DATA_ANALYSIS

April 15, 2023

1 DATA ANALYSIS PROJECT

Reading data using pandas

[1]:
```
import pandas as pd
```

[2]:
```
df = pd.read_excel('C:\DATA_SET\defaultOfCreditCardClients.xlsx')
```

[3]:
```
df.head()
```

[3]:
```
   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
1   2     120000    2          2         2   26     -1      2      0      0
2   3      90000    2          2         2   34      0      0      0      0
3   4      50000    2          2         1   37      0      0      0      0
4   5      50000    1          2         1   57     -1      0     -1      0

   …  BILL_AMT4  BILL_AMT5  BILL_AMT6  PAY_AMT1  PAY_AMT2  PAY_AMT3  \
0  …          0          0          0         0       689         0
1  …       3272       3455       3261         0      1000      1000
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         0         0                           1
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]
```

2 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	
1	2	120000	2	2	2	26	-1	2	0	0	
2	3	90000	2	2	2	34	0	0	0	0	
3	4	50000	2	2	1	37	0	0	0	0	
4	5	50000	1	2	1	57	-1	0	-1	0	

	…	BILL_AMT4	BILL_AMT5	BILL_AMT6	PAY_AMT1	PAY_AMT2	PAY_AMT3	\
0	…	0	0	0	0	689	0	
1	…	3272	3455	3261	0	1000	1000	
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	0	0	1
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]

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()
```

[11]:

	ID	LIMIT_BAL	SEX	EDUCATION	MARRIAGE	AGE	PAY_0	PAY_2	PAY_3	\
0	1	20000	FEMALE	UNIVERSITY	MARRIED	24	2	2	-1	

```
1   2     120000  FEMALE  UNIVERSITY  UNMARRIED   26  PAY DULY      2      0
2   3      90000  FEMALE  UNIVERSITY  UNMARRIED   34  PAY DULY      0      0
3   4      50000  FEMALE  UNIVERSITY    MARRIED   37  PAY DULY      0      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  \
0     -1  …          0          0          0         0       689         0
1      0  …       3272       3455       3261         0      1000      1000
2      0  …      14331      14948      15549      1518      1500      1000
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
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]
```

[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()
```

[17]:
```
   ID  LIMIT_BAL     SEX   EDUCATION   MARRIAGE  AGE     PAY_0     PAY_2  \
0   1      20000  FEMALE  UNIVERSITY    MARRIED   24         2         2
1   2     120000  FEMALE  UNIVERSITY  UNMARRIED   26  PAY DULY         2
2   3      90000  FEMALE  UNIVERSITY  UNMARRIED   34  PAY DULY  PAY DULY
3   4      50000  FEMALE  UNIVERSITY    MARRIED   37  PAY DULY  PAY DULY
4   5      50000    MALE  UNIVERSITY    MARRIED   57  PAY DULY  PAY DULY

       PAY_3      PAY_4  … BILL_AMT4 BILL_AMT5  BILL_AMT6  PAY_AMT1  \
0  PREPAID 1  PREPAID 1  …         0         0          0         0
```

```
1   PAY DULY   PAY DULY  …      3272      3455       3261         0
2   PAY DULY   PAY DULY  …     14331     14948      15549      1518
3   PAY DULY   PAY DULY  …     28314     28959      29547      2000
4  PREPAID 1   PAY DULY  …     20940     19146      19131      2000

   PAY_AMT2  PAY_AMT3  PAY_AMT4  PAY_AMT5  PAY_AMT6  \
0       689         0         0         0         0
1      1000      1000      1000         0      2000
2      1500      1000      1000      1000      5000
3      2019      1200      1100      1069      1000
4     36681     10000      9000       689       679

   DEFAULT PAYMENT NEXT MONTH
0                           1
1                           1
2                           0
3                           0
4                           0

[5 rows x 25 columns]
```

[18]:
```
df = df.rename(columns={'PAY_0': 'PAID_APRL2005', 'PAY_2':␣
 ↪'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  \
0   1      20000  FEMALE  UNIVERSITY    MARRIED   24             2
1   2     120000  FEMALE  UNIVERSITY  UNMARRIED   26      PAY DULY
2   3      90000  FEMALE  UNIVERSITY  UNMARRIED   34      PAY DULY
3   4      50000  FEMALE  UNIVERSITY    MARRIED   37      PAY DULY
4   5      50000    MALE  UNIVERSITY    MARRIED   57      PAY DULY

  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      PAY DULY  …     14331     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      29547      2000      2019      1200      1100      1069      1000
4      19131      2000     36681     10000      9000       689       679
```

```
   DEFAULT PAYMENT NEXT MONTH
0                           1
1                           1
2                           0
3                           0
4                           0

[5 rows x 25 columns]
```

[20]:
```
df = df.rename(columns={'BILL_AMT6': 'BILL_APRL2005', 'BILL_AMT5':␣
↪'BILL_MAY2005','BILL_AMT4': 'BILL_JUNE2005', 'BILL_AMT3': 'BILL_JULY2005',
                        'BILL_AMT2': 'BILL_AUG2005', 'BILL_AMT1':␣
↪'BILL_SEP2005'})
```

[21]:
```
df.head()
```

[21]:
```
   ID  LIMIT_BAL     SEX   EDUCATION   MARRIAGE  AGE PAID_APRL2005  \
0   1      20000  FEMALE  UNIVERSITY    MARRIED   24             2
1   2     120000  FEMALE  UNIVERSITY  UNMARRIED   26      PAY DULY
2   3      90000  FEMALE  UNIVERSITY  UNMARRIED   34      PAY DULY
3   4      50000  FEMALE  UNIVERSITY    MARRIED   37      PAY DULY
4   5      50000    MALE  UNIVERSITY    MARRIED   57      PAY DULY

  PAID_MAY2005 PAID_JUNE2005 PAID_JULY2005  … BILL_JUNE2005 BILL_MAY2005  \
0            2     PREPAID 1     PREPAID 1  …             0            0
1            2      PAY DULY      PAY DULY  …          3272         3455
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  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          29547      2000      2019      1200      1100      1069      1000
4          19131      2000     36681     10000      9000       689       679

   DEFAULT PAYMENT NEXT MONTH
0                           1
1                           1
2                           0
3                           0
4                           0

[5 rows x 25 columns]
```

[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	\
0	1	20000	FEMALE	UNIVERSITY	MARRIED	24	2	
1	2	120000	FEMALE	UNIVERSITY	UNMARRIED	26	PAY DULY	
2	3	90000	FEMALE	UNIVERSITY	UNMARRIED	34	PAY DULY	
3	4	50000	FEMALE	UNIVERSITY	MARRIED	37	PAY DULY	
4	5	50000	MALE	UNIVERSITY	MARRIED	57	PAY DULY	

	PAID_MAY2005	PAID_JUNE2005	PAID_JULY2005	…	BILL_JUNE2005	BILL_MAY2005	\
0	2	PREPAID 1	PREPAID 1	…	0	0	
1	2	PAY DULY	PAY DULY	…	3272	3455	
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	0	0	689	0	
1	3261	0	1000	1000	
2	15549	1518	1500	1000	
3	29547	2000	2019	1200	
4	19131	2000	36681	10000	

	BILL_PAID_JUNE2005	BILL_PAID_MAY2005	BILL_PAID_APRL2005	\
0	0	0	0	
1	1000	0	2000	
2	1000	1000	5000	
3	1100	1069	1000	
4	9000	689	679	

	DEFAULT PAYMENT NEXT MONTH
0	1
1	1
2	0
3	0
4	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	
1	2	120000	FEMALE	UNIVERSITY	UNMARRIED	26	PAY DULY	
2	3	90000	FEMALE	UNIVERSITY	UNMARRIED	34	PAY DULY	
3	4	50000	FEMALE	UNIVERSITY	MARRIED	37	PAY DULY	
4	5	50000	MALE	UNIVERSITY	MARRIED	57	PAY DULY	

	PAID_MAY2005	PAID_JUNE2005	PAID_JULY2005	…	BILL_JUNE2005	BILL_MAY2005	\
0	2	PREPAID 1	PREPAID 1	…	0	0	
1	2	PAY DULY	PAY DULY	…	3272	3455	
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	0	0	689	0	
1	3261	0	1000	1000	
2	15549	1518	1500	1000	
3	29547	2000	2019	1200	
4	19131	2000	36681	10000	

	BILL_PAID_JUNE2005	BILL_PAID_MAY2005	BILL_PAID_APRL2005	\
0	0	0	0	
1	1000	0	2000	
2	1000	1000	5000	
3	1100	1069	1000	
4	9000	689	679	

	DEFAULT_PAYMENT_NEXT_MONTH
0	1
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()
```

[27]:

	ID	LIMIT_BAL	SEX	EDUCATION	MARRIAGE	AGE	PAID_APRL2005	\
0	1	20000	FEMALE	UNIVERSITY	MARRIED	24	2	
1	2	120000	FEMALE	UNIVERSITY	UNMARRIED	26	PAY DULY	
2	3	90000	FEMALE	UNIVERSITY	UNMARRIED	34	PAY DULY	

```
3   4      50000  FEMALE  UNIVERSITY    MARRIED   37      PAY DULY
4   5      50000    MALE  UNIVERSITY    MARRIED   57      PAY DULY

  PAID_MAY2005 PAID_JUNE2005 PAID_JULY2005  … BILL_JULY2005 BILL_JUNE2005  \
0            2     PREPAID 1     PREPAID 1  …           689             0
1            2      PAY DULY      PAY DULY  …          2682          3272
2     PAY DULY      PAY DULY      PAY DULY  …         13559         14331
3     PAY DULY      PAY DULY      PAY DULY  …         49291         28314
4     PAY DULY     PREPAID 1      PAY DULY  …         35835         20940

   BILL_MAY2005  BILL_APRL2005  BILL_PAID_SEP2005  BILL_PAID_AUG2005  \
0             0              0                  0                689
1          3455           3261                  0               1000
2         14948          15549               1518               1500
3         28959          29547               2000               2019
4         19146          19131               2000              36681

   BILL_PAID_JULY2005  BILL_PAID_JUNE2005  BILL_PAID_MAY2005  \
0                   0                   0                  0
1                1000                1000                  0
2                1000                1000               1000
3                1200                1100               1069
4               10000                9000                689

   BILL_PAID_APRL2005
0                   0
1                2000
2                5000
3                1000
4                 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()
```

[30]:
```
   ID  LIMIT_BAL     SEX   EDUCATION   MARRIAGE  AGE PAID_APRL2005  \
0   1      20000  FEMALE  UNIVERSITY    MARRIED   24             2
1   2     120000  FEMALE  UNIVERSITY  UNMARRIED   26      PAY DULY
2   3      90000  FEMALE  UNIVERSITY  UNMARRIED   34      PAY DULY
```

```
3   4      50000  FEMALE  UNIVERSITY    MARRIED   37      PAY DULY
4   5      50000    MALE  UNIVERSITY    MARRIED   57      PAY DULY

  PAID_MAY2005 PAID_JUNE2005 PAID_JULY2005  … BILL_MAY2005 BILL_APRL2005  \
0            2     PREPAID 1     PREPAID 1  …            0             0
1            2      PAY DULY      PAY DULY  …         3455          3261
2     PAY DULY      PAY DULY      PAY DULY  …        14948         15549
3     PAY DULY      PAY DULY      PAY DULY  …        28959         29547
4     PAY DULY     PREPAID 1      PAY DULY  …        19146         19131

   BILL_PAID_SEP2005  BILL_PAID_AUG2005  BILL_PAID_JULY2005  \
0                  0                689                   0
1                  0               1000                1000
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                   0        7704
1                1000                  0                2000       17077
2                1000               1000                5000      101653
3                1100               1069                1000      231334
4                9000                689                 679      109339

