

Datasets

Bank account closing

metastatic cancer diagnosis

Diabetes Prediction

Higgs Boson Search

Marketing Response

Default Prediction

Customer Satisfaction

Product Recommendation

Student Performance from game play

Horse Health

Software Defects

Cirrhosis Outcomes

Mushroom Classification

Is the first row the column names?

Comma separated?

Yes

Yes

Yes

Yes

Yes

Yes

Yes

Yes

Yes

yes

Yes

yes

Yes

Yes

Yes

Yes

Yes

Yes

Yes

Yes

Yes

Yes

Yes

Yes

Yes

Yes

Number of rows	Number of columns	Size on the disk
165035	14	12 M
12906	83	11 M
601	10	20 K
250001	32	53 M
145232	1934	921 M
400000	2	29 M
76020	371	57 M
13647310	48	2.1 G
26296947	20	4.4 G
1236	29	224 K
101764	23	10 M
7906	20	716 K
8124	23	365 K

Datasets**Is the first row the column names?**

Bank account closing	Yes
metastatic cancer diagnosis	Yes
Diabetes Prediction	Yes
Higgs Boson Search	Yes
Marketing Response	Yes
Default Prediction	Yes
Customer Satisfaction	Yes
Product Recommendation	Yes
Student Performance from game play	Yes
Horse Health	Yes
Software Defects	Yes
Cirrhosis Outcomes	Yes
Mushroom Classification	Yes

Comma separated?	Number of rows	Number of columns
Yes	165035	14
Yes	12906	83
Yes	601	10
Yes	250001	32
yes	145232	1934
yes	400000	2
Yes	76020	371
Yes	13647310	48
Yes	26296947	20
Yes	1236	29
Yes	101764	23
Yes	7906	20
Yes	8124	23

Size on the disk

12 M
11 M
20 K
53 M
921 M
29 M
57 M
2.1 G
4.4 G
224 K
10 M
716 K
365 K