# datacleaning-file-1

### November 1, 2024

```
[1]: import pandas as pd
     df = pd.read_csv("train.csv")
[2]: print(df)
           Loan_ID
                     Gender Married Dependents
                                                      Education Self_Employed
          LP001002
    0
                       Male
                                  No
                                                       Graduate
    1
          LP001003
                       Male
                                                1
                                 Yes
                                                       Graduate
                                                                             No
    2
          LP001005
                       Male
                                 Yes
                                               0
                                                       Graduate
                                                                            Yes
    3
          LP001006
                                                   Not Graduate
                       Male
                                 Yes
                                               0
                                                                             No
    4
          LP001008
                       Male
                                  No
                                               0
                                                       Graduate
                                                                             No
                                               0
    609
          LP002978
                     Female
                                  No
                                                       Graduate
                                                                             No
    610
          LP002979
                       Male
                                 Yes
                                              3+
                                                       Graduate
                                                                             No
          LP002983
                       Male
                                                       Graduate
                                 Yes
                                               1
                                                                             No
    612
          LP002984
                       Male
                                 Yes
                                                2
                                                       Graduate
                                                                             No
    613
          LP002990
                    Female
                                  No
                                               0
                                                       Graduate
                                                                            Yes
          ApplicantIncome
                             CoapplicantIncome
                                                 LoanAmount
                                                              Loan_Amount_Term
    0
                                            0.0
                      5849
                                                         NaN
                                                                           360.0
                      4583
                                         1508.0
                                                       128.0
    1
                                                                           360.0
    2
                      3000
                                            0.0
                                                        66.0
                                                                           360.0
    3
                                         2358.0
                      2583
                                                       120.0
                                                                           360.0
    4
                      6000
                                            0.0
                                                       141.0
                                                                           360.0
                                            0.0
                                                                           360.0
    609
                      2900
                                                        71.0
    610
                      4106
                                            0.0
                                                        40.0
                                                                           180.0
    611
                      8072
                                          240.0
                                                       253.0
                                                                           360.0
    612
                      7583
                                            0.0
                                                       187.0
                                                                           360.0
    613
                      4583
                                            0.0
                                                       133.0
                                                                           360.0
          Credit_History Property_Area Loan_Status
    0
                      1.0
                                   Urban
    1
                      1.0
                                   Rural
                                                     N
                                                     Y
    2
                      1.0
                                   Urban
                                                     Y
    3
                      1.0
                                   Urban
    4
                      1.0
                                   Urban
```

• •	•••	•••	•••	
609	1.0	Rural		Y
610	1.0	Rural		Y
611	1.0	Urban		Y
612	1.0	Urban		Y
613	0.0	Semiurban		N

[614 rows x 13 columns]

# [3]: df.drop(['Dependents'], axis=1) #drop the column

[3]:		${\tt Loan\_ID}$	Gender M	larried	]	Education	Self_Empl	oyed	ApplicantI	ncome	\
	0	LP001002	Male	No		${\tt Graduate}$		No		5849	
	1	LP001003	Male	Yes		${\tt Graduate}$		No		4583	
	2	LP001005	Male	Yes		Graduate		Yes		3000	
	3	LP001006	Male	Yes	Not	Graduate		No		2583	
	4	LP001008	Male	No		Graduate		No		6000	
		•••		•			•••		•••		
	609	LP002978	Female	No		${\tt Graduate}$		No		2900	
	610	LP002979	Male	Yes		${\tt Graduate}$		No		4106	
	611	LP002983	Male	Yes		${\tt Graduate}$		No		8072	
	612	LP002984	Male	Yes		Graduate		No		7583	
	613	LP002990	Female	No		${\tt Graduate}$		Yes		4583	
		Coapplica	ntIncome	LoanAmo	unt	Loan_Amo	ount_Term	Cred	it_History	\	
	0		0.0		NaN		360.0		1.0		
	1		1508.0	12	28.0		360.0		1.0		
	2		0.0	6	6.0		360.0		1.0		
	3		2358.0	12	20.0		360.0		1.0		
	4		0.0	14	1.0		360.0		1.0		
			•••	•••					•••		
	609		0.0	7	1.0		360.0		1.0		
	610		0.0	4	10.0		180.0		1.0		
	611		240.0	25	3.0		360.0		1.0		
	612		0.0	18	37.0		360.0		1.0		
	613		0.0	13	33.0		360.0		0.0		

## Property\_Area Loan\_Status

0	Urban	Y
1	Rural	N
2	Urban	Y
3	Urban	Y
4	Urban	Y
	•••	•••
609	Rural	Y
610	Rural	Y
611	Urban	Y

612	Urban	Y
613	Semiurban	N

[614 rows x 12 columns]

ΓΛ	۱. ا	df .	dron (	΄ Γ 🔾	11)	#drop	the	manne	
14	I : I	aI.	arop(	. 10.	11)	#arov	tne	rows	

Γ47.		I aan ID	C	M	D	Edu oo to	: 0-14	E ] d	`
[4]:	0	Loan_ID LP001005			Dependents 0			_Employed Yes	\
	2		Male	Yes					
	3 4	LP001006	Male Male	Yes No	0			No No	
	<del>4</del> 5	LP001008				Gradua		No	
		LP001011	Male	Yes	2	Gradua		Yes	
	6	LP001013	Male	Yes	0	Not Gradua	ate	No	
	609	 LP002978	 Female	 No		 Cmoday		No	
				No	0	Gradua Gradua		No No	
	610	LP002979	Male	Yes	3+			No No	
	611	LP002983	Male	Yes	1	Gradua		No No	
	612	LP002984	Male	Yes	2			No	
	613	LP002990	Female	No	0	Gradua	ate	Yes	
		Applicant	Income	Coapplio	cantIncome	LoanAmount	Loan A	mount_Term	\
	2	11	3000	11	0.0	66.0	_	360.0	
	3		2583		2358.0	120.0		360.0	
	4		6000		0.0	141.0		360.0	
	5		5417		4196.0	267.0		360.0	
	6		2333		1516.0	95.0		360.0	
			•••		•••	***		•••	
	609		2900		0.0	71.0		360.0	
	610		4106		0.0	40.0		180.0	
	611		8072		240.0	253.0		360.0	
	612		7583		0.0	187.0		360.0	
	613		4583		0.0	133.0		360.0	
	_	Credit_Hi	•	-	Area Loan_St				
	2		1.0		rban	Y			
	3		1.0		rban	Y			
	4		1.0		rban	Y			
	5		1.0		rban	Y			
	6		1.0	Uı	rban	Y			
	• •		•••	•••	•••				
	609		1.0		ıral	Y			
	610		1.0		ıral	Y			
	611		1.0		rban	Y			
	612		1.0		rban	Y			
	613		0.0	Semiu	rban	N			

[612 rows x 13 columns]

```
[5]: df.columns[0]
                     #displays 1st column name
[5]: 'Loan ID'
     import pandas as pd
     import numpy as np
     df = pd.read_csv("train.csv")
     print(df.replace(np.NaN,0))
     #df['DataFrame Column'] = df['DataFrame Column'].replace(np.nan, 0)
           Loan ID
                    Gender Married Dependents
                                                     Education Self_Employed
    0
         LP001002
                       Male
                                 No
                                                      Graduate
    1
          LP001003
                       Male
                                Yes
                                               1
                                                      Graduate
                                                                            No
    2
         LP001005
                       Male
                                Yes
                                               0
                                                      Graduate
                                                                           Yes
    3
                       Male
         LP001006
                                Yes
                                              0
                                                  Not Graduate
                                                                            No
    4
                                                      Graduate
         LP001008
                       Male
                                 No
                                              0
                                                                            No
    . .
         LP002978
                                              0
                                                                            No
    609
                    Female
                                 No
                                                      Graduate
                       Male
                                             3+
    610
         LP002979
                                Yes
                                                      Graduate
                                                                            No
         LP002983
                       Male
                                                      Graduate
    611
                                Yes
                                              1
                                                                            No
         LP002984
                                              2
    612
                       Male
                                Yes
                                                      Graduate
                                                                            No
                                                      Graduate
         LP002990 Female
                                 No
                                               0
    613
                                                                           Yes
                            CoapplicantIncome LoanAmount Loan_Amount_Term
          ApplicantIncome
    0
                      5849
                                           0.0
                                                        0.0
                                                                          360.0
    1
                      4583
                                        1508.0
                                                      128.0
                                                                          360.0
    2
                      3000
                                           0.0
                                                       66.0
                                                                          360.0
    3
                      2583
                                        2358.0
                                                      120.0
                                                                          360.0
    4
                      6000
                                           0.0
                                                      141.0
                                                                          360.0
    609
                      2900
                                           0.0
                                                       71.0
                                                                          360.0
    610
                                                       40.0
                      4106
                                           0.0
                                                                          180.0
    611
                                         240.0
                                                      253.0
                                                                          360.0
                      8072
    612
                      7583
                                           0.0
                                                      187.0
                                                                          360.0
    613
                      4583
                                           0.0
                                                      133.0
                                                                          360.0
          Credit_History Property_Area Loan_Status
    0
                      1.0
                                  Urban
    1
                      1.0
                                   Rural
                                                    N
    2
                      1.0
                                  Urban
                                                    Y
    3
                      1.0
                                   Urban
                                                    Y
    4
                      1.0
                                   Urban
                                                    Y
    . .
    609
                      1.0
                                  Rural
                                                    Y
                                                    Y
    610
                      1.0
                                   Rural
                                                    Y
                      1.0
                                   Urban
    611
                      1.0
                                                    Y
    612
                                   Urban
```

0.0 Semiurban N

[614 rows x 13 columns]

```
[7]: print ("NaN replaced with '0':")
     print( df.fillna(method='pad'))
    NaN replaced with '0':
           Loan_ID Gender Married Dependents
                                                      Education Self_Employed \
    0
          LP001002
                       Male
                                  No
                                                       Graduate
                                                                             No
    1
          LP001003
                       Male
                                               1
                                                       Graduate
                                                                             No
                                 Yes
    2
                       Male
                                               0
          LP001005
                                                       Graduate
                                                                            Yes
                                 Yes
    3
          LP001006
                       Male
                                 Yes
                                               0
                                                  Not Graduate
                                                                             No
    4
          LP001008
                                               0
                                                       Graduate
                       Male
                                  No
                                                                             No
     . .
          LP002978
                                               0
                                                       Graduate
    609
                     Female
                                  No
                                                                             No
    610
          LP002979
                       Male
                                 Yes
                                              3+
                                                       Graduate
                                                                             No
    611
          LP002983
                       Male
                                 Yes
                                               1
                                                       Graduate
                                                                             No
                                               2
    612
          LP002984
                       Male
                                 Yes
                                                       Graduate
                                                                             No
    613
          LP002990
                    Female
                                  No
                                               0
                                                       Graduate
                                                                            Yes
          ApplicantIncome
                            CoapplicantIncome
                                                 LoanAmount
                                                               Loan_Amount_Term
    0
                      5849
                                            0.0
                                                         NaN
                                                                           360.0
                      4583
                                         1508.0
                                                       128.0
                                                                           360.0
    1
    2
                      3000
                                            0.0
                                                        66.0
                                                                           360.0
    3
                      2583
                                         2358.0
                                                       120.0
                                                                           360.0
    4
                      6000
                                            0.0
                                                       141.0
                                                                           360.0
     . .
                       •••
                                            0.0
                                                                           360.0
    609
                      2900
                                                        71.0
    610
                      4106
                                            0.0
                                                        40.0
                                                                           180.0
    611
                      8072
                                          240.0
                                                       253.0
                                                                           360.0
    612
                      7583
                                            0.0
                                                       187.0
                                                                           360.0
    613
                      4583
                                            0.0
                                                       133.0
                                                                           360.0
          Credit_History Property_Area Loan_Status
    0
                      1.0
                                   Urban
                                                     Y
                      1.0
                                   Rural
                                                     N
    1
    2
                      1.0
                                   Urban
                                                     Y
    3
                      1.0
                                   Urban
                                                     Y
    4
                      1.0
                                   Urban
                                                     Y
    . .
                                                     Y
    609
                      1.0
                                   Rural
                                                     Y
                                   Rural
    610
                      1.0
                                                     Y
    611
                      1.0
                                   Urban
    612
                      1.0
                                   Urban
                                                     Υ
                      0.0
                               Semiurban
                                                     N
    613
```

[614 rows x 13 columns]

```
future version. Use obj.ffill() or obj.bfill() instead.
       print( df.fillna(method='pad'))
 [8]: print (df['Loan_ID'].isnull())
     0
            False
     1
            False
     2
            False
     3
            False
     4
            False
     609
            False
     610
            False
     611
             False
     612
             False
     613
             False
     Name: Loan_ID, Length: 614, dtype: bool
 [9]: print (df['Dependents'].notnull())
     0
             True
     1
             True
     2
             True
     3
             True
     4
             True
     609
             True
     610
             True
     611
             True
     612
             True
     613
             True
     Name: Dependents, Length: 614, dtype: bool
[10]: print( df['Self_Employed'].isnull())
     0
             False
     1
             False
     2
             False
     3
             False
     4
             False
     609
            False
     610
            False
            False
     611
     612
            False
     613
            False
```

C:\Users\Prateek\AppData\Local\Temp\ipykernel\_17908\2002809902.py:2:

FutureWarning: DataFrame.fillna with 'method' is deprecated and will raise in a

Name: Self\_Employed, Length: 614, dtype: bool [11]: print(df) print ("NaN replaced with '0':") print( df.fillna(0)) Loan\_ID Gender Married Dependents Education Self\_Employed \ 0 LP001002 Male No 0 Graduate No LP001003 Male Graduate 1 Yes 1 No 2 LP001005 Male Yes 0 Graduate Yes Male Not Graduate 3 LP001006 Yes 0 No 4 0 Graduate LP001008 Male No No . . LP002978 0 609 Female No Graduate No LP002979 Male 3+ Graduate 610 Yes No 611 LP002983 Male Yes 1 Graduate No 612 LP002984 Male Yes 2 Graduate No 613 LP002990 Female No 0 Graduate Yes ApplicantIncome CoapplicantIncome LoanAmount Loan\_Amount\_Term 0 5849 NaN 360.0 1 4583 1508.0 128.0 360.0 2 3000 66.0 360.0 0.0 3 2583 2358.0 120.0 360.0 4 6000 0.0 141.0 360.0 . . 609 2900 0.0 71.0 360.0 610 4106 0.0 40.0 180.0 611 240.0 253.0 8072 360.0 612 7583 0.0 187.0 360.0 613 4583 0.0 133.0 360.0 Credit\_History Property\_Area Loan\_Status 1.0 Urban Y 0 1 1.0 Rural N 2 1.0 Urban Y 3 Y 1.0 Urban 4 1.0 Y Urban Y 609 1.0 Rural Rural Y 610 1.0 Y 611 1.0 Urban Y 612 1.0 Urban 613 0.0 Semiurban N [614 rows x 13 columns]

NaN replaced with '0':

