

In [1]:

```
## Perform the following operations using 'python' Language on the Amazon
## book review Data set.
## 1) Create data subsets(Row wise)
## 2) Merge Data (Row wise)
## 3) Sort Data
## 4) Transposing Data
## 5) Reshape Data
```

In [3]:

```
import pandas as pd
import numpy as np
df = pd.read_csv(r"C:\Users\Vinayak\Downloads\dsbda\datasets\Amazonbookreviews.csv", sep=
df.head()
```

Out[3]:

	reviewerID	asin	reviewername	helpful	reviewerText	overallRating	Summary
0	A2SUAM1J3GNN3B	23456.0	SNEHA	2	I bought this for my husband who plays the pia...	4.0	SUBWAY SURF
1	A2SUAN1J3GNN0C	879877.0	KARTHII	3	good book to read	3.0	PILOT
2	A2SUAN1J3GNN7B	23456.0	SHASHANK	3	high cost	2.5	DRIVE IN HEAVEN
3	A5SUAN1J3GNN9B	67578.0	NITIN	3	good useof book	2.2	SPORTS HEAVEN
4	A1SUAM1J3GNN3B	123.0	PRANOTI	3	I m happy with the book	3.0	HEAVEN OF PEARLS

In [ ]:

```
#Row Wise
```

In [4]:

```
#creating the subset1  
df1 = df[['asin', 'helpful', 'overallRating']].loc[0:15]  
df1
```

Out[4]:

	asin	helpful	overallRating
0	23456.0	2	4.0
1	879877.0	3	3.0
2	23456.0	3	2.5
3	67578.0	3	2.2
4	123.0	3	3.0
5	236789.0	3	3.0
6	NaN	3	3.0
7	224442.0	3	3.0
8	98767.0	2	3.0
9	78761.0	2	5.0
10	321456.0	2	4.0
11	876231.0	2	4.0
12	987.0	2	4.0
13	234.0	2	3.0
14	124.0	2	3.0
15	156.0	2	3.0

In [5]:

```
#creating the subset2
```

```
df2 = df[['asin', 'helpful', 'overallRating']].loc[16:30]  
df2
```

Out[5]:

	asin	helpful	overallRating
16	876.0	2	3.0
17	123.0	2	2.0
18	987.0	2	2.0
19	22456.0	2	2.0
20	9876.0	2	2.0
21	3451.0	2	4.0
22	879.0	2	4.0
23	321.0	2	4.0
24	2.0	3	4.0
25	1.0	3	4.0
26	23.0	3	4.0
27	11.0	3	4.0
28	11.0	3	4.0
29	23.0	3	4.0
30	21221.0	3	4.0

In [8]:

```
#merge  
merging = pd.concat([df1,df2])  
merging
```

Out[8]:

	asin	helpful	overallRating
0	23456.0	2	4.0
1	879877.0	3	3.0
2	23456.0	3	2.5
3	67578.0	3	2.2
4	123.0	3	3.0
5	236789.0	3	3.0
6	NaN	3	3.0
7	224442.0	3	3.0
8	98767.0	2	3.0
9	78761.0	2	5.0
10	321456.0	2	4.0
11	876231.0	2	4.0
12	987.0	2	4.0
13	234.0	2	3.0
14	124.0	2	3.0
15	156.0	2	3.0
16	876.0	2	3.0
17	123.0	2	2.0
18	987.0	2	2.0
19	22456.0	2	2.0
20	9876.0	2	2.0
21	3451.0	2	4.0
22	879.0	2	4.0
23	321.0	2	4.0
24	2.0	3	4.0
25	1.0	3	4.0
26	23.0	3	4.0
27	11.0	3	4.0
28	11.0	3	4.0
29	23.0	3	4.0
30	21221.0	3	4.0

In [ ]:

```
#row wise
```

In [15]:

```
#creating the subset3  
df3 = df[['asin','helpful','overallRating']]  
df3
```

Out[15]:

	asin	helpful	overallRating
0	23456.0	2	4.0
1	879877.0	3	3.0
2	23456.0	3	2.5
3	67578.0	3	2.2
4	123.0	3	3.0
5	236789.0	3	3.0
6	NaN	3	3.0
7	224442.0	3	3.0
8	98767.0	2	3.0
9	78761.0	2	5.0
10	321456.0	2	4.0
11	876231.0	2	4.0
12	987.0	2	4.0
13	234.0	2	3.0
14	124.0	2	3.0
15	156.0	2	3.0
16	876.0	2	3.0
17	123.0	2	2.0
18	987.0	2	2.0
19	22456.0	2	2.0
20	9876.0	2	2.0
21	3451.0	2	4.0
22	879.0	2	4.0
23	321.0	2	4.0
24	2.0	3	4.0
25	1.0	3	4.0
26	23.0	3	4.0
27	11.0	3	4.0
28	11.0	3	4.0
29	23.0	3	4.0
30	21221.0	3	4.0
31	21.0	3	4.0
32	12212.0	3	3.0
33	2121.0	3	3.0
34	21212.0	3	3.0
35	21221.0	3	3.0
36	22245.0	3	2.0

	asin	helpful	overallRating
37	678.0	3	2.0
38	111.0	3	2.0
39	223.0	3	2.0
40	2312.0	3	4.0
41	1222.0	2	4.0
42	11111.0	2	4.0
43	9800.0	2	4.0
44	NaN	2	4.0



In [16]:

```
#creating the subset4  
df4 = df[['Summary', 'reviewerText', 'reviewerID']]  
df4
```

Out[16]:

	Summary	reviewerText	reviewerID
0	SUBWAY SURF	I bought this for my husband who plays the pia...	A2SUAM1J3GNN3B
1	PILOT	good book to read	A2SUAN1J3GNN0C
2	DRIVE IN HEAVEN	high cost	A2SUAN1J3GNN7B
3	SPORTS HEAVEN	good useof book	A5SUAN1J3GNN9B
4	HEAVEN OF PEARLS	I m happy with the book	A1SUAM1J3GNN3B
5	GODS GIFT	im not happy with the delay of product	BA2345WERSDFRER
6	GODESS LIFE	delay in book	CADERVG3450EREW
7	LIFE OF PIE	Wonderful book	ASDF345671QWERT
8	INDIA A MASSIVE POWER	The music is at times hard to read because we ...	MNHJ879WERFGH
9	DATA IS DATA	many errors	1234ZSDFREQWEG
10	INDIA A MASSIVE POWER	no mini	9345GHIKLFLLD
11	FRIENDS HELP	high vat	NaN
12	MOTIVATIONAL	the dealer is charged more money	HGKIUYE345SFDG
13	DARK KNIGHT	helpful	NIKRT2398701222
14	HELL OF HELL	staff is good	MAHI2348900122
15	DEAA	good buy for money	LIHJKD87634O21
16	DEAR TO ALL	money is dis	SISISWER4657893
17	MOTHER	stare the raeder	MIKI23412341QW
18	WINE OF WINE	reader is giood	AAAJI37DFG378I
19	INDIA IS MY COUNTRY	faskjas	FAJI234GAJI987
20	ONE FOR ALLL	fdghi	VABI9876ERTT1
21	DATA IS DATA	dvdasaewb	NIKI12NIKI1231
22	INDIA A MASSIVE POWER	NA	MIKISDF1232XAD
23	FRIENDS HELP	NA	GHAU8768QWER
24	MOTIVATIONAL	NA	NIKI12NIKI1231
25	DARK KNIGHT	0.888888888	MIKISDF1232XAD
26	HELL OF HELL	9.666666667	GHAU8768QWER
27	DEAA	helpful	BA2345WERSDFRER
28	DEAR TO ALL	staff is good	CADERVG3450EREW
29	MOTHER	good buy for money	ASDF345671QWERT
30	WINE OF WINE	money is dis	MNHJ879WERFGH
31	INDIA IS MY COUNTRY	stare the raeder	1234ZSDFREQWEG
32	ONE FOR ALLL	reader is giood	9345GHIKLFLLD
33	helpful	faskjas	NaN

	Summary	reviewerText	reviewerID
34	staff is good	fdghi	HGKIUYE345SFDG
35	good buy for money	8.78E+21	NIKRT2398701222
36	money is dis	67667	MAHI2348900122
37	stare the raeder	3434567	LIHJKD87634O21
38	reader is glood	098O	SISISWER4657893
39	faskjas	GOOD LUCK	8786112GHAUI
40	HELL OF HELL	BYE MY LOVE	67FDASAA123
41	DEAA	LOVE U	ASWER1234
42	DEAR TO ALL	HELLO DEAR	QWCVFED123
43	MOTHER	DAER PLEASE READ BOOK	123LKIUHNFG
44	WINE OF WINE	BOOK	987DFHGGJI657



In [18]:

```
#merge  
merging = pd.concat([df3,df4],sort=False)  
merging
```

Out[18]:

	asin	helpful	overallRating	Summary	reviewerText	reviewerID
0	23456.0	2.0	4.0	NaN	NaN	NaN
1	879877.0	3.0	3.0	NaN	NaN	NaN
2	23456.0	3.0	2.5	NaN	NaN	NaN
3	67578.0	3.0	2.2	NaN	NaN	NaN
4	123.0	3.0	3.0	NaN	NaN	NaN
5	236789.0	3.0	3.0	NaN	NaN	NaN
6	NaN	3.0	3.0	NaN	NaN	NaN
7	224442.0	3.0	3.0	NaN	NaN	NaN
8	98767.0	2.0	3.0	NaN	NaN	NaN
9	78761.0	2.0	5.0	NaN	NaN	NaN
10	321456.0	2.0	4.0	NaN	NaN	NaN
11	876231.0	2.0	4.0	NaN	NaN	NaN
12	987.0	2.0	4.0	NaN	NaN	NaN
13	234.0	2.0	3.0	NaN	NaN	NaN
14	124.0	2.0	3.0	NaN	NaN	NaN
15	156.0	2.0	3.0	NaN	NaN	NaN
16	876.0	2.0	3.0	NaN	NaN	NaN
17	123.0	2.0	2.0	NaN	NaN	NaN
18	987.0	2.0	2.0	NaN	NaN	NaN
19	22456.0	2.0	2.0	NaN	NaN	NaN
20	9876.0	2.0	2.0	NaN	NaN	NaN
21	3451.0	2.0	4.0	NaN	NaN	NaN
22	879.0	2.0	4.0	NaN	NaN	NaN
23	321.0	2.0	4.0	NaN	NaN	NaN
24	2.0	3.0	4.0	NaN	NaN	NaN
25	1.0	3.0	4.0	NaN	NaN	NaN
26	23.0	3.0	4.0	NaN	NaN	NaN
27	11.0	3.0	4.0	NaN	NaN	NaN
28	11.0	3.0	4.0	NaN	NaN	NaN
29	23.0	3.0	4.0	NaN	NaN	NaN
...	...	...	...	...	...	...
15	NaN	NaN	NaN	DEAA	good buy for money	LIHJKD87634O21
16	NaN	NaN	NaN	DEAR TO ALL	money is dis	SISISWER4657893
17	NaN	NaN	NaN	MOTHER	stare the raeder	MIKI23412341QW
18	NaN	NaN	NaN	WINE OF WINE	reader is good	AAAJI37DFG378I

	asin	helpful	overallRating	Summary	reviewerText	reviewerID
19	NaN	NaN	NaN	INDIA IS MY COUNTRY	faskjas	FAJI234GAJI987
20	NaN	NaN	NaN	ONE FOR ALLL	fdghi	VABI9876ERTT1
21	NaN	NaN	NaN	DATA IS DATA	dvdasaewb	NIKI12NIKI1231
22	NaN	NaN	NaN	INDIA A MASSIVE POWER	NA	MIKISDF1232XAD
23	NaN	NaN	NaN	FRIENDS HELP	NA	GHAU8768QWER
24	NaN	NaN	NaN	MOTIVATIONAL	NA	NIKI12NIKI1231
25	NaN	NaN	NaN	DARK KNIGHT	0.888888888	MIKISDF1232XAD
26	NaN	NaN	NaN	HELL OF HELL	9.666666667	GHAU8768QWER
27	NaN	NaN	NaN	DEAA	helpful	BA2345WERSDFRER
28	NaN	NaN	NaN	DEAR TO ALL	staff is good	CADERVG3450EREW
29	NaN	NaN	NaN	MOTHER	good buy for money	ASDF345671QWERT
30	NaN	NaN	NaN	WINE OF WINE	money is dis	MNHJ879WERFGH
31	NaN	NaN	NaN	INDIA IS MY COUNTRY	stare the raeder	1234ZSDFREQWEG
32	NaN	NaN	NaN	ONE FOR ALLL	reader is glood	9345GHIKLFLLD
33	NaN	NaN	NaN	helpful	faskjas	NaN
34	NaN	NaN	NaN	staff is good	fdghi	HGKIUYE345SFDG
35	NaN	NaN	NaN	good buy for money	8.78E+21	NIKRT2398701222
36	NaN	NaN	NaN	money is dis	67667	MAHI2348900122
37	NaN	NaN	NaN	stare the raeder	3434567	LIHJKD87634O21
38	NaN	NaN	NaN	reader is glood	098O	SISISWER4657893
39	NaN	NaN	NaN	faskjas	GOOD LUCK	8786112GHAUI
40	NaN	NaN	NaN	HELL OF HELL	BYE MY LOVE	67FDASAA123
41	NaN	NaN	NaN	DEAA	LOVE U	ASWER1234
42	NaN	NaN	NaN	DEAR TO ALL	HELLO DEAR	QWCVFED123
43	NaN	NaN	NaN	MOTHER	DAER PLEASE READ BOOK	123LKIUHNFG
44	NaN	NaN	NaN	WINE OF WINE	BOOK	987DFHGI657

90 rows × 6 columns

In [10]:

```
#sort
sort_values=df.sort_values('overallRating',ascending=False)
sort_values
```

Out[10]:

	reviewerID	asin	reviewername	helpful	reviewerText	overallRating	Summary	Un
9	1234ZSDFREQWEG	78761.0	SATISH	2	many errors	5.0	DATA IS DATA	
0	A2SUAM1J3GNN3B	23456.0	SNEHA	2	I bought this for my husband who plays the pia...	4.0	SUBWAY SURF	
21	NIKI12NIKI1231	3451.0	SHREYA	2	dvdasaweb	4.0	DATA IS DATA	
43	123LKIUHNFG	9800.0	MAHI	2	DAER PLEASE READ BOOK	4.0	MOTHER	
42	QWCVFED123	11111.0	DHONI	2	HELLO DEAR	4.0	DEAR TO ALL	

In [11]:

```
#transpose
df.transpose()
```

Out[11]:

	0	1	2	3
reviewerID	A2SUAM1J3GNN3B	A2SUAN1J3GNN0C	A2SUAN1J3GNN7B	A5SUAN1J3GNN9E
asin	23456	879877	23456	67571
reviewername	SNEHA	KARTHI	SHASHANK	NITIN
helpful	2	3	3	3
reviewerText	I bought this for my husband who plays the pia...	good book to read	high cost	good useof boo
overallRating	4	3	2.5	2.0
Summary	SUBWAY SURF	PILOT	DRIVE IN HEAVEN	SPORTS HEAVEN
UnixreviewTime	NaN	NaN	NaN	NaN
reviewTime	NaN	NaN	NaN	NaN

9 rows × 45 columns

In [12]:

```
df.shape
```

Out[12]:

```
(45, 9)
```

In [13]:

```
#reshape  
pivot_table = pd.pivot_table(df, index=['helpful', 'overallRating'], values='asin')  
pivot_table
```

Out[13]:

		asin
helpful	overallRating	
2	2.0	8360.500
	3.0	20031.400
	4.0	124891.400
	5.0	78761.000
3	2.0	5814.250
	2.2	67578.000
	2.5	23456.000
	3.0	174749.625
	4.0	2625.000