



In [3]:

```
import pandas as pd

data=pd.read_csv(r'C:\Users\Sid\Desktop\python files\glass prediction KNN from scratch\ufo.
print(data.head())
```

	City	Colors	Reported Shape	Reported State	Tim
0	Ithaca	NaN	TRIANGLE	NY	6/1/1930 22:0
1	Willingboro	NaN	OTHER	NJ	6/30/1930 20:0
2	Holyoke	NaN	OVAL	CO	2/15/1931 14:0
3	Abilene	NaN	DISK	KS	6/1/1931 13:0
4	New York Worlds Fair	NaN	LIGHT	NY	4/18/1933 19:0



In [4]:

#1st Method

```
print(data.isnull().head())
print(data.isnull())
```

	City	Colors	Reported	Shape	Reported	State	Time
0	False		True		False	False	False
1	False		True		False	False	False
2	False		True		False	False	False
3	False		True		False	False	False
4	False		True		False	False	False
5	False		True		False	False	False
6	False		True		False	False	False
7	False		True		False	False	False
8	False		True		False	False	False
9	False		True		False	False	False
10	False		True		False	False	False
11	False		True		False	False	False
12	False		False		False	False	False
13	False		True		False	False	False
14	False		True		False	False	False
15	False		True		False	False	False
16	False		True		True	False	False
17	False		True		True	False	False
18	False		True		False	False	False
19	False		False		False	False	False
20	False		True		False	False	False
21	True		True		True	False	False
22	True		True		False	False	False
23	False		True		False	False	False
24	False		True		False	False	False
25	False		True		False	False	False
26	False		True		False	False	False
27	False		True		False	False	False
28	False		True		False	False	False
29	False		True		False	False	False
...
18211	False		True		False	False	False
18212	False		True		False	False	False
18213	False		False		False	False	False
18214	False		True		False	False	False
18215	False		True		False	False	False
18216	False		False		False	False	False
18217	False		True		False	False	False
18218	False		True		False	False	False
18219	False		True		False	False	False
18220	False		False		False	False	False
18221	False		True		False	False	False
18222	False		True		False	False	False
18223	False		True		True	False	False

18224	False	True	False	False	False
18225	False	True	False	False	False
18226	False	True	False	False	False
18227	False	True	False	False	False
18228	False	True	False	False	False
18229	False	True	False	False	False
18230	False	True	False	False	False
18231	False	True	False	False	False
18232	False	True	True	False	False
18233	False	False	False	False	False
18234	False	True	False	False	False
18235	False	True	True	False	False
18236	False	True	False	False	False
18237	False	True	False	False	False
18238	False	True	True	False	False
18239	False	False	False	False	False
18240	False	True	False	False	False

[18241 rows x 5 columns]



In [5]:

```
#2nd method: Dropping missing values
print('no: of rows and columns before removing null values')
print('rows, columns: ', data.shape)
print('First 5 rows')
print(data.head())
print('-----')

print('after null values have been removed')
print(data.dropna(subset=['City', 'Shape Reported', 'Colors Reported'], how='any').head())

print('no: of rows and columns after removing null values')
print('rows, columns', data.shape)

print((data))
```

no: of rows and columns before removing null values

rows, columns: (18241, 5)

First 5 rows

	City	Colors Reported	Shape Reported	State	Time
0	Ithaca	NaN	TRIANGLE	NY	6/1/1930 22:0
1	Willingboro	NaN	OTHER	NJ	6/30/1930 20:0
2	Holyoke	NaN	OVAL	CO	2/15/1931 14:0
3	Abilene	NaN	DISK	KS	6/1/1931 13:0
4	New York Worlds Fair	NaN	LIGHT	NY	4/18/1933 19:0

after null values have been removed

	City	Colors Reported	Shape Reported	State	Time
12	Belton	RED	SPHERE	SC	6/30/1939 20:00
19	Bering Sea	RED	OTHER	AK	4/30/1943 23:00
36	Portsmouth	RED	FORMATION	VA	7/10/1945 1:30
44	Blairsdon	GREEN	SPHERE	CA	6/30/1946 19:00
82	San Jose	BLUE	CHEVRON	CA	7/15/1947 21:00

no: of rows and columns after removing null values

rows, columns (18241, 5)

	City	Colors Reported	Shape Reported	State	\
0	Ithaca	NaN	TRIANGLE	NY	
1	Willingboro	NaN	OTHER	NJ	
2	Holyoke	NaN	OVAL	CO	
3	Abilene	NaN	DISK	KS	
4	New York Worlds Fair	NaN	LIGHT	NY	
5	Valley City	NaN	DISK	ND	
6	Crater Lake	NaN	CIRCLE	CA	
7	Alma	NaN	DISK	MI	
8	Eklutna	NaN	CIGAR	AK	
9	Hubbard	NaN	CYLINDER	OR	
10	Fontana	NaN	LIGHT	CA	
11	Waterloo	NaN	FIREBALL	AL	
12	Belton	RED	SPHERE	SC	
13	Keokuk	NaN	OVAL	IA	
14	Ludington	NaN	DISK	MI	

15	Forest Home	NaN	CIRCLE	CA
16	Los Angeles	NaN	NaN	CA
17	Hapeville	NaN	NaN	GA
18	Oneida	NaN	RECTANGLE	TN
19	Bering Sea	RED	OTHER	AK
20	Nebraska	NaN	DISK	NE
21	NaN	NaN	NaN	LA
22	NaN	NaN	LIGHT	LA
23	Owensboro	NaN	RECTANGLE	KY
24	Wilderness	NaN	DISK	WV
25	San Diego	NaN	CIGAR	CA
26	Wilderness	NaN	DISK	WV
27	Clovis	NaN	DISK	NM
28	Los Alamos	NaN	DISK	NM
29	Ft. Duschene	NaN	DISK	UT
...
18211	Holyoke	NaN	DIAMOND	MA
18212	Carson	NaN	DISK	CA
18213	Pasadena	GREEN	FIREBALL	CA
18214	Austin	NaN	FORMATION	TX
18215	El Campo	NaN	OTHER	TX
18216	Garden Grove	ORANGE	LIGHT	CA
18217	Berthoud Pass	NaN	TRIANGLE	CO
18218	Sisterdale	NaN	DIAMOND	TX
18219	Garden Grove	NaN	CHEVRON	CA
18220	Shasta Lake	BLUE	DISK	CA
18221	Franklin	NaN	DISK	NH
18222	Albrightsville	NaN	OTHER	PA
18223	Greenville	NaN	NaN	SC
18224	Eufaula	NaN	DISK	OK
18225	Simi Valley	NaN	FORMATION	CA
18226	San Francisco	NaN	FORMATION	CA
18227	San Francisco	NaN	TRIANGLE	CA
18228	Kingsville	NaN	LIGHT	TX
18229	Chicago	NaN	DISK	IL
18230	Pismo Beach	NaN	OVAL	CA
18231	Pismo Beach	NaN	OVAL	CA
18232	Lodi	NaN	NaN	WI
18233	Anchorage	RED	VARIOUS	AK
18234	Capitola	NaN	TRIANGLE	CA
18235	Fountain Hills	NaN	NaN	AZ
18236	Grant Park	NaN	TRIANGLE	IL
18237	Spirit Lake	NaN	DISK	IA
18238	Eagle River	NaN	NaN	WI
18239	Eagle River	RED	LIGHT	WI
18240	Ybor	NaN	OVAL	FL

Time

0	6/1/1930 22:00
1	6/30/1930 20:00
2	2/15/1931 14:00
3	6/1/1931 13:00
4	4/18/1933 19:00
5	9/15/1934 15:30
6	6/15/1935 0:00
7	7/15/1936 0:00
8	10/15/1936 17:00
9	6/15/1937 0:00
10	8/15/1937 21:00
11	6/1/1939 20:00
12	6/30/1939 20:00

```
13      7/7/1939  2:00
14      6/1/1941 13:00
15      7/2/1941 11:30
16      2/25/1942 0:00
17      6/1/1942 22:30
18      7/15/1942 1:00
19      4/30/1943 23:00
20      6/1/1943 15:00
21      8/15/1943 0:00
22      8/15/1943 0:00
23     10/15/1943 11:00
24      1/1/1944 10:00
25      1/1/1944 12:00
26      1/1/1944 12:00
27      4/2/1944 11:00
28      6/1/1944 12:00
29      6/30/1944 10:00
...
18211 12/28/2000 18:00
18212 12/28/2000 18:20
18213 12/28/2000 19:10
18214 12/29/2000 0:00
18215 12/29/2000 9:00
18216 12/29/2000 16:10
18217 12/29/2000 19:30
18218 12/29/2000 20:00
18219 12/29/2000 20:30
18220 12/29/2000 20:30
18221 12/29/2000 20:50
18222 12/29/2000 21:00
18223 12/29/2000 22:00
18224 12/29/2000 23:30
18225 12/30/2000 10:00
18226 12/30/2000 22:00
18227 12/30/2000 22:00
18228 12/31/2000 4:00
18229 12/31/2000 11:45
18230 12/31/2000 20:00
18231 12/31/2000 20:00
18232 12/31/2000 20:30
18233 12/31/2000 21:00
18234 12/31/2000 22:00
18235 12/31/2000 23:00
18236 12/31/2000 23:00
18237 12/31/2000 23:00
18238 12/31/2000 23:45
18239 12/31/2000 23:45
18240 12/31/2000 23:59
```

[18241 rows x 5 columns]



In []:

