

Lending Club Case Study

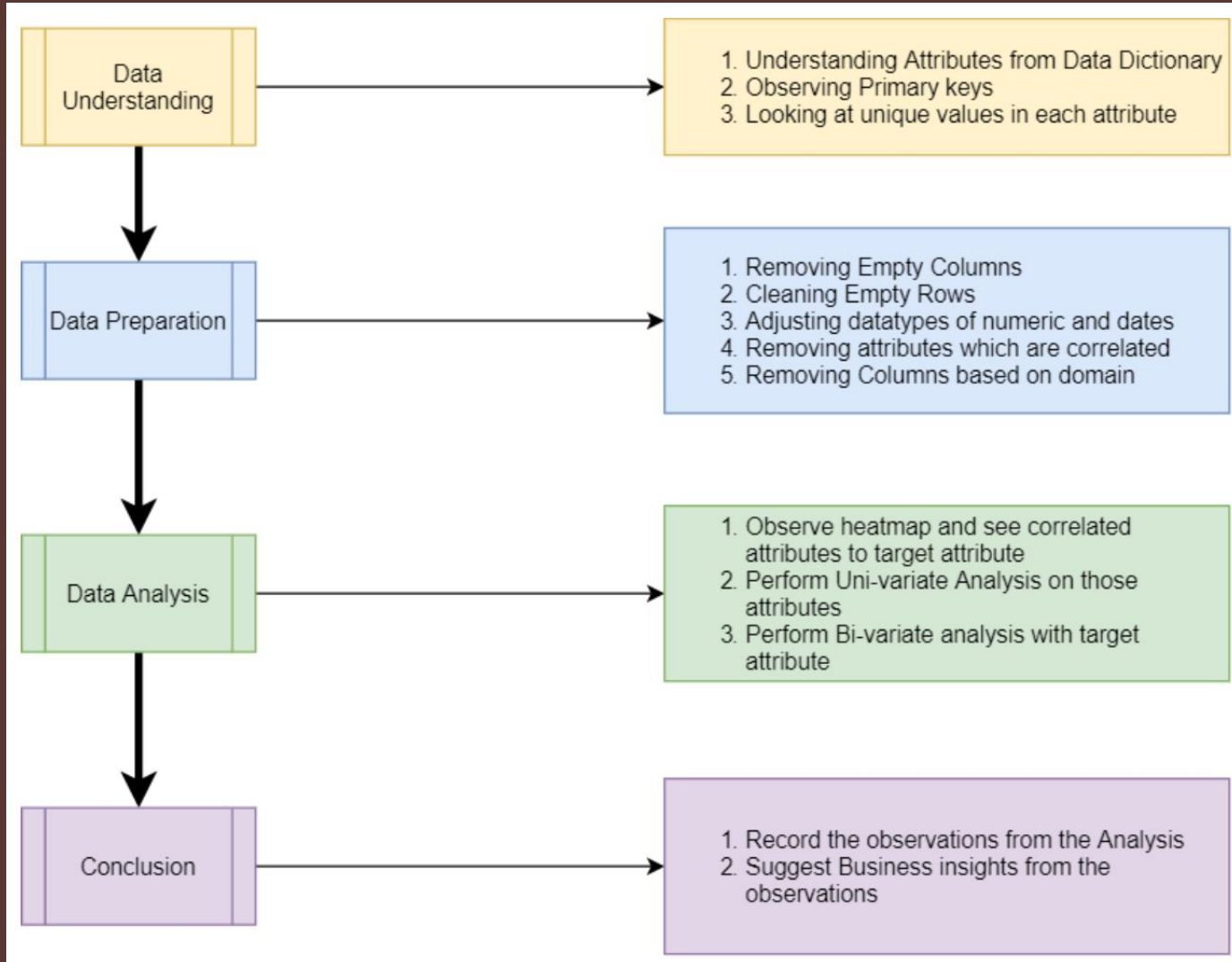
By Pooja Reddy Karnati



Objectives

- This company is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures. Borrowers can easily access lower interest rate loans through a fast online interface.
- The objective is to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default. The company can utilise this knowledge for its portfolio and risk assessment.

Problem Solving Methodology



Data Cleaning / Data Preparation

- Remove Columns with are empty
- Remove Columns that are redundant
- Fix Datatypes like Date and Numeric
- Remove percentage symbols and make them numeric
- Remove empty Rows as we want our analysis to be clean
- Remove correlated attributes as they add same redundancy
- Remove post loan approved attributes (Domain D riven)
- Create Derived Columns (Domain Driven)

Data Analysis

Univariate Analysis

- Perform Univariate Analysis on these 8 attributes from categorical and continuous variables with domain knowledge and correlation:
 1. Loan Amount (Most of the loan amounts are distributed between 5000 to
 2. Interest Rate (Most of the interest rate are distributed between 8.5% to
 3. Annual Income (Most of the applicants earns between 50,000 to 1,00,000
 4. Loan Status (14% of the applicants Charged
 5. Purpose of Loan (Approx. 40% of the applicants applied loan for paying their other loans(Debt Consolidation))
 6. Home ownership status (48% of applicants are living in rented home whereas 44% applicants were mortgaged their home)
 7. Year wise analysis (loan applicants are increasing year on year, approx 47% of loan applicants received loans in 2011)
 8. Loan Term (73% of applicants applied loan for 36 months term

Bivariate/Multivariate Analysis with Probability of Charge off

- Perform Bivariate/Multivariate Analysis on attributes with respect to the probability of being defaulted.
 1. Location vs Probability Charge off There are multiple States/Provinces with high probability of charge, highest being 'NV' at 21%
 2. Purpose of Loan vs Probability Charge Off Applicants who has taken the Loan for 'small business' has the highest probability of charge off of 25.6%. So bank should take extra caution like take some asset or guarantee while approving the loan for purpose of 'small business'
 3. Grade/Subgrade vs Probability Charge Off As we move from Grade A to G, probability that person will charged off is increasing.

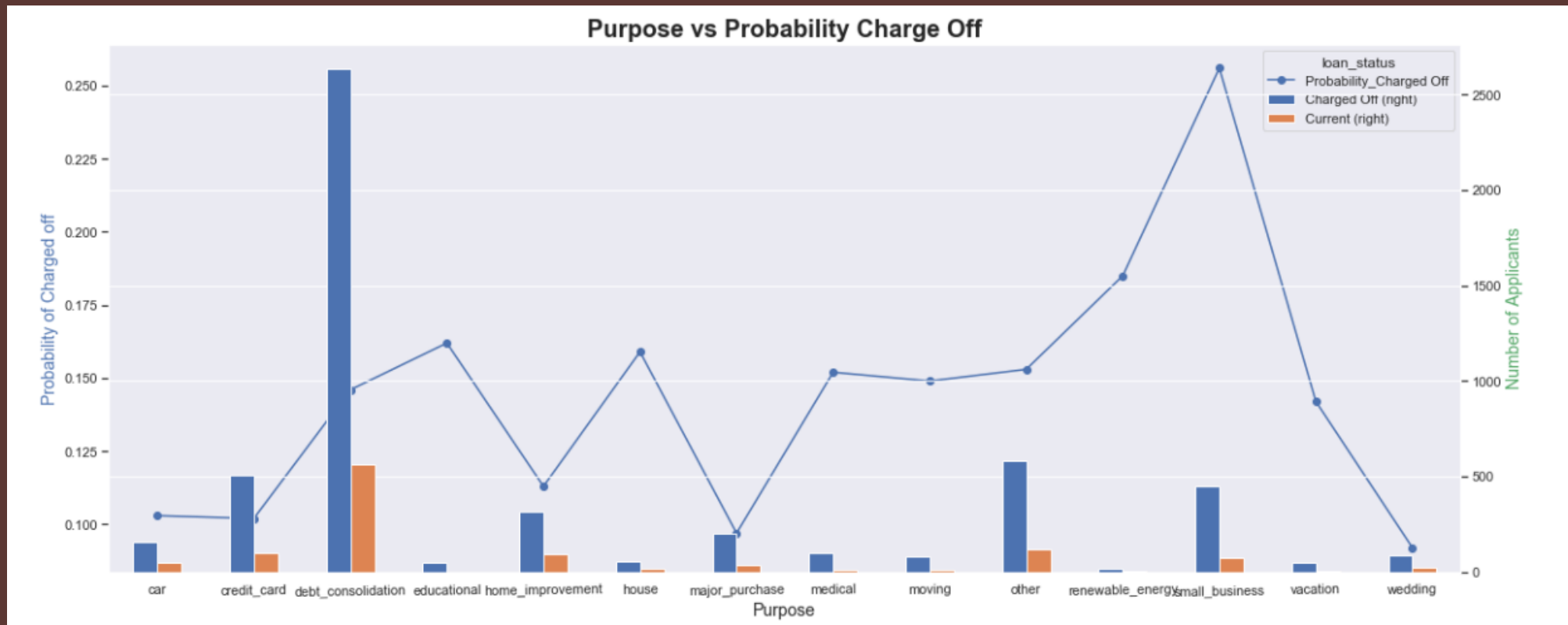
Bivariate/Multivariate Analysis with Probability of Charge off (continued)

- 4. Annual Income Range vs Probability Charge Off As the annual income is decreasing the probability that person will default is increasing with highest of 17% at (0 to 25000) salary bracket.
- 5. Interest rate Range vs Probability Charge Off As the interest rate is increasing the probability that person will default is increasing with highest of 23% at 15% & above bracket.
- 6. Employment Length vs Probability Charge Off Applicants who are 1 year or less than 1 year of experience are more probable of charged off aggregating to 26%

Graphs

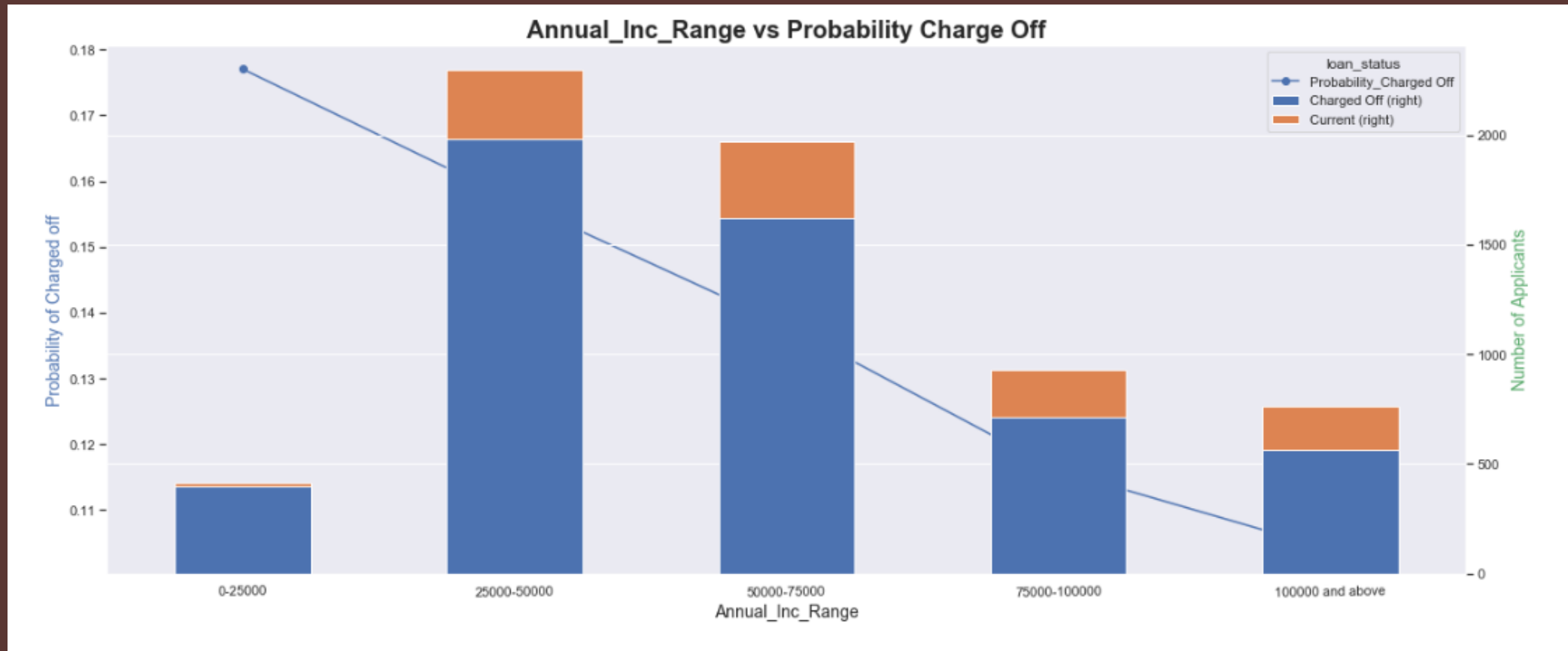
Refer Python Notebook for More

- Insights: Applicants who has taken the Loan for 'small business' has the highest probability of charge off of 25.6%. Therefore, bank should take extra caution like take some asset or guarantee while approving the loan for purpose of 'small business'



Graphs Refer Python Notebook for More

- insights: As the annual income is decreasing the probability that person will default is increasing with highest of 17% at (0 to 25000) salary bracket.



Conclusion

- From above Analysis, we can conclude the top factors for being defaulted or charged off are:
 1. Purpose of Loan
 2. Employment Length
 3. Annual Income
 4. Interest Rate
 5. Grade
 6. Location
 7. Term