

```
In [5]: # Import Libraries
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Set up visuals
sns.set(style="whitegrid")
plt.rcParams["figure.figsize"] = (12, 6)

# Load dataset
file_path = 'C:/Users/shazi/data/Hospital_General_Information.csv' # Adjust path i
df = pd.read_csv(file_path)

# Preview data
print("Dataset shape:", df.shape)
print(df.head())

# Clean column names
df.columns = df.columns.str.lower().str.replace(' ', '_').str.replace('(', '').str.

# Check missing values
print(df.isnull().sum())

# Top 10 states by number of hospitals
top_states = df['state'].value_counts().head(10)
print("Top 10 states by number of hospitals:\n", top_states)

# Plot: Top states by number of hospitals
top_states.plot(kind='bar', color='skyblue')
plt.title("Top 10 States by Number of Hospitals")
plt.xlabel("State")
plt.ylabel("Number of Hospitals")
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()

# Hospital types distribution
hospital_types = df['hospital_type'].value_counts()
hospital_types.plot(kind='pie', autopct='%1.1f%%', startangle=140, colors=sns.color
plt.title("Distribution of Hospital Types")
plt.ylabel("")
plt.tight_layout()
plt.show()

# Ownership distribution
ownership = df['hospital_ownership'].value_counts()
ownership.plot(kind='bar', color='lightgreen')
plt.title("Hospital Ownership Types")
plt.xlabel("Ownership")
plt.ylabel("Number of Hospitals")
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```

```
# Average hospital rating (if available)
if 'hospital_overall_rating' in df.columns:
    rating_counts = df['hospital_overall_rating'].value_counts().sort_index()
    rating_counts.plot(kind='bar', color='coral')
    plt.title("Hospital Overall Rating Distribution")
    plt.xlabel("Rating")
    plt.ylabel("Count")
    plt.xticks(rotation=0)
    plt.tight_layout()
    plt.show()
else:
    print("No overall rating column in dataset.")

# Export cleaned data
df.to_csv('C:/Users/shazi/data/output/cleaned_cms_hospital_info.csv', index=False)
print("Cleaned data exported to outputs/cleaned_cms_hospital_info.csv")
```

Dataset shape: (5396, 38)

	Facility ID	Facility Name	Address \
0	010001	SOUTHEAST HEALTH MEDICAL CENTER	1108 ROSS CLARK CIRCLE
1	010005	MARSHALL MEDICAL CENTERS	2505 U S HIGHWAY 431 NORTH
2	010006	NORTH ALABAMA MEDICAL CENTER	1701 VETERANS DRIVE
3	010007	MIZELL MEMORIAL HOSPITAL	702 N MAIN ST
4	010008	CRENSHAW COMMUNITY HOSPITAL	101 HOSPITAL CIRCLE

	City/Town	State	ZIP Code	County/Parish	Telephone Number \
0	DOTHAN	AL	36301	HOUSTON	(334) 793-8701
1	BOAZ	AL	35957	MARSHALL	(256) 593-8310
2	FLORENCE	AL	35630	LAUDERDALE	(256) 629-1000
3	OPP	AL	36467	COVINGTON	(334) 493-3541
4	LUVERNE	AL	36049	CRENSHAW	(334) 335-3374

	Hospital Type	Hospital Ownership ... \
0	Acute Care Hospitals	Government - Hospital District or Authority ...
1	Acute Care Hospitals	Government - Hospital District or Authority ...
2	Acute Care Hospitals	Proprietary ...
3	Acute Care Hospitals	Voluntary non-profit - Private ...
4	Acute Care Hospitals	Proprietary ...

	Count of READM Measures Better	Count of READM Measures No Different \
0	1	8
1	0	8
2	0	7
3	0	7
4	0	2

	Count of READM Measures Worse	READM Group Footnote \
0	2	NaN
1	1	NaN
2	2	NaN
3	0	NaN
4	0	NaN

	Pt Exp Group Measure Count	Count of Facility Pt Exp Measures \
0	8	8
1	8	8
2	8	8
3	8	8
4	8	Not Available

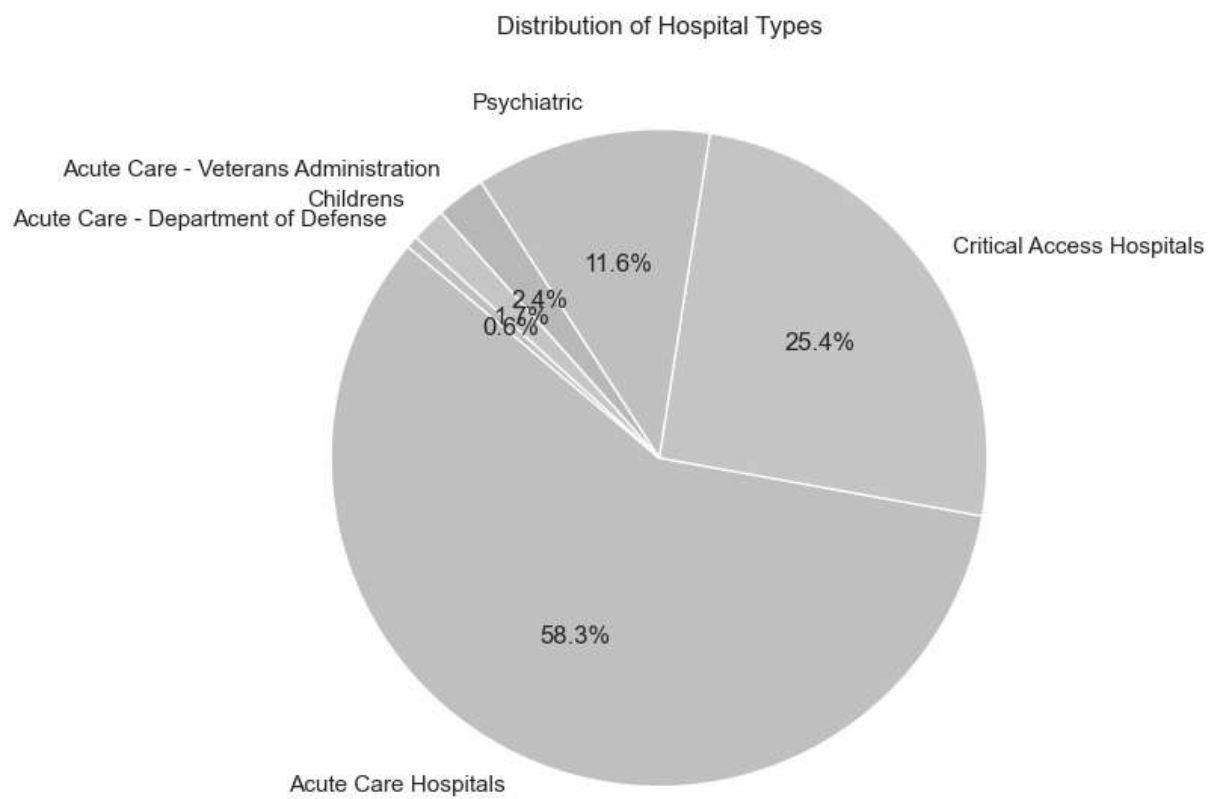
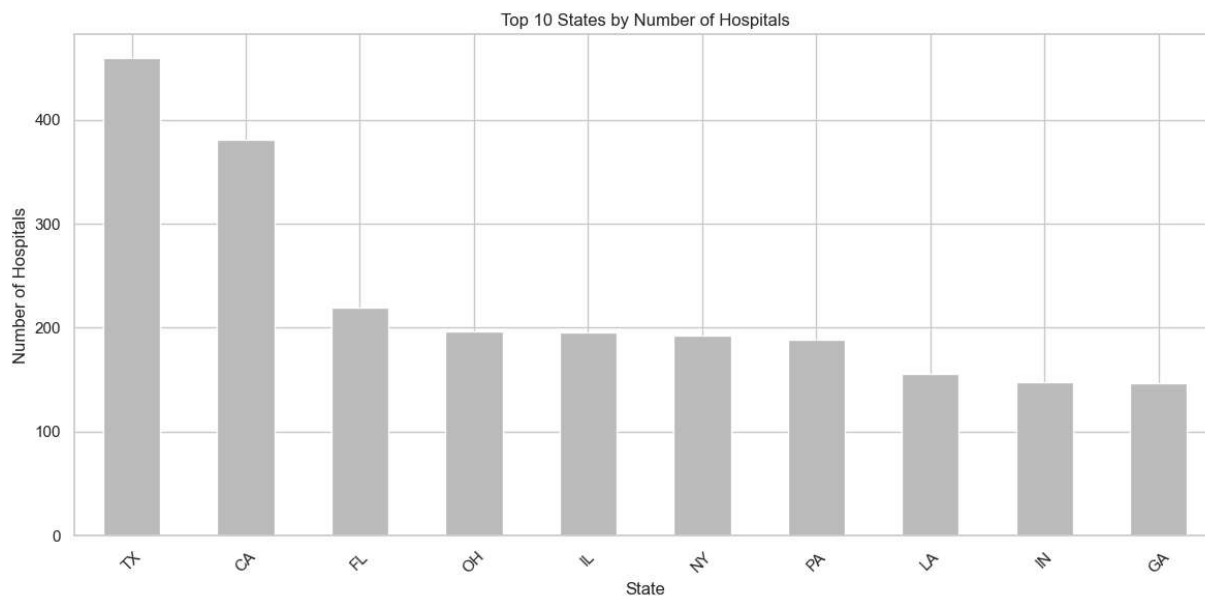
	Pt Exp Group Footnote	TE Group Measure Count	Count of Facility TE Measures \
0	NaN	12	10
1	NaN	12	12
2	NaN	12	11
3	NaN	12	7
4	5.0	12	6

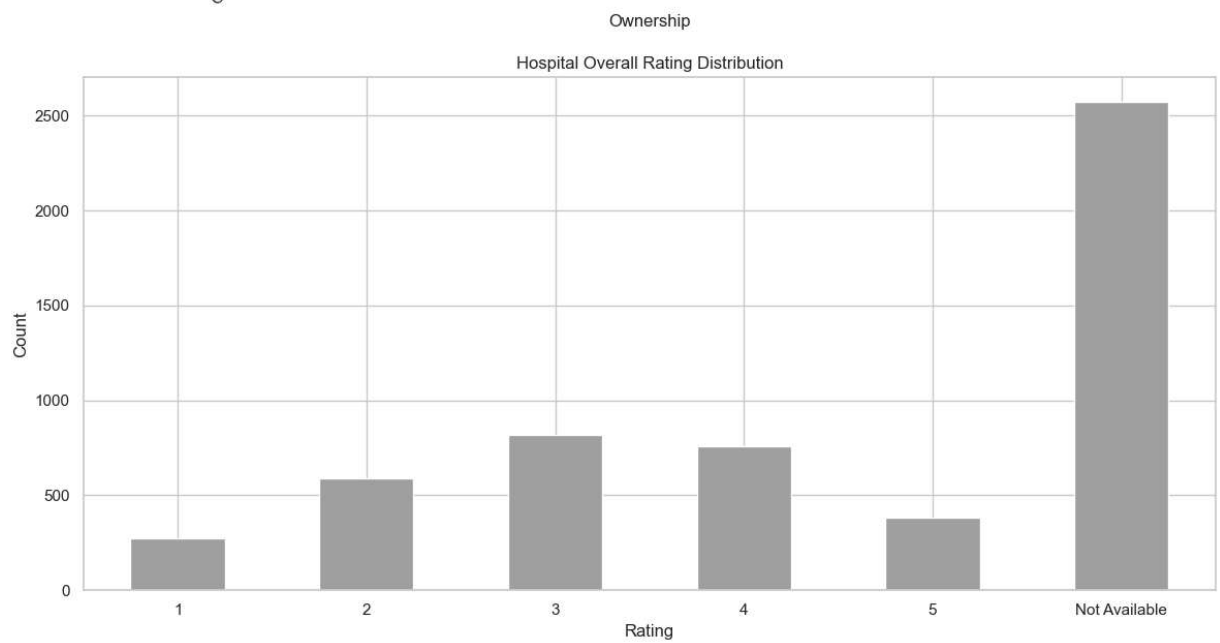
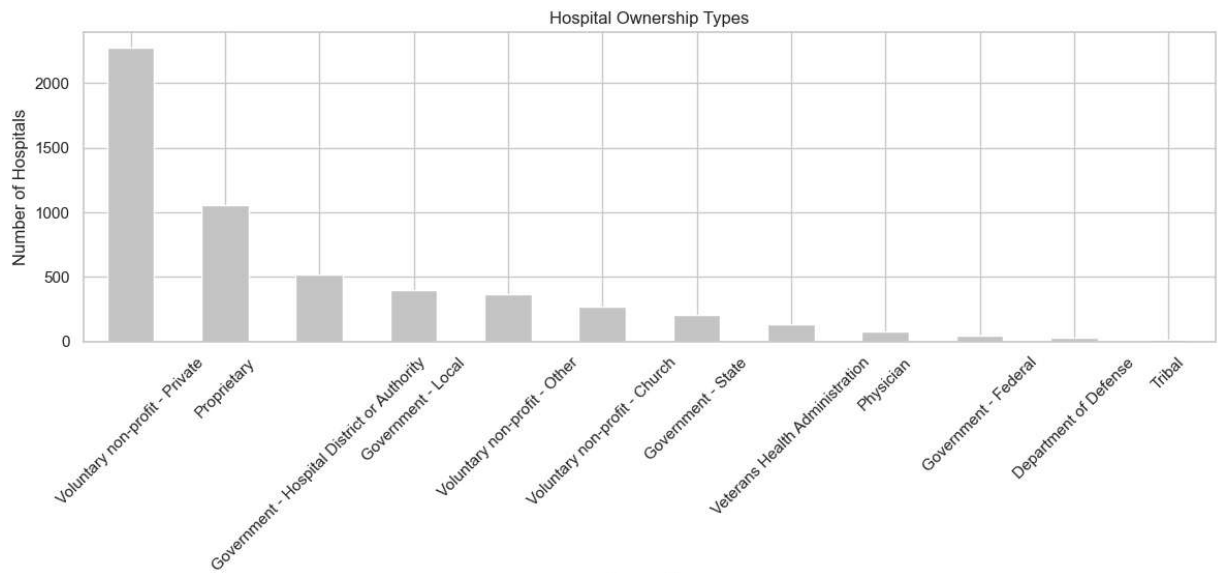
	TE Group Footnote
0	NaN
1	NaN
2	NaN
3	NaN
4	NaN

```

[5 rows x 38 columns]
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 3163
hospital_overall_rating  0
hospital_overall_rating_footnote 2777
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      3527
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    3406
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     4315
pt_exp_group_measure_count 0
count_of_facility_pt_exp_measures 0
pt_exp_group_footnote    3207
te_group_measure_count   0
count_of_facility_te_measures 0
te_group_footnote        4509
dtype: int64
Top 10 states by number of hospitals:
state
TX    459
CA    381
FL    219
OH    196
IL    195
NY    192
PA    188
LA    156
IN    148
GA    147
Name: count, dtype: int64

```





Cleaned data exported to outputs/cleaned\_cms\_hospital\_info.csv

In [ ]: