

```
In [1]: import numpy as np
import pandas as pd
```

```
In [2]: import matplotlib.pyplot as plt
import warnings
from mlxtend.plotting import plot_confusion_matrix
from plotly.offline import plot, iplot, init_notebook_mode
import plotly.graph_objs as go
from plotly.subplots import make_subplots
import plotly.express as px
init_notebook_mode(connected=True)
warnings.filterwarnings("ignore")
%matplotlib inline
import datetime as dt
```

```
In [3]: # import data
AIRBNB = pd.read_csv('AB_NYC_2019.csv')
AIRBNB.head()
```

Out[3]:

	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
0	2539	Clean & quiet apt home by the park	2787	John	Brooklyn	Kensington	40.64749	-73.97237	Private room	149	1	9	2018-10-19	0.21	6	365
1	2595	Skylit Midtown Castle	2845	Jennifer	Manhattan	Midtown	40.75362	-73.98377	Entire home/apt	225	1	45	2019-05-21	0.38	2	355
2	3647	THE VILLAGE OF HARLEM....NEW YORK !	4632	Elisabeth	Manhattan	Harlem	40.80902	-73.94190	Private room	150	3	0	NaN	NaN	1	365
3	3831	Cozy Entire Floor of Brownstone	4869	LisaRoxanne	Brooklyn	Clinton Hill	40.68514	-73.95976	Entire home/apt	89	1	270	2019-07-05	4.64	1	194
4	5022	Entire Apt: Spacious Studio/Loft by central park	7192	Laura	Manhattan	East Harlem	40.79851	-73.94399	Entire home/apt	80	10	9	2018-11-19	0.10	1	0

```
In [4]: AIRBNB.tail()
```

Out[4]:

	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
48890	36484665	Charming one bedroom - newly renovated rowhouse	8232441	Sabrina	Brooklyn	Bedford-Stuyvesant	40.67853	-73.94995	Private room	70	2	0	NaN	NaN	2	9
48891	36485057	Affordable room in Bushwick/East Williamsburg	6570630	Marisol	Brooklyn	Bushwick	40.70184	-73.93317	Private room	40	4	0	NaN	NaN	2	36
48892	36485431	Sunny Studio at Historical Neighborhood	23492952	Ilgar & Aysel	Manhattan	Harlem	40.81475	-73.94867	Entire home/apt	115	10	0	NaN	NaN	1	27
48893	36485609	43rd St. Time Square-cozy single bed	30985759	Taz	Manhattan	Hell's Kitchen	40.75751	-73.99112	Shared room	55	1	0	NaN	NaN	6	2
48894	36487245	Trendy duplex in the very heart of Hell's Kitchen	68119814	Christophe	Manhattan	Hell's Kitchen	40.76404	-73.98933	Private room	90	7	0	NaN	NaN	1	23

```
In [5]: print("El tamaño del dataset es: ",AIRBNB.shape)
print("Info del dataset: ",AIRBNB.info())
```

El tamaño del dataset es: (48895, 16)
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 48895 entries, 0 to 48894
Data columns (total 16 columns):
Column Non-Null Count Dtype
--- ---
0 id 48895 non-null int64
1 name 48879 non-null object
2 host_id 48895 non-null int64
3 host_name 48874 non-null object
4 neighbourhood_group 48895 non-null object
5 neighbourhood 48895 non-null object
6 latitude 48895 non-null float64
7 longitude 48895 non-null float64
8 room_type 48895 non-null object
9 price 48895 non-null int64
10 minimum_nights 48895 non-null int64
11 number_of_reviews 48895 non-null int64
12 last_review 38843 non-null object
13 reviews_per_month 38843 non-null float64
14 calculated_host_listings_count 48895 non-null int64
15 availability_365 48895 non-null int64
dtypes: float64(3), int64(7), object(6)
memory usage: 4.8+ MB
Info del dataset: None

```
In [6]: total_of_all = AIRBNB.isnull().sum().sort_values(ascending=False)
percent_of_all = (AIRBNB.isnull().sum()/AIRBNB.isnull().count()).sort_values(ascending=False)
missing_data_test = pd.concat([total_of_all, percent_of_all], axis=1, keys=['Total Nulos', 'Percent'])
missing_data_test.head()
```

Out[6]:

	Total Nulos	Percent
last_review	10052	0.205583
reviews_per_month	10052	0.205583
host_name	21	0.000429
name	16	0.000327
id	0	0.000000

```
In [9]: dimHost = pd.DataFrame(AIRBNB, columns = ['id','host_id', 'host_name', 'price', 'minimum_nights', 'number_of_reviews'])
dimHousing = pd.DataFrame(AIRBNB, columns = ['id', 'name', 'neighbourhood_group', 'neighbourhood', 'room_type', 'last_review'])
tableFact = pd.DataFrame(AIRBNB, columns = ['id', 'host_id', 'latitude', 'longitude', 'reviews_per_month', 'calculated_host_listings_count', 'availability_365'])
```

```
In [10]: dimHost.to_csv('dimhost.csv',index=False)
dimHousing.to_csv('dimhousing.csv',index=False)
tableFact.to_csv('tableFact.csv',index=False)
```

```
In [ ]:
```