

## Top K keywords for category

You will be provided with a dataset like following:

```
1306,lampu depan vario 125
485,metal figure tomy
467,nicotine gum
363,hdd docking
69,hadcase samsung j7 plus
636,led rg
1819,baju tidur dewasa
321,hardisk 1 tb
667,exhaust fan maspion 301 nex
1311,spoiler helm
356,routerboard
2091,atomizer tsunami
1122,id card yoyo
1917,tas ransel tribal
24,galaxy s6 edge
958,timbangan digital kue
24,hp xioami
1765,long dress
1065,hot air gun
```

Format is:

```
|categoryId,searchKeywords
```

Task is to write a **dataflow job** in **spark** to get “**top 10 categories for a given category**”.

Output can be a text file in following format:

```
|1917|tas ransel tribal_12,tas wanita_10,tas_2
|1122|id card yoyo_5,yoyo_3,card holder_3
```

Format is:

```
|categoryId,searchKeyword1_count1,searchKeyword2_count2,...
```

Deliverables:

1. Output file
2. Runnable spark code