

```
In [1]: import pandas as pd
import numpy as np

In [18]: df=pd.read_csv("D://datascience//data.csv",encoding= 'unicode_escape')

In [19]: df
```

Out[19]:

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
0	536365	85123A	WHITE HANGING HEART T- LIGHT HOLDER	6	12/1/2010 8:26	2.55	17850.0	United Kingdom
1	536365	71053	WHITE METAL LANTERN	6	12/1/2010 8:26	3.39	17850.0	United Kingdom
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	12/1/2010 8:26	2.75	17850.0	United Kingdom
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	12/1/2010 8:26	3.39	17850.0	United Kingdom
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	12/1/2010 8:26	3.39	17850.0	United Kingdom
...
541904	581587	22613	PACK OF 20 SPACEBOY NAPKINS	12	12/9/2011 12:50	0.85	12680.0	France
541905	581587	22899	CHILDREN'S APRON DOLLY GIRL	6	12/9/2011 12:50	2.10	12680.0	France
541906	581587	23254	CHILDRENS CUTLERY DOLLY GIRL	4	12/9/2011 12:50	4.15	12680.0	France
541907	581587	23255	CHILDRENS CUTLERY CIRCUS PARADE	4	12/9/2011 12:50	4.15	12680.0	France
541908	581587	22138	BAKING SET 9 PIECE RETROSPOT	3	12/9/2011 12:50	4.95	12680.0	France
541909 rows × 8 columns								

In [20]:

df.isnull()

Out[20]:

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
0	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False
...
541904	False	False	False	False	False	False	False	False
541905	False	False	False	False	False	False	False	False
541906	False	False	False	False	False	False	False	False
541907	False	False	False	False	False	False	False	False
541908	False	False	False	False	False	False	False	False

541909 rows × 8 columns



In [21]:

df.notnull()

Out[21]:

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
0	True	True	True	True	True	True	True	True
1	True	True	True	True	True	True	True	True
2	True	True	True	True	True	True	True	True
3	True	True	True	True	True	True	True	True
4	True	True	True	True	True	True	True	True
...
541904	True	True	True	True	True	True	True	True
541905	True	True	True	True	True	True	True	True
541906	True	True	True	True	True	True	True	True
541907	True	True	True	True	True	True	True	True
541908	True	True	True	True	True	True	True	True

541909 rows × 8 columns



In [22]:

df.dropna(axis=0)

Out[22]:

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
0	536365	85123A	WHITE HANGING HEART T-	6	12/1/2010 8:26	2.55	17850.0	United Kingdom

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
			LIGHT HOLDER					
1	536365	71053	WHITE METAL LANTERN	6	12/1/2010 8:26	3.39	17850.0	United Kingdom
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	12/1/2010 8:26	2.75	17850.0	United Kingdom
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	12/1/2010 8:26	3.39	17850.0	United Kingdom
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	12/1/2010 8:26	3.39	17850.0	United Kingdom
...
541904	581587	22613	PACK OF 20 SPACEBOY NAPKINS	12	12/9/2011 12:50	0.85	12680.0	France
541905	581587	22899	CHILDREN'S APRON DOLLY GIRL	6	12/9/2011 12:50	2.10	12680.0	France
541906	581587	23254	CHILDRENS CUTLERY DOLLY GIRL	4	12/9/2011 12:50	4.15	12680.0	France
541907	581587	23255	CHILDRENS CUTLERY CIRCUS PARADE	4	12/9/2011 12:50	4.15	12680.0	France
541908	581587	22138	BAKING SET 9 PIECE RETROSPOT	3	12/9/2011 12:50	4.95	12680.0	France

406829 rows × 8 columns

In [23]:

df.replace()

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
0	536365	85123A	WHITE HANGING HEART T- LIGHT HOLDER	6	12/1/2010 8:26	2.55	17850.0	United Kingdom
1	536365	71053	WHITE METAL LANTERN	6	12/1/2010 8:26	3.39	17850.0	United Kingdom

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	12/1/2010 8:26	2.75	17850.0	United Kingdom
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	12/1/2010 8:26	3.39	17850.0	United Kingdom
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	12/1/2010 8:26	3.39	17850.0	United Kingdom
...
541904	581587	22613	PACK OF 20 SPACEBOY NAPKINS	12	12/9/2011 12:50	0.85	12680.0	France
541905	581587	22899	CHILDREN'S APRON DOLLY GIRL	6	12/9/2011 12:50	2.10	12680.0	France
541906	581587	23254	CHILDRENS CUTLERY DOLLY GIRL	4	12/9/2011 12:50	4.15	12680.0	France
541907	581587	23255	CHILDRENS CUTLERY CIRCUS PARADE	4	12/9/2011 12:50	4.15	12680.0	France
541908	581587	22138	BAKING SET 9 PIECE RETROSPOT	3	12/9/2011 12:50	4.95	12680.0	France

541909 rows × 8 columns

```
In [24]: df.head()
```

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
0	536365	85123A	WHITE HANGING HEART T- LIGHT HOLDER	6	12/1/2010 8:26	2.55	17850.0	United Kingdom
1	536365	71053	WHITE METAL LANTERN	6	12/1/2010 8:26	3.39	17850.0	United Kingdom
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	12/1/2010 8:26	2.75	17850.0	United Kingdom

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	12/1/2010 8:26	3.39	17850.0	United Kingdom
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	12/1/2010 8:26	3.39	17850.0	United Kingdom

In [25]: `df.tail()`

Out[25]:

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
541904	581587	22613	PACK OF 20 SPACEBOY NAPKINS	12	12/9/2011 12:50	0.85	12680.0	France
541905	581587	22899	CHILDREN'S APRON DOLLY GIRL	6	12/9/2011 12:50	2.10	12680.0	France
541906	581587	23254	CHILDRENS CUTLERY DOLLY GIRL	4	12/9/2011 12:50	4.15	12680.0	France
541907	581587	23255	CHILDRENS CUTLERY CIRCUS PARADE	4	12/9/2011 12:50	4.15	12680.0	France
541908	581587	22138	BAKING SET 9 PIECE RETROSPOT	3	12/9/2011 12:50	4.95	12680.0	France



In [26]: `df.shape`

Out[26]: (541909, 8)

In [27]: `df.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 541909 entries, 0 to 541908
Data columns (total 8 columns):
#   Column          Non-Null Count  Dtype
---  ---
0   InvoiceNo        541909 non-null object
1   StockCode        541909 non-null object
2   Description      540455 non-null object
3   Quantity         541909 non-null int64
4   InvoiceDate      541909 non-null object
5   UnitPrice        541909 non-null float64
6   CustomerID       406829 non-null float64
7   Country          541909 non-null object
dtypes: float64(2), int64(1), object(5)
memory usage: 33.1+ MB
```

In [29]: `df.values`

Out[29]: array([['536365', '85123A', 'WHITE HANGING HEART T-LIGHT HOLDER', ...,
2.55, 17850.0, 'United Kingdom'],
['536365', '71053', 'WHITE METAL LANTERN', ..., 3.39, 17850.0,

```
'United Kingdom'],
['536365', '84406B', 'CREAM CUPID HEARTS COAT HANGER', ..., 2.75,
17850.0, 'United Kingdom'],
...,
['581587', '23254', 'CHILDRENS CUTLERY DOLLY GIRL ', ..., 4.15,
12680.0, 'France'],
['581587', '23255', 'CHILDRENS CUTLERY CIRCUS PARADE', ..., 4.15,
12680.0, 'France'],
['581587', '22138', 'BAKING SET 9 PIECE RETROSPOT ', ..., 4.95,
12680.0, 'France']], dtype=object)
```

In [30]: `df.value_counts()`

```
Out[30]: InvoiceNo  StockCode  Description  Quantity  InvoiceDate
UnitPrice  CustomerID  Country
555524      22698      PINK REGENCY TEACUP AND SAUCER      1      6/5/2011 11:37
2.95      16923.0      United Kingdom      20
22697      GREEN REGENCY TEACUP AND SAUCER      1      6/5/2011 11:37
2.95      16923.0      United Kingdom      12
572861      22775      PURPLE DRAWERKNOB ACRYLIC EDWARDIAN      12      10/26/2011 12:4
6 1.25      14102.0      United Kingdom      8
541266      21755      LOVE BUILDING BLOCK WORD      1      1/16/2011 16:25
5.95      15673.0      United Kingdom      6
540524      21756      BATH BUILDING BLOCK WORD      1      1/9/2011 12:53
5.95      16735.0      United Kingdom      6

..
569360      22725      ALARM CLOCK BAKELIKE CHOCOLATE      1      10/3/2011 14:59
3.75      14653.0      United Kingdom      1
22726      ALARM CLOCK BAKELIKE GREEN      1      10/3/2011 14:59
3.75      14653.0      United Kingdom      1
22727      ALARM CLOCK BAKELIKE RED      1      10/3/2011 14:59
3.75      14653.0      United Kingdom      1
22728      ALARM CLOCK BAKELIKE PINK      1      10/3/2011 14:59
3.75      14653.0      United Kingdom      1
536365      21730      GLASS STAR FROSTED T-LIGHT HOLDER      6      12/1/2010 8:26
4.25      17850.0      United Kingdom      1
Length: 401604, dtype: int64
```

In [31]: `df['Quantity'].unique()`

```
Out[31]: array([ 6, 8, 2, 32, 3, 4, 24, 12,
48, 18, 20, 36, 80, 64, 10, 120,
96, 23, 5, 1, -1, 50, 40, 100,
192, 432, 144, 288, -12, -24, 16, 9,
128, 25, 30, 28, 7, 56, 72, 200,
600, 480, -6, 14, -2, 11, 33, 13,
-4, -5, -7, -3, 70, 252, 60, 216,
384, -10, 27, 15, 22, 19, 17, 21,
34, 47, 108, 52, -9360, -38, 75, 270,
42, 240, 90, 320, 1824, 204, 69, -36,
-192, -144, 160, 2880, 1400, 39, -48, -50,
26, 1440, 31, 82, 78, 97, 98, 35,
57, -20, 110, -22, -30, -70, -130, -80,
-120, -40, -25, -14, -15, -69, -140, -320,
-8, 720, 156, 324, 38, 37, 49, 95,
-9, -11, 29, 41, -72, -35, -21, -43,
-19, -18, -44, 402, 378, 150, 300, 54,
104, 67, 258, 66, 44, 55, 46, 99,
61, 408, 972, 208, 1008, 1000, -77, 1488,
250, 1394, 400, -223, -150, -13, -33, -723,
-177, 79, 84, -32, -100, -28, 272, -145,
-47, -96, 113, 45, 106, 68, 267, 115,
65, 1728, -60, -16, 53, -240, 76, 460,
71, 43, 213, 58, 576, 2400, 500, 180,
-300, -500, -23, 752, 960, 1296, 210, 172,
215, 129, 138, 116, 135, 197, -106, -54,
-17, -939, 147, 168, 256, -201, -53, -29,
-2600, -990, -290, -45, 860, 1010, 1356, 1284,
```

```
186, 114, 360, 1930, 2000, 3114, 1300, 670,
111, 211, 59, -310, -61, -41, 176, 648,
62, 74215, -74215, -64, -84, 89, -1400, 73,
-57, 112, 456, -59, -31, 5568, 2560, 136,
900, -600, -42, -94, -207, -52, 130, -206,
2592, 420, 800, 101, 1200, 864, -217, 94,
-1430, 1287, -162, -230, -173, -390, -234, 504,
123, 118, -76, -200, 1056, 1500, 280, 407,
141, 124, -99, 51, -92, -741, 3906, -400,
-114, 102, 1152, -88, 198, 117, 86, -720,
125, -86, -391, -87, -278, 140, 228, -154,
-3000, 81, -675, -210, -345, -975, -1200, -1121,
-27, -541, -1277, -197, 3000, -1100, -63, -5368,
219, -259, 232, -82, -178, 170, -110, 768,
88, 700, -102, 2160, -323, -450, -232, -83,
-153, -524, -2472, -49, -105, 342, 304, 167,
-34, 640, 175, 220, 74, 93, 164, 163,
63, 637, 122, 158, 165, 350, -46, -75,
-690, -39, -66, 83, 312, -180, 392, -624,
-194, -26, -62, -37, -91, -139, -158, -1479,
-55, -576, -750, 330, 151, -93, -432, -58,
-1092, -670, -1300, -1930, -2000, -3114, 462, -65,
-68, 492, -620, 273, -90, -170, -1512, -51,
85, -56, -160, -360, 105, -960, -2376, 1350,
428, -1350, 336, -786, -590, -168, -101, -71,
132, -413, -664, 227, 2700, 222, 246, 906,
756, 888, 552, -97, 224, 738, 608, -212,
4300, 146, 143, 276, -125, -116, -108, 4000,
-304, -272, -1206, -95, 1600, 323, -161, -472,
-618, -204, 1515, -1515, -9058, -9600, 660, -420,
-126, -220, -271, -1440, 264, 188, 588, 612,
152, -324, -480, -164, -78, -118, 430, -1681,
87, 155, 701, 828, 540, 696, 560, -250,
-408, -179, 121, -124, 512, -251, -3100, 3100,
-169, 126, 291, 3186, -2834, -109, 109, -121,
-530, -227, 261, -346, 352, 142, 107, -188,
-1060, -342, -288, 348, 1900, 157, -343, -455,
425, 968, 684, 824, -828, -701, 196, 248,
410, 236, 230, -156, -553, 145, 448, 245,
-252, -334, -318, -113, -115, 171, -242, 840,
-967, -203, -3167, -443, -1897, 225, -434, 750,
-682, -484, 682, 344, -635, -117, -3667, 450,
310, 494, -384, 92, 1788, -138, 624, 744,
416, 496, 396, 306, 1878, 1944, 666, 708,
1428, 852, 1412, 528, -756, -752, -152, -85,
-312, -79, -147, -67, -131, 183, -209, -186,
-231, -129, -458, -275, -2880, 1540, 672, -800,
-430, -380, -74, -840, -1296, -365, -104, -270,
-73, -306, 91, 255, 468, -468, -111, -184,
-103, -335, 4800, -112, -1000, 912, 1992, 184,
148, -657, -1671, -1158, -2618, -2003, -674, -4830,
-905, -1128, 832, 992, 630, 1020, 2100, 162,
-1560, -1284, -81, 314, 370, 131, 133, 484,
149, 153, 257, 139, 137, 628, 179, 1820,
478, 335, 253, 242, 375, -276, -256, 281,
193, 181, 404, 244, 207, 199, 1130, 326,
654, 688, 268, 249, -864, 234, -1510, -550,
-244, -132, -327, -313, 279, -398, -280, 374,
-504, -696, -149, -224, -428, 212, -267, -175,
12540, 760, -98, 774, -151, 2040, -1131, -135,
1120, 77, -900, 177, 490, 388, 620, 1404,
-155, -355, -337, -1050, -338, 205, -235, 698,
80995, -80995], dtype=int64)
```

In [32]:

df.describe()

Out[32]:

	Quantity	UnitPrice	CustomerID
count	541909.000000	541909.000000	406829.000000

	Quantity	UnitPrice	CustomerID
mean	9.552250	4.611114	15287.690570
std	218.081158	96.759853	1713.600303
min	-80995.000000	-11062.060000	12346.000000
25%	1.000000	1.250000	13953.000000
50%	3.000000	2.080000	15152.000000
75%	10.000000	4.130000	16791.000000
max	80995.000000	38970.000000	18287.000000

In []: df.median()

In []: df.mean()

In []: df.count()

In []:

In []: