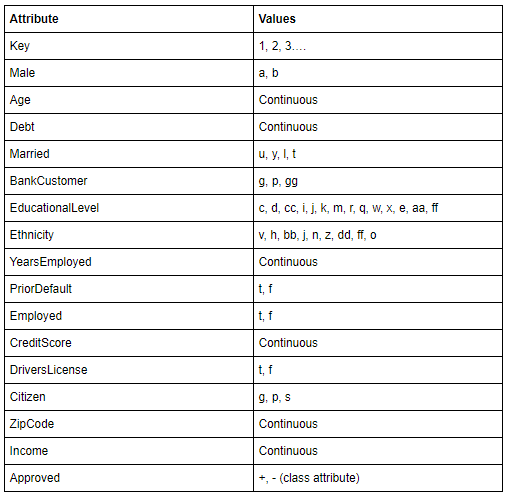
**Exploratory Data analysis**

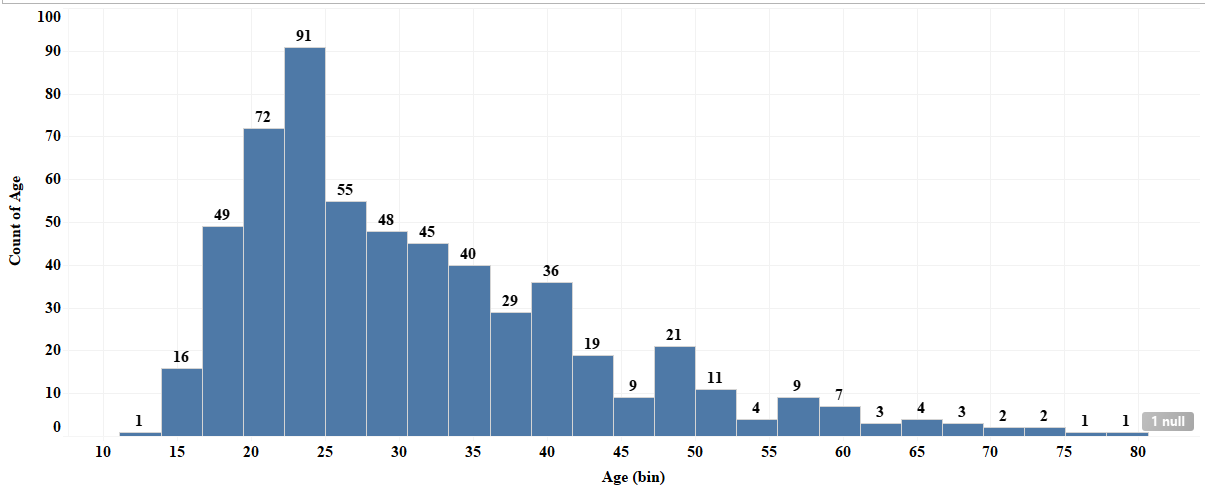
**Credit Card**

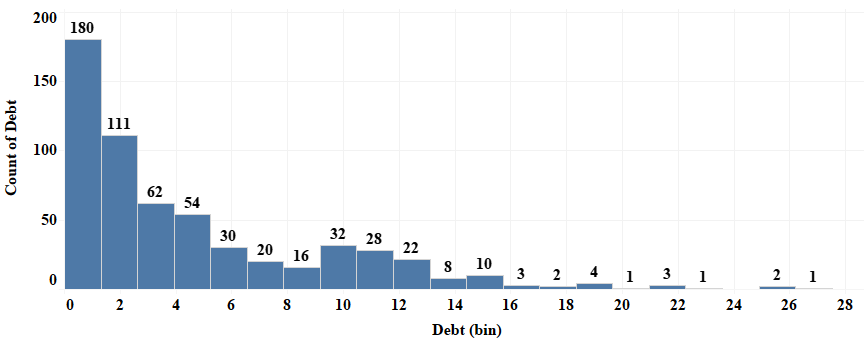
**Summary:**

* **Train data:** 590 records
* **Test data:** 100 records
* **Number of features:** 17

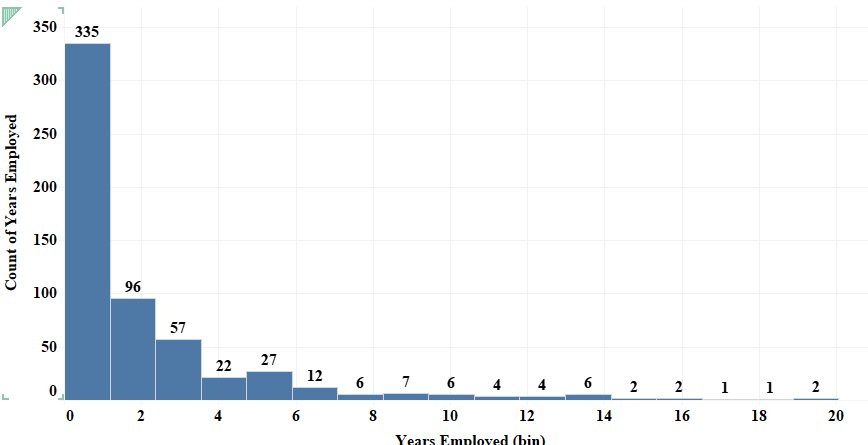


**Univariate analysis:**

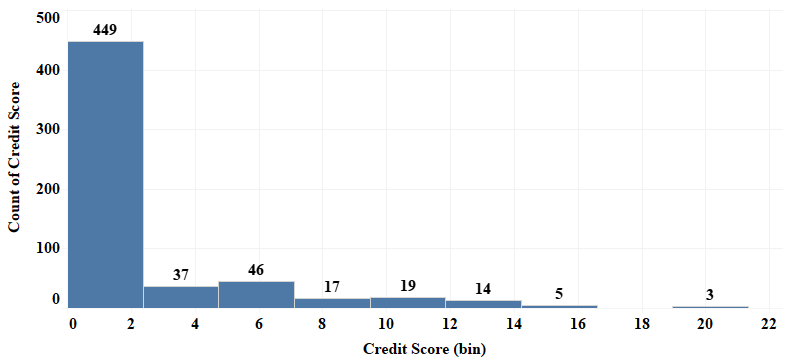
* **Continuous data:** 7 Attributes
  + **Key**: 590 unique values (Serial number)
  + **Age**: Mean= 31.3, Median = 28.1, mode = 22 ,min=13.75, max =80.25, SD=12.01, variance = 144.3
  + **Debt**: Mean= 4.6, Median = 2.7, mode = 0 ,min=0, max =26.34, SD=4.9, variance = 23.5



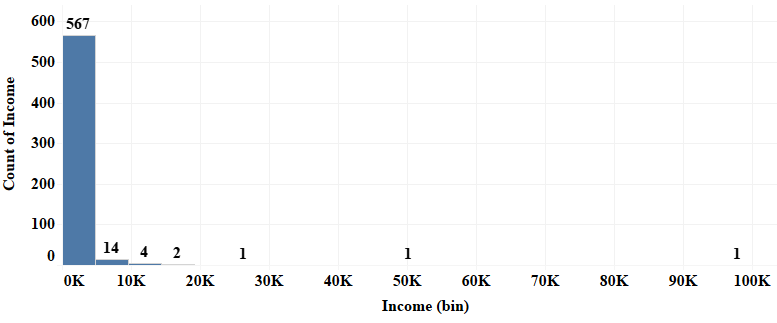
* + **YearsEmployed**: Mean= 2.1, Median = 1, mode = 0 ,min=0, max =20, SD=3.2, variance = 10.2



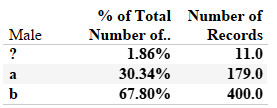
* + **CreditScore:** Mean= 2.1, Median = 0, mode = 0 ,min=0, max =20, SD=3.7, variance = 13.7



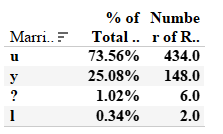
* + **Income:** Mean= 918.7, Median = 4, mode = 0 ,min=0, max =100,000, SD=5,025, variance = 25,254,800



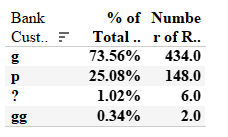
* **Discrete data:** 10 Attributes
  + Male:



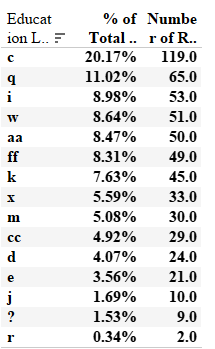
* + Married:



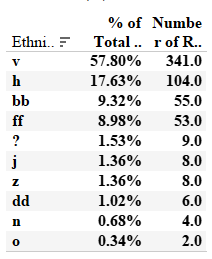
* + BankCustomer



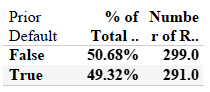
* + EducationLevel



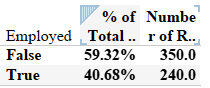
* + Ethnicity:



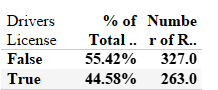
* + PriorDefault



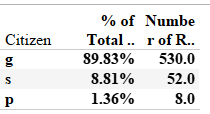
* + Employed



* + DriversLicense

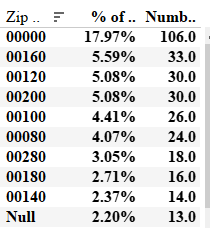


* + Citizen

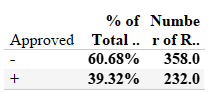


* + **ZipCode:** unique value: 155 code and null values are available

Top 10 zip codes:



* + Approved



**Bivariate analysis**

**o Male:**

Approved no yes

**Male**

a 102 77

b 248 152

no 8 3

dof=2

Expected: [[108.61355932 70.38644068]

[242.71186441 157.28813559]

[ 6.67457627 4.32542373]]

probability=0.950, critical=5.991, stat=1.986

Independent (fail to reject H0)

significance=0.050, p=0.370

Independent (fail to reject H0)

* + **Married:**

Approved no yes

Married

l 0 2

nol 2 4

u 245 189

y 111 37

dof=3

Expected: [[ 1.21355932 0.78644068]

[ 3.64067797 2.35932203]

[263.34237288 170.65762712]

[ 89.80338983 58.19661017]]

probability=0.950, critical=7.815, stat=20.939

Dependent (reject H0)

significance=0.050, p=0.000

Dependent (reject H0)

* + **BankCustomer**

Approved no yes

BankCustomer

g 245 189

gg 0 2

no 2 4

p 111 37

dof=3

Expected: [[263.34237288 170.65762712]

[ 1.21355932 0.78644068]

[ 3.64067797 2.35932203]

[ 89.80338983 58.19661017]]

probability=0.950, critical=7.815, stat=20.939

Dependent (reject H0)

significance=0.050, p=0.000

Dependent (reject H0)

* + **EducationLevel**

Approved no yes

EducationLevel

aa 32 18

c 68 49

cc 10 18

d 20 4

e 9 12

ff 43 6

i 43 10

j 7 3

k 36 9

m 19 11

no 10 4

q 25 39

r 1 1

w 29 21

x 6 27

dof=14

Expected: [[30.33898305 19.66101695]

[70.99322034 46.00677966]

[16.98983051 11.01016949]

[14.56271186 9.43728814]

[12.74237288 8.25762712]

[29.73220339 19.26779661]

[32.15932203 20.84067797]

[ 6.06779661 3.93220339]

[27.30508475 17.69491525]

[18.20338983 11.79661017]

[ 8.49491525 5.50508475]

[38.83389831 25.16610169]

[ 1.21355932 0.78644068]

[30.33898305 19.66101695]

[20.02372881 12.97627119]]

probability=0.950, critical=23.685, stat=86.102

Dependent (reject H0)

significance=0.050, p=0.000

Dependent (reject H0)

* + **Ethnicity:**

Approved no yes

Ethnicity

bb 34 21

dd 4 2

ff 46 7

h 44 60

j 5 3

n 2 2

no 5 4

o 1 1

v 215 126

z 2 6

dof=9

Expected: [[ 33.37288136 21.62711864]

[ 3.64067797 2.35932203]

[ 32.15932203 20.84067797]

[ 63.10508475 40.89491525]

[ 4.85423729 3.14576271]

[ 2.42711864 1.57288136]

[ 5.46101695 3.53898305]

[ 1.21355932 0.78644068]

[206.91186441 134.08813559]

[ 4.85423729 3.14576271]]

probability=0.950, critical=16.919, stat=35.447

Dependent (reject H0)

significance=0.050, p=0.000

Dependent (reject H0)

**o PriorDefault**

Approved no yes

PriorDefault

f 281 18

t 77 214

dof=1

Expected: [[181.42711864 117.57288136]

[176.57288136 114.42711864]]

probability=0.950, critical=3.841, stat=278.953

Dependent (reject H0)

significance=0.050, p=0.000

Dependent (reject H0)

* **Employed**

Approved no yes

Employed

f 277 73

t 81 159

dof=1

Expected: [[212.37288136 137.62711864]

[145.62711864 94.37288136]]

probability=0.950, critical=3.841, stat=121.057

Dependent (reject H0)

significance=0.050, p=0.000

Dependent (reject H0)

**o DriversLicense**

Approved no yes

DriversLicense

f 198 129

t 160 103

dof=1

Expected: [[198.41694915 128.58305085]

[159.58305085 103.41694915]]

probability=0.950, critical=3.841, stat=0.000

Independent (fail to reject H0)

significance=0.050, p=0.989

Independent (fail to reject H0)

* **Citizen**

Approved no yes

Citizen

g 315 215

p 3 5

s 40 12

dof=2

Expected: [[321.59322034 208.40677966]

[ 4.85423729 3.14576271]

[ 31.55254237 20.44745763]]

probability=0.950, critical=5.991, stat=7.897

Dependent (reject H0)

significance=0.050, p=0.019

Dependent (reject H0)

* **ZipCode:**

probability=0.950, critical=185.052, stat=205.005

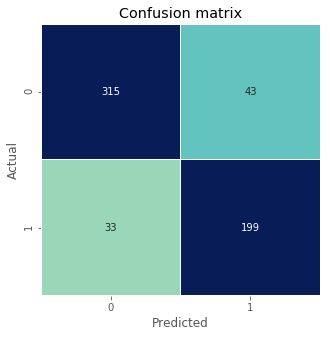
Dependent (reject H0)

significance=0.050, p=0.004

Dependent (reject H0)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **ACCURACY      MATRIX** | **Logistic Regression** | **Logistic Regression L2** | **Multinomial NB** | **SVM** | **Random Forest** |
| **Years employed + Income** | 0.72 | 0.72 | 0.65 | 0.86 | 0.61 |
| **Years employed + Income + married** | 0.73 | 0.73 | 0.68 | 0.85 | 0.60 |
| **Years employed + Income + married + bankCustomer** | 0.73 | 0.73 | 0.70 | 0.84 | 0.61 |
| **Years employed + Income + married + bankCustomer + EducationLevel** | 0.77 | 0.76 | 0.71 | 0.83 | 0.61 |
| **Years employed + Income + married + bankCustomer + EducationLevel + Ethnicity** | 0.78 | 0.77 | 0.71 | 0.82 | 0.61 |
| **Years employed + Income + married + bankCustomer + EducationLevel + Ethnicity +PrioriDefault** | 0.85 | 0.86 | 0.72 | 0.85 | 0.61 |
| **Years employed + Income + married + bankCustomer + EducationLevel + Ethnicity +PrioriDefault +Employed** | 0.87 | 0.87 | 0.72 | 0.86 | 0.61 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Input** | **Performance metrics** | **Logistic Regression** | **Logistic Regression L2** | **Multinomial NB** | **SVM** | **Random Forest** |
| **Years employed + Income + married + EducationLevel + Ethnicity +PrioriDefault +Employed +Citizen** | **Accuracy** | 0.87 | 0.87 | 0.73 | 0.86 | 0.61 |
| **Precision** | [0.90,  0.82] | [0.91, 0.82] | [0.71,0.77] | [0.85, 0.87] | [0.61, 0  ] |
| **Recall** | [0.87, 0.86] | [0.88, 0.86] | [0.92, 0.43] | [0.93, 0.75] | [1, 0] |
| **F1 score** | [0.89, 0.84] | [0.89, 0.84] | [0.80, 0.55] | [0.89, 0.81] | [0.76, 0 ] |

****