```
Absenteeism-Data-Cleaning-Exercise
In [1]:
           import pandas as pd
In [2]:
           raw_data = pd.read_csv('Absenteeism-data.csv')
In [3]:
           raw_data
Out[3]:
                                                                                                  Body
                                                                                                                                    Absenteeism
                                             Transportation
                                                                                  Daily Work
                     Reason for
                                                               Distance
               ID
                                                                                                  Mass
                                                                                                         Education
                                                                                                                   Children
                                                                                                                             Pets
                                                                                                                                         Time in
                                      Date
                                                                         Age
                                                               to Work
                       Absence
                                                   Expense
                                                                                Load Average
                                                                                                  Index
                                                                                                                                          Hours
            0
               11
                            26
                                07/07/2015
                                                       289
                                                                     36
                                                                          33
                                                                                     239.554
                                                                                                     30
                                                                                                                 1
                                                                                                                           2
                                                                                                                                 1
                                                                                                                                              4
            1
                                14/07/2015
                                                                          50
                                                                                     239.554
                                                                                                                                 0
                                                                                                                                              0
               36
                             0
                                                        118
                                                                     13
                                                                                                     31
                                                                                                                                 0
                                                                                                                                              2
            2
                            23
                                15/07/2015
                                                       179
                                                                     51
                                                                          38
                                                                                     239.554
                                                                                                     31
                                                                                                                 1
                                                                                                                           0
                3
            3
                                16/07/2015
                                                       279
                                                                      5
                                                                          39
                                                                                     239.554
                                                                                                     24
                                                                                                                           2
                                                                                                                                 0
                                                                                                                                              4
            4
               11
                            23
                                23/07/2015
                                                       289
                                                                     36
                                                                          33
                                                                                     239.554
                                                                                                     30
                                                                                                                 1
                                                                                                                           2
                                                                                                                                 1
                                                                                                                                              2
                                                                                                                 2
          695
               17
                            10 23/05/2018
                                                       179
                                                                     22
                                                                          40
                                                                                     237.656
                                                                                                     22
                                                                                                                           2
                                                                                                                                 0
                                                                                                                                              8
                                23/05/2018
                                                       225
                                                                          28
                                                                                     237.656
                                                                                                                                 2
                                                                                                                                              3
               28
                             6
                                                                     26
                                                                                                     24
          696
          697
               18
                            10
                                24/05/2018
                                                       330
                                                                     16
                                                                          28
                                                                                     237.656
                                                                                                     25
                                                                                                                 2
                                                                                                                           0
                                                                                                                                 0
                                                                                                                                              8
          698
               25
                            23
                                24/05/2018
                                                       235
                                                                     16
                                                                          32
                                                                                     237.656
                                                                                                     25
                                                                                                                 3
                                                                                                                           0
                                                                                                                                 0
                                                                                                                                              2
          699
               15
                            28
                                31/05/2018
                                                       291
                                                                     31
                                                                          40
                                                                                     237.656
                                                                                                     25
                                                                                                                                 1
                                                                                                                                              2
         700 rows × 12 columns
In [4]:
           df = raw_data.copy()
In [5]:
           df
                                                                                                                                    Absenteeism
Out[5]:
                                                                                                  Body
                     Reason for
                                             Transportation
                                                               Distance
                                                                                  Daily Work
               ID
                                      Date
                                                                                                  Mass
                                                                                                         Education Children Pets
                                                                                                                                         Time in
                                                                         Age
                       Absence
                                                   Expense
                                                               to Work
                                                                                Load Average
                                                                                                  Index
                                                                                                                                          Hours
            0
               11
                            26 07/07/2015
                                                       289
                                                                          33
                                                                                     239.554
                                                                                                     30
                                                                                                                 1
                                                                                                                           2
                                                                                                                                 1
                                                                                                                                              4
                                                                     36
               36
                              0
                                 14/07/2015
                                                        118
                                                                     13
                                                                          50
                                                                                     239.554
                                                                                                     31
                                                                                                                                 0
                                                                                                                                              0
            2
                3
                                15/07/2015
                                                       179
                                                                          38
                                                                                     239.554
                                                                                                     31
                                                                                                                 1
                                                                                                                           0
                                                                                                                                 0
                                                                                                                                              2
                            23
                                                                     51
            3
                                                                                     239.554
                                                                                                                           2
                                                                                                                                 0
                                                                                                                                              4
                                16/07/2015
                                                       279
                                                                      5
                                                                          39
                                                                                                     24
            4
               11
                            23
                                23/07/2015
                                                       289
                                                                     36
                                                                          33
                                                                                     239.554
                                                                                                     30
                                                                                                                 1
                                                                                                                           2
                                                                                                                                 1
                                                                                                                                              2
          695
               17
                            10
                                23/05/2018
                                                        179
                                                                     22
                                                                          40
                                                                                     237.656
                                                                                                     22
                                                                                                                 2
                                                                                                                           2
                                                                                                                                 0
                                                                                                                                              8
               28
                                23/05/2018
                                                       225
                                                                          28
                                                                                     237.656
                                                                                                     24
                                                                                                                                 2
                                                                                                                                              3
          696
                              6
                                                                     26
                                                                                                     25
                                                                                                                 2
                                                                                                                           0
                                                                                                                                 0
                                                                                                                                              8
          697
               18
                            10
                                24/05/2018
                                                       330
                                                                     16
                                                                          28
                                                                                     237.656
          698
               25
                            23
                                24/05/2018
                                                       235
                                                                     16
                                                                          32
                                                                                     237.656
                                                                                                     25
                                                                                                                 3
                                                                                                                           0
                                                                                                                                 0
                                                                                                                                              2
                                                                                                                                              2
                            28 31/05/2018
                                                       291
                                                                                     237.656
                                                                                                     25
          699 15
                                                                     31
                                                                          40
                                                                                                                 1
                                                                                                                                 1
         700 rows × 12 columns
In [6]:
           df.info()
          <class 'pandas.core.frame.DataFrame'>
          RangeIndex: 700 entries, 0 to 699
          Data columns (total 12 columns):
```

#	Column	Non-Null Count	Dtype
0	ID	700 non-null	int64
1	Reason for Absence	700 non-null	int64
2	Date	700 non-null	object
3	Transportation Expense	700 non-null	int64
4	Distance to Work	700 non-null	int64
5	Age	700 non-null	int64
6	Daily Work Load Average	700 non-null	float64
7	Body Mass Index	700 non-null	int64
8	Education	700 non-null	int64

int64

700 non-null

Children

```
700 non-null
                                                                  int64
            11 Absenteeism Time in Hours 700 non-null
                                                                  int64
           dtypes: float64(1), int64(10), object(1)
           memory usage: 65.8+ KB
 In [7]:
            df.drop(['ID'], axis=1, inplace=True)
 In [8]:
            df
 Out[8]:
                                                                                                                                   Absenteeism
                  Reason for
                                                                                 Daily Work Body Mass
                                           Transportation Distance to
                                                                       Age
                                                                                                         Education Children Pets
                                    Date
                                                                                                                                        Time in
                     Absence
                                                 Expense
                                                                Work
                                                                               Load Average
                                                                                                 Index
                                                                                                                                         Hours
             0
                          26 07/07/2015
                                                     289
                                                                        33
                                                                                    239.554
                                                                                                     30
                                                                                                                          2
                                                                                                                                              4
                                                                   36
                                                                                                                                1
             1
                           0 14/07/2015
                                                      118
                                                                   13
                                                                        50
                                                                                    239.554
                                                                                                     31
                                                                                                                                0
                                                                                                                                              0
                                                                                    239.554
                                                                                                                                0
             2
                          23 15/07/2015
                                                      179
                                                                   51
                                                                        38
                                                                                                     31
                                                                                                                 1
                                                                                                                          O
                                                                                                                                              2
             3
                              16/07/2015
                                                      279
                                                                    5
                                                                        39
                                                                                    239.554
                                                                                                     24
                                                                                                                                0
                                                                                                                                              2
             4
                          23 23/07/2015
                                                     289
                                                                   36
                                                                        33
                                                                                    239.554
                                                                                                     30
                                                                                                                 1
                                                                                                                          2
                                                                                                                                1
           695
                          10 23/05/2018
                                                      179
                                                                   22
                                                                        40
                                                                                    237.656
                                                                                                     22
                                                                                                                 2
                                                                                                                          2
                                                                                                                                0
                                                                                                                                              8
           696
                           6 23/05/2018
                                                      225
                                                                   26
                                                                        28
                                                                                    237.656
                                                                                                     24
                                                                                                                                2
                                                                                                                                              3
           697
                          10 24/05/2018
                                                      330
                                                                   16
                                                                        28
                                                                                    237.656
                                                                                                     25
                                                                                                                 2
                                                                                                                          0
                                                                                                                                0
                                                                                                                                              8
                                                                                                                                0
                                                                                                                                              2
           698
                          23 24/05/2018
                                                      235
                                                                   16
                                                                        32
                                                                                    237.656
                                                                                                     25
                                                                                                                 3
                                                                                                                          0
                                                                                                                                              2
                          28 31/05/2018
                                                      291
                                                                                                     25
           699
                                                                   31
                                                                        40
                                                                                    237.656
                                                                                                                 1
                                                                                                                                1
          700 rows × 11 columns
 In [9]:
            df['Reason for Absence'].min()
 Out[9]: 0
In [10]:
            df['Reason for Absence'].max()
Out[10]: 28
In [11]:
            df['Reason for Absence'].nunique()
Out[11]: 28
In [12]:
            df['Reason for Absence'].unique()
Out[12]: array([26, 0, 23, 7, 22, 19, 1, 11, 14, 21, 10, 13, 28, 18, 25, 24, 6, 27, 17, 8, 12, 5, 9, 15, 4, 3, 2, 16], dtype=int64)
In [13]:
            df['Reason for Absence'].value_counts()
          23
                  147
Out[13]:
           28
                  110
           27
                   66
           13
           0
                   38
           19
           22
                   32
           26
                   31
           25
                   29
           11
           10
                   22
           18
           14
                   18
                   16
                   13
           12
           21
           8
           9
           24
                    3
                    3
           16
```

```
4
                          17
                           Name: Reason for Absence, dtype: int64
In [14]:
                             sorted(df['Reason for Absence'].unique())
Out[14]: [0,
                              2,
                             6,
7,
8,
9,
                              11,
                              12,
                              13,
                              15,
                              16,
                              17,
                              18,
                              19,
                              21,
                              22,
                              23,
                              24,
                              25,
                              26,
In [15]:
                             reason_columns = pd.get_dummies(df['Reason for Absence'])
In [16]:
                             reason_columns
Out[16]:
                                        0 1 2 3 4 5 6 7 8 9 ... 18 19 21 22 23 24 25
                                                                                                                                                                                                      26 27 28
                                 0 0 0 0 0 0 0 0 0 0 0 ...
                                                                                                                                  0
                                                                                                                                            0
                                                                                                                                  0
                                 1 1 0 0 0 0 0 0 0 0 0 ...
                                                                                                                                            0
                                                                                                                                                      0
                                                                                                                                                                            0
                                                                                                                                                                                                           0
                                                                                                                                                                                      0
                                 2 0 0 0 0 0 0 0 0 0 0 ...
                                3 0 0 0 0 0 0 0 1 0 0 ... 0
                                                                                                                                            0
                                                                                                                                                       0
                                                                                                                                                                 0
                                                                                                                                                                            0
                                                                                                                                                                                                          0
                                                                                                                                                                                                                               0
                                 4 0 0 0 0 0 0 0 0 0 0 ... 0 0
                                                                                                                                                      0
                                                                                                                                                                 0
                                                                                                                                                                                      0
                                                                                                                                                                                                0
                                                                                                                                                                                                          0
                                                                                                                                                                                                                     0
                                                                                                                                                                                                                               0
                                                                                                                                                                           1
                           695 0 0 0 0 0 0 0 0 0 0 ... 0 0
                                                                                                                                                       0
                                                                                                                                                                                                0
                                                                                                                                                                                                          0
                                                                                                                                                                                                                     0
                                                                                                                                                                                                                               0
                                                                                                                                                                 0
                                                                                                                                                                           0
                                                                                                                                                                                      0
                           696 0 0 0 0 0 0 1 0 0 0 ... 0
                           697 0 0 0 0 0 0 0 0 0 0 ...
                                                                                                                                 0
                                                                                                                                            0
                                                                                                                                                       0
                                                                                                                                                                 0
                                                                                                                                                                            0
                           698 0 0 0 0 0 0 0 0 0 0 ... 0 0
                                                                                                                                                     0
                                                                                                                                                                0
                           699 0 0 0 0 0 0 0 0 0 0 ... 0 0 0
                         700 rows × 28 columns
In [17]:
                             reason_columns['Check'] = reason_columns.sum(axis=1)
In [18]:
                             reason_columns
Out[18]:
                                        0 1 2 3 4 5 6 7 8 9 ... 19 21 22 23 24 25 26 27 28 Check
                                 0 0 0 0 0 0 0 0 0 0 0 ... 0
                                                                                                                                            0
                                 1 1 0 0 0 0 0 0 0 0 0 ... 0
                                                                                                                                            0
                                                                                                                                                      0
                                                                                                                                                                 0
                                                                                                                                                                           0
                                                                                                                                                                                      0
                                                                                                                                                                                                0
                                                                                                                                                                                                           0
                                 2 0 0 0 0 0 0 0 0 0 0 ...
                                 \begin{picture}(20,10) \put(0,0){$3$} \put(0,0){$0$} \put(0,
                                                                                                                                                      0
                                                                                                                                                                 0
                                                                                                                                                                           0
                                                                                                                                                                                                          0
                                                                                                                                                                                      0
                                 4 0 0 0 0 0 0 0 0 0 0 ... 0 0
                                                                                                                                                      0
                                                                                                                                                                 1
                                                                                                                                                                           0
                                                                                                                                                                                      0
                                                                                                                                                                                                0
                                                                                                                                                                                                          0
```

```
0 1 2 3 4 5 6 7 8 9 ... 19 21 22 23 24 25 26 27 28 Check
695 0 0 0 0 0 0 0 0 0 0 ...
                             0
696 0 0 0 0 0 0 1 0 0 0 ... 0 0
                                  0
                                     0
                                        0
                                                 0
697 0 0 0 0 0 0 0 0 0 0 ... 0
                               0
                                   0
                                        0
                                              0
                                                 0
   0 0 0 0 0 0 0 0 0 0 ... 0
                                0
                                   0
699 0 0 0 0 0 0 0 0 0 0 ... 0 0
                                  0
                                        0
```

700 rows × 29 columns

```
In [19]:
        reason_columns['Check'].sum(axis=0)
Out[19]: 700
In [20]:
        reason_columns['Check'].unique()
Out[20]: array([1], dtype=int64)
In [21]:
        reason_columns = reason_columns.drop(['Check'], axis=1)
In [22]:
        reason_columns
Out[22]: 0 1 2 3 4 5 6 7 8 9 ... 18 19 21 22 23 24 25 26 27 28
         0 0 0 0 0 0 0 0 0 0 0 ... 0 0
                                            0
                                               0
                                                  0
                                                     0
                                                        0
         1 1 0 0 0 0 0 0 0 0 0 ...
         2 0 0 0 0 0 0 0 0 0 0 ... 0
                                         0
                                            0
         3 0 0 0 0 0 0 1 0 0 ... 0
                                         0
                                            0
         4 0 0 0 0 0 0 0 0 0 0 ...
                                     0 0
                                            0
        695 0 0 0 0 0 0 0 0 0 0 ... 0 0
                                            0
                                               0
                                                  0
        696 0 0 0 0 0 0 1 0 0 0 ... 0 0
                                            0
                                                  0
                                                           0
        697 0 0 0 0 0 0 0 0 0 0 ... 0
                                         0
                                            0
                                                  0
        698 0 0 0 0 0 0 0 0 0 0 ... 0 0
        699 0 0 0 0 0 0 0 0 0 0 ... 0 0
       700 rows × 28 columns
In [23]:
        reason_columns = reason_columns.drop([0], axis=1)
In [24]:
        reason_columns
Out[24]: 1 2 3 4 5 6 7 8 9 10 ... 18 19 21 22 23 24 25 26 27 28
         0 0 0 0 0 0 0 0 0 0 ...
                                      0
                                          0
                                             0
                                               0
                                                   0
                                                      0
         1 0 0 0 0 0 0 0 0 0 0 ...
                                       0
                                          0
                                             0
                                                0
                                                   0
         2 0 0 0 0 0 0 0 0 0 0 ...
                                      0
                                          0
                                             0
                                                0
                                                   1
                                                      0
                                                                  0
         3 0 0 0 0 0 0 1 0 0 ... 0
                                          0
                                            0
         4 0 0 0 0 0 0 0 0 0 0 ...
                                       0
                                          0
                                             0
                                                0
                                                      0
        695 0 0 0 0 0 0 0 0 0 1 ... 0
                                          0
                                            0
                                               0
                                                   0
                                                      0
                                                                  0
        696 0 0 0 0 0 1 0 0 0 ...
                                       0
                                          0
                                             0
                                                0
                                                   0
                                                      0
```

700 rows × 27 columns

0 0 0

0 0 0

0 0 0

0 0

0 0 0 0 0 0 0 0 0 1 ...

