```
# top 50 domains statistics from email addresses
+-------------------+-----+
|             domain|count|
+-------------------+-----+
|          gmail.com| 4646|
|         redhat.com|  158|
|         apache.org|  129|
|            163.com|  110|
|        hotmail.com|   83|
|            gnu.org|   77|
|        freebsd.org|   76|
|             qq.com|   65|
|         google.com|   59|
|          intel.com|   54|
|          yahoo.com|   54|
|        outlook.com|   50|
|       mirantis.com|   35|
|         oracle.com|   28|
|   openjdk.java.net|   27|
|         python.org|   24|
|   googlegroups.com|   23|
|            126.com|   23|
|         huawei.com|   21|
|          yandex.ru|   21|
|      ververica.com|   20|
|   lists.debian.org|   20|
|             gmx.de|   19|
|      linux.ibm.com|   19|
|            gmx.net|   19|
|            arm.com|   18|
|         debian.org|   16|
|         golang.org|   16|
|    linux.intel.com|   16|
|      googlemail.com|  15|
|             web.de|   15|
|         icloud.com|   14|
|         vmware.com|   14|
|syzkaller.appspot...|   14|
|      protonmail.com|  14|
|       openstack.org|  14|
|           pobox.com|  14|
|           cisco.com|  14|
|          linaro.org|  13|
|          cygwin.com|  13|
|             hpe.com|  13|
|          kernel.org|  13|
|         t-online.de|  13|
|       canonical.com|  12|
|           nokia.com|  12|
|        ericsson.com|  12|
|             gmx.com|  11|
|       rackspace.com|  10|
|         foxmail.com|  10|
|            dell.com|  10|
+-------------------+-----+
only showing top 50 rows

# how many email addresses we have
>>> df.count()
10200
```