



Data Collection and Preprocessing Phase

Date	18 JULY 2024
Team ID	739977
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	Dosavintio								
Section	Description								
	Dimension: 614 rows ×		lumns						
	Descriptive	statist	tics:						
	D df.describe()								
	hotel_id	log_price	accommodates	bathrooms	latitude	longitude	number_of_reviews	bedrooms	beds
	count 7.411100e+04	74111.000000	74111.000000	74111.000000	74111.000000	74111.000000	74111.000000	74111.000000	74111.000000
	mean 1.126662e+07	4.782069	3.155146	1.235302	38.445958		20.900568	1.265467	1.710677
	std 6.081735e+06	0.717394	2.153589	0.582192	3.080167	21.705322	37.828641	0.851671	1.253673
	min 3.440000e+02	0.000000	1.000000	0.000000	33.338905	-122.511500	0.000000	0.000000	0.000000
	25% 6.261964e+06	4.317488	2.000000	1.000000	34.127908	-118.342374	1.000000	1.000000	1.000000
	50% 1.225415e+07	4.709530	2.000000	1.000000	40.662138	-76.996965	6.000000	1.000000	1.000000
	75% 1.640226e+07	5.220356	4.000000	1.000000	40.746096	-73.954660	23.000000	1.000000	2.000000
	max 2.123090e+07	7.600402	16.000000	8.000000	42.390437	-70.985047	605.000000	10.000000	18.000000
Data Overview									



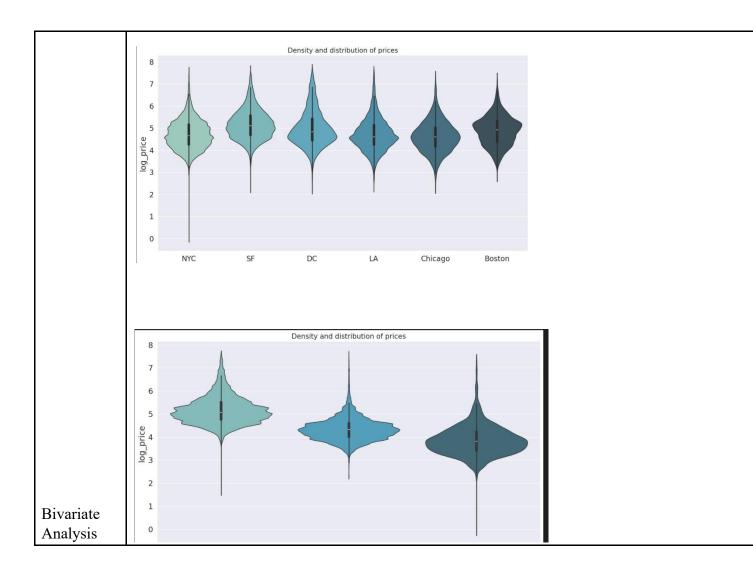




Univariate Analysis

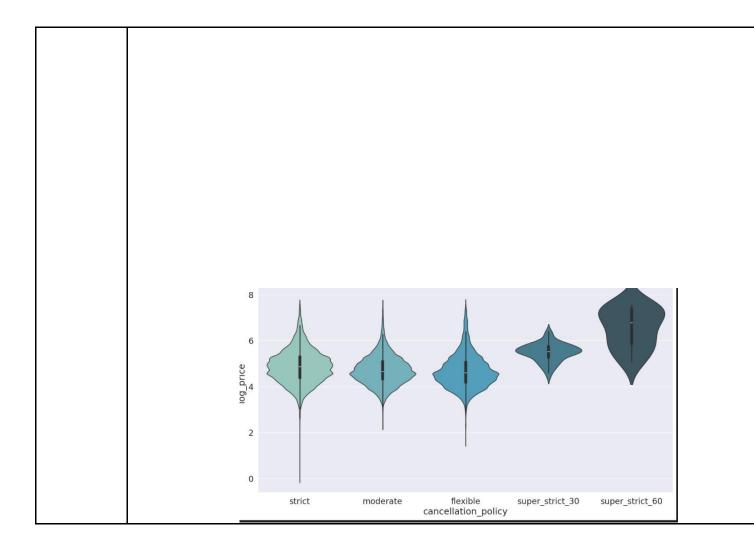






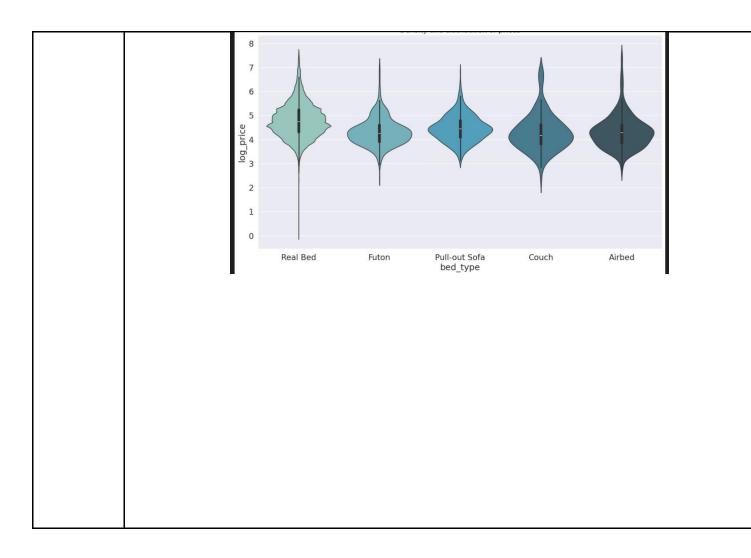






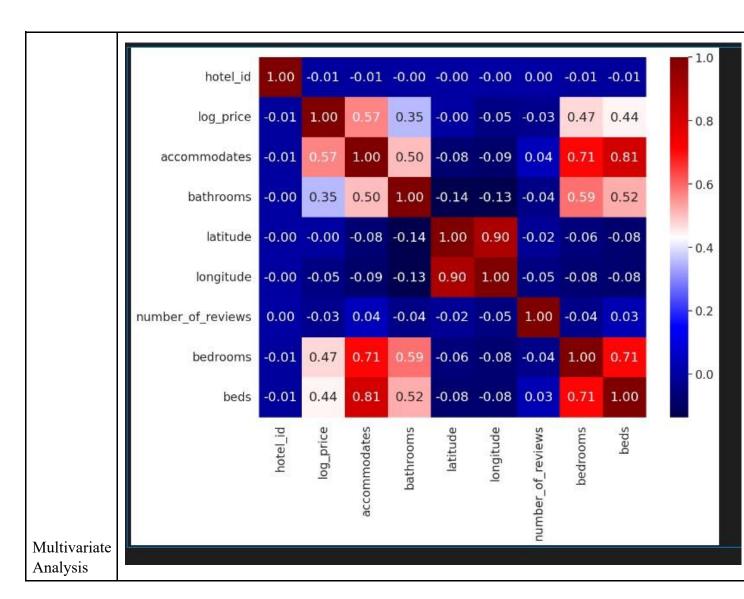
















Outliers and Anomalies	-
Data	

sing Code Screenshots

Preproces





	·
	If log_price_property_type_coses_type assection accommodate betterooms ber_type_concentions_price_standing_fee_substander_berteaming_fee_substander_
	1 650425 513909 Apartment Unity Types (Confidence No. 7 to New Bed on 12 Year 43790115 -77389040 Income New Hell's Octom 6 513 (Confidence No. 1001) 3.5 (Confidence No. 1001)
	4 3508199 4.748932 Apertment Street (TV35mmer-Yostenia 2 10 Feel Bed modernine Tow 38875077 -77.7585198 Great Notice in Columnia 4 46.5 Feel 20099 6.5 Research Feel 20099 6.5
Loading Data	
Handling Missing Data	hotel_id
	<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)</pre>
Data Transformation	<pre>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	
Troccssed Data	<u></u>