

LENDING CLUB CASE STUDY

Tanikella V S S Pavan Kumar

Hari Haran J

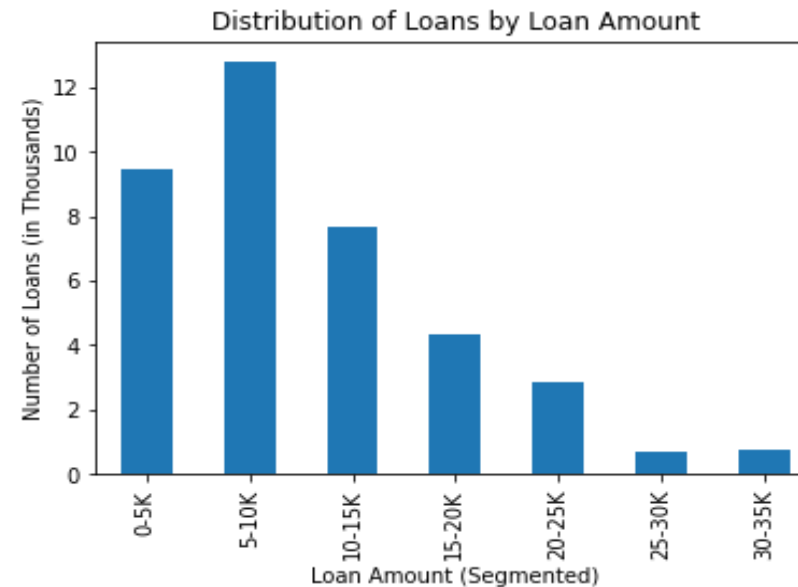
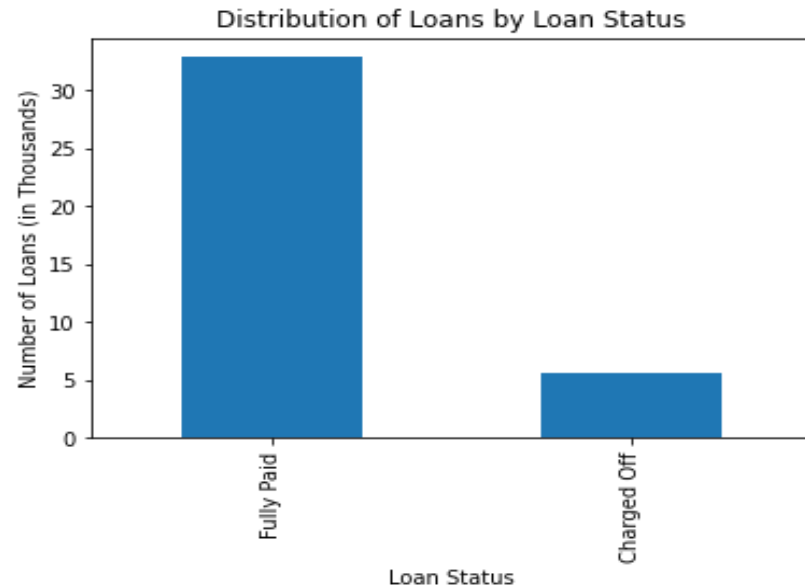
Problem Statement

- ▶ The client is consumer finance company which specializes in lending various types of loans to urban customers
- ▶ Approving loans to customers who are likely to default, will lead to financial loss to the company
- ▶ The company has provided historical data of applicants whose loans were approved and their respective consumer and loan attributes
- ▶ The aim of this analysis is understand driving factors which indicate whether a person is likely to default, using historical data

Approach

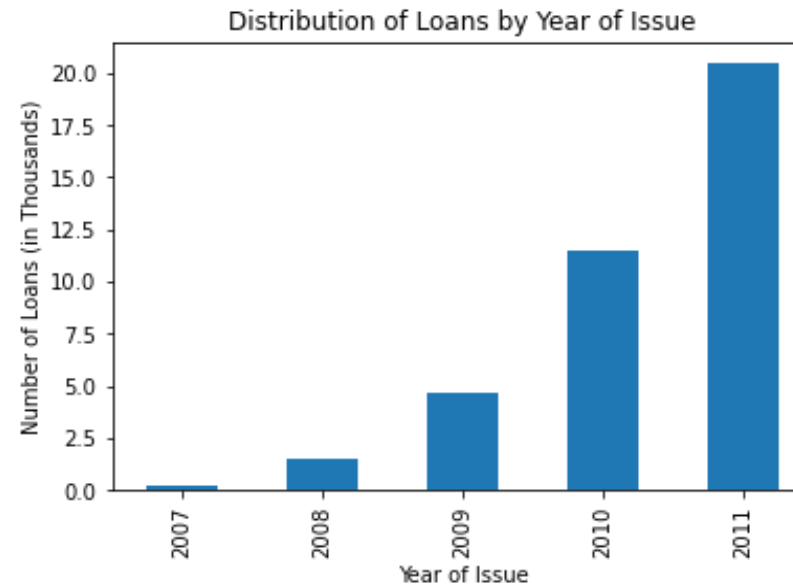
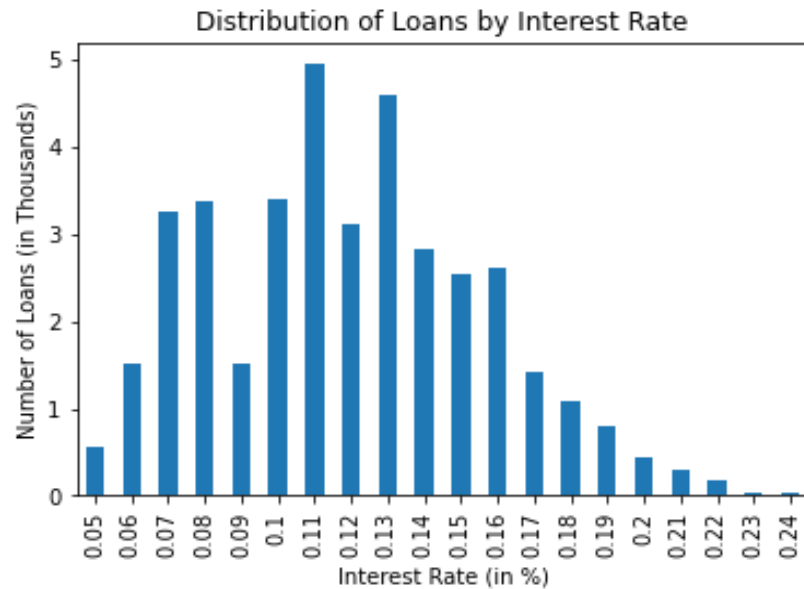
- ▶ Data Cleaning
 - ▶ Drop constant columns
 - ▶ Impute missing values
 - ▶ Remove irrelevant rows [Ex: loan status = 'current']
 - ▶ Ensure proper datatypes
- ▶ Create Derived Metrics
 - ▶ Credit History, Issue Year etc
- ▶ Univariate Analysis
 - ▶ Bar plot for categorical variables
 - ▶ Histograms for Numeric variables
- ▶ Segmented Univariate Analysis
- ▶ Bi Variate Analysis
 - ▶ Correlation Analysis
 - ▶ 100% bar plots against Loan status etc

Univariate Analysis 1/3



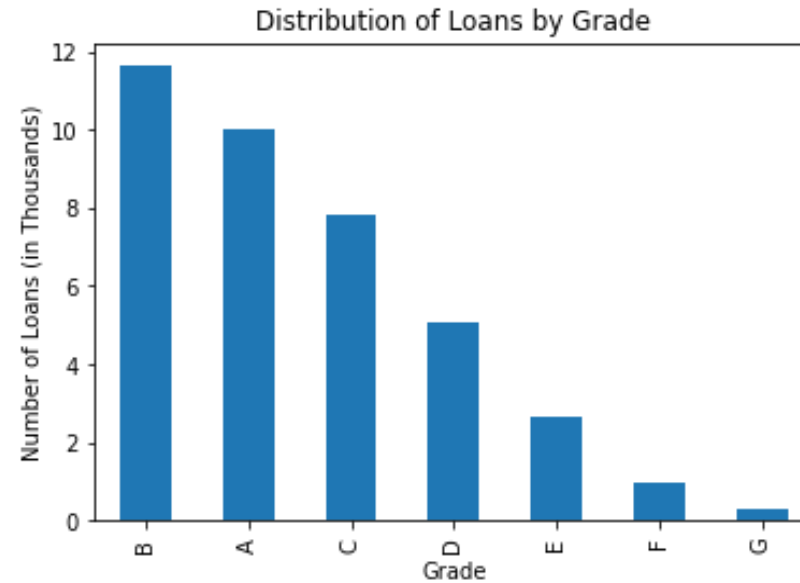
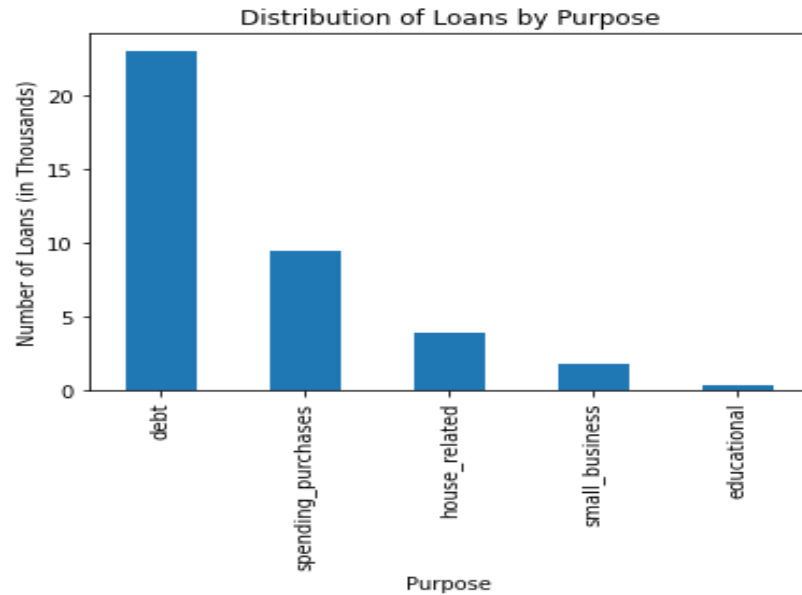
- ▶ Nearly 30K (~85%) of the loans are fully paid
- ▶ Most of the loan are for less than 15K Amount

Univariate Analysis 2/3



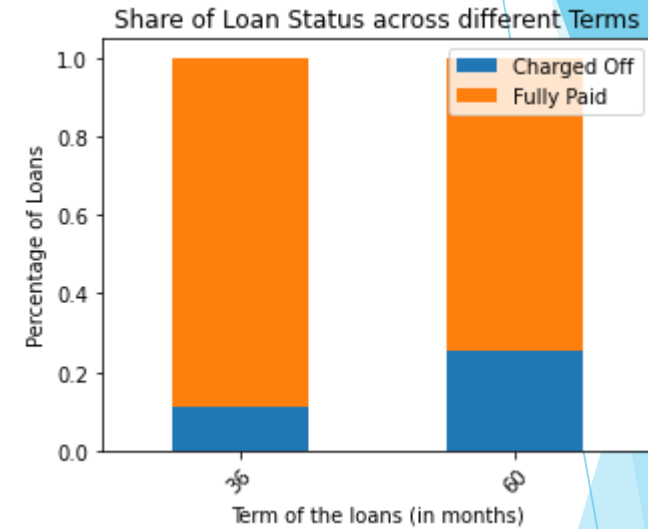
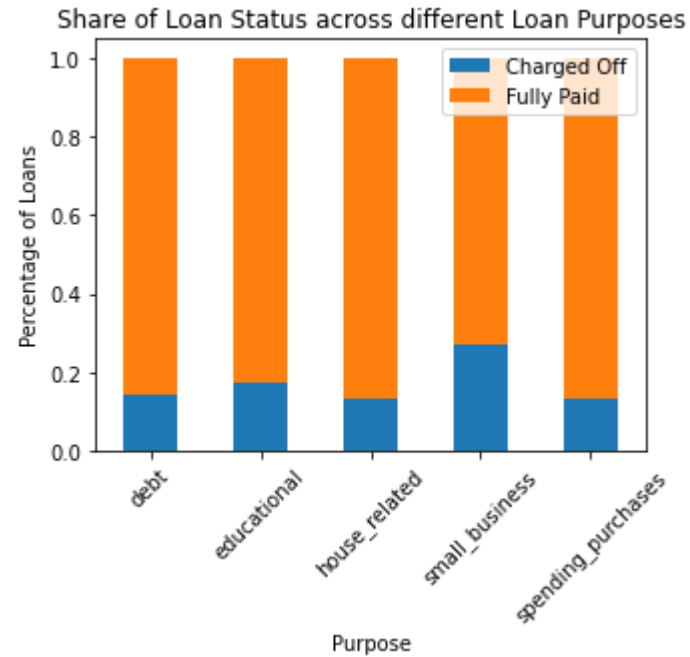
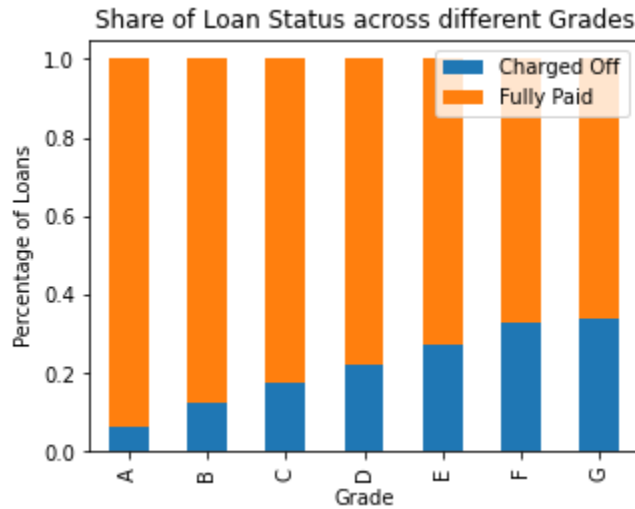
- ▶ >50% of loans are approved at rates between 10 and 15 %
- ▶ >60% of loans are issued in 2011

Univariate Analysis 3/3



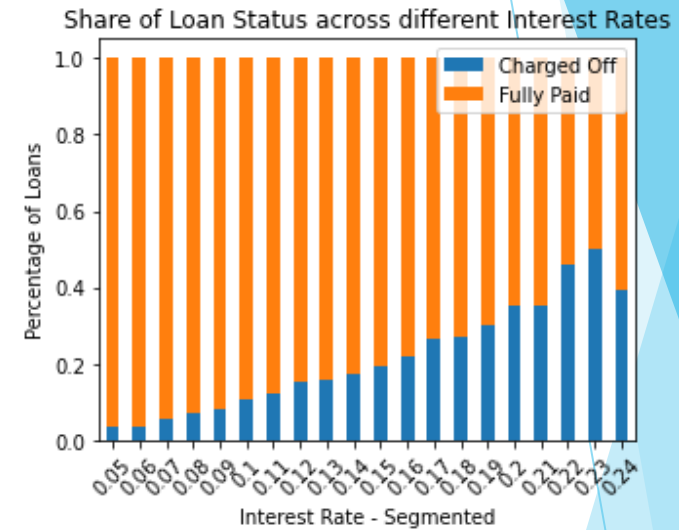
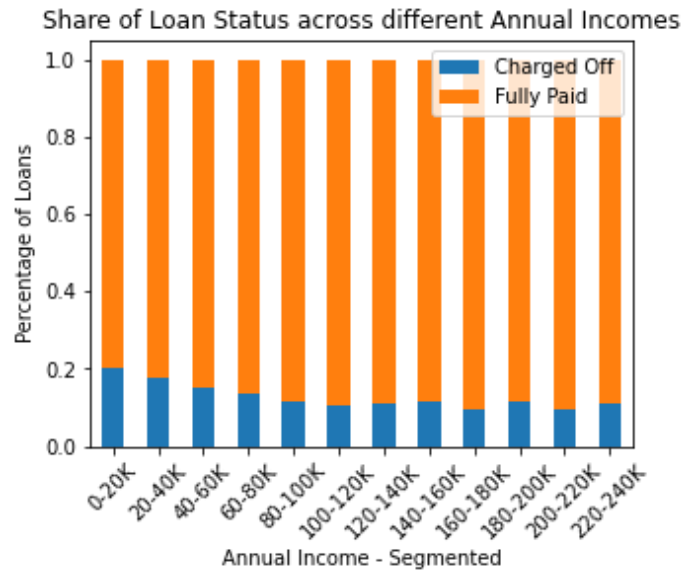
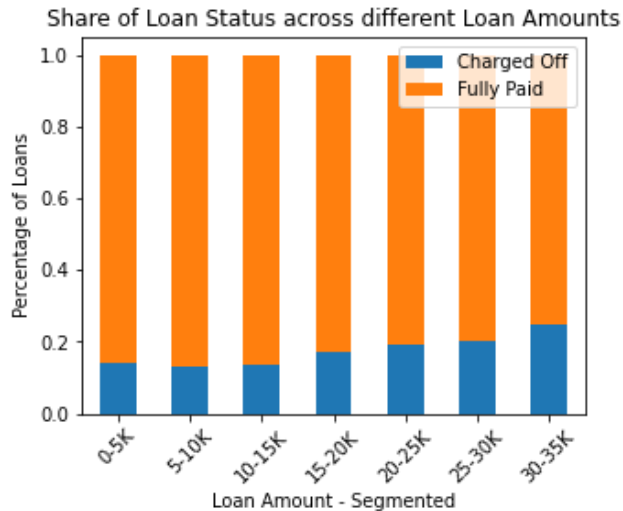
- ▶ >50% of loans are for debt consolidated purposes
- ▶ Most of the loans are issues for Grades A and B

Multi Variate Analysis - 1 / 4



- ▶ Loans with low quality grades are more likely to default
- ▶ Loans for small businesses are more likely to default
- ▶ Loans for longer terms are more likely to default

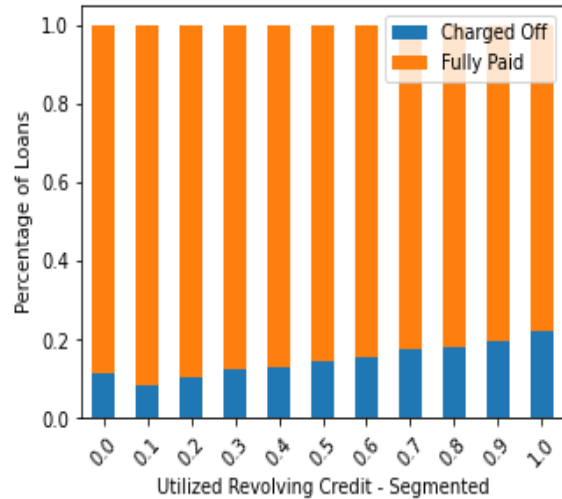
Multi Variate Analysis - 2/4



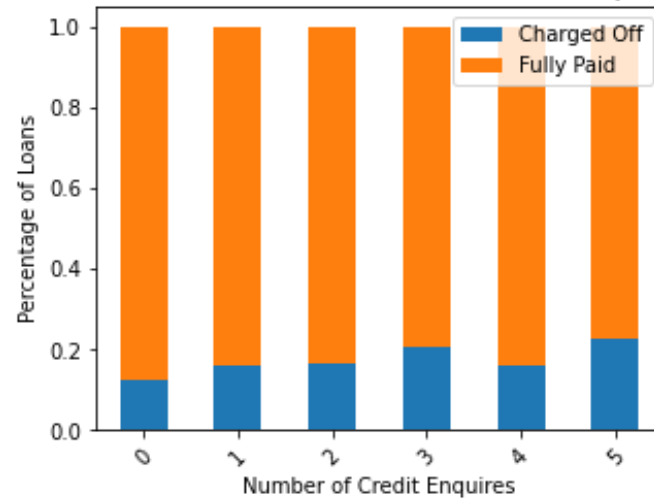
- ▶ Loans with higher loan amount or interest rate are more likely to default
- ▶ Loans with lower Annual Incomes are more likely to default

Multi Variate Analysis - 3/4

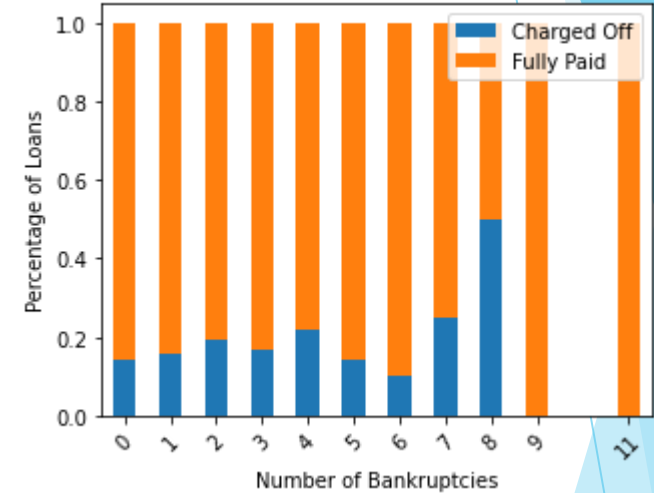
Share of Loan Status across different of Credit Utilization



Share of Loan Status across # of Credit Enquires

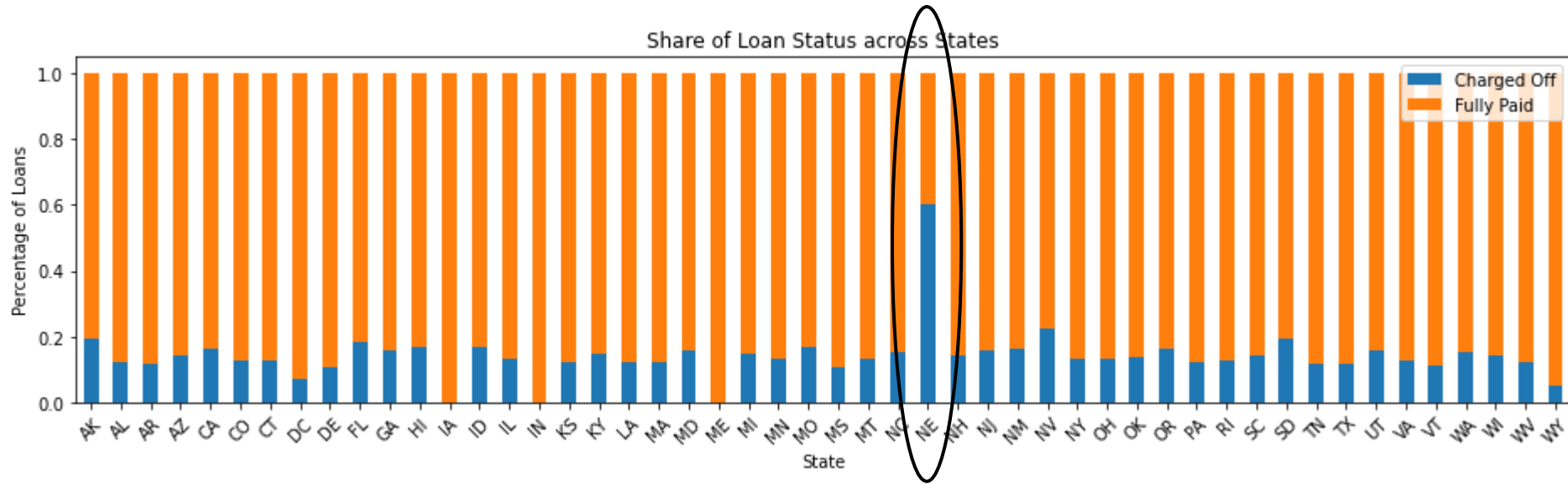


Share of Loan Status across # of Bankruptcies



- ▶ Loans with higher credit utilisation, credit enquires and number of bankruptcies are at higher risk to default

Multi Variate Analysis - 4/4



- ▶ Nevada in particular, seems to have a very high default rate, this need localised deep drive to understand the local trends