DSBDA practical3

May 12, 2023

```
[2]: import pandas as pd
     import numpy as np
     import statistics as st
```

```
[6]: # Load the data
     df=pd.read_csv("D:\College Practicals\DSBDApractical3\Loan_Default.csv")
```

```
[7]: print(df.shape)
```

```
(148670, 34)
```

```
[8]: print(df.info())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 148670 entries, 0 to 148669
Data columns (total 34 columns):
 #   Column                 Non-Null Count   Dtype
---  ------                 --------------   -----
 0   ID                     148670 non-null  int64
 1   year                   148670 non-null  int64
 2   loan_limit             145326 non-null  object
 3   Gender                 148670 non-null  object
 4   approv_in_adv          147762 non-null  object
 5   loan_type              148670 non-null  object
 6   loan_purpose           148536 non-null  object
 7   Credit_Worthiness      148670 non-null  object
 8   open_credit            148670 non-null  object
 9   business_or_commercial 148670 non-null  object
 10  loan_amount            148670 non-null  int64
 11  rate_of_interest       112231 non-null  float64
 12  Interest_rate_spread   112031 non-null  float64
 13  Upfront_charges        109028 non-null  float64
 14  term                   148629 non-null  float64
 15  Neg_ammortization      148549 non-null  object
 16  interest_only          148670 non-null  object
 17  lump_sum_payment       148670 non-null  object
 18  property_value         133572 non-null  float64
 19  construction_type      148670 non-null  object
 20  occupancy_type         148670 non-null  object
```

1

```
21  Secured_by               148670 non-null  object
22  total_units              148670 non-null  object
23  income                   139520 non-null  float64
24  credit_type              148670 non-null  object
25  Credit_Score             148670 non-null  int64
26  co-applicant_credit_type 148670 non-null  object
27  age                      148470 non-null  object
28  submission_of_application 148470 non-null object
29  LTV                      133572 non-null  float64
30  Region                   148670 non-null  object
31  Security_Type            148670 non-null  object
32  Status                   148670 non-null  int64
33  dtir1                    124549 non-null  float64
dtypes: float64(8), int64(5), object(21)
memory usage: 38.6+ MB
None
```

[9]: `df.mean`

[9]:
```
<bound method NDFrame._add_numeric_operations.<locals>.mean of            ID
       year loan_limit        Gender approv_in_adv loan_type  \
0       24890  2019      cf  Sex Not Available        nopre     type1
1       24891  2019      cf               Male        nopre     type2
2       24892  2019      cf               Male          pre     type1
3       24893  2019      cf               Male        nopre     type1
4       24894  2019      cf              Joint          pre     type1
...       ...   ...     ...                ...          ...       ...
148665 173555  2019      cf  Sex Not Available        nopre     type1
148666 173556  2019      cf               Male        nopre     type1
148667 173557  2019      cf               Male        nopre     type1
148668 173558  2019      cf             Female        nopre     type1
148669 173559  2019      cf             Female        nopre     type1

        loan_purpose Credit_Worthiness open_credit business_or_commercial  … \
0               p1                l1        nopc                  nob/c  …
1               p1                l1        nopc                    b/c  …
2               p1                l1        nopc                  nob/c  …
3               p4                l1        nopc                  nob/c  …
4               p1                l1        nopc                  nob/c  …
...            ...               ...         ...                    ...  …
148665          p3                l1        nopc                  nob/c  …
148666          p1                l1        nopc                  nob/c  …
148667          p4                l1        nopc                  nob/c  …
148668          p4                l1        nopc                  nob/c  …
148669          p3                l1        nopc                  nob/c  …

        credit_type  Credit_Score  co-applicant_credit_type     age  \
```

```
0                EXP       758                            CIB   25-34
1                EQUI      552                            EXP   55-64
2                EXP       834                            CIB   35-44
3                EXP       587                            CIB   45-54
4                CRIF      602                            EXP   25-34
...              ...       ...                            ...   ...
148665           CIB       659                            EXP   55-64
148666           CIB       569                            CIB   25-34
148667           CIB       702                            EXP   45-54
148668           EXP       737                            EXP   55-64
148669           CIB       830                            CIB   45-54

        submission_of_application        LTV  Region  Security_Type  Status  \
0                          to_inst  98.728814   south         direct       1
1                          to_inst        NaN   North         direct       1
2                          to_inst  80.019685   south         direct       0
3                         not_inst  69.376900   North         direct       0
4                         not_inst  91.886544   North         direct       0
...                            ...        ...     ...            ...     ...
148665                     to_inst  71.792763   south         direct       0
148666                    not_inst  74.428934   south         direct       0
148667                    not_inst  61.332418   North         direct       0
148668                     to_inst  70.683453   North         direct       0
148669                    not_inst  72.849462   North         direct       0

        dtir1
0        45.0
1         NaN
2        46.0
3        42.0
4        39.0
...       ...
148665   48.0
148666   15.0
148667   49.0
148668   29.0
148669   44.0

[148670 rows x 34 columns]>
```

```python
[10]: print(df.loc[:,'year'].median())
```

```
2019.0
```

```python
[11]: print(df.loc[:,'income'].median())
```

```
5760.0
```

```
[12]: df.median(axis=1)[0:5]
```

C:\Users\lalit\AppData\Local\Temp\ipykernel_18120\512305079.py:1: FutureWarning:
Dropping of nuisance columns in DataFrame reductions (with 'numeric_only=None')
is deprecated; in a future version this will raise TypeError. Select only valid
columns before calling the reduction.
  df.median(axis=1)[0:5]

```
[12]: 0    1249.0
      1    2019.0
      2     595.0
      3     473.5
      4     360.0
      dtype: float64
```

```
[13]: df.mode()
```

```
[13]:             ID     year loan_limit Gender approv_in_adv loan_type loan_purpose  \
      0        24890  2019.0         cf   Male         nopre     type1           p3
      1        24891     NaN        NaN    NaN           NaN       NaN          NaN
      2        24892     NaN        NaN    NaN           NaN       NaN          NaN
      3        24893     NaN        NaN    NaN           NaN       NaN          NaN
      4        24894     NaN        NaN    NaN           NaN       NaN          NaN
      ...        ...     ...        ...    ...           ...       ...          ...
      148665  173555     NaN        NaN    NaN           NaN       NaN          NaN
      148666  173556     NaN        NaN    NaN           NaN       NaN          NaN
      148667  173557     NaN        NaN    NaN           NaN       NaN          NaN
      148668  173558     NaN        NaN    NaN           NaN       NaN          NaN
      148669  173559     NaN        NaN    NaN           NaN       NaN          NaN

             Credit_Worthiness open_credit business_or_commercial  ... credit_type  \
      0                     l1        nopc                  nob/c  ...         CIB
      1                    NaN         NaN                    NaN  ...         NaN
      2                    NaN         NaN                    NaN  ...         NaN
      3                    NaN         NaN                    NaN  ...         NaN
      4                    NaN         NaN                    NaN  ...         NaN
      ...                  ...         ...                    ...  ... ...      ...
      148665               NaN         NaN                    NaN  ...         NaN
      148666               NaN         NaN                    NaN  ...         NaN
      148667               NaN         NaN                    NaN  ...         NaN
      148668               NaN         NaN                    NaN  ...         NaN
      148669               NaN         NaN                    NaN  ...         NaN

             Credit_Score  co-applicant_credit_type    age  \
      0             763.0                       CIB  45-54
      1               NaN                       NaN    NaN
      2               NaN                       NaN    NaN
      3               NaN                       NaN    NaN
```

```
4                    NaN                        NaN      NaN
…                     …                          …        …
148665               NaN                        NaN      NaN
148666               NaN                        NaN      NaN
148667               NaN                        NaN      NaN
148668               NaN                        NaN      NaN
148669               NaN                        NaN      NaN

        submission_of_application     LTV  Region  Security_Type  Status  dtir1
0                      to_inst      81.25   North         direct     0.0   37.0
1                          NaN       NaN     NaN            NaN      NaN    NaN
2                          NaN       NaN     NaN            NaN      NaN    NaN
3                          NaN       NaN     NaN            NaN      NaN    NaN
4                          NaN       NaN     NaN            NaN      NaN    NaN
…                           …         …       …              …        …      …
148665                     NaN       NaN     NaN            NaN      NaN    NaN
148666                     NaN       NaN     NaN            NaN      NaN    NaN
148667                     NaN       NaN     NaN            NaN      NaN    NaN
148668                     NaN       NaN     NaN            NaN      NaN    NaN
148669                     NaN       NaN     NaN            NaN      NaN    NaN

[148670 rows x 34 columns]
```

[14]:
```python
df.std()
```

```
C:\Users\lalit\AppData\Local\Temp\ipykernel_18120\3390915376.py:1:
FutureWarning: Dropping of nuisance columns in DataFrame reductions (with
'numeric_only=None') is deprecated; in a future version this will raise
TypeError.  Select only valid columns before calling the reduction.
  df.std()
```

[14]:
```
ID                    42917.476598
year                      0.000000
loan_amount          183909.310127
rate_of_interest          0.561391
Interest_rate_spread      0.513043
Upfront_charges        3251.121510
term                     58.409084
property_value       359935.315562
income                 6496.586382
Credit_Score            115.875857
LTV                      39.967603
Status                    0.430942
dtir1                    10.545435
dtype: float64
```

[15]:
```python
print(df.loc[:,'year'].std())
```

```
      0.0
```

```
[16]: print(df.loc[:,'year'].std())
```

```
      0.0
```

```
[17]: df.std(axis=1)[0:5]
```

C:\Users\lalit\AppData\Local\Temp\ipykernel_18120\696587434.py:1: FutureWarning:
Dropping of nuisance columns in DataFrame reductions (with 'numeric_only=None')
is deprecated; in a future version this will raise TypeError. Select only valid
columns before calling the reduction.
  df.std(axis=1)[0:5]

```
[17]: 0      48456.424600
      1      76496.488376
      2     171800.879225
      3     219726.466641
      4     272175.749915
      dtype: float64
```

```
[18]: df.var()
```

C:\Users\lalit\AppData\Local\Temp\ipykernel_18120\1568254755.py:1:
FutureWarning: Dropping of nuisance columns in DataFrame reductions (with
'numeric_only=None') is deprecated; in a future version this will raise
TypeError. Select only valid columns before calling the reduction.
  df.var()

```
[18]: ID                   1.841910e+09
      year                 0.000000e+00
      loan_amount          3.382263e+10
      rate_of_interest     3.151601e-01
      Interest_rate_spread 2.632128e-01
      Upfront_charges      1.056979e+07
      term                 3.411621e+03
      property_value       1.295534e+11
      income               4.220563e+07
      Credit_Score         1.342721e+04
      LTV                  1.597409e+03
      Status               1.857112e-01
      dtir1                1.112062e+02
      dtype: float64
```

```
[19]: from scipy.stats import iqr
      iqr(df['income'])
```

```
[19]: nan
```

```
[20]: print(df.skew())
```

```
ID                     0.000000
year                   0.000000
loan_amount            1.666998
rate_of_interest       0.388406
Interest_rate_spread   0.280762
Upfront_charges        1.754076
term                  -2.174822
property_value         4.586276
income                17.307695
Credit_Score           0.004767
LTV                  120.615337
Status                 1.176762
dtir1                 -0.551465
dtype: float64
```

```
C:\Users\lalit\AppData\Local\Temp\ipykernel_18120\1926848427.py:1:
FutureWarning: Dropping of nuisance columns in DataFrame reductions (with
'numeric_only=None') is deprecated; in a future version this will raise
TypeError.  Select only valid columns before calling the reduction.
  print(df.skew())
```

```
[ ]:
```