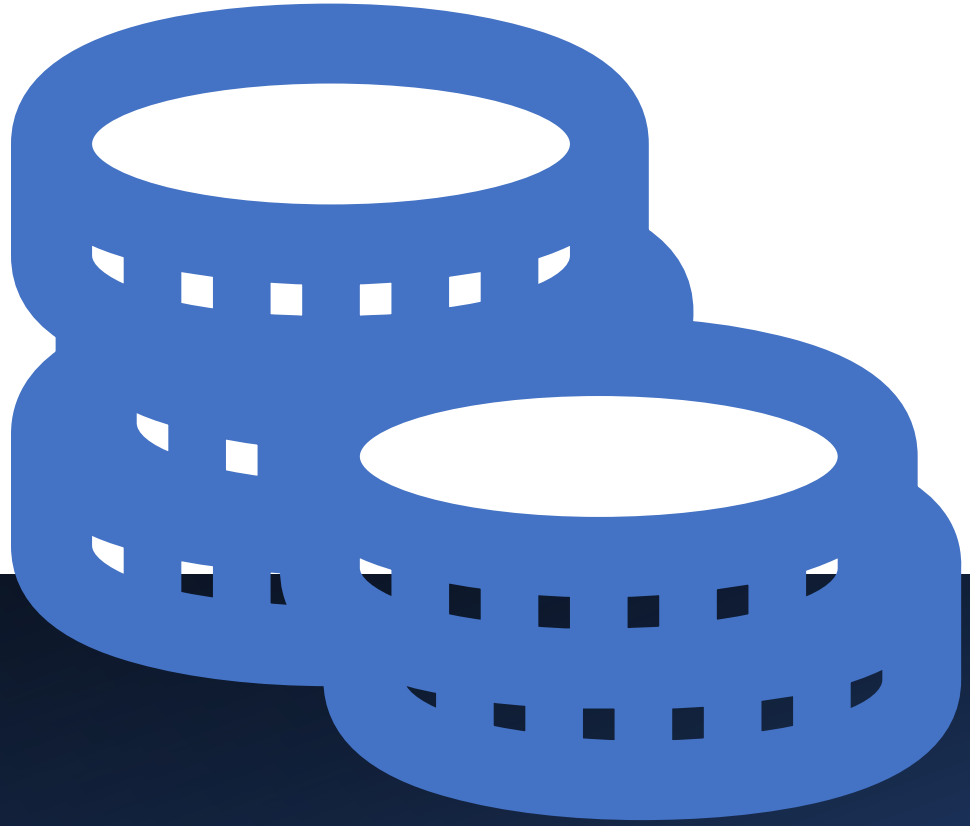
Lending Club Case Study



By

Damodar V Avadhani

Mallikarjuna

Business Problem

- 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 company wants 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.

Univariate and Segmented Univariate Analysis

We used univariate analysis to clean the data. We analyzed columns with null values and removed those with all nulls. We also removed columns with single values that would not contribute to our analysis.

Using univariate analysis, we also removed the outliers for loan amount and annual income columns. The reason being that it is very unlikely that people might have suck high incomes or request for unusually high loan amounts.

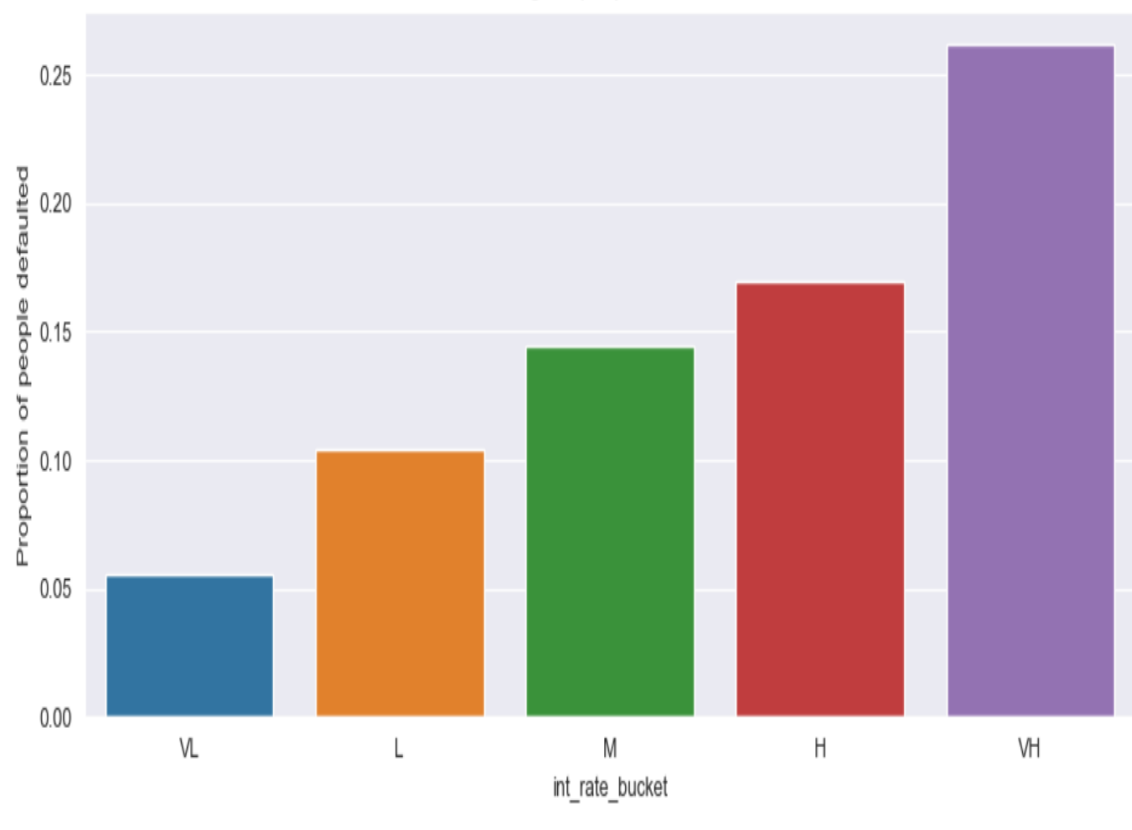
There is an imbalance in the dataset, which is the dataset has around 5000 rows of data for people who default. But 30000 rows for people who paid the loans.

Hence, it would be impractical to do analysis based on the number of people defaulted, instead, we will do analysis based on the proportion of people who have defaulted. This would be segmented univariate analysis

Convention

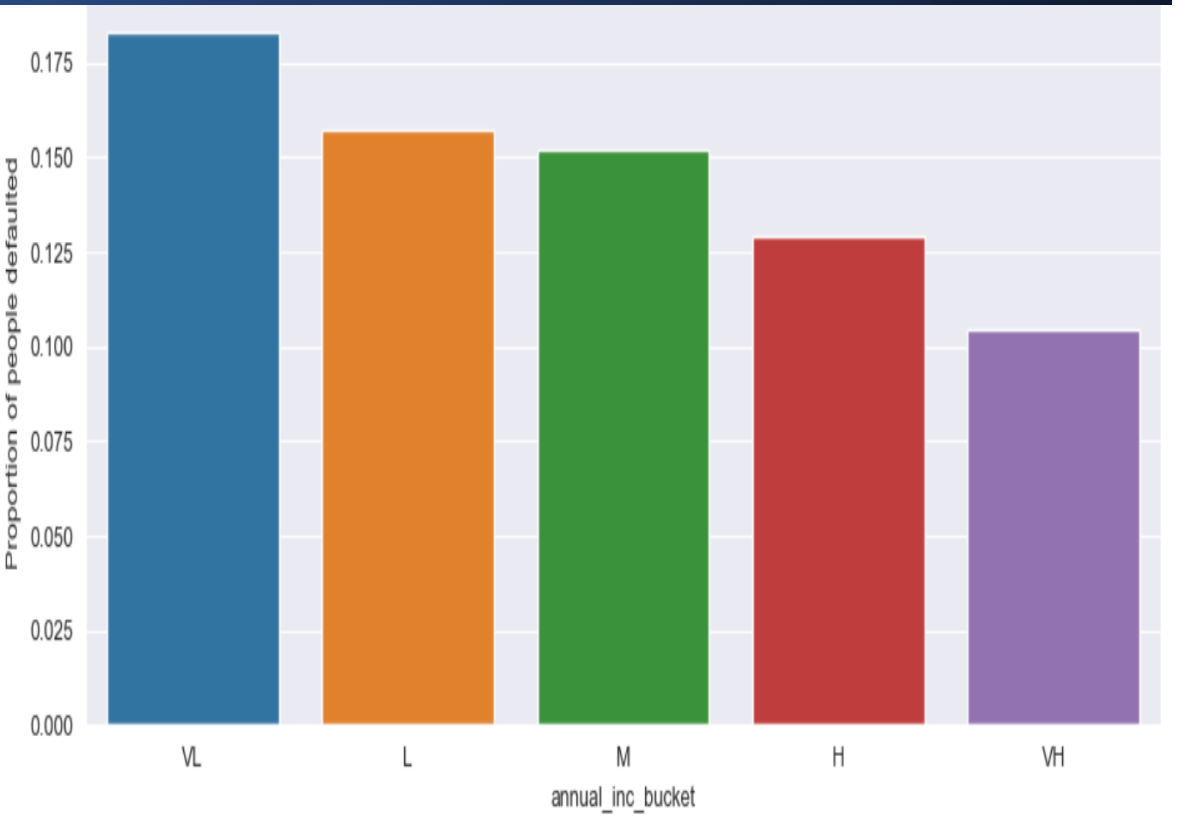
- We convert numeric variables into 5 categories from Very Low(VL), Low(L), Medium(M), High(H), Very High(VH). The division is made for every 20% quartile.

Interest Rate



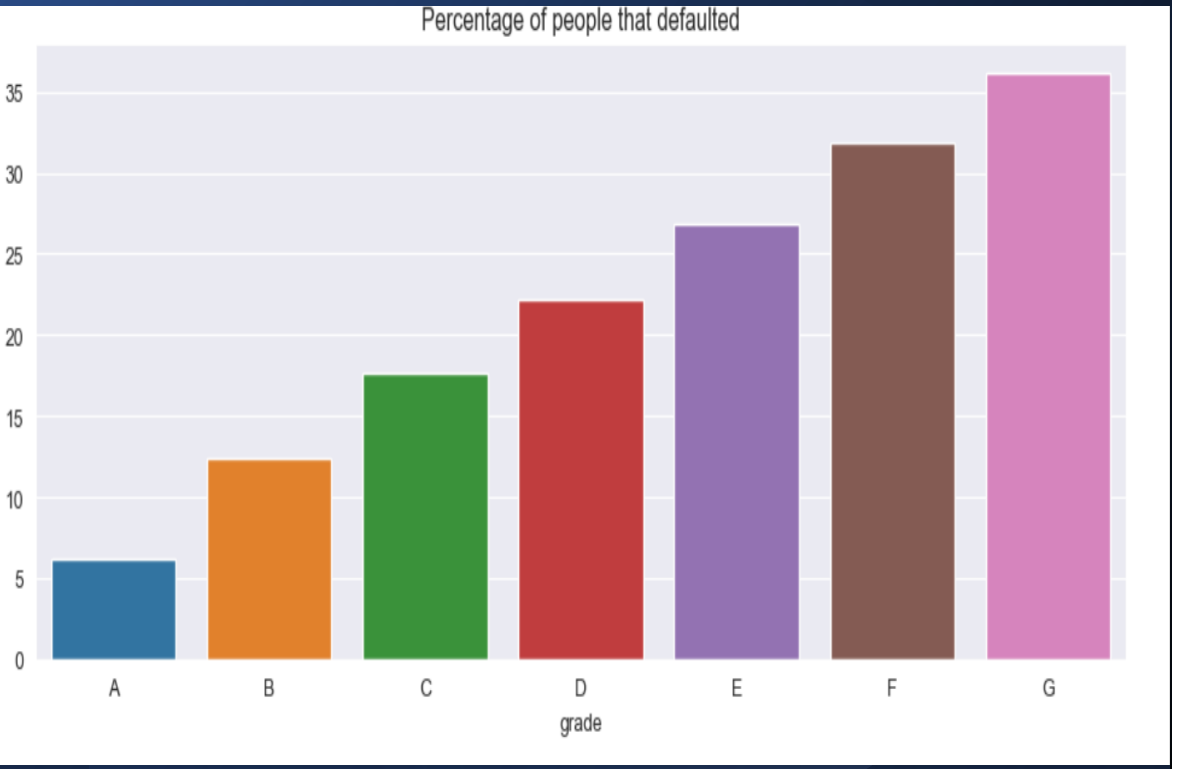
- On analysis we observed that higher interest rate on any loan amount has a higher probability of default.
- The IQR for people who default is 11-16%, while it is 8-13% for people who pay their loans.
- The graph also shows that as the interest rate range increases, the percentage of default also increases.
- The banks should not pose outrageous interest rates on customers in the hope of making business. This can lead to default which is one of the major loss-making reasons. Reasonable interest rates need to offered to ensure that the loan is repayable. Based on data, an interest rate of 10% seems to be fair.

Annual Income



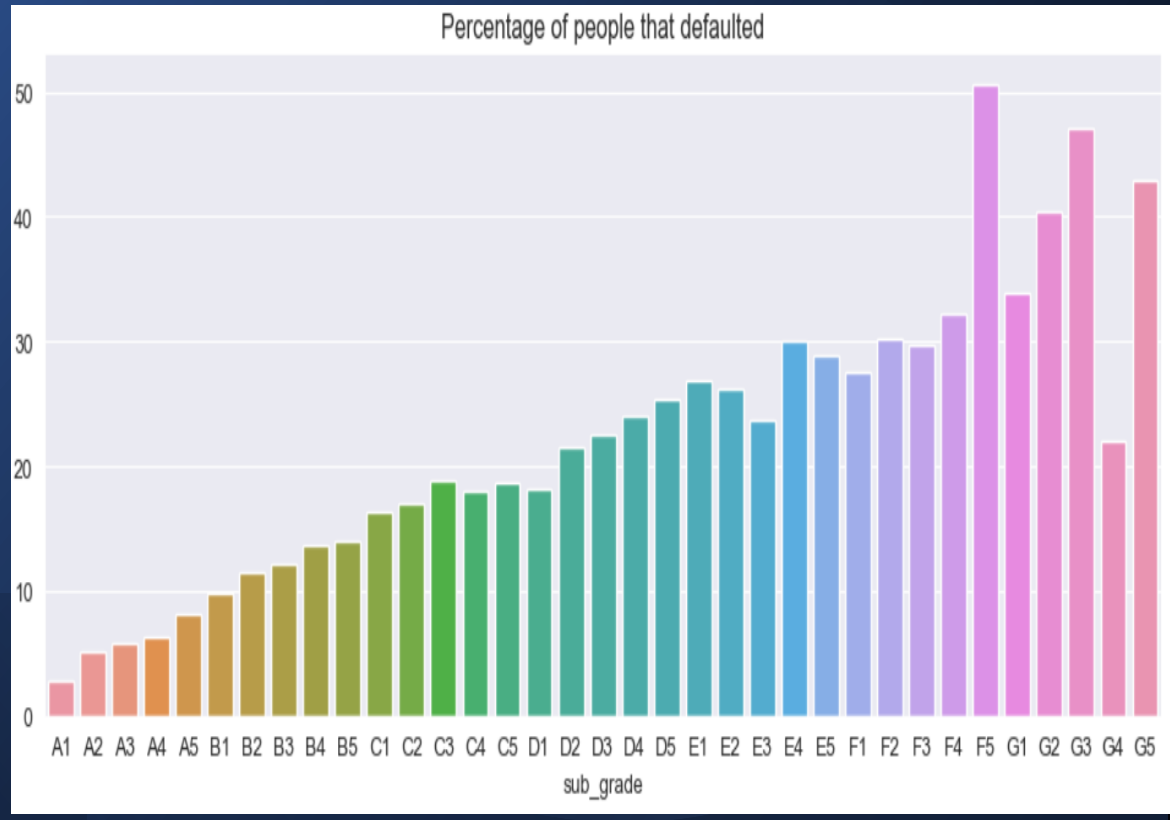
- On analysis we see that people with higher annual incomes can repay their loans and lower annual incomes struggle to repay and default.
- The median annual income for people who default is 50000$. For people who have paid their loans it is 56000$.
- The graph also shows that with the increase of annual income the proportion of people who default decreases.
- Provision of loans to customers with low annual income can result into default. Banks must provide loans to people with higher annual incomes. People with lower annual income might not be able to save enough to pay the loan back.

Grade



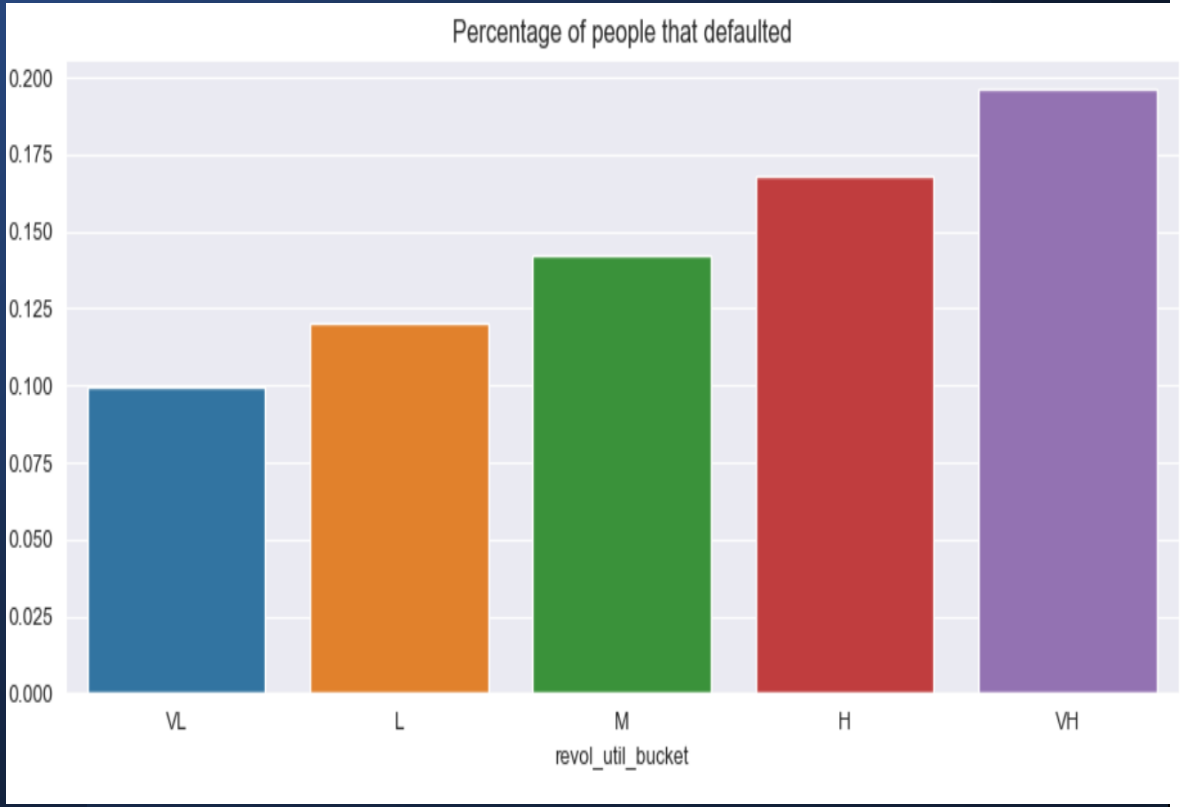
- On analysis we see that, higher the grade alphabetically, higher the risk of default.
- The graph also shows that as the grade varies from A->G, the proportion of people that default also increases.
- Banks must be very careful while lending out loans to high-risk customers. Especially while lending high loan amounts to them. Grade is typically calculated based on financial stability and credit worthiness of the customer. If the customer isn't financially stable, higher their grade and higher the risk of default.

Sub Grade



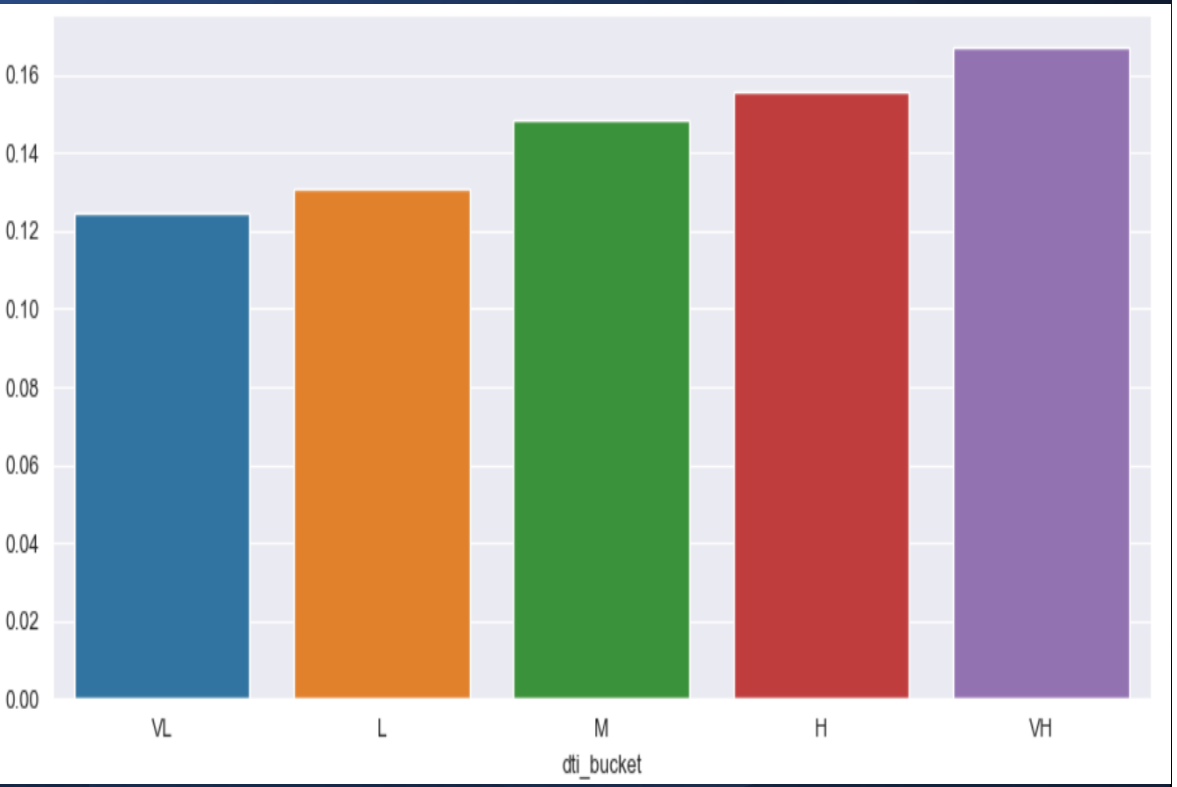
- On analysis we see that, higher the subgrade within a grade numerically, higher the risk of default.
- Although there might be a few exceptions on the graph shown. The trend is still convincing enough.
- A5 is riskier than A4 is riskier than A3 is riskier than A2 and so on.
- The recommendation is like that of grade. Minimize loans to customers with higher grades as they might be financially unstable.

Revolving Credit Utilization



- Revolving utilization compares the balance on each of your credit cards to your credit limit. So higher your credit utilization, the more you are in debt and the assumption is the more difficult for you to repay your loans. The data too proves the same.
- The graph on the left shows that an increasing credit utilization leads to a higher proportion of defaults.
- The median credit utilization for people who have defaulted is 58% and people who have paid their loans is 47% which is a significant difference.
- For an individual, a bad credit utilization might mean bad spending habits. For an organization it might mean that it is not as profitable. In both these cases the bank must minimize the provisioning of loans as these individuals and organizations might have hard time repaying these loans.

Debt-To-Income Ratio

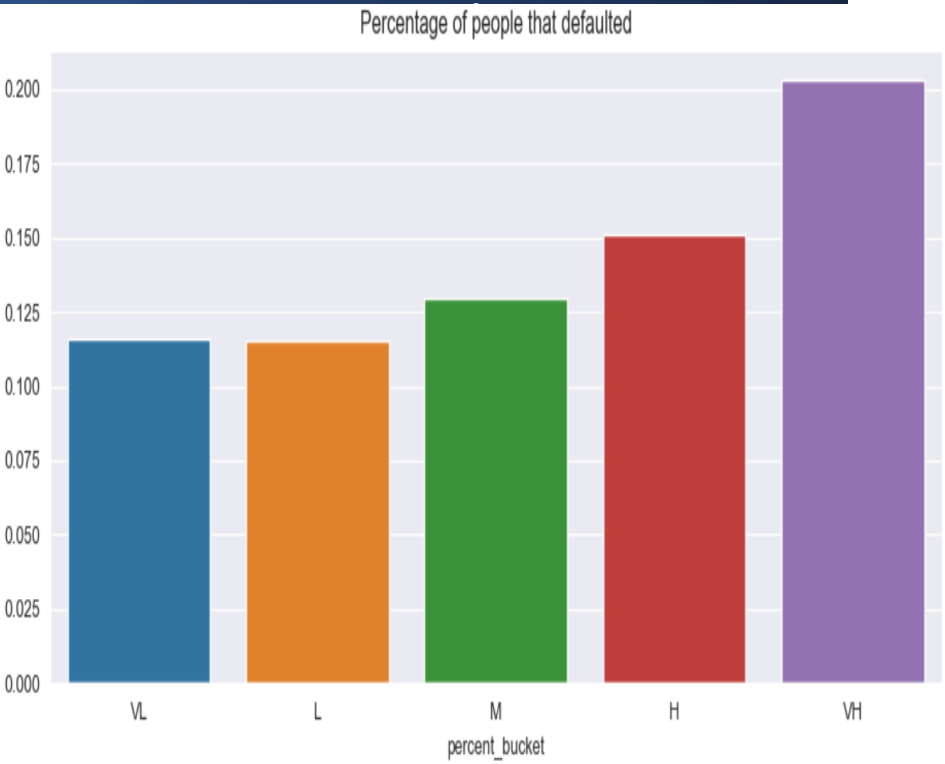


- Debt to Income ratio compares how much you owe to how much you spend every year. A low DTI means that your earnings are much high er than your expenditure.
- The graph shows as the debt-to-income range increases the percentage of people who default also increases.
- A high DTI might mean higher bills owed and expenses made. This might also be an indicator of already existing debts. A high DTI might be risky and can probably lead to defaults.

Driven Metrics

- There are 2 possible derivable metrics here: -
  - Percent of income provided as loan (Data driven metric) – We expect to see that expressing a new column that holds the value (loan_amnt/annual_inc)*100 might be helpful to analysis. Providing a large percentage of loan amount with respect to annual income might not be a good idea.
  - Date – Derive month, day and year from the loan issued date. This is a type driven metric.

Percent of loan amount wrt



- On analysis we see that, the larger the percentage of loan amount with respect annual income, the more difficult it might get for the customer to repay the loan. The graph shows the same.
- The business recommendation here is that the bank can strategize as to what is the ideal loan amount to annual income ratio which is safe to maintain, and they must ensure that this amount is not exceeded.

Bi-Variate analysis

- Bi-Variate analysis was done on a lot of variables. In the next slide we will not only the most crucial observations based on the analysis done.

Observations

- **Interest Rate vs Annual Income**
  - Interest rate is inversely proportional to the annual income. Lower annual income and High interest rates result in defaults. Higher annual income, lower interest rates can repay the loan.
- **Interest Rates vs Grade**
  - Higher interest rates on customers with riskier grades or sub-grades can lead to a high proportion of defaults.
- **Interest Rates vs Purpose**
  - Most small businesses are not able to repay the loans when interest is more than 10.75%.
  - very high interest rates for educational and house loans that are greater than 14.96% have a very high rate of defaults.
- **Annual_Inc vs Term**
  - For a low-income range less than 49000 dollars, when term is 60 months, there is a fair percentage of people that default.
- **Annual_Inc vs Purpose**
  - For renewable energy when income < 36000 dollars, there are a fair percentage of values that have defaulted.
  - For small business when income < 49000 dollars, there are a fair percentage of people that have defaulted.

Observations

- **Annual Income vs Credit Utilization**
  - People with low income and high credit utilization have a default rate and people with high income and low credit utilization have a low default rate. These 2 variables are inversely proportional.
- **Annual Income vs DTI**
  - a low income with high debt-to-income ratio reflects financial instability and can lead to defaults
- **Revol_Util vs Purpose**
  - People with housing loans with a high credit utilization have a very high default ratio.
- **Purpose vs Term**
  - Educational and Small business loans have a very high default ratio when the term given to fulfill the loan is 60 months.
- **Revol_Util vs Purpose**
  - People with housing loans with a high credit utilization have a very high default ratio.
- **Revol_Util vs DTI**
  - For education loans, a high dti can cause a higher default ratio

These are scenarios that the bank can avoid. in order to avoid defaults

Conclusions

THANK YOU