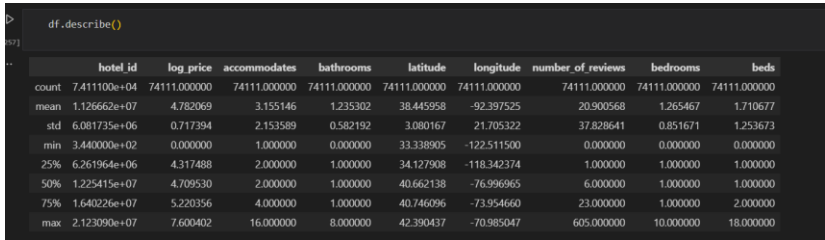


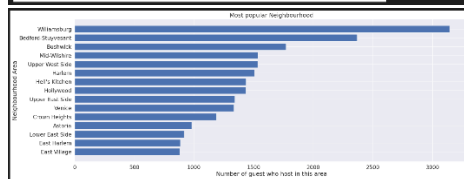
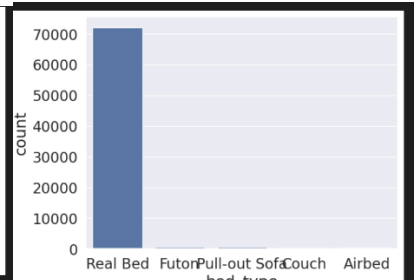
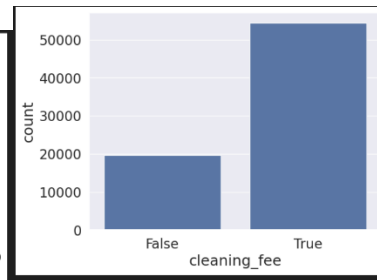
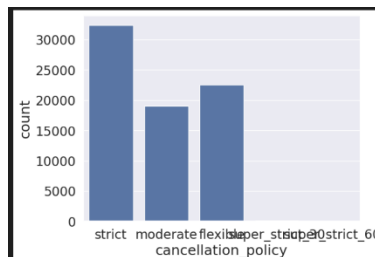
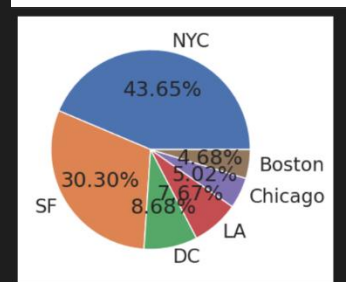
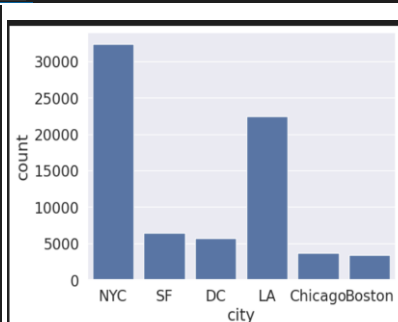
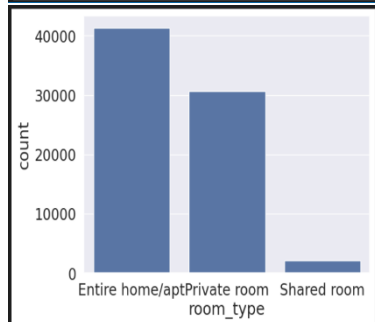
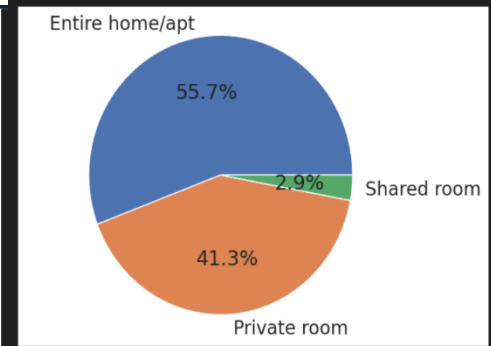
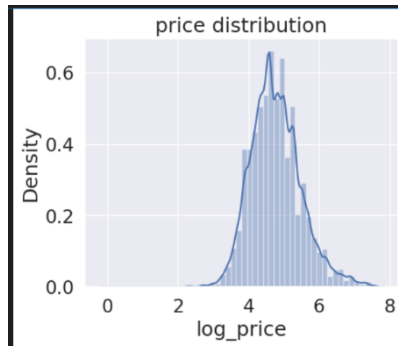
## Data Collection and Preprocessing Phase

Date	18 JULY 2024
Team ID	740111
Project Title	Unveiling Airbnb Price Patterns: Machine Learning For Forecasting
Maximum Marks	6 Marks

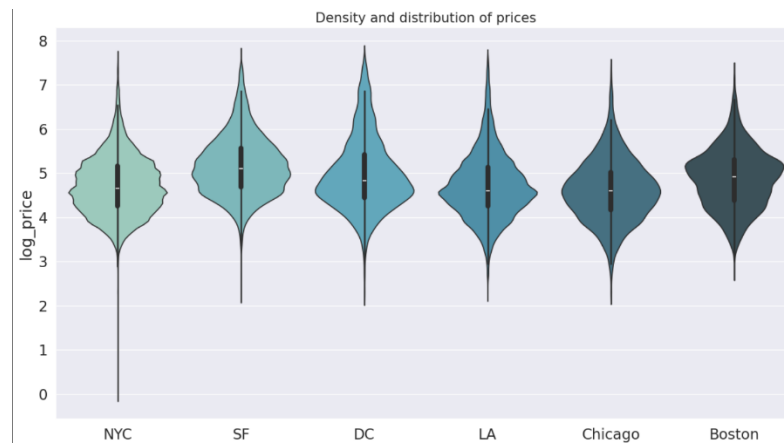
### Data Exploration and Preprocessing Report

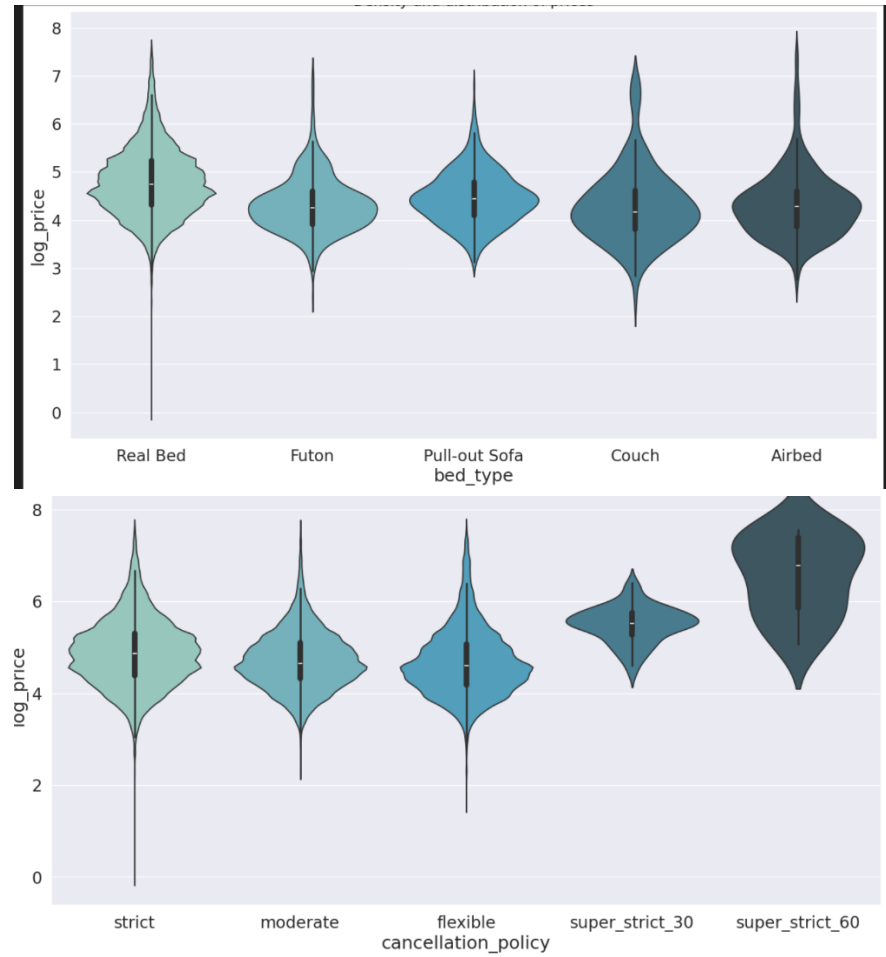
Dataset variables will be statistically analyzed to identify patterns and outliers, with Python employed for preprocessing tasks like normalization and feature engineering. Data cleaning will address missing values and outliers, ensuring quality for subsequent analysis and modeling, and forming a strong foundation for insights and predictions.

Section	Description
Data Overview	<p><u>Dimension:</u> 614 rows × 13 columns</p> <p><u>Descriptive statistics:</u></p> 
Univariate Analysis	

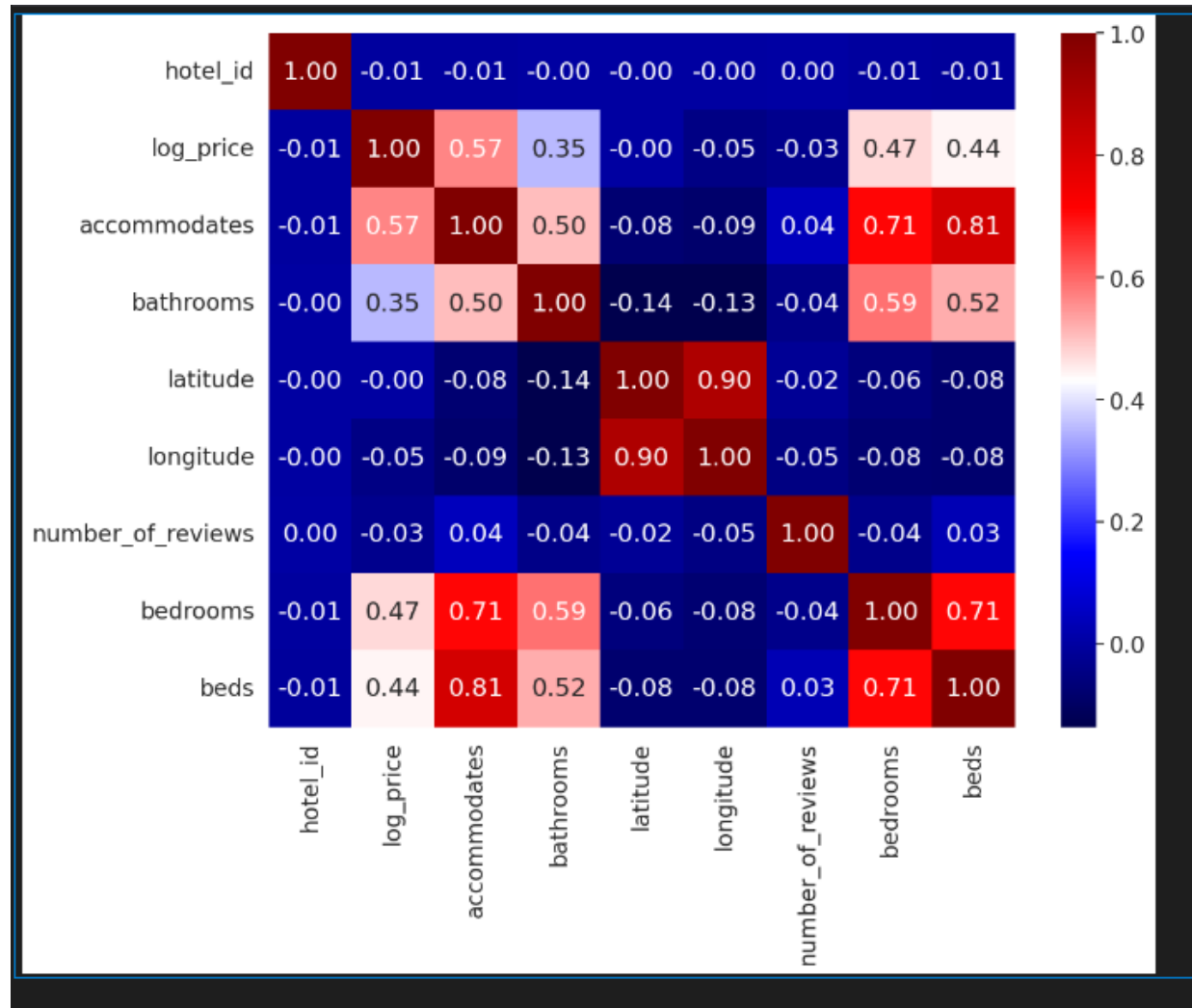


## Bivariate Analysis





## Multivariate Analysis



## Outliers and Anomalies

-

## Data Preprocessing Code Screenshots

## Loading Data

	id	log_price	property_type	room_type	amenities	accommodates	bathrooms	bed_type	cancellation_policy	cleaning_fee	latitude	longitude	name	neighbourhood	number_of_reviews	review_scores_rating	thumbnail_url	zipcode	bedrooms	beds
0	6901257	5.010635	Apartment	Entire home/apt	["Wireless Internet","Air conditioning","Stic...	3	1.0	Real Bed	strict	True	-40.696524	-73.991617	Beautiful brownstone 1+ bedroom	Brooklyn Heights	2	100.0	https://a0.muscache.com/im/pictures/8d7c3bf7-c...	11201	1.0	
1	6304938	5.129899	Apartment	Entire home/apt	["Wireless Internet","Air conditioning","Stic...	7	1.0	Real Bed	strict	True	-40.766115	-73.989040	Superb 3BR Apt Located Near Times Square	Hell's Kitchen	6	93.0	https://a0.muscache.com/im/pictures/348a55b4-4...	10019	3.0	
2	7918400	4.976734	Apartment	Entire home/apt	["TV","Cable TV","Wireless Internet","Air condi...	5	1.0	Real Bed	moderate	True	-40.808110	-73.943756	The Carlini Oasis	Harlem	10	92.0	https://a0.muscache.com/im/pictures/9a653658-8...	10027	1.0	
3	13418776	6.620073	House	Entire home/apt	["TV","Cable TV","Internet","Wireless Internet","St...	4	1.0	Real Bed	flexible	True	-37.772004	-122.431819	Beautiful Flat in the Heart of SF	Lower Haight	0	NaN	https://a0.muscache.com/im/pictures/72208da9-9...	94117.0	2.0	
4	3008709	4.744832	Apartment	Entire home/apt	["TV","Internet","Wireless Internet","Air conditio...	2	1.0	Real Bed	moderate	True	-38.925927	-77.034596	Great studio in midtown DC	Columbia Heights	4	40.0	NaN	20009	0.0	

Handling Missing Data	<pre> hotel_id            0 log_price           0 property_type       0 room_type           0 accommodates        0 bathrooms           200 bed_type            0 cancellation_policy 0 cleaning_fee        0 city                0 description          0 host_identity_verified 188 host_since          188 instant_bookable     0 latitude            0 longitude            0 name                0 neighbourhood        6872 number_of_reviews    0 bedrooms            91 beds                131 dtype: int64 </pre>
Data Transformation	<pre> from sklearn.preprocessing import StandardScaler import numpy as np  # Generate random data with matching dimensions x_train = np.random.rand(55583, 13) y_train = np.random.rand(55583) x_test = np.random.rand(18528, 13) y_test = np.random.rand(18528)  # Scale the data scaler = StandardScaler() x_train_scaled = scaler.fit_transform(x_train) x_test_scaled = scaler.transform(x_test) </pre>
Feature Engineering	Attached the codes in final submission.
Save Processed Data	-