

1.2 Assignment: R/Python Refresher

Introduction to the Hospital Financial and Operational Efficiency Analysis

This analysis project utilizes the `HospitalFinancialData21-22.xlsx` dataset, sourced from the California Health and Human Services Agency's website. The dataset provides a detailed view of financial and operational metrics from 439 hospitals across California, covering the fiscal years 2021-2022.

Dataset Overview:

- **Composition:** The dataset includes 439 rows, each representing a unique hospital, and contains 246 columns with key metrics such as Net Income, Total Assets, Total Liabilities, and the Number of Licensed Beds.
- **Focus:** The analysis primarily concentrates on financial indicators like net income and operational metrics such as the number of licensed beds, offering insights into hospital revenue and bed utilization efficiency.

Project Context:

- **Objective:** The primary aim of this project is to analyze hospital revenue and operational efficiency, focusing on how hospital size and geographical location influence these factors.
- **Analytical Approach:** The project involves exploring relationships between hospital size, revenue, and bed utilization efficiency across different counties in California.

This project seeks to provide valuable insights into the financial and operational aspects of California's healthcare landscape, aiding in understanding the dynamics of hospital revenue generation and resource utilization.

Data Loading

Import necessary libraries and load the dataset for analysis. Display the first few rows to understand the basic structure of the data.

In [20]:  import pandas as pd

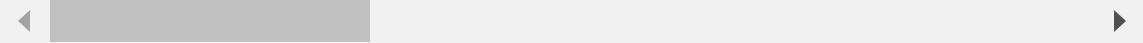
```
# Load the dataset
file_path = 'HospitalFinancialData21-22.xlsx'
data = pd.read_excel(file_path)

data.head()
```

Out[20]:

	Unnamed: 0.1	Unnamed: 0	Facility_ID	Hospital_Name	Financial_Year_Start	Financial_Year_E
0	0	0	106580996	ADVENTIST HEALTH AND RIDEOUT	2021-01-01	2021-12-
1	1	1	106150788	ADVENTIST HEALTH BAKERSFIELD	2021-01-01	2021-12-
2	2	2	106171049	ADVENTIST HEALTH CLEARLAKE	2021-01-01	2021-12-
3	3	3	106150706	ADVENTIST HEALTH DELANO	2021-01-01	2021-12-
4	4	4	106190323	ADVENTIST HEALTH GLENDALE	2021-01-01	2021-12-

5 rows × 248 columns



Initial Data Exploration

Display the shape of the dataset and basic statistics to gain an initial understanding of the data's scale and distribution.

In [21]: # Initial Data Exploration

```
# Display the shape of the dataset
print('Dataset Shape:', data.shape)

# Display basic statistics
print('\nBasic Statistics:')
data.describe()
```

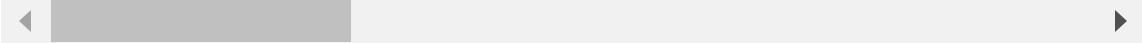
Dataset Shape: (439, 248)

Basic Statistics:

Out[21]:

	Unnamed: 0.1	Unnamed: 0	Facility_ID	Reporting_Period	Health_Service_Area	Health_
count	439.00000	439.00000	4.390000e+02	439.000000	439.000000	
mean	219.00000	219.00000	1.062846e+08	360.806378	8.353075	
std	126.87264	126.87264	1.356147e+05	30.974675	4.140863	
min	0.00000	0.00000	1.060107e+08	15.000000	1.000000	
25%	109.50000	109.50000	1.061904e+08	365.000000	5.000000	
50%	219.00000	219.00000	1.063013e+08	365.000000	10.000000	
75%	328.50000	328.50000	1.063740e+08	365.000000	11.000000	
max	438.00000	438.00000	1.065810e+08	395.000000	14.000000	

8 rows × 225 columns



Analysis Questions

Questions that the analysis aims to answer

Hospital Bed Efficiency: How does the ratio of staffed beds (BED_STF) to licensed beds (BED_LIC) vary across hospitals?

Average Net Patient Revenue by County: Which counties have the highest and lowest average Net Patient Revenue?

Correlation between Hospital Size and Net Patient Revenue: Is there a correlation between the size of the hospital (in terms of licensed bed count) and its Net Patient Revenue?

Revenue Distribution Across Different Hospital Sizes: How is the Net Patient Revenue distributed across different hospital sizes?

Analysis Question for Santa Clara and Sutter Counties: How does the staffed bed efficiency (ratio of staffed beds to licensed beds) compare between Santa Clara and Sutter counties?

These questions are explored to gain insights into operational efficiency, regional economic

Data Selection

Select relevant columns for the analysis to focus on specific aspects of the dataset that are most pertinent to our questions.

In [22]:

```
# Selecting Relevant Columns for Analysis
relevant_columns = ['Net_Patient_Revenue', 'BED_LIC', 'BED_STF', 'County_Name']
selected_data = data[relevant_columns]

# Display the first few rows of the selected data
selected_data.head()
```

Out[22]:

	Net_Patient_Revenue	BED_LIC	BED_STF	County_Name
0	442709734	221	192	Yuba
1	478276664	254	195	Kern
2	153326934	25	18	Lake
3	98896954	156	77	Kern
4	485170098	515	294	Los Angeles

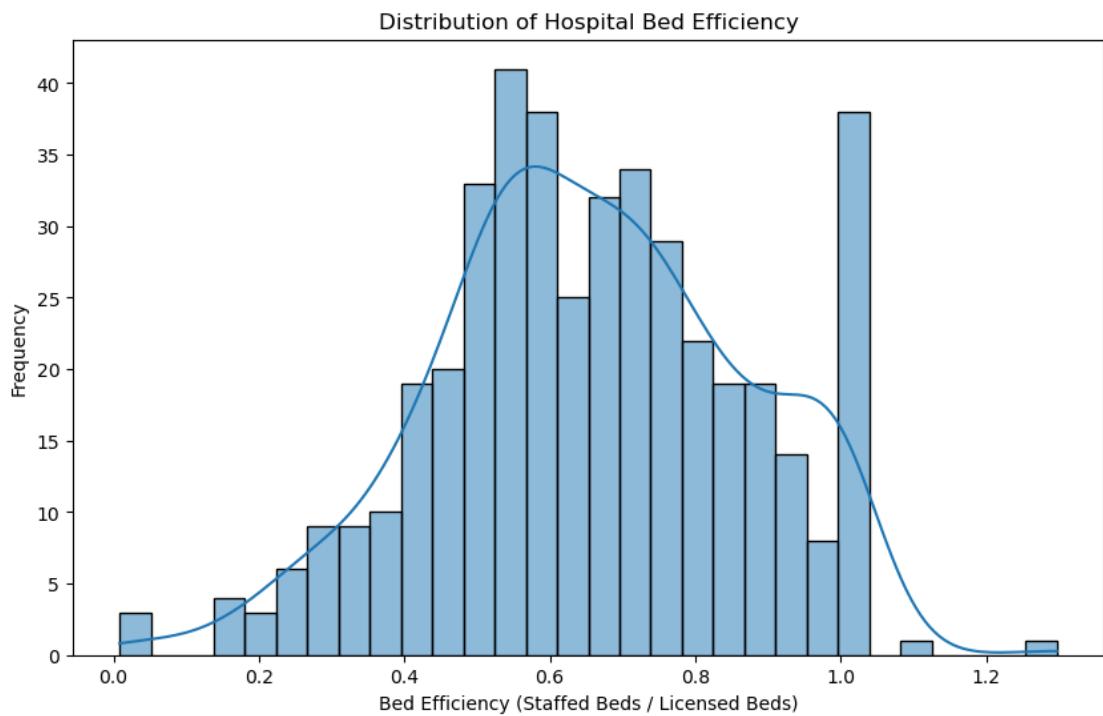
Histogram

To visualize the efficiency of hospital bed usage to understand how well hospitals are utilizing their available resources.

```
In [23]: ┌─ import matplotlib.pyplot as plt
      ┌─ import seaborn as sns

      # Hospital Efficiency: BED_STF to BED_LIC Ratio
      selected_data = data[relevant_columns].copy()
      selected_data['Bed_Efficiency'] = selected_data['BED_STF'] / selected_data['BED_LIC']

      # Histogram for Bed Efficiency
      plt.figure(figsize=(10, 6))
      sns.histplot(data=selected_data, x='Bed_Efficiency', bins=30, kde=True)
      plt.title('Distribution of Hospital Bed Efficiency')
      plt.xlabel('Bed Efficiency (Staffed Beds / Licensed Beds)')
      plt.ylabel('Frequency')
      plt.show()
```

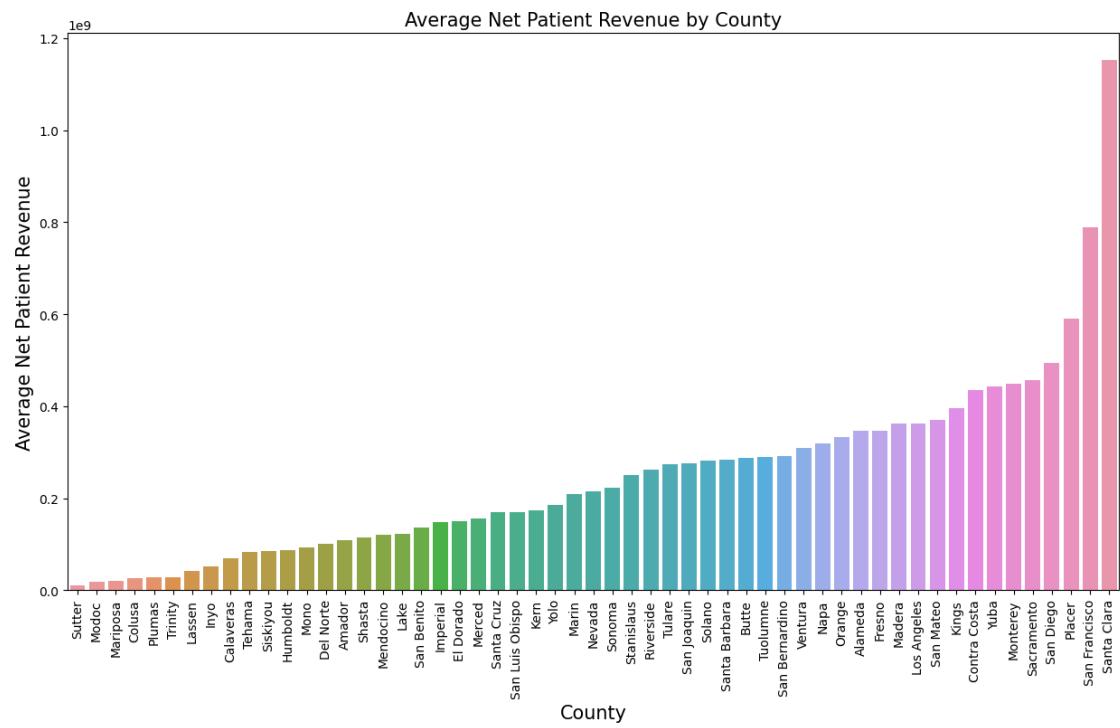


Bar Graph

To analyze and visualize the average net patient revenue by county to identify geographical variations in hospital revenues.

```
In [24]: # Average Net Patient Revenue by County
avg_revenue_by_county = selected_data.groupby('County_Name')[['Net_Patient_Revenue']].mean()

# Bar Graph for Average Revenue by County
plt.figure(figsize=(15, 8))
sns.barplot(x=avg_revenue_by_county.index, y=avg_revenue_by_county.values)
plt.xticks(rotation='vertical')
plt.title('Average Net Patient Revenue by County', fontsize=15) # Increase font size
plt.xlabel('County', fontsize=15) # Increase font size for x-label
plt.ylabel('Average Net Patient Revenue', fontsize=15) # Increase font size for y-label
plt.show()
```

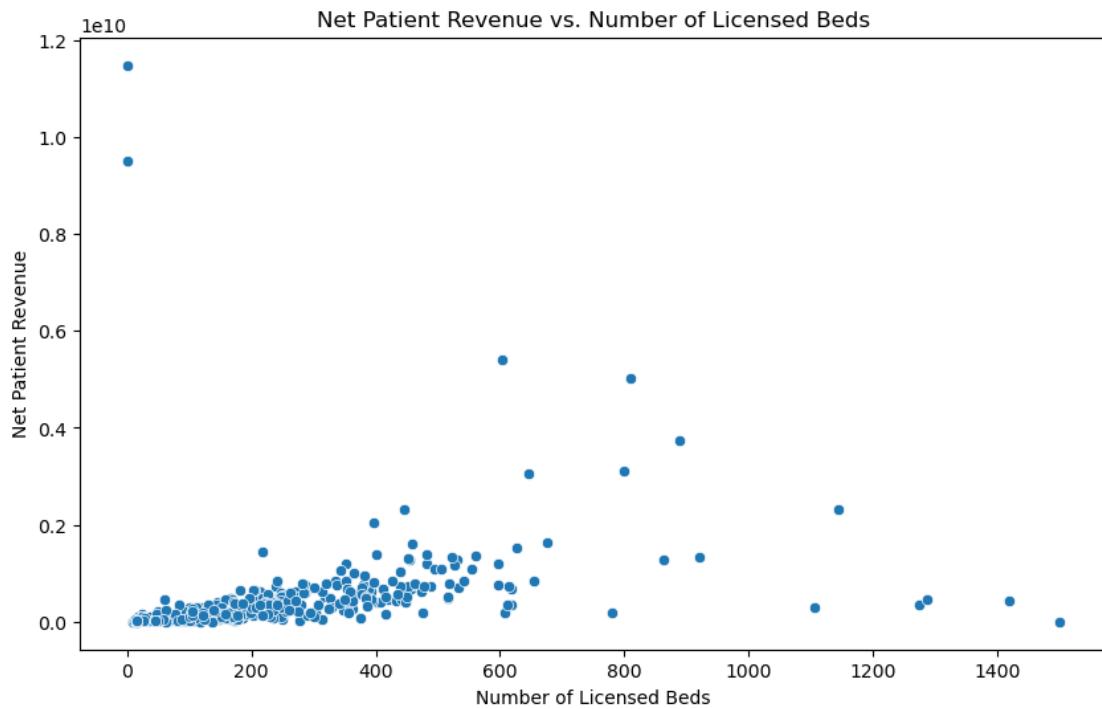


Scatter Plot

To explore the relationship between hospital size (in terms of licensed beds) and net patient revenue.

In [25]:

```
# Correlation between Hospital Size and Net Patient Revenue
plt.figure(figsize=(10, 6))
sns.scatterplot(data=selected_data, x='BED_LIC', y='Net_Patient_Revenue')
plt.title('Net Patient Revenue vs. Number of Licensed Beds')
plt.xlabel('Number of Licensed Beds')
plt.ylabel('Net Patient Revenue')
plt.show()
```



Histogram

To investigate how net patient revenue is distributed across hospitals of different sizes, categorizing hospitals based on their bed count.

In [26]:

```

import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Categorizing hospitals based on BED_LIC
selected_data['Hospital_Size_Category'] = pd.cut(selected_data['BED_LIC'], bins=[0, 100000000, 200000000, 300000000, 400000000, 500000000], labels=['Very Small', 'Small', 'Medium', 'Large'])

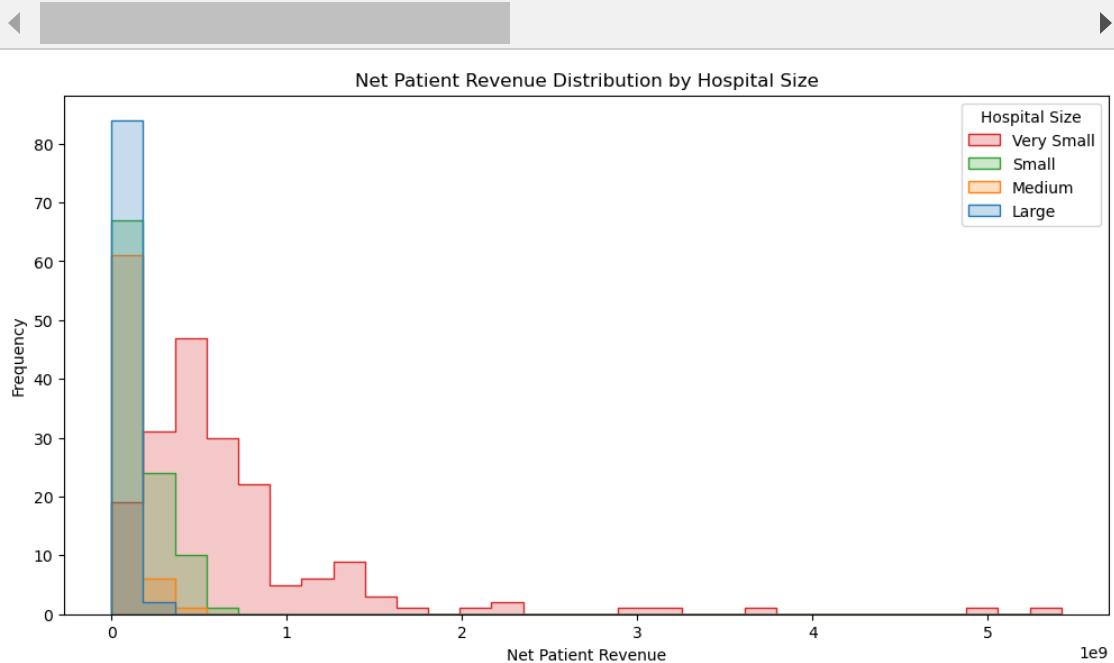
# Histogram for Net Patient Revenue by Hospital Size
plt.figure(figsize=(12, 6))
sns.histplot(data=selected_data, x='Net_Patient_Revenue', hue='Hospital_Size_Category', kde=False)

plt.title('Net Patient Revenue Distribution by Hospital Size')
plt.xlabel('Net Patient Revenue')
plt.ylabel('Frequency')

# Manually adding Legend
plt.legend(title='Hospital Size', labels=['Very Small', 'Small', 'Medium', 'Large'])

plt.show()

```



Box Plot

To compare the staffed bed efficiency between Santa Clara and Sutter counties to understand regional differences in operational efficiency.

```
In [27]: # import pandas as pd
# import matplotlib.pyplot as plt
# import seaborn as sns

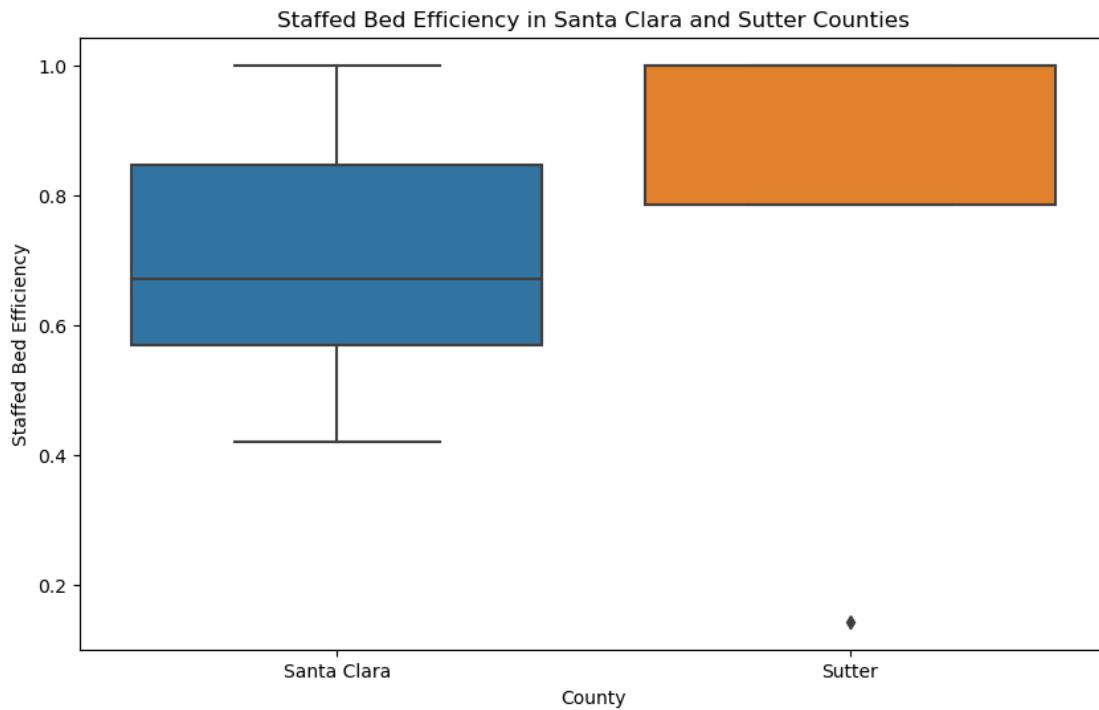
# Filter data for Santa Clara and Sutter counties
# Explicitly create a copy to avoid SettingWithCopyWarning
filtered_data = selected_data[selected_data['County_Name'].isin(['Santa Clara', 'Sutter'])]

# Display the first few rows of the filtered data
filtered_data.head()

# Calculate staffed bed efficiency for the filtered data
filtered_data['Staffed_Bed_Efficiency'] = filtered_data['BED_STF'] / filtered_data['LICENSED_BEDS']

# Boxplot for Staffed Bed Efficiency in Santa Clara and Sutter Counties
plt.figure(figsize=(10, 6))
sns.boxplot(x='County_Name', y='Staffed_Bed_Efficiency', data=filtered_data)
plt.title('Staffed Bed Efficiency in Santa Clara and Sutter Counties')
plt.xlabel('County')
plt.ylabel('Staffed Bed Efficiency')
```

Out[27]: Text(0, 0.5, 'Staffed Bed Efficiency')



Answers

Based on the visualizations created, here are the answers to the identified questions along with a summary of the results and conclusions:

- Hospital Bed Efficiency:** The histogram of the ratio of staffed beds to licensed beds shows a varied distribution. Most hospitals have a ratio close to or below 1, indicating that they are utilizing most of their licensed beds. However, there are some with a ratio

significantly lower than 1, suggesting underutilization of licensed bed capacity. This could be due to various factors such as staffing limitations or patient demand.

2. **Average Net Patient Revenue by County:** The bar graph reveals significant variations in average Net Patient Revenue across counties. Some counties have notably higher average revenues (Santa Clara), and others show low revenue (Sutter), which could be influenced by factors like local economic conditions, population health needs, and hospital services offered. This suggests that geographical location is a significant factor in hospital revenue.
3. **Correlation between Hospital Size and Net Patient Revenue:** The scatter plot indicates a general trend where larger hospitals (in terms of licensed beds) tend to have higher Net Patient Revenue. However, the relationship is not strictly linear, and there are exceptions. This suggests that while size is an important factor in revenue, it is not the only determinant.
4. **Revenue Distribution Across Different Hospital Sizes:** The histogram categorizing hospitals by size shows that larger hospitals generally have a wider range and higher peaks in Net Patient Revenue. This supports the idea that larger hospitals, possibly due to more extensive services and capacity, tend to generate more revenue. However, there is considerable overlap between categories, indicating that size alone does not dictate revenue.
5. **Santa Clara and Sutter Comparison:** Sutter County shows a wider variation in efficiency. This means there's a significant variation in how hospitals utilize their bed capacity. Some hospitals are maximizing their bed usage effectively, but others are not, leading to an overall inconsistency in bed utilization across the county. In contrast, Santa Clara County shows a more consistent and narrower range of staffed bed efficiency. This demonstrates more uniformity in bed usage among its hospitals. Here, bed utilization is more balanced.

Conclusion

The analysis suggests that both hospital size and geographical location play significant roles in determining hospital revenue and operational efficiency. Larger hospitals generally have higher revenues, but this is not a universal rule. Geographical location also impacts revenue, with some counties showing notably higher average revenues than others. Additionally, the efficiency in bed utilization varies across hospitals, which could impact their operational effectiveness.

In particular, the comparison between Santa Clara and Sutter counties revealed that while Santa Clara demonstrates a more consistent and moderate use of bed capacity, Sutter County exhibits a wider range of efficiency in bed utilization. This inconsistency in Sutter County suggests that some hospitals are maximizing bed usage effectively, whereas others are underutilizing their resources, leading to a mixed scenario of operational efficiency within the county.