



# Lending Club: Case Study

Team:

1. Megha CR
2. G Mothy

# Table of Content



1. Problem Statement
2. Data Understanding
3. Data Analysis
4. Driver Columns
5. Conclusions

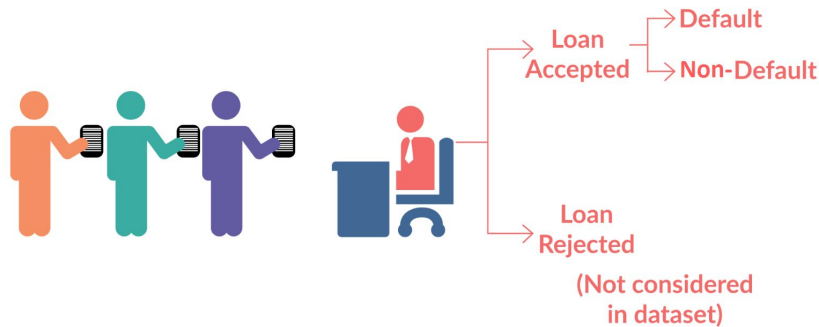
---

# Problem Statement

# Introduction

Solving this assignment will give you an idea about how real business problems are solved using EDA. In this case study, apart from applying the techniques you have learnt in EDA, you will also develop a basic understanding of risk analytics in banking and financial services and understand how data is used to minimise the risk of losing money while lending to customers.

## LOAN DATASET



# Business Understanding



You work for a consumer finance company which specialises in lending various types of loans to urban customers. When the company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile. Two types of risks are associated with the bank's decision:

- If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company
- If the applicant is not likely to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company

## Business Objectives

The company wants to understand the **driving factors (or driver variables) behind loan default**, i.e. the variables which are strong indicators of default. The company can utilise this knowledge for its portfolio and risk assessment.

---

# Data Understanding

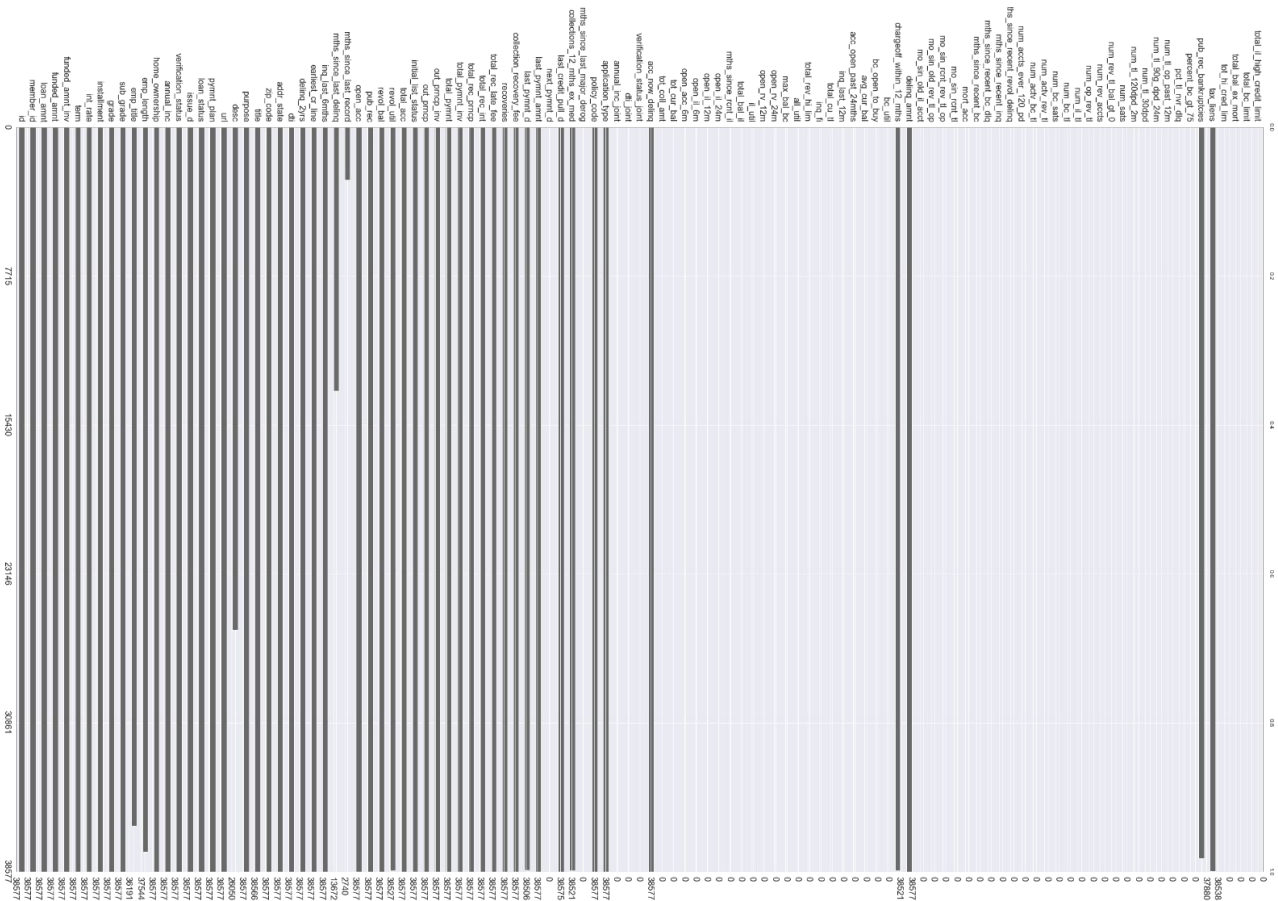
# Significant Features



Features considered for analysis post data cleaning :

1. loan\_amnt → loan amount applied for the loan by borrower
2. term → number of payments on the loan (in months)
3. int\_rate → Interest Rate on the loan
4. installment → monthly payment owed by the borrower if the loan originates
5. emp\_length → employment length in years. Possible values are between 0 and 10, higher is 10+
6. dti → borrower's total monthly debt payments to monthly income ratio
7. grade → loan grade assigned by lending club
8. home\_ownership → home ownership status provided by the borrower during registration
9. verification\_status → Indicates if whether income was verified , not verified, or only source was verified
10. purpose → category for which the loan was requested
11. addr\_state → state provided by the borrower in the loan application
12. loan\_status → status of the loan (target variable)

# Missing Values



## Observations:

There were 57 features with more than 60% null values. These features were removed to proceed further for the analysis



# Correlation Plot

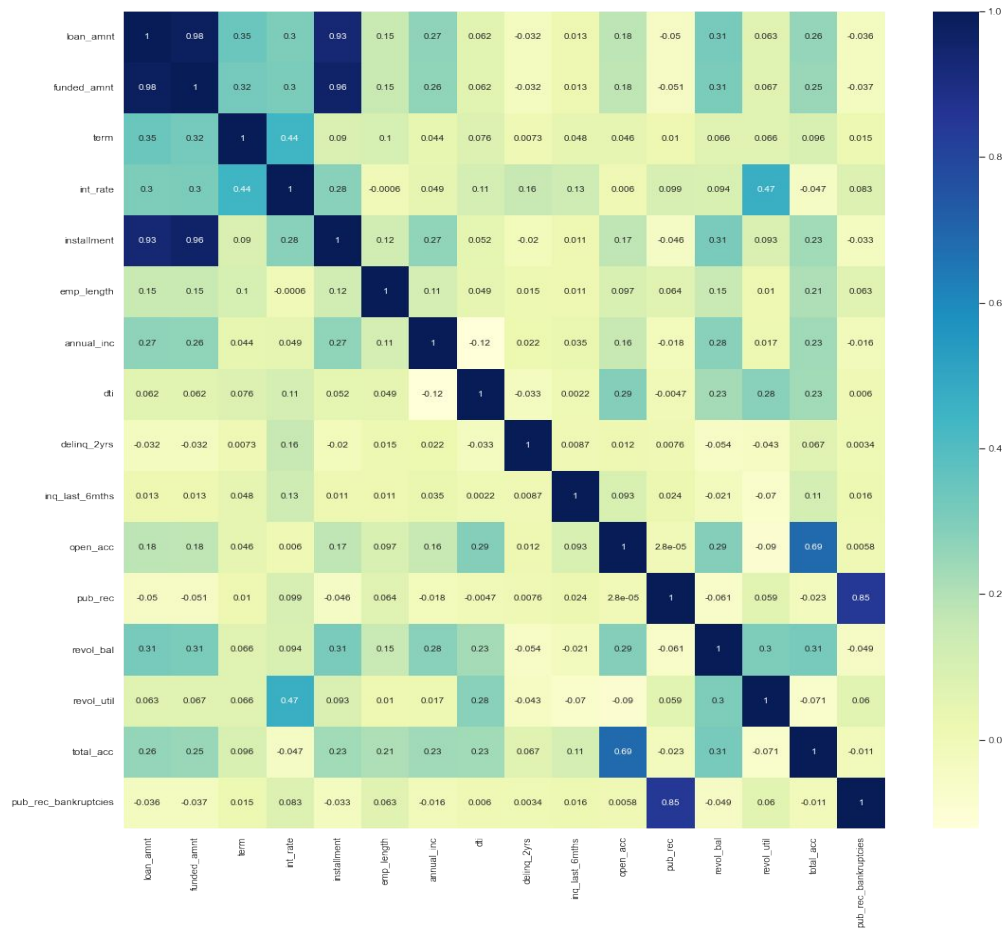


## Observations:

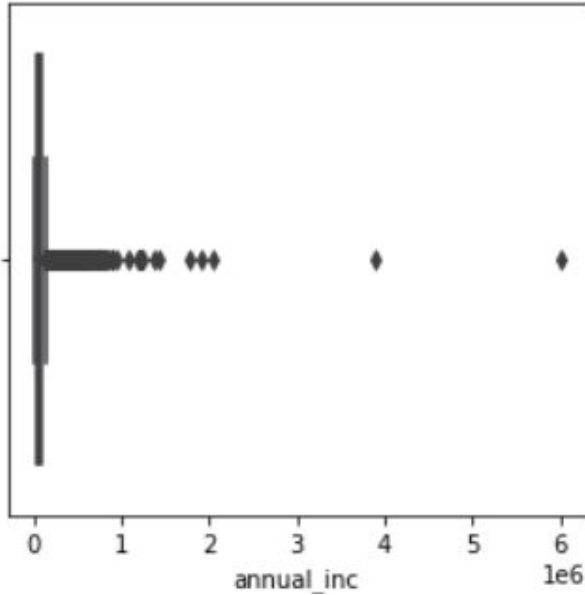
There are features which are highly correlated.

- loan\_amnt & funded\_amt are highly correlated.
- pub\_rec\_bankruptcies & pub\_rec are highly correlated.
- open\_acc & total\_acc

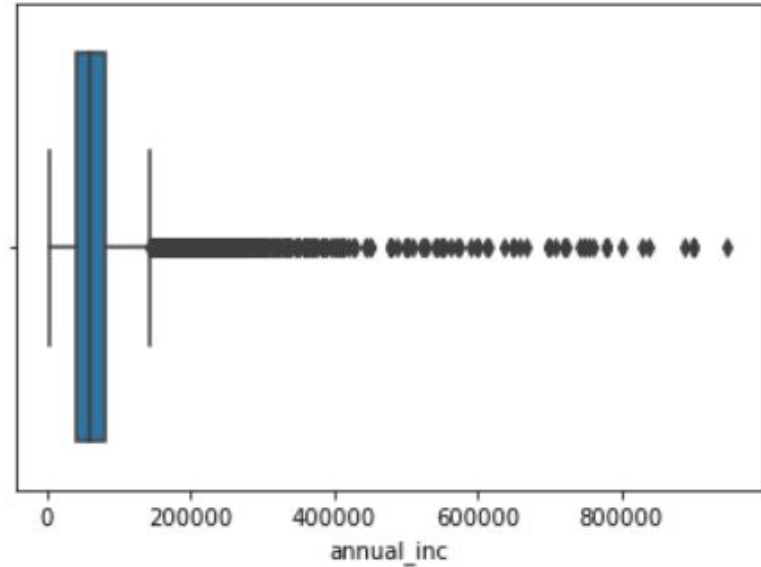
Only one of the two features from highly correlated pair are considered for further analysis



# Outlier Treatment



Before Outlier Treatment



After Outlier Treatment

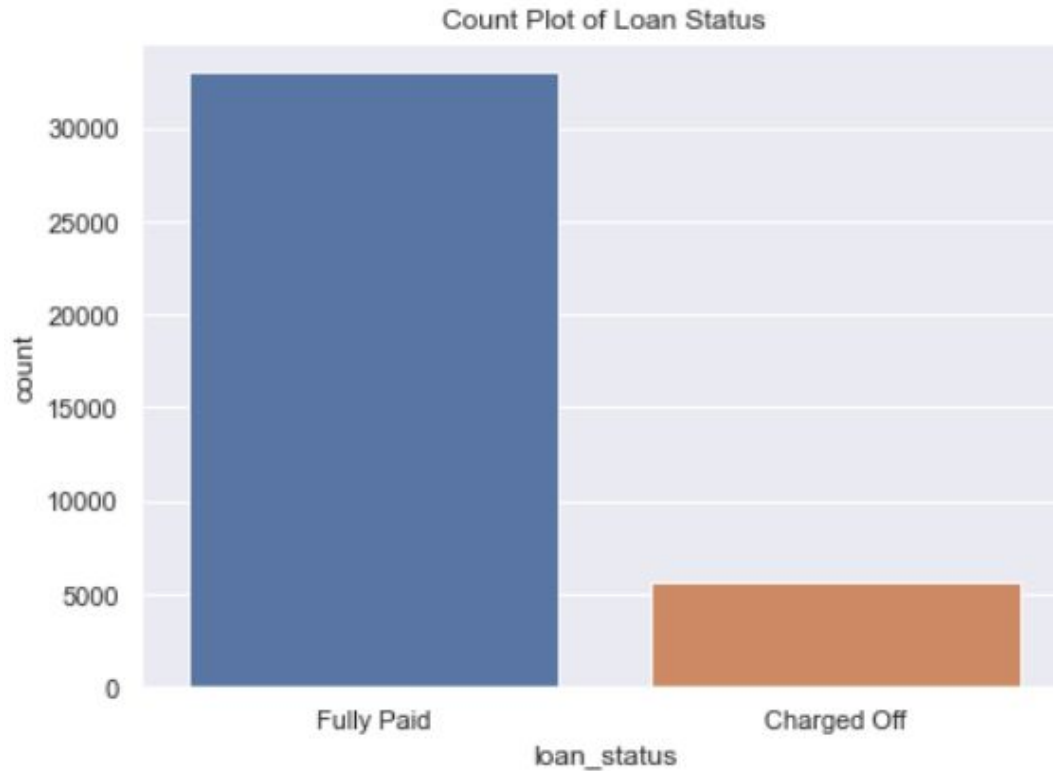
## Observations:

1. Annual Income has few extreme outliers, so outliers were removed.

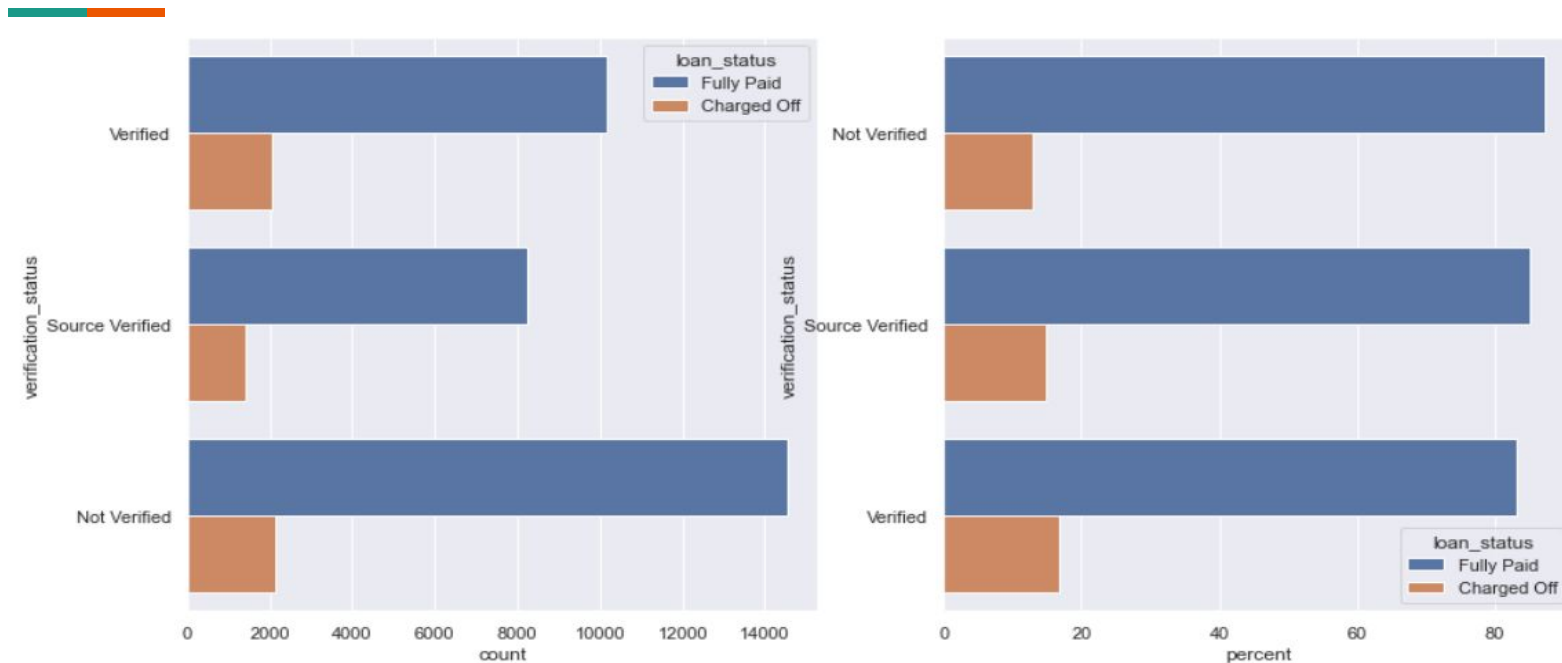
---

# Data Analysis

# Understanding Loan Status (Target column)



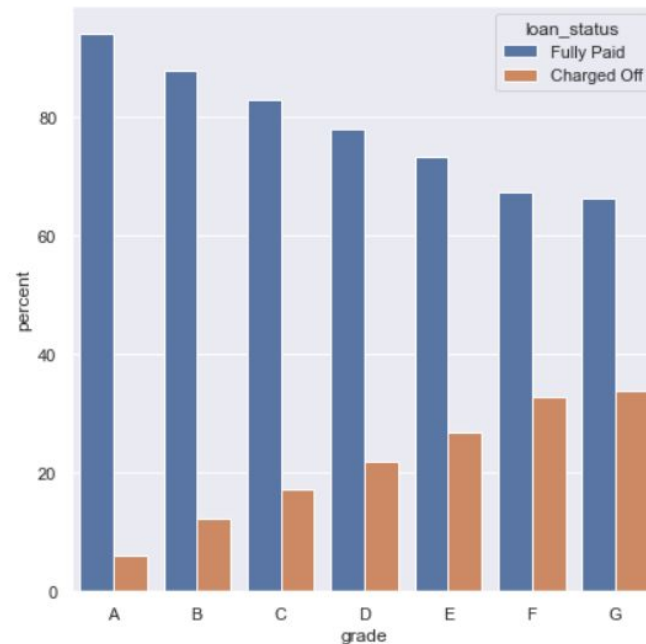
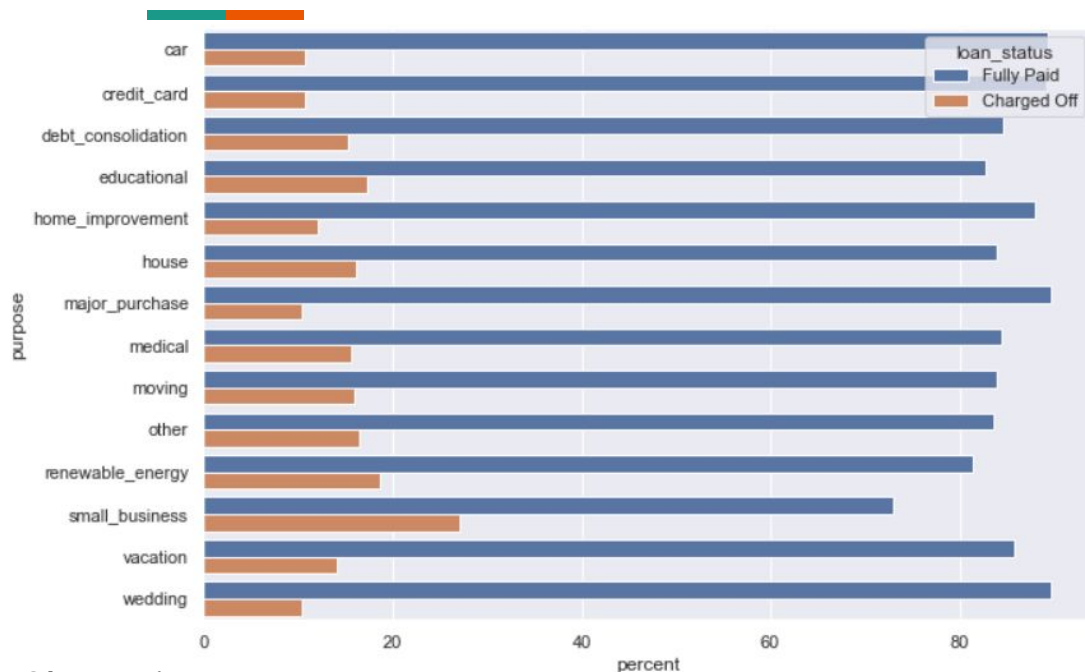
# Verification Status



## Observations:

1. Contradictory to belief, here the Verified applicants tend to default more. Extra care should be taken in the verification process as verification is mostly done for higher amount loans which have higher default risk.

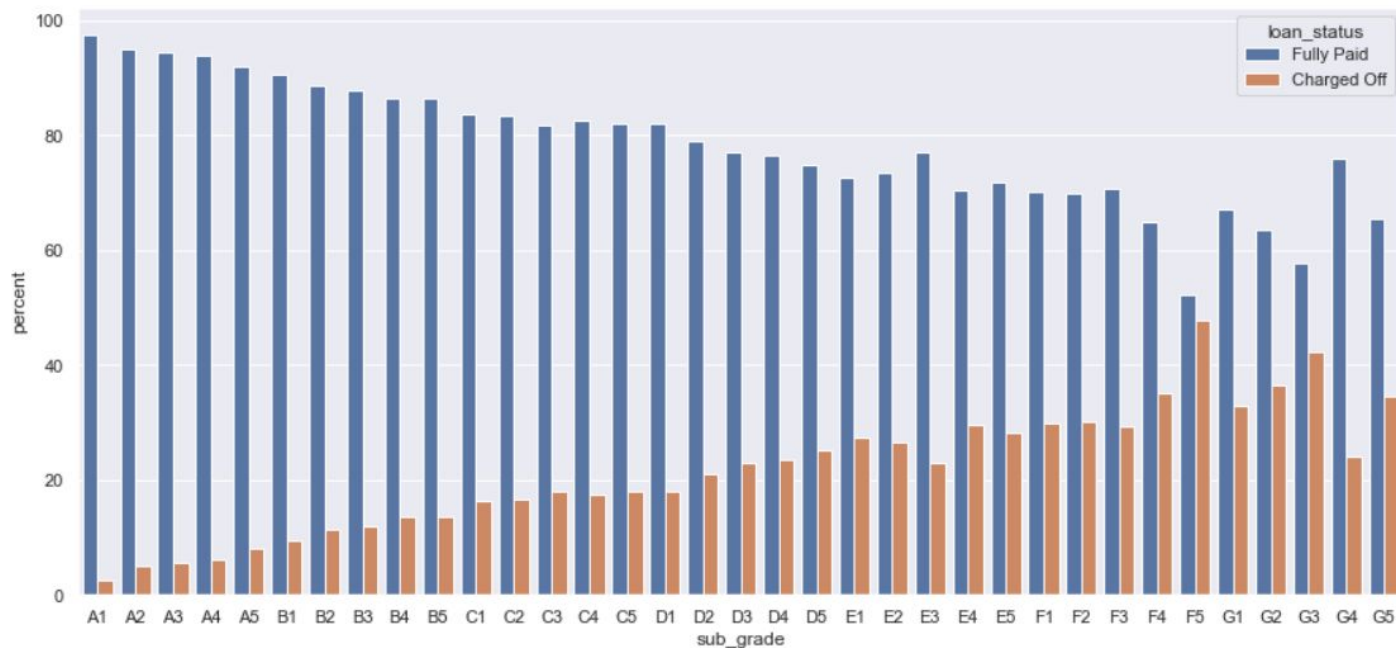
# Purpose of Loan & Grade of the Applicant



## Observations:

1. Loans with purpose of Small business, renewable energy or educational have higher default chances than other purposes. But by significance, loans for small business should be considered carefully
2. Loans having grade E, F or G tend to default at higher rate. The trend of defaulting is increasing with the grade level.

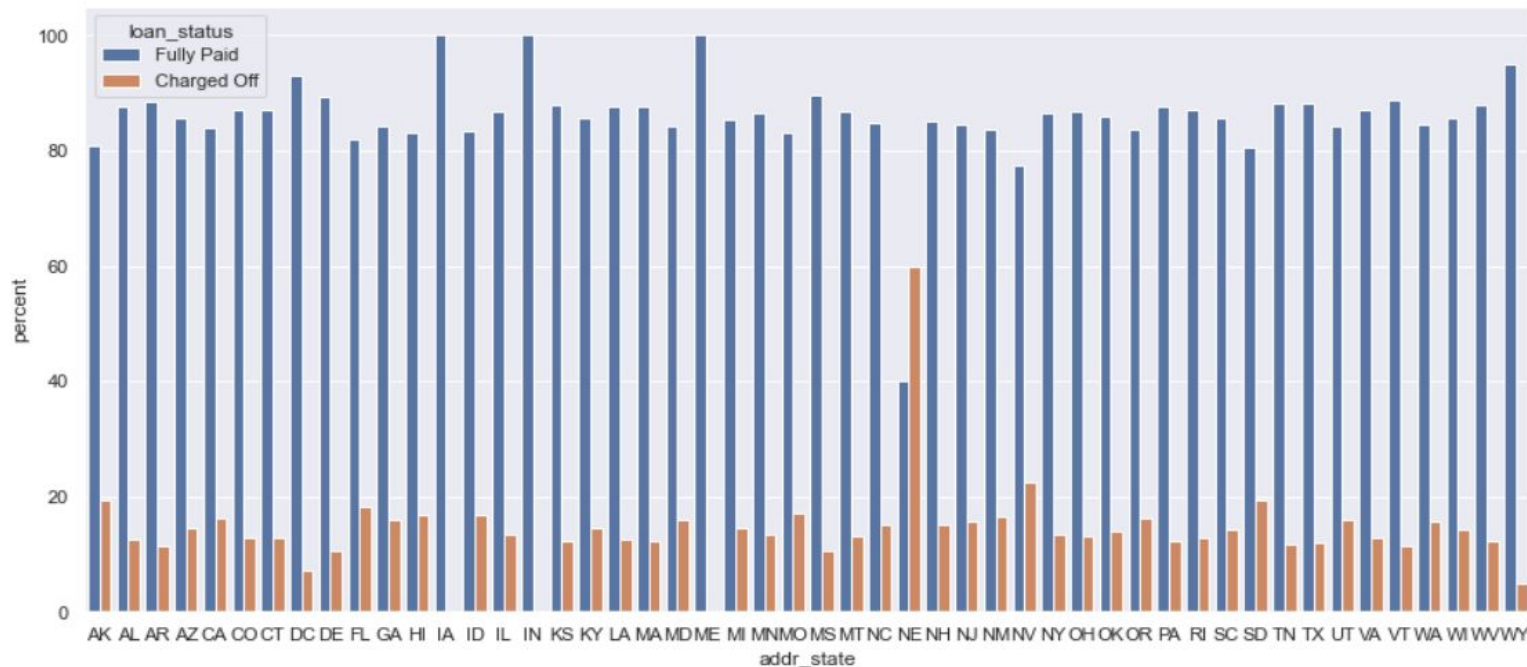
# Sub Grade



## Observations:

As grade increases, chances of loan default also increase. Need to be careful with loans graded with F5,G3,G5.

# Address State

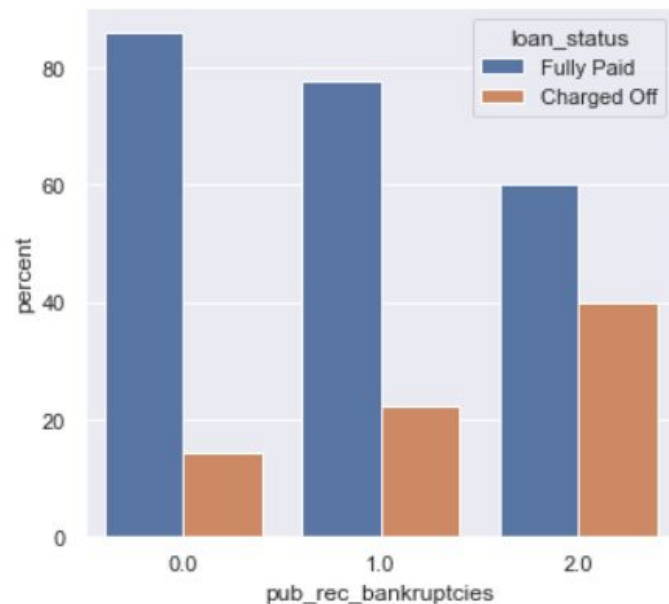
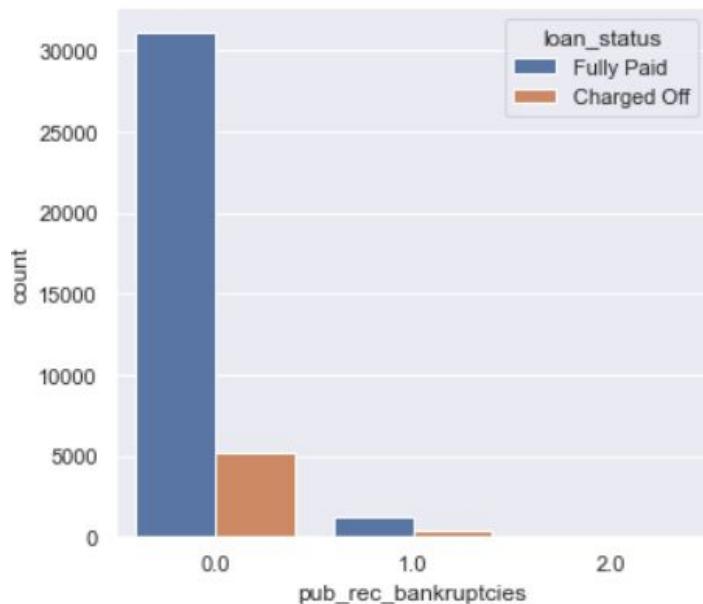


## Observations:

- 1.State 'NE' loans are very suspicious. It's default ratio(60%) is higher than fully-paid ratio(40%).
2. IA,IN and ME have got 0 defaults.



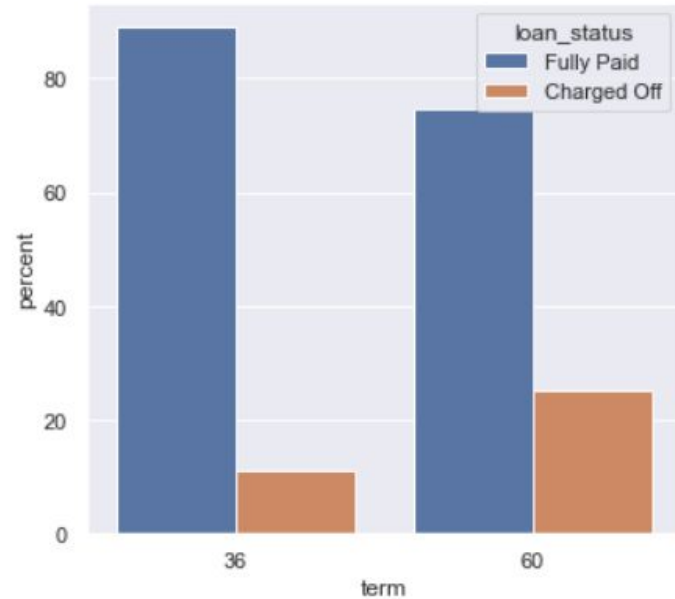
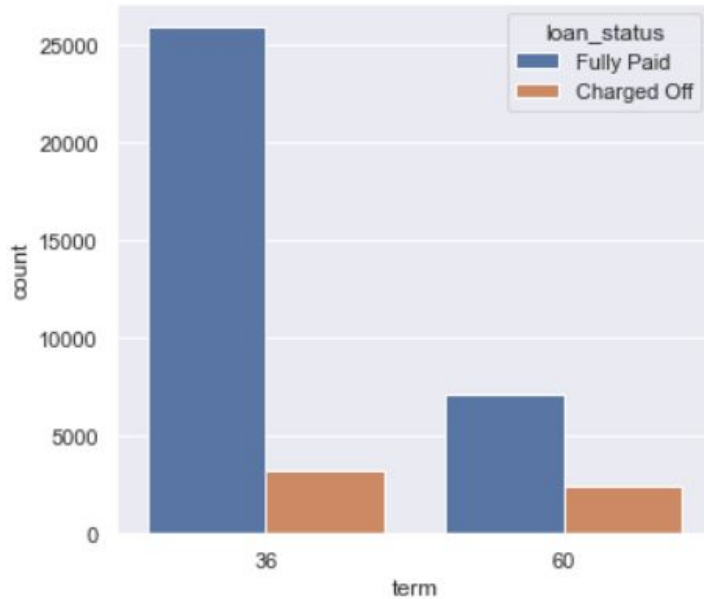
# Public Record Bankruptcies



## Observations:

1. Higher the number of public record bankruptcies, higher chances of loan getting defaulted.

# Term



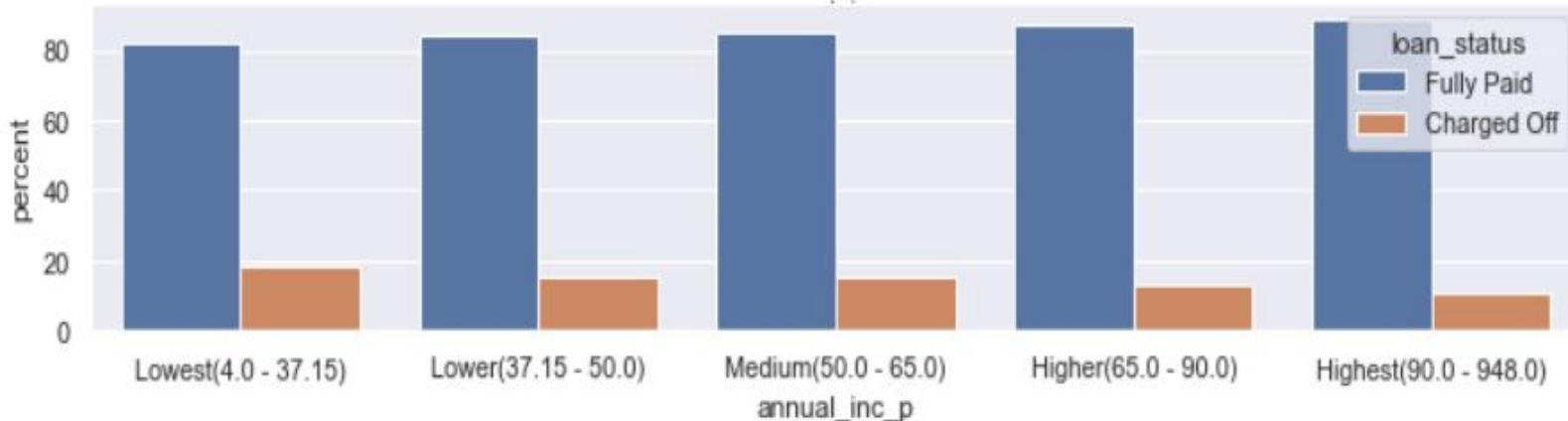
## Observations:

Significantly higher chance of defaulting with term loan of 60 months than 36 months.

# Annual Income



	annual_inc_p	Charged off %
0	Lowest(4.0 - 37.15)	0.183197
1	Lower(37.15 - 50.0)	0.156865
2	Medium(50.0 - 65.0)	0.149629
3	Higher(65.0 - 90.0)	0.129651
4	Highest(90.0 - 948.0)	0.109356



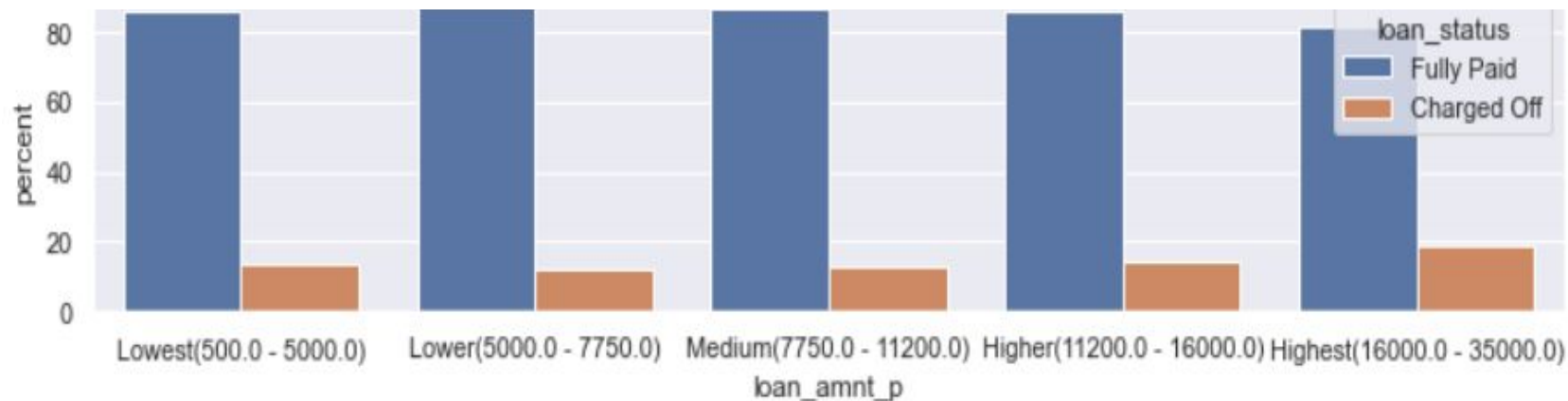
## Observations:

1. Lower income applicants tend to default higher percentage.

# Loan amount



	loan_amnt_p	Charged off %
0	Highest(16000.0 - 35000.0)	0.187673
1	Higher(11200.0 - 16000.0)	0.145295
2	Lowest(500.0 - 5000.0)	0.138769
3	Medium(7750.0 - 11200.0)	0.131663
4	Lower(5000.0 - 7750.0)	0.123582



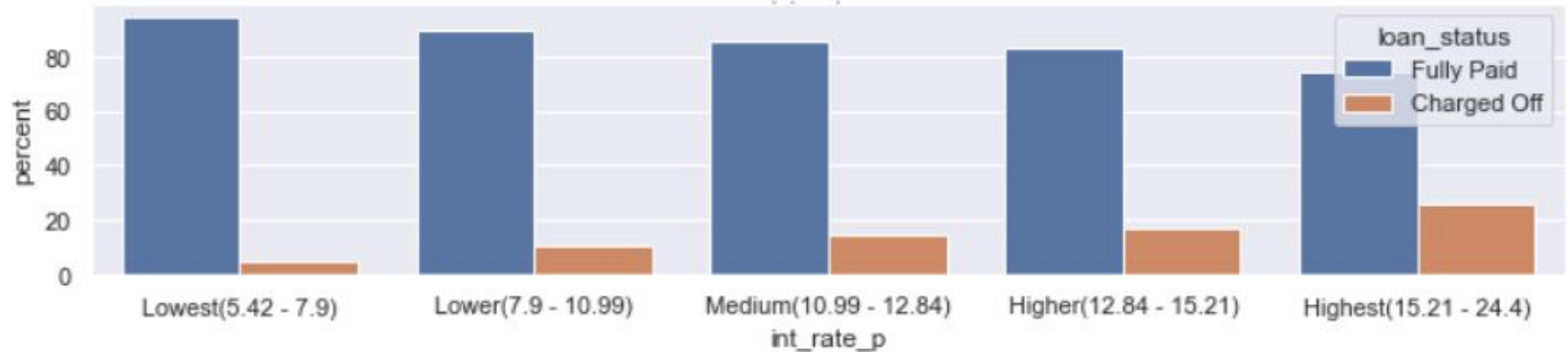
## Observations:

The chances of default from higher loan amount applicants is higher.

# Interest Rate



	int_rate_p	Charged off %
0	Highest(15.21 - 24.4)	0.262150
1	Higher(12.84 - 15.21)	0.171578
2	Medium(10.99 - 12.84)	0.147038
3	Lower(7.9 - 10.99)	0.102888
4	Lowest(5.42 - 7.9)	0.053585



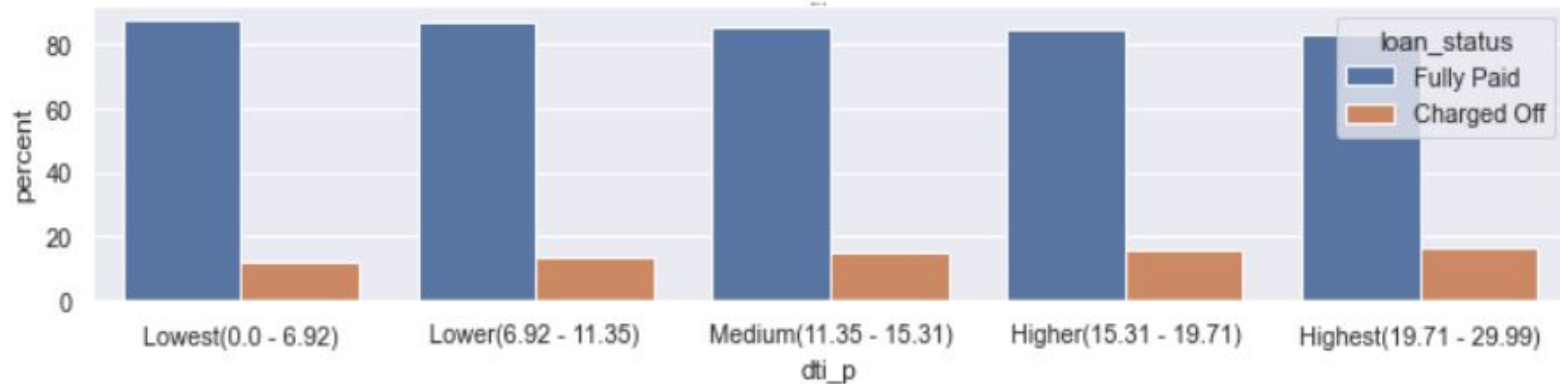
## Observations:

The percentage of loan being defaulted are more for higher interest rates.

# DTI (Debt to Income Ratio)



	dti_p	Charged off %
0	Highest(19.71 - 29.99)	0.168745
1	Higher(15.31 - 19.71)	0.157908
2	Medium(11.35 - 15.31)	0.147609
3	Lower(6.92 - 11.35)	0.132539
4	Lowest(0.0 - 6.92)	0.122716



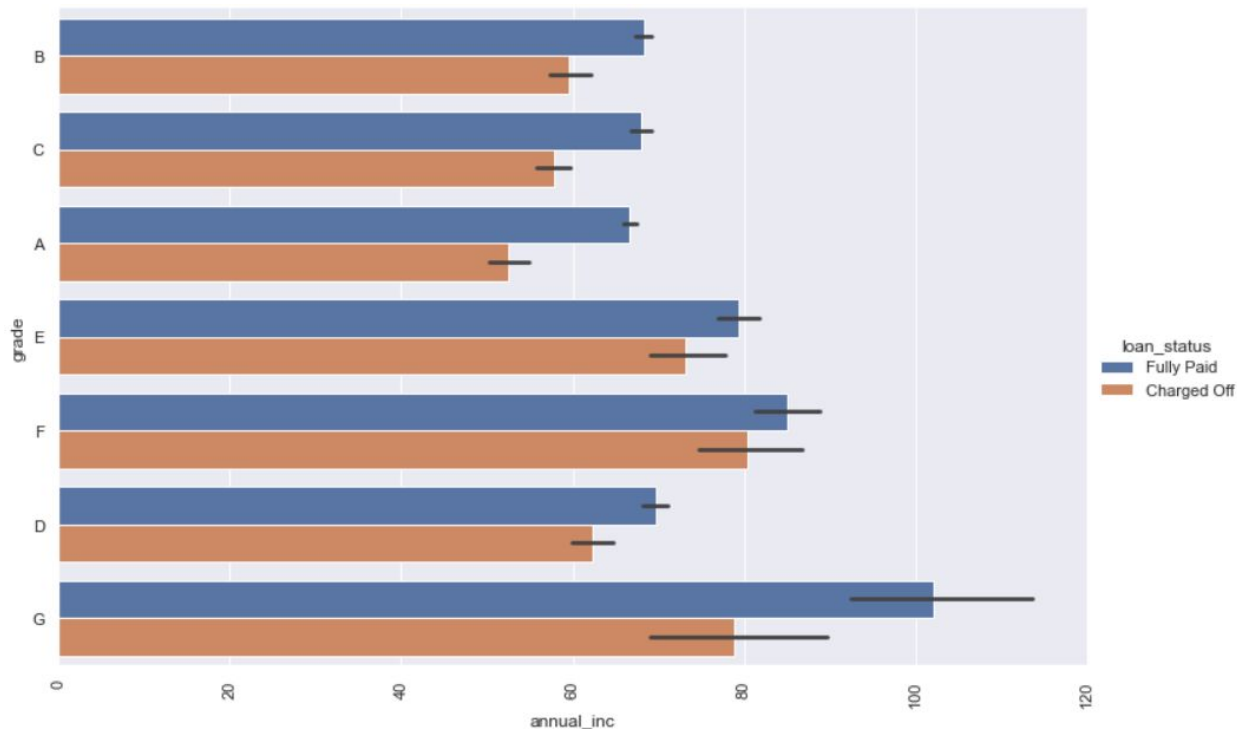
## Observations:

1. Higher dti applicants tend to have higher loan defaults

# Grade vs Annual Income

## Observations:

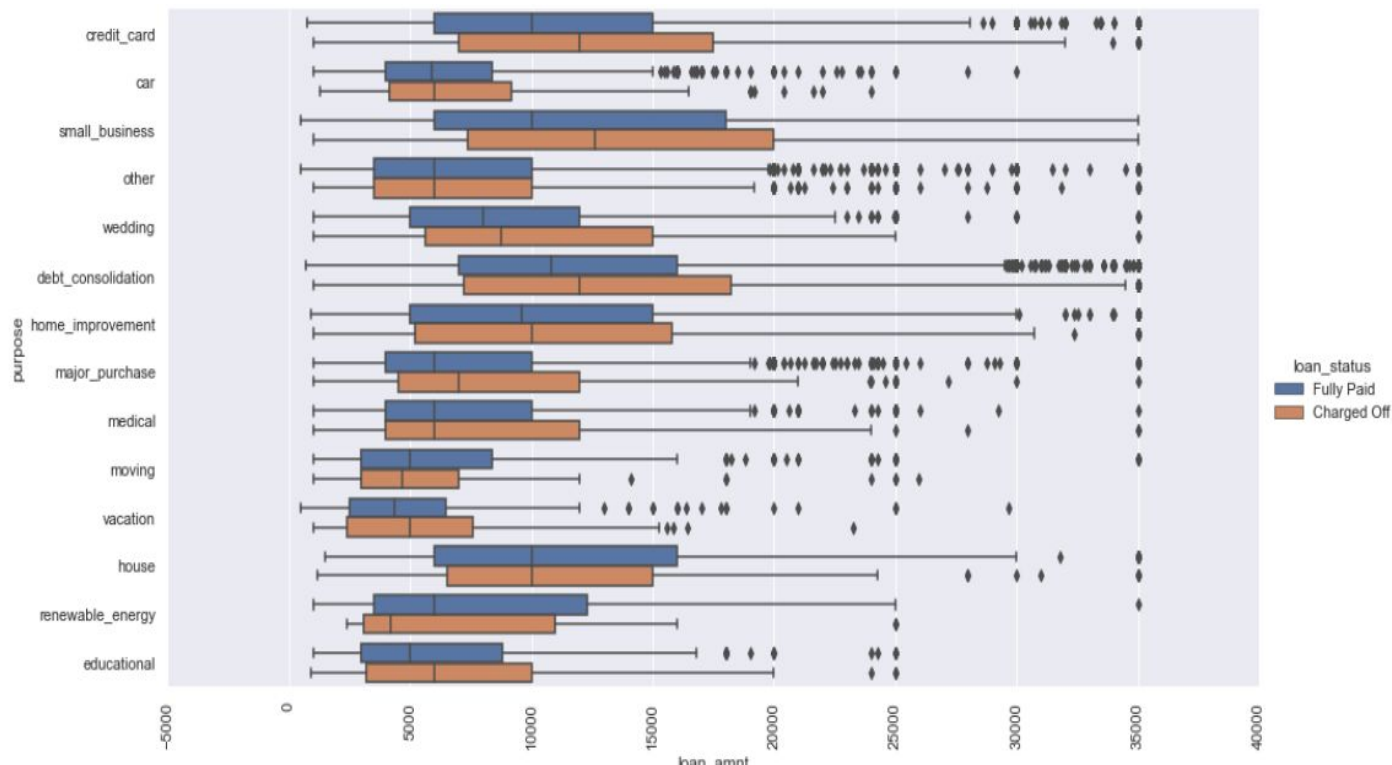
1. Higher incomes have grade in increasing manner, Lowest income being A, highest G.
2. Grades E, F and G have defaulted more.
3. Though the income is high, the loan approval/rejection should be evaluated by considering other factors like grade too.



# Purpose vs Loan amount

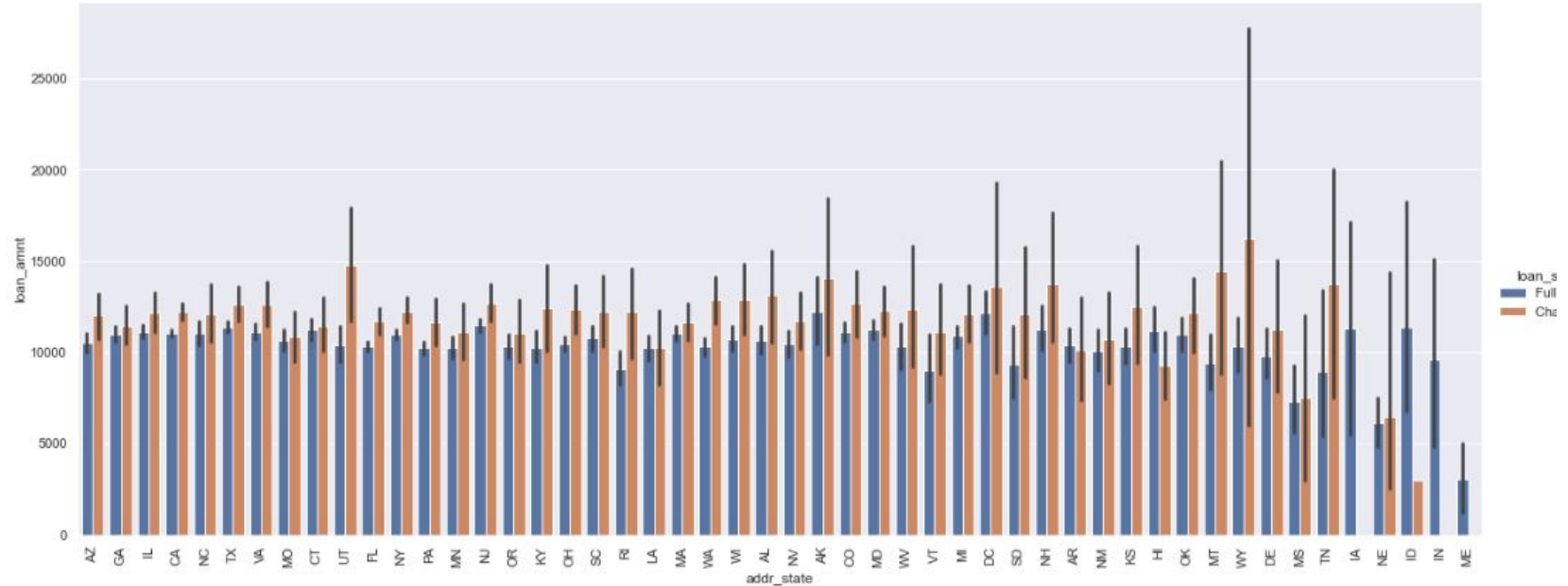
## Observations:

High loans were taken for small businesses followed by debt consolidation and loans against credit cards, which have higher tendency of defaulting.





# Address State vs Loan Amount



## Observations:

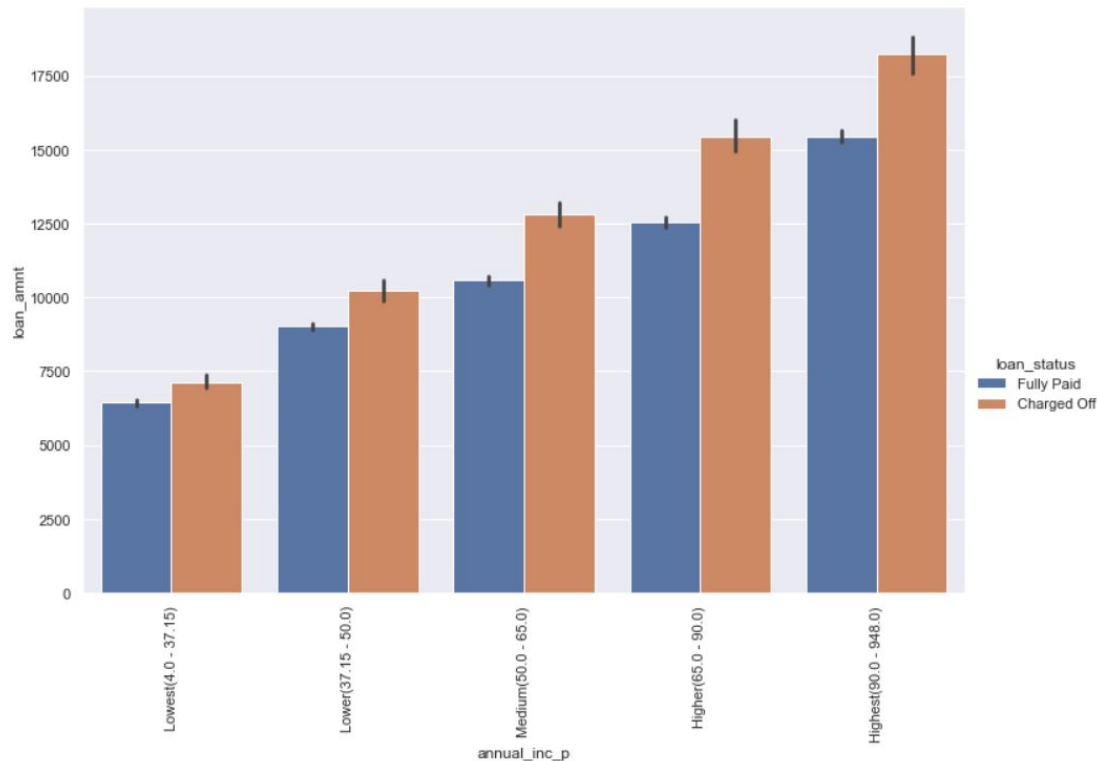
1. State 'WY' with higher loan amount has seen significantly high default tendency.

# Annual Income range vs loan Amount



## Observations:

Across all income groups, higher the loan amount higher the default rates.

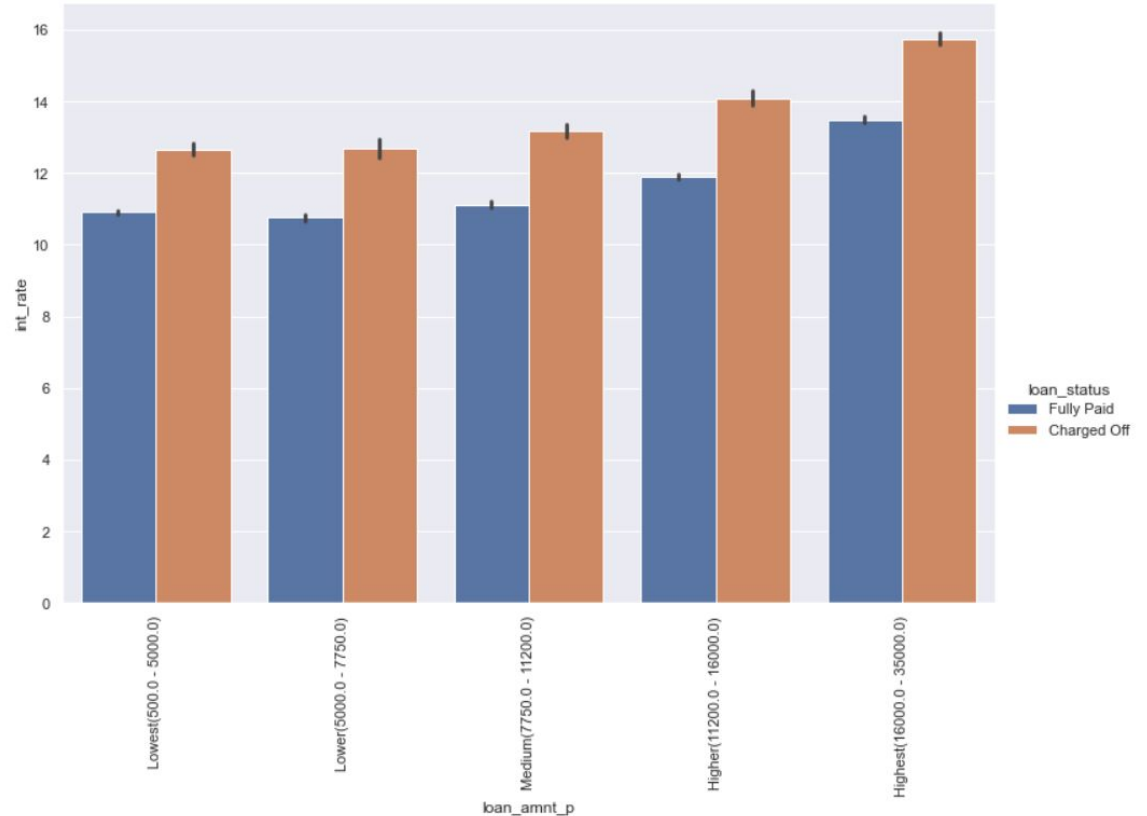


# Loan amount range vs Interest Rate



## Observations:

Across all loan amount range, higher the interest rates higher the default rates.



---

# Driver Columns

# Driver columns



1. Grade
2. Term
3. Interest Rate
4. Loan Amount
5. Purpose of the Loan
6. Annual Income
7. Verification Status

---

# Conclusions

# Key Observations



1. **State:** States IA, IN and ME have 0 default cases and hence no possible risk foreseen. Whereas, state NE has higher default ratio than loan completion. And state WE has higher default cases with higher loan amounts.
2. **Grade/Subgrade:** Customers with high annual income have higher loan amount approvals and are assigned higher grades. And as grades go higher, default cases increase exponentially. Subgrades F5, G3, G5 have comparatively defaulted more.
3. **Verification status:** Most verified loans are the ones with higher loan amount. But contradictory to belief, verified loans have defaulted more. This might mean that some customers might have submitted fake documents to get high loan amounts approved.
4. **Purpose:** Small business, Renewable energy have high default ratio. With higher loan amounts, small businesses have higher tendency of defaulting, followed by debt consolidation and credit cards. So, loans for small business seem to be risky and needs to be considered with additional verification.



5. **Interest rate:** Higher interest rate of more than 15% have more chances of default. Across all loan amount range, higher the interest rates higher the default rates.
6. **Term:** Loans of term 60 months have seen significantly higher defaults than 36 months.
7. **annual income:** Across all income groups, higher the loan amount higher the default rates. Even though the income is high, the loan approval/rejection should be evaluated by considering other factors like grade,etc.





**Thank You**