

Kmeans

1. Change directory

```
• text = sc.textFile('file:/Users/wangmengyuan/Desktop/rr/listings.txt')
```

2. #spark-submit --master local[*] --packages com.databricks:spark-csv_2.10:1.2.0 cluster.py

```
wangmengyuans-MacBook-Pro:rr wangmengyuan$ spark-submit --master local[*] --packages com.databricks:spark-csv_2.10:1.2.0 cluster.py
Ivy Default Cache set to: /Users/wangmengyuan/.ivy2/cache
The jars for the packages stored in: /Users/wangmengyuan/.ivy2/jars
:: loading settings :: url = jar:file:/Users/wangmengyuan/spark/lib/spark-assembly-1.6.1-hadoop1.2.1.jar!/org/apache/ivy/core/settings/ivysettings.xml
com.databricks#spark-csv_2.10 added as a dependency
:: resolving dependencies :: org.apache.spark#spark-submit-parent;1.0
  confs: [default]
  found com.databricks#spark-csv_2.10;1.2.0 in central
  found org.apache.commons#commons-csv;1.1 in central
  found com.univocity#univocity-parsers;1.5.1 in central
:: resolution report :: resolve 180ms :: artifacts dl 6ms
  :: modules in use:
    com.databricks#spark-csv_2.10;1.2.0 from central in [default]
    com.univocity#univocity-parsers;1.5.1 from central in [default]
    org.apache.commons#commons-csv;1.1 from central in [default]

-----
|      conf      | number | modules search | dwl | dded | evicted | | artifacts number | dwl | dded |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
|      default   |      3 |      0      | 0   | 0   |    0    | |      3      | 0   | 0   |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|

:: retrieving :: org.apache.spark#spark-submit-parent
  confs: [default]
  0 artifacts copied, 3 already retrieved (0KB/6ms)

+-----+
| houseid | prediction |
+-----+
| 12147973 |           6 |
| 3075044  |           8 |
| 6976     |           8 |
| 1436513  |           8 |
| 7651065  |           8 |
| 12386020 |           8 |
| 5706985  |           8 |
| 2843445  |           8 |
| 753446   |           8 |
| 849408   |           4 |
| 12023024 |           1 |
| 1668313  |           8 |
| 2684840  |           1 |
| 13547301 |           9 |
| 5434353  |           8 |
| 225979   |           8 |
| 3420384  |           8 |
| 13512930 |           1 |
| 7482195  |           1 |
| 7252607  |           1 |
+-----+
only showing top 20 rows
```