

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import train_test_split
```

```
df=pd.read_csv("/content/Amazonbookreviews.csv")
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 45 entries, 0 to 44
Data columns (total 9 columns):
#   Column                Non-Null Count  Dtype
---  -
0   reviewerID            43 non-null    object
1   asin                  43 non-null    float64
2   reviewername          44 non-null    object
3   helpful               45 non-null    int64
4   reviewerText          45 non-null    object
5   overallRating         45 non-null    float64
6   Summary               45 non-null    object
7   UnixreviewTime        0 non-null     float64
8   reviewTime            0 non-null     float64
dtypes: float64(4), int64(1), object(4)
memory usage: 3.3+ KB
```

```
df
```

	reviewerID	asin	reviewername	helpful	reviewerText	overallRating	Summary	Ur
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	
5	BA2345WERSDFRER	236789.0	SHALINI	3	im not happy with the delay of product	3.0	GODS GIFT	
6	CADERVG3450EREW	NaN	SNEHAL	3	delay in book	3.0	GODESS LIFE	
7	ASDF345671QWERT	224442.0	SHIKHA	3	Wonderful book	3.0	LIFE OF PIE	
8	MNHJ879WERFGH	98767.0	AKSHAY	2	The music is at times hard to read because we ...	3.0	INDIA A MASSIVE POWER	
9	1234ZSDFREQWEG	78761.0	SATISH	2	many errors	5.0	DATA IS DATA	
10	9345GHIKLFLLED	321456.0	RAJU	2	no mini	4.0	INDIA A MASSIVE POWER	
11	NaN	876231.0	ROHIT	2	high vat	4.0	FRIENDS HELP	
12	HGKIUYE345SFDG	987.0	MOHIT	2	the dealer is charged more	4.0	MOTIVATIONAL	

```
df.describe()
```

	asin	helpful	overallRating	UnixreviewTime	reviewTime
count	43.000000	45.000000	45.000000	0.0	0.0
mean	69702.023256	2.533333	3.282222	NaN	NaN
std	193175.237815	0.504525	0.816113	NaN	NaN
min	1.000000	2.000000	2.000000	NaN	NaN
25%	140.000000	2.000000	3.000000	NaN	NaN
50%	2121.000000	3.000000	3.000000	NaN	NaN
75%	22350.500000	3.000000	4.000000	NaN	NaN
max	879877.000000	3.000000	5.000000	NaN	NaN

```
df1=df[['reviewerID','asin','reviewername']].loc[0:14]
df1
```

```
reviewerID    asin  reviewername
0    A2SUAM1J3GNN3B  23456.0    SNEHA
1    A2SUAM1J3GNN3B  879877.0    KARTHIK
df2=df[['reviewerID','asin','reviewername']].loc[15:30]
df2
```

	reviewerID	asin	reviewername
15	LIHJKD87634O21	156.0	SONAM
16	SISISWER4657893	876.0	RAJESH
17	MIKI23412341QW	123.0	SALMAN
18	AAAJI37DFG378I	987.0	PANKAJ
19	FAJI234GAJI987	22456.0	DIPTI
20	VABI9876ERTT1	9876.0	MANISH
21	NIKI12NIKI1231	3451.0	SHREYA
22	MIKISDF1232XAD	879.0	AISHWARYA
23	GHAU8768QWER	321.0	SUMEET
24	NIKI12NIKI1231	2.0	PRAJWAL
25	MIKISDF1232XAD	1.0	JACOB
26	GHAU8768QWER	23.0	MARYLS
27	BA2345WERSDFRER	11.0	CHRIS
28	CADERVG3450EREW	11.0	MORRIS
29	ASDF345671QWERT	23.0	HARRY
30	MNHJ879WERFGH	21221.0	JOHN

```
df3=df[['reviewerID','asin','reviewername']].loc[31:40]
df3
```

	reviewerID	asin	reviewername
31	1234ZSDFREQWEG	21.0	VARSHA
32	9345GHIKLFLLLED	12212.0	SHRADDHA
33	NaN	2121.0	SAIF
34	HGKIUYE345SFDG	21212.0	KRISSH
35	NIKRT2398701222	21221.0	NEIL
36	MAHI2348900122	22245.0	KANGANA
37	LIHJKD87634O21	678.0	KHANN
38	SISISWER4657893	111.0	RIYAA
39	8786112GHAUI	223.0	NaN
40	67FDASAA123	2312.0	NANAJJI

```
mer=pd.concat([df1,df2,df3])
mer
```

	reviewerID	asin	reviewername
0	A2SUAM1J3GNN3B	23456.0	SNEHA
1	A2SUAN1J3GNN0C	879877.0	KARTHII
2	A2SUAN1J3GNN7B	23456.0	SHASHANK
3	A5SUAN1J3GNN9B	67578.0	NITIN
4	A1SUAM1J3GNN3B	123.0	PRANOTI
5	BA2345WERSDFRER	236789.0	SHALINI
6	CADERVG3450EREW	NaN	SNEHAL
7	ASDF345671QWERT	224442.0	SHIKHA
8	MNHJ879WERFGH	98767.0	AKSHAY
9	1234ZSDFREQWEG	78761.0	SATISH
10	9345GHIKLFILLED	321456.0	RAJU
11	NaN	876231.0	ROHIT
12	HGKIUYE345SFDG	987.0	MOHIT
13	NIKRT2398701222	234.0	NILESH
14	MAHI2348900122	124.0	KAILASH
15	LIHJKD87634O21	156.0	SONAM
16	SISISWER4657893	876.0	RAJESH
17	MIKI23412341QW	123.0	SALMAN
18	AAAJI37DFG378I	987.0	PANKAJ
19	FAJI234GAJI987	22456.0	DIPTI
20	VABI9876ERTT1	9876.0	MANISH
21	NIKI12NIKI1231	3451.0	SHREYA
22	MIKISDF1232XAD	879.0	AISHWARYA
23	GHAU8768QWER	321.0	SUMEET
24	NIKI12NIKI1231	2.0	PRAJWAL
25	MIKISDF1232XAD	1.0	JACOB
26	GHAU8768QWER	23.0	MARYLS
27	BA2345WERSDFRER	11.0	CHRIS
28	CADERVG3450EREW	11.0	MORRIS
29	ASDF345671QWERT	23.0	HARRY

```
sv=df.sort_values('asin')
sv
```

	reviewerID	asin	reviewername	helpful	reviewerText	overallRat
25	MIKISDF1232XAD	1.0	JACOB	3	0.888888888	
24	NIKI12NIKI1231	2.0	PRAJWAL	3	NA	
28	CADERVG3450EREW	11.0	MORRIS	3	staff is good	
27	BA2345WERSDFRER	11.0	CHRIS	3	helpful	
31	1234ZSDFREQWEG	21.0	VARSHA	3	stare the raeder	
29	ASDF345671QWERT	23.0	HARRY	3	good buy for money	
26	GHAU8768QWER	23.0	MARYLS	3	9.666666667	
38	SISISWER4657893	111.0	RIYAA	3	0980	
4	A1SUAM1J3GNN3B	123.0	PRANOTI	3	I m happy with the book	
17	MIKI23412341QW	123.0	SALMAN	2	stare the raeder	
14	MAHI2348900122	124.0	KAILASH	2	staff is good	
15	LIHJKD87634O21	156.0	SONAM	2	good buy for money	
39	8786112GHAUI	223.0	NaN	3	GOOD LUCK	
13	NIKRT2398701222	234.0	NILESH	2	helpful	
23	GHAU8768QWER	321.0	SUMEET	2	NA	
37	LIHJKD87634O21	678.0	KHANN	3	3434567	
16	SISISWER4657893	876.0	RAJESH	2	money is dis	
22	MIKISDF1232XAD	879.0	AISHWARYA	2	NA	
18	AAAJI37DFG378I	987.0	PANKAJ	2	reader is giood	
12	HGKIUYE345SFDG	987.0	MOHIT	2	the dealer is charged more money	
41	ASWER1234	1222.0	PREETI	2	LOVE U	
33	NaN	2121.0	SAIF	3	faskjas	
40	67FDASAA123	2312.0	NANAJJI	3	BYE MY LOVE	
21	NIKI12NIKI1231	3451.0	SHREYA	2	dvdasaewb	
43	123LKIUHNFG	9800.0	MAHI	2	DAER PLEASE READ BOOK	
20	VABI9876ERTT1	9876.0	MANISH	2	fdghi	
42	QWCVFED123	11111.0	DHONI	2	HELLO DEAR	
32	9345GHIKLFLED	12212.0	SHRADDHA	3	reader is giood	

```
df.transpose()
```

	0	1	2	3	
reviewerID	A2SUAM1J3GNN3B	A2SUAN1J3GNN0C	A2SUAN1J3GNN7B	A5SUAN1J3GNN9B	A1SUAM1J3GNN3B
asin	23456.0	879877.0	23456.0	67578.0	
reviewername	SNEHA	KARTHII	SHASHANK	NITIN	PRA

```
pt=pd.pivot_table(df,index=['reviewerID','asin'],values='helpful')
pt
```

		helpful	
reviewerID	asin		
1234ZSDFREQWEG	21.0	3	
	78761.0	2	
123LKIUHNFG	9800.0	2	
67FDASAA123	2312.0	3	
8786112GHAUI	223.0	3	
9345GHIKLFILLED	12212.0	3	
	321456.0	2	
A1SUAM1J3GNN3B	123.0	3	
A2SUAM1J3GNN3B	23456.0	2	
A2SUAN1J3GNN0C	879877.0	3	
A2SUAN1J3GNN7B	23456.0	3	
A5SUAN1J3GNN9B	67578.0	3	
AAAJI37DFG378I	987.0	2	
ASDF345671QWERT	23.0	3	
	224442.0	3	
ASWER1234	1222.0	2	
BA2345WERSDFRER	11.0	3	
	236789.0	3	
CADERVG3450EREW	11.0	3	
FAJI234GAJI987	22456.0	2	
GHAU8768QWER	23.0	3	
	321.0	2	
HGKIUYE345SFDG	987.0	2	
	21212.0	3	
LIHJKD87634O21	156.0	2	
	678.0	3	
MAHI2348900122	124.0	2	
	22245.0	3	
MIKI23412341QW	123.0	2	
MIKISDF1232XAD	1.0	3	
	879.0	2	
MNHJ879WERFGH	21221.0	3	
	98767.0	2	
NIKI12NIKI1231	2.0	3	
	3451.0	2	
NIKRT2398701222	234.0	2	
	21221.0	3	
QWCVFED123	11111.0	2	
SISISWER4657893	111.0	3	
	876.0	2	
VABI9876ERTT1	9876.0	2	

[Colab paid products](#) - [Cancel contracts here](#)

✓ 0s completed at 7:58 PM

