

# Quiz2

MSAN 694, Diane Woodbridge

## Description

Use the given *file\_name* and *app\_name* in “user\_definition.py” to complete the python code “quiz2.py”.

Using the input file, generate a pair RDD with the format:

*(The number of digits in a zip code, The number of respective occurrences)*

and print the pair RDD **ordered by key**.

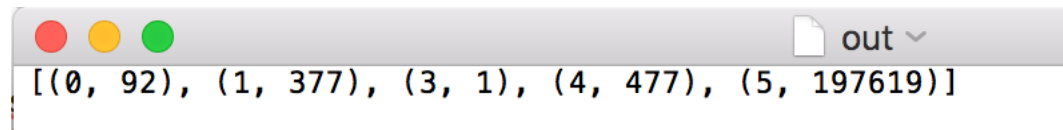
For example, when an input file has

```
9412,Tournahu George L,3301 Broderick St,San Francisco,CA
94124,Stephens Institute Inc,2225 Jerrold Ave,San Francisco,CA
9415,Stephens Institute Inc,180 New Montgomery St,San Francisco,CA
```

“quiz2.py” should **print** the output [(4, 2), (5, 1)].

Submit the quiz2.py file (**ONLY**)- the name of your file should be quiz2\_LastName\_Firstname.py on the link below.

If you run `spark-submit quiz2_Woodbridge_Diane.py > out`, the output should look like as below.



```
[(0, 92), (1, 377), (3, 1), (4, 477), (5, 197619)]
```