In [1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

In [2]:
```
data=pd.read_excel("D:/proposal-mehran-11/main-data.xlsx")
data
```

Out[2]:

JMBER	Year/Month	State	Contracting-Party	...	Sonog	PRT-Surgeon-Orga	PRT - Orga-Cons	PRT-Orga-Drug	PRT-Orga-Tool	PR-O
925084	1401/01	تهران	اسناد بیمارستانی	...	NaN	64208438.0	NaN	8213085.0	7090558.0	138850
6956-1	1401/12	زنجان	اسناد بیمارستانی	...	NaN	NaN	NaN	0.0	NaN	
083404	1401/06	تهران	اسناد بیمارستانی	...	NaN	NaN	NaN	NaN	3091500.0	
427688	1401/11	یزد	اسناد بیمارستانی	...	NaN	NaN	NaN	NaN	525492.0	
134454	1401/03	تهران	اسناد بیمارستانی	...	3313305.0	NaN	NaN	NaN	8656200.0	
...	...	...	...	...	...	...	...	...	...	
087160	1401/10	فارس	اسناد بیمارستانی	...	NaN	25434900.0	NaN	13742838.0	12475947.0	80717
005941	1401/05	قم	اسناد بیمارستانی	...	NaN	13530132.0	1475100.0	8590829.0	NaN	13572
115525	1401/07	آذربایجان شرقی	اسناد بیمارستانی	...	NaN	NaN	NaN	NaN	NaN	
695523	1401/03	آذربایجان شرقی	اسناد بیمارستانی	...	NaN	NaN	NaN	NaN	386253.0	
473748	1401/02	تهران	اسناد بیمارستانی	...	NaN	NaN	NaN	NaN	1771200.0	

In [4]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 41913 entries, 0 to 41912
Data columns (total 34 columns):
 #   Column              Non-Null Count  Dtype  
---  ------              --------------  -----  
 0   File-Type           41913 non-null  object 
 1   Day-of-Bedridden    41913 non-null  int64  
 2   Date-of-Admission   41913 non-null  object 
 3   Date-of-Release     41913 non-null  object 
 4   Dr.Name             41913 non-null  object 
 5   Dr.ID               41913 non-null  int64  
 6   CLAIM_NUMBER        41913 non-null  object 
 7   Year/Month          41913 non-null  object 
 8   State               41913 non-null  object 
 9   Contracting-Party   41913 non-null  object 
 10  Center-Name         41913 non-null  object 
 11  Owner               41913 non-null  object 
 12  Dr.Name2            13782 non-null  object 
 13  Bill-Type           41913 non-null  object 
 14  National ID         41913 non-null  int64  
 15  Age                 41913 non-null  int64  
 16  Cash-Desk           41913 non-null  object 
 17  Month               41913 non-null  int64  
 18  Gender              41913 non-null  object 
 19  Service-Name        15613 non-null  object 
 20  ServiceID           15613 non-null  object 
 21  NumBed/Day          16002 non-null  float64
 22  Num-Bed-ICU         3923 non-null   float64
 23  PRT-Hotl-Orga       16002 non-null  float64
 24  Sonog               5414 non-null   float64
 25  PRT-Surgeon-Orga    15613 non-null  float64
 26  PRT -Orga-Cons      9024 non-null   float64
 27  PRT-Orga-Drug       19269 non-null  float64
 28  PRT-Orga-Tool       21835 non-null  float64
 29  PR-Orga-Pthl        15305 non-null  float64
 30  Prot                41913 non-null  int64  
 31  PRT-Orga-Totl       41913 non-null  float64
 32  Obli-Insu-Totl      41913 non-null  float64
 33  Totl-Amou           41913 non-null  float64
dtypes: float64(12), int64(6), object(16)
memory usage: 10.9+ MB
```

In [5]:
```
data.describe()
```

Out[5]:

	Day-of-Bedridden	Dr.ID	National ID	Age	Month	NumBed/Day	Num-Bed-ICU	
count	41913.000000	4.191300e+04	4.191300e+04	41913.000000	41913.000000	16002.000000	3923.000000	1.6
mean	3.080619	1.998060e+08	4.106949e+09	50.635364	6.798177	3.626032	2.794290	1.7
std	4.804914	1.076292e+08	1.518453e+10	15.843874	3.352792	4.860545	4.787983	3.9
min	0.000000	1.520680e+05	3.903000e+04	1.000000	1.000000	1.000000	1.000000	0.0
25%	1.000000	8.134745e+07	1.501581e+09	39.000000	4.000000	1.000000	1.000000	3.0
50%	2.000000	2.645886e+08	2.870633e+09	50.000000	7.000000	2.000000	1.000000	6.1
75%	3.000000	2.711638e+08	4.432575e+09	62.000000	10.000000	4.000000	3.000000	1.8
max	429.000000	2.779699e+08	8.001981e+11	102.000000	12.000000	85.000000	85.000000	1.0

In [6]:
```
data["Target"]=1
i=0
#data['Target']= data['File Type']
for i in range(len(data)) :
    if data["File-Type"][i]=='glob':
        data["Target"][i]=0
```

```
C:\Users\Naeim\AppData\Local\Temp\ipykernel_9432\454407751.py:6: SettingWithCopyWarnin
g:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user
_guide/indexing.html#returning-a-view-versus-a-copy (https://pandas.pydata.org/pandas-d
ocs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)
  data["Target"][i]=0
```

In [7]:
```
data.head(10)
```

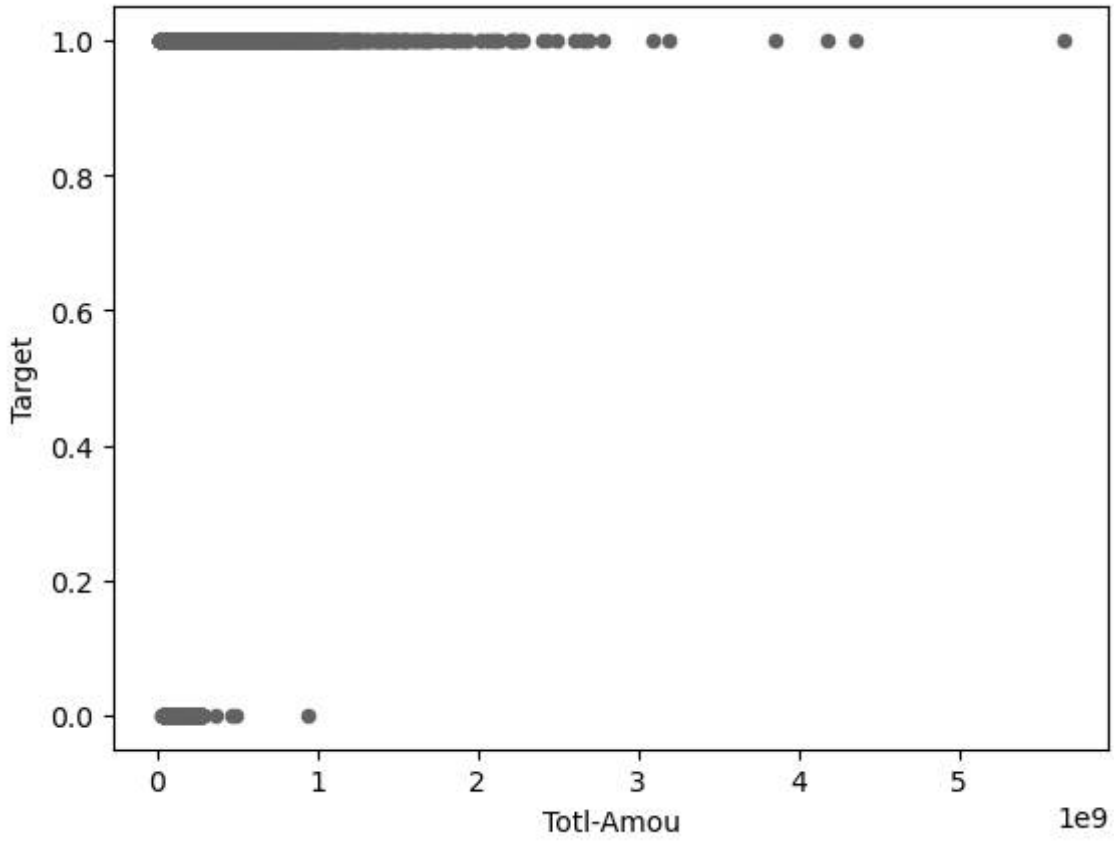
Out[7]:

	File-Type	Day-of-Bedridden	Date-of-Admission	Date-of-Release	Dr.Name	Dr.ID	CLAIM_NUMBER	Year/Month	State	Contr
0	not glob	5	1401/01/16	1401/01/21	حسام الدین اقلیمی	274674103	5925084	1401/01	تهران	ستانی
1	glob	1	1401/12/22	1401/12/23	حمیدرضا سلامت	270858572	U86956-1	1401/12	زنجان	ستانی
2	glob	5	1401/06/08	1401/06/13	حسام الدین اقلیمی	248153954	6083404	1401/06	تهران	ستانی
3	glob	1	1401/11/02	1401/11/03	علی اکبر سلمانی زارچ	269913913	1427688	1401/11	یزد	ستانی
4	glob	7	1401/03/19	1401/03/26	امیر زمانی	275471278	6134454	1401/03	تهران	ستانی
5	glob	3	1401/03/07	1401/03/10	سیدحمیدرضا حقیقی کاخکی	275762540	6228829	1401/03	خراسان رضوی	ستانی
6	glob	3	1401/03/23	1401/03/26	احسان عبدالهی منصورخانی	4502068	676656	1401/03	فارس	ستانی
7	glob	2	1401/09/12	1401/09/14	محمدکاظم شاهمرادی	249327067	60-35-37-1	1401/09	لرستان	ستانی
8	glob	3	1401/06/07	1401/06/10	فاطمه رامزی	118950934	2599255	1401/06	اردبیل	ستانی
9	glob	3	1401/04/07	1401/04/10	دلیر حمزه ئی	5490341	187803-2	1401/04	کردستان	ستانی

10 rows × 35 columns

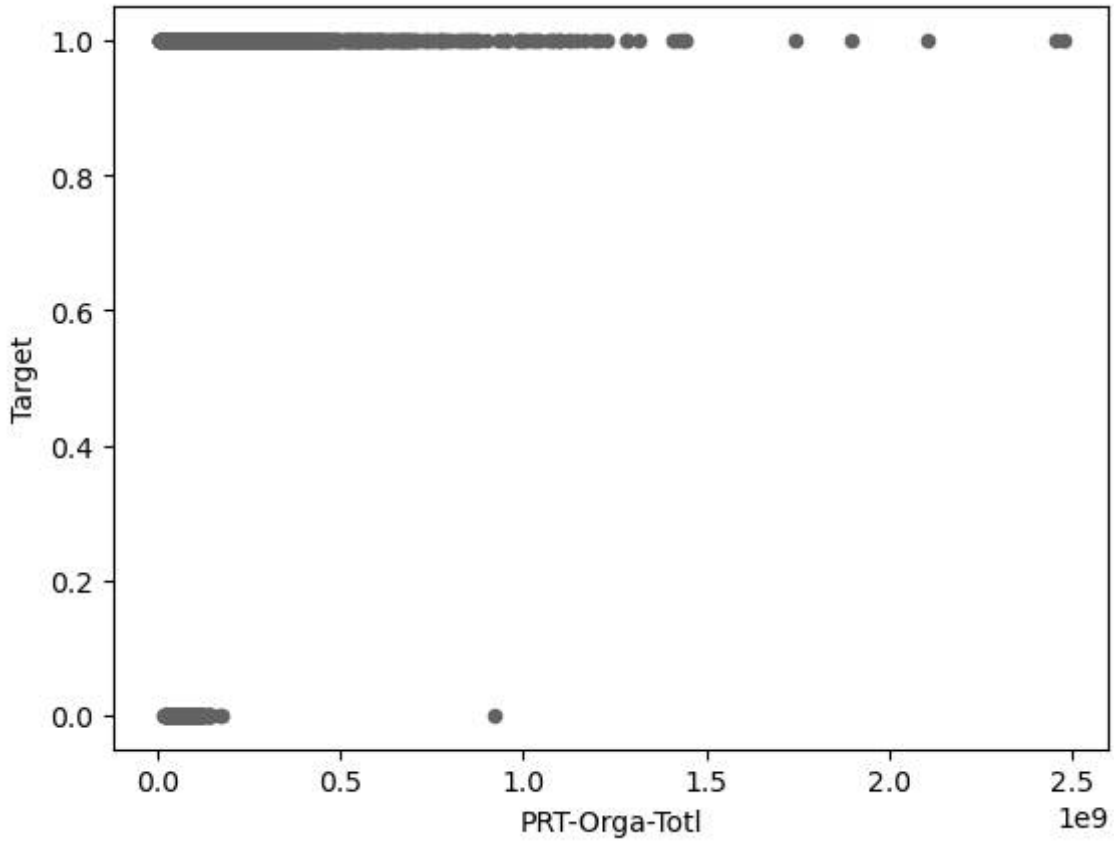
In [8]:
```
data.plot(kind='scatter',x="Totl-Amou",y="Target")

plt.show()
```



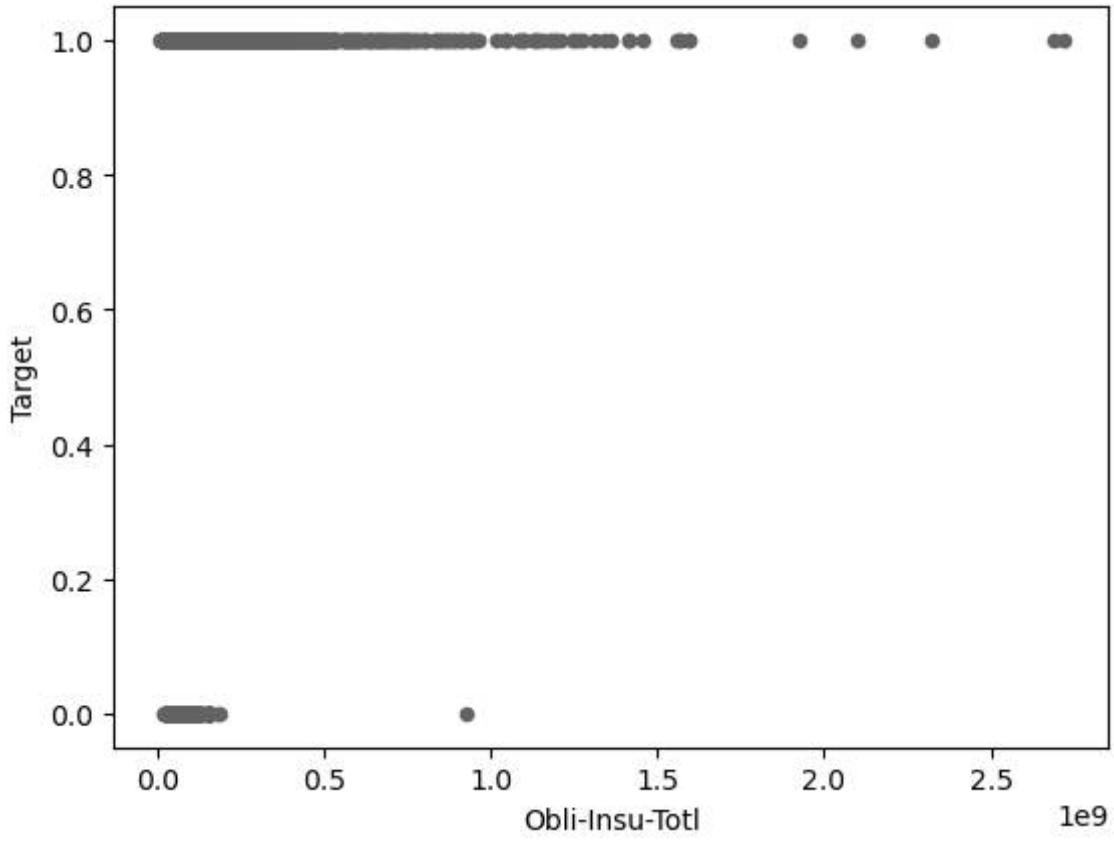
In [9]:
```
data.plot(kind='scatter',x="PRT-Orga-Totl",y="Target")

plt.show()
```



In [10]:
```
data.plot(kind='scatter',x="Obli-Insu-Totl",y="Target")

plt.show()
```



In [11]:
```
print(data.corr())

#Num of Bed/Day WITH Day of Bedridden=.67 ...Num Bed ICU WITH  Day of Bedridden=.58 ...P
#PRT-Orga-Drug WITH Day of Bedridden=.59...PR-Orga-Pthl WITH Day of Bedridden=.59...
#PRT-Orga-Totl VA Obli-Insu-Totl WITH Day of Bedridden=.55
```

	Day-of-Bedridden	Dr.ID	National ID	Age	Month	\
Day-of-Bedridden	1.000000	-0.009218	0.012013	0.116428	-0.006323	
Dr.ID	-0.009218	1.000000	0.006069	0.008867	0.145424	
National ID	0.012013	0.006069	1.000000	-0.006400	0.004055	
Age	0.116428	0.008867	-0.006400	1.000000	-0.013869	
Month	-0.006323	0.145424	0.004055	-0.013869	1.000000	
NumBed/Day	0.761217	-0.017061	0.027233	0.146424	-0.006071	
Num-Bed-ICU	0.583404	0.000924	0.024653	0.036584	0.000605	
PRT-Hotl-Orga	0.678752	-0.015568	0.015748	0.160627	0.004170	
Sonog	0.357045	-0.020690	0.021894	0.062047	0.046195	
PRT-Surgeon-Orga	0.405588	-0.024230	0.032778	0.010358	0.046639	
PRT -Orga-Cons	0.485676	-0.010602	0.039748	0.098417	0.018696	
PRT-Orga-Drug	0.596510	0.001631	0.007068	0.135836	-0.012031	
PRT-Orga-Tool	0.372345	0.023195	-0.005555	0.129993	0.104203	
PR-Orga-Pthl	0.599904	-0.024528	0.018126	0.092870	0.017999	
Prot	-0.014230	0.008522	-0.009915	0.013183	0.015404	
PRT-Orga-Totl	0.558261	-0.021939	0.015897	0.096647	0.061453	
Obli-Insu-Totl	0.559836	-0.021613	0.015530	0.099245	0.060927	
Totl-Amou	0.433479	0.008134	-0.004630	0.205897	0.067073	
Target	0.102707	0.002076	0.010208	0.200744	0.014574	

In [15]:
```
data_Noglob=data[(data["Target"]>0)]
#data_Noglob
data_Noglob2=data[(data["File-Type"]=='noglob')]
data_Noglob2.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 0 entries
Data columns (total 35 columns):
 #   Column              Non-Null Count  Dtype  
---  ------              --------------  -----  
 0   File-Type           0 non-null      object 
 1   Day-of-Bedridden    0 non-null      int64  
 2   Date-of-Admission   0 non-null      object 
 3   Date-of-Release     0 non-null      object 
 4   Dr.Name             0 non-null      object 
 5   Dr.ID               0 non-null      int64  
 6   CLAIM_NUMBER        0 non-null      object 
 7   Year/Month          0 non-null      object 
 8   State               0 non-null      object 
 9   Contracting-Party   0 non-null      object 
 10  Center-Name         0 non-null      object 
 11  Owner               0 non-null      object 
 12  Dr.Name2            0 non-null      object 
 13  Bill-Type           0 non-null      object 
 14  National ID         0 non-null      int64  
 15  Age                 0 non-null      int64  
 16  Cash-Desk           0 non-null      object 
 17  Month               0 non-null      int64  
 18  Gender              0 non-null      object 
 19  Service-Name        0 non-null      object 
 20  ServiceID           0 non-null      object 
 21  NumBed/Day          0 non-null      float64
 22  Num-Bed-ICU         0 non-null      float64
 23  PRT-Hotl-Orga       0 non-null      float64
 24  Sonog               0 non-null      float64
 25  PRT-Surgeon-Orga    0 non-null      float64
 26  PRT -Orga-Cons      0 non-null      float64
 27  PRT-Orga-Drug       0 non-null      float64
 28  PRT-Orga-Tool       0 non-null      float64
 29  PR-Orga-Pthl        0 non-null      float64
 30  Prot                0 non-null      int64  
 31  PRT-Orga-Totl       0 non-null      float64
 32  Obli-Insu-Totl      0 non-null      float64
 33  Totl-Amou           0 non-null      float64
 34  Target              0 non-null      int64  
dtypes: float64(12), int64(7), object(16)
memory usage: 0.0+ bytes
```

In [16]:
```
data_Noglob.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 15603 entries, 0 to 41909
Data columns (total 35 columns):
 #   Column              Non-Null Count  Dtype  
---  ------              --------------  -----  
 0   File-Type           15603 non-null  object 
 1   Day-of-Bedridden    15603 non-null  int64  
 2   Date-of-Admission   15603 non-null  object 
 3   Date-of-Release     15603 non-null  object 
 4   Dr.Name             15603 non-null  object 
 5   Dr.ID               15603 non-null  int64  
 6   CLAIM_NUMBER        15603 non-null  object 
 7   Year/Month          15603 non-null  object 
 8   State               15603 non-null  object 
 9   Contracting-Party   15603 non-null  object 
 10  Center-Name         15603 non-null  object 
 11  Owner               15603 non-null  object 
 12  Dr.Name2            13771 non-null  object 
 13  Bill-Type           15603 non-null  object 
 14  National ID         15603 non-null  int64  
 15  Age                 15603 non-null  int64  
 16  Cash-Desk           15603 non-null  object 
 17  Month               15603 non-null  int64  
 18  Gender              15603 non-null  object 
 19  Service-Name        15602 non-null  object 
 20  ServiceID           15602 non-null  object 
 21  NumBed/Day          15083 non-null  float64
 22  Num-Bed-ICU         3784 non-null   float64
 23  PRT-Hotl-Orga       15083 non-null  float64
 24  Sonog               3848 non-null   float64
 25  PRT-Surgeon-Orga    15602 non-null  float64
 26  PRT -Orga-Cons      8715 non-null   float64
 27  PRT-Orga-Drug       15509 non-null  float64
 28  PRT-Orga-Tool       15312 non-null  float64
 29  PR-Orga-Pthl        14899 non-null  float64
 30  Prot                15603 non-null  int64  
 31  PRT-Orga-Totl       15603 non-null  float64
 32  Obli-Insu-Totl      15603 non-null  float64
 33  Totl-Amou           15603 non-null  float64
 34  Target              15603 non-null  int64  
dtypes: float64(12), int64(7), object(16)
memory usage: 4.3+ MB
```

In [17]:
```
data_Noglob.describe()
```

Out[17]:

	Day-of-Bedridden	Dr.ID	National ID	Age	Month	NumBed/Day	Num-Bed-ICU	
count	15603.000000	1.560300e+04	1.560300e+04	15603.00000	15603.000000	15083.000000	3784.000000	1.5(
mean	3.727681	1.993761e+08	3.903700e+09	56.82279	6.861629	3.724906	2.842759	1.85
std	6.196552	1.079837e+08	1.705416e+10	16.04230	3.343170	4.980971	4.864417	4.02
min	0.000000	1.520680e+05	3.903000e+04	1.00000	1.000000	1.000000	1.000000	0.0(
25%	1.000000	7.341807e+07	1.287854e+09	47.00000	4.000000	1.000000	1.000000	3.06
50%	2.000000	2.666520e+08	2.659202e+09	58.00000	7.000000	2.000000	1.000000	6.13
75%	4.000000	2.714244e+08	4.284875e+09	68.00000	10.000000	4.000000	3.000000	2.03
max	429.000000	2.779699e+08	8.001981e+11	102.00000	12.000000	85.000000	85.000000	1.04

In [18]:
```
data_Noglob.to_excel("D:/proposal-mehran-11/data_Noglob.xlsx")
```

In [19]:
```
data_glob=data[(data["Target"]==0)]
data_glob2=data[(data["File-Type"]=='glob')]
data_glob2.info()
#data_glob
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 26310 entries, 1 to 41912
Data columns (total 35 columns):
 #   Column              Non-Null Count  Dtype  
---  ------              --------------  -----  
 0   File-Type           26310 non-null  object 
 1   Day-of-Bedridden    26310 non-null  int64  
 2   Date-of-Admission   26310 non-null  object 
 3   Date-of-Release     26310 non-null  object 
 4   Dr.Name             26310 non-null  object 
 5   Dr.ID               26310 non-null  int64  
 6   CLAIM_NUMBER        26310 non-null  object 
 7   Year/Month          26310 non-null  object 
 8   State               26310 non-null  object 
 9   Contracting-Party   26310 non-null  object 
 10  Center-Name         26310 non-null  object 
 11  Owner               26310 non-null  object 
 12  Dr.Name2            11 non-null     object 
 13  Bill-Type           26310 non-null  object 
 14  National ID         26310 non-null  int64  
 15  Age                 26310 non-null  int64  
 16  Cash-Desk           26310 non-null  object 
 17  Month               26310 non-null  int64  
 18  Gender              26310 non-null  object 
 19  Service-Name        11 non-null     object 
 20  ServiceID           11 non-null     object 
 21  NumBed/Day          919 non-null    float64
 22  Num-Bed-ICU         139 non-null    float64
 23  PRT-Hotl-Orga       919 non-null    float64
 24  Sonog               1566 non-null   float64
 25  PRT-Surgeon-Orga    11 non-null     float64
 26  PRT -Orga-Cons      309 non-null    float64
 27  PRT-Orga-Drug       3760 non-null   float64
 28  PRT-Orga-Tool       6523 non-null   float64
 29  PR-Orga-Pthl        406 non-null    float64
 30  Prot                26310 non-null  int64  
 31  PRT-Orga-Totl       26310 non-null  float64
 32  Obli-Insu-Totl      26310 non-null  float64
 33  Totl-Amou           26310 non-null  float64
 34  Target              26310 non-null  int64  
dtypes: float64(12), int64(7), object(16)
memory usage: 7.2+ MB
```

In [20]:
```
data_glob.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 26310 entries, 1 to 41912
Data columns (total 35 columns):
 #   Column              Non-Null Count  Dtype  
---  ------              --------------  -----  
 0   File-Type           26310 non-null  object 
 1   Day-of-Bedridden    26310 non-null  int64  
 2   Date-of-Admission   26310 non-null  object 
 3   Date-of-Release     26310 non-null  object 
 4   Dr.Name             26310 non-null  object 
 5   Dr.ID               26310 non-null  int64  
 6   CLAIM_NUMBER        26310 non-null  object 
 7   Year/Month          26310 non-null  object 
 8   State               26310 non-null  object 
 9   Contracting-Party   26310 non-null  object 
 10  Center-Name         26310 non-null  object 
 11  Owner               26310 non-null  object 
 12  Dr.Name2            11 non-null     object 
 13  Bill-Type           26310 non-null  object 
 14  National ID         26310 non-null  int64  
 15  Age                 26310 non-null  int64  
 16  Cash-Desk           26310 non-null  object 
 17  Month               26310 non-null  int64  
 18  Gender              26310 non-null  object 
 19  Service-Name        11 non-null     object 
 20  ServiceID           11 non-null     object 
 21  NumBed/Day          919 non-null    float64
 22  Num-Bed-ICU         139 non-null    float64
 23  PRT-Hotl-Orga       919 non-null    float64
 24  Sonog               1566 non-null   float64
 25  PRT-Surgeon-Orga    11 non-null     float64
 26  PRT -Orga-Cons      309 non-null    float64
 27  PRT-Orga-Drug       3760 non-null   float64
 28  PRT-Orga-Tool       6523 non-null   float64
 29  PR-Orga-Pthl        406 non-null    float64
 30  Prot                26310 non-null  int64  
 31  PRT-Orga-Totl       26310 non-null  float64
 32  Obli-Insu-Totl      26310 non-null  float64
 33  Totl-Amou           26310 non-null  float64
 34  Target              26310 non-null  int64  
dtypes: float64(12), int64(7), object(16)
memory usage: 7.2+ MB
```

In [21]:
```
data_glob.describe()
```

Out[21]:

	Day-of-Bedridden	Dr.ID	National ID	Age	Month	NumBed/Day	Num-Bed-ICU	
count	26310.000000	2.631000e+04	2.631000e+04	26310.000000	26310.000000	919.000000	139.000000	9.19
mean	2.696883	2.000609e+08	4.227484e+09	46.965945	6.760547	2.003264	1.474820	1.51
std	3.689607	1.074196e+08	1.395691e+10	14.529814	3.357982	1.175731	1.023853	1.52
min	0.000000	1.563170e+05	1.006036e+07	1.000000	1.000000	1.000000	1.000000	0.00
25%	2.000000	8.180023e+07	1.602659e+09	36.000000	4.000000	1.000000	1.000000	0.00
50%	2.000000	2.629895e+08	2.993629e+09	46.000000	7.000000	2.000000	1.000000	0.00
75%	3.000000	2.709714e+08	4.539995e+09	57.000000	10.000000	2.000000	2.000000	0.00
max	190.000000	2.773739e+08	5.434532e+11	102.000000	12.000000	15.000000	10.000000	2.84

In [20]:
```
data_glob.to_excel("D:/proposal-mehran-11/data_glob.xlsx")
```

In [22]:
```
data_Noglob["Totl-Amou"].mean()
```

Out[22]: 193461376.94163942

In [23]:
```
data_Noglob["PRT-Orga-Totl"].mean()
```

Out[23]: 74773442.48449016

In [24]:
```
data_glob["Totl-Amou"].mean()
```

Out[24]: 87331162.58928165

In [25]:
```
data_glob["PRT-Orga-Totl"].mean()
```

Out[25]: 66158453.61337894

In [ ]:
```

```