

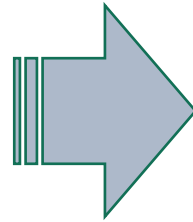


# *LENDING CLUB CASE STUDY*

AMANDEEP MINHAS  
ARUN VINCENT J A

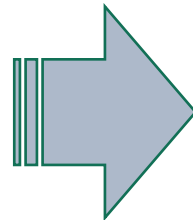
# PROBLEM

COMPANY  
PROFILE



Lending Club is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures.

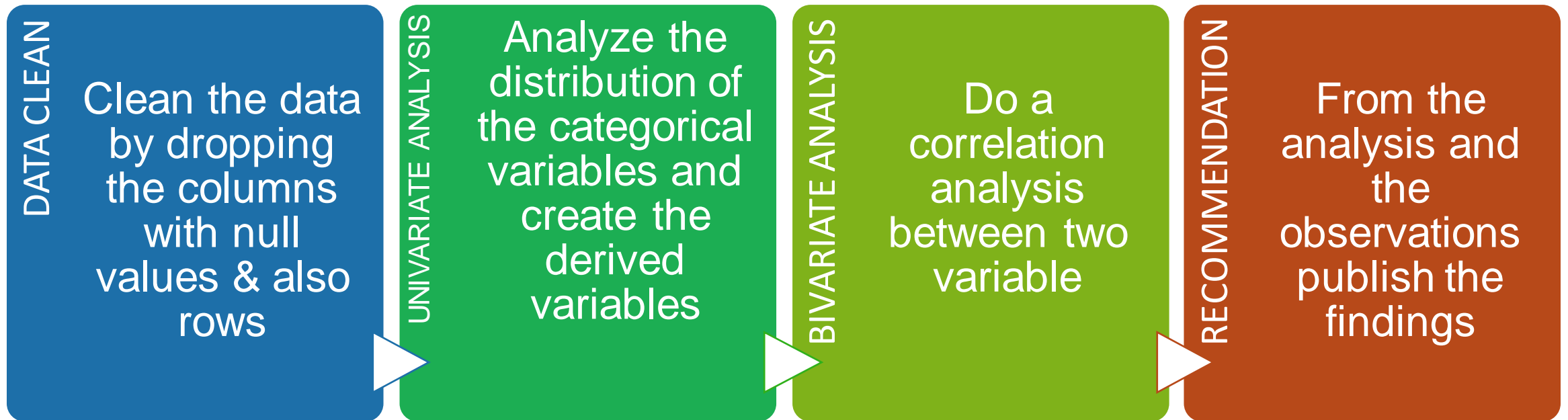
PROBLEM



Lending Club wants to understand the driving factors behind loan default, i.e. the driver variables which are strong indicators of default.

So as a data scientist working for Lending Club, we have to analyze the dataset containing information about past loan applicants using EDA to understand how consumer attributes and loan attributes influence the tendency of defaulters.

# APPROACH



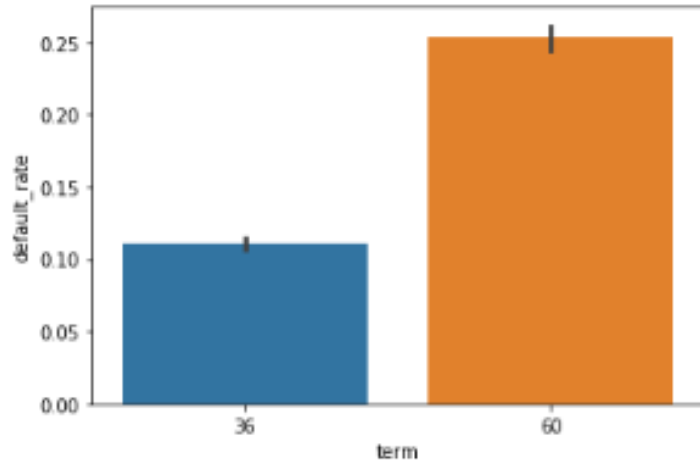
# DATA CLEANUP

- Drop columns with all null values
- Drop columns with high missing values and what won't contribute much to our analysis.
- As emp\_title has ~73% unique values, it's impact on our analysis won't be of much significance. So let's drop it.
- Since loan\_amnt, funded\_amnt and funded\_amnt\_inv represent similar identity, it would be safe to just use 1 columns. We will use funded\_amnt\_inv as that represents the loan granted. We will also rename funded\_amnt\_inv to loan\_amnt

# DATA CLEANUP

- As loan\_status is our target column. We only want to figure out if a customer defaulted or not.
- So We will create a new column named default\_rate with will hold value of either 0 or 1 as such :-
  - 0: Not Defaulted
  - 1: Defaulted
- We can also drop rows where loan\_status is eq to Current as it will not contribute to analysis
- Delete columns that are of data that are either collected after a loan is grated to the customer/have singular for all rows/ do not contribute to analysis.
- Convert term to numeric field
- Drop column member\_id as it has all unique values which do not contribute to analysis.
- Remove Outliers from **Annual Income, Loan Amount, Interest Rate**
- Since there are very few records for Grade E, we will consider it an outlier and drop that column.

# ANALYSIS

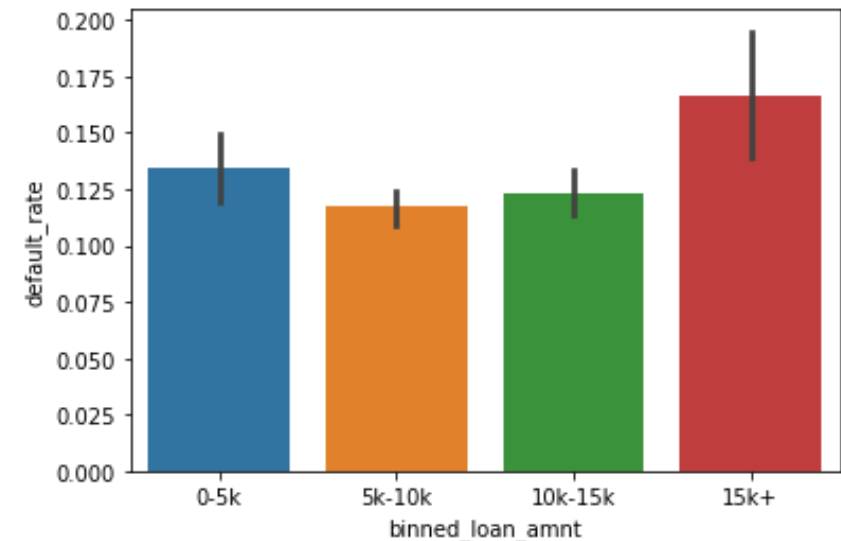


## Term Analysis

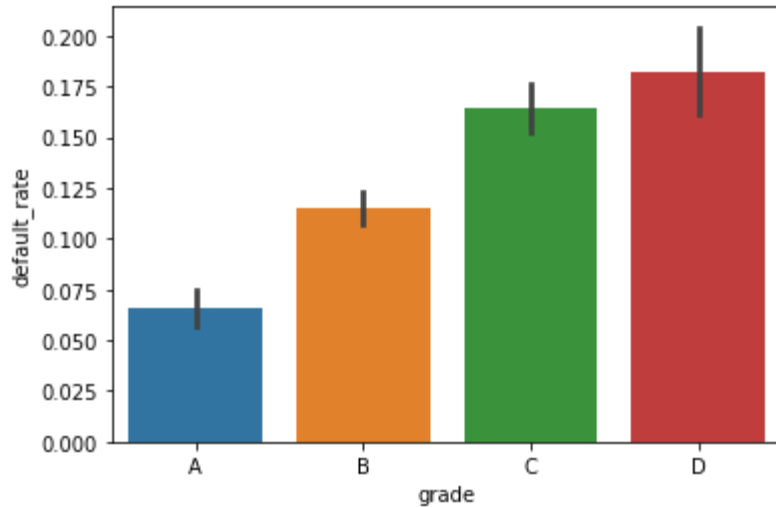
- 10% of the loans taken for a term of 36 months defaulted.
- 21% of the loans taken for a term of 60 months defaulted.

## Analysis with respect to Loan Amounts

- ~13% of the customers who borrow less than 5k tend to default.
- Lowest Risk** ~12% of the customers who borrow in the range of 5k to 10k tend to default.
- ~12% of the customers who borrow in the range of 10k to 15k tend to default.
- Highest Risk** ~17% of the customers who borrow more than 15k tend to default.



# ANALYSIS



## Grade:

As expected, as the grade increases, the risk or default increases as well.

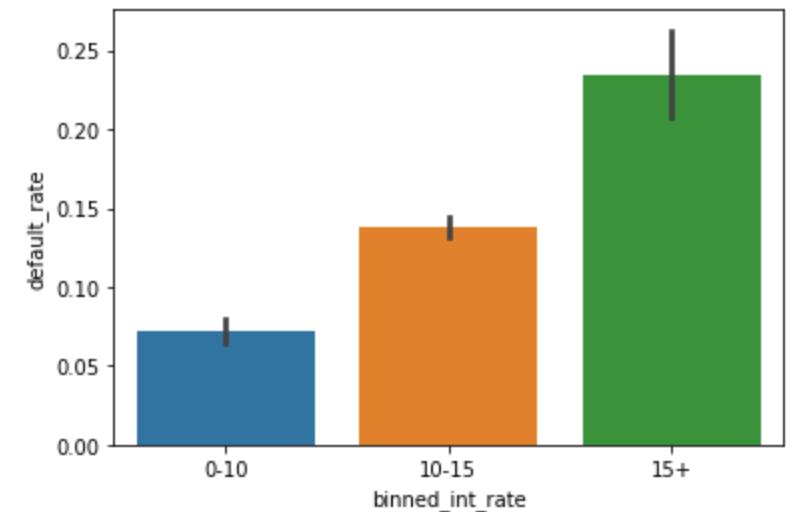
**Lowest Risk** Chances of default are minimum for loans given out to customers with grade A (~7%)

**Highest Risk** Chances of default are maximum for loans given out to customers with grade D (~18%)

## Analysis with respect to Interest Rate

- ~**low** ~7% default when interest rate is **less than 10%**.
- ~ **Medium** ~14% default when interest rate is **between 10% to 15%**.
- ~ **High** ~23% default when interest rate is **above 15%**.

Interest Rate and Tendency to Default are directly proportional. As one increases, other increases as well.

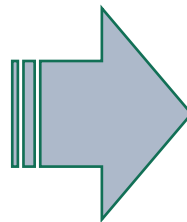
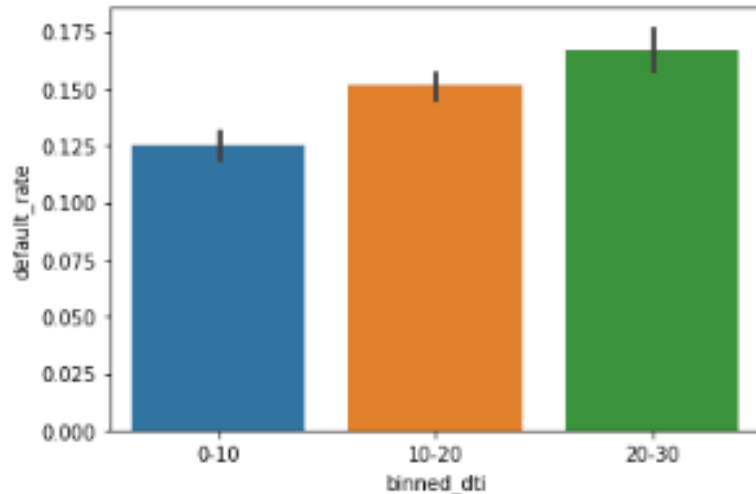
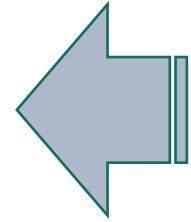
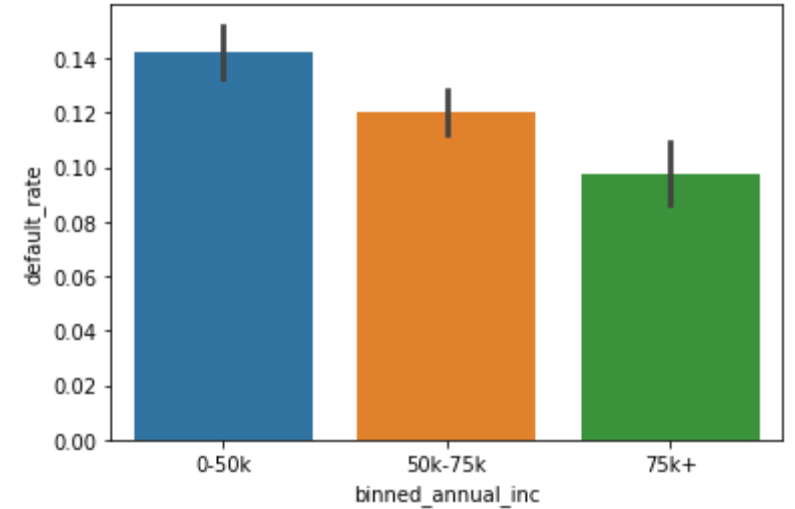


# ANALYSIS

## Analysis with respect to Annual Income

As income increases, the chances of customer defaulting decreases.

- ~14% of customers with annual income of less than USD 50k tend to default.
- ~12% of customers with annual income between USD 50k to USD 75k tend to default.
- ~10% of customers with annual income above USD 75k tend to default.



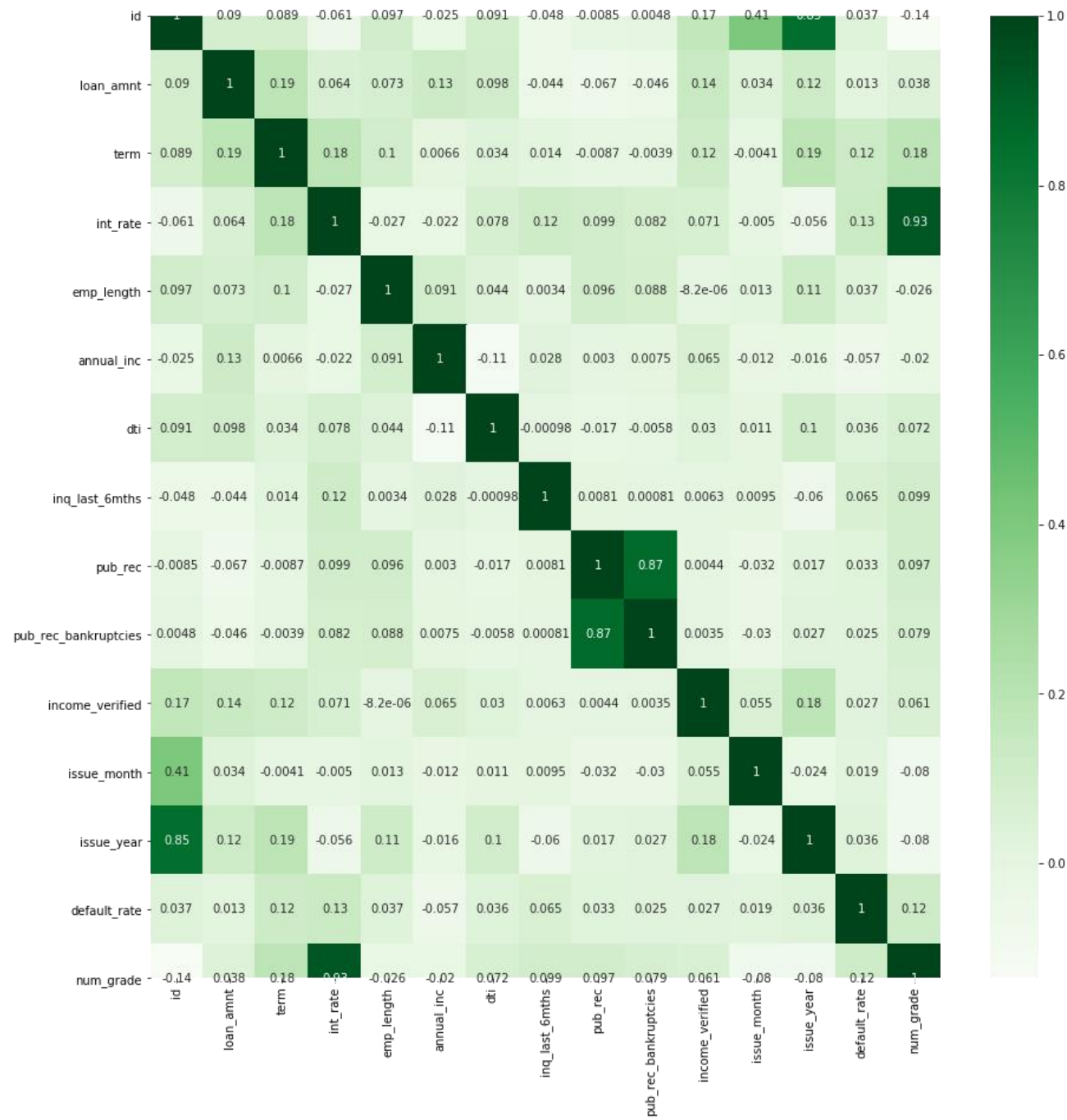
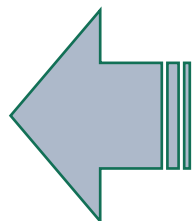
## Analyze DTI:

After Creating buckets for DTI ranging from 0-10, 10 to 20 and 20 - 30 and plot against default rate it is observed that **as DTI increases, the tendency to default increases**

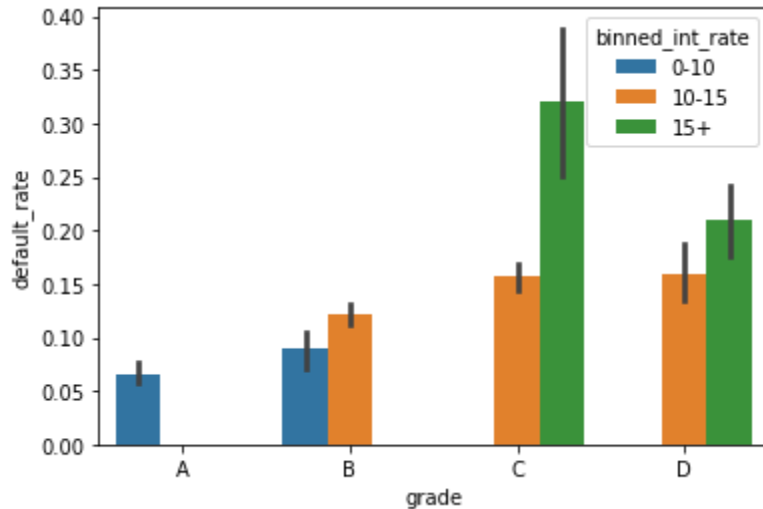


# ANALYSIS

Correlation map



# ANALYSIS



## Interest Rate | Grade

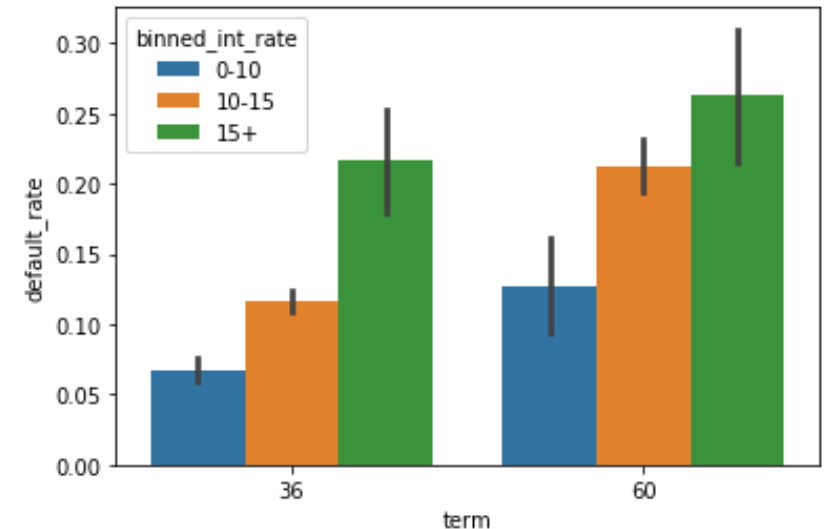
**Lowest Risk** → Chances of default are minimum for loans given out at interest rate of less than 10% to customer with Grade A. (7%)

**Highest Risk** → Chances of default are maximum for loans given out at interest rate of more than 15% to customer with Grade C. (32%)

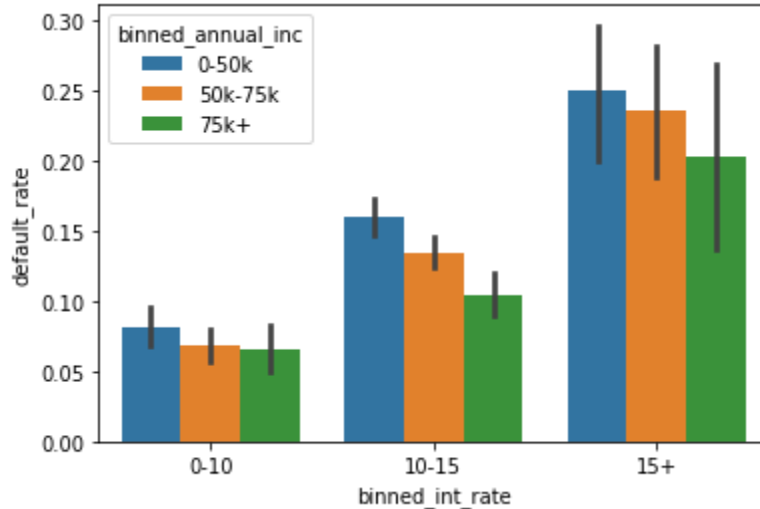
## Interest Rate | Term

**Lowest Risk** → Chances of default are minimum when granting a loan for tenure of 36 months with interest rate of less than 10% (7%)

**Highest Risk** → Chances of default are maximum when granting a loan for tenure of 60 months with interest rate of more than 15% (26%).



# ANALYSIS



## Interest Rate | Annual Income

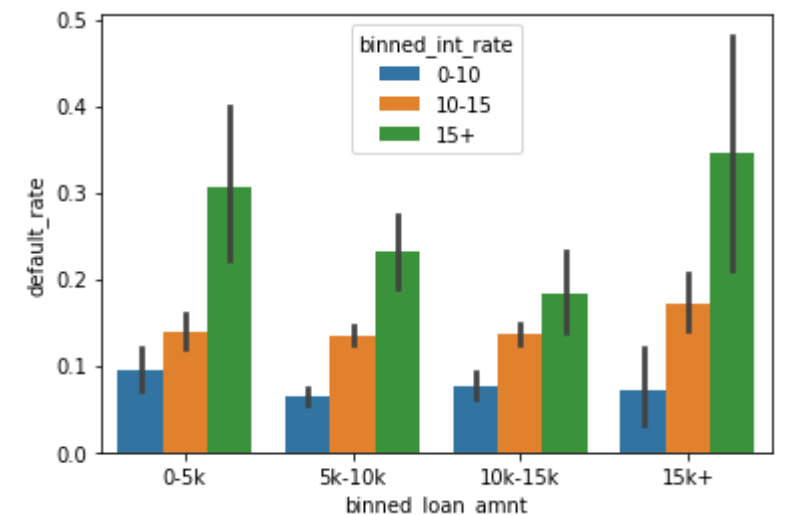
**Lowest Risk** → Chances of default are minimum (~7%) for loans given out at an interest rate of less than 10% to customers with annual income of more than 75k

• **Highest Risk** → Chances of default are maximum (~25%) for loans given out at an interest rate of more than 15% to customers with annual income of less than 50k

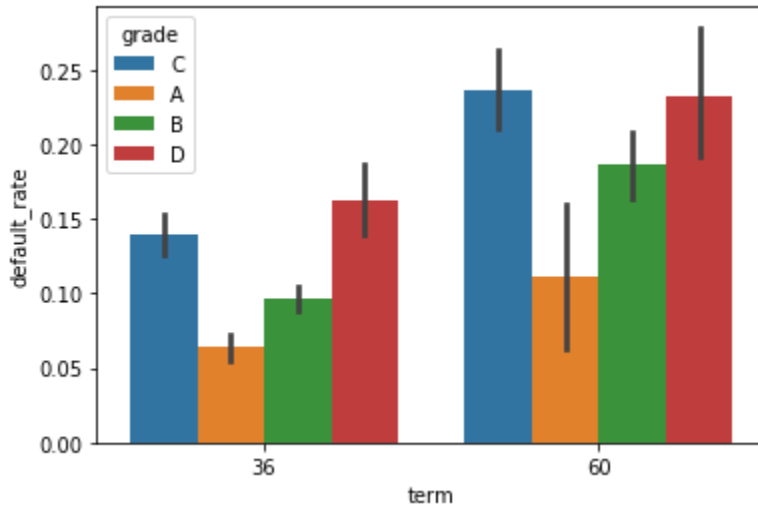
## Interest Rate | Loan Amount

**Lowest Risk** → Chances of default are minimum for a loan of over 5k-10k with interest rate of less than 10% (6%).

**Highest Risk** → Chances of default are maximum for a loan of over 15k with interest rate of more than 15% (35%).



# ANALYSIS



## Grade | Term

**Lowest Risk** → Chances of default are minimum when giving out loan to customer with A grade for a term of 36 months (~6%)

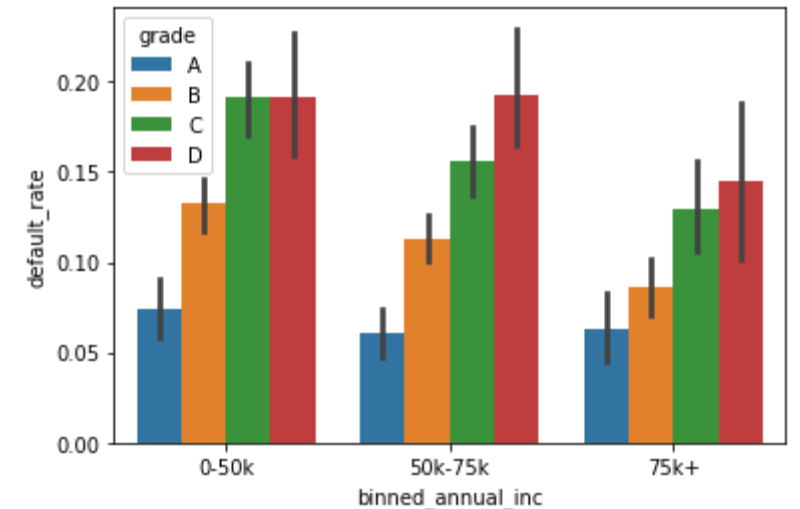
**Highest Risk** → Chances of default are maximum when giving out loan to customer with C grade for a term of 60 months (~24%)

## Grade | Annual Income

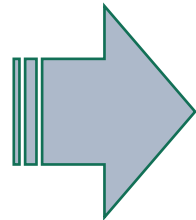
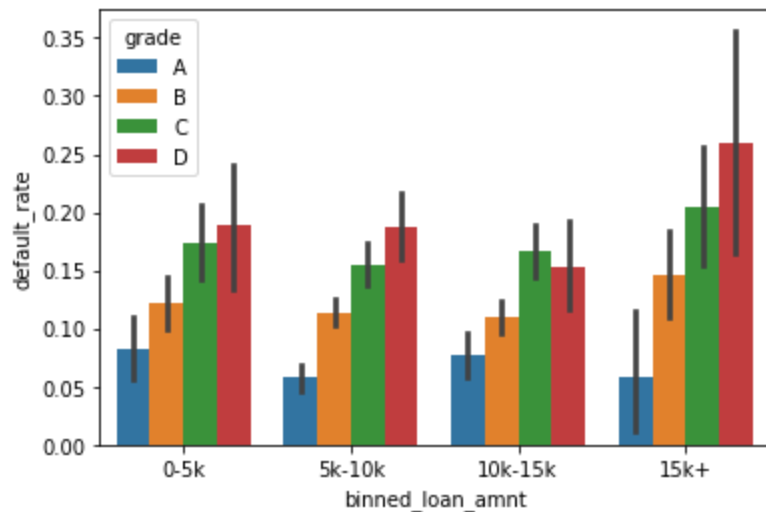
**Lowest Risk** → Chances of default are minimum when granting loans to customers with annual income of 50k - 75k and with grade A (~1%)

**Highest Risk** → Chances of default are maximum when granting loans to

- customers with annual income of less than 50k and with grade C or D (~19%)
- customers with annual income of 50k-75k and with grade D (~19%)



# ANALYSIS



## Grade | Loan Amount

**Lowest Risk** → Chances of default are minimum when granting loan of 5k - 10k to customers with grade A (~6%)

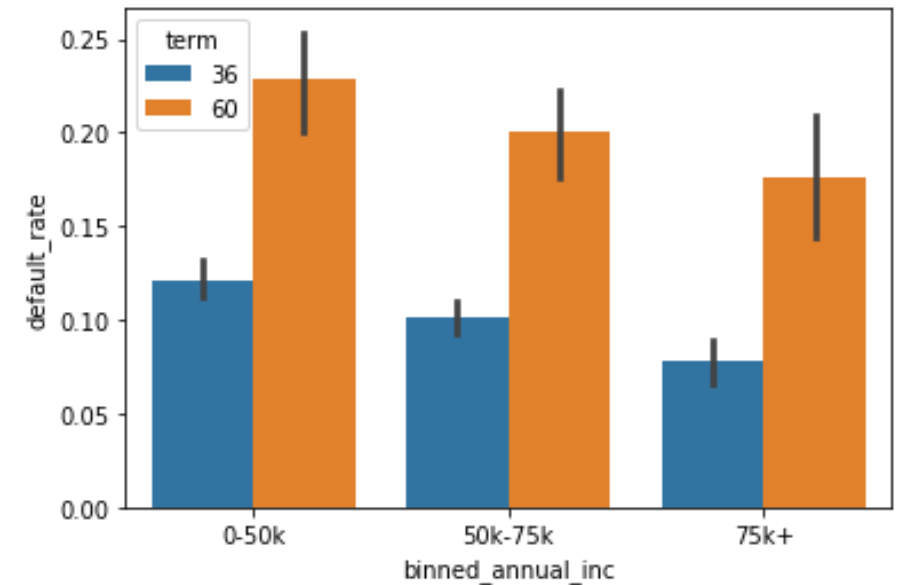
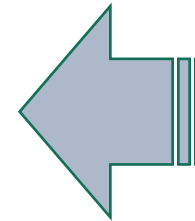
**Highest Risk** → Chances of default are maximum when granting loan of more than 15k to customers with grade D (~26%)

# ANALYSIS

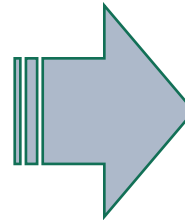
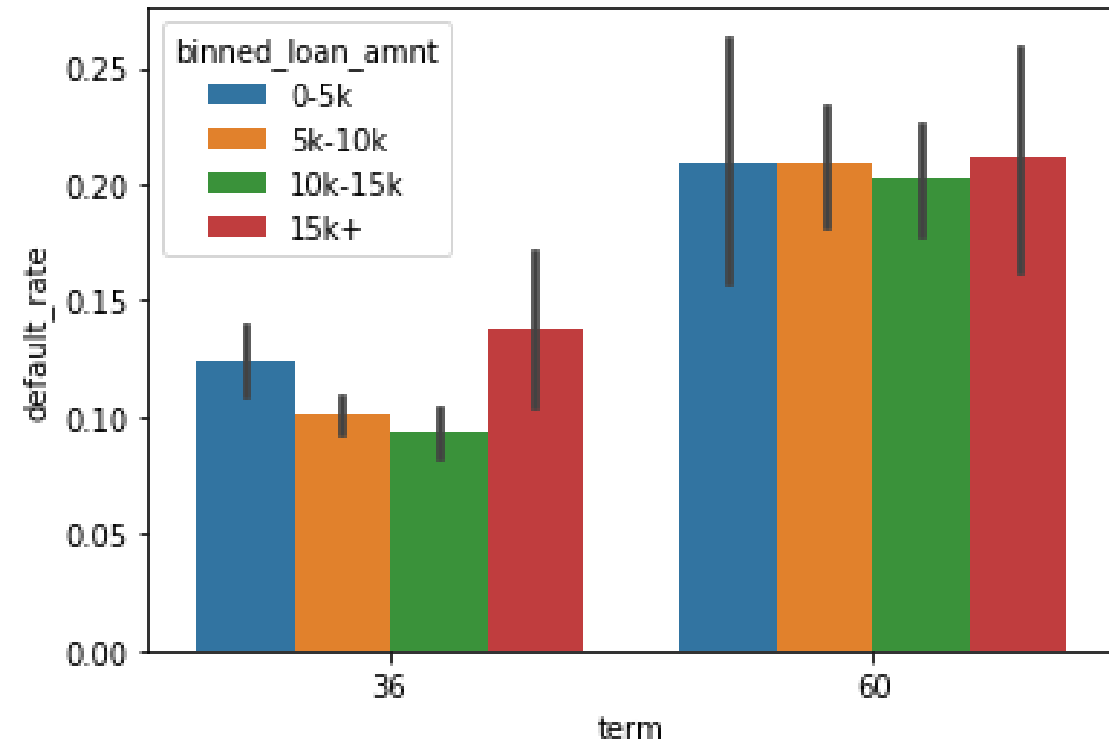
## Term | Annual Income

**Lowest Risk** → Chances of default are minimum when giving out a loan to a customer with annual income of 75k+ for a term of 36 months (~8%)

**Highest Risk** → Chances of default are maximum when giving out a loan to a customer with annual income of less than 50k for a term of 60 months (~23%)



# ANALYSIS



## Term | Loan Amount

Loans given for USD 10-15k tend to do better and have less change of defaulting.

### For 36 months

- chances of default are minimum for loans ranging between 10k to 15k. (~9%)
- chances of default are maximum for loan of more than 15k. (14%)

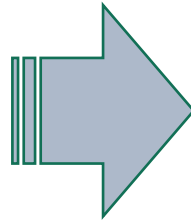
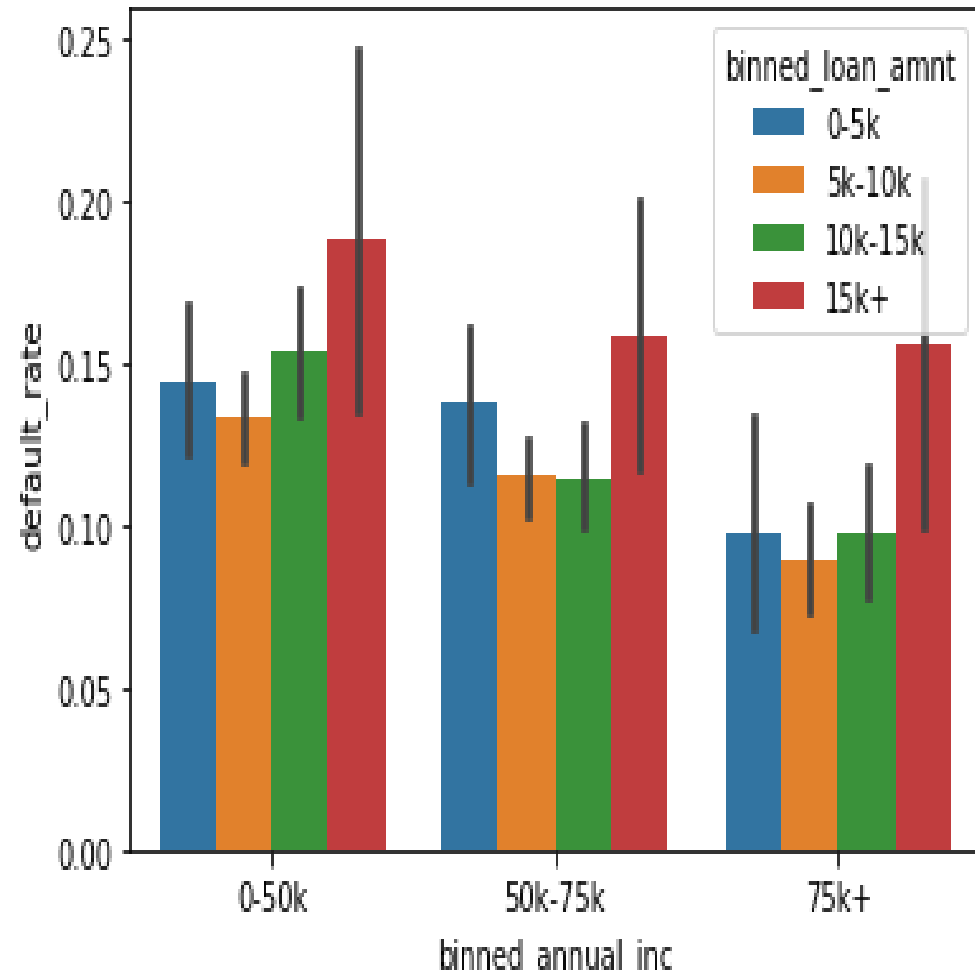
### For 60 months

- chances of default are minimum for loans ranging between 10k to 15k. (~20%)
- chances of default of all others loan amounts are 21%

### Overall

- **Lowest Risk** Chances of default are minimum when giving out loan in the range of USD 10k to 15k for a term of 36 months (~9%)
- **Highest Risk** Chances of default are maximum when giving out loan for any amount for a term of 60 months

# ANALYSIS



## Loan Amount | Annual Income

### For Income 0-50k

- chances of default are minimum for loans ranging between 5k to 10k. (~13%)
- chances of default are maximum for loans greater than 15k. (19%)

### For Income 50k-75k

- chances of default are minimum for loans ranging between 5k to 15k. (~11%)
- chances of default are maximum for loans greater than 15k. (16%)

### For Income 75k+

- chances of default are minimum for loans ranging between 5k to 19k. (~9%)
- chances of default are maximum for loans greater than 15k. (16%)

**Lowest Risk** → chances of default are minimum for loans ranging between 5k to 10k granted to customer with income of greater than 75k (~9%)

**Highest Risk** → chances of default are maximum for loans greater than 15k granted to customer with income in the range of 0 to 50k (19%)



# RECOMMENDATION

## GOOD LOAN

	binned_int_rate	grade	term	binned_annual_inc	binned_loan_amnt	default_rate_mean	default_rate_count
275	15+	D	36	75k+	5k-10k	0.03	36.00
152	10-15	C	36	75k+	0-5k	0.03	67.00
179	10-15	D	36	75k+	5k-10k	0.04	47.00
7	0-10	A	36	50k-75k	5k-10k	0.05	682.00
11	0-10	A	36	75k+	5k-10k	0.06	315.00
27	0-10	B	36	0-50k	5k-10k	0.06	172.00
129	10-15	B	36	75k+	10k-15k	0.06	251.00
131	10-15	B	36	75k+	5k-10k	0.06	337.00
37	0-10	B	60	0-50k	10k-15k	0.06	32.00
9	0-10	A	36	75k+	10k-15k	0.06	173.00
31	0-10	B	36	50k-75k	5k-10k	0.07	196.00
5	0-10	A	36	50k-75k	10k-15k	0.07	306.00
4	0-10	A	36	50k-75k	0-5k	0.07	170.00
3	0-10	A	36	0-50k	5k-10k	0.07	595.00
33	0-10	B	36	75k+	10k-15k	0.08	66.00
35	0-10	B	36	75k+	5k-10k	0.08	92.00
29	0-10	B	36	50k-75k	10k-15k	0.08	118.00
25	0-10	B	36	0-50k	10k-15k	0.08	91.00
1	0-10	A	36	0-50k	10k-15k	0.08	200.00

## BAD LOAN

	binned_int_rate	grade	term	binned_annual_inc	binned_loan_amnt	default_rate_mean	default_rate_count
625	15+	F	60	0-50k	10k-15k	0.40	52.00
638	15+	F	60	50k-100k	15k+	0.40	257.00
626	15+	F	60	0-50k	15k+	0.39	41.00
627	15+	F	60	0-50k	5k-10k	0.38	52.00
562	15+	D	60	0-50k	15k+	0.37	115.00
594	15+	E	60	0-50k	15k+	0.37	107.00
560	15+	D	60	0-50k	0-5k	0.37	119.00
622	15+	F	36	50k-100k	15k+	0.37	38.00
542	15+	C	60	50k-100k	15k+	0.36	55.00
306	10-15	C	60	0-50k	15k+	0.35	142.00
593	15+	E	60	0-50k	10k-15k	0.35	147.00
572	15+	D	60	50k-100k	0-5k	0.35	52.00
565	15+	D	60	100k-200k	10k-15k	0.34	58.00
574	15+	D	60	50k-100k	15k+	0.34	357.00
639	15+	F	60	50k-100k	5k-10k	0.34	47.00
563	15+	D	60	0-50k	5k-10k	0.34	184.00
592	15+	E	60	0-50k	0-5k	0.34	131.00
529	15+	C	60	0-50k	10k-15k	0.33	33.00
670	15+	G	60	50k-100k	15k+	0.33	82.00
606	15+	E	60	50k-100k	15k+	0.33	474.00
595	15+	E	60	0-50k	5k-10k	0.32	161.00
305	10-15	C	60	0-50k	10k-15k	0.32	158.00
630	15+	F	60	100k-200k	15k+	0.32	114.00

# RECOMMENDATION

From the tables formulated in the last slide

GOOD LOAN - Loan which has a default rate of less than 8%

BAD LOAN - Loan which has a default rate of more than 30%

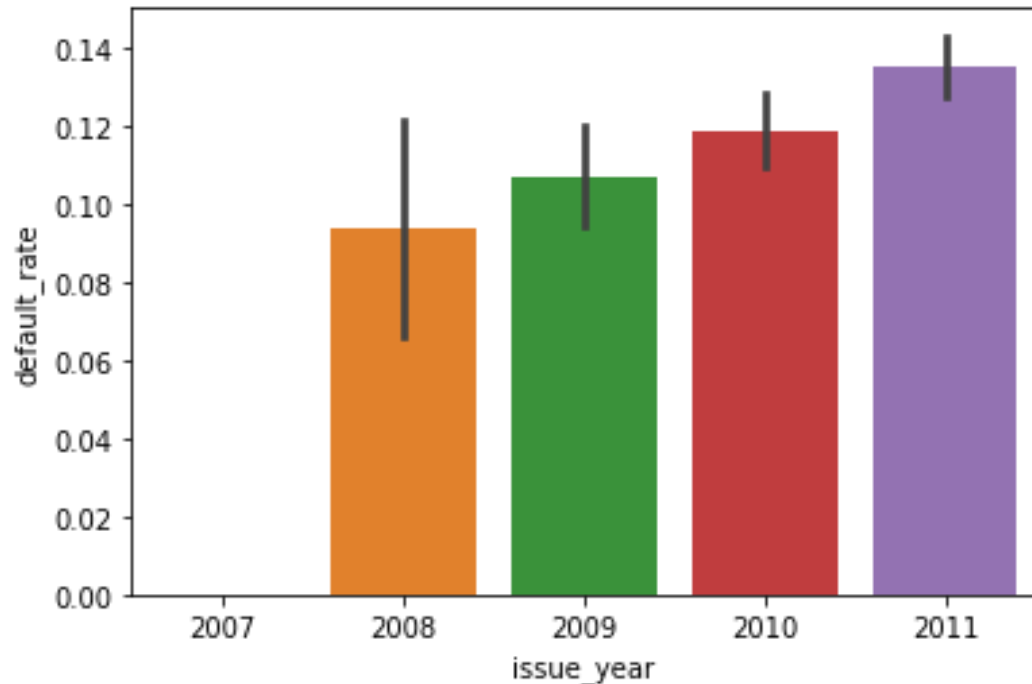
Good/Bad	GRADE	TERM	INTEREST RATE	LOAN AMOUNT	ANNUAL INCOME
Good Loan	A	36	less than 10%		
	B		less than 15%	5k -15k	
	C		10% - 15%	less than 5k	above 75k
	D		above 10%	5k -15k	above 75k
Bad Loan	C	60	10% - 15%	above 10k	below 50k
	D	60	above 15%	5k -10k	below 50k

We strongly recommend the club avoid giving loans to the people with combination in the Bad Loan category.

Also we found that there is high chance of not default for people who come under Good Loan category

# Interesting Fact

## RECESSION FACTOR



### Default by Year:

During the analysis it was observed that there was an upward trend year on year, going from 12% in 2009 to 16% in 2011 which may be due to the Recession.

This can be taken as separate case study to find whether

- People defaulted more due to Recession, or
- There was a Recession as more people Defaulted in those years.