upGrad

#LifeKoKaroLift

Lending Club Case Study: Analysis

upGrad

Case Study : Lending Club

Developed By : Rajesh Sushir

Aditya Bhosale

upGrad

Agenda

- Problem Statement
- Analysis Approach
- Analysis
- Insights

Problem Statement

- About Lending Club and Business Objective

What is Lending Club?

Lending Club is a marketplace for personal loans that matches borrowers who are seeking a loan with investors looking to lend money and make a return.

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:

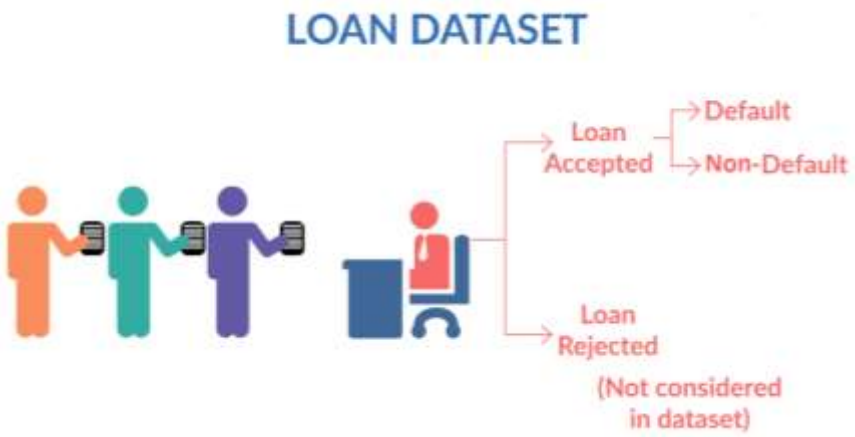
- If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company
- If the applicant is not likely to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company

Lending Club: EDA Case Study

upGrad

If one is able to identify these risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss. Identification of such applicants using EDA is the aim of this case study.

In other words, 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.



Fully paid: Applicant has fully paid the loan (the principal and the interest rate)

Current: Applicant is in the process of paying the instalments, i.e. the tenure of the loan is not yet completed. These candidates are not labelled as 'defaulted'.

Charged-off: Applicant has not paid the instalments in due time for a long period of time, i.e. he/she has defaulted on the loan

Lending Club: EDA Case Study

upGrad

The aim is to identify patterns which indicate if a person is likely to default, which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.

Analysis Approach

- About Pre Analysis Approach and Methods

Before Starting Actual Analysis, we have performed the below steps:

Data Understanding:

The given dataset is having lot of columns. Hence, first step is to understand about all those columns and check whether those columns are related to actual analysis or not.

There are 3 types of data available in given data set:

1. Customer Demographic Information
2. Loan Characteristics
3. Customer Behavioral Information

Data Cleaning:

The below steps are applied for Data Cleaning process:

1. Removal of customer behavioral columns
2. Removal of columns which are having all NULL values
3. Removal of columns which are having only 1 unique value
4. Removal of records for which we have NULL values
5. Filtered out data set based on Loan Status “Fully Paid” and “Charged Off”

After applying all the above steps, we have cleaned our data set and prepared for analysis and finding insights and patterns.

We have performed EDA on cleaned data set and tried to find relations and patterns.

upGrad

Analysis

- About Univariate, Bivariate and Multivariate Analysis

Univariate Analysis:

Univariate analysis is the simplest form of analyzing data. “Uni” means “one”, so in other words your data has only one variable. It doesn’t deal with causes or relationships (unlike regression) and it’s major purpose is to describe; It takes data, summarizes that data and finds patterns in the data.

Here, we are performing univariate analysis for the below variables

- Loan Amount
- Interest Rate
- Home Ownership
- Installment Per Month
- Year Month
- Purpose of Loan
- Grade
- Term
- Employee Length
- Annual Income
- Verification Status
- Loan Status
- Address State

Observation from Univariate Analysis:

1. Less people are requesting high loan amount.
2. Most of the Loans are shorter, with a term of 36 months.
3. Most of the borrowers are getting Interest Rate between 7.5% to approx. 14%.
4. People are taking less loan as Employee Length increases.
5. Most of the borrowers either have existing mortgages or renting their homes.
6. Around 42% borrowers have not verified.
7. Loan requests were increased after 2009 or 2010.
8. Maximum loans were taken at the end of Year i.e. from September month.
9. Around 85% borrowers had fully paid loans.
10. Most of the borrowers are taking loans for debt consolidation.
11. Loan amounts for the high-grade loans tend to be smaller than low grade loans.

Bivariate Analysis:

It is a methodical statistical technique applied to a pair of variables (features/ attributes) of data to determine the empirical relationship between them. In order words, it is meant to determine any concurrent relations (usually over and above a simple correlation analysis).

Here, we are performing bivariate analysis for the below variables

- Loan Status vs Purpose
- Loan Status vs Employee Length
- Loan Amount vs Grade
- Installment vs Income
- Loan Status vs State
- Loan Status vs Grade
- Loan Status vs Year
- Loan Status vs Home Ownership
- Loan Status vs Verification Status

Observation from bivariate Analysis:

1. There is good ration for loans taken for debt consolidation, credit card w.r.t Fully Paid & Charged Off.
2. Borrowers defaulted their loans for Higher Loan Grade.
3. There are more Charged Off for Verified Status as compare to Not Verified.
4. More loan requests from borrowers who were renting or mortgaging their homes to get extra income or pay their debts.
5. Charged Off ratio increases with Employee length also we can see decrease for loan requests for higher employee length.
6. Loan Grade increases with Loan Amount.
7. Higher installments as compared to Annual Income.

Multivariate Analysis:

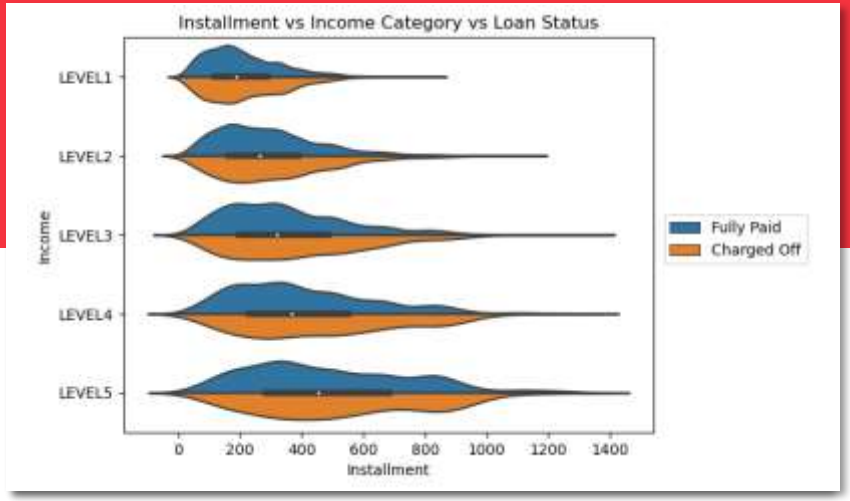
Multivariate analysis takes a whole host of variables into consideration. This makes it a complicated as well as essential tool. The greatest virtue of such a model is that it considers as many factors into consideration as possible. This results in tremendous reduction of bias and gives a result closest to reality.

Here, we are performing multivariate analysis for the below variables

- Installment vs Income Category vs Loan Status
- Income Category vs DTI vs Loan Amount
- Loan Amount vs Grade vs Loan Status
- Term vs Month vs Loan Status
- Loan Amount vs Funded Amount Inv vs Loan Status
- Loan Amount vs Verification Status vs Loan Status
- Home Ownership vs Purpose vs Loan Status
- DTI vs Income Category vs Loan Status
- Loan Amount vs Term vs Loan Status
- Purpose vs Month vs Loan Status
- Loan Amount vs Funded Amount vs Loan Status
- Income Category vs Purpose vs Loan Status
- Loan Amount vs Home Ownership vs Loan Status

upGrad

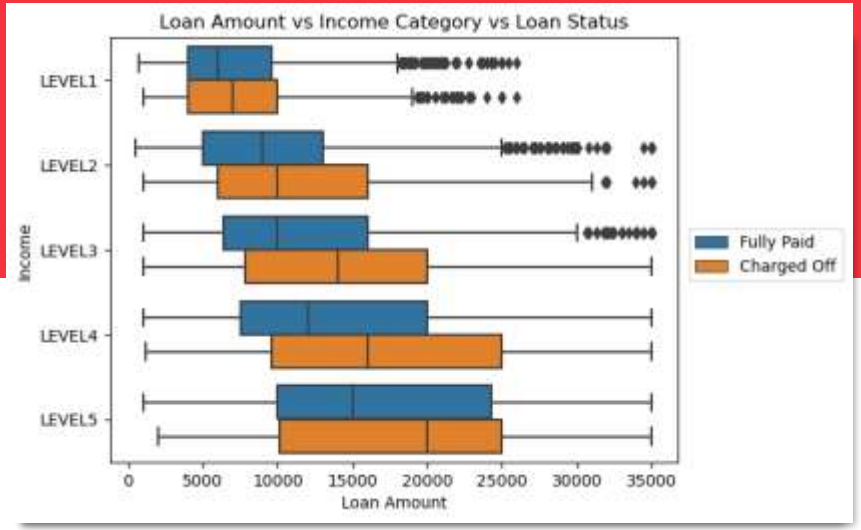
Lending Club: EDA Case Study



Observation

Higher Installments for any Income have more number of defaults.

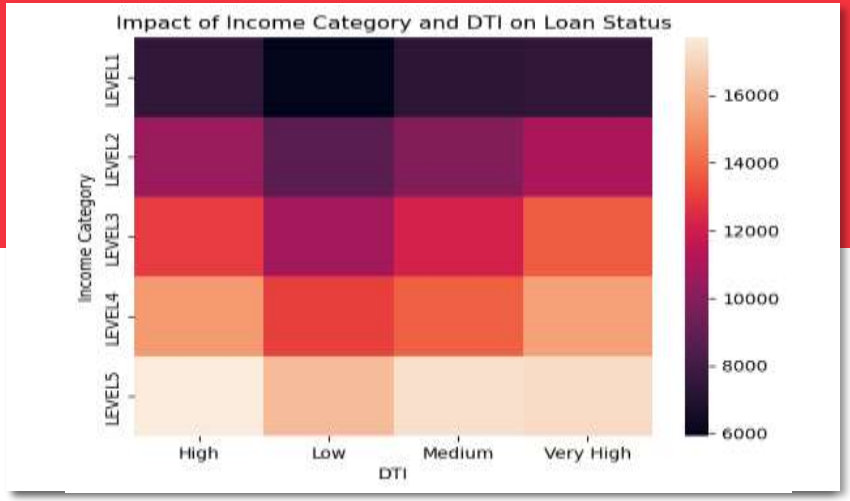
Lending Club: EDA Case Study



Observation

Loan Amount is dependent on Annual Income. If company lends more money without considering his/her annual income, then it may results in more defaults

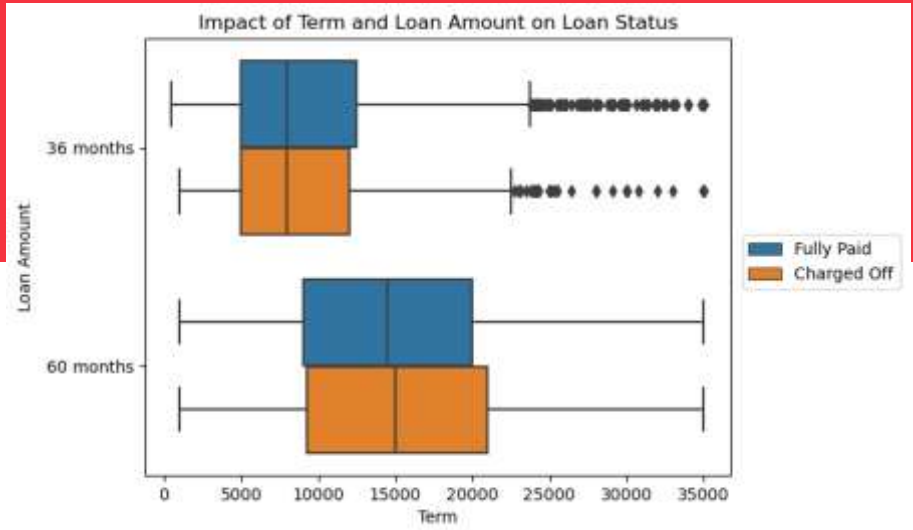
Lending Club: EDA Case Study



Observation

High DTI for lower income may result in more defaults.

Lending Club: EDA Case Study

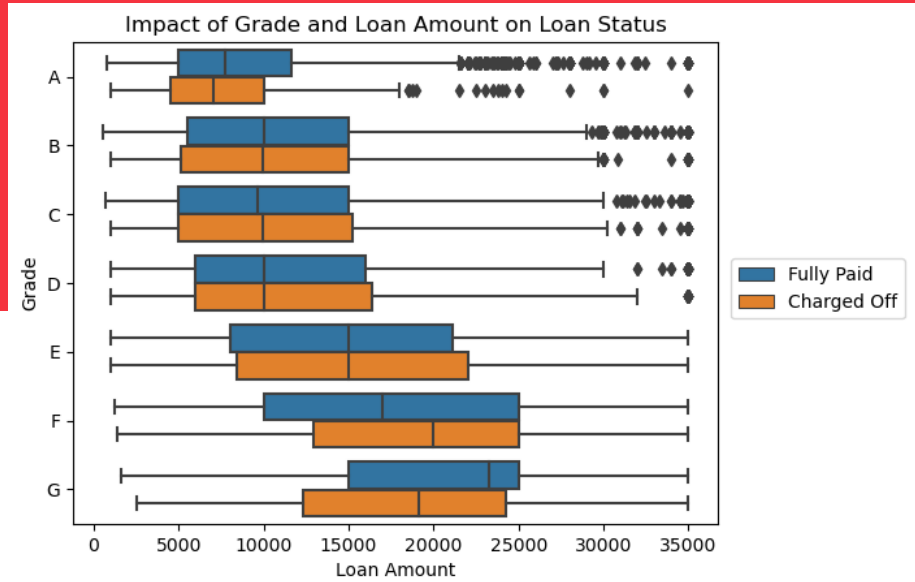


Observation

Loan Term for 60 months and for high amount are more likely to go default.

upGrad

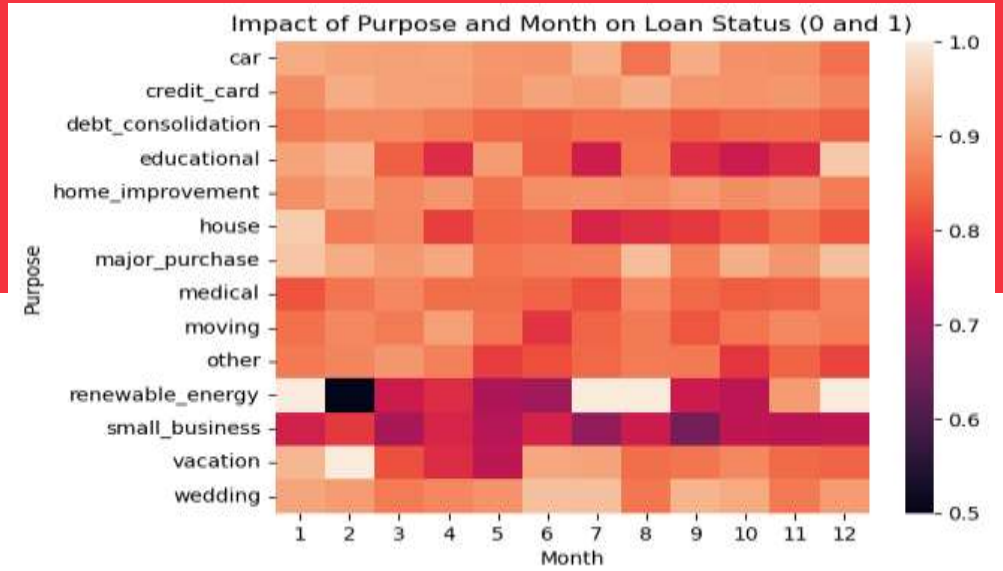
Lending Club: EDA Case Study



Observation

High Loan Grade may result more defaults due high interest rates.

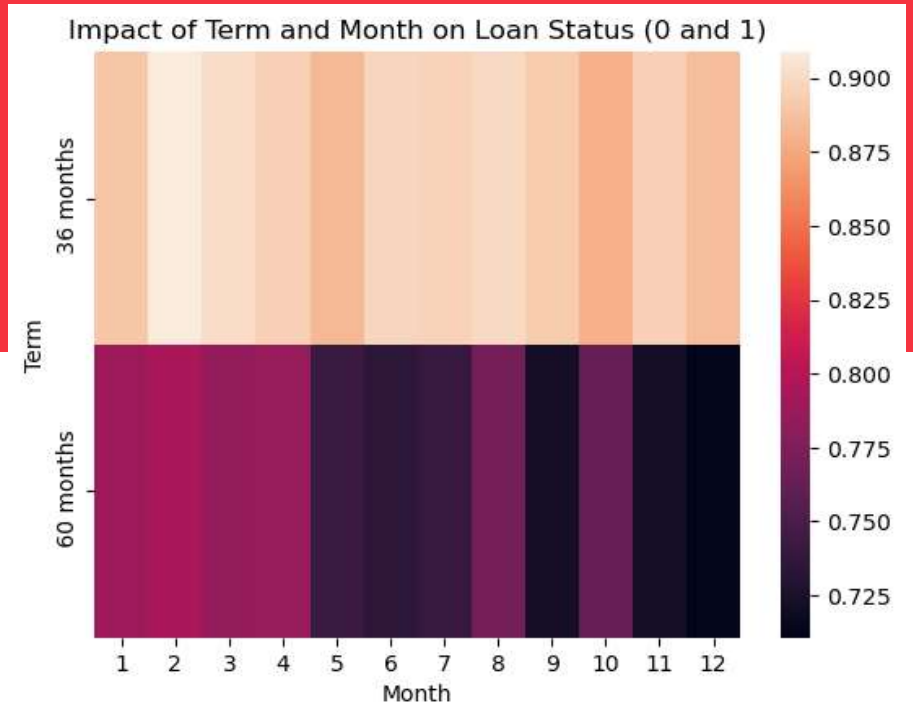
Lending Club: EDA Case Study



Observation

Loan requested for renewable_energy, small_business based on loan month are most likely driver for defaults.

Lending Club: EDA Case Study

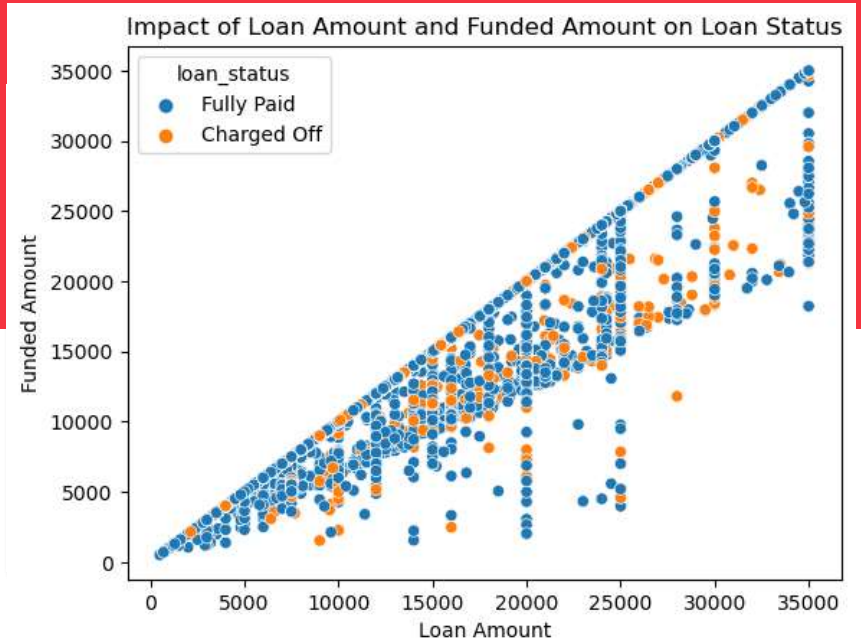


Observation

Borrowers are applying Loan for Term 60 months are more likely to defaults. Also there is relation with month of Loan Application.

upGrad

Lending Club: EDA Case Study

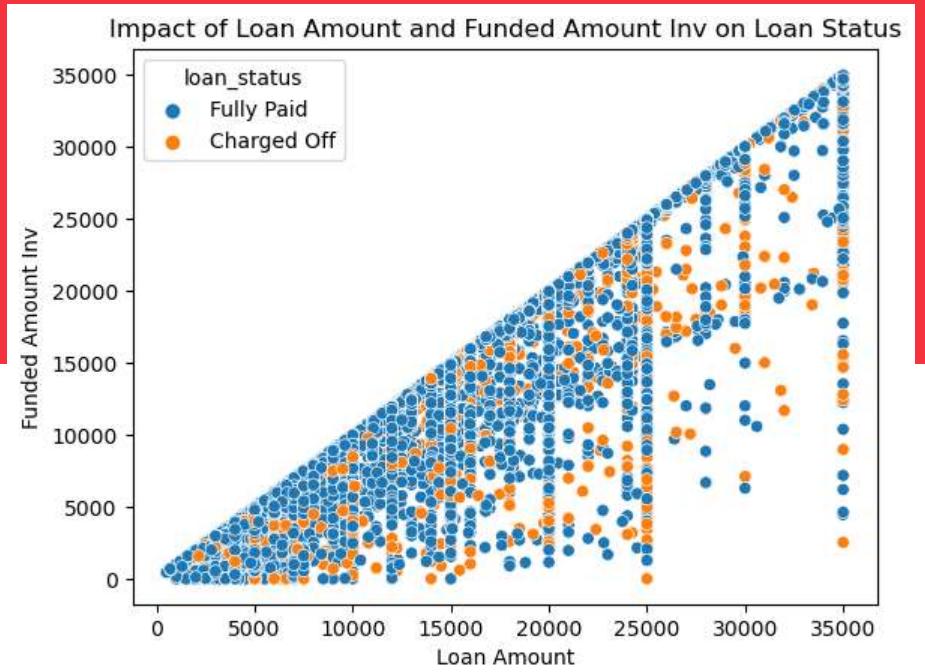


Observation

There is correlation between Loan Amount and Funded Amount. If both are same, then those are less likely to default.

upGrad

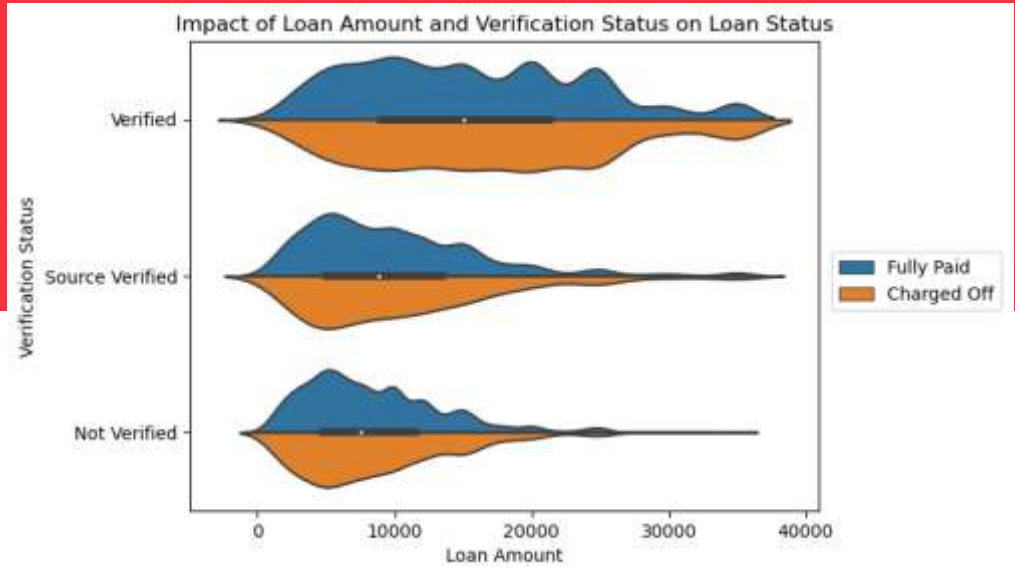
Lending Club: EDA Case Study



Observation

There is correlation between Loan Amount and Funded Amount Inverstor. If both are same, then those are less likely to default.

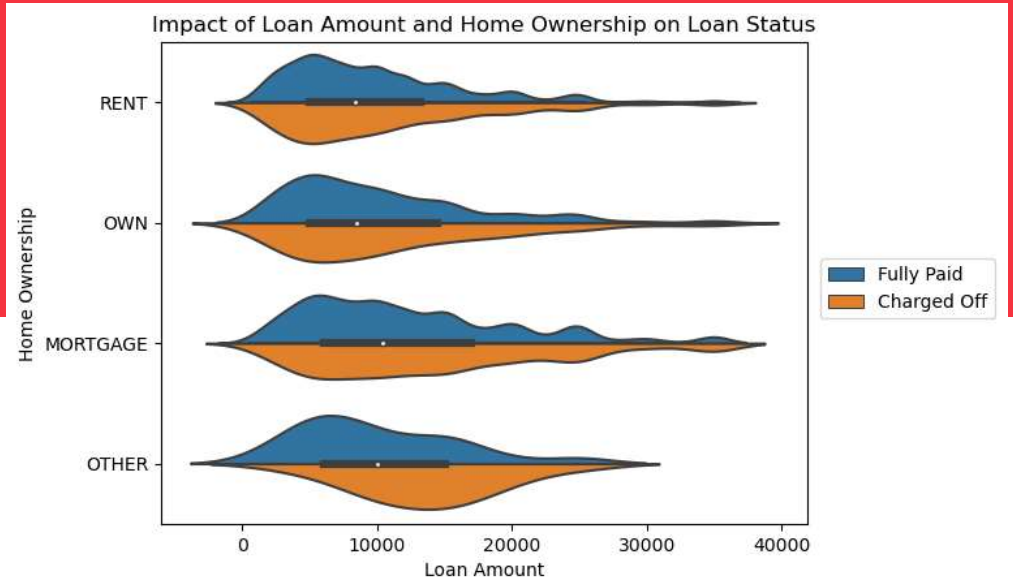
Lending Club: EDA Case Study



Observation

As we can see more defaults for Verified Status that means something is went wrong during verification.

Lending Club: EDA Case Study

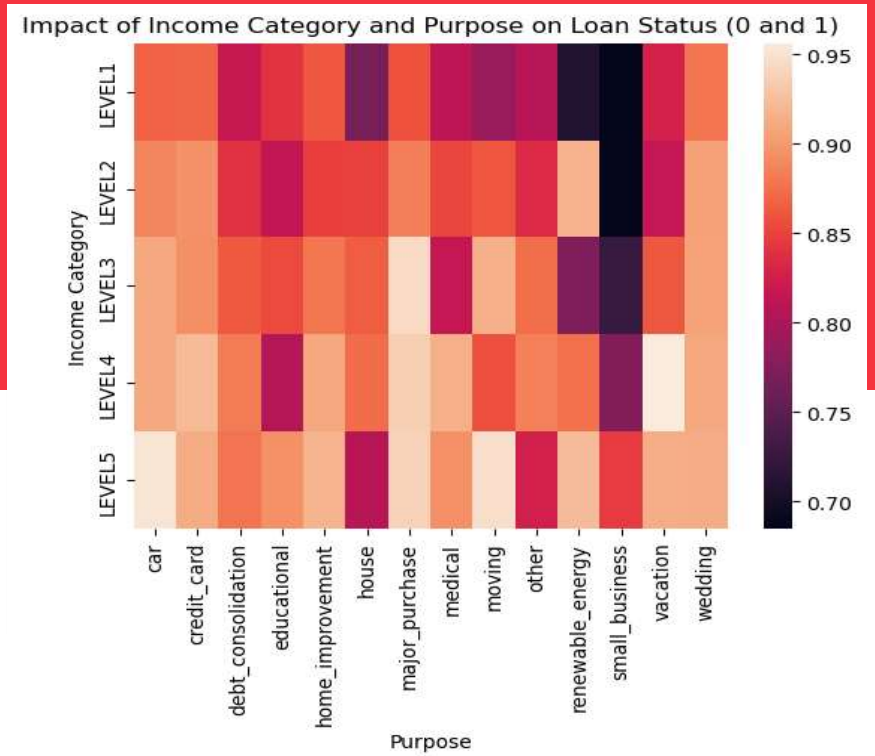


Observation

We can see more defaults for home ownership category OTHER.

upGrad

Lending Club: EDA Case Study

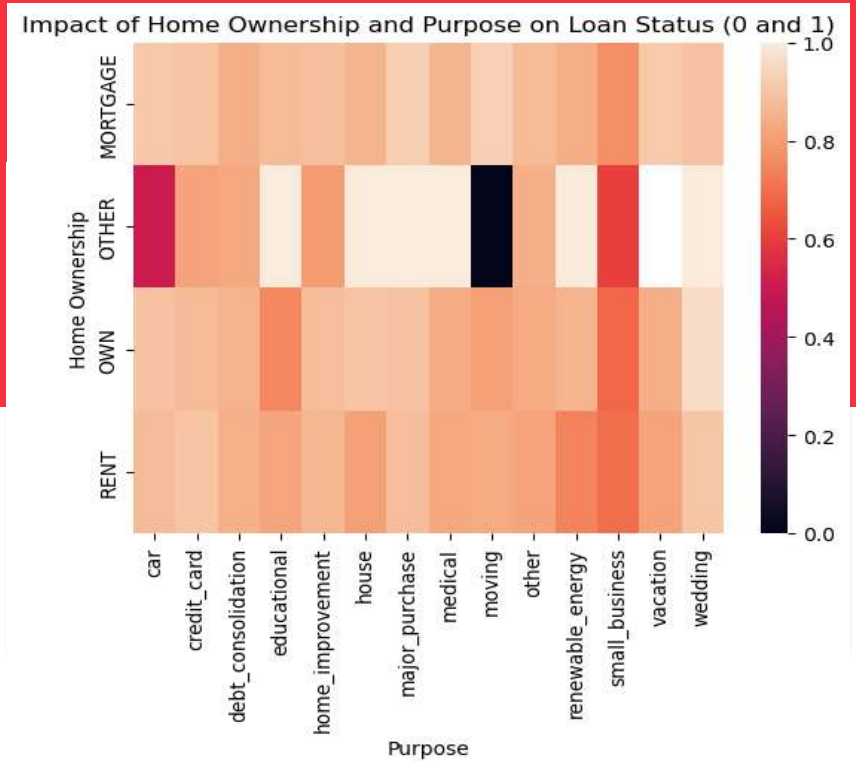


Observation

Defaults are more for lower income who are taking loan small business and renewable energy.

upGrad

Lending Club: EDA Case Study



Observation

Home Ownership - OTHER and loan purpose Car, moving, small business may go on more defaults.

Lending Club: EDA Case Study

upGrad

There are many factors which contributes for getting loan or rejecting loan. Please find the below drivers:

- Higher Installment
- Higher Loan amount
- Higher Debt to Income Ratio
- Loan Issue Month
- Higher Interest Rate
- Loan Terms (60 months)
- Loan Purpose (small business, renewable energy)
- Home Ownership (other)

Hence, giving Higher Loan Amount for High Interest to Lower or Middle level Income and for longer term whose have Home Ownership as OTHER may have more chances to go on default.

upGrad

#LifeKoKaroLift

Thank You!