

```
In [2]: import pandas as pd
ufo = pd.read_csv('C:\\Users\\Apeh\\Desktop\\CODE\\DATASET\\ufo.csv')
ufo.head()
```

```
Out[2]:
```

	City	Colors Reported	Shape Reported	State	Time
0	Ithaca	NaN	TRIANGLE	NY	6/1/1930 22:00
1	Willingboro	NaN	OTHER	NJ	6/30/1930 20:00
2	Holyoke	NaN	OVAl	CO	2/15/1931 14:00
3	Abilene	NaN	DISK	KS	6/1/1931 13:00
4	New York Worlds Fair	NaN	LIGHT	NY	4/18/1933 19:00

```
In [3]: # EXPLORE THE DATA
ufo.columns
```

```
Out[3]: Index(['City', 'Colors Reported', 'Shape Reported', 'State', 'Time'], dtype='object')
```

```
In [4]: ufo.shape
```

```
Out[4]: (18241, 5)
```

```
In [5]: # CHECKING THE MISSING DATA
ufo.isnull().head(10)
```

```
Out[5]:
```

	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

```
In [6]: # CHECKING THE VALUE OF MISSING DATA
ufo.isnull().sum()
```

```
Out[6]: City                25
Colors Reported    15359
```

Shape Reported 2644  
State 0  
Time 0  
dtype: int64

```
In [8]: # PRINT COL THAT INVOLVE IN MISSING DATA
ufo[ufo.City.isnull()]
```

Out[8]:

	City	Colors Reported	Shape Reported	State	Time
21	NaN	NaN	NaN	LA	8/15/1943 0:00
22	NaN	NaN	LIGHT	LA	8/15/1943 0:00
204	NaN	NaN	DISK	CA	7/15/1952 12:30
241	NaN	BLUE	DISK	MT	7/4/1953 14:00
613	NaN	NaN	DISK	NV	7/1/1960 12:00
1877	NaN	YELLOW	CIRCLE	AZ	8/15/1969 1:00
2013	NaN	NaN	NaN	NH	8/1/1970 9:30
2546	NaN	NaN	FIREBALL	OH	10/25/1973 23:30
3123	NaN	RED	TRIANGLE	WV	11/25/1975 23:00
4736	NaN	NaN	SPHERE	CA	6/23/1982 23:00
5269	NaN	NaN	NaN	AZ	6/30/1985 21:30
6735	NaN	NaN	FORMATION	TX	4/1/1992 2:00
7208	NaN	NaN	CIRCLE	MI	10/4/1993 17:30
8828	NaN	NaN	TRIANGLE	WA	10/30/1995 21:30
8967	NaN	NaN	VARIOUS	CA	12/8/1995 18:00
9273	NaN	NaN	TRIANGLE	OH	5/1/1996 3:00
9388	NaN	NaN	OVAL	CA	6/12/1996 12:00
9587	NaN	NaN	EGG	FL	8/24/1996 15:00
10399	NaN	NaN	TRIANGLE	IL	6/15/1997 23:00
11625	NaN	NaN	CIRCLE	TX	6/7/1998 7:00
12441	NaN	RED	FIREBALL	WA	10/26/1998 17:58
15767	NaN	NaN	RECTANGLE	NV	1/21/2000 11:30
15812	NaN	NaN	LIGHT	NV	2/2/2000 3:00
16054	NaN	GREEN	NaN	FL	3/11/2000 3:30
16608	NaN	NaN	SPHERE	NY	6/15/2000 15:00

```
In [9]: # HANDLING THE MISSING DATA START WITH COLUMN
ufo.columns
```

```
Out[9]: Index(['City', 'Colors Reported', 'Shape Reported', 'State', 'Time'], dtype='object')
```

```
In [10]: ufo.columns = ufo.columns.str.strip().str.lower().str.replace(' ', '_')
         ufo.columns
```

```
Out[10]: Index(['city', 'colors_reported', 'shape_reported', 'state', 'time'], dtype='object')
```

```
In [11]: ufo[ufo.shape_reported.isnull()].head(3)
```

```
Out[11]:
```

	city	colors_reported	shape_reported	state	time
16	Los Angeles	NaN	NaN	CA	2/25/1942 0:00
17	Hapeville	NaN	NaN	GA	6/1/1942 22:30
21	NaN	NaN	NaN	LA	8/15/1943 0:00

```
In [12]: ufo.shape
```

```
Out[12]: (18241, 5)
```

```
In [13]: ufo.dropna().shape
```

```
Out[13]: (2486, 5)
```

```
In [18]: # ANOTHER APPROACH
         ufo['shape_reported'].value_counts(dropna=False)
```

```
Out[18]:
```

VARIOUS	2977
LIGHT	2803
DISK	2122
TRIANGLE	1889
OTHER	1402
CIRCLE	1365
SPHERE	1054
FIREBALL	1039
OVAl	845
CIGAR	617
FORMATION	434
RECTANGLE	303
CYLINDER	294
CHEVRON	248
DIAMOND	234
EGG	197
FLASH	188
TEARDROP	119
CONE	60
CROSS	36
DELTA	7
CRESCENT	2
ROUND	2
FLARE	1
PYRAMID	1
HEXAGON	1
DOME	1

Name: shape\_reported, dtype: int64

```
In [19]: ufo['shape_reported'].fillna(value='VARIOUS',inplace=True)
ufo['shape_reported'].value_counts()
```

```
Out[19]: VARIOUS      2977
LIGHT        2803
DISK         2122
TRIANGLE     1889
OTHER        1402
CIRCLE       1365
SPHERE       1054
FIREBALL     1039
OVAL         845
CIGAR        617
FORMATION    434
RECTANGLE    303
CYLINDER     294
CHEVRON      248
DIAMOND      234
EGG          197
FLASH        188
TEARDROP     119
CONE         60
CROSS        36
DELTA        7
CRESCENT     2
ROUND        2
FLARE        1
PYRAMID      1
HEXAGON      1
DOME         1
Name: shape_reported, dtype: int64
```

```
In [20]: # CHECK ANY MISSING VA AFTER CLEANING
ufo.isnull().sum()
```

```
Out[20]: city          25
colors_reported    15359
shape_reported      0
state              0
time              0
dtype: int64
```

```
In [21]: # COLORS
ufo['colors_reported'].value_counts(dropna=False)
```

```
Out[21]: NaN          15359
RED          780
GREEN        531
ORANGE       528
BLUE         450
YELLOW       169
RED GREEN    89
RED BLUE     78
RED ORANGE   44
GREEN BLUE   34
RED GREEN BLUE 33
ORANGE YELLOW 26
RED YELLOW   25
ORANGE GREEN 23
YELLOW GREEN 17
ORANGE BLUE  10
```

```

RED YELLOW GREEN          9
YELLOW BLUE               6
ORANGE GREEN BLUE        5
YELLOW GREEN BLUE         5
RED ORANGE YELLOW         4
RED YELLOW GREEN BLUE     4
RED ORANGE BLUE           3
RED ORANGE GREEN          3
RED YELLOW BLUE           3
ORANGE YELLOW GREEN       1
ORANGE YELLOW BLUE        1
RED ORANGE YELLOW BLUE    1
Name: colors_reported, dtype: int64

```

```

In [22]: ufo['colors_reported'].fillna(value='UNIDENTIFIED COLORS',inplace=True)
         ufo['colors_reported'].value_counts()

```

```

Out[22]: UNIDENTIFIED COLORS    15359
         RED                   780
         GREEN                 531
         ORANGE                528
         BLUE                  450
         YELLOW                169
         RED GREEN             89
         RED BLUE              78
         RED ORANGE            44
         GREEN BLUE            34
         RED GREEN BLUE        33
         ORANGE YELLOW         26
         RED YELLOW            25
         ORANGE GREEN          23
         YELLOW GREEN          17
         ORANGE BLUE           10
         RED YELLOW GREEN      9
         YELLOW BLUE           6
         ORANGE GREEN BLUE     5
         YELLOW GREEN BLUE     5
         RED ORANGE YELLOW     4
         RED YELLOW GREEN BLUE 4
         RED YELLOW BLUE       3
         RED ORANGE GREEN      3
         RED ORANGE BLUE       3
         ORANGE YELLOW GREEN   1
         ORANGE YELLOW BLUE    1
         RED ORANGE YELLOW BLUE 1
Name: colors_reported, dtype: int64

```

```

In [23]: ufo.isnull().sum()

```

```

Out[23]: city                25
         colors_reported      0
         shape_reported      0
         state                0
         time                 0
         dtype: int64

```