```
>>> df.select(split(col("email"),'@')[1].alias("domain"),"count").groupBy("domai
n").count().orderBy("count",ascending=False).show(50)
+--------------------+-----+

|              domain|count|
+--------------------+-----+
|           gmail.com| 4748|
|          redhat.com|  160|
|          apache.org|  130|
|             163.com|  110|
|             gnu.org|   86|
|         hotmail.com|   85|
|         freebsd.org|   76|
|              qq.com|   65|
|          google.com|   61|
|           yahoo.com|   56|
|           intel.com|   55|
|         outlook.com|   53|
|        mirantis.com|   35|
|          oracle.com|   29|
|    openjdk.java.net|   27|
|          debian.org|   26|
|    googlegroups.com|   25|
|          python.org|   24|
|             126.com|   23|
|           yandex.ru|   21|
|          huawei.com|   21|
|              gmx.de|   21|
|       ververica.com|   20|
|    lists.debian.org|   20|
|             gmx.net|   20|
|       linux.ibm.com|   19|
|             arm.com|   18|
|          golang.org|   16|
|     linux.intel.com|   16|
|      googlemail.com|   16|
|              web.de|   16|
|syzkaller.appspot...|   15|
|          icloud.com|   15|
|          vmware.com|   14|
|       openstack.org|   14|
|      protonmail.com|   14|
|           cisco.com|   14|
|           pobox.com|   14|
|         t-online.de|   14|
|          linaro.org|   13|
|           nokia.com|   13|
|          cygwin.com|   13|
|          kernel.org|   13|
|             hpe.com|   13|
|       canonical.com|   12|
|             free.fr|   12|
|             gmx.com|   12|
|        ericsson.com|   12|
|           posteo.de|   11|
|         comcast.net|   11|
+--------------------+-----+
only showing top 50 rows

>>> df.select(reverse(split(col("email"),'\.'))[0].alias("TLD"),"count").groupBy
```

```
("TLD").count().orderBy("count",ascending=False).show(50)
+----+-----+
| TLD|count|
+----+-----+
| com| 7427|
| org|  929|
| net|  376|
|  de|  317|
| edu|  115|
|  uk|   95|
|  io|   88|
|  fr|   72|
|  ru|   57|
|  ca|   55|
|  au|   49|
|  ch|   43|
|  nl|   42|
|  it|   39|
|  jp|   37|
|  eu|   37|
|  at|   36|
|  se|   33|
|  in|   33|
|  cn|   30|
|  us|   29|
|  cz|   27|
|  br|   26|
|  pl|   24|
|  be|   18|
|  ai|   18|
|  nz|   17|
|name|   16|
|  co|   16|
|  es|   15|
|  me|   14|
|  no|   13|
|  fi|   12|
|  dk|   11|
|  cc|   10|
| gov|   10|
|info|   10|
|  gr|    9|
|  hu|    8|
| dev|    8|
| xyz|    7|
|  sk|    7|
| biz|    7|
|  kr|    7|
|  ie|    7|
|  tr|    6|
|  fm|    6|
|  mx|    6|
|  is|    5|
|  ro|    5|
+----+-----+
only showing top 50 rows

>>> df.count()
10514
```