Lending Club
Case Study

Presented By
Anuraj Kumar and Sunny Kumar

Problem Statement

As an online lending marketplace, company aims to :

- Identify 'risky' loan applicants who are likely to default and cause credit loss.
- Analyzing the driving factors behind loan defaults using exploratory data analysis (EDA).
- Understanding the key indicators of loan defaults, to minimize financial loss and for better risk assessment.

Approach

1. Understanding and Data cleaning
   a. Understanding what each columns represents
   b. Dropping columns which cannot be used, having null values, no variation etc.
   c. imputing missing values, removing duplicates, and correcting inconsistencies.
2. Analysing the data
   a. Univariate Analysis
   b. Bivariate Analysis
   c. Multivariate and segmented Analysis
3. Conclusion
   a. Drawing observations and recommendations based on the data

Analysis

Upon data cleaning we narrowed down our data to these columns, these will undergo further analysis to figure out if they are relevant:

- loan_amnt
- funded_amnt
- funded_amnt_inv
- term
- int_rate
- installment
- grade
- emp_length
- home_ownership

- annual_inc
- verification_status
- issue_d
- loan_status
- purpose
- zip_code
- addr_state
- dti
- delinq_2yrs

- open_acc
- pub_rec
- revol_bal
- revol_util
- total_acc
- total_rec_prncp
- last_credit_pull_d
- pub_rec_bankruptcies
- earliest_cr_line
- inq_last_6mths

Observation via graphs

Observation via graphs

Observation via graphs

Heatmap of numerical values
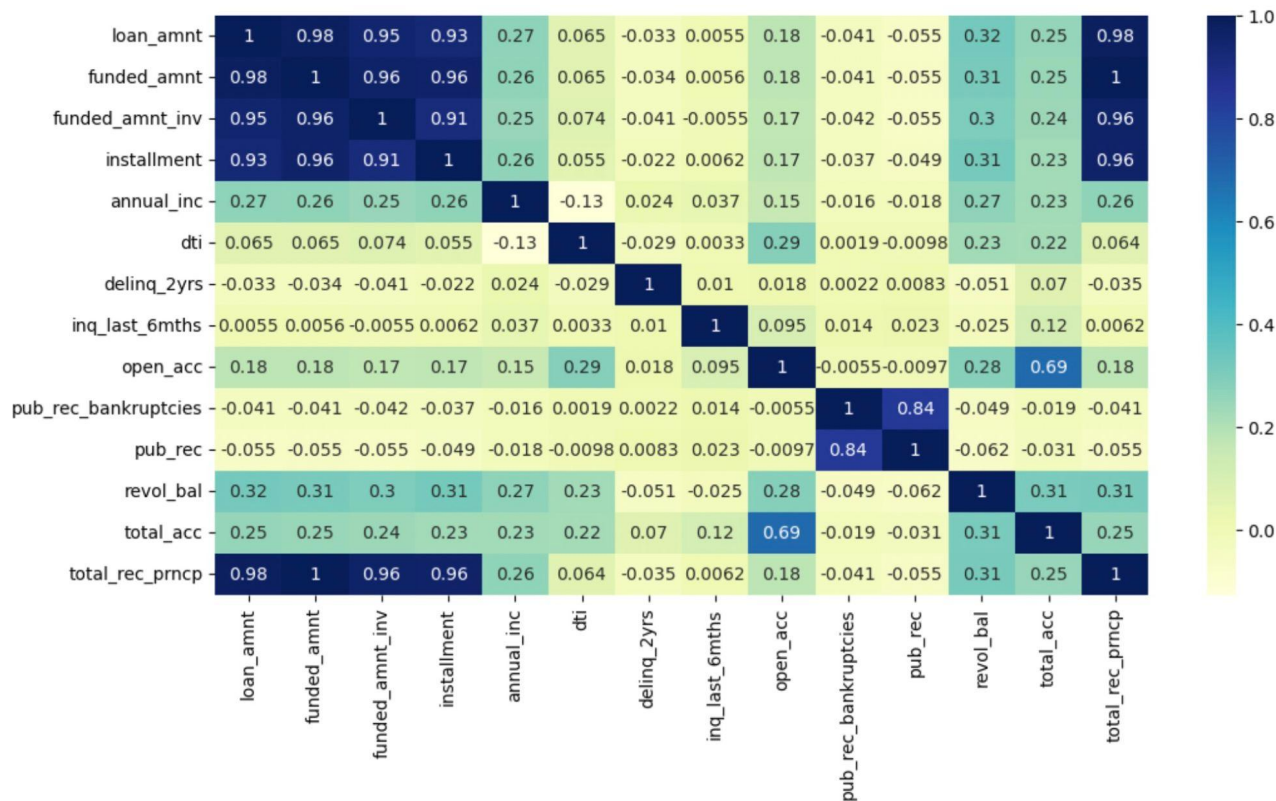
Recommendations & Conclusion

The following segments seem to have **strong** influence on defaults :

1. purpose (small business is more likely to default)
2. grade (lower grades are more likely to default)
3. addr_state (some states are more likely to have high defaults)
4. Home_ownership ()
5. Delinq_2yrs
6. Inq_last_6_months (probability of default increases with number of inq)
7. Term
8. Pub_rec
9. Pub_rec_bankrupcis
10. Dti
11. Loan_amnt
12. Int_rates
13.

Recommendations & Conclusion

The following segments seem to have **Low/No** influence on loan defaults :

1. Verfication_status
2. Emp_length
3. Home_ownership
4. Installments
5. Issue_d
6. Total_acc
7. total_rec