

Lending Club Case Study

Sudheendra Gogi (Sudhy)

Agenda

Objective

Features & Enrichment

Feature Analysis Outcomes

Recommendations

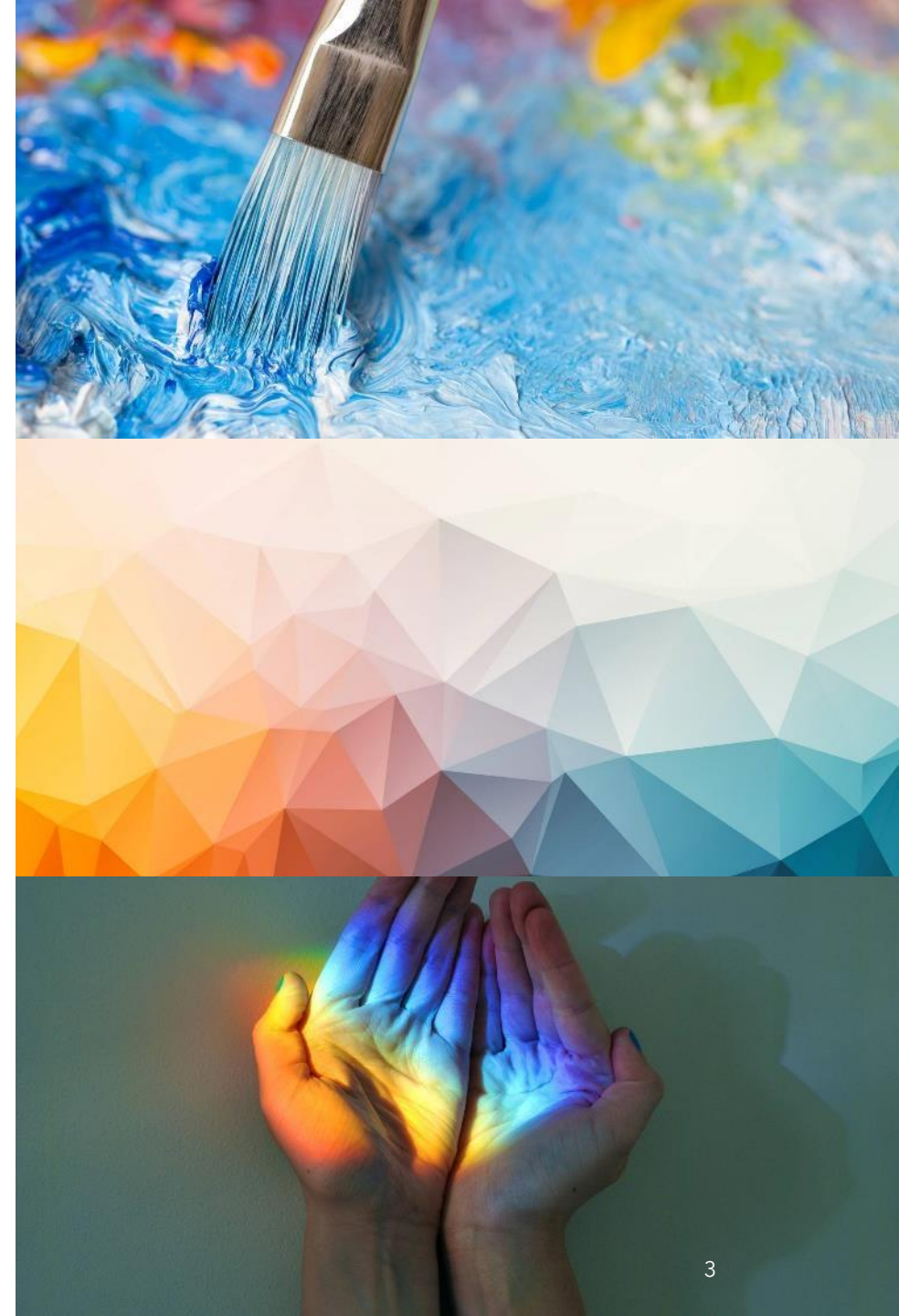
DataVisualization



Objective

Analyze, identify and perform risk classification of consumer loan applications to minimize the amount of credit loss.

Also, the analysis should list out strong indicators of default which will be leveraged for portfolio and risk assessment.



Features & Enrichment

Data Catalog

Data Enrichment

- Standardization, Mutations



Features & Enrichment

Data Catalog

- Reviewed and understood around 115 data elements, while many of the elements don't provide insights or has NA (Not Applicable), around 30% of them were helpful.
- Applicants FICO credit scores, employment tenure, DTI, Days past dues (DPD), last payment dates, Recovered Payments by buckets (Principal, interest, fee) were provided.

Data Enrichment

- Features like id, member_id, grade, sub_grade, employee_title, desc, title, zip, mths_since_last_record, pub_rec, revol_bal and values that has NA (payment plan, chargeoff_within_12_mths) have been mutated
- Features like mths_since_last_delinq, delinq_2yrs missing values have been filled with zero in the code.
- Loan accounts that are currently in active (Loan status ='Current') are excluded from analysis.

Analysis & Outcomes

Univariate, Bivariate Analysis

Binning

Derived Variables



Feature Analysis

Univariate & Bivariate Analysis

- Below are drivers to classify
- Recoveries - The due amount to be recovered from customer.
- delinq_2yrs - The frequency with which the borrower is delinquent in last 2 years, higher the number, higher the chances of default.
- annual_inc - The annual income of the borrower
- mths_since_last_delinq - The number of months after which the borrower did not repay the loan amount.
- DTI - Debt to income ratio. A value of lower number reflects the good ability to repay the loan.

Data Binning

- Below features are binned based on various ranges
- recoveries (RecoveryCategory (High, Medium, Low))
- delinq_2yrs (DelqFreqCategory (High, Medium, Low))
- annual_inc (IncomeCategory (High, Medium, Low))
- mths_since_last_delinq (DPDCategory (High, Medium, Low))

Feature Analysis

Derived Variables, Functions

- Derived RiskCategory variable based on below
- recoveries - veryhighBal (>10001), HighBal (3501-10000), MedBal (1001-3500), LowBal (0-1000)
- delinq_2yrs - DelqShortFrqncy (1-2), DelqMediumFrqncy (3-5), DelqLongFrqncy (6-8)
- mths_since_last_delinq -> DPDSHORT (1-3), DPDMedium (4-6), DPDLONG (7-12), DPDVeryLong (>12)
- annual_inc -> veryhighSal (>150001), HighSal (60001-150000), Med Sal (18000-60000), LowSal (9500-18000)

Function to derive RiskCategory variable

- The function
EvalRisk(DPDCategory,IncomeCategory,De
lqFreqCategory,RecoveryCategory) in
python will evaluate the Risk category of
each borrower (Risk types are High,
Medium , Low)

Reccomendations

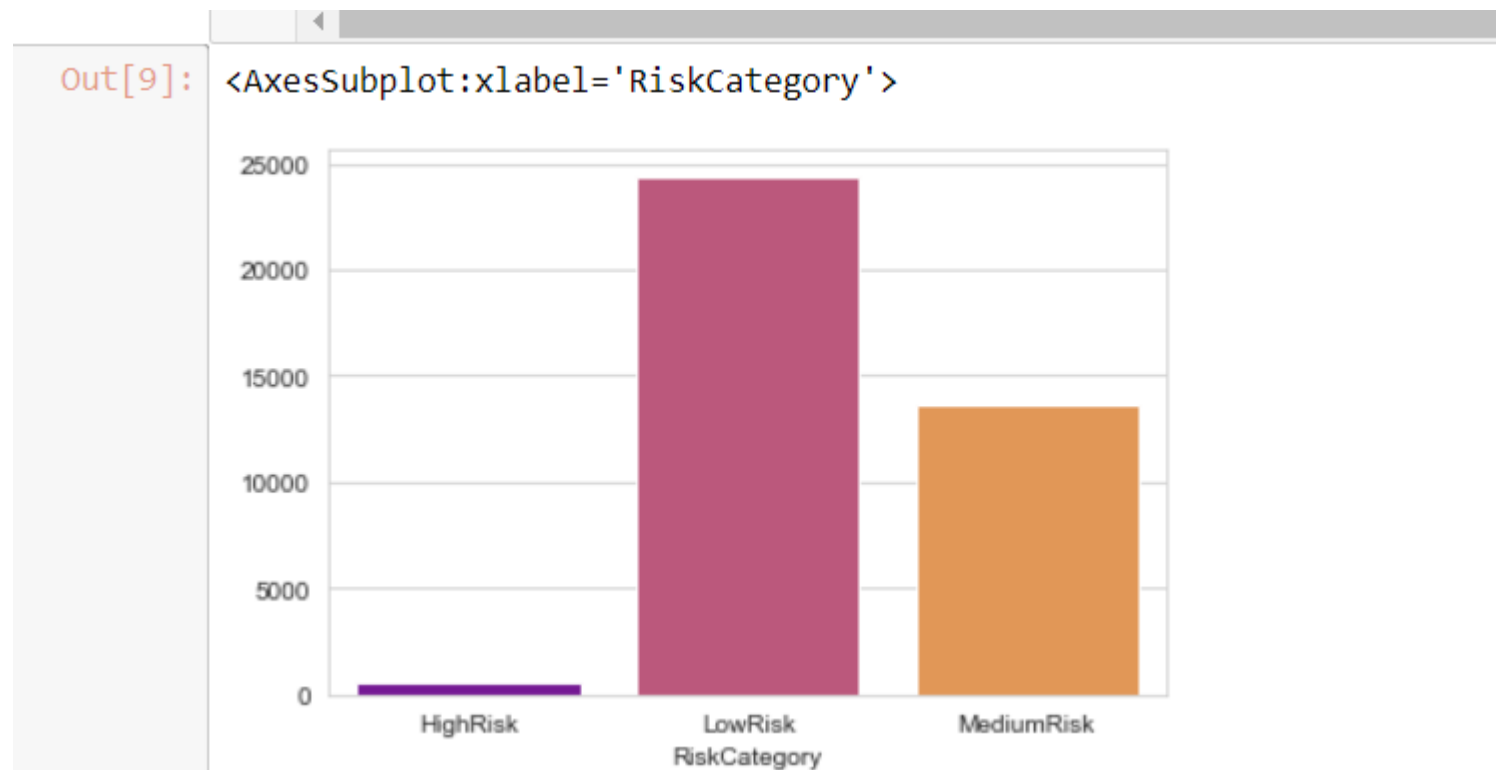


Recommendations

Using Risk Category (Derived Variable), Home Ownership, DTI

- A High Risk application with OWN Home ownership and a DTI (0-10), a loan with 80% funding amount can be sanctioned with higher interest rates for long duration. Rest of High Risk categories can be rejected.
- A Medium Risk application with OWN Home ownership and a DTI (0-10), a loan with 90% funding amount can be sanctioned with higher interest rates for long duration
- A Medium Risk application with Lease/Rent Home ownership and DTI (10-30), a loan with 85% funding amount can be sanctioned with medium interest rates for medium/long duration
- A Low Risk application with OWN Home ownership and DTI (0-30), a loan with 100% funding amount can be sanctioned with medium interest rates for medium/long duration
- A Low Risk application with Lease/Rent Home ownership and DTI (0-30), a loan with 95% funding amount can be sanctioned with medium interest rates for medium/long duration

Data Visualization





Thank you

Sudheendra
sudgo9@gmail.com