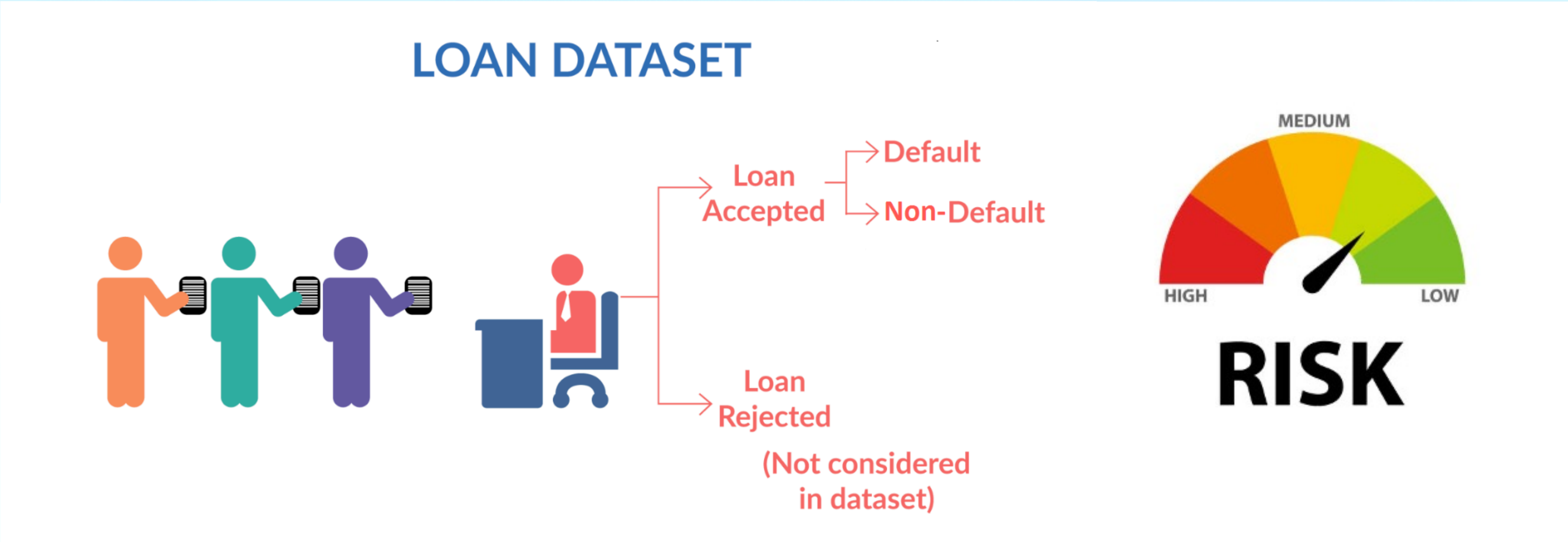
Lending Club Case Study

Driving Factors Behind Loan Default

Anirudh Gahlot & Minakshi Garg

Business Objective

- Understanding the driving factors behind loan default.
- Simplifying the loan application decision making process for the Lending Club



Data Understanding

1. Shape of dataset: rows-39717, columns-111

2. Columns with all NA values: 54 columns

3. Columns with same value: pymnt_plan, initial_list_status, mths_since_last_major_derog, policy_cd, application_type, dti_joint

4. Id columns: id, member_id, url

5. Behavior columns having no impact on lending decision: emp_title, issue_d, desc, earliest_cr_line, inq_last_6mths, mths_since_last_delinq, mths_since_last_record, open_acc, pub_rec, revol_bal, revol_util, total_acc, out_prncp, out_prncp_inv, total_pymnt, total_pymnt_inv, total_rec_prncp, total_rec_int, total_rec_late_fee, recoveries, collection_recovery_fee, last_pymnt_d, last_pymnt_amnt, next_pymnt_d, last_credit_pull_d, collections_12_mths_ex_med, chargeoff_within_12_mths, pub_rec_bankruptcies, tax_liens, title, delinq_2yrs

6. Interesting columns for lending analysis: loan_amnt, funded_amnt, funded_amnt_inv, term, int_rate, installment, grade, sub_grade, emp_length, home_ownership, annual_inc, verification_status, loan_status, purpose, zip_code, addr_state, dti

Data Cleaning and Manipulation

Removed unnecessary columns

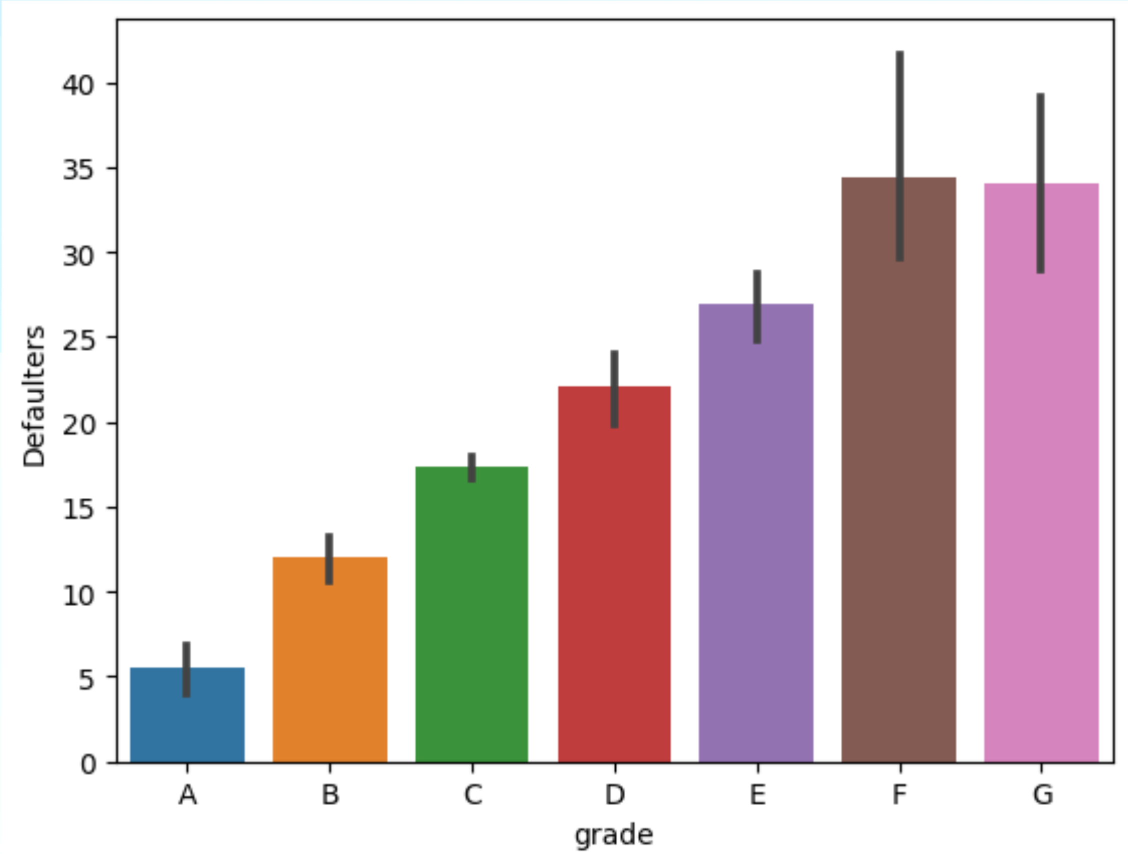
1. Columns with all NA values
2. Columns with all similar values
3. Columns with behavior values which are not helpful in forming lending decision
4. ID columns

Manipulation for analysis

1. Rounded column funded_amnt_inv to nearest integer
2. Removed % sign from int_rate and convert column to float type
3. Converted emp_length to continuous categorical column from 0 to 10

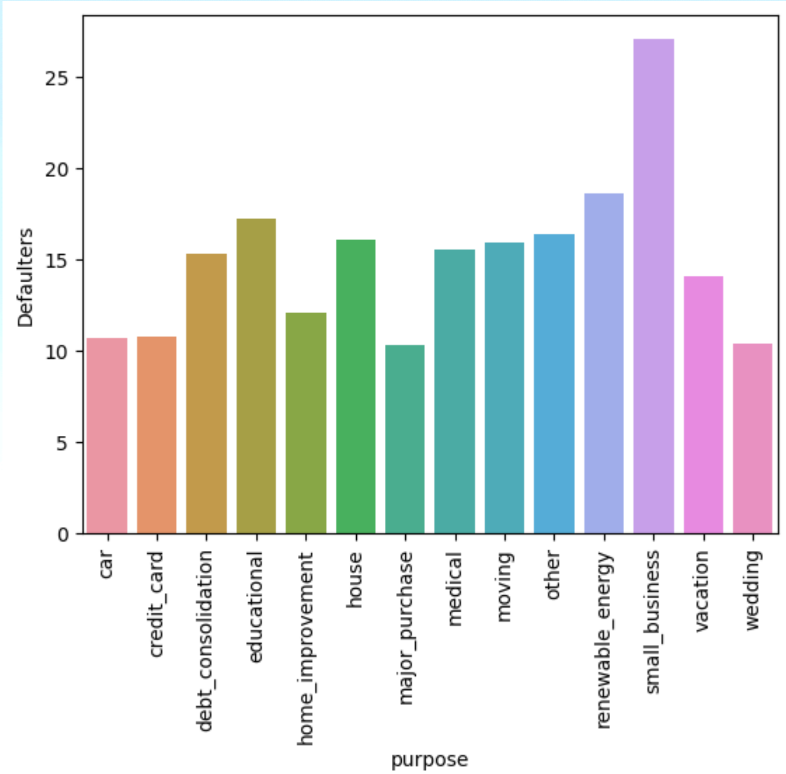
Analysis based on LC assigned loan grade

Individual with LC assigned loan grade F and G are more likely to default.



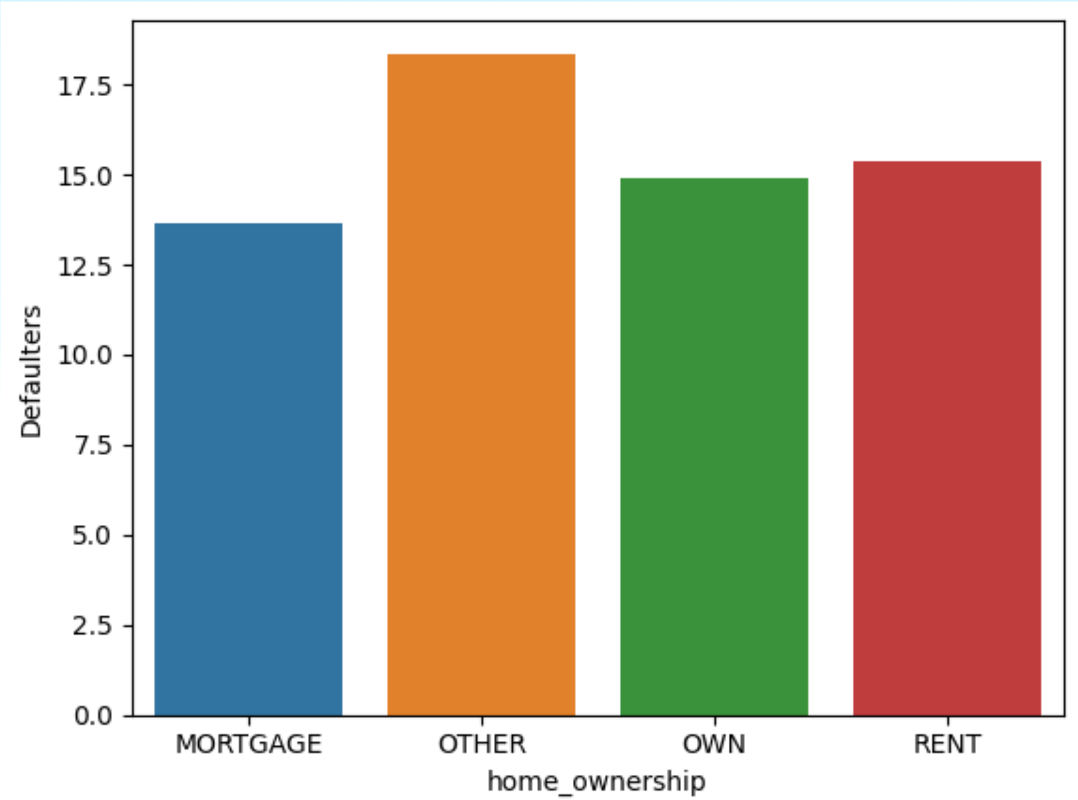
Analysis based on purpose of loan

Small business owners are more likely to default.



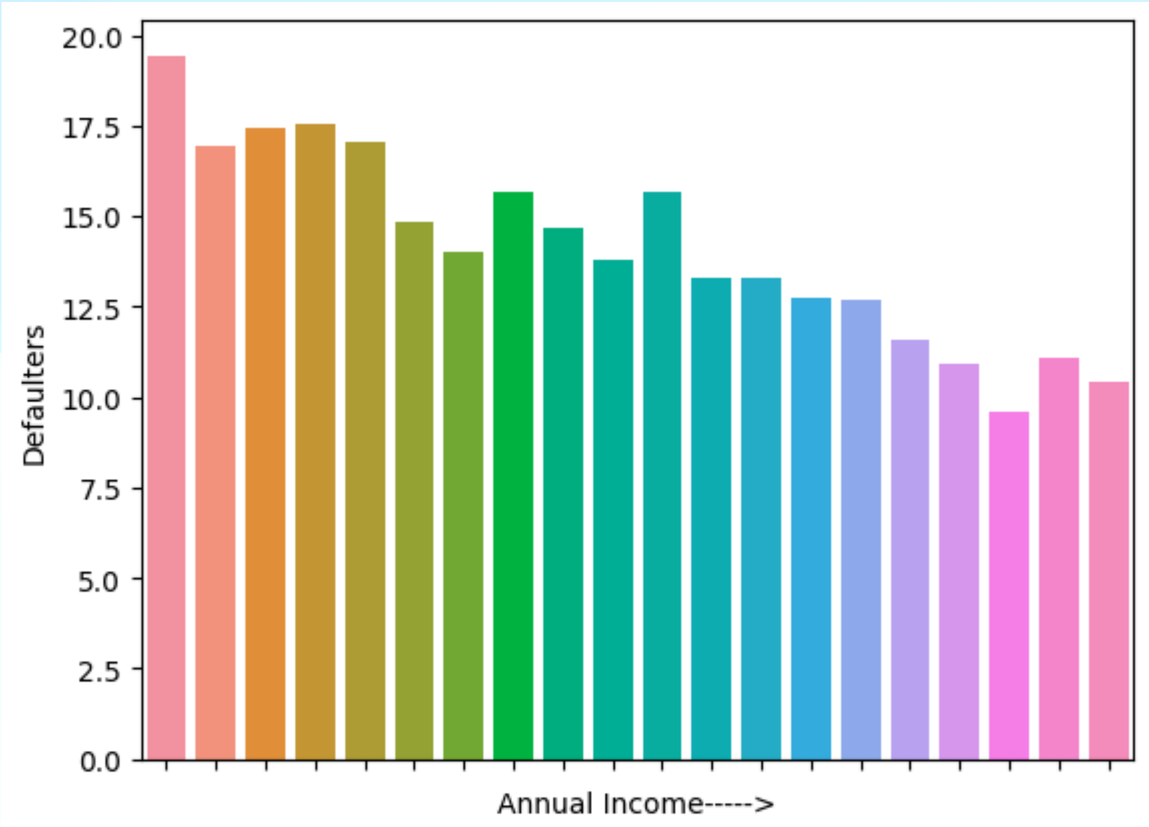
Analysis based on ownership of home

Individuals with no home ownership are likely to default



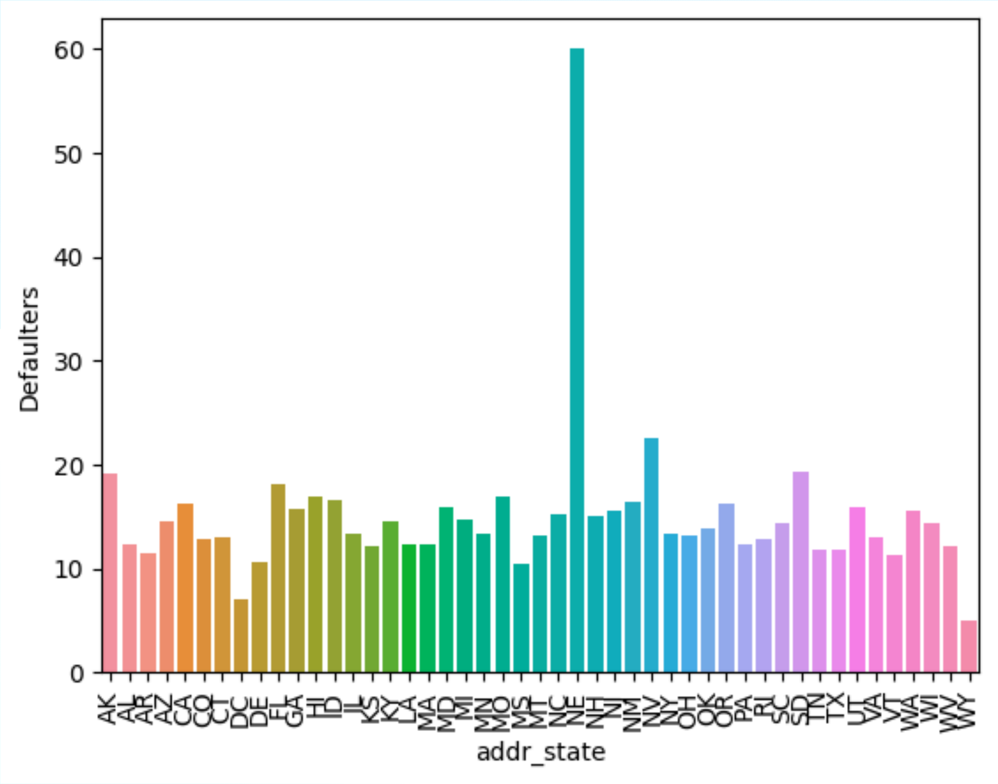
Analysis based on annual income

Individuals with low income are likely to default.



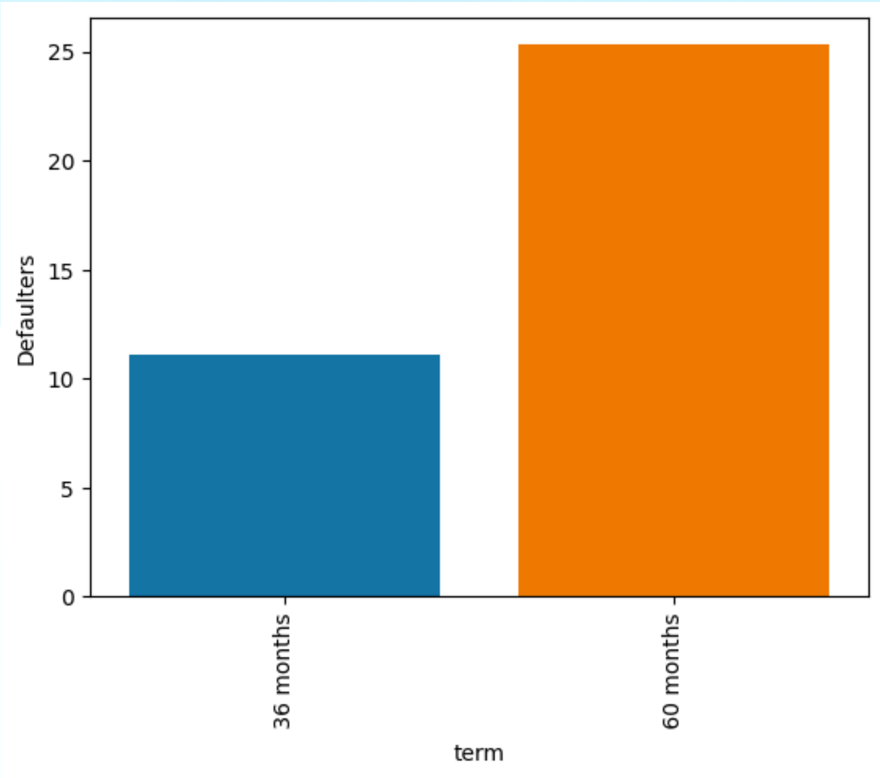
Analysis based on address state

Individuals living in NE address state are more likely to default.



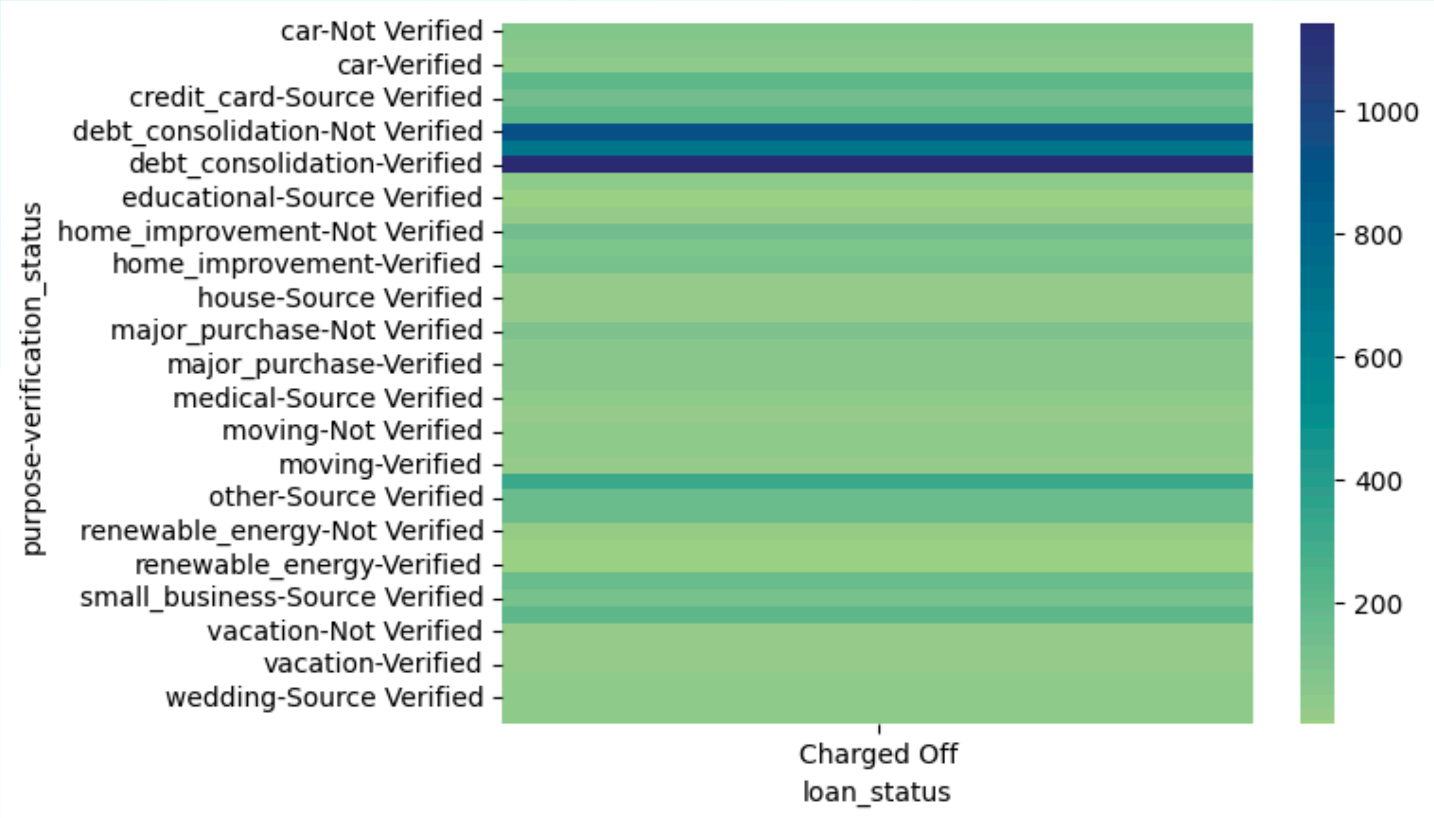
Analysis based on term of loan

Individuals opting for longer term are more likely to default



Multivariate analysis based on purpose and verification-status

Individuals with debt consolidation purpose are maximum defaulted.



Summary

- Individual with LC assigned loan grade F and G are more likely to default.
- Small business owners are more likely to default.
- Individuals with no home ownership are likely to default.
- Individuals opting for longer term are more likely to default
- Individuals with low income are likely to default.
- Individuals living in NE address state are more likely to default.
- Individuals with debt consolidation purpose are maximum defaulted.

While scrutinizing the loan application special attention should be given to above factors.