

STRUCTURED PYRAMID ANALYSIS PLAN

Classification & Prediction of Loan Defaulters

Team 8:

Adarsh Lalchandani

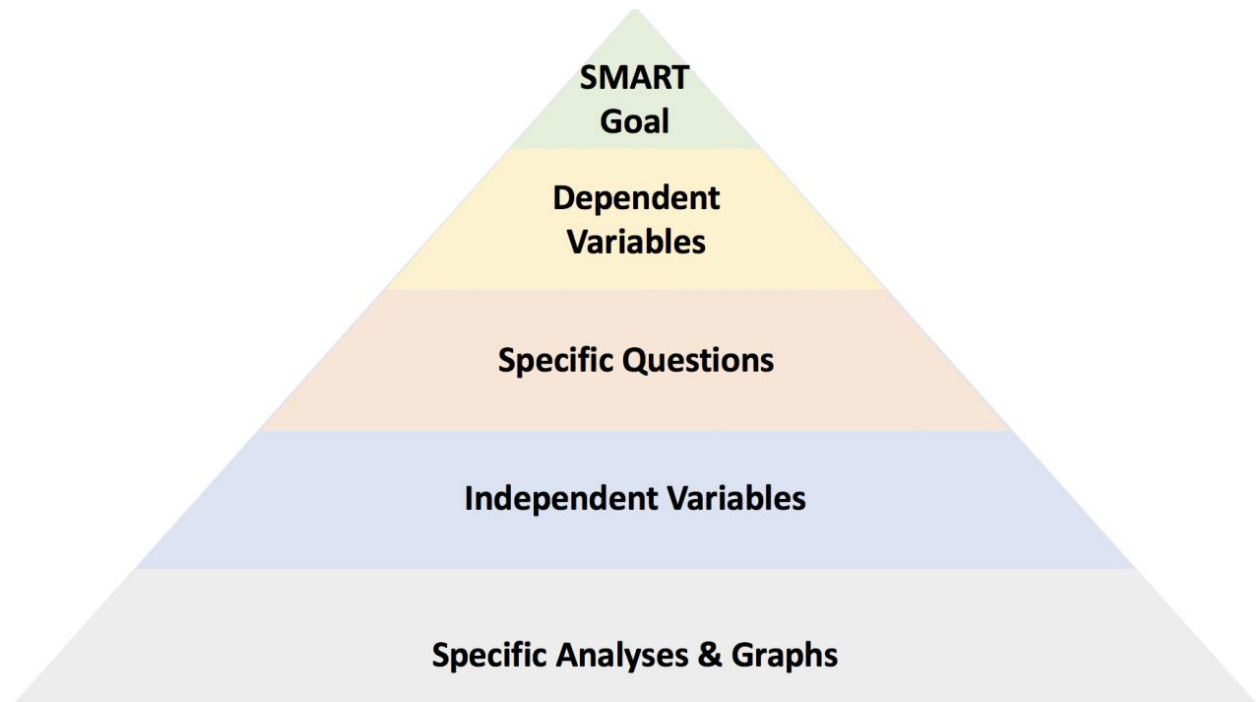
Karthik Grandhi

Richard Britto

GitHub - <https://github.com/AdarshRL2109>

GitHub - <https://github.com/karthii24>

GitHub - <https://github.com/richardbritto97>



GOAL:

To classify & predict whether the loan applicant is going to default or not (Boolean) based on the information we receive by the applicant at the time of application.

DEPENDENT VARIABLES:

default_flag taking values of 0 or 1 to determine if the borrower is going to default

SPECIFIC QUESTIONS:

What is the effect of loan amount to the number of loan defaults?

Who will default a loan based on the information they provide?

What is the minimum information needed to identify if the applicant will default?

INDEPENDENT VARIABLES:

loan_status, int_rate, loan_amt, Annual_inc, purpose, open_acc, No_of_Ac, emp_length, dti, revol_util, total_pymt, total_rec_late_fee, inq_last_6mths, delinq_2yrs

SPECIFIC ANALYSIS & GRAPHS:

- Univariate Analysis of specific variables vs number of defaults
- Bivariate Analysis on continuous variables vs default percentages
- Bivariate Analysis on categorical variables vs default percentages
- Multivariate Analysis to create the co-relation plot and find significant of each feature

HYPOTHESIS:

There exists some co-relation between the information we get from a user and the probability of that person to be defaulting with an outcome being either a defaulter or not a defaulter. The strength of co-relation will vary for each distinct variable and we will determine the effect of each variable over the outcome of this prediction.

DATA DICTIONARY:

Variable	Description
member_id	Unique identifier
loan_status	Current status of the loan
int_rate	Interest Rate on the loan
default_flag	A Boolean value where 0 means no default & 1 means default
inq_last_6mths	The number of inquiries in past 6 months (excluding auto and mortgage inquiries)
Annual_inc	The self-reported annual income provided by the borrower during registration.
Purpose	A category provided by the borrower for the loan request.
open_acc	The number of open credit lines in the borrower's credit file.
No_of_Ac	Categorical value for open_acc variable
emp_length	The job title supplied by the Borrower when applying for the loan.
loan_amnt	The listed amount of the loan applied for by the borrower. If at some point in time, the credit department reduces the loan amount, then it will be reflected in this value.
dti	A ratio calculated using the borrower's total monthly debt payments on the total debt obligations, excluding mortgage and the requested LC loan, divided by the borrower's self-reported monthly income.
revol_util	Revolving line utilization rate, or the amount of credit the borrower is using relative to all available revolving credit.
total_pymnt	Payments received to date for total amount funded
total_rec_late_fee	Late fees received to date
delinq_2yrs	The past-due amount owed for the accounts on which the borrower is now delinquent.