   TOTAL_BILL_PAID
0              689
1             5000
2            11018
3             8388
4            59049

[5 rows x 26 columns]
```

[31]:
```
df.loc[:, 'DIFF_PERCENTAGE'] = (df['TOTAL_BILL_PAID'] / df['TOTAL_BILL'] * 100)
```

[32]:
```
df.head(10)
```

[32]:
```
   ID  LIMIT_BAL     SEX    EDUCATION   MARRIAGE  AGE PAID_APRL2005  \
0   1      20000  FEMALE   UNIVERSITY    MARRIED   24             2
1   2     120000  FEMALE   UNIVERSITY  UNMARRIED   26      PAY DULY
2   3      90000  FEMALE   UNIVERSITY  UNMARRIED   34      PAY DULY
3   4      50000  FEMALE   UNIVERSITY    MARRIED   37      PAY DULY
4   5      50000    MALE   UNIVERSITY    MARRIED   57      PAY DULY
5   6      50000    MALE     GRADUATE  UNMARRIED   37      PAY DULY
6   7     500000    MALE     GRADUATE  UNMARRIED   29      PAY DULY
7   8     100000  FEMALE   UNIVERSITY  UNMARRIED   23      PAY DULY
8   9     140000  FEMALE  HIGH SCHOOL    MARRIED   28      PAY DULY
```

```
9  10      20000    MALE  HIGH SCHOOL  UNMARRIED   35           -2

  PAID_MAY2005 PAID_JUNE2005 PAID_JULY2005  … BILL_APRL2005  \
0            2     PREPAID 1     PREPAID 1  …             0
1            2      PAY DULY      PAY DULY  …          3261
2     PAY DULY      PAY DULY      PAY DULY  …         15549
3     PAY DULY      PAY DULY      PAY DULY  …         29547
4     PAY DULY     PREPAID 1      PAY DULY  …         19131
5     PAY DULY      PAY DULY      PAY DULY  …         20024
6     PAY DULY      PAY DULY      PAY DULY  …        473944
7    PREPAID 1     PREPAID 1      PAY DULY  …           567
8     PAY DULY             2      PAY DULY  …          3719
9    PREPAID 2     PREPAID 2     PREPAID 2  …         13912

  BILL_PAID_SEP2005  BILL_PAID_AUG2005  BILL_PAID_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             55000              40000               38000
7               380                601                   0
8              3329                  0                 432
9                 0                  0                   0

   BILL_PAID_JUNE2005  BILL_PAID_MAY2005  BILL_PAID_APRL2005  TOTAL_BILL  \
0                   0                  0                   0        7704
1                1000                  0                2000       17077
2                1000               1000                5000      101653
3                1100               1069                1000      231334
4                9000                689                 679      109339
5                1000               1000                 800      238114
6               20239              13750               13770     2724595
7                 581               1687                1542       13486
8                1000               1000                1000       65212
9               13007               1122                   0       26919

   TOTAL_BILL_PAID  DIFF_PERCENTAGE
0              689         8.943406
1             5000        29.279147
2            11018        10.838834
3             8388         3.625926
4            59049        54.005433
5             7772         3.263983
6           180759         6.634344
7             4791        35.525730
```

```
8             6761        10.367724
9            14129        52.487091

[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:
        return 'SUBSTANDARD 1'
    elif DIFF_PERCENTAGE >= 25 and DIFF_PERCENTAGE < 50:
        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  \
0   1      20000  FEMALE  UNIVERSITY    MARRIED   24             2
1   2     120000  FEMALE  UNIVERSITY  UNMARRIED   26      PAY DULY
2   3      90000  FEMALE  UNIVERSITY  UNMARRIED   34      PAY DULY
3   4      50000  FEMALE  UNIVERSITY    MARRIED   37      PAY DULY
4   5      50000    MALE  UNIVERSITY    MARRIED   57      PAY DULY

  PAID_MAY2005 PAID_JUNE2005 PAID_JULY2005  … BILL_PAID_SEP2005  \
0            2     PREPAID 1     PREPAID 1  …                 0
1            2      PAY DULY      PAY DULY  …                 0
2     PAY DULY      PAY DULY      PAY DULY  …              1518
3     PAY DULY      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                   0
1              1000                1000                1000
2              1500                1000                1000
3              2019                1200                1100
4             36681               10000                9000

   BILL_PAID_MAY2005  BILL_PAID_APRL2005  TOTAL_BILL  TOTAL_BILL_PAID  \
0                  0                   0        7704              689
1                  0                2000       17077             5000
2               1000                5000      101653            11018
3               1069                1000      231334             8388
4                689                 679      109339            59049
```

```
   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)
```