```
import pandas as pd
df=pd.read_csv("C:/Users/Sruth/Downloads/sales_data_sample.csv")

 print(df.head())

   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  3884.34

3        10145               45      83.26                6  3746.70

4        10159               49     100.00               14  5205.27


          ORDERDATE   STATUS  QTR_ID  MONTH_ID  YEAR_ID  ...  \
0    2/24/2003 0:00  Shipped       1         2     2003  ...
1  05-07-2003 00:00  Shipped       2         5     2003  ...
2  07-01-2003 00:00  Shipped       3         7     2003  ...
3    8/25/2003 0:00  Shipped       3         8     2003  ...
4  10-10-2003 00:00  Shipped       4        10     2003  ...

                    ADDRESSLINE1  ADDRESSLINE2           CITY STATE  \
0        897 Long Airport Avenue          NaN            NYC    NY
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
0      10022     USA       NaN              Yu             Kwai
Small
1      51100  France      EMEA         Henriot             Paul
Small
2      75508  France      EMEA        Da Cunha           Daniel
Medium
3      90003     USA       NaN           Young            Julie
Medium
4        NaN     USA       NaN           Brown            Julie
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
---  ------            --------------  -----
 0   ORDERNUMBER       2823 non-null   int64
 1   QUANTITYORDERED   2823 non-null   int64
 2   PRICEEACH         2823 non-null   float64
 3   ORDERLINENUMBER   2823 non-null   int64
 4   SALES             2823 non-null   float64
 5   ORDERDATE         2823 non-null   object
 6   STATUS            2823 non-null   object
 7   QTR_ID            2823 non-null   int64
 8   MONTH_ID          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
 16  ADDRESSLINE2      302 non-null    object
 17  CITY              2823 non-null   object
 18  STATE             1337 non-null   object
 19  POSTALCODE        2747 non-null   object
 20  COUNTRY           2823 non-null   object
 21  TERRITORY         1749 non-null   object
 22  CONTACTLASTNAME   2823 non-null   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  \
count   2823.000000      2823.000000  2823.000000      2823.000000
mean   10258.725115        35.092809    83.658544         6.466171
std       92.085478         9.741443    20.174277         4.225841
min    10100.000000         6.000000    26.880000         1.000000
25%    10180.000000        27.000000    68.860000         3.000000
50%    10262.000000        35.000000    95.700000         6.000000
75%    10333.500000        43.000000   100.000000         9.000000
max    10425.000000        97.000000   100.000000        18.000000

              SALES       QTR_ID     MONTH_ID     YEAR_ID         MSRP

count   2823.000000  2823.000000  2823.000000  2823.00000  2823.000000

mean    3553.889072     2.717676     7.092455  2003.81509   100.715551

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     3.000000     8.000000  2004.00000    99.000000
75%     4508.000000     4.000000    11.000000  2004.00000   124.000000
max    14082.800000     4.000000    12.000000  2005.00000   214.000000
```

```
df.shape
```

```
(2823, 25)
```

```
#GIVES NUMBER OF NULL VALUES IN EACH COLUMN
df.isnull().sum()
```

```
ORDERNUMBER            0
QUANTITYORDERED        0
PRICEEACH              0
ORDERLINENUMBER        0
SALES                  0
ORDERDATE              0
STATUS                 0
QTR_ID                 0
MONTH_ID               0
YEAR_ID                0
PRODUCTLINE            0
MSRP                   0
PRODUCTCODE            0
CUSTOMERNAME           0
PHONE                  0
ADDRESSLINE1           0
ADDRESSLINE2        2521
CITY                   0
STATE               1486
POSTALCODE            76
COUNTRY                0
TERRITORY           1074
CONTACTLASTNAME        0
CONTACTFIRSTNAME       0
DEALSIZE               0
dtype: int64
```

```
#THIS WILLL REMOVE ROWS WHICH CONTAINS TOTAL NULL VALUES
df.dropna(how="all")
```

```
      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
3884.34
3           10145               45      83.26                 6
3746.70
4           10159               49     100.00                14
5205.27
...           ...              ...        ...               ...     ..
.
2818        10350               20     100.00                15
2244.40
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     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         1     2005  ...
2820 2005-01-03  Resolved       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          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

2819                     Torikatu 38          null            Oulu  null

2820              C/ Moralzarzal, 86          null          Madrid  null

2821           1 rue Alsace-Lorraine          null        Toulouse  null

2822              8616 Spinnaker Dr.          null          Boston    MA


     POSTALCODE  COUNTRY TERRITORY CONTACTLASTNAME CONTACTFIRSTNAME
DEALSIZE
0         10022      USA      null              Yu             Kwai
Small
1         51100   FRANCE      EMEA         Henriot             Paul
Small
2         75508   FRANCE      EMEA        Da Cunha           Daniel
Medium
3         90003      USA      null           Young            Julie
Medium
4          null      USA      null           Brown            Julie
Medium
...         ...      ...       ...             ...              ...
...
2818      28034    SPAIN      EMEA          Freyre            Diego
Small
2819      90110  FINLAND      EMEA       Koskitalo           Pirkko
Medium
2820      28034    SPAIN      EMEA          Freyre            Diego
Medium
2821      31000   FRANCE      EMEA          Roulet          Annette
Small
2822      51003      USA      null         Yoshido             Juri
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
3884.34
3           10145               45      83.26                6
```

```
3746.70
4           10159               49      100.00               14
5205.27
...           ...              ...        ...               ...       ..
.
2818        10350               20      100.00               15
2244.40
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     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         1     2005  ...
2820 2005-01-03  Resolved       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          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

2819                    Torikatu 38          null           Oulu  null

2820             C/ Moralzarzal, 86          null         Madrid  null

2821          1 rue Alsace-Lorraine          null       Toulouse  null
```

```
2822             8616 Spinnaker Dr.          null         Boston    MA

     POSTALCODE  COUNTRY TERRITORY CONTACTLASTNAME CONTACTFIRSTNAME
DEALSIZE
0         10022      USA      null              Yu             Kwai
Small
1         51100   FRANCE      EMEA         Henriot             Paul
Small
2         75508   FRANCE      EMEA        Da Cunha           Daniel
Medium
3         90003      USA      null           Young            Julie
Medium
4          null      USA      null           Brown            Julie
Medium
...         ...      ...       ...             ...              ...
...
2818      28034    SPAIN      EMEA          Freyre            Diego
Small
2819      90110  FINLAND      EMEA       Koskitalo           Pirkko
Medium
2820      28034    SPAIN      EMEA          Freyre            Diego
Medium
2821      31000   FRANCE      EMEA          Roulet          Annette
Small
2822      51003      USA      null         Yoshido             Juri
Medium

[2823 rows x 25 columns]
```

```
#HANDLING MISSING VALUES
df['ADDRESSLINE2']=df['ADDRESSLINE2'].fillna('null')
df['STATE']=df['STATE'].fillna('null')
df['TERRITORY']=df['TERRITORY'].fillna('null')
df['POSTALCODE']=df['POSTALCODE'].fillna('null')

df.isnull().sum()
```

```
ORDERNUMBER         0
QUANTITYORDERED     0
PRICEEACH           0
ORDERLINENUMBER     0
SALES               0
ORDERDATE           0
STATUS              0
QTR_ID              0
MONTH_ID            0
YEAR_ID             0
PRODUCTLINE         0
```

```
MSRP                0
PRODUCTCODE         0
CUSTOMERNAME        0
PHONE               0
ADDRESSLINE1        0
ADDRESSLINE2        0
CITY                0
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'])
```

```
0      2003-02-24
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  \
0           10107               30      95.70                2
2871.00
1           10121               34      81.35                5
2765.90
2           10134               41      94.74                2
3884.34
```

```
3           10145               45      83.26                 6
3746.70
4           10159               49     100.00                14
5205.27
...           ...              ...        ...               ...       ..
.
2818        10350               20     100.00                15
2244.40
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     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         1     2005  ...
2820 2005-01-03  Resolved       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          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

2819                    Torikatu 38          null           Oulu  null

2820             C/ Moralzarzal, 86          null         Madrid  null
```

```
2821          1 rue Alsace-Lorraine          null       Toulouse  null

2822             8616 Spinnaker Dr.          null         Boston    MA


     POSTALCODE  COUNTRY TERRITORY CONTACTLASTNAME CONTACTFIRSTNAME 
DEALSIZE  
0         10022      USA      null              Yu             Kwai    
Small  
1         51100   FRANCE      EMEA         Henriot             Paul    
Small  
2         75508   FRANCE      EMEA        Da Cunha           Daniel   
Medium  
3         90003      USA      null           Young            Julie   
Medium  
4          null      USA      null           Brown            Julie   
Medium  
...         ...      ...       ...             ...              ...      
...  
2818      28034    SPAIN      EMEA          Freyre            Diego    
Small  
2819      90110  FINLAND      EMEA       Koskitalo           Pirkko   
Medium  
2820      28034    SPAIN      EMEA          Freyre            Diego   
Medium  
2821      31000   FRANCE      EMEA          Roulet          Annette    
Small  
2822      51003      USA      null         Yoshido             Juri   
Medium  

[2823 rows x 25 columns]
```