

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

```
data = pd.read_csv('/home/mridul/PYTHON/Data Cleaning/Datasets/Airbnb_Open_Data.csv')
data
```

/tmp/ipykernel_22272/12675833.py:1: DtypeWarning: Columns (25) have mixed types. Specify dtype option on import or set low_mem

data = pd.read_csv('/home/mridul/PYTHON/Data Cleaning/Datasets/Airbnb_Open_Data.csv')

	id	NAME	host id	host_identity_verified	host name	neighbourhood group	neighbourhood	lat	long	cou
0	1001254	Clean & quiet apt home by the park	80014485718	unconfirmed	Madaline	Brooklyn	Kensington	40.64749	-73.97237	U
1	1002102	Skylit Midtown Castle	52335172823	verified	Jenna	Manhattan	Midtown	40.75362	-73.98377	U
2	1002403	THE VILLAGE OF HARLEM....NEW YORK !	78829239556	NaN	Elise	Manhattan	Harlem	40.80902	-73.94190	U
3	1002755	NaN	85098326012	unconfirmed	Garry	Brooklyn	Clinton Hill	40.68514	-73.95976	U
4	1003689	Entire Apt: Spacious Studio/Loft by central park	92037596077	verified	Lyndon	Manhattan	East Harlem	40.79851	-73.94399	U
...
102594	6092437	Spare room in Williamsburg	12312296767	verified	Krik	Brooklyn	Williamsburg	40.70862	-73.94651	U
102595	6092990	Best Location near Columbia U	77864383453	unconfirmed	Mifan	Manhattan	Morningside Heights	40.80460	-73.96545	U
102596	6093542	Comfy, bright room in Brooklyn	69050334417	unconfirmed	Megan	Brooklyn	Park Slope	40.67505	-73.98045	U
102597	6094094	Big Studio-One Stop from Midtown	11160591270	unconfirmed	Christopher	Queens	Long Island City	40.74989	-73.93777	U
102598	6094647	585 sf Luxury Studio	68170633372	unconfirmed	Rebecca	Manhattan	Upper West Side	40.76807	-73.98342	U

102599 rows × 26 columns

```
data.columns

Index(['id', 'NAME', 'host id', 'host_identity_verified', 'host name',
      'neighbourhood group', 'neighbourhood', 'lat', 'long', 'country',
      'country code', 'instant_bookable', 'cancellation_policy', 'room type',
      'Construction year', 'price', 'service fee', 'minimum nights',
      'number of reviews', 'last review', 'reviews per month',
      'review rate number', 'calculated host listings count',
      'availability 365', 'house_rules', 'license'],
      dtype='object')
```

```
data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 102599 entries, 0 to 102598
Data columns (total 26 columns):
 #   Column                                Non-Null Count  Dtype
---  -
 0   id                                    102599 non-null  int64
 1   NAME                                102349 non-null  object
 2   host id                             102599 non-null  int64
 3   host_identity_verified              102310 non-null  object
 4   host name                           102193 non-null  object
 5   neighbourhood group                 102570 non-null  object
 6   neighbourhood                       102583 non-null  object
 7   lat                                 102591 non-null  float64
 8   long                                102591 non-null  float64
 9   country                             102067 non-null  object
```

```
10 country code 102468 non-null object
11 instant_bookable 102494 non-null object
12 cancellation_policy 102523 non-null object
13 room type 102599 non-null object
14 Construction year 102385 non-null float64
15 price 102352 non-null object
16 service fee 102326 non-null object
17 minimum nights 102190 non-null float64
18 number of reviews 102416 non-null float64
19 last review 86706 non-null object
20 reviews per month 86720 non-null float64
21 review rate number 102273 non-null float64
22 calculated host listings count 102280 non-null float64
23 availability 365 102151 non-null float64
24 house_rules 50468 non-null object
25 license 2 non-null object
dtypes: float64(9), int64(2), object(15)
memory usage: 20.4+ MB
```

▼ Checklist

Renaming the columns

Deleting Redundant Columns

Dropping Duplicates

cleaning individual columns

Remove the NaN values from dataset

Check for more transformations on dataset

```
cols_to_drop = ['reviews per month', 'review rate number', 'calculated host listings count', 'availability 365', 'house rules']
df = data.drop(columns = cols_to_drop)
```

df.shape df											
	NAME	host id	host_identity_verified	host name	neighbourhood group	neighbourhood	lat	long	country	code	id
0	Clean & quiet apt home by the park	80014485718	unconfirmed	Madaline	Brooklyn	Kensington	40.64749	-73.97237	United States		
1	Skylit Midtown Castle	52335172823	verified	Jenna	Manhattan	Midtown	40.75362	-73.98377	United States		
2	THE VILLAGE OF HARLEM....NEW YORK !	78829239556	NaN	Elise	Manhattan	Harlem	40.80902	-73.94190	United States		
3	NaN	85098326012	unconfirmed	Garry	Brooklyn	Clinton Hill	40.68514	-73.95976	United States		
4	Entire Apt: Spacious Studio/Loft by central park	92037596077	verified	Lyndon	Manhattan	East Harlem	40.79851	-73.94399	United States		
...	
102594	Spare room in Williamsburg	12312296767	verified	Krik	Brooklyn	Williamsburg	40.70862	-73.94651	United States		
102595	Best Location near Columbia U	77864383453	unconfirmed	Mifan	Manhattan	Morningside Heights	40.80460	-73.96545	United States		
102596	Comfy, bright room in Brooklyn	69050334417	unconfirmed	Megan	Brooklyn	Park Slope	40.67505	-73.98045	United States		
102597	Big Studio-One Stop from Midtown	11160591270	unconfirmed	Christopher	Queens	Long Island City	40.74989	-73.93777	United States		
102598	585 sf Luxury Studio	68170633372	unconfirmed	Rebecca	Manhattan	Upper West Side	40.76807	-73.98342	United States		
102599 rows × 19 columns											

```
new_cols = []
for i in df.columns:
    new_cols.append(i.capitalize())

for i in new_cols:
```

df

	Name	Host id	Host_identity_verified	Host name	Neighbourhood group	Neighbourhood	Lat	Long	Country	Col
0	Clean & quiet apt home by the park	80014485718	unconfirmed	Madaline	Brooklyn	Kensington	40.64749	-73.97237	United States	
1	Skylit Midtown Castle	52335172823	verified	Jenna	Manhattan	Midtown	40.75362	-73.98377	United States	
2	THE VILLAGE OF HARLEM....NEW YORK !	78829239556	NaN	Elise	Manhattan	Harlem	40.80902	-73.94190	United States	
3	NaN	85098326012	unconfirmed	Garry	Brooklyn	Clinton Hill	40.68514	-73.95976	United States	
4	Entire Apt: Spacious Studio/Loft by central park	92037596077	verified	Lyndon	Manhattan	East Harlem	40.79851	-73.94399	United States	
...
102594	Spare room in Williamsburg	12312296767	verified	Krik	Brooklyn	Williamsburg	40.70862	-73.94651	United States	
102595	Best Location near Columbia U	77864383453	unconfirmed	Mifan	Manhattan	Morningside Heights	40.80460	-73.96545	United States	
102596	Comfy, bright room in Brooklyn	69050334417	unconfirmed	Megan	Brooklyn	Park Slope	40.67505	-73.98045	United States	
102597	Big Studio-One Stop from Midtown	11160591270	unconfirmed	Christopher	Queens	Long Island City	40.74989	-73.93777	United States	
102598	585 sf Luxury Studio	68170633372	unconfirmed	Rebecca	Manhattan	Upper West Side	40.76807	-73.98342	United States	

102599 rows × 19 columns

df

	Name	Host id	Host_identity_verified	Host name	Neighbourhood group	Neighbourhood	Lat	Long	Country	Co
0	Clean & quiet apt home by the park	80014485718	unconfirmed	Madaline	Brooklyn	Kensington	40.64749	-73.97237	United States	
1	Skylit Midtown Castle	52335172823	verified	Jenna	Manhattan	Midtown	40.75362	-73.98377	United States	
2	THE VILLAGE OF HARLEM....NEW YORK !	78829239556	NaN	Elise	Manhattan	Harlem	40.80902	-73.94190	United States	
3	NaN	85098326012	unconfirmed	Garry	Brooklyn	Clinton Hill	40.68514	-73.95976	United States	
4	Entire Apt: Spacious Studio/Loft by central park	92037596077	verified	Lyndon	Manhattan	East Harlem	40.79851	-73.94399	United States	
...
102594	Spare room in Williamsburg	12312296767	verified	Krik	Brooklyn	Williamsburg	40.70862	-73.94651	United States	
102595	Best Location near Columbia U	77864383453	unconfirmed	Mifan	Manhattan	Morningside Heights	40.80460	-73.96545	United States	
102596	Comfy, bright room in Brooklyn	69050334417	unconfirmed	Megan	Brooklyn	Park Slope	40.67505	-73.98045	United States	
102597	Big Studio-One Stop from Midtown	11160591270	unconfirmed	Christopher	Queens	Long Island City	40.74989	-73.93777	United States	
102598	585 sf Luxury Studio	68170633372	unconfirmed	Rebecca	Manhattan	Upper West Side	40.76807	-73.98342	United States	
102599 rows × 19 columns										

```
# checking for duplicates
df.duplicated().sum()

np.int64(541)
```

```
df.drop_duplicates(inplace = True)
df.duplicated().sum()

np.int64(0)
```

```
# checking for missing values :
df.isna().sum()

Name                250
Host id              0
Host_identity_verified  289
Host name            404
Neighbourhood group   29
Neighbourhood        16
Lat                  8
Long                 8
Country              532
Country code         131
Instant_bookable     105
Cancellation_policy   76
Room type            0
Construction year     214
Price                247
Service fee           273
Minimum nights        400
Number of reviews     183
Last review          15832
dtype: int64
```

```
df.drop(columns = 'Last review', inplace = True)
```

```
df
```

	Name	Host id	Host_identity_verified	Host name	Neighbourhood group	Neighbourhood	Lat	Long	Country	Coun c
0	Clean & quiet apt home by the park	80014485718	unconfirmed	Madaline	Brooklyn	Kensington	40.64749	-73.97237	United States	
1	Skylit Midtown Castle	52335172823	verified	Jenna	Manhattan	Midtown	40.75362	-73.98377	United States	
2	THE VILLAGE OF HARLEM....NEW YORK !	78829239556	NaN	Elise	Manhattan	Harlem	40.80902	-73.94190	United States	
3	NaN	85098326012	unconfirmed	Garry	Brooklyn	Clinton Hill	40.68514	-73.95976	United States	
4	Entire Apt: Spacious Studio/Loft by central park	92037596077	verified	Lyndon	Manhattan	East Harlem	40.79851	-73.94399	United States	
...
102053	Cozy bright room near Prospect Park	77326652202	unconfirmed	Mariam	Brooklyn	Flatbush	40.64945	-73.96108	United States	
102054	Private Bedroom with Amazing Rooftop View	45936254757	verified	Trey	Brooklyn	Bushwick	40.69872	-73.92718	United States	
102055	Pretty Brooklyn One-Bedroom for 2 to 4 people	23801060917	verified	Michael	Brooklyn	Bedford-Stuyvesant	40.67810	-73.90822	United States	
102056	Room & private bathroom in historic Harlem	15593031571	unconfirmed	Shireen	Manhattan	Harlem	40.81248	-73.94317	United States	
102057	Rosalee Stewart	93578954226	verified	Stanley	Manhattan	Harlem	40.81315	-73.94747	United States	
102058 rows × 18 columns										

```
df.dropna(inplace = True)
df.isna().sum()

Name 0
Host_id 0
Host_identity_verified 0
Host_name 0
Neighbourhood_group 0
Neighbourhood 0
Lat 0
Long 0
Country 0
Country_code 0
Instant_bookable 0
Cancellation_policy 0
Room_type 0
Construction_year 0
Price 0
Service_fee 0
Minimum_nights 0
Number_of_reviews 0
dtype: int64
```

		Name	Host id	Host_identity_verified	Host name	Neighbourhood group	Neighbourhood	Lat	Long	Country	Count	co
	0	Clean & quiet apt home by the park	80014485718	UNCONFIRMED	Madaline	Brooklyn	Kensington	40.64749	-73.97237	United States		U
	1	Skylit Midtown Castle	52335172823	VERIFIED	Jenna	Manhattan	Midtown	40.75362	-73.98377	United States		U
	4	Entire Apt: Spacious Studio/Loft by central park	92037596077	VERIFIED	Lyndon	Manhattan	East Harlem	40.79851	-73.94399	United States		U
	5	Large Cozy 1 BR Apartment In Midtown East	45498551794	VERIFIED	Michelle	Manhattan	Murray Hill	40.74767	-73.97500	United States		U
	7	BlissArtsSpace!	90821839709	UNCONFIRMED	Emma	Brooklyn	Bedford-Stuyvesant	40.68688	-73.95596	United States		U
...			
	102029	Cozy room in a 2B with backyard access	5155155913	UNCONFIRMED	Cristina	Queens	Astoria	40.76987	-73.92679	United States		U
	102030	An East Williamsburg Delight	71564706996	VERIFIED	Petra	Brooklyn	Bedford-Stuyvesant	40.68335	-73.91222	United States		U
	102031	Cozy bedroom in williamsburg	83101851929	VERIFIED	Russell	Brooklyn	Williamsburg	40.70870	-73.96710	United States		U
	102032	Spacious Room w/ 2 Bed	69037598918	UNCONFIRMED	Christine	Queens	Bayside	40.75047	-73.75349	United States		U
	102040	Room in Queens, NY, near LGA.	56457739998	VERIFIED	Sonia	Queens	East Elmhurst	40.76245	-73.87938	United States		U
99342 rows × 18 columns												

```
df['Instant_bookable'] = df['Instant_bookable'].apply(lambda x : 1 if x == True else 0)
df['Instant_bookable']
df
```

	Name	Host id	Host_identity_verified	Host name	Neighbourhood group	Neighbourhood	Lat	Long	Country	Count	co
0	Clean & quiet apt home by the park	80014485718	UNCONFIRMED	Madaline	Brooklyn	Kensington	40.64749	-73.97237	United States		U
1	Skylit Midtown Castle	52335172823	VERIFIED	Jenna	Manhattan	Midtown	40.75362	-73.98377	United States		U
4	Entire Apt: Spacious Studio/Loft by central park	92037596077	VERIFIED	Lyndon	Manhattan	East Harlem	40.79851	-73.94399	United States		U
5	Large Cozy 1 BR Apartment In Midtown East	45498551794	VERIFIED	Michelle	Manhattan	Murray Hill	40.74767	-73.97500	United States		U
7	BlissArtsSpace!	90821839709	UNCONFIRMED	Emma	Brooklyn	Bedford-Stuyvesant	40.68688	-73.95596	United States		U
...		
102029	Cozy room in a 2B with backyard access	5155155913	UNCONFIRMED	Cristina	Queens	Astoria	40.76987	-73.92679	United States		U
102030	An East Williamsburg Delight	71564706996	VERIFIED	Petra	Brooklyn	Bedford-Stuyvesant	40.68335	-73.91222	United States		U
102031	Cozy bedroom in williamsburg	83101851929	VERIFIED	Russell	Brooklyn	Williamsburg	40.70870	-73.96710	United States		U
102032	Spacious Room w/ 2 Bed	69037598918	UNCONFIRMED	Christine	Queens	Bayside	40.75047	-73.75349	United States		U
102040	Room in Queens, NY, near LGA.	56457739998	VERIFIED	Sonia	Queens	East Elmhurst	40.76245	-73.87938	United States		U
99342 rows × 18 columns											

```
df['Instant_bookable']

0      0
1      0
4      0
5      1
7      0

..
102029  0
102030  1
102031  1
102032  0
102040  1
Name: Instant_bookable, Length: 99342, dtype: int64
```

```
df.reset_index(inplace=True,drop = True)
```

```
df.head(20)
```

	Name	Host id	Host_identity_verified	Host name	Neighbourhood group	Neighbourhood	Lat	Long	Country	Country code
0	Clean & quiet apt home by the park	80014485718	UNCONFIRMED	Madaline	Brooklyn	Kensington	40.64749	-73.97237	United States	US
1	Skylit Midtown Castle	52335172823	VERIFIED	Jenna	Manhattan	Midtown	40.75362	-73.98377	United States	US
2	Entire Apt: Spacious Studio/Loft by central park	92037596077	VERIFIED	Lyndon	Manhattan	East Harlem	40.79851	-73.94399	United States	US
3	Large Cozy 1 BR Apartment In Midtown East	45498551794	VERIFIED	Michelle	Manhattan	Murray Hill	40.74767	-73.97500	United States	US
4	BlissArtsSpace!	90821839709	UNCONFIRMED	Emma	Brooklyn	Bedford-Stuyvesant	40.68688	-73.95596	United States	US
5	Large Furnished Room Near B'way	79384379533	VERIFIED	Evelyn	Manhattan	Hell's Kitchen	40.76489	-73.98493	United States	US
6	Cozy Clean Guest Room - Family Apt	75527839483	UNCONFIRMED	Carl	Manhattan	Upper West Side	40.80178	-73.96723	United States	US
7	Cute & Cozy Lower East Side 1 bdrm	1280143094	VERIFIED	Miranda	Manhattan	Chinatown	40.71344	-73.99037	United States	US
8	Beautiful 1br on Upper West Side	18824631834	VERIFIED	Alan	Manhattan	Upper West Side	40.80316	-73.96545	United States	US
9	Lovely Room 1, Garden, Best Area, Legal rental	26802410424	VERIFIED	Darcy	brookln	South Slope	40.66829	-73.98779	United States	US
10	Wonderful Guest Bedroom in Manhattan for SINGLES	88920244552	VERIFIED	Leonardo	Manhattan	Upper West Side	40.79826	-73.96113	United States	US
11	Large B&B Style rooms	57938703679	UNCONFIRMED	Jared	Brooklyn	Bedford-Stuyvesant	40.68310	-73.95473	United States	US
12	Lovely room 2 & garden; Best area, Legal rental	53343148512	VERIFIED	Brad	Brooklyn	South Slope	40.66869	-73.98780	United States	US
13	Clean and Quiet in Brooklyn	26207748876	VERIFIED	Arthur	Brooklyn	Bedford-Stuyvesant	40.68876	-73.94312	United States	US
14	Cute apt in artist's home	88653822946	VERIFIED	Joyce	Brooklyn	Bushwick	40.70186	-73.92745	United States	US
15	Country space in the city	49725315867	VERIFIED	Deanna	Brooklyn	Flatbush	40.63702	-73.96327	United States	US
16	ENJOY Downtown NYC!	63131487263	UNCONFIRMED	Byron	Manhattan	East Village	40.72290	-73.98199	United States	US
17	Beautiful Sunny Park Slope Brooklyn	50357575975	VERIFIED	Alina	Brooklyn	South Slope	40.66278	-73.97966	United States	US
18	West Side Retreat	55430108992	UNCONFIRMED	Alford	Manhattan	Upper West Side	40.79009	-73.97927	United States	US
19	BEST BET IN HARLEM	2107606817	UNCONFIRMED	Chester	Manhattan	Harlem	40.81175	-73.94478	United States	US

```
df['Total Price'] = df['Price'] + df['Service fee']
df.drop(columns = 'Total Price', inplace = True)
df
```


	Name	Host id	Host_identity_verified	Host name	Neighbourhood group	Neighbourhood	Lat	Long	Country	Country code
0	Clean & quiet apt home by the park	80014485718	UNCONFIRMED	Madaline	Brooklyn	Kensington	40.64749	-73.97237	United States	US
1	Skylit Midtown Castle	52335172823	VERIFIED	Jenna	Manhattan	Midtown	40.75362	-73.98377	United States	US
2	Entire Apt: Spacious Studio/Loft by central park	92037596077	VERIFIED	Lyndon	Manhattan	East Harlem	40.79851	-73.94399	United States	US
3	Large Cozy 1 BR Apartment In Midtown East	45498551794	VERIFIED	Michelle	Manhattan	Murray Hill	40.74767	-73.97500	United States	US
4	BlissArtsSpace!	90821839709	UNCONFIRMED	Emma	Brooklyn	Bedford-Stuyvesant	40.68688	-73.95596	United States	US
...
99337	Cozy room in a 2B with backyard access	5155155913	UNCONFIRMED	Cristina	Queens	Astoria	40.76987	-73.92679	United States	US
99338	An East Williamsburg Delight	71564706996	VERIFIED	Petra	Brooklyn	Bedford-Stuyvesant	40.68335	-73.91222	United States	US
99339	Cozy bedroom in williamsburg	83101851929	VERIFIED	Russell	Brooklyn	Williamsburg	40.70870	-73.96710	United States	US
99340	Spacious Room w/ 2 Bed	69037598918	UNCONFIRMED	Christine	Queens	Bayside	40.75047	-73.75349	United States	US
99341	Room in Queens, NY, near LGA.	56457739998	VERIFIED	Sonia	Queens	East Elmhurst	40.76245	-73.87938	United States	US
99342 rows × 18 columns										

df.dtypes	
Name	object
Host_id	int64
Host_identity_verified	object
Host name	object
Neighbourhood group	object
Neighbourhood	object
Lat	float64
Long	float64
Country	object
Country code	object
Instant_bookable	int64
Cancellation_policy	object
Room type	object
Construction year	float64
Price	object
Service fee	object
Minimum nights	float64
Number of reviews	float64
dtype:	object

```
df["Price"] = df["Price"].str.replace("$", "")
df["Price"] = df["Price"].str.replace(",", "")
```

```
df["Price"] = df["Price"].astype(float)
```

```
df['Service fee'] = df['Service fee'].astype(float)
df.dtypes
```

Name	object
Host id	int64
Host_identity_verified	object
Host name	object
Neighbourhood group	object
Neighbourhood	object
Lat	float64
Long	float64
Country	object
Country code	object
Instant_bookable	int64
Cancellation_policy	object
Room type	object
Construction year	float64
Price	float64

Service feefloat64 Minimum nightsfloat64 Number of reviewsfloat64 dtype: object										
df['Total charge'] = df['Price'] + df['Service fee'] df										
	Name	Host id	Host_identity_verified	Host name	Neighbourhood group	Neighbourhood	Lat	Long	Country	Country code
0	Clean & quiet apt home by the park	80014485718	UNCONFIRMED	Madaline	Brooklyn	Kensington	40.64749	-73.97237	United States	US
1	Skyliit Midtown Castle	52335172823	VERIFIED	Jenna	Manhattan	Midtown	40.75362	-73.98377	United States	US
2	Entire Apt: Spacious Studio/Loft by central park	92037596077	VERIFIED	Lyndon	Manhattan	East Harlem	40.79851	-73.94399	United States	US
3	Large Cozy 1 BR Apartment In Midtown East	45498551794	VERIFIED	Michelle	Manhattan	Murray Hill	40.74767	-73.97500	United States	US
4	BlissArtsSpace!	90821839709	UNCONFIRMED	Emma	Brooklyn	Bedford-Stuyvesant	40.68688	-73.95596	United States	US
...
99337	Cozy room in a 2B with backyard access	5155155913	UNCONFIRMED	Cristina	Queens	Astoria	40.76987	-73.92679	United States	US
99999	An East	74561760000	VERIFIED	Don	Brooklyn	Bedford-	40.68688	-73.95596	United	US