CHOOSING THE NY BEST HOLIDAY STAY

CHOOSING OUR DESIRES

- Generally when choosing a trip, voyagers try to stay at calm places while they are near to visit interesting places BUT, the Price of the residence is a general research among visitors world wide.

- So in a trip research you must look for an interesting place to go, a way to make it cheap and try to make it safe.

- For this reasons, our research will be centered in the city of New york and will be focused in 2 main points:
    - Price of airbnbs
    - Neighbourhood security

DATA ACQUISITION AND CLEANING

- Airbnb pricing with it's location csv:

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

- Criminality per neighbourhood csv:

- https://data.world/data-society/nyc-crime-data

FIRST WE GET THE TABLE OF AIRBNBS

	neighbourhood_group	latitude	longitude	price	minimum_nights	availability_365
0	Brooklyn	40.64749	-73.97237	149	1	365
1	Manhattan	40.75362	-73.98377	225	1	355
2	Manhattan	40.80902	-73.94190	150	3	365
3	Brooklyn	40.68514	-73.95976	89	1	194
4	Manhattan	40.79851	-73.94399	80	10	0
...	...	...	...	...	...	...
48890	Brooklyn	40.67853	-73.94995	70	2	9
48891	Brooklyn	40.70184	-73.93317	40	4	36
48892	Manhattan	40.81475	-73.94867	115	10	27
48893	Manhattan	40.75751	-73.99112	55	1	2
48894	Manhattan	40.76404	-73.98933	90	7	23

48895 rows × 6 columns

	neighbourhood_group	latitude	longitude	price	minimum_nights	availability_365
0	Brooklyn	40.64749	-73.97237	149	1	365
1	Manhattan	40.75362	-73.98377	225	1	355
3	Brooklyn	40.68514	-73.95976	89	1	194
9	Manhattan	40.71344	-73.99037	150	1	188
17	Manhattan	40.74192	-73.99501	140	1	12
...	...	...	...	...	...	...
48878	Queens	40.74477	-73.87727	45	1	172
48880	Brooklyn	40.71232	-73.94220	120	1	365
48887	Brooklyn	40.71232	-73.94220	170	1	365
48889	Queens	40.69137	-73.80844	65	1	163
48893	Manhattan	40.75751	-73.99112	55	1	2

	neighbourhood_group	latitude	longitude	price	availability_365
0	Brooklyn	40.64749	-73.97237	149	365
1	Manhattan	40.75362	-73.98377	225	355
3	Brooklyn	40.68514	-73.95976	89	194
9	Manhattan	40.71344	-73.99037	150	188
22	Brooklyn	40.67989	-73.97798	80	364
...	...	...	...	...	...
48851	Queens	40.69137	-73.86591	35	341
48867	Brooklyn	40.59080	-73.97116	25	338
48873	Brooklyn	40.63116	-73.92616	170	363
48880	Brooklyn	40.71232	-73.94220	120	365
48887	Brooklyn	40.71232	-73.94220	170	365

AND ASSURING AVAILABILITY

NOW FROM THE CRIMINALITY WE GET THE CRIME INTENSITY

	CMPLNT_NUM	CMPLNT_FR_DT	CMPLNT_FR_TM	CMPLNT_TO_DT	CMPLNT_TO_TM	RPT_DT
0	736216184	09/30/2016	23:25:00	09/30/2016	23:25:00	09/30/201
1	294332956	09/30/2016	23:16:00	09/30/2016	23:21:00	09/30/201
2	852981427	09/30/2016	23:00:00	09/30/2016	23:05:00	09/30/201
3	369976063	09/30/2016	23:00:00	NaN	NaN	09/30/201
4	117213771	09/30/2016	23:00:00	09/30/2016	23:10:00	09/30/201
...	...	...	...	...	...	...
24281	317497826	09/12/2016	19:40:00	09/12/2016	19:43:00	09/12/201
24282	827964069	09/12/2016	19:35:00	09/12/2016	19:40:00	09/12/201
24283	585540160	09/12/2016	19:35:00	09/12/2016	19:40:00	09/12/201
24284	924661366	09/12/2016	19:30:00	09/12/2016	19:45:00	09/12/201
24285	162692438	09/12/2016	19:30:00	09/12/2016	19:50:00	09/12/201

24286 rows × 24 columns

	BORO_NM
0	BRONX
1	BROOKLYN
2	BRONX
3	QUEENS
4	QUEENS
...	...
24281	BROOKLYN
24282	MANHATTAN
24283	BRONX
24284	MANHATTAN
24285	BRONX

24286 rows × 1 columns

	0
BORO_NM	
BRONX	5289
BROOKLYN	7345
MANHATTAN	5780
QUEENS	4736
STATEN ISLAND	1136

AFTER NORMALIZATION OF CRIME INTENSITY AND MERGING TABLES WE GET THIS

	neighbourhood_group	latitude	longitude	price	criminality_index
0	Brooklyn	40.64749	-73.97237	149	1.000000
1	Brooklyn	40.68514	-73.95976	89	1.000000
2	Brooklyn	40.67989	-73.97798	80	1.000000
3	Brooklyn	40.68310	-73.95473	70	1.000000
4	Brooklyn	40.63702	-73.96327	150	1.000000
...	...	...	...	...	...
3490	Queens	40.75104	-73.81459	45	0.579804
3491	Queens	40.73710	-73.88086	50	0.579804
3492	Queens	40.69185	-73.86431	37	0.579804
3493	Queens	40.69183	-73.86523	34	0.579804
3494	Queens	40.69137	-73.86591	35	0.579804

3495 rows × 5 columns

WE WILL FIT THE MODEL WITH PRICE AND CRIMINALITY INDEX FROM THIS

	price	criminality_index
0	149	1.000000
1	89	1.000000
2	80	1.000000
3	70	1.000000
4	150	1.000000
...	...	...
3490	45	0.579804
3491	50	0.579804
3492	37	0.579804
3493	34	0.579804
3494	35	0.579804

3495 rows × 2 columns

THEN WE GET THE CLUSTER LABELS FOR EACH AIRBNB. DUE TO EXCESSIVE DATA WE SELECT ARBITRARLY AN 40% OF THE SAMPLES

	Cluster_Labels	neighbourhood_group	latitude	longitude	price	criminality_index
2949	9	Queens	40.68090	-73.78196	50	0.579804
1348	0	Manhattan	40.77550	-73.94750	130	0.747947
2899	0	Queens	40.72363	-73.80285	100	0.579804
1699	4	Manhattan	40.79099	-73.94500	400	0.747947
228	9	Brooklyn	40.68517	-73.93473	30	1.000000
...	...	...	...	...	...	...
1948	0	Manhattan	40.72546	-73.98396	179	0.747947
93	9	Brooklyn	40.69023	-73.94628	75	1.000000
542	4	Brooklyn	40.67476	-73.93016	550	1.000000
2757	9	Queens	40.71293	-73.84915	70	0.579804
1236	0	Manhattan	40.80034	-73.96638	105	0.747947

1398 rows × 6 columns

AS FINAL RESULT WE CONCLUDE WITH THE BEST ÁREA BEING MANHATTAN AS SEEN IN THE PICTURE