```
import pandas as pd
df=pd.read csv("C:/Users/Sruth/Downloads/sales data sample.csv")
 print(df.head())
   ORDERNUMBER QUANTITYORDERED
                                  PRICEEACH ORDERLINENUMBER
SALES \
         10107
                              30
                                      95.70
                                                             2
                                                                2871.00
         10121
                              34
                                      81.35
                                                             5
                                                                2765.90
                                      94.74
         10134
                              41
                                                                3884.34
3
         10145
                              45
                                      83.26
                                                             6
                                                                3746.70
                              49
                                     100.00
                                                            14
                                                                5205.27
         10159
          ORDERDATE
                       STATUS
                               QTR ID
                                        MONTH ID
                                                  YEAR ID
                      Shipped
                                                     2003
     2/24/2003 0:00
                                               2
                                    1
   05-07-2003 00:00
                      Shipped
                                     2
                                               5
                                                     2003
1
                                               7
2
   07-01-2003 00:00
                                    3
                                                     2003
                      Shipped
3
     8/25/2003 0:00
                      Shipped
                                     3
                                               8
                                                     2003
4
   10-10-2003 00:00
                                     4
                     Shipped
                                              10
                                                     2003
                     ADDRESSLINE1
                                   ADDRESSLINE2
                                                            CITY STATE \
                                                                    NY
0
         897 Long Airport Avenue
                                             NaN
                                                             NYC
1
              59 rue de l'Abbaye
                                             NaN
                                                           Reims
                                                                   NaN
2
   27 rue du Colonel Pierre Avia
                                             NaN
                                                           Paris
                                                                   NaN
3
              78934 Hillside Dr.
                                             NaN
                                                       Pasadena
                                                                    CA
4
                 7734 Strong St.
                                             NaN
                                                  San Francisco
                                                                    CA
  POSTALCODE COUNTRY TERRITORY CONTACTLASTNAME CONTACTFIRSTNAME
DEALSIZE
       10022
                 USA
                                              Yu
                                                              Kwai
                            NaN
Small
       51100 France
                           EMEA
