## DATA ANALYSIS

April 15, 2023

## DATA ANALYSIS PROJECT

Reading data using pandas

```
[1]: import pandas as pd
[2]:
    df = pd.read_excel('C:\DATA_SET\defaultOfCreditCardClients.xlsx')
     df.head()
[3]:
[3]:
             LIMIT_BAL
                               EDUCATION
                                           MARRIAGE
                                                       AGE
                                                            PAY_0
                                                                    PAY_2
                                                                            PAY_3
                                                                                    PAY_4
         ID
                         SEX
     0
          1
                 20000
                            2
                                        2
                                                   1
                                                        24
                                                                 2
                                                                         2
                                                                                -1
                                                                                        -1
          2
                120000
                                        2
                                                   2
                                                        26
                                                                         2
                                                                                 0
                                                                                        0
     1
                            2
                                                                -1
     2
          3
                 90000
                            2
                                        2
                                                   2
                                                        34
                                                                 0
                                                                         0
                                                                                 0
                                                                                        0
                                        2
     3
          4
                 50000
                            2
                                                   1
                                                        37
                                                                 0
                                                                         0
                                                                                 0
                                                                                        0
     4
                 50000
                                        2
                                                   1
          5
                            1
                                                        57
                                                                -1
                                                                         0
                                                                                -1
                                                                                        0
                        BILL_AMT5
                                                PAY_AMT1
                                                                       PAY_AMT3
            BILL_AMT4
                                    BILL_AMT6
                                                            PAY_AMT2
     0
                     0
                                 0
                                              0
                                                         0
                                                                  689
                                                                               0
                 3272
                              3455
                                          3261
                                                         0
                                                                 1000
                                                                            1000
     1
                                                                            1000
     2
                14331
                             14948
                                         15549
                                                      1518
                                                                 1500
     3
                28314
                             28959
                                         29547
                                                      2000
                                                                 2019
                                                                            1200
     4
                20940
                             19146
                                         19131
                                                      2000
                                                                36681
                                                                           10000
        PAY_AMT4
                    PAY_AMT5
                               PAY_AMT6
                                          default payment next month
     0
                0
                            0
                                       0
     1
             1000
                            0
                                    2000
                                                                       1
     2
             1000
                        1000
                                    5000
                                                                      0
     3
             1100
                        1069
                                    1000
                                                                      0
     4
             9000
                         689
                                     679
                                                                      0
     [5 rows x 25 columns]
```

## Cleaning the data

Capitalize all the column names

```
[4]: df.columns = df.columns.str.upper()
```

```
[5]: df.head()
 [5]:
         ID
            LIMIT BAL
                        SEX EDUCATION MARRIAGE AGE PAY 0 PAY 2 PAY 3 PAY 4 \
      0
          1
                 20000
                          2
                                     2
                                                1
                                                    24
                                                            2
                                                                   2
                                                                         -1
                                                                                 -1
                120000
      1
          2
                          2
                                     2
                                                2
                                                    26
                                                           -1
                                                                   2
                                                                          0
                                                                                 0
      2
                 90000
                          2
                                     2
                                                2
                                                            0
                                                                   0
                                                                          0
                                                                                 0
          3
                                                    34
                                     2
          4
                 50000
                          2
                                                1
                                                    37
                                                            0
                                                                   0
                                                                          0
                                                                                 0
      3
      4
                 50000
                                     2
                                                1
                                                                   0
          5
                          1
                                                    57
                                                           -1
                                                                         -1
                                                                                 0
            BILL_AMT4 BILL_AMT5 BILL_AMT6 PAY_AMT1 PAY_AMT2 PAY_AMT3 \
                                                             689
      0
                    0
                               0
                                          0
                                                     0
                                                                         0
                 3272
                            3455
                                                     0
                                                            1000
                                                                      1000
      1
                                       3261
      2
                14331
                           14948
                                       15549
                                                  1518
                                                            1500
                                                                      1000
      3
                28314
                           28959
                                       29547
                                                  2000
                                                            2019
                                                                      1200
      4
                20940
                           19146
                                       19131
                                                  2000
                                                           36681
                                                                     10000
         PAY_AMT4 PAY_AMT5 PAY_AMT6 DEFAULT PAYMENT NEXT MONTH
      0
                          0
             1000
                          0
                                 2000
      1
                                                                 1
             1000
                                 5000
      2
                       1000
                                                                 0
      3
             1100
                       1069
                                                                 0
                                 1000
      4
             9000
                                                                 0
                        689
                                  679
      [5 rows x 25 columns]
     Generating GENDER column using SEX column when value is 1 then MALE if value
     is 2 then FEMALE
 [6]: df.loc[df["SEX"] == 1, "SEX"] = "MALE"
 [7]: df.loc[df["SEX"] == 2, "SEX"] = "FEMALE"
 [8]: df.loc[df["MARRIAGE"] == 1, "MARRIAGE"] = "MARRIED"
      df.loc[df["MARRIAGE"] == 2, "MARRIAGE"] = "UNMARRIED"
      df.loc[df["MARRIAGE"] == 3, "MARRIAGE"] = "OTHER"
 [9]: df.loc[df["EDUCATION"] == 1, "EDUCATION"] = "GRADUATE"
      df.loc[df["EDUCATION"] == 2, "EDUCATION"] = "UNIVERSITY"
      df.loc[df["EDUCATION"] == 3, "EDUCATION"] = "HIGH SCHOOL"
      df.loc[df["EDUCATION"] == 4, "EDUCATION"] = "OTHERS"
[10]: pay_map = {0: 'PAY DULY',-1: 'PAY DULY'}
      df['PAY_0'] = df['PAY_0'].replace(pay_map)
[11]: df.head()
                                 EDUCATION
[11]:
         ID LIMIT_BAL
                           SEX
                                              MARRIAGE AGE
                                                                PAY_O PAY_2 PAY_3 \
                 20000 FEMALE UNIVERSITY
      0
          1
                                               MARRIED
                                                         24
                                                                    2
                                                                           2
                                                                                  -1
```

```
1
         2
               120000 FEMALE UNIVERSITY UNMARRIED
                                                      26 PAY DULY
                                                                               0
     2
         3
                90000 FEMALE UNIVERSITY
                                                      34 PAY DULY
                                                                        0
                                                                               0
                                          UNMARRIED
     3
                                                                               0
        4
                50000 FEMALE UNIVERSITY
                                             MARRIED
                                                      37 PAY DULY
                                                                        0
     4
         5
                50000
                         MALE UNIVERSITY
                                             MARRIED
                                                      57 PAY DULY
                                                                        0
                                                                              -1
        PAY_4 ... BILL_AMT4 BILL_AMT5 BILL_AMT6 PAY_AMT1 PAY_AMT2 PAY_AMT3 \
           -1 ...
     0
                          0
                                     0
                                               0
                                                         0
                                                                 689
     1
            0 ...
                       3272
                                  3455
                                             3261
                                                         0
                                                                1000
                                                                          1000
     2
            0 ...
                                 14948
                                            15549
                                                                          1000
                      14331
                                                      1518
                                                                1500
     3
            0 ...
                      28314
                                 28959
                                            29547
                                                      2000
                                                                2019
                                                                          1200
     4
            0 ...
                      20940
                                 19146
                                            19131
                                                      2000
                                                               36681
                                                                         10000
        PAY_AMT4 PAY_AMT5 PAY_AMT6 DEFAULT PAYMENT NEXT MONTH
     0
               0
                         0
                                   0
                                                               1
            1000
                         0
                                2000
                                                              1
     1
     2
            1000
                      1000
                                5000
                                                              0
     3
            1100
                      1069
                                1000
                                                              0
     4
            9000
                       689
                                 679
                                                              0
     [5 rows x 25 columns]
[12]: pay map = {0: 'PAY DULY',-1: 'PREPAID 1', -2: 'PREPAID 2'}
     df['PAY_2'] = df['PAY_2'].replace(pay_map)
[13]: pay_map = {0: 'PAY DULY',-1: 'PREPAID 1', -2: 'PREPAID 2'}
     df['PAY_3'] = df['PAY_3'].replace(pay_map)
[14]: pay_map = {0: 'PAY DULY',-1: 'PREPAID 1', -2: 'PREPAID 2'}
     df['PAY_4'] = df['PAY_4'].replace(pay_map)
[15]: pay_map = {0: 'PAY DULY',-1: 'PREPAID 1', -2: 'PREPAID 2'}
     df['PAY 5'] = df['PAY 5'].replace(pay map)
[16]: pay map = {0: 'PAY DULY',-1: 'PREPAID 1', -2: 'PREPAID 2'}
     df['PAY_6'] = df['PAY_6'].replace(pay_map)
[17]: df.head()
                                                                       PAY 2 \
[17]:
        ID
           LIMIT BAL
                          SEX
                                EDUCATION
                                            MARRIAGE AGE
                                                             PAY 0
     0
         1
                20000 FEMALE UNIVERSITY
                                             MARRIED
                                                      24
                                                                 2
                                                                           2
         2
               120000 FEMALE UNIVERSITY UNMARRIED
     1
                                                      26 PAY DULY
     2
                90000 FEMALE UNIVERSITY UNMARRIED
         3
                                                      34 PAY DULY PAY DULY
                50000 FEMALE UNIVERSITY
     3
                                             MARRIED
                                                          PAY DULY PAY DULY
                                                      37
                50000
        5
                         MALE UNIVERSITY
                                             MARRIED
                                                      57 PAY DULY PAY DULY
                       PAY_4 ... BILL_AMT4 BILL_AMT5 BILL_AMT6 PAY_AMT1 \
            PAY 3
     O PREPAID 1 PREPAID 1 ... O
                                                 0
                                                            0
                                                                      0
```

```
PAY DULY ...
      2
         PAY DULY
                     PAY DULY
                                      14331
                                                14948
                                                           15549
                                                                       1518
      3
          PAY DULY
                     PAY DULY ...
                                      28314
                                                28959
                                                           29547
                                                                       2000
      4 PREPAID 1
                     PAY DULY ...
                                      20940
                                                                       2000
                                                19146
                                                           19131
         PAY_AMT2 PAY_AMT3 PAY_AMT4 PAY_AMT5 PAY_AMT6
      0
                          0
                                     0
                                               0
              689
      1
             1000
                       1000
                                 1000
                                               0
                                                      2000
      2
                       1000
                                  1000
                                                      5000
             1500
                                            1000
      3
             2019
                       1200
                                  1100
                                            1069
                                                      1000
      4
            36681
                      10000
                                  9000
                                             689
                                                       679
         DEFAULT PAYMENT NEXT MONTH
      0
      1
                                   1
      2
                                  0
      3
                                  0
      4
      [5 rows x 25 columns]
[18]: df = df.rename(columns={'PAY 0': 'PAID APRL2005', 'PAY 2':11
       ↔ 'PAID_MAY2005', 'PAY_3': 'PAID_JUNE2005', 'PAY_4': 'PAID_JULY2005',
                               'PAY_5': 'PAID_AUG2005', 'PAY_6': 'PAID_SEP2005'})
[19]: df.head()
[19]:
         ID
            LIMIT_BAL
                           SEX
                                 EDUCATION
                                              MARRIAGE AGE PAID_APRL2005 \
                 20000 FEMALE UNIVERSITY
                                                         24
      0
          1
                                               MARRIED
          2
      1
                120000 FEMALE UNIVERSITY UNMARRIED
                                                         26
                                                                 PAY DULY
      2
                 90000 FEMALE UNIVERSITY
                                                                 PAY DULY
          3
                                             UNMARRIED
                                                         34
      3
          4
                 50000
                        FEMALE UNIVERSITY
                                                         37
                                                                 PAY DULY
                                               MARRIED
      4
          5
                 50000
                          MALE UNIVERSITY
                                                         57
                                                                 PAY DULY
                                               MARRIED
        PAID_MAY2005 PAID_JUNE2005 PAID_JULY2005 ... BILL_AMT4 BILL_AMT5 \
      0
                   2
                         PREPAID 1
                                       PREPAID 1 ...
                                                             0
                                                                        0
      1
                   2
                          PAY DULY
                                        PAY DULY ...
                                                          3272
                                                                     3455
      2
                          PAY DULY
                                         PAY DULY ...
                                                         14331
            PAY DULY
                                                                   14948
      3
            PAY DULY
                          PAY DULY
                                         PAY DULY ...
                                                         28314
                                                                   28959
      4
            PAY DULY
                         PREPAID 1
                                        PAY DULY ...
                                                         20940
                                                                   19146
         BILL_AMT6 PAY_AMT1 PAY_AMT2 PAY_AMT3 PAY_AMT4 PAY_AMT5 PAY_AMT6 \
      0
                 0
                           0
                                   689
                                                0
                                                          0
                                                                     0
                                                                               0
      1
              3261
                           0
                                  1000
                                             1000
                                                       1000
                                                                     0
                                                                            2000
      2
             15549
                        1518
                                  1500
                                             1000
                                                       1000
                                                                  1000
                                                                            5000
      3
                        2000
                                                                            1000
             29547
                                  2019
                                             1200
                                                       1100
                                                                  1069
      4
             19131
                        2000
                                  36681
                                            10000
                                                       9000
                                                                   689
                                                                             679
```

PAY DULY

```
0
      1
                                  1
      2
                                  0
      3
                                  0
                                  0
      [5 rows x 25 columns]
[20]: df = df.rename(columns={'BILL_AMT6': 'BILL_APRL2005', 'BILL_AMT5':
       G'BILL MAY2005', 'BILL_AMT4': 'BILL_JUNE2005', 'BILL_AMT3': 'BILL_JULY2005',
                              'BILL AMT2': 'BILL AUG2005', 'BILL AMT1':

¬'BILL_SEP2005'})
[21]: df.head()
         ID LIMIT_BAL
[21]:
                           SEX
                                 EDUCATION
                                             MARRIAGE AGE PAID_APRL2005 \
      0
          1
                 20000 FEMALE UNIVERSITY
                                              MARRIED
                                                        24
                                                                        2
                120000 FEMALE UNIVERSITY UNMARRIED
                                                                PAY DULY
      1
          2
                                                        26
      2
                90000 FEMALE UNIVERSITY UNMARRIED
                                                        34
                                                                PAY DULY
          3
      3
         4
                50000 FEMALE UNIVERSITY
                                              MARRIED
                                                        37
                                                                PAY DULY
          5
                50000
                          MALE UNIVERSITY
                                              MARRIED
                                                        57
                                                                PAY DULY
       PAID MAY2005 PAID JUNE2005 PAID JULY2005 ... BILL JUNE2005 BILL MAY2005 \
                                       PREPAID 1 ...
      0
                   2
                         PREPAID 1
                                                                0
                                       PAY DULY ...
      1
                   2
                         PAY DULY
                                                             3272
                                                                          3455
      2
           PAY DULY
                          PAY DULY
                                        PAY DULY ...
                                                            14331
                                                                          14948
      3
           PAY DULY
                         PAY DULY
                                        PAY DULY ...
                                                                          28959
                                                            28314
      4
           PAY DULY
                                        PAY DULY ...
                         PREPAID 1
                                                            20940
                                                                          19146
         BILL_APRL2005 PAY_AMT1 PAY_AMT2 PAY_AMT3 PAY_AMT4 PAY_AMT5 PAY_AMT6 \
      0
                     0
                               0
                                       689
                                                                       0
                                                   0
                                                             0
                                                                                  0
                               0
                                      1000
      1
                  3261
                                                1000
                                                          1000
                                                                       0
                                                                               2000
                                      1500
                                                1000
                                                          1000
                                                                    1000
                                                                               5000
      2
                 15549
                            1518
      3
                            2000
                                      2019
                                                1200
                                                          1100
                                                                    1069
                                                                               1000
                 29547
                 19131
                            2000
                                     36681
                                               10000
                                                          9000
                                                                     689
                                                                               679
         DEFAULT PAYMENT NEXT MONTH
      0
      1
                                  1
      2
                                  0
      3
                                  0
                                  0
```

[5 rows x 25 columns]

DEFAULT PAYMENT NEXT MONTH

```
[22]: df = df.rename(columns={'PAY_AMT6': 'BILL_PAID_APRL2005', 'PAY_AMT5':
       ↔ 'BILL_PAID_MAY2005', 'PAY_AMT4': 'BILL_PAID_JUNE2005', 'PAY_AMT3': ⊔
       ⇔'BILL_PAID_JULY2005',
                               'PAY AMT2': 'BILL PAID AUG2005', 'PAY AMT1':
       ⇔'BILL_PAID_SEP2005'})
[23]: df.head()
[23]:
         ID
             LIMIT_BAL
                           SEX
                                 EDUCATION
                                              MARRIAGE AGE PAID_APRL2005 \
                 20000
                        FEMALE UNIVERSITY
                                               MARRIED
                                                         24
      0
          1
          2
                120000 FEMALE UNIVERSITY UNMARRIED
                                                                 PAY DULY
      1
                                                         26
      2
          3
                 90000 FEMALE UNIVERSITY
                                             UNMARRIED
                                                         34
                                                                 PAY DULY
                 50000 FEMALE UNIVERSITY
                                               MARRIED
                                                         37
                                                                 PAY DULY
      3
          5
                 50000
                          MALE UNIVERSITY
                                               MARRIED
                                                                 PAY DULY
                                                         57
        PAID MAY2005 PAID JUNE2005 PAID JULY2005 ... BILL JUNE2005 BILL MAY2005 \
                         PREPAID 1
      0
                   2
                                        PREPAID 1
                                                                  0
                                         PAY DULY ...
                   2
                          PAY DULY
                                                              3272
                                                                            3455
      1
      2
            PAY DULY
                          PAY DULY
                                         PAY DULY
                                                              14331
                                                                           14948
            PAY DULY
                          PAY DULY
                                         PAY DULY ...
      3
                                                              28314
                                                                           28959
            PAY DULY
                         PREPAID 1
                                         PAY DULY ...
                                                              20940
                                                                           19146
         BILL APRL2005
                        BILL_PAID_SEP2005 BILL_PAID_AUG2005 BILL_PAID_JULY2005
      0
                                                          689
      1
                  3261
                                         0
                                                         1000
                                                                              1000
      2
                 15549
                                      1518
                                                         1500
                                                                              1000
      3
                                      2000
                                                         2019
                                                                              1200
                 29547
      4
                                      2000
                 19131
                                                        36681
                                                                             10000
         BILL_PAID_JUNE2005 BILL_PAID_MAY2005
                                                 BILL PAID APRL2005
      0
                                              0
                                                                   0
                          0
                       1000
                                              0
      1
                                                                2000
                                           1000
      2
                       1000
                                                                5000
      3
                       1100
                                           1069
                                                                1000
                       9000
                                            689
                                                                 679
         DEFAULT PAYMENT NEXT MONTH
      0
                                   1
      1
      2
                                   0
      3
                                   0
                                   0
      [5 rows x 25 columns]
[24]: df = df.rename(columns={'DEFAULT PAYMENT NEXT MONTH':
       ⇔'DEFAULT_PAYMENT_NEXT_MONTH'})
```

```
[25]: df.head()
[25]:
         ID
             LIMIT BAL
                            SEX
                                  EDUCATION
                                              MARRIAGE AGE PAID APRL2005
      0
          1
                 20000 FEMALE UNIVERSITY
                                                MARRIED
                                                          24
          2
                120000
                        FEMALE UNIVERSITY
                                                          26
                                                                  PAY DULY
      1
                                             UNMARRIED
      2
                 90000 FEMALE UNIVERSITY
                                             UNMARRIED
                                                          34
                                                                  PAY DULY
                 50000
                         FEMALE UNIVERSITY
                                                                  PAY DULY
      3
                                                MARRIED
                                                          37
      4
                                                MARRIED
          5
                 50000
                           MALE UNIVERSITY
                                                                  PAY DULY
                                                          57
        PAID_MAY2005 PAID_JUNE2005 PAID_JULY2005
                                                   ... BILL_JUNE2005 BILL_MAY2005 \
      0
                   2
                          PREPAID 1
                                        PREPAID 1
                                                                   0
                                                                                0
                                         PAY DULY
                   2
                           PAY DULY
                                                               3272
                                                                             3455
      1
      2
            PAY DULY
                           PAY DULY
                                         PAY DULY ...
                                                               14331
                                                                            14948
      3
            PAY DULY
                           PAY DULY
                                         PAY DULY
                                                               28314
                                                                            28959
      4
            PAY DULY
                          PREPAID 1
                                         PAY DULY ...
                                                               20940
                                                                            19146
         BILL_APRL2005 BILL_PAID_SEP2005 BILL_PAID_AUG2005 BILL_PAID_JULY2005
      0
                                                           689
                  3261
                                         0
                                                          1000
                                                                               1000
      1
      2
                 15549
                                      1518
                                                          1500
                                                                               1000
      3
                 29547
                                      2000
                                                          2019
                                                                               1200
      4
                                      2000
                                                         36681
                 19131
                                                                              10000
                                                 BILL_PAID_APRL2005
         BILL_PAID_JUNE2005
                             BILL_PAID_MAY2005
      0
                           0
                                                                    0
                        1000
                                               0
                                                                 2000
      1
      2
                        1000
                                            1000
                                                                5000
      3
                        1100
                                            1069
                                                                 1000
                        9000
                                             689
                                                                 679
         DEFAULT PAYMENT NEXT MONTH
      0
      1
                                   1
      2
                                   0
      3
                                   0
      4
                                   0
      [5 rows x 25 columns]
[26]: df = df.drop('DEFAULT_PAYMENT_NEXT_MONTH', axis=1)
[27]:
      df.head()
                                                         AGE PAID_APRL2005 \
[27]:
         ID LIMIT_BAL
                            SEX
                                  EDUCATION
                                               MARRIAGE
      0
          1
                 20000
                         FEMALE UNIVERSITY
                                                          24
                                                                          2
                                                MARRIED
      1
          2
                120000
                         FEMALE UNIVERSITY
                                              UNMARRIED
                                                          26
                                                                   PAY DULY
      2
          3
                                                          34
                 90000
                         FEMALE UNIVERSITY
                                              UNMARRIED
                                                                   PAY DULY
```

```
3
         4
                 50000 FEMALE UNIVERSITY
                                               MARRIED
                                                         37
                                                                 PAY DULY
      4
         5
                 50000
                          MALE UNIVERSITY
                                                         57
                                                                 PAY DULY
                                               MARRIED
        PAID MAY2005 PAID JUNE2005 PAID JULY2005 ... BILL JULY2005 BILL JUNE2005
      0
                   2
                         PREPAID 1
                                       PREPAID 1
                                                               689
                   2
                          PAY DULY
                                        PAY DULY
                                                              2682
                                                                             3272
      1
      2
            PAY DULY
                          PAY DULY
                                        PAY DULY ...
                                                             13559
                                                                            14331
            PAY DULY
                          PAY DULY
                                        PAY DULY ...
      3
                                                             49291
                                                                            28314
            PAY DULY
                                        PAY DULY ...
                         PREPAID 1
                                                             35835
                                                                            20940
         BILL MAY2005 BILL APRL2005 BILL PAID SEP2005 BILL PAID AUG2005 \
      0
                    0
                                   0
                                                       0
                                                                         689
                 3455
                                3261
                                                       0
                                                                        1000
      1
                                15549
      2
                14948
                                                    1518
                                                                        1500
      3
                28959
                               29547
                                                    2000
                                                                        2019
      4
                                                    2000
                19146
                                19131
                                                                       36681
         BILL_PAID_JULY2005
                            BILL_PAID_JUNE2005
                                                 BILL_PAID_MAY2005
      0
                          0
                                               0
                                                                  0
                       1000
                                            1000
                                                                  0
      1
      2
                       1000
                                            1000
                                                                1000
      3
                       1200
                                            1100
                                                                1069
      4
                      10000
                                            9000
                                                                689
         BILL_PAID_APRL2005
      0
                          0
                       2000
      1
      2
                       5000
      3
                       1000
                        679
      [5 rows x 24 columns]
[28]: df['TOTAL_BILL'] =
       →df[['BILL SEP2005','BILL AUG2005','BILL JULY2005','BILL JUNE2005','BILL MAY2005','BILL APRL
       ⇒sum(axis=1)
[29]: df['TOTAL_BILL_PAID'] =
       →df[['BILL_PAID_SEP2005','BILL_PAID_AUG2005','BILL_PAID_JULY2005','BILL_PAID_JUNE2005','BILL
       ⇒sum(axis=1)
[30]: df.head()
         ID LIMIT_BAL
                           SEX
                                              MARRIAGE AGE PAID APRL2005 \
[30]:
                                 EDUCATION
          1
                 20000 FEMALE UNIVERSITY
                                               MARRIED
                                                         24
                                                                         2
      0
      1
          2
                120000 FEMALE UNIVERSITY UNMARRIED
                                                         26
                                                                 PAY DULY
      2
                 90000 FEMALE UNIVERSITY
                                                         34
                                                                 PAY DULY
          3
                                            UNMARRIED
```

```
3 4
                50000 FEMALE UNIVERSITY
                                            MARRIED
                                                        37
                                                                PAY DULY
      4 5
                50000
                          MALE UNIVERSITY
                                                        57
                                                                PAY DULY
                                              MARRIED
        PAID_MAY2005 PAID_JUNE2005 PAID_JULY2005 ... BILL_MAY2005 BILL_APRL2005 \
      0
                   2
                        PREPAID 1
                                      PREPAID 1
                                                 ...
                   2
                         PAY DULY
                                       PAY DULY
                                                            3455
                                                                          3261
      1
                                       PAY DULY ...
      2
           PAY DULY
                         PAY DULY
                                                           14948
                                                                         15549
           PAY DULY
                        PAY DULY
                                      PAY DULY ...
      3
                                                           28959
                                                                         29547
           PAY DULY
                       PREPAID 1
                                      PAY DULY ...
                                                           19146
                                                                         19131
        BILL PAID SEP2005 BILL PAID AUG2005 BILL PAID JULY2005 \
     0
                         0
                                          689
                         0
                                         1000
                                                             1000
      1
      2
                      1518
                                         1500
                                                             1000
      3
                      2000
                                         2019
                                                             1200
      4
                      2000
                                        36681
                                                            10000
        BILL_PAID_JUNE2005
                            BILL_PAID_MAY2005
                                                BILL_PAID_APRL2005
                                                                    TOTAL_BILL \
     0
                                             0
                                                                 0
                                                                          7704
                          0
                       1000
                                             0
                                                              2000
     1
                                                                         17077
      2
                       1000
                                          1000
                                                              5000
                                                                        101653
      3
                       1100
                                          1069
                                                              1000
                                                                        231334
      4
                       9000
                                           689
                                                              679
                                                                        109339
        TOTAL BILL PAID
     0
                     689
                    5000
      1
      2
                   11018
      3
                   8388
                   59049
      [5 rows x 26 columns]
[31]: df.loc[:, 'DIFF_PERCENTAGE'] = (df['TOTAL_BILL_PAID'] / df['TOTAL_BILL'] * 100)
[32]: df.head(10)
                                 EDUCATION
[32]:
        ID LIMIT_BAL
                          SEX
                                             MARRIAGE AGE PAID APRL2005 \
      0
          1
                20000 FEMALE
                                 UNIVERSITY
                                              MARRIED
                                                         24
                                                                        2
          2
                120000 FEMALE
                                                         26
      1
                                 UNIVERSITY UNMARRIED
                                                                 PAY DULY
      2
                                                         34
                                                                 PAY DULY
          3
                90000 FEMALE
                                 UNIVERSITY UNMARRIED
      3
          4
                50000 FEMALE
                                 UNIVERSITY
                                              MARRIED
                                                         37
                                                                 PAY DULY
                50000
                         MALE
                                                         57
                                                                 PAY DULY
      4
          5
                                 UNIVERSITY
                                              MARRIED
      5
          6
                50000
                         MALE
                                   GRADUATE UNMARRIED
                                                         37
                                                                 PAY DULY
                                 GRADUATE UNMARRIED
      6
         7
                500000
                        MALE
                                                         29
                                                                PAY DULY
                100000 FEMALE
                                                         23
                                                                 PAY DULY
     7
         8
                                UNIVERSITY UNMARRIED
      8
                140000 FEMALE HIGH SCHOOL
                                               MARRIED
                                                         28
                                                                 PAY DULY
```

			III DOILOGE O				
	PAID_MAY2005 PAI	D HINESOOS	OCA IIII ALVO	15 B	TII ADRIANA	\	
0	2	PREPAID 1	PREPAID		0	`	
1	2	PAY DULY	PAY DU		3261		
2	PAY DULY	PAY DULY	PAY DU		15549		
3	PAY DULY	PAY DULY	PAY DU		29547		
4	PAY DULY	PREPAID 1	PAY DU		19131		
5	PAY DULY	PAY DULY	PAY DU		20024		
6	PAY DULY	PAY DULY	PAY DU		473944		
7	PREPAID 1	PREPAID 1	PAY DU		567		
8	PAY DULY	2	PAY DU		3719		
9		PREPAID 2	PREPAID		13912		
Ü	110211112	110211112	1 1021 1112		10012		
	BILL_PAID_SEP200	)5 BTI.I. PAT	D AUG2005	BTI.I. PA	ID JULY2005	\	
0		0	689	,	0	•	
1	0		1000		1000		
2	1518		1500		1000		
3	2000		2019		1200		
4	2000		36681		10000		
5	2500		1815		657		
6	5500		40000		38000		
7	38		601		0		
8	332		0		432		
9		0	0		0		
			•				
	BILL_PAID_JUNE2			BILL_		TOTAL_BILL	\
0				BILL_		<del>-</del>	\
0	BILL_PAID_JUNE2	005 BILL_F	PAID_MAY2005	BILL_	PAID_APRL2005	7704	\
	BILL_PAID_JUNE2	005 BILL_F	PAID_MAY2005 0	BILL_	PAID_APRL2005 0	7704 17077	\
1	BILL_PAID_JUNE2	0005 BILL_F 0	PAID_MAY2005 0 0	BILL_	PAID_APRL2005 0 2000	7704 17077 101653	\
1 2	BILL_PAID_JUNE2  1 1 1	005 BILL_F 0 000	PAID_MAY2005 0 0 1000	BILL_	PAID_APRL2005 0 2000 5000	7704 17077 101653 231334	\
1 2 3	BILL_PAID_JUNE2  1 1 1 9	0005 BILL_F 0000 0000 1000	PAID_MAY2005 0 0 1000 1069	BILL_	PAID_APRL2005 0 2000 5000 1000	7704 17077 101653 231334 109339	\
1 2 3 4	BILL_PAID_JUNE2  1 1 1 9	005 BILL_F 0.000 000 100	PAID_MAY2005 0 0 1000 1069 689	BILL_	PAID_APRL2005 0 2000 5000 1000 679	7704 17077 101653 231334 109339 238114	\
1 2 3 4 5	BILL_PAID_JUNE2  1 1 1 9 1	0005 BILL_F 0000 0000 1000 0000	PAID_MAY2005 0 0 1000 1069 689 1000	BILL_	PAID_APRL2005 0 2000 5000 1000 679 800	7704 17077 101653 231334 109339 238114 2724595	\
1 2 3 4 5 6	BILL_PAID_JUNE2  1 1 1 9 1 20	0005 BILL_F 0.000 .000 .000 .100 .000 .000 .000	PAID_MAY2005 0 1000 1069 689 1000 13750	BILL_	PAID_APRL2005 0 2000 5000 1000 679 800 13770	7704 17077 101653 231334 109339 238114 2724595 13486	\
1 2 3 4 5 6 7	BILL_PAID_JUNE2  1 1 1 9 1 20	005 BILL_F 0.000 000 100 0000 0000 0239 581	PAID_MAY2005 0 1000 1069 689 1000 13750 1687	BILL_	PAID_APRL2005 0 2000 5000 1000 679 800 13770 1542	7704 17077 101653 231334 109339 238114 2724595 13486 65212	
1 2 3 4 5 6 7 8	BILL_PAID_JUNE2  1 1 1 9 1 20 1 13	005 BILL_F 0 000 000 100 000 000 0239 581 000 000	PAID_MAY2005 0 1000 1069 689 1000 13750 1687 1000 1122	BILL_	PAID_APRL2005 0 2000 5000 1000 679 800 13770 1542 1000	7704 17077 101653 231334 109339 238114 2724595 13486 65212	\
1 2 3 4 5 6 7 8 9	BILL_PAID_JUNE2  1 1 1 9 1 20 1 13 TOTAL_BILL_PAID	0005 BILL_F 0 0000 0000 1000 0000 0239 581 0000 0007	PAID_MAY2005  0  1000 1069 689 1000 13750 1687 1000 1122 CENTAGE	BILL_	PAID_APRL2005 0 2000 5000 1000 679 800 13770 1542 1000	7704 17077 101653 231334 109339 238114 2724595 13486 65212	\
1 2 3 4 5 6 7 8 9	BILL_PAID_JUNE2  1 1 1 9 1 20 1 TOTAL_BILL_PAID 689	0005 BILL_F 0 0000 0000 100 0000 0000 2339 581 0000 0007	PAID_MAY2005 0 1000 1069 689 1000 13750 1687 1000 1122 CENTAGE 943406	BILL_	PAID_APRL2005 0 2000 5000 1000 679 800 13770 1542 1000	7704 17077 101653 231334 109339 238114 2724595 13486 65212	\
1 2 3 4 5 6 7 8 9	BILL_PAID_JUNE2  1 1 1 9 1 20 1 1 3 TOTAL_BILL_PAID 689 5000	005 BILL_F 0 000 000 100 000 239 581 000 000 007	PAID_MAY2005 0 1000 1069 689 1000 13750 1687 1000 1122 CENTAGE 943406 279147	BILL_	PAID_APRL2005 0 2000 5000 1000 679 800 13770 1542 1000	7704 17077 101653 231334 109339 238114 2724595 13486 65212	\
1 2 3 4 5 6 7 8 9	BILL_PAID_JUNE2  1 1 1 9 1 20 1 13 TOTAL_BILL_PAID 689 5000 11018	0005 BILL_F 0 0000 0000 1000 0000 0239 581 0000 0007 0 DIFF_PERC 0 8.	PAID_MAY2005  0  1000 1069 689 1000 13750 1687 1000 1122 CENTAGE 943406 279147 838834	BILL_	PAID_APRL2005 0 2000 5000 1000 679 800 13770 1542 1000	7704 17077 101653 231334 109339 238114 2724595 13486 65212	\
1 2 3 4 5 6 7 8 9	BILL_PAID_JUNE2  1 1 1 9 1 20 1 1 3 TOTAL_BILL_PAID 689 5000 11018 8388	0005 BILL_F 0 0000 0000 100 0000 0000 239 581 0000 0007 0 DIFF_PERC 0 8. 0 29. 8 10.	PAID_MAY2005  0  1000 1069 689 1000 13750 1687 1000 1122 CENTAGE 943406 279147 838834 625926	BILL_	PAID_APRL2005 0 2000 5000 1000 679 800 13770 1542 1000	7704 17077 101653 231334 109339 238114 2724595 13486 65212	\
1 2 3 4 5 6 7 8 9	BILL_PAID_JUNE2  1 1 1 9 1 20 1 1 3 TOTAL_BILL_PAID 689 5000 11018 8388 59049	0005 BILL_F 0 0000 0000 0000 0000 0000 0000 000	PAID_MAY2005  0  1000 1069 689 1000 13750 1687 1000 1122  CENTAGE 943406 279147 838834 625926 005433	BILL_	PAID_APRL2005 0 2000 5000 1000 679 800 13770 1542 1000	7704 17077 101653 231334 109339 238114 2724595 13486 65212	
1 2 3 4 5 6 7 8 9 0 1 2 3 4 5	BILL_PAID_JUNE2  1 1 1 9 1 20 1 13 TOTAL_BILL_PAID 689 5000 11018 8388 59049 7772	0005 BILL_F 0 0000 0000 1000 0000 0000 239 581 0000 0007 0 DIFF_PERC 0 8. 0 29. 8 10. 8 3.	PAID_MAY2005  0  1000 1069 689 1000 13750 1687 1000 1122 CENTAGE 943406 279147 838834 625926 005433 263983	BILL_	PAID_APRL2005 0 2000 5000 1000 679 800 13770 1542 1000	7704 17077 101653 231334 109339 238114 2724595 13486 65212	
1 2 3 4 5 6 7 8 9	BILL_PAID_JUNE2  1 1 1 9 1 20 1 1 3 TOTAL_BILL_PAID 689 5000 11018 8388 59049	0005 BILL_F 0 0000 0000 100 0000 0000 239 581 0000 0007 0 DIFF_PERC 8 29 8 10 8 3 54 8 6 6	PAID_MAY2005  0  1000 1069 689 1000 13750 1687 1000 1122  CENTAGE 943406 279147 838834 625926 005433	BILL	PAID_APRL2005 0 2000 5000 1000 679 800 13770 1542 1000	7704 17077 101653 231334 109339 238114 2724595 13486 65212	

MALE HIGH SCHOOL UNMARRIED

-2

35

9 10

20000

```
9
                                52.487091
                   14129
      [10 rows x 27 columns]
[33]: def get_grade(DIFF_PERCENTAGE):
          if DIFF_PERCENTAGE >= 75:
              return 'STANDARD'
          elif DIFF_PERCENTAGE >= 50 and DIFF_PERCENTAGE < 75:</pre>
              return 'SUBSTANDARD 1'
          elif DIFF_PERCENTAGE >= 25 and DIFF_PERCENTAGE < 50:</pre>
              return 'SUBSTANDARD 2'
          else:
              return 'DEFAULTER'
      df['CLASSIFICATION'] = df['DIFF_PERCENTAGE'].apply(lambda x: get_grade(x))
[36]:
     df.head()
[36]:
         ID LIMIT_BAL
                           SEX
                                 EDUCATION
                                              MARRIAGE AGE PAID_APRL2005 \
                 20000 FEMALE UNIVERSITY
                                                         24
                                                                         2
      0
          1
                                               MARRIED
          2
                120000 FEMALE UNIVERSITY UNMARRIED
                                                         26
                                                                 PAY DULY
      1
      2
          3
                90000 FEMALE UNIVERSITY UNMARRIED
                                                         34
                                                                 PAY DULY
                 50000 FEMALE UNIVERSITY
      3
        4
                                                         37
                                                                 PAY DULY
                                               MARRIED
                          MALE UNIVERSITY
                                                                 PAY DULY
                 50000
                                               MARRIED
                                                         57
        PAID_MAY2005 PAID_JUNE2005 PAID_JULY2005 ... BILL_PAID_SEP2005
                   2
                         PREPAID 1
                                       PREPAID 1 ...
                   2
                          PAY DULY
                                        PAY DULY
                                                                      0
      1
      2
            PAY DULY
                          PAY DULY
                                        PAY DULY ...
                                                                  1518
            PAY DULY
      3
                          PAY DULY
                                         PAY DULY ...
                                                                  2000
      4
            PAY DULY
                         PREPAID 1
                                        PAY DULY ...
                                                                  2000
        BILL_PAID_AUG2005 BILL_PAID_JULY2005 BILL_PAID_JUNE2005 \
      0
                      689
                                                                 0
      1
                     1000
                                          1000
                                                              1000
      2
                     1500
                                          1000
                                                              1000
      3
                     2019
                                          1200
                                                              1100
      4
                                                              9000
                    36681
                                         10000
         BILL PAID MAY2005 BILL PAID APRL2005
                                                 TOTAL BILL TOTAL BILL PAID \
      0
                         0
                                                       7704
                                                                          689
      1
                         0
                                           2000
                                                      17077
                                                                         5000
      2
                                           5000
                                                                        11018
                      1000
                                                     101653
      3
                      1069
                                           1000
                                                     231334
                                                                        8388
      4
                       689
                                            679
                                                     109339
                                                                       59049
```

8

6761

10.367724

```
DIFF_PERCENTAGE CLASSIFICATION
0 8.943406 DEFAULTER
1 29.279147 SUBSTANDARD 2
2 10.838834 DEFAULTER
3 3.625926 DEFAULTER
4 54.005433 SUBSTANDARD 1
```

[5 rows x 28 columns]

```
[35]: df.to_excel('C:/DATA_SET/DEFAULT.xlsx', index=False)
```