



# **Lending Club Case Study**

---

- **Srinivasan P**

# Agenda

---

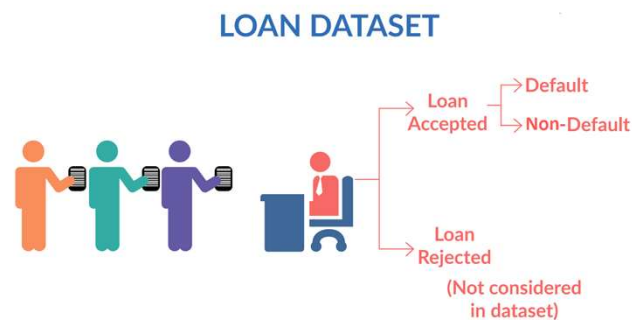
- Understanding the business domain and objective and its data structure
- Prepare the data for analysis
- Univariate analysis
- Bivariant analysis
- Summary of observations



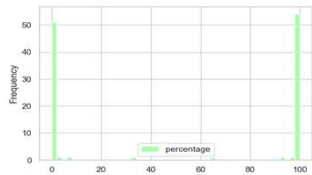
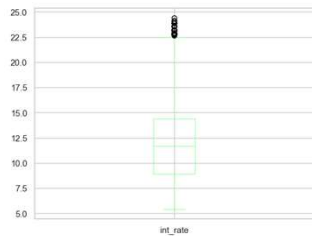
# Business domain and objectives

---

- Loan Accepted (Fully Paid, Current, Charged-off)
- Customers labelled as 'charged-off' are the 'defaulters'.
- Identify risky loan applicants (moving to Charged-Off state)
- Understand the driving factors behind the loan defaults



# Preparing the data



- There are many columns having null values between 10 to 100 %, columns with more than 10k null values may not add value to problem so lets list out those columns, which can be removed
- Check for columns having missing values and unique values, columns with one value would not be of any use in the analysis, so drop them
- Identify outliers in the variables and filter them out
- Check for ordered and unordered categorical variables like grade, grade\_sub, emp\_title, home\_ownership
- Derive new columns like from issue\_d we can derive, issue\_month and issue\_year.
- Convert the earliest\_cr\_line to days\_from\_earliest\_cr\_line format



# Univariate Analysis

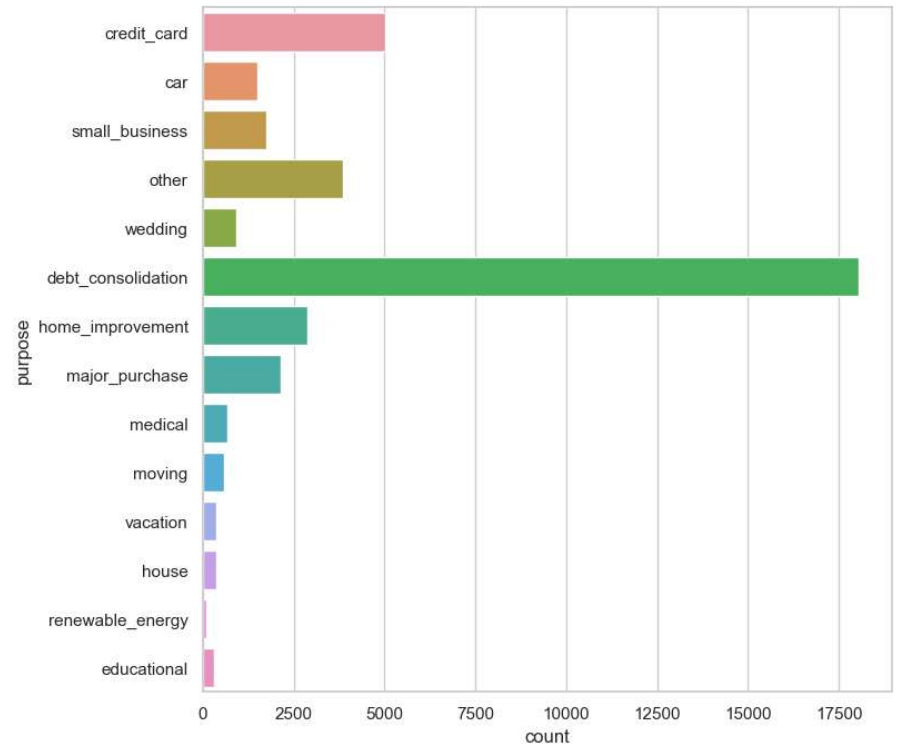
---

Distribution of variables and Ordered and unordered categorical variable analysis and Segmented analysis

# Purpose

---

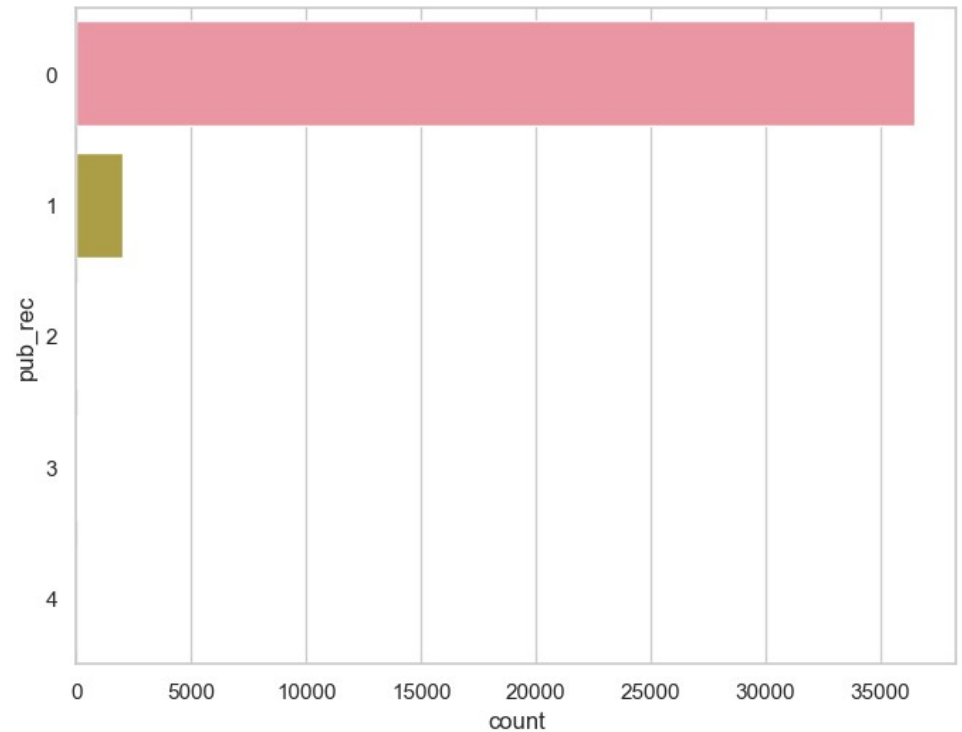
From the above analysis it is evident that the loans taken for small business, renewable energy and educational are the riskier ones



# Derogatory public records

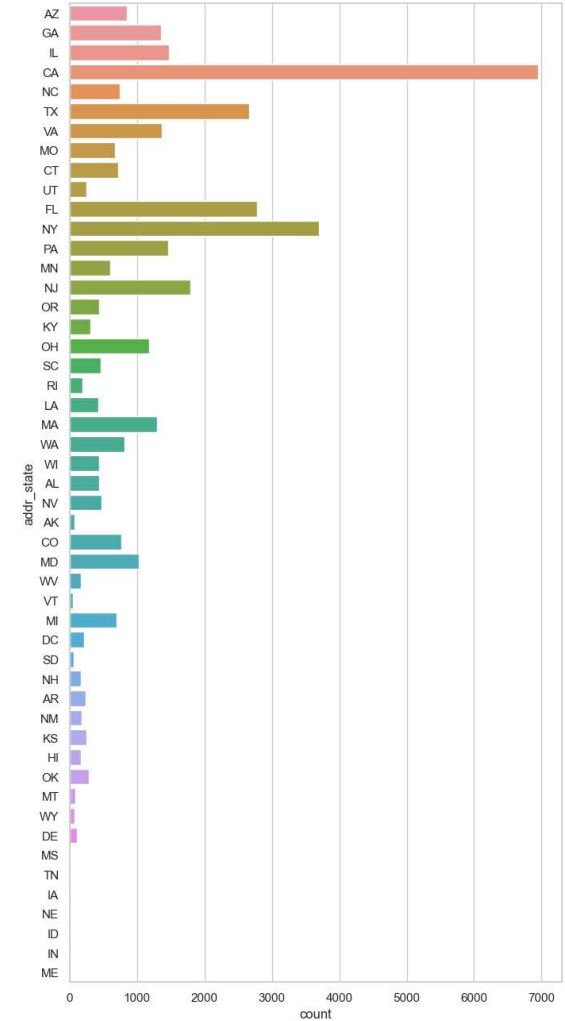
graph LR

graph clearly shows that individuals with non-zero derogatory public records have higher chances of charged off



# Addr\_state

From the above table we can see that the some of the risky states are NE, NV, SD, AK, FL, MO

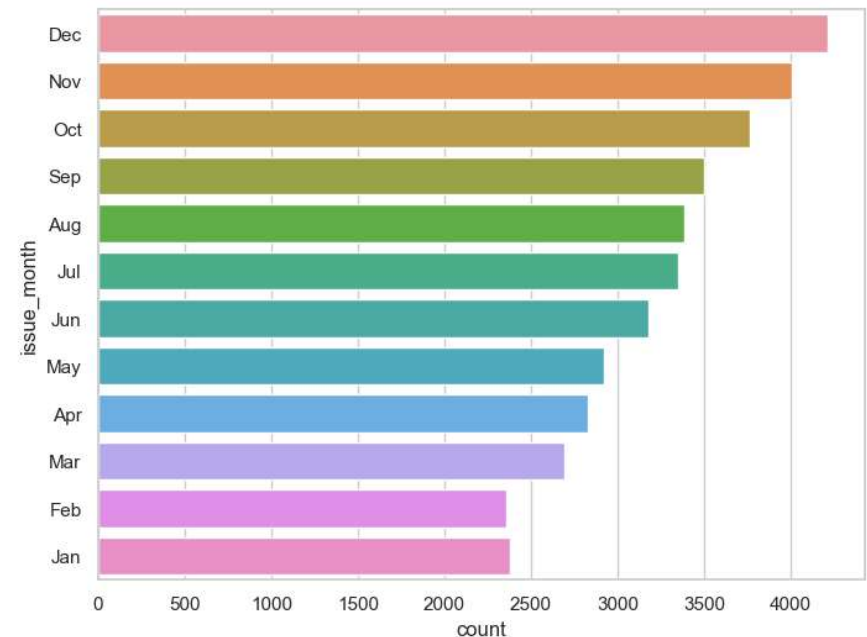




# Issue\_month

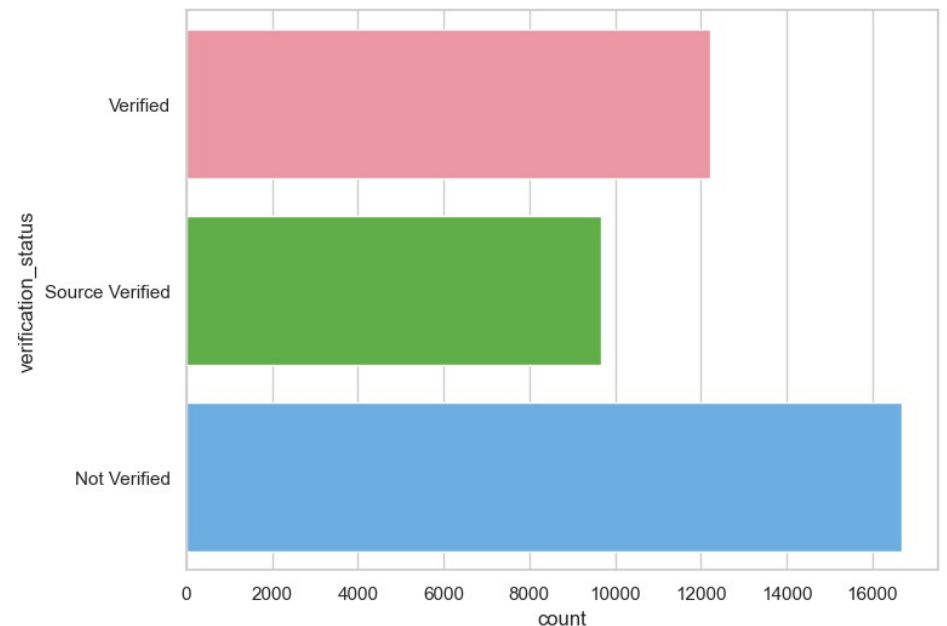


we can clearly see December is the month which has the highest number of loan applications per year and also have the biggest default ratio. May be people take loan for travel/party during Christmas and then are not able to pay back. Month of May is also another one, which is during the summer break and right before the Memorial day and Independence day breaks in US where people love to travel



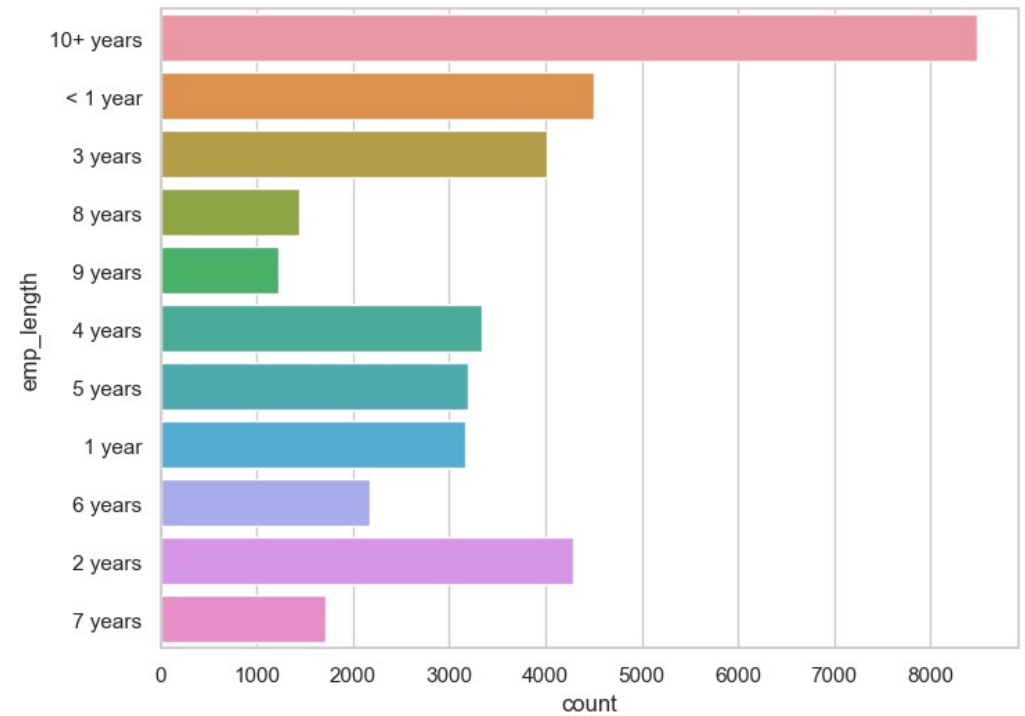
# Verification status

Interestingly verified applicants have more defaulted % which means the verification process is not correct.



# emp\_length

we can see having no employment will lead to Defaulted payment





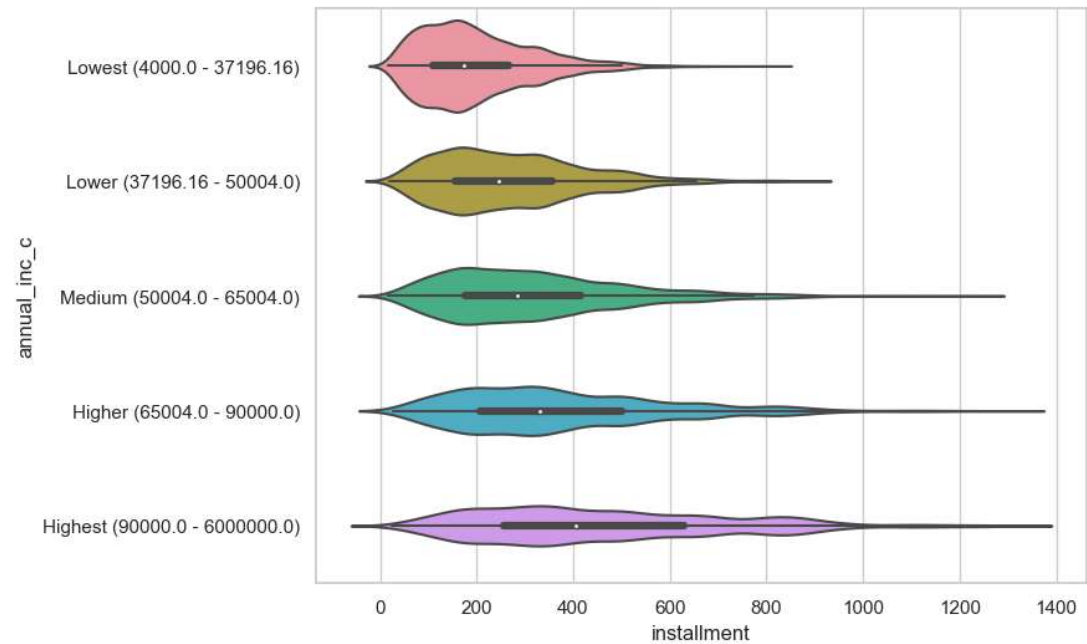
# Bivariate Analysis

---

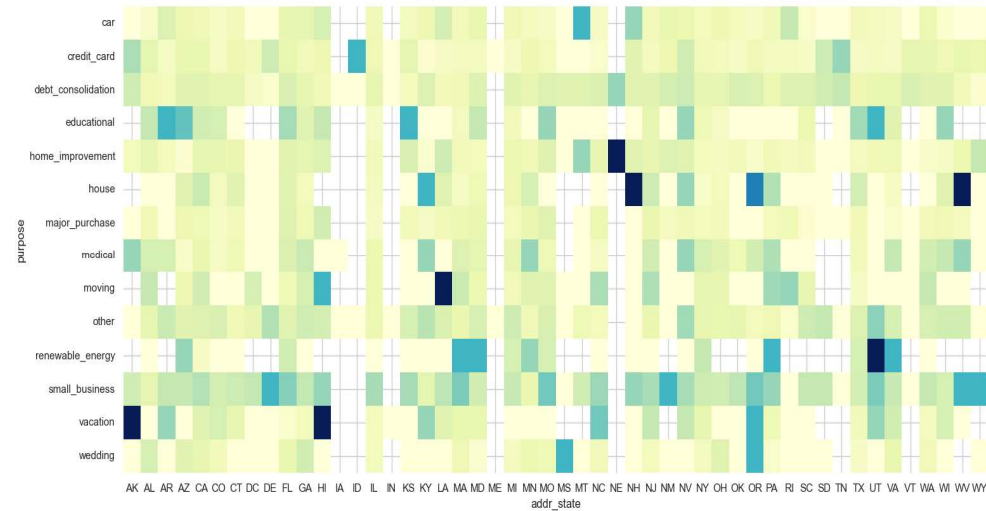
Correlation of variables

# Annual\_inc\_c and installment

Higher installments for any income group have more number of defaults.

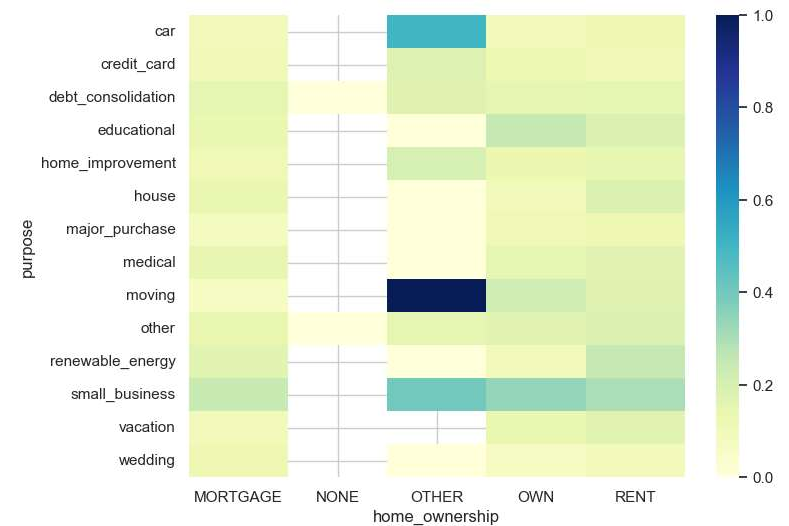


home\_ownership  
 -> purpose  
 addr\_state  
 -> purpose



As per the above plot, the darker the intersection of addr\_state has with the purpose of the loan, the riskier the loan application is. Some of the examples are below:

- vacation loans in AK, HI, OR
- education loans in AR, KS, UT
- small business loans in DE, NM, WV, WY
- wedding loans in MS, OR

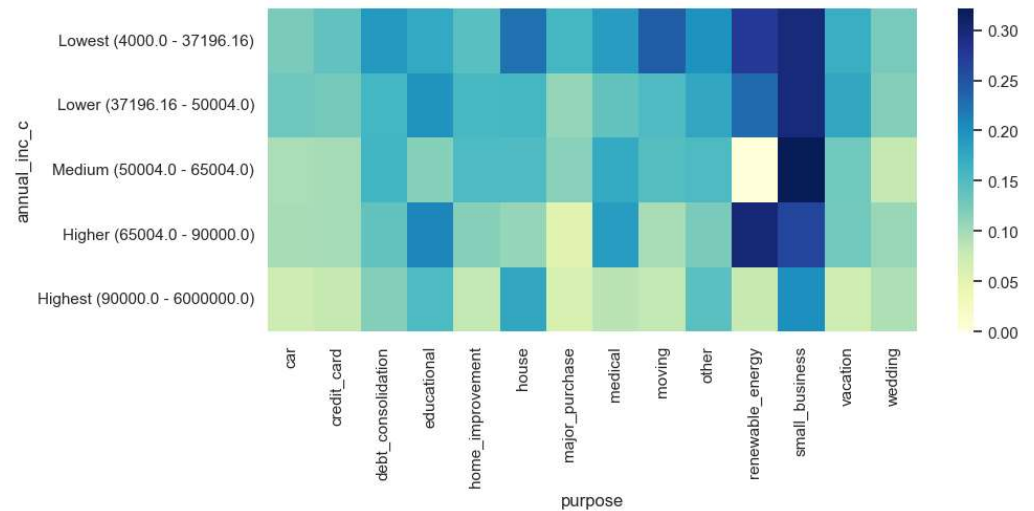


# purpose and annual\_inc\_c

\_\_\_\_\_

Plot of various income groups versus the risky purposes of loans for them. Some examples are:

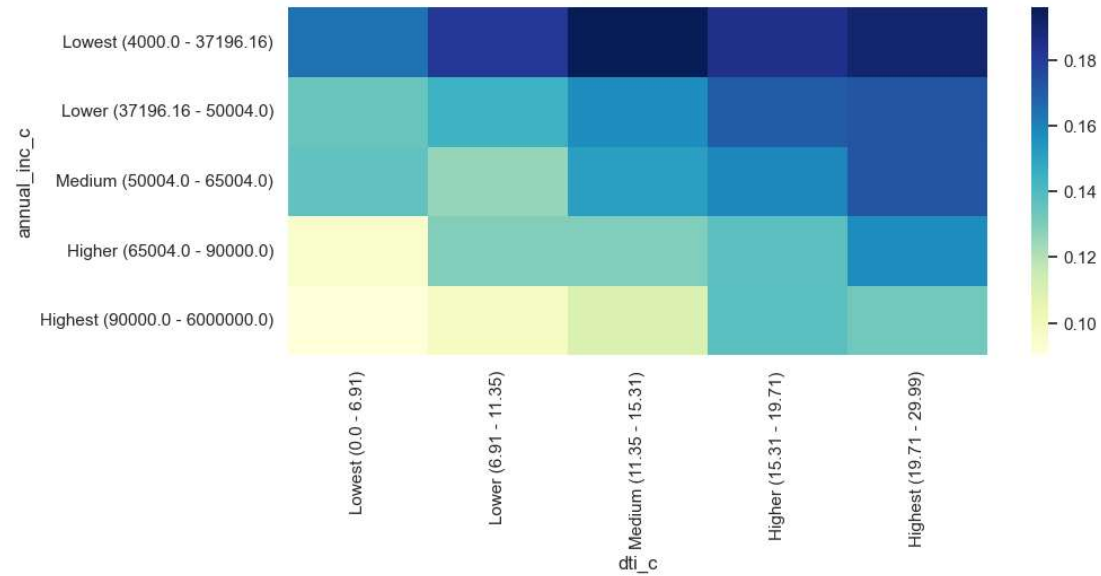
- small business loans for lowest and medium income groups
- renewable energy loans for higher income group



# purpose and annual\_inc\_c

\_\_\_\_\_

Medium debt-to-income group in the lowest income range is the most risky when it comes to loan repayment.





# Low and Medium: Risky loan applicants factor

---

## Low Factors:

.Higher loan amount (above 16K)

- Higher installment amount (above 327)
- Lower annual income (below 37K)
- Higher debt to income ratio (above 15%)
- Applicant's address state (NV, SD, AK, FL, etc.)
- Loan issue month (Dec, May, Sep)

## Medium Factors:

Higher interest rate (above 13%)

- Higher revolving line utilization rate (above 58%)
- Repayment term (5 years)
- Loan grade & sub-grade (D to G)
- Missing employment record
- Loan purpose (small business, renewable energy, educational)
- Derogatory public records (1 or 2)
- Public bankruptcy records (1 or 2)

# **OVERALL: Risky loan applicants factor**

---

- High loan amount & interest rate for lower income group
- High installment and longer repayment term
- Home ownership (other) and loan purpose (car, moving or small business)
- Residential state and loan purpose
- Income group and loan purpose

# Thank you

---

Srinivasan Paneer

9884454884

[srinivasan\\_.panneer@hotmail.com](mailto:srinivasan_.panneer@hotmail.com)

