**Name :** Siddhant Mangale

**Class :** B.Tech (A)

**Roll No. :** 31 **Experiment No. :**7

julia> using CSV

julia> using DataFrames

julia> train=CSV.read("C:/Users/siddhant/OneDrive/Desktop/loan.csv",DataFrame, normalizenames=true)

614×13 DataFrame

Row │ Loan\_ID Gender Married Dependents Education Self\_Employed ApplicantIncome C ⋯

│ String15 String7? String3? String3? String15 String3? Int64 F

⋯─────┼─────────────────────────────────────────────────────────────────────────

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| ─────────────────── 1 │ LP001002 Male No 0 | | | | | Graduate No | 5849 ⋯ |
| 2 │ LP001003 Male Yes 1 Graduate No | | | | | 4583 |  |
| 3 │ LP001005 Male | Yes | 0 | Graduate | Yes | 3000 | |
| 4 │ LP001006 Male Yes 0 Not Graduate No 2583 | | | | | | |
| 5 │ LP001008 Male  Yes 2 Graduate | No  Yes | 0 | Graduate  5417 | No | 6000 ⋯ 6 │ LP001011 Male | |

1. │ LP001013 Male Yes 0 Not Graduate No 2333
2. │ LP001014 Male Yes 3+ Graduate No 3036
3. │ LP001018 Male Yes 2 Graduate No 4006 ⋯ 10 │ LP001020 Male Yes 1 Graduate No 12841

0

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 11 │ LP001024 Male | | Yes | 2 | Graduate | No 3200 | |
| 12 │ LP001027 Male | | Yes | 2 | Graduate | missing 250 | |
| 13 │ LP001028 Male  No 0 Graduate | | Yes  No | 2 | Graduate  1853 | No 3073 | |
| 15 │ LP001030 | Male | Yes | 2 | Graduate | No | 1299 |
| 16 │ LP001032 | Male | No | 0 | Graduate | No | 4950 |

⋯ 14 │ LP001029 Male

17 │ LP001034 Male No 1 Not Graduate No 3596 ⋯ 18 │ LP001036 Female No 0 Graduate No 3510

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 19 │ LP001038 | Male | Yes | | 0 Not Graduate No 4887 | | | | | |
| 20 │ LP001041 | Male | Yes | | 0 Graduate missing 2600 | | | | | |
| 21 │ LP001043 | Male | Yes | | 0 Not Graduate No 7660 ⋯ 22 │ LP001 | | | | | |
| Male Yes 1  ⋮ │ ⋮ ⋮ ⋮  Graduate No | | Gra  ⋮ | duate  ⋮  3859 | | No  ⋮ | 5  ⋮ | 955  ⋱ 594 │ LP002936 Male | Yes | 0 |

046

Graduate Yes 16120

|  |  |  |  |
| --- | --- | --- | --- |
| 595 │ LP002938 | Male | Yes | 0 |
| 596 │ LP002940 | Male | No | 0 |

Not Graduate No 3833 ⋯ 597 │ LP002941

Male Yes 2 Not Graduate Yes 6383

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 598 │ LP002943 | Male | No | missing | Graduate | No | 2987 |
| 599 │ LP002945 | Male | Yes | 0 | Graduate | Yes | 9963 |
| 600 │ LP002948 | Male | Yes | 2 | Graduate | No | 5780 ⋯ 601 │ LP002949 |

Female No 3+ Graduate missing 416

602 │ LP002950 Male Yes 0 Not Graduate missing 2894

4

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 603 │ LP002953 | Male | Yes | 3+ | Graduate | No | 5703 |
| 604 │ LP002958 | Male | No | 0 | Graduate | No | 3676 ⋯ 605 │ LP002959 |
| Female Yes  606 │ LP002960 | 1  Male | Graduat  Yes | e No 12000  0 Not Graduate No 2400 | | | |
| 607 │ LP002961 | Male | Yes | 1 Graduate No 3400 | | | |
| 608 │ LP002964 | Male | Yes | 2 Not Graduate No 3987 ⋯ 609 │ LP00297 | | | |

Male Yes 0 Graduate No 3232

610 │ LP002978 Female No 611 │ LP002979 Male Yes

|  |  |  |  |
| --- | --- | --- | --- |
| 0 | Graduate | No | 2900 |
| 3+ | Graduate | No | 4106 |

612 │ LP002983 Male Yes 1 Graduate No 8072 ⋯ 613 │ LP002984

Male Yes 2 Graduate No 7583

614 │ LP002990 Female No 0 Graduate Yes 4583

6 columns and 571 rows omitted

julia> size(train) (614, 13)

julia> names(train)

13-element Vector{String}:

"Loan\_ID" "Gender" "Married" "Dependents" "Education" "Self\_Employed" "ApplicantIncome"

"CoapplicantIncome" "LoanAmount" "Loan\_Amount\_Term" "Credit\_History" "Property\_Area" "Loan\_Status"

julia> train[1:10,1:7] 10×7 DataFrame

Row │ Loan\_ID Gender Married Dependents Education Self\_Employed ApplicantIncome

│ String15 String7? String3? String3? String15 String3? Int64

─────┼───────────────────────────────────────────────────────────────────────────

─────────────

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 1 │ LP001002 Male | No | 0 | Graduate | No | 5849 |
| 2 │ LP001003 Male | Yes | 1 | Graduate | No | 4583 |
| 5 │ LP001008 Male | No | 0 | Graduate | No | 6000 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 6 │ LP001011 Male | Yes | 2 | Graduate | Yes | 5417 |
| 7 │ LP001013 Male | Yes | 0 | Not Graduate No | | 2333 |
| 8 │ LP001014 Male | Yes | 3+ | Graduate No | | 3036 |
| 9 │ LP001018 Male | Yes | 2 | Graduate No | | 4006 |
| 10 │ LP001020 Male | Yes | 1 | Graduate No | | 12841 |

julia> train[1:10,8:13]

│ Float64 Int64? Int64? Int64? String15 String1

⋯─────┼─────────────────────────────────────────────────────────────────────────

─────────────────── 1 │ 0.0 missing 360 1 Urban Y ⋯ 2 │

1508.0 128 360 1 Rural N

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 3 │ | 0.0 | 66 | 360 | | 1 | Urban | Y | | | | |
| 5 │  360 | 0.0  1 Urban | 141 | Y | 360 | 1 Urban | | Y ⋯ 6 │ | | | 4196.0 | 267 |
| 7 │ | 1516.0 | 95 | | 360 | 1 Urban | | | Y |  |  |  |
| 8 │ | 2504.0 | 158 | | 360 | 0 Semiurban | | |  | N |  |  |
| 9 │ | 1526.0 | 168 | | 360 | 1 Urban | | | Y | ⋯ 10 │ | 10968.0 | 349 |

360 1 Semiurban N

1 column omitted

julia> describe(train) 13×7 DataFrame

Row │ variable mean min median max nmissing eltype ⋯ │ Symbol Union… Any Union… Any Int64 Type

⋯─────┼─────────────────────────────────────────────────────────────────────────

─────────────────── 1 │ Loan\_ID LP001002 LP002990 0 String15 ⋯

2 │ Gender Female Male 13 Union{Missing, Str 3 │ Married No Yes 3 Union{Missing, Str

1. │ Dependents 0 3+ 15 Union{Missing, Str
2. │ Education Graduate Not Graduate 0 String15 ⋯ 6 │ Self\_Employed No Yes 32 Union{Missing, Str

7 │ ApplicantIncome 5403.46 150 3812.5 81000 0 Int64

8 │ CoapplicantIncome 1621.25 0.0 1188.5 41667.0 0 Float64

9 │ LoanAmount 146.412 9 128.0 700 22 Union{Missing, Int ⋯ 10 │

Loan\_Amount\_Term 342.0 12 360.0 480 14 Union{Missing, Int

11 │ Credit\_History 0.842199 0 1.0 1 50 Union{Missing, Int 12 │ Property\_Area Rural Urban 0 String15

13 │ Loan\_Status N Y 0 String1 ⋯

1 column omitted

julia> using Pkg

julia> Pkg.add("Plots")

Updating registry at `C:\Users\netra\.julia\registries\General.toml` Resolving package versions...

No Changes to `C:\Users\netra\.julia\environments\v1.8\Project.toml` No Changes to `C:\Users\netra\.julia\environments\v1.8\Manifest.toml`

julia> Pkg.add("PyPlot") Resolving package versions...

No Changes to `C:\Users\netra\.julia\environments\v1.8\Project.toml` No Changes to `C:\Users\netra\.julia\environments\v1.8\Manifest.toml`

julia> Pkg.add("StatsPlots") Resolving package versions...

No Changes to `C:\Users\netra\.julia\environments\v1.8\Project.toml` No Changes to `C:\Users\netra\.julia\environments\v1.8\Manifest.toml`

julia> using Plots

julia> pyplot()

[ Info: Precompiling PyPlot [d330b81b-6aea-500a-939a-2ce795aea3ee] ERROR: LoadError: PyCall not properly installed. Please run Pkg.build("PyCall")

julia> train=dropmissing(train) 480×13 DataFrame

Row │ Loan\_ID Gender Married Dependents Education Self\_Employed Applica ⋯

│ String15 String7 String3 String3 String15 String3 Int64 ⋯

─────┼───────────────────────────────────────────────────────────────────────────

────

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 1 │ LP001003 Male | | Yes | 1 | Graduate | No | ⋯ |
| 2 │ LP001005 Male | | Yes | 0 | Graduate | Yes |  |
| 3 │ LP001006 Male Yes | | | 0 | Not Graduate No | |  |
| 4 │ LP001008 Male No | | | 0 | Graduate No | |  |
| 5 │ LP001011 Male Yes | | | 2 | Graduate Yes | | ⋯ |
| 6 │ LP001013 Male Yes | | | 0 | Not Graduate No | |  |
| 7 │ LP001014 Male Yes | | | 3+ | Graduate No | |  |
| 8 │ LP001018 Male Yes | | | 2 | Graduate No | |  |
| 9 │ LP001020 Male Yes | | | 1 | Graduate No | | ⋯ |
| 10 │ LP001024 Male Yes | | | 2 | Graduate No | |  |
| 11 │ LP001028 Male Yes | | | 2 | Graduate No | |  |
| 12 │ LP001029 Male No | | | 0 | Graduate No | |  |
| 13 │ LP001030 Male Yes | | | 2 | Graduate No | | ⋯ |
| 14 │ LP001032 Male No | | | 0 | Graduate No | |  |
| 15 │ LP001036 Female No | | | 0 | Graduate No | |  |
| 16 │ LP001038 | Male | Yes | 0 | Not Graduate No | | |
| 17 │ LP001043 | Male | Yes | 0 | Not Graduate No | | |
| 18 │ LP001046 | Male | Yes | 1 | Graduate No | | |
| 19 │ LP001047 | Male | Yes | 0 | Not Graduate No | | |
| 20 │ LP001066 | Male | Yes | 0 | Graduate | Yes |  |
| 21 │ LP001068 | Male | Yes | 0 | Graduate | No | ⋯ |

⋯

22 │ LP001073 Male Yes 2 Not Graduate No

⋮ │ ⋮ ⋮ ⋮ ⋮ ⋮ ⋮ ⋱

460 │ LP002917 Female No 0 Not Graduate No

⋯

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 461 │ LP002926 | Male | Yes | 2 | Graduate | Yes |
| 462 │ LP002928 | Male | Yes | 0 | Graduate | No |
| 463 │ LP002931 | Male | Yes | 2 | Graduate | Yes |
| 464 │ LP002936 | Male | Yes | 0 | Graduate | No |
| 465 │ LP002938 | Male | Yes | 0 | Graduate | Yes |
| 466 │ LP002940 | Male | No | 0 | Not Graduate No | |
| 467 │ LP002941 | Male | Yes | 2 | Not Graduate Yes | |
| 468 │ LP002945 | Male | Yes | 0 | Graduate Yes | |
| 469 │ LP002948 | Male | Yes | 2 | Graduate No | |
| 470 │ LP002953 | Male | Yes | 3+ | Graduate No | |
| 471 │ LP002958 | Male | No | 0 | Graduate No | |
| 472 │ LP002959 | Female | Yes | 1 | Graduate No | |
| 473 │ LP002961 | Male | Yes | 1 | Graduate No | |
| 474 │ LP002964 Male Yes 2 Not Graduate No | | | | | |
| 475 │ LP002974 | Male | Yes | 0 | Graduate | No |
| 476 │ LP002978 | Female | No | 0 | Graduate | No |
| 477 │ LP002979 | Male | Yes | 3+ | Graduate | No |
| 478 │ LP002983 | Male | Yes | 1 | Graduate | No |
| 479 │ LP002984 | Male | Yes | 2 | Graduate | No |
| 480 │ LP002990 | Female | No | 0 | Graduate | Yes |

⋯

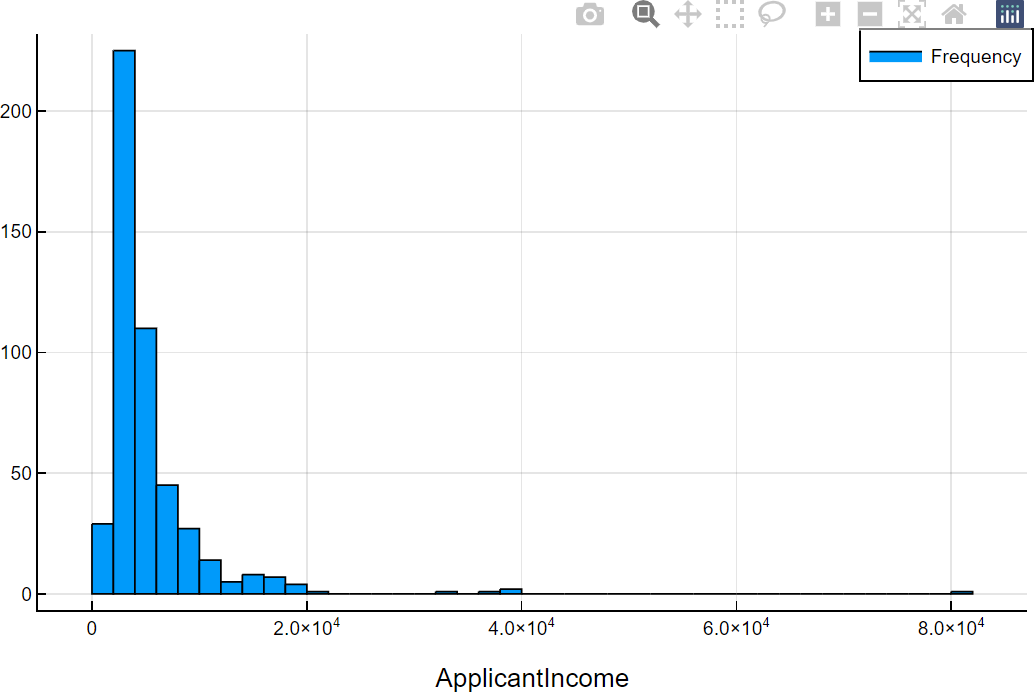
⋯

⋯

⋯

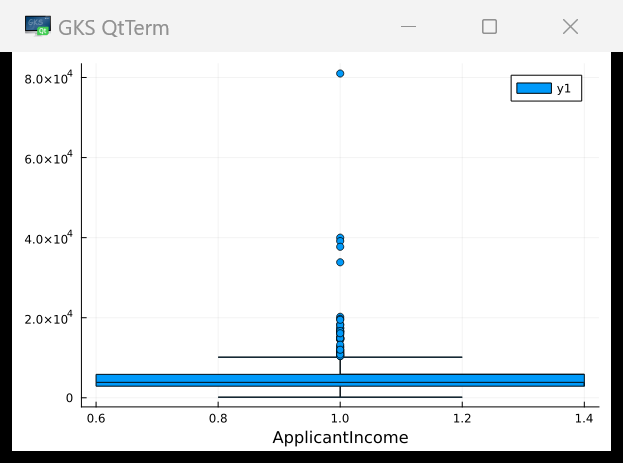
7 columns and 437 rows omitted

julia> Plots.histogram(train[!,"ApplicantIncome"],bins=50,xlabel="ApplicantIncome",labels="Frequency")

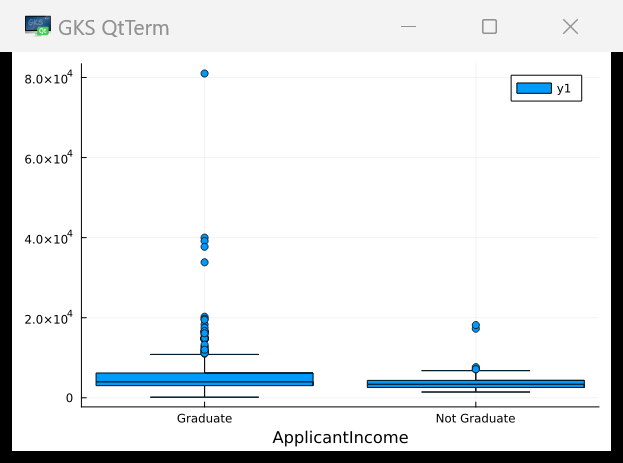


julia> using StatsPlots

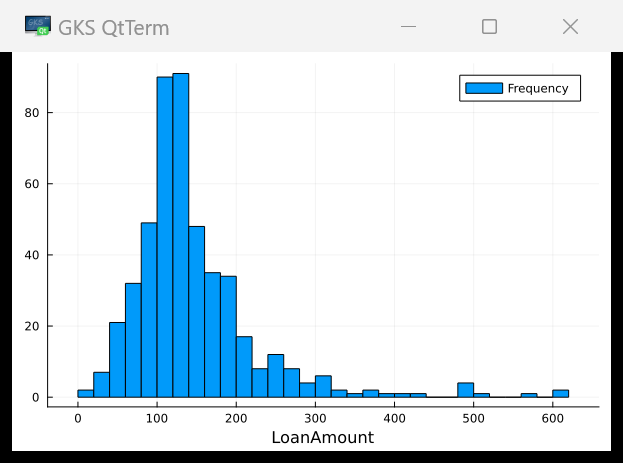
julia> StatsPlots.boxplot(train[!,"ApplicantIncome"],xlabel="ApplicantIncome")



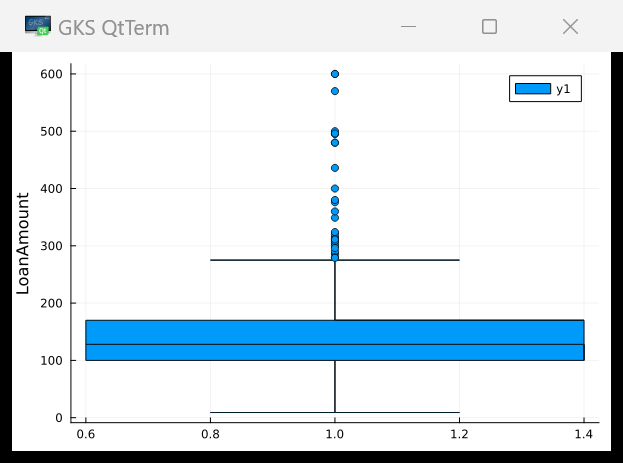
julia> StatsPlots.boxplot(train[!,"Education"],train[!,"ApplicantIncome"],xlabel="ApplicantIncome")



julia> Plots.histogram(train[!,"LoanAmount"],bins=50,xlabel="LoanAmount",labels="Frequency")



julia> StatsPlots.boxplot(train[!,"LoanAmount"],ylabel="LoanAmount")



julia> StatsPlots.scatter(train[!,"LoanAmount"],train[!,"ApplicantIncome"],xlabel="LoanAmount",ylabel="A pplicantIncome")