0 0 0 0 0 0 0 0 0 0 ... 0

0 0 0 0 0 0 0 0 0 0 ... 0

```
In [25]:
           df.drop(['Reason for Absence'], axis=1, inplace=True)
In [26]:
Out[26]:
                               Transportation
                                                 Distance to
                                                                      Daily Work Load
                                                                                         Body Mass
                                                                                                                                  Absenteeism
                     Date
                                                             Age
                                                                                                    Education Children Pets
                                                      Work
                                                                                             Index
                                                                                                                                 Time in Hours
                                     Expense
                                                                             Average
             0 07/07/2015
                                         289
                                                         36
                                                              33
                                                                              239.554
                                                                                                30
                                                                                                                      2
                                                                                                                                            4
               14/07/2015
                                                              50
                                                                              239.554
                                                                                                                           0
                                                                                                                                            0
                                         118
                                                         13
                                                                                                31
               15/07/2015
                                         179
                                                         51
                                                              38
                                                                              239.554
                                                                                                31
                                                                                                                      0
                                                                                                                           0
                                                                                                                                            2
               16/07/2015
                                         279
                                                          5
                                                              39
                                                                              239.554
                                                                                                24
                                                                                                                     2
                                                                                                                           0
                                                                                                                                            4
               23/07/2015
                                         289
                                                         36
                                                              33
                                                                              239.554
                                                                                                30
                                                                                                            1
                                                                                                                     2
                                                                                                                           1
                                                                                                                                            2
           695
               23/05/2018
                                         179
                                                         22
                                                              40
                                                                              237.656
                                                                                                22
                                                                                                            2
                                                                                                                     2
                                                                                                                           0
                                                                                                                                            8
                23/05/2018
                                         225
                                                         26
                                                              28
                                                                              237.656
                                                                                                                      1
                                                                                                            2
                                                                              237.656
                                                                                                                     0
                                                                                                                                            8
           697
                24/05/2018
                                         330
                                                         16
                                                              28
                                                                                                25
                                                                                                                           0
                24/05/2018
                                         235
                                                                              237.656
                                                                                                            3
                                                                                                                     0
                                                                                                                           0
                                                                                                                                            2
           698
                                                         16
                                                              32
                                                                                                25
           699
               31/05/2018
                                         291
                                                         31
                                                              40
                                                                              237.656
                                                                                                25
                                                                                                            1
                                                                                                                      1
                                                                                                                                            2
          700 rows × 10 columns
In [27]:
           reason_type_1 = reason_columns.loc[:, 1:14].max(axis=1)
           reason_type_2 = reason_columns.loc[:, 15:17].max(axis=1)
           reason_type_3 = reason_columns.loc[:, 18:21].max(axis=1)
           reason_type_4 = reason_columns.loc[:, 22:].max(axis=1)
In [28]:
           reason_type_4
          0
                   1
Out[28]:
                  0
           2
                   1
           3
                  0
                  1
           695
                  0
           696
                  0
           697
                  0
           698
          699
           Length: 700, dtype: uint8
In [29]:
Out[29]:
                                                                      Daily Work Load
                                                                                         Body Mass
                                                                                                                                 Absenteeism
                               Transportation
                                                 Distance to
                     Date
                                                             Age
                                                                                                    Education Children Pets
                                                                                                                                Time in Hours
                                     Expense
                                                      Work
                                                                             Average
                                                                                             Index
             0 07/07/2015
                                         289
                                                         36
                                                              33
                                                                              239.554
                                                                                                30
                                                                                                            1
                                                                                                                      2
                                                                                                                           1
                                                                                                                                            4
                14/07/2015
                                         118
                                                         13
                                                              50
                                                                              239.554
                                                                                                31
                                                                                                                                            0
                                                                                                                      1
             2 15/07/2015
                                         179
                                                                              239.554
                                                                                                                     0
                                                                                                                           0
                                                                                                                                            2
                                                         51
                                                              38
                                                                                                31
                                                                                                            1
                16/07/2015
                                         279
                                                              39
                                                                              239.554
                                                                                                                     2
                                                                                                                           0
                                                                                                                                            4
                                                          5
                                                                                                24
             4 23/07/2015
                                         289
                                                         36
                                                              33
                                                                              239.554
                                                                                                30
                                                                                                            1
                                                                                                                     2
                                                                                                                           1
                                                                                                                                            2
           695
                23/05/2018
                                         179
                                                         22
                                                              40
                                                                              237.656
                                                                                                22
                                                                                                            2
                                                                                                                     2
                                                                                                                           0
                                                                                                                                            8
                                                                                                                                            3
           696
                23/05/2018
                                         225
                                                         26
                                                              28
                                                                              237.656
                                                                                                24
                                                                                                                     1
                                                                                                                           2
                                                                                                            2
                24/05/2018
                                         330
                                                              28
                                                                              237.656
                                                                                                25
                                                                                                                     0
                                                                                                                           0
                                                                                                                                            8
           697
                                                         16
                24/05/2018
                                         235
                                                         16
                                                              32
                                                                              237.656
                                                                                                25
                                                                                                            3
                                                                                                                     0
                                                                                                                           0
                                                                                                                                            2
                                                                                                                                            2
           699 31/05/2018
                                         291
                                                         31
                                                                              237.656
                                                                                                25
                                                                                                                     1
                                                                                                                           1
                                                              40
                                                                                                            1
          700 rows × 10 columns
In [30]:
           df = pd.concat([df, reason_type_1, reason_type_2, reason_type_3, reason_type_4], axis=1)
```

```
In [31]:
Out[31]
```

]:		Date	Transportation Expense	Distance to Work	Age	Daily Work Load Average	Body Mass Index	Education	Children	Pets	Absenteeism Time in Hours	0	1	2	3
	0	07/07/2015	289	36	33	239.554	30	1	2	1	4	0	0	0	1
	1	14/07/2015	118	13	50	239.554	31	1	1	0	0	0	0	0	0
	2	15/07/2015	179	51	38	239.554	31	1	0	0	2	0	0	0	1
	3	16/07/2015	279	5	39	239.554	24	1	2	0	4	1	0	0	0
	4	23/07/2015	289	36	33	239.554	30	1	2	1	2	0	0	0	1
6	595	23/05/2018	179	22	40	237.656	22	2	2	0	8	1	0	0	0
6	596	23/05/2018	225	26	28	237.656	24	1	1	2	3	1	0	0	0
6	597	24/05/2018	330	16	28	237.656	25	2	0	0	8	1	0	0	0
6	598	24/05/2018	235	16	32	237.656	25	3	0	0	2	0	0	0	1
e	599	31/05/2018	291	31	40	237.656	25	1	1	1	2	0	0	0	1

700 rows × 14 columns

```
In [32]:
          df.rename(columns={0: "Reason_1", 1: "Reason_2", 2: "Reason_3", 3: "Reason_4"}, inplace=True)
In [33]:
```

Out[33]: Daily Body Absenteeism Transportation Distance Work Reason_1 Reason_2 Reason_3 **Education Children Pets** Date Mass Time in Age to Work Load **Expense** Index Hours Average 0 07/07/2015 239.554 14/07/2015 239,554 15/07/2015 239,554 16/07/2015 239.554 4 23/07/2015 239 554 ••• 23/05/2018 237.656 23/05/2018 237.656 24/05/2018 237.656 24/05/2018 237.656 31/05/2018 237.656

700 rows × 14 columns

```
In [34]:
    df.columns
'Reason_4'],
       dtype='object')
In [35]:
    'Pets', 'Absenteeism Time in Hours']
In [36]:
    df = df[col_reordered]
In [37]:
    df
```

Out[37]:

	Reas	son_1 R	eason_2	Reason_3	Reason_4	Date	Transporta Exp		ance Vork	Age	Work Load	Body Mass Eo	ducation	Chile	dren	Pets	Abs
	0	0	0	0	1	07/07/2015		289	36	33 23	9.554	30	1		2	1	
	1	0	0	0	0	14/07/2015		118	13	50 23	9.554	31	1		1	0	
	2	0	0	0	1	15/07/2015		179	51	38 23	9.554	31	1		0	0	
	3	1	0	0		16/07/2015		279	5		9.554	24	1		2	0	
	4	0	0	0	1	23/07/2015		289	36		9.554	30	1		2	1	
6						23/05/2018		179	22	40 23	7.656	22	2				
	96	1	0	0		23/05/2018		225	26		7.656	24	1		1	2	
6	97	1	0	0	0	24/05/2018		330	16	28 23	7.656	25	2		0	0	
6	98	0	0	0	1	24/05/2018		235	16	32 23	7.656	25	3		0	0	
6	i99	0	0	0	1	31/05/2018		291	31	40 23	7.656	25	1		1	1	
70	00 rows	× 14 col	umns														
4																	•
	df['Dat	e'] = p	d.to_dat	tetime(df	['Date'],	format =	'%d/%m/%Y')									
	df																
	Reas	son_1 R	eason_2	Reason_3	Reason_4	Date Trai	nsportation Expense	Distance to Work	Age	Daily Work Load Average	Body Mass Index	Educati	ion Chil	dren	Pets	Abse	entee Tim He
	0	0	0	0	1	2015- 07-07	289	36	33	239.554	30		1	2	1		
	1	0	0	0	0	2015- 07-14	118	13	50	239.554	31		1	1	0		
	2	0	0	0	1	2015- 07-15 2015-	179	51	38	239.554	31		1	0	0		
	3	0	0	0	0	07-16 2015-	279	36	39	239.554	30		1	2	0		
6	95	1	0	0	0	2018- 05-23	179	22	40	237.656	22		2	2	0		
	96	1	0	0	0	2018- 05-23	225	26	28	237.656	24		1	1	2		
6						2018-					2.5				0		
6	597	1	0	0	0	2018- 05-24 2018-	330	16	28	237.656	25		2	0	0		
6	698	0	0	0	1	05-24 2018- 05-24 2018-	235	16	32	237.656	25		3	0	0		
6		0	0			05-24 2018- 05-24					25				0		
6	598 599 00 rows :	0	0	0	1	05-24 2018- 05-24 2018-	235	16	32	237.656	25		3	0			•
66 66 70	598 599 00 rows :	0 0 × 14 col	0 0 umns	0	1	05-24 2018- 05-24 2018-	235	16	32	237.656	25		3	0			>
66 66 70	598 599 00 rows :	0 0 × 14 col	0 0 umns	0	1	05-24 2018- 05-24 2018-	235	16	32	237.656	25		3	0			>

	Reason_1	Reason_2	Reason_3	Reason_4	Date	Transportation Expense		Age	Daily Work Load Average	Body Mass Index	Education	Children	Pets	Absentee Tim He
0	0	0	0	1	2015- 07-07	289	36	33	239.554	30	1	2	1	
1	0	0	0	0	2015- 07-14	118	13	50	239.554	31	1	1	0	
2	0	0	0	1	2015- 07-15	179	51	38	239.554	31	1	0	0	
3	1	0	0	0	2015- 07-16	279	5	39	239.554	24	1	2	0	
4	0	0	0	1	2015- 07-23	289	36	33	239.554	30	1	2	1	
695	1	0	0	0	2018- 05-23	179	22	40	237.656	22	2	2	0	
696	1	0	0	0	2018- 05-23	225	26	28	237.656	24	1	1	2	
697	1	0	0	0	2018- 05-24	330	16	28	237.656	25	2	0	0	
698	0	0	0	1	2018- 05-24	235	16	32	237.656	25	3	0	0	
699	0	0	0	1	2018- 05-31	291	31	40	237.656	25	1	1	1	

700 rows × 15 columns

In [44]: df['Day of Week'] = df['Date'].dt.day_name()
In [45]: df

Out[45]:

]:		Reason_1	Reason_2	Reason_3	Reason_4	Date	Transportation Expense	Distance to Work	Age	Daily Work Load Average	Body Mass Index	Education	Children	Pets	Absented Tim He
	0	0	0	0	1	2015- 07-07	289	36	33	239.554	30	1	2	1	
	1	0	0	0	0	2015- 07-14	118	13	50	239.554	31	1	1	0	
	2	0	0	0	1	2015- 07-15	179	51	38	239.554	31	1	0	0	
	3	1	0	0	0	2015- 07-16	279	5	39	239.554	24	1	2	0	
	4	0	0	0	1	2015- 07-23	289	36	33	239.554	30	1	2	1	
	695	1	0	0	0	2018- 05-23	179	22	40	237.656	22	2	2	0	
	696	1	0	0	0	2018- 05-23	225	26	28	237.656	24	1	1	2	
	697	1	0	0	0	2018- 05-24	330	16	28	237.656	25	2	0	0	
	698	0	0	0	1	2018- 05-24	235	16	32	237.656	25	3	0	0	
	699	0	0	0	1	2018- 05-31	291	31	40	237.656	25	1	1	1	

700 rows × 16 columns

```
In [46]:
df['Education'].unique()
```

```
Out[46]: array([1, 3, 2, 4], dtype=int64)
In [47]:
           df['Education'].value counts()
Out[47]: 1
                583
                 40
          Name: Education, dtype: int64
In [48]:
           df['Education'] = df['Education'].map({1:0, 2:1, 3:1, 4:1})
In [49]:
           df
                                                                                               Daily
Out[49]:
                                                                                                      Body
                                                                                                                                      Absentee
                                                              Transportation Distance
                                                                                               Work
                Reason_1 Reason_2 Reason_3 Reason_4
                                                        Date
                                                                                                      Mass
                                                                                                            Education Children Pets
                                                                                                                                           Tim
                                                                                       Age
                                                                    Expense
                                                                              to Work
                                                                                               Load
                                                                                                     Index
                                                                                                                                           He
                                                                                             Average
                                                        2015-
             0
                       0
                                 0
                                           0
                                                                         289
                                                                                        33
                                                                                             239.554
                                                                                                        30
                                                                                                                    0
                                                                                                                             2
                                                     1
                                                                                   36
                                                                                                                                   1
                                                        07-07
                                                        2015-
                                           0
                                                                         118
                                                                                   13
                                                                                        50
                                                                                             239.554
                                                                                                                             1
                                                        07-14
                                                        2015-
             2
                       0
                                 0
                                           0
                                                                         179
                                                                                             239.554
                                                                                                                    0
                                                                                                                             0
                                                                                                                                   0
                                                                                   51
                                                                                        38
                                                                                                        31
                                                        07-15
                                                        2015-
                                           0
            3
                                 0
                                                     0
                                                                         279
                                                                                    5
                                                                                        39
                                                                                             239.554
                                                                                                        24
                                                                                                                    0
                                                                                                                             2
                                                                                                                                   0
                                                        07-16
                                                        2015-
             4
                       0
                                 0
                                           0
                                                                         289
                                                                                   36
                                                                                             239.554
                                                                                                         30
                                                                                                                    0
                                                                                                                             2
                                                                                                                                   1
                                                                                        33
                                                        07-23
                                                        2018-
          695
                                 0
                                           0
                                                     0
                                                                         179
                                                                                   22
                                                                                        40
                                                                                             237.656
                                                                                                        22
                                                                                                                             2
                                                                                                                                  0
                                                        05-23
                                                        2018-
          696
                                 0
                                           0
                                                                         225
                                                                                   26
                                                                                        28
                                                                                             237.656
                                                                                                        24
                                                                                                                    0
                                                                                                                                   2
                                                        05-23
                                                        2018-
          697
                                 0
                                           0
                                                     0
                                                                         330
                                                                                             237.656
                                                                                                        25
                                                                                                                             0
                                                                                                                                   0
                                                                                   16
                                                                                        28
                                                        05-24
                                                        2018-
                                           0
          698
                       0
                                 0
                                                                         235
                                                                                             237.656
                                                                                                        25
                                                                                                                             0
                                                                                                                                   0
                                                                                   16
                                                                                        32
                                                        05-24
                                                        2018-
                       0
                                 0
                                           0
                                                                         291
          699
                                                                                   31
                                                                                        40
                                                                                             237.656
                                                                                                        25
                                                                                                                    0
                                                                                                                                   1
                                                        05-31
          700 rows × 16 columns
          4
```

In []: