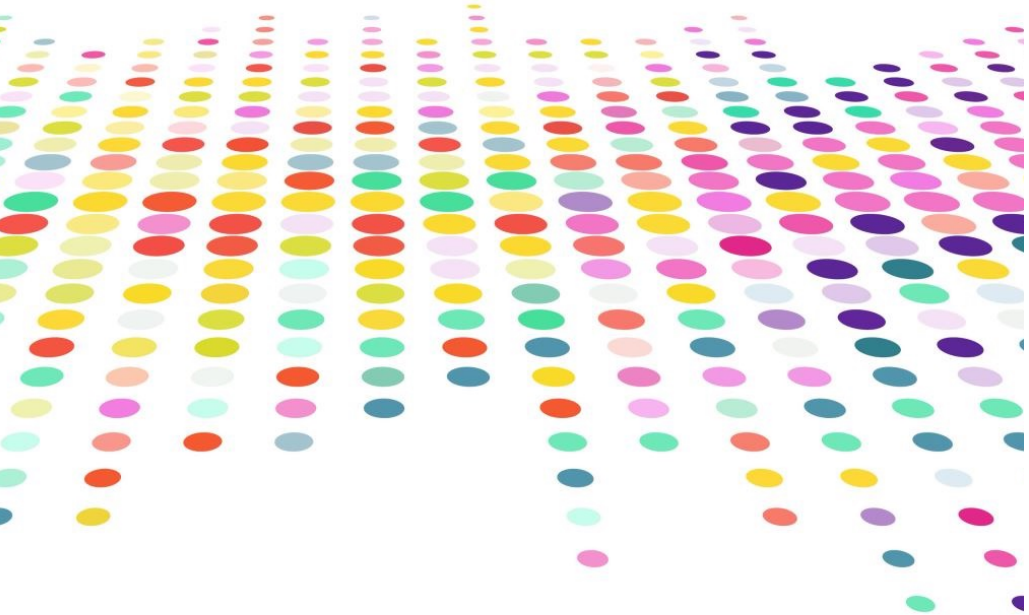
LENDING CASE STUDY



By Pankaj Singh

- Introduction
- Problem Statement
- Assumption
- Data Driven Analysis (Exploratory Data Analysis)
  - Data cleansing
  - Univariate analysis
  - Bi-variate analysis
  - Multivariate analysis
- Summary

INTRODUCTION

A consumer finance company, specializes in lending various types of loans to urban customers. It aims to cut down the financial loss by analyzing existing datasets across different customers.

Existing dataset contains customers who:

a) Fully paid the loans
b) Still paying the loan
c) Who are charged-off (defaulted)

PROBLEM STATEMENT

Identify patterns & attributes which indicate customers with certain characteristics are more likely to be charged-off.

ASSUMPTION

- Case study is limited to identify behavior for new customers by analyzing existing dataset.
- For existing charged off customers, lending back may make appropriate details.

DATA CLEANSING

Clean datasets by running a series of operations to reduce error.
Operations performed on listed below:

- Drop duplicate records
- Dropped columns with all values as NA
- Dropped columns which is not required for applicants applying for loans
- Fix data types & Data imputation (emp_length, int_rate, pub_bank_ruptcies)
- Filter rows, columns containing zero/na values
- Filter rows where loan amount < funded amount & funded amount by investor
- Derive the columns for further analysis (For ex: Extract year and month from date column)

Results:

Data set	Records count before data cleansing	Records count after data cleansing
Rows	39717	37868
Columns	111	44

UNIVARIATE ANALYSIS

- A single attribute of dataset is analyzed, and behavior is noticed.
- Attributes analyzed for univariate analysis are listed below:
  - Loan status
  - Loan issued for the purpose
  - Consumer housing ownership impact on loan
  - Interest rate
  - Annual Income
  - Loan amount issued to consumer
  - Loan amount issued in year, month

UNIVARIATE ANALYSIS (SUMMARY)

- Out of all the loans 14% of loans are defaulted.
- Consumers with salary < 250K contributed most to the defaulted loan:
  - People with higher salary can easily pay up the loan

- Most of the loans were issued in month of dec.
  - Consumers defaulted most in month of dec as the financial year tax statement starts towards the end.

- Rent & Mortgage contributed most into the defaulted loan.
  - In Urban areas, consumers renting out the flat or paying the mortages are defaulted most due to high proportion of their earnings going into this.

- Debt_consolidation, other, credit card, small_business, home_improvement contributed most into the charged off loans.
  - Consumers having small businesses, home improvement, and when debt is consolidated, they tend to default as they require large sum of money.

- Majority of the loan amounts were in range of 100,000- 150,000

BI-VARIATE ANALYSIS

- In bi-variate analysis, two variables are analyzed further to understand the relationships.
- Datasets attributes analyzed to understand the behaviors are listed below:

Behavior analysis against charged-off

- Annual income
- Purpose
- Grade/Sub-grade
- Interest rate
- Employees working experience
- Address state where loan was taken in
- Verification (i.e. Income verification status)
- Public bankruptcies (any record available)

Behavior analysis between different attributes

- Loan amount vs Purpose
- Interest rate vs Term of loan
- Grade vs Interest date
- Year vs Interest rate
- Loan amount vs Interest rate
- Current Debt ratio vs Interest rate
- Annual Income across grade

BI-VARIATE ANALYSIS (SUMMARY)

Most commonly observed behavior between variables are listed below:

- Almost all the grades & sub-grades contributed to the charged off loans except G.
- Consumers from interest rate range 10-15, 15-20 contributed most to the charged off loan.
- Consumers with higher exp. tend to default less compared with less exp.
  - ➢ It may be due to consumer with less exp. earns less
- Loans issued for higher tenure has high interest rate
  - ➢ Higher risk for consumers defaulting to loan in long term.
- Consumers from states GA, CA, TX, VA, FL, NY, PA, CH, IL, WA, NJ contributed most to the charged off loan
- Bad loans have high interest rate.
  - ➢ Consumers with bad loans may tend to default more.

MULTI-VARIATE ANALYSIS

- Consumer behavior is observed across different data sets attribute in pair plot.
- Consumer attributes used for analysis are:
  - Annual income
  - Loan amount
  - Interest rate
  - Loan issued year
  - Debt ratio (dti)
  - Employee experience

MULTI-VARIATE ANALYSIS

- High interest rate for high loan amount.
- Higher annual income has less dti
  - Consumer with higher income can easily settle their debt.
- Consumers had high income as the year progresses.
  - As Country economy progresses, consumers can earn more with the time.

- Consumer can borrow higher loan amount as the year progresses.
  - Economy increased as the year progresses forward, and consumer had benefit of higher loan approved

SUMMARY

- Consumers with more experience, can earn more over the period of time; can borrow high amount of loan with higher interest rates.
- Consumer loan amount should be reduced and examined further, where consumers are:
  - With salary < 250K
  - Across certain purposes (loan borrowed against) and states.
  - Almost all the Grades (Except G)