

## StackOverflow: step by step guide

```
'1,100,102,,2,Python',  
'2,101,,100,5,',  
'2,102,,100,9,',  
'2,103,,100,3,',  
'1,104,,2,Python',  
'2,105,,104,8,',  
'2,106,,104,5',  
...
```

↓  
map  
↓

```
(100, (1, 0, True, u'Python')),  
(100, (0, 1)),  
(100, (0, 1)),  
(100, (0, 1)),  
(104, (1, 0, False, u'Python')),  
(104, (0, 1)),  
(104, (0, 1))
```

↓  
reduceByKey  
↓

```
(100, (1, 3, True, u'Python')),  
(104, (1, 2, False, u'Python')),  
(300, (1, 0, False, u'Java')),  
(206, (1, 2, False, u'C#')),  
(400, (1, 2, True, u'PHP')),  
(200, (1, 5, False, u'C#')),  
(305, (1, 1, True, u'Java')),  
(307, (1, 2, True, u'Java')),  
(403, (1, 1, True, u'PHP')),  
(301, (1, 3, True, u'Java'))
```

↓  
map  
↓

```
((u'Python', True), 3),  
((u'Python', False), 2),  
((u'Java', False), 0),  
((u'C#', False), 2),  
((u'PHP', True), 2),  
((u'C#', False), 5),  
((u'Java', True), 1),  
((u'Java', True), 2),  
((u'PHP', True), 1),  
((u'Java', True), 3)
```

↓  
combineByKey  
↓

```
((u'C#', False), (7, 2)),  
((u'Python', False), (2, 1)),  
((u'PHP', True), (3, 2)),  
((u'Java', False), (0, 1)),  
((u'Python', True), (3, 1)),  
((u'Java', True), (6, 3))
```

→map→

```
('Java', (3, 2.0, 1, 0.0)),  
('PHP', (2, 1.5, 0, nan)),  
('Python', (1, 3.0, 1, 2.0)),  
('C#', (0, nan, 2, 3.5))
```

↑  
sortBy  
↑

```
('Java', (3, 2.0, 1, 0.0)),  
('PHP', (2, 1.5, 0, nan)),  
('Python', (1, 3.0, 1, 2.0)),  
('C#', (0, nan, 2, 3.5))
```

↑  
reduceByKey  
↑

```
(u'C#', (0, nan, 2, 3.5)),  
(u'Python', (0, nan, 1, 2.0)),  
(u'PHP', (2, 1.5, 0, nan)),  
(u'Java', (0, nan, 1, 0.0)),  
(u'Python', (1, 3.0, 0, nan)),  
(u'Java', (3, 2.0, 0, nan))
```