```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import os
import seaborn as sns
os.environ['KAGGLE_CONFIG_DIR']='/content'
```

```python
!kaggle datasets download -d thedevastator/hospitals-in-the-united-states-a-comprehensive-d
```

```
Traceback (most recent call last):
  File "/usr/local/bin/kaggle", line 5, in <module>
    from kaggle.cli import main
  File "/usr/local/lib/python3.9/dist-packages/kaggle/__init__.py", line 23, in <module>
    api.authenticate()
  File "/usr/local/lib/python3.9/dist-packages/kaggle/api/kaggle_api_extended.py", line 164, in authenticate
    raise IOError('Could not find {}. Make sure it\'s located in'
OSError: Could not find kaggle.json. Make sure it's located in /content. Or use the environment method.
```

```python
df=pd.read_csv('Hospital_General_Information.csv')
```

```python
df.shape
```

```
(4818, 30)
```

```python
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4818 entries, 0 to 4817
Data columns (total 30 columns):
 #   Column                                                          Non-Null Count  Dtype
---  ------                                                          --------------  -----
 0   index                                                           4818 non-null   int64
 1   Provider ID                                                     4818 non-null   int64
 2   Hospital Name                                                   4818 non-null   object
 3   Address                                                         4818 non-null   object
 4   City                                                            4818 non-null   object
 5   State                                                           4818 non-null   object
 6   ZIP Code                                                        4818 non-null   int64
 7   County Name                                                     4803 non-null   object
 8   Phone Number                                                    4818 non-null   int64
 9   Hospital Type                                                   4818 non-null   object
 10  Hospital Ownership                                              4818 non-null   object
 11  Emergency Services                                              4818 non-null   bool
 12  Meets criteria for meaningful use of EHRs                       0 non-null      float64
 13  Hospital overall rating                                         3648 non-null   float64
 14  Hospital overall rating footnote                                1342 non-null   object
 15  Mortality national comparison                                   4818 non-null   object
 16  Mortality national comparison footnote                          1337 non-null   object
 17  Safety of care national comparison                              4818 non-null   object
 18  Safety of care national comparison footnote                     1892 non-null   object
 19  Readmission national comparison                                 4818 non-null   object
 20  Readmission national comparison footnote                        1002 non-null   object
 21  Patient experience national comparison                          4818 non-null   object
 22  Patient experience national comparison footnote                 1355 non-null   object
 23  Effectiveness of care national comparison                       4818 non-null   object
 24  Effectiveness of care national comparison footnote              1116 non-null   object
 25  Timeliness of care national comparison                          4818 non-null   object
 26  Timeliness of care national comparison footnote                 1331 non-null   object
 27  Efficient use of medical imaging national comparison            4818 non-null   object
 28  Efficient use of medical imaging national comparison footnote   2026 non-null   object
 29  Location                                                        4818 non-null   object
dtypes: bool(1), float64(2), int64(4), object(23)
memory usage: 1.1+ MB
```
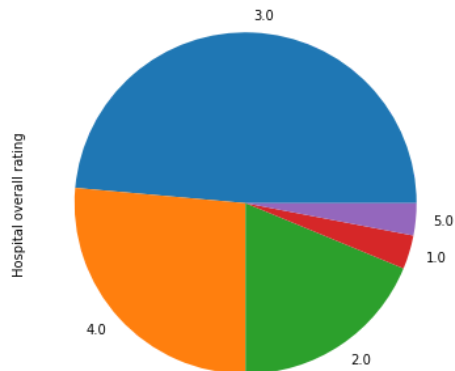
```python
df.describe()
```

|       | index        | Provider ID    | ZIP Code      | Phone Number | Meets criteria for meaningful use of EHRs | Hospital overall rating |
|-------|--------------|----------------|---------------|--------------|-------------------------------------------|-------------------------|
| count | 4818.000000  | 4818.000000    | 4818.000000   | 4.818000e+03 | 0.0                                       | 3648.000000             |
| mean  | 2408.500000  | 267983.546492  | 53956.188667  | 5.849487e+09 | NaN                                       | 3.073465                |
| std   | 1390.981129  | 156646.082755  | 26915.868321  | 2.329768e+09 | NaN                                       | 0.834595                |
| min   | 0.000000     | 10001.000000   | 603.000000    | 9.369338e+08 | NaN                                       | 1.000000                |
| 25%   | 1204.250000  | 140185.250000  | 33013.750000  | 4.017728e+09 | NaN                                       | 3.000000                |
| 50%   | 2408.500000  | 260037.000000  | 55395.500000  | 6.053298e+09 | NaN                                       | 3.000000                |
| 75%   | 3612.750000  | 390209.250000  | 76019.250000  | 7.877564e+09 | NaN                                       | 4.000000                |

```
df['Hospital overall rating'].value_counts()
```

```
    3.0    1772
    4.0     964
    2.0     684
    1.0     117
    5.0     111
    Name: Hospital overall rating, dtype: int64
```

```
a=df['Hospital overall rating'].value_counts().plot(kind='pie', figsize=(6, 6))
```



```
df['Mortality national comparison'].value_counts()
```
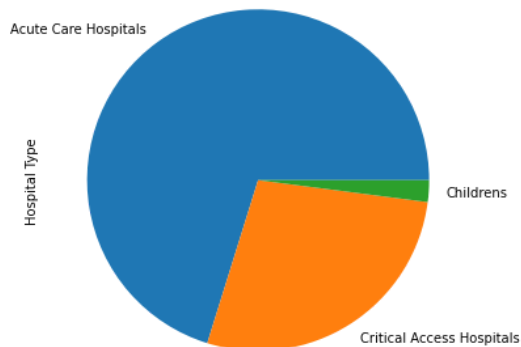
```
    Same as the National average    2736
    Not Available                   1337
    Above the National average       402
    Below the National average       343
    Name: Mortality national comparison, dtype: int64
```

```
df['Hospital Type'].value_counts()
```

```
    Acute Care Hospitals        3382
    Critical Access Hospitals   1337
    Childrens                     99
    Name: Hospital Type, dtype: int64
```

```
df['Hospital Type'].value_counts().plot(kind='pie', figsize=(6, 6))
```

```
    <Axes: ylabel='Hospital Type'>
```



```
df['Meets criteria for meaningful use of EHRs'].value_counts()
```
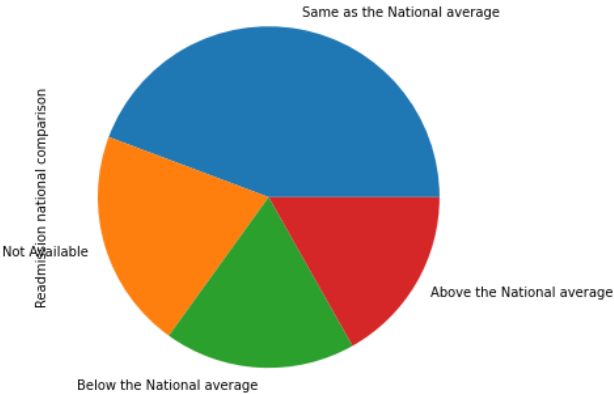
```
    Series([], Name: Meets criteria for meaningful use of EHRs, dtype: int64)
```

```
df['Readmission national comparison'].value_counts()
```

```
        Same as the National average     2133
        Not Available                    1002
        Below the National average        872
        Above the National average        811
        Name: Readmission national comparison, dtype: int64
```

```python
df['Readmission national comparison'].value_counts().plot(kind='pie', figsize=(6, 6))
```

```
<Axes: ylabel='Readmission national comparison'>
```



```python
df['City'].value_counts()
```

```
        CHICAGO          29
        HOUSTON          26
        LOS ANGELES      20
        DALLAS           19
        PHILADELPHIA     19
                         ..
        BARBOURVILLE      1
        CROWLEY           1
        MOUNT CARMEL      1
        ALBIA             1
        ELLENSBURG        1
        Name: City, Length: 2949, dtype: int64
```
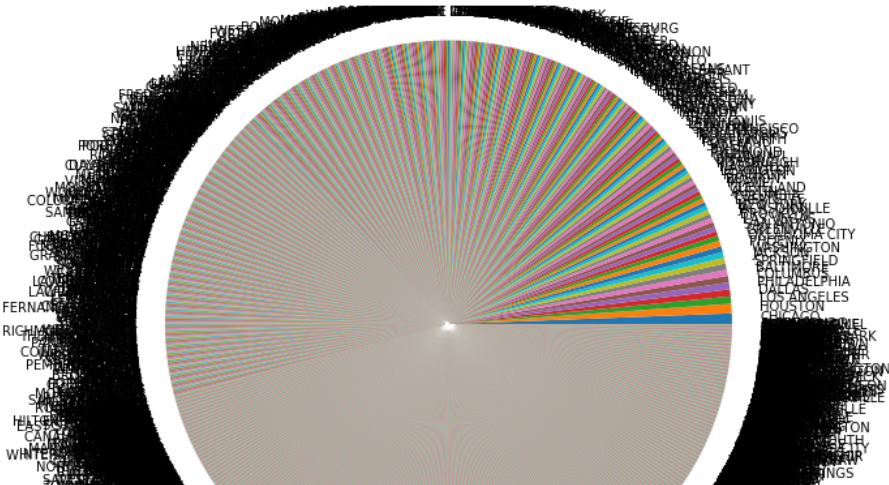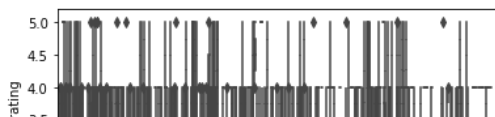
```python
df['City'].value_counts().plot(kind='pie', figsize=(10, 10))
```

```
<Axes: ylabel='City'>
```



```python
sns.boxplot( x='City', y='Hospital overall rating', data=df, )
plt.show()
```

```python
print('The Dataframe:\n',df)
print('\n'+'-'*250+'\n')
print(df.shape)
print('\n'+'-'*250+'\n')
print(df.head())
print('\n'+'-'*250+'\n')
print(df.describe())
print('\n'+'-'*250+'\n')
print(df.info())
print('\n'+'-'*250+'\n')

print(df.sort_values(by=['State'],ascending=True))
print('\n'+'-'*250+'\n')
print(df.sort_values(by=['Hospital overall rating'],ascending=True))
print('\n'+'-'*250+'\n')
print(df.sort_values(by=['Mortality national comparison'],ascending=True))
print('\n'+'-'*250+'\n')
print(df.sort_values(by=['Location'],ascending=True))
print('\n'+'-'*250+'\n')
print(df.fillna(0,inplace=True))
print('\n'+'-'*250+'\n')
print(df.fillna(' '))
print('\n'+'-'*250+'\n')
```

```
    The Dataframe:
          index  Provider ID                    Hospital Name  \
    0         0        10005      MARSHALL MEDICAL CENTER SOUTH
    1         1        10032                   WEDOWEE HOSPITAL
    2         2        10131            CRESTWOOD MEDICAL CENTER
    3         3        20001    PROVIDENCE ALASKA MEDICAL CENTER
    4         4        20018  YUKON KUSKOKWIM DELTA REG HOSPITAL
    ...     ...          ...                                ...
    4813   4813       501333  KITTITAS VALLEY COMMUNITY HOSPITAL
    4814   4814       450827         KELL WEST REGIONAL HOSPITAL
    4815   4815       501310         NEWPORT COMMUNITY HOSPITAL
    4816   4816       670107    BAYLOR EMERGENCY MEDICAL CENTER
    4817   4817       450213            UNIVERSITY HEALTH SYSTEM

                             Address         City State  ZIP Code  \
    0        2505 U S HIGHWAY 431 NORTH         BOAZ    AL     35957
    1            209 NORTH MAIN STREET      WEDOWEE    AL     36278
    2              ONE HOSPITAL DR SE    HUNTSVILLE    AL     35801
    3                    BOX 196604      ANCHORAGE    AK     99508
    4                    PO BOX 287         BETHEL    AK     99559
    ...                          ...          ...   ...       ...
    4813          603 SOUTH CHESTNUT    ELLENSBURG    WA     98926
    4814     5402 KELL WEST BOULEVARD  WICHITA FALLS   TX     76310
    4815           714 WEST PINE STREET      NEWPORT    WA     99156
    4816  12500 SOUTH FREEWAY SUITE 100     BURLESON    TX     76028
    4817               4502 MEDICAL DR   SAN ANTONIO    TX     78229

           County Name  Phone Number          Hospital Type  ...  \
    0         MARSHALL    2565938310       Acute Care Hospitals  ...
    1         RANDOLPH    2563572111       Acute Care Hospitals  ...
    2          MADISON    2568823100       Acute Care Hospitals  ...
    3        ANCHORAGE    9072123675       Acute Care Hospitals  ...
    4           BETHEL    9075436300       Acute Care Hospitals  ...
    ...            ...          ...                     ...  ...
    4813       KITTITAS    5099629841  Critical Access Hospitals  ...
    4814        WICHITA    9406925888       Acute Care Hospitals  ...
    4815  PEND OREILLE    5094472441  Critical Access Hospitals  ...
    4816        JOHNSON    2142946250       Acute Care Hospitals  ...
    4817          BEXAR    2103584000       Acute Care Hospitals  ...

             Readmission national comparison footnote  \
    0                                            NaN
    1                                            NaN
    2                                            NaN
    3                                            NaN
    4                                            NaN
    ...                                          ...
    4813                                         NaN
    4814                                         NaN
    4815                                         NaN
    4816  Results are not available for this reporting p...
    4817                                         NaN

             Patient experience national comparison  \
    0                 Same as the National average
    1                              Not Available
    2                 Same as the National average
```

Colab paid products  -  Cancel contracts here