autoencoders

July 16, 2024

1 IMPORTING DEPENDENCIES

```python
from sklearn.preprocessing import StandardScaler
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib
from sklearn.preprocessing import StandardScaler
from tensorflow.keras.callbacks import ReduceLROnPlateau
```

```python
df = pd.read_csv("/content/HealthcareProviders.csv")
```

```python
df_org = df
```

2 BASIC EXPLORATION OF THE DATASET

```python
df.describe(include='all')
```

```
               index  National Provider Identifier  \
count   1.000000e+05                  1.000000e+05
unique           NaN                           NaN
top              NaN                           NaN
freq             NaN                           NaN
mean    4.907646e+06                  1.498227e+09
std     2.839633e+06                  2.874125e+08
min     2.090000e+02                  1.003001e+09
25%     2.458791e+06                  1.245669e+09
50%     4.901266e+06                  1.497847e+09
75%     7.349450e+06                  1.740374e+09
max     9.847440e+06                  1.993000e+09


        Last Name/Organization Name of the Provider First Name of the Provider  \
count                                        100000                      95745
unique                                        42820                      13022
top                                           PATEL                    MICHAEL
```

```
freq                                               557                    2350
mean                                               NaN                     NaN
std                                                NaN                     NaN
min                                                NaN                     NaN
25%                                                NaN                     NaN
50%                                                NaN                     NaN
75%                                                NaN                     NaN
max                                                NaN                     NaN

        Middle Initial of the Provider Credentials of the Provider  \
count                            70669                         92791
unique                              29                          1854
top                                  A                            MD
freq                              8152                         32874
mean                               NaN                           NaN
std                                NaN                           NaN
min                                NaN                           NaN
25%                                NaN                           NaN
50%                                NaN                           NaN
75%                                NaN                           NaN
max                                NaN                           NaN

        Gender of the Provider Entity Type of the Provider  \
count                    95746                        100000
unique                       2                             2
top                          M                             I
freq                     66641                         95746
mean                       NaN                           NaN
std                        NaN                           NaN
min                        NaN                           NaN
25%                        NaN                           NaN
50%                        NaN                           NaN
75%                        NaN                           NaN
max                        NaN                           NaN

        Street Address 1 of the Provider Street Address 2 of the Provider  … \
count                            100000                             40637  …
unique                            51928                             10024  …
top                        200 1ST ST SW                         SUITE 200  …
freq                                244                              1624  …
mean                                NaN                               NaN  …
std                                 NaN                               NaN  …
min                                 NaN                               NaN  …
25%                                 NaN                               NaN  …
50%                                 NaN                               NaN  …
75%                                 NaN                               NaN  …
max                                 NaN                               NaN  …
```

```
          HCPCS Code                                          HCPCS Description  \
count         100000                                                    100000
unique          2631                                                      2455
top            99213   Established patient office or other outpatient…
freq            4578                                                      4578
mean             NaN                                                       NaN
std              NaN                                                       NaN
min              NaN                                                       NaN
25%              NaN                                                       NaN
50%              NaN                                                       NaN
75%              NaN                                                       NaN
max              NaN                                                       NaN


          HCPCS Drug Indicator Number of Services  \
count                   100000            100000
unique                       2              2748
top                          N                13
freq                     93802              3018
mean                       NaN               NaN
std                        NaN               NaN
min                        NaN               NaN
25%                        NaN               NaN
50%                        NaN               NaN
75%                        NaN               NaN
max                        NaN               NaN


          Number of Medicare Beneficiaries  \
count                             100000
unique                              1274
top                                   11
freq                                4791
mean                                 NaN
std                                  NaN
min                                  NaN
25%                                  NaN
50%                                  NaN
75%                                  NaN
max                                  NaN


          Number of Distinct Medicare Beneficiary/Per Day Services  \
count                                              100000
unique                                               1979
top                                                    12
freq                                                 3210
mean                                                  NaN
std                                                   NaN
```

```
min                                        NaN
25%                                        NaN
50%                                        NaN
75%                                        NaN
max                                        NaN

        Average Medicare Allowed Amount Average Submitted Charge Amount  \
count                            100000                          100000
unique                            49629                           38088
top                                   3                             150
freq                               1017                             970
mean                                NaN                             NaN
std                                 NaN                             NaN
min                                 NaN                             NaN
25%                                 NaN                             NaN
50%                                 NaN                             NaN
75%                                 NaN                             NaN
max                                 NaN                             NaN

        Average Medicare Payment Amount Average Medicare Standardized Amount
count                            100000                               100000
unique                            83367                                76237
top                                2.94                                25.32
freq                                623                                 1630
mean                                NaN                                  NaN
std                                 NaN                                  NaN
min                                 NaN                                  NaN
25%                                 NaN                                  NaN
50%                                 NaN                                  NaN
75%                                 NaN                                  NaN
max                                 NaN                                  NaN

[11 rows x 27 columns]
```

** ADDING A NEW COLUMN "MONEY DIFFERENCE" IN THE DATASET WHICH CALCULATES THE DIFFERENCE BETWEEN "AVERAGE SUBMITTED CHARGE AMOUNT" COLUMN AND THE "AVERAGE MEDICARE PAYMENT AMOUNT" COLUMN**

```python
df["Average Submitted Charge Amount"] = df["Average Submitted Charge Amount"].replace(',', '', regex=True)
```

```python
df["Average Medicare Payment Amount"] = df["Average Medicare Payment Amount"].replace(',', '', regex=True)
```

```python
df["Money difference"] = df["Average Submitted Charge Amount"].astype(float) - df["Average Medicare Payment Amount"].astype(float)
```

3 Basic exploration of the dataset with the new column

```
[ ]: df.nunique()
```

```
[ ]: index                                                       100000
      National Provider Identifier                                 89508
      Last Name/Organization Name of the Provider                  42820
      First Name of the Provider                                   13022
      Middle Initial of the Provider                                  29
      Credentials of the Provider                                   1854
      Gender of the Provider                                           2
      Entity Type of the Provider                                     2
      Street Address 1 of the Provider                             51928
      Street Address 2 of the Provider                             10024
      City of the Provider                                          5846
      Zip Code of the Provider                                     47827
      State Code of the Provider                                      58
      Country Code of the Provider                                    4
      Provider Type                                                  90
      Medicare Participation Indicator                                2
      Place of Service                                                2
      HCPCS Code                                                   2631
      HCPCS Description                                             2455
      HCPCS Drug Indicator                                            2
      Number of Services                                           2748
      Number of Medicare Beneficiaries                             1274
      Number of Distinct Medicare Beneficiary/Per Day Services     1979
      Average Medicare Allowed Amount                             49629
      Average Submitted Charge Amount                             38088
      Average Medicare Payment Amount                             83367
      Average Medicare Standardized Amount                        76237
      Money difference                                            92772
      dtype: int64
```

```
[ ]: (df.isnull().sum()/(len(df)))*100
```

```
[ ]: index                                            0.000
      National Provider Identifier                     0.000
      Last Name/Organization Name of the Provider      0.000
      First Name of the Provider                       4.255
      Middle Initial of the Provider                  29.331
      Credentials of the Provider                      7.209
      Gender of the Provider                           4.254
      Entity Type of the Provider                      0.000
      Street Address 1 of the Provider                 0.000
      Street Address 2 of the Provider                59.363
      City of the Provider                             0.000
```

```
Zip Code of the Provider                                    0.000
State Code of the Provider                                  0.000
Country Code of the Provider                                0.000
Provider Type                                               0.000
Medicare Participation Indicator                            0.000
Place of Service                                            0.000
HCPCS Code                                                  0.000
HCPCS Description                                           0.000
HCPCS Drug Indicator                                        0.000
Number of Services                                          0.000
Number of Medicare Beneficiaries                            0.000
Number of Distinct Medicare Beneficiary/Per Day Services    0.000
Average Medicare Allowed Amount                             0.000
Average Submitted Charge Amount                             0.000
Average Medicare Payment Amount                             0.000
Average Medicare Standardized Amount                        0.000
Money difference                                            0.000
dtype: float64
```

[ ]: `df.describe(include='all').T`

```
[ ]:                                                    count  unique  \
     index                                           100000.0    NaN
     National Provider Identifier                    100000.0    NaN
     Last Name/Organization Name of the Provider       100000  42820
     First Name of the Provider                         95745  13022
     Middle Initial of the Provider                     70669     29
     Credentials of the Provider                        92791   1854
     Gender of the Provider                             95746      2
     Entity Type of the Provider                       100000      2
     Street Address 1 of the Provider                  100000  51928
     Street Address 2 of the Provider                   40637  10024
     City of the Provider                              100000   5846
     Zip Code of the Provider                        100000.0    NaN
     State Code of the Provider                        100000     58
     Country Code of the Provider                      100000      4
     Provider Type                                     100000     90
     Medicare Participation Indicator                  100000      2
     Place of Service                                  100000      2
     HCPCS Code                                        100000   2631
     HCPCS Description                                 100000   2455
     HCPCS Drug Indicator                              100000      2
     Number of Services                               100000   2748
     Number of Medicare Beneficiaries                 100000   1274
     Number of Distinct Medicare Beneficiary/Per Day…  100000   1979
     Average Medicare Allowed Amount                   100000  49629
     Average Submitted Charge Amount                   100000  38088
```

```
Average Medicare Payment Amount                          100000  83367
Average Medicare Standardized Amount                     100000  76237
Money difference                                         100000.0    NaN

                          top  \
index
NaN
National Provider Identifier
NaN
Last Name/Organization Name of the Provider
PATEL
First Name of the Provider
MICHAEL
Middle Initial of the Provider
A
Credentials of the Provider
MD
Gender of the Provider
M
Entity Type of the Provider
I
Street Address 1 of the Provider
200 1ST ST SW
Street Address 2 of the Provider
SUITE 200
City of the Provider
NEW YORK
Zip Code of the Provider
NaN
State Code of the Provider
CA
Country Code of the Provider
US
Provider Type
Diagnostic Radiology
Medicare Participation Indicator
Y
Place of Service
O
HCPCS Code
99213
HCPCS Description                             Established patient office
or other outpatient…
HCPCS Drug Indicator
N
Number of Services
13
```

Number of Medicare Beneficiaries
11
Number of Distinct Medicare Beneficiary/Per Day…
12
Average Medicare Allowed Amount
3
Average Submitted Charge Amount
150
Average Medicare Payment Amount
2.94
Average Medicare Standardized Amount
25.32
Money difference
NaN

	freq	mean
index	NaN	4907645.74603
National Provider Identifier	NaN	1498226858.04662
Last Name/Organization Name of the Provider	557	NaN
First Name of the Provider	2350	NaN
Middle Initial of the Provider	8152	NaN
Credentials of the Provider	32874	NaN
Gender of the Provider	66641	NaN
Entity Type of the Provider	95746	NaN
Street Address 1 of the Provider	244	NaN
Street Address 2 of the Provider	1624	NaN
City of the Provider	1061	NaN
Zip Code of the Provider	NaN	416381950.78367
State Code of the Provider	7775	NaN
Country Code of the Provider	99994	NaN
Provider Type	12537	NaN
Medicare Participation Indicator	99969	NaN
Place of Service	61616	NaN
HCPCS Code	4578	NaN
HCPCS Description	4578	NaN
HCPCS Drug Indicator	93802	NaN
Number of Services	3018	NaN
Number of Medicare Beneficiaries	4791	NaN
Number of Distinct Medicare Beneficiary/Per Day…	3210	NaN
Average Medicare Allowed Amount	1017	NaN
Average Submitted Charge Amount	970	NaN
Average Medicare Payment Amount	623	NaN
Average Medicare Standardized Amount	1630	NaN
Money difference	NaN	277.191655

	std
index	2839632.695465

```
National Provider Identifier                                287412506.095332
Last Name/Organization Name of the Provider                            NaN
First Name of the Provider                                             NaN
Middle Initial of the Provider                                         NaN
Credentials of the Provider                                            NaN
Gender of the Provider                                                 NaN
Entity Type of the Provider                                            NaN
Street Address 1 of the Provider                                       NaN
Street Address 2 of the Provider                                       NaN
City of the Provider                                                   NaN
Zip Code of the Provider                                    308256603.986241
State Code of the Provider                                             NaN
Country Code of the Provider                                           NaN
Provider Type                                                          NaN
Medicare Participation Indicator                                      NaN
Place of Service                                                       NaN
HCPCS Code                                                            NaN
HCPCS Description                                                     NaN
HCPCS Drug Indicator                                                  NaN
Number of Services                                                   NaN
Number of Medicare Beneficiaries                                     NaN
Number of Distinct Medicare Beneficiary/Per Day…            NaN
Average Medicare Allowed Amount                                      NaN
Average Submitted Charge Amount                                      NaN
Average Medicare Payment Amount                                      NaN
Average Medicare Standardized Amount                                 NaN
Money difference                                            924.898491

                                                            min  \
index                                                       209.0
National Provider Identifier                                1003001298.0
Last Name/Organization Name of the Provider                           NaN
First Name of the Provider                                            NaN
Middle Initial of the Provider                                        NaN
Credentials of the Provider                                           NaN
Gender of the Provider                                                NaN
Entity Type of the Provider                                           NaN
Street Address 1 of the Provider                                      NaN
Street Address 2 of the Provider                                      NaN
City of the Provider                                                  NaN
Zip Code of the Provider                                     601.0
State Code of the Provider                                            NaN
Country Code of the Provider                                          NaN
Provider Type                                                         NaN
Medicare Participation Indicator                                     NaN
Place of Service                                                      NaN
HCPCS Code                                                           NaN
```

```
HCPCS Description                                       NaN
HCPCS Drug Indicator                                    NaN
Number of Services                                      NaN
Number of Medicare Beneficiaries                        NaN
Number of Distinct Medicare Beneficiary/Per Day…        NaN
Average Medicare Allowed Amount                         NaN
Average Submitted Charge Amount                         NaN
Average Medicare Payment Amount                         NaN
Average Medicare Standardized Amount                    NaN
Money difference                                 -16.984065

                                                      25%  \
index                                         2458790.75
National Provider Identifier                1245669407.25
Last Name/Organization Name of the Provider            NaN
First Name of the Provider                             NaN
Middle Initial of the Provider                         NaN
Credentials of the Provider                            NaN
Gender of the Provider                                 NaN
Entity Type of the Provider                            NaN
Street Address 1 of the Provider                       NaN
Street Address 2 of the Provider                       NaN
City of the Provider                                   NaN
Zip Code of the Provider                       142630001.0
State Code of the Provider                             NaN
Country Code of the Provider                           NaN
Provider Type                                          NaN
Medicare Participation Indicator                       NaN
Place of Service                                       NaN
HCPCS Code                                             NaN
HCPCS Description                                      NaN
HCPCS Drug Indicator                                   NaN
Number of Services                                     NaN
Number of Medicare Beneficiaries                       NaN
Number of Distinct Medicare Beneficiary/Per Day…       NaN
Average Medicare Allowed Amount                        NaN
Average Submitted Charge Amount                        NaN
Average Medicare Payment Amount                        NaN
Average Medicare Standardized Amount                   NaN
Money difference                                 33.911165

                                                      50%  \
index                                           4901266.0
National Provider Identifier                   1497846612.0
Last Name/Organization Name of the Provider            NaN
First Name of the Provider                             NaN
Middle Initial of the Provider                         NaN
```

```
Credentials of the Provider                             NaN
Gender of the Provider                                  NaN
Entity Type of the Provider                             NaN
Street Address 1 of the Provider                        NaN
Street Address 2 of the Provider                        NaN
City of the Provider                                    NaN
Zip Code of the Provider                         363302500.0
State Code of the Provider                              NaN
Country Code of the Provider                            NaN
Provider Type                                           NaN
Medicare Participation Indicator                        NaN
Place of Service                                        NaN
HCPCS Code                                              NaN
HCPCS Description                                       NaN
HCPCS Drug Indicator                                    NaN
Number of Services                                      NaN
Number of Medicare Beneficiaries                        NaN
Number of Distinct Medicare Beneficiary/Per Day…        NaN
Average Medicare Allowed Amount                         NaN
Average Submitted Charge Amount                         NaN
Average Medicare Payment Amount                         NaN
Average Medicare Standardized Amount                    NaN
Money difference                                   89.912401

                                                      75%  \
index                                            7349450.5
National Provider Identifier                  1740373949.25
Last Name/Organization Name of the Provider             NaN
First Name of the Provider                              NaN
Middle Initial of the Provider                          NaN
Credentials of the Provider                             NaN
Gender of the Provider                                  NaN
Entity Type of the Provider                             NaN
Street Address 1 of the Provider                        NaN
Street Address 2 of the Provider                        NaN
City of the Provider                                    NaN
Zip Code of the Provider                         681988102.0
State Code of the Provider                              NaN
Country Code of the Provider                            NaN
Provider Type                                           NaN
Medicare Participation Indicator                        NaN
Place of Service                                        NaN
HCPCS Code                                              NaN
HCPCS Description                                       NaN
HCPCS Drug Indicator                                    NaN
Number of Services                                      NaN
Number of Medicare Beneficiaries                        NaN
```

```
Number of Distinct Medicare Beneficiary/Per Day…            NaN
Average Medicare Allowed Amount                             NaN
Average Submitted Charge Amount                             NaN
Average Medicare Payment Amount                             NaN
Average Medicare Standardized Amount                        NaN
Money difference                                     211.301618

                                                           max
index                                                9847440.0
National Provider Identifier                      1992999874.0
Last Name/Organization Name of the Provider                NaN
First Name of the Provider                                 NaN
Middle Initial of the Provider                             NaN
Credentials of the Provider                                NaN
Gender of the Provider                                     NaN
Entity Type of the Provider                                NaN
Street Address 1 of the Provider                           NaN
Street Address 2 of the Provider                           NaN
City of the Provider                                       NaN
Zip Code of the Provider                           999016573.0
State Code of the Provider                                 NaN
Country Code of the Provider                               NaN
Provider Type                                              NaN
Medicare Participation Indicator                           NaN
Place of Service                                           NaN
HCPCS Code                                                 NaN
HCPCS Description                                          NaN
HCPCS Drug Indicator                                       NaN
Number of Services                                         NaN
Number of Medicare Beneficiaries                           NaN
Number of Distinct Medicare Beneficiary/Per Day…           NaN
Average Medicare Allowed Amount                            NaN
Average Submitted Charge Amount                            NaN
Average Medicare Payment Amount                            NaN
Average Medicare Standardized Amount                       NaN
Money difference                                  57038.775556
```

Dropping different irrelevant fields throughout the process

```python
df = df.drop(columns=['index', 'National Provider Identifier','Street Address 1
 of the Provider','Street Address 2 of the Provider','Country Code of the
 Provider'])
```

```python
df['Last Name/Organization Name of the Provider'] = df['First Name of the
 Provider'].astype(str) + " " + df['Last Name/Organization Name of the
 Provider']
```

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

```
[ ]:    Last Name/Organization Name of the Provider First Name of the Provider  \
    0                     SATYASREE UPADHYAYULA                    SATYASREE
    1                             WENDY JONES                         WENDY
    2                         RICHARD DUROCHER                       RICHARD
    3                           JASPER FULLARD                        JASPER
    4                         ANTHONY PERROTTI                       ANTHONY


       Middle Initial of the Provider Credentials of the Provider  \
    0                              NaN                         M.D.
    1                                P                         M.D.
    2                                W                          DPM
    3                              NaN                           MD
    4                                E                           DO


       Gender of the Provider Entity Type of the Provider City of the Provider  \
    0                        F                           I          SAINT LOUIS
    1                        F                           I         FAYETTEVILLE
    2                        M                           I          NORTH HAVEN
    3                        M                           I          KANSAS CITY
    4                        M                           I              JUPITER


       Zip Code of the Provider State Code of the Provider  \
    0             631041004.0                           MO
    1             283043815.0                           NC
    2              64732343.0                           CT
    3             641183998.0                           MO
    4             334585700.0                           FL


                 Provider Type  …  \
    0          Internal Medicine  …
    1  Obstetrics & Gynecology  …
    2                  Podiatry  …
    3          Internal Medicine  …
    4          Internal Medicine  …


                                    HCPCS Description HCPCS Drug Indicator  \
    0  Initial hospital inpatient care, typically 70 …                    N
    1  Screening mammography, bilateral (2-view study…                    N
    2  Established patient home visit, typically 25 m…                    N
    3                        Urinalysis, manual test                      N
    4  Injection beneath the skin or into muscle for …                    N


       Number of Services Number of Medicare Beneficiaries  \
    0                  27                               24
    1                 175                              175
```

```
2                            32                                  13
3                            20                                  18
4                            33                                  24

   Number of Distinct Medicare Beneficiary/Per Day Services  \
0                                                 27
1                                                175
2                                                 32
3                                                 20
4                                                 31

   Average Medicare Allowed Amount Average Submitted Charge Amount  \
0                     200.58777778                       305.21111111
1                           123.73                              548.8
2                            90.65                                155
3                              3.5                                  5
4                            26.52                                 40

   Average Medicare Payment Amount Average Medicare Standardized Amount  \
0                      157.26222222                           160.90888889
1                            118.83                           135.31525714
2                         64.4396875                             60.5959375
3                              3.43                                   3.43
4                       19.539393939                           19.057575758

   Money difference
0        147.948889
1        429.970000
2         90.560312
3          1.570000
4         20.460606

[5 rows x 23 columns]
```

```python
df=df.drop(columns=['First Name of the Provider'])
```

```python
df.head()
```

```
   Last Name/Organization Name of the Provider Middle Initial of the Provider  \
0                         SATYASREE UPADHYAYULA                            NaN
1                                   WENDY JONES                              P
2                               RICHARD DUROCHER                             W
3                                 JASPER FULLARD                           NaN
4                               ANTHONY PERROTTI                             E

   Credentials of the Provider Gender of the Provider  \
0                        M.D.                      F
```

```
1                          M.D.                        F
2                          DPM                         M
3                           MD                         M
4                           DO                         M
```

```
   Entity Type of the Provider City of the Provider  Zip Code of the Provider  \
0                           I           SAINT LOUIS                631041004.0
1                           I          FAYETTEVILLE                283043815.0
2                           I           NORTH HAVEN                 64732343.0
3                           I           KANSAS CITY                641183998.0
4                           I               JUPITER                334585700.0
```

```
   State Code of the Provider             Provider Type  \
0                          MO         Internal Medicine
1                          NC  Obstetrics & Gynecology
2                          CT                   Podiatry
3                          MO         Internal Medicine
4                          FL         Internal Medicine
```

```
   Medicare Participation Indicator  … \
0                                 Y  …
1                                 Y  …
2                                 Y  …
3                                 Y  …
4                                 Y  …
```

```
                                 HCPCS Description HCPCS Drug Indicator  \
0  Initial hospital inpatient care, typically 70 …                    N
1  Screening mammography, bilateral (2-view study…                    N
2  Established patient home visit, typically 25 m…                    N
3                           Urinalysis, manual test                   N
4  Injection beneath the skin or into muscle for …                    N
```

```
   Number of Services Number of Medicare Beneficiaries  \
0                  27                               24
1                 175                              175
2                  32                               13
3                  20                               18
4                  33                               24
```

```
   Number of Distinct Medicare Beneficiary/Per Day Services  \
0                                                 27
1                                                175
2                                                 32
3                                                 20
4                                                 31
```

```
      Average Medicare Allowed Amount Average Submitted Charge Amount  \
0                       200.58777778                       305.21111111
1                             123.73                             548.8
2                              90.65                               155
3                                3.5                                 5
4                              26.52                                40

      Average Medicare Payment Amount Average Medicare Standardized Amount  \
0                        157.26222222                            160.90888889
1                              118.83                            135.31525714
2                           64.4396875                              60.5959375
3                                3.43                                    3.43
4                         19.539393939                            19.057575758

      Money difference
0            147.948889
1            429.970000
2             90.560312
3              1.570000
4             20.460606

[5 rows x 22 columns]
```

```python
df.rename(columns = {'Last Name/Organization Name of the Provider':'Full
name'}, inplace = True)
```

```python
df.head()
```

```
                 Full name Middle Initial of the Provider  \
0  SATYASREE UPADHYAYULA                              NaN
1            WENDY JONES                                P
2        RICHARD DUROCHER                               W
3          JASPER FULLARD                             NaN
4        ANTHONY PERROTTI                               E

   Credentials of the Provider Gender of the Provider  \
0                         M.D.                      F
1                         M.D.                      F
2                          DPM                      M
3                           MD                      M
4                           DO                      M

   Entity Type of the Provider City of the Provider  Zip Code of the Provider  \
0                            I          SAINT LOUIS               631041004.0
1                            I         FAYETTEVILLE               283043815.0
2                            I          NORTH HAVEN                64732343.0
3                            I          KANSAS CITY               641183998.0
```

```
4                              I              JUPITER              334585700.0

   State Code of the Provider            Provider Type  \
0                     MO          Internal Medicine
1                     NC  Obstetrics & Gynecology
2                     CT                   Podiatry
3                     MO          Internal Medicine
4                     FL          Internal Medicine


   Medicare Participation Indicator  …  \
0                                Y  …
1                                Y  …
2                                Y  …
3                                Y  …
4                                Y  …


                              HCPCS Description HCPCS Drug Indicator  \
0  Initial hospital inpatient care, typically 70 …                 N
1  Screening mammography, bilateral (2-view study…                 N
2  Established patient home visit, typically 25 m…                 N
3                          Urinalysis, manual test                 N
4  Injection beneath the skin or into muscle for …                 N


   Number of Services Number of Medicare Beneficiaries  \
0                 27                               24
1                175                              175
2                 32                               13
3                 20                               18
4                 33                               24


   Number of Distinct Medicare Beneficiary/Per Day Services  \
0                                                27
1                                               175
2                                                32
3                                                20
4                                                31


   Average Medicare Allowed Amount Average Submitted Charge Amount  \
0                     200.58777778                     305.21111111
1                           123.73                            548.8
2                            90.65                              155
3                              3.5                                5
4                            26.52                               40


   Average Medicare Payment Amount Average Medicare Standardized Amount  \
0                     157.26222222                         160.90888889
1                           118.83                         135.31525714
```

```
2                    64.4396875                        60.5959375
3                          3.43                              3.43
4                    19.539393939                      19.057575758

   Money difference
0       147.948889
1       429.970000
2        90.560312
3         1.570000
4        20.460606

[5 rows x 22 columns]
```

[ ]: `df=df.drop(columns=['HCPCS Description'])`

[ ]: `df.head()`

[ ]:
```
              Full name Middle Initial of the Provider  \
0  SATYASREE UPADHYAYULA                            NaN
1           WENDY JONES                              P
2       RICHARD DUROCHER                             W
3         JASPER FULLARD                           NaN
4       ANTHONY PERROTTI                             E

  Credentials of the Provider Gender of the Provider  \
0                        M.D.                      F
1                        M.D.                      F
2                         DPM                      M
3                          MD                      M
4                          DO                      M

  Entity Type of the Provider City of the Provider  Zip Code of the Provider  \
0                           I          SAINT LOUIS               631041004.0
1                           I          FAYETTEVILLE             283043815.0
2                           I          NORTH HAVEN               64732343.0
3                           I          KANSAS CITY              641183998.0
4                           I              JUPITER              334585700.0

  State Code of the Provider              Provider Type  \
0                         MO           Internal Medicine
1                         NC   Obstetrics & Gynecology
2                         CT                    Podiatry
3                         MO           Internal Medicine
4                         FL           Internal Medicine

  Medicare Participation Indicator  … HCPCS Code HCPCS Drug Indicator  \
0                                Y  …      99223                     N
```

```
1                                       Y  …     G0202                       N
2                                       Y  …     99348                       N
3                                       Y  …     81002                       N
4                                       Y  …     96372                       N

   Number of Services Number of Medicare Beneficiaries  \
0                  27                                24
1                 175                               175
2                  32                                13
3                  20                                18
4                  33                                24

   Number of Distinct Medicare Beneficiary/Per Day Services  \
0                                                 27
1                                                175
2                                                 32
3                                                 20
4                                                 31

   Average Medicare Allowed Amount Average Submitted Charge Amount  \
0                       200.58777778                      305.21111111
1                             123.73                            548.8
2                              90.65                              155
3                                3.5                                5
4                              26.52                               40

   Average Medicare Payment Amount Average Medicare Standardized Amount  \
0                      157.26222222                          160.90888889
1                            118.83                           135.31525714
2                         64.4396875                            60.5959375
3                              3.43                                  3.43
4                       19.539393939                          19.057575758

   Money difference
0         147.948889
1         429.970000
2          90.560312
3           1.570000
4          20.460606

[5 rows x 21 columns]
```

```python
plt.figure(figsize=(16,5))
plt.subplot(1,2,1)
sns.histplot(df['Money difference'], kde=True)

plt.show()
```
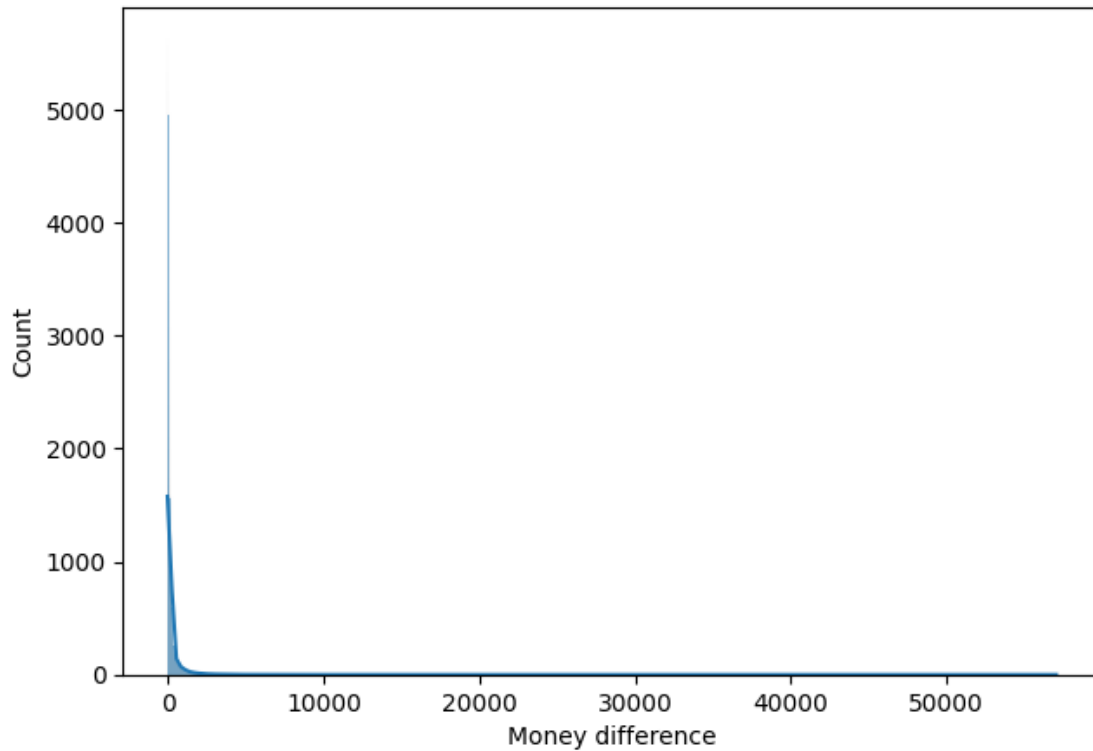
3.0.1 Scaling the numerical fields using the formula of standard scaler

```
[ ]: scaler = StandardScaler()
```

```
[ ]: df['Money difference'] = scaler.fit_transform(df[['Money difference']])

display(df)
```

```
                   Full name Middle Initial of the Provider  \
0        SATYASREE UPADHYAYULA                            NaN
1                 WENDY JONES                              P
2             RICHARD DUROCHER                             W
3               JASPER FULLARD                            NaN
4             ANTHONY PERROTTI                             E
...                      ...                              ...
99995              JOAN PAPES                             NaN
99996         MARGARET HAYNER                               S
99997           DANA VALENCIA                             NaN
99998   RAFAELA GONZALEZ-LAMOS                            NaN
99999           ELIIAN RAMEZANI                           NaN

        Credentials of the Provider Gender of the Provider  \
```

```
                                 M.D.                    F
0
1                                M.D.                    F
2                                 DPM                    M
3                                  MD                    M
4                                  DO                    M
…                                   …                    …
99995                              PT                    F
99996                            ARNP                    F
99997                            M.D.                    M
99998                             NaN                    F
99999                             NaN                    F

         Entity Type of the Provider City of the Provider  \
0                                   I          SAINT LOUIS
1                                   I         FAYETTEVILLE
2                                   I          NORTH HAVEN
3                                   I          KANSAS CITY
4                                   I              JUPITER
…                                   …                    …
99995                               I           WILMINGTON
99996                               I              REDMOND
99997                               I          SAINT LOUIS
99998                               I            LARCHMONT
99999                               I           GREAT NECK

         Zip Code of the Provider State Code of the Provider  \
0                     631041004.0                          MO
1                     283043815.0                          NC
2                      64732343.0                          CT
3                     641183998.0                          MO
4                     334585700.0                          FL
…                               …                           …
99995                 604811236.0                          IL
99996                 977561502.0                          OR
99997                 631312322.0                          MO
99998                 105383500.0                          NY
99999                     11023.0                          NY

                                 Provider Type  \
0                             Internal Medicine
1                         Obstetrics & Gynecology
2                                       Podiatry
3                             Internal Medicine
4                             Internal Medicine
…                                             …
99995   Physical Therapist in Private Practice
99996                       Nurse Practitioner
99997                               Cardiology
```

```
99998                          Internal Medicine
99999  Physical Therapist in Private Practice


       Medicare Participation Indicator  … HCPCS Code HCPCS Drug Indicator  \
0                                     Y  …      99223                    N
1                                     Y  …      G0202                    N
2                                     Y  …      99348                    N
3                                     Y  …      81002                    N
4                                     Y  …      96372                    N
…                                     …  …        …                      …
99995                                 Y  …      97162                    N
99996                                 Y  …      99213                    N
99997                                 Y  …      93320                    N
99998                                 Y  …      G0008                    N
99999                                 Y  …      97112                    N


       Number of Services Number of Medicare Beneficiaries  \
0                      27                               24
1                     175                              175
2                      32                               13
3                      20                               18
4                      33                               24
…                       …                                …
99995                  20                               20
99996                 136                              107
99997                  11                               11
99998                  12                               12
99999                 589                               76


       Number of Distinct Medicare Beneficiary/Per Day Services  \
0                                                     27
1                                                    175
2                                                     32
3                                                     20
4                                                     31
…                                                      …
99995                                                 20
99996                                                136
99997                                                 11
99998                                                 12
99999                                                587


       Average Medicare Allowed Amount Average Submitted Charge Amount  \
0                          200.58777778                     305.21111111
1                                123.73                            548.8
2                                 90.65                              155
3                                   3.5                                5
4                                 26.52                               40
```

```
...                 ...                         ...
99995             85.3725                       214
99996               61.27                144.05147059
99997               17.98                109.54545455
99998               30.54                          65
99999         38.601052632                46.867572156


      Average Medicare Payment Amount Average Medicare Standardized Amount  \
0                       157.26222222                            160.90888889
1                             118.83                            135.31525714
2                          64.4396875                              60.5959375
3                               3.43                                    3.43
4                        19.539393939                            19.057575758
...                              ...                                     ...
99995                         60.7255                                 62.2485
99996                     30.006176471                            37.040220588
99997                           14.09                                   14.62
99998                           29.93                                   25.32
99999                     29.736876061                            25.912139219


      Money difference
0            -0.139738
1             0.165185
2            -0.201787
3            -0.298004
4            -0.277579
...                ...
99995        -0.133980
99996        -0.176395
99997        -0.196494
99998        -0.261783
99999        -0.281179

[100000 rows x 21 columns]
```

```
[ ]: df['Average Medicare Payment Amount'] = scaler.fit_transform(df[['Average␣
     ↪Medicare Payment Amount']])

     display(df)
```

```
              Full name Middle Initial of the Provider  \
0      SATYASREE UPADHYAYULA                         NaN
1               WENDY JONES                           P
2          RICHARD DUROCHER                           W
3            JASPER FULLARD                         NaN
4          ANTHONY PERROTTI                           E
...                 ...                               ...
99995            JOAN PAPES                         NaN
```

```
99996        MARGARET HAYNER                          S
99997         DANA VALENCIA                         NaN
99998  RAFAELA GONZALEZ-LAMOS                       NaN
99999         ELIIAN RAMEZANI                       NaN

       Credentials of the Provider Gender of the Provider  \
0                         M.D.                      F
1                         M.D.                      F
2                          DPM                      M
3                           MD                      M
4                           DO                      M
…                            …                      …
99995                       PT                      F
99996                     ARNP                      F
99997                     M.D.                      M
99998                      NaN                      F
99999                      NaN                      F

       Entity Type of the Provider City of the Provider  \
0                            I         SAINT LOUIS
1                            I        FAYETTEVILLE
2                            I         NORTH HAVEN
3                            I         KANSAS CITY
4                            I             JUPITER
…                            …                   …
99995                        I          WILMINGTON
99996                        I             REDMOND
99997                        I         SAINT LOUIS
99998                        I           LARCHMONT
99999                        I          GREAT NECK

       Zip Code of the Provider State Code of the Provider  \
0                631041004.0                          MO
1                283043815.0                          NC
2                 64732343.0                          CT
3                641183998.0                          MO
4                334585700.0                          FL
…                          …                           …
99995            604811236.0                          IL
99996            977561502.0                          OR
99997            631312322.0                          MO
99998            105383500.0                          NY
99999             11023.0                             NY

                         Provider Type  \
0                    Internal Medicine
1               Obstetrics & Gynecology
2                             Podiatry
```

```
3                           Internal Medicine
4                           Internal Medicine
…                                   …
99995  Physical Therapist in Private Practice
99996                      Nurse Practitioner
99997                             Cardiology
99998                      Internal Medicine
99999  Physical Therapist in Private Practice


       Medicare Participation Indicator  … HCPCS Code HCPCS Drug Indicator  \
0                                   Y    …     99223                    N
1                                   Y    …     G0202                    N
2                                   Y    …     99348                    N
3                                   Y    …     81002                    N
4                                   Y    …     96372                    N
…                                 …  …        …                     …
99995                               Y    …     97162                    N
99996                               Y    …     99213                    N
99997                               Y    …     93320                    N
99998                               Y    …     G0008                    N
99999                               Y    …     97112                    N


       Number of Services Number of Medicare Beneficiaries  \
0                      27                               24
1                     175                              175
2                      32                               13
3                      20                               18
4                      33                               24
…                       …                                …
99995                  20                               20
99996                 136                              107
99997                  11                               11
99998                  12                               12
99999                 589                               76


       Number of Distinct Medicare Beneficiary/Per Day Services  \
0                                                       27
1                                                      175
2                                                       32
3                                                       20
4                                                       31
…                                                        …
99995                                                   20
99996                                                  136
99997                                                   11
99998                                                   12
99999                                                  587
```

```
       Average Medicare Allowed Amount Average Submitted Charge Amount  \
0                       200.58777778                        305.21111111
1                             123.73                               548.8
2                              90.65                                 155
3                                3.5                                   5
4                              26.52                                  40
...                              ...                                 ...
99995                        85.3725                                 214
99996                          61.27                         144.05147059
99997                          17.98                         109.54545455
99998                          30.54                                  65
99999                    38.601052632                         46.867572156

       Average Medicare Payment Amount  Average Medicare Standardized Amount  \
0                             0.400082                           160.90888889
1                             0.207649                           135.31525714
2                            -0.064687                             60.5959375
3                            -0.370166                                   3.43
4                            -0.289505                           19.057575758
...                                ...                                    ...
99995                        -0.083284                                62.2485
99996                        -0.237098                           37.040220588
99997                        -0.316791                                  14.62
99998                        -0.237479                                  25.32
99999                        -0.238446                           25.912139219

       Money difference
0             -0.139738
1              0.165185
2             -0.201787
3             -0.298004
4             -0.277579
...                 ...
99995         -0.133980
99996         -0.176395
99997         -0.196494
99998         -0.261783
99999         -0.281179

[100000 rows x 21 columns]
```

```
[ ]: df['Average Submitted Charge Amount'] = scaler.fit_transform(df[['Average␣
     ↪Submitted Charge Amount']])

     display(df)
```

```
                 Full name Middle Initial of the Provider  \
0        SATYASREE UPADHYAYULA                         NaN
```

```
1                WENDY JONES                              P
2             RICHARD DUROCHER                            W
3              JASPER FULLARD                           NaN
4             ANTHONY PERROTTI                            E
…                      …                                  …
99995            JOAN PAPES                             NaN
99996         MARGARET HAYNER                             S
99997          DANA VALENCIA                            NaN
99998  RAFAELA GONZALEZ-LAMOS                           NaN
99999          ELIIAN RAMEZANI                          NaN


       Credentials of the Provider Gender of the Provider  \
0                           M.D.                          F
1                           M.D.                          F
2                            DPM                          M
3                             MD                          M
4                             DO                          M
…                              …                          …
99995                         PT                          F
99996                       ARNP                          F
99997                       M.D.                          M
99998                        NaN                          F
99999                        NaN                          F


       Entity Type of the Provider City of the Provider  \
0                              I          SAINT LOUIS
1                              I          FAYETTEVILLE
2                              I          NORTH HAVEN
3                              I          KANSAS CITY
4                              I              JUPITER
…                              …                  …
99995                          I          WILMINGTON
99996                          I             REDMOND
99997                          I          SAINT LOUIS
99998                          I           LARCHMONT
99999                          I          GREAT NECK


       Zip Code of the Provider State Code of the Provider  \
0                   631041004.0                          MO
1                   283043815.0                          NC
2                    64732343.0                          CT
3                   641183998.0                          MO
4                   334585700.0                          FL
…                            …                            …
99995               604811236.0                          IL
99996               977561502.0                          OR
99997               631312322.0                          MO
99998               105383500.0                          NY
```

```
99999                    11023.0                        NY

                              Provider Type  \
0                        Internal Medicine
1                   Obstetrics & Gynecology
2                                  Podiatry
3                         Internal Medicine
4                         Internal Medicine
…                                        …
99995   Physical Therapist in Private Practice
99996                     Nurse Practitioner
99997                             Cardiology
99998                      Internal Medicine
99999   Physical Therapist in Private Practice


        Medicare Participation Indicator  … HCPCS Code HCPCS Drug Indicator  \
0                                     Y   …      99223                    N
1                                     Y   …      G0202                    N
2                                     Y   …      99348                    N
3                                     Y   …      81002                    N
4                                     Y   …      96372                    N
…                                    … …        …                        …
99995                                 Y   …      97162                    N
99996                                 Y   …      99213                    N
99997                                 Y   …      93320                    N
99998                                 Y   …      G0008                    N
99999                                 Y   …      97112                    N


        Number of Services Number of Medicare Beneficiaries  \
0                       27                               24
1                      175                              175
2                       32                               13
3                       20                               18
4                       33                               24
…                        …                                …
99995                   20                               20
99996                  136                              107
99997                   11                               11
99998                   12                               12
99999                  589                               76


        Number of Distinct Medicare Beneficiary/Per Day Services  \
0                                                       27
1                                                      175
2                                                       32
3                                                       20
4                                                       31
…                                                        …
```

```
99995                                               20
99996                                              136
99997                                               11
99998                                               12
99999                                              587


       Average Medicare Allowed Amount Average Submitted Charge Amount  \
0                       200.58777778                       -0.046433
1                             123.73                        0.182805
2                              90.65                       -0.187794
3                                3.5                       -0.328957
4                              26.52                       -0.296019
…                                  …                               …
99995                        85.3725                       -0.132270
99996                          61.27                       -0.198097
99997                          17.98                       -0.230571
99998                          30.54                       -0.272492
99999                   38.601052632                       -0.289556


       Average Medicare Payment Amount  Average Medicare Standardized Amount  \
0                             0.400082                           160.90888889
1                             0.207649                           135.31525714
2                            -0.064687                             60.5959375
3                            -0.370166                                   3.43
4                            -0.289505                            19.057575758
…                                    …                                      …
99995                        -0.083284                                62.2485
99996                        -0.237098                           37.040220588
99997                        -0.316791                                  14.62
99998                        -0.237479                                  25.32
99999                        -0.238446                           25.912139219


       Money difference
0              -0.139738
1               0.165185
2              -0.201787
3              -0.298004
4              -0.277579
…                      …
99995          -0.133980
99996          -0.176395
99997          -0.196494
99998          -0.261783
99999          -0.281179


[100000 rows x 21 columns]
```

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

```
[ ]:                Full name Middle Initial of the Provider  \
     0  SATYASREE UPADHYAYULA                          NaN
     1           WENDY JONES                             P
     2      RICHARD DUROCHER                             W
     3        JASPER FULLARD                           NaN
     4      ANTHONY PERROTTI                             E

        Credentials of the Provider Gender of the Provider  \
     0                        M.D.                        F
     1                        M.D.                        F
     2                         DPM                        M
     3                          MD                        M
     4                          DO                        M

        Entity Type of the Provider City of the Provider  Zip Code of the Provider  \
     0                           I          SAINT LOUIS                631041004.0
     1                           I          FAYETTEVILLE               283043815.0
     2                           I          NORTH HAVEN                 64732343.0
     3                           I          KANSAS CITY                641183998.0
     4                           I              JUPITER                334585700.0

        State Code of the Provider              Provider Type  \
     0                         MO          Internal Medicine
     1                         NC  Obstetrics & Gynecology
     2                         CT                    Podiatry
     3                         MO          Internal Medicine
     4                         FL          Internal Medicine

        Medicare Participation Indicator  … HCPCS Code HCPCS Drug Indicator  \
     0                                Y  …      99223                     N
     1                                Y  …      G0202                     N
     2                                Y  …      99348                     N
     3                                Y  …      81002                     N
     4                                Y  …      96372                     N

        Number of Services Number of Medicare Beneficiaries  \
     0                  27                               24
     1                 175                              175
     2                  32                               13
     3                  20                               18
     4                  33                               24

        Number of Distinct Medicare Beneficiary/Per Day Services  \
     0                                                 27
     1                                                175
```

```
2                                                           32
3                                                           20
4                                                           31

   Average Medicare Allowed Amount  Average Submitted Charge Amount  \
0                      200.58777778                        -0.046433
1                            123.73                         0.182805
2                             90.65                        -0.187794
3                               3.5                        -0.328957
4                             26.52                        -0.296019

   Average Medicare Payment Amount  Average Medicare Standardized Amount  \
0                         0.400082                           160.90888889
1                         0.207649                           135.31525714
2                        -0.064687                             60.5959375
3                        -0.370166                                   3.43
4                        -0.289505                           19.057575758

   Money difference
0         -0.139738
1          0.165185
2         -0.201787
3         -0.298004
4         -0.277579

[5 rows x 21 columns]
```

```python
df["Average Medicare Allowed Amount"] = df["Average Medicare Allowed Amount"].
 ↪replace('.', '', regex=True)
```

```python
df["Number of Services"] = df["Number of Services"].replace(',', '', regex=True)
```

```python
df["Number of Medicare Beneficiaries"] = df["Number of Medicare Beneficiaries"].
 ↪replace(',', '', regex=True)
```

```python
df["Number of Distinct Medicare Beneficiary/Per Day Services"] = df["Number of␣
 ↪Distinct Medicare Beneficiary/Per Day Services"].replace(',', '', regex=True)
```

```python
df.drop(columns=['Average Medicare Allowed Amount','Average Medicare␣
 ↪Standardized Amount',], inplace=True)
```

```python
df.head()
```

```
             Full name Middle Initial of the Provider  \
0  SATYASREE UPADHYAYULA                           NaN
1           WENDY JONES                             P
2       RICHARD DUROCHER                            W
```

```
3          JASPER FULLARD                              NaN
4          ANTHONY PERROTTI                            E


   Credentials of the Provider Gender of the Provider  \
0                        M.D.                        F
1                        M.D.                        F
2                         DPM                        M
3                          MD                        M
4                          DO                        M


   Entity Type of the Provider City of the Provider  Zip Code of the Provider  \
0                            I        SAINT LOUIS                 631041004.0
1                            I        FAYETTEVILLE                283043815.0
2                            I        NORTH HAVEN                  64732343.0
3                            I        KANSAS CITY                 641183998.0
4                            I        JUPITER                     334585700.0


   State Code of the Provider            Provider Type  \
0                        MO          Internal Medicine
1                        NC   Obstetrics & Gynecology
2                        CT                   Podiatry
3                        MO          Internal Medicine
4                        FL          Internal Medicine


   Medicare Participation Indicator Place of Service HCPCS Code  \
0                                Y                 F      99223
1                                Y                 O      G0202
2                                Y                 O      99348
3                                Y                 O      81002
4                                Y                 O      96372


   HCPCS Drug Indicator Number of Services Number of Medicare Beneficiaries  \
0                    N                27                              24
1                    N               175                             175
2                    N                32                              13
3                    N                20                              18
4                    N                33                              24


   Number of Distinct Medicare Beneficiary/Per Day Services  \
0                                      27
1                                     175
2                                      32
3                                      20
4                                      31


   Average Submitted Charge Amount  Average Medicare Payment Amount  \
0                      -0.046433                         0.400082
```

```
1                     0.182805                     0.207649
2                    -0.187794                    -0.064687
3                    -0.328957                    -0.370166
4                    -0.296019                    -0.289505


    Money difference
0          -0.139738
1           0.165185
2          -0.201787
3          -0.298004
4          -0.277579
```

```python
mean = np.mean(df["Money difference"])
std = np.std(df["Money difference"])
print('mean of the dataset is', mean)
print('std. deviation is', std)
```

```
mean of the dataset is -3.19211324040225e-17
std. deviation is 0.9999999999999999
```

```python
threshold = 3
outlier = []
for i in df["Money difference"]:
    z = (i-mean)/std
    if z > threshold:
        outlier.append(i)
print('outlier in dataset is', outlier)
```

```
outlier in dataset is [5.001480953472853, 7.460672545959053, 5.70618723372476,
5.472844079938339, 5.5955541655733585, 4.102689153875423, 3.125975413181884,
7.71562222857764, 7.4408227836147, 10.440100269481231, 4.9144463476674005,
14.65648528005027, 3.520356080519061, 9.839336685639676, 5.291678258126511,
12.288456819070694, 3.3148067941615165, 6.618849955479128, 5.087215423724219,
4.032193802311867, 6.61781104974683, 6.05097357153305, 5.551040769304619,
4.986930703016156, 31.079025660093187, 3.5079951528708033, 8.144963742937621,
3.6452169325732102, 5.282640091792032, 3.7323507792897472, 3.4044834638271375,
8.132835264219686, 5.934169435970785, 3.3895289080275064, 6.783230562633316,
4.36565587624966, 6.419579009759744, 5.385542319706993, 5.227544784104939,
10.285939937668005, 12.982379682037807, 4.163789831507433, 11.579963469712071,
5.543206894132582, 3.3951877693604637, 6.33314268937909, 3.8314872505974136,
5.379308938119263, 5.982010060303651, 6.12386259998153, 3.032327516620086,
9.853421941729525, 4.066363721557715, 3.718869232436207, 3.0758240989994223,
3.952362755338471, 3.228471343257186, 6.6943155789678706, 4.361554504799258,
4.980040823527655, 7.387769878918673, 4.485574500407166, 5.307126061363896,
4.987948538523319, 4.214340136085704, 23.274412827554034, 16.62027047040995,
4.131057893621092, 6.440416665853624, 3.627356975693066, 5.209639538060813,
48.799238014256694, 3.236222502884505, 3.0973498536123163, 5.991330364480495,
4.596385052490389, 3.683870812951039, 6.147252123154096, 5.568957526010002,
```

61.37091609842292, 5.826591069446645, 24.79156068183084, 6.781868244160666,
3.286072546932091, 4.355913099458368, 6.024402101668438, 9.904327591468782,
4.140542479490628, 4.25486226305052, 5.481015768681434, 10.821290983486758,
5.261054247084487, 5.506553517181567, 3.595362088428921, 7.0566192882296175,
3.2612286898642937, 15.207591756467998, 3.942020835453162, 4.875267665388059,
4.606371063168371, 3.032022245993767, 3.5063555282453045, 5.288482972239081,
6.625025473433072, 4.349071734057384, 11.325719328911216, 3.626594014887162,
5.389509961149619, 6.68300374743195, 3.9690114469478663, 4.2079689356902445,
4.89605691389529, 4.694130155354416, 15.350800013926037, 3.2399611136852373,
3.72098472259157, 11.13851106208972, 5.816454279868506, 4.138935808657011,
5.1904496109709655, 3.3112672821726674, 5.0794675077123115, 9.258271302049751,
4.275894225741565, 7.999327060856401, 3.175764538996557, 3.3068399536251145,
3.2461003930163477, 12.346431433363177, 3.907680034049186, 5.04864513946782,
5.24409315338187, 6.347318925129438, 4.482301016704186, 4.204448190319515,
3.4427824194922207, 3.315122537243484, 3.2497736072649235, 20.04754205241074,
4.4622288537619035, 3.6243606859570905, 4.087838509090286, 18.197298754770923,
5.763525532207272, 6.86295967571961, 3.8826923535928093, 6.16950324543789,
11.468576801255804, 12.708752109750089, 5.718620482860273, 9.986448545802356,
46.51604890924597, 3.3952567605412782, 7.291753774729393, 6.093118311755289,
3.185482626976711, 3.2765802866360576, 3.9414426903067983, 5.125595650562889,
3.542996852321498, 6.657399908786567, 4.166101775803666, 4.055598679105545,
3.9379053146819905, 6.016212826377757, 11.739276966954346, 12.012368924921573,
6.51380505453853, 4.158719490987006, 3.6134214906039452, 3.6181827248579586,
7.1768047748335695, 4.298006672870726, 3.2952848471815837, 7.377956962858079,
3.972370921984396, 7.173028560799377, 7.551233396890593, 7.731791768496339,
6.659306433819623, 3.9107313295450936, 5.331720389087539, 3.9468835025778715,
13.660637909829086, 7.021457896499301, 4.094091126593226, 4.011622134574951,
3.218713466895553, 4.803003354175504, 5.84740658520558, 5.569119238197061,
5.770448018332147, 3.447331881793767, 6.714336171302203, 5.534642920217079,
8.369775963929364, 5.834829852590762, 3.5311350631112437, 3.7468808657809674,
12.493567014323226, 10.147217373054154, 5.516957740363576, 9.26241736338456,
3.473369712818604, 15.223342802531468, 3.090787183096497, 3.447015595081857,
3.08706592941162, 4.121063622892661, 5.04636409552786, 24.004617615263072,
5.287721456528715, 4.850922365848368, 4.519922225200116, 5.069804764968876,
3.9277769499624005, 7.1184365912645875, 9.228128744432793, 4.046293827696323,
3.1949154660764636, 3.3246275112144827, 3.2902487789718906, 13.401773748663384,
3.4701801576643856, 3.7981725822277816, 8.423894485757652, 3.2258207453583747,
7.372776743687087, 3.1962892723536935, 13.291271108774433, 7.77017218129148,
3.481915738426095, 15.225492459017838, 25.35525755943711, 6.58993358271504,
3.8677290150989787, 3.8172340321871574, 9.097179060933147, 3.4169798747502838,
7.100872856891512, 3.603693314612252, 3.481837514867051, 3.0999044680765038,
24.796532585056713, 3.6333304087844045, 8.137710298039007, 4.331329963591836,
6.449677426355267, 15.891557573734584, 19.298013518084623, 6.016607852351124,
4.389779665151913, 4.124244393255141, 5.261697916895002, 5.71607604683962,
3.01731806177547, 3.0368704660770605, 7.113322490966469, 4.360064860841789,
3.5100044342885752, 5.046352511236534, 8.9390645351078, 3.994952002613629,
3.631263373150806, 3.701984360722436, 3.838231035948146, 9.044012741353786,
3.830148134376165, 3.8639550685682, 3.3161817073225057, 8.087045025318105,

5.153724676523058, 4.840889527883977, 5.256826956818772, 17.665852192157836,
5.4203315298796735, 8.565435686210208, 4.118551397670534, 8.618099968035116,
8.033959880349903, 6.498790268703865, 4.706054392551485, 3.8581283557924233,
3.4670816039028898, 3.068250932767786, 8.549508898576995, 3.541885682351758,
5.849576814906546, 3.5693679252360906, 44.48671066045172, 4.3367131731347355,
9.449150860953328, 4.639186585099081, 6.4395715370901545, 3.0610403388398835,
6.335841778106534, 6.54766948063523, 3.014269370470257, 4.743055700310074,
3.0322170251100142, 8.234620180508868, 6.198635301673518, 5.074667970533745,
3.279787796347654, 3.7284102024130124, 4.4893931506617815, 3.792429934405607,
3.4713478592123708, 3.4376683191941013, 3.524656832997509, 3.03647153569895,
36.6703678785862, 17.340455642890134, 4.589843761411399, 4.3174411675138,
13.623520137475083, 3.588209015446836, 3.5491165150450703, 6.187151358541109,
4.602441698880855, 3.4026763624828957, 4.36985265935016, 6.2662044444285705,
13.377478086808239, 12.76785092320456, 4.83652071342322, 3.994731744455034,
4.589768297704152, 7.506282977829831, 5.346858085889697, 11.921435154055782,
6.8660029270601095, 4.563321499041436, 3.779939251134239, 6.286018620473587,
9.965748318903934, 4.4430737680415975, 3.0730828080351573, 5.695223189779983,
6.947471444793494, 13.190300199948172, 3.434851179116206, 4.0109690750600455,
3.808975787383011, 3.270153916742962, 3.0663052461603058, 3.5039173619962742,
7.2680645701772155, 5.565166479365829, 3.440141783871601, 10.298190260517517,
3.3408306725043615, 3.039439479133862, 13.602516301103105, 4.225985796649015,
5.830759115249869, 4.620344958772768, 4.258856398618449, 7.180427487795807,
6.125351835319143, 4.052256553926739, 9.63649598799689, 3.035580292293001,
3.7344915654610533, 4.210146141391626, 5.265843727065861, 3.3452387049460146,
10.942664151087865, 3.0012838346472352, 3.737809593226307, 3.532162321377398,
3.245560391165411, 3.3790222312414677, 13.707018098777453, 7.712824125148849,
9.384949688299246, 5.888613861529406, 4.004259970045243, 3.323997018451017,
8.165160744961602, 4.469095312311229, 16.284875722824793, 3.3515424667496108,
9.706554397642565, 4.177795571284504, 7.1706382564852005, 3.734635448910576,
3.2034252262580463, 10.39209037619155, 5.5006409884512815, 5.145856683144455,
12.9021959395123, 3.2453998858843973, 5.311707397830746, 4.773165286868304,
6.780846789782063, 3.626068314320377, 15.936226710359238, 20.166099477791644,
7.689189978546376, 26.90324721287978, 4.846861457635256, 4.739720417505542,
21.045739887564178, 3.206621362707933, 12.127580951115577, 4.839322898942007,
7.858867493200748, 4.09026350195755, 17.185711031172595, 6.002140400916832,
3.4715332408828004, 5.884262551455458, 8.921442974305208, 3.652905988231564,
6.162434792867734, 5.76646525177679, 3.698337079308817, 7.592707224619772,
3.5528638068310356, 3.1401206654853433, 9.47223514509239, 25.96158339376447,
4.1171232558616255, 3.686075199425159, 7.271860583120944, 3.262253597763473,
3.3004576752775066, 3.6205698455960458, 4.165036235815391, 14.441242204987121,
7.991325258784238, 3.907635756103933, 5.125068509592997, 4.243525646984372,
15.514272504430156, 15.287270087835779, 4.721507233404716, 3.113793320398831,
8.305061814006187, 7.407788710034711, 3.3179619872865977, 4.028778431680334,
12.94974211443884, 5.650787114166689, 3.3238931026251604, 5.906551144849444,
6.914988923643794, 5.769328537387196, 4.18861507917515, 3.1124311079275335,
4.645828969863903, 4.101260593110335, 7.124837325675765, 4.578264054393883,
4.7495481020930885, 17.28809497496091, 3.417769904618811, 4.729616139690189,
7.712193699398478, 3.3542209067186577, 3.6777965318236756, 9.013308961180787,

14.33484161835653, 5.096163022196799, 30.288498141601842, 4.611259534410565,
10.522143343960064, 4.794769885478998, 7.940836945286304, 3.953588843182373,
5.56781362993937, 29.43625971470472, 3.4389415972449733, 7.1028287397932255,
4.700468855026193, 3.3494952085636274, 3.366206714385642, 9.231041642130211,
3.9580409035405917, 5.583158118987496, 8.293813712798814, 5.1951692553124715,
7.999656620118212, 14.075393053531775, 8.62040207784079, 3.1922563771925976,
3.5384202608855713, 12.389395968361018, 3.49626051495241, 3.419685605839865,
6.330162362015155, 6.7324247332446, 3.2853683220222174, 30.996614177289654,
3.8300700995738954, 3.5339661727271188, 6.203161754113323, 3.6962399138722217,
3.1369534852659613, 4.72227941038228, 5.799496068669598, 3.404208881619262,
10.959082666946681, 36.00795193383489, 6.108450436617197, 6.960894190748951,
4.014038054606617, 3.201542265928771, 9.529232296378101, 8.295686186378907,
3.7852332948571887, 3.658679099440957, 40.69931958944351, 17.939363135279905,
5.215598606303949, 6.4705420206427, 3.6084516211898077, 3.5557536828437377,
4.33800705827436, 3.928931338880211, 3.1002616777433145, 3.066634640040978,
6.75303250315626, 3.7007832653861388, 4.8934432735342295, 8.204133054416072,
3.332558661461237, 4.489852108330581, 10.99809156508577, 7.09550285004185,
4.3631317537597525, 3.1540573997015713, 8.232265670615208, 3.029510534719258,
3.3068811934055575, 3.20552802416793, 16.735057105727098, 14.972236340770474,
6.324210092681846, 4.618376764977439, 6.602206294544243, 7.884946161105747,
6.065461119651585, 3.1070357882924364, 11.363582746668161, 3.713907221621051,
3.8894503181812587, 4.238029961746982, 7.944195758282549, 3.151856912226169,
3.045493591893081, 12.862331373054502, 3.0483263493520822, 3.2026956973183367,
21.756054314627427, 3.5711663392700492, 3.551573771895341, 3.0990835245096315,
9.264334835608818, 17.648257223909138, 12.886527104049128, 3.5800812455050823,
5.238314787677692, 4.305023260744054, 18.747433255892613, 3.9374803446674806,
3.9366890896421647, 3.1498345213474788, 7.873375300259983, 3.147488052062871,
5.043917482806991, 3.2842072759838223, 3.060355914937084, 6.343540113175542,
7.0913415905307895, 3.1848933702094704, 7.423340043886186, 4.550769007759385,
3.3600279267159943, 11.549771462280772, 4.476112393765513, 6.760642642711244,
3.5785801213114667, 3.04236645179217, 3.1356825534518604, 5.859621629001818,
3.1686861052054938, 5.08513012186505, 4.415961540141988, 6.165951242675327,
11.528185030576143, 3.055385717928602, 5.122944207875256, 6.331368554466018,
4.139128533472685, 4.766054369944768, 25.004048115418307, 13.624934816582545,
3.5366634615399364, 15.701505920839983, 4.042600791313925, 4.835602924690417,
7.117550003052229, 4.560398941473492, 8.688934625816673, 7.915112685429367,
3.0715527981056123, 4.22688062259575, 11.806315688071194, 3.559462216388043,
5.239171335589216, 12.63563997913021, 4.672926567501072, 4.76000798438567,
3.905814182882992, 3.449527594970022, 10.133996781547665, 4.427981946373356,
16.518475127526592, 3.273180650047664, 6.986583251940217, 3.1653287888566486,
22.62877892862111, 3.4388694398630357, 3.4005358733294533, 4.86152767767057,
3.1535288567005413, 6.035217917600341, 4.160317555081746, 3.659194531705247,
3.406836901100432, 3.046985654982173, 5.897894468989344, 7.148313109729566,
7.385239501148884, 7.0345910132122995, 7.1817648553840066, 3.4965242123221847,
3.0047099935220016, 18.635152597499047, 7.222513397754432, 8.251619190864691,
4.596385052490389, 3.1626421684895347, 3.756230208370719, 5.288181320103619,
4.529987917388208, 43.87112273176157, 4.903099417736773, 4.668434400411335,
3.212118115559712, 8.11718614128857, 3.2918780494472064, 4.088272698346382,

4.643336746006974, 3.047200015653288, 8.464437876340604, 6.262911079741773,
3.002646742863227, 10.507743313342619, 4.410371244341008, 3.208024672910942,
4.285062845335957, 3.716807207896493, 12.298931731001268, 3.8169480534284146,
3.771528806639889, 4.99354954149877, 4.116935446507242, 3.4506312479475434,
10.490445112355387, 15.117389269173513, 6.824277041285825, 3.3830422316581745,
17.08723757030537, 3.877199831496733, 11.293949705969252, 4.59089508596321,
3.247875477454583, 5.6327363944040885, 6.394530763121657, 3.179779269911351,
5.210299636318469, 11.128346028741202, 3.5084249782532217, 6.041352737922534,
6.715837399983483, 13.344651444585384, 3.287341944472855, 4.077379008203629,
4.24498846838318, 3.076294353896342, 26.9742444675445, 9.198687830824172,
15.65415110216128, 5.401925355473252, 11.310126643179949, 3.606257547057831,
3.637190500842642, 7.479646144043834, 4.4647516307449955, 4.719083972195599,
3.791004860527087, 5.079821988493041, 4.194003748695872, 3.9038299140216113,
12.264992123629726, 4.558871095651714, 4.366310323318365, 3.3719079314432987,
9.208573292073364, 5.636269232189311, 3.042880454227539, 3.2437635288338162,
5.354012242531254, 4.5733344797760855, 15.17951285905951, 3.7494446325060826,
5.442508390291414, 4.091596021205675, 3.7666851535454704, 6.698881630999926,
3.4974265270849316, 3.396325609049709, 3.5706781751506833, 3.331688591663097,
4.304786106279233, 6.53414756879351, 3.279865401049745, 3.8481484854730312,
4.687907681180349, 7.596385896435081, 3.3295372706830424, 5.724936594945788,
8.752498334377679, 12.230485420303914, 5.684806461498291, 9.690929052693756,
3.082071920114193, 6.833886120174648, 6.466579558730582, 12.733799433082238,
7.456116084562549, 39.79275100110514, 4.501843123802668, 9.727914377418474,
4.236105876731552, 18.305508768580186, 10.359434214275312, 3.4860764474199364,
5.45887140030486, 14.392440860184715, 3.622674039136333, 3.2809347630949355,
5.1859536415638745, 6.427627140953049, 3.189551285084441, 8.359198173729819,
9.037813574610738, 5.669892008937887, 4.259155324257293, 10.361086576243357,
5.253665224640707, 3.2121065874261796, 4.16217104220228, 5.074867655607489,
5.133688400832481, 6.359768938664148, 3.5539281451679217, 3.3527586586829354,
3.077169575626148, 7.301700861990003, 6.427728962073698, 14.103327216885036,
3.234823423437197, 3.096267613740353, 4.531791891779074, 3.57562622170414,
5.720688733390023, 5.553066592544702, 3.1935371138722353, 4.22551547241441,
5.263846500933999, 3.2608101889055474, 5.379553498406791, 4.990770072002603,
14.984052025328586, 3.925381772602611, 7.736245388687584, 5.184343685164087,
25.24303955284834, 3.910256932854031, 5.701517259195812, 3.5710390273651647,
16.810014513693105, 3.6415633354403387, 4.52913600098107, 3.069909698973346,
4.748667072021165, 4.099097225180463, 8.555517155523432, 5.157636322301903,
4.142134179752488, 18.81256754843317, 6.078082504211325, 4.362444047393843,
11.403791243852947, 6.266621901959487, 3.328434995896085, 14.594162748386623,
5.201304910622362, 3.032915244494066, 4.839243469936544, 3.0874630265415486,
9.084202166155642, 7.915863058142283, 3.054244545954312, 3.326293378029351,
3.9288827090140024, 17.433842203758537, 5.427432657964378, 3.8205317078550776,
8.361499120212827, 3.6761328424370205, 6.350769230926139, 5.6636997926788135,
3.253323757025687, 3.0620806302014376, 5.38206583105746, 4.982401825020206,
7.242404245526383, 3.057384353303241, 3.603300837147513, 3.0597263060155693,
8.213437365233537, 3.5671838552984667, 5.205992295289465, 3.2548444593575625,
3.5978416522094667, 3.438961212911863, 3.6560137649648046, 7.576647949568925,
8.222271351805583, 4.899086123008387, 5.2495332124530085, 31.89272540423009,

5.2594196867373535, 4.111840498990468, 9.724886102034134, 3.234334386039254,
3.8893603001770454, 3.088155784189739, 7.040714160043742, 8.549577702579578,
6.977764405213402, 4.103885300557433, 5.159109078105433, 12.818978934641066,
3.8592359261491485, 3.2799379553604693, 6.204223587654703, 12.334678338678433,
3.231316594690268, 3.1513886250577263, 3.702736401648382, 6.773042006261474,
3.9224054043740226, 3.068722683587697, 5.601030053760886, 3.21165804712887,
3.0320006666529196, 7.379110597365004, 4.156451528461503, 3.7810974880935673,
5.893637154446253, 16.475607517065953, 3.1999987888864605, 5.679382673126462,
8.233742629398279, 3.156346465480863, 4.140177452634991, 4.04051713939542,
4.862339892095331, 5.015466763164255, 12.717566924641321, 4.582152461203617,
3.5550539292577077, 7.088846580959636, 3.4270841525980726, 3.2470194173828175,
3.7244597838016444, 6.3363822403238075, 11.20530158099847, 4.119650956416127,
9.358553754261852, 3.2223541794689865, 3.734709772335638, 5.021074945086651,
4.928036772235037, 6.2959520485561535, 17.333442393767452, 6.142400277919723,
7.114738869695969, 4.770508382537793, 16.664591651323633, 4.95573884380166,
17.934453281335646, 8.814581354657424, 5.303482400052128, 4.715658515217584,
5.215737051042956, 3.0967405529777023, 5.481193828787474, 3.0881846093241214,
4.611961891754806, 3.5789181596326394, 3.244430346932767, 6.185386291154919,
4.107670419505254, 3.4089974203899414, 9.106192919127192, 3.1168934959247845,
3.200606979674814, 6.9595816077126065, 4.898400245825327, 4.971574047088776,
11.073928912922936, 7.418312214783999, 19.882512349807296, 4.472571232398228,
3.004635832677757, 7.9897481473314125, 4.881089190509152, 3.092576696844384,
4.245785134029302, 4.783121263678621, 3.5688491851652397, 29.644538321269238,
4.855753444797344, 4.0518616409825565, 3.9306494424886544, 3.6485869967847275,
3.847084406938928, 4.372058530600047, 4.819521465656375, 3.5442990867218076,
3.428953805789695, 3.911613046463392, 3.913487683059235, 19.668686499386922,
3.5059604792638916, 7.263183882984908, 8.110528800894246, 5.627335774744658,
5.2201431080903875, 3.045823359459874, 4.445609332698415, 4.194109166196731,
3.0983604320511375, 3.1660695887738983, 3.469176854741428, 3.900113271363095,
15.834939413122914, 3.6107097953489635, 5.198690106062859, 4.137059377141707,
6.160159072252079, 4.45173824921889, 5.28595133450851, 12.240662561711964,
11.259424463640627, 6.762381255570188, 3.2876874858062637, 9.85069348870313,
4.239961274045716, 5.536363174515533, 4.187927375825802, 4.266068566164744,
6.031836240607741, 4.000673458144771, 7.704849821431807, 3.1265018055794003,
3.2395873334036236, 5.339116842081187, 10.26612766897697, 5.378443553539779,
21.70893410765693, 3.5197464625939876, 4.26785005955455, 8.259095997445565,
25.75680378348063, 4.404488089305343, 3.507404969326954, 5.090567325949153,
11.417307849431781, 5.504272023103475, 6.213134069490195, 11.652721856284403,
5.282343398052517, 4.50968753740549, 3.012434497312489, 5.673794743651045,
4.007495862559387, 4.582337921539289, 5.448084438468453, 5.040187325084262,
3.3769823021126264, 4.822954046262556, 7.229996747550928, 4.308907022620836,
5.6091127829731444, 6.4820927388097545, 5.075543634108843, 15.246830354281423,
6.942351522649005, 3.578445089515408, 6.970256526191647, 6.676678779249219,
4.0029589594177795, 11.370634470659445, 4.943045677254404, 5.279159338988372,
3.808790380704307, 3.134906580525937, 21.639129701998552, 7.2329628850056515,
4.646521154494995, 5.202269804251806, 3.451062713193793, 3.417037502540793,
3.0533431411878693, 6.872722743487708, 3.43000875309231, 16.284595124748282,
3.211505204337852, 3.720221974523588, 6.016207661385133, 8.753526901875134,

```
5.062145056807543, 6.642866709700375, 4.841062606694146, 24.55632015574802,
3.2231494488467956, 3.5075583202951996, 3.299468678579808, 10.719487608582362,
6.385340191797744, 9.478966894618905, 3.322492169667886, 4.685345947304696,
21.605673887779993, 7.787737366317487, 3.005474606708849, 4.721020800262338,
3.4694010540008677, 3.844071706098345, 3.190074370075443, 5.706339434106936,
5.539985211724561, 3.8193827471600947, 4.46755887591459, 3.3720067312226885,
3.954767490256503, 5.834224377714028, 4.091655887501675, 35.16340578255477]
```

3.0.2 Appending a column called Z-score to the dataset to store the Z-score of each row

```
[ ]: df["Z-score"] = (df["Money difference"] - mean)/std
```

3.0.3 Making a column called Fraud that represents if we consider the data to be a fraudulent data or not depending on the Z-score. If the Z-score is above 3 then we consider it to be anomalous/fraudulent or if the Z-score is less then we consider it normal.

```
[ ]: df['Fraud'] = df['Z-score'].apply(lambda x: 0 if x <= 3 else 1)
```

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

```
[ ]:                 Full name Middle Initial of the Provider  \
     0  SATYASREE UPADHYAYULA                            NaN
     1           WENDY JONES                              P
     2       RICHARD DUROCHER                             W
     3         JASPER FULLARD                           NaN
     4       ANTHONY PERROTTI                             E

       Credentials of the Provider Gender of the Provider  \
     0                      M.D.                       F
     1                      M.D.                       F
     2                       DPM                       M
     3                        MD                       M
     4                        DO                       M

       Entity Type of the Provider City of the Provider  Zip Code of the Provider  \
     0                           I          SAINT LOUIS                631041004.0
     1                           I         FAYETTEVILLE                283043815.0
     2                           I          NORTH HAVEN                 64732343.0
     3                           I          KANSAS CITY                641183998.0
     4                           I              JUPITER                334585700.0

       State Code of the Provider             Provider Type  \
     0                         MO         Internal Medicine
     1                         NC   Obstetrics & Gynecology
     2                         CT                  Podiatry
```

```
3                              MO       Internal Medicine
4                              FL       Internal Medicine

   Medicare Participation Indicator  … HCPCS Code HCPCS Drug Indicator  \
0                               Y  …     99223                     N
1                               Y  …     G0202                     N
2                               Y  …     99348                     N
3                               Y  …     81002                     N
4                               Y  …     96372                     N

   Number of Services Number of Medicare Beneficiaries  \
0                  27                                 24
1                 175                                175
2                  32                                 13
3                  20                                 18
4                  33                                 24

   Number of Distinct Medicare Beneficiary/Per Day Services  \
0                                                 27
1                                                175
2                                                 32
3                                                 20
4                                                 31

   Average Submitted Charge Amount  Average Medicare Payment Amount  \
0                        -0.046433                         0.400082
1                         0.182805                         0.207649
2                        -0.187794                        -0.064687
3                        -0.328957                        -0.370166
4                        -0.296019                        -0.289505

   Money difference    Z-score  Fraud
0         -0.139738 -0.139738      0
1          0.165185  0.165185      0
2         -0.201787 -0.201787      0
3         -0.298004 -0.298004      0
4         -0.277579 -0.277579      0

[5 rows x 21 columns]
```

```
[ ]: df['Fraud'].value_counts()
```

```
[ ]: Fraud
     0     98933
     1      1067
     Name: count, dtype: int64
```

```
[ ]: df.nunique()
```

```
[ ]: Full name                                                84197
     Middle Initial of the Provider                              29
     Credentials of the Provider                               1854
     Gender of the Provider                                       2
     Entity Type of the Provider                                 2
     City of the Provider                                      5846
     Zip Code of the Provider                                 47827
     State Code of the Provider                                  58
     Provider Type                                               90
     Medicare Participation Indicator                             2
     Place of Service                                             2
     HCPCS Code                                                2631
     HCPCS Drug Indicator                                         2
     Number of Services                                        2748
     Number of Medicare Beneficiaries                          1274
     Number of Distinct Medicare Beneficiary/Per Day Services  1979
     Average Submitted Charge Amount                          38088
     Average Medicare Payment Amount                          83367
     Money difference                                         91947
     Z-score                                                  91947
     Fraud                                                        2
     dtype: int64
```

```
[ ]: df.drop(columns=['Zip Code of the Provider'], inplace=True)
```

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

```
[ ]:             Full name Middle Initial of the Provider  \
     0  SATYASREE UPADHYAYULA                            NaN
     1            WENDY JONES                              P
     2       RICHARD DUROCHER                              W
     3         JASPER FULLARD                            NaN
     4       ANTHONY PERROTTI                              E

       Credentials of the Provider Gender of the Provider  \
     0                        M.D.                       F
     1                        M.D.                       F
     2                         DPM                       M
     3                          MD                       M
     4                          DO                       M

       Entity Type of the Provider City of the Provider State Code of the Provider  \
     0                           I          SAINT LOUIS                          MO
     1                           I         FAYETTEVILLE                          NC
     2                           I          NORTH HAVEN                          CT
```

```
3                         I          KANSAS CITY                    MO
4                         I           JUPITER                       FL


            Provider Type Medicare Participation Indicator Place of Service  \
0          Internal Medicine                               Y                 F
1  Obstetrics & Gynecology                                 Y                 O
2                 Podiatry                                 Y                 O
3          Internal Medicine                               Y                 O
4          Internal Medicine                               Y                 O


  HCPCS Code HCPCS Drug Indicator Number of Services  \
0     99223                    N                  27
1     G0202                    N                 175
2     99348                    N                  32
3     81002                    N                  20
4     96372                    N                  33


  Number of Medicare Beneficiaries  \
0                               24
1                              175
2                               13
3                               18
4                               24


  Number of Distinct Medicare Beneficiary/Per Day Services  \
0                                                 27
1                                                175
2                                                 32
3                                                 20
4                                                 31


  Average Submitted Charge Amount  Average Medicare Payment Amount  \
0                      -0.046433                         0.400082
1                       0.182805                         0.207649
2                      -0.187794                        -0.064687
3                      -0.328957                        -0.370166
4                      -0.296019                        -0.289505


  Money difference    Z-score   Fraud
0        -0.139738  -0.139738       0
1         0.165185   0.165185       0
2        -0.201787  -0.201787       0
3        -0.298004  -0.298004       0
4        -0.277579  -0.277579       0
```

3.0.4 Applying One hot encoding the the columns with few categories

```
df = pd.get_dummies(df, columns=['Gender of the Provider','Entity Type of the
    Provider','Place of Service','Medicare Participation Indicator','HCPCS Drug
    Indicator'], dtype='int')
```

```
df.head()
```

```
                 Full name Middle Initial of the Provider  \
0  SATYASREE UPADHYAYULA                               NaN
1           WENDY JONES                                   P
2        RICHARD DUROCHER                                W
3          JASPER FULLARD                              NaN
4        ANTHONY PERROTTI                                 E

  Credentials of the Provider City of the Provider State Code of the Provider  \
0                        M.D.          SAINT LOUIS                          MO
1                        M.D.         FAYETTEVILLE                          NC
2                         DPM          NORTH HAVEN                          CT
3                          MD          KANSAS CITY                          MO
4                          DO              JUPITER                          FL

            Provider Type HCPCS Code Number of Services  \
0       Internal Medicine      99223                 27
1  Obstetrics & Gynecology     G0202                175
2                Podiatry      99348                 32
3       Internal Medicine      81002                 20
4       Internal Medicine      96372                 33

   Number of Medicare Beneficiaries  \
0                                24
1                               175
2                                13
3                                18
4                                24

   Number of Distinct Medicare Beneficiary/Per Day Services  ...  \
0                                                 27          ...
1                                                175          ...
2                                                 32          ...
3                                                 20          ...
4                                                 31          ...

   Gender of the Provider_F  Gender of the Provider_M  \
0                         1                         0
1                         1                         0
2                         0                         1
```

```
        3                          0                           1
        4                          0                           1

            Entity Type of the Provider_I  Entity Type of the Provider_O  \
        0                              1                              0
        1                              1                              0
        2                              1                              0
        3                              1                              0
        4                              1                              0

            Place of Service_F  Place of Service_O  Medicare Participation Indicator_N  \
        0                    1                   0                                   0
        1                    0                   1                                   0
        2                    0                   1                                   0
        3                    0                   1                                   0
        4                    0                   1                                   0

            Medicare Participation Indicator_Y  HCPCS Drug Indicator_N  \
        0                                    1                       1
        1                                    1                       1
        2                                    1                       1
        3                                    1                       1
        4                                    1                       1

            HCPCS Drug Indicator_Y
        0                        0
        1                        0
        2                        0
        3                        0
        4                        0

        [5 rows x 25 columns]
```

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

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 100000 entries, 0 to 99999
Data columns (total 25 columns):
 #   Column                                      Non-Null Count
Dtype
---  ------                                      --------------
-----
 0   Full name                                   100000 non-null
object
 1   Middle Initial of the Provider              70669 non-null
object
 2   Credentials of the Provider                 92791 non-null
```

```
 object
 3   City of the Provider                                  100000 non-null
 object
 4   State Code of the Provider                            100000 non-null
 object
 5   Provider Type                                         100000 non-null
 object
 6   HCPCS Code                                            100000 non-null
 object
 7   Number of Services                                    100000 non-null
 object
 8   Number of Medicare Beneficiaries                      100000 non-null
 object
 9   Number of Distinct Medicare Beneficiary/Per Day Services  100000 non-null
 object
 10  Average Submitted Charge Amount                       100000 non-null
 float64
 11  Average Medicare Payment Amount                       100000 non-null
 float64
 12  Money difference                                      100000 non-null
 float64
 13  Z-score                                               100000 non-null
 float64
 14  Fraud                                                 100000 non-null
 int64
 15  Gender of the Provider_F                              100000 non-null
 int64
 16  Gender of the Provider_M                              100000 non-null
 int64
 17  Entity Type of the Provider_I                         100000 non-null
 int64
 18  Entity Type of the Provider_O                         100000 non-null
 int64
 19  Place of Service_F                                    100000 non-null
 int64
 20  Place of Service_O                                    100000 non-null
 int64
 21  Medicare Participation Indicator_N                    100000 non-null
 int64
 22  Medicare Participation Indicator_Y                    100000 non-null
 int64
 23  HCPCS Drug Indicator_N                                100000 non-null
 int64
 24  HCPCS Drug Indicator_Y                                100000 non-null
 int64
dtypes: float64(4), int64(11), object(10)
memory usage: 19.1+ MB
```

```
df.drop(columns=['State Code of the Provider'], inplace=True)
```

3.0.5 Applying label encoding the fields with large no. of categories

```python
from sklearn import preprocessing

label_encoder = preprocessing.LabelEncoder()

df['Provider Type'] = label_encoder.fit_transform(df['Provider Type'])
```

```python
df_frequencies = df['Provider Type'].value_counts()

df_frequency_map = df_frequencies.to_dict()

df['Provider Type'] = df['Provider Type'].map(df_frequency_map)
```

```python
df.head()
```

```
               Full name Middle Initial of the Provider  \
0  SATYASREE UPADHYAYULA                             NaN
1           WENDY JONES                               P
2       RICHARD DUROCHER                              W
3         JASPER FULLARD                            NaN
4       ANTHONY PERROTTI                              E

  Credentials of the Provider City of the Provider  Provider Type HCPCS Code  \
0                        M.D.          SAINT LOUIS          11366     99223
1                        M.D.         FAYETTEVILLE           1028     G0202
2                         DPM          NORTH HAVEN           2027     99348
3                          MD          KANSAS CITY          11366     81002
4                          DO              JUPITER          11366     96372

  Number of Services Number of Medicare Beneficiaries  \
0                 27                               24
1                175                              175
2                 32                               13
3                 20                               18
4                 33                               24

  Number of Distinct Medicare Beneficiary/Per Day Services  \
0                                                 27
1                                                175
2                                                 32
3                                                 20
4                                                 31

  Average Submitted Charge Amount  …  Gender of the Provider_F  \
```

```
0                            -0.046433  …                              1
1                             0.182805  …                              1
2                            -0.187794  …                              0
3                            -0.328957  …                              0
4                            -0.296019  …                              0

   Gender of the Provider_M  Entity Type of the Provider_I  \
0                         0                              1
1                         0                              1
2                         1                              1
3                         1                              1
4                         1                              1

   Entity Type of the Provider_O  Place of Service_F  Place of Service_O  \
0                              0                   1                   0
1                              0                   0                   1
2                              0                   0                   1
3                              0                   0                   1
4                              0                   0                   1

   Medicare Participation Indicator_N  Medicare Participation Indicator_Y  \
0                                   0                                   1
1                                   0                                   1
2                                   0                                   1
3                                   0                                   1
4                                   0                                   1

   HCPCS Drug Indicator_N  HCPCS Drug Indicator_Y
0                       1                       0
1                       1                       0
2                       1                       0
3                       1                       0
4                       1                       0

[5 rows x 24 columns]
```

```python
df.drop(columns=['City of the Provider'], inplace=True)
```

```python
df.head()
```

```
             Full name Middle Initial of the Provider  \
0  SATYASREE UPADHYAYULA                           NaN
1           WENDY JONES                             P
2       RICHARD DUROCHER                           W
3         JASPER FULLARD                           NaN
4       ANTHONY PERROTTI                           E
```

```
   Credentials of the Provider  Provider Type HCPCS Code Number of Services  \
0                        M.D.          11366      99223                  27
1                        M.D.           1028      G0202                 175
2                         DPM           2027      99348                  32
3                          MD          11366      81002                  20
4                          DO          11366      96372                  33


   Number of Medicare Beneficiaries  \
0                                24
1                               175
2                                13
3                                18
4                                24


   Number of Distinct Medicare Beneficiary/Per Day Services  \
0                                                27
1                                               175
2                                                32
3                                                20
4                                                31


   Average Submitted Charge Amount  Average Medicare Payment Amount  …  \
0                        -0.046433                         0.400082  …
1                         0.182805                         0.207649  …
2                        -0.187794                        -0.064687  …
3                        -0.328957                        -0.370166  …
4                        -0.296019                        -0.289505  …


   Gender of the Provider_F  Gender of the Provider_M  \
0                         1                         0
1                         1                         0
2                         0                         1
3                         0                         1
4                         0                         1


   Entity Type of the Provider_I  Entity Type of the Provider_O  \
0                              1                              0
1                              1                              0
2                              1                              0
3                              1                              0
4                              1                              0


   Place of Service_F  Place of Service_O  Medicare Participation Indicator_N  \
0                   1                   0                                   0
1                   0                   1                                   0
2                   0                   1                                   0
3                   0                   1                                   0
```

```
4                           0                         1                                     0

    Medicare Participation Indicator_Y  HCPCS Drug Indicator_N  \
0                                   1                        1
1                                   1                        1
2                                   1                        1
3                                   1                        1
4                                   1                        1

    HCPCS Drug Indicator_Y
0                        0
1                        0
2                        0
3                        0
4                        0

[5 rows x 23 columns]
```

```
[ ]: df.drop(columns=['Full name','Middle Initial of the Provider','Credentials of␣
     ↪the Provider'], inplace=True)
```

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

```
[ ]:    Provider Type HCPCS Code  Number of Services  \
0             11366      99223                  27
1              1028      G0202                 175
2              2027      99348                  32
3             11366      81002                  20
4             11366      96372                  33

   Number of Medicare Beneficiaries  \
0                                24
1                               175
2                                13
3                                18
4                                24

   Number of Distinct Medicare Beneficiary/Per Day Services  \
0                                                27
1                                               175
2                                                32
3                                                20
4                                                31

   Average Submitted Charge Amount  Average Medicare Payment Amount  \
0                        -0.046433                         0.400082
1                         0.182805                         0.207649
```

```
2                    -0.187794                    -0.064687
3                    -0.328957                    -0.370166
4                    -0.296019                    -0.289505

   Money difference   Z-score  Fraud  Gender of the Provider_F  \
0         -0.139738 -0.139738      0                         1
1          0.165185  0.165185      0                         1
2         -0.201787 -0.201787      0                         0
3         -0.298004 -0.298004      0                         0
4         -0.277579 -0.277579      0                         0

   Gender of the Provider_M  Entity Type of the Provider_I  \
0                         0                              1
1                         0                              1
2                         1                              1
3                         1                              1
4                         1                              1

   Entity Type of the Provider_O  Place of Service_F  Place of Service_O  \
0                              0                   1                   0
1                              0                   0                   1
2                              0                   0                   1
3                              0                   0                   1
4                              0                   0                   1

   Medicare Participation Indicator_N  Medicare Participation Indicator_Y  \
0                                   0                                   1
1                                   0                                   1
2                                   0                                   1
3                                   0                                   1
4                                   0                                   1

   HCPCS Drug Indicator_N  HCPCS Drug Indicator_Y
0                       1                       0
1                       1                       0
2                       1                       0
3                       1                       0
4                       1                       0
```

```python
[ ]: df['Fraud'].value_counts()
```

```
[ ]: Fraud
     0    98933
     1     1067
     Name: count, dtype: int64
```

```python
[ ]: df['Number of Services'] = scaler.fit_transform(df[['Number of Services']])
```

```
df['Number of Medicare Beneficiaries'] = scaler.fit_transform(df[['Number of
  ↪Medicare Beneficiaries']])
```

```
df['Number of Distinct Medicare Beneficiary/Per Day Services'] = scaler.
  ↪fit_transform(df[['Number of Distinct Medicare Beneficiary/Per Day
  ↪Services']])
```

```
df.head()
```

```
     Provider Type HCPCS Code  Number of Services  \
0            11366      99223           -0.085301
1             1028      G0202           -0.025939
2             2027      99348           -0.083296
3            11366      81002           -0.088109
4            11366      96372           -0.082895

     Number of Medicare Beneficiaries  \
0                            -0.059308
1                             0.076775
2                            -0.069222
3                            -0.064716
4                            -0.059308

     Number of Distinct Medicare Beneficiary/Per Day Services  \
0                                            -0.070183
1                                             0.020049
2                                            -0.067135
3                                            -0.074451
4                                            -0.067744

     Average Submitted Charge Amount  Average Medicare Payment Amount  \
0                          -0.046433                         0.400082
1                           0.182805                         0.207649
2                          -0.187794                        -0.064687
3                          -0.328957                        -0.370166
4                          -0.296019                        -0.289505

     Money difference   Z-score  Fraud  Gender of the Provider_F  \
0           -0.139738 -0.139738      0                         1
1            0.165185  0.165185      0                         1
2           -0.201787 -0.201787      0                         0
3           -0.298004 -0.298004      0                         0
4           -0.277579 -0.277579      0                         0

     Gender of the Provider_M  Entity Type of the Provider_I  \
0                           0                              1
1                           0                              1
```

```
2                         1                           1
3                         1                           1
4                         1                           1

   Entity Type of the Provider_O  Place of Service_F  Place of Service_O  \
0                               0                   1                   0
1                               0                   0                   1
2                               0                   0                   1
3                               0                   0                   1
4                               0                   0                   1

   Medicare Participation Indicator_N  Medicare Participation Indicator_Y  \
0                                    0                                    1
1                                    0                                    1
2                                    0                                    1
3                                    0                                    1
4                                    0                                    1

   HCPCS Drug Indicator_N  HCPCS Drug Indicator_Y
0                        1                       0
1                        1                       0
2                        1                       0
3                        1                       0
4                        1                       0
```

3.0.6 Dropping the Fraud column to avoid bias in the model while training

```python
df.drop(columns=['Fraud'], inplace=True)
```

```python
df.drop(columns=['HCPCS Code'], inplace=True)
```

```python
df.head()
```

```
   Provider Type  Number of Services  Number of Medicare Beneficiaries  \
0          11366           -0.085301                         -0.059308
1           1028           -0.025939                          0.076775
2           2027           -0.083296                         -0.069222
3          11366           -0.088109                         -0.064716
4          11366           -0.082895                         -0.059308

   Number of Distinct Medicare Beneficiary/Per Day Services  \
0                                          -0.070183
1                                           0.020049
2                                          -0.067135
3                                          -0.074451
4                                          -0.067744
```

```
    Average Submitted Charge Amount  Average Medicare Payment Amount  \
0                        -0.046433                         0.400082
1                         0.182805                         0.207649
2                        -0.187794                        -0.064687
3                        -0.328957                        -0.370166
4                        -0.296019                        -0.289505


    Money difference   Z-score  Gender of the Provider_F  \
0         -0.139738 -0.139738                         1
1          0.165185  0.165185                         1
2         -0.201787 -0.201787                         0
3         -0.298004 -0.298004                         0
4         -0.277579 -0.277579                         0


    Gender of the Provider_M  Entity Type of the Provider_I  \
0                         0                              1
1                         0                              1
2                         1                              1
3                         1                              1
4                         1                              1


    Entity Type of the Provider_O  Place of Service_F  Place of Service_O  \
0                              0                   1                   0
1                              0                   0                   1
2                              0                   0                   1
3                              0                   0                   1
4                              0                   0                   1


    Medicare Participation Indicator_N  Medicare Participation Indicator_Y  \
0                                  0                                   1
1                                  0                                   1
2                                  0                                   1
3                                  0                                   1
4                                  0                                   1


    HCPCS Drug Indicator_N  HCPCS Drug Indicator_Y
0                        1                       0
1                        1                       0
2                        1                       0
3                        1                       0
4                        1                       0
```

3.0.7 Autoencoders

```python
import tensorflow as tf
from tensorflow.keras.models import Model
from tensorflow.keras.layers import Input, Dense
from tensorflow.keras import regularizers
from sklearn.metrics import mean_squared_error
```

```python
X = df
```

```python
input_dim = X.shape[1]
encoding_dim = 32

input_layer = Input(shape=(input_dim,))
encoder = Dense(encoding_dim, activation="relu",
  activity_regularizer=regularizers.l1(10e-5))(input_layer)
encoder = Dense(16, activation="relu")(encoder)
encoder = Dense(2, activation="relu")(encoder)
encoder = Dense(16, activation="relu")(encoder)
decoder = Dense(32, activation="relu")(encoder)
decoder = Dense(input_dim, activation="relu")(decoder)

autoencoder = Model(inputs=input_layer, outputs=decoder)
autoencoder.compile(optimizer='adam', loss='mean_squared_error')
```

```python
autoencoder.summary()
```

```
Model: "model_7"

_____
 Layer (type)                Output Shape              Param #
=================================================================
 input_8 (InputLayer)        [(None, 18)]              0

 dense_40 (Dense)            (None, 32)                608

 dense_41 (Dense)            (None, 16)                528

 dense_42 (Dense)            (None, 2)                 34

 dense_43 (Dense)            (None, 16)                48

 dense_44 (Dense)            (None, 32)                544

 dense_45 (Dense)            (None, 18)                594

=================================================================
Total params: 2356 (9.20 KB)
Trainable params: 2356 (9.20 KB)
```

Non-trainable params: 0 (0.00 Byte)

```
[ ]: tf.keras.utils.plot_model(
         autoencoder,
         to_file='model.png',
         show_shapes=True,
         show_dtype=True,
         show_layer_names=True,
         rankdir='TB',
         expand_nested=False,
         dpi=200,
         show_layer_activations=True,
         show_trainable=False,
     )
```

[ ]:

input_8	input:	[(None, 18)]
InputLayer		
float32	output:	[(None, 18)]

dense_40	input:	(None, 18)
Dense	relu	
float32	output:	(None, 32)

dense_41	input:	(None, 32)
Dense	relu	
float32	output:	(None, 16)

dense_42	input:	(None, 16)
Dense	relu	
float32	output:	(None, 2)

dense_43	input:	(None, 2)
Dense	relu	
float32	output:	(None, 16)

dense_44	input:	(None, 16)
Dense	relu	
float32	output:	(None, 32)

dense_45	input:	(None, 32)
Dense	relu	56
float32	output:	(None, 18)

```
history = autoencoder.fit(X, X, epochs=50, batch_size=32, shuffle=True,
    validation_split=0.2)
```

```
Epoch 1/50
2500/2500 [==============================] - 9s 3ms/step - loss: 2218371.2500 -
val_loss: 50.1618
Epoch 2/50
2500/2500 [==============================] - 7s 3ms/step - loss: 25.1066 -
val_loss: 1.9095
Epoch 3/50
2500/2500 [==============================] - 8s 3ms/step - loss: 1.7546 -
val_loss: 1.8728
Epoch 4/50
2500/2500 [==============================] - 6s 2ms/step - loss: 7.8894 -
val_loss: 12.8511
Epoch 5/50
2500/2500 [==============================] - 8s 3ms/step - loss: 12.5431 -
val_loss: 1.8378
Epoch 6/50
2500/2500 [==============================] - 6s 3ms/step - loss: 9.3502 -
val_loss: 1.7894
Epoch 7/50
2500/2500 [==============================] - 7s 3ms/step - loss: 10.7063 -
val_loss: 1.7205
Epoch 8/50
2500/2500 [==============================] - 7s 3ms/step - loss: 9.9165 -
val_loss: 1.7072
Epoch 9/50
2500/2500 [==============================] - 8s 3ms/step - loss: 17.6724 -
val_loss: 1.6992
Epoch 10/50
2500/2500 [==============================] - 7s 3ms/step - loss: 8.6495 -
val_loss: 1.6919
Epoch 11/50
2500/2500 [==============================] - 8s 3ms/step - loss: 9.0176 -
val_loss: 1.6868
Epoch 12/50
2500/2500 [==============================] - 6s 3ms/step - loss: 8.1592 -
val_loss: 1.6789
Epoch 13/50
2500/2500 [==============================] - 8s 3ms/step - loss: 5.5013 -
val_loss: 72.0353
Epoch 14/50
2500/2500 [==============================] - 6s 2ms/step - loss: 6.8135 -
val_loss: 1.6794
```

```
Epoch 15/50
2500/2500 [==============================] - 8s 3ms/step - loss: 11.2588 -
val_loss: 1.6531
Epoch 16/50
2500/2500 [==============================] - 6s 2ms/step - loss: 3.6358 -
val_loss: 1.6450
Epoch 17/50
2500/2500 [==============================] - 8s 3ms/step - loss: 6.9054 -
val_loss: 12.1671
Epoch 18/50
2500/2500 [==============================] - 6s 3ms/step - loss: 3.5173 -
val_loss: 1.6199
Epoch 19/50
2500/2500 [==============================] - 7s 3ms/step - loss: 10.1462 -
val_loss: 1.6058
Epoch 20/50
2500/2500 [==============================] - 6s 2ms/step - loss: 9.6839 -
val_loss: 1.5904
Epoch 21/50
2500/2500 [==============================] - 8s 3ms/step - loss: 6.6740 -
val_loss: 1.6174
Epoch 22/50
2500/2500 [==============================] - 6s 3ms/step - loss: 2.9106 -
val_loss: 1.5789
Epoch 23/50
2500/2500 [==============================] - 7s 3ms/step - loss: 5.7660 -
val_loss: 1.5735
Epoch 24/50
2500/2500 [==============================] - 6s 2ms/step - loss: 5.7230 -
val_loss: 1.5712
Epoch 25/50
2500/2500 [==============================] - 9s 4ms/step - loss: 4.8066 -
val_loss: 2.2808
Epoch 26/50
2500/2500 [==============================] - 7s 3ms/step - loss: 3.7358 -
val_loss: 1.7014
Epoch 27/50
2500/2500 [==============================] - 6s 3ms/step - loss: 7.2950 -
val_loss: 1.5342
Epoch 28/50
2500/2500 [==============================] - 7s 3ms/step - loss: 5.0996 -
val_loss: 1.5214
Epoch 29/50
2500/2500 [==============================] - 7s 3ms/step - loss: 7.8400 -
val_loss: 1.4995
Epoch 30/50
2500/2500 [==============================] - 7s 3ms/step - loss: 2.5595 -
val_loss: 1.4965
```

```
Epoch 31/50
2500/2500 [==============================] - 6s 2ms/step - loss: 4.3813 -
val_loss: 1.4671
Epoch 32/50
2500/2500 [==============================] - 6s 2ms/step - loss: 8.3217 -
val_loss: 1.4599
Epoch 33/50
2500/2500 [==============================] - 6s 3ms/step - loss: 4.3808 -
val_loss: 1.4488
Epoch 34/50
2500/2500 [==============================] - 6s 3ms/step - loss: 4.7826 -
val_loss: 1.4364
Epoch 35/50
2500/2500 [==============================] - 7s 3ms/step - loss: 3.1755 -
val_loss: 1.4297
Epoch 36/50
2500/2500 [==============================] - 6s 2ms/step - loss: 4.1724 -
val_loss: 1.4193
Epoch 37/50
2500/2500 [==============================] - 7s 3ms/step - loss: 3.2009 -
val_loss: 1.4115
Epoch 38/50
2500/2500 [==============================] - 6s 2ms/step - loss: 3.7699 -
val_loss: 1.6326
Epoch 39/50
2500/2500 [==============================] - 7s 3ms/step - loss: 4.1135 -
val_loss: 17.8680
Epoch 40/50
2500/2500 [==============================] - 6s 3ms/step - loss: 4.2767 -
val_loss: 1.3954
Epoch 41/50
2500/2500 [==============================] - 8s 3ms/step - loss: 2.8079 -
val_loss: 1.3907
Epoch 42/50
2500/2500 [==============================] - 6s 3ms/step - loss: 4.3047 -
val_loss: 1.3879
Epoch 43/50
2500/2500 [==============================] - 8s 3ms/step - loss: 5.5976 -
val_loss: 1.3861
Epoch 44/50
2500/2500 [==============================] - 6s 2ms/step - loss: 2.6621 -
val_loss: 1.3811
Epoch 45/50
2500/2500 [==============================] - 7s 3ms/step - loss: 4.9479 -
val_loss: 1.3831
Epoch 46/50
2500/2500 [==============================] - 6s 2ms/step - loss: 2.6820 -
val_loss: 1.3811
```

```
Epoch 47/50
2500/2500 [==============================] - 8s 3ms/step - loss: 4.2590 -
val_loss: 1.3806
Epoch 48/50
2500/2500 [==============================] - 6s 2ms/step - loss: 3.0875 -
val_loss: 1.7594
Epoch 49/50
2500/2500 [==============================] - 8s 3ms/step - loss: 3.3513 -
val_loss: 1.3767
Epoch 50/50
2500/2500 [==============================] - 6s 2ms/step - loss: 3.2688 -
val_loss: 1.3785
```

```python
[ ]: labels = autoencoder.predict(X)
     mse = np.mean(np.power(X - labels, 2), axis=1)
```

```
3125/3125 [==============================] - 5s 2ms/step
```

```python
[ ]: threshold = np.percentile(mse, 99)
     anomalies = mse > threshold

     print(f"Threshold: {threshold}")
     print(f"Number of anomalies detected: {np.sum(anomalies)}")
```

```
Threshold: 3.242576424762769
Number of anomalies detected: 1000
```

```python
[ ]: df_org['anomaly'] = anomalies
```

```python
[ ]: df_org['anomaly'] = df_org['anomaly'].apply(lambda x: 1 if x == True else 0)
```

```python
[ ]: df['anomaly'] = df_org['anomaly']
```

```python
[ ]: df_org.head()
```

```
[ ]:      index  National Provider Identifier  \
     0  8774979                    1891106191
     1  3354385                    1346202256
     2  3001884                    1306820956
     3  7594822                    1770523540
     4   746159                    1073627758

       Last Name/Organization Name of the Provider First Name of the Provider  \
     0                                 UPADHYAYULA                   SATYASREE
     1                                       JONES                       WENDY
     2                                    DUROCHER                     RICHARD
     3                                     FULLARD                      JASPER
     4                                    PERROTTI                     ANTHONY
```

```
   Middle Initial of the Provider Credentials of the Provider  \
0                            NaN                          M.D.
1                              P                          M.D.
2                              W                           DPM
3                            NaN                            MD
4                              E                            DO

  Gender of the Provider Entity Type of the Provider  \
0                      F                            I
1                      F                            I
2                      M                            I
3                      M                            I
4                      M                            I

  Street Address 1 of the Provider Street Address 2 of the Provider  …  \
0                 1402 S GRAND BLVD                   FDT 14TH FLOOR  …
1                  2950 VILLAGE DR                              NaN  …
2                20 WASHINGTON AVE                          STE 212  …
3                5746 N BROADWAY ST                             NaN  …
4                  875 MILITARY TRL                       SUITE 200  …

  HCPCS Drug Indicator  Number of Services Number of Medicare Beneficiaries  \
0                    N                   27                               24
1                    N                  175                              175
2                    N                   32                               13
3                    N                   20                               18
4                    N                   33                               24

  Number of Distinct Medicare Beneficiary/Per Day Services  \
0                                                 27
1                                                175
2                                                 32
3                                                 20
4                                                 31

  Average Medicare Allowed Amount Average Submitted Charge Amount  \
0                     200.58777778                     305.21111111
1                           123.73                            548.8
2                            90.65                              155
3                              3.5                                5
4                            26.52                               40

  Average Medicare Payment Amount Average Medicare Standardized Amount  \
0                     157.26222222                          160.90888889
1                           118.83                           135.31525714
2                        64.4396875                            60.5959375
```

```
3                           3.43                                          3.43
4                  19.539393939                                  19.057575758

   Money difference  anomaly
0        147.948889        0
1        429.970000        0
2         90.560312        0
3          1.570000        0
4         20.460606        0

[5 rows x 29 columns]
```

[ ]: `list(df_org['anomaly']).count(0)`

[ ]: 99000

```python
inliers = df_org[df_org['anomaly'] == 0]
anomalies = df_org[df_org['anomaly'] == 1]

# Function to plot categorical feature distributions
def plot_categorical_feature(df_org, feature):
    fig, ax = plt.subplots(1, 2, figsize=(14, 6), sharey=True)

    # Plot for inliers
    inliers[feature].value_counts().plot(kind='bar', ax=ax[0], color='green',
 ↪alpha=0.6)
    ax[0].set_title(f'Inliers - {feature}')
    ax[0].set_ylabel('Count')

    # Plot for anomalies
    anomalies[feature].value_counts().plot(kind='bar', ax=ax[1], color='red',
 ↪alpha=0.6)
    ax[1].set_title(f'Anomalies - {feature}')

    plt.show()

# Plot the categorical features
plot_categorical_feature(df_org, 'Gender of the Provider')
plot_categorical_feature(df_org, 'Entity Type of the Provider')
plot_categorical_feature(df_org, 'Place of Service')
plot_categorical_feature(df_org, 'Medicare Participation Indicator')
plot_categorical_feature(df_org, 'HCPCS Drug Indicator')
```

Inliers - Gender of the Provider

Anomalies - Gender of the Provider

Inliers - Entity Type of the Provider

Anomalies - Entity Type of the Provider

Inliers - Place of Service

Anomalies - Place of Service

Inliers - Medicare Participation Indicator

Anomalies - Medicare Participation Indicator

Inliers - HCPCS Drug Indicator   Anomalies - HCPCS Drug Indicator
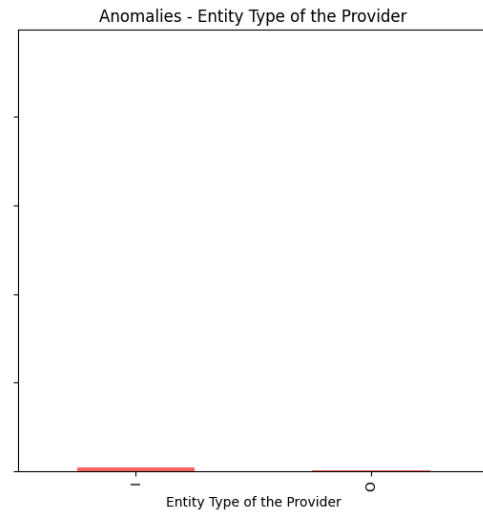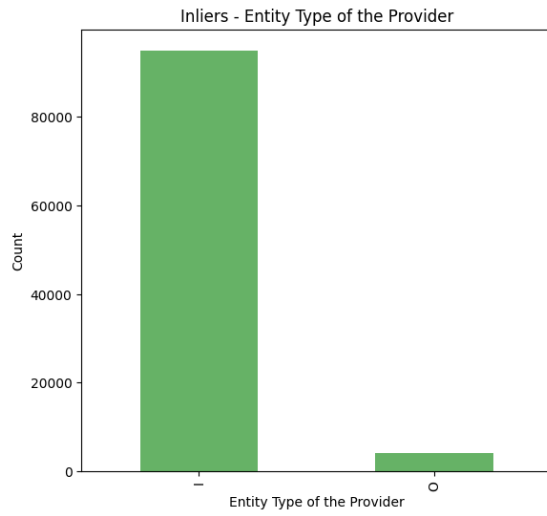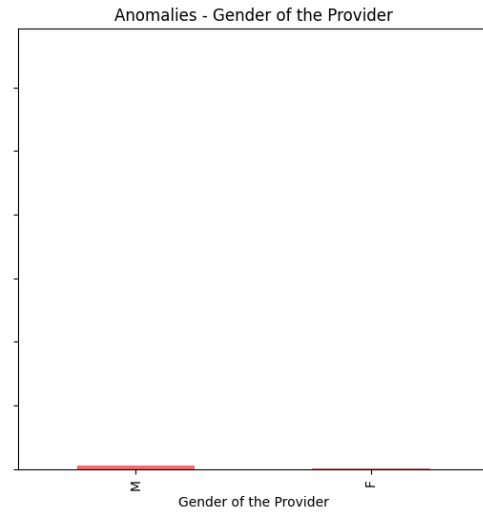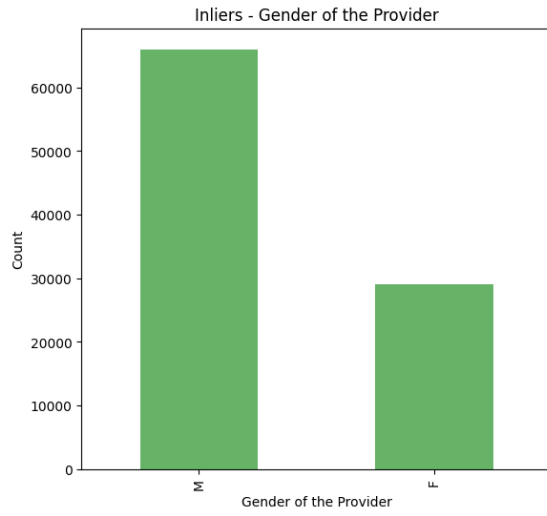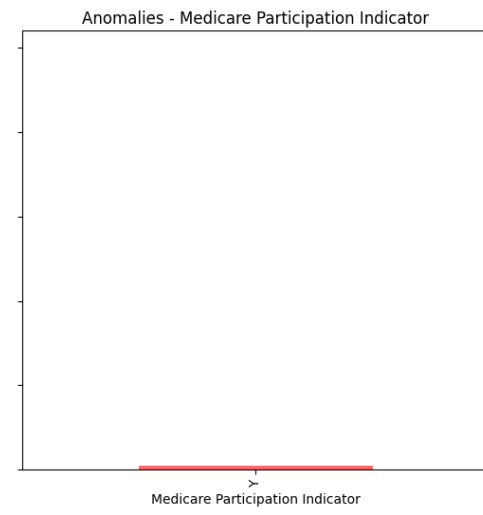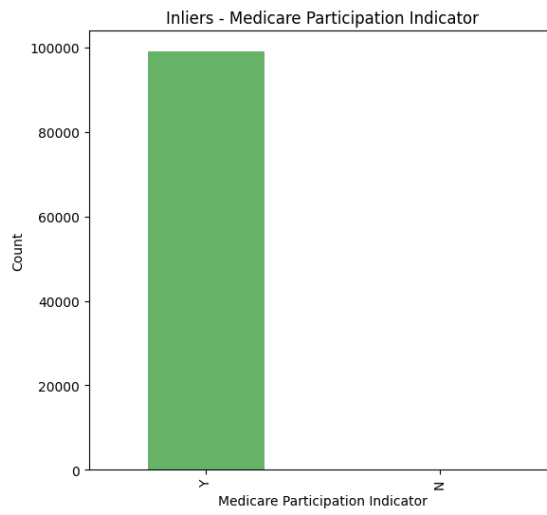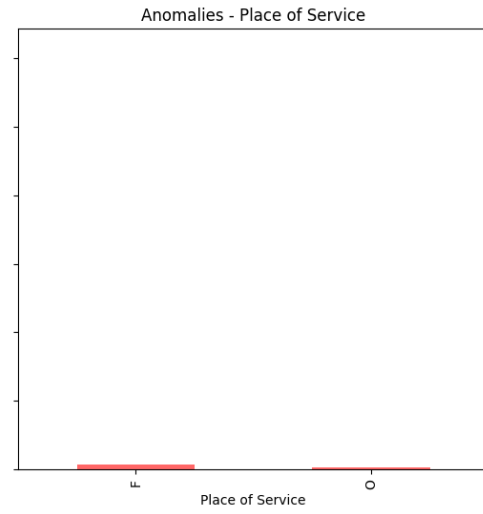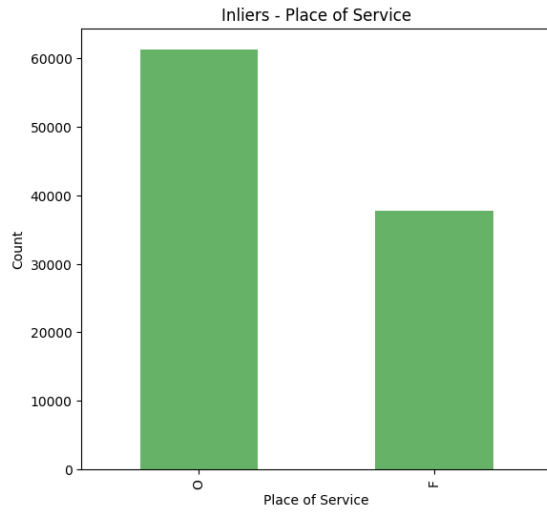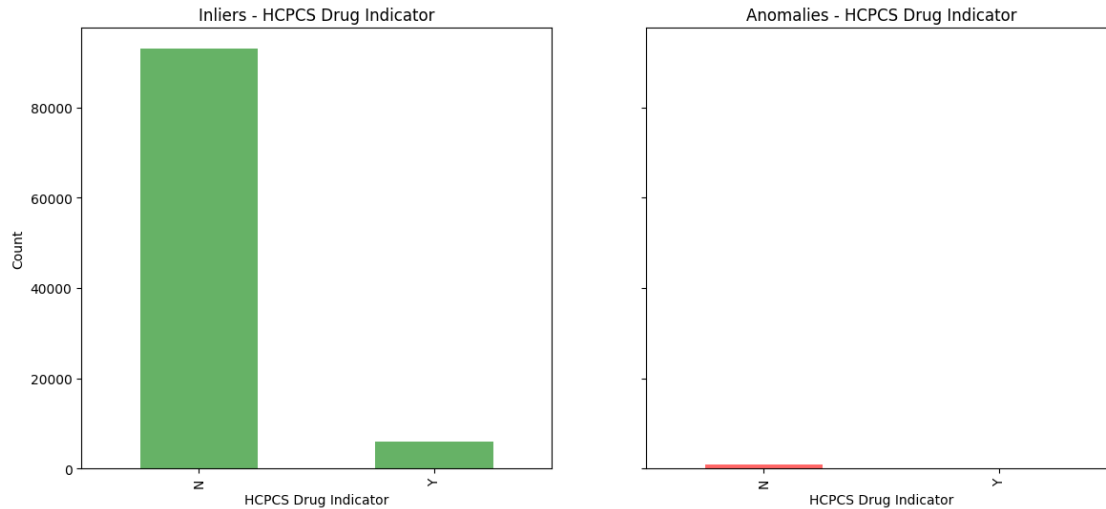
```
normal_data = df[df['anomaly'] == 0]
anomalous_data = df[df['anomaly'] == 1]
```

```
plt.scatter(normal_data['Average Submitted Charge Amount'],
 →normal_data['Average Medicare Payment Amount'], c='green', label='Normal')
plt.scatter(anomalous_data['Average Submitted Charge Amount'],
 →anomalous_data['Average Medicare Payment Amount'], c='red', label='Anomaly')
plt.xlabel('Average Submitted Charge Amount')
plt.ylabel('Average Medicare Payment Amount')
plt.title('Scatter Plot of Anomalies vs Normal Data')
plt.legend()
plt.show()
```
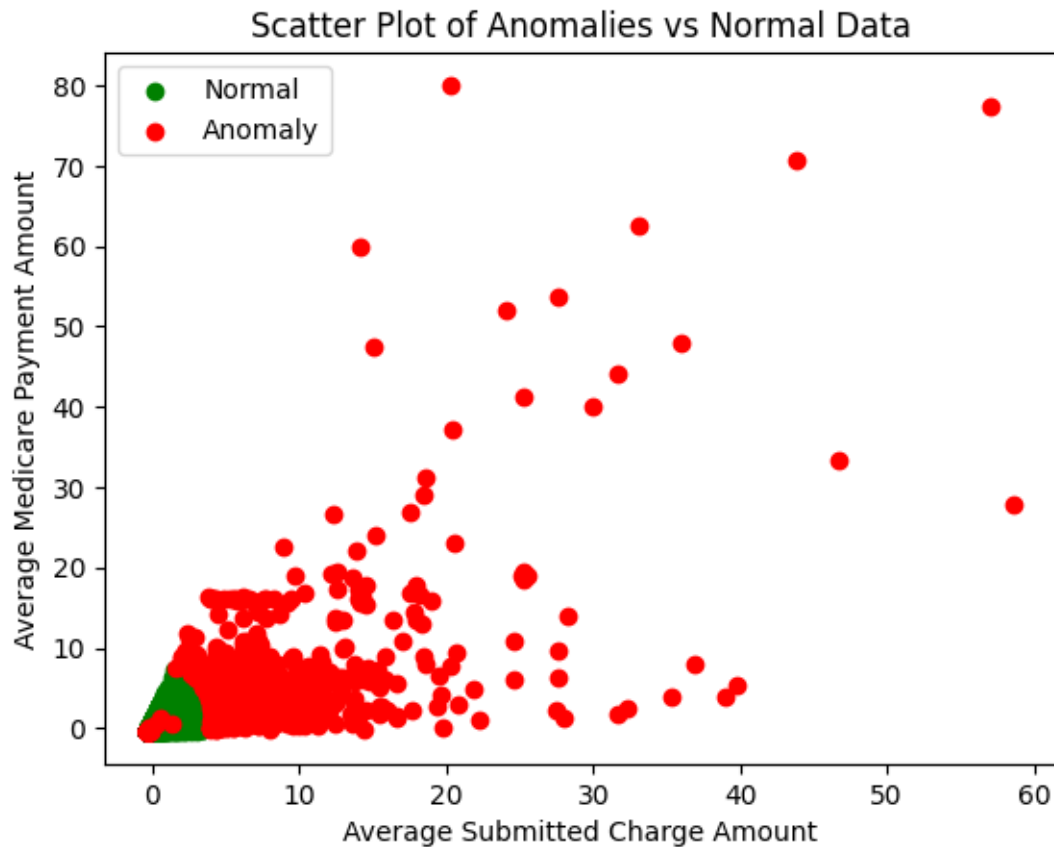
Scatter Plot of Anomalies vs Normal Data

```
plt.scatter(normal_data['Average Submitted Charge Amount'], normal_data['Money␣
 ↪difference'], c='green', label='Normal')
plt.scatter(anomalous_data['Average Submitted Charge Amount'],␣
 ↪anomalous_data['Money difference'], c='red', label='Anomaly')
plt.xlabel('Average Submitted Charge Amount')
plt.ylabel('Money difference')
plt.title('Scatter Plot of Anomalies vs Normal Data')
plt.legend()
plt.show()
```
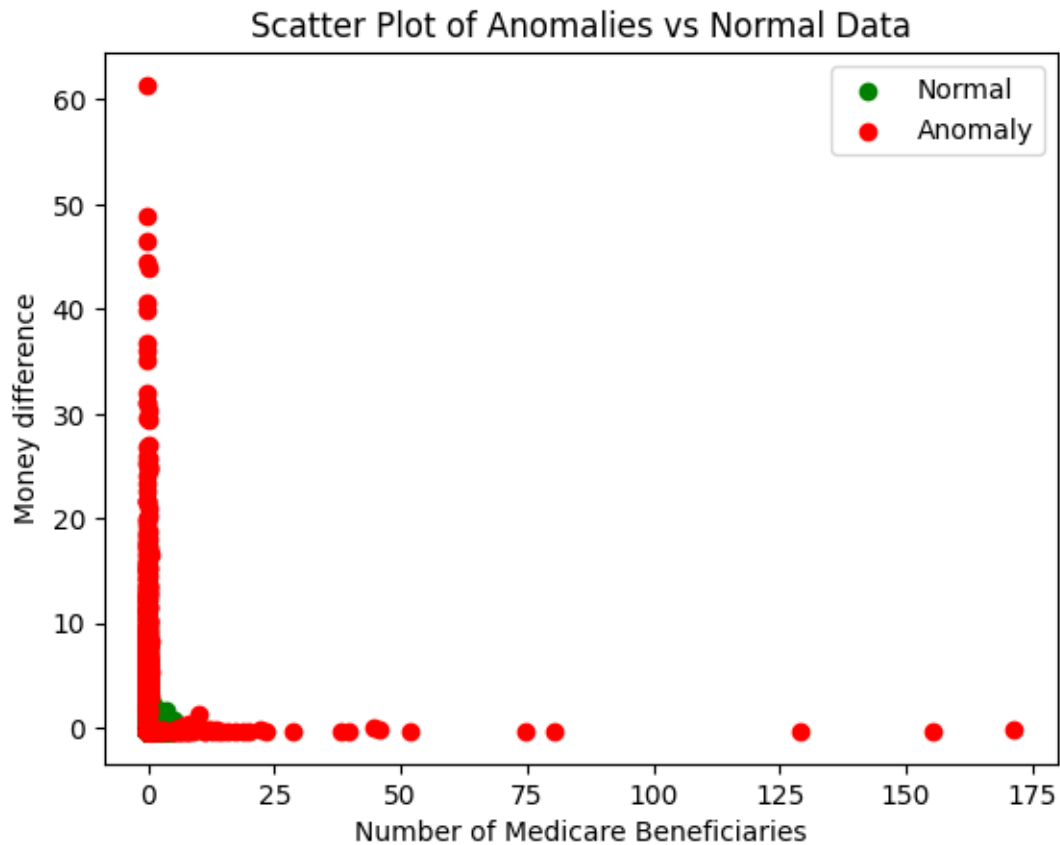
Scatter Plot of Anomalies vs Normal Data
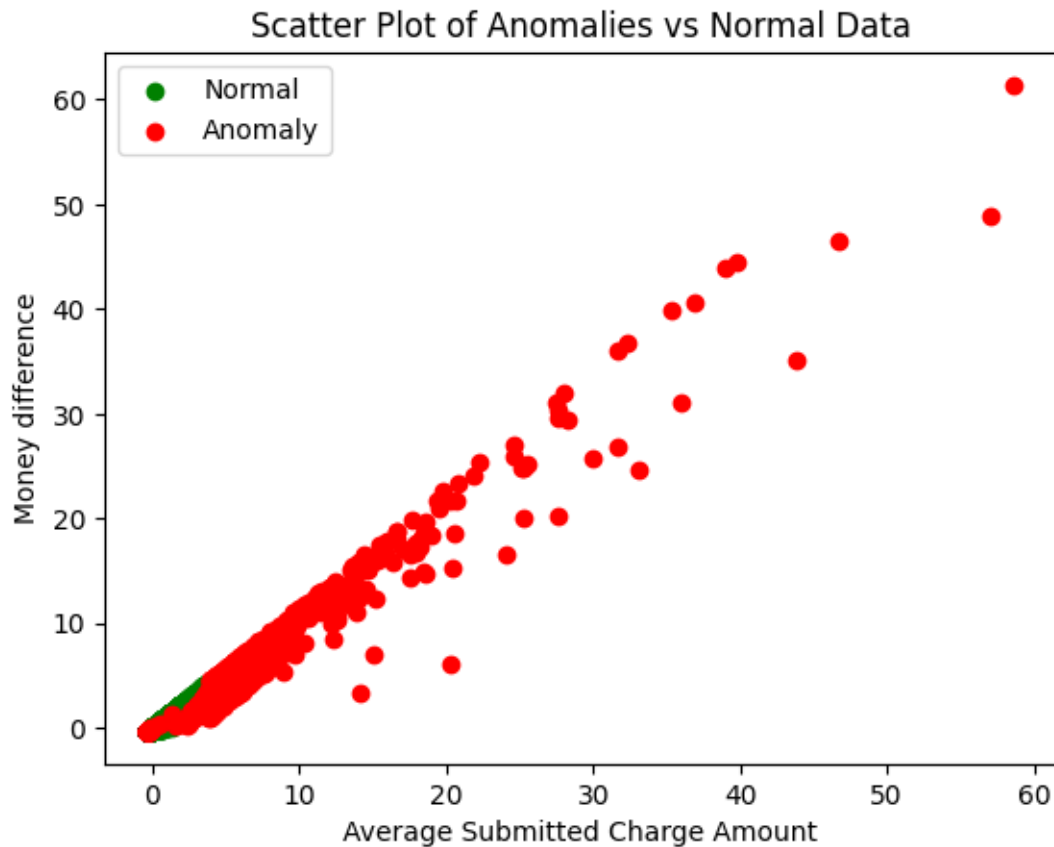
```
plt.scatter(normal_data['Average Medicare Payment Amount'], normal_data['Money␣
 ↪difference'], c='green', label='Normal')
plt.scatter(anomalous_data['Average Medicare Payment Amount'],␣
 ↪anomalous_data['Money difference'], c='red', label='Anomaly')
plt.xlabel('Average Medicare Payment Amount')
plt.ylabel('Money difference')
plt.title('Scatter Plot of Anomalies vs Normal Data')
plt.legend()
plt.show()
```

Scatter Plot of Anomalies vs Normal Data
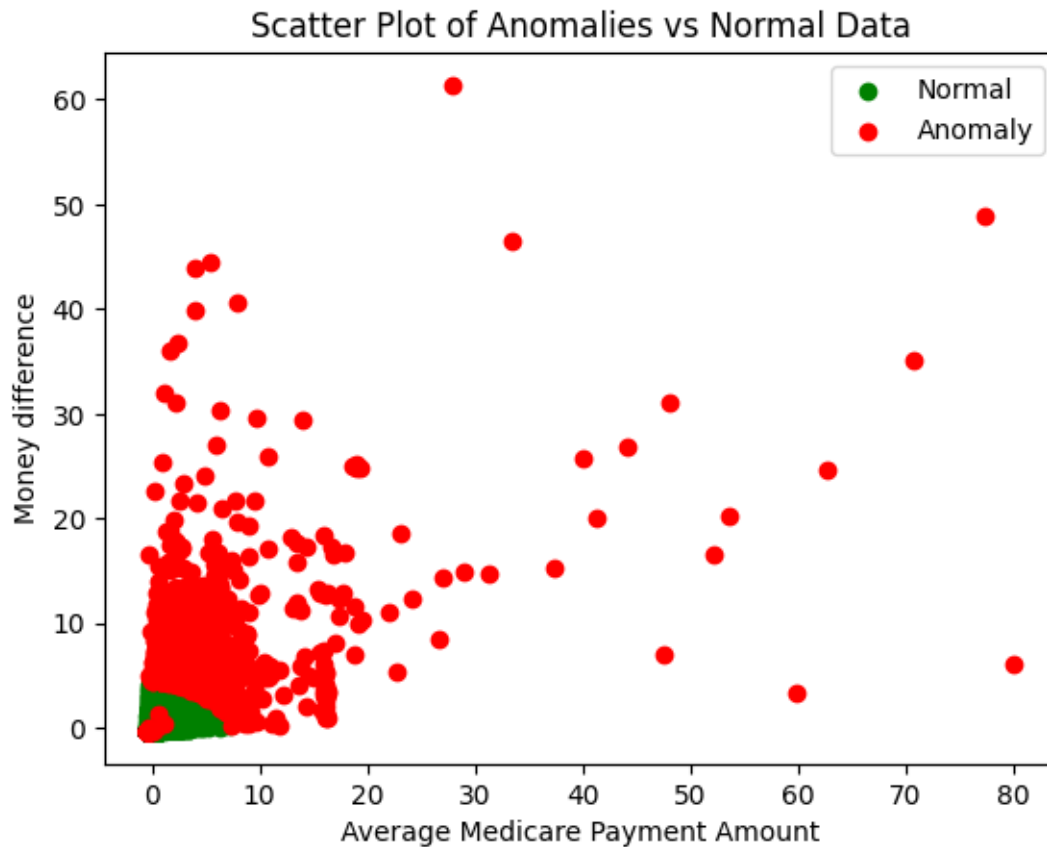
```python
plt.scatter(normal_data['Number of Services'], normal_data['Money difference'],
 ↪c='green', label='Normal')
plt.scatter(anomalous_data['Number of Services'], anomalous_data['Money
 ↪difference'], c='red', label='Anomaly')
plt.xlabel('Number of Services')
plt.ylabel('Money difference')
plt.title('Scatter Plot of Anomalies vs Normal Data')
plt.legend()
plt.show()
```

Scatter Plot of Anomalies vs Normal Data
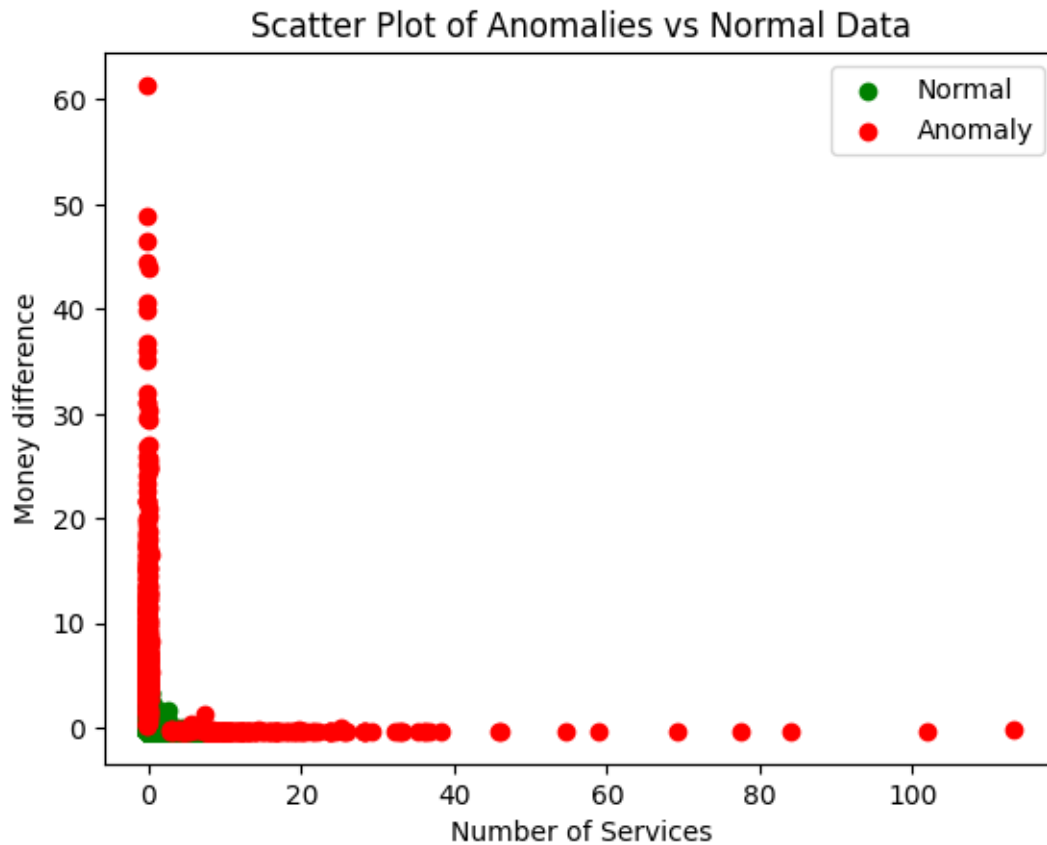
```
plt.scatter(normal_data['Number of Medicare Beneficiaries'], normal_data['Money␣
 ↪difference'], c='green', label='Normal')
plt.scatter(anomalous_data['Number of Medicare Beneficiaries'],␣
 ↪anomalous_data['Money difference'], c='red', label='Anomaly')
plt.xlabel('Number of Medicare Beneficiaries')
plt.ylabel('Money difference')
plt.title('Scatter Plot of Anomalies vs Normal Data')
plt.legend()
plt.show()
```

Scatter Plot of Anomalies vs Normal Data

```
plt.scatter(normal_data['Number of Distinct Medicare Beneficiary/Per Day␣
 ↪Services'], normal_data['Money difference'], c='green', label='Normal')
plt.scatter(anomalous_data['Number of Distinct Medicare Beneficiary/Per Day␣
 ↪Services'], anomalous_data['Money difference'], c='red', label='Anomaly')
plt.xlabel('Number of Distinct Medicare Beneficiary/Per Day Services')
plt.ylabel('Money difference')
plt.title('Scatter Plot of Anomalies vs Normal Data')
plt.legend()
plt.show()
```

Scatter Plot of Anomalies vs Normal Data