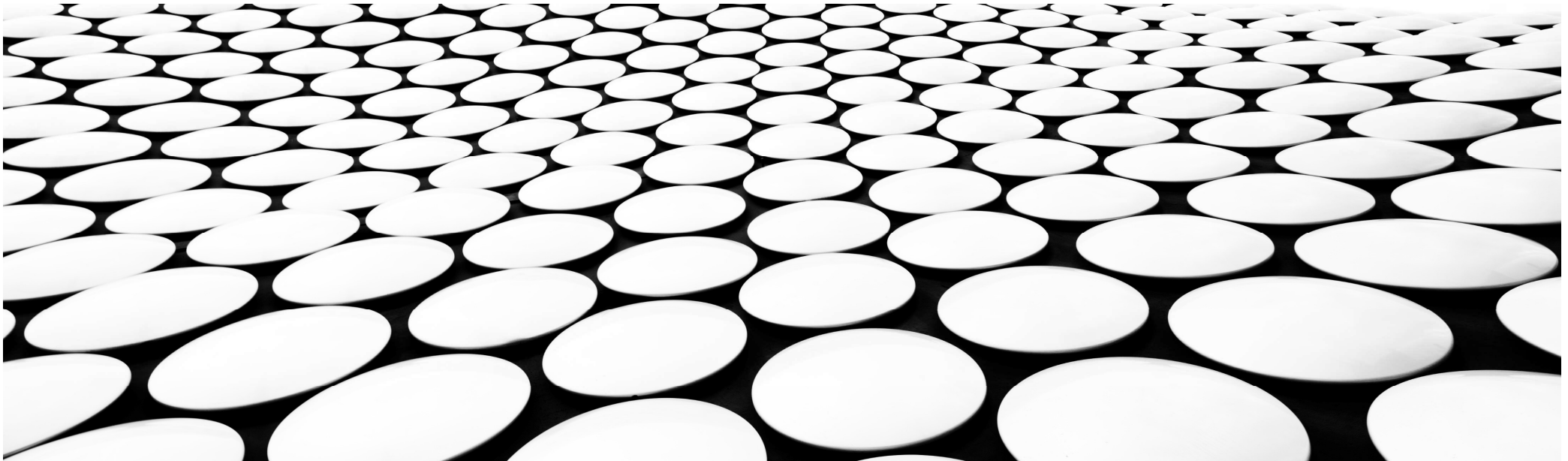


---

# LENDING CLUB CASE STUDY

SUJITH KUMAR (APR – 2022 BATCH)

DATE OF SUBMISSION: 08 JUNE 2022



## PURPOSE | DATA SET

[Lending Club](#) is a full-spectrum fintech marketplace bank which through technology driven platform offers customers ability to borrow money to meet variety of purposes.

Purpose of this case study is to perform EDA analysis on the data set provided so as to determine the critical attributes which can help determine whether a loan application would be at risk of default or would bring in good business

### Initial observations of the data set:

- The provided data set has 39,717 rows and 111 columns
- There are several columns with NA values
- High correlation between loan\_amnt, funded\_amnt, funded\_amnt\_inv
- Int\_rate and revol\_util are of type object due to % sign
- Emp\_Title has over 28K unique values.

Correlation b/w amount columns

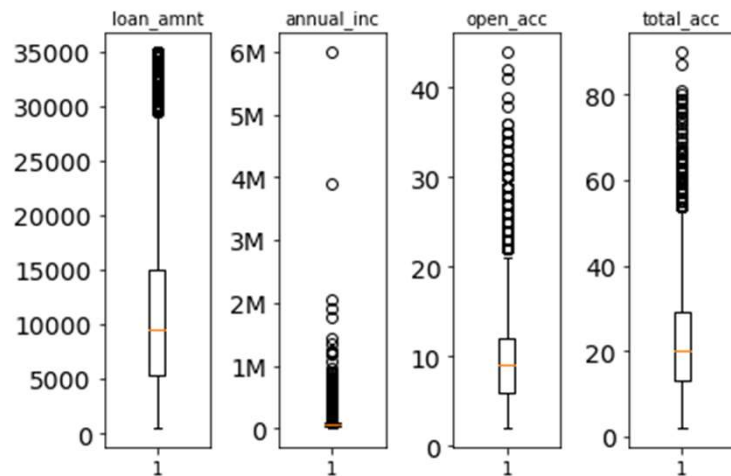


# CLEANSING DATA

Type of Cleansing	Description
Drop columns without data	54 columns from the data set had all values as null
Drop behavioural attributes	Few attributes in the data set are attributes specific to the behaviour of a customer during the tenure of the loan. This data is not available beforehand during the loan application and hence can be disregarded for the analysis. Ex: delinq_2_yrs, earliest_cr_line etc
Drop other attributes which don't aid in analysis	<ol style="list-style-type: none"><li>1. Attributes such as id, url which are unique and don't aid in analysis</li><li>2. Attributes such as tax_liens which have one unique value (0) and rest null</li><li>3. Attributes such as mths_since_last_record where data is mostly null</li><li>4. Columns with high correlation – funded_amnt and funded_amnt_inv are highly correlated to loan_amnt and hence can be dropped</li><li>5. Other attributes such as zip_code, title which are either masked or have too many variations</li></ol>
Cleanse for Data Type Issue	<ul style="list-style-type: none"><li>• Character '%' was observed in int_rate and revol_util</li><li>• The 50 rows where revol_util is null was dropped and the % character was removed to correct datatype</li></ul>
Rows excluded from further analysis	Loan_status=Current , are the ongoing loans and would not aid in analysis
Remove Outliers	<ul style="list-style-type: none"><li>• 397 rows(1%) had annual income which is way above the remaining 99% of the rows</li><li>• There were outliers in the total_acc (number of credit lines in borrowers credit file)</li></ul>

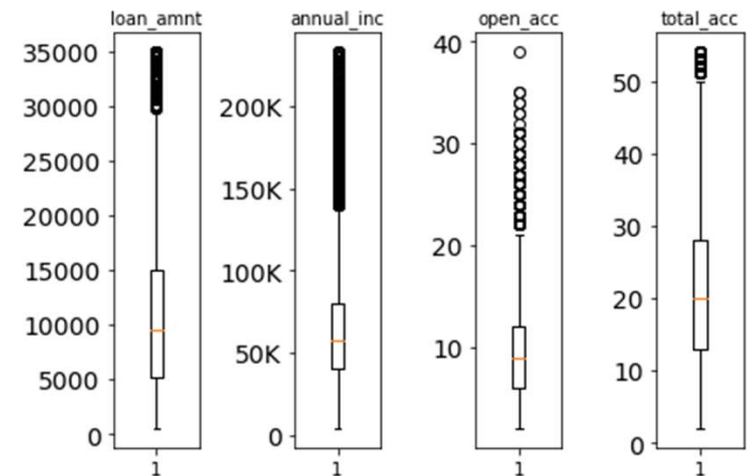
# OUTLIERS - ANALYSIS

Before outlier removal



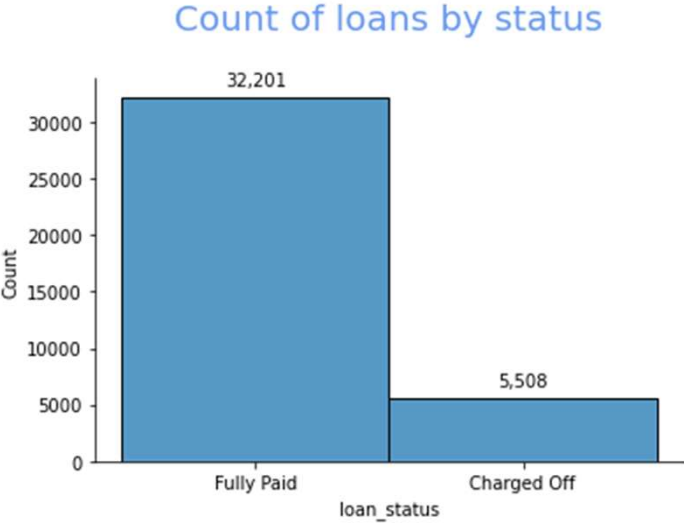
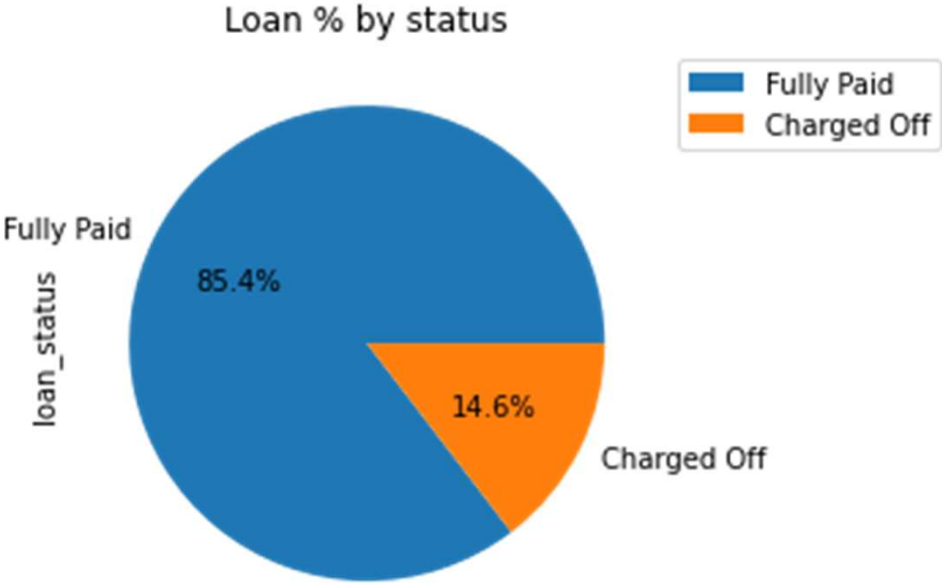
601 rows in dataset which fall above 99.5 percentile of the loan\_amnt i.e., 1.56 percent of rows  
389 rows in dataset which fall above 99 percentile of the annual\_inc i.e., 1.01 percent of rows  
265 rows in dataset which fall above 99.5 percentile of the open\_acc i.e., 0.69 percent of rows  
206 rows in dataset which fall above 99.5 percentile of the total\_acc i.e., 0.53 percent of rows

After outlier removal



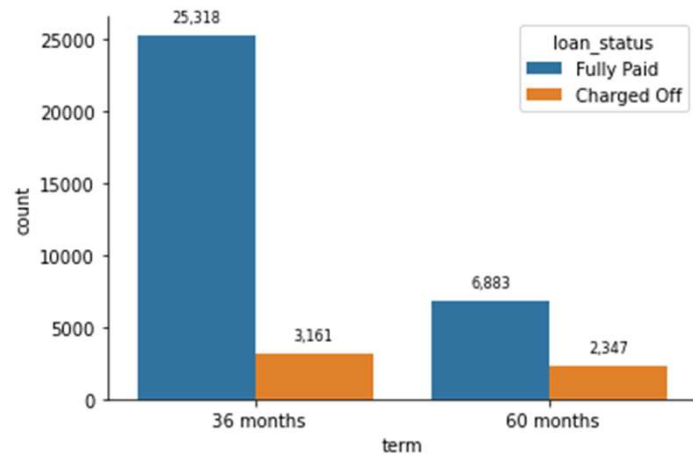
**37709** rows remain post exclusion of current ongoing loans and outliers

# LOAN STATUS - ANALYSIS

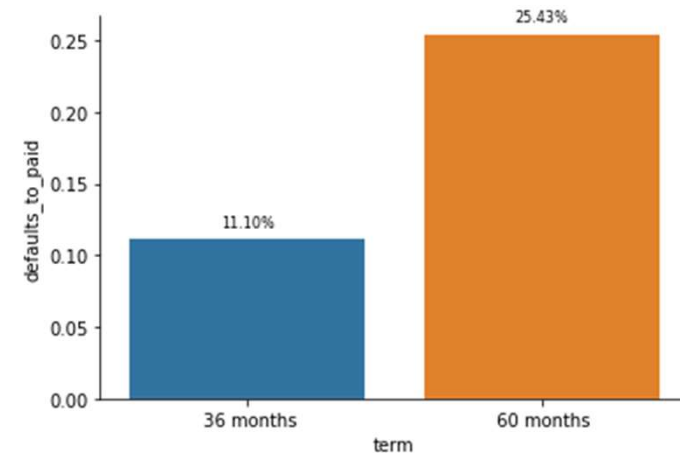


## TERM – DEFAULT VS PAID

Counts by term



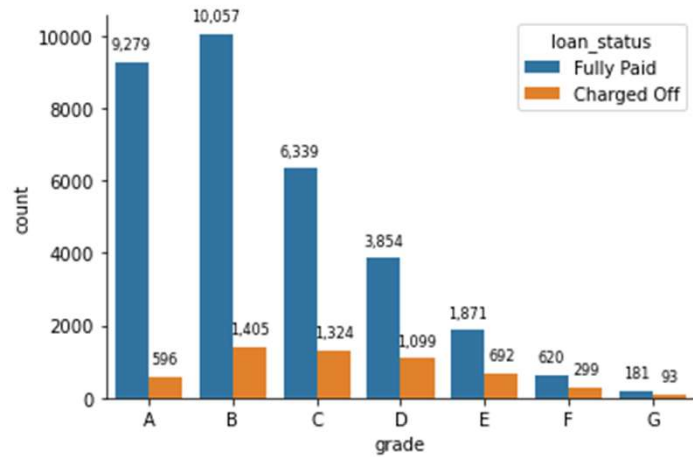
By term : Default Percent to Total



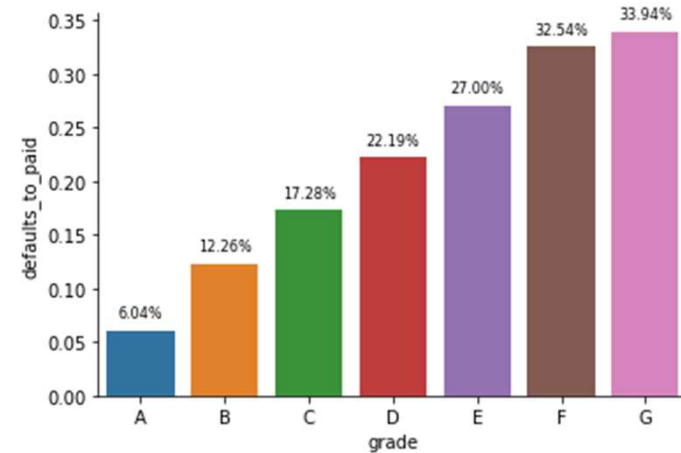
**Observation:** When compared to the total number of loans, loans taken for 60 months have a higher percent of defaults

## GRADE – DEFAULT VS PAID

Counts by grade



By grade : Default Percent to Total

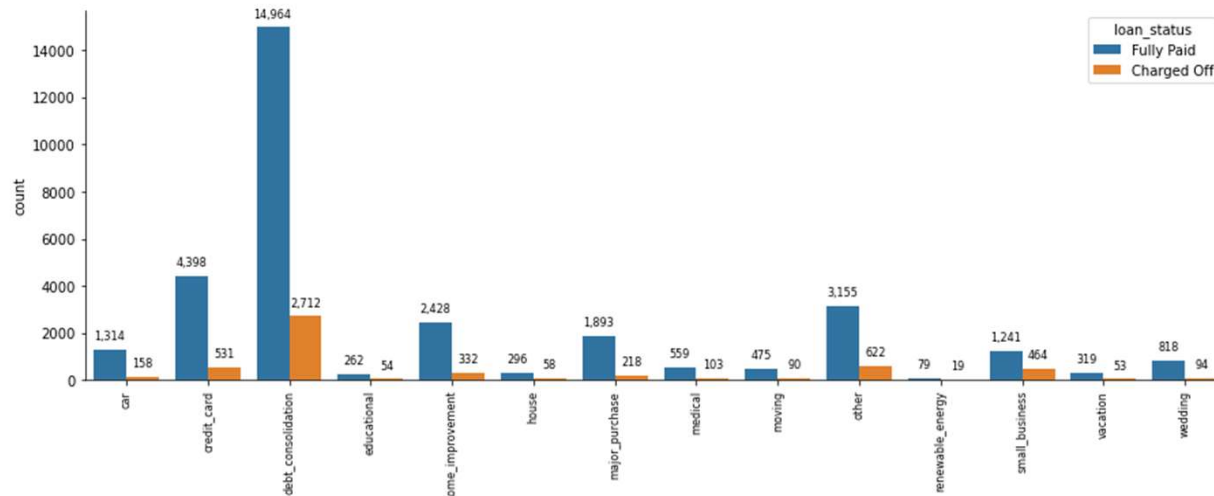


### Observation:

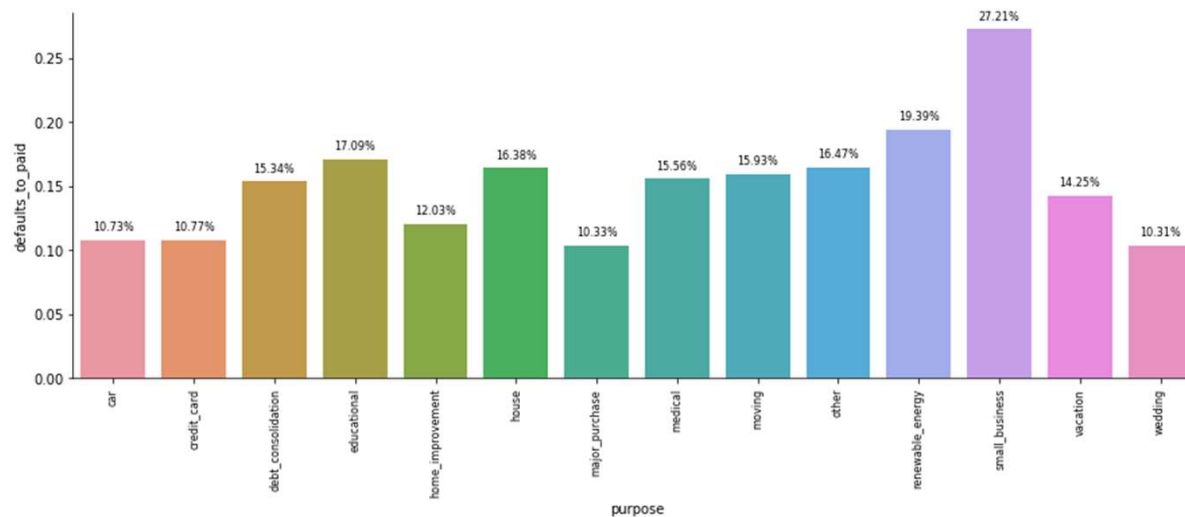
- 1) When compared to the total number of loans, loans classified as grades E, F, G have a higher risk of defaulting
- 2) Similar behaviour is observed when plotted against subgrade

# PURPOSE – DEFAULT VS PAID

Counts by purpose



By purpose : Default Percent to Total



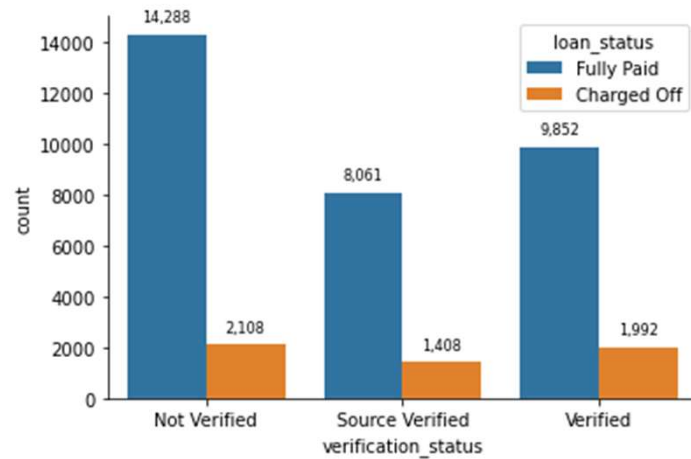
## Observation:

1. Maximum loan is taken for 'debt consolidation'. While 15% of loans taken for the purpose default, it is also a good business opportunity
2. Loans taken for small business have a higher percentage of defaulters and should be evaluated

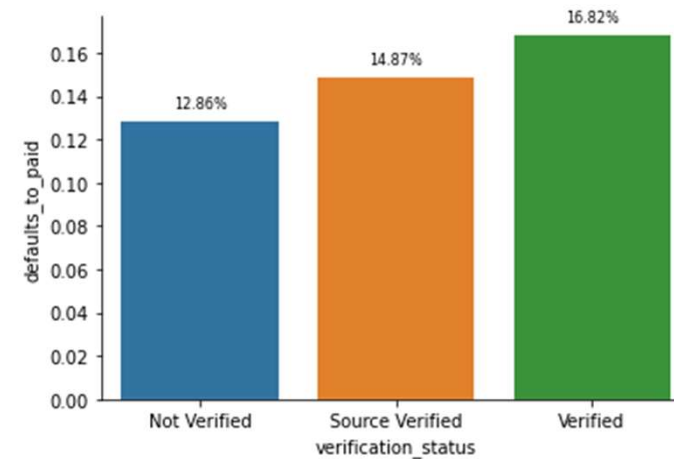


## VERIFICATION STATUS – DEFAULT VS PAID

Counts by verification\_status



By verification\_status : Default Percent to Total

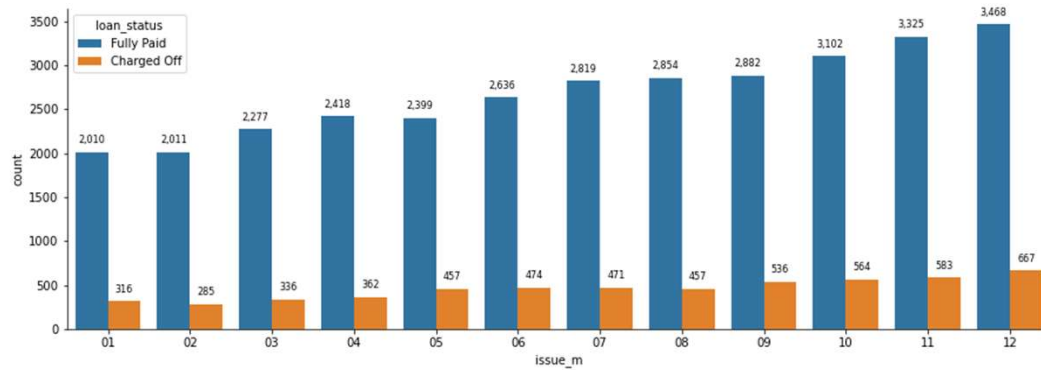


### Observation:

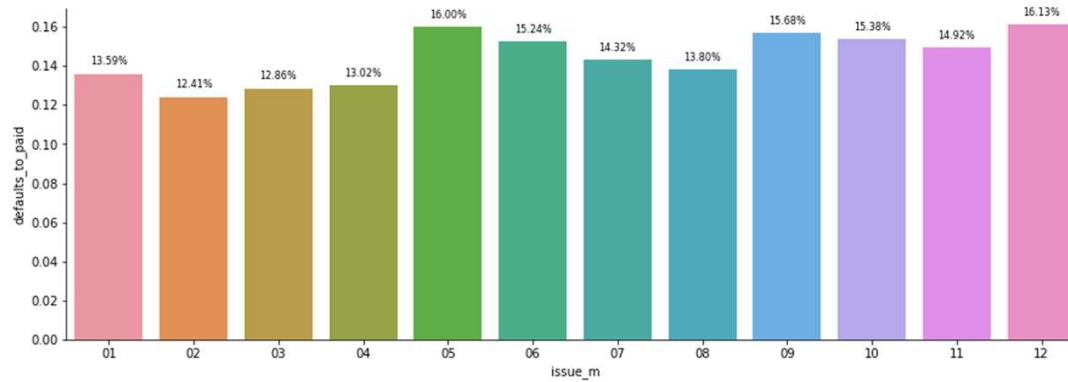
- 1) Maximum loans are disbursed without income verification
- 2) However, the default percent for loans where income is verified is less when compared with loans where income is verified

# ISSUED MONTH – DEFAULT VS PAID

Counts by issue\_m



By issue\_m : Default Percent to Total

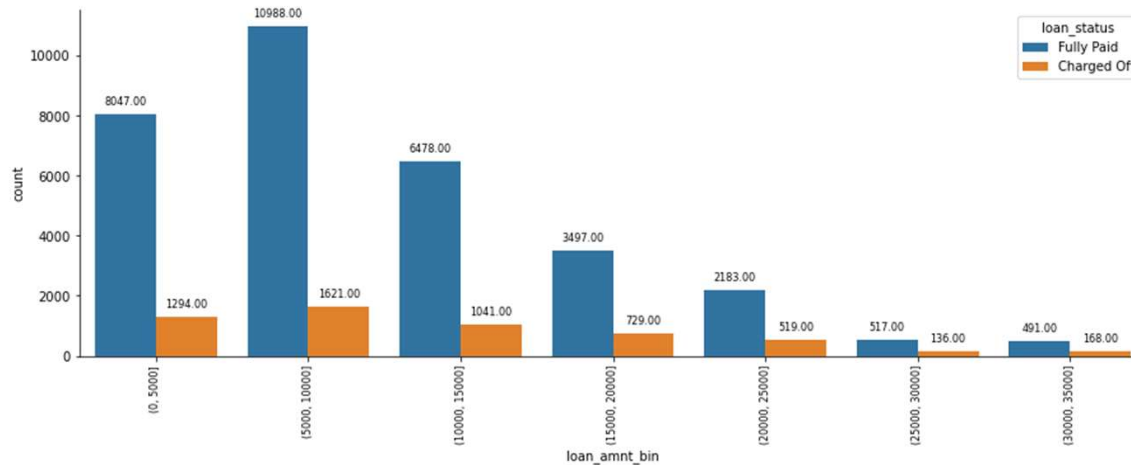


## Observation:

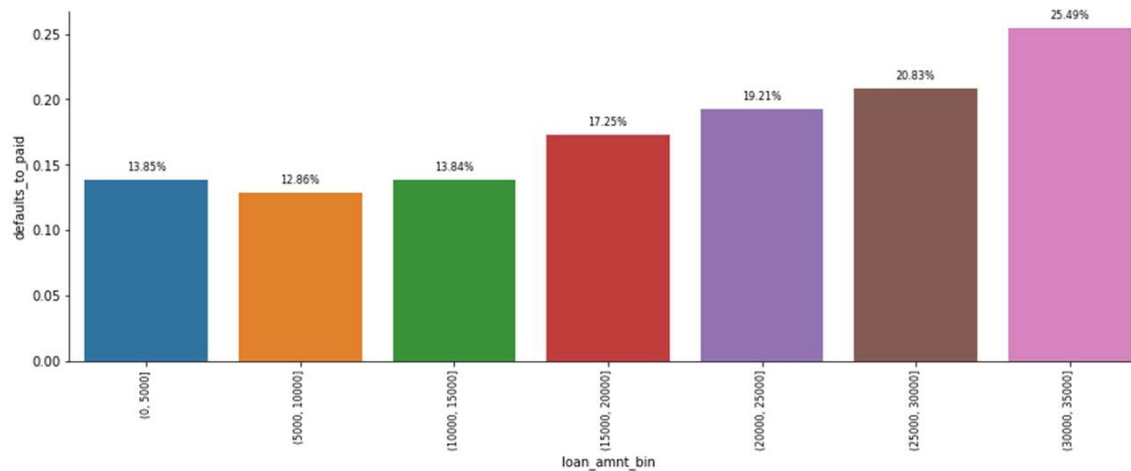
1. Maximum loans are availed towards the end of the year
2. Relatively higher defaults are observed during Financial Year close and holiday season

# LOAN AMOUNT – DEFAULT VS PAID

Counts by loan\_amnt\_bin



By loan\_amnt\_bin : Default Percent to Total

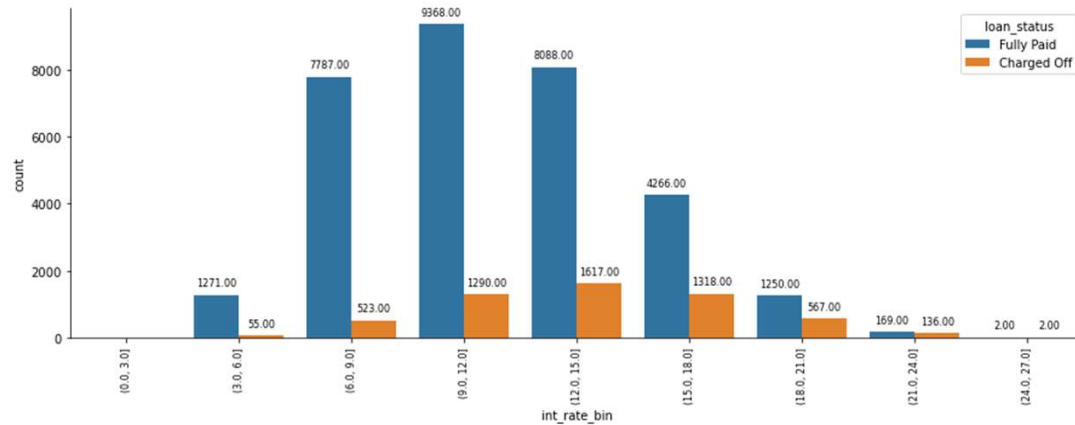


## Observation:

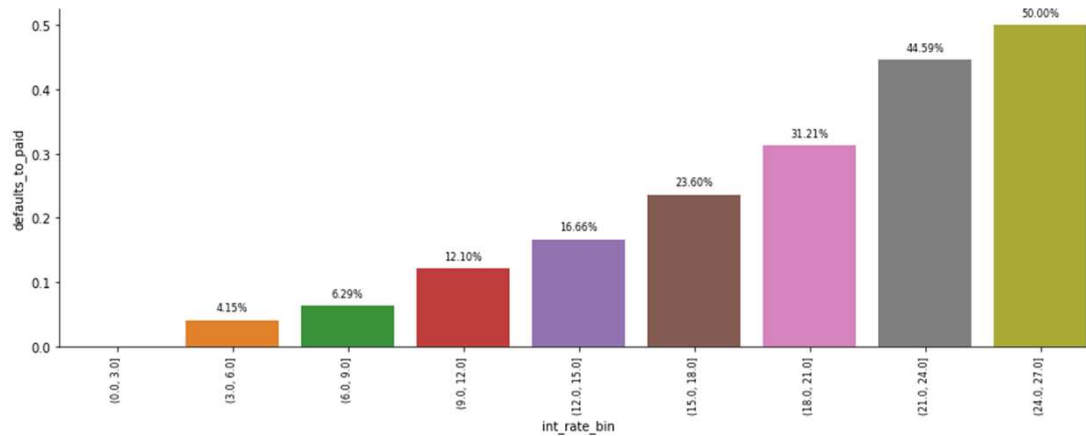
1. Maximum loans availed are of low to mid value
2. Percentage of defaults is higher as the loan amount increases

# INTEREST RATE – DEFAULT VS PAID

Counts by int\_rate\_bin



By int\_rate\_bin : Default Percent to Total

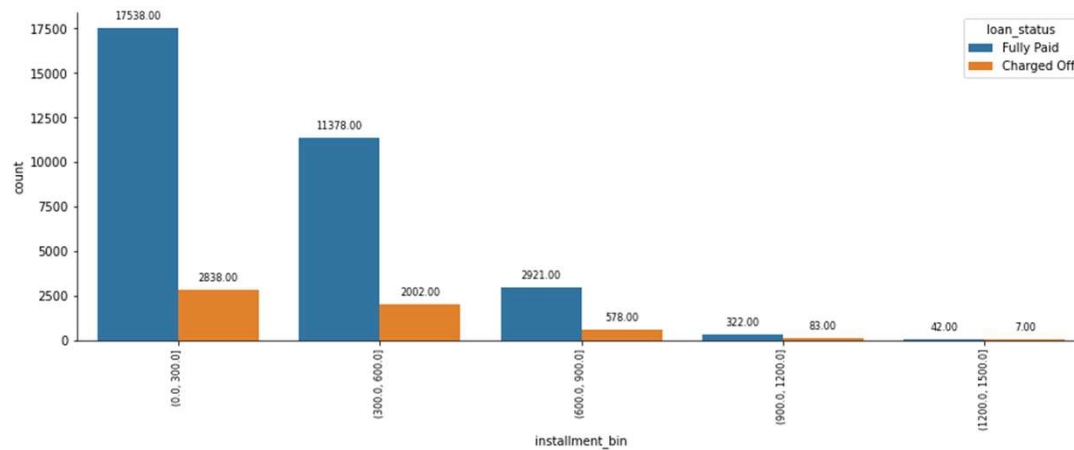


## Observation:

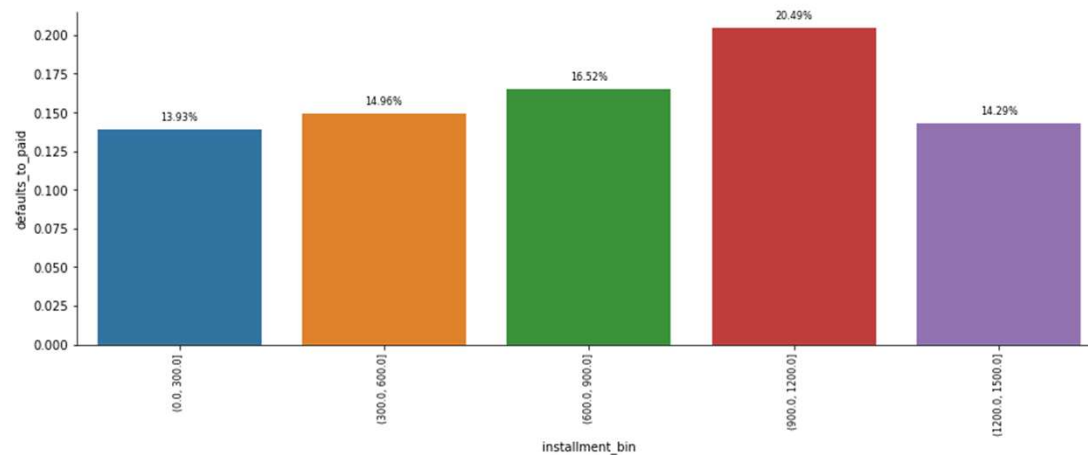
1. Low interest rate is driving higher number of loans
2. As the interest rate increases percentage of defaults also increases

# INSTALMENT – DEFAULT VS PAID

Counts by installment\_bin



By installment\_bin : Default Percent to Total

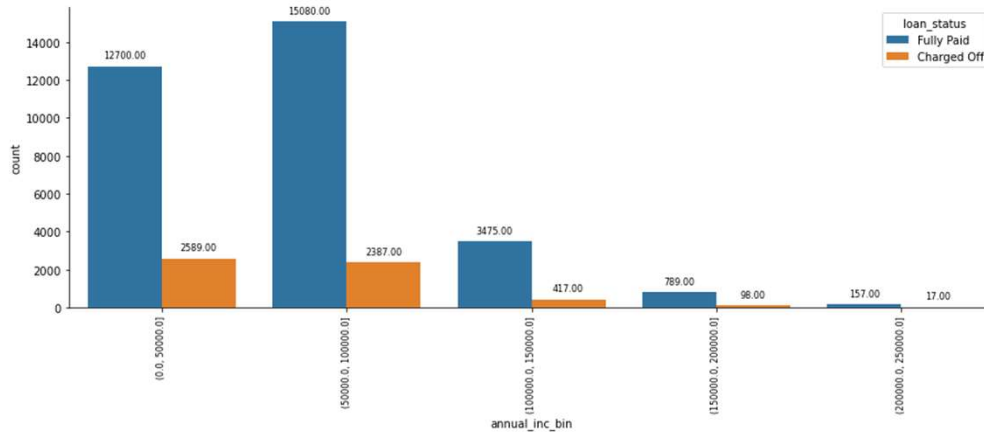


## Observation:

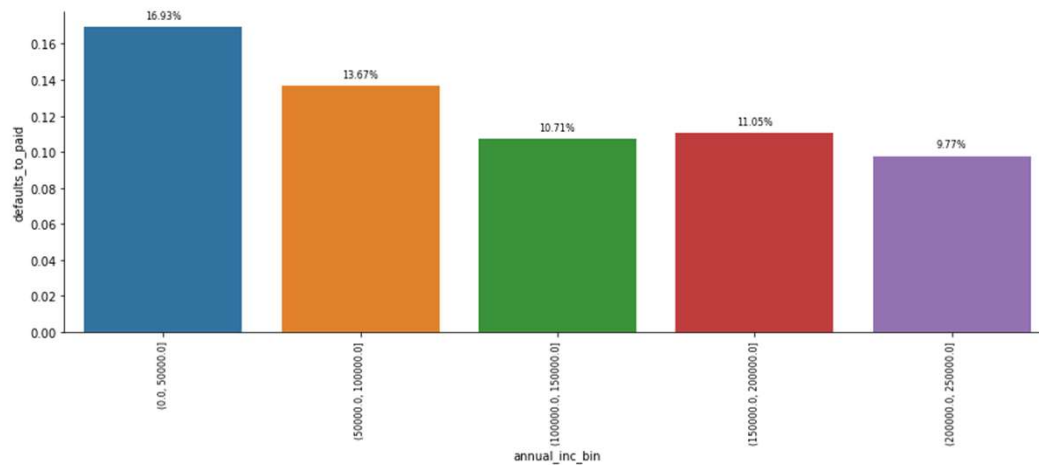
1. Loans with Instalment below 300\$ is most beneficial to business with higher loan disbursement and relatively lower default percentage

# ANNUAL INCOME BIN – DEFAULT VS PAID

Counts by annual\_inc\_bin



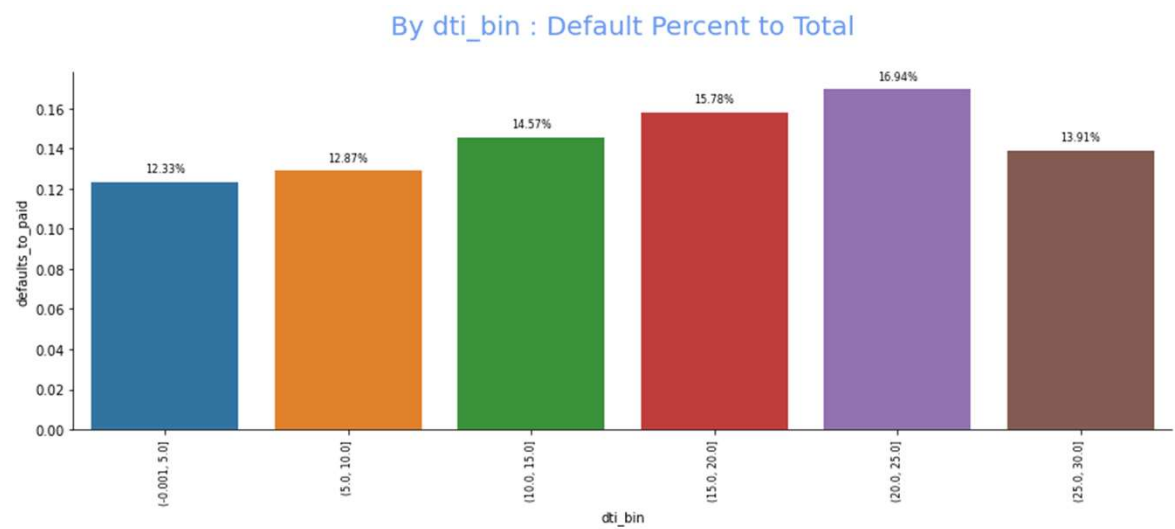
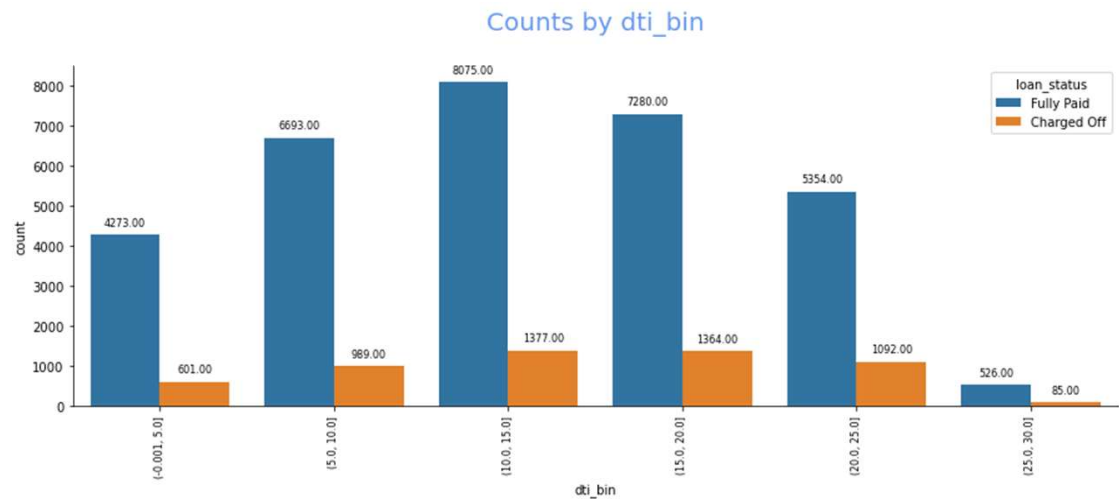
By annual\_inc\_bin : Default Percent to Total



## Observation:

1. Maximum loans are availed by low income groups
2. Relatively higher defaults are also observed by low income groups

# DEBT TO INCOME BIN – DEFAULT VS PAID

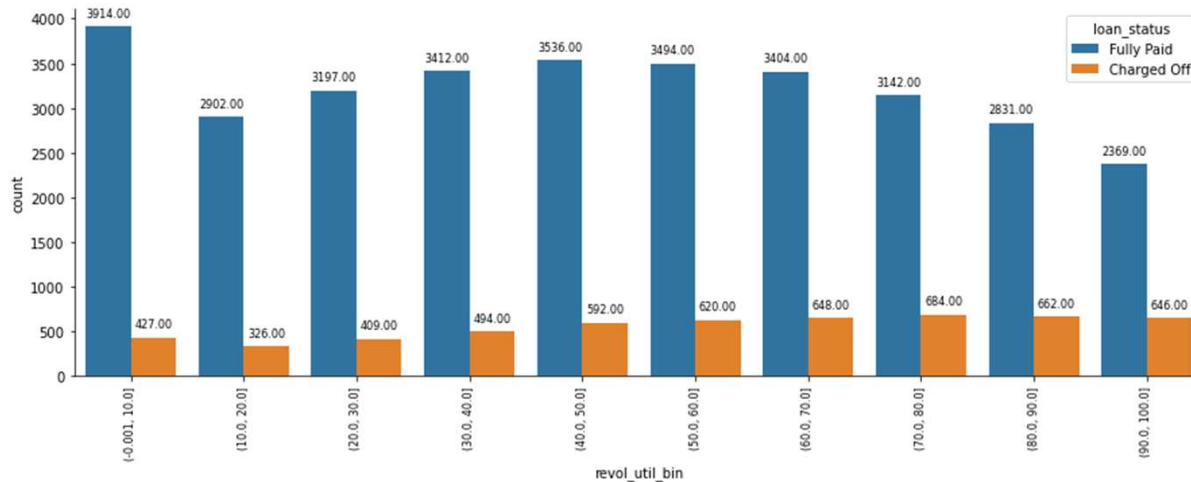


Observation:

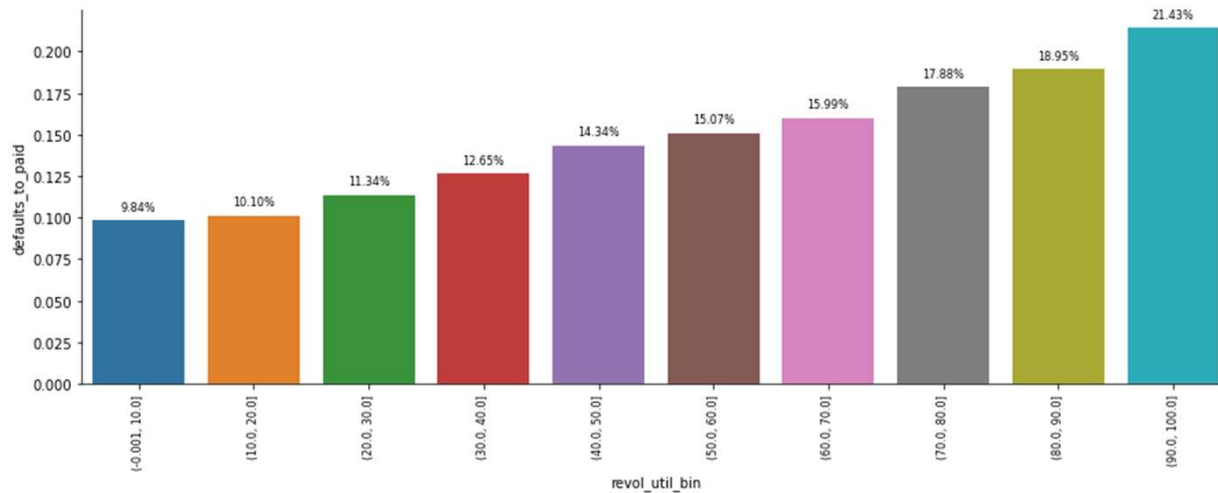
1. Gradual increase in default percent is observed as the Debt to Income ratio increases

# REVOLVING UTILIZATION – DEFAULT VS PAID

Counts by revol\_util\_bin



By revol\_util\_bin : Default Percent to Total



## Observation:

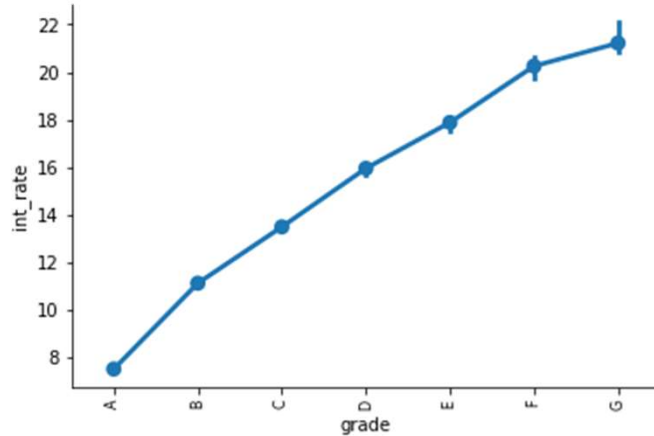
1. Gradual increase in default percent is observed as the revolving utilization increases



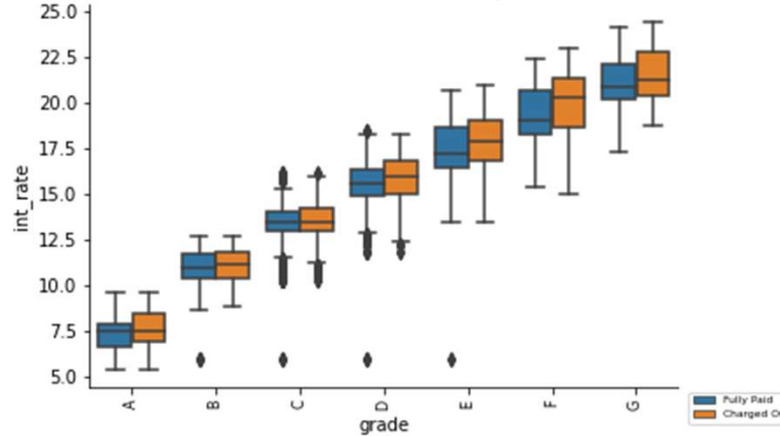
# BIVARIATE ANALYSIS – DEFAULT VS PAID

## Grade vs Interest Rate

Defaulters Relation: Grade vs Interest Rate



Relation: Grade vs Interest Rate by loan status

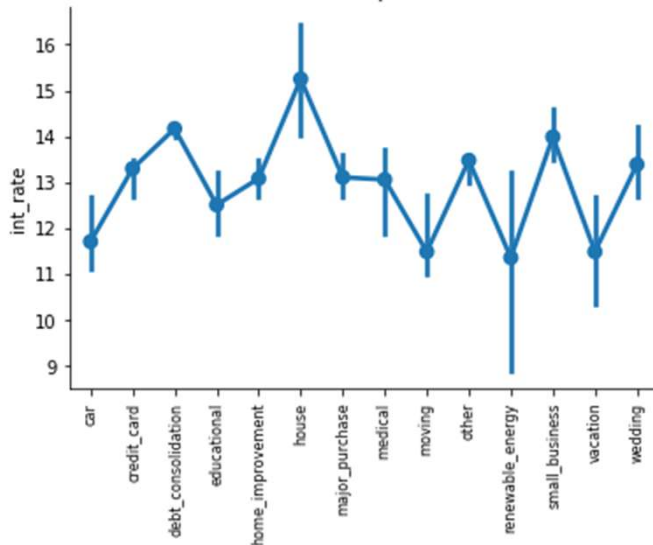


### Observation:

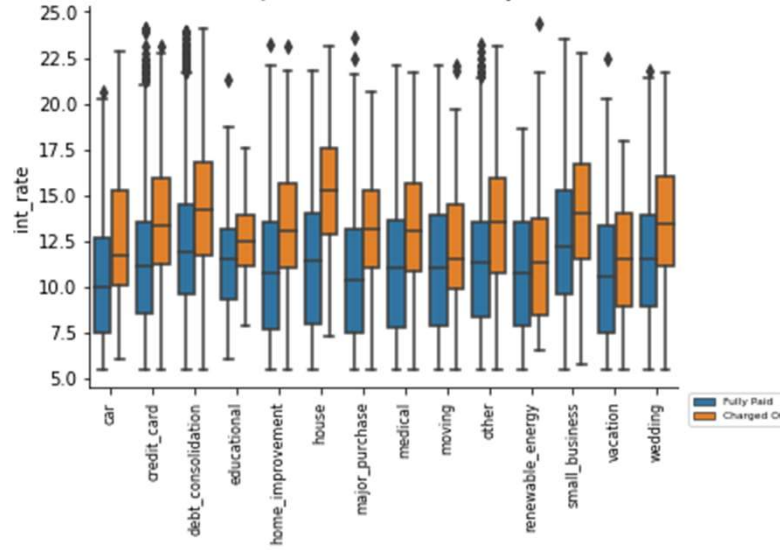
1. Interest Rate increases with Grade
2. Median for default increases with Grade and default rate

## Purpose vs Interest Rate

Defaulters Relation: Purpose vs Interest Rate



Relation: Purpose vs Interest Rate by loan status



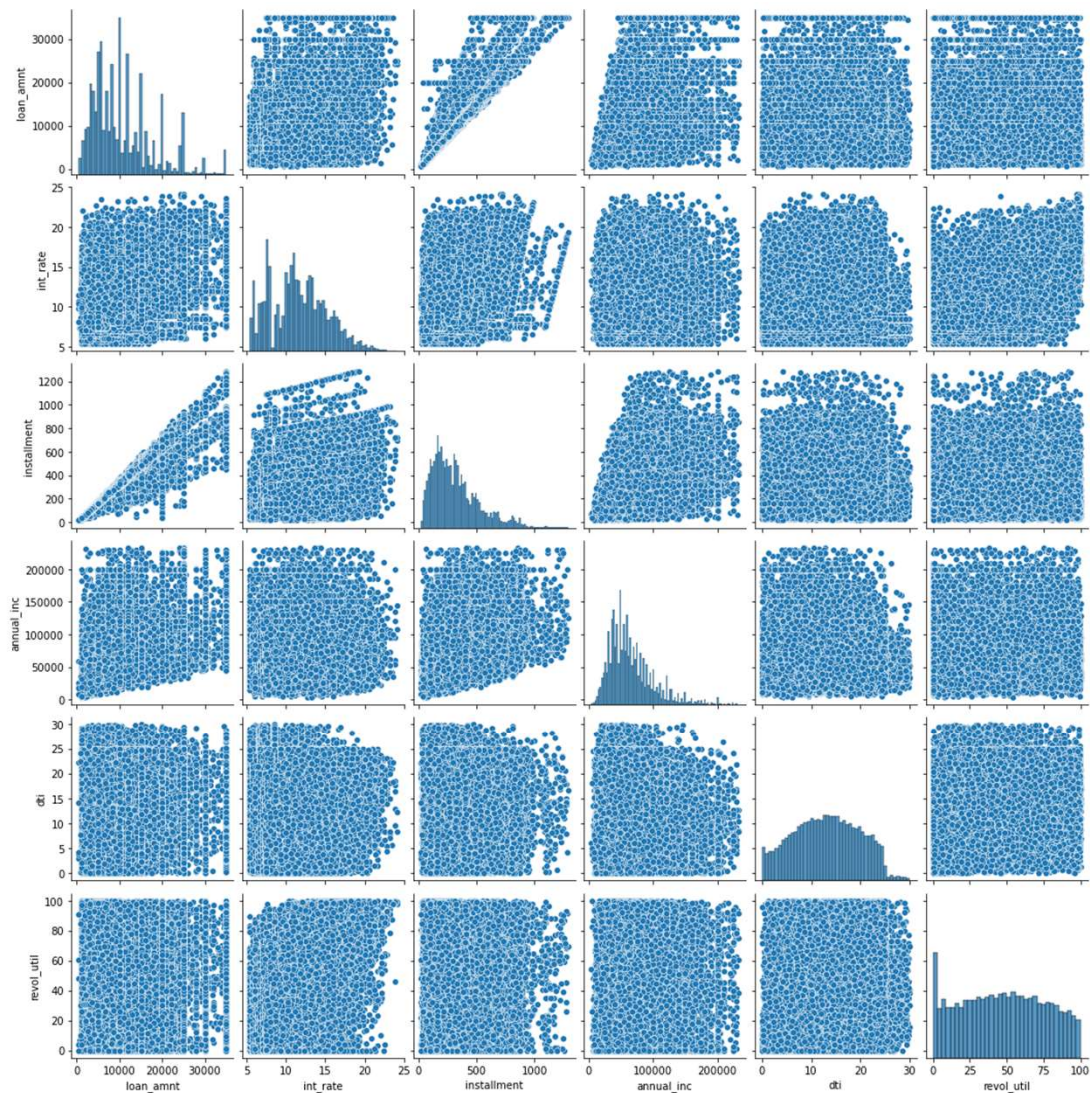
### Observation:

1. Highest interest rate is observed for house, debt consolidation and small business
2. Median of defaults for these are observed to be high as well

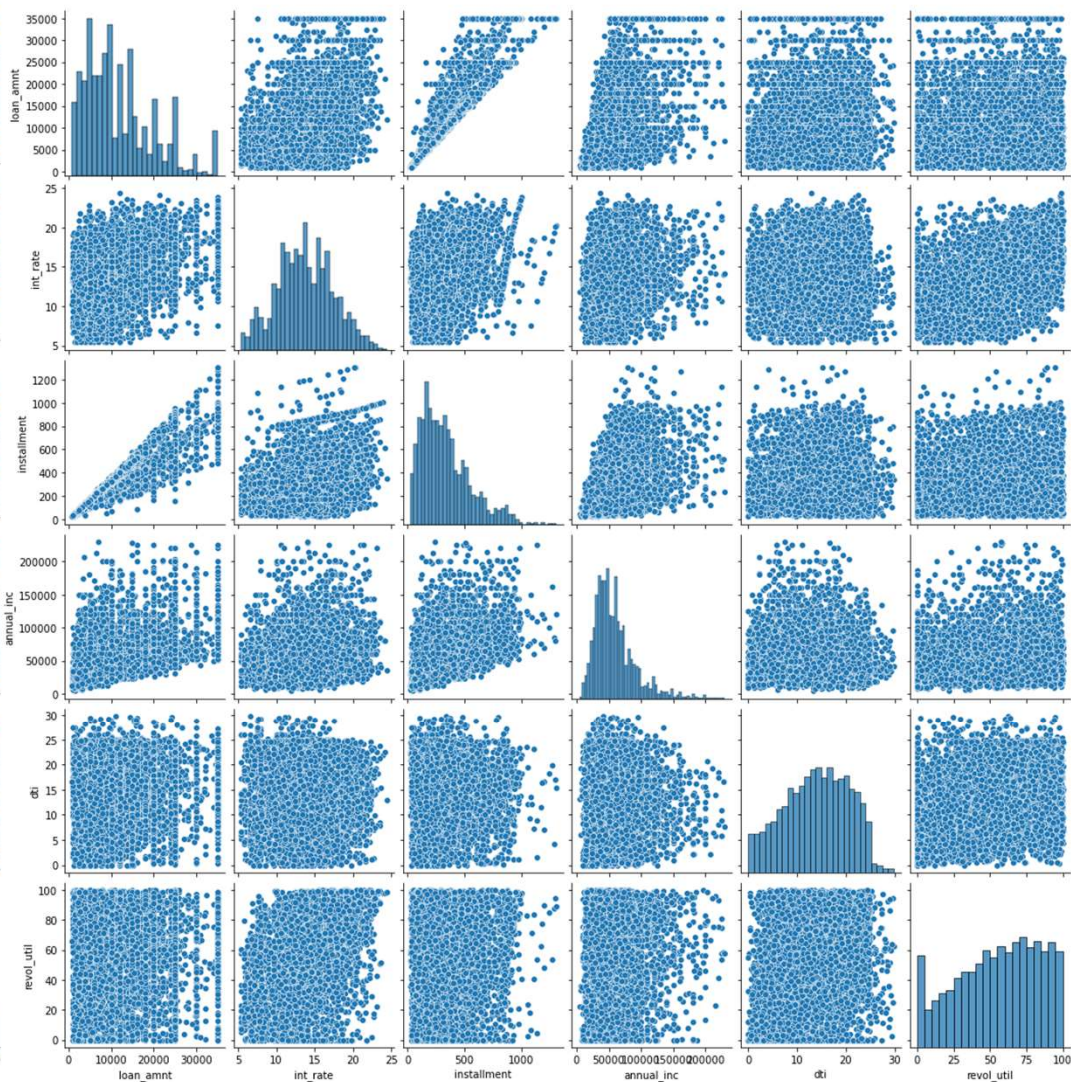


# RELATION ACROSS VARIOUS ATTRIBUTES (BIRDS VIEW)

Paid Loans



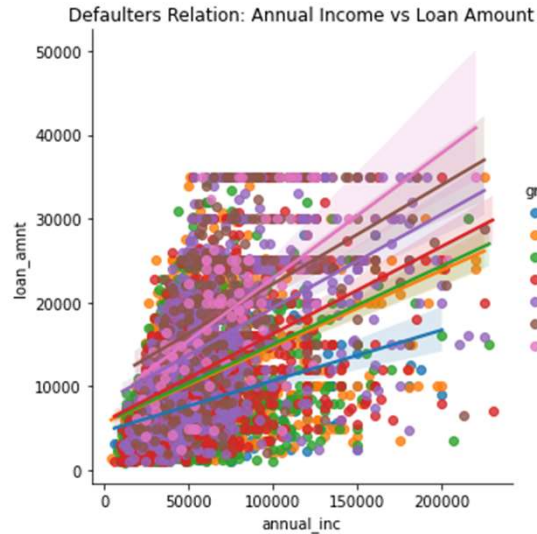
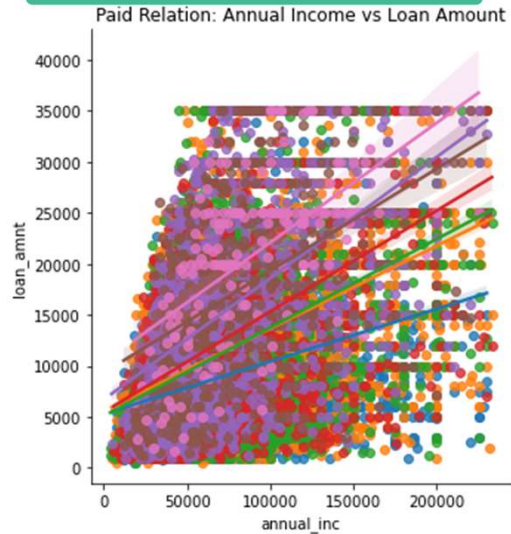
Defaulted Loans





# BIVARIATE ANALYSIS – DEFAULT VS PAID

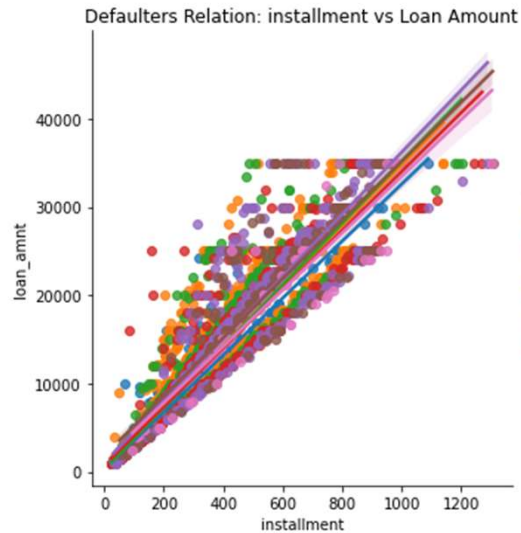
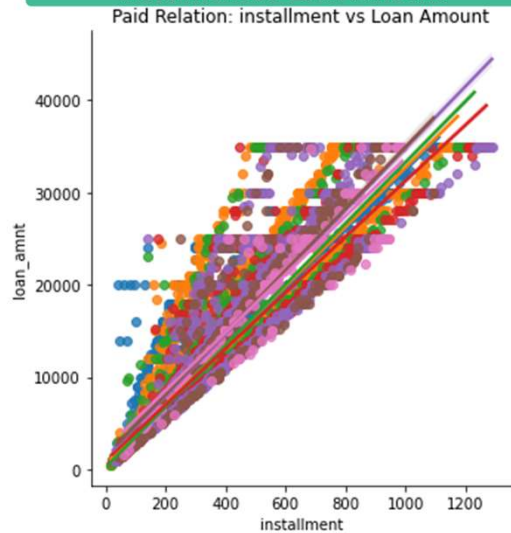
## Annual Income vs Loan Amount



### Observation:

1. Density of loans availed is high for low-mid range annual income and loan amount
2. High density of defaults are observed for low-mid range annual income and loan amount

## Instalment vs Loan Amount



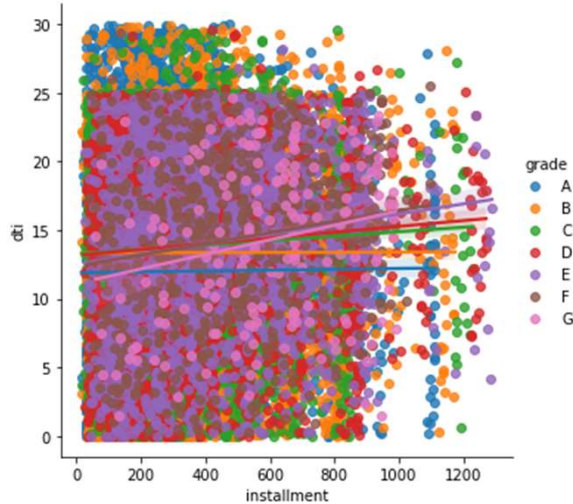
### Observation:

1. As the loan amount increases, instalment increases
2. Density of defaults is high for mid range loan amount and instalment

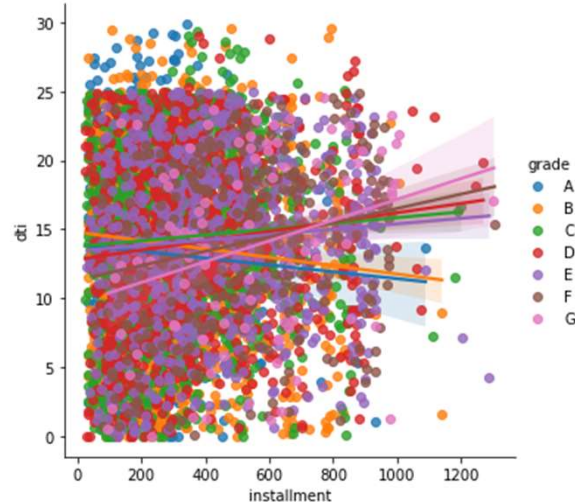
# BIVARIATE ANALYSIS – DEFAULT VS PAID

## DTI vs Instalment

Paid Relation: installment vs DTI

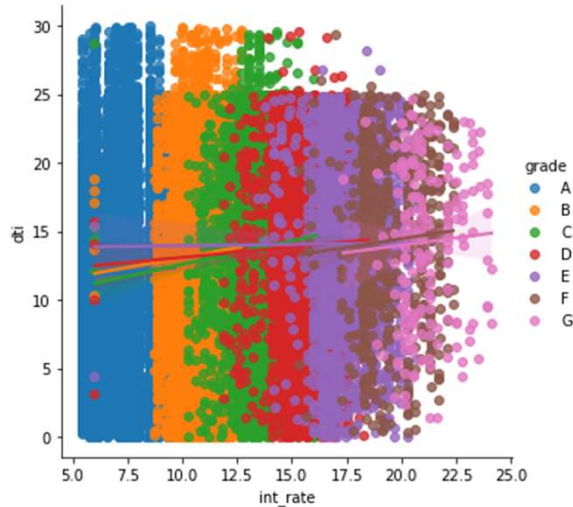


Defaulters Relation: installment vs DTI

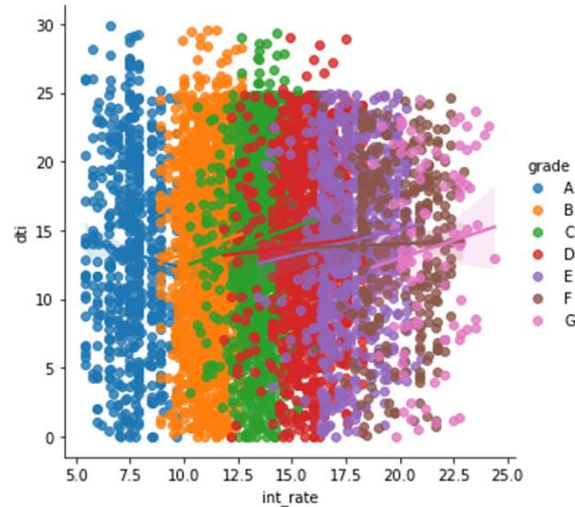


## DTI vs Interest Rate

Paid Relation: interest rate vs dti



Defaulters Relation: interest rate vs dti



### Observation:

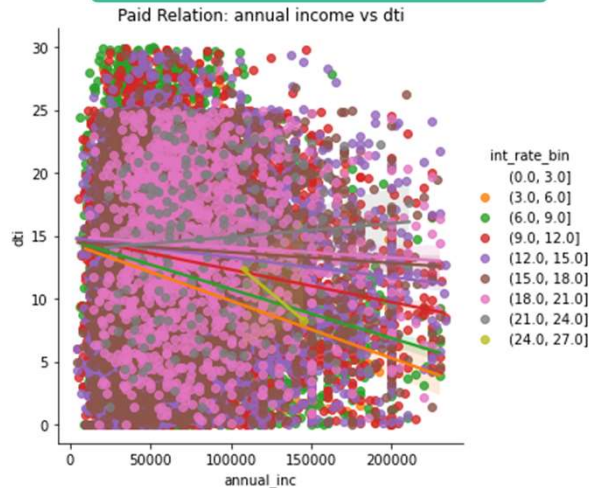
1. Customers with low dti having loans with low instalment exhibit higher payment behaviour
2. Loans in grade A and B tend to show lower risk of default even if the instalment is high

### Observation:

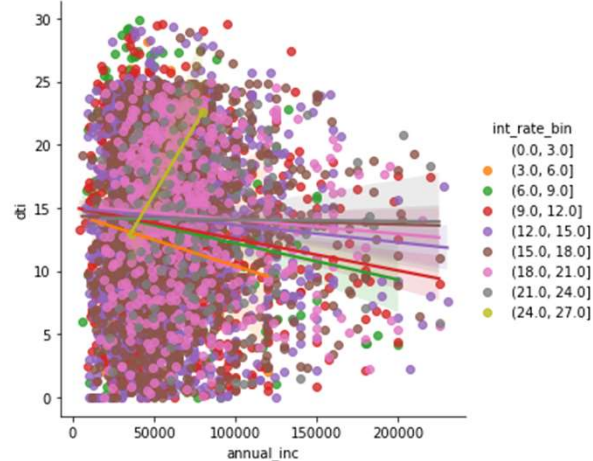
1. With high dti, loans availed at higher interest rate are showing higher defaults

# BIVARIATE ANALYSIS – DEFAULT VS PAID

Annual Income vs DTI



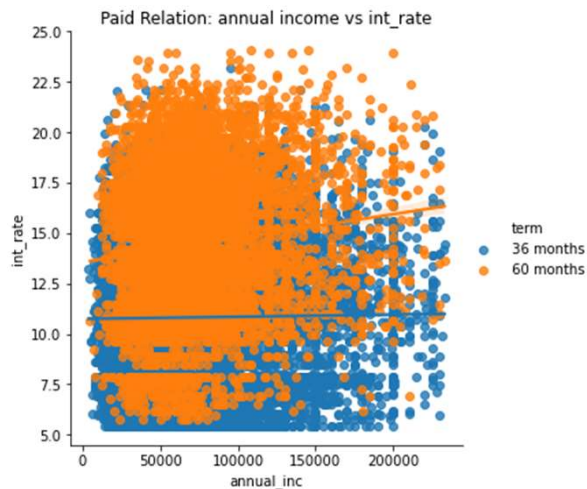
Defaulters Relation: annual income vs dti



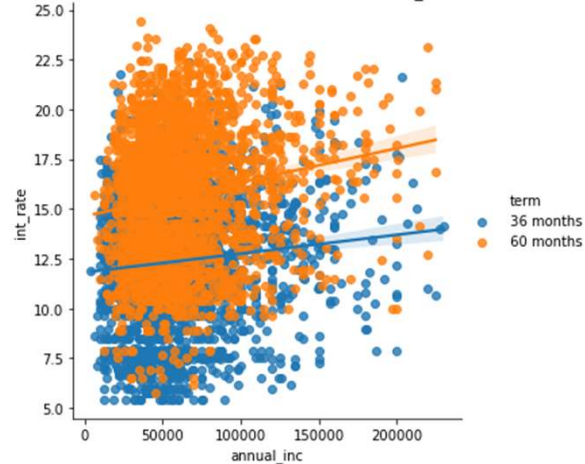
## Observation:

1. Customers with high dti, low income exhibit high risk of defaulting higher interest rate loans

Annual Income vs Interest Rate



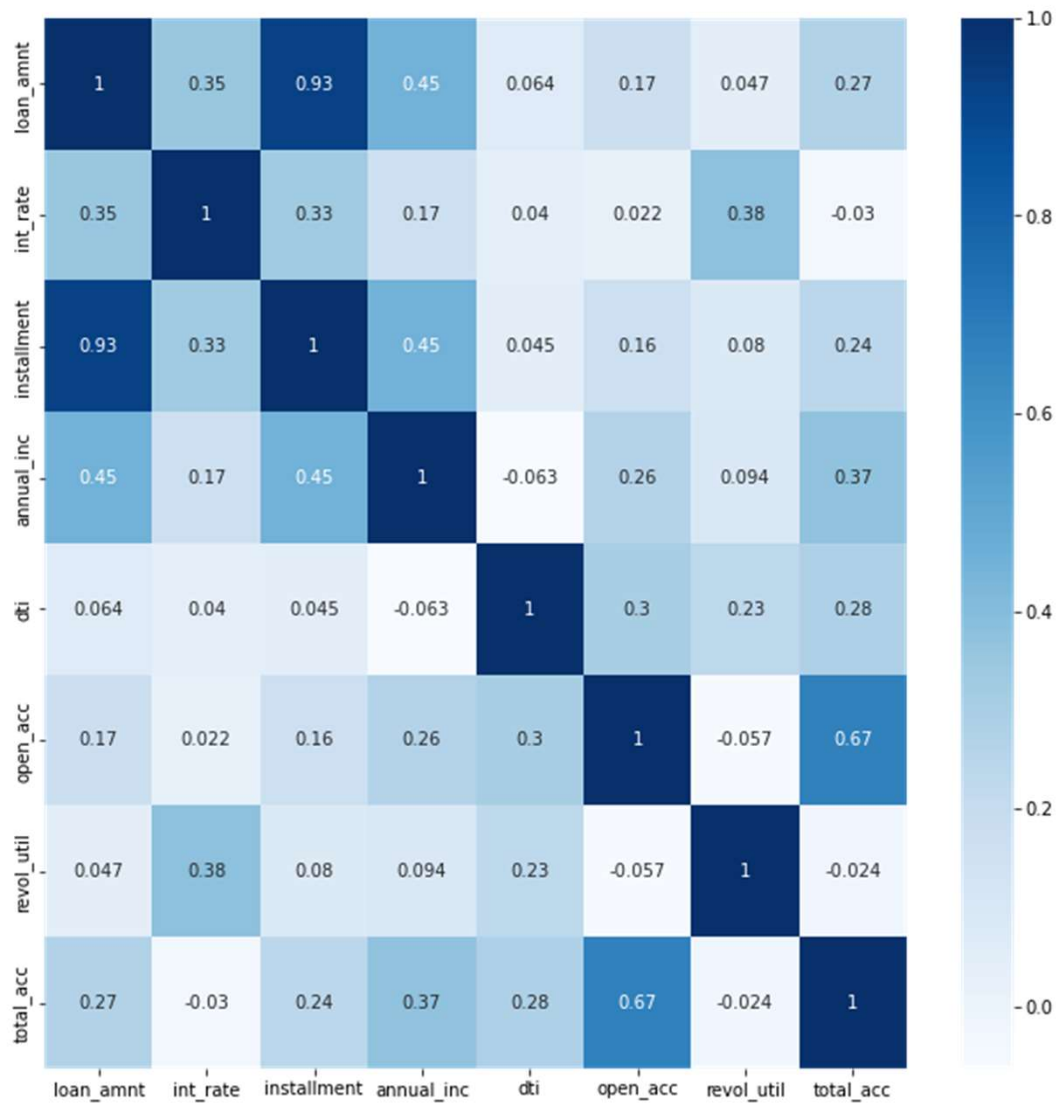
Defaulters Relation: annual income vs int\_rate



## Observation:

1. Long term loans at higher interest rate show a higher trend of defaulting

## HEAT MAP – DEFAULTED LOANS



### Observation:

1. Total number of credit lines in a borrowers file shows high correlation with the number of open accounts
2. As the annual income increases the loan instalment increases. Same is the case with loan amount

**Note:** Heat map shows similar correlation with attributes in case of paid or defaulted loans. Hence only defaulted loans is displayed here



## OBSERVATIONS, RECOMMENDATIONS

Below listed attributes are critical in determining whether a loan brings in good business or offers a high risk of default

Attribute	Observations
Term	<ul style="list-style-type: none"><li>• <b>Short Term loans (36 months) have a low probability of default</b> when compared with Long Term(60 months)</li><li>• Long term loans at higher interest rate show a higher trend of defaulting</li></ul>
Grade, Sub Grade	<ul style="list-style-type: none"><li>• Terms with <b>Grade D,E,F,G and corresponding sub grades have a higher percentage of default</b> when compared with A,B,C;</li><li>• One of the reasons could be the higher interest rate associated with it</li><li>• For customers <b>with low DTI, loans in grade A and B tend to show lower risk</b> of default even if the installment is high</li></ul>
Purpose	<ul style="list-style-type: none"><li>• Most loans are taken for purpose of '<b>Debt Consolidation</b>'. While 15% of these default, they are still a <b>good business opportunity</b>.</li><li>• Loans taken for the purpose of '<b>Small Business</b>' show a <b>higher percent</b> of default and should be scrutinized before approval</li><li>• Loans taken for purpose of '<b>Housing</b>' have a <b>higher interest rate and the default percentage is high</b> for the housing loans at higher interest</li></ul>
Income Verification Status	<ul style="list-style-type: none"><li>• The loans <b>disbursed post verification of income status have shown higher percent of defaults</b> when compared to loans for which income status has not been verified</li></ul>
Month of Issue	<p>Maximum Loans are being issued towards the end of the year (holiday season) and this is a good time for business</p> <p>While defaults are relatively similar throughout the year, higher defaults are observed during Financial Year close and holiday season</p>

## OBSERVATIONS, RECOMMENDATIONS

Below listed attributes are critical in determining whether a loan brings in good business or offers a high risk of default

Attribute	Observations
Loan Amount	<ul style="list-style-type: none"><li>• <b>Most loans</b> availed are <b>low-mid value</b> (\$15,000).</li><li>• As the <b>loan amount increases</b> the percentage of <b>corresponding defaults also increases</b>.</li><li>• <b>Most of the loans</b> are availed by the customers who are in <b>low-mid annual income group</b></li></ul>
Interest Rate	<ul style="list-style-type: none"><li>• Most loans availed are low-mid rate (15%).</li><li>• Loans availed at <b>higher interest</b> rate display <b>higher percentage of defaults</b></li></ul>
Instalment	<ul style="list-style-type: none"><li>• Loans with <b>Instalment below 300\$</b> is <b>most beneficial to business</b> with higher loan disbursement and relatively lower default percentage</li></ul>
Annual Income	<ul style="list-style-type: none"><li>• People in <b>low-mid income</b> groups bring in <b>high volume of business</b> by availing most loans but they also account for <b>higher percent of defaults</b></li></ul>
Debt To Income Ratio	<ul style="list-style-type: none"><li>• As the <b>DTI increases</b>, corresponding <b>default percent also increases</b>.</li><li>• Customers with high dti have availed loans availed at higher interest rate and are showing higher defaults</li><li>• Customers with <b>high dti, low income exhibit high risk of defaulting</b> higher interest rate loans</li></ul>
Revolving Utilization	<ul style="list-style-type: none"><li>• Customers with higher revolving utilization have higher risk of defaulting</li></ul>