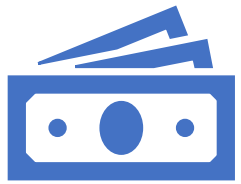
Lending Club Case Study

EDA presented by Ramanan Santhanagopalan (3/7/2022)



Problem Statement & Goal

Background: Customer is the largest online lending marketplace for personal loans, business loans and financing of medical procedures. Borrowers can easily access lower interest rate loans through a fast online interface.

Problem: Like most other lending companies, lending loans to ‘risky’ applicants is the largest source of financial loss (called credit loss). The credit loss is the amount of money lost by the lender when the borrower refuses to pay or runs away with the money owed.

Goal: Identify risky loan applicants early thereby cutting down the amount of credit loss. Conduct analysis to understand the driving factors behind loan default - variables which are strong indicators of default. Customer shall then utilize this knowledge for its ongoing portfolio and risk assessment.

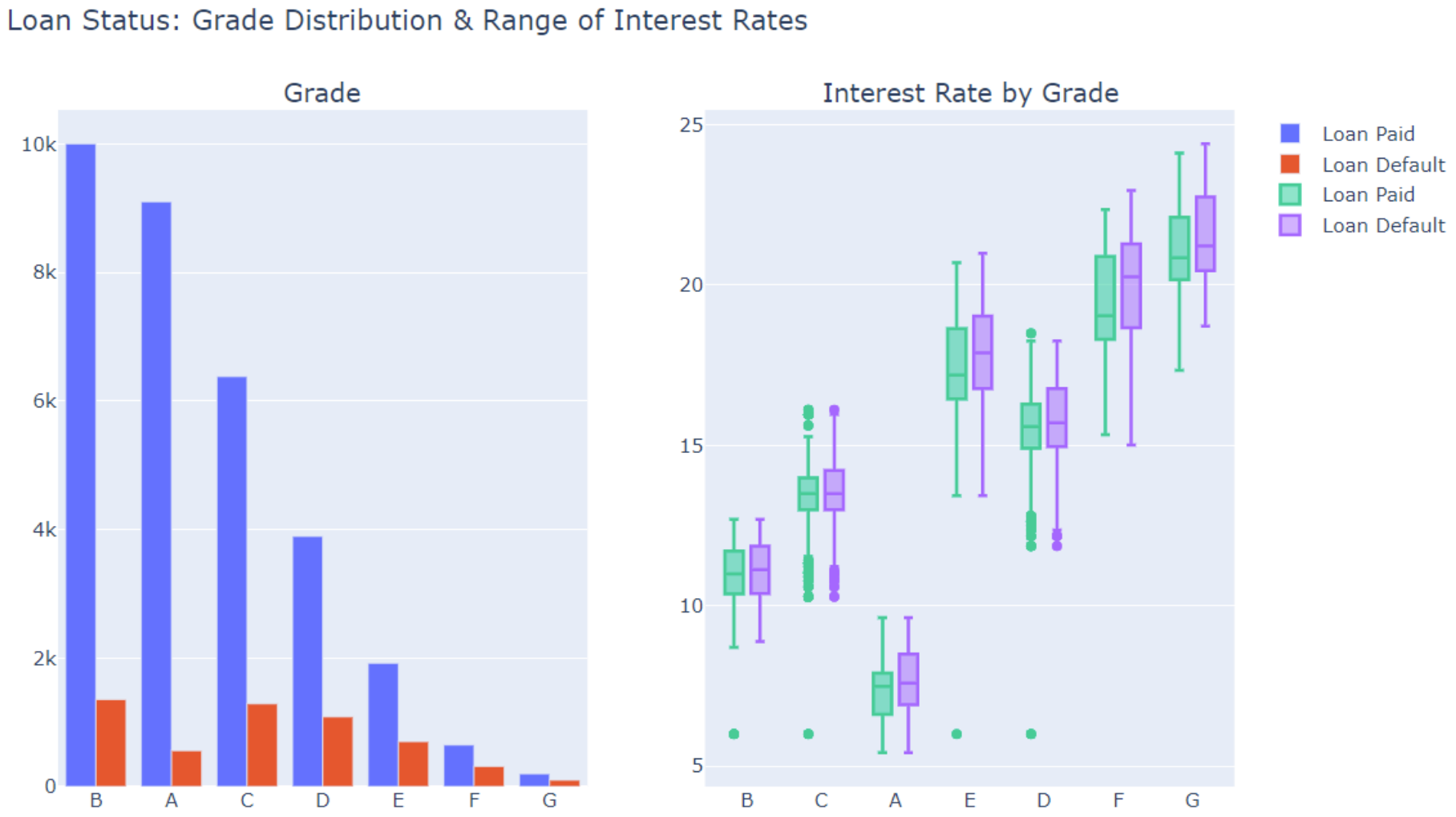
Exploratory Analysis

Type of analysis conducted	Column	Observations/Comments
Univariate	Loan status	Analyze the dataset provided by loan status, Charged Off 5627 Current 1140 Fully Paid 32950
Univariate	Home Ownership	Analyze the home ownership status of the members, RENT 18899 MORTGAGE 17659 OWN 3058 OTHER 98 NONE 3
Bivariate	Loan Status & Home Ownership	Study the relationship to see if there is a pattern between defaulters & home ownership
Univariate	Application type	All applicants are individual, and no joint applicants found
Univariate	Debt to Income ratio	Analyze the debt-to-income ratio, 25^{th} 8.2, 50^{th} 13.41, 75^{th} 18.59
Univariate	Interest Rate	Plot the density of interest rates in the dataset, 25^{th} 9.25, 50^{th} 11.86, 75^{th} 14.59
Bivariate	Length of Employment and Loan Status	Convert the length of employment to buckets, Fresher(up to 1year), Early Career (2-5 years), Mid Career (6-9 years) and Late Career (10+ years) to study the distribution against loan status
Univariate	Funded amount by the investor	25^{th} 5K, 50^{th} 9K, 75^{th} 14425 to see if there are outliers that needs to be cleansed. There are a few outliers above 25K and a few zeros. For rows with zeros, the imputation strategy will be to impute using Funded Amount column
Bivariate	Term and Loan status	Assess the relationship between Term of the loans and status. The terms are bucketed as short (36 months) & long (60 months) respectively for analysis.

Exploratory Analysis (Continued)

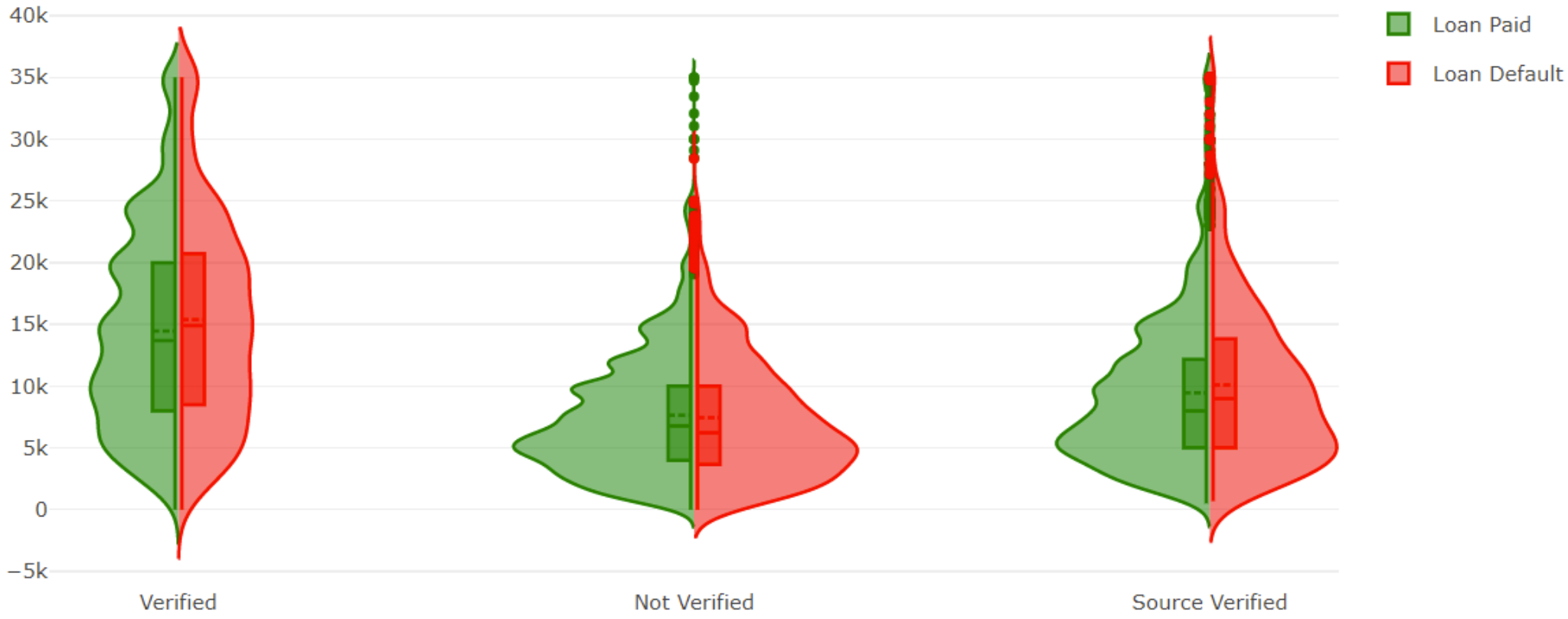
Type of analysis conducted	Column	Observations/Comments
Multivariate	Term, Interest Rate and Loan status	Analyze the bucketed terms with status of loans to see if defaulters are in a particular interest bracket
Multivariate	Interest Rate buckets, Loan Status & Loan Amount	Analyze the interest rate buckets (Very low – up to 5%, Low – 5 to 10%, Nominal – 10 to 15%, High – 15 to 20% and Very High > 20%) to see if defaulters fall in a certain bucket of interest and within a stipulated range of loan amount rendered
Multivariate	Interest Rate buckets, Employment length & Loan Status	Analyze the interest rate buckets (Very low – up to 5%, Low – 5 to 10%, Nominal – 10 to 15%, High – 15 to 20% and Very High > 20%) and employment buckets (Fresher, Early, Mid and Late career) to see if defaulters fall in a certain bucket
Bivariate	Income verification and loan status	Analyze the relationship between loan status and income verification to see if defaulters fall in a certain category of verification.
Bivariate	Grade and Loan Status	Analyze the relationship of grade of the loans against loan status to see if default is high in certain grades of the loan
Multivariate	Grade, Loan Amount and Loan Status	Analyze the relationship of grade of the loans against loan status to see if default is high in certain grades of the loan and in a stipulated range of loan amounts
Univariate	Annual Income	Analyze the annual income profiles and clean up the outliers using the imputation of IQR range in all outlier values
Bivariate	Subgrade and Loan Status	Analyze the relationship between sub grades and loan status. Sub grades B5, B3, B4, C3, D2, D3 and D4 appear to be most at risk
Bivariate	Issue Date and Loan Status	Analyze the relationship between loan defaults and issue date. The defaults have a growth trend in 2011 particularly the last 4 months of 2011 have seen a significant uptick In defaults

Loan Status and Interest Distribution by Grade



Analysis: Loan grades B, C, D and E are most at risk of default.

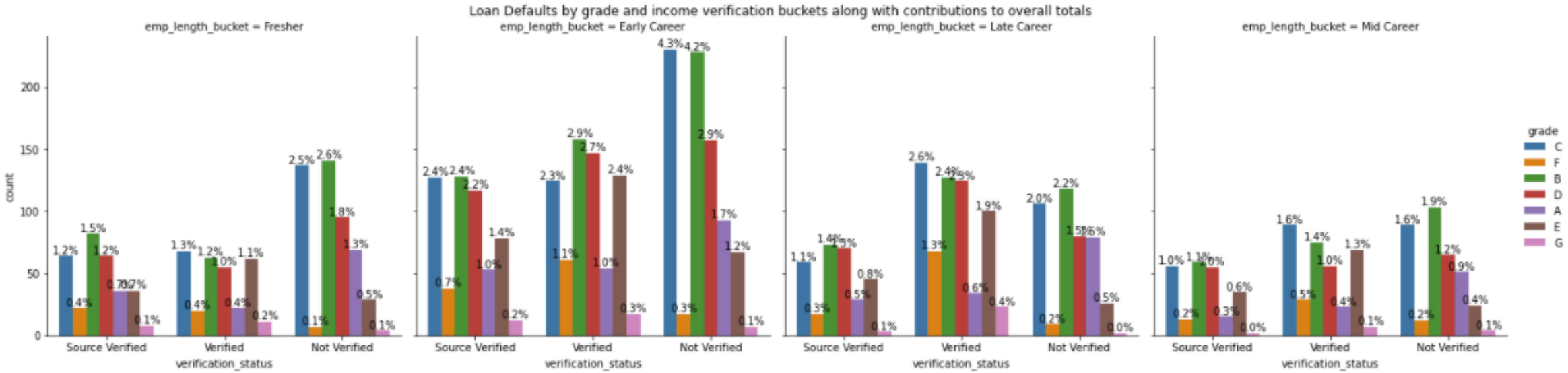
Income verification, Funded Amount by Investors and Loan Status



Analysis: Default risk of loans with income not verified is much higher for amounts over 6K and 3P income verified loans over 8K.

Proposal: Mandate an income verification process by the lender for loans over 6K

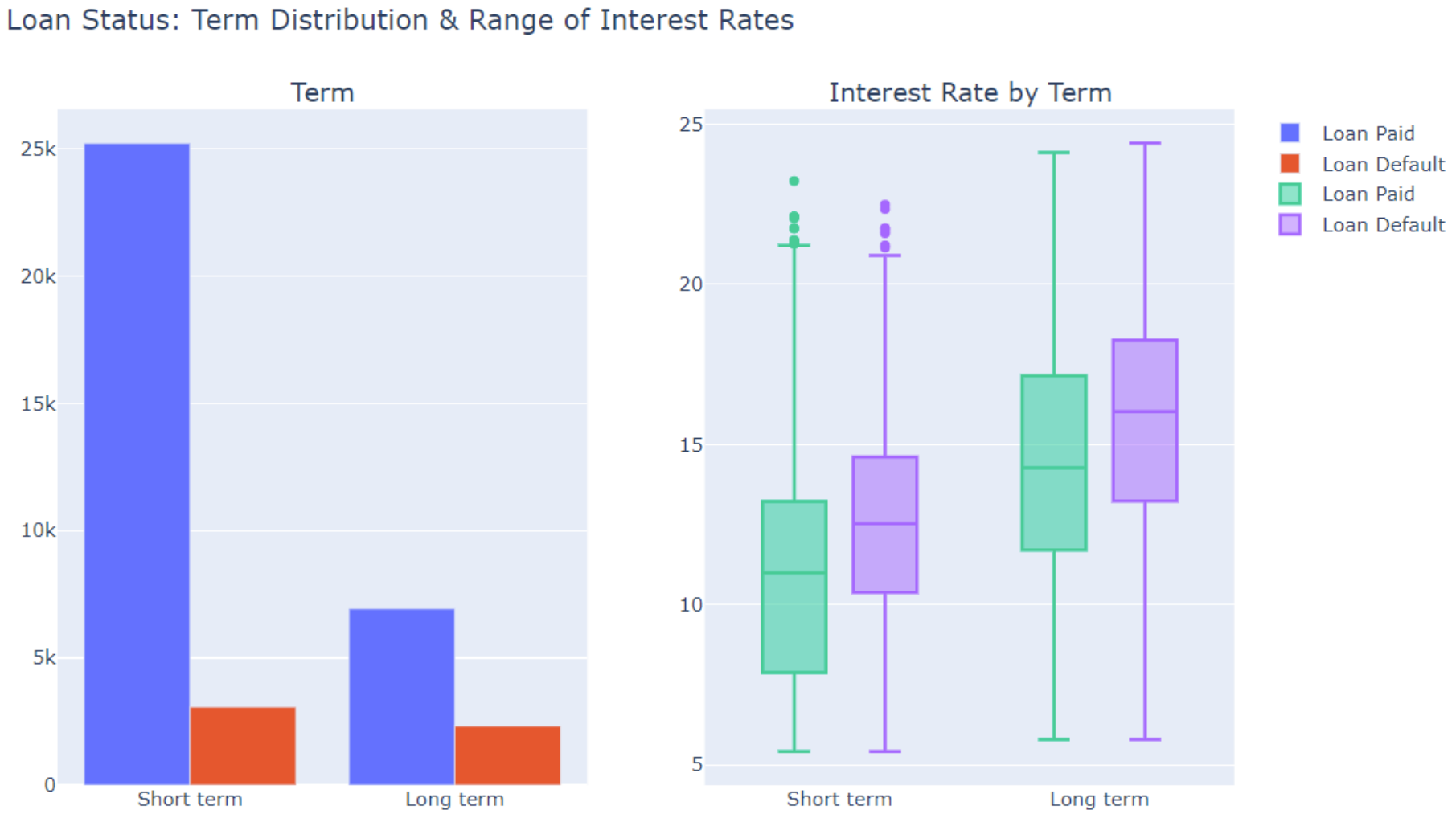
Loan Defaulters by employment bucket, income verification and grade



1. Most at risk: Early career (2 to 5 years) member with incomes not verified
2. Next most at risk: Freshers (<=1 year) with incomes not verified

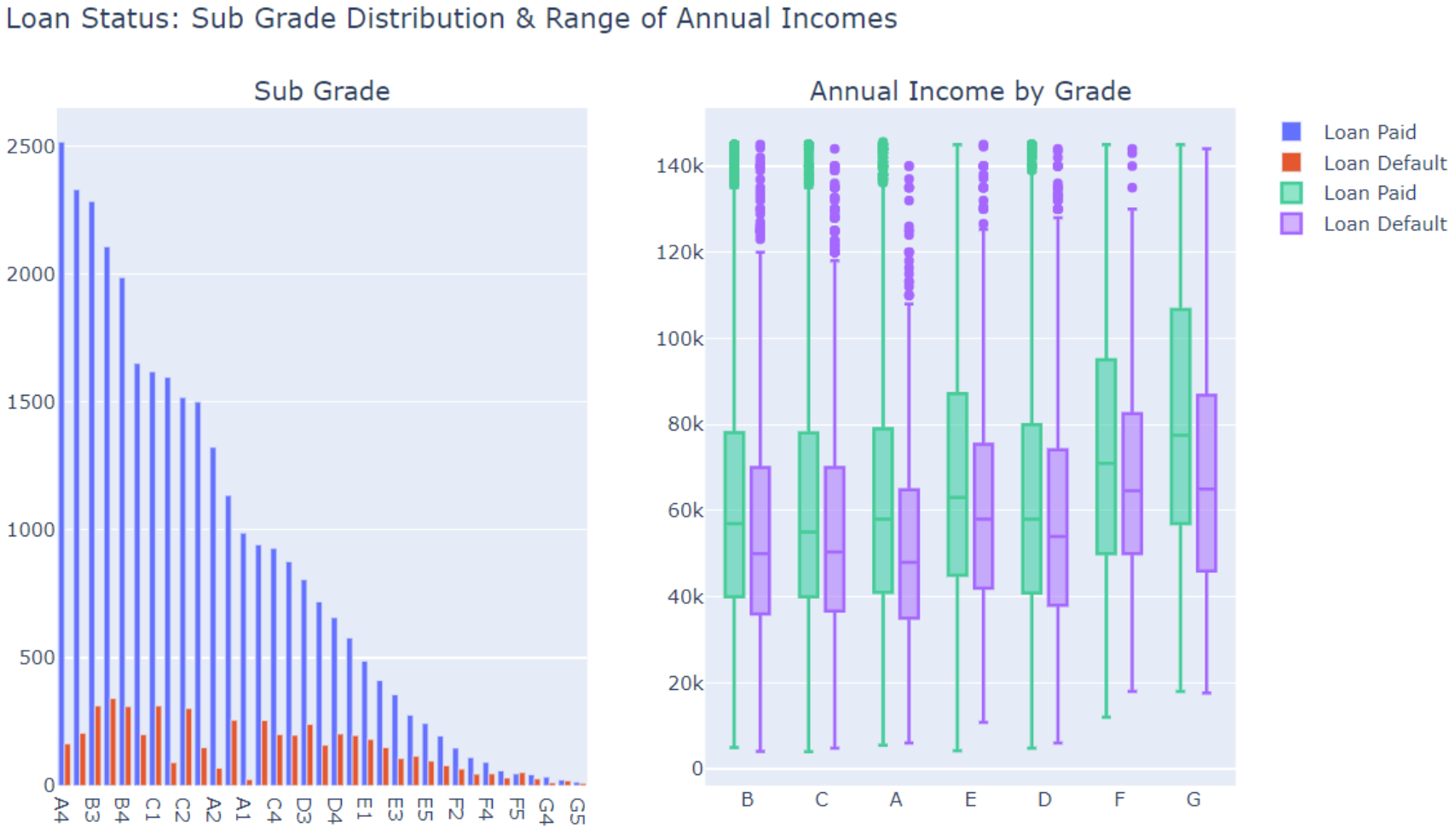
Proposal: Mandate income verification process for loan applicants up to 5 years of employment experience particularly for loan grades B, C and D

Term and Interest Rate impact on loan defaults



Analysis: Short term loans with higher interest rate (> 10%) are at risk of defaults.

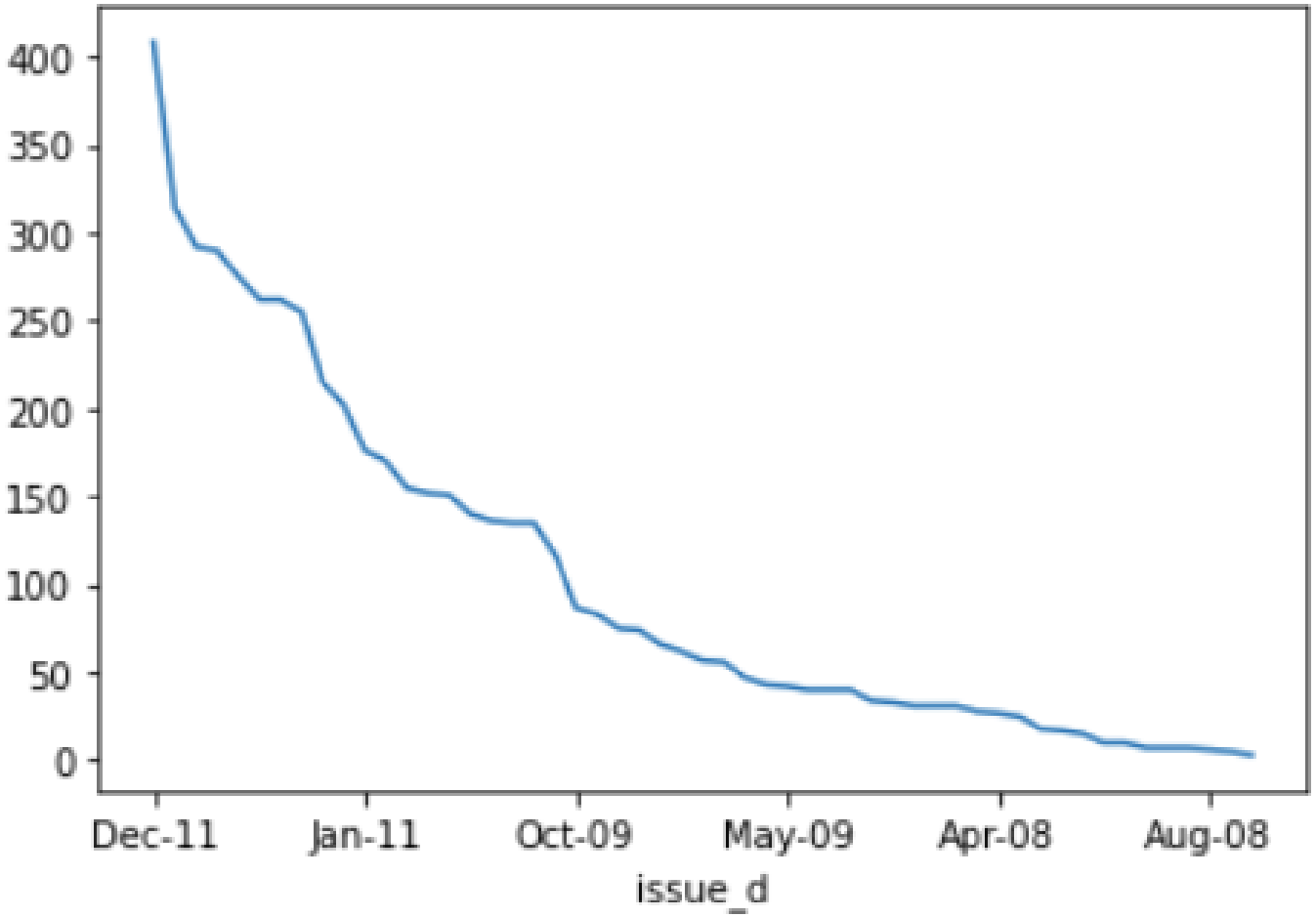
Sub Grades and Annual Income based analysis on Loan Defaults



Analysis: Sub grades B3, B4, B5, C3, D2, D3& D4 are most at risk of defaults and annual incomes below 40K will need extra scrutiny.

Proposal: Mandate an Annual Income > 40K as eligibility for all the loans

Defaults by issue date of the loan



Analysis: Loan defaults have been on an increasing trend with a significant spike in 2011. The back half and especially the last 2-3 months of 2011 have seen a significant spike in loan defaults.

Proposal: Avoid loan issuance during the Nov/Dec timeframe as there is a significant risk of default.