**AV : Healthcare Analytics II**

## **About Dataset**

|  |  |
| --- | --- |
| Column | Description |
| case\_id | Case\_ID registered in Hospital |
| Hospital\_code | Unique code for the Hospital |
| Hospital\_type\_code | Unique code for the type of Hospital |
| City\_Code\_Hospital | City Code of the Hospital |
| Hospital\_region\_code | Region Code of the Hospital |
| Available Extra Rooms in Hospital | Number of Extra rooms available in the Hospital |
| Department | Department overlooking the case |
| Ward\_Type | Code for the Ward type |
| Ward\_Facility\_Code | Code for the Ward Facility |
| Bed Grade | Condition of Bed in the Ward |
| patientid | Unique Patient Id |
| City\_Code\_Patient | City Code for the patient |
| Type of Admission | Admission Type registered by the Hospital |
| Severity of Illness | Severity of the illness recorded at the time of admission |
| Visitors with Patient | Number of Visitors with the patient |
| Age | Age of the patient |
| Admission\_Deposit | Deposit at the Admission Time |
| Stay | Stay Days by the patient |

data info

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 case\_id 318438 non-null int64

1 Hospital\_code 318438 non-null int64

2 Hospital\_type\_code 318438 non-null object

3 City\_Code\_Hospital 318438 non-null int64

4 Hospital\_region\_code 318438 non-null object

5 Available Extra Rooms in Hospital 318438 non-null int64

6 Department 318438 non-null object

7 Ward\_Type 318438 non-null object

8 Ward\_Facility\_Code 318438 non-null object

9 Bed Grade 318325 non-null float64

10 patientid 318438 non-null int64

11 City\_Code\_Patient 313906 non-null float64

12 Type of Admission 318438 non-null object

13 Severity of Illness 318438 non-null object

14 Visitors with Patient 318438 non-null int64

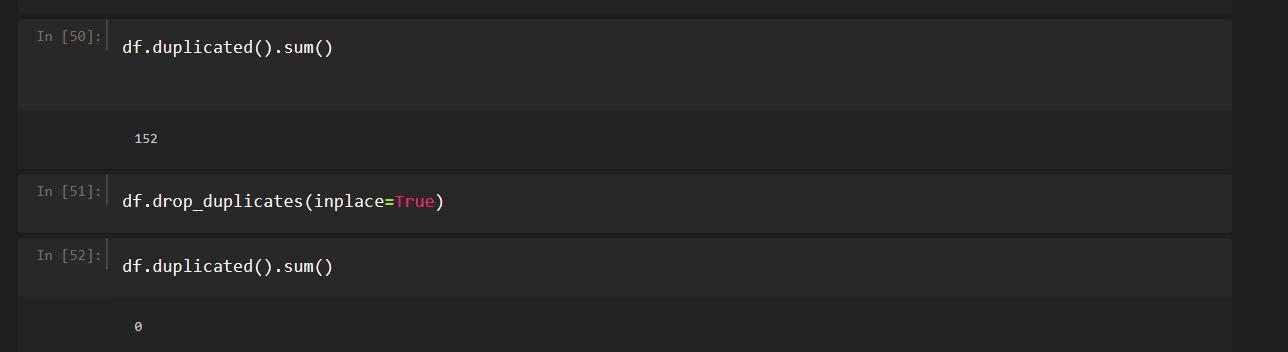
15 Age 318438 non-null object

16 Admission\_Deposit 318438 non-null float64

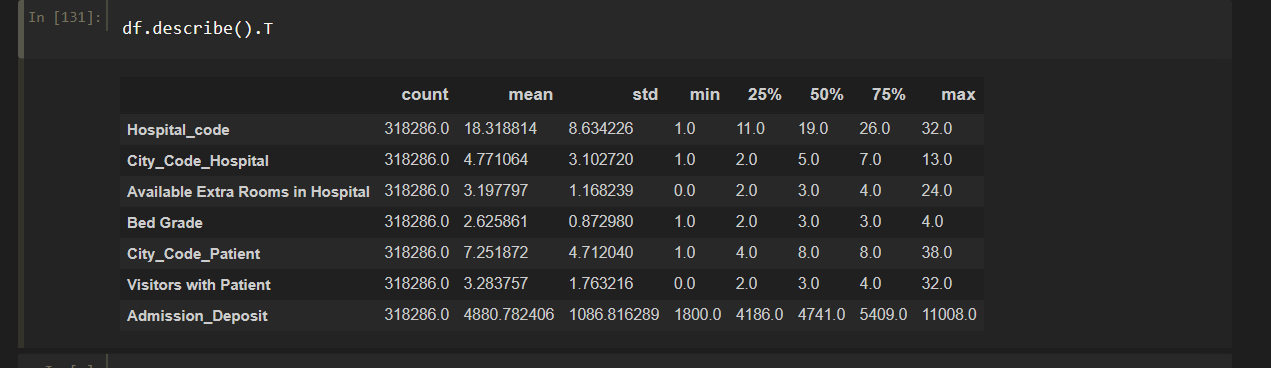
17 Stay 318438 non-null object

**We see that we have missing values in ['Bed Grade', 'City\_Code\_Patient'] so we filled them with mean value of each column**

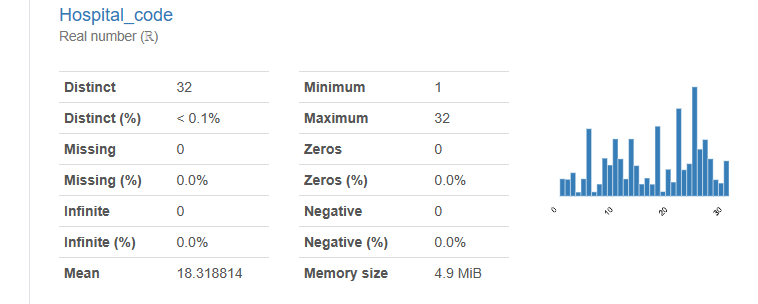
**I also found that data has 152 duplicated rows so I deleted them**

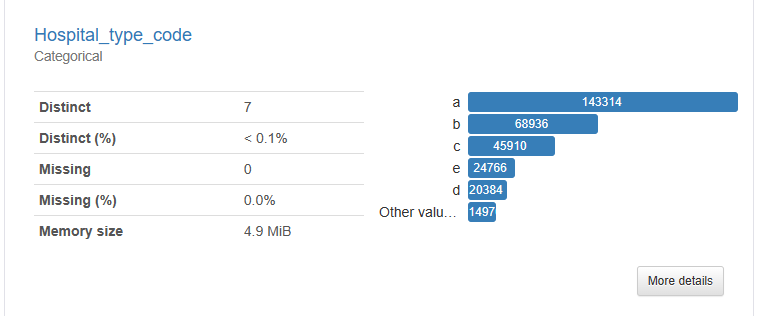


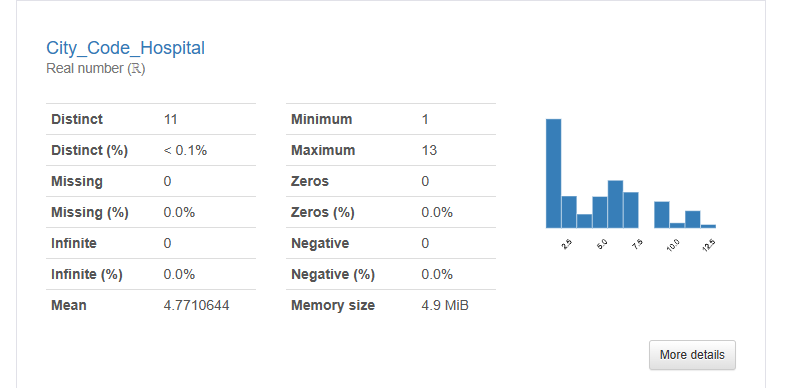
**Here is the descriptive analysis for the data**

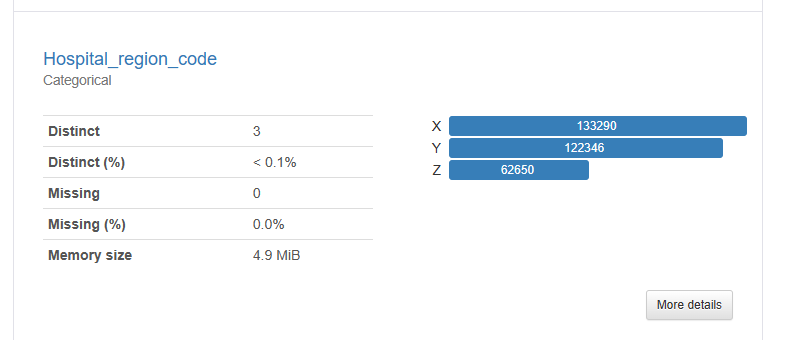


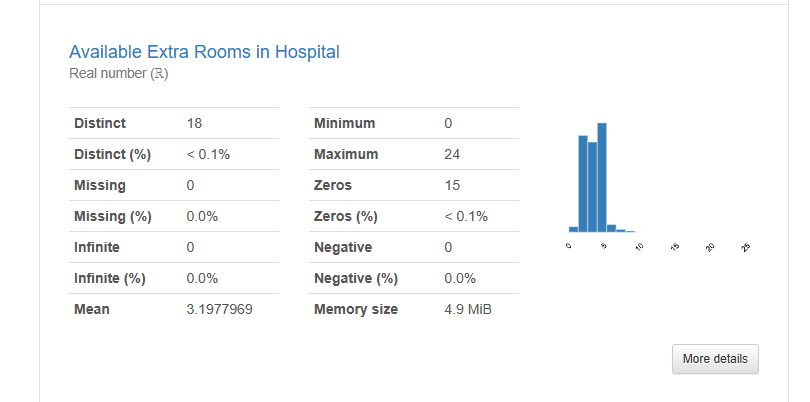
Here is detailed report for every column

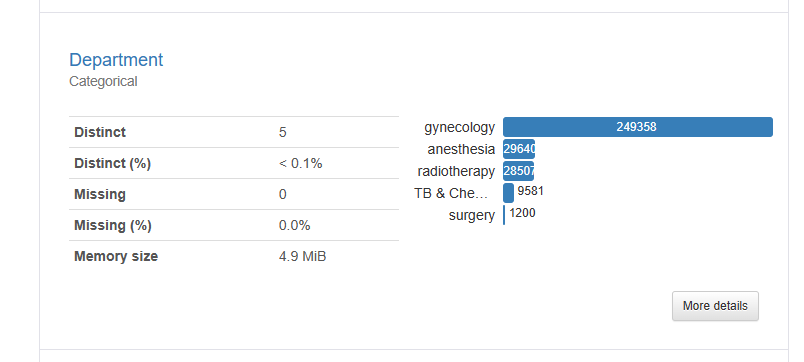


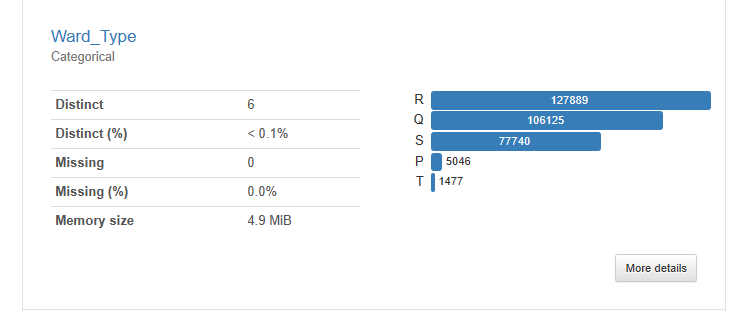


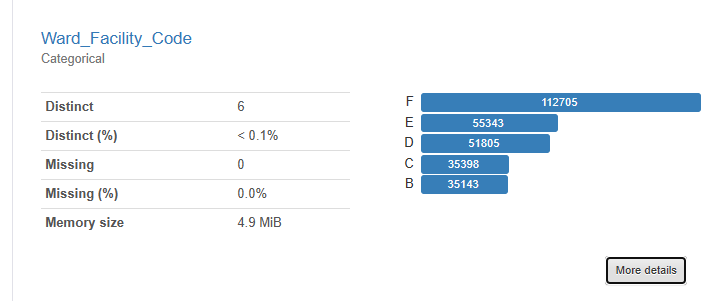


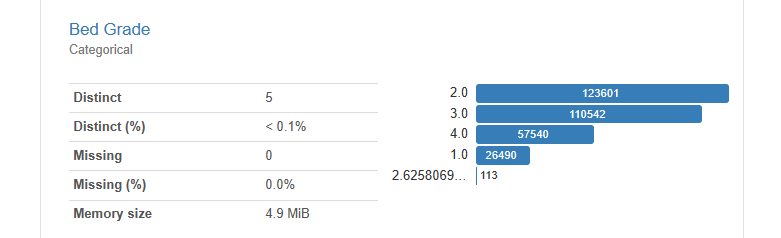


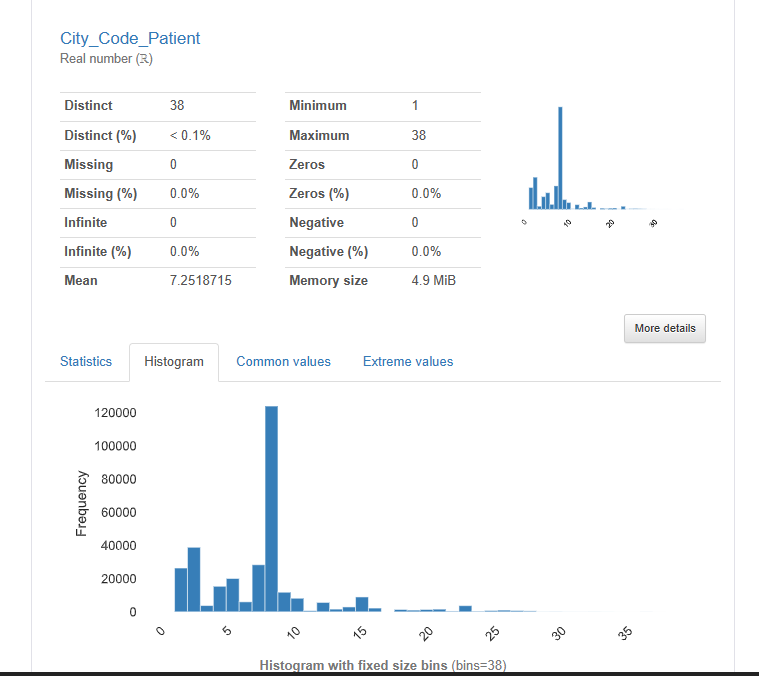


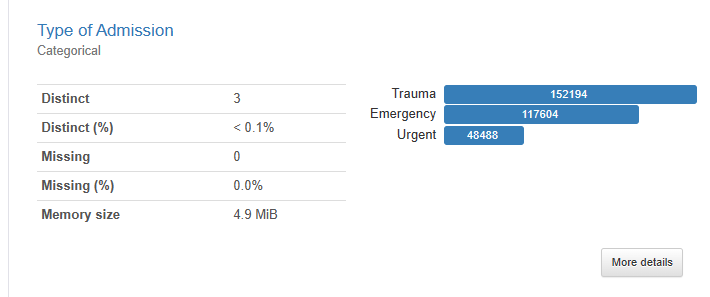


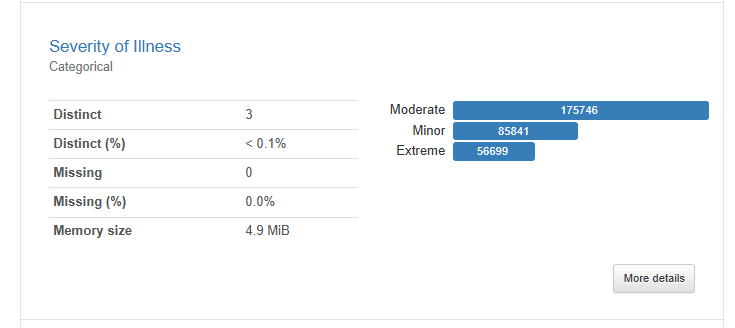


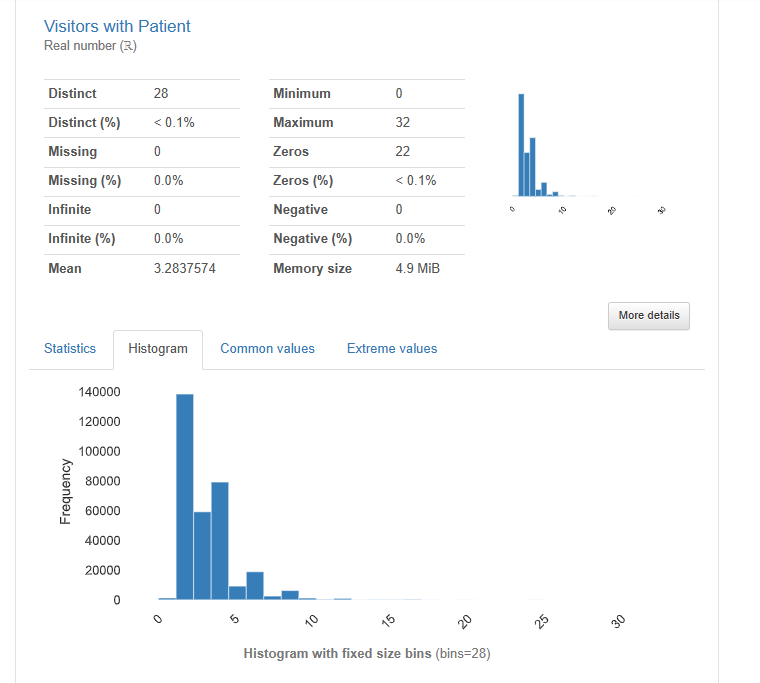


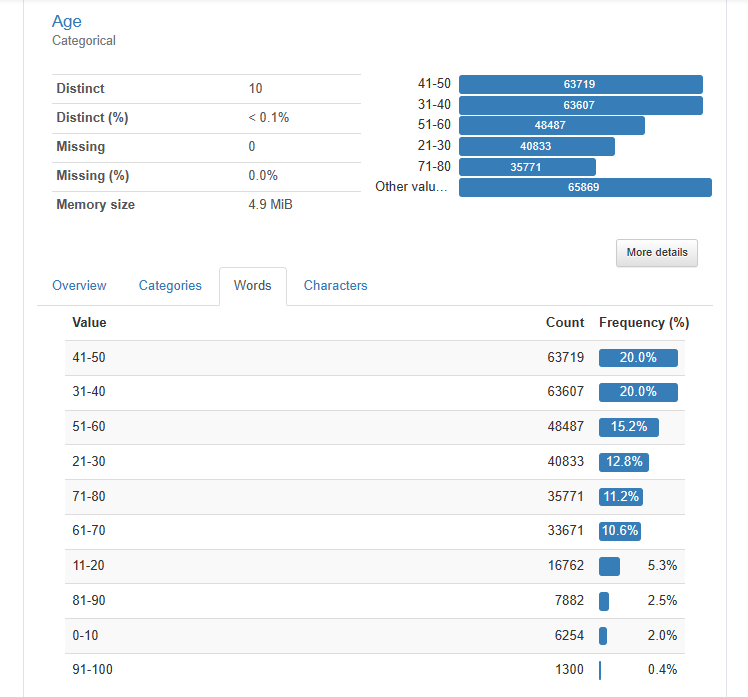


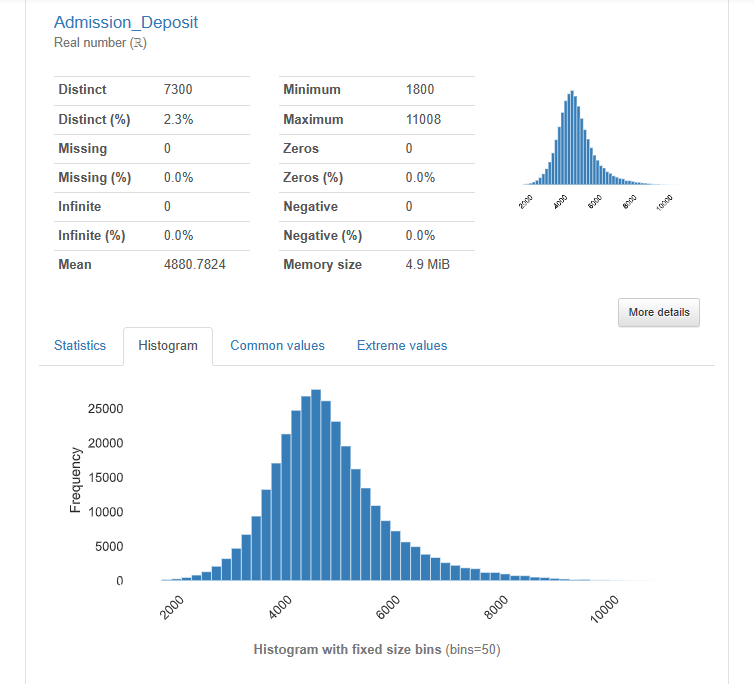


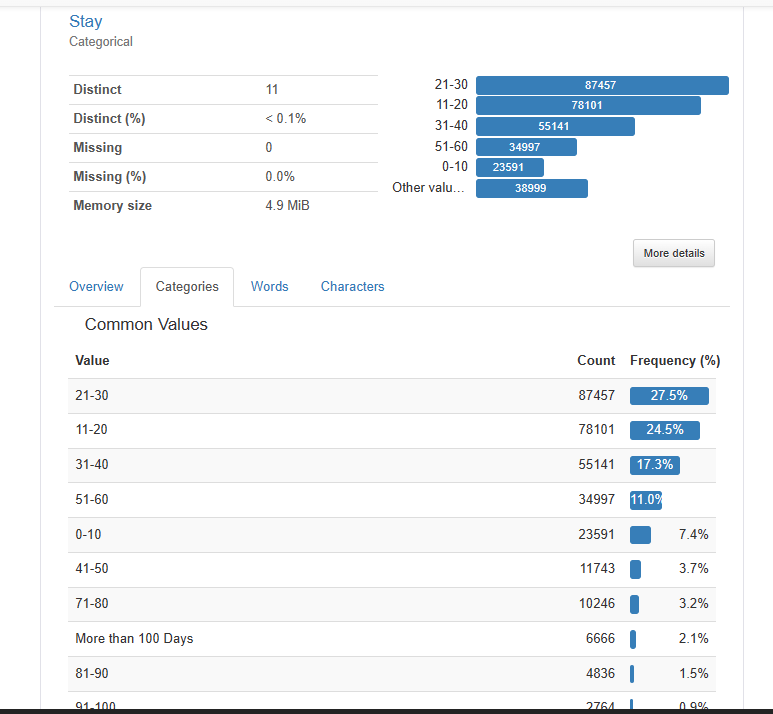




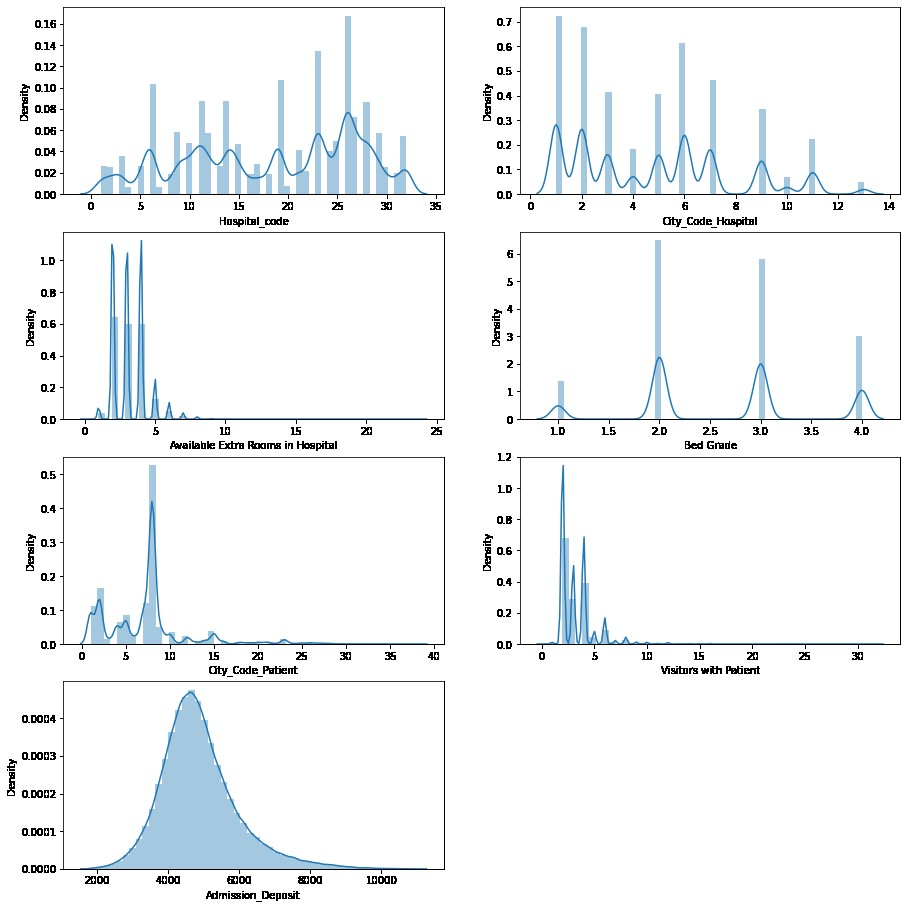




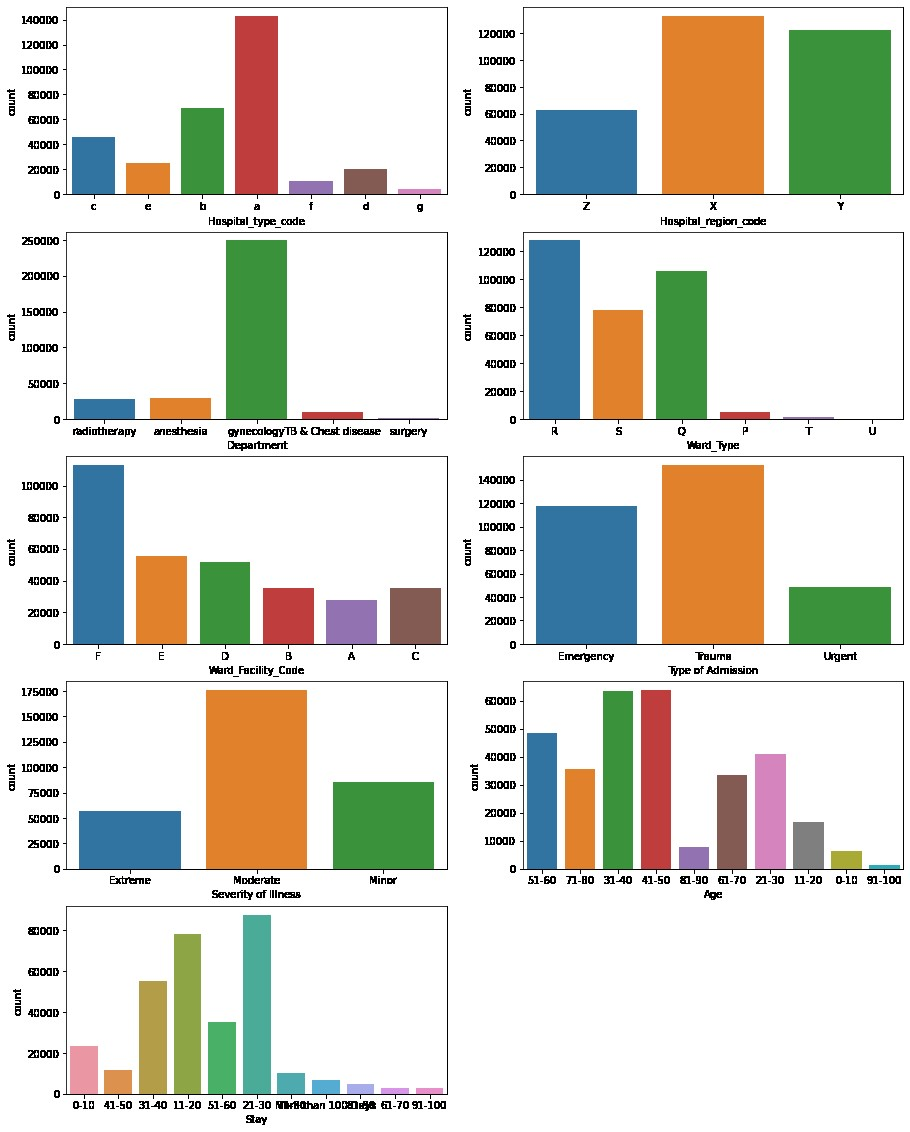


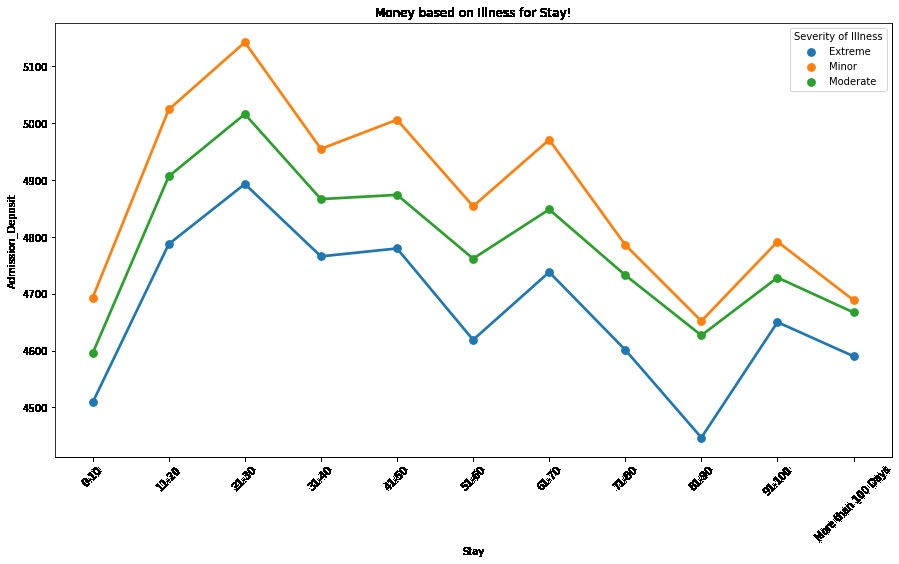


**Here is the distribution plot for the numeric columns:**

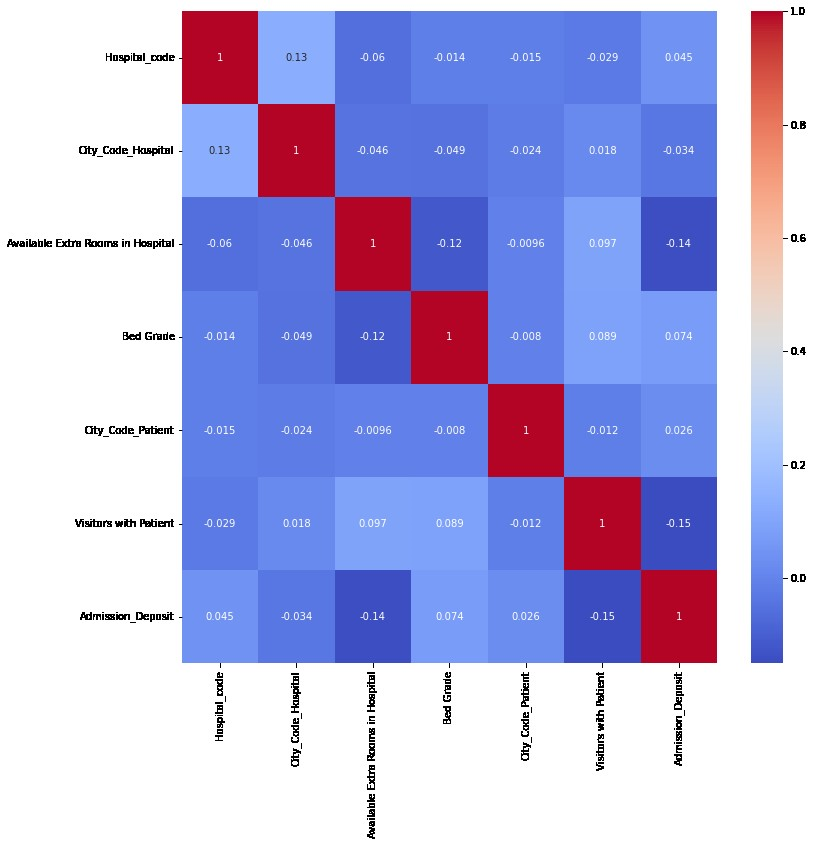


**Count plot for categorical columns**





***Here is the correlation matrix***



the distribution of categories within each categorical column and their relationship with the 'Stay' column.

Hospital\_type\_code

a 143425

b 68946

c 45928

e 24770

d 20389

f 10703

g 4277

Name: Hospital\_type\_code, dtype: int64

7

Stay 0-10 11-20 21-30 31-40 41-50 \

Hospital\_type\_code

a 0.033159 0.125007 0.126511 0.075286 0.012910

b 0.016157 0.043754 0.061365 0.038893 0.012225

c 0.011267 0.032402 0.039336 0.025820 0.005879

d 0.004126 0.013890 0.017357 0.011638 0.002387

e 0.005668 0.020864 0.020488 0.012750 0.002399

f 0.003213 0.007323 0.005957 0.006573 0.000641

g 0.000534 0.002142 0.003737 0.002258 0.000437

Stay 51-60 61-70 71-80 81-90 91-100 \

Hospital\_type\_code

a 0.045902 0.002902 0.011688 0.005935 0.003021

b 0.021697 0.002880 0.008190 0.003222 0.002459

c 0.016254 0.001350 0.005059 0.002277 0.001379

d 0.008520 0.000584 0.002390 0.001322 0.000565

e 0.008893 0.000619 0.002588 0.001240 0.000685

f 0.006155 0.000138 0.001514 0.000798 0.000433

g 0.002547 0.000144 0.000773 0.000399 0.000141

Stay More than 100 Days

Hospital\_type\_code

a 0.008080

b 0.005671

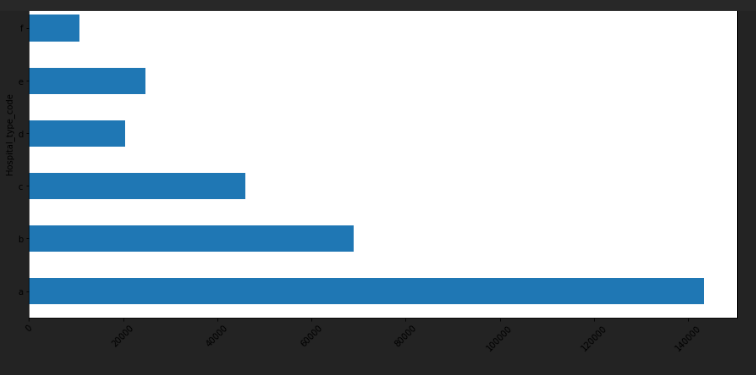
c 0.003206

d 0.001250

e 0.001592

f 0.000867

g 0.000320



-----------------

Hospital\_region\_code

X 133336

Y 122428

Z 62674

Name: Hospital\_region\_code, dtype: int64

3

Stay 0-10 11-20 21-30 31-40 41-50 \

Hospital\_region\_code

X 0.025578 0.106410 0.121744 0.071694 0.013833

Y 0.032807 0.093538 0.099219 0.066933 0.015815

Z 0.015739 0.045434 0.053788 0.034591 0.007229

Stay 51-60 61-70 71-80 81-90 91-100 \

Hospital\_region\_code

X 0.046235 0.003178 0.012847 0.005998 0.003602

Y 0.040897 0.003718 0.012778 0.006092 0.003354

Z 0.022836 0.001721 0.006576 0.003103 0.001727

Stay More than 100 Days

Hospital\_region\_code

X 0.007600

Y 0.009314

Z 0.004073



-----------------

Department

gynecology 249486

anesthesia 29649

radiotherapy 28516

TB & Chest disease 9586

surgery 1201

Name: Department, dtype: int64

5

Stay 0-10 11-20 21-30 31-40 41-50 \

Department

TB & Chest disease 0.002484 0.007471 0.008221 0.005505 0.000942

anesthesia 0.004770 0.027707 0.028106 0.015570 0.002672

gynecology 0.059192 0.187446 0.214745 0.136714 0.030370

radiotherapy 0.007471 0.021945 0.022808 0.014800 0.002760

surgery 0.000207 0.000813 0.000870 0.000628 0.000132

Stay 51-60 61-70 71-80 81-90 91-100 \

Department

TB & Chest disease 0.003288 0.000239 0.000923 0.000440 0.000201

anesthesia 0.007936 0.000543 0.002208 0.001206 0.000490

gynecology 0.087895 0.006990 0.025641 0.011770 0.007125

radiotherapy 0.010225 0.000829 0.003275 0.001658 0.000804

surgery 0.000625 0.000016 0.000154 0.000119 0.000063

Stay More than 100 Days

Department

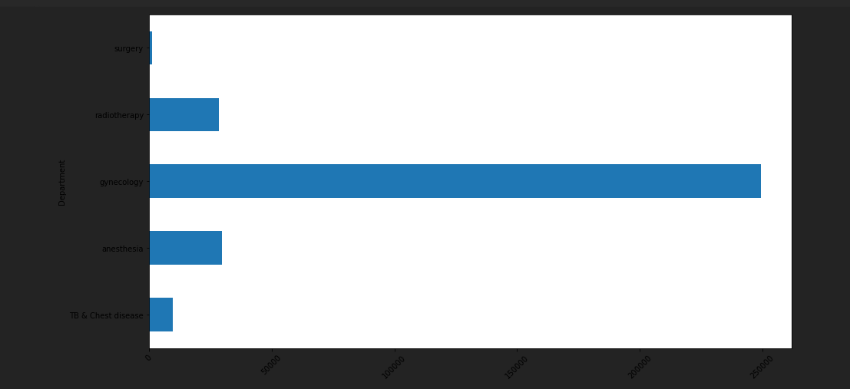
TB & Chest disease 0.000389

anesthesia 0.001900

gynecology 0.015579

radiotherapy 0.002974

surgery 0.000144



-----------------

Ward\_Type

R 127947

Q 106165

S 77794

P 5046

T 1477

U 9

Name: Ward\_Type, dtype: int64

6

Stay 0-10 11-20 21-30 31-40 41-50 51-60 \

Ward\_Type

P 0.001448 0.004374 0.005659 0.002340 0.000895 0.000584

Q 0.026112 0.091710 0.109082 0.053941 0.015532 0.020434

R 0.029236 0.099300 0.115244 0.069235 0.013406 0.043126

S 0.016980 0.049143 0.043531 0.046781 0.006827 0.045227

T 0.000345 0.000845 0.001225 0.000914 0.000217 0.000597

U 0.000003 0.000009 0.000009 0.000006 0.000000 0.000000

Stay 61-70 71-80 81-90 91-100 More than 100 Days

Ward\_Type

P 0.000229 0.000166 0.000044 0.000050 0.000057

Q 0.003096 0.006133 0.002487 0.001448 0.003417

R 0.003282 0.012580 0.005313 0.003511 0.007562

S 0.001969 0.013145 0.007226 0.003611 0.009857

T 0.000041 0.000176 0.000122 0.000063 0.000094

U 0.000000 0.000000 0.000000 0.000000 0.000000



-----------------

Ward\_Facility\_Code

F 112753

E 55351

D 51809

C 35463

B 35156

A 27906

Name: Ward\_Facility\_Code, dtype: int64

6

Stay 0-10 11-20 21-30 31-40 41-50 \

Ward\_Facility\_Code

A 0.007480 0.019266 0.023477 0.015711 0.003762

B 0.008187 0.024727 0.027098 0.020164 0.003423

C 0.012313 0.037897 0.024705 0.017558 0.001868

D 0.012307 0.030913 0.047416 0.029211 0.010523

E 0.010947 0.047359 0.051366 0.028938 0.005351

F 0.022890 0.085219 0.100688 0.061635 0.011949

Stay 51-60 61-70 71-80 81-90 91-100 \

Ward\_Facility\_Code

A 0.009908 0.000801 0.003037 0.001281 0.000914

B 0.015887 0.000785 0.004337 0.002412 0.001065

C 0.010128 0.000433 0.002264 0.001256 0.000490

D 0.014882 0.002500 0.006177 0.002424 0.001799

E 0.017611 0.001291 0.004745 0.002308 0.001212

F 0.041553 0.002807 0.011641 0.005511 0.003203

Stay More than 100 Days

Ward\_Facility\_Code

A 0.001997

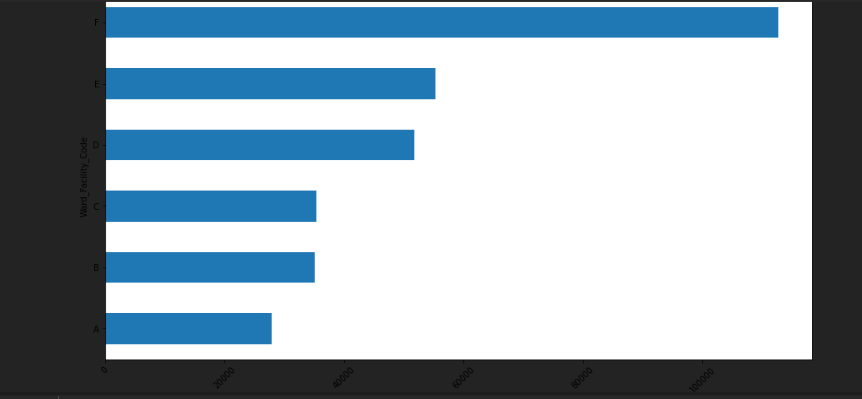
B 0.002318

C 0.002453

D 0.004544

E 0.002691

F 0.006984



-----------------

Type of Admission

Trauma 152261

Emergency 117676

Urgent 48501

Name: Type of Admission, dtype: int64

3

Stay 0-10 11-20 21-30 31-40 41-50 51-60 \

Type of Admission

Emergency 0.044649 0.099106 0.086042 0.059418 0.013017 0.037323

Trauma 0.016732 0.105970 0.145221 0.088265 0.018741 0.057499

Urgent 0.012743 0.040306 0.043487 0.025534 0.005119 0.015146

Stay 61-70 71-80 81-90 91-100 More than 100 Days

Type of Admission

Emergency 0.003065 0.010599 0.005364 0.002936 0.008024

Trauma 0.004645 0.017432 0.008002 0.004767 0.010875

Urgent 0.000908 0.004170 0.001828 0.000980 0.002088



-----------------

Severity of Illness

Moderate 175843

Minor 85872

Extreme 56723

Name: Severity of Illness, dtype: int64

3

Stay 0-10 11-20 21-30 31-40 41-50 \

Severity of Illness

Extreme 0.010674 0.033030 0.048681 0.031673 0.007383

Minor 0.024702 0.085043 0.067627 0.045368 0.009421

Moderate 0.038749 0.127309 0.158442 0.096176 0.020073

Stay 51-60 61-70 71-80 81-90 91-100 \

Severity of Illness

Extreme 0.024422 0.002032 0.008086 0.003495 0.002528

Minor 0.022384 0.001630 0.006055 0.003093 0.001335

Moderate 0.063161 0.004955 0.018060 0.008605 0.004820

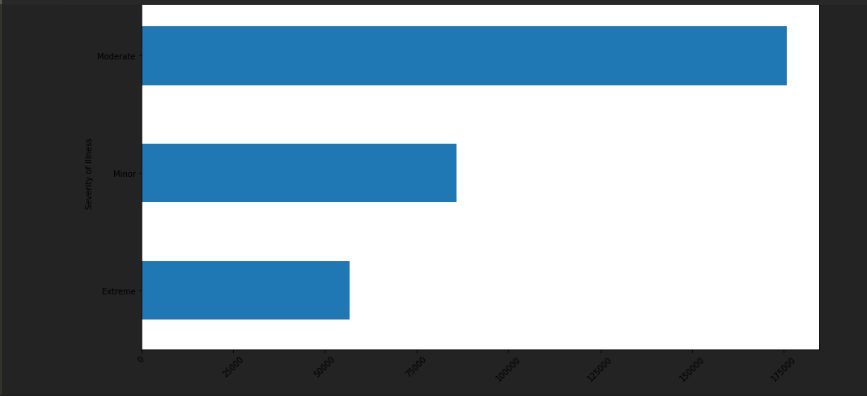
Stay More than 100 Days

Severity of Illness

Extreme 0.006124

Minor 0.003008

Moderate 0.011855



-----------------

Age

41-50 63749

31-40 63639

51-60 48514

21-30 40843

71-80 35792

61-70 33687

11-20 16768

81-90 7890

0-10 6254

91-100 1302

Name: Age, dtype: int64

10

Stay 0-10 11-20 21-30 31-40 41-50 51-60 61-70 \

Age

0-10 0.001931 0.006152 0.004676 0.003184 0.000587 0.001828 0.000082

11-20 0.004874 0.016779 0.013541 0.008419 0.001602 0.004488 0.000279

21-30 0.010888 0.035398 0.035781 0.021706 0.004390 0.011911 0.000826

31-40 0.015438 0.049592 0.058253 0.034267 0.007452 0.020466 0.001598

41-50 0.014844 0.046976 0.056231 0.034490 0.007873 0.022576 0.001765

51-60 0.010762 0.035630 0.041006 0.026909 0.005448 0.018022 0.001407

61-70 0.006890 0.024714 0.028367 0.018622 0.003784 0.012816 0.001021

71-80 0.006912 0.024991 0.029940 0.020161 0.004343 0.013921 0.001187

81-90 0.001325 0.004371 0.006029 0.004723 0.001190 0.003398 0.000361

91-100 0.000261 0.000779 0.000926 0.000735 0.000207 0.000543 0.000091

Stay 71-80 81-90 91-100 More than 100 Days

Age

0-10 0.000480 0.000264 0.000110 0.000345

11-20 0.001099 0.000700 0.000223 0.000653

21-30 0.003222 0.001715 0.000725 0.001699

31-40 0.005675 0.002515 0.001520 0.003071

41-50 0.006739 0.002779 0.001815 0.004104

51-60 0.005370 0.002462 0.001567 0.003765

61-70 0.003863 0.001884 0.001036 0.002792

71-80 0.004293 0.002104 0.001212 0.003335

81-90 0.001262 0.000678 0.000415 0.001024

91-100 0.000198 0.000091 0.000060 0.000198



-----------------

Stay

21-30 87491

11-20 78139

31-40 55159

51-60 35018

0-10 23604

41-50 11743

71-80 10254

More than 100 Days 6683

81-90 4838

91-100 2765

61-70 2744

Name: Stay, dtype: int64

11

Stay 0-10 11-20 21-30 31-40 41-50 \

Stay

0-10 0.074124 0.000000 0.000000 0.000000 0.000000

11-20 0.000000 0.245382 0.000000 0.000000 0.000000

21-30 0.000000 0.000000 0.274751 0.000000 0.000000

31-40 0.000000 0.000000 0.000000 0.173217 0.000000

41-50 0.000000 0.000000 0.000000 0.000000 0.036877

51-60 0.000000 0.000000 0.000000 0.000000 0.000000

61-70 0.000000 0.000000 0.000000 0.000000 0.000000

71-80 0.000000 0.000000 0.000000 0.000000 0.000000

81-90 0.000000 0.000000 0.000000 0.000000 0.000000

91-100 0.000000 0.000000 0.000000 0.000000 0.000000

More than 100 Days 0.000000 0.000000 0.000000 0.000000 0.000000

Stay 51-60 61-70 71-80 81-90 91-100 \

Stay

0-10 0.000000 0.000000 0.000000 0.000000 0.000000

11-20 0.000000 0.000000 0.000000 0.000000 0.000000

21-30 0.000000 0.000000 0.000000 0.000000 0.000000

31-40 0.000000 0.000000 0.000000 0.000000 0.000000

41-50 0.000000 0.000000 0.000000 0.000000 0.000000

51-60 0.109968 0.000000 0.000000 0.000000 0.000000

61-70 0.000000 0.008617 0.000000 0.000000 0.000000

71-80 0.000000 0.000000 0.032201 0.000000 0.000000

81-90 0.000000 0.000000 0.000000 0.015193 0.000000

91-100 0.000000 0.000000 0.000000 0.000000 0.008683

More than 100 Days 0.000000 0.000000 0.000000 0.000000 0.000000

Stay More than 100 Days

Stay

0-10 0.000000

11-20 0.000000

21-30 0.000000

31-40 0.000000

41-50 0.000000

51-60 0.000000

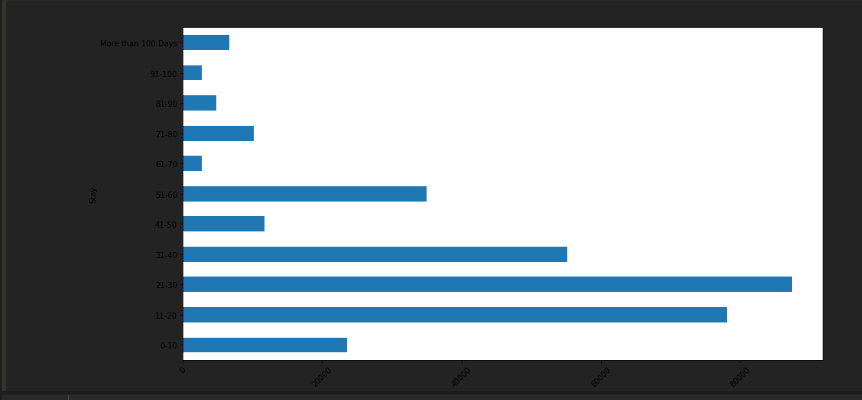
61-70 0.000000

71-80 0.000000

81-90 0.000000

91-100 0.000000

More than 100 Days 0.020987



-----------------

**The distribution of categories within each column of the DataFrame, with bars colored according to the severity of illness.**

