Lab 8: Define and Solve an ML Problem of Your Choosing

In [1]:
```
import pandas as pd
import numpy as np
import os
import matplotlib.pyplot as plt
import seaborn as sns
```

In this lab assignment, you will follow the machine learning life cycle and implement a model to solve a machine learning problem of your choosing. You will select a data set and choose a predictive problem that the data set supports. You will then inspect the data with your problem in mind and begin to formulate a project plan. You will then implement the machine learning project plan.

You will complete the following tasks:

1. Build Your DataFrame
2. Define Your ML Problem
3. Perform exploratory data analysis to understand your data.
4. Define Your Project Plan
5. Implement Your Project Plan:
    - Prepare your data for your model.
    - Fit your model to the training data and evaluate your model.
    - Improve your model's performance.

Part 1: Build Your DataFrame

You will have the option to choose one of four data sets that you have worked with in this program:

- The "census" data set that contains Census information from 1994: `censusData.csv`
- Airbnb NYC "listings" data set: `airbnbListingsData.csv`
- World Happiness Report (WHR) data set: `WHR2018Chapter2OnlineData.csv`
- Book Review data set: `bookReviewsData.csv`

Note that these are variations of the data sets that you have worked with in this program. For example, some do not include some of the preprocessing necessary for specific models.

Load a Data Set and Save it as a Pandas DataFrame

The code cell below contains filenames (path + filename) for each of the four data sets available to you.

Task: In the code cell below, use the same method you have been using to load the data using `pd.read_csv()` and save it to DataFrame `df`.

You can load each file as a new DataFrame to inspect the data before choosing your data set.

In [2]:
```
# File names of the four data sets
adultDataSet_filename = os.path.join(os.getcwd(), "data", "censusData.csv")
airbnbDataSet_filename = os.path.join(os.getcwd(), "data", "airbnbListingsData.csv"
WHRDataSet_filename = os.path.join(os.getcwd(), "data", "WHR2018Chapter2OnlineData.
bookReviewDataSet_filename = os.path.join(os.getcwd(), "data", "bookReviewsData.csv


df = pd.read_csv(airbnbDataSet_filename)

df.head()
```

Out[2]:

	name	description	neighborhood_overview	host_name	host_location	host_abo
0	Skylit Midtown Castle	Beautiful, spacious skylit studio in the heart...	Centrally located in the heart of Manhattan ju...	Jennifer	New York, New York, United States	A New York since 2000! M passion creati
1	Whole flr w/private bdrm, bath & kitchen(pls r...	Enjoy 500 s.f. top floor in 1899 brownstone, w...	Just the right mix of urban center and local n...	LisaRoxanne	New York, New York, United States	Laid-ba Native Ne York (formerly b coas
2	Spacious Brooklyn Duplex, Patio + Garden	We welcome you to stay in our lovely 2 br dupl...	NaN	Rebecca	Brooklyn, New York, United States	Rebecca is artist/design and Henoch
3	Large Furnished Room Near B'way	Please don't expect the luxury here just a bas...	Theater district, many restaurants around here.	Shunichi	New York, New York, United States	I used to wo for a financ industry b no
4	Cozy Clean Guest Room - Family Apt	Our best guests are seeking a safe, clean, spa...	Our neighborhood is full of restaurants and ca...	MaryEllen	New York, New York, United States	Welcome family life wi my oldest tw away

5 rows × 50 columns

Part 2: Define Your ML Problem

Next you will formulate your ML Problem. In the markdown cell below, answer the following questions:

1. List the data set you have chosen.
2. What will you be predicting? What is the label?
3. Is this a supervised or unsupervised learning problem? Is this a clustering, classification or regression problem? Is it a binary classificaiton or multi-class classifiction problem?
4. What are your features? (note: this list may change after your explore your data)
5. Explain why this is an important problem. In other words, how would a company create value with a model that predicts this label?

1. Data Set I Have Chosen: I'm using the dataset airbnbListingsData.csv for this project.

2. Prediction Target / Label: The goal here is to predict the price of an Airbnb listing. So, the label I'm predicting is price, which is a continuous numeric value representing how much a listing costs per night.

3. Type of Machine Learning Problem: This is a supervised learning problem because I have data with known prices that I can train the model on. Since price is a continuous number, this is a regression problem.

4. Features (Initial Assumptions): Some features I plan to use as inputs include:

- room_type
- accommodates
- bathrooms
- bedrooms
- beds
- amenities
- minimum_nights
- maximum_nights
- availability_365

5. Why This Problem Matters: Predicting the price of Airbnb listings matters because:

- Hosts can use the predictions to price their listings competitively.

- Accurate price predictions improve customer satisfaction by setting expectations.
- It supports dynamic pricing tools and market analysis for the platform.

Part 3: Understand Your Data

The next step is to perform exploratory data analysis. Inspect and analyze your data set with your machine learning problem in mind. Consider the following as you inspect your data:

1. What data preparation techniques would you like to use? These data preparation techniques may include:
    - addressing missingness, such as replacing missing values with means
    - finding and replacing outliers
    - renaming features and labels
    - finding and replacing outliers
    - performing feature engineering techniques such as one-hot encoding on categorical features
    - selecting appropriate features and removing irrelevant features
    - performing specific data cleaning and preprocessing techniques for an NLP problem
    - addressing class imbalance in your data sample to promote fair AI
2. What machine learning model (or models) you would like to use that is suitable for your predictive problem and data?
    - Are there other data preparation techniques that you will need to apply to build a balanced modeling data set for your problem and model? For example, will you need to scale your data?
3. How will you evaluate and improve the model's performance?
    - Are there specific evaluation metrics and methods that are appropriate for your model?

Think of the different techniques you have used to inspect and analyze your data in this course. These include using Pandas to apply data filters, using the Pandas `describe()` method to get insight into key statistics for each column, using the Pandas `dtypes` property to inspect the data type of each column, and using Matplotlib and Seaborn to detect outliers and visualize relationships between features and labels. If you are working on a classification problem, use techniques you have learned to determine if there is class imbalance.

Task: Use the techniques you have learned in this course to inspect and analyze your data. You can import additional packages that you have used in this course that you will need to perform this task.

Note: You can add code cells if needed by going to the **Insert** menu and clicking on **Insert Cell Below** in the drop-drown menu.

In [3]:
```
df.shape
df.info()
df.head()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 28022 entries, 0 to 28021
Data columns (total 50 columns):
 #   Column                                        Non-Null Count  Dtype
---  ------                                        --------------  -----
 0   name                                          28017 non-null  object
 1   description                                   27452 non-null  object
 2   neighborhood_overview                         18206 non-null  object
 3   host_name                                     28022 non-null  object
 4   host_location                                 27962 non-null  object
 5   host_about                                    17077 non-null  object
 6   host_response_rate                            16179 non-null  float64
 7   host_acceptance_rate                          16909 non-null  float64
 8   host_is_superhost                             28022 non-null  bool
 9   host_listings_count                           28022 non-null  float64
 10  host_total_listings_count                     28022 non-null  float64
 11  host_has_profile_pic                          28022 non-null  bool
 12  host_identity_verified                        28022 non-null  bool
 13  neighbourhood_group_cleansed                  28022 non-null  object
 14  room_type                                     28022 non-null  object
 15  accommodates                                  28022 non-null  int64
 16  bathrooms                                     28022 non-null  float64
 17  bedrooms                                      25104 non-null  float64
 18  beds                                          26668 non-null  float64
 19  amenities                                     28022 non-null  object
 20  price                                         28022 non-null  float64
 21  minimum_nights                                28022 non-null  int64
 22  maximum_nights                                28022 non-null  int64
 23  minimum_minimum_nights                        28022 non-null  float64
 24  maximum_minimum_nights                        28022 non-null  float64
 25  minimum_maximum_nights                        28022 non-null  float64
 26  maximum_maximum_nights                        28022 non-null  float64
 27  minimum_nights_avg_ntm                        28022 non-null  float64
 28  maximum_nights_avg_ntm                        28022 non-null  float64
 29  has_availability                              28022 non-null  bool
 30  availability_30                               28022 non-null  int64
 31  availability_60                               28022 non-null  int64
 32  availability_90                               28022 non-null  int64
 33  availability_365                              28022 non-null  int64
 34  number_of_reviews                             28022 non-null  int64
 35  number_of_reviews_ltm                         28022 non-null  int64
 36  number_of_reviews_l30d                        28022 non-null  int64
 37  review_scores_rating                          28022 non-null  float64
 38  review_scores_cleanliness                     28022 non-null  float64
 39  review_scores_checkin                         28022 non-null  float64
 40  review_scores_communication                   28022 non-null  float64
 41  review_scores_location                        28022 non-null  float64
 42  review_scores_value                           28022 non-null  float64
 43  instant_bookable                              28022 non-null  bool
 44  calculated_host_listings_count                28022 non-null  int64
 45  calculated_host_listings_count_entire_homes   28022 non-null  int64
 46  calculated_host_listings_count_private_rooms  28022 non-null  int64
 47  calculated_host_listings_count_shared_rooms   28022 non-null  int64
 48  reviews_per_month                             28022 non-null  float64
 49  n_host_verifications                          28022 non-null  int64
```

```
dtypes: bool(5), float64(21), int64(15), object(9)
memory usage: 9.8+ MB
```

Out[3]:

	name	description	neighborhood_overview	host_name	host_location	host_abo
0	Skylit Midtown Castle	Beautiful, spacious skylit studio in the heart...	Centrally located in the heart of Manhattan ju...	Jennifer	New York, New York, United States	A New York since 2000! M passion creati
1	Whole flr w/private bdrm, bath & kitchen(pls r...	Enjoy 500 s.f. top floor in 1899 brownstone, w...	Just the right mix of urban center and local n...	LisaRoxanne	New York, New York, United States	Laid-ba Native Ne York (formerly b coas
2	Spacious Brooklyn Duplex, Patio + Garden	We welcome you to stay in our lovely 2 br dupl...	NaN	Rebecca	Brooklyn, New York, United States	Rebecca is artist/design and Henoch
3	Large Furnished Room Near B'way	Please don't expect the luxury here just a bas...	Theater district, many restaurants around here.	Shunichi	New York, New York, United States	I used to wo for a financ industry b no
4	Cozy Clean Guest Room - Family Apt	Our best guests are seeking a safe, clean, spa...	Our neighborhood is full of restaurants and ca...	MaryEllen	New York, New York, United States	Welcome family life wi my oldest tw away

5 rows × 50 columns

In [4]:
```
df = df.drop(columns=[
    'name',
    'description',
    'neighborhood_overview',
    'host_name',
    'host_location',
    'host_about',
    'amenities'
])
```

In [5]:
```
num_cols = df.select_dtypes(include=['float64', 'int64']).columns
df[num_cols] = df[num_cols].fillna(df[num_cols].mean())
```

In [6]:
```
df['host_identity_verified'] = df['host_identity_verified'].fillna(False)
df['host_has_profile_pic'] = df['host_has_profile_pic'].fillna(False)
```

In [7]:
```
df = pd.get_dummies(df, columns=[
    'room_type',
    'neighbourhood_group_cleansed',
], drop_first=True)
```

In [8]:
```
df['host_is_superhost'] = df['host_is_superhost'].astype(int)
```

In [9]:
```
cat_cols = df.select_dtypes(include='object').columns
cat_cols
```

Out[9]:
```
Index([], dtype='object')
```

In [10]:
```
num_features = ['price', 'bathrooms', 'bedrooms', 'beds', 'number_of_reviews']
```

In [11]:
```
df[num_features].describe()
```

Out[11]:

	price	bathrooms	bedrooms	beds	number_of_reviews
count	28022.000000	28022.000000	28022.000000	28022.000000	28022.000000
mean	154.228749	1.142174	1.329708	1.629556	31.589965
std	140.816605	0.421132	0.663238	1.070269	56.299006
min	29.000000	0.000000	1.000000	1.000000	1.000000
25%	70.000000	1.000000	1.000000	1.000000	3.000000
50%	115.000000	1.000000	1.000000	1.000000	9.000000
75%	180.000000	1.000000	1.329708	2.000000	33.000000
max	1000.000000	8.000000	12.000000	21.000000	1009.000000

In [13]:
```
df['price']
```

Out[13]:
```
0         150.0
1          75.0
2         275.0
3          68.0
4          75.0
          ...
28017      89.0
28018    1000.0
28019      64.0
28020      84.0
28021      70.0
Name: price, Length: 28022, dtype: float64
```

Part 4: Define Your Project Plan

Now that you understand your data, in the markdown cell below, define your plan to implement the remaining phases of the machine learning life cycle (data preparation,

modeling, evaluation) to solve your ML problem. Answer the following questions:

- Do you have a new feature list? If so, what are the features that you chose to keep and remove after inspecting the data?
- Explain different data preparation techniques that you will use to prepare your data for modeling.
- What is your model (or models)?
- Describe your plan to train your model, analyze its performance and then improve the model. That is, describe your model building, validation and selection plan to produce a model that generalizes well to new data.

Feature Selection: After exploring the data, I kept features that are relevant and have good data quality, including: accommodates, bathrooms, bedrooms, beds, room_type, host_response_rate, host_identity_verified, host_listings_count, amenities, minimum_nights, maximum_nights, availability_365, review_scores_location.

I removed features with many missing values, low correlation with price, or complex text data.

Data Preparation :

- Fill missing values (median for numbers, mode or "Unknown" for categories).
- Scale numerical features if needed.
- Detect and handle outliers.
- Split data into train/test sets (80/20).

Modeling:

- Main model: Random Forest Regressor for its ability to handle non-linearity and outliers.
- Baseline: Linear Regression.

Training & Evaluation

- Train models with cross-validation and tune hyperparameters (using GridSearchCV).
- Evaluate using MSE and R^2 on validation and test sets.
- Choose the best model balancing performance and simplicity, then test on unseen data.

Part 5: Implement Your Project Plan

Task: In the code cell below, import additional packages that you have used in this course that you will need to implement your project plan.

In [14]:
```
import pandas as pd
import numpy as np
```

```
import os
import matplotlib.pyplot as plt
import seaborn as sns

from sklearn.model_selection import train_test_split, GridSearchCV
from sklearn.linear_model import LinearRegression
from sklearn.metrics import accuracy_score, confusion_matrix, precision_recall_curv
```

Task: Use the rest of this notebook to carry out your project plan.

You will:

1. Prepare your data for your model.
2. Fit your model to the training data and evaluate your model.
3. Improve your model's performance by performing model selection and/or feature selection techniques to find best model for your problem.

Add code cells below and populate the notebook with commentary, code, analyses, results, and figures as you see fit.

In [22]:
```
X = df.drop(columns=['price', 'host_is_superhost'])
y = df['price']

X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.2, random_state=1234
)
```

In [23]:
```
lr_model = LinearRegression()
lr_model.fit(X_train, y_train)
```

Out[23]:
```
▾ LinearRegression
LinearRegression()
```

In [25]:
```
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score

y_pred = lr_model.predict(X_test)

mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)

print(f"Mean Squared Error (MSE): {mse:.2f}")
print(f"R-squared (R2) score: {r2:.2f}")
```

```
Mean Squared Error (MSE): 11089.41
R-squared (R2) score: 0.47
```

In [28]:
```
from sklearn.ensemble import RandomForestRegressor
from sklearn.model_selection import GridSearchCV
```

```
rf = RandomForestRegressor(random_state=1234)

param_grid = {
    'n_estimators': [50, 100, 200],         # number of trees
    'max_depth': [None, 10, 20, 30],        # max depth of each tree
    'min_samples_split': [2, 5, 10],        # min samples required to split a node
    'min_samples_leaf': [1, 2, 4],          # min samples required at leaf node
}
```

In [29]:
```
grid_search = GridSearchCV(estimator=rf, param_grid=param_grid,
                           cv=3, n_jobs=-1, verbose=2, scoring='neg_mean_squared_er
```

In [41]:
```
grid_search.fit(X_train, y_train)
```

```
Fitting 3 folds for each of 108 candidates, totalling 324 fits
```

Out[41]:

```
▸ GridSearchCV
▸ estimator: RandomForestRegressor
  ▸ RandomForestRegressor
```

In [42]:
```
print("Best parameters found:", grid_search.best_params_)
print("Best negative MSE:", grid_search.best_score_)
```

```
Best parameters found: {'max_depth': 20, 'min_samples_leaf': 2, 'min_samples_split':
5, 'n_estimators': 200}
Best negative MSE: -7962.801539215525
```

In [43]:
```
best_rf = grid_search.best_estimator_
y_pred = best_rf.predict(X_test)

from sklearn.metrics import mean_squared_error, r2_score

mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)

print(f"Test MSE: {mse:.2f}")
print(f"Test R2: {r2:.2f}")
```

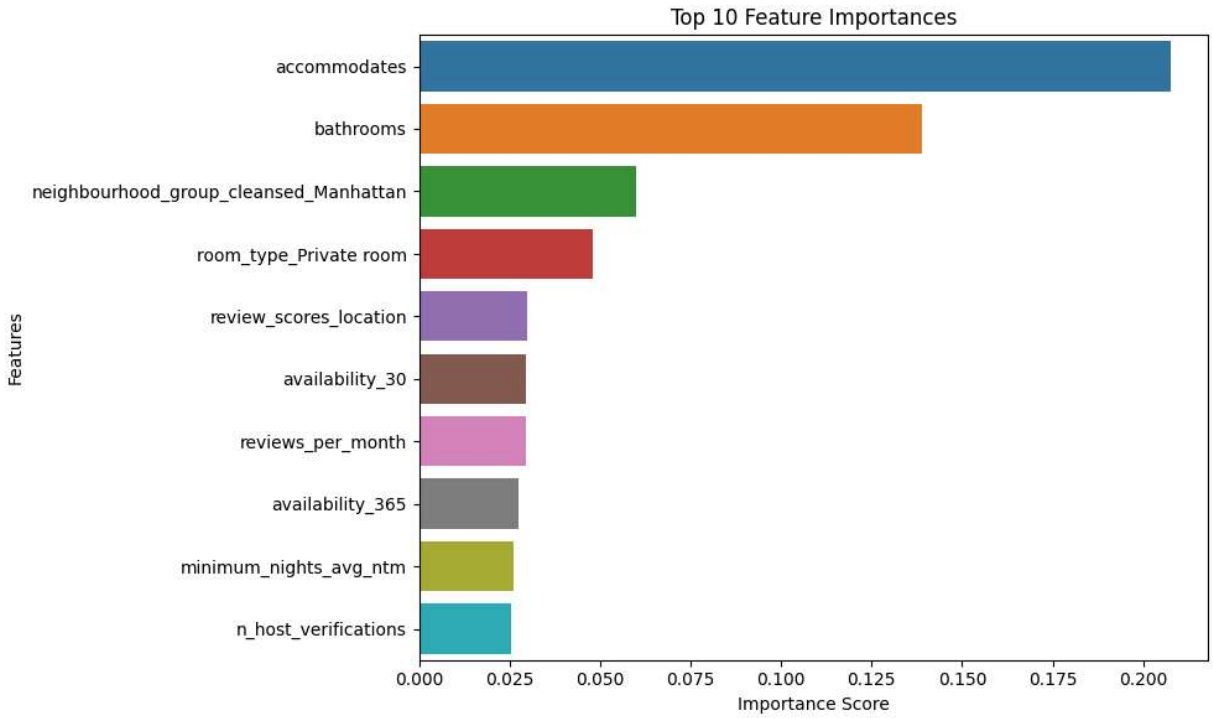
```
Test MSE: 7840.03
Test R2: 0.63
```

In [44]:
```
import matplotlib.pyplot as plt
import seaborn as sns

feature_importances = pd.Series(best_rf.feature_importances_, index=X.columns)
top_features = feature_importances.sort_values(ascending=False).head(10)

plt.figure(figsize=(10, 6))
sns.barplot(x=top_features, y=top_features.index)
plt.title("Top 10 Feature Importances")
plt.xlabel("Importance Score")
plt.ylabel("Features")
```

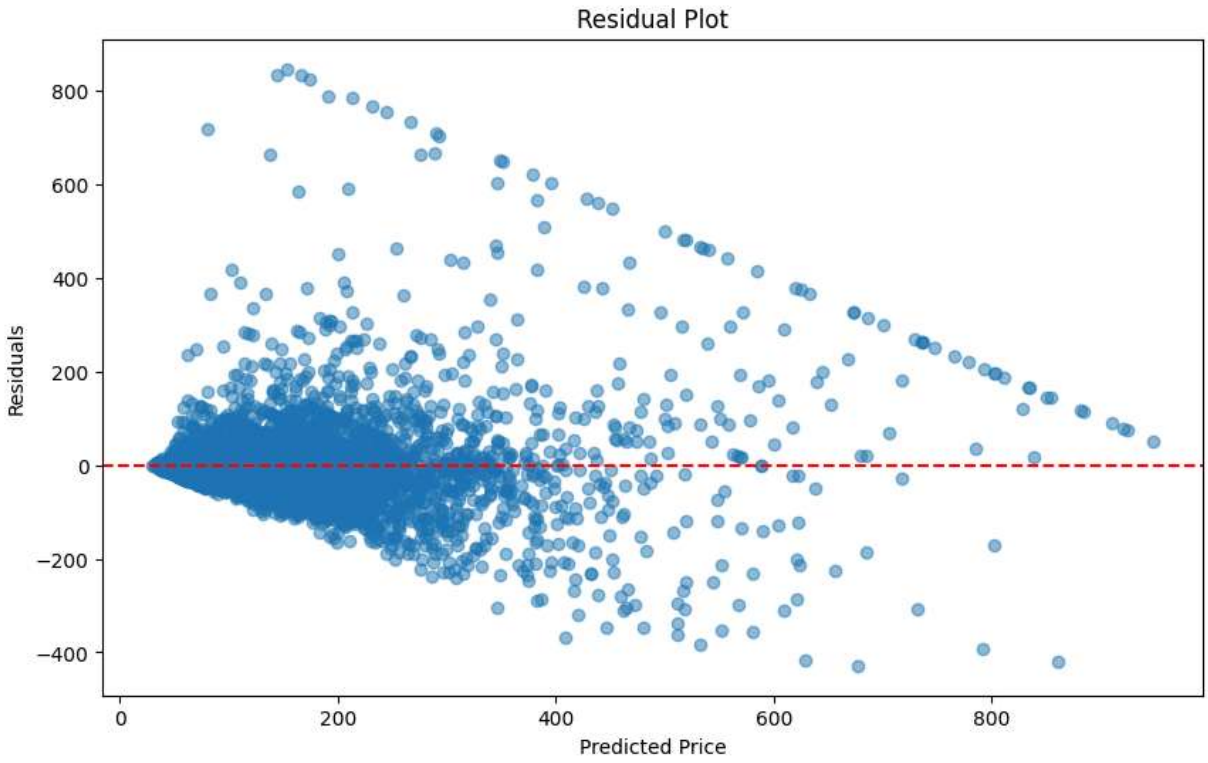
```
plt.tight_layout()
plt.show()
```



In [45]:
```
import matplotlib.pyplot as plt

residuals = y_test - y_pred

plt.figure(figsize=(10,6))
plt.scatter(y_pred, residuals, alpha=0.5)
plt.axhline(0, linestyle='--', color='red')
plt.xlabel("Predicted Price")
plt.ylabel("Residuals")
plt.title("Residual Plot")
plt.show()
```


Residual Plot
Residuals
Predicted Price
800
600
400
200
0
−200
−400
0
200
400
600
800

```
[CV] END max_depth=None, min_samples_leaf=1, min_samples_split=2, n_estimators=50; t
otal time=   9.8s
[CV] END max_depth=None, min_samples_leaf=1, min_samples_split=5, n_estimators=50; t
otal time=   8.8s
[CV] END max_depth=None, min_samples_leaf=1, min_samples_split=5, n_estimators=50; t
otal time=   8.9s
[CV] END max_depth=None, min_samples_leaf=1, min_samples_split=5, n_estimators=200;
total time=  34.9s
[CV] END max_depth=None, min_samples_leaf=2, min_samples_split=2, n_estimators=100;
total time=  16.7s
[CV] END max_depth=None, min_samples_leaf=2, min_samples_split=2, n_estimators=200;
total time=  33.7s
[CV] END max_depth=None, min_samples_leaf=2, min_samples_split=10, n_estimators=200;
total time=  29.9s
[CV] END max_depth=None, min_samples_leaf=4, min_samples_split=5, n_estimators=100;
total time=  14.2s
[CV] END max_depth=None, min_samples_leaf=4, min_samples_split=5, n_estimators=200;
total time=  28.5s
[CV] END max_depth=10, min_samples_leaf=1, min_samples_split=2, n_estimators=200; to
tal time=  19.0s
[CV] END max_depth=10, min_samples_leaf=1, min_samples_split=5, n_estimators=200; to
tal time=  18.8s
[CV] END max_depth=10, min_samples_leaf=2, min_samples_split=2, n_estimators=100; to
tal time=   9.4s
[CV] END max_depth=10, min_samples_leaf=2, min_samples_split=5, n_estimators=50; tot
al time=   4.7s
[CV] END max_depth=10, min_samples_leaf=2, min_samples_split=5, n_estimators=200; to
tal time=  18.9s
[CV] END max_depth=10, min_samples_leaf=4, min_samples_split=2, n_estimators=50; tot
al time=   4.7s
[CV] END max_depth=10, min_samples_leaf=4, min_samples_split=2, n_estimators=100; to
tal time=   9.3s
[CV] END max_depth=10, min_samples_leaf=4, min_samples_split=5, n_estimators=50; tot
al time=   4.7s
[CV] END max_depth=10, min_samples_leaf=4, min_samples_split=5, n_estimators=200; to
tal time=  18.6s
[CV] END max_depth=20, min_samples_leaf=1, min_samples_split=2, n_estimators=50; tot
al time=   8.9s
[CV] END max_depth=20, min_samples_leaf=1, min_samples_split=2, n_estimators=200; to
tal time=  35.5s
[CV] END max_depth=20, min_samples_leaf=1, min_samples_split=10, n_estimators=50; to
tal time=   7.7s
[CV] END max_depth=20, min_samples_leaf=1, min_samples_split=10, n_estimators=200; t
otal time=  30.4s
[CV] END max_depth=20, min_samples_leaf=2, min_samples_split=5, n_estimators=50; tot
al time=   7.8s
[CV] END max_depth=20, min_samples_leaf=2, min_samples_split=5, n_estimators=200; to
tal time=  31.3s
[CV] END max_depth=20, min_samples_leaf=4, min_samples_split=2, n_estimators=50; tot
al time=   7.0s
[CV] END max_depth=20, min_samples_leaf=4, min_samples_split=2, n_estimators=100; to
tal time=  13.7s
[CV] END max_depth=20, min_samples_leaf=4, min_samples_split=5, n_estimators=50; tot
al time=   7.0s
[CV] END max_depth=20, min_samples_leaf=4, min_samples_split=5, n_estimators=200; to
tal time=  27.5s
```

```
[CV] END max_depth=30, min_samples_leaf=1, min_samples_split=2, n_estimators=50; tot
al time=   9.7s
[CV] END max_depth=30, min_samples_leaf=1, min_samples_split=2, n_estimators=200; to
tal time=  39.2s
[CV] END max_depth=30, min_samples_leaf=1, min_samples_split=10, n_estimators=50; to
tal time=   8.0s
[CV] END max_depth=30, min_samples_leaf=1, min_samples_split=10, n_estimators=200; t
otal time=  31.6s
[CV] END max_depth=30, min_samples_leaf=2, min_samples_split=5, n_estimators=50; tot
al time=   8.1s
[CV] END max_depth=30, min_samples_leaf=2, min_samples_split=5, n_estimators=100; to
tal time=  16.4s
[CV] END max_depth=30, min_samples_leaf=2, min_samples_split=10, n_estimators=100; t
otal time=  15.1s
[CV] END max_depth=30, min_samples_leaf=2, min_samples_split=10, n_estimators=200; t
otal time=  30.5s
[CV] END max_depth=30, min_samples_leaf=4, min_samples_split=5, n_estimators=200; to
tal time=  28.0s
[CV] END max_depth=None, min_samples_leaf=1, min_samples_split=2, n_estimators=100;
total time=  19.7s
[CV] END max_depth=None, min_samples_leaf=1, min_samples_split=5, n_estimators=200;
total time=  35.0s
[CV] END max_depth=None, min_samples_leaf=1, min_samples_split=10, n_estimators=200;
total time=  32.2s
[CV] END max_depth=None, min_samples_leaf=2, min_samples_split=5, n_estimators=100;
total time=  16.3s
[CV] END max_depth=None, min_samples_leaf=2, min_samples_split=10, n_estimators=50;
total time=   7.5s
[CV] END max_depth=None, min_samples_leaf=2, min_samples_split=10, n_estimators=100;
total time=  14.9s
[CV] END max_depth=None, min_samples_leaf=4, min_samples_split=2, n_estimators=50; t
otal time=   7.1s
[CV] END max_depth=None, min_samples_leaf=4, min_samples_split=2, n_estimators=200;
total time=  28.3s
[CV] END max_depth=None, min_samples_leaf=4, min_samples_split=10, n_estimators=50;
total time=   6.9s
[CV] END max_depth=None, min_samples_leaf=4, min_samples_split=10, n_estimators=100;
total time=  13.6s
[CV] END max_depth=10, min_samples_leaf=1, min_samples_split=2, n_estimators=50; tot
al time=   4.7s
[CV] END max_depth=10, min_samples_leaf=1, min_samples_split=2, n_estimators=200; to
tal time=  18.9s
[CV] END max_depth=10, min_samples_leaf=1, min_samples_split=10, n_estimators=50; to
tal time=   4.7s
[CV] END max_depth=10, min_samples_leaf=1, min_samples_split=10, n_estimators=100; t
otal time=   9.4s
[CV] END max_depth=10, min_samples_leaf=2, min_samples_split=2, n_estimators=50; tot
al time=   4.7s
[CV] END max_depth=10, min_samples_leaf=2, min_samples_split=2, n_estimators=100; to
tal time=   9.4s
[CV] END max_depth=10, min_samples_leaf=2, min_samples_split=5, n_estimators=100; to
tal time=   9.4s
[CV] END max_depth=10, min_samples_leaf=2, min_samples_split=5, n_estimators=200; to
tal time=  18.9s
[CV] END max_depth=10, min_samples_leaf=4, min_samples_split=2, n_estimators=200; to
tal time=  18.7s
```

```
[CV] END max_depth=10, min_samples_leaf=4, min_samples_split=10, n_estimators=50; to
tal time=   4.7s
[CV] END max_depth=10, min_samples_leaf=4, min_samples_split=10, n_estimators=100; t
otal time=   9.3s
[CV] END max_depth=20, min_samples_leaf=1, min_samples_split=2, n_estimators=50; tot
al time=   8.9s
[CV] END max_depth=20, min_samples_leaf=1, min_samples_split=2, n_estimators=200; to
tal time=  35.6s
[CV] END max_depth=20, min_samples_leaf=1, min_samples_split=10, n_estimators=100; t
otal time=  15.3s
[CV] END max_depth=20, min_samples_leaf=2, min_samples_split=2, n_estimators=50; tot
al time=   7.9s
[CV] END max_depth=20, min_samples_leaf=2, min_samples_split=2, n_estimators=200; to
tal time=  32.0s
[CV] END max_depth=20, min_samples_leaf=2, min_samples_split=10, n_estimators=50; to
tal time=   7.4s
[CV] END max_depth=20, min_samples_leaf=2, min_samples_split=10, n_estimators=100; t
otal time=  14.8s
[CV] END max_depth=20, min_samples_leaf=4, min_samples_split=2, n_estimators=100; to
tal time=  14.0s
[CV] END max_depth=20, min_samples_leaf=4, min_samples_split=2, n_estimators=200; to
tal time=  28.1s
[CV] END max_depth=20, min_samples_leaf=4, min_samples_split=10, n_estimators=100; t
otal time=  13.8s
[CV] END max_depth=30, min_samples_leaf=1, min_samples_split=2, n_estimators=100; to
tal time=  19.6s
[CV] END max_depth=30, min_samples_leaf=1, min_samples_split=5, n_estimators=50; tot
al time=   8.7s
[CV] END max_depth=30, min_samples_leaf=1, min_samples_split=5, n_estimators=200; to
tal time=  35.0s
[CV] END max_depth=30, min_samples_leaf=1, min_samples_split=10, n_estimators=200; t
otal time=  32.2s
[CV] END max_depth=30, min_samples_leaf=2, min_samples_split=5, n_estimators=200; to
tal time=  32.8s
[CV] END max_depth=30, min_samples_leaf=4, min_samples_split=2, n_estimators=50; tot
al time=   7.1s
[CV] END max_depth=30, min_samples_leaf=4, min_samples_split=2, n_estimators=100; to
tal time=  14.0s
[CV] END max_depth=30, min_samples_leaf=4, min_samples_split=5, n_estimators=50; tot
al time=   7.1s
[CV] END max_depth=30, min_samples_leaf=4, min_samples_split=5, n_estimators=200; to
tal time=  28.5s
[CV] END max_depth=None, min_samples_leaf=1, min_samples_split=2, n_estimators=100;
total time=  19.5s
[CV] END max_depth=None, min_samples_leaf=1, min_samples_split=5, n_estimators=100;
total time=  17.7s
[CV] END max_depth=None, min_samples_leaf=1, min_samples_split=10, n_estimators=50;
total time=   8.0s
[CV] END max_depth=None, min_samples_leaf=1, min_samples_split=10, n_estimators=100;
total time=  16.1s
[CV] END max_depth=None, min_samples_leaf=2, min_samples_split=2, n_estimators=50; t
otal time=   8.4s
[CV] END max_depth=None, min_samples_leaf=2, min_samples_split=2, n_estimators=200;
total time=  33.6s
[CV] END max_depth=None, min_samples_leaf=2, min_samples_split=10, n_estimators=50;
total time=   7.5s
```

```
[CV] END max_depth=None, min_samples_leaf=2, min_samples_split=10, n_estimators=100;
total time=  15.2s
[CV] END max_depth=None, min_samples_leaf=4, min_samples_split=2, n_estimators=100;
total time=  14.1s
[CV] END max_depth=None, min_samples_leaf=4, min_samples_split=2, n_estimators=200;
total time=  28.5s
[CV] END max_depth=None, min_samples_leaf=4, min_samples_split=10, n_estimators=100;
total time=  13.9s
[CV] END max_depth=10, min_samples_leaf=1, min_samples_split=2, n_estimators=100; to
tal time=   9.5s
[CV] END max_depth=10, min_samples_leaf=1, min_samples_split=5, n_estimators=50; tot
al time=   4.7s
[CV] END max_depth=10, min_samples_leaf=1, min_samples_split=5, n_estimators=100; to
tal time=   9.4s
[CV] END max_depth=10, min_samples_leaf=1, min_samples_split=10, n_estimators=50; to
tal time=   4.7s
[CV] END max_depth=10, min_samples_leaf=1, min_samples_split=10, n_estimators=200; t
otal time=  18.7s
[CV] END max_depth=10, min_samples_leaf=2, min_samples_split=5, n_estimators=50; tot
al time=   4.7s
[CV] END max_depth=10, min_samples_leaf=2, min_samples_split=5, n_estimators=100; to
tal time=   9.4s
[CV] END max_depth=10, min_samples_leaf=2, min_samples_split=10, n_estimators=50; to
tal time=   4.7s
[CV] END max_depth=10, min_samples_leaf=2, min_samples_split=10, n_estimators=200; t
otal time=  18.6s
[CV] END max_depth=10, min_samples_leaf=4, min_samples_split=5, n_estimators=50; tot
al time=   4.7s
[CV] END max_depth=10, min_samples_leaf=4, min_samples_split=5, n_estimators=200; to
tal time=  18.6s
[CV] END max_depth=20, min_samples_leaf=1, min_samples_split=2, n_estimators=50; tot
al time=   8.9s
[CV] END max_depth=20, min_samples_leaf=1, min_samples_split=2, n_estimators=100; to
tal time=  17.8s
[CV] END max_depth=20, min_samples_leaf=1, min_samples_split=5, n_estimators=200; to
tal time=  32.8s
[CV] END max_depth=20, min_samples_leaf=1, min_samples_split=10, n_estimators=200; t
otal time=  30.7s
[CV] END max_depth=20, min_samples_leaf=2, min_samples_split=5, n_estimators=100; to
tal time=  15.7s
[CV] END max_depth=20, min_samples_leaf=2, min_samples_split=10, n_estimators=100; t
otal time=  14.6s
[CV] END max_depth=20, min_samples_leaf=2, min_samples_split=10, n_estimators=200; t
otal time=  29.5s
[CV] END max_depth=20, min_samples_leaf=4, min_samples_split=5, n_estimators=200; to
tal time=  27.9s
[CV] END max_depth=30, min_samples_leaf=1, min_samples_split=2, n_estimators=50; tot
al time=   9.8s
[CV] END max_depth=30, min_samples_leaf=1, min_samples_split=2, n_estimators=100; to
tal time=  19.4s
[CV] END max_depth=30, min_samples_leaf=1, min_samples_split=5, n_estimators=100; to
tal time=  17.5s
[CV] END max_depth=30, min_samples_leaf=1, min_samples_split=10, n_estimators=50; to
tal time=   7.9s
[CV] END max_depth=30, min_samples_leaf=1, min_samples_split=10, n_estimators=100; t
otal time=  15.8s
```

```
[CV] END max_depth=30, min_samples_leaf=2, min_samples_split=2, n_estimators=50; tot
al time=   8.4s
[CV] END max_depth=30, min_samples_leaf=2, min_samples_split=2, n_estimators=200; to
tal time=  33.2s
[CV] END max_depth=30, min_samples_leaf=2, min_samples_split=10, n_estimators=50; to
tal time=   7.5s
[CV] END max_depth=30, min_samples_leaf=2, min_samples_split=10, n_estimators=100; t
otal time=  14.9s
[CV] END max_depth=30, min_samples_leaf=4, min_samples_split=2, n_estimators=50; tot
al time=   7.1s
[CV] END max_depth=30, min_samples_leaf=4, min_samples_split=2, n_estimators=200; to
tal time=  28.3s
[CV] END max_depth=30, min_samples_leaf=4, min_samples_split=10, n_estimators=50; to
tal time=   6.9s
[CV] END max_depth=30, min_samples_leaf=4, min_samples_split=10, n_estimators=100; t
otal time=  13.2s
[CV] END max_depth=None, min_samples_leaf=1, min_samples_split=2, n_estimators=50; t
otal time=   9.9s
[CV] END max_depth=None, min_samples_leaf=1, min_samples_split=5, n_estimators=50; t
otal time=   8.7s
[CV] END max_depth=None, min_samples_leaf=1, min_samples_split=5, n_estimators=100;
total time=  17.6s
[CV] END max_depth=None, min_samples_leaf=1, min_samples_split=5, n_estimators=200;
total time=  35.5s
[CV] END max_depth=None, min_samples_leaf=2, min_samples_split=2, n_estimators=200;
total time=  33.4s
[CV] END max_depth=None, min_samples_leaf=2, min_samples_split=10, n_estimators=100;
total time=  15.0s
[CV] END max_depth=None, min_samples_leaf=2, min_samples_split=10, n_estimators=200;
total time=  30.3s
[CV] END max_depth=None, min_samples_leaf=4, min_samples_split=5, n_estimators=200;
total time=  27.8s
[CV] END max_depth=10, min_samples_leaf=1, min_samples_split=2, n_estimators=50; tot
al time=   4.7s
[CV] END max_depth=10, min_samples_leaf=1, min_samples_split=2, n_estimators=100; to
tal time=   9.4s
[CV] END max_depth=10, min_samples_leaf=1, min_samples_split=5, n_estimators=50; tot
al time=   4.7s
[CV] END max_depth=10, min_samples_leaf=1, min_samples_split=5, n_estimators=100; to
tal time=   9.4s
[CV] END max_depth=10, min_samples_leaf=1, min_samples_split=10, n_estimators=100; t
otal time=   9.3s
[CV] END max_depth=10, min_samples_leaf=1, min_samples_split=10, n_estimators=200; t
otal time=  18.6s
[CV] END max_depth=10, min_samples_leaf=2, min_samples_split=5, n_estimators=100; to
tal time=   9.4s
[CV] END max_depth=10, min_samples_leaf=2, min_samples_split=10, n_estimators=100; t
otal time=   9.3s
[CV] END max_depth=10, min_samples_leaf=2, min_samples_split=10, n_estimators=200; t
otal time=  18.6s
[CV] END max_depth=10, min_samples_leaf=4, min_samples_split=5, n_estimators=100; to
tal time=   9.3s
[CV] END max_depth=10, min_samples_leaf=4, min_samples_split=10, n_estimators=100; t
otal time=   9.3s
[CV] END max_depth=10, min_samples_leaf=4, min_samples_split=10, n_estimators=200; t
otal time=  18.5s
```

```
[CV] END max_depth=20, min_samples_leaf=1, min_samples_split=5, n_estimators=50; tot
al time=   8.2s
[CV] END max_depth=20, min_samples_leaf=1, min_samples_split=5, n_estimators=100; to
tal time=  16.5s
[CV] END max_depth=20, min_samples_leaf=1, min_samples_split=10, n_estimators=50; to
tal time=   7.6s
[CV] END max_depth=20, min_samples_leaf=1, min_samples_split=10, n_estimators=200; t
otal time=  30.5s
[CV] END max_depth=20, min_samples_leaf=2, min_samples_split=5, n_estimators=50; tot
al time=   7.8s
[CV] END max_depth=20, min_samples_leaf=2, min_samples_split=5, n_estimators=100; to
tal time=  15.5s
[CV] END max_depth=20, min_samples_leaf=2, min_samples_split=10, n_estimators=50; to
tal time=   7.4s
[CV] END max_depth=20, min_samples_leaf=2, min_samples_split=10, n_estimators=200; t
otal time=  29.3s
[CV] END max_depth=20, min_samples_leaf=4, min_samples_split=5, n_estimators=50; tot
al time=   7.0s
[CV] END max_depth=20, min_samples_leaf=4, min_samples_split=5, n_estimators=100; to
tal time=  13.7s
[CV] END max_depth=20, min_samples_leaf=4, min_samples_split=10, n_estimators=50; to
tal time=   6.7s
[CV] END max_depth=20, min_samples_leaf=4, min_samples_split=10, n_estimators=200; t
otal time=  27.4s
[CV] END max_depth=30, min_samples_leaf=1, min_samples_split=5, n_estimators=50; tot
al time=   8.7s
[CV] END max_depth=30, min_samples_leaf=1, min_samples_split=5, n_estimators=100; to
tal time=  17.4s
[CV] END max_depth=30, min_samples_leaf=1, min_samples_split=5, n_estimators=200; to
tal time=  35.3s
[CV] END max_depth=30, min_samples_leaf=2, min_samples_split=2, n_estimators=200; to
tal time=  33.1s
[CV] END max_depth=30, min_samples_leaf=2, min_samples_split=10, n_estimators=50; to
tal time=   7.4s
[CV] END max_depth=30, min_samples_leaf=2, min_samples_split=10, n_estimators=100; t
otal time=  15.2s
[CV] END max_depth=30, min_samples_leaf=4, min_samples_split=2, n_estimators=100; to
tal time=  14.2s
[CV] END max_depth=30, min_samples_leaf=4, min_samples_split=2, n_estimators=200; to
tal time=  28.4s
[CV] END max_depth=30, min_samples_leaf=4, min_samples_split=10, n_estimators=100; t
otal time=  13.4s
[CV] END max_depth=None, min_samples_leaf=1, min_samples_split=2, n_estimators=200;
total time=  39.2s
[CV] END max_depth=None, min_samples_leaf=1, min_samples_split=10, n_estimators=50;
total time=   8.0s
[CV] END max_depth=None, min_samples_leaf=1, min_samples_split=10, n_estimators=200;
total time=  31.8s
[CV] END max_depth=None, min_samples_leaf=2, min_samples_split=5, n_estimators=50; t
otal time=   8.1s
[CV] END max_depth=None, min_samples_leaf=2, min_samples_split=5, n_estimators=100;
total time=  16.4s
[CV] END max_depth=None, min_samples_leaf=2, min_samples_split=10, n_estimators=50;
total time=   7.6s
[CV] END max_depth=None, min_samples_leaf=2, min_samples_split=10, n_estimators=200;
total time=  30.0s
```

```
[CV] END max_depth=None, min_samples_leaf=4, min_samples_split=5, n_estimators=50; t
otal time=   7.1s
[CV] END max_depth=None, min_samples_leaf=4, min_samples_split=5, n_estimators=100;
total time=  13.8s
[CV] END max_depth=None, min_samples_leaf=4, min_samples_split=10, n_estimators=50;
total time=   6.8s
[CV] END max_depth=None, min_samples_leaf=4, min_samples_split=10, n_estimators=200;
total time=  27.7s
[CV] END max_depth=10, min_samples_leaf=1, min_samples_split=5, n_estimators=50; tot
al time=   4.7s
[CV] END max_depth=10, min_samples_leaf=1, min_samples_split=5, n_estimators=200; to
tal time=  18.8s
[CV] END max_depth=10, min_samples_leaf=2, min_samples_split=2, n_estimators=50; tot
al time=   4.7s
[CV] END max_depth=10, min_samples_leaf=2, min_samples_split=2, n_estimators=200; to
tal time=  18.8s
[CV] END max_depth=10, min_samples_leaf=2, min_samples_split=10, n_estimators=50; to
tal time=   4.7s
[CV] END max_depth=10, min_samples_leaf=2, min_samples_split=10, n_estimators=100; t
otal time=   9.3s
[CV] END max_depth=10, min_samples_leaf=4, min_samples_split=2, n_estimators=50; tot
al time=   4.6s
[CV] END max_depth=10, min_samples_leaf=4, min_samples_split=2, n_estimators=200; to
tal time=  18.6s
[CV] END max_depth=10, min_samples_leaf=4, min_samples_split=10, n_estimators=50; to
tal time=   4.6s
[CV] END max_depth=10, min_samples_leaf=4, min_samples_split=10, n_estimators=200; t
otal time=  18.5s
[CV] END max_depth=20, min_samples_leaf=1, min_samples_split=2, n_estimators=200; to
tal time=  35.6s
[CV] END max_depth=20, min_samples_leaf=1, min_samples_split=10, n_estimators=100; t
otal time=  15.2s
[CV] END max_depth=20, min_samples_leaf=2, min_samples_split=2, n_estimators=50; tot
al time=   8.0s
[CV] END max_depth=20, min_samples_leaf=2, min_samples_split=2, n_estimators=200; to
tal time=  31.7s
[CV] END max_depth=20, min_samples_leaf=2, min_samples_split=10, n_estimators=50; to
tal time=   7.3s
[CV] END max_depth=20, min_samples_leaf=2, min_samples_split=10, n_estimators=100; t
otal time=  14.5s
[CV] END max_depth=20, min_samples_leaf=4, min_samples_split=2, n_estimators=50; tot
al time=   7.0s
[CV] END max_depth=20, min_samples_leaf=4, min_samples_split=2, n_estimators=200; to
tal time=  28.0s
[CV] END max_depth=20, min_samples_leaf=4, min_samples_split=10, n_estimators=50; to
tal time=   6.9s
[CV] END max_depth=20, min_samples_leaf=4, min_samples_split=10, n_estimators=100; t
otal time=  13.5s
[CV] END max_depth=30, min_samples_leaf=1, min_samples_split=2, n_estimators=50; tot
al time=   9.9s
[CV] END max_depth=30, min_samples_leaf=1, min_samples_split=2, n_estimators=200; to
tal time=  39.1s
[CV] END max_depth=30, min_samples_leaf=1, min_samples_split=10, n_estimators=100; t
otal time=  15.9s
[CV] END max_depth=30, min_samples_leaf=2, min_samples_split=2, n_estimators=50; tot
al time=   8.3s
```

```
[CV] END max_depth=30, min_samples_leaf=2, min_samples_split=2, n_estimators=100; to
tal time=  16.5s
[CV] END max_depth=30, min_samples_leaf=2, min_samples_split=5, n_estimators=50; tot
al time=   8.2s
[CV] END max_depth=30, min_samples_leaf=2, min_samples_split=5, n_estimators=200; to
tal time=  32.4s
[CV] END max_depth=30, min_samples_leaf=4, min_samples_split=2, n_estimators=50; tot
al time=   7.0s
[CV] END max_depth=30, min_samples_leaf=4, min_samples_split=2, n_estimators=100; to
tal time=  14.3s
[CV] END max_depth=30, min_samples_leaf=4, min_samples_split=5, n_estimators=100; to
tal time=  14.2s
[CV] END max_depth=30, min_samples_leaf=4, min_samples_split=5, n_estimators=200; to
tal time=  27.4s
[CV] END max_depth=None, min_samples_leaf=1, min_samples_split=2, n_estimators=200;
total time=  39.2s
[CV] END max_depth=None, min_samples_leaf=1, min_samples_split=10, n_estimators=100;
total time=  15.9s
[CV] END max_depth=None, min_samples_leaf=2, min_samples_split=2, n_estimators=50; t
otal time=   8.3s
[CV] END max_depth=None, min_samples_leaf=2, min_samples_split=2, n_estimators=100;
total time=  16.6s
[CV] END max_depth=None, min_samples_leaf=2, min_samples_split=5, n_estimators=50; t
otal time=   8.1s
[CV] END max_depth=None, min_samples_leaf=2, min_samples_split=5, n_estimators=200;
total time=  32.6s
[CV] END max_depth=None, min_samples_leaf=4, min_samples_split=2, n_estimators=50; t
otal time=   7.1s
[CV] END max_depth=None, min_samples_leaf=4, min_samples_split=2, n_estimators=100;
total time=  13.9s
[CV] END max_depth=None, min_samples_leaf=4, min_samples_split=5, n_estimators=50; t
otal time=   6.9s
[CV] END max_depth=None, min_samples_leaf=4, min_samples_split=5, n_estimators=100;
total time=  14.2s
[CV] END max_depth=None, min_samples_leaf=4, min_samples_split=10, n_estimators=50;
total time=   7.0s
[CV] END max_depth=None, min_samples_leaf=4, min_samples_split=10, n_estimators=200;
total time=  27.2s
[CV] END max_depth=10, min_samples_leaf=1, min_samples_split=5, n_estimators=100; to
tal time=   9.4s
[CV] END max_depth=10, min_samples_leaf=1, min_samples_split=10, n_estimators=50; to
tal time=   4.7s
[CV] END max_depth=10, min_samples_leaf=1, min_samples_split=10, n_estimators=200; t
otal time=  18.6s
[CV] END max_depth=10, min_samples_leaf=2, min_samples_split=2, n_estimators=200; to
tal time=  18.8s
[CV] END max_depth=10, min_samples_leaf=2, min_samples_split=10, n_estimators=200; t
otal time=  18.6s
[CV] END max_depth=10, min_samples_leaf=4, min_samples_split=5, n_estimators=50; tot
al time=   4.6s
[CV] END max_depth=10, min_samples_leaf=4, min_samples_split=5, n_estimators=100; to
tal time=   9.3s
[CV] END max_depth=10, min_samples_leaf=4, min_samples_split=10, n_estimators=50; to
tal time=   4.6s
[CV] END max_depth=10, min_samples_leaf=4, min_samples_split=10, n_estimators=200; t
otal time=  18.4s
```

```
[CV] END max_depth=20, min_samples_leaf=1, min_samples_split=5, n_estimators=50; tot
al time=   8.2s
[CV] END max_depth=20, min_samples_leaf=1, min_samples_split=5, n_estimators=100; to
tal time=  16.4s
[CV] END max_depth=20, min_samples_leaf=1, min_samples_split=5, n_estimators=200; to
tal time=  32.9s
[CV] END max_depth=20, min_samples_leaf=2, min_samples_split=2, n_estimators=100; to
tal time=  16.1s
[CV] END max_depth=20, min_samples_leaf=2, min_samples_split=5, n_estimators=100; to
tal time=  15.7s
[CV] END max_depth=20, min_samples_leaf=2, min_samples_split=5, n_estimators=200; to
tal time=  31.5s
[CV] END max_depth=20, min_samples_leaf=4, min_samples_split=2, n_estimators=200; to
tal time=  27.5s
[CV] END max_depth=20, min_samples_leaf=4, min_samples_split=10, n_estimators=50; to
tal time=   6.8s
[CV] END max_depth=20, min_samples_leaf=4, min_samples_split=10, n_estimators=200; t
otal time=  26.9s
[CV] END max_depth=30, min_samples_leaf=1, min_samples_split=2, n_estimators=200; to
tal time=  39.3s
[CV] END max_depth=30, min_samples_leaf=1, min_samples_split=10, n_estimators=200; t
otal time=  31.7s
[CV] END max_depth=30, min_samples_leaf=2, min_samples_split=5, n_estimators=50; tot
al time=   8.1s
[CV] END max_depth=30, min_samples_leaf=2, min_samples_split=5, n_estimators=100; to
tal time=  16.1s
[CV] END max_depth=30, min_samples_leaf=2, min_samples_split=10, n_estimators=50; to
tal time=   7.6s
[CV] END max_depth=30, min_samples_leaf=2, min_samples_split=10, n_estimators=200; t
otal time=  30.0s
[CV] END max_depth=30, min_samples_leaf=4, min_samples_split=5, n_estimators=50; tot
al time=   7.1s
[CV] END max_depth=30, min_samples_leaf=4, min_samples_split=5, n_estimators=100; to
tal time=  13.9s
[CV] END max_depth=30, min_samples_leaf=4, min_samples_split=10, n_estimators=50; to
tal time=   6.8s
[CV] END max_depth=30, min_samples_leaf=4, min_samples_split=10, n_estimators=200; t
otal time=  24.8s
[CV] END max_depth=None, min_samples_leaf=1, min_samples_split=2, n_estimators=100;
total time=  19.6s
[CV] END max_depth=None, min_samples_leaf=1, min_samples_split=5, n_estimators=100;
total time=  17.4s
[CV] END max_depth=None, min_samples_leaf=1, min_samples_split=10, n_estimators=50;
total time=   8.0s
[CV] END max_depth=None, min_samples_leaf=1, min_samples_split=10, n_estimators=100;
total time=  15.8s
[CV] END max_depth=None, min_samples_leaf=2, min_samples_split=2, n_estimators=50; t
otal time=   8.3s
[CV] END max_depth=None, min_samples_leaf=2, min_samples_split=2, n_estimators=100;
total time=  16.9s
[CV] END max_depth=None, min_samples_leaf=2, min_samples_split=5, n_estimators=100;
total time=  16.4s
[CV] END max_depth=None, min_samples_leaf=2, min_samples_split=5, n_estimators=200;
total time=  33.0s
[CV] END max_depth=None, min_samples_leaf=4, min_samples_split=2, n_estimators=200;
total time=  27.9s
```

```
[CV] END max_depth=None, min_samples_leaf=4, min_samples_split=10, n_estimators=100;
total time=  13.9s
[CV] END max_depth=None, min_samples_leaf=4, min_samples_split=10, n_estimators=200;
total time=  27.9s
[CV] END max_depth=10, min_samples_leaf=1, min_samples_split=5, n_estimators=200; to
tal time=  18.8s
[CV] END max_depth=10, min_samples_leaf=2, min_samples_split=2, n_estimators=100; to
tal time=   9.4s
[CV] END max_depth=10, min_samples_leaf=2, min_samples_split=5, n_estimators=50; tot
al time=   4.7s
[CV] END max_depth=10, min_samples_leaf=2, min_samples_split=5, n_estimators=200; to
tal time=  18.7s
[CV] END max_depth=10, min_samples_leaf=4, min_samples_split=2, n_estimators=50; tot
al time=   4.6s
[CV] END max_depth=10, min_samples_leaf=4, min_samples_split=2, n_estimators=100; to
tal time=   9.3s
[CV] END max_depth=10, min_samples_leaf=4, min_samples_split=5, n_estimators=100; to
tal time=   9.3s
[CV] END max_depth=10, min_samples_leaf=4, min_samples_split=5, n_estimators=200; to
tal time=  18.5s
[CV] END max_depth=20, min_samples_leaf=1, min_samples_split=2, n_estimators=100; to
tal time=  17.8s
[CV] END max_depth=20, min_samples_leaf=1, min_samples_split=5, n_estimators=100; to
tal time=  16.3s
[CV] END max_depth=20, min_samples_leaf=1, min_samples_split=10, n_estimators=50; to
tal time=   7.6s
[CV] END max_depth=20, min_samples_leaf=1, min_samples_split=10, n_estimators=100; t
otal time=  15.3s
[CV] END max_depth=20, min_samples_leaf=2, min_samples_split=2, n_estimators=100; to
tal time=  15.9s
[CV] END max_depth=20, min_samples_leaf=2, min_samples_split=2, n_estimators=200; to
tal time=  32.0s
[CV] END max_depth=20, min_samples_leaf=2, min_samples_split=10, n_estimators=200; t
otal time=  29.1s
[CV] END max_depth=20, min_samples_leaf=4, min_samples_split=5, n_estimators=50; tot
al time=   6.8s
[CV] END max_depth=20, min_samples_leaf=4, min_samples_split=5, n_estimators=100; to
tal time=  14.0s
[CV] END max_depth=20, min_samples_leaf=4, min_samples_split=10, n_estimators=100; t
otal time=  13.7s
[CV] END max_depth=20, min_samples_leaf=4, min_samples_split=10, n_estimators=200; t
otal time=  27.5s
[CV] END max_depth=30, min_samples_leaf=1, min_samples_split=5, n_estimators=50; tot
al time=   8.8s
[CV] END max_depth=30, min_samples_leaf=1, min_samples_split=5, n_estimators=200; to
tal time=  34.8s
[CV] END max_depth=30, min_samples_leaf=2, min_samples_split=2, n_estimators=50; tot
al time=   8.2s
[CV] END max_depth=30, min_samples_leaf=2, min_samples_split=2, n_estimators=100; to
tal time=  16.8s
[CV] END max_depth=30, min_samples_leaf=2, min_samples_split=5, n_estimators=100; to
tal time=  16.3s
[CV] END max_depth=30, min_samples_leaf=2, min_samples_split=5, n_estimators=200; to
tal time=  32.9s
[CV] END max_depth=30, min_samples_leaf=4, min_samples_split=2, n_estimators=200; to
tal time=  27.8s
```

```
[CV] END max_depth=30, min_samples_leaf=4, min_samples_split=10, n_estimators=50; to
tal time=   7.0s
[CV] END max_depth=30, min_samples_leaf=4, min_samples_split=10, n_estimators=200; t
otal time=  24.3s
[CV] END max_depth=None, min_samples_leaf=1, min_samples_split=2, n_estimators=50; t
otal time=   9.7s
[CV] END max_depth=None, min_samples_leaf=1, min_samples_split=2, n_estimators=200;
total time=  39.5s
[CV] END max_depth=None, min_samples_leaf=1, min_samples_split=10, n_estimators=200;
total time=  31.6s
[CV] END max_depth=None, min_samples_leaf=2, min_samples_split=5, n_estimators=50; t
otal time=   8.2s
[CV] END max_depth=None, min_samples_leaf=2, min_samples_split=5, n_estimators=200;
total time=  32.3s
[CV] END max_depth=None, min_samples_leaf=4, min_samples_split=2, n_estimators=50; t
otal time=   6.9s
[CV] END max_depth=None, min_samples_leaf=4, min_samples_split=2, n_estimators=100;
total time=  14.2s
[CV] END max_depth=None, min_samples_leaf=4, min_samples_split=5, n_estimators=50; t
otal time=   7.1s
[CV] END max_depth=None, min_samples_leaf=4, min_samples_split=5, n_estimators=200;
total time=  28.3s
[CV] END max_depth=10, min_samples_leaf=1, min_samples_split=2, n_estimators=50; tot
al time=   4.7s
[CV] END max_depth=10, min_samples_leaf=1, min_samples_split=2, n_estimators=100; to
tal time=   9.5s
[CV] END max_depth=10, min_samples_leaf=1, min_samples_split=2, n_estimators=200; to
tal time=  18.9s
[CV] END max_depth=10, min_samples_leaf=1, min_samples_split=10, n_estimators=100; t
otal time=   9.3s
[CV] END max_depth=10, min_samples_leaf=2, min_samples_split=2, n_estimators=50; tot
al time=   4.7s
[CV] END max_depth=10, min_samples_leaf=2, min_samples_split=2, n_estimators=200; to
tal time=  18.7s
[CV] END max_depth=10, min_samples_leaf=2, min_samples_split=10, n_estimators=50; to
tal time=   4.6s
[CV] END max_depth=10, min_samples_leaf=2, min_samples_split=10, n_estimators=100; t
otal time=   9.3s
[CV] END max_depth=10, min_samples_leaf=4, min_samples_split=2, n_estimators=100; to
tal time=   9.3s
[CV] END max_depth=10, min_samples_leaf=4, min_samples_split=2, n_estimators=200; to
tal time=  18.5s
[CV] END max_depth=10, min_samples_leaf=4, min_samples_split=10, n_estimators=100; t
otal time=   9.2s
[CV] END max_depth=20, min_samples_leaf=1, min_samples_split=2, n_estimators=100; to
tal time=  17.7s
[CV] END max_depth=20, min_samples_leaf=1, min_samples_split=5, n_estimators=50; tot
al time=   8.2s
[CV] END max_depth=20, min_samples_leaf=1, min_samples_split=5, n_estimators=200; to
tal time=  32.7s
[CV] END max_depth=20, min_samples_leaf=2, min_samples_split=2, n_estimators=50; tot
al time=   8.0s
[CV] END max_depth=20, min_samples_leaf=2, min_samples_split=2, n_estimators=100; to
tal time=  15.8s
[CV] END max_depth=20, min_samples_leaf=2, min_samples_split=5, n_estimators=50; tot
al time=   7.9s
```

```
[CV] END max_depth=20, min_samples_leaf=2, min_samples_split=5, n_estimators=200; to
tal time=  31.0s
[CV] END max_depth=20, min_samples_leaf=4, min_samples_split=2, n_estimators=50; tot
al time=   6.8s
[CV] END max_depth=20, min_samples_leaf=4, min_samples_split=2, n_estimators=100; to
tal time=  14.0s
[CV] END max_depth=20, min_samples_leaf=4, min_samples_split=5, n_estimators=100; to
tal time=  14.0s
[CV] END max_depth=20, min_samples_leaf=4, min_samples_split=5, n_estimators=200; to
tal time=  27.9s
[CV] END max_depth=30, min_samples_leaf=1, min_samples_split=2, n_estimators=100; to
tal time=  19.6s
[CV] END max_depth=30, min_samples_leaf=1, min_samples_split=5, n_estimators=100; to
tal time=  17.6s
[CV] END max_depth=30, min_samples_leaf=1, min_samples_split=10, n_estimators=50; to
tal time=   7.9s
[CV] END max_depth=30, min_samples_leaf=1, min_samples_split=10, n_estimators=100; t
otal time=  16.0s
[CV] END max_depth=30, min_samples_leaf=2, min_samples_split=2, n_estimators=100; to
tal time=  16.7s
[CV] END max_depth=30, min_samples_leaf=2, min_samples_split=2, n_estimators=200; to
tal time=  33.6s
[CV] END max_depth=30, min_samples_leaf=2, min_samples_split=10, n_estimators=200; t
otal time=  29.9s
[CV] END max_depth=30, min_samples_leaf=4, min_samples_split=5, n_estimators=50; tot
al time=   7.0s
[CV] END max_depth=30, min_samples_leaf=4, min_samples_split=5, n_estimators=100; to
tal time=  14.2s
[CV] END max_depth=30, min_samples_leaf=4, min_samples_split=10, n_estimators=100; t
otal time=  13.9s
[CV] END max_depth=30, min_samples_leaf=4, min_samples_split=10, n_estimators=200; t
otal time=  23.5s
```

Analysis

The goal of my project was to predict how much an Airbnb listing would cost based on different features of the listing. I tried out a few models, but the best one turned out to be a Random Forest Regressor that I tuned using GridSearchCV to get the best settings.

Here's how it did:

The Mean Squared Error (MSE) was about 7,840, which basically means on average, the squared difference between predicted prices and actual prices was that much.

The R^2 score was 0.63, so the model explains about 63% of the variation in listing prices. The model is capturing a good chunk of what affects price, but there's still room for improvement.

Some of the most important features the model found were:

- accommodates (how many people the place fits)
- bathrooms (number of bathrooms)

- room_type_Private room (whether it’s a private room versus an entire place)
- review_scores_location (how guests rate the location)
- availability_30 (how many days it’s available in the next 30 days)

It makes sense that bigger places with more bathrooms and good locations cost more. Also, whether it’s a private room or entire home definitely changes the price. The availability feature probably helps show how in-demand a place is, which also impacts price.

For next steps, I think I could try transforming the price variable (like using a log scale) to handle the fact that errors seem to get bigger with higher prices. I might also try other models like Gradient Boosting to see if I can get better predictions.

I think this model is a great start and gives some good insights on what affects Airbnb prices.

In [ ]: