Lending Club - Case Study

EDA - Presentation

CONTRIBUTORS:

- UDAY KUMAR ADAVI
- BIBHU SUNDAR SAHOO

Lending Club business

Business Understanding:

Lending Club is a consumer finance company which specializes in lending various types of loans. The company receives a loan application, the company must decide for loan approval based on the applicant's profile. Certain factors of risks are associated while making this decision.

> **Loss of business -** Applicant is likely to pay the loan, and company is not approving the loan.
> **Financial Loss** - Applicant may default on the loan, i.e. not pay.

Business Objective:

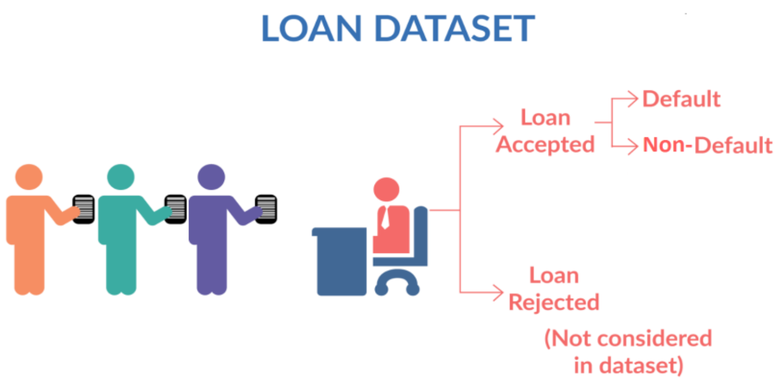
- Improve loan approval process to balance revenue from customer who fully pay their loans and financial loss due to customers defaulting

Problem statement

When a person applies for a loan, there are **two types of decisions** that could be taken by the company:

- **Loan accepted**
- **Loan rejected**

Lending Club would like for us to do an exploratory data analysis to the past loan applicant's dataset to understand consumer attributes and loan attributes that drive ***loan default behavior*** & thereby the ability to identify risky loan applicants.



Structured process was followed to solve this problem

1. Business problem framing
2. Data Sourcing & understandin g
3. Data cleaning & handling
4. Exploratory Analysis - univariates
5. Exploratory analysis of Target variable relationships
6. Summary of findings, insights & recommenda tions

Data considered for the EDA

Consumer Attributes
- Annual income
- Employment Length
- Home ownership
- Debt-to-income ratio
- Past delinquencies/bankruptcies
- Earliest credit line open
- State/Postcode
- Revolving balance & utilisation
- Employment title

Loan Attributes
- Loan amount
- Funded amount
- Instalment
- Interest rate
- Term
- Grade
- Verification status
- Title
- Loan Purpose

Input variables of interest for analysis

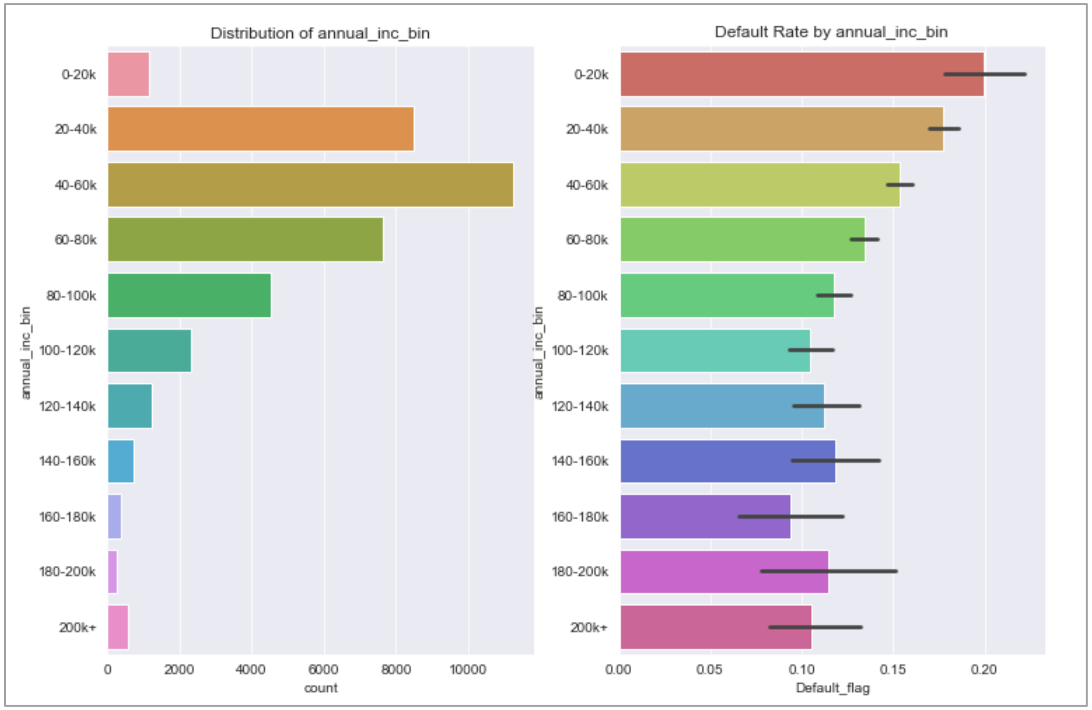
Loan Status
- Fully Paid
- Charge-off

Default Flag
- Loans for which status was Charge-off

Target variable for analysis

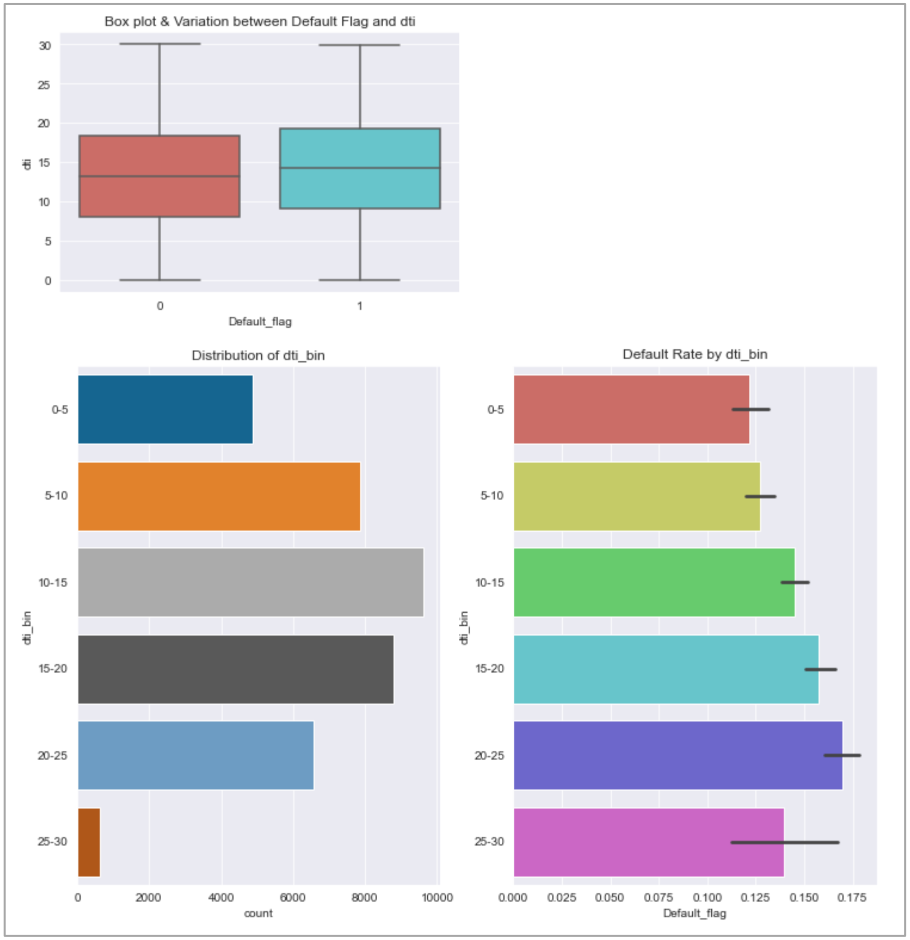
Data analysis illustration - Numerical variable example#1

Annual Income of customer



- As expected lower incomes have higher default rate, atleast until 80k
- And then onwards with higher annual income the default rate is on lower side

Data analysis illustration - Numerical variable example#2

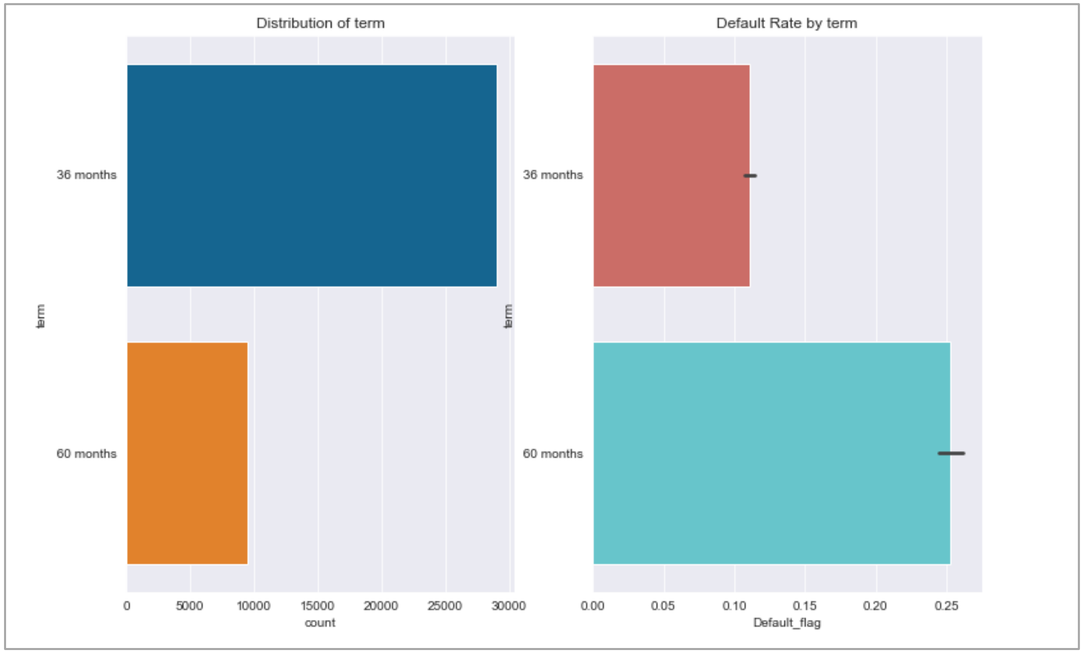


Debt to income ratio of customer

- As expected, higher the DTI, higher the default rate
- Thus it's risky to offer loan to customers with high DTI, especially >10

Data analysis illustration - Categorical variable example#1

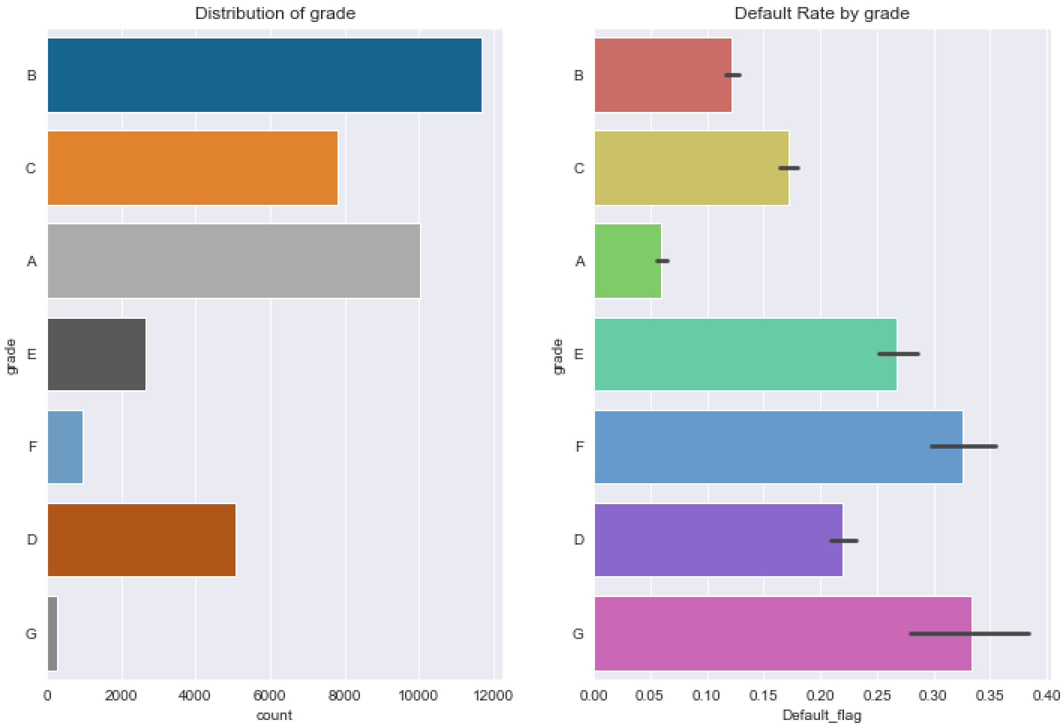
Term of loan



- Clearly, '60 months' term group has much higher default rate than 36 months

Data analysis illustration - Categorical variable example#2

Grade of loan



- It is unclear how the grade and subgrade are calculated, but there is definitely a positive correlation between default rate and grade, subgrade
- The lending club grade and sub grade are related. The grades are in an ordinal scale where starting scale like A1,B1 have lower default rate which further increases by the end of the ordinal scale. Lower alphabet grade(A,B...), subgrade(A1,B1) have lower default rates

Key findings & recommendations

- Any loan applicant with **annual income less than 100K** should go through heightened scrutiny
- It's **risky** to offer loan to applicants with **higher DTI and Loan to income ratio**.
- The higher default rate in **high interest rate** category evidently says it's **risky to offer high interest** loan to customers.
- Any loan applicant who has a credit age below 5 years should go through heightened scrutiny.

Summary of analysis across all data -1

Varialbe/Feature	Type	Attribute name	Finding/Insight	Recommendation & Actionability
Loan amount	Continuous & Bins	Loan_amnt	For loan amount 15K and above we see an increasing trend of high default rate than low loan amount value	Any loan application with **loan amount above 15K** should go through **hightened scrutiny**
Annual income	Continuous & Bins	annual_inc	The lower annual income applicant have higher default rate than the higher income applicant. In addition, for income below 80K we found an increasing trend of high default rate	Any loan applicant with **annual income less than 80K** should go through heightened scrutiny
Debt to Income Ratio-DTI	Continuous & Bins	dti	As expected, we found higher default rate for high DTI applicant.	It's **risky** to offer loan to applicants with **higher DTI(>10)**
Interest rate	Continuous & Bins	int_rate	Lower interest rates show lower default rates. We assume this could be because customers are offered lower interest rate if they have good outstanding financial profile and capable of paying repayment	The higher default rate in high interest rate category evidently says it's risky to offer high interest loan to customers. Review the vost benefit analysis of Interest rate offerred to potential loss due to default
Loan to Income Ratio-LIR	Continuous & Bins	loan_to_income_ratio- derived from loan_amt and annual_inc	As expected, we found higher default rate for high Loan to Income Ratio of applicant	It's **risky** to offer loan to applicants with **higher Loan to Income Ratio(>0.25)**
Credit Age	Continuous & Bins	credit_age - derived from	We found slightly higher default rate for credit age below 5 years incomparion to higher credit age category. One hypothesis for the high default rate for this category can be young applicants who don't have good outstanding financial status. The above 5 years of credit age category showed a mixed trend.	Any loan applicant who has a credit age below 5 years should go through heightened scrutiny
open_acc	Continuous & Bins	open_acc	The customers with open credit line below 4 have higher default rate	Consider criteria for approving loan applications where customers have btween 0 and 4 open credit lines

Summary of analysis across all data -2

Varialbe/Feature	Type	Attribute name	Finding/Insight	Recommendation & Actionability
Term	Categorical	term	The 60 month term group have higher default rate than 36 month group	It's **risky** to offer **Loans of term 60 month** to applicants. Consider offerring variable rates with varied terms
Lending Club Loan Grade and Sub Grade	Categorical	grade, sub_grade	The lending club grade and sub grade are related. The grades are in an ordinal scale where starting scale like A1,B1 have lower default rate which further increases by the end of the ordinal scale G5,F5.	It's **risky** to offer loan to applications with **higher ordinal value Lending club grades and subgrades**.
Home Ownership	Categorical	home_ownership	Home ownership 'others' have the lowest frequency however the default rates are significantly high. 'Mortgage' type Home ownership has slightly lower default rate	Although the number of application with home ownership type 'Others' are significantly less but they should be marked as risk and highly scrutinized. Dig deeper into the cases that are under the 'Other' and ensure the data is capture appropriately going forward
Lending club Income Verification Status	Categorical	verification_status	We found the 'Not verified Status' has lower default rate than 'Verified status' which looks incorrect.	Further investigation should be done on the income verfication status reported by lending club
Delinq_2yrs	Categorical	Delinq_2yrs	The increase trend in delinq_2yrs count from 0 to 11 also increases the Default rate. However for 5 and 6 the default rate is less than others which may be because of low count.	Application which have **higher Delinq_2yrs** marked by Lending club should be **scrutinized further**
pub_rec_bankruptcies	Categorical	pub_rec_bankruptcies	Higher the bankruptcies, the higher is the defualt rate	Applicant with higher bankruptices record should be scrutinized further.
State Address	Categorical	addr_state	A state wide distribution chart would be difficult to read as there are lot of State values However, it can be noted, that there are some state that have higher default rate marginally it would interesting to map the states to see if there is any geographical relationship with default behavior	
Purpose	Categorical	purpose	debt_consolidation has the largest share of all the approved loan. small_business loans has the higest default rates. major purchases loans has the lowest default rate among all loan purpose.	
Loan issue date	Categorical	issue_y - derived from issue_d	The count of loans issued have gone up from 2007 to 2011 and the default rate went down during these year until 2011, where we see spike in default rate again. 2010 has lower default rate compared with 2011	From a business perspective, understand the reasons behind this trend to validate the finding