Name: Hosain Ahmed Student ID: 301209637

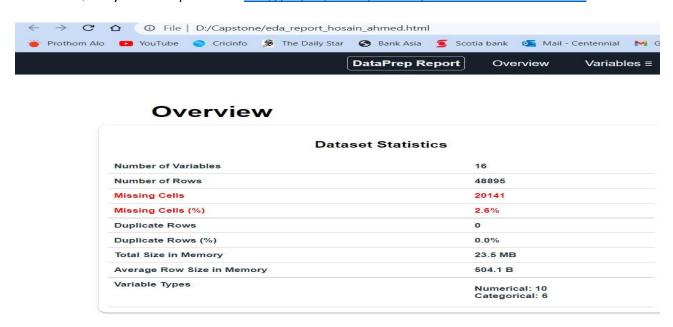
Data Source: https://www.kaggle.com/datasets/dgomonov/new-york-city-airbnb-open-data

Data Structure:Rows: 48895Columns: 16

- Column names: id, name, host_id, host_name, neighbourhood_group, neighbourhood, latitude, longitude, room_type, price, minimum_nights, number_of_reviews, last_review, reviews_per_month, calculated_host_listings_count, availability_365
- Data types:

Numerical: 10Categorical: 6

• Data Quality: EDA Report Link: file:///D:/Capstone/eda_report_hosain_ahmed.html



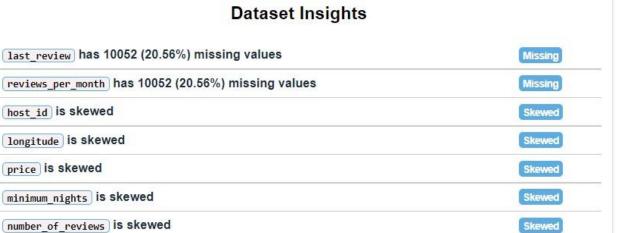
Name: Hosain Ahmed Student ID: 301209637

price is skewed

reviews per month is skewed

availability_365 is skewed

calculated_host_listings_count is skewed



Skewed

Skewed

Skewed

1 2

Dataset Insights

name has a high cardinality: 47905 distinct values	High Cardinality
host_name has a high cardinality: 11452 distinct values	High Cardinality
neighbourhood has a high cardinality: 221 distinct values	High Cardinality
last_review has a high cardinality: 1764 distinct values	High Cardinality
last_review has constant length 10	Constant Length
longitude has 48895 (100.0%) negatives	Negatives
number_of_reviews has 10052 (20.56%) zeros	Zeros
availability_365 has 17533 (35.86%) zeros	Zeros

Name: Hosain Ahmed Student ID: 301209637

Excluded Variables:

EXCLUSION	VARIABLE	DEFINITION
	id	Property ID
Irrelevant variables	Name	Property Tag Name
variables	host_id	Property host ID
	last_review	When the last time host recieved a review
Variables with	calculated_host_listings_count	Number of property a single host is offering
more		
skewness		
value of		
7.56774264		

Selected Variables:

VARIABLE	DEFINITION	
neighbourhood_group	Larger neighborhood groups	
neighbourhood	Specific áreas under the neighborhood groups	
latitude	Latitude of the property	
longitude	Longitude of the property	
room_type	Types of rooms offered	
price	Price per night stay in the property.	
minimum_nights	Number of night stayed in the property	
number_of_reviews	Total number of reviews each host received	
reviews_per_month	Reviews received per month by each host	
availability_365	The property was available for booking for how many days during the whole year	

Issues Found in variables:

Issues Found	Variables	Potential Actions
Skewness	host_id (Rejected), longitude,	Cap & Floor and Log
	price, minimum_nights,	Transformation.

Name: Hosain Ahmed Student ID: 301209637

	number_of_reviews,	
	reviews_per_month,	
	calculated_host_listings_cou	nt,
	availability_365	
Missing Values	last_review,	Imputation/ Drop Missing
	reviews_per_month	Values

o Potential actions

- Cap & Floor for Skewness
- Imputation for missing values
- Log Transform for Skewness