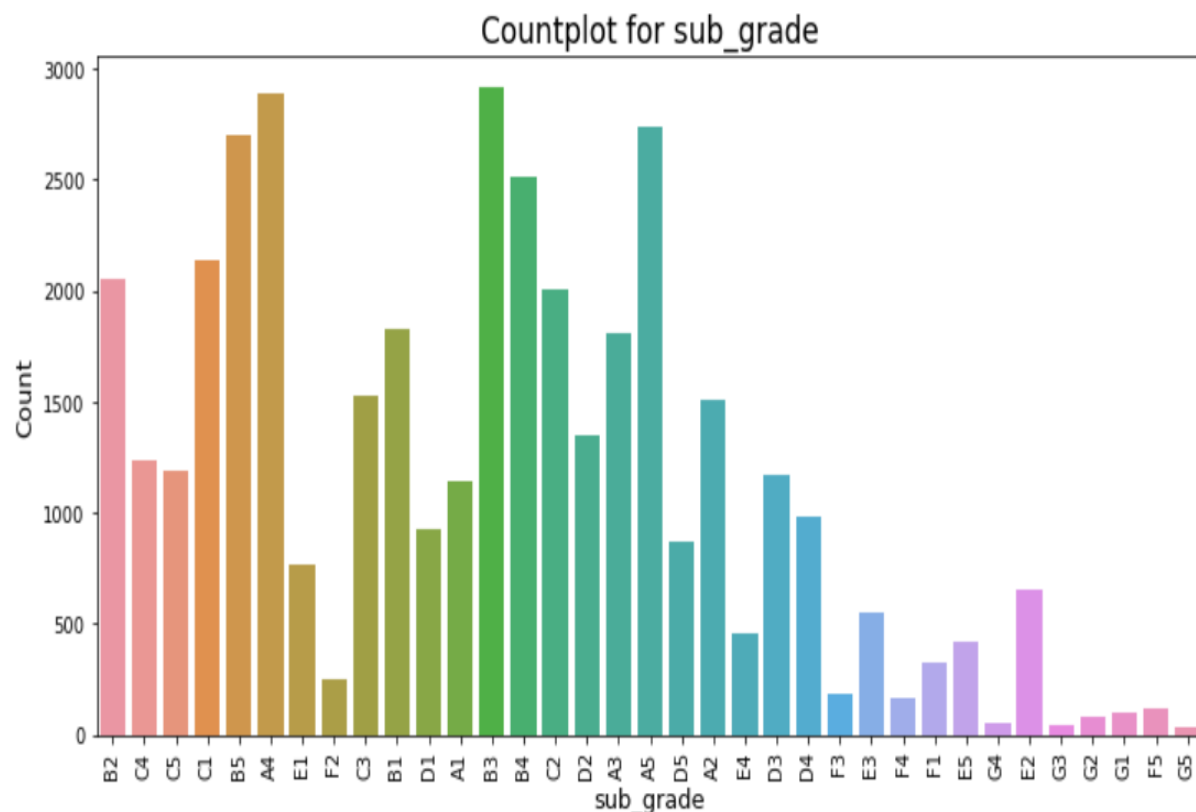
Based on my analysis I have shortlisted 9 drivers which are strong indicators of loan default:-

1. Loan Amount
2. Interest Rate
3. Annual Income
4. DTI
5. Sub Grade
6. Home Ownership
7. Purpose
8. Verification Status
9. Loan Status



# Univariate Analysis of Categorical Variables

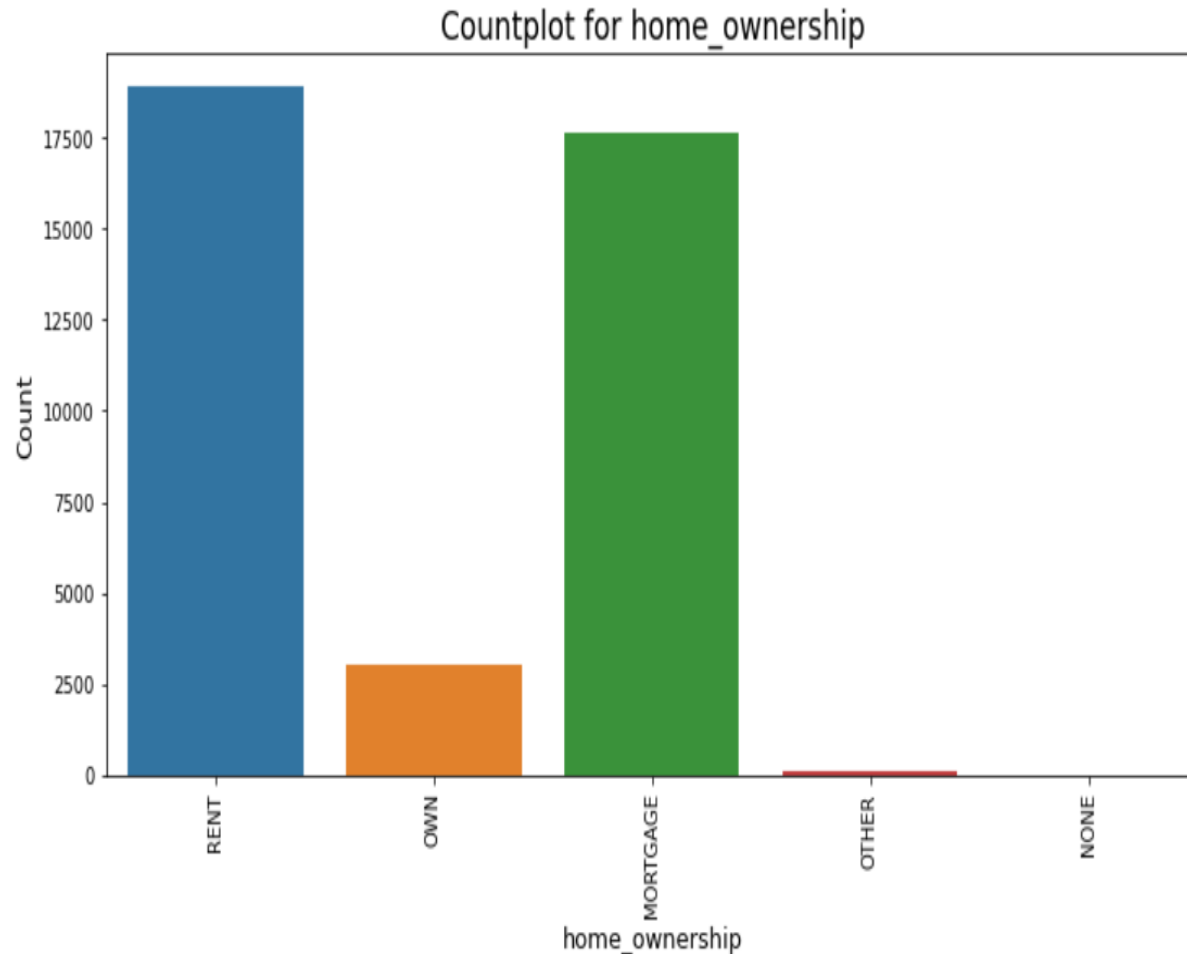
## 1. Sub Grade



Sub Grade has been shortlisted one of important categorical driver over grade as it gives the info of both loan grades & its sub grades and LC assigned loan sub grade. It shows majority of the applicants are having a loan grade & sub grade of B5, B4, B3, B4, A5 which gives a valuable info on the quality of applicants.



## 2. House Ownership

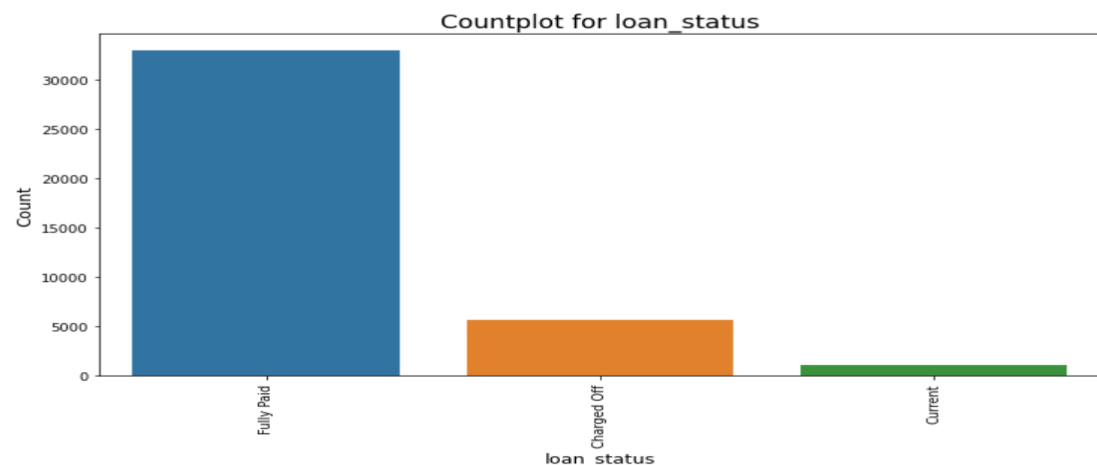
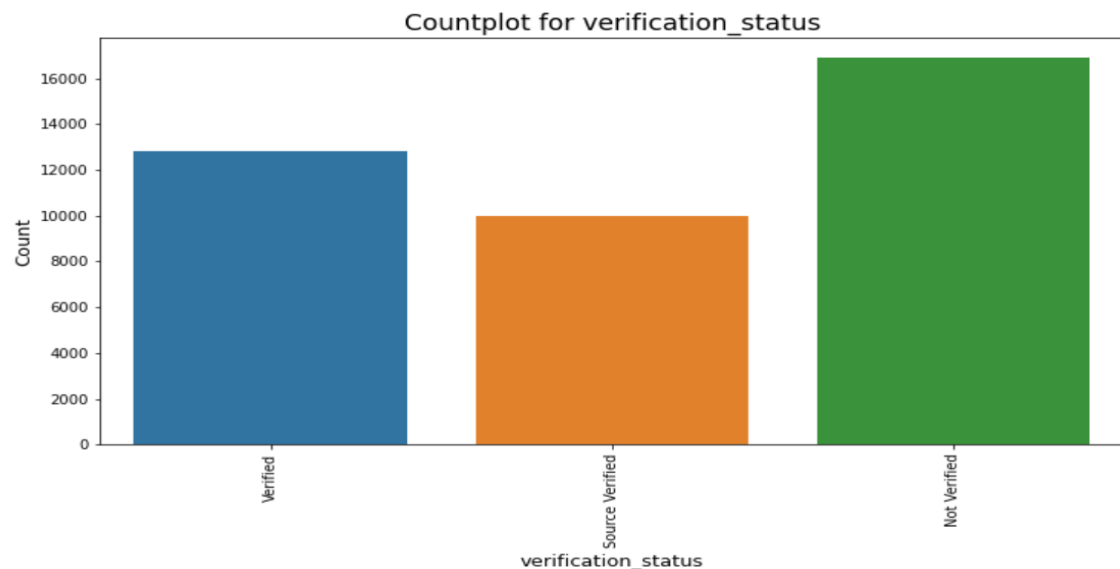


House Ownership has been considered as an important driver as it shows if an applicant has a financial asset that can be used in case of default to recover the loan & if a person is in mortgage, then there is already a financial obligation on the applicant & if applicant is on rent then it shows that client doesn't have a strong income to own a house.

It shows that majority of the applicants are either in rented house or in mortgage house which shows either they don't have strong income to own a house or already having a huge financial obligation.



### 3. Loan Status & Verification Status

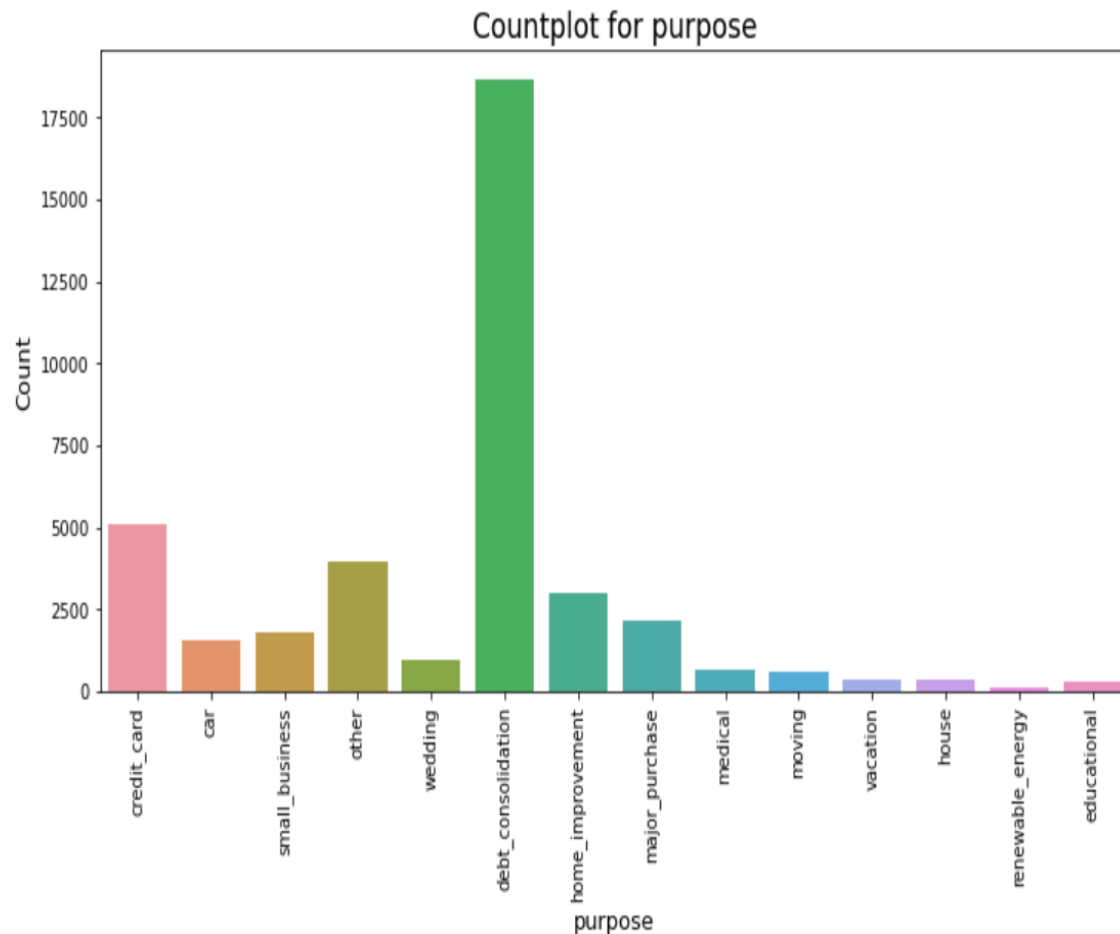


**Verification Status:** It is an important driver as it gives an insight about the applicant authenticity & if income was verified by LC, not verified, or if the income source was verified. It shows majority of the applicants are not verified which is a red flag or signal to keep an eye on other drivers for a better understanding of his authenticity of income or source of income.

**Loan Status:** It is a very important driver to consider applicant scenario as if the applicant has fully paid there is chance that he might not default on the loan & if the applicant has been charged off it means there is chance, he has financial irregularities or was not discipline in paying off his debts. This can be used with other drivers to get a complete picture of the applicant financial stability. It shows majority of the applicants are fully paid but some followed with charged-off so it gives a valuable insight about the financial stability & debt repaying discipline.

# upGrad

## 4. Purpose



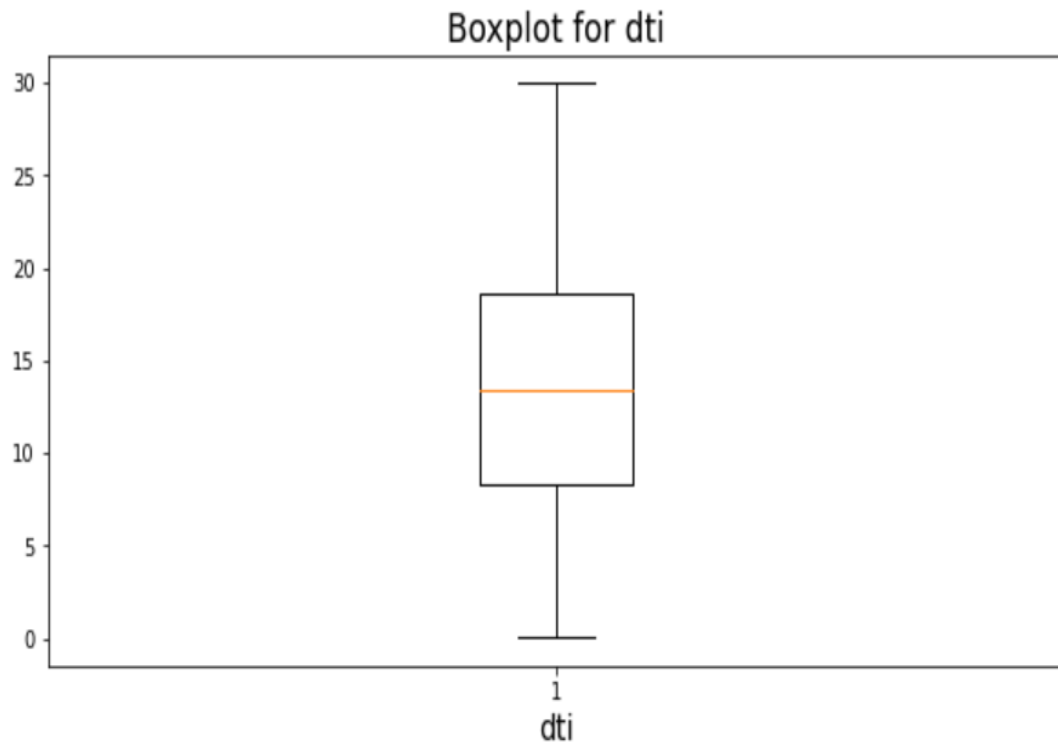
Purpose is an important driver or more of a factor that needs to be understood as this will help to understand the client intention & implementation of loan amount is in fruitful purpose or not.

It shows majority of the applicants applying for the purpose of debt consolidation which shows high number of applicants are finding it tough to pay their current obligations which is a red flag or signal to keep an eye on other drivers for a better understanding of the applicant financial scenarios.





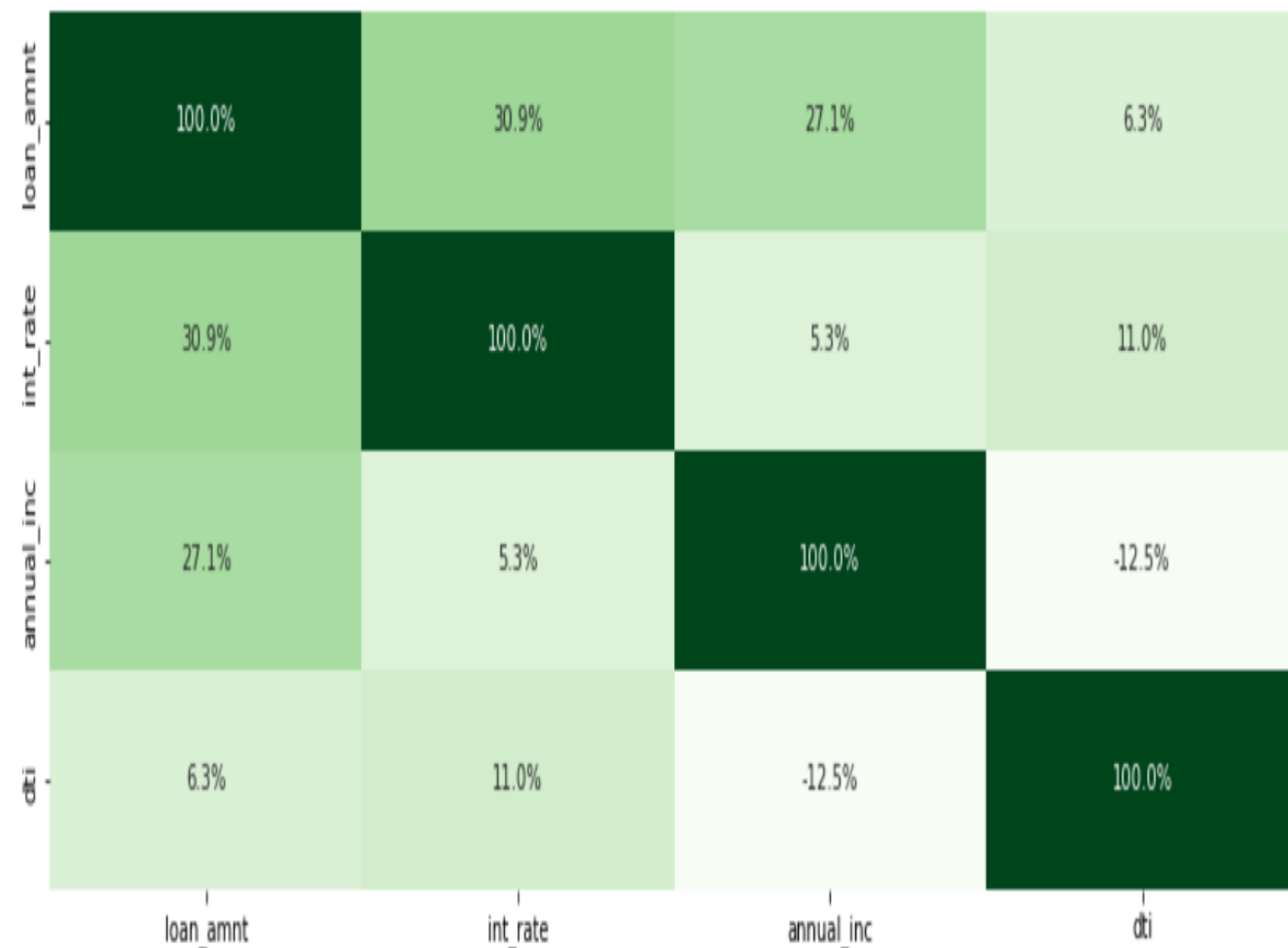
## Transformation of DTI driver



DTI driver had some null values but as we analyzed with the help of box plot that DTI is having stationary or static data with literally no outliers, so we imputed those null values with mean. Had it been with outliers then we might have considered median or interquartile difference method for imputation of null values.



## Bivariate Analysis of Continuous Variables



Bivariate analysis has been done with the support of heat map & correlation as multiple drivers as can be grouped in one plot.

Loan amount is one of the relevant drivers to understand the requirement of the consumer & interest rate and annual income are the other drivers that can clearly justify that the client is going to default on the applied loan amount as they are correlated.

DTI is another driver that has been shortlisted as the relevant driver as it helps to understand FOIR, or the current financial stability based upon his current obligations & income even if it is not having good correlation with any of the other drivers. As we can clearly see that int\_rate & annual\_income has good correlation with loan amount.



The code of this analysis can be accessed through the following GitHub link:

<https://github.com/theshivmalhotra/LendingClub>