Loan\_ID Gender Married Dependents Education Self\_Employed \

```
0
           LP001002
                        Male
                                   No
                                                 0
                                                         Graduate
                                                                               No
           LP001003
                        Male
     1
                                   Yes
                                                 1
                                                         Graduate
                                                                               No
     2
                        Male
           LP001005
                                   Yes
                                                 0
                                                         Graduate
                                                                              Yes
     3
           LP001006
                        Male
                                                 0
                                                    Not Graduate
                                                                               No
                                   Yes
     4
                                                 0
           LP001008
                        Male
                                    No
                                                         Graduate
                                                                               No
      . .
     609
           LP002978
                      Female
                                    No
                                                 0
                                                         Graduate
                                                                               No
     610
           LP002979
                        Male
                                   Yes
                                                3+
                                                         Graduate
                                                                               No
           LP002983
                        Male
                                   Yes
                                                         Graduate
                                                                               No
     611
                                                 1
           LP002984
                                                 2
     612
                        Male
                                   Yes
                                                         Graduate
                                                                               No
           LP002990
                                                 0
     613
                      Female
                                    No
                                                         Graduate
                                                                              Yes
           ApplicantIncome
                              CoapplicantIncome LoanAmount
                                                                 Loan_Amount_Term
     0
                                                                             360.0
                       5849
                                              0.0
                                                           0.0
                                          1508.0
     1
                       4583
                                                         128.0
                                                                             360.0
     2
                       3000
                                              0.0
                                                          66.0
                                                                             360.0
     3
                       2583
                                          2358.0
                                                         120.0
                                                                             360.0
     4
                       6000
                                              0.0
                                                         141.0
                                                                             360.0
                        ...
      . .
     609
                       2900
                                              0.0
                                                          71.0
                                                                             360.0
     610
                       4106
                                              0.0
                                                          40.0
                                                                             180.0
                                           240.0
     611
                       8072
                                                         253.0
                                                                             360.0
     612
                       7583
                                              0.0
                                                         187.0
                                                                             360.0
     613
                       4583
                                              0.0
                                                         133.0
                                                                             360.0
           Credit_History Property_Area Loan_Status
     0
                       1.0
                                     Urban
                                                      Y
                       1.0
                                                      N
     1
                                     Rural
                                                       Y
     2
                       1.0
                                     Urban
     3
                       1.0
                                     Urban
                                                       Y
     4
                       1.0
                                     Urban
                                                       Υ
                                                      Y
     609
                       1.0
                                     Rural
     610
                       1.0
                                     Rural
                                                      Y
                                                      Y
     611
                       1.0
                                     Urban
                                                       Y
     612
                       1.0
                                     Urban
                       0.0
                                Semiurban
                                                       N
     613
      [614 rows x 13 columns]
[12]: df = df.dropna() #drops rows with null values
```

8

1

0

Education Self\_Employed

No

Yes

No

Graduate

Graduate

Not Graduate

Gender Married Dependents

Yes

Yes

Yes

Male

Male

Male

[13]: print(df)

1

2

3

Loan\_ID

LP001003

LP001005

LP001006

```
LP001008
                                                      Graduate
     4
                       Male
                                  No
                                               0
                                                                            No
     5
           LP001011
                       Male
                                 Yes
                                               2
                                                      Graduate
                                                                           Yes
      . .
     609
          LP002978
                    Female
                                  No
                                               0
                                                      Graduate
                                                                            No
                       Male
                                                                            No
     610
          LP002979
                                 Yes
                                              3+
                                                      Graduate
     611 LP002983
                       Male
                                 Yes
                                               1
                                                      Graduate
                                                                            No
                       Male
                                               2
                                                                            No
     612
          LP002984
                                 Yes
                                                      Graduate
     613 LP002990 Female
                                  No
                                                      Graduate
                                               0
                                                                           Yes
           ApplicantIncome
                            CoapplicantIncome LoanAmount Loan_Amount_Term
     1
                      4583
                                         1508.0
                                                      128.0
                                                                          360.0
     2
                      3000
                                            0.0
                                                       66.0
                                                                          360.0
     3
                      2583
                                         2358.0
                                                      120.0
                                                                          360.0
     4
                                            0.0
                      6000
                                                      141.0
                                                                          360.0
     5
                      5417
                                        4196.0
                                                      267.0
                                                                          360.0
      . .
                       •••
     609
                      2900
                                            0.0
                                                       71.0
                                                                          360.0
     610
                      4106
                                            0.0
                                                       40.0
                                                                          180.0
     611
                      8072
                                          240.0
                                                      253.0
                                                                          360.0
     612
                                            0.0
                                                      187.0
                      7583
                                                                          360.0
     613
                                            0.0
                                                      133.0
                      4583
                                                                          360.0
           Credit_History Property_Area Loan_Status
                      1.0
                                   Rural
     1
     2
                      1.0
                                   Urban
                                                    Υ
     3
                                                    Y
                      1.0
                                   Urban
     4
                      1.0
                                   Urban
                                                    Y
     5
                      1.0
                                                    Y
                                   Urban
      . .
     609
                      1.0
                                   Rural
                                                    Y
                                                    Y
     610
                                   Rural
                      1.0
                                                    Y
     611
                      1.0
                                   Urban
                      1.0
                                   Urban
                                                    Y
     612
     613
                      0.0
                               Semiurban
                                                    N
      [480 rows x 13 columns]
[14]: import pandas as pd
      import numpy as np
      df = pd.read_csv("train.csv")
      df.head()
          Loan ID Gender Married Dependents
[14]:
                                                   Education Self Employed \
      0 LP001002
                     Male
                               No
                                            0
                                                    Graduate
                                                                         No
      1 LP001003
                     Male
                              Yes
                                            1
                                                    Graduate
                                                                         No
      2 LP001005
                     Male
                              Yes
                                            0
                                                    Graduate
                                                                        Yes
      3 LP001006
                     Male
                              Yes
                                               Not Graduate
                                                                         No
```

```
4 LP001008
                     Male
                               No
                                            0
                                                    Graduate
                                                                         No
         ApplicantIncome
                           CoapplicantIncome
                                               LoanAmount
                                                            Loan_Amount_Term \
      0
                     5849
                                          0.0
                                                       NaN
                                                                        360.0
      1
                     4583
                                       1508.0
                                                     128.0
                                                                        360.0
      2
                     3000
                                          0.0
                                                      66.0
                                                                        360.0
      3
                     2583
                                       2358.0
                                                     120.0
                                                                        360.0
      4
                     6000
                                          0.0
                                                     141.0
                                                                        360.0
         Credit_History Property_Area Loan_Status
      0
                     1.0
                                 Urban
      1
                     1.0
                                 Rural
                                                  N
      2
                                                  Y
                     1.0
                                 Urban
                                                  Y
      3
                     1.0
                                 Urban
      4
                     1.0
                                 Urban
                                                  Y
[15]: to_drop = ['Gender', 'Married']
      #df.drop(columns=to_drop, inplace=True)
      df.drop(to_drop, inplace=True, axis=1)
[16]: df.head()
[16]:
          Loan_ID Dependents
                                   Education Self_Employed
                                                             ApplicantIncome
        LP001002
                                    Graduate
                                                                         5849
                                                         No
      1 LP001003
                            1
                                    Graduate
                                                         No
                                                                         4583
      2 LP001005
                            0
                                    Graduate
                                                        Yes
                                                                         3000
      3 LP001006
                            0
                               Not Graduate
                                                         No
                                                                         2583
      4 LP001008
                            0
                                    Graduate
                                                         No
                                                                         6000
         CoapplicantIncome LoanAmount Loan_Amount_Term
                                                             Credit_History \
      0
                        0.0
                                     NaN
                                                      360.0
                                                                         1.0
                     1508.0
                                   128.0
                                                      360.0
                                                                         1.0
      1
      2
                                    66.0
                        0.0
                                                      360.0
                                                                         1.0
      3
                     2358.0
                                   120.0
                                                      360.0
                                                                         1.0
      4
                        0.0
                                   141.0
                                                      360.0
                                                                         1.0
        Property_Area Loan_Status
                Urban
      0
                                 Y
      1
                Rural
                                 N
      2
                Urban
                                 Y
      3
                Urban
                                 Y
      4
                Urban
                                 Y
[17]: df.info()
```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 614 entries, 0 to 613

```
#
          Column
                               Non-Null Count
                                               Dtype
           _____
                               _____
      0
          Loan_ID
                               614 non-null
                                                object
          Dependents
                                                object
      1
                               599 non-null
      2
          Education
                               614 non-null
                                                object
      3
          Self Employed
                               582 non-null
                                                object
                                                int64
      4
          ApplicantIncome
                               614 non-null
      5
          CoapplicantIncome
                               614 non-null
                                               float64
      6
          LoanAmount
                                               float64
                               592 non-null
      7
          Loan_Amount_Term
                                               float64
                               600 non-null
      8
          Credit_History
                               564 non-null
                                               float64
      9
          Property_Area
                               614 non-null
                                               object
      10 Loan_Status
                               614 non-null
                                                object
     dtypes: float64(4), int64(1), object(6)
     memory usage: 52.9+ KB
[18]:
     df.shape
[18]: (614, 11)
      df.count()
[19]:
[19]: Loan_ID
                            614
      Dependents
                            599
      Education
                            614
      Self_Employed
                            582
      ApplicantIncome
                            614
      CoapplicantIncome
                            614
      LoanAmount
                            592
      Loan_Amount_Term
                            600
      Credit_History
                            564
      Property_Area
                            614
      Loan_Status
                            614
      dtype: int64
[20]: df.isnull()
[20]:
           Loan_ID
                    Dependents
                                 Education
                                             Self_Employed
                                                             ApplicantIncome
      0
             False
                          False
                                      False
                                                     False
                                                                       False
      1
             False
                          False
                                     False
                                                     False
                                                                       False
      2
             False
                          False
                                     False
                                                     False
                                                                       False
      3
             False
                          False
                                     False
                                                     False
                                                                       False
                          False
      4
             False
                                     False
                                                     False
                                                                       False
               •••
      . .
             False
                                                                       False
      609
                          False
                                     False
                                                     False
      610
             False
                          False
                                     False
                                                     False
                                                                       False
```

Data columns (total 11 columns):

611	F-1	E-l-	- Falas	E-1	F-1	
611		Fals			False	
612		Fals			False	
613	B False	Fals	e False	False	False	
	Coapplicar	ntIncome	LoanAmount	Loan_Amount_Term	Credit_History	\
0		False	True	False	False	
1		False	False	False	False	
2		False	False	False	False	
3		False	False	False	False	
4		False	False	False	False	
		•••	***	•••	•••	
609	)	False	False	False	False	
610	)	False	False	False	False	
611	L	False	False	False	False	
612	2	False	False	False	False	
613	3	False	False	False	False	
	Property_A	rea Loa	n_Status			
0	Fa	alse	False			
1	Fa	alse	False			
2	Fa	alse	False			
3	Fa	alse	False			
4	Fa	alse	False			
			•••			
609	) Fa	alse	False			
610	) Fa	alse	False			
611	L Fa	alse	False			
612	2 Fa	alse	False			
613	B Fa	alse	False			
[61	4 rows x 11	columns]				
[21]: mis	ssing_values=	df.isnul	1()			
[00].	adam l	d+				
[22]: mis	ssing_values	atypes				
[22]: Loa	an_ID	bo	ol			
	endents	bo	ol			
_	cation	ро	ol			

#### [22]: Education bool Self\_Employed bool ApplicantIncome bool CoapplicantIncome bool LoanAmount bool Loan\_Amount\_Term bool Credit\_History bool Property\_Area bool Loan\_Status bool

### dtype: object

```
no_missing_values=missing_values.sum()
[23]:
[24]: missing_values.sum()
[24]: Loan_ID
                            0
      Dependents
                            15
      Education
                            0
      Self_Employed
                            32
      ApplicantIncome
                            0
      CoapplicantIncome
                            0
      LoanAmount
                            22
      Loan_Amount_Term
                            14
      Credit_History
                            50
      Property_Area
                             0
                             0
      Loan_Status
      dtype: int64
[25]: len(df)
[25]: 614
[26]: no_missing_values/len(df)
[26]: Loan_ID
                            0.000000
      Dependents
                            0.024430
      Education
                            0.000000
      Self_Employed
                            0.052117
      ApplicantIncome
                            0.000000
      CoapplicantIncome
                            0.000000
      LoanAmount
                            0.035831
      Loan_Amount_Term
                            0.022801
      Credit_History
                            0.081433
      Property_Area
                            0.000000
      Loan_Status
                            0.000000
      dtype: float64
[27]: no_missing_values/len(df)*100
[27]: Loan_ID
                            0.000000
      Dependents
                            2.442997
      Education
                            0.000000
      Self_Employed
                            5.211726
      ApplicantIncome
                            0.000000
      CoapplicantIncome
                            0.000000
      LoanAmount
                            3.583062
```

 Loan\_Amount\_Term
 2.280130

 Credit\_History
 8.143322

 Property\_Area
 0.000000

 Loan\_Status
 0.000000

dtype: float64

```
[28]: df.isnull().mean().round(4) * 100
```

0.00 [28]: Loan\_ID Dependents 2.44 Education 0.00 Self\_Employed 5.21 ApplicantIncome 0.00 CoapplicantIncome 0.00 LoanAmount 3.58 Loan\_Amount\_Term 2.28 Credit\_History 8.14 Property\_Area 0.00 Loan\_Status 0.00 dtype: float64

[29]: # https://towardsdatascience.com/  $\Rightarrow data-cleaning-in-python-the-ultimate-guide-2020-c63b88bf0a0d$