Lending Club Case Study

March 9, 2022

Subham Sahoo & Gaurav Kalra
Batch: C37

Problem Statement

The aim is to analyze information about past loan applicants of a Consumer Finance Company, using EDA to identify patterns which indicate if a person is likely to default.

The company can further utilise this knowledge of driving factors behind loan defaults for its portfolio and risk assessment.

Analysis Approach

The Past Loan Applicants Data of approximately 40k records with 111 variables was Analyzed using EDA Techniques:

1. Data Cleanup
2. Univariate and Segmented Univariate Analysis
3. Bivariate Analysis

Data Cleanup

Step-1

Removing Irrelevant Columns

1. Removed variables (Columns) with more than 85% values as NA, 0, Null or Empty Strings to avoid **Biased** view.
2. Column with Only One Unique Value removed
3. Also Removed variables that do not represent buyer or loan behavior

Step-2

Data Formatting and Derived Columns

1. Values were stripped off any leading & trailing spaces, any suffix such as %, years, any prefixes such as < etc.
2. Numeric columns were converted from string to numeric data types.
3. Creating Derived column. For instance Issue Month column was created from issue date.

Step-3

Removing irrelevant Rows

1. Removed Rows with status Current - as this data is about current loans which are running and is irrelevant until converted to Fully Paid or Charged off.
2. It was checked if there were any duplicates records representing same information

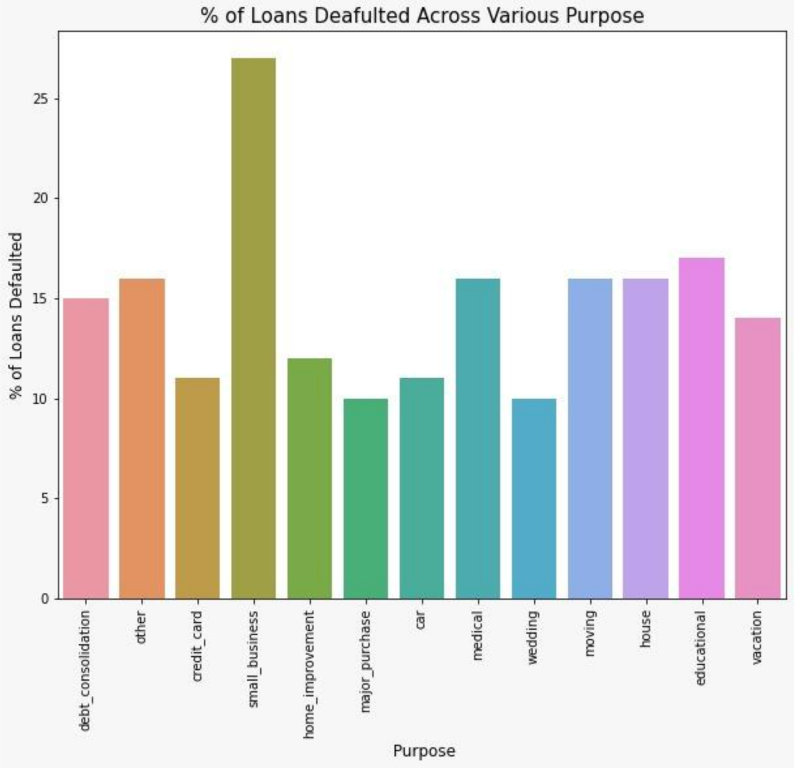
Data Understanding

After Applying the Data Cleanup techniques on given 111 variables for around 40k loans, following variables were selected to analyze the data

int_rate: Rate of Interest of Loan	**home_ownership**: Applicant’s Home Ownership details
purpose: Purpose for which Loan was taken	**annual_inc**: Self reported Annual Income of Applicant
grade & sub_grade: Grade assessment by LC	**open_acc**: Number of open credit lines of Applicant in credit file
addr_state: Applicant's Address State	**loan_status**: Status of Loan
emp_length: Length of Applicant's Employment	**revol_bal and revol_util**: Revolving Balance(the unpaid portion carries over to the next month)
dti: Debt to Income Ratio of the applicant	**funded_amnt_inv**: Final Amount approved by Investors

About Data Used for Analysis and Graphs

1. It is important to note that **absolute** higher number of defaults for a particular variable may give **misleading** impression about it's impact on default behavior. For instance, It appears that Loans of applicants from state CA has highest number of Defaults. But the state CA also contributes to maximum number of total loans. So arriving at conclusion that CA is biggest default loan generator may lead to loss of business to the Lending Company.
2. To overcome this issue - instead of looking at absolute numbers, The **percentage of Loans which are defaulted** in a category, was used to draw the graphs. With this approach, **actual insights** show up. Continuing with above example with this technique, we figure out that around 23+% loan applicants default from state "NV". Whereas only 16% of loans default from state "CA".
3. To remove **outliers**, a particular **segment** of the variable was considered Only if it had a **minimum** number loans in that segment. For instance, if there is only one person with with 1M+ annual salary and that person defaulted, it can't be concluded that 100% loan applicants with annual income above 1M+ will default.

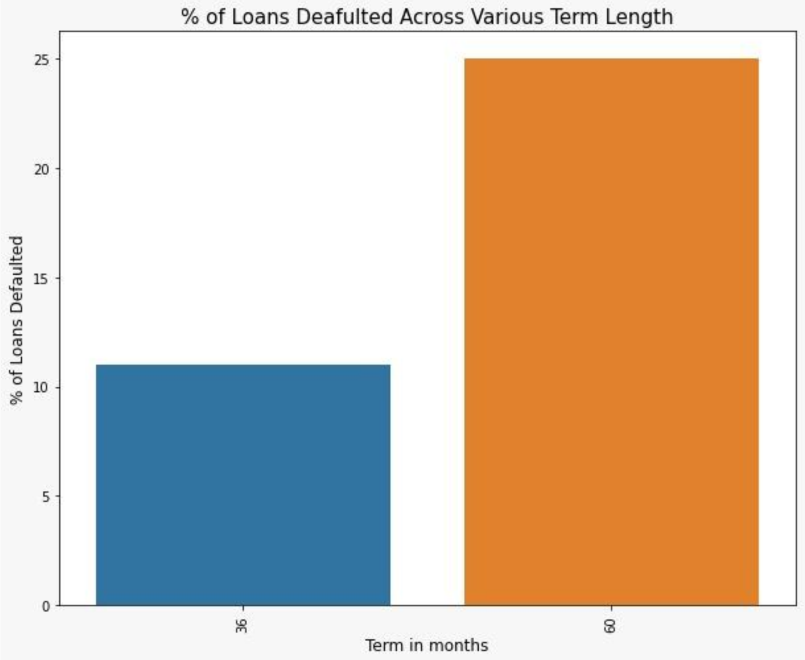


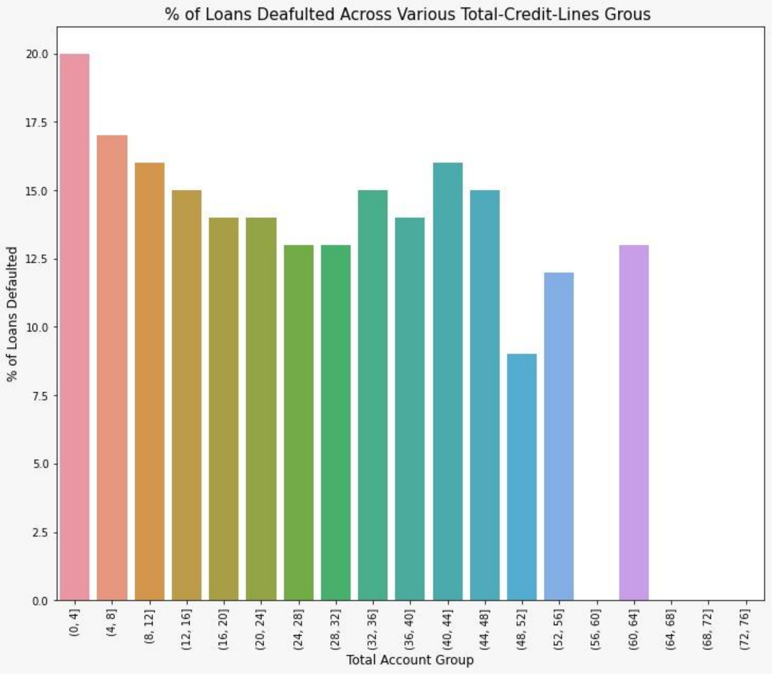
Purpose of Loan

Around 27% of Loans default which are taken for Small Business Purposes

Term - in Months

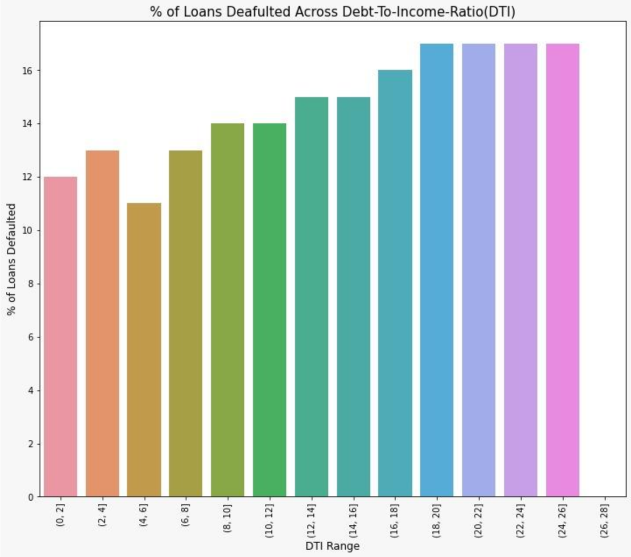
25% of the Loans taken for 60 month tend to default. This is significantly higher than 36 month Period loans which default only for around 12% of the times.





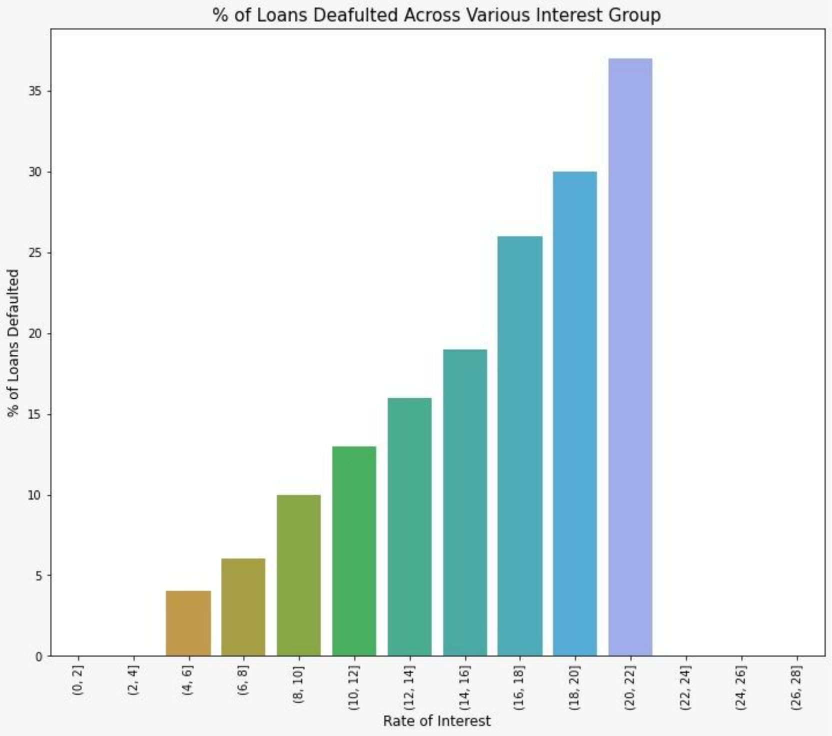
DTI - Debt to Income Ratio

DTI ratio above 10-12 points has a bit higher default rate compared to the lower DTI range. At higher ranges the default rate becomes quite steady. But certainly lower DTI rate has lower default rate.



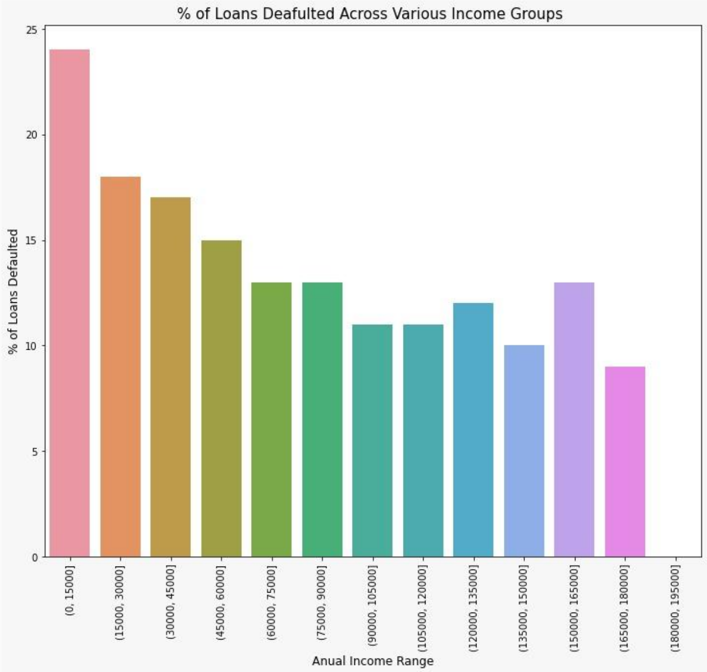
Total Account - Number of Open Credit Lines

With increase in open credit lines up to 28, the percentage of default applicants continue to decline. So lower the number of credit lines, higher the chances of defaults. Beyond 28 the impact is not steady.



Annual Income

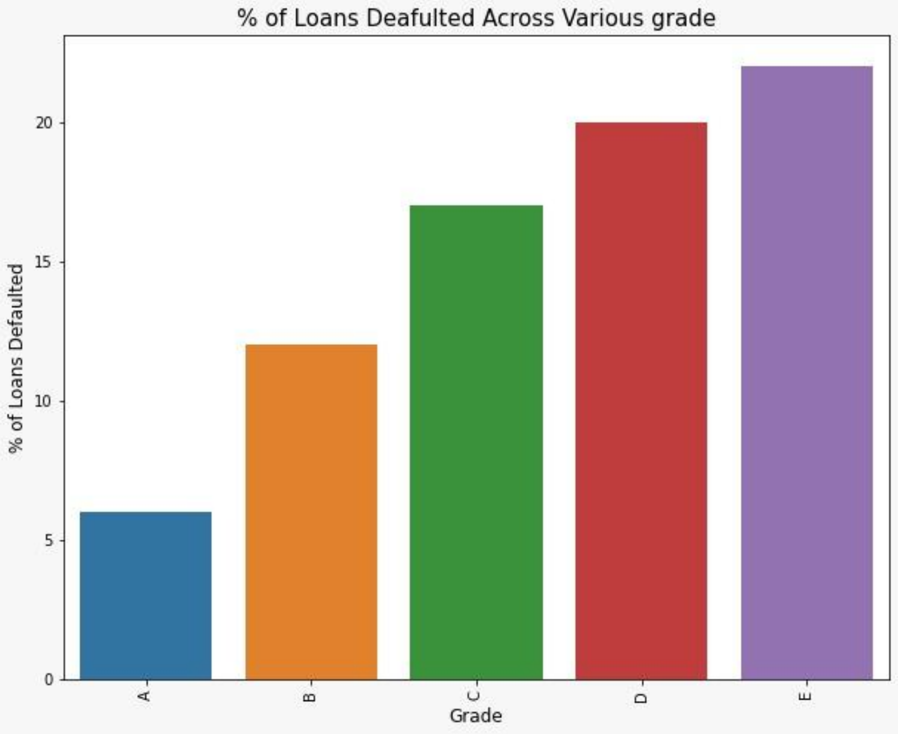
Applicants with lowest annual income tend to default most with 24% of such loans defaulting. The trend continues and after a point, the impact annual income on default loans is steady.



Interest Rate

Increase in Interest rate directly leads to increase in Default loans.

Interest Rate of 20% to 24% has maximum of around 37% defaulted loans

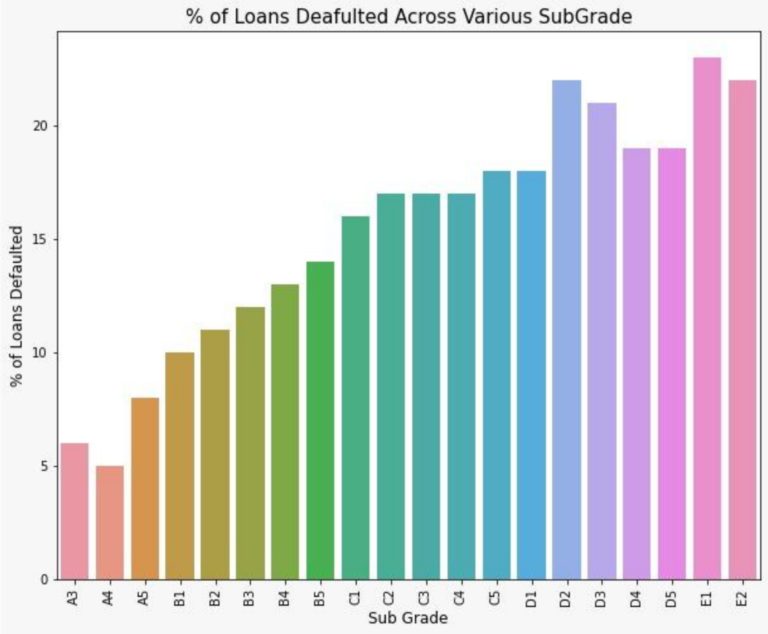


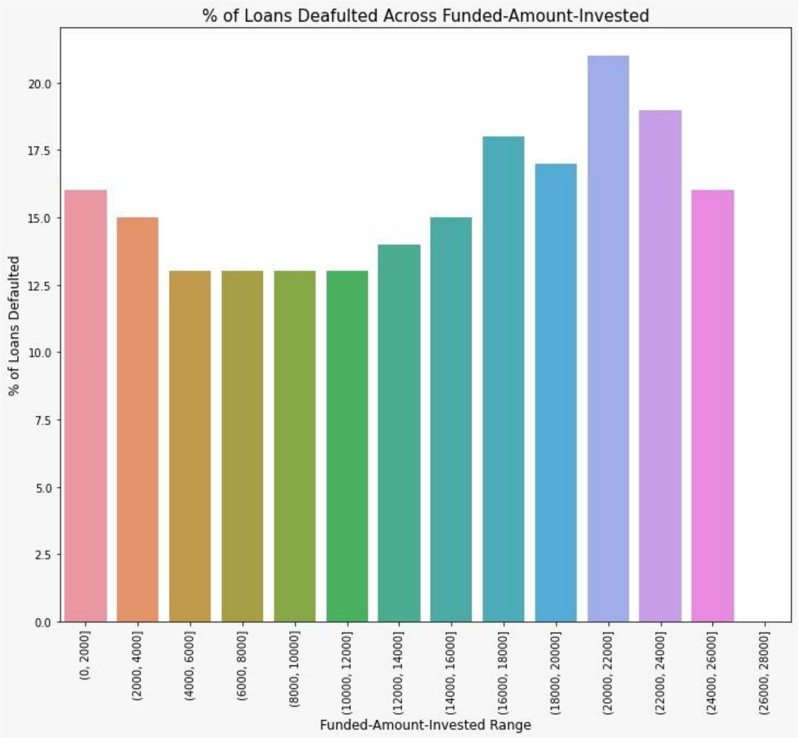
Grade

Category E Loans tend to default the most with around 33% loans defaulting. Category A Loans perform the best. Category seems to have direct impact on the % of default loans.

Subgrade

To further understand the Grade Behavior, Subgrades were analyzed and it leads to a very similar conclusion that Lower the grade higher the default rate. E4 being the biggest contributor for default loans. A3 being the one with lowest default rate.



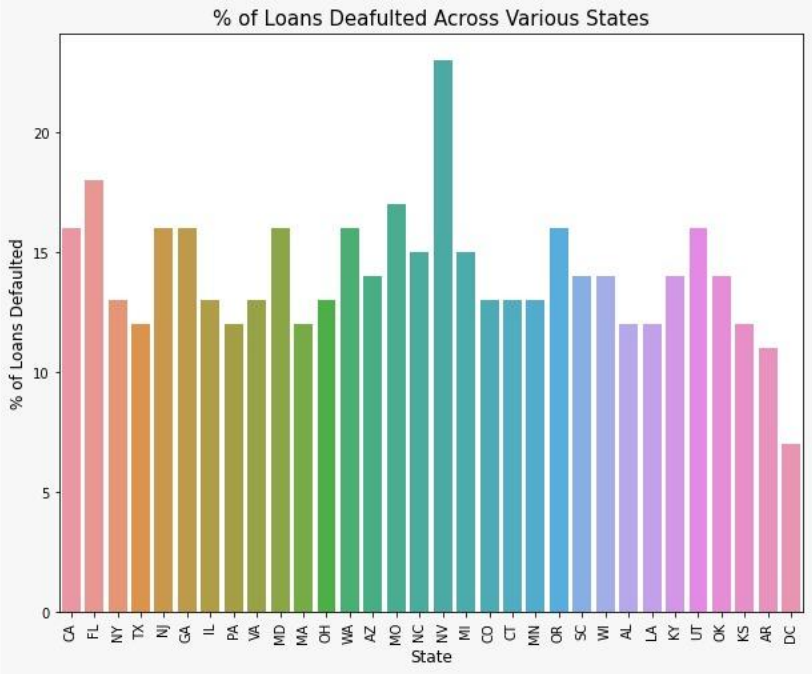


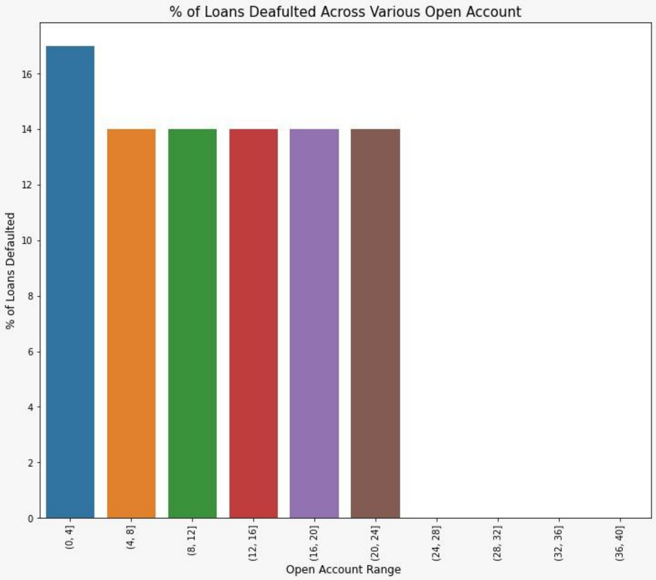
Funded Amount Committed by Investors

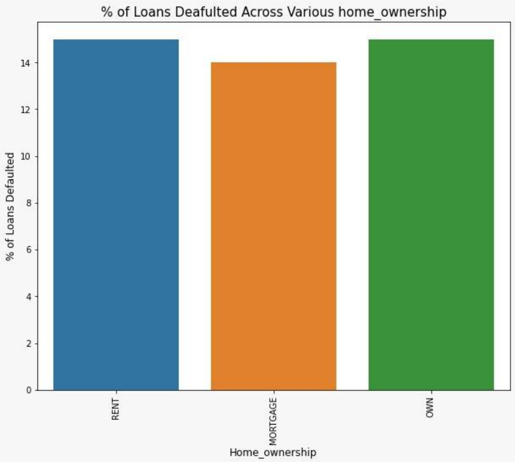
Loans with amount of 16k and above have higher Default rate. 20k to 22k Category has highest around 21% of loans defaulting.

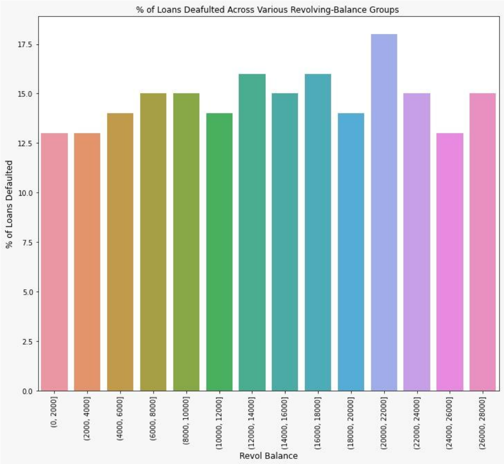
State

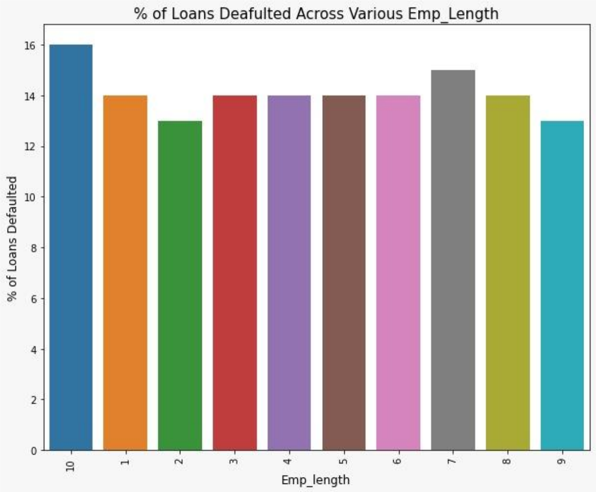
23% of Applicants from State “NV” default. On the other hand only 7% of Loans from “DC” default.











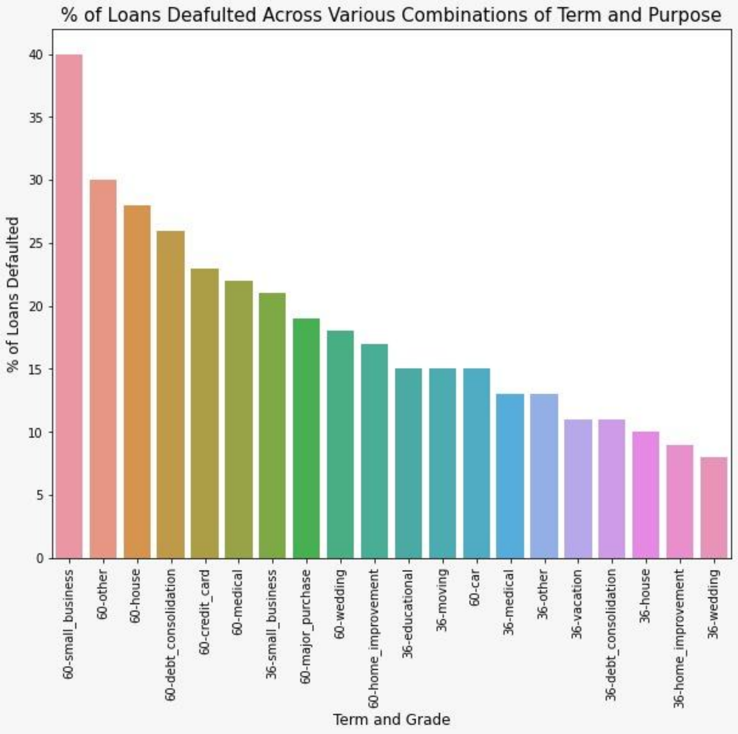
Other Variables

Variables such as Open Account Range, Home Ownership, Credit Revolving Balance, and Employment Length were also analyzed. The different segments of these variables show similar behavior for Default Ratio.

Impact of Combination of Two Variables on Default Rate

As a next step, we further analyzed impact of combinations of various parameters on Default Rate.

1. Grade and Annual Income
2. Term and Purpose
3. State and Annual Income
4. Grade and Interest Rate
5. Term and Annual income
6. Purpose and Annual Income

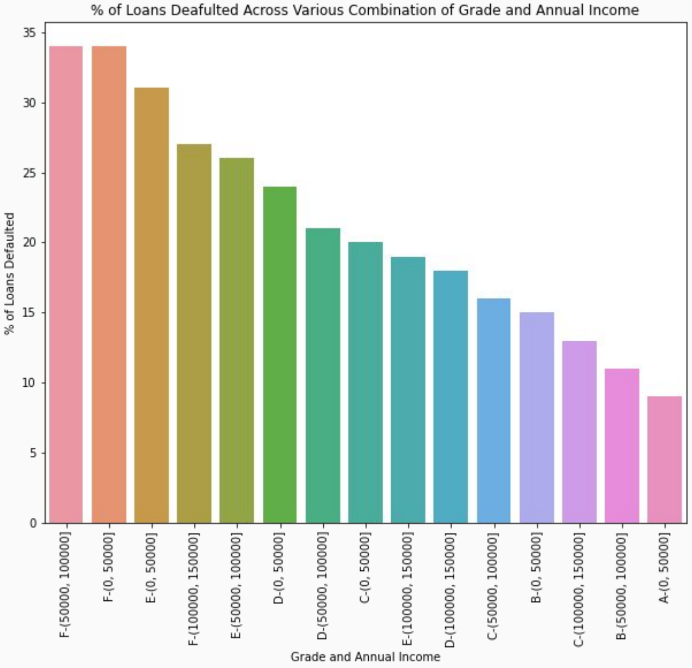


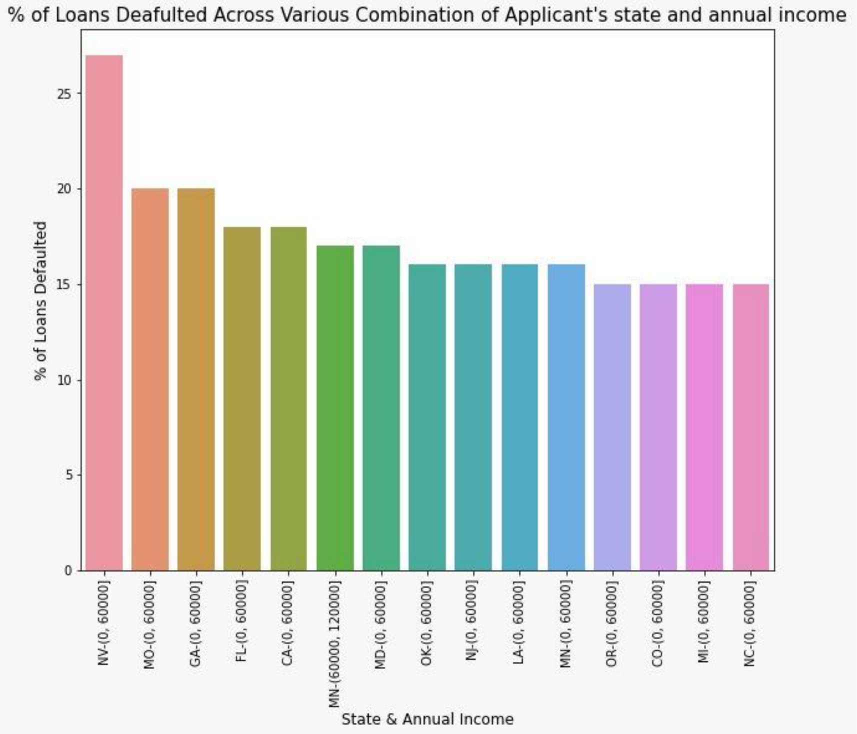
Term and Loan Purpose

60 Month Loan taken for Small Business has highest chances of defaulting. Short term Wedding Loans have lowest rate of default.

Grade and Annual Income

Grade A In Lower Income Groups have lowest % of default. Category F continues to lead in this category as well.



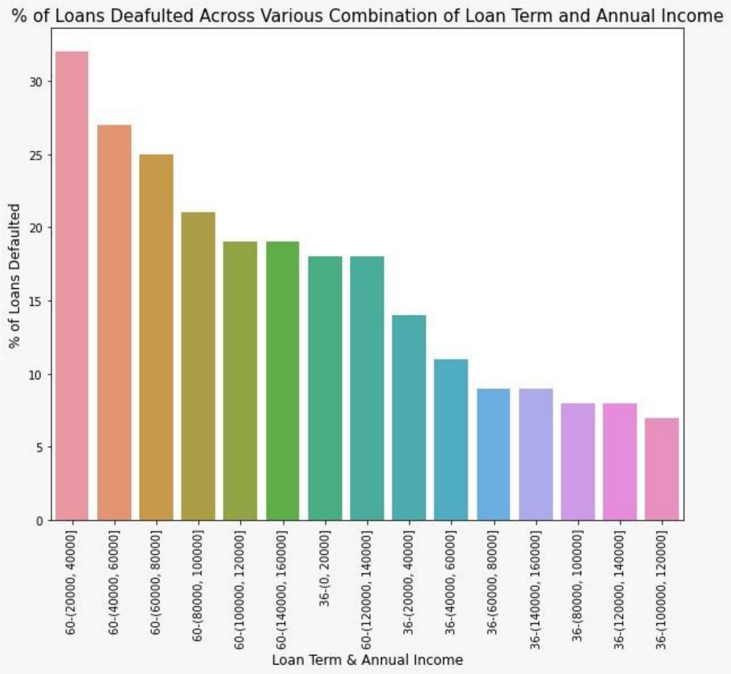


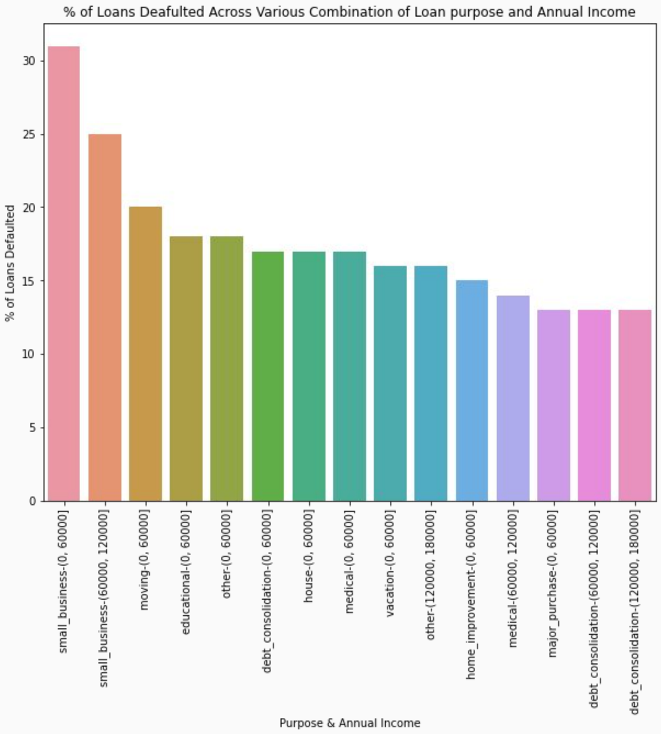
State and Annual Income

State NV Applicants with Under 60K annual income have the highest default rate.

Term and Annual Income

Long Term Loans default more across all income groups.



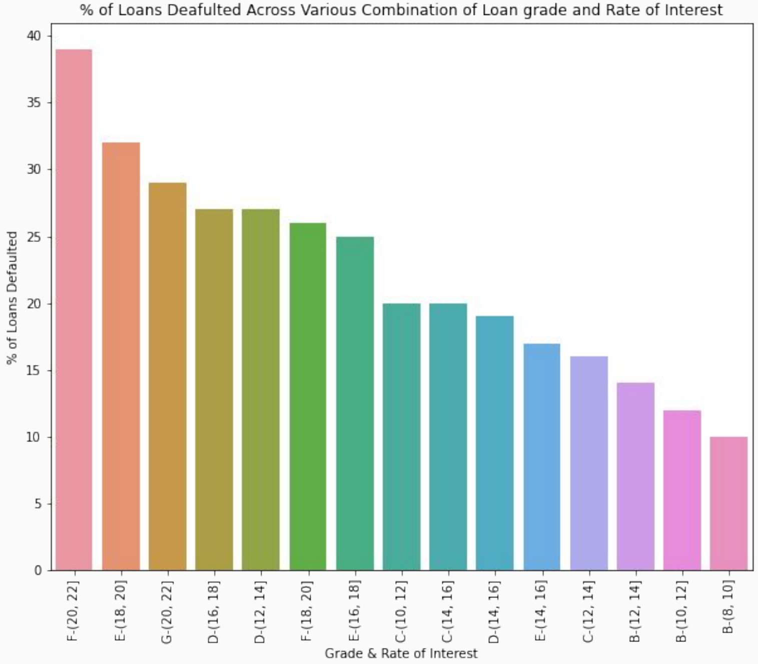


Purpose and Annual Income

Small Business Loans across all Income Groups have higher default rate.

Grade and Rate of Interest

Giving High Interest Loan to Category F Applicant has around 38% chances of getting default.



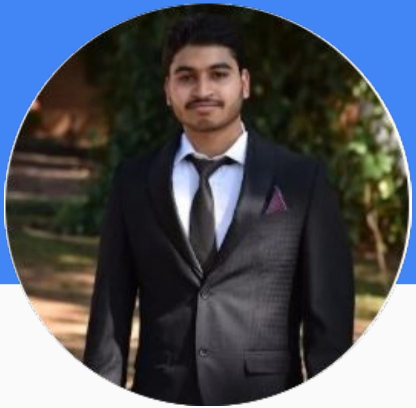
Recommendations

After carefully analyzing Various parameters that depict buyer behavior and Loan Conditions, following are few of the major driving factors for default loans.

1. Applicants with **Grade F**
2. Higher **Interest Rate (20% to 22%)** leads to significantly higher rate of defaults.
3. Loans taken for a purpose of **Small Business** have very high default rates.
4. Loans taken for a term for **60 month Period** have double the default rate than 36 month default rate.
5. Applicants with lower **Annual Income (0 to 15k)**
6. ~23% Loan Applicants from **State NV** default
7. Lower number of **Total Open Credit Lines** (0 to 7) have Higher Default Rate

Presence of More than one variables from list mentioned above, significantly increase the Default rate. For Instance, 40% of the Small business loans taken for a 60 month period default.

Thank You



Subham Sahoo

https://www.linkedin.com/in/subham-sahoo-650871157

Gaurav Kalra

https://www.linkedin.com/in/gaurav-kalra-a648697