CMS - Healthcare Analysis

CMS

CMS is the federal agency that provides health coverage to more than 160 million through Medicare, Medicaid, the Children's Health Insurance Program, and the Health Insurance Marketplace. CMS works in partnership with the entire health care community to improve quality, equity and outcomes in the health care system.

In [43]:

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')
```

In [44]:

```
pd.set_option('display.max_columns', None)
pd.set_option('display.max_rows', None)
```

HCAHPS File handling

HCAHPS

HCAHPS stands for the **Hospital Consumer Assessment of Healthcare Providers and Systems**. It is a national survey that measures patients' perceptions of their care during a recent inpatient hospital stay, and the results are publicly reported to help people compare hospitals. The survey is used by the Centers for Medicare and Medicaid Services (CMS) to assess patient satisfaction and also affects hospital reimbursement.

In [45]:

```
hcahps= pd.read_csv("/content/HCAHPS-Hospital.csv")
```

In [46]:

```
hcahps.shape
```

Out[46]:

```
(445563, 22)
```

In [47]:

```
hcahps.head()
```

Out[47]:

	Facility ID	Facility Name	Address	City/Town	State	ZIP Code	County/Parish	Telephone Number
0	010001	SOUTHEAST HEALTH MEDICAL CENTER	1108 ROSS CLARK CIRCLE	DOTHAN	AL	36301	HOUSTON	(334) 793-8701
1	010001	SOUTHEAST HEALTH MEDICAL CENTER	1108 ROSS CLARK CIRCLE	DOTHAN	AL	36301	HOUSTON	(334) 793-8701
2	010001	SOUTHEAST HEALTH MEDICAL CENTER	1108 ROSS CLARK CIRCLE	DOTHAN	AL	36301	HOUSTON	(334) 793-8701
3	010001	SOUTHEAST HEALTH MEDICAL CENTER	1108 ROSS CLARK CIRCLE	DOTHAN	AL	36301	HOUSTON	(334) 793-8701
4	010001	SOUTHEAST HEALTH MEDICAL CENTER	1108 ROSS CLARK CIRCLE	DOTHAN	AL	36301	HOUSTON	(334) 793-8701

In [48]:

```
hcahps.describe()
```

Out[48]:

	ZIP Code
count	445563.000000
mean	54114.258192
std	26956.222954
min	603.000000
25%	33021.000000
50%	55705.000000
75%	76132.000000
max	99929.000000

In [49]:

```
hcahps.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 445563 entries, 0 to 445562
Data columns (total 22 columns):
 #   Column                                 Non-Null Count   Dtype 
---  ------                                 --------------   ----- 
 0   Facility ID                            445563 non-null  object
 1   Facility Name                          445563 non-null  object
 2   Address                                445563 non-null  object
 3   City/Town                              445563 non-null  object
 4   State                                  445563 non-null  object
 5   ZIP Code                               445563 non-null  int64 
 6   County/Parish                          445563 non-null  object
 7   Telephone Number                       445563 non-null  object
 8   HCAHPS Measure ID                      445563 non-null  object
 9   HCAHPS Question                        445563 non-null  object
 10  HCAHPS Answer Description              445563 non-null  object
 11  Patient Survey Star Rating             445563 non-null  object
 12  Patient Survey Star Rating Footnote    17875 non-null   object
 13  HCAHPS Answer Percent                  445563 non-null  object
 14  HCAHPS Answer Percent Footnote         123264 non-null  object
 15  HCAHPS Linear Mean Value               445563 non-null  object
 16  Number of Completed Surveys            445563 non-null  object
 17  Number of Completed Surveys Footnote   159216 non-null  object
 18  Survey Response Rate Percent           445563 non-null  object
 19  Survey Response Rate Percent Footnote  159216 non-null  object
 20  Start Date                             445563 non-null  object
 21  End Date                               445563 non-null  object
dtypes: int64(1), object(21)
memory usage: 74.8+ MB
```

In [50]:

```
cols= ["Patient Survey Star Rating",
    "HCAHPS Answer Percent",
    "HCAHPS Linear Mean Value",
    "Number of Completed Surveys",
    "Survey Response Rate Percent"]
hcahps[cols] = hcahps[cols].apply(pd.to_numeric, errors='coerce')
```

Observation

Due to the presence of "Not Available" at Numerical columns earlier, it is now converted as null.

In [51]:

```
hcahps.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 445563 entries, 0 to 445562
Data columns (total 22 columns):
 #   Column                                 Non-Null Count   Dtype
---  ------                                 --------------   -----
 0   Facility ID                            445563 non-null  object
 1   Facility Name                          445563 non-null  object
 2   Address                                445563 non-null  object
 3   City/Town                              445563 non-null  object
 4   State                                  445563 non-null  object
 5   ZIP Code                               445563 non-null  int64
 6   County/Parish                          445563 non-null  object
 7   Telephone Number                       445563 non-null  object
 8   HCAHPS Measure ID                      445563 non-null  object
 9   HCAHPS Question                        445563 non-null  object
 10  HCAHPS Answer Description              445563 non-null  object
 11  Patient Survey Star Rating             34826 non-null   float64
 12  Patient Survey Star Rating Footnote    17875 non-null   object
 13  HCAHPS Answer Percent                  270831 non-null  float64
 14  HCAHPS Answer Percent Footnote         123264 non-null  object
 15  HCAHPS Linear Mean Value               31660 non-null   float64
 16  Number of Completed Surveys            368652 non-null  float64
 17  Number of Completed Surveys Footnote   159216 non-null  object
 18  Survey Response Rate Percent           368652 non-null  float64
 19  Survey Response Rate Percent Footnote  159216 non-null  object
 20  Start Date                             445563 non-null  object
 21  End Date                               445563 non-null  object
dtypes: float64(5), int64(1), object(16)
memory usage: 74.8+ MB
```

In [52]:

```
hcahps.isnull().sum()
```

Out[52]:

	0
Facility ID	0
Facility Name	0
Address	0
City/Town	0
State	0
ZIP Code	0
County/Parish	0
Telephone Number	0
HCAHPS Measure ID	0
HCAHPS Question	0
HCAHPS Answer Description	0
Patient Survey Star Rating	410737
Patient Survey Star Rating Footnote	427688
HCAHPS Answer Percent	174732
HCAHPS Answer Percent Footnote	322299
HCAHPS Linear Mean Value	413903
Number of Completed Surveys	76911
Number of Completed Surveys Footnote	286347
Survey Response Rate Percent	76911
Survey Response Rate Percent Footnote	286347
Start Date	0
End Date	0

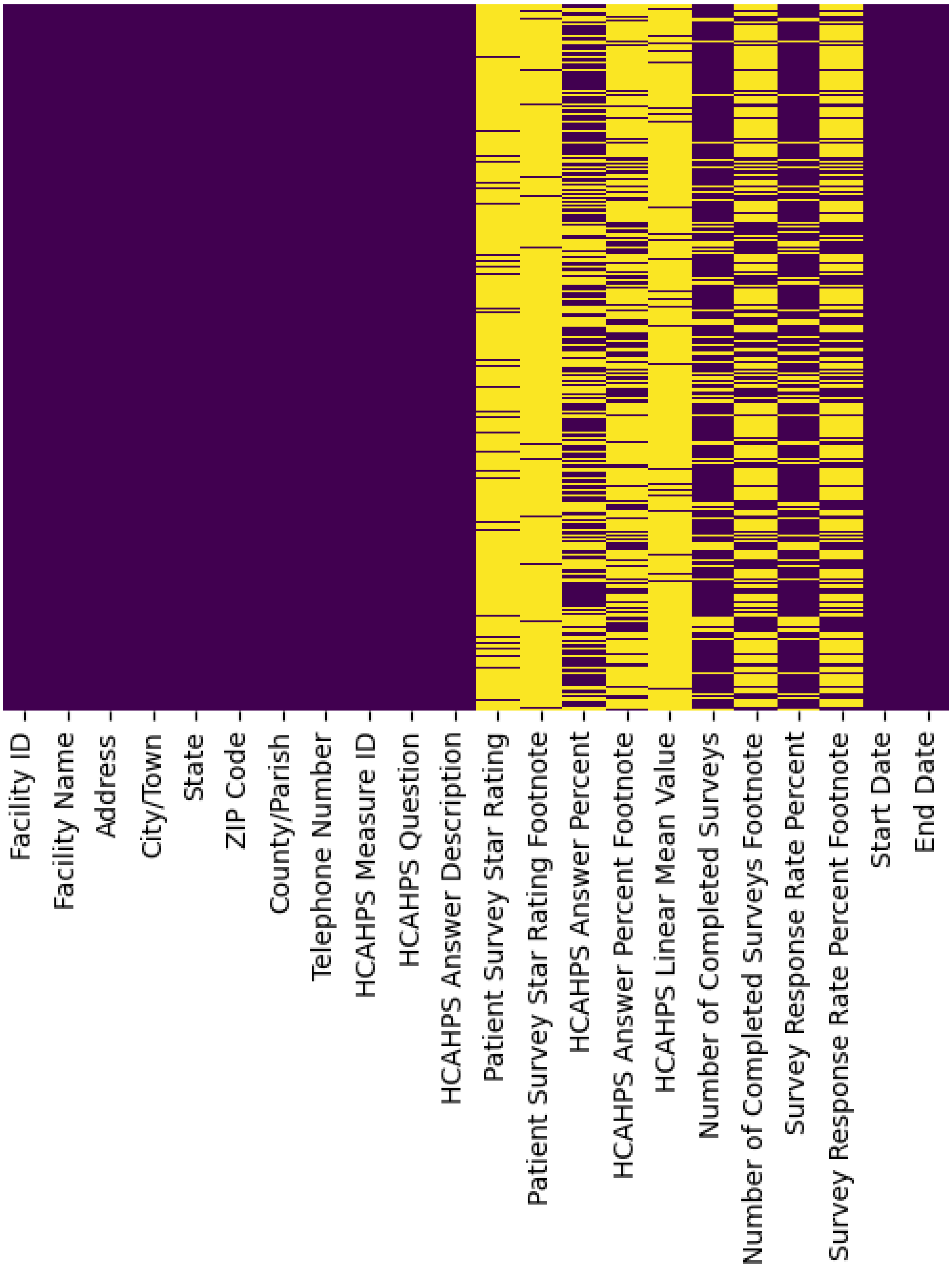
dtype: int64

In [53]:

```
sns.heatmap(hcahps.isnull(),yticklabels=False,cbar=False,cmap='viridis')
```

Out[53]:

```
<Axes: >
```



```
In [121…  #null_columns = [col for col in hcahps.columns if hcahps[col].isnull().sum() > 0
          #for col in null_columns:
              #print(f"Column: {col}")
              #print(hcahps[col].value_counts(dropna=True))
              #print("-------------------------------------------------------")
```

```
In [55]:  hcahps.drop(columns=['Patient Survey Star Rating Footnote','HCAHPS Answer Percen
                               'Survey Response Rate Percent Footnote'],inplace=True)
          hcahps.columns
```

```
Out[55]:  Index(['Facility ID', 'Facility Name', 'Address', 'City/Town', 'State',
                 'ZIP Code', 'County/Parish', 'Telephone Number', 'HCAHPS Measure ID',
                 'HCAHPS Question', 'HCAHPS Answer Description',
                 'Patient Survey Star Rating', 'HCAHPS Answer Percent',
                 'HCAHPS Linear Mean Value', 'Number of Completed Surveys',
                 'Survey Response Rate Percent', 'Start Date', 'End Date'],
                dtype='object')
```

In [56]:
```
hcahps.head()
```

Out[56]:

	Facility ID	Facility Name	Address	City/Town	State	ZIP Code	County/Parish	Telephone Number
0	010001	SOUTHEAST HEALTH MEDICAL CENTER	1108 ROSS CLARK CIRCLE	DOTHAN	AL	36301	HOUSTON	(334) 793-8701
1	010001	SOUTHEAST HEALTH MEDICAL CENTER	1108 ROSS CLARK CIRCLE	DOTHAN	AL	36301	HOUSTON	(334) 793-8701
2	010001	SOUTHEAST HEALTH MEDICAL CENTER	1108 ROSS CLARK CIRCLE	DOTHAN	AL	36301	HOUSTON	(334) 793-8701
3	010001	SOUTHEAST HEALTH MEDICAL CENTER	1108 ROSS CLARK CIRCLE	DOTHAN	AL	36301	HOUSTON	(334) 793-8701
4	010001	SOUTHEAST HEALTH MEDICAL CENTER	1108 ROSS CLARK CIRCLE	DOTHAN	AL	36301	HOUSTON	(334) 793-8701

In [57]:
```
hcahps[hcahps.duplicated()].shape
```

Out[57]:
```
(0, 18)
```

In [58]:
```
hcahps.describe()
```

Out[58]:

	ZIP Code	Patient Survey Star Rating	HCAHPS Answer Percent	HCAHPS Linear Mean Value	Number of Completed Surveys	R Rate
count	445563.000000	34826.000000	270831.000000	31660.000000	368652.000000	36865.
mean	54114.258192	3.218802	34.713530	85.276216	537.799193	2.
std	26956.222954	1.031572	28.956781	6.065431	695.393950	[illegible]
min	603.000000	1.000000	0.000000	53.000000	25.000000	[illegible]
25%	33021.000000	3.000000	10.000000	81.000000	126.000000	1
50%	55705.000000	3.000000	22.000000	86.000000	376.500000	2
75%	76132.000000	4.000000	61.000000	90.000000	617.250000	2
max	99929.000000	5.000000	100.000000	100.000000	12627.000000	7

Replacing Null :

Typically we will replace the numerical value columns with 'MEAN' but here it won't work because :

1. Artificially inflate performance of hospitals with no data
2. Bias the results toward the average
3. Rankings and insights become misleading

Case i) Go with '0' -> replacing with 0 means artificially labeling them as the worst hospital, which creates false negative insights.

Case ii) Go with 'MEDIAN' -> Represents central tendency without extreme bias

In [59]:

```
for col3 in cols:
    hcahps[col3] = hcahps[col3].fillna(hcahps[col3].median())
```

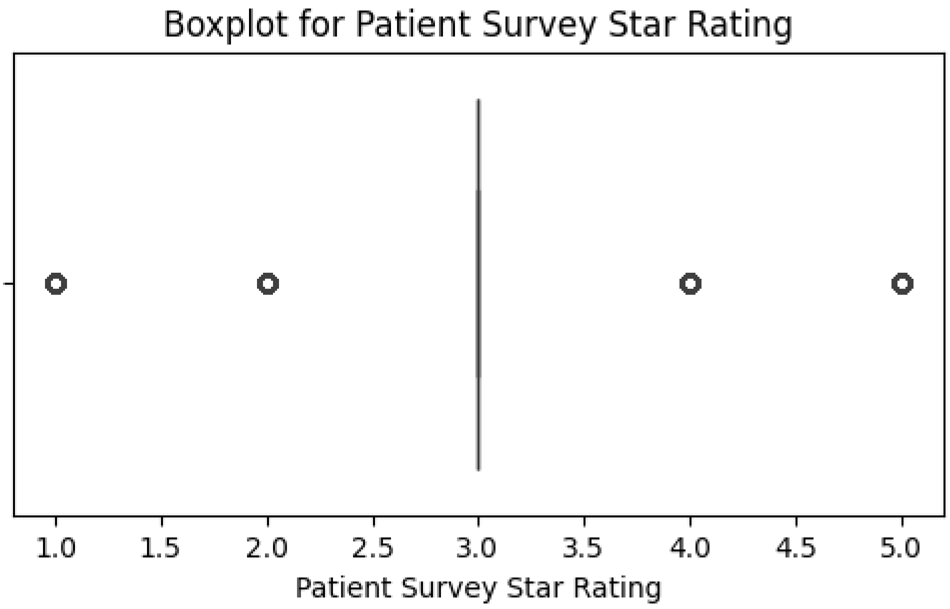
In [60]:

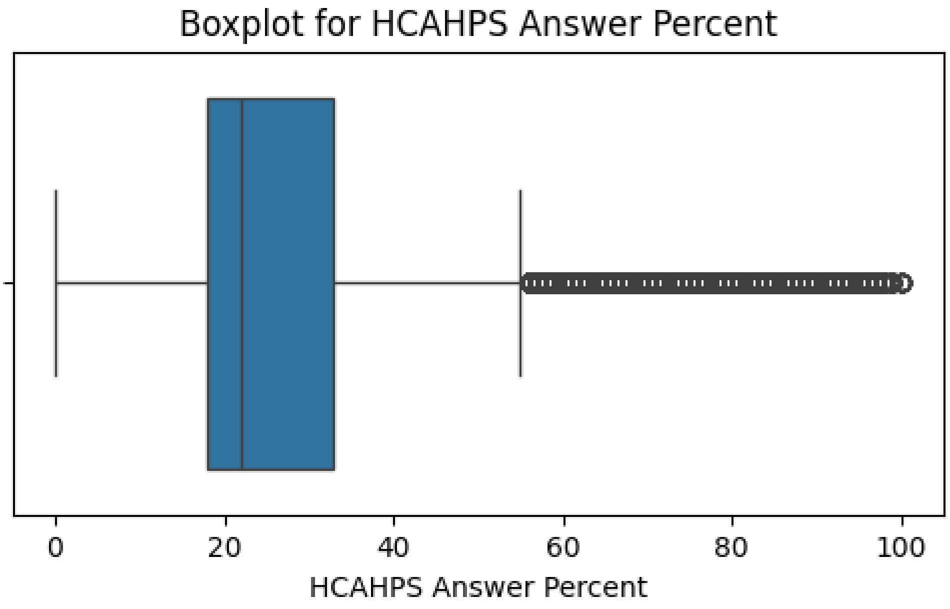
```
hcahps.isnull().sum()
```

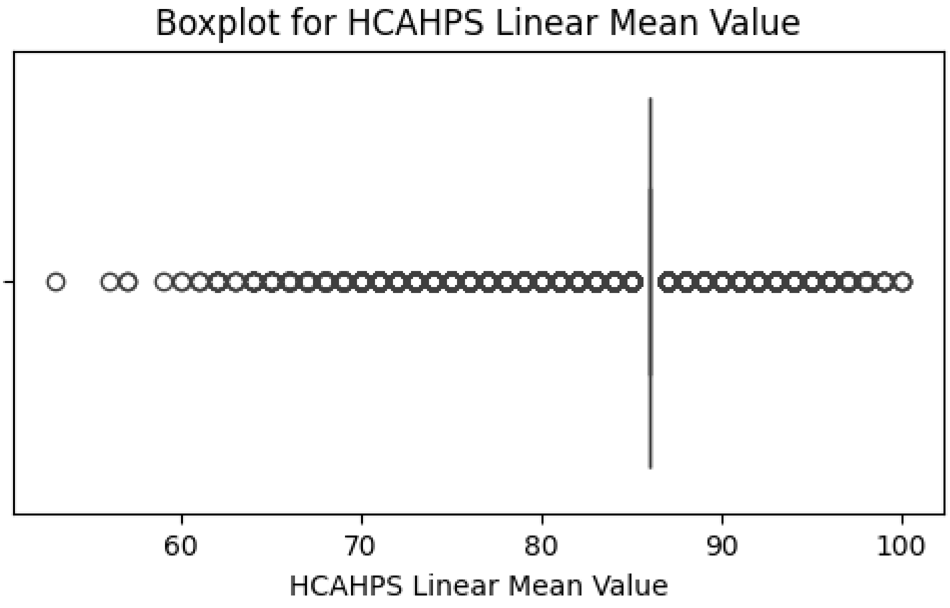
```
Out[60]:                                  0

                         Facility ID      0
                       Facility Name      0
                             Address      0
                           City/Town      0
                               State      0
                            ZIP Code      0
                       County/Parish      0
                    Telephone Number      0
                   HCAHPS Measure ID      0
                     HCAHPS Question      0
          HCAHPS Answer Description       0
          Patient Survey Star Rating      0
               HCAHPS Answer Percent      0
           HCAHPS Linear Mean Value       0
       Number of Completed Surveys        0
        Survey Response Rate Percent      0
                          Start Date      0
                            End Date      0

dtype: int64
```

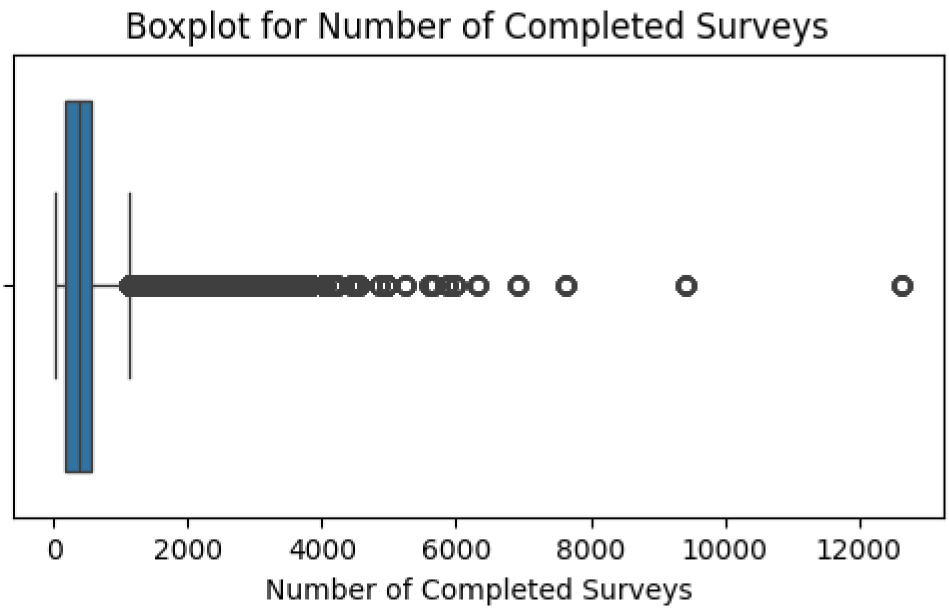
Outlier Detection:

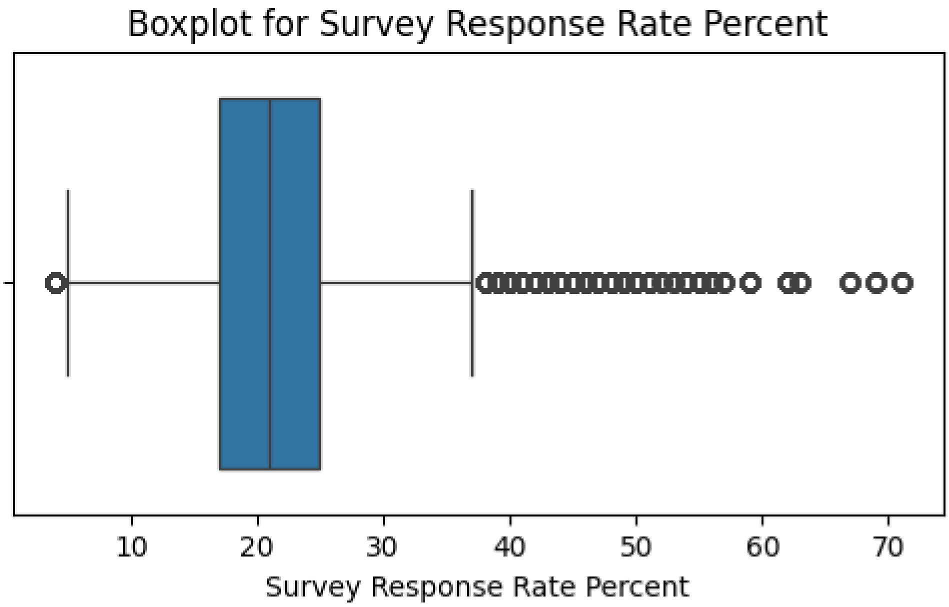
```
In [61]: for col3 in cols:
             plt.figure(figsize=(6,3))
             sns.boxplot(x=hcahps[col3])
             plt.title(f"Boxplot for {col3}")
             plt.show()
```


Boxplot for Patient Survey Star Rating
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
Patient Survey Star Rating


Boxplot for HCAHPS Answer Percent
0
20
40
60
80
100
HCAHPS Answer Percent


Boxplot for HCAHPS Linear Mean Value
60
70
80
90
100
HCAHPS Linear Mean Value





Observation:

We are not going to remove outliers here.Because in healthcare:

i) Outliers often represent real situations, not mistakes

ii) Extreme values may signal critical insights

iii) Removing them would hide true problems or distort findings

Hospital Gen_Info File Handling

In [62]:
```
Hosp= pd.read_csv("/content/Hospital_General_Information.csv")
```

In [63]:
```
Hosp.head()
```

Out[63]:

	Facility ID	Facility Name	Address	City/Town	State	ZIP Code	County/Parish	Telephone Number
0	010001	SOUTHEAST HEALTH MEDICAL CENTER	1108 ROSS CLARK CIRCLE	DOTHAN	AL	36301	HOUSTON	(334) 793-8701
1	010005	MARSHALL MEDICAL CENTERS	2505 U S HIGHWAY 431 NORTH	BOAZ	AL	35957	MARSHALL	(256) 593-8310
2	010006	NORTH ALABAMA MEDICAL CENTER	1701 VETERANS DRIVE	FLORENCE	AL	35630	LAUDERDALE	(256) 768-8400
3	010007	MIZELL MEMORIAL HOSPITAL	702 N MAIN ST	OPP	AL	36467	COVINGTON	(334) 493-3541
4	010008	CRENSHAW COMMUNITY HOSPITAL	101 HOSPITAL CIRCLE	LUVERNE	AL	36049	CRENSHAW	(334) 335-3374

In [64]:

```
Hosp.shape
```

Out[64]:

```
(5421, 38)
```

In [65]:

```
Hosp.isnull().sum()
```

Out[65]:

	0
Facility ID	0
Facility Name	0
Address	0
City/Town	0
State	0
ZIP Code	0
County/Parish	0
Telephone Number	0
Hospital Type	0
Hospital Ownership	0
Emergency Services	0
Meets criteria for birthing friendly designation	3157
Hospital overall rating	0
Hospital overall rating footnote	2857
MORT Group Measure Count	0
Count of Facility MORT Measures	0
Count of MORT Measures Better	0
Count of MORT Measures No Different	0
Count of MORT Measures Worse	0
MORT Group Footnote	3643
Safety Group Measure Count	0
Count of Facility Safety Measures	0
Count of Safety Measures Better	0
Count of Safety Measures No Different	0
Count of Safety Measures Worse	0
Safety Group Footnote	3350
READM Group Measure Count	0
Count of Facility READM Measures	0
Count of READM Measures Better	0
Count of READM Measures No Different	0
Count of READM Measures Worse	0
READM Group Footnote	4271
Pt Exp Group Measure Count	0

	0
Count of Facility Pt Exp Measures	0
Pt Exp Group Footnote	3154
TE Group Measure Count	0
Count of Facility TE Measures	0
TE Group Footnote	4493

dtype: int64

```
In [66]: Hosp.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5421 entries, 0 to 5420
Data columns (total 38 columns):
 #   Column                                         Non-Null Count  Dtype  
---  ------                                         --------------  -----  
 0   Facility ID                                    5421 non-null   object 
 1   Facility Name                                  5421 non-null   object 
 2   Address                                        5421 non-null   object 
 3   City/Town                                      5421 non-null   object 
 4   State                                          5421 non-null   object 
 5   ZIP Code                                       5421 non-null   int64  
 6   County/Parish                                  5421 non-null   object 
 7   Telephone Number                               5421 non-null   object 
 8   Hospital Type                                  5421 non-null   object 
 9   Hospital Ownership                             5421 non-null   object 
 10  Emergency Services                             5421 non-null   object 
 11  Meets criteria for birthing friendly designation  2264 non-null   object 
 12  Hospital overall rating                        5421 non-null   object 
 13  Hospital overall rating footnote               2564 non-null   object 
 14  MORT Group Measure Count                       5421 non-null   object 
 15  Count of Facility MORT Measures                5421 non-null   object 
 16  Count of MORT Measures Better                  5421 non-null   object 
 17  Count of MORT Measures No Different            5421 non-null   object 
 18  Count of MORT Measures Worse                   5421 non-null   object 
 19  MORT Group Footnote                            1778 non-null   float64
 20  Safety Group Measure Count                     5421 non-null   object 
 21  Count of Facility Safety Measures              5421 non-null   object 
 22  Count of Safety Measures Better                5421 non-null   object 
 23  Count of Safety Measures No Different          5421 non-null   object 
 24  Count of Safety Measures Worse                 5421 non-null   object 
 25  Safety Group Footnote                          2071 non-null   float64
 26  READM Group Measure Count                      5421 non-null   object 
 27  Count of Facility READM Measures               5421 non-null   object 
 28  Count of READM Measures Better                 5421 non-null   object 
 29  Count of READM Measures No Different           5421 non-null   object 
 30  Count of READM Measures Worse                  5421 non-null   object 
 31  READM Group Footnote                           1150 non-null   float64
 32  Pt Exp Group Measure Count                     5421 non-null   object 
 33  Count of Facility Pt Exp Measures              5421 non-null   object 
 34  Pt Exp Group Footnote                          2267 non-null   float64
 35  TE Group Measure Count                         5421 non-null   object 
 36  Count of Facility TE Measures                  5421 non-null   object 
 37  TE Group Footnote                              928 non-null    float64
dtypes: float64(5), int64(1), object(32)
memory usage: 1.6+ MB
```

In [67]:
```
clmn=Hosp.columns[12:]
Hosp[clmn] = Hosp[clmn].apply(pd.to_numeric, errors='coerce')
```

In [68]:
```
Hosp.drop(columns=['Hospital overall rating footnote','MORT Group Footnote',
                   'Safety Group Footnote','READM Group Footnote','TE Group Foot
```

In [69]:
```
Hosp[Hosp.duplicated()].shape
```

Out[69]:
```
(0, 33)
```

Null Handling

In [70]:
```
Hosp.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5421 entries, 0 to 5420
Data columns (total 33 columns):
 #   Column                                        Non-Null Count  Dtype  
---  ------                                        --------------  -----  
 0   Facility ID                                   5421 non-null   object 
 1   Facility Name                                 5421 non-null   object 
 2   Address                                       5421 non-null   object 
 3   City/Town                                     5421 non-null   object 
 4   State                                         5421 non-null   object 
 5   ZIP Code                                      5421 non-null   int64  
 6   County/Parish                                 5421 non-null   object 
 7   Telephone Number                              5421 non-null   object 
 8   Hospital Type                                 5421 non-null   object 
 9   Hospital Ownership                            5421 non-null   object 
 10  Emergency Services                            5421 non-null   object 
 11  Meets criteria for birthing friendly designation  2264 non-null   object 
 12  Hospital overall rating                       2869 non-null   float64
 13  MORT Group Measure Count                      4558 non-null   float64
 14  Count of Facility MORT Measures               3644 non-null   float64
 15  Count of MORT Measures Better                 3644 non-null   float64
 16  Count of MORT Measures No Different           3644 non-null   float64
 17  Count of MORT Measures Worse                  3644 non-null   float64
 18  Safety Group Measure Count                    4558 non-null   float64
 19  Count of Facility Safety Measures             3356 non-null   float64
 20  Count of Safety Measures Better               3356 non-null   float64
 21  Count of Safety Measures No Different         3356 non-null   float64
 22  Count of Safety Measures Worse                3356 non-null   float64
 23  READM Group Measure Count                     4558 non-null   float64
 24  Count of Facility READM Measures              4271 non-null   float64
 25  Count of READM Measures Better                4271 non-null   float64
 26  Count of READM Measures No Different          4271 non-null   float64
 27  Count of READM Measures Worse                 4271 non-null   float64
 28  Pt Exp Group Measure Count                    4558 non-null   float64
 29  Count of Facility Pt Exp Measures             3154 non-null   float64
 30  Pt Exp Group Footnote                         2267 non-null   float64
 31  TE Group Measure Count                        4558 non-null   float64
 32  Count of Facility TE Measures                 4493 non-null   float64
dtypes: float64(21), int64(1), object(11)
memory usage: 1.4+ MB
```

In [71]:
```
Hosp.isnull().sum()
```

Out[71]:

	0
Facility ID	0
Facility Name	0
Address	0
City/Town	0
State	0
ZIP Code	0
County/Parish	0
Telephone Number	0
Hospital Type	0
Hospital Ownership	0
Emergency Services	0
Meets criteria for birthing friendly designation	3157
Hospital overall rating	2552
MORT Group Measure Count	863
Count of Facility MORT Measures	1777
Count of MORT Measures Better	1777
Count of MORT Measures No Different	1777
Count of MORT Measures Worse	1777
Safety Group Measure Count	863
Count of Facility Safety Measures	2065
Count of Safety Measures Better	2065
Count of Safety Measures No Different	2065
Count of Safety Measures Worse	2065
READM Group Measure Count	863
Count of Facility READM Measures	1150
Count of READM Measures Better	1150
Count of READM Measures No Different	1150
Count of READM Measures Worse	1150
Pt Exp Group Measure Count	863
Count of Facility Pt Exp Measures	2267
Pt Exp Group Footnote	3154
TE Group Measure Count	863
Count of Facility TE Measures	928

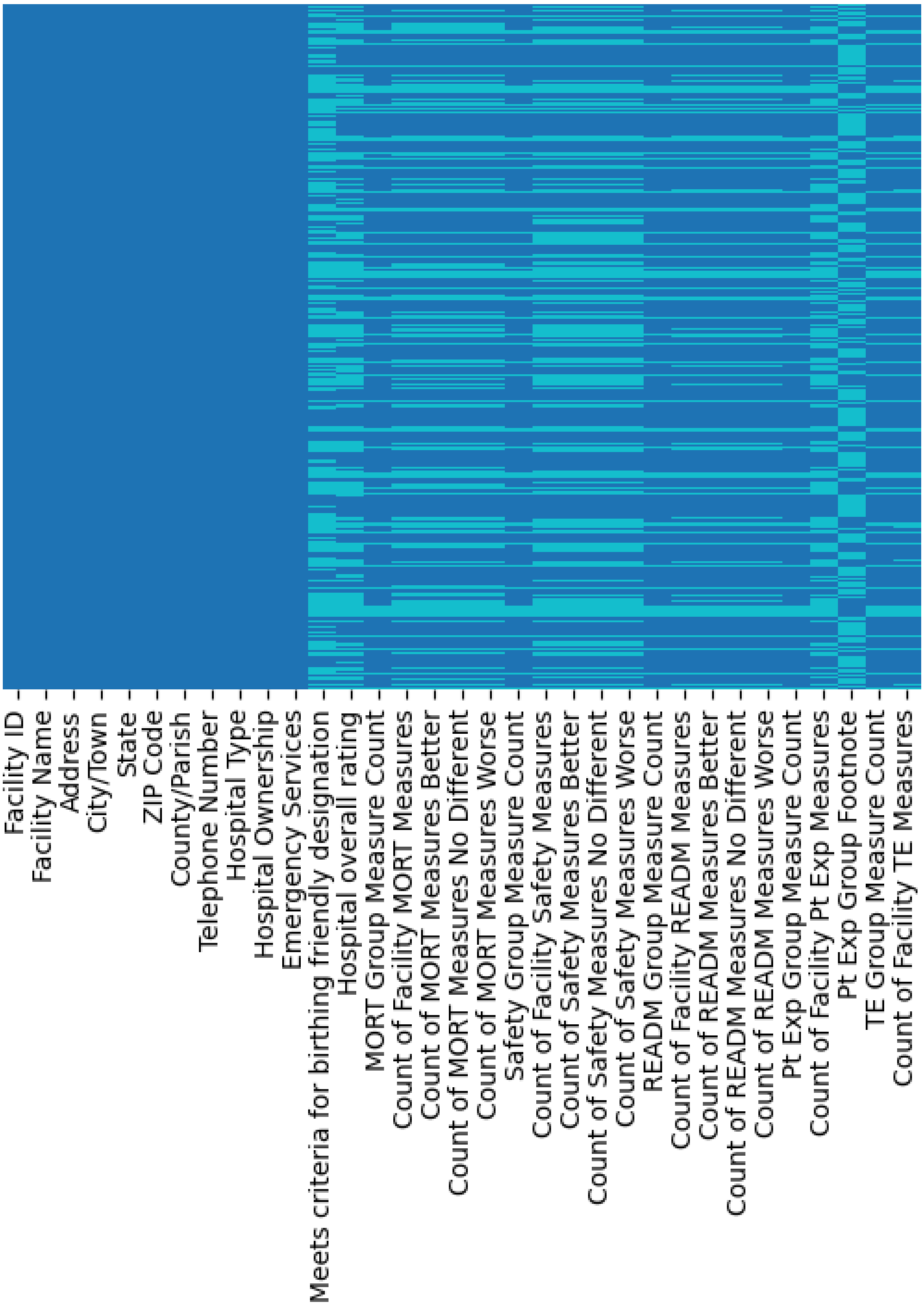
dtype: int64

In [72]:

```
sns.heatmap(Hosp.isnull(),yticklabels=False,cbar=False,cmap='tab10')
```

Out[72]:

```
<Axes: >
```



In [73]:

```
Clm =Hosp.columns[12:]
for cl in Clm:
  Hosp[cl]=Hosp[cl].fillna(Hosp[cl].median())
```

In [74]:

```
Hosp.isnull().sum()
```

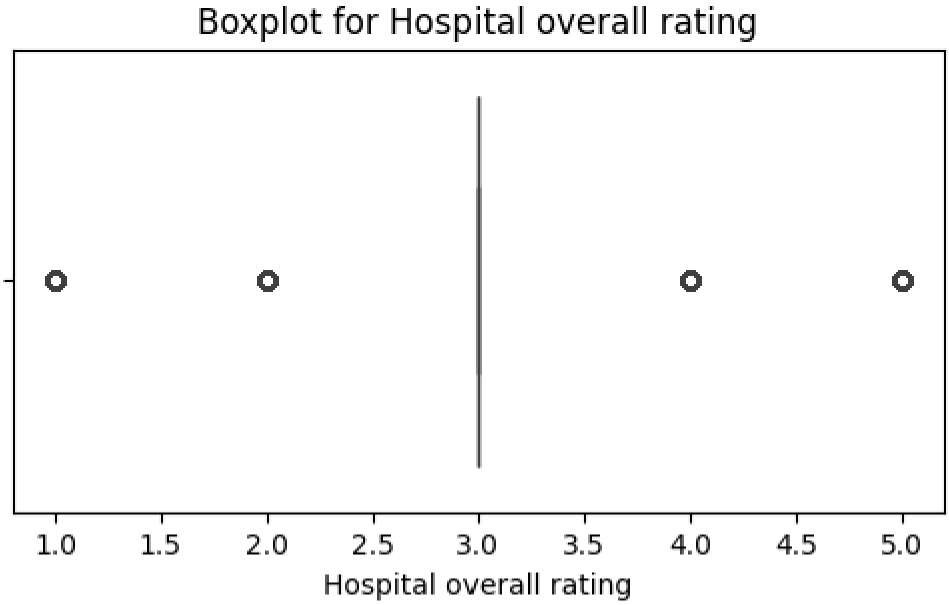
Out[74]:

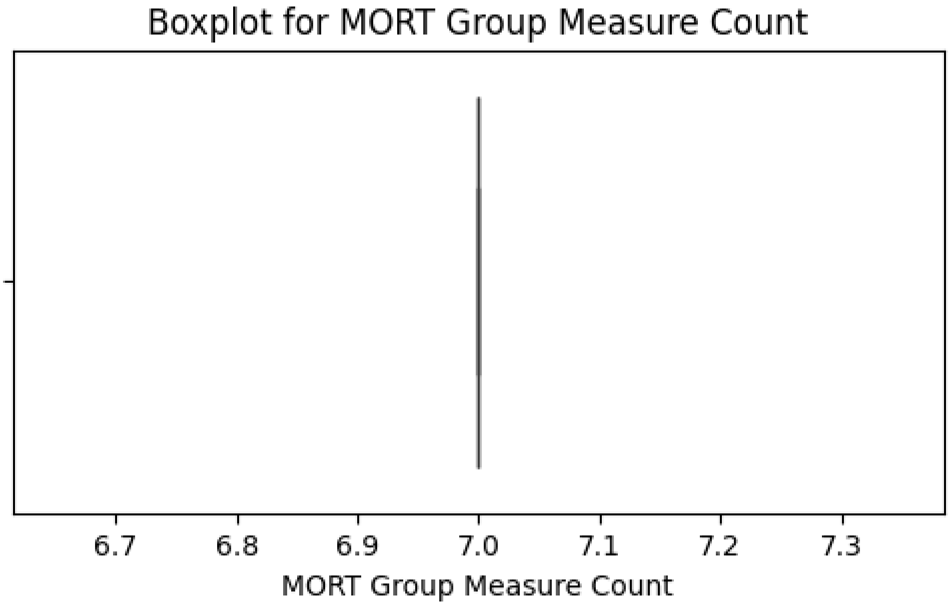
	0
Facility ID	0
Facility Name	0
Address	0
City/Town	0
State	0
ZIP Code	0
County/Parish	0
Telephone Number	0
Hospital Type	0
Hospital Ownership	0
Emergency Services	0
Meets criteria for birthing friendly designation	3157
Hospital overall rating	0
MORT Group Measure Count	0
Count of Facility MORT Measures	0
Count of MORT Measures Better	0
Count of MORT Measures No Different	0
Count of MORT Measures Worse	0
Safety Group Measure Count	0
Count of Facility Safety Measures	0
Count of Safety Measures Better	0
Count of Safety Measures No Different	0
Count of Safety Measures Worse	0
READM Group Measure Count	0
Count of Facility READM Measures	0
Count of READM Measures Better	0
Count of READM Measures No Different	0
Count of READM Measures Worse	0
Pt Exp Group Measure Count	0
Count of Facility Pt Exp Measures	0
Pt Exp Group Footnote	0
TE Group Measure Count	0
Count of Facility TE Measures	0

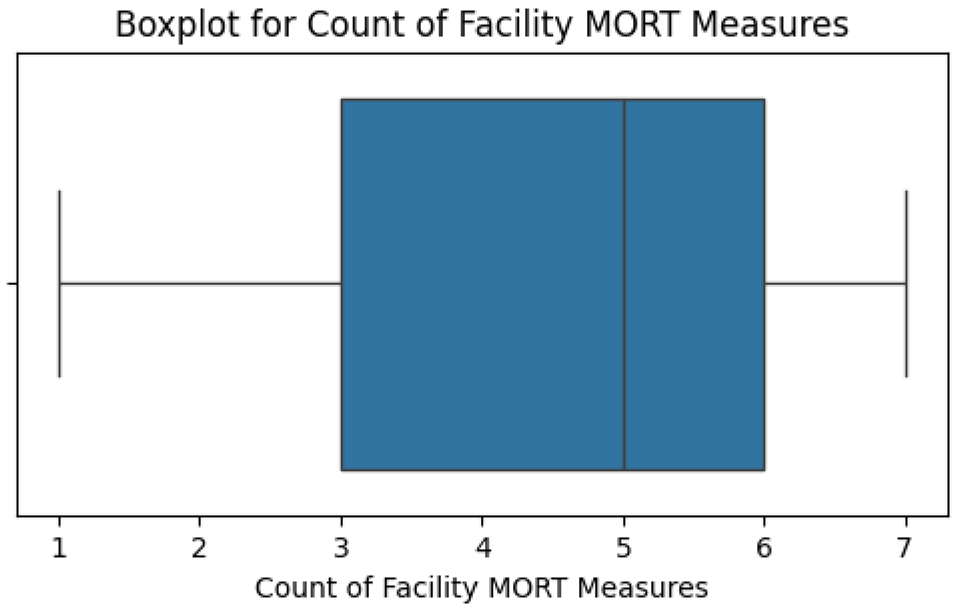
dtype: int64

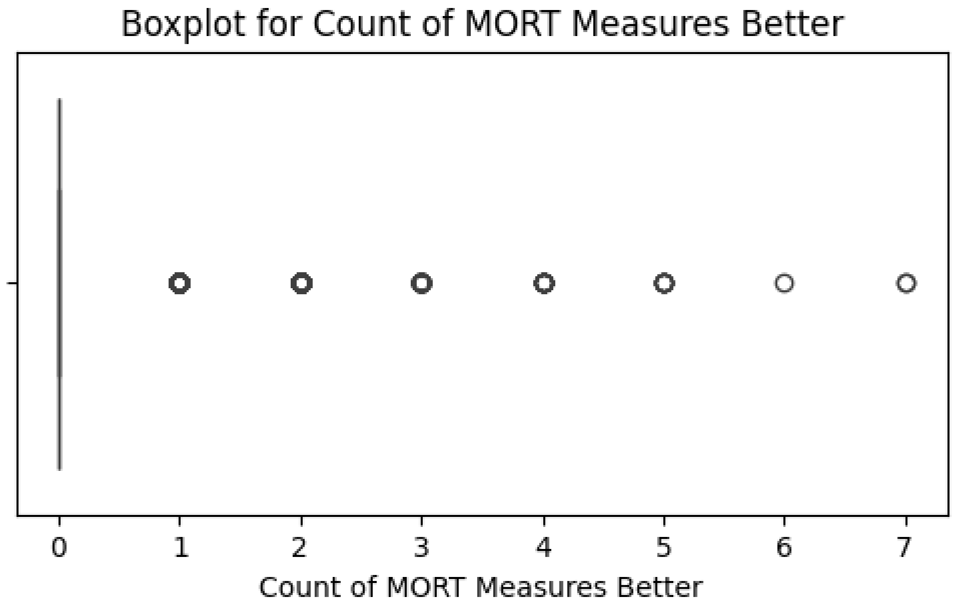
Outlier Detection

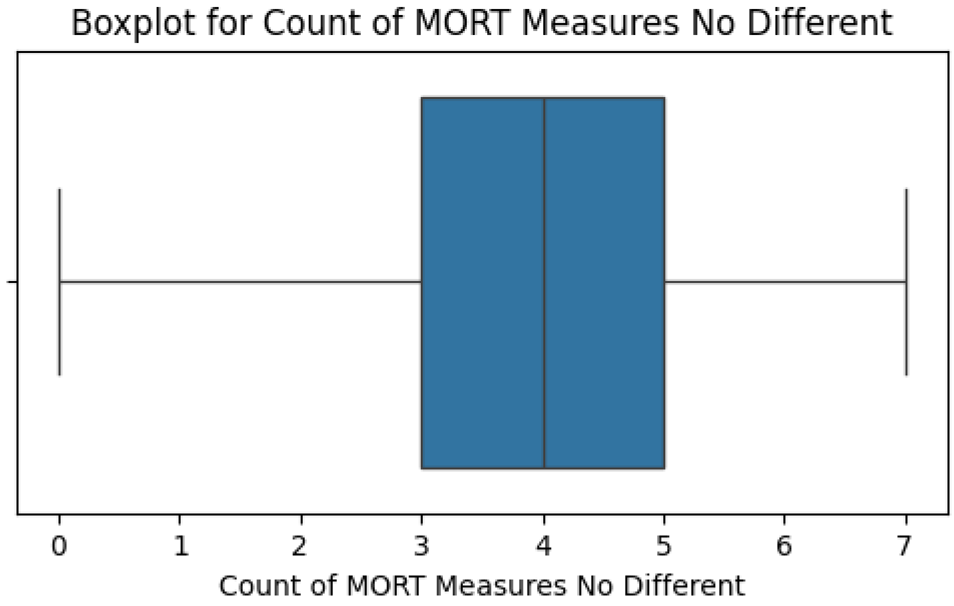
In [75]:
```
for cl in Clm:
  plt.figure(figsize=(6,3))
  sns.boxplot(x=Hosp[cl])
  plt.title(f"Boxplot for {cl}")
  plt.show()
```

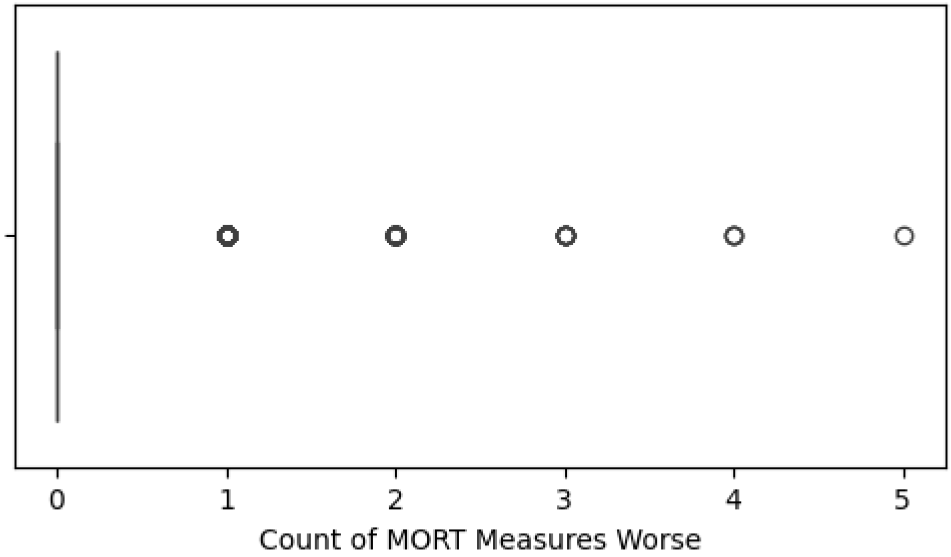




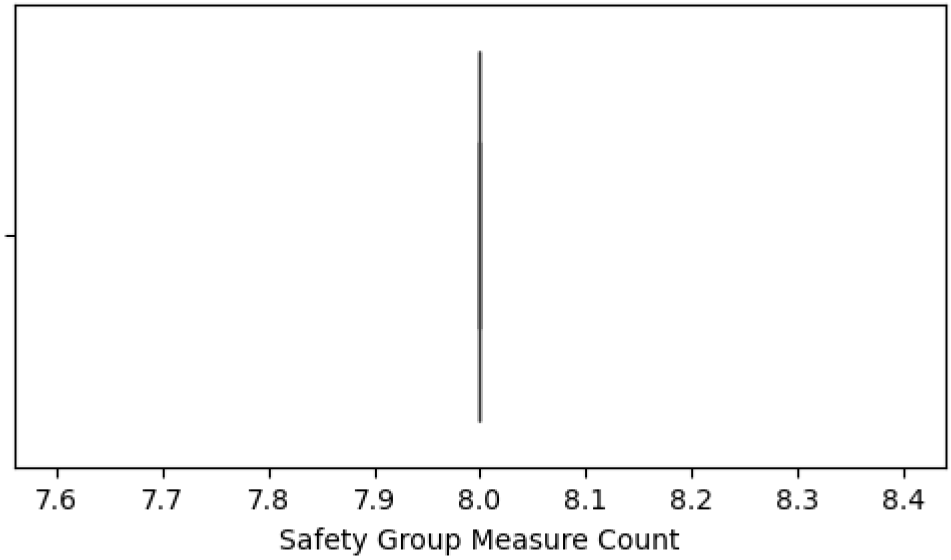

Boxplot for Count of Facility MORT Measures
1
2
3
4
5
6
7
Count of Facility MORT Measures


Boxplot for Count of MORT Measures Better
0
1
2
3
4
5
6
7
Count of MORT Measures Better

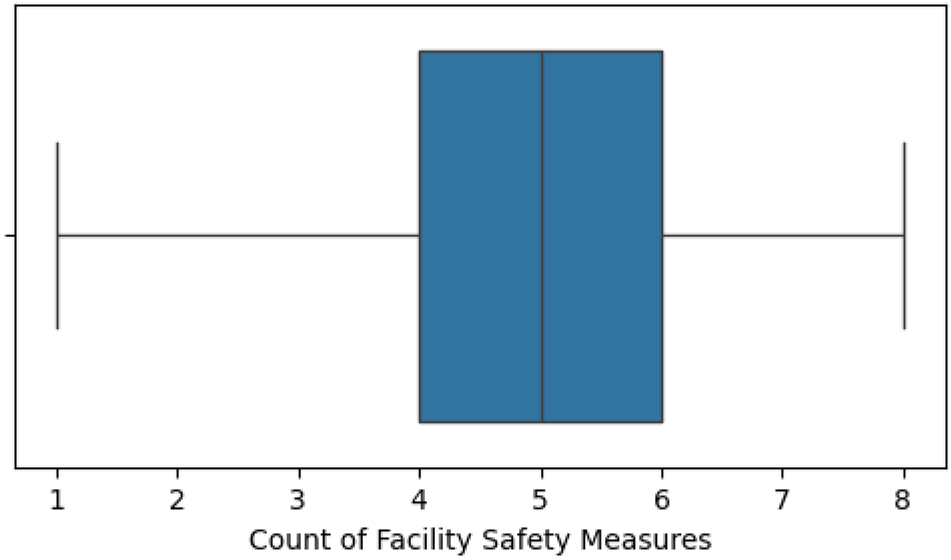

Boxplot for Count of MORT Measures No Different
0
1
2
3
4
5
6
7
Count of MORT Measures No Different



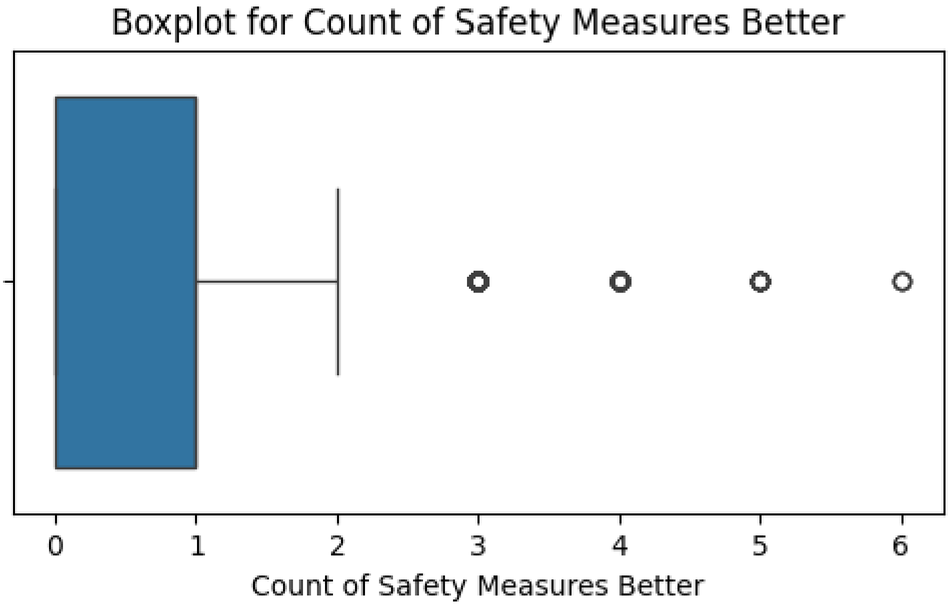
Boxplot for Count of MORT Measures Worse

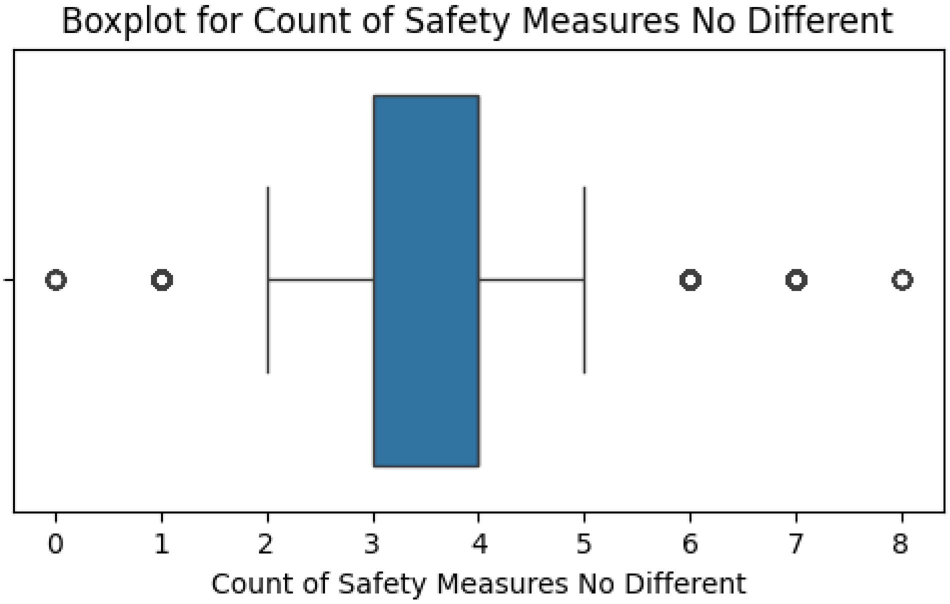


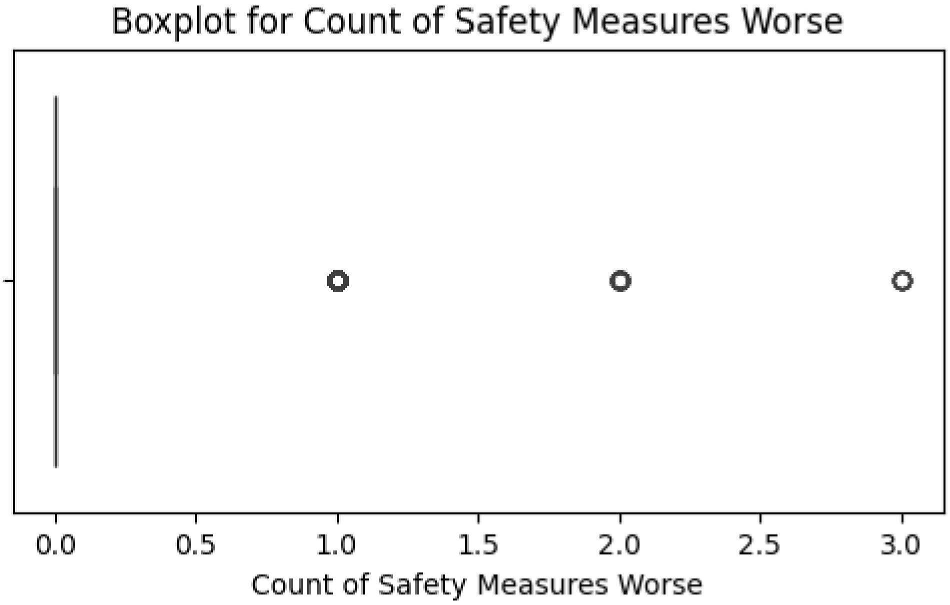
Boxplot for Safety Group Measure Count

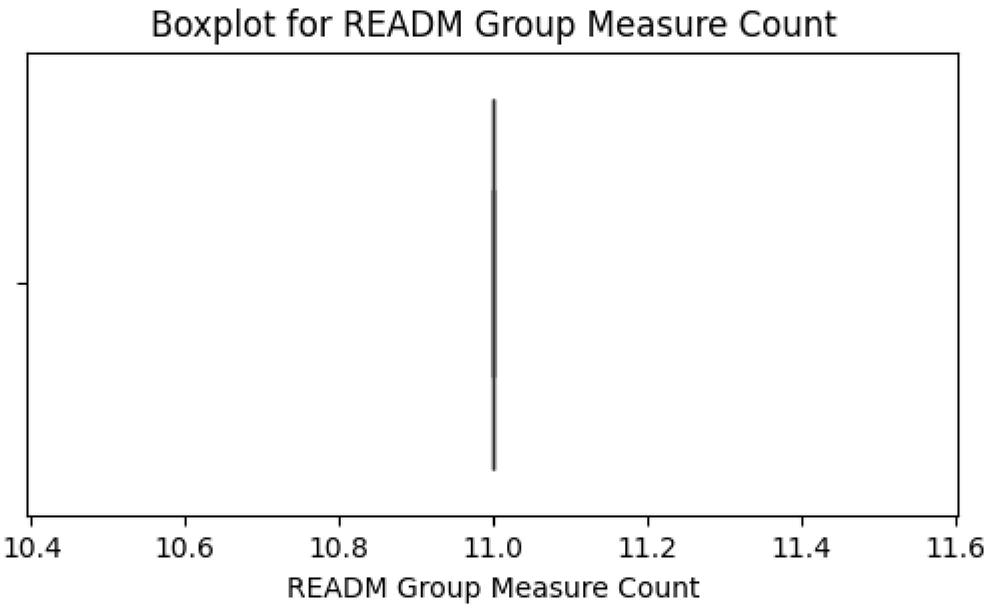


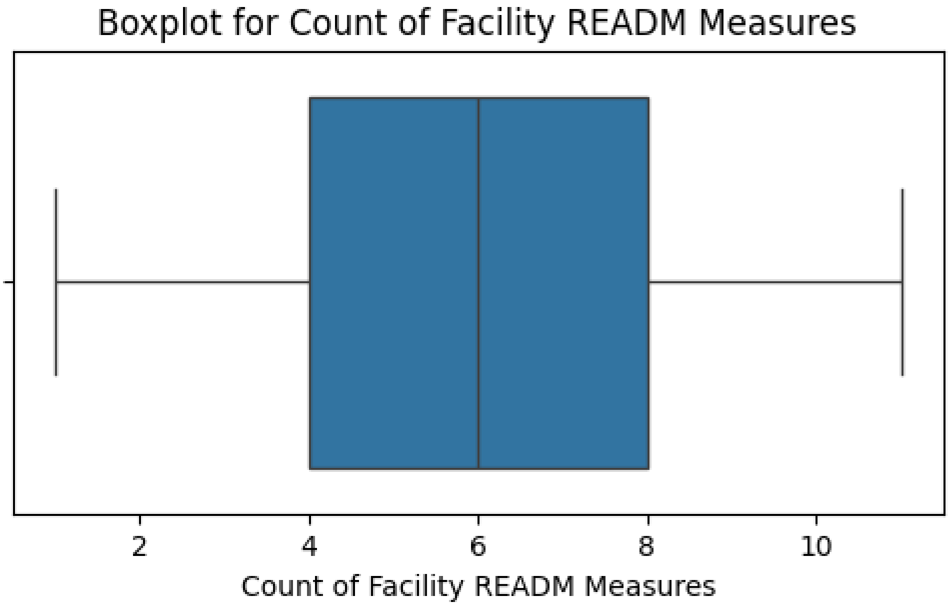
Boxplot for Count of Facility Safety Measures

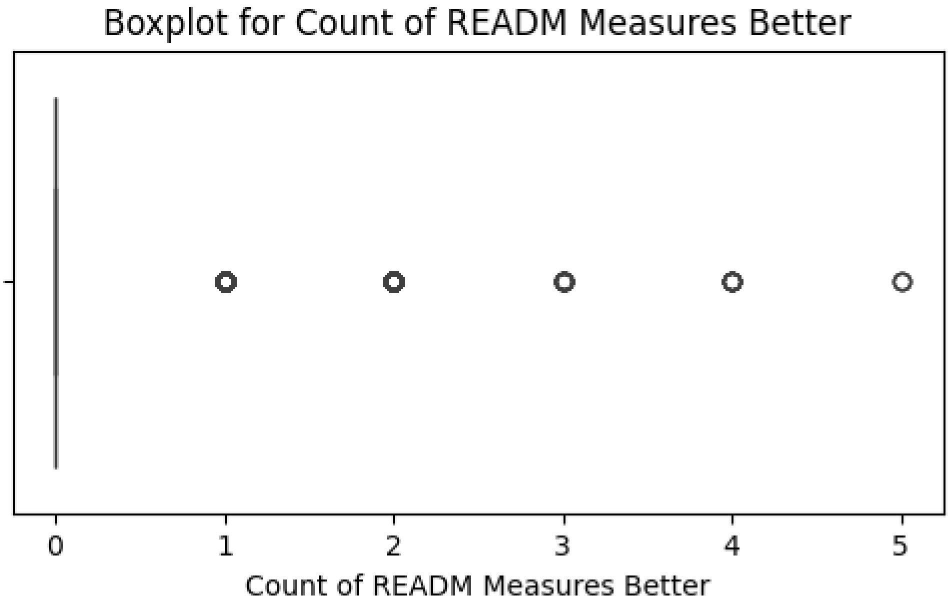

Boxplot for Count of Safety Measures Better
0
1
2
3
4
5
6
Count of Safety Measures Better

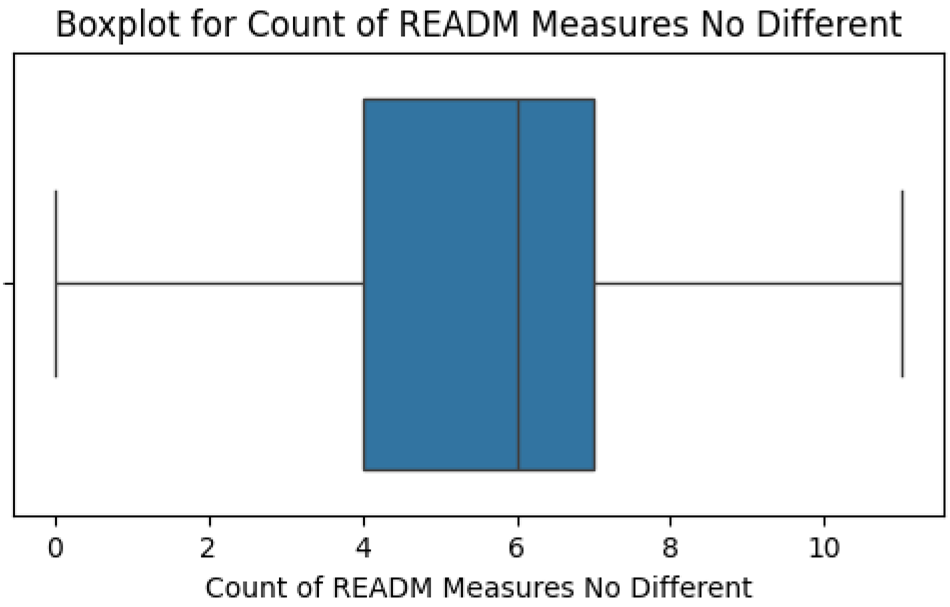

Boxplot for Count of Safety Measures No Different
0
1
2
3
4
5
6
7
8
Count of Safety Measures No Different

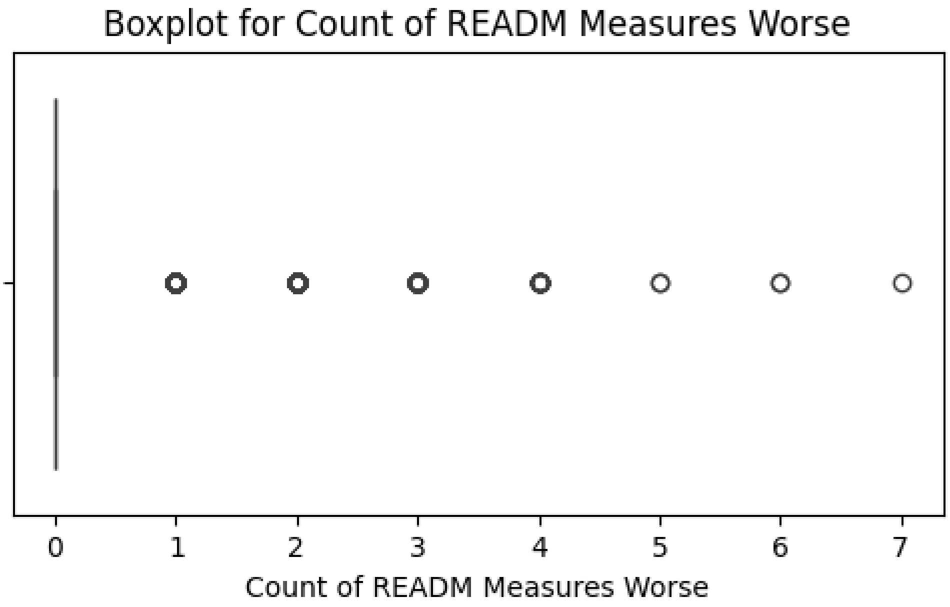

Boxplot for Count of Safety Measures Worse
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Count of Safety Measures Worse

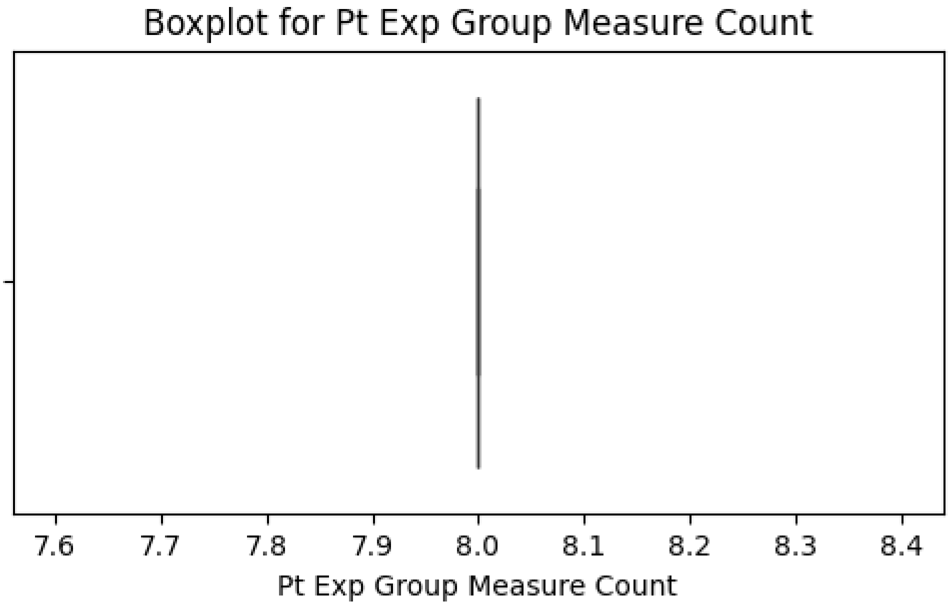

Boxplot for READM Group Measure Count
10.4
10.6
10.8
11.0
11.2
11.4
11.6
READM Group Measure Count

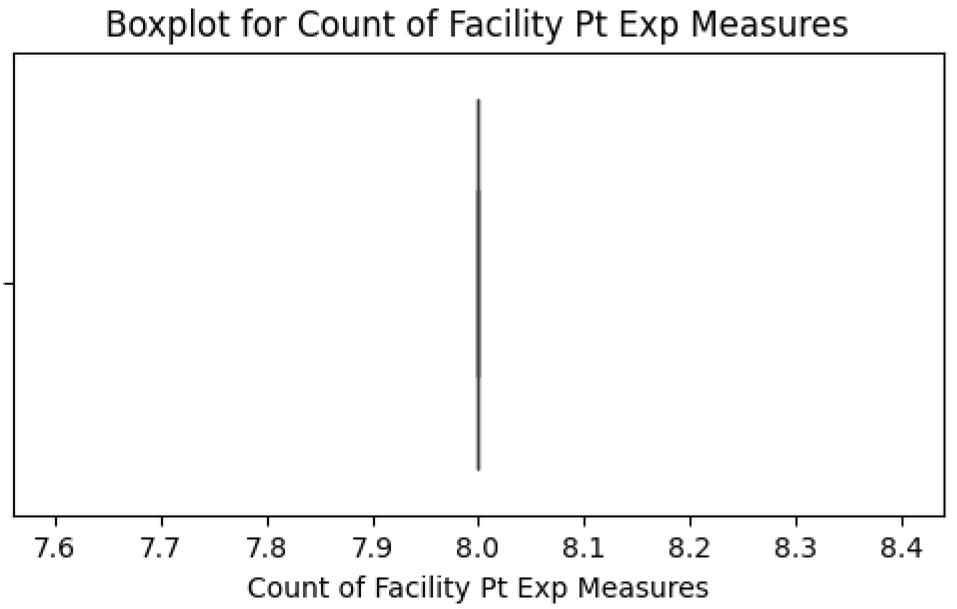

Boxplot for Count of Facility READM Measures
2
4
6
8
10
Count of Facility READM Measures

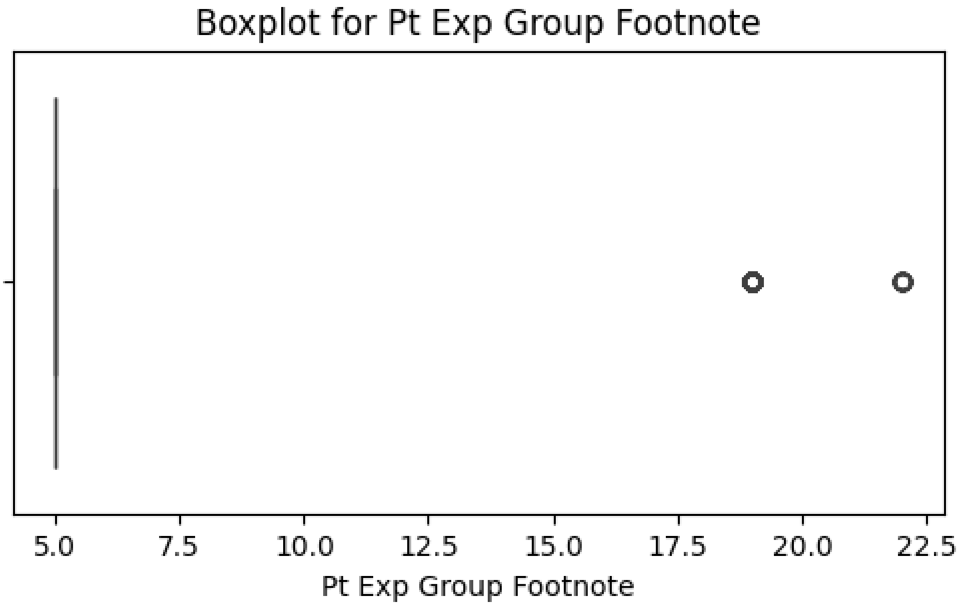

Boxplot for Count of READM Measures Better
0
1
2
3
4
5
Count of READM Measures Better

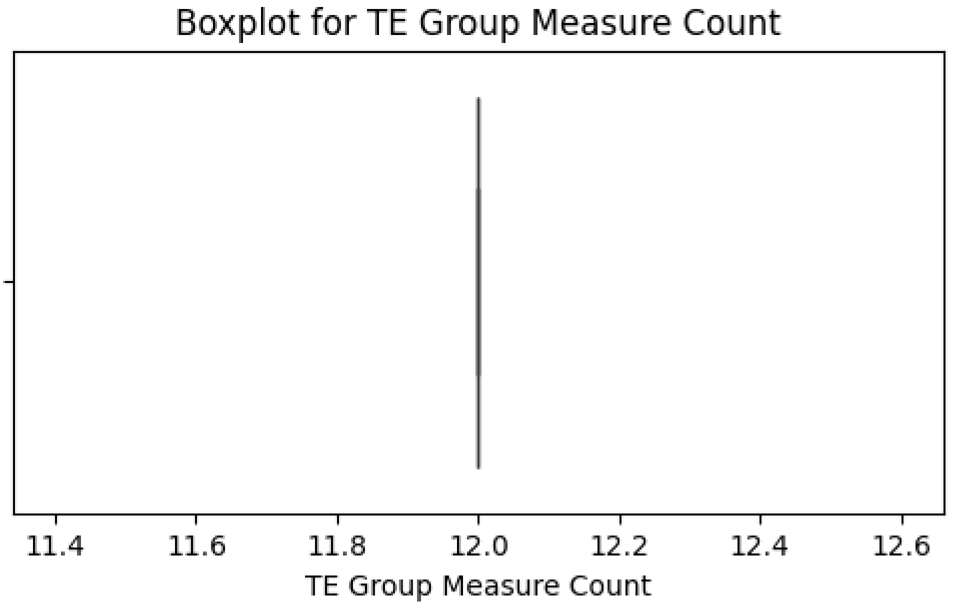

Boxplot for Count of READM Measures No Different
0
2
4
6
8
10
Count of READM Measures No Different

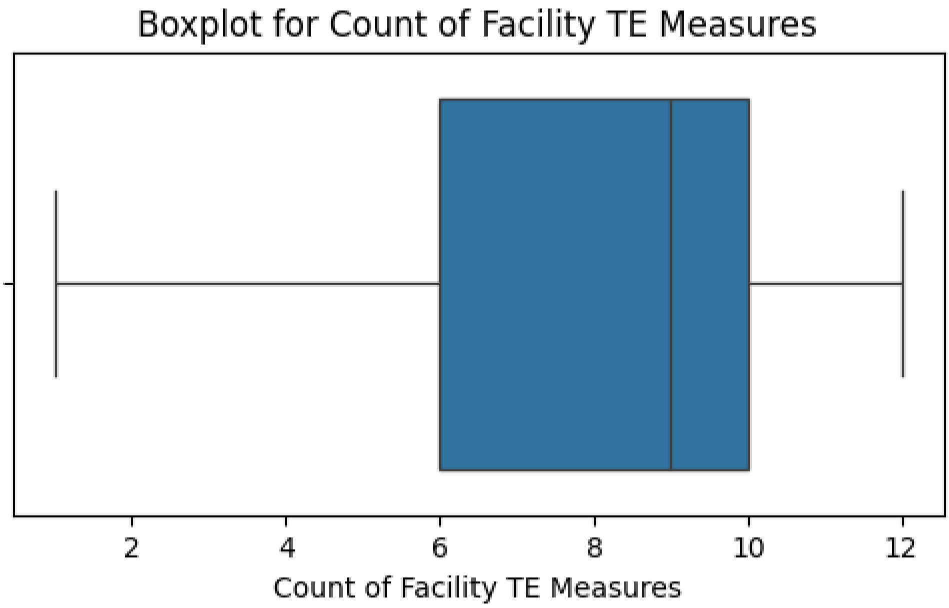

Boxplot for Count of READM Measures Worse
0
1
2
3
4
5
6
7
Count of READM Measures Worse


Boxplot for Pt Exp Group Measure Count
7.6
7.7
7.8
7.9
8.0
8.1
8.2
8.3
8.4
Pt Exp Group Measure Count


Boxplot for Count of Facility Pt Exp Measures


Boxplot for Pt Exp Group Footnote


Boxplot for TE Group Measure Count



```
In [75]:
```

Complications File Handling

```
In [76]: comp = pd.read_csv('/content/Complications_and_Deaths-Hospital.csv')
```

```
In [77]: comp.head()
```

Out[77]:

	Facility ID	Facility Name	Address	City/Town	State	ZIP Code	County/Parish	Telephone Number
0	010001	SOUTHEAST HEALTH MEDICAL CENTER	1108 ROSS CLARK CIRCLE	DOTHAN	AL	36301	HOUSTON	(334) 793-8701
1	010001	SOUTHEAST HEALTH MEDICAL CENTER	1108 ROSS CLARK CIRCLE	DOTHAN	AL	36301	HOUSTON	(334) 793-8701
2	010001	SOUTHEAST HEALTH MEDICAL CENTER	1108 ROSS CLARK CIRCLE	DOTHAN	AL	36301	HOUSTON	(334) 793-8701
3	010001	SOUTHEAST HEALTH MEDICAL CENTER	1108 ROSS CLARK CIRCLE	DOTHAN	AL	36301	HOUSTON	(334) 793-8701
4	010001	SOUTHEAST HEALTH MEDICAL CENTER	1108 ROSS CLARK CIRCLE	DOTHAN	AL	36301	HOUSTON	(334) 793-8701

In [78]:
```
comp.shape
```
Out[78]:
```
(95820, 18)
```
In [79]:
```
comp[comp.duplicated].shape
```
Out[79]:
```
(0, 18)
```
In [80]:
```
comp.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 95820 entries, 0 to 95819
Data columns (total 18 columns):
 #   Column               Non-Null Count  Dtype 
---  ------               --------------  ----- 
 0   Facility ID          95820 non-null  object
 1   Facility Name        95820 non-null  object
 2   Address              95820 non-null  object
 3   City/Town            95820 non-null  object
 4   State                95820 non-null  object
 5   ZIP Code             95820 non-null  int64 
 6   County/Parish        95820 non-null  object
 7   Telephone Number     95820 non-null  object
 8   Measure ID           95820 non-null  object
 9   Measure Name         95820 non-null  object
 10  Compared to National 95820 non-null  object
 11  Denominator          95820 non-null  object
 12  Score                95820 non-null  object
 13  Lower Estimate       95820 non-null  object
 14  Higher Estimate      95820 non-null  object
 15  Footnote             44867 non-null  object
 16  Start Date           95820 non-null  object
 17  End Date             95820 non-null  object
dtypes: int64(1), object(17)
memory usage: 13.2+ MB
```

In [81]:

```
comp.isnull().sum()
```

Out[81]:

	0
Facility ID	0
Facility Name	0
Address	0
City/Town	0
State	0
ZIP Code	0
County/Parish	0
Telephone Number	0
Measure ID	0
Measure Name	0
Compared to National	0
Denominator	0
Score	0
Lower Estimate	0
Higher Estimate	0
Footnote	50953
Start Date	0
End Date	0

dtype: int64

In [82]:

```
comp.drop('Footnote',inplace=True,axis=1)
```

In [83]:

```
colu = comp.columns[11:15]
comp[colu] = comp[colu].apply(pd.to_numeric, errors='coerce')
```

In [84]:

```
comp.info()
```

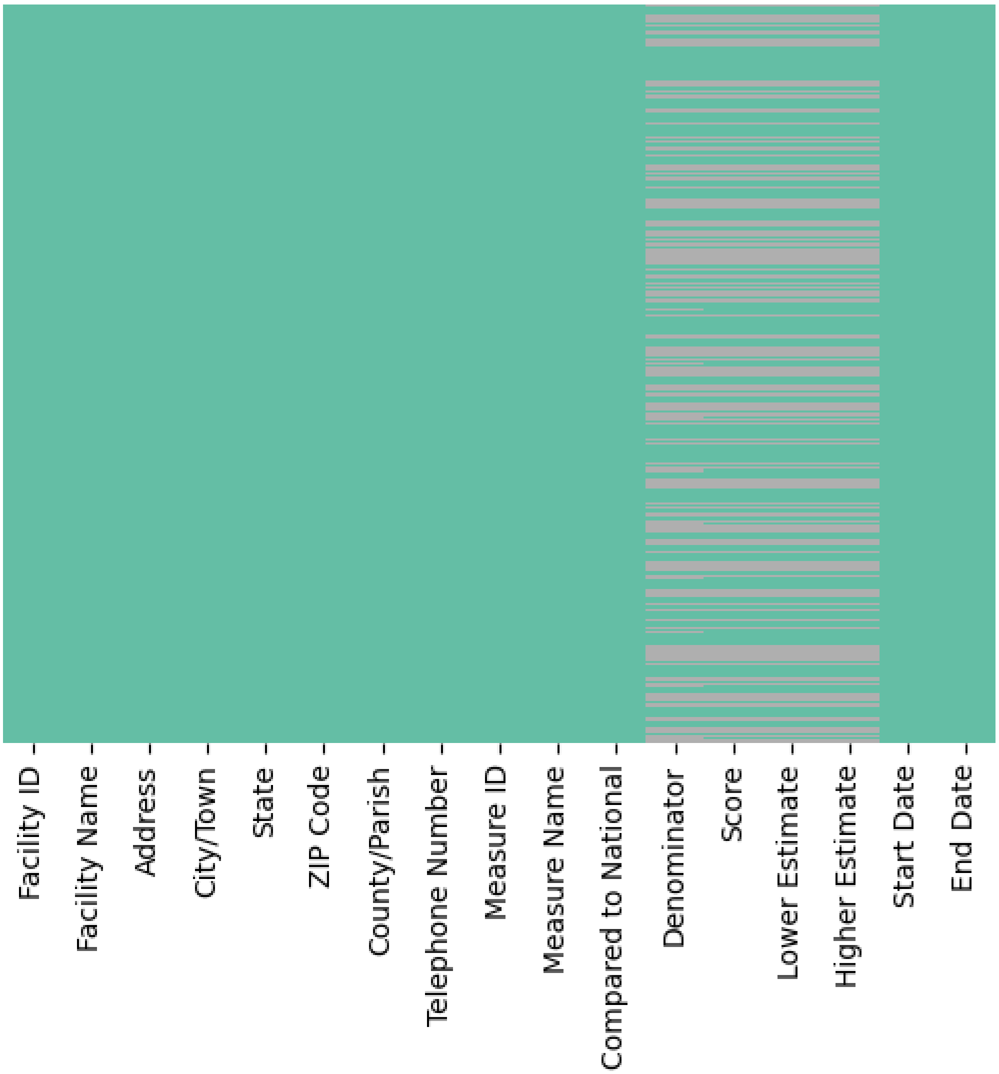
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 95820 entries, 0 to 95819
Data columns (total 17 columns):
 #   Column                Non-Null Count  Dtype  
---  ------                --------------  -----  
 0   Facility ID           95820 non-null  object 
 1   Facility Name         95820 non-null  object 
 2   Address               95820 non-null  object 
 3   City/Town             95820 non-null  object 
 4   State                 95820 non-null  object 
 5   ZIP Code              95820 non-null  int64  
 6   County/Parish         95820 non-null  object 
 7   Telephone Number      95820 non-null  object 
 8   Measure ID            95820 non-null  object 
 9   Measure Name          95820 non-null  object 
 10  Compared to National  95820 non-null  object 
 11  Denominator           49256 non-null  float64
 12  Score                 52179 non-null  float64
 13  Lower Estimate        52179 non-null  float64
 14  Higher Estimate       52179 non-null  float64
 15  Start Date            95820 non-null  object 
 16  End Date              95820 non-null  object 
dtypes: float64(4), int64(1), object(12)
memory usage: 12.4+ MB
```

In [85]:

```
sns.heatmap(comp.isnull(),yticklabels=False,cbar=False,cmap='Set2')
```

Out[85]:

```
<Axes: >
```



In [86]:
```
for c in colu:
  comp[c]=comp[c].fillna(comp[c].median())
```

In [87]:
```
comp.isnull().sum()
```

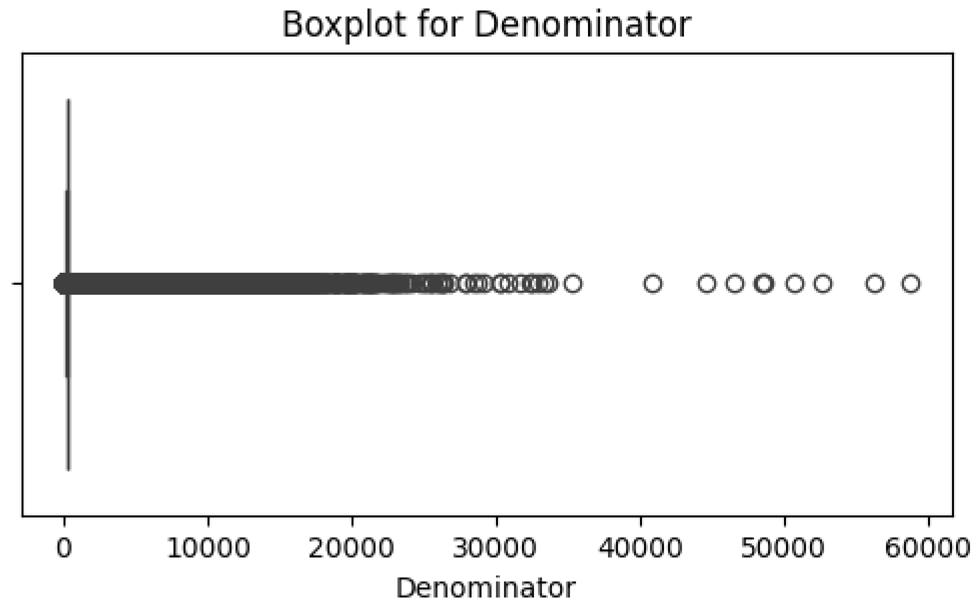
Out[87]:

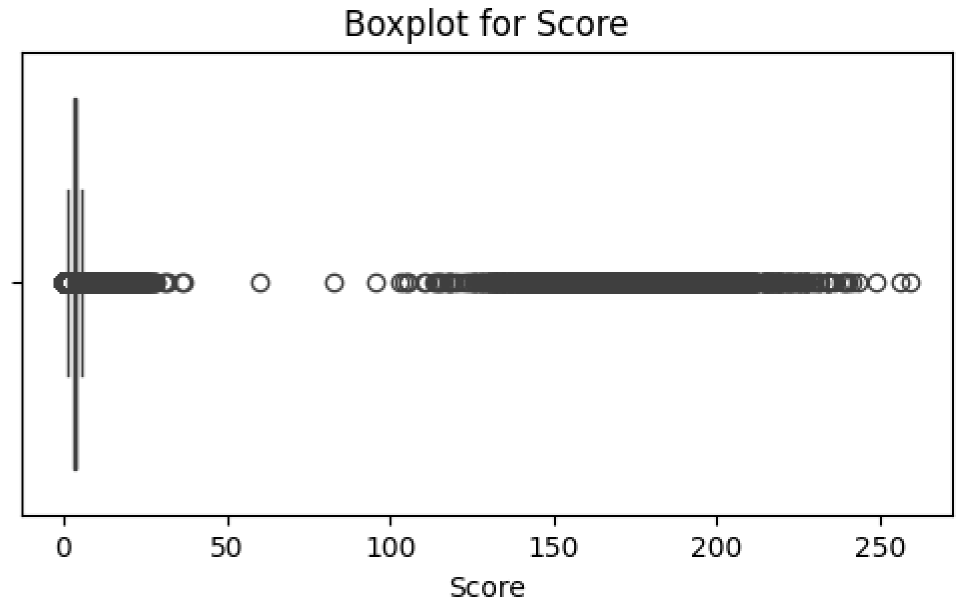
```
                         0
           Facility ID   0
         Facility Name   0
               Address   0
             City/Town   0
                 State   0
              ZIP Code   0
         County/Parish   0
      Telephone Number   0
            Measure ID   0
          Measure Name   0
  Compared to National   0
           Denominator   0
                 Score   0
        Lower Estimate   0
       Higher Estimate   0
            Start Date   0
              End Date   0

dtype: int64
```

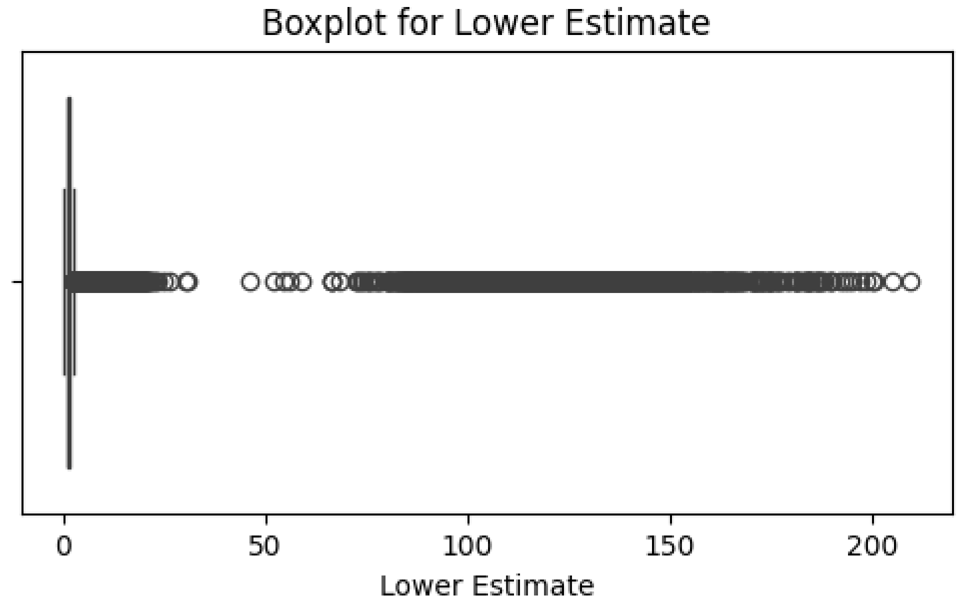
Outlier Detection:

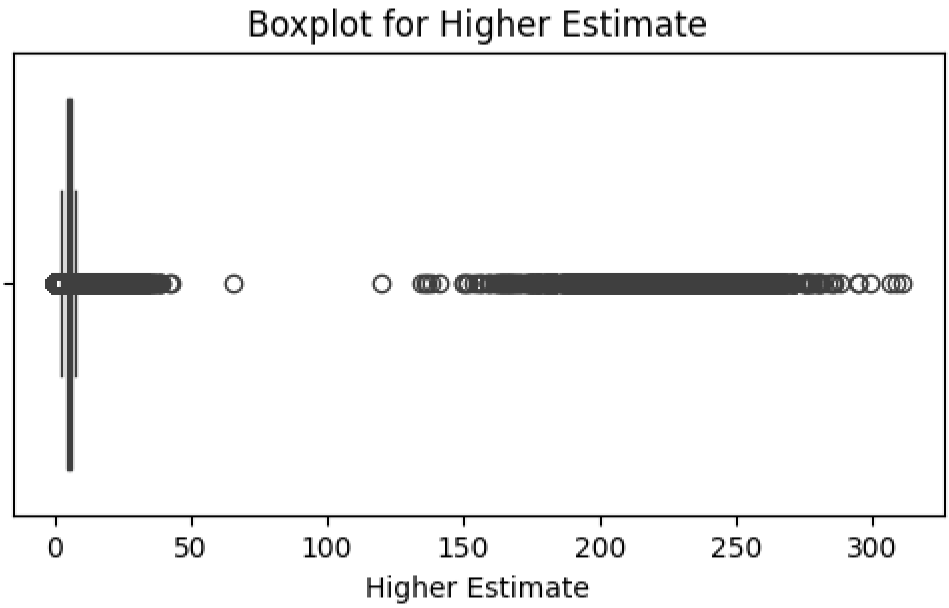
In [88]:

```python
for c in colu:
  plt.figure(figsize=(6,3))
  sns.boxplot(x=comp[c])
  plt.title(f"Boxplot for {c}")
  plt.show()
```


Boxplot for Denominator
0
10000
20000
30000
40000
50000
60000
Denominator


Boxplot for Score
0
50
100
150
200
250
Score


Boxplot for Lower Estimate
0
50
100
150
200
Lower Estimate



```
In [89]: ### Notes : Due to the presence of Hospital details in all the tables -> Going t
```

```
In [90]: hosp_dim = Hosp[['Facility ID','Facility Name','Address','City/Town',
                          'State','ZIP Code','County/Parish','Telephone Number']]
```

```
In [91]: hosp_dim.shape
```

```
Out[91]: (5421, 8)
```

```
In [92]: hosp_dim['Facility ID'].nunique()
```

```
Out[92]: 5421
```

```
In [122…  ### Observation: "Facility ID" doen't have any null values, So we can use it as
```

Normalization:

```
In [94]: drop_cols = ['Facility Name','Address','City/Town','State',
                      'ZIP Code','County/Parish','Telephone Number']
         for df in [hcahps,Hosp,comp]:
           df.drop(columns=drop_cols,inplace=True)
```

```
In [95]: hcahps.columns
```

```
Out[95]: Index(['Facility ID', 'HCAHPS Measure ID', 'HCAHPS Question',
                'HCAHPS Answer Description', 'Patient Survey Star Rating',
                'HCAHPS Answer Percent', 'HCAHPS Linear Mean Value',
                'Number of Completed Surveys', 'Survey Response Rate Percent',
                'Start Date', 'End Date'],
               dtype='object')
```

Insights and visualization:

```
In [96]: hcahps.head(2)
```

Out[96]:

	Facility ID	HCAHPS Measure ID	HCAHPS Question	HCAHPS Answer Description	Patient Survey Star Rating	HCAHPS Answer Percent	HCAHPS Linear Mean Value	Num Com Su
0	010001	H_COMP_1_A_P	Patients who reported that their nurses "Alway...	Nurses "always" communicated well	3.0	75.0	86.0	
1	010001	H_COMP_1_SN_P	Patients who reported that their nurses "Somet...	Nurses "sometimes" or "never" communicated well	3.0	6.0	86.0	

In [97]:

```
hcahps['HCAHPS Question'].nunique()
```

Out[97]:

```
93
```

In [123…

```
# hcahps[hcahps["Facility ID"] == "010001"]
```

In [99]:

```
hcahps.tail()
```

Out[99]:

	Facility ID	HCAHPS Measure ID	HCAHPS Question	HCAHPS Answer Description	Patient Survey Star Rating	HCAHPS Answer Percent
445558	671301	H_RECMND_DY	Patients who reported YES, they would definite...	"YES", patients would definitely recommend the...	3.0	22.0
445559	671301	H_RECMND_PY	Patients who reported YES, they would probably...	"YES", patients would probably recommend the h...	3.0	22.0
445560	671301	H_RECMND_LINEAR_SCORE	Recommend hospital - linear mean score	Recommend hospital - linear mean score	3.0	22.0
445561	671301	H_RECMND_STAR_RATING	Recommend hospital - star rating	Recommend hospital - star rating	3.0	22.0
445562	671301	H_STAR_RATING	Summary star rating	Summary star rating	3.0	22.0

In [100…

```
hcahps['HCAHPS Measure ID'].nunique()
```

Out[100…

```
93
```

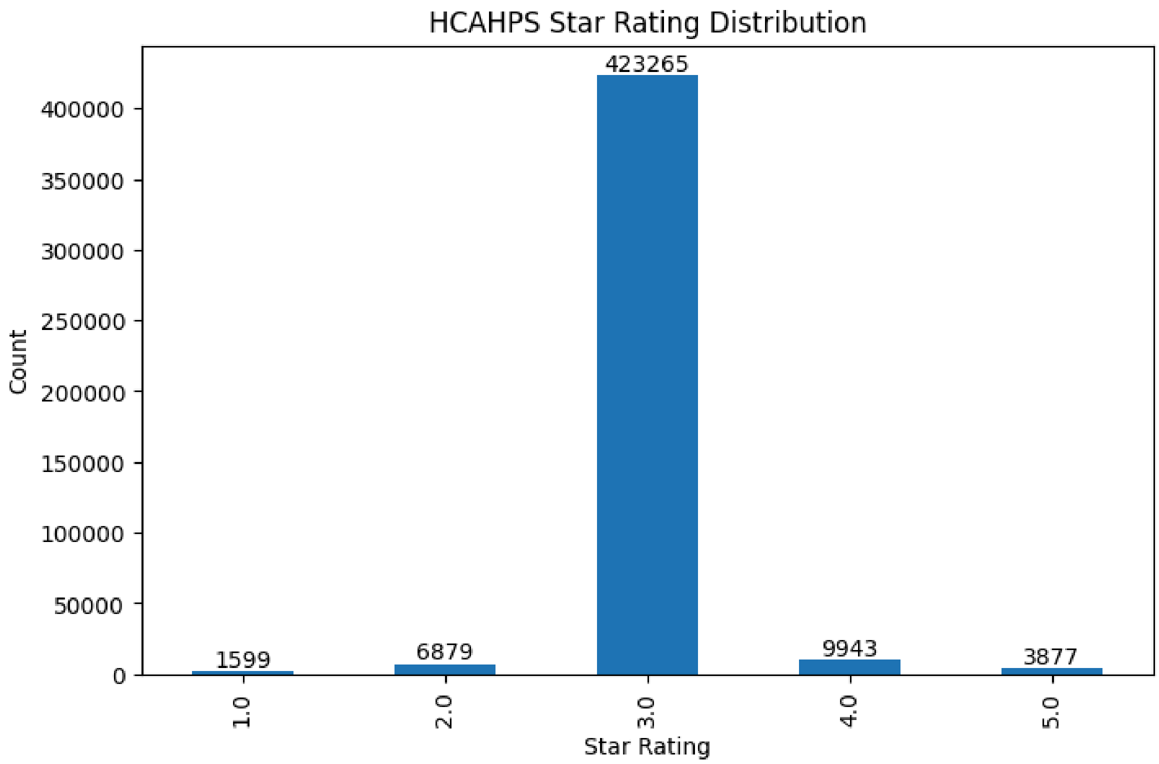
Observation:

From this we got to know that each 'Facility ID' had atmax of 93 Distinct Questions on survey.

HCAHPS Star Rating Distribution

In [101…

```
rating_counts = hcahps["Patient Survey Star Rating"].value_counts().sort_index()

plt.figure(figsize=(8,5))
ax= rating_counts.plot(kind='bar')
plt.title("HCAHPS Star Rating Distribution")
plt.xlabel("Star Rating")
plt.ylabel("Count")

for i in ax.containers:
    ax.bar_label(i)

plt.show()
```



In [102…

```
hcahps.describe()
```

Out[102…

	Patient Survey Star Rating	HCAHPS Answer Percent	HCAHPS Linear Mean Value	Number of Completed Surveys	Survey Response Rate Percent
count	445563.000000	445563.000000	445563.000000	445563.000000	445563.000000
mean	3.017102	29.727792	85.948571	509.956481	22.002713
std	0.294317	23.413649	1.627458	635.465414	7.805713
min	1.000000	0.000000	53.000000	25.000000	4.000000
25%	3.000000	18.000000	86.000000	169.000000	17.000000
50%	3.000000	22.000000	86.000000	376.500000	21.000000
75%	3.000000	33.000000	86.000000	548.000000	25.000000
max	5.000000	100.000000	100.000000	12627.000000	71.000000

In [103…

```
hosp_dim.head()
```

Out[103…

	Facility ID	Facility Name	Address	City/Town	State	ZIP Code	County/Parish	Telephone Number
0	010001	SOUTHEAST HEALTH MEDICAL CENTER	1108 ROSS CLARK CIRCLE	DOTHAN	AL	36301	HOUSTON	(334) 793-8701
1	010005	MARSHALL MEDICAL CENTERS	2505 U S HIGHWAY 431 NORTH	BOAZ	AL	35957	MARSHALL	(256) 593-8310
2	010006	NORTH ALABAMA MEDICAL CENTER	1701 VETERANS DRIVE	FLORENCE	AL	35630	LAUDERDALE	(256) 768-8400
3	010007	MIZELL MEMORIAL HOSPITAL	702 N MAIN ST	OPP	AL	36467	COVINGTON	(334) 493-3541
4	010008	CRENSHAW COMMUNITY HOSPITAL	101 HOSPITAL CIRCLE	LUVERNE	AL	36049	CRENSHAW	(334) 335-3374

In [104…

```
hosp_dim['State'].nunique()
```

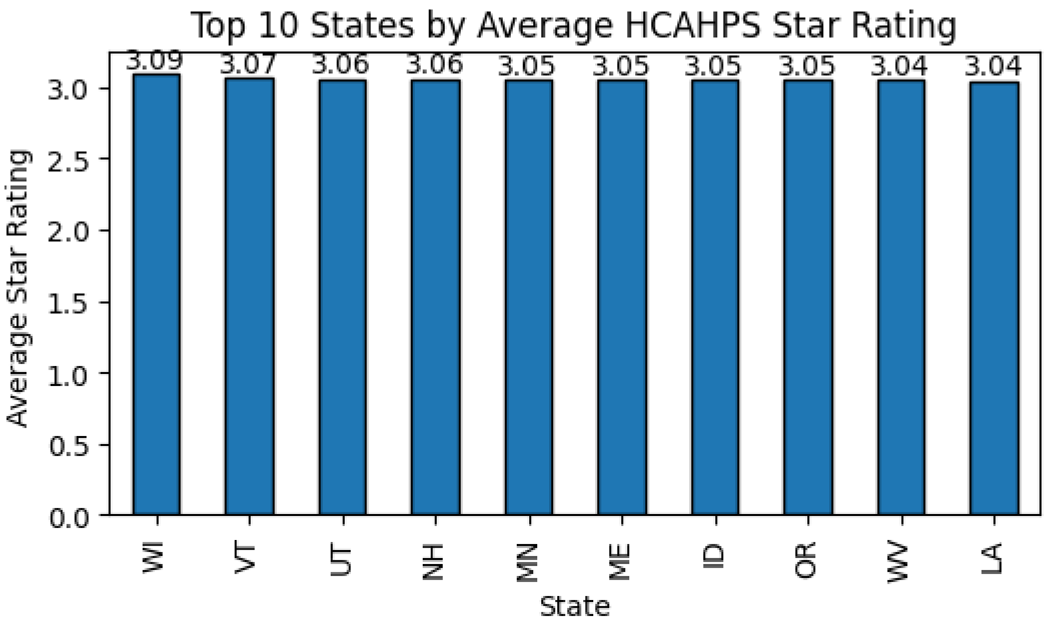
Out[104…

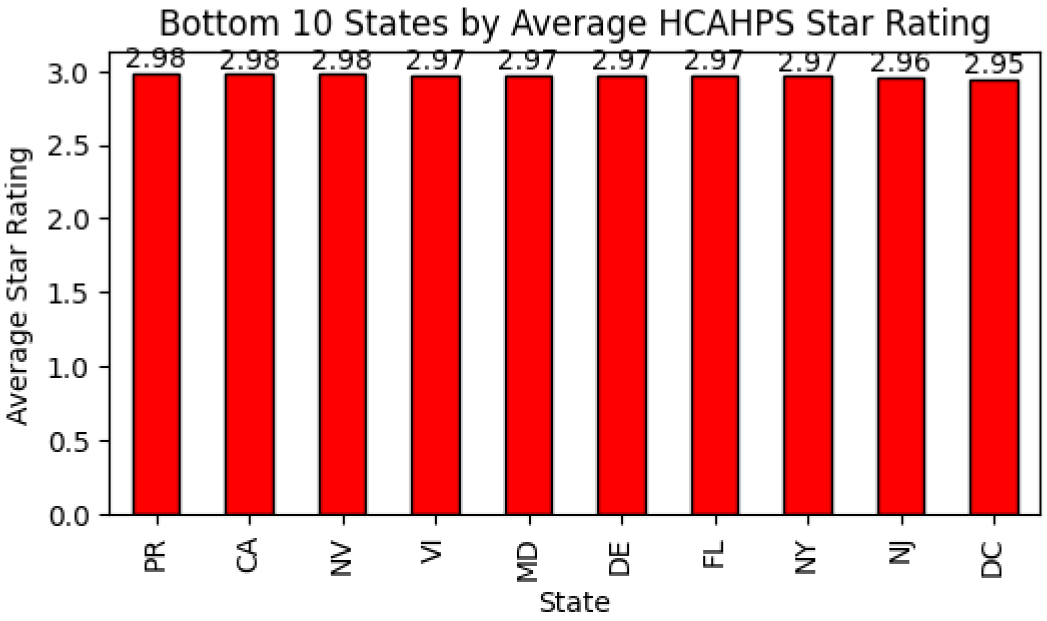
```
56
```

Top 10 & Bottom 10 States by rating

In [105…

```
merged = hcahps.merge(hosp_dim[["Facility ID", "State"]], on="Facility ID", how=
# Calculate state average rating
state_avg = merged.groupby("State")["Patient Survey Star Rating"].mean().sort_va

# Split into Top 10 & Bottom 10
top_10 = state_avg.head(10)
bottom_10 = state_avg.tail(10)

# ---- Plot Top 10 -----
plt.figure(figsize=(6,3))
ax = top_10.plot(kind='bar', edgecolor='black')
plt.title("Top 10 States by Average HCAHPS Star Rating")
plt.xlabel("State")
plt.ylabel("Average Star Rating")

for container in ax.containers:
    ax.bar_label(container,fmt='%.2f')

plt.show()

# ---- Plot Bottom 10 -----
plt.figure(figsize=(6,3))
ax2 = bottom_10.plot(kind='bar', edgecolor='black', color='red')
plt.title("Bottom 10 States by Average HCAHPS Star Rating")
plt.xlabel("State")
plt.ylabel("Average Star Rating")

for container in ax2.containers:
    ax2.bar_label(container,fmt='%.2f')

plt.show()
```

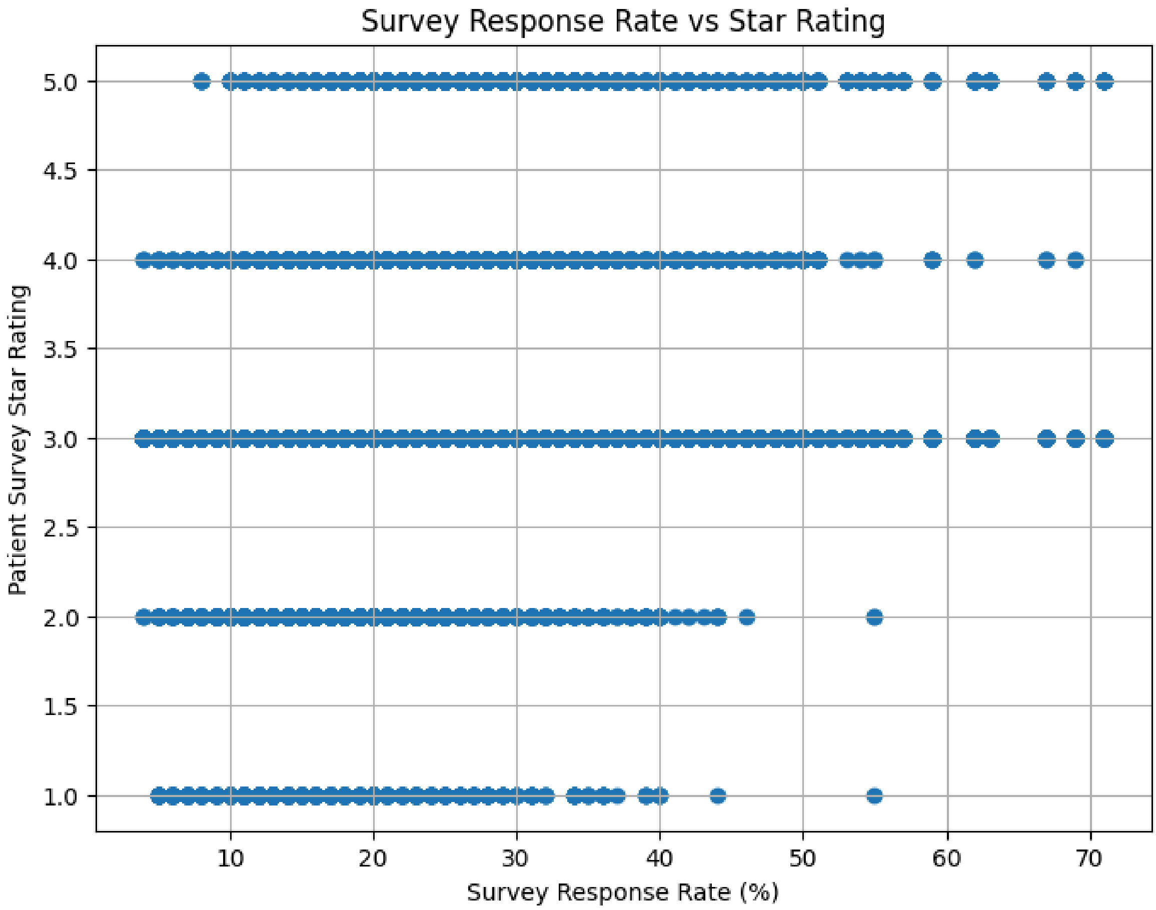




Survey Response Rate vs Star Rating

In [106…

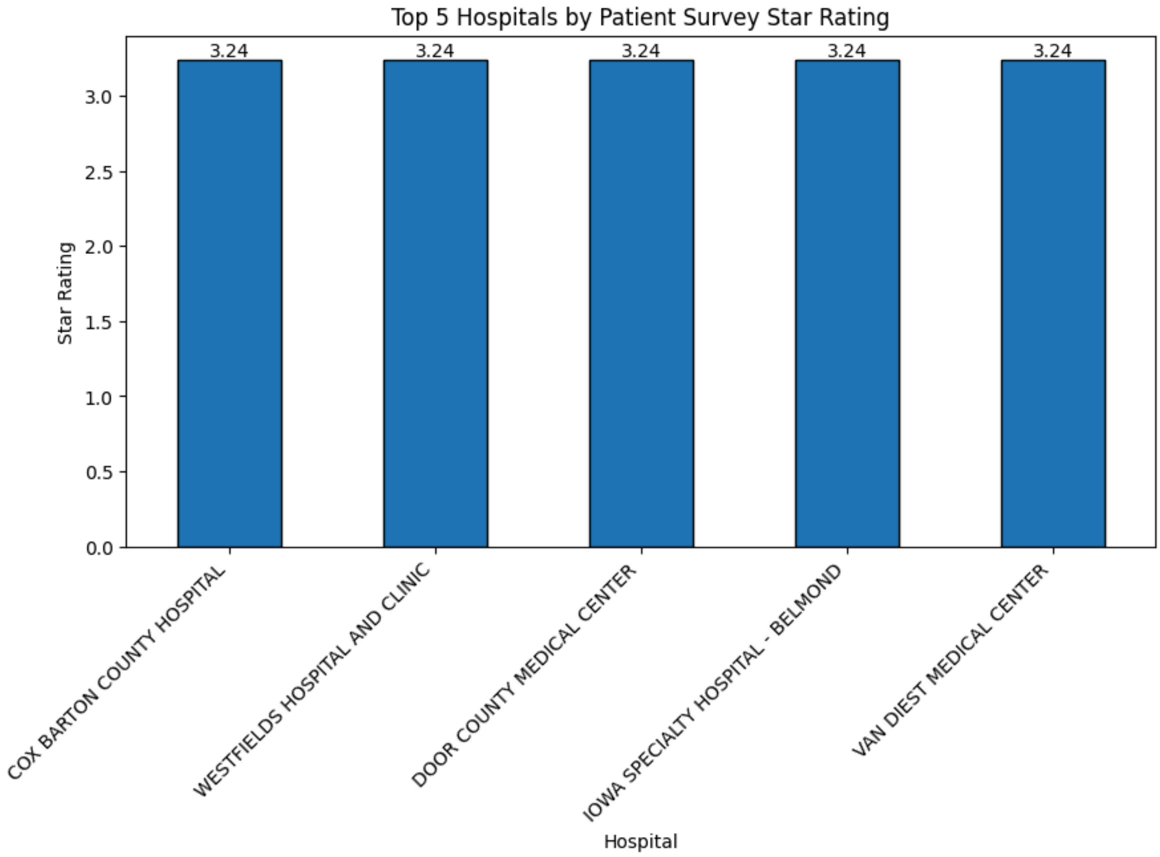
```
plt.figure(figsize=(8,6))

plt.scatter(
    merged["Survey Response Rate Percent"],
    merged["Patient Survey Star Rating"]
)

plt.title("Survey Response Rate vs Star Rating")
plt.xlabel("Survey Response Rate (%)")
plt.ylabel("Patient Survey Star Rating")
plt.grid(True)
plt.show()
```



Top 5 Hospitals by Rating

In [107…

```
merged1 = hcahps.merge(hosp_dim[["Facility ID", "Facility Name"]], on="Facility
# Top 5 hospitals by star rating
Hos_avg = merged1.groupby("Facility Name")["Patient Survey Star Rating"].mean().

plt.figure(figsize=(10,5))
ax = Hos_avg.plot(
    kind='bar',
    edgecolor='black')
plt.title("Top 5 Hospitals by Patient Survey Star Rating")
plt.xlabel("Hospital")
plt.ylabel("Star Rating")
plt.xticks(rotation=45, ha='right')

for con in ax.containers:
    ax.bar_label(con, fmt="%.2f")

plt.show()
```



```
In [108…  Hosp.head()
```

Out[108…

	Facility ID	Hospital Type	Hospital Ownership	Emergency Services	Meets criteria for birthing friendly designation	Hospital overall rating	MORT Group Measure Count	Count c Facilit MOR Measure
0	010001	Acute Care Hospitals	Government - Hospital District or Authority	Yes	Y	4.0	7.0	7
1	010005	Acute Care Hospitals	Government - Hospital District or Authority	Yes	Y	3.0	7.0	6
2	010006	Acute Care Hospitals	Proprietary	Yes	Y	2.0	7.0	7
3	010007	Acute Care Hospitals	Voluntary non-profit - Private	Yes	NaN	1.0	7.0	3
4	010008	Acute Care Hospitals	Proprietary	Yes	NaN	3.0	7.0	1

```
In [109…  comp.head()
```

Out[109…

	Facility ID	Measure ID	Measure Name	Compared to National	Denominator	Score	Lower Estimate	E
0	010001	COMP_HIP_KNEE	Rate of complications for hip/knee replacement…	No Different Than the National Rate	27.0	3.2	1.7	
1	010001	Hybrid_HWM	Hybrid Hospital-Wide All-Cause Risk Standardiz…	No Different Than the National Rate	1835.0	4.5	2.6	
2	010001	MORT_30_AMI	Death rate for heart attack patients	No Different Than the National Rate	270.0	11.4	9.1	
3	010001	MORT_30_CABG	Death rate for CABG surgery patients	No Different Than the National Rate	144.0	3.0	1.6	
4	010001	MORT_30_COPD	Death rate for COPD patients	No Different Than the National Rate	112.0	9.4	6.4	

In [110…

```
comp.shape
```

Out[110…

```
(95820, 10)
```

In [111…

```
comp[['Measure ID','Measure Name']].nunique()
```

Out[111…

	0
Measure ID	20
Measure Name	20

dtype: int64

In [112…

```
## Helps to identify the reasons behind the deaths in different hospitals
comp[['Measure ID','Measure Name']].value_counts()
```

Out[112…

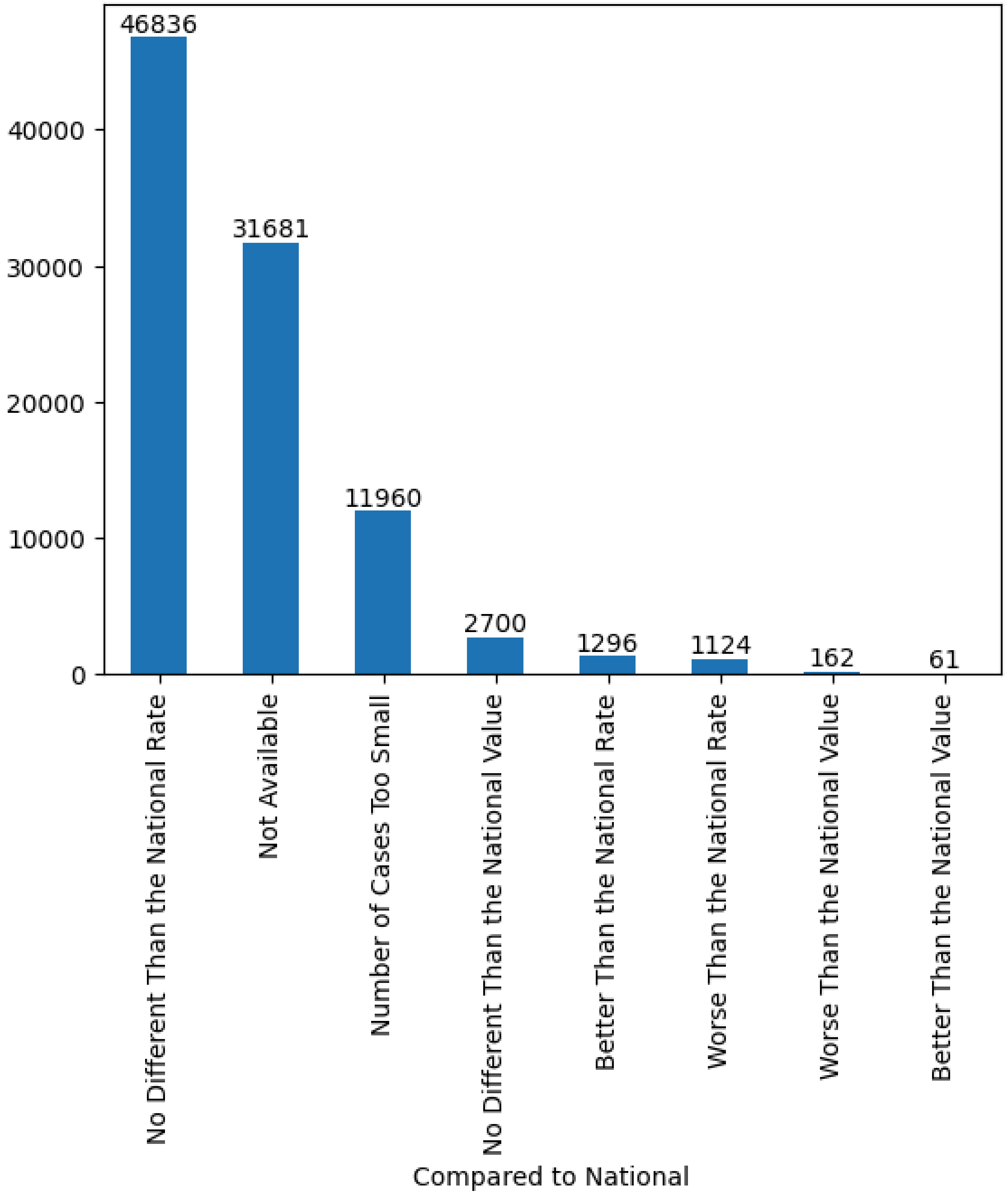
		count
Measure ID	**Measure Name**	
COMP_HIP_KNEE	**Rate of complications for hip/knee replacement patients**	4791
Hybrid_HWM	**Hybrid Hospital-Wide All-Cause Risk Standardized Mortality Rate**	4791
MORT_30_AMI	**Death rate for heart attack patients**	4791
MORT_30_CABG	**Death rate for CABG surgery patients**	4791
MORT_30_COPD	**Death rate for COPD patients**	4791
MORT_30_HF	**Death rate for heart failure patients**	4791
MORT_30_PN	**Death rate for pneumonia patients**	4791
MORT_30_STK	**Death rate for stroke patients**	4791
PSI_03	**Pressure ulcer rate**	4791
PSI_04	**Death rate among surgical inpatients with serious treatable complications**	4791
PSI_06	**Iatrogenic pneumothorax rate**	4791
PSI_08	**In-hospital fall-associated fracture rate**	4791
PSI_09	**Postoperative hemorrhage or hematoma rate**	4791
PSI_10	**Postoperative acute kidney injury requiring dialysis rate**	4791
PSI_11	**Postoperative respiratory failure rate**	4791
PSI_12	**Perioperative pulmonary embolism or deep vein thrombosis rate**	4791
PSI_13	**Postoperative sepsis rate**	4791
PSI_14	**Postoperative wound dehiscence rate**	4791
PSI_15	**Abdominopelvic accidental puncture or laceration rate**	4791
PSI_90	**CMS Medicare PSI 90: Patient safety and adverse events composite**	4791

dtype: int64

Distribution of No.Of Hospitals by comparison with National Rate

In [113…

```
bx= comp['Compared to National'].value_counts().plot(kind='bar')
for i in bx.containers:
    bx.bar_label(i)
plt.show()
```



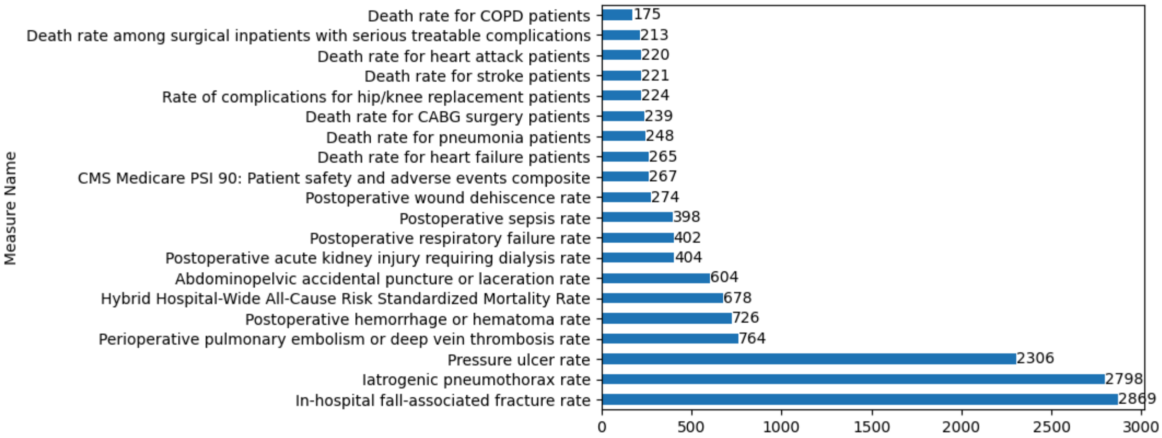
Average Patient sample Volume Across Complication Measures

In [114…

```
### Helps to understanding reliability and impact based on sample size used in p
```

In [115…

```
cx=comp.groupby('Measure Name')['Denominator'].mean().round().sort_values(ascend
for i in cx.containers:
    cx.bar_label(i)
plt.show()
```



Best_2 & Worst_2 Hospitals for each Complication Measures

In [116…

```
##Higher Score   -- More complications/deaths  → worse hospital
##Lower Score    -- Better performance         → safer hospital
```

In [117…

```
top2 = (comp.merge(hosp_dim, on='Facility ID')
            .sort_values(['Measure Name','Score'], ascending=[True, True])
            .groupby('Measure Name')
            .head(2))
top2['Rank'] = top2.groupby('Measure Name').cumcount() + 1

top2[['Measure Name','Rank','Facility ID','Facility Name','Score']]
```

Out[117…

	Measure Name	Rank	Facility ID	Facility Name	Score
57358	Abdominopelvic accidental puncture or lacerati...	1	330101	NEW YORK-PRESBYTERIAN HOSPITAL	0.35
3078	Abdominopelvic accidental puncture or lacerati...	2	030103	MAYO CLINIC HOSPITAL	0.36
57359	CMS Medicare PSI 90: Patient safety and advers...	1	330101	NEW YORK-PRESBYTERIAN HOSPITAL	0.40
95219	CMS Medicare PSI 90: Patient safety and advers...	2	670122	HOUSTON METHODIST THE WOODLANDS HOSPITAL	0.55
69	Death rate among surgical inpatients with seri...	1	010007	MIZELL MEMORIAL HOSPITAL	3.43
89	Death rate among surgical inpatients with seri...	2	010008	CRENSHAW COMMUNITY HOSPITAL	3.43
40143	Death rate for CABG surgery patients	1	220071	MASSACHUSETTS GENERAL HOSPITAL	1.00
64663	Death rate for CABG surgery patients	2	360180	CLEVELAND CLINIC	1.00
84	Death rate for COPD patients	1	010008	CRENSHAW COMMUNITY HOSPITAL	3.43
164	Death rate for COPD patients	2	010018	CALLAHAN EYE HOSPITAL	3.43
22	Death rate for heart attack patients	1	010005	MARSHALL MEDICAL CENTERS	3.43
62	Death rate for heart attack patients	2	010007	MIZELL MEMORIAL HOSPITAL	3.43
85	Death rate for heart failure patients	1	010008	CRENSHAW COMMUNITY HOSPITAL	3.43
165	Death rate for heart failure patients	2	010018	CALLAHAN EYE HOSPITAL	3.43
166	Death rate for pneumonia patients	1	010018	CALLAHAN EYE HOSPITAL	3.43
226	Death rate for pneumonia patients	2	010022	FLOYD CHEROKEE MEDICAL CENTER	3.43
67	Death rate for stroke patients	1	010007	MIZELL MEMORIAL HOSPITAL	3.43
87	Death rate for stroke patients	2	010008	CRENSHAW COMMUNITY HOSPITAL	3.43

	Measure Name	Rank	Facility ID	Facility Name	Score
58441	Hybrid Hospital-Wide All-Cause Risk Standardiz...	1	330214	NYU LANGONE HOSPITALS	1.80
40141	Hybrid Hospital-Wide All-Cause Risk Standardiz...	2	220071	MASSACHUSETTS GENERAL HOSPITAL	2.00
56750	Iatrogenic pneumothorax rate	1	330024	MOUNT SINAI HOSPITAL	0.08
70290	Iatrogenic pneumothorax rate	2	390049	ST LUKE'S HOSPITAL BETHLEHEM	0.08
57351	In-hospital fall-associated fracture rate	1	330101	NEW YORK-PRESBYTERIAN HOSPITAL	0.14
71471	In-hospital fall-associated fracture rate	2	390174	THOMAS JEFFERSON UNIVERSITY HOSPITAL	0.15
79095	Perioperative pulmonary embolism or deep vein ...	1	450040	COVENANT MEDICAL CENTER	1.31
57355	Perioperative pulmonary embolism or deep vein ...	2	330101	NEW YORK-PRESBYTERIAN HOSPITAL	1.48
57353	Postoperative acute kidney injury requiring di...	1	330101	NEW YORK-PRESBYTERIAN HOSPITAL	0.48
25113	Postoperative acute kidney injury requiring di...	2	140281	NORTHWESTERN MEMORIAL HOSPITAL	0.65
57352	Postoperative hemorrhage or hematoma rate	1	330101	NEW YORK-PRESBYTERIAN HOSPITAL	0.94
80512	Postoperative hemorrhage or hematoma rate	2	450358	HOUSTON METHODIST HOSPITAL	1.12
25114	Postoperative respiratory failure rate	1	140281	NORTHWESTERN MEMORIAL HOSPITAL	1.71
14994	Postoperative respiratory failure rate	2	090011	MEDSTAR WASHINGTON HOSPITAL CENTER	1.94
25116	Postoperative sepsis rate	1	140281	NORTHWESTERN MEMORIAL HOSPITAL	1.15
57356	Postoperative sepsis rate	2	330101	NEW YORK-PRESBYTERIAN HOSPITAL	1.66
57357	Postoperative wound dehiscence rate	1	330101	NEW YORK-PRESBYTERIAN HOSPITAL	0.84
31357	Postoperative wound dehiscence rate	2	170040	UNIVERSITY OF KANSAS HOSPITAL	1.16
51428	Pressure ulcer rate	1	280013	THE NEBRASKA MEDICAL CENTER	0.05

	Measure Name	Rank	Facility ID	Facility Name	Score
59868	Pressure ulcer rate	2	340002	MEMORIAL MISSION HOSPITAL AND ASHEVILLE SURGER...	0.05
8180	Rate of complications for hip/knee replacement...	1	050290	SAINT JOHN'S HEALTH CENTER	1.40
8660	Rate of complications for hip/knee replacement...	2	050351	TORRANCE MEMORIAL MEDICAL CENTER	1.60

In [118…

```
Bot2 = (comp.merge(hosp_dim, on='Facility ID')
            .sort_values(['Measure Name','Score'], ascending=[True, False])
            .groupby('Measure Name')
            .head(2))
Bot2['Rank'] = Bot2.groupby('Measure Name').cumcount() + 1

Bot2[['Measure Name','Rank','Facility ID','Facility Name','Score']]
```

Out[118…

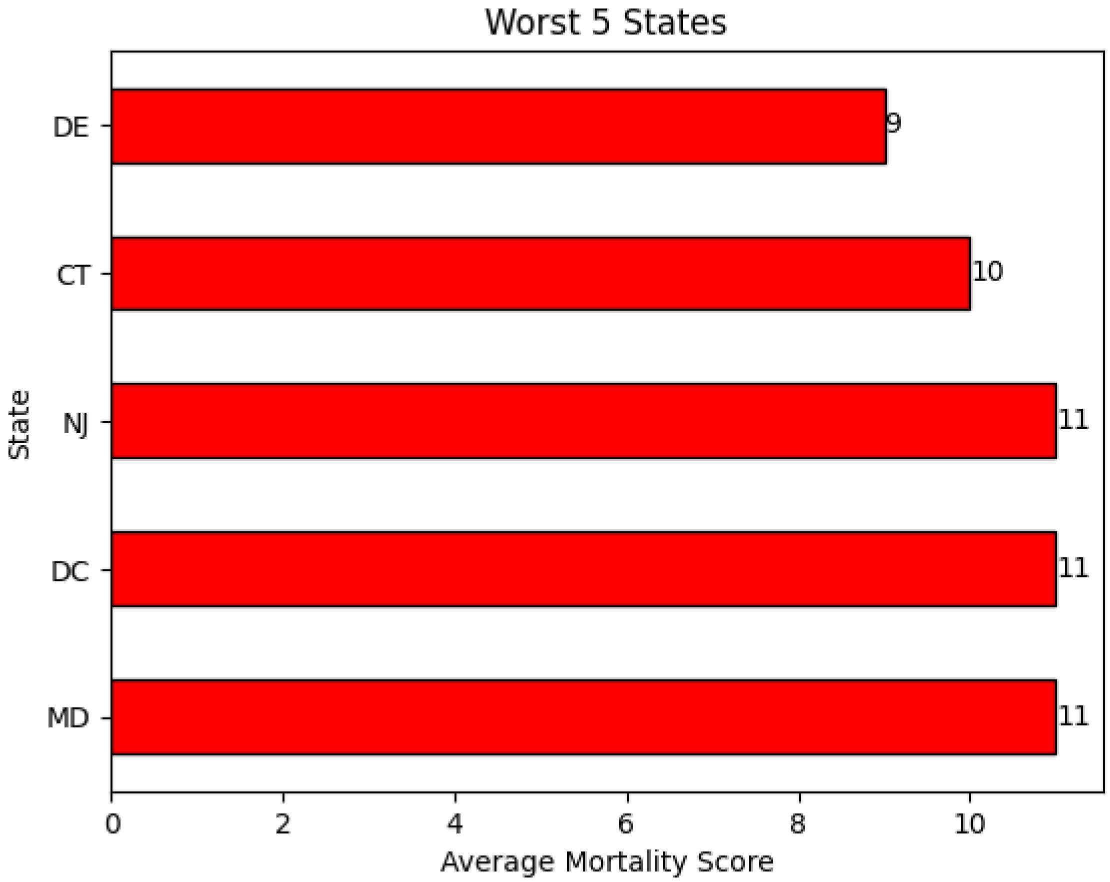
	Measure Name	Rank	Facility ID	Facility Name	Score
98	Abdominopelvic accidental puncture or lacerati…	1	010008	CRENSHAW COMMUNITY HOSPITAL	3.43
178	Abdominopelvic accidental puncture or lacerati…	2	010018	CALLAHAN EYE HOSPITAL	3.43
99	CMS Medicare PSI 90: Patient safety and advers…	1	010008	CRENSHAW COMMUNITY HOSPITAL	3.43
239	CMS Medicare PSI 90: Patient safety and advers…	2	010022	FLOYD CHEROKEE MEDICAL CENTER	3.43
70589	Death rate among surgical inpatients with seri…	1	390073	UPMC ALTOONA	259.01
81849	Death rate among surgical inpatients with seri…	2	450686	UNIVERSITY MEDICAL CENTER	255.81
80183	Death rate for CABG surgery patients	1	450231	BSA HOSPITAL	6.50
31183	Death rate for CABG surgery patients	2	170013	HAYS MEDICAL CENTER	6.20
1884	Death rate for COPD patients	1	020006	MAT-SU REGIONAL MEDICAL CENTER	16.50
69724	Death rate for COPD patients	2	381325	GOOD SHEPHERD MEDICAL CENTER	16.50
382	Death rate for heart attack patients	1	010039	HUNTSVILLE HOSPITAL	17.10
79662	Death rate for heart attack patients	2	450133	MIDLAND MEMORIAL HOSPITAL	16.70
92005	Death rate for heart failure patients	1	520116	WATERTOWN MEMORIAL HOSPITAL	20.20
51485	Death rate for heart failure patients	2	280032	MARY LANNING HEALTHCARE	19.80
46686	Death rate for pneumonia patients	1	250082	DELTA HEALTH SYSTEM - THE MEDICAL CENTER	32.20
46326	Death rate for pneumonia patients	2	250036	GEORGE REGIONAL HEALTH SYSTEM	30.40
68827	Death rate for stroke patients	1	380027	MERCY MEDICAL CENTER	22.10
77447	Death rate for stroke patients	2	440073	MAURY REGIONAL HOSPITAL	21.40
73741	Hybrid Hospital-Wide All-Cause Risk Standardiz…	1	400117	HOSPITAL METROPOLITANO DR SUSONI	7.80

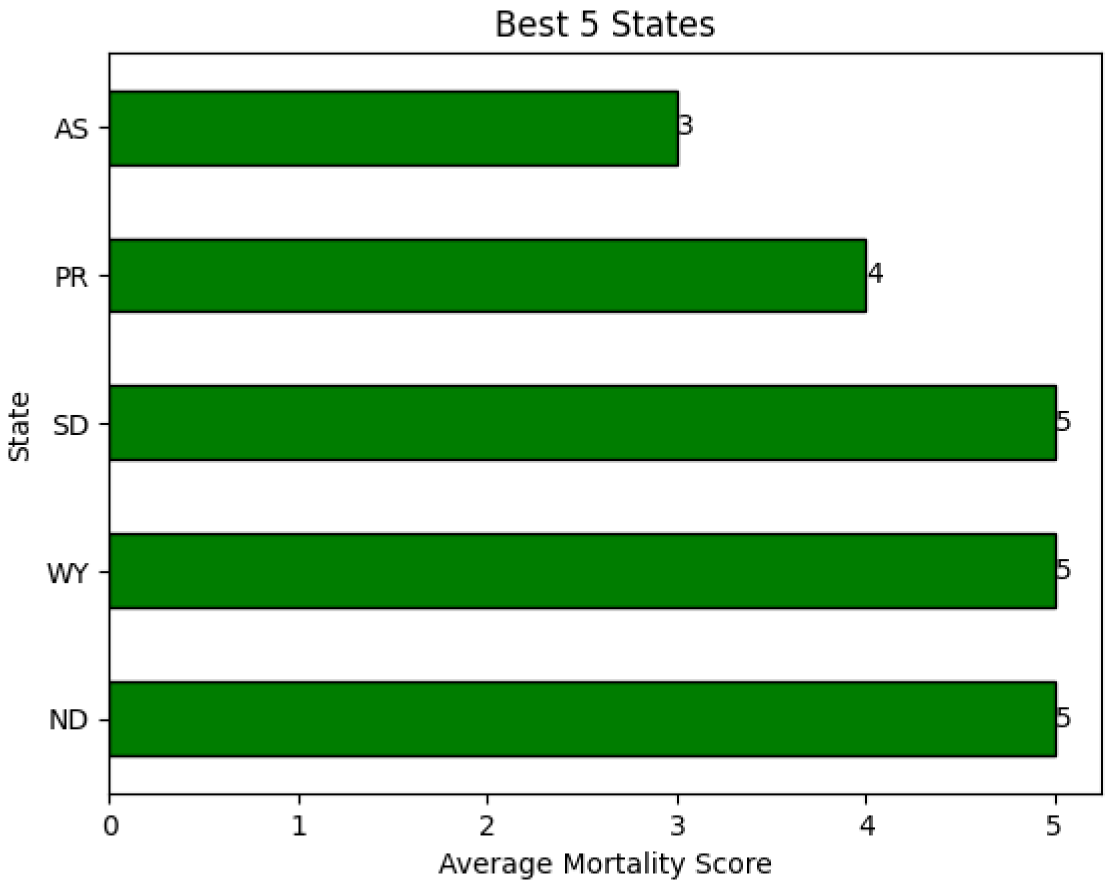
	Measure Name	Rank	Facility ID	Facility Name	Score
73541	Hybrid Hospital-Wide All-Cause Risk Standardiz...	2	400104	HOSPITAL MENONITA CAGUAS INC	7.20
1090	Iatrogenic pneumothorax rate	1	010110	BULLOCK COUNTY HOSPITAL	3.43
1610	Iatrogenic pneumothorax rate	2	01021F	TUSCALOOSA VA MEDICAL CENTER	3.43
1091	In-hospital fall-associated fracture rate	1	010110	BULLOCK COUNTY HOSPITAL	3.43
1611	In-hospital fall-associated fracture rate	2	01021F	TUSCALOOSA VA MEDICAL CENTER	3.43
88875	Perioperative pulmonary embolism or deep vein ...	1	500064	HARBORVIEW MEDICAL CENTER	8.14
975	Perioperative pulmonary embolism or deep vein ...	2	010100	THOMAS HOSPITAL	8.11
2333	Postoperative acute kidney injury requiring di...	1	030002	BANNER - UNIVERSITY MEDICAL CENTER PHOENIX	3.98
2433	Postoperative acute kidney injury requiring di...	2	030012	YAVAPAI REGIONAL MEDICAL CENTER	3.97
51512	Postoperative hemorrhage or hematoma rate	1	280040	THE NEBRASKA METHODIST HOSPITAL	5.52
40212	Postoperative hemorrhage or hematoma rate	2	220075	MASSACHUSETTS EYE AND EAR INFIRMARY -	4.95
17154	Postoperative respiratory failure rate	1	100212	MARION COMMUNTIY HOSPITAL	59.92
27174	Postoperative respiratory failure rate	2	150058	MEMORIAL HOSPITAL OF SOUTH BEND	36.65
19596	Postoperative sepsis rate	1	110064	PIEDMONT COLUMBUS REGIONAL MIDTOWN	12.27
7836	Postoperative sepsis rate	2	050242	DOMINICAN HOSPITAL	11.66
74417	Postoperative wound dehiscence rate	1	420007	SPARTANBURG MEDICAL CENTER	3.54
8577	Postoperative wound dehiscence rate	2	050335	ADVENTIST HEALTH SONORA	3.50
14808	Pressure ulcer rate	1	080007	BEEBE MEDICAL CENTER	8.31
4248	Pressure ulcer rate	2	040020	ST BERNARDS MEDICAL CENTER	8.19
18980	Rate of complications for hip/knee replacement...	1	110006	ST MARY'S HOSPITAL	9.30

	Measure Name	Rank	Facility ID	Facility Name	Score
7180	Rate of complications for hip/knee replacement...	2	050150	SIERRA NEVADA MEMORIAL HOSPITAL	6.80

Complication & Mortality Rates by State

```
In [119…
Mrg = comp.merge(hosp_dim, on='Facility ID')
dx=Mrg.groupby('State')['Score'].mean().round().sort_values(ascending=False)

wrst= dx.head(5).plot(kind='barh',edgecolor='black',color='red')
plt.title('Worst 5 States')
plt.xlabel('Average Mortality Score')
plt.ylabel('State')

for i in wrst.containers:
    wrst.bar_label(i)
plt.show()

#--- Best 5 states

bst = dx.tail(5).plot(kind='barh',edgecolor='black',color='green')
plt.title('Best 5 States')
plt.xlabel('Average Mortality Score')
plt.ylabel('State')

for i in bst.containers:
    bst.bar_label(i)
plt.show()
```




Best 5 States
AS
PR
SD
WY
ND
State
3
4
5
5
5
0
1
2
3
4
5
Average Mortality Score

```
In [ ]:
```