                                         Henriot
                                                              Paul
1
Small
       75508 France
                           EMEA
                                        Da Cunha
                                                            Daniel
Medium
       90003
                 USA
                            NaN
                                           Young
                                                             Julie
Medium
         NaN
                 USA
                            NaN
                                                             Julie
                                           Brown
Medium
[5 rows x 25 columns]
df.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2823 entries, 0 to 2822
```

```
Data columns (total 25 columns):
#
     Column
                        Non-Null Count
                                         Dtype
                                         int64
 0
     ORDERNUMBER
                        2823 non-null
 1
     QUANTITYORDERED
                        2823 non-null
                                         int64
 2
     PRICEEACH
                        2823 non-null
                                         float64
 3
     ORDERLINENUMBER
                        2823 non-null
                                         int64
 4
                        2823 non-null
                                         float64
     SALES
 5
                                         object
     ORDERDATE
                        2823 non-null
 6
     STATUS
                        2823 non-null
                                         object
 7
     QTR ID
                        2823 non-null
                                         int64
     MONTH ID
 8
                        2823 non-null
                                         int64
 9
     YEAR ID
                        2823 non-null
                                         int64
 10
     PRODUCTLINE
                        2823 non-null
                                         object
 11
     MSRP
                        2823 non-null
                                         int64
 12
     PRODUCTCODE
                        2823 non-null
                                         object
 13
     CUSTOMERNAME
                        2823 non-null
                                         object
 14
     PHONE
                        2823 non-null
                                         object
 15
    ADDRESSLINE1
                        2823 non-null
                                         object
    ADDRESSLINE2
                        302 non-null
                                         object
 16
                        2823 non-null
 17
     CITY
                                         object
 18
    STATE
                        1337 non-null
                                         object
                                         object
 19
    POSTALCODE 
                        2747 non-null
 20
    COUNTRY
                        2823 non-null
                                         object
 21
     TERRITORY
                        1749 non-null
                                         object
                        2823 non-null
 22
     CONTACTLASTNAME
                                         object
 23
     CONTACTFIRSTNAME
                        2823 non-null
                                         object
24
     DEALSIZE
                        2823 non-null
                                         object
dtypes: float64(2), int64(7), object(16)
memory usage: 551.5+ KB
df.describe()
        ORDERNUMBER
                      QUANTITYORDERED
                                          PRICEEACH
                                                      ORDERLINENUMBER \
                          2823.000000
count
        2823.000000
                                        2823.000000
                                                          2823.000000
       10258.725115
                            35.092809
                                          83.658544
                                                             6.466171
mean
                             9.741443
                                          20.174277
                                                             4.225841
std
          92.085478
       10100.000000
                             6.000000
                                          26.880000
                                                             1.000000
min
                            27.000000
25%
       10180.000000
                                          68.860000
                                                             3.000000
       10262.000000
                            35.000000
                                          95.700000
50%
                                                             6.000000
75%
       10333.500000
                            43,000000
                                         100.000000
                                                             9.000000
                            97.000000
                                         100.000000
max
       10425.000000
                                                            18.000000
                                                                      MSRP
              SALES
                           QTR ID
                                       MONTH ID
                                                     YEAR ID
                      2823,000000
                                    2823,000000
                                                 2823,00000
                                                              2823,000000
count
        2823.000000
                         2.717676
                                       7.092455
                                                 2003.81509
                                                               100.715551
        3553.889072
mean
std
        1841.865106
                         1.203878
                                       3.656633
                                                     0.69967
                                                                40.187912
```

```
min
         482,130000
                         1.000000
                                       1.000000
                                                 2003.00000
                                                                33,000000
25%
        2203.430000
                         2.000000
                                       4.000000
                                                 2003.00000
                                                                68.000000
50%
        3184.800000
                                                                99.000000
                         3.000000
                                       8.000000
                                                 2004.00000
75%
        4508.000000
                         4.000000
                                      11.000000
                                                 2004.00000
                                                               124.000000
       14082.800000
                                      12.000000
                         4.000000
                                                 2005.00000
                                                               214.000000
max
df.shape
(2823, 25)
#GIVES NUMBER OF NULL VALUES IN EACH COLUMN
df.isnull().sum()
ORDERNUMBER
                        0
                        0
OUANTITYORDERED
                        0
PRICEEACH
                        0
ORDERLINENUMBER
SALES
                        0
                        0
ORDERDATE
STATUS
                        0
QTR ID
                        0
MONTH ID
                        0
YEAR ID
                        0
PRODUCTLINE
                        0
MSRP
                        0
                        0
PRODUCTCODE
                        0
CUSTOMERNAME
                        0
PHONE
ADDRESSLINE1
                        0
ADDRESSLINE2
                     2521
CITY
                        0
STATE
                     1486
POSTALCODE
                       76
COUNTRY
                        0
TERRITORY
                     1074
CONTACTLASTNAME
                        0
                        0
CONTACTFIRSTNAME
DEALSIZE
                        0
dtype: int64
#THIS WILLL REMOVE ROWS WHICH CONTAINS TOTAL NULL VALUES
df.dropna(how="all")
      ORDERNUMBER QUANTITYORDERED PRICEEACH ORDERLINENUMBER
SALES \
```

0 2871.0	1016	97	30	95.7	0		2	
1	1012	21	34	81.3	5		5	
2765.9 2	1013	34	41	94.7	4		2	
3884.3 3	34 1014	15	45	83.2	6		6	
3746.7	70							
4	1015	59	49	100.0	0		14	
5205.2	21				_			
		•		• • •	•			• •
2818	1035	50	20	100.0	0		15	
2244.4		70	20	100 0	0		1	
2819 3978.5	1037 51	7.5	29	100.0	U		1	
2820	1038	36	43	100.0	0		4	
5417.5 2821	5/ 1039	97	34	62.2	4		1	
2116.3	16							
2822	1041	L4	47	65.5	2		9	
3079.4	44							
1 2 2 3 4 4 2 4 4 2 4 4 4 4 4 4 4 4 4 4 4	ORDERDATE 2003-02-24 2003-07-05 2003-01-07 2003-08-25 2003-10-10 2004-02-12 2005-01-31 2005-01-03 2005-03-28 2005-06-05	STATUS Shipped Shipped Shipped Shipped Shipped Shipped Shipped Shipped Resolved Shipped On Hold	1 2 3 3 4 4 1 1	MONTH_ID	YEAR	03 03 03 03 04 05 05		
		ADD	RESSLINE	1 ADDRESS	LINE2		CITY	STATE
0	897	Long Airpo	rt Avenue	Э	null		NYC	NY
1		59 rue de	e l'Abbaye	е	null		Reims	null
2	27 rue du	Colonel Pi	erre Avia	Э	null		Paris	null
3		78934 Hil	lside Dr		null		Pasadena	CA
4		7734 S	trong St		null	San F	rancisco	CA
				•				

2818		C/ Mora	lzarzal, 86	null	Madrid	null
2819		٦	Γorikatu 38	null	0ulu	null
2820		C/ Mora	lzarzal, 86	null	Madrid	null
2821	1	rue Alsac	ce-Lorraine	null	Toulouse	null
2822		8616 Sp	innaker Dr.	null	Boston	MA
200						
POS DEALSIZE	TALCODE	COUNTRY	TERRITORY	CONTACTLASTNAME	CONTACTETRSTNA	ME
0	10022	USA	null	Yu	Kw	ai
Small					_	_
1	51100	FRANCE	EMEA	Henriot	Pa	ul
Small 2	75508	FRANCE	EMEA	Da Cunha	Dani	ല
Medium	75500	TTOUTCE	LITEA	Da Callila	Danii	
3	90003	USA	null	Young	Jul	ie
Medium					_	
4	null	USA	null	Brown	Jul	ie
Medium						
					•	• •
2818	28034	SPAIN	EMEA	Freyre	Die	ao
Small				,		3
2819	90110	FINLAND	EMEA	Koskitalo	Pirk	ko
Medium	20024	CDATN	E14E 4	_	<u> </u>	
2820 Medium	28034	SPAIN	EMEA	Freyre	Die	go
2821	31000	FRANCE	EMEA	Roulet	Annet	tο
Small	31000	TRANCE	LIILA	Noutet	Ailic	
2822	51003	USA	null	Yoshido	Ju	ri
Medium						

[2823 rows x 25 columns]

#REMOVE DUPLICATES
df.drop_duplicates()

	ORDERNUMBER	QUANTITYORDERED	PRICEEACH	ORDERLINENUMBER
SALES	\			
0	10107	30	95.70	2
2871.	00			
1	10121	34	81.35	5
2765.	90			
2	10134	41	94.74	2
3884.				
3	10145	45	83.26	6

4 10159 49 100.00 14 5205.27	2746 70								
		10159		49	100.0	00		14	
2244.40 2819									
2819 10373 29 100.00 1 3978.51 2820 10386 43 100.00 4 5417.57 2821 10397 34 62.24 1 2116.16 2822 10414 47 65.52 9 3079.44 ORDERDATE STATUS QTR_ID MONTH_ID YEAR_ID \ 0 2003.02.24 Shipped 1 2 2003 \ 1 2003.07.05 Shipped 2 5 5003 \ 2 2003.01.07 Shipped 3 7 2003 \ 3 2003.08.25 Shipped 3 8 2003 \ 4 2003.10.10 Shipped 4 10 2003 \ 2818 2004.02.21 Shipped 1 1 20005 \ 2821 2005.06.05 On Hold 2 5 2005 \ ADDRESSLINE1 ADDRESSLINE2 CITY STATE \ 0 897 Long Airport Avenue nul NYC NY 1 59 rue de l'Abbaye null Reims null 3 78934 Hillside Dr. null San Francisco CA		10350		20	100.0	00		15	
2820 10386 43 100.00 4 5417.57 2821 10397 34 62.24 1 2116.16 2822 10414 47 65.52 9 3079.44 ORDERDATE STATUS QTR_ID MONTH_ID YEAR_ID\ 0 2003-02-24 Shipped 1 2 2003 1 2003-07-05 Shipped 2 5 2003 2 2003-01-07 Shipped 3 7 2003 3 2003-08-25 Shipped 3 8 2003 4 2003-10-10 Shipped 4 10 2003 2818 2004-02-12 Shipped 4 12 2004 2818 2005-01-31 Shipped 1 1 2005 2829 2005-01-03 Resolved 1 3 2005 2820 2005-06-05 On Hold 2 5 2005 ADDRESSLINE1 ADDRESSLINE2 CITY STATE 0 897 Long Airport Avenue null NYC NY 1 59 rue de l'Abbaye null Reims null 2 27 rue du Colonel Pierre Avia null Paris null 3 78934 Hillside Dr. null San Francisco CA 2818 C/ Moralzarzal, 86 null Madrid null 2820 C/ Moralzarzal, 86 null Madrid null	2819	10373		29	100.0	00		1	
5417.57 2821 10397 34 62.24 1 2116.16 2822 10414 47 65.52 9 3079.44		10386		43	100 0	۱0		4	
2116.16 2822		10300		43	100.0	,,,		7	
2822 10414 47 65.52 9 3079.44 ORDERDATE STATUS QTR_ID MONTH_ID YEAR_ID \ 0 2003-02-24 Shipped 1 2 2003 \ 1 2003-07-05 Shipped 2 5 2003 \ 2 2003-01-07 Shipped 3 7 2003 \ 3 2003-08-25 Shipped 3 8 2003 \ 4 2003-10-10 Shipped 4 10 2003 \ \ 2818 2004-02-12 Shipped 1 1 2005 \ 2820 2005-01-31 Shipped 1 1 2005 \ 2821 2005-03-28 Shipped 1 3 2005 \ 2822 2005-03-28 Shipped 1 3 2005 \ 2822 2005-06-05 On Hold 2 5 2005 \ ADDRESSLINE1 ADDRESSLINE2 CITY STATE \ 0 897 Long Airport Avenue null NYC NY \ 1 59 rue de l'Abbaye null Reims null \ 2 27 rue du Colonel Pierre Avia null Paris null \ 3 78934 Hillside Dr. null San Francisco CA \		10397		34	62.2	24		1	
ORDERDATE STATUS QTR_ID MONTH_ID YEAR_ID \ 0		10414		47	65.5	52		9	
0 2003-02-24 Shipped 1 2 2003 1 2003-07-05 Shipped 2 5 2003 2 2003-01-07 Shipped 3 7 2003 3 2003-08-25 Shipped 3 8 2003 4 2003-10-10 Shipped 4 10 2003 2818 2004-02-12 Shipped 4 12 2004 2819 2005-01-31 Shipped 1 3 2005 2821 2005-03-28 Shipped 1 3 2005 2822 2005-06-05 On Hold 2 5 2005 ADDRESSLINE1 ADDRESSLINE2 CITY STATE 0 897 Long Airport Avenue null NYC NY 1 59 rue de l'Abbaye null Paris null 2 27 rue du Colonel Pierre Avia null Pasadena CA 4 7734		10 11 1		.,	0313	, _		J	
NYC NY 1 59 rue de l'Abbaye null Reims null 2 27 rue du Colonel Pierre Avia null Paris null 3 78934 Hillside Dr. null Pasadena CA 4 7734 Strong St. null San Francisco CA 2818 C/ Moralzarzal, 86 null Madrid null 2820 C/ Moralzarzal, 86 null Madrid null	0 2003-1 1 2003-1 2 2003-1 3 2003-1 4 2003-1 2818 2004-1 2819 2005-1 2820 2005-1 2821 2005-1	02-24 07-05 01-07 08-25 10-10 02-12 01-31 01-03 R	Shipped Shipped Shipped Shipped Shipped Shipped Shipped desolved Shipped	- 1 2 3 4 4 1 1	- 2 5 7 8 10 12 1 3 3	20 20 20 20 20 20 20 20	03 03 03 03 04 05 05		
0 897 Long Airport Avenue null NYC NY 1 59 rue de l'Abbaye null Reims null 2 27 rue du Colonel Pierre Avia null Paris null 3 78934 Hillside Dr. null Pasadena CA 4 7734 Strong St. null San Francisco CA 2818 C/ Moralzarzal, 86 null Madrid null 2820 C/ Moralzarzal, 86 null Madrid null			ADD	RESSLINE	1 ADDRESS	SLINE2		CITY	STATE
2 27 rue du Colonel Pierre Avia null Paris null 3 78934 Hillside Dr. null Pasadena CA 4 7734 Strong St. null San Francisco CA 2818 C/ Moralzarzal, 86 null Madrid null 2819 Torikatu 38 null Oulu null 2820 C/ Moralzarzal, 86 null Madrid null	0	897 Lo	ng Airpo	rt Avenu	е	null		NYC	NY
378934 Hillside Dr.nullPasadenaCA47734 Strong St.nullSan FranciscoCA2818C/ Moralzarzal, 86nullMadridnull2819Torikatu 38nullOulunull2820C/ Moralzarzal, 86nullMadridnull	1	5	9 rue de	l'Abbay	е	null		Reims	null
4 7734 Strong St. null San Francisco CA 2818 C/ Moralzarzal, 86 null Madrid null 2819 Torikatu 38 null Oulu null 2820 C/ Moralzarzal, 86 null Madrid null	2 27 r	ue du Co	lonel Pi	erre Avi	а	null		Paris	null
	3	7	8934 Hil	lside Dr		null	Pā	asadena	CA
2818 C/ Moralzarzal, 86 null Madrid null 2819 Torikatu 38 null Oulu null 2820 C/ Moralzarzal, 86 null Madrid null	4		7734 S	trong St		null	San Fra	ancisco	CA
Torikatu 38 null Oulu null 2820 C/ Moralzarzal, 86 null Madrid null									
2820 C/ Moralzarzal, 86 null Madrid null	2818	C	/ Moralz	arzal, 8	6	null		Madrid	null
	2819		То	rikatu 3	8	null		Oulu	null
2821 1 rue Alsace-Lorraine null Toulouse null	2820	С	/ Moralz	arzal, 8	6	null		Madrid	null

2822		8616 Spi	innaker Dr	. null	Boston	MA
2022		0010 001			203 (0.1.	, .
DOCT	AL CODE	COUNTDY	TEDDITODY	CONTACTI ACTNAME	CONTACTETECTNAME	
DEALSIZE	TALCODE	COUNTRY	TERRITORY	CUNTACTLASTNAME	CONTACTFIRSTNAME	
0 0	10022	USA	null	Yu	Kwai	
Small				. •		
1	51100	FRANCE	EMEA	Henriot	Paul	
Small						
2 Madium	75508	FRANCE	EMEA	Da Cunha	Daniel	
Medium 3	90003	USA	null	Young	Julie	
Medium	90003	UJA	nacc	roung	Juite	
4	null	USA	null	Brown	Julie	
Medium						
2010	20024	CDATN	ГМГΛ	Гиалия	Diana	
2818 Small	28034	SPAIN	EMEA	Freyre	Diego	
2819	90110	FINLAND	EMEA	Koskitalo	Pirkko	
Medium	30110	1 1112/1110	211271	NOSKILUCO	1 21 11110	
2820	28034	SPAIN	EMEA	Freyre	Diego	
Medium						
2821	31000	FRANCE	EMEA	Roulet	Annette	
Small 2822	51003	USA	null	Yoshido	Juri	
Medium	31003	UJA	nacc	10311140	Juil	
[2823 row	ıs x 25	columns]				
#HANDLING	MTCCTN	IC VALUES				
			DRESSLINE	2'].fillna('null'	')	
			fillna('nu		,	
				llna('null')		
df['POSTA	(LCODE ']	=df['P0S]	TALCODE'].	fillna('null')		
df.isnull	(). <mark>sum</mark> ()				
ORDERNUMB	DED	0				
QUANTITYO		0 0				
PRICEEACH		0				
ORDERLINE		0				
SALES		0				
ORDERDATE		0				
STATUS		0				
QTR_ID MONTH ID		0 0				
YEAR ID		0				
PRODUCTLI	NE	0				

```
MSRP
                    0
PRODUCTCODE
                    0
CUSTOMERNAME
                    0
PHONE
                    0
ADDRESSLINE1
                    0
ADDRESSLINE2
                    0
                    0
CITY
STATE
                    0
POSTALCODE
                    0
COUNTRY
                    0
TERRITORY
                    0
CONTACTLASTNAME
                    0
CONTACTFIRSTNAME
                    0
DEALSIZE
                    0
dtype: int64
#FIXING INCONSISTENCIES
df['COUNTRY'] = df['COUNTRY'].str.upper()
#CONVERTING DATE COLUMNS TO PROPER FORMAT
df['ORDERDATE'] = pd.to datetime(df['ORDERDATE'], format='mixed',
errors='coerce', dayfirst=True)
print(df['ORDERDATE'])
       2003-02-24
0
1
       2003-07-05
2
       2003-01-07
3
       2003-08-25
4
       2003 - 10 - 10
2818
       2004-02-12
2819
       2005-01-31
2820
       2005-01-03
2821
       2005-03-28
2822
       2005-06-05
Name: ORDERDATE, Length: 2823, dtype: datetime64[ns]
#CLEANED DATA SET
df.to csv("cleaned sales data.csv", index=False)
df
      ORDERNUMBER QUANTITYORDERED
                                     PRICEEACH ORDERLINENUMBER
SALES \
                                         95.70
                                                               2
            10107
                                 30
2871.00
                                 34
                                         81.35
                                                               5
            10121
2765.90
                                 41
                                         94.74
                                                               2
            10134
3884.34
```

3 3746.7	1014	15	45	83.2	26		6	
4	1015	59	49	100.0	00		14	
5205.2								
					•			
2818	1035	50	20	100.0	00		15	
2244.4	10							
2819	1037	73	29	100.0	00		1	
3978.5 2820	1038	36	43	100.0	00		4	
5417.5		,,	15	10010	, 0		•	
2821	1039	97	34	62.2	24		1	
2116.1 2822	1041	1.4	47	65.5	52		9	
3079.4		LT	7/	03.3	, _		5	
		CTATUC	OTD TD	MONTH TO	\/EAD	T.D.	,	
0 2	ORDERDATE 2003-02-24	STATUS Shipped	QTR_ID 1	MONTH_ID 2	YEAR_	03	. \	
1 2	2003-02-24	Shipped		5		03		
2 2	2003-01-07	Shipped	2 3 3	7	20	03		
	2003-08-25	Shipped		8		03		
	2003 - 10 - 10	Shipped	4	10		03	•	
2818 2	2004-02-12	Shipped	4	12		04	•	
	2005-01-31	Shipped	i	1		05		
	2005-01-03	Resolved	1	3	20	05		
	2005-03-28	Shipped	1	3	20			
2822 2	2005 - 06 - 05	On Hold	2	5	20	05		
		ADD	RESSLINE:	1 ADDRESS	SLINE2		CITY	STATE
\	007	1 a a a . A d		_	11		NVC	NIV
0	897	Long Airpo	rt Avenue	2	null		NYC	NY
1		59 rue de	l'Abbaye	9	null		Reims	null
2	27 rue du	Colonel Pi	erre Avia	3	null		Paris	null
3		78934 Hil	lside Dr		null		Pasadena	CA
4		7734 S	trong St		null	San F	rancisco	CA
2818		C/ Moralz	arzal, 80	5	null		Madrid	null
2819		То	rikatu 38	3	null		0ulu	null
2820		C/ Moralz	arzal 04	ŝ	null		Madrid	null
2020		c/ Huratz	ui Zat, 00	,	Hutt		Hauliu	Hutt

2821	1	rue Alsac	ce-Lorraine	e nul	l	Toulouse	null
2822		8616 Spi	innaker Dr	. nul	l	Boston	MA
POST	ALCODE	COUNTRY	TERRITORY	CONTACTLASTNA	ME CONT	ACTFIRSTNA	ME
DEALSIZE							
0	10022	USA	null		Yu	Kw	ai
Small							
1	51100	FRANCE	EMEA	Henri	ot	Pa	ul
Small							
2	75508	FRANCE	EMEA	Da Cun	ha	Dani	el
Medium							
3	90003	USA	null	You	ng	Jul	ie
Medium					3		
4	null	USA	null	Bro	wn	Jul	ie
Medium							
2818	28034	SPAIN	EMEA	Frey	re	Die	qo
Small				,			
2819	90110	FINLAND	EMEA	Koskita	lo	Pirk	ko
Medium							
2820	28034	SPAIN	EMEA	Frey	re	Die	go
Medium							
2821	31000	FRANCE	EMEA	Roul	et	Annet	te
Small							
2822	51003	USA	null	Yoshi	do	Ju	ri
Medium							
[2823 row	ıs x 25	columns]					