Lending Club Case Study:

VIGNESHWAR MOHANASUNDRAM

SANTOSH KUMAR RAMARATHNAM

Problem statement

The finance association which provides various types of loans to different clients When the party accepts a loan application, the party has to resolve for loan authorization established the # claimant's profile. Two types of risks are guide the bank's conclusion,

- If the seeker is inclined compensate the loan, before not approving the loan results in a misfortune of trade to the association
- If the candidate is with difficulty to restore the loan, i.e. he/she is inclined default, before authorizing the loan concede possibility bring about a financial misfortune for the guest
- The applicant likely beneath holds the information about past loan claimants and either they 'defaulted' a suggestion of correction.
- The aim search out recognize patterns which display if one is inclined default, that grant permission be secondhand for taking # conduct to a degree declining the loan, lowering the amount of loan, loaning (to dangerous applicants) at a bigger interest, etc.

Analysis approach

Data cleaning:

- Analyse the whole data check for null value columns and look for any deprecation in actual data.
- Filter out null or invalid column which are not much helpful for analysis.

Eg: we have cleared around 54 null valued column like

mths_since_rcnt_il,total_bal_il,il_util,open_rv_12m,open_rv_24m,etc..,

- Modifying special characters in the values to analytical data for best understating.

Eg: Removing % symbol in revol_util at the end.

Analysis approach

- Examine the values to check whether the type assigned is suitable for analysis, if not convert it accordingly based on input criteria.

Eg: emp_length has been converted to int64 for analysis.

- Analyse the whole data check for null value rows and look for any deprecation in actual data.
- Filter out null or invalid column which are not much helpful for analysis.
- Remove inappropriate columns and rows which diverts the analysis.

Eg; Dropped zipcode column which had incomplete values.

Insights

Univarient analysis:

1. Annual income is plotted using box plot, It has been identified with outliers.These values can tilt our analysis and often provide us with a biased perspective of the data available.
2. The spread of a annual income values are spotted enormous thus it has been removed.
   Eg: Removing outliers we could see 50 percentile is around 57000.
3. Analyzing loan amount, funded amount and funded amount invested in subplot we could see loan amount request and fund amount along with funded amount invested looks similar.
4. Rate_of_interest has been plotted in distplot, shows most of the Interest Rates on loans are in range between 10 – 15%.

Insights

Bivarient analysis:

1. On deriving boxplot with variables of grade and loan amount, people with lower grade like g or f more likely prefer for larger loan amount
2. Accompanying the middle loan amount for a grade G loan being nearly 10000 above that of a grade A, B, or C loan.
3. Calculate chargedoff ratio by dividing chargeoff value with sum of all three purpose.
4. Using linechart with variables purpose and charge off ratio we could clearly see small_business takes up high chargedoff ratio. It is highly risk to provide loan to small_business.

Insights

Multivariate analysis:

1. On deriving heatmap with some integer values like loan_amnt, int_rate, installment,emp_length, annual_inc,dti, delinq_2yrs, inq_last_6mths, open_acc, pub_rec, revol_bal, revol_util,total_acc, last_pymnt_amnt, pub_rec_bankruptcies below points are observed.

- on comparing int_rate with revol_util the corr value is around 0.4 therefore investors/bank can gain more profit based on high interest rate.
- on correlating the last_pymnt_amnt with pub_rec_bankruptcies, we can say if last payment amount is low then there is high probability of bankruptcies.
- as expected when annual income increases then loan amount is also increasing.
- on seeing heat map, it is clear that whenever the loan amount is higher then number of installment is also in propotion.
- pub_rec is directly propotional to pub_rec_bankruptcies with corr value of 0.84, if public record is higher then risk of defaulters will also be higher.

Thank you