## Univariate Analysis (including Segments)

Following variables are subjected to analysis.

|  |  |  |
| --- | --- | --- |
| Variable | Type | Observations |
| Loan\_status | Categorical unordered | Out of the three catogeries, ‘Charged off’ is of mainly of interest |
| Annual\_inc | Numerical . For analysis it is divided into segments of size 25000 | Min 4000 and Max 6000000, with mean of 68900. 90% of applicants have income less than 116000 |
| Term in months (36 & 60) | Categorical unordered |  |
| loan\_amnt | Numerical. Divided into segments for analysis | Range 500-35000. Divided into bins of 5K each |
| mths\_since\_last\_delinq | Numerical (after excluding NAs). Divided into segments for analysis | It is in the range of 0-120.  Divided into segments of size 20 for analysis |
| delinq\_2yrs (number of 30+ days past-due incidences of delinquency for past 2 yrs) | Numerical |  |
| pub\_rec\_bankruptcies | Numerical | only 3 values – 0,1,2 |
| Home\_ownership | Unordered Categorical | Mortgage, Rent, Owned, Other |
| emp\_length | Numerial | 1-10 years |
| verification\_status | Unordered categorical | Verified, Source Verified and Unverified |
| Interest\_rate | Numerical. Divided into segments for analysis | Range 5.4 to 24.6. Divided into buckets of size 2 |
| installment | Instalment amount. Divided into segments for analysis | Range 16-1305, 95th percentile at 762. Divided into buckets of 200 |

## Derived Metrics

|  |  |  |
| --- | --- | --- |
| Derived Metric | Origin metric(s) | Notes |
| %defaults | Loans with status=Charged Off, Total #loans | Aggregate count of entries with status=’Charged Off’, divided by Total entries |
| Annual Income Buckets | Annual Income | Bucket size of 25000 |
| Loan Amount Buckets | Loan Amount | Bucket size of 5000 |
| Months Since Last Delinquency reported Buckets | mths\_since\_last\_delinq | Bucket size of 20 |
| Interest rate bucket | Interest rate | Bucket size of 2 |
| Installment amount bucket | Installment | Bucket size of 200 |

## Bivariate Analysis

|  |  |  |
| --- | --- | --- |
| **Variable 1** | **Variable 2** | **Observations** |
| Percentage of defaulters | Income range | %defaults are almost inversely related to the income range |
| Percentage of defaulters | Term (of loan) | %defaults are almost two times more for loans of 60month term compared to 30month term. |
| Percentage of defaulters | Loan Amount | Nearly linear but small increase of defaults with higher loan amounts |
| Percentage of defaulters | Loan Installment | No significant variation |
| Percentage of defaulters | income verification status | no significant variation of defaults in relation to this |
| Percentage of defaulters | The number of 30+ days past-due incidences of delinquency in the borrower's credit file for the past 2 years | No significant correlation between the two variables is observed |
| Percentage of defaulters | Interest rate | defaults are higher for higher interest rate loans, almost linearly so |
| Percentage of defaulters | The number of months since the borrower's last delinquency. | There is no significant variation of defaults with this variable |
| Percentage of defaulters | The number of open credit lines in the borrower's credit file | There is no significant variation of defaults in relation to this variable |
| Percentage of defaulters | home ownership category | There is no significant variation of defaults in relation to this variable |
| Percentage of defaulters | Employment length | Defaults are much higher for applicants employed for <= 1 year. For others, i.e, >1 years, there is no significant variation |
| Percentage of defaulters | Number of inquiries | default rate is higher for applications who have made more inquiries in past 6 months. Value > 4 are ignored because they contain very few samples |
| Percentage of defaulters | Public record of bankruptcies | There is no significant variation of defaults in relation to 'Number of public records of bankruptcies.' |
| Percentage of defaulters | Number of derogatory public records | There is no significant variation of defaults in relation to 'Number of public records of bankruptcies.' |

## Bivariate categorial variable analysis

|  |  |  |  |
| --- | --- | --- | --- |
| **Metric** | **Category variable 1** | **Category variable 2** | **Observation** |
| Percentage of defaulters | Income range | Term | For loan term of 36 months default rate is consistently lower for all income ranges.  For loan term of 60 months it is higher except for income above 80K |
| Percentage of defaulters | Income range | Verification status | Income verification is not useful as it does not help in reducing rate of loan default across all income ranges |
| Percentage of defaulters | Income range | Instalment bucket | For income ranges upto 60K, reducing the installment amount by 2-3 buckets could reduce default rate significantly |
| Percentage of defaulters | Income range | Loan amount | For income ranges upto 60K, reducing the loan amount by 1-2 buckets could reduce default rates significantly |

Note: For other combinations i.e., term – loan amount, verification status – installment bucket, or verification status – loan amount, there is no useful inference.