LEANDING CLUB CASESTUDY

TEAM MEMBERS

SHAMILY JENNYMOOR AND SHIVANG PATEL

CONTENT

- Business Objectives
- Data Description
- Data Understanding
- Data Cleaning and Pre-processing
- Univariate Analysis
- Bivariate Analysis
- Multivariate Analysis
- Final Inference and Suggestions
- Useful Links

BUSINESS OBJECTIVES

- Lending Club is a marketplace that connects borrowers with investors for personal loans. The primary challenge is minimizing financial losses caused by lending to risky applicants who may default on their loans.
- This study aims to identify applicants likely to default through Exploratory Data Analysis (EDA) of the provided dataset. Patterns indicating default can guide actions like denying loans, reducing loan amounts, or increasing interest rates for high-risk borrowers.
- The analysis will focus on the influence of consumer and loan attributes on default tendencies. The objective is to uncover key driving factors behind loan defaults.
- These insights will enable Lending Club to enhance its portfolio management, improve risk assessment, and reduce credit losses effectively.

DATA DESCRIPTION

- The dataset contains historical loan application data with 111 columns and over 39,000 records, including customer credit history and Lending Club details.
- Analysis will focus on identifying relationships between variables and their impact on loan success or default.
- Only relevant columns influencing loan status will be selected, followed by data cleaning and transformation to create the final dataset for analysis.

PROBLEM SOLVING METHODOLOGY

The data analysis consists four main parts:

- Data understanding
- Data cleaning (cleaning missing values, removing duplicate columns and so)
- Data Analysis
- Recommendations

DATA UNDERSTANDING

Leading Attribute

- Loan Status - Key Attribute (*loan_status*). The column has three distinct values
- Fully-Paid - The customer has paid the loan completely
- Charged-Off - The customer is "Charged-Off" or has "Defaulted“
- Current - These are in progress loan payments and cannot contribute to conclusive evidence if the customer will default of pay in future

DATA CLEANING AND PREPROCESSING

- Load the dataset into a loan_dataset dataframe and remove columns with only NA values.Exclude current loan records and columns with more than 60% missing values.
- Drop ignored, behavioral, granular, and textual columns, including fund_amnt and fund_amnt_inv due to high correlation with loan_amnt.
- Fix data discrepancies like incorrect data types, blanks, and duplicates, and create bins for loan characteristics.
- Derive new columns from issue_date for year/month analysis and include a concluded column for better insights.

UNIVARIATE ANALYSIS

Univariate analysis is deals analysis of a single variable to understand its distribution, central tendency and dispersion.

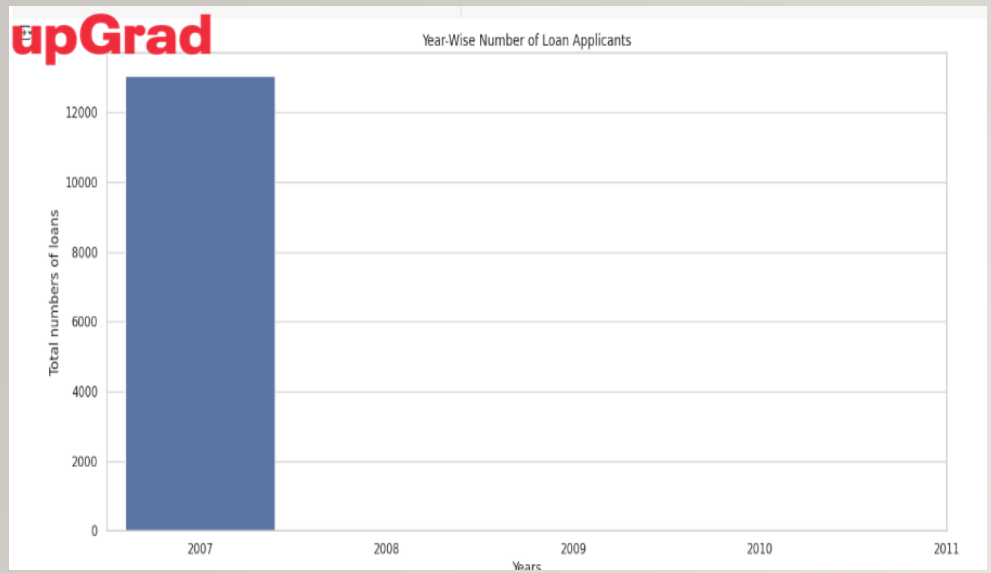
- *Categorical variable*

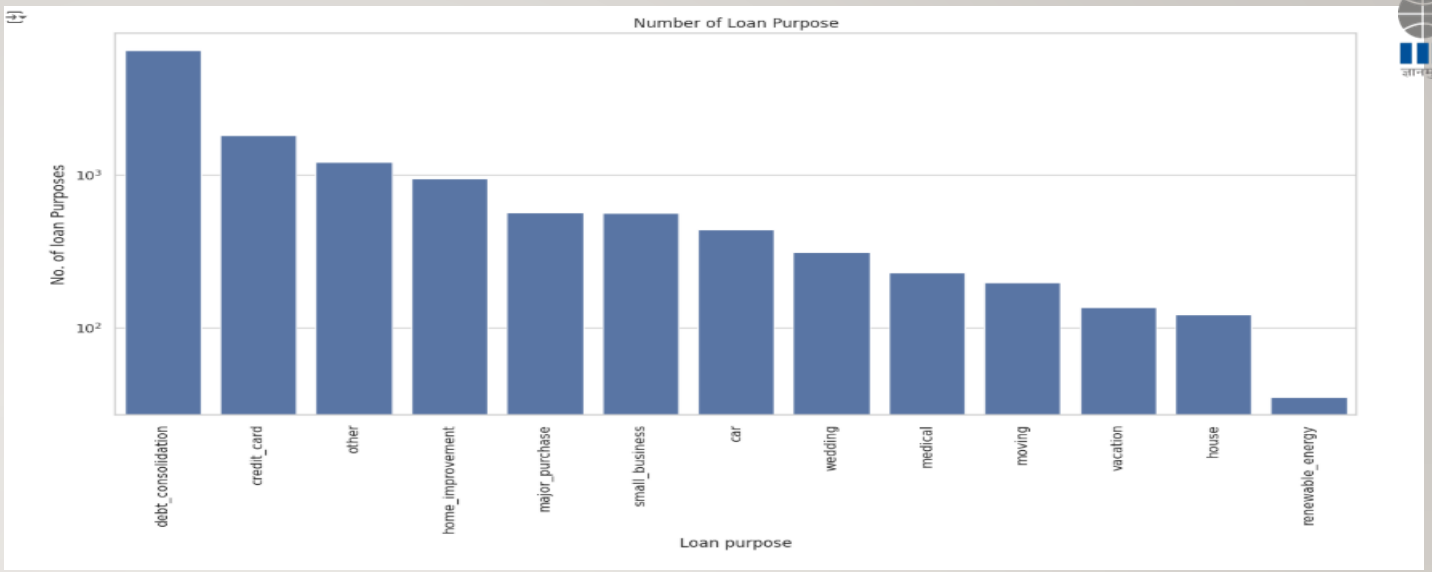
Ordered	UnOrdered
Grade(grade)	Loan Status(loan_status)
Employment tenure(emp_length)	Address State(addr_state)
Term(term)	Home Ownership(home_ownership)
Issue Year(issue_year)	Verification status(verification_status)
Issue Month(issue_month)	Purpose(purpose)

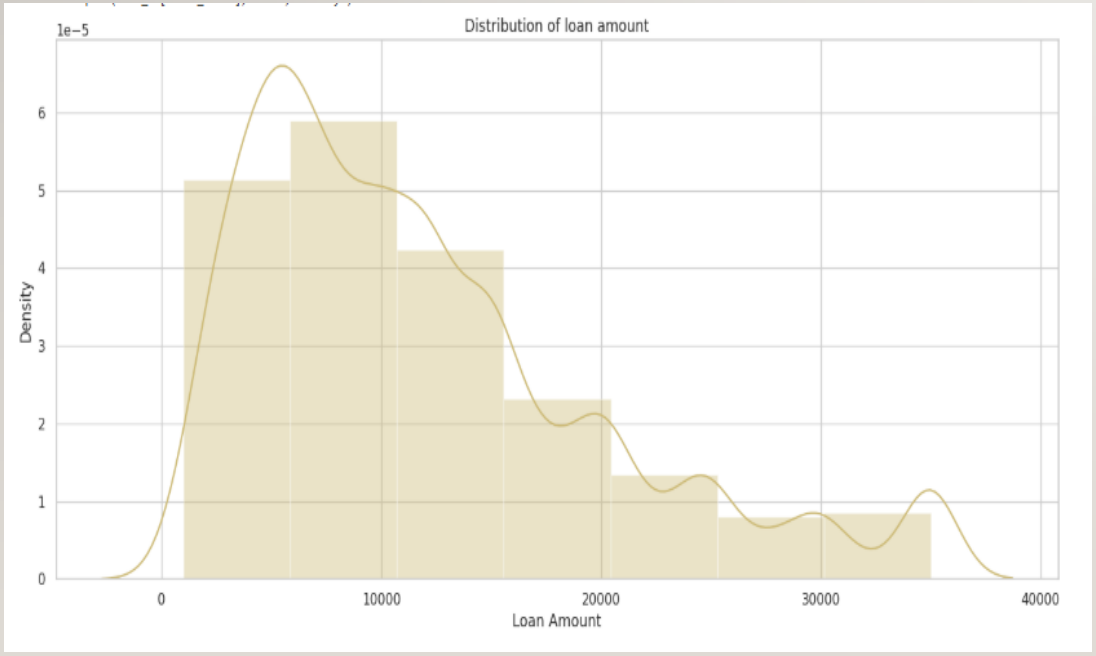
UNIVARIATE ANALYSIS

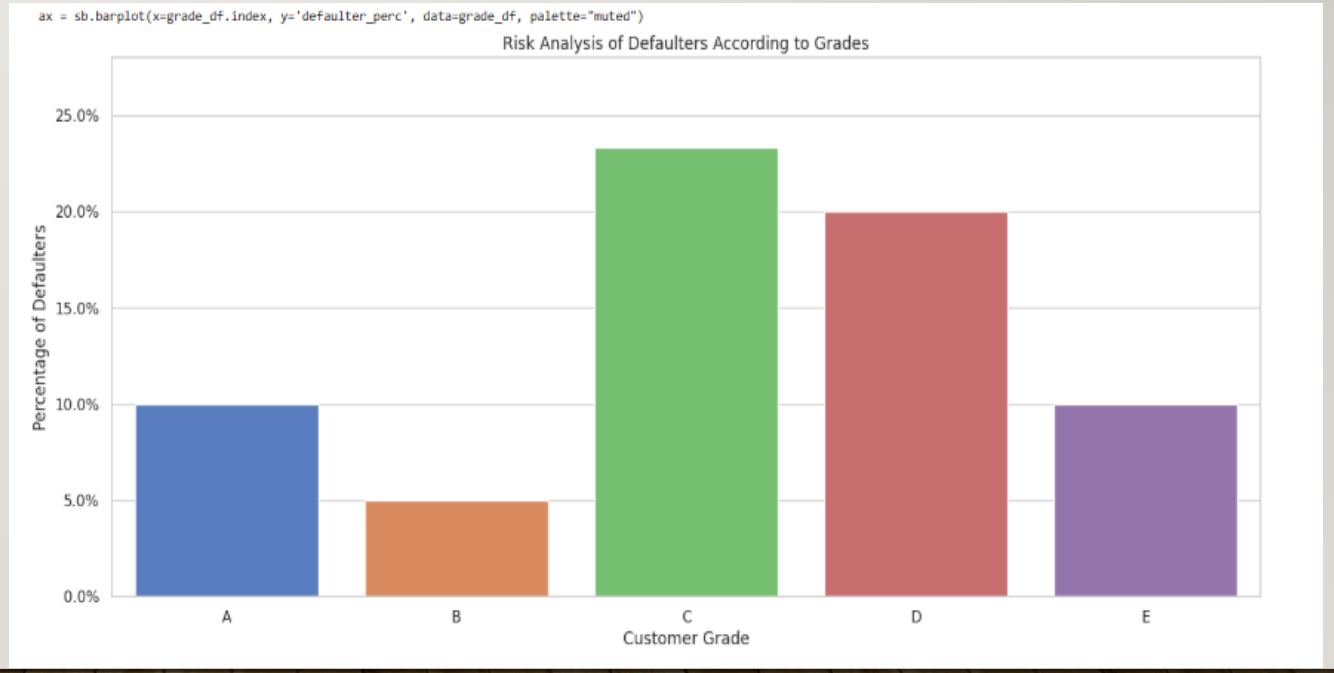
Quantitative variable

- Interest Rate (int_rate)
- *Debt to income ratio(dti)*
- *loan amount (loan_amnt)*
- *Annual income (annual_amnt)*
- *Monthly instalments(installment)*
- *Public record of bankruptcy(pub_rec_bankruptcies)*


Year-Wise Number of Loan Applicants
Total numbers of loans
0
2000
4000
6000
8000
10000
12000
2007
2008
2009
2010
2011
Years

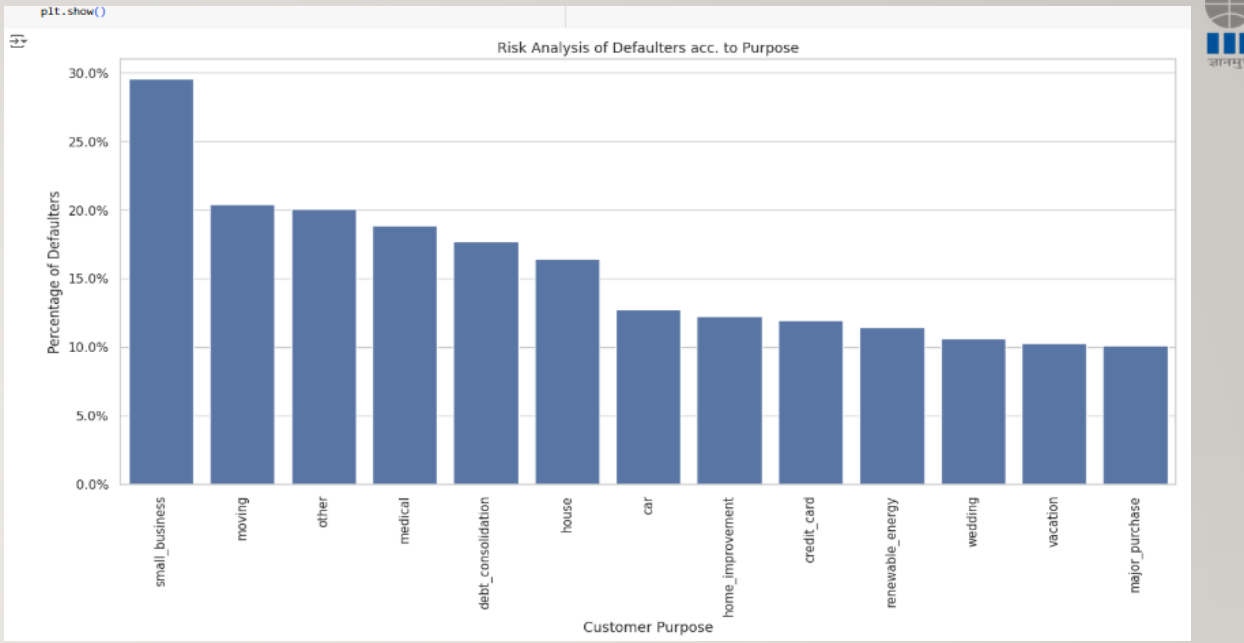

Number of Loan Purpose
No. of loan Purposes
10^3
10^2
debt_consolidation
credit_card
other
home_improvement
major_purchase
small_business
car
wedding
medical
moving
vacation
house
renewable_energy
Loan purpose

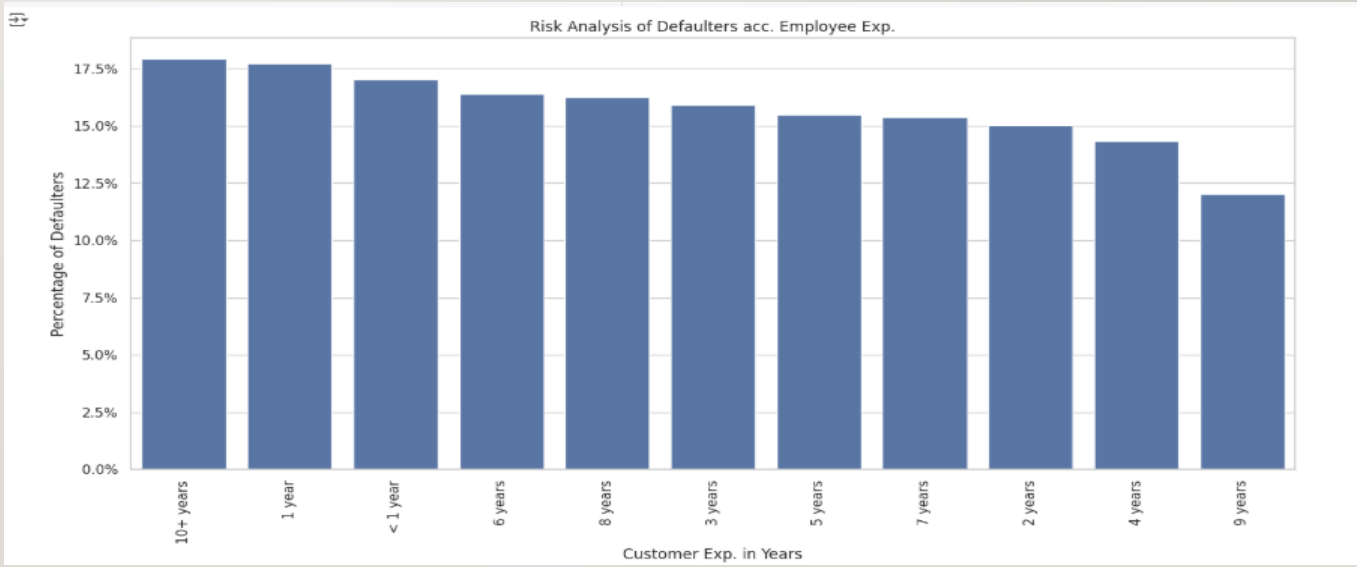

Distribution of loan amount
1e−5
Density
0
1
2
3
4
5
6
0
10000
20000
30000
40000
Loan Amount

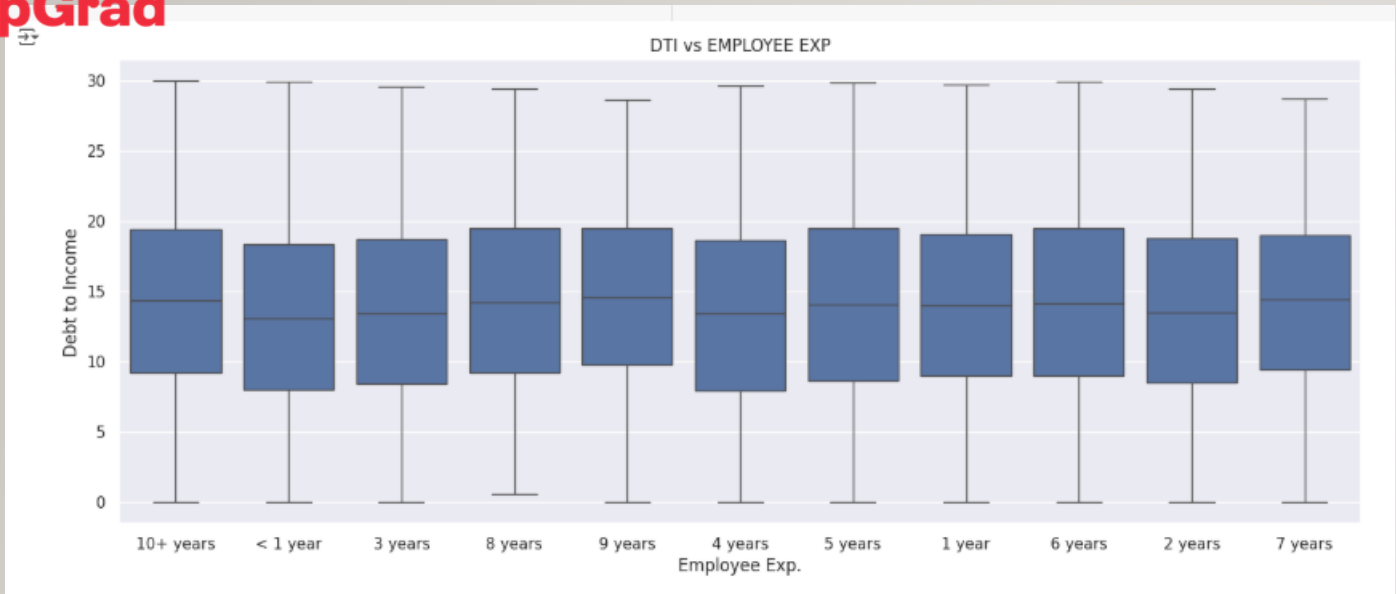

ax = sb.barplot(x=grade_df.index, y='defaulter_perc', data=grade_df, palette="muted")
Risk Analysis of Defaulters According to Grades
Percentage of Defaulters
0.0%
5.0%
10.0%
15.0%
20.0%
25.0%
A
B
C
D
E
Customer Grade

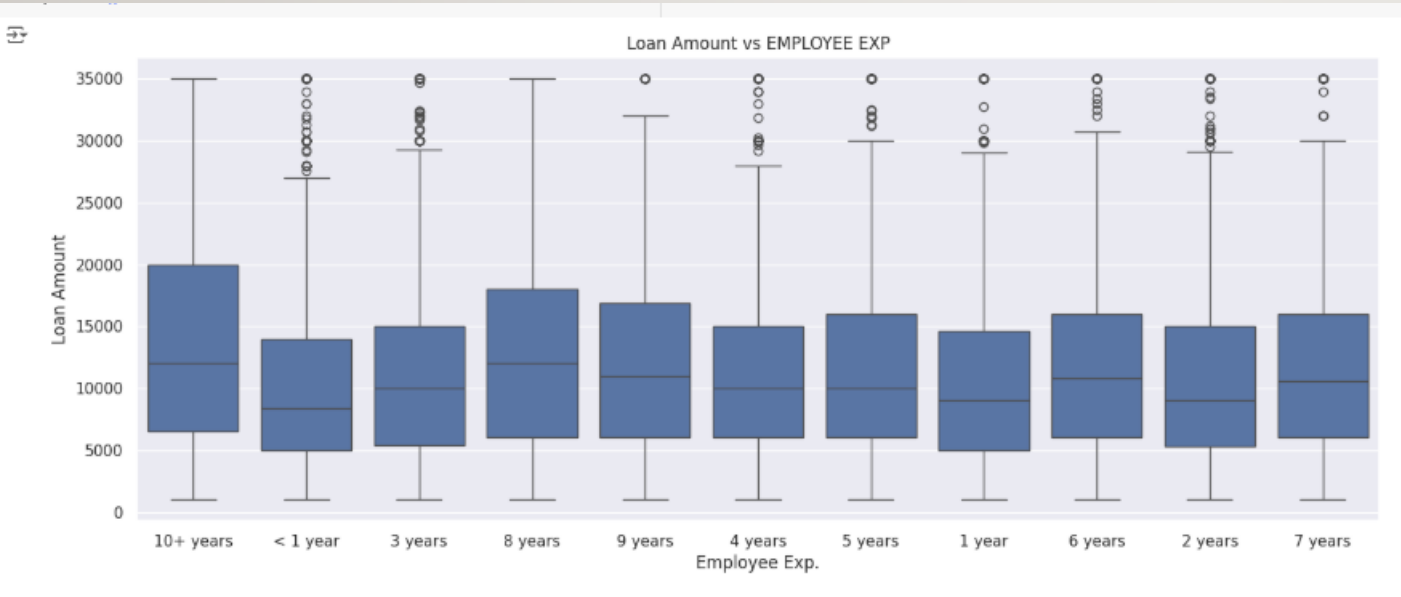
upGrad

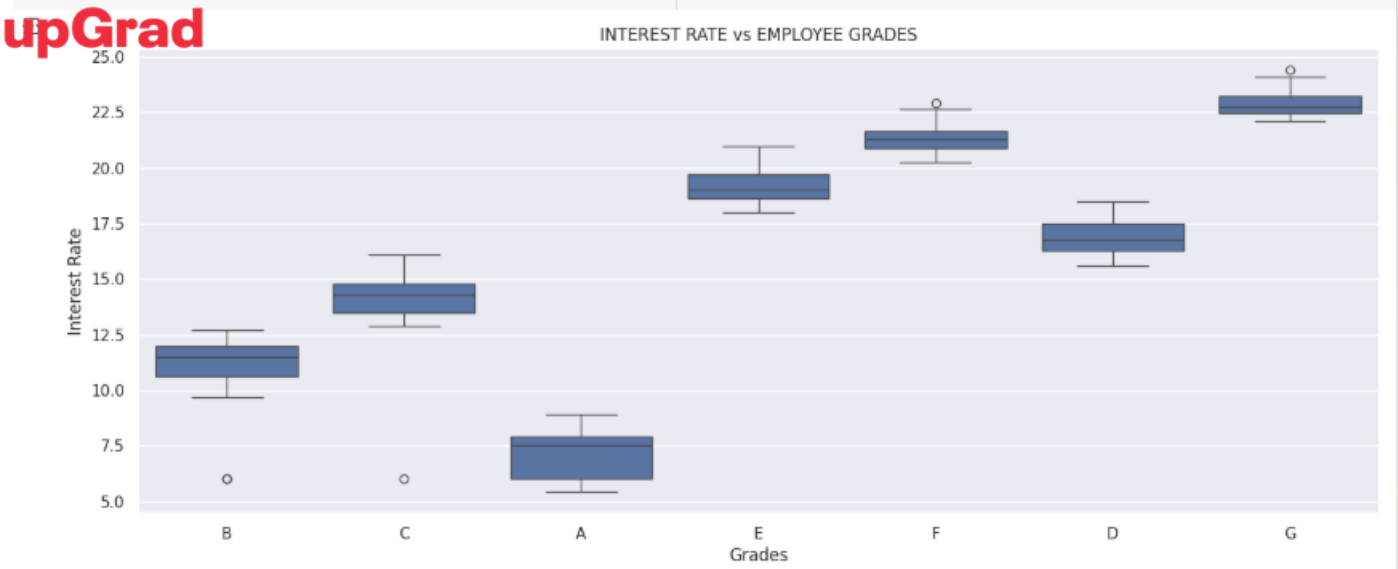
```
plt.show()
```

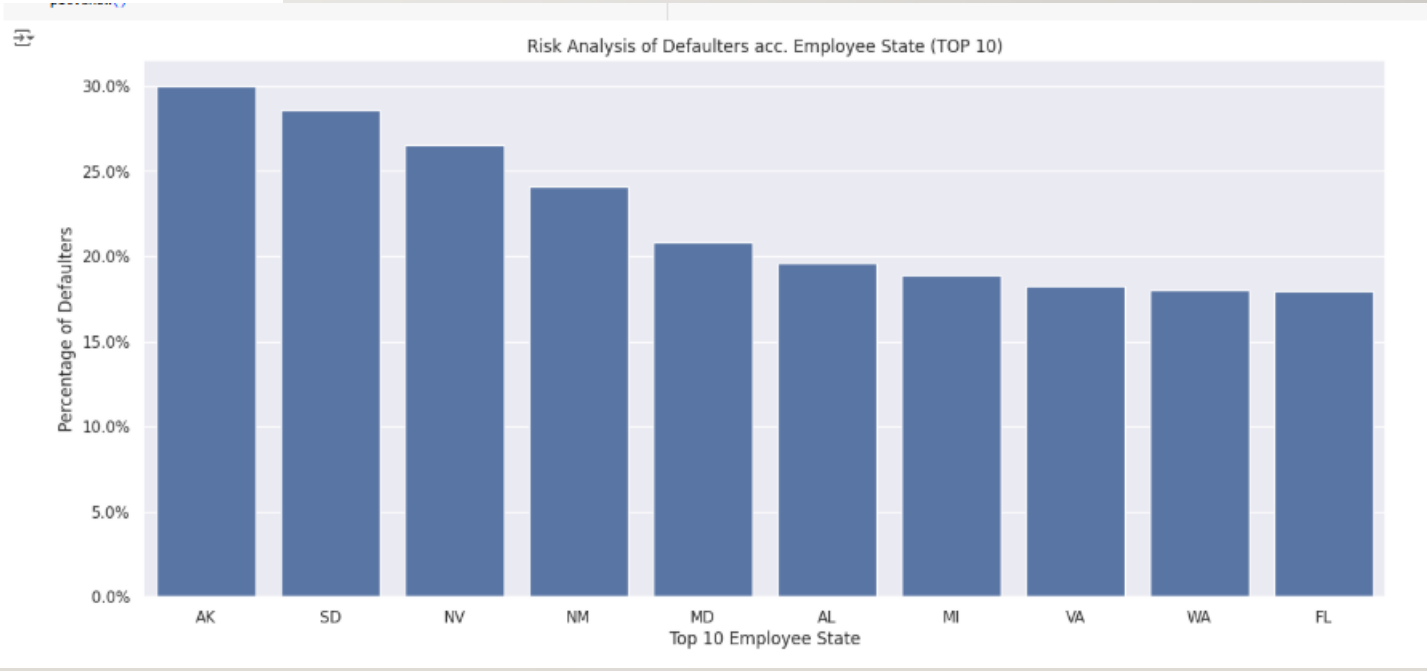


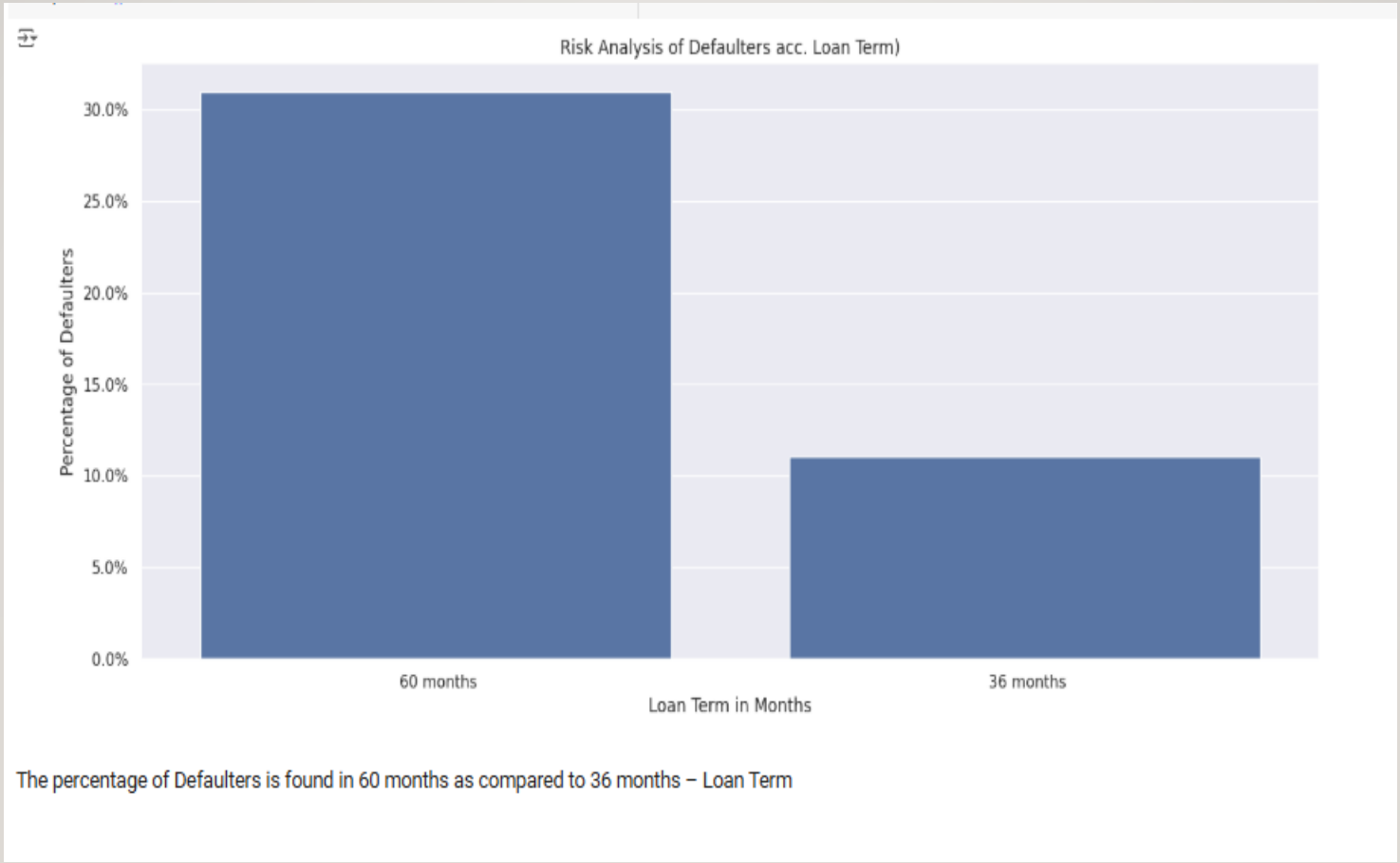



DTI vs EMPLOYEE EXP
Debt to Income
30
25
20
15
10
5
0
10+ years
< 1 year
3 years
8 years
9 years
4 years
5 years
1 year
6 years
2 years
7 years
Employee Exp.


Loan Amount vs EMPLOYEE EXP
Loan Amount
35000
30000
25000
20000
15000
10000
5000
0
10+ years
< 1 year
3 years
8 years
9 years
4 years
5 years
1 year
6 years
2 years
7 years
Employee Exp.


INTEREST RATE vs EMPLOYEE GRADES
Interest Rate
25.0
22.5
20.0
17.5
15.0
12.5
10.0
7.5
5.0
B
C
A
E
F
D
G
Grades


Risk Analysis of Defaulters acc. Employee State (TOP 10)
Percentage of Defaulters
30.0%
25.0%
20.0%
15.0%
10.0%
5.0%
0.0%
AK
SD
NV
NM
MD
AL
MI
VA
WA
FL
Top 10 Employee State



The percentage of Defaulters is found in 60 months as compared to 36 months – Loan Term

FINAL INFERENCE AND SUGGESTIONS

Inference-I

- **Verification Status**: Verified loan are showing higher charged off rate.
- **High-Risk Loan Purposes**: Debt consolidation and credit card loans exhibit a higher percentage of defaults, indicating that borrowers seeking these loans might be in a financially vulnerable position.
- **Loan Amount and Interest Rate:** Most defaults occur for loan amounts between $5,600 and $16,500, particularly in the $5,000 to $10,000 range. These loan amounts often carry high to very high-interest rates, increasing the likelihood of default.
- **High Interest Rate**: High interest rate generally have high defaulting rates.
- **Grade and Homeownership**: Loan grade is negatively correlated with defaulting, implying that lower-grade loans are riskier. Additionally, loan applicants who do not own a house have a higher defaulting rate.

  1) Low grade loan had a much higher rate of defaulting during 2008 financial crisis. So, keep watch during when market is in instability

FINAL INFERENCE AND SUGGESTIONS

Inference - II

- **Employment Tenure**: Unexpectedly, customers with longer employment tenures have a higher defaulting rate. This further investigated using pair plot and details mentioned in below point.

  1) Loan Amount and Employment Tenure Interaction: As per pair plot, customer with higher amount of loan.Higher amount is linked to higher interest. All these factor result in higher defaulting

  2) Confounding Factors for employee tenure: It's possible that other confounding factors are at play, such as:

  3) Age: Longer-tenured employees are likely older and might have different financial priorities or life events that impact their ability to repay (e.g., children's education, medical expenses).

  4) Complacency: Lenders might be more lenient with underwriting standards for longer-tenured employees, assuming lower risk, which could lead to approving some riskier loans.

- **Bankruptcies and DTI**: A higher count of bankruptcies increases the chance of defaulting. High DTI is also a cause of defaulting, and it is negatively correlated with annual income, suggesting that high DTI often implies lower annual income.

- **Installment and Loan Term**: Installment and loan amount have a high positive correlation. Large-term loans with high-interest rates increase the likelihood of defaulting.

FINAL INFERENCE AND SUGGESTIONS

Suggestion - I

- Implement risk-based pricing, adjusting interest rates based on borrower risk profiles.
- Continuously monitor the relationship between interest rates and defaults to optimize pricing strategies.
- Perform thorough background checks for bankruptcies and consider stricter criteria for applicants with a history of bankruptcies.
- Enforce stricter DTI limits, especially for higher loan amounts or longer loan terms.
- Encourage shorter loan terms, especially for higher-risk borrowers, to reduce overall interest paid and the risk of default.
- For customer with long employment tenure, following factor need to taken into consideration

  - we need to check DTI value (to see whether existing financial burden is there).

  - Make sure that proper verification and risk analysis is done and no Complacency is done due to tenure

https://github.com/Shamilyjennymoor/Case-Study