In [2]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

In [3]:
```
dataset = pd.read_csv(r'C:\Users\aishw\OneDrive\Desktop\global\global terrorism data
```

```
C:\ProgramData\Anaconda3\lib\site-packages\IPython\core\interactiveshell.py:3146: Dt
ypeWarning: Columns (4,6,31,33,61,62,63,76,79,90,92,94,96,114,115,121) have mixed ty
pes.Specify dtype option on import or set low_memory=False.
  has_raised = await self.run_ast_nodes(code_ast.body, cell_name,
```

In [4]:
```
dataset.head()
```

Out[4]:

	eventid	iyear	imonth	iday	approxdate	extended	resolution	country	country_txt	regi
0	197000000001	1970	7	2	NaN	0	NaN	58	Dominican Republic	
1	197000000002	1970	0	0	NaN	0	NaN	130	Mexico	
2	197001000001	1970	1	0	NaN	0	NaN	160	Philippines	
3	197001000002	1970	1	0	NaN	0	NaN	78	Greece	
4	197001000003	1970	1	0	NaN	0	NaN	101	Japan	

5 rows × 135 columns

Missing Values

In [5]:
```
features_with_na=[features for features in dataset.columns if dataset[features].isnu

for feature in features_with_na:
    print(feature, np.round(dataset[feature].isnull().mean(), 4),  ' % missing value
```

```
approxdate 0.9491  % missing values
resolution 0.9878  % missing values
provstate 0.0023  % missing values
city 0.0024  % missing values
latitude 0.0251  % missing values
longitude 0.0251  % missing values
specificity 0.0  % missing values
location 0.6946  % missing values
summary 0.364  % missing values
alternative 0.8403  % missing values
alternative_txt 0.8403  % missing values
attacktype2 0.9652  % missing values
attacktype2_txt 0.9652  % missing values
attacktype3 0.9976  % missing values
attacktype3_txt 0.9976  % missing values
targsubtype1 0.0571  % missing values
targsubtype1_txt 0.0571  % missing values
corp1 0.2342  % missing values
target1 0.0035  % missing values
natlty1 0.0086  % missing values
natlty1_txt 0.0086  % missing values
targtype2 0.9387  % missing values
targtype2_txt 0.9387  % missing values
targsubtype2 0.9412  % missing values
targsubtype2_txt 0.9412  % missing values
corp2 0.9443  % missing values
```

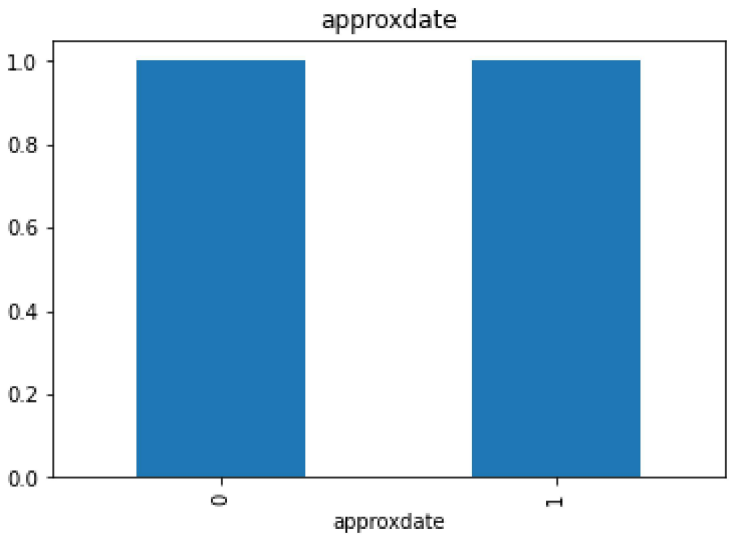
```
target2 0.9393  % missing values
natlty2 0.9404  % missing values
natlty2_txt 0.9404  % missing values
targtype3 0.9935  % missing values
targtype3_txt 0.9935  % missing values
targsubtype3 0.994  % missing values
targsubtype3_txt 0.994  % missing values
corp3 0.9944  % missing values
target3 0.9935  % missing values
natlty3 0.9937  % missing values
natlty3_txt 0.9937  % missing values
gsubname 0.9676  % missing values
gname2 0.9889  % missing values
gsubname2 0.9991  % missing values
gname3 0.9982  % missing values
gsubname3 0.9999  % missing values
motive 0.7217  % missing values
guncertain1 0.0021  % missing values
guncertain2 0.9892  % missing values
guncertain3 0.9982  % missing values
nperps 0.3914  % missing values
nperpcap 0.3825  % missing values
claimed 0.3639  % missing values
claimmode 0.895  % missing values
claimmode_txt 0.895  % missing values
claim2 0.9896  % missing values
claimmode2 0.9966  % missing values
claimmode2_txt 0.9966  % missing values
claim3 0.9982  % missing values
claimmode3 0.9993  % missing values
claimmode3_txt 0.9993  % missing values
compclaim 0.9734  % missing values
weapsubtype1 0.1143  % missing values
weapsubtype1_txt 0.1143  % missing values
weaptype2 0.9278  % missing values
weaptype2_txt 0.9278  % missing values
weapsubtype2 0.9365  % missing values
weapsubtype2_txt 0.9365  % missing values
weaptype3 0.9897  % missing values
weaptype3_txt 0.9897  % missing values
weapsubtype3 0.9907  % missing values
weapsubtype3_txt 0.9907  % missing values
weaptype4 0.9996  % missing values
weaptype4_txt 0.9996  % missing values
weapsubtype4 0.9996  % missing values
weapsubtype4_txt 0.9996  % missing values
weapdetail 0.3724  % missing values
nkill 0.0568  % missing values
nkillus 0.3547  % missing values
nkillter 0.3685  % missing values
nwound 0.0898  % missing values
nwoundus 0.3561  % missing values
nwoundte 0.3806  % missing values
propextent 0.6474  % missing values
propextent_txt 0.6474  % missing values
propvalue 0.7854  % missing values
propcomment 0.681  % missing values
ishostkid 0.001  % missing values
nhostkid 0.9253  % missing values
nhostkidus 0.9256  % missing values
nhours 0.9776  % missing values
ndays 0.9553  % missing values
divert 0.9982  % missing values
kidhijcountry 0.9818  % missing values
ransom 0.5741  % missing values
ransomamt 0.9926  % missing values
ransomamtus 0.9969  % missing values
ransompaid 0.9957  % missing values
ransompaidus 0.997  % missing values
```

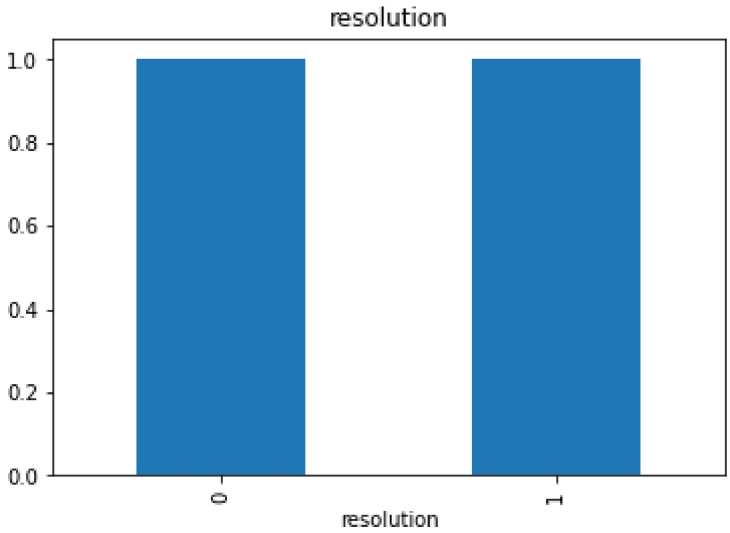
```
ransomnote 0.9972  % missing values
hostkidoutcome 0.9395  % missing values
hostkidoutcome_txt 0.9395  % missing values
nreleased 0.9428  % missing values
addnotes 0.8443  % missing values
scite1 0.3643  % missing values
scite2 0.5766  % missing values
scite3 0.7605  % missing values
related 0.8622  % missing values
```

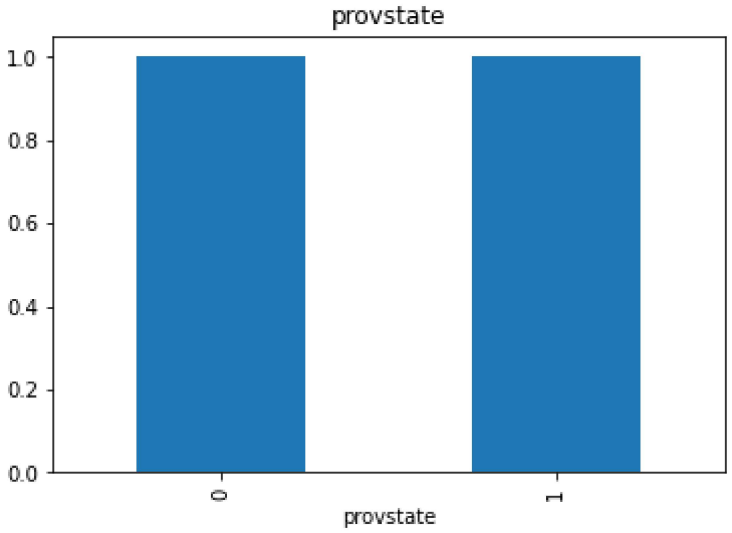
In [6]:
```
dataset.success.value_counts()
```

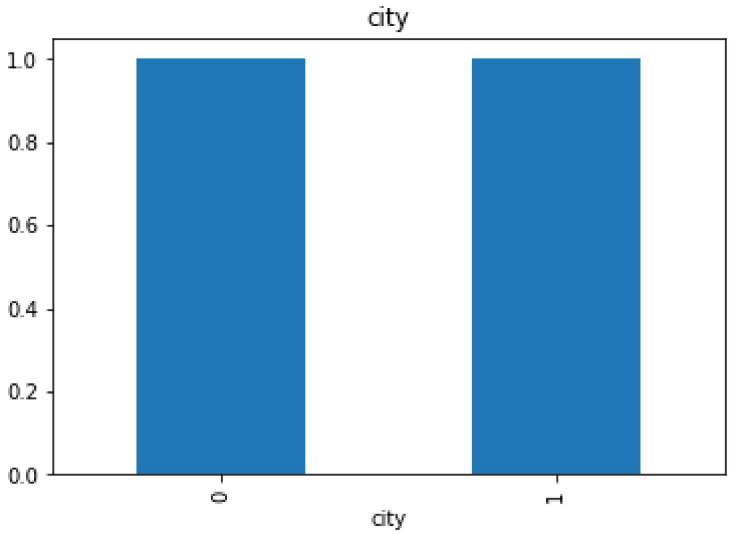
Out[6]:
```
1    161632
0     20059
Name: success, dtype: int64
```

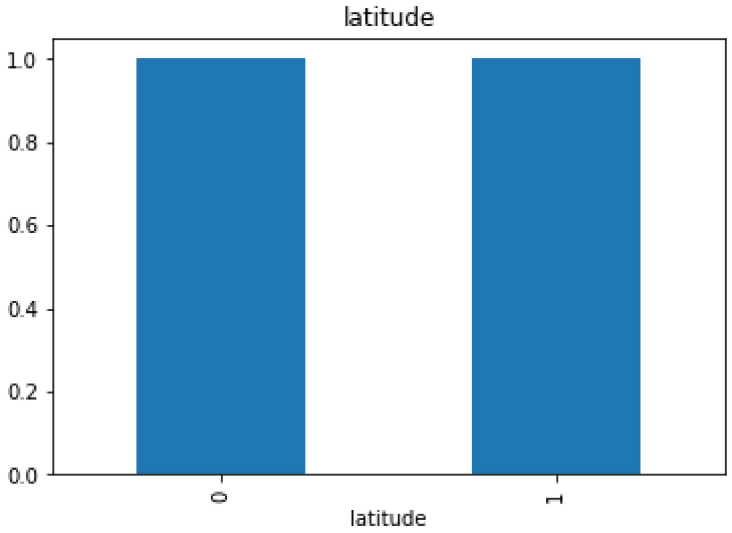
In [7]:
```
for feature in features_with_na:
    data = dataset.copy()

    # let's make a variable that indicates 1 if the observation was missing or zero
    data[feature] = np.where(data[feature].isnull(), 1, 0)

    # let's calculate the mean SalePrice where the information is missing or present
    data.groupby(feature)['success'].median().plot.bar()
    plt.title(feature)
    plt.show()
```

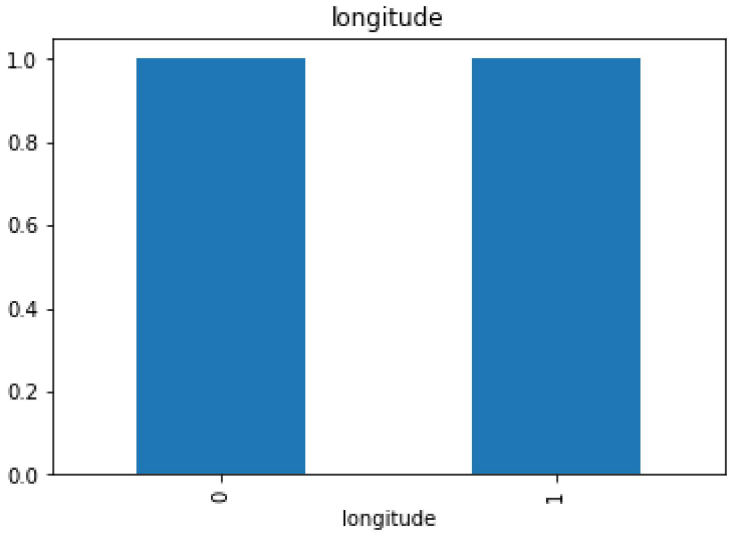


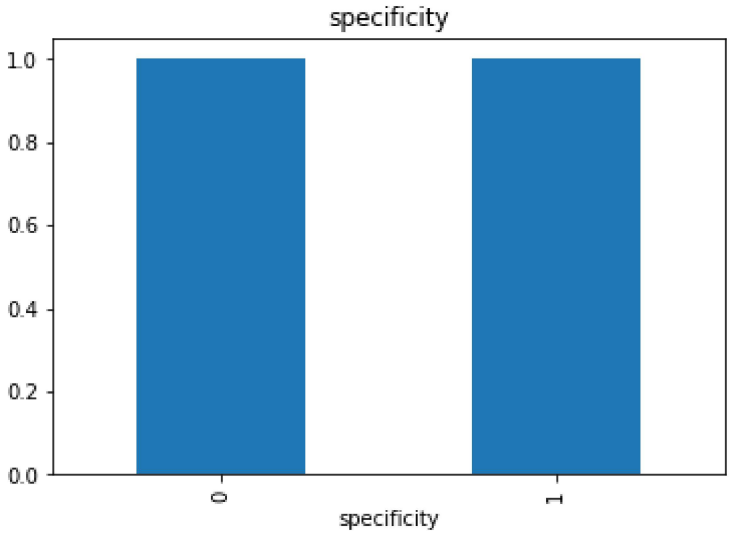


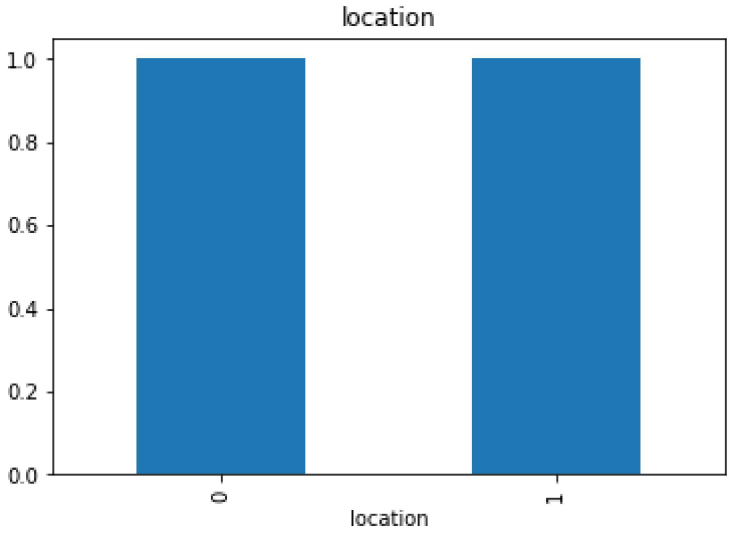

provstate
1.0
0.8
0.6
0.4
0.2
0.0
0
1
provstate

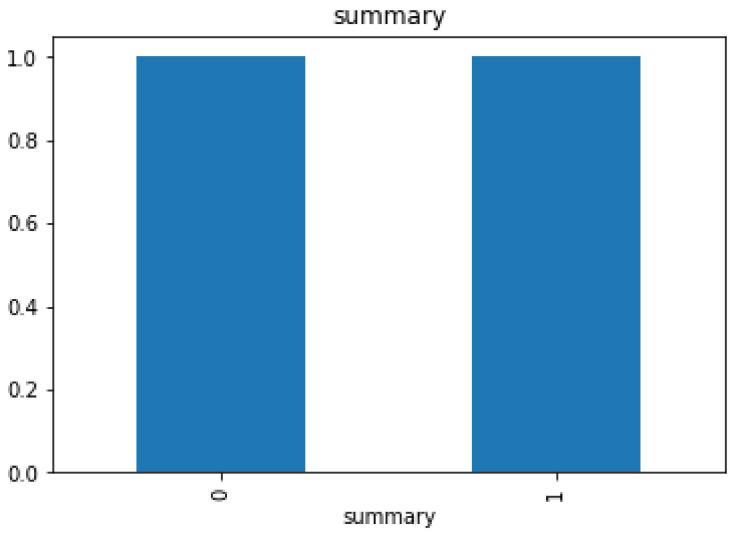

city
1.0
0.8
0.6
0.4
0.2
0.0
0
1
city

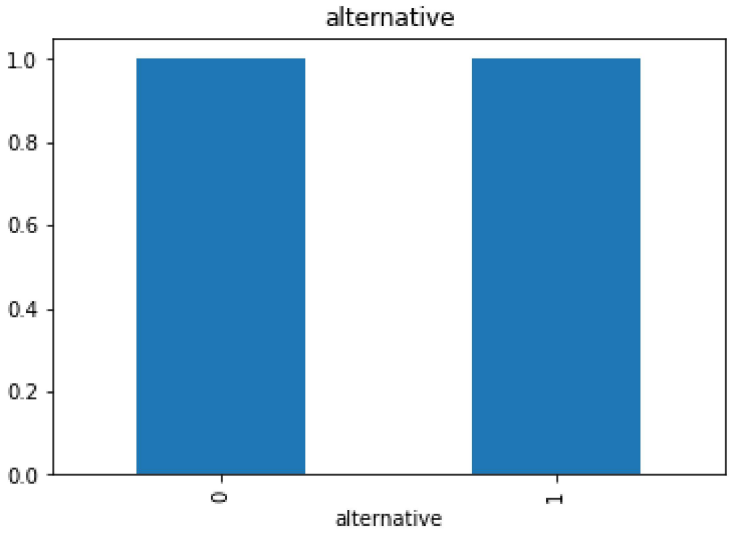

latitude
1.0
0.8
0.6
0.4
0.2
0.0
0
1
latitude

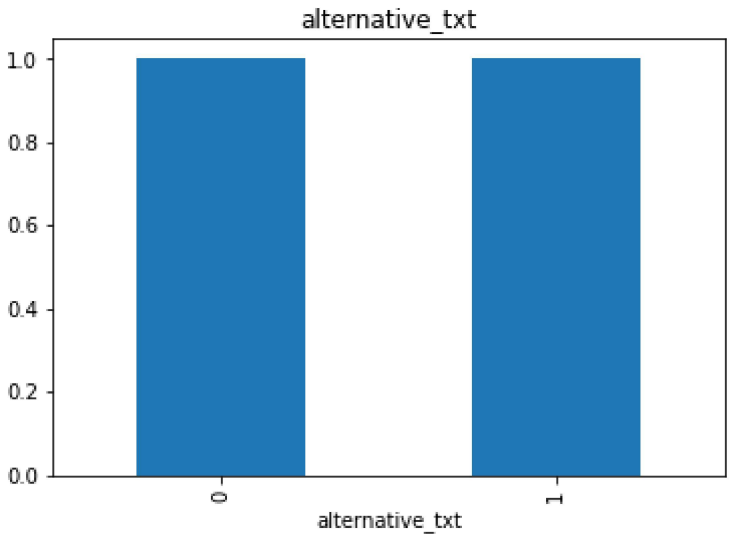

longitude
1.0
0.8
0.6
0.4
0.2
0.0
0
1
longitude

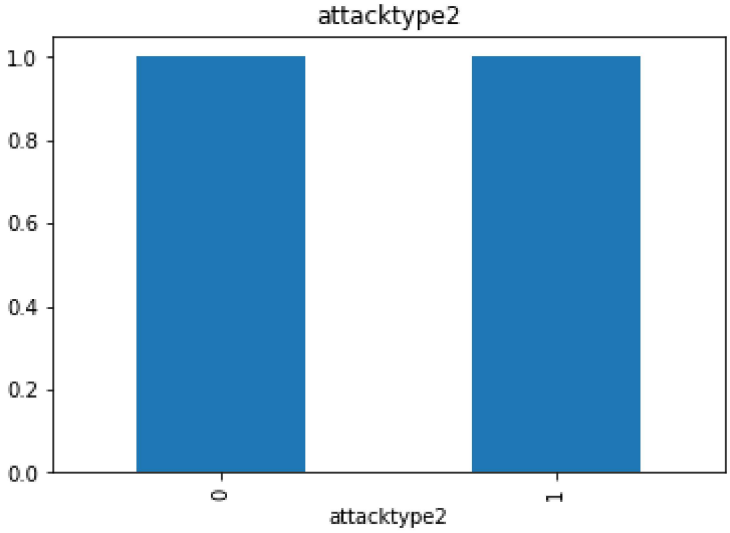

specificity
1.0
0.8
0.6
0.4
0.2
0.0
0
1
specificity

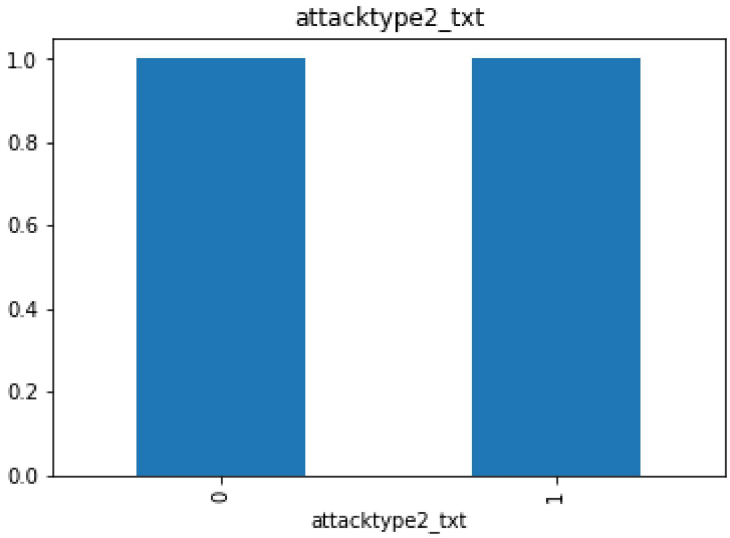

location
1.0
0.8
0.6
0.4
0.2
0.0
0
1
location

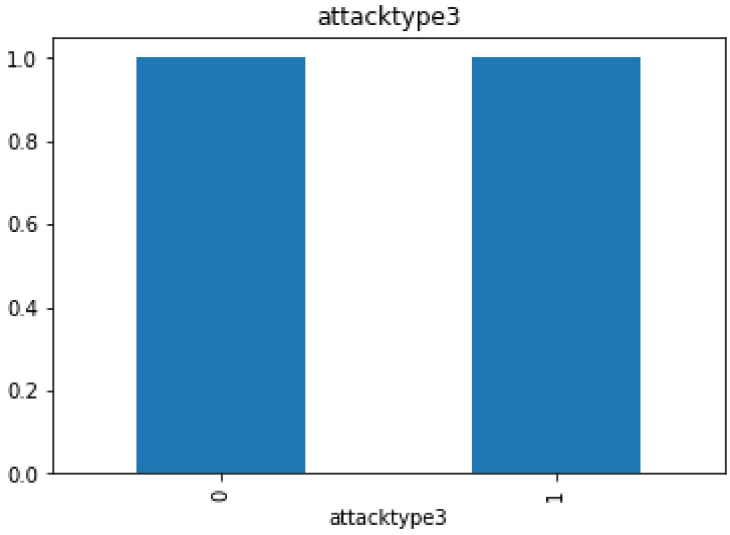

summary
1.0
0.8
0.6
0.4
0.2
0.0
0
1
summary

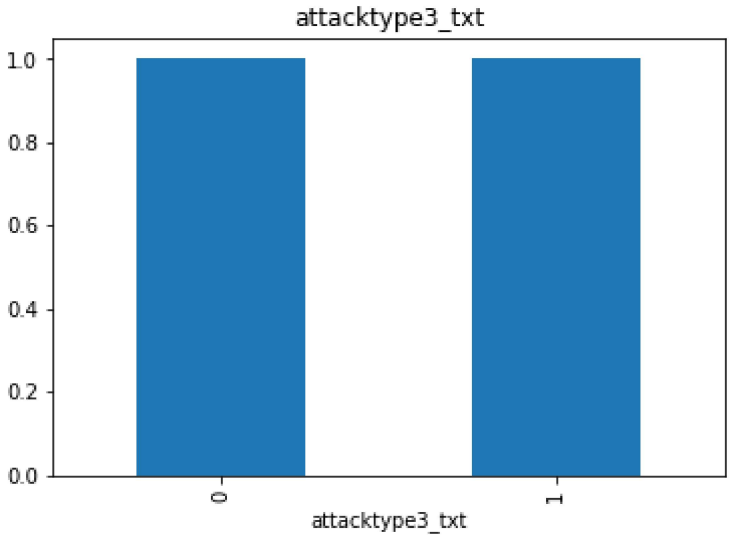

alternative
1.0
0.8
0.6
0.4
0.2
0.0
0
1
alternative

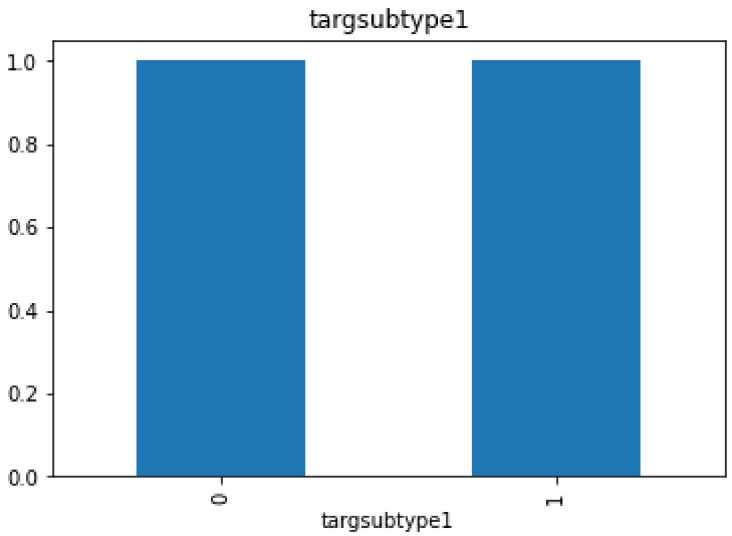

alternative_txt
1.0
0.8
0.6
0.4
0.2
0.0
0
1
alternative_txt

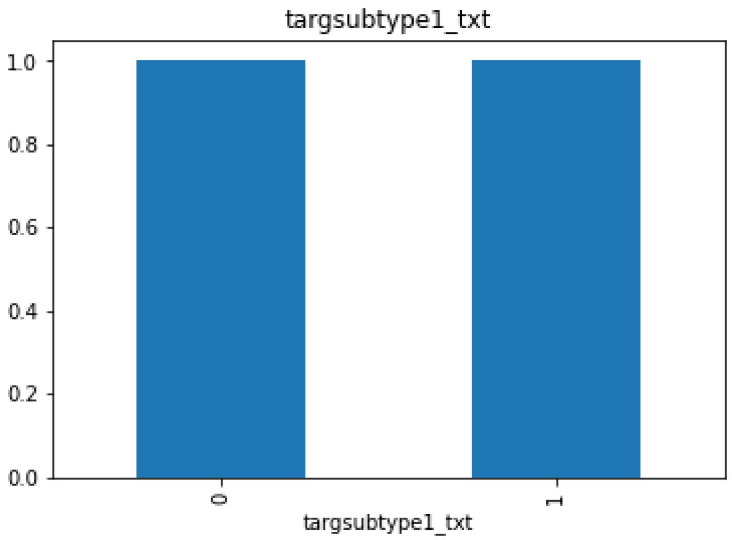

attacktype2
1.0
0.8
0.6
0.4
0.2
0.0
0
1
attacktype2

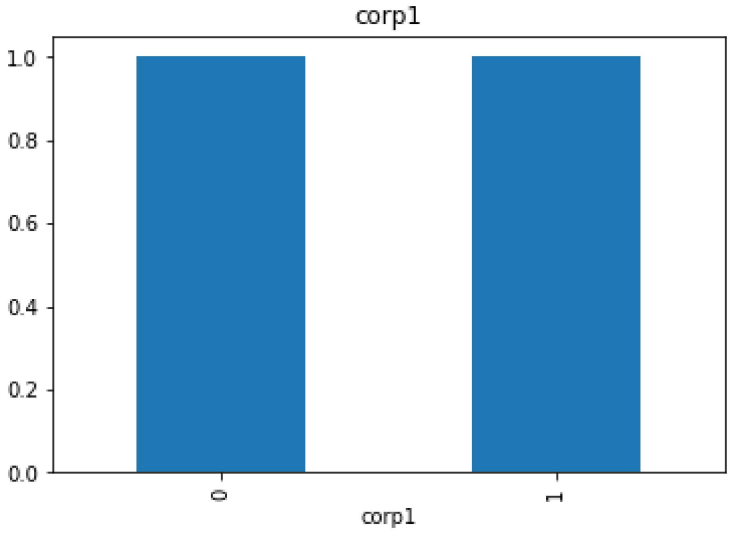

attacktype2_txt
1.0
0.8
0.6
0.4
0.2
0.0
0
1
attacktype2_txt

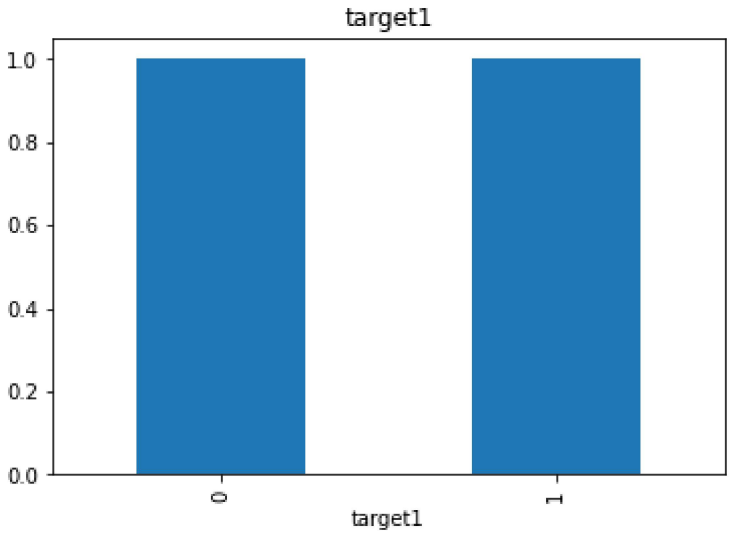

attacktype3
1.0
0.8
0.6
0.4
0.2
0.0
0
1
attacktype3

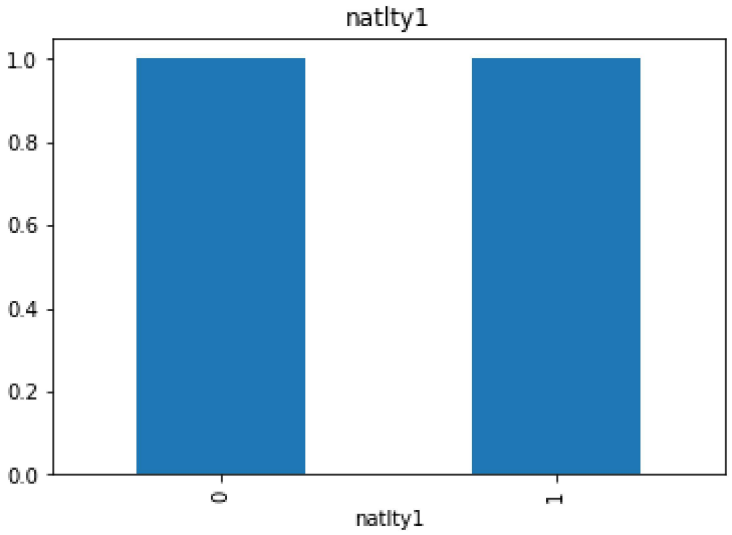

attacktype3_txt
1.0
0.8
0.6
0.4
0.2
0.0
0
1
attacktype3_txt

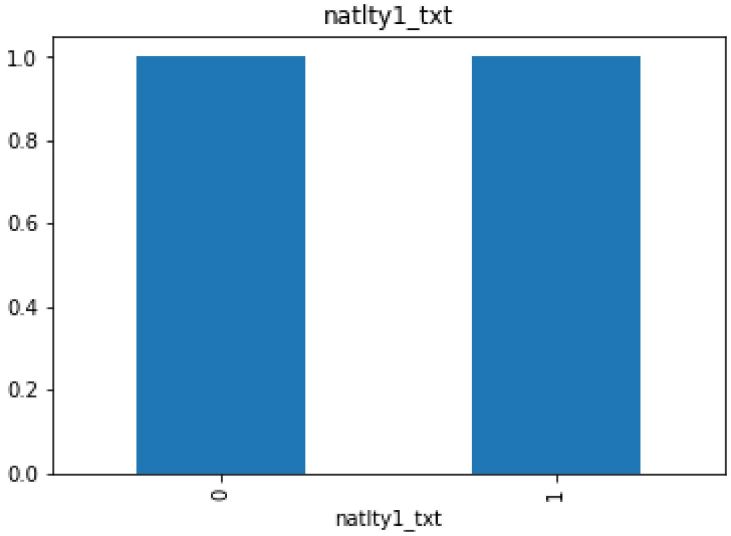

targsubtype1
1.0
0.8
0.6
0.4
0.2
0.0
0
1
targsubtype1

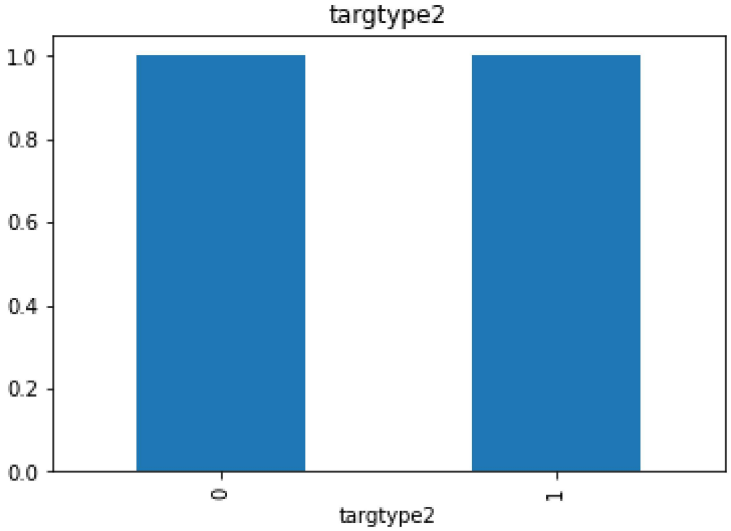

targsubtype1_txt
1.0
0.8
0.6
0.4
0.2
0.0
0
1
targsubtype1_txt

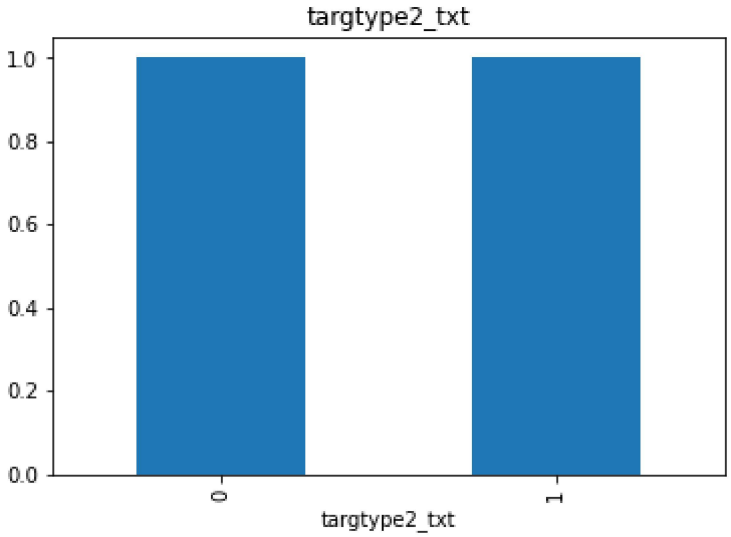

corp1
1.0
0.8
0.6
0.4
0.2
0.0
0
1
corp1

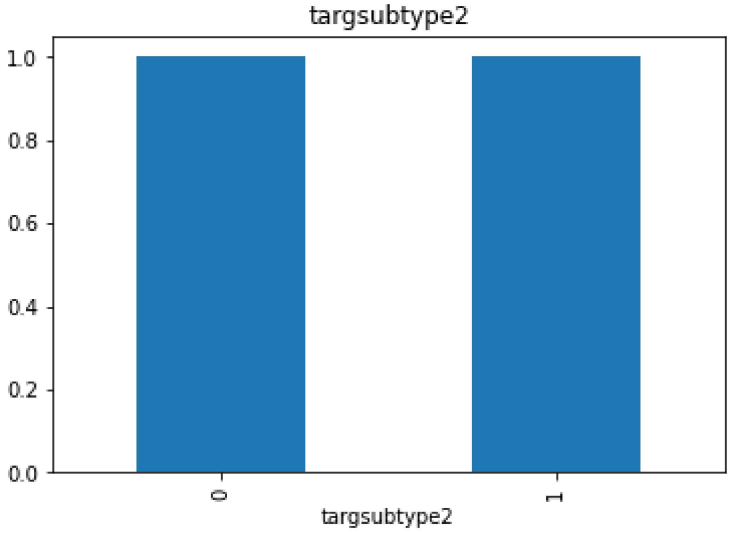

target1
1.0
0.8
0.6
0.4
0.2
0.0
0
1
target1

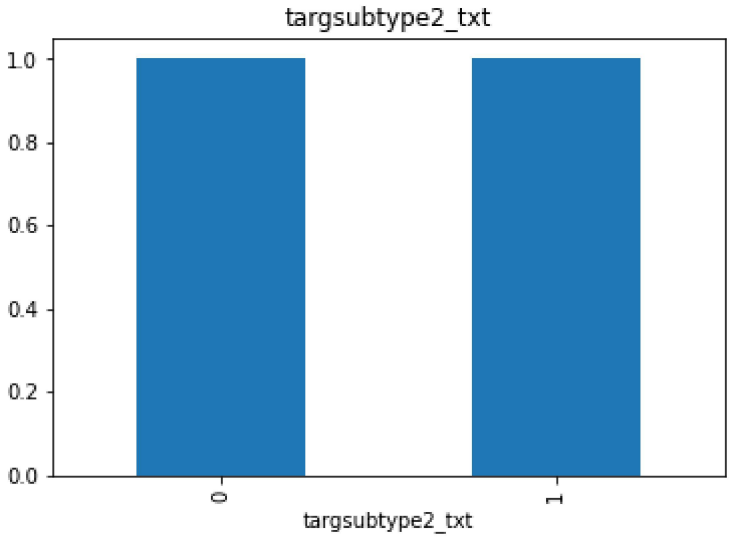

natlty1
1.0
0.8
0.6
0.4
0.2
0.0
0
1
natlty1

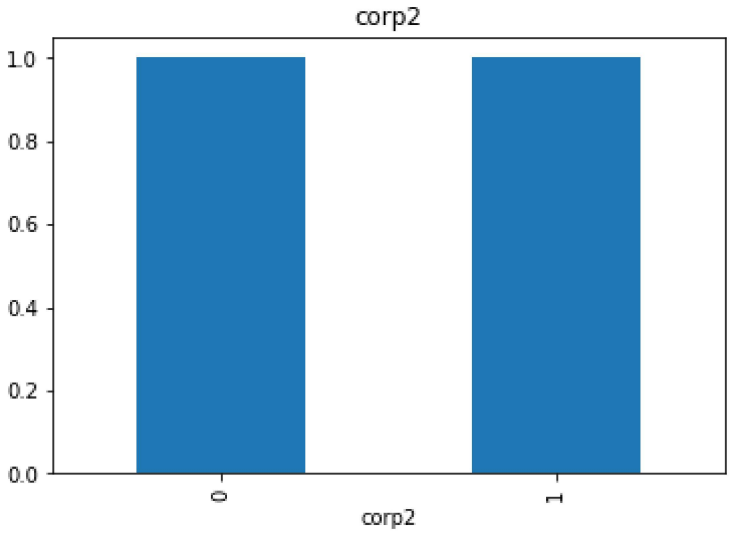

natlty1_txt
1.0
0.8
0.6
0.4
0.2
0.0
0
1
natlty1_txt

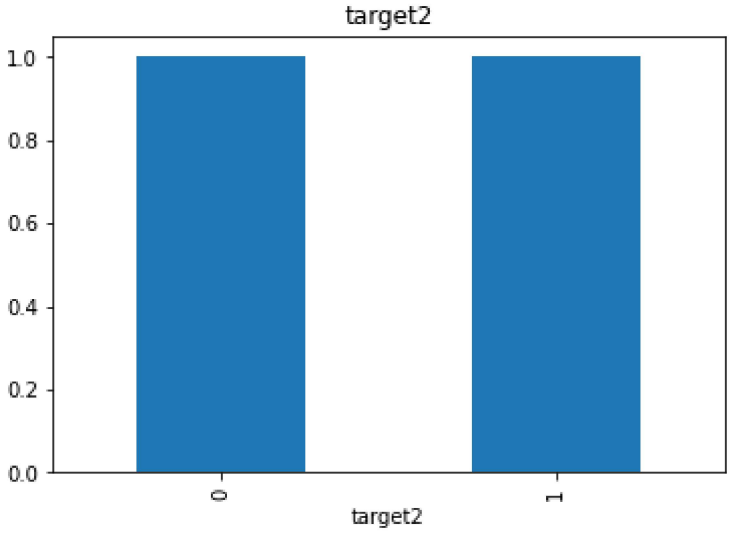

targtype2
1.0
0.8
0.6
0.4
0.2
0.0
0
1
targtype2

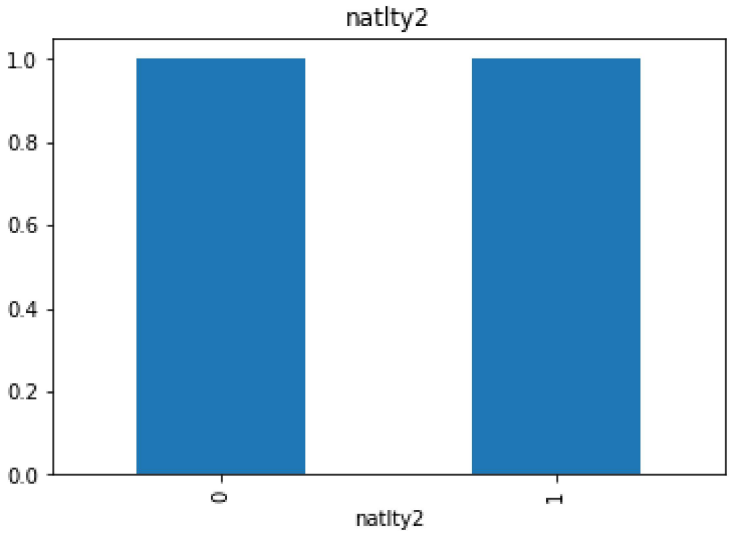

targtype2_txt
1.0
0.8
0.6
0.4
0.2
0.0
0
1
targtype2_txt

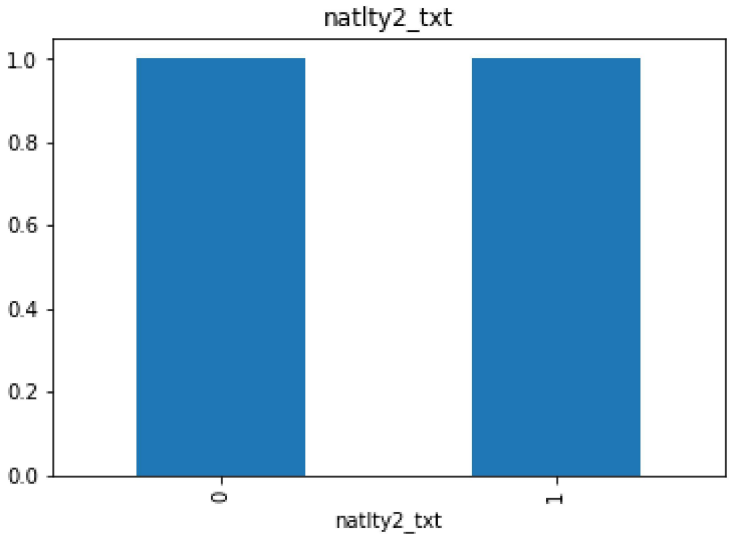

targsubtype2
1.0
0.8
0.6
0.4
0.2
0.0
0
1
targsubtype2

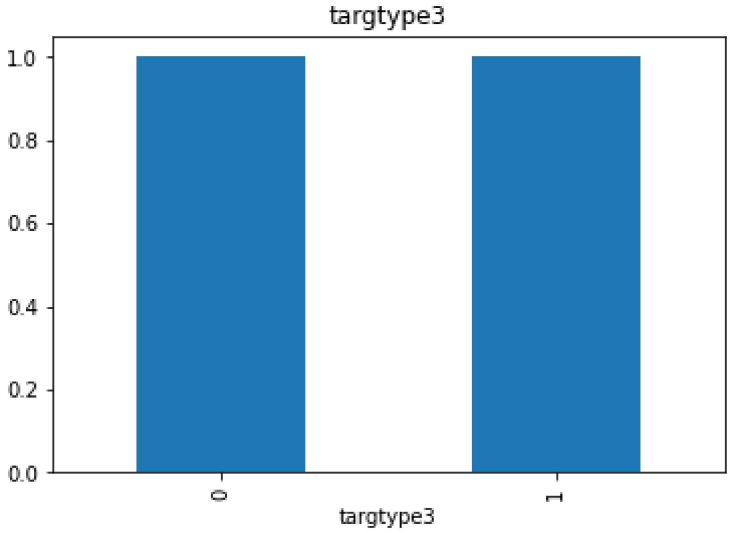

targsubtype2_txt
1.0
0.8
0.6
0.4
0.2
0.0
0
1
targsubtype2_txt

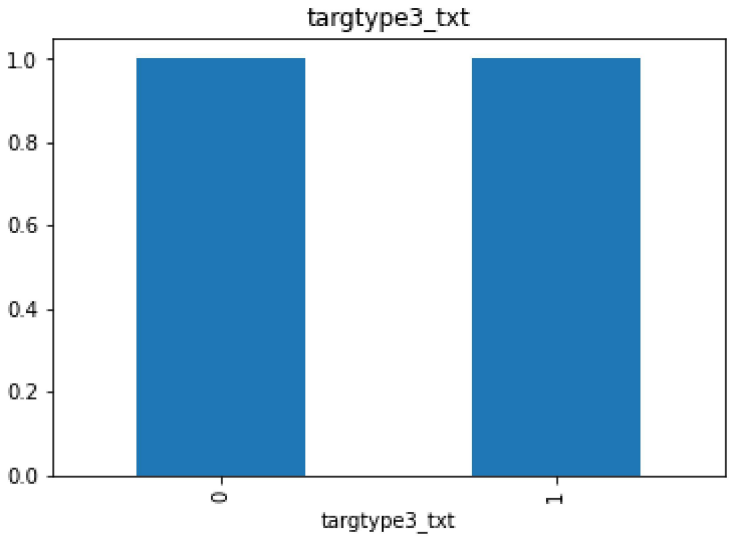

corp2
1.0
0.8
0.6
0.4
0.2
0.0
0
1
corp2

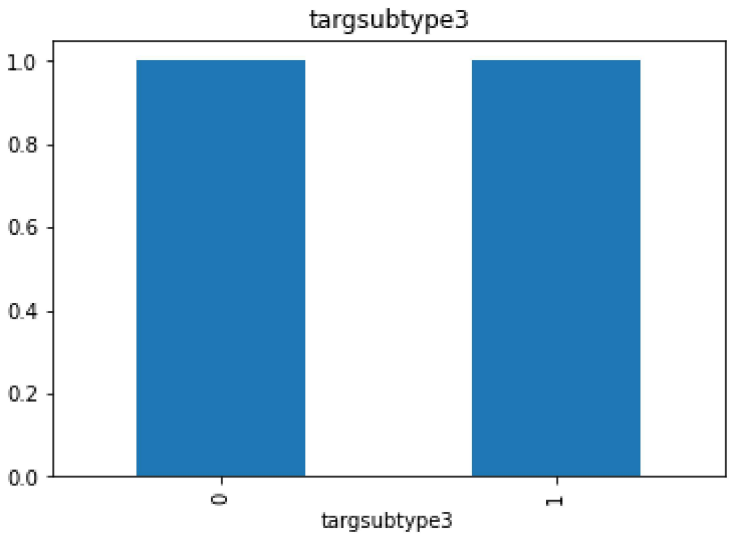

target2
1.0
0.8
0.6
0.4
0.2
0.0
0
1
target2

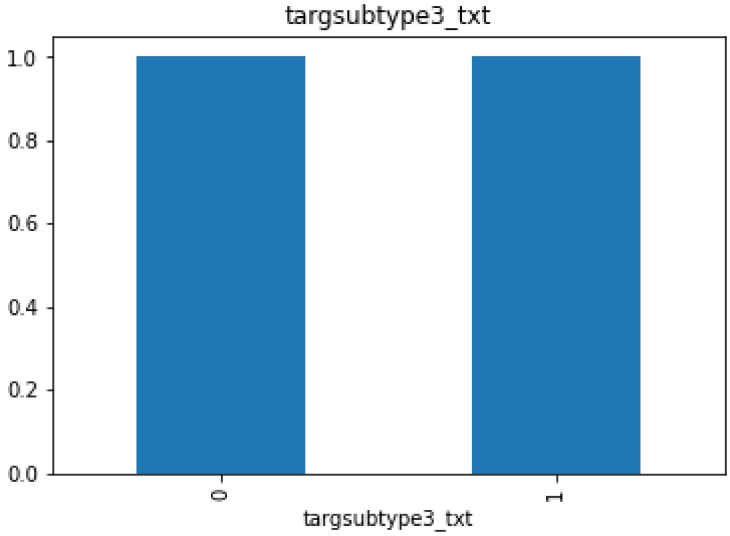

natlty2
1.0
0.8
0.6
0.4
0.2
0.0
0
1
natlty2

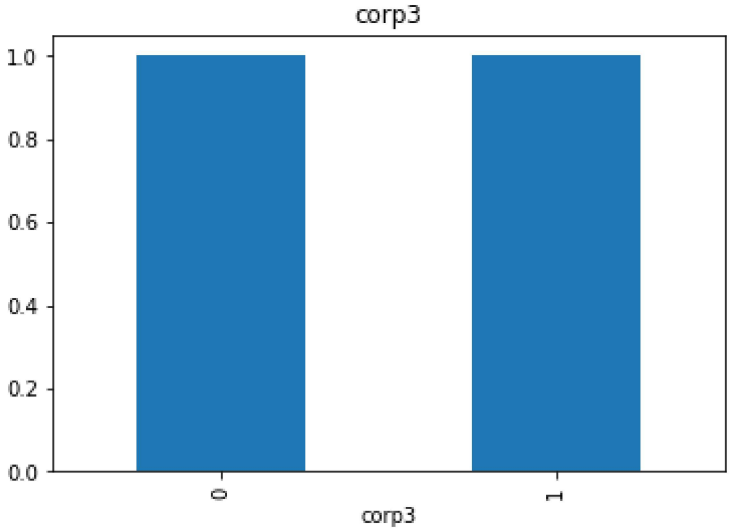

natlty2_txt
1.0
0.8
0.6
0.4
0.2
0.0
0
1
natlty2_txt

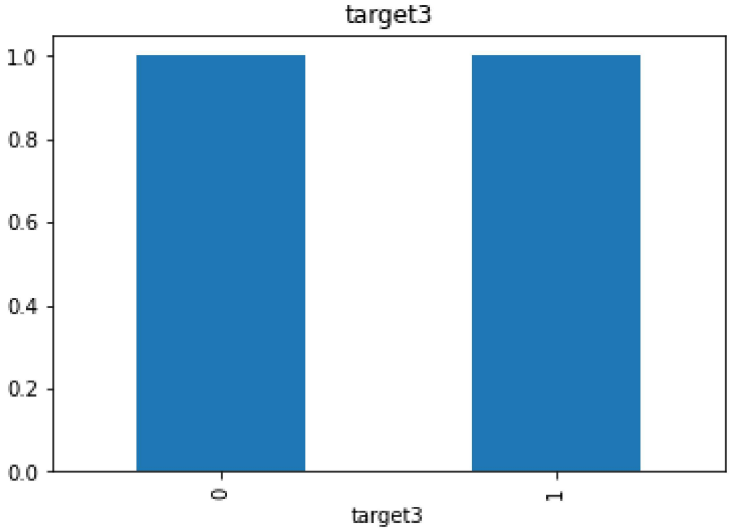

targtype3
1.0
0.8
0.6
0.4
0.2
0.0
0
1
targtype3

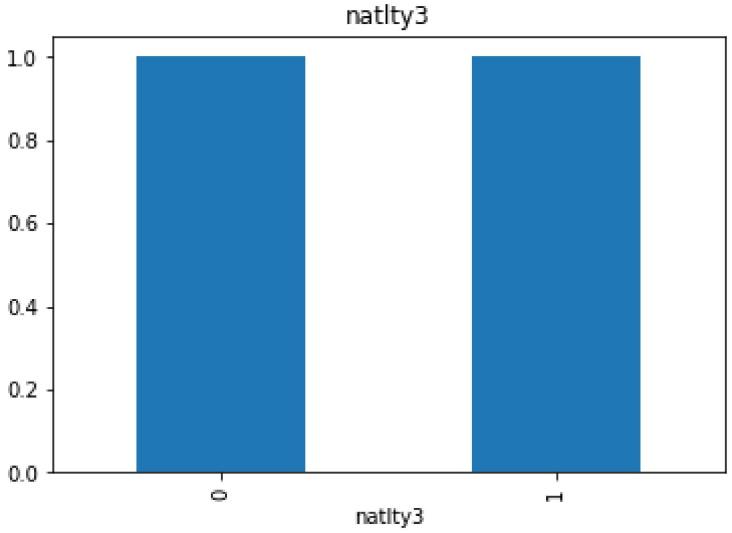

targtype3_txt
1.0
0.8
0.6
0.4
0.2
0.0
0
1
targtype3_txt

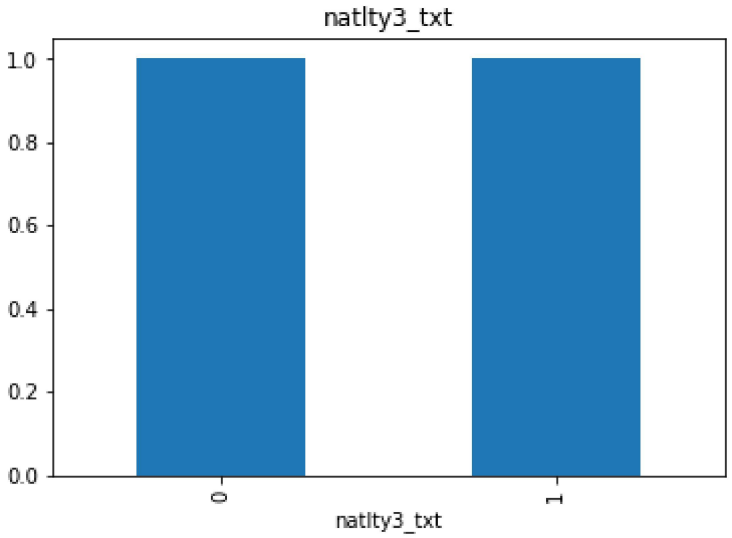

targsubtype3
1.0
0.8
0.6
0.4
0.2
0.0
0
1
targsubtype3

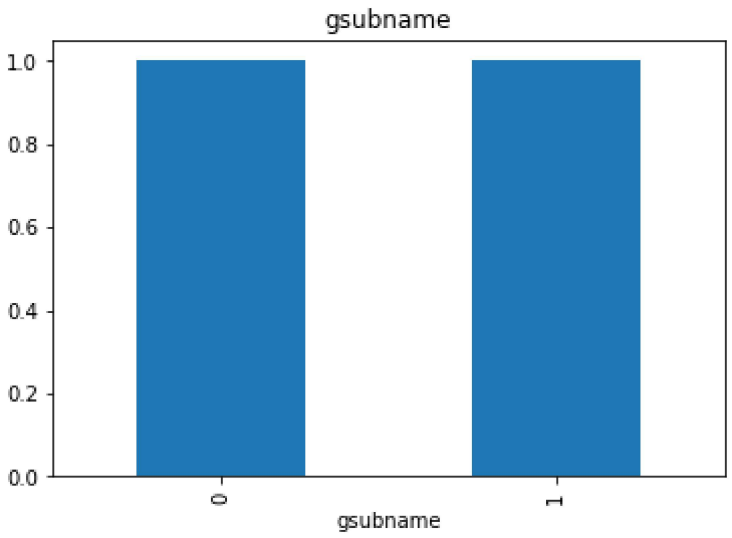

targsubtype3_txt
1.0
0.8
0.6
0.4
0.2
0.0
0
1
targsubtype3_txt

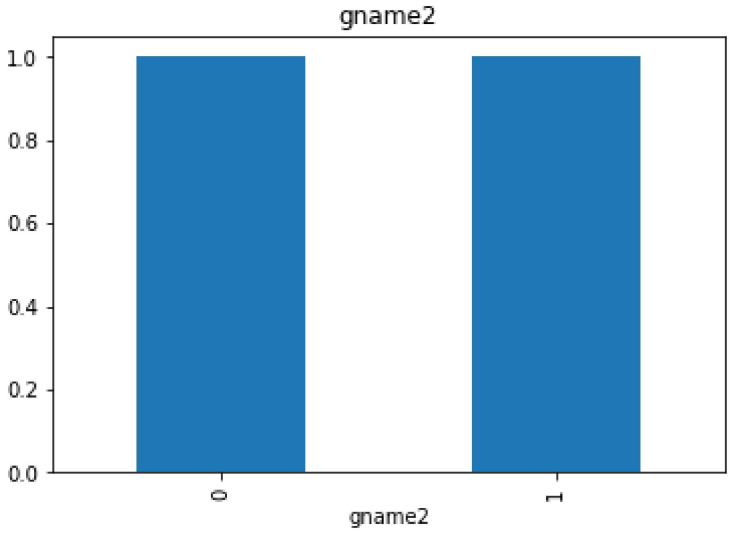

corp3
1.0
0.8
0.6
0.4
0.2
0.0
0
1
corp3

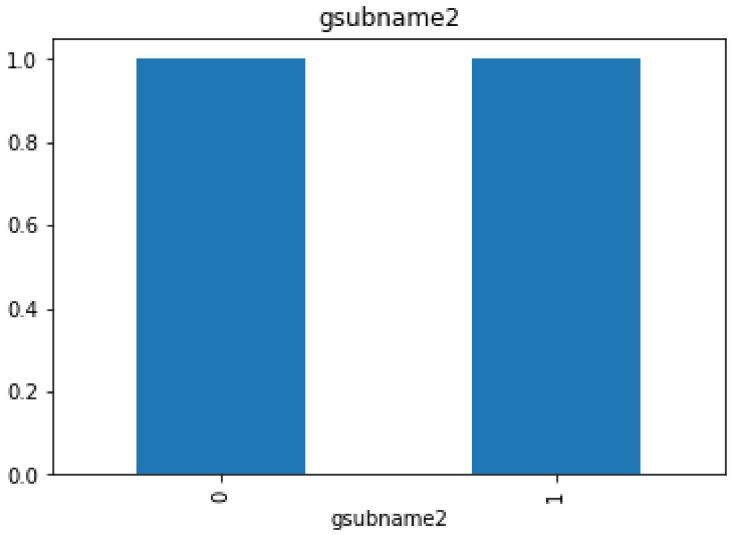

target3
1.0
0.8
0.6
0.4
0.2
0.0
0
1
target3

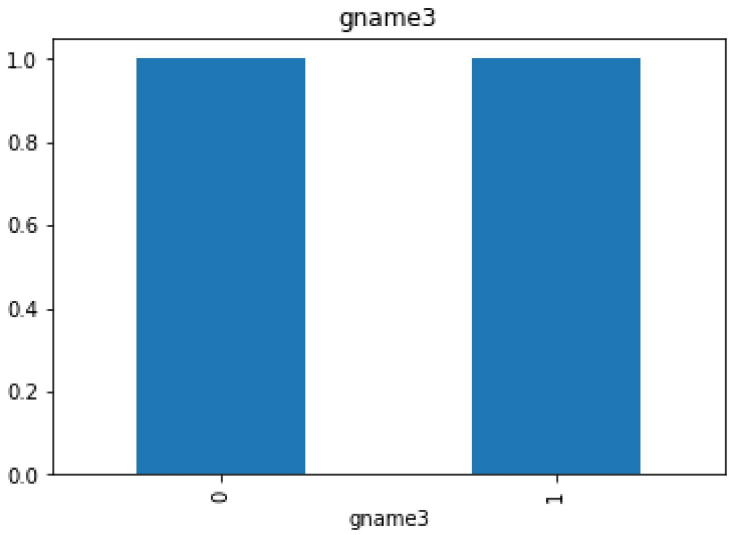

natlty3
1.0
0.8
0.6
0.4
0.2
0.0
0
1
natlty3

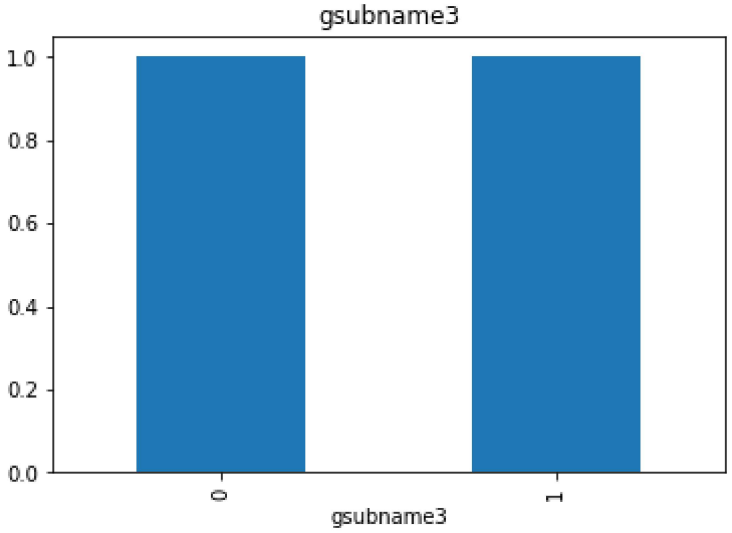

natlty3_txt
1.0
0.8
0.6
0.4
0.2
0.0
0
1
natlty3_txt

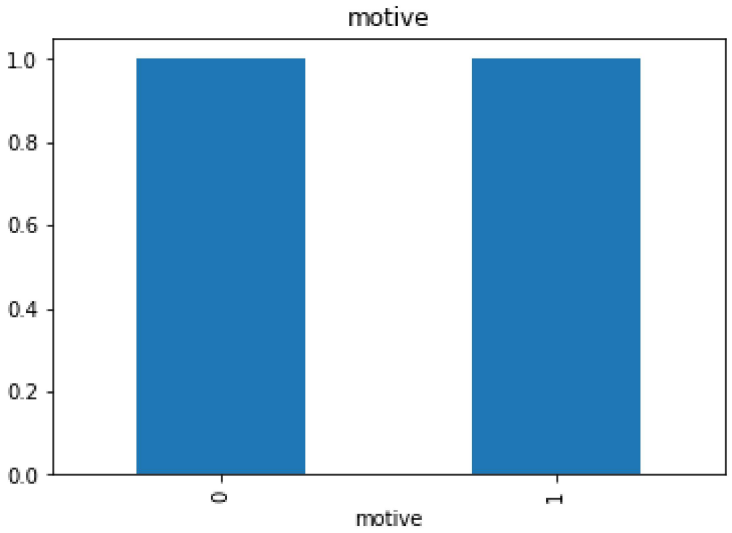

gsubname
1.0
0.8
0.6
0.4
0.2
0.0
0
1
gsubname

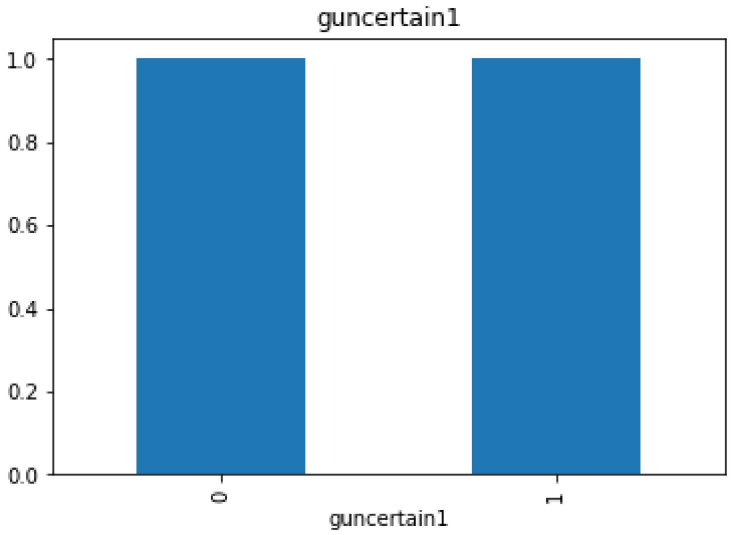

gname2
1.0
0.8
0.6
0.4
0.2
0.0
0
1
gname2

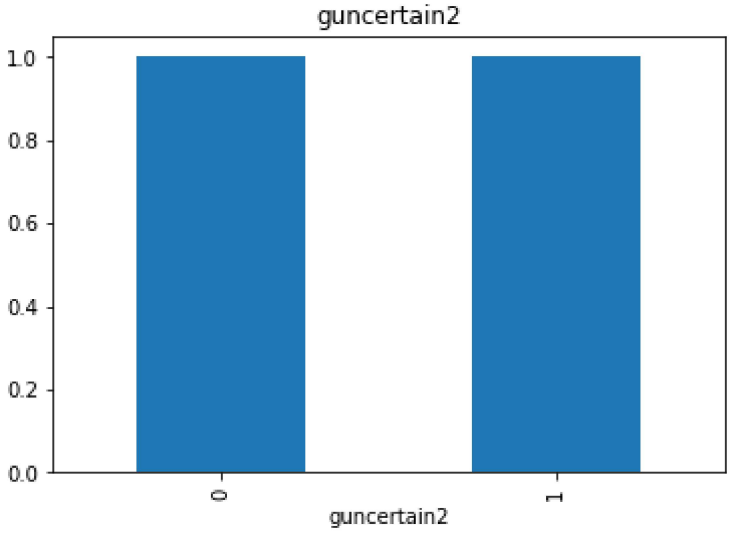

gsubname2
1.0
0.8
0.6
0.4
0.2
0.0
0
1
gsubname2

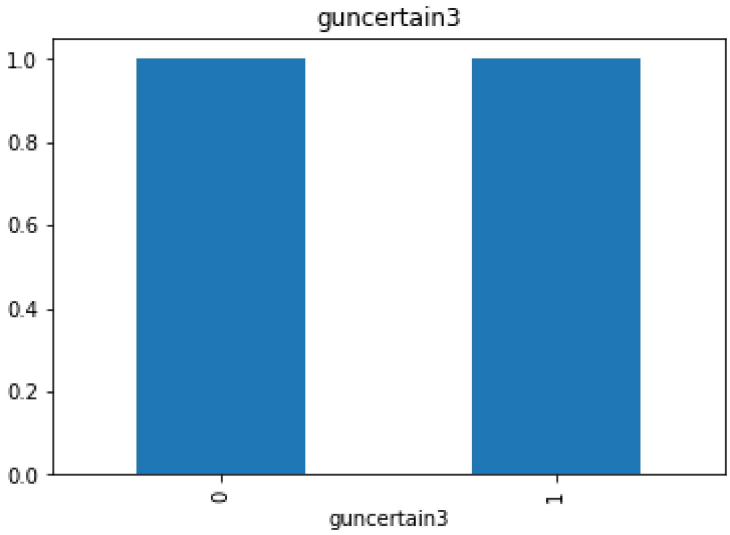

gname3
1.0
0.8
0.6
0.4
0.2
0.0
0
1
gname3

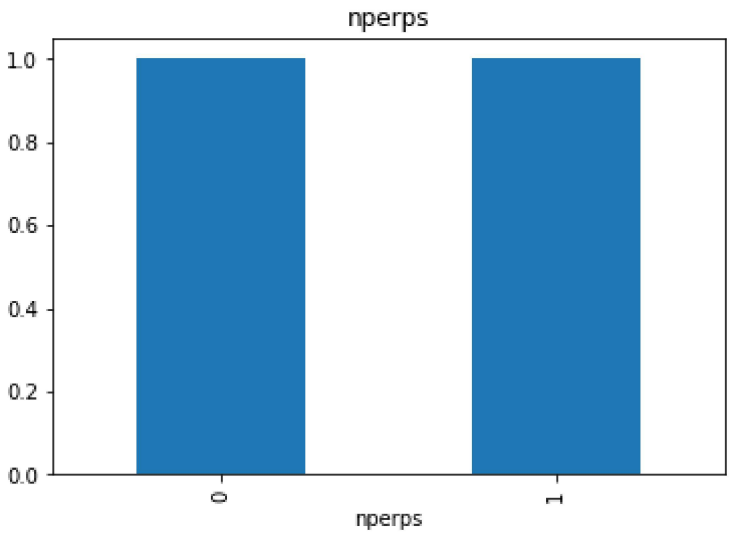

gsubname3
1.0
0.8
0.6
0.4
0.2
0.0
0
1
gsubname3

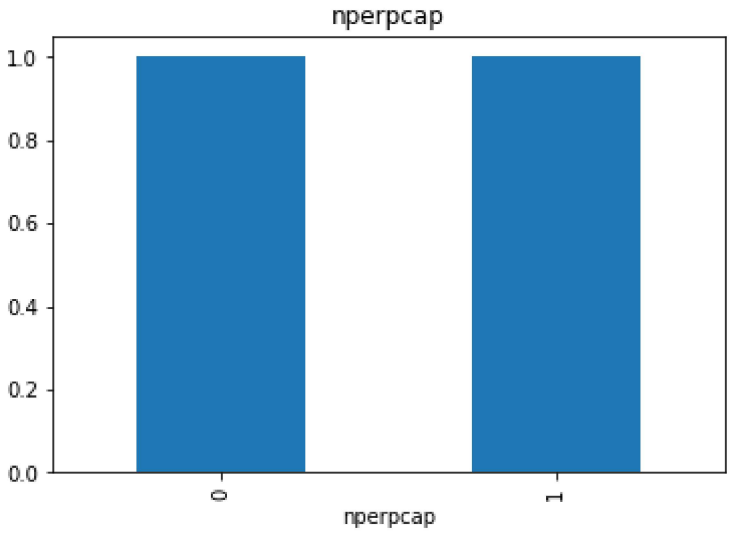

motive
1.0
0.8
0.6
0.4
0.2
0.0
0
1
motive

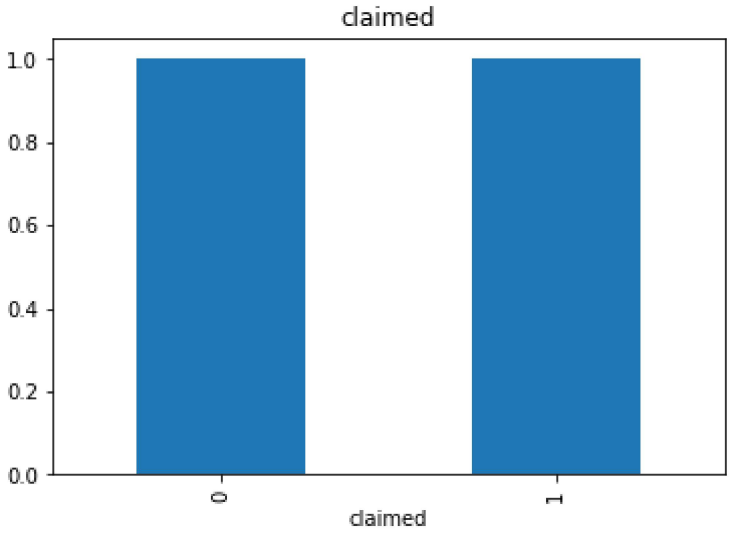

guncertain1
1.0
0.8
0.6
0.4
0.2
0.0
0
1
guncertain1

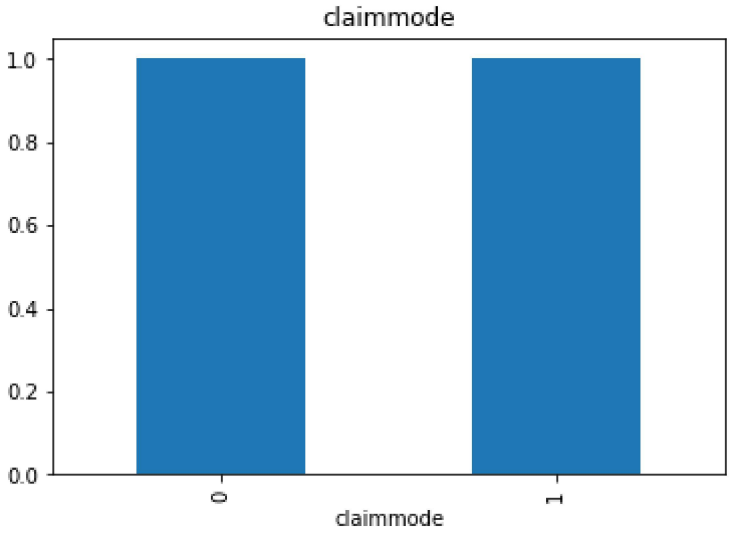

guncertain2
1.0
0.8
0.6
0.4
0.2
0.0
0
1
guncertain2

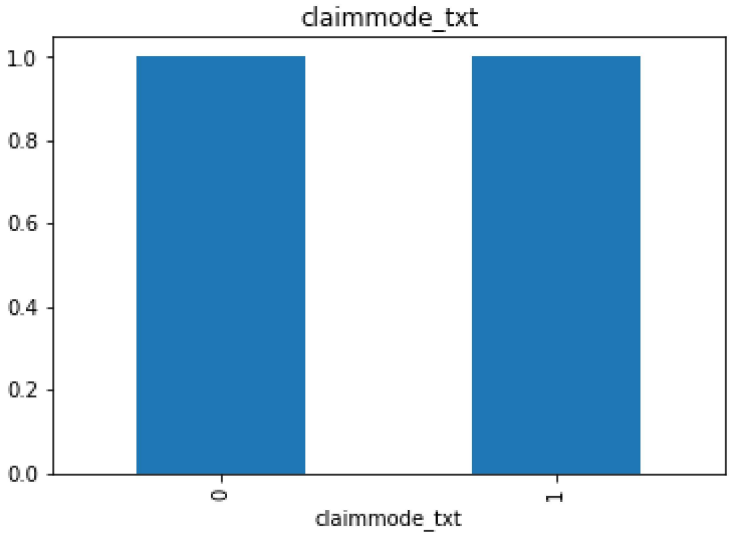

guncertain3
1.0
0.8
0.6
0.4
0.2
0.0
0
1
guncertain3

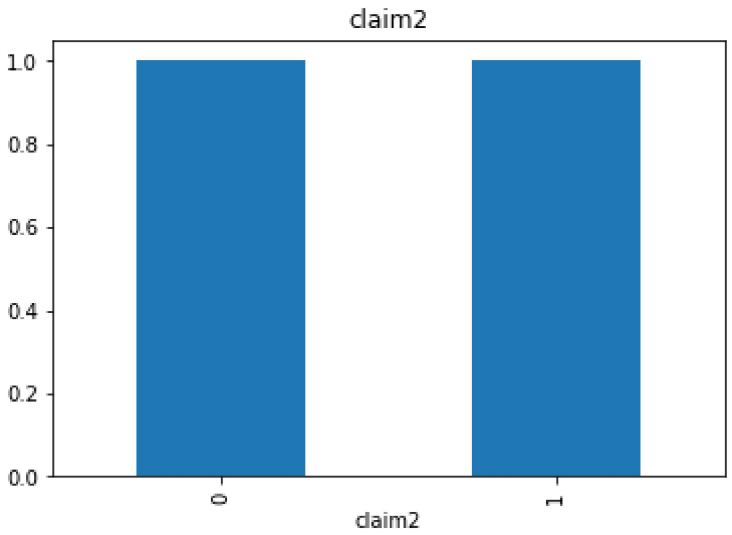

nperps
1.0
0.8
0.6
0.4
0.2
0.0
0
1
nperps

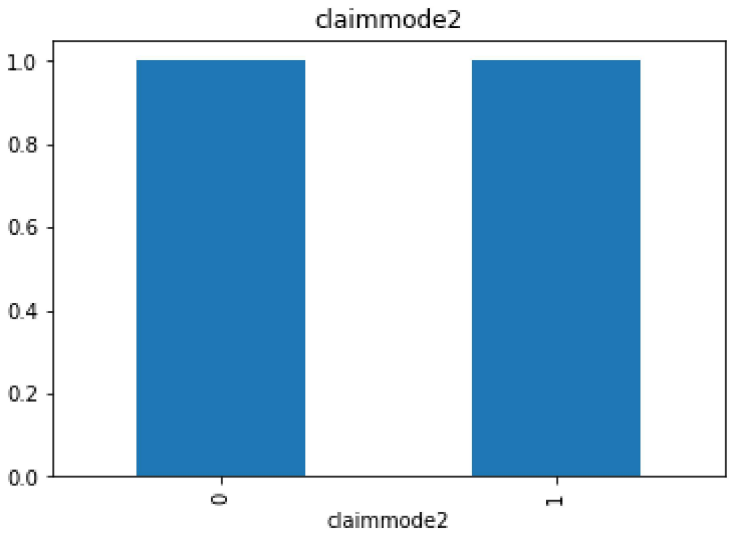

nperpcap
1.0
0.8
0.6
0.4
0.2
0.0
0
1
nperpcap

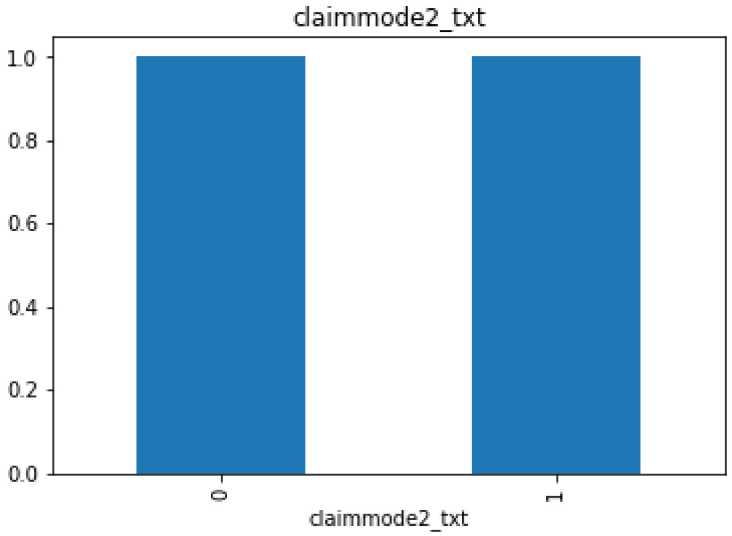

claimed
1.0
0.8
0.6
0.4
0.2
0.0
0
1
claimed

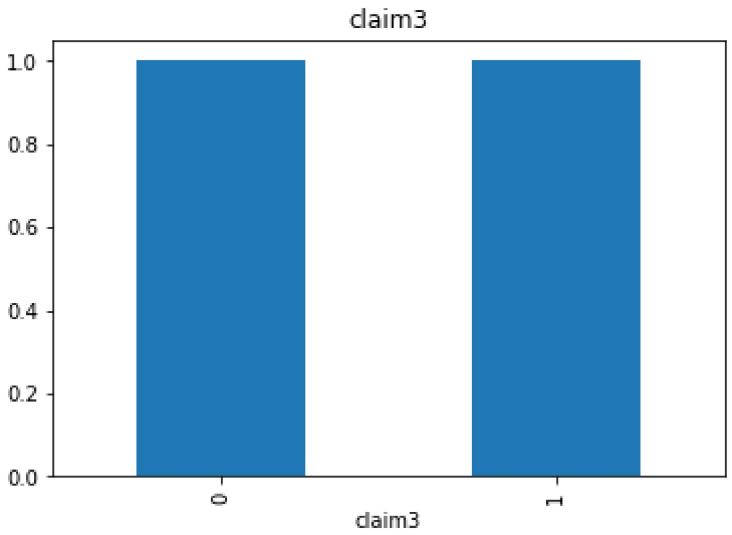

claimmode
1.0
0.8
0.6
0.4
0.2
0.0
0
1
claimmode

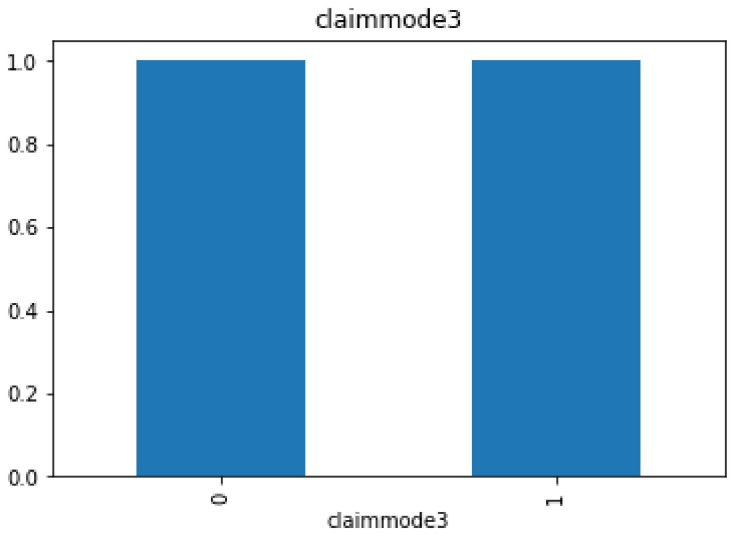

claimmode_txt
1.0
0.8
0.6
0.4
0.2
0.0
0
1
claimmode_txt

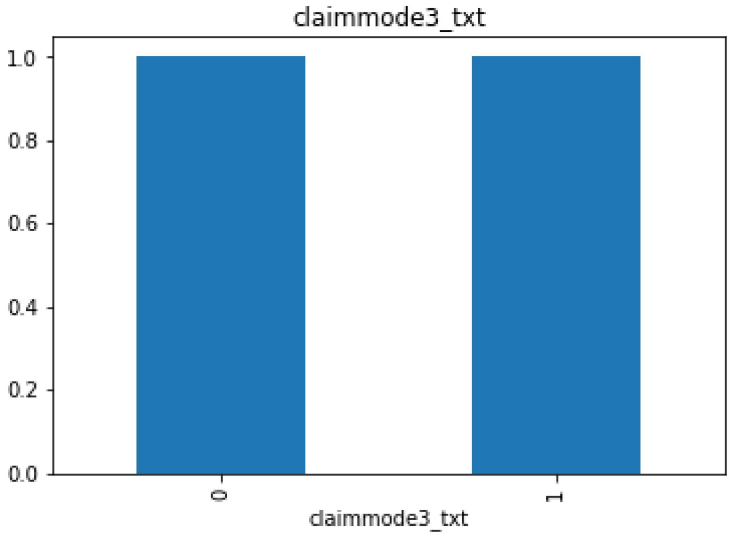

claim2
1.0
0.8
0.6
0.4
0.2
0.0
0
1
claim2

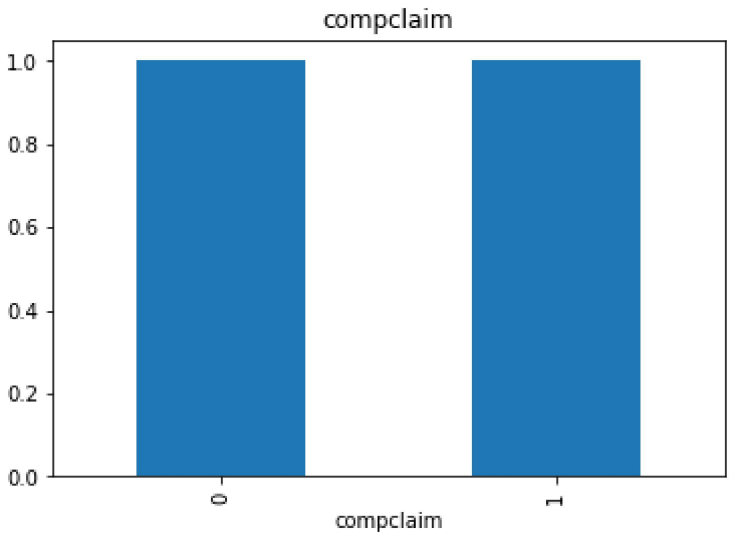

claimmode2
1.0
0.8
0.6
0.4
0.2
0.0
0
1
claimmode2

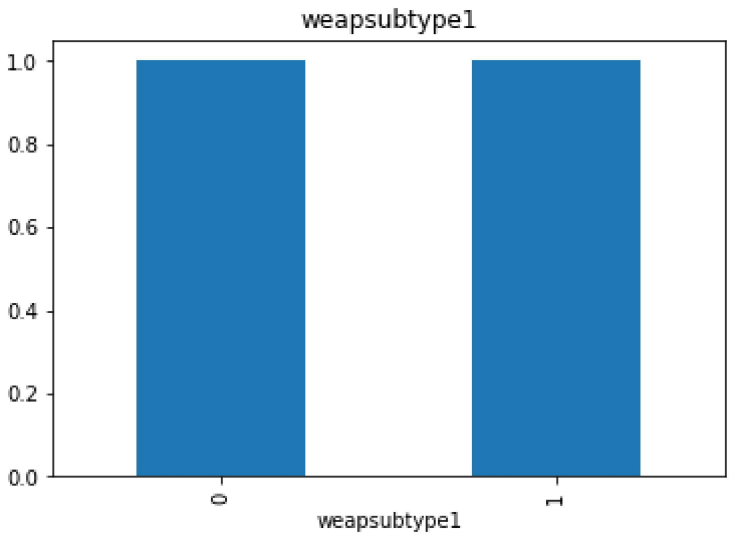

claimmode2_txt
1.0
0.8
0.6
0.4
0.2
0.0
0
1
claimmode2_txt

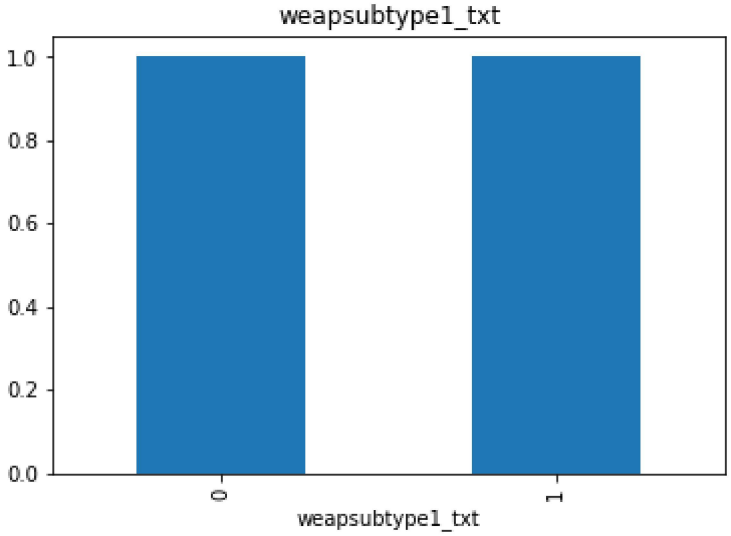

claim3
1.0
0.8
0.6
0.4
0.2
0.0
0
1
claim3

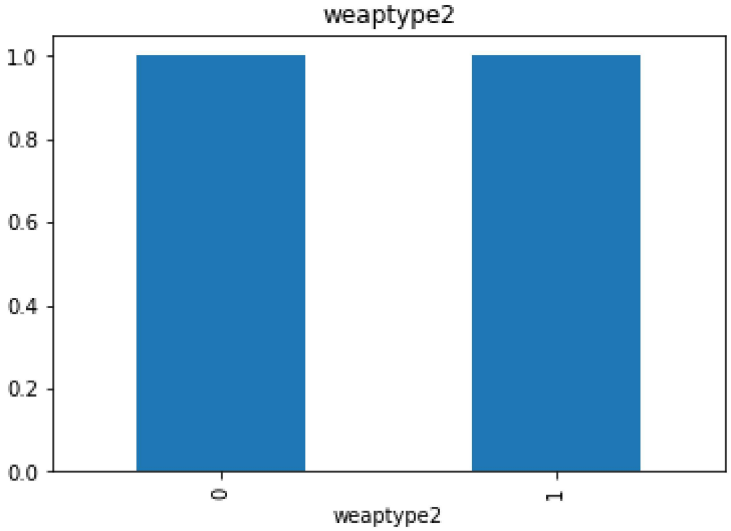

claimmode3
1.0
0.8
0.6
0.4
0.2
0.0
0
1
claimmode3

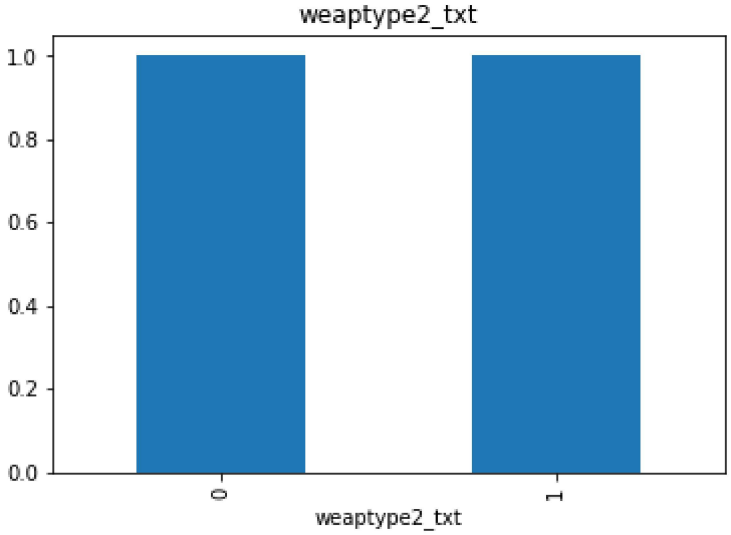

claimmode3_txt
1.0
0.8
0.6
0.4
0.2
0.0
0
1
claimmode3_txt

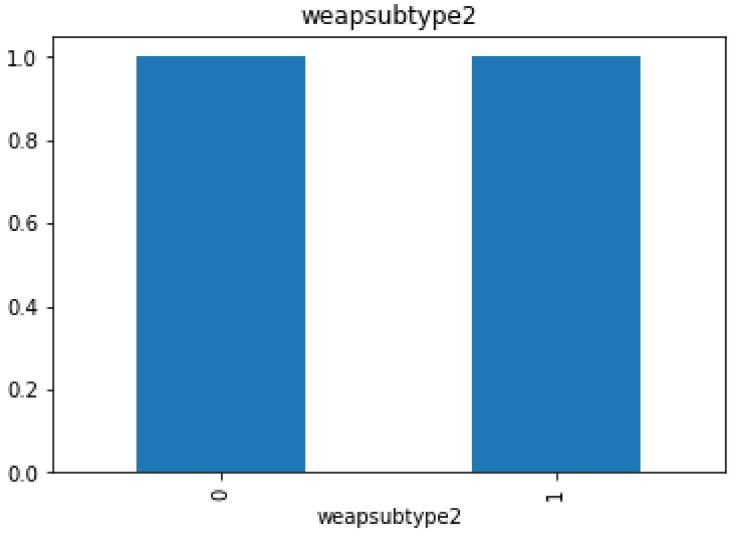

compclaim
1.0
0.8
0.6
0.4
0.2
0.0
0
1
compclaim

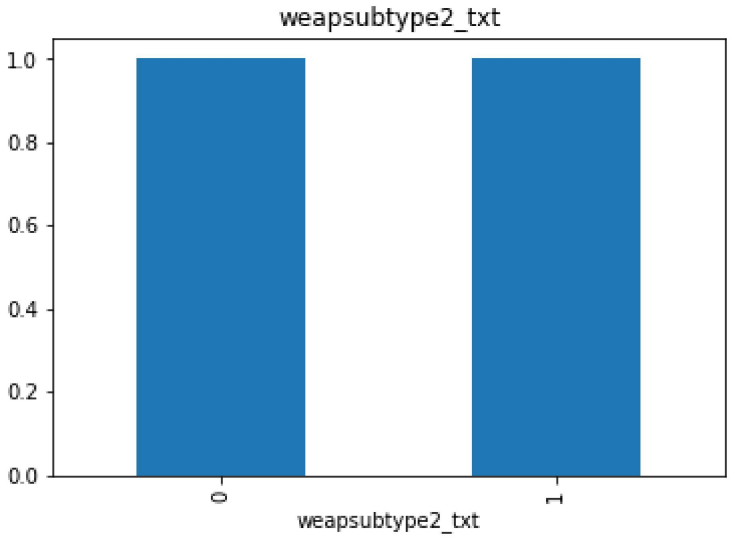

weapsubtype1
1.0
0.8
0.6
0.4
0.2
0.0
0
1
weapsubtype1

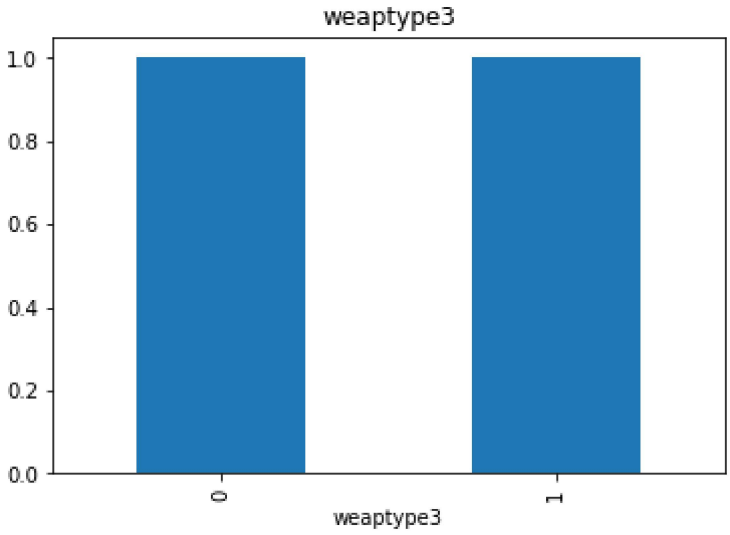

weapsubtype1_txt
1.0
0.8
0.6
0.4
0.2
0.0
0
1
weapsubtype1_txt

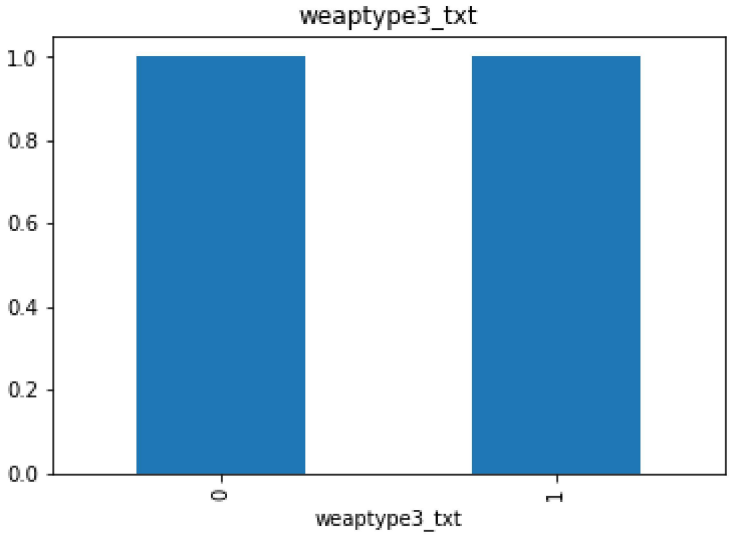

weaptype2
1.0
0.8
0.6
0.4
0.2
0.0
0
1
weaptype2

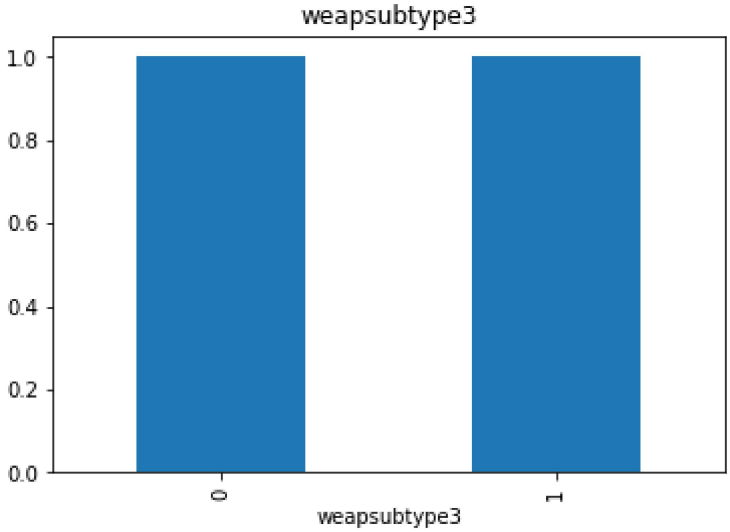

weaptype2_txt
1.0
0.8
0.6
0.4
0.2
0.0
0
1
weaptype2_txt

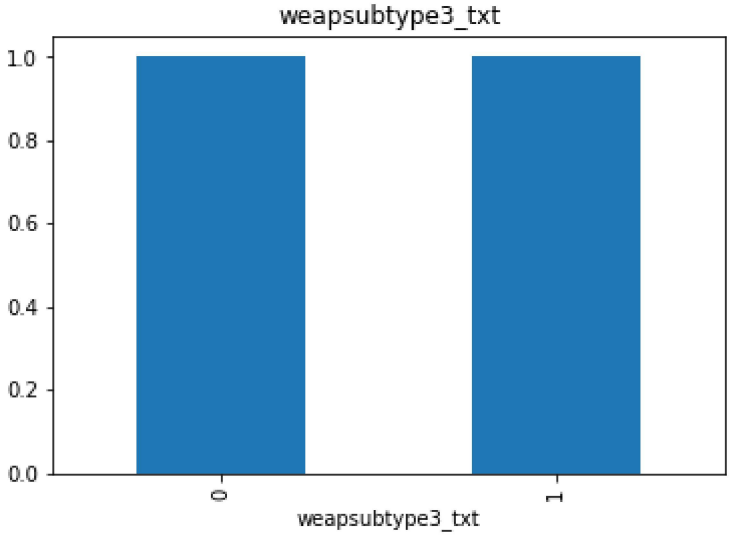

weapsubtype2
1.0
0.8
0.6
0.4
0.2
0.0
0
1
weapsubtype2

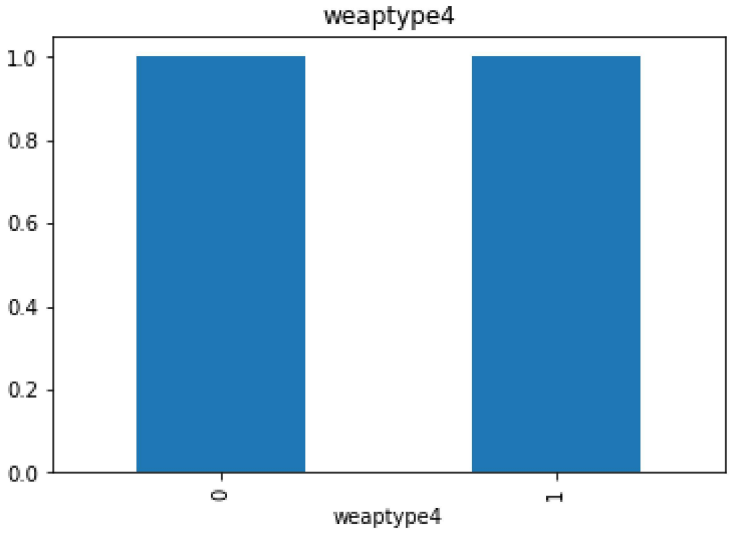

weapsubtype2_txt
1.0
0.8
0.6
0.4
0.2
0.0
0
1
weapsubtype2_txt

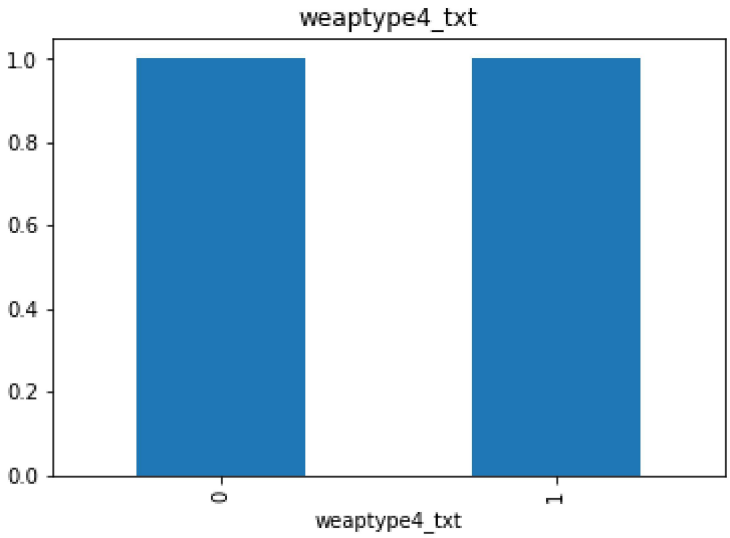

weaptype3
1.0
0.8
0.6
0.4
0.2
0.0
0
1
weaptype3

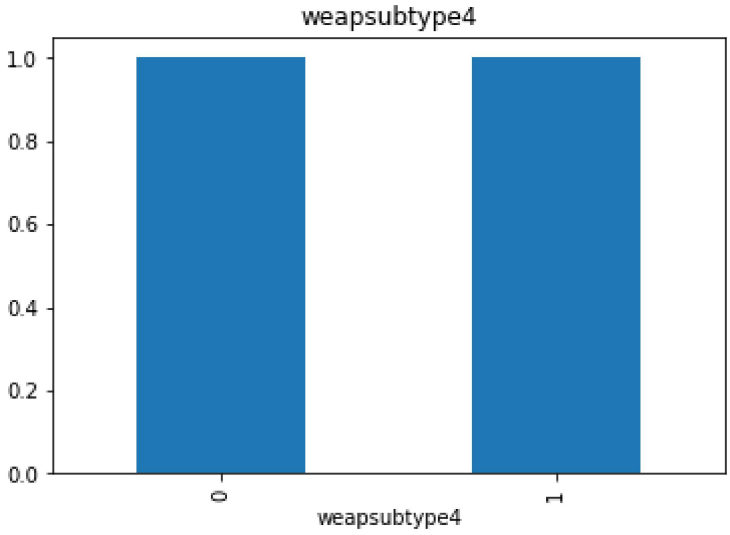

weaptype3_txt
1.0
0.8
0.6
0.4
0.2
0.0
0
1
weaptype3_txt

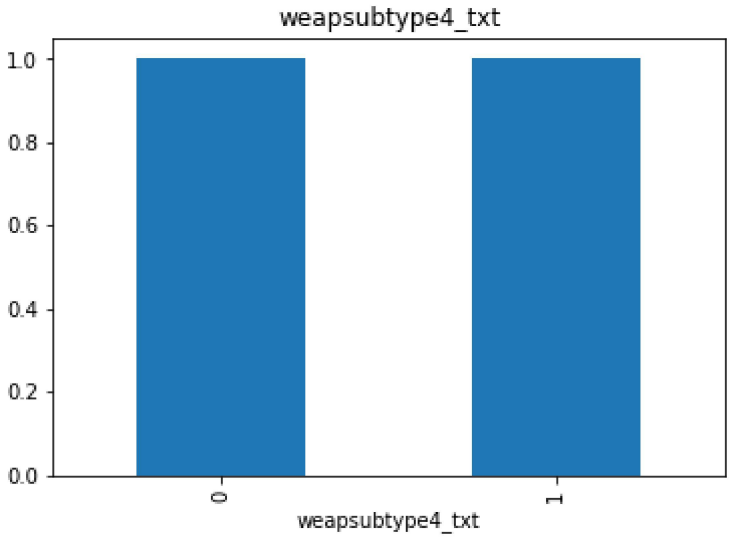

weapsubtype3
1.0
0.8
0.6
0.4
0.2
0.0
0
1
weapsubtype3

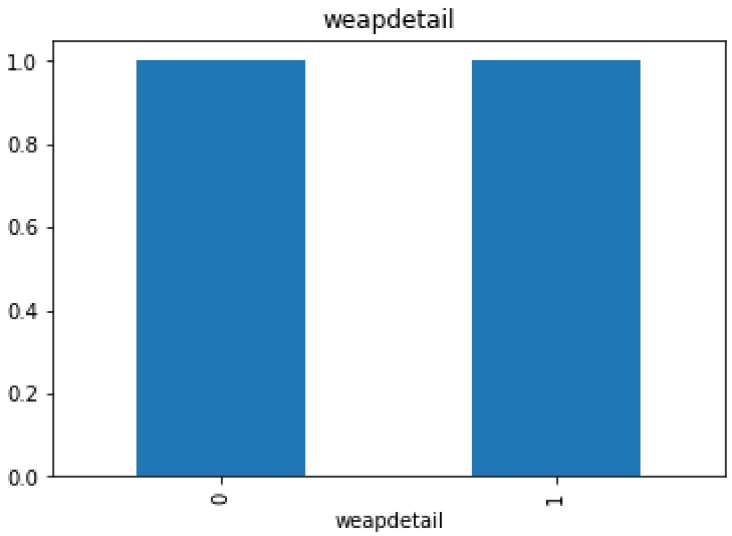

weapsubtype3_txt
1.0
0.8
0.6
0.4
0.2
0.0
0
1
weapsubtype3_txt

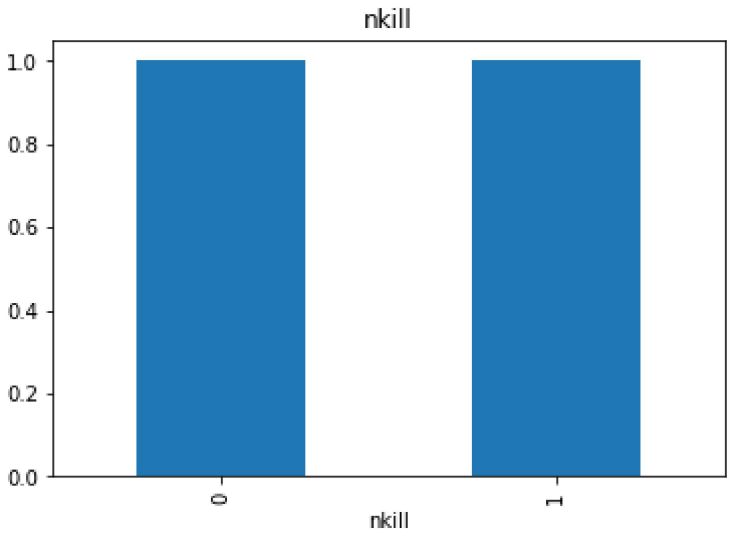

weaptype4
1.0
0.8
0.6
0.4
0.2
0.0
0
1
weaptype4

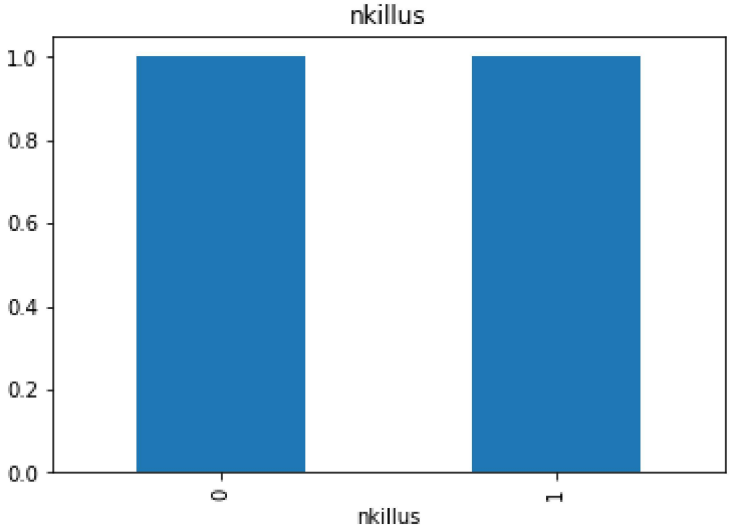

weaptype4_txt
1.0
0.8
0.6
0.4
0.2
0.0
0
1
weaptype4_txt

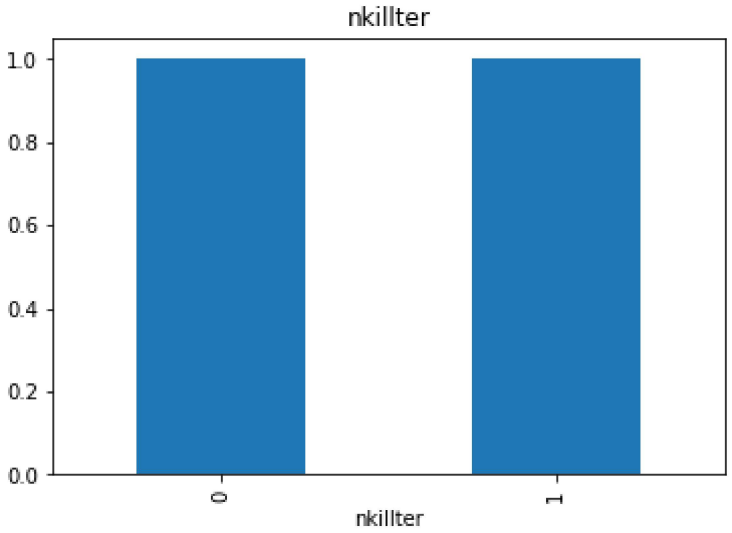

weapsubtype4
1.0
0.8
0.6
0.4
0.2
0.0
0
1
weapsubtype4

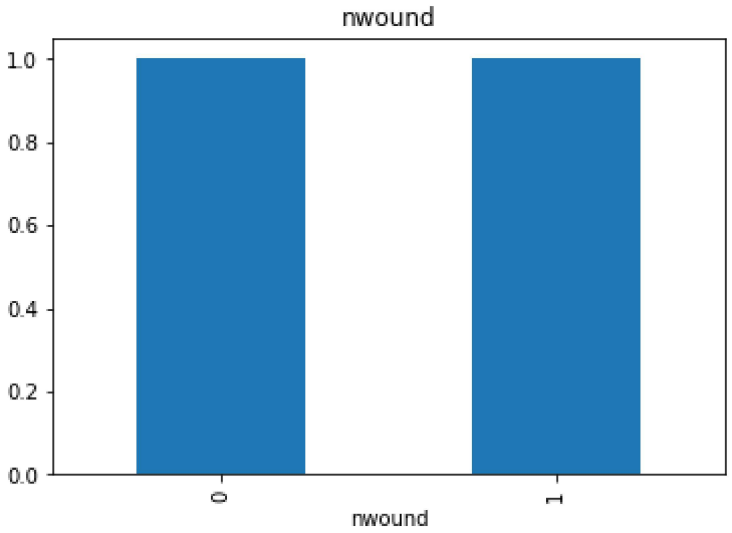

weapsubtype4_txt
1.0
0.8
0.6
0.4
0.2
0.0
0
1
weapsubtype4_txt

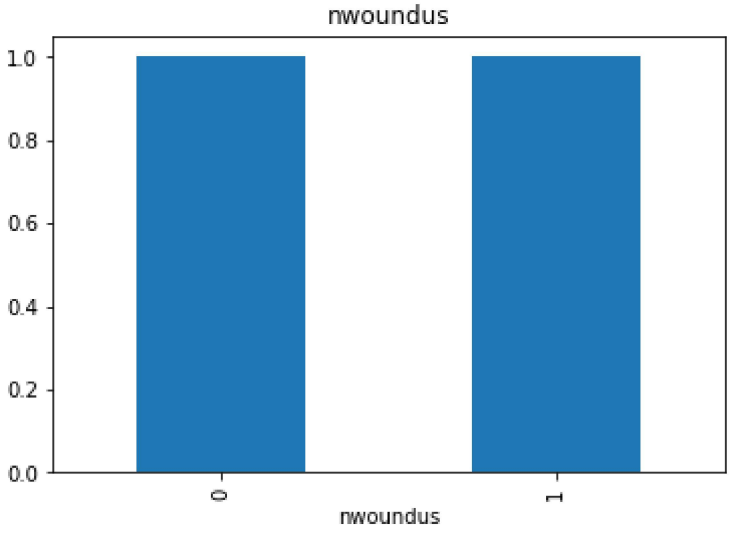

weapdetail
1.0
0.8
0.6
0.4
0.2
0.0
0
1
weapdetail

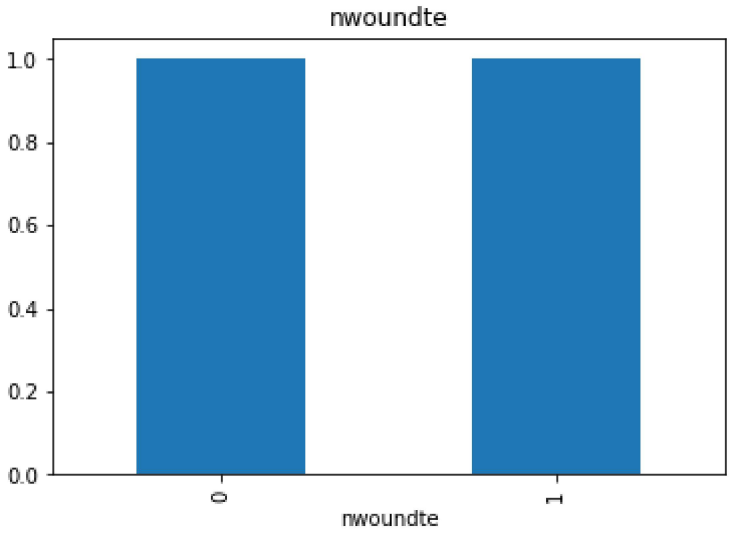

nkill
1.0
0.8
0.6
0.4
0.2
0.0
0
1
nkill

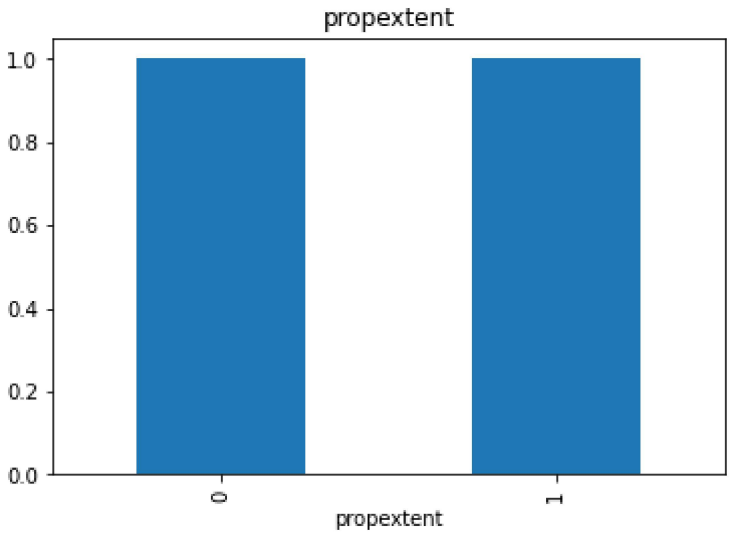

nkillus
1.0
0.8
0.6
0.4
0.2
0.0
0
1
nkillus

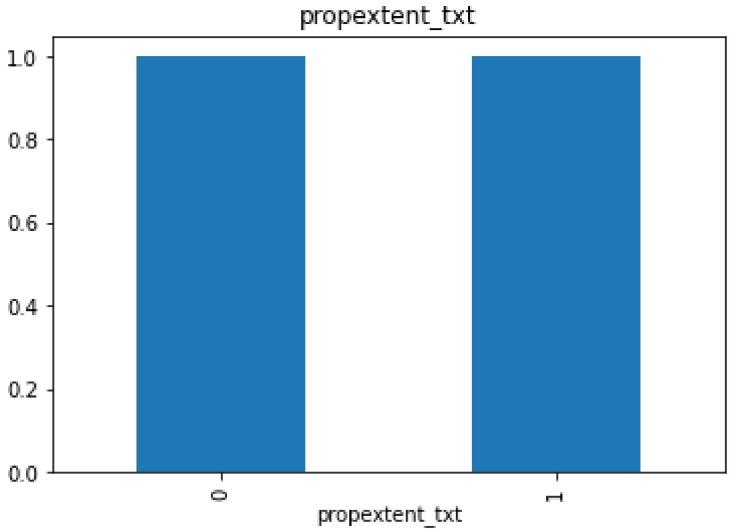

nkillter
1.0
0.8
0.6
0.4
0.2
0.0
0
1
nkillter

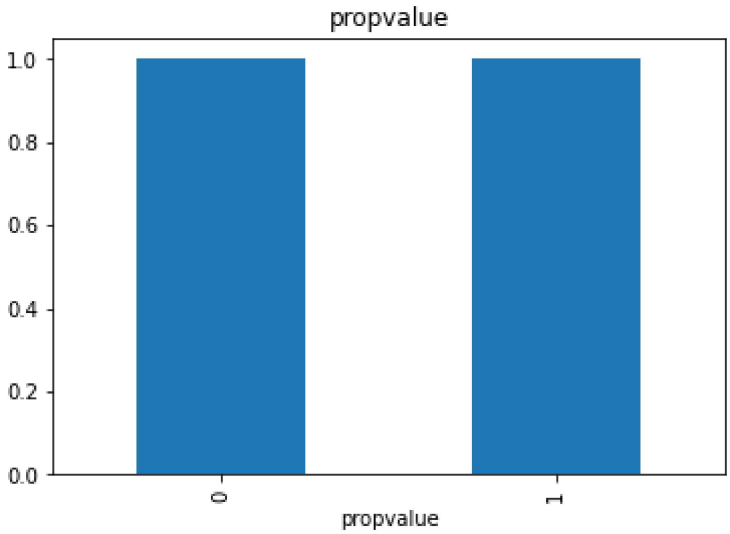

nwound
1.0
0.8
0.6
0.4
0.2
0.0
0
1
nwound

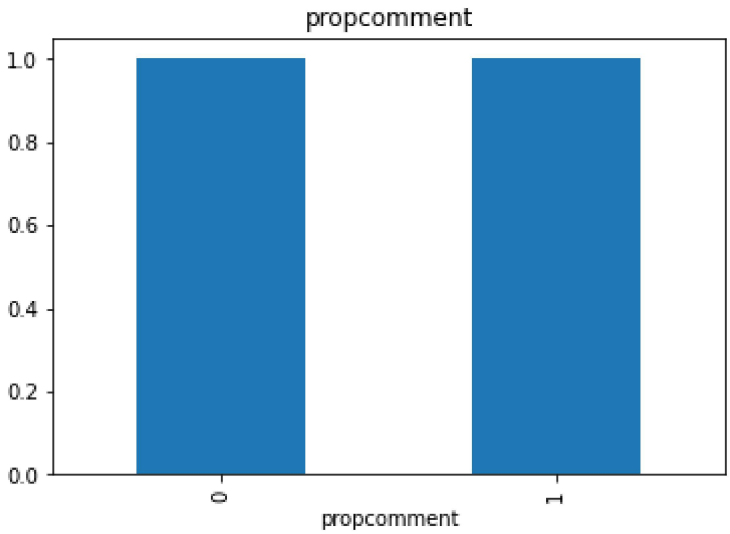

nwoundus
1.0
0.8
0.6
0.4
0.2
0.0
0
1
nwoundus

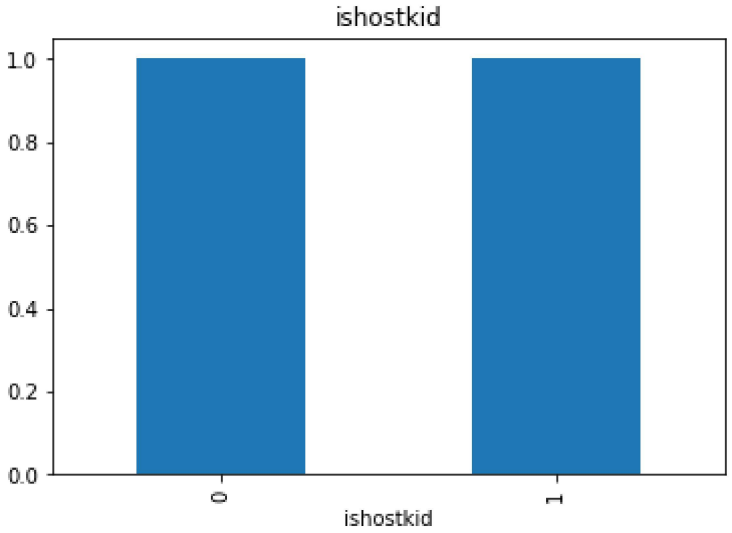

nwoundte
1.0
0.8
0.6
0.4
0.2
0.0
0
1
nwoundte

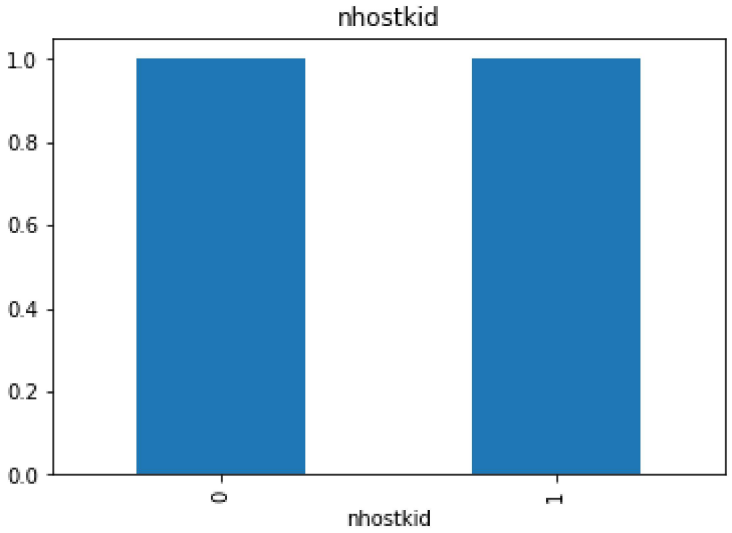

propextent
1.0
0.8
0.6
0.4
0.2
0.0
0
1
propextent

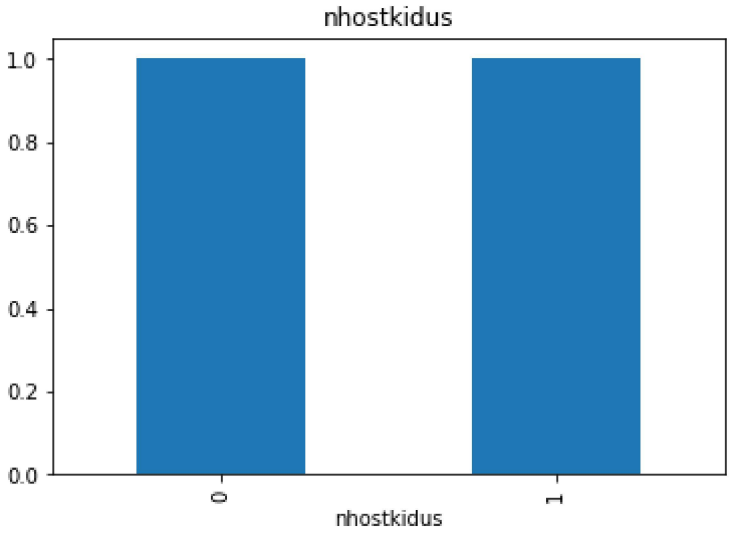

propextent_txt
1.0
0.8
0.6
0.4
0.2
0.0
0
1
propextent_txt

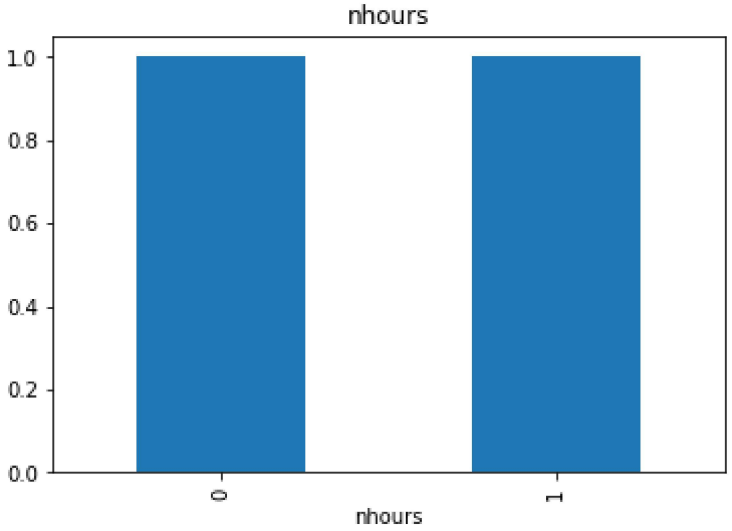

propvalue
1.0
0.8
0.6
0.4
0.2
0.0
0
1
propvalue

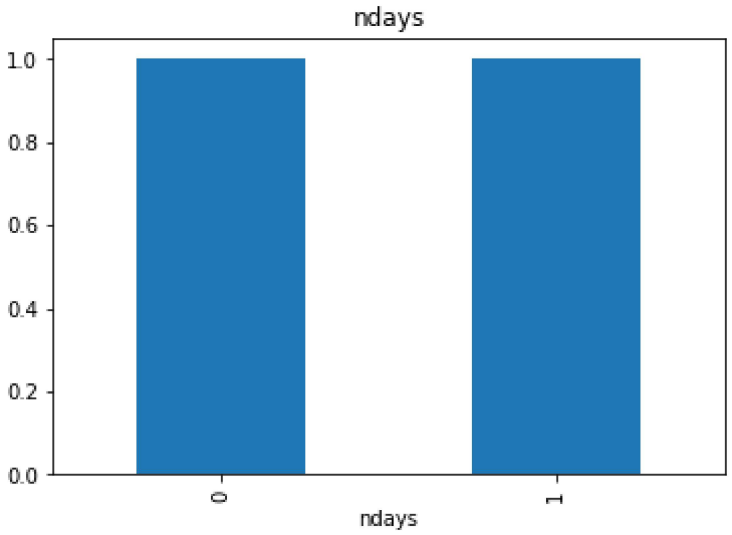

propcomment
1.0
0.8
0.6
0.4
0.2
0.0
0
1
propcomment

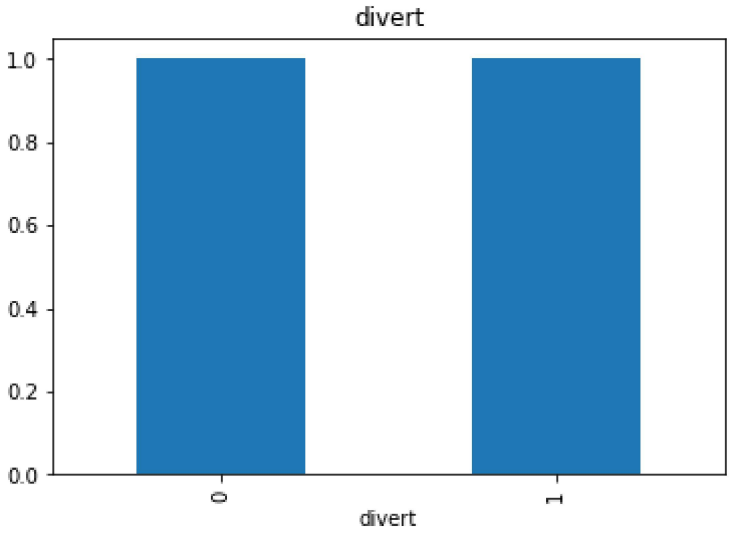

ishostkid
1.0
0.8
0.6
0.4
0.2
0.0
0
1
ishostkid

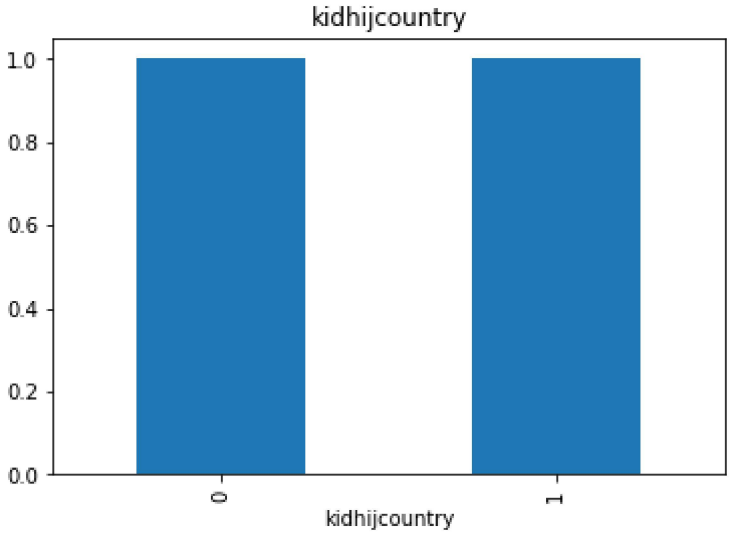

nhostkid
1.0
0.8
0.6
0.4
0.2
0.0
0
1
nhostkid

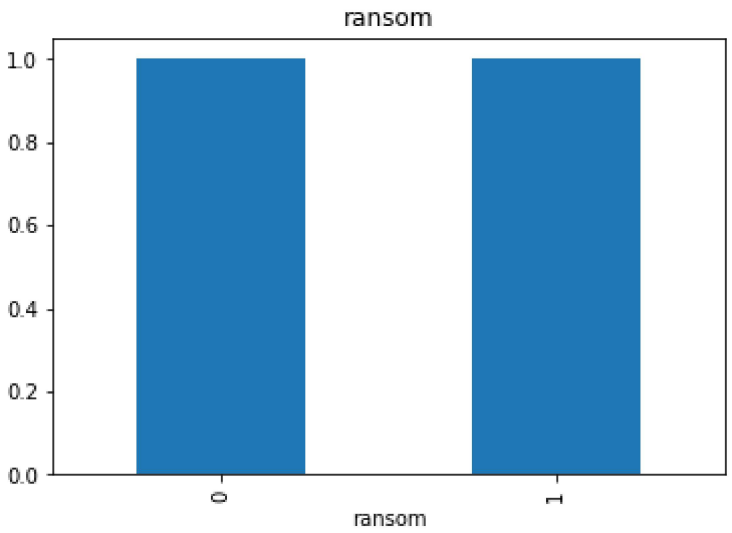

nhostkidus
1.0
0.8
0.6
0.4
0.2
0.0
0
1
nhostkidus

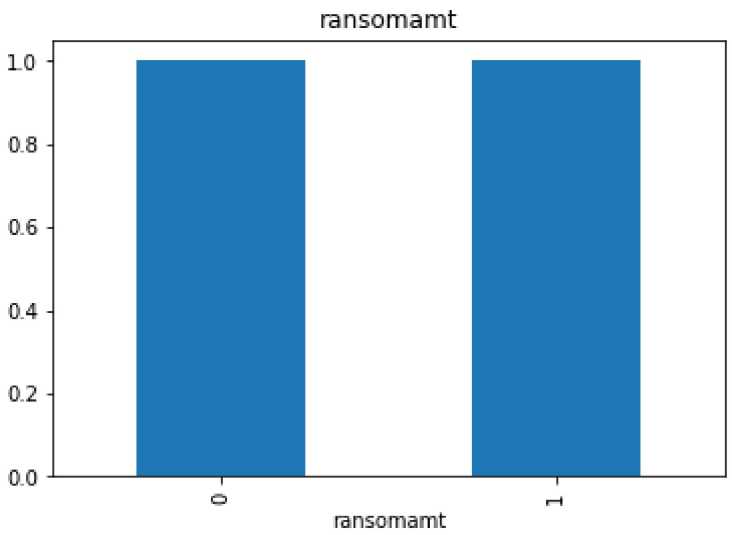

nhours
1.0
0.8
0.6
0.4
0.2
0.0
0
1
nhours

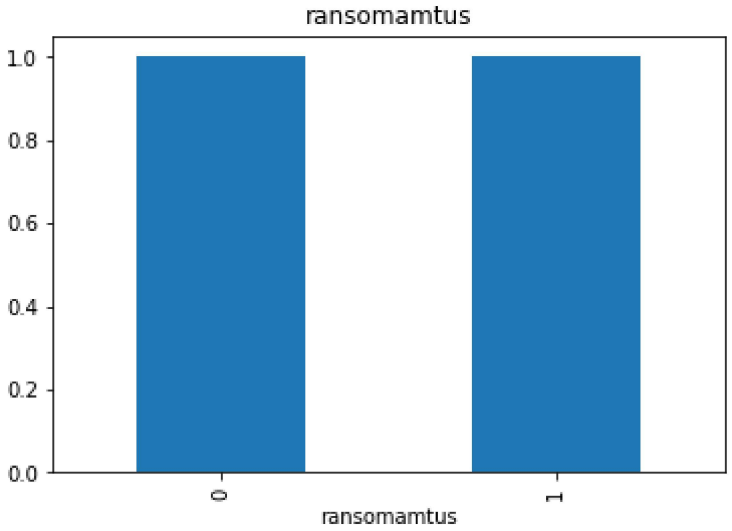

ndays
1.0
0.8
0.6
0.4
0.2
0.0
0
1
ndays

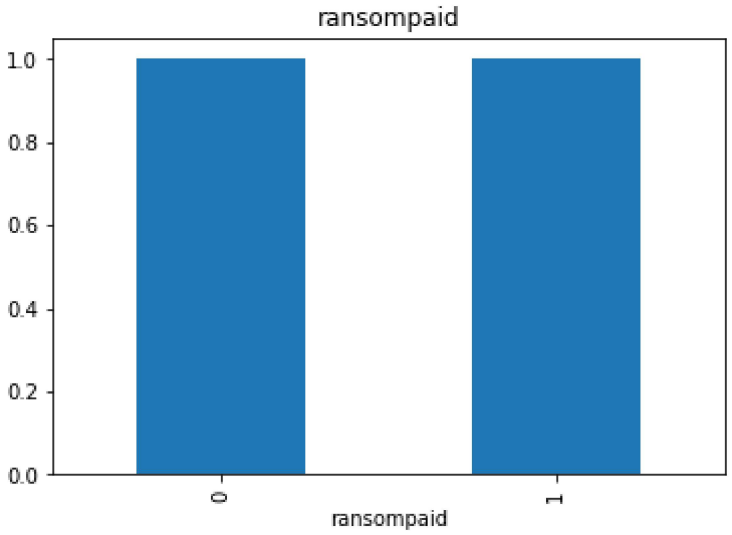

divert
1.0
0.8
0.6
0.4
0.2
0.0
0
1
divert

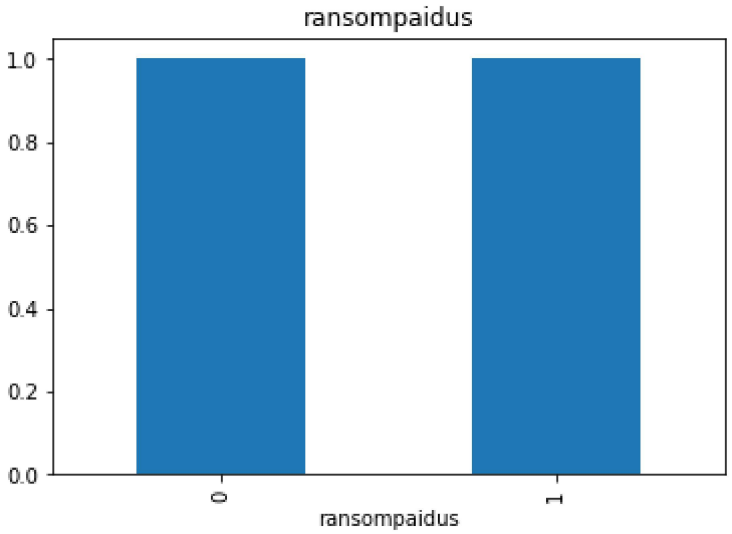

kidhijcountry
1.0
0.8
0.6
0.4
0.2
0.0
0
1
kidhijcountry

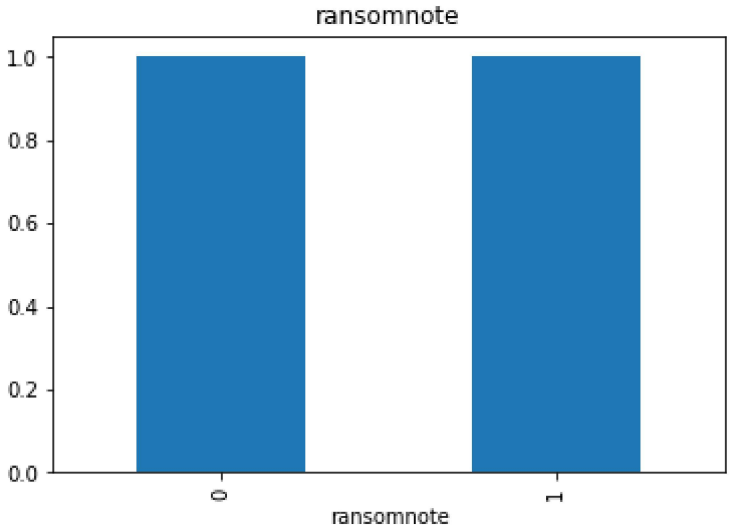

ransom
1.0
0.8
0.6
0.4
0.2
0.0
0
1
ransom

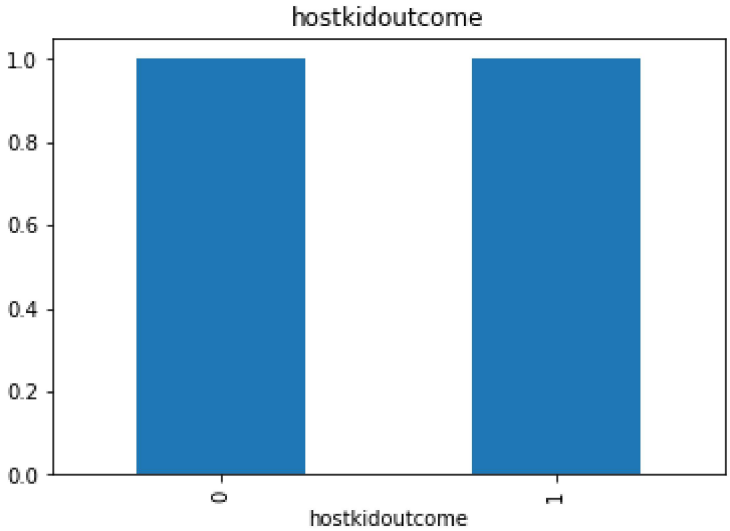

ransomamt
1.0
0.8
0.6
0.4
0.2
0.0
0
1
ransomamt

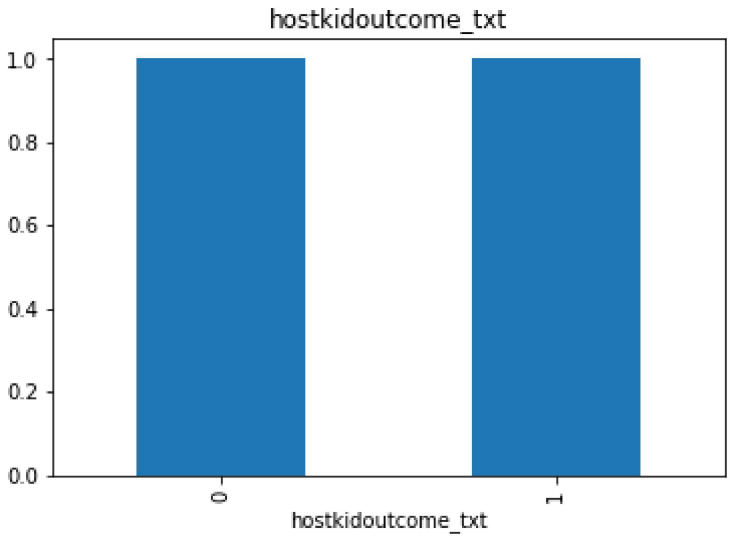

ransomamtus
1.0
0.8
0.6
0.4
0.2
0.0
0
1
ransomamtus

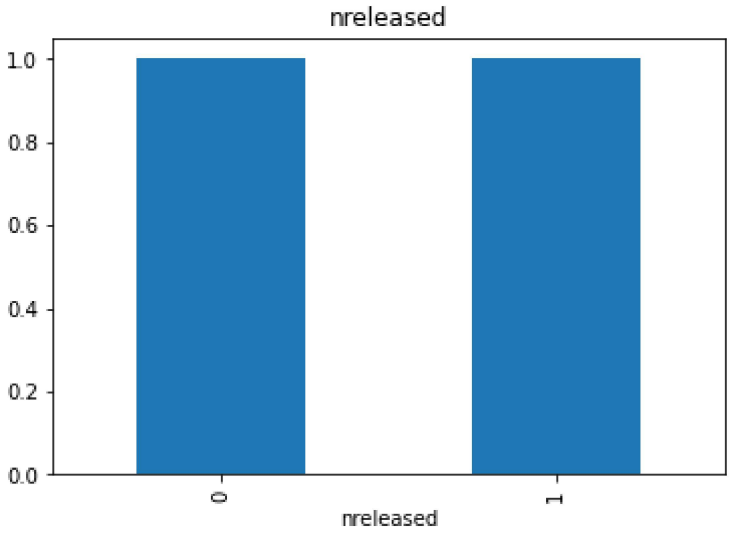

ransompaid
1.0
0.8
0.6
0.4
0.2
0.0
0
1
ransompaid

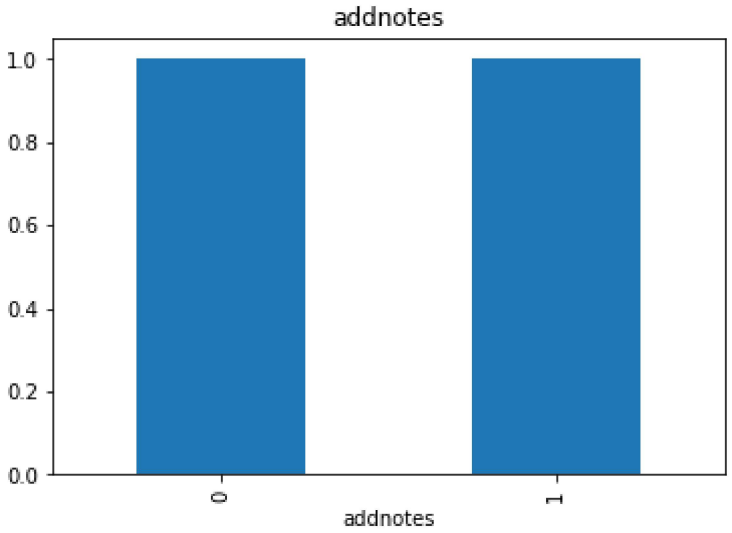

ransompaidus
1.0
0.8
0.6
0.4
0.2
0.0
0
1
ransompaidus

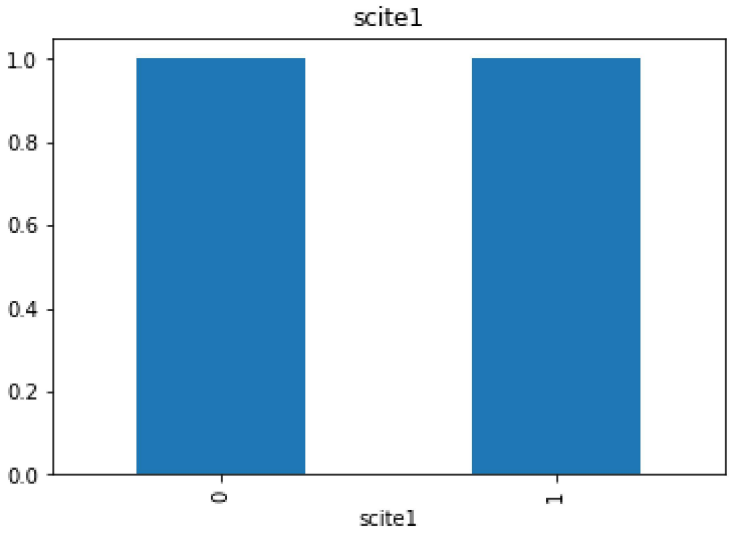

ransomnote
1.0
0.8
0.6
0.4
0.2
0.0
0
1
ransomnote

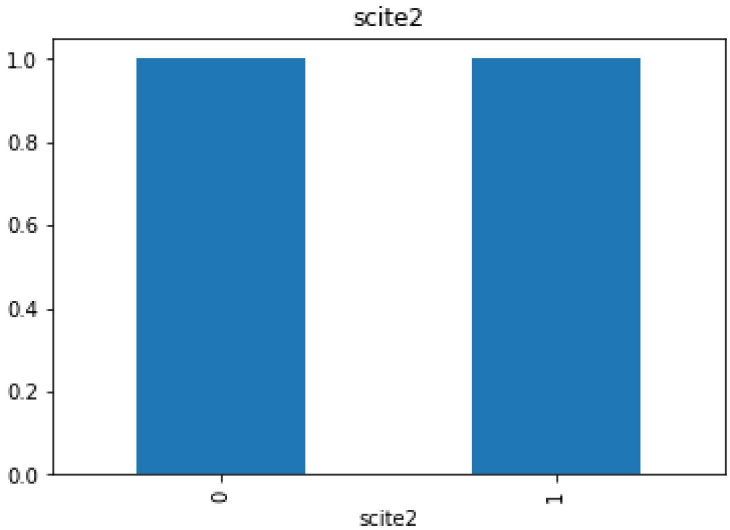

hostkidoutcome
1.0
0.8
0.6
0.4
0.2
0.0
0
1
hostkidoutcome

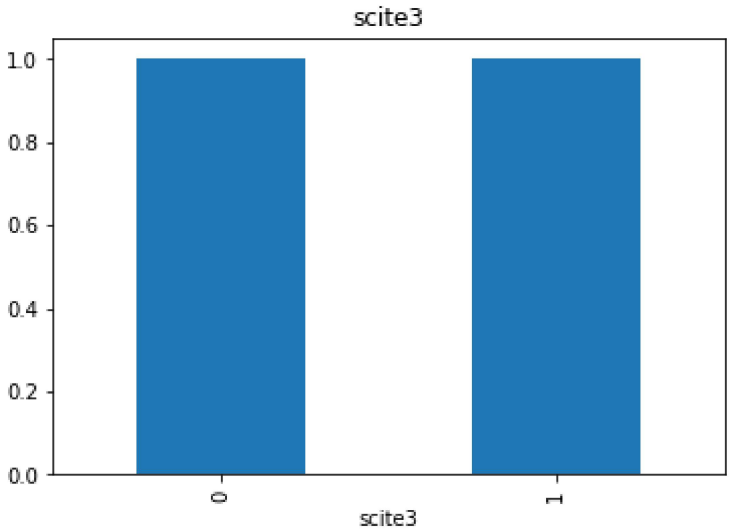

hostkidoutcome_txt
1.0
0.8
0.6
0.4
0.2
0.0
0
1
hostkidoutcome_txt

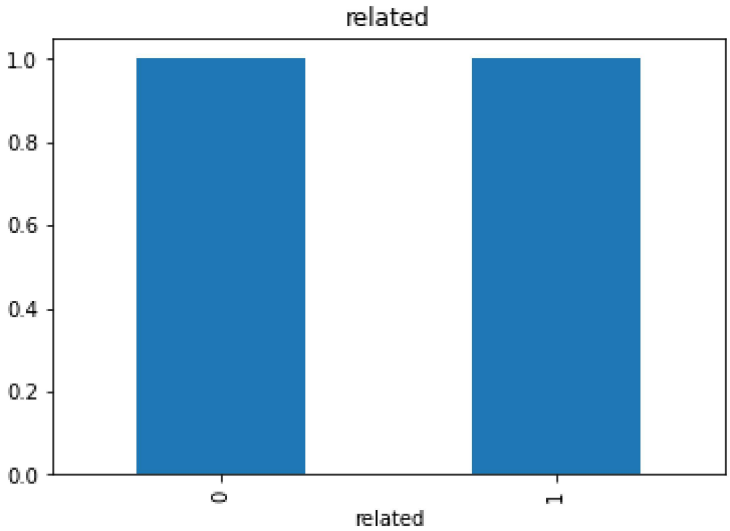

nreleased
1.0
0.8
0.6
0.4
0.2
0.0
0
1
nreleased


addnotes
1.0
0.8
0.6
0.4
0.2
0.0
0
1
addnotes


scite1
1.0
0.8
0.6
0.4
0.2
0.0
0
1
scite1







Numerical Variables

In [8]:
```
# list of numerical variables
numerical_features = [feature for feature in dataset.columns if dataset[feature].dty

print('Number of numerical variables: ', len(numerical_features))

# visualise the numerical variables
dataset[numerical_features].head()
```

```
Number of numerical variables:  77
```

Out[8]:

	eventid	iyear	imonth	iday	extended	country	region	latitude	longitude	specificity
0	197000000001	1970	7	2	0	58	2	18.456792	-69.951164	1.0
1	197000000002	1970	0	0	0	130	1	19.371887	-99.086624	1.0
2	197001000001	1970	1	0	0	160	5	15.478598	120.599741	4.0
3	197001000002	1970	1	0	0	78	8	37.997490	23.762728	1.0
4	197001000003	1970	1	0	0	101	4	33.580412	130.396361	1.0

5 rows × 77 columns

Temporal Variable

In [9]:
```
year_feature = [feature for feature in numerical_features if 'Yr' in feature or 'Yea

year_feature
```

Out[9]:
```
['iyear']
```

In [10]:
```
for feature in year_feature:
    print(feature, dataset[feature].unique())
```

```
iyear [1970 1971 1972 1973 1974 1975 1976 1977 1978 1979 1980 1981 1986 1982
 1983 1984 1985 1987 1988 1989 1990 1991 1992 1994 1995 1996 1997 1998
 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012
 2013 2014 2015 2016 2017]
```

In [11]:
```
discrete_feature=[feature for feature in numerical_features if len(dataset[feature].
print("Discrete Variables Count: {}".format(len(discrete_feature)))
```

```
Discrete Variables Count: 47
```

In [12]:
```
discrete_feature
```

Out[12]:
```
['imonth',
 'extended',
 'region',
 'specificity',
 'vicinity',
 'crit1',
 'crit2',
 'crit3',
 'doubtterr',
 'alternative',
 'multiple',
 'success',
 'suicide',
 'attacktype1',
 'attacktype2',
 'attacktype3',
 'targtype1',
 'targtype2',
 'targtype3',
 'guncertain1',
 'guncertain2',
 'guncertain3',
 'individual',
 'claimed',
 'claimmode',
 'claim2',
 'claimmode2',
 'claim3',
```

```
 'claimmode3',
 'compclaim',
 'weaptype1',
 'weaptype2',
 'weaptype3',
 'weapsubtype3',
 'weaptype4',
 'weapsubtype4',
 'property',
 'propextent',
 'ishostkid',
 'ransom',
 'ransomamtus',
 'ransompaidus',
 'hostkidoutcome',
 'INT_LOG',
 'INT_IDEO',
 'INT_MISC',
 'INT_ANY']
```

In [13]:
```
dataset[discrete_feature].head()
```

Out[13]:

	imonth	extended	region	specificity	vicinity	crit1	crit2	crit3	doubtterr	alternative	...	pro
0	7	0	2	1.0	0	1	1	1	0.0	NaN	...	
1	0	0	1	1.0	0	1	1	1	0.0	NaN	...	
2	1	0	5	4.0	0	1	1	1	0.0	NaN	...	
3	1	0	8	1.0	0	1	1	1	0.0	NaN	...	
4	1	0	4	1.0	0	1	1	1	-9.0	NaN	...	

5 rows × 47 columns

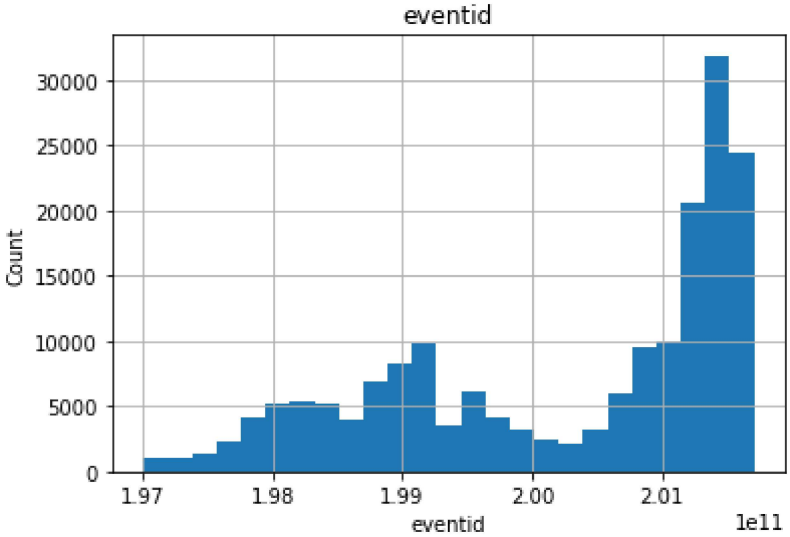
In [ ]:
```
for feature in discrete_feature:
    data=dataset.copy()
    data.groupby(feature)['success'].plot.bar()
    plt.xlabel(feature)
    plt.ylabel('success')
    plt.title(feature)
    plt.show()
```

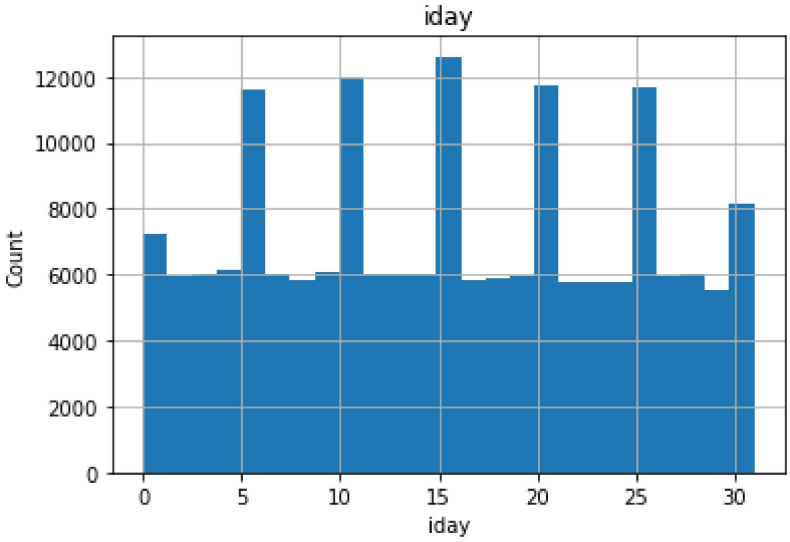
Continuous Variable

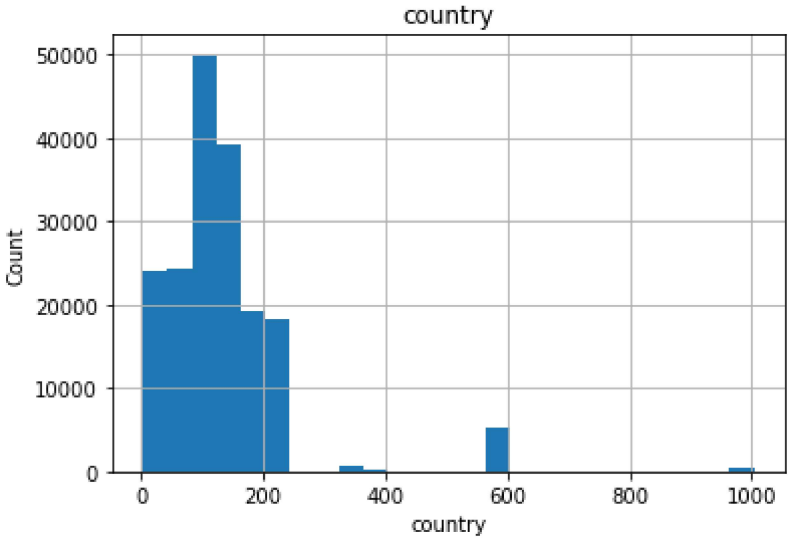
In [14]:
```
continuous_feature=[feature for feature in numerical_features if feature not in disc
print("Continuous feature Count {}".format(len(continuous_feature)))
```

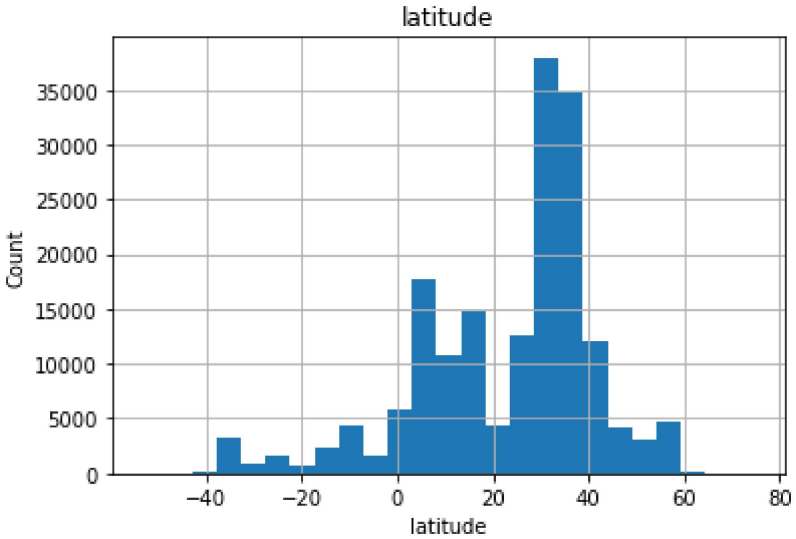
```
Continuous feature Count 29
```

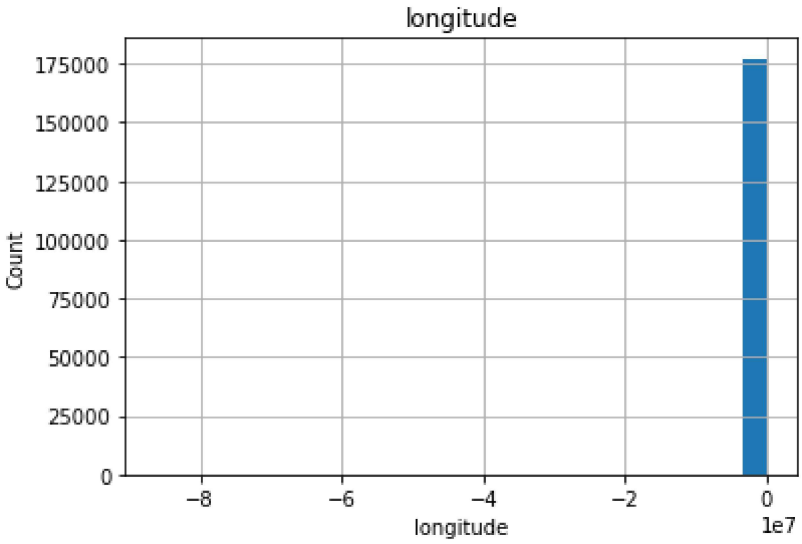
In [15]:
```
for feature in continuous_feature:
    data=dataset.copy()
    data[feature].hist(bins=25)
    plt.xlabel(feature)
    plt.ylabel("Count")
    plt.title(feature)
    plt.show()
```

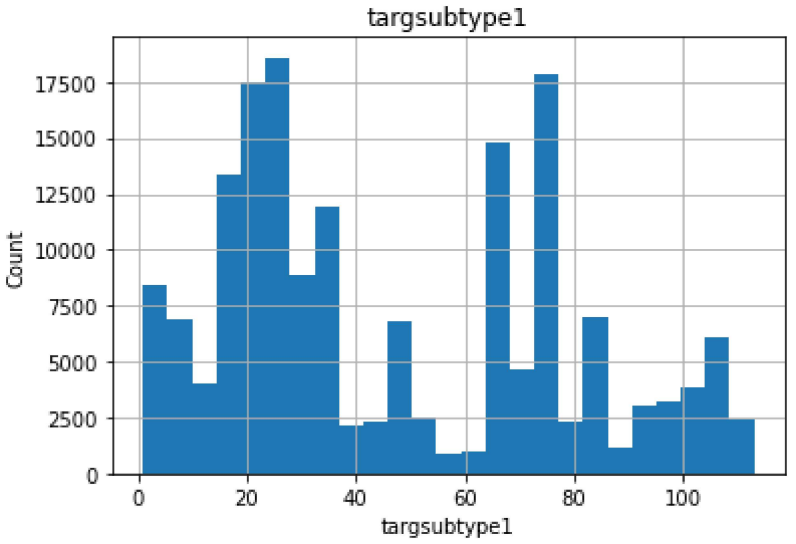

eventid
Count
30000
25000
20000
15000
10000
5000
0
1.97
1.98
1.99
2.00
2.01
eventid
1e11

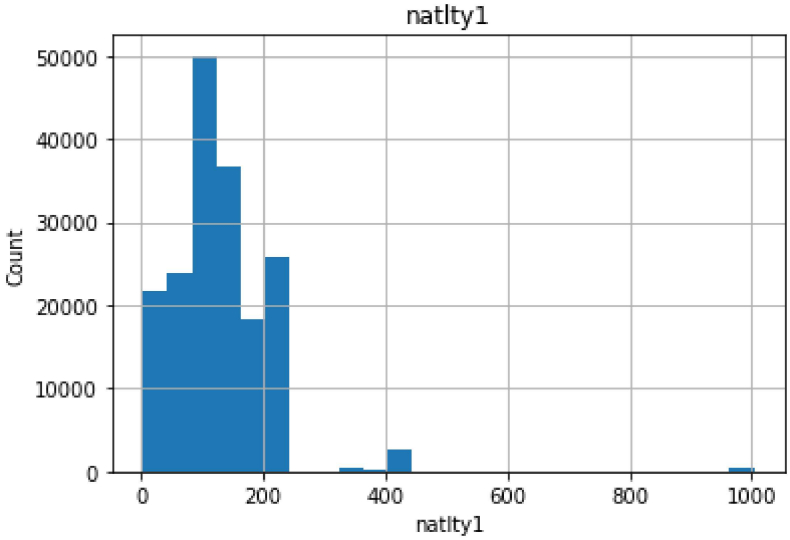

iday
Count
12000
10000
8000
6000
4000
2000
0
0
5
10
15
20
25
30
iday

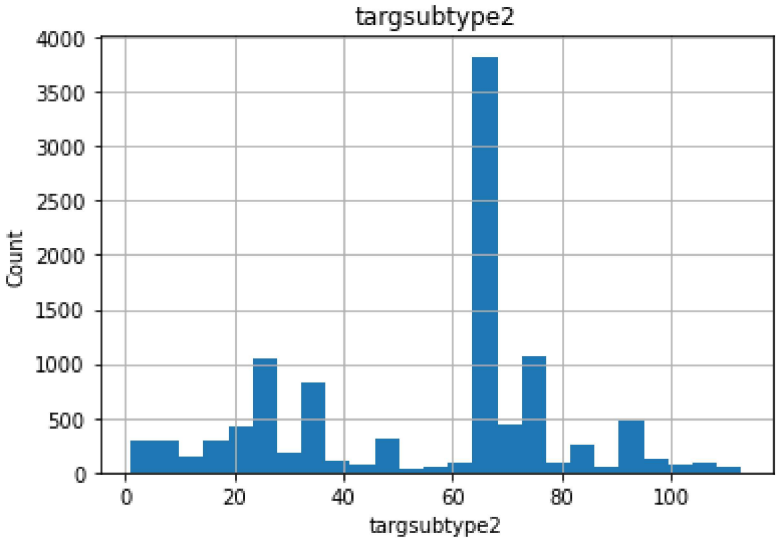

country
Count
50000
40000
30000
20000
10000
0
0
200
400
600
800
1000
country

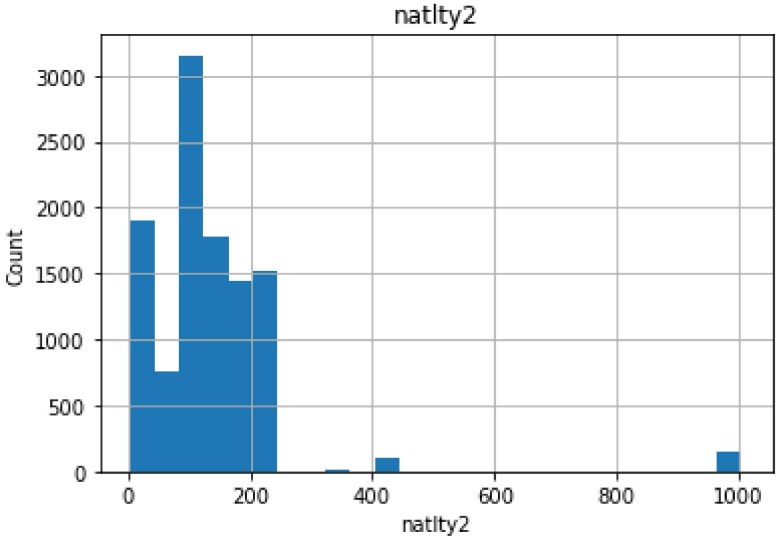

latitude
Count
35000
30000
25000
20000
15000
10000
5000
0
−40
−20
0
20
40
60
80
latitude

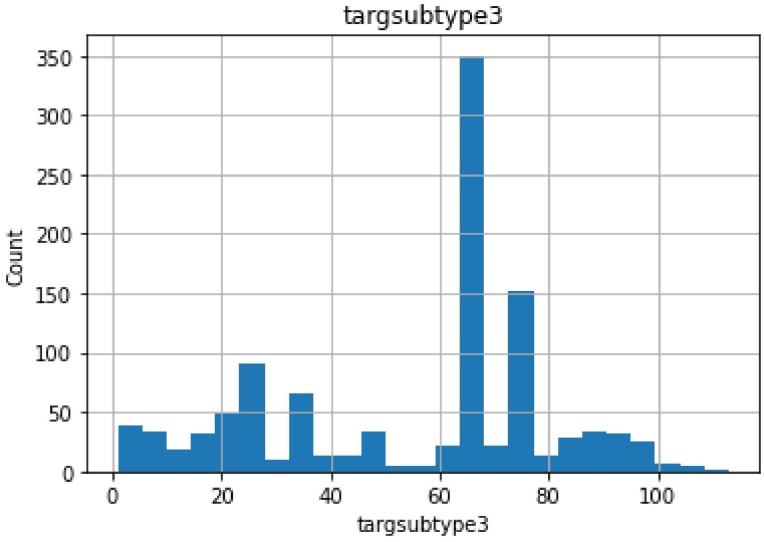

longitude
Count
175000
150000
125000
100000
75000
50000
25000
0
−8
−6
−4
−2
0
1e7
longitude

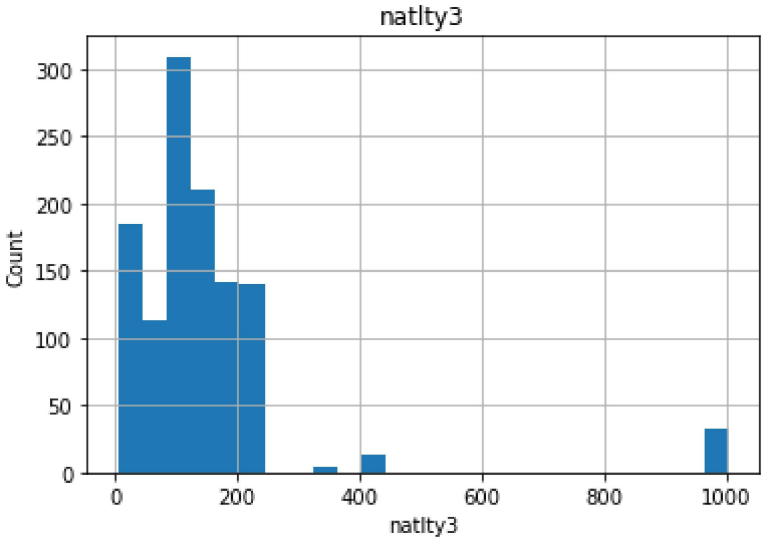

targsubtype1
Count
17500
15000
12500
10000
7500
5000
2500
0
0
20
40
60
80
100
targsubtype1

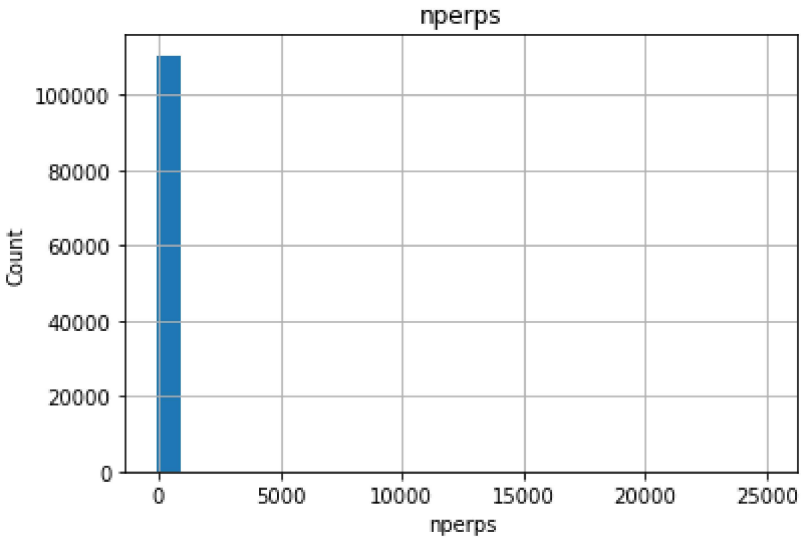

natlty1
Count
natlty1
0
10000
20000
30000
40000
50000
0
200
400
600
800
1000

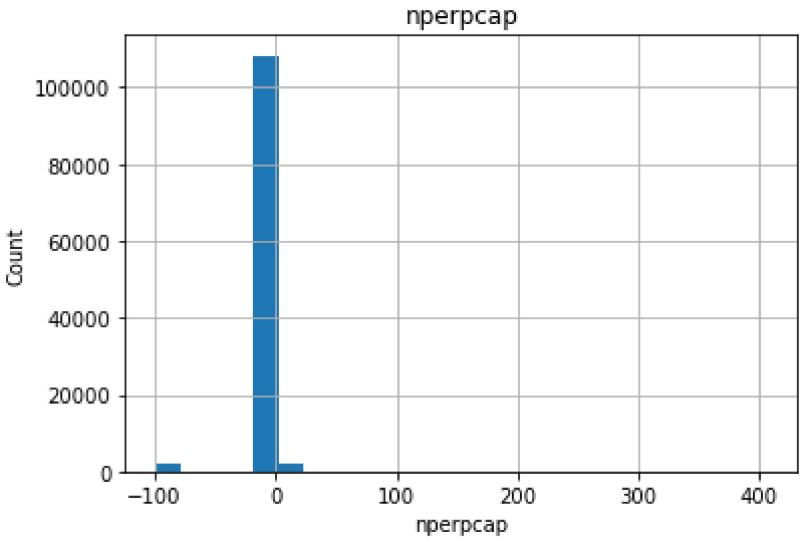

targsubtype2
Count
targsubtype2
0
500
1000
1500
2000
2500
3000
3500
4000
0
20
40
60
80
100

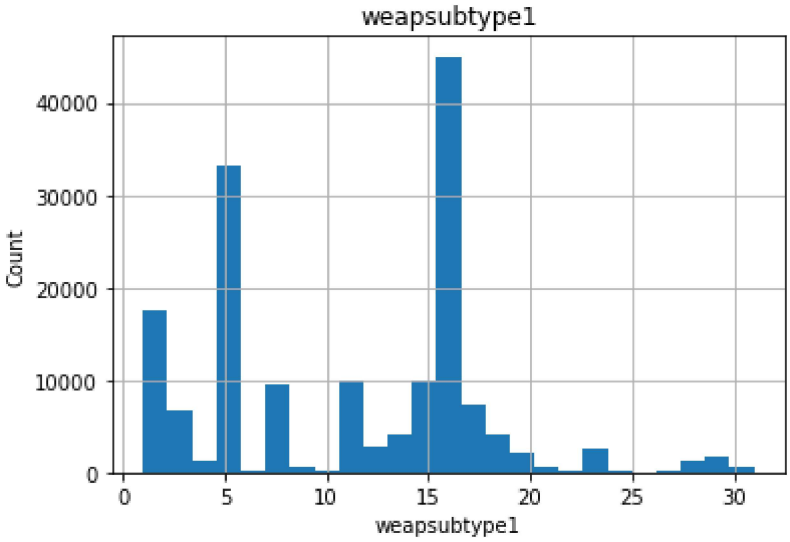

natlty2
Count
natlty2
0
500
1000
1500
2000
2500
3000
0
200
400
600
800
1000

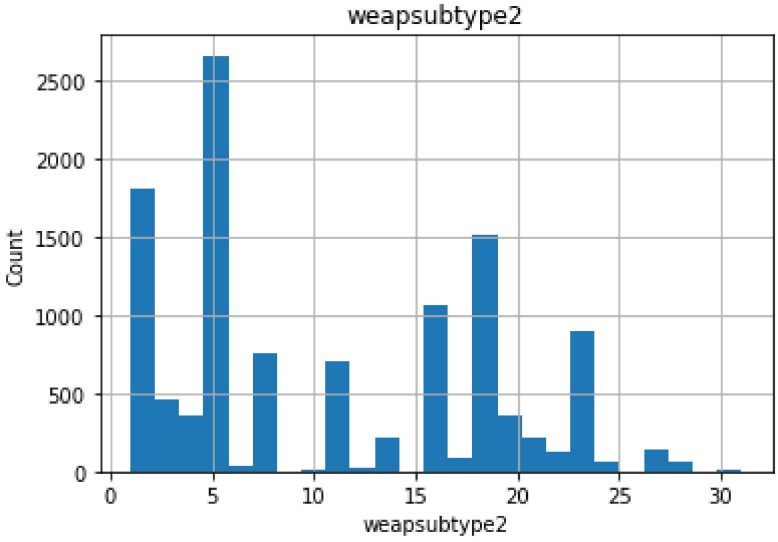

targsubtype3
Count
0
50
100
150
200
250
300
350
0
20
40
60
80
100
targsubtype3

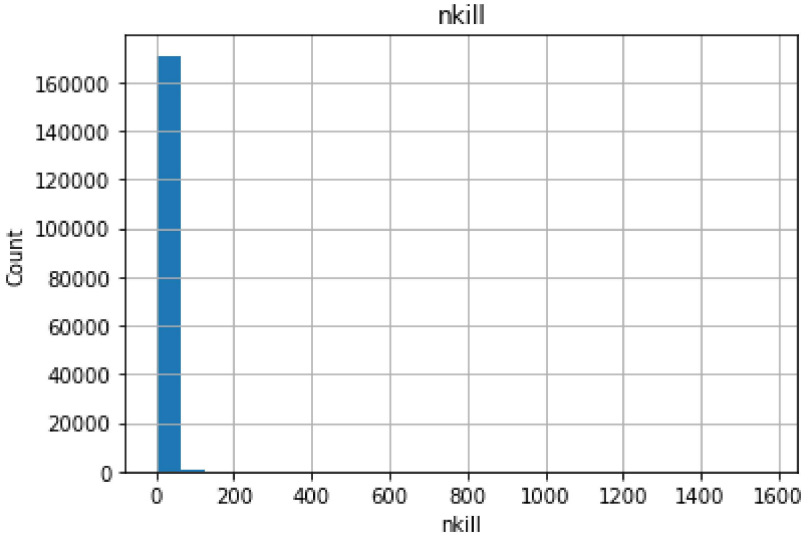

natlty3
Count
0
50
100
150
200
250
300
0
200
400
600
800
1000
natlty3

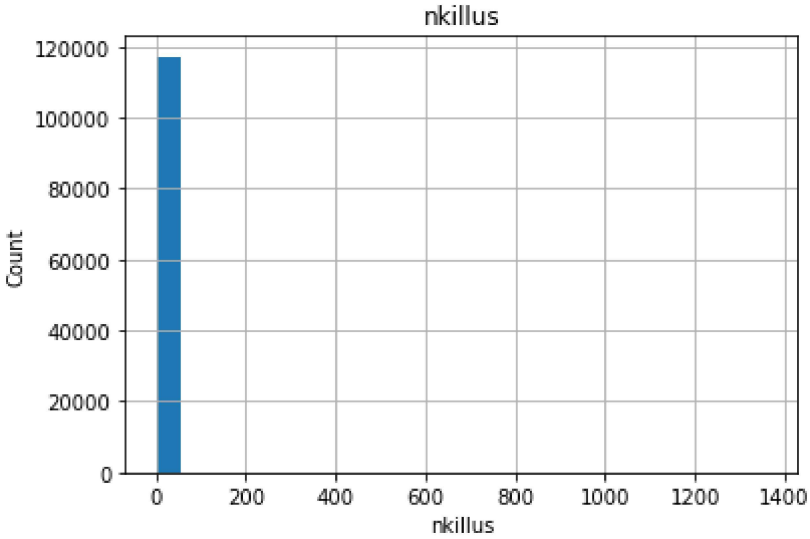

nperps
Count
0
20000
40000
60000
80000
100000
0
5000
10000
15000
20000
25000
nperps

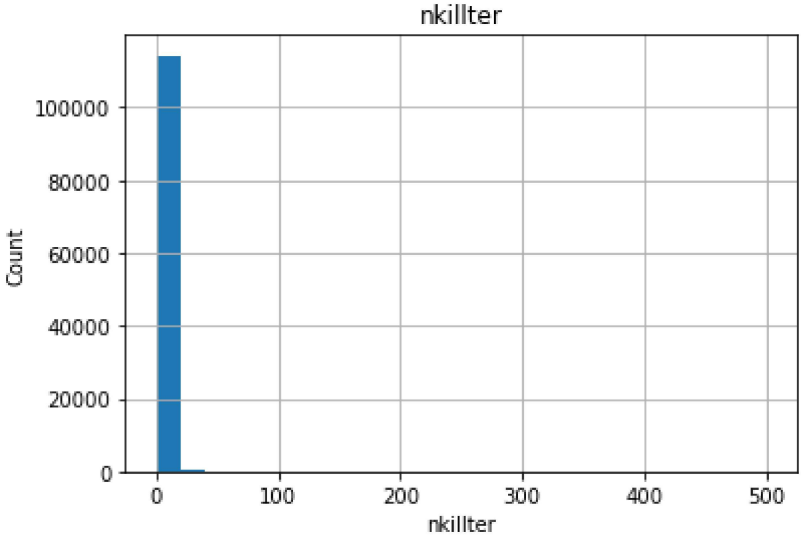

nperpcap
Count
100000
80000
60000
40000
20000
0
−100
0
100
200
300
400
nperpcap

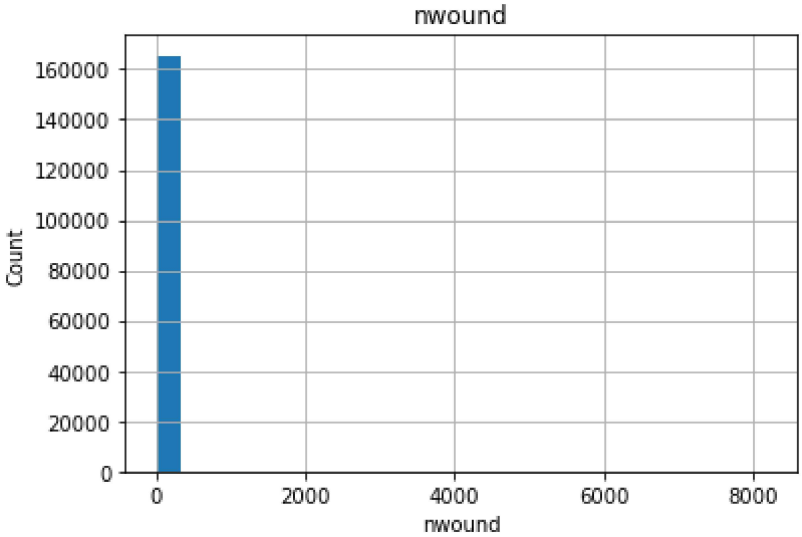

weapsubtype1
Count
40000
30000
20000
10000
0
0
5
10
15
20
25
30
weapsubtype1

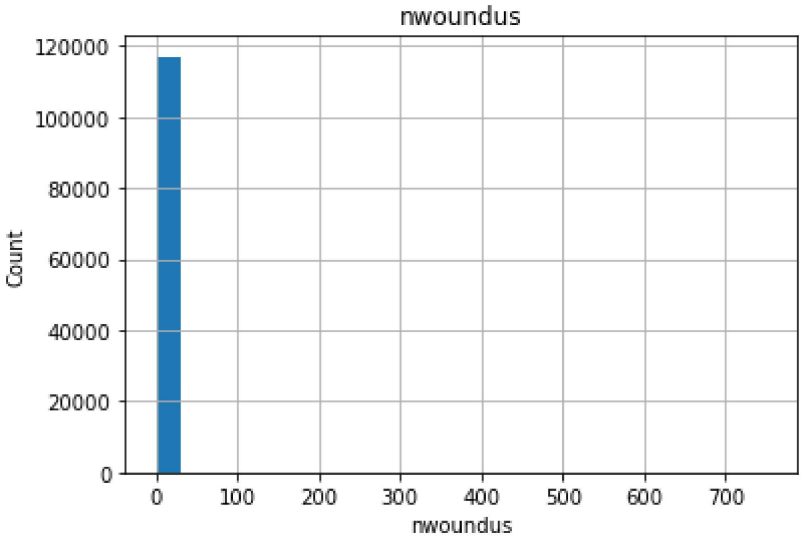

weapsubtype2
Count
2500
2000
1500
1000
500
0
0
5
10
15
20
25
30
weapsubtype2

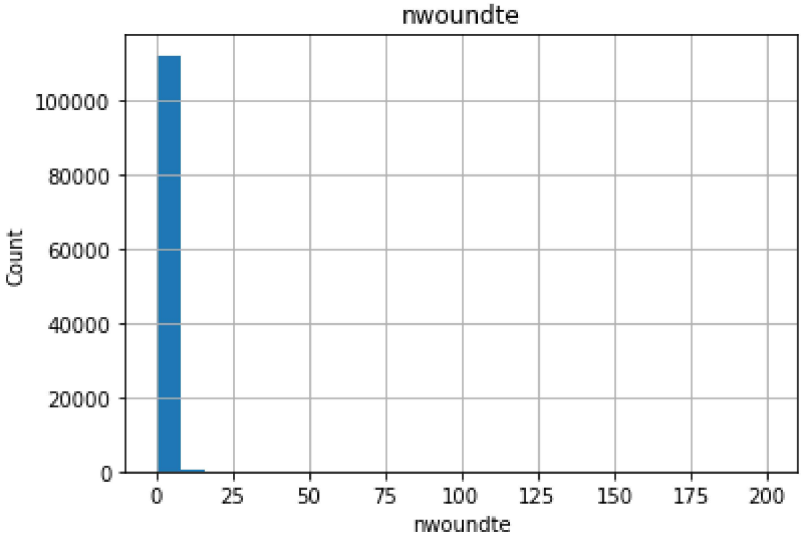

nkill
Count
160000
140000
120000
100000
80000
60000
40000
20000
0
0
200
400
600
800
1000
1200
1400
1600
nkill

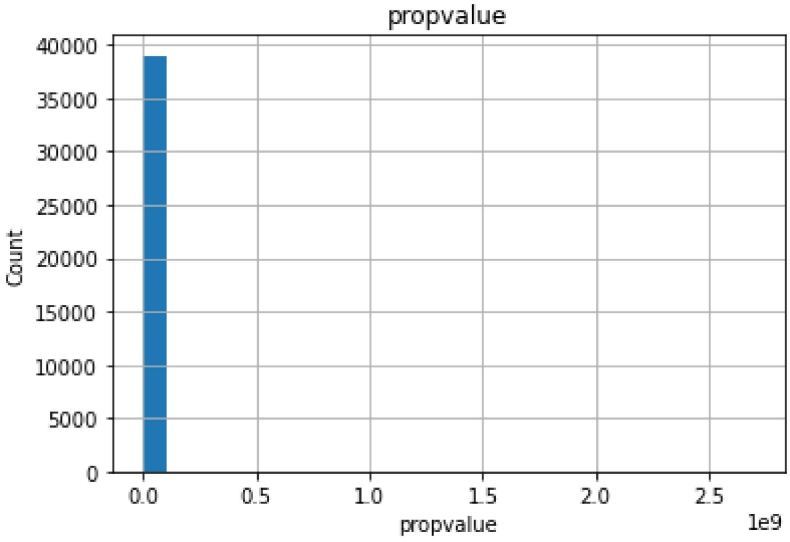

nkillus
Count
120000
100000
80000
60000
40000
20000
0
0
200
400
600
800
1000
1200
1400
nkillus

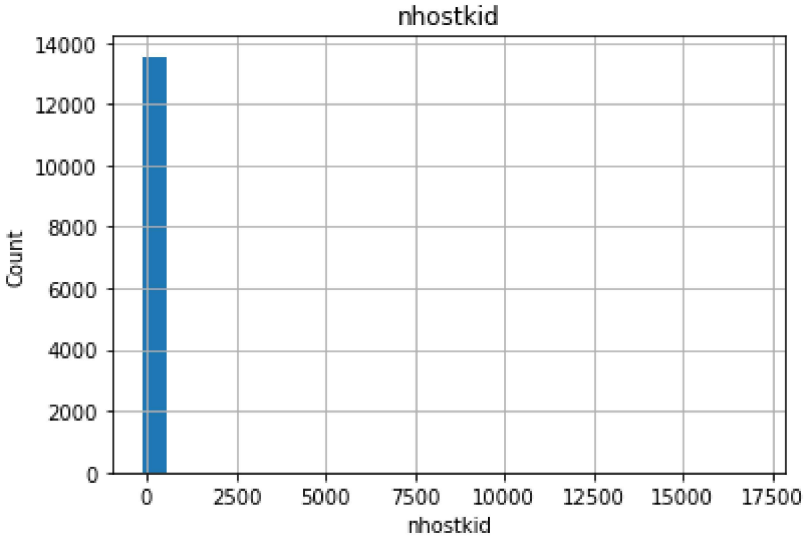

nkillter
Count
100000
80000
60000
40000
20000
0
0
100
200
300
400
500
nkillter

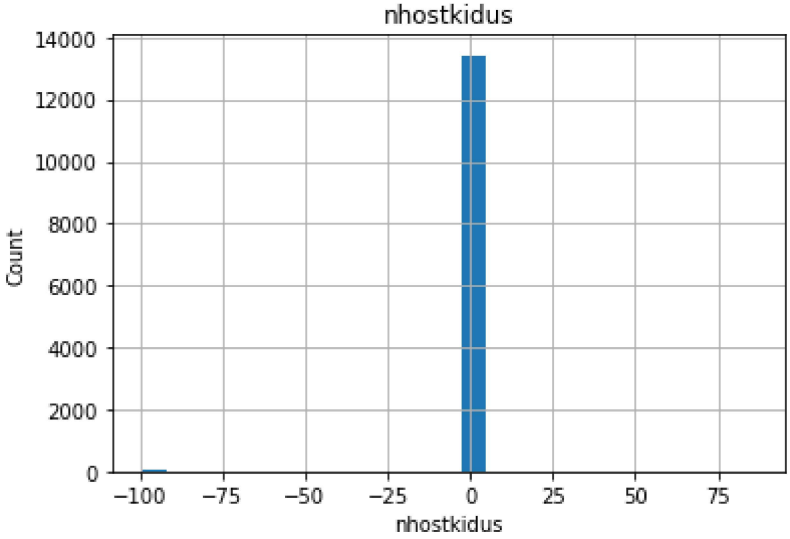

nwound
Count
160000
140000
120000
100000
80000
60000
40000
20000
0
0
2000
4000
6000
8000
nwound

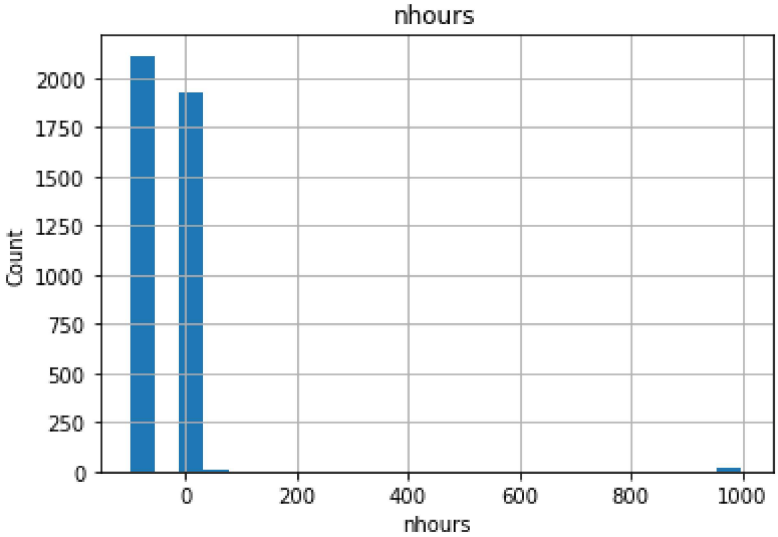

nwoundus
Count
120000
100000
80000
60000
40000
20000
0
0
100
200
300
400
500
600
700
nwoundus

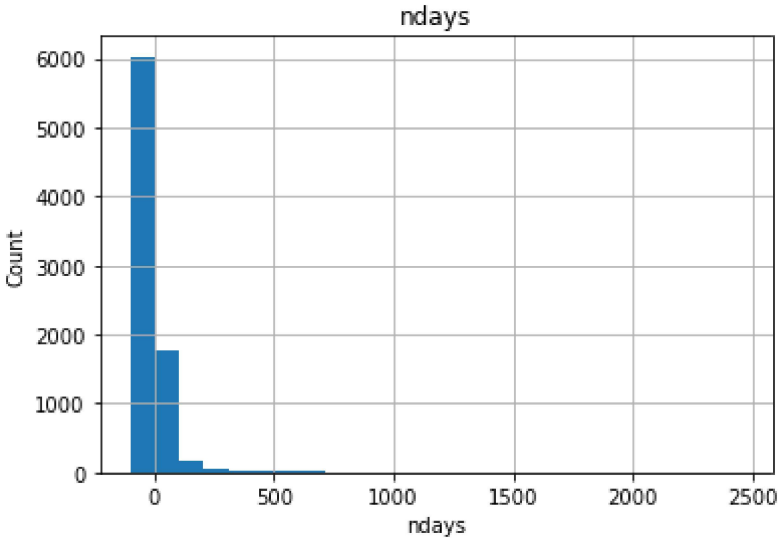

nwoundte
Count
100000
80000
60000
40000
20000
0
0
25
50
75
100
125
150
175
200
nwoundte

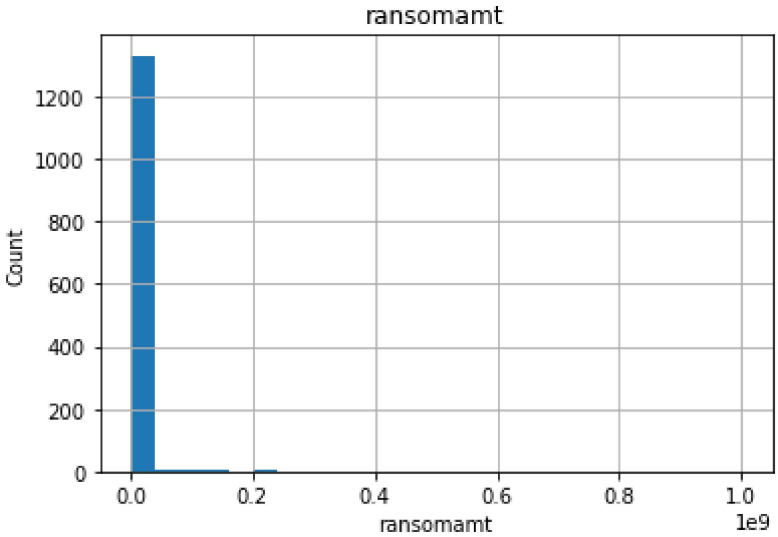

propvalue
Count
40000
35000
30000
25000
20000
15000
10000
5000
0
0.0
0.5
1.0
1.5
2.0
2.5
propvalue
1e9

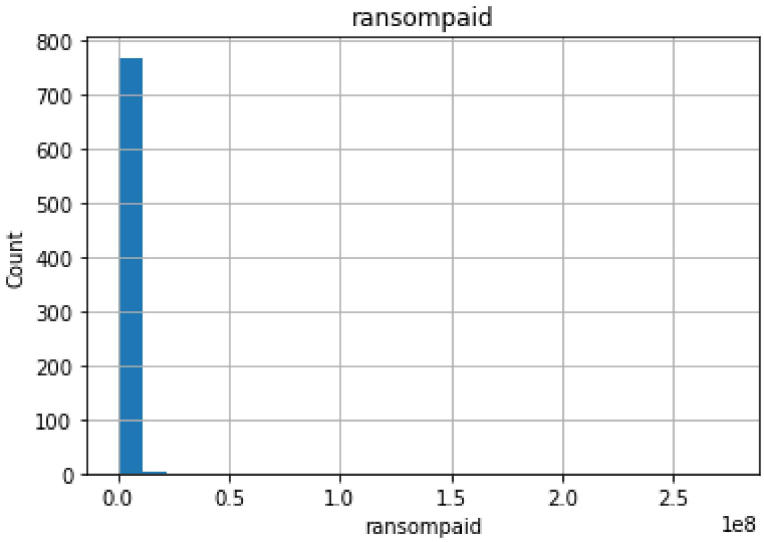

nhostkid
Count
14000
12000
10000
8000
6000
4000
2000
0
0
2500
5000
7500
10000
12500
15000
17500
nhostkid

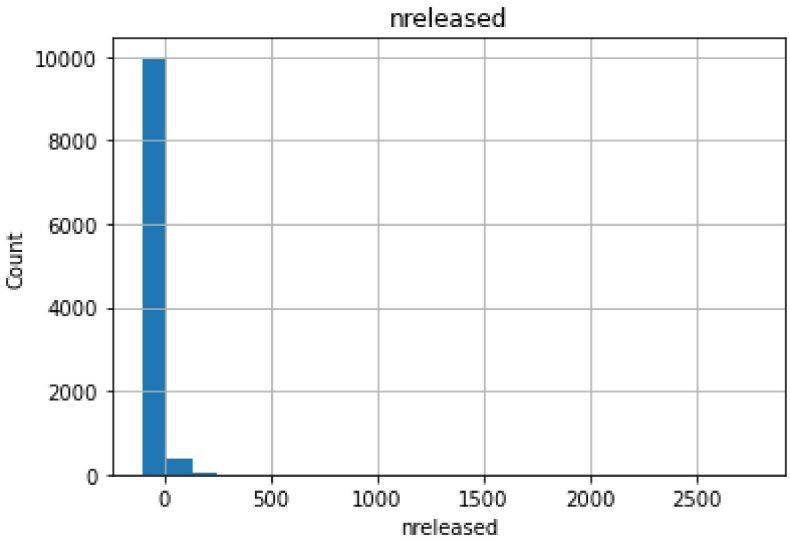

nhostkidus
Count
14000
12000
10000
8000
6000
4000
2000
0
−100
−75
−50
−25
0
25
50
75
nhostkidus


nhours
Count
2000
1750
1500
1250
1000
750
500
250
0
0
200
400
600
800
1000
nhours


ndays
Count
6000
5000
4000
3000
2000
1000
0
0
500
1000
1500
2000
2500
ndays

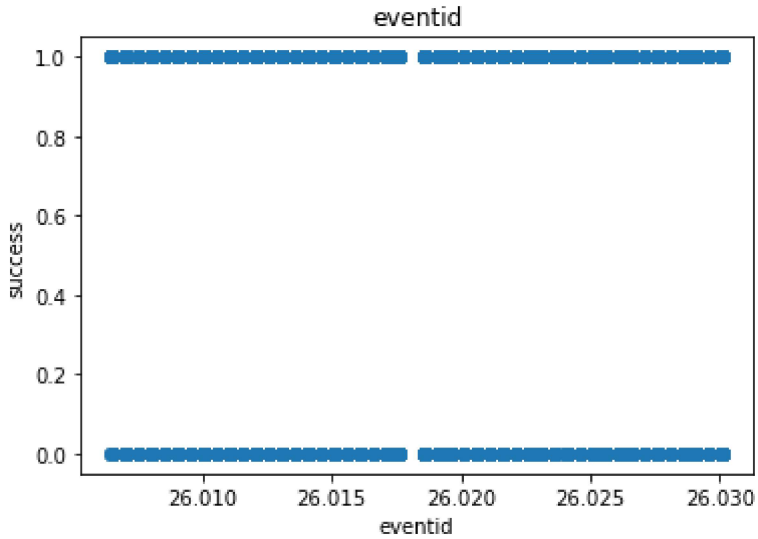

ransomamt
Count
1200
1000
800
600
400
200
0
0.0
0.2
0.4
0.6
0.8
1.0
1e9
ransomamt

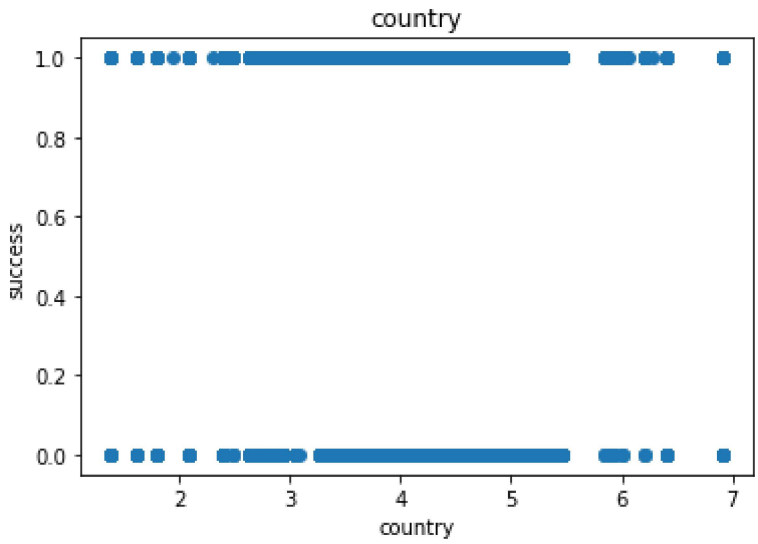




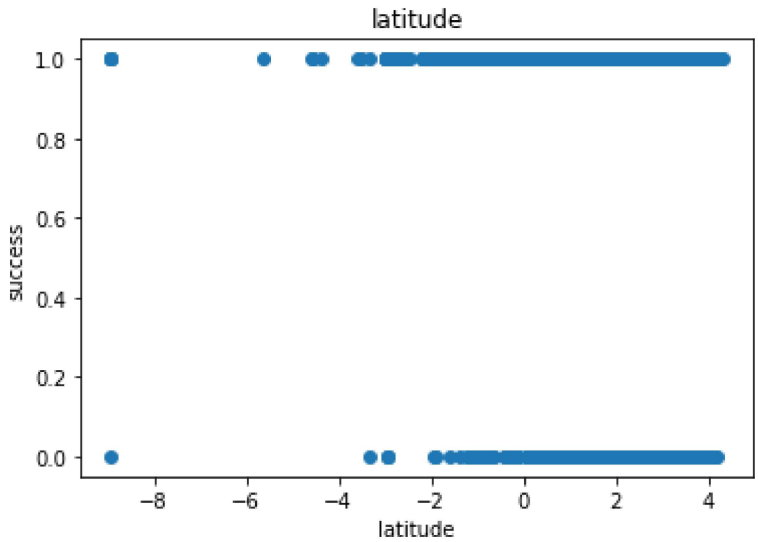
Log Trasnsformation

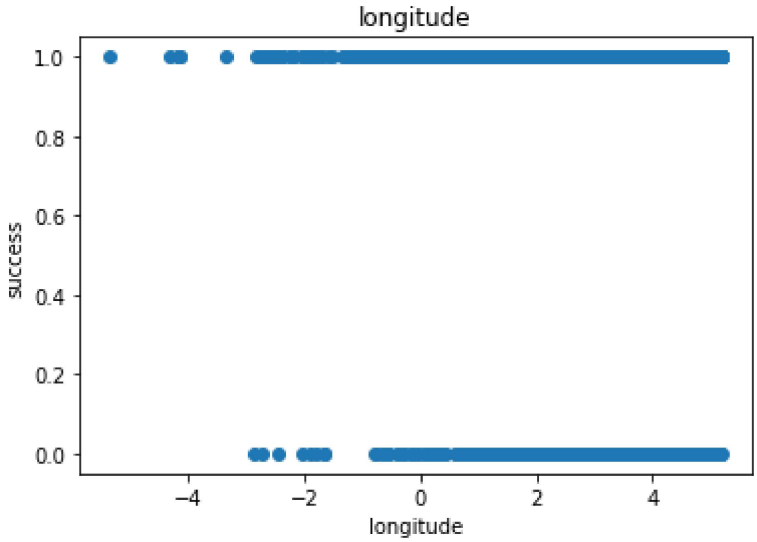
In [16]:
```
for feature in continuous_feature:
    data=dataset.copy()
    if 0 in data[feature].unique():
        pass
    else:
        data[feature]=np.log(data[feature])
        plt.scatter(data[feature],data['success'])
        plt.xlabel(feature)
        plt.ylabel('success')
        plt.title(feature)
        plt.show()
```

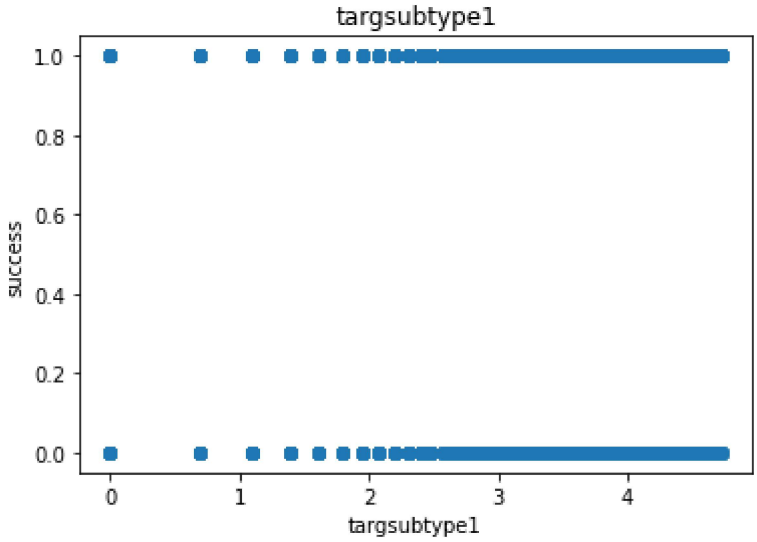


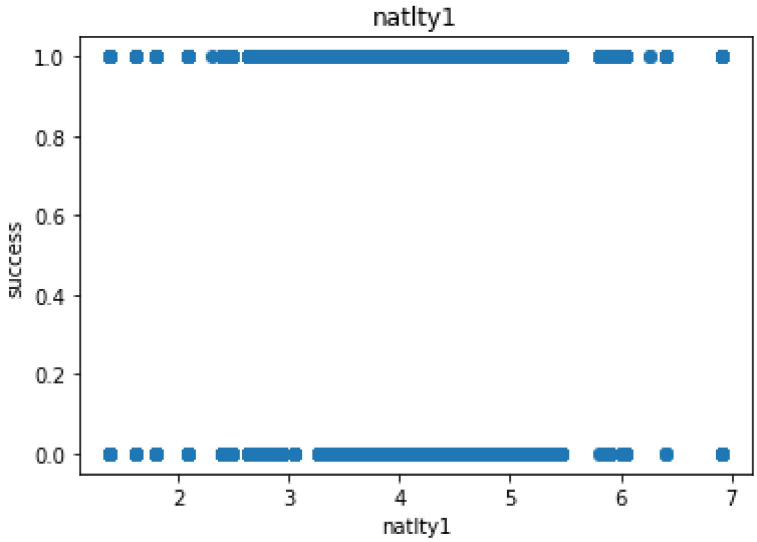


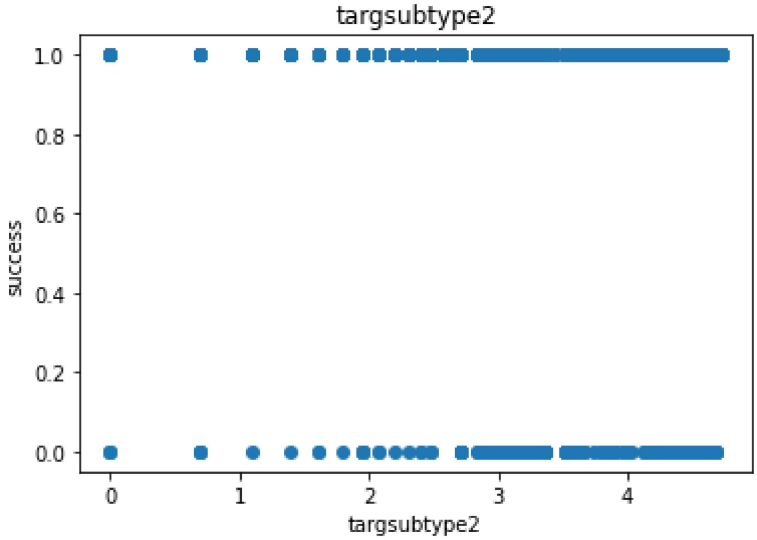

```
C:\ProgramData\Anaconda3\lib\site-packages\pandas\core\series.py:726: RuntimeWarnin
g: invalid value encountered in log
  result = getattr(ufunc, method)(*inputs, **kwargs)
```

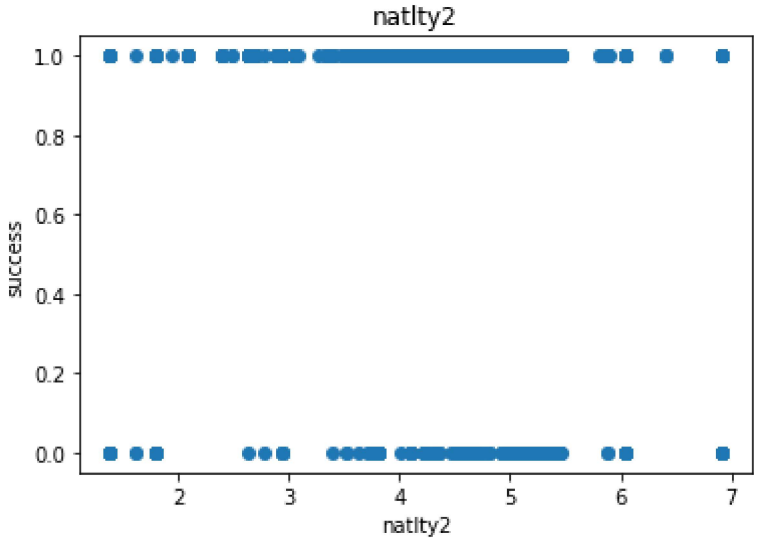


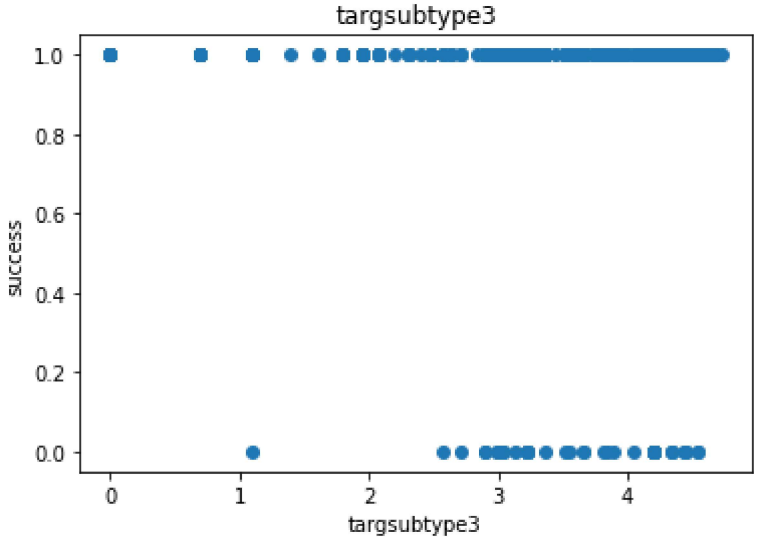

longitude
success
longitude
1.0
0.8
0.6
0.4
0.2
0.0
−4
−2
0
2
4

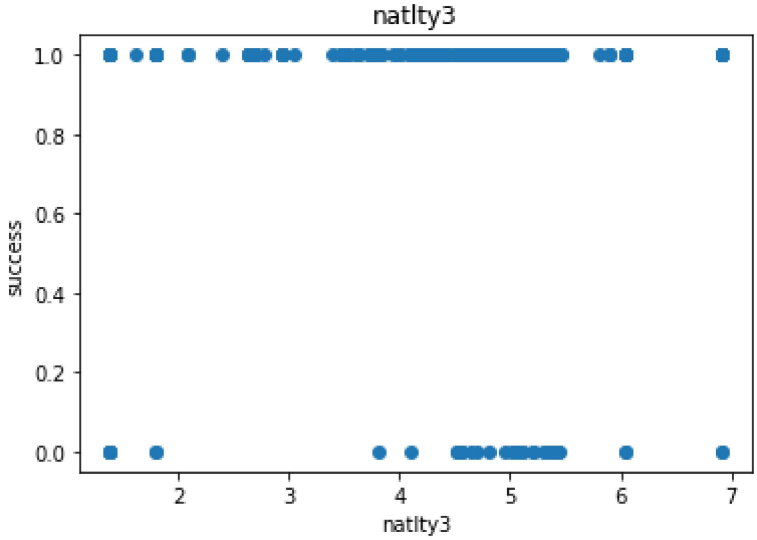

targsubtype1
success
targsubtype1
1.0
0.8
0.6
0.4
0.2
0.0
0
1
2
3
4

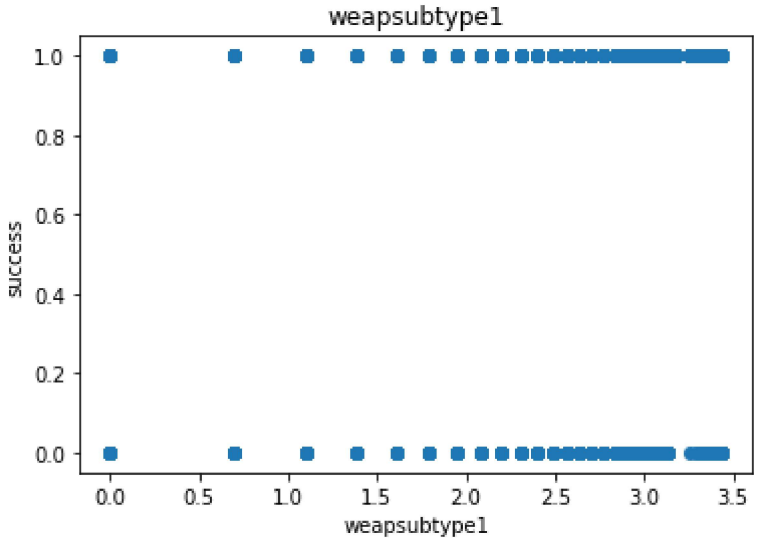

natlty1
success
natlty1
1.0
0.8
0.6
0.4
0.2
0.0
2
3
4
5
6
7

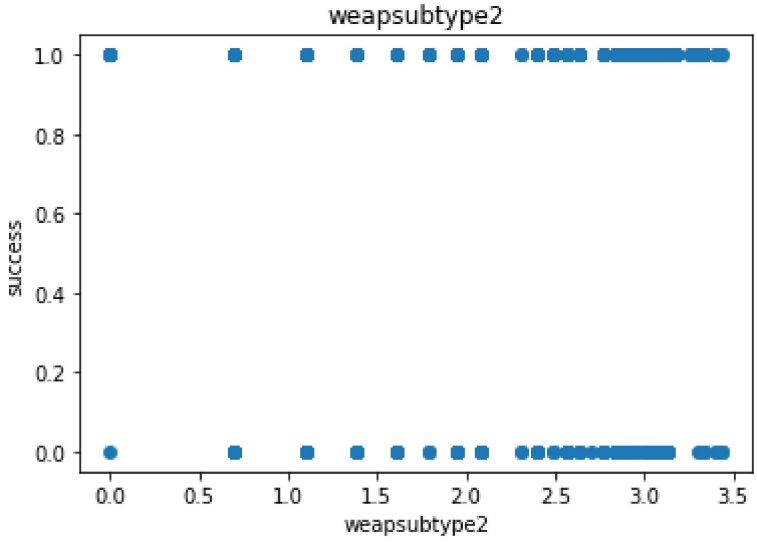

targsubtype2
success
1.0
0.8
0.6
0.4
0.2
0.0
0
1
2
3
4
targsubtype2


natlty2
success
1.0
0.8
0.6
0.4
0.2
0.0
2
3
4
5
6
7
natlty2


targsubtype3
success
1.0
0.8
0.6
0.4
0.2
0.0
0
1
2
3
4
targsubtype3



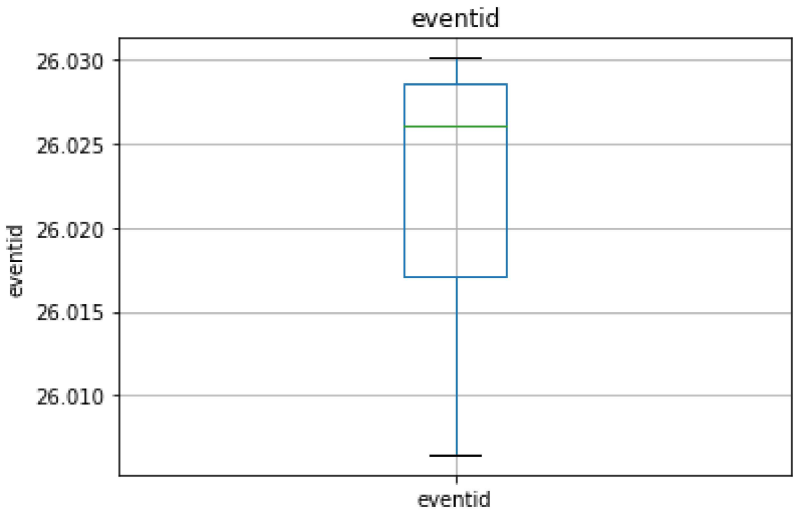


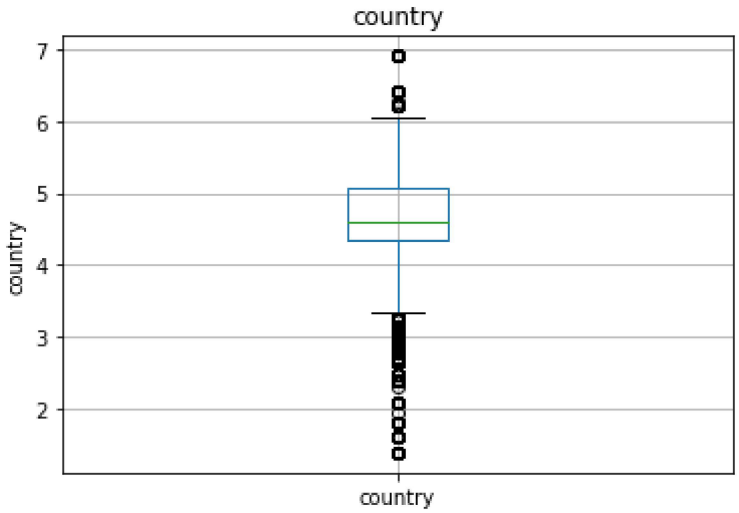


Outliers

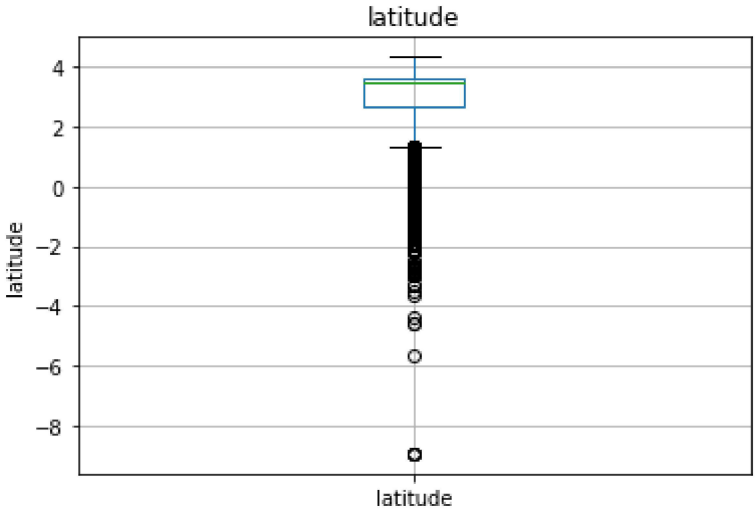
In [65]:
```
for feature in continuous_feature:
    data=dataset.copy()
    if 0 in data[feature].unique():
        pass
    else:
        data[feature]=np.log(data[feature])
        data.boxplot(column=feature)
        plt.ylabel(feature)
```

```
        plt.title(feature)
        plt.show()
```

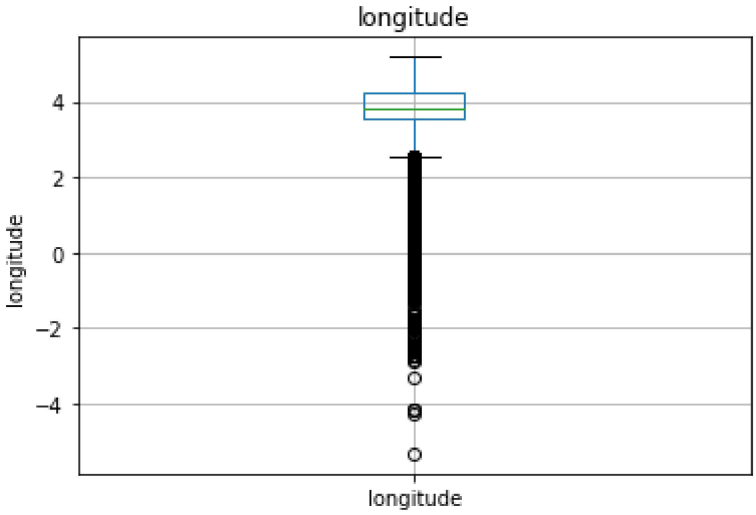


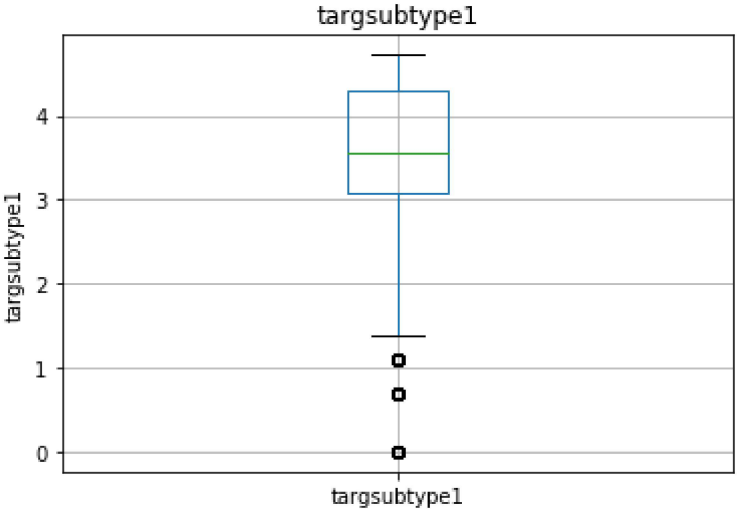


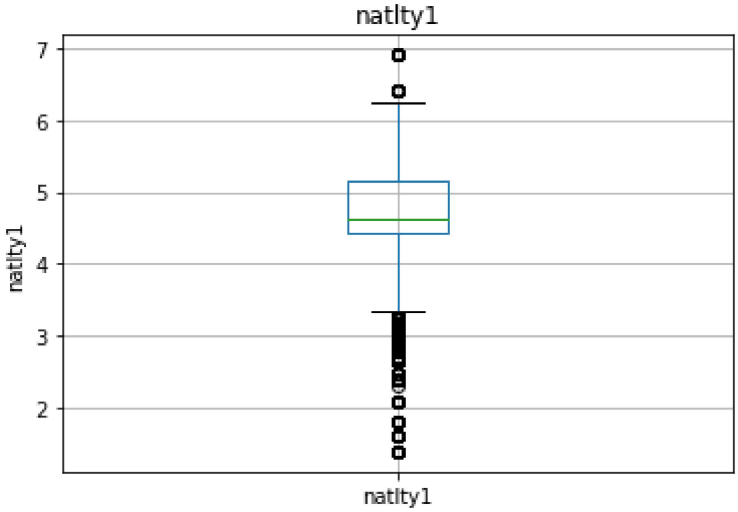

```
C:\Users\Omkar\Anaconda3\lib\site-packages\ipykernel_launcher.py:6: RuntimeWarning:
invalid value encountered in log
```

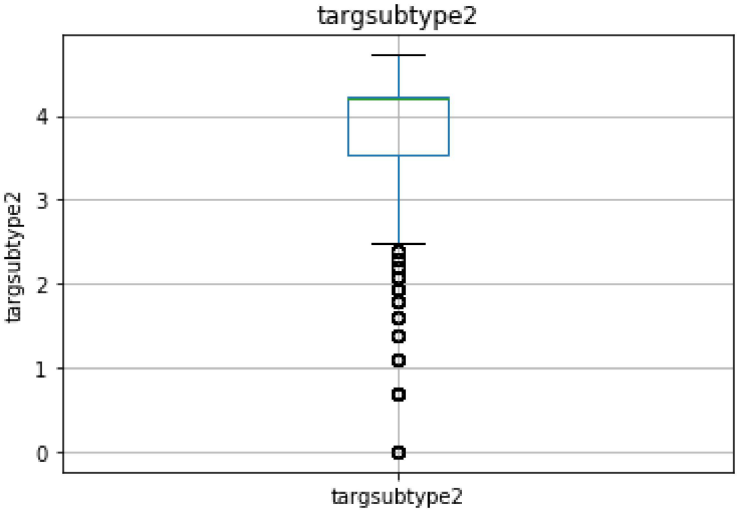


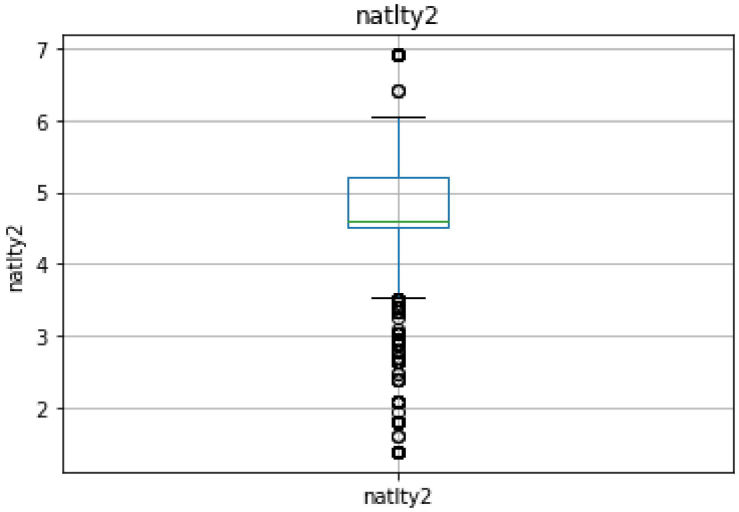

```
C:\Users\Omkar\Anaconda3\lib\site-packages\ipykernel_launcher.py:6: RuntimeWarning:
invalid value encountered in log
```

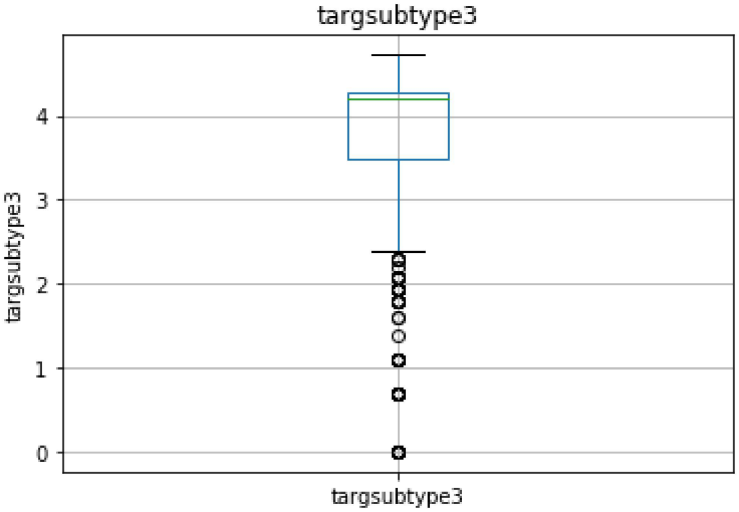

longitude
longitude
4
2
0
−2
−4
longitude

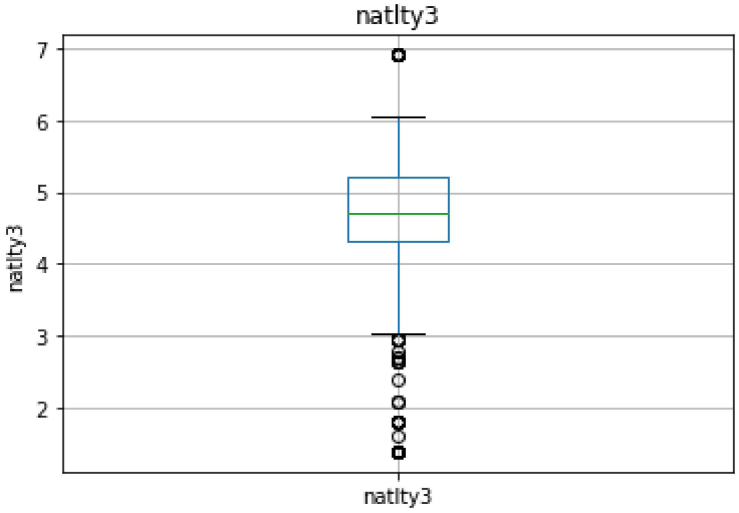

targsubtype1
targsubtype1
4
3
2
1
0
targsubtype1

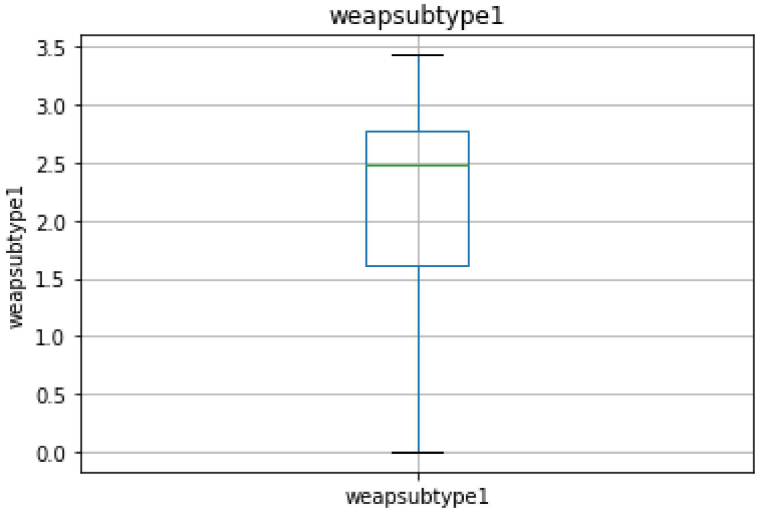

natlty1
natlty1
7
6
5
4
3
2
natlty1

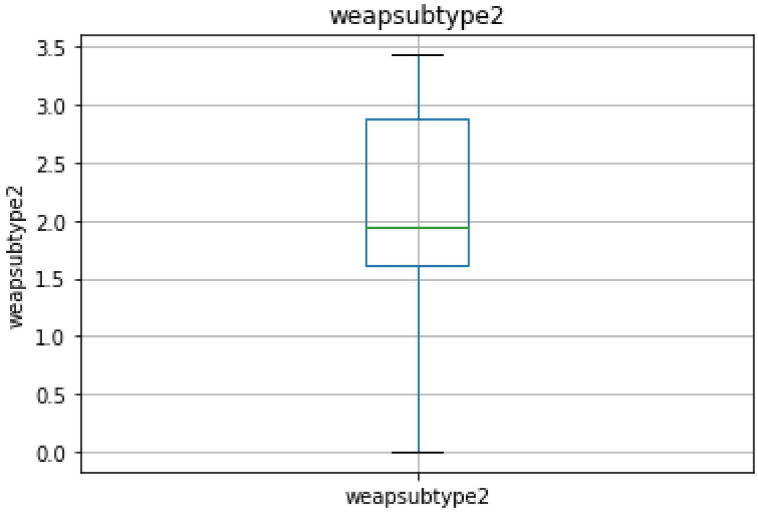

targsubtype2
targsubtype2
0
1
2
3
4
targsubtype2


natlty2
natlty2
2
3
4
5
6
7
natlty2


targsubtype3
targsubtype3
0
1
2
3
4
targsubtype3







Categorical Variables

In [17]:
```
categorical_features=[feature for feature in dataset.columns if data[feature].dtypes
categorical_features
```

Out[17]:
```
['region_txt',
 'alternative_txt',
 'attacktype1_txt',
 'attacktype2_txt',
 'attacktype3_txt',
 'gsubname3',
 'claimmode_txt',
 'claimmode2_txt',
 'claimmode3_txt',
```

```
 'weaptype1_txt',
 'weaptype2_txt',
 'weaptype3_txt',
 'weaptype4_txt',
 'weapsubtype4_txt',
 'propextent_txt',
 'hostkidoutcome_txt']
```

In [18]:
```
dataset[categorical_features].head()
```

Out[18]:

	region_txt	alternative_txt	attacktype1_txt	attacktype2_txt	attacktype3_txt	gsubname3	clai
0	Central America & Caribbean	NaN	Assassination	NaN	NaN	NaN	
1	North America	NaN	Hostage Taking (Kidnapping)	NaN	NaN	NaN	
2	Southeast Asia	NaN	Assassination	NaN	NaN	NaN	
3	Western Europe	NaN	Bombing/Explosion	NaN	NaN	NaN	
4	East Asia	NaN	Facility/Infrastructure Attack	NaN	NaN	NaN	

In [19]:
```
for feature in categorical_features:
    print('The feature is {} and number of categories are {}'.format(feature,len(dat
```

```
The feature is region_txt and number of categories are 12
The feature is alternative_txt and number of categories are 6
The feature is attacktype1_txt and number of categories are 9
The feature is attacktype2_txt and number of categories are 10
The feature is attacktype3_txt and number of categories are 9
The feature is gsubname3 and number of categories are 15
The feature is claimmode_txt and number of categories are 11
The feature is claimmode2_txt and number of categories are 10
The feature is claimmode3_txt and number of categories are 9
The feature is weaptype1_txt and number of categories are 12
The feature is weaptype2_txt and number of categories are 12
The feature is weaptype3_txt and number of categories are 11
The feature is weaptype4_txt and number of categories are 6
The feature is weapsubtype4_txt and number of categories are 17
The feature is propextent_txt and number of categories are 5
The feature is hostkidoutcome_txt and number of categories are 8
```

In [ ]:
```
for feature in categorical_features:
    data=dataset.copy()
    data.groupby(feature)['success'].plot.bar()
    plt.xlabel(feature)
    plt.ylabel('success')
    plt.title(feature)
    plt.show()
```

In [ ]: