CSV File Processing

CSV file (Comma Separated Values)

A commonly used text format to store data.
Each column is a field and each row is a subject.
Fields are in general separated by comma.

Company	hispanic	white	black	asian	total_female	white_female	asian_female	black_female	hispanic_female	women_of_color	underrepresented_minorities_female	people_of_color	underrepresented_minorities
Nvidia	3.4	42.9	1	52.1	17.2	4.8	11.1	0.3	0.8	12.3	1.3	57.1	5
Pinterest	3.7	50.2	2.1	39.6	43.1	24.2	14.6	0.8	1.7	19	4.3	49.8	10.2
Salesforce	3.9	66.6	2.3	24.3	31.9	20.5	8.1	0.8	1.4	11.4	3.3	33.4	9.1
NetApp	4.1	63.7	3.2	27.5	24	12.7	8.5	1.1	1.3	11.3	2.8	36.3	8.9
eBay	4.3	52.7	2.2	39.4	35.9	19.6	12.5	1	2	16.3	3.7	47.3	7.9
View	4.3	56.5	15.4	20.7	17	8.5	6.5	1.3	0.4	8.5	2	43.5	22.8
Facebook	4.5	51.8	2.3	38.3	31.7	15.1	12.5	1	1.9	16.6	4.1	48.2	9.9
Twitter	4.5	58.6	2.6	31.5	35.4	20.3	10.9	0.9	1.8	15.1	4.2	41.4	9.9
Adobe	4.7	66.4	1.6	25.7	32.2	18.7	10.1	0.5	2.3	13.5	3.4	33.6	7.9
LinkedIn	4.7	53.5	2.5	36.7	40.2	22.9	13	1.1	2.2	17.3	4.3	46.5	9.8
Uber	4.8	53.3	4.8	33.7	29.5	15.1	8.8	2.5	1.8	14.4	5.6	46.7	13
Cisco	5.1	54.2	3.4	36	26.2	12.9	10	1.2	1.7	13.3	3.3	45.8	9.8
Google	5.2	56.9	2.4	33.5	29.2	14.9	10.8	1	1.7	14.3	3.5	43.1	9.7
HPE	5.7	72.7	7.2	12.8	33.1	23.2	4	3.4	1.9	9.9	5.9	27.3	14.6
MobileIro	5.7	52.4	1.8	37.7	30.8	10.9	16.6	0.8	1.6	20	3.4	47.6	9.9
Square	5.9	60.2	4.9	24.4	34.6	19	9.2	2	2.6	15.6	6.4	39.8	15.4
23andMe	6.4	62.6	1.7	23.6	50.2	28.3	13.1	1	3.7	21.9	8.8	37.4	13.8
PayPal	6.4	55.8	8.4	27.8	42.9	24	9.9	5.2	3	18.9	9	44.2	16.4
Airbnb	6.5	56.6	2.9	30.4	42.9	24.9	11.7	1.3	3.5	17.9	6.2	43.4	13
Lyft	7	64.1	6.7	18.4	42.5	27.6	6.3	3.8	3.1	14.9	8.7	35.9	17.5
HP Inc.	8	73.9	3.8	12.3	31	22	4.4	1.4	2.6	9	4.6	26.1	13.8
Intel	8.1	49.7	3.9	35.9	26	11	10.9	1	2.1	15	4	50.3	14.3
Intuit	8.4	54.2	4	30	42.9	23	12.3	2.2	3.7	20	7.7	45.8	15.8
Sanmina	11.4	57.8	6.3	23.4	39	22.8	9.4	3	3.5	16.2	6.8	42.2	18.8
Apple	12.3	56.1	9.2	19.5	30.7	15.8	5.9	3.8	4.1	14.9	9.1	43.9	24.4

How many columns are in the data in the CSV file below?

```
City,State,Mon,Tue,Wed,Thu,Fri
Chennai,Tamil Nadu,25,27,30,29,28
Madison,Wisconsin,-10,-12,-20,-3,-17
San Diego,California,10,12,8,9,11
```

A. 5 B. 6 C. 7 D. 8

One function that is widely used is CSE8ACSV.py is

`def get_csv(filename): given a csv file name, returns a list of dictionaries.`

DEMO: read in tech_diversity.csv and see the result.

What will be printed?

```
weather = [{"City":"Chennai", "Mon":25, "Tue":27, "Wed":30, "Thu":29, "Fri":28},
           {"City":"Madison", "Mon":-10, "Tue":-12, "Wed":-20, "Thu":-3, "Fri":-17},
           {"City":"San Diego", "Mon":10, "Tue":12, "Wed":8, "Thu":9, "Fri":11}]
```

```
>>> weather["City"]
```

A). ["Chennai", "Madison", "San Diego"] B). “Chennai” C). Error: D). “City”:”Chennai”

```
>>> weather[1]["City"]
```

A)’Madison’ B). 25 C). Error D). ‘Chennai’

```
>>> weather[1]["City"]["Wed"]
```

A)30 B). -20 C). 8 D). Error

If we have `for row in weather`, what is the type of each row?

A). dict B). list C). key D). value E). str

What will be printed?

```
for row in weather:
  print(row["City"][0])
```

A)	B)	C)	D)
Chennai Madison San Diego	Chennai	C M S	Error: A string cannot be indexed!

Coding challenge:
Write a function that takes in a column name and returns name of the company that has the maximum percentage. For example, find_best(‘total_female’) returns 23andMe