Empoweringpatientvoices

November 19, 2024

[1]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.ensemble import RandomForestRegressor
from sklearn.model_selection import train_test_split, GridSearchCV
from sklearn.metrics import mean_absolute_error, r2_score
from sklearn.preprocessing import LabelEncoder
# Import necessary libraries for Linear Regression and Support Vector Regression
from sklearn.linear_model import LinearRegression
from sklearn.svm import SVR
```

[2]:

```
hospital_beds = pd.read_csv("Hospital Beds.csv")
HCAHPS = pd.read_csv("v1 HCAHPS 2022.csv")
```

[3]:

```
hospital_beds
```

[3]:

```
      Provider CCN                        Hospital Name  \
0           441314         LAUDERDALE COMMUNITY HOSPITAL
1           341317   LIFEBRITE COMMUNITY HOPITAL OF STOK
2           511320              JACKSON GENERAL HOSPITAL
3           520011  LAKEVIEW MEDICAL CENTER OF RICE LAKE
4           520037             MARSHFIELD MEDICAL CENTER
…              …                                    …
6045        450379                 DALLAS MEDICAL CENTER
6046        450675                MEDICAL CITY ARLINGTON
6047        450775        HCA HOUSTON HEALTHCARE KINGWOOD
6048        450869        DOCTORS HOSPITAL AT RENAISSANCE
6049        462003       WESTERN PEAKS SPECIALTY HOSPITAL

     Fiscal Year Begin Date Fiscal Year End Date  number_of_beds
0                 10/1/2020           12/31/2020            25.0
1                 10/1/2020           12/31/2020            65.0
2                 10/1/2020           12/31/2020            25.0
3                 10/1/2020           12/31/2020            40.0
4                 10/1/2020           12/31/2020           198.0
…                      …                    …              …
```

```
6045               1/1/2021           12/31/2021            119.0
6046               6/1/2021            5/31/2022            358.0
6047              10/1/2020            9/30/2021            573.0
6048               1/1/2021           12/31/2021            519.0
6049               1/1/2021           12/31/2021            124.0

[6050 rows x 5 columns]
```

[4]:
```
HCAHPS
```

[4]:
```
       Facility ID                    Facility Name                  Address  \
0            10001  SOUTHEAST HEALTH MEDICAL CENTER   1108 ROSS CLARK CIRCLE
1            10001  SOUTHEAST HEALTH MEDICAL CENTER   1108 ROSS CLARK CIRCLE
2            10001  SOUTHEAST HEALTH MEDICAL CENTER   1108 ROSS CLARK CIRCLE
3            10001  SOUTHEAST HEALTH MEDICAL CENTER   1108 ROSS CLARK CIRCLE
4            10001  SOUTHEAST HEALTH MEDICAL CENTER   1108 ROSS CLARK CIRCLE
…            …                              …                      …
299920      670309  TEXAS HEALTH HOSPITAL MANSFIELD      2300 LONE STAR ROAD
299921      670309  TEXAS HEALTH HOSPITAL MANSFIELD      2300 LONE STAR ROAD
299922      670309  TEXAS HEALTH HOSPITAL MANSFIELD      2300 LONE STAR ROAD
299923      670309  TEXAS HEALTH HOSPITAL MANSFIELD      2300 LONE STAR ROAD
299924      670309  TEXAS HEALTH HOSPITAL MANSFIELD      2300 LONE STAR ROAD

        City/Town State  ZIP Code County/Parish Telephone Number  \
0          DOTHAN    AL     36301       HOUSTON   (334) 793-8701
1          DOTHAN    AL     36301       HOUSTON   (334) 793-8701
2          DOTHAN    AL     36301       HOUSTON   (334) 793-8701
3          DOTHAN    AL     36301       HOUSTON   (334) 793-8701
4          DOTHAN    AL     36301       HOUSTON   (334) 793-8701
…           …   …       …            …              …
299920  MANSFIELD    TX     76063       TARRANT   (682) 341-5000
299921  MANSFIELD    TX     76063       TARRANT   (682) 341-5000
299922  MANSFIELD    TX     76063       TARRANT   (682) 341-5000
299923  MANSFIELD    TX     76063       TARRANT   (682) 341-5000
299924  MANSFIELD    TX     76063       TARRANT   (682) 341-5000

            HCAHPS Measure ID  \
0                H_COMP_1_A_P
1               H_COMP_1_SN_P
2                H_COMP_1_U_P
3       H_COMP_1_LINEAR_SCORE
4        H_COMP_1_STAR_RATING
…                       …
299920            H_RECMND_DY
299921            H_RECMND_PY
299922  H_RECMND_LINEAR_SCORE
299923   H_RECMND_STAR_RATING
```

```
299924          H_STAR_RATING

                                          HCAHPS Question  \
0       Patients who reported that their nurses "Alway…
1       Patients who reported that their nurses "Somet…
2       Patients who reported that their nurses "Usual…
3                 Nurse communication - linear mean score
4                       Nurse communication - star rating
…                                                    …
299920  Patients who reported YES, they would definite…
299921  Patients who reported YES, they would probably…
299922            Recommend hospital - linear mean score
299923                  Recommend hospital - star rating
299924                               Summary star rating

                                HCAHPS Answer Description  \
0                       Nurses "always" communicated well
1         Nurses "sometimes" or "never" communicated well
2                      Nurses "usually" communicated well
3                 Nurse communication - linear mean score
4                       Nurse communication - star rating
…                                                    …
299920  "YES", patients would definitely recommend the…
299921  "YES", patients would probably recommend the h…
299922            Recommend hospital - linear mean score
299923                  Recommend hospital - star rating
299924                               Summary star rating

        HCAHPS Answer Percent  Number of Completed Surveys  \
0                        74.0                          536
1                         8.0                          536
2                        18.0                          536
3                         NaN                          536
4                         NaN                          536
…                       …                          …
299920                   72.0                          174
299921                   19.0                          174
299922                    NaN                          174
299923                    NaN                          174
299924                    NaN                          174

        Survey Response Rate Percent Start Date    End Date
0                                 15   1/1/2022  12/31/2022
1                                 15   1/1/2022  12/31/2022
2                                 15   1/1/2022  12/31/2022
3                                 15   1/1/2022  12/31/2022
4                                 15   1/1/2022  12/31/2022
```

```
…                              …        …         …
299920                            12   1/1/2022  12/31/2022
299921                            12   1/1/2022  12/31/2022
299922                            12   1/1/2022  12/31/2022
299923                            12   1/1/2022  12/31/2022
299924                            12   1/1/2022  12/31/2022

[299925 rows x 16 columns]
```

[5]:

```python
# Data Cleaning
hospital_beds.rename(columns={'Provider CCN': 'Facility ID'}, inplace=True)
hospital_beds['Facility ID'] = hospital_beds['Facility ID'].astype(str)
HCAHPS['Facility ID'] = HCAHPS['Facility ID'].astype(str)
```

[6]:

```python
# Merge Datasets
merged_data = pd.merge(hospital_beds, HCAHPS, on='Facility ID', how='inner')
```

[7]:

```python
merged_data
```

[7]:

```
       Facility ID                    Hospital Name Fiscal Year Begin Date  \
0           511320         JACKSON GENERAL HOSPITAL              10/1/2020
1           511320         JACKSON GENERAL HOSPITAL              10/1/2020
2           511320         JACKSON GENERAL HOSPITAL              10/1/2020
3           511320         JACKSON GENERAL HOSPITAL              10/1/2020
4           511320         JACKSON GENERAL HOSPITAL              10/1/2020
…            …                              …                    …
289969      450869  DOCTORS HOSPITAL AT RENAISSANCE               1/1/2021
289970      450869  DOCTORS HOSPITAL AT RENAISSANCE               1/1/2021
289971      450869  DOCTORS HOSPITAL AT RENAISSANCE               1/1/2021
289972      450869  DOCTORS HOSPITAL AT RENAISSANCE               1/1/2021
289973      450869  DOCTORS HOSPITAL AT RENAISSANCE               1/1/2021

       Fiscal Year End Date  number_of_beds                    Facility Name  \
0                12/31/2020            25.0         JACKSON GENERAL HOSPITAL
1                12/31/2020            25.0         JACKSON GENERAL HOSPITAL
2                12/31/2020            25.0         JACKSON GENERAL HOSPITAL
3                12/31/2020            25.0         JACKSON GENERAL HOSPITAL
4                12/31/2020            25.0         JACKSON GENERAL HOSPITAL
…                     …             …                              …
289969           12/31/2021           519.0  DOCTORS HOSPITAL AT RENAISSANCE
289970           12/31/2021           519.0  DOCTORS HOSPITAL AT RENAISSANCE
289971           12/31/2021           519.0  DOCTORS HOSPITAL AT RENAISSANCE
289972           12/31/2021           519.0  DOCTORS HOSPITAL AT RENAISSANCE
289973           12/31/2021           519.0  DOCTORS HOSPITAL AT RENAISSANCE

                  Address City/Town State  ZIP Code County/Parish  \
0          122 PINNELL ST    RIPLEY    WV     25271       JACKSON
```

```
1          122 PINNELL ST    RIPLEY    WV     25271       JACKSON
2          122 PINNELL ST    RIPLEY    WV     25271       JACKSON
3          122 PINNELL ST    RIPLEY    WV     25271       JACKSON
4          122 PINNELL ST    RIPLEY    WV     25271       JACKSON
…                      …        …   …       …           …
289969  5501 SOUTH MCCOLL  EDINBURG    TX     78539       HIDALGO
289970  5501 SOUTH MCCOLL  EDINBURG    TX     78539       HIDALGO
289971  5501 SOUTH MCCOLL  EDINBURG    TX     78539       HIDALGO
289972  5501 SOUTH MCCOLL  EDINBURG    TX     78539       HIDALGO
289973  5501 SOUTH MCCOLL  EDINBURG    TX     78539       HIDALGO

       Telephone Number      HCAHPS Measure ID  \
0        (304) 372-2731           H_COMP_1_A_P
1        (304) 372-2731          H_COMP_1_SN_P
2        (304) 372-2731           H_COMP_1_U_P
3        (304) 372-2731  H_COMP_1_LINEAR_SCORE
4        (304) 372-2731   H_COMP_1_STAR_RATING
…                   …                    …
289969   (956) 362-8677            H_RECMND_DY
289970   (956) 362-8677            H_RECMND_PY
289971   (956) 362-8677  H_RECMND_LINEAR_SCORE
289972   (956) 362-8677   H_RECMND_STAR_RATING
289973   (956) 362-8677          H_STAR_RATING

                                             HCAHPS Question  \
0       Patients who reported that their nurses "Alway…
1       Patients who reported that their nurses "Somet…
2       Patients who reported that their nurses "Usual…
3                 Nurse communication - linear mean score
4                       Nurse communication - star rating
…                                                    …
289969  Patients who reported YES, they would definite…
289970  Patients who reported YES, they would probably…
289971             Recommend hospital - linear mean score
289972                   Recommend hospital - star rating
289973                                Summary star rating

                                   HCAHPS Answer Description  \
0                       Nurses "always" communicated well
1         Nurses "sometimes" or "never" communicated well
2                      Nurses "usually" communicated well
3                 Nurse communication - linear mean score
4                       Nurse communication - star rating
…                                                    …
289969  "YES", patients would definitely recommend the…
289970  "YES", patients would probably recommend the h…
289971             Recommend hospital - linear mean score
```

```
289972                   Recommend hospital - star rating
289973                                Summary star rating

        HCAHPS Answer Percent  Number of Completed Surveys  \
0                        88.0                          138
1                         2.0                          138
2                        10.0                          138
3                         NaN                          138
4                         NaN                          138
…                         …                            …
289969                   59.0                         1237
289970                   31.0                         1237
289971                    NaN                         1237
289972                    NaN                         1237
289973                    NaN                         1237

        Survey Response Rate Percent Start Date   End Date
0                                 31   1/1/2022  12/31/2022
1                                 31   1/1/2022  12/31/2022
2                                 31   1/1/2022  12/31/2022
3                                 31   1/1/2022  12/31/2022
4                                 31   1/1/2022  12/31/2022
…                                …        …         …
289969                             9   1/1/2022  12/31/2022
289970                             9   1/1/2022  12/31/2022
289971                             9   1/1/2022  12/31/2022
289972                             9   1/1/2022  12/31/2022
289973                             9   1/1/2022  12/31/2022

[289974 rows x 20 columns]
```

[8]:
```
# Handle Missing Values
merged_data['number_of_beds'].fillna(merged_data['number_of_beds'].median(),␣
 ↪inplace=True)
merged_data['HCAHPS Answer Percent'].fillna(merged_data['HCAHPS Answer␣
 ↪Percent'].mean(), inplace=True)
```

[9]:
```
# Standardize Text fields
merged_data['Facility Name'] = merged_data['Facility Name'].str.title()
merged_data['City/Town'] = merged_data['City/Town'].str.title()
merged_data['State'] = merged_data['State'].str.upper()
```

[10]:
```
merged_data
```

[10]:
```
       Facility ID                    Hospital Name Fiscal Year Begin Date  \
0           511320         JACKSON GENERAL HOSPITAL              10/1/2020
1           511320         JACKSON GENERAL HOSPITAL              10/1/2020
```

```
2           511320         JACKSON GENERAL HOSPITAL              10/1/2020
3           511320         JACKSON GENERAL HOSPITAL              10/1/2020
4           511320         JACKSON GENERAL HOSPITAL              10/1/2020
…             …                                  …                  …
289969      450869  DOCTORS HOSPITAL AT RENAISSANCE               1/1/2021
289970      450869  DOCTORS HOSPITAL AT RENAISSANCE               1/1/2021
289971      450869  DOCTORS HOSPITAL AT RENAISSANCE               1/1/2021
289972      450869  DOCTORS HOSPITAL AT RENAISSANCE               1/1/2021
289973      450869  DOCTORS HOSPITAL AT RENAISSANCE               1/1/2021

       Fiscal Year End Date  number_of_beds               Facility Name  \
0                12/31/2020            25.0         Jackson General Hospital
1                12/31/2020            25.0         Jackson General Hospital
2                12/31/2020            25.0         Jackson General Hospital
3                12/31/2020            25.0         Jackson General Hospital
4                12/31/2020            25.0         Jackson General Hospital
…                     …              …                              …
289969           12/31/2021           519.0  Doctors Hospital At Renaissance
289970           12/31/2021           519.0  Doctors Hospital At Renaissance
289971           12/31/2021           519.0  Doctors Hospital At Renaissance
289972           12/31/2021           519.0  Doctors Hospital At Renaissance
289973           12/31/2021           519.0  Doctors Hospital At Renaissance

                  Address City/Town State  ZIP Code County/Parish  \
0          122 PINNELL ST    Ripley    WV     25271       JACKSON
1          122 PINNELL ST    Ripley    WV     25271       JACKSON
2          122 PINNELL ST    Ripley    WV     25271       JACKSON
3          122 PINNELL ST    Ripley    WV     25271       JACKSON
4          122 PINNELL ST    Ripley    WV     25271       JACKSON
…                     …        …    …        …             …
289969  5501 SOUTH MCCOLL  Edinburg    TX     78539       HIDALGO
289970  5501 SOUTH MCCOLL  Edinburg    TX     78539       HIDALGO
289971  5501 SOUTH MCCOLL  Edinburg    TX     78539       HIDALGO
289972  5501 SOUTH MCCOLL  Edinburg    TX     78539       HIDALGO
289973  5501 SOUTH MCCOLL  Edinburg    TX     78539       HIDALGO

       Telephone Number      HCAHPS Measure ID  \
0        (304) 372-2731           H_COMP_1_A_P
1        (304) 372-2731          H_COMP_1_SN_P
2        (304) 372-2731           H_COMP_1_U_P
3        (304) 372-2731  H_COMP_1_LINEAR_SCORE
4        (304) 372-2731   H_COMP_1_STAR_RATING
…                   …                    …
289969   (956) 362-8677            H_RECMND_DY
289970   (956) 362-8677            H_RECMND_PY
289971   (956) 362-8677  H_RECMND_LINEAR_SCORE
289972   (956) 362-8677   H_RECMND_STAR_RATING
```

```
289973   (956) 362-8677          H_STAR_RATING

                                          HCAHPS Question  \
0       Patients who reported that their nurses "Alway…
1       Patients who reported that their nurses "Somet…
2       Patients who reported that their nurses "Usual…
3                 Nurse communication - linear mean score
4                       Nurse communication - star rating
…                                                   …
289969  Patients who reported YES, they would definite…
289970  Patients who reported YES, they would probably…
289971             Recommend hospital - linear mean score
289972                   Recommend hospital - star rating
289973                                Summary star rating

                                HCAHPS Answer Description  \
0                       Nurses "always" communicated well
1         Nurses "sometimes" or "never" communicated well
2                      Nurses "usually" communicated well
3                 Nurse communication - linear mean score
4                       Nurse communication - star rating
…                                                   …
289969  "YES", patients would definitely recommend the…
289970  "YES", patients would probably recommend the h…
289971             Recommend hospital - linear mean score
289972                   Recommend hospital - star rating
289973                                Summary star rating

        HCAHPS Answer Percent  Number of Completed Surveys  \
0                   88.000000                          138
1                    2.000000                          138
2                   10.000000                          138
3                   34.722222                          138
4                   34.722222                          138
…                       …                          …
289969              59.000000                         1237
289970              31.000000                         1237
289971              34.722222                         1237
289972              34.722222                         1237
289973              34.722222                         1237

        Survey Response Rate Percent Start Date    End Date
0                                 31   1/1/2022  12/31/2022
1                                 31   1/1/2022  12/31/2022
2                                 31   1/1/2022  12/31/2022
3                                 31   1/1/2022  12/31/2022
4                                 31   1/1/2022  12/31/2022
```

```
…                               …        …         …
289969                              9  1/1/2022  12/31/2022
289970                              9  1/1/2022  12/31/2022
289971                              9  1/1/2022  12/31/2022
289972                              9  1/1/2022  12/31/2022
289973                              9  1/1/2022  12/31/2022

[289974 rows x 20 columns]
```

[11]:
```
# Check for and Remove Outliers
```

[12]:
```
Q1 = merged_data['number_of_beds'].quantile(0.25)
Q3 = merged_data['number_of_beds'].quantile(0.75)
IQR = Q3 - Q1
merged_data = merged_data[(merged_data['number_of_beds'] >= (Q1 - 1.5 * IQR)) &
                          (merged_data['number_of_beds'] <= (Q3 + 1.5 * IQR))]
```

[13]:
```
merged_data
```

[13]:
```
       Facility ID                    Hospital Name Fiscal Year Begin Date  \
0           511320         JACKSON GENERAL HOSPITAL              10/1/2020
1           511320         JACKSON GENERAL HOSPITAL              10/1/2020
2           511320         JACKSON GENERAL HOSPITAL              10/1/2020
3           511320         JACKSON GENERAL HOSPITAL              10/1/2020
4           511320         JACKSON GENERAL HOSPITAL              10/1/2020
…             …                              …                    …
289969      450869  DOCTORS HOSPITAL AT RENAISSANCE               1/1/2021
289970      450869  DOCTORS HOSPITAL AT RENAISSANCE               1/1/2021
289971      450869  DOCTORS HOSPITAL AT RENAISSANCE               1/1/2021
289972      450869  DOCTORS HOSPITAL AT RENAISSANCE               1/1/2021
289973      450869  DOCTORS HOSPITAL AT RENAISSANCE               1/1/2021

       Fiscal Year End Date  number_of_beds                    Facility Name  \
0                12/31/2020            25.0         Jackson General Hospital
1                12/31/2020            25.0         Jackson General Hospital
2                12/31/2020            25.0         Jackson General Hospital
3                12/31/2020            25.0         Jackson General Hospital
4                12/31/2020            25.0         Jackson General Hospital
…                     …             …                              …
289969           12/31/2021           519.0  Doctors Hospital At Renaissance
289970           12/31/2021           519.0  Doctors Hospital At Renaissance
289971           12/31/2021           519.0  Doctors Hospital At Renaissance
289972           12/31/2021           519.0  Doctors Hospital At Renaissance
289973           12/31/2021           519.0  Doctors Hospital At Renaissance

                  Address City/Town State  ZIP Code County/Parish  \
0          122 PINNELL ST    Ripley    WV     25271       JACKSON
```

```
1          122 PINNELL ST    Ripley    WV     25271       JACKSON
2          122 PINNELL ST    Ripley    WV     25271       JACKSON
3          122 PINNELL ST    Ripley    WV     25271       JACKSON
4          122 PINNELL ST    Ripley    WV     25271       JACKSON
…                   …       …   …       …           …
289969  5501 SOUTH MCCOLL  Edinburg    TX     78539       HIDALGO
289970  5501 SOUTH MCCOLL  Edinburg    TX     78539       HIDALGO
289971  5501 SOUTH MCCOLL  Edinburg    TX     78539       HIDALGO
289972  5501 SOUTH MCCOLL  Edinburg    TX     78539       HIDALGO
289973  5501 SOUTH MCCOLL  Edinburg    TX     78539       HIDALGO

       Telephone Number      HCAHPS Measure ID  \
0        (304) 372-2731           H_COMP_1_A_P
1        (304) 372-2731          H_COMP_1_SN_P
2        (304) 372-2731           H_COMP_1_U_P
3        (304) 372-2731  H_COMP_1_LINEAR_SCORE
4        (304) 372-2731   H_COMP_1_STAR_RATING
…                 …                    …
289969   (956) 362-8677            H_RECMND_DY
289970   (956) 362-8677            H_RECMND_PY
289971   (956) 362-8677  H_RECMND_LINEAR_SCORE
289972   (956) 362-8677   H_RECMND_STAR_RATING
289973   (956) 362-8677          H_STAR_RATING

                                          HCAHPS Question  \
0       Patients who reported that their nurses "Alway…
1       Patients who reported that their nurses "Somet…
2       Patients who reported that their nurses "Usual…
3                 Nurse communication - linear mean score
4                       Nurse communication - star rating
…                                                    …
289969  Patients who reported YES, they would definite…
289970  Patients who reported YES, they would probably…
289971             Recommend hospital - linear mean score
289972                   Recommend hospital - star rating
289973                                Summary star rating

                                HCAHPS Answer Description  \
0                       Nurses "always" communicated well
1         Nurses "sometimes" or "never" communicated well
2                      Nurses "usually" communicated well
3                 Nurse communication - linear mean score
4                       Nurse communication - star rating
…                                                    …
289969  "YES", patients would definitely recommend the…
289970  "YES", patients would probably recommend the h…
289971             Recommend hospital - linear mean score
```

```
289972                   Recommend hospital - star rating
289973                               Summary star rating

        HCAHPS Answer Percent  Number of Completed Surveys  \
0                   88.000000                          138
1                    2.000000                          138
2                   10.000000                          138
3                   34.722222                          138
4                   34.722222                          138
…                         …                           …
289969              59.000000                         1237
289970              31.000000                         1237
289971              34.722222                         1237
289972              34.722222                         1237
289973              34.722222                         1237

        Survey Response Rate Percent Start Date    End Date
0                                 31   1/1/2022  12/31/2022
1                                 31   1/1/2022  12/31/2022
2                                 31   1/1/2022  12/31/2022
3                                 31   1/1/2022  12/31/2022
4                                 31   1/1/2022  12/31/2022
…                               …         …         …
289969                             9   1/1/2022  12/31/2022
289970                             9   1/1/2022  12/31/2022
289971                             9   1/1/2022  12/31/2022
289972                             9   1/1/2022  12/31/2022
289973                             9   1/1/2022  12/31/2022

[274722 rows x 20 columns]
```

[14]:
```
#Convert Categorical Data to Consistent Codes
# Map categorical variables, like HCAHPS Measure ID and Bed Category, to␣
 ↪integer codes.
```

[15]:
```
le = LabelEncoder()
# Applying Label Encoding to the 'HCAHPS Measure ID' column using .loc to avoid␣
 ↪SettingWithCopyWarning
merged_data.loc[:, 'HCAHPS Measure ID'] = le.fit_transform(merged_data['HCAHPS␣
 ↪Measure ID'])
```

[16]:
```
# Fill Missing Values with More Sophisticated Methods
#Purpose: Improve the accuracy of missing value imputation.
#Steps: Use regression or k-Nearest Neighbors (KNN) #imputation for fields␣
 ↪where relationships exist with other features.
```

[17]:
```
merged_data = merged_data.drop(columns=['ZIP Code', 'Telephone Number'])
```

[18]:

```
merged_data
```

[18]:

```
       Facility ID                    Hospital Name Fiscal Year Begin Date  \
0           511320         JACKSON GENERAL HOSPITAL              10/1/2020
1           511320         JACKSON GENERAL HOSPITAL              10/1/2020
2           511320         JACKSON GENERAL HOSPITAL              10/1/2020
3           511320         JACKSON GENERAL HOSPITAL              10/1/2020
4           511320         JACKSON GENERAL HOSPITAL              10/1/2020
…            …                              …                    …
289969      450869  DOCTORS HOSPITAL AT RENAISSANCE               1/1/2021
289970      450869  DOCTORS HOSPITAL AT RENAISSANCE               1/1/2021
289971      450869  DOCTORS HOSPITAL AT RENAISSANCE               1/1/2021
289972      450869  DOCTORS HOSPITAL AT RENAISSANCE               1/1/2021
289973      450869  DOCTORS HOSPITAL AT RENAISSANCE               1/1/2021

       Fiscal Year End Date  number_of_beds                    Facility Name  \
0                12/31/2020            25.0         Jackson General Hospital
1                12/31/2020            25.0         Jackson General Hospital
2                12/31/2020            25.0         Jackson General Hospital
3                12/31/2020            25.0         Jackson General Hospital
4                12/31/2020            25.0         Jackson General Hospital
…                     …             …                             …
289969           12/31/2021           519.0  Doctors Hospital At Renaissance
289970           12/31/2021           519.0  Doctors Hospital At Renaissance
289971           12/31/2021           519.0  Doctors Hospital At Renaissance
289972           12/31/2021           519.0  Doctors Hospital At Renaissance
289973           12/31/2021           519.0  Doctors Hospital At Renaissance

                  Address City/Town State County/Parish HCAHPS Measure ID  \
0          122 PINNELL ST    Ripley    WV       JACKSON                11
1          122 PINNELL ST    Ripley    WV       JACKSON                13
2          122 PINNELL ST    Ripley    WV       JACKSON                15
3          122 PINNELL ST    Ripley    WV       JACKSON                12
4          122 PINNELL ST    Ripley    WV       JACKSON                14
…                   …       …   …           …               …
289969  5501 SOUTH MCCOLL  Edinburg    TX       HIDALGO                83
289970  5501 SOUTH MCCOLL  Edinburg    TX       HIDALGO                85
289971  5501 SOUTH MCCOLL  Edinburg    TX       HIDALGO                84
289972  5501 SOUTH MCCOLL  Edinburg    TX       HIDALGO                86
289973  5501 SOUTH MCCOLL  Edinburg    TX       HIDALGO                90

                                          HCAHPS Question  \
0       Patients who reported that their nurses "Alway…
1       Patients who reported that their nurses "Somet…
2       Patients who reported that their nurses "Usual…
3                 Nurse communication - linear mean score
4                       Nurse communication - star rating
```

```
…                                                   …
289969  Patients who reported YES, they would definite…
289970  Patients who reported YES, they would probably…
289971             Recommend hospital - linear mean score
289972                   Recommend hospital - star rating
289973                                Summary star rating

                                HCAHPS Answer Description  \
0                       Nurses "always" communicated well
1         Nurses "sometimes" or "never" communicated well
2                      Nurses "usually" communicated well
3                 Nurse communication - linear mean score
4                       Nurse communication - star rating
…                                                   …
289969  "YES", patients would definitely recommend the…
289970  "YES", patients would probably recommend the h…
289971             Recommend hospital - linear mean score
289972                   Recommend hospital - star rating
289973                                Summary star rating

        HCAHPS Answer Percent  Number of Completed Surveys  \
0                   88.000000                          138
1                    2.000000                          138
2                   10.000000                          138
3                   34.722222                          138
4                   34.722222                          138
…                       …                          …
289969              59.000000                         1237
289970              31.000000                         1237
289971              34.722222                         1237
289972              34.722222                         1237
289973              34.722222                         1237

        Survey Response Rate Percent Start Date    End Date
0                                 31   1/1/2022  12/31/2022
1                                 31   1/1/2022  12/31/2022
2                                 31   1/1/2022  12/31/2022
3                                 31   1/1/2022  12/31/2022
4                                 31   1/1/2022  12/31/2022
…                              …        …         …
289969                             9   1/1/2022  12/31/2022
289970                             9   1/1/2022  12/31/2022
289971                             9   1/1/2022  12/31/2022
289972                             9   1/1/2022  12/31/2022
289973                             9   1/1/2022  12/31/2022

[274722 rows x 18 columns]
```

[19]:

```
#Standardize Date Format
```

[20]:

```
merged_data['Fiscal Year Begin Date'] = pd.to_datetime(merged_data['Fiscal Year␣
  ↪Begin Date'], errors='coerce')
merged_data['Fiscal Year End Date'] = pd.to_datetime(merged_data['Fiscal Year␣
  ↪End Date'], errors='coerce')
```

[21]:

```
merged_data
```

[21]:

```
       Facility ID                   Hospital Name Fiscal Year Begin Date  \
0           511320         JACKSON GENERAL HOSPITAL             2020-10-01
1           511320         JACKSON GENERAL HOSPITAL             2020-10-01
2           511320         JACKSON GENERAL HOSPITAL             2020-10-01
3           511320         JACKSON GENERAL HOSPITAL             2020-10-01
4           511320         JACKSON GENERAL HOSPITAL             2020-10-01
…            …                              …                    …
289969      450869  DOCTORS HOSPITAL AT RENAISSANCE             2021-01-01
289970      450869  DOCTORS HOSPITAL AT RENAISSANCE             2021-01-01
289971      450869  DOCTORS HOSPITAL AT RENAISSANCE             2021-01-01
289972      450869  DOCTORS HOSPITAL AT RENAISSANCE             2021-01-01
289973      450869  DOCTORS HOSPITAL AT RENAISSANCE             2021-01-01

       Fiscal Year End Date  number_of_beds                    Facility Name  \
0                2020-12-31            25.0         Jackson General Hospital
1                2020-12-31            25.0         Jackson General Hospital
2                2020-12-31            25.0         Jackson General Hospital
3                2020-12-31            25.0         Jackson General Hospital
4                2020-12-31            25.0         Jackson General Hospital
…                     …             …                              …
289969           2021-12-31           519.0  Doctors Hospital At Renaissance
289970           2021-12-31           519.0  Doctors Hospital At Renaissance
289971           2021-12-31           519.0  Doctors Hospital At Renaissance
289972           2021-12-31           519.0  Doctors Hospital At Renaissance
289973           2021-12-31           519.0  Doctors Hospital At Renaissance

                  Address City/Town State County/Parish HCAHPS Measure ID  \
0          122 PINNELL ST    Ripley    WV       JACKSON                11
1          122 PINNELL ST    Ripley    WV       JACKSON                13
2          122 PINNELL ST    Ripley    WV       JACKSON                15
3          122 PINNELL ST    Ripley    WV       JACKSON                12
4          122 PINNELL ST    Ripley    WV       JACKSON                14
…                   …       …   …           …               …
289969  5501 SOUTH MCCOLL  Edinburg    TX       HIDALGO                83
289970  5501 SOUTH MCCOLL  Edinburg    TX       HIDALGO                85
289971  5501 SOUTH MCCOLL  Edinburg    TX       HIDALGO                84
289972  5501 SOUTH MCCOLL  Edinburg    TX       HIDALGO                86
289973  5501 SOUTH MCCOLL  Edinburg    TX       HIDALGO                90
```

```
                                          HCAHPS Question  \
0       Patients who reported that their nurses "Alway…
1       Patients who reported that their nurses "Somet…
2       Patients who reported that their nurses "Usual…
3                 Nurse communication - linear mean score
4                       Nurse communication - star rating
…                                                     …
289969  Patients who reported YES, they would definite…
289970  Patients who reported YES, they would probably…
289971             Recommend hospital - linear mean score
289972                   Recommend hospital - star rating
289973                                Summary star rating

                                HCAHPS Answer Description  \
0                       Nurses "always" communicated well
1         Nurses "sometimes" or "never" communicated well
2                      Nurses "usually" communicated well
3                 Nurse communication - linear mean score
4                       Nurse communication - star rating
…                                                     …
289969  "YES", patients would definitely recommend the…
289970  "YES", patients would probably recommend the h…
289971             Recommend hospital - linear mean score
289972                   Recommend hospital - star rating
289973                                Summary star rating

        HCAHPS Answer Percent  Number of Completed Surveys  \
0                   88.000000                          138
1                    2.000000                          138
2                   10.000000                          138
3                   34.722222                          138
4                   34.722222                          138
…                         …                         …
289969              59.000000                         1237
289970              31.000000                         1237
289971              34.722222                         1237
289972              34.722222                         1237
289973              34.722222                         1237

        Survey Response Rate Percent Start Date    End Date
0                                 31   1/1/2022  12/31/2022
1                                 31   1/1/2022  12/31/2022
2                                 31   1/1/2022  12/31/2022
3                                 31   1/1/2022  12/31/2022
4                                 31   1/1/2022  12/31/2022
…                              …        …          …
```

```
289969                                9   1/1/2022  12/31/2022
289970                                9   1/1/2022  12/31/2022
289971                                9   1/1/2022  12/31/2022
289972                                9   1/1/2022  12/31/2022
289973                                9   1/1/2022  12/31/2022

[274722 rows x 18 columns]
```

[22]:

```
#Remove Duplicates
```

[23]:

```
merged_data.drop_duplicates(subset=['Facility ID', 'HCAHPS Measure ID'],␣
 ↪inplace=True)
```

[24]:

```
#Handle Highly Skewed Data
merged_data['number_of_beds_log'] = np.log1p(merged_data['number_of_beds'])
```

[25]:

```
#Detect and Correct Anomalies in Numerical Columns
merged_data = merged_data[(merged_data['number_of_beds'] > 0) &␣
 ↪(merged_data['number_of_beds'] < 5000)]
```

[26]:

```
#\Impute Missing Categories with "Unknown"
```

[27]:

```
merged_data['County/Parish'].fillna('Unknown', inplace=True)
```

[28]:

```
#Normalize Numerical Variables
```

[29]:

```
from sklearn.preprocessing import MinMaxScaler
scaler = MinMaxScaler()
merged_data[['number_of_beds', 'HCAHPS Answer Percent']] = scaler.
 ↪fit_transform(merged_data[['number_of_beds', 'HCAHPS Answer Percent']])
```

[30]:

```
# Feature Engineering
# 1. Bed Category
merged_data['Bed Category'] = pd.cut(
    merged_data['number_of_beds'],
    bins=[0, 50, 150, 300, 1000, 3000],
    labels=['Small (0-50)', 'Medium (51-150)', 'Large (151-300)', 'X-Large␣
 ↪(301-1000)', 'Mega (1001+)']
)
```

[31]:

```
merged_data
```

[31]:

```
       Facility ID                      Hospital Name Fiscal Year Begin Date  \
0           511320        JACKSON GENERAL HOSPITAL               2020-10-01
1           511320        JACKSON GENERAL HOSPITAL               2020-10-01
2           511320        JACKSON GENERAL HOSPITAL               2020-10-01
3           511320        JACKSON GENERAL HOSPITAL               2020-10-01
```

```
4           511320         JACKSON GENERAL HOSPITAL             2020-10-01
…            …                              …                    …
289969      450869  DOCTORS HOSPITAL AT RENAISSANCE             2021-01-01
289970      450869  DOCTORS HOSPITAL AT RENAISSANCE             2021-01-01
289971      450869  DOCTORS HOSPITAL AT RENAISSANCE             2021-01-01
289972      450869  DOCTORS HOSPITAL AT RENAISSANCE             2021-01-01
289973      450869  DOCTORS HOSPITAL AT RENAISSANCE             2021-01-01

       Fiscal Year End Date  number_of_beds                    Facility Name  \
0                2020-12-31        0.026154         Jackson General Hospital
1                2020-12-31        0.026154         Jackson General Hospital
2                2020-12-31        0.026154         Jackson General Hospital
3                2020-12-31        0.026154         Jackson General Hospital
4                2020-12-31        0.026154         Jackson General Hospital
…                     …              …                              …
289969           2021-12-31        0.786154  Doctors Hospital At Renaissance
289970           2021-12-31        0.786154  Doctors Hospital At Renaissance
289971           2021-12-31        0.786154  Doctors Hospital At Renaissance
289972           2021-12-31        0.786154  Doctors Hospital At Renaissance
289973           2021-12-31        0.786154  Doctors Hospital At Renaissance

                  Address City/Town State County/Parish HCAHPS Measure ID  \
0          122 PINNELL ST    Ripley    WV       JACKSON                11
1          122 PINNELL ST    Ripley    WV       JACKSON                13
2          122 PINNELL ST    Ripley    WV       JACKSON                15
3          122 PINNELL ST    Ripley    WV       JACKSON                12
4          122 PINNELL ST    Ripley    WV       JACKSON                14
…                    …       …   …           …               …
289969  5501 SOUTH MCCOLL  Edinburg    TX       HIDALGO                83
289970  5501 SOUTH MCCOLL  Edinburg    TX       HIDALGO                85
289971  5501 SOUTH MCCOLL  Edinburg    TX       HIDALGO                84
289972  5501 SOUTH MCCOLL  Edinburg    TX       HIDALGO                86
289973  5501 SOUTH MCCOLL  Edinburg    TX       HIDALGO                90

                                          HCAHPS Question  \
0       Patients who reported that their nurses "Alway…
1       Patients who reported that their nurses "Somet…
2       Patients who reported that their nurses "Usual…
3                 Nurse communication - linear mean score
4                       Nurse communication - star rating
…                                                   …
289969  Patients who reported YES, they would definite…
289970  Patients who reported YES, they would probably…
289971             Recommend hospital - linear mean score
289972                   Recommend hospital - star rating
289973                                Summary star rating
```

```
                                HCAHPS Answer Description  \
0                       Nurses "always" communicated well   
1         Nurses "sometimes" or "never" communicated well   
2                      Nurses "usually" communicated well   
3                 Nurse communication - linear mean score   
4                       Nurse communication - star rating   
…                                                    …   
289969  "YES", patients would definitely recommend the…   
289970  "YES", patients would probably recommend the h…   
289971             Recommend hospital - linear mean score   
289972                   Recommend hospital - star rating   
289973                                Summary star rating   

        HCAHPS Answer Percent  Number of Completed Surveys  \
0                    0.880000                          138   
1                    0.020000                          138   
2                    0.100000                          138   
3                    0.347222                          138   
4                    0.347222                          138   
…                       …                           …   
289969               0.590000                         1237   
289970               0.310000                         1237   
289971               0.347222                         1237   
289972               0.347222                         1237   
289973               0.347222                         1237   

        Survey Response Rate Percent Start Date    End Date  \
0                                 31   1/1/2022  12/31/2022   
1                                 31   1/1/2022  12/31/2022   
2                                 31   1/1/2022  12/31/2022   
3                                 31   1/1/2022  12/31/2022   
4                                 31   1/1/2022  12/31/2022   
…                              …        …         …   
289969                             9   1/1/2022  12/31/2022   
289970                             9   1/1/2022  12/31/2022   
289971                             9   1/1/2022  12/31/2022   
289972                             9   1/1/2022  12/31/2022   
289973                             9   1/1/2022  12/31/2022   

        number_of_beds_log  Bed Category  
0                 3.258097  Small (0-50)  
1                 3.258097  Small (0-50)  
2                 3.258097  Small (0-50)  
3                 3.258097  Small (0-50)  
4                 3.258097  Small (0-50)  
…                    …            …  
289969            6.253829  Small (0-50)  
```

```
289970           6.253829  Small (0-50)
289971           6.253829  Small (0-50)
289972           6.253829  Small (0-50)
289973           6.253829  Small (0-50)

[271467 rows x 20 columns]
```

[32]:

```
# 2. Interaction Terms - Example of interaction with nurse communication
# Assuming 'H_Nurse_Comm' column exists; adjust based on actual column names
if 'H_Nurse_Comm' in merged_data.columns:
    merged_data['Beds_Nurse_Comm'] = merged_data['number_of_beds'] *␣
 ↪merged_data['H_Nurse_Comm']
```

[33]:

```
# 3. Regional Feature - Approximate region based on state
regions = {
    'Northeast': ['NY', 'NJ', 'PA', 'MA', 'CT', 'RI', 'VT', 'NH', 'ME'],
    'Midwest': ['IL', 'IN', 'OH', 'MI', 'WI', 'MN', 'IA', 'MO', 'ND', 'SD',␣
 ↪'NE', 'KS'],
    'South': ['TX', 'FL', 'GA', 'NC', 'SC', 'VA', 'AL', 'TN', 'KY', 'MS', 'LA',␣
 ↪'AR', 'OK', 'WV'],
    'West': ['CA', 'WA', 'OR', 'NV', 'AZ', 'UT', 'CO', 'ID', 'MT', 'WY', 'NM',␣
 ↪'AK', 'HI']
}
merged_data['Region'] = merged_data['State'].map(
    lambda x: next((region for region, states in regions.items() if x in␣
 ↪states), 'Other')
)
```

[34]:

```
merged_data
```

[34]:

```
       Facility ID                    Hospital Name Fiscal Year Begin Date  \
0           511320          JACKSON GENERAL HOSPITAL             2020-10-01
1           511320          JACKSON GENERAL HOSPITAL             2020-10-01
2           511320          JACKSON GENERAL HOSPITAL             2020-10-01
3           511320          JACKSON GENERAL HOSPITAL             2020-10-01
4           511320          JACKSON GENERAL HOSPITAL             2020-10-01
…              …                                …                   …
289969      450869  DOCTORS HOSPITAL AT RENAISSANCE             2021-01-01
289970      450869  DOCTORS HOSPITAL AT RENAISSANCE             2021-01-01
289971      450869  DOCTORS HOSPITAL AT RENAISSANCE             2021-01-01
289972      450869  DOCTORS HOSPITAL AT RENAISSANCE             2021-01-01
289973      450869  DOCTORS HOSPITAL AT RENAISSANCE             2021-01-01

       Fiscal Year End Date  number_of_beds             Facility Name  \
0                2020-12-31        0.026154  Jackson General Hospital
1                2020-12-31        0.026154  Jackson General Hospital
2                2020-12-31        0.026154  Jackson General Hospital
```

```
3               2020-12-31       0.026154        Jackson General Hospital
4               2020-12-31       0.026154        Jackson General Hospital
…                    …              …                              …
289969          2021-12-31       0.786154  Doctors Hospital At Renaissance
289970          2021-12-31       0.786154  Doctors Hospital At Renaissance
289971          2021-12-31       0.786154  Doctors Hospital At Renaissance
289972          2021-12-31       0.786154  Doctors Hospital At Renaissance
289973          2021-12-31       0.786154  Doctors Hospital At Renaissance

                 Address City/Town State County/Parish  …  \
0          122 PINNELL ST    Ripley    WV       JACKSON  …
1          122 PINNELL ST    Ripley    WV       JACKSON  …
2          122 PINNELL ST    Ripley    WV       JACKSON  …
3          122 PINNELL ST    Ripley    WV       JACKSON  …
4          122 PINNELL ST    Ripley    WV       JACKSON  …
…                     …        …   …            …   …
289969  5501 SOUTH MCCOLL  Edinburg    TX       HIDALGO  …
289970  5501 SOUTH MCCOLL  Edinburg    TX       HIDALGO  …
289971  5501 SOUTH MCCOLL  Edinburg    TX       HIDALGO  …
289972  5501 SOUTH MCCOLL  Edinburg    TX       HIDALGO  …
289973  5501 SOUTH MCCOLL  Edinburg    TX       HIDALGO  …

                                           HCAHPS Question  \
0       Patients who reported that their nurses "Alway…
1       Patients who reported that their nurses "Somet…
2       Patients who reported that their nurses "Usual…
3                 Nurse communication - linear mean score
4                       Nurse communication - star rating
…                                                     …
289969  Patients who reported YES, they would definite…
289970  Patients who reported YES, they would probably…
289971             Recommend hospital - linear mean score
289972                   Recommend hospital - star rating
289973                                Summary star rating

                                 HCAHPS Answer Description  \
0                       Nurses "always" communicated well
1         Nurses "sometimes" or "never" communicated well
2                      Nurses "usually" communicated well
3                 Nurse communication - linear mean score
4                       Nurse communication - star rating
…                                                     …
289969  "YES", patients would definitely recommend the…
289970  "YES", patients would probably recommend the h…
289971             Recommend hospital - linear mean score
289972                   Recommend hospital - star rating
289973                                Summary star rating
```

```
       HCAHPS Answer Percent  Number of Completed Surveys  \
0                   0.880000                          138
1                   0.020000                          138
2                   0.100000                          138
3                   0.347222                          138
4                   0.347222                          138
…                        …                          …
289969              0.590000                         1237
289970              0.310000                         1237
289971              0.347222                         1237
289972              0.347222                         1237
289973              0.347222                         1237

        Survey Response Rate Percent  Start Date    End Date  \
0                                 31    1/1/2022  12/31/2022
1                                 31    1/1/2022  12/31/2022
2                                 31    1/1/2022  12/31/2022
3                                 31    1/1/2022  12/31/2022
4                                 31    1/1/2022  12/31/2022
…                                …         …          …
289969                             9    1/1/2022  12/31/2022
289970                             9    1/1/2022  12/31/2022
289971                             9    1/1/2022  12/31/2022
289972                             9    1/1/2022  12/31/2022
289973                             9    1/1/2022  12/31/2022

       number_of_beds_log  Bed Category Region
0                3.258097  Small (0-50)  South
1                3.258097  Small (0-50)  South
2                3.258097  Small (0-50)  South
3                3.258097  Small (0-50)  South
4                3.258097  Small (0-50)  South
…                     …           …     …
289969           6.253829  Small (0-50)  South
289970           6.253829  Small (0-50)  South
289971           6.253829  Small (0-50)  South
289972           6.253829  Small (0-50)  South
289973           6.253829  Small (0-50)  South

[271467 rows x 21 columns]
```

[35]:

```
# Drop unnecessary columns 'Hospital Name' and 'Facility Name'
merged_data = merged_data.drop(columns=['Hospital Name'])
```

[36]:

```
merged_data
```

[36]:

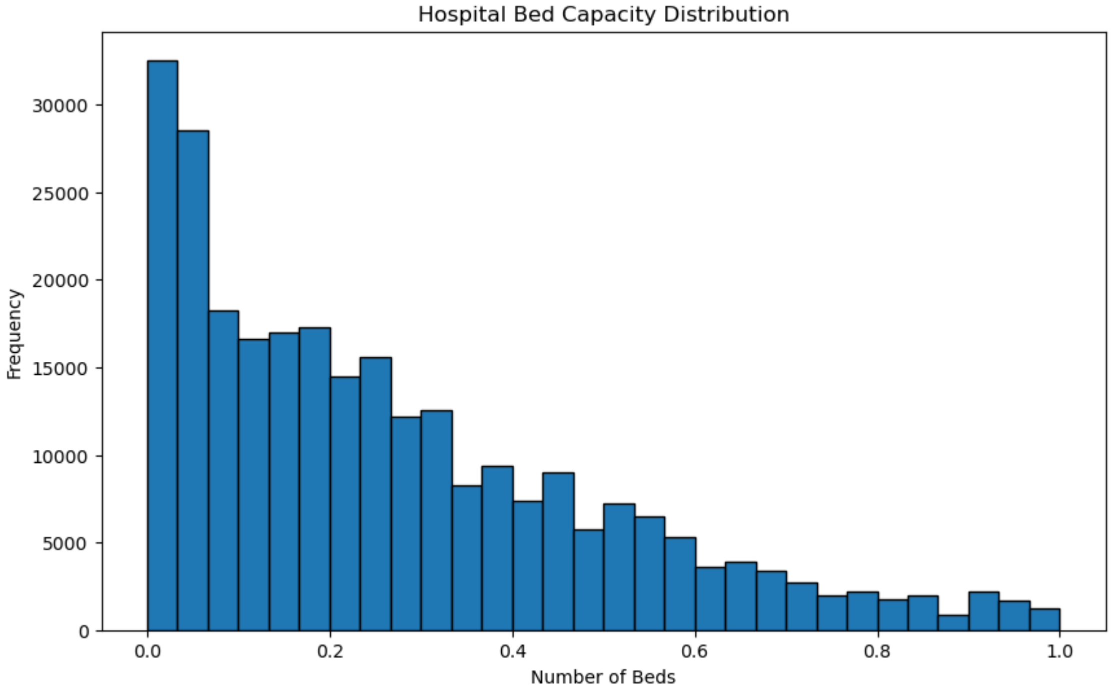
```
       Facility ID Fiscal Year Begin Date Fiscal Year End Date  \
0           511320             2020-10-01           2020-12-31
1           511320             2020-10-01           2020-12-31
2           511320             2020-10-01           2020-12-31
3           511320             2020-10-01           2020-12-31
4           511320             2020-10-01           2020-12-31
…             …                    …                    …
289969      450869             2021-01-01           2021-12-31
289970      450869             2021-01-01           2021-12-31
289971      450869             2021-01-01           2021-12-31
289972      450869             2021-01-01           2021-12-31
289973      450869             2021-01-01           2021-12-31

        number_of_beds                    Facility Name            Address  \
0             0.026154         Jackson General Hospital     122 PINNELL ST
1             0.026154         Jackson General Hospital     122 PINNELL ST
2             0.026154         Jackson General Hospital     122 PINNELL ST
3             0.026154         Jackson General Hospital     122 PINNELL ST
4             0.026154         Jackson General Hospital     122 PINNELL ST
…                    …                              …                …
289969        0.786154  Doctors Hospital At Renaissance  5501 SOUTH MCCOLL
289970        0.786154  Doctors Hospital At Renaissance  5501 SOUTH MCCOLL
289971        0.786154  Doctors Hospital At Renaissance  5501 SOUTH MCCOLL
289972        0.786154  Doctors Hospital At Renaissance  5501 SOUTH MCCOLL
289973        0.786154  Doctors Hospital At Renaissance  5501 SOUTH MCCOLL

       City/Town State County/Parish HCAHPS Measure ID  \
0         Ripley    WV       JACKSON                11
1         Ripley    WV       JACKSON                13
2         Ripley    WV       JACKSON                15
3         Ripley    WV       JACKSON                12
4         Ripley    WV       JACKSON                14
…           …   …            …                …
289969  Edinburg    TX       HIDALGO                83
289970  Edinburg    TX       HIDALGO                85
289971  Edinburg    TX       HIDALGO                84
289972  Edinburg    TX       HIDALGO                86
289973  Edinburg    TX       HIDALGO                90

                                          HCAHPS Question  \
0       Patients who reported that their nurses "Alway…
1       Patients who reported that their nurses "Somet…
2       Patients who reported that their nurses "Usual…
3                 Nurse communication - linear mean score
4                       Nurse communication - star rating
…                                                    …
289969  Patients who reported YES, they would definite…
```

```
289970  Patients who reported YES, they would probably…
289971             Recommend hospital - linear mean score
289972                   Recommend hospital - star rating
289973                                Summary star rating

                                HCAHPS Answer Description  \
0                       Nurses "always" communicated well
1         Nurses "sometimes" or "never" communicated well
2                      Nurses "usually" communicated well
3                 Nurse communication - linear mean score
4                       Nurse communication - star rating
…                                                    …
289969  "YES", patients would definitely recommend the…
289970  "YES", patients would probably recommend the h…
289971             Recommend hospital - linear mean score
289972                   Recommend hospital - star rating
289973                                Summary star rating

        HCAHPS Answer Percent  Number of Completed Surveys  \
0                    0.880000                          138
1                    0.020000                          138
2                    0.100000                          138
3                    0.347222                          138
4                    0.347222                          138
…                       …                           …
289969               0.590000                         1237
289970               0.310000                         1237
289971               0.347222                         1237
289972               0.347222                         1237
289973               0.347222                         1237

        Survey Response Rate Percent Start Date    End Date  \
0                                 31   1/1/2022  12/31/2022
1                                 31   1/1/2022  12/31/2022
2                                 31   1/1/2022  12/31/2022
3                                 31   1/1/2022  12/31/2022
4                                 31   1/1/2022  12/31/2022
…                              …        …         …
289969                             9   1/1/2022  12/31/2022
289970                             9   1/1/2022  12/31/2022
289971                             9   1/1/2022  12/31/2022
289972                             9   1/1/2022  12/31/2022
289973                             9   1/1/2022  12/31/2022

        number_of_beds_log  Bed Category Region
0                 3.258097  Small (0-50)  South
1                 3.258097  Small (0-50)  South
```

```
2                 3.258097  Small (0-50)  South
3                 3.258097  Small (0-50)  South
4                 3.258097  Small (0-50)  South
…                    …           …    …
289969            6.253829  Small (0-50)  South
289970            6.253829  Small (0-50)  South
289971            6.253829  Small (0-50)  South
289972            6.253829  Small (0-50)  South
289973            6.253829  Small (0-50)  South

[271467 rows x 20 columns]
```

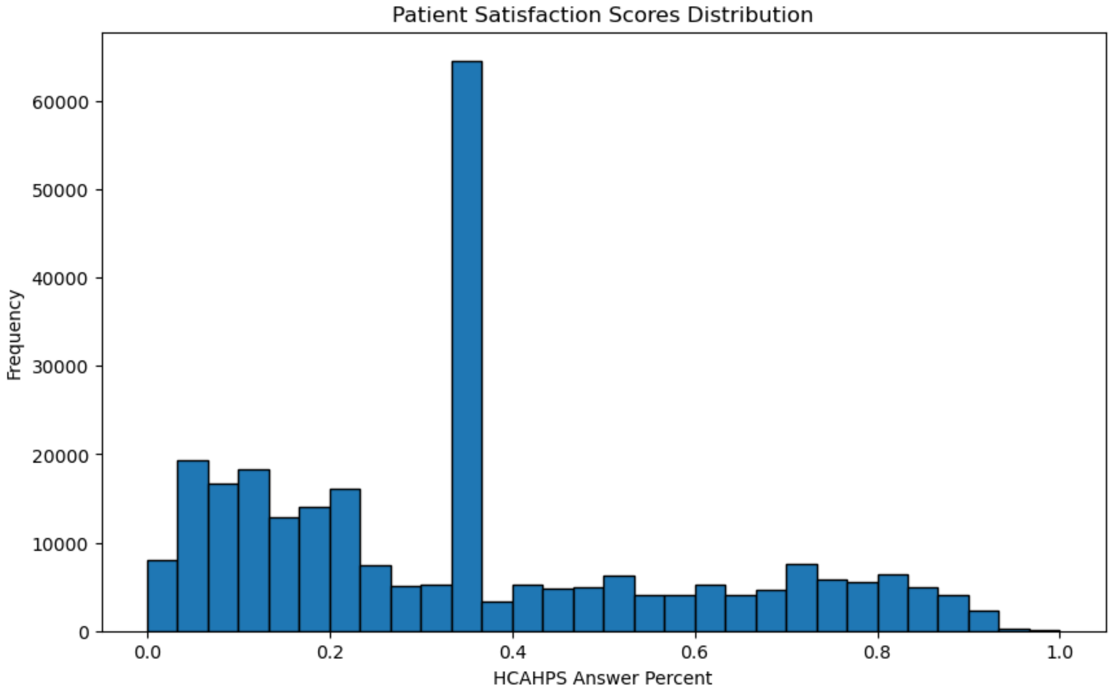
[37]:

```python
#Data Exploration & Visualization

# 1. Hospital Bed Capacity Distribution
plt.figure(figsize=(10, 6))
plt.hist(merged_data['number_of_beds'], bins=30, edgecolor='black')
plt.title('Hospital Bed Capacity Distribution')
plt.xlabel('Number of Beds')
plt.ylabel('Frequency')
plt.show()
```

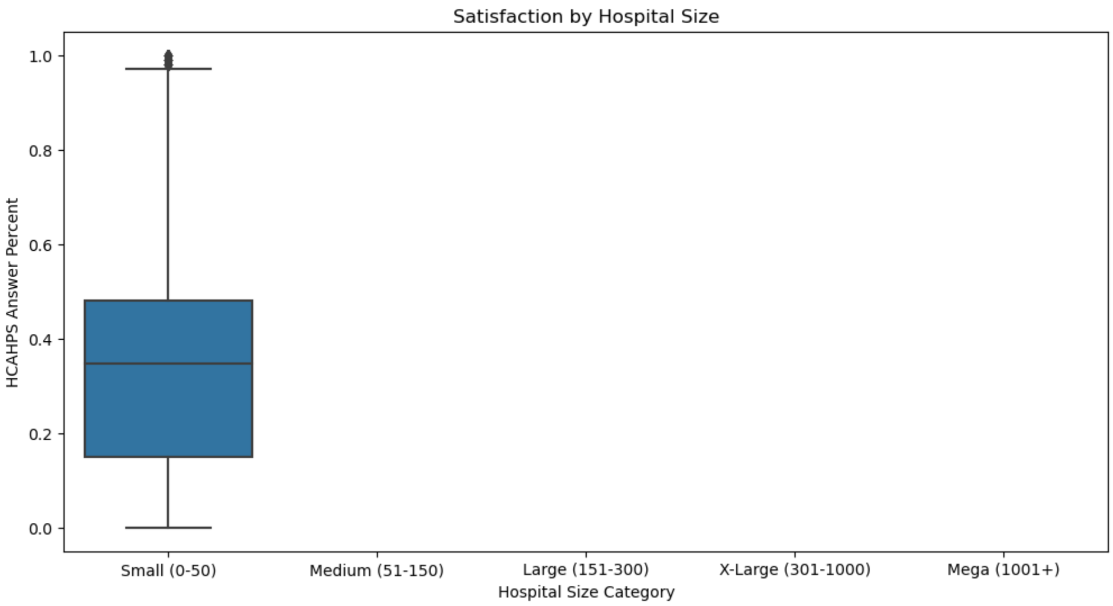


[38]:

```python
# 2. Patient Satisfaction Scores Distribution
plt.figure(figsize=(10, 6))
plt.hist(merged_data['HCAHPS Answer Percent'].dropna(), bins=30,␣
 ↪edgecolor='black')
plt.title('Patient Satisfaction Scores Distribution')
plt.xlabel('HCAHPS Answer Percent')
plt.ylabel('Frequency')
plt.show()
```

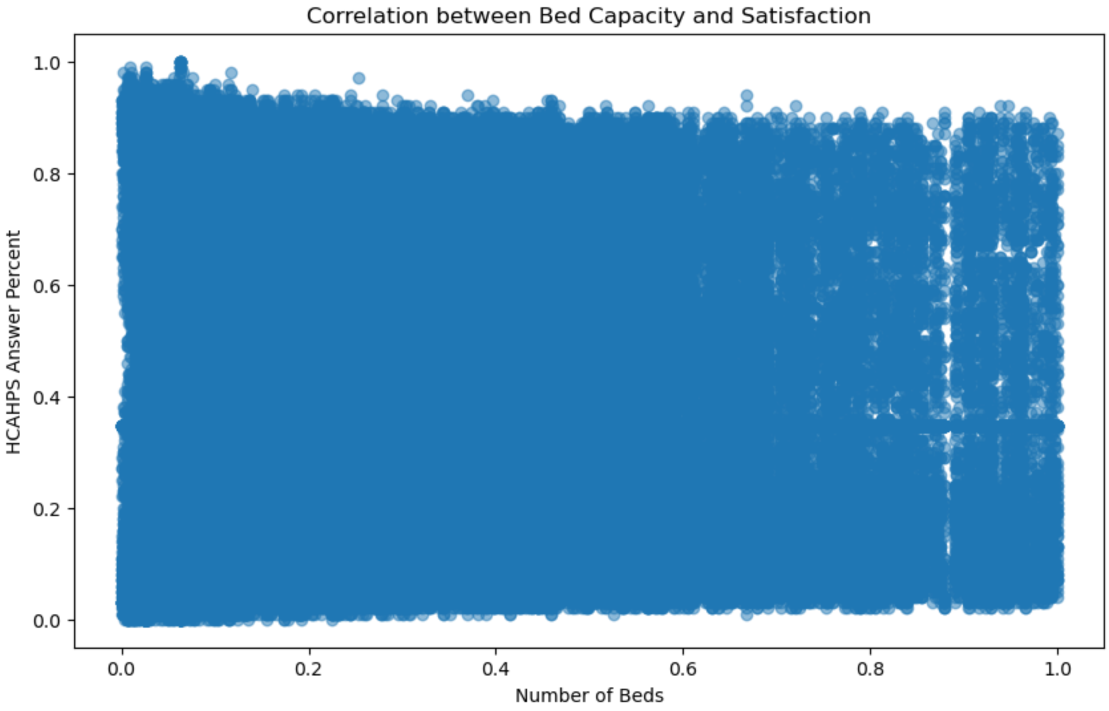


[39]:
```
# 3. Satisfaction by Hospital Size (Box Plot)
plt.figure(figsize=(12, 6))
sns.boxplot(x='Bed Category', y='HCAHPS Answer Percent', data=merged_data)
plt.title('Satisfaction by Hospital Size')
plt.xlabel('Hospital Size Category')
plt.ylabel('HCAHPS Answer Percent')
plt.show()
```

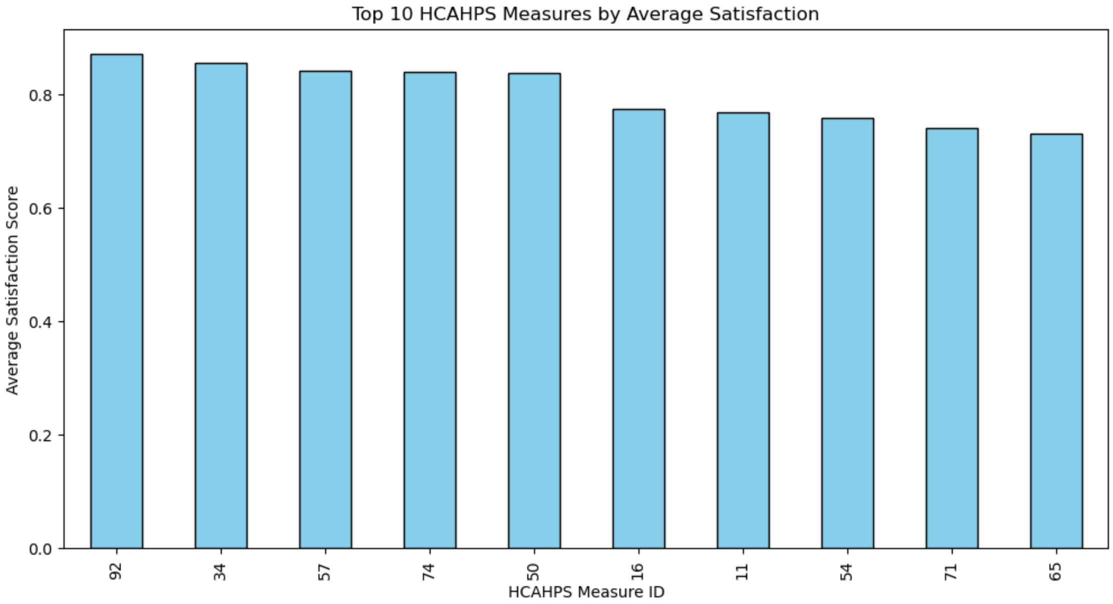


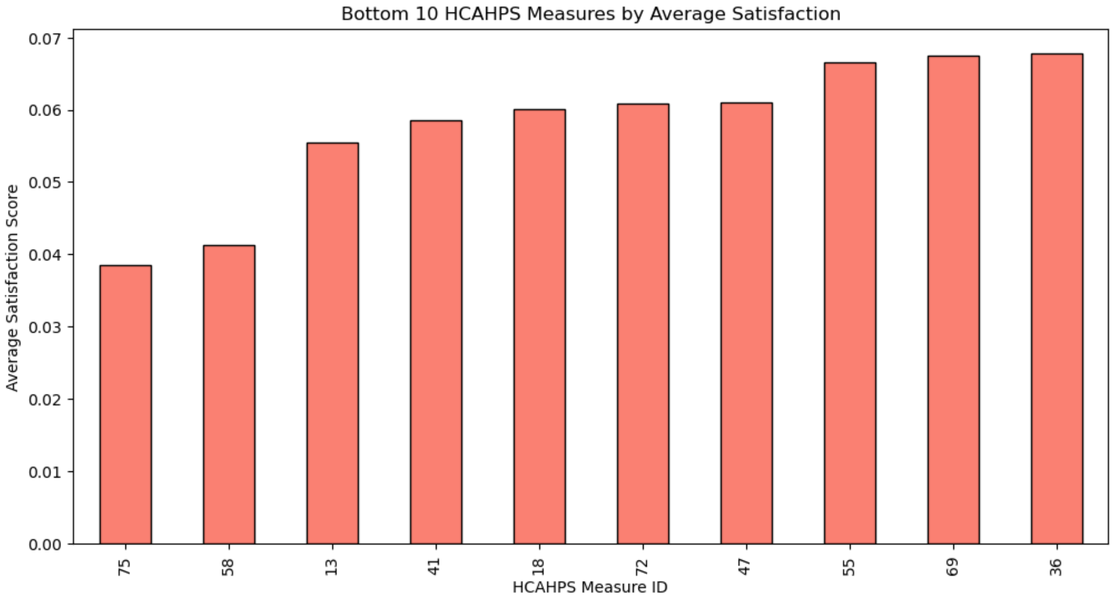
[40]:

```
# 4. Correlation Analysis (Scatter Plot)
plt.figure(figsize=(10, 6))
plt.scatter(merged_data['number_of_beds'], merged_data['HCAHPS Answer␣
  ↪Percent'], alpha=0.5)
plt.title('Correlation between Bed Capacity and Satisfaction')
plt.xlabel('Number of Beds')
plt.ylabel('HCAHPS Answer Percent')
plt.show()
```



```
[41]: # 5. Top & Bottom HCAHPS Measures (Bar Chart)
top_hcahps_measures = merged_data.groupby('HCAHPS Measure ID')['HCAHPS Answer␣
 ↪Percent'].mean().sort_values(ascending=False).head(10)
bottom_hcahps_measures = merged_data.groupby('HCAHPS Measure ID')['HCAHPS␣
 ↪Answer Percent'].mean().sort_values().head(10)

plt.figure(figsize=(12, 6))
top_hcahps_measures.plot(kind='bar', color='skyblue', edgecolor='black')
plt.title('Top 10 HCAHPS Measures by Average Satisfaction')
plt.xlabel('HCAHPS Measure ID')
plt.ylabel('Average Satisfaction Score')
plt.show()

plt.figure(figsize=(12, 6))
bottom_hcahps_measures.plot(kind='bar', color='salmon', edgecolor='black')
plt.title('Bottom 10 HCAHPS Measures by Average Satisfaction')
plt.xlabel('HCAHPS Measure ID')
plt.ylabel('Average Satisfaction Score')
plt.show()
```





[42]:
```
# Machine Learning Model (Random Forest)

# Categorical Encoding
le = LabelEncoder()
merged_data['Bed Category Encoded'] = le.fit_transform(merged_data['Bed Category'])
merged_data['Region Encoded'] = le.fit_transform(merged_data['Region'])
```

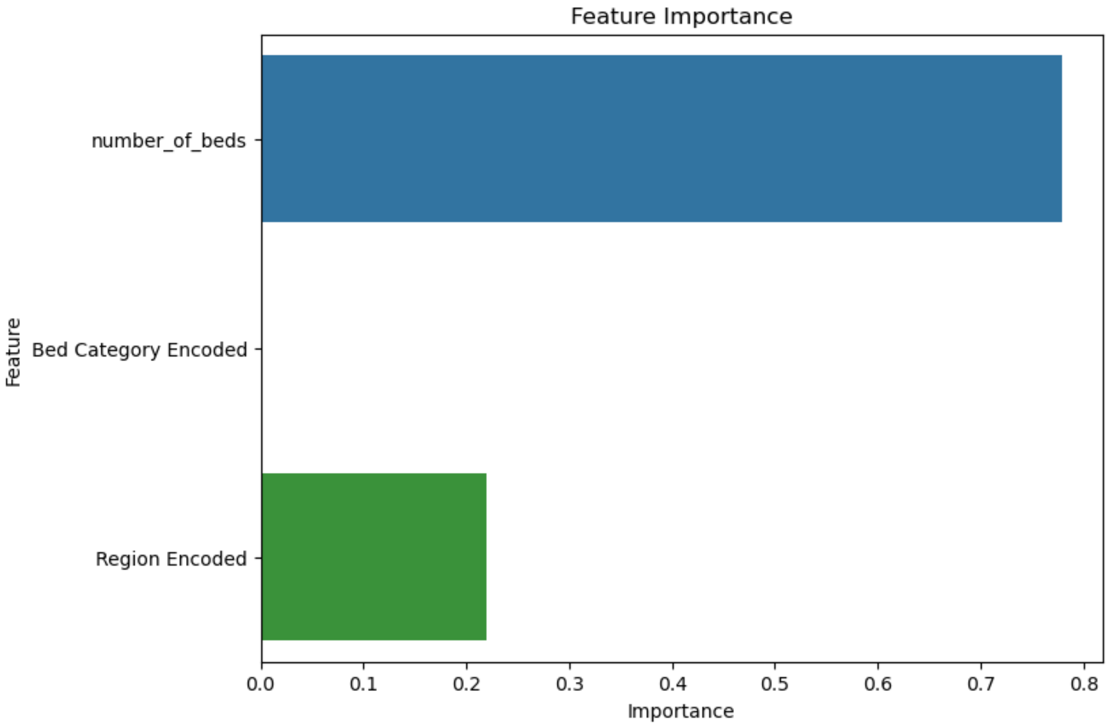
[43]:

```
# Feature Selection
features = merged_data[['number_of_beds', 'Bed Category Encoded', 'Region␣
 ↪Encoded']]
target = merged_data['HCAHPS Answer Percent']

# Split the data
X_train, X_test, y_train, y_test = train_test_split(features, target,␣
 ↪test_size=0.3, random_state=42)

# Train Random Forest model with default parameters
default_rf = RandomForestRegressor(random_state=42)
default_rf.fit(X_train, y_train)

# Predictions and Evaluation
y_pred_rf = default_rf.predict(X_test)
mae = mean_absolute_error(y_test, y_pred_rf)
r2 = r2_score(y_test, y_pred_rf)
mae_rf = mean_absolute_error(y_test, y_pred_rf)
r2_rf = r2_score(y_test, y_pred_rf)

# Display results for SVR
print("Random Forest Results:")
print(f"Mean Absolute Error (MAE): {mae_rf}")
print(f"R-squared (R2): {r2_rf}\n")
```

```
Random Forest Results:
Mean Absolute Error (MAE): 0.1895826267099196
R-squared (R2): -0.012881246198060081
```

[44]:

```
# Feature Importance
plt.figure(figsize=(8, 6))
sns.barplot(x=default_rf.feature_importances_, y=features.columns)
plt.title('Feature Importance')
plt.xlabel('Importance')
plt.ylabel('Feature')
plt.show()
```



[45]:

```
# 2. Linear Regression Model
linear_model = LinearRegression()
linear_model.fit(X_train, y_train)
y_pred_lr = linear_model.predict(X_test)
mae_lr = mean_absolute_error(y_test, y_pred_lr)
r2_lr = r2_score(y_test, y_pred_lr)
# Display results for SVR
print("Linear Regression Results:")
print(f"Mean Absolute Error (MAE): {mae_lr}")
print(f"R-squared (R2): {r2_lr}\n")
```

```
Linear Regression Results:
Mean Absolute Error (MAE): 0.18672883815425434
R-squared (R2): -0.00011557418350482962
```

[46]:

```

!pip install xgboost
!pip install catboost
!pip install lightgbm
from xgboost import XGBRegressor

xgb_model = XGBRegressor(n_estimators=100, learning_rate=0.1, max_depth=3)
xgb_model.fit(X_train, y_train)
```

```
y_pred_xgb = xgb_model.predict(X_test)
```

```
Requirement already satisfied: xgboost in
/srv/conda/envs/notebook/lib/python3.11/site-packages (2.1.2)
Requirement already satisfied: numpy in
/srv/conda/envs/notebook/lib/python3.11/site-packages (from xgboost) (1.24.2)
Requirement already satisfied: nvidia-nccl-cu12 in
/srv/conda/envs/notebook/lib/python3.11/site-packages (from xgboost) (2.23.4)
Requirement already satisfied: scipy in
/srv/conda/envs/notebook/lib/python3.11/site-packages (from xgboost) (1.10.1)
Requirement already satisfied: catboost in
/srv/conda/envs/notebook/lib/python3.11/site-packages (1.2.7)
Requirement already satisfied: graphviz in
/srv/conda/envs/notebook/lib/python3.11/site-packages (from catboost) (0.20.3)
Requirement already satisfied: matplotlib in
/srv/conda/envs/notebook/lib/python3.11/site-packages (from catboost) (3.7.1)
Requirement already satisfied: numpy<2.0,>=1.16.0 in
/srv/conda/envs/notebook/lib/python3.11/site-packages (from catboost) (1.24.2)
Requirement already satisfied: pandas>=0.24 in
/srv/conda/envs/notebook/lib/python3.11/site-packages (from catboost) (2.0.2)
Requirement already satisfied: scipy in
/srv/conda/envs/notebook/lib/python3.11/site-packages (from catboost) (1.10.1)
Requirement already satisfied: plotly in
/srv/conda/envs/notebook/lib/python3.11/site-packages (from catboost) (5.13.1)
Requirement already satisfied: six in
/srv/conda/envs/notebook/lib/python3.11/site-packages (from catboost) (1.16.0)
Requirement already satisfied: python-dateutil>=2.8.2 in
/srv/conda/envs/notebook/lib/python3.11/site-packages (from
pandas>=0.24->catboost) (2.9.0)
Requirement already satisfied: pytz>=2020.1 in
/srv/conda/envs/notebook/lib/python3.11/site-packages (from
pandas>=0.24->catboost) (2024.1)
Requirement already satisfied: tzdata>=2022.1 in
/srv/conda/envs/notebook/lib/python3.11/site-packages (from
pandas>=0.24->catboost) (2024.2)
Requirement already satisfied: contourpy>=1.0.1 in
/srv/conda/envs/notebook/lib/python3.11/site-packages (from
matplotlib->catboost) (1.3.0)
Requirement already satisfied: cycler>=0.10 in
/srv/conda/envs/notebook/lib/python3.11/site-packages (from
matplotlib->catboost) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in
/srv/conda/envs/notebook/lib/python3.11/site-packages (from
matplotlib->catboost) (4.54.1)
Requirement already satisfied: kiwisolver>=1.0.1 in
/srv/conda/envs/notebook/lib/python3.11/site-packages (from
matplotlib->catboost) (1.4.7)
```

```
Requirement already satisfied: packaging>=20.0 in
/srv/conda/envs/notebook/lib/python3.11/site-packages (from
matplotlib->catboost) (24.0)
Requirement already satisfied: pillow>=6.2.0 in
/srv/conda/envs/notebook/lib/python3.11/site-packages (from
matplotlib->catboost) (10.0.1)
Requirement already satisfied: pyparsing>=2.3.1 in
/srv/conda/envs/notebook/lib/python3.11/site-packages (from
matplotlib->catboost) (3.1.4)
Requirement already satisfied: tenacity>=6.2.0 in
/srv/conda/envs/notebook/lib/python3.11/site-packages (from plotly->catboost)
(9.0.0)
Requirement already satisfied: lightgbm in
/srv/conda/envs/notebook/lib/python3.11/site-packages (4.5.0)
Requirement already satisfied: numpy>=1.17.0 in
/srv/conda/envs/notebook/lib/python3.11/site-packages (from lightgbm) (1.24.2)
Requirement already satisfied: scipy in
/srv/conda/envs/notebook/lib/python3.11/site-packages (from lightgbm) (1.10.1)
```

[47]:

```python
# Calculate evaluation metrics
mae_xgb = mean_absolute_error(y_test, y_pred_xgb)
r2_xgb = r2_score(y_test, y_pred_xgb)

# Display results
print("XgBoost Model Results:")
print(f"Mean Absolute Error (MAE): {mae_xgb}")
print(f"R-squared (R2): {r2_xgb}")
```

```
XgBoost Model Results:
Mean Absolute Error (MAE): 0.18705400866521646
R-squared (R2): -0.0010606349277038074
```

[48]:

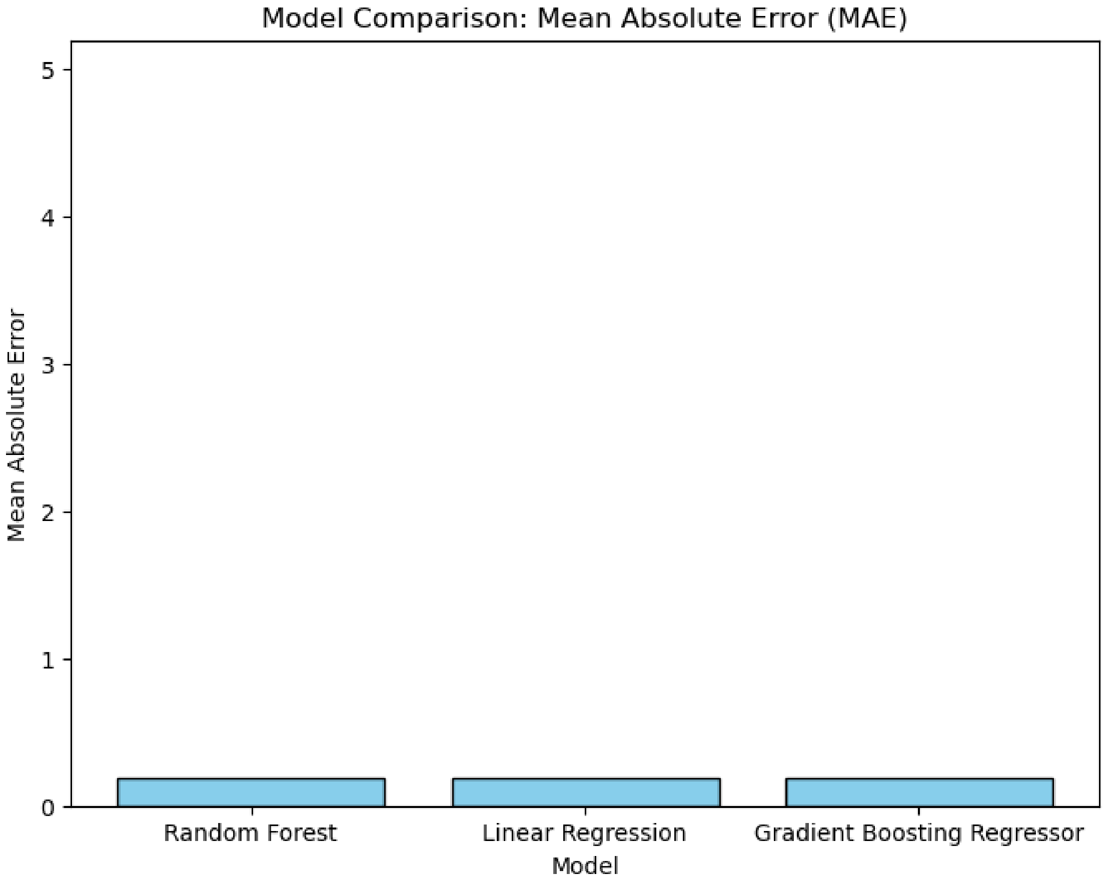
```python
# Random Forest results (already trained in previous code)
print("Random Forest Results:")
print(f"Mean Absolute Error (MAE): {mae_rf}")
print(f"R-squared (R2): {r2_rf}\n")

# Compare model performances
models = {
    "Random Forest": {"MAE": mae_rf, "R2": r2_rf},
    "Linear Regression": {"MAE": mae_lr, "R2": r2_lr},
    "Gradient Boosting Regressor": {"MAE": mae_xgb, "R2": r2_xgb}
}

# Choose the best model based on lowest MAE
best_model = min(models, key=lambda x: models[x]["MAE"])
```

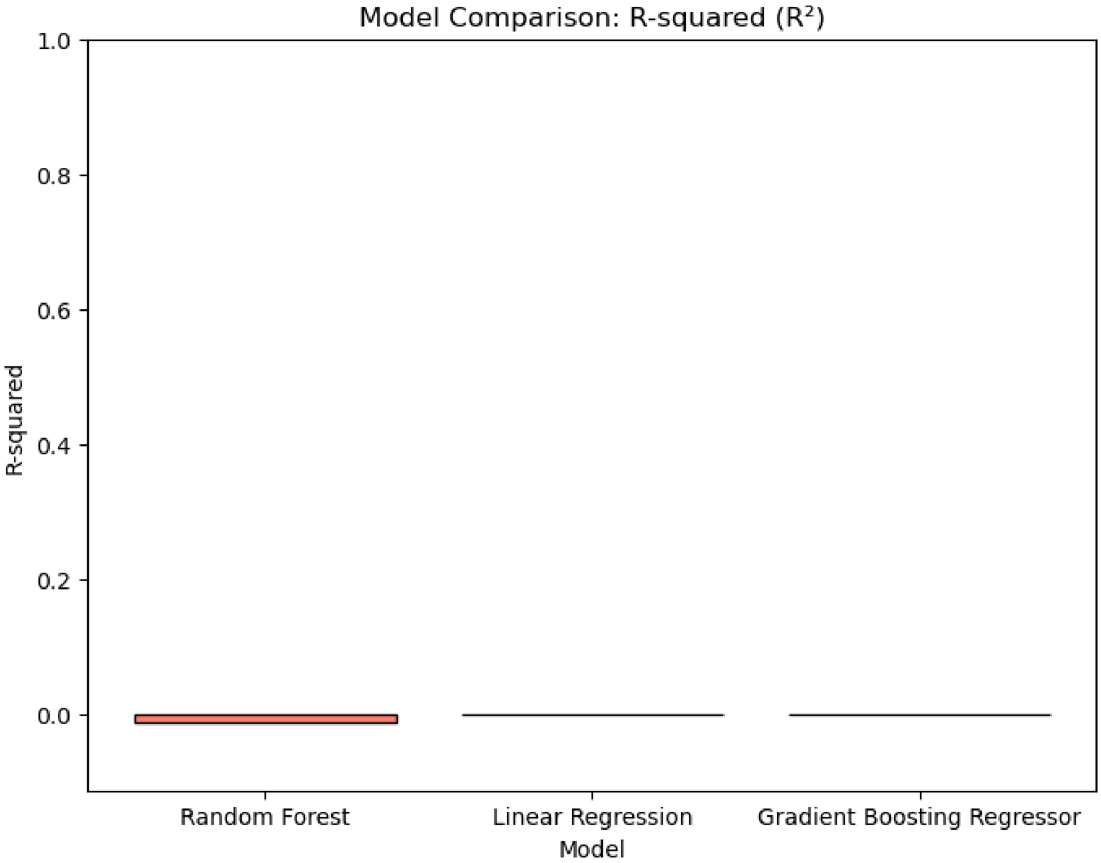
```
# Display comparison
print("Model Performance Comparison:")
for model_name, metrics in models.items():
    print(f"{model_name} - MAE: {metrics['MAE']}, R2: {metrics['R2']}")

print(f"\nBest Model based on MAE: {best_model} with MAE =␣
 ↪{models[best_model]['MAE']} and R2 = {models[best_model]['R2']}")
```

```
Random Forest Results:
Mean Absolute Error (MAE): 0.1895826267099196
R-squared (R2): -0.012881246198060081

Model Performance Comparison:
Random Forest - MAE: 0.1895826267099196, R2: -0.012881246198060081
Linear Regression - MAE: 0.18672883815425434, R2: -0.00011557418350482962
Gradient Boosting Regressor - MAE: 0.18705400866521646, R2:
-0.0010606349277038074

Best Model based on MAE: Linear Regression with MAE = 0.18672883815425434 and R2
= -0.00011557418350482962
```

[49]:

```
# Define the model names and their performance metrics
model_names = list(models.keys())
mae_values = [metrics["MAE"] for metrics in models.values()]
r2_values = [metrics["R2"] for metrics in models.values()]

# Plotting Mean Absolute Error (MAE) for each model
plt.figure(figsize=(8, 6))
plt.bar(model_names, mae_values, color='skyblue', edgecolor='black')
plt.title("Model Comparison: Mean Absolute Error (MAE)")
plt.xlabel("Model")
plt.ylabel("Mean Absolute Error")
plt.ylim(0, max(mae_values) + 5)  # Adjust y-limit for better visibility
plt.show()

# Plotting R-squared (R²) for each model
plt.figure(figsize=(8, 6))
plt.bar(model_names, r2_values, color='salmon', edgecolor='black')
plt.title("Model Comparison: R-squared (R²)")
plt.xlabel("Model")
plt.ylabel("R-squared")
plt.ylim(min(r2_values) - 0.1, 1)  # Adjust y-limit for R²
plt.show()
```


Model Comparison: Mean Absolute Error (MAE)
Mean Absolute Error
0
1
2
3
4
5
Random Forest
Linear Regression
Gradient Boosting Regressor
Model


Model Comparison: R-squared (R²)
R-squared
1.0
0.8
0.6
0.4
0.2
0.0
Random Forest
Linear Regression
Gradient Boosting Regressor
Model

```
[ ]:
```

```
[ ]:
```