# Lending Case Study

**FACILITATOR NAME**

1. AYUSHI GANDHI
2. SUNIL MALGUNKAR

# Problem Statement

- **Consumer Finance 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 utilize this knowledge for its portfolio and risk assessment.
- If one is able to identify these risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss. Identification of such applicants using EDA is the aim of this case study.

# Analysis Approach

- Analyse the data
- Data Cleaning
- Univariate Analysis
  - Unordered categorical method for purpose field
  - Ordered categorical method for Interest rate and Annual income field
- Segmented Univariate
  - Grouping fields like Verification status, Grade and loan status
- Bivariate Analysis
  - Continuous Variables – Annual income, interest rates (Derived metrics)
  - Categorical Variables – loan status, purpose
- Derived Metrics
  - Derived the customer status field basis loan status

# Results of analysis

- Data Cleaning
  - Replacing all 0,0.0,null values with nan
  - Dropping columns with more than 50% null values

```
null_percentage = (loan_data.isnull().sum() / len(loan_data)) * 100
columns_to_drop = null_percentage[null_percentage > 50].index
loan_data.drop(columns=columns_to_drop,inplace=True)
loan_data
```

| | id | member_id | loan_amnt | funded_amnt | funded_amnt_inv | term | int_rate | installment | grade | sub_grade | ... | initial_list_status | total_pymnt | tot |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 1077501 | 1296599 | 5000 | 5000 | 4975.0 | 36 months | 10.65% | 162.87 | B | B2 | ... | f | 5863.155187 | |
| **1** | 1077430 | 1314167 | 2500 | 2500 | 2500.0 | 60 months | 15.27% | 59.83 | C | C4 | ... | f | 1008.710000 | |
| **2** | 1077175 | 1313524 | 2400 | 2400 | 2400.0 | 36 months | 15.96% | 84.33 | C | C5 | ... | f | 3005.666844 | |
| **3** | 1076863 | 1277178 | 10000 | 10000 | 10000.0 | 36 months | 13.49% | 339.31 | C | C1 | ... | f | 12231.890000 | |
| **4** | 1075358 | 1311748 | 3000 | 3000 | 3000.0 | 60 months | 12.69% | 67.79 | B | B5 | ... | f | 3513.330000 | |

  - Renaming some of column name for better understanding

# Results of analysis

- Univariate Analysis
  - Loan_purpose dataframe to group data based on purpose for which loan has been taken

| purpose | id | member_id | loan_amnt | funded_amnt | term | interest_rate | installment | grade | sub_grade | employment_time |
|---|---|---|---|---|---|---|---|---|---|---|
| renewable_energy | 103 | 103 | 103 | 103 | 103 | 103 | 103 | 103 | 103 | 95 |
| educational | 325 | 325 | 325 | 325 | 325 | 325 | 325 | 325 | 325 | 317 |
| house | 381 | 381 | 381 | 381 | 381 | 381 | 381 | 381 | 381 | 368 |
| vacation | 381 | 381 | 381 | 381 | 381 | 381 | 381 | 381 | 381 | 352 |
| moving | 583 | 583 | 583 | 583 | 583 | 583 | 583 | 583 | 583 | 559 |
| medical | 693 | 693 | 693 | 693 | 693 | 693 | 693 | 693 | 693 | 668 |
| wedding | 947 | 947 | 947 | 947 | 947 | 947 | 947 | 947 | 947 | 934 |
| car | 1549 | 1549 | 1549 | 1549 | 1549 | 1549 | 1549 | 1549 | 1549 | 1497 |
| small_business | 1828 | 1828 | 1828 | 1828 | 1828 | 1828 | 1828 | 1828 | 1828 | 1783 |
| major_purchase | 2187 | 2187 | 2187 | 2187 | 2187 | 2187 | 2187 | 2187 | 2187 | 2115 |
| home_improvement | 2976 | 2976 | 2976 | 2976 | 2976 | 2976 | 2976 | 2976 | 2976 | 2879 |
| other | 3993 | 3993 | 3993 | 3993 | 3993 | 3993 | 3993 | 3993 | 3993 | 3833 |
| credit_card | 5130 | 5130 | 5130 | 5130 | 5130 | 5130 | 5130 | 5130 | 5130 | 5000 |
| debt_consolidation | 18641 | 18641 | 18641 | 18641 | 18641 | 18641 | 18641 | 18641 | 18641 | 18242 |

# Results of analysis

- Univariate Analysis
  - Ordered categorical method for Interest rate – created field interest_rate_range

| orrower_open_credit_lines | revolving_balance | revolving_utilization_rate | last_pymnt_d | last_pymnt_amnt | application_type | customer_status | interest_rate_float | interest_rate_range |
|---|---|---|---|---|---|---|---|---|
| 3 | 13648.0 | 83.70% | Jan-15 | 171.62 | INDIVIDUAL | Non-default | 11.0 | 10-15 |
| 3 | 1687.0 | 9.40% | Apr-13 | 119.66 | INDIVIDUAL | Default | 15.0 | 15-20 |
| 2 | 2956.0 | 98.50% | Jun-14 | 649.91 | INDIVIDUAL | Non-default | 16.0 | 15-20 |
| 10 | 5598.0 | 21% | Jan-15 | 357.48 | INDIVIDUAL | Non-default | 13.0 | 10-15 |
| 15 | 27783.0 | 53.90% | May-16 | 67.79 | INDIVIDUAL | Non-default | 13.0 | 10-15 |
| 9 | 7963.0 | 28.30% | Jan-15 | 161.03 | INDIVIDUAL | Non-default | 8.0 | 5-10 |
| 7 | 17726.0 | 85.60% | May-16 | 1313.76 | INDIVIDUAL | Non-default | 16.0 | 15-20 |

# Results of analysis

- Segmented Univariate
  - Grouping fields like Verification status, Grade and loan status

| grade | verification_status | loan_status | |
|---|---|---|---|
| A | Not Verified | Fully Paid | 4855 |
| | | Charged Off | 316 |
| | | Current | 20 |
| | Source Verified | Fully Paid | 2394 |
| | | Charged Off | 140 |
| | | | ... |
| G | Source Verified | Charged Off | 26 |
| | | Current | 3 |
| | Verified | Fully Paid | 131 |
| | | Charged Off | 58 |
| | | Current | 13 |

# Results of analysis

- Bivariate Analysis
  - Continuous Variables – Annual income, interest rates (Derived metrics)

| grade | sub_grade | ... | zip_code | addr_state | dti | delinq_2yrs | earliest_cr_line | inq_last_6mths | decisions | interest_rate_float | interest_rate_range | Salary_range |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| B | B2 | ... | 860xx | AZ | 27.65 | 0 | Jan-85 | 1 | Non-Default | 11.0 | 10-15 | less than 1 lac |
| C | C4 | ... | 309xx | GA | 1.00 | 0 | Apr-99 | 5 | Default | 15.0 | 15-20 | less than 1 lac |
| C | C5 | ... | 606xx | IL | 8.72 | 0 | Nov-01 | 2 | Non-Default | 16.0 | 15-20 | less than 1 lac |
| C | C1 | ... | 917xx | CA | 20.00 | 0 | Feb-96 | 1 | Non-Default | 13.0 | 10-15 | less than 1 lac |
| B | B5 | ... | 972xx | OR | 17.94 | 0 | Jan-96 | 0 | Non-Default | 13.0 | 10-15 | less than 1 lac |
| A | A4 | ... | 852xx | AZ | 11.20 | 0 | Nov-04 | 3 | Non-Default | 8.0 | 5-10 | less than 1 lac |
| C | C5 | ... | 280xx | NC | 23.51 | 0 | Jul-05 | 1 | Non-Default | 16.0 | 15-20 | less than 1 lac |

# Results of analysis

- Bivariate Analysis
  - Categorical Variables – loan status, purpose

```
#% of loan taken for different purpose
round(data.purpose.value_counts(normalize=True)*100,2)
```

```
debt_consolidation    46.93
credit_card           12.92
other                 10.05
home_improvement       7.49
major_purchase         5.51
small_business         4.60
car                    3.90
wedding                2.38
medical                1.74
moving                 1.47
vacation               0.96
house                  0.96
educational            0.82
renewable_energy       0.26
Name: purpose, dtype: float64
```

# Results of analysis

- Derived Metrics
  - Derived the customer status field basis loan status

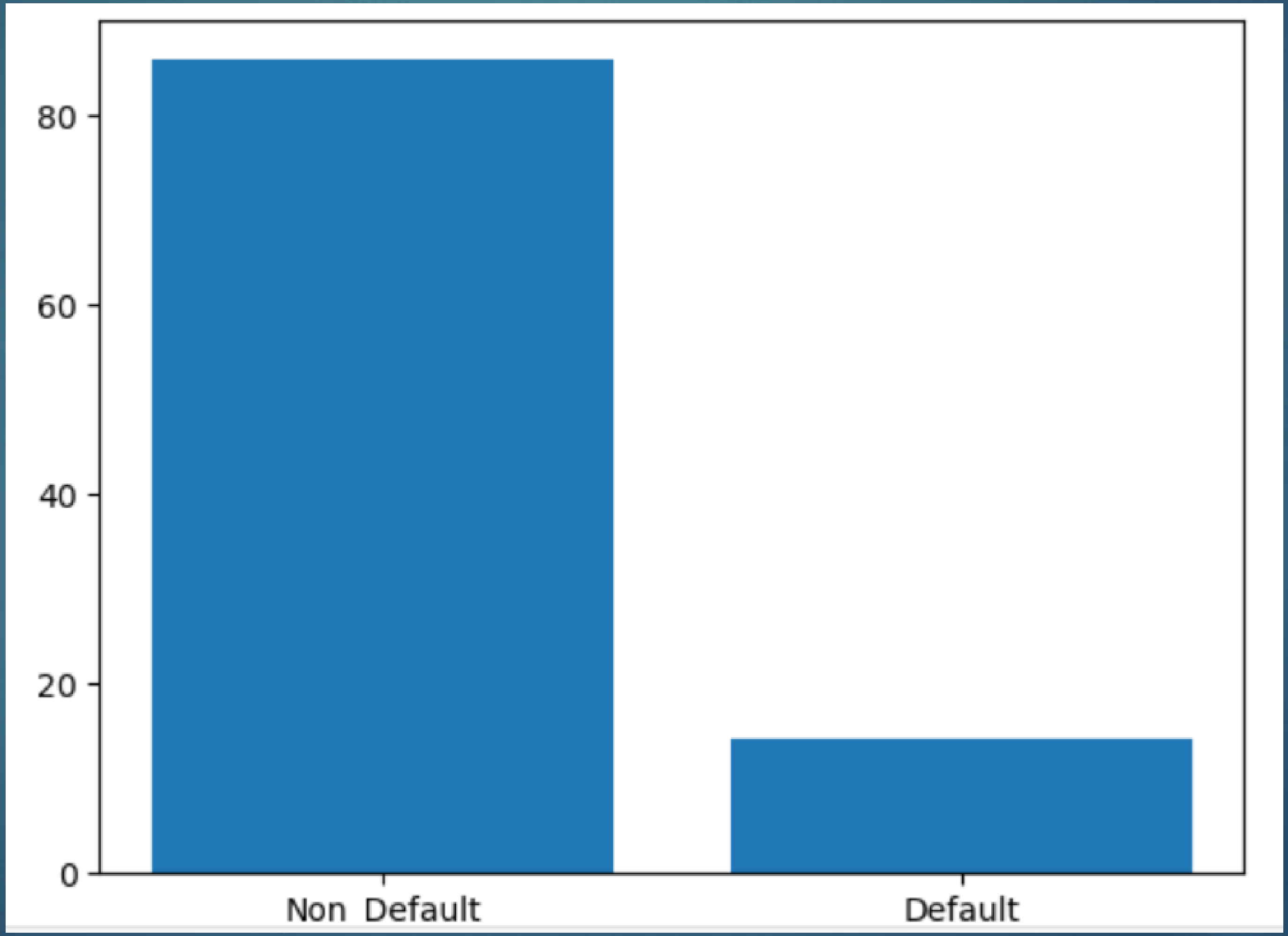
```
round(loan_data.customer_status.value_counts(normalize=True)*100,2)
```

```
Non-default    85.83
Default        14.17
Name: customer_status, dtype: float64
```

- non-default customer in percentage is 85.83
- default customer is percentage is 14.17

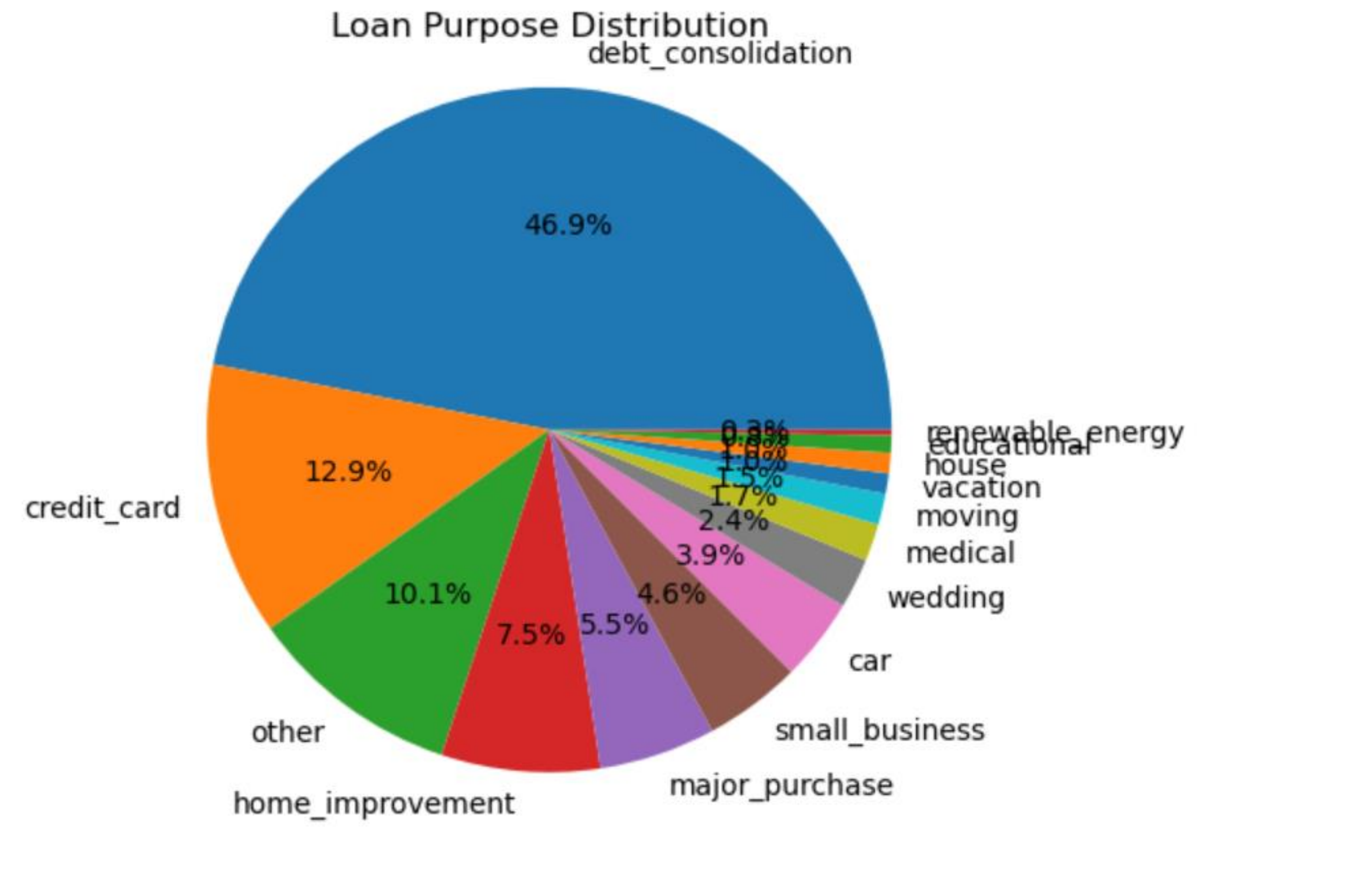
# Results of analysis

- Default and Non-default Percentage Graph



# Results of analysis

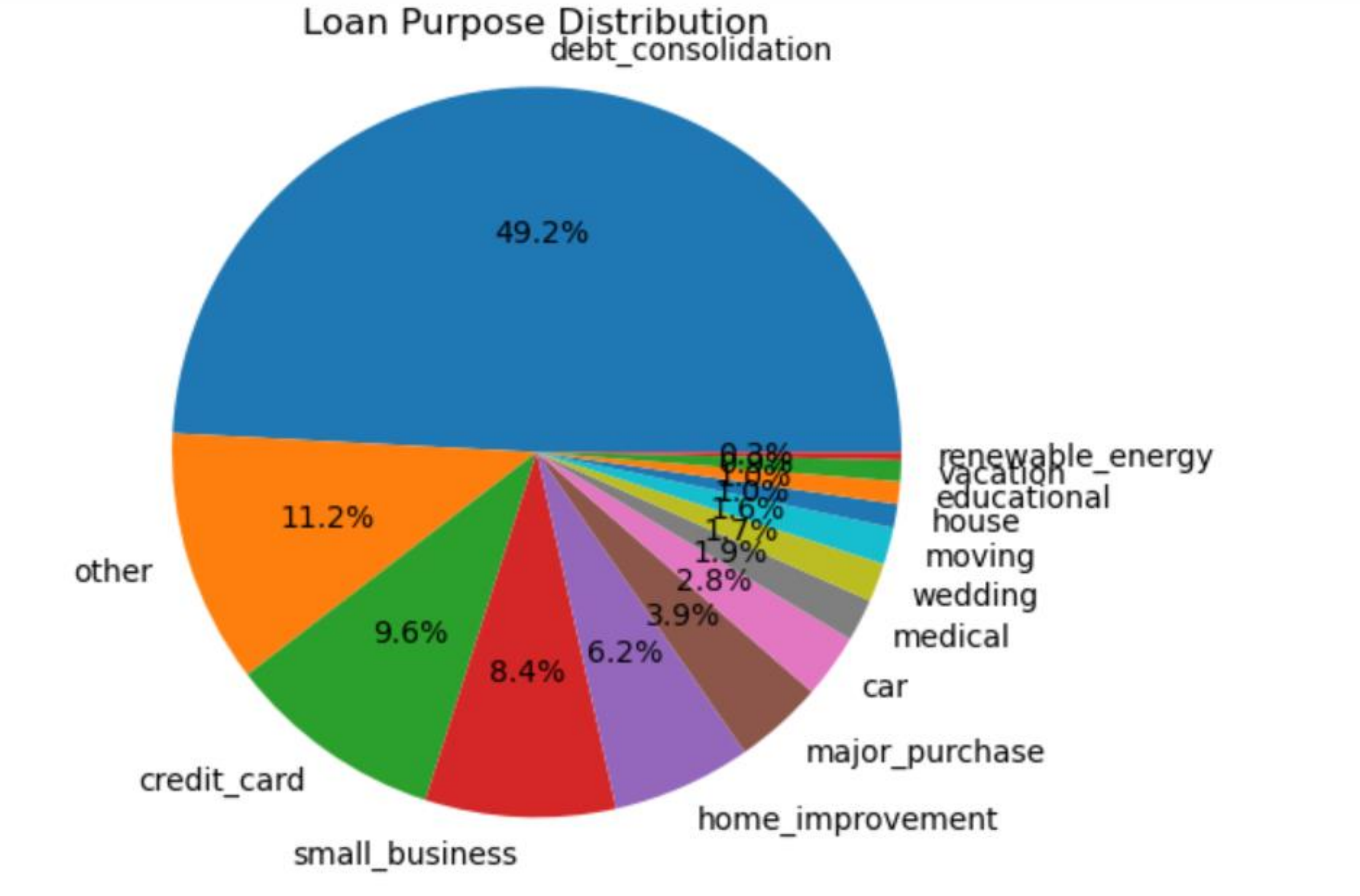
- Overall Loan purpose distribution




- debt_consolidation, credit_card, other are the main cause for which customers are applying for loan.

# Results of analysis

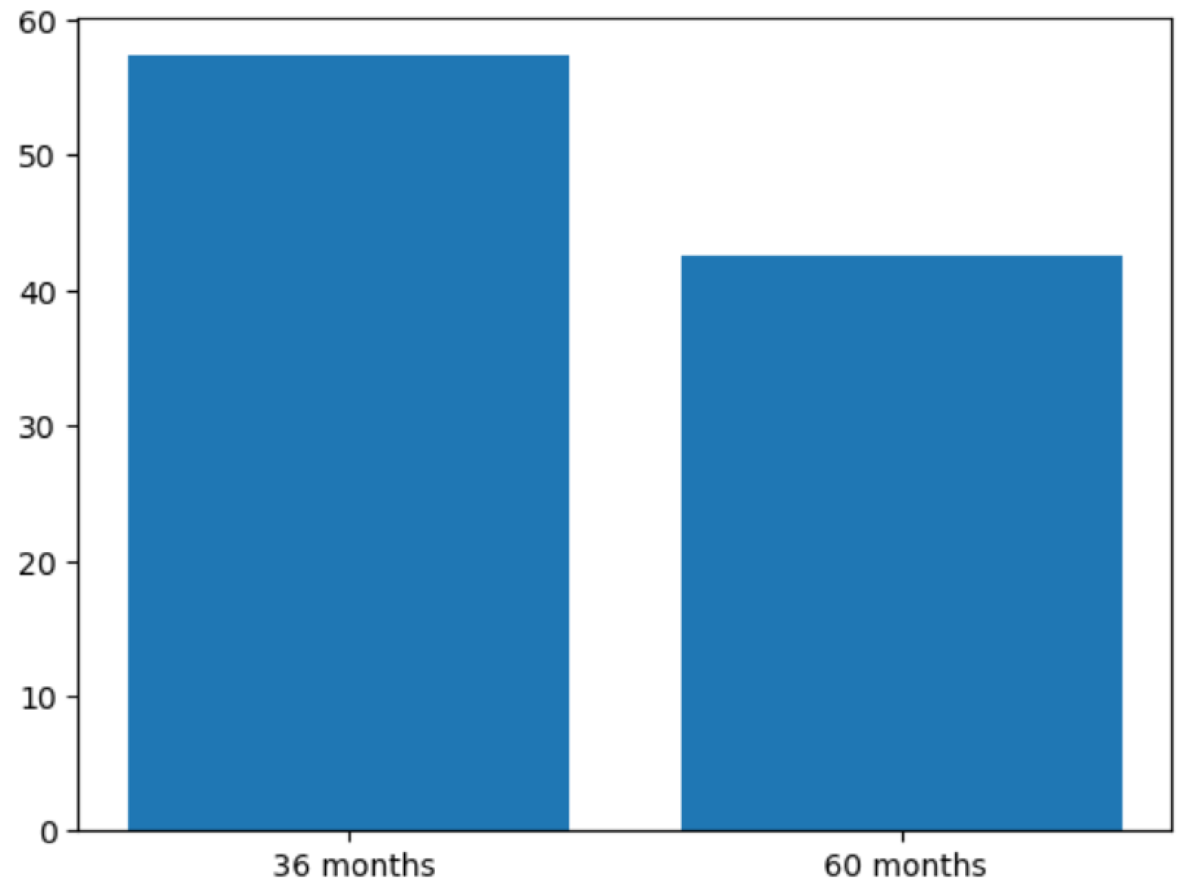
- Default status loan purpose distribution




- customer who takes loan with the purpose of debt_consolidation are more likely to be defaulty.

# Results of analysis

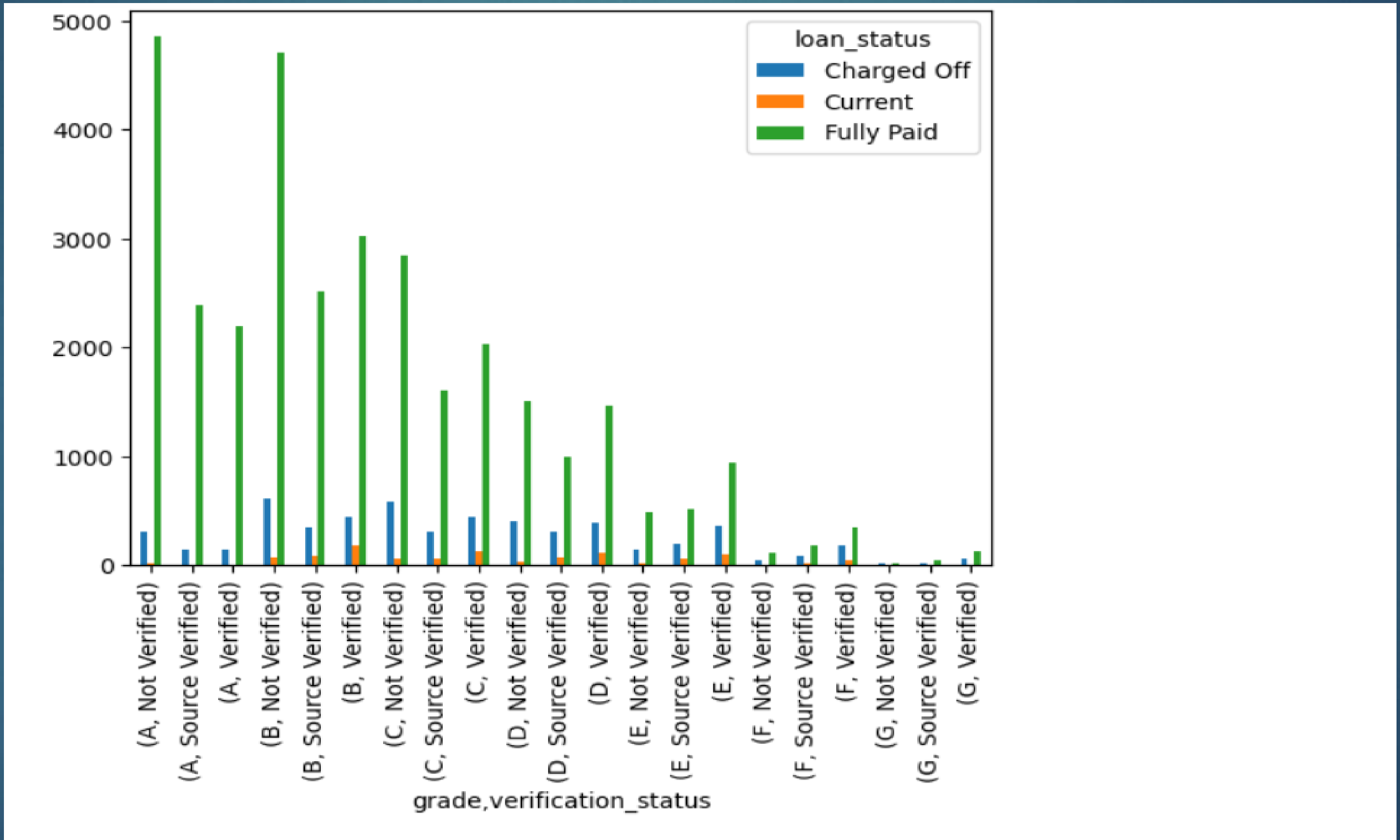
- Customers with loan term




- cutomer with term 36 are most likely to default

# Results of analysis

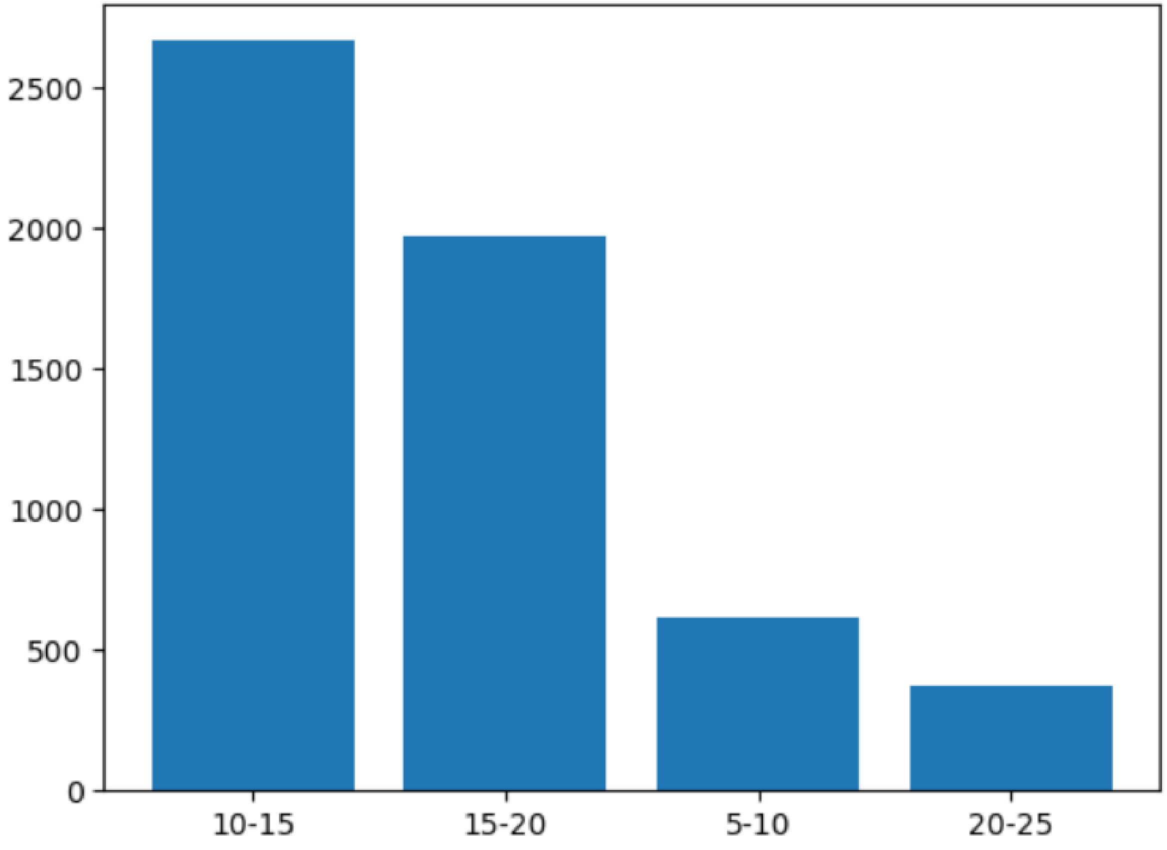
- Loan and verification Status graph




- From above plot it is summerise that cutomer with grade 'B' and non-verfied have a higher probability to br defaulty
- Cutomer who have grade 'A' are most likely to complete the loan amount

# Results of analysis

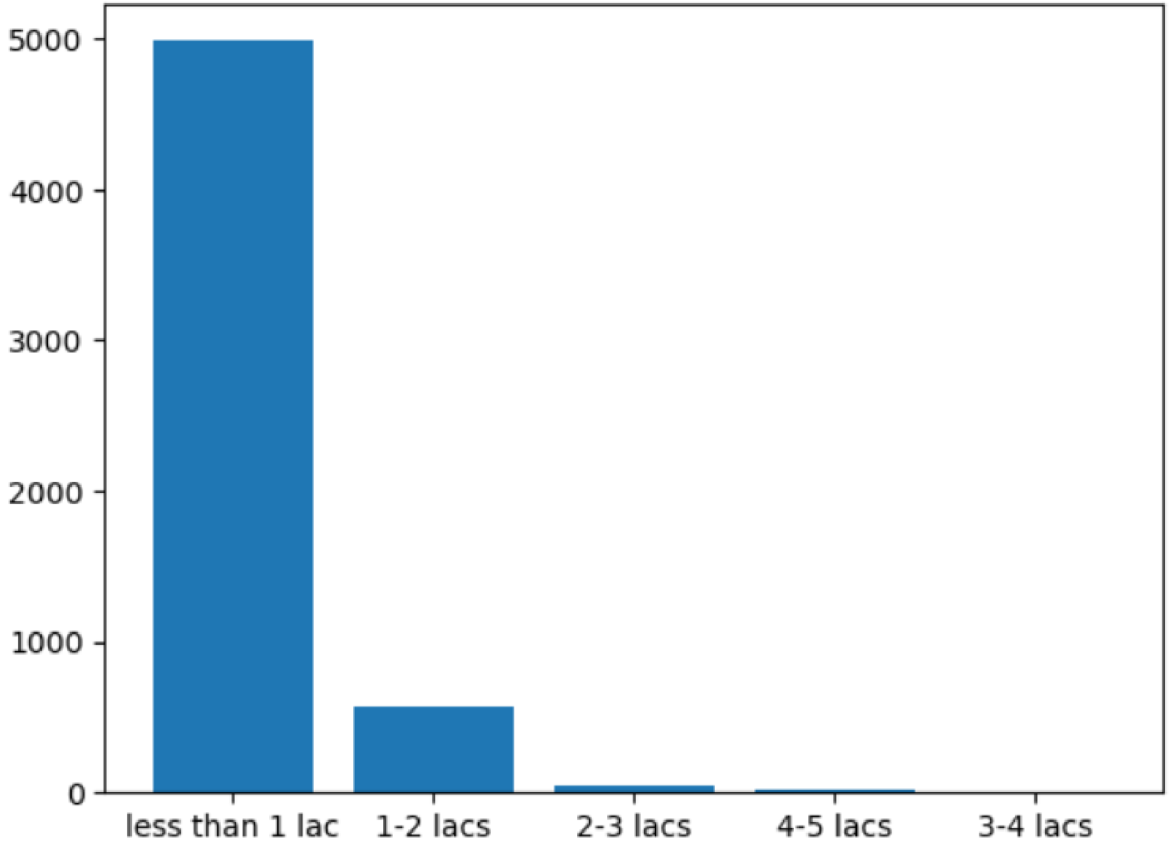
- Interest Rate rangewise graph of default status




- A bar chart study reveals that customers who are in interest rates of 10-15% are more likely to default on their loan obligations. It results in a 20–25 percent interest rate range with a lower client default rate.

# Results of analysis

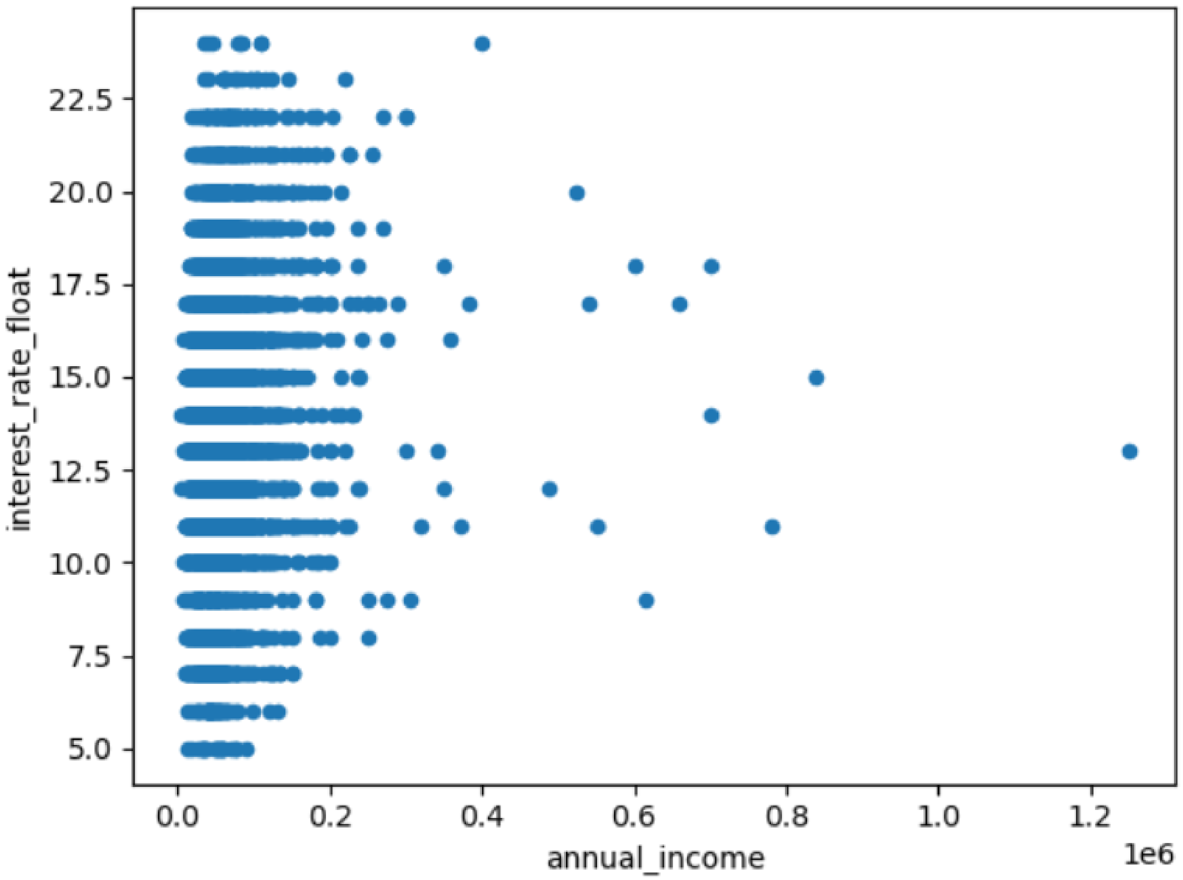
- Annual income of default status customers graph




- Customers who make less than $1 lac in annual income are more likely to default.
- There have been no defaults in the 3–4 lac range.

# Results of analysis

- Scatterplot basis Annual income and interest rate of default status customers graph




- The scatter plot reveals that the highest default customers are those with salaries between 0 and 1 lac, with the maximum default customers being those with salaries under 20,000 that too in interest rate 10-17.5%.

# Final Conclusion

- The grade, annual income, and term of the customer have a significant impact on whether they can repay the entire loan amount.
- Consumers with fewer loan payments (term)are more likely to default than those with more payments(term) (perhaps because they make smaller monthly payments over a longer term).
- The annual income and interest rate have a significant impact on a customer's ability to make loan payments since customers with lower salaries but higher interest rates are more likely to default than those with higher salaries.