



## CASE STUDY

Cassandra Pratt Romero

# ABOUT LENDING CLUB

This company 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.

APPLICANTS

There is a considerable number of **“risky”** customers

CREDIT LOSS

A risky customer **refuses to pay** or **runs away with the money owed**

COSTS

Operational costs are the same for all customers so there is a **big financial loss** (credit loss)

ANALYSIS

**How can we identify risky loan applicants** to reduce defaulted loans?

PROBLEM

# HOW TO FIND A SOLUTION?

## GET THE DATA

Obtain historical records of given loans that contain complete data for a loan at a given time.

## CLEAN AND ANALYSE THE DATA

Identify the fields that do not add information for our objectives, clean them out, and perform Exploratory Data Analysis.

## OBTAIN DATA INSIGHTS

Compare fields to review how they affect the final loan status, and point out relationships.

## IDENTIFY DEFAULT INDICATORS

Pick the fields that negatively affect the most the result of the loan status as indicators for default, **to prevent future loss.**



## DATASET DESCRIPTION

### 111 COLUMNS

To be cleaned up according to their content and importance for the analysis.

### 39,717 RECORDS

Each record describes a loan.

### 3 DIFFERENT LOAN STATUS

Fully Paid, Charged Off, and Current.

Current loans will be discarded as we are interested in terminated records.

# DATA CLEANING

## ALL-EMPTY VALUES COLUMNS

Drop variables for which all values are empty (null values).

## SINGLE VALUE COLUMNS

These kind of variables do not add information for the results as their variance is zero.

## COLUMN DESCRIPTION

Drop variables that were updated after loan approval or meaningless variables.

## RECORDS WITH MISSING VALUES

Assign column median value to column's missing values.

## “CURRENT” STATUS LOANS

Drop records with loan status “Current” and preserve only the paid and defaulted ones.

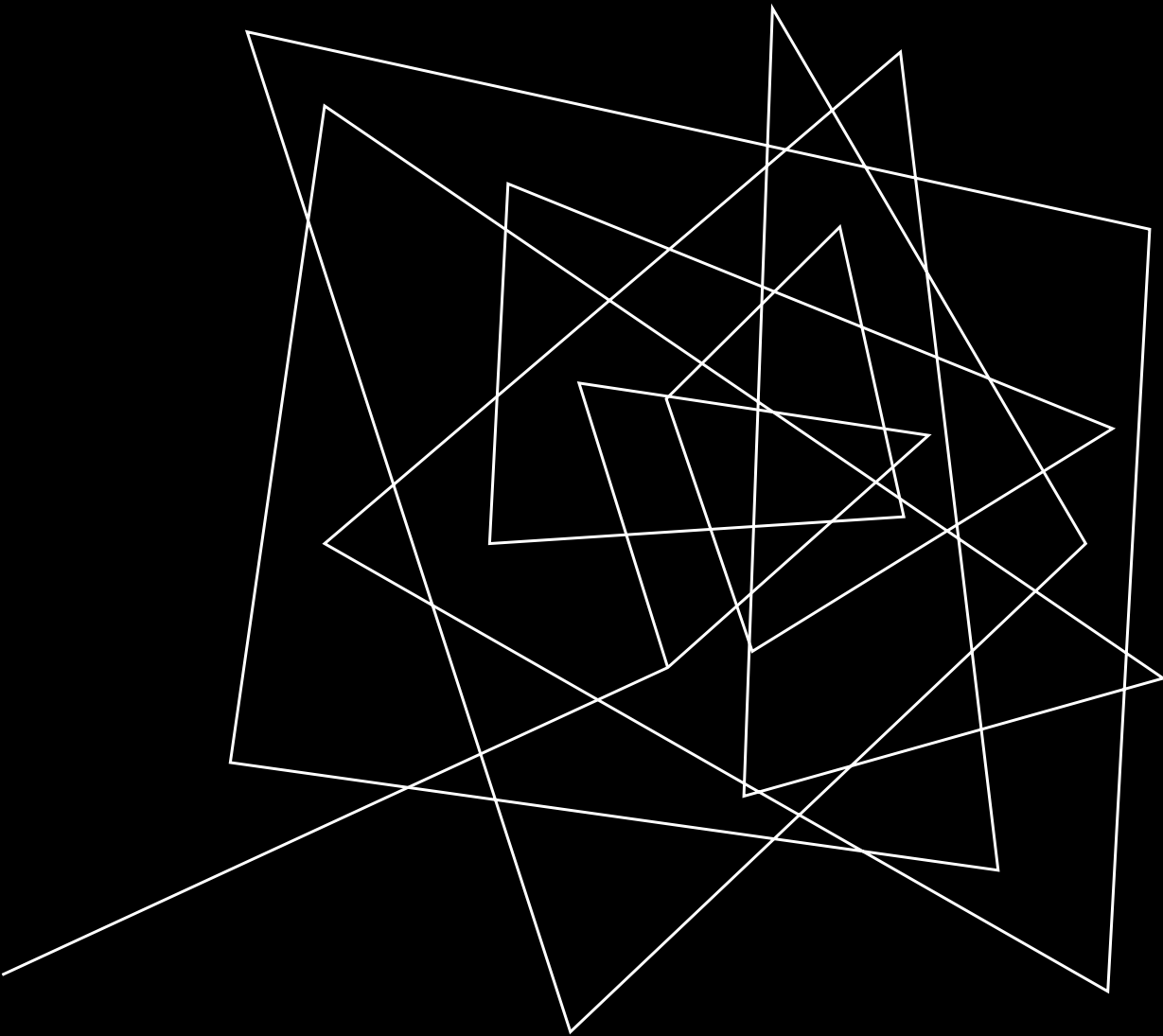
# VARIABLES TO ANALYSE

## Categorical variables

- earliest\_cr\_line
- emp\_length
- grade
- home\_ownership
- loan\_status
- purpose
- term
- verification\_status

## Quantitative variables

- annual\_inc
- delinq\_2yrs
- dti
- inq\_last\_6mths
- int\_rate
- installment
- loan\_amnt
- open\_acc
- pub\_rec
- pub\_rec\_bankruptcies
- revol\_bal
- revol\_util
- total\_acc



# ANALYSIS RESULTS



# HANDLING OUTLIERS

An outlier is an **unusual value in the dataset**.

For **quantitative variables** we identified and removed outliers:

- Borrowers with annual income greater than \$145,000.
- More than \$28,857 for total credit revolving balance.
- Borrowers with installment greater than \$749.345.
- Borrowers that requested loans greater than \$25,000.
- More than 3 incidence of delinquency in last 2 years.
- More than 3 inquiries in the past 6 months.
- Loans with interest rates greater than 22.13%.
- Borrowers with more than 25 open credit lines (and that is already a big number!)
- More than 1 derogatory public record or more than 1 public record bankruptcy.
- Borrowers with total credit lines greater than 48.

## SIMPLIFYING VARIABLES

For **categorical variables** we identified areas of improvement for analysis:

- Kept only the numerical values for the employment length (i.e. “10+ years” was converted to 10).
- From 3 verification status available we went down to 2: Verified, and Not Verified.

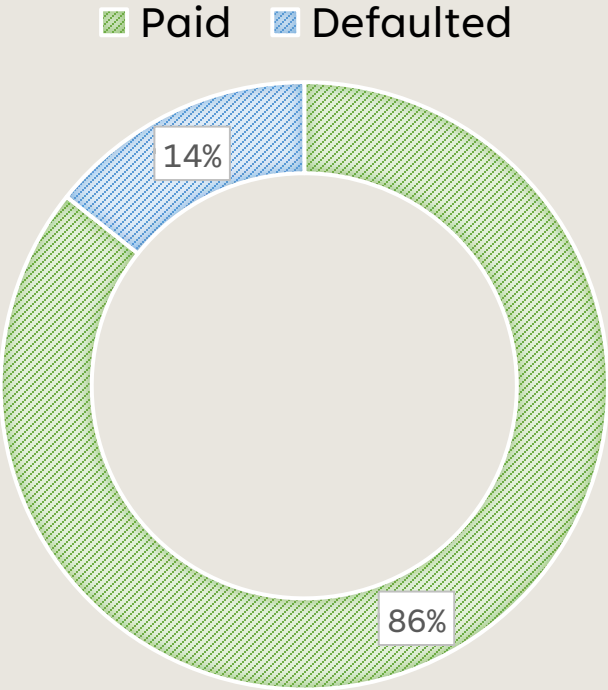
## DATA ENRICHMENT

- Month and year were extracted from the earliest reported credit line date.
- For interest rate and revolving utilization rate, 3 categories were created: Low, Medium, and High.

# TOTAL LOANS INSIGHTS

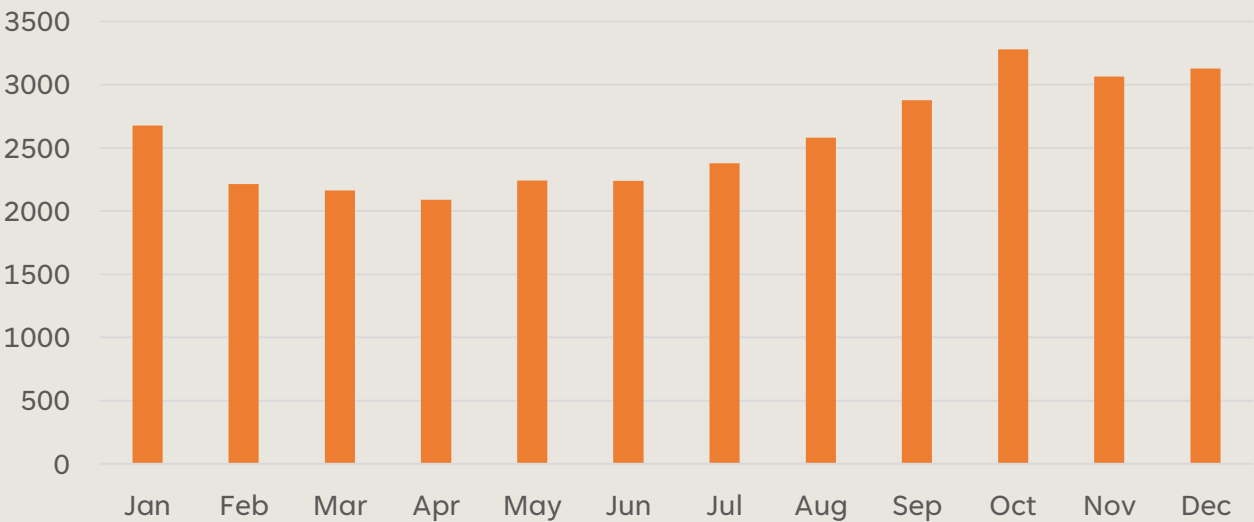
After cleaning we have 30,925 loan records.

## LOANS BY STATUS



Towards the end of the year and in January the total loans requests increase. Most likely, it is due to the **expenses that are made at the end of the year and the debts accumulated throughout the year.**

## LOANS BY REQUEST MONTH



## INDICATORS OF DEFAULT (1)

Following variables were found to have an important impact over the total amount of defaulted loans.

**Marked in bold** are the **strong indicators of default**, accordingly to their distributions and behavior by loan status.

### CATEGORICAL VARIABLES

- **emp\_length**
- **purpose**
- grade
- home\_ownership

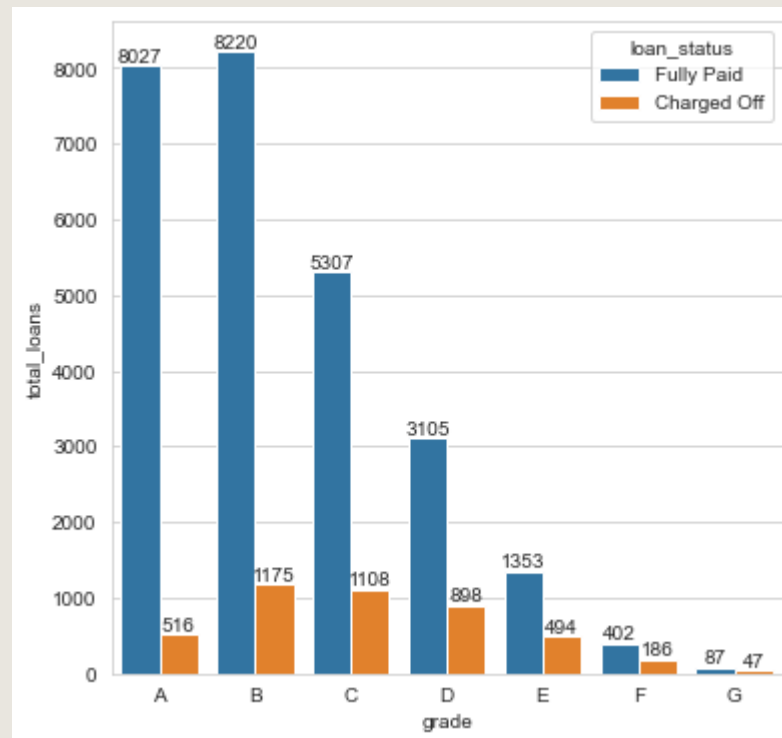
### QUANTITATIVE VARIABLES

- annual\_inc
- dti
- **inq\_last\_6mths**
- installment
- **int\_rate**
- loan\_amnt
- **revol\_util**

## INDICATORS OF DEFAULT (2)

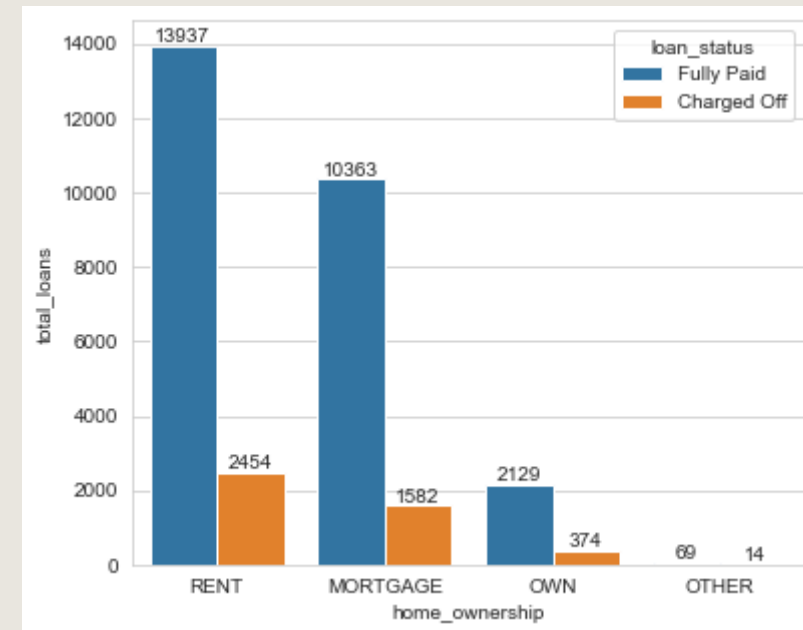
### GRADE

The lower/worst the grade assigned the more percentage of defaulted loans within the grade counts.



### HOME OWNERSHIP

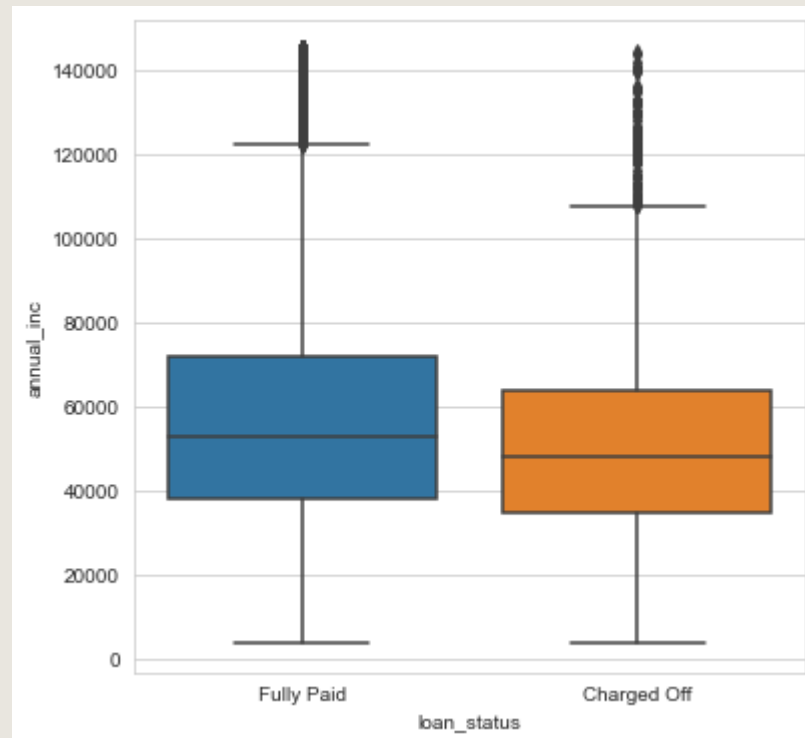
Mortgage and Rent categories of home ownership have the highest percentage of defaulted loans (13.05% of total).



# INDICATORS OF DEFAULT (3)

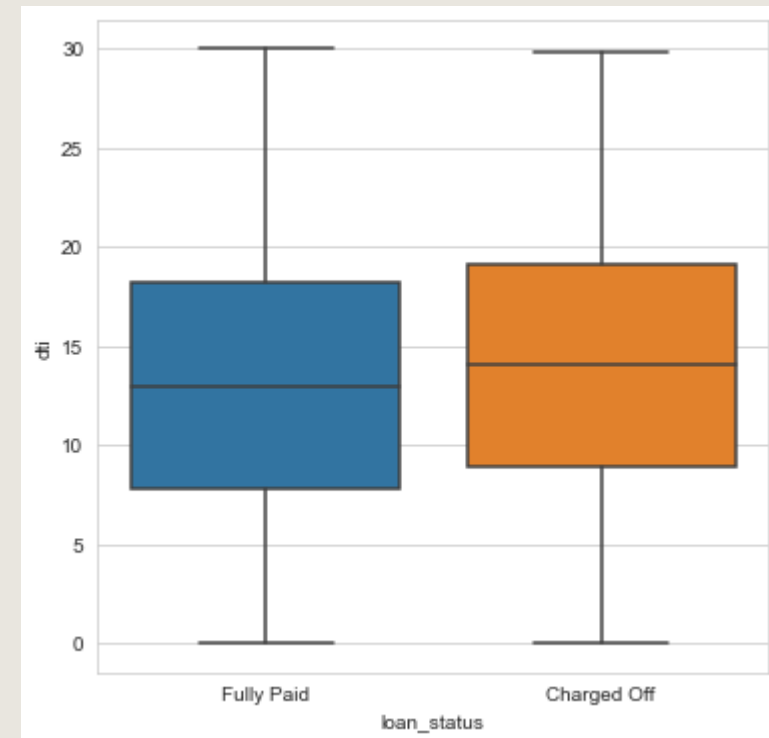
## ANNUAL INCOME

Borrowers with defaulted loans have **an annual income in general smaller** than people who paid the loan (common values between 4,080 and 144,240).



## DTI

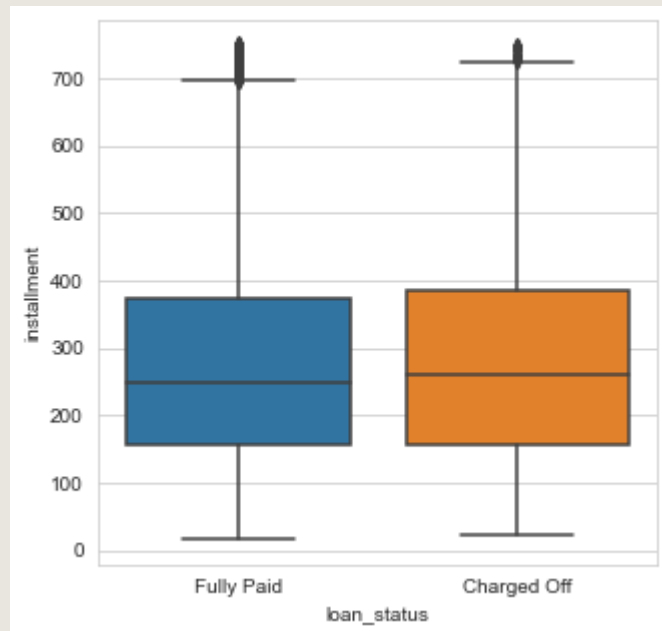
Borrowers with defaulted loans **make in general greater monthly debt payments** than the ones who paid the loan (75% of time having up to 19.15 DTI).



# INDICATORS OF DEFAULT (4)

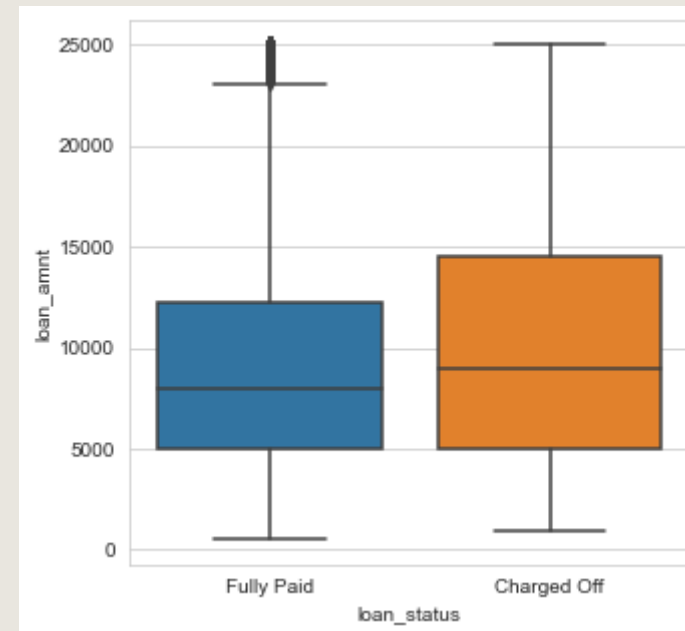
## INSTALLMENT

Borrowers with defaulted loans **were given in general higher monthly payments** than the ones who were able to pay. Installments for defaulted loans start at 22.79 while for paid ones starts at 15.69.



## LOAN AMOUNT

Borrowers with defaulted loans **asked in general for greater loan amounts** than the ones who were able to pay (75% of time going up to 14,500 and with a maximum of 25,000)



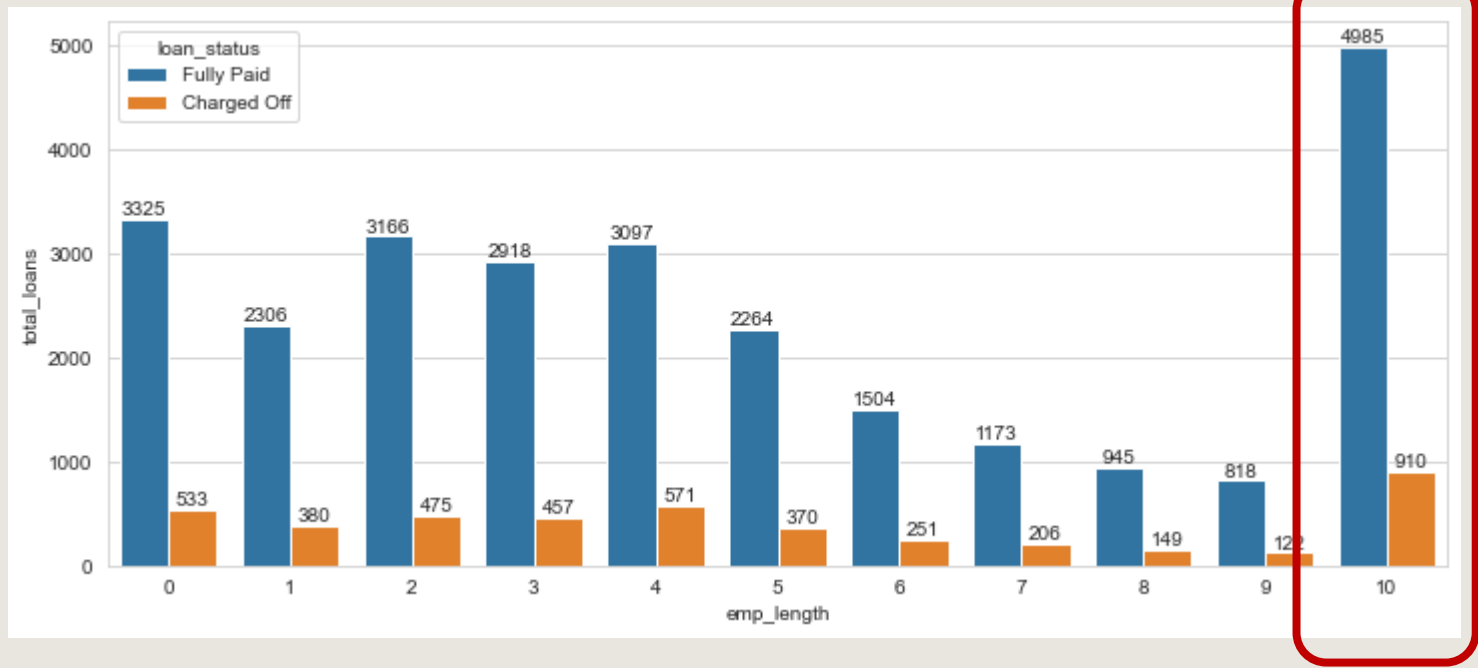
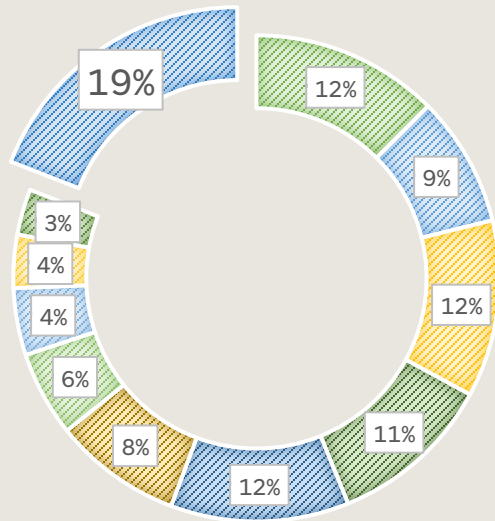
# STRONG INDICATORS OF DEFAULT (1)

## EMPLOYMENT LENGTH

People that has been **employed for more than 10 years** are the ones who mostly request loans (**19.06%**) and that end up in default (2.94% of total loans).

### LOANS BY EMPLOYMENT LENGTH

0 1 2 3 4 5 6 7 8 9 10

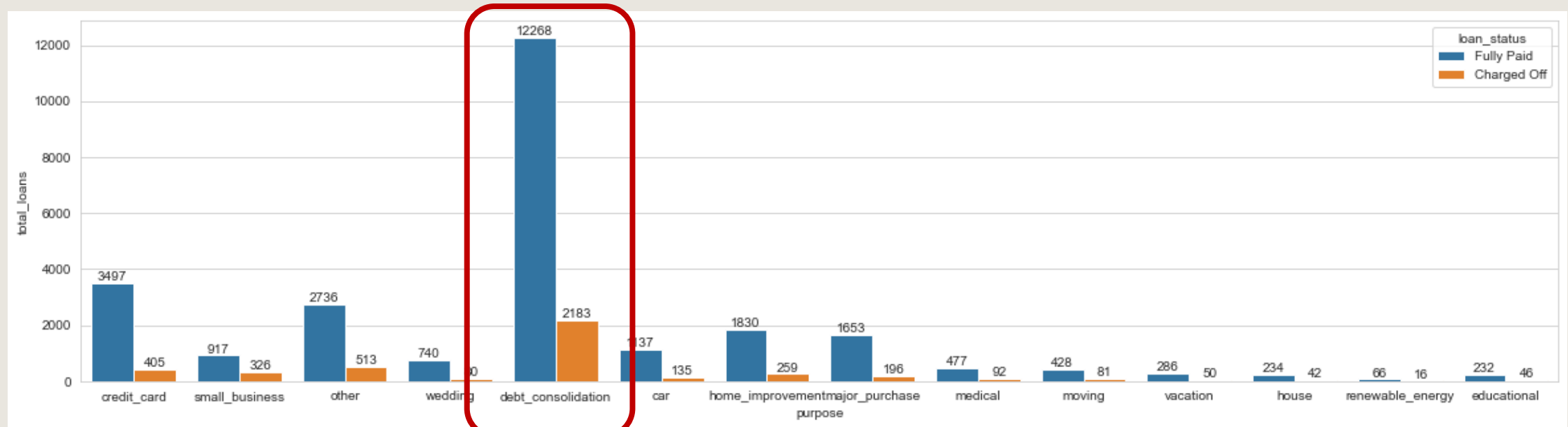




## STRONG INDICATORS OF DEFAULT (2)

### PURPOSE

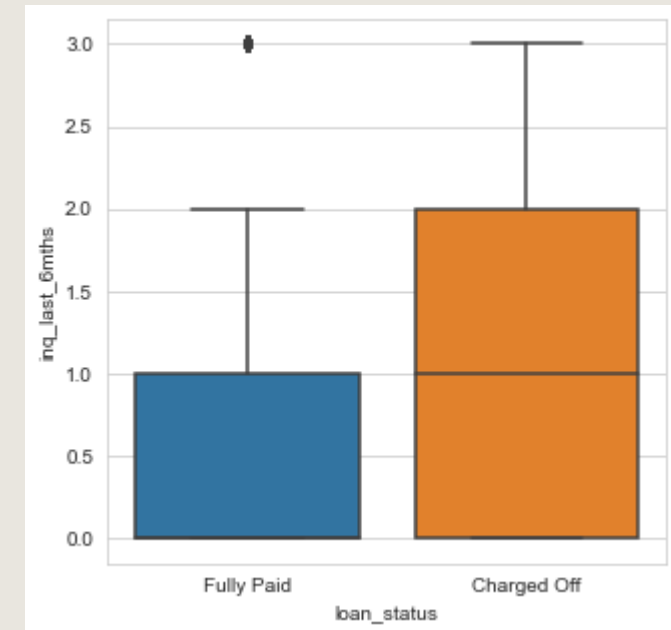
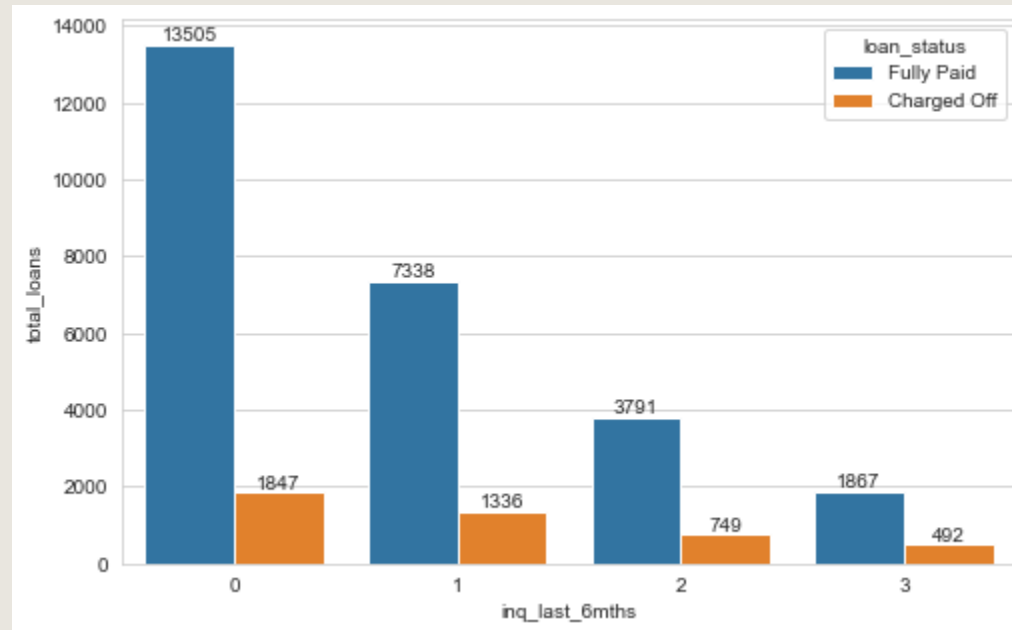
**"debt\_consolidation"** has the maximum count for loans (47.72% of total), and the category also includes the maximum value for defaulted loans (7.05% of total).



# STRONG INDICATORS OF DEFAULT (3)

## NUMBER OF INQUIRIES IN PAST 6 MONTHS

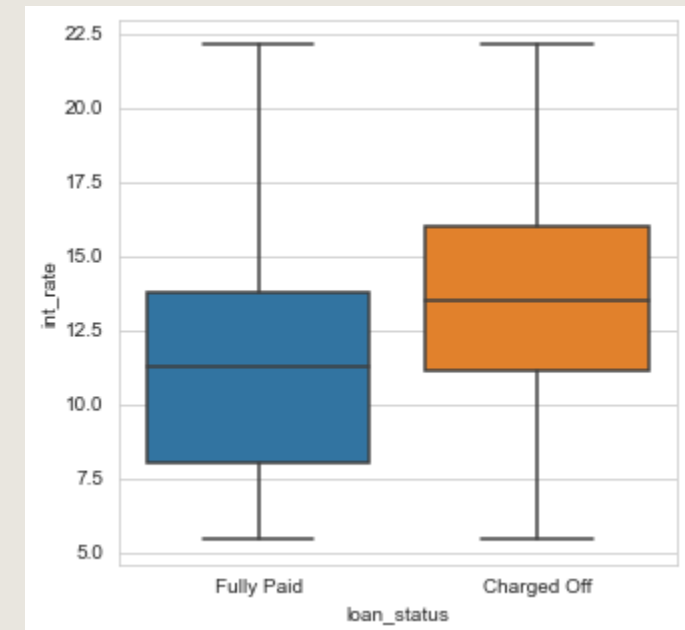
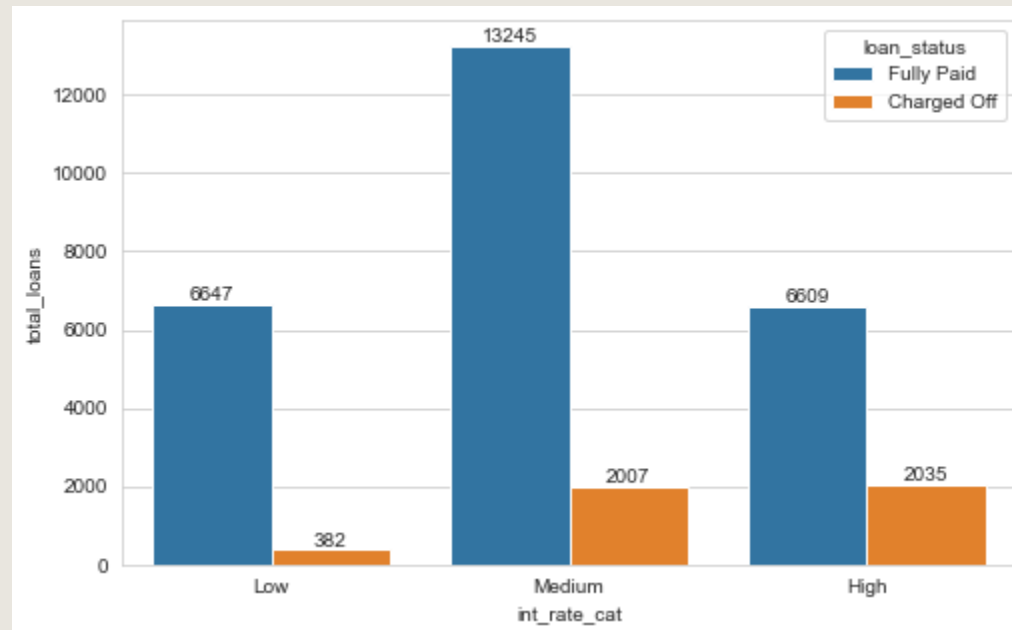
Borrowers with defaulted loans **have more inquiries done to their debts in last 6 months**, which is a strong indicator of them looking more for lenders: 75% of inquiries for defaulted loans are between zero and two, while for paid ones goes up to one.



# STRONG INDICATORS OF DEFAULT (4)

## INTEREST RATE

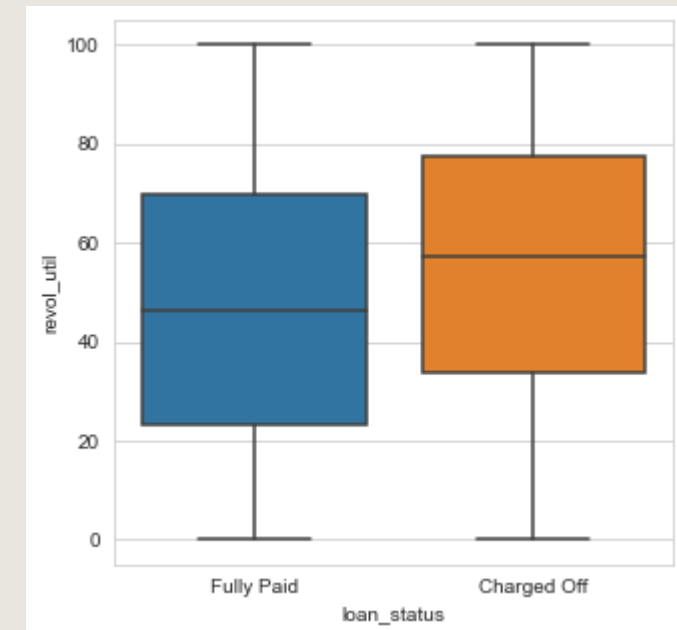
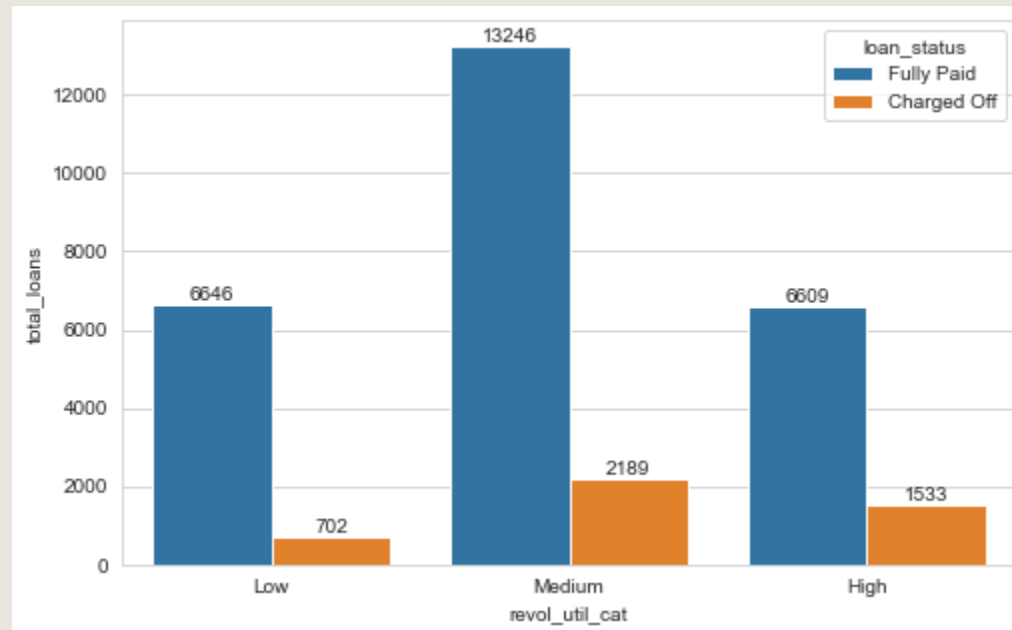
Borrowers with defaulted loans were given **in general higher interest rates** (medium and high categories) than the ones who were able to pay (75% of time going up to 15.99% with a maximum of 22.11%)



# STRONG INDICATORS OF DEFAULT (5)

## REVOLVING LINE UTILIZATION RATE

Borrowers with defaulted loans **use more credit relative to all available revolving credit** (medium and high categories) than the ones who were able to pay: 75% of borrowers have up to 77.5 utilization rate, while for paid loans they have up to 69.6.



## RECOMMENDATIONS

### DEFINE A “RISKY PROFILE”

According to the data analysis, a **risky profile** follows these characteristics:

- Employed more than 10 years
- Purpose of loan is “debt consolidation”
- Has more than one inquiry in the past 6 months
- The interest rate for the loan would be greater than 23.20%
- The borrowers revolving line utilization rate is greater than 8.00%

### KEEP ON UPDATING DATA

In order to keep valuable results, it is necessary to set up a scheduled check of data to run the complete analysis and update the risky profile.



## SUMMARY

It is possible to obtain indicators of default with historical dataset of loans.

In order to prevent future credit loss, it is important for Lending Club to toughen up the verification of borrowers with a risky profile and when possible, avoid lending money to them.

# THANK YOU

Cassandra Pratt Romero

[cprattr@gmail.com](mailto:cprattr@gmail.com)