

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.simplefilter(action='ignore', category=pd.errors.DataWarning)

In [2]: df=pd.read_csv("C:\Users\vajnu\Downloads\Airbnb_Data.csv")

In [3]: df.shape
Out[5]: (182599, 26)

In [7]: df.head()

Out[7]:
```

	id	NAME	host id	host_identity_verified	host name	neighbourhood group	neighbourhood	lat	long	country	...	service fee	minimum nights	number of reviews	last review	reviews per month	review rate number	calculated host listings count	availability 365	host type
0	1001254	Clean & quiet apt home by the park	80014485718	unconfirmed	Madaline	Brooklyn	Kensington	40.64749	-73.97237	United States	...	\$193	10.0	9.0	10/19/2021	0.21	4.0	6.0	286.0	CI
1	1002102	Skyline Midtown Castle	52335172823	verified	Jenna	Manhattan	Midtown	40.75362	-73.98377	United States	...	\$28	30.0	45.0	5/21/2022	0.38	4.0	2.0	228.0	C
2	1002403	THE VILLAGE HARLEM_NYC	78829239556	unconfirmed	Elise	Manhattan	Harlem	40.80902	-73.94190	United States	...	\$124	3.0	0.0	NaN	NaN	5.0	1.0	352.0	F
3	1002755	Entire Apt: Spacious Studio!off by central park	85098326012	unconfirmed	Garry	Brooklyn	Clinton Hill	40.68514	-73.95976	United States	...	\$74	30.0	270.0	7/5/2019	4.64	4.0	1.0	322.0	
4	1003689	Entire Apt: Spacious Studio!off by central park	92037596077	verified	Lyndon	Manhattan	East Harlem	40.79851	-73.94399	United States	...	\$41	10.0	9.0	11/19/2018	0.10	3.0	1.0	289.0	

5 rows x 26 columns

```
In [9]: df.columns
Out[9]: Index(['id', 'NAME', 'host id', 'host_identity_verified', 'host name', 'neighbourhood group', 'neighbourhood', 'lat', 'long', 'country', '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')

In [11]: df.isnull().sum()

Out[11]:
```

	id	NAME	host id	host_identity_verified	host name	neighbourhood group	neighbourhood	lat	long	country	...	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
id	0																					
NAME	210																					
host id	0																					
host_identity_verified	289																					
host name	486																					
neighbourhood group	29																					
neighbourhood	8																					
lat	8																					
long	8																					
country	532																					
country code	131																					
instant_bookable	185																					
cancellation_policy	76																					
room type	0																					
construction year	214																					
price	247																					
service fee	273																					
minimum nights	489																					
number of reviews	183																					
last review	15893																					
reviews per month	15879																					
review rate number	326																					
calculated host listings count	310																					
availability 365	448																					
house_rules	52133																					
license	182097																					
dtype:	int64																					
memory usage:	21.8+ MB																					

```
In [10]: df.info()
<class 'pandas.core.frame.DataFrame'>
Index: 182599 entries, 0 to 182598
Data columns (total 26 columns):
#   Column                Non-Null Count  Dtype  ---
#   0   id                    182599 non-null  int64   0
#   1   NAME                  182593 non-null  object  1
#   2   host id              182593 non-null  int64   2
#   3   host_identity_verified 182593 non-null  int64   3
#   4   host name            182593 non-null  object  4
#   5   neighbourhood group   182593 non-null  object  5
#   6   neighbourhood         182597 non-null  object  6
#   7   lat                  182593 non-null  float64 7
#   8   long                 182592 non-null  float64 8
#   9   country              182409 non-null  object  9
#  10  country code         182581 non-null  object  10
#  11  instant_bookable     182593 non-null  object  11
#  12  cancellation_policy   182593 non-null  object  12
#  13  room type            182593 non-null  object  13
#  14  construction year     182593 non-null  float64 14
#  15  price                182593 non-null  float64 15
#  16  service fee          182593 non-null  float64 16
#  17  minimum nights       182593 non-null  float64 17
#  18  number of reviews    182593 non-null  float64 18
#  19  last review          182593 non-null  datetime64[ns] 19
#  20  reviews per month    182593 non-null  float64 20
#  21  review rate number    182593 non-null  float64 21
#  22  calculated host listings count 182593 non-null  float64 22
#  23  availability 365       182593 non-null  float64 23
#  24  house_rules           182593 non-null  object  24
#  25  license               182593 non-null  object  25
dtypes: datetime64[ns](1), float64(12), int64(2), object(13)
memory usage: 21.8+ MB

In [24]: df.info()
<class 'pandas.core.frame.DataFrame'>
Index: 181949 entries, 0 to 182598
Data columns (total 26 columns):
#   Column                Non-Null Count  Dtype  ---
#   0   id                    181949 non-null  int64   0
#   1   NAME                  182593 non-null  object  1
#   2   host id              182593 non-null  int64   2
#   3   host_identity_verified 182593 non-null  int64   3
#   4   host name            182593 non-null  object  4
#   5   neighbourhood group   182593 non-null  object  5
#   6   neighbourhood         182597 non-null  object  6
#   7   lat                  182593 non-null  float64 7
#   8   long                 182592 non-null  float64 8
#   9   country              182409 non-null  object  9
#  10  country code         182581 non-null  object  10
#  11  instant_bookable     182593 non-null  object  11
#  12  cancellation_policy   182593 non-null  object  12
#  13  room type            182593 non-null  object  13
#  14  construction year     182593 non-null  float64 14
#  15  price                182593 non-null  float64 15
#  16  service fee          182593 non-null  float64 16
#  17  minimum nights       182593 non-null  float64 17
#  18  number of reviews    182593 non-null  float64 18
#  19  last review          182593 non-null  datetime64[ns] 19
#  20  reviews per month    182593 non-null  float64 20
#  21  review rate number    182593 non-null  float64 21
#  22  calculated host listings count 182593 non-null  float64 22
#  23  availability 365       182593 non-null  float64 23
#  24  house_rules           182593 non-null  object  24
#  25  license               182593 non-null  object  25
dtypes: datetime64[ns](1), float64(12), int64(2), object(13)
memory usage: 21.8+ MB

In [30]: df.fillna({'host_identity_verified':'unconfirmed',
                  'instant_bookable':False,
                  'number of reviews':0,
                  'review rate number':0,
                  'minimum nights':0},inplace=True)
df.fillna({'price':df['price'].mean(),
          'service fee':df['service fee'].mean(),
          'cancellation_policy':'strict',
          'construction year':df['construction year'].mean(),
          'availability 365':df['availability 365'].min(),
          'calculated host listings count':df['calculated host listings count'].min()},inplace=True)
# filling null values

In [32]: df.isnull().sum()

Out[32]:
```

	id	NAME	host id	host_identity_verified	host name	neighbourhood group	neighbourhood	lat	long	country	...	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
id	0																					
NAME	0																					
host id	0																					
host_identity_verified	0																					
host name	0																					
neighbourhood group	26																					
neighbourhood	16																					
lat	8																					
long	8																					
country	526																					
country code	122																					
instant_bookable	0																					
cancellation_policy	0																					
room type	0																					
price	0																					
service fee	0																					
minimum nights	0																					
number of reviews	0																					
last review	0																					
reviews per month	0																					
review rate number	0																					
calculated host listings count	0																					
availability 365	0																					
dtype:	int64																					

```
In [34]: df['service fee']=df['service fee'].replace({'N/A':''},regex=True,astype=float)
df['price']=df['price'].replace({'N/A':''},regex=True,astype=float)

In [34]: df['cancellation_policy'].isnull().sum()

Out[34]: 0

In [36]: df.head()

Out[36]:
```

	id	NAME	host id	host_identity_verified	host name	neighbourhood group	neighbourhood	lat	long	country	...	Construction year	price	service fee	minimum nights	number of reviews	last review	reviews per month	review rate number	calculated host listings count	house_type
0	1001254	Clean & quiet apt home by the park	80014485718	unconfirmed	Madaline	Brooklyn	Kensington	40.64749	-73.97237	United States	...	2020.0	966.0	193.0	10.0	9.0	2021-10-19	0.21	4.0	6.0	
1	1002102	Skyline Midtown Castle	52335172823	verified	Jenna	Manhattan	Midtown	40.75362	-73.98377	United States	...	2007.0	142.0	28.0	30.0	45.0	2022-05-21	0.38	4.0	2.0	
2	1002403	THE VILLAGE HARLEM_NYC	78829239556	unconfirmed	Elise	Manhattan	Harlem	40.80902	-73.94190	United States	...	2005.0	620.0	124.0	3.0	0.0	2012-07-11	0.00	5.0	1.0	
4	1003689	Entire Apt: Spacious Studio!off by central park	92037596077	verified	Lyndon	Manhattan	East Harlem	40.79851	-73.94399	United States	...	2009.0	204.0	41.0	10.0	9.0	2018-11-19	0.10	3.0	1.0	
5	1004098	Large Cozy 1 BR Apartment in Midtown East	45498551794	verified	Michelle	Manhattan	Murray Hill	40.74767	-73.97500	United States	...	2013.0	577.0	115.0	3.0	74.0	2019-06-22	0.59	3.0	1.0	

5 rows x 24 columns



