Lending Club Case Study

Team

- Aswin J Shaji
- Lokesh Gaddam

Nov 8, 2023

Agenda

Topic
Case Study Business Understanding & Objective
Case Study Approach
Data Cleaning Summary
Univariate Analysis: Categorical Columns (4 slides)
Univariate Analysis: Continuous Columns (1 slide)
Bivariate Analysis: Between Continuous & Categorical columns(4 slides)
Correlation
EDA Summary
Additional analysis based on IV & WOE : Categorical Columns (2 slides)
Additional analysis based on IV & WOE : Numerical Columns (2 slides)
IV & WOE conclusion and summary

Aswin J Shaji
Lokesh Gaddam

Assignment : Lending Club Case Study

(Team : Aswin J Shaji & Lokesh Gaddam)
Group Facilitator : Name: **Aswin J Shaji**
Team Member Detail: Name: **Lokesh Gaddam**
Date : 8 Nov 2023

Business Understanding :

Lending Club is a financial institution that specializes in offering different loan types to urban clientele. Upon receiving a loan application, the company needs to evaluate the applicant's profile to decide on loan approval.
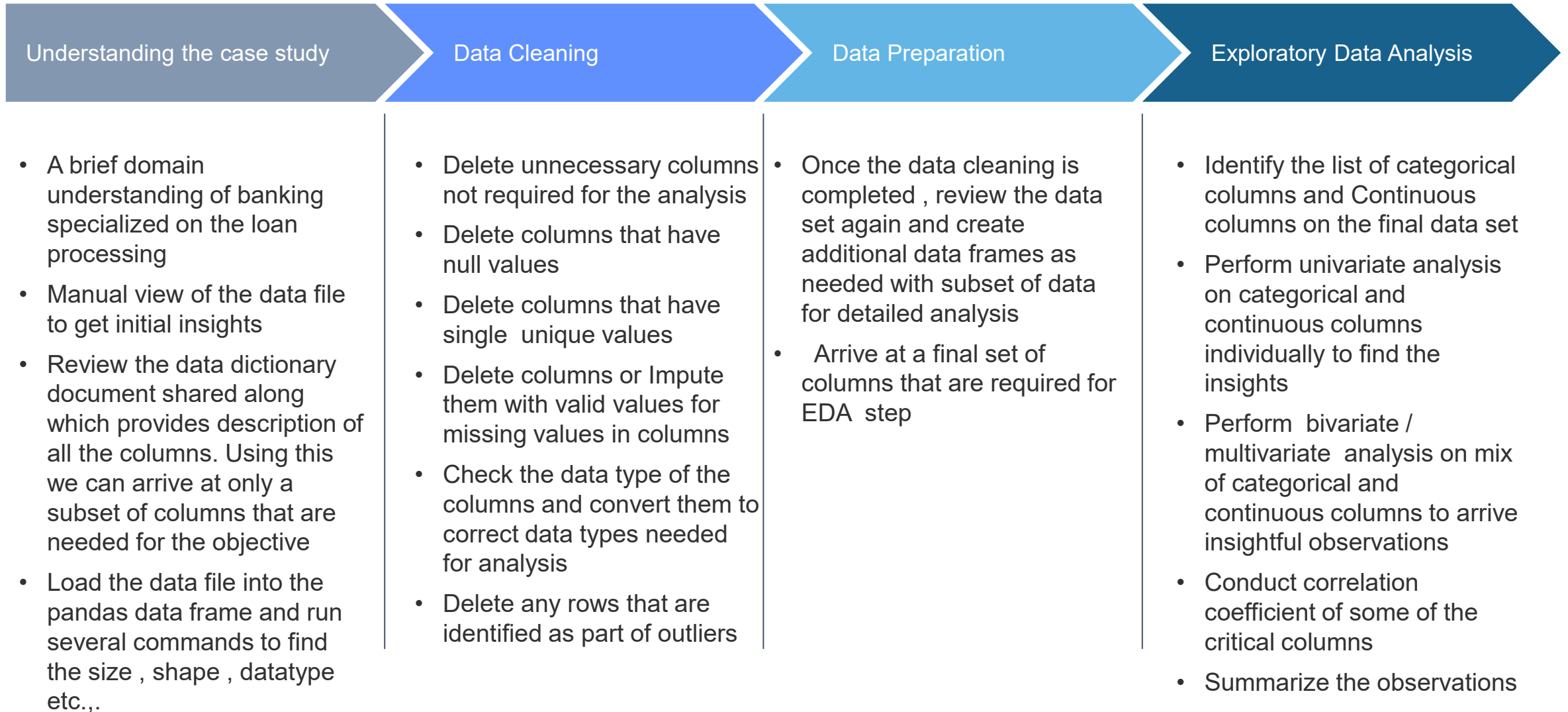
This decision involves managing two types of risks:

• Denying the loan to a potentially creditworthy applicant results in a loss of business for the company.

• Approving a loan for an applicant at risk of defaulting may lead to financial loss for the company.

Objective:

The company aims to comprehend the key factors influencing loan default (specifically, when the loan_status = 'Charged Off'), referring to the robust indicators that strongly correlate with defaults. This understanding can be employed by the company for portfolio management and assessing risks.

Case Study Approach

Understanding the case study	Data Cleaning	Data Preparation	Exploratory Data Analysis

Understanding the case study

- A brief domain understanding of banking specialized on the loan processing
- Manual view of the data file to get initial insights
- Review the data dictionary document shared along which provides description of all the columns. Using this we can arrive at only a subset of columns that are needed for the objective
- Load the data file into the pandas data frame and run several commands to find the size , shape , datatype etc.,.

Data Cleaning

- Delete unnecessary columns not required for the analysis
- Delete columns that have null values
- Delete columns that have single unique values
- Delete columns or Impute them with valid values for missing values in columns
- Check the data type of the columns and convert them to correct data types needed for analysis
- Delete any rows that are identified as part of outliers

Data Preparation

- Once the data cleaning is completed , review the data set again and create additional data frames as needed with subset of data for detailed analysis
- Arrive at a final set of columns that are required for EDA step

Exploratory Data Analysis

- Identify the list of categorical columns and Continuous columns on the final data set
- Perform univariate analysis on categorical and continuous columns individually to find the insights
- Perform bivariate / multivariate analysis on mix of categorical and continuous columns to arrive insightful observations
- Conduct correlation coefficient of some of the critical columns
- Summarize the observations

Aswin J Shaji

Lokesh Gaddam

Data Cleaning Summary

- Delete Columns
  - Identify the columns that have all NaN or max NULL values and drop those columns
  - Identify the columns that are not useful for the analysis and drop them (like IDs, Description, title etc.,.)
  - Identify the columns that have single unique value and drop the columns
- Duplicate data removal:
  - Identify the duplicate records / identical rows and remove the records
- Remove outliers:
  - Often cases where there will be skewed distribution of data below 25% or above 75% that affects the analysis and should be removed from the data set
- Missing values treatment :
  - Identify the columns where there more than 30-40% of missing values and handle them with appropriate approach by imputing with valid methods of mean/ medium or mode

Univariate Analysis: Categorical Columns

Grade

- Observation: Loan applicants who are part o GRADE B,D,C are more defaulted when compared to other grades
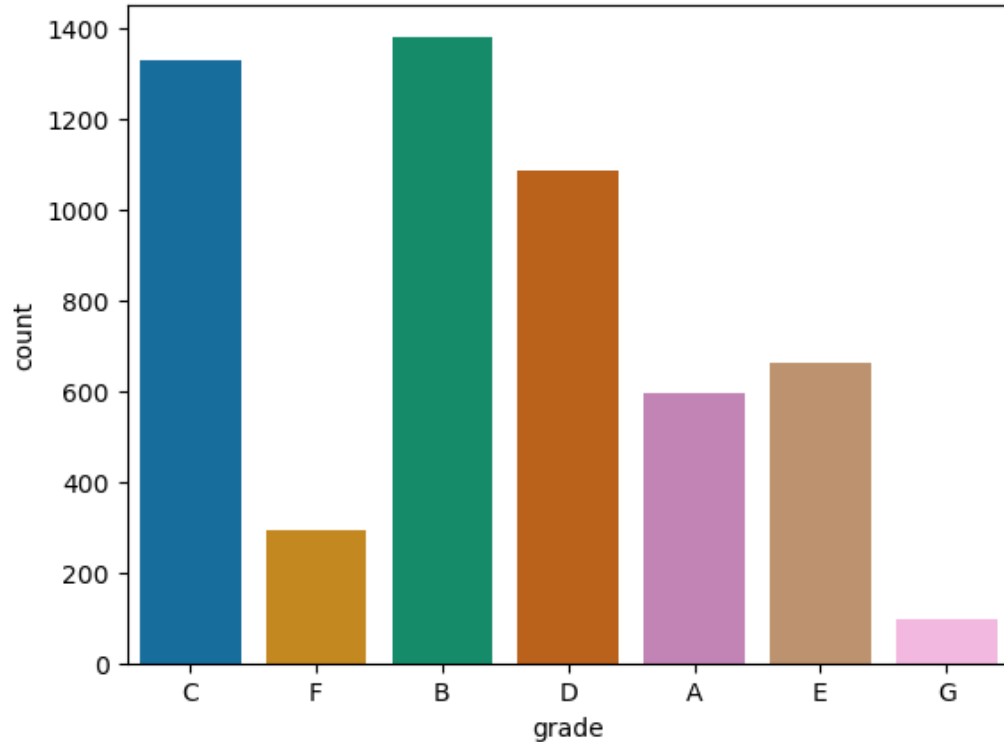
Term

- Loan applicants whose term is 36months have high chances of defaulting

Aswin J Shaji

Lokesh Gaddam

Univariate Analysis: Categorical Columns
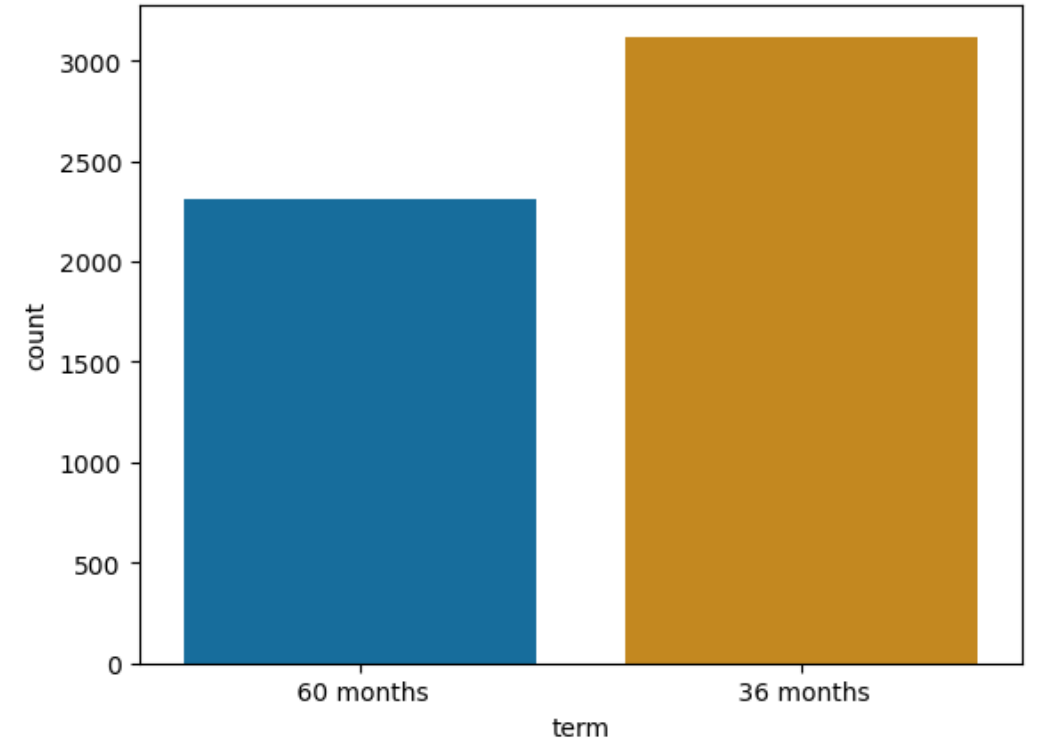
Emp Length

- Applicants who have employment tenure as 10years are more likely to be defaulted
- which ideally does not make any sense because generally with more experience the annual income will be high and hence minimal defaulting
- we will have to do bi-variate analysis on this for further insights combining with other categorical or numerical values
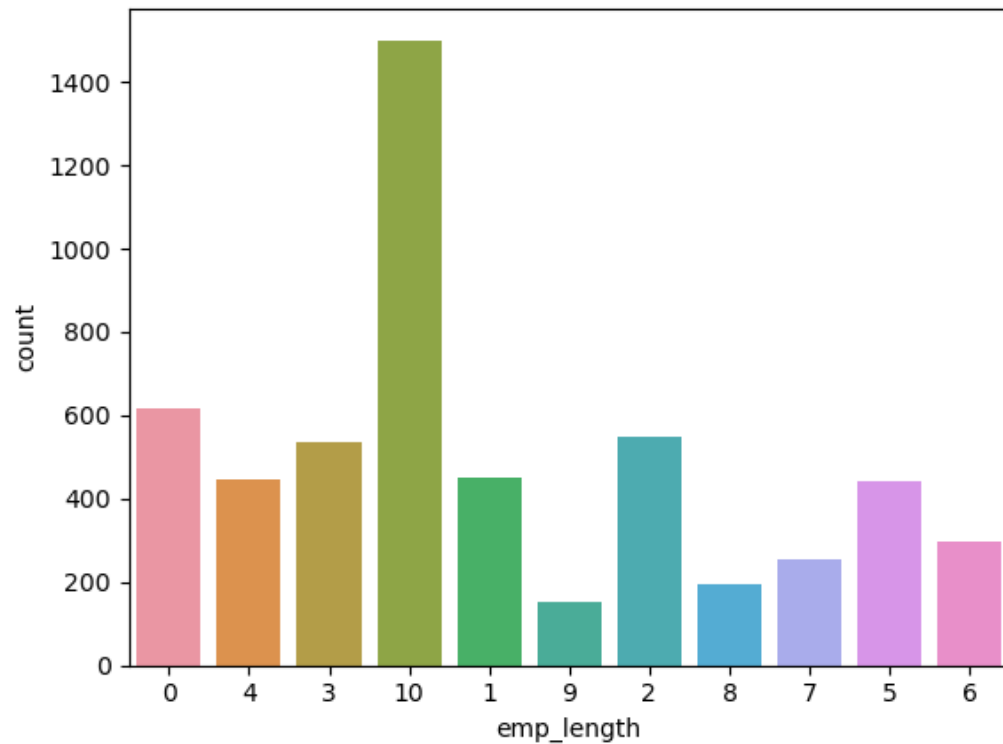
Verification status

- Applicants who have not be verified have high changes for default

- However, the verified status also is close to defaulters, so this required additional analysis with other columns to get more insights

Aswin J Shaji

Lokesh Gaddam
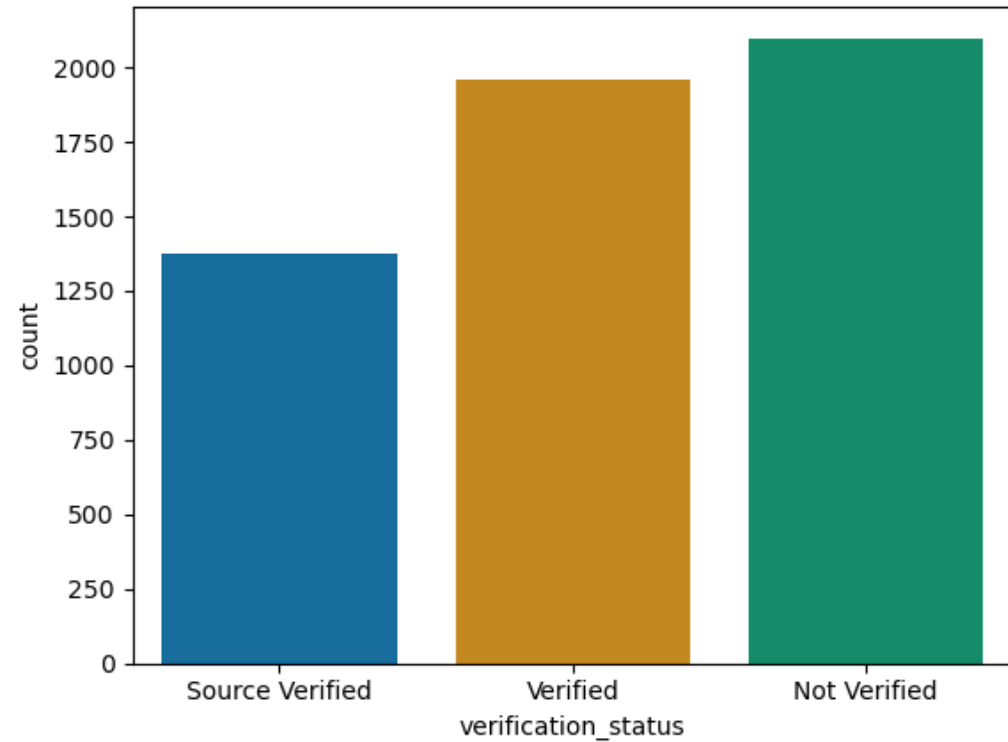
Purpose

- Applicants who plan to clear other loans/ debts consolidation are more susceptible for defaulting
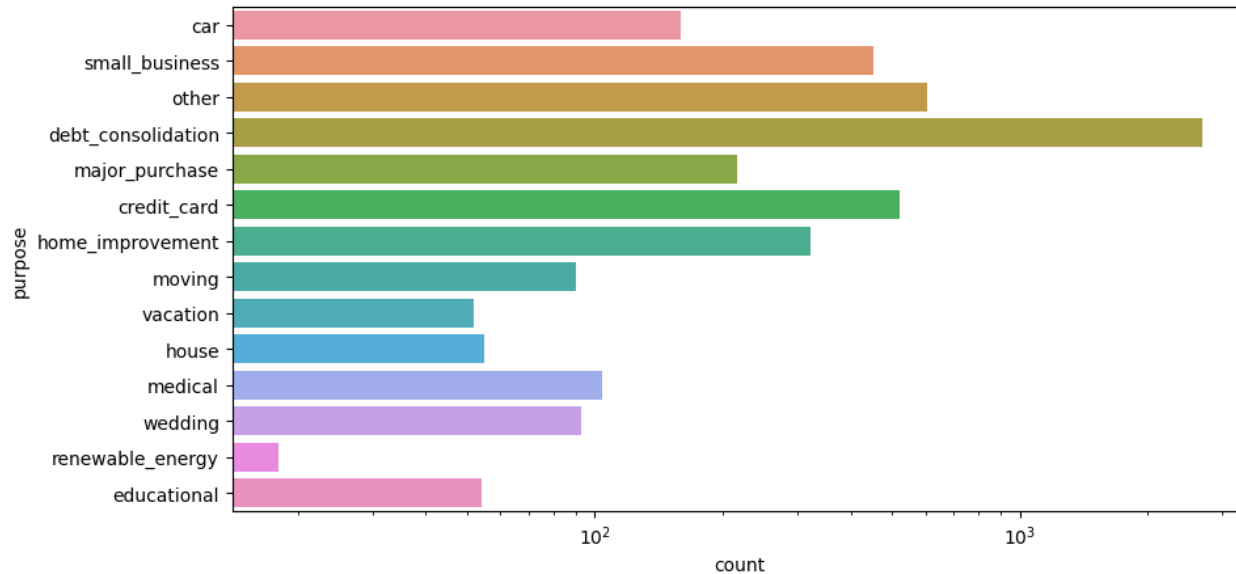
Home Ownership

- Applicants on RENT have most possibility of defaulting on the loan. And the next immediate category is those applications who have mortgages have possibility of defaulting
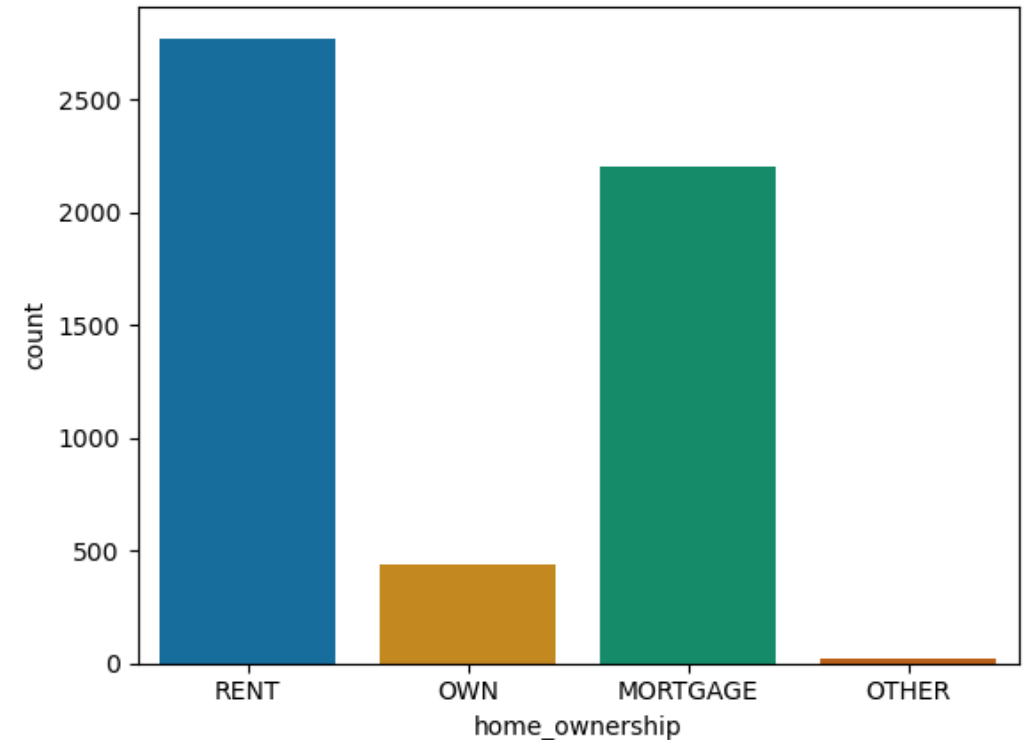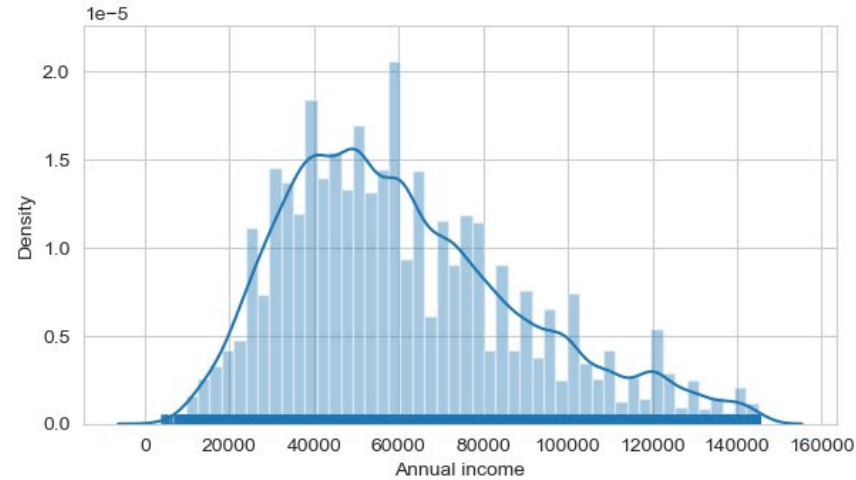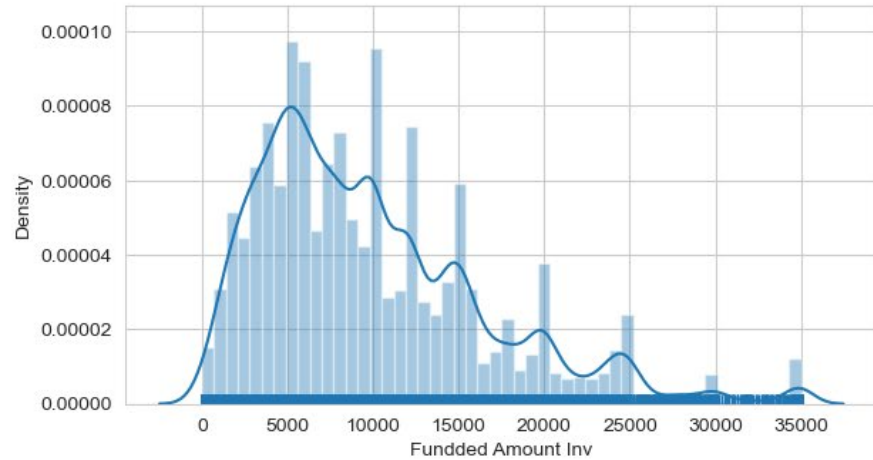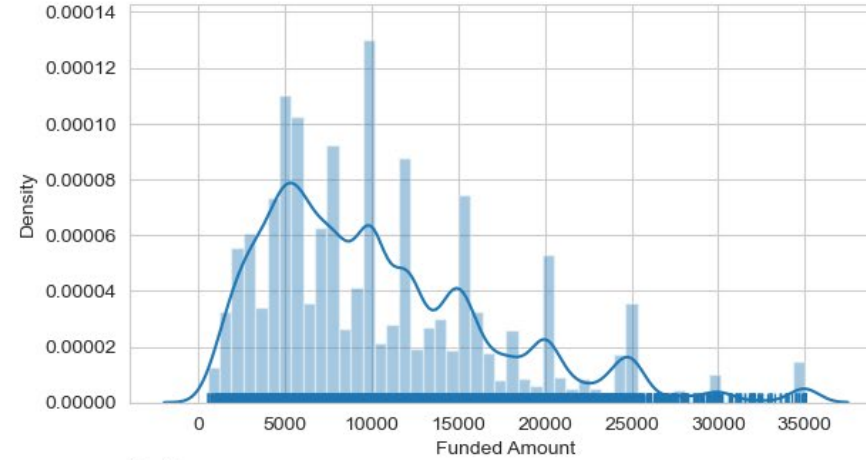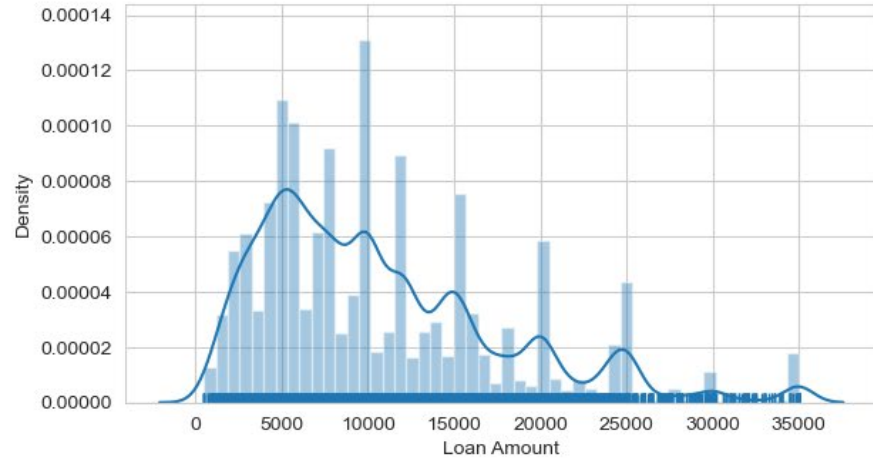
Inference:

- The distribution for all the 3 columns loan_amnt, funded_amnt, funded_amnt_inv are all almost the same

- Majority of the loan amounts taken are within the range of 4000 – 15000

- Majority of the annual income are ranging from 4ok – 80k

Aswin J Shaji
Lokesh Gaddam

Annual income & Purpose

- Applicants with annual income between 60-70K taken loan for home improvements tend to default in high number with next immediate inline with the small business, credit card and wedding
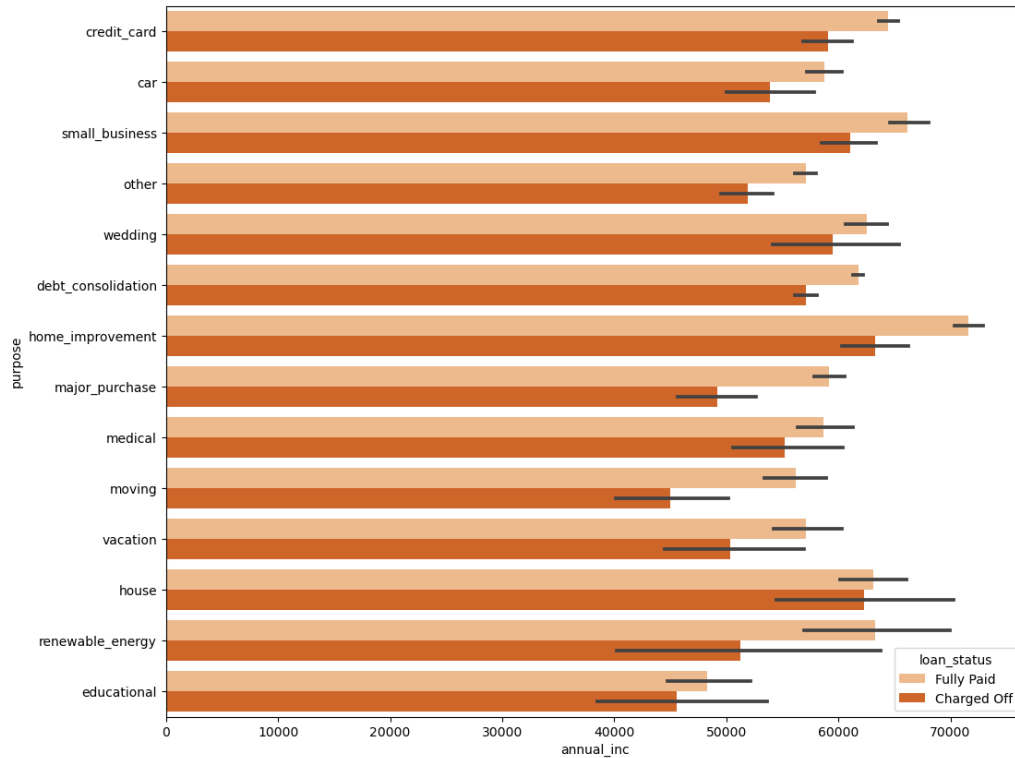
Annual income with Home Owner

- Applicants with annual income of 60-70 k and with mortgage are the ones who are defaulting maximum

Aswin J Shaji

Lokesh Gaddam
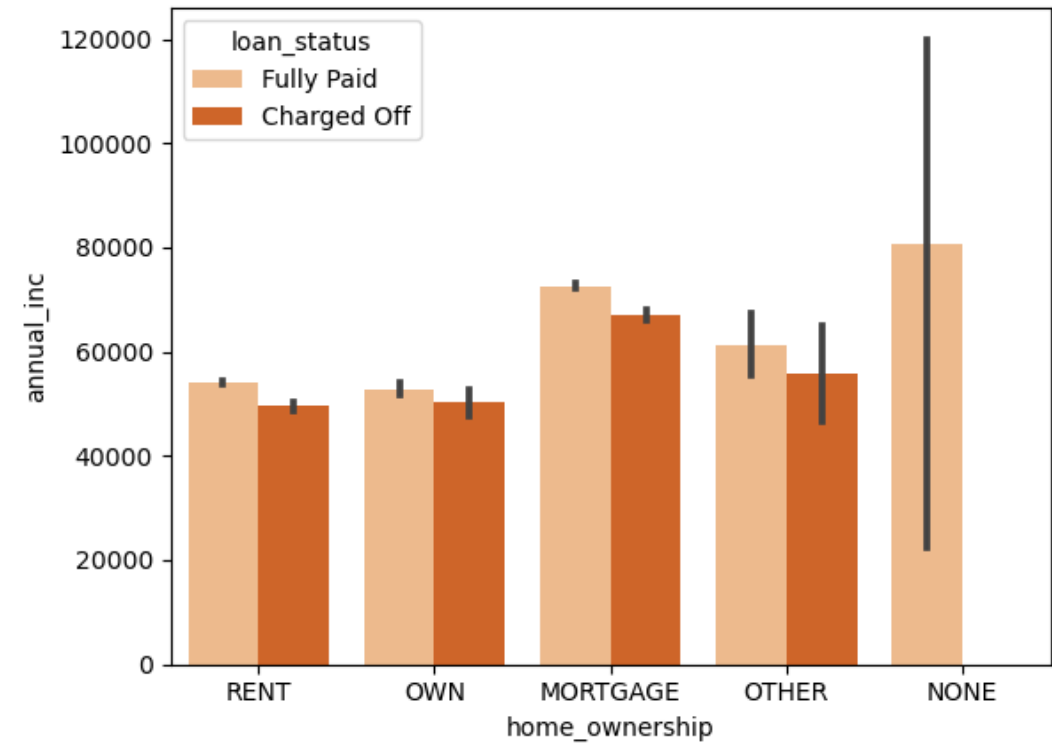
Annual income with Term

- Applicants those who had taken loan to repay in 60 months had more % of number of applicants getting charged off as compared to applicants who had taken loan for 36 months
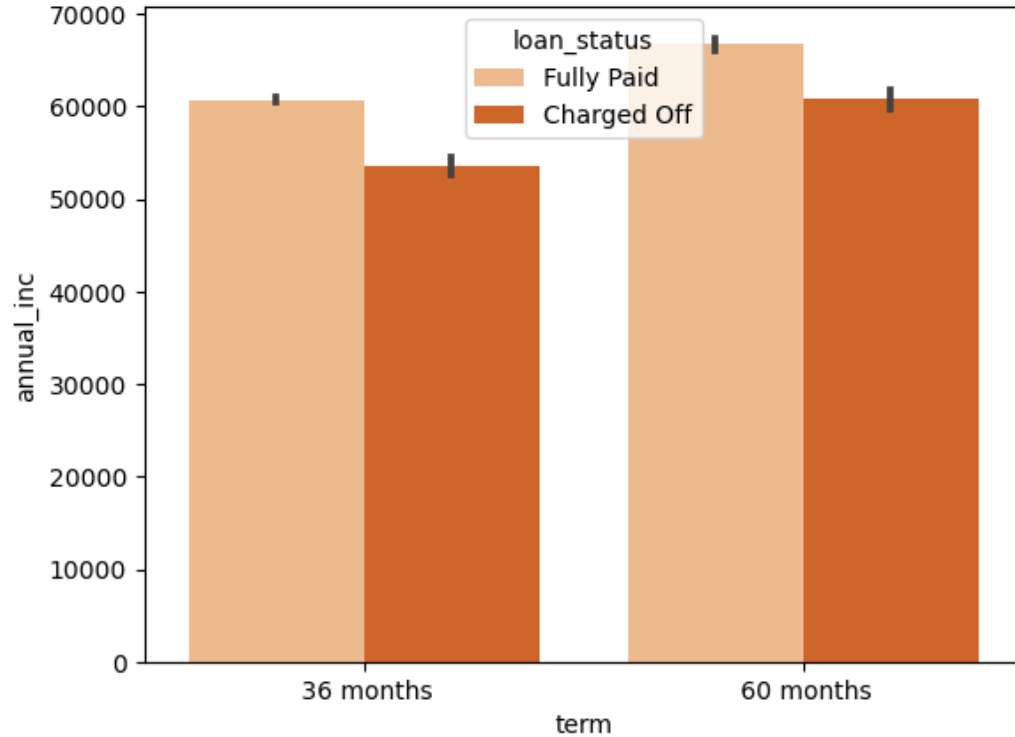
Loan amount with Home ownership

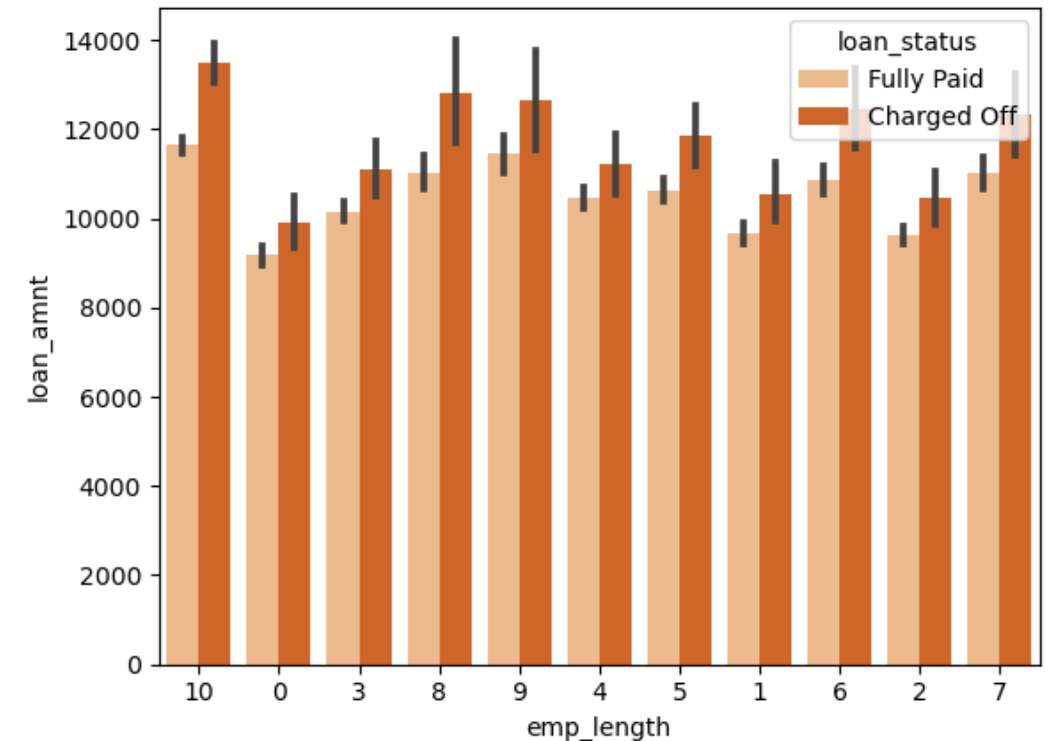- Applicants whose home ownership is 'MORTGAGE and have loan of 13-14k are possible defaulters
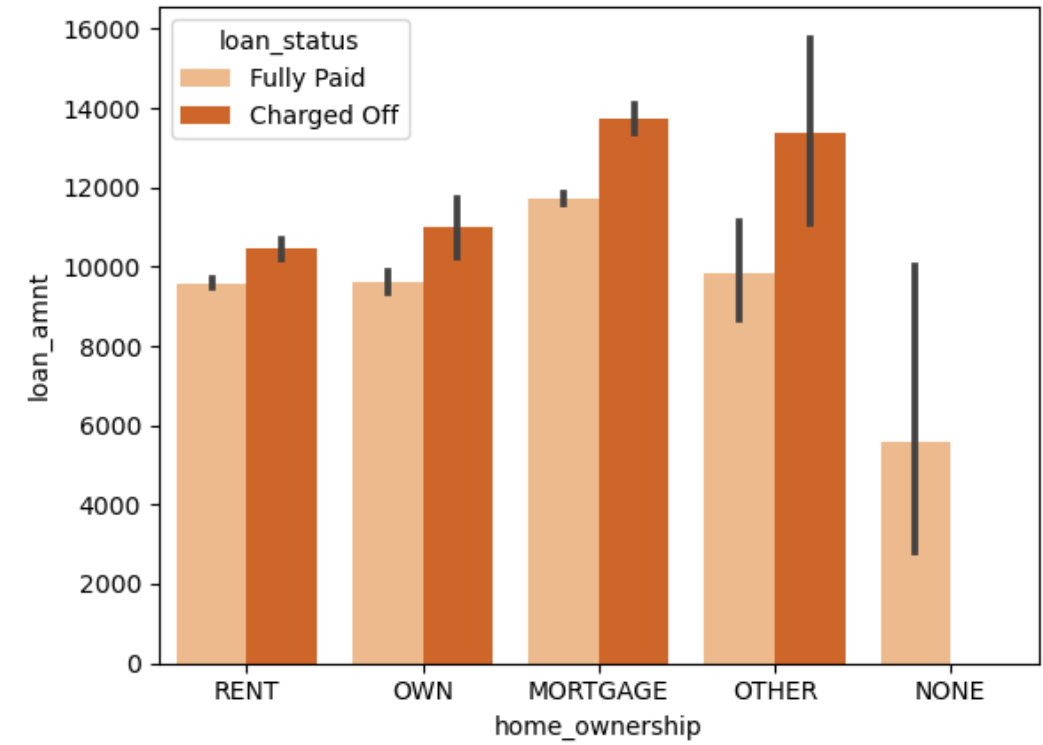
Loan amount with purpose

- Borrowers seeking loan amounts between 11k and 14k for the purpose of small businesses tend to have a higher defaulter rate.
- The subsequent group with a high likelihood of defaulter primarily consists of those seeking loans for credit card and debt consolidation purposes.
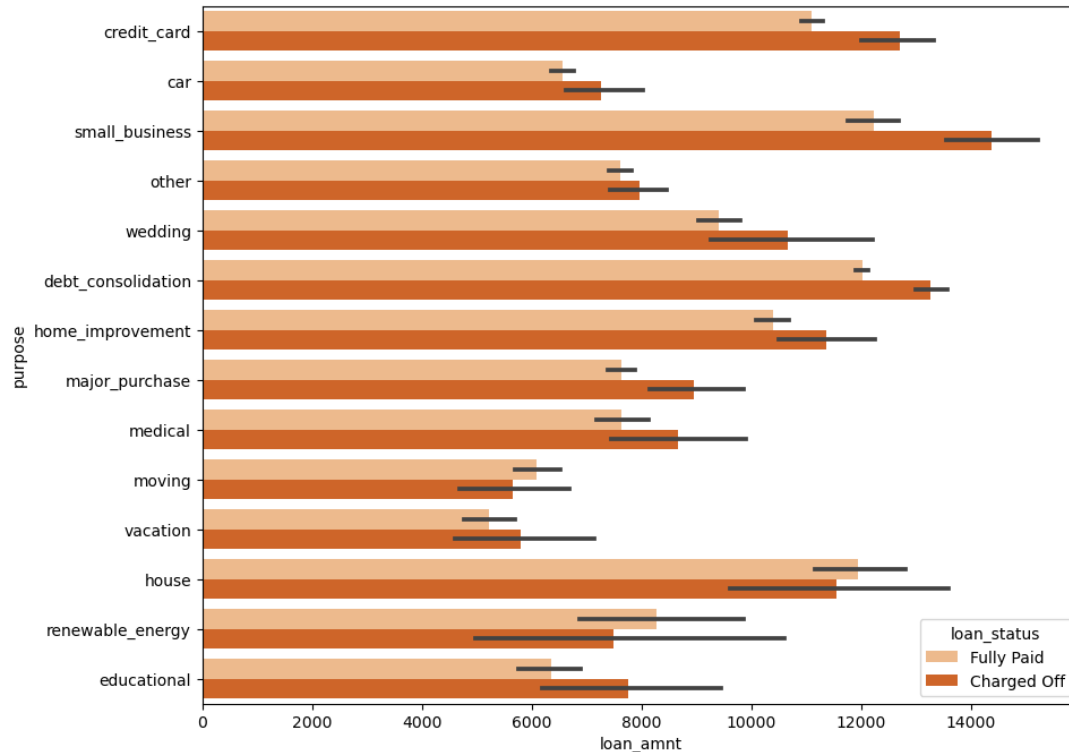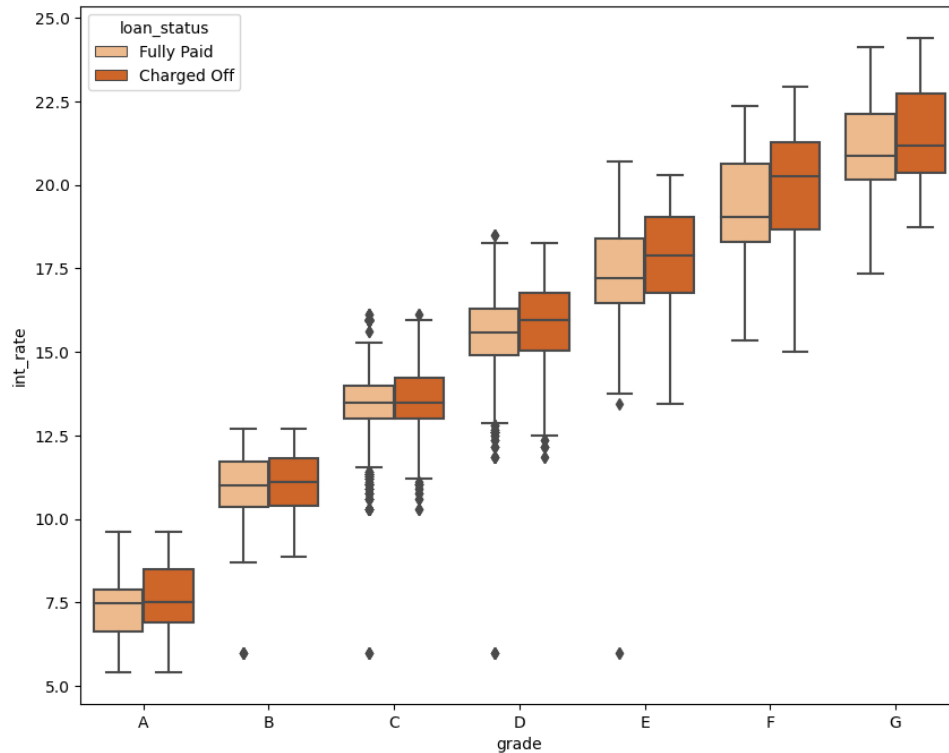
Loan amount with employee length

- Borrowers with a decade of work experience who have borrowed between 13,000 and 14,000 dollars tend to default on their loans.
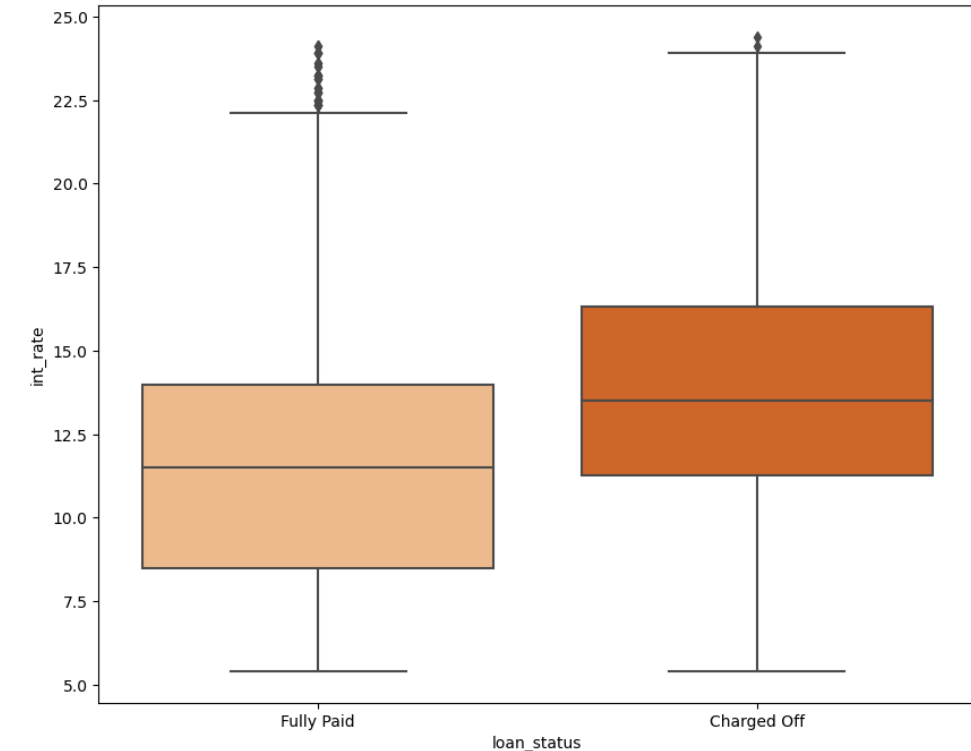
Aswin J Shaji

Lokesh Gaddam

Grade with Interest rate

- A grade, used by banks to assess borrowers, ranges from A for high ratings to F for low ratings.
- Interest rates rise as the grade shifts from A to F, signifying that higher-graded applicants receive lower interest rates, while rates increase with a move from A to F.
- Furthermore, with an increase in grade and interest rates, there's a corresponding increase in the likelihood of default, particularly in immediate concerns such as defaults in credit card and debt consolidation.
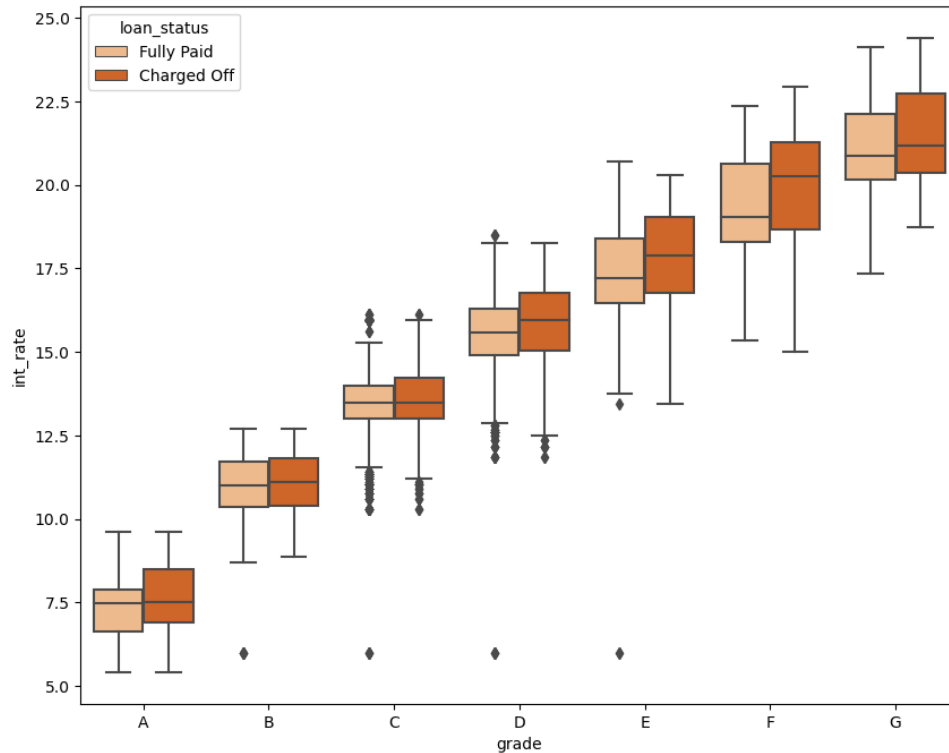
Loan Status with Interest Rate

- As the interest rates are increasing beyond 11% there is a trend in the defaulters, and it increases as it moves beyond 14%

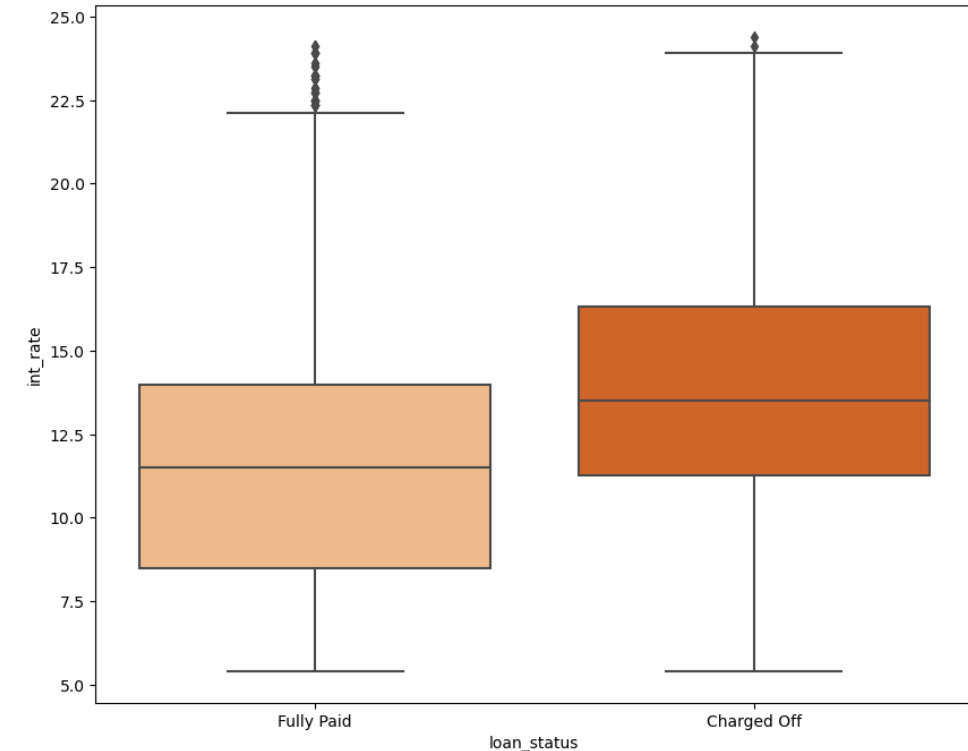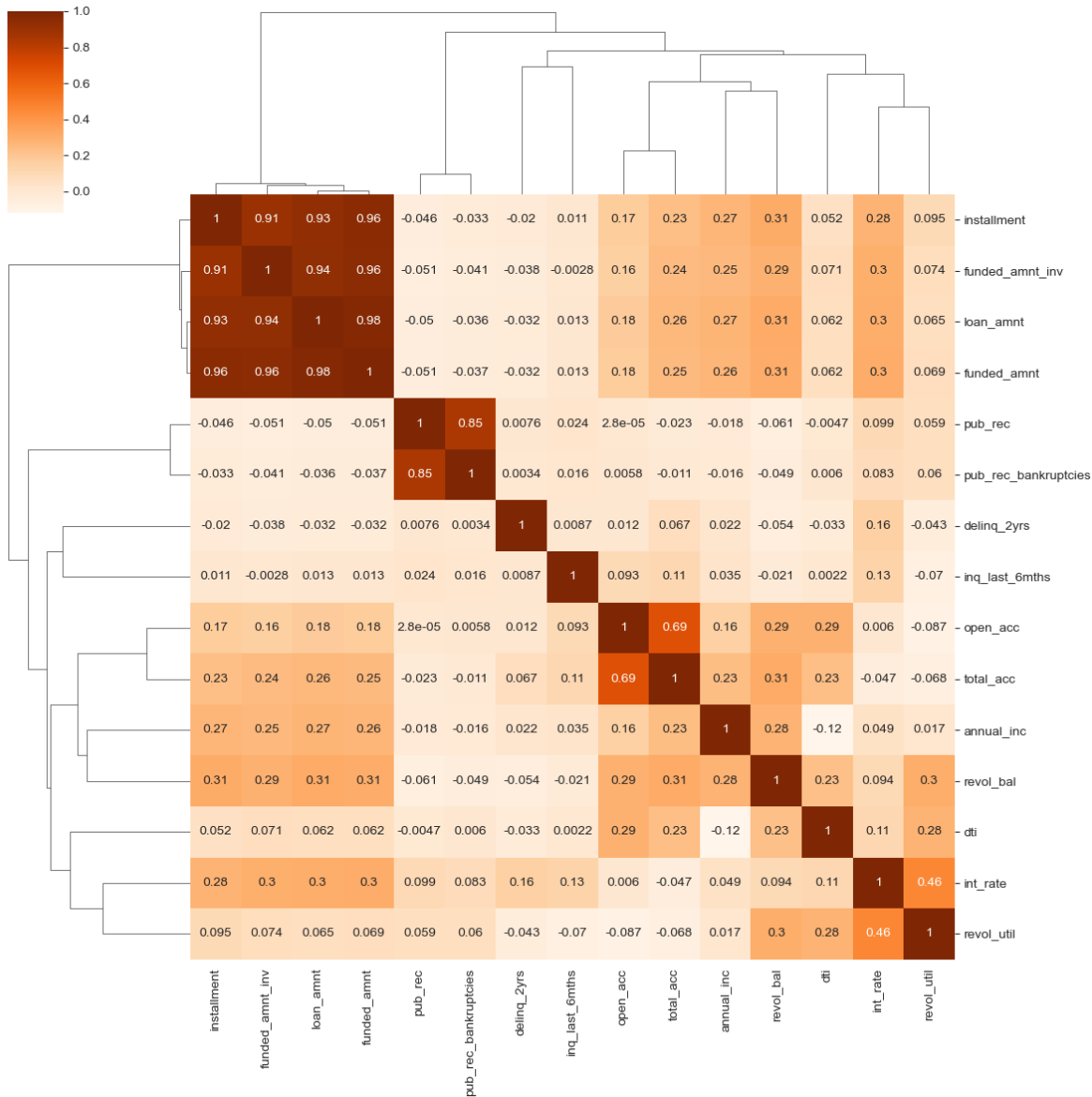Aswin J Shaji

Lokesh Gaddam

Grade with Interest rate

- A grade, used by banks to assess borrowers, ranges from A for high ratings to F for low ratings.
- Interest rates rise as the grade shifts from A to F, signifying that higher-graded applicants receive lower interest rates, while rates increase with a move from A to F.
- Furthermore, with an increase in grade and interest rates, there's a corresponding increase in the likelihood of default, particularly in immediate concerns such as defaults in credit card and debt consolidation.

Loan Status with Interest Rate

- As the interest rates are increasing beyond 11% there is a trend in the defaulters, and it increases as it moves beyond 14%

Aswin J Shaji

Lokesh Gaddam

Inference:

- Columns *installment, funded_amnt, loan_amnt* have high correlation as per the cluster map

- Public records related fields *pub_rec & pub_rec_bankrupcies* and number of accounts related fields *open_acc & total_acc* are correlated

- Also int_rate has a bit of an impact on installment , *funded_amnt, loan_amt, pub_rec*

EDA Summary

❖ **Observation on Grades:** Loan applicants falling within grades B, D, and C demonstrate higher defaulter rates compared to other grade categories.

❖ **Term Impact on Defaulting:** Loan applicants with a term of 36 months show a higher likelihood of defaulting.

❖ **Employment Length and Defaulting:** Applicants with a tenure of 10 years in employment are more prone to defaulting. However, this contradicts the typical understanding that longer job experience correlates with higher annual income and lower default rates. Bivariate analysis is necessary to explore this in conjunction with other factors for deeper insights.

❖ **Verification Status and Default Probability:** Applicants who have not been verified exhibit a higher chance of default. However, verified status also shows a proximity to defaulters, necessitating further analysis with additional columns for comprehensive insights.

❖ **Loan Purpose and Default Probability:** Applicants intending to clear other loans or consolidate debts are more susceptible to defaulting.

❖ **Home Ownership and Defaulting Probability:** Those who rent their homes have the highest likelihood of defaulting on a loan, followed by applicants with mortgages.

❖ **Annual Income, Purpose, and Defaulting:** Applicants earning an annual income between $60,000 and $70,000, seeking loans for home improvements, tend to default in higher numbers, followed closely by those seeking loans for small businesses, credit cards, and weddings.

❖ **Annual Income and Home Ownership Impact on Defaulting:** Applicants with an annual income of $60,000 to $70,000 and with mortgages are the most frequent defaulters.

❖ **Loan Amount and Home Ownership Impact on Defaulting:** Applicants owning homes with mortgages and having loans within the range of $13,000 to $14,000 show a propensity for defaulting.

❖ **Loan Amount and Purpose Impact on Defaulting:** Borrowers seeking loan amounts between $11,000 and $14,000 for small business purposes exhibit a higher rate of defaulting, followed by those seeking loans for credit cards and debt consolidation.

❖ **Loan Amount and Employee Length Impact on Defaulting:** Borrowers with a decade of work experience, borrowing between $13,000 and $14,000, tend to default on their loans.

❖ **Relationship between Grade and Interest Rate:** Grading, used by banks to evaluate borrowers from A to F, influences interest rates. Interest rates rise with a shift from A to F, indicating that higher-graded applicants receive lower rates. This increase in grade and interest rates corresponds with a higher likelihood of default, particularly in credit card and debt consolidation defaults.

❖ **Loan Status in Relation to Interest Rates:** There's an observable trend in defaulters as interest rates exceed 11%, further increasing as it surpasses 14%.

Aswin J Shaji   Lokesh Gaddam

Graphs Plotting
refer next slide for summary

IV – Information Value

WoE – weight of evidence
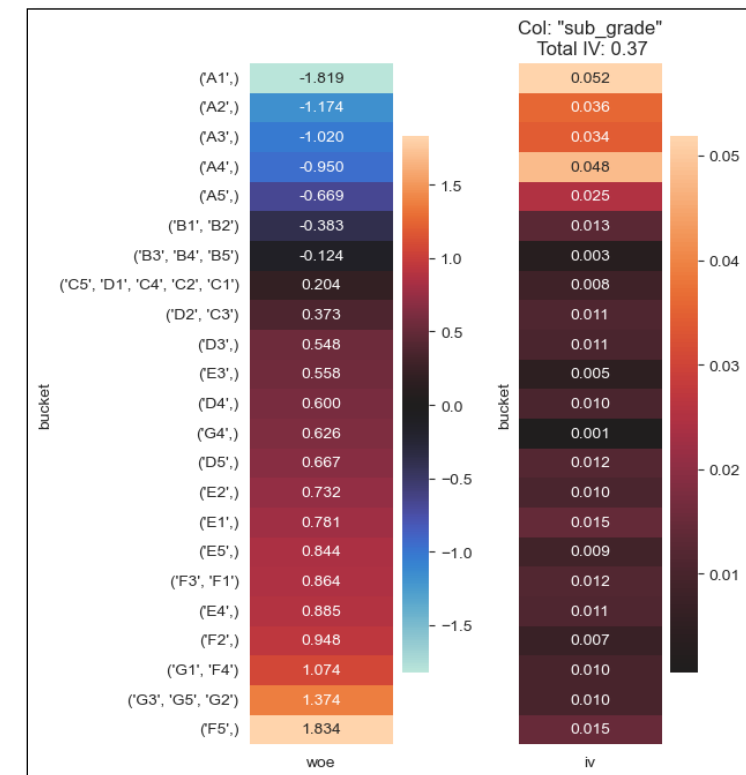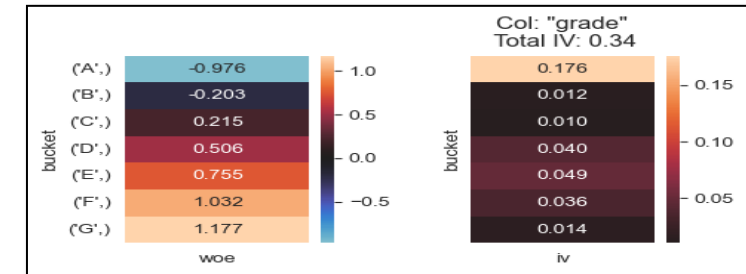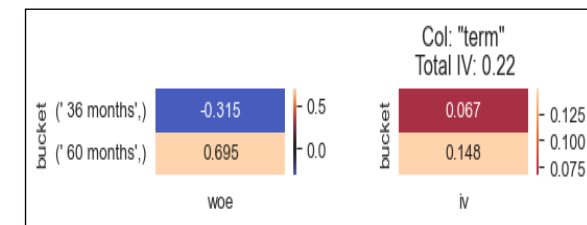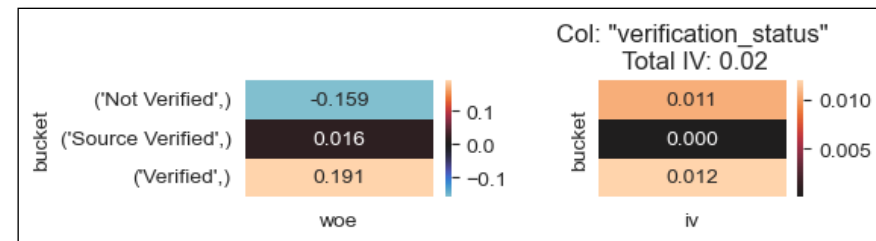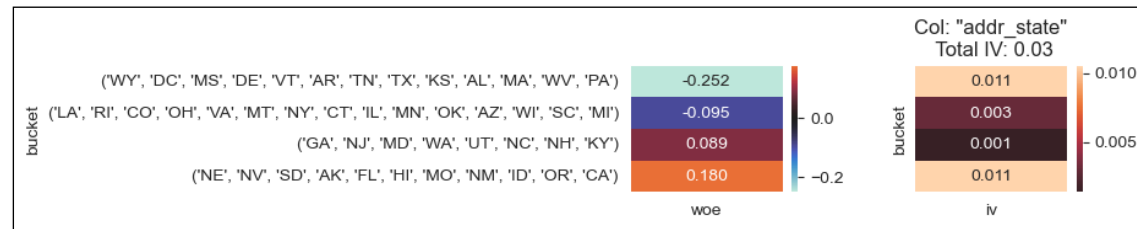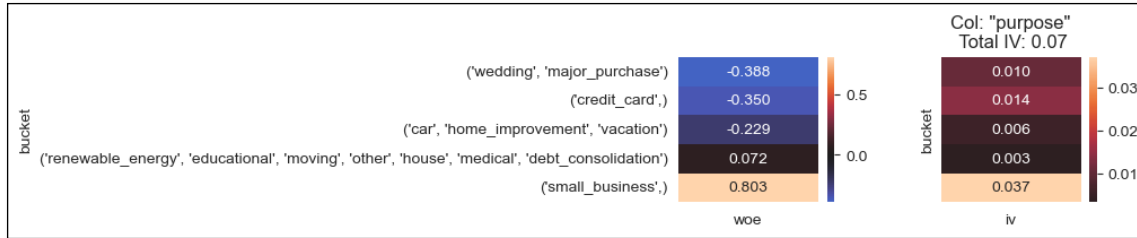
Summary
refer previous slide for graphs

Customers with Lower Default Probability:

- Shorter loan terms associate with a decreased likelihood of default.

- Higher credit grades (above B5) correspond to reduced default probabilities, with an inverse relationship—higher grades indicate lower default chances.

- Verified income is linked to a lower likelihood of default.

- Ideal loan purposes with lower default chances include weddings, major purchases, and credit cards.

- Following closely, car loans, home improvement, and vacation purposes exhibit lower default probabilities.

- States with the most favorable customers include WY, DC, MS, DE, VT, AR, TN, TX, KS, AL, MA, WV, and PA.

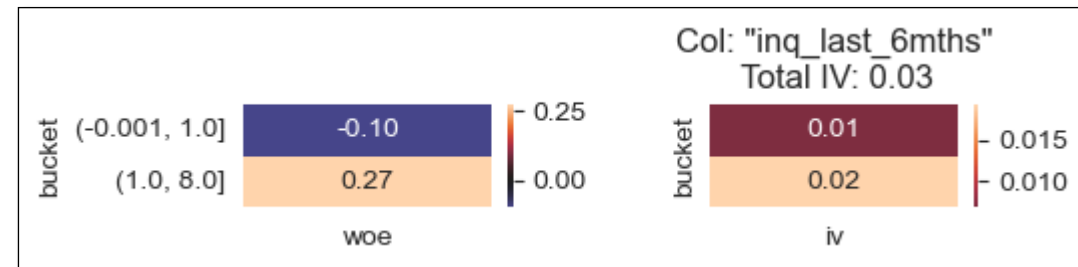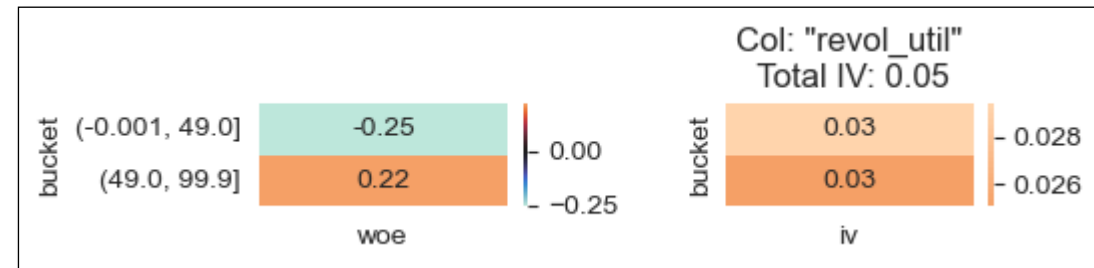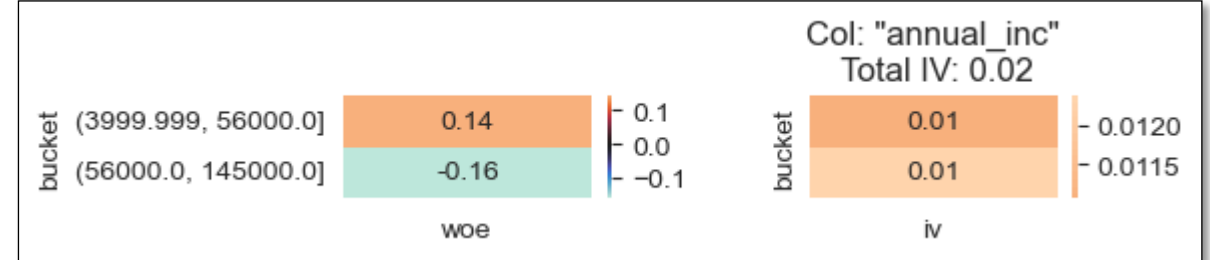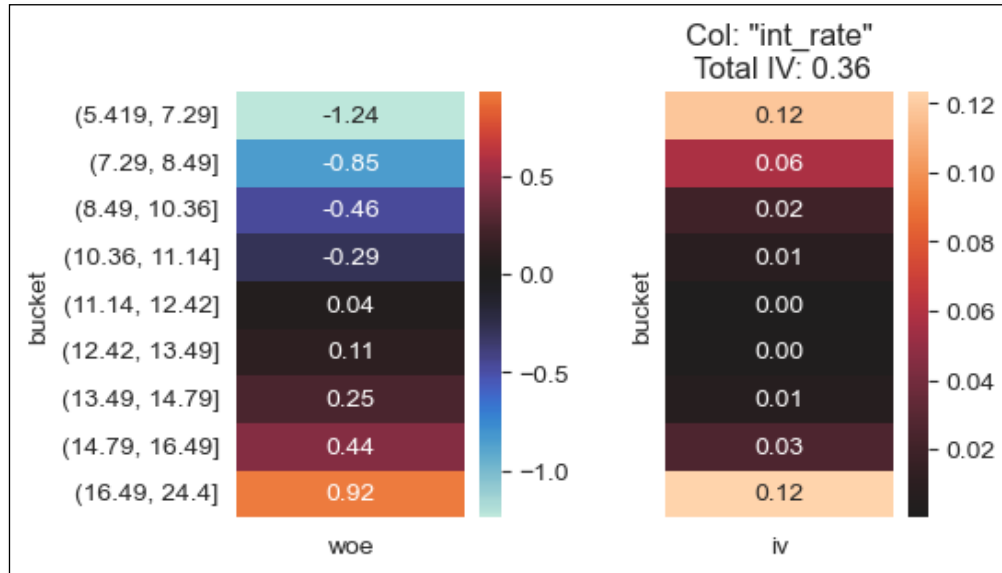## Customers with Higher Default Probability:

- Longer loan terms are associated with an increased likelihood of default.

- Lower credit grades (below C1) are likely to result in defaults, with a higher chance of defaulting as the grade decreases.

- Unverified income correlates with a higher chance of default.

- Loans intended for funding small businesses represent the least favorable purposes in terms of default probability.

- States with less desirable customers in terms of default likelihood include NE, NV, SD, AK, FL, HI, MO, NM, ID, OR, and CA.

Graphs Plotting
refer next slide for summary

IV – Information Value

WoE – weight of evidence

Summary
refer previous slide for graphs

Good Customers with Lower Default Probability:

- The interest rate for the loan plays a significant role in default probability; lower interest rates correlate with a reduced chance of default.

- Optimal interest rate: Below 8% offers the lowest default likelihood, while below 10% stands as the next favorable option.

- Higher income levels are associated with a decreased likelihood of default.

- A lower credit utilization percentage is linked to a reduced chance of default.

Bad Customers with Higher Default Probability:

- The interest rate for the loan greatly impacts the chance of default; higher interest rates correspond to an increased probability of default.

- Loans above 14% pose a high chance of default and should be avoided.

- Lower income levels are linked to a higher probability of default.

- Higher credit utilization percentages are associated with an increased chance of default.

- A higher likelihood of default is indicated if more than 2 inquiries have been made in the last 6 months.

Preferred Customer Profile:

Loan Request Details:

- Seeks a shorter-term loan duration.

- Holds a credit grade higher than B5.

- Verification of income is confirmed.

- Optimal loan purposes: weddings, major purchases, and credit cards. Subsequently, car, home improvement, and vacation rank as the next best options.

- Resides in one of the following states: WY, DC, MS, DE, VT, AR, TN, TX, KS, AL, MA, WV, PA.

- Prefers a lower interest rate, ideally not exceeding 8%; otherwise, 10% is acceptable.

- Higher income bracket.

- Maintains a lower credit utilization.

Less Preferred Customer Profile:

Loan Request Details:

- Seeks a long-term loan duration.

- Holds a credit grade lower than C1.

- Lacks verified income.

- Seeks a loan specifically for funding a small business.

- Resides in one of the following states: NE. NV, SD, AK, FL, HI, MO, NM, ID, OR, CA.

- Prefers a higher interest rate, exceeding 14%.

- Lower income status.

- Maintains a higher credit utilization percentage.

- Has two or more inquiries in the past 6 months.

Key consideration for Loan Approval:

When approving loans, Information value (IV) and Weight of Evidence (WOE) serve as the sole objective analysis criteria, focusing on the outlined